SEGUNDA EDICIÓN,
REVISADA Y COMPLETADA

DATA DRIVEN CONSTRUCTION

NAVEGANDO LA ERA DE LOS DATOS EN LA INDUSTRIA DE LA CONSTRUCCIÓN

CON APLICACIONES DE IA Y LLM

Artem Boiko

LIBROS EN OTROS IDIOMAS EN EL SITIO WEB DE DATADRIVENCONSTRUCTION

DATA-DRIVEN CONSTRUCTION

Navegando la Era de los Datos en la Industria de la Construcción

Segunda edición, revisada y completada

ARTEM BOIKO

"Boiko es el James Carville de la informática: en el tan citado "Es la economía, estúpido" de este último, sólo hay que cambiar una palabra por este famoso libro. "Son los datos, estúpido". (Y para orientarse en el universo de los datos, sigue siendo válido un dicho de los antiguos romanos que se remonta a la época griega: "Navigare necesse est". El autor guía a sus lectores por todas las profundidades y bajíos del océano de los datos con mano segura y una brújula inquebrantable, sin olvidar un exhaustivo enfoque histórico y, por último, pero no por ello menos importante, unos gráficos muy originales y un buen sentido del humor que no sólo se aprecia a segunda vista. La respuesta internacional al libro de Boiko oscila entre la aprobación eufórica y el escepticismo más bien bilioso, lo que ha sentado muy bien a la segunda edición alemana del libro. Boiko es un pensador de datos original y nada dogmático. Presenta al lector ideas apasionantes y tesis siempre valientes, incluso provocadoras, que inspiran a seguir pensando. Excelente medicina para la enfermedad alemana del consensualismo latente. Por cierto, el proverbio latino anterior tiene un complemento: "vivere non est necesse". No se aplica al planteamiento de Boiko sobre el mundo de los datos: los datos viven y su vida es necesaria, por no decir crucial."

- Dr. Burkhard Talebitari, redactor freelance - entre otros para la revista: BIM, publicada anualmente por Ernst & Sohn desde 2013.

"El libro de Artem Boiko es un hito en la democratización de la digitalización en el sector de la construcción y un auténtico revulsivo para las pequeñas y medianas empresas (PYME). Especialmente innovador: mediante el uso de modernas herramientas de código abierto de bajo y ningún código, las empresas ya pueden integrar eficazmente los datos en sus procesos empresariales y analizarlos de forma rentable, sin necesidad de conocimientos profundos de programación. Esto hace innecesario el costoso uso de engorrosos paquetes de software comerciales. Este libro es una llamada a la acción. Es una guía valiosa para cualquiera que no sólo quiera entender la transformación digital en la industria de la construcción, sino que también quiera darle forma de forma activa, pragmática, eficiente y con visión de futuro. Ahora es el momento de trabajar juntos para compartir estos conocimientos y aumentar de forma sostenible la productividad del sector de la construcción".

-Dr. Michael Max Buehler, catedrático de Gestión de la Construcción en HTWG Konstanz, copropietario de GemeinWerk Ventures y director independiente de DevvStream.

"DataDrivenConstruction" libro es uno de los primeros pasos más allá de los límites del mundo habitual de los constructores, con sus complejos sistemas de diseño y gestión, cuando, al parecer, la complejidad y la saturación de datos ni siquiera da una oportunidad para la simplificación radical y una mayor transparencia del trabajo con los datos de construcción. En su libro, Artem muestra en un lenguaje sencillo qué oportunidades abren ante nosotros las modernas tecnologías de trabajo con datos, y da literalmente pasos concretos que puedes aplicar inmediatamente en tu trabajo. Insto a todos los que quieran entender hacia dónde irán los sistemas de automatización en el sector de la construcción a que estudien detenidamente este libro para darse cuenta de que la revolución de los datos en la construcción ya está llamando a nuestra puerta. Ahora sólo interesa a los frikis, pero dentro de unos años, al igual que el BIM, estos enfoques y programas serán omnipresentes".

- Ihor Rogachew, Jefe del Centro de Competencia IMT, BIM y Transformación Digital en RGD, y Fundador de InfraBIM.Pro.

"Si "los datos son el nuevo petróleo", tenemos que aprender a definirlos, encontrarlos, extraerlos, refinarlos, para hacerlos valiosos. He encontrado el libro DataDrivenConstruction muy informativo y perspicaz. El libro proporciona un trasfondo histórico útil y explica el trabajo con datos en un lenguaje sencillo. Para quienes estén interesados en la transformación digital, ofrece una buena comprensión de los datos: cómo funcionan, cómo se estructuran y cómo pueden utilizarse

- Ralph Montague, Director de ArcDox, Director de la Cumbre de Coordinadores BIM y Presidente del Comité Espejo Nacional BIM de la Autoridad Nacional de Normalización de Irlanda.

"Recomiendo encarecidamente el libro DataDrivenConstruction que aborda, como dice el título, un enfoque de gestión de la información basada en datos para AECO. Actualmente lo estoy utilizando para ayudar a iniciar una serie de debates con diversos grupos. Me ha parecido una referencia muy accesible. Además de una visión exhaustiva del contexto histórico de las herramientas en AECO, los datos y la introducción de varias tecnologías clave, el libro contiene una serie de diagramas muy útiles que describen el alcance de las fuentes de datos y los artefactos del usuario final con flujos de trabajo de muestra. Me parece que estos son los tipos de diagramas que necesitamos más a la hora de desarrollar y supervisar las estrategias de información y contribuir a las BEP, definiendo el modelo general de datos de la empresa en el que se pueden superponer los límites de un PIM y AIM ".

- Paul Ransley, consultor principal de Acmena e ingeniero de integración de sistemas en Transport for London.

"Como se subraya en el libro, la información es un activo crucial para el sector de la construcción y disponer de ella en formatos accesibles facilita enormemente la toma de decisiones acertadas y agiliza los plazos de los proyectos. El libro ofrece un enfoque neutral y eficaz para acceder a esta fuente y aprovecharla en la toma de decisiones. La metodología presentada en el libro aprovecha un enfoque contemporáneo que combina la programación basada en inteligencia artificial con herramientas accesibles de código abierto. Al aprovechar el poder de la IA y utilizar software de código abierto, la metodología pretende mejorar la automatización, optimizar los procesos y promover la accesibilidad y la colaboración en este campo. El lenguaje del libro es claro y fácil de seguir".

- Dr. Salih Ofluoğlu, Decano de la Facultad de Bellas Artes y Arquitectura de la Universidad Bilim de Antalya y organizador del Foro Euroasiático BIM.

El libro "DATA DRIVEN CONSTRUCTION" cambia las reglas del juego para cualquiera que sienta curiosidad por saber hacia dónde se dirige el sector de la construcción en la era de los datos. Artem no se limita a rascar la superficie, sino que profundiza en los avances actuales, los retos y las oportunidades prometedoras de la construcción. Lo que distingue a este libro es su accesibilidad: Artem explica ideas complejas con analogías que facilitan la comprensión del contenido. El libro me ha parecido increíblemente informativo y atractivo. En resumen, Artem ha creado un valioso recurso que no sólo informa, sino que también inspira. Tanto si eres un profesional experimentado como un recién llegado a la construcción, este libro ampliará tu perspectiva y profundizará tu comprensión de hacia dónde se dirige la industria. Muy recomendable".

- Moayad Saleh, arquitecto y director de implantación de BIM en TMM GROUP Gesamtplanungs GmbH.

"Para cualquiera que trabaje en el sector de la construcción, desde novatos hasta profesionales experimentados, este libro es un revulsivo. No es la típica lectura polvorienta, sino que está repleto de ideas, estrategias y un toque de humor para mantener el interés. Desde los antiguos métodos de registro de datos hasta las tecnologías digitales más avanzadas, abarca la evolución del uso de datos en la construcción. Es como viajar en una máquina del tiempo a través de la evolución de los datos de la construcción. Tanto si es arquitecto, ingeniero, gestor de proyectos o analista de datos, esta completa guía revolucionará su forma de abordar los proyectos. Prepárese para optimizar procesos, mejorar la toma de decisiones y gestionar proyectos como nunca antes".

- Pierpaolo Vergati, profesor de la Universidad Sapienza de Roma y director de proyectos de construcción en Fintecna.

"Leí el libro de un tirón, en menos de 6 horas. La calidad de fabricación del libro es excelente, papel satinado denso, esquemas de colores, un tipo de letra agradable. La gran cantidad de ejemplos prácticos sobre cómo trabajar con LLM específicos del sector de la construcción le ahorrará meses, si no años, de autoestudio. Los ejemplos de trabajo son muy variados, desde los más sencillos a los más complejos, sin necesidad de adquirir complejos y costosos programas informáticos. El libro permitirá a los propietarios de cualquier empresa del sector de la construcción echar un nuevo vistazo a su estrategia empresarial, su digitalización y sus perspectivas de desarrollo. Y a las empresas más pequeñas, aumentar su eficiencia con herramientas asequibles y gratuitas".

- Mikhail Kosarev, profesor y consultor sobre transformación digital en el sector de la construcción en TIM-ASG.

"Debo decir que Data-Driven Construction es digno de ser enseñado como libro de texto en las universidades y es un libro que hará valiosas contribuciones al campo en desarrollo de BIM. El libro Data-Driven Construction contiene un glosario técnico que explica muy bien los conceptos. Temas que son extremadamente difíciles de explicar se hacen sencillos y comprensibles con un lenguaje visual muy bonito. Creo que lo que se pretende explicar en los visuales debe expresarse al lector, aunque sea brevemente. La comprensibilidad de algunos visuales, es decir, la lectura del visual requiere información aparte. También me gustaría decir que me complace presentar la valiosa obra de Artem Boiko en mis conferencias y seminarios en universidades".

- Dr. Ediz Yazicioglu, propietario de ArchCube y profesor de Gestión de Proyectos de Construcción en el Departamento de Arquitectura de la Universidad Técnica de Estambul y en la Universidad Medipol.

"Data-Driven Construction", de Artem Boiko, es una obra impresionante que ofrece una base sólida para el sector de la construcción en tiempos de tecnologías y posibilidades de información en constante crecimiento. Boiko consigue presentar temas complejos de forma comprensible al tiempo que introduce ideas visionarias. El libro es un compendio bien pensado que no sólo destaca los avances actuales, sino que también ofrece una perspectiva de las innovaciones futuras. Resulta muy recomendable para cualquiera que quiera familiarizarse con la planificación y ejecución de la construcción basada en datos".

- Markus Eiberger, profesor de la Universidad de Ciencias Aplicadas de Stuttgart, director de proyectos senior y subdirector de sucursal de Konstruktionsgruppe Bauen, miembro de la Junta Directiva de la Asociación BIM Cluster Baden-Württemberg.

"Todo lo que puedo decir es: ¡WOW! La forma en que incorporaste la historia, el LLM, los gráficos y la facilidad general para entender tus puntos es realmente notable. La fluidez del libro es asombrosa. Hay tantos aspectos brillantes en este libro que realmente cambia las reglas del juego. Es una gran fuente de información, y te felicito por el esfuerzo y la pasión que has puesto en él. Enhorabuena por haber creado una obra tan extraordinaria. Podría seguir, pero basta con decir que estoy increíblemente impresionado".

- Natasha Prinsloo, Digital Practice Lead EnergyLab_.

"Data Driven Construction transmite vívidamente los fundamentos del trabajo basado en la información con datos de construcción. Un libro que aborda los flujos de información y los conceptos económicos fundamentales y que, por tanto, se diferencia de otros libros de BIM, ya que no sólo representa la perspectiva de un fabricante de software, sino que también intenta transmitir conceptos fundamentales. Un libro que merece la pena leer y ver".

- Jakob Hirn, director general y cofundador de Build Informed GmbH, e iniciador del foro de innovación "On Top With BIM ".

Los datos son el nuevo petróleo", como suele decirse, por lo que sus prospectores o mineros deben contar con las herramientas y la mentalidad adecuadas para extraer valor de este recurso del siglo XXI. El sector de la construcción lleva demasiado tiempo en la resbaladiza pendiente de los procesos basados en "información 3D", en los que la entrega del proyecto se basa en la información horneada de otra persona (por ejemplo, ya han trazado el gráfico circular o de barras), mientras que los "datos" subyacentes (por ejemplo, la hoja de cálculo sin procesar) son capaces de ofrecer mucho más, especialmente porque la fusión de múltiples datos y la IA aportan un potencial ilimitado. Si se dedica a la construcción (o a la enseñanza o la investigación), este libro es su mejor -y hasta ahora único- recurso para navegar por el mundo de los datos en el que nos encontramos.

- Dr. Zulfikar Adamu, Profesor Asociado de TI Estratégica en la Construcción en la LSBU, Reino Unido.

Segunda edición, abril de 2025.
2025 | Artem Boiko | Karlsruhe

ISBN 978-3-9827303-4-9

REGISTRIERTES URHEBERWERK
ID-CHECK D9A2A2C55EAC5
WWW.PRIORMART.COM

REGISTRIERTES URHEBERWERK
ID-CHECK 8ECEC0192F499
WWW.PRIORMART.COM

Artem Boiko Derechos de autor

boikoartem@gmail.com
info@datadrivenconstruction.io

Ninguna parte de este libro puede ser reproducida o transmitida de ninguna forma o por ningún medio, electrónico o mecánico, incluyendo fotocopia, grabación o cualquier sistema de almacenamiento y recuperación de información, sin el permiso escrito del autor - excepto para distribución no comercial en forma inalterada. El libro se distribuye gratuitamente y puede compartirse libremente con otros usuarios con fines personales, educativos o de investigación, siempre que se conserven la autoría y las referencias al original. El autor conserva todos los derechos no patrimoniales sobre el texto y no ofrece garantías expresas ni implícitas. Las empresas, productos y nombres mencionados en el libro pueden ser ficticios o utilizarse a modo de ejemplo. El autor no se hace responsable de las consecuencias del uso de la información facilitada. La información contenida en el libro se proporciona "tal cual", sin garantía de integridad o pertinencia. El autor no se hace responsable de los daños incidentales o consecuentes derivados del uso de la información, el código o los programas contenidos en este libro. Los ejemplos de código presentados en este libro tienen únicamente fines educativos. Los lectores los utilizan bajo su propia responsabilidad. El autor recomienda probar todas las soluciones de software antes de utilizarlas en un entorno de producción. Todas las marcas comerciales y nombres de productos mencionados en el texto son marcas comerciales, marcas registradas o marcas de servicio de sus respectivas empresas y son propiedad de sus respectivos dueños. El uso de estos nombres en el libro no implica relación alguna con sus propietarios ni la aprobación por parte de los mismos. La mención de productos o servicios de terceros no constituye una recomendación ni implica su aprobación. Los nombres de empresas y productos utilizados en los ejemplos pueden ser marcas registradas de sus propietarios. Los enlaces a sitios web de terceros se facilitan para mayor comodidad y no implican que el autor apruebe la información proporcionada en dichos sitios. Todas las estadísticas, citas y estudios citados estaban actualizados en el momento de su redacción. Los datos pueden cambiar con el tiempo

Este libro se distribuye bajo una licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional (CC BY-NC-ND 4.0). Puede copiarlo y distribuirlo con fines no comerciales, siempre que conserve la autoría y no realice cambios.

CC BY-NC-ND 4.0

© 2024 Artem Boiko. Primera edición.
© 2025 Artem Boiko. Segunda edición, revisada y completada.
Todos los derechos reservados

PREFACIO A LA SEGUNDA EDICIÓN

Este libro es el resultado de un animado diálogo con la comunidad profesional. Se basa en numerosos debates profesionales sobre la gestión de datos en el sector de la construcción, que tuvieron lugar en diversas plataformas profesionales y plataformas de medios sociales. Estos debates se convirtieron en la base de artículos, publicaciones y materiales visuales que han generado una amplia respuesta en la comunidad profesional. El contenido del autor atrae millones de visitas cada año en diversas plataformas e idiomas, reuniendo a profesionales del ámbito de la digitalización de la construcción.

Un año después de la publicación de la primera edición, el libro fue encargado por expertos de más de 50 países, desde Brasil y Perú hasta Mauricio y Japón. La segunda edición del libro, que ahora tiene en sus manos, ha sido revisada y ampliada a partir de las opiniones de los expertos, las críticas a la primera edición y los debates en círculos profesionales. Gracias a los comentarios recibidos, la segunda edición se ha ampliado considerablemente: se han añadido nuevos capítulos sobre las tecnologías CAD (BIM) y la creación de procesos ETL eficaces. También se ha aumentado considerablemente el número de ejemplos prácticos y estudios de casos. Especialmente valiosos son los comentarios de los líderes de la industria de la construcción, las empresas de consultoría y las principales empresas de TI, que se dirigieron al autor con preguntas sobre la digitalización y la interoperabilidad tanto antes como después de la publicación de la primera versión del libro. Muchos de ellos ya han aplicado los enfoques descritos en el libro o tienen previsto hacerlo en un futuro próximo

Tiene en sus manos un libro creado mediante el debate y el intercambio activo de opiniones. El progreso nace en el diálogo, en el choque de opiniones y en la apertura a nuevos planteamientos. Gracias por participar en este diálogo. Su crítica constructiva es la base de futuras mejoras. Si hay errores en el texto o si desea compartir ideas y sugerencias, cualquier comentario será bienvenido. Los datos de contacto figuran al final del libro

¿POR QUÉ ES GRATUITO EL LIBRO?

Este libro se concibió como un recurso educativo abierto destinado a difundir enfoques modernos de la gestión de datos en el sector de la construcción. La primera versión del libro sirvió de base para recoger comentarios y sugerencias de la comunidad profesional, lo que permitió mejorar la estructura y el contenido del material. Todos los comentarios, sugerencias e ideas se han analizado cuidadosamente y se han incorporado a esta versión revisada. El objetivo del libro es ayudar a los profesionales de la construcción a comprender la importancia de trabajar con datos: de forma sistemática, consciente y con la vista puesta en el valor a largo plazo de la información. El autor ha recopilado ejemplos, ilustraciones y observaciones prácticas de más de 10 años de trabajo en la digitalización de la construcción. La mayor parte de este material procede de proyectos reales, conversaciones con ingenieros y promotores, participación en iniciativas internacionales y seminarios de formación. El libro es un intento de estructurar la experiencia acumulada y compartirla de forma accesible. Si desea apoyar la difusión de las ideas del libro y obtener un formato cómodo para leer, trabajar con ejemplos y materiales visuales, puede adquirir una versión impresa.

DERECHOS DE USO

Todos los materiales, ilustraciones y fragmentos de este libro pueden reproducirse, citarse o utilizarse en cualquier formato y soporte siempre que se cite como fuente a Artem Boiko y el título del libro sea "Data-Driven Construction". Gracias por el respeto al trabajo y la difusión del conocimiento.

Dedico este con sincera gratitud a mi familia, que me inculcó desde pequeño un profundo amor por la construcción, a mi pueblo minero natal, que me dio lecciones de resistencia, y a mi esposa, topógrafa, cuyo apoyo inquebrantable ha sido mi inspiración constante.

A QUIÉN VA DIRIGIDO ESTE LIBRO

Escrito en un lenguaje accesible, este libro está dirigido a un amplio abanico de lectores del sector de la construcción: desde estudiantes y novatos que desean comprender los fundamentos de los procesos de construcción modernos hasta profesionales que necesitan una metodología actualizada para gestionar los datos en la construcción. Tanto si es usted arquitecto, ingeniero, capataz, jefe de obra o analista de datos, esta completa guía con numerosas ilustraciones y gráficos únicos le ofrece valiosos conocimientos sobre cómo utilizar los datos en la empresa para optimizar y automatizar procesos, mejorar la toma de decisiones y gestionar proyectos de construcción a distintos niveles utilizando herramientas modernas

El libro es una guía completa que combina fundamentos teóricos y recomendaciones prácticas para integrar las técnicas de gestión de datos en los procesos de construcción. El libro se centra en el uso estratégico de la información para optimizar las operaciones, automatizar los procesos, mejorar la toma de decisiones y gestionar eficazmente los proyectos mediante modernas herramientas digitales.

Este libro abarca los aspectos teóricos y prácticos del trabajo con información en el sector de la construcción. A través de ejemplos detallados, explora la metodología de la parametrización de tareas, la recopilación de requisitos, el tratamiento de datos no estructurados y multiformato y su transformación en soluciones eficaces para las empresas de construcción.

El lector recorre sucesivamente el camino desde la formación de requisitos y el desarrollo de modelos de datos básicos hasta procesos más complejos de integración de fuentes de información heterogéneas, creación de procesos ETL, construcción de pipelines de información y modelos de aprendizaje automático. El enfoque secuencial permite demostrar claramente los mecanismos de organización y automatización de los procesos empresariales y los sistemas de apoyo a la toma de decisiones en el sector de la construcción. Cada parte del libro concluye con un capítulo práctico que contiene instrucciones paso a paso que permiten la aplicación inmediata de los conocimientos adquiridos en proyectos reales.

SINOPSIS DE PARTES DEL LIBRO

Este libro se estructura en torno al concepto de transformación de datos en la cadena de valor: desde la recogida de datos y la garantía de calidad hasta el tratamiento analítico y la extracción de valiosas soluciones prácticas mediante herramientas y metodologías modernas.

Parte 1: Evolución digital en la construcción - recorre la transformación histórica de la gestión de datos desde las tablillas de arcilla hasta los modernos sistemas digitales, analizando la aparición de sistemas modulares y la creciente importancia de la digitalización de la información en el contexto de las revoluciones industriales.

Parte 2: Retos de la información para el sector de la construcción - explora los problemas de la fragmentación de datos, los "silos de información", el impacto del enfoque HiPPO en la toma de decisiones y las limitaciones de los formatos propietarios, sugiriendo la consideración del paso a la IA y los ecosistemas LLM.

Parte 3: Sistematización de datos en la construcción - forma una tipología de los datos de la construcción, describe métodos de su organización, integración con los sistemas corporativos y analiza la creación de centros de competencia para la normalización de los procesos de información.

Parte 4: Aseguramiento de la calidad de los datos - revela metodologías para convertir información dispar en datos estructurados de calidad, incluida la extracción de datos de diversas fuentes, la validación y el modelado mediante LLM.

Parte 5: Cálculos de costes y plazos: aborda la digitalización de los cálculos de costes y planificación, la automatización de la obtención de volúmenes a partir de modelos CAD (BIM), las tecnologías de modelización 4D-8D y el cálculo ESG de proyectos de construcción.

Parte 6: CAD y BIM - analiza de forma crítica la evolución de las tecnologías de diseño, los problemas de interoperabilidad de los sistemas, las tendencias hacia formatos de datos abiertos y las perspectivas de aplicación de la inteligencia artificial al diseño.

Parte 7: Análisis de datos y automatización: examina los principios de la visualización de la información, los indicadores clave de rendimiento, los procesos ETL, las herramientas de orquestación de flujos de trabajo y la aplicación de modelos lingüísticos para automatizar tareas rutinarias.

Parte 8: Almacenamiento y gestión de datos: explora los formatos de almacenamiento de datos, los conceptos de almacén de datos y lago de datos, los principios de gestión de datos y los nuevos enfoques, incluidas las bases de datos vectoriales y las metodologías DataOps y VectorOps.

Parte 9: Big Data y aprendizaje automático: se centra en la transición al análisis objetivo basado en datos históricos, el Internet de las cosas en las obras de construcción y la aplicación de algoritmos de aprendizaje automático para predecir los costes y plazos de los proyectos.

Parte 10: La industria de la construcción en la era de los datos digitales - presenta una mirada al futuro de la industria de la construcción, analizando el cambio del análisis causal al trabajo con correlaciones, el concepto de "Uberización" de la construcción y las estrategias para la transformación digital.

What is meant by **data-driven construction**?

Led by CEO and project managers

A data-driven approach to construction is the **process** of implementing **connected and automated systems** that enable companies to **continuously improve** **project delivery** and operational efficiency while **reducing costs and risks**, ultimately creating a **sustainable competitive advantage.**

Focuses on real-time analytics

It's a mindset shift!

Streamline operations & enhance collaboration

The ultimate goal

INTRODUCCIÓN

¿Cuánto tiempo puede seguir siendo competitiva su empresa en un mundo en el que la tecnología evoluciona rápidamente y todos los aspectos de la actividad empresarial, desde el cálculo de plazos y costes hasta el análisis de riesgos, se automatizan mediante modelos de aprendizaje automático?

La industria de la construcción, que existe desde hace tanto tiempo como la propia humanidad, se encuentra en la cúspide de un cambio revolucionario que promete modificar por completo nuestra concepción de la construcción tradicional. Ya en otros sectores de la economía, la digitalización no solo está cambiando las reglas, sino que está expulsando sin contemplaciones del mercado a las empresas que no han sabido adaptarse al nuevo entorno de procesamiento de datos y son incapaces de mejorar la velocidad de toma de decisiones (Fig. 1).

Fig. 1 La rapidez de la toma de decisiones en el sector de la construcción depende del factor humano con más frecuencia que en otros sectores.

La banca, el comercio minorista, la logística y la agroindustria avanzan rápidamente hacia la plena digitalización, donde las imprecisiones y las opiniones subjetivas ya no tienen cabida. Los algoritmos modernos son capaces de analizar enormes cantidades de datos y ofrecer a los clientes predicciones precisas, ya se trate de la probabilidad de devolución de un préstamo, las rutas de reparto óptimas o la previsión de riesgos.

La construcción es uno de los últimos sectores en hacer la inevitable transición de soluciones basadas en las opiniones de especialistas muy bien pagados a soluciones basadas en datos. Esta transición está impulsada no solo por las nuevas capacidades tecnológicas, sino también por la mayor exigencia de transparencia, precisión y rapidez por parte del mercado y los clientes.

Robotización, automatización de procesos, datos abiertos y previsiones basadas en ellos: todo esto ya no

son solo posibilidades, sino una fatalidad. La mayoría de las empresas del sector de la construcción, que eran hasta hace poco responsables ante el cliente de calcular el volumen, el coste, el plazo de los proyectos y el control de calidad, corren ahora el riesgo de convertirse en meras ejecutoras de pedidos, sin tomar decisiones clave (Fig. 2).

Gracias a los avances en potencia informática, algoritmos de aprendizaje automático y acceso democratizado a los datos, ahora es posible combinar automáticamente datos de distintas fuentes, lo que permite profundizar en el análisis de procesos, la predicción de riesgos y la optimización de costes en las fases de discusión de un proyecto de construcción. Estas tecnologías tienen el potencial de mejorar radicalmente la eficiencia y reducir los costes en todo el sector.

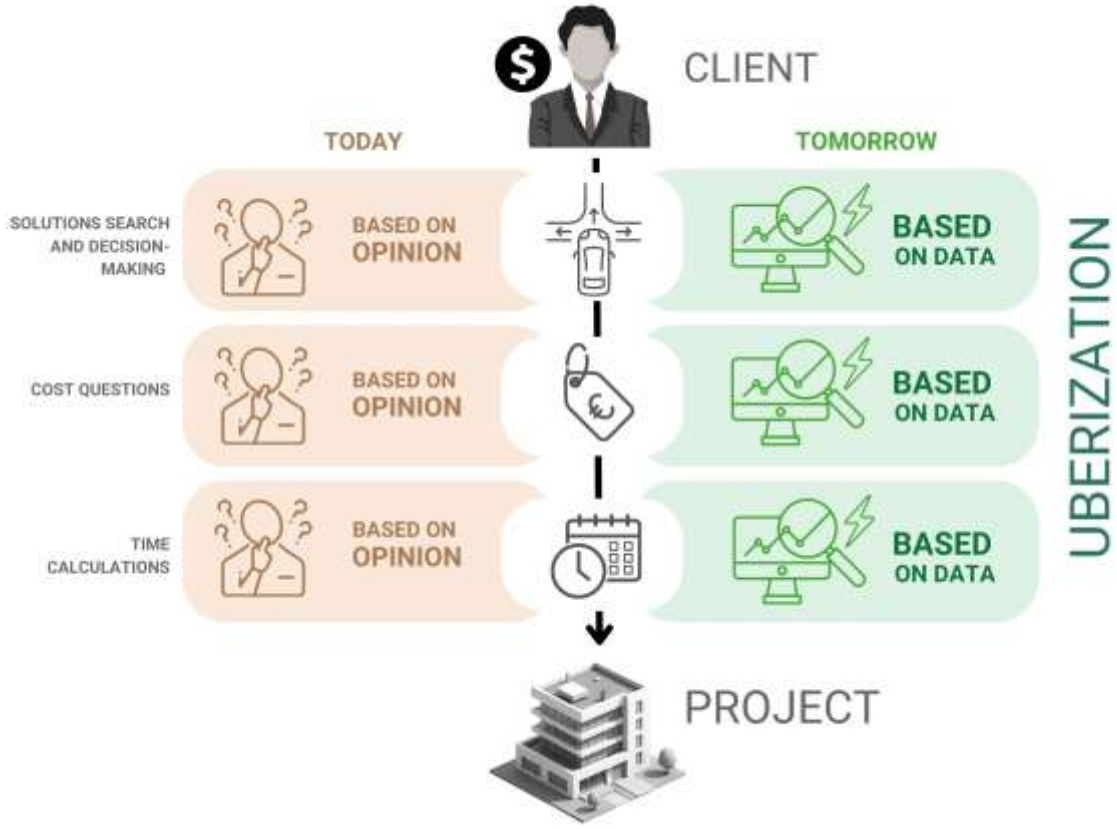


Fig. 2 Al cliente no le interesan excesivos factores humanos en el camino hacia la realización de su proyecto.

A pesar de todas las ventajas de las nuevas herramientas y conceptos, la industria de la construcción va muy por detrás de otros sectores de la economía en la adopción de nuevas tecnologías.

> Según el informe IT Metrics Key Data 2017, el sector de la construcción ocupa el último lugar en gasto en TI entre otros 19 sectores[1].

El rápido crecimiento del volumen de datos y la complejidad de los procesos se están convirtiendo en un quebradero de cabeza para la dirección de las empresas, y el principal problema a la hora de utilizar las nuevas tecnologías es que los datos, a pesar de su abundancia, siguen estando fragmentados, desestructurados y a menudo son incompatibles entre distintos sistemas y productos de software. Por eso, muchas empresas del sector de la construcción se preocupan ahora principalmente por los problemas de calidad

de los datos, que solo pueden resolverse con la implantación de sistemas de gestión y análisis eficaces y automatizados.

> Según una encuesta realizada por KPMG® entre directivos de la construcción en 2023 [2], los sistemas de información para la gestión de proyectos (PMIS), la analítica de datos avanzada y básica y el modelado de información para la construcción (BIM) son los que tienen mayor potencial para mejorar el ROI de los proyectos (Fig, 3).

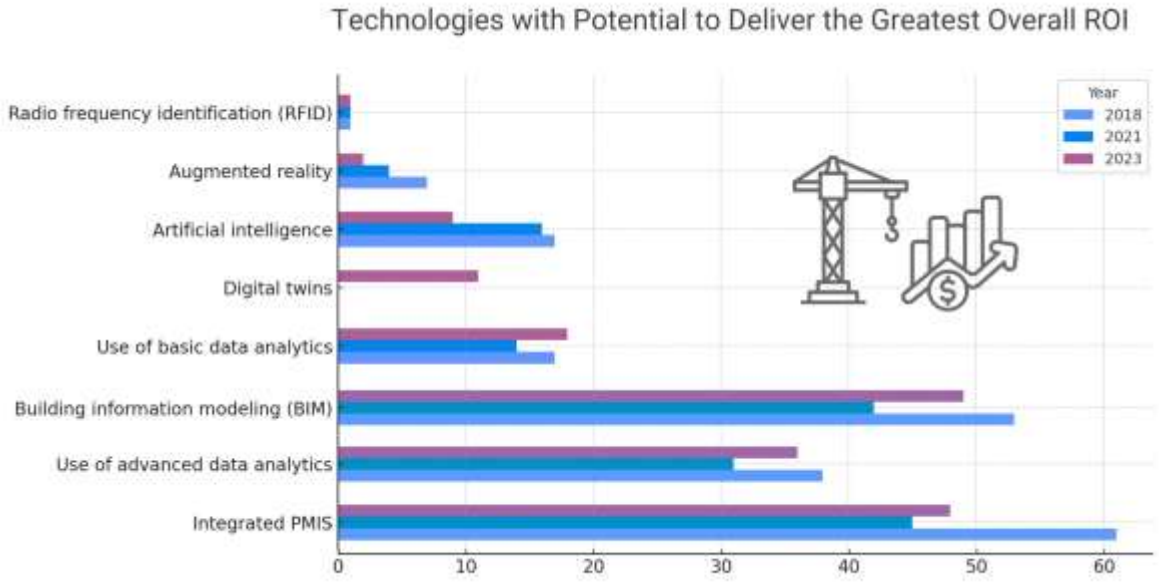


Fig. 3 Encuesta entre directivos de empresas de construcción: ¿qué tecnologías proporcionarán el mayor rendimiento de la inversión (ROI) en proyectos de capital? (basada en materiales [2]).

La solución a los retos asociados a la integración de datos en los procesos empresariales pasa por garantizar una información de alta calidad, utilizar formatos de datos adecuados y aplicar métodos eficaces de creación, almacenamiento, análisis y tratamiento de datos.

La toma de conciencia del valor de los datos está obligando a varias industrias a abandonar las aplicaciones aisladas y las complejas estructuras burocráticas de gestión. En su lugar, la atención se está desplazando hacia la creación de nuevos enfoques de la arquitectura de la información, transformando las empresas en modernas empresas impulsadas por los datos. Tarde o temprano, el propio sector de la construcción dará este paso, pasando de una evolución digital gradual a una auténtica revolución digital que afectará a todas las empresas.

> La transición a procesos empresariales basados en datos no será fácil. Muchas empresas se enfrentarán a desafíos porque los ejecutivos no siempre entienden cómo utilizar conjuntos de datos caóticos para mejorar la eficiencia y el crecimiento empresarial.

Este libro se adentra en el mundo de los datos, donde la información se está convirtiendo en un recurso estratégico clave que determina la eficacia y sostenibilidad de los procesos empresariales. Con el rápido crecimiento de la información, las empresas se enfrentan a nuevos retos. La transformación digital ya no es solo una palabra de moda: se está convirtiendo en una necesidad.

Fig. 4 Los datos y los procesos son la base de la construcción.

Entender la transformación significa ser capaz de explicar lo complejo con palabras sencillas. Por eso el libro está escrito en un lenguaje accesible y va acompañado de las ilustraciones del autor, creadas específicamente para explicar con claridad los conceptos clave. Estos diagramas, gráficos y visualizaciones están diseñados para eliminar las barreras de percepción y hacer comprensible el material incluso a quienes antes consideraban estos temas demasiado complejos. Todas las ilustraciones, diagramas y gráficos de este libro han sido creados por el autor y diseñados específicamente para visualizar los conceptos clave descritos en el texto.

> *Una imagen vale más que mil palabras [3].*
>
> - Fred R. Barnard, ilustrador inglés, 1927.

Para conectar la teoría con la práctica, utilizaremos herramientas de inteligencia artificial (en particular, modelos lingüísticos) que le permitirán desarrollar soluciones sin necesidad de profundos conocimientos de programación. Si te orientas hacia el material práctico y te interesa más el trabajo práctico con datos, puedes saltarte la primera parte introductoria e ir directamente a la segunda parte del libro, donde comienza la descripción de ejemplos y casos concretos.

Sin embargo, no ponga excesivas expectativas en las herramientas de IA (Inteligencia Artificial), aprendizaje automático y LLM (Large Language Models) en general. Sin datos de entrada de calidad y un profundo conocimiento de la materia, ni siquiera los algoritmos más avanzados pueden ofrecer resultados fiables y significativos.

> Satya Nadella, consejero delegado de Microsoft, advierte del riesgo de una burbuja en la inteligencia artificial a principios de 2025 [4], comparando el actual bombo publicitario con la burbuja de las puntocom. Nadella subraya que las afirmaciones de que se han alcanzado hitos de la AGI (Inteligencia Artificial General) sin una justificación adecuada son una "manipulación sin sentido de las métricas". Nadella cree que el verdadero éxito de la IA debe medirse por su contribución al crecimiento del PIB mundial, en lugar de hacer demasiado hincapié en las palabras de moda.

Detrás de todas las palabras de moda sobre nuevas tecnologías y conceptos se esconde un trabajo complejo y minucioso para garantizar la calidad de los datos, la parametrización de los procesos empresariales y la adaptación de las herramientas a las tareas reales.

> Un enfoque basado en datos no es un producto que se pueda descargar o comprar sin más. Es una estrategia que hay que construir. Comienza con una nueva mirada a los procesos y problemas existentes, y luego requiere un movimiento disciplinado en la dirección elegida.

Los principales desarrolladores de software y proveedores de aplicaciones no serán el motor del cambio en el sector de la construcción: para muchos de ellos, un enfoque basado en los datos supone una amenaza para su modelo de negocio establecido.

> *Otras industrias [a diferencia de la construcción], como la automovilística, ya han experimentado un cambio radical y disruptivo, y su transformación digital está muy avanzada. Las empresas de la construcción deben actuar con rapidez y decisión: las empresas ágiles cosecharán enormes recompensas, mientras que para las que duden, los riesgos serán graves. Pensemos en la convulsión que la fotografía digital ha causado en este sector [5].*
>
> - Informe del Foro Económico Mundial Shaping the Future of Construction, 2016.

Las empresas que reconozcan a tiempo las oportunidades y ventajas del nuevo enfoque obtendrán una ventaja competitiva sostenible y podrán desarrollarse y crecer sin depender de las soluciones de los grandes proveedores.

Ésta es su oportunidad no sólo de capear el temporal de la digitalización de la información que se avecina, sino de tomar las riendas. En este libro encontrará no solo un análisis del estado actual del sector, sino también recomendaciones concretas para repensar y reestructurar sus procesos y su negocio con el fin de convertirse en líder de la nueva era de la construcción y mejorar su experiencia profesional.

El futuro digital de la construcción no consiste solo en utilizar nuevas tecnologías y programas, sino en replantearse fundamentalmente el tratamiento de los datos y los modelos de negocio.

¿Está preparada su empresa para este cambio estratégico?

ÍNDICE

INTRODUCCIÓN **1**

VISIÓN GENERAL ..... **I**

I PARTE DE LAS TABLILLAS DE ARCILLA A LA REVOLUCIÓN DIGITAL: EVOLUCIÓN DE LA INFORMACIÓN EN LA CONSTRUCCIÓN **2**

CAPÍTULO 1.1. EVOLUCIÓN DE LA UTILIZACIÓN DE DATOS EN EL SECTOR DE LA CONSTRUCCIÓN 3

El nacimiento de la era de los datos en la construcción 3

De la arcilla y el papiro a la tecnología digital 4

El proceso como herramienta para una experiencia basada en datos 5

Digitalización de la información del proceso de construcción 8

CAPÍTULO 1.2. TECNOLOGÍAS Y SISTEMAS DE GESTIÓN EN LA CONSTRUCCIÓN MODERNA 11

La revolución digital y la aparición de sistemas MRP/ERP modulares 11

Sistemas de gestión de datos: de la minería de datos a los retos empresariales 13

Micelio empresarial: cómo los datos conectan con los procesos empresariales 17

CAPÍTULO 1.3. LA REVOLUCIÓN DIGITAL Y LA EXPLOSIÓN DE DATOS 20

El inicio del auge del volumen de datos como ola evolutiva 20

La cantidad de datos generados en una empresa moderna 22

El coste del almacenamiento de datos: el aspecto económico 23

Fronteras de la acumulación de datos: de la masa al significado 25

Próximos pasos: de la teoría de los datos al cambio práctico 27

PARTE II CÓMO EL SECTOR DE LA CONSTRUCCIÓN SE AHOGA EN EL CAOS DE LOS DATOS **29**

CAPÍTULO 2.1. FRAGMENTACIÓN Y SILOS DE DATOS 30

Cuantas más herramientas, ¿más eficaz será la empresa? 30

Silos de datos y su impacto en el rendimiento de la empresa 32

Duplicación y falta de calidad de los datos como consecuencia de la fragmentación 36

HiPPO o el peligro de las opiniones en la toma de decisiones 37

Aumento continuo de la complejidad y el dinamismo de los procesos empresariales 40

La Cuarta Revolución Industrial (Industria 4.0) y la Quinta Revolución Industrial (Industria 5.0) en el sector de la construcción 43

CAPÍTULO 2.2. CONVERTIR EL CAOS EN ORDEN Y REDUCIR LA COMPLEJIDAD 46

Código redundante y sistemas cerrados como barrera para mejorar la productividad 46
De los silos a un único almacén de datos 48
Los sistemas de almacenamiento integrados permiten la transición a los agentes de IA 50
De la recogida de datos a la toma de decisiones: el camino hacia la automatización 52
Próximos pasos: convertir el caos en un sistema gestionable 54
III PARTE MARCO DE DATOS EN LOS PROCESOS EMPRESARIALES DE LA CONSTRUCCIÓN 56
CAPÍTULO 3.1. TIPOS DE DATOS EN LA CONSTRUCCIÓN 57
Los tipos de datos más importantes en el sector de la construcción 57
Datos estructurados 61
RDBMS bases de datos relacionales y lenguaje de consulta SQL 63
Consultas SQL en bases de datos y nuevas tendencias 65
Datos no estructurados 67
Datos de texto: entre el caos desestructurado y la estructura 68
Datos semiestructurados y poco estructurados 69
Datos geométricos y sus aplicaciones 70
Datos CAD: del diseño al almacenamiento de datos 73
La aparición del concepto BIM (BOM) y el uso del CAD en los procesos 76
CAPÍTULO 3.2. ARMONIZACIÓN Y ESTRUCTURACIÓN DE LOS DATOS 82
Poblar los sistemas con datos en el sector de la construcción 82
Transformación de datos: la base fundamental del análisis empresarial moderno 85
Modelos de datos: relaciones en los datos y relaciones entre elementos 88
Formatos propietarios y su impacto en los procesos digitales 92
Los formatos abiertos están cambiando el enfoque de la digitalización 96
Cambio de paradigma: el código abierto como fin de la era del dominio de los proveedores de software 97
Datos abiertos estructurados: la base de la transformación digital 100
CAPÍTULO 3.3. LAS LLM Y SU PAPEL EN EL TRATAMIENTO DE DATOS Y LOS PROCESOS EMPRESARIALES 103
Chats LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok para automatizar los procesos de tratamiento de datos 103
Grandes modelos lingüísticos LLM: cómo funcionan 104
Utilización de LLM locales para datos sensibles de la empresa 107
Control total de la IA en la empresa y cómo desplegar su propio LLM 109

RAG: Asistentes LLM inteligentes con acceso a datos corporativos 111
CAPÍTULO 3.4. IDE CON SOPORTE LLM Y FUTUROS CAMBIOS DE PROGRAMACIÓN 114
Elegir una IDE: de los experimentos LLM a las soluciones empresariales..... 114
IDE con soporte LLM y futuros cambios de programación 116
Python Pandas: una herramienta indispensable para trabajar con datos..... 117
DataFrame: formato universal para datos tabulares 121
Próximos pasos: creación de un marco de datos sostenible 124
PARTE IV CALIDAD DE LOS DATOS: ORGANIZACIÓN, ESTRUCTURACIÓN, MODELIZACIÓN 126
CAPÍTULO 4.1. CONVERSIÓN DE DATOS A UN FORMATO ESTRUCTURADO 127
Aprenda a convertir documentos, PDF, imágenes y textos en formatos estructurados 127
Ejemplo de conversión de un documento PDF en una tabla 128
Convierte imágenes JPEG, PNG a formato estructurado 132
Conversión de datos de texto en formato estructurado 135
Traducción de datos CAD (BIM) a un formato estructurado..... 138
Los proveedores de soluciones CAD se pasan a los datos estructurados 143
CAPÍTULO 4.2. CLASIFICACIÓN E INTEGRACIÓN: UN LENGUAJE COMÚN PARA LOS DATOS DE LA CONSTRUCCIÓN 146
La rapidez en la toma de decisiones depende de la calidad de los datos 146
Normalización e integración de datos 147
La compatibilidad digital empieza por los requisitos 150
Un lenguaje común de construcción: el papel de los clasificadores en la transformación digital 152
Masterformat, OmniClass, Uniclass y CoClass: evolución de los sistemas de clasificación 155
CAPÍTULO 4.3. MODELIZACIÓN DE DATOS Y CENTRO DE EXCELENCIA..... 160
Modelización de datos: modelo conceptual, lógico y físico..... 160
Modelización práctica de datos en el contexto de la construcción..... 164
Creación de una base de datos con LLM..... 166
Centro de Excelencia (CoE) para la Modelización de Datos 168
CAPÍTULO 4.4. SISTEMATIZACIÓN DE REQUISITOS Y VALIDACIÓN DE LA INFORMACIÓN..... 172
Recopilación y análisis de requisitos: transformar las comunicaciones en datos estructurados..... 172
Los organigramas de procesos y la eficacia de los marcos conceptuales..... 176
Requisitos estructurados y expresiones regulares RegEx..... 178
Recogida de datos para el proceso de verificación 183

Validación de datos y resultados de la validación 185
Visualización de los resultados de las inspecciones 190
Comparación de los controles de calidad de los datos con las necesidades de la vida humana 192
Próximos pasos: convertir los datos en cálculos y planes precisos 194
V PARTE CÁLCULO DE COSTES Y PLAZOS: INCORPORACIÓN DE DATOS A LOS PROCESOS DE CONSTRUCCIÓN 196
CAPÍTULO 5.1. CÁLCULOS Y ESTIMACIONES DE COSTES PARA PROYECTOS DE CONSTRUCCIÓN 197
Conceptos básicos de la construcción: estimación de cantidades, costes y plazos 197
Métodos de cálculo del coste estimado de los proyectos 198
Método basado en los recursos para realizar estimaciones y cálculos en la construcción 199
Base de datos de recursos de la construcción: catálogo de materiales y obras de construcción 200
Cálculo del coste de las obras en función de la base de recursos 201
Cálculo del coste final del proyecto: de la estimación al presupuesto 206
CAPÍTULO 5.2. DESPEGUE DE CANTIDADES Y CREACIÓN AUTOMÁTICA DE PRESUPUESTOS Y CALENDARIOS 210
De las 3D a las 4D y 5D: volumetría y cuantificación 210
Atributos 5D y obtención de volúmenes de atributos a partir de CAD 210
QTO Quantity Take-Off: agrupación de datos de proyectos por atributos 214
Automatización de QTO mediante LLM y datos estructurados 219
Cálculo del QTO de todo el proyecto mediante reglas de grupo a partir de una hoja de cálculo Excel 223
CAPÍTULO 5.3. 4D, 6D-8D Y CÁLCULO DE LAS EMISIONES DE DIÓXIDO DE CARBONO CO_2 229
Modelo 4D: integración del tiempo en las estimaciones de construcción 229
Calendario de construcción y su automatización basada en datos de cálculo de costes 230
Capas avanzadas de atributos 6D-8D: de la eficiencia energética a la seguridad 232
Estimación del CO_2 y cálculo de las emisiones de dióxido de carbono de los proyectos de construcción 235
CAPÍTULO 5.4. SISTEMAS ERP Y PMIS PARA LA CONSTRUCCIÓN 240
Sistemas ERP para la construcción sobre el ejemplo de cálculos y estimaciones 240
PMIS: el punto intermedio entre la ERP y la obra 245
Especulación, beneficio, cierre y falta de transparencia en ERP y PMIS 246
El fin de la era de los ERP/PMIS cerrados: el sector de la construcción necesita nuevos enfoques 249
Próximos pasos: utilización eficaz de los datos del proyecto 251

PARTE VI CAD Y BIM: MARKETING, REALIDAD Y FUTURO DE LOS DATOS DE DISEÑO EN LA CONSTRUCCIÓN 254

CAPÍTULO 6.1. APARICIÓN DE LOS CONCEPTOS BIM EN EL SECTOR DE LA CONSTRUCCIÓN 255

Historia de la aparición del BIM y el BIM abierto como conceptos de marketing de los proveedores de CAD 255

La realidad del BIM: sistemas modulares cerrados en lugar de bases de datos integradas 258

La aparición del formato abierto IFC en el sector de la construcción 260

Problema de formato IFC en función del núcleo geométrico 262

La aparición del tema de la semántica y la ontología en la construcción 265

Por qué las tecnologías semánticas no cumplen las expectativas en el sector de la construcción .. 267

CAPÍTULO 6.2. FORMATOS DE PROYECTOS CERRADOS Y PROBLEMAS DE INTEROPERABILIDAD 271

Datos cerrados y caída de la productividad: el callejón sin salida de la industria CAD (BIM) 271

El mito de la interoperabilidad entre sistemas CAD 273

Transición a USD y datos granulares 277

CAPÍTULO 6.3. GEOMETRÍA EN LA CONSTRUCCIÓN: DE LAS LÍNEAS A LOS METROS CÚBICOS 281

Cuando las líneas se convierten en dinero o por qué los constructores necesitan geometría 281

De líneas a volúmenes: cómo el área y el volumen se convierten en datos 281

Pasar a MESH, USD y polígonos: utilizar el teselado para la geometría 284

LOD, LOI, LOMD: una clasificación única de los detalles en CAD (BIM) 285

Nuevas normas CAD (BIM) - AIA, BEP, IDS, LOD, COBie 288

CAPÍTULO 6.4.PARAMETRIZACIÓN DEL DISEÑO Y UTILIZACIÓN DE LLM PARA EL FUNCIONAMIENTO DE CAD 293

La ilusión de unicidad de los datos CAD (BIM): el camino hacia la analítica y los formatos abiertos 293

Diseñar a través de parámetros: el futuro de CAD y BIM 296

La aparición del LLM en los procesos de datos CAD de diseño 299

Análisis automatizado de archivos DWG con LLM y Pandas 302

Próximos pasos: pasar de los formatos cerrados a los datos abiertos 308

PARTE VII TOMA DE DECISIONES BASADA EN DATOS, ANÁLISIS, AUTOMATIZACIÓN Y APRENDIZAJE AUTOMÁTICO 311

CAPÍTULO 7.1. ANÁLISIS DE DATOS Y TOMA DE DECISIONES BASADA EN DATOS 312

Los datos como recurso en la toma de decisiones 312

Visualización de datos: la clave para comprender y tomar decisiones 316

Indicadores de rendimiento KPI y ROI 318

Cuadros de mando y cuadros de mando: visualización de métricas para una gestión eficaz.....320
El análisis de datos y el arte de hacer preguntas 321
CAPÍTULO 7.2. FLUJO DE DATOS SIN ESFUERZO MANUAL: POR QUÉ ES NECESARIA LA ETL 324
Automatización ETL: menores costes y tratamiento de datos más rápido.....324
ETL Extract: recogida de datos 328
ETL Transform: aplicación de reglas de validación y transformación 331
Carga ETL: visualización de resultados en tablas y gráficos.....333
Carga ETL: Creación automática de PDF 339
ETL Load: generación automática de documentos a partir de FPDF.....340
Carga ETL: generación de informes y carga en otros sistemas 344
ETL con LLM: Visualización de datos a partir de documentos PDF.....345
CAPÍTULO 7.3. TRANSPORTADOR ETL AUTOMÁTICO (TUBERÍA) 350
Canalización: canalización automática de datos ETL.....350
Proceso de validación de datos Pipeline-ETL con LLM 354
Pipeline-ETL: verificación de datos e información de elementos de proyecto en CAD (BIM) 356
CAPÍTULO 7.4. ORQUESTACIÓN DE ETL Y FLUJOS DE TRABAJO: SOLUCIONES PRÁCTICAS 362
DAG y Apache Airflow: automatización y orquestación de flujos de trabajo 362
Apache Airflow: una aplicación práctica sobre automatización ETL.....363
Apache NiFi para enrutamiento y transformación de datos 367
n8n Orquestación de procesos Low-Code, No-Code.....368
Próximos pasos: pasar de las operaciones manuales a las soluciones basadas en análisis 371
PARTE VIII ALMACENAMIENTO Y GESTIÓN DE DATOS EN LA CONSTRUCCIÓN 373
CAPÍTULO 8.1. INFRAESTRUCTURA DE DATOS: DE LOS FORMATOS DE ALMACENAMIENTO A REPOSITORIOS DIGITALES 374
Átomos de datos: la base de una gestión eficaz de la información 374
Almacenamiento de información: ficheros o datos 375
Almacenamiento de big data: análisis de los formatos más populares y su eficiencia 377
Optimizar el almacenamiento con Apache Parquet.....380
DWH: Almacén de datos almacenes de datos 382
Data Lake - la evolución de ETL en ELT: de la limpieza tradicional al procesamiento flexible.....384
Arquitectura Data Lakehouse: la sinergia de los almacenes y los lagos de datos 386
CDE, PMIS, ERP o DWH y Data Lake.....389

CAPÍTULO 8.2. GESTIÓN DE ALMACENES DE DATOS Y PREVENCIÓN DEL CAOS 392
Bases de datos vectoriales y Bounding Box 392
Gobernanza de datos, minimalismo de datos y pantanos de datos 395
DataOps y VectorOps: nuevas normas para trabajar con datos 398
El camino a seguir: del almacenamiento caótico al almacenamiento estructurado 400
IX PARTE BIG DATA, APRENDIZAJE AUTOMÁTICO Y PREDICCIONES 402
CAPÍTULO 9.1. BIG DATA Y SU ANÁLISIS 403
Big data en la construcción: de la intuición a la previsibilidad 403
Cuestionando la viabilidad del big data: correlación, estadística y muestreo de datos 404
Big data: análisis de los datos del millón de permisos de construcción de San Francisco 407
Ejemplo de big data basado en datos CAD (BIM) 413
IoT Internet de las cosas y contratos inteligentes 417
CAPÍTULO 9.2. APRENDIZAJE AUTOMÁTICO Y PREDICCIONES 421
El aprendizaje automático y la inteligencia artificial cambiarán nuestra forma de construir 421
De la evaluación subjetiva a la previsión estadística 424
Titanic dataset: Hello World en el mundo del análisis de datos y big data 425
Aprendizaje automático en acción: de los pasajeros del Titanic a la gestión de proyectos 430
Predicciones y previsiones basadas en datos históricos 434
Conceptos clave del aprendizaje automático 437
CAPÍTULO 9.3. PREDICCIÓN DE COSTES Y PLAZOS MEDIANTE APRENDIZAJE AUTOMÁTICO APRENDIZAJE 440
Ejemplo de uso del aprendizaje automático para determinar el coste y el calendario de un proyecto 440
Previsión de costes y plazos de proyectos mediante regresión lineal 442
Predicción de costes y plazos de proyectos mediante el algoritmo K-nearest neighbour (k-NN) 445
Próximos pasos: del almacenamiento al análisis y la previsión 449
PARTE X LA INDUSTRIA DE LA CONSTRUCCIÓN EN LA ERA DIGITAL. OPORTUNIDADES Y RETOS 452
CAPÍTULO 10.1. ESTRATEGIAS DE SUPERVIVENCIA: CREACIÓN DE VENTAJAS COMPETITIVAS 453
Correlaciones en lugar de cálculos: el futuro de la analítica de la construcción 453
Enfoque basado en datos en la construcción: un nuevo nivel de infraestructura 456
La oficina digital de nueva generación: cómo la IA está cambiando el espacio de trabajo 458
Los datos abiertos y la uberización son una amenaza para el actual negocio de la construcción 460

Los problemas no resueltos de la uberización como última oportunidad de aprovechar el tiempo para la transformación 463

CAPÍTULO 10.2. GUÍA PRÁCTICA PARA APLICAR UN ENFOQUE BASADO EN DATOS 468

De la teoría a la práctica: hoja de ruta para la transformación digital en la construcción 468

Sentar las bases digitales: 1-5 pasos hacia la madurez digital 470

Liberar el potencial de los datos: 5-10 pasos hacia la madurez digital 475

Hoja de ruta de la transformación: del caos a una empresa basada en datos 482

Industria 5.0 de la construcción: cómo ganar dinero cuando ya no puedes esconderte 485

CONCLUSIÓN 487

SOBRE EL AUTOR 490

RELACIÓN ULTERIOR 491

COMENTARIO DE TRADUCCIÓN 491

OTRAS COMPETENCIAS Y CONCEPTOS 492

GLOSARIO 496

LISTA DE REFERENCIAS Y MATERIAL EN LÍNEA 503

ÍNDICE DE MATERIAS 519

MÁXIMA COMODIDAD CON LA VERSIÓN IMPRESA

Tiene en sus manos una versión digital gratuita de **Data-Driven Construction**. Para trabajar con mayor comodidad y acceder rápidamente a los materiales, le recomendamos que preste atención a la **edición impresa**:

- **Siempre a mano**: el libro en formato impreso se convertirá en una herramienta de trabajo fiable, que le permitirá encontrar y utilizar rápidamente las visualizaciones y diagramas adecuados en cualquier situación de trabajo.
- **Ilustraciones de alta calidad**: todas las imágenes y gráficos de la edición impresa se presentan con la máxima calidad.
- **Acceso rápido a la información**: navegación sencilla, posibilidad de tomar notas, marcar favoritos y trabajar con el libro en cualquier lugar.

Al adquirir la versión impresa completa del libro, obtendrá una cómoda herramienta para trabajar con la información de forma cómoda y eficaz: la posibilidad de utilizar rápidamente materiales visuales en las tareas cotidianas, encontrar rápidamente los esquemas necesarios y tomar notas. Además, su compra apoya la difusión del conocimiento abierto.

Puede solicitar la versión impresa del libro en: datadrivenconstruction.io/books

I PARTE
DE LAS TABLILLAS DE ARCILLA A LA REVOLUCIÓN DIGITAL: CÓMO HA EVOLUCIONADO LA INFORMACIÓN EN LA CONSTRUCCIÓN

La primera parte del libro examina la evolución histórica de la gestión de datos en el sector de la construcción, desde los primitivos registros en soportes físicos hasta los modernos ecosistemas digitales. Analiza la transformación de las tecnologías de gestión de la información, la aparición de los sistemas ERP y el impacto de la fragmentación de los datos en la eficiencia de los procesos empresariales. Se presta especial atención al proceso de digitalización de la información y a la creciente importancia del análisis objetivo en lugar del juicio subjetivo de los expertos. Se examina en detalle el crecimiento exponencial de los volúmenes de información a los que se enfrenta la industria moderna de la construcción y los retos asociados para los sistemas empresariales. Se explora el posicionamiento de la industria de la construcción en el contexto de la cuarta y quinta revoluciones industriales, así como el potencial del uso de la inteligencia artificial y los enfoques centrados en los datos para crear una ventaja competitiva sostenible.

CAPÍTULO 1.1.
EVOLUCIÓN DEL USO DE DATOS EN EL SECTOR DE LA CONSTRUCCIÓN

El nacimiento de la era de los datos en la construcción

Hace unos 10.000 años, en el Neolítico, la humanidad realizó una transición revolucionaria en su desarrollo, abandonando el estilo de vida nómada en favor de la vida sedentaria, lo que dio lugar a la aparición de las primeras construcciones primitivas de arcilla, madera y piedra [6]. A partir de este momento comienza la historia de la industria de la construcción.

Con el desarrollo de las civilizaciones, la arquitectura se hizo cada vez más compleja, dando lugar a los primeros templos rituales y edificios públicos. La creciente complejidad de los diseños arquitectónicos obligó a los ingenieros y gestores de la Antigüedad a crear los primeros registros y cálculos. Los primeros registros en tablillas de arcilla y papiros incluían a menudo una descripción de la lógica que había detrás del cálculo de la cantidad de materiales de construcción necesarios, su coste y el cálculo del pago por el trabajo realizado [7]. Así comenzó la era del uso de datos en la construcción, mucho antes de la llegada de las modernas tecnologías digitales (Fig, 1.1-1).

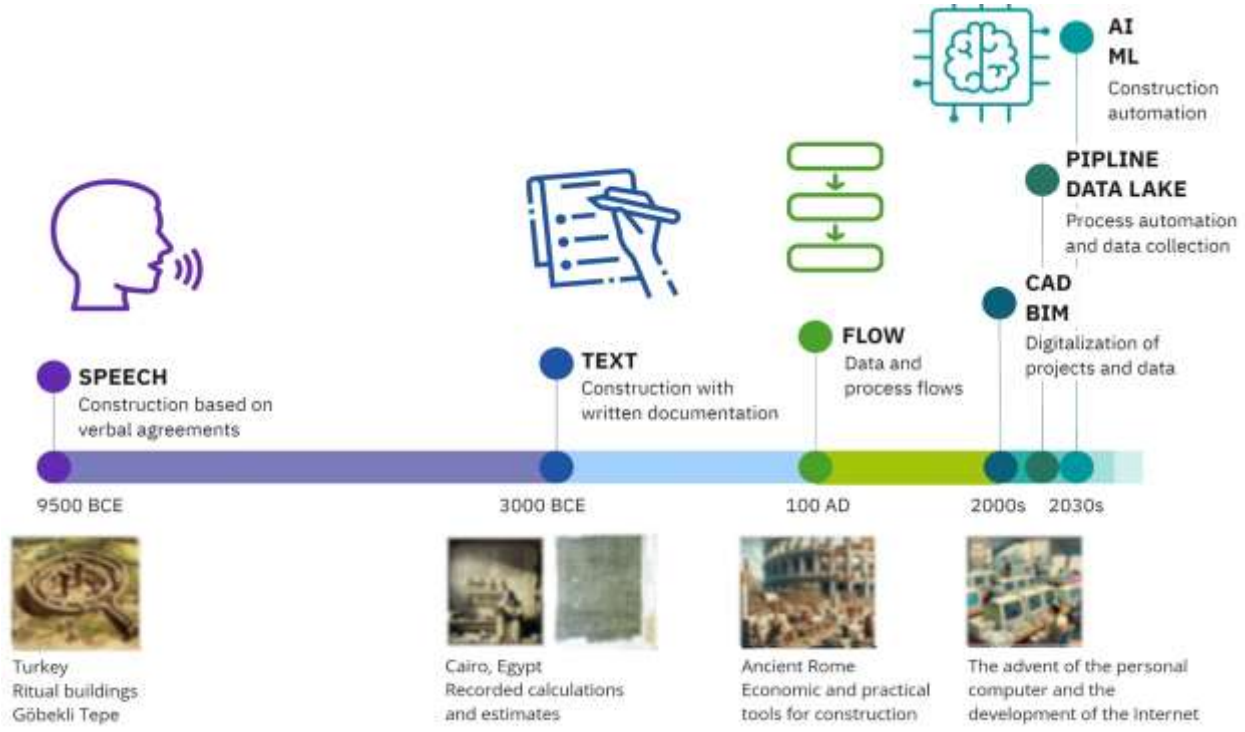


Fig. 1.1-1 Cronología del desarrollo de las tecnologías de la información en la construcción: de la información verbal a la inteligencia artificial.

De la arcilla y el papiro a la tecnología digital

Los primeros testimonios documentales en construcción se remontan a la época de la construcción de pirámides, en torno al 3000-4000 a.C.[7]. Desde entonces, el mantenimiento de registros escritos ha facilitado y acompañado el progreso de la industria de la construcción, permitiendo la acumulación y sistematización de conocimientos que, a lo largo de los 10.000 años siguientes, dieron lugar a importantes innovaciones en los métodos de construcción y la arquitectura.

El uso de los primeros soportes físicos en la construcción, como las tablillas de arcilla, los papiros de hace miles de años (Fig, 1.1-2) o el papel "A0" en la década de 1980, para registrar datos no estaba pensado originalmente para aplicar esta información a nuevos proyectos. El principal objetivo de tales registros era detallar el estado actual del proyecto, incluidos los cálculos de los materiales necesarios y el coste de la obra. Del mismo modo, en el mundo actual, la disponibilidad de datos y modelos digitales de diseño no siempre garantiza su aplicación en futuros proyectos y a menudo sirve principalmente como información para los cálculos actuales de los materiales necesarios y los costes de construcción.

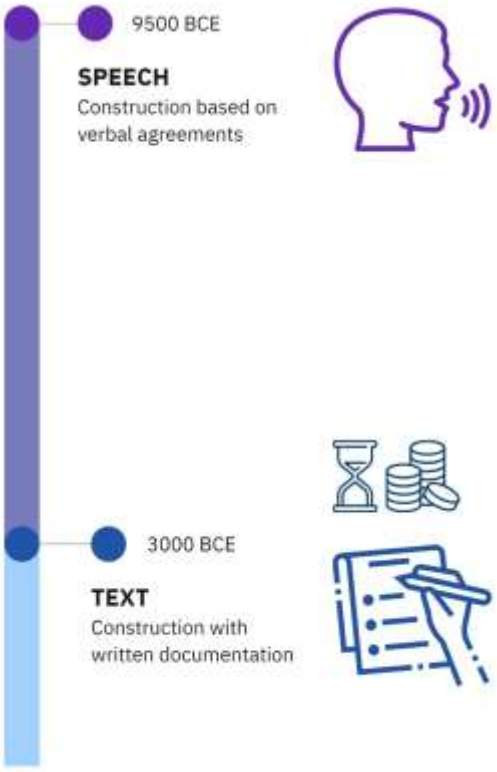


Fig. 1.1-2 Papiro del siglo III a.C. en el que se describe el coste de pintar varios tipos de ventanas de un palacio real con la técnica de la encáustica.

> La humanidad tardó unos 5.000 años en pasar de las conversaciones verbales a los documentos escritos en la gestión de proyectos de construcción, y otro tanto en pasar del papel a los datos digitales como principal recurso de planificación y control.

Del mismo modo que el desarrollo de las relaciones comerciales y monetarias estimuló la aparición de la escritura y de los primeros abogados para resolver litigios, los primeros registros de costes materiales y alcances de obra en la construcción propiciaron la aparición de los primeros gestores en la industria de la construcción, entre cuyas funciones figuraban documentar, supervisar y responsabilizarse de la información clave sobre plazos y costes de los proyectos.

Hoy en día, los datos desempeñan un papel mucho más importante: no sólo registran las decisiones tomadas, sino que también se convierten en una herramienta para predecir y modelizar el futuro. Esta es la base sobre la que se asienta el moderno enfoque de procesos en la gestión de proyectos: convertir la experiencia acumulada en un sistema de toma de decisiones basado en datos estructurados y verificables.

El proceso como herramienta para una experiencia basada en datos

En el centro de cualquier proceso está la transformación de la experiencia pasada en una herramienta para planificar el futuro. La experiencia en sentido moderno es un conjunto estructurado de datos que pueden analizarse para hacer predicciones con conocimiento de causa.

> Son los datos históricos los que sirven de base a las previsiones, ya que demuestran claramente los resultados del trabajo realizado y permiten conocer los factores que afectan a esos resultados.

Pongamos un ejemplo concreto de construcción monolítica: normalmente, al planificar el calendario de las obras, se tienen en cuenta el volumen de hormigón, la complejidad de la estructura y las condiciones meteorológicas. Supongamos que un determinado capataz de obra o los datos históricos de la empresa de los últimos tres años (2023-2025) muestran que verter una estructura monolítica de 200 m² en tiempo lluvioso llevó entre 4,5 y 6 días (Fig. 1.1-3). Son estas estadísticas acumuladas las que se convierten en la base para predecir los plazos y calcular los recursos a la hora de planificar trabajos similares en futuros proyectos. A partir de estos datos históricos, el capataz o el estimador pueden hacer una previsión informada, basada en la experiencia, del tiempo necesario para completar futuros trabajos similares en 2026 en condiciones similares.

En este caso de la analítica temporal, el proceso analítico actúa como mecanismo para transformar datos dispares en experiencias estructuradas y luego en una herramienta de planificación precisa. Los datos y los procesos constituyen un ecosistema único en el que uno no puede existir sin el otro.

> *Contar lo que es contable, medir lo que es mensurable y hacer mensurable lo que no es mensurable [8].*
>
> - Galileo Galilei

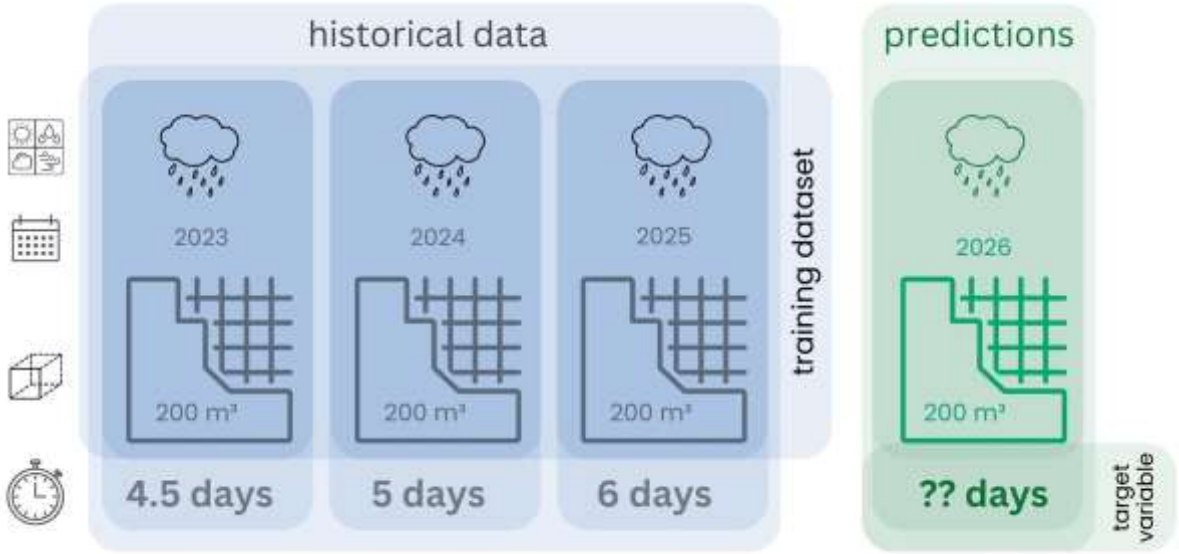


Fig. 1.1-3 Los datos históricos actúan como conjunto de datos de entrenamiento para predecir uno de los valores en el futuro

En el panorama empresarial actual, el análisis de datos se está convirtiendo en un componente esencial de la gestión eficaz de proyectos, la optimización de procesos y la toma de decisiones estratégicas. El sector de la construcción está dominando gradualmente cuatro niveles clave de analítica, cada uno de los cuales responde a una pregunta específica y aporta ventajas únicas (Fig. 1.1-4):

- **Análisis descriptivo** - responde a la pregunta "*¿qué ha pasado*?" y proporciona datos históricos e informes sobre acontecimientos y resultados pasados: en los últimos tres años (2023-2025), se tardó entre 4,5 y 6 días en verter una estructura monolítica de 200 m² en tiempo lluvioso.
- **Análisis de diagnóstico**: responde a la pregunta "*¿por qué ha ocurrido esto?"* identificando las causas de los problemas: el análisis muestra que el tiempo de vertido de la estructura monolítica aumentó debido al tiempo lluvioso, que ralentizó el proceso de curado del hormigón.
- **Análisis predictivos**: con visión de futuro, predicen posibles riesgos y plazos respondiendo a la pregunta "*¿qué ocurrirá*?": basándose en datos históricos, se prevé que verter una estructura monolítica similar de 200 m² en tiempo lluvioso en 2026 llevará aproximadamente 5,5 días, teniendo en cuenta todos los factores y tendencias conocidos.
- **Análisis prescriptivo** - proporciona recomendaciones automatizadas y responde a la pregunta "*¿qué hacer*?", permitiendo a las empresas elegir las acciones óptimas: Para optimizar el trabajo, por ejemplo, se recomienda: utilizar aditivos especiales para acelerar el curado del hormigón en

condiciones de alta humedad; planificar el vertido para los periodos con menor probabilidad de precipitaciones; disponer refugios temporales para la estructura, que reducirán el tiempo de trabajo a 4-4,5 días incluso en condiciones meteorológicas adversas.

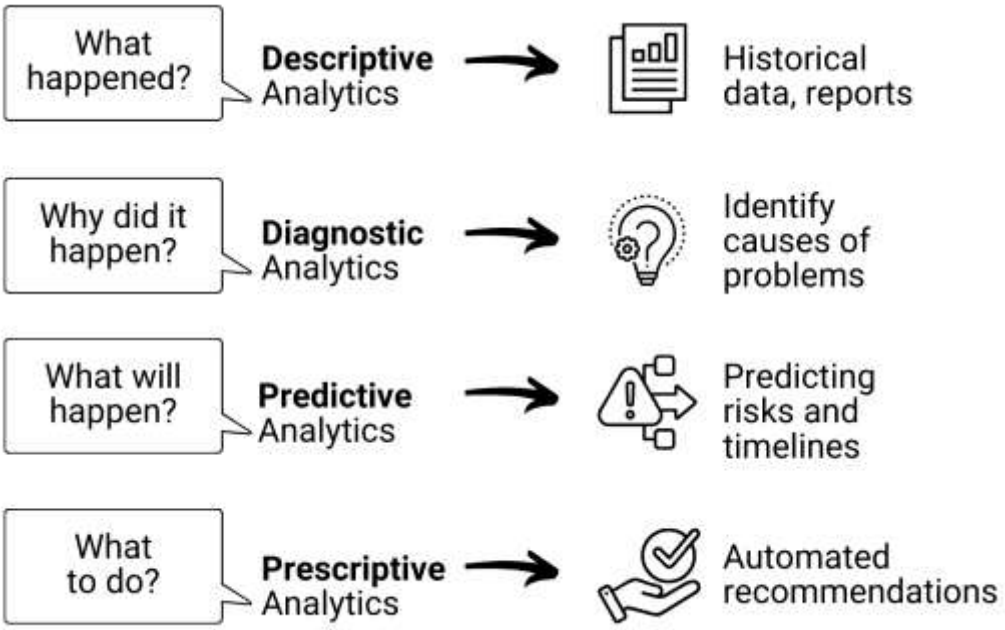


Fig. 1.1-4 Principales tipos de análisis: de la descripción del pasado a la toma de decisiones automatizada.

La transformación digital en toda regla, que implica una transición hacia la analítica de sistemas y la gestión basada en datos, requiere no sólo la externalización, sino la formación de un equipo interno competente. Los miembros clave de ese equipo deben ser gestores de productos, ingenieros de datos, analistas y desarrolladores, que trabajarán en estrecha colaboración con las unidades de negocio (Fig. 4.3-9). Esta colaboración es necesaria para formular preguntas analíticas inteligentes y parametrizar eficazmente las tareas de toma de decisiones empresariales. En una sociedad de la información, los datos no se convierten en una mera herramienta auxiliar, sino en la base de la previsión y la optimización.

En el sector de la construcción, la transformación digital está cambiando radicalmente la forma de diseñar, gestionar y explotar las instalaciones. Este proceso se conoce como digitalización de la información: todos los aspectos del proceso de construcción se digitalizan y se convierten en un formato digital apto para el análisis.

Digitalización de la información del proceso de construcción

Durante milenios, la cantidad de información registrada en la construcción apenas ha variado, pero ha crecido rápidamente en las últimas décadas (Fig, 1.1-5).

> Según el estudio de PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World"(2015) [9], el 90% de todos los datos del mundo se crearon en los dos últimos años (en 2015). Sin embargo, la mayoría de las empresas no aprovechan plenamente estos datos, ya que permanecen en sistemas aislados o simplemente se archivan sin un análisis real.

El aumento del volumen de datos no ha hecho más que acelerarse en los últimos años, duplicándose de 15 zettabytes en 2015 a 181 zettabytes en 2025 [10]. Cada día, los servidores de las empresas de construcción y diseño se llenan de documentación de proyectos, programas de trabajo, cálculos e informes financieros. Para los planos 2D/3D -se utilizan los formatos DWG, DXF y DGN, y para los modelos 3D - RVT, NWC, PLN e IFC™. Los documentos de texto, tablas y presentaciones se guardan en DOC, XLSX y PPT. Además de vídeo e imágenes de la obra - en MPG y JPEG, datos en tiempo real de componentes IoT, etiquetas RFID® (identificación y seguimiento) y sistemas de gestión de edificios BMS (supervisión y control)

Fig. 1.1-5 Crecimiento parabólico de datos 2010-2025 (basado en [10]).

Con el rápido crecimiento de la información, el sector de la construcción se enfrenta a la necesidad no sólo de recopilar y almacenar datos, sino también de garantizar su verificación, validación, mensurabilidad y tratamiento analítico. En la actualidad, el sector atraviesa una fase activa de digitalización de la información: la transformación sistemática de todos los aspectos de la actividad de la construcción en un formato digital apto para el análisis, la interpretación y la automatización.

> Digitalizar la información significa tomar información sobre todas las entidades y elementos de un proyecto de construcción y del propio proceso de construcción -incluidos aquellos que antes no considerábamos información en absoluto- y convertirla a un formato de datos para que la información sea cuantificable y fácil de analizar.

En el contexto de la construcción, esto significa capturar y digitalizar información sobre todos los elementos de los proyectos y todos los procesos -desde el movimiento de maquinaria y personas en la obra hasta las condiciones meteorológicas y climáticas en el lugar de construcción, los precios actuales de los materiales y los tipos de interés de los bancos centrales- para generar modelos analíticos.

> *Si puedes medir aquello de lo que hablas y expresarlo en números, entonces sabes algo sobre el tema. Pero si no puede expresarlo cuantitativamente, su conocimiento es extremadamente limitado e insatisfactorio. Puede ser un punto de partida, pero no es el nivel del verdadero conocimiento científico. [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, científico británico

La digitalización de la información va mucho más allá del enfoque tradicional de recopilación de información, en el que solo se registraban métricas básicas como las horas de trabajo o los costes reales de material. Hoy en día, prácticamente cualquier evento puede transformarse en un flujo de datos apto para un análisis en profundidad mediante herramientas avanzadas de análisis y técnicas de aprendizaje automático. El sector de la construcción ha experimentado un cambio fundamental, pasando de los planos en papel, las hojas de cálculo de Excel y las instrucciones verbales a sistemas digitales (Fig, 1.2-4) en los que cada elemento de un proyecto se convierte en una fuente de datos. Incluso los empleados -desde los ingenieros hasta los trabajadores de la construcción a pie de obra- se ven ahora como una colección de variables digitales y conjuntos de datos.

> Según el informe de KPMG "Desafíos conocidos - Nuevos enfoques: Encuesta mundial de la construcción 2023", los gemelos digitales, la IA (Inteligencia Artificial) y Big Data, se perfilan como impulsores clave para mejorar la rentabilidad de los proyectos [2].

Las tecnologías modernas no sólo simplifican la recogida de información, haciéndola en gran medida automática, sino que también reducen radicalmente el coste del almacenamiento de datos. Como consecuencia, las empresas están abandonando el enfoque selectivo y prefieren almacenar todo el conjunto de información para su posterior análisis (Fig. 2.1-5), lo que abre oportunidades potenciales para optimizar los procesos en el futuro.

> La digitalización de la información y la digitalización permiten sacar a la luz el valor oculto de la información, hasta ahora desaprovechado. Si se organizan adecuadamente, los datos pueden reutilizarse, reinterpretarse e integrarse en nuevos servicios y soluciones.

En el futuro, es probable que la digitalización de la información conduzca a la automatización total de la gestión de documentos, la introducción de procesos de construcción autogestionados y la aparición de nuevas profesiones: analistas de datos de construcción, expertos en gestión de proyectos de IA e ingenieros digitales. Los proyectos de construcción se convertirán en fuentes dinámicas de información, y la toma de decisiones no se basará en la intuición o la experiencia subjetiva, sino en hechos digitales fiables y reproducibles

> *La información es el petróleo del siglo XXI, y la analítica es el motor de combustión interna [12].*
>
> - Peter Sondergaard, Vicepresidente Senior de Gartner®

> Según IoT Analytics 2024 [13], se espera que el gasto mundial en gestión y análisis de datos aumente drásticamente de 185.500 millones de dólares en 2023 a 513.300 millones de dólares en 2030, a una tasa de crecimiento anual compuesta del 16%. Sin embargo, no todos los componentes crecen al mismo ritmo: la analítica crece rápidamente, mientras que el crecimiento del almacenamiento se ralentiza. La analítica proporcionará el crecimiento más rápido en el ecosistema de gestión de datos: se prevé que pase de 60.600 millones de dólares en 2023 a 227.900 millones de dólares en 2030, una tasa de crecimiento anual compuesto del 27%.

Con la aceleración de la digitalización de la información y el rápido crecimiento de los volúmenes de información, la gestión de proyectos y empresas de construcción se enfrenta a la necesidad de almacenar, analizar y procesar sistemáticamente datos diversos, a menudo heterogéneos. En respuesta a este reto, a partir de mediados de los noventa, la industria inició un cambio masivo hacia la creación, el almacenamiento y la gestión electrónicos de la documentación: desde hojas de cálculo y cálculos de diseño hasta planos y contratos.

Los documentos tradicionales en papel, que requieren firmas, almacenamiento físico, revisión periódica y archivo en armarios, están siendo sustituidos gradualmente por sistemas digitales que almacenan los datos de forma estructurada: en bases de datos de aplicaciones especializadas.

CAPÍTULO 1.2.
TECNOLOGÍAS Y SISTEMAS DE GESTIÓN EN LA CONSTRUCCIÓN MODERNA

La revolución digital y la aparición de sistemas modulares MRP/ERP -systems

La era del almacenamiento y procesamiento de datos digitales modernos comenzó con la llegada de la cinta magnética en los años 50, que abrió la posibilidad de almacenar y utilizar grandes cantidades de información. El siguiente gran avance fue la aparición de las unidades de disco, que cambiaron radicalmente el planteamiento de la gestión de datos en la industria de la construcción.

Con el desarrollo del almacenamiento de datos, un gran número de empresas se han introducido en el mercado de soluciones y han empezado a desarrollar software modular para crear, almacenar, procesar datos y automatizar tareas rutinarias

> El crecimiento exponencial de la información y las herramientas ha llevado a la necesidad de soluciones integradas y modulares que no trabajen con archivos individuales, sino que ayuden a gestionar y controlar el flujo de datos en todos los procesos y proyectos.

Las primeras herramientas integrales de la plataforma no sólo debían almacenar documentos, sino también documentar todas las solicitudes de cambio y operaciones en los procesos: quién las inició, cuál era el alcance de la solicitud y qué se registró finalmente como valor o atributo. Para ello, se necesitaba un sistema que pudiera realizar un seguimiento preciso de los cálculos y las decisiones tomadas (Fig. 1.2-1). Este tipo de plataformas fueron los primeros sistemas MRP (Material Requirements Planning) y ERP (Enterprise Resource Planning) que se popularizaron a partir de principios de los noventa [14]

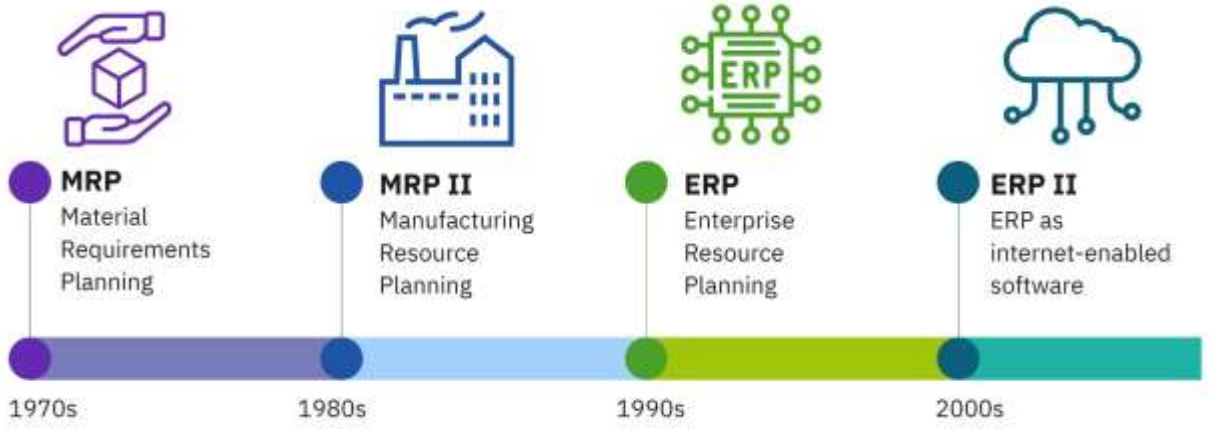


Fig. 1.2-1 Los avances en la tecnología de almacenamiento de datos propiciaron la aparición de los sistemas ERP en la década de 1980.

Los primeros sistemas MRP -y ERP - sentaron las bases de la era de la digitalización en la gestión de procesos empresariales y proyectos de construcción. Los sistemas modulares, diseñados originalmente para automatizar procesos empresariales clave, se integraron con el tiempo con soluciones de software adicionales, más flexibles y adaptables.

Estas soluciones adicionales se diseñaron para el procesamiento de datos y la gestión de contenidos de proyectos (Fig. 1.2-2), o bien sustituyeron a determinados módulos de los grandes sistemas, o bien los complementaron eficazmente, ampliando la funcionalidad de todo el sistema.

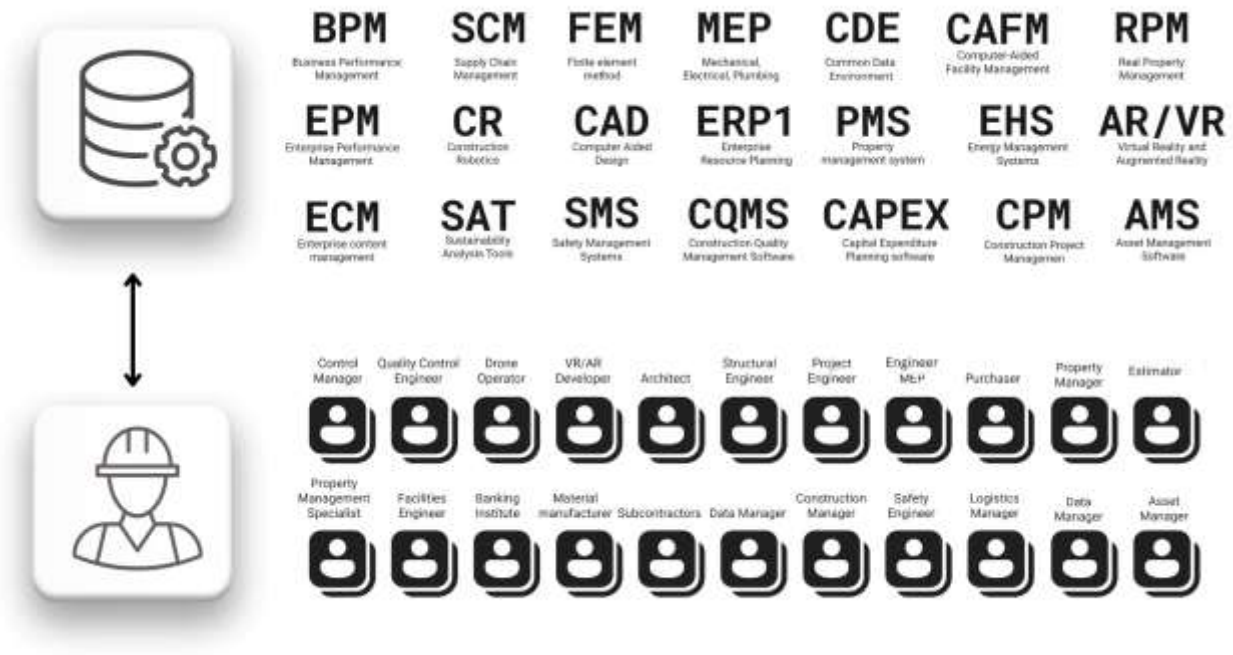


Fig. 1.2-2 Las nuevas soluciones informáticas han atraído a la empresa a un ejército de gestores para gestionar los flujos de datos.

En las últimas décadas, las empresas han invertido mucho en sistemas modulares [15], percibiéndolos como soluciones integradas a largo plazo.

> Según el informe Software Path para 2022 [16], el presupuesto medio por usuario de un sistema ERP es de 9.000 dólares. Por término medio, alrededor del 26% de los empleados de la empresa utilizan este tipo de sistemas. Así, para una organización con 100 usuarios, el coste total de la implantación de un ERP alcanza aproximadamente los 900.000 dólares.

Las inversiones en soluciones propietarias, cerradas y modulares están cada vez menos justificadas ante el rápido desarrollo de tecnologías modernas, flexibles y abiertas. Si ya se han realizado tales inversiones, es importante reevaluar objetivamente el papel de los sistemas existentes: si siguen siendo esenciales a largo plazo, o si sus funciones pueden revisarse y aplicarse de forma más eficiente y transparente.

Uno de los principales problemas de las plataformas de datos modulares actuales es que centralizan la gestión de datos en aplicaciones cerradas. Como resultado, los datos -el activo principal de una empresa- pasan a depender de soluciones de software específicas, en lugar de viceversa. Esto limita la reutilización de la información, complica la migración a y reduce la agilidad empresarial en un panorama digital en rápida

evolución.

Si es probable que la pertinencia o relevancia de la arquitectura modular cerrada disminuya en el futuro, tiene sentido reconocer los costes incurridos hoy como costes irrecuperables y centrarse en un cambio estratégico hacia un ecosistema digital más abierto, escalable y adaptable.

> El software propietario se caracteriza por el control exclusivo del desarrollador sobre el código fuente y los datos del usuario creados como parte del uso de dichas soluciones. A diferencia del software de código abierto, los usuarios no tienen acceso a la estructura interna de la aplicación y no pueden revisarla, modificarla o adaptarla a sus necesidades de forma independiente. En su lugar, están obligados a adquirir licencias que les conceden el derecho a utilizar el software dentro de los límites establecidos por el vendedor.

Un enfoque moderno centrado en los datos ofrece un paradigma diferente: los datos deben considerarse un activo estratégico importante, independiente, duradero y separado de soluciones de software específicas. Las aplicaciones, a su vez, se convierten en meras herramientas de datos que pueden sustituirse libremente sin riesgo de perder información crítica.

El desarrollo de los sistemas ERP y MRP en la década de 1990 (Fig, 1.2-1) proporcionó a las empresas potentes herramientas para la gestión de procesos, pero también tuvo la consecuencia imprevista de aumentar considerablemente el número de personas implicadas en el mantenimiento de los flujos de información. En lugar de automatizar y simplificar las tareas operativas, estos sistemas crearon a menudo nuevos niveles de complejidad, burocracia y dependencia de los recursos informáticos internos.

Sistemas de gestión de datos: de la minería de datos a los retos empresariales

Las empresas de hoy se enfrentan a la necesidad de integrar múltiples sistemas de gestión de datos. Seleccionar sistemas de gestión de datos, gestionar estos sistemas de forma inteligente e integrar fuentes de datos dispares se está convirtiendo en algo fundamental para el rendimiento empresarial.

A mediados de la década de 2020, se pueden encontrar cientos (y miles en las grandes empresas de construcción) de sistemas diferentes (Fig, 1.2-3) que deben funcionar en armonía para que todos los aspectos del proceso de construcción se desarrollen de forma fluida y cohesionada.

> Según el estudio de Deloitte de 2016® "Data-Driven Management in Digital Capital Projects" - el profesional medio de la construcción utiliza 3,3 aplicaciones de software al día, pero solo 1,7 de ellas están integradas entre sí[17].

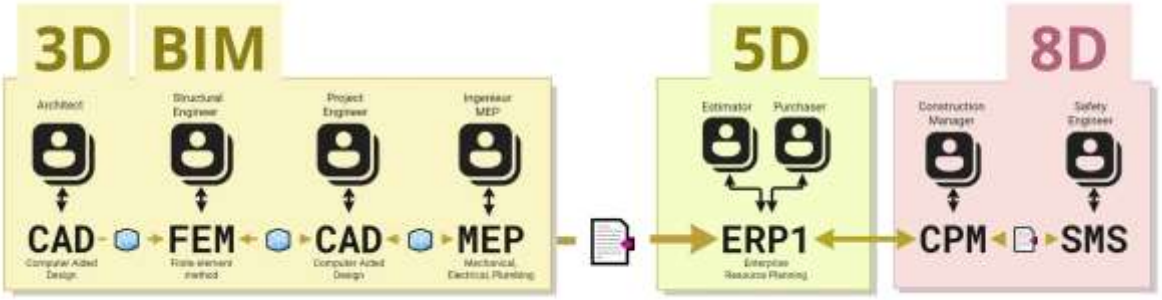


Fig. 1.2-3 Todo sistema empresarial requiere un equipo profesional y un gestor responsable para una gestión de datos de calidad.

A continuación se ofrece una lista de sistemas populares para empresas medianas y grandes del sector de la construcción que se utilizan en la gestión eficaz de proyectos de construcción:

- **ERP (Planificación de Recursos Empresariales**): proporciona integración de procesos empresariales, como contabilidad, compras y gestión de proyectos.
- **CAPEX (Capital Expenditure Planning Software)** - utilizado para presupuestar y gestionar inversiones financieras en proyectos de construcción, ayuda a determinar el coste de los activos fijos y las inversiones en activos a largo plazo.
- **CAD (Computer-Aided Design)** y **BIM (Building Information Modeling)** - se utilizan para crear dibujos técnicos detallados y precisos y modelos 3D - de proyectos. Estos sistemas se centran en el trabajo con información geométrica.
- **MEP (Mechanical, Electrical, Plumbing**) - Sistemas de ingeniería que incluyen componentes mecánicos, eléctricos y de fontanería, y detallan el sistema "circulatorio" interno de un proyecto.
- **SIG (Sistemas de Información Geográfica**): se utilizan para el análisis y la planificación del terreno, incluida la cartografía y el análisis espacial.
- **CQMS (software de gestión de la calidad de la construcción)**: garantiza que los procesos de construcción cumplen las normas y reglamentos establecidos, lo que ayuda a eliminar defectos.
- **CPM (gestión de proyectos de construcción)**: incluye la planificación, coordinación y control de los procesos de construcción.
- **CAFM (Gestión de instalaciones asistida por ordenador)**: sistemas de gestión y mantenimiento de edificios.
- **La gestión de la cadena de suministro (SCM**) es necesaria para optimizar el flujo de materiales e información entre los proveedores y la obra.
- **EPM (Enterprise Performance Management)** - destinado a mejorar los procesos y el rendimiento empresariales.
- **AMS (Asset Management Software)**: se utiliza para optimizar el uso, la gestión y el mantenimiento de equipos e infraestructuras a lo largo del ciclo de vida de los activos.
- **RPM (Real Property Management)** - incluye tareas y procesos relacionados con la gestión y explotación de edificios y terrenos, así como de los recursos y activos asociados.

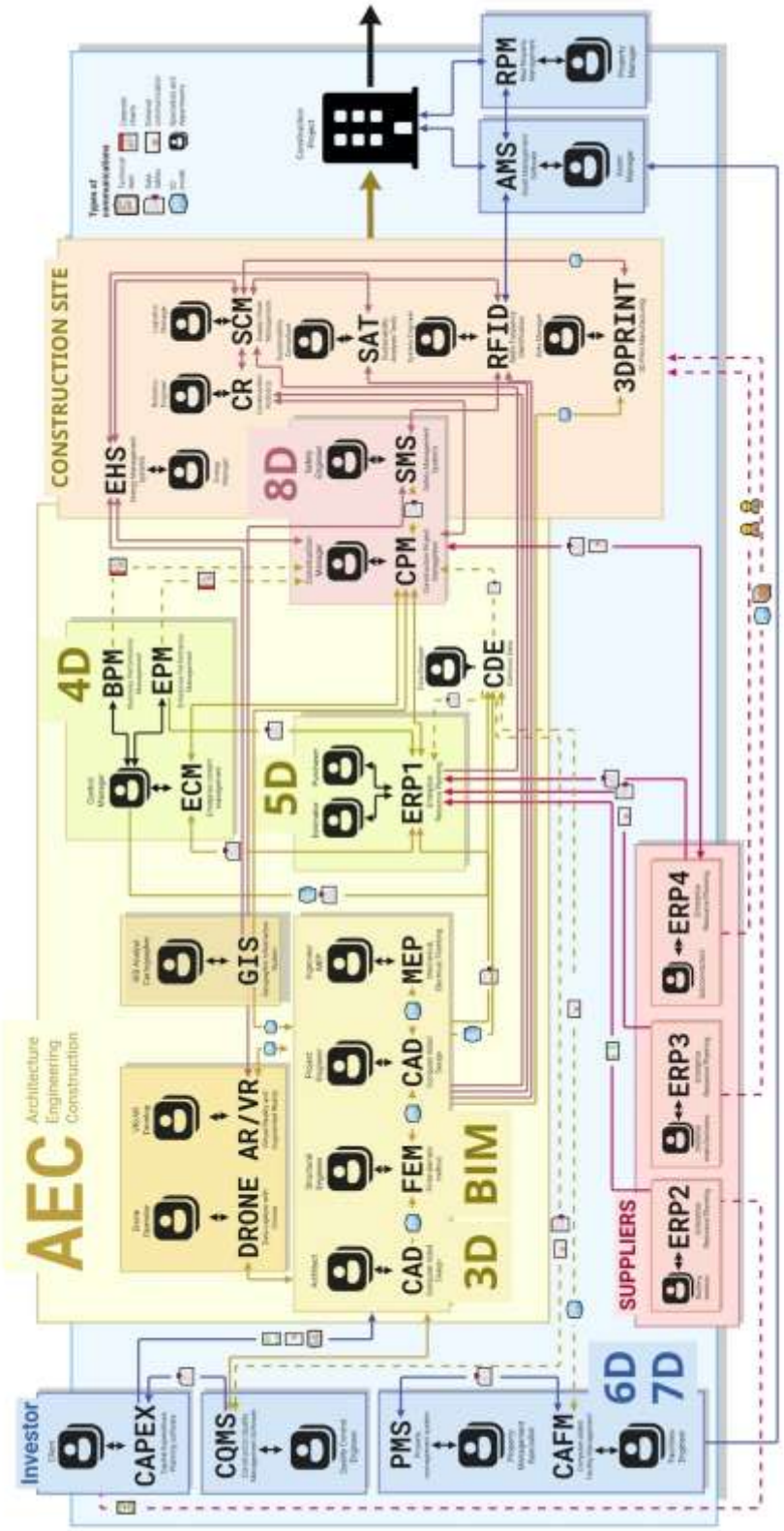


Fig. 1.2-4 Interconectividad de los sistemas que conecta los procesos de la empresa con el flujo de información entre los distintos departamentos.

- **CAE (Ingeniería asistida por ordenador)** - La ingeniería asistida por ordenador incluye sistemas de cálculo y simulación como el análisis de elementos finitos (FEA) y la dinámica de fluidos computacional (CFD).
- **CFD (**Computational **Fluid** Dynamics) - Dinámica de fluidos computacional, modelización de flujos de fluidos y gases. Subcategoría CAE.
- **CAPP (Computer-Aided Process Planning)** - Planificación de procesos asistida por ordenador. Se utiliza para crear mapas de rutas y procesos.
- **CAM (Fabricación asistida por ordenador)** - fabricación asistida por ordenador, generación de programas de control para máquinas CNC.
- **PDM (Product Data Management)** - Gestión de datos de productos, un sistema para almacenar y gestionar documentación técnica.
- **MES (Manufacturing Execution System)** es un sistema de control de procesos de fabricación en tiempo real.
- **PLM (Product Lifecycle Management)** - gestión del ciclo de vida de un elemento de proyecto, integra PDM, CAPP, CAM y otros sistemas para un control completo del producto desde su desarrollo hasta su eliminación.

Estos y muchos otros sistemas, que incorporan diversas soluciones informáticas, se han convertido en parte integrante de la industria moderna de la construcción (Fig. 1.2-4). En esencia, estos sistemas son bases de datos especializadas con interfaces intuitivas que permiten introducir, procesar y analizar información de forma eficaz en todas las fases del diseño y la construcción. La integración de las herramientas digitales entre sí no sólo ayuda a optimizar los procesos de trabajo, sino que también mejora notablemente la precisión de las decisiones, lo que repercute positivamente en los plazos y la calidad de la ejecución de los proyectos.

Pero en la mitad de los casos no hay integración. Según las estadísticas, sólo una de cada dos aplicaciones o sistemas está integrada con otras soluciones [17]. Esto indica la continua fragmentación del entorno digital y subraya la necesidad de desarrollar normas abiertas e interfaces unificadas para garantizar el intercambio de información de extremo a extremo dentro de un proyecto de construcción.

Uno de los principales retos de la integración para las empresas modernas sigue siendo la elevada complejidad de los sistemas digitales y los requisitos de competencia de los usuarios necesarios para una recuperación e interpretación eficaces de la información. Se forma un equipo de especialistas, encabezado por un gestor clave (Fig. 1.2-2), para dar soporte a cada sistema implantado en la empresa.

> El gestor de sistemas clave desempeña un papel decisivo en la buena dirección del flujo de datos y es responsable de la calidad de la información final, igual que los primeros gestores de hace miles de años eran responsables de los números escritos en papiros o tablillas de arcilla.

Para convertir flujos de información dispares en una herramienta de gestión, es esencial la capacidad de integrar y gestionar sistemáticamente los datos. En esta arquitectura, los directivos deben actuar como elementos de una red única, como un micelio que conecta las partes individuales de la empresa en un organismo vivo coherente capaz de adaptarse y evolucionar.

Micelio empresarial: cómo los datos conectan con los procesos empresariales

El proceso de integración de datos en aplicaciones y bases de datos se basa en la agregación de información procedente de diversas fuentes, incluidos distintos departamentos y especialistas (Fig, 1.2-4). Los especialistas buscan los datos pertinentes, los procesan y los transfieren a sus sistemas y aplicaciones para su uso posterior.

Cada sistema de la empresa, formado por un conjunto de herramientas, tecnologías y bases de datos, es un árbol del conocimiento enraizado en el suelo de los datos históricos y que crece para dar nuevos frutos en forma de soluciones acabadas: documentos, cálculos, tablas, gráficos y cuadros de mando (Fig. 1.2-5). Los sistemas de una empresa, como los árboles de un determinado bosque, interactúan y se comunican entre sí, constituyendo un sistema complejo y bien estructurado apoyado y gestionado por gestores expertos.

> El sistema de recuperación y transferencia de información de una empresa funciona como una compleja red forestal formada por árboles (sistemas) y hongos micelio (gestores) que actúan como conductores y recicladores, garantizando que la información se transfiera y fluya a los sistemas adecuados. Esto ayuda a mantener un flujo y una distribución de datos sanos y eficientes dentro de la empresa.

Los expertos, como las raíces, absorben los datos en bruto en las fases iniciales de un proyecto, convirtiéndolos en nutrientes para el ecosistema corporativo. Los sistemas de gestión de datos y contenidos (Fig. 1.2-4 - ERP, CPM, BIM, etc.) actúan como potentes autopistas de la información a través de las cuales circulan estos conocimientos por todos los niveles de la empresa.

Al igual que en la naturaleza, donde cada elemento del ecosistema desempeña su papel, en el paisaje empresarial de una compañía, cada participante en el proceso -desde el ingeniero hasta el analista - contribuye al crecimiento y la fertilidad del entorno de información. Estos "árboles de datos" sistémicos (Fig. 1.2-5) no son meros mecanismos de recopilación de información, sino una ventaja competitiva que garantiza la sostenibilidad de una empresa.

Los ecosistemas forestales son un reflejo sorprendentemente preciso de cómo se organizan las estructuras corporativas digitales. Al igual que la estructura escalonada de un bosque -desde la maleza hasta las copas de los árboles-, la gobernanza corporativa asigna tareas a niveles de responsabilidad y departamentos funcionales.

> Las raíces profundas y ramificadas de un árbol proporcionan resistencia y acceso a los nutrientes. Del mismo modo, una estructura organizativa sólida y unos procesos estables para trabajar con datos de calidad sustentan todo el ecosistema de información de una empresa, contribuyendo a su crecimiento y desarrollo sostenibles incluso en periodos de inestabilidad y crisis del mercado (de vientos fuertes).

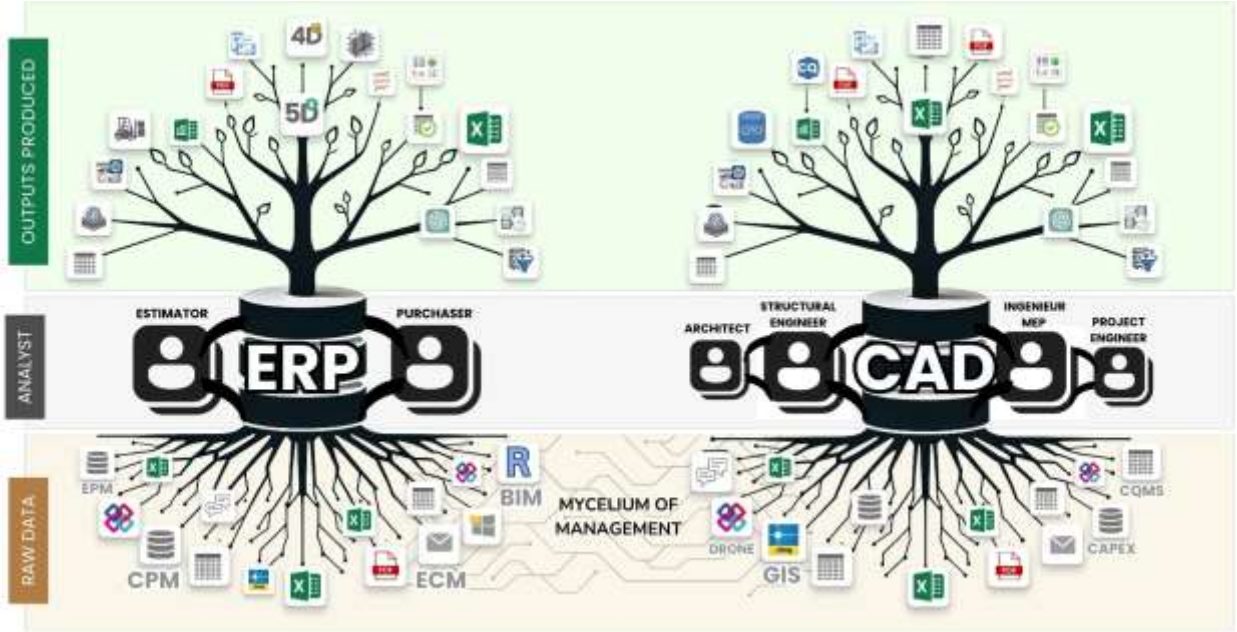


Fig. 1.2-5 La integración de datos a través de distintos sistemas es como un micelio que conecta a gestores y especialistas en una única red de información.

La comprensión moderna de la escala en los negocios ha evolucionado. Hoy en día, el valor de una empresa viene determinado no sólo por su parte visible -las "coronas" en forma de documentos finales e informes-, sino también por la profundidad del "sistema raíz" de datos recopilados cualitativamente y procesados sistemáticamente. Cuanta más información pueda recopilarse y procesarse, mayor será su valor empresarial. Las empresas que acumulan metódicamente un "abono" de datos ya procesados y son capaces de extraer de ellos ideas útiles obtienen una ventaja estratégica

> La información histórica se está convirtiendo en un nuevo tipo de capital, que permite el crecimiento, la optimización de procesos y la ventaja competitiva. En un mundo basado en los datos, no gana quien más tiene, sino quien más sabe.

Para el sector de la construcción, esto significa avanzar hacia una gestión de proyectos en tiempo real, en la que todos los procesos -desde el diseño y la contratación hasta la coordinación de contratistas- se basarán en datos pertinentes actualizados diariamente. La integración de la información procedente de distintas fuentes (sistemas ERP, modelos CAD, sensores IoT en las obras, RFID) permitirá realizar previsiones más precisas, reaccionar con rapidez a los cambios y evitar los retrasos causados por la falta de datos actualizados.

> Según el informe Data-Driven Enterprise 2025 de McKinsey & Company (McKinsey & Company®, 2022 [18]), las empresas de éxito del futuro dependerán de los datos en todos los aspectos clave de sus operaciones, desde las decisiones estratégicas hasta las interacciones operativas.

Los datos dejarán de ser solo una herramienta de análisis y se convertirán en parte integrante de todos los procesos empresariales, aportando transparencia, control y automatización de la gestión. El movimiento impulsado por los datos en permitirá a las organizaciones minimizar el impacto del factor humano, reducir

los riesgos operativos y aumentar la transparencia y la eficiencia de la toma de decisiones.

El siglo XXI está dando la vuelta al paradigma económico: mientras que antes se llamaba "oro negro" al petróleo por su capacidad para alimentar maquinaria y transporte, hoy, comprimidos bajo la presión del tiempo, los datos históricos se están convirtiendo en un nuevo recurso estratégico, que no alimenta máquinas, sino algoritmos de toma de decisiones que impulsarán los negocios.

CAPÍTULO 1.3.
LA REVOLUCIÓN DIGITAL Y LA EXPLOSIÓN DE DATOS

El inicio del auge de los datos como ola evolutiva

El sector de la construcción está experimentando una explosión de información sin precedentes. Si pensamos en la empresa como un árbol del conocimiento (Fig. 1.2-5) alimentado por datos, la fase actual de digitalización puede compararse con el rápido crecimiento de la vegetación durante el periodo Carbonífero, una época en la que la biosfera de la Tierra se transformó por la rápida acumulación de biomasa (Fig. 1.3-1).

Con la digitalización mundial, la cantidad de información en el sector de la construcción se duplica cada año. La tecnología moderna permite recopilar datos en segundo plano, analizarlos en tiempo real y utilizarlos a una escala que parecía imposible hace poco tiempo.

> Según la Ley de Moore, formulada por Gordon Moore (cofundador de Intel®), la densidad y complejidad de los circuitos integrados y la cantidad de datos procesados y almacenados se duplica aproximadamente cada dos años [19].

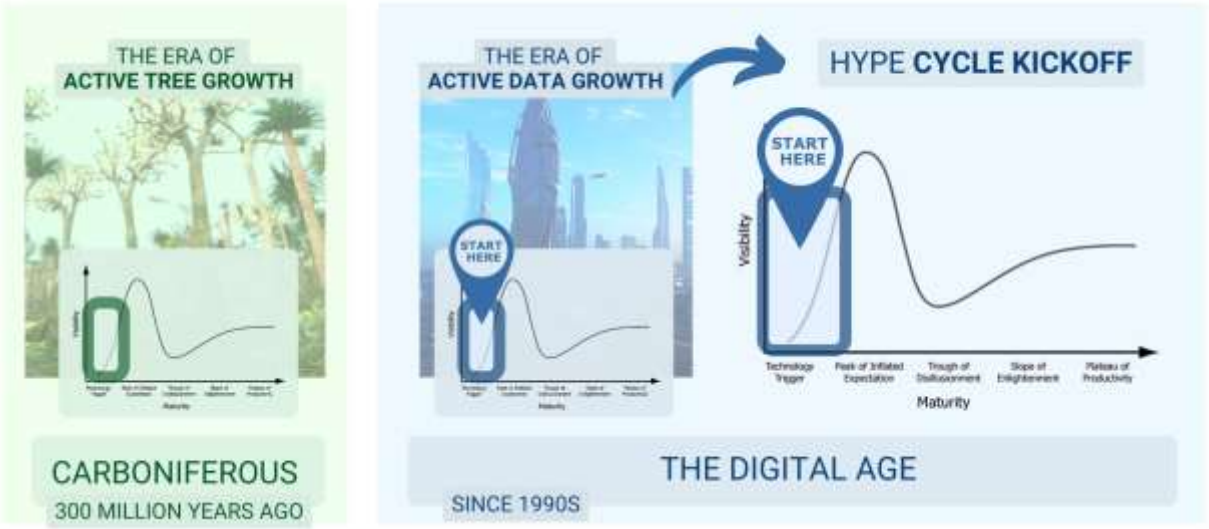


Fig. 1.3-1 El inicio de la digitalización ha provocado un crecimiento exponencial de los datos, muy parecido al auge de la vegetación en la era del carbón.

Mientras que antiguas estructuras megalíticas como Göbekli Tepe (Turquía) no dejaron tras de sí conocimientos documentados aptos para su reutilización, hoy las tecnologías digitales permiten acumular y reutilizar información. Esto puede compararse con la transición evolutiva de las plantas con esporas a las plantas con semillas (angiospermas): la aparición de la semilla dio lugar a la propagación generalizada de la vida en el planeta. (Fig. 1.3-2).

Del mismo modo, los datos de proyectos anteriores se convierten en una especie de "semillas digitales",

portadoras de conocimientos de ADN que pueden ampliarse y utilizarse en nuevos proyectos y pro ductos. La aparición de modernas herramientas de inteligencia artificial -aprendizaje automático y grandes modelos lingüísticos (LLM) como ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok- permiten extraer, interpretar y aplicar automáticamente los datos en nuevos contextos

> Al igual que las semillas revolucionaron la propagación de la vida en un planeta inicialmente sin vida, las "semillas de datos" se están convirtiendo en la base de la aparición automática de nuevas estructuras de información y conocimiento, permitiendo a los ecosistemas digitales evolucionar de forma independiente y adaptarse a las cambiantes necesidades de los usuarios.

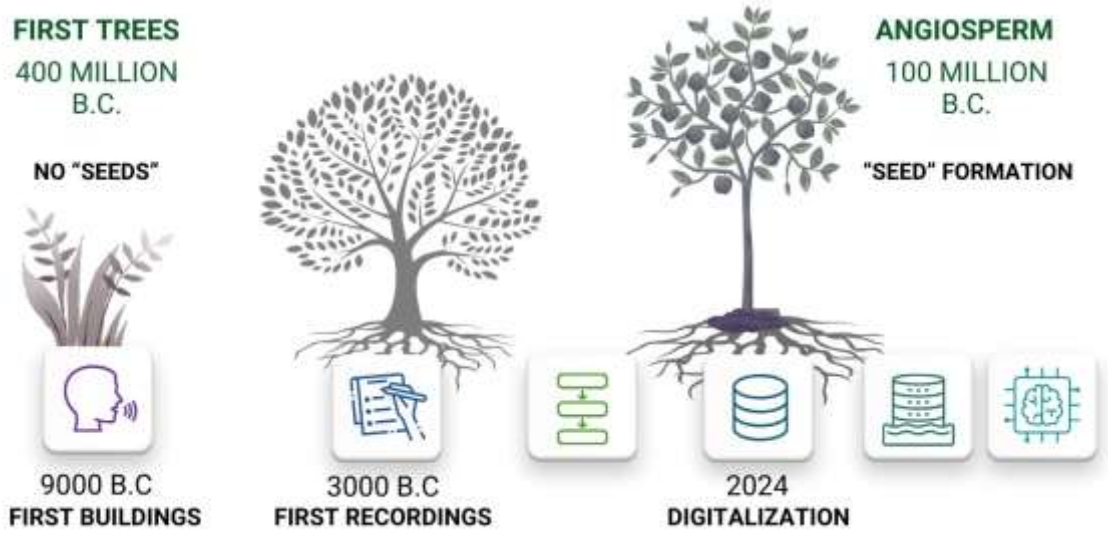


Fig. 1.3-2 Las "semillas de datos" digitales desempeñan el mismo papel evolutivo que las angiospermas, las plantas con flores que transformaron el ecosistema terrestre.

Nos encontramos en el umbral de una nueva era en la construcción, en la que la explosión de datos y la difusión activa de "semillas de datos" -información estructurada de proyectos pasados y en curso- están formando los cimientos del futuro digital del sector. Su "polinización" mediante modelos lingüísticos de big data (LLM) nos permite no limitarnos a observar el cambio digital, sino participar activamente en la creación de ecosistemas adaptativos de autoaprendizaje. Esto no es evolución: es una revolución digital en la que los datos se están convirtiendo en el principal pilar de una nueva realidad

> *La cantidad de datos en el sector de la construcción está aumentando drásticamente debido a la información procedente de diversas disciplinas a lo largo del ciclo de vida de los proyectos de construcción. Esta enorme acumulación de datos ha empujado al sector de la construcción hacia la era de los Big Data [20].*
>
> - Prof. Hang Yang, Departamento de Ingeniería Civil y Arquitectura, Universidad Tecnológica de Wuhan, Wuhan,

El crecimiento de los datos en la era de la información recuerda a los procesos evolutivos de la naturaleza: igual que el desarrollo de los bosques cambió el antiguo paisaje del planeta, la actual explosión de la información está cambiando el paisaje de todo el sector de la construcción.

La cantidad de datos generados en una empresa moderna

En los dos últimos años se ha creado el 90% de todos los datos existentes en el mundo [21]. A partir de 2023, cada persona, incluidos los profesionales del sector de la construcción, generará unos 1,7 megabytes de datos por segundo [22], y la cantidad total de datos en el mundo alcanzará los 64 zettabytes en 2023 y se prevé que supere los 180 zettabytes, o 180*10^15 megabytes, en 2025 [23].

> Esta explosión de la información tiene un precedente histórico: la invención de la imprenta por Johannes Gutenberg en el siglo XV. Sólo cincuenta años después de su introducción, el número de libros en Europa se duplicó: en pocas décadas se imprimieron tantos libros como los que se habían creado a mano en los 1.200 años anteriores [24]. Hoy asistimos a un crecimiento aún más rápido: la cantidad de datos en el mundo se duplica cada tres años.

Dado el ritmo actual de crecimiento de los datos, el sector de la construcción tiene el potencial de generar tanta información en las próximas décadas como la que ha acumulado en toda su historia anterior

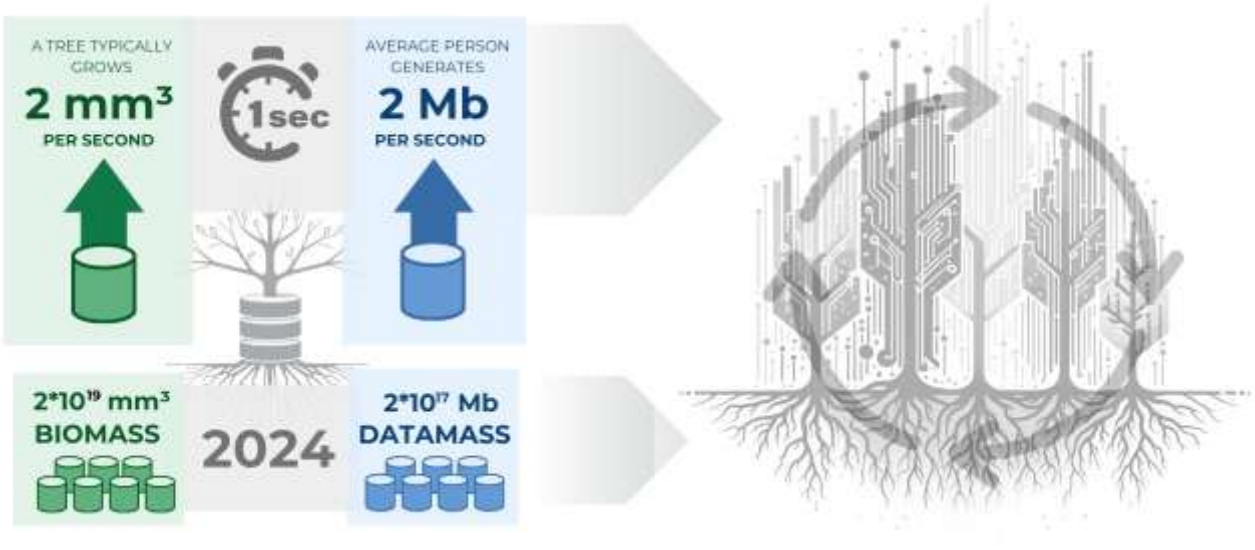


Fig. 1.3-3 El almacenamiento diario de datos por parte de cada empleado en los servidores de la empresa contribuye al crecimiento constante de los datos.

En el mundo actual de la construcción, incluso las pequeñas empresas generan a diario una enorme cantidad de información multiformato, y la huella digital incluso de una pequeña empresa de construcción puede alcanzar decenas de gigabytes al día, desde maquetas y planos hasta registros fotográficos y sensores in situ. Si suponemos que cada técnico genera de media unos 1,7 MB de datos por segundo, esto equivale a unos 146 GB al día, o 53 TB al año (Fig. 1.3-3).

Cuando un equipo de 10 personas trabaja activamente durante sólo 3 horas diarias, la cantidad acumulada de información generada al día alcanza los 180 gigabytes (Fig. 1.3-4).

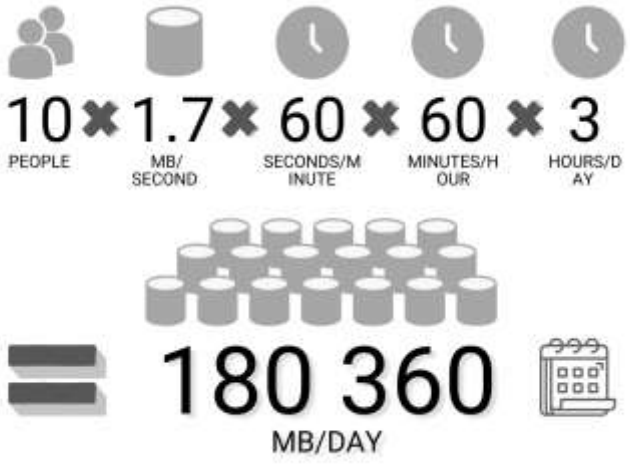


Fig. 1.3-4 Una empresa de 10 personas genera aproximadamente 50-200 gigabytes de datos al día.

Suponiendo que el 30% de los datos de trabajo sean nuevos (el resto se sobrescribe o se borra), una empresa de 10 personas puede crear del orden de varios cientos de gigabytes de datos nuevos al mes (las cifras reales dependen del tipo de negocio que haga la empresa)

Así pues, está claro: no solo estamos generando cada vez más datos, sino que nos enfrentamos a una necesidad creciente de su gestión eficiente, almacenamiento y disponibilidad a largo plazo. Y mientras que antes los datos podían "reposar" en servidores locales sin coste alguno, en el contexto de la transformación digital, cada vez son más las empresas que empiezan a utilizar soluciones en la nube como base de su infraestructura de información.

El coste del almacenamiento de datos: el aspecto económico

En los últimos años, cada vez más empresas externalizan el almacenamiento de datos a servicios en la nube. Por ejemplo, si una empresa aloja la mitad de sus datos en la nube, a un precio medio de 0,015 dólares por gigabyte al mes, sus costes de almacenamiento pueden aumentar entre 10 y 50 dólares [25] cada mes.

Para una pequeña empresa con patrones típicos de generación de datos, los costes de almacenamiento en la nube pueden oscilar entre cientos y potencialmente más de mil dólares al mes (Fig, 1.3-5) en unos pocos años, creando una carga financiera potencialmente significativa.

> Según el estudio de Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], que encuestó a 214 responsables de infraestructuras tecnológicas [26], en el que se encuestó a 214 responsables de infraestructuras tecnológicas, más de un tercio de las organizaciones están externalizando el almacenamiento para hacer frente al creciente volumen y complejidad de las operaciones de datos, y casi dos tercios de las empresas prefieren un modelo basado en suscripciones.

Fig. 1.3-5 Trasladar datos a la nube puede incrementar los costes mensuales de almacenamiento hasta en 2.000 dólares, incluso para una empresa con tan sólo 10 empleados.

La situación se complica aún más por la acelerada adopción de tecnologías en la nube como CAD (BIM), CAFM, PMIS y ERP -sistemas que aumentan aún más los costes de almacenamiento y procesamiento de datos. Como consecuencia, las empresas se ven obligadas a buscar formas de optimizar los costes y reducir la dependencia de los proveedores de la nube.

Desde 2023, con el desarrollo activo de grandes modelos lingüísticos (LLM), los enfoques del almacenamiento de datos han empezado a cambiar. Cada vez son más las empresas que piensan en recuperar el control de sus datos a medida que resulta más seguro y rentable procesar la información en sus propios servidores.

En este contexto, la tendencia a abandonar el almacenamiento y procesamiento en la nube de sólo los datos necesarios en favor del despliegue local de soluciones empresariales LLM y AI pasa a primer plano. Como señaló el CEO de Microsoft en una de sus entrevistas [27], en lugar de depender de varias aplicaciones independientes o de soluciones SaaS basadas en la nube para realizar diferentes tareas, los agentes de IA gestionarán los procesos en bases de datos, automatizando las funciones de diferentes sistemas.

> *[...] el antiguo enfoque de esta cuestión [el tratamiento de datos] era el siguiente: si nos remontamos a la forma en que las distintas aplicaciones empresariales gestionaban la integración, utilizaban conectores. Las empresas vendían licencias para estos conectores, y el modelo de negocio se formó en torno a ello. SAP [ERP] es uno de los ejemplos clásicos: sólo se podía acceder a los datos de SAP si se tenía el conector adecuado. Así que me parece que algo similar surgirá en el caso de la interacción entre agentes [AI] [...]. El enfoque, al menos el que nosotros adoptamos, es el siguiente: creo que el concepto de la existencia de aplicaciones empresariales probablemente se derrumbará en la era de los agentes [de IA]. Porque si lo piensas, son esencialmente bases de datos con un montón de lógica empresarial*
>
> - Satya Nadella, CEO de Microsoft, entrevista con el canal BG2, 2024. [28]

En este paradigma, el enfoque LLM basado en datos va más allá de los sistemas clásicos. La inteligencia artificial se convierte en un intermediario entre el usuario y los datos (Fig. 2.2-3, Fig. 2.2-4), eliminando la necesidad de múltiples interfaces intermediarias y aumentando la eficiencia de los procesos empresariales. Hablaremos más sobre este enfoque del trabajo con datos en el capítulo "Convertir el caos en orden y reducir la complejidad".

Mientras la arquitectura del futuro sigue tomando forma, las empresas ya se enfrentan a las consecuencias de decisiones pasadas. La digitalización masiva de las últimas décadas, acompañada de la introducción de sistemas dispares y la acumulación incontrolada de datos, ha provocado un nuevo problema: la sobrecarga de información.

Fronteras de la acumulación de datos: de la masa al significado

Los sistemas de las empresas modernas se desarrollan y funcionan con éxito bajo un crecimiento controlado, cuando el volumen de datos y el número de aplicaciones están en equilibrio con las capacidades de los departamentos de TI y los directivos. Sin embargo, en las últimas décadas, la digitalización ha provocado un aumento incontrolable del volumen y la complejidad de los datos, lo que ha causado un efecto de sobresaturación en el ecosistema de información de las empresas.

Hoy en día, los servidores y las instalaciones de almacenamiento están sometidos a una afluencia sin precedentes de información sin procesar y multiformato que no tiene tiempo de convertirse en abono y se vuelve rápidamente irrelevante. Los limitados recursos de las empresas no pueden hacer frente a este diluvio, y los datos se acumulan en silos aislados (los llamados "silos ") que requieren un tratamiento manual para extraer información útil.

> Como resultado, al igual que un bosque cubierto de hiedra y moho, los sistemas de gestión de las empresas modernas sufren a menudo una sobrecarga de información. En lugar de alimentar el humus de información en el corazón del ecosistema corporativo, se forman zonas aisladas de datos de distinto formato, lo que conduce inevitablemente a una disminución de la eficacia global de los procesos empresariales.

Al largo periodo de crecimiento exponencial de los datos de los últimos 40 años seguirá inevitablemente una fase de saturación y posterior enfriamiento. Cuando el almacenamiento alcance sus límites, se producirá un cambio cualitativo: los datos dejarán de ser un mero objeto de almacenamiento para convertirse en un recurso estratégico.

Con el desarrollo de la inteligencia artificial y el aprendizaje automático, las empresas tienen la oportunidad de reducir los costes de procesamiento de la información y pasar del crecimiento cuantitativo al uso cualitativo de los datos. En la próxima década, el sector de la construcción tendrá que pasar de centrarse en crear cada vez más datos a garantizar su estructura, integridad y valor analítico.

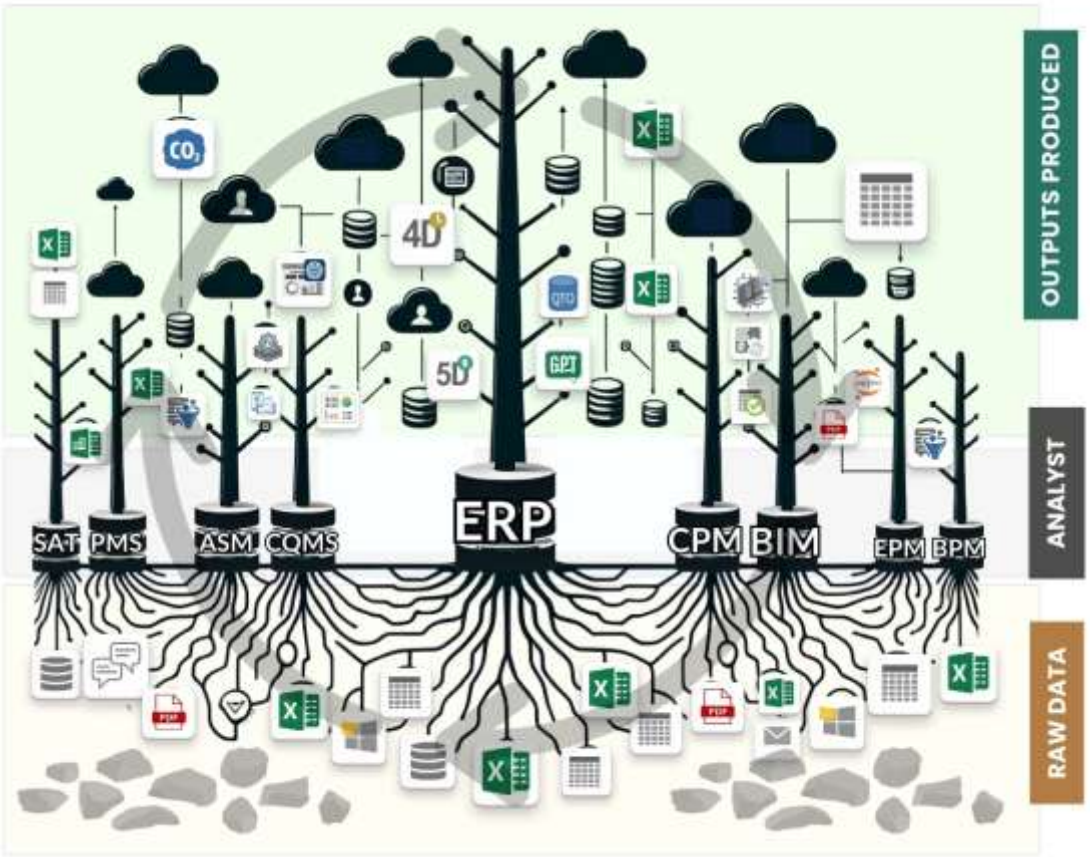


Fig. 1.3-6 Las fuentes de datos aisladas impiden compartir información entre sistemas de datos.

> El principal valor ya no está en la cantidad de información, sino en la capacidad de interpretarla automáticamente y convertirla en conocimiento aplicado útil para tomar decisiones de gestión. Para que los datos sean realmente útiles, deben gestionarse adecuadamente: recopilarse, verificarse, estructurarse, almacenarse y analizarse en el contexto de tareas empresariales específicas.

El proceso de análisis de datos en una empresa es similar al ciclo de vida y descomposición de los árboles en un bosque y la aparición de nuevos árboles jóvenes y fuertes: los árboles maduros mueren, se descomponen y se convierten en un entorno nutritivo para el nuevo crecimiento. Los procesos acabados y completados, cuando finalizan, pasan a formar parte del ecosistema de información de la empresa, convirtiéndose con el tiempo en el humus de información que alimenta el futuro crecimiento de nuevos sistemas y datos.

En la práctica, sin embargo, este ciclo suele romperse. En lugar de una renovación orgánica, se forma un caos de capas, similar a los estratos geológicos, donde los nuevos sistemas se superponen a los antiguos sin una integración y estructuración profundas. Como resultado, surgen "silos" de información dispares que obstaculizan la circulación del conocimiento y complican la gestión de los datos.

Próximos pasos: de la teoría de los datos al cambio práctico

La evolución de los datos en la construcción es un viaje desde las tablillas de arcilla hasta las modernas plataformas modulares. El reto actual no es recopilar información, sino crear un marco que convierta datos dispares y diversos en un recurso estratégico. Tanto si su función es la de director de empresa como la de ingeniero, comprender el valor de los datos y cómo trabajar con ellos será una habilidad clave en el futuro.

Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar los planteamientos comentados en sus tareas cotidianas:

- Realizar una auditoría personal de los flujos de información
  - ☐ Haz una lista de todos los sistemas y aplicaciones con los que trabajas a diario
  - ☐ Marque dónde pasa más tiempo buscando o volviendo a comprobar los datos
  - ☐ Identifique sus principales fuentes de información
  - ☐ Analice su panorama actual de aplicaciones para detectar redundancias y duplicidad de funciones.
- Esforzarse por avanzar en los procesos por niveles de madurez analítica
  - ☐ Comience sus tareas con análisis descriptivos (¿qué ha pasado?)
  - ☐ Introducir gradualmente un diagnóstico (¿por qué ha ocurrido esto?)
  - ☐ Piense en cómo en los procesos puede pasar a un análisis predictivo (¿qué ocurrirá?) y prescriptivo (¿qué hacer?).
- Empiece a estructurar sus datos de trabajo
  - ☐ Implante un sistema unificado para nombrar los archivos y carpetas que utiliza con frecuencia en su trabajo.
  - ☐ Cree plantillas para documentos e informes de uso frecuente
  - ☐ Archivar periódicamente los proyectos finalizados con una estructura clara

Aunque no puedas cambiar toda la infraestructura de información de tu equipo o empresa, empieza por tus propios procesos y pequeñas mejoras en tu trabajo diario. Recuerde que el valor real de los datos no reside en su volumen, sino en la capacidad de extraer de ellos información procesable. Incluso los conjuntos de datos pequeños pero adecuadamente estructurados y analizados pueden tener un impacto significativo cuando se integran en los procesos de toma de decisiones.

En las siguientes partes del libro, pasaremos a métodos y herramientas específicos para trabajar con datos, estudiaremos formas de transformar información no estructurada en conjuntos estructurados, exploraremos tecnologías de automatización de análisis y detallaremos cómo crear un ecosistema de análisis eficaz en una empresa de construcción.

II PARTE

CÓMO EL SECTOR DE LA CONSTRUCCIÓN SE AHOGA EN EL CAOS DE LOS DATOS

La segunda parte se dedica a analizar críticamente los retos a los que se enfrentan las empresas de construcción a la hora de gestionar volúmenes de datos cada vez mayores. Se examinan en detalle las consecuencias de la fragmentación de la información y el fenómeno de los "datos en silos", que dificulta la toma de decisiones eficaz. Se investigan los problemas del enfoque HiPPO (Highest Paid Person's Opinion) y su impacto en la calidad de las decisiones de gestión en los proyectos de construcción. Se evalúa el impacto de los procesos empresariales dinámicos y su creciente complejidad en los flujos de información y la eficiencia operativa. Se dan ejemplos concretos de cómo la excesiva complejidad de los sistemas aumenta los costes y reduce la flexibilidad de las organizaciones. Se presta especial atención a las limitaciones creadas por los formatos propietarios y a las perspectivas de uso de estándares abiertos en el sector de la construcción. Se presenta el concepto de avanzar hacia ecosistemas de software basados en IA y LLM, que minimizan la complejidad excesiva y las barreras técnicas.

CAPÍTULO 2.1.
FRAGMENTACIÓN Y SILOS DE DATOS

Cuantas más herramientas, ¿más eficaz será la empresa?

A primera vista, puede parecer que más herramientas digitales conducen a una mayor eficiencia. En la práctica, sin embargo, no es así. Con cada nueva solución, ya sea un servicio en la nube, un sistema heredado u otro informe de Excel, una empresa añade otra capa a su panorama digital, una capa que a menudo no está integrada con el resto (Fig, 2.1-1).

Los datos pueden compararse con el carbón o el petróleo: tardan años en acumularse, compactados bajo capas de caos, errores, procesos desestructurados y formatos olvidados. Para extraer de ellos información realmente útil, las empresas deben vadear literalmente capas de soluciones obsoletas y ruido digital.

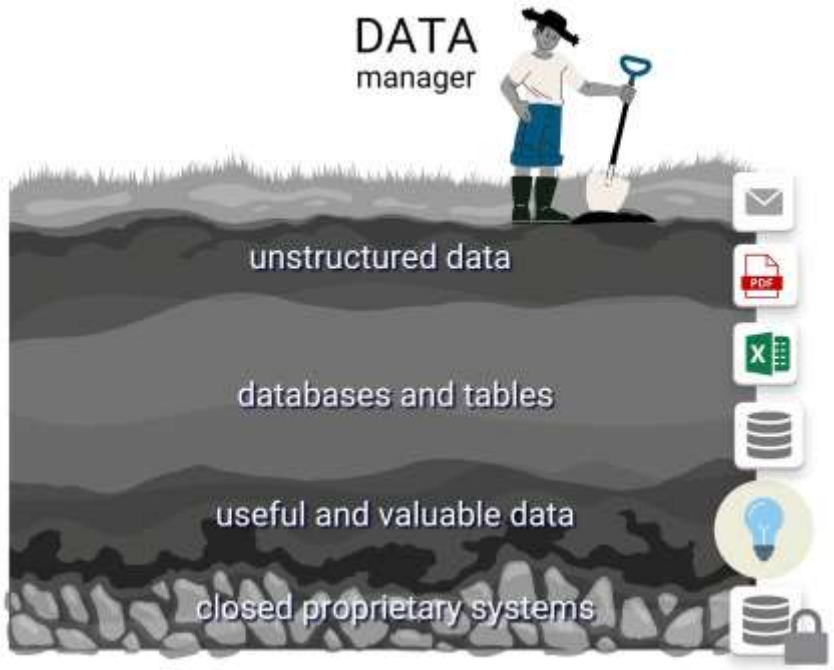


Fig. 2.1-1 Los datos diversos forman capas compartimentadas: incluso las ideas "de oro" se pierden en las rocas geológicas de la complejidad sistémica.

Cada nueva aplicación deja tras de sí un rastro: un archivo, una tabla o todo un "silo" aislado en el servidor. Una capa es de arcilla (datos obsoletos y olvidados), otra de arena (tablas e informes dispares) y la tercera de granito (formatos propietarios cerrados que no se pueden integrar). Con el tiempo, el entorno digital de una empresa se parece cada vez más a un depósito de acumulación incontrolada de información, donde el valor se pierde en las profundidades de los servidores de la empresa.

Con cada nuevo proyecto y cada nuevo sistema, no sólo la infraestructura se vuelve más compleja, sino

también el camino hacia los datos útiles de calidad. Llegar a la "roca" valiosa requiere una limpieza profunda, estructurar la información, "trocearla", agruparla en trozos significativos y extraer ideas estratégicamente importantes mediante análisis y modelización de datos.

> *Los datos son algo valioso, y durarán más que los propios sistemas [que procesan los datos] [29].*
>
> - Tim Berners-Lee, padre de la World Wide Web y creador del primer sitio web

Antes de que los datos puedan convertirse en "algo valioso" y en una base fiable para la toma de decisiones, deben someterse a una cuidadosa preparación. Es un preprocesamiento adecuado lo que convierte los datos dispares en experiencia estructurada, en humus de información útil, que luego se convierte en una herramienta de previsión y optimización.

> Existe la idea errónea de que se necesitan datos perfectamente limpios para empezar a analizarlos, pero en la práctica, poder trabajar con datos sucios es una parte esencial del proceso.

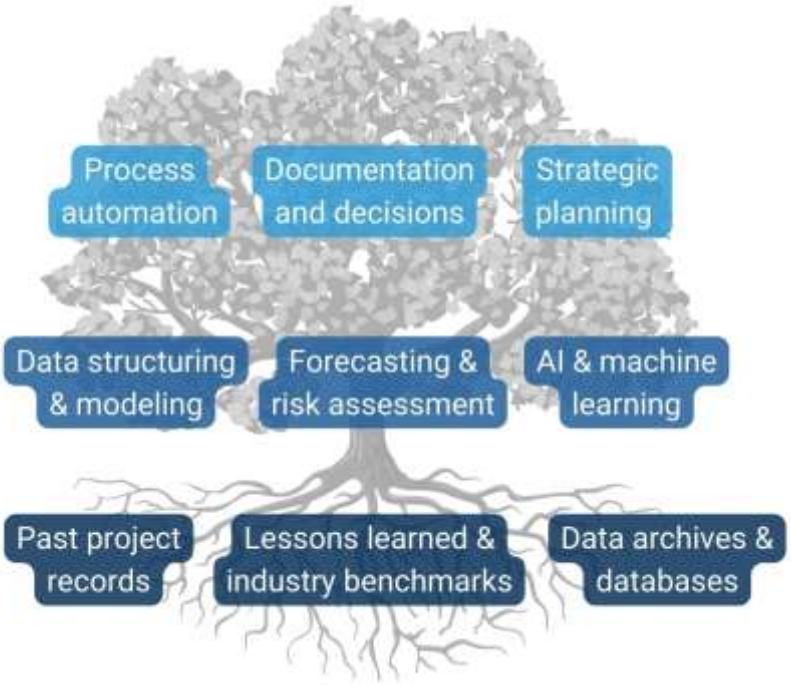


Fig. 2.1-2 Los datos son el sistema raíz y los cimientos de la empresa, que a su vez se basa en los procesos de toma de decisiones.

A medida que la tecnología sigue avanzando, su empresa también debe avanzar y aprender a crear valor a partir de los datos. Al igual que las empresas petroleras y carboneras están construyendo la infraestructura

necesaria para extraer minerales, las empresas también deben aprender a gestionar el flujo de nueva información en sus propios servidores y extraer ideas valiosas de los datos no utilizados, sin formato y obsoletos, convirtiéndolos en un recurso estratégico.

Crear campos (almacenes de datos) es el primer paso. Ni siquiera las herramientas más potentes resuelven el problema del aislamiento de los datos y de los datos multiformato si las empresas siguen funcionando con sistemas aislados. Cuando los datos existen separados unos de otros, sin cruzarse ni compartir información, las empresas se enfrentan al efecto "silo de datos". En lugar de una infraestructura única y coherente, las empresas se ven obligadas a gastar recursos en fusionar y sincronizar datos.

Silos de datos y su impacto en el rendimiento de la empresa

Imagina que estás construyendo una urbanización, pero cada equipo tiene su propio proyecto. Unos construyen los muros, otros las comunicaciones y otros las carreteras sin consultarse entre sí. El resultado es que las tuberías no coinciden con las aberturas de los muros, los huecos de los ascensores no se corresponden con las plantas y hay que desmontar y volver a colocar las calles.

Esta situación no es sólo un escenario hipotético, sino una realidad de muchos proyectos de construcción modernos. Debido al gran número de contratistas generales y subcontratistas que trabajan con sistemas diferentes y sin un único centro de coordinación, el proceso se convierte en una serie de aprobaciones interminables, repeticiones y conflictos. Todo ello provoca importantes retrasos y múltiples costes del proyecto.

Una situación clásica en una obra es simple: el encofrado está listo, pero la entrega de armadura no ha llegado a tiempo. Al comprobar la información en varios sistemas, la comunicación es aproximadamente la siguiente:

- **El capataz** de la obra del día 20 escribe al jefe de proyecto**:** *"Hemos terminado de colocar el encofrado, ¿dónde están las barras de refuerzo?*".
- **Jefe de proyecto** (PMIS) al departamento de compras**:** - "*El encofrado está listo. En mi sistema [PMIS] pone que las barras de refuerzo debían llegar el día 18. ¿Dónde están? ¿Dónde están las barras de refuerzo?"*
- **Especialista en la cadena de suministro** (ERP): - "Nuestro ERP dice que la entrega será el día 25".
- **Ingeniero de datos** o departamento informático (responsable de las integraciones): - En PMIS la fecha es el 18, en ERP es el 25. No existe un vínculo OrderID entre ERP y PMIS, por lo que los datos no están sincronizados. Este es un ejemplo típico de brecha de información.
- **De Jefe de Proyecto** a **Director General**: "*La entrega de los accesorios se retrasa, la obra está parada y no está claro quién es el responsable*".

La causa del incidente fue el aislamiento de los datos en sistemas dispares. Integrando y unificando las fuentes de datos, creando un único repositorio de información y automatizando mediante ETL -tools (Apache NiFi, Airflow o n8n), pueden eliminarse los silos entre sistemas. Estos y otros métodos y herramientas se tratarán en detalle en secciones posteriores del libro.

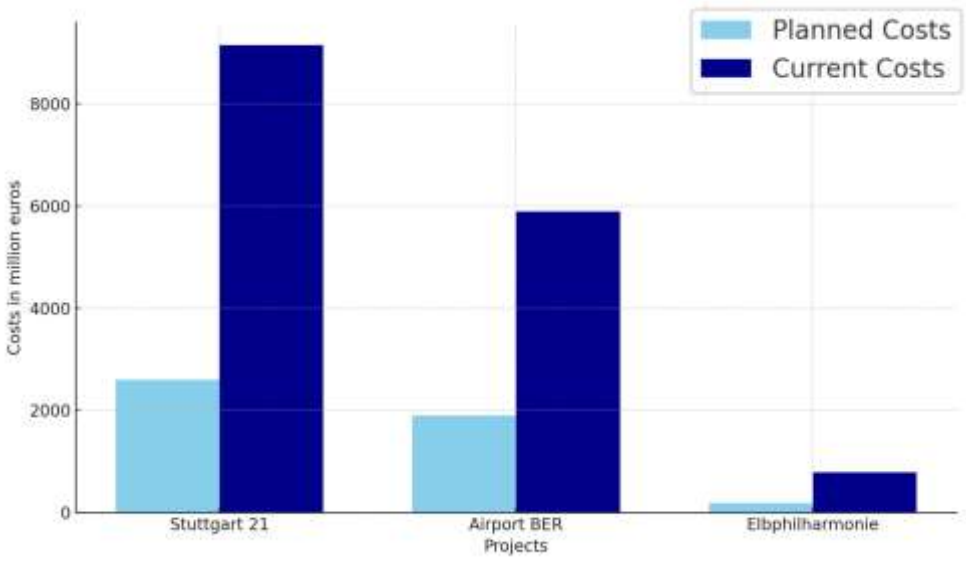


Fig. 2.1-3 Comparación de los costes previstos y reales de los grandes proyectos de infraestructuras en Alemania.

Lo mismo ocurre con los sistemas empresariales: primero se crean soluciones aisladas y luego hay que gastar enormes presupuestos para integrarlas y armonizarlas. Si los modelos de datos y comunicación se hubieran pensado desde el principio, no habría ninguna necesidad de integración. Los datos aislados crean el caos en el mundo digital, como un proceso de construcción descoordinado.

> Según el estudio de KPMG para 2023 "Cue construction 4.0: Time to make or break", sólo el 36% de las empresas comparten datos de forma eficaz entre departamentos, mientras que el 61% se enfrenta a graves problemas debido a "silos" de datos aislados [30].

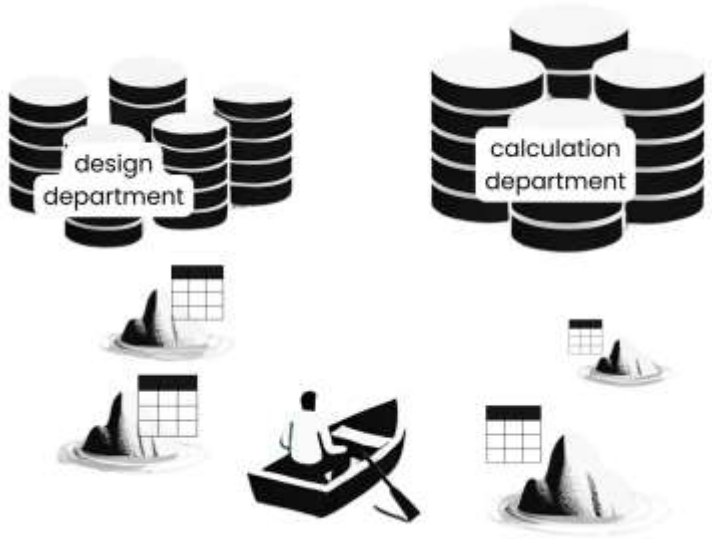


Fig. 2.1-4 Años de datos difíciles de recopilar se acumulan en "silos" de almacenamiento aislados con el riesgo de no ser utilizados nunca.

Los datos de las empresas se almacenan en sistemas aislados, como árboles individuales dispersos por el paisaje. Cada uno contiene información valiosa, pero la falta de conexiones entre ellos impide crear un ecosistema único e interconectado. Esta compartimentación obstaculiza el flujo de datos y limita la capacidad de la organización para ver el panorama completo. Conectar estos silos es un proceso extremadamente largo y complejo de cultivo de micelio de hongos a nivel de gestión para aprender a transferir piezas individuales de información entre sistemas.

Según un estudio del FEM de 2016, uno de los principales obstáculos para la transformación digital es la falta de estándares de datos comunes y la fragmentación.

> *El sector de la construcción es uno de los más fragmentados del mundo y depende de la interacción fluida de todos los participantes en la cadena de valor [5].*
>
> - Foro Económico Mundial 2016: Configurar el futuro de

Diseñadores, gestores, coordinadores y desarrolladores prefieren a menudo trabajar de forma autónoma, evitando las complejidades de la coordinación. Esta inclinación natural conduce a la creación de "silos" de información en los que los datos quedan aislados dentro de sistemas separados. Cuantos más sistemas aislados haya, más difícil será conseguir que trabajen juntos. Con el tiempo, cada sistema adquiere su propia base de datos y un departamento de apoyo especializado de gestores (Fig, 1.2-4), lo que complica aún más la integración.

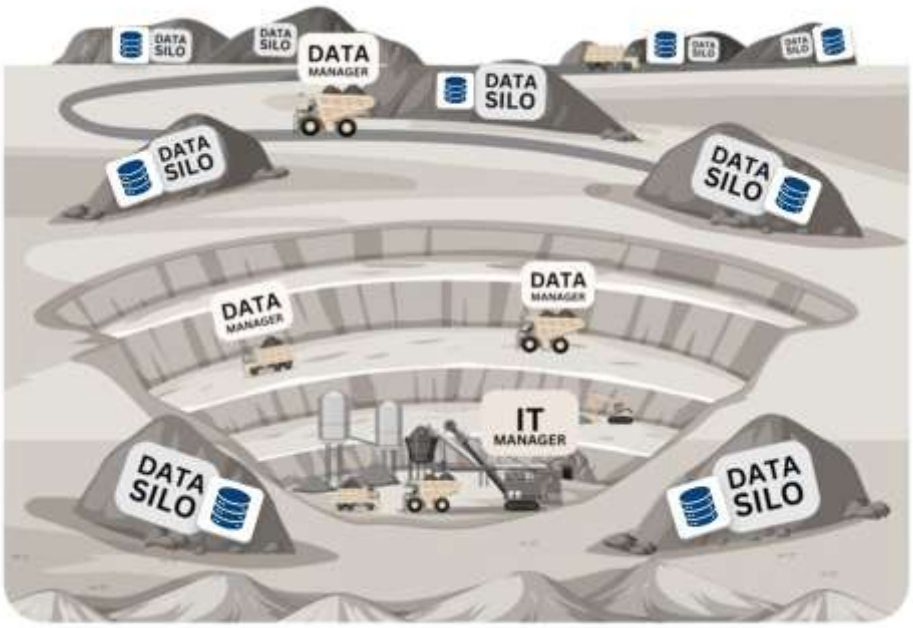


Fig. 2.1-5 Cada sistema tiende a crear su propio silo de datos que deben ser procesados por las herramientas adecuadas [31].

El círculo vicioso de los sistemas corporativos es el siguiente: las empresas invierten en complejas soluciones iso leadas, luego afrontan elevados costes para su integración y los desarrolladores, conscientes

de la complejidad de combinar sistemas, prefieren trabajar en sus ecosistemas cerrados. Todo ello aumenta la fragmentación del panorama informático y dificulta la migración a nuevas soluciones (Fig, 2.1-5). Los directivos acaban criticando los silos de datos, pero rara vez analizan sus causas y cómo evitarlos. Los directivos se quejan de la obsolescencia de los sistemas informáticos, pero sustituirlos requiere una inversión considerable y rara vez da los resultados esperados. En consecuencia, incluso los intentos de combatir el problema suelen empeorar las cosas.

La principal razón de esta desconexión es la priorización de las aplicaciones sobre los datos. Las empresas primero desarrollan sistemas separados o compran soluciones listas para usar a los proveedores, y luego intentan unificarlos creando almacenamiento y bases de datos duplicados e incompatibles.

Superar el problema de la fragmentación exige un nuevo enfoque radical: dar prioridad a los datos sobre las aplicaciones. Las empresas deben desarrollar primero estrategias de gestión de datos y modelos de datos, y después construir sistemas o adquirir soluciones que funcionen con un único conjunto de información en lugar de crear nuevas barreras.

> *Estamos entrando en un nuevo mundo en el que los datos pueden ser más importantes que el software.*
>
> - Tim O'Reilly, director general de O'Reilly Media, Inc.

> El estudio de McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) demuestra que el sector de la construcción va a la zaga de otros sectores en transformación digital [32]. Según el informe, la adopción de una gestión de datos automatizada y de plataformas digitales puede mejorar significativamente la productividad y reducir las pérdidas asociadas a la incoherencia de los procesos. Esta necesidad de transformación digital también se subraya en el informe Egan (Reino Unido, 1998) [33], que destaca el papel clave de los procesos integrados y de un enfoque colaborativo en la construcción.

> Como consecuencia, mientras que en los últimos 10.000 años el principal problema para los gestores de datos ha sido la falta de datos, con la avalancha de datos y sistemas de gestión de datos, los usuarios y gestores se enfrentan a un problema: la sobreabundancia de datos, que dificulta encontrar información legalmente correcta y de calidad.

Los silos de datos dispares conducen inevitablemente al grave problema de la reducción de la calidad de los datos. Con múltiples sistemas independientes, los mismos datos pueden existir en diferentes versiones, a menudo con valores contradictorios, lo que crea dificultades adicionales para los usuarios que necesitan determinar qué información es relevante y fiable.

Duplicación y falta de calidad de los datos como consecuencia de la desunión

Debido al problema de los silos de datos, los directivos tienen que dedicar un tiempo considerable a buscar

y conciliar datos. Para protegerse de los problemas de calidad, las empresas crean complejas estructuras de gestión de la información en las que una vertical de directivos se encarga de buscar, verificar y conciliar los datos. Sin embargo, este enfoque sólo aumenta la burocracia y ralentiza la toma de decisiones. Cuantos más datos haya, más difícil será analizarlos e interpretarlos, sobre todo si no existe una norma uniforme para almacenarlos y procesarlos.

Con la plétora de aplicaciones y sistemas informáticos que han ido creciendo como setas después de la lluvia en la última década, el problema de los silos y la calidad inadecuada de los datos ha cobrado cada vez más importancia para los usuarios finales. Los mismos datos, pero con valores diferentes, pueden encontrarse ahora en distintos sistemas y aplicaciones (Fig, 2.1-6). Esto genera dificultades a los usuarios finales cuando intentan determinar qué versión de los datos es relevante y correcta entre las muchas disponibles. Esto conduce a errores en los análisis y, en última instancia, en la toma de decisiones.

Para evitar problemas a la hora de encontrar los datos correctos, los directivos de las empresas crean una burocracia multinivel de gestores de verificación. Su tarea consiste en encontrar, comprobar y enviar rápidamente los datos necesarios en forma de tablas e informes, navegando por el laberinto de sistemas dispares.

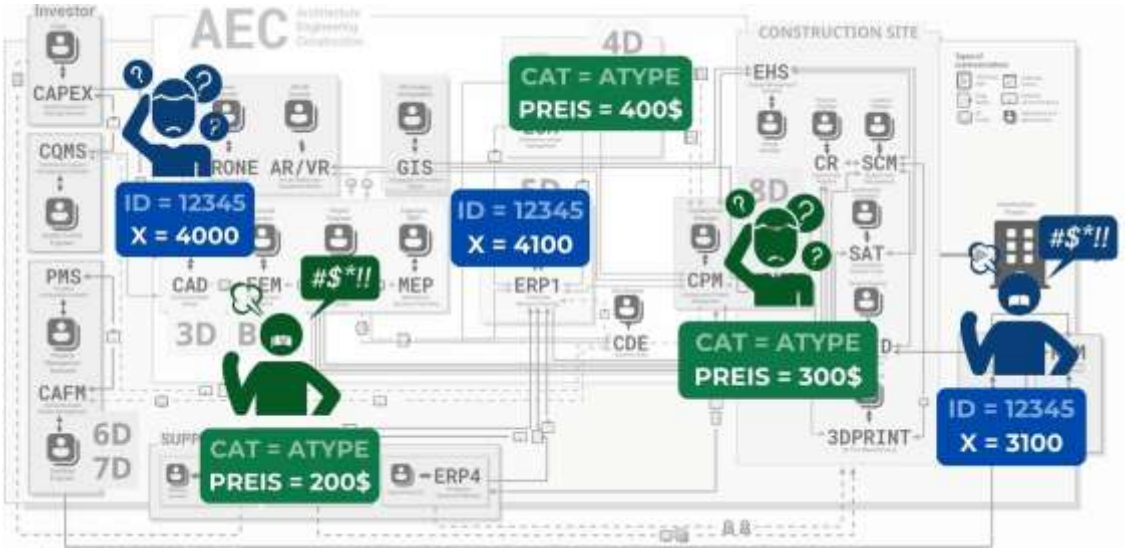


Fig. 2.1-6 Al tratar de encontrar los datos correctos, los gestores deben garantizar la calidad y la fiabilidad jurídica de los datos entre los distintos sistemas.

En la práctica, sin embargo, este modelo genera nuevas complejidades. Cuando los datos se gestionan manualmente y la información está dispersa en muchas decisiones no relacionadas entre sí, cada intento de obtener información precisa y actualizada a través de una pirámide de responsables de la toma de decisiones (Fig, 2.1-7) se convierte en un cuello de botella, que consume mucho tiempo y es propenso a errores.

La situación se ve agravada por la avalancha de soluciones digitales. El mercado del software sigue inundado de nuevas herramientas que parecen prometedoras. Pero sin una estrategia clara de gestión de datos, estas soluciones no se integran en un sistema unificado, sino que crean capas adicionales de complejidad y duplicación. Como resultado, en lugar de simplificar los procesos, las empresas se encuentran en un

entorno de información aún más fragmentado y caótico.

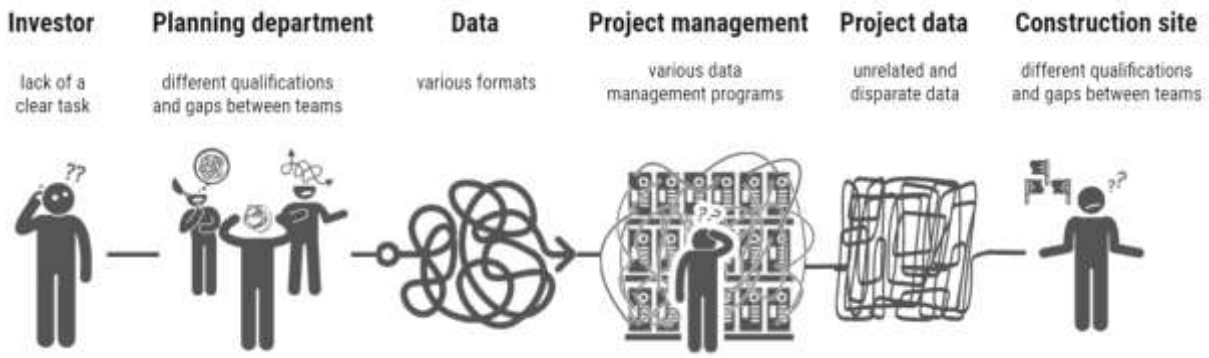


Fig. 2.1-7 La complejidad de los sistemas y la variedad de formatos de datos provocan una pérdida de coherencia en el proceso de construcción.

Todos estos problemas asociados a la gestión de una multitud de soluciones dispares llevan tarde o temprano a la dirección de la empresa a una importante constatación: no se trata del volumen de datos ni de la búsqueda de la próxima herramienta de tratamiento de datos "de talla única". La verdadera razón reside en la calidad de los datos y en cómo la organización los crea, recibe, almacena y utiliza.

> La clave del éxito sostenible no está en perseguir nuevas aplicaciones "mágicas", sino en crear una cultura de datos dentro de la empresa. Esto significa tratar los datos como un activo estratégico y hacer de la calidad, integridad y relevancia de los datos una prioridad a todos los niveles de la organización.

La solución al dilema calidad frente a cantidad radica en la creación de una estructura de datos unificada que elimine duplicaciones, suprima incoherencias y unifique los flujos de información. Esta arquitectura proporciona una fuente de datos única y fiable sobre la que tomar decisiones informadas, precisas y oportunas.

De lo contrario, como sigue ocurriendo a menudo, las empresas siguen confiando en las opiniones subjetivas y las evaluaciones intuitivas de los expertos de HiPPO en lugar de basarse en hechos fiables. En el sector de la construcción, en el que los conocimientos especializados desempeñan tradicionalmente un papel importante, esto es especialmente notable.

HiPPO o el peligro de las opiniones en la toma de decisiones

Tradicionalmente, en el sector de la construcción, las decisiones clave se toman basándose en la experiencia y el juicio subjetivo. Sin datos oportunos y fiables, los directivos de las empresas tienen que actuar a ciegas, basándose en la intuición de los empleados mejor pagados (HiPPO - Highest Paid Person's Opinion) y no en hechos objetivos (Fig. 2.1-8).

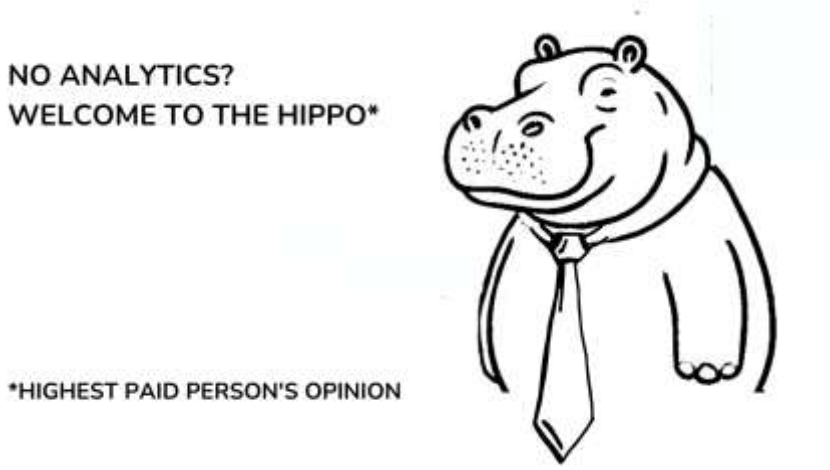


Fig. 2.1-8 A falta de análisis el negocio depende de la opinión subjetiva de profesionales experimentados.

Este enfoque puede estar justificado en un entorno estable y de cambios lentos, pero en la era de la transformación digital se convierte en un grave riesgo. Las decisiones basadas en la intuición y las conjeturas son propensas a la distorsión, a menudo se basan en hipótesis sin fundamento y no tienen en cuenta el complejo panorama que reflejan los datos

> Lo que se hace pasar por debate inteligente en el nivel decisorio de una empresa no suele basarse en nada concreto. El éxito de una empresa no debe depender de la autoridad y el sueldo de los expertos, sino de la capacidad de trabajar eficazmente con datos, identificar patrones y tomar decisiones con conocimiento de causa.

Es importante abandonar la idea de que la autoridad o la experiencia significan automáticamente que una decisión es correcta. El enfoque basado en los datos cambia las reglas del juego: los datos y los análisis, y no el cargo y el salario, son ahora la base de la toma de decisiones. Los macrodatos, el aprendizaje automático y los análisis visuales nos permiten identificar patrones y basarnos en hechos en lugar de en conjeturas (Fig. 1.1-4).

> *Sin datos, no eres más que otra persona con una opinión [34].*
>
> - W. Edwards Deming, académico y consultor de gestión

Los métodos modernos de gestión de datos también garantizan la continuidad de los conocimientos dentro de la empresa. Los procesos claramente descritos, la automatización y un enfoque sistemático permiten transferir incluso funciones clave sin perder eficacia.

Sin embargo, la confianza ciega en los datos también puede conducir a graves errores. Los datos en sí no son más que una colección de números. Sin el análisis adecuado, el contexto y la capacidad de identificar patrones, no tienen ningún valor y no pueden impulsar los procesos. La clave del éxito reside no en elegir

entre la intuición HiPPO y la analítica, sino en crear herramientas inteligentes que transformen información dispar en decisiones manejables y fundamentadas.

> En un entorno de construcción digital, no son la antigüedad y el lugar en la jerarquía los factores de éxito decisivos, sino la capacidad de respuesta, la precisión de las decisiones y la eficiencia de los recursos

Los datos son herramientas, no verdades absolutas. Deben complementar el pensamiento humano, no sustituirlo. A pesar de las ventajas de la analítica, los datos no pueden suplantar por completo la intuición y la experiencia humanas. Su función es ayudar a tomar decisiones más precisas e informadas.

La ventaja competitiva se logrará no sólo cumpliendo las normas, sino superando a los competidores en el uso eficiente de unos recursos que son iguales para todos. En el futuro, las competencias en materia de datos serán tan importantes como lo fueron en su día la alfabetización o las matemáticas. Los profesionales capaces de analizar e interpretar los datos podrán tomar decisiones más precisas, desplazando a los que sólo se basan en la experiencia personal (Fig, 2.1-9).

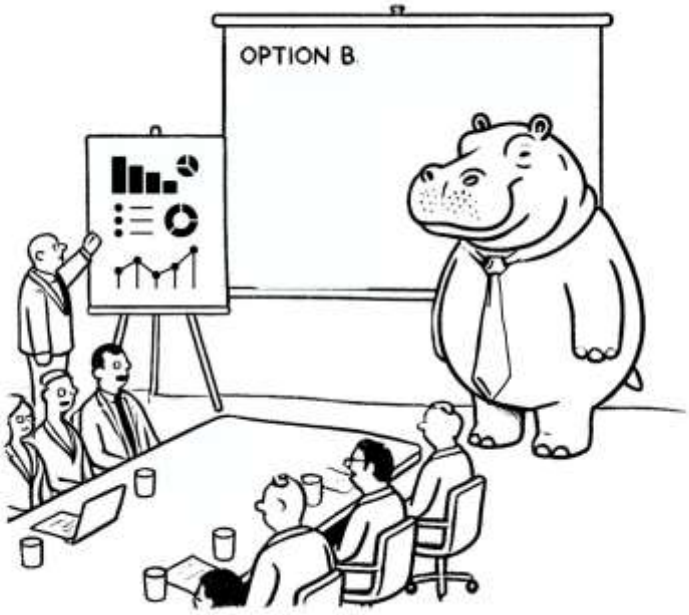


Fig. 2.1-9 Las decisiones deben basarse en un análisis objetivo, no en la opinión del empleado mejor pagado.

Gestores, especialistas e ingenieros actuarán como analistas de datos en, estudiando la estructura, la dinámica y los indicadores clave de los proyectos. Los recursos humanos se convertirán en elementos del sistema, que requerirán una personalización flexible basada en datos para maximizar la eficiencia.

> *Los errores cuando se utilizan datos inadecuados son mucho menores que cuando no se utilizan datos [35].*
>
> - Charles Babbage, inventor de la primera máquina de calcular analítica

La aparición de los big data y la introducción de los LLM (Large Language Models) han cambiado radicalmente no sólo nuestra forma de analizar, sino también la propia naturaleza de la toma de decisiones. Mientras que antes la atención se centraba en la causalidad (por qué ocurría algo -análisis de diagnóstico) (Fig. 1.1-4), hoy en día está cobrando protagonismo la capacidad de predecir el futuro (análisis predictivo) y, en el futuro, el análisis prescriptivo, en el que el aprendizaje automático y la IA sugieren la mejor opción en el proceso de toma de decisiones.

> Según el nuevo estudio de SAP™, "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Themselves" 2025 [36], el 44% de los altos ejecutivos estaría dispuesto a cambiar su decisión anterior basándose en el asesoramiento de la IA, y el 38% confiaría en la IA para tomar decisiones empresariales en su nombre. Mientras tanto, el 74% de los ejecutivos afirma que confía más en los consejos de la IA que en sus amigos y familiares, y el 55% trabaja en empresas en las que los conocimientos derivados de la IA sustituyen o a menudo eluden los métodos tradicionales de toma de decisiones, especialmente en organizaciones con ingresos anuales superiores a 5.000 millones de dólares. Además, el 48% de los encuestados utiliza herramientas de IA generativa a diario, incluido un 15% que las utiliza varias veces al día.

Con el desarrollo del LLM y los sistemas automatizados de gestión de datos, surge un nuevo reto: cómo utilizar la información de forma eficaz sin perder su valor en el caos de formatos incompatibles y fuentes heterogéneas, que se complementa con la creciente complejidad y dinámica de los procesos empresariales.

Aumento continuo de la complejidad y el dinamismo de los procesos empresariales

El sector de la construcción se enfrenta hoy a serios retos en la gestión de datos y procesos. Los principales retos son los sistemas de información aislados, la excesiva burocracia y la falta de integración entre las herramientas digitales. Estos retos se intensifican a medida que los propios procesos empresariales se vuelven más complejos, impulsados por la tecnología, las cambiantes exigencias de los clientes y la evolución de la normativa.

La singularidad de los proyectos de construcción se debe no sólo a sus peculiaridades técnicas, sino también a las diferencias en las normas nacionales y los requisitos reglamentarios de los distintos países (Fig. 4.2-10, Fig. 5.1-7). Esto requiere un enfoque flexible e individualizado de cada proyecto, difícil de aplicar en los sistemas de control modulares tradicionales. Debido a la complejidad de los procesos y a la gran cantidad de datos, muchas empresas recurren a proveedores que ofrecen soluciones especializadas. Pero el mercado está saturado: muchas empresas de nueva creación ofrecen productos similares, centrados en tareas limitadas. Como resultado, a menudo se pierde el enfoque holístico de la gestión de datos.

> La adaptación al flujo continuo de nuevas tecnologías y requisitos del mercado se está convirtiendo en un factor crítico para la competitividad. Sin embargo, las aplicaciones propietarias y los sistemas modulares existentes tienen poca capacidad de adaptación: cualquier cambio suele requerir revisiones largas y costosas por parte de desarrolladores que no siempre entienden las particularidades de los procesos de construcción.

Las empresas se encuentran rehenes del retraso tecnológico, a la espera de nuevas actualizaciones en lugar de aplicar con prontitud enfoques integrados innovadores. Como consecuencia, la estructura interna de las organizaciones de la construcción suele ser un complejo ecosistema de sistemas jerárquicos interconectados, y a menudo cerrados, coordinados a través de una red multinivel de directivos (Fig, 2.1-10).

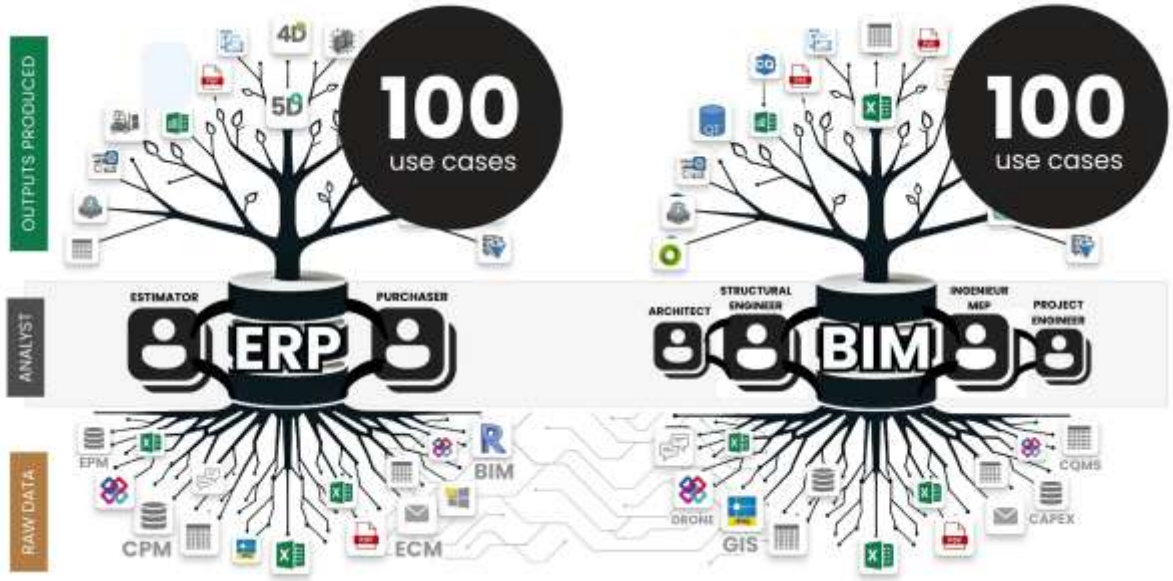


Fig. 2.1-10 Las empresas están formadas por sistemas interconectados cuya interconexión forma procesos que requieren automatización.

> Según una encuesta realizada por la Asociación Canadiense de la Construcción y KPMG Canadá en 2021 [37], solo el 25% de las empresas cree estar en una posición significativa o diferente en comparación con sus competidores en lo que respecta a la adopción de tecnología o soluciones digitales. Sólo el 23% de los encuestados afirma que sus soluciones se basan de forma significativa o importante en los datos. Al mismo tiempo, la mayoría de los encuestados calificaron su uso de otras tecnologías de puramente experimental o admitieron no utilizarlas en absoluto.

Esta reticencia a participar en experimentos tecnológicos es especialmente evidente en los grandes proyectos de infraestructuras, donde los errores pueden costar millones de dólares. Incluso las tecnologías más avanzadas -gemelos digitales, análisis predictivo - suelen encontrar resistencia no por su eficacia, sino por la falta de fiabilidad demostrada en proyectos reales.

> Según el informe del Foro Económico Mundial (FEM) "Shaping the Future of Construction" [5], la introducción de nuevas tecnologías en la construcción se enfrenta no sólo a dificultades técnicas, sino también a barreras psicológicas por parte de los clientes. (5), la introducción de nuevas tecnologías en la construcción se enfrenta no sólo a dificultades técnicas, sino también a barreras psicológicas por parte de los clientes. Muchos clientes temen que el uso de soluciones avanzadas convierta sus proyectos en un lugar de experimentación y los convierta en "conejillos de indias", y las consecuencias imprevisibles pueden acarrear costes y riesgos adicionales.

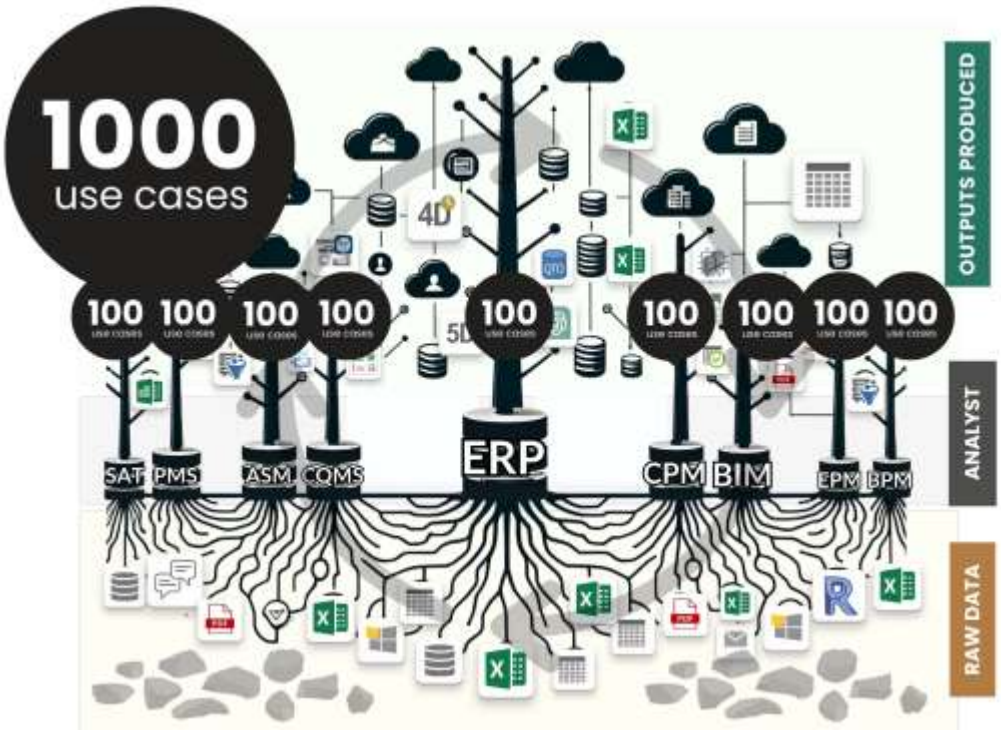


Fig. 2.1-11 Para cada caso de uso de datos, el mercado de soluciones ofrece aplicaciones para optimizar y automatizar procesos.

> El sector de la construcción es muy diverso: los distintos proyectos tienen diferentes requisitos, peculiaridades regionales, normativas legales de clasificación (Fig. 4.2-10), normas de cálculo (Fig. 5.1-7), etc. Por lo tanto, es prácticamente imposible crear una aplicación o sistema universal propio que se adapte perfectamente a todos estos requisitos y especificidades de los proyectos.

En un intento por hacer frente a la creciente complejidad de los sistemas y a la dependencia de los proveedores de software, cada vez se es más consciente de que la clave para una gestión eficaz de los datos no es sólo la apertura y la normalización, sino también la simplificación de la propia arquitectura de procesos. La creciente complejidad y dinamismo de los procesos empresariales exige nuevos planteamientos, en los que la prioridad pase de acumular datos a estructurarlos y organizarlos. Este cambio será el siguiente paso

en el desarrollo de la industria de la construcción, que marcará el fin de la era del dominio de los proveedores de software y el comienzo de la era de la organización significativa de la información.

La constatación de las limitaciones de las soluciones de talla única y la vulnerabilidad ante una complejidad cada vez mayor está provocando un cambio de prioridades: de las plataformas cerradas y el acaparamiento de datos a la transparencia, la adaptabilidad y el tratamiento estructurado de la información. Este cambio de mentalidad refleja cambios más amplios en la economía y la tecnología mundiales, descritos a través del prisma de las llamadas "revoluciones industriales". Para entender hacia dónde se dirige la construcción y su dirección futura, es necesario considerar el lugar que ocupa el sector en el contexto de la Cuarta y Quinta Revoluciones Industriales: de la automatización y la digitalización a la personalización, los estándares abiertos y el modelo de datos basado en los servicios.

La Cuarta Revolución Industrial (Industria 4.0) y la Quinta Revolución Industrial (Industria 5.0) en la construcción

Las etapas tecnológicas y económicas son conceptos teóricos utilizados para describir y analizar la evolución de la sociedad y la economía en distintas fases de desarrollo. Pueden ser interpretados de forma diferente por distintos investigadores y expertos.

- **La cuarta revolución industrial** (4IR o Industria 4.0) está relacionada con las tecnologías de la información, la automatización, la digitalización y la globalización. Uno de sus elementos clave es la creación de soluciones informáticas propias, es decir, productos digitales especializados diseñados para tareas y empresas específicas. Estas soluciones suelen convertirse en una parte importante de la infraestructura informática, pero son poco escalables sin modificaciones adicionales.
- **La Quinta Revolución Industrial (5IR)** se encuentra ahora en una fase de conceptualización y desarrollo más temprana que la 4IR. Entre sus principios básicos figura una mayor personalización de productos y servicios. La 5IR es un movimiento hacia una actividad económica más adaptable, flexible y personalizada, centrada en la personalización, la consultoría y los modelos orientados al servicio. Un aspecto clave del quinto modo económico es el uso de datos para la toma de decisiones, que es prácticamente imposible sin el uso de datos abiertos y herramientas abiertas (Fig, 2.1-12).

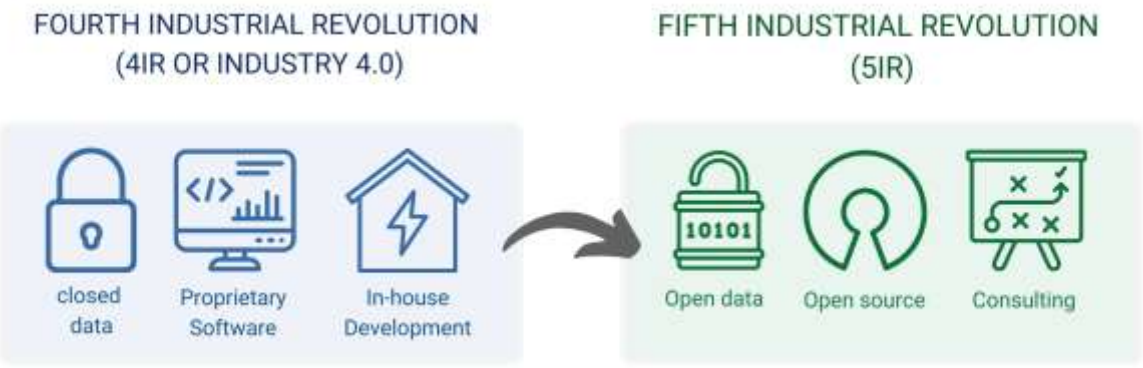


Fig. 2.1-12 El cuarto patrón se centra en las soluciones, mientras que el quinto se

centra en la personalización y los datos.

> La creación de una aplicación para empresas del sector de la construcción para su uso en diez o cien organizaciones no garantiza su ampliación con éxito a otras empresas, regiones o países sin modificaciones y mejoras significativas. La probabilidad de escalar con éxito este tipo de soluciones sigue siendo baja, ya que cada organización tiene procesos, requisitos y condiciones únicos que pueden requerir adaptaciones personalizadas.

Es importante comprender que ya hoy en día el éxito de la integración de soluciones tecnológicas implica un enfoque profundamente personalizado de cada proceso, proyecto y empresa. Esto significa que incluso después de que se haya desarrollado un marco, herramienta o programa universal, será necesario adaptarlo y personalizarlo detalladamente para satisfacer los requisitos únicos y las condiciones de cada empresa y proyecto concretos.

> Según el informe de PwC "Decoding the Fifth Industrial Revolution" [38] [38], cerca del 50% de los altos ejecutivos de diversas industrias confían este año en la integración de tecnología avanzada y conocimientos humanos. Este enfoque les permite adaptarse rápidamente a los cambios en el diseño de los productos o a las exigencias de los clientes, creando una producción personalizada.

Cada proceso requiere el desarrollo de una función o aplicación única, lo que, dado el tamaño de la industria global de la construcción y la diversidad de proyectos, conduce a la existencia de un enorme número de casos de negocio que representan cada vez una lógica única de Pipeline (Fig. 2.1-13). Cada uno de estos casos tiene sus propias peculiaridades y requiere un enfoque personalizado. En el capítulo dedicado al aprendizaje automático y al análisis sintáctico del conjunto de datos del Titanic (Fig. 9.2-9) veremos con más detalle la variedad de posibles soluciones al mismo problema analítico en el contexto de distintos enfoques.

> **Pipeline** en el contexto de los procesos digitales es una secuencia de actividades, procesos y herramientas que permiten un flujo automatizado o estructurado de datos y trabajo a través de las diferentes etapas del ciclo de vida del proyecto.

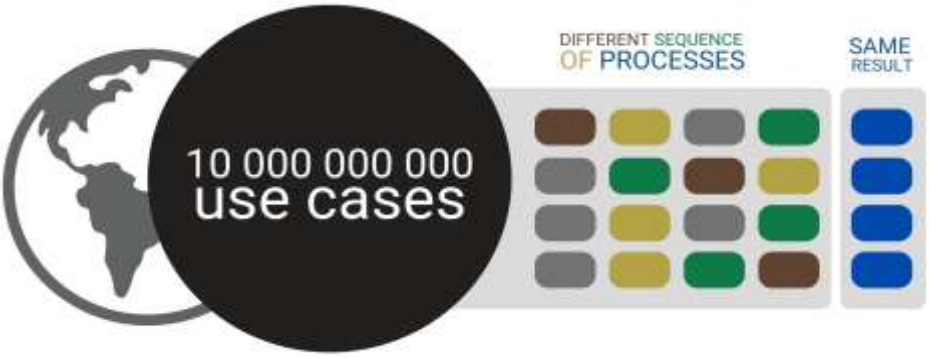


Fig. 2.1-13 La individualidad y variabilidad de los casos empresariales imposibilita los

intentos de crear plataformas y herramientas cerradas escalables.

Nuestras vidas ya han cambiado en muchos aspectos bajo la influencia de la transformación digital, y hoy podemos hablar del advenimiento de una nueva etapa en el desarrollo económico del sector de la construcción. En esta "nueva economía", la competencia se organizará según reglas diferentes: quien sea capaz de convertir eficazmente el conocimiento público y los datos abiertos en productos y servicios demandados obtendrá una ventaja clave en las condiciones de la quinta revolución industrial.

> Como señala la economista Kate Maskus en el libro "Derechos privados y problemas públicos: la economía mundial de la propiedad intelectual en el siglo XXI"[39] 2012[39], "vivimos en una economía global del conocimiento, y el futuro pertenece a quienes saben convertir los descubrimientos científicos en mercancías".

La transición al quinto modo económico implica un cambio de enfoque, de las soluciones informáticas cerradas a los estándares y plataformas abiertos. Las empresas empezarán a abandonar los productos de software tradicionales en favor de modelos orientados a los servicios en los que los datos, y no las tecnologías patentadas, se convertirán en el principal activo.

> El estudio Harvard Business School 2024 [40] muestra el enorme valor económico del software de código abierto (Open Source Software, OSS). Según el estudio, el OSS está presente en el 96% de todos los códigos de software, y algunos programas comerciales constan de un 99,9% de componentes OSS. Sin OSS, las empresas gastarían 3,5 veces más en software.

> La creación de ecosistemas empresariales, siguiendo las tendencias mundiales, cambiará gradualmente hacia un quinto paradigma económico, en el que los servicios de análisis y consultoría centrados en los datos serán más prioritarios que las soluciones aisladas y cerradas con escenarios de uso rígidamente definidos.

La era de la digitalización cambiará el equilibrio de poder en el sector: en lugar de depender de las soluciones de los proveedores, las empresas basarán su competitividad en su capacidad para utilizar los datos con eficacia. Como resultado, el sector de la construcción pasará de sistemas rígidos heredados a ecosistemas flexibles y adaptables en los que los estándares abiertos y las herramientas interoperables serán la base de la gestión de proyectos. El fin de la era del dominio de los proveedores de aplicaciones creará un nuevo entorno en el que el valor no vendrá definido por la posesión de código fuente cerrado y conectores especializados, sino por la capacidad de convertir los datos en una ventaja estratégica.

CAPÍTULO 2.2.
TRANSFORMAR EL CAOS EN ORDEN Y REDUCIR LA COMPLEJIDAD

Código redundante y sistemas cerrados como barrera a la mejora de la productividad

En las últimas décadas, los cambios tecnológicos en el ámbito de las TI han sido impulsados principalmente por los proveedores de software. Ellos marcan el rumbo del desarrollo, determinando qué tecnologías deben adoptar las empresas y cuáles deben dejar atrás. En la época en que se pasaba de las soluciones en silos a las bases de datos centralizadas y los sistemas integrados, los vendedores promovían los productos con licencia, que proporcionaban control sobre el acceso y la escalabilidad. Más tarde, con la llegada de las tecnologías en la nube y los modelos de software como servicio (SaaS), este control evolucionó hacia un modelo de suscripción, consolidando a los usuarios como clientes fieles de los servicios digitales.

Este enfoque ha dado lugar a una paradoja: a pesar de los volúmenes sin precedentes de código de programa creado, sólo una pequeña parte se utiliza realmente. Tal vez haya cientos o miles de veces más código del necesario, porque los mismos procesos empresariales se describen y duplican en docenas o cientos de programas de diferentes maneras, incluso dentro de la misma empresa. Al mismo tiempo, ya se han pagado los costes de desarrollo, que son irrecuperables. Sin embargo, la industria sigue reproduciendo este ciclo, creando nuevos productos con un valor añadido mínimo para el usuario final, más a menudo bajo la presión de las expectativas del mercado que de las necesidades reales.

> Según la Guía de Estimación de Costes de Desarrollo de Software de la Universidad de Adquisiciones de Defensa (DAU) [41], el coste del desarrollo de software puede variar significativamente en función de varios factores, como la complejidad del sistema y la tecnología elegida. Históricamente, los costes de desarrollo en 2008 han rondado los 100 dólares por línea de código fuente (SLOC), mientras que los costes de mantenimiento pueden ascender a 4.000 dólares por SLOC.

Sólo uno de los componentes de las aplicaciones CAD -el núcleo geométrico- puede tener decenas de millones de líneas de código (Fig. 6.1-5). Una situación similar se observa en los sistemas ERP (Fig. 5.4-4), sobre cuya complejidad volveremos en la quinta parte del libro. Sin embargo, una mirada más atenta revela que gran parte de este código no añade valor, sino que simplemente actúa como "cartero", moviendo mecánicamente los datos entre la base de datos, la API, la interfaz de usuario y otras tablas del sistema. A pesar del mito popular sobre la importancia crítica de la llamada lógica de negocio, la cruda realidad es mucho más prosaica: las bases de código modernas están llenas de bloques de plantillas obsoletos (código heredado) cuyo único propósito es garantizar la transferencia de datos entre tablas y componentes sin afectar a la toma de decisiones ni a la eficiencia empresarial.

Como resultado, las soluciones cerradas que procesan datos de varias fuentes se convierten inevitablemente en confusos "ecosistemas espagueti". Estos sistemas complejos y entrelazados sólo pueden ser manejados por un ejército de gestores que trabajan de forma semirrutinaria. Esta organización de la gestión de datos no sólo es ineficaz en términos de recursos, sino que también crea vulnerabilidades críticas

en los procesos empresariales, haciendo que la empresa dependa de un estrecho círculo de especialistas que entienden cómo funciona este laberinto tecnológico.

El continuo aumento de la cantidad de código, el número de aplicaciones y la creciente complejidad de los conceptos ofrecidos por los proveedores ha conducido a un resultado natural: el aumento de la complejidad del ecosistema informático en la construcción. Esto ha hecho ineficaz la aplicación práctica de la digitalización mediante el aumento del número de aplicaciones en el sector. Los productos de software creados sin prestar la debida atención a las necesidades de los usuarios suelen requerir importantes recursos para su implantación y asistencia, pero no aportan los beneficios esperados.

> Según el estudio de McKinsey "Aumentar la productividad de la construcción" [42], en las dos últimas décadas, el crecimiento mundial de la productividad laboral en la construcción ha sido, por término medio, de sólo el 1% anual, frente a un crecimiento del 2,8% en el conjunto de la economía mundial y del 3,6% en la industria manufacturera. En Estados Unidos, la productividad laboral de la construcción por trabajador se ha reducido a la mitad desde los años sesenta [43].

La creciente complejidad de los sistemas, el aislamiento y los datos cerrados han dificultado la comunicación entre profesionales, lo que convierte al sector de la construcción en uno de los menos eficientes (Fig, 2.2-1). a 22 billones de dólares en 2040, lo que exigirá importantes mejoras de eficiencia.

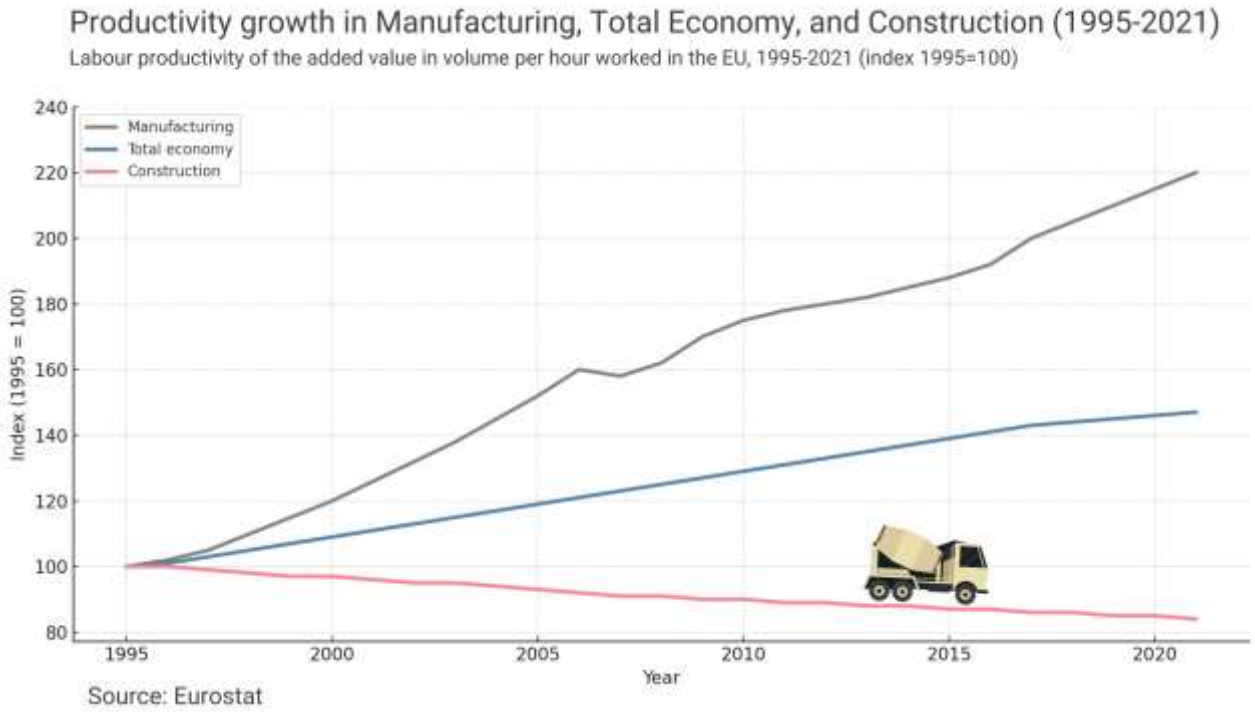


Fig. 2.2-1 Los datos cerrados y complejos y, como consecuencia, la escasa comunicación entre especialistas han llevado a la industria de la construcción a ser uno de los sectores menos eficientes de la economía (basado en [44], [45]).

> Como se subraya en el estudio de McKinsey (2024) "Garantizar la productividad de la construcción ya no es opcional", con la creciente escasez de recursos y el impulso del sector para duplicar su tasa de crecimiento, la construcción ya no puede permitirse permanecer en los niveles de productividad actuales [44]. Se prevé que los costes mundiales de la construcción aumenten de 13 billones de dólares en 2023 a niveles mucho más altos a finales de la década, lo que hace que la cuestión de la eficiencia no solo sea relevante, sino crítica.

Una de las vías clave para mejorar la eficiencia será la inevitable unificación y simplificación de las estructuras de las aplicaciones y las arquitecturas de los ecosistemas de datos. Este enfoque de la racionalización eliminará las capas redundantes de abstracción y la complejidad innecesaria que se han acumulado a lo largo de los años en los sistemas empresariales.

De los silos a un único almacén de datos

Cuantos más datos acumula una organización, más difícil resulta extraer de ellos un valor real. Debido a la naturaleza fragmentada del almacenamiento de información en silos aislados, los procesos empresariales de las empresas modernas son como constructores intentando levantar un rascacielos con materiales almacenados en miles de almacenes distintos. El exceso de información no sólo dificulta el acceso a la información legalmente relevante, sino que también ralentiza la toma de decisiones: cada paso tiene que comprobarse y confirmarse repetidamente.

> Cada tarea o proceso está conectado a una tabla o base de datos distinta, y el intercambio de datos entre sistemas requiere complejas integraciones. Los errores e incoherencias en un sistema pueden provocar fallos en cadena en los demás. Valores incorrectos, actualizaciones tardías e información duplicada obligan a los empleados a dedicar mucho tiempo a conciliar y reconciliar datos manualmente. Como resultado, la organización dedica más tiempo a lidiar con las consecuencias de la fragmentación que a desarrollar y optimizar los procesos

Este problema es universal: algunas empresas siguen luchando contra el caos, mientras que otras encuentran una solución en la integración: trasladar los flujos de información a un sistema de almacenamiento centralizado. Piense en ello como una gran tabla donde almacenar cualquier entidad relacionada con tareas, proyectos y objetos. En lugar de decenas de tablas y formatos dispares, surge un único repositorio cohesionado (Fig, 2.2-2) que permite:

- minimizar la pérdida de datos;
- eliminar la necesidad de armonizar constantemente la información;
- mejorar la disponibilidad y calidad de los datos;
- simplificar el tratamiento analítico y el aprendizaje automático

Llevar los datos a una norma común significa que, independientemente de la fuente, la información se convierte en un formato unificado y legible por máquina. Esta organización de los datos permite comprobar su integridad, analizarlos en tiempo real y utilizarlos rápidamente para tomar decisiones de gestión.

El concepto de sistemas de almacenamiento integrados y su aplicación en la analítica y el aprendizaje

automático se tratarán con más detalle en el capítulo "Almacenamiento de Big Data y aprendizaje automático". Los temas de modelado y estructuración de datos se tratarán en detalle en los capítulos "Transformación de datos en una forma estructurada" y "Cómo cambian las reglas del juego los estándares: de los archivos aleatorios a un modelo de datos elaborado".

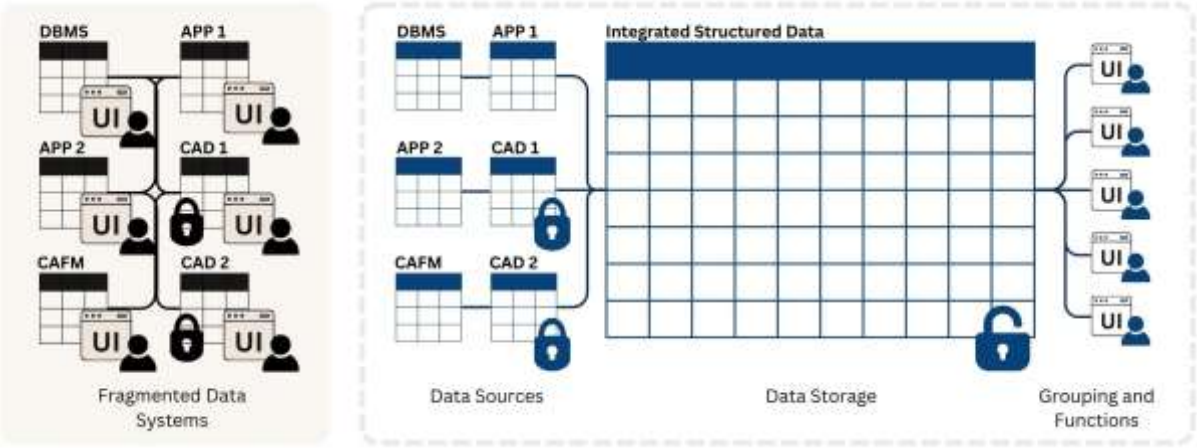


Fig. 2.2-2 La integración de datos elimina los silos, mejora la disponibilidad de la información y optimiza los procesos empresariales.

Una vez estructurados y fusionados los datos, el siguiente paso lógico es validarlos. Con un único repositorio integrado, este proceso se simplifica enormemente: se acabaron los múltiples esquemas incoherentes, las estructuras duplicadas y las complejas relaciones entre tablas. Toda la información se ajusta a un único modelo de datos, lo que elimina las incoherencias internas y acelera el proceso de validación. Validar y garantizar la calidad de los datos son aspectos fundamentales de todos los procesos empresariales, y los veremos con más detalle en los capítulos correspondientes del libro.

En la última etapa, los datos se agrupan, se filtran y se analizan. Se les aplican diversas funciones: agregación (suma, multiplicación), cálculos entre tablas, columnas o filas (Fig. 2.2-4). El trabajo con datos se convierte en una secuencia de pasos: recopilación, estructuración, validación, transformación, procesamiento analítico y descarga en aplicaciones finales donde la información se utiliza para resolver problemas prácticos. En los capítulos sobre ETL -procesos y enfoque de canalización de datos- trataremos con más detalle la construcción de este tipo de escenarios, la automatización de los pasos y la creación de flujos de procesamiento.

Así pues, la transformación digital no consiste solo en simplificar el manejo de la información. Se trata de eliminar la complejidad excesiva en la gestión de datos, de pasar del caos a la previsibilidad, de múltiples sistemas a un proceso manejable. Cuanto menor sea la complejidad de la arquitectura, menos código se necesitará para soportarla. Y en el futuro, el código como tal puede desaparecer por completo, dando paso a agentes inteligentes que analicen, sistematicen y transformen los datos de forma independiente.

Los sistemas de almacenamiento integrados permiten la transición a los agentes de IA

Cuanto menos complejos sean los datos y los sistemas, menos código habrá que escribir y mantener. Y la forma más fácil de ahorrar en desarrollo es deshacerse por completo del código, sustituyéndolo por datos. Cuando el desarrollo de aplicaciones pasa del código a los modelos de datos, se produce inevitablemente un cambio hacia un enfoque centrado en los datos (data-driven), porque detrás de estos conceptos hay una forma de pensar completamente distinta.

> Cuando se opta por trabajar con los datos en el centro, se empieza a ver su papel de otra manera. Los datos ya no son solo "materia prima" para las aplicaciones: ahora son la base en torno a la cual se construyen la arquitectura, la lógica y la interacción.

El enfoque tradicional de la gestión de datos suele empezar en el nivel de las aplicaciones y, en su construcción, se asemeja a un engorroso sistema burocrático: aprobaciones a varios niveles, comprobaciones manuales, versiones interminables de documentos a través de los correspondientes productos de software. Con el desarrollo de las tecnologías digitales, cada vez más empresas se verán obligadas a pasar al principio del minimalismo: almacenar y utilizar sólo lo que sea realmente necesario y vaya a utilizarse.

La lógica de la minimización ha sido asumida por los proveedores. Para simplificar el almacenamiento y procesamiento de datos, el trabajo de los usuarios se está trasladando de las aplicaciones y herramientas offline a los servicios en la nube y a las llamadas soluciones SaaS.

> El concepto SaaS (Software as a Service, o "software como servicio") es una de las tendencias clave de las infraestructuras informáticas modernas, que permite a los usuarios acceder a aplicaciones a través de Internet sin tener que instalar y mantener software en sus propios ordenadores.

Por un lado, SaaS ha facilitado el escalado, el control de versiones y la reducción de los costes de soporte y mantenimiento, pero por otro, además de la dependencia de la lógica de una aplicación concreta, también ha hecho al usuario completamente dependiente de la infraestructura en nube del proveedor. Si un servicio se cae, el acceso a los datos y a los procesos empresariales puede bloquearse temporal o incluso permanentemente. Además, todos los datos del usuario cuando trabaja con aplicaciones SaaS se almacenan en los servidores del proveedor, lo que crea riesgos de seguridad y de cumplimiento de la normativa. Los cambios en las tarifas o las condiciones de uso también pueden provocar un aumento de los costes o la necesidad de una migración urgente.

El desarrollo de los agentes de IA, LLM y el enfoque centrado en los datos han cuestionado el futuro de las aplicaciones en su forma tradicional y la ejecución de SaaS. Mientras que antes las aplicaciones y los servicios debían gestionar la lógica empresarial y procesar los datos, con la llegada de los agentes de IA, estas funciones pueden pasar a sistemas inteligentes que trabajen directamente con los datos.

Por ello, cada vez se habla más de arquitecturas híbridas en los departamentos de TI y a nivel directivo, donde los -agentes de IA y las soluciones locales complementan los servicios en la nube, reduciendo la

dependencia de las -plataformas SaaS.

> *El enfoque que adoptamos reconoce que las aplicaciones empresariales tradicionales o las aplicaciones SaaS pueden cambiar drásticamente en la era de los agentes. Estas aplicaciones son esencialmente bases de datos CRUD [crear, leer, actualizar y borrar] con lógica empresarial. Pero en el futuro, esta lógica será asumida por agentes de IA [46].*
>
> - Satya Nadella, CEO de Microsoft, 2024.

Un enfoque centrado en los datos y el uso de agentes de IA/LLM pueden reducir los procesos redundantes y, por tanto, la carga de trabajo de los empleados. Cuando los datos se organizan adecuadamente, resulta más fácil analizarlos, visualizarlos y aplicarlos a la toma de decisiones. En lugar de interminables informes y comprobaciones, los especialistas acceden a información actualizada en unos pocos clics o con la ayuda de agentes LLM automáticamente en forma de documentos listos y cuadros de mando.

Para la manipulación de los datos contaremos con la ayuda de herramientas de inteligencia artificial (AI) y chats LLM. En los últimos años, se ha observado una tendencia a abandonar las operaciones CRUD tradicionales (crear, leer, actualizar, borrar) en favor del uso de grandes modelos lingüísticos (LLM) para la gestión de datos. Los LLM son capaces de interpretar el lenguaje natural y generar automáticamente consultas adecuadas a la base de datos, lo que simplifica la interacción con los sistemas de gestión de datos (Fig. 2.2-3).

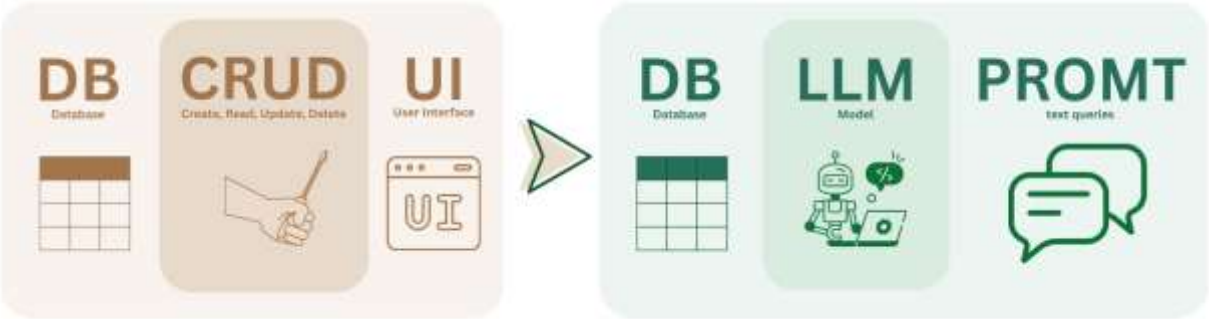


Fig. 2.2-3 La IA sustituirá e integrará soluciones de almacenamiento y bases de datos, desplazando gradualmente a las aplicaciones y operaciones CRUD tradicionales.

> *En los próximos 3-6 meses, la IA escribirá el 90% del código, y en 12 meses casi todo el código podría ser generado por la IA [47].*
>
> - Dario Amodei, CEO de LLM Anthropic, marzo de 2025.

A pesar del rápido desarrollo de herramientas de desarrollo de IA (por ejemplo, GitHub Copilot), en 2025 los

desarrolladores seguirán desempeñando un papel clave en este proceso. agentes de IA se están convirtiendo en asistentes cada vez más útiles: interpretan automáticamente las consultas del usuario, generan consultas SQL y Pandas (más sobre esto en los siguientes capítulos) o escriben código para analizar datos. De este modo, la inteligencia artificial está sustituyendo gradualmente a las interfaces de usuario tradicionales de las aplicaciones.

La proliferación de modelos de inteligencia artificial, como los modelos lingüísticos, impulsará el desarrollo de arquitecturas híbridas. En lugar de abandonar por completo las soluciones en la nube y los productos SaaS, es posible que asistamos a la integración de servicios en la nube con sistemas locales de gestión de datos. Por ejemplo, el aprendizaje federado permite potentes modelos de IA sin tener que trasladar datos sensibles a la nube. De este modo, las empresas pueden mantener el control de sus datos al tiempo que acceden a tecnologías avanzadas.

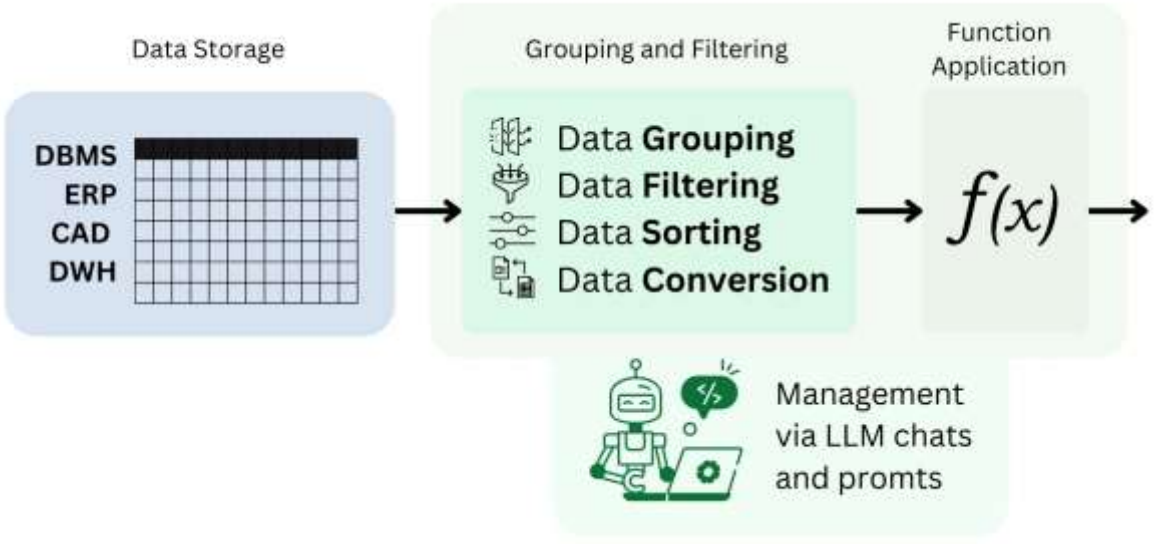


Fig. 2.2-4 Las operaciones básicas de agrupación, filtrado y clasificación, seguidas de la aplicación de funciones, serán gestionadas por los chats LLM.

El futuro del sector de la construcción se basará en una combinación de soluciones locales, potencia de la nube y modelos inteligentes que trabajen juntos para crear sistemas de gestión de datos eficientes y seguros. LLM permitirá a usuarios sin profundos conocimientos técnicos interactuar con bases de datos y almacenes de datos formulando sus consultas en lenguaje natural. Hablaremos más sobre los agentes LLM y AI y su funcionamiento en el capítulo "Agentes LLM y formatos de datos estructurados".

Unos datos bien organizados y unas herramientas de análisis sencillas y fáciles de usar no sólo facilitarán el trabajo con la información, sino que también contribuirán a minimizar los errores, aumentar la eficacia y automatizar los procesos.

De la recogida de datos a la toma de decisiones: el camino hacia la automatización

En partes posteriores del libro, veremos en detalle cómo interactúan los especialistas entre sí y cómo los datos se convierten en la base para la toma de decisiones, la automatización y la eficiencia operativa. La Fig. 2.2-5 ofrece un diagrama de ejemplo que muestra la secuencia de los pasos del procesamiento de datos en un enfoque centrado en los datos. Este diagrama ilustra el proceso de mejora continua (Continuous Improvement Pipeline), algunas de cuyas partes se tratarán en detalle más adelante en el libro.

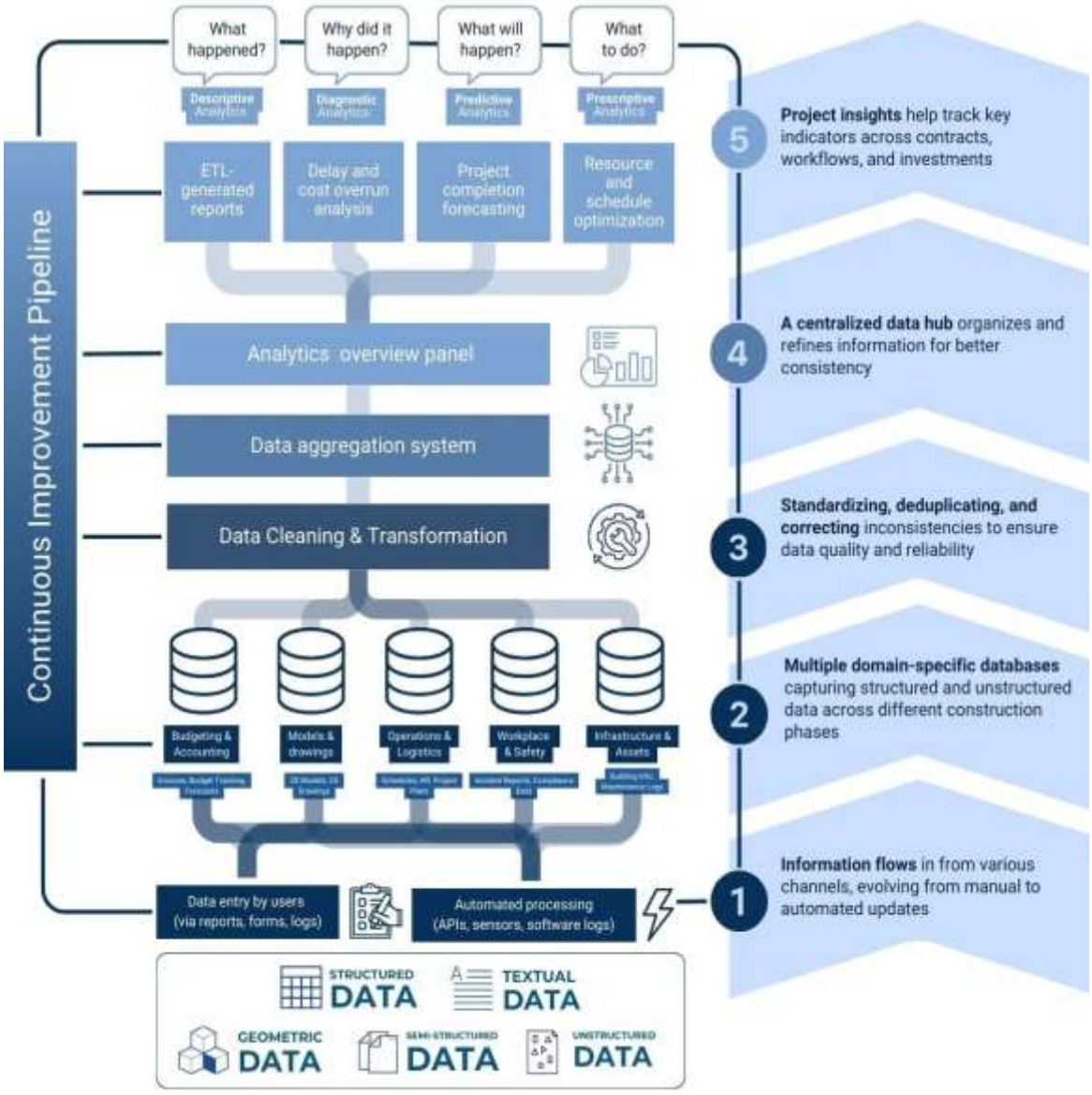


Fig. 2.2-5 Un ejemplo de proceso de mejora continua de datos: el flujo de procesamiento y análisis de datos en proyectos de construcción.

El sistema que describe los procesos empresariales de una mediana empresa se basa en un principio multinivel. Incluye: recogida de datos, depuración, agregación, tratamiento analítico y toma de decisiones basada en los resultados. Estudiaremos todas estas etapas más adelante en el libro - tanto en un contexto teórico como a través de ejemplos prácticos:

- En el primer nivel tiene lugar la **entrada de datos (**Fig. 3.1-1**)**. La información se recibe tanto manualmente (a través de informes, formularios, registros) como de forma automatizada (desde API, sensores, sistemas de software). Los datos pueden tener distintas estructuras: geométricos, de texto, no estructurados. En esta fase es necesario normalizar, estructurar y unificar los flujos de información.
- El siguiente nivel es el **tratamiento y la transformación de los datos**. Incluye los procesos de limpieza, eliminación de duplicados, corrección de errores y preparación de la información para su posterior análisis (Fig, 4.2-5). Esta etapa es crítica porque la calidad del análisis depende directamente de la limpieza y exactitud de los datos.
- A continuación, **los datos fluyen hacia tablas especializadas, marcos de datos o bases de datos** divididas por áreas funcionales: presupuestación y contabilidad, modelos y planos, logística, seguridad e infraestructuras. Esta división facilita el acceso y el análisis cruzado de la información.
- A continuación, los datos se **agregan y se muestran en un cuadro de mando analítico** (escaparate). Aquí se aplican métodos de análisis descriptivos, de diagnóstico, predictivos y prescriptivos. Esto ayuda a responder preguntas clave (Fig. 1.1-4): qué ha pasado, por qué ha pasado, qué pasará en el futuro y qué medidas hay que tomar. Por ejemplo, el sistema puede identificar retrasos, predecir la finalización de proyectos u optimizar recursos.
- Finalmente, el último nivel genera **conclusiones analíticas e indicadores clave** que ayudan a supervisar el cumplimiento de los contratos, gestionar las inversiones y mejorar los procesos empresariales (Fig. 7.4-2). Esta información se convierte en la base para la toma de decisiones y la estrategia de desarrollo de la empresa.

Del mismo modo, los datos pasan de ser recopilados a ser utilizados en la gestión estratégica. En las siguientes partes del libro, veremos cada etapa en detalle, centrándonos en los tipos de datos, las técnicas de procesamiento de datos, las herramientas de análisis y casos reales de cómo se utilizan estos enfoques en el sector de la construcción.

Próximos pasos: convertir el caos en un sistema gestionable

En esta parte, exploramos los retos que plantean los silos de información y examinamos el impacto de la excesiva complejidad de los sistemas en el rendimiento empresarial, analizando la transición de la cuarta revolución industrial a la quinta, en la que los datos, más que las aplicaciones, ocupan un lugar central. Hemos visto cómo los sistemas de información en silos están creando barreras al intercambio de conocimientos, y cómo la continua complejidad del panorama informático está reduciendo la productividad e inhibiendo la innovación en el sector de la construcción.

Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar los planteamientos comentados en sus tareas cotidianas:

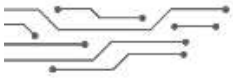
- Visualice su panorama informativo

- ☐ Crea un mapa visual de las fuentes de datos (Miro, Figma, Canva) con las que trabajas habitualmente.
- ☐ Añade a este mapa los sistemas y aplicaciones que utilizas en tu trabajo
- ☐ Identificar posibles funciones duplicadas y soluciones redundantes
- ☐ Identificar los puntos críticos en los que puede producirse una pérdida o corrupción de datos durante la transmisión entre sistemas.

- Aplicar prácticas personalizadas de gestión de datos
  - ☐ Cambiar el enfoque de las aplicaciones a los datos como activo clave en los procesos.
  - ☐ Documentar las fuentes de datos y la metodología de tratamiento para garantizar la transparencia
  - ☐ Desarrollar mecanismos para evaluar y mejorar la calidad de los datos
  - ☐ Esfuércese por garantizar que los datos se introducen una vez y se utilizan repetidamente: ésta es la base de una organización eficaz de los procesos.
- Promueva en su equipo un enfoque centrado en los datos (data-driven)
  - ☐ Sugerir el uso de formatos normalizados y uniformes para el intercambio de datos entre pares.
  - ☐ Plantear regularmente cuestiones relacionadas con la calidad y la disponibilidad de los datos en las reuniones de equipo.
  - ☐ Conozca las alternativas de código abierto a las herramientas que utiliza para resolver sus problemas

Empiece poco a poco: elija un proceso o conjunto de datos específico que sea crítico para su trabajo y aplíquele un enfoque centrado en los datos, desplazando la atención de las herramientas a los datos. Si consigue el éxito en un único proyecto piloto, no sólo obtendrá experiencia práctica, sino también una clara demostración de las ventajas de la nueva metodología para su equipo. Al completar la mayoría de estos pasos, si tienes dudas, puedes pedir aclaraciones y ayuda a cualquier LLM actualizado.

En las siguientes partes del libro, pasaremos a un examen más detallado de las técnicas de estructuración y armonización de datos y exploraremos enfoques prácticos para integrar información heterogénea. Se prestará especial atención a la transición de silos dispares a ecosistemas de datos unificados, que desempeñan un papel clave en la transformación digital del sector de la construcción.



III PARTE

MARCO DE DATOS EN LOS PROCESOS EMPRESARIALES DE LA CONSTRUCCIÓN

En la tercera parte, se forma una comprensión global de la tipología de los datos en la construcción y de los métodos para su organización eficaz. Se analizan las características y especificidades del trabajo con datos estructurados, no estructurados, semiestructurados, textuales y geométricos en el contexto de los proyectos de construcción. Se revisan los modernos formatos de almacenamiento y los protocolos de intercambio de información entre los distintos sistemas utilizados en la industria. Se describen herramientas y técnicas prácticas para convertir datos multiformato en un único entorno estructurado, incluido el modo de integrar datos CAD (BIM). Se proponen enfoques para garantizar la calidad de los datos de mediante la normalización y la validación, fundamentales para la precisión de los cálculos de construcción. Se analizan en detalle los aspectos prácticos del uso de tecnologías modernas (Python Pandas, LLM -modelos) con ejemplos de código para resolver problemas típicos del sector de la construcción. Se corrobora el valor de la creación de un centro de competencia (CoE) como estructura organizativa para la coordinación y normalización de los enfoques de gestión de la información.

CAPÍTULO 3.1.
TIPOS DE DATOS EN LA CONSTRUCCIÓN

Los tipos de datos más importantes en el sector de la construcción

En el sector moderno de la construcción, los sistemas, aplicaciones y almacenes de datos de las empresas de se llenan activamente de información y datos de diversos tipos y formatos (Fig. 3.1-1). Veamos más de cerca los principales tipos de datos que conforman el panorama informativo de una empresa moderna del sector de la construcción:

- Datos **estructurados**: estos datos tienen una estructura organizativa clara, por ejemplo, las hojas de cálculo de Excel y las bases de datos relacionales.
- Datos **no estructurados**: se trata de información que no está organizada según reglas estrictas. Ejemplos de este tipo de datos son el texto, el vídeo, las fotos y las grabaciones de audio.
- Datos **poco estructurados**: estos datos ocupan una posición intermedia entre los datos estructurados y los no estructurados. Contienen elementos de estructura, pero esta estructura no siempre es clara o a menudo se describe a través de diferentes esquemas. Ejemplos de datos semiestructurados en la construcción son: las especificaciones técnicas, la documentación del proyecto o los informes de progreso.
- **Datos textuales**: incluye todo lo derivado de comunicaciones orales y escritas, como correos electrónicos, transcripciones de reuniones y citas.
- Datos **geométricos**: estos datos proceden de los programas CAD en los que los especialistas crean datos geométricos de los elementos del proyecto para su visualización, confirmación de los valores de volumen o comprobación de colisiones.

Es importante señalar que los datos geométricos y de texto (alfanuméricos) no constituyen una categoría separada, sino que pueden estar presentes en los tres tipos de datos. Los datos geométricos, por ejemplo, pueden formar parte tanto de datos estructurados (formatos CAD paramétricos) como de datos no estructurados (dibujos escaneados). Del mismo modo, los datos textuales pueden estar organizados en bases de datos (datos estructurados) o existir como documentos sin una estructura clara.

Cada tipo de dato en una empresa de construcción es un elemento único en el mosaico de los activos de información de la empresa. Desde los datos no estructurados, como las imágenes de las obras y las grabaciones de audio de las reuniones, hasta los registros estructurados, incluidas las tablas y las bases de datos, cada elemento desempeña un papel importante en la configuración del panorama informativo de la empresa.

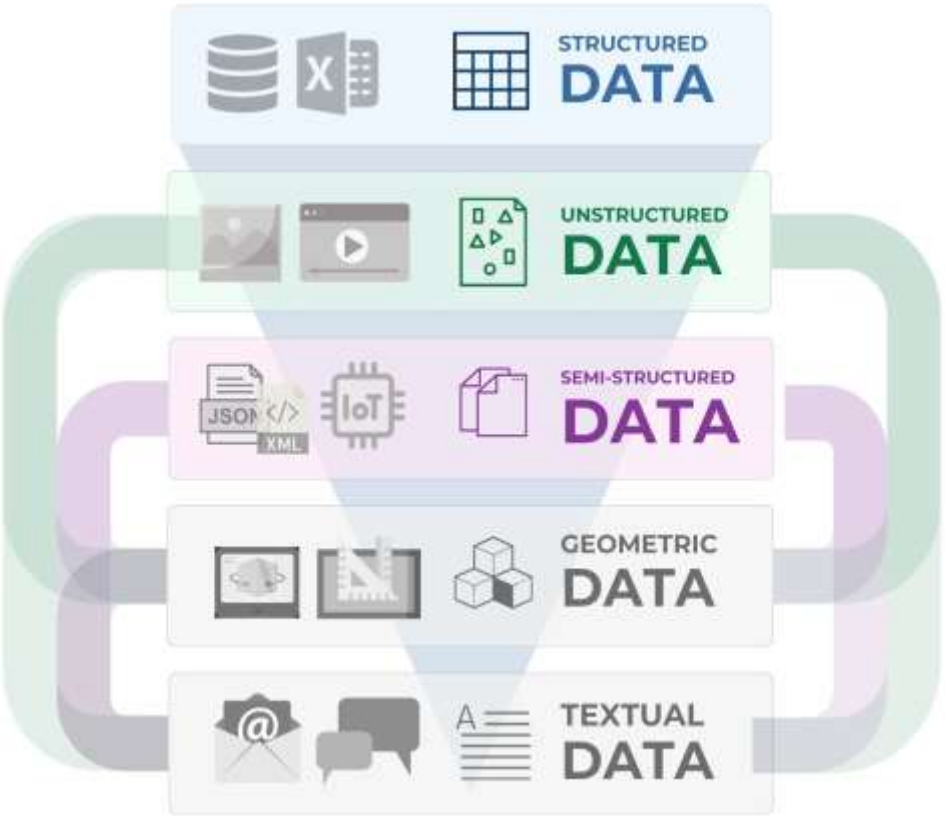


Fig. 3.1-1 Los ingenieros y gestores de datos deben aprender a trabajar con todos los tipos de datos utilizados en el sector de la construcción.

A continuación se ofrece una lista de muestra de algunos de los sistemas y tipos de datos asociados (Fig, 3.1-2) utilizados en la construcción:

- **ERP** (Enterprise Resource Planning) - maneja datos generalmente estructurados para ayudar a gestionar los recursos de la empresa e integrar diversos procesos empresariales.
- **CAD** (Computer-Aided Design) combinado con **BIM** (Building Information Modeling): utiliza datos geométricos y semiestructurados para diseñar y modelar proyectos de construcción, garantizando la precisión y coherencia de la información durante la fase de diseño.
- **SIG** (Sistemas de Información Geográfica): trabaja con datos geométricos y estructurados para crear y analizar datos cartográficos y relaciones espaciales.
- **RFID** (Identificación por radiofrecuencia): utiliza datos semiestructurados para realizar un seguimiento eficaz de materiales y equipos en una obra mediante identificación por radiofrecuencia.
- **ECM** (Engineering Content Management) es un sistema de gestión de datos y documentación de ingeniería, incluidos datos semiestructurados y no estructurados como dibujos técnicos y documentos de diseño.

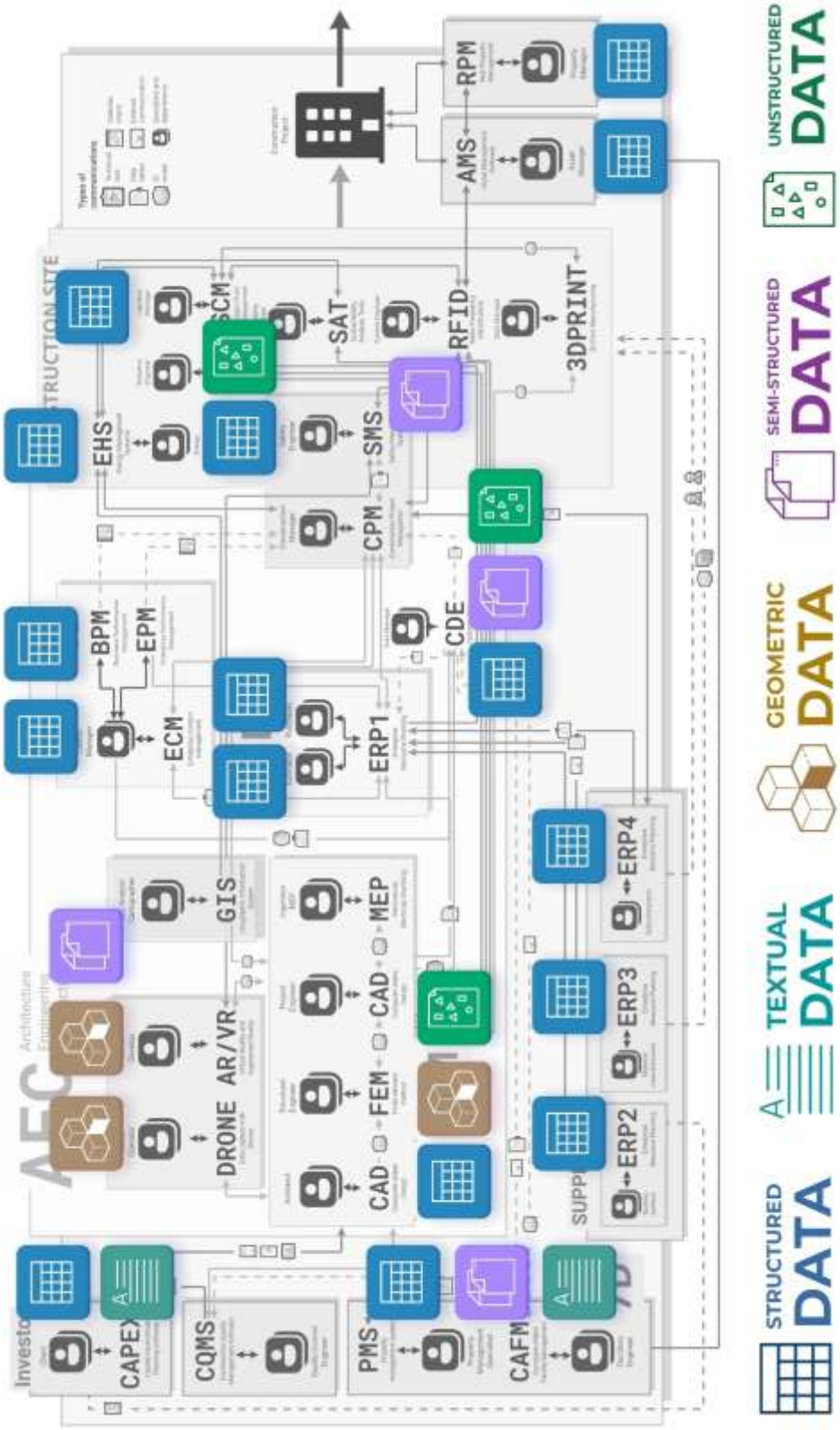


Fig. 3.1-2 Diferentes formatos y datos pueblan distintos sistemas, lo que exige su traducción a un formato adecuado para una integración compleja.

Estos y muchos otros sistemas de la empresa gestionan una amplia gama de datos, desde datos tabulares estructurados hasta complejos modelos geométricos, proporcionando una colaboración integrada en los procesos de diseño, planificación y gestión de la construcción.

En el ejemplo de un diálogo simplificado (Fig. 3.1-3), se intercambian distintos tipos de datos entre los especialistas del proyecto de construcción:

- **Arquitecto**: "*Teniendo en cuenta los deseos del cliente, he añadido una zona de asientos en el tejado. Eche un vistazo al nuevo diseño*" (datos geométricos - modelo).
- **Ingeniero de estructuras**: "*Se ha recibido el proyecto. Estoy calculando la capacidad de carga de la cubierta para la nueva zona de recreo*" (datos estructurados y semiestructurados - tablas de cálculo).
- **Jefe de compras**: "*Necesito especificaciones y cantidades de materiales para el área recreativa para organizar la compra*" (datos textuales y semiestructurados - listas y especificaciones).
- **Ingeniero de Salud y Seguridad**: "*He recibido datos sobre la nueva zona. Estoy evaluando los riesgos y actualizando el plan de seguridad*" (datos semiestructurados - documentos y planos).
- **Especialista en BIM** -modelling: "*Realización de cambios en el modelo global del proyecto para ajustar la documentación de trabajo*" (datos geométricos y datos semiestructurados).
- **Jefe de proyecto**: "*Estoy incorporando la nueva área de descanso al plan de trabajo. Estoy actualizando los calendarios y los recursos en el sistema de gestión de proyectos*" (datos estructurados y semiestructurados: calendarios y planes).
- **Especialista en mantenimiento de instalaciones (FM)**: *"Preparo datos para el futuro mantenimiento de la zona recreativa y los introduzco en el sistema de gestión de la propiedad*" (datos estructurados y semiestructurados: instrucciones y planes de mantenimiento).

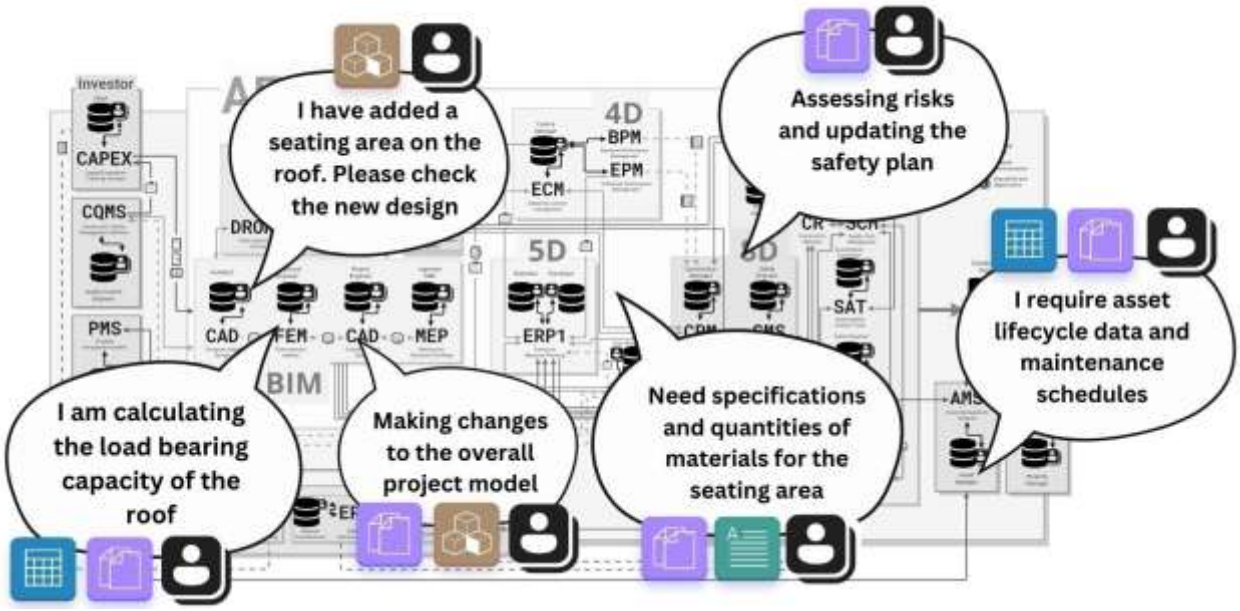


Fig. 3.1-3 La comunicación entre especialistas tiene lugar tanto a nivel de texto como de datos.

Cada profesional trabaja con distintos tipos de datos para garantizar una colaboración eficaz en equipo y el éxito del proyecto. Comprender las diferencias entre datos estructurados, semiestructurados y no estructurados permite reconocer el papel único que desempeña cada tipo en los procesos empresariales digitales. No sólo es importante saber que existen distintas formas de datos, sino también comprender

cómo, dónde y por qué se utilizan.

> No hace mucho, la idea de combinar datos tan diversos parecía ambiciosa pero difícil de realizar. Hoy ya forma parte de la práctica cotidiana. La integración de datos de esquemas y estructuras diferentes se ha convertido en parte integrante de la arquitectura de los sistemas de información modernos.

En los capítulos siguientes examinaremos en detalle las principales normas y enfoques que permiten combinar datos estructurados, semiestructurados y no estructurados en una única visión coherente. Se prestará especial atención a los datos estructurados y las bases de datos relacionales como principales mecanismos de almacenamiento, tratamiento y análisis de la información en el sector de la construcción.

Datos estructurados

En el sector de la construcción, la información procede de muchas fuentes: planos, especificaciones, calendarios e informes. Para gestionar eficazmente este flujo de información, es necesario estructurarla. Los datos estructurados permiten organizar la información de forma cómoda, legible y accesible.

> Según el 5º Informe Anual sobre Tecnología de la Construcción de JB Knowledge [17], el 67% de los profesionales de la gestión de proyectos de construcción realizan el seguimiento y la evaluación del rendimiento laboral manualmente o mediante hojas de cálculo.

Algunos de los formatos de datos estructurados más comunes son XLSX y CSV. Son muy utilizados para almacenar, procesar y analizar información en hojas de cálculo. En estas hojas de cálculo, los datos se presentan en forma de filas y columnas, lo que facilita su lectura, edición y análisis.

XLSX, formato creado por Microsoft, se basa en el uso de estructuras XML y se archiva mediante el algoritmo ZIP. Las principales características del formato:

- Soporte para fórmulas complejas, gráficos y macros.
- Capacidad para almacenar datos en diferentes hojas, así como para dar formato a la información.
- Optimizado para Microsoft Excel, pero compatible con otras suites ofimáticas.

Formato CSV es un archivo de texto sin formato en el que los valores están separados por comas, punto y coma u otros caracteres delimitadores. Principales ventajas:

- Compatibilidad universal con diversos programas y sistemas operativos.
- Fácil importación y exportación a bases de datos y sistemas analíticos.
- Fácil procesamiento incluso en editores de texto.

Sin embargo, CSV no admite fórmulas ni formatos, por lo que su principal aplicación es el intercambio de datos entre sistemas y la actualización masiva de información. Debido a su versatilidad e independencia de plataformas, CSV se ha convertido en una herramienta popular para la transferencia de datos en entornos informáticos heterogéneos.

Los dos formatos XLSX y CSV sirven de enlace entre distintos sistemas que tratan datos estructurados

(Fig, 3.1-4). Son especialmente útiles en tareas en las que la legibilidad, la edición manual y la compatibilidad básica son importantes

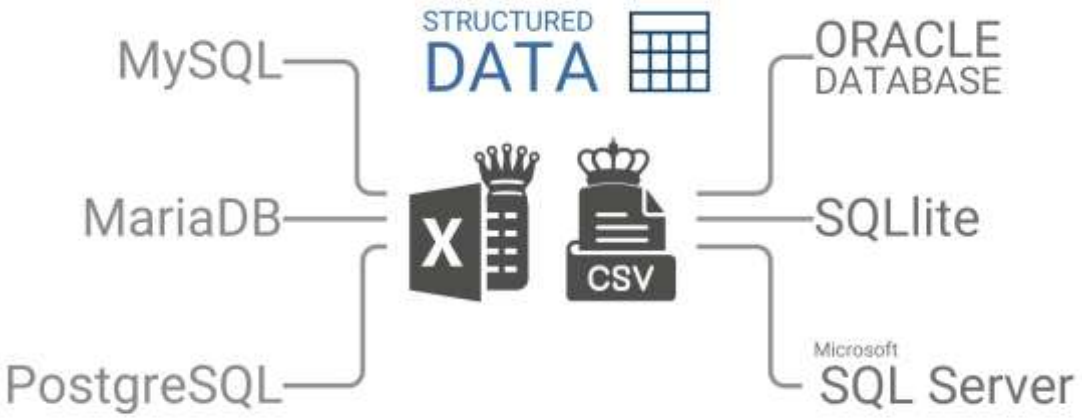


Fig. 3.1-4 Los formatos XLSX y CSV son el nexo de unión entre distintos sistemas que trabajan con datos estructurados.

> La independencia de plataformas convierte a CSV en el formato más popular para la transferencia de datos en entornos y sistemas informáticos heterogéneos.

Sin embargo, XLSX y CSV no están diseñados para la computación de alto rendimiento o el almacenamiento a largo plazo de grandes cantidades de datos. Para estos fines se utilizan formatos estructurados más modernos, como Apache Parquet, Apache ORC, Feather, HDF5. Estos formatos se tratarán con más detalle en el capítulo "Almacenamiento de Big Data: análisis de los formatos más populares y su eficacia", en la Parte 9 de este libro

En la práctica, Excel con formato XLSX se utiliza más a menudo para pequeñas tareas y automatización de procesos rutinarios. Escenarios más complejos requieren el uso de sistemas de gestión de datos como ERP, PMIS CAFM, CPM, SCM y otros (Fig. 3.2-1). Estos sistemas almacenan datos estructurados en los que se basa la organización y gestión de los flujos de información de la empresa.

Los modernos sistemas de información para la gestión de datos utilizados en el sector de la construcción se basan en datos estructurados organizados en forma de tablas. Para una gestión fiable, escalable y holística de grandes volúmenes de información, los desarrolladores de aplicaciones y sistemas recurren a los sistemas de gestión de bases de datos relacionales (RDBMS).

Bases de datos relacionales RDBMS y lenguaje de consulta SQL

Para almacenar, procesar y analizar datos de forma eficaz, **las bases de datos relacionales (RDBMS**) son

sistemas de almacenamiento de datos que organizan la información en tablas con relaciones definidas entre ellas.

> Los datos organizados en bases de datos (RDBMS) no son sólo información digital; son la base de las transacciones e interacciones entre distintos sistemas.

He aquí algunos de los sistemas de gestión de bases de datos relacionales (RDBMS) más comunes (Fig, 3.1-5):

- **MySQL** (Open Source) es uno de los RDBMS más populares, que forma parte de la pila LAMP (Linux, Apache, MySQL, PHP /Perl/Python). Es ampliamente utilizado en el desarrollo web debido a su simplicidad y alto rendimiento.
- **PostgreSQL** (código abierto) es un potente sistema relacional de objetos conocido por su fiabilidad y sus funciones avanzadas. Es adecuado para soluciones empresariales complejas.
- **Microsoft SQL Server** es un sistema comercial de Microsoft muy utilizado en entornos corporativos por su integración con otros productos de la empresa y su alto nivel de seguridad.
- **Oracle Database** es uno de los DBMS más potentes y fiables utilizados en grandes empresas y aplicaciones de misión crítica.
- **IBM DB2**: dirigido a grandes empresas, ofrece un alto rendimiento y tolerancia a fallos.
- **SQLite** (código abierto) es una base de datos integrada ligera, ideal para aplicaciones móviles y sistemas autónomos como el software de diseño CAD (BIM).

Los sistemas de gestión de bases de datos más populares en el sector de la construcción - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 y SQLite - trabajan con datos estructurados. Todos estos SGBD son soluciones potentes y flexibles para gestionar una amplia gama de procesos y aplicaciones empresariales, desde pequeños sitios web hasta sistemas empresariales a gran escala (Fig. 3.2-1).

> Según Statista [48], los sistemas de gestión de bases de datos relacionales (SGBDR) representan alrededor del 72% del total de SGBDR en uso en 2022.

Rank			DBMS	Database Model	Open Source vs Commercial
Mar2025	Feb2025	Mar2024			
1.	1.	1.	**Oracle®**	Relational, Multi-model	Commercial
2.	2.	2.	**MySQL**	Relational, Multi-model	**Open Source**
3.	3.	3.	**Microsoft® SQL Server**	Relational, Multi-model	Commercial
4.	4.	4.	**PostgreSQL**	Relational, Multi-model	**Open Source**
5.	5.	5.	MongoDB	Document, Multi-model	**Open Source**
6.	7.	9.	**Snowflake®**	Relational	Commercial
7.	6.	6.	Redis®	Key-value, Multi-model	**Open Source**
8.	8.	7.	Elasticsearch®	Multi-model	**Open Source**
9.	9.	8.	**IBM Db2**	Relational, Multi-model	Commercial
10.	10.	10.	**SQLite**	Relational	**Open Source**
11.	11.	12.	Apache Cassandra®	Multi-model	**Open Source**
12.	12.	11.	**Microsoft Access®**	Relational	**Open Source**
13.	13.	17.	Databricks®	Multi-model	Commercial
14.	14.	13.	**MariaDB**	Relational, Multi-model	**Open Source**
15.	15.	14.	Splunk	Search engine	Commercial
16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
17.	17.	15.	**Microsoft Azure SQL**	Relational, Multi-model	Commercial

Fig. 3.1-5 Popularidad del uso de bases de datos estructuradas (marcadas en azul) en la clasificación de SGBD (basado en [49]).

Es bastante fácil instalar bases de datos de código abierto - incluso sin amplios conocimientos técnicos. Los sistemas de código abierto, como PostgreSQL, MySQL o SQLite, están disponibles de forma gratuita y funcionan en la mayoría de sistemas operativos: Windows, macOS y Linux. Basta con ir a la web oficial del proyecto, descargar el instalador y seguir las instrucciones. En la mayoría de los casos, la instalación no lleva más de 10-15 minutos. En la cuarta parte del libro modelaremos y crearemos una base de datos de este tipo (Fig. 4.3-8).

Si tu empresa utiliza servicios en la nube (por ejemplo, Amazon Web Services, Google Cloud o Microsoft Azure), puedes desplegar la base de datos en un par de clics: la plataforma te ofrecerá plantillas ya preparadas para la instalación. Gracias a la apertura del código, este tipo de bases de datos son fáciles de personalizar para tus tareas, y una enorme comunidad de usuarios siempre te ayudará a encontrar una solución a cualquier problema.

RDBMS sigue siendo la base de una gran cantidad de aplicaciones empresariales y plataformas analíticas (Fig, 3.1-6) que permiten a las empresas almacenar, procesar y analizar datos de forma eficiente y, por tanto, tomar decisiones informadas y oportunas.

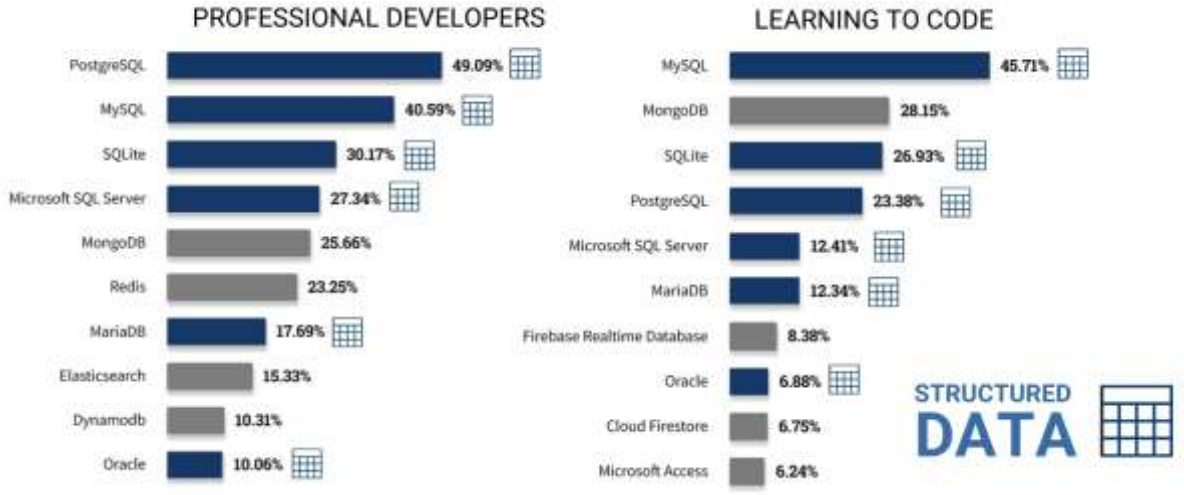


Fig. 3.1-6 Encuesta a desarrolladores en StackOverFlow (el mayor foro de TI) sobre qué bases de datos utilizaron el año pasado y cuáles quieren utilizar el año que viene (los RDBMS aparecen resaltados en azul) (basado en [50]).

RDBMS proporciona fiabilidad, coherencia de datos, soporte de transacciones y utiliza un potente lenguaje de consulta - SQL (Structured Query Language), que se utiliza a menudo en analítica y permite obtener, modificar y analizar fácilmente la información almacenada en las bases de datos. SQL es la principal herramienta para trabajar con datos en sistemas relacionales.

SQL -consultas en bases de datos y nuevas tendencias

La principal ventaja del lenguaje SQL, utilizado a menudo en bases de datos relacionales, frente a otros tipos de gestión de la información (por ejemplo, con ayuda de las clásicas hojas de cálculo Excel) es el soporte de volúmenes muy grandes de bases de datos a gran velocidad de procesamiento de consultas.

> El lenguaje de consulta estructurado (SQL) es un lenguaje de programación especializado diseñado para almacenar, procesar y analizar información en bases de datos relacionales. SQL se utiliza para crear, gestionar y acceder a datos, permitiendo encontrar, filtrar, combinar y agregar información de forma eficaz. Es una herramienta clave para acceder a los datos, ya que proporciona una forma cómoda y formalizada de interactuar con los almacenes de información.

La evolución de los sistemas SEQUEL-SQL pasa por importantes productos y empresas como Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL y MySQL, y culmina con la aparición de SQLite y MariaDB [51]. SQL proporciona capacidades de hoja de cálculo que no se encuentran en Excel, lo que hace que la manipulación de datos sea más escalable, segura y fácil de automatizar:

- **Creación y gestión de estructuras de datos (DDL):** En SQL se pueden crear, modificar y eliminar

tablas de una base de datos, establecer vínculos entre ellas y definir estructuras de almacenamiento de datos. Excel, en cambio, trabaja con hojas y celdas fijas, sin relaciones claramente definidas entre hojas y conjuntos de datos.

- **Manipulación de datos (DML):** SQL permite añadir, modificar, eliminar y recuperar datos de forma masiva y a gran velocidad realizando consultas complejas con filtrado, ordenación y uniones de tablas (Fig, 3.1-7). En Excel, el procesamiento de grandes cantidades de información requiere acciones manuales o macros especiales, lo que ralentiza el proceso y aumenta la probabilidad de errores.
- **Control de acceso (DCL):** SQL permite diferenciar los derechos de acceso a los datos para distintos usuarios, limitando la capacidad de editar o ver la información. En Excel, en cambio, el acceso es compartido (al transferir un archivo) o requiere configuraciones complejas con permisos compartidos a través de servicios en la nube.

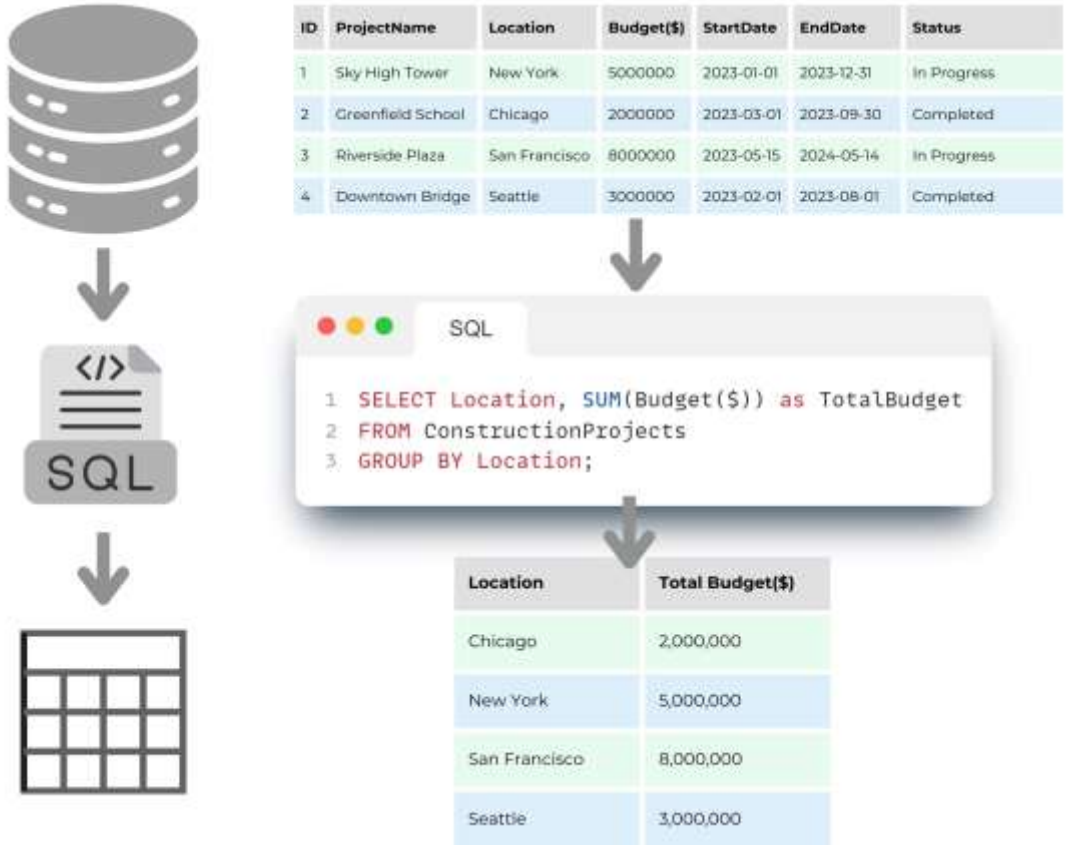


Fig. 3.1-7 Ejemplo de DML en SQL: procesamiento rápido, agrupación y agregación con unas pocas líneas de código para el tratamiento automático de datos.

Excel facilita el trabajo con datos gracias a su estructura visual e intuitiva. Sin embargo, a medida que aumenta la cantidad de datos, el rendimiento de Excel disminuye. Excel también se enfrenta a límites en la cantidad de datos que puede almacenar -un máximo de un millón de filas- y el rendimiento se degrada mucho antes de alcanzar este límite. Así pues, mientras que Excel parece preferible para visualizar y manipular pequeñas cantidades de datos, SQL es más adecuado para tratar grandes conjuntos de datos.

La siguiente etapa en el desarrollo de los datos estructurados fue la aparición de las bases de datos columnares (Columnar Databases), que constituyen una alternativa a las bases de datos relacionales tradicionales, especialmente cuando se trata de volúmenes de datos y cálculos analíticos considerablemente mayores. A diferencia de las bases de datos por filas, en las que los datos se almacenan línea por línea, las bases de datos columnares registran la información por columnas. En comparación con las bases de datos clásicas, esto permite:

- Reduzca el espacio de almacenamiento comprimiendo eficazmente datos uniformes en columnas.
- Acelere las consultas analíticas, ya que sólo se leen las columnas necesarias y no toda la tabla.
- Optimice Big Data y el almacenamiento de datos, por ejemplo, Data Lakehouse Architecture.

Hablaremos más sobre bases de datos columnares, Pandas DataFrame, Apache Parquet, HDF5, así como sobre la creación de Big Data -almacenes basados en ellos con fines de análisis y procesamiento de datos en los siguientes capítulos de este libro - "DataFrame: un formato universal de datos tabulares" y "Formatos de almacenamiento de datos y trabajo con Apache Parquet: DWH -almacenes de datos y arquitectura Data Lakehouse ".

Datos no estructurados

Aunque la mayoría de los datos utilizados en aplicaciones y sistemas de información tienen forma estructurada, la mayor parte de la información generada en la construcción tiene forma de datos no estructurados: imágenes, vídeos, documentos de texto, grabaciones de audio y otras formas de contenido. Esto es especialmente cierto en las fases de construcción, explotación y supervisión técnica, en las que predomina la información visual y textual.

> **Los datos no estructurados** son información que no tiene un modelo o estructura predefinidos, ni está organizada en filas y columnas tradicionales como en las bases de datos o tablas.

En términos generales, los datos no estructurados pueden clasificarse en dos categorías:

- Datos no estructurados generados por el ser humano, que incluyen varios tipos de contenidos generados por el ser humano: documentos de texto, correos electrónicos, imágenes, vídeos, etc.
- Los datos no estructurados generados por máquinas son creados por dispositivos y sensores: incluyen archivos de registro, datos GPS, resultados del Internet de las Cosas (IoT) y otra información de telemetría de una obra, por ejemplo.

A diferencia de los datos estructurados, que se organizan cómodamente en tablas y bases de datos, los datos no estructurados requieren pasos adicionales de procesamiento antes de su integración en sistemas de información (Fig. 3.1-8). El uso de tecnologías para la recogida, el análisis y la transformación automatizados de tales datos abre nuevas oportunidades para mejorar la eficiencia de la construcción, reducir los errores y minimizar la influencia del factor humano.

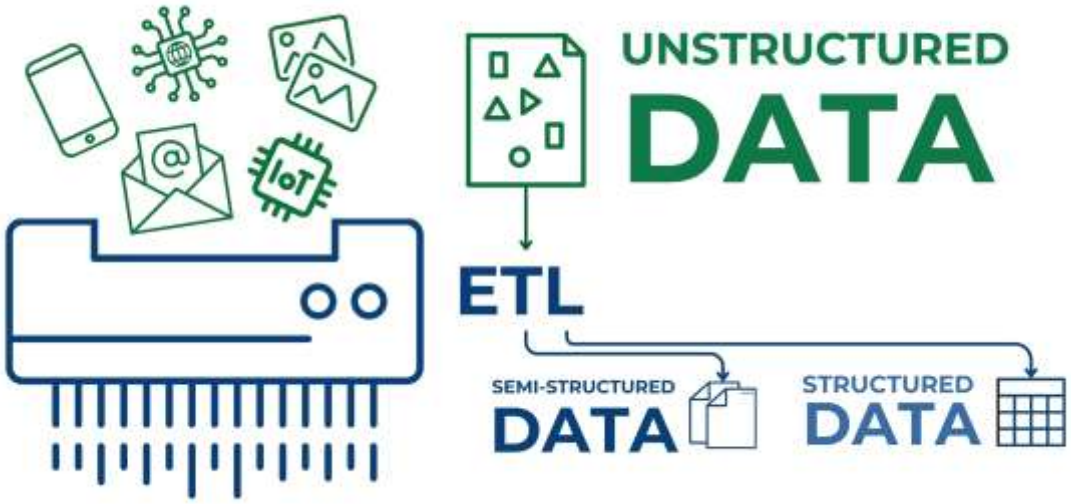


Fig. 3.1-8 El tratamiento de datos no estructurados comienza con su transformación en datos semiestructurados y estructurados.

Los datos no estructurados suponen hasta el 80% de toda la información [52] con la que se encuentran los profesionales en las empresas, por lo que trataremos en detalle sus tipos y tratamiento con ejemplos en los siguientes capítulos del libro.

Para facilitar el debate, los datos textuales se clasifican por separado. Aunque se trata de un tipo de datos no estructurados, su importancia y prevalencia en el sector de la construcción requieren una atención especial.

Datos de texto: entre el caos no estructurado y el caos estructurado y

Los datos textuales en el sector de la construcción abarcan una amplia gama de formatos y tipos de información, desde documentos en papel hasta métodos informales de comunicación como cartas, conversaciones, correspondencia de trabajo y reuniones verbales en la obra. Todos estos datos textuales contienen información importante para la gestión de los proyectos de construcción, desde detalles sobre decisiones de diseño y cambios en los planos hasta debates sobre cuestiones de seguridad y negociaciones con contratistas y clientes (Fig, 3.1-9).

Fig. 3.1-9 Datos de texto, uno de los tipos de información más utilizados en la comunicación entre los participantes en un proyecto.

La información textual puede ser tanto formalizada como no estructurada. Los datos formalizados incluyen documentos de Word (.doc,.docx), PDF, así como archivos de texto de actas de reuniones (.txt). Los datos informales incluyen correspondencia por mensajería y correo electrónico, transcripciones de reuniones (Teams, Zoom, Google Meet) y grabaciones de audio de debates (.mp3,.wav) que requieren conversión a texto.

Pero mientras que los documentos escritos, como las solicitudes formales, las condiciones contractuales y los correos electrónicos, suelen tener ya cierta estructura, las comunicaciones verbales y la correspondencia de trabajo a menudo siguen sin estar estructuradas, lo que dificulta su análisis e integración en los sistemas de gestión de proyectos.

La clave para una gestión eficaz de los datos de texto es convertirlos a un formato estructurado. Esto permite integrar automáticamente la información procesada en los sistemas existentes que ya trabajan con datos estructurados.

Fig. 3.1-10 Conversión de contenido textual en datos estructurados.

Para hacer un uso eficaz de la información textual, ésta debe convertirse automáticamente en un formato estructurado (Fig, 3.1-10). Este proceso suele implicar varios pasos:

- **Reconocimiento de texto (OCR)**: conversión de imágenes de documentos y dibujos en un formato legible por máquina.
- **Análisis de textos (NLP)**: identificación automática de parámetros clave (fechas, importes y cifras relevantes para el proyecto).

- **Clasificación de datos** - categorización de la información (finanzas, logística, gestión de riesgos).

Tras el reconocimiento y la clasificación, los datos ya estructurados pueden integrarse en bases de datos y utilizarse en sistemas automatizados de información y gestión.

Datos semiestructurados y poco estructurados

Datos semiestructurados contiene cierto nivel de organización, pero no tiene un esquema o estructura estrictos. Aunque dicha información incluye elementos estructurados (por ejemplo, fechas, nombres de empleados y listas de tareas realizadas), el formato de presentación puede variar considerablemente de un proyecto a otro o incluso de un empleado a otro. Ejemplos de estos datos son los registros de tiempo, los informes de progreso y los calendarios, que pueden presentarse en diversos formatos.

Los datos semiestructurados son más fáciles de analizar que los no estructurados, pero requieren un tratamiento adicional para su integración en sistemas normalizados de gestión de proyectos.

> Trabajar con datos semiestructurados caracterizados por una estructura en constante cambio plantea importantes retos. Esto se debe a que la variabilidad de la estructura de los datos exige enfoques individuales separados para procesar y analizar cada fuente de datos semiestructurados.

Pero mientras que tratar datos no estructurados requiere mucho esfuerzo, procesar datos semiestructurados puede hacerse con métodos y herramientas relativamente sencillos.

> Datos **débilmente estructurados** es un término más general que describe los datos con una estructura mínima o incompleta. La mayoría de las veces se trata de documentos de texto, chats, correos electrónicos, en los que se encuentran algunos metadatos (por ejemplo, fecha, remitente), pero la mayor parte de la información se presenta de forma caótica.

En la construcción, los datos poco estructurados se encuentran en diversos procesos. Por ejemplo:

- Presupuestos y ofertas: tablas con datos sobre materiales, cantidades y costes, pero sin un formato uniforme.
- Dibujos y diagramas de ingeniería: archivos en formato PDF o DWG, que contienen anotaciones de texto y metadatos, pero sin una estructura estrictamente fija.
- Planes de trabajo: datos de MS Project, Primavera P6 u otros sistemas, que pueden tener una estructura de exportación diferente.
- CAD (BIM -modelos) - contienen elementos de la estructura, pero la representación de los datos depende del software y del estándar del proyecto.

Los datos geométricos, producidos por los sistemas CAD, pueden clasificarse del mismo modo que los datos semiestructurados. Sin embargo, clasificaremos los datos geométricos CAD (BIM) como un tipo de datos independiente porque, al igual que los datos de texto, a menudo pueden tratarse como un tipo de

datos independiente en los procesos de la empresa.

Datos geométricos y su aplicación

Mientras que los metadatos sobre los elementos del proyecto casi siempre se almacenan en forma de tablas, formatos estructurados o poco estructurados, los datos geométricos de los elementos del proyecto en la mayoría de los casos se crean utilizando herramientas especiales de CAD (Fig. 3.1-11), que permiten visualizar los elementos del proyecto en detalle como un conjunto de líneas (2D) o cuerpos geométricos (3D).

Fig. 3.1-11 Las herramientas CAD han ayudado a trasladar la información geométrica del soporte físico a la forma de base de datos.

Cuando se trabaja con datos geométricos en construcción y arquitectura, pueden identificarse tres aplicaciones principales de los datos geométricos (Fig, 3.1-12):

- **Confirmación de volúmenes**: los datos geométricos, generados en programas CAD (BIM) utilizando núcleos geométricos especiales, son necesarios para determinar automáticamente y con precisión los volúmenes y dimensiones de los elementos del proyecto. Estos datos incluyen áreas, volúmenes, longitudes y otros atributos importantes calculados automáticamente y necesarios para la planificación, presupuestación y pedido de recursos y materiales
- **Visualización del proyecto**: en caso de que se produzcan cambios en el proyecto, la visualización de elementos permite generar automáticamente dibujos actualizados en distintos planos. La visualización del proyecto en las fases iniciales ayuda a acelerar el entendimiento entre todos los participantes para ahorrar tiempo y recursos durante el proceso de construcción.
- **Comprobación de** colisiones: en proyectos complejos de construcción e ingeniería en los que es fundamental la interacción de varias categorías de elementos (por ejemplo, tuberías y muros) sin que se produzcan "conflictos geométricos", la comprobación de colisiones desempeña un papel fundamental. El uso del software de detección de colisiones permite identificar de forma proactiva posibles conflictos geométricos entre los elementos del proyecto, evitando así costosos errores durante el proceso de construcción.

Desde los inicios de las oficinas de diseño de ingeniería, desde la época de la construcción de las primeras estructuras complejas, los ingenieros estructurales proporcionaban información geométrica en forma de dibujos, líneas y elementos geométricos planos (en papiro, papel "A0" o en formatos DWG, PDF, PLT), sobre cuya base los capataces y los estimadores (Fig. 3.1-11), durante los últimos milenios, con la ayuda de reglas y transportes, recopilaban volúmenes atributivos o cantidades de elementos y grupos de elementos. 3.1-11), durante los últimos milenios, con la ayuda de reglas y transportes, recogieron volúmenes atributivos o cantidades de elementos y grupos de elementos.

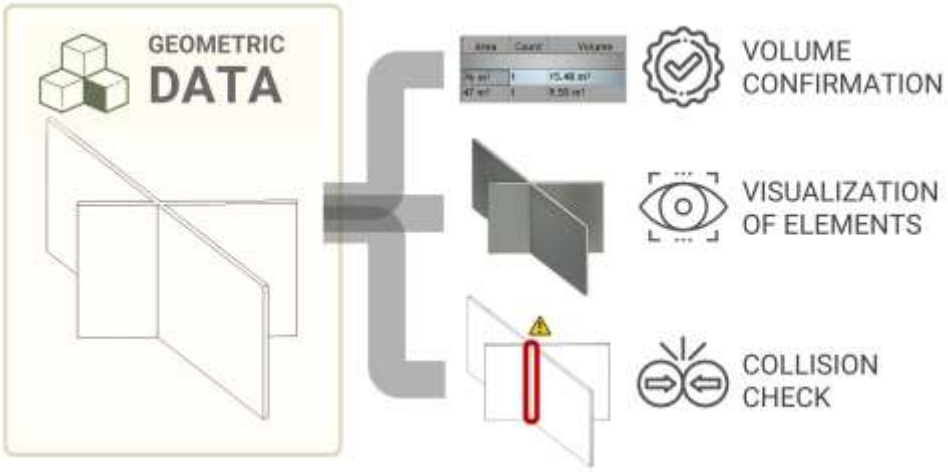


Fig. 3.1-12 La geometría es la base para obtener los parámetros volumétricos de los elementos, que luego se utilizan para calcular el coste y el calendario del proyecto.

Hoy en día, esta tarea manual y lenta se resuelve de forma totalmente automatizada gracias a la aparición del modelado volumétrico en las modernas herramientas CAD (BIM), que permite obtener automáticamente, con ayuda de un núcleo geométrico especial, los atributos volumétricos de cualquier elemento sin necesidad de calcular manualmente los parámetros volumétricos.

Las herramientas CAD modernas también permiten clasificar y categorizar los elementos del proyecto, de modo que se pueden cargar tablas de especificaciones de la base de datos del proyecto para utilizarlas en diversos sistemas, como la estimación de costes, la programación o el cálculo de CO_2 (Fig. 3.1-13). En el capítulo "Obtención de cantidades y cuantificación" trataremos la obtención de especificaciones, tablas QTO y cantidades, así como ejemplos prácticos.

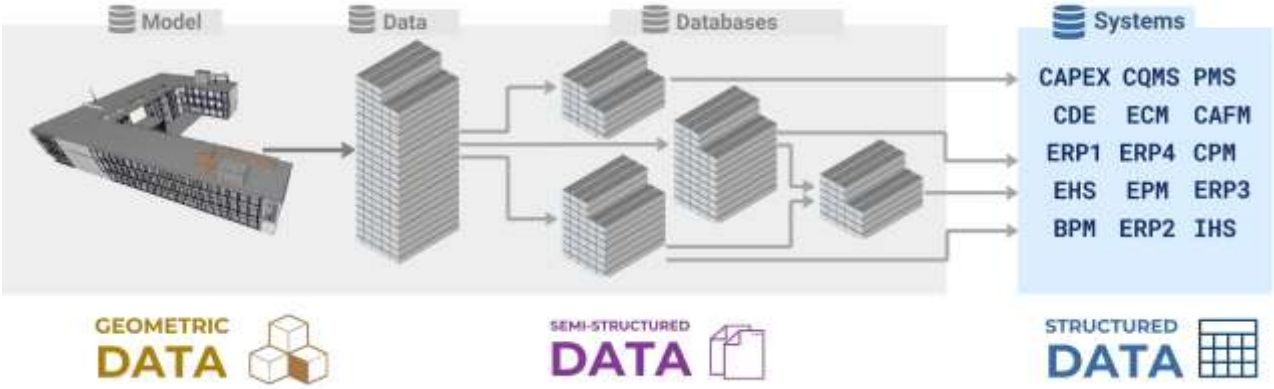


Fig. 3.1-13 Las herramientas CAD (BIM) almacenan datos en bases de datos diseñadas para integrarse e interactuar con otros sistemas.

Debido a la naturaleza cerrada de las bases de datos y los formatos utilizados en el entorno CAD, los datos geométricos creados en las soluciones CAD se han convertido en realidad en un tipo de información independiente. Combina tanto la geometría de los elementos como la metainformación (estructurada o semi-estructurada), encerrada en archivos y formatos especializados.

CAD datos: del diseño al almacenamiento de datos

Los modernos sistemas CAD y BIM almacenan los datos en sus propios formatos, a menudo propietarios: DWG, DXF, RVT, DGN, PLN y otros. Estos formatos admiten representaciones de objetos tanto en 2D como en 3D y conservan no sólo la geometría, sino también los atributos asociados a los objetos. He aquí los más comunes:

- **DWG** es un formato de archivo binario utilizado para almacenar datos y metadatos de diseño bidimensionales (y, con menor frecuencia, tridimensionales).
- **DXF** es un formato de texto para el intercambio de dibujos 2D y 3D -drawings between CAD -systems. Contiene datos de geometría, capas y atributos, y admite representación ASCII y binaria.
- **RVT** es un formato binario de almacenamiento de modelos CAD que incluye la geometría 3D, los atributos de los elementos, las relaciones y los parámetros de diseño.
- **IFC** es un formato de texto abierto para el intercambio de datos de construcción entre sistemas CAD (BIM). Incluye geometría, propiedades de los objetos e información sobre sus relaciones.

Además de éstos, se utilizan otros formatos: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Aunque difieren en su finalidad y nivel de apertura (Fig, 3.1-14), todos ellos pueden representar el mismo modelo de información del proyecto de diferentes formas. En proyectos complejos, estos formatos suelen utilizarse en paralelo, desde el dibujo hasta la coordinación de los modelos del proyecto.

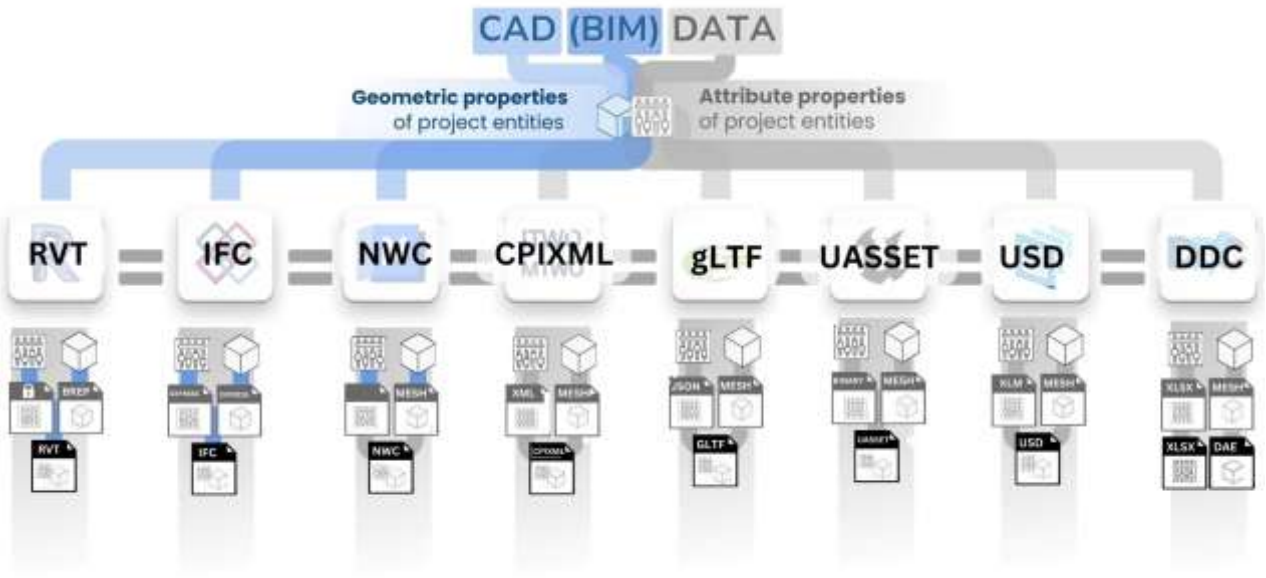


Fig. 3.1-14 Los formatos de almacenamiento CAD más populares describen la geometría mediante parámetros BREP o MESH, complementados con datos de atributos.

Todos los formatos anteriores permiten almacenar datos sobre cada elemento de un proyecto de construcción y todos ellos contienen dos tipos fundamentales de datos:

- **Parámetros geométricos**: describen la forma, la ubicación y las dimensiones de un objeto. La geometría y su uso se tratarán en detalle en la sexta parte del libro, dedicada a las soluciones CAD (BIM);
- **Propiedades de los atributos** - contienen información diversa: materiales, tipos de elementos, características técnicas, identificadores únicos y otras propiedades que pueden tener los elementos del proyecto.

Los datos de atributos revisten especial importancia en los proyectos modernos, ya que definen las características operativas de los objetos, permiten realizar cálculos de ingeniería y costes y facilitan la interacción de extremo a extremo entre los participantes en el diseño, la construcción y la explotación. Por ejemplo:

- Para ventanas y puertas: tipo de construcción, tipo de acristalamiento, dirección de apertura (Fig. 3.2-1).
- En el caso de los muros, se registra información sobre los materiales, el aislamiento térmico y el rendimiento acústico.
- Para los sistemas de ingeniería, se almacenan los parámetros de tuberías, conductos, rutas de cables y sus conexiones.

Estos parámetros pueden almacenarse tanto en los propios archivos CAD-(BIM -)como en bases de datos externas -como resultado de la exportación, conversión o acceso directo a estructuras CAD internas mediante herramientas de ingeniería inversa. Este enfoque facilita la integración de la información de diseño con otros sistemas y plataformas corporativos

La ingeniería inversa en el contexto de CAD (BIM) es el proceso de extracción y análisis de la estructura interna de un modelo digital para recrear su lógica, estructura de datos y dependencias sin acceso a los algoritmos o documentación originales.

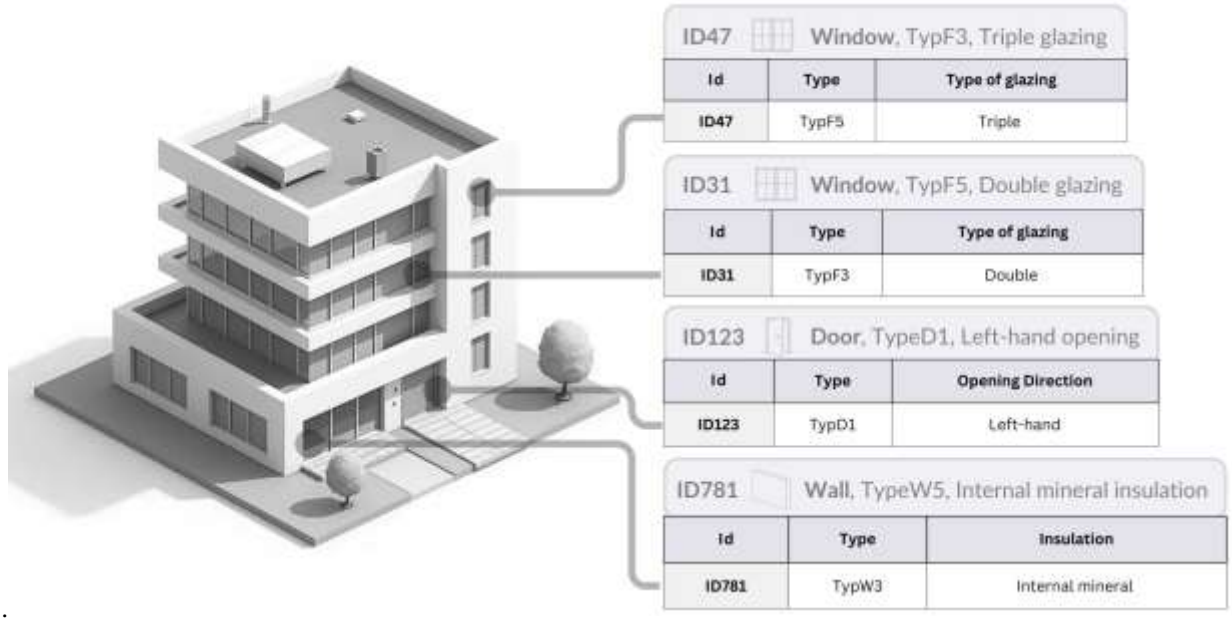


Fig. 3.1-15 Un elemento de proyecto, además de describir geometría paramétrica o poligonal, contiene información sobre parámetros y propiedades de elementos.

Como resultado, se forma un conjunto único de parámetros y propiedades en torno a cada elemento, que incluye tanto características únicas de cada objeto (por ejemplo, identificador y dimensiones) como atributos comunes para grupos de elementos. Esto permite no sólo analizar elementos-entidades individuales del proyecto, sino también combinarlos en grupos lógicos, que luego pueden utilizar otros especialistas para sus tareas y cálculos en sistemas y bases de datos.

Una entidad es un objeto concreto o abstracto del mundo real que puede identificarse, describirse y representarse unívocamente en forma de datos.

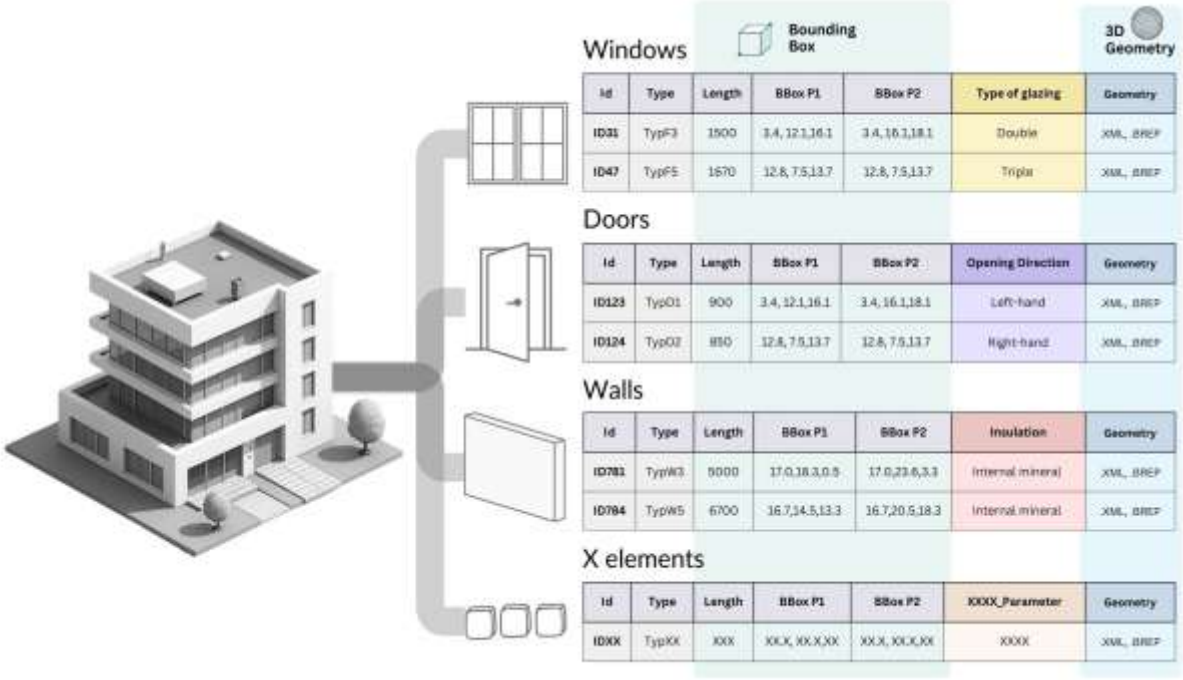


Fig. 3.1-16 Cada elemento del proyecto contiene atributos introducidos por el diseñador o calculados en el programa CAD.

En las últimas décadas, el sector de la construcción ha desarrollado muchos nuevos formatos CAD (BIM) que simplifican la creación, el almacenamiento y la transferencia de datos. Estos formatos pueden ser cerrados o abiertos, tabulares, paramétricos o gráficos. Sin embargo, su diversidad y fragmentación complican considerablemente la gestión de los datos en todas las fases del ciclo de vida del proyecto. En la Fig. 3.1-17 se presenta una tabla comparativa de los principales formatos utilizados para el intercambio de información en la construcción (versión completa disponible mediante código QR).

Para resolver los problemas de interoperabilidad y acceso a los datos CAD, intervienen gestores BIM (BIM) y coordinadores, cuya tarea es controlar las exportaciones, comprobar la calidad de los datos e integrar partes de los datos CAD (BIM) en otros sistemas.

Sin embargo, debido a la naturaleza cerrada y a la complejidad de los formatos, es difícil automatizar este proceso, lo que obliga a los especialistas a realizar muchas operaciones manualmente, sin poder construir procesos completos de tratamiento de datos en línea (pipeline).

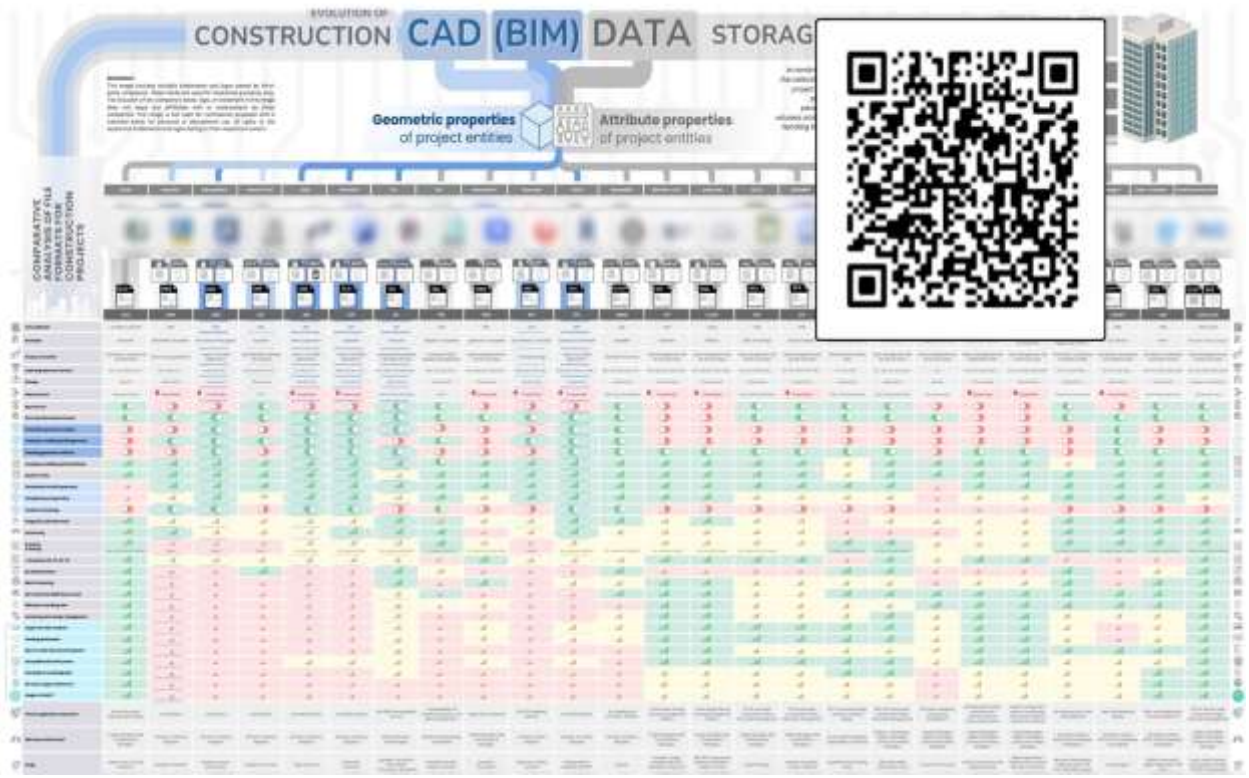


Fig. 3.1-17 Tabla comparativa de los principales formatos de datos en los que se almacena la información de los elementos del proyecto [53].

Para entender por qué existen tantos formatos de datos diferentes y por qué la mayoría de ellos son cerrados, es importante profundizar en los procesos que tienen lugar dentro de los programas CAD (BIM), que se explorarán en detalle en la sexta parte del libro.

Los desarrolladores de sistemas CAD introdujeron una capa de información adicional a la geometría en forma del concepto BIM (Building Information Modeling), un término de marketing promovido activamente en el sector de la construcción desde 2002 [54].

La aparición del concepto BIM (BOM) y el uso de CAD en los procesos

El concepto de Building Information Modelling (BIM), esbozado por primera vez en el BIM Whitepaper de 2002 [54], surgió de las iniciativas de marketing de los fabricantes de software de CAD. Surgió de las iniciativas de marketing de los desarrolladores de software CAD y fue un intento de adaptar principios ya bien establecidos en ingeniería mecánica a las necesidades del sector de la construcción.

La inspiración para BIM vino del concepto de BOM (Bill of Materials) - Lista de Materiales - que se viene utilizando ampliamente en la industria desde finales de los años ochenta. En ingeniería mecánica, BOM permitió vincular datos de sistemas CAD con PDM (Product Data Management), PLM (Product Lifecycle Management) y sistemas ERP, proporcionando una gestión holística de la información de ingeniería a lo largo de todo el ciclo de vida del producto (Fig. 3.1-8).

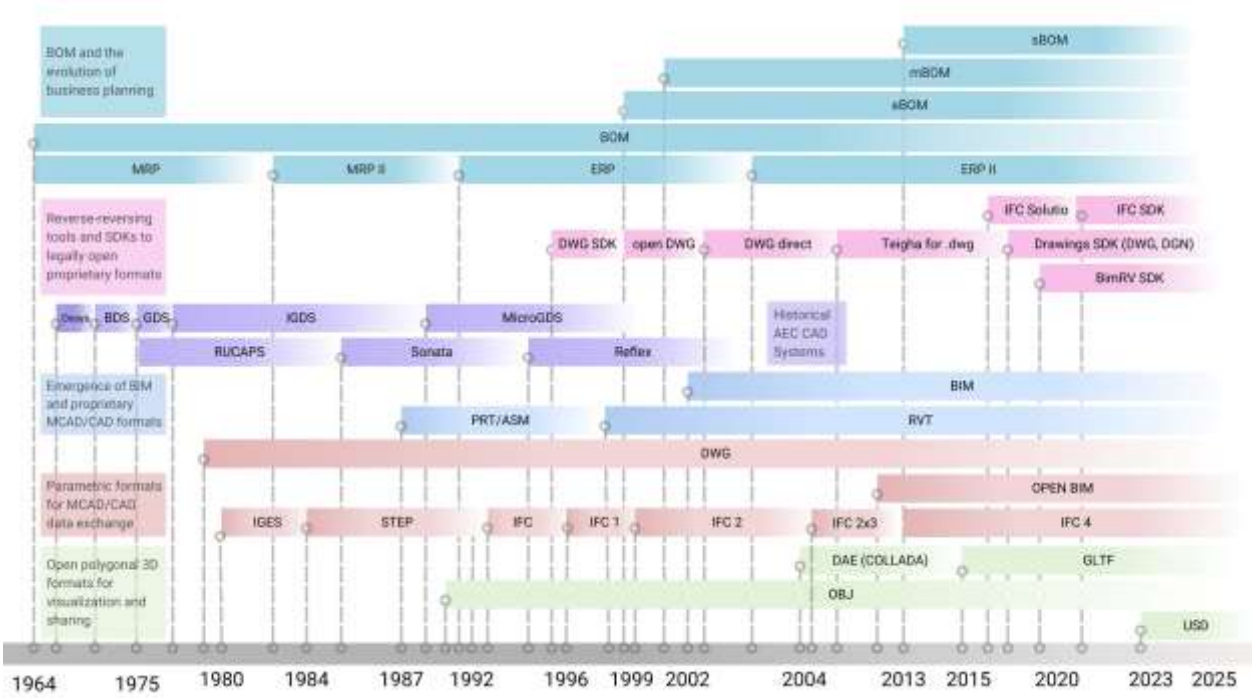


Fig. 3.1-18 Evolución de las especificaciones (lista de materiales), la modelización de la información (BIM) y los formatos digitales en la industria de la construcción de ingeniería.

El desarrollo moderno del concepto de lista de materiales ha dado lugar a la aparición de un marco ampliado: XBOM (Extended BOM), que incluye no sólo la composición del producto, sino también escenarios de comportamiento, requisitos operativos, parámetros de sostenibilidad y datos para análisis predictivos. XBOM cumple esencialmente la misma función que BIM en la construcción: ambos enfoques pretenden convertir el modelo digital en una fuente única de verdad para todos los participantes en el proyecto a lo largo de todo su ciclo de vida.

Un hito clave en la aparición de la lista de materiales en la construcción fue la introducción del primer CAD paramétrico (MCAD) adaptado específicamente al sector de la construcción en 2002. Fue desarrollado por el equipo que previamente había creado Pro-E®, un revolucionario sistema MCAD para ingeniería mecánica que había surgido a finales de la década de 1980 y se había convertido en un estándar del sector [55].

Ya a finales de la década de 1980, el objetivo era eliminar las limitaciones [56] de los programas CAD existentes entonces. El objetivo principal era reducir la mano de obra necesaria para modificar los parámetros de los elementos de diseño y posibilitar la actualización del modelo a partir de datos externos a los programas CAD mediante una base de datos [57]. En este sentido, el papel más importante debía desempeñarlo la parametrización: la recuperación automática de características de la base de datos y su utilización para actualizar el modelo dentro de los sistemas CAD.

> Pro-E y el concepto de modelado paramétrico elemental c BOM subyacente han tenido un impacto significativo en el desarrollo del mercado CAD - y MCAD - [58]. Este modelo lleva 25 años en la industria y muchos sistemas modernos se han convertido en sus sucesores conceptuales.

> *El objetivo es crear un sistema lo suficientemente flexible como para animar al ingeniero a considerar fácilmente diferentes diseños. Y el coste de realizar cambios en el diseño debe ser lo más cercano posible a cero. El software tradicional de CAD / CAM restringe de forma poco realista la realización de cambios poco costosos únicamente al principio del proceso de diseño [59].*
>
> - Samuel Heisenberg, fundador de Parametric Technology Corporation®, desarrollador del producto MCAD -Pro-E y profesor del creador de un producto CAD que utiliza el formato RVT

En ingeniería mecánica, los sistemas PDM, PLM, MRP y ERP se han convertido en plataformas clave. Desempeñan un papel central en la gestión de datos y procesos, recopilando información de los sistemas CAx (CAD, CAM, CAE) y organizando las actividades de diseño en función de la estructura del producto (lista de materiales: eBOM, pBOM, mBOM) (Fig. 3.1-18). Esta integración reduce los errores, evita la duplicación de datos y garantiza la trazabilidad de principio a fin, desde el diseño hasta la producción.

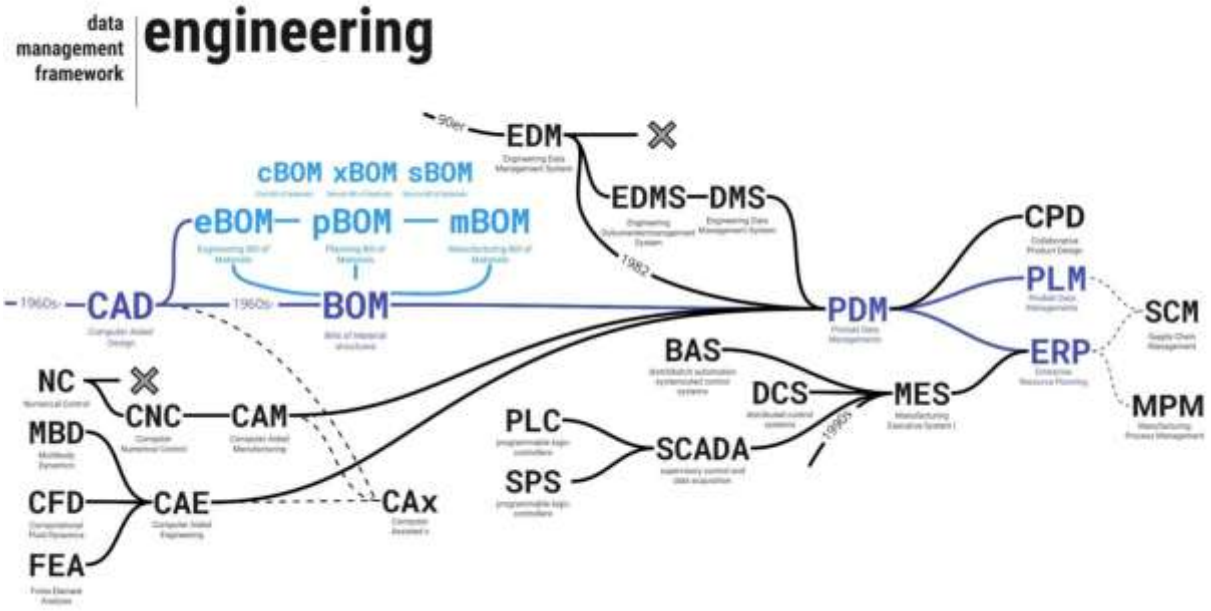


Fig. 3.1-19 Históricamente, la lista de materiales surgió en la década de 1960 como una forma de estructurar los datos de los sistemas CAx y pasarlos a los sistemas de control.

La compra por parte de uno de los principales proveedores de una solución CAD desarrollada por el antiguo equipo de Pro-E y basada en el enfoque de la lista de materiales estuvo marcada casi de inmediato por la publicación de la serie de libros blancos sobre BIM (2002-2003)[60][61]. Ya desde mediados de la década de 2000, el concepto BIM empezó a promoverse activamente en el sector de la construcción, lo que aumentó notablemente el interés por el software paramétrico. La popularidad creció tan rápidamente que la horquilla de construcción de ingeniería mecánica Pro-E - CAD paramétrico promovido por este proveedor - ha desplazado de hecho a sus competidores en el segmento del diseño arquitectónico y estructural (Fig.

3.1-20). A principios de la década de 2020, ha consolidado de facto el dominio mundial en el mercado de BIM (CAD) [62].

Fig. 3.1-20 Popularidad de las búsquedas en Google (RVT frente a IFC): el CAD paramétrico creado por el antiguo equipo de Pro-E con soporte de listas de materiales -BIM ha ganado popularidad en casi todos los países del mundo.

En los últimos 20 años, la abreviatura BIM ha adquirido multitud de interpretaciones, cuyos múltiples significados tienen su origen en los conceptos iniciales de marketing surgidos a principios de la década de 2000. La norma ISO 19650, que desempeñó un papel importante en la popularización del término, aseguró de hecho el estatus de BIM como enfoque "basado en la ciencia" de la gestión de la información. Sin embargo, en el texto de la propia norma, dedicado a la gestión de datos a lo largo del ciclo de vida de los objetos mediante BIM, se menciona la abreviatura BIM, pero nunca se define claramente

El sitio web original del proveedor, que publicó una serie de libros blancos sobre BIM en 2002[60] y 2003[61], reproducía en realidad materiales de marketing sobre los conceptos BOM (listas de materiales) y PLM (gestión del ciclo de vida del producto) utilizados anteriormente en el software de ingeniería mecánica Pro-E en los años noventa [63].

> *Building Information Modelling (BIM), un nuevo e innovador enfoque del diseño, la construcción y la gestión de edificios introducido por..... [nombre de la empresa proveedora de CAD] en 2002, ha cambiado la forma de pensar de los profesionales del sector de todo el mundo sobre cómo aplicar la tecnología al diseño, la construcción y la gestión de edificios.*
>
> - BIM Whitepaper, 2003 [61]

Estas primeras publicaciones vinculaban directamente el BIM con el concepto de base de datos integrada centralizada. Como se afirma en el Libro Blanco de 2003, BIM es la gestión de la información de construcción en la que todas las actualizaciones se producen en un único repositorio, garantizando que todos los planos, cortes y especificaciones (BOM - Bills of Materials) estén sincronizados.

> *BIM se describe como gestión de la información de construcción, donde todas las actualizaciones y todos los cambios tienen lugar en una base de datos. De este modo, ya se trate de esquemas, secciones o planos, todo está siempre coordinado, es coherente y está actualizado.*
>
> - Página web de una empresa de CAD vendor with BIM Whitepaper, 2003 [54].

La idea de gestionar el diseño a través de una única base de datos integrada se ha debatido ampliamente ya en investigaciones de los años ochenta. Por ejemplo, el concepto de BDS de Charles Eastman [57] incluía 43 referencias al término "base de datos" (Fig, 6.1-2). En 2004, este número se había reducido casi a la mitad, a 23, en el Libro Blanco de 2002 sobre BIM [64]. Y a mediados de la década de 2000, el tema de las bases de datos prácticamente había desaparecido de los materiales de marketing de los proveedores y de la agenda de digitalización en general.

Aunque en un principio la base de datos y el acceso a ella se concibieron como el núcleo del sistema BIM, con el tiempo el énfasis se ha desplazado a la geometría, la visualización y el 3D. El propio registrador de la norma IFC en 1994, que publicó el Libro Blanco BIM en 2002 -el mismo vendedor-, señalaba explícitamente en el Libro Blanco de principios de la década de 2000 las limitaciones de formatos neutros como IGES, STEP e IFC y la necesidad de acceder directamente a las bases de datos CAD:

> *Las distintas aplicaciones pueden ser incompatibles y los datos reintroducidos pueden ser inexactos [...]. El resultado del diseño tradicional asistido por ordenador [CAD]: costes más elevados, plazos de comercialización más largos y menor calidad del producto. Hoy en día, todas las aplicaciones importantes utilizan interfaces estándar de la industria para el intercambio de datos de bajo nivel. Si se utilizan las antiguas normas IGES o las nuevas STEP [IFC es una copia de facto y de jure del formato STEP/IGES] para intercambiar datos entre aplicaciones de distintos proveedores, los usuarios pueden lograr cierta compatibilidad de datos entre los mejores productos. Pero IGES y STEP sólo funcionan a niveles bajos y no pueden intercambiar datos tan ricos como la información generada por las principales aplicaciones actuales [...]. Y aunque estas y otras normas mejoran casi a diario, siempre irán por detrás de los productos de los proveedores actuales en cuanto a riqueza de datos. [...] los programas de una aplicación deben poder intercambiar y conservar la riqueza de datos sin recurrir a traductores neutrales como IGES, STEP [IFC] o PATRAN. En su lugar, las aplicaciones marco deben poder acceder directamente a la base de datos CAD subyacente para no perder el detalle y la precisión de la información.*
>
> - Whitepaper del proveedor de CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Así, ya en los años 80 y principios de los 2000, se consideraba que el elemento clave del diseño digital en el entorno CAD era la base de datos y no el formato-archivo o el formato neutro IFC. Se sugería abandonar los traductores y que las aplicaciones tuvieran acceso directo a los datos. Sin embargo, en realidad, a mediados de la década de 2020, el concepto de BIM empezó a parecerse a una estrategia de "divide y vencerás", en la que se priorizan los intereses de los proveedores de software que utilizan núcleos geométricos cerrados frente al desarrollo de un intercambio de información abierto.

Hoy en día, BIM se percibe como parte integrante del sector de la construcción. Pero en las dos últimas décadas, las promesas de simplificar la colaboración y la integración de datos han quedado en gran medida incumplidas. La mayoría de las soluciones siguen ligadas a formatos cerrados o neutros y a herramientas especializadas. Analizaremos en detalle la historia del BIM, el BIM abierto y el IFC, así como los problemas de interoperabilidad y los núcleos geométricos en la parte 6 del libro "CAD y BIM: marketing, realidad y el futuro de los datos de diseño en la construcción".

> Hoy en día, el sector se enfrenta al reto clave de pasar de la concepción tradicional del CAD (BIM) como herramienta de modelado a su uso como base de datos de pleno derecho. Para ello es necesario adoptar nuevos enfoques a la hora de trabajar con la información, abandonar la dependencia de los ecosistemas cerrados y aplicar soluciones abiertas.

Con el desarrollo de herramientas de ingeniería inversa que permiten acceder a bases de datos CAD, así como la difusión de las tecnologías de código abierto y LLM, los usuarios y desarrolladores del sector de la construcción se alejan cada vez más de los términos vagos de los proveedores de software. En su lugar, la atención se está desplazando hacia lo que realmente importa: los datos (bases de datos) y los procesos.

> Detrás de las siglas y visualizaciones de moda se esconden prácticas estándar de gestión de datos: almacenamiento, transferencia y transformación, es decir, el clásico proceso ETL (Extraer, Transformar, Cargar). Al igual que en otros sectores, la digitalización de la construcción requiere no sólo normas de intercambio, sino también una gestión claramente estructurada de la información heterogénea.

Para aprovechar plenamente el potencial de los datos CAD (BIM), las empresas deben replantearse su enfoque de la gestión de la información. Esto conducirá inevitablemente a un elemento clave de la transformación digital: la unificación, normalización y estructuración significativa de los datos con los que los profesionales de la construcción trabajan a diario.

CAPÍTULO 3.2.
ARMONIZACIÓN Y ESTRUCTURACIÓN DE DATOS

Sistemas de llenado de datos en el sector de la construcción

Ya se trate de grandes corporaciones o de medianas empresas, los especialistas se dedican a diario a llenar de información multiformato sistemas informáticos y bases de datos con diversas interfaces (Fig. 3.2-1) que, con la ayuda de los gestores, deben cooperar entre sí. Es este complejo de sistemas y procesos en interacción el que, en última instancia, genera ingresos y beneficios para la empresa.

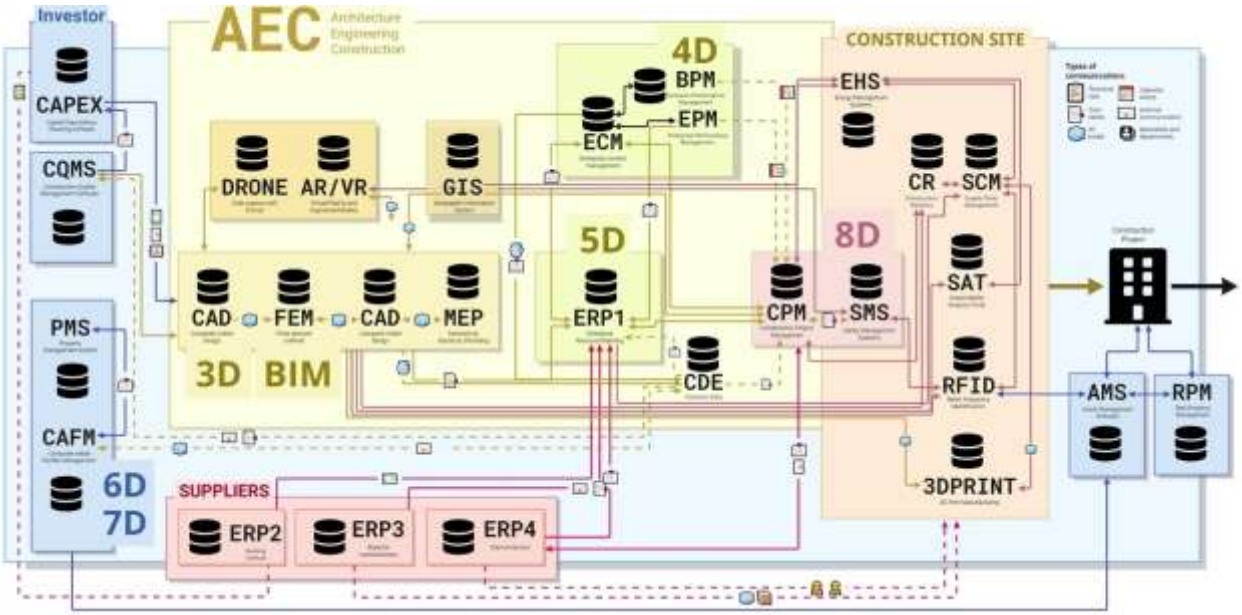


Fig. 3.2-1 Prácticamente todos los sistemas o aplicaciones utilizados en el sector de la construcción tienen en su núcleo una de las populares bases de datos RDBMS.

Cada una de las categorías de sistemas mencionadas anteriormente y aplicadas en el sector de la construcción trabaja con sus propios tipos de datos, correspondientes al papel funcional de estos sistemas. Para pasar del nivel abstracto al concreto, pasamos de los tipos de datos a su representación en forma de formatos y documentos.

A la lista de sistemas facilitada anteriormente (Fig, 1.2-4), añadimos ahora los tipos concretos de formatos y documentos con los que suelen trabajar:

- **Inversor (CAPEX)**
  - Datos financieros: presupuestos, previsiones de gastos (datos estructurados).
  - Datos sobre tendencias del mercado: análisis de mercado (datos estructurados y no estructurados).
  - Datos legales y contractuales: contratos (datos de texto).

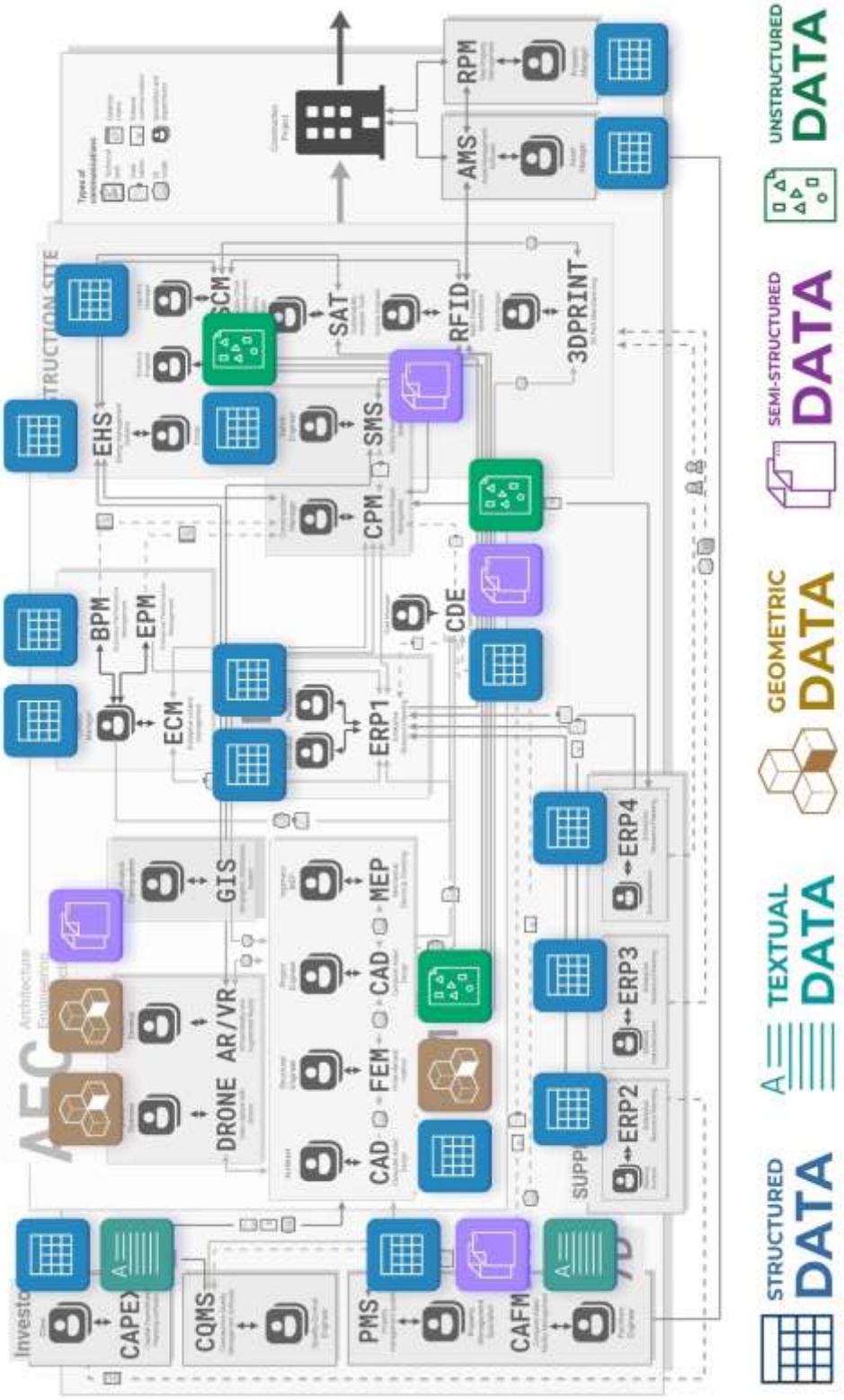


Fig. 3.2-2 El sector de la construcción utiliza muchos sistemas con diferentes interfaces que tratan distintos tipos de datos.

- **Sistemas de gestión (PMS, CAFM, CQMS)**
  - Datos del proyecto: gráficos, tareas (datos estructurados).
  - Datos de mantenimiento de las instalaciones: planes de mantenimiento (texto y datos semiestructurados).
  - Datos de control de calidad: normas, informes de inspección (datos textuales y no estructurados).
- **CAD, MEF y BIM**
  - Dibujos técnicos: planos arquitectónicos, estructurales (datos geométricos, datos no estructurados).
  - Modelos de edificios: 3D -modelos, datos de materiales (datos geométricos y semiestructurados).
  - Cálculos de ingeniería: análisis de cargas (datos estructurados).
- **Sistemas de gestión de obras (EHS, SCM)**
  - Datos de seguridad y salud: protocolos de seguridad (datos textuales y estructurados).
  - Datos de la cadena de suministro: inventarios, pedidos (datos estructurados).
  - Informes diarios: horas de trabajo, productividad (datos estructurados).
- **Drones, AR/VR, GIS, 3D -printing**
  - Geodatos: mapas topográficos (datos geométricos y estructurados).
  - Datos en tiempo real: vídeo y fotos (datos no estructurados).
  - Modelos para impresión 3D: dibujos digitales (datos geométricos).
- **Sistemas de gestión adicionales (4D BPM, 5D ERP1)**
  - Datos sobre plazos y costes: calendarios, estimaciones (datos estructurados).
  - Gestión de cambios: registros de cambios del proyecto (texto y datos estructurados).
  - Informes de resultados: indicadores de éxito (datos estructurados).
- **Integración y comunicación de datos (CDE, RFID, AMS, RPM)**
  - Intercambio de datos: intercambio de documentos, modelos de datos (datos estructurados y textuales).
  - RFID y datos de seguimiento: logística, gestión de activos (datos estructurados).
  - Vigilancia y control: sensores en las obras (datos estructurados y no estructurados).

Así, cada sistema del sector de la construcción -desde los sistemas de gestión de obras hasta las bases de datos operativas- funciona con su propio tipo de información: estructurada, textual, geométrica y otras. El "paisaje de datos" con el que los profesionales tienen que trabajar a diario es extremadamente diverso. Sin embargo, una simple enumeración de formatos no revela la complejidad del trabajo real con la información.

En la práctica, las empresas se enfrentan al hecho de que los datos, incluso cuando se recuperan de los sistemas, no están listos para ser utilizados "tal cual". Esto es especialmente cierto en el caso de textos, imágenes, PDF, archivos CAD y otros formatos difíciles de analizar con herramientas estándar. Por eso, el siguiente paso clave es la transformación de datos, un proceso sin el cual el tratamiento, el análisis, la visualización y la toma de decisiones no pueden automatizarse eficazmente.

Transformación de datos: la base fundamental del análisis empresarial moderno

Hoy en día, la mayoría de las empresas se encuentran ante una paradoja: alrededor del 80% de sus procesos diarios siguen basándose en datos estructurados clásicos: las conocidas hojas de cálculo de Excel y las bases de datos relacionales (RDBMS) [66]. Sin embargo, al mismo tiempo, el 80% de la nueva información que entra en el ecosistema digital de las empresas está desestructurada o poco estructurada (Fig. 3.2-3) [52]. Esto incluye texto, gráficos, geometría, imágenes, modelos CAD, documentación en PDF, grabaciones de audio y vídeo, correspondencia electrónica y mucho más.

Además, el volumen de datos no estructurados sigue creciendo rápidamente: se calcula que la tasa de crecimiento anual es del 55-65% [67]. Esta dinámica crea serias dificultades para integrar la nueva información en los procesos empresariales existentes. Ignorar este flujo de datos multiformato conduce a la formación de vacíos de información y reduce la capacidad de gestión de todo el entorno digital de la empresa.

Fig. 3.2-3 El crecimiento anual de los datos no estructurados plantea retos a la hora de integrar la información en flujo en los procesos empresariales.

> Ignorar los complejos datos no estructurados y los confusos datos poco estructurados en los procesos de automatización puede provocar importantes lagunas en el panorama informativo de una empresa. En el mundo actual, en el que la información se mueve de forma incontrolable y como una avalancha, las empresas deben adoptar un enfoque híbrido de la gestión de datos que incluya métodos eficaces para tratar todo tipo de datos.

La clave de una gestión de datos eficaz reside en organizar, estructurar y clasificar distintos tipos de datos "Babel" (incluidos los formatos no estructurados, textuales y geométricos, en datos estructurados o poco estructurados). Este proceso transforma conjuntos de datos caóticos en estructuras organizadas para su integración en sistemas, lo que permite tomar decisiones basadas en ellos (Fig, 3.2-4).

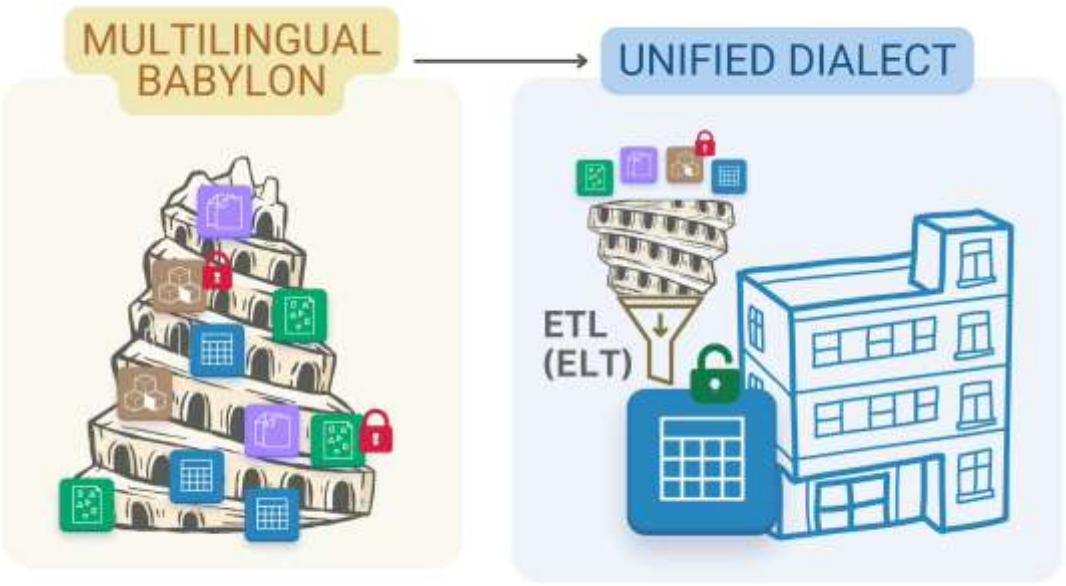


Fig. 3.2-4 La principal tarea de los departamentos de gestión de datos es traducir la "Babilonia" de datos diversos y multiformato a un sistema estructurado y categorizado.

Uno de los principales obstáculos a esta armonización sigue siendo el bajo nivel de interoperabilidad entre las distintas plataformas digitales, los "silos" de los que hemos hablado en los capítulos anteriores.

> Según el informe, el Instituto Nacional de Normas y Tecnología (NIST, EE.UU.) subraya [68] que la escasa compatibilidad de datos entre distintas plataformas de construcción provoca pérdidas de información y costes adicionales significativos. Sólo en 2002, los problemas de interoperabilidad de los programas informáticos causaron pérdidas en la construcción de capital en EE.UU. por valor de 15.800 millones de dólares al año, de los cuales dos tercios corresponden a los propietarios y operadores de los edificios, especialmente durante su funcionamiento y mantenimiento [68]. El estudio también señala que la normalización de los formatos de datos puede reducir estas pérdidas y mejorar la eficiencia a lo largo del ciclo de vida de las instalaciones.

Según el estudio CrowdFlower de 2016 [69], que abarcó a 16.000 científicos de datos de todo el mundo, el principal problema siguen siendo los datos "sucios" y multiformato. Según este estudio, el recurso más valioso no son las bases de datos finales o los modelos de aprendizaje automático, sino el tiempo dedicado a preparar la información.

La limpieza, el formateo y la organización ocupan hasta el 60% del tiempo de un analista y de un gestor de datos. Casi una quinta parte se dedica a buscar y recopilar los conjuntos de datos adecuados, a menudo ocultos en silos e inaccesibles para el análisis. Y sólo un 9% del tiempo se dedica directamente a modelizar, analizar, hacer predicciones y probar hipótesis. El resto se dedica a comunicar, visualizar, elaborar informes y buscar fuentes de información de apoyo

Por término medio, el trabajo de datos de un directivo se distribuye de la siguiente manera (Fig, 3.2-5):

- **Limpieza y organización de datos (60%):** disponer de datos limpios y estructurados puede reducir

considerablemente el tiempo de trabajo del analista de y acelerar el proceso de realización de tareas.

- **Recopilación de datos (19%):** Un reto importante para los profesionales de la ciencia de datos es encontrar conjuntos de datos relevantes. A menudo, los datos de las empresas se apilan en "silos" organizados caóticamente, lo que dificulta el acceso a la información que necesitan.
- **Modelización/aprendizaje automático (9%):** A menudo se complica por la falta de claridad de los objetivos empresariales por parte de los clientes. La falta de una declaración de objetivos clara puede anular el potencial incluso del mejor modelo.
- **Otras tareas (5%):** además de procesar datos, los analistas tienen que ocuparse de investigar, explorar los datos desde distintos ángulos, comunicar los resultados mediante visualizaciones e informes y recomendar la optimización de procesos y estrategias.

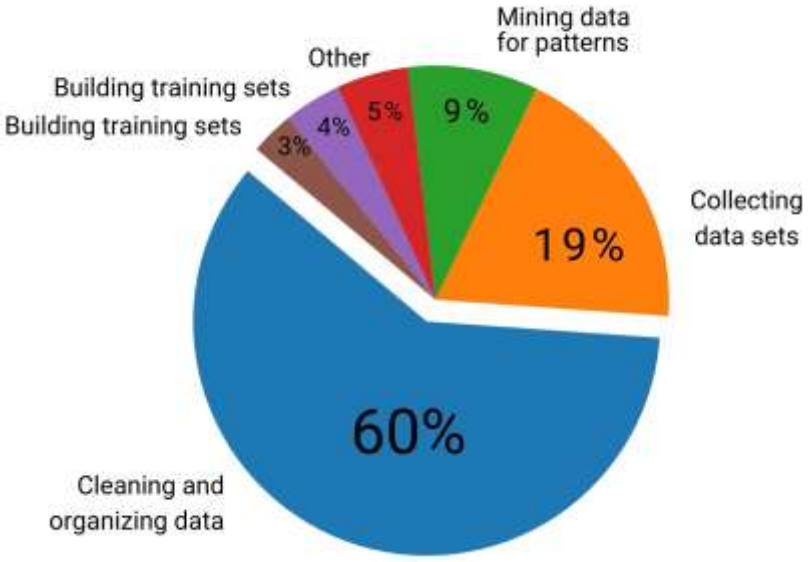


Fig. 3.2-5 A qué dedican más tiempo los gestores de datos que trabajan con datos (basado en [70]).

> Estas estimaciones están respaldadas por otros estudios. Según el estudio Xplenty publicado en BizReport en 2015 [71], entre el 50% y el 90% del tiempo de los profesionales de inteligencia empresarial (BI) (BI) se dedica a preparar los datos para el análisis.

La limpieza, validación y organización de los datos de representa una base fundamental para todos los procesos posteriores de datos y análisis, y ocupa hasta el 90% del tiempo de los científicos de datos.

Este minucioso trabajo, invisible para el usuario final, es crucial. Los errores en los datos brutos distorsionan inevitablemente los análisis, son engañosos y pueden dar lugar a costosos errores de gestión. Por eso los procesos de limpieza y normalización de datos -desde eliminar duplicados y rellenar omisiones hasta armonizar unidades de medida y alinearse con un modelo común- se están convirtiendo en piedra angular de la estrategia digital moderna.

Así, la transformación, limpieza y normalización exhaustivas de los datos no sólo ocupan la mayor parte

del tiempo de los especialistas (hasta el 80% del trabajo con datos), sino que también determinan la posibilidad de su utilización eficaz en el marco de los procesos empresariales modernos. Sin embargo, la organización y la limpieza de los datos no agotan por sí solas la tarea de gestión óptima de los flujos de información de una empresa. Durante la etapa de organización y estructuración, se convierte en la elección de un modelo de datos adecuado, que afecta directamente a la comodidad y eficacia del trabajo con la información en las etapas posteriores de procesamiento.

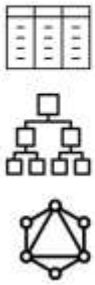
Dado que los datos y los objetivos empresariales son diferentes, es importante comprender las características de los modelos de datos y ser capaz de seleccionar o crear la estructura adecuada. Según el grado de estructuración y la forma de describir las relaciones entre elementos, existen tres modelos principales: estructurado, poco estructurado y gráfico. Cada uno es adecuado para tareas diferentes y tiene sus propios puntos fuertes y débiles.

Modelos de datos: relaciones en los datos y relaciones entre elementos

En los sistemas de información, los datos se organizan de distintas maneras, en función de las tareas y necesidades de almacenamiento, tratamiento y transmisión de la información. La diferencia clave entre los tipos de modelos de datos, la forma en que se almacena la información, es el grado de estructuración y la manera en que se describen las relaciones entre los elementos.

> Los datos estructurados tienen un esquema claro y repetible: se organizan en forma de tablas con columnas fijas. Este formato proporciona previsibilidad, facilidad de procesamiento y eficacia a la hora de realizar consultas SQL, filtrado y agregación. Ejemplos: bases de datos (RDBMS), Excel, CSV.

Datos poco estructurados permite una estructura flexible: distintos elementos pueden contener diferentes conjuntos de atributos y almacenarse como jerarquías. Algunos ejemplos son JSON, XML u otros formatos de documento. Estos datos son convenientes cuando es necesario modelar objetos anidados y relaciones entre ellos, pero, por otro lado, complican el análisis y la normalización de los datos (Fig. 3.2-6).



Data Model	Storage Format	Example
Relational	CSV, SQL	A table of doors in Excel
Hierarchical	JSON, XML	Nested door objects inside a room
Graph-based	RDF, GraphDB	Relationships between building elements

Fig. 3.2-6 Un modelo de datos es una estructura lógica que describe cómo se organizan, almacenan y procesan los datos en un sistema.

La elección del formato adecuado depende de los objetivos:

- Si la velocidad de filtrado y análisis es importante - tablas relacionales (SQL, CSV, RDBMS, bases de datos columnares) serán suficientes.
- Si se requiere flexibilidad en la estructura, es mejor utilizar JSON o XML.
- Si los datos tienen relaciones complejas, las bases de datos gráficas proporcionan visibilidad y escalabilidad.

En las bases de datos relacionales clásicas (RDBMS), cada entidad (por ejemplo, una puerta) se representa mediante una fila y sus propiedades mediante columnas de la tabla. Por ejemplo, una tabla de elementos de la categoría "Puertas" puede contener los campos ID, Altura, Anchura, Resistencia al fuego e ID de la sala que indica la sala (Fig. 3.2-7).

En las bases de datos relacionales clásicas (RDBMS) las relaciones se forman en forma de tablas, donde cada registro representa un objeto y las columnas sus parámetros. En el formato tabular, los datos sobre puertas del proyecto tienen este aspecto, en el que cada fila representa un elemento independiente: una puerta con su identificador único y sus atributos, y la conexión con la habitación se realiza a través del parámetro "ID de habitación".

Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Fig. 3.2-7 Información sobre los tres elementos de la categoría "Puertas" del proyecto en forma de tabla estructurada.

En formatos poco estructurados como JSON o XML, los datos se almacenan de forma jerárquica o anidada, donde los elementos pueden contener otros objetos y su estructura puede variar. Esto permite modelizar relaciones complejas entre elementos. Una información similar sobre las puertas del proyecto, que se registró de forma estructurada (Fig, 3.2-7), se representa en un formato poco estructurado (JSON) de tal manera (Fig, 3.2-8) que se convierten en objetos anidados dentro de Rooms (Rooms - ID), lo que refleja lógicamente la jerarquía.

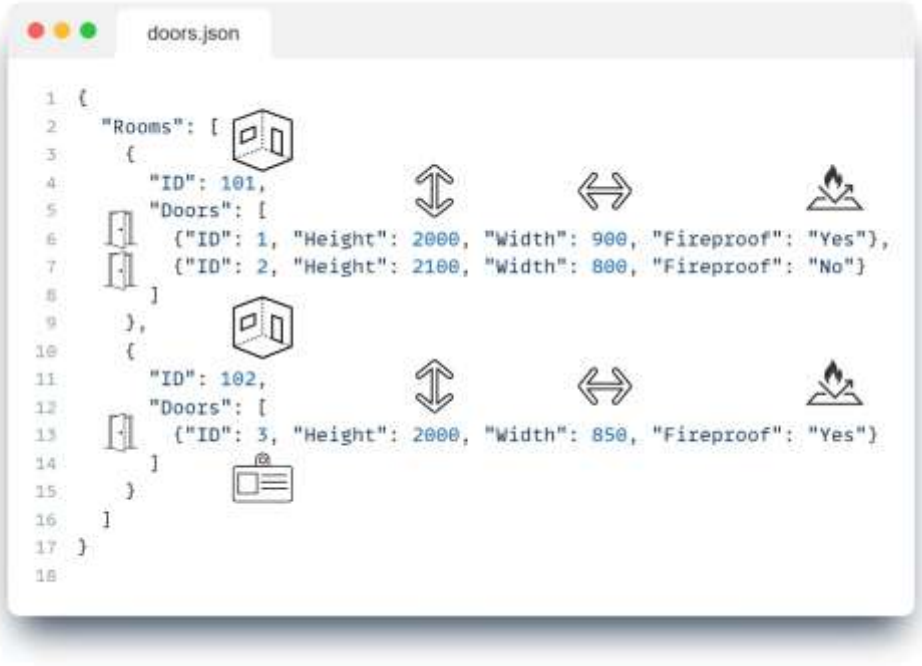


Fig. 3.2-8 Información sobre los elementos de la categoría "Puertas" del proyecto en formato JSON.

En un modelo gráfico, los datos se representan como nodos (vértices) y enlaces (aristas) entre ellos. Esto permite visualizar las complejas relaciones entre los objetos y sus atributos. En el caso de los datos de puertas y habitaciones del proyecto, la representación gráfica de es la siguiente:

- **Los nodos (nodos**) representan las entidades principales: salas (Sala 101, Sala 102) y puertas (ID1001, ID1002, ID1003)
- **Las nervaduras (enlaces)** muestran las relaciones entre estas entidades, por ejemplo, si una puerta pertenece a una determinada habitación
- **Los atributos** se asignan a los nodos y contienen las propiedades de la entidad (altura, anchura, resistencia al fuego de las puertas).

Fig. 3.2-9 Información de la entidad de la puerta del proyecto en vista gráfica.

En el modelo gráfico de datos de descripción de puertas, cada habitación y cada puerta son nodos independientes. Las puertas están vinculadas a las habitaciones mediante aristas, que indican que la puerta pertenece a una habitación determinada. Los atributos de las puertas (altura, anchura, resistencia al fuego) se almacenan como propiedades de los nodos correspondientes. En el capítulo "Surgimiento de la semántica y la ontología en la construcción" se tratarán más detalles sobre los formatos de grafos y cómo surgió la semántica de grafos en el sector de la construcción.

Las bases de datos de grafos son eficaces cuando lo importante son las relaciones entre los datos y no los datos en sí, como en los sistemas de recomendación, los sistemas de enrutamiento o al modelizar relaciones complejas en proyectos de gestión de instalaciones. El formato gráfico simplifica la creación de nuevas relaciones al permitir añadir nuevos tipos de datos al gráfico sin cambiar la estructura de almacenamiento. Sin embargo, en comparación con las tablas relacionales y los formatos estructurados, en un grafo no hay conectividad de datos adicional: la transferencia de datos de bases de datos bidimensionales a un grafo no aumenta el número de relaciones ni permite obtener nueva información.

> La forma y el esquema de los datos deben adaptarse al caso de uso específico y a las tareas que se vayan a realizar. Para trabajar eficazmente en los procesos empresariales, es importante utilizar aquellas herramientas y aquellos modelos de datos que ayuden a obtener resultados de la forma más rápida y sencilla posible.

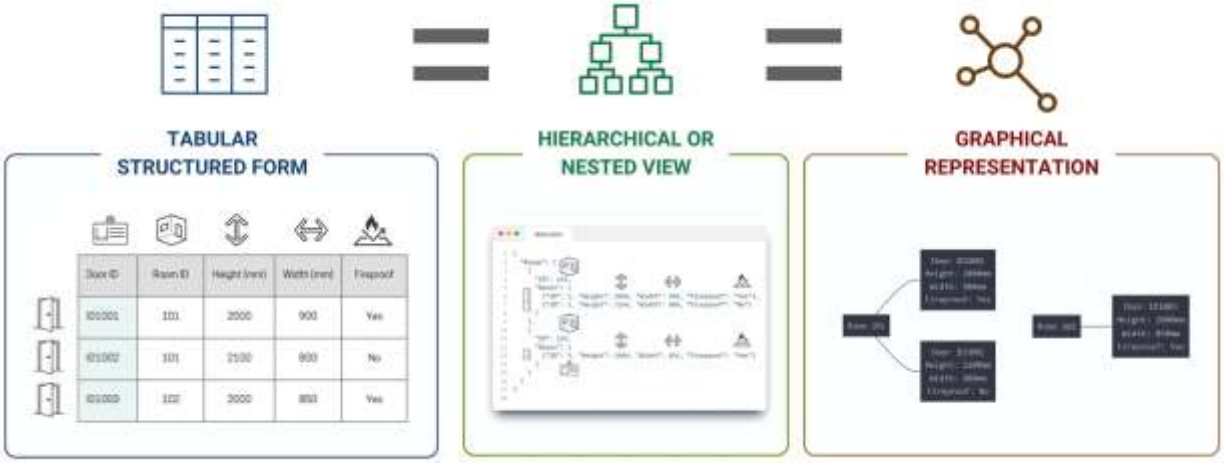


Fig. 3.2-10 La misma información sobre los elementos del proyecto puede almacenarse en distintos formatos utilizando diferentes modelos de datos.

> Hoy en día, la mayoría de las grandes empresas se enfrentan al problema de la excesiva complejidad de los datos. Cada una de los cientos o miles de aplicaciones utiliza su propio modelo de datos, lo que crea una complejidad excesiva: un modelo individual es a menudo decenas de veces más complejo de lo necesario, y el conjunto de todos los modelos es miles de veces más complejo. Esta complejidad excesiva dificulta considerablemente el trabajo tanto de los desarrolladores como de los usuarios finales.

Tal complejidad impone serias limitaciones al desarrollo y mantenimiento de los sistemas de la empresa. Cada nuevo elemento del modelo requiere código adicional, implementación de nueva lógica, pruebas exhaustivas y adaptación a las soluciones existentes. Todo ello aumenta los costes y ralentiza el trabajo del equipo de automatización de la empresa, convirtiendo incluso las tareas más sencillas en procesos costosos y lentos.

La complejidad afecta a todos los niveles de la arquitectura de datos. En las bases de datos relacionales se manifiesta en el creciente número de tablas y columnas, a menudo redundantes. En los sistemas orientados a objetos, la complejidad aumenta por la multiplicidad de clases y propiedades interrelacionadas. En formatos como XML o JSON, la complejidad se manifiesta a través de confusas estructuras anidadas, claves únicas y esquemas incoherentes.

> La excesiva complejidad de los modelos de datos hace que los sistemas no sólo sean menos eficientes, sino también difíciles de entender por los usuarios finales y, en el futuro, grandes modelos lingüísticos y agentes LLM. Es el problema de la comprensión y la complejidad de los modelos de datos y el procesamiento de datos lo que plantea la siguiente cuestión: cómo hacer que los datos sean lo suficientemente fáciles de usar como para que empiecen a ser útiles rápidamente.

Incluso cuando los modelos de datos se eligen sabiamente, su utilidad se reduce drásticamente si el acceso a los datos está restringido. Los formatos propietarios y las plataformas cerradas dificultan la integración, complican la automatización y quitan el control de la información propietaria, creando no solo un silo de datos nuevos, sino un silo cerrado al que solo se puede acceder con autorización del proveedor. Para comprender la magnitud del problema, es importante considerar exactamente cómo afectan los sistemas cerrados a los procesos digitales en la construcción.

Formatos propietarios y su impacto en los procesos digitales

Uno de los principales retos a los que se enfrentan las empresas de construcción durante la digitalización es el acceso limitado a los datos. Esto dificulta la integración de sistemas, reduce la calidad de la información y dificulta la organización de procesos eficientes. El uso de formatos propietarios y soluciones de software cerradas suele ser la causa de estas dificultades.

Por desgracia, hasta ahora, muchos programas utilizados en el sector de la construcción sólo permiten al usuario guardar datos en formatos propietarios o en la nube, a los que sólo se puede acceder a través de interfaces estrictamente limitadas. Y no es raro que estas soluciones se construyan en dependencia de sistemas aún más cerrados de proveedores más grandes. Como resultado, incluso los desarrolladores que desearían ofrecer arquitecturas más abiertas se ven obligados a cumplir las normas dictadas por los grandes proveedores.

Mientras que los modernos sistemas de gestión de datos de construcción admiten cada vez más formatos y estándares abiertos (Fig. 3.1-5), las bases de datos basadas en CAD (BIM) y los sistemas ERP y CAFM relacionados siguen siendo "islas" propietarias aisladas en el paisaje digital de la industria (Fig. 3.2-11).

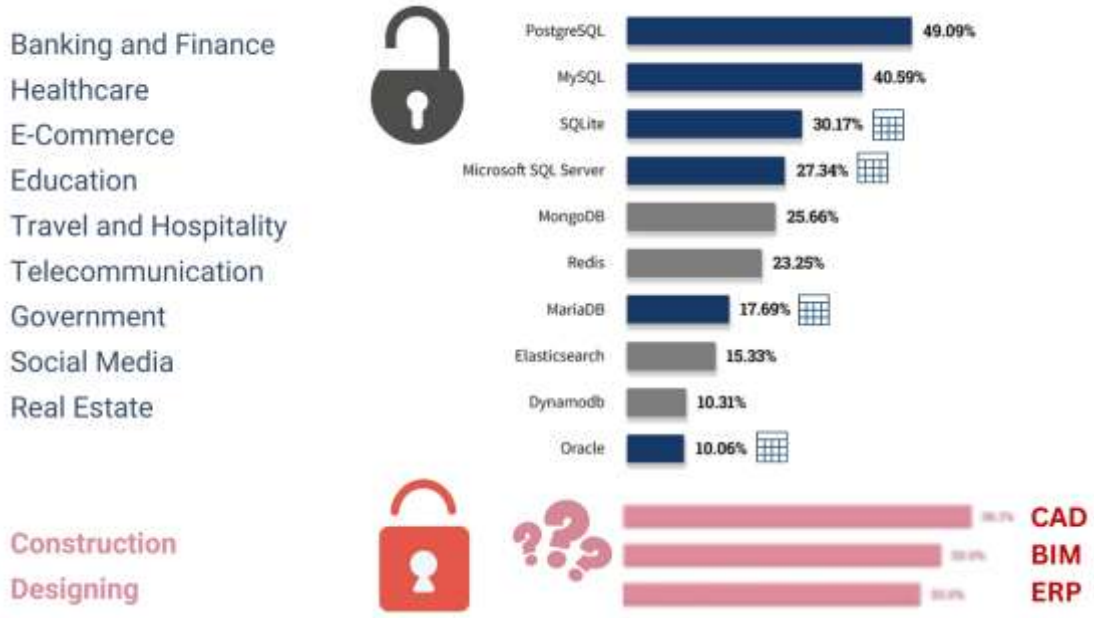


Fig. 3.2-11 La naturaleza cerrada y propietaria de los datos crea barreras a su integración y acceso.

Los formatos y protocolos cerrados y monopolizados no son un problema exclusivo del sector de la construcción. En muchos sectores de la economía, la lucha contra las normas cerradas y el acceso limitado a los datos comenzó con la ralentización de la innovación (Fig, 3.2-12), la existencia de barreras artificiales a la entrada de nuevos actores y la profundización de la dependencia de los grandes proveedores. Con el rápido crecimiento de la importancia de los datos, las autoridades de competencia simplemente no tienen tiempo para responder a los retos que plantean los nuevos mercados digitales y, como resultado, los formatos cerrados y el acceso restringido a los datos se convierten esencialmente en "fronteras" digitales que limitan el flujo de información y el crecimiento [63].

Si las máquinas producen todo lo que necesitamos, nuestra situación dependerá de cómo se distribuyan estos bienes. Todo el mundo sólo podrá disfrutar de una vida de prosperidad si se reparte la riqueza producida por las máquinas. O la mayoría de la gente acabará viviendo en la más absoluta pobreza si los propietarios de automóviles consiguen presionar contra la redistribución de la riqueza. De momento, las cosas parecen ir por el segundo camino, ya que la tecnología conduce a una desigualdad cada vez mayor [72].

> - Stephen Hawking, astrofísico, 2015

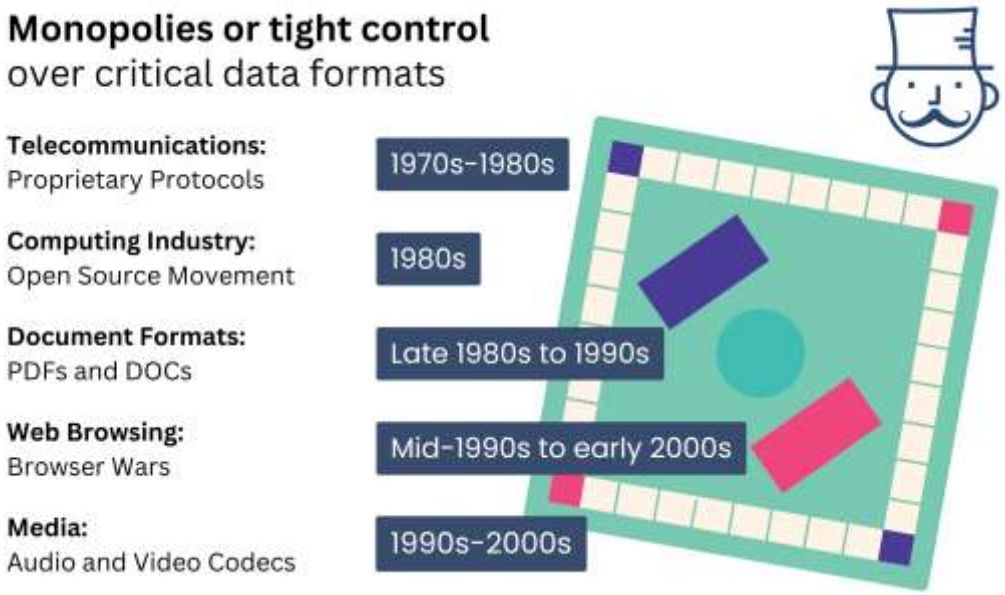


Fig. 3.2-12 La propiedad monopolística de formatos y protocolos de datos clave no es un problema exclusivo del sector de la construcción.

En consecuencia, debido al acceso cerrado a las bases de datos programas, los gestores de datos, los analistas de datos, los especialistas en TI y los desarrolladores que crean aplicaciones para el acceso, el tratamiento y la automatización de datos en el sector de la construcción se enfrentan hoy a numerosas dependencias de los proveedores de software (Fig. 3.2-13). Estas dependencias en forma de capas de acceso adicionales requieren la creación de soluciones con API especializadas -conexiones y herramientas y software especiales.

> Una API (Interfaz de Programación de Aplicaciones) es una interfaz formalizada a través de la cual un programa puede interactuar con otro, intercambiando datos y funcionalidades sin tener que acceder al código fuente. Una API describe qué peticiones puede hacer un sistema externo, qué formato deben tener y qué respuestas recibirá. Es un "contrato" normalizado entre módulos de software.

El gran número de dependencias de soluciones cerradas hace que toda la arquitectura de código y la lógica de los procesos empresariales de una empresa se conviertan en una "arquitectura espagueti" de herramientas dependientes de la política del proveedor de software para ofrecer un acceso de calidad a los datos.

La dependencia de soluciones y plataformas cerradas no sólo conlleva una pérdida de flexibilidad, sino también riesgos empresariales reales. Cambiar los términos de la licencia, cerrar el acceso a los datos, cambiar los formatos o la estructura de la API: todo esto puede bloquear procesos críticos. De repente, resulta que la actualización de una tabla requiere rehacer todo un bloque de integraciones y conectores (Fig. 3.2-13), y cualquier actualización a gran escala del software o de su proveedor de API se convierte en

una amenaza potencial para la estabilidad de todo el sistema de la empresa.

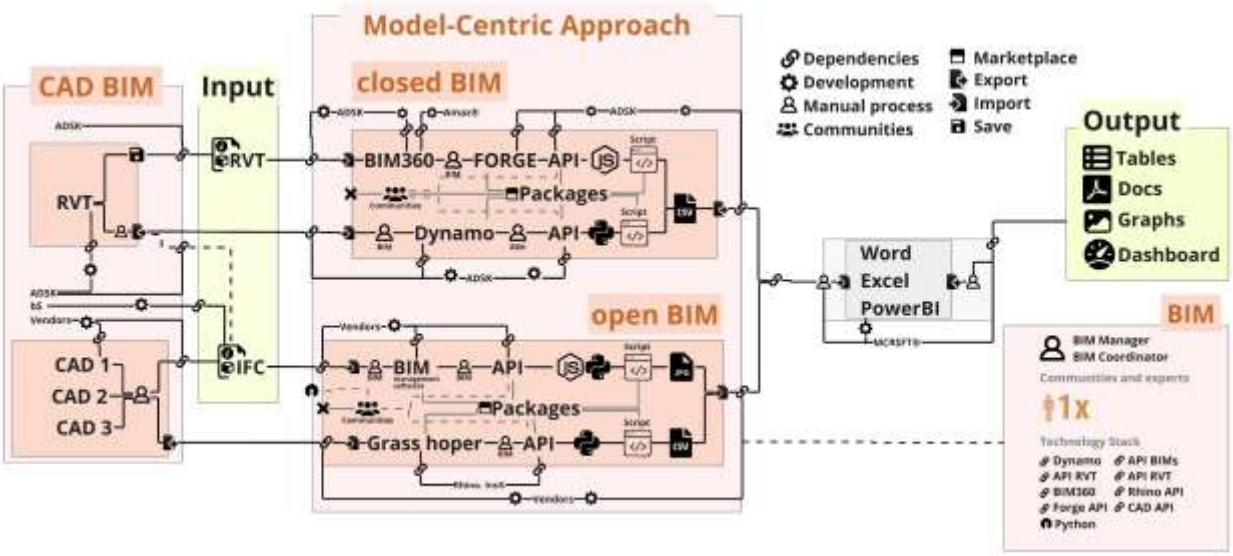


Fig. 3.2-13 Un ejemplo del gran número de dependencias en el procesamiento CAD -data crea barreras a la integración de datos en el ecosistema de las empresas de construcción.

> En tales condiciones, los desarrolladores y arquitectos de sistemas se ven obligados a trabajar no para anticiparse, sino para sobrevivir. En lugar de implantar nuevas soluciones, se adaptan. En lugar de desarrollar, intentan mantener la compatibilidad. En lugar de automatizar y acelerar los procesos, dedican su tiempo a estudiar las próximas interfaces cerradas, la documentación de la API y la interminable reconstrucción del código.

Trabajar con formatos y sistemas cerrados no es solo un reto técnico: es una limitación estratégica. A pesar de las evidentes oportunidades que ofrecen la automatización moderna, la IA, el LLM y el análisis predictivo, muchas empresas no aprovechan todo su potencial. Y las barreras erigidas por los formatos propietarios (Fig, 3.2-13) niegan a las empresas el acceso a sus propios datos. Esta es quizá la ironía de la transformación digital en la construcción.

> La transparencia de los datos y los sistemas abiertos no son un lujo, sino un requisito previo para la velocidad y la eficiencia. Sin apertura, los procesos empresariales se llenan de burocracia innecesaria, cadenas de aprobación de varios niveles y una dependencia cada vez mayor del principio HiPPO: tomar decisiones basándose en la opinión de la persona mejor pagada.

Sin embargo, se vislumbra un cambio de paradigma en el horizonte. A pesar del predominio de las soluciones propietarias, cada vez más empresas se dan cuenta de las limitaciones de las arquitecturas inspiradas en la Cuarta Revolución Industrial. En la actualidad, el vector se está desplazando hacia los principios de la Quinta Revolución, en la que los datos como activo estratégico, las interfaces abiertas (API) y la verdadera interoperabilidad entre sistemas ocupan un lugar central.

Esta transición marca el paso de ecosistemas cerrados a arquitecturas digitales flexibles y modulares en

las que los formatos abiertos, las normas y el intercambio transparente de datos desempeñan un papel fundamental.

Los formatos abiertos están cambiando el enfoque de la digitalización

El sector de la construcción fue uno de los últimos en abordar el problema de los datos cerrados y patentados. A diferencia de otros sectores de la economía, aquí la digitalización ha tardado en desarrollarse. Las razones son, entre otras, el tradicional carácter conservador del sector, la prevalencia de soluciones locales dispares y el arraigo de la gestión de documentos en papel. Durante décadas, los procesos clave de la construcción se basaron en planos físicos, llamadas telefónicas y bases de datos no sincronizadas. En este contexto, los formatos cerrados se han percibido durante mucho tiempo como la norma más que como un obstáculo.

La experiencia de otras industrias demuestra que eliminar las barreras a los datos cerrados conduce a un aumento de la innovación, un desarrollo acelerado y una mayor competencia [73]. En ciencia, el intercambio de datos abiertos permite acelerar los descubrimientos y desarrollar la cooperación internacional. En medicina, puede mejorar la eficacia del diagnóstico y el tratamiento. En ingeniería de software, para crear ecosistemas de cocreación y mejora rápida de productos.

> Según el informe de McKinsey "Open Data: Unlock Innovation and Productivity with Information Flow" 2013. [74], los datos abiertos tienen el potencial de desbloquear entre 3 y 5 billones de dólares anuales en siete sectores clave, como la construcción, el transporte, la sanidad y la energía. Según el mismo estudio, los ecosistemas de datos descentralizados permiten a las grandes constructoras y contratistas reducir los costes de desarrollo y mantenimiento de software, acelerando la adopción digital.

La transición a las arquitecturas abiertas, iniciada hace tiempo en otros sectores de la economía, se está extendiendo poco a poco a la industria de la construcción. Las grandes empresas y los clientes públicos, y especialmente las organizaciones financieras que controlan las inversiones en proyectos de construcción, exigen cada vez más el uso de datos abiertos y el acceso al código fuente de cálculos, cálculos y aplicaciones. Ya no sólo se espera que los desarrolladores creen soluciones digitales y muestren las cifras finales de un proyecto, sino que sean transparentes, reproducibles e independientes de terceros proveedores de aplicaciones.

> El uso de soluciones de código abierto proporciona al cliente la confianza de que, aunque los desarrolladores externos dejen de colaborar o abandonen el proyecto, ello no afectará a la capacidad de seguir desarrollando herramientas y sistemas. Una de las principales ventajas de los datos abiertos es su capacidad para eliminar la dependencia de los desarrolladores de aplicaciones de plataformas específicas para acceder a los datos.

Si una empresa no puede abandonar por completo las soluciones propietarias, un posible compromiso es el uso de técnicas de ingeniería inversa. Estos métodos legales y técnicamente sólidos permiten convertir formatos cerrados en otros más accesibles, estructurados y aptos para la integración. Esto es especialmente importante cuando hay que conectar con sistemas heredados o migrar información de un entorno

de software a otro.

> Uno de los ejemplos más brillantes en la historia de la transición a formatos abiertos y la aplicación de la ingeniería inversa (pirateo legal de sistemas propietarios) en la construcción es la historia de la lucha por la apertura del formato DWG, muy utilizado en los sistemas de diseño asistido por ordenador (CAD). En 1998, en respuesta al monopolio de un proveedor de software, los otros 15 proveedores de CAD formaron una nueva alianza denominada "Open DWG" para ofrecer a los desarrolladores herramientas libres e independientes que les permitieran trabajar con el formato DWG (el estándar de facto para la transferencia de dibujos) sin necesidad de software propietario ni API cerradas. Este acontecimiento supuso un punto de inflexión que permitió a decenas de miles de empresas acceder libremente al formato cerrado de una solución CAD popular desde finales de los 80 hasta hoy y crear soluciones compatibles que fomentaron la competencia en el mercado CAD [75]. Hoy en día, el SDK "Open DWG", que se creó por primera vez en 1996, se utiliza en casi todas las soluciones en las que es posible importar, editar y exportar el formato DWG, al margen de la aplicación oficial del desarrollador del formato DWG.

Otros gigantes tecnológicos están forzando transformaciones similares. Microsoft, antaño símbolo del enfoque propietario, abrió el código fuente de.NET Framework, empezó a utilizar Linux en la infraestructura de servicios en la nube Azure y adquirió GitHub para reforzar su posición en la comunidad de código abierto. [76]. Meta (antes Facebook) ha publicado modelos de IA de código abierto , como la serie Llama, para fomentar la innovación y la colaboración en el desarrollo de agentes de IA. Su consejero delegado, Mark Zuckerberg, prevé que las plataformas de código abierto liderarán los avances tecnológicos en la próxima década [77].

> Open Source es un modelo de desarrollo y distribución de software en el que el código fuente está abierto para su libre uso, estudio, modificación y distribución.

Los datos abiertos y las soluciones de código abierto se están convirtiendo no sólo en una tendencia, sino en la base de la sostenibilidad digital. Ofrecen a las empresas flexibilidad, capacidad de recuperación, control sobre sus propias decisiones y la posibilidad de ampliar los procesos digitales sin depender de las políticas de los proveedores. Y, lo que es igual de importante, devuelven a las empresas el control sobre el recurso más valioso del siglo XXI: sus datos.

Cambio de paradigma: el código abierto como fin de la era del dominio de los proveedores de software

El sector de la construcción está experimentando un cambio que no puede monetizarse de la forma habitual. El concepto de enfoque basado en los datos, centrado en ellos, y el uso de herramientas de Open Source están llevando a replantearse las reglas del juego en las que se apoyan los gigantes del software del mercado.

> A diferencia de anteriores transformaciones tecnológicas, esta transición no será promovida activamente por los proveedores. El cambio de paradigma amenaza sus modelos de negocio tradicionales basados en licencias, suscripciones y consultoría. La nueva realidad no consiste en un producto listo para usar o una suscripción de pago, sino que requiere un reajuste de los procesos y la mentalidad.

Para gestionar y desarrollar soluciones de centros de datos basadas en tecnologías abiertas, las empresas tendrán que replantearse los procesos internos. Los especialistas de distintos departamentos no solo tendrán que colaborar, sino también replantearse cómo trabajan juntos.

El nuevo paradigma implica el uso de datos abiertos y soluciones de código abierto, en las que las herramientas basadas en inteligencia artificial y grandes modelos de lenguaje (LLM), más que los programadores, desempeñarán un papel especial en la creación de código de software. Ya a mediados de 2024, más del 25% del nuevo código en Google se crea con IA [78]. En el futuro, la codificación con LLM hará el 80% del trabajo en sólo el 20% del tiempo (Fig, 3.2-14).

> Según el estudio 2020 de McKinsey [79], las GPU están sustituyendo cada vez más a las CPU en el análisis debido a su alto rendimiento y al soporte que ofrecen las modernas herramientas de código abierto. Esto permite a las empresas acelerar el procesamiento de datos sin realizar inversiones significativas en software caro ni contratar a los escasos especialistas.

Destacadas consultoras como McKinsey, PwC y Deloitte subrayan la creciente importancia de los estándares abiertos y las aplicaciones de código abierto en todos los sectores.

> Según el informe de PwC Open Source Monitor 2019 [80], el 69% de las empresas con 100 o más empleados utilizan conscientemente soluciones de código abierto. El OSS se utiliza de forma especialmente activa en las grandes empresas: 71% de las empresas con 200-499 empleados, 78% en la categoría de 500-1999 empleados, y hasta 86% entre las empresas con más de 2000 empleados. Según el informe OSSRA 2023 de Synopsys, el 96% de las bases de código analizadas contenían componentes de código abierto [81].

En el futuro, el papel del desarrollador no consistirá en escribir código manualmente, sino en diseñar modelos de datos, arquitecturas de flujo y gestionar agentes de IA que creen los cálculos adecuados a demanda. Las interfaces de usuario se volverán minimalistas y la interacción se basará en el diálogo. La programación clásica dará paso al diseño de alto nivel y a la orquestación de soluciones digitales (Fig, 3.2-14). Las tendencias actuales, como las plataformas de bajo código (Fig, 7.4-6) y los ecosistemas habilitados para LLM (Fig, 7.4-4), reducirán significativamente el coste de desarrollo y mantenimiento de los sistemas informáticos.

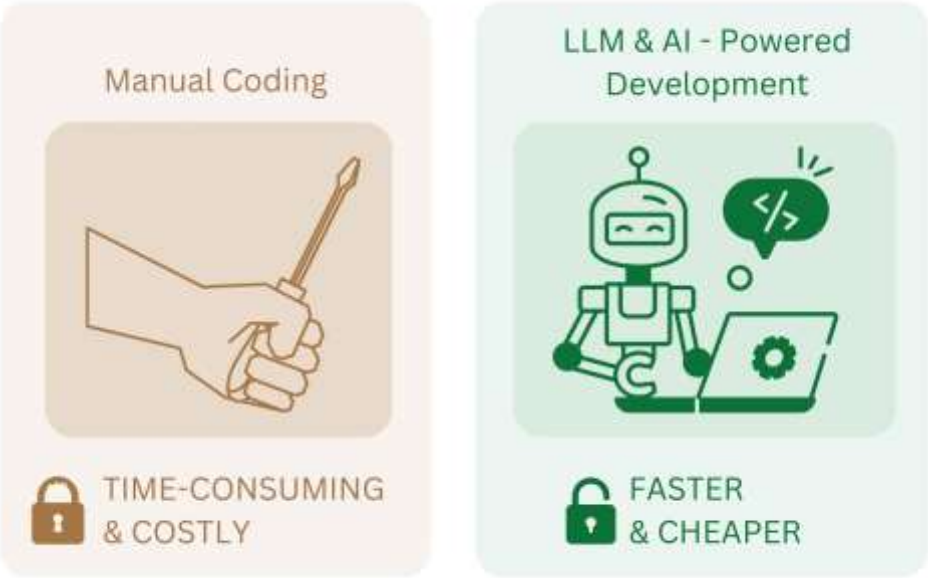


Fig. 3.2-14 Aunque hoy en día las aplicaciones son creadas manualmente por programadores, en el futuro una parte importante del código será generado por soluciones basadas en IA y LLM.

Esta transición será distinta de las anteriores y es poco probable que los grandes proveedores de software la catalicen.

> En el estudio de la Harvard Business School "The Value of Open Source Software" 2024 [40], se estima el valor total del software de código abierto desde dos puntos de vista. Por un lado, si calculamos cuánto costaría construir desde cero todas las soluciones de código abierto existentes, la cantidad ascendería a unos 4.150 millones de dólares. Por otro lado, si imaginamos que cada empresa desarrolla por su cuenta sus propios análogos de soluciones de código abierto (lo que ocurre en todas partes), sin tener acceso a las herramientas existentes, entonces el coste total de las empresas alcanzaría la friolera de 8,8 billones de dólares: es el coste de la demanda.

> No es difícil adivinar que ningún gran proveedor de software está interesado en reducir un mercado de software con un valor potencial de 8,8 billones de dólares a sólo 4.150 millones. Esto significaría reducir el volumen de la demanda en más de 2.000 veces. Semejante transformación simplemente no es rentable para los proveedores cuyos modelos de negocio se basan en años de mantener la dependencia de los clientes de soluciones cerradas. Así que las empresas que esperan que alguien les ofrezca una solución llave en mano cómoda y abierta pueden llevarse una decepción: esos vendedores simplemente no aparecerán.

El cambio a una arquitectura digital abierta no significa pérdidas de empleo o ingresos. Al contrario, crea las condiciones para modelos de negocio flexibles y adaptables que pueden acabar desplazando al mercado tradicional de licencias y software en caja.

En lugar de vender licencias, servicios; en lugar de formatos cerrados, plataformas abiertas; en lugar de

depender de un proveedor, independencia y capacidad de crear soluciones para necesidades reales. Los que antes se limitaban a utilizar herramientas podrán convertirse en sus coautores. Y quienes sepan trabajar con datos, modelos, escenarios y lógica se encontrarán en el centro de la nueva economía digital de la industria. Hablaremos más de estos cambios y de los nuevos roles, modelos de negocio y formatos de cooperación que están surgiendo en torno a los datos abiertos en la última y décima parte del libro.

> Las soluciones basadas en datos y código abiertos permitirán a las empresas centrarse en la eficiencia de los procesos empresariales en lugar de luchar con API obsoletas e integrar sistemas cerrados. Una transición consciente a la arquitectura abierta puede mejorar significativamente la productividad y reducir la dependencia de los proveedores.

La transición a una nueva realidad no sólo supone un cambio en los planteamientos del desarrollo de software, sino también un replanteamiento del principio mismo de trabajar con datos. En el centro de esta transformación no está el código, sino la información: su estructura, accesibilidad e interpretabilidad. Aquí es donde los datos abiertos y estructurados pasan a primer plano, convirtiéndose en parte integrante de la nueva arquitectura digital.

Datos abiertos estructurados: la base de la transformación digital

Mientras que en décadas pasadas la sostenibilidad empresarial venía determinada en gran medida por la elección de soluciones de software y la dependencia de proveedores específicos, en la economía digital actual el factor clave es la calidad de los datos y la capacidad de trabajar con ellos de forma eficaz. El código fuente abierto es una parte importante del nuevo paradigma tecnológico, pero su potencial sólo se despliega realmente cuando los datos son comprensibles, organizados y legibles por máquina. Entre todos los tipos de modelos de datos, los datos abiertos estructurados se están convirtiendo en la piedra angular de la transformación digital sostenible.

La principal ventaja de los datos abiertos estructurados es su interpretación inequívoca y la posibilidad de procesamiento automatizado. Esto permite aumentar considerablemente la eficiencia tanto a nivel de operaciones individuales como en toda la organización.

> Según el informe de Deloitte "El proceso de transferencia de datos en la transformación empresarial" [82], es fundamental trabajar con TI para gestionar la transferencia de datos estructurados. Según el informe del Gobierno británico "Data Analytics and AI in Government Project Delivery" (2024) [83], la eliminación de barreras al intercambio de datos entre distintos proyectos y organizaciones es clave para mejorar la eficiencia en la gestión de proyectos. El documento subraya que la normalización de los formatos de datos y la introducción de principios de datos abiertos pueden evitar la duplicación de información, minimizar la pérdida de tiempo y mejorar la precisión de las previsiones.

Para el sector de la construcción, en el que tradicionalmente prevalece un alto grado de fragmentación y diversidad de formatos, el proceso de unificación estructurado y los datos abiertos estructurados desempeñan un papel crucial en la configuración de procesos coherentes y manejables (Fig. 4.1-14). Permiten a los participantes en el proyecto centrarse en mejorar la productividad en lugar de en resolver problemas

técnicos relacionados con incompatibilidades entre plataformas cerradas, modelos de datos y formatos.

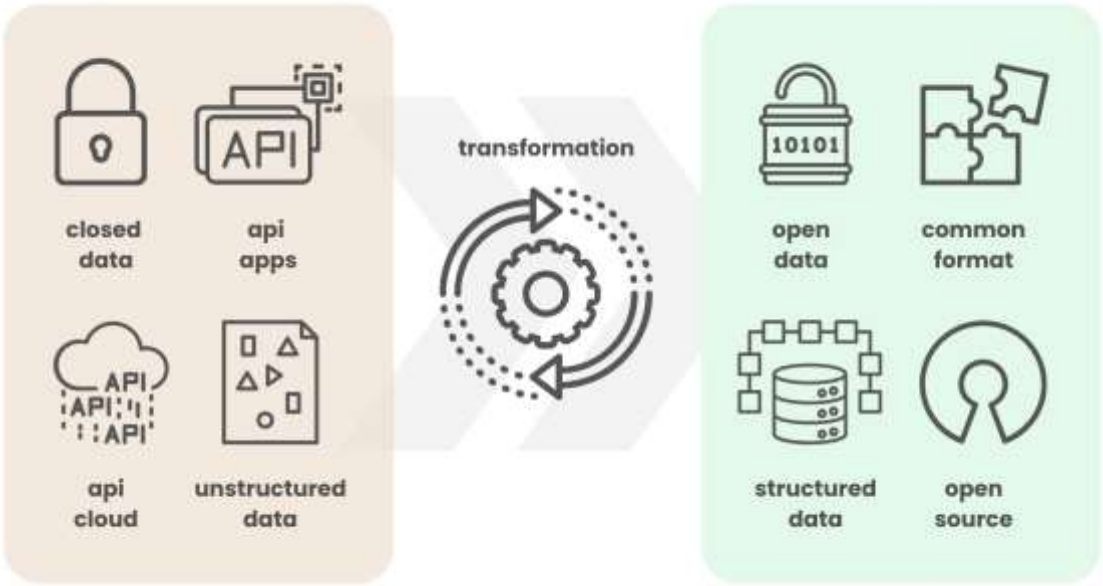


Fig. 3.2-15 Los datos estructurados abiertos reducen la dependencia de soluciones y plataformas de software y aceleran la innovación.

Las herramientas tecnológicas modernas, de las que hablaremos en detalle más adelante en el libro, permiten no sólo recopilar información, sino también limpiarla automáticamente: eliminar duplicidades, corregir errores y normalizar valores. Esto significa que analistas e ingenieros no trabajan con documentos dispares, sino con una base de conocimientos organizada y apta para el análisis, la automatización y la toma de decisiones.

> *Hazlo tan sencillo como puedas, pero no más.*
>
> - Albert Einstein, físico teórico (la autoría de la cita es discutida [84])

Hoy en día, la mayoría de las interfaces de usuario para trabajar con datos pueden crearse automáticamente, sin necesidad de escribir manualmente código para cada caso de negocio. Para ello se requiere una capa de infraestructura que comprenda la estructura, el modelo y la lógica de los datos sin instrucciones adicionales (Fig, 4.1-15). Son los datos estructurados los que hacen posible este enfoque: los formularios, tablas, filtros y vistas pueden generarse automáticamente con un mínimo esfuerzo de programación.

Las interfaces más importantes para el usuario pueden seguir requiriendo un perfeccionamiento manual. Pero en la mayoría de los casos -y se trata de entre el 50 y el 90 por ciento de los escenarios de trabajo- basta con la generación automática de aplicaciones y cálculos sin necesidad de utilizar aplicaciones especiales para este fin (Fig. 3.2-16), lo que reduce considerablemente los costes de desarrollo y mantenimiento, disminuye los errores y acelera la implantación de soluciones digitales.

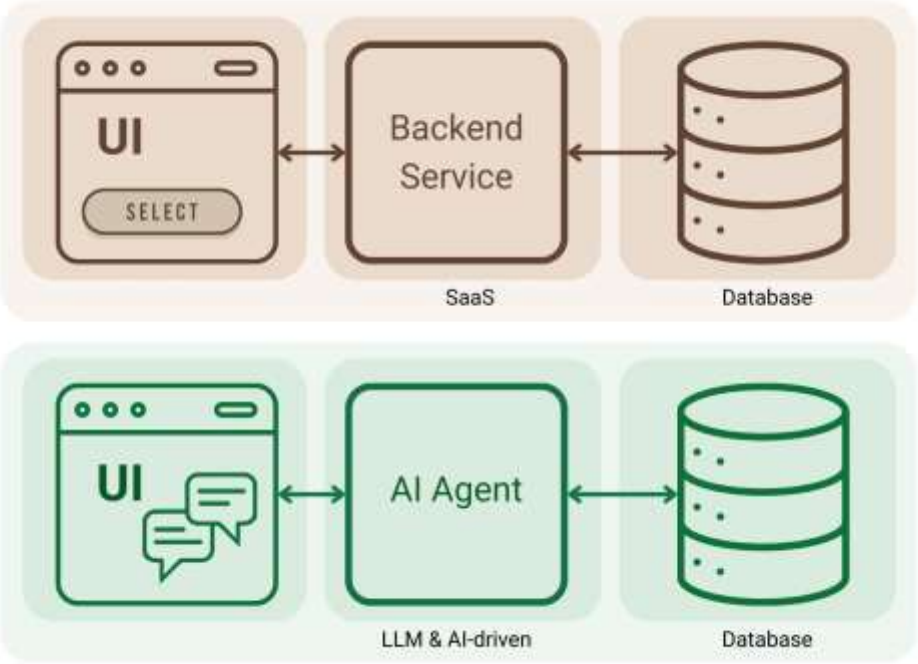


Fig. 3.2-16 Modelos arquitectónicos para trabajar con datos: arquitectura de aplicación tradicional y modelo orientado a la IA con LLM.

El paso de arquitecturas basadas en aplicaciones individuales a sistemas gestionados de forma inteligente y basados en modelos lingüísticos (LLM) es el siguiente paso en la evolución digital. En una arquitectura de este tipo, los datos estructurados se convierten no solo en un objeto de almacenamiento, sino también en la base de la interacción con herramientas de IA capaces de analizar, interpretar y recomendar acciones basadas en el contexto.

En los siguientes capítulos, veremos ejemplos reales de implementación de una arquitectura basada en datos estructurados abiertos y mostraremos cómo se aplican los modelos de lenguaje para interpretar, validar y procesar datos automáticamente. Estos casos prácticos le ayudarán a comprender mejor cómo funciona la nueva lógica digital en acción - y qué beneficios aporta a las empresas que están preparadas para la transformación.

CAPÍTULO 3.3.
LLM Y SU PAPEL EN EL TRATAMIENTO DE DATOS Y LOS PROCESOS EMPRESARIALES

LLM salas de chat: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok para automatizar los procesos de datos.

La aparición de los grandes modelos lingüísticos (LLM) fue una extensión natural del movimiento hacia los datos abiertos estructurados y la filosofía del código abierto. Cuando los datos se vuelven organizados, accesibles y legibles por máquina, el siguiente paso es una herramienta que pueda interactuar con esta información sin necesidad de escribir código complejo o poseer conocimientos técnicos especializados.

> Los LLM son un producto directo de la apertura: grandes conjuntos de datos abiertos, publicaciones y el movimiento Open Source. Sin artículos académicos abiertos, datos textuales a disposición del público y una cultura de desarrollo colaborativo, no habría ChatGPT ni otros LLM. El LLM es, en cierto sentido, un "destilado" del conocimiento digital acumulado de la humanidad, reunido y educado a través de los principios de la apertura.

Los grandes modelos lingüísticos modernos (LLM - Large Language Models) como ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ ofrecen a los usuarios la posibilidad de formular consultas a los datos en lenguaje natural. Esto hace que el trabajo con la información sea accesible no sólo a los desarrolladores, sino también a analistas, ingenieros, planificadores, gestores y otros profesionales antes alejados de la programación

> **LLM** (Large Language Model) es una inteligencia artificial entrenada para comprender y generar textos a partir de ingentes cantidades de datos recogidos en todo Internet. Es capaz de analizar contextos, responder preguntas, dialogar, escribir textos y generar código de software.

Si antes la visualización, el tratamiento o el análisis de datos requerían el conocimiento de un lenguaje de programación especial: Python, SQL, R o Scala, así como la capacidad de trabajar con bibliotecas como Pandas, Polars o DuckDB y muchas otras, a partir de 2023 la situación ha cambiado radicalmente. Ahora el usuario puede simplemente describir lo que quiere obtener, y el propio modelo generará el código, lo ejecutará, mostrará una tabla o un gráfico y explicará el resultado. Por primera vez en décadas, el desarrollo de la tecnología no ha tomado el camino de la complicación, sino el de la simplificación radical y la accesibilidad.

Este principio - "procesar datos con palabras (prompts)". - marcó una nueva etapa en la evolución del trabajo con la información, llevando de hecho la creación de soluciones a un nivel de abstracción aún mayor. Al igual que antes ya no era necesario que los usuarios entendieran los fundamentos técnicos de Internet para gestionar tiendas online o crear sitios web con WordPress, Joomla y otros sistemas modulares de

código abierto (av tor el libro lleva trabajando con este tipo de sistemas desde 2005, incluidas plataformas online educativas y de ingeniería). - Esto, a su vez, ha provocado un auge de los contenidos digitales y los negocios en línea: hoy en día, ingenieros, analistas y gestores pueden automatizar flujos de trabajo sin necesidad de conocer lenguajes de programación. Esto se ve facilitado por potentes LLM -tanto gratuitos como de código abierto, como LLaMA, Mistral, Qwen, DeepSeek y otros- que ponen las tecnologías avanzadas al alcance del público más amplio posible.

Grandes modelos lingüísticos LLM: cómo funciona

Los grandes modelos lingüísticos (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) son redes neuronales entrenadas con enormes cantidades de datos textuales de Internet, libros, artículos y otras fuentes. Su principal tarea es comprender el contexto del habla humana y generar respuestas significativas.

El LLM moderno se basa en la arquitectura Transformer propuesta por investigadores de Google en 2017 [85]. El componente clave de esta arquitectura es el mecanismo de atención, que permite al modelo considerar las relaciones entre palabras independientemente de su posición en el texto.

> El proceso de aprendizaje del LLM es remotamente similar a la forma en que los humanos aprenden un idioma, sólo que millones de veces mayor. El modelo analiza miles de millones de ejemplos de palabras y expresiones, identificando patrones en la estructura del lenguaje y en la lógica de las transiciones semánticas. Todo el texto se descompone en tokens, unidades semánticas mínimas (palabras o sus partes), que luego se transforman en vectores en un espacio multidimensional (Fig. 8.2-2). Estas representaciones vectoriales permiten a la máquina "entender" las relaciones ocultas entre conceptos, en lugar de limitarse a operar sobre el texto como una secuencia de símbolos.

Los Big Language Models no son sólo herramientas para generar texto. Son capaces de reconocer significados, encontrar conexiones entre conceptos y trabajar con datos, aunque se presenten en distintos formatos. Lo principal es que la información se descomponga en modelos comprensibles y se represente como tokens con los que el LLM pueda trabajar.

El mismo planteamiento puede aplicarse a los proyectos de construcción. Si pensamos en un proyecto como una especie de texto en el que cada edificio, elemento o construcción es un token, podemos empezar a procesar esa información de forma similar. Los proyectos de construcción pueden compararse con libros organizados en categorías, capítulos y grupos de párrafos formados por tokens mínimos: los elementos de un proyecto de construcción (Fig, 3.3-1). Al traducir los modelos de datos a un formato estructurado, también podemos traducir los datos estructurados a bases vectoriales (Fig. 8.2-2), que son una fuente ideal para el aprendizaje automático y tecnologías como LLM.

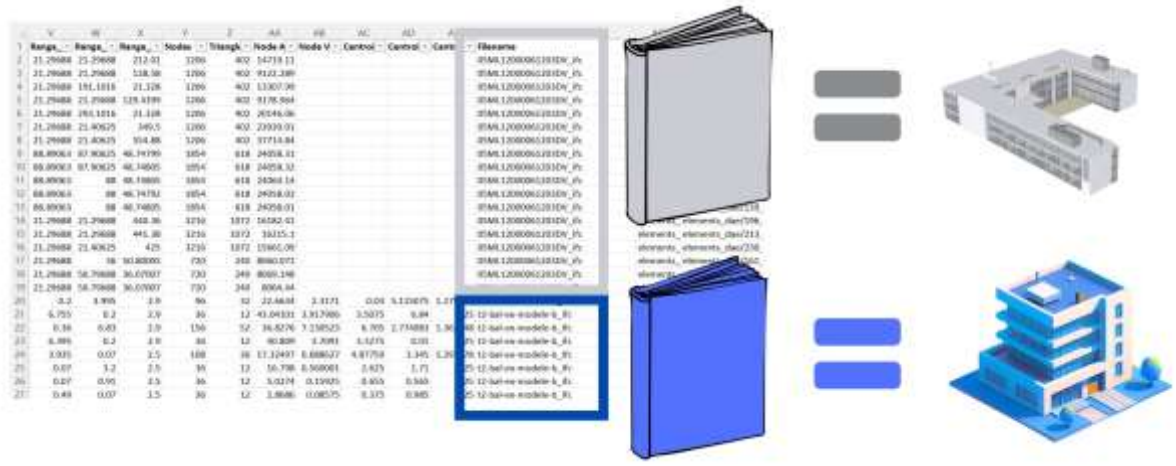


Fig. 3.3-1 Un elemento de un proyecto de construcción es como una ficha en un texto: una unidad mínima a partir de la cual se forman grupos (párrafos) de secciones (categorías) de todo el proyecto.

Si un proyecto de construcción se digitaliza y sus elementos se representan como tokens o vectores, resulta posible acceder a ellos en lenguaje natural en lugar de mediante rígidas consultas formales. Aquí es donde entra en juego una de las principales ventajas del LLM: la capacidad de comprender el significado de una consulta y relacionarla con los datos pertinentes.

El ingeniero ya no tiene que escribir la consulta SQL o el código Python para obtener los datos requeridos; simplemente, entendiendo el LLM y la estructura de datos, puede formular la tarea de la forma habitual: "*Encontrar todas las estructuras de hormigón armado con clase de hormigón superior a B30 y calcular su volumen total*". El modelo reconocerá el significado de la consulta, la convertirá en una forma legible por la máquina, encontrará los datos (agrupará y transformará) y devolverá el resultado final.

Los documentos, tablas y modelos de proyectos se convierten en representaciones vectoriales (incrustación) y se almacenan en la base de datos. Cuando un usuario formula una pregunta, la consulta también se convierte en un vector y el sistema encuentra los datos más relevantes. Esto permite al LLM basarse no sólo en su propio conocimiento entrenado, sino también en datos corporativos reales, aunque hayan aparecido después de finalizar el entrenamiento del modelo.

> Una de las ventajas más importantes del LLM en la construcción es la capacidad de generar código de programa. En lugar de pasar la tarea técnica a un programador, los especialistas pueden describir la tarea en lenguaje natural, y el modelo creará el código necesario, que puede utilizarse (copiándolo del chat) en la creación de código de automatización de procesos. Los modelos de LLM permiten a especialistas sin grandes conocimientos de programación contribuir a la automatización y mejora de los procesos de negocio de la empresa.

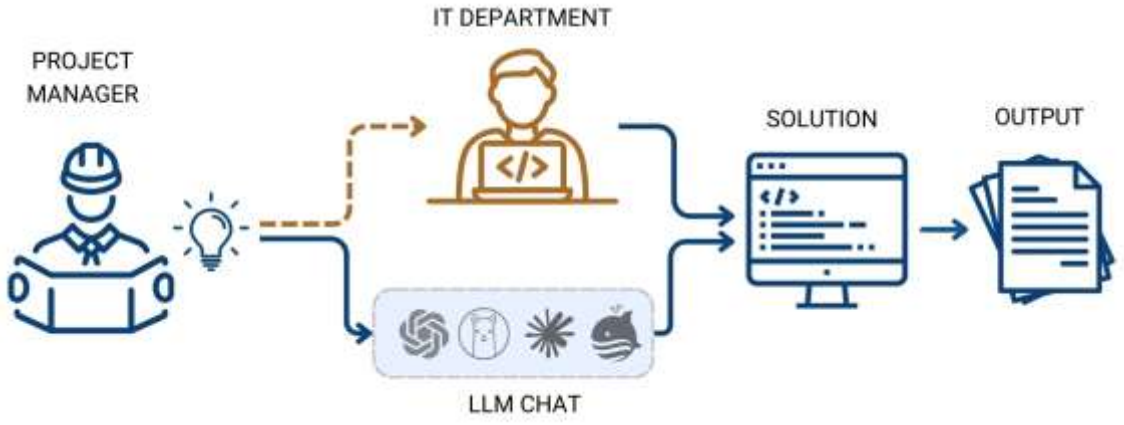


Fig. 3.3-2 LLMs ofrecen a los usuarios la posibilidad de escribir código y obtener resultados sin necesidad de conocimientos de programación.

> Según un estudio realizado por Wakefield Research y patrocinado por SAP en 2024 [36], en el que se encuestó a 300 altos ejecutivos de empresas con ingresos anuales de al menos 1.000 millones de dólares en EE.UU.: el 52% de los altos ejecutivos confía en la IA para analizar datos y ofrecer recomendaciones para la toma de decisiones. Otro 48% utiliza la IA para identificar riesgos que antes no se tenían en cuenta, y un 47% para sugerir planes alternativos. Además, el 40% utiliza la IA para el desarrollo de nuevos productos, la planificación de presupuestos y la investigación de mercados. El estudio también mostró el impacto positivo de la IA en la vida personal, con un 39% de los encuestados que declararon haber mejorado el equilibrio entre la vida laboral y personal, un 38% que declararon haber mejorado su salud mental y un 31% que declararon haber reducido sus niveles de estrés.

Sin embargo, a pesar de toda su potencia, los LLM siguen siendo una herramienta que es importante utilizar conscientemente. Como cualquier tecnología, tienen limitaciones. Uno de los problemas más conocidos son las llamadas "alucinaciones", casos en los que el modelo produce con confianza una respuesta plausible pero incorrecta desde el punto de vista de los hechos. Por eso es fundamental entender cómo funciona el modelo: qué datos y modelos de datos puede interpretar sin errores, cómo interpreta las consultas y de dónde obtiene la información. También conviene recordar que los conocimientos del LLM se limitan a la fecha de su formación y, sin conexión con datos externos, es posible que el modelo no tenga en cuenta las normas, estándares, precios o tecnologías actuales.

La solución a estos problemas pasa por actualizar periódicamente las bases de datos de vectores, conectarse a las fuentes pertinentes y desarrollar IA autónoma -agentes que no se limiten a responder preguntas, sino que utilicen proactivamente los datos para formarse, gestionar tareas, identificar riesgos, sugerir opciones de optimización y supervisar el rendimiento de los proyectos.

> La transición a LLM -interfaces en la construcción no es sólo una novedad tecnológica. Es un cambio de paradigma que elimina las barreras entre las personas y los datos. Es la capacidad de trabajar con la información con la misma facilidad con la que hablamos entre nosotros y, aun así, obtener resultados precisos, verificados y procesables.

Las empresas que empiecen a utilizar estas herramientas antes que las demás obtendrán una importante ventaja competitiva. Esto incluye acelerar el trabajo, reducir costes y mejorar la calidad de las soluciones de diseño mediante un acceso rápido al análisis de datos y la capacidad de encontrar rápidamente respuestas a preguntas complejas. Pero también hay que tener en cuenta cuestiones de seguridad. El uso de servicios LLM basados en la nube puede conllevar riesgos de fuga de datos. Por ello, las organizaciones buscan cada vez más soluciones alternativas que les permitan desplegar herramientas LLM en su propia infraestructura, localmente, con total protección y control de la información.

Uso de LLM locales para datos sensibles de la empresa

La aparición de los primeros chat-LLM en 2022 marcó una nueva etapa en el desarrollo de la inteligencia artificial. Sin embargo, inmediatamente después de la adopción generalizada de estos modelos, surgió una pregunta legítima: ¿hasta qué punto es seguro transferir a la nube datos y consultas relacionados con la empresa? La mayoría de los modelos lingüísticos basados en la nube almacenaban el historial de comunicaciones y los documentos cargados en sus servidores y, para las empresas que manejan información sensible, esto suponía un serio obstáculo para la adopción de la IA.

> Una de las soluciones más sostenibles y lógicas a este problema ha sido el despliegue de Open Source LLM localmente, dentro de la infraestructura informática corporativa. A diferencia de los servicios en la nube, los modelos locales funcionan sin conexión a Internet, no transfieren datos a servidores externos y ofrecen a las empresas un control total sobre la información

> *El mejor modelo abierto [Open Source LLM] es actualmente comparable en rendimiento a los modelos cerrados [como ChatGPT, Claude], pero con un retraso de aproximadamente un año [77].*
>
> - Ben Cottier, investigador principal de Epoch AI, organización de investigación sin ánimo de lucro, 2024

Las principales empresas tecnológicas han empezado a poner sus LLM a disposición del uso local. La serie LLaMA de código abierto de Meta y el proyecto DeepSeek de China, en rápido crecimiento, son ejemplos del paso a la arquitectura abierta. Junto a ellos, Mistral y Falcon también han lanzado potentes modelos libres de las restricciones de las plataformas propietarias. Estas iniciativas no sólo han acelerado el desarrollo de la IA mundial, sino que también han ofrecido a las empresas preocupadas por la privacidad alternativas reales de independencia, flexibilidad y cumplimiento de las normas de seguridad.

En un entorno empresarial, especialmente en el sector de la construcción, la protección de datos no es sólo una cuestión de comodidad, sino de cumplimiento de la normativa. Trabajar con documentos de licitación, presupuestos, planos y correspondencia confidencial requiere controles estrictos. Y aquí es donde

el LLM local ofrece la garantía necesaria de que los datos permanecen dentro del perímetro de la empresa.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Fig. 3.3-3 Los modelos locales proporcionan un control y una seguridad completos, mientras que las soluciones basadas en la nube ofrecen una integración sencilla y actualizaciones automáticas.

Principales ventajas de la fuente abierta local LLM:

- Control total de los datos. Toda la información permanece dentro de la empresa, lo que elimina el acceso no autorizado y la fuga de datos.
- Funcionamiento sin conexión. Sin dependencia de la conexión a Internet, lo que es especialmente importante para infraestructuras informáticas aisladas. Esto también garantiza un funcionamiento ininterrumpido ante sanciones o servicios en la nube bloqueados.
- Flexibilidad de aplicación. El modelo puede utilizarse para la generación de textos, el análisis de

datos, la escritura de código de programa, el apoyo al diseño y la gestión de procesos empresariales.

- Adaptación a los objetivos de la empresa. El LLM puede formarse a partir de documentos internos, lo que permite tener en cuenta las especificidades del trabajo de la empresa y las características de su sector. El LLM local puede conectarse a plataformas CRM, ERP o BI, lo que permite automatizar el análisis de consultas de clientes, la creación de informes o incluso la previsión de tendencias.

> Desplegar el modelo gratuito y de código abierto de DeepSeek -R1-7B en un servidor, para acceso de todo un equipo de usuarios, a un coste de 1000 dólares al mes puede costar potencialmente menos que las cuotas anuales de las API en la nube, como ChatGPT o Claude y permite a las empresas tomar el control total de sus datos, elimina su transferencia a internet y ayuda a cumplir con requisitos normativos como el GDPR

En otros sectores, los LLM locales ya están cambiando su enfoque de la automatización. En los servicios de asistencia, responden a las frecuentes consultas de los clientes, reduciendo la carga de trabajo de los operadores. En los departamentos de RRHH, analizan los CV y seleccionan a los candidatos pertinentes. En el comercio electrónico, generan ofertas personalizadas sin revelar los datos del usuario.

Se espera un efecto similar en el sector de la construcción. Gracias a la integración del LLM con los datos y normas del proyecto, es posible acelerar la preparación de la documentación, automatizar la elaboración de presupuestos y realizar análisis predictivos de costes. El uso de LLM junto con tablas estructuradas y marcos de datos se está convirtiendo en un área especialmente prometedora.

Control total de AI en la empresa y cómo desplegar su propio LLM

Las herramientas modernas permiten a las empresas desplegar un gran modelo lingüístico (LLM) localmente en tan solo unas horas. Esto proporciona un control total sobre los datos y la infraestructura, eliminando la dependencia de servicios externos en la nube y minimizando el riesgo de fuga de información. Esta solución es especialmente relevante para las organizaciones que trabajan con documentación sensible de proyectos o datos sensibles desde el punto de vista comercial.

En función de las tareas y los recursos, existen diferentes escenarios de implantación, desde soluciones listas para usar hasta arquitecturas más flexibles y escalables. Una de las herramientas más sencillas es Ollama, que permite ejecutar modelos lingüísticos literalmente en un clic, sin necesidad de profundos conocimientos técnicos. Un comienzo rápido con Ollama:

1. Descargue la distribución para su sistema operativo (Windows / Linux / macOS) desde el sitio web oficial: ollama.com
2. Instale el modelo a través de la línea de comandos. Por ejemplo, para el modelo *Mistral*:

```
ollama run mistral
```

3. Después de ejecutar el modelo, está listo para trabajar: puedes enviar consultas de texto a través del terminal o integrarlo en otras herramientas. Ejecute el modelo y ejecute una consulta:

```
ollama run mistral "¿Cómo crear un cálculo con todos los recursos de la obra para
instalar un tabique de pladur de 100 mm de ancho?".
```

Para quienes prefieran trabajar en un entorno visual familiar, existe LM Studio, una aplicación gratuita con una interfaz que recuerda a ChatGPT

- Instale LM Studio descargando el kit de distribución del sitio web oficial - lmstudio.ai
- A través del catálogo incorporado, seleccione un modelo (por ejemplo, Falcon o GPT-Neo-X) y descárguelo
- Trabaje con el modelo a través de una interfaz intuitiva que recuerda a ChatGPT, pero completamente localizada.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Fig. 3.3-4 Comparación de los modelos LLM locales de código abierto más conocidos -models.

La elección del modelo depende de los requisitos de velocidad, precisión y capacidades de hardware disponibles (Fig, 3.3-4). Los modelos pequeños, como Mistral 7B y Baichuan 7B, son adecuados para tareas

ligeras y dispositivos móviles, mientras que los modelos potentes, como DeepSeek -V3, requieren importantes recursos informáticos pero ofrecen un alto rendimiento y compatibilidad con varios idiomas. En los próximos años, el mercado de los LLM crecerá rápidamente: veremos cada vez más modelos ligeros y especializados. En lugar de LLM de uso general que cubran todos los contenidos humanos de, surgirán modelos entrenados en un dominio específico. Por ejemplo, podemos esperar la aparición de modelos diseñados exclusivamente para manejar cálculos de ingeniería, estimaciones de construcción o datos en formatos CAD. Estos modelos especializados serán más rápidos, precisos y seguros de utilizar, sobre todo en entornos profesionales en los que son importantes la fiabilidad y la profundidad temática.

Una vez puesto en marcha el LLM local, puede adaptarse a las tareas específicas de la empresa. Para ello se utiliza la técnica del ajuste fino, mediante la cual el modelo se entrena aún más con documentos internos, instrucciones técnicas, plantillas de contratos o documentación de proyectos.

RAG: LLM inteligente -asistentes con acceso a datos corporativos

La siguiente etapa en la evolución de la aplicación de LLM en los negocios es la integración de modelos con datos corporativos reales en tiempo real. Este enfoque se denomina RAG (Retrieval-Augmented Generation) - Generación Recuperada-Aumentada. En esta arquitectura, el modelo lingüístico se convierte no sólo en una interfaz de diálogo, sino en un completo asistente inteligente capaz de navegar por documentos, planos, bases de datos y proporcionar respuestas precisas y contextuales.

> La principal ventaja del GAR es la posibilidad de utilizar datos internos de la empresa sin necesidad de entrenar previamente el modelo, manteniendo al mismo tiempo una gran precisión y flexibilidad en el tratamiento de la información.

La tecnología RAG combina dos componentes principales:

- Recuperación: el modelo se conecta a los almacenes de datos -documentos, tablas, PDF -archivos, dibujos- y recupera la información pertinente solicitada por el usuario.
- Generación aumentada: a partir de los datos extraídos, el modelo genera una respuesta precisa e informada, teniendo en cuenta el contexto y la especificidad de la consulta.

Para ejecutar LLM con soporte RAG, hay que seguir algunos pasos:

- **Preparación de datos**: recopilar los documentos necesarios, planos, especificaciones, tablas. Pueden estar en distintos formatos y estructuras, desde PDF hasta Excel.
- **Indexación y vectorización**: mediante herramientas como LlamaIndex o LangChain, los datos se convierten en representaciones vectoriales que permiten encontrar vínculos semánticos entre fragmentos de texto (más información sobre las bases de datos vectoriales y la conversión de grandes matrices en representaciones vectoriales, incluidos los proyectos CAD, en la Parte 8).
- **Consulta al asistente**: una vez cargados los datos, puedes hacer preguntas al modelo y éste buscará respuestas en el marco corporativo en lugar de en conocimientos generales recogidos de

internet.

Supongamos que una empresa tiene una carpeta constructionsite_docs, donde se almacenan contratos, instrucciones, presupuestos y tablas. Utilizando un script de Python (Fig. 3.3-5), podemos escanear esta carpeta y construir una indexación vectorial: cada documento se convertirá en un conjunto de vectores que reflejarán el contenido semántico del texto. Esto convierte los documentos en una especie de "mapa de significados" en el que el modelo puede navegar eficazmente y encontrar conexiones entre términos y frases.

Por ejemplo, el modelo "recuerda" que las palabras "devolución" y "reclamación" suelen encontrarse en la sección del contrato relativa al envío de materiales a la obra. Entonces, si se formula una pregunta -por ejemplo, "¿Cuál es nuestro plazo de devolución?" (Fig. 3.3-5 - línea 11 de código) - el LLM analizará los documentos internos y encontrará la información exacta, actuando como un asistente inteligente capaz de leer y comprender el contenido de todos los archivos corporativos.

llama.py

```
1  from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3  # Load documents from the folder
4  documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6  # Creating a vector index for semantic search
7  index = VectorStoreIndex.from_documents(documents)
8
9  # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)
```

Fig. 3.3-5 LM lee una carpeta con archivos, de forma similar a como una persona la abre y busca el documento deseado

El código puede ejecutarse en cualquier ordenador con Python instalado. Hablaremos más sobre el uso de Python e IDEs para ejecutar el código en el próximo capítulo.

El despliegue local de LLM no es sólo una tendencia, sino una solución estratégica para las empresas que valoran la seguridad y la flexibilidad. Sin embargo, desplegar LLM, ya sea en ordenadores locales de la empresa o utilizando soluciones en línea, es sólo el primer paso. Para aplicar las capacidades de LLM a tareas del mundo real, las empresas deben utilizar herramientas que les permitan no sólo recibir respuestas de chat, sino también almacenar la lógica creada en forma de código que pueda ejecutarse fuera del contexto de uso de LLM. Esto es importante para escalar las soluciones: unos procesos bien organizados permiten aplicar los desarrollos de IA a varios proyectos a la vez o incluso a toda la empresa.

En este contexto, la elección de un entorno de desarrollo (IDE) adecuado desempeña un papel importante.

Las modernas herramientas de programación permiten no sólo desarrollar soluciones basadas en LLM, sino también integrarlas en los procesos empresariales existentes, convirtiéndolos en ETL automatizados -Pipeline

CAPÍTULO 3.4.
IDE CON SOPORTE LLM Y FUTUROS CAMBIOS DE PROGRAMACIÓN

Elegir una IDE: de los experimentos LLM a las soluciones empresariales

Al sumergirse en el mundo de la automatización, el análisis de datos y la inteligencia artificial -especialmente cuando se trabaja con grandes modelos lingüísticos (LLM)- es fundamental elegir el entorno de desarrollo integrado (IDE) adecuado. Este IDE será su principal herramienta de trabajo: el lugar donde se ejecutará el código generado por el LLM, tanto en un ordenador local como dentro de la red corporativa. La elección del IDE no sólo determina la comodidad de tu trabajo, sino también la rapidez con la que podrás pasar de consultas experimentales del LLM a soluciones completas integradas en procesos de negocio reales.

> **IDE** (Entorno de Desarrollo Integrado) es un bloque de construcción versátil en su ordenador para la automatización de procesos y el procesamiento de datos. En lugar de guardar una sierra, un martillo, un taladro y otras herramientas por separado, dispone de un dispositivo que puede hacerlo todo: cortar, sujetar, taladrar e incluso comprobar la calidad de los materiales. IDE para programadores es un espacio único donde puedes escribir código (por analogía con la construcción: crear planos), probar su funcionamiento (montaje de modelos de edificios), encontrar errores (como comprobar la resistencia de las estructuras en construcción) y ejecutar el proyecto terminado (puesta en marcha de la casa).

Una visión general de los IDE más populares:

- **PyCharm®** (JetBrains) es un potente IDE profesional para Python. Es muy adecuado para proyectos serios debido al gran número de funciones incorporadas. Sin embargo, el soporte básico para archivos Jupyter interactivos (IPYNB) solo está disponible en la versión de pago, y los principiantes pueden encontrar la interfaz abrumadora.

> Un archivo con la extensión IPYNB (Interactive Python Notebook) es un formato para cuadernos Jupyter® interactivos (Fig, 3.4-1) en el que se combinan código, visualizaciones y explicaciones en un único documento. Este formato es ideal para crear informes, análisis y escenarios de formación.

- **VS Code®** (Microsoft) es una herramienta rápida, flexible y personalizable con soporte gratuito para IPYNB y numerosos plug-ins. Adecuada tanto para principiantes como para profesionales. Permite integrar GitHub Copilot y plugins de modelos lingüísticos, lo que la convierte en una gran opción para proyectos de IA y ciencia de datos.
- **Jupyter Notebook** - Una opción clásica y popular para la experimentación y el aprendizaje. Permite escribir código, añadir explicaciones y visualizar resultados en una única interfaz (Fig. 3.4-1). Ideal para probar rápidamente hipótesis, trabajar con LLM y crear pasos reproducibles de análisis de datos wild. Para gestionar dependencias y bibliotecas, recomendamos utilizar Anaconda Navigator,

una interfaz visual para gestionar el entorno Python.

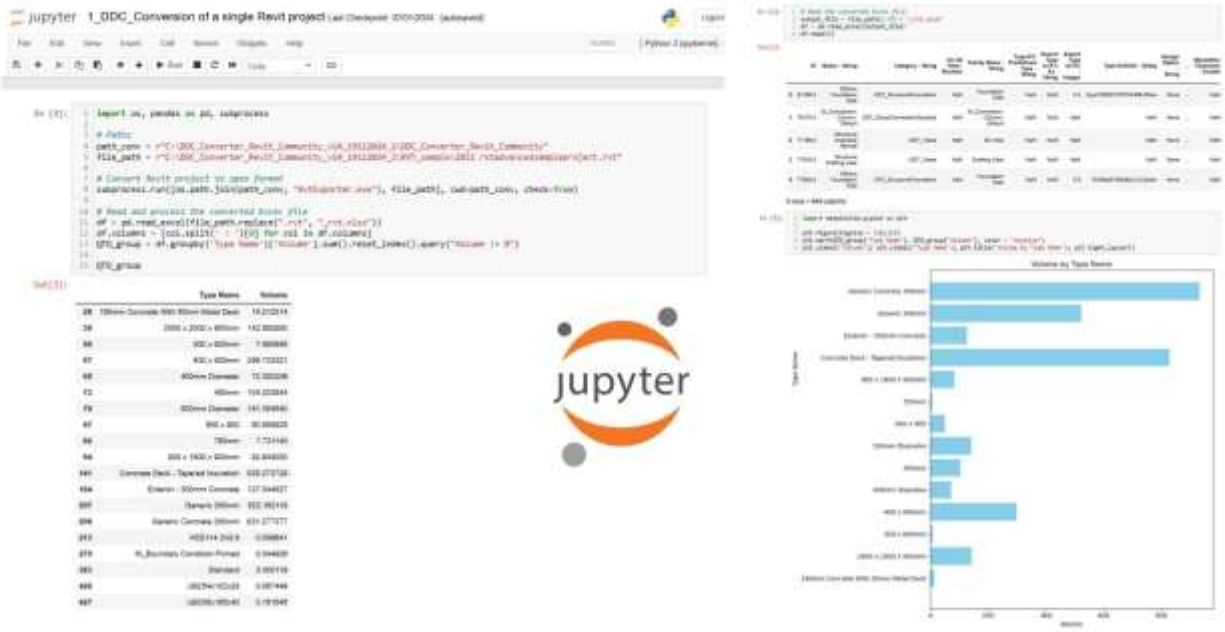


Fig. 3.4-1 Jupyter Notebook una de las herramientas más cómodas y populares para crear procesos Pipeline.

- **Google Collab**™ (y la plataforma Kaggle (Fig, 9.2-5)) es una alternativa a Jupyter basada en la nube que proporciona acceso gratuito a la GPU/TPU. Es una gran solución para empezar: no requiere instalación de software local y permite trabajar directamente desde un navegador. Admite la integración con Google Drive y, recientemente, con Gemini (el LLM de Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Fig. 3.4-2 Comparación de IDEs: Jupyter Notebook una de las herramientas más cómodas y sencillas para crear procesos Pipeline.

La elección del IDE depende de tus tareas. Si quieres empezar a trabajar rápidamente con IA, prueba Jupyter Notebook o Google Collab. Para proyectos serios es mejor utilizar PyCharm o VS Code. Lo principal es

empezar. Las herramientas modernas te permiten convertir rápidamente tus experimentos en soluciones prácticas.

Todos los IDEs descritos anteriormente permiten crear pipelines de procesamiento de datos, es decir, cadenas de módulos de bloques de código (que podrían ser generados por el LLM), cada uno de los cuales es responsable de una etapa diferente, por ejemplo:

- escenarios analíticos,
- cadenas de extracción de información de documentos,
- respuestas automáticas basadas en RAG,
- generación de informes y visualizaciones.

Gracias a su estructura modular, cada paso puede representarse como un bloque independiente: carga de datos→ filtrado→ análisis→ visualización→ exportación de resultados. Estos bloques pueden reutilizarse, - adaptarse y ensamblarse en nuevas cadenas como un constructor, sólo para datos.

Para ingenieros, gestores y analistas, esto abre la posibilidad de documentar la lógica de la toma de decisiones en forma de código que puede generarse con LLM. Este enfoque ayuda a agilizar tareas rutinarias, automatizar operaciones típicas y crear procesos repetibles en los que cada paso está claramente documentado y es transparente para todos los miembros del equipo.

Las herramientas ETL automatizadas Pipelines (Fig. 7.2-3), Apache Airflow (Fig. 7.4-4), Apache NiFi (Fig. 7.4-5) y n8n (Fig. 7.4-6) para construir bloques de lógica para la automatización de procesos se tratarán con más detalle en la Parte 7 y la Parte 8 del libro.

IDE compatible con LLM y futuros cambios de programación

La integración de la inteligencia artificial en los procesos de desarrollo está cambiando el panorama de la programación. Los entornos modernos ya no son meros editores de texto con resaltado de sintaxis: se están convirtiendo en asistentes inteligentes capaces de entender la lógica del proyecto, completar el código e incluso explicar cómo funciona un fragmento de código concreto. Están apareciendo en el mercado productos que utilizan la IA para superar los límites del desarrollo convencional:

- **GitHub Copilot** (se integra con VS Code, PyCharm): AI es un asistente que genera código basado en comentarios o descripciones parciales, convirtiendo pistas textuales en soluciones acabadas.
- **Cursor** (un fork de VS Code con AI -kernel): permite no sólo terminar de escribir código, sino también hacer preguntas al proyecto, buscar dependencias y aprender de la base de código.
- **JetBrains AI Assistant**: un plugin para JetBrains IDE (incluyendo PyCharm) con la función de explicar código complejo, optimización y creación de pruebas.
- **Amazon CodeWhisperer**: un análogo de Copilot centrado en la seguridad y la compatibilidad con los servicios AWS de Amazon.

La programación experimentará un cambio drástico en los próximos años. El foco principal pasará de la escritura rutinaria de código al diseño de modelos y arquitectura de datos: los desarrolladores participarán más en el diseño de sistemas, mientras que la IA se encargará de las tareas de plantilla: generación de

código, pruebas, documentación y funciones básicas. El futuro de la programación es una colaboración entre humanos e IA, en la que las máquinas se encargan de la rutina técnica y los humanos se centran en la creatividad.

La programación en lenguaje natural se convertirá en algo cotidiano. La personalización de los IDE alcanzará un nuevo nivel: los entornos de desarrollo aprenderán a adaptarse al estilo de trabajo del usuario y sus empresas anticipando patrones, ofreciendo soluciones contextuales y aprendiendo de proyectos anteriores

> Esto no suprime el papel del desarrollador, pero lo transformará radicalmente: de la escritura de código a la gestión del conocimiento, la calidad y los procesos. Esta evolución también afectará a la inteligencia empresarial, donde la creación de informes, visualizaciones y aplicaciones de apoyo a la toma de decisiones se hará cada vez más mediante la generación de código y lógica con ayuda de IA y LLM, chat e interfaces de agentes.

Una vez que una empresa ha creado chats LLM y seleccionado un entorno de desarrollo adecuado, el siguiente paso importante es organizar los datos. Este proceso implica extraer información de fuentes dispares, limpiarla, transformarla en un formato estructurado e integrarla en los sistemas corporativos.

En un enfoque moderno de la gestión de datos centrado en los datos, uno de los objetivos clave es reunirlos en un único formato universal que sea compatible con un gran número de herramientas y aplicaciones. Se necesitan bibliotecas especializadas para gestionar los procesos de estructuración y los datos estructurados. Una de las más potentes, flexibles y populares es la biblioteca Pandas para Python. Permite procesar cómodamente datos tabulares: filtrar, agrupar, limpiar, anexar, agregar e informar.

Python Pandas: una herramienta indispensable para trabajar con datos

Pandas ocupa un lugar especial en el mundo del análisis de datos y la automatización. Se trata de una de las bibliotecas más populares y utilizadas del lenguaje de programación Python [86], diseñada para trabajar con datos estructurados.

Una biblioteca es como un conjunto de herramientas ya preparadas: funciones, módulos, clases. Igual que en una obra no hace falta inventar un martillo o un nivel cada vez, en programación las bibliotecas permiten resolver problemas rápidamente sin reinventar funciones y soluciones básicas.

> Pandas es una biblioteca Python de código abierto , que proporciona estructuras de datos intuitivas y de alto rendimiento, en particular DataFrame, un formato universal para trabajar con tablas. Pandas es una navaja suiza para analistas, ingenieros y desarrolladores que trabajan con datos.

Python es un lenguaje de programación de alto nivel con una sintaxis sencilla que se utiliza activamente en análisis, automatización, aprendizaje automático y desarrollo web. Su popularidad se debe a la legibilidad de su código, su naturaleza multiplataforma y su rico ecosistema de bibliotecas. Hasta la fecha, se han creado más de 137.000 paquetes de código abierto para Python [87], y este número sigue creciendo casi a diario. Cada una de estas bibliotecas es una especie de repositorio de funciones listas para usar: desde

sencillas operaciones matemáticas hasta complejas herramientas de procesamiento de imágenes, análisis de big data, redes neuronales e integración con servicios externos.

> En otras palabras, imagina que tienes acceso libre y gratuito a cientos de miles de soluciones de software listas para usar, bibliotecas y herramientas que puedes integrar directamente en tus procesos de negocio. Es como un enorme catálogo de aplicaciones de automatización, análisis, visualización, integración y mucho más, y todo ello disponible inmediatamente después de instalar Python.

Pandas es uno de los paquetes más populares del ecosistema Python. En 2022, el número medio de descargas de la biblioteca Pandas alcanzó los 4 millones al día (Fig, 3.4-3), mientras que a principios de 2025, esta cifra ha aumentado hasta los 12 millones de descargas diarias, lo que refleja su creciente popularidad y su uso generalizado en el análisis de datos y en los chats de LLM [86]

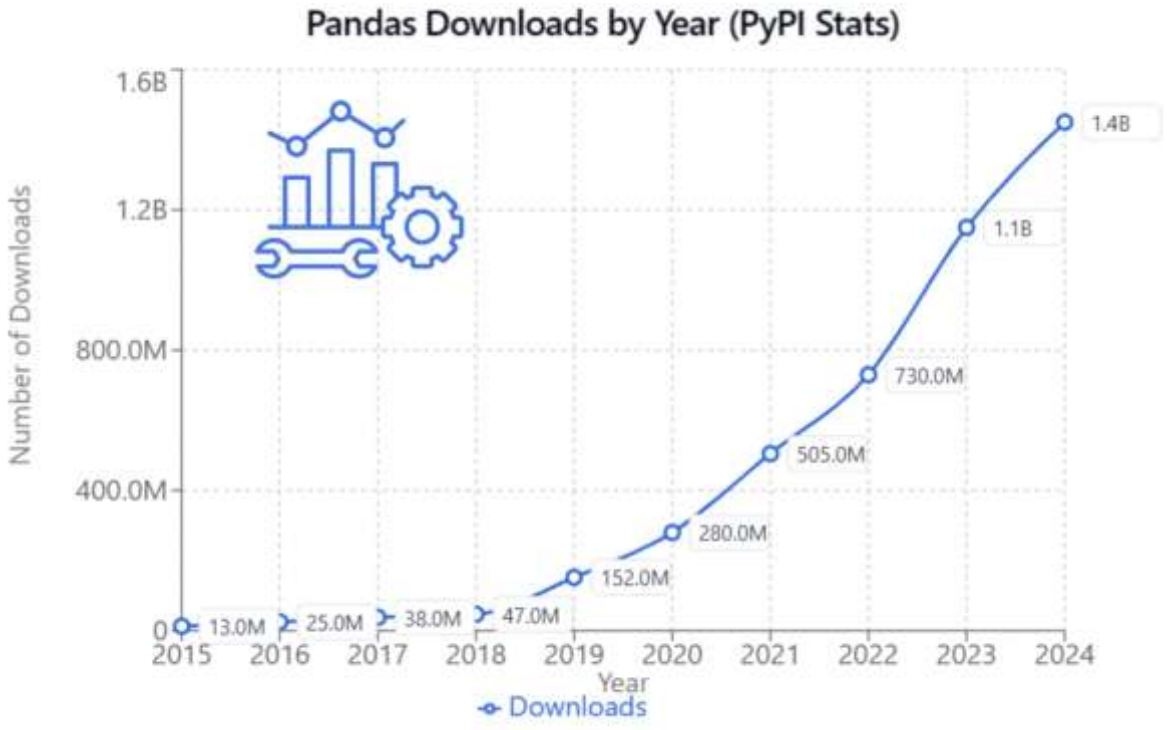


Fig. 3.4-3 Pandas es una de las bibliotecas más descargadas. En 2024, su número anual de descargas superó los 1.400 millones.

El lenguaje de consulta de la biblioteca de Pandas es similar en funcionalidad al lenguaje de consulta SQL, del que hablamos en el capítulo "Bases de datos relacionales y lenguaje de consulta SQL".

> En el mundo de la analítica y la gestión de datos estructurados, Pandas destaca por su sencillez, rapidez y potencia, proporcionando a los usuarios una amplia gama de herramientas para analizar y procesar la información con eficacia.

Ambas herramientas - SQL y Pandas - proporcionan potentes capacidades de manipulación de datos, especialmente si se comparan con el Excel tradicional. Admiten operaciones como la selección, el filtrado (Fig, 3.4-4), con la única diferencia de que SQL está optimizado para trabajar con bases de datos relacionales, mientras que Pandas procesa los datos en RAM, lo que le permite ejecutarse en cualquier ordenador, sin necesidad de crear bases de datos y desplegar una infraestructura independiente.

Fig. 3.4-4 Pandas, a diferencia de SQL, tiene la flexibilidad de trabajar con una gran variedad de formatos de datos, no limitados a bases de datos.

Pandas suele preferirse para la investigación científica, la automatización de procesos, la creación de pipelines (incluyendo ETL) y la manipulación de datos en Python, mientras que SQL es un estándar de gestión de bases de datos y se utiliza a menudo en entornos empresariales para manejar grandes cantidades de datos.

> La biblioteca Pandas del lenguaje de programación Python permite realizar no sólo operaciones básicas, como leer y escribir tablas, sino también tareas más complejas, como combinar datos, agruparlos y realizar cálculos analíticos complejos.

En la actualidad, la biblioteca Pandas se utiliza no solo en la investigación académica y la analítica empresarial, sino también en conjunción con los modelos LLM. Por ejemplo, la división Meta® (Facebook™), al publicar un nuevo modelo de código abierto LlaMa 3.1 en 2024, prestó especial atención al trabajo con

datos estructurados, haciendo que uno de los casos clave y primeros en su lanzamiento fuera exactamente el procesamiento de dataframes estructurados (Fig. 3.4-5) en formato CSV y la integración con la biblioteca Pandas directamente en el chat.

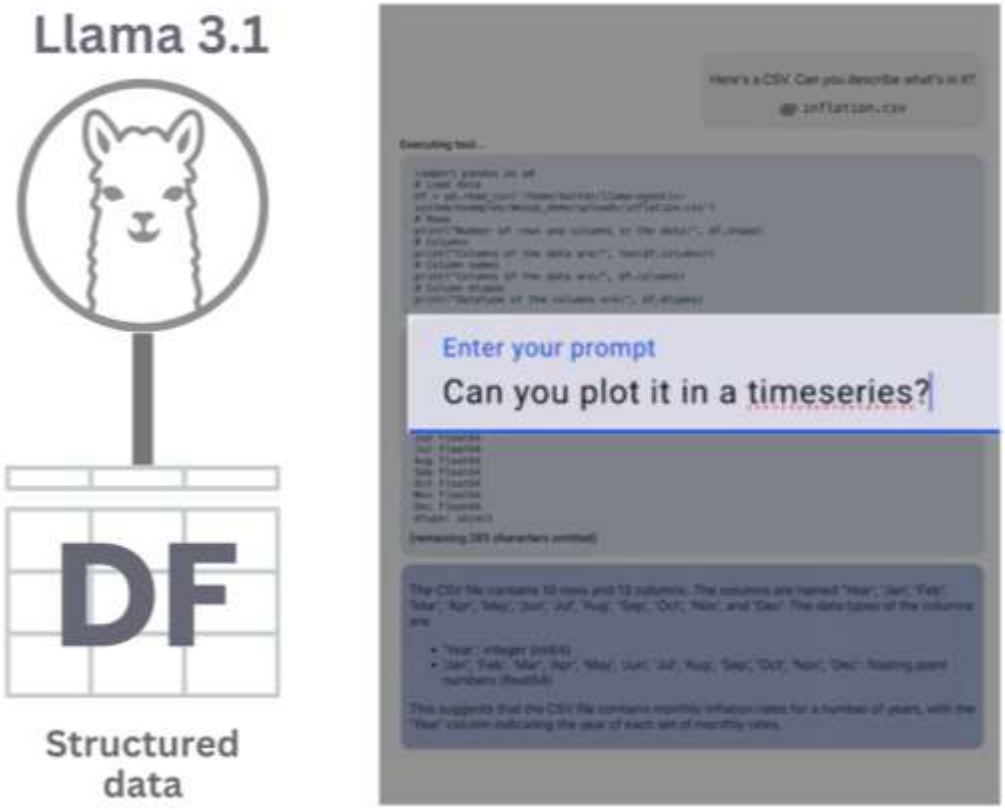


Fig. 3.4-5 Uno de los primeros y principales casos del equipo Meta presentados en LlaMa 3.1 en 2024 fue la construcción de aplicaciones utilizando Pandas.

> *Pandas es una herramienta esencial para millones de científicos de datos que procesan y preparan datos para la IA generativa. Acelerar Pandas sin cambiar el código será un gran paso adelante. Los científicos de datos podrán procesar datos en minutos en lugar de horas y obtener órdenes de magnitud de datos para entrenar modelos de IA generativa [88].*
>
> - Jensen Huang, fundador y consejero delegado de NVIDIA

Usando Pandas, es posible gestionar y analizar conjuntos de datos mucho más allá de las capacidades de Excel. Mientras que Excel suele ser capaz de manejar hasta 1 millón de filas de datos, Pandas puede manejar fácilmente conjuntos de datos (Fig, 9.1-2, Fig. 9.1-10) que contienen decenas de millones de filas [89]. Esta capacidad permite a los usuarios realizar sofisticados análisis y visualizaciones de datos en grandes conjuntos de datos, proporcionando una visión profunda y facilitando la toma de decisiones basada en

datos. Además, Pandas cuenta con un fuerte apoyo de la comunidad [90]: cientos de millones de desarrolladores y analistas de todo el mundo (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) lo utilizan en línea o fuera de línea todos los días, proporcionando un gran número de soluciones listas para usar para cualquier problema empresarial.

En el corazón de la mayoría de los procesos analíticos de Python se encuentra una forma estructurada de datos llamada DataFrame, proporcionada por la biblioteca Pandas. Se trata de una herramienta potente y flexible para organizar, analizar y visualizar datos tabulares.

DataFrame: formato universal de datos tabulares

DataFrame es la estructura central de la librería Pandas, que es una tabla bidimensional (Fig, 3.4-6) donde las filas corresponden a objetos individuales o registros y las columnas corresponden a sus características, parámetros o categorías. Esta estructura se parece visualmente a las hojas de cálculo de Excel, pero es muy superior en flexibilidad, escalabilidad y funcionalidad.

> Un **DataFrame** es una forma de representar y procesar datos tabulares almacenados en la memoria RAM del ordenador.

DataFrame es una forma de representar y procesar datos tabulares almacenados en la memoria RAM del ordenador. En una tabla, las filas pueden reflejar, por ejemplo, los elementos de un proyecto de construcción, y las columnas, sus propiedades: categorías, dimensiones, coordenadas, coste, términos, etcétera. Además, una tabla de este tipo puede contener tanto información sobre un proyecto (Fig. 4.1-13) como datos sobre millones de objetos de miles de proyectos diferentes (Fig. 9.1-10). Gracias a las operaciones vectorizadas de Pandas, es fácil filtrar, agrupar y agregar tales volúmenes de información a gran velocidad.

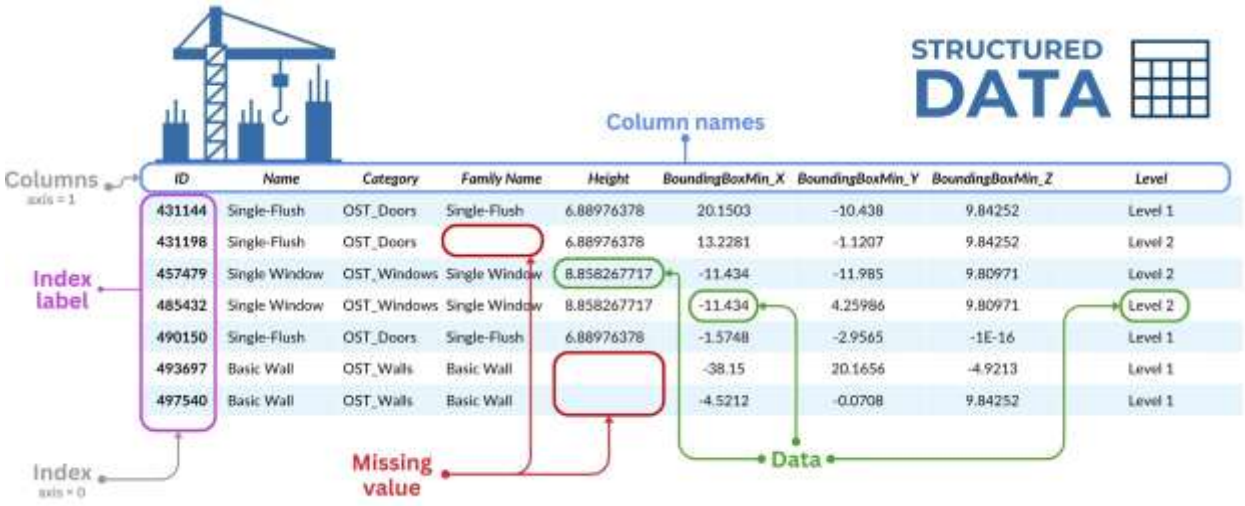


Fig. 3.4-6 Proyecto de construcción como DataFrame es una tabla bidimensional con elementos en filas y atributos en columnas.

Nvidia calcula que, en la actualidad, hasta el 30% de todos los recursos informáticos se utilizan para procesar datos estructurados -tramas de datos- y esta proporción sigue creciendo.

> *El tratamiento de datos es lo que probablemente se hace en un tercio de la informática mundial en cada empresa. El procesamiento de datos y los datos de la mayoría de las empresas están en DataFrame, en formato de tabla*
>
> - Jensen Huang, consejero delegado de Nvidia [91]

Vamos a enumerar algunas características clave de DataFrame en Pandas:

- Columnas: en DataFrame, los datos se organizan en columnas, cada una con un nombre único. Las columnas de atributos pueden contener datos de distintos tipos, de forma similar a las columnas de las bases de datos o las columnas de las tablas.
- **Pandas Series** es una estructura de datos unidimensional en Pandas, similar a una lista o columna en una tabla, donde cada valor corresponde a un índice diferente

> La serie Pandas tiene más de 400 atributos y métodos, lo que hace que trabajar con datos sea increíblemente flexible. Puede aplicar directamente una de las cuatrocientas funciones disponibles a una columna, realizar operaciones matemáticas, filtrar datos, reemplazar valores, trabajar con fechas, cadenas y mucho más. Además, Series admite operaciones vectorizadas, lo que acelera enormemente el procesamiento de grandes conjuntos de datos en comparación con los cálculos cíclicos. Por ejemplo, puede multiplicar fácilmente todos los valores por un número, sustituir los datos que faltan o aplicar transformaciones complejas sin escribir complicados bucles.

- **Filas**: en DataFrame se pueden indexar con valores únicos. Este índice permite cambiar y ajustar rápidamente los datos de filas específicas.
- **Índice**: Por defecto, cuando se crea un DataFrame Pandas asigna a cada fila un índice de 0 a N-1 (donde N es el número de todas las filas del DataFrame). Sin embargo, el índice se puede cambiar para incluir designaciones especiales como fechas o características únicas.
- **Indexar** filas en un DataFrame significa que a cada fila se le asigna un nombre o etiqueta única, que se denomina índice del DataFrame.
- **Tipos de** datos**:** DataFrame admite una gran variedad de tipos de datos, entre los que se incluyen: `int`, `float`, `bool`, `datetime64` y `obect` para datos de texto. Cada columna DataFrame tiene su propio tipo de datos que determina qué operaciones se pueden realizar en su contenido.
- **Operaciones con datos**: DataFrame admite una amplia gama de operaciones para el tratamiento de datos, como la agregación (`groupby`), la fusión (`merge` y `join`), la concatenación (`concat`), la división-aplicación-combinación y muchas otras técnicas de transformación de datos.
- **Manipulación del tamaño**: DataFrame permite añadir y eliminar columnas y filas, lo que la convierte en una estructura dinámica que puede modificarse en función de sus necesidades de análisis de datos.

- **Visualización de datos:** mediante técnicas de visualización integradas o la interconexión con conocidas bibliotecas de visualización de datos como Matplotlib o Seaborn, DataFrame puede convertirse fácilmente en gráficos y diagramas para presentar los datos de forma gráfica.
- **Entrada y salida de datos:** Pandas proporciona funciones para leer, importar y exportar datos a varios formatos de archivo como CSV, Excel, JSON, HTML y SQL, convirtiendo potencialmente a DataFrame en un eje central para la recogida y distribución de datos.

A diferencia de CSV y XLSX, Pandas DataFrame proporciona mayor flexibilidad y rendimiento a la hora de trabajar con datos: puede manejar grandes cantidades de información en RAM, soporta tipos de datos extendidos (incluyendo fechas, valores lógicos y series temporales) y proporciona amplias capacidades para filtrar, agregar, fusionar y visualizar datos. Mientras que CSV no almacena información sobre los tipos y la estructura de los datos, y XLSX suele estar sobrecargado de formato y tiene poca escalabilidad, DataFrame sigue siendo la opción óptima para el análisis rápido, la automatización de procesos y la integración con modelos de IA (Fig, 3.4-7). Los siguientes capítulos explorarán cada uno de estos aspectos de los datos en detalle, también en la Parte 8 del libro, formatos similares como Parquet, Apache Orc, JSON, Feather, HDF5 y almacenes de datos serán discutidos en detalle (Fig. 8.1-2).

	XLSX	CSV	Pandas DataFrame
Storage	Tabular	Tabular	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
Compression	Built-in	None	None (in-memory)
Performance	Low	Medium	High (memory dependent)
Complexity	High (formatting, styles)	Low	Low
Data Type Support	Limited	Very limited	Extended
Scalability	Low	Low	Medium (memory limited)

Fig. 3.4-7 DataFrame es la elección óptima de manipulación de datos con alto rendimiento y soporte avanzado de tipos de datos.

Gracias a su flexibilidad, potencia y facilidad de uso, la biblioteca Pandas y el formato DataFrame se han convertido en el estándar de facto en el análisis de datos en Python. Son ideales tanto para crear informes sencillos como para construir complejos pipelines analíticos, especialmente en conjunción con modelos LLM.

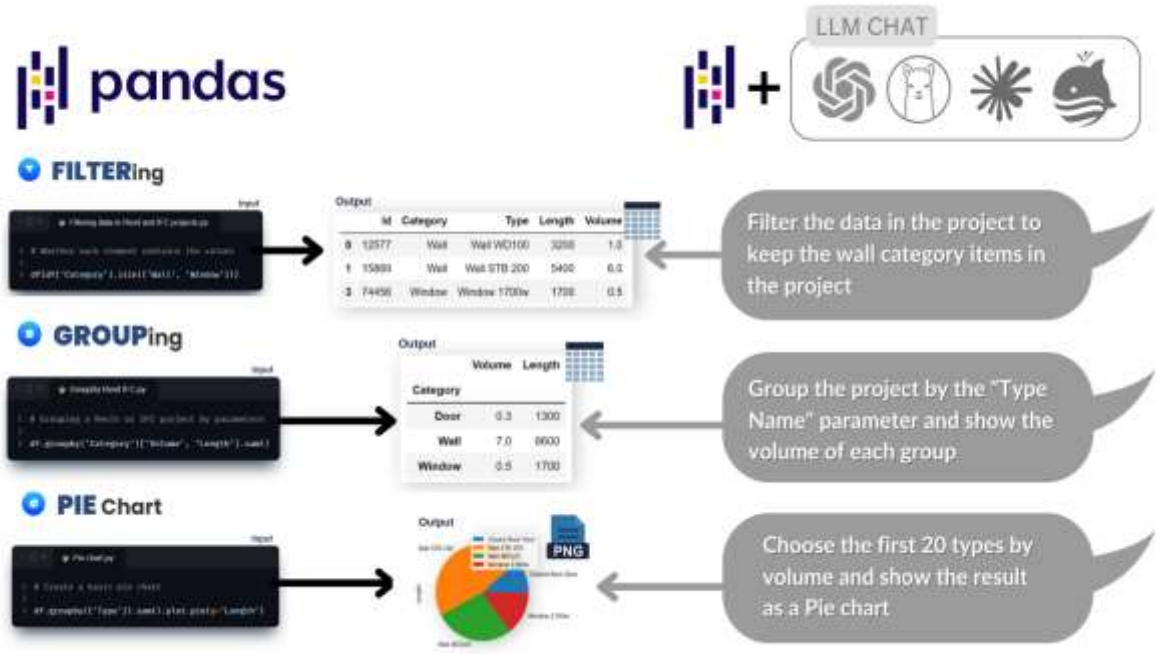


Fig. 3.4-8 Los LLM simplifican la interacción con Pandas: en lugar de código, basta con una consulta de texto.

En la actualidad, Pandas se utiliza activamente en salas de chat basadas en LLM como ChatGPT, LlaMa, DeepSeek, QWEN y otras. En muchos casos, cuando un modelo recibe una consulta relacionada con el procesamiento de tablas, la validación de datos o la analítica, genera código utilizando exactamente la biblioteca Pandas. Esto convierte a DataFrame en un "lenguaje" natural para representar datos en diálogos con IA (Fig, 3.4-8).

Las tecnologías de datos modernas, como Pandas, facilitan el análisis, la automatización y la integración de los datos en los procesos empresariales. Proporcionan resultados con rapidez, reducen la carga de trabajo de los especialistas y garantizan la repetibilidad de las operaciones.

Próximos pasos: crear un marco de datos sostenible

En esta parte hemos estudiado los principales tipos de datos utilizados en el sector de la construcción, hemos conocido los distintos formatos para almacenarlos y hemos analizado el papel de las herramientas modernas, como LLM e IDEs, en el tratamiento de la información. Hemos aprendido que una gestión eficaz de los datos es la base de una toma de decisiones informada y de la automatización de los procesos empresariales. Las organizaciones que son capaces de estructurar y organizar sus datos obtienen una importante ventaja competitiva en las fases de procesamiento y transformación de datos.

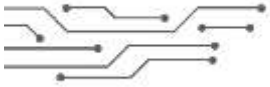
Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar los planteamientos comentados en sus tareas cotidianas:

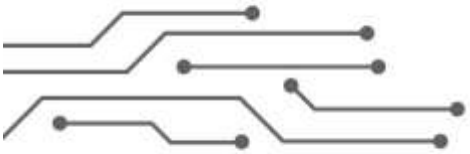
- Realice una auditoría de datos de sus procesos

  - ☐ Haz un inventario de todos los tipos de datos que utilizas en tus proyectos
  - ☐ Determine qué tipos y modelos de datos son los más importantes para sus procesos empresariales.
  - ☐ Identificar las áreas problemáticas en las que la información suele estar desestructurada, mal estructurada o inaccesible.
- Empezar a elaborar una estrategia de gestión de datos
  - ☐ Plantear cuestiones políticas y normas para tratar distintos tipos de datos
  - ☐ Analice cuáles de sus flujos de trabajo pueden mejorarse convirtiendo los datos no estructurados en datos estructurados.
  - ☐ Crear una política de almacenamiento y acceso a los datos que tenga en cuenta la seguridad y la confidencialidad.
- Instalar y dominar las herramientas básicas para trabajar con datos
  - ☐ Elija un IDE adecuado a sus tareas (por ejemplo, instale VS Code o Jupyter Notebook)
  - ☐ Pruebe a crear un LLM local para tratar sus datos personales de forma confidencial.
  - ☐ Empieza a experimentar con la librería Pandas para procesar datos tabulares XLSX
  - ☐ Describe al LLM las tareas típicas que manejas en herramientas de hojas de cálculo o bases de datos y pídele que automatice el trabajo usando Pandas

La aplicación de estos pasos le permitirá transformar gradualmente su forma de trabajar con los datos, pasando de conjuntos de datos dispares y desestructurados a un ecosistema unificado en el que los datos se convierten en un activo accesible y comprensible. Empiece poco a poco: cree su primer DataFrame en Pandas, ejecute un LLM local, automatice su primera tarea rutinaria utilizando Python (por ejemplo, hojas de cálculo de Excel).

La cuarta parte del libro se centrará en la calidad de los datos, la organización, estructuración y modelización de datos. Nos centraremos en las metodologías que transforman fuentes de datos dispares -desde PDF y textos hasta imágenes y modelos CAD- en conjuntos de datos estructurados aptos para el análisis y la automatización. También exploraremos cómo se formalizan los requisitos de los datos, cómo se construyen los modelos conceptuales y lógicos en los proyectos de construcción y cómo pueden ayudar en este proceso los modelos de lenguaje modernos (LLM).





IV PARTE

CALIDAD DE LOS DATOS: ORGANIZACIÓN, ESTRUCTURACIÓN, MODELIZACIÓN

La cuarta parte se centra en las metodologías y tecnologías para transformar información dispar en conjuntos de datos estructurados de alta calidad. Se analizan en detalle los procesos de formación y documentación de requisitos de datos como base para una arquitectura de la información eficaz en proyectos de construcción. Se presentan métodos prácticos para extraer información estructurada de diversas fuentes (documentos PDF, imágenes, archivos de texto, modelos CAD) con ejemplos de aplicación. Se analiza el uso de expresiones regulares (RegEx) y otras herramientas para la validación y verificación automáticas de datos. Se describe paso a paso el proceso de modelización de datos a nivel conceptual, lógico y físico, teniendo en cuenta las especificidades del sector de la construcción. Se muestran ejemplos concretos de utilización de modelos lingüísticos (LLM) para automatizar los procesos de estructuración y validación de la información. Se proponen enfoques eficaces para la visualización de los resultados de los análisis, aumentando la disponibilidad de información analítica para todos los niveles de la gestión de proyectos de construcción

CAPÍTULO 4.1.
CONVERSIÓN DE DATOS A UN FORMATO ESTRUCTURADO

En la era de la economía impulsada por los datos, éstos se están convirtiendo en la base de la toma de decisiones, en lugar de ser un obstáculo. En lugar de adaptar constantemente la información a cada nuevo sistema y sus formatos, las empresas buscan cada vez más formar un único modelo de datos estructurados que sirva de fuente universal de verdad para todos los procesos. Los sistemas de información modernos no se diseñan en torno a formatos e interfaces, sino en torno al significado de los datos, porque la estructura puede cambiar, pero el significado de la información sigue siendo el mismo durante mucho más tiempo.

> La clave para trabajar eficazmente con los datos no es convertirlos y transformarlos sin cesar, sino organizarlos correctamente desde el principio: crear una estructura universal capaz de aportar transparencia, automatización e integración en todas las fases del ciclo de vida del proyecto.

El enfoque tradicional obliga a realizar ajustes manuales con cada implantación de una nueva plataforma: migrar datos, cambiar nombres de atributos y ajustar formatos. Estos pasos no mejoran la calidad de los datos en sí, sino que solo enmascaran los problemas, creando un círculo vicioso de transformaciones interminables. Como resultado, las empresas se vuelven dependientes de soluciones de software específicas, y la transformación digital se ralentiza.

En los capítulos siguientes, veremos cómo estructurar correctamente los datos y, a continuación, cómo crear modelos universales, minimizar la dependencia de las plataformas y centrarnos en lo más importante: los datos como recurso estratégico en torno al cual se construyen procesos sostenibles.

Aprender a convertir documentos, PDF, imágenes y textos en formatos estructurados.

En los proyectos de construcción, la inmensa mayoría de la información existe en forma no estructurada: documentos técnicos, declaraciones de trabajo, planos, especificaciones, calendarios, protocolos. Su diversidad, tanto de formato como de contenido, dificulta la integración y la automatización.

> El proceso de conversión a formatos estructurados o semiestructurados puede variar en función del tipo de datos de entrada y de los resultados de tratamiento deseados.

Transformar los datos de no estructurados a estructurados es tanto un arte como una ciencia. Este proceso varía en función del tipo de datos de entrada y de la finalidad del análisis y suele ocupar una parte importante del trabajo del ingeniero de datos (Fig, 3.2-5) y del analista, con el objetivo de producir un conjunto de datos limpios y organizados.

Fig. 4.1-1 Conversión de un documento escaneado no estructurado en un formato de tabla estructurado.

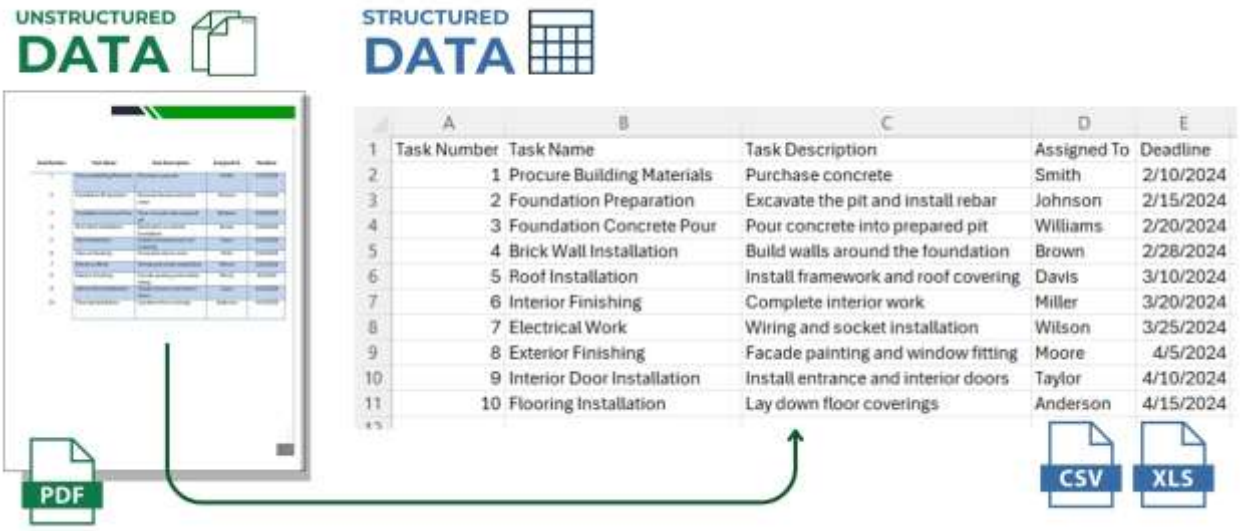
La conversión de documentos, PDF, imágenes y textos a un formato estructurado (Fig, 4.1-1) es un proceso paso a paso que incluye las siguientes etapas:

- Extraer): En este paso, se carga un documento fuente o una imagen que contenga datos no estructurados. Puede ser, por ejemplo, un documento PDF, una fotografía, un dibujo o un esquema.
- **Transformación de datos** (Transform): A continuación se procede a transformar los datos no estructurados en un formato estructurado. Por ejemplo, puede consistir en reconocer e interpretar el texto de las imágenes mediante el reconocimiento óptico de caracteres (OCR) u otros métodos de procesamiento.
- **Cargar y guardar los datos** (Load): el último paso consiste en guardar los datos procesados en varios formatos, como CSV, XLSX, XML, JSON, para seguir trabajando, donde la elección del formato depende de los requisitos y preferencias específicos.

Este proceso, conocido como ETL (Extraer, Transformar, Cargar), desempeña un papel clave en el tratamiento automatizado de datos y se tratará con más detalle en el capítulo "ETL y Pipeline: Extraer, Transformar, Cargar". A continuación, veremos ejemplos de transformación de documentos de distintos formatos en datos estructurados.

Ejemplo de conversión de un documento PDF en una tabla

Una de las tareas más habituales en los proyectos de construcción es procesar especificaciones en formato PDF. Para demostrar la transición de datos no estructurados a un formato estructurado, consideremos un ejemplo práctico: extraer una tabla de un documento PDF y convertirla a formato CSV o Excel (Fig. 4.1-2).



	A	B	C	D	E
1	Task Number	Task Name	Task Description	Assigned To	Deadline
2	1	Procure Building Materials	Purchase concrete	Smith	2/10/2024
3	2	Foundation Preparation	Excavate the pit and install rebar	Johnson	2/15/2024
4	3	Foundation Concrete Pour	Pour concrete into prepared pit	Williams	2/20/2024
5	4	Brick Wall Installation	Build walls around the foundation	Brown	2/28/2024
6	5	Roof Installation	Install framework and roof covering	Davis	3/10/2024
7	6	Interior Finishing	Complete interior work	Miller	3/20/2024
8	7	Electrical Work	Wiring and socket installation	Wilson	3/25/2024
9	8	Exterior Finishing	Facade painting and window fitting	Moore	4/5/2024
10	9	Interior Door Installation	Install entrance and interior doors	Taylor	4/10/2024
11	10	Flooring Installation	Lay down floor coverings	Anderson	4/15/2024

Fig. 4.1-2 A diferencia del PDF, los formatos CSV y XLSX están muy extendidos y se integran fácilmente en diversos sistemas de gestión de datos.

Los modelos de lenguaje LLM, como ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN simplifican enormemente la forma en que los científicos de datos trabajan con los datos, reduciendo la necesidad de un aprendizaje profundo de los lenguajes de programación y permitiendo resolver muchas tareas con consultas de texto.

Por lo tanto, en lugar de perder tiempo buscando soluciones en Internet (normalmente en el sitio web StackOverFlow o en foros y chats temáticos) o poniéndonos en contacto con especialistas en procesamiento de datos, podemos utilizar las capacidades de los modernos LLM en línea o locales. Basta con hacer una consulta y el modelo proporcionará un código listo para convertir el documento PDF en un formato tabular.

- Envíe la siguiente solicitud de texto a cualquier modelo LLM (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otro):

> Por favor, escriba un código para extraer texto de un archivo PDF que contenga una tabla. El código debe tomar la ruta del archivo como argumento y devolver la tabla extraída como un DataFrame⏎

- La respuesta al modelo LLM se presentará en la mayoría de los casos en forma de código en Python, ya que este lenguaje se utiliza ampliamente para el tratamiento de datos, la automatización y el trabajo con diversos formatos de archivo:

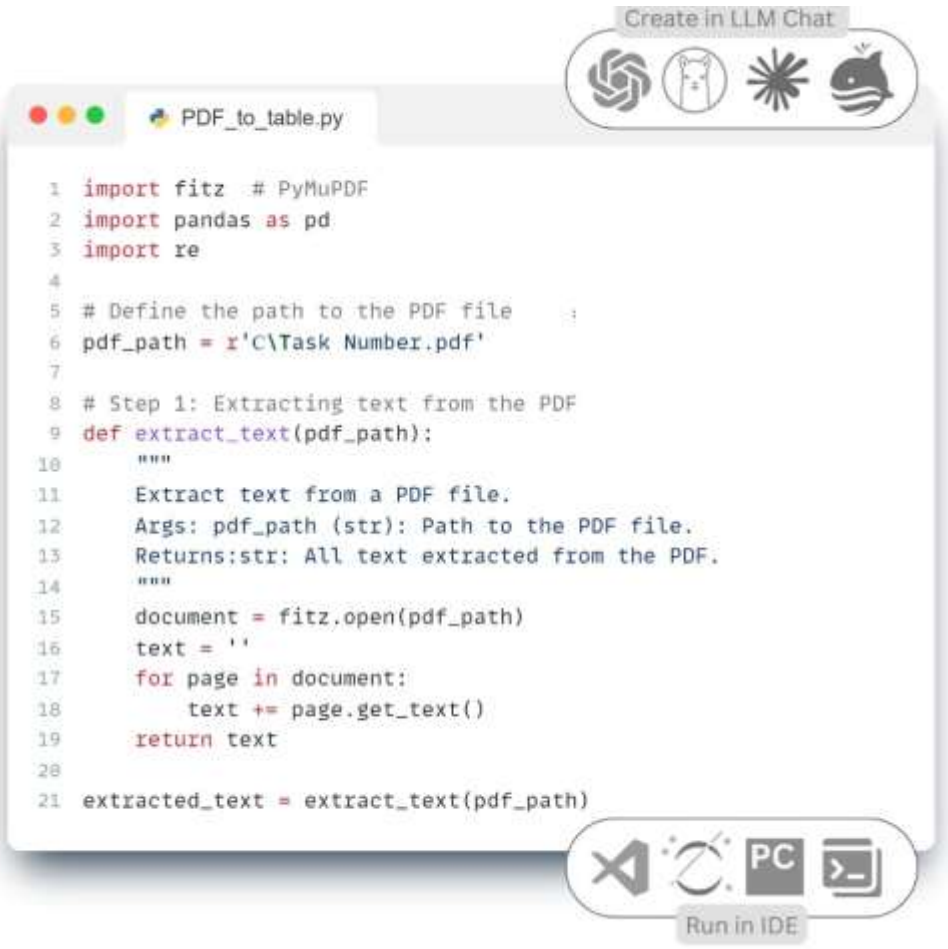


Fig. 4.1-3 La respuesta LLM en forma de código Python y sus bibliotecas y paquetes (Pandas, Fitz) extrae texto de un archivo PDF.

Este código (Fig, 4.1-3) se puede ejecutar offline en uno de los IDEs populares que hemos mencionado anteriormente: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con plugin Python, JupyterLab o herramientas online populares: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- En el paso "Convertir", utilizamos la popular biblioteca Pandas (de la que hablamos en detalle en el capítulo "Python Pandas: una herramienta indispensable para trabajar con datos") para leer el texto extraído en el DataFrame y guardar el DataFrame en un archivo de tabla CSV o XLXS:

> Necesito un código que convierta la tabla resultante de un archivo PDF en un DataFrame. También añadir código para guardar el DataFrame a un archivo CSV ⏎

Respuesta de LLM:

Create in LLM Chat

PDF_to_table.py

```python
# Step 2: Converting the extracted text to a DataFrame
def text_to_df(text):
    """
    Convert extracted text to a pandas DataFrame.
    Args: text (str): Extracted text from a PDF file.
    Returns: pandas.DataFrame: DataFrame representation of the text.
    """
    lines = text.split('\n')
    data = []
    for line in lines:
        if line.strip():  # Check if line is not empty
            data.append(re.split(r'\s{2,}', line))

    # Determine the number of columns based on the longest row
    num_columns = max(len(row) for row in data)

    # Create DataFrame with dynamic columns
    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
    return df

df = text_to_df(extracted_text)

# Save the DataFrame to a CSV file
csv_path = 'C:\extracted_table.csv'
df.to_csv(csv_path, index=False)
```

Run in IDE

Fig. 4.1-4 Conversión de la tabla extraída de PDF a DataFrame y guardado de la tabla en un archivo CSV.

Si se produce un error durante la ejecución del código (Fig. 4.1-3, Fig. 4.1-4) -por ejemplo, debido a la falta de bibliotecas o a una ruta de archivo incorrecta-, basta con copiar el texto del error junto con el código fuente y volver a enviarlo al modelo LLM. El modelo analizará el mensaje de error, explicará cuál es el problema y sugerirá soluciones o pasos adicionales.

> De este modo, la interacción con la IA LLM se convierte en un ciclo completo: →→prueba→ retroalimentación→ corrección - sin necesidad de profundos conocimientos técnicos.

Utilizando una consulta de texto plano en LLM chat y una docena de líneas de Python que podemos ejecutar localmente en cualquier IDE, convertimos un documento PDF -en un formato CSV tabular, que, a diferencia de un documento PDF, es fácilmente legible por máquina y se integra rápidamente en cualquier sistema de gestión de datos.

Podemos aplicar este código (Fig, 4.1-3, Fig. 4.1-4), copiándolo desde cualquier sala de chat LLM, a decenas o miles de nuevos documentos PDF en el servidor, automatizando así el proceso de conversión de un flujo de documentos no estructurados a un formato de tabla CSV estructurada.

Pero los documentos PDF no siempre contienen texto, la mayoría de las veces son documentos escaneados que deben procesarse como imágenes. Aunque las imágenes son intrínsecamente desestructuradas, el desarrollo y la aplicación de bibliotecas de reconocimiento nos permiten extraer, procesar y analizar su contenido, lo que nos permite aprovechar al máximo estos datos en nuestros procesos empresariales.

Convierte imágenes JPEG, PNG a formato estructurado

Las imágenes son una de las formas más comunes de datos no estructurados. En la construcción y muchas otras industrias, se almacena una enorme cantidad de información en forma de documentos escaneados, esquemas, fotografías y dibujos. Estos datos contienen información valiosa pero no pueden procesarse directamente, como una hoja de cálculo Excel o una base de datos. Las imágenes contienen mucha información compleja porque su contenido, colores y texturas son variados y se requiere un tratamiento especial para extraer información útil.

La dificultad de utilizar imágenes como fuente de datos radica en la falta de estructura. Las imágenes no transmiten un significado de forma directa y fácilmente cuantificable que un ordenador pueda entender o procesar inmediatamente, como hace una hoja de cálculo de Excel o una tabla de una base de datos. Para convertir los datos de imágenes no estructurados en una forma estructurada, deben utilizarse bibliotecas especiales capaces de interpretar la información visual que contienen (Fig, 4.1-5).

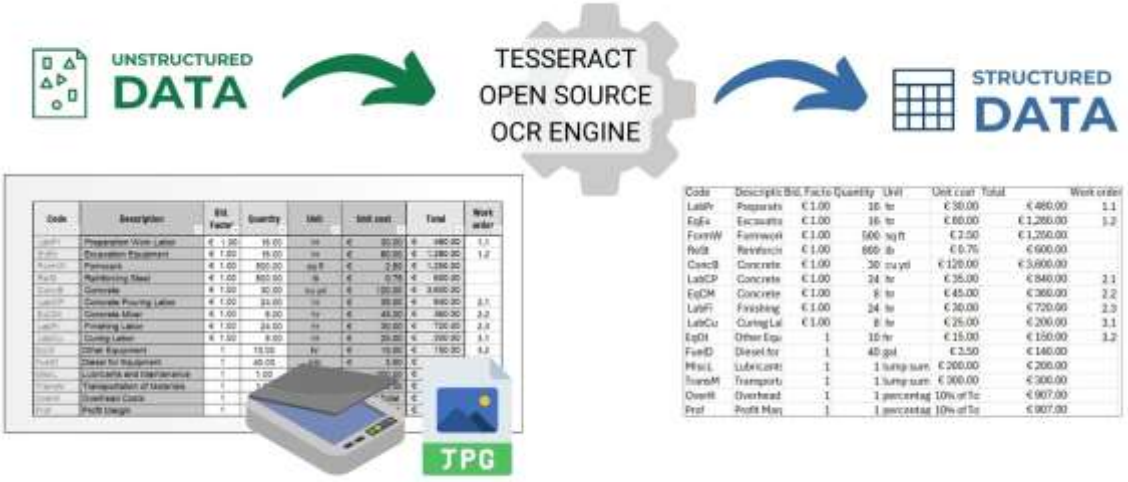


Fig. 4.1-5 La conversión de documentos e imágenes escaneados a formatos estructurados es posible con herramientas especiales de OCR.

La tecnología OCR (Optical Character Recognition) se utiliza para extraer texto de imágenes. Permite reconocer letras y números en documentos escaneados, fotos y archivos PDF, convirtiéndolos en texto editable y legible por máquina. Las tecnologías de OCR se utilizan desde hace tiempo en la automatización de documentos y hoy en día se integran fácilmente en cualquier proceso empresarial y en las aplicaciones Python -applications. Una de las herramientas de OCR más populares es Tesseract, un código abierto, desarrollado originalmente por HP™ y ahora respaldado por Google™. Es compatible con más de 100 idiomas y tiene una gran precisión de reconocimiento.

Pidamos al chat LLM que escriba un código de ejemplo para recuperar datos de una tabla escaneada o fotografiada de forma estructurada.

- Envíe una solicitud de texto al chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otro):

> Escriba código para convertir una imagen JPEG que contenga una tabla en una tabla DataFrame ⏎

- La respuesta de LLM sugerirá en la mayoría de los casos el uso de la biblioteca Pytesseract para reconocer texto en imágenes:

Create in LLM Chat

group.py

```python
import pytesseract
from PIL import Image
import pandas as pd

# Load the image from file
image_path = '/mnt/data/ConstructiontDDC,
image = Image.open(image_path)

# Use Tesseract to convert the image to text
text = pytesseract.image_to_string(image, lang='eng')

# Split the text into lines and then columns
rows = text.split('\n')
data = [row.split('\t') for row in rows if row.strip()]

# Create a DataFrame using pandas
# Assuming the first row is the header
df = pd.DataFrame(data[1:], columns=data[0])

# Display the DataFrame
df.head()  # Show only the first few rows for brevity
```

Run in IDE

Fig. 4.1-6 Convierte el texto extraído de una tabla de imágenes o fotos en una representación tabular estructurada.

En este ejemplo - el código (Fig, 4.1-6) derivado del LLM, utiliza la biblioteca pytesseract (Tesseract para Python) para convertir una imagen en texto utilizando OCR (reconocimiento óptico de caracteres) y la biblioteca Pandas para convertir este texto en una forma estructurada, es decir, un DataFrame.

> El proceso de conversión suele implicar un tratamiento previo para mejorar la calidad de la imagen, tras lo cual se aplican diversos algoritmos para la detección de patrones, la extracción de características o el reconocimiento de objetos. Como resultado, la información visual no estructurada se convierte en datos estructurados.

Aunque los PDF y las imágenes son fuentes clave de información no estructurada, el verdadero campeón en términos de volumen es el texto generado en correos electrónicos, chats, reuniones, mensajeros. Estos datos no sólo son numerosos, sino que están dispersos, informalizados y muy poco estructurados.

Convertir datos de texto en un formato estructurado

Además de los documentos PDF con tablas (Fig. 4.1-2) y las versiones escaneadas de formularios tabulares (Fig. 4.1-5), una parte importante de la información de la documentación del proyecto se presenta en forma de texto. Puede tratarse tanto de frases coherentes en documentos de texto como de registros fragmentarios dispersos en dibujos y esquemas. En las condiciones modernas de tratamiento de datos, una de las tareas más habituales es convertir ese texto en un formato estructurado adecuado para el análisis, la visualización y la toma de decisiones.

Para ello es fundamental la taxonomía, un sistema de clasificación que permite organizar la información en categorías y subcategorías basadas en características comunes.

> **Una taxonomía** es una estructura de clasificación jerárquica utilizada para agrupar y organizar objetos. En el contexto del tratamiento de textos, proporciona la base para clasificar sistemáticamente los elementos en categorías semánticas, lo que simplifica el análisis y mejora la calidad del tratamiento de datos.

La creación de taxonomías va acompañada de las etapas de extracción de entidades, categorización y contextualización. Para modelizar el proceso de extracción de información a partir de datos textuales, los pasos siguientes son similares a los que ya hemos aplicado a la estructuración de datos a partir de documentos PDF:

- Extraer): necesita analizar datos de texto para extraer información sobre retrasos y cambios en el calendario del proyecto.
- **Categorización y clasificación** (Transofrm): categoriza la información recibida, por ejemplo, los motivos de los retrasos y los cambios de horario.
- **Integración** (Load): al final preparamos los datos estructurados para su integración en sistemas externos de gestión de datos.

Pongámonos en situación: tenemos un diálogo entre un jefe de proyecto y un ingeniero en el que se discuten problemas con el calendario. Nuestro objetivo es extraer los elementos clave (motivos del retraso, ajustes del calendario) y presentarlos de forma estructurada (Fig, 4.1-7).

Vamos a realizar la extracción basándonos en las palabras clave esperadas, crearemos un DataFrame para simular la extracción de datos y tras la transformación una nueva tabla DataFrame que contendrá columnas para la fecha, el evento (por ejemplo, motivo del retraso) y la acción (por ejemplo, cambio de horario).

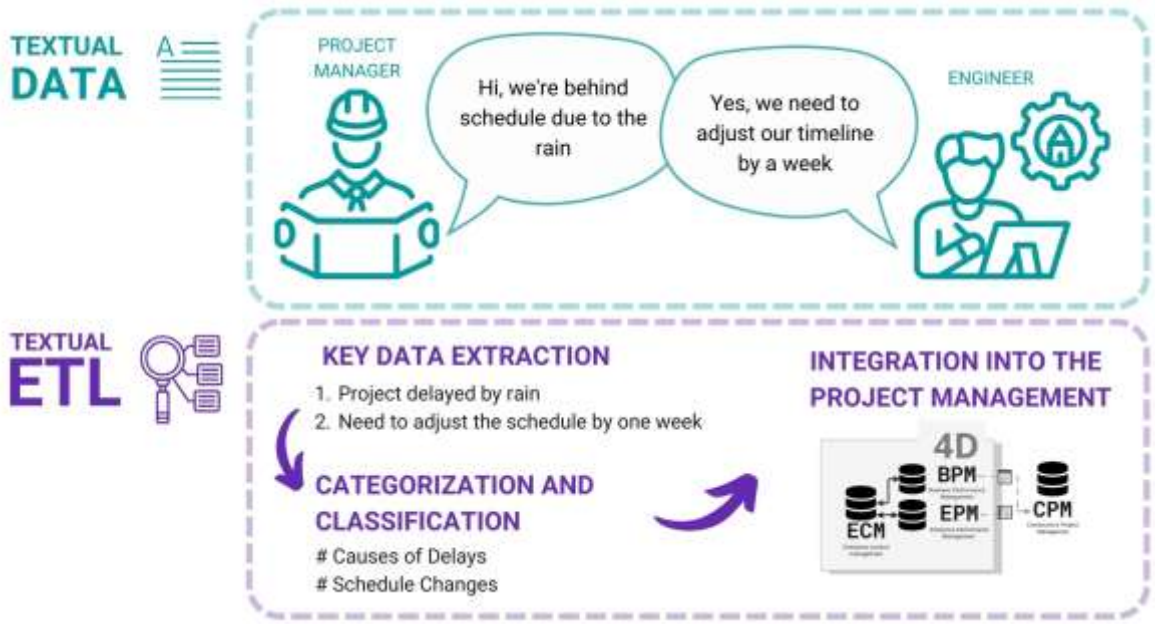


Fig. 4.1-7 Resaltar la información clave del texto sobre la necesidad de ajustar los plazos e integrar los cambios en el sistema de gestión del proyecto.

A continuación se muestra el código para resolver el problema utilizando una consulta de texto en uno de los modelos lingüísticos como en los ejemplos anteriores.

- Envía una solicitud de texto a cualquier sala de chat de LLM:

> Tengo una conversación entre un gerente: "Hola, llevamos retraso debido a la lluvia" y un ingeniero: "Sí, tenemos que ajustar el plazo una semana". Necesito un script que analice futuros diálogos de texto similares, extraiga de ellos los motivos de los retrasos y los ajustes de plazos necesarios y, a continuación, genere un DataFrame a partir de estos datos. El DataFrame debe guardarse en un archivo CSV. ⏎

- La respuesta del LLM suele incluir código Python que utiliza expresiones regulares (re - Regex) y la biblioteca Pandas (pd):

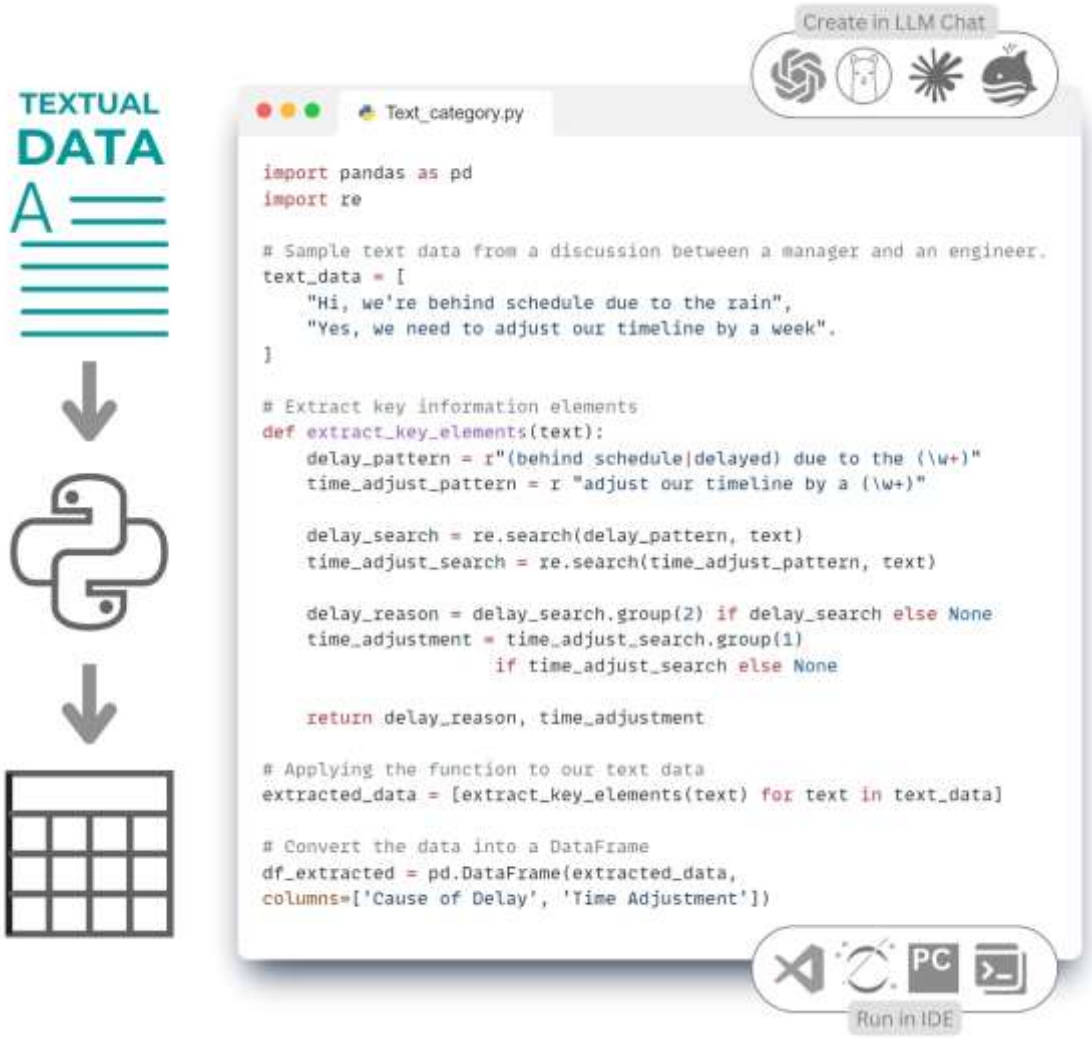
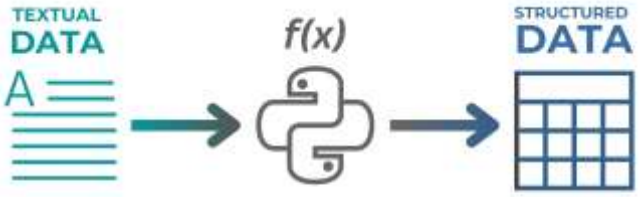


Fig. 4.1-8 Resaltar la información clave del texto sobre la necesidad de ajustar los plazos en una tabla.

En este ejemplo (Fig, 4.1-7), se analizan datos de texto que contienen correspondencia entre un director de proyecto y un ingeniero para identificar y extraer información específica que pueda afectar a la gestión de futuros proyectos con diálogos similares. Utilizando expresiones regulares (hablaremos más sobre expresiones regulares en el capítulo "Requisitos estructurados y expresiones regulares RegEx "), se identifican mediante patrones las causas de los retrasos del proyecto y los ajustes necesarios en el calendario. La función escrita en este ejemplo extrae la causa del retraso o el ajuste de tiempo de las cadenas basándose en los patrones: resaltando la palabra después de "debido a" como la causa del retraso o la palabra después de "por" como el ajuste de tiempo.

Si una fila menciona un retraso debido a las condiciones meteorológicas, se identifica "lluvia" como la

causa; si una fila menciona un ajuste de la programación para un periodo determinado, ese periodo se extrae como un ajuste de tiempo (Fig, 4.1-9). La ausencia de cualquiera de estas palabras en una fila da lugar a un valor de "Ninguno" para el atributo-columna correspondiente.



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Fig. 4.1-9 La tabla resumen obtenida como DataFrame, tras la ejecución del código, contiene información sobre la existencia de retrasos y los ajustes de tiempo necesarios.

La estructuración y parametrización de las condiciones a partir del texto (diálogo, carta, documento) permite eliminar rápidamente los retrasos en la construcción: por ejemplo, la falta de trabajadores puede afectar al ritmo de trabajo con mal tiempo, por lo que las empresas, conociendo de antemano los parámetros de retraso a partir de los diálogos (Fig. 4.1-9) entre el capataz de la obra y el director del proyecto, pueden reforzar la cuadrilla en caso de previsión desfavorable.

La conversión de documentos e imágenes a un formato estructurado puede realizarse mediante herramientas de categorización relativamente sencillas, abiertas y gratuitas.

La categorización de los elementos también es una parte clave del trabajo con los datos del proyecto, especialmente en el contexto del uso de software CAD (BIM).

Conversión de datos CAD (BIM) en un formato estructurado

La estructuración y categorización de los datos CAD (BIM) es una tarea más compleja porque los datos almacenados de las bases de datos CAD (BIM) casi siempre están en formatos paramétricos cerrados o complejos, y a menudo combinan elementos de datos geométricos (semiestructurados) y elementos de metainformación (datos semiestructurados o estructurados) al mismo tiempo.

Los formatos de datos nativos de los sistemas CAD (BIM) suelen estar protegidos y ser inaccesibles para su uso directo, a menos que se trate de software especializado o API - interfaces del propio desarrollador (Fig. 4.1-10). Este aislamiento de los datos forma silos de almacenamiento cerrados que limitan el libre intercambio de información e inhiben la creación de procesos digitales integrales en la empresa.

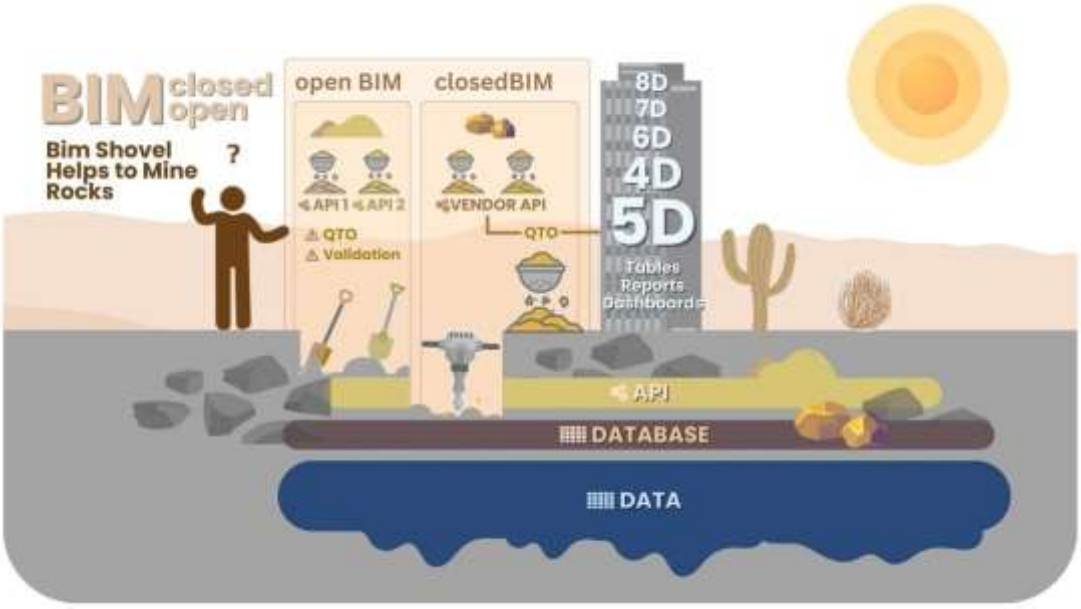


Fig. 4.1-10 Los profesionales de CAD (BIM) pueden acceder a los datos nativos a través de API -conexiones o herramientas de proveedores.

En los formatos especiales de CAD (BIM), la información sobre las características y atributos de los elementos del proyecto se recoge en un sistema de clasificación jerárquica, en el que las entidades con propiedades correspondientes se sitúan, como la fruta de un árbol frutal, en los nodos más recientes de las ramas de clasificación de datos (Fig. 4.1-11).

La extracción de datos de tales jerarquías es posible de dos maneras: manualmente, haciendo clic en cada nodo, como si se tratara de un árbol, cortando con un hacha las ramas seleccionadas de categorías y tipos. Por otro lado, el uso de interfaces de programación de aplicaciones (API) proporciona un enfoque más eficiente y automatizado para recuperar y agrupar datos, transformándolos finalmente en una tabla estructurada para su uso en otros sistemas.

Para extraer tablas de datos estructurados de proyectos CAD (BIM) se pueden utilizar diferentes herramientas como Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC, o soluciones de código abierto, como IfcOpSh o IFCjs para el formato IFC.

Las modernas herramientas de exportación y conversión de datos permiten, con el fin de simplificar el tratamiento y la preparación de los datos, separar el contenido de los modelos CAD en dos componentes clave: la información geométrica y los datos de atributos (Fig. 4.1-13) - metainformación que describe las propiedades de los elementos de diseño (Fig. 3.1-16). Estas dos capas de datos permanecen vinculadas mediante identificadores únicos, gracias a los cuales es posible asignar con precisión cada elemento con descripción geométrica (mediante parámetros o polígonos) a sus atributos: nombre, material, fase de realización, coste, etc. Este enfoque garantiza la integridad del modelo y permite un uso flexible de los datos tanto para visualización (datos del modelo geométrico) como para tareas analíticas o de gestión (estructurados o no estructurados), trabajando con los dos tipos de datos por separado o en paralelo.

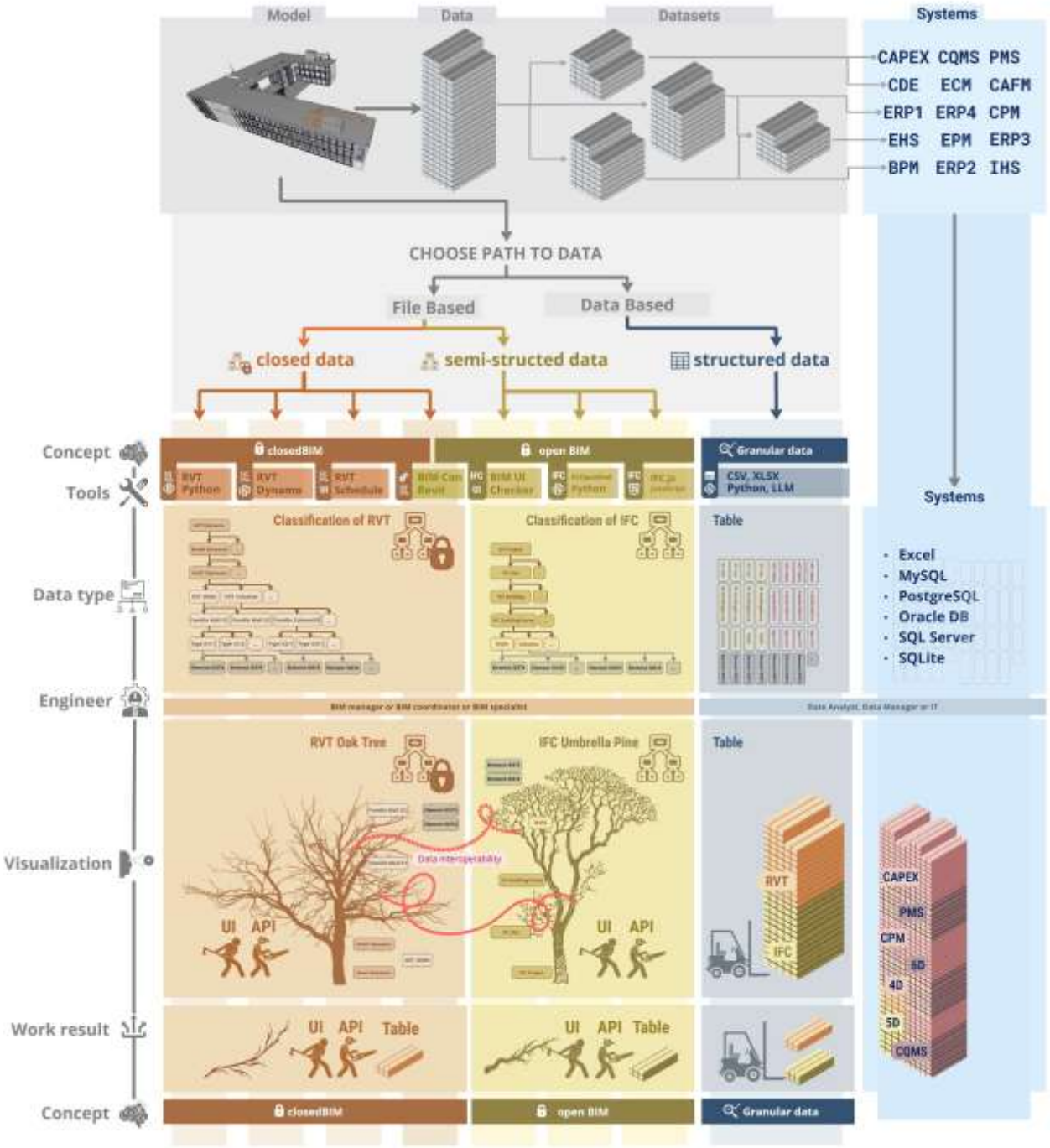


Fig. 4.1-11 La visión de la información de las bases de datos CAD (BIM) se presenta al usuario en forma de árboles de clasificación.

Con el desarrollo de las tecnologías de ingeniería inversa y la llegada de los SDK (Software Development Kit) para la conversión de datos CAD, la disponibilidad y conversión de datos de formatos cerrados de software CAD (BIM) se ha vuelto mucho más fácil. Ahora es posible convertir de forma legal y segura datos de formatos cerrados a formatos universales aptos para su análisis y uso en otros sistemas. La historia de las primeras herramientas de ingeniería inversa ("Open DWG") y la lucha por el dominio de los formatos de

los proveedores de CAD se trataron en el capítulo "Datos estructurados: la base de la transformación digital".

> Las herramientas de ingeniería inversa permiten recuperar legítimamente los datos de formatos propietarios cerrados, descomponiendo la información del formato mixto CAD (BIM) en los tipos de datos y formatos requeridos por el usuario, lo que facilita su tratamiento y análisis.

Gracias a la ingeniería inversa y al acceso directo a la información de las bases de datos CADhace que la información sea accesible, permitiendo datos abiertos y herramientas abiertas, así como el análisis de datos mediante herramientas estándar, la creación de informes, las visualizaciones y la integración con otros sistemas digitales (Fig. 4.1-12).

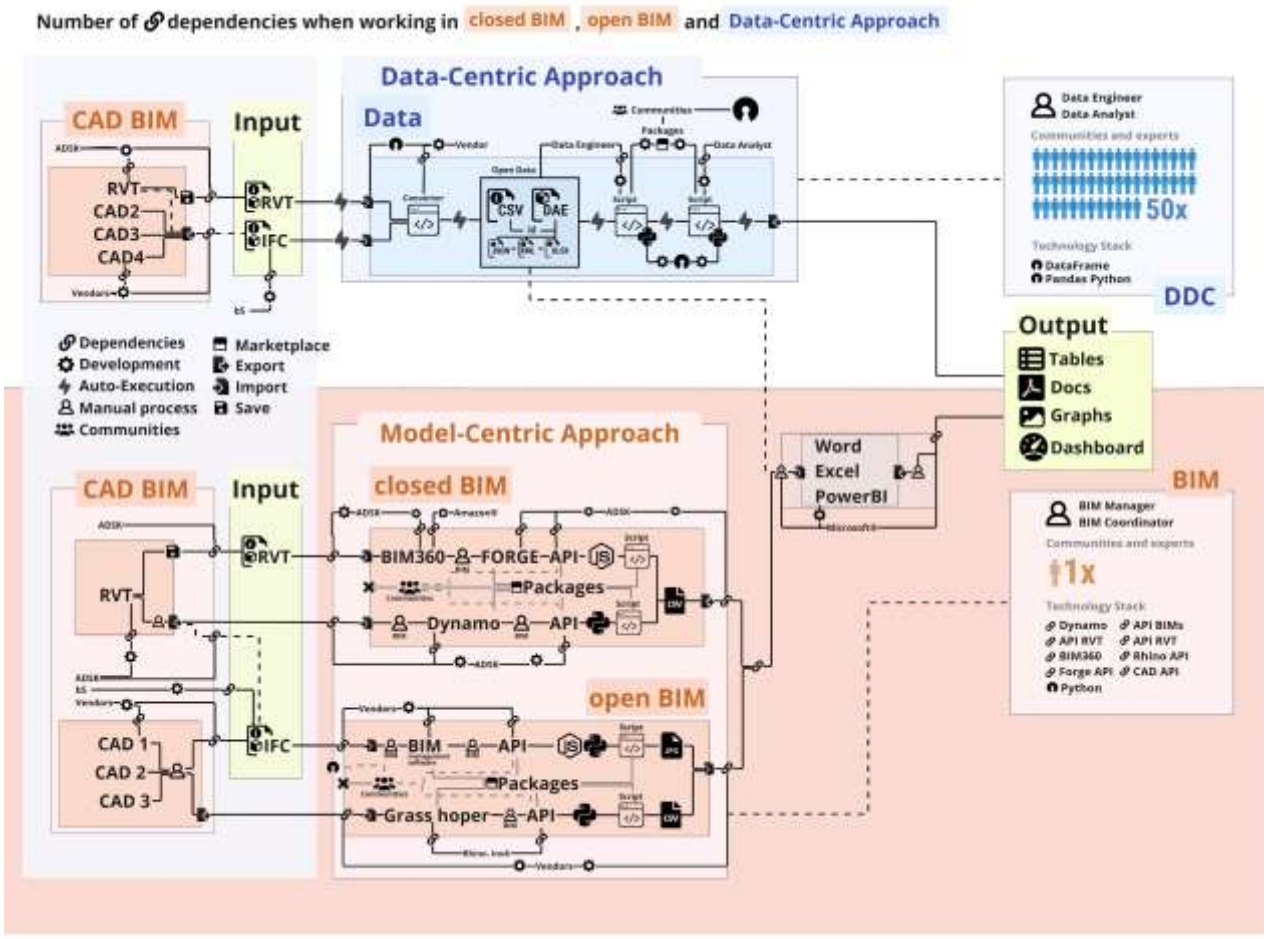


Fig. 4.1-12 El acceso directo a los datos CAD minimiza las dependencias de las plataformas de software y avanza hacia un enfoque centrado en los datos.

Desde 1996 para el formato DWG, desde 2008 para el formato DGN y desde 2018 para RVT ha sido posible convertir formatos de datos CAD inicialmente cerrados en cualquier otro formato, incluyendo formatos estructurados, de forma cómoda y eficiente con la ayuda de herramientas de ingeniería inversa (Fig. 4.1-13).

Hoy en día, casi todas las grandes empresas de CAD (BIM) y grandes empresas de ingeniería del mundo utilizan SDKs - herramientas de ingeniería inversa para extraer datos de formatos cerrados de proveedores de CAD (BIM) [92].

Fig. 4.1-13 El uso de herramientas de ingeniería inversa permite convertir las bases de datos de los programas CAD (BIM) en cualquier modelo de datos conveniente.

Convertir los datos de formatos cerrados y propietarios a formatos abiertos y separar los formatos mixtos CAD (BIM) en datos geométricos y de atributos de metainformación simplifica el proceso de trabajo con ellos, haciéndolos disponibles para el análisis, la manipulación y la integración con otros sistemas (Fig. 4.1-14).

> En el trabajo actual con datos CAD (BIM), hemos llegado a un punto en el que no necesitamos pedir permiso a los proveedores de CAD (BIM) para acceder a la información de los formatos CAD.

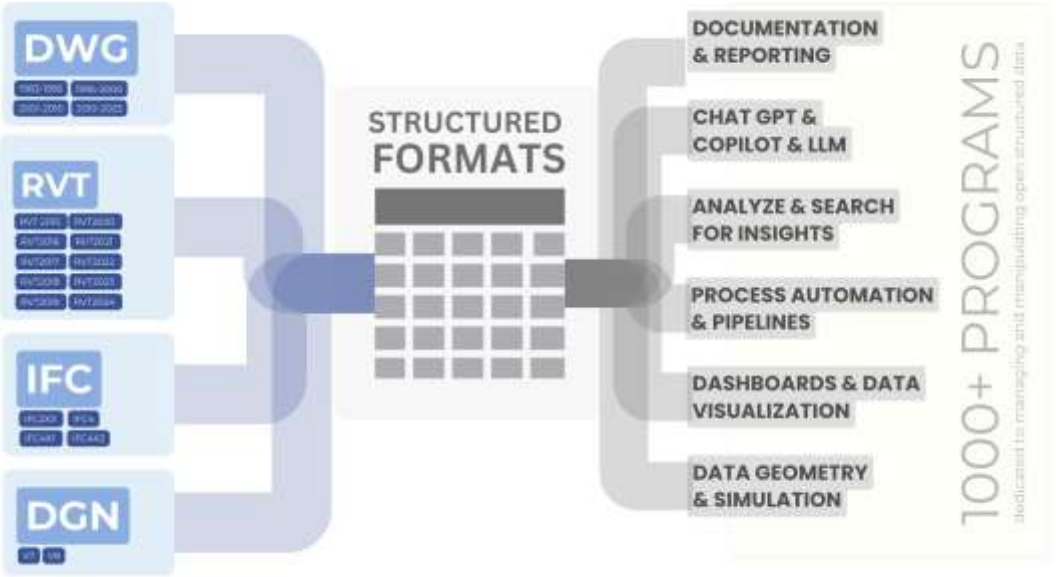


Fig. 4.1-14 Las modernas herramientas SDK permiten la conversión legal de datos de formatos de bases de datos CAD propietarios (BIM).

Las tendencias actuales en el tratamiento de datos de diseño CAD siguen estando marcadas por los principales agentes del mercado -los proveedores de CAD -, que trabajan para reforzar su posición en el mundo de los datos y crear nuevos formatos y conceptos.

CAD Los proveedores de soluciones se orientan hacia los datos estructurados

A partir de 2024, el sector del diseño y la construcción experimentará un importante cambio tecnológico en el uso y tratamiento de datos. En lugar del libre acceso a los datos de diseño, los proveedores de sistemas CAD se están centrando en promover los próximos nuevos conceptos. Enfoques como BIM (creado en 2002) y BIM abierto (creado en 2012) están dando paso gradualmente a soluciones tecnológicas modernas que los vendedores de CAD están empezando a promover [93]:

- Avanzar hacia el uso de datos "granulares" que permitan una gestión eficaz de la información y una evolución hacia el análisis de datos.
- Surgimiento del formato USD y aplicación del enfoque Entidad-componente-sistema (ECS) para una organización flexible de los datos.
- Uso activo de la inteligencia artificial en el tratamiento de datos, la automatización de procesos y el análisis de datos
- Desarrollo de la interoperabilidad - mejora de la interacción entre distintos programas, sistemas y bases de datos.

Cada uno de estos aspectos se tratará con más detalle en la sexta parte del libro "CAD y BIM: marketing, realidad y futuro de los datos de diseño en la construcción". En este capítulo sólo esbozaremos brevemente el vector general del cambio: los principales proveedores de CAD intentan ahora replantearse la forma en que se estructura la información de diseño. Uno de los cambios clave es el abandono del modelo clásico de almacenamiento basado en archivos en favor de una arquitectura de datos granular y orientada al análisis que proporcione un acceso continuo a los componentes individuales del modelo [93].

> La esencia de lo que está ocurriendo es que la industria está abandonando los formatos engorrosos, especializados y paramétricos que requieren núcleos geométricos en favor de soluciones más universales, legibles por máquina y flexibles.

Uno de estos motores del cambio es el formato USD (Universal Scene Description), desarrollado originalmente en la industria de los gráficos por ordenador, pero ya reconocido en aplicaciones de ingeniería gracias al desarrollo de la plataforma NVIDIA Omniverse (e Isaac Sim) para simulaciones y visualizaciones [93]. A diferencia del IFC paramétrico, USD ofrece una estructura más sencilla y permite describir la geometría y las propiedades de los objetos en formato JSON (Fig, 4.1-15), lo que facilita el procesamiento de la información y agiliza su integración en los procesos digitales. El nuevo formato permite almacenar la geometría (además de BREP -NURBS - más detalles en la Parte 6 del libro) en forma de polígonos MESH, y las propiedades de los objetos en JSON, lo que lo hace más conveniente para los procesos automatizados y el trabajo en ecosistemas en la nube [94].

Algunos proveedores de CAD y ERP ya utilizan formatos similares (por ejemplo, NWD, SVF, CP2, CPIXML),

pero la mayoría siguen siendo cerrados y no están disponibles para uso externo, lo que limita las posibilidades de integración y reutilización de datos. En este contexto, USD puede desempeñar el mismo papel que desempeñó DXF en su momento: una alternativa abierta a formatos propietarios como DWG.

General Information			Comparison / Notes
Year of format creation	1991	2016	IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar	IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF	IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK	ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK	ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD	Different alliances for each format
Year of Alliance Formation	1994	2023	The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co	ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage			Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification	IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing	IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction	USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera	USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

Fig. 4.1-15 USD formato como un intento de los proveedores de CAD de satisfacer la demanda de interoperabilidad e independencia de los datos de diseño de los núcleos geométricos.

La transición de los principales desarrolladores a los formatos abiertos y simplificados USD, GLTF, OBJ, XML (NWD cerrado, CP2, SVF, SVF2, CPIXML) y otros similares (Fig. 3.1-17) refleja la tendencia mundial y la demanda del sector de simplificar los datos y aumentar su accesibilidad. En los próximos años, cabe esperar un alejamiento gradual de las normas y formatos paramétricos complejos con dependencia de núcleos geométricos en favor de soluciones más ligeras y estructuradas. Esta transición acelerará la digitalización del sector de la construcción, facilitará la automatización de procesos y simplificará el intercambio de datos.

A pesar de los planes estratégicos de los proveedores de CAD para promover nuevos formatos abiertos,

los profesionales del sector de la construcción también pueden obtener pleno acceso a los datos de sistemas CAD cerrados, sin necesidad de herramientas CAD (BIM), utilizando herramientas de ingeniería inversa.

> Todas estas tendencias conducen inevitablemente a pasar de modelos 3D voluminosos y monolíticos a datos universales estructurados y al uso de formatos que llevan mucho tiempo demostrando su eficacia en otros sectores. Una vez que los equipos de proyecto empiezan a ver los modelos CAD no sólo como objetos visuales o un conjunto de archivos, sino como bases de datos que contienen conocimientos e información, el planteamiento del diseño y la gestión cambia radicalmente.

Una vez que los equipos han aprendido a extraer datos estructurados de documentos, textos, dibujos y modelos CAD, y tienen acceso a bases de datos, el siguiente paso clave es el modelado de datos y la garantía de calidad. Este paso es el que determina en gran medida la velocidad de procesamiento y transformación de la información que, en última instancia, se utilizará para tomar decisiones en tareas de aplicación específicas.

CAPÍTULO 4.2.
CLASIFICACIÓN E INTEGRACIÓN: UN LENGUAJE COMÚN PARA LOS DATOS DE CONSTRUCCIÓN

La rapidez en la toma de decisiones depende de la calidad de los datos

La arquitectura actual de los datos de diseño está experimentando cambios fundamentales. El sector está pasando de modelos voluminosos y aislados y formatos cerrados a estructuras más flexibles y legibles por máquinas, centradas en el análisis, la integración y la automatización de procesos. Sin embargo, la transición a nuevos formatos no garantiza por sí sola la eficiencia: la calidad de los propios datos es inevitablemente el centro de atención.

> En las páginas de este libro hablamos mucho de formatos, sistemas y procesos. Pero todos estos esfuerzos carecen de sentido sin un elemento clave: datos en los que se pueda confiar. La calidad de los datos es la piedra angular de la digitalización, sobre la que volveremos a lo largo de las siguientes partes.

Las empresas de construcción modernas, sobre todo las grandes, utilizan docenas y a veces miles de sistemas y bases de datos diferentes (Fig. 4.2-1). Estos sistemas no sólo deben llenarse de información nueva con regularidad, sino que también deben interactuar eficazmente entre sí. Todos los nuevos datos generados como resultado del procesamiento de la información entrante se integran en estos entornos y sirven para resolver tareas empresariales específicas.

Y mientras que en el pasado las decisiones sobre tareas empresariales específicas las tomaban los altos directivos -los llamados HiPPO (Fig. 2.1-9)- basándose en la experiencia y la intuición, hoy, con el fuerte aumento del volumen de información, este enfoque se está volviendo controvertido. La analítica automatizada, que trabaja con datos en tiempo real, lo está sustituyendo.

> Los debates sobre los procesos empresariales "tradicionales-manuales" a nivel ejecutivo se desplazarán hacia la analítica operativa, que requiere respuestas rápidas a las consultas empresariales.

La época en que contables, capataces y estimadores generaban manualmente informes y cuadros resumen y mostraban los datos de los proyectos a lo largo de días y semanas es cosa del pasado. Hoy en día, la rapidez y la puntualidad en la toma de decisiones se están convirtiendo en un factor clave de la ventaja competitiva.

Fig. 4.2-1 La industria de la construcción tarda días en calcular y tomar decisiones, a diferencia de otras industrias en las que esto ocurre en horas o minutos.

La principal diferencia entre el sector de la construcción y otras industrias más avanzadas digitalmente (Fig, 4.2-1) es el bajo nivel de calidad y normalización de los datos. Los enfoques anticuados de la generación, transmisión y procesamiento de la información ralentizan los procesos y crean caos. La falta de normas uniformes de calidad de los datos dificulta la implantación de la automatización de extremo a extremo

> Uno de los principales retos sigue siendo la escasa calidad de los datos de entrada, así como la falta de procesos formalizados para su preparación y validación. Sin datos fiables y coherentes, no puede haber una integración eficaz entre sistemas. Esto provoca retrasos, errores y mayores costes en todas las fases del ciclo de vida del proyecto.

En las siguientes secciones del libro, analizamos en detalle cómo puede mejorar la calidad de los datos, normalizar los procesos y acortar el camino desde la información hasta los datos de calidad, validados y coherentes.

Normalización e integración de datos

Una gestión de datos eficaz requiere una estrategia de normalización clara. Sólo con unos requisitos claros en cuanto a estructura y calidad de los datos se puede automatizar su validación, reducir las operaciones manuales y acelerar la toma de decisiones informadas en todas las fases de un proyecto.

En la práctica diaria, una empresa de construcción tiene que procesar cientos de archivos cada día: correos electrónicos, documentos PDF, archivos de diseño CAD, datos procedentes de sensores IoT, que deben integrarse en los procesos empresariales de la empresa.

El bosque del ecosistema de bases de datos y herramientas de una empresa (Fig, 4.2-2) debe aprender a extraer nutrientes de los datos multiformato entrantes para lograr los resultados que desea la empresa.

> Para gestionar eficazmente el flujo de datos, no hace falta necesariamente contratar a un ejército de gestores, primero hay que desarrollar requisitos y normas estrictos para los datos y utilizar herramientas adecuadas para validarlos, armonizarlos y procesarlos automáticamente.

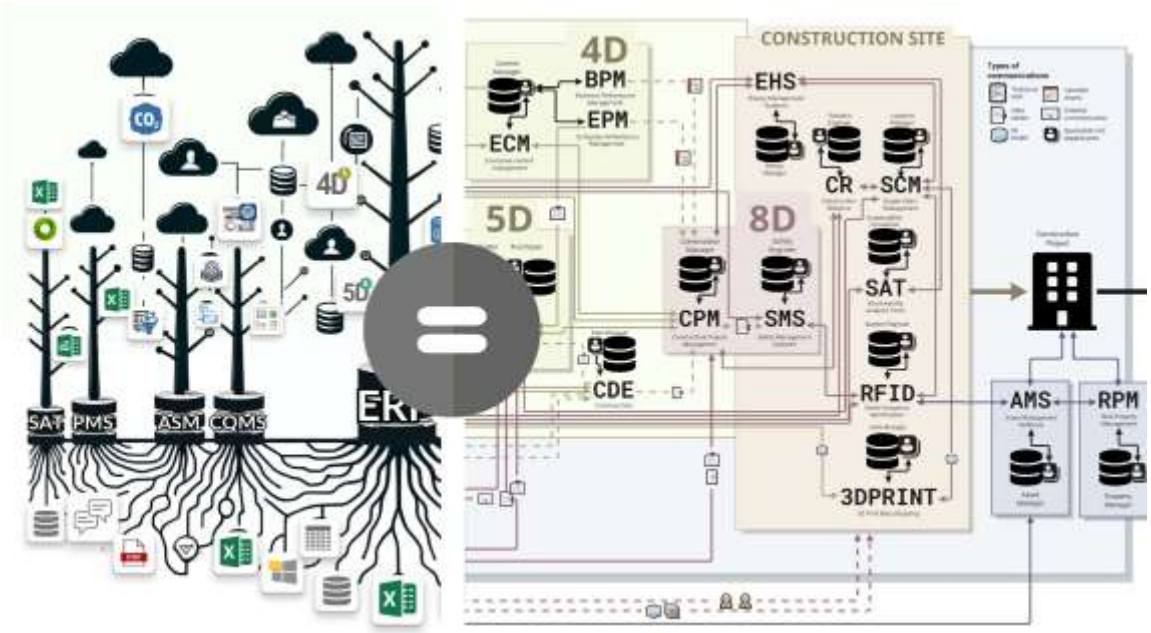


Fig. 4.2-2 Garantizar la salud del ecosistema de la empresa requiere una dotación de recursos de calidad y oportuna de sus sistemas.

Con el fin de automatizar el proceso de validación y armonización de datos (para su posterior integración automática) hay que empezar por describir los requisitos mínimos de datos necesarios para cada sistema específico. Estos requisitos se definen en:

- ¿Qué necesita conseguir exactamente?
- ¿De qué forma (estructura, formato)?
- ¿Qué atributos son obligatorios?
- ¿Qué tolerancias de precisión e integridad son aceptables?

Los requisitos de los datos describen los criterios de calidad, estructura e integridad de la información recibida y procesada. Por ejemplo, en el caso de los textos de los documentos PDF, es importante garantizar un formato preciso conforme a las normas del sector (Fig. 7.2-14 - Fig. 7.2-16). Los objetos de los modelos CAD deben tener atributos correctos (dimensiones, códigos, enlaces a clasificadores) (Fig. 7.3-9, Fig. 7.3-10). Y para los escaneados de contratos, son importantes unas fechas claras y la capacidad de extraer automáticamente el importe y los términos clave (Fig. 4.1-7 - Fig. 4.1-10).

> Formular los requisitos de los datos y comprobar automáticamente su cumplimiento es uno de los pasos que más tiempo lleva, pero también uno de los más importantes. Es el paso que más tiempo consume en los procesos empresariales.

Como se menciona en la Parte 3 de este libro, entre el 50% y el 90% del tiempo de los profesionales de inteligencia de negocio (BI) se dedica a la preparación de los datos y no a su análisis (Fig, 3.2-5). Este proceso incluye la recopilación, verificación, validación, armonización y estructuración de los datos.

Según una encuesta de 2016 [95], los científicos de datos de una gran variedad de campos de amplio espectro afirmaron que dedican la mayor parte de su tiempo de trabajo (alrededor del 80%) a hacer lo que menos les gusta (Fig, 4.2-3): recopilar conjuntos de datos existentes y organizarlos (unificarlos, estructurarlos). Así, menos del 20% de su tiempo queda para tareas creativas, como encontrar patrones y regularidades que conduzcan a nuevas percepciones y descubrimientos.

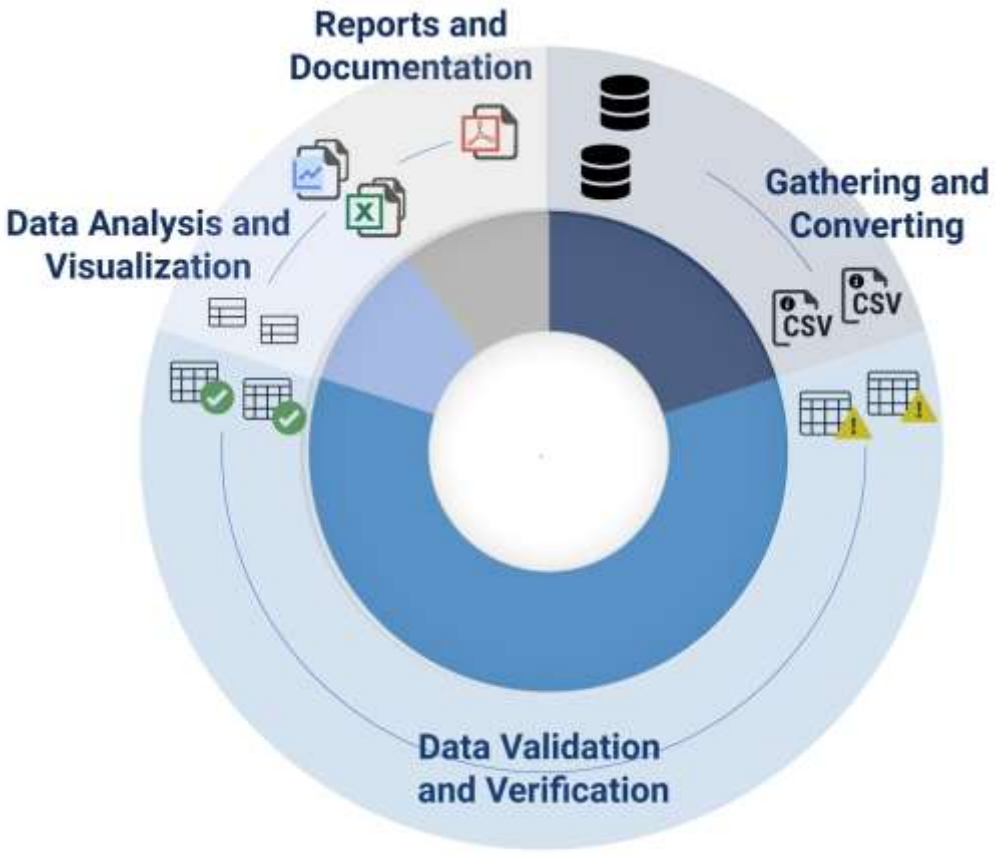


Fig. 4.2-3 Verificar y garantizar la calidad de los datos es el paso más costoso, largo y complejo en la preparación de los datos para su integración en otros sistemas.

El éxito de la gestión de datos en una empresa de construcción requiere un enfoque global que incluya la parametrización de las tareas, la formulación de requisitos de calidad de los datos y el uso de herramientas adecuadas para la validación automatizada.

La interoperabilidad digital empieza por los requisitos

A medida que aumenta el número de sistemas digitales en las empresas, también lo hace la necesidad de coherencia de datos entre ellos. Los directivos responsables de los distintos sistemas informáticos se ven a menudo incapaces de seguir el ritmo del creciente volumen de información y la variedad de formatos. En tales circunstancias, se ven obligados a pedir a especialistas que creen datos en un formato adecuado para su uso en otras aplicaciones y plataformas.

Esto, a su vez, obliga a los ingenieros y al personal de generación de datos a adaptarse a multitud de requisitos, a menudo sin transparencia ni una comprensión clara de dónde y cómo se utilizarán los datos en el futuro. La falta de enfoques normalizados para tratar la información genera ineficacia y mayores costes durante la fase de verificación, que suele ser manual debido a la complejidad y la naturaleza no normalizada de los datos.

> El problema de la normalización de datos no es sólo una cuestión de comodidad o automatización. Se trata de una pérdida financiera directa. Según un informe de IBM de 2016, las pérdidas anuales derivadas de la mala calidad de los datos de en Estados Unidos ascienden a 3,1 billones de dólares[96]. Además, estudios realizados por el MIT y otras consultoras analíticas muestran que el coste de una mala calidad de los datos de puede llegar al 15-25% de los ingresos de una empresa [97].

En estas condiciones, resulta fundamental contar con requisitos de datos claramente definidos y descripciones de qué parámetros, en qué formato y con qué nivel de detalle deben incluirse en los objetos creados. Sin formalizar estos requisitos, es imposible garantizar la calidad de los datos y la compatibilidad entre sistemas y fases del proyecto (Fig. 4.2-4).

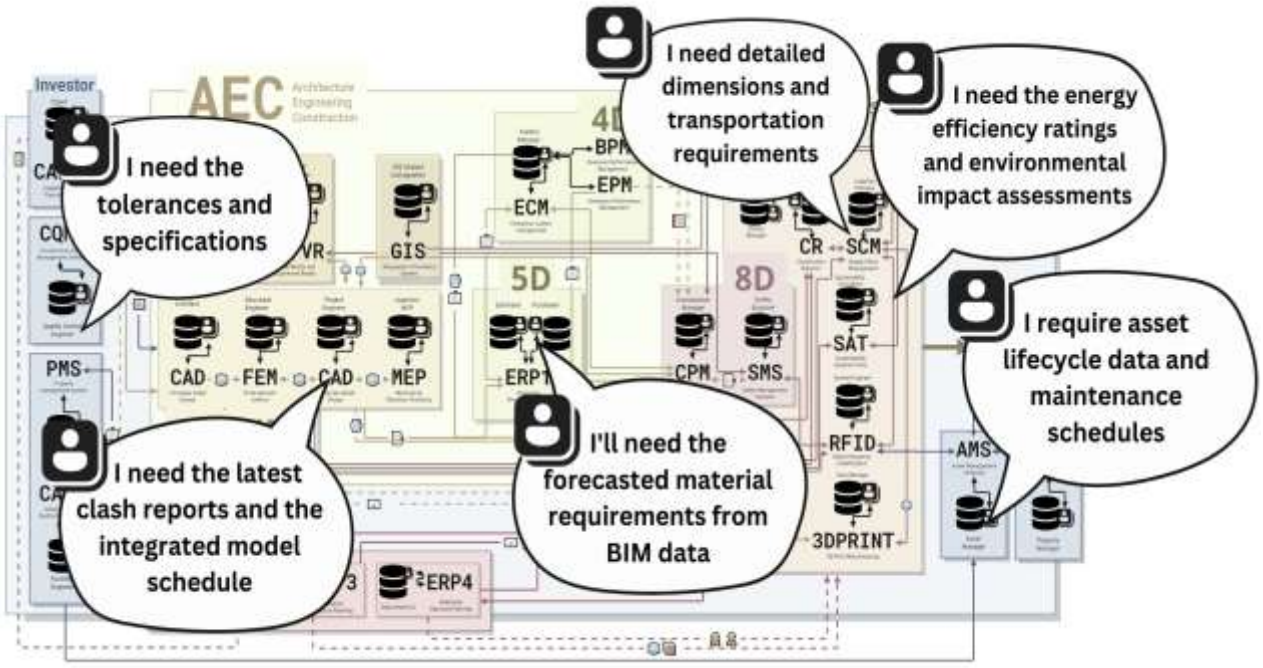


Fig. 4.2-4 Un negocio se basa en la interacción de diferentes roles, cada uno de los cuales requiere ciertos parámetros y valores que son críticos para cumplir los objetivos empresariales.

Para formular los requisitos de datos correctos, es necesario comprender los procesos empresariales a nivel de datos. Los proyectos de construcción varían en tipo, escala y número de participantes, y cada sistema, ya sea de modelado (CAD (BIM)), programación (ERP 4D), cálculo de costes (ERP 5D) o logística (SCM), requiere sus propios parámetros exclusivos para las entradas (elementos-entidad de entrada).

En función de estas necesidades, los responsables de la empresa deben diseñar nuevas estructuras de datos para satisfacer los requisitos o adaptar las tablas y bases de datos existentes. La calidad de los datos creados dependerá directamente de la precisión y corrección con que se formulen los requisitos (Fig. 4.2-5).

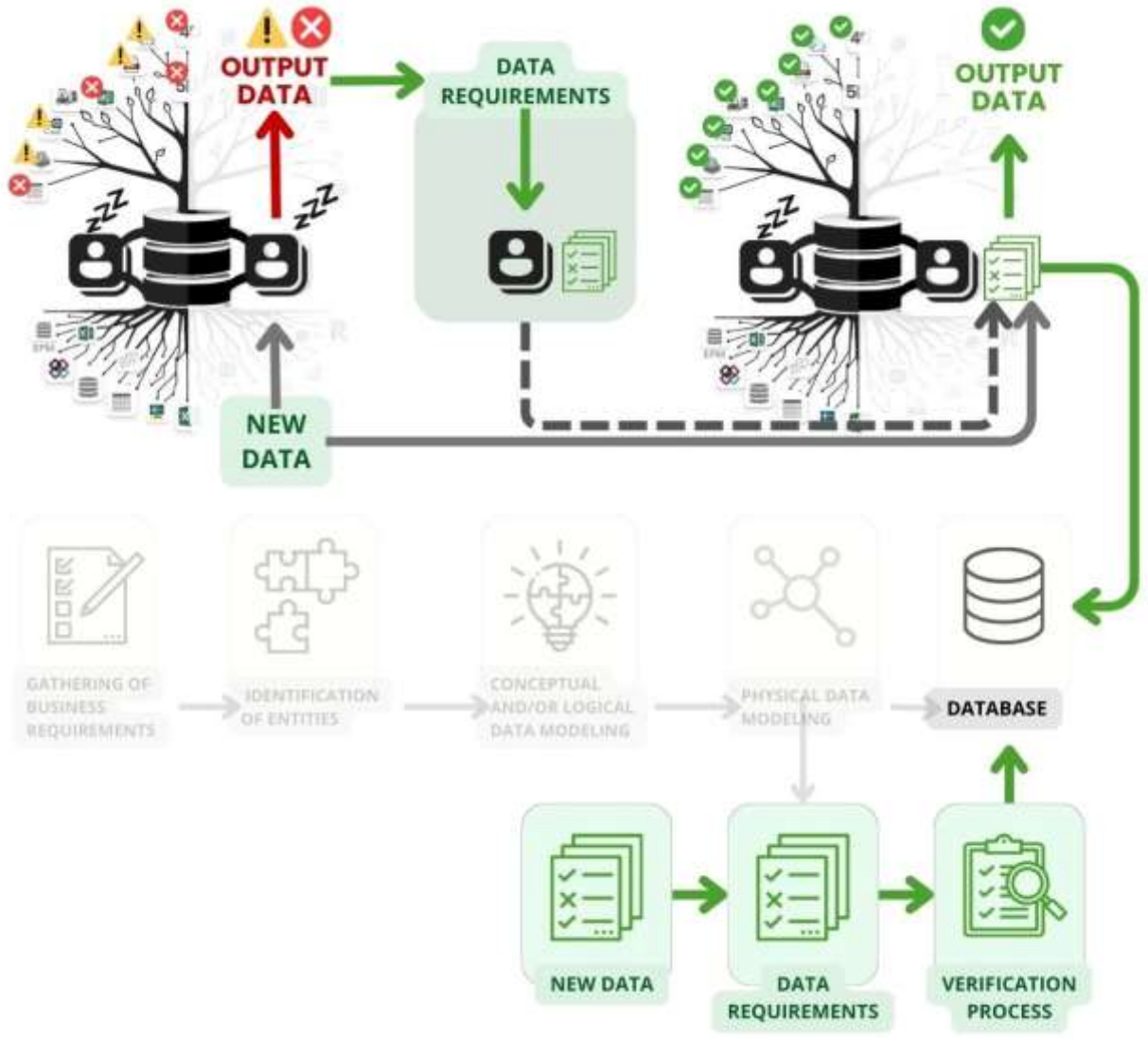


Fig. 4.2-5 La calidad de los datos depende de la calidad de los requisitos que se crean para casos específicos de uso de datos.

Dado que cada sistema tiene sus propios requisitos específicos en materia de datos, el primer paso para formular requisitos generales consiste en categorizar todos los elementos que intervienen en los procesos empresariales. Esto implica la necesidad de dividir los objetos en clases y grupos de clases correspondientes a sistemas o tareas de aplicación específicos. Para cada uno de estos grupos se elaboran requisitos independientes en materia de estructura, atributos y calidad de los datos.

En la práctica, sin embargo, la aplicación de este enfoque se enfrenta a un reto importante: la falta de un lenguaje común para agrupar los datos. Clasificaciones dispares, identificadores duplicados y formatos incompatibles hacen que cada empresa, cada software e incluso cada proyecto formen sus propios modelos y clases de datos aislados. El resultado es una "Torre de Babel" digital en la que la transferencia de información entre sistemas requiere múltiples conversiones a los modelos y clases de datos adecuados, a menudo de forma manual. Este obstáculo sólo puede superarse avanzando hacia clasificadores universales y conjuntos normalizados de requisitos.

Un lenguaje común de construcción: el papel de los clasificadores en la transformación digital

En el contexto de la digitalización y automatización de los procesos de inspección y procesamiento, los sistemas de clasificación desempeñan un papel especial: una especie de "diccionarios digitales" que garantizan la uniformidad en la descripción y parametrización de los objetos. Los clasificadores constituyen el "lenguaje común" que permite agrupar los datos por su significado e integrarlos entre distintos sistemas, niveles de gestión y fases del ciclo de vida del proyecto.

El impacto más tangible de los clasificadores se produce en la economía del ciclo de vida del edificio, donde el aspecto más importante es la optimización de los costes de explotación a largo plazo. Los estudios demuestran que los costes de explotación suponen hasta el 80% del coste total de propiedad de un edificio, es decir, tres veces más que los costes iniciales de construcción (Fig. 4.2-6) [98]. Esto significa que la decisión sobre los costes futuros se toma en gran medida en la fase de diseño

> Por ello, los requisitos de los ingenieros de operaciones (CAFM, AMS, PMS, RPM) deben convertirse en el punto de partida para generar requisitos de datos durante la fase de diseño (Fig. 1.2-4). Estos sistemas no deben considerarse la fase final del proyecto, sino parte integrante de todo el ecosistema digital del proyecto, desde su concepción hasta su desmantelamiento

Un clasificador moderno no es sólo un sistema de códigos para agrupar. Es un mecanismo de comprensión mutua entre arquitectos, ingenieros, estimadores, logistas, responsables de mantenimiento y sistemas informáticos. Al igual que el piloto automático de un coche debe reconocer los objetos de la carretera sin ambigüedades y con precisión, los sistemas digitales de construcción y sus usuarios deben interpretar el mismo elemento del proyecto sin ambigüedades para los distintos sistemas a través de la clase de elemento.

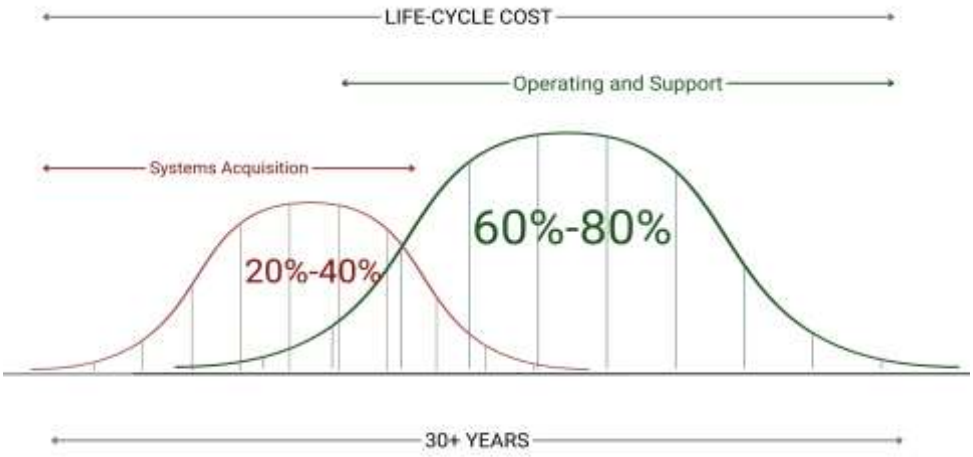


Fig. 4.2-6 Los costes de explotación y mantenimiento superan tres veces el coste de construcción y representan entre el 60% y el 80% de todos los costes del ciclo de vida de un edificio (según [99]).

El nivel de desarrollo de los clasificadores está directamente relacionado con la profundidad de la digitalización de una empresa y su madurez digital. Las organizaciones con un bajo nivel de madurez digital se enfrentan a datos fragmentados, sistemas de información incompatibles y, en consecuencia, clasificadores incompatibles e ineficaces. En estas empresas, un mismo elemento puede tener a menudo diferentes identificadores de agrupación en distintos sistemas, lo que dificulta gravemente la integración final e imposibilita la automatización de los procesos.

Por ejemplo, la misma ventana de un proyecto puede etiquetarse de forma diferente en el modelo CAD y en el sistema de estimación y mantenimiento (Fig. 4.2-7) debido a la percepción multidimensional de los elementos por parte de los distintos participantes en el proceso. Para el estimador, en la categoría de elemento de ventana son importantes el volumen y el coste; para el servicio de mantenimiento, la disponibilidad y la facilidad de mantenimiento; para el arquitecto, las características estéticas y funcionales. En consecuencia, un mismo elemento puede requerir parámetros diferentes.

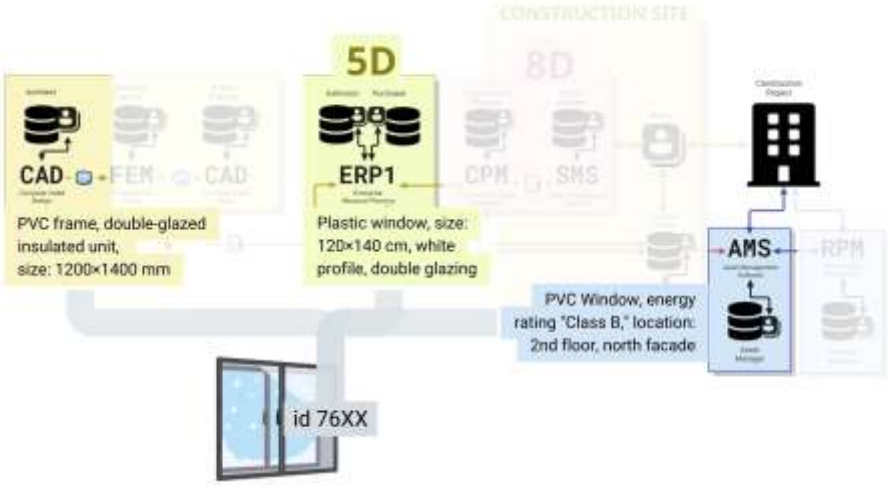


Fig. 4.2-7 Con una categorización incoherente entre sistemas un elemento perderá parte de la información de atributo en cada paso de su transición a otro sistema.

Debido a la dificultad de definir sin ambigüedades la clasificación de los elementos de construcción, los especialistas de distintos campos suelen asignar clases incompatibles a un mismo elemento. El resultado es la pérdida de una visión unificada del objeto, lo que requiere una intervención manual posterior para armonizar los distintos sistemas de clasificación y establecer una coherencia entre los tipos y clases definidos por los distintos especialistas.

Como consecuencia de esta incoherencia, la documentación operativa que recibe el departamento de compras (ERP) cuando se adquiere un elemento de construcción a un fabricante a menudo no puede vincularse correctamente con la clasificación de dicho elemento en la obra (PMIS, SCM). En consecuencia, es probable que la información crítica no se integre en los sistemas de gestión de infraestructuras y activos (CAFM, AMS), lo que crea graves problemas durante la puesta en servicio, así como durante el mantenimiento posterior (AMS, RPM) o la sustitución del elemento.

> En las empresas con una elevada madurez digital, los clasificadores desempeñan el papel de un sistema nervioso que unifica todos los flujos de información. Un mismo elemento recibe un identificador único, lo que permite transferirlo entre los sistemas CAD, ERP, AMS y CAFM y sus clasificadores sin distorsiones ni pérdidas.

Para construir clasificadores eficaces, hay que entender cómo se utilizan los datos. Un mismo ingeniero puede nombrar y clasificar un elemento de forma diferente en distintos proyectos. Sólo recopilando estadísticas de uso a lo largo de los años se puede desarrollar un sistema de clasificación estable. El aprendizaje automático ayuda a ello: los algoritmos analizan miles de proyectos (Fig. 9.1-10), identificando clases y parámetros probables mediante aprendizaje automático (Fig. 10.1-6). La clasificación automática es especialmente valiosa en entornos en los que la clasificación manual no es posible debido al volumen de datos. Los sistemas de clasificación automática podrán distinguir categorías básicas basándose en parámetros de elementos mínimamente poblados (más detalles en las partes novena y décima del libro).

Los sistemas de clasificación desarrollados se convierten en catalizadores de una mayor digitalización, creando la base para:

- Estimación automatizada de los costes y plazos del proyecto.
- Análisis predictivo de riesgos y conflictos potenciales
- Optimización de los procesos de adquisición y las cadenas logísticas
- Creación de gemelos digitales de edificios y estructuras
- Integraciones con sistemas de ciudades inteligentes e Internet de las cosas

> El tiempo para la transformación es limitado: con el desarrollo de las tecnologías de aprendizaje automático y visión por ordenador, el problema de la clasificación automática, irresoluble durante décadas, se resolverá en los próximos años, y las empresas de construcción y diseño que no se adapten a tiempo corren el riesgo de repetir el destino de las flotas de taxis desplazadas por las plataformas digitales.

En las partes quinta y novena del libro se aborda la automatización del cálculo de costes y la programación, así como los macrodatos y el aprendizaje automático de. El riesgo de que se repita el destino de las flotas de taxis y la Uberización del sector de la construcción se tratan en detalle en la décima parte del libro.

Para comprender el papel clave de los clasificadores en la transformación digital del sector de la construcción, es necesario recurrir a la historia de su evolución. Es el contexto histórico el que nos permite darnos cuenta de cómo han evolucionado los enfoques de la clasificación y qué tendencias definen su estado actual.

Masterformat, OmniClass, Uniclass y CoClass: la evolución de los sistemas de clasificación

Históricamente, los clasificadores de elementos y trabajos de construcción han evolucionado en tres generaciones, cada una de las cuales refleja el nivel de tecnología disponible y las necesidades actuales de la industria en un periodo de tiempo concreto (Fig, 4.2-8):

- **Primera generación** (de principios de los 50 a finales de los 80): directorios en papel, clasificadores jerárquicos utilizados localmente (por ejemplo, Masterformat, SfB).
- **La segunda generación** (de finales de los 90 a mediados de los 2010) son las hojas de cálculo y las bases de datos estructuradas implementadas en Excel y Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Tercera generación** (década de 2010 hasta la actualidad) - servicios digitales y API -interfaces, integración con CAD (BIM), automatización (Uniclass 2015, CoClass).

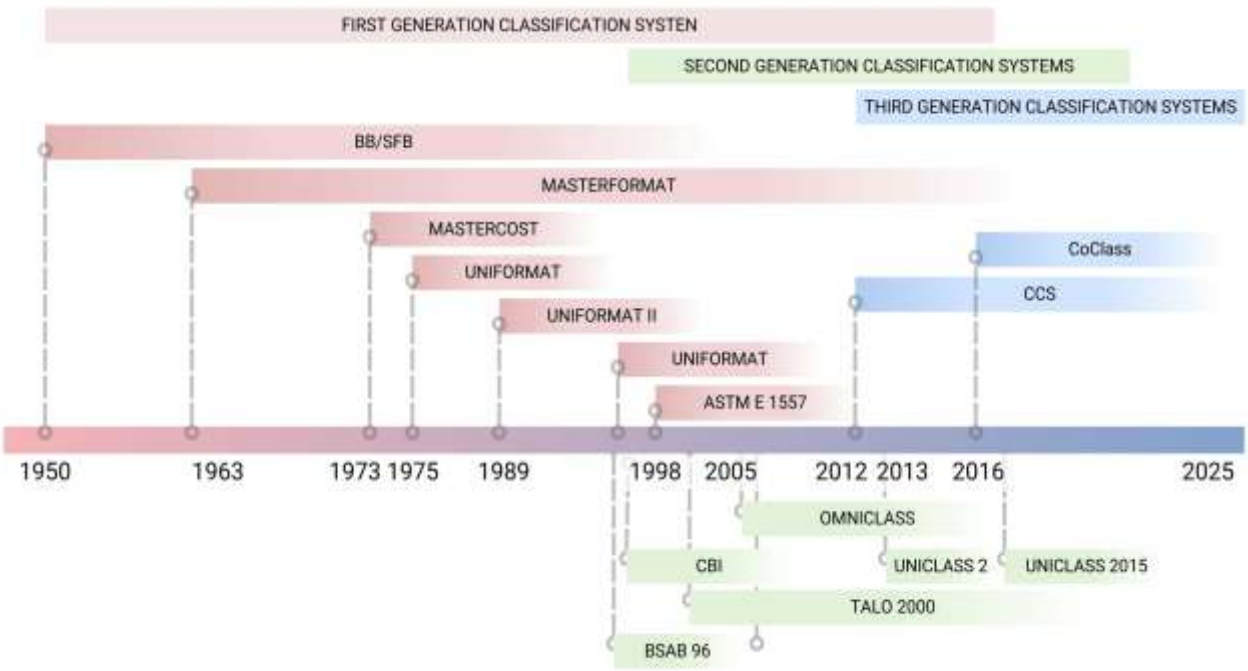


Fig. 4.2-8 Tres generaciones de clasificadores de la industria de la construcción.

En las últimas décadas se ha producido una reducción de la complejidad jerárquica (Fig. 4.2-9) de los clasificadores: mientras que los primeros sistemas, como OmniClass, utilizaban hasta 7 niveles de anidamiento para describir 6887 clases, las soluciones modernas, como CoClass, se limitan a 3 niveles con 750 clases. Esto facilita el trabajo con los datos manteniendo la granularidad necesaria. Uniclass 2015, utilizada a menudo como estándar en el Reino Unido, combina 7210 clases en solo 4 niveles, lo que resulta conveniente para los proyectos CAD y la contratación pública.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Fig. 4.2-9 Con cada nueva generación de clasificadores, la complejidad de la categorización disminuye drásticamente.

En los sistemas de estimación de la construcción de los distintos países, debido a las diferencias en las clasificaciones, incluso un elemento típico como un muro de cimentación de hormigón puede describirse de formas muy diferentes (Fig, 4.2-10). Estas diferencias reflejan las prácticas nacionales de construcción, los sistemas de medición utilizados, los enfoques de la clasificación de materiales y los requisitos normativos y técnicos vigentes en cada país.

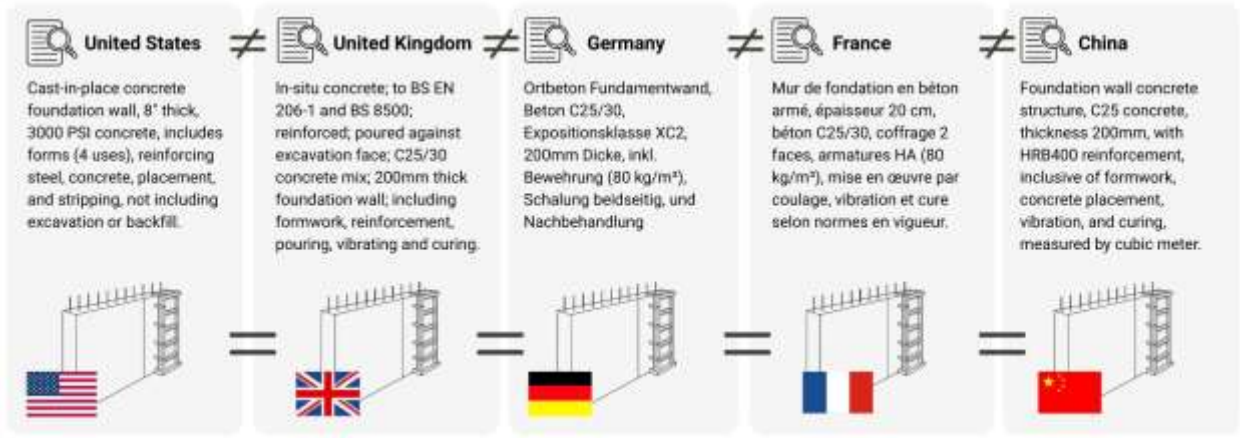


Fig. 4.2-10 El mismo elemento se utiliza en proyectos de distintos países mediante descripciones y clasificaciones diferentes.

La diversidad de clasificaciones de los mismos elementos complica la cooperación internacional y hace que la comparación de costes y alcance de los trabajos en proyectos internacionales lleve mucho tiempo y a veces resulte casi imposible. Por el momento, no existe un clasificador universal a nivel mundial: cada país o región desarrolla sus propios sistemas basados en las normas locales, el idioma y la cultura empresarial:

- **CCS** (Dinamarca): Sistema de clasificación de costes: un sistema para clasificar los costes de a lo largo del ciclo de vida de una instalación (diseño, construcción, funcionamiento). Se hace hincapié en la lógica de funcionamiento y mantenimiento, pero también incluye la gestión de presupuestos y recursos.
- **NS 3451** (Noruega): clasifica las instalaciones por funciones, elementos de diseño y fases del ciclo de vida. Se utiliza para la gestión de proyectos, la estimación de costes y la planificación a largo plazo.
- **MasterFormat** (EE.UU.): sistema para estructurar las especificaciones de construcción en secciones (por ejemplo, hormigón, electricidad, acabados). Se centra en disciplinas y tipos de obra más que en elementos funcionales (a diferencia de UniFormat).
- **Uniclass 2** (UK): uno de los clasificadores más detallados, utilizado en la contratación pública y en proyectos BIM. Unifica los datos sobre objetos, obras, materiales y espacios en un único sistema.
- **OmniClass**: una norma internacional (desarrollada por CSI en EE.UU.) para gestionar la información sobre objetos, desde bibliotecas de componentes hasta especificaciones electrónicas. Adecuado para el almacenamiento de datos a largo plazo, compatible con CAD (BIM) y otras herramientas digitales.
- **COBie**: Construction-Operation Building information exchange es un estándar internacional para el intercambio de datos entre las fases de diseño, construcción y explotación. Incluido en BS 1192-4:2014 como parte del concepto de "BIM -modelo listo para usar". Se centra en la transferencia de información (por ejemplo, especificaciones de equipos, garantías, contactos con contratistas).

Es probable que la globalización del sector de la construcción conduzca a una unificación gradual de los sistemas de clasificación de los elementos de construcción, lo que reducirá significativamente la dependencia de las normas nacionales locales. Este proceso puede desarrollarse a semejanza de la evolución de

las comunicaciones por Internet, donde los protocolos universales de transferencia de datos acabaron desplazando a los formatos locales dispares, garantizando la interoperabilidad global de los sistemas.

Una vía de desarrollo alternativa puede ser la transición directa a sistemas de clasificación automática basados en tecnologías de aprendizaje automático. Estas tecnologías, que se están desarrollando hoy en día principalmente en el campo del transporte autónomo, tienen un potencial significativo para su aplicación a grandes conjuntos de datos de diseño CAD (Fig. 10.1-6).

> Hoy en día, la situación no se limita únicamente a la agrupación nacional de clasificadores. Debido a las numerosas peculiaridades que no se tienen en cuenta a nivel nacional, cada empresa tiene que unificar y normalizar las categorías de elementos y recursos con los que trabaja.

Por regla general, este proceso se inicia a pequeña escala, con tablas locales de objetos o sistemas internos de etiquetado. Sin embargo, el objetivo estratégico es pasar a un lenguaje común para describir todos los elementos, que sea comprensible no sólo dentro de la empresa, sino también fuera de ella -idealmente, armonizado con clasificadores internacionales o industriales (Fig. 4.2-8). Este enfoque facilita la integración con socios externos y sistemas digitales y promueve la formación de procesos unificados de extremo a extremo dentro del ciclo de vida de los objetos.

Antes de pasar a la automatización y los sistemas informáticos escalables, es necesario utilizar clasificadores de ámbito nacional o construir una estructura de identificación de elementos propia, lógica e inequívoca. Cada objeto -ya sea una ventana (Fig, 4.2-11), una puerta o un sistema de ingeniería- debe describirse de forma que pueda reconocerse inequívocamente en cualquier sistema digital de la empresa. Esto es fundamental en la transición de los dibujos planos a los modelos digitales, que abarcan tanto la fase de diseño como la explotación de los edificios.

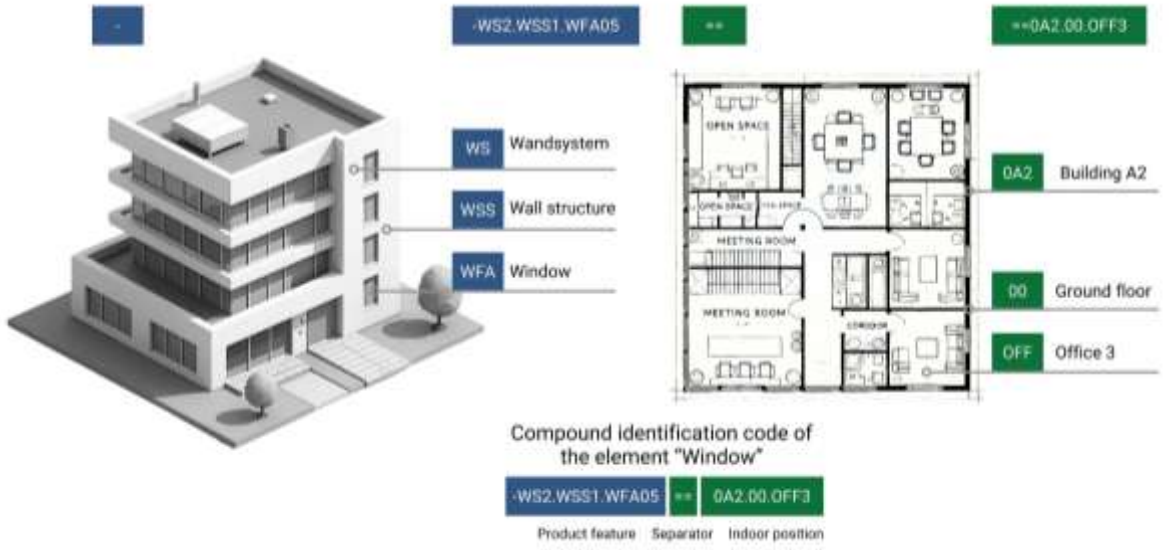


Fig. 4.2-11 Ejemplo de identificador de elemento constructivo de ventana compuesta basado en la clasificación y la posición en el edificio.

Un ejemplo de clasificadores internos podría ser el desarrollo de un código de identificación compuesto (Fig, 4.2-11). Dicho código combina varios niveles de información: la finalidad funcional del elemento (por ejemplo, "ventana en la pared"), su tipo y la referencia espacial exacta -edificio A2, planta 0, habitación 3-. Esta estructura multinivel permite crear un sistema unificado de navegación por los modelos digitales y la documentación, especialmente en las fases de verificación y transformación de datos, en las que se requiere una agrupación inequívoca de los elementos. El reconocimiento inequívoco de elementos garantiza la coherencia entre departamentos y reduce los riesgos de duplicación, errores y pérdida de información.

Un clasificador bien construido no es sólo un documento técnico, es la base del ecosistema digital de una empresa:

- garantiza la compatibilidad de los datos entre sistemas;
- reduce el coste de búsqueda y tratamiento de la información;
- aumenta la transparencia y la capacidad de gestión;
- crea una base para la ampliación y la automatización.

La descripción normalizada de los objetos, mediante el uso de clasificadores nacionales o códigos de identificación compuestos propios, se convierte en la base de unos datos coherentes, un intercambio de información fiable y la posterior implantación de servicios inteligentes, desde la contratación automatizada hasta los gemelos digitales.

Tras completar la fase de estructuración de los datos multiformato y seleccionar el clasificador que se utilizará para reconocer y agrupar los elementos, el siguiente paso consiste en modelizar correctamente los datos. Este proceso implica identificar los parámetros clave, construir una estructura lógica de los datos y describir las relaciones entre los elementos.

CAPÍTULO 4.3.
MODELIZACIÓN DE DATOS Y CENTRO DE EXCELENCIA

Modelización de datos: modelo conceptual, lógico y físico

La gestión eficaz de los datos (antes estructurados y clasificados por categorías) es imposible sin una estructura de almacenamiento y tratamiento bien pensada. Para garantizar el acceso y la coherencia de la información en las fases de almacenamiento y tratamiento, las empresas recurren a la modelización de datos, una metodología que permite diseñar tablas, bases de datos y vínculos entre ellas en función de las necesidades de la empresa.

> La modelización de datos es la base sobre la que se construye cualquier ecosistema digital. Sin una descripción de sistemas, requisitos y modelización de datos, los ingenieros y profesionales que crean datos no saben ni entienden dónde se utilizarán los datos que crean.

Al igual que en la construcción de un edificio, donde no se puede empezar a poner ladrillos sin un plan, la creación de un sistema de almacén de datos requiere una comprensión clara de qué datos se utilizarán, cómo se vincularán y quién trabajará con ellos. Sin una descripción de los procesos y requisitos, los ingenieros y profesionales que crean los datos pierden de vista dónde y cómo se utilizarán en el futuro.

El modelo de datos sirve de puente entre la empresa y las TI. Permite formalizar los requisitos, estructurar la información y facilitar la comunicación entre las partes interesadas. En este sentido, el modelado de datos es similar al trabajo de un arquitecto que, según el plano del cliente, desarrolla un plan de construcción y luego lo entrega a los constructores -administradores y desarrolladores de bases de datos- para su implementación (creación de la base de datos).

Así pues, toda empresa de construcción, además de estructurar y categorizar elementos y recursos (Fig. 4.2-11), debe dominar el arte de "construir" bases de datos (tablas) y aprender a crear vínculos entre ellas, como si conectara los ladrillos en un muro fiable y sólido de conocimiento a partir de los datos de la empresa. Entre los conceptos clave de la modelización de datos (Fig, 4.3-1) se incluyen:

- **Las entidades** son objetos sobre los que hay que recopilar datos. En la fase inicial de diseño, una entidad puede ser un único elemento (por ejemplo, "puerta"), y en el modelo de estimación puede ser un grupo de elementos categorizados (por ejemplo, "puertas interiores").
- **Los atributos** son características de las entidades que describen detalles importantes: dimensiones, propiedades, costes de montaje, logística y otros parámetros.
- **Relaciones (enlaces)**: muestran cómo interactúan las entidades entre sí. Pueden ser de uno de los tipos "uno a uno", "muchos a uno", "muchos a muchos".
- **Los diagramas ER** (Entity-Relationship diagrams) son diagramas visuales que muestran entidades, atributos y las relaciones entre ellos. Los diagramas ER pueden ser conceptuales, lógicos y físicos, cada uno de los cuales refleja un nivel de detalle diferente.

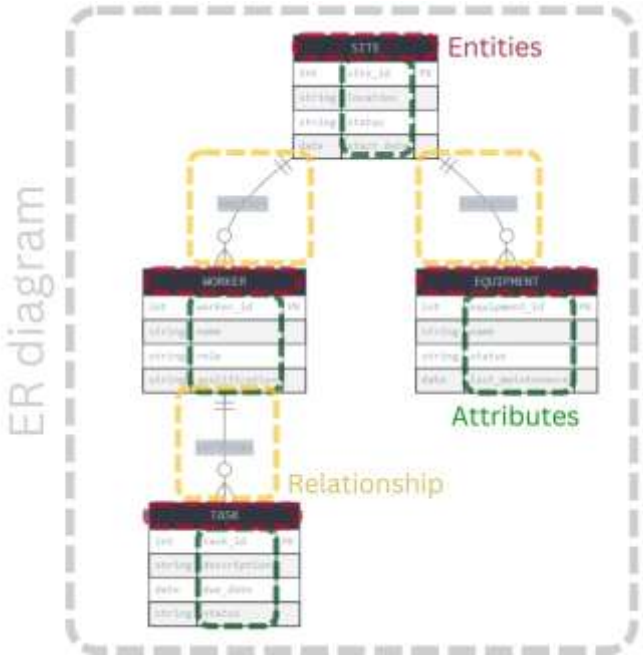


Fig. 4.3-1 Diagrama ER de una estructura conceptual de base de datos con entidades, atributos y relaciones.

El proceso de diseño de datos y definición de relaciones entre ellos se divide tradicionalmente en tres modelos principales. Cada uno de ellos cumple determinadas funciones, diferenciándose en el nivel de detalle y el grado de abstracción a la hora de representar la estructura de datos:

- **Modelo conceptual de datos**: este modelo describe las principales entidades y sus relaciones sin entrar en los detalles de los atributos. Suele utilizarse en las fases iniciales de la planificación. En esta fase podemos hacer esquemas a partir de bases de datos y sistemas para mostrar la relación entre los distintos departamentos y especialistas.

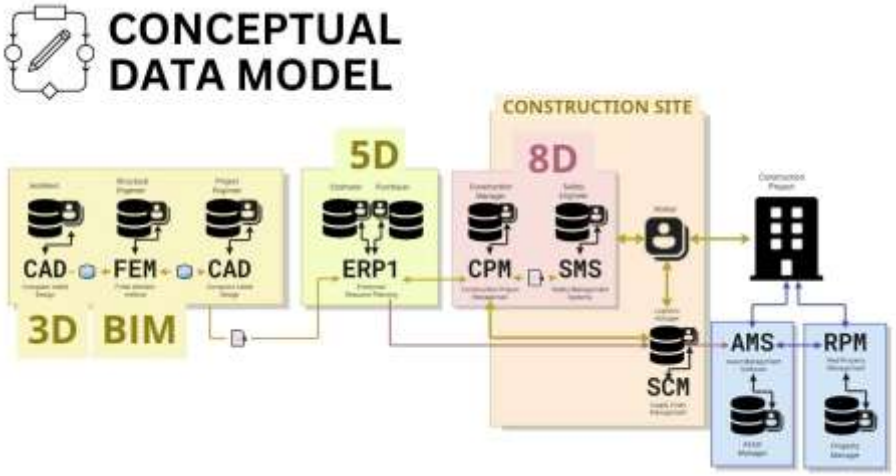


Fig. 4.3-2 El diagrama conceptual describe el contenido del sistema: una representación de alto nivel de las relaciones, sin detalles técnicos.

- **Modelo Lógico de Datos**: Basado en el modelo conceptual, el modelo lógico de datos incluye descripciones detalladas de entidades, atributos, claves y relaciones, mapeando información y reglas de negocio.

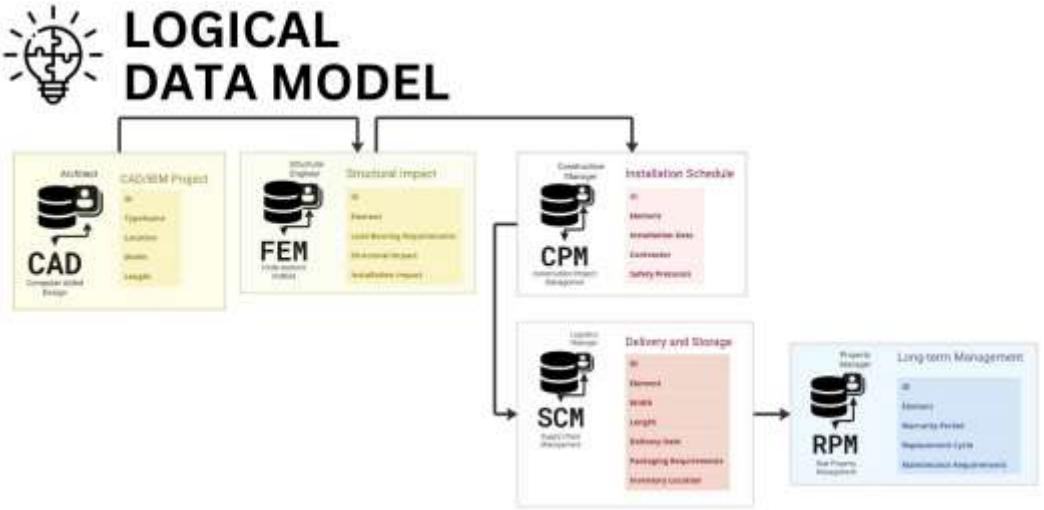


Fig. 4.3-3 Modelo lógico de datos detalla los tipos de datos, las relaciones y las claves, pero sin la implementación del sistema.

- **Modelo físico de datos**: Este modelo describe las estructuras necesarias para implementar una base de datos, incluidas tablas, columnas y relaciones. Se centra en el rendimiento de la base de datos, las estrategias de indexación y el almacenamiento físico para optimizar el despliegue físico de las bases de datos.

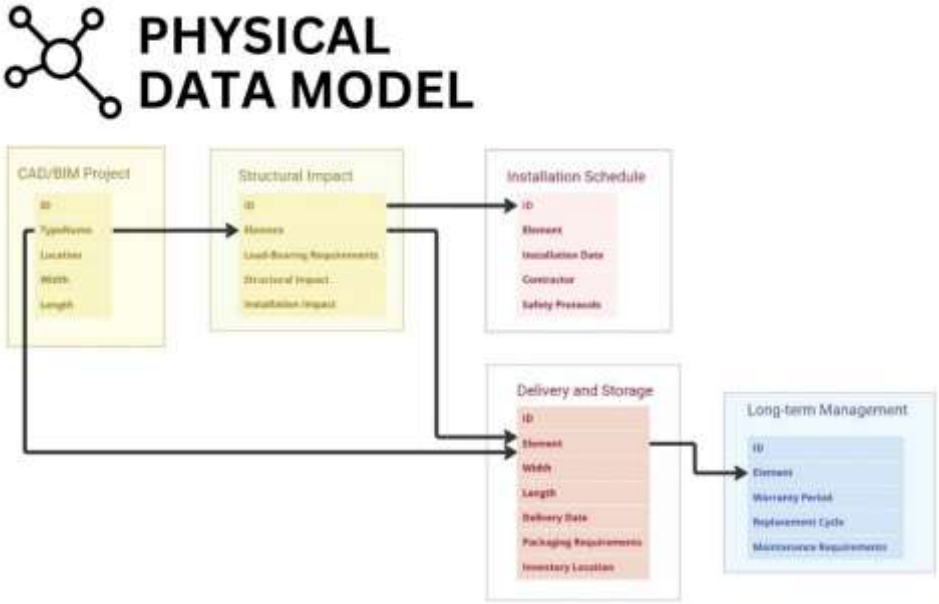


Fig. 4.3-4 El modelo físico de datos define cómo se implementará el sistema, incluidas las tablas y los detalles específicos de la base de datos.

> A la hora de diseñar bases de datos y relaciones tabulares, la comprensión de los niveles de abstracción desempeña un papel clave en la construcción de una arquitectura de sistema eficaz.

Una metodología eficaz de modelado de datos permite combinar los objetivos empresariales con la implementación técnica, lo que hace que toda la cadena del proceso sea más transparente y manejable. El modelado de datos no es una tarea puntual, sino un proceso que implica pasos secuenciales (Fig, 4.3-5):

- **Recopilación de requisitos empresariales:** se definen las tareas clave, los objetivos y los flujos de información. Esta es la fase de interacción activa con expertos y usuarios.
- **Identificación de entidades:** se destacan los principales objetos, categorías y tipos de datos que es importante tener en cuenta en el futuro sistema.
- **Desarrollo de un modelo conceptual y lógico:** primero se capturan las entidades clave y sus relaciones, y después los atributos, las reglas y la estructura detallada.
- **Modelización física:** se diseña la implementación técnica del modelo: tablas, campos, relaciones, restricciones, índices.
- **Creación de la base de datos:** la última etapa consiste en implantar el modelo físico en el SGBD elegido, realizar pruebas y prepararse para el funcionamiento.

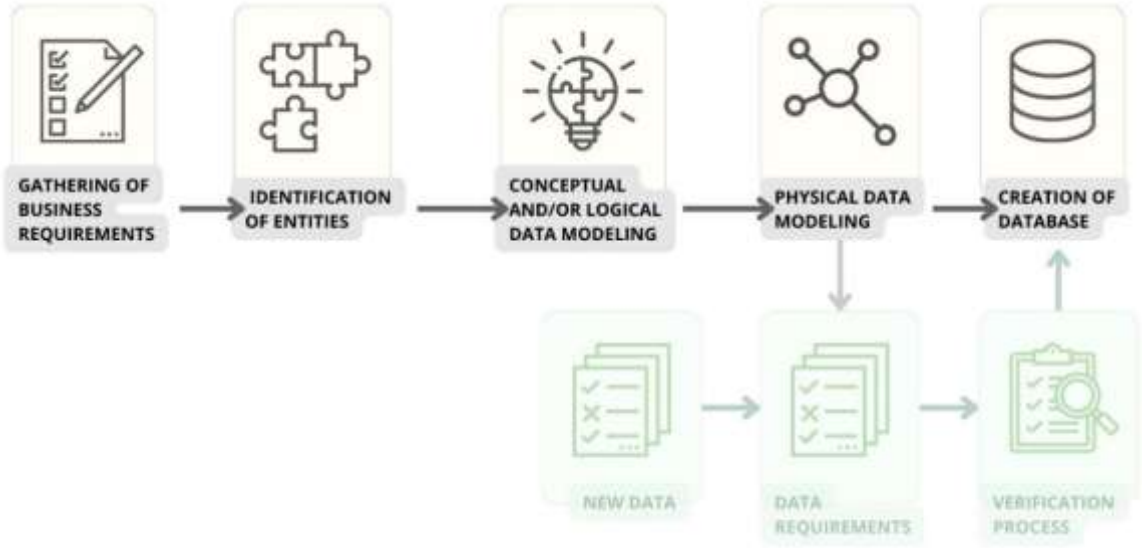


Fig. 4.3-5 La creación de bases de datos y sistemas de gestión de datos para procesos empresariales comienza con la generación de requisitos y el modelado de datos.

Unos procesos de modelización de datos correctamente diseñados permiten flujos de información transparentes, lo que es especialmente importante en proyectos complejos como la gestión de proyectos de construcción o de obras. Veamos cómo pasar de un modelo conceptual a un modelo lógico y luego a un modelo físico puede ayudar a agilizar los procesos.

Modelización práctica de datos en el contexto de la construcción

Tomemos una tarea de gestión de una obra de construcción como ejemplo de modelado de datos y convirtamos los requisitos de foreman en un modelo lógico estructurado. Partiendo de las necesidades básicas de la gestión de obras, definimos entidades clave para: obra (SITE), trabajadores (WORKER), equipos (EQUIPMENT), tareas (TASK) y uso de equipos (EQUIPMENT_USAGE). Cada entidad contiene un conjunto de atributos que reflejan características importantes. Por ejemplo, para la TAREA, podría ser una descripción de la tarea, fecha de vencimiento, estado, prioridad; para el TRABAJADOR, podría ser el nombre, su papel en el sitio, empleo actual, etc.

El modelo lógico establece relaciones entre estas entidades, mostrando cómo interactúan entre sí en los procesos de trabajo reales (Fig, 4.3-6). Por ejemplo, la relación entre sitio y trabajadores indica que muchos trabajadores pueden trabajar en un sitio, mientras que la relación entre trabajadores y tareas refleja que un trabajador puede realizar múltiples tareas.

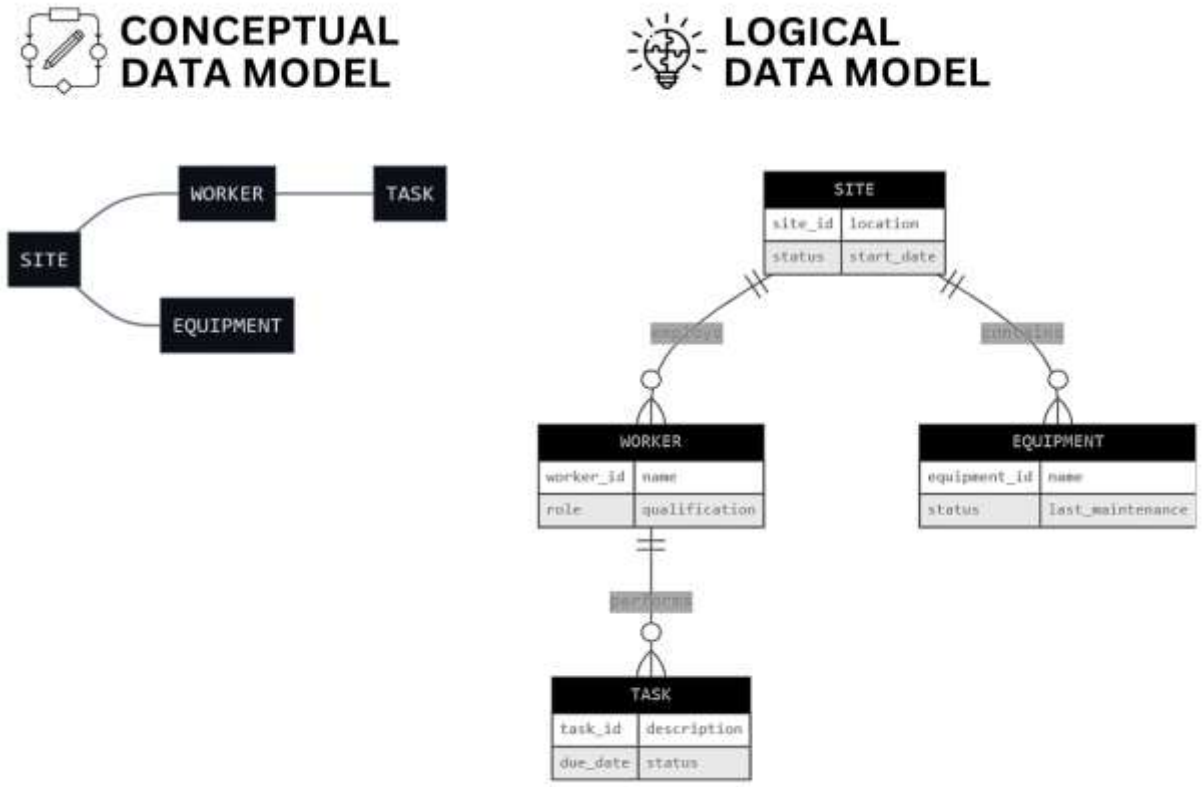


Fig. 4.3-6 Modelo de datos conceptual y lógico generado por los requisitos del capataz para describir los procesos de la obra de construcción.

Al pasar al modelo físico, se añaden detalles técnicos de implementación: tipos de datos específicos (VARCHAR, INT, DATE), claves primarias y foráneas para las relaciones entre tablas, e índices para optimizar el rendimiento de la base de datos (Fig, 4.3-7).

Por ejemplo, habría que definir tipos específicos con valores posibles para los estados, y añadir índices en campos clave como estado y trabajador_id para mejorar el rendimiento de las búsquedas. De este modo,

una descripción lógica del sistema se convierte en un plan concreto de implantación de la base de datos, listo para ser creado e implantado.

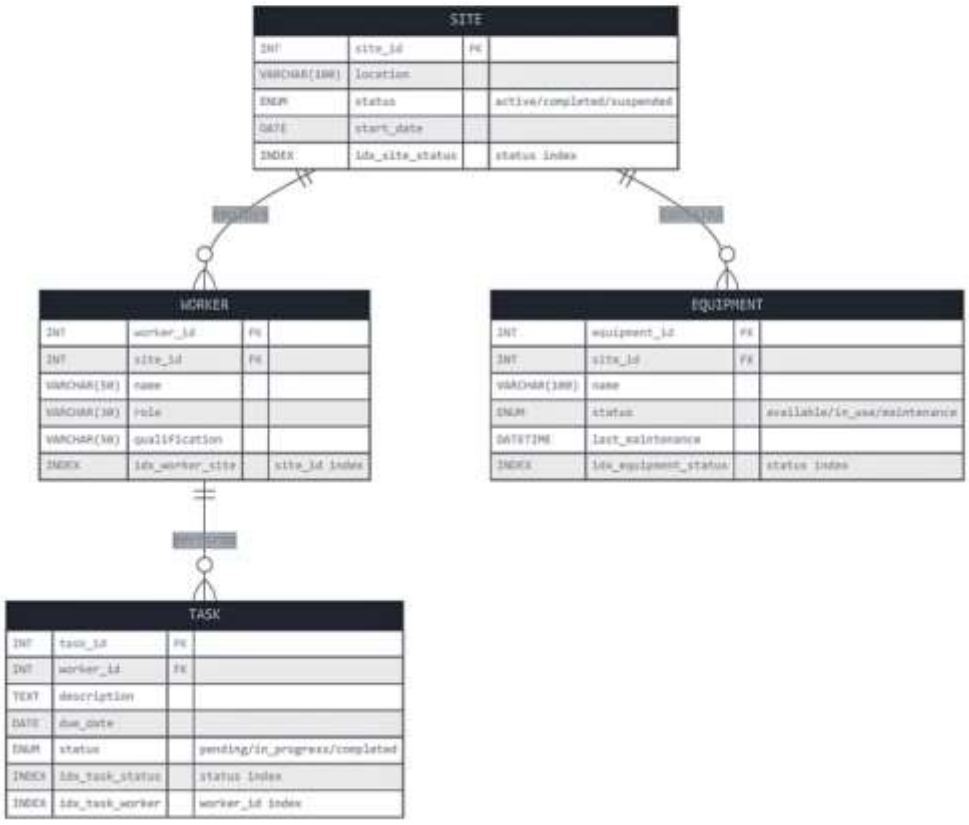


Fig. 4.3-7 El modelo de datos físicos describe las entidades de una obra mediante los parámetros mínimos requeridos.

> El modelo físico suele ser diferente del modelo lógico. Por término medio, la distribución del tiempo de modelización es la siguiente: aproximadamente el 50% se dedica al modelo conceptual (recopilación de requisitos, discusión de procesos, identificación de entidades), el 10% al modelo lógico (especificación de atributos y relaciones) y el 40% al modelo físico (implementación, pruebas, adaptación al SGBD).

Este equilibrio se debe a que la etapa conceptual sienta las bases de la estructura de datos, mientras que el modelo lógico sólo especifica relaciones y atributos. El modelo físico es el que requiere más recursos, ya que es en esta fase cuando se implementan los datos en plataformas y herramientas específicas

Creación de una base de datos con LLM

Teniendo un modelo de datos y la descripción de las entidades a través de parámetros, estamos listos para crear bases de datos - almacenes, donde almacenaremos la información que viene después de la etapa de estructuración en procesos específicos.

Intentemos crear un ejemplo de base de datos sencilla pero funcional con una cantidad mínima de código utilizando SQLite tomando como ejemplo el lenguaje de programación Python. Las bases de datos relacionales se trataron en detalle en el capítulo "Bases de datos relacionales estructuradas y lenguaje de consulta SQL ".

> SQLite es una base de datos ligera, de código abierto, integrable y multiplataforma, compatible con las principales funciones de SQL. SQLite se utiliza ampliamente en CAD (BIM), ERP, EPM y está presente en casi todos los smartphones.

Pidamos al modelo LLM que nos cree una pequeña base de datos SQLlite basada en el modelo físico creado anteriormente (Fig, 4.3-4) para los sistemas FEM (impacto estructural) y CPM (programa de montaje).

- Envíe una solicitud de texto al chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otro):

> Cree bases de datos SQLlite para los sistemas "Impacto Constructivo" y "Programa de Instalación" a partir de la figura del archivo adjunto (pegue el esquema en el chat como imagen (Fig. 4.3-4)⏎

Respuesta de LLM:

Create in LLM Chat

Database SQLite

```python
import sqlite3

# Connect to SQLite database (or create if it doesn't exist)
conn = sqlite3.connect('construction.db')

# Create a cursor object using the cursor() method
cursor = conn.cursor()

# Create table - Structural Impact
cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    LoadBearingRequirements TEXT,
                    StructuralImpact TEXT,
                    InstallationImpact TEXT
                )''')

# Create table - Installation Schedule
cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    InstallationDate DATE,
                    Contractor TEXT,
                    SafetyProtocols TEXT
                )''')

# Commit the changes
conn.commit()

# Close the connection
conn.close()

print("Database and tables created successfully.")
```

Run in IDE

Fig. 4.3-8 El código generado crea bases de datos y tablas a partir de la imagen cargada en las que ahora podemos cargar datos.

Este código de base de datos puede ejecutarse en el entorno Python con soporte para SQLite, que suele incluirse en el entorno Python estándar. Puede ejecutarse en cualquier máquina Python local capaz de ejecutar scripts y crear archivos. Una vez ejecutado, este código creará un archivo en su disco duro - una base de datos SQLite llamada construction.db (similar a un archivo Excel con hojas de trabajo) que contendrá las tablas "Structural Impact" y "Installation Schedule".

Una vez creada una nueva base de datos **construction.db** en estas tablas, se pueden añadir datos a estas tablas a través de consultas SQL _o de importación, lo que permitirá empezar a crear procesamientos automáticos de datos más adelante. Los datos pueden importarse a la base de datos SQLite desde archivos CSV, hojas de cálculo Excel o exportarse desde otras bases de datos y repositorios a través de la API.

Para establecer procesos sostenibles de modelización de datos y gestión eficaz de bases de datos, una empresa necesita una estrategia claramente definida, así como coordinación entre los equipos técnicos y los de negocio. Con proyectos dispares y múltiples fuentes de datos, suele ser difícil garantizar la coherencia, la normalización y el control de calidad a todos los niveles. Una solución clave puede ser crear un Centro de Excelencia (COE) dedicado al modelado de datos dentro de la empresa.

Centro de Excelencia (CdE) de Modelización de Datos

Dado que los datos se están convirtiendo en uno de los principales activos estratégicos, las empresas no sólo deben recopilar y almacenar correctamente la información, sino que es importante aprender a gestionarlos sistemáticamente. El Centro de Excelencia de Clasificación y Modelización de Datos (CoE) es una unidad estructural que garantiza la coherencia, calidad y eficiencia de todo el tratamiento de datos en la organización.

> El Centro de Excelencia (CoE) es el núcleo de apoyo experto y la base metodológica para la transformación digital en una empresa. Construye una cultura impulsada por los datos y permite a las organizaciones crear procesos que toman decisiones basadas en datos estructurados, validados y representativos, en lugar de en la intuición o la información local.

Un centro de datos de excelencia suele estar formado por equipos interfuncionales que trabajan según el principio de las "dos pizzas". Este principio, propuesto por Jeff Bezos, significa que el tamaño del equipo debe ser tal que pueda alimentarse con dos pizzas, es decir, no más de 6-10 personas. Este planteamiento ayuda a evitar una burocracia excesiva y aumenta la flexibilidad del trabajo. El equipo del CdE debe incluir empleados con diversas competencias técnicas, desde analistas de datos y aprendizaje automático hasta expertos en áreas empresariales específicas. Con sus profundos conocimientos técnicos, los ingenieros de datos no sólo deben optimizar los procesos y modelar los datos, sino también apoyar a sus colegas reduciendo el tiempo dedicado a tareas rutinarias (Fig. 4.3-9).

> Al igual que en la naturaleza la biodiversidad garantiza la resistencia de los ecosistemas, en el mundo digital la flexibilidad y la adaptabilidad se consiguen gracias a la diversidad de enfoques para tratar los datos. Sin embargo, esta diversidad debe sustentarse en normas y conceptos comunes.

Un Centro de Excelencia (CoE) puede compararse con las "condiciones climáticas" de un ecosistema forestal, que determinan qué tipos de datos prosperarán y cuáles se descartarán automáticamente. Al crear un "clima" favorable a los datos de calidad, el CdE facilita la selección natural de las mejores prácticas y metodologías, que luego se convierten en normas para la organización.

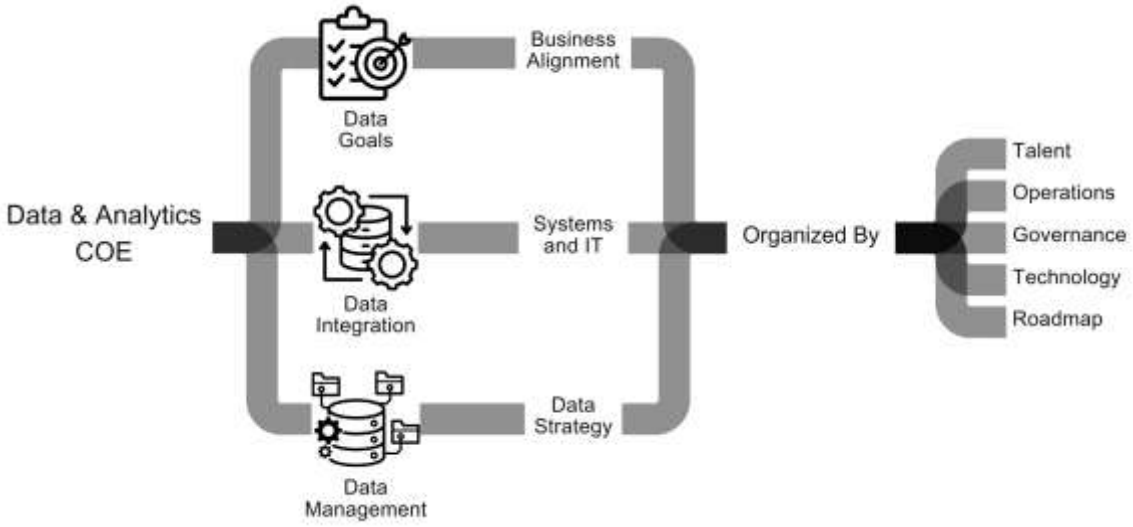


Fig. 4.3-9 El Centro de Excelencia (CoE) para Datos y Análisis reúne conocimientos especializados sobre aspectos clave de la gestión, integración y estrategia de datos.

Para acelerar los ciclos de integración y lograr mejores resultados, el CoE debe proporcionar a sus miembros un grado suficiente de autonomía en la toma de decisiones. Esto es especialmente importante en un entorno dinámico en el que el ensayo y error, la retroalimentación constante y los lanzamientos frecuentes pueden reportar importantes beneficios. Sin embargo, esta autonomía sólo es eficaz si existe una comunicación clara y el apoyo de la alta dirección. Sin visión estratégica y coordinación al más alto nivel, incluso el equipo más competente puede encontrar obstáculos a la hora de poner en práctica sus iniciativas.

> El COE o la alta dirección de la empresa son los responsables de garantizar que el enfoque de modelización de datos no se limite a uno o dos proyectos, sino que se integre en el sistema general de gestión de la información y de los procesos empresariales.

El Centro de Experiencia (CoE), además de sus tareas relacionadas con el modelado de datos y la Gobernanza de Datos, es responsable de desarrollar normas y enfoques comunes para el despliegue y funcionamiento de la infraestructura de datos. Además, fomenta una cultura de mejora continua, optimización de procesos y uso eficiente de los datos en la organización (Fig, 4.3-10).

El enfoque sistemático de la gestión de datos y modelos en el seno del CdE puede dividirse a grandes rasgos en varios bloques clave:

- **Normalización de procesos y gestión del ciclo de vida de los modelos:** CoE desarrolla e implanta metodologías para unificar la creación y gestión de modelos de datos. Esto incluye: el establecimiento de plantillas estructurales, métodos de control de calidad y sistemas de control de versiones para garantizar la continuidad de los datos en todas las fases de trabajo.

- **Gestión de funciones y asignación de responsabilidades:** El COE define las funciones clave en el proceso de modelado de datos. A cada participante en el proyecto se le asignan funciones y áreas de responsabilidad claramente definidas, lo que facilita el trabajo en equipo y reduce el riesgo de incoherencias en los datos.
- **Control de calidad y auditoría:** la gestión eficaz de los datos de construcción requiere un seguimiento continuo de su calidad. Se están implantando mecanismos automatizados para comprobar los datos, identificar errores y atributos que faltan.
- **Gestión de metadatos y arquitectura de la información:** CoE es responsable de crear un sistema unificado de clasificación e identificadores, normas de denominación y descripción de entidades, que es fundamental para la integración entre sistemas.

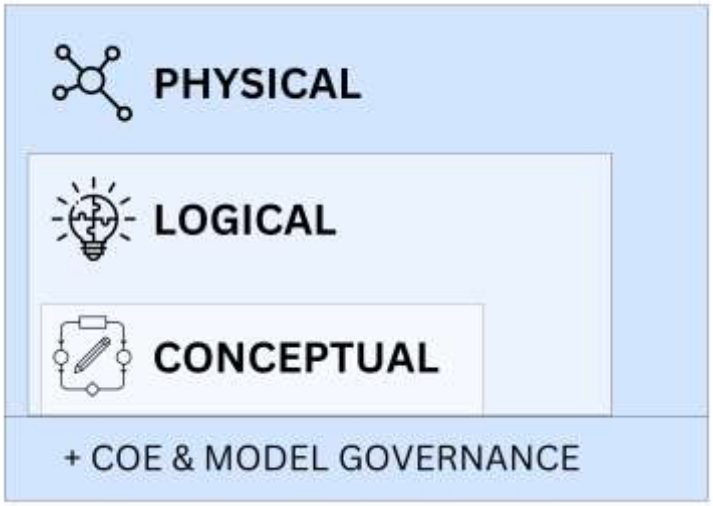


Fig. 4.3-10 La modelización y la gestión de la calidad de los datos es uno de los principales retos del CdE

El Centro de Excelencia (CoE) de Datos no es sólo un grupo de expertos, sino un mecanismo sistémico que crea una nueva cultura basada en los datos y garantiza un enfoque unificado del trabajo con datos en toda la empresa. Mediante una integración competente de los procesos de modelización en el sistema general de gestión de la información, la normalización, la clasificación y el control de la calidad de los datos, el CoE ayuda a las empresas a mejorar continuamente sus productos y procesos empresariales, reaccionar más rápidamente a los cambios del mercado y tomar decisiones fundamentadas basadas en análisis fiables.

Estos centros son especialmente eficaces cuando se combinan con los principios modernos de DataOps - en un movimiento que garantiza la entrega continua, la automatización y el control de calidad de los datos. Hablaremos más sobre DataOps en la Parte 8, en el capítulo "Tecnologías modernas de datos en el sector de la construcción".

En los capítulos siguientes pasaremos de la estrategia a la práctica: "transformémonos" condicionalmente en un centro de datos: veremos varios ejemplos de cómo se lleva a cabo la parametrización de tareas, la recopilación de requisitos y el proceso de validación automática.

CAPÍTULO 4.4.
SISTEMATIZACIÓN DE LOS REQUISITOS Y VALIDACIÓN DE LA INFORMACIÓN

Recopilación y análisis de requisitos: transformar las comunicaciones en datos estructurados

Recopilar y gestionar los requisitos es el primer paso para garantizar la calidad de los datos. A pesar del desarrollo de herramientas digitales, la mayoría de los requisitos siguen formulándose de forma no estructurada: a través de cartas, actas de reuniones, llamadas telefónicas y conversaciones verbales. Esta forma de comunicación dificulta la automatización, validación y reutilización de la información. En este capítulo veremos cómo traducir los requisitos textuales en estructuras formales, garantizando que los requisitos empresariales sean transparentes y sistemáticos.

> La investigación de Gartner, "Data Quality: Best Practices for Accurate Insights", destaca la importancia crítica de la calidad de los datos para el éxito de las iniciativas de datos y análisis [100]. Señalan que la mala calidad de los datos cuesta a las organizaciones una media de al menos 12,9 millones de dólares al año y que unos datos fiables y de alta calidad son esenciales para crear una empresa basada en datos.

La falta de requisitos estructurados hace que un mismo elemento (entidad) y sus parámetros puedan almacenarse en distintos sistemas con variaciones diferentes. Esto no sólo reduce la eficacia de los procesos, sino que también provoca pérdidas de tiempo, duplicación de información y la necesidad de volver a validar los datos antes de poder utilizarlos. Como resultado, incluso una sola omisión -un parámetro perdido o un solo elemento descrito incorrectamente- puede ralentizar la toma de decisiones y provocar un uso ineficaz de los recursos.

> *A falta de un clavo, se perdió la herradura.*
> *A falta de herradura, se perdió el caballo.*
> *A falta de caballo, se perdió el jinete.*
> *A falta de jinete, se perdió el mensaje.*
> *A falta de mensaje, se perdió la batalla.*
> *A falta de batalla, se perdió un reino.*
> *Todo por falta de un clavo en la herradura.*
>
> - Proverbio [101]

El análisis y la recopilación de requisitos para el proceso de cumplimentación y almacenamiento de datos comienza con la identificación de todas las partes interesadas. Al igual que la pérdida proverbial de un solo clavo conduce a una cadena de consecuencias críticas, en los negocios, la pérdida de una sola parte intere-

sada, un requisito pasado por alto o la pérdida de incluso un solo parámetro puede afectar significativamente no solo a un proceso empresarial individual, sino a todo el ecosistema de un proyecto y a la organización en su conjunto. Por lo tanto, es crucial identificar incluso aquellos elementos, parámetros y funciones que parecen insignificantes a primera vista, pero que más tarde pueden resultar críticos para la sostenibilidad empresarial.

Imaginemos que una empresa tiene un proyecto en el que el cliente presenta una nueva solicitud: "*añadir una ventana adicional en el lado norte del edificio*". En el pequeño proceso "*solicitud del cliente de añadir una nueva ventana al proyecto actual*" participan el arquitecto, el cliente, el especialista en CAD (BIM), el jefe de obra, el jefe de logística, el analista de ERP, el ingeniero de control de calidad, el ingeniero de seguridad, el jefe de control y el administrador de la propiedad.

> Incluso un proceso pequeño puede implicar a docenas de especialistas diferentes. Cada participante en el proceso debe comprender los requisitos de los especialistas de con los que está conectado a nivel de datos.

En el nivel de texto (Fig. 4.4-1), la comunicación entre el cliente y los especialistas de la cadena de procesos tiene lugar del siguiente modo:

- **Cliente:** "*Hemos decidido añadir una ventana adicional en el lado norte para mejorar la iluminación. ¿Se puede realizar*?"
- **Arquitecto**: "*Claro, revisaré el proyecto para incluir la nueva ventana y enviaré los planos CAD actualizados (BIM*)".
- **Especialista en CAD (BIM):** *"Recibo un nuevo proyecto. Actualizo el modelo CAD (BIM) con la ventana adicional y, tras la coordinación con el ingeniero FEM, proporciono la ubicación y las dimensiones exactas de la nueva ventana*".
- **Jefe de obra**: "*Se ha recibido un nuevo proyecto. Estamos ajustando las fechas de instalación de 4D e informando a todos los subcontratistas pertinentes*".
- **Ingeniero de instalaciones (CAFM**): *"Introduciré los datos de 6D de la nueva ventana en el sistema CAFM para la futura gestión de las instalaciones y la planificación del mantenimiento*."
- **Responsable de logística**: "*Necesito las dimensiones y el peso de la nueva ventana para organizar la entrega de la ventana en la obra*".
- **ERP** -analista: "*Necesito las tablas de alcance y el tipo exacto de ventana para la actualización del presupuesto 5D en nuestro sistema ERP para reflejar el coste de la nueva ventana en la estimación global del proyecto*."
- **Ingeniero de control de calidad**: "*Una vez listas las especificaciones de las ventanas, me aseguraré de que cumplen nuestras normas de calidad y materiales*".
- **Ingeniero de seguridad**: "*Evaluaré los aspectos de seguridad de la nueva ventana, con especial atención a la conformidad y la evacuación según el esquema 8D ".*
- **Responsable de control**: "*Basándonos en el alcance exacto del trabajo de ERP, actualizaremos nuestro calendario 4D para reflejar la instalación de la nueva ventana, y almacenaremos los nuevos datos en el sistema de gestión de contenidos del proyecto*."
- **Trabajador (instalador)**: *"Necesita instrucciones sobre la instalación, el montaje y el ritmo de trabajo. Además, ¿se ha introducido alguna norma de seguridad especial que tenga que seguir?*".

- **Administrador de fincas**: "*Una vez instalada, documentaré la información sobre la garantía y el mantenimiento para la gestión a largo plazo*".
- **Gestor de activos**: *"Ingeniero de equipos, envíe los datos finales para el seguimiento de activos y la gestión del ciclo de vida*".
- **Cliente**: "*Espera, a lo mejor tengo prisa y no necesito la ventana. Quizá debería hacer un balcón*".

En este tipo de situaciones, que se dan con frecuencia, incluso un pequeño cambio provoca una reacción en cadena entre múltiples sistemas y roles. En este caso, casi toda la comunicación en la fase inicial es en forma de texto: correos electrónicos, chats, actas de reuniones (Fig. 4.4-1).

En un sistema de comunicación basado en texto para un proyecto de construcción de este tipo, es muy importante contar con un sistema de confirmación legal y registro de todas las operaciones de intercambio de datos y de todas las decisiones tomadas. Así se garantiza que cada decisión, instrucción o cambio realizado sea legalmente válido y rastreable, reduciendo el riesgo de futuros "malentendidos

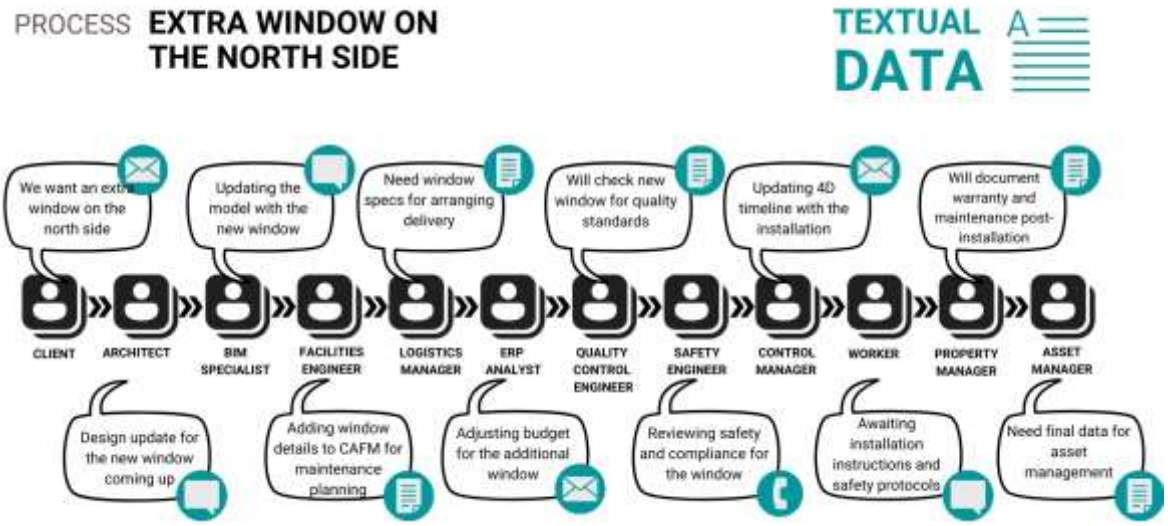


Fig. 4.4-1 La comunicación entre el cliente y el contratista en las fases iniciales de un proyecto suele contener datos textuales multiformato.

> La falta de control legal y validación de las decisiones en los sistemas pertinentes de un proyecto de construcción puede acarrear graves problemas a todos los implicados. Cada decisión, orden o cambio realizado sin la documentación y validación adecuadas puede dar lugar a disputas (y litigios).

La consolidación legal de todas las decisiones en la comunicación textual sólo puede garantizarse mediante un gran número de documentos firmados, que recaerán sobre los hombros de la dirección, obligada a registrar todas las transacciones. Como resultado, si se exige a cada participante que firme documentos para cada acción, el sistema pierde flexibilidad y se convierte en un laberinto burocrático. La falta de confirmaciones de las transacciones no sólo retrasará la ejecución del proyecto, sino que también puede provocar pérdidas financieras y el deterioro de las relaciones entre los participantes, hasta llegar a problemas legales.

Un proceso de aprobación de transacciones de este tipo, que suele comenzar con discusiones basadas en texto, evoluciona gradualmente hacia un intercambio de documentos multiformato en las siguientes etapas (Fig. 4.4-2), lo que complica considerablemente la comunicación que antes sólo tenía lugar a través de texto. Sin unos requisitos claramente definidos, la automatización de este tipo de procesos, repletos de datos multiformato y un gran número de requisitos textuales, resulta casi imposible.

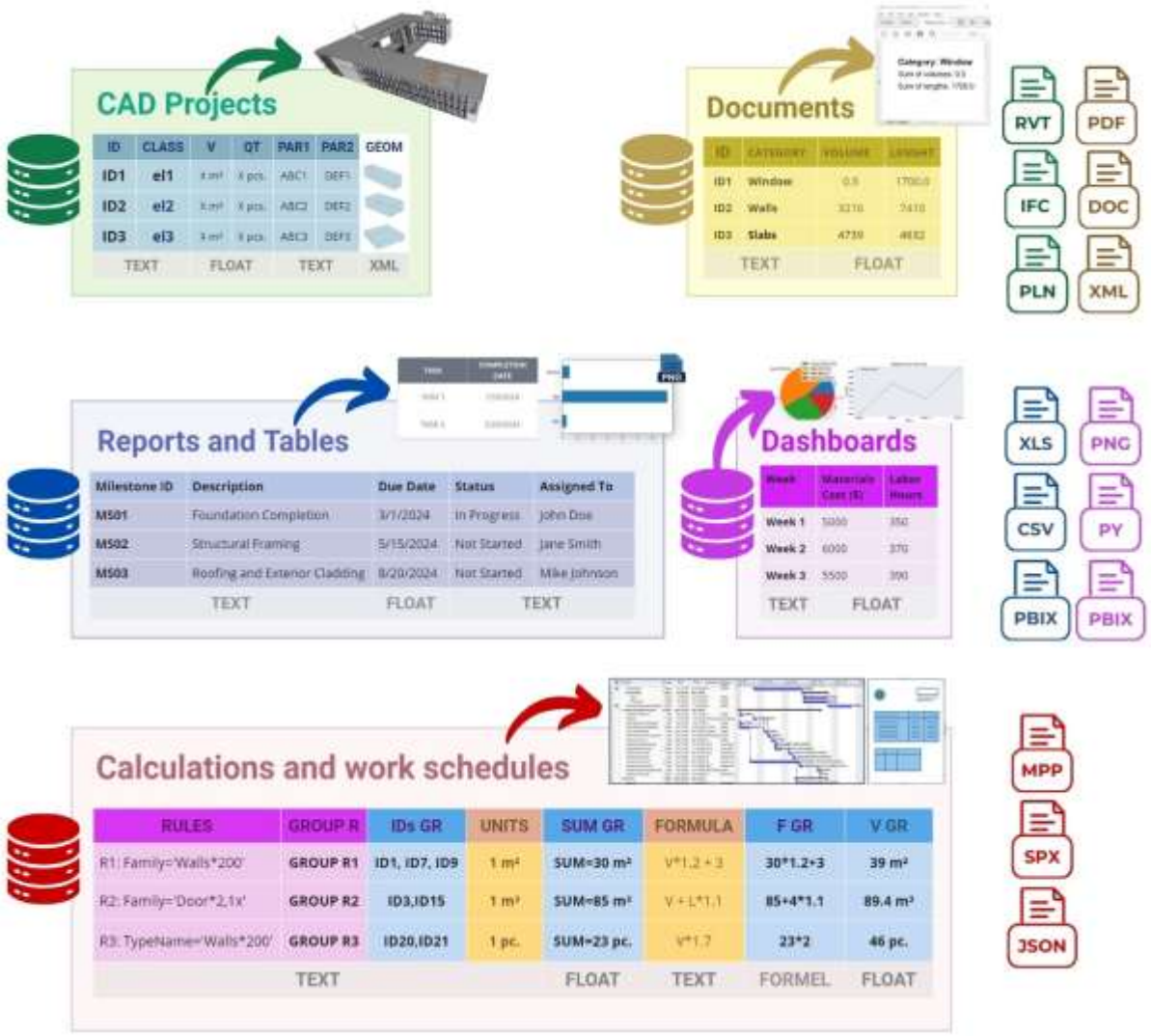


Fig. 4.4-2 Cada sistema del panorama de las empresas de construcción sirve como fuente de documentos legalmente relevantes en diversos formatos.

Las comunicaciones textuales exigen que cada profesional se familiarice con la correspondencia completa o asista regularmente a todas las reuniones para conocer el estado actual del proyecto.

Para superar esta limitación, es necesario pasar de la comunicación textual a un modelo de requisitos estructurado. Esto sólo es posible mediante el análisis sistemático, la visualización del proceso y la descripción de las interacciones en forma de diagramas de flujo y modelos de datos (Fig. 4.4-3). Al igual que en la modelización de datos (Fig. 4.3-7), pasamos del nivel contextual-ideal al conceptual añadiendo los sistemas y herramientas utilizados por los participantes y los vínculos entre ellos.

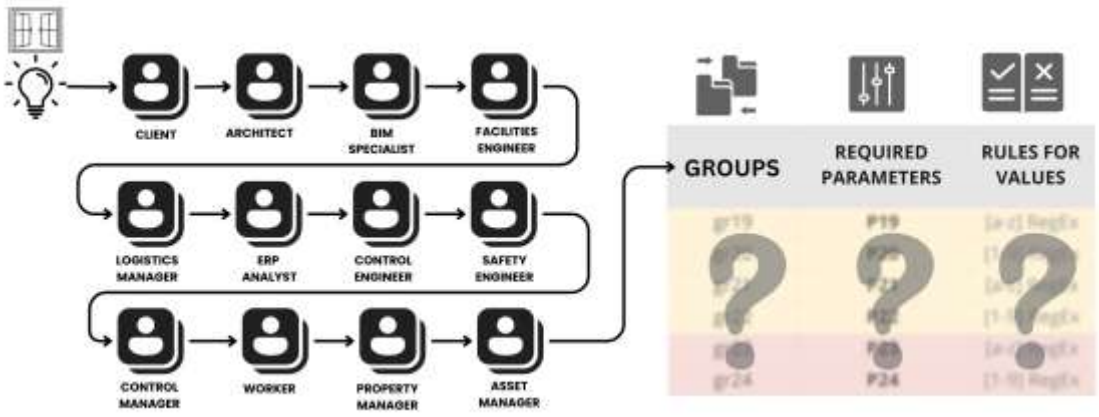


Fig. 4.4-3 Para aprender a gestionar y automatizar el proceso de validación, es necesario visualizar los procesos y estructurar los requisitos.

El primer paso para sistematizar los requisitos y las relaciones consiste en visualizar todos los vínculos y relaciones mediante diagramas de flujo conceptuales. El nivel conceptual no sólo facilitará a todos los participantes en el proceso la comprensión de toda la cadena de procesos, sino que también mostrará claramente por qué y para quién son necesarios los datos (y requisitos) en cada paso del proceso.

Organigramas de procesos y eficacia de los marcos conceptuales

Para salvar la brecha entre los enfoques tradicionales y modernos de la gestión de datos, las empresas deben pasar conscientemente de las descripciones textuales fragmentadas a las representaciones estructuradas de los procesos. La evolución de los datos -de las tablillas de arcilla a los ecosistemas digitales- requiere nuevas herramientas de pensamiento. Una de ellas es la modelización conceptual mediante diagramas de flujo. La creación de diagramas visuales -flujogramas, diagramas de procesos, diagramas de interacción- permite a los participantes en un proyecto darse cuenta de cómo sus acciones y decisiones afectan a todo el sistema de toma de decisiones.

> Si los procesos requieren no sólo almacenar datos, sino analizarlos o automatizarlos, entonces hay que empezar a abordar el tema de la creación de una capa de requisitos conceptual y visual.

En nuestro ejemplo (Fig. 4.4-1), cada especialista puede formar parte de un pequeño equipo, pero también de un departamento mayor con hasta una docena de expertos bajo el control de un director general. Cada departamento utiliza una base de datos de aplicaciones especializadas (Fig. 1.2-4 por ejemplo ERP, CAD, MEP, CDE, ECM, CPM etc.), que se actualiza regularmente con la información entrante necesaria para crear documentos, registrar el estado legal de las decisiones y gestionar los procesos.

> El proceso de transacción es similar al trabajo de los antiguos gestores de hace 4.000 años, cuando se utilizaban tablillas de arcilla y papiros para confirmar legalmente las decisiones. La diferencia entre los sistemas modernos y sus predecesores de arcilla y papel es que los métodos modernos incluyen además el proceso de convertir la información textual en formato digital para su posterior tratamiento automático en otros sistemas y herramientas.

Crear una visualización del proceso en forma de diagramas de flujo conceptuales ayudará a describir cada paso y las interacciones entre las distintas funciones, haciendo que un flujo de trabajo complejo resulte claro y sencillo.

> Visualizar los procesos de garantiza que la lógica del proceso sea transparente y accesible para todos los miembros del equipo.

El mismo proceso comunicativo para añadir una ventana a un proyecto que se describió en forma de texto, mensajes (Fig, 4.4-1) y diagrama de bloques es similar al modelo conceptual del que hablamos en el capítulo sobre modelado de datos (Fig, 4.4-4).

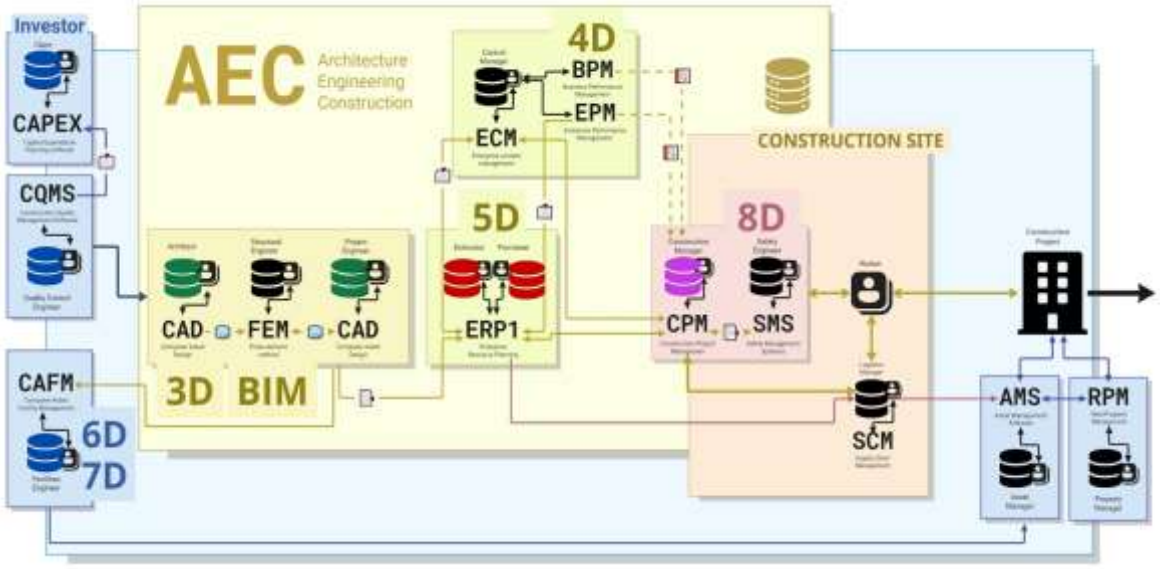


Fig. 4.4-4 El diagrama conceptual muestra a los participantes en el proyecto como usuarios de una base de datos en la que sus consultas enlazan distintos sistemas.

Aunque los diagramas conceptuales son un paso importante, muchas empresas se limitan a este nivel, creyendo que un diagrama visual es suficiente para entender los procesos. Esto crea una ilusión de manejabilidad: los directivos pueden percibir más fácilmente el panorama general en un diagrama de flujo de este tipo y ver las conexiones entre participantes y etapas. Sin embargo, tales esquemas no dan una idea clara de qué datos se necesitan para cada participante, en qué formato deben transmitirse y qué parámetros y atributos son obligatorios para la realización de la automatización. Un diagrama de flujo conceptual se parece más a una hoja de ruta: indica quién interactúa con quién, pero no revela qué se transfiere en esas interacciones.

> Aunque un proceso se describa detalladamente a nivel conceptual mediante diagramas de flujo, esto no garantiza su eficacia. La visualización suele simplificar el trabajo de los gestores, permitiéndoles seguir el proceso más fácilmente con informes paso a paso. Sin embargo, para los ingenieros de bases de datos, la representación conceptual puede no ser lo suficientemente clara y no proporcionar una comprensión clara de cómo implementar el proceso a nivel de parámetros y requisitos.

A medida que avanzamos hacia ecosistemas de datos más complejos, la implementación inicial de herramientas conceptuales y visuales se vuelve fundamental para garantizar que los procesos de datos no sólo sean eficientes, sino que también estén alineados con los objetivos estratégicos de la organización. Para trasladar plenamente este proceso de adición de ventanas (Fig, 4.4-1) al nivel de los requisitos de datos, tenemos que ir un nivel más allá y trasladar la visualización conceptual del proceso al nivel lógico y físico de los datos, los atributos necesarios y sus valores límite.

Requisitos estructurados y expresiones regulares RegEx

Hasta el 80% de los datos creados en las empresas se encuentran en formatos no estructurados o semiestructurados [52] -texto, documentos, cartas, PDF -archivos, conversaciones. Estos datos (Fig. 4.4-1) son difíciles de analizar, verificar, transferir entre sistemas y utilizar en la automatización.

Para garantizar la manejabilidad, la transparencia y la validación automática, es necesario traducir los requisitos textuales y semiestructurados a formatos bien definidos y estructurados. El proceso de estructuración no sólo afecta a los datos (que hemos tratado en detalle en los primeros capítulos de esta parte del libro), sino también a los propios requisitos, que los participantes en el proyecto suelen formular en forma de texto libre a lo largo del ciclo de vida del proyecto, a menudo sin pensar que estos procesos pueden automatizarse.

> Al igual que ya hemos convertido datos de una forma textual no estructurada a una forma estructurada, en el flujo de trabajo de requisitos convertiremos requisitos textuales a un formato estructurado de "capa lógica y física".

Como parte del ejemplo de añadir una ventana (Fig, 4.4-1), el siguiente paso es describir los requisitos de datos en forma de tabla. Estructuraremos la información de cada sistema utilizado por los participantes en el proyecto especificando los atributos clave y sus valores límite

Consideremos, por ejemplo, uno de estos sistemas (Fig. 4.4-5): el sistema de gestión de la calidad de la construcción (CQMS), que utiliza el ingeniero de control de calidad por parte del cliente. Con su ayuda comprueba si un nuevo elemento del proyecto -en este caso, una "nueva ventana"- cumple las normas y requisitos establecidos.

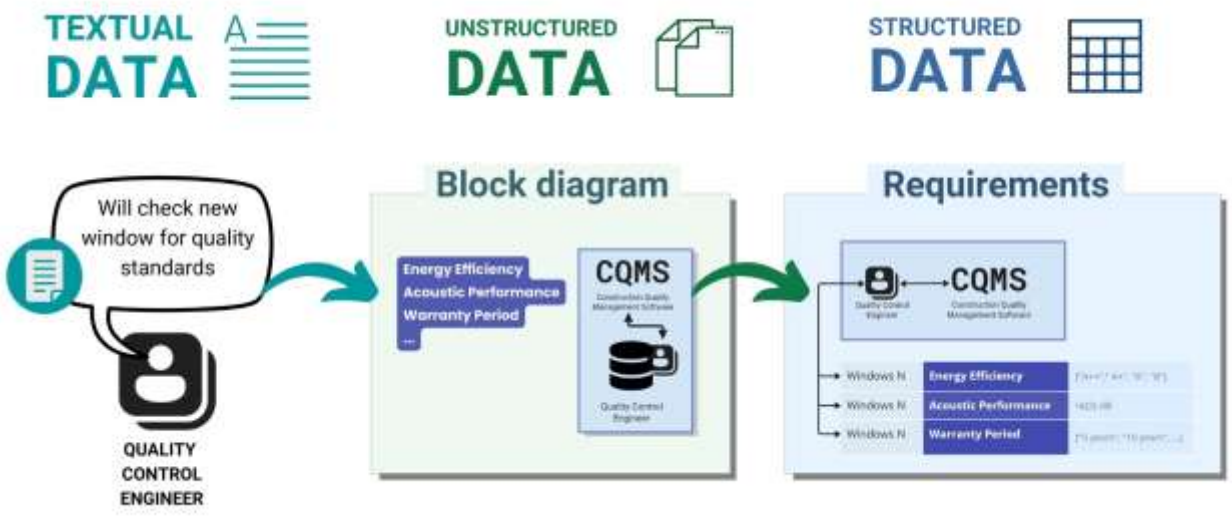


Fig. 4.4-5 Convertir los requisitos textuales en un formato de tabla con descripciones de los atributos de las entidades simplifica la comprensión a otros especialistas.

A modo de ejemplo, consideremos algunos requisitos importantes para atributos de entidades del tipo "sistemas de ventanas" en CQMS -system (Fig. 4.4-6): eficiencia energética, rendimiento acústico y periodo de garantía. Cada categoría incluye ciertas normas y especificaciones que deben tenerse en cuenta a la hora de diseñar e instalar sistemas de ventanas.

Fig. 4.4-6 El Ingeniero de Control de Calidad debe inspeccionar los nuevos elementos del Tipo Ventana en cuanto a eficiencia energética, aislamiento acústico y normas de garantía.

Los requisitos de datos que un ingeniero de control de calidad especifica en forma de tabla tienen, por ejemplo, los siguientes valores límite:

- **La clase de eficiencia energética de las ventanas** va de "A++", que denota la mayor eficiencia, a "B", considerada el nivel mínimo aceptable, y estas clases están representadas por una lista de valores aceptables **["A++", "A+", "A", "A", "A", "B"].**

- **El aislamiento acústico de las ventanas**, medido en decibelios y que muestra su capacidad para reducir el ruido de la calle, se define mediante la expresión regular \d{2}dB**.**
- **El atributo "Periodo de garantía"** para la entidad "Tipo de ventana" comienza en cinco años, estableciendo este periodo como el mínimo permitido al seleccionar un producto; también se especifican valores de periodo de garantía como **["5 años", "10 años", etc.]** o la condición lógica **">5 (años)".**

De acuerdo con los requisitos recogidos, dentro de los atributos establecidos, los nuevos elementos de categoría o clase de ventana con calificaciones inferiores a "B", como "C" o "D", no superarán el ensayo de eficiencia energética. El aislamiento acústico de las ventanas en los datos o documentos dirigidos al ingeniero de control de calidad se etiquetará con un número de dos dígitos seguido del postfijo "dB", como "35 dB" o "40 dB", y no se aceptarán valores fuera de este formato, como "9 D B" o "100 decibelios" (ya que no superarán el patrón para cadenas RegEx). El periodo de garantía debe comenzar con un mínimo de "5 años" y las ventanas con periodos de garantía más cortos, como "3 años" o "4 años", no cumplirán los requisitos que el ingeniero de calidad ha descrito en el formato de tabla.

Para cotejar estos valores de parámetros de atributos con los valores límite de los requisitos en el proceso de validación, utilizamos una lista de valores permitidos (["A", "B", "C"]), diccionarios (["A": "H1", "H2"; "B": W1", "W2"]), operaciones lógicas (por ejemplo, ">", "<", "<=", ">=" "==") para valores numéricos) y expresiones regulares (para valores de cadena y texto, como en el atributo "Rendimiento acústico"). Las expresiones regulares son una herramienta muy importante cuando se trabaja con valores de cadena.

> **Las expresiones regulares** (RegEx) se utilizan en lenguajes de programación, incluido Python (biblioteca Re), para encontrar y modificar cadenas. Regex es como un detective en el mundo de las cadenas, capaz de identificar patrones de texto con precisión.

En las expresiones regulares, las letras se describen directamente utilizando los caracteres correspondientes del alfabeto, mientras que los números pueden representarse mediante el carácter especial \d, que corresponde a cualquier dígito del 0 al 9. Los corchetes se utilizan para indicar un rango de letras o dígitos, por ejemplo [a-z] para cualquier letra minúscula del alfabeto latino o [0-9], que equivale a \d. Para los caracteres no numéricos y los que no son letras, se utilizan \D y \W respectivamente.

Casos de uso populares de RegEx (Fig, 4.4-7):

- **Verificación de la dirección de correo electrónico:** para comprobar si una cadena es una dirección de correo electrónico válida, puede utilizar la plantilla "^ **[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$**".
- **Extracción** de fecha: La plantilla "\b\d{2}d{2}d{2}d{2}d{2}d{2}d{2}d{2}d{2}d{2}**.\d{4}b**" se puede utilizar para extraer la fecha de un texto en formato DD.MM.AAAA.
- Verificación de números de teléfono: para verificar números de teléfono con el formato +49(000)000-0000, el patrón tendrá el aspecto **"\+\d{2}(\d{3})\d{3}-\d{4}".**

Al traducir los requisitos de un ingeniero de control de calidad al formato de atributos y sus valores límite (Fig. 4.4-6), los hemos transformado de su formato de texto original (conversaciones, cartas y documen-

tos normativos) a una tabla organizada y estructurada, lo que permite comprobar y analizar automáticamente cualquier dato entrante (por ejemplo, nuevos elementos de la categoría Ventana). La presencia de requisitos permite descartar automáticamente los datos que no se han comprobado y transferir automáticamente los datos comprobados a los sistemas para su tratamiento posterior.

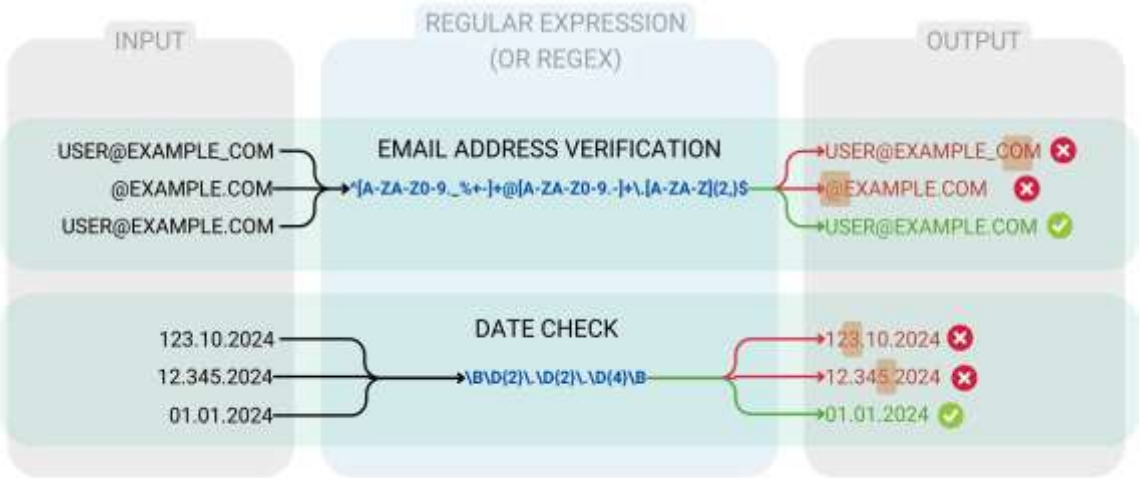


Fig. 4.4-7 El uso de expresiones regulares es una herramienta extremadamente importante en el proceso de validación de datos de texto.

Ahora, pasando del nivel conceptual al lógico del trabajo con requisitos, convertiremos todos los requisitos de todos los especialistas de nuestro proceso de instalación de una nueva ventana (Fig. 4.4-4) en una lista ordenada en formato de atributos y añadiremos estas listas con los atributos necesarios a nuestro diagrama de flujo para cada especialista (Fig. 4.4-8).

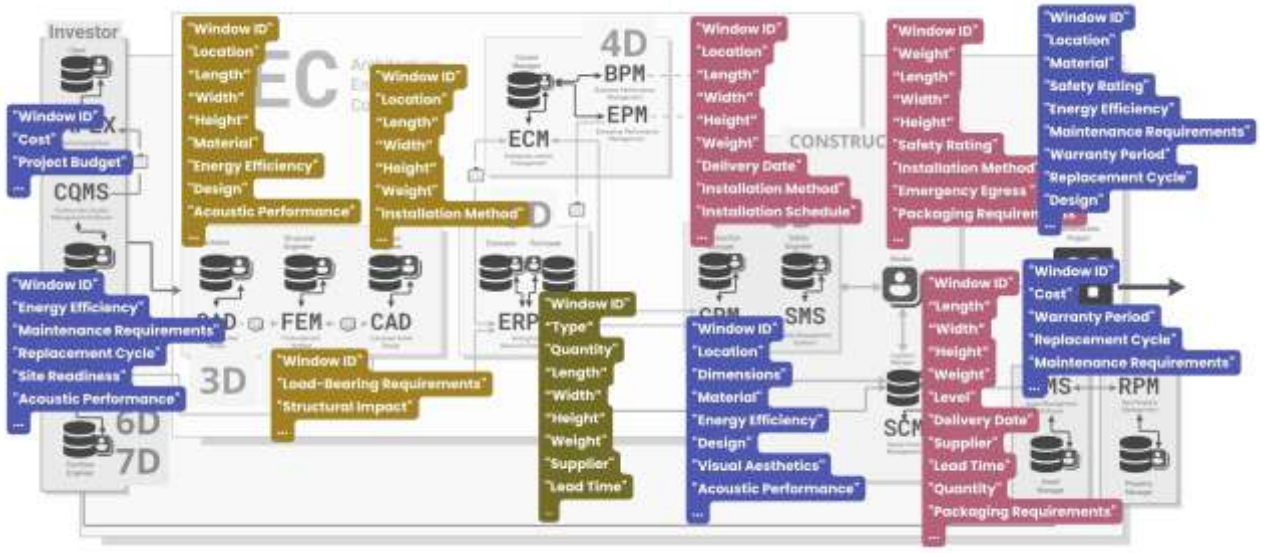


Fig. 4.4-8 A nivel del proceso lógico, los atributos que maneja cada especialista se añaden a sus respectivos sistemas.

Al añadir todos los atributos a una tabla de procesos común, transformamos la información presentada

anteriormente como texto y diálogo en el nivel conceptual (Fig. 4.4-1) en la forma estructurada y sistemática de las tablas del nivel físico (Fig. 4.4-9).

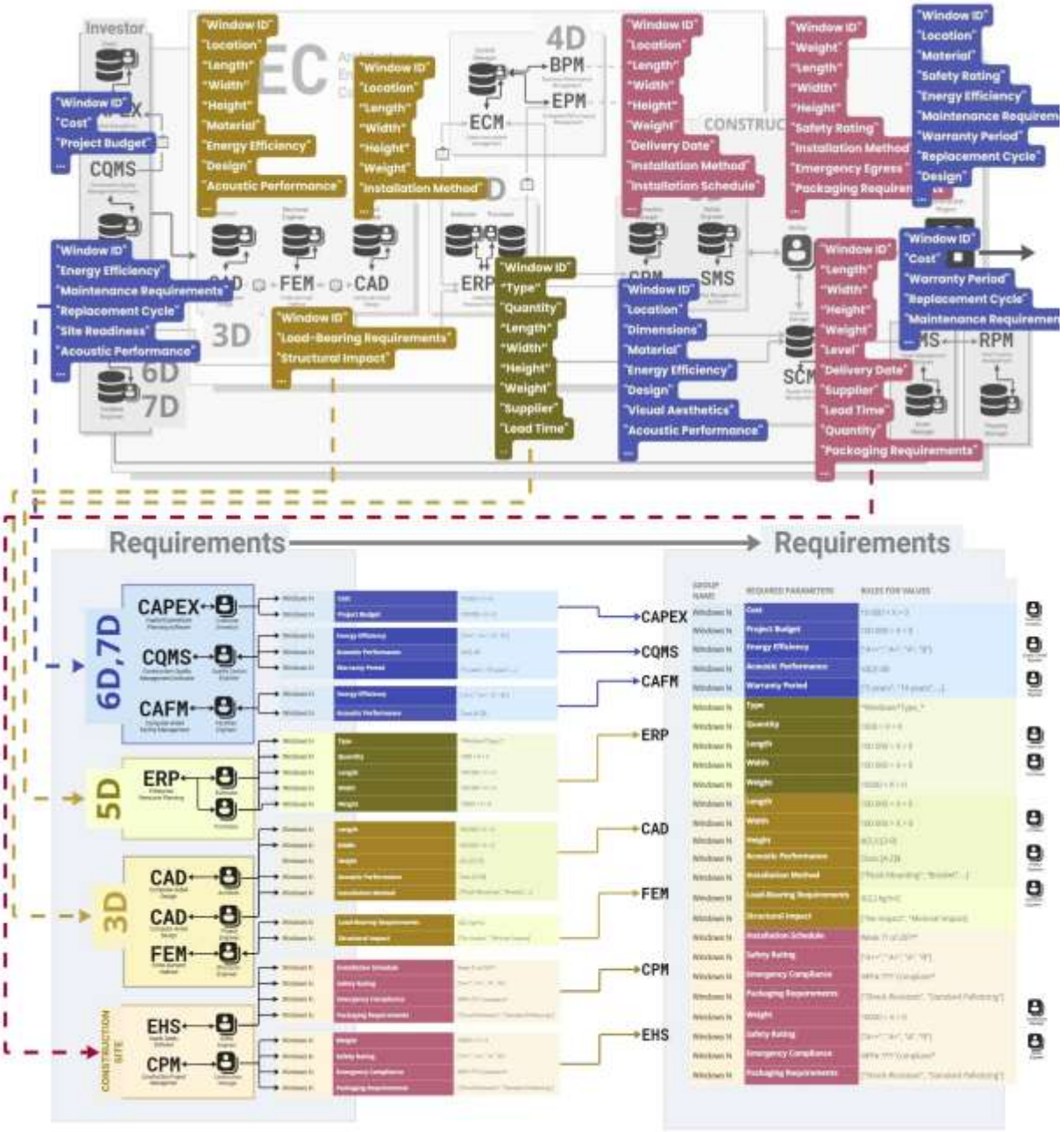


Fig. 4.4-9 Convertir el diálogo no estructurado de los especialistas en tablas estructuradas ayuda a comprender los requisitos a nivel físico.

Ahora hay que comunicar los requisitos de los datos a los especialistas que crean información para sistemas específicos. Por ejemplo, si se trabaja en una base de datos CAD, antes de empezar a modelar elementos hay que recopilar todos los parámetros necesarios en función de los escenarios de uso final de los datos. Esto suele empezar por la fase operativa, seguida de la obra, el departamento de logística, el de estimaciones, el de cálculos estructurales, etcétera. Sólo cuando haya tenido en cuenta los requisitos de

todas estas áreas, podrá empezar a crear datos basados en los parámetros recopilados. Esto le permitirá automatizar la verificación y transferencia de datos a lo largo de la cadena en el futuro.

Cuando los nuevos datos cumplen los requisitos, se integran automáticamente en el ecosistema de datos de la empresa, dirigiéndose directamente a los usuarios y sistemas para los que estaban previstos. La verificación de los datos con respecto a los atributos y sus valores garantiza que la información cumple las normas de calidad exigidas y está lista para su uso en los escenarios de la empresa.

Se han definido los requisitos de los datos y ahora, antes de que pueda comenzar la verificación, hay que crear, obtener o recopilar los datos que se van a verificar, o registrar el estado actual de la información en las bases de datos para utilizarla en el proceso de verificación.

Recogida de datos para el proceso de verificación

Antes de empezar la validación, es importante asegurarse de que los datos están disponibles en una forma adecuada para el proceso de validación. Esto no significa simplemente disponer de la información, sino prepararla: los datos deben recopilarse y transformarse de formatos no estructurados, poco estructurados, textuales y geométricos a una forma estructurada. Este proceso se describe con detalle en los capítulos anteriores, en los que se analizaron métodos para transformar distintos tipos de datos. Como resultado de todas las transformaciones, los datos entrantes adoptan la forma de tablas estructuradas abiertas (Fig. 4.1-2, Fig. 4.1-9, Fig. 4.1-13).

Con los requisitos y las tablas estructuradas con los parámetros y valores límite necesarios (Fig. 4.4-9), podemos empezar a validar los datos, ya sea como un único proceso automatizado (Pipeline) o como una validación paso a paso de cada documento entrante.

Para iniciar la comprobación, es necesario recibir un nuevo archivo como entrada o fijar el estado actual de los datos: crear una instantánea o exportar los datos actuales y entrantes, o establecer una conexión con una base de datos externa o interna. En el ejemplo que nos ocupa, dicha instantánea se crea convirtiendo automáticamente los datos CAD de a un formato estructurado registrado, digamos, a las 23:00:00 del viernes 29 de marzo de 2024, después de que todos los diseñadores se hayan ido a casa.

Fig. 4.4-10 Instantánea de la base de datos CAD (BIM) que muestra la información de atributos actual para una nueva entidad de clase "Ventana" en la versión actual del modelo de proyecto.

Gracias a las herramientas de ingeniería inversa comentadas en el capítulo "Traducción de datos CAD (BIM) a un formato estructurado", esta información procedente de diferentes herramientas y editores CAD (BIM) puede organizarse en tablas independientes (Fig. 4.4-11) o combinarse en una tabla común que combine diferentes secciones del proyecto (Fig. 9.1-10).

Dicha tabla - base de datos muestra identificadores únicos de ventanas y puertas (atributo ID), nombres de tipos (TypeName), dimensiones (Ancho, Largo), materiales (Material), así como indicadores de rendimiento energético y acústico y otras características. Una tabla de este tipo rellenada en el programa CAD (BIM) es recopilada por un ingeniero de diseño a partir de varios departamentos y documentos, formando un modelo de información del proyecto.

Fig. 4.4-11 Los datos estructurados de sistemas CAD pueden ser una tabla bidimensional con columnas que denotan atributos de elementos.

Los proyectos CAD reales (BIM) incluyen decenas o cientos de miles de elementos (Fig. 9.1-10). Los elementos dentro de los formatos CAD se categorizan automáticamente por tipo y categoría: desde ventanas y puertas hasta losas, forjados y muros. Los identificadores únicos (por ejemplo, el ID nativo, que establece automáticamente la solución CAD) o los atributos de tipo (Nombre de tipo, Tipo, Familia) permiten realizar un seguimiento del mismo objeto en diferentes sistemas. Por ejemplo, una ventana nueva en la pared norte de un edificio puede identificarse unívocamente mediante un único identificador "W-NEW" en todos los sistemas pertinentes de la organización.

Aunque los nombres e identificadores de las entidades deben ser coherentes en todos los sistemas, el conjunto de atributos y valores asociados a estas entidades puede variar significativamente en función del contexto de uso. Arquitectos, ingenieros de estructuras, profesionales de la construcción, la logística y el mantenimiento de inmuebles perciben los mismos elementos de distintas maneras. Cada uno de ellos se basa en sus propios clasificadores, normas y objetivos: algunos miran una ventana desde un punto de vista puramente estético, evaluando su forma y proporciones, mientras que otros lo hacen desde un punto de vista de ingeniería u operativo, analizando la conductividad térmica, el método de instalación, el peso o los requisitos de mantenimiento. Por eso, a la hora de modelizar los datos y describir los elementos, es importante tener en cuenta la versatilidad de su uso y garantizar la coherencia de los datos, teniendo en cuenta al mismo tiempo las especificidades del sector.

Para cada función en los procesos de la empresa existen bases de datos especializadas con su propia interfaz de usuario, desde el diseño y los cálculos hasta la logística, la instalación y la explotación del edificio (Fig. 4.4-12). Cada uno de estos sistemas está gestionado por un equipo profesional de especialistas a través de una interfaz de usuario especial o mediante consultas a la base de datos, donde detrás de la suma de todas las decisiones tomadas sobre los valores introducidos al final de la cadena se encuentra el director del sistema o director de departamento, que es responsable de la validez legal y la calidad de los datos introducidos ante sus homólogos que prestan servicio a otros sistemas.

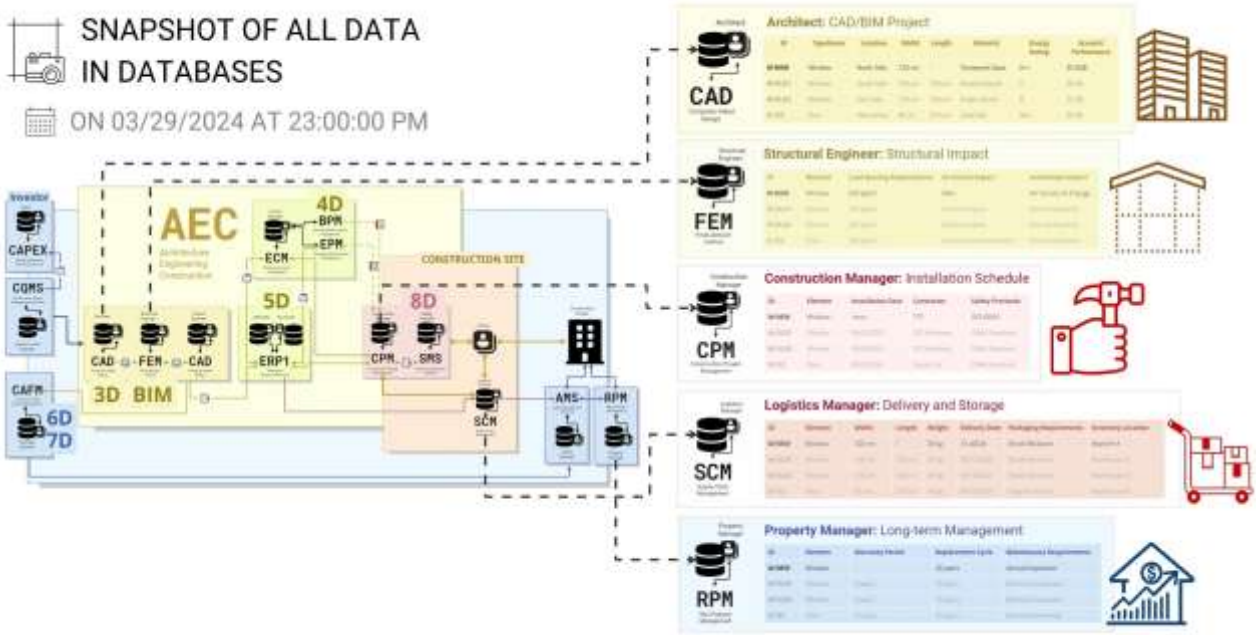


Fig. 4.4-12 La misma entidad tiene el mismo identificador en diferentes sistemas, pero diferentes atributos que sólo son importantes en ese sistema.

Una vez organizada la recopilación de requisitos y datos estructurados a nivel lógico y físico, nos queda establecer un proceso para cotejar automáticamente los datos de los distintos documentos entrantes y los distintos sistemas con los requisitos recopilados previamente.

Verificación de los datos de y resultados de la verificación

Todos los datos nuevos que se introduzcan en el sistema -ya sean documentos, tablas o entradas en la base de datos procedentes del cliente, el arquitecto, el ingeniero, el capataz, el responsable de logística o el administrador de la propiedad- deben validarse con respecto a los requisitos formulados anteriormente (Fig. 4.4-9). El proceso de validación es fundamental: cualquier error en los datos puede dar lugar a cálculos incorrectos, retrasos en los plazos e incluso pérdidas económicas. Para minimizar estos riesgos, debe organizarse un procedimiento de validación de datos sistemático, repetible e iterativo.

> Para validar los nuevos datos que se introducen en el sistema (no estructurados, textuales o geométricos), hay que convertirlos a un formato poco estructurado o estructurado. A continuación, el proceso de validación debe cotejar los datos con una lista completa de atributos requeridos y sus valores permitidos.

La conversión de distintos tipos de datos: texto, imágenes, documentos PDF y datos mixtos CAD (BIM) a un formato estructurado se trató en detalle en el capítulo "Conversión de datos a un formato estructurado".

Un ejemplo es una tabla obtenida de un proyecto CAD (BIM) (Fig. 4.4-11). Incluye datos geométricos semi-estructurados e información estructurada de atributos sobre entidades del proyecto (Fig. 3.1-14) - por ejemplo, un elemento de la clase "Ventanas".

Para realizar la validación, comparamos los valores de los atributos (Fig. 4.4-11) con los valores límite de referencia que han definido los expertos en forma de requisito (Fig. 4.4-9). La tabla de comparación final (Fig. 4.4-13) dejará claro qué valores son aceptables y cuáles deben corregirse antes de que los datos puedan utilizarse fuera de las aplicaciones CAD (BIM).

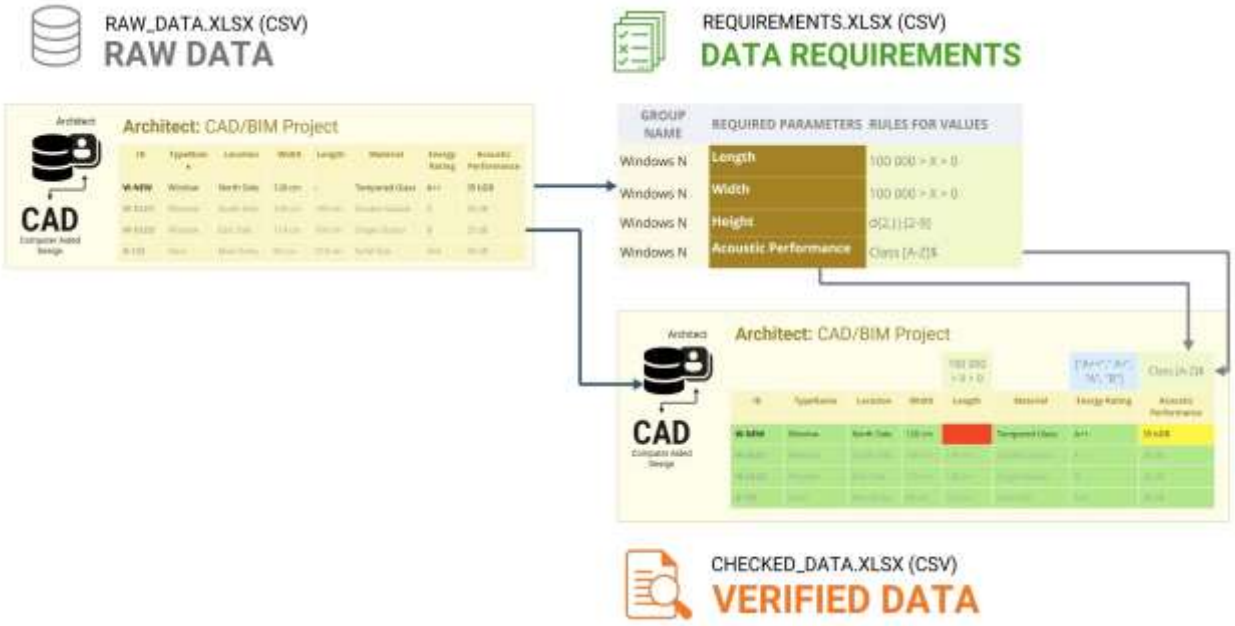


Fig. 4.4-13 La tabla de validación final destaca aquellos valores de atributo para la nueva entidad de clase "Windows" a los que debe prestar atención.

Implementando una solución similar utilizando la biblioteca Pandas, que describimos anteriormente en el capítulo "Pandas: Una herramienta indispensable para el análisis de datos", validaremos los datos de un archivo tabular extraído de un archivo CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Fig, 4.4-11) utilizando los requisitos de otro archivo tabular de requisitos (Fig, 4.4-9).

Para obtener el código, necesitamos describir en el prompt para LLM que necesitamos cargar los datos del archivo **raw_data.xlsx** (un conjunto completo de datos de la base de datos CAD (BIM)), **comprobarlos** y guardar el resultado en un nuevo archivo **checked_data.xlsx** (Fig. 4.4-13).

Vamos a obtener el código utilizando LLM sin mencionar la biblioteca Pandas:

> Escriba código para comprobar la tabla desde el archivo raw_data.xlsx y validarlos utilizando las siguientes reglas de validación: los valores de las columnas 'Width' y 'Length' son mayores que cero, 'Energy Rating' está incluido en la lista ['A++', 'A+', 'A', 'B'], y 'Acoustic Performance' como una variable que especificaremos más tarde - con la columna de validación final añadida, y guarde la tabla final en un nuevo archivo Excel checked_data.xlsx ⏎

La respuesta de LLM describirá un breve ejemplo de código Python que puede refinarse y aumentarse con las siguientes indicaciones:

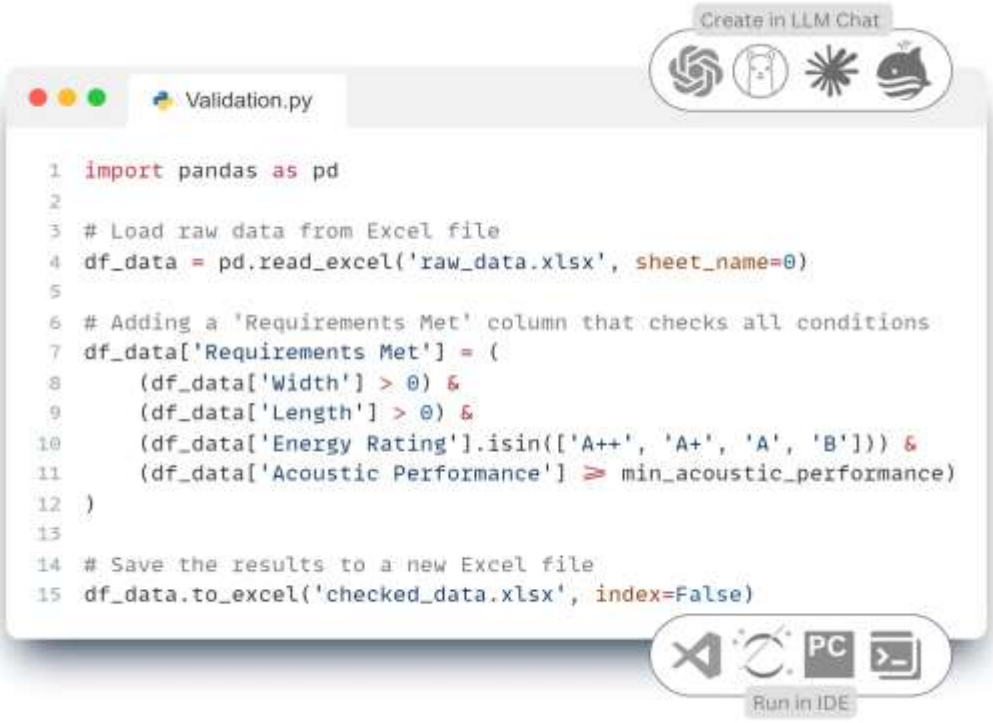


Fig. 4.4-14 El código generado por el modelo LLM- comprueba el diseño CAD (BIM) convertido comparándolo con los requisitos de atributos en forma de valores límite.

El código generado por el modelo de lenguaje LLM, se puede utilizar en cualquier IDE popular o herramienta online: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con plugin Python, JupyterLab o herramientas online populares Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

La ejecución del código (Fig. 4.4-14) mostrará que los "elementos de entidad" W-OLD1, W-OLD2, D-122 (y otros elementos de) de la base de datos CAD (BIM) cumplen los requisitos de atributo: la anchura y la longitud son mayores que cero, y la clase de eficiencia energética es uno de los valores de la lista "A++", "A", "B", "C" (Fig. 4.4-15).

El elemento W-NEW que necesitamos y que hemos añadido recientemente, responsable de la nueva clase de elemento "Ventana" en el lado norte, no es conforme (atributo "Requrments Met") porque su longitud es cero (un valor "0.0" se considera inaceptable por nuestra regla "Width'>0") y no especifica una clase de eficiencia energética.

CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Fig. 4.4-15 Comprobación identifica las entidades que no han pasado por el proceso de verificación y añade un nuevo atributo con los valores "Falso" o "Verdadero" a los resultados.

Del mismo modo, comprobamos la coherencia de todos los elementos del proyecto (entidades) y los atributos necesarios para cada uno de los sistemas, tablas o bases de datos en todos los datos que recibimos de distintos especialistas (Fig, 4.4-1) durante el proceso de añadir una ventana al proyecto.

En la tabla final es conveniente resaltar en color los resultados de la comprobación para su visualización: en verde se marcan los atributos que han superado la comprobación, en amarillo los valores con desviaciones no críticas y en rojo las discrepancias críticas (Fig. 4.4-16).

Como resultado de la validación (Fig, 4.4-16), obtenemos una lista de elementos de confianza y validados con sus identificadores que se ha comprobado que cumplen los requisitos de atributos. Los elementos validados proporcionan la garantía de que estos elementos cumplen las normas y especificaciones establecidas para todos los sistemas implicados en el proceso de adición de elementos de la clase Ventana o de cualquier otra clase (hablaremos más sobre la automatización de la validación de datos y la creación de un proceso ETL automatizado en el capítulo "Automatización de ETL y validación de datos").

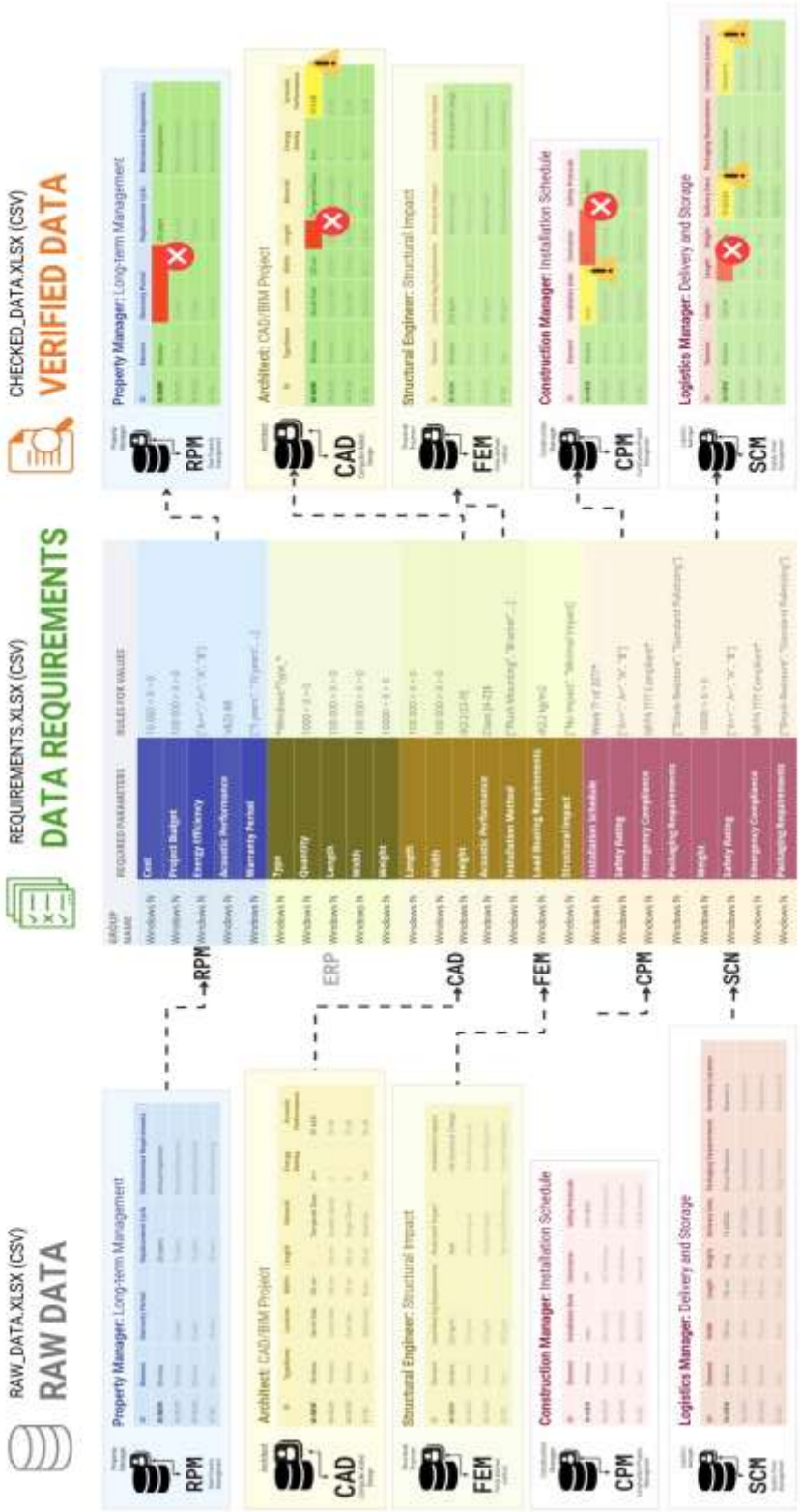


Fig. 4.4-16 El resultado de la comprobación realizada para todos los sistemas permite determinar qué datos no cumplen los requisitos de la empresa.

Las entidades que se han validado con éxito no suelen requerir mucha atención. Pasan sin obstáculos a las siguientes fases de tratamiento e integración en otros sistemas. A diferencia de los elementos "de calidad", son los elementos que no superan la validación los que revisten mayor interés. La información sobre estas desviaciones es fundamental: debe comunicarse no sólo en informes tabulares, sino también utilizando diversas herramientas de visualización. Una representación gráfica de los resultados de la verificación ayuda a evaluar rápidamente el estado general de la calidad de los datos, identificar las áreas problemáticas y tomar rápidamente medidas correctivas o correctoras.

Visualización de los resultados de la verificación

Visualización es una herramienta esencial para interpretar los resultados de las inspecciones. Además de las habituales tablas resumen, puede incluir paneles informativos, diagramas y documentos PDF generados automáticamente que agrupan los elementos del proyecto según su estado de inspección. La codificación por colores puede desempeñar aquí un papel de apoyo: el verde puede indicar elementos que se han validado con éxito, el amarillo puede indicar elementos que requieren atención adicional y el rojo puede indicar elementos que tienen errores críticos o en los que faltan datos clave.

En nuestro ejemplo (Fig. 4.4-1), analizamos los datos de cada sistema paso a paso: desde el CAD (BIM) y la gestión de la propiedad hasta la logística y los programas de instalación (Fig. 4.4-16). Tras la auditoría, se generan automáticamente alertas individuales o documentos de informe para cada especialista, por ejemplo, en formato PDF (Fig. 4.4-17). Si los datos son correctos, el especialista recibe un breve mensaje: "Gracias por colaborar". Si se encuentran discrepancias, se envía un informe detallado con el texto: "Este documento enumera elementos, sus identificadores, atributos y valores cuya conformidad no se ha comprobado".

Fig. 4.4-17 La validación y la generación automática de informes aceleran el proceso de búsqueda y comprensión de las deficiencias de los datos para el profesional que los crea.

Gracias al proceso de validación automatizado, en cuanto se detecta un error o una laguna de datos, se envía una notificación instantánea en forma de mensaje de chat, correo electrónico o documento PDF a la persona responsable de crear o procesar las entidades pertinentes y sus atributos (Fig. 4.4-18), con una lista de elementos y descripciones de atributos que no se han validado.

Fig. 4.4-18 Los informes de inspección automáticos facilitan la comprensión de los errores y agilizan el trabajo de completar los datos del proyecto.

Por ejemplo, si un documento llega al sistema de gestión de la propiedad (tras su estructuración) con el atributo "Periodo de garantía" incorrectamente cumplimentado, el administrador de la propiedad recibe una alerta con una lista de atributos que deben comprobarse y corregirse.

Del mismo modo, cualquier deficiencia en el programa de instalación o en los datos logísticos hará que se genere un informe automático y que, por ejemplo, se envíe una notificación por chat o un correo electrónico con los resultados de la inspección al especialista correspondiente.

Además de los documentos PDF y los gráficos con los resultados, es posible crear cuadros de mando y modelos interactivos 3D (Fig. 7.1-6, Fig. 7.2-12) en los que se destacan los elementos a los que faltan atributos, lo que permite a los usuarios utilizar visualmente las geometrías 3D de los elementos para filtrar y evaluar la calidad e integridad de estos elementos en el proyecto.

> La visualización de los resultados de las inspecciones en en forma de documentos, gráficos o cuadros de mando generados automáticamente simplifica enormemente la interpretación de los datos y facilita una comunicación eficaz entre los participantes en el proyecto.

El proceso de comprobación automática de datos procedentes de diversos sistemas y fuentes de información puede compararse a la toma de decisiones informadas en la vida cotidiana. Al igual que las empresas del sector de la construcción tienen en cuenta muchas variables -desde la fiabilidad de los datos de entrada hasta su repercusión en los plazos, el coste y la calidad de realización del proyecto-, al tomar decisiones importantes, por ejemplo, al elegir un lugar para vivir, una persona sopesa toda una serie de factores: accesibilidad del transporte, infraestructuras, coste, seguridad, calidad de vida. Todas estas consideraciones forman un sistema de criterios que constituyen la base de las decisiones finales que conforman nuestras

vidas.

Comparación de los controles de calidad de los datos con las necesidades de la vida humana

A pesar de la constante evolución de los métodos y herramientas de control de la calidad de los datos, el principio fundamental de la conformidad de la información permanece inalterado. Este principio está integrado en los cimientos de un sistema de gestión maduro, ya sea en la empresa o en la vida cotidiana.

> El proceso de validación iterativa de datos se parece mucho al proceso de toma de decisiones al que todos nos enfrentamos a diario. En ambos casos, nos basamos en la experiencia, los datos y la nueva información disponible. Y cada vez son más las decisiones vitales y profesionales -desde las estratégicas hasta las cotidianas- que se toman basándose en datos.

Por ejemplo, al elegir un lugar para vivir o una pareja para toda la vida, formamos intuitivamente en nuestra mente una tabla de criterios y características con los que comparamos alternativas (Fig. 4.4-19). Estas características -ya sean cualidades personales de una persona o parámetros de una propiedad- representan atributos que influyen en la decisión final.

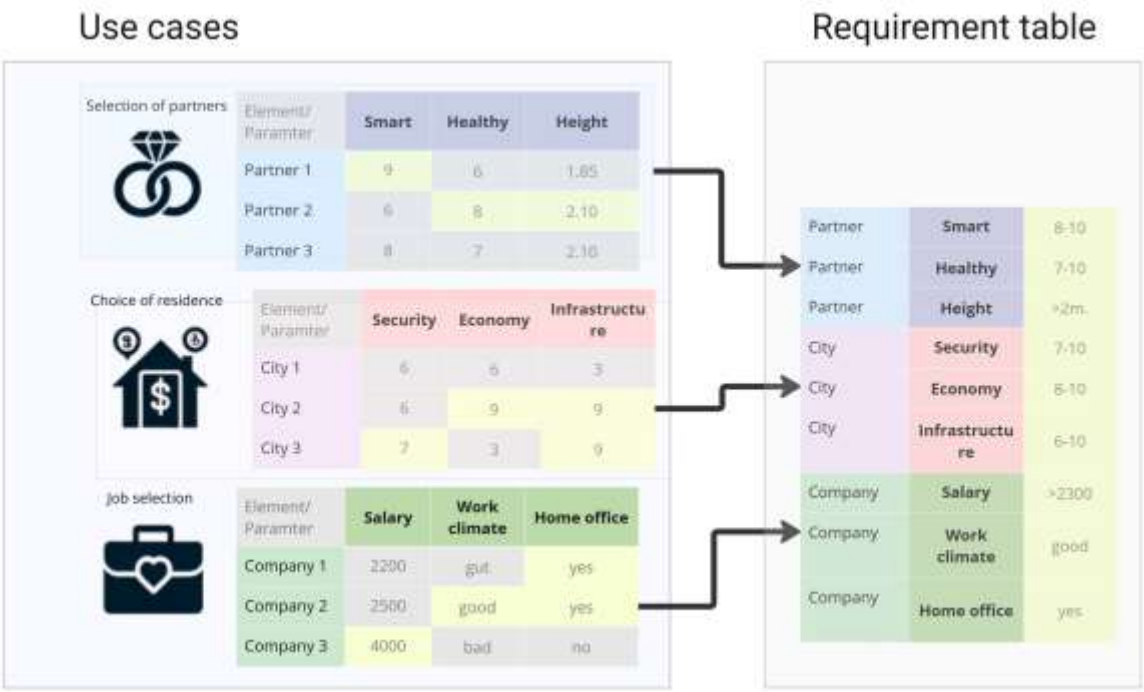


Fig. 4.4-19 La elección de residencia, empleo o asociación se basa en los requisitos de atributos individuales.

El uso de datos estructurados y de un enfoque formalizado para describir los requisitos (Fig, 4.4-20) contri-

buye a tomar decisiones más informadas y fundamentadas tanto en la vida profesional como en la personal.

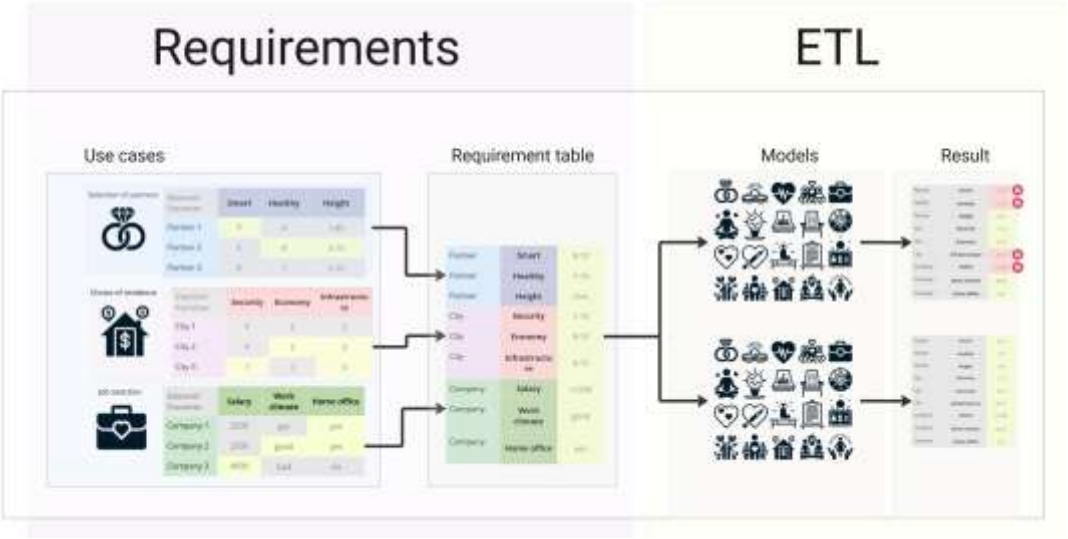


Fig. 4.4-20 La formalización de los requisitos permite sistematizar la percepción de las decisiones vitales y empresariales.

El enfoque de toma de decisiones basado en datos no es exclusivamente una herramienta empresarial. También se integra perfectamente en la vida cotidiana, siguiendo pasos comunes de procesamiento de datos (Fig. 4.4-21) similares al proceso ETL (Extraer, Transformar, Cargar) que ya hemos tratado al principio de esta parte al estructurar los datos y que examinaremos en detalle en el contexto de la automatización de tareas en la séptima parte del libro:

- **Los datos como base (Extracto**): En cualquier ámbito -ya sea el trabajo o la vida personal- recopilamos información. En la empresa, pueden ser informes, cifras, datos de mercado; en la vida personal, pueden ser experiencias personales, consejos de seres queridos, comentarios, observaciones.
- **Criterios de evaluación (Transform): la** información recogida se interpreta en función de criterios predefinidos. En el trabajo, son indicadores de rendimiento (KPI), limitaciones presupuestarias y normas; en la vida personal, parámetros como el precio, la comodidad, la fiabilidad, el carisma, etc.
- **Predicción y análisis de riesgos (Load):** la última etapa consiste en la toma de decisiones basada en el análisis de los datos transformados y la comparación de las posibles consecuencias. Es similar a los procesos empresariales, en los que los datos pasan por una lógica empresarial y un filtro de riesgos.

> Las decisiones que tomamos -desde preferencias triviales como qué desayunar hasta acontecimientos vitales importantes como elegir carrera o pareja- son intrínsecamente el resultado de procesar y evaluar datos.

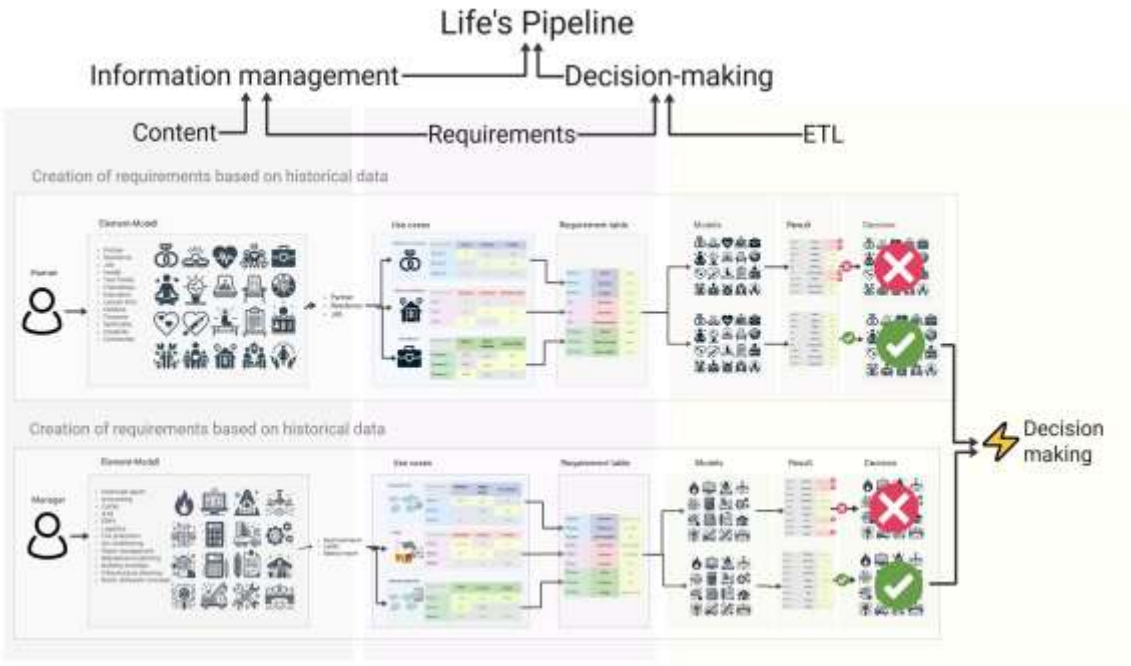


Fig. 4.4-21 Los negocios y la vida en general son una serie de decisiones basadas en datos en las que la calidad de los datos utilizados para tomar decisiones es un factor clave.

Todo en nuestras vidas está interconectado, y al igual que los organismos vivos, incluidos los humanos, siguen las leyes de la naturaleza, evolucionando y adaptándose a las condiciones cambiantes, los procesos humanos, incluida la forma en que recopilamos y analizamos los datos, reflejan estos principios naturales. La estrecha relación entre la naturaleza y la actividad humana confirma no sólo nuestra dependencia de la naturaleza, sino también nuestro deseo de aplicar leyes perfeccionadas por millones de años de evolución para crear arquitecturas de datos, procesos y sistemas para la toma de decisiones.

Las nuevas tecnologías, especialmente en la construcción, son un excelente ejemplo de cómo la humanidad se inspira en la naturaleza una y otra vez para crear soluciones mejores, más sostenibles y eficientes.

Próximos pasos: convertir los datos en cálculos y planes precisos

En esta parte hemos visto cómo convertir datos no estructurados en un formato estructurado, desarrollar modelos de datos y organizar procesos para comprobar la calidad de la información en proyectos de construcción. La gestión, normalización y clasificación de datos es un proceso fundamental que requiere un enfoque sistemático y una comprensión clara de los requisitos empresariales. Las técnicas y herramientas analizadas en esta parte permiten una integración fiable entre distintos sistemas a lo largo de todo el ciclo de vida de un objeto.

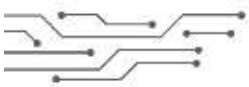
Para resumir esta parte, destaquemos los principales pasos prácticos que le ayudarán a aplicar los enfoques comentados en sus tareas diarias:

- Empiece por sistematizar los requisitos

- ☐ Cree un registro de atributos y parámetros para los elementos clave de sus proyectos y procesos
- ☐ Documentar los valores límite de cada atributo
- ☐ Visualizar procesos y relaciones entre clases, sistemas y atributos mediante diagramas de flujo (por ejemplo, en Miro, Canva o Visio).

- Automatizar la conversión de datos
  - ☐ Compruebe cuáles de sus documentos de uso frecuente en los procesos pueden digitalizarse mediante bibliotecas OCR y convertirse en tablas.
  - ☐ Consulte las herramientas de ingeniería inversa para extraer datos de CAD (BIM)
  - ☐ Intente configurar la recuperación automática de datos desde documentos o formatos que utilice con frecuencia en su trabajo a una hoja de cálculo.
  - ☐ Establezca conversiones automáticas entre distintos formatos de datos
- Crear una base de conocimientos para la categorización
  - ☐ Desarrollar un clasificador de elementos interno o utilizar uno ya existente en que se ajuste a las normas del sector.
  - ☐ Documentar las interrelaciones entre los distintos sistemas de clasificación
  - ☐ Debatir con su equipo el tema de la utilización de un sistema unificado de identificación y clasificación inequívoca de los elementos.
  - ☐ Empezar a crear un proceso para validar automáticamente los datos, tanto los que se manejan dentro del equipo como los que se transmiten a sistemas externos.

Utilizando estos enfoques, puede mejorar notablemente la calidad de sus datos y simplificar su posterior tratamiento y transformación. En las siguientes partes de este libro, veremos cómo aplicar datos ya estructurados y preparados para cálculos automatizados, estimación de costes, programación y gestión de proyectos de construcción.



V PARTE

CÁLCULO DE COSTES Y PLAZOS: INCORPORACIÓN DE DATOS A LOS PROCESOS DE CONSTRUCCIÓN

La quinta parte está dedicada a los aspectos prácticos de la utilización de datos para optimizar el cálculo de costes y la planificación de proyectos de construcción. Se analizan en detalle el método basado en los recursos de estimación de costes y la automatización de los procesos de estimación. Se estudian métodos de adquisición automatizada de Quantity Take-Off) a partir de modelos CAD (BIM)y su integración con sistemas de cálculo. Se exploran las tecnologías de modelización 4D y 5D para la programación temporal y la gestión de costes de construcción, con ejemplos concretos de su aplicación. Se presenta un análisis de las capas de información ampliadas 6D -8D, que ofrecen un enfoque integrado para evaluar la sostenibilidad, el funcionamiento y la seguridad de los objetos inmobiliarios. Se analizan en detalle los métodos de cálculo de la huella de carbono y los indicadores ESG de los proyectos de construcción en el contexto de los requisitos y normas medioambientales modernos. Se evalúan críticamente las posibilidades y limitaciones de los sistemas tradicionales ERP y PMIS en la gestión del proceso de construcción, con un análisis de su impacto en la transparencia de los precios. Se pronostican perspectivas para la transición de soluciones cerradas a estándares abiertos y herramientas flexibles de análisis de datos capaces de garantizar una mayor eficiencia de los procesos de construcción.

CAPÍTULO 5.1.
CÁLCULOS Y ESTIMACIONES DE COSTES PARA PROYECTOS DE CONSTRUCCIÓN

Conceptos básicos de la construcción: estimación de cantidades, costes y plazos

Entre los muchos procesos empresariales que determinan la sostenibilidad de una empresa del sector de la construcción, revisten especial importancia -como hace miles de años- los procesos para estimar con precisión el número de elementos, el coste del proyecto y el plazo de entrega (Fig. 5.1-1).

El desarrollo de la escritura fue el resultado de un conjunto de factores, entre ellos la necesidad de registrar las transacciones económicas, el comercio y la gestión de los recursos en las sociedades primitivas. Los primeros documentos jurídicamente significativos, tablillas de arcilla con cálculos de costes de materiales y mano de obra, se utilizaron en el contexto del comercio y la construcción. Estas tablillas registraban las obligaciones de las partes en la construcción de estructuras y se conservaban como prueba de acuerdos y relaciones monetarias y comerciales.

> Durante miles de años, el enfoque de la estimación ha permanecido prácticamente inalterado: los cálculos se realizaban manualmente, confiando en la experiencia y la intuición del ingeniero estimador. Sin embargo, con la llegada de los sistemas ERP modulares y las herramientas CAD, el enfoque tradicional para estimar la cantidad, el coste y el tiempo empezó a transformarse rápidamente. Las tecnologías digitales actuales permiten automatizar por completo los cálculos clave de tiempo y coste, lo que hace posible una mayor precisión, rapidez y transparencia en la planificación de recursos de los proyectos de construcción.

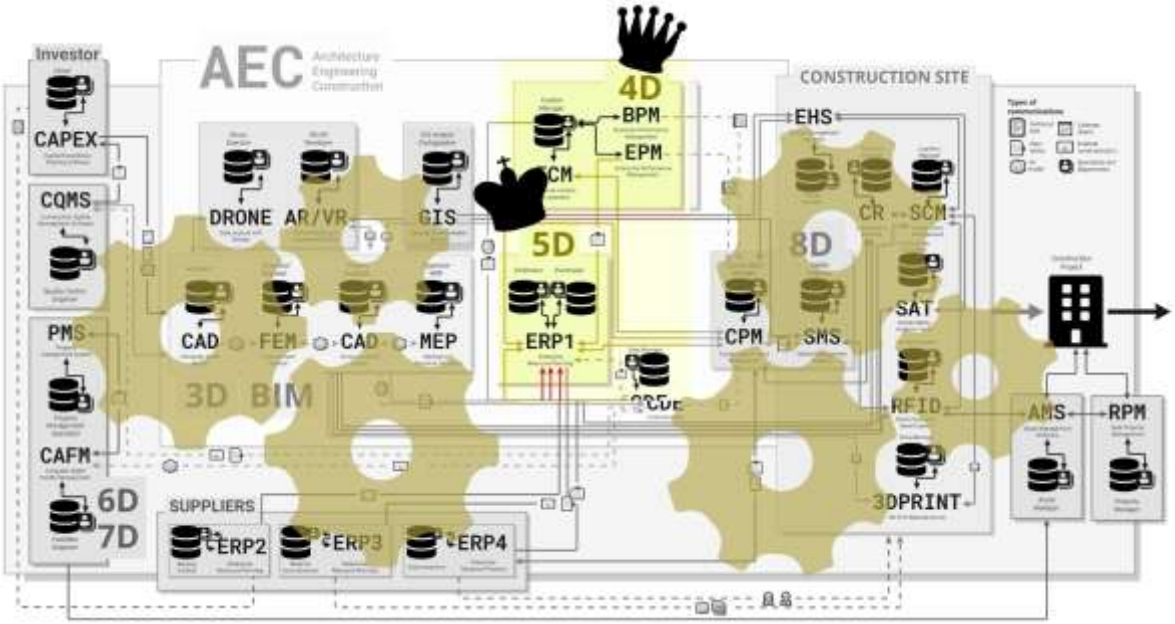


Fig. 5.1-1 De los muchos sistemas diferentes, las herramientas responsables de las métricas de volumen, coste y tiempo son las más importantes en las empresas.

Las empresas de construcción se centran en datos precisos sobre plazos y costes. Estos, a su vez, dependen de la cantidad de materiales y mano de obra utilizados, y su transparencia afecta a la rentabilidad. Sin embargo, la complejidad de los procesos de cálculo y su falta de transparencia suelen provocar un aumento de los costes de los proyectos, el incumplimiento de los plazos e incluso la quiebra.

> Según el informe de KPMG "Problemas conocidos - Nuevos enfoques" (2023), sólo el 50% de los proyectos de construcción se terminan a tiempo, y el 87% de las empresas señalan un mayor control sobre la economía de los proyectos de capital. Los principales problemas están relacionados con la falta de personal cualificado y la dificultad de prever los riesgos [2].

Los datos históricos sobre costes y tiempos de proceso se recopilan durante la construcción de proyectos anteriores a lo largo de la vida de la empresa constructora y se introducen en bases de datos de diversos sistemas (ERP, PMIS BPM, EPM, etc.).

> Disponer de datos históricos de costes de buena calidad es una ventaja competitiva importante para una organización de construcción, ya que afecta directamente a su supervivencia.

Los departamentos de estimación y cálculo de costes de las empresas de construcción e ingeniería se crean para recopilar, almacenar y actualizar datos históricos sobre cálculos de proyectos. Su función principal es acumular y sistematizar la experiencia de la empresa, lo que permite mejorar con el tiempo la precisión de la estimación del alcance, el calendario y el coste de los nuevos proyectos. Este enfoque ayuda a minimizar los errores en los cálculos futuros basados en la práctica y los resultados de proyectos ya ejecutados.

Métodos de cálculo del coste estimado de los proyectos

Los especialistas en cálculo de costes utilizan diversos métodos de estimación, cada uno de ellos centrado en un tipo específico de datos, disponibilidad de información y nivel de detalle del proyecto. Entre los más comunes figuran:

- **Método basado en los recursos**: estimación del coste estimado de un proyecto a partir de un análisis detallado de todos los recursos necesarios, como materiales, equipos y mano de obra. Este método requiere una lista detallada de todas las tareas y recursos necesarios para realizar cada tarea, seguida de un cálculo de su coste. Este método es muy preciso y se utiliza ampliamente en la estimación de costes.
- **Método paramétrico**: utiliza modelos estadísticos para estimar los costes en función de los parámetros del proyecto. Puede consistir en analizar el coste por unidad de medida, como la superficie del edificio o el alcance de las obras, y adaptar estos costes a las condiciones específicas del proyecto. Este método es especialmente eficaz en las primeras fases, cuando aún no se dispone de información detallada.
- Método unitario (**método de** costes unitarios): calcula el coste estimado de un proyecto basándose en un coste por unidad de medida (por ejemplo, por metro cuadrado o metro cúbico). Permite comparar y analizar rápida y fácilmente el coste de distintos proyectos o partes de proyectos.
- Juicio de **expertos** (método Delphi): se basa en las opiniones de expertos que utilizan su experiencia y conocimientos para estimar el valor de un proyecto. Este método es útil cuando no se dispone de datos de referencia precisos o el proyecto es único.

Cabe señalar que el método paramétrico y los juicios de expertos pueden adaptarse a modelos de aprendizaje automático. Esto permite generar automáticamente previsiones de costes y plazos del proyecto a partir de muestras de entrenamiento. En el capítulo "Ejemplo de utilización del aprendizaje automático para hallar el coste y el calendario del proyecto" (Fig. 9.3-5) se tratan con más detalle ejemplos de aplicación de tales modelos.

No obstante, el método basado en los recursos sigue siendo el más popular y utilizado en la práctica mundial. No sólo proporciona una evaluación precisa del coste estimado, sino que también permite calcular la duración de los procesos individuales en la obra y de todo el proyecto en su conjunto (más detalles en el capítulo "Cronogramas de construcción y 4D -datos del proyecto").

Método basado en los recursos cálculo de costes y estimaciones en la construcción

El cálculo de costes basado en los recursos es un método de contabilidad de gestión en el que el coste de un proyecto se basa en una contabilidad directa de todos los recursos implicados. En la construcción, este enfoque implica un análisis y una evaluación detallados de todos los recursos materiales, laborales y técnicos necesarios para realizar la obra.

El método basado en los recursos, ofrece un alto grado de transparencia y precisión en la planificación presupuestaria, ya que se centra en los precios reales de los recursos en el momento de la estimación. Esto es especialmente importante en un entorno económico inestable en el que las fluctuaciones de precios pueden afectar significativamente al coste global de un proyecto.

En los capítulos siguientes analizaremos en detalle el proceso de cálculo de costes basado en los recursos. Para entender mejor sus principios en la construcción, haremos una analogía con el cálculo de costes de una cena en un restaurante. El gerente del restaurante, al planificar la velada, hace una lista de los productos necesarios, tiene en cuenta el tiempo de cocción de cada plato y, a continuación, multiplica los costes por el número de comensales. En la construcción, el proceso es similar: para cada categoría de elementos del proyecto (objetos), se generan estimaciones pormenorizadas Recetas, y el coste total del proyecto se determina sumando todos los costes en una factura total: la estimación final por categoría.

La etapa clave e inicial del enfoque basado en los recursos es la creación de la base de datos inicial de la empresa. En la primera fase del cálculo de costes, se elabora una lista estructurada de todos los artículos, materiales, tipos de trabajo y recursos de que dispone la empresa en sus proyectos de construcción -desde un clavo en el almacén hasta una descripción de las personas a través de su cualificación y tarifa horaria-. Esta información se sistematiza en una única "Base de datos de recursos y materiales de construcción", un registro tabular que contiene datos sobre nombres, características, unidades de medida y precios actuales. Esta base de datos se convierte en la fuente de información principal y primaria para todos los cálculos de recursos posteriores, tanto de costes como de plazos de las obras.

Base de datos de recursos de construcción: catálogo de materiales y obras de construcción

Una base de datos o tabla de recursos y materiales de construcción - incluye información detallada sobre cada elemento que puede utilizarse en un proyecto de construcción - un producto, artículo, material o servicio, incluyendo su nombre, descripción, unidad de medida y coste unitario, registrados de forma estructurada. En esta tabla se pueden encontrar desde distintos tipos de combustibles y materiales utilizados en los proyectos hasta listas detalladas de especialistas en forma de distintas categorías con descripciones de las tarifas horarias (Fig. 5.1-2).

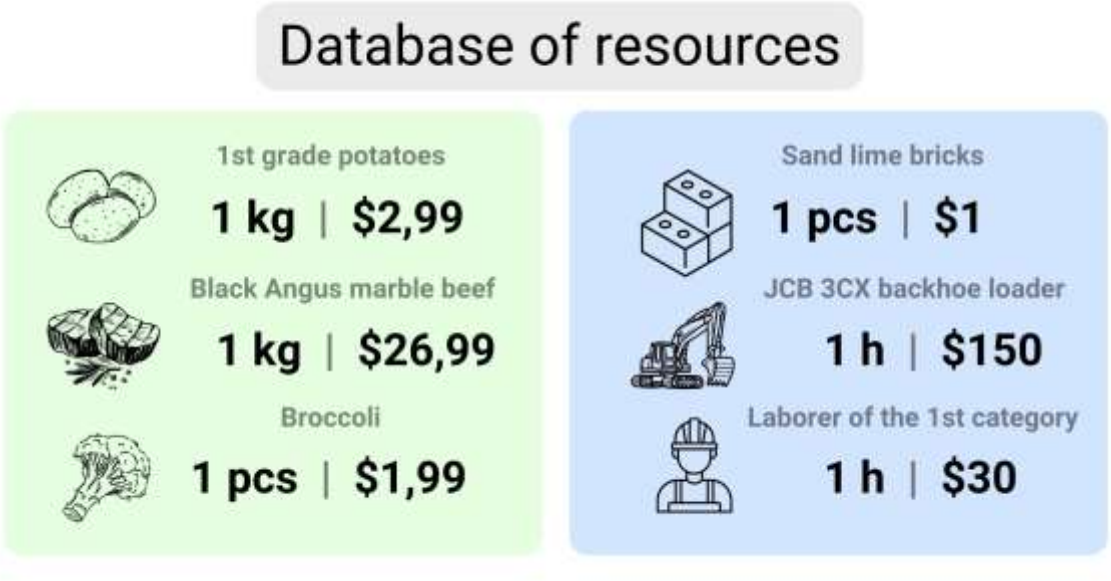


Fig. 5.1-2 Una tabla de recursos es una lista de ingredientes que describe un material y un servicio con un coste unitario.

> "Base de datos de recursos" es similar al catálogo de productos de una tienda online, donde cada artículo tiene una descripción detallada de sus atributos. Esto facilita a los estimadores de costes la selección de los recursos adecuados (como la selección de productos al añadirlos a la cesta de la compra) necesarios para calcular procesos de construcción específicos en forma de cálculos (pedido final en la tienda online).

Una base de datos de recursos también puede considerarse como una lista de todos los ingredientes de un libro de cocina de restaurante. Cada material de construcción, equipamiento y servicio es similar a los ingredientes utilizados en las recetas. "Base de datos de recursos" es una lista detallada de todos los ingredientes: materiales de construcción y servicios, incluido su coste por unidad: pieza, metro, hora, litro, etc.

Pueden añadirse nuevos elementos de entidad a la tabla "Bases de datos de recursos de construcción" de dos formas: manualmente (Fig. 5.1-3) o automáticamente mediante la integración con los sistemas de gestión de inventarios o las bases de datos de proveedores de la empresa.

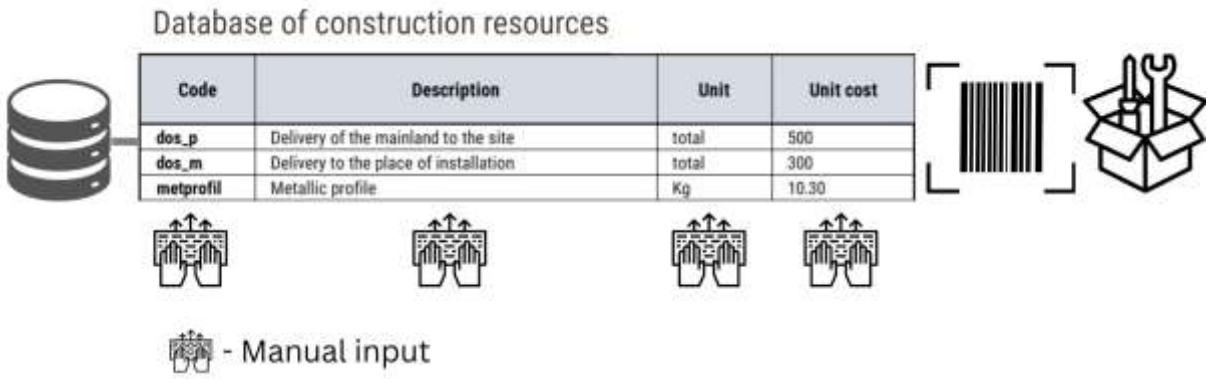


Fig. 5.1-3 La base de datos de recursos se rellena manualmente o adopta automáticamente datos de otras bases de datos.

Una empresa de construcción mediana típica utiliza una base de datos que contiene miles y a veces decenas de miles de elementos con descripciones detalladas que pueden utilizarse en proyectos de construcción. Estos datos se utilizan automáticamente en los contratos y la documentación de los proyectos para describir con precisión la composición de las obras y los procesos

> Para mantenerse al día de las condiciones cambiantes del mercado, como la inflación, el atributo "coste unitario" de cada producto (bien o servicio) en la base de datos de recursos (Fig, 5.1-3) se actualiza periódicamente de forma manual o descargando automáticamente los precios actuales de otros sistemas o plataformas en línea.

La actualización del coste unitario de un recurso puede hacerse mensual, trimestral o anualmente, según la naturaleza del recurso, la inflación y la coyuntura económica exterior. Estas actualizaciones son necesarias para mantener la exactitud de los cálculos y estimaciones, ya que estos elementos básicos son el punto de partida del trabajo de los estimadores de costes. Los datos actualizados sirven para generar estimaciones, presupuestos y calendarios que reflejen las condiciones reales del mercado y reduzcan el riesgo de errores en los cálculos posteriores del proyecto.

Compilación de cálculos y cálculo del coste de las obras en función de la base de recursos

Una vez rellenada la "Base de datos de recursos de construcción" (Fig. 5.1-3) con entidades unitarias mínimas, puede empezar a crear cálculos, que se calculan para cada proceso o trabajo en la obra para determinadas unidades de medida: por ejemplo, para un metro cúbico de hormigón, un metro cuadrado de pared de pladur, por metro de bordillo o por instalación de ventanas.

Por ejemplo, para construir un muro de ladrillos de 1 m² (Fig, 5.1-4), según la experiencia de proyectos anteriores, se necesitan aproximadamente 65 ladrillos (entidad "Ladrillo de silicato") a un coste de 1 $ por pieza (atributo "Coste por pieza"), lo que suma un total de 65 $. Además, según mi experiencia, se requiere utilizar equipo de construcción (entidad "Cargadora JCB 3CX") durante 10 minutos que colocará ladrillos cerca del área de trabajo. Dado que el alquiler del equipo cuesta 150 $ por hora, 6 minutos de uso costarían

aproximadamente 15 $. Además, se necesitará un contratista de colocación de ladrillos durante 2 horas, con una tarifa horaria de 30 $ y un total de 60 $.

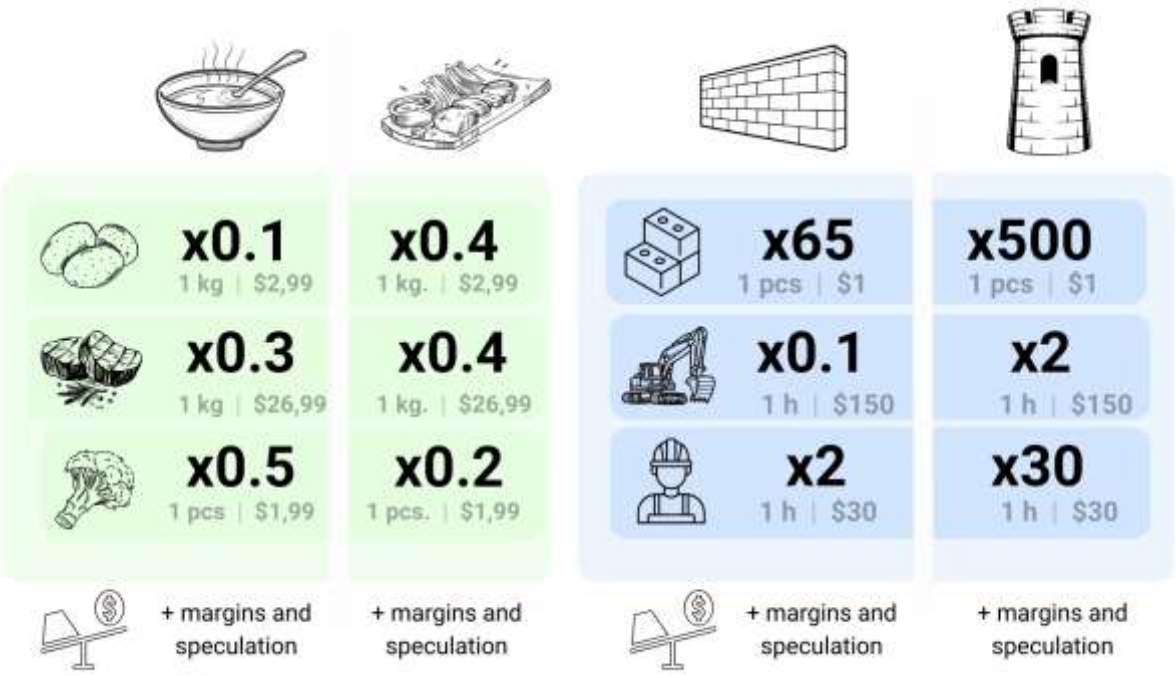


Fig. 5.1-4 Los cálculos de costes proporcionan una lista detallada de los materiales y servicios de construcción necesarios para realizar las obras y procesos.

La composición de los cálculos (las llamadas "recetas") se forma a partir de la experiencia histórica acumulada por la empresa en el proceso de realización de un gran volumen de trabajos similares. Esta experiencia práctica suele acumularse a través de la información obtenida en la obra. En concreto, el capataz recopila información directamente en la obra, registrando los costes reales de mano de obra, el consumo de material y los matices de las operaciones tecnológicas. En colaboración con el departamento de estimaciones, esta información se perfecciona de forma iterativa: se afinan las descripciones de los procesos, se ajusta la combinación de recursos y se actualizan los cálculos de costes para reflejar los datos reales de proyectos recientes.

> Del mismo modo que una receta describe los ingredientes y las cantidades necesarias para preparar un plato, una estimación de costes proporciona una lista detallada de todos los materiales, recursos y servicios de construcción necesarios para completar un trabajo o proceso concreto.

El trabajo realizado con regularidad permite a los trabajadores, capataces y estimadores orientarse en la cantidad necesaria de recursos: materiales, combustible, tiempo de mano de obra y otros parámetros necesarios para realizar una unidad de obra (Fig. 5.1-5). Estos datos se introducen en los sistemas de estimación en forma de tablas, donde cada tarea y operación se describe a través de los elementos mínimos

de la base de recursos (con precios actualizados constantemente), lo que garantiza la exactitud de los cálculos.

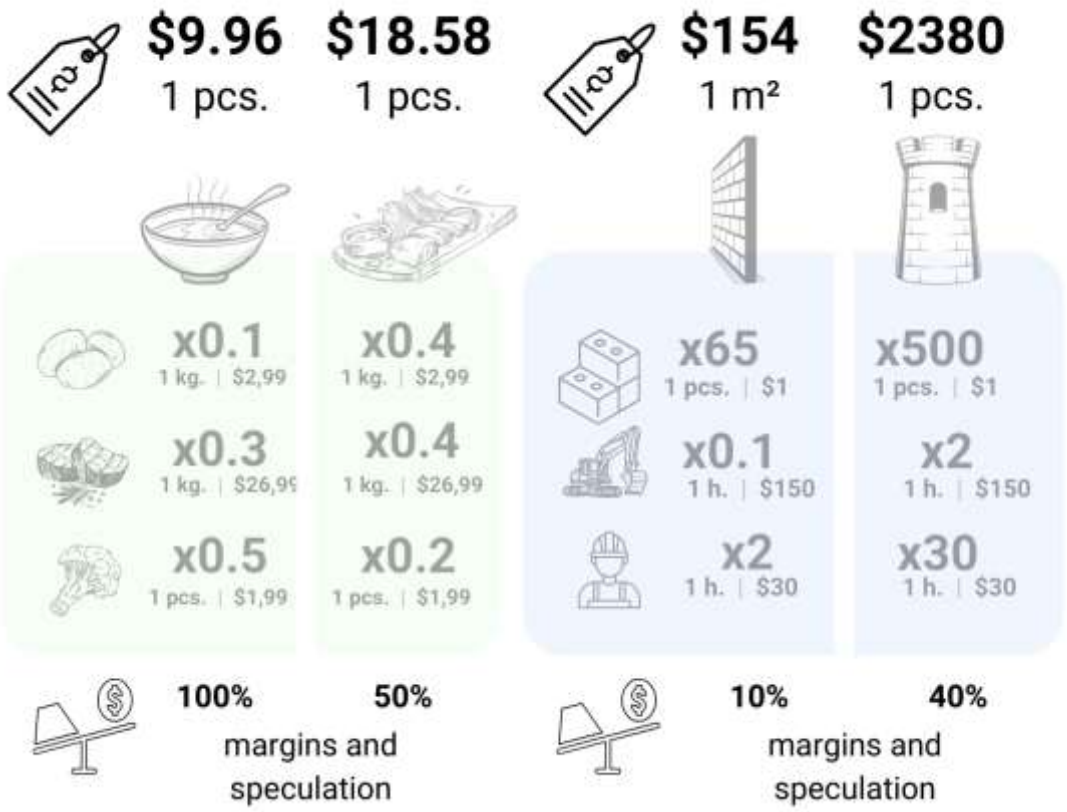


Fig. 5.1-5 Las tarifas unitarias se recogen para cada trabajo, donde el atributo de volumen de la entidad se multiplica por su cantidad con un porcentaje de beneficio añadido.

Para obtener el coste total de cada proceso o actividad (objeto de cálculo del coste), el atributo de coste se multiplica por su número y sus coeficientes. Los coeficientes pueden tener en cuenta diversos factores, como la complejidad del trabajo, las características regionales, la tasa de inflación, los riesgos potenciales (porcentaje de gastos generales previstos) o la especulación (factor de beneficio adicional).

> El estimador, como analista, convierte la experiencia y las recomendaciones del capataz en estimaciones normalizadas, que describen los procesos de construcción a través de entidades de recursos en forma tabular. En esencia, la tarea del estimador consiste en recopilar y estructurar, mediante parámetros y coeficientes, la información procedente de la obra.

Así, el coste final por unidad de obra (por ejemplo, metro cuadrado o cúbico, o una instalación de una sola unidad) incluye no sólo los costes directos de materiales y mano de obra, sino también los márgenes de la empresa, los gastos generales, los seguros y otros factores (Fig, 5.1-6)

Al mismo tiempo, ya no tenemos que preocuparnos de los precios reales en los cálculos (de recetas), ya que los precios reales se reflejan siempre en la "base de recursos" (tabla de ingredientes). En el nivel de

cálculos los datos de la base de recursos se cargan automáticamente en la tabla (por ejemplo, por código de artículo o su identificador único), que carga la descripción y el precio real por unidad, que a su vez puede cargarse automáticamente desde plataformas online o tienda online de materiales de construcción. El estimador a nivel de cálculos obras sólo tiene que describir la obra o proceso mediante el atributo "cantidad de recursos" y factores adicionales.

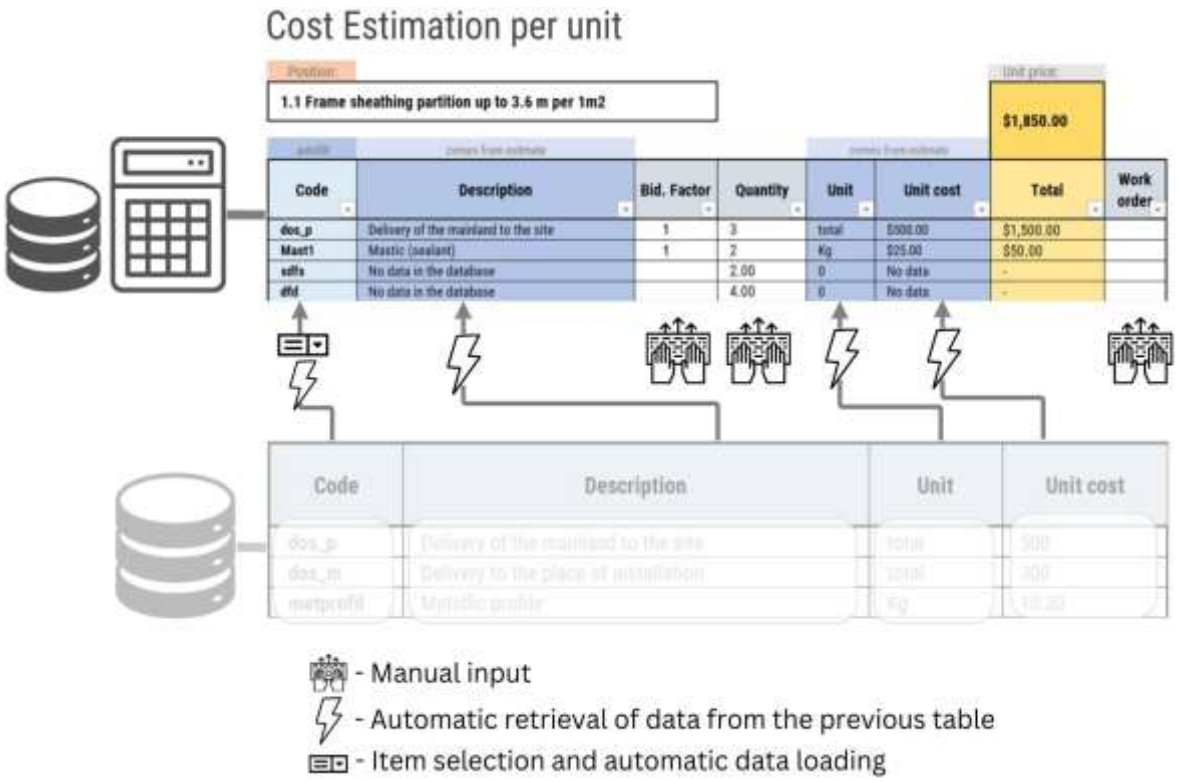


Fig. 5.1-6 En la fase de cálculo del coste unitario del trabajo, sólo se rellenan los atributos del número de recursos necesarios, todo lo demás se carga automáticamente de la base de datos de recursos.

Los cálculos de costes de los trabajos creados se almacenan en forma de tablas de plantillas de proyectos típicos, que están directamente vinculadas a la base de datos de recursos y materiales de construcción. Estas plantillas representan recetas estandarizadas de tipos de trabajo repetitivos para futuros proyectos, lo que garantiza la uniformidad de los cálculos en toda la empresa.

Cuando el coste de cualquier recurso cambia en la base de datos (Fig. 5.1-3) - ya sea manual o automáticamente mediante la descarga de los precios de mercado actuales (por ejemplo, en condiciones inflacionistas) - las actualizaciones se reflejan inmediatamente en todos los cálculos de costes vinculados (Fig. 5.1-6). Esto significa que sólo es necesario modificar la base de recursos, mientras que las plantillas de cálculo de costes y las estimaciones permanecen inalteradas a lo largo del tiempo. Este enfoque garantiza la estabilidad y reproducibilidad de los cálculos ante cualquier fluctuación de precios, que sólo se contabilizan en una tabla de recursos relativamente sencilla (Fig, 5.1-3).

Para cada nuevo proyecto, se crea una copia de la plantilla de cálculo de costes estándar, lo que permite introducir cambios y ajustar las actividades para satisfacer requisitos específicos sin modificar la plantilla original adoptada por la empresa. Este enfoque ofrece flexibilidad a la hora de adaptar los cálculos: se pueden tener en cuenta las particularidades de la obra, los deseos del cliente, introducir coeficientes de riesgo o de rentabilidad (especulación)... todo ello sin violar las normas de la empresa. Esto ayuda a la empresa a encontrar un equilibrio entre la maximización del beneficio, la satisfacción del cliente y el mantenimiento de su competitividad.

En algunos países, estas plantillas de cálculo de costes, acumuladas durante décadas, se normalizan a nivel nacional y pasan a formar parte de las normas del sistema nacional de cálculo de costes de la construcción (Fig, 5.1-7).

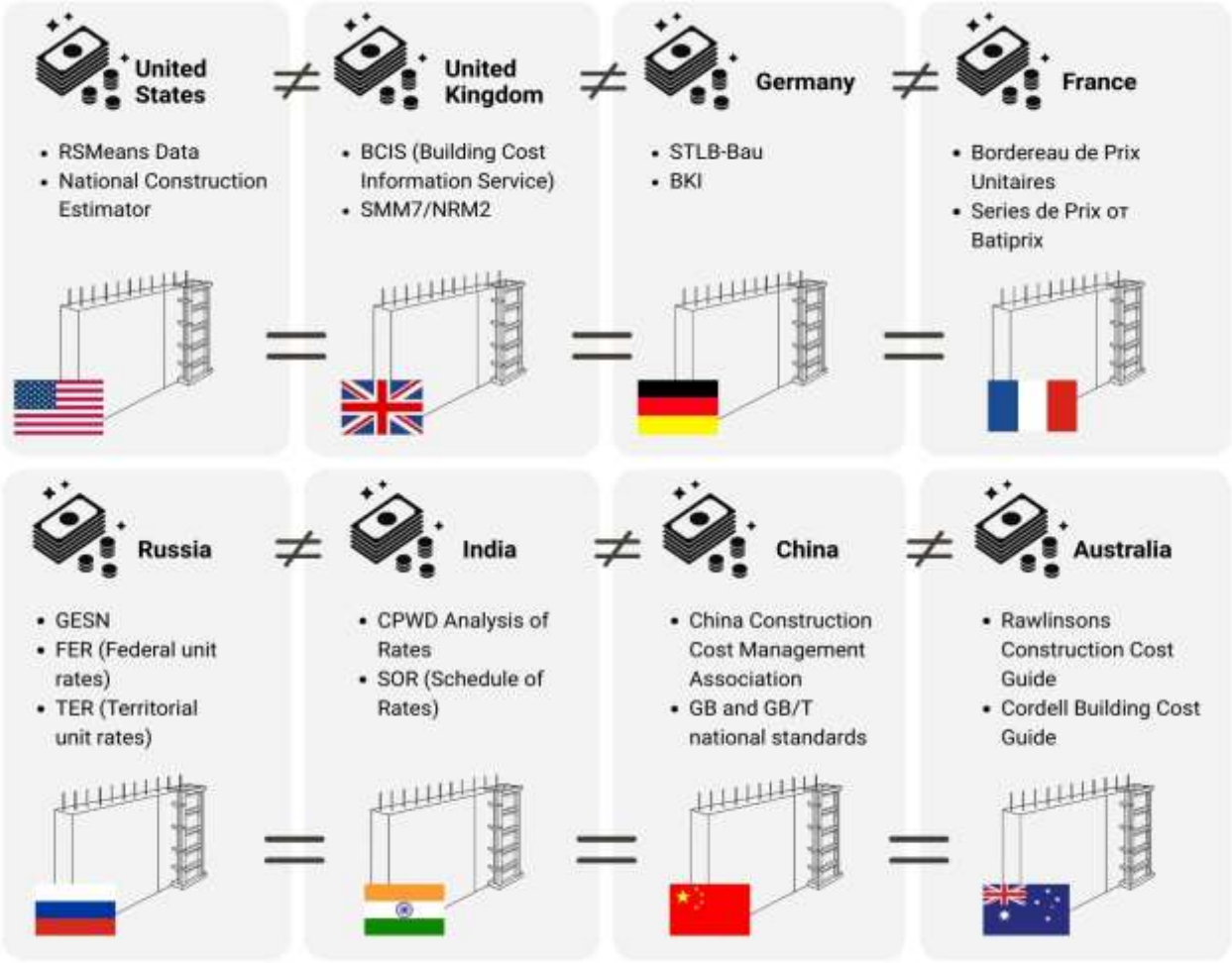


Fig. 5.1-7 Diferentes países de todo el mundo tienen sus propias normas de cálculo de costes con sus propios compendios y normas (prescriptivas) de obras de construcción para calcular el coste de un mismo elemento.

Estas bases normalizadas de estimaciones de recursos (Fig, 5.1-7) son de uso obligatorio todos los participantes en el mercado, especialmente para los proyectos financiados con fondos públicos. Dicha normalización garantiza la transparencia, comparabilidad y equidad en la formación de precios y obligaciones contractuales para el cliente

Cálculo del coste final del proyecto: de la estimación al presupuesto

Las normas de estimación estatales y específicas del sector desempeñan diferentes papeles en la práctica de la construcción en los distintos países. Mientras que algunos países exigen el cumplimiento estricto de una única norma, la mayoría de las economías desarrolladas adoptan un enfoque más flexible. En las economías de mercado, las normas de construcción estatales suelen servir sólo de referencia. Las empresas constructoras adaptan estas normas a sus modelos operativos o las revisan por completo, complementándolas con sus propios factores personalizados. Estos ajustes reflejan la experiencia empresarial, la eficiencia en la gestión de los recursos y, a menudo, factores en los que, por ejemplo, pueden tenerse en cuenta los beneficios especulativos de una empresa.

> En consecuencia, el nivel de competencia, la demanda del mercado, los márgenes previstos e incluso las relaciones con clientes específicos pueden dar lugar a desviaciones significativas de las normas estandarizadas. Esta práctica aporta flexibilidad al mercado, pero también dificulta la comparación transparente de las ofertas de distintos contratistas, introduciendo un elemento de precios especulativos en el sector de la construcción en esta fase del proceso de cálculo.

Una vez preparadas las plantillas de cálculo de las distintas actividades y procesos -o, lo que es más frecuente, simplemente copiadas de los presupuestos públicos estándar (Fig. 5.1-7) con coeficientes añadidos para reflejar las "peculiaridades" de una empresa concreta-, el último paso consiste en multiplicar el coste de cada partida por el atributo correspondiente del ámbito de trabajo o los procesos del nuevo proyecto.

> A la hora de calcular el coste total de un proyecto de nueva construcción, el paso clave consiste en resumir los costes de todos los elementos de cálculo, multiplicados por el volumen de estos elementos de trabajo en el proyecto.

Para crear el coste total del proyecto, en nuestro ejemplo simplificado, empezamos calculando el coste de construcción de un metro cuadrado de muro y multiplicamos el coste de su cálculo (por ejemplo, el trabajo "1m² de instalación estándar de elementos de muro") por el número total de metros cuadrados de muros del proyecto (por ejemplo, el atributo "Área" o "Cantidad" (Fig. 5.1-8) de una entidad de tipo "Elementos de muro" de CAD del proyecto o los cálculos del capataz).

Del mismo modo, calculamos el coste de todos los elementos del proyecto (Fig. 5.1-8): tomamos el coste por unidad de obra y lo multiplicamos por el volumen de un elemento o grupo de elementos concretos del proyecto. El estimador sólo tiene que introducir el número de estos elementos, actividades o procesos del

proyecto en en forma de volumen o cantidad. Esto permite generar automáticamente un presupuesto de construcción completo.

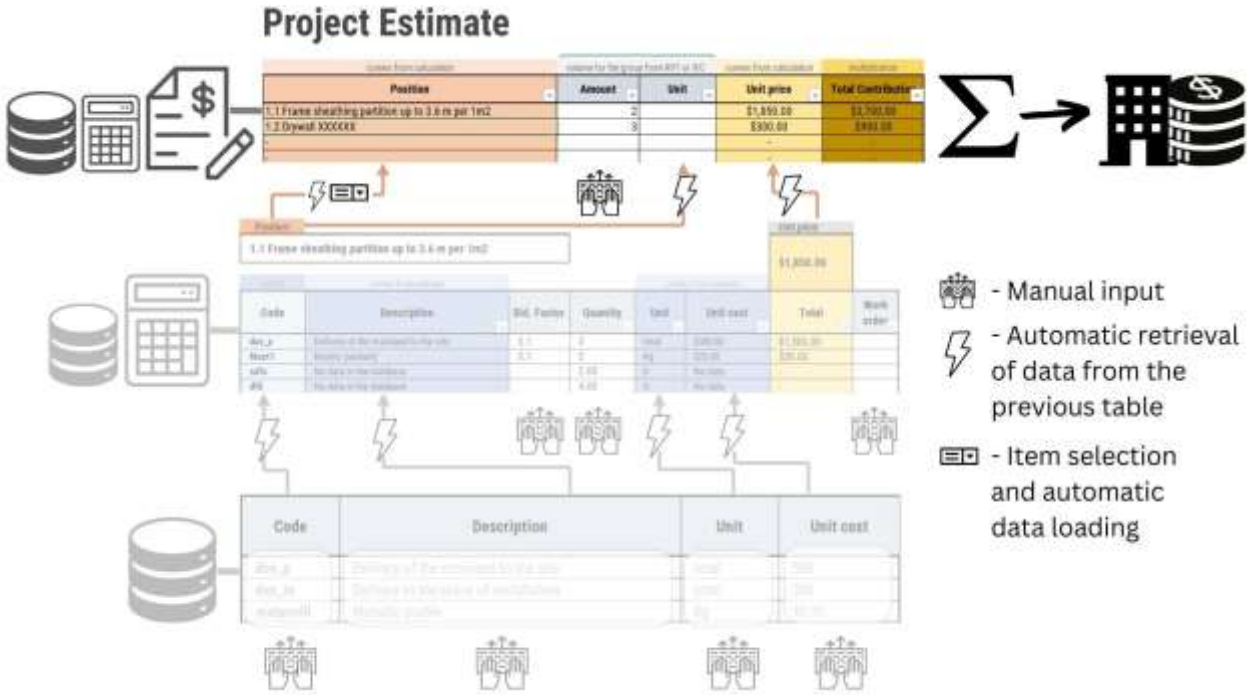


Fig. 5.1-8 En la fase de creación de un presupuesto, sólo introducimos el alcance del trabajo.

Como en el caso de los cálculos, en este nivel cargamos automáticamente elementos calculados listos (de la plantilla de cálculo o nuevos copiados de la plantilla y editados), que automáticamente traen consigo el coste actual por unidad de trabajo (que se actualiza automáticamente desde la base de datos de recursos (Fig. 5.1-8 tabla inferior)). Por consiguiente, en caso de cualquier cambio de datos en la base de datos de recursos o en las tablas de cálculo de costes, los datos del presupuesto se actualizarán automáticamente para el día en curso, sin necesidad de cambiar el cálculo de costes o el propio presupuesto.

En el contexto de un restaurante, el coste final de un evento se calcula de forma similar y equivale al coste final de toda la cena, donde el coste de cada plato multiplicado por el número de comensales suma el coste total del cheque (Fig, 5.1-9). Y al igual que en la construcción, las recetas para cocinar en un restaurante pueden no cambiar durante décadas. A diferencia de los precios, donde el coste de los ingredientes puede cambiar cada hora.

> Al igual que el propietario de un restaurante multiplica el coste de cada comida por el número de raciones y personas para determinar el coste total del evento, el gestor de estimaciones de costes suma el coste de todos los componentes del proyecto para llegar a una estimación completa de la construcción.

Así, para cada actividad del proyecto se determina su coste final (Fig. 5.1-9), que, multiplicado por el volumen de atributos de la entidad correspondiente a esta actividad, da el coste de los grupos de actividades,

a partir del cual se obtiene el coste final de todo el proyecto.

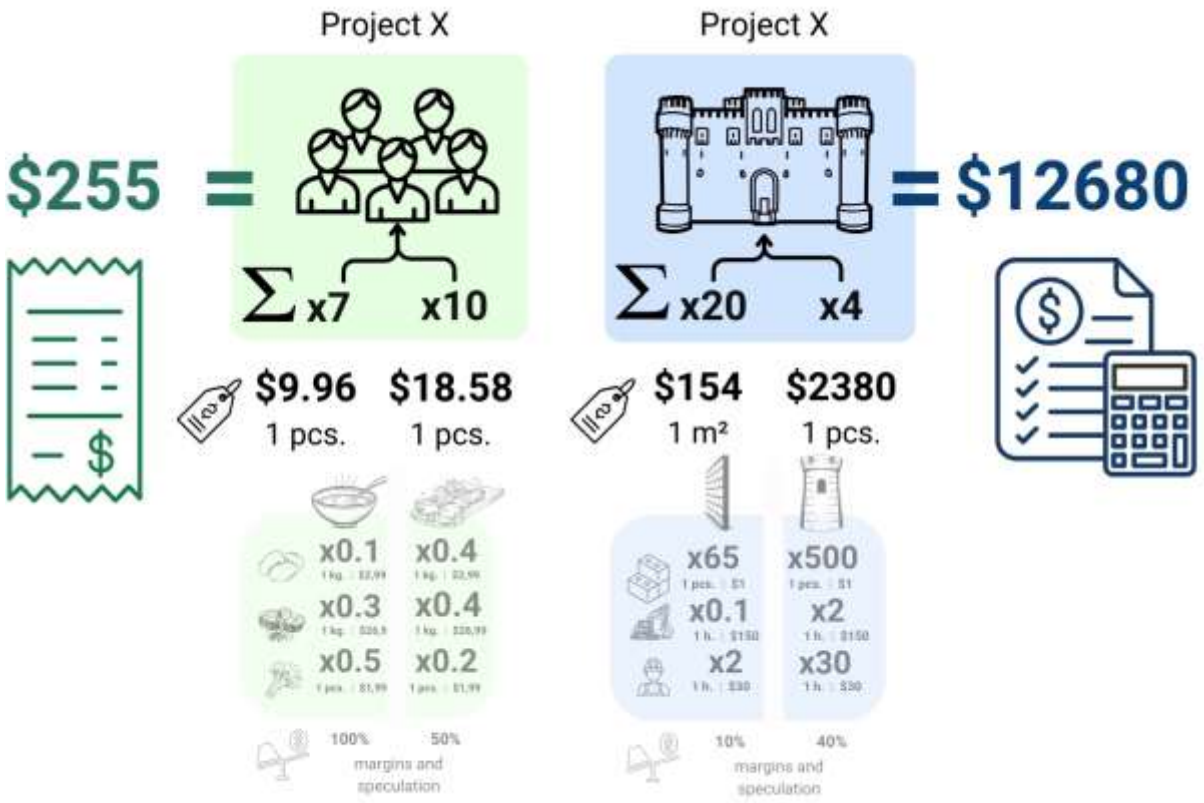


Fig. 5.1-9 La estimación final se calcula sumando el atributo de coste de trabajo de cada elemento por su atributo de alcance.

El coste total del proyecto (Fig, 5.1-8) ofrece una imagen financiera del proyecto que permite a los clientes, inversores u organizaciones de financiación comprender el presupuesto total y los recursos financieros necesarios para entregar el proyecto en un día determinado, teniendo en cuenta los precios actuales.

Y si los procesos de compilación de bases de recursos, cálculos y estimaciones (recetas de procesos) ya han sido elaborados, semiautomatizados y perfeccionados por decenas de miles de años y registrados a nivel estatal, entonces la obtención automática de información cualitativa sobre el volumen y la cantidad de elementos para la última etapa de la estimación final - hoy en día sigue siendo un cuello de botella en los procesos de todos los cálculos de atributos de costes y plazos del proyecto, y en general del presupuesto global del proyecto.

> Durante miles de años, el método tradicional de cálculo de volúmenes ha sido la medición manual de volúmenes y cantidades a partir de dibujos planos. Con la llegada de la era digital, las empresas han descubierto que ahora es posible extraer automáticamente información sobre volúmenes y cantidades a partir de los datos geométricos contenidos en modelos CAD, revolucionando formas milenarias de obtener datos cuantitativos.

Los enfoques modernos de estimación y cálculo de procesos implican la extracción automática de atributos volumétricos y cuantitativos a partir de bases de datos CAD, que pueden cargarse y conectarse al proceso de cálculo de costes para obtener volúmenes de grupos de proyectos actualizados en cualquier fase del diseño hasta la explotación.

CAPÍTULO 5.2.
DESPEGUE DE CANTIDADES Y CREACIÓN AUTOMÁTICA DE PRESUPUESTOS Y CALENDARIOS

Pasar de 3D a 4D y 5D: utilizar parámetros volumétricos y cuantitativos

Con las tablas de cálculo de costes con los procesos descritos a través de los recursos (Fig. 5.1-8) en la mano, el siguiente paso es obtener automáticamente los parámetros de volumen o cantidad para un grupo de elementos que son necesarios para los cálculos y para la estimación final.

Las características volumétricas de los elementos del proyecto -por ejemplo, muros o losas- pueden extraerse automáticamente de las bases de datos CAD. Los objetos paramétricos creados en programas CAD se convierten mediante un núcleo geométrico en valores numéricos de longitud, anchura, área, volumen y otros parámetros. El proceso de obtención de volúmenes basados en geometría 3D se tratará con más detalle en la siguiente y sexta parte (Fig. 6.3-3), dedicada al trabajo con CAD (BIM). Además de los volúmenes, también se puede obtener el número de elementos similares a partir de la base de datos del modelo CAD filtrando y agrupando los objetos por categorías y propiedades. Estos parámetros, que permiten la agrupación, se convierten en la base para vincular los elementos del proyecto mediante el cálculo de recursos con los cálculos, las estimaciones finales y el presupuesto de todo el proyecto.

Así, el modelo de datos extraído del modelo 3D (CAD) se aumenta con nuevas capas de parámetros, denominadas 4D y 5D. En las nuevas capas de atributos de entidad, 4D (tiempo) y 5D (coste), los datos geométricos 3D se utilizan como fuente de valores de atributos de volumen de entidad.

- **4D** es una capa de información de parámetros que añade información sobre la duración de las operaciones de construcción a los parámetros 3D de los elementos. Estos datos son esenciales para planificar los calendarios y gestionar los plazos de los proyectos
- **5D** es el siguiente nivel de ampliación del modelo de datos, en el que los elementos se complementan con características de coste. Esto añade un aspecto financiero a la información geométrica: el coste de los materiales, obras y equipos, lo que permite calcular presupuestos, análisis de rentabilidad y gestión de costes durante el proceso de construcción.

Los datos de costes y atributos 3D, 4D y 5D de grupos de entidades del proyecto se describen de forma similar a los cálculos de los sistemas modulares ERP, PIMS (o herramientas similares a Excel) y se utilizan para el cálculo automático de costes y la planificación presupuestaria tanto de grupos individuales como del presupuesto completo del proyecto.

Atributos 5D y obtención de volúmenes de atributos a partir de CAD

Al elaborar el presupuesto final de un proyecto de construcción, cuya preparación hemos tratado en capítulos anteriores (Fig, 5.1-8), los atributos de volumen de cada categoría de elementos del proyecto se recopilan manualmente o se extraen de las especificaciones de atributos de volumen proporcionadas por el software de CAD.

El método manual tradicional de cálculo de cantidades implica que el capataz y el estimador analicen dibujos que se han presentado durante miles de años como líneas sobre papel y durante los últimos 30 años en formatos digitales como PDF (PLT) o DWG. Basándose en su experiencia profesional, miden las cantidades de trabajo y materiales necesarios, a menudo con regla y transportador. Este método requiere un esfuerzo y un tiempo considerables, así como una especial atención a los detalles.

Determinar los atributos de alcance de esta manera puede llevar desde unos pocos días hasta varios meses, dependiendo del alcance del proyecto. Además, dado que todas las mediciones y cálculos se realizan manualmente, existe el riesgo de que se produzcan errores humanos que pueden dar lugar a datos inexactos, lo que posteriormente repercute en errores en la estimación del tiempo y el coste del proyecto, de los que toda la empresa será responsable.

Los métodos modernos basados en el uso de bases de datos CAD simplifican enormemente el cálculo de volúmenes. En los modelos CAD, la geometría de los elementos ya incluye atributos de volumen que pueden calcularse automáticamente (a través del núcleo de geometría (Fig. 6.3-3)) y presentarse o exportarse en forma tabular.

En este caso, el departamento de presupuestos solicita al diseñador CAD datos sobre las características cuantitativas y de volumen de los elementos del proyecto. Estos datos se exportan en forma de hojas de cálculo o se integran directamente en bases de datos de cálculo de costes, ya sean sistemas Excel, ERP o PMIS. Este proceso no suele comenzar con una solicitud formal, sino con un breve diálogo entre el cliente (iniciador) y el arquitecto estimador de la empresa constructora o de diseño. A continuación se muestra un ejemplo simplificado de cómo se forma una tabla estructurada para cálculos automáticos (QTO) a partir de la comunicación cotidiana:

- Cliente - "*Quiero añadir otra planta al edificio, con la misma configuración que la primera planta*"
- Arquitecto (CAD) - "*Añadiendo una tercera planta, la configuración es la misma que en la segunda planta*". Y después de este mensaje envía una nueva versión CAD del proyecto al estimador.
- *Estimator* realiza automáticamente la agrupación y el cálculo (ERP, PMIS, Excel) - "*Ejecutaré el proyecto a través de una hoja de cálculo Excel con reglas QTO (ERP, PMIS), obtendré los volúmenes por categoría para la nueva planta y generaré una estimación*"

  Como resultado, el diálogo de texto se transforma en una estructura de tabla con reglas de agrupación:

Elemento	Categoría	Suelo
Solapamiento	OST_Suelos	3
Columna	OST_ColumnasEstructurales	3
escalera	Escaleras_OST	3

Tras el proceso de agrupación automática del modelo CAD del diseñador según las reglas QTO del estimador y la multiplicación automática de volúmenes por cálculos de recursos (Fig. 5.1-8) obtenemos los siguientes resultados, que se envían al cliente:

Elemento	Volumen	Suelo	Precio por unidad.	Coste total
Solapamiento	420 м²	3	150 €/м²	**63 000 €**
Columna	4 piezas	3	2450 €/unidad.	**9 800 €**
escalera	2 piezas	3	4.300 euros/unidad.	**8 600 €**
TOTAL:	-	-	-	**81 400 €**

- El cliente: "*Gracias, es mucho, tenemos que reducir algunas habitaciones*". Y el ciclo se repite muchas veces.

Este escenario puede repetirse muchas veces, sobre todo en la fase de aprobación, en la que el cliente espera una respuesta instantánea. En la práctica, sin embargo, estos procesos pueden alargarse durante días o incluso semanas. Hoy en día, gracias a la introducción de reglas automáticas de agrupación y cálculo, actividades que antes llevaban un tiempo considerable deberían completarse en cuestión de minutos. La adquisición automatizada de cantidades, mediante reglas de agrupación, no sólo acelera los cálculos y estimaciones, sino que, al minimizar el factor humano, reduce la probabilidad de errores, proporcionando una evaluación transparente y precisa de los costes del proyecto.

> Si los requisitos del departamento de estimación se tuvieron en cuenta inicialmente al crear el modelo 3D en el sistema CAD (lo que aún es poco frecuente en la práctica), y los nombres, identificadores de grupos de elementos y sus atributos de clasificación se establecen en forma de parámetros que coinciden con las estructuras de grupos y clases de estimación, entonces los atributos volumétricos pueden transferirse automáticamente a los sistemas de estimación sin transformaciones adicionales.

La extracción automática de atributos volumétricos del CAD en forma de tablas de especificaciones permite obtener rápidamente datos actualizados sobre el coste de las obras individuales y del proyecto en su conjunto (Fig. 5.2-1). Actualizando únicamente el archivo CAD con los volúmenes del proyecto en el proceso de cálculo o en el sistema de cálculo, la empresa puede recalcular rápidamente el presupuesto teniendo en cuenta los últimos cambios, garantizando una gran precisión y coherencia de todos los cálculos posteriores.

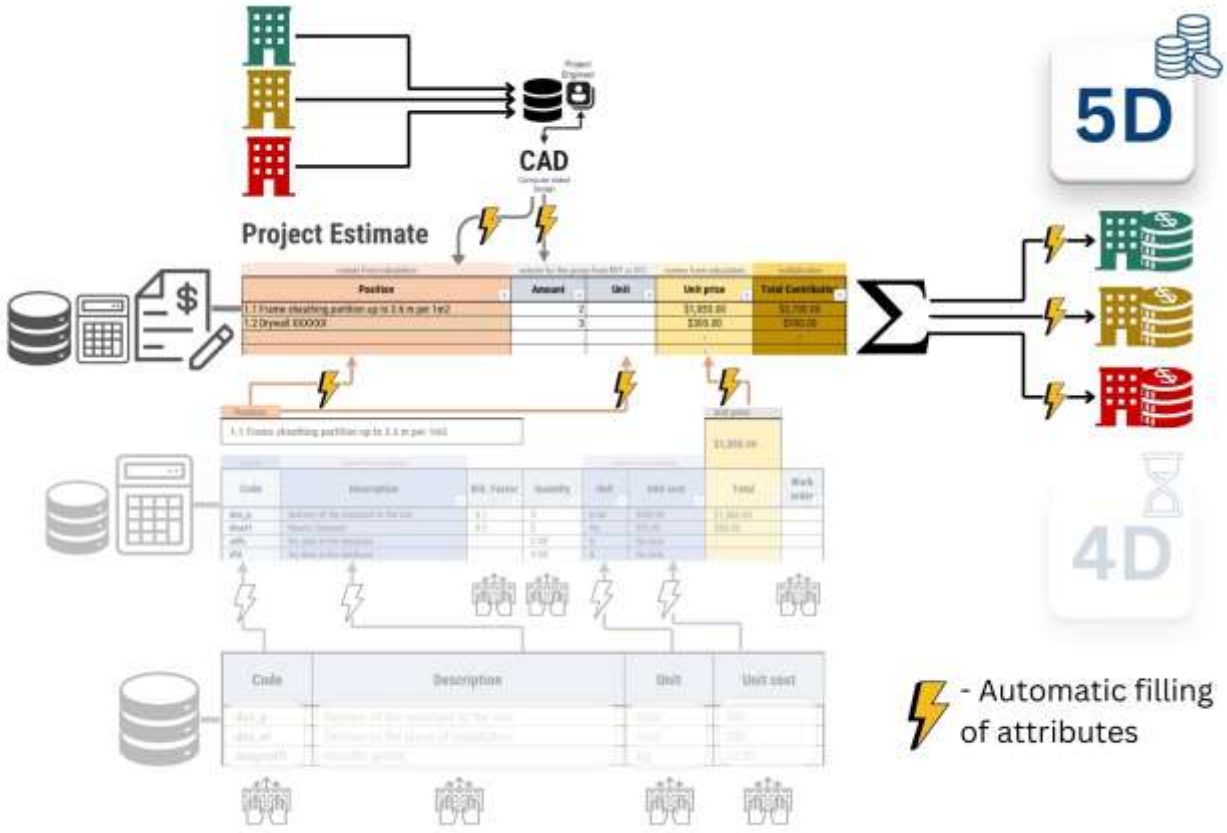


Fig. 5.2-1 Los atributos de volumen de las tablas CAD o bases de datos se introducen automáticamente en el presupuesto, lo que le permite calcular instantáneamente el coste total del proyecto.

Con la creciente complejidad de los proyectos de capital, calcular el presupuesto completo y analizar el coste total de los proyectos en tal escenario (Fig, 5.2-1) - se convierte en una herramienta clave para tomar decisiones con conocimiento de causa.

> Según el estudio de Accenture Creating More Value through Capital Projects (2024) [20], las empresas líderes están integrando activamente la analítica de datos en las iniciativas digitales, utilizando información histórica para predecir y optimizar los resultados. La investigación muestra que cada vez más propietarios-operadores aplican la analítica de big data para predecir las tendencias del mercado y evaluar la viabilidad comercial antes de iniciar el diseño. Esto se consigue analizando los almacenes de datos de una cartera de proyectos existente. Además, el 79% de los propietarios-operadores están aplicando análisis predictivos "sólidos" para evaluar el rendimiento de los proyectos y apoyar la toma de decisiones operativas en tiempo real.

La gestión moderna y eficaz de los proyectos de construcción está inextricablemente ligada al procesamiento y análisis de grandes cantidades de información en todas las fases del diseño y en los procesos que preceden al diseño. El uso de almacenes de datos, cálculos de recursos, modelos predictivos y aprendizaje automático permite no sólo minimizar los riesgos en los cálculos, sino también tomar decisiones estratégicas sobre la financiación de proyectos en las primeras fases del diseño. Hablaremos más sobre los almacenes de datos y los modelos predictivos que complementarán los cálculos en la novena parte del

libro.

La obtención automática de parámetros volumétricos de elementos a partir de proyectos CAD, necesarios para la preparación de presupuestos, se realiza con la ayuda de herramientas de agrupación QTO (Quantity Take-Off). Las herramientas QTO funcionan agrupando todos los objetos del proyecto por identificadores especiales de elementos o parámetros de atributos de elementos, utilizando especificaciones y tablas creadas en la base de datos CAD.

QTO Quantity Take-Off: agrupación de datos de proyecto por atributos

QTO (Quantity Take-Off) en construcción es el proceso de extracción de las cantidades de elementos necesarios para realizar un proyecto. En la práctica, el QTO suele ser un proceso semimanual que implica la recopilación de datos de diversas fuentes: documentos PDF, planos DWG y modelos digitales CAD.

Cuando se trabaja con datos extraídos de bases de datos CAD, el proceso QTO se realiza como una secuencia de operaciones de filtrado, clasificación, agrupación y agregación. Los elementos del modelo se seleccionan según los parámetros de clases, categorías y tipos, tras lo cual sus atributos cuantitativos - como volumen, área, longitud o cantidad- se resumen según la lógica de cálculo (Fig. 5.2-2).

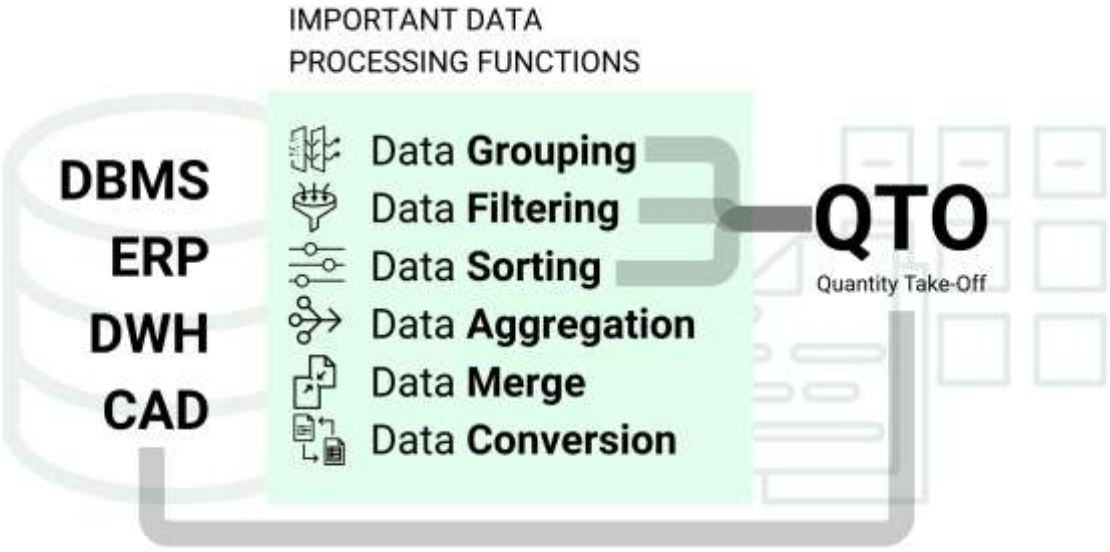


Fig. 5.2-2 La agrupación y el filtrado de datos son las funciones más populares aplicadas a bases de datos y almacenes de datos.

El proceso QTO (filtrado y agrupación) permite sistematizar datos, formar especificaciones y preparar información de entrada para calcular presupuestos, compras y calendarios de trabajo. La base del QTO es la clasificación de los elementos según el tipo de atributos que deben medirse. Para cada elemento o grupo de elementos, se selecciona el parámetro de medición cuantitativa adecuado. Por ejemplo:

- **Atributo de longitud** (bordillo - en metros)
- **Atributo de superficie** (trabajos en pladur - en metros cuadrados)

- **Atributo de volumen** (obras de hormigón - en metros cúbicos)
- **Atributo de cantidad** (ventanas - por pieza)

Además de las características volumétricas generadas matemáticamente a partir de la geometría, en los cálculos posteriores a la agrupación QTO suelen aplicarse factores de rebasamiento (Fig. 5.2-12 por ejemplo, 1,1 para tener en cuenta un 10% para logística e instalación), valores de corrección que tienen en cuenta las pérdidas, la instalación, el almacenamiento o el transporte. Esto permite predecir con mayor exactitud el consumo real de materiales y evitar tanto la escasez como el exceso de existencias en la obra.

> Un proceso automatizado de absorción de cantidades (QTO) es esencial para realizar cálculos y estimaciones precisos en, reducir los errores humanos en los procesos de especificación de volúmenes y evitar los pedidos excesivos o insuficientes de materiales.

Como ejemplo del proceso QTO, consideremos un caso común en el que es necesario mostrar desde la base de datos CAD una tabla-especificación de volúmenes por tipos de elementos para una determinada categoría, clases de elementos. Agrupemos todos los elementos del proyecto por tipo desde la categoría de muro de proyecto CAD y resumamos los atributos de volumen para cada tipo para presentar el resultado como una tabla de volúmenes QTO (Fig. 5.2-3).

En el ejemplo de un proyecto CAD típico (Fig. 5.2-3), todos los elementos de categoría de muro dentro de la base de datos CAD están agrupados por tipo de muro, por ejemplo "Lamelle 11.5", "MW 11.5" y "STB 20.0", y tienen atributos de volumen bien definidos representados en cubos métricos.

El objetivo del gestor, que se encuentra en la interfaz entre diseñadores y especialistas en cálculo, es obtener una tabla automatizada de volúmenes por tipo de elemento en la categoría seleccionada. No sólo para un proyecto concreto, sino también de forma universal aplicable a otros proyectos con una estructura de modelo similar. Esto permite que el enfoque sea escalable y que los datos puedan reutilizarse sin duplicar esfuerzos.

> Atrás quedaron los días en que los diseñadores experimentados y los estimadores se armaban con una regla y medían cuidadosamente cada línea en papel o en planos PDF -una tradición que no ha cambiado en los últimos milenios. Con el desarrollo de la modelización 3D, en la que la geometría de cada elemento está ahora directamente vinculada a atributos volumétricos calculados automáticamente, el proceso de determinación de volúmenes y cantidades QTO se ha automatizado.

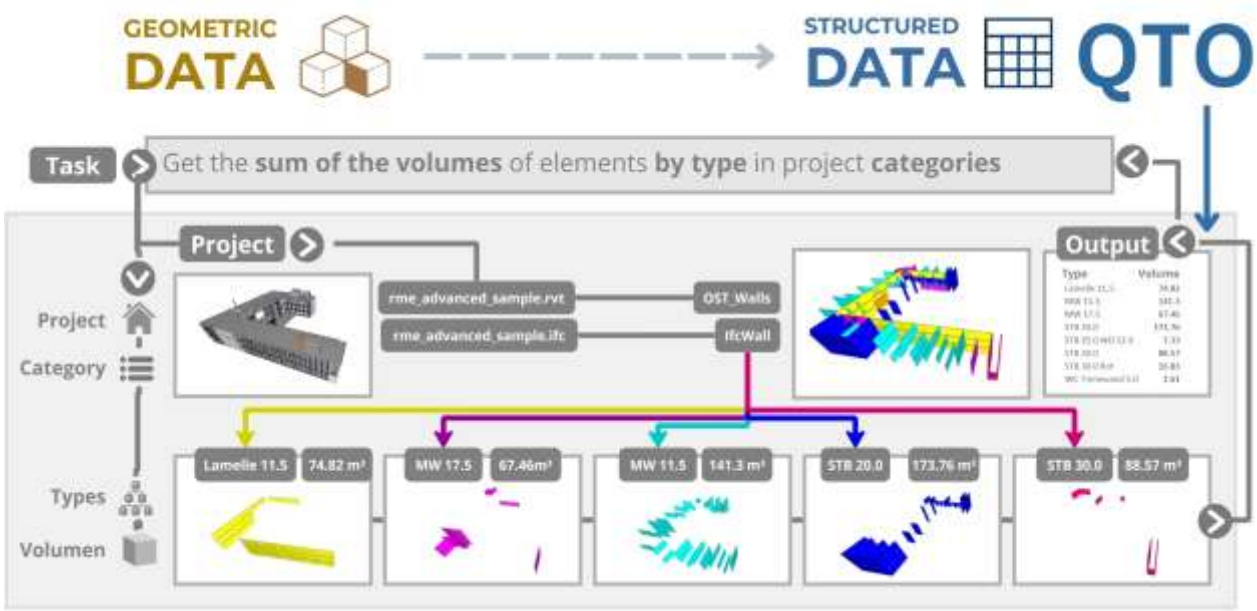


Fig. 5.2-3 La obtención de atributos de volumen y cantidad QTO de un proyecto implica agrupar y filtrar elementos del proyecto.

En nuestro ejemplo, la tarea consiste en "seleccionar una categoría de muros en un proyecto, agrupar todos los elementos por tipo y presentar la información sobre los atributos de volumen en un formato tabular estructurado", de modo que esta tabla pueda ser utilizada por docenas de otros especialistas para cálculos de costes, logística, calendarios de trabajo y otros casos empresariales (Fig. 6.1-3).

Debido a la naturaleza cerrada de los datos CAD , hoy en día no todos los especialistas pueden utilizar el acceso directo a la base de datos CAD (las razones y soluciones al problema del acceso se detallan en la sexta parte del libro). Por ello, muchos tienen que recurrir a herramientas BIM especializadas basadas en los conceptos de BIM abierto y BIM cerrado [63]. Cuando se trabaja con herramientas BIM especializadas o directamente en el entorno de software CAD, la tabla con los resultados de QTO (Quantity Take-Off) puede generarse de distintas formas, dependiendo de si se utiliza una interfaz manual o la automatización del software.

Por ejemplo, utilizando la interfaz de usuario del software CAD (BIM), basta con realizar unas 17 acciones (clics en botones) para obtener una tabla de volúmenes lista (Fig. 5.2-4). No obstante, el usuario debe conocer bien la estructura del modelo y las funciones del software CAD (BIM).

Si la automatización se aplica mediante código de programa o mediante plug-ins y herramientas API dentro de los programas CAD, se reduce el número de pasos manuales para obtener las tablas de volumen, pero será necesario escribir entre 40 y 150 líneas de código, en función de la biblioteca o herramienta utilizada:

- **IfcOpSh (BIM abierto**) o **Dynamo IronPython (BIM cerrado)** - permiten obtener una tabla QTO a partir de un formato CAD o un programa CAD en sólo ~40 líneas de código.
- **IFC_js (open BIM**) - requiere aproximadamente 150 líneas de código para extraer voluminosos atributos del modelo IFC.

- **Interfaz de herramientas CAD (BIM)**: permite obtener el mismo resultado manualmente, en 17 clics de ratón.

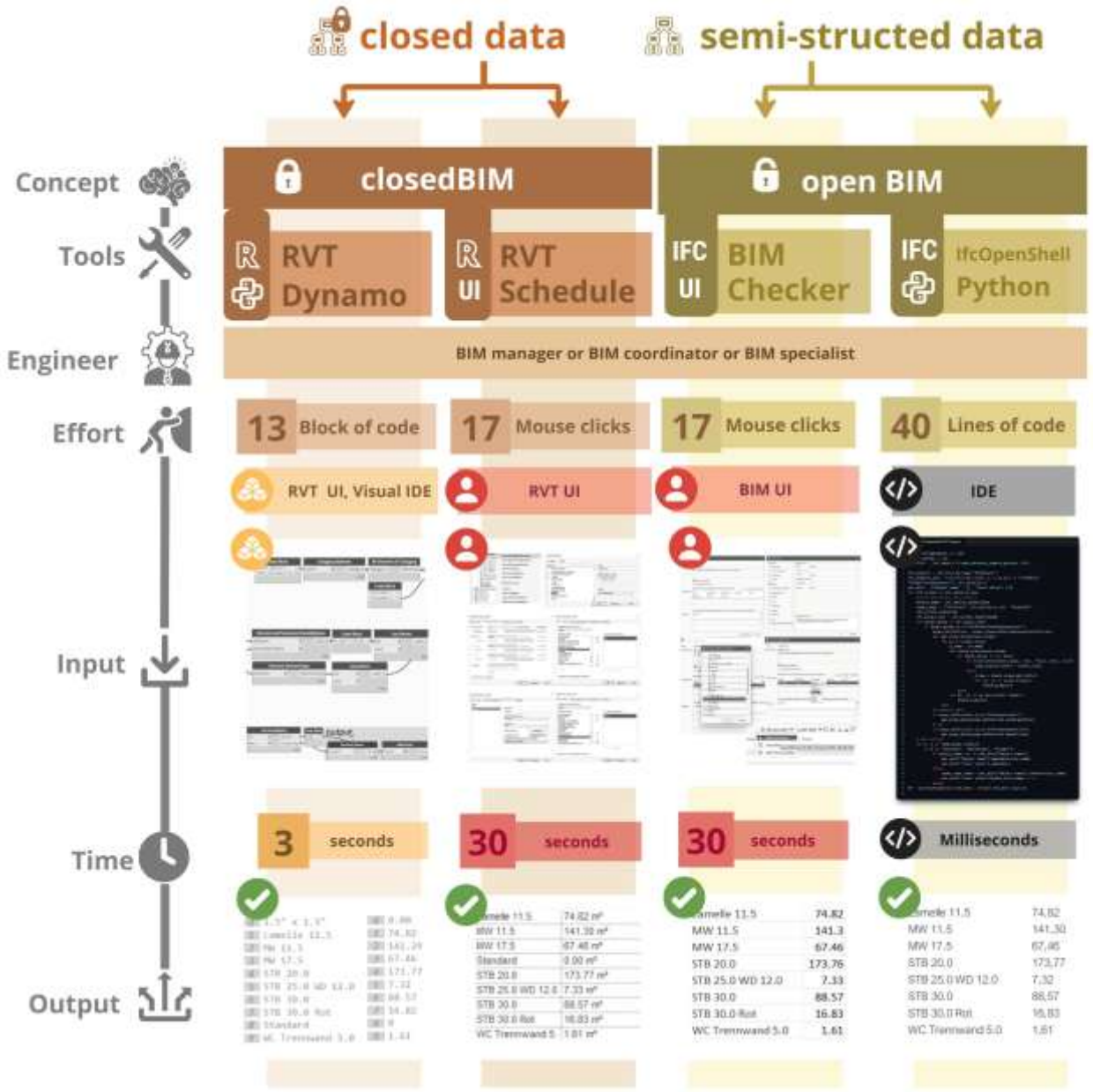


Fig. 5.2-4 Diseñadores y gestores de CAD (BIM), utilizan entre 40 y 150 líneas de código o una docena de pulsaciones de teclas para crear tablas QTO

El resultado es el mismo: una tabla estructurada con atributos de volumen para un grupo de elementos. La única diferencia es el coste de mano de obra y el nivel de conocimientos técnicos requeridos por el usuario (Fig. 5.2-4). Las herramientas modernas, en relación con la recogida manual de volúmenes, aceleran considerablemente el proceso QTO y reducen la probabilidad de errores. Permiten extraer los datos

directamente del modelo del proyecto, eliminando la necesidad de recalcular manualmente los volúmenes a partir de los planos, como se hacía en el pasado.

Independientemente del método utilizado, ya sea BIM abierto o BIM cerrado, es posible obtener una tabla QTO - idéntica con los volúmenes de los elementos del proyecto (Fig. 5.2-4). Sin embargo, cuando se trabaja con datos de proyecto en conceptos CAD - (BIM-), los usuarios dependen de herramientas especializadas y API proporcionadas por los proveedores (Fig. 3.2-13). Esto crea capas adicionales de dependencia y requiere el aprendizaje de esquemas de datos únicos, al tiempo que limita el acceso directo a los datos.

Debido a la naturaleza cerrada de los datos CAD, la obtención de tablas QTO y otros parámetros complica la automatización de los cálculos y la integración con sistemas externos. Utilizando herramientas de acceso directo a las bases de datos y transfiriendo los datos CAD del proyecto mediante herramientas de ingeniería inversa a un formato de marco de datos estructurado abierto (Fig. 4.1-13), puede obtenerse una tabla QTO idéntica con una sola línea de código (Fig. 5.2-5 - variante con datos granulares).

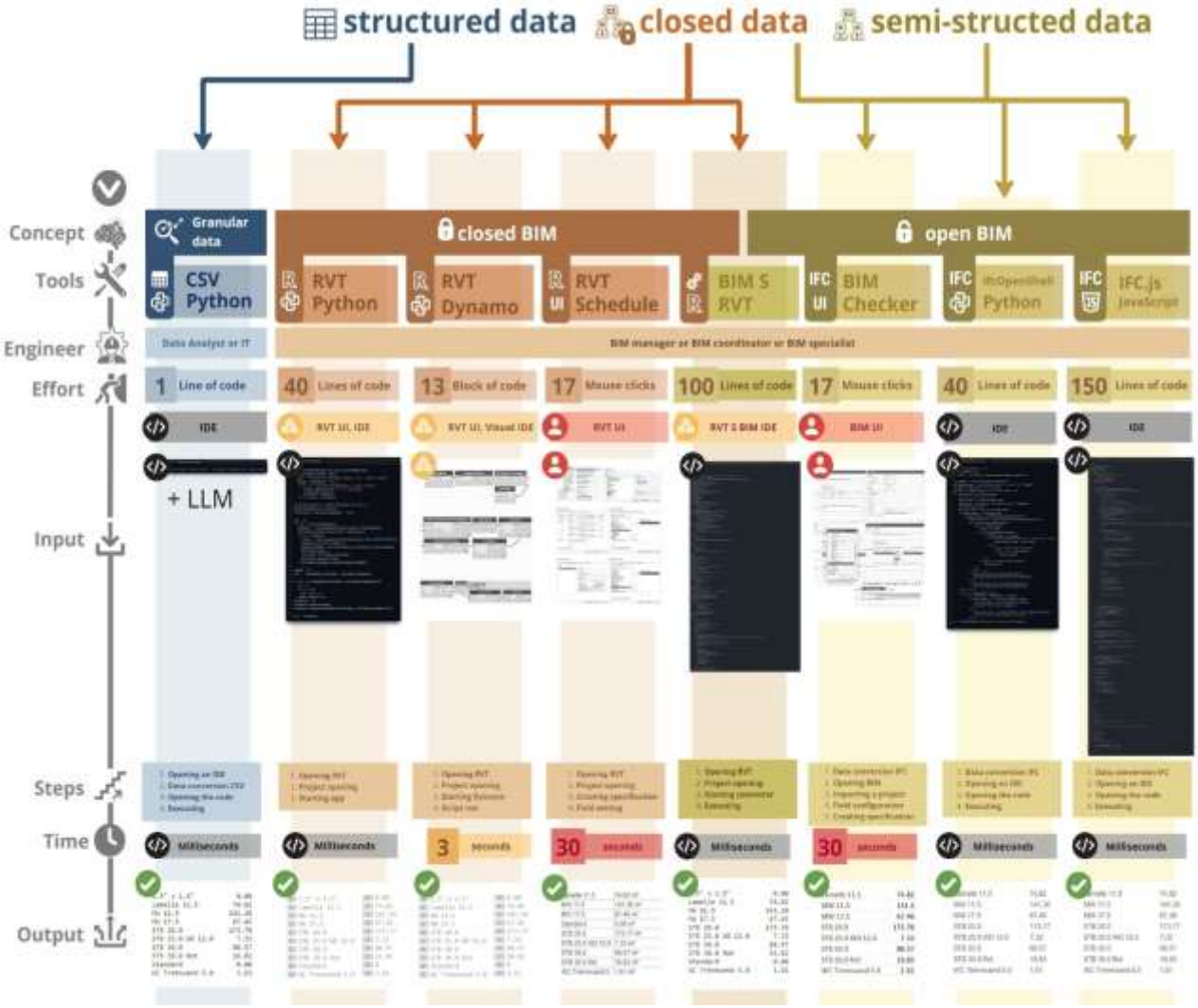


Fig. 5.2-5 Distintas herramientas producen los mismos resultados en forma de tablas de atributos de entidades de proyecto, pero con distintos costes de mano de obra.

Cuando se utilizan datos estructurados abiertos de proyectos CAD, como se menciona en el capítulo "Conversión de datos CAD (BIM) en forma estructurada", el proceso de agrupación, QTO, se simplifica enormemente.

Los enfoques basados en el uso de datos estructurados abiertos o el acceso directo a bases de datos de modelos CAD están libres de las restricciones de marketing asociadas al acrónimo BIM. Se basan en herramientas probadas y utilizadas desde hace tiempo en otros sectores (Fig. 7.3-10 Proceso ETL).

> Según el estudio de McKinsey "Open Data: Unleash Innovation and Productivity with Streaming Information" [102] realizado en 2013, el uso de datos abiertos puede crear oportunidades de ahorro de entre 30.000 y 50.000 millones de dólares al año en el diseño, la ingeniería, la adquisición y la construcción de instalaciones de generación de energía. Esto se traduce en un ahorro del 15% en los costes de capital de construcción.

Trabajar con datos estructurados abiertos (granulares) simplifica la recuperación y el procesamiento de la información, reduce la dependencia de plataformas BIM especializadas y abre el camino a la automatización sin necesidad de utilizar sistemas propietarios ni modelos de datos paramétricos y complejos procedentes de formatos CAD.

Automatización de QTO mediante LLM y datos estructurados

La traducción de datos no estructurados a una forma estructurada mejora significativamente la eficacia de varios procesos: simplifica el procesamiento de datos (Fig. 4.1-1, Fig. 4.1-2) y acelera el proceso de validación al hacer que los requisitos sean claros y transparentes, como ya hemos comentado en los capítulos anteriores. Del mismo modo, la traducción de los datos CAD (BIM) a un formato abierto estructurado (Fig. 4.1-12, Fig. 4.1-13) facilita el proceso de agrupación de atributos y el proceso QTO.

La tabla de atributos de QTO tiene una forma estructurada, por lo que al utilizar datos estructurados de CAD, trabajamos con un único modelo de datos (Fig, 5.2-5), lo que elimina la necesidad de convertir y traducir los modelos de datos del proyecto y las reglas de agrupación a un denominador común. Esto nos permite agrupar datos por uno o varios atributos con una sola línea de código. En cambio, en el BIM abierto y el BIM cerrado, donde los datos se almacenan en formatos semiestructurados, paramétricos o cerrados, el procesamiento requiere decenas o incluso cientos de líneas de código, así como el uso de la API para interactuar con la información sobre geometría y atributos.

- Ejemplo de agrupación de un proyecto estructurado QTO por un atributo. Consulta de texto en cualquier sala de chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otra):

```
Tengo CAD -project como DataFrame - por favor, filtre los datos del proyecto para obtener elementos con el parámetro "Tipo" que contenga sólo el valor "Tipo 1⏎
```

- Es muy probable que la respuesta de LLM sea en forma de código Python utilizando Pandas:

Fig. 5.2-6 Una línea de código escrita con LLM, permite agrupar todo un proyecto CAD por el atributo "Tipo" y obtener el grupo de elementos deseado.

Gracias a la sencilla estructura del DataFrame bidimensional no necesitamos explicar el esquema LLM ni el modelo de datos, lo que acorta los pasos de interpretación y acelera la creación de soluciones finales. Antes, escribir incluso un código sencillo requería aprender lenguajes de programación, pero ahora los modernos modelos de lenguaje (LLM) nos permiten traducir automáticamente la lógica del proceso en código cuando trabajamos con datos estructurados mediante consultas de texto.

> La automatización y los modelos lingüísticos de LLM pueden eliminar por completo la necesidad de que los profesionales que trabajan con la agrupación y el procesamiento de datos CAD (BIM) tengan que aprender lenguajes de programación o herramientas BIM, ya que ofrecen la posibilidad de resolver problemas mediante consultas de texto.

La misma consulta -agrupar todos los elementos del proyecto de la categoría "muros" y calcular los volúmenes de cada tipo (Fig. 5.2-5)- que en un entorno CAD (BIM) requiere 17 clics en la interfaz o escribir 40 líneas de código, en herramientas abiertas de tratamiento de datos (por ejemplo, SQL o Pandas) parece una consulta sencilla e intuitiva:

- Con una sola línea en Pandas:

```
df[df['Categoría'].isin(['OST_Paredes'])].groupby('Tipo')['Volumen'].sum()
```

Decodificación del código: tomar de df (DataFrame) los elementos, cuyo atributo-columna "Categoría", tiene valores "OST_Paredes", agrupar todos los elementos obtenidos por atributo-columna "Tipo" y sumar para el grupo de elementos obtenidos el atributo "Volumen".

- Agrupación de un proyecto estructurado recuperado de CAD mediante SQL:

```
SELECT Tipo, SUM(Volumen) COMO VolumenTotal
FROM elementos
WHERE Categoría = 'OST_Paredes'
GROUP BY Tipo;
```

- Con la ayuda de LLM podemos escribir una petición de agrupación a la base de datos del proyecto como una simple referencia de texto - un prompt (Fig. 5.2-7):

> Para el marco de datos del proyecto, agrupe los elementos por el parámetro "Tipo", pero sólo los elementos cuyo parámetro "Categoría" sea igual a "OST_Paredes" u "OST_Columnas" y sume el parámetro de columna "Volumen" para el resultante. ⏎

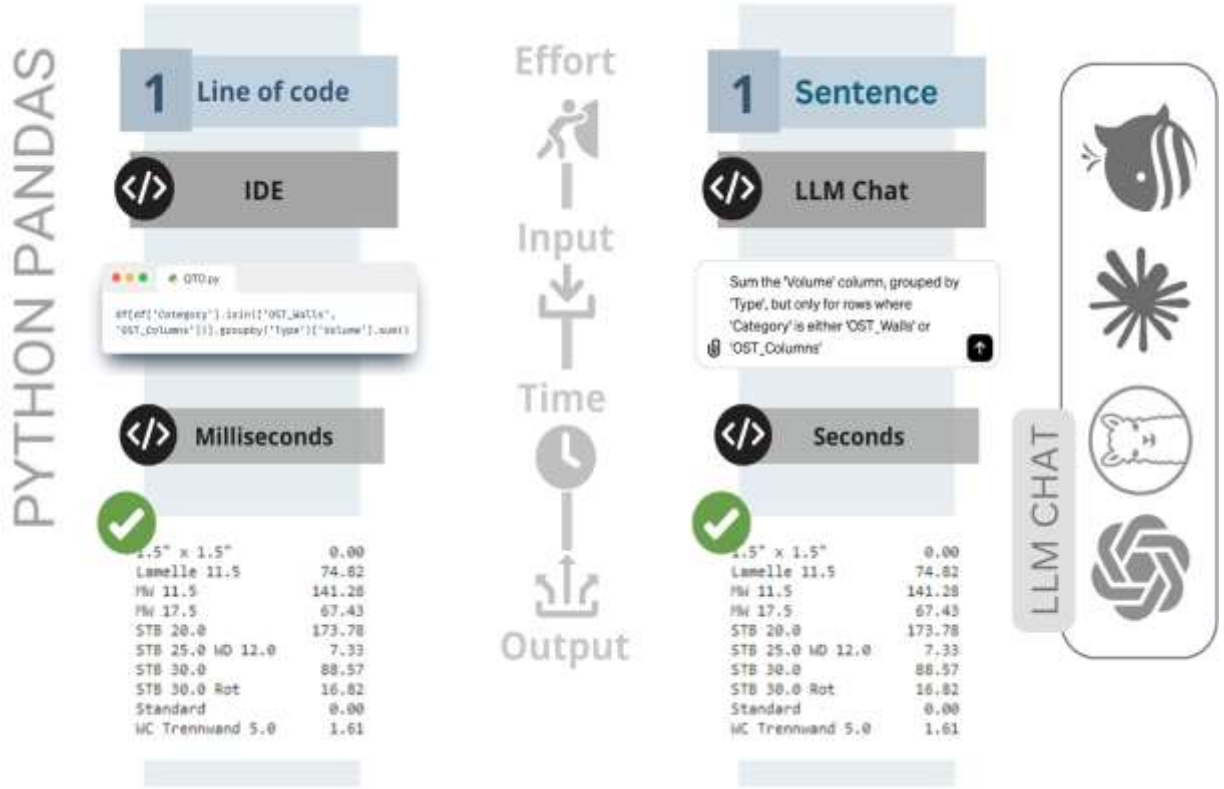


Fig. 5.2-7 Utilizando SQL, Pandas y LLM la automatización del procesamiento de datos es ahora posible con unas pocas líneas de código y consultas de texto.

La obtención de QTO a partir de datos CAD mediante herramientas LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), cambia radicalmente los métodos tradicionales de extracción de información de atributos, datos cuantitativos y volumétricos de objetos individuales y grupos de objetos.

Ahora, incluso los jefes de proyecto, especialistas en cálculo de costes o logística que no tengan profundos conocimientos de diseño y no dispongan de software CAD especializado - (BIM-) vendedores, teniendo acceso a la base de datos CAD pueden obtener el volumen total de elementos de la categoría de muros u otros objetos en cuestión de segundos, simplemente escribiendo o dictando una consulta.

En las consultas de texto (Fig. 5.2-8) el agente LLM del modelo procesa la petición del usuario de aplicar una determinada función a uno o más parámetros - columnas de la tabla. Como resultado, el usuario en comunicación con el LLM recibe o bien un nuevo parámetro-columna con nuevos valores o bien un valor específico después de la agrupación

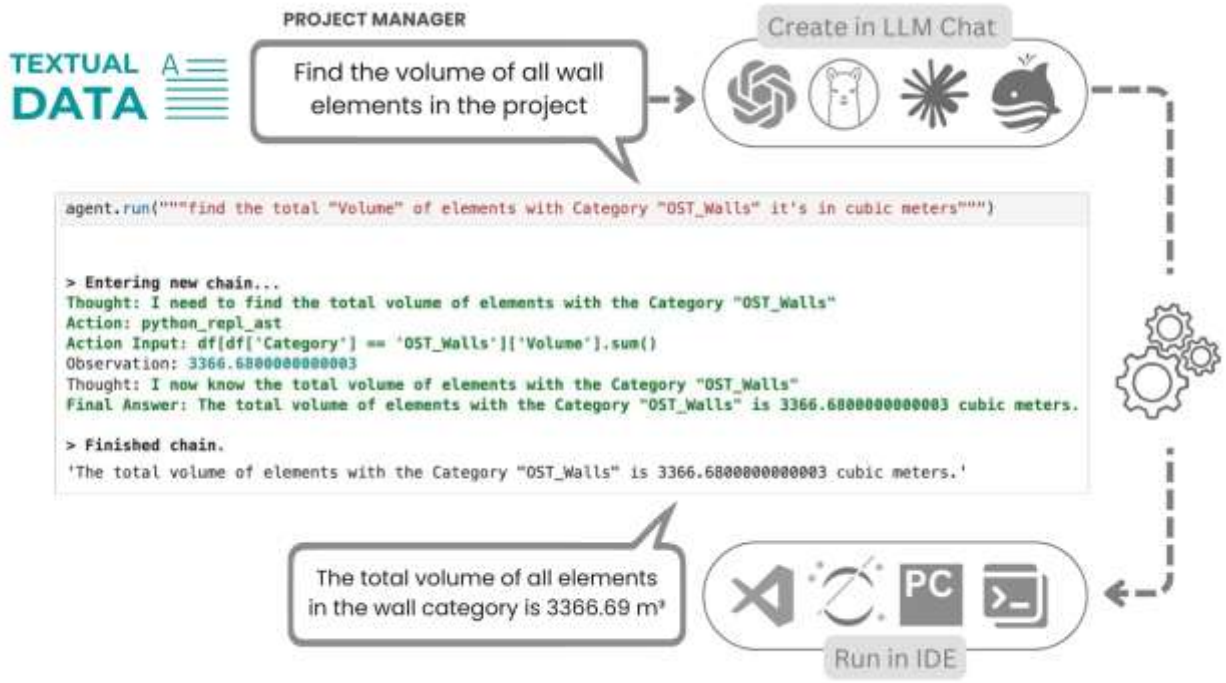


Fig. 5.2-8 El modelo LLM, cuando trabaja con datos estructurados, entiende a partir del contexto de una consulta de texto sobre qué agrupación y atributos pregunta el usuario.

Si sólo es necesario obtener cantidades para un grupo de elementos, basta con realizar una simple consulta QTO (Fig, 5.2-7) sobre los datos del modelo CAD. Sin embargo, cuando se calcula un presupuesto o una estimación para un proyecto completo que consta de muchos grupos de elementos, a menudo es necesario extraer cantidades para todos los tipos de elementos (clases0, donde cada categoría de elementos se procesa por separado - con agrupación por atributos relevantes.

En la práctica de los estimadores y evaluadores, se utilizan reglas de agrupación y cálculo individuales para distintos tipos de objetos. Por ejemplo, las ventanas suelen agruparse por plantas o zonas (parámetro de agrupación - atributo Nivel, Habitaciones), y las paredes - por material o tipo de construcción (parámetro Material, Tipo). Para automatizar el proceso de agrupación, dichas reglas se describen de antemano en en

forma de tablas de reglas de agrupación. Estas tablas actúan como plantillas de configuración que definen qué atributos deben utilizarse en los cálculos para cada grupo de elementos del proyecto.

QTO cálculo de todo el proyecto mediante reglas de grupo a partir de una hoja de cálculo Excel

En proyectos de construcción reales, a menudo es necesario realizar la agregación por varios atributos simultáneamente dentro de un grupo de elementos. Por ejemplo, cuando se trabaja con la categoría "Ventanas" (donde el atributo Categoría contiene valores como OST_Ventanas o IfcVentanas), los elementos pueden agruparse no sólo por tipo -por ejemplo, por el valor del campo Nombre de tipo o Tipo-, sino también por características adicionales, como el nivel de conductividad térmica especificado en el atributo correspondiente. Esta agrupación multidimensional permite obtener resultados más precisos para un grupo concreto. Del mismo modo, al calcular las categorías de paredes o suelos, pueden utilizarse combinaciones arbitrarias de atributos -como material, nivel, suelo, resistencia al fuego y otros parámetros- como filtros o criterios de agrupación (Fig, 5.2-9).

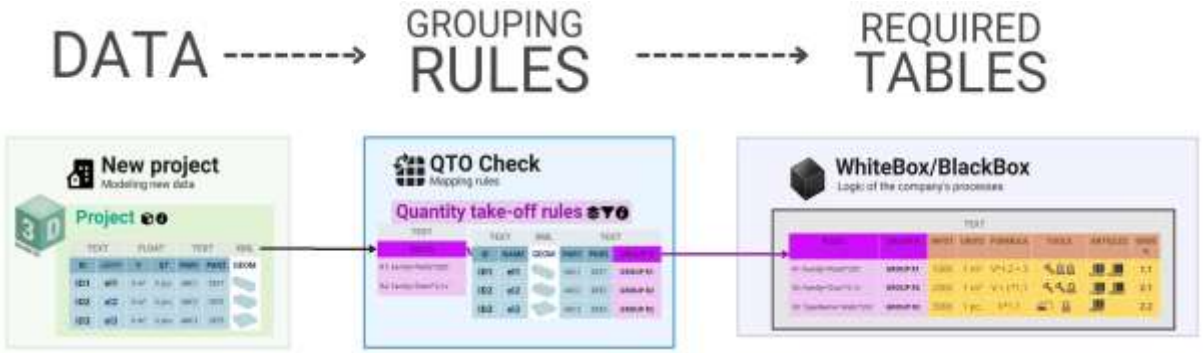


Fig. 5.2-9 Para cada grupo o categoría de entidades de un proyecto, existe una fórmula de agrupación diferente que consta de uno o varios criterios.

El proceso de definición de dichas reglas de agrupación es similar al proceso de creación de requisitos de datos descrito en el capítulo "Creación de requisitos y comprobación de la calidad data " (Fig, 4.4-5), donde tratamos en detalle el trabajo con modelos de datos. Estas reglas de agrupación y cálculo garantizan la precisión y pertinencia de los resultados para calcular automáticamente los atributos totales de la cantidad o volumen de una categoría de entidades, teniendo en cuenta todas las condiciones necesarias que deben tenerse en cuenta en los cálculos y cómputos.

- El siguiente ejemplo de código filtra la tabla de proyectos para que el conjunto de datos resultante contenga únicamente entidades en las que la columna de atributos "Categoría" contenga los valores "OST_Windows" o "IfcWindows" y, al mismo tiempo, la columna de atributos "Tipo" contenga el valor "Tipo 1":

> Tengo un proyecto DataFrame - filtrar los datos para que sólo permanezcan en el conjunto de datos los elementos que tengan el atributo "Categoría" que contenga los valores "OST_Windows" o "IfcWindows" y al mismo tiempo el atributo Tipo contenga el valor "Tipo 1⏎

- Respuesta de LLM:

Fig. 5.2-10 Una única línea de código, similar a la fórmula de Excel, permite agrupar todas las entidades del proyecto por varios atributos.

El código resultante (Fig. 5.2-10) tras la traducción de los datos CAD en formatos abiertos estructurados (Fig. 4.1-13) se puede ejecutar en uno de los IDEs populares (entornos de desarrollo integrados) que hemos mencionado anteriormente en modo offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con plugin Python, JupyterLab o herramientas online populares: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Para recuperar las entidades de proyecto en el formulario QTO DataFrame bajo la categoría "Windows" sólo con un valor específico de conductividad térmica, podemos utilizar la siguiente consulta al LLM:

> Tengo un proyecto DataFrame - filtrar los datos para que sólo queden en el conjunto de datos los registros con "Categoría" que contenga los valores "OST_Windows" o "IfcWindows", y al mismo tiempo la columna ConductividadTérmica debe tener un valor de 0.⏎

- Respuesta de LLM:

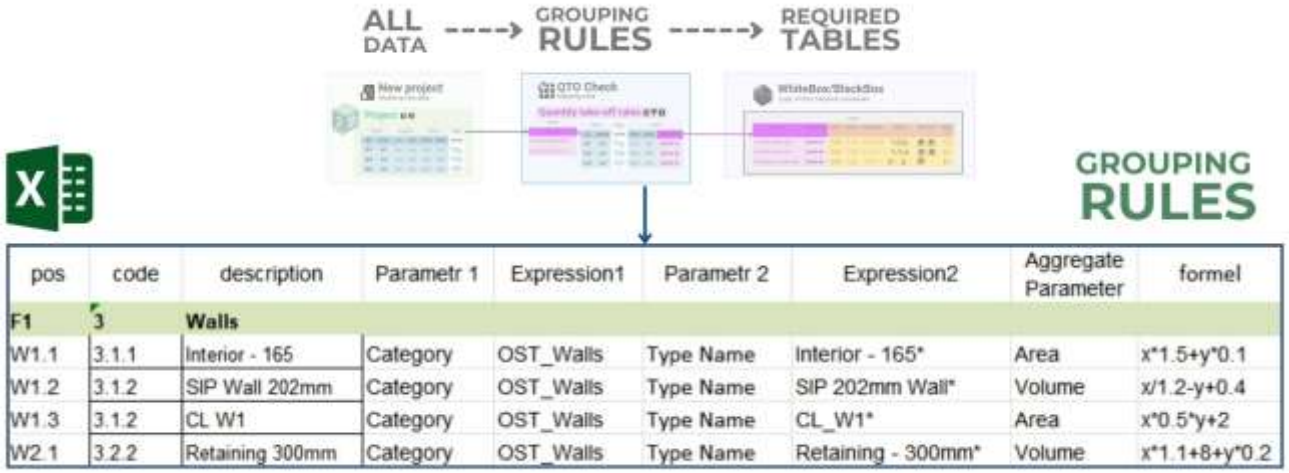
Fig. 5.2-11 El extremadamente sencillo lenguaje de consulta Pandas Python permite ejecutar QTOs para cualquier número de proyectos simultáneamente.

En la respuesta recibida de LLM (Fig. 5.2-11), se utiliza la condición lógica "&" para combinar dos criterios: valor de conductividad térmica y pertenencia a una de las dos categorías. El método "isin" comprueba si el valor del atributo-columna "Categoría" figura en la lista proporcionada.

En proyectos con un gran número de grupos de elementos, con una lógica de agrupación diferente - para cada categoría de entidades de proyecto (por ejemplo: ventanas, puertas, forjados) deben establecerse reglas de agrupación individuales, que pueden incluir coeficientes adicionales o fórmulas de cálculo de atributos totales. Estas fórmulas (Fig, 5.2-12 "formel" de atributo, por ejemplo, valor x de cantidad y volumen y de grupo) y coeficientes tienen en cuenta las características únicas de cada grupo, por ejemplo:

- % de adiciones al volumen de material para tener en cuenta los rebasamientos
- cantidad adicional fija de material
- ajustes relacionados con posibles riesgos y errores de cálculo en forma de fórmulas

Una vez formuladas las reglas de filtrado y agrupación en forma de fórmulas de parámetros para cada categoría de elementos, pueden almacenarse como una tabla línea por línea, por ejemplo, en formato Excel (Fig, 5.2-12). Al almacenar estas reglas de forma estructurada, el proceso de extracción, filtrado y agrupación de los datos del proyecto puede automatizarse por completo. En lugar de escribir manualmente muchas consultas separadas, el sistema simplemente lee la tabla de parámetros y aplica las reglas apropiadas al modelo (el marco de datos global del proyecto (Fig, 4.1-13)), generando tablas QTO finales para cada categoría de elementos del proyecto.



pos	code	description	Parametr 1	Expression1	Parametr 2	Expression2	Aggregate Parameter	formel
F1	3	**Walls**						
W1.1	3.1.1	Interior - 165	Category	OST_Walls	Type Name	Interior - 165*	Area	x*1.5+y*0.1
W1.2	3.1.2	SIP Wall 202mm	Category	OST_Walls	Type Name	SIP 202mm Wall*	Volume	x/1.2-y+0.4
W1.3	3.1.2	CL W1	Category	OST_Walls	Type Name	CL_W1*	Area	x*0.5*y+2
W2.1	3.2.2	Retaining 300mm	Category	OST_Walls	Type Name	Retaining - 300mm*	Volume	x*1.1+8+y*0.2

Fig. 5.2-12 Tabla de agrupación de atributos de QTO establece reglas para agrupar los elementos del proyecto, proporcionando un número total y un alcance precisos para cada categoría.

Las reglas recopiladas le permitirán agrupar todo el proyecto y realizar todos los cálculos necesarios, incluidos los ajustes de los atributos de volumen. Como resultado, los volúmenes se llevan al "volumen real", que se utiliza para los cálculos y cálculos, y no los que estaban originalmente en la etapa de diseño en el modelo CAD.

> En el proceso de creación automática de tablas de volumen QTO para todo el proyecto, la aplicación debe recorrer todas las categorías de la tabla de reglas de agrupación, tomar los atributos de agrupación, agrupar por ellos todos los elementos del proyecto y agregar el atributo de volumen para ese grupo, multiplicándolo además por un factor o coeficiente de refino.

Pidamos a LLM que nos escriba código para una solución de este tipo, en la que el código tendría que cargar dos tablas: una tabla de reglas de agrupación (Fig, 5.2-12) y una tabla de datos del propio proyecto (Fig, 4.1-13), y luego aplicar las reglas de agrupación, agrupar los elementos según las reglas dadas, calcular los valores agregados y guardar los resultados en un nuevo archivo de Excel.

- Envíe una solicitud de texto al chat LLM:

> Necesito un código para leer los datos del proyecto del archivo 'basic_sample_project.xlsx' y luego las reglas de 'Grouping_rules_QTO.xlsx' para agrupar todos los datos por 'Parameter 1' y 'Parameter 2', agregar 'Aggregate Parameter', filtrar por 'Expression2', realizar cálculos desde 'Formel1' y guardar la tabla QTO en 'QTQ_table2.xlsx⏎

- Respuesta de LLM:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Reading CSV and Excel files with project data and grouping rules respectively
df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])

# Cleaning numerical columns in the dataframe by removing non-numeric characters
numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
df[numeric_cols] = df[numeric_cols
].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+.\d+)')[0].fillna(0)))

# Define a function to process each row in the grouping rules dataframe
def process_row(row):
    # Group by specified parameters and calculate sum and count for the aggregate parameter
    group = df.groupby([row["Parameter 1"], row["Parameter 2"]
    ][row["Aggregate Parameter"]].agg(['sum', 'count'])
    group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]

    # Perform calculation defined in the rules
    group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
    return group.reset_index()

# Apply the function to each row in the grouping rules dataframe and concatenate the results
dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)

# Save the processed data to an Excel file using the OpenPyXL engine
dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)
```

Run in IDE

Fig. 5.2-13 Código Python Pandas procesa datos de diseño de un proyecto CAD , agrupándolos según reglas definidas en Excel.

El resultado final de la ejecución del código (Fig. 5.2-13) será una tabla entidad-grupo que contiene no sólo los atributos de volumen generalizados del modelo original CAD- (BIM-), sino también un nuevo atributo de volumen real que tiene en cuenta todos los requisitos para la correcta creación de cálculos y estimaciones (Ejemplo Fig. 5.2-14).

QTO TABLE FOR ALL GROUPS OF ELEMENTS ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id´s	After Calculation	Measure
0	OST_Walls	Interior - 165 Partitior	17	3	424922, 425745,	25.8	m2
1	OST_Walls	SIP 202mm Wall - con	21.95	4	198694, 428588,	13.9	m3
2	OST_Walls	CL_W1	230	6	493612, 493697,	692	m2
3	OST_Walls	Retaining - 300mm Cc	57.93	10	599841, 599906,	72.7	m3

Fig. 5.2-14 El atributo "Después del cálculo" se añade a la tabla resumen después de ejecutar el código que calculará automáticamente el volumen real.

El código resultante (Fig. 5.2-13) puede ejecutarse en uno de los IDE populares (que hemos mencionado anteriormente) y aplicarse a cualquier número de proyectos entrantes existentes o nuevos (RVT, IFC, DWG, NWS, DGN etc.), ya sean unos pocos proyectos o quizá cientos de proyectos en diferentes formatos de forma estructurada (Fig. 5.2-15).

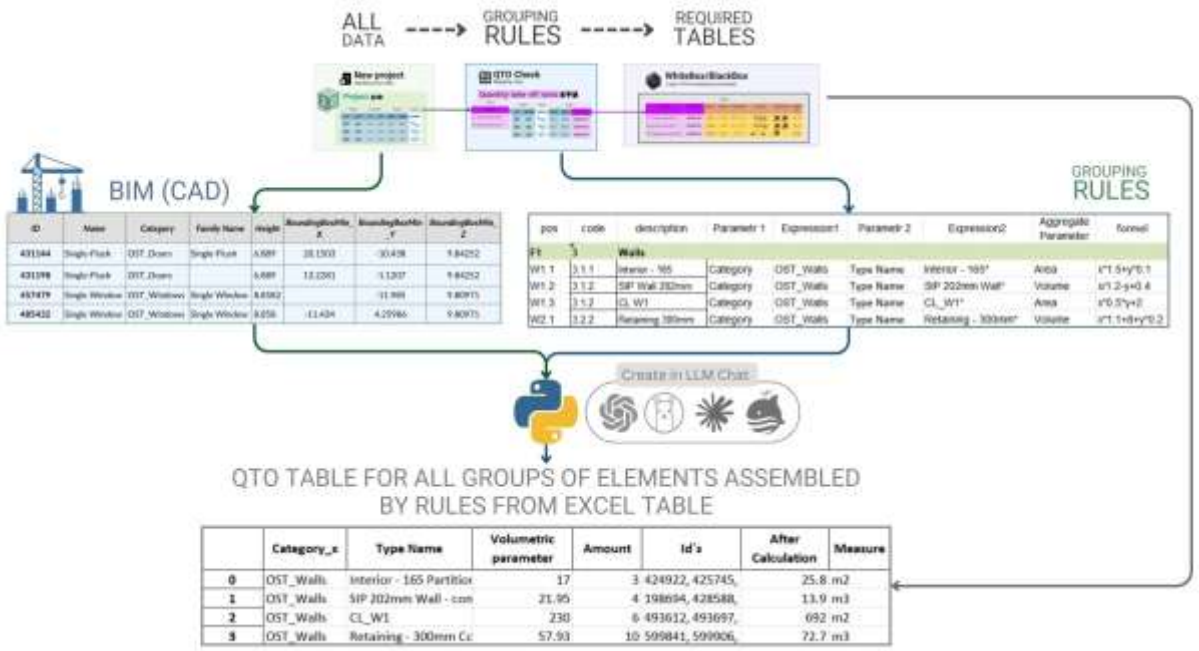


Fig. 5.2-15 El proceso automático de agrupación de datos de construcción vincula los datos BIM (CAD) a las tablas QTO mediante reglas de una hoja de cálculo Excel.

El proceso de recopilación de datos volumétricos personalizado y parametrizado (Fig, 5.2-15) permite la recopilación totalmente automatizada de datos sobre atributos cuantitativos y volúmenes de los elementos del proyecto para seguir trabajando con ellos, incluida la estimación de costes, la logística, los calendarios de trabajo y el cálculo de la huella de carbono y otras tareas analíticas.

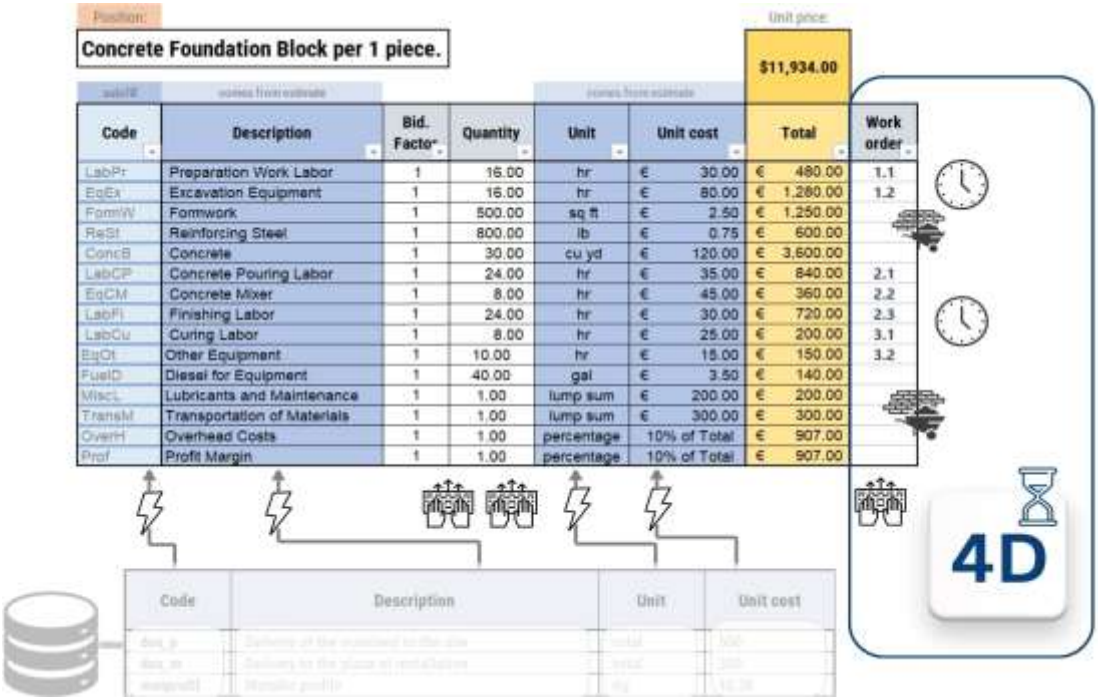
Una vez aprendidas las herramientas que nos permiten organizar y agrupar fácilmente grupos de elementos de proyectos en función de determinados atributos, ya estamos preparados para integrar los proyectos agrupados y filtrados con diversos cálculos de la empresa y escenarios empresariales.

CAPÍTULO 5.3.
4D, 6D -8D Y CÁLCULO DE LAS EMISIONES DE DIÓXIDO DE CARBONO CO_2

Modelo 4D: integración del tiempo en las estimaciones de construcción

Además del cálculo de costes, una de las aplicaciones clave de los datos de diseño en la construcción es la determinación de los parámetros de tiempo, tanto para operaciones de construcción individuales como para todo el proyecto. El método de estimación basado en los recursos y la base de datos de cálculo asociada, que se trató en detalle en el capítulo anterior "Cálculos y estimaciones para proyectos de construcción", se utilizan a menudo como base para el cálculo automatizado del tiempo y la creación de un calendario para la ejecución de los trabajos.

El enfoque basado en los recursos no sólo tiene en cuenta los costes de material, sino también los recursos de tiempo. En el cálculo de costes, puede asignarse a cada proceso un atributo de orden de trabajo (Fig, 5.3-1 - Parámetro de orden de trabajo) y la cantidad de tiempo y coste asociados a la ejecución de ese proceso. Estos parámetros son especialmente importantes para describir operaciones que no tienen un precio de mercado fijo y no son directamente adquiribles, como el uso de equipos de construcción, el empleo de mano de obra o los procesos logísticos (que suelen expresarse normalmente en horas). En estos casos, el coste no lo determina el departamento de contratación, sino directamente la empresa ejecutora en función de normas internas o índices de producción (Fig, 5.3-1).



Code	Description	Bid. Factor	Quantity	Unit	Unit cost	Total	Work order
LabPr	Preparation Work Labor	1	16.00	hr	€ 30.00	€ 480.00	1.1
EqEx	Excavation Equipment	1	16.00	hr	€ 80.00	€ 1,280.00	1.2
FormW	Formwork	1	500.00	sq ft	€ 2.50	€ 1,250.00	
ReSt	Reinforcing Steel	1	800.00	lb	€ 0.75	€ 600.00	
ConcB	Concrete	1	30.00	cu yd	€ 120.00	€ 3,600.00	
LabCP	Concrete Pouring Labor	1	24.00	hr	€ 35.00	€ 840.00	2.1
EqCM	Concrete Mixer	1	8.00	hr	€ 45.00	€ 360.00	2.2
LabFi	Finishing Labor	1	24.00	hr	€ 30.00	€ 720.00	2.3
LabCu	Curing Labor	1	8.00	hr	€ 25.00	€ 200.00	3.1
EqOt	Other Equipment	1	10.00	hr	€ 15.00	€ 150.00	3.2
FuelD	Diesel for Equipment	1	40.00	gal	€ 3.50	€ 140.00	
MiscL	Lubricants and Maintenance	1	1.00	lump sum	€ 200.00	€ 200.00	
TransM	Transportation of Materials	1	1.00	lump sum	€ 300.00	€ 300.00	
OverH	Overhead Costs	1	1.00	percentage	10% of Total	€ 907.00	
Prof	Profit Margin	1	1.00	percentage	10% of Total	€ 907.00	

Fig. 5.3-1 Los cálculos del trabajo en el método de estimación basado en los recursos incluyen la temporización de las horas de trabajo.

Así, los cálculos a nivel de costes incluyen no sólo los costes de combustible y material (coste de compra),

sino también el tiempo empleado por los conductores, técnicos y trabajadores auxiliares en la obra. En el ejemplo dado (Fig. 5.3-1), la tabla de costes es un cálculo del coste de instalación de un bloque de cimentación, que incluye las fases constituyentes del trabajo, como la preparación, la instalación del armazón y el vertido del hormigón, así como los materiales y la mano de obra necesarios. Sin embargo, algunas operaciones, como los trabajos de preparación, pueden no tener costes de material, pero sí importantes costes de mano de obra temporal expresados en horas/hombre.

Para planificar la secuencia de trabajo (para el plan de trabajo) en la obra, se añade manualmente el atributo "Orden de trabajo" a la tabla de cálculo (Fig. 5.3-1). Este atributo se especifica en una columna adicional sólo para los elementos cuya unidad de medida se expresa en tiempo (hora, día). Este atributo se añade al código de trabajo, la descripción, la cantidad, la unidad de medida (parámetro "Unidad") y los costes. La secuencia numérica (parámetro "Orden de trabajo") de las actividades permite establecer el orden de ejecución de las tareas en la obra y utilizarlo para la programación.

Calendario de construcción y su automatización basada en datos de cálculo de costes

El calendario de construcción es una representación visual del plan de trabajo y los procesos que se llevarán a cabo en el marco de la ejecución del proyecto. Se crea sobre la base de cálculos detallados de recursos (Fig, 5.3-1), donde cada tarea-trabajo se programa, además del coste de los recursos, por tiempo y secuencia.

A diferencia de los enfoques de promediación, en los que las estimaciones de tiempo se basan en las horas típicas de instalación de materiales o equipos, en el método basado en los recursos la planificación se basa en los datos reales del cálculo de costes. Cada partida de la estimación relacionada con los costes laborales se basa en el calendario aplicado, que tiene en cuenta las condiciones reales de utilización de los recursos durante el periodo de trabajo. El ajuste de las horas productivas mediante coeficientes a nivel del cálculo de costes (Fig. 5.3-1 parámetro "Factor Bid."), permite tener en cuenta las diferencias de productividad y las peculiaridades estacionales que afectan al calendario de trabajo.

Para determinar las fechas de inicio y fin del proceso para el calendario de construcción en el diagrama de Gantt, tomamos los valores del atributo cantidad de tiempo para cada elemento del cálculo de costes de los bloques de cimentación y los multiplicamos por el número de bloques (en este caso, el número de bloques de cimentación de hormigón). Este cálculo da la duración de cada tarea. A continuación, trazamos estas duraciones en una línea de tiempo, a partir de la fecha de inicio del proyecto, para crear un gráfico y el resultado es una representación visual que muestra cuándo debe empezar y terminar cada tarea. El parámetro "Orden de trabajo" de los procesos nos permite además saber si el proceso de trabajo se está ejecutando en paralelo ("Orden de trabajo", por ejemplo 1.1-1.1) o secuencialmente (1.1-1.2).

> **Diagrama de Gantt** es una herramienta gráfica de planificación y gestión de proyectos que representa las tareas como barras horizontales en una línea de tiempo. Cada barra representa la duración de la tarea, su inicio y su fin.

Un calendario, o diagrama de Gantt, ayuda a los directores de proyecto y a los trabajadores a entender

claramente cuándo y en qué secuencia deben realizarse las distintas fases de la construcción, garantizando un uso eficiente de los recursos y el cumplimiento de los plazos.

Presentemos un calendario para la instalación de tres bloques de cimentación de hormigón utilizando los cálculos de la tabla anterior. Utilizando la tabla de costes (Fig. 5.3-1) del ejemplo anterior, pida al LLM que programe la instalación de los 3 elementos de bloques de cimentación, por ejemplo, para el primero de mayo de 2024.

Para enviar un cálculo de costes a LLM, podemos cargar la tabla de costes en formato XLSX o simplemente insertar una captura de pantalla de una imagen JPEG del cálculo de costes directamente en el chat de LLM (Fig. 5.3-1). LLM encontrará independientemente una biblioteca para visualizar la imagen de la tabla y, multiplicando los atributos de tiempo de las actividades de la tabla por su volumen, añadirá todos los datos en un gráfico.

- Envíe una solicitud de texto a LLM:

> Cree un programa de trabajo en forma de diagrama de Gantt utilizando los valores de la tabla de tiempos (Fig. 5.3-1 como JPEG en el apéndice) para la instalación de 3 bloques de cimentación. Los trabajos para cada bloque se realizarán secuencialmente. El inicio de los trabajos se fijará a partir del 01/05/2024 ⏎

- Respuesta de LLM

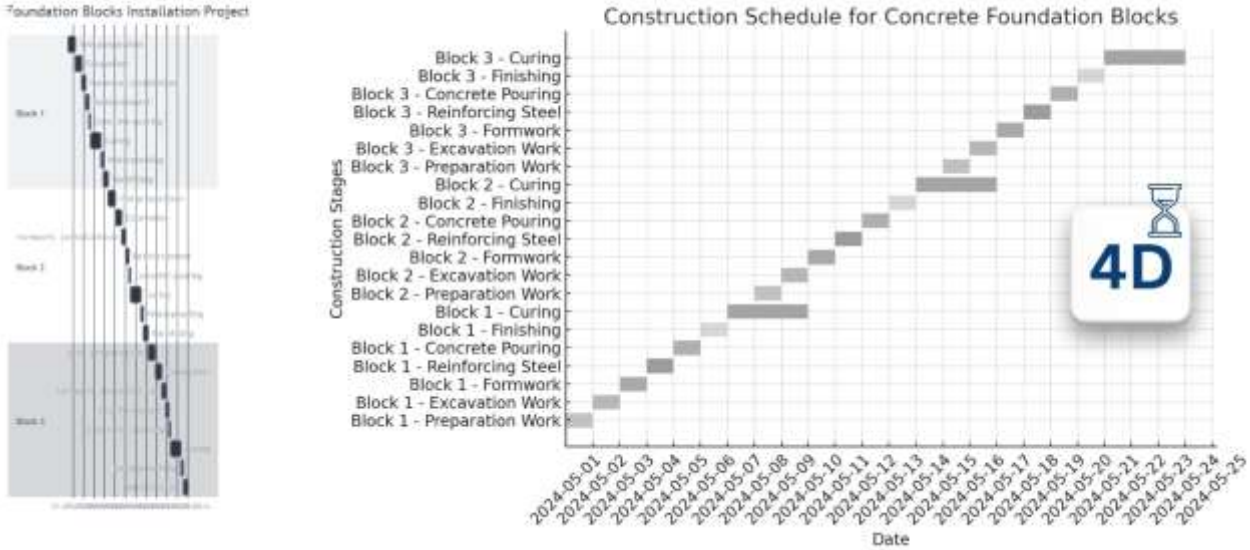


Fig. 5.3-2 Generado automáticamente por varios LLM El diagrama de Gantt muestra las etapas de construcción de tres bloques de hormigón, según las condiciones del prompt.

El gráfico resultante (Fig. 5.3-2) es un diagrama de tiempo en el que cada barra horizontal corresponde a una determinada fase de trabajo en el bloque de cimentación y muestra la secuencia de operaciones (parámetro "Orden de trabajo"), como preparación, excavación, instalación de encofrados, armadura, vertido

de hormigón y acabado, es decir, aquellos procesos que han rellenado los parámetros de tiempo y secuencia en los cálculos.

Un calendario de este tipo (Fig. 5.3-2) no tiene en cuenta las limitaciones relacionadas con los días laborables, los turnos o las normas de tiempo de trabajo, sino que está pensado únicamente para la visualización conceptual del proceso. Un calendario preciso que refleje el paralelismo del trabajo puede complementarse con promts apropiados o instrucciones adicionales dentro de la sala de chat

Con un cálculo de costes único (Fig. 5.3-1), gracias a los atributos de volumen de la geometría 3D, es posible estimar automáticamente tanto el coste del proyecto mediante estimaciones automatizadas como, al mismo tiempo, calcular las características temporales de los grupos en forma de tablas o gráficos para las distintas variantes del proyecto (Fig. 5.3-3).

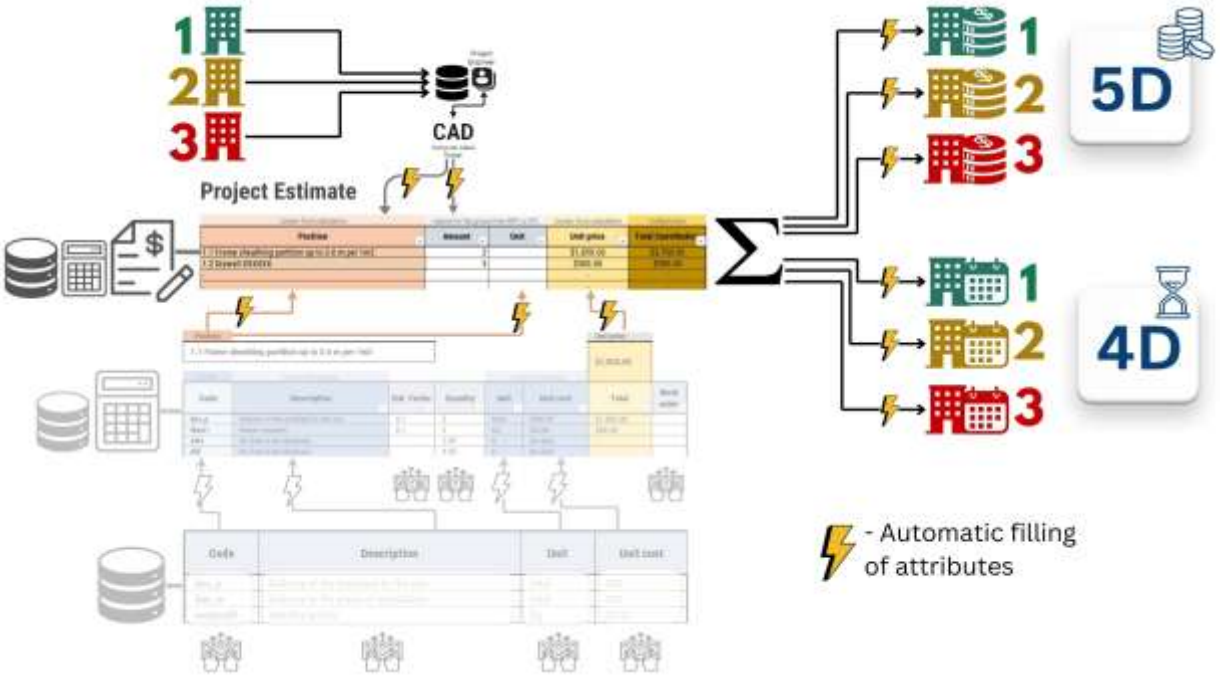


Fig. 5.3-3 Cálculo automático, permite la previsión instantánea y automática de costes y plazos para diferentes opciones de proyecto.

Los modernos sistemas modulares ERP -systems (Fig. 5.4-4), que cargan datos de modelos CAD utilizan métodos similares de cálculo automatizado del tiempo que reducen considerablemente el proceso de toma de decisiones. Esto permite planificar de forma instantánea y precisa los calendarios de trabajo y calcular el tiempo total necesario para completar todas las tareas de la realización del proyecto, teniendo en cuenta los precios reales.

Capas de atributos ampliadas 6D -8D: de la eficiencia energética a la garantía de seguridad

6D, 7D y 8D son niveles ampliados de modelización de la información, cada uno de los cuales aporta capas

adicionales de atributos al modelo de información global del proyecto, cuya base son los atributos del modelo 3D con su número y alcance. Cada capa adicional aporta parámetros específicos que son necesarios para la agrupación o identificación posterior en otros sistemas, como por ejemplo los sistemas de gestión de la propiedad (PMS), la gestión de instalaciones asistida por ordenador (CAFM), la gestión de proyectos de construcción (CPM) y los sistemas de gestión de la seguridad (SMS).

Fig. 5.3-4 Los atributos 6D, 7D y 8D del modelo de información de datos amplían la consideración de diversos aspectos del proyecto, desde la eficiencia energética hasta la seguridad.

- En **6D**además de la base de datos del proyecto (o marco de datos (Fig, 4.1-13)) con atributos geométricos y volumétricos de los elementos, se añade información (atributos-columnas) sobre sostenibilidad medioambiental. Esto incluye información relacionada con la eficiencia energética, la huella de carbono, la reciclabilidad de los materiales y el uso de tecnologías respetuosas con el medio ambiente. Estos datos permiten evaluar el impacto ambiental del proyecto, optimizar las decisiones sobre el mismo y alcanzar los Objetivos de Desarrollo Sostenible (ODS).
- **7D** complementan los atributos necesarios para gestionar el mantenimiento de los edificios. Se trata de datos sobre los programas de mantenimiento, los ciclos de vida de los componentes, la documentación técnica y el historial de reparaciones. Este conjunto de información garantiza que el modelo pueda integrarse con los sistemas de mantenimiento (CAFM, AMS), permite una planificación eficaz del mantenimiento, la sustitución de equipos y proporciona apoyo a lo largo de todo el ciclo de vida de la instalación.
- **8D** capa de atributos adicional, - incluye información relacionada con la seguridad, tanto en la fase de construcción como durante la posterior explotación. El modelo incluye medidas de seguridad para el personal, instrucciones de emergencia, sistemas de evacuación y requisitos de protección contra incendios. La integración de estos datos en el modelo digital ayuda a tener en cuenta los riesgos de antemano y a desarrollar soluciones arquitectónicas, de ingeniería y organizativas que tengan en cuenta los requisitos de salud y seguridad.

En forma tabular estructurada, las capas 4D a 8D representan atributos adicionales en forma de columnas con valores rellenados (Fig, 5.3-5) que se añaden a los atributos del modelo 3D ya rellenados, como el nombre, la categoría, el tipo y las características volumétricas. Los valores de las capas de atributos 6D, 7D y 8D contienen datos textuales y numéricos adicionales como el porcentaje de reciclado, la huella de carbono, el periodo de garantía, el ciclo de sustitución, la fecha de instalación, los protocolos de seguridad, etc.

ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO2e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO2e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO2e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO2e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Fig. 5.3-5 6D -8D añadir capas de atributos al modelo de información de datos, que ya contiene atributos geométricos y volumétricos del modelo 3D.

Para nuestra nueva ventana (Fig, 4.4-1), el elemento con el identificador W-NEW (Fig, 5.3-5) puede tener los siguientes atributos 3D -8D:

3D -attributes - información geométrica obtenida de los sistemas CAD:

- "Nombre de tipo" - elemento "Ventana"
- "Anchura" - 120 cm
- Además, puede añadir los puntos "Bounding Box " de un elemento o su "geometría BREP / MESH " como atributo independiente

Atributos de 6D - sostenibilidad medioambiental:

- Tasa de reciclaje del 90
- " Huella de carbono - 1.622 kg de CO_2

Atributos 7D - datos de gestión de objetos:

- "Periodo de garantía" - 8 años
- "Ciclo de sustitución" cumple 20 años
- "Mantenimiento" - obligatorio anualmente

Atributos de la 8D - garantizar el uso y funcionamiento seguros de los edificios:

- Ventana "Instalado" - por la empresa "XYZ Windows

- "Norma de seguridad": cumple la norma ISO 45001

Todos los parámetros registrados en una base o conjunto de datos (Fig. 5.3-5) son necesarios para que los especialistas de distintos departamentos puedan agruparlos, buscarlos o realizar cálculos. Esta descripción multidimensional basada en atributos de los objetos del proyecto proporciona una imagen completa de su ciclo de vida, sus requisitos operativos y muchos otros aspectos necesarios para el diseño, la construcción y la explotación del proyecto.

Estimación del CO_2 y cálculo de las emisiones de dióxido de carbono de los proyectos de construcción.

Junto al tema de la sostenibilidad de los proyectos de construcción en la fase 6D (Fig, 5.3-5), la construcción moderna se está centrando en la sostenibilidad medioambiental de los proyectos, donde uno de los aspectos clave pasa a ser la evaluación y minimización de las emisiones de dióxido de carbono CO_2 que se producen durante las fases del ciclo de vida de un proyecto (por ejemplo, fabricación e instalación).

> Estimar y calcular las emisiones de carbono de los materiales de construcción es un proceso mediante el cual se determinan las emisiones totales de carbono multiplicando los atributos volumétricos de un elemento o grupo de elementos utilizados en un proyecto por un factor de emisión de carbono adecuado para la categoría.

La consideración de las emisiones de carbono en la evaluación de los proyectos de construcción como parte de los criterios más amplios ESG (medioambientales, sociales y de gobernanza) añade un nuevo nivel de complejidad al análisis. Esto es especialmente importante para el cliente-inversor a la hora de obtener certificaciones relevantes como LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) o DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). La obtención de una de estas certificaciones puede aumentar significativamente la comerciabilidad de un inmueble, simplificar la puesta en servicio y garantizar el cumplimiento de los requisitos de los inquilinos orientados a la sostenibilidad (ESG). En función de los requisitos del proyecto, también pueden utilizarse las certificaciones HQE (Haute Qualité Environnementale, la norma francesa de construcción ecológica), WELL (WELL Building Standard, centrada en la salud y el confort de los usuarios) y GRESB (Global Real Estate Sustainability Benchmark)

> Medio ambiente, sociedad y gobernanza **ESG** (environ-mental, social and governance) es un amplio conjunto de principios que pueden utilizarse para evaluar la gobernanza corporativa y el impacto social y medioambiental de una empresa, tanto interna como externamente.

ESG, desarrollado originalmente a principios de la década de 2000 por fondos financieros para proporcionar a los inversores información sobre amplios criterios medioambientales, sociales y de gobernanza, ha evolucionado hasta convertirse en un indicador clave para evaluar tanto empresas como proyectos, incluidos los de construcción. Según las investigaciones de las principales consultoras, las consideraciones medioambientales, sociales y de gobernanza (ESG) se están convirtiendo en parte integrante del sector de la

construcción.

> Según EY (2023) "The Path to Carbon Neutrality", las empresas que aplican activamente los principios ESG no sólo reducen los riesgos a largo plazo, sino que también aumentan la eficiencia de sus modelos de negocio, lo que es especialmente importante en la transformación global de los mercados [103]. El informe ESG Awareness de PwC señala que la concienciación de las empresas sobre la importancia de los factores ESG oscila entre el 67% y el 97%, y que la mayoría de las organizaciones consideran que estas tendencias son clave para la sostenibilidad futura [104] y que las empresas, en su mayoría, están recibiendo una presión significativa de las partes interesadas para integrar los principios ESG.

> Así, la integración de los principios ESG en los proyectos de construcción no sólo contribuye a la obtención de certificados internacionales de sostenibilidad como LEED, BREEAM, DGNB, sino que también garantiza la sostenibilidad a largo plazo y la competitividad de las empresas del sector.

Uno de los factores que más influyen en la huella de carbono global de un proyecto de construcción son las fases de producción y logística de los materiales y componentes de construcción. Los materiales utilizados en la obra suelen tener un impacto decisivo en las emisiones totales de CO_2, especialmente en las primeras fases del ciclo de vida del proyecto: desde la extracción de las materias primas hasta la entrega en la obra.

El cálculo de las emisiones por categoría o tipo de elemento de construcción requiere el uso de factores de emisión de carbono de referencia que reflejen la cantidad de CO_2 generada por la producción de diferentes materiales. Tales materiales incluyen el hormigón, los ladrillos, el acero reciclado, el aluminio y otros. Estos valores suelen extraerse de fuentes acreditadas y bases de datos internacionales como UK ICE 2015 (Inventory of Carbon and Energy) y US EPA 2006 (U.S. Environmental Protection Agency) [105]. La siguiente tabla (Fig, 5.3-6) resume los factores de emisión de referencia para una serie de materiales de construcción comunes. Se proporcionan dos parámetros clave para cada material: emisiones específicas de CO_2 (en kilogramos por kilogramo de material) y factores de conversión de volumen a peso (en kilogramos por metro cúbico), que son necesarios para integrar los cálculos en el modelo de diseño y vincularlos a la agrupación de datos QTO.

Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m3 to kg
		Process Emissions (kg CO2e/ kg of product) (K1)	Process Emissions (kg CO2e/ kg of product) (K2)	Kg / m3 (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

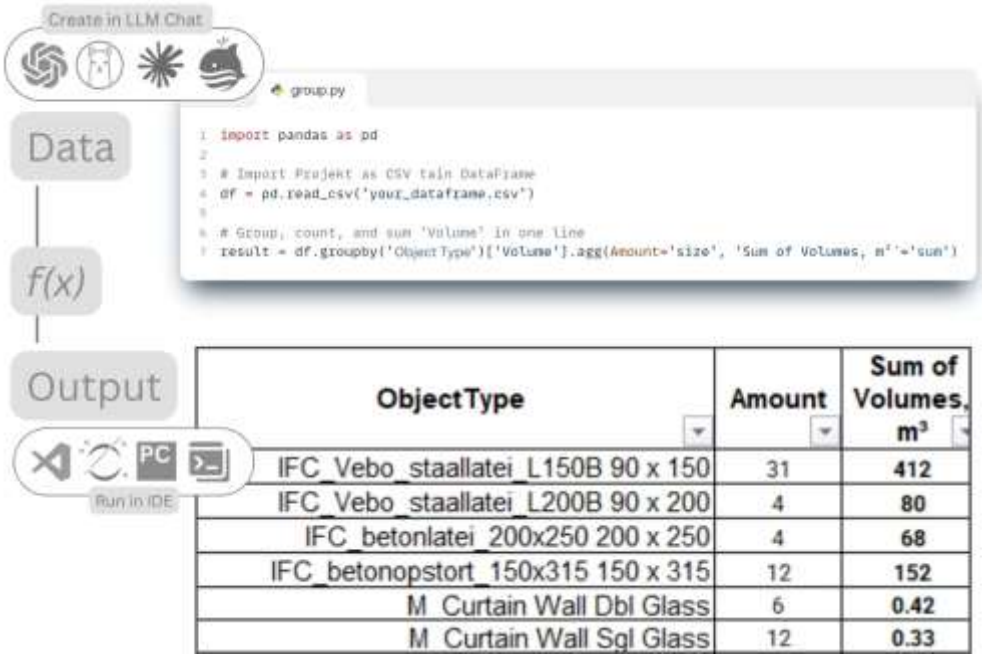
Fig. 5.3-6 Cantidad de carbono emitida durante la producción de distintos materiales de construcción, según la base de datos del ICE británico y la EPA estadounidense.

Para calcular las emisiones totales de CO_2 de un proyecto, al igual que en los cálculos 4D y 5D, es necesario determinar los volúmenes de atributos de cada grupo de instalaciones. Esto puede hacerse utilizando herramientas de análisis cuantitativo (QTO) obteniendo los volúmenes de atributos en metros cúbicos, como se explica detalladamente en la sección sobre Quantity take-off. A continuación, estos volúmenes se multiplican por los coeficientes adecuados para el atributo "Emisiones de CO_2 de proceso" de cada grupo de materiales.

- Extraigamos automáticamente la tabla de volúmenes por tipo de elemento del proyecto CAD (BIM) agrupando todos los datos del proyecto como ya se ha hecho en los capítulos anteriores. Para realizar esta tarea, vamos a referirnos a LLM.

> Por favor, agrupe la tabla DataFrame del proyecto CAD (BIM) por el parámetro de columna "Nombre del objeto" (o "Tipo") y muestre el número de elementos en cada grupo, y resuma el parámetro "Volumen" para todos los elementos del tipo. ⏎

- Respuesta de LLM:



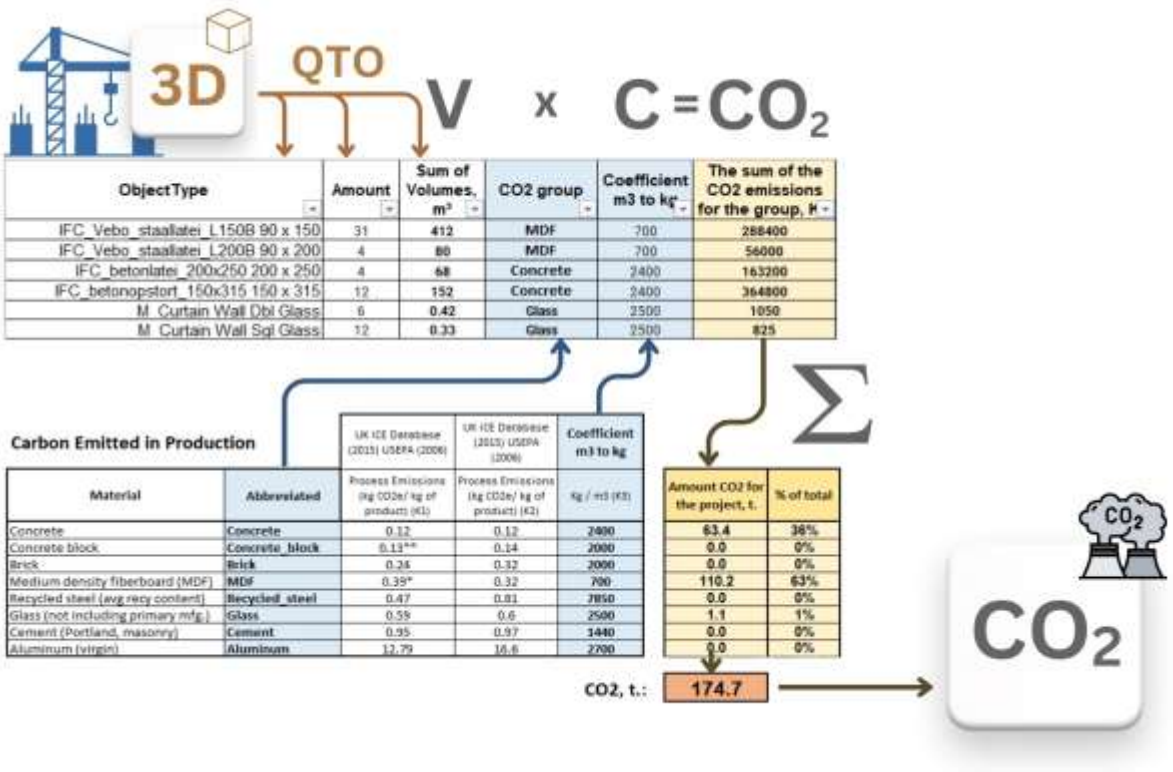
ObjectType	Amount	Sum of Volumes, m³
IFC_Vebo_staallatei_L150B 90 x 150	31	412
IFC_Vebo_staallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Fig. 5.3-7 El código generado en LLM nos ha agrupado las entidades de proyecto por tipo (ObjectType) con el atributo resumido "Volumen".

Para automatizar el cálculo de las emisiones totales de CO_2 de todo el proyecto, basta con configurar la asignación automática de datos en la tabla o vincular manualmente los tipos de elementos (Fig, 5.3-7) con los tipos de materiales correspondientes (Fig, 5.3-6) a partir de la tabla de factores de emisión. La tabla terminada con los factores de emisión y las fórmulas, así como el código para recuperar volúmenes de formatos CAD (BIM) y automatizar la determinación del CO_2 se pueden encontrar en GitHub buscando "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction". [106].

Así, la integración de datos tras agrupar los elementos QTO de la base de datos CAD permite calcular automáticamente las emisiones de dióxido de carbono (Fig. 5.3-8) de las distintas opciones de diseño. De este modo, es posible analizar el impacto de los distintos materiales en diferentes variantes y seleccionar únicamente las soluciones que cumplan los requisitos de emisiones de CO_2 cliente para obtener un certificado determinado cuando se ponga en servicio el edificio

> La estimación de las emisiones de CO_2 multiplicando los factores por los volúmenes de los elementos agrupados del proyecto es un ejemplo típico de tarea en el proceso de obtención de una calificación ESG por parte de una empresa de construcción (por ejemplo, la certificación LEED) para una instalación.



ObjectType	Amount	Sum of Volumes, m³	CO2 group	Coefficient m3 to kg*	The sum of the CO2 emissions for the group, K
IFC_Vebo_staallatei_L150B 90 x 150	31	412	MDF	700	288400
IFC_Vebo_staallatei_L200B 90 x 200	4	80	MDF	700	56000
IFC_betonlatei_200x250 200 x 250	4	68	Concrete	2400	163200
IFC_betonopstort_150x315 150 x 315	12	152	Concrete	2400	364800
M_Curtain Wall Dbl Glass	6	0.42	Glass	2500	1050
M_Curtain Wall Sgl Glass	12	0.33	Glass	2500	825

Material	Abbreviated	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K1)	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K2)	Coefficient m3 to kg kg / m3 (K3)	Amount CO2 for the project, t.	% of total
Concrete	Concrete	0.12	0.12	2400	63.4	36%
Concrete block	Concrete_block	0.13**	0.14	2000	0.0	0%
Brick	Brick	0.24	0.32	2000	0.0	0%
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700	110.2	63%
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850	0.0	0%
Glass (not including primary mfg.)	Glass	0.59	0.6	2500	1.1	1%
Cement (Portland, masonry)	Cement	0.95	0.97	1440	0.0	0%
Aluminum (virgin)	Aluminum	12.79	16.6	2700	0.0	0%

Fig. 5.3-8 La integración de grupos QTO a partir de bases de datos CAD proporciona precisión y automatización a la hora de derivar estimaciones de las emisiones finales de CO_2.

Del mismo modo, al definir los volúmenes de los grupos de elementos, podemos realizar cálculos para el control y la logística de materiales, el seguimiento y la gestión de la calidad, la modelización y el análisis energético, y un sinfín de otras tareas para obtener un nuevo estado de atributo (parámetro en la tabla) tanto para los grupos de elementos individuales como para todo el proyecto.

Si el número de estos procesos de cálculo en la empresa empieza a crecer, se plantea la cuestión de la necesidad de automatizar dichos cálculos e implantar los resultados de los cálculos en los procesos y sistemas de gestión de datos de la empresa.

Debido a la complejidad de una solución completa, las medianas y grandes empresas del sector de la construcción subcontratan dicha automatización a empresas de desarrollo de sistemas ERP (o PMIS). Las empresas de desarrollo crean un único sistema modular integral para que los grandes clientes gestionen muchas capas de información diferentes, incluidos los cálculos de materiales y recursos.

CAPÍTULO 5.4.
CONSTRUCCIÓN DE SISTEMAS ERP Y PMIS

ERP de la construcción -sistemas sobre el ejemplo de cálculos y estimaciones

Los sistemas ERP modulares integran varias capas de atributos (información) y flujos de datos en un único sistema integral, lo que permite a los gestores de proyectos gestionar recursos, finanzas, logística y otros aspectos de un proyecto de forma sincronizada dentro de una única plataforma. Un sistema ERP para la construcción actúa como el "cerebro" de los proyectos de construcción, simplificando los procesos repetitivos mediante la automatización y proporcionando transparencia y control a lo largo de todo el proceso de construcción.

> **ERP para la construcción** -systems (Enterprise Resource Planning) son soluciones informáticas integrales diseñadas para gestionar y optimizar diversos aspectos del proceso de construcción. En el núcleo de los sistemas ERP para la construcción se encuentran módulos para gestionar el cálculo de costes y la programación, lo que los convierte en una herramienta importante para una planificación eficaz de los recursos.

Los módulos ERP -systems permiten a los usuarios introducir, procesar y analizar datos de forma estructurada sobre diversos aspectos de un proyecto, que pueden incluir la contabilidad de costes de material y mano de obra, la utilización de equipos, la gestión logística, los recursos humanos, los contactos y otras actividades de construcción.

Uno de los bloques funcionales del sistema es el módulo de automatización de la lógica empresarial - BlackBox/WhiteBox, que desempeña el papel de centro de control del proceso.

BlackBox /WhiteBox permite a los especialistas que utilizan un sistema ERP gestionar de forma flexible, mediante derechos de acceso, diversos aspectos de la empresa que ya han sido preconfigurados por otros usuarios o administradores. En el contexto de los sistemas ERP, los términos *BlackBox* y *WhiteBox* se refieren a los niveles de transparencia y controlabilidad de la lógica interna del sistema:

- **BlackBox** ("caja negra") - el usuario interactúa con el sistema a través de la interfaz, sin acceso a la lógica interna de ejecución del proceso. El sistema realiza los cálculos por sí solo, basándose en reglas predefinidas ocultas al usuario final. Éste introduce los datos y obtiene el resultado sin saber qué atributos o coeficientes se han utilizado en su interior.
- **WhiteBox** ("caja blanca") - la lógica del proceso está disponible para su visualización, personalización y modificación. Los usuarios avanzados, administradores o integradores pueden definir manualmente algoritmos de tratamiento de datos, reglas de cálculo y escenarios de interacción entre las entidades del proyecto.

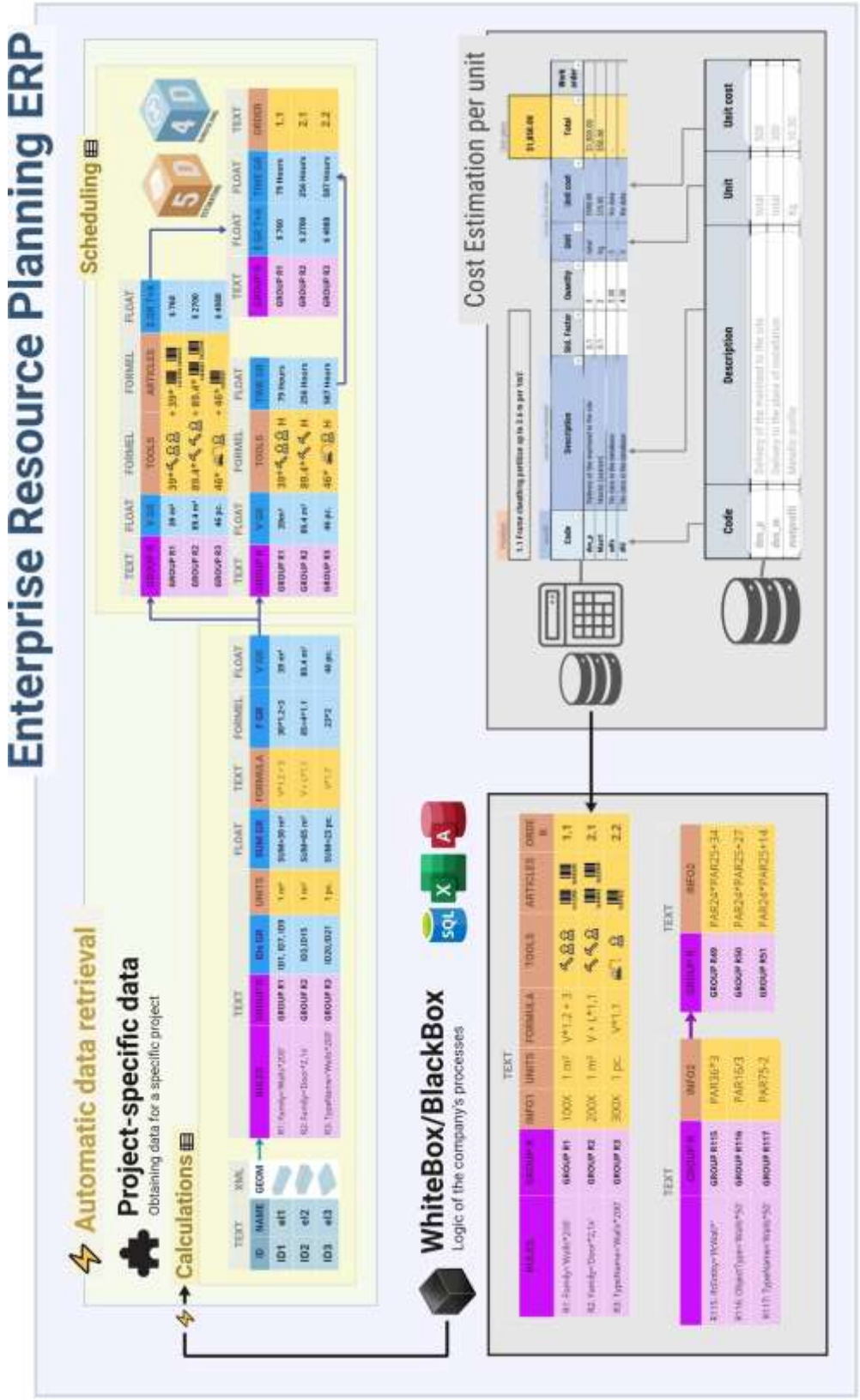

Enterprise Resource Planning ERP
Automatic data retrieval
Project-specific data
Obtaining data for a specific project
Calculations
Scheduling
WhiteBox/BlackBox
Logic of the company's processes
Cost Estimation per unit
Code
Description
Unit
Unit cost

Fig. 5.4-1 Arquitectura del sistema ERP de la construcción -obtener presupuestos y calendarios de trabajo rellenando manualmente los atributos de volumen.

Un ejemplo es cuando un usuario o administrador experimentado establece una regla: qué atributos de un presupuesto deben multiplicarse entre sí o agruparse por una determinada característica, y dónde debe registrarse el resultado final. Posteriormente, los profesionales menos formados, como los ingenieros de presupuestos, sólo tienen que cargar los nuevos datos en el ERP a través de la interfaz de usuario y obtener presupuestos, calendarios o especificaciones listos sin tener que escribir código ni comprender los detalles técnicos de la lógica.

> En los capítulos anteriores, los módulos de cálculo y lógica se trataron en el contexto de la interacción LLM. En un entorno ERP, dichos cálculos y transformaciones tienen lugar dentro de módulos ocultos tras una interfaz de botones y formularios.

En el siguiente ejemplo (Fig. 5.4-1), el administrador del sistema ERP en el módulo BlackBox /WhiteBox ha definido reglas para emparejar atributos de entidades de las estimaciones con atributos para agrupar QTO. Gracias a este módulo BlackBox/WhiteBox configurado (por el gestor o el administrador), el usuario (estimador o ingeniero), añadiendo manualmente un atributo de cantidad o volumen a través de la interfaz de usuario del ERP, recibe automáticamente las estimaciones y los programas de trabajo terminados. De este modo, los procesos de cálculo y generación de presupuestos tratados en los capítulos anteriores con la ayuda de código, dentro del ERP, se convierten en un transportador semiautomatizado.

La conexión de este proceso semiautomatizado con los atributos volumétricos de los modelos CAD (BIM) (Fig. 4.1-13), a través, por ejemplo, de la carga del proyecto CAD en el módulo ERP preconfigurado a tal efecto, convierte el flujo de datos en un mecanismo sincronizado capaz de actualizar de forma autónoma e instantánea el valor de grupos de elementos individuales o de todo el proyecto en respuesta a cualquier cambio que se produzca en el mismo durante la fase de diseño, al cargar el modelo CAD en el ERP.

Para crear un flujo de datos automatizado (Fig. 5.4-2) entre los sistemas CAD (BIM) y ERP, los procesos y requisitos básicos para los datos de las bases de datos de modelos CAD (BIM) deben definirse de forma estructurada, como ya hemos comentado en el capítulo anterior "Requisitos y garantía de calidad de los datos ". Este proceso en ERP se divide en pasos similares:

- **Creación de reglas de validación (1),** que desempeñan un papel importante para garantizar la exactitud de los datos que entran en el sistema ERP. Las reglas de validación sirven como filtros que validan las entidades y sus atributos, permitiendo que sólo los elementos que pasan los requisitos entren en el sistema. Usted puede aprender más sobre verificación y validación en el capítulo "Creando requerimientos y validando calidad de datos ".
- A continuación, tiene **lugar un proceso de verificación (2)** dentro de ERP, que confirma que todos los elementos de la entidad del proyecto con sus atributos y valores se han creado correctamente y están listos para los siguientes pasos de procesamiento.
- Si hay problemas con datos de atributos incompletos, **se genera un informe (3)** y el proyecto, junto con las instrucciones para su corrección, se envía a revisión hasta que esté listo para la siguiente iteración.
- Una vez que los datos del proyecto han sido validados y verificados, se utilizan en otro módulo

ERP **(4) para crear tablas de Extracción de Cantidades (QTO**) que crean atributos de cantidad para grupos de entidades, materiales y recursos de acuerdo con reglas generadas previamente (WhiteBox/BlackBox).

- Los datos agrupados por reglas de concordancia o QTO **se integran automáticamente en los cálculos (por ejemplo, coste y tiempo) (5).**
- En el último paso del sistema ERP, el usuario, multiplicando los atributos de alcance de la tabla QTO por los atributos de las tablas de procesos (por ejemplo, elementos estimados)**, genera automáticamente resultados de cálculo (6)** (por ejemplo, estimaciones de costes, calendarios de trabajo o emisiones de CO_2) para cada grupo de entidades y para el proyecto en su conjunto.

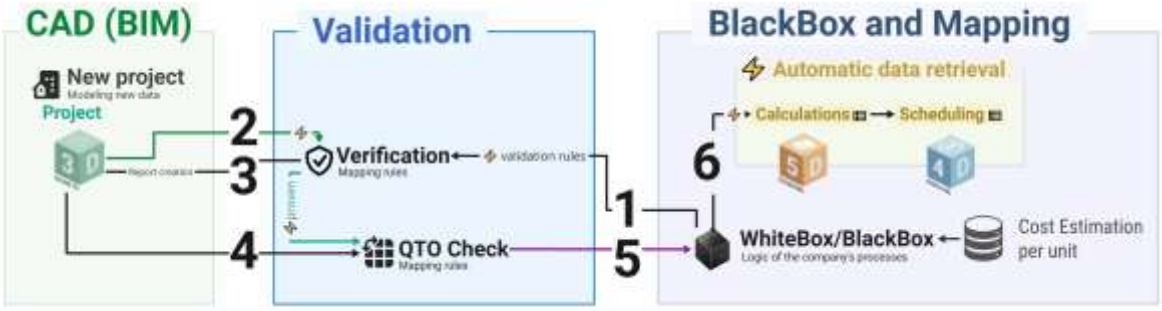


Fig. 5.4-2 Arquitectura de un sistema ERP de construcción con CAD (BIM), desde la creación de reglas de validación (1) hasta el cálculo automático de costes y calendarios de trabajo (5-6).

En un sistema ERP modular, los procesos se integran mediante un software que incluye una interfaz de usuario. Detrás de la interfaz se encuentra el back-end, donde tablas estructuradas procesan los datos realizando diversas operaciones que han sido preconfiguradas por el gestor o administrador. Como resultado, el usuario, gracias a la lógica de automatización predefinida y personalizada (en los módulos BlackBox /WhiteBox), recibe documentos preparados de forma semiautomática que responden a sus tareas

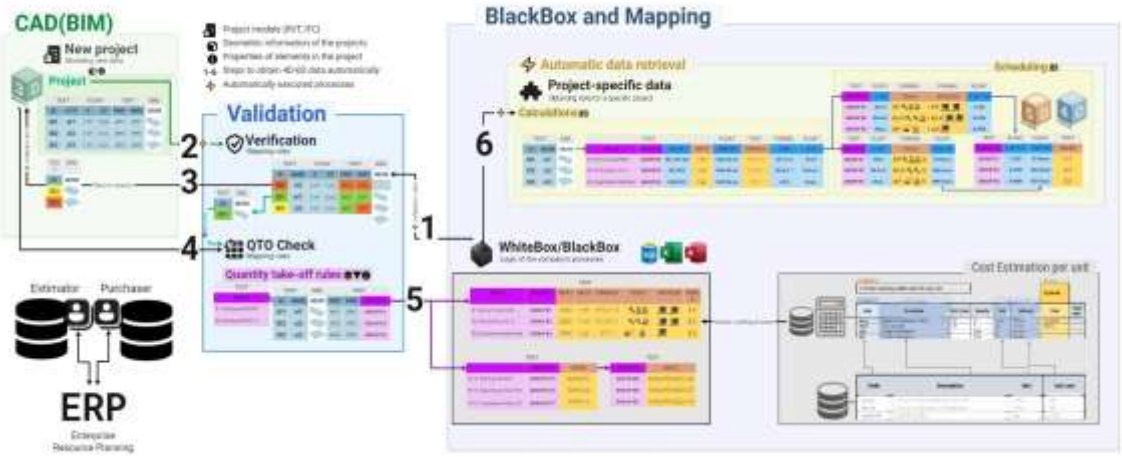


Fig. 5.4-3 ERP -sistema ayuda a los gestores y usuarios a moverse entre tablas especializadas para generar nuevos datos.

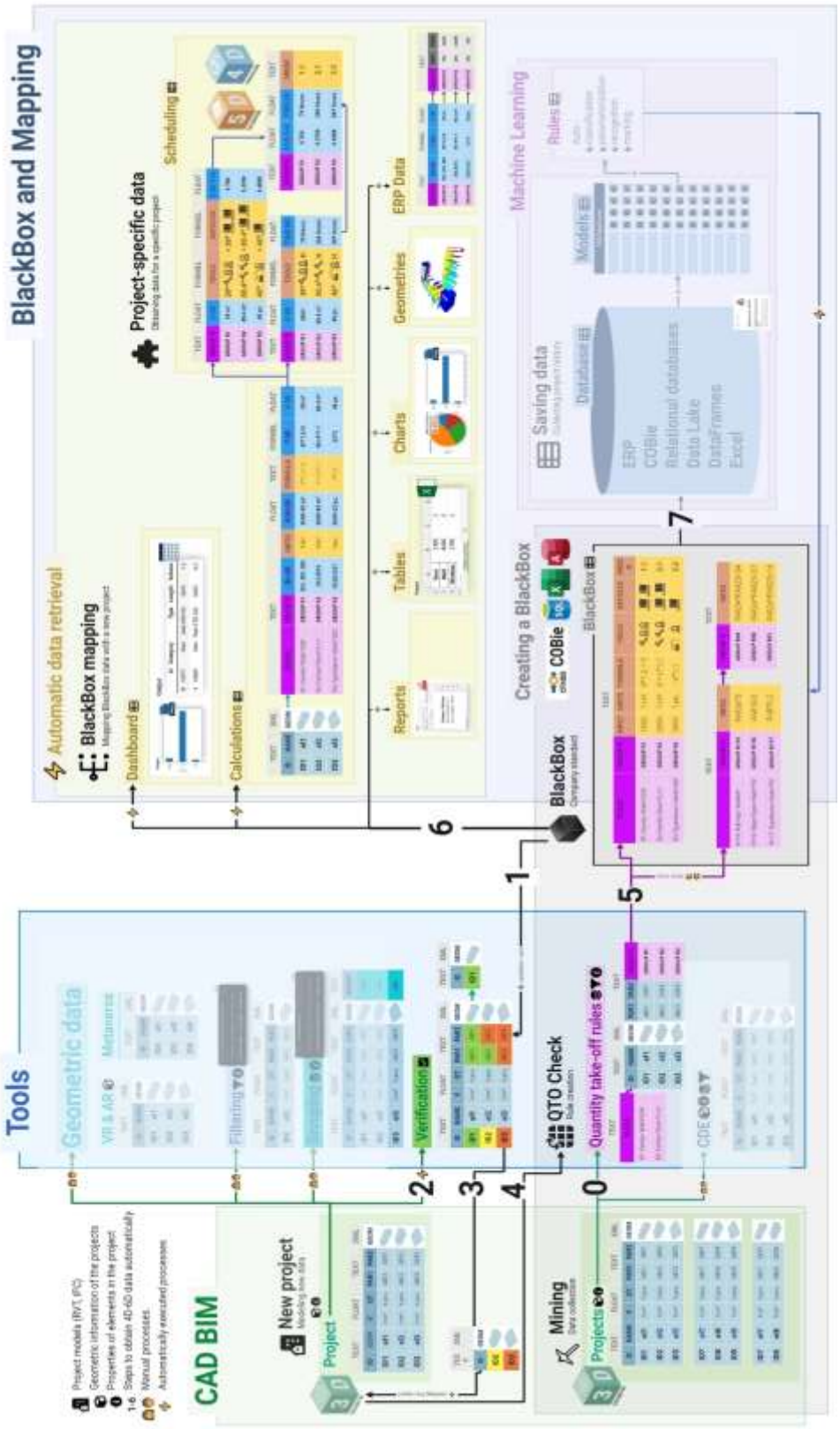


Fig. 5.4-4 ERP -sistema está integrado con herramientas analíticas y automatiza el proceso de toma de decisiones de la empresa.

Del mismo modo, los procesos de los sistemas ERP, desde el inicio hasta el cálculo final (pasos 1-6 Fig. 5.4-3) son una cadena de pasos interrelacionados que, en última instancia, aportan transparencia, eficacia y precisión a la planificación.

Los modernos sistemas ERP para la construcción incluyen no sólo módulos de cálculo de costes y plazos, sino también docenas de otros módulos preconfigurados, que suelen abarcar la gestión de documentos, el seguimiento del progreso del proyecto, la gestión de contratos, la cadena de suministro y la logística, así como la integración con otros sistemas y plataformas empresariales. Las herramientas analíticas integradas del ERP permiten a los usuarios automatizar la creación de cuadros de mando para supervisar los KPI de los proyectos (KPI - indicadores clave de rendimiento). De este modo, se consigue una gestión centralizada y coherente de todos los aspectos de un proyecto de construcción, tratando de combinar un gran número de aplicaciones y sistemas en una única plataforma.

En el futuro, los sistemas ERP -analytics se utilizarán en combinación con el aprendizaje automático para aumentar la precisión y optimizar el proceso de cálculo de los futuros atributos de los proyectos. Los datos y atributos analizados y recopilados de los sistemas ERP en Big Data (Fig. 5.4-4) constituirán en el futuro la base para crear modelos predictivos que puedan anticipar con precisión posibles retrasos, riesgos o, por ejemplo, posibles cambios en los costes de material.

Como alternativa al ERP, el sector de la construcción suele utilizar el PMIS (Project Management Information System), un sistema de gestión de proyectos diseñado para el control detallado de las tareas a nivel de proyecto de construcción individual.

PMIS: Intermediario entre ERP y la obra

A diferencia de ERP, que abarca toda la cadena de procesos empresariales de una empresa, PMIS se centra en la gestión de un proyecto concreto, el seguimiento de plazos, presupuestos, recursos y documentación.

> PMIS (Project Management Information System) es un software de gestión de proyectos de construcción diseñado para planificar, seguir, analizar e informar sobre todos los aspectos de un proyecto.

PMIS permite gestionar documentos, calendarios, presupuestos y a primera vista, PMIS puede parecer una solución duplicada de ERP, pero la diferencia clave es el nivel de gestión:

- **ERP** se centra en los procesos de negocio de la empresa en su conjunto: costes, contratos, compras, recursos humanos y gestión de recursos a nivel corporativo.
- **PMIS** se centra en la gestión de proyectos individuales, proporcionando planificación detallada, control de cambios, elaboración de informes y coordinación de participantes.

En muchos casos, son los sistemas ERP los que ya disponen de funcionalidad suficiente, y la implantación del PMIS se convierte más en una cuestión de comodidad y preferencia de la empresa. Muchos contratistas y clientes utilizan el PMIS no porque sea necesario, sino porque lo impone el proveedor o un gran cliente que quiere agregar datos en una plataforma concreta.

Cabe mencionar que en la terminología internacional para la gestión de proyectos de construcción existen

otros conceptos populares separados como PLM (Gestión del Ciclo de Vida del Producto) y EPC y EPC-M (Ingeniería, Adquisición y Gestión de la Construcción) - métodos de contratación en la industria de la construcción.

Si una empresa ya utiliza el ERP con módulos de gestión de proyectos, la introducción del PMIS puede ser un enlace innecesario que duplica funcionalidades. Sin embargo, si los procesos no están automatizados y los datos están fragmentados, el PMIS puede ser una herramienta más cómoda y fácil de mantener.

Especulación, lucro, insularidad y falta de transparencia en ERP y PMIS

A pesar de la simplicidad externa de las interfaces y los procedimientos, los sistemas ERP y PMIS para la construcción son, en la mayoría de los casos, soluciones cerradas e inflexibles. Estos sistemas suelen suministrarse como un paquete de software preconfigurado por un único proveedor, con acceso limitado a las bases de datos internas y a la lógica de los procesos.

Los proveedores de CAD-(BIM-asumen cada vez más el desarrollo y control de estos sistemas, ya que sus bases de datos contienen la información que necesitan los sistemas ERP: atributos cuantitativos y volumétricos de los elementos del proyecto. Sin embargo, en lugar de facilitar el acceso a estos datos en un formato abierto o legible por máquina, los proveedores sólo ofrecen escenarios de usuario limitados y una lógica de procesamiento cerrada, predefinida dentro de los módulos BlackBox. Esto reduce la flexibilidad del sistema e impide adaptarlo a las condiciones específicas de cada proyecto.

> La escasa transparencia de los datos sigue siendo uno de los principales retos de los procesos digitales en la construcción. La arquitectura cerrada de las bases de datos, la falta de acceso a conjuntos completos de atributos de los elementos de construcción, la concentración en *módulos de* automatización *BlackBox* y la falta de interfaces abiertas aumentan significativamente los riesgos de burocracia documental. Estas limitaciones crean cuellos de botella en el proceso de toma de decisiones, dificultan la verificación de la información y abren la puerta a la ocultación de datos o a la especulación dentro de los sistemas ERP/PMIS. Por lo general, los usuarios sólo reciben un acceso limitado -ya sea una interfaz reducida o una API parcial- sin la posibilidad de interactuar directamente con las fuentes de datos primarias. Esto es especialmente crítico cuando se trata de parámetros generados automáticamente a partir de proyectos CAD, como volúmenes, áreas y cantidades utilizadas para cálculos QTO.

Como consecuencia, en lugar de buscar la eficiencia a través de la automatización de procesos, los datos abiertos, la reducción de los costes de transacción y la creación de nuevos modelos de negocio, muchas empresas constructoras se centran en la gestión de parámetros externos - manipulando factores, factores de ajuste y métodos de cálculo que afectan a los costes de los proyectos en plataformas ERP/PMIS cerradas. Esto crea espacio para la especulación, distorsiona los costes reales de producción y reduce la confianza entre todos los participantes en el proceso de construcción

> En la construcción, el beneficio se forma como la diferencia entre los ingresos de un proyecto terminado y los costes variables, que incluyen el diseño, los materiales, la mano de obra y otros costes directos directamente relacionados con la ejecución del proyecto. Sin embargo, el factor clave que influye en el valor de estos costes no es sólo la tecnología o la logística, sino también la rapidez y exactitud de los cálculos y la calidad de las decisiones de gestión dentro de la empresa.

El problema se agrava por el hecho de que en la mayoría de las empresas de construcción los procesos de cálculo de costes siguen siendo poco transparentes no sólo para los clientes, sino también para los propios empleados, que no forman parte de los departamentos de estimación o financiero. Esta cerrazón favorece la formación dentro de la empresa de un grupo privilegiado de especialistas, portadores de "conocimientos financieros", que tienen el derecho exclusivo de editar atributos y factores de corrección en los sistemas ERP/PMIS. Estos empleados, junto con los jefes de empresa, pueden controlar realmente la lógica financiera del proyecto.

> En tales condiciones, los estimadores se convierten en "malabaristas financieros", que se equilibran entre la maximización del beneficio de la empresa y la necesidad de mantener un precio competitivo para el cliente. Al mismo tiempo, tienen que evitar manipulaciones flagrantes y groseras para no socavar la reputación de la empresa. Es en esta fase cuando se establecen coeficientes para ocultar volúmenes o costes de materiales y obras sobreestimados.

Como resultado, el principal esquema para aumentar la eficiencia y la rentabilidad de las empresas que operan en la industria de la construcción no es la automatización y la aceleración de los procesos de toma de decisiones, sino la especulación con los precios de los materiales y las obras (Fig. 5.4-5). La sobreestimación del coste de las obras y materiales se lleva a cabo una contabilidad "gris" en sistemas cerrados ERP /PMIS - inflando los porcentajes sobre los precios medios de mercado de los materiales o volúmenes de obras mediante coeficientes (Fig. 5.1-6), que se trataron en el capítulo "Recopilación de cálculos y cálculo del coste de las obras a partir de la base de recursos".

> Como resultado, el cliente recibe un cálculo que no refleja el coste real ni el alcance del trabajo, sino que es una derivada de muchos coeficientes internos ocultos. Al mismo tiempo, los subcontratistas, en un intento de ajustarse a las tarifas infravaloradas fijadas por el contratista general, se ven obligados a menudo a comprar materiales más baratos y de baja calidad, lo que empeora la calidad final de la construcción.

El proceso especulativo de buscar beneficios de la nada acaba perjudicando tanto a los clientes, que reciben datos poco fiables, como a los ejecutivos, que se ven obligados a buscar más y más modelos especulativos

En consecuencia, cuanto mayor es el proyecto, mayor es el nivel de burocracia en la gestión de datos y procesos. Cada paso y cada módulo esconden a menudo coeficientes y recargos opacos incrustados en algoritmos de cálculo y procedimientos internos. Esto no sólo dificulta la auditoría, sino que distorsiona

considerablemente la imagen financiera del proyecto. En los grandes proyectos de construcción, estas prácticas suelen multiplicar (a veces hasta por diez) el coste final, mientras que los volúmenes y costes reales quedan fuera del control efectivo del cliente (Fig. 2.1-3 Comparación de los costes previstos y reales de grandes proyectos de infraestructuras en Alemania.

> Según el informe de McKinsey & Company Imagining the Digital Future of Construction (2016), los grandes proyectos de construcción se completan de media un 20% más tarde de lo previsto y hasta un 80% por encima del presupuesto [107].

Los departamentos de estimaciones y presupuestos se están convirtiendo en el eslabón más vigilado de una empresa. El acceso a los mismos está estrictamente limitado incluso para los especialistas internos y, debido a la lógica cerrada y a las estructuras de las bases de datos, es imposible evaluar objetivamente la eficacia de las decisiones sobre los proyectos sin distorsiones. La falta de transparencia lleva a que las empresas se vean obligadas no a optimizar los procesos, sino a luchar por la supervivencia mediante una gestión "creativa" de cifras y factores (Fig. 5.3-1, Fig. 5.1-6 - por ejemplo, el parámetro "Factor de oferta").

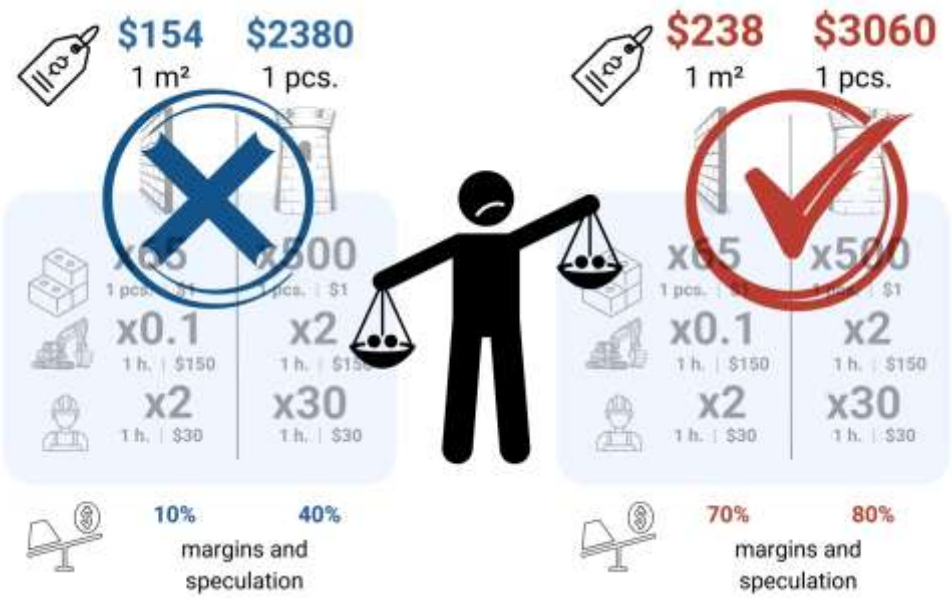


Fig. 5.4-5 Los ratios de especulación a nivel de liquidación son el principal beneficio de las empresas y el arte de hacer malabarismos entre la calidad del trabajo y la reputación.

Todo esto arroja dudas sobre el uso continuado de sistemas cerrados ERP/PMIS en la construcción. En el contexto de la transformación digital y las crecientes demandas de transparencia por parte de los clientes (Fig. 10.2-3), es poco probable que la realización de proyectos a largo plazo siga dependiendo de soluciones propietarias que limitan la flexibilidad, dificultan la integración y obstaculizan el desarrollo empresarial.

Y por muy favorable que sea para las empresas de la construcción trabajar con silos de datos y datos opacos en bases de datos cerradas - inevitablemente, el futuro de la industria de la construcción implicará una transición hacia plataformas abiertas, estructuras de datos legibles por máquinas y transparentes, y una automatización basada en la confianza. Esta transformación se impulsará desde arriba, bajo la presión de los clientes, los reguladores y la sociedad, que exigen cada vez más responsabilidad, sostenibilidad,

transparencia y viabilidad económica.

El fin de la era de los ERP cerrados /PMIS: el sector de la construcción necesita nuevos enfoques

El uso de voluminosos sistemas modulares ERP/PMIS -systems que constan de decenas de millones de líneas de código dificulta enormemente cualquier cambio en ellos. En este caso, la transición a una nueva plataforma en presencia de módulos ya preconfigurados para la empresa, decenas de miles de artículos en bases de datos de recursos (Fig. 5.1-3) y miles de cálculos ya hechos (Fig. 5.1-6) se convierte en un proceso costoso y largo. Cuanto mayor sea el código y la arquitectura heredada, mayor será el nivel de ineficacia interna, y cada nuevo proyecto no hará sino empeorar las cosas. En muchas empresas, la migración de datos y la integración de nuevas soluciones se convierten en epopeyas de varios años acompañadas de constantes retrabajos e interminable búsqueda de compromisos. El resultado suele ser la vuelta a plataformas antiguas y conocidas, a pesar de sus limitaciones.

Como se destaca en el informe alemán Black Book [108] sobre fallos sistémicos en la gestión de datos de la construcción, la fragmentación de la información y la falta de un enfoque centralizado de su gestión es una causa clave de ineficacia. Sin normalización e integración, los datos pierden su valor, convirtiéndose en un archivo más que en una herramienta de gestión.

> Una de las principales causas de la pérdida de calidad de los datos es la planificación y el control inadecuados de los proyectos de construcción, que a menudo provocan importantes aumentos de costes. La sección "Focus: The Cost Explosion" del Libro Negro analiza los factores clave que contribuyen a estas consecuencias indeseables. Entre ellos figuran los análisis inadecuados de las necesidades, la falta de estudios de viabilidad y una planificación descoordinada que genera costes adicionales que podrían haberse evitado.

> En un ecosistema informático maduro de una empresa, sustituir un sistema obsoleto es comparable a sustituir una columna de carga en un edificio ya construido. No basta con retirar el antiguo e instalar uno nuevo: es importante hacerlo de forma que el edificio permanezca estable, los techos no se derrumben y todas las comunicaciones sigan funcionando. Aquí es donde radica la dificultad: cualquier error puede tener graves consecuencias para todo el sistema de la empresa.

Sin embargo, los desarrolladores de grandes productos ERP para el sector de la construcción siguen utilizando la cantidad de código escrito como argumento a favor de su plataforma. En conferencias especializadas, todavía se oyen frases como: "Se necesitarían 150 años-hombre para recrear un sistema así", a pesar de que la mayor parte de la funcionalidad de estos sistemas esconde bases de datos y funciones bastante sencillas para trabajar con tablas, empaquetadas en una interfaz de usuario fija y especial. En la práctica, el volumen de código de "150 años-hombre" se convierte en una carga más que en una ventaja competitiva. Cuanto más código, mayor es el coste de soporte, más difícil es adaptarse a las nuevas condiciones y más alto es el umbral de entrada para nuevos desarrolladores y clientes.

> Hoy en día, muchos sistemas de construcción modular parecen engorrosas y anticuadas "construcciones Frankenstein" en las que cualquier cambio descuidado puede provocar fallos. Cada nuevo módulo aumenta la complejidad de un sistema ya sobrecargado, convirtiéndolo en un laberinto que solo unos pocos especialistas pueden entender, lo que dificulta aún más su mantenimiento y modernización.

De la complejidad también se dan cuenta los propios desarrolladores, que periódicamente se detienen a refactorizar, es decir, a revisar la arquitectura para tener en cuenta la aparición de nuevas tecnologías. Sin embargo, aunque la refactorización se haga con regularidad, la complejidad crece inevitablemente. Los arquitectos de este tipo de sistemas se acostumbran a la creciente complejidad, pero para los nuevos usuarios y especialistas se convierte en una barrera infranqueable. Como resultado, toda la experiencia se concentra en manos de unos pocos desarrolladores, y el sistema deja de ser escalable. A corto plazo, estos expertos son útiles, pero a largo plazo se convierten en parte del problema.

> *Las organizaciones seguirán integrando datos "pequeños" con sus homólogos de big data, y es absurdo que alguien crea que una sola aplicación -por cara o robusta que sea- puede encargarse de todo [109].*
>
> - Phil Simon, presentador del podcast Conversaciones sobre colaboración

Surge una pregunta legítima: ¿necesitamos realmente sistemas tan engorrosos y cerrados para calcular el coste y el calendario del trabajo en forma de tablas, si otras industrias hace tiempo que pueden hacer frente a tareas similares utilizando herramientas analíticas con datos abiertos y una lógica transparente?

Por el momento, las plataformas modulares cerradas siguen siendo muy demandadas en el sector de la construcción, debido sobre todo a las particularidades de la contabilidad de costes (Fig. 5.1-7). Estos sistemas suelen utilizarse para ejecutar esquemas grises u opacos, que permiten ocultar al cliente los costes reales. Sin embargo, a medida que la industria madure digitalmente, principalmente los clientes, y se adentre en la llamada "Era Uberizada", los intermediarios, es decir, las empresas de construcción con sus ERP, perderán importancia en los cálculos de tiempo y costes. Esto cambiará para siempre la faz del sector de la construcción. Más información en la última parte del libro y en el capítulo "Construcción 5.0: cómo ganar dinero cuando ya no puedes esconderte".

> Los miles de soluciones heredadas acumuladas en los últimos 30 años, con miles de años-hombre invertidos en su desarrollo, empezarán a desaparecer rápidamente. El paso a una gestión de datos abierta, transparente y flexible es inevitable. La única cuestión es qué empresas serán capaces de adaptarse a estos cambios y cuáles seguirán siendo rehenes del viejo modelo.

Una situación similar se observa en el ámbito de las herramientas CAD (BIM -), cuyos datos rellenan hoy los parámetros volumétricos de las entidades de diseño en los sistemas ERP/PMIS -. Inicialmente, la idea de BIM (desarrollada ya en 2002 [110] se basaba en el concepto de una única base de datos integrada, pero

en la práctica hoy en día trabajar con BIM requiere todo un conjunto de software especializado y formatos. Lo que se suponía que iba a simplificar la gestión del diseño y la construcción se ha convertido en otra capa de soluciones propietarias que complican la integración y reducen la flexibilidad empresarial.

Próximos pasos: uso eficiente de los datos del proyecto

En esta parte, hemos mostrado cómo los datos estructurados se convierten en la base de cálculos precisos de costes y plazos para proyectos de construcción. La automatización de los procesos de QTO, programación y estimación reduce los costes de mano de obra y mejora notablemente la precisión de los resultados.

Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar los enfoques comentados a sus tareas diarias. Estos enfoques son universales: son útiles tanto para la transformación digital de una empresa como para el trabajo diario de los profesionales implicados en los cálculos:

- Automatice los cálculos rutinarios
  - Intente encontrar cálculos de costes estándar que pueda utilizar en su trabajo.
  - Analice qué métodos se utilizan en su país para calcular el coste de las obras o procesos de construcción (Fig, 5.1-7).
  - Si trabaja con un sistema CAD - aprenda las funciones de extracción automática de especificaciones y datos QTO en su software CAD (BIM-).
  - Utilice el LLM para escribir código de borrador para automatizar los cálculos
- Desarrolle sus propias herramientas para QTO
  - Crear secuencias de comandos o tablas para automatizar el recuento de volúmenes
  - Normalizar categorías y grupos de elementos para un enfoque coherente de la evaluación.
  - Documentar la metodología de cálculo para garantizar la reproducibilidad de los resultados en nuevos proyectos.
- Integrar diferentes aspectos del proyecto en su trabajo
  - Si trabaja con sistemas modulares, intente visualizar sus procesos no sólo en forma de diagramas o gráficos, sino también a nivel de datos, especialmente en forma de tablas.
  - Dominar la fusión automática de datos extraídos de bases de datos CAD con cálculos - con código Python utilizando agrupación, filtrado y agregación.
  - Crear visualizaciones claras de grupos QTO para presentar información compleja a colegas y clientes.

Estos pasos ayudarán a construir un sistema de cálculo sostenible basado en la automatización y la normalización de los datos. Este planteamiento mejorará la precisión y reducirá la rutina de los problemas de cálculo cotidianos.

Los siguientes capítulos se centran en los aspectos técnicos de los productos CAD - (BIM-) y en las razones por las que las bases de datos CAD siguen siendo difíciles de integrar en los procesos de negocio de las empresas. Si ahora no te interesa la historia de la implantación de BIM en la construcción, la evolución de las herramientas CAD y los aspectos técnicos del trabajo con estas tecnologías, puedes pasar directamente a la séptima parte del libro "Toma de decisiones basada en datos".

MÁXIMA COMODIDAD CON LA VERSIÓN IMPRESA

Tiene en sus manos una versión digital gratuita de **Data-Driven Construction**. Para trabajar con mayor comodidad y acceder rápidamente a los materiales, le recomendamos que preste atención a la **edición impresa**:

- **Siempre a mano**: el libro en formato impreso se convertirá en una herramienta de trabajo fiable, que le permitirá encontrar y utilizar rápidamente las visualizaciones y diagramas adecuados en cualquier situación de trabajo.
- **Ilustraciones de alta calidad**: todas las imágenes y gráficos de la edición impresa se presentan con la máxima calidad.
- **Acceso rápido a la información**: navegación sencilla, posibilidad de tomar notas, marcar favoritos y trabajar con el libro en cualquier lugar.

Al adquirir la versión impresa completa del libro, obtendrá una cómoda herramienta para trabajar con la información de forma cómoda y eficaz: la posibilidad de utilizar rápidamente materiales visuales en las tareas cotidianas, encontrar rápidamente los esquemas necesarios y tomar notas. Además, su compra apoya la difusión del conocimiento abierto.

Solicite la versión impresa del libro en: datadrivenconstruction.io/books

VI PARTE

CAD Y BIM: MARKETING, REALIDAD Y FUTURO DE LOS DATOS DE DISEÑO EN LA CONSTRUCCIÓN

La sexta parte del libro presenta un análisis crítico de la evolución de las tecnologías CAD y BIM y su impacto en los procesos de gestión de datos en la construcción. Traza la transformación histórica del concepto BIM desde la idea original de una base de datos integrada hasta las actuales construcciones de marketing promovidas por los vendedores de software. Se evalúan las repercusiones de los formatos patentados y los sistemas cerrados en la eficacia de la gestión de datos de proyectos y el rendimiento general del sector de la construcción. Se analizan en detalle los problemas de compatibilidad de los distintos sistemas CAD y las dificultades de su integración con los procesos empresariales de las empresas constructoras. Se debaten las tendencias actuales hacia formatos de datos abiertos simplificados, como USD, y su posible impacto en la industria. Se presentan enfoques alternativos para extraer información de sistemas cerrados, incluidas técnicas de ingeniería inversa. Se analizan las perspectivas de aplicación de la inteligencia artificial y el aprendizaje automático para automatizar los procesos de diseño y análisis de datos en la construcción. Se formulan previsiones para el desarrollo de tecnologías de diseño orientadas a las necesidades reales de los usuarios y no a los intereses de los vendedores de software.

CAPÍTULO 6.1.
APARICIÓN DE LOS CONCEPTOS BIM- EN EL SECTOR DE LA CONSTRUCCIÓN

Originalmente, esta sexta parte, dedicada al CAD (BIM), no se incluyó en la primera versión del libro. Los temas de formatos propietarios, núcleos geométricos y sistemas cerrados son excesivamente técnicos, sobrecargados de detalles y aparentemente inútiles para quienes sólo quieren entender cómo trabajar con datos. Sin embargo, los comentarios y las peticiones para añadir aclaraciones a la primera versión del libro demostraron que, sin comprender las complejidades del funcionamiento interno de los sistemas CAD, los núcleos geométricos, la variedad de formatos y los esquemas de almacenamiento incompatibles para los mismos datos, es imposible entender realmente por qué los conceptos promovidos por los proveedores a menudo dificultan el trabajo con la información y obstaculizan la transición al diseño paramétrico abierto. Por eso esta parte ha ocupado un lugar propio en la estructura del libro. Si el CAD (BIM) no es una prioridad para usted, puede pasar directamente a la siguiente parte: "PARTE VII: Toma de decisiones basada en datos, análisis, automatización y aprendizaje automático".

Historia de la aparición de BIM y BIM abierto como conceptos de marketing de los proveedores de CAD-

Con la llegada de los datos digitales en la década de 1990, la informática se introdujo no sólo en los procesos empresariales, sino también en los de diseño, dando lugar a conceptos como CAD (sistemas de diseño asistido por ordenador) y, más tarde, BIM (modelado de información para la construcción)

Sin embargo, como cualquier innovación, no son el punto final del desarrollo. Conceptos como BIM se han convertido en un hito importante en la historia del sector de la construcción, pero tarde o temprano pueden dar paso a mejores herramientas y enfoques que respondan mejor a los retos del futuro.

> Abrumado por la influencia de los proveedores de CAD y confundido por las complejidades de su propia aplicación, el concepto de BIM, que apareció en 2002, puede que no llegue a cumplir treinta años, como una estrella de rock que brilló con fuerza pero se desvaneció rápidamente. La razón es sencilla: las exigencias de los científicos de datos cambian más rápido de lo que los proveedores de CAD pueden adaptarse a ellas.

Ante la falta de datos de calidad, los profesionales actuales del sector de la construcción exigen interoperabilidad entre plataformas y acceso a datos abiertos de proyectos CAD- para simplificar su análisis y procesamiento. La complejidad de los datos CAD y su confuso procesamiento repercuten negativamente en todos los implicados en el proceso de construcción: diseñadores, jefes de proyecto, trabajadores de la construcción a pie de obra y, en última instancia, el cliente.

En lugar de un conjunto de datos completo para su explotación, el cliente y el inversor reciben hoy contenedores en formatos CAD- que requieren complejos núcleos geométricos, comprensión de los esquemas de datos, documentación API actualizada anualmente y software CAD especializado (BIM) para trabajar

con los datos. Al mismo tiempo, gran parte de los datos de diseño siguen sin utilizarse.

En el mundo actual del diseño y la construcción, la complejidad del acceso a los datos CAD conduce a una sobreingeniería de la gestión de proyectos. Las medianas y grandes empresas que trabajan con datos CAD o desarrollan soluciones BIM se ven obligadas a mantener estrechas relaciones con los proveedores de soluciones CAD para acceder a los datos a través de API, o bien a eludir las restricciones de los proveedores CAD utilizando costosos conversores SDK para realizar ingeniería inversa, con el fin de obtener datos abiertos [75].

> El enfoque de los datos patentados está anticuado y ya no satisface las exigencias del entorno digital actual. El futuro dividirá a las empresas en dos tipos: las que utilicen los datos abiertos con eficacia, y las que abandonen el mercado.

El concepto de BIM (Building Information Modeling), apareció en la industria de la construcción con la publicación de uno de los principales proveedores de CAD - Whitepaper BIM [54] en 2002 y, complementado por el concepto de ingeniería mecánica BOM (Bills of Materials), se originó a partir del enfoque paramétrico para la creación y el procesamiento de los datos del proyecto (Fig. 6.1-1). El enfoque paramétrico para la creación y el procesamiento de datos de proyecto fue uno de los primeros en implementarse en el sistema Pro-E para el diseño de ingeniería mecánica (MCAD). Este sistema se convirtió en un prototipo [111] para muchas soluciones modernas de CAD, incluidas las que se utilizan actualmente en el sector de la construcción.

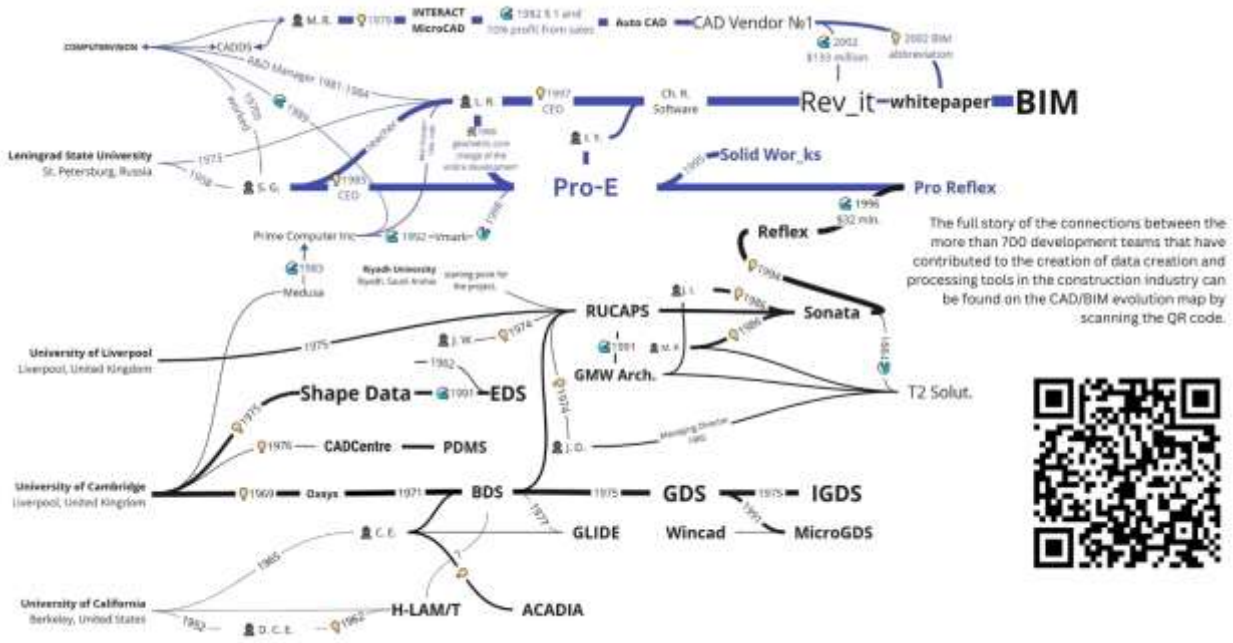


Fig. 6.1-1 Mapa de la historia del concepto BIM y conceptos similares.

Los periodistas y consultores de AEC, que promocionaron las herramientas CAD -vendors hasta principios de la década de 2000, centraron su atención en el Whitepaper BIM a partir de 2002. Fueron el Whitepaper BIM 2002-2004 y los artículos publicados en 2002, 2003, 2005 y 2007 los que desempeñaron un papel clave en la popularización del concepto BIM en el sector de la construcción [112].

> *Building Information Modelling es una estrategia..... [CAD] para aplicar las tecnologías de la información al sector de la construcción.*
>
> - BIM Whitepaper, 2002 [60]

A mediados de la década de 2000, los "investigadores" empezaron a relacionar el concepto BIM- publicado por el proveedor CAD- en 2002 con trabajos científicos anteriores, como el BDS de Charles Eastman, que se convirtió en la base de sistemas como GLIDE, GBM, BPM, RUCAPS. En su obra pionera Building Description System (1974), Charles Eastman sentó las bases teóricas del modelado moderno de la información. El término "base de datos " aparece 43 veces en su obra (Fig. 6.1-2), más que ningún otro, salvo la palabra "edificio".

> La idea clave de Eastman era que toda la información sobre un edificio -desde la geometría hasta las propiedades de los elementos y sus interrelaciones- se almacenara en una única base de datos estructurada. A partir de esta base de datos pueden generarse y analizarse automáticamente planos, especificaciones, cálculos y cumplimiento de códigos. Eastman criticó explícitamente los planos como método de comunicación anticuado y redundante, señalando la duplicación de información, los problemas de actualización y la necesidad de actualizaciones manuales cuando se introducen cambios. En su lugar, propuso un único modelo digital en una base de datos donde cualquier cambio se hace una vez y se refleja automáticamente en todas las vistas.

Cabe destacar que Eastman no situó la visualización en el centro de su concepto. La información era fundamental en su sistema: parámetros, relaciones, atributos, capacidades de análisis y automatización. En su opinión, los dibujos eran sólo una de las formas de mostrar los datos de la base de datos, no la fuente principal de información sobre el diseño.

En el primer Libro Blanco sobre BIM del principal proveedor de CAD, la frase "base de datos " se utilizaba tan a menudo como en el BDS de Charles Eastman: 23 veces [60] a lo largo de siete páginas y era una de las palabras más populares del documento después de "Edificio", "Información", "Modelado" y "Diseño". Sin embargo, en 2003, el término "base de datos" sólo aparece dos veces en documentos similares [61], y a finales de la década de 2000 el tema de las bases de datos prácticamente había desaparecido del debate sobre los datos de diseño. Como resultado, el concepto de "una única base de datos integrada para el análisis visual y cuantitativo" nunca llegó a materializarse del todo.

Así, la industria de la construcción ha pasado del concepto progresista de SDE de Charles Eastman con su énfasis en las bases de datos y las ideas de Samuel Geisberg sobre la actualización automática de los datos de diseño a partir de bases de datos en el producto de ingeniería mecánica Pro-E (el predecesor de las populares soluciones CAD -utilizadas en la construcción hoy en día) al actual BIM comercializado, en el que apenas se menciona la gestión de datos a través de bases de datos, a pesar de que éste era el concepto en el que se basaba la teoría original.

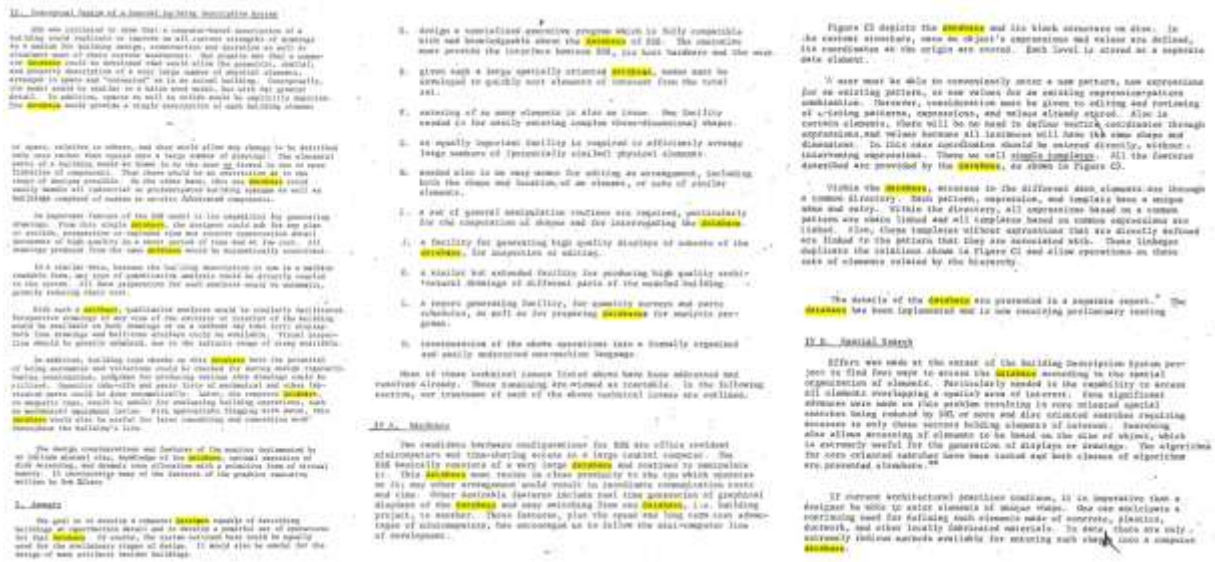


Fig. 6.1-2 En el concepto de SDE, descrito por Charles Eastman en 1974, la frase "Base de datos" (resaltada en amarillo) se utilizó 43 veces.

BDS y conceptos similares hasta la década de 2000 se desarrollaron como base de datos digital de edificios más que como herramienta de visualización. En 2002, BIM se convirtió en una herramienta de diseño en la que la base de datos pasó a un segundo plano. Qué hemos perdido en la transición de BDS y conceptos similares en la década de 1990 a BIM a mediados de la década de 2010:

- Bases de datos abiertas: BDS y otros conceptos similares hacían hincapié en la analítica, BIM hacía hincapié en el diseño.
- Flexibilidad para trabajar con datos: BDS hacía hincapié en el análisis de datos, BIM en que los procesos se basaran en datos oscuros.
- Transparencia: el SDE pretendía ser una base de datos integrada abierta, mientras que los proveedores de CAD en BIM han hecho sus bases de datos completamente cerradas y llevan 20 años luchando sin éxito contra las herramientas de ingeniería inversa que abren formatos propietarios.

En los últimos 30 años, los diseñadores nunca han tenido acceso a una "base de datos integrada" y, tras veinte años de euforia comercial en torno a las herramientas BIM, el sector de la construcción empieza a darse cuenta de las consecuencias de esta moda.

La realidad de BIM: en lugar de bases de datos integradas, sistemas modulares cerrados

En lugar de centrarse en los datos, estructurarlos e integrarlos en procesos unificados, los usuarios de los sistemas CAD - (BIM-) se ven obligados a trabajar con un conjunto fragmentado de soluciones patentadas, cada una de las cuales dicta sus propias reglas del juego:

- **La base de datos unificada**, comentada en el primer Libro Blanco sobre BIM**, sigue siendo un mito.** A pesar de las grandes afirmaciones, el acceso a los datos sigue siendo limitado y distribuido entre sistemas cerrados.
- **BIM -los modelos se han convertido en un ecosistema cerrado** más que en una herramienta. En

lugar de un intercambio de información transparente, los usuarios se ven obligados a pagar suscripciones y utilizar API patentadas.

- **Los datos pertenecen a los proveedores, no a los usuarios.** La información de los proyectos está bloqueada en formatos propietarios o servicios en la nube, en lugar de estar disponible en formatos abiertos e independientes.

Los ingenieros de diseño y los jefes de proyecto no suelen tener acceso a los sistemas de bases de datos CAD, ni al formato en el que se almacenan los datos de sus propios proyectos. Esto hace imposible verificar rápidamente la información o formular requisitos de estructura y calidad de los datos (Fig. 6.1-3). El acceso a estos datos requiere todo un conjunto de programas especializados conectados a través de APIs y plug-ins, lo que conduce a una excesiva burocratización de los procesos en la industria de la construcción. Mientras tanto, estos datos son utilizados simultáneamente por decenas de sistemas de información y cientos de especialistas.

> *Tenemos que ser capaces de gestionar todos estos datos [CAD (BIM)] almacenarlos digitalmente y vender software de gestión del ciclo de vida y los procesos, porque **por cada ingeniero** [diseñador] que crea algo [en software CAD], **hay diez personas** que trabajan con esos datos" [41].*
>
> - CEO de CAD - el proveedor que creó el concepto BIM, 2005.

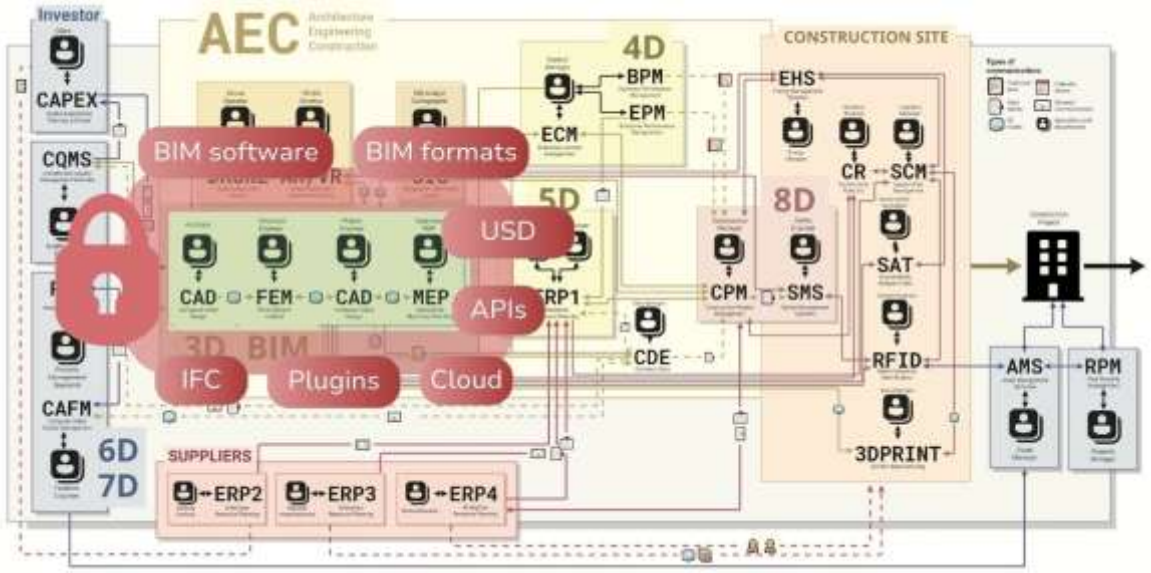


Fig. 6.1-3 Las bases de datos CAD- (BIM-) siguen siendo uno de los últimos sistemas cerrados para los departamentos informáticos y los gestores de datos en el ecosistema empresarial de la construcción.

> Cuando se hace evidente que BIM es más un medio de comercializar bases de datos que una herramienta de gestión de bases de datos en toda regla, surge una pregunta lógica: ¿cómo podemos recuperar el control de los datos? La respuesta es utilizar estructuras de datos abiertas en las que el usuario, y no el vendedor del software, se convierte en propietario de la información.

Los usuarios y desarrolladores de soluciones del sector de la construcción, al igual que sus homólogos de otras industrias, se alejarán inevitablemente de la vaga terminología de los proveedores de software que ha dominado los últimos 30 años, centrándose en los aspectos clave de la digitalización: "datos" y "procesos".

A finales de los años 80, el área clave del desarrollo digital en la construcción se presentaba como una cuestión de acceso a los datos y gestión de la información del proyecto. Con el tiempo, sin embargo, el enfoque ha cambiado. En lugar de desarrollar enfoques transparentes y accesibles para trabajar con datos, el formato IFC y el concepto de BIM abierto se promovieron activamente como intentos de desviar la atención de los especialistas de los temas de gestión de bases de datos de proyectos.

La aparición del formato abierto IFC en el sector de la construcción

El llamado formato abierto IFC (Industry Foundation Classes) se posiciona como un estándar para garantizar la interoperabilidad entre distintos sistemas CAD (BIM -). Su desarrollo se llevó a cabo en el marco de organizaciones creadas y controladas por los principales proveedores de CAD. Basándose en el formato IFC, dos empresas de CAD- desarrollaron en 2012 el concepto de marketing OPEN BIM[63].

> **IFC** (Industry Foundation Classes) es un estándar abierto para el intercambio de datos en el sector de la construcción, diseñado para garantizar la interoperabilidad entre distintos sistemas CAD - (BIM-).

Open BIM - concepto implica trabajar con información de bases de datos CAD e intercambiar información entre sistemas mediante un formato abierto de intercambio de datos CAD - IFC.

> *El Programa Open BIM es una campaña de marketing iniciada por... [1 proveedor de CAD],... [2 proveedor de CAD] y otras empresas para fomentar y facilitar la promoción coordinada global del concepto OPEN BIM en toda la industria AEC, con una comunicación coherente y una marca común disponible para los participantes en el programa.*
>
> - Del sitio web del proveedor de CAD, Programa OPEN BIM, 2012 [113].

IFC fue adaptado por la Universidad Técnica de Múnich a partir del formato de ingeniería mecánica STEP a finales de la década de 1980, y posteriormente fue registrado por una importante empresa de diseño y un importante proveedor de CAD- para formar la IAI (Industry Alliance for Interoperability) en 1994 [114] (Fig.

6.1-4). El formato IFC se desarrolló para garantizar la interoperabilidad entre diferentes sistemas CAD y se basó en los principios establecidos en el formato de ingeniería mecánica STEP, que, a su vez, surgió del formato IGES creado en 1979 por un grupo de usuarios y proveedores de CAD con el apoyo del NIST (The National Institute of Standards and Technology) y el Departamento de Defensa de EE.UU. [115].

Sin embargo, la compleja estructura de IFC, su estrecha dependencia del núcleo geométrico, así como las diferencias en la implementación del formato por parte de distintas soluciones de software han provocado numerosos problemas en su aplicación práctica. Dificultades similares -pérdida de detalle, limitación de la precisión y necesidad de utilizar formatos intermedios- encontraron anteriormente los especialistas en ingeniería mecánica al trabajar con los formatos IGES y STEP, de los que surgió IFC.

Fig. 6.1-4 Mapa de vínculos entre equipos de desarrollo y productos CAD (BIM) [116].

> En 2000, el mismo proveedor de CAD -que registró el formato IFC y creó la organización IAI (más tarde bS), publica el libro blanco "Diseño y fabricación integrados: ventajas y justificación" [65]. El documento destacaba la importancia de mantener la granularidad total de los datos en el intercambio entre programas del mismo sistema, sin utilizar formatos neutros como IGES, STEP [idéntico a IFC]. En su lugar, se proponía que las aplicaciones tuvieran acceso directo a la base de datos CAD subyacente para evitar la pérdida de precisión de la información.

En 2002, el mismo vendedor de CAD compra el producto paramétrico BOM (Fig. 3.1-18, más detalles en de la tercera parte) y sobre su base forma el concepto BIM. Como resultado, en el intercambio de datos de

proyectos de construcción sólo se utilizan formatos CAD cerrados o el formato IFC (STEP), cuyas limitaciones fueron expuestas por escrito por el propio vendedor de CAD en 2000, que introdujo este formato en el sector de la construcción.

En el mapa "La evolución del CAD (BIM)" [116] se presenta una historia detallada de la interacción de más de 700 equipos de desarrollo implicados en la creación de herramientas para crear y procesar datos de construcción. [116].

El formato abierto IFC consta de una descripción geométrica de los elementos de diseño y una descripción de la metainformación. Se utilizan varios métodos para representar la geometría en formato IFC, como CSG y Swept Solids: sin embargo, la representación paramétrica BREP se ha convertido en el principal estándar para transferir la geometría de los elementos en formato IFC, ya que este formato es compatible cuando se exporta desde programas CAD- (BIM-) y permite la edición potencial de los elementos cuando se vuelve a importar IFC a programas CAD.

IFC problema de formato en función del núcleo geométrico

En la mayoría de los casos, cuando la geometría en IFC se define de forma paramétrica (BREP), resulta imposible visualizar o recuperar propiedades geométricas como el volumen o el área de las entidades del proyecto sólo con un archivo IFC, ya que para trabajar con la geometría y visualizarla en este caso se necesita un núcleo de geometría (Fig, 6.1-5), del que se carece inicialmente.

> **El núcleo de geometría** es un componente de software que proporciona algoritmos básicos para crear, editar y analizar objetos geométricos en CAD (CAD), BIM y otras aplicaciones de ingeniería. Se encarga de construir geometría 2D y 3D y de realizar operaciones con ella, como: operaciones booleanas, suavizado, intersecciones, transformaciones y visualización.

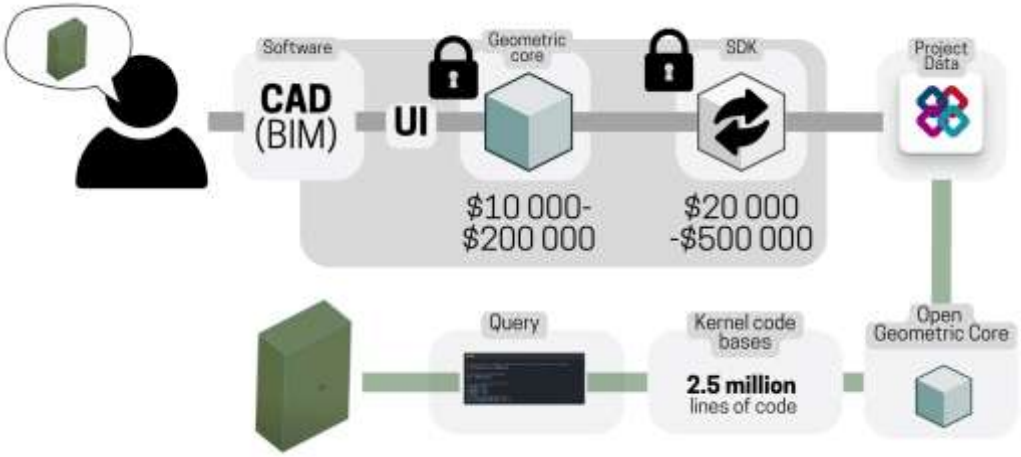


Fig. 6.1-5 La creación de geometría a través de software CAD- pasa hoy en día por kernels de geometría propietarios y SDKs, que a menudo no son propiedad de los vendedores de CAD.

Cada programa CAD y cualquier programa que trabaje con formatos paramétricos o IFC tienen su propio kernel geométrico o adquirido. Y si con elementos primitivos en formato IFC -BREP no puede haber problemas y en programas con diferentes kernels geométricos estos elementos se pueden visualizar de forma

similar, pero además de problemas con diferentes motores de kernels geométricos, hay bastantes elementos que tienen sus propias peculiaridades para su correcta visualización. Este problema se trata en detalle en el estudio internacional " A reference study of IFC software support" publicado 2019 [117].

> *Los mismos conjuntos de datos normalizados producen resultados contradictoriosencontrándose pocos patrones comunes, y se han detectado graves problemas en el apoyo a la norma [IFC], probablemente debido a la gran complejidad del modelo de datos normalizado. Las propias normas tienen parte de culpa, ya que a menudo dejan algunos detalles sin definir, con altos grados de libertad y varias interpretaciones posibles. Permiten una gran complejidad en la organización y el almacenamiento de objetos, lo que no favorece una comprensión universal eficaz, unas implementaciones únicas y un modelado de datos coherente [117].*
>
> - Estudio de referencia sobre el soporte de software IFC, 2021

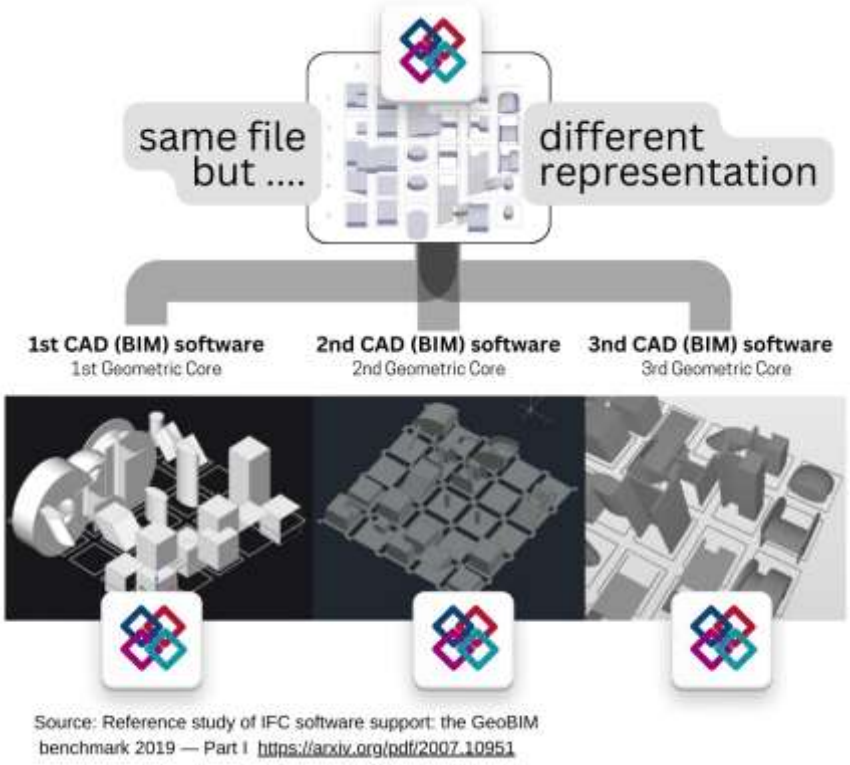


Fig. 6.1-6 Diferentes kernels geométricos dan diferentes representaciones de la misma geometría descrita paramétricamente (basado en [117]).

La comprensión correcta de "ciertas disposiciones" está al alcance de los miembros de pago de las organizaciones especiales que desarrollan IFC. En consecuencia, quien quiera acceder a conocimientos importantes sobre determinadas características de IFC intentará cooperar con grandes proveedores de CAD-, o llegar a una consideración cualitativa de las características mediante su propia investigación

> *Te tropiezas con una pregunta sobre la importación y exportación de datos a través del formato IFC y preguntas a tus compañeros: "¿Por qué aparece en el archivo IFC información sobre la transferencia paramétrica de locales? La especificación abierta no dice nada al respecto". Respuesta de los proveedores europeos "más informados": "Sí, no se dice, pero está permitido".*
>
> - De la entrevista al desarrollador de CAD 2021 [118]

El IFC describe la geometría mediante primitivas paramétricas, pero no contiene un núcleo integrado: su función la desempeña el programa CAD, que compila la geometría a través del núcleo de geometría. El núcleo geométrico realiza los cálculos matemáticos y define las intersecciones, mientras que el IFC sólo proporciona los datos para su interpretación. Si el IFC contiene caras incorrectas, distintos programas con distintos kernels de geometría pueden ignorarlas o producir errores, dependiendo del kernel.

> En consecuencia, para trabajar con el formato IFC es necesario responder a la pregunta principal, para la que es difícil encontrar una respuesta inequívoca: ¿qué herramienta, con qué núcleo geométrico debe utilizarse para obtener la calidad de datos que el proyecto tenía originalmente en el programa CAD del que se obtuvo IFC?

Los problemas de calidad de los datos y la complejidad del formato IFC no permiten el uso directo de los datos de diseño para la automatización de procesos, el análisis y el procesamiento de datos, lo que a menudo lleva a los desarrolladores a la inevitable necesidad de utilizar soluciones CAD cerradas -soluciones con acceso "de calidad" a los datos[63], sobre lo que escribió el propio vendedor que registró IFC en 1994 [65].

Todas las peculiaridades del mapeado y la generación de parámetros IFC en el núcleo geométrico sólo pueden ser realizadas por grandes equipos de desarrolladores con experiencia en el trabajo con núcleos geométricos. Por lo tanto, la práctica actual de las peculiaridades y complejidad del formato IFC beneficia principalmente a los vendedores de CAD- y tiene mucho en común con la estrategia de los grandes vendedores de software "adoptar, extender, destruir", cuando la creciente complejidad del estándar en realidad crea barreras para los pequeños actores del mercado [94].

La estrategia de los grandes vendedores en este sentido puede consistir en adaptar los estándares abiertos, añadir extensiones y funciones propias para crear dependencia de los usuarios hacia sus productos y así expulsar a los competidores.

> El formato IFC, destinado a ser un puente universal entre distintos sistemas CAD- (BIM-), cumple en realidad la función de indicador de problemas de compatibilidad entre los núcleos geométricos de distintas plataformas CAD, de forma similar al formato STEP del que surgió originalmente.

Como resultado, hoy en día una implementación completa y de alta calidad de la ontología IFC es factible para los grandes proveedores de CAD, que pueden invertir importantes recursos para dar soporte a todas

las entidades y su mapeo a su propio núcleo geométrico interno, que no existe para IFC como estándar. Los grandes proveedores también tienen la capacidad de coordinar entre ellos los detalles técnicos de las características que pueden no estar disponibles ni siquiera para el participante más activo en las organizaciones de desarrollo del formato IFC.

> Para los pequeños equipos independientes y los proyectos de código abierto, que se esfuerzan por apoyar el desarrollo de formatos interoperables, la falta de un núcleo geométrico propio se convierte en un grave problema. Sin él, es prácticamente imposible tener en cuenta todas las sutilezas y matices asociados al intercambio de datos entre plataformas.

Con el desarrollo del formato paramétrico IFC y el concepto de BIM abierto, en el sector de la construcción se han intensificado los debates sobre el papel de la ontología y la semántica en la gestión de datos y procesos.

Aparición en la construcción del tema de semántica y ontología

Gracias a las ideas de la Internet semántica finales de los noventa y a los esfuerzos de las organizaciones implicadas en el desarrollo del formato IFC, la semántica y las ontologías se han convertido en uno de los elementos clave de la normalización que se debatirá en el sector de la construcción a mediados de la década de 2020.

> Las tecnologías semánticas permiten unificar, normalizar y modificar grandes conjuntos de datos heterogéneos y realizar búsquedas complejas.

OWL (Web Ontology Language), representado como grafos RDF -triplets (Resource Description Framework) (Fig. 6.1-7), se utiliza para almacenar datos semánticos. OWL hace referencia a modelos de datos de grafos, cuyos tipos tratamos con más detalle en el capítulo "Modelos de datos: relaciones de datos y relaciones entre elementos".

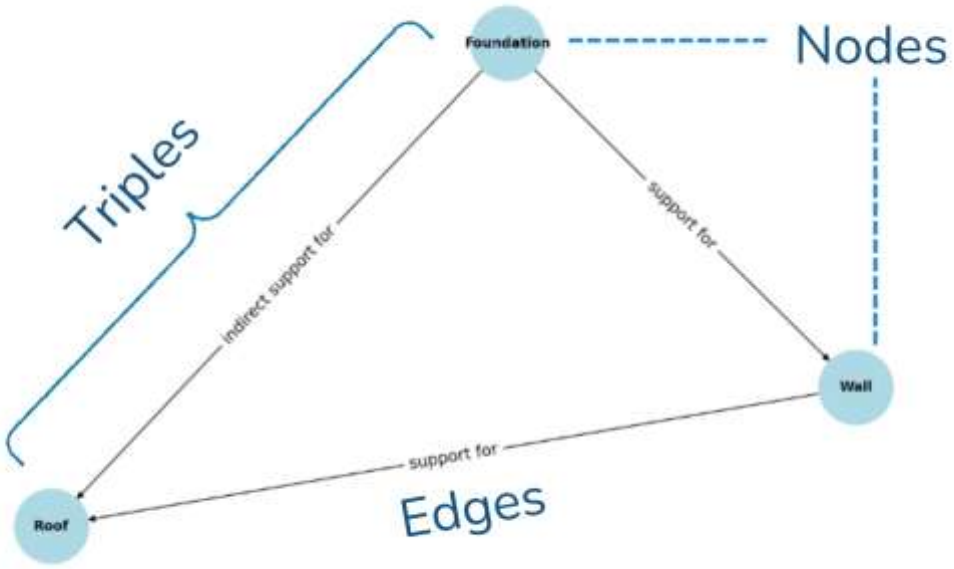


Fig. 6.1-7 Modelo de datos RDF: Nodos, Aristas y Triplas que ilustran las relaciones entre los bloques de construcción.

Teóricamente, la inferencia lógica de los risoners (software de inferencia lógica automática) permite derivar nuevas afirmaciones de las ontologías. Por ejemplo, si la ontología de la construcción registra que "un cimiento es un soporte para un muro" y "un muro es un soporte para un tejado" (Fig. 6.1-7), el risoner es capaz de inferir automáticamente que "un cimiento es un soporte para un tejado".

Este mecanismo es útil para optimizar el análisis de datos porque evita prescribir explícitamente todas las dependencias. Sin embargo, no crea nuevos conocimientos, sino que se limita a identificar y estructurar hechos ya conocidos.

> La semántica no crea nuevos significados o conocimientos en sí misma y no es superior a otras tecnologías de almacenamiento y tratamiento de datos en este aspecto. Representar los datos de las bases de datos relacionales como tripletas no los hace más significativos. Sustituir tablas por estructuras de grafos puede ser útil para unificar modelos de datos, facilitar su recuperación y editarlos con seguridad, pero no hace que los datos sean "más inteligentes": el ordenador no empieza a entender mejor su contenido.

Las relaciones lógicas en los datos pueden organizarse sin necesidad de complejas tecnologías semánticas (Fig. 6.1-8). Las bases de datos relacionales tradicionales (SQL), así como los formatos CSV o XLSX permiten construir dependencias similares. Por ejemplo, en una base de datos columnar, se puede añadir un campo "soporte del tejado" y asociar automáticamente el tejado con los cimientos al crear un muro. Este enfoque se aplica sin utilizar RDF, OWL, grafos ni risoners, por lo que sigue siendo una solución sencilla y eficaz para almacenar y analizar datos.

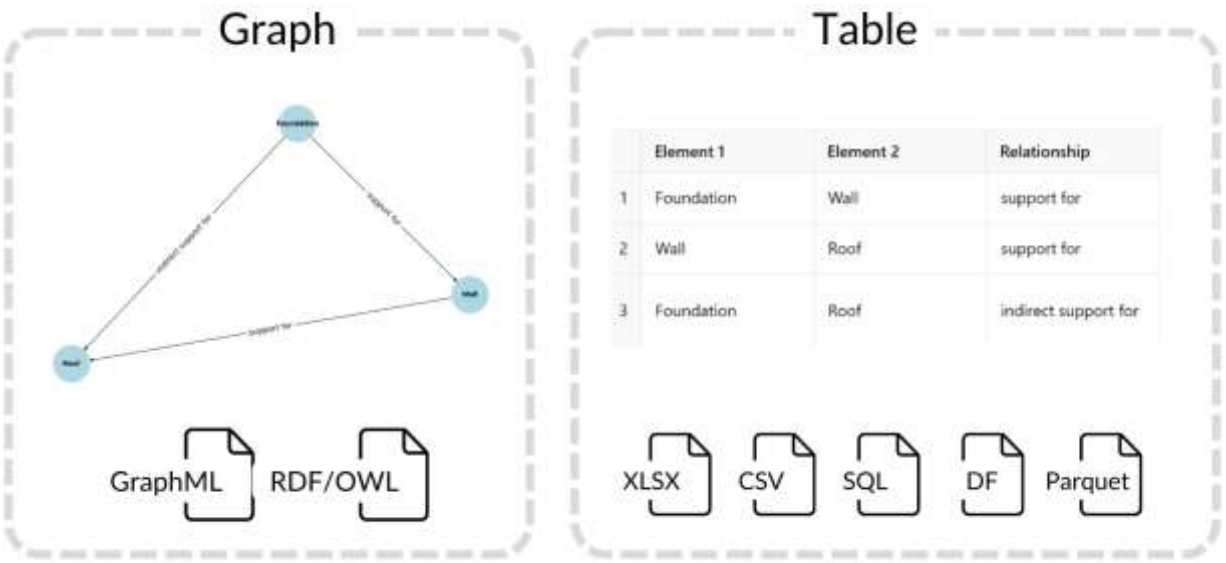


Fig. 6.1-8 Comparación de los modelos de datos de gráficos y tablas para representar las mismas relaciones lógicas.

La decisión de varias grandes empresas constructoras y de la organización de desarrollo de formatos IFC [94] de seguir el concepto de web semántica, que parecía prometedor a finales de los noventa, ha tenido un impacto significativo en el desarrollo de normas en el sector de la construcción.

Sin embargo, lo paradójico es que el propio concepto de web semántica, pensado originalmente para Internet, no ha sido ampliamente adoptado ni siquiera en su entorno nativo. A pesar del desarrollo de RDF y OWL, la web semántica propiamente dicha no ha aparecido en su concepción original, y su creación es ya improbable.

Por qué las tecnologías semánticas no cumplen las expectativas en el sector de la construcción

Otras industrias se han enfrentado a las limitaciones de las tecnologías para utilizar la semántica. En la industria del juego, los intentos de describir los objetos del juego y sus interacciones mediante ontologías han resultado ineficaces debido a la elevada dinámica de cambio. En consecuencia, se ha optado por formatos de datos más sencillos, como XML y JSON, junto con soluciones algorítmicas. La situación fue similar en el sector inmobiliario: debido a las diferencias regionales en la terminología y a los frecuentes cambios del mercado, el uso de ontologías resultó excesivamente complejo, mientras que bases de datos sencillas y estándares como RETS [119] pudieron afrontar mejor los retos del intercambio de datos.

> Las dificultades técnicas, como la complejidad del marcado, el elevado coste de mano de obra y la escasa motivación de los desarrolladores, frenaron la adopción de la web semántica y en otros sectores de la economía. RDF (Resource Description Framework) no se convirtió en un estándar de masas, y las ontologías resultaron demasiado complejas y económicamente injustificadas.

Como resultado, la ambiciosa idea de crear una web semántica global no llegó a materializarse. Aunque algunos elementos de la tecnología, como las ontologías y SPARQL, se han abierto camino en las soluciones empresariales, no se ha logrado el objetivo original de crear una única estructura de datos global.

El concepto de una Internet en la que los ordenadores sean capaces de dar sentido a los contenidos ha demostrado ser técnicamente difícil y comercialmente poco rentable. Por eso, las empresas que apoyaron la idea acabaron reduciendo su uso a herramientas útiles individuales, dejando RDF y OWL para necesidades corporativas muy especializadas y no para Internet en su conjunto. Un análisis de las tendencias de Google (Fig, 6.1-9) en los últimos 20 años sugiere que quizá ya no haya perspectivas para la web semántica.

> *No es necesario multiplicar las entidades innecesariamente. Si hay varias explicaciones lógicamente coherentes de un fenómeno que lo explican igual de bien, uno debería, en igualdad de condiciones, preferir la más sencilla de ellas.*
>
> - La navaja de

Aquí surge una pregunta lógica: ¿por qué utilizar tripletas, risers y SPARQL en la construcción en absoluto, cuando se pueden procesar datos utilizando consultas estructuradas populares (SQL, Pandas, Apache®)? En las aplicaciones empresariales, SQL es el estándar para trabajar con bases de datos. SPARQL, por el contrario, requiere complejas estructuras de grafos y software especializado y, según las tendencias de Google, no despierta el interés de los desarrolladores.

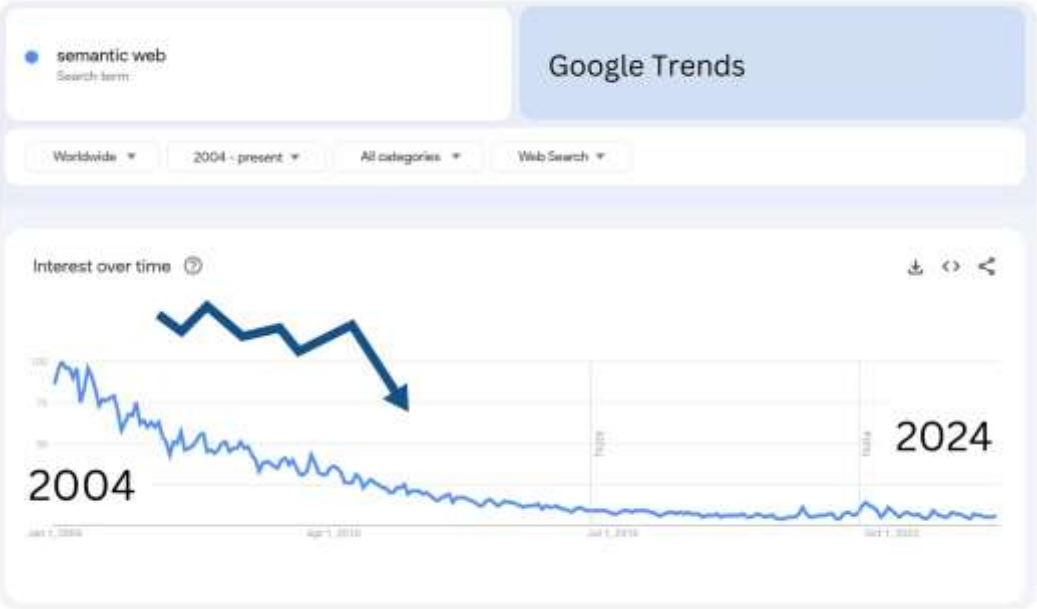


Fig. 6.1-9 Interés de las consultas sobre "Internet semántico" según las estadísticas de Google.

Las bases de datos de grafos y los árboles de clasificación pueden ser útiles en algunos casos, pero su aplicación no siempre está justificada para la mayoría de las tareas cotidianas. En consecuencia, la creación de grafos de conocimiento y el uso de tecnologías de web semántica sólo tienen sentido cuando es necesario unificar datos de distintas fuentes o realizar conclusiones lógicas complejas.

Pasar de las tablas a los modelos de datos gráficos mejora la búsqueda y unifica el flujo de información, pero no hace que los datos tengan más sentido para las máquinas. La cuestión no es si hay que utilizar tecnologías semánticas, sino dónde marcan realmente la diferencia. Antes de implantar la ontología, la semántica y las bases de datos de grafos en su empresa, averigüe qué empresas utilizan ya con éxito estas tecnologías y dónde han fracasado.

A pesar de las ambiciosas expectativas, las tecnologías semánticas nunca llegaron a convertirse en una solución universal para estructurar los datos en el sector de la construcción. En la práctica, estas tecnologías no han conducido a una solución universal, sino que sólo han añadido nuevas complejidades, y estos esfuerzos se hacen eco de las ambiciones no realizadas del concepto de Internet Semántica, donde las expectativas superaron con creces la realidad.

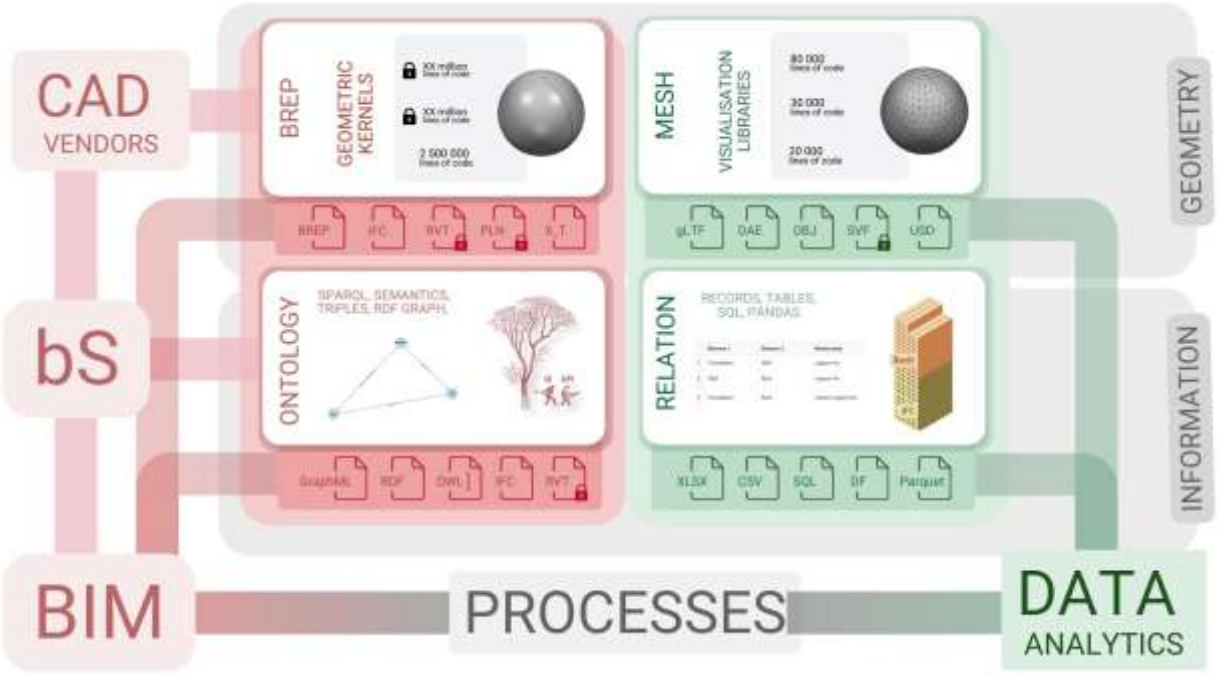


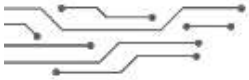
Fig. 6.1-10 Geometría e información en los procesos de construcción: de los complejos sistemas CAD y BIM- a los datos simplificados para análisis.

Mientras que en TI los fracasos de la web semántica se han visto compensados por la aparición de nuevas tecnologías (big data, IoT, machine learning, AR/VR), el sector de la construcción no tiene estas ocasiones.

Además de los retos que plantea el uso de conceptos para comunicar las relaciones de datos entre los elementos del proyecto, sigue existiendo un problema fundamental: la propia disponibilidad de esos datos. El sector de la construcción sigue dominado por sistemas cerrados, lo que dificulta trabajar con datos, compartir información y mejorar la eficiencia de los procesos.

Es el carácter cerrado de los datos lo que se está convirtiendo en una de las principales barreras que obstaculizan el desarrollo de soluciones digitales en la construcción. A diferencia de la industria informática, donde los formatos de datos abiertos y armonizados se han convertido en la norma, en el sector CAD (BIM) cada software utiliza su propio formato, creando ecosistemas cerrados y limitando artificialmente a los

usuarios.



CAPÍTULO 6.2.
FORMATOS DE PROYECTOS CERRADOS Y PROBLEMAS DE INTEROPERABILIDAD

Datos cerrados y caída de la productividad: el callejón sin salida de la industria CAD (BIM)

La naturaleza propietaria de los sistemas CAD ha llevado a que cada programa tenga su propio formato de datos, que o bien es cerrado e inaccesible desde el exterior - RVT, PLN, DWG, NDW, NWD, SKP, o bien está disponible en forma semiestructurada mediante un proceso de conversión bastante complejo - JSON, XML (CPIXML), IFC, STEP e ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, etc...

Los distintos formatos de datos en los que pueden almacenarse los mismos datos sobre los mismos proyectos no sólo difieren en su estructura, sino que también incluyen distintas versiones del marcado interno, que los desarrolladores deben tener en cuenta para garantizar la compatibilidad de las aplicaciones. Por ejemplo, un formato CAD de 2025 se abrirá en un programa CAD de 2026, pero el mismo proyecto nunca se abrirá en todas las versiones del programa CAD que puedan haber estado disponibles antes de 2025.

> Al no proporcionar acceso directo a las bases de datos, un proveedor de software del sector de la construcción suele crear su propio formato y herramientas exclusivos que un profesional (ingeniero de diseño o gestor de datos) debe utilizar para acceder a los datos, importarlos y exportarlos.

Como consecuencia, los vendedores de CAD básico (BIM) y soluciones relacionadas (por ejemplo, ERP/PMIS)) suben constantemente los precios por utilizar los productos, y los usuarios de a pie se ven obligados a pagar una "comisión" en cada etapa de la transferencia de datos por formatos [63]: por conectar, importar, exportar y trabajar con datos que los propios usuarios han creado.

El coste de acceder a los datos almacenados en la nube de los productos CAD más populares - (BIM-) alcanzará un dólar por transacción en 2025 [120], y las suscripciones a los productos ERP de la construcción para empresas medianas alcanzan sumas de cinco y seis cifras al año [121].

> La esencia del software de construcción moderno es que no es la automatización ni el aumento de la eficacia, sino la capacidad de los ingenieros para entender un determinado software altamente especializado lo que afecta a la calidad y el coste del procesamiento de datos de los proyectos de construcción, así como a los beneficios y la supervivencia a largo plazo de las empresas que los acometen.

La falta de acceso a las bases de datos CAD -sistemas que se utilizan en docenas de otros sistemas y cientos de procesos [63], y la consiguiente falta de comunicación de calidad entre los distintos especialistas ha llevado a la industria de la construcción a la categoría de uno de los sectores más ineficaces de la economía en términos de productividad [44].

En los últimos 20 años de aplicaciones de diseño CAD- (BIM-), aparición de nuevos sistemas (ERP), nuevas tecnologías y materiales de construcción, la productividad de todo el sector de la construcción ha descendido un 20% (Fig, 2.2-1), mientras que la productividad global de todos los sectores de la economía que no tienen grandes problemas para acceder a bases de datos y conceptos BIM similares a los de marketing ha aumentado un 70% (96% en la industria manufacturera) [122].

Fig. 6.2-1 Debido al aislamiento y la complejidad de los datos de los proyectos de los que dependen decenas de departamentos y cientos de procesos en el sector de la construcción, la velocidad de la toma de decisiones es varias veces menor que en otros sectores.

Sin embargo, también hay ejemplos aislados de enfoques alternativos para crear interoperabilidad entre soluciones CAD. La mayor empresa de construcción de Europa con el proyecto SCOPE [123], iniciado ya en 2018, demuestra cómo es posible ir más allá de la lógica clásica de los sistemas CAD- (BIM-). En lugar de intentar subyugar IFC o confiar en kernels de geometría propietarios, los desarrolladores de SCOPE utilizan APIs y SDKs ingeniería inversa para extraer datos de varios programas CAD, convertirlos a formatos neutrales como OBJ o CPIXML basados en el único kernel de geometría Open Source OCCT, y aplicarlos además a cientos de procesos de negocio de empresas de construcción y diseño. Sin embargo, a pesar de lo progresivo de la idea, estos proyectos se enfrentan a las limitaciones y complejidad de los kernels geométricos libres y formando parte de ecosistemas cerrados de una sola empresa que reproducen la lógica de las soluciones monovendedor.

Debido a las limitaciones de los sistemas cerrados y a las diferencias en los formatos de datos, así como a la falta de herramientas eficaces para su unificación, las empresas que tienen que trabajar con formatos CAD se a la acumulación de importantes cantidades de datos con distintos grados de estructura y cierre. Estos datos no se utilizan adecuadamente y desaparecen en archivos, donde permanecen para siempre olvidados y sin uso.

> Los datos obtenidos mediante un gran esfuerzo en la fase de diseño se vuelven inaccesibles para su uso posterior debido a su complejidad y naturaleza cerrada.

Como resultado, en los últimos 30 años, los desarrolladores del sector de la construcción se han visto obligados a enfrentarse al mismo problema una y otra vez: cada nuevo formato cerrado o solución propietaria genera la necesidad de integrarse con los sistemas CAD abiertos y cerrados existentes. Estos constantes intentos de garantizar la interoperabilidad entre diferentes soluciones CAD y BIM sólo sirven para complicar el ecosistema de datos, en lugar de contribuir a su simplificación y normalización.

El mito de la interoperabilidad entre sistemas CAD

Mientras que a mediados de la década de 1990 la dirección clave del desarrollo de la interoperabilidad en el entorno CAD era la ruptura del formato propietario DWG -que culminó con la victoria de la alianza Open DWG [75] y la apertura real del formato de dibujo más popular para todo el sector de la construcción-, a mediados de la década de 2020 el enfoque ha cambiado. Una nueva tendencia está cobrando impulso en el sector de la construcción: numerosos equipos de desarrollo se están centrando en crear los llamados "puentes" entre los sistemas CAD cerrados (BIM cerrado), el formato IFC y las soluciones abiertas (BIM abierto). La mayoría de estas iniciativas se basan en el uso del formato IFC y el núcleo geométrico OCCT, que proporcionan un puente técnico entre plataformas dispares. Este enfoque se considera una dirección prometedora que puede mejorar significativamente el intercambio de datos y la interoperabilidad de las herramientas de software.

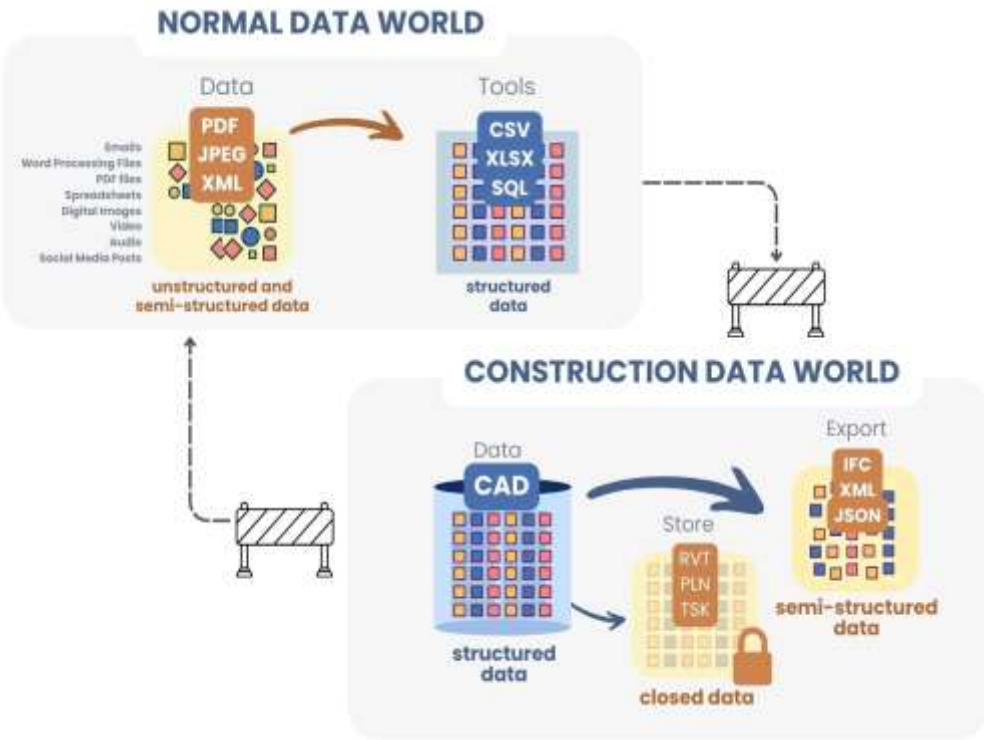


Fig. 6.2-2 Mientras otros sectores trabajan con datos abiertos, el de la construcción tiene que hacerlo con formatos CAD cerrados o poco estructurados (BIM).

Este enfoque tiene paralelismos históricos. En la década de 2000, los desarrolladores, tratando de superar el dominio del mayor vendedor de editores gráficos (2D world), intentaron crear una integración perfecta entre su solución propietaria y el código abierto gratuito, una alternativa a GIMP (Fig. 6.2-3). Entonces, como hoy en día en la construcción, se trataba de intentar tender un puente entre sistemas cerrados y abiertos, preservando al mismo tiempo los parámetros complejos, las capas y la lógica interna del software.

Sin embargo, los usuarios buscaban en realidad soluciones sencillas: datos planos y abiertos sin excesiva complejidad de capas y parámetros de programa (análogos del núcleo geométrico en CAD). Los usuarios buscaban formatos de datos sencillos y abiertos, libres de una lógica excesiva. JPEG, PNG y GIF se convirtieron en tales formatos en gráficos. Hoy en día se utilizan en redes sociales, páginas web y aplicaciones: son fáciles de procesar e interpretar, independientemente de la plataforma o el proveedor de software.

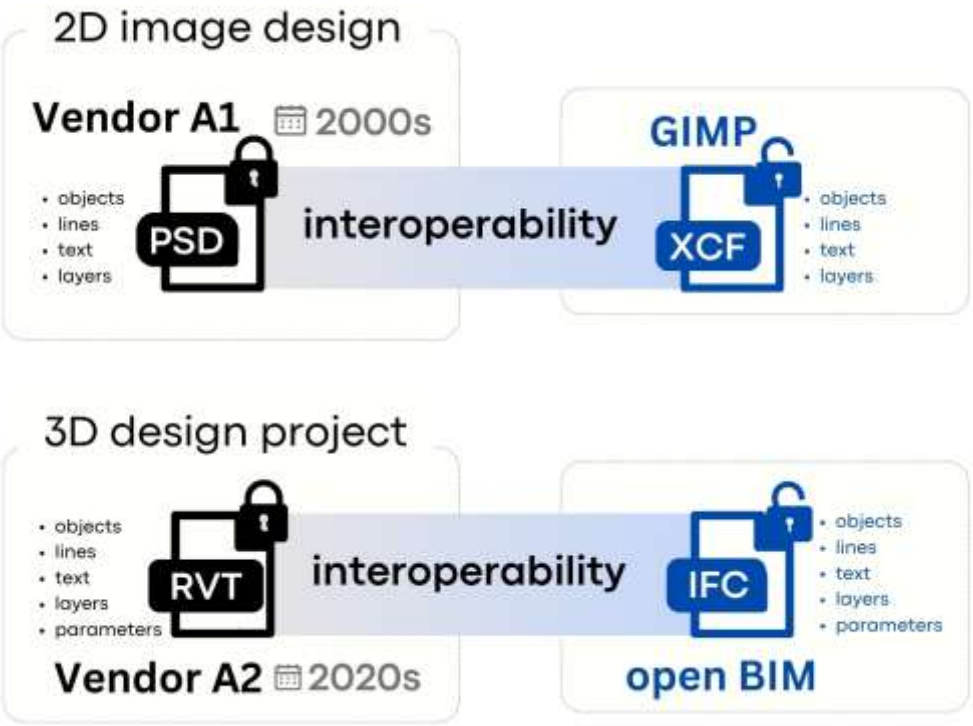


Fig. 6.2-3 La interoperabilidad de los formatos de datos en la construcción es similar a la trayectoria de los intentos de fusionar un producto propietario de un vendedor popular y el GIMP de código abierto en la década de 2000.

Como resultado, hoy en día casi nadie en la industria de la imagen utiliza formatos cerrados como PSD o abiertos XCF para aplicaciones, redes sociales como Facebook e Instagram, o como contenido en sitios web. En su lugar, la mayoría de las tareas utilizan formatos planos y abiertos JPEG, PNG y GIF por su facilidad de uso y amplia compatibilidad. Los formatos abiertos como JPEG y PNG se han convertido en el estándar para compartir imágenes debido a su versatilidad y amplia compatibilidad, lo que facilita su uso en diversas plataformas. Una transición similar puede observarse en otros formatos de intercambio, como el vídeo y el audio, donde destacan formatos universales como MPEG y MP3 por su eficacia de compresión

y amplia compatibilidad. Este movimiento hacia la normalización ha simplificado el intercambio y la reproducción de contenidos e información, haciéndolos accesibles a todos los usuarios a través de múltiples plataformas (Fig, 6.2-4).

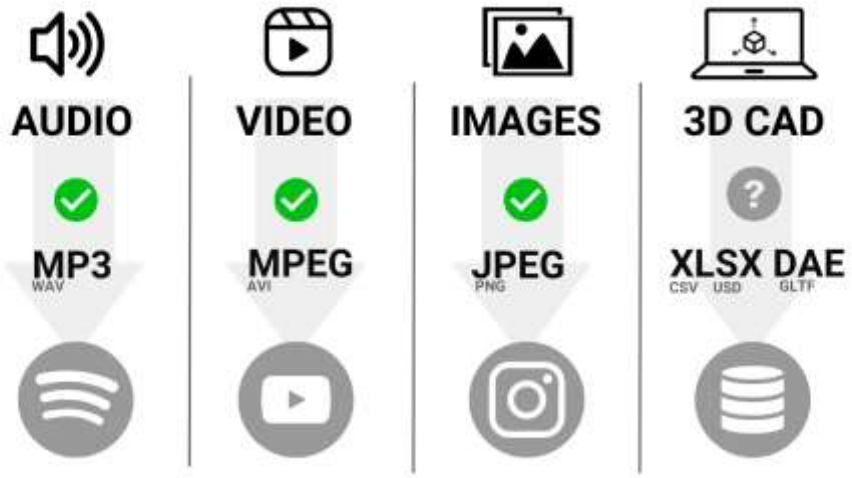


Fig. 6.2-4 Los formatos simplificados, sin funciones de edición complejas, se han popularizado para compartir y utilizar datos.

Procesos similares ocurren en la modelización 3D. Los formatos sencillos y abiertos como USD, OBJ, glTF, DAE, DXF, SQL y XLSX se utilizan cada vez más en proyectos para el intercambio de datos fuera del entorno CAD (BIM). Estos formatos almacenan toda la información necesaria, incluida la geometría y los metadatos, sin necesidad de manejar una compleja estructura BREP, núcleos de geometría o clasificadores internos específicos del proveedor. Formatos propietarios como NWC, SVF, SVF2, CPIXML y CP2 proporcionados por los principales proveedores de software también realizan funciones similares, pero siguen siendo cerrados, a diferencia de los estándares abiertos.

> Cabe destacar (y vale la pena recordarlo de nuevo, como ya se ha mencionado en el capítulo anterior) que esta idea -el rechazo de formatos intermedios neutros y paramétricos como IGES, STEP e IFC - fue apoyada ya en 2000 por el principal proveedor de CAD que creó el Libro Blanco BIM y registró el formato IFC en 1994. En el Libro Blanco 2000 "Diseño y fabricación integrados" [65], el proveedor de CAD subraya la importancia del acceso nativo a la base de datos CAD dentro del entorno de software, sin necesidad de utilizar traductores intermedios ni formatos paramétricos, para mantener la integridad y precisión de la información.

El sector de la construcción aún no se ha puesto de acuerdo ni sobre las herramientas de acceso a las bases de datos CAD ni sobre su ingeniería inversa forzada, ni sobre la adopción de un formato de datos simplificado común para su uso fuera de las plataformas CAD (BIM). Por ejemplo, muchas grandes empresas de Europa Central y regiones de habla alemana que operan en el sector de la construcción utilizan el formato CPIXML en sus sistemas ERP [121]. Este formato propietario, que es una especie de XML, combina datos de proyectos CAD (BIM), incluidos los geométricos y los metadatos, en una única estructura simplificada y organizada. Las grandes constructoras también están creando nuevos formatos y sistemas propios, como en el proyecto SCOPE, del que hablamos en el capítulo anterior

La lógica cerrada de los formatos CAD paramétricos o los complejos archivos paramétricos IFC (STEP) son redundantes en la mayoría de los procesos empresariales. Los usuarios buscan formatos simplificados y planos como USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE &XLSX, que contienen toda la información necesaria de los elementos, pero no están cargados con la lógica redundante de la geometría BREP, la dependencia de los núcleos de geometría y las clasificaciones internas de productos específicos CAD y BIM (Fig. 6.2-5).

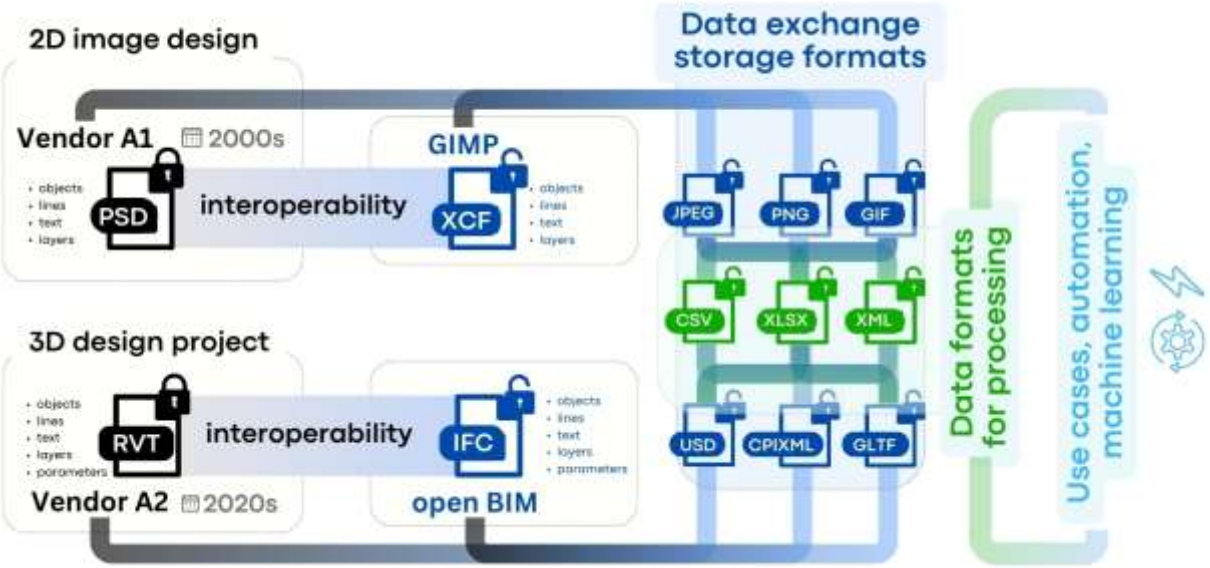


Fig. 6.2-5 Para la mayoría de los casos de uso, los usuarios eligen los formatos más sencillos posibles, independientes de los programas de los proveedores.

La aparición de formatos de imagen planos como JPEG, PNG y GIF, libres de la lógica redundante de los motores internos de los proveedores, ha impulsado el desarrollo de miles de soluciones interoperables para procesar y utilizar gráficos. Esto ha propiciado la aparición de aplicaciones que van desde herramientas de retoque y filtrado hasta redes sociales como Instagram, Snapchat y Canva, donde estos datos simplificados pueden utilizarse sin estar atados a un desarrollador de software específico.

> La normalización y simplificación de los formatos de diseño CAD estimulará la aparición de muchas herramientas nuevas, fáciles de usar e independientes, para trabajar con proyectos de construcción.

Pasar de la compleja lógica de las aplicaciones de proveedores vinculadas a núcleos geométricos cerrados a formatos abiertos universales basados en bibliotecas de elementos simplificados crea los requisitos previos para un tratamiento de datos más flexible, transparente y eficaz. Esto también abre el acceso a la información a todas las partes implicadas en el proceso de construcción, desde los diseñadores hasta los clientes y los servicios de mantenimiento.

No obstante, es muy probable que en los próximos años los proveedores de CAD intenten cambiar de nuevo el debate sobre la interoperabilidad de y el acceso a las bases de datos CAD. Se tratará ya de conceptos "nuevos" -como datos granulares, gráficos inteligentes, "modelos federados", gemelos digitales en

repositorios en la nube-, así como de la creación de alianzas y normas industriales que continúen la senda del BIM y el BIM abierto. A pesar de lo atractivo de la terminología, estas iniciativas pueden volver a convertirse en herramientas para retener a los usuarios dentro de ecosistemas propietarios. Un ejemplo es la promoción activa del formato USD (Universal Scene Description) como "nuevo estándar" para la colaboración entre plataformas CAD (BIM) a partir de 2023.

Ir a USD y datos granulares m

La aparición de la alianza AOUSD [124] en 2023 marca un giro importante en el sector de la construcción. Asistimos al comienzo de una nueva realidad, configurada por los proveedores de CAD, en el tratamiento de los datos de construcción a través de varios cambios significativos. El primer cambio importante se refiere a la percepción de los datos de CAD. Los profesionales que participan en las primeras fases del diseño conceptual son cada vez más conscientes de que la creación de un diseño en un entorno CAD es sólo el punto de partida. Los datos generados durante el proceso de diseño acaban convirtiéndose en la base del análisis, el funcionamiento y la gestión de los objetos. Esto significa que deben ser accesibles y utilizables en sistemas que vayan más allá de las herramientas CAD tradicionales.

Paralelamente, se está produciendo una revolución en el planteamiento de los principales desarrolladores. El principal proveedor de CAD- del sector, creador del concepto BIM y del formato IFC, está dando un giro inesperado a su estrategia. A partir de 2023, la empresa se aleja del almacenamiento tradicional de datos en archivos individuales, centrándose en el trabajo con datos granulares (normalizados y estructurados) y pasando a un enfoque centrado en los datos [125].

> Los proveedores están siguiendo las tendencias históricas de otros sectores: la mayoría de los usuarios no necesitan formatos CAD cerrados (similares a PSD) ni complejos archivos IFC paramétricos (similares a GIMP con lógica de capas). Necesitan imágenes de objetos sencillas que puedan utilizarse en CAFM (Instagram de la construcción), ERP (Facebook) y miles de otros procesos repletos de hojas de cálculo Excel y documentos PDF.

Las tendencias actuales en el sector de la construcción podrían sentar las bases para un abandono gradual de los formatos paramétricos y complejos en favor de los formatos más universales e independientes USD, GLTF, DAE, OBJ (con metainformación tanto dentro del híbrido como en formatos separados estructurados o poco estructurados). Los líderes históricos, incluidas las principales empresas de diseño que en su día promovieron activamente IFC a mediados de los noventa, promueven ahora abiertamente el nuevo formato USD [93], destacando su sencillez y versatilidad (Fig. 6.2-6). La adopción masiva de USD en productos, la compatibilidad con GLTF y la integración activa en herramientas como Blender, Unreal Engine y Omniverse muestran el potencial del inicio de un nuevo paradigma para trabajar con datos. Junto a la popularidad de soluciones localizadas como el formato plano europeo del USD - CPIXML, utilizado en los ERP europeos más populares podría reforzar potencialmente la posición del USD en Europa Central. Las organizaciones implicadas en el desarrollo del formato IFC ya están adaptando su estrategia al USD [126], lo que no hace sino confirmar la inevitabilidad del cambio.

Technical Specifications			Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON
Export/Import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement

Fig. 6.2-6 Comparación de las especificaciones técnicas de los formatos IFC y USD.

> En este contexto, USD tiene potencial para convertirse en la norma de facto, ya que promete superar muchas de las limitaciones actuales, relacionadas principalmente con la complejidad de los formatos CAD existentes - (BIM-) y la dependencia de su interpretación de los núcleos geométricos.

En lugar de los paramétricos y complejos formatos CAD e IFC - los formatos de datos simplificados USD, gLTF, DAE, OBJ con metainformación de elementos en CSV, XLSX, JSON, XML ganarán un lugar en la industria de la construcción debido a su simplicidad y flexibilidad.

Los cambios actuales en el sector de la construcción parecen a primera vista un avance tecnológico asociado a la transición del anticuado IFC al más moderno USD. Sin embargo, conviene tener en cuenta que ya

en el año 2000 el mismo proveedor de CAD, que desarrolló IFC, escribió sobre sus problemas y la necesidad de acceder a la base de datos [65], y ahora promueve activamente la transición a un nuevo estándar: USD.

> Detrás de otra fachada de "datos abiertos" USD y "nuevos" conceptos para la gestión granular de datos, a través de las aplicaciones basadas en la nube que los proveedores de CAD están empezando a promocionar, puede esconderse la intención de los proveedores de monopolizar la gestión de datos de proyectos, donde los usuarios se encuentran en una posición en la que la elección del formato está más relacionada con los intereses corporativos que con las necesidades reales.

Un análisis de los hechos clave [93] muestra que el principal objetivo de estos cambios no es tanto la comodidad del usuario como el mantenimiento del control sobre los ecosistemas y los flujos de datos en beneficio de los vendedores que, en 40 años, nunca han sido capaces de proporcionar acceso a las bases de datos CAD.

Quizá haya llegado el momento de que las empresas dejen de esperar nuevos conceptos de los proveedores de software y se centren en el autodesarrollo en la dirección centrada en los datos. Una vez liberada de los problemas de acceso a los datos mediante herramientas de ingeniería inversa, la industria podrá avanzar de forma autónoma hacia herramientas modernas, gratuitas y cómodas para trabajar y analizar los datos sin imponer nuevos conceptos.

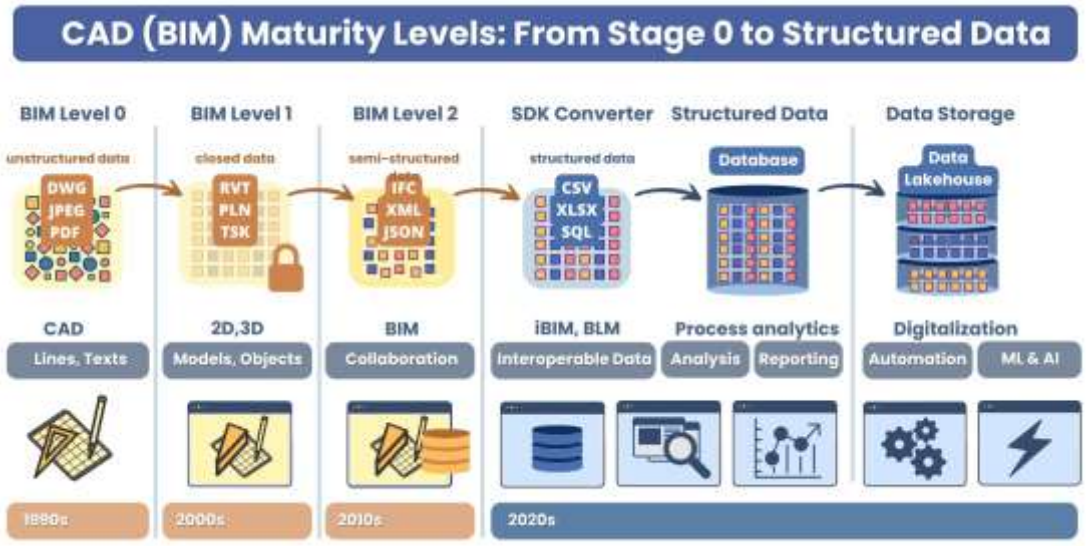


Fig. 6.2-7 Nivel de madurez CAD (BIM): de los datos no estructurados a los datos estructurados y los repositorios.

El acceso a las bases de datos, los datos abiertos y los formatos se convertirá inevitablemente en un estándar en el sector de la construcción, independientemente de los intentos de los proveedores de paralizar el proceso: es sólo cuestión de tiempo (Fig. 6.2-7). El ritmo de esta transición puede aumentar considerablemente si cada vez más profesionales se familiarizan con los formatos abiertos, las herramientas de bases de datos y los SDK de ingeniería inversa disponibles, que permiten el acceso directo a los datos CAD -sistemas [92].

> El futuro está en los datos abiertos, unificados y accesibles analíticamente. Para no depender de las soluciones de los proveedores y no ser rehenes de ecosistemas cerrados, las empresas de construcción e ingeniería tendrán que apostar tarde o temprano por la apertura y la independencia, eligiendo formatos y soluciones que proporcionen un control total sobre los datos.

Los datos que se están creando hoy en el sector de la construcción serán un recurso clave para las decisiones empresariales en el futuro. Actuarán como el "combustible" estratégico que impulse el desarrollo y la eficiencia de las empresas de construcción. El futuro del sector de la construcción reside en la capacidad de trabajar con datos, no en la elección de formatos o modelos de datos.

Para entender la diferencia entre los formatos abiertos USD, glTF, DAE, OBJ y los formatos CAD paramétricos propietarios, es importante considerar uno de los elementos de datos más complejos y clave en la visualización y los cálculos de diseño: la geometría y sus procesos de generación. Y para entender cómo los datos geométricos se convierten en la base de la analítica y los cálculos en la construcción, es necesario profundizar en los mecanismos de generación, transformación y almacenamiento de la geometría.

CAPÍTULO 6.3.
GEOMETRÍA EN LA CONSTRUCCIÓN: DE LAS LÍNEAS A LOS METROS CÚBICOS

Cuando las líneas se convierten en dinero o por qué los constructores necesitan geometría

La geometría en la construcción no es sólo una visualización, sino también la base de cálculos cuantitativos precisos. En el modelo de proyecto, la geometría complementa las listas de parámetros de los elementos (Fig. 3.1-16) con importantes características volumétricas como la longitud, el área y el volumen. Estos valores de los parámetros volumétricos se calculan automáticamente utilizando los núcleos de geometría y son el punto de partida para las estimaciones, los calendarios y los modelos de recursos. Como ya hemos comentado en la Parte 5 de este libro y en el capítulo "Cálculo de costes y estimaciones de proyectos de construcción ", son los parámetros volumétricos de los grupos de objetos de los modelos CAD los que constituyen la base de los modernos sistemas ERP, PMIS / La geometría desempeña un papel fundamental no sólo en la fase de diseño, sino también en la gestión de la ejecución del proyecto, el control del calendario, la elaboración de presupuestos y la explotación. Al igual que hace miles de años, cuando se construían las pirámides egipcias, la precisión de un proyecto dependía de medidas de longitud como los codos y los codos, hoy en día la precisión de la interpretación de la geometría en los programas CAD -afecta directamente al resultado: desde el presupuesto y los plazos hasta la selección del contratista y la logística de entrega

> En un entorno altamente competitivo y de presupuesto limitado, la precisión de los cálculos volumétricos, que depende directamente de la geometría, se convierte en un factor de supervivencia. Los modernos sistemas ERP dependen directamente de las características volumétricas correctas obtenidas de los modelos CAD - y BIM -. Por eso, la descripción geométrica exacta de los elementos no es sólo una visualización, sino una herramienta clave para gestionar el coste y el tiempo de construcción.

Históricamente, la geometría ha sido el lenguaje principal de la comunicación en ingeniería. Desde las líneas en papiro hasta los modelos digitales, los dibujos y las representaciones geométricas han servido como medio de intercambio de información entre diseñadores, capataces y estimadores. Antes de la llegada de los ordenadores, los cálculos se hacían manualmente, con reglas y transportadores. Hoy en día, esta tarea está automatizada gracias a la modelización volumétrica: los núcleos geométricos de los programas de CAD convierten líneas y puntos en cuerpos tridimensionales de los que se extraen automáticamente todas las características necesarias.

En los programas CAD, la creación de elementos geométricos para los cálculos se realiza a través de la interfaz de usuario de los programas CAD- (BIM-). Para transformar puntos y líneas en cuerpos volumétricos, se utiliza el núcleo geométrico, que cumple la tarea clave: la transformación de la geometría en modelos volumétricos, a partir de los cuales se calculan automáticamente las características volumétricas del elemento tras la aproximación.

De líneas a volúmenes: cómo el área y el volumen se convierten en datos

En la práctica de la ingeniería, los volúmenes y las áreas se calculan a partir de superficies geométricas descritas analíticamente o mediante modelos paramétricos como NURBS (nonuniform rational B-splines) dentro del marco BREP (boundary element representation).

> NURBS (Non-Uniform Rational B-Splines) es una forma matemática de describir curvas y superficies, mientras que BREP es un marco para describir la geometría tridimensional completa de un objeto, incluidos sus límites, que puede definirse mediante NURBS.

A pesar de la precisión de BREP y NURBS, requieren potentes recursos informáticos y algoritmos complejos. Sin embargo, el cálculo directo a partir de descripciones tan precisas desde el punto de vista matemático suele ser difícil, por lo que en la práctica casi siempre se recurre a la teselación (transformación de superficies en una rejilla de triángulos), que simplifica los cálculos posteriores. Teselación es la partición de una superficie compleja en triángulos o polígonos. En los entornos CAD /CAE, este método se utiliza para la visualización, el cálculo de volúmenes, la búsqueda de colisiones, la exportación a formatos como MESH y el análisis de colisiones. Un ejemplo de la naturaleza es un panal de abejas, donde una forma compleja se descompone en una cuadrícula regular (Fig. 6.3-1).

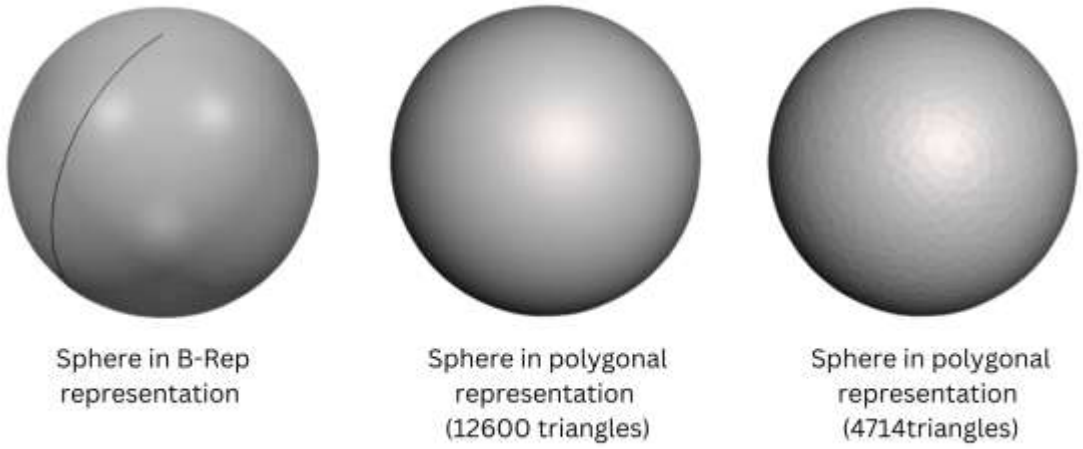


Fig. 6.3-1 La misma esfera en descripción paramétrica BREP y representación poligonal con diferente número de triángulos.

BREP (NURBS), utilizado en CAD, no es un modelo fundamental de geometría. Se creó como una herramienta cómoda para representar círculos y splines racionales y para minimizar el almacenamiento de datos geométricos. Sin embargo, tiene limitaciones - por ejemplo, la incapacidad de describir con precisión la sinusoide que subyace a las líneas y superficies helicoidales, y la necesidad de utilizar núcleos geométricos complejos.

Por el contrario, las mallas triangulares y la teselación de formas paramétricas se caracterizan por su sencillez, el uso eficiente de la memoria y la capacidad de procesar grandes cantidades de datos (Fig. 6.3-2). Estas ventajas permiten prescindir de los complejos y costosos núcleos geométricos, y de las decenas de millones de líneas de código que llevan incorporadas, a la hora de calcular formas geométricas.

En la mayoría de los casos de construcción no importa cómo se definan exactamente las características volumétricas: mediante modelos paramétricos (BREP, IFC) o mediante polígonos (USD, glTF, DAE, OBJ). La geometría sigue siendo la forma de aproximación: ya sea mediante NURBS o MESH, siempre se trata de una descripción aproximada de la forma.

La geometría definida como polígonos o BREP (NURBS) sigue siendo hasta cierto punto sólo una forma de aproximarse con una descripción aproximada de una forma continua. Al igual que las integrales de Fresnel no tienen una expresión analítica exacta, la discretización de la geometría mediante polígonos o NURBS es siempre una aproximación, al igual que las MESH triangulares.

La geometría paramétrica en formato BREP es necesaria principalmente cuando el tamaño mínimo de los datos es importante y es posible utilizar núcleos geométricos caros y que consumen muchos recursos para su procesamiento y visualización. La mayoría de las veces es característico de los desarrolladores de programas CAD -programmes, que para este fin aplican en sus productos kernels geométricos de proveedores MCAD -vendors. En este caso, incluso dentro de estos programas, los modelos BREP en el proceso de teselación para su visualización y cálculo se convierten a menudo en triángulos (de forma similar a como los archivos PSD se simplifican en JPEG).

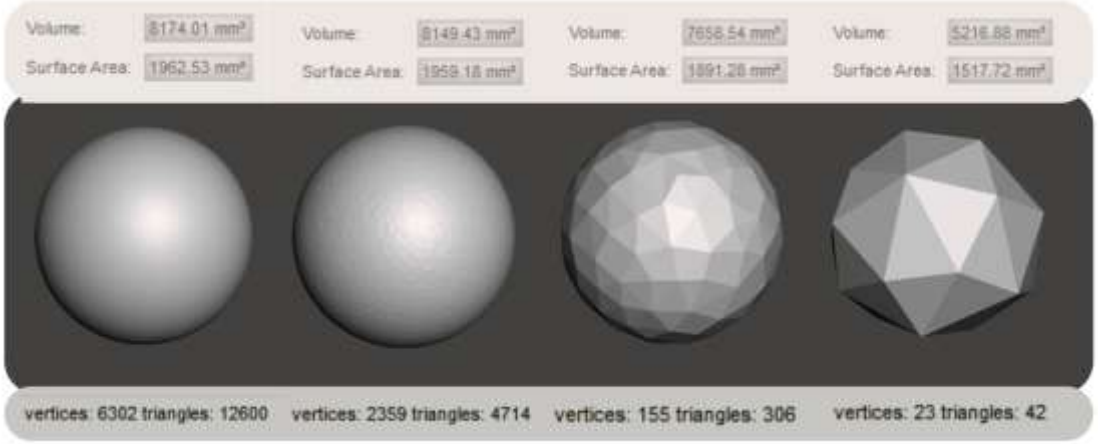


Fig. 6.3-2 Diferencia de características volumétricas en figuras con diferente número de polígonos.

El MESH poligonal, así como el BREP paramétrico, tienen sus propias ventajas y limitaciones, pero el objetivo es el mismo: describir la geometría teniendo en cuenta las tareas del usuario. En última instancia, la precisión de un modelo geométrico depende no sólo del método de su representación, sino también de los requisitos de una tarea concreta.

En la mayoría de los problemas de construcción, la necesidad de geometría paramétrica y núcleos geométricos complejos puede ser redundante.

En cada tarea concreta de automatización de cálculos, merece la pena considerar si la importancia de la geometría paramétrica es exagerada por los desarrolladores de CAD interesados en promocionar y vender sus propios productos de software.

Pasar a MESH, USD y polígonos: utilizar la teselación para la geometría

En el sector de la construcción, a la hora de transmitir, desarrollar sistemas, bases de datos o automatizar procesos que trabajen con información de diseño y geometría de elementos, es importante esforzarse por lograr la independencia de editores CAD y núcleos de geometría específicos.

> El formato de intercambio que se utilice tanto en los departamentos de cálculo como en la obra no debe basarse en un programa específico de CAD- (BIM-). La información geométrica debería representarse en el formato directamente mediante teselación, sin referencia al núcleo geométrico ni a la arquitectura CAD.

La geometría paramétrica procedente de CAD puede considerarse una fuente intermedia, pero no la base de un formato universal. En cualquier caso, la mayoría de las descripciones paramétricas (incluidas BREP y NURBS) se convierten a MESH poligonal para su posterior procesamiento. Si el resultado es el mismo (teselación y polígonos) y el proceso es más sencillo, la elección es obvia. Esto es similar a la elección entre ontologías gráficas y tablas estructuradas (que analizamos en la cuarta parte): la complejidad excesiva rara vez está justificada (Fig. 3.2-10, Fig. 6.1-8).

Formatos abiertos como: OBJ, STL, glTF, SVF, CPIXML, USD y DAE, utilizan una estructura de malla triangular universal, lo que les confiere importantes ventajas. Estos formatos presentan una excelente interoperabilidad: son fáciles de leer y visualizar utilizando las bibliotecas de código abierto disponibles, sin necesidad de complejos núcleos geométricos especializados que contengan millones de líneas de código (Fig, 6.3-3). Estos versátiles formatos geométricos se utilizan en aplicaciones que van desde las relativamente sencillas herramientas de diseño de cocinas de IKEA™ hasta complejos sistemas de visualización de objetos en aplicaciones de cine y RV. Una ventaja importante es la disponibilidad de un gran número de bibliotecas gratuitas y de código abierto para trabajar con estos formatos, disponibles para la mayoría de plataformas y lenguajes de programación.

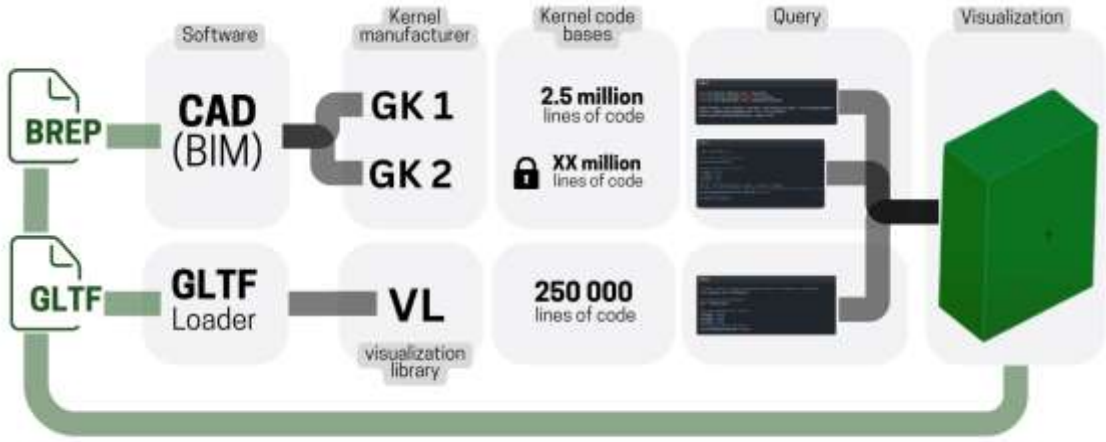


Fig. 6.3-3 La misma representación geométrica se consigue mediante el uso de formatos paramétricos y núcleos geométricos, o mediante el uso de formatos triangulados y bibliotecas de visualización de código abierto.

Además de los propios usuarios, los proveedores de CAD tienen problemas para interpretar formatos CAD paramétricos extranjeros o IFC abierto debido a los diferentes núcleos geométricos. En la práctica, todos los proveedores de CAD, sin excepción, utilizan el SDK de ingeniería inversa para transferir datos entre sistemas, y ninguno de ellos confía en formatos como IFC o USD [93] para fines de interoperabilidad.

> En lugar de utilizar conceptos promovidos por alianzas de proveedores de CAD- que ellos mismos no utilizan, es más productivo que los desarrolladores y usuarios de soluciones CAD se centren en comprender las ventajas de cada enfoque en un contexto específico y elijan uno u otro tipo de geometría en función del caso de uso. Elegir entre distintas representaciones geométricas es un compromiso entre precisión, eficiencia computacional y necesidades prácticas de una tarea concreta.

La complejidad asociada al uso de núcleos geométricos, tradicionalmente impuesta al sector de la construcción por los grandes proveedores a la hora de procesar los datos de diseño, resulta a menudo redundante. El formato USD basado en la geometría MESH puede convertirse en una especie de "caja de Pandora" para el sector, al abrir nuevas posibilidades para que los desarrolladores organicen el intercambio de datos, fuera del marco de las estructuras IFC y BREP paramétricas típicas de los proveedores de CAD.

Tras un examen más detallado de la estructura de USD, DAE, gLTF, OBJ, etc., resulta obvio que existen formatos más sencillos y abiertos que permiten organizar eficazmente la transferencia y el uso de información geométrica sin necesidad de depender de paramétricas complejas y núcleos geométricos cerrados. Este enfoque no sólo reduce el umbral técnico de entrada para los desarrolladores, sino que también favorece el desarrollo de soluciones flexibles, escalables y verdaderamente abiertas para la construcción digital.

LOD, LOI, LOMD - clasificación única de detalle en CAD (BIM)

Además de los formatos de representación geométrica, en un mundo en el que las distintas industrias utilizan diferentes niveles de detalle y profundidad de los datos, las metodologías CAD - (BIM-) ofrecen sus propios sistemas de clasificación, que estructuran el enfoque para informar a los modelos de construcción.

Uno de los ejemplos de nuevos enfoques de normalización es la introducción de niveles de desarrollo de modelos, que reflejan el grado de preparación y fiabilidad de los componentes gráficos y de información. Para la diferenciación del contenido de la información en el trabajo con datos CAD - (BIM-) aparecieron LOD (Level Of Detail) - nivel de detalle de la parte gráfica del modelo, y LOI (Level Of Information) - nivel de elaboración de los datos. Además, para el enfoque integrado se introdujo el concepto de LOA (Level of Accuracy) - precisión de los elementos presentados y LOG (Level of Geometry) para determinar la precisión de la representación gráfica.

Los niveles de detalle (LOD) se indican con números del 100 al 500, que reflejan el grado de desarrollo del modelo. LOD 100 es un modelo conceptual con formas y dimensiones generales. LOD 200 incluye dimensiones y formas más precisas, pero con detalles condicionales. LOD 300 es un modelo detallado con dimensiones, formas y ubicaciones de elementos precisas. LOD 400 contiene información detallada necesaria para la fabricación e instalación de elementos. LOD 500 refleja el estado real de la instalación después de la construcción y se utiliza para el funcionamiento y el mantenimiento. Estos niveles describen la estructura de la saturación de información del modelo CAD (BIM) en las distintas fases del ciclo de vida, incluidas

3D, 4D, 5D y más allá.

En los proyectos reales, el alto nivel de detalle (LOD400) suele ser excesivo y basta con utilizar geometría LOD100 o incluso planos, mientras que el resto de los datos pueden obtenerse computacionalmente o a partir de elementos relacionados que pueden no tener una geometría definida. Por ejemplo, los espacios y elementos de sala (categorías de elementos de sala) pueden no tener geometría visual, pero contienen cantidades significativas de información y bases de datos en torno a las cuales se construyen muchos procesos empresariales.

> Por tanto, es importante definir claramente el nivel de detalle necesario antes de iniciar el diseño. Para los casos de uso 4D -7D, a menudo bastan incluso dibujos DWG y una geometría mínima LOD100. La tarea clave en el proceso de requisitos es encontrar un equilibrio entre la riqueza y la practicidad del modelo.

En esencia, si consideramos los datos CAD (BIM) como una base de datos (que lo es), la descripción de la saturación del modelo a través de nuevas siglas no es más que un modelado de datos paso a paso para los sistemas de información, que parte del nivel conceptual y llega hasta el físico (Fig. 6.3-4), que se trató en detalle en la tercera y cuarta partes del libro. Cada aumento de LOD y LOI significa la adición de información necesaria para nuevas tareas: cálculos, gestión de la construcción, explotación, y se caracteriza por el enriquecimiento sucesivo del modelo con capas de información adicionales (3D -8D) en forma de diversos parámetros, de los que hablamos en la quinta parte del libro.

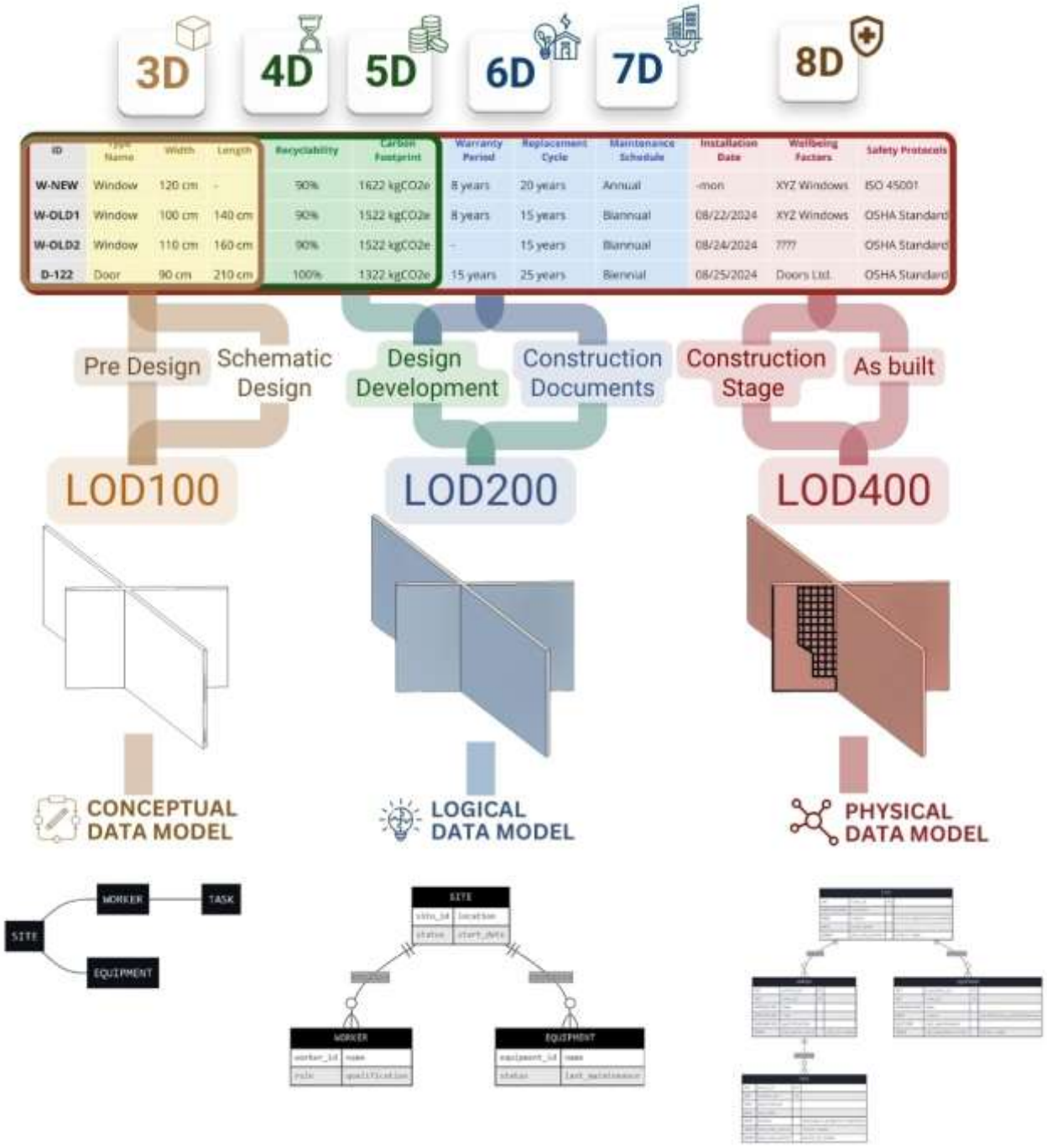


Fig. 6.3-4 El proceso de concretar los detalles de un proyecto es idéntico al modelado de datos desde un modelo de datos conceptual a uno físico.

La geometría es sólo una parte de los datos de diseño, cuya necesidad no siempre está justificada en los proyectos de construcción y la cuestión clave de trabajar con datos CAD no es tanto cómo se visualizan los modelos, sino más bien cómo pueden utilizarse los datos de estos modelos fuera de los programas CAD- (BIM -).

A mediados de la década de 2000, el sector de la construcción se enfrentó a un reto sin precedentes con

el rápido aumento de la cantidad de datos en los sistemas de gestión y procesamiento de datos, especialmente los procedentes de los departamentos de CAD (BIM). Este espectacular aumento del volumen de datos cogió por sorpresa a los directivos de las empresas, que no estaban preparados para las crecientes exigencias en materia de calidad y gestión de datos.

Nuevas normas CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Aprovechando la falta de acceso abierto a las bases de datos CAD y la escasa competencia en el mercado del tratamiento de datos, y utilizando las campañas de marketing asociadas al nuevo acrónimo BIM, las organizaciones implicadas en el desarrollo de enfoques para tratar los datos CAD- han empezado a crear nuevas normas y conceptos que de iure deberían tener como objetivo mejorar las prácticas de gestión de datos.

> Aunque casi todas las iniciativas apoyadas directa o indirectamente por los proveedores y desarrolladores de CAD (BIM) han tenido como objetivo optimizar los flujos de trabajo, han dado lugar a una plétora de normas defendidas por diversas partes interesadas, lo que ha llevado al sector de la construcción a cierta ambigüedad y confusión sobre los procesos de datos.

Vamos a enumerar algunas de las nuevas normas de datos, además de LOD, LOI, LOA, LOG, que han surgido en los últimos años en el sector de la construcción:

- **BEP** (BIM Execution Plan): describe cómo integrar y utilizar CAD (BIM) en un proyecto, definiendo métodos y procesos de tratamiento de datos.
- **Documento EIR /AIA** (Requisitos de información del cliente) - preparado por el cliente antes de la licitación y contiene los requisitos para que el contratista prepare y facilite información. Sirve de base para las BEP del respectivo proyecto.
- **AIM** (Asset Information Model) forma parte del proceso BIM. Una vez entregado y completado el proyecto, el modelo de datos se denomina Modelo de Información de Activos o AIM. La finalidad del AIM es gestionar, mantener y explotar el activo realizado.
- **IDS** (Information Delivery Specification) - define los requisitos de y qué datos y en qué formato se requieren en las distintas fases de un proyecto de construcción.
- **iLOD** es el nivel de detalle LOD, con el que se representa la información en el modelo BIM. Define lo detallada y completa que es la información del modelo, desde representaciones geométricas básicas hasta especificaciones y datos detallados.
- **eLOD** - LOD nivel de detalle de los elementos individuales de un modelo CAD (BIM). Define hasta qué punto se modela cada elemento y la información asociada, como dimensiones, materiales, características de rendimiento y otros atributos relevantes.
- **APS** (Platform Services) y otros productos de los principales proveedores de CAD (BIM) - describen en las herramientas y la infraestructura necesarias para crear modelos de datos abiertos y enlazados.

Aunque el propósito declarado de la implantación de estándares CAD (BIM) -como LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD- es mejorar la calidad de la gestión de datos y ampliar las capacidades de automatización, en la práctica su uso suele conducir a una excesiva complejidad y fragmentación de los procesos. Si consideramos el modelo CAD (BIM) como una especie de base de datos, resulta obvio que muchas de estas normas duplican enfoques ya establecidos y eficaces utilizados en otros sectores para trabajar con sistemas de información. En lugar de simplificar y unificar, estas iniciativas suelen crear una carga terminológica adicional y dificultan la aplicación de soluciones verdaderamente abiertas y flexibles.

En particular, muchos de estos nuevos conceptos están sustituyendo en realidad a los procesos de modelización y validación de datos que se analizaron en detalle en las primeras partes del libro y que se utilizan desde hace tiempo en otros sectores de la economía. En la construcción, en cambio, el proceso de normalización suele ir en dirección contraria: se crean nuevos formatos de descripción de datos, nuevas normas y nuevos conceptos de validación de datos, que no siempre conducen a una uniformidad real y a una aplicabilidad práctica. Como resultado, en lugar de simplificar y automatizar el procesamiento, la industria se enfrenta a niveles adicionales de regulación y burocracia (Fig. 6.3-1), lo que no siempre favorece una mayor eficiencia.

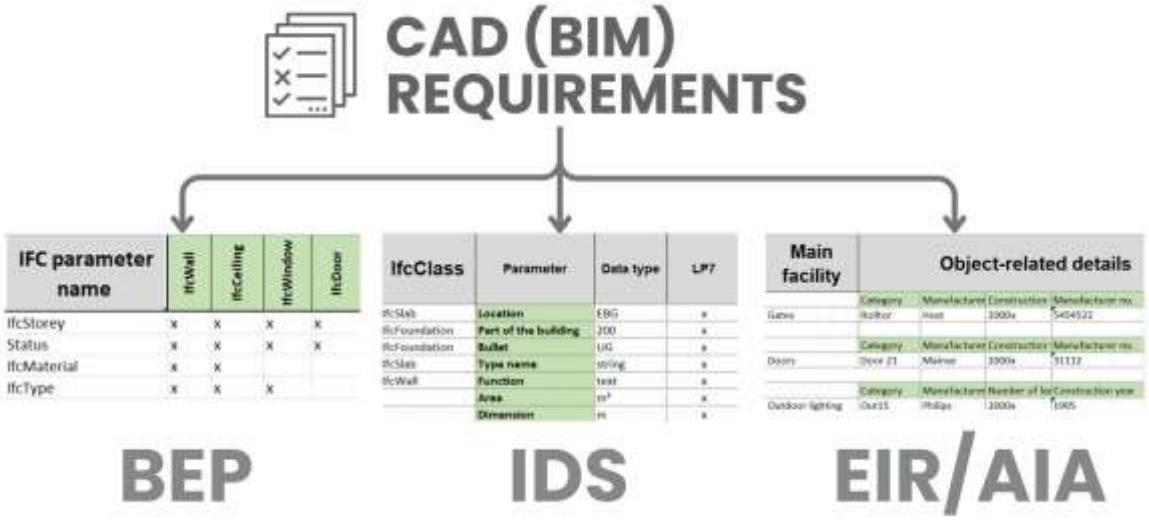


Fig. 6.3-1 Los requisitos de contenido de datos e información se reducen a la descripción de atributos y sus valores límite, descritos mediante tablas.

> En lugar de simplificar el tratamiento de datos, los nuevos conceptos relacionados con los datos CAD (BIM) suelen generar complejidades y disputas adicionales ya en la fase de interpretación y definiciones básicas.

Uno de los últimos ejemplos de nuevos conceptos es el formato IDS (introducido en 2020) que permite describir los requisitos para la composición de atributos de un modelo de información en el concepto de BIM abierto. Los requisitos IDS describen información sobre atributos y sus valores límite en forma de tabla estructurada (Excel o MySQL), que luego se traduce en el marcado de un formato XML semiestructurado, renombrado de XML a la abreviatura especial IDS.

> Contrariamente a la opinión promovida por los proveedores y apoyada por BIM y BIM abierto, de que el tratamiento de datos en la construcción es único debido al uso de herramientas especializadas como CAD y BIM, los formatos de datos y las prácticas de gestión de datos de esta industria no son diferentes de los otras industrias.

El número de requisitos para proyectos y formatos CAD (BIM) puede simplificarse utilizando una única tabla de requisitos con columnas de atributos, detallada en el capítulo "Traducción de requisitos a formato estructurado", sin tener que traducir los requisitos originalmente estructurados a formatos que no sean tablas (IDS se describe inicialmente mediante una tabla).

El enfoque simplificado (Fig, 6.3-2), que incluye columnas para identificadores de entidad, propiedades y valores límite que se analizaron en detalle en capítulos anteriores (Fig, 4.4-9, Fig. 4.4-16, Fig. 7.3-10), elimina la necesidad de convertir los requisitos al formato IDS-XML. Este método proporciona un mecanismo directo, menos engorroso y más transparente para el control de la calidad de los datos. Se basa en herramientas ampliamente utilizadas, desde expresiones regulares (RegEx) a dataframes, Pandas y ETL estándar -payplanes-, exactamente igual que las que utilizan los profesionales de otros sectores de la economía para trabajar con datos.

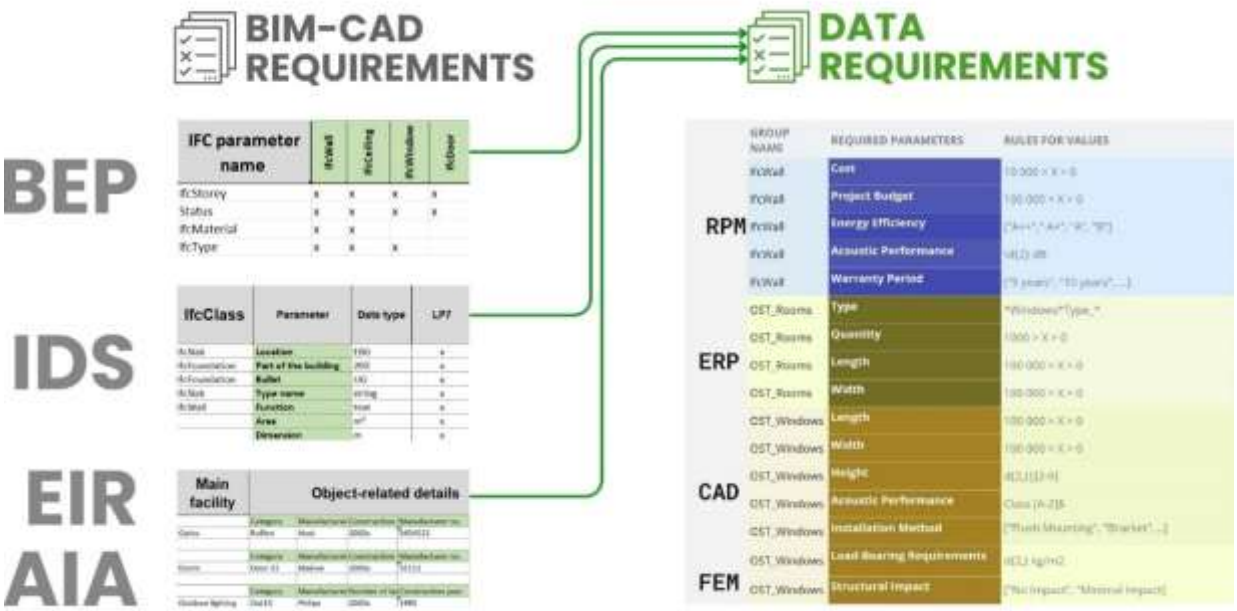


Fig. 6.3-2 Los requisitos de datos en otras industrias se simplifican a una descripción estructurada de atributos y sus valores límite.

Con el tiempo, en el sector de la construcción, debido a la naturaleza cerrada de los datos, están surgiendo cada vez más enfoques y técnicas nuevos para controlar y gestionar esta diversidad de datos, aunque los datos en los proyectos de construcción son esencialmente los mismos que en otros campos. Mientras que otras industrias se las han arreglado con éxito con enfoques estandarizados para el tratamiento de datos, la construcción sigue desarrollando nuevos y exclusivos formatos de datos, requisitos de y conceptos de validación.

> Los métodos y herramientas utilizados para recopilar, preparar y analizar datos en la construcción no deben ser fundamentalmente diferentes de los utilizados por los especialistas en otros sectores de la economía.

La industria ha desarrollado un ecosistema terminológico distinto que requiere una reflexión crítica y una reevaluación:

- El formato STEP se posiciona bajo el nuevo nombre IFC, complementado con la categorización de construcción, sin tener en cuenta las limitaciones del propio formato STEP.
- El formato paramétrico IFC se utiliza en los procesos de comunicación de datos a pesar de la falta de un núcleo geométrico unificado necesario para la visualización y el cálculo.
- El acceso a las bases de datos CAD -systems se promociona bajo el término "BIM ", sin que se hable de las particularidades de estas bases de datos ni del acceso a ellas.
- Los vendedores promueven la interoperabilidad a través de los formatos IFC y USD, a menudo sin ponerlos en práctica, recurriendo a una costosa ingeniería inversa que a ellos mismos les ha costado.
- Los términos LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD se utilizan universalmente para describir los mismos parámetros de entidad, sin referencia a herramientas de modelización y verificación utilizadas desde hace tiempo en otros sectores.

El sector de la construcción demuestra que todo lo anterior, aunque suene extraño, es posible en la industria de la construcción, sobre todo si el objetivo principal es monetizar cada etapa del procesamiento de datos mediante la venta de servicios y software especializados. Desde un punto de vista empresarial, no hay nada malo en ello. Sin embargo, sigue siendo una incógnita si estos acrónimos y enfoques relacionados con CAD (BIM) realmente añaden valor y simplifican los procesos profesionales.

> En el sector de la construcción, este sistema funciona porque el propio sector obtiene la mayor parte de sus beneficios especulativos en este laberinto de sistemas y siglas. Las empresas interesadas en procesos transparentes y datos abiertos son escasas. Es probable que esta compleja situación se prolongue indefinidamente, hasta que los clientes, inversores, bancos y fondos de capital riesgo empiecen a exigir enfoques más claros e informados de la gestión de la información.

El sector ha acumulado un número excesivo de siglas, pero describen los mismos procesos y requisitos de datos en distintos grados. Su utilidad real para simplificar los flujos de trabajo sigue siendo cuestionable.

Aunque los conceptos y las siglas de marketing van y vienen, los propios procesos de validación de requisitos de datos seguirán siendo para siempre parte integrante de los procesos empresariales. En lugar de creando formatos y normativas cada vez más especializados, el sector de la construcción debería recurrir a herramientas que ya han demostrado su eficacia en otros ámbitos como las finanzas, la industria y la informática.

La abundancia de términos, acrónimos y formatos crea la ilusión de procesos de construcción digital profundamente elaborados. Sin embargo, los conceptos de marketing y la compleja terminología ocultan a menudo una verdad simple pero incómoda: los datos siguen siendo de difícil acceso, mal documentados y rígidamente ligados a soluciones de software específicas.

Para salir de este círculo vicioso de siglas y formatos por formatos, es necesario considerar los sistemas CAD (BIM) no como herramientas mágicas de gestión de la información, sino como lo que realmente son: bases de datos especializadas. Y es a través de este prisma como se puede entender dónde acaba el marketing y empieza el verdadero trabajo con la información.

CAPÍTULO 6.4.
PARAMETRIZACIÓN DEL DISEÑO Y UTILIZACIÓN DE LLM PARA EL FUNCIONAMIENTO DE CAD

La ilusión de unicidad de los datos CAD (BIM): el camino hacia la analítica y los formatos abiertos

Las modernas plataformas CAD (BIM) han transformado considerablemente el planteamiento de la gestión de la información sobre diseño y construcción. Mientras que antes estas herramientas se utilizaban principalmente para crear planos y modelos 3D, hoy funcionan como auténticos repositorios de datos del proyecto. Bajo el concepto de fuente única de la verdad, el modelo paramétrico se está convirtiendo cada vez más en la principal y a menudo única fuente de información del proyecto, garantizando su integridad y relevancia a lo largo de todo el ciclo de vida del proyecto.

> La diferencia clave entre las plataformas CAD - (BIM -) y otros sistemas de gestión de datos de construcción es la necesidad de herramientas y API especializadas para acceder a la información (la única fuente de verdad). Estas bases de datos no son universales en el sentido tradicional: en lugar de una estructura abierta y una integración flexible, son un entorno cerrado, vinculado a una plataforma y un formato específicos.

A pesar de la complejidad de trabajar con datos CAD hay una cuestión más importante que va más allá de la realización técnica: ¿qué son realmente las bases de datos CAD (BIM)? Para responder a esta pregunta, es necesario ir más allá de los acrónimos y conceptos habituales impuestos por los desarrolladores de software. En su lugar, conviene centrarse en la esencia del trabajo con la información del proyecto: los datos y su tratamiento.

> El proceso empresarial en la construcción no empieza con el trabajo en herramientas CAD -o BIM -, sino con la formación de los requisitos del proyecto y la modelización de datos. En primer lugar, se definen los parámetros de la tarea: la lista de entidades, sus características iniciales y los valores límite que deben tenerse en cuenta al resolver una tarea concreta. Sólo entonces se crean modelos y elementos en los sistemas CAD (BIM) a partir de los parámetros especificados

El proceso que precede a la creación de información en las bases de datos CAD - (BIM-) es completamente igual al proceso de modelado de datos que se trató en detalle en la cuarta parte del libro y en el capítulo "Modelado de datos: modelo conceptual, lógico y físico" (Fig. 4.3-1).

Al igual que en el modelado de datos creamos requisitos para los datos que más tarde queremos procesar en la base de datos, para las bases de datos CAD los gestores crean requisitos de diseño en forma de varias columnas de tabla o listas de pares clave-valor (Fig. 6.4-1, pasos 1-2). Y sólo sobre la base de estos parámetros iniciales utilizando la API automática o manualmente, el diseñador crea (o más bien refina)

objetos en bases de datos CAD- (BIM) (pasos 3-4), después de lo cual se comprueba de nuevo el cumplimiento de los requisitos iniciales (pasos 5-6). Este proceso -definición→ creación→ validación→ ajuste (pasos 2-6)- se repite iterativamente hasta que la calidad de los datos, al igual que en el modelado de datos, alcanza el nivel deseado para el sistema de destino -documentos, tablas o cuadros de mando (paso 7).

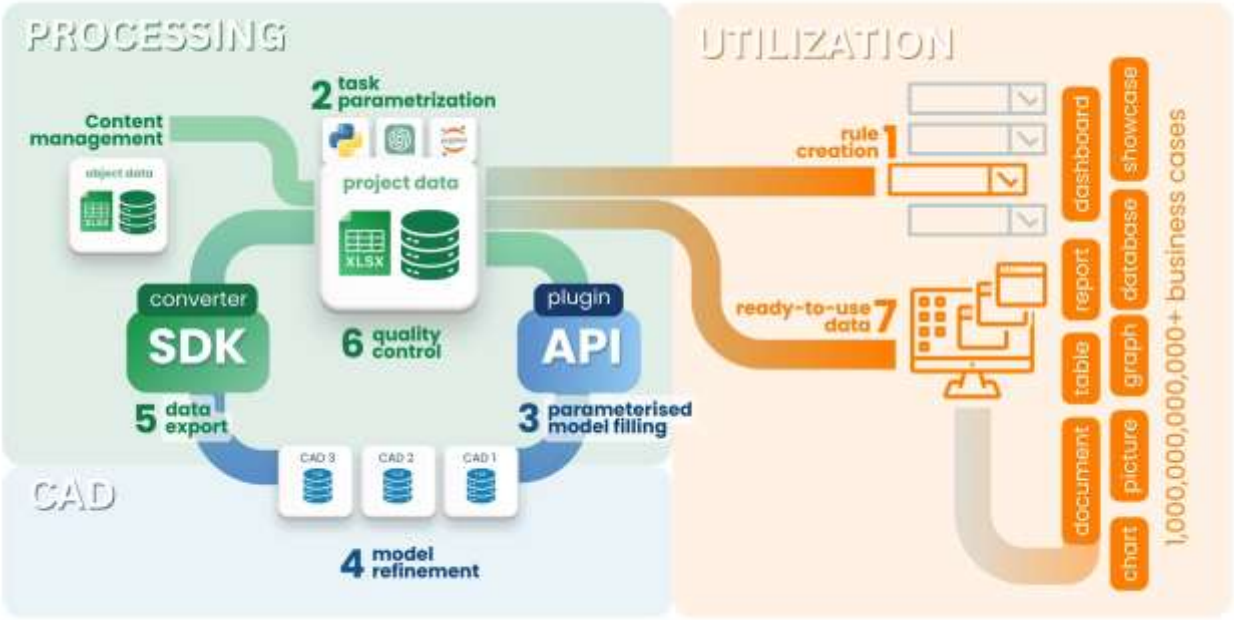


Fig. 6.4-1 El ciclo de saturación de información de las bases de datos para los procesos empresariales en la ejecución de proyectos de construcción.

Si consideramos CAD (BIM) como un mecanismo de transferencia de parámetros en forma de un conjunto de pares clave-valor generados a partir de requisitos definidos fuera del entorno de diseño (Fig. 6.4-1, pasos 1-2), el centro del debate se desplaza de las soluciones de software específicas y sus limitaciones a aspectos más fundamentales: estructura de datos, modelos de datos y requisitos de datos. En esencia, estamos hablando de la saturación de parámetros de la base de datos y del proceso clásico de modelado de datos (pasos 2-3 y 5-6). La única diferencia es que, debido a la naturaleza cerrada de las bases de datos CAD y a las peculiaridades de los formatos utilizados, este proceso va acompañado del uso de herramientas BIM especializadas. La cuestión que se plantea es la siguiente: ¿cuál es la singularidad de BIM si no existen enfoques similares en otros sectores?

En los últimos 20 años, BIM se ha posicionado como algo más que una simple fuente de datos. El paquete CAD -BIM se comercializa a menudo como una herramienta paramétrica con una base de datos inherentemente integrada [64], capaz de automatizar los procesos de diseño, modelado y gestión del ciclo de vida de los proyectos de construcción. Sin embargo, en realidad, el BIM se ha convertido más en una herramienta para mantener a los usuarios en la plataforma de los proveedores que en un método cómodo de gestión de datos y procesos.

Como resultado, los datos CAD- (BIM-) quedan aislados dentro de sus plataformas, ocultando la información del proyecto tras API y kernels de geometría propietarios. Esto ha privado a los usuarios de la capaci-

dad de acceder de forma independiente a las bases de datos de y extraer, analizar, automatizar y transferir datos a otros sistemas, pasando por alto los ecosistemas de los proveedores.

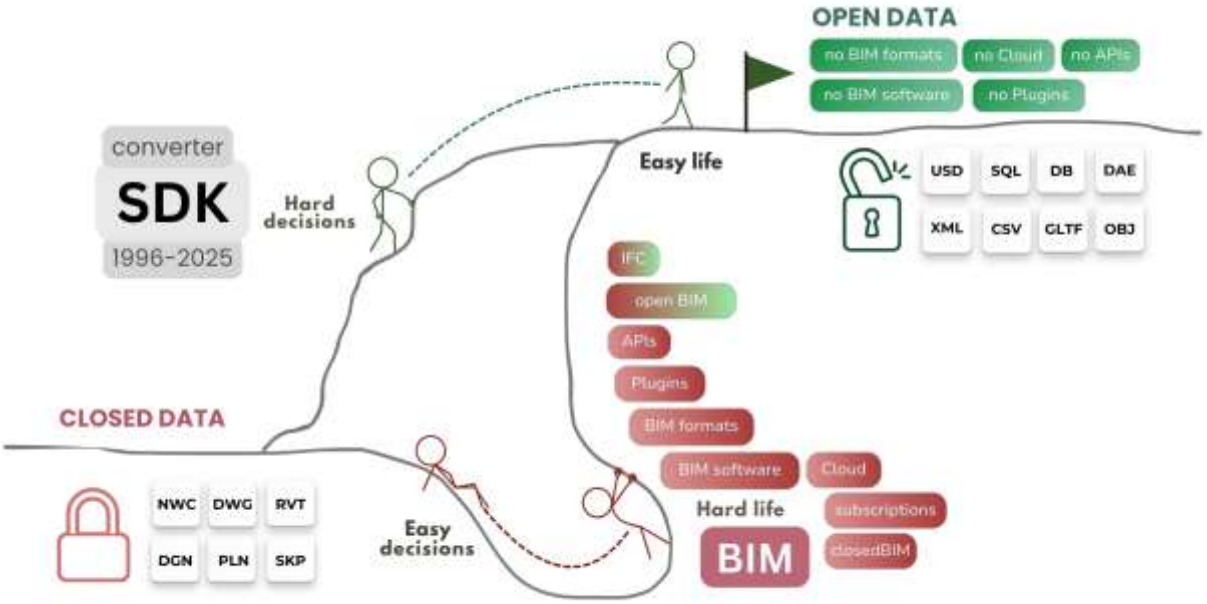


Fig. 6.4-2 En la construcción, los formatos modernos requieren sofisticados núcleos geométricos, una API actualizada anualmente y licencias especiales para el software CAD - (BIM -).

Las empresas que trabajan con herramientas CAD modernas deberían utilizar el mismo enfoque para trabajar con datos que todos los propios vendedores CAD sin excepción utilizan en la práctica: transformación de datos utilizando SDK - herramientas de ingeniería inversa, contra las que los vendedores CAD llevan luchando desde 1995 [75]. Teniendo acceso completo a la base de datos CAD y utilizando herramientas de ingeniería inversa, podemos obtener [127] un conjunto plano de entidades con atributos y exportarlos a cualquier formato abierto conveniente (Fig. 6.4-2), incluyendo tanto la geometría como los parámetros de los elementos de diseño. Este enfoque cambia fundamentalmente el paradigma de trabajo con la información: de una arquitectura orientada a archivos a una arquitectura centrada en datos:

- Formatos de datos como RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET y otros contienen información idéntica sobre elementos de un mismo proyecto. Esto significa que el conocimiento de un formato concreto y de su esquema no debería ser un obstáculo para trabajar con los datos en sí.
- Los datos de cualquier formato pueden combinarse en una única estructura abierta, estructurada y granular (Fig, 9.1-10) que contiene la geometría triangular MESH y las propiedades de todas las entidades objeto, sin las restricciones de los núcleos geométricos.
- El análisis de datos aspira a la universalidad: utilizando datos abiertos, puede trabajar con los datos del proyecto independientemente del formato utilizado.
- Minimización de la dependencia de las API de y de los plugins de proveedores: trabajar con datos

ya no depende de los conocimientos de API.

Cuando y CAD -los requisitos de datos se transforman en formatos de representación estructurados fáciles de analizar- los desarrolladores dejan de depender de esquemas de datos específicos y ecosistemas cerrados.

Diseño mediante parámetros: el futuro de CAD y BIM

Ningún proyecto de construcción en el mundo ha empezado nunca en un programa de CAD. Antes de que un dibujo o modelo tome forma en CAD, pasa por la etapa de conceptualización (Fig. 6.4-1, etapas 1-2), en la que la atención se centra en los parámetros que definen la idea básica y la lógica del futuro objeto. Esta etapa corresponde al nivel conceptual en el modelado de datos (Fig. 4.3-6). Los parámetros pueden existir únicamente en la mente del diseñador, pero lo ideal es que estén dispuestos en forma de listas estructuradas, tablas o almacenados en bases de datos (Fig. 6.4-3), lo que permite transparencia, reproducibilidad y una mayor automatización del proceso de diseño.

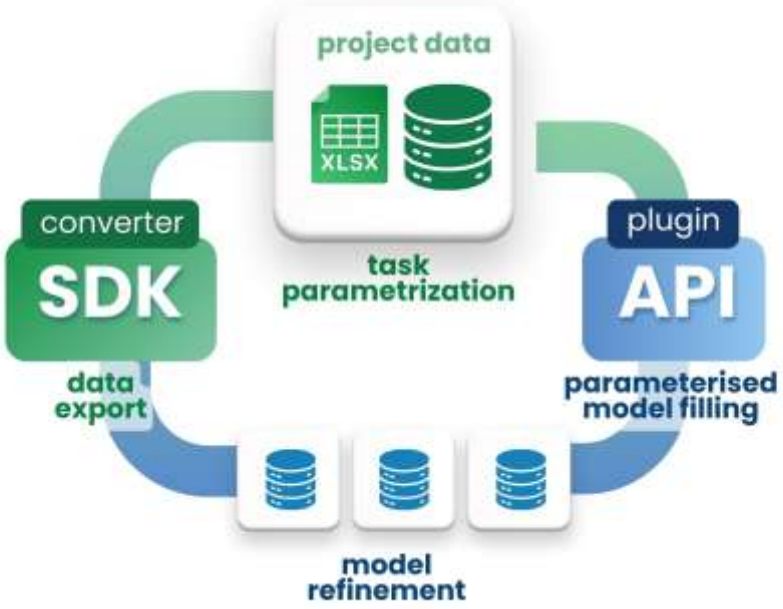


Fig. 6.4-3 El proceso de diseño es un proceso iterativo de poblar la base de datos CAD con información del exterior utilizando los requisitos de la cadena de valor.

Antes de iniciar el modelado CAD propiamente dicho (la fase lógica y física del modelado de datos (Fig, 4.3-7)), es importante definir los parámetros límite que sirven de base al proyecto. Estos atributos, al igual que ocurre con otros requisitos, se recogen desde el final de la cadena de utilización de los datos (por ejemplo, los sistemas) y a través de ellos ya se definen las limitaciones, los objetivos y las características clave de los futuros objetos del proyecto.

La propia modelización puede automatizarse por completo en un 60-100% con ayuda de herramientas de modelización paramétrica (Fig. 6.4-3), si los requisitos están bien definidos. Una vez descrito el proyecto en forma de parámetros, su formación se hace técnicamente factible, por ejemplo, con la ayuda de lengua-

jes de programación visual como Grasshopper Dynamo, integrados en los modernos entornos CAD o soluciones gratuitas en Blender, UE, Omniverse.

Fig. 6.4-4 La mayoría de los proyectos tipificados ya se crean hoy de forma totalmente automática gracias a las herramientas de programación paramétrica.

Ya hoy, los grandes proyectos industriales y tipificados no se crean de la mano del departamento de diseño, sino mediante herramientas paramétricas y programación visual. Esto permite construir un modelo basado en datos y no en las decisiones subjetivas de un diseñador o un gestor concreto.

> *El contenido precede al diseño. El diseño sin contenido no es diseño, sino decoración [128].*
>
> - Jeffrey Zeldman, diseñador web y empresario

El proceso no empieza con el dibujo o el modelado 3D, sino con la formación de requisitos. Son los requisitos los que determinan qué elementos se utilizarán en el proyecto, qué datos hay que transferir a otros departamentos y sistemas. Sólo la existencia de requisitos estructurados permite comprobar automáticamente los modelos de forma periódica (por ejemplo, incluso cada 10 minutos sin distraer al diseñador de su trabajo).

> Quizá en el futuro el sistema CAD- (BIM-) se convierta en una mera interfaz para rellenar la base de datos, y no importe en qué herramienta CAD realice el modelado (nivel físico).

Del mismo modo, en ingeniería mecánica, la modelización 3D se utiliza a menudo, pero no es un elemento necesario u obligatorio del proyecto. En la mayoría de los casos, la documentación clásica en 2D es suficiente y sobre su base se crea el modelo de información necesario. Este modelo se ensambla a partir de

componentes estructurados según normas industriales y contiene toda la información necesaria para comprender el diseño y la organización de la producción. A continuación, se crea un modelo de información de la fábrica, al que se añaden productos específicos y diagramas de flujo ya orientados a las necesidades de los tecnólogos. Todo el proceso puede organizarse sin complejidades innecesarias, sin sobrecargar el sistema con gráficos 3D cuando no aportan ventajas reales.

> Es importante entender que el propio modelo 3D y el sistema CAD no deben desempeñar el papel principal, sino que son sólo una herramienta para el análisis cuantitativo y geométrico. Todos los demás parámetros, excepto la geometría, que describen la entidad, deben almacenarse y procesarse fuera del entorno CAD si es posible (BIM).

El diseño mediante parámetros no es sólo una tendencia, sino el futuro inevitable del sector de la construcción. En lugar de crear manualmente complejos modelos 3D, los diseñadores trabajarán con datos, los validarán y automatizarán los procesos, acercando la construcción al mundo de la programación. Con el tiempo, los procesos de diseño se basarán en los principios del desarrollo de software:

- Crear requisitos → Crear modelo→ Subir al servidor→ Validar cambios→ Pull request
- Las solicitudes Pull ejecutan automáticamente comprobaciones del modelo con respecto a los requisitos que se crearon antes o durante el diseño como parte de la solicitud Pull.
- Tras la comprobación de la calidad de los datos y su aprobación, los cambios se implementan en el proyecto, en la base de datos común o se transfieren automáticamente a otros sistemas.

Ya en la ingeniería mecánica, los cambios de diseño se inician con la creación de un aviso de cambio. A la industria de la construcción le espera un esquema similar: el diseño será un proceso iterativo en el que cada paso se apoyará en requisitos paramétricos. Un sistema de este tipo permitirá a los diseñadores crear comprobaciones automatizadas y pull request automatizadas para requisitos específicos.

> El diseñador del futuro es ante todo un operador de datos, no un modelador manual. Su tarea consiste en llenar el proyecto de entidades paramétricas, donde la geometría es solo uno de los atributos.

La comprensión de la importancia del modelado de datos, la clasificación y la normalización, que se han tratado en detalle en los capítulos anteriores del libro, es lo que desempeñará un papel importante en la transformación. Las normas de diseño del futuro se formalizarán como pares de parámetros clave-valor en forma de esquemas XLSX o XML.

> El futuro del sector de la construcción pasa por recopilar datos, analizarlos, validarlos y automatizar los procesos mediante herramientas analíticas. BIM (o CAD) no es el objetivo final, sino solo una etapa de la evolución. Cuando los profesionales se den cuenta de que pueden trabajar directamente con los datos, prescindiendo de las herramientas CAD tradicionales, el propio término "BIM" dará paso gradualmente a los conceptos de utilización de datos estructurados y granulares de los proyectos de construcción.

Uno de los factores clave que han acelerado la transformación ha sido la aparición de los grandes modelos

lingüísticos (LLM) y las herramientas basadas en ellos. Estas tecnologías están cambiando el modo en que se manejan los datos de diseño, permitiendo acceder a la información sin necesidad de conocer en profundidad las API de o las soluciones de los proveedores. Con los LLM, el proceso de crear un requisito e interactuar con datos CAD se vuelve intuitivo y accesible.

Aparición de LLM en los procesos de tratamiento de datos de diseño CAD

Además del desarrollo de herramientas de acceso a bases de datos CAD y de formatos CAD abiertos y simplificados, la aparición de herramientas LLM (Large Language Models) está revolucionando el tratamiento de los datos de diseño. Mientras que en el pasado el acceso a la información se realizaba principalmente a través de interfaces complejas y requería conocimientos de programación y de API, ahora es posible interactuar con los datos utilizando el lenguaje natural.

Los ingenieros, gestores y planificadores sin formación técnica pueden obtener la información necesaria a partir de los datos del proyecto formulando consultas en lenguaje corriente. Siempre que los datos estén estructurados y sean accesibles (Fig. 4.1-13), basta con formular una pregunta en LLM chat del tipo: "*Mostrar en una tabla con agrupación por tipo todos los muros con un volumen superior a 10 metros cúbicos*" - y el modelo convertirá automáticamente esta consulta en SQL o en código en Pandas, generando una tabla resumen, un gráfico o incluso un documento acabado.

A continuación se muestran algunos ejemplos reales de cómo los modelos LLM interactúan con los datos de diseño representados en diferentes formatos CAD- (BIM-).

- Ejemplo de una consulta en LLM chat a un proyecto CAD en formato RVT tras su conversión (Fig. 4.1-13) a un marco de datos tabular (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otro):

> Agrupe los datos en Dataframe obtenidos del archivo RVT por "Nombre de tipo" al sumar el parámetro "Volumen" y muestre el número de elementos en el grupo. Y por favor muestre todo esto como un histograma horizontal sin valores cero.

Respuesta LLM en forma de gráfico de barras horizontales (formato PNG):

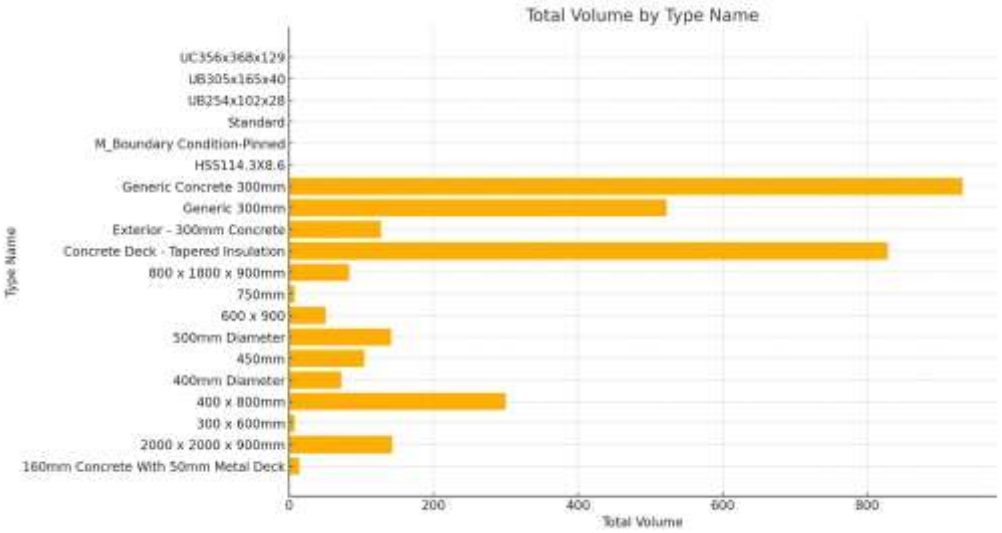


Fig. 6.4-5 En lugar de 17 clics de ratón o 40 líneas de código mediante el uso de plug-ins, en LLM recuperamos instantáneamente la tabla QTO mediante una consulta de texto.

Para generar una tabla QTO de tipos de muros con superficie total y cantidad a partir de la categoría "Muros" formulemos una consulta de texto para LLM -chat:

> Tome sólo los elementos del marco de datos del proyecto que tengan "OST_Walls" en el parámetro "Category", agrúpelos por "Type Name", sume el valor de la columna "Area", añada la cantidad y muéstrelos en la tabla eliminando los valores nulos.

Respuesta LLM en forma de tabla QTO acabada:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Fig. 6.4-6 La creación de una tabla QTO en lenguaje natural proporciona el mismo resultado de calidad que cuando se utilizan herramientas CAD - (BIM-).

Consultemos el proyecto en formato IFC después de convertirlo en un marco de datos tabular e introduzcamos una consulta de texto similar en cualquier chat LLM:

> Tome sólo aquellos elementos del proyecto que tengan valores de Nivel 1 y Nivel 2 en el parámetro "Padre", y tome los elementos que tengan valores de IfcSlab en el parámetro "Categoría", luego agrupe estos elementos por el parámetro "Tipo de objeto", sume los valores en el parámetro "PSet_RVT _Dimensiones Área", y muéstrelos como un gráfico circular.

- Respuesta LLM como gráfico circular acabado de grupos de elementos a partir de datos IFC:

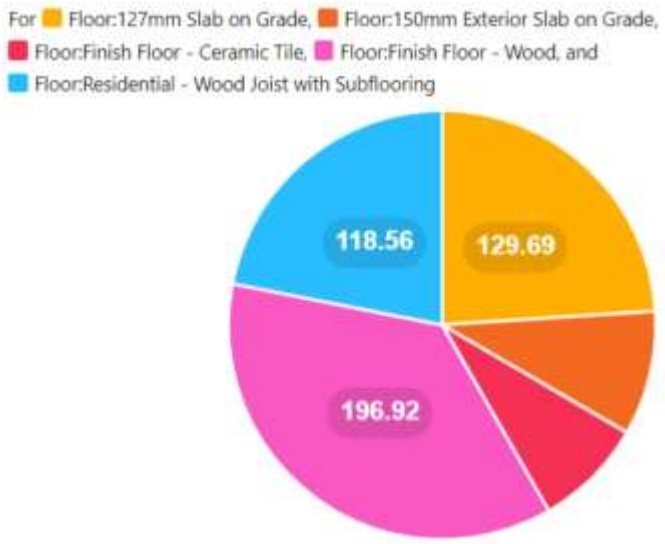


Fig. 6.4-7 El resultado de una consulta de datos IFC en un formato estructurado puede ser cualquier tipo de gráfico que resulte conveniente para comprender los datos.

Detrás de cada una de las soluciones prefabricadas resultantes (Fig. 6.4-5 - Fig. 6.4-7) hay una docena de líneas de código Python utilizando la biblioteca Pandas. El código resultante puede copiarse de la sala de chat de LLM y utilizarse en cualquier IDE local o en línea para obtener resultados idénticos fuera de la sala de chat de LLM.

En el mismo chat LLM podemos trabajar no sólo con proyectos obtenidos a partir de formatos CAD 3D (BIM, sino también con dibujos planos en formato DWG, a los que podemos consultar en el chat LLM para visualizar, por ejemplo, datos sobre grupos de elementos en forma de líneas o geometrías 3D tras su conversión a un formato estructurado.

Análisis automatizado de archivos DWG con LLM y Pandas

El proceso de tratamiento de datos a partir de archivos DWG -debido a la naturaleza no estructurada de la información- siempre ha sido una tarea compleja, que ha requerido software especializado y, a menudo, análisis manuales. Sin embargo, con el desarrollo de la inteligencia artificial y las herramientas LLM, se ha hecho posible automatizar muchos pasos, este, hoy en día, proceso mayoritariamente manual. Consideremos una tubería real de solicitudes a LLM (en este ejemplo ChatGPT) para trabajar con dibujos DWG, que le permiten trabajar con el proyecto:

- Filtrar datos DWG por capa, ID y coordenadas
- Visualizar la geometría de los elementos
- Anotación automática de dibujos en función de parámetros

- Expandir polilíneas de muro al plano horizontal
- Cree visualizaciones interactivas en 3D de datos planos en
- Estructure y analice los datos de construcción sin complejas herramientas CAD -tools

En nuestro caso, el proceso de construcción de Pipeline comienza con la generación de código secuencial a través del LLM. En primer lugar, se genera una consulta que describe la tarea. ChatGPT genera código Python, que se ejecuta y analiza, mostrando el resultado dentro de la sala de chat. Si el resultado no es el esperado, se corrige la consulta y se repite el proceso

> **Pipeline** es una secuencia de pasos automatizados para procesar y analizar datos. En un proceso de este tipo, cada paso toma datos como entrada, realiza transformaciones y pasa el resultado al siguiente paso.

Tras obtener el resultado deseado, el código se copia desde LLM y se pega en el código en forma de bloques en cualquiera de los IDEs convenientes, en nuestro caso en la plataforma Kaggle.com. Los fragmentos de código resultantes se combinan en un único Pipeline, que automatiza todo el proceso, desde la carga de datos hasta su análisis final. Este enfoque permite un rápido desarrollo y escalado de los procesos analíticos sin necesidad de profundos conocimientos de programación. El código completo de todos los fragmentos que se muestran a continuación, junto con ejemplos de consultas, se puede encontrar en la plataforma Kaggle.com buscando "DWG Analyse with ChatGPT | DataDrivenConstruction" [129].

Vamos a empezar el proceso de trabajo con datos DWG, después de la conversión en forma estructurada (Fig. 4.1-13), con un paso clásico - agrupación y filtrado de todos los datos de dibujo, necesarios para nuestra tarea elementos de pared, específicamente polilíneas (parámetro 'ParentID' permite agrupar líneas en grupos), que en el parámetro (columna dataframe) "Layer" tiene un valor de cadena que contiene la siguiente combinación de letras (RegEx) - "wall".

- Para obtener el código de una tarea similar y el resultado en forma de imagen debe escribir la siguiente consulta en LLM:

> En primer lugar, compruebe si el marco de datos obtenido de DWG contiene las columnas definidas: 'Capa', 'ID', 'ParentID' y 'Punto'. A continuación, filtre los identificadores de la columna 'Layer' que contengan la cadena 'wall'. Busque los elementos de la columna 'ParentID' que coincidan con estos identificadores. Defina una función para limpiar y dividir los datos de la columna "Punto". Esto incluye eliminar los paréntesis y dividir los valores en coordenadas "x", "y" y "z". Trazar los datos utilizando matplotlib. Para cada "ParentID" único, dibuje una polilínea separada que conecte las coordenadas del "Punto". Asegúrate de que el primer y el último punto están conectados si es posible. Establece las etiquetas y los títulos apropiados, asegurándote de que los ejes x e y están igualmente escalados.

- La respuesta LLM le dará una imagen ya hecha detrás de la cual se esconde el código Python que la ha generado:

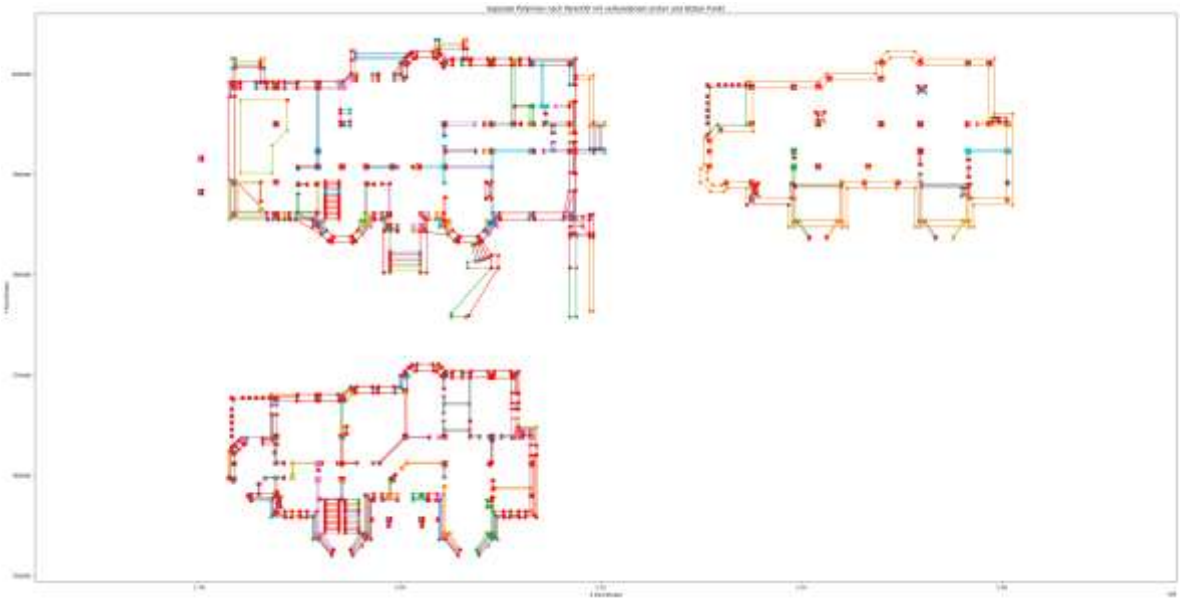


Fig. 6.4-8 El código LLM extrajo todas las líneas de la capa "muro" del archivo DWG, borró sus coordenadas y construyó las polilíneas utilizando una de las librerías Python.

Ahora vamos a añadir a las líneas el parámetro de área que cada polilínea tiene en sus propiedades (en una de las columnas del dataframe):

> Ahora obtenga sólo un "ParentID" de cada polilínea - encuentre este ID en la columna "ID", tome el valor "Area", divídalo por 1.000.000 y añada este valor al gráfico

La respuesta LLM mostrará un nuevo gráfico donde cada polilínea tendrá una leyenda con su área:

Fig. 6.4-9 LLM ha añadido un código que toma los valores de área de cada polilínea y los añade a la imagen con visualización de líneas.

A continuación, convertiremos cada polilínea en una línea horizontal, añadiremos una línea paralela a una altura de 3000 mm y las conectaremos en un único plano para mostrar de este modo la disposición de las superficies de los elementos del muro:

> Necesita tomar todos los elementos de la columna "Layer" con el valor "wall". Tome estos IDs como una lista de la columna "ID" y encuentre estos IDs de todo el marco de datos en la columna "ParentID". Todos los elementos son líneas que se combinan en una única polilínea. Cada línea tiene una geometría x, y diferente del primer punto en la columna "Punto". Debe tomar cada polilínea a su vez y desde el punto 0,0 horizontalmente dibujar la longitud de cada segmento de la polilínea. la longitud de cada segmento de la polilínea en una línea. Luego dibuje exactamente las mismas líneas sólo que 3000 más arriba, conecte todos los puntos en un plano.

La respuesta LLM emitirá un código que le permitirá trazar dibujos de muros en el plano:

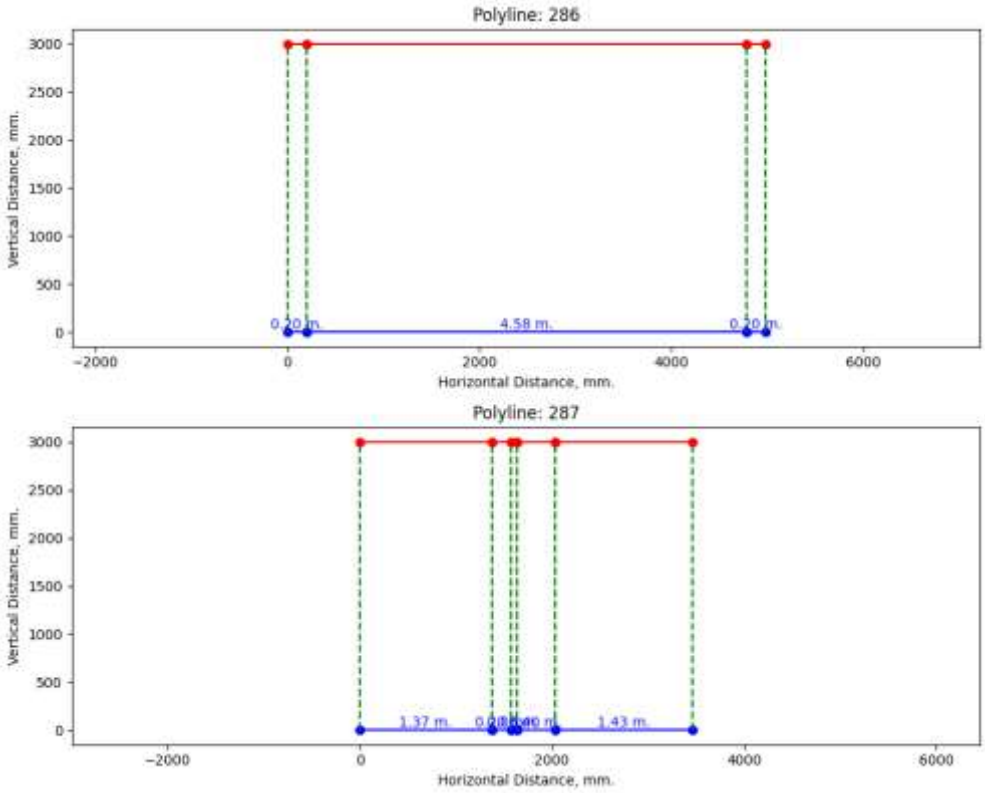


Fig. 6.4-10 Convertimos cada polilínea que utiliza indicaciones en un trazado que visualiza los planos de la pared directamente en el chat LLM.

Ahora pasemos de la proyección 2D a la 3D -modelando muros a partir de líneas planas conectando las capas superior e inferior de polilíneas:

> Visualizar los elementos de las paredes en 3D, conectando polilíneas a las alturas z = 0 y z = 3000 mm. Crear una geometría cerrada que represente los muros del edificio. Utilizar la herramienta de gráficos 3D Matplotlib.

- LLM generará un gráfico 3D interactivo en el que cada polilínea se representará como un conjunto de planos. El usuario podrá moverse libremente entre los elementos con un ratón de ordenador, explorando el modelo en modo 3D copiando el código del chat al IDE:

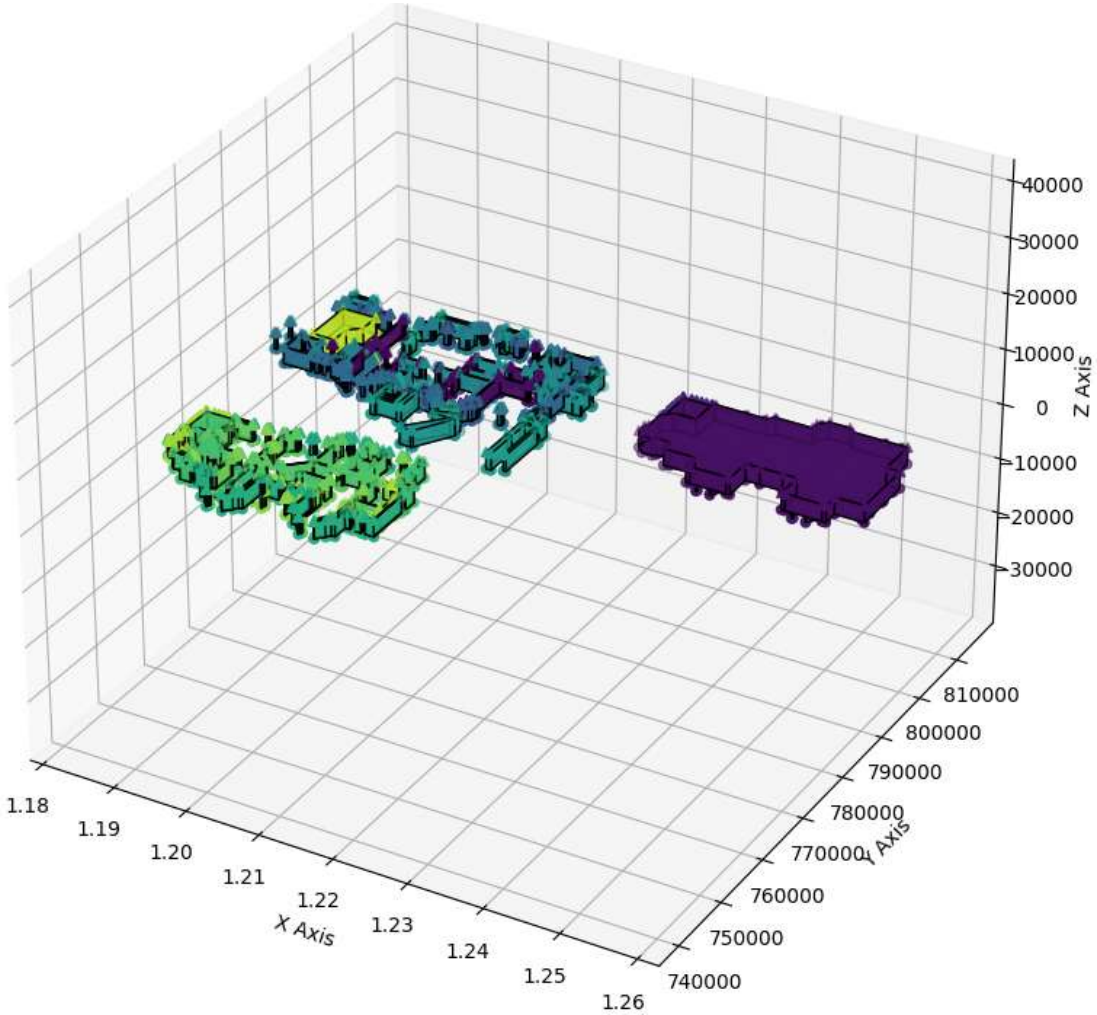


Fig. 6.4-11 LLM ayudó a construir código [129] para visualizar líneas de dibujo planas en una vista 3D que puede ser explorada en el visor 3D dentro del IDE.

Para construir un Pipeline lógico y reproducible - desde la conversión inicial y la carga del archivo DWG - hasta el resultado final - se recomienda copiar el bloque de código LLM -generado en el IDE después de cada paso. De esta manera, no sólo comprueba el resultado en el chat, sino que también lo ejecuta en su entorno de desarrollo inmediatamente. Esto le permite construir el proceso secuencialmente, depurándolo y adaptándolo según sea necesario.

Puede encontrar el código completo de Pipeline de todos los fragmentos (Fig,s 6.4-8 a Figuras 6.4-11) junto con ejemplos de consultas en la plataforma Kaggle.com buscando "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. En Kaggle no sólo se puede ver el código y las indicaciones utilizadas, sino también copiar y probar todo el Pipeline con los marcos de datos originales DWG en la nube de

forma gratuita sin tener que instalar ningún software adicional o el propio IDE.

El enfoque presentado en este capítulo permite automatizar completamente la comprobación, el procesamiento y la generación de documentos basados en proyectos DWG. El Pipeline desarrollado es adecuado tanto para el procesamiento de dibujos individuales como para el procesamiento por lotes de decenas, cientos y miles de archivos DWG con generación automática de los informes y visualizaciones necesarios para cada proyecto.

El proceso puede organizarse de forma secuencial y transparente: primero, los datos del archivo CAD se convierten automáticamente al formato XLSX, luego se cargan en un marco de datos, a lo que sigue la agrupación, la comprobación y la generación de resultados; todo ello se implementa en un único cuaderno Jupyter o en un script Python, en cualquier IDE popular. Si es necesario, el proceso puede ampliarse fácilmente mediante la integración con sistemas de gestión de documentación de proyectos: los archivos CAD pueden recuperarse automáticamente según criterios especificados, los resultados pueden devolverse al sistema de almacenamiento y los usuarios pueden recibir una notificación cuando los resultados estén listos, por correo electrónico o mensajería.

> El uso de los chats y agentes de LLM para trabajar con datos de diseño reduce la dependencia de programas especializados de CAD -y permite realizar análisis y visualizaciones de diseños arquitectónicos sin necesidad de interactuar manualmente con la interfaz -sin clics de ratón y sin tener que recordar complejos menús de navegación.

Cada día que pasa, el sector de la construcción oirá hablar más de LLM, datos estructurados granulares, DataFrames y bases de datos columnares. Los DataFrames bidimensionales unificados formados a partir de diversas bases de datos y formatos CAD, serán el combustible ideal para las modernas herramientas analíticas que manejan activamente los especialistas de otras industrias.

El propio proceso de automatización se simplificará significativamente: en lugar de estudiar la API de productos de nicho cerrados y escribir complejos scripts para analizar o transformar parámetros, ahora bastará con formular una tarea en forma de conjunto de comandos de texto individuales, que se plegarán en el Pipeline o Workflow-process necesario para el lenguaje de programación requerido, que se ejecuta gratuitamente en casi cualquier dispositivo. Ya no habrá que esperar a nuevos productos, formatos, complementos o actualizaciones de los proveedores de herramientas CAD- (BIM-). Los ingenieros y constructores podrán trabajar de forma independiente con los datos utilizando herramientas sencillas, gratuitas y fáciles de entender, asistidos por chats y agentes de LLM.

Próximos pasos: pasar de los formatos cerrados a los datos abiertos

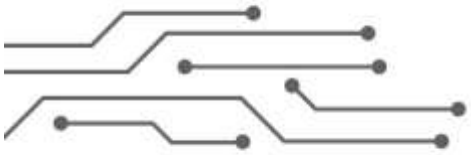
Cuando se trabaje con los datos de diseño del futuro, es poco probable que alguien necesite realmente comprender los núcleos geométricos de las herramientas propietarias o aprender cientos de formatos incompatibles que contienen la misma información. Sin embargo, sin entender por qué es importante el paso a los datos estructurados abiertos, es difícil defender el uso de nuevas herramientas gratuitas, datos abiertos y enfoques que probablemente no promuevan los proveedores de software.

En este capítulo hemos hablado de las características clave de los datos CAD (BIM), de sus limitaciones y

oportunidades, y de que, a pesar de las promesas de marketing de los proveedores, los ingenieros y diseñadores se enfrentan cada día a dificultades para extraer, transferir y analizar la información de diseño. Comprender la arquitectura de estos sistemas y conocer enfoques alternativos -basados en formatos abiertos y en la automatización mediante LLM - puede facilitar mucho la vida incluso a un solo profesional, por no hablar de las empresas. Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar a sus tareas diarias los enfoques tratados en esta sección:

- Amplíe sus herramientas para trabajar con datos de proyectos
  - ☐ Explore los plug-ins y utilidades disponibles para extraer datos de los sistemas CAD - (BIM-) que utilice.
  - ☐ Explore los SDK y API disponibles que le permiten automatizar la extracción de datos de formatos cerrados sin tener que abrir manualmente software especializado.
  - ☐ Dominio de los conocimientos básicos para trabajar con formatos abiertos de geometría no paramétrica (OBJ, glTF, USD, DAE) y las bibliotecas de código abierto correspondientes.
  - ☐ Intentar pensar en un sistema para almacenar los metadatos del proyecto separados de la geometría fuera de las soluciones CAD (BIM) para simplificar el análisis y la integración con otros sistemas.
  - ☐ Utilice LLM para automatizar los problemas de conversión de datos entre formatos.
- Cree sus propios procesos para gestionar la información del proyecto
  - ☐ Empezar a describir tareas y requisitos de modelización mediante parámetros y sus valores en formatos sencillos y estructurados.
  - ☐ Cree una biblioteca personal de secuencias de comandos o bloques de código para las operaciones más frecuentes.
- Promueva el uso de normas abiertas en su trabajo
  - ☐ Invitar a colegas y socios a compartir datos en formatos abiertos que no estén restringidos por el ecosistema de proveedores de software.
  - ☐ Demostrar las ventajas de utilizar datos estructurados con ejemplos concretos.
  - ☐ Iniciar debates sobre los problemas que plantean los formatos cerrados y sus posibles soluciones.

Aunque no pueda cambiar la política de su empresa en relación con las plataformas CAD - (BIM-), una comprensión personal de los principios del trabajo con datos de proyectos en formatos abiertos le permitirá aumentar considerablemente la eficacia de su trabajo. Al crear tus propias herramientas y métodos para extraer y transformar datos de distintos formatos, no sólo optimizarás tus flujos de trabajo, sino que también ganarás en flexibilidad para eludir las limitaciones de las soluciones de software estándar.



VII PARTE

TOMA DE DECISIONES BASADA EN DATOS, ANÁLISIS, AUTOMATIZACIÓN Y APRENDIZAJE AUTOMÁTICO

La séptima parte se centra en el análisis de datos y la automatización de procesos en el sector de la construcción. Se analiza cómo los datos se convierten en la base para la toma de decisiones y se explican los principios de visualización de la información para un análisis eficaz. Se describen con detalle los indicadores clave de rendimiento (KPI), los métodos para evaluar el retorno de la inversión (ROI) y la creación de cuadros de mando para el seguimiento de proyectos. Se presta especial atención a los procesos ETL (Extract, Transform, Load) y su automatización mediante pipelines (Pipeline) para convertir datos dispares en información estructurada para su análisis. Se discuten herramientas de orquestación de flujos de trabajo como Apache Airflow, Apache NiFi y n8n, que permiten construir pipelines de datos automatizados sin profundos conocimientos de programación. Los modelos de grandes lenguajes (LLM) y su uso para simplificar el análisis de datos y automatizar tareas rutinarias están desempeñando un papel importante

CAPÍTULO 7.1.
ANÁLISIS DE DATOS Y TOMA DE DECISIONES BASADA EN DATOS

Tras las fases de recopilación, estructuración, limpieza y verificación de la información, ha surgido un conjunto de datos coherente y analizable. En las partes anteriores del libro se abordó la sistematización y estructuración de fuentes heterogéneas, desde documentos PDF y registros de texto de reuniones hasta modelos CAD y datos geométricos. Se describe con detalle el proceso de comprobación y alineación de la información con los requisitos de diversos sistemas y clasificadores, eliminando duplicados e incoherencias.

Todos los cálculos realizados con estos datos (tercera, cuarta partes del libro) -desde simples transformaciones hasta cálculos de tiempo, coste e indicadores ESG (quinta parte)- son tareas de análisis agregado. Constituyen la base para comprender el estado actual de un proyecto, evaluar sus parámetros y, a continuación, tomar decisiones. Gracias a los cálculos, los datos pasan de ser un conjunto de registros dispares a un recurso manejable capaz de responder a las preguntas clave de la empresa.

En los capítulos anteriores se han detallado los procesos de recopilación de datos y control de calidad para su uso en casos y procesos empresariales típicos específicos del sector de la construcción. La analítica en este contexto es similar en muchos aspectos a las aplicaciones en otras industrias, pero tiene una serie de características específicas.

En los capítulos siguientes se detallará el proceso de análisis de datos, incluidos los pasos de automatización desde la adquisición inicial de información y su transformación hasta su posterior transferencia a sistemas y documentos de destino. En primer lugar, se presentará una parte teórica centrada en determinados aspectos de la toma de decisiones basada en datos. A continuación, en los capítulos siguientes, la parte práctica relacionada con la automatización y construcción de ETL -Pipeline.

Los datos como recurso en la toma de decisiones

La toma de decisiones basada en datos suele ser un proceso iterativo y comienza con la recopilación sistemática de información procedente de diversas fuentes de información. Como en un ciclo natural, los elementos de datos individuales y los sistemas de información enteros caen gradualmente al suelo, acumulándose en los depósitos de información de las empresas (Fig, 1.3-2). Con el tiempo, estos datos, como las hojas y ramas caídas, se transforman en material valioso. El micelio de ingenieros y analistas de datos organiza y prepara la información para su uso futuro y convierte los datos y sistemas caídos en valioso abono, para que crezcan nuevos brotes y nuevos sistemas (Fig, 1.2-5).

Las tendencias en el uso generalizado de la analítica en diversas industrias, marca el inicio de una nueva era en la que trabajar con datos se convierte en la base de la actividad profesional (Fig. 7.1-1). Es importante que los profesionales del sector de la construcción se adapten a estos cambios y estén preparados para entrar en una nueva era: la era de los datos y la analítica

Mover datos manualmente entre tablas y realizar cálculos manualmente se están convirtiendo poco a poco en cosa del pasado, dando paso a la automatización, el análisis del flujo de datos, la analítica y el aprendizaje automático. Estas herramientas se están convirtiendo en elementos clave de los sistemas modernos de apoyo a la toma de decisiones.

En el libro de "Rebooting. Guía de McKinsey para superar la competencia en la era de la tecnología digital y la inteligencia artificial" [130], se cita un estudio realizado en 2022 con 1.330 altos ejecutivos de distintas regiones, industrias y áreas funcionales [130], cita un estudio realizado en 2022 con 1.330 altos ejecutivos de distintas regiones, industrias y áreas funcionales. Según sus resultados, el 70% de los líderes utiliza la analítica avanzada para generar sus propias ideas, y el 50% aplica la inteligencia artificial para mejorar y automatizar los procesos de toma de decisiones.

Fig. 7.1-1 El análisis de datos y la analítica es la principal herramienta para aumentar la velocidad de la toma de decisiones en una empresa.

El análisis de datos, como la propagación del micelio, penetra en el humus de decisiones pasadas, ayudando a conectar sistemas individuales y guiando a los directivos hacia valiosos conocimientos. Estos conocimientos, como los nutrientes de los árboles de sistemas de datos en descomposición, alimentan las nuevas decisiones de la empresa, lo que conduce a un cambio eficaz y a un crecimiento de la información de calidad, como los nuevos brotes y retoños que surgen de un suelo rico y sano (Fig, 1.2-5).

Los números tienen una historia importante que contar. Cuentan contigo para que les des una voz clara y convincente [131].

- Stephen Few, experto en visualización de datos

En las empresas medianas y pequeñas, el trabajo de extracción y preparación de información para su posterior análisis es hoy un proceso extremadamente laborioso (Fig, 7.1-2), comparable a la minería del carbón del siglo XVIII. Hasta hace poco, la labor de extracción y preparación de datos estaba más bien reservada a aventureros que trabajaban en un nicho muy especializado con un conjunto reducido y limitado de herra-

mientas para trabajar con distintos tipos de datos procedentes de fuentes no estructuradas, poco estructuradas, mixtas y cerradas.

> Los responsables de la toma de decisiones y los gestores no suelen tener experiencia con datos y sistemas heterogéneos, y sin embargo necesitan tomar decisiones basadas en ellos. Como resultado, la toma de decisiones basada en datos en la industria de la construcción moderna en las últimas décadas se ha parecido menos a un proceso automatizado y más al trabajo manual de varios días de un minero en las primeras minas de carbón.

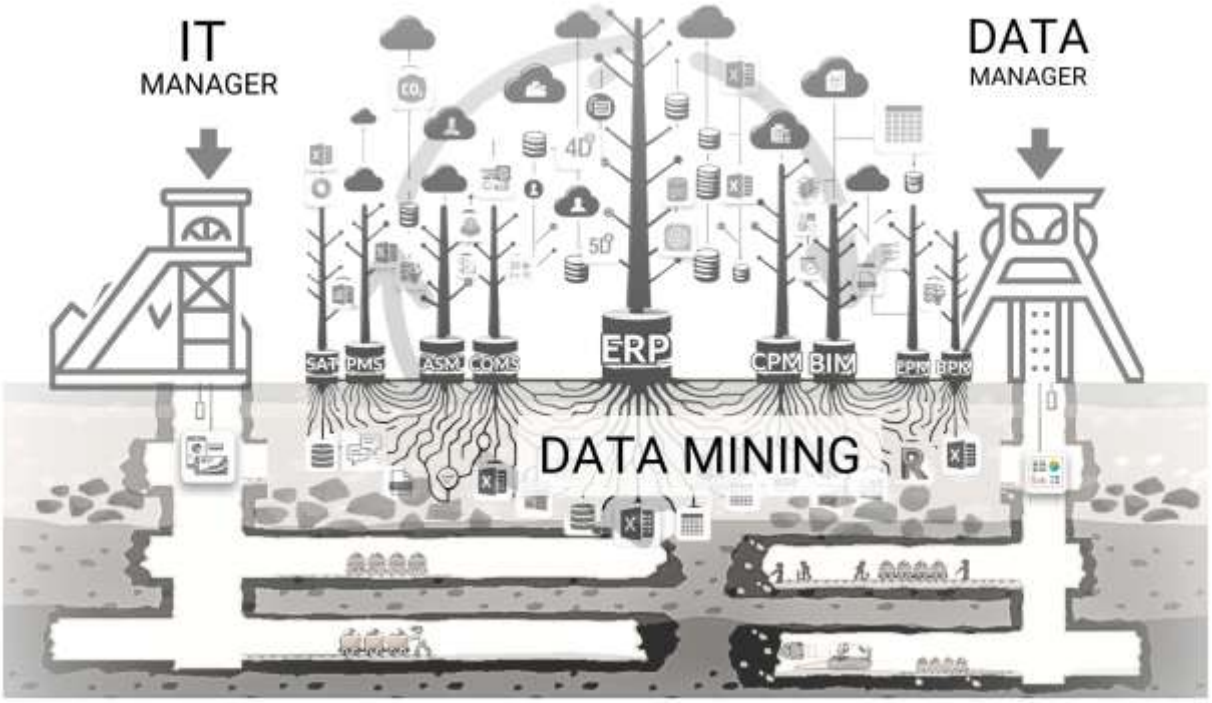


Fig. 7.1-2 En el proceso de minería de datos, los expertos recorren un complejo camino de preparación de los datos, desde la limpieza hasta la estructuración para su posterior análisis.

Aunque los métodos modernos de extracción de datos en el sector de la construcción son sin duda más avanzados que las primitivas técnicas de los mineros del siglo XII, sigue siendo una tarea compleja y de alto riesgo, que requiere importantes recursos y conocimientos especializados que sólo pueden permitirse las grandes empresas. Los procesos de extracción y análisis de datos de a partir del legado acumulado de proyectos pasados han sido realizados hasta hace poco predominantemente por grandes empresas tecnológicamente avanzadas que llevan décadas recopilando y almacenando datos de forma sistemática

> Antes, el protagonismo de la analítica correspondía a empresas tecnológicamente maduras que llevaban décadas acumulando datos. Hoy, la situación está cambiando: el acceso a los datos y a las herramientas de procesamiento de datos se está democratizando: soluciones antes complejas están ahora al alcance de todos de forma gratuita.

La aplicación de la analítica permite a las empresas tomar decisiones más precisas e informadas en tiempo real. El siguiente caso práctico ilustra cómo los datos históricos pueden ayudar a tomar decisiones financieramente sólidas:

- **Jefe de proyecto** - "*Ahora el precio medio del hormigón en la ciudad es de 82€ /m³, tenemos 95 €/m³ en el presupuesto*".
- **Estimador** - "*En proyectos anteriores, el rebasamiento era de un 15%, así que me eché atrás*".
- **Gestor de datos o ingeniero de control del cliente**: "*Veamos los análisis de las tres últimas licitaciones*".

Tras analizar el DataFrame de proyectos anteriores, obtenemos:

- **Precio medio de compra real:** 84,80 /m³€
- **Ratio medio de gasto excesivo:** +4,7%**.**
- **Tasa recomendada en el presupuesto:** ~ 85 /m³€

Esa decisión ya no se basará en sensaciones subjetivas, sino en estadísticas históricas concretas, lo que contribuye a reducir riesgos y aumentar la validez de la oferta de licitación. El análisis de los datos de proyectos anteriores se convierte en una especie de "abono orgánico" del que germinan nuevas soluciones más precisas.

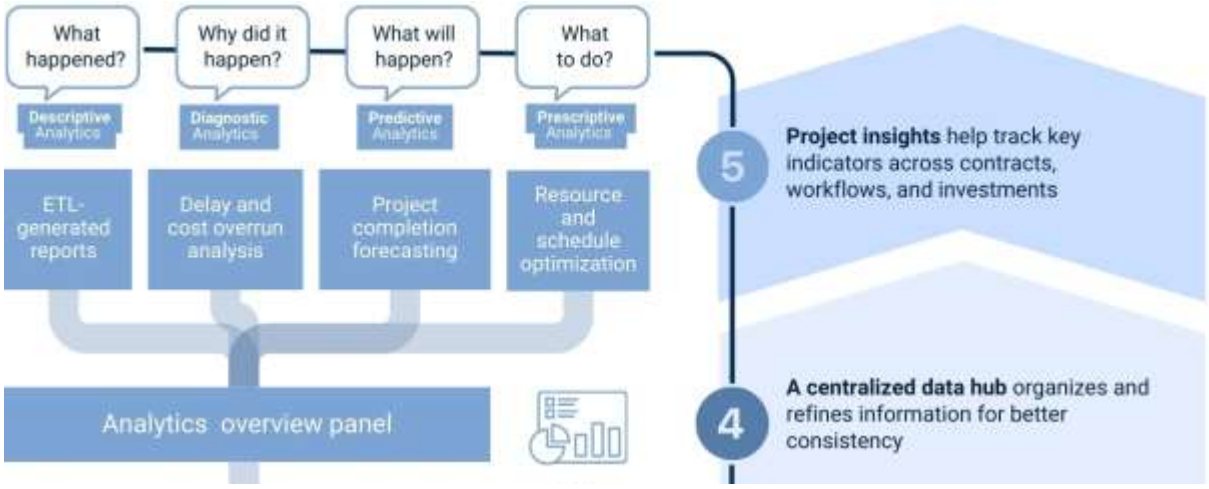


Fig. 7.1-3 El análisis de datos responde a tres preguntas clave: qué ha ocurrido, por qué ha ocurrido y qué debe hacerse a continuación.

Los responsables de la toma de decisiones y los gestores se enfrentan a menudo a la necesidad de trabajar con datos y sistemas heterogéneos sin contar con suficientes conocimientos técnicos. En tales situaciones, la visualización, uno de los primeros y más importantes pasos del proceso analítico, es una ayuda clave para comprender los datos. Permite presentar la información de forma visual y comprensible.

Visualización de datos: la clave para comprender y tomar decisiones

En el sector de la construcción actual, en el que los datos de los proyectos se caracterizan por su complejidad y su estructura multinivel, la visualización desempeña un papel fundamental. La visualización de los datos de permite a los gestores de proyectos e ingenieros visualizar patrones y tendencias complejos ocultos en grandes volúmenes heterogéneos de datos.

> Visualizar los datos de facilita la comprensión del estado de un proyecto: asignación de recursos, tendencias de costes o uso de materiales. Los gráficos y diagramas hacen que la información compleja y árida sea accesible y comprensible, lo que permite identificar rápidamente las áreas clave que necesitan atención y detectar posibles problemas.

La visualización de los datos de no sólo facilita la interpretación de la información, sino que es un paso crucial en el proceso analítico y la toma de decisiones de gestión fundamentadas, ya que ayuda a *responder a las pregun*tas *"¿qué ha pasad*o?" y "¿cómo ha sucedido?" (Fig, 2.2-5). (Fig, 2.2-5).

> *Los gráficos son herramientas visuales para resolver problemas lógicos [132].*
>
> - Jacques Bertin, "Gráficos y tratamiento gráfico de la información

Antes de tomar decisiones clave, es más probable que los jefes de proyecto utilicen representaciones visuales de los datos en lugar de cifras áridas y difíciles de interpretar procedentes de hojas de cálculo o mensajes de texto.

> Los datos sin visualización son como materiales de construcción esparcidos al azar por una obra: su potencial no está claro. Sólo cuando se visualizan con claridad, como una casa hecha de ladrillos y hormigón, se hace patente su valor. Hasta que no se construye la casa, es imposible saber si el montón de materiales se convertirá en una pequeña cabaña, un chalet de lujo o un rascacielos.

Las empresas disponen de datos procedentes de diversos sistemas (Fig, 1.2-4 a Fig. 2.1-10), transacciones financieras y numerosos datos de texto. Sin embargo, la utilización de estos datos en beneficio de la empresa resulta a menudo complicada. En tales situaciones, la visualización se convierte en una herramienta importante para comunicar el significado de los datos, ayudando a presentar la información en formatos comprensibles para cualquier experto, como cuadros de mando, gráficos y diagramas.

> El estudio de PwC "What Students Need to Succeed in a Fast Changing Business World" (2015) destaca [9] que las empresas de éxito van más allá del análisis de datos y utilizan activamente herramientas de visualización interactiva como gráficos, infografías y cuadros de mando analíticos para apoyar la toma de decisiones. Según el informe, la visualización de datos ayuda a los clientes a comprender la historia que cuentan los datos a través de gráficos, cuadros de mando y modelos de datos interactivos.

El proceso de convertir la información en formas gráficas visuales como cuadros, gráficos y diagramas mejora la comprensión e interpretación de los datos por parte del cerebro humano (Fig, 7.1-4). Esto permite a los gestores y analistas de proyectos evaluar con mayor rapidez escenarios complejos y tomar decisiones fundamentadas basadas en tendencias y patrones visualmente reconocibles, en lugar de en la intuición.

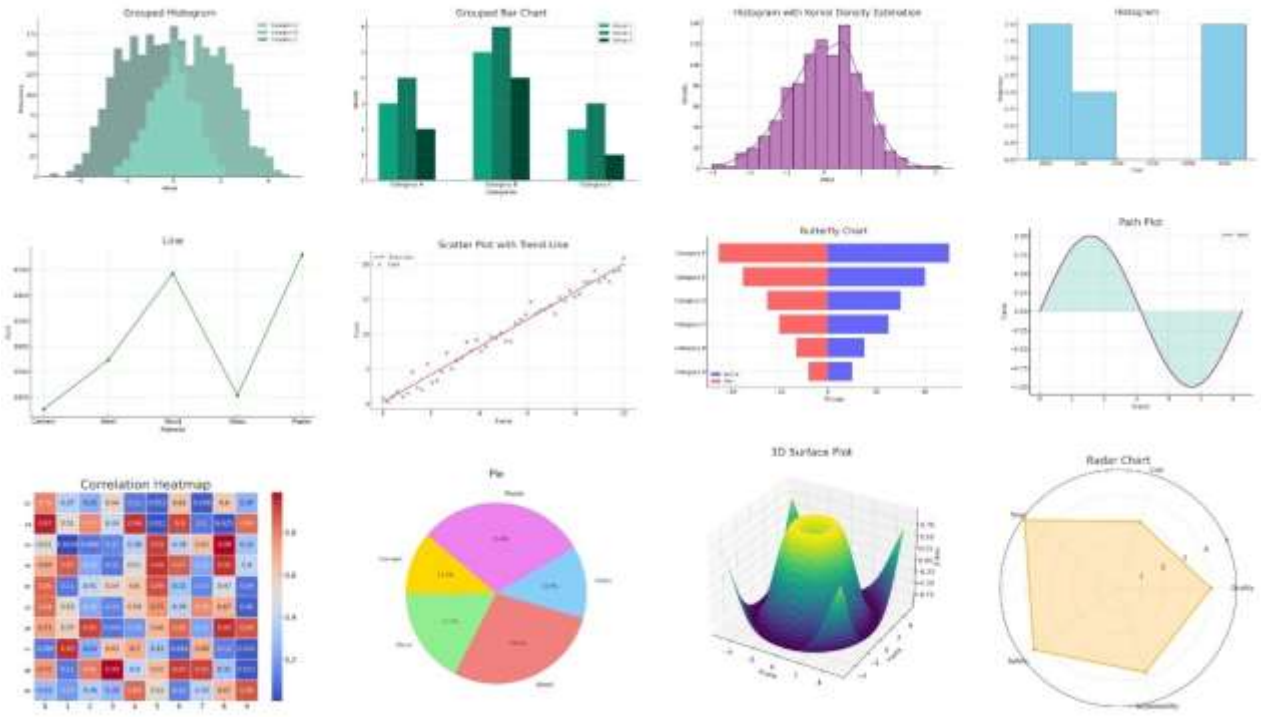


Fig. 7.1-4 Varios tipos de visualización están diseñados para ayudar al cerebro humano a comprender mejor y dar sentido a la árida información de los números.

La creación de visualizaciones a partir de datos y el uso de varias bibliotecas de visualización gratuitas se tratarán con más detalle en el próximo capítulo sobre procesos ETL.

La visualización se está convirtiendo en un elemento integral del trabajo con datos en el sector de la construcción: no sólo ayuda a "ver" los datos, sino también a comprender su significado en el contexto de las tareas de gestión. Sin embargo, para que la visualización sea realmente útil, es necesario determinar de antemano qué es exactamente lo que hay que visualizar y qué métricas son realmente importantes para evaluar el rendimiento del proyecto. Aquí es donde entran en juego métricas de rendimiento como los KPI y el ROI. Sin ellos, incluso los cuadros de mando más bonitos corren el riesgo de no ser más que "ruido informativo".

KPIs y ROI

En el sector de la construcción actual, la gestión de los indicadores de rendimiento (KPI y ROI) y su visualización a través de informes y cuadros de mando desempeñan un papel fundamental en la mejora de la productividad y la eficacia de la gestión de proyectos.

Como en cualquier negocio, en la construcción es necesario definir claramente las métricas con las que se miden el éxito, el rendimiento de la inversión y el rendimiento. A la hora de obtener datos sobre diversos procesos, una organización impulsada por los datos debe aprender primero a identificar **los KPI clave (Key Performance Indicators)**, medidas cuantitativas que reflejan el grado de consecución de los objetivos estratégicos y operativos.

Para calcular un KPI, suele utilizarse una fórmula (Fig, 7.1-5) que incluye indicadores reales y planificados. Por ejemplo, para calcular un KPI individual para un proyecto, empleado o proceso, hay que dividir el rendimiento real por el rendimiento previsto y multiplicar el resultado por 100%.

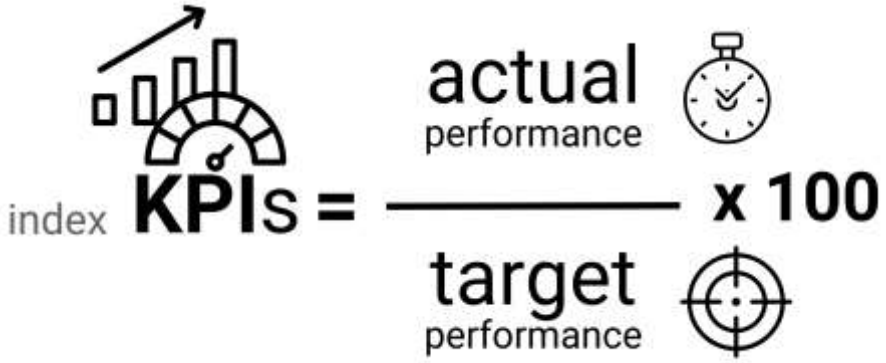


Fig. 7.1-5 Los KPI se utilizan para medir el éxito de un proyecto o proceso en la consecución de objetivos clave.

A nivel de obra pueden utilizarse indicadores más detallados (KPI):

- **Calendario de hitos clave** (cimentación, instalación, acabado): permite controlar el cumplimiento de los planes de trabajo.
- **Porcentaje de exceso de material**: ayuda a gestionar las compras y a minimizar el despilfarro.
- **Número de paradas de máquina no programadas**: afecta a la productividad y a los costes.

Elegir los parámetros equivocados puede llevar a decisiones erróneas sobre "*qué hacer*" (Fig, 2.2-5). Por ejemplo, si una empresa se centra únicamente en el coste por metro cuadrado pero no tiene en cuenta el coste de las reformas, el ahorro en materiales puede dar lugar a una peor calidad y mayores costes en futuros proyectos.

A la hora de fijar objetivos, es importante dejar claro qué se está midiendo. Una redacción imprecisa lleva a conclusiones erróneas y complica el control. Veamos ejemplos de KPI acertados y fallidos en la construcción.

Buenos indicadores clave de rendimiento:

- *"Para finales de año, reduzca un 10% el porcentaje de trabajos de redecoración"*.
- *"Aumentar la velocidad de instalación de fachadas en un 15% sin comprometer la calidad para el próximo trimestre"*

- *"Reducir el tiempo de inactividad de la maquinaria en un 20% optimizando los planes de trabajo para finales de año*"

Estas métricas son claramente mensurables, tienen valores y plazos específicos.

KPIs malos:

- *"Construiremos más rápido*" (¿Cuánto más rápido? ¿Qué significa "más rápido"?).
- *"Mejoraremos la calidad del trabajo concreto*" (¿Cómo se mide exactamente la calidad?)
- *"Mejoraremos la interacción de los contratistas in situ*" (¿Qué criterios demostrarán la mejora?)

> Un buen KPI es aquel que puede medirse y evaluarse objetivamente. En la construcción, esto es especialmente importante, ya que sin indicadores claros es imposible controlar el rendimiento y lograr resultados estables.

Además del KPI, existe una métrica adicional para evaluar la eficacia de las inversiones: el **ROI (Return on Investment)**, un indicador del rendimiento de la inversión que refleja la relación entre el beneficio y los fondos invertidos. El ROI ayuda a evaluar si la introducción de nuevos métodos, tecnologías o herramientas está justificada: desde soluciones digitales y automatización (por ejemplo, la Fig. 7.3-2) hasta el uso de nuevos materiales de construcción. Este indicador ayuda a tomar decisiones informadas sobre nuevas inversiones en función de su impacto real en la rentabilidad de la empresa

> En el contexto de la gestión de proyectos de construcción, el ROI (retorno de la inversión) puede utilizarse como uno de los indicadores clave de rendimiento (KPI) si el objetivo de una empresa es medir el retorno de la inversión de un proyecto, una tecnología o la mejora de un proceso. Por ejemplo, si se está implantando una nueva técnica de gestión de la construcción, el ROI puede mostrar en qué medida ha mejorado la rentabilidad.

La medición periódica de los indicadores clave de rendimiento (KPI) y el rendimiento de la inversión (ROI) a partir de datos recogidos de diversas fuentes, como el consumo de material, las horas de trabajo y los costes, permite a la dirección de proyectos gestionar eficazmente los recursos y tomar decisiones rápidas. Almacenar estos datos a largo plazo permite analizar tendencias futuras y optimizar los procesos.

Se utilizan diversos cuadros y gráficos para visualizar los KPI, los ROI y otras métricas, que suelen combinarse en cuadros de mando.

Cuadros de mando y paneles de control: visualización de métricas para una gestión eficaz

Para visualizar indicadores y métricas se utilizan diversos cuadros y gráficos, que suelen combinarse en vitrinas de datos y cuadros de mando. Estos cuadros de mando ofrecen una visión centralizada del estado de un proyecto o de partes de un proyecto, mostrando indicadores clave (idealmente en tiempo real). Los

cuadros de mando actualizados de forma continua permiten al equipo responder rápidamente a los cambios.

> **Cuadros de mando** son herramientas que visualizan evaluaciones cuantitativas, haciéndolas fácilmente accesibles y comprensibles para todos los participantes en el proyecto.

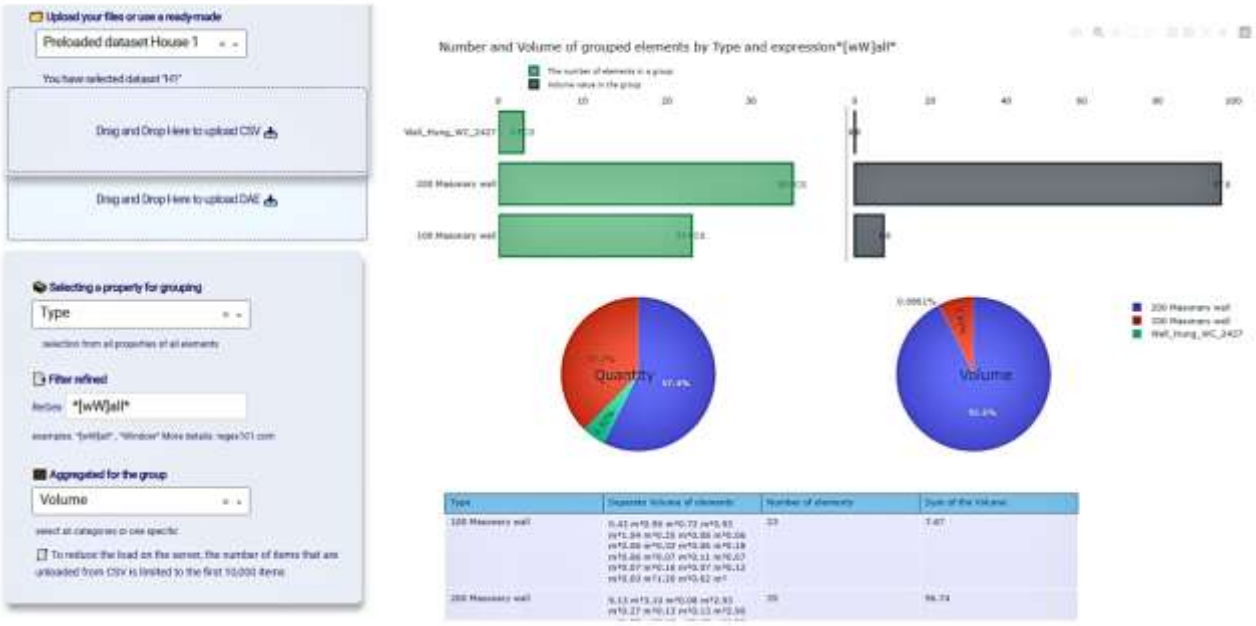


Fig. 7.1-6 Gestionar los KPI y visualizarlos a través de cuadros de mando es clave para mejorar la productividad y la eficiencia de los proyectos.

He aquí algunos ejemplos de herramientas populares en las que se pueden crear cuadros de mando:

- **Power BI** es una herramienta de Microsoft para crear informes y cuadros de mando interactivos.
- **Tableau y Google Data Studio** son potentes herramientas para visualizar datos y crear cuadros de mando sin tener que escribir código.
- **Plotly** (Fig, 7.1-6, Fig. 7.2-12) es una biblioteca para crear gráficos interactivos, y Dash es un marco para crear aplicaciones web de análisis de datos. Pueden combinarse para crear cuadros de mando interactivos.
- **Numerosas bibliotecas de Python** (Fig, 7.2-9 - Fig. 7.2-11) - Python dispone de muchas bibliotecas de código abierto y gratuitas para la visualización de datos, como Matplotlib, Seaborn, Plotly, Bokeh y otras. Se pueden utilizar para crear gráficos e integrarlos en una aplicación web utilizando marcos como Flask o Django.
- **Bibliotecas JavaScript**: permite crear cuadros de mando interactivos utilizando bibliotecas JavaScript de código abierto como D3.js o Chart.js e integrarlos en una aplicación web.

> Para evaluar los KPI y crear cuadros de mando, necesita datos actualizados y un calendario claro de recogida y análisis de la información.

En general, los KPI, el ROI y los cuadros de mando del sector de la construcción constituyen la base de un enfoque analítico de la gestión de proyectos. No solo ayudan a supervisar y evaluar el estado actual, sino que también proporcionan información valiosa para los futuros procesos de planificación y optimización, procesos que dependen directamente de la interpretación de los datos y de la formulación de las preguntas adecuadas y oportunas.

Análisis de datos y el arte de hacer preguntas

La interpretación de los datos es la fase final del análisis, en la que la información cobra sentido y empieza a "hablar". Aquí es donde se formulan las respuestas a las preguntas clave: *"¿qué hacer?"* y *"¿cómo hacer?"* (Fig, 2.2-5). Esta etapa permite resumir los resultados, identificar patrones, establecer relaciones causa-efecto y extraer conclusiones basadas en la visualización y el análisis estadístico.

> *Tal vez no esté lejos el momento en que se llegue a comprender que, para llegar a ser plenamente un ciudadano efectivo de uno de los nuevos y grandes Estados mundiales complejos que se están desarrollando en la actualidad, es tan necesario saber calcular, pensar en términos de medias, máximos y mínimos, como lo es ahora saber leer y escribir* [133].
>
> - Samuel S. Wilkes, citado en un discurso presidencial de 1951 ante la American Statistical Association

> Según el informe "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) publicado por el Gobierno británico [83], la aplicación de la analítica de datos y la inteligencia artificial (AI) puede mejorar notablemente los procesos de gestión de proyectos, aumentando la precisión de las previsiones de plazos y costes, además de reducir el riesgo y la incertidumbre. El documento destaca que las organizaciones públicas que utilizan herramientas analíticas avanzadas logran un mayor rendimiento en las iniciativas de infraestructuras.

Las empresas de construcción modernas que operan en el entorno altamente competitivo y de bajos márgenes de la cuarta revolución industrial pueden compararse con las operaciones militares. En este caso, la supervivencia y el éxito de la empresa dependen de la rapidez con que se obtengan recursos e información de calidad y, por tanto, de que se tomen decisiones a tiempo y con conocimiento de causa (Fig. 7.1-7).

Si la visualización de datos es la "inteligencia" que proporciona la visión de conjunto, la analítica de datos es la "munición" necesaria para la acción. Responde a las preguntas: *¿qué hacer?* y *¿cómo hacerlo?*, constituyendo la base para obtener una ventaja competitiva en el mercado.

> La analítica convierte datos dispares en información estructurada y significativa en la que basar las decisiones.

La tarea de los analistas y gestores no consiste sólo en interpretar la información, sino en ofrecer decisiones fundamentadas, identificar tendencias, determinar relaciones entre distintos tipos de datos y categorizarlos de acuerdo con los objetivos y especificidades del proyecto. Utilizando herramientas de visualización y métodos de análisis estadístico, convierten los datos en un activo estratégico para la empresa.

Fig. 7.1-7 Es el análisis de los datos el que, en última instancia, convierte la información recopilada en una fuente para la toma de decisiones.

Para tomar decisiones realmente fundamentadas en el proceso analítico, es necesario aprender a formular correctamente las preguntas que se hacen a los datos. La calidad de estas preguntas afecta directamente a la profundidad de los conocimientos obtenidos y, en consecuencia, a la calidad de las decisiones de gestión.

> *El pasado sólo existe en la medida en que está presente en los registros de hoy. Y lo que estos registros representan viene determinado por las preguntas que nos hacemos. No hay más historia que ésta [134].*
>
> - John Archibald Wheeler, físico 1982

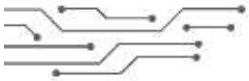
El arte de hacer preguntas profundas y pensar de forma crítica es una habilidad esencial para trabajar con datos. La mayoría de la gente tiende a hacer preguntas simples y superficiales que requieren poco esfuerzo para responderlas. Sin embargo, el verdadero análisis comienza con preguntas significativas y reflexivas que pueden descubrir relaciones ocultas y relaciones causa-efecto en la información que puede estar oculta tras múltiples capas de razonamiento.

> Según el estudio "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135], para que la transformación digital tenga éxito es necesario invertir en capacidades analíticas, programas de gestión del cambio y alinear los objetivos empresariales con las iniciativas de TI. Las empresas que crean una cultura impulsada por los datos deben invertir en capacidades de análisis de datos y poner en marcha programas de gestión del cambio para inculcar nuevas formas de pensar, comportamientos y formas de trabajar.

Si no se invierte en el desarrollo de una cultura analítica, la mejora de las herramientas de datos y la formación de especialistas, las empresas seguirán corriendo el riesgo de tomar decisiones basadas en información obsoleta o incompleta, o de confiar en las opiniones subjetivas de los directivos de HiPPO (Fig, 2.1-9).

> La toma de conciencia de la relevancia y la necesidad de actualizar constantemente los análisis y los cuadros de mando lleva inevitablemente a los directivos a comprender la importancia de automatizar los procesos analíticos. La automatización aumenta la velocidad de la toma de decisiones, reduce el impacto del factor humano y garantiza la pertinencia de los datos. Con el crecimiento exponencial de los volúmenes de información, la velocidad se convierte no sólo en una ventaja competitiva, sino en un factor clave para el éxito sostenible.

La automatización de los procesos de análisis y tratamiento de datos en general está indisolublemente ligada al tema del ETL (Extraer, Transformar, Cargar). Al igual que en el proceso de automatización necesitamos transformar los datos, en el proceso ETL los datos se extraen de diversas fuentes, se transforman según los requisitos necesarios y se cargan en los sistemas de destino para su uso posterior.



CAPÍTULO 7.2.
FLUJO DE DATOS SIN ESFUERZO MANUAL: POR QUÉ ES NECESARIA LA ETL

Automatización ETL: reducción de costes y agilización del tratamiento de datos

Cuando los indicadores clave de rendimiento (KPI) dejan de crecer a pesar del aumento del volumen de datos y del tamaño de los equipos, la dirección de la empresa se da cuenta inevitablemente de la necesidad de automatizar los procesos. Tarde o temprano, esta toma de conciencia se convierte en un incentivo para poner en marcha una automatización compleja, cuyo principal objetivo es reducir la complejidad de los procesos, acelerar el procesamiento y reducir la dependencia del factor humano.

> Según el estudio de McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136], las empresas que utilizan arquitecturas de streaming de datos obtienen una ventaja significativa porque pueden analizar la información en tiempo real. Las tecnologías de streaming permiten el análisis directo de mensajes en tiempo real y la aplicación del mantenimiento predictivo en la fabricación mediante el análisis de datos de sensores en tiempo real.

> La simplificación de procesos es la automatización, en la que las funciones manuales tradicionales se sustituyen por algoritmos y sistemas.

La cuestión de la automatización, o más bien de "minimizar el papel del ser humano en el tratamiento de datos", es un proceso irreversible y muy delicado para toda empresa. Los especialistas de cualquier campo profesional suelen dudar a la hora de revelar plenamente sus métodos y sutilezas de trabajo a sus colegas optimizadores, conscientes del riesgo de perder su empleo en un entorno tecnológico en rápida evolución.

> *Si quieres hacer enemigos, intenta cambiar las cosas [137].*
>
> - Woodrow Wilson, discurso ante un congreso de vendedores, Detroit, 1916

A pesar de las evidentes ventajas de la automatización, muchas empresas siguen teniendo una elevada proporción de trabajo manual en sus prácticas cotidianas, especialmente en el ámbito de los datos de ingeniería. Para ilustrar la situación actual, veamos un ejemplo típico de procesamiento secuencial de datos dentro de este tipo de procesos.

El tratamiento manual de datos puede ilustrarse con el ejemplo de la interacción con información obtenida de bases de datos CAD. El tratamiento tradicional de datos (proceso "manual" ETL) en los departamentos de CAD (BIM) para crear tablas de atributos o crear documentación basada en datos de diseño tiene lugar en el siguiente orden (Fig. 7.2-1):

1. **Extracción** manual **(Extract**): el usuario abre manualmente el proyecto - lanzando la aplicación CAD (BIM) (Fig. 7.2-1 paso 1).
2. **Verificación**: el siguiente paso suele implicar la ejecución manual de varios plug-ins o aplicaciones de ayuda de para preparar los datos y evaluar su calidad (Fig. 7.2-1 paso 2-3).
3. **Transformación** manual **(Transform):** tras la preparación, comienza el tratamiento de los datos, que requiere el manejo manual de diversas herramientas informáticas en las que se preparan los datos para su carga (Fig. 7.2-1 paso 4).
4. **Carga** manual **(Load):** carga manual de datos convertidos a sistemas externos, formatos de datos y documentos (Fig. 7.2-1 paso 5).

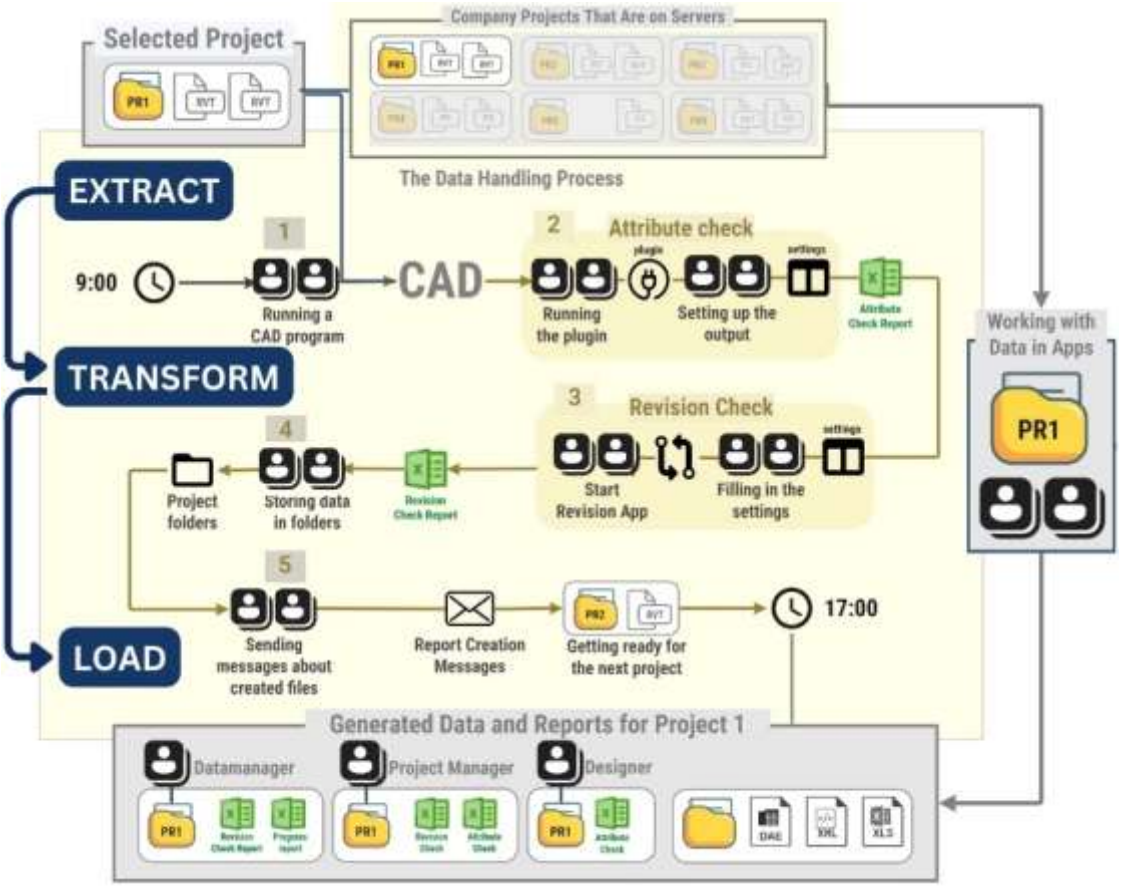


Fig. 7.2-1 El procesamiento ETL manual tradicional está limitado por los deseos y las capacidades físicas de cada técnico.

Un flujo de trabajo de este tipo es un ejemplo de un proceso ETL clásico - extracción, transformación y carga (ETL). A diferencia de otros sectores, en los que las canalizaciones ETL automáticas son la norma desde hace tiempo, en el de la construcción sigue predominando el trabajo manual, que ralentiza los procesos y aumenta los costes

> **ETL (Extraer, Transformar, Cargar**) es el proceso de extraer datos de diversas fuentes, transformarlos al formato deseado y cargarlos en el sistema de destino para su posterior análisis y uso.

ETL es un proceso que denota tres componentes clave del procesamiento de datos: Extraer, Transformar y Cargar (Fig, 7.2-2):

- Extraer - extraer datos de diferentes fuentes (archivos, bases de datos, API).
- **Transform** - limpieza de datos, agregación, normalización y procesamiento lógico.
- **Cargar**: carga información estructurada en un almacén de datos, un informe o un sistema de BI.

Anteriormente en el libro, el concepto de ETL sólo se había tocado ocasionalmente: en la conversión de un documento escaneado no estructurado en un formato tabular estructurado (Fig. 4.1-1), en el contexto de la formalización de requisitos para sistematizar la percepción de los procesos tanto vitales como empresariales (Fig. 4.4-20) y en la automatización de la validación y el tratamiento de datos a partir de soluciones CAD. Veamos ahora con más detalle el ETL en el contexto de los flujos de trabajo típicos.

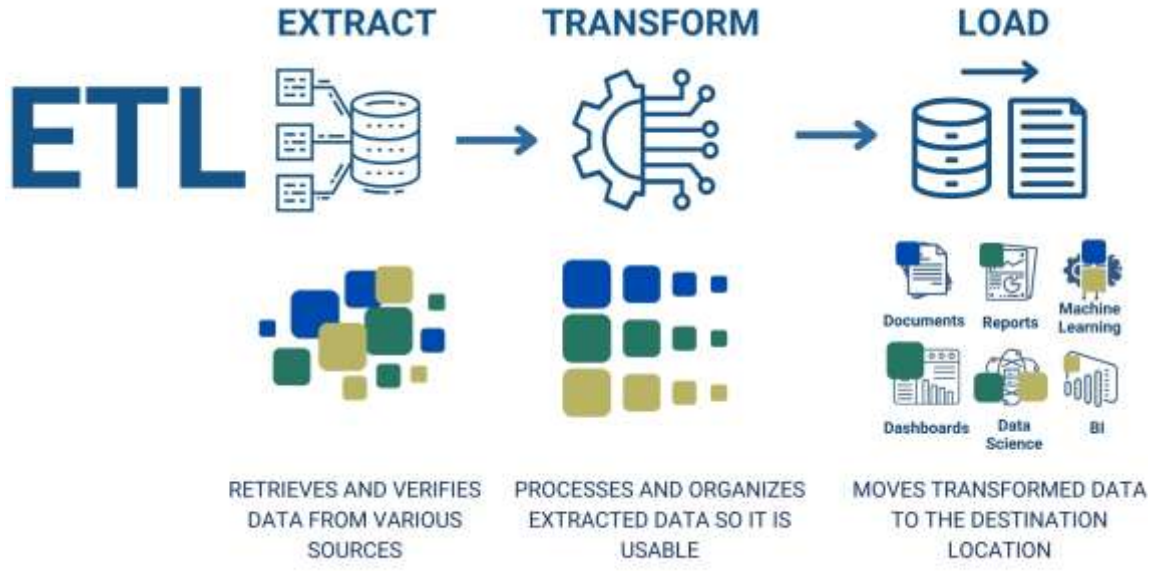


Fig. 7.2-2 ETL automatiza las tareas repetitivas de procesamiento de datos.

ETL manual o semiautomatizado -el proceso implica un gestor o técnico que gestiona manualmente todos los pasos, desde la recogida de datos hasta la generación de informes. Este proceso requiere un tiempo considerable, sobre todo cuando el horario de trabajo es limitado (por ejemplo, de 9:00 a 17:00).

A menudo, las empresas intentan resolver el problema de la escasa eficacia y la lentitud adquiriendo soluciones modulares integradas (ERP, PMIS, CPM, CAFM, etc.), que luego desarrollan proveedores y consultores externos. Pero estos vendedores y desarrolladores externos se convierten a menudo en un punto crítico de dependencia: sus limitaciones técnicas afectan directamente al rendimiento de todo el sistema y de la

empresa en su conjunto, como se ha detallado en capítulos anteriores sobre sistemas y formatos propietarios. Los problemas creados por la fragmentación y la dependencia se trataron en detalle en el capítulo "Cómo las empresas de la construcción se ahogan en el caos de los datos".

> Si una empresa no está preparada para implantar una gran plataforma modular de uno de los proveedores, empieza a buscar formas alternativas de automatización. Una de ellas es desarrollar su propio ETL modular abierto -conveyors, donde cada etapa (extracción, transformación, validación, carga) se implementa como scripts ejecutados según un calendario.

En la versión automatizada del mismo flujo de trabajo ETL (Fig. 7.2-1), el flujo de trabajo parece un código modular que comienza con el tratamiento de los datos y su traducción a una forma estructurada abierta. Una vez recibidos los datos estructurados, se ejecutan automáticamente varios scripts o módulos, de forma programada, para comprobar los cambios, transformar y enviar mensajes (Fig. 7.2-3).

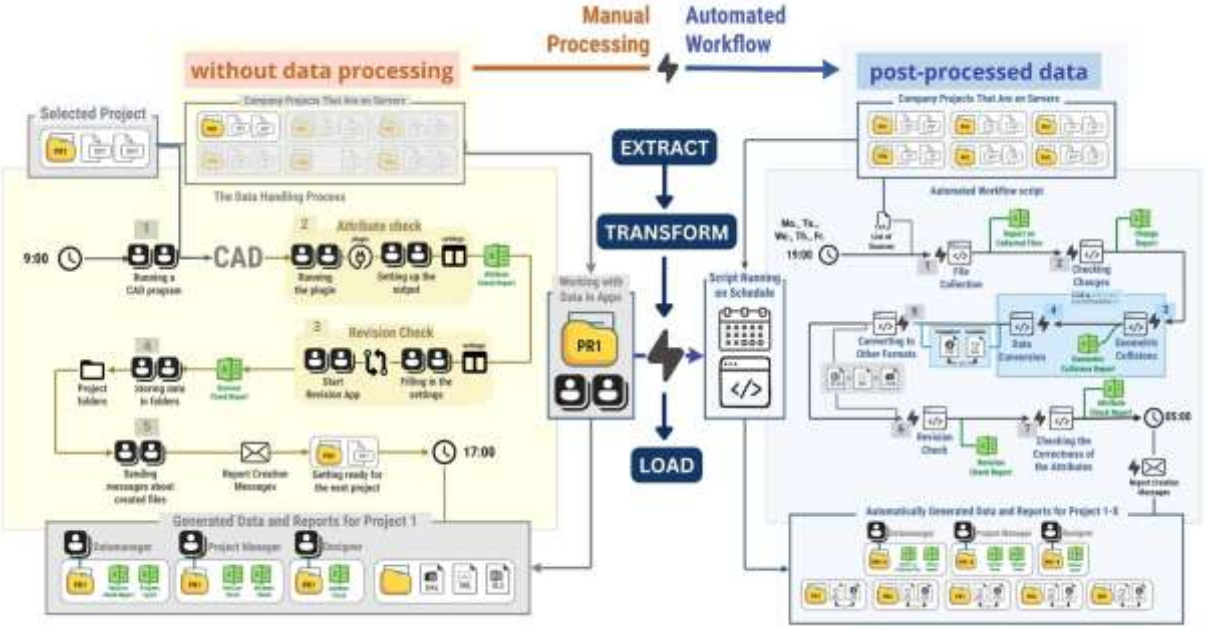


Fig. 7.2-3 A la izquierda, el mecanizado manual; a la derecha, un proceso automático que, a diferencia del mecanizado manual tradicional, no está limitado por las capacidades del usuario.

En un flujo de trabajo automatizado, el procesamiento de datos se simplifica mediante el preprocesamiento de datos ET(L): estructuración y unificación.

> En los métodos de tratamiento tradicionales, los especialistas trabajan con los datos "tal cual", es decir, tal y como se recuperan de los sistemas o programas informáticos. En cambio, en los procesos automatizados, los datos suelen pasar primero por un ETL -payplane, donde se les da una estructura y un formato coherentes y adecuados para su posterior uso y análisis.

Tomemos un ejemplo práctico de ETL, que demuestra el proceso de validación de tablas de datos descrito en el capítulo "Validación de datos y resultados de la validación" (Fig, 4.4-13). Para ello, utilizamos la biblioteca Pandas junto con el LLM para los procesos automatizados de análisis y procesamiento de datos.

ETL Extraer: recogida de datos

La primera etapa del proceso ETL - Extraer) - comienza con la escritura de código para recopilar conjuntos de datos que se comprobarán y procesarán posteriormente. Para ello, escanearemos todas las carpetas del servidor de producción, recopilaremos documentos de un determinado formato y contenido y, a continuación, los convertiremos a un formato estructurado. Este proceso se trata en detalle en los capítulos "Conversión de datos no estructurados y textuales en forma estructurada" y "Conversión de datos CAD (BIM) en forma estructurada" (Fig. 4.1-1 - Fig. 4.1-12).

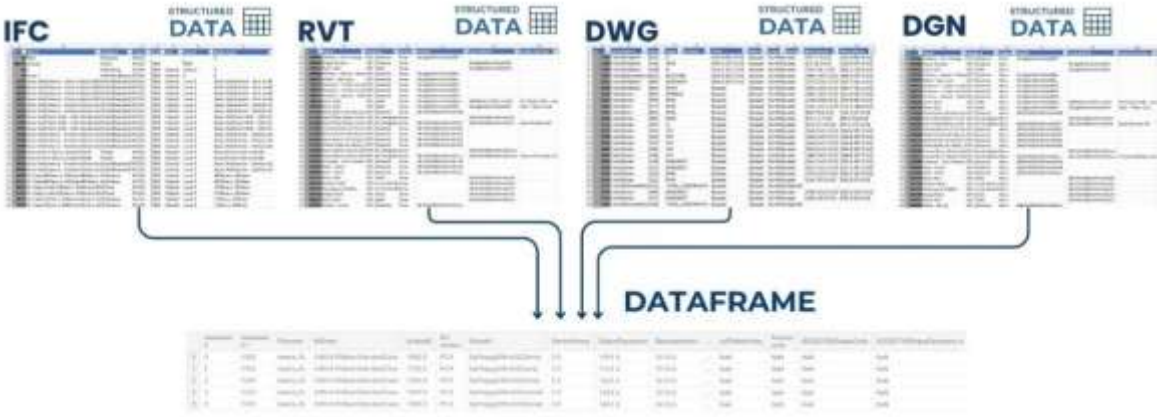


Fig. 7.2-4 Convertir los datos CAD (BIM) en un gran marco de datos que contendrá todas las secciones del proyecto.

Como ejemplo ilustrativo, utilizamos el paso de carga de datos Extract y obtenemos una tabla de todos los proyectos CAD- (BIM-) (Fig. 7.2-4) utilizando convertidores habilitados para ingeniería inversa [138] para formatos RVT e IFC para obtener tablas estructuradas de todos los proyectos y combinarlas en una gran tabla DataFrame.

Create in LLM Chat

Extract data from CAD-BIM.py

```python
import os
import subprocess
import time
import pandas as pd

path_conv = r'C:\DDC_2023\\'  # Where RvtExporter.exe|IfcExporter.exe is located
path = r'C:\IFCprojects\\'    # Where Revit|IFC files are stored

def convert_and_wait(path_conv, exporter_name, file_path, extension):
    # Start the conversion process
    subprocess.Popen([os.path.join(path_conv, exporter_name),
    file_path], cwd=path_conv)
    output_file = os.path.join(path,
    f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")

# Conversion process for RVT and IFC files
for file in os.listdir(path):
    full_path = os.path.join(path, file)
    if file.endswith('.ifc'):
        convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
    elif file.endswith('.rvt'):
        convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')

# Combine converted Excel files into one dataframe
df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
if f.endswith('.xlsx')], ignore_index=True)
```

PC

Run in IDE

Fig. 7.2-5 Conversión mediante código Python y la herramienta de ingeniería inversa SDK de archivos RVT e IFC en un gran DataFrame estructurado (df).

Pandas DataFrame puede cargar datos desde una gran variedad de fuentes, incluyendo archivos de texto CSV, hojas de cálculo Excel, archivos JSON - y XML -, formatos de almacenamiento de big data como Parquet y HDF5, y desde MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle y otras bases de datos. Además, Pandas soporta la carga de datos desde APIs, páginas web, servicios en la nube y sistemas de almacenamiento como Google BigQuery, Amazon Redshift y Snowflake.

- Para escribir código que conecte y recopile información de las bases de datos, envíe una solicitud de texto similar a la sala de chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otra):

> Por favor, escriba un ejemplo de conexión a MySQL y la conversión de datos a ⏎

- Respuesta de LLM:

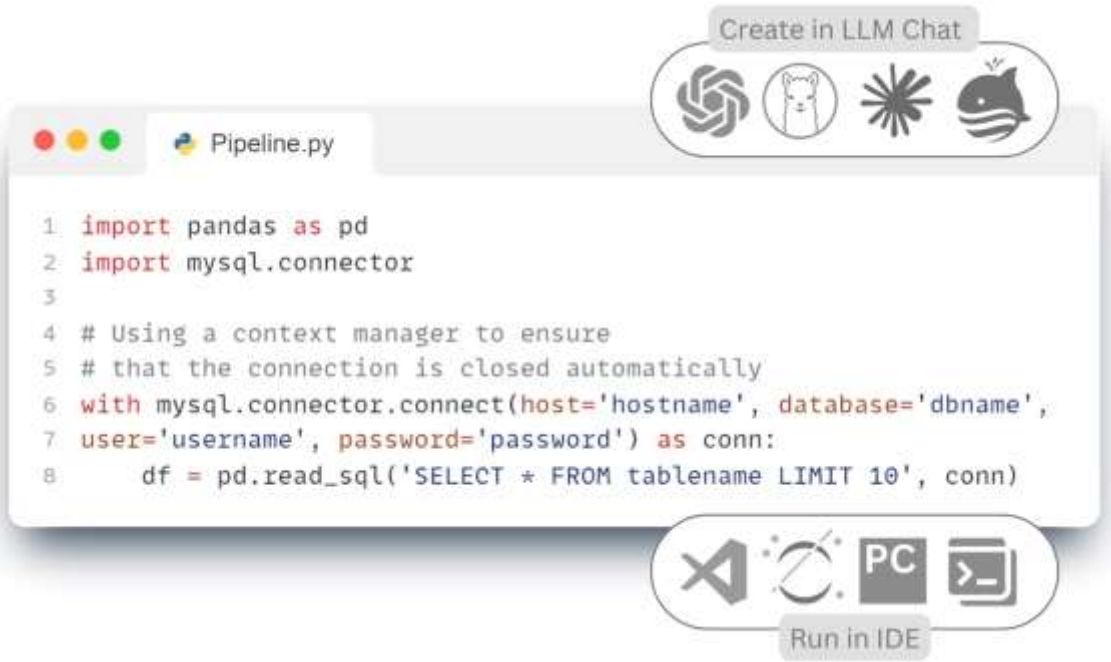


Fig. 7.2-6 Ejemplo de conexión mediante Python a una base de datos MySQL e importación de datos de la base de datos MySQL a un DataFrame.

El código resultante (Fig. 7.2-5, Fig. 7.2-6) se puede ejecutar en uno de los IDEs (entornos de desarrollo integrados) populares que hemos mencionado anteriormente en modo offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con plugin Python, JupyterLab o herramientas online populares: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Al cargar los datos multiformato en la variable "df" (Fig, 7.2-5 - fila 25; Fig. 7.2-6 - fila 8), convertimos los datos al formato Pandas DataFrame, una de las estructuras más populares para el procesamiento de datos, que es una tabla bidimensional con filas y columnas. Hablaremos más sobre otros formatos de almacenamiento utilizados en ETL -Pipelines como Parquet, Apache ORC, JSON, Feather, HDF5, así como los almacenes de datos modernos en el capítulo "Almacenamiento y gestión de datos en el sector de la construcción" (Fig, 8.1-2).

Tras la etapa de extracción y estructuración de los datos (Extract), se forma una única matriz de información (Fig. 7.2-5, Fig. 7.2-6), lista para su tratamiento posterior. Sin embargo, antes de cargar estos datos en los sistemas de destino o de utilizarlos para el análisis, es necesario garantizar su calidad, integridad y conformidad con los requisitos especificados. En esta fase tiene lugar la transformación de los datos (Transform), un paso clave para garantizar la fiabilidad de las conclusiones y decisiones posteriores.

ETL Transform: aplicación de reglas de validación y transformación

Transformación es la etapa en la que se procesan y transforman los datos. Este proceso puede incluir la comprobación de la corrección, la normalización, el relleno de los valores que faltan y la validación mediante herramientas automatizadas

> Según el estudio de PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], las empresas de auditoría modernas se están alejando de la verificación aleatoria de datos y avanzan hacia el análisis de cantidades masivas de información mediante herramientas automatizadas. Este enfoque permite no solo identificar discrepancias en los informes, sino también ofrecer recomendaciones para optimizar los procesos empresariales.

En la construcción, pueden utilizarse métodos similares, por ejemplo, para la validación automática de los datos de diseño, el control de calidad de la construcción y la evaluación del rendimiento de los contratistas. Una herramienta para automatizar y acelerar el procesamiento de datos es el uso de expresiones regulares (RegEx) en la fase de transformación de datos (Transform) del proceso ETL. RegEx permite validar eficazmente cadenas de datos, detectar incoherencias y garantizar la integridad de la información con un consumo mínimo de recursos. Más detalles sobre RegEx (Fig. 4.4-7) en el capítulo "Traducir requisitos a formato estructurado".

Veamos un ejemplo práctico: en el sistema de gestión de objetos inmuebles (RPM), el gestor establece los requisitos para los atributos clave de los objetos (Fig. 7.2-7). En la fase de transformación, es necesario validar los siguientes parámetros:

- verificación de los formatos de los identificadores de objetos (atributo "ID")
- control de los valores del periodo de garantía de sustitución (atributo "Periodo de garantía")
- verificación del ciclo de sustitución de elementos (atributo "Requisitos de mantenimiento")

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

Fig. 7.2-7 Verificación de la calidad comienza con el establecimiento de requisitos para los atributos y sus valores límite.

Para establecer valores límite para la validación de parámetros, por ejemplo, supongamos que sabemos

por nuestra experiencia que los valores válidos para el atributo "ID" sólo pueden incluir valores de cadena "W-NEW", "W-OLD1" o "D-122" o valores similares en los que el primer carácter es una letra seguida de un guión y, a continuación, tres caracteres alfabéticos "NEW", "OLD" o cualquier número de tres dígitos (Fig, 7.2-7). Para validar estos identificadores puede utilizarse la siguiente expresión regular (RegEx):

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Este modelo se asegura de que todos los identificadores de los datos coinciden con los criterios especificados. Si algún valor no supera la comprobación, el sistema registra un error. Para crear código Python para transformar los datos y utilizar los datos resultantes para crear una tabla de resultados, basta con formular una consulta en el chat LLM.

- Solicitud por SMS a LLM:

> Escribir código para validar las columnas de DataFrame utilizando expresiones regulares, que comprueba identificadores en formato 'W-NEW' o 'W-OLD' mediante RegEx, eficiencia energética con letras de la 'A' a la 'G', periodo de garantía y ciclo de sustitución con valores numéricos en años. ⏎

Respuesta de LLM:

Create in LLM Chat

Pipeline.py

```python
# RegEx
import re

# Define regex patterns for each column
patterns = {
    'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
    'Energy_Efficiency': r'^[A-G]\+*$',
    'Warranty_Period': r'^\d+\syears$',
    'Replacement_Cycle': r'^\d+\syears$'
}

# Define a verification function using regex
def verify_data(column, pattern):
    # Apply the regex pattern to the series and return the boolean results
    return column.apply(lambda x: bool(re.match(pattern, str(x))))

# Apply the verification function to the DataFrame for each column with its respective pattern
for column, pattern in patterns.items():
    df['verified_' + column] = verify_data(df[column], pattern)
```

Run in IDE

Fig. 7.2-8 El código automatiza el proceso de validación aplicando plantillas RegEx a las columnas de parámetros del marco de datos.

El código Python generado automáticamente (Fig, 7.2-8) utiliza la biblioteca "re" (expresiones regulares RegEx) para definir una función que comprueba cada atributo de un elemento de datos en un DataFrame. Para cada columna (atributo) especificada, la función aplica un patrón RegEx para comprobar que cada registro se ajusta al formato esperado y añade los resultados como nuevos valores (Falso/Verdadero) a la nueva columna-atributo del DataFrame.

> Esta validación automatizada garantiza la conformidad formal de los datos y puede utilizarse como parte de un sistema de control de calidad durante la fase de transformación.

Tras completar con éxito el paso Transformar y la comprobación de calidad, los datos están listos para cargarse en los sistemas de destino. Los datos transformados y validados pueden cargarse en CSV, JSON, Excel, bases de datos y otros formatos para su uso posterior. En función de la tarea, los resultados también pueden presentarse en informes, gráficos o cuadros de mando analíticos.

ETL Carga: Visualice los resultados de en tablas y gráficos

Una vez finalizada la etapa Transformar, cuando los datos se han estructurado y verificado, la etapa final es Cargar, en la que los datos pueden cargarse tanto en el sistema de destino como visualizarse para su

análisis. La presentación visual de los datos permite identificar rápidamente las desviaciones, analizar las distribuciones y comunicar las conclusiones clave a todos los participantes en el proyecto, incluidos los que no tienen formación técnica.

En lugar de presentar la información en forma de tablas y números, podemos utilizar infografías, gráficos y cuadros de mando (dashboards). Una de las herramientas más comunes y flexibles para visualizar datos estructurados en Python es la biblioteca Matplotlib (Fig, 7.2-9, Fig. 7.2-10). Permite crear gráficos estáticos, animados e interactivos, y admite una amplia gama de tipos de gráficos.

- Para visualizar los resultados de la comprobación de atributos del sistema RPM (Fig, 7.2-7), puede utilizar la siguiente consulta al modelo lingüístico:

> Escriba código para visualizar los datos de DataFrame, arriba (Fig, 7.2-7), con un histograma para que los resultados muestren la tasa de error del atributo ⏎

- Respuesta LLM en forma de código y visualización directamente en la sala de chat LLM de los resultados de la ejecución del código:

Create in LLM Chat

Pipeline.py

```python
# Re-importing necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Visualization 1: Bar Chart
plt.figure(figsize=(10, 6))
df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
plt.title('Data Verification Summary - Bar Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Run in IDE

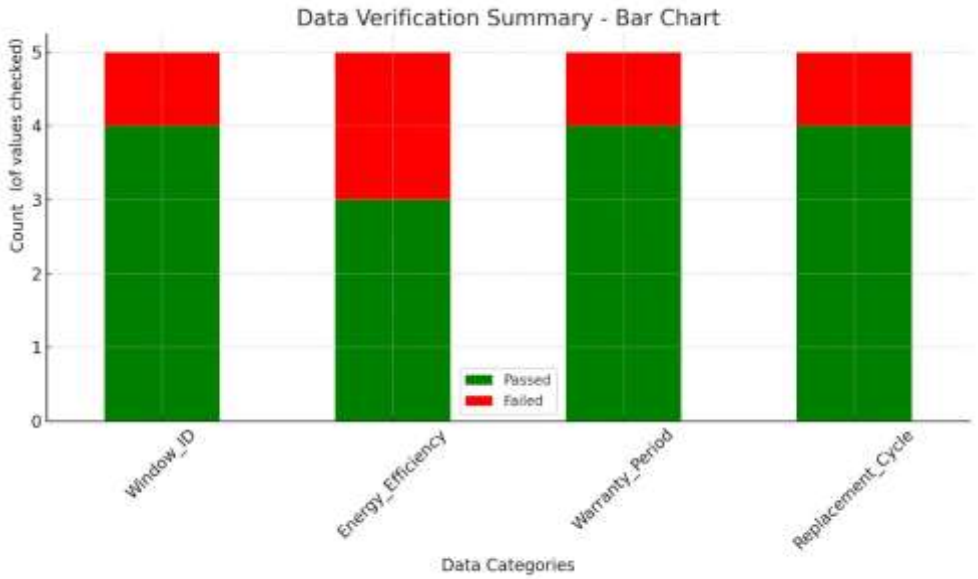


Fig. 7.2-9 Visualización de los resultados del paso Transformar de comprobación de valores de atributos del sistema RPM (Fig. 7.2-7) como histograma en el paso Cargar.

Existen muchas bibliotecas de visualización de código abierto y gratuitas que permiten presentar datos estructurados en diferentes formatos. Sigamos visualizando los resultados con otro tipo de gráfico con la siguiente indicación en el chat:

> Representar los mismos datos como un gráfico de líneas ⏎

Respuesta de LLM:

Create in LLM Chat

Pipeline.py

```
# Visualization 3: Line Chart
df_visual.plot(kind='line', marker='o', figsize=(10, 6))
plt.title('Data Verification Trend - Line Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```

Run in IDE

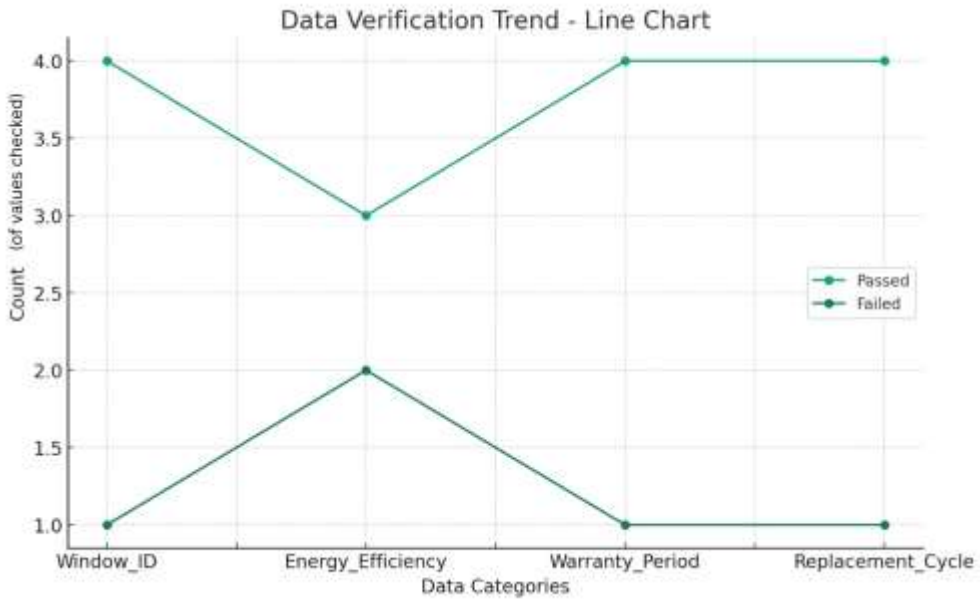


Fig. 7.2-10 Visualización de los datos de validación (Fig, 7.2-8) como diagrama de líneas obtenido mediante la biblioteca Matplotlib.

Existen muchas bibliotecas de visualización de código abierto y gratuitas, como:

- Seaborn - para gráficos estadísticos (Fig. 7.2-11)
- Plotly - para visualizaciones web interactivas (Fig. 7.2-12, Fig. 7.1-6)
- Altair: visualización declarativa
- Dash o Streamlit - para crear cuadros de mando completos

No es necesario conocer bibliotecas específicas para la visualización: las herramientas modernas, como LLM, permiten generar automáticamente código para gráficos y aplicaciones completas a partir de la descripción de la tarea.

La elección de la herramienta depende de los objetivos del proyecto: si se trata de un informe, una presentación o un cuadro de mando en línea. Por ejemplo, la biblioteca de código abierto Seaborn es especialmente buena para trabajar con datos categóricos, ayudando a identificar patrones y tendencias.

- Para ver la biblioteca Seaborn en funcionamiento, puede pedir directamente al LLM que utilice la biblioteca deseada o enviarle una solicitud de texto similar a:

> Mostrar mapa de calor de ⏎

- La respuesta de LLM en forma de código y un gráfico terminado, cuyo código de trazado puede copiarse ahora en el IDE, y el propio gráfico puede copiarse o guardarse para pegarlo en un documento:

Fig. 7.2-11 Visualización de los resultados de la validación (Fig, 7.2-8) de los datos mediante la biblioteca Seaborn.

Para quienes prefieran un enfoque interactivo, existen herramientas que permiten crear diagramas y paneles dinámicos con capacidad de interacción. La biblioteca Plotly (Fig, 7.1-6, Fig. 7.2-12) ofrece la posibilidad de crear gráficos y paneles altamente interactivos que pueden incrustarse en páginas web y permiten al usuario interactuar con los datos en tiempo real.

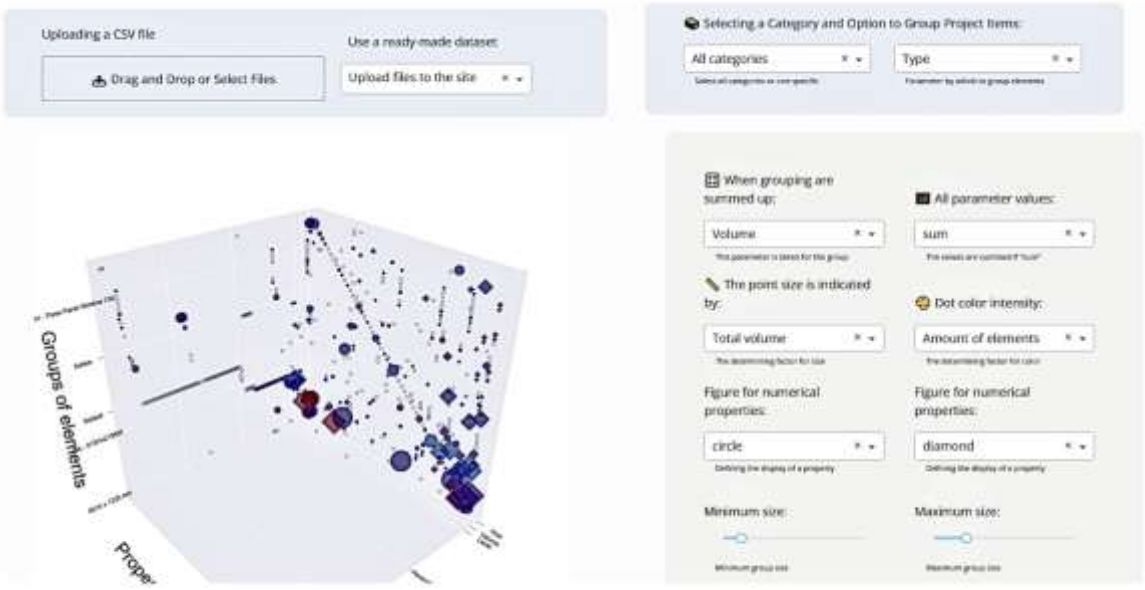


Fig. 7.2-12 Visualización interactiva en 3D de los atributos de los elementos de un proyecto CAD- (BIM-) mediante la biblioteca Plotly.

Las bibliotecas especializadas de código abierto Bokeh, Dash y Streamlit ofrecen una forma cómoda de presentar datos sin necesidad de tener profundos conocimientos de desarrollo web. Bokeh es adecuada para gráficos interactivos complejos, Dash se utiliza para crear cuadros de mando analíticos completos y Streamlit permite crear rápidamente aplicaciones web para el análisis de datos.

Con estas herramientas de visualización, los desarrolladores y analistas de pueden difundir eficazmente los resultados a colegas y partes interesadas, permitiendo una interacción intuitiva con los datos y simplificando la toma de decisiones.

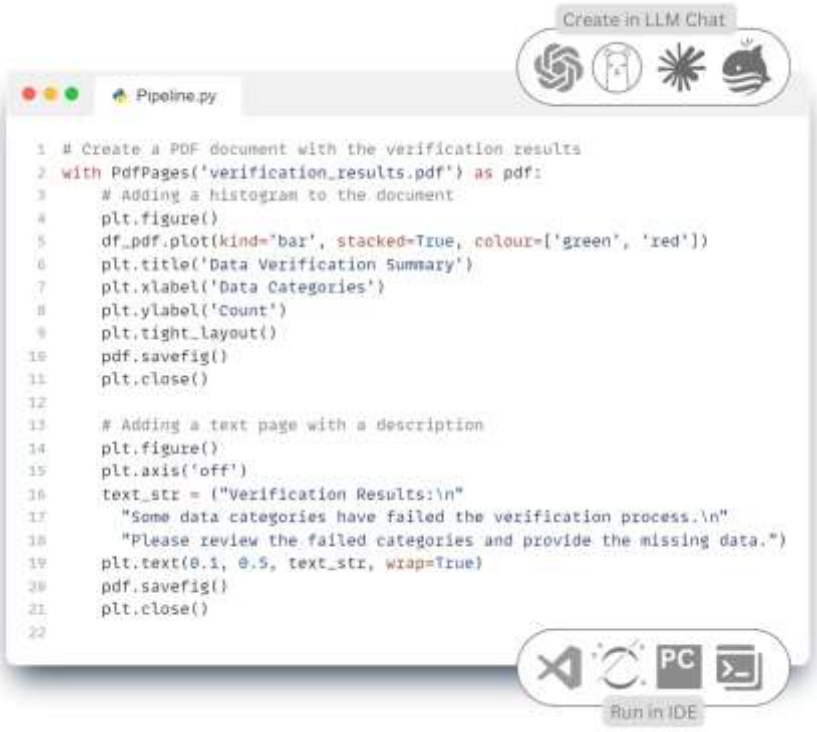
ETL Carga: Creación automática de documentos PDF

En la fase de carga de datos, no sólo se pueden visualizar los datos, cargarlos en tablas o bases de datos, sino también generar automáticamente informes que incluyan los gráficos, cuadros e indicadores analíticos clave necesarios para que los reciba el gestor o especialista que espera los resultados de la auditoría. Los informes automatizados pueden contener tanto comentarios como interpretaciones textuales de los datos y elementos de visualización: tablas, gráficos.

- Para crear un PDF -informe con un histograma (Fig. 7.2-9) y una descripción del análisis basado en la validación que hemos realizado en los capítulos anteriores, basta con formular una solicitud en una continuación del diálogo con el LLM, por ejemplo:

> Escriba código para crear un archivo PDF con un histograma y una descripción de los resultados de la validación de datos anterior (en el chat), y escriba un texto advirtiendo de que algunas categorías no han sido validadas y que es necesario rellenar los que faltan. ⏎

- Respuesta de LLM como código y PDF acabado con resultados:



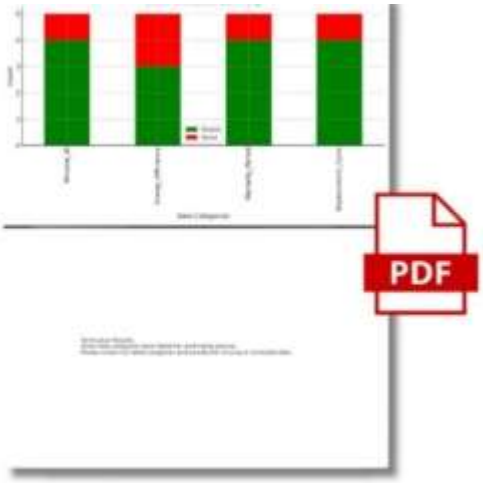


Fig. 7.2-13 El código automatizado crea un documento PDF que contiene un histograma con los datos de la prueba y un texto con los resultados de la misma.

Una solución automática de sólo 20 líneas de código que utiliza LLM crea instantáneamente el documento PDF (o DOC) deseado con visualización en forma de histograma de atributos (Fig. 7.2-13) que muestra el número de datos que han superado y no superado la validación, y con la adición de un bloque de texto que resume los resultados y las recomendaciones futuras acciones.

La generación automatizada de documentos es un elemento clave de la fase de carga, especialmente en un entorno de proyecto en el que la rapidez de los informes y la precisión son fundamentales.

ETL Carga: generación automática de documentos a partir de FPDF

Automatización de la generación de informes en la fase ETL Load es un paso importante en el tratamiento de datos, sobre todo cuando los resultados de los análisis deben presentarse en un formato fácil de comunicar y comprender. En el sector de la construcción, esto suele ser relevante para los informes de progreso, las estadísticas de datos de proyectos, los informes de control de calidad o la documentación financiera.

Una de las herramientas más cómodas para estas tareas es la biblioteca de código abierto, FPDF, disponible tanto para Python como para PHP.

> La biblioteca **FPDF** de código abierto ofrece una forma flexible de generar documentos mediante código, permitiendo añadir encabezados, texto, tablas e imágenes. El uso de código en lugar de la edición manual reduce los errores y agiliza el proceso de elaboración de informes en formato PDF.

Uno de los pasos clave en la creación de un documento PDF es añadir títulos y el texto principal en forma de comentarios o descripciones. Sin embargo, al crear un informe, no sólo es importante añadir texto, sino

también estructurarlo adecuadamente. Encabezamientos, sangrías, interlineado - todo esto afecta a la legibilidad del documento. Con FPDF, puede establecer parámetros de formato, controlar la disposición de los elementos y personalizar el estilo del documento.

FPDF es, en principio, muy similar a HTML. Quienes ya estén familiarizados con HTML pueden generar fácilmente documentos PDF de cualquier complejidad utilizando FPDF, ya que la estructura del código es muy similar a la del marcado HTML: los encabezados, el texto, las imágenes y las tablas se añaden de forma similar. Quienes no estén familiarizados con HTML no tienen por qué preocuparse: pueden utilizar LLM, que les ayudará instantáneamente a componer el código para generar el diseño de documento deseado.

- El siguiente ejemplo muestra cómo generar un informe con una cabecera y un cuerpo de texto. Ejecutando este código en cualquier IDE compatible con Python se crea un archivo PDF que contiene la cabecera y el texto deseados:

```
from fpdf import FPDF     # Importar la librería FPDF
pdf = FPDF()    # Crear PDF -documento
pdf.add_page()     # Añadir una página

pdf.set_font("Arial", style='B', size=16)  # Establecer fuente: Arial, negrita, tamaño 16
pdf.cell(200, 10, "Informe del proyecto", ln=True, align='C')  # Crear un encabezado y centrarlo
pdf.set_font("Arial", size=12)  # Cambiar la fuente a Arial normal, tamaño 12
pdf.multi_cell(0, 10, "Este documento contiene datos sobre los resultados de la verificación de los archivos del proyecto...")  # Añadir texto multilínea
pdf.output(r "C:\informes\informe.pdf") # Guardar PDF -archivo
```

Fig. 7.2-14 Con unas pocas líneas de código Python, podemos generar automáticamente el documento de texto PDF que necesitamos.

Al preparar informes, es importante tener en cuenta que los datos a partir de los cuales se forma el documento rara vez son estáticos. Las cabeceras, los bloques de texto (Fig. 7.2-14) suelen formarse dinámicamente, recibiendo valores en la etapa de Transformación del proceso ETL.

El uso del código permite crear documentos que contienen información actualizada: nombre del proyecto, fecha de generación del informe, así como información sobre los participantes o el estado actual. El uso de variables en el código permite insertar automáticamente estos datos en los lugares requeridos del informe, eliminando por completo la necesidad de editarlos manualmente antes de enviarlos.

Además del texto simple y los encabezados, las tablas ocupan un lugar especial en la documentación de los proyectos. Casi todos los documentos contienen datos estructurados: desde descripciones de objetos hasta resultados de inspecciones. La generación automática de tablas a partir de los datos de la fase Transform permite no sólo agilizar el proceso de preparación del documento, sino también minimizar los errores al transferir la información. FPDF permite insertar tablas en archivos PDF (como texto o imágenes), estableciendo los bordes de las celdas, el tamaño de las columnas y los tipos de letra (Fig. 7.2-15). Resulta especialmente práctico cuando se trabaja con datos dinámicos, en los que el número de filas y columnas puede variar en función de las tareas del documento.

- El siguiente ejemplo muestra cómo automatizar la creación de tablas, por ejemplo, con listas de materiales, estimaciones o resultados de pruebas de parámetros:

```
datos = [
    ["Artículo", "Cantidad", "Precio"],   # Encabezados de columna
    ["Hormigón", "10 m³", "$ 500."],   # Datos de la primera fila.
    ["Varilla", "2 toneladas", "$ 600"],   # Datos de la segunda fila.
    ["Ladrillo", "5.000 piezas", "$ 750."], # Datos de la tercera fila.
]

pdf = FPDF ()   # Crear PDF -documento
pdf.add_page() # Añadir una página
pdf.set_font("Arial", size=12)   # Establecer la fuente

for row in data:   # Buscar en las filas de la tabla
    for item in fila:   # Recorre las celdas de la fila
        pdf.cell(60, 10, item, border=1)   # Crea una celda con borde, ancho 60 y alto 10
    pdf.ln()   # Pasar a la línea siguiente
pdf.output(r "C:\informes\table.pdf")   # Guardar PDF -file
```

Item	Quantity	Price
Concrete	10 m³	$500
Rebar	2 t.	$600
Brick	5000 pcs.	$750

Fig. 7.2-15 Puede generar automáticamente no sólo texto sino también cualquier información de tabla del paso Transformar en PDF.

En los escenarios reales de elaboración de informes, las tablas suelen ser información generada dinámicamente que se obtiene en la fase de transformación de los datos. En el ejemplo dado (Fig. 7.2-15) la tabla se inserta en el documento PDF de forma estática: los datos para el ejemplo se colocaron en el diccionario de datos (primera línea del código), en condiciones reales dicha variable de datos se rellena automáticamente después de, por ejemplo, agrupar el marco de datos.

En la práctica, estas tablas se construyen a menudo a partir de datos estructurados procedentes de diversas fuentes dinámicas: bases de datos, archivos Excel, interfaces API o resultados de cálculos analíticos. La mayoría de las veces, en la fase de transformación (ETL), los datos se agregan, agrupan o filtran, y sólo después se transforman en totales en forma de gráficos o tablas bidimensionales que se muestran en los informes. Esto significa que el contenido de la tabla puede cambiar en función de los parámetros seleccionados, el periodo de análisis, los filtros del proyecto o la configuración del usuario.

> El uso de marcos de datos y conjuntos de datos dinámicos en la etapa Transformar hace que el proceso de elaboración de informes en la etapa Cargar sea lo más flexible, escalable y fácilmente repetible posible sin necesidad de intervención manual.

Además de tablas y texto, FPDF también permite añadir gráficos de datos tabulares, lo que permite incrustar en el informe imágenes generadas con Matplotlib u otras librerías de visualización de las que hemos hablado anteriormente. Cualquier gráfico, tabla o diagrama puede añadirse al documento utilizando el código.

Utilizando la biblioteca de Python FPDF, vamos a añadir un gráfico pregenerado con Matplotlib. al documento PDF:

```
import matplotlib.pyplot as plt  # Importar Matplotlib para crear gráficos

fig, ax = plt.subplots()  # Crear la figura y los ejes del gráfico
categories = ["Concrete", "Rebar", "Brick"]  # Nombres de las categorías
values = [50000, 60000, 75000]  # Valores de las categorías
ax.bar(categories, values)  # Crear un gráfico de barras
plt.ylabel("Valor,$.")  # Señalar el eje Y
plt.title("Distribución de costes")  # Añadir un título
plt.savefig(r "C:\reports\chart\chart\chart.png")  # Guardar el gráfico como imagen

pdf = FPDF ()  # Crear PDF -documento
pdf.add_page() # Añadir una página
pdf.set_font("Arial", size=12)  # Establecer el tipo de letra
pdf.cell(200, 10, "Gráfico de costes", ln=True, align='C')  # Añadir una cabecera

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100)  # Insertar la
imagen en el PDF (x, y - coordenadas, w - anchura)
pdf.output(r "C:informes_gráficos_informe.pdf")  # Guardar el archivo PDF
```

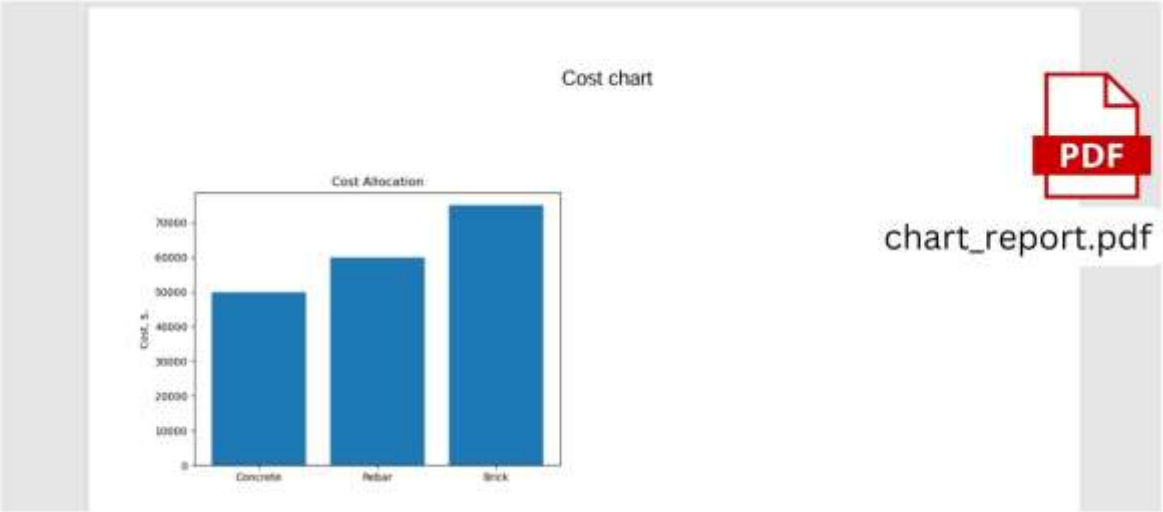


Fig. 7.2-16 Con una docena de líneas de código, puede generar un gráfico, guardarlo y, a continuación, pegarlo en un documento PDF.

FPDF hace que el proceso de preparación y lógica de documentos sea transparente, rápido y cómodo. Las plantillas integradas en el código permiten generar documentos con datos actualizados, eliminando la necesidad de rellenarlos manualmente.

> Uso de la automatización ETL: en lugar de realizar informes manuales que requieren mucho tiempo, los profesionales pueden centrarse en analizar los datos y tomar decisiones, en lugar de elegir la herramienta adecuada para trabajar con un silo de datos concreto con una interfaz de usuario clara.

De este modo, la biblioteca FPDF proporciona una herramienta flexible para la creación automatizada de documentos de cualquier complejidad, desde breves informes técnicos hasta complejos resúmenes analíticos con tablas y gráficos, lo que permite no sólo agilizar el flujo de documentos, sino también reducir significativamente la probabilidad de errores asociados a la introducción manual de datos y al formateo.

ETL Load: Informes y carga a otros sistemas

En la fase de carga, los resultados se generaron en forma de tablas, gráficos e informes finales en PDF elaborados de acuerdo con los requisitos establecidos. Además, es posible exportar estos datos a formatos legibles por máquina (por ejemplo, CSV), necesarios para la integración con sistemas externos como ERP, CAFM, CPM, plataformas de BI y otras soluciones corporativas o industriales. Además de CSV, las cargas pueden realizarse a XLSX, JSON, XML o directamente a bases de datos que admitan el intercambio automático de información.

- Para generar el código apropiado para automatizar el paso Load, basta con consultar la interfaz LLM, por ejemplo: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude o QWEN:

> Escriba código para generar un informe de resultados de validación de datos en DataFrame, donde las columnas prefijadas con 'verified_' se cuentan, se renombran a 'Passed' y 'Failed', los valores que faltan se sustituyen por 0, y luego sólo las filas que pasan todas las validaciones se exportan a un archivo CSV.

Respuesta de LLM:

Fig. 7.2-17 Los datos validados obtenidos en el paso Transformar del marco de datos final se exportan a un archivo CSV para su integración con otros sistemas.

El código anterior (Fig. 7.2-17) implementa la etapa final del ETL -proceso - Carga, durante la cual los datos verificados se guardan en formato CSV, compatible con la mayoría de los sistemas y bases de datos externos. De este modo, hemos completado el ciclo completo del proceso ETL, que incluye la extracción, transformación, visualización, documentación y exportación de los datos a los sistemas y formatos que necesitemos, lo que garantiza la reproducibilidad, transparencia y automatización del trabajo con la información.

El canal ETL puede utilizarse tanto para procesar proyectos individuales como para aplicaciones a gran escala, al analizar cientos o miles de datos entrantes en forma de documentos, imágenes, escaneados, proyectos CAD, nubes de puntos, archivos PDF u otras fuentes procedentes de sistemas distribuidos. La capacidad de automatizar totalmente el proceso hace del ETL no sólo una herramienta de procesamiento técnico, sino la base de una infraestructura de información digital para la construcción.

ETL con LLM: Visualización de datos a partir de documentos PDF

Es hora de pasar a construir un proceso ETL completo que cubra todas las etapas clave del tratamiento de datos en un único escenario: extracción, transformación y carga. Construyamos una línea de producción ETL automatizada que le permita procesar documentos PDF sin trabajo manual: extraiga datos de los documentos, visualícelos, analícelos y transfiéralos a otros sistemas.

El proceso ETL de nuestro ejemplo se describirá mediante prompts, que tendrán que explicar al modelo de lenguaje (LLM) todos los procesos ETL con una descripción del resultado final que debe obtenerse. En este caso, la tarea consiste en encontrar todos los archivos PDF de la carpeta especificada y sus subcarpetas, extraer de ellos la información pertinente -por ejemplo, nombres de materiales, cantidades y costes- y presentar el resultado como una tabla estructurada (DataFrame) para su posterior análisis

- Primera consulta de texto en LLM para extraer automáticamente datos de varios documentos PDF y crear un marco de datos de los datos para el paso Extraer:

> Escriba código para extraer información sobre materiales de los archivos PDF de una carpeta determinada y sus subcarpetas. Los datos del PDF incluyen el nombre del material, la cantidad y el coste. El resultado debe guardarse en un DataFrame ⏎

- Respuesta de LLM:

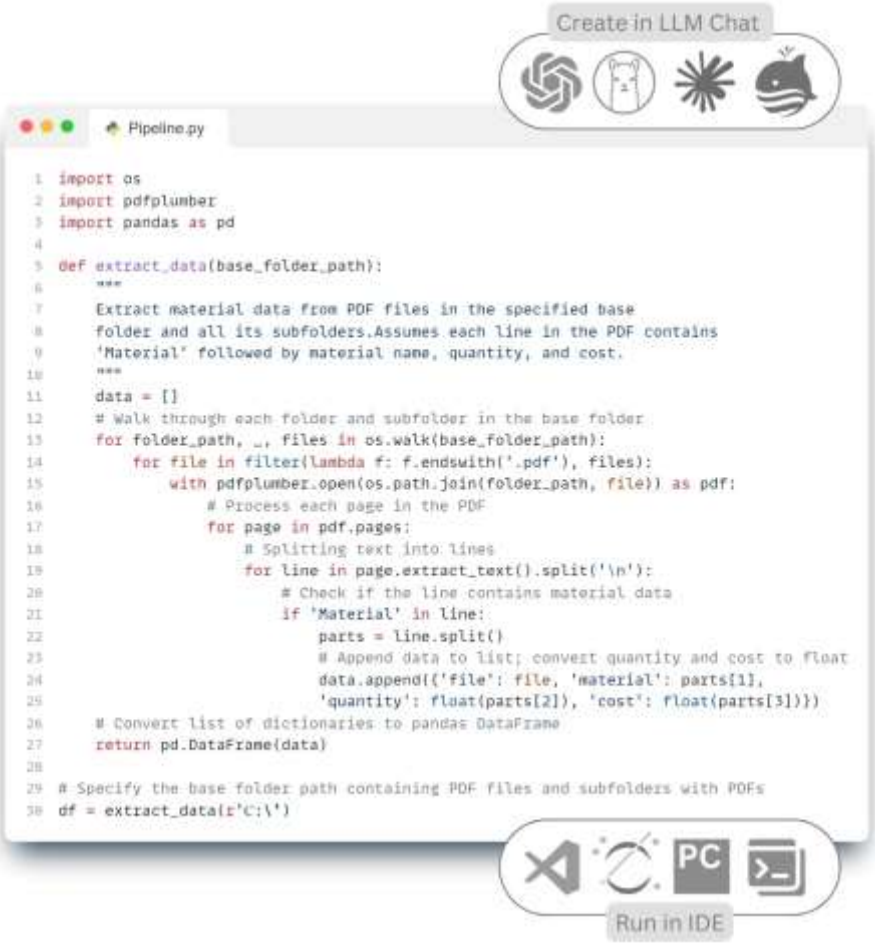


Fig. 7.2-18 LLM crea código Python -para extraer datos de archivos PDF -en una carpeta específica y todas sus subcarpetas.

La respuesta LLM (Fig, 7.2-18) es un script de Python listo para usar que recorre automáticamente todas las carpetas, abre los archivos PDF encontrados, extrae información textual de ellos y la convierte en una tabla. El código de chat resultante puede ejecutarse en uno de los IDEs populares PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con el plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con el plugin Python, JupyterLab o las populares herramientas online Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternativamente, para simplificar el proceso de recopilación, en lugar de copiar el código de la LLM y utilizar el código en el IDE, también podemos cargar una docena de archivos PDF directamente en el chat de LLM (Fig, 7.2-19) y obtener una tabla como salida, sin tener que ver el código ni ejecutarlo. El resultado de ejecutar este código será una tabla con los atributos que hayamos seleccionado.

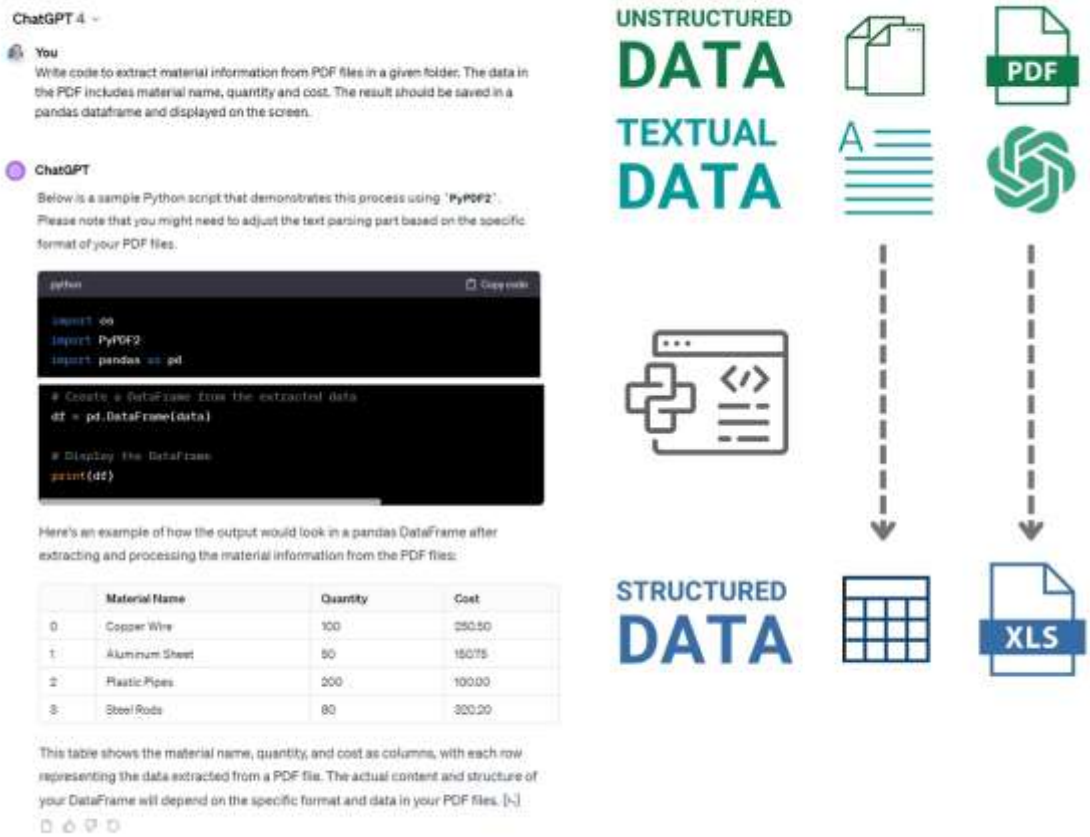


Fig. 7.2-19 Resultado de la ejecución de código en LLM, que extrae datos de archivos PDF en una vista de marco de datos estructurado con atributos seleccionados.

En el siguiente paso, solicitamos un modelo lingüístico sobre los datos obtenidos, por ejemplo, para comparar el coste y el volumen de uso de materiales y crear algunas visualizaciones de ejemplo que sirvan de base para análisis posteriores.

- Pida en una charla continua con el LLM que trace algunos gráficos a partir de las tablas que se produjeron en el paso Transformar (Fig, 7.2-18):

Visualice el coste total y la cantidad de cada material desde DataFrame (Fig, 7.2-18⏎

Create in LLM Chat

Pipeline.py

```python
import matplotlib.pyplot as plt

# Plotting total cost per material
# Group data by 'material' and sum up the 'cost' for each material
df.groupby('material')['cost'].sum().plot(kind='bar')
plt.title('Total Cost per Material')
plt.xlabel('Material')
plt.ylabel('Total Cost')
plt.show()

# Plotting total quantity of materials used
# Group data by 'material' and sum up the 'quantity' for each material
df.groupby('material')['quantity'].sum().plot(kind='barh')
plt.title('Total Quantity of Materials Used')
plt.xlabel('Quantity')
plt.ylabel('Material')
plt.show()
```

Run in IDE

Fig. 7.2-20 Respuesta LLM -modelos como código Python para visualizar datos de un marco de datos utilizando la biblioteca matplotlib.

LLM genera y ejecuta automáticamente código Python (Fig. 7.2-20) utilizando la biblioteca matplotlib. Tras ejecutar este código, obtenemos gráficos de costes y uso de materiales en proyectos de construcción directamente en chat (Fig. 7.2-21), lo que simplifica enormemente el trabajo analítico.

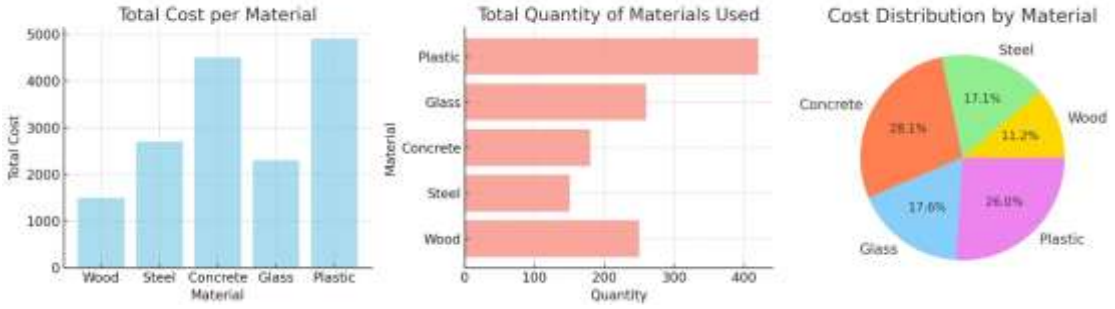


Fig. 7.2-21 Visualización de la respuesta LLM como gráficos basados en los datos recogidos en el DataFrame.

La ayuda en el desarrollo de ideas para escribir código ETL, analizar y ejecutar código, visualizar resultados

está disponible a través de simples consultas de texto en LLM, sin necesidad de aprender los fundamentos de la programación. La aparición de herramientas de IA como LLM está cambiando definitivamente el enfoque de la programación y la automatización del procesamiento de datos (Fig, 7.2-22).

> Según el informe de PwC "¿Cuál es el valor real de la inteligencia artificial para su empresa y cómo puede capitalizarlo?" (2017) [139], la automatización de procesos y las mejoras de la productividad serán los principales motores del crecimiento económico. Y se espera que las mejoras de productividad representen más del 55% de todo el crecimiento del PIB impulsado por la IA entre 2017 y 2030."

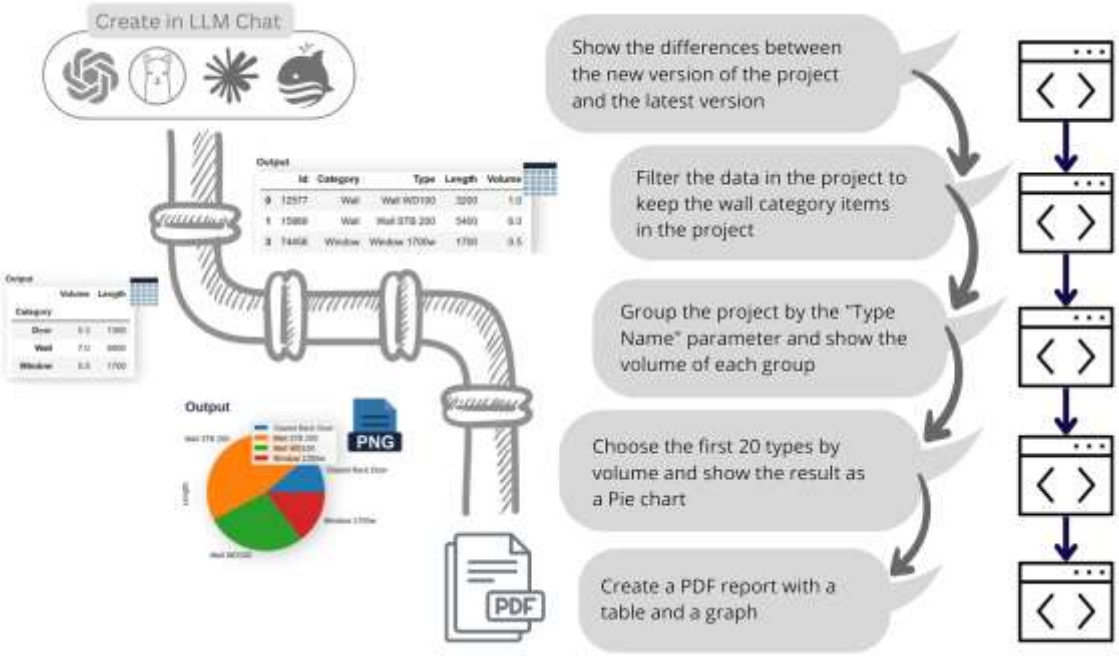


Fig. 7.2-22 AI LLM ayuda a generar borradores de código que se aplican a futuros proyectos sin necesidad de un LLM.

> Con herramientas como ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, así como datos abiertos y software de código abierto, podemos automatizar procesos que antes sólo se hacían con sistemas propietarios modulares especializados, de alto coste y difícil mantenimiento.

En el contexto de la construcción, esto significa que las empresas que sean las primeras en implantar procesos automatizados de Pipeline -data cosecharán importantes beneficios, desde la mejora de la eficiencia en la gestión de proyectos hasta la reducción de pérdidas financieras, pasando por la eliminación de aplicaciones fragmentadas y almacenes de datos aislados.

La lógica descrita de ejecución de tareas de negocio en el proceso ETL es una parte crucial de la automatización de la analítica y de los procesos de procesamiento de datos, que es una variación específica de un concepto más amplio - pipelines (Tuberías).

CAPÍTULO 7.3.
TRANSPORTADOR AUTOMÁTICO ETL (TUBERÍA)

Canalización: ETL automático transportador datos

El proceso ETL se ha utilizado tradicionalmente para procesar datos en sistemas analíticos, abarcando tanto fuentes estructuradas como no estructuradas. Sin embargo, en el entorno digital actual se utiliza cada vez más un término más amplio: Pipeline (transportador), que describe cualquier cadena secuencial de procesamiento en la que el resultado de una etapa se convierte en la entrada de la siguiente.

Este enfoque se aplica no sólo a los datos, sino también a otros tipos de automatización: procesamiento de tareas, creación de informes, integración con software y flujo de trabajo digital (Fig. 7.3-1).

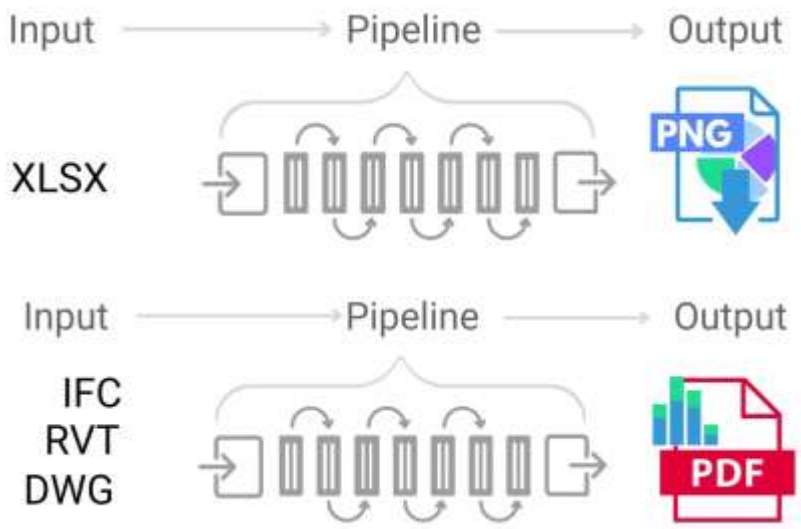


Fig. 7.3-1 Pipeline es una secuencia de procesamiento en la que la salida de una etapa se convierte en la entrada de la etapa siguiente.

El uso de Pipeline es uno de los elementos principales de la automatización, especialmente cuando se trabaja con grandes cantidades de datos heterogéneos. La arquitectura pipeline permite organizar pasos de procesamiento complejos en un formato modular, coherente y manejable, lo que aumenta la legibilidad, simplifica el mantenimiento del código y permite la depuración incremental y la realización de pruebas escalables.

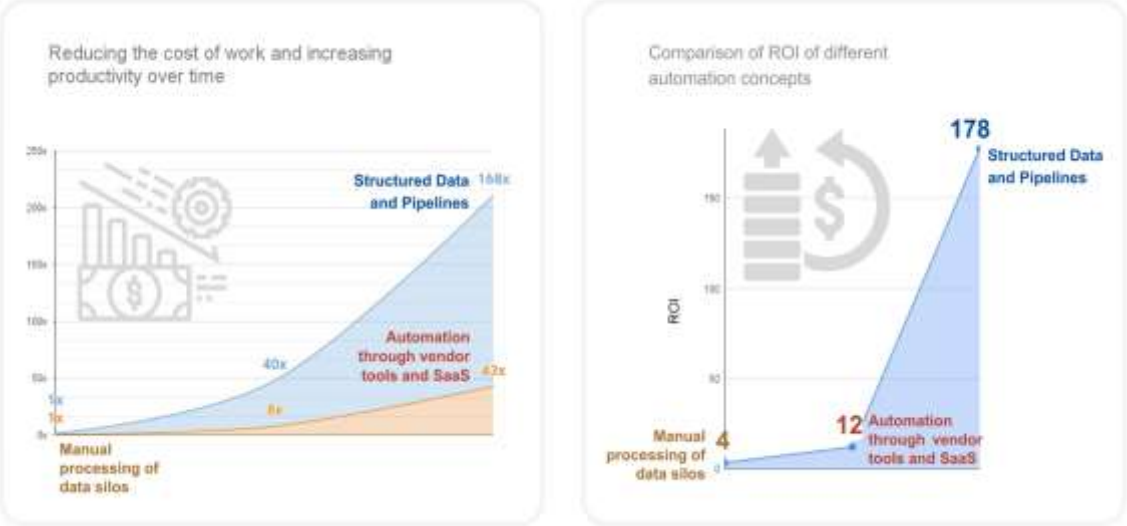


Fig. 7.3-2 ROI Pipeline proceso de validación de datos reduce el tiempo de ejecución en decenas y cientos de veces en comparación con el procesamiento mediante herramientas clásicas [74].

A diferencia del trabajo manual en sistemas propietarios (ERP, PMIS, CAD, etc.), el pipelining permite aumentar considerablemente (Fig. 7.3-2) la velocidad de las tareas, evitar el trabajo repetitivo y automatizar el inicio de los procesos en el momento adecuado (Fig. 7.3-3).

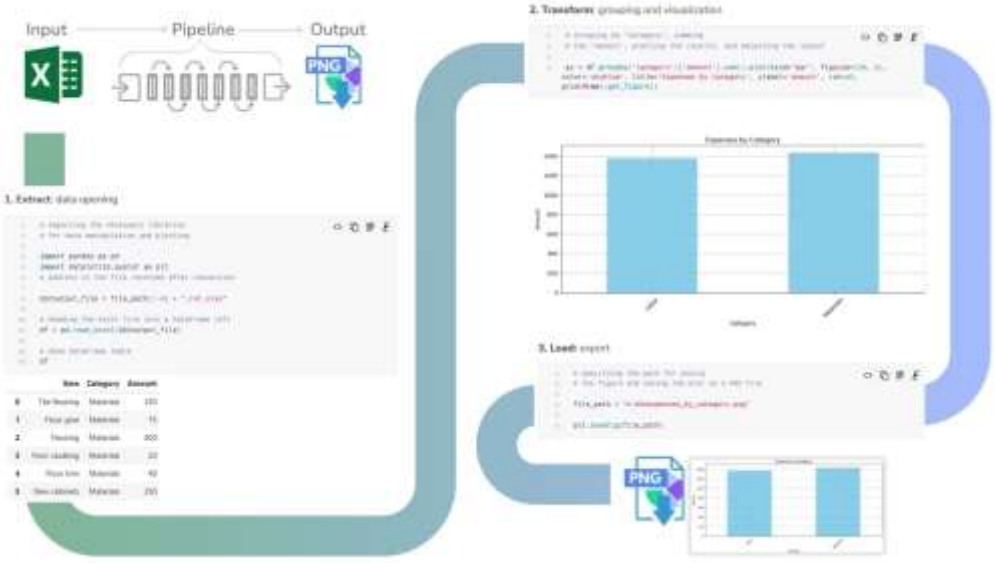


Fig. 7.3-3 Ejemplo ETL Pipeline sobre la obtención automática de un gráfico a partir de datos tabulares en un archivo XLSX sin abrir Excel.

Para procesar datos en flujo y crear una canalización automatizada, de forma similar al proceso ETL, es necesario identificar previamente las fuentes de datos, así como el plazo para recopilarlos, ya sea para un proceso empresarial específico o para toda la empresa.

En los proyectos de construcción, los datos proceden de muchas fuentes heterogéneas con distintos intervalos de actualización. Para crear un escaparate de datos fiable, es fundamental registrar cuándo se recupera y actualiza la información. Esto permite tomar decisiones a tiempo y mejora la eficacia de la gestión de proyectos.

Una opción es iniciar el proceso de ensamblaje a una hora fija, por ejemplo, a las 19:00, al final de la jornada laboral. En ese momento, se activa el primer script encargado de agregar los datos procedentes de diversos sistemas y almacenes (Fig. 7.3-4 paso 1). A continuación, se automatiza el procesamiento y la transformación de los datos en un formato estructurado adecuado para el análisis (Fig. 7.3-4 paso 2-4). En la etapa final, utilizando los datos preparados, se generan automáticamente informes, cuadros de mando y otros productos descritos en los capítulos anteriores (Fig. 7.3-4 paso 6-7). Como resultado, a las 05:00 de la mañana, los gestores ya disponen de informes actualizados sobre el estado del proyecto en el formato requerido (Fig. 7.3-5).

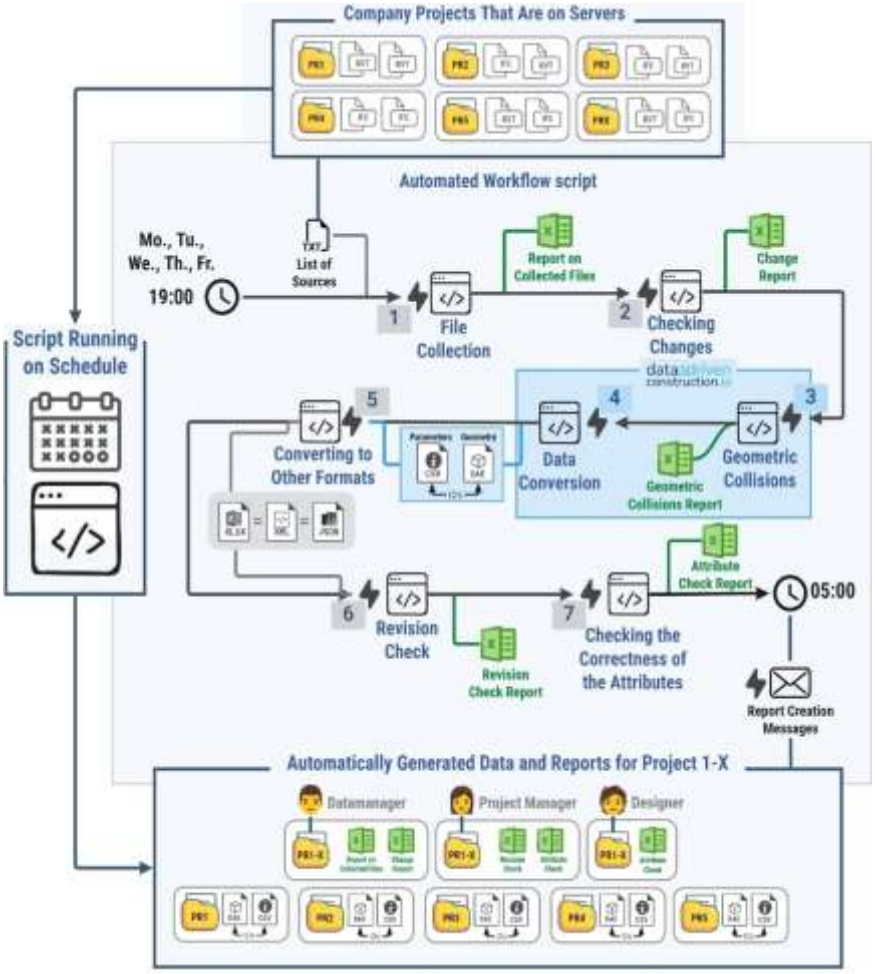


Fig. 7.3-4 Los datos de Pipeline, recogidos automáticamente por la tarde, se procesan durante la noche para que los gestores dispongan de informes actualizados e informes frescos por la mañana.

> La recopilación oportuna de datos, la definición de KPI, la automatización de los procesos de transformación y la visualización a través de cuadros de mando son elementos clave para el éxito de la toma de decisiones basada en datos.

Estos procesos automatizados (Fig. 7.3-4) pueden ejecutarse con total autonomía: se ejecutan según un calendario, procesan los datos sin supervisión y pueden desplegarse en la nube o en el propio servidor de la empresa (Fig. 7.3-5). Esto permite integrar estos ETL pipelines en la infraestructura informática existente, manteniendo el control sobre los datos y proporcionando flexibilidad en el escalado.

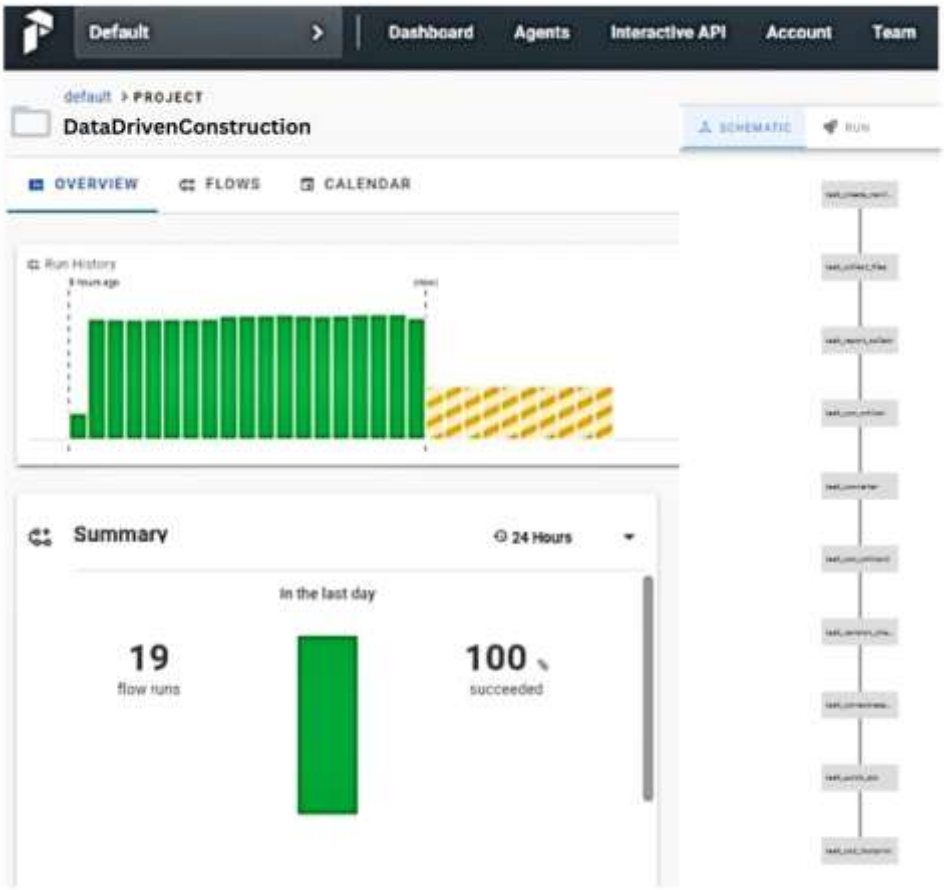


Fig. 7.3-5 Procesos automáticos ETL -conveyor (Fig. 7.3-4) en la plataforma Prefect, en la que 10 scripts python se ejecutan alternativamente después de las 19:00 todos los días laborables.

La automatización de los flujos de trabajo no sólo aumenta la productividad del equipo al liberar tiempo para tareas más significativas y menos rutinarias, sino que también sirve como un primer paso importante hacia la incorporación de tecnologías de inteligencia artificial (AI) en los procesos empresariales, de las que hablaremos con más detalle en el capítulo sobre análisis predictivo y aprendizaje automático.

Tubería -ETL proceso de validación de datos con LLM

En los capítulos anteriores sobre creación de requisitos de datos y automatización de ETL, desglosamos paso a paso el proceso de preparación, transformación, validación y visualización de datos. Estas actividades se implementaron como bloques de código independientes (Fig, 7.2-18 - Fig. 7.2-20), cada uno de los cuales realiza una tarea específica.

Ahora nos queda el siguiente objetivo: combinar estos elementos en un único, coherente y automatizado pipeline de procesamiento de datos - un pipeline, ETL -Pipeline - en el que todas las etapas (carga, validación, visualización, exportación) se ejecutan secuencialmente en un único script autoejecutable.

En el siguiente ejemplo, se implementará un ciclo completo de procesamiento de datos: desde la carga del archivo CSV de origen →hasta la comprobación de la estructura y los valores mediante expresiones regulares →el cálculo de los resultados→ la generación de un informe visual en formato PDF.

Puede utilizar la siguiente consulta de texto a la LLM. para recuperar el código apropiado:

> Por favor, escriba un ejemplo de código que cargue datos desde CSV, valide DataFrame usando expresiones regulares, compruebe identificadores en formato 'W-NEW' o 'W-OLD', eficiencia energética con letras 'A' a 'G', periodo de garantía y ciclo de sustitución con valores numéricos en años y al final cree un informe con un recuento de valores aprobados y fallidos, genere un PDF con un histograma de resultados y añada una descripción de texto. ⏎

Respuesta de LLM:

Create in LLM Chat

Pipeline.py

```python
# Importing necessary libraries
import pandas as pd
import re
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages


# Load data
df = pd.read_csv('path_to_csv_file.csv')

# Define patterns and verify data
patterns = {
    'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
    'Energy_Efficiency': r'^[A-G]\+*$',
    'Warranty_Period': r'^\d+\syears$',
    'Replacement_Cycle': r'^\d+\syears$'
}

for column, pattern in patterns.items():
    df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))

# Create report
report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
report.columns = ['Passed', 'Failed']

# Generate PDF
pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
with PdfPages(pdf_path) as pdf:
    # Plot and save bar chart
    report.plot(kind='bar', stacked=True, color=['green', 'red'])
    plt.title('Data Verification Summary')
    plt.xlabel('Data Categories')
    plt.ylabel('Count')
    plt.tight_layout()
    pdf.savefig()
    plt.close()

    # Add text page
    plt.figure()
    plt.axis('off')
    text_str = "Verification Results Summary"
    plt.text(0.1, 0.5, text_str, wrap=True)
    pdf.savefig()
    plt.close()

pdf_path

```

Run in IDE

Fig. 7.3-6 Pipeline (ETL) automatiza el ciclo completo de procesamiento de datos: desde la descarga y la validación hasta la creación de un informe estructurado en formato PDF.

El código automatizado (Fig. 7.3-6) dentro de la sala de chat LLM o en DIE, después de copiar el código, validará los datos del archivo CSV utilizando las expresiones regulares especificadas, creará un informe sobre el número de registros aprobados y no aprobados y, a continuación, guardará los resultados de la validación como un archivo PDF.

Esta estructura ETL -conveyor structure, en la que cada paso -desde la carga de datos hasta la generación de informes- se implementa como un módulo independiente, garantiza la transparencia, la escalabilidad y la reproducibilidad. La presentación de la lógica de validación como un código Python fácil de leer hace que el proceso sea transparente y comprensible no solo para los desarrolladores, sino también para los especialistas en gestión de datos, calidad y análisis.

> El enfoque de Pipeline para automatizar el procesamiento de datos permite estandarizar los procesos, aumentar su repetibilidad y simplificar la adaptación a nuevos proyectos. De este modo se crea una metodología unificada para analizar los datos, independientemente de la fuente o el tipo de tarea, ya se trate de pruebas de conformidad, elaboración de informes o transferencia de datos a sistemas externos.

Esta automatización reduce el error humano, disminuye la dependencia de soluciones patentadas y aumenta la precisión y fiabilidad de los resultados, lo que las hace adecuadas tanto para la analítica operativa a nivel de proyecto como para la analítica estratégica a nivel de empresa.

Tubería -ETL: verificación de datos e información de los elementos del proyecto en CAD (BIM)

Los datos de los sistemas CAD y las bases de datos (BIM) son algunas de las fuentes de datos más sofisticadas y actualizadas dinámicamente en el negocio de las empresas de construcción. Estas aplicaciones no solo describen el proyecto mediante la geometría, sino que la complementan con información textual de varias capas: volúmenes, propiedades de los materiales, asignación de espacios, niveles de eficiencia energética, tolerancias, expectativas de vida útil y otros atributos.

> Los atributos asignados a las entidades en los modelos CAD se forman en la fase de diseño y se convierten en la base de otros procesos empresariales, como el cálculo de costes, la programación, la evaluación del ciclo de vida y la integración con los sistemas ERP y CAFM, en los que la eficacia de los procesos depende en gran medida de la calidad de los datos procedentes de los departamentos de diseño.

El enfoque tradicional para la validación de atributos en modelos CAD- (BIM-) implica la validación manual (Fig. 7.2-1), que se convierte en un proceso largo y costoso cuando el volumen de modelos es grande. Teniendo en cuenta el volumen y el número de proyectos de construcción modernos y sus actualizaciones periódicas, el proceso de validación y transformación de datos resulta insostenible e inasequible.

Los contratistas generales y los gestores de proyectos se enfrentan a la necesidad de procesar grandes cantidades de datos de proyectos, incluidas múltiples versiones y fragmentos de los mismos modelos. Los datos proceden de organizaciones de diseño en formatos RVT, DWG, DGN, IFC, NWD y otros (Fig, 3.1-14) y requieren comprobaciones periódicas para verificar su conformidad con las normas industriales y corporativas

La dependencia de acciones manuales y software especializado convierte el proceso de validación de datos en un cuello de botella en los flujos de trabajo relacionados con datos procedentes de modelos de toda la empresa. La automatización y el uso de requisitos estructurados pueden eliminar esta dependencia, aumentando drásticamente la velocidad y fiabilidad de la validación de datos (Fig, 7.3-7).

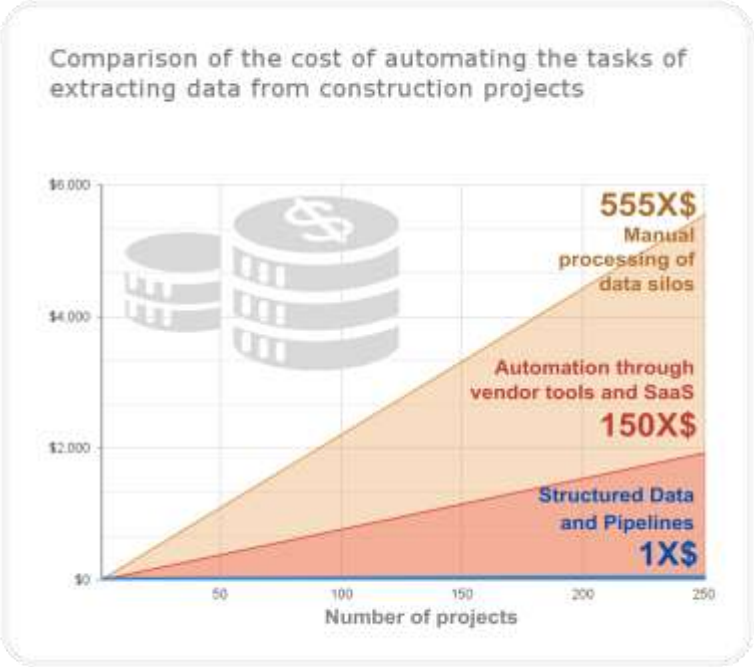


Fig. 7.3-7 La automatización aumenta la velocidad de verificación y tratamiento de los datos, lo que reduce decenas de veces el coste del trabajo [140].

El proceso de validación de datos CAD incluye la extracción de datos (etapa ETL Extract) a partir de diversos formatos cerrados (RVT, DWG, DGN, NWS, etc.) o abiertos semiestructurados y paramétricos (IFC, CPXML, USD).) o abiertos semiestructurados y paramétricos (IFC, CPXML, USD), en los que se pueden aplicar tablas de reglas a cada atributo y sus valores (etapa Transformar) mediante expresiones regulares RegEx (Fig. 7.3-8), proceso que tratamos en detalle en la cuarta parte del libro.

La creación de un informe de errores en PDF de y de los registros validados correctamente debe completarse con la salida (paso Cargar) en formatos estructurados que sólo tengan en cuenta las entidades validadas que puedan utilizarse para procesos posteriores.

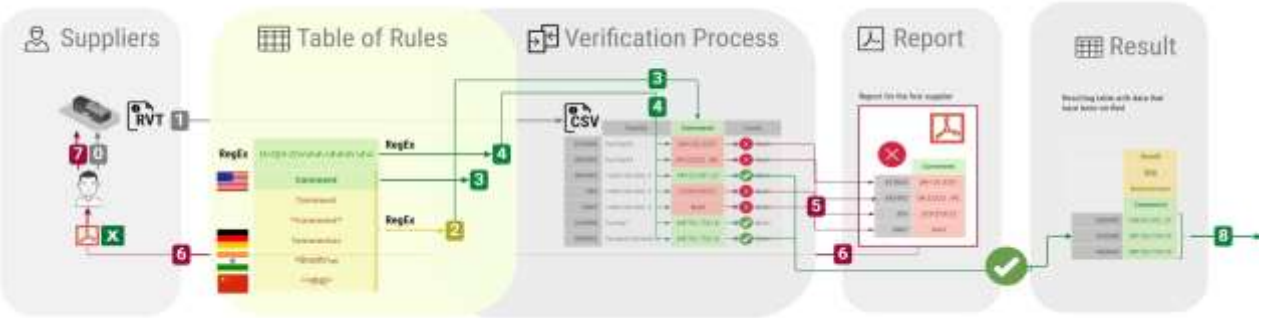


Fig. 7.3-8 Proceso de validación de datos desde los proveedores de datos del proyecto hasta el informe final validado mediante expresiones regulares.

La automatización de la validación de datos procedentes de sistemas CAD (BIM) con requisitos estructurados y el flujo de nuevos datos que se procesan a través de ETL-Pipelines (Fig, 7.3-9) reduce la necesidad de intervención manual en el proceso de validación (cada uno de los procesos de validación y requisitos de datos se han tratado en capítulos anteriores).

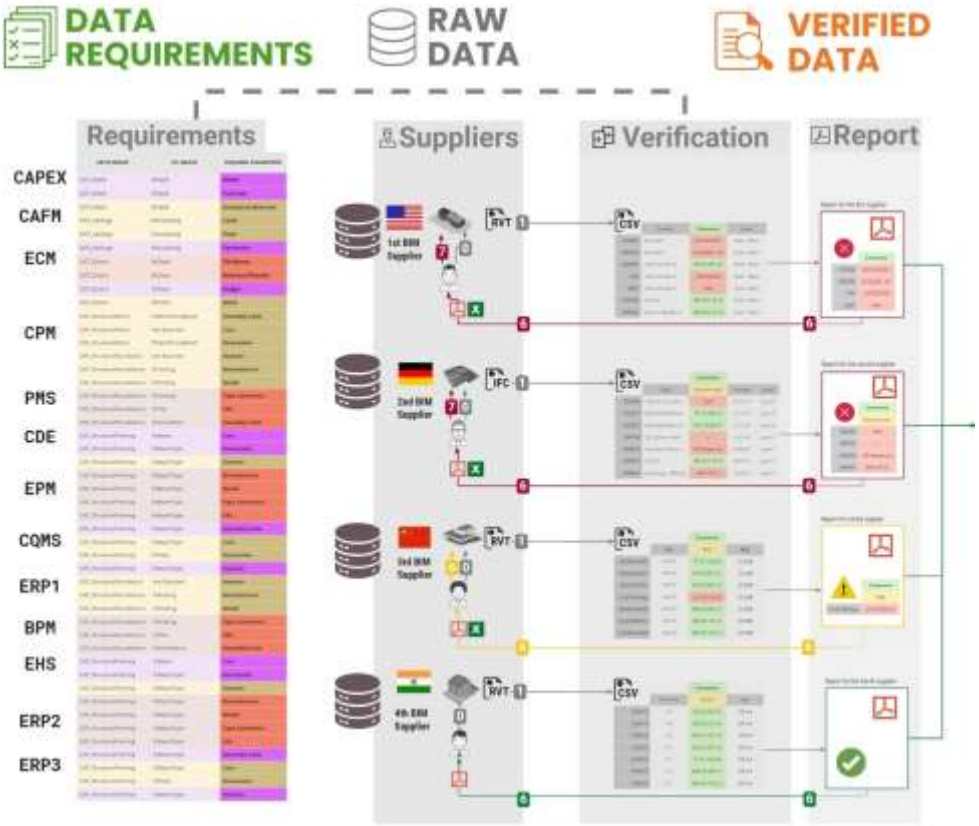


Fig. 7.3-9 La automatización de la validación de datos mediante ETL simplifica la gestión de los proyectos de construcción al agilizar los procesos.

Tradicionalmente, la validación de modelos proporcionados por contratistas y especialistas en CAD (BIM) puede llevar de días a semanas. Sin embargo, con la introducción de procesos ETL automatizados, esto puede reducirse a unos minutos. En una situación típica, el contratista afirma: *"El modelo está validado y es conforme"*. Esta declaración inicia una cadena de verificación de la afirmación del contratista sobre la calidad de los datos:

- Jefe de proyecto - "*El contratista afirma: 'El modelo ha sido probado, todo está bien'*".
- Gestor de datos - Validación de la carga:
  - Un simple script en Pandas detecta una infracción en segundos. La automatización elimina las disputas:
    - Categoría: OST_StructuralColumns, Parámetro: FireRating IS NULL.
    - Generar una lista de ID de infracciones→ exportar a Excel/PDF.

  Un sencillo script en Pandas detecta la brecha en cuestión de segundos:

```
df = datos_modelo[datos_modelo["Categoría"] == "OST_Columnas_estructurales"]  #
Filtrado
issues = df[df["FireRating"].isnull()]  # Valores vacíos
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Exportar IDs
```

- Responsable de datos a responsable de proyecto - *"Una comprobación de muestra que 18 columnas no tienen rellenado el parámetro FireRating"*.
- Jefe de proyecto a contratista - "El *modelo se devuelve para revisión: el parámetro FireRating es obligatorio, sin él la aceptación es imposible"*

Como resultado, el modelo CAD no se somete a validación, la automatización elimina las disputas y el contratista recibe casi instantáneamente un informe estructurado con una lista de identificaciones de los elementos problemáticos. De este modo, el proceso de validación se vuelve transparente, repetible y protegido de errores humanos (Fig. 7.3-10).

> Este enfoque convierte el proceso de validación de datos en una función de ingeniería en lugar de un control de calidad manual. Esto no solo aumenta la productividad, sino que también permite aplicar la misma lógica a todos los proyectos de la empresa, lo que posibilita la transformación digital integral de los procesos, desde el diseño hasta las operaciones.

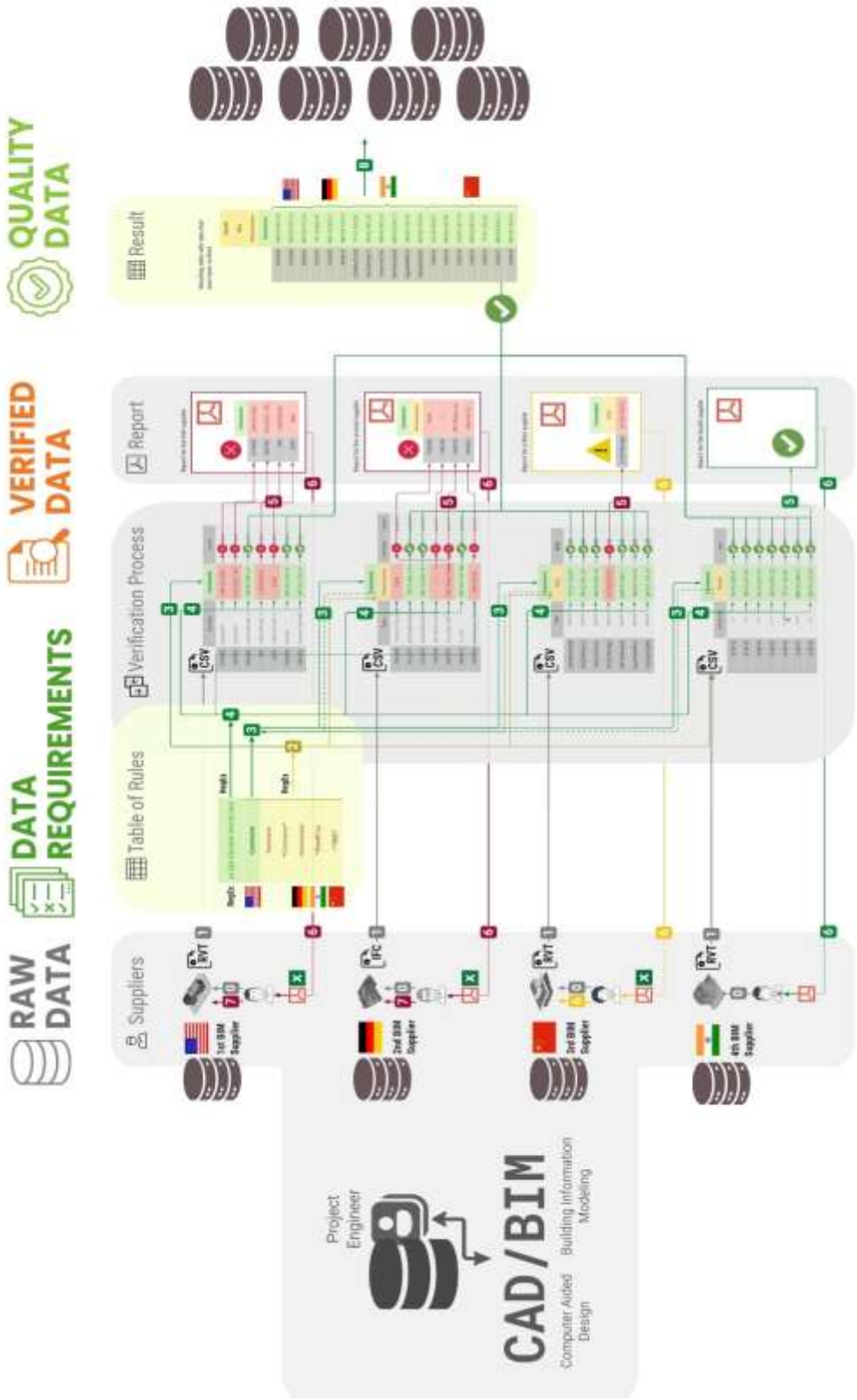


Fig. 7.3-10 La automatización de la comprobación de los atributos de los elementos elimina el error humano y reduce la probabilidad de errores.

Mediante el uso de canalizaciones automatizadas (Fig. 7.3-10), los usuarios del sistema que esperan datos de calidad de los sistemas CAD- (BIM-) pueden obtener instantáneamente los datos de salida que necesitan - tablas, documentos, imágenes - e integrarlos rápidamente en sus tareas de trabajo.

> La automatización del control, el procesamiento y el análisis está impulsando un cambio en la forma de enfocar la gestión de los proyectos de construcción, especialmente la interoperabilidad de los distintos sistemas, sin recurrir a complejos y costosos sistemas modulares patentados o a soluciones cerradas de proveedores.

Aunque los conceptos y las siglas de marketing van y vienen, los propios procesos de validación de requisitos de datos seguirán siendo para siempre parte integrante de los procesos empresariales. En lugar de crear formatos y normas cada vez más especializados, el sector de la construcción debería recurrir a herramientas que ya han demostrado su eficacia en otras industrias. Hoy en día, existen potentes plataformas para automatizar el tratamiento de datos y la integración de procesos que permiten a las empresas reducir significativamente el tiempo dedicado a operaciones rutinarias y minimizar los errores en Extraer, Transformar y Cargar.

Uno de los ejemplos más populares de soluciones para la automatización y orquestación de procesos ETL es Apache Airflow, que permite organizar procesos computacionales complejos y gestionar pipelines ETL. Junto con Airflow, también se utilizan activamente otras soluciones similares como Apache NiFi para el enrutamiento y streaming de datos y n8n para la automatización de procesos empresariales.

CAPÍTULO 7.4.
ORQUESTACIÓN DE ETL Y FLUJOS DE TRABAJO: SOLUCIONES PRÁCTICAS

DAG y Apache Airflow: automatización y orquestación de flujos de trabajo

Apache Airflow es una plataforma gratuita y de código abierto, diseñada para automatizar, orquestar y supervisar flujos de trabajo (ETL -conveyors).

Cada día hay que trabajar con grandes cantidades de datos:

- Descargar archivos de distintas fuentes - Extraer (por ejemplo, de proveedores o clientes).
- Transformar estos datos en el formato requerido - Transformar (estructurar, limpiar y validar)
- Enviar resultados para su verificación y crear informes - Cargar (cargar en los sistemas, documentos, bases de datos o cuadros de mando necesarios.

La ejecución manual de estos procesos ETL requiere un tiempo considerable y conlleva el riesgo de error humano. Un cambio en la fuente de datos o un fallo en uno de los pasos puede provocar retrasos y resultados incorrectos.

Las herramientas de automatización, como Apache Airflow, permiten construir un ETL fiable -conveyor, minimizar los errores, reducir el tiempo de procesamiento y garantizar que los datos son correctos en cada etapa. En el corazón de Apache Airflow se encuentra el concepto de DAG (Directed Acyclic Graph) - un grafo acíclico dirigido en el que cada tarea (operador) está conectada a otras dependencias y se ejecuta estrictamente en una secuencia especificada. DAG elimina los ciclos, lo que proporciona una estructura lógica y predecible de la ejecución de tareas.

Airflow se encarga de la orquestación: gestiona las dependencias entre tareas, controla los calendarios de ejecución, realiza un seguimiento del estado y reacciona automáticamente ante los fallos. Este enfoque minimiza la intervención manual y garantiza la fiabilidad de todo el proceso.

> **Task Orchestrator** es una herramienta o sistema diseñado para gestionar y controlar la ejecución de tareas en entornos informáticos y de información complejos. Facilita el proceso de despliegue, automatización y gestión de la ejecución de tareas para mejorar el rendimiento y optimizar los recursos.

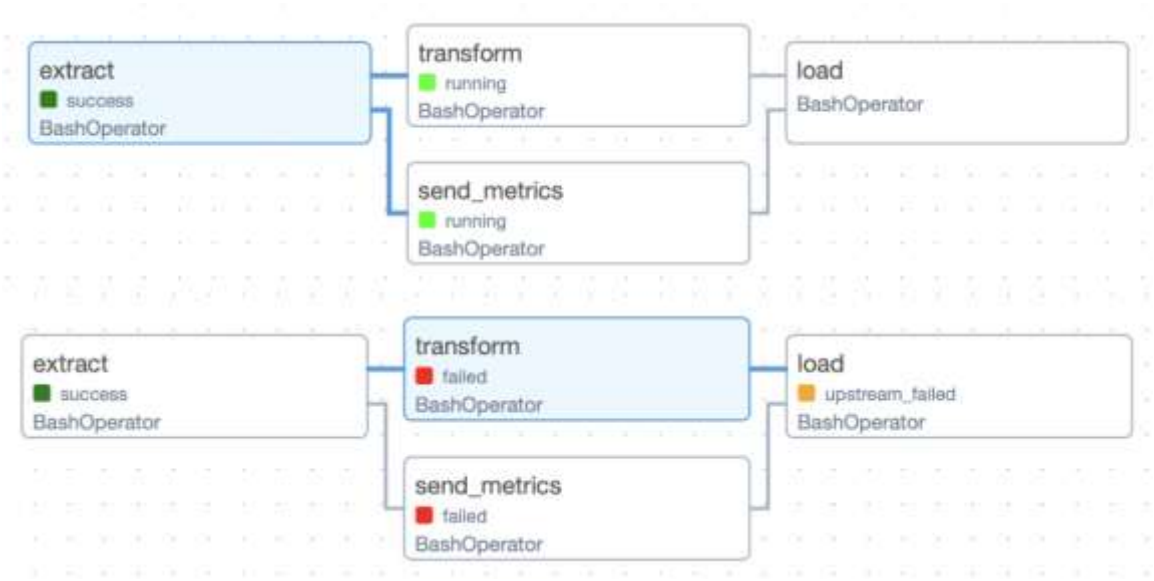


Fig. 7.4-1 Apache Airflow proporciona una interfaz fácil de usar en la que se puede visualizar el DAG -ETL, ver los registros de ejecución, el estado de inicio de las tareas, etc.

Airflow se utiliza ampliamente para la orquestación y automatización de la computación distribuida, el procesamiento de datos, la gestión de procesos ETL (Extract, Transform, Load), la programación de tareas y otros escenarios de datos. Por defecto, Apache Airflow utiliza SQLite como base de datos.

Un ejemplo de DAG simple, similar a ETL, consta de tareas - Extraer, Transformar y Cargar. En el grafo, que se controla a través de la interfaz de usuario (Fig. 7.4-1), se define el orden de ejecución de las tareas (fragmentos de código): por ejemplo, primero se ejecuta Extract, luego Transform (y sending_metrics), y la tarea Load completa el trabajo. Cuando se completan todas las tareas, se considera que el proceso de carga de datos se ha realizado correctamente.

Apache Airflow: aplicación práctica en la automatización ETL

Apache Airflow se utiliza ampliamente para organizar procesos complejos de procesamiento de datos, permitiendo construir ETL flexibles -conveyors. Apache Airflow puede ejecutarse a través de una interfaz web o mediante programación con código Python (Fig. 7.4-2). En la interfaz web (Fig. 7.4-3), los administradores y desarrolladores pueden supervisar visualmente los DAG, ejecutar tareas y analizar los resultados de la ejecución.

Utilizando DAG, puede establecer una secuencia clara de tareas, gestionar las dependencias entre ellas y reaccionar automáticamente a los cambios en los datos de origen. Veamos un ejemplo de uso de Airflow para automatizar el procesamiento de informes (Fig. 7.4-2).

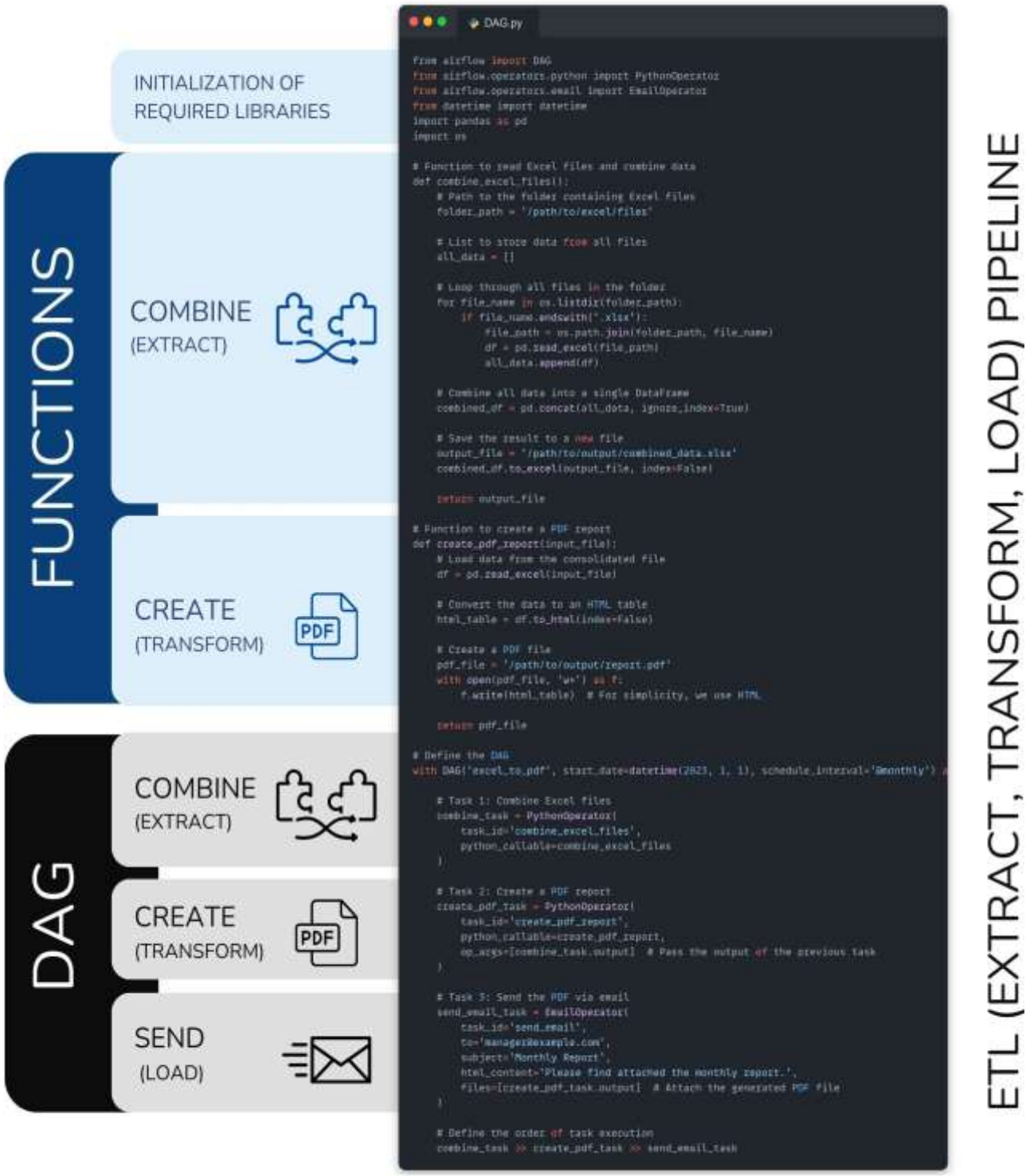


Fig. 7.4-2 Concepto ETL -transportador para el procesamiento de datos mediante Apache Airflow.

Este ejemplo (Fig, 7.4-2) considera el DAG, que realiza tareas clave dentro del ETL -conveyor:

- **Leer Excel -ficheros (Extraer):**
  - Recorrido secuencial de todos los ficheros de un directorio dado.
  - Leer los datos de cada fichero usando la librería pandas.
  - Fusiona todos los datos en un único DataFrame.
- **Crear documento PDF (Transformar):**
  - Transforme el DataFrame combinado en una tabla HTML.
  - Guardar la tabla como PDF (en la versión de demostración - a través de HTML).
- **Envío de un informe por correo electrónico (Load):**
  - Aplicar EmailOperator para enviar PDF -documento por correo electrónico.
- **Configuración del DAG:**
  - Definición de la secuencia de tareas: extracción de datos→ generación de informe→ envío.
  - Asignación de un calendario de lanzamiento (@mensual - primer día de cada mes).

El ejemplo de ETL automatizado (Fig, 7.4-2) muestra cómo recopilar datos de archivos Excel, crear un documento PDF y enviarlo por correo electrónico. Éste es sólo uno de los muchos casos de uso posibles de Airflow. Este ejemplo puede adaptarse a cualquier tarea específica para simplificar y automatizar el procesamiento de datos.

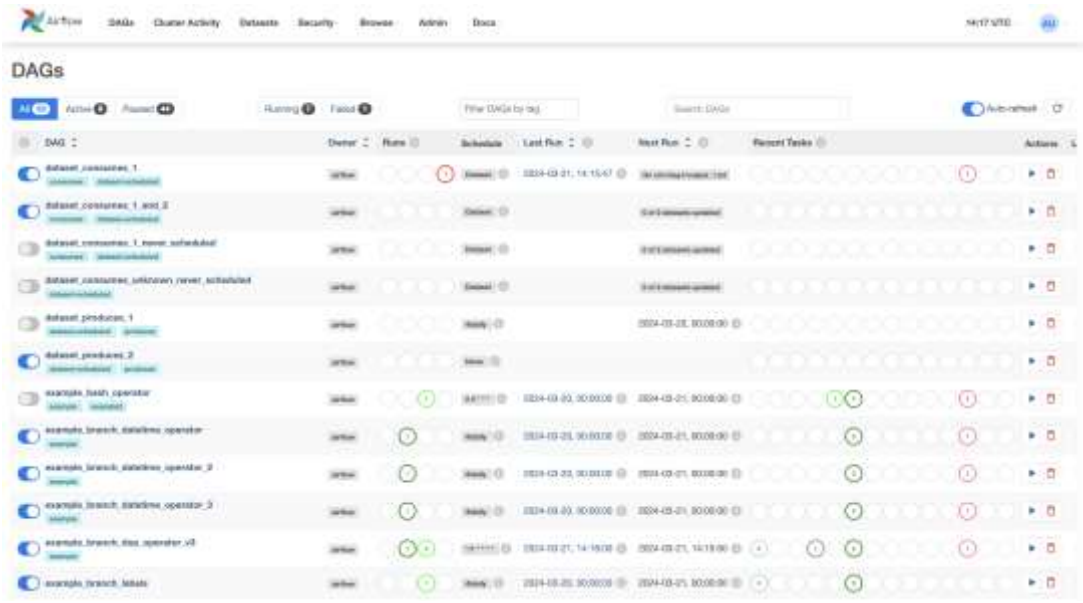


Fig. 7.4-3 Visión general de todos los DAGs en el entorno con información sobre ejecuciones recientes.

La interfaz web de Apache Airflow (Fig, 7.4-3) proporciona un entorno visual completo para gestionar flujos de trabajo de datos. Muestra los DAG como gráficos interactivos en los que los nodos representan las tareas y las aristas las dependencias entre ellas, lo que facilita el seguimiento de flujos de trabajo de datos complejos. La interfaz incluye un panel de control con información sobre el estado de ejecución de las

tareas, el historial de ejecución, registros detallados y métricas de rendimiento. Los administradores pueden iniciar tareas manualmente, reiniciar operaciones fallidas, suspender DAG y configurar variables de entorno, todo ello a través de una interfaz de usuario intuitiva.

Dicha arquitectura puede complementarse con validación de datos, notificaciones sobre el estado de ejecución, integración con API externas o bases de datos. Airflow permite adaptar el DAG de forma flexible: añadir nuevas tareas, cambiar su orden, combinar cadenas, lo que lo convierte en una herramienta eficaz para automatizar procesos complejos de tratamiento de datos. Al ejecutar DAG en la interfaz web de Airflow (Fig. 7.4-3, Fig. 7.4-4), puede supervisar el estado de las tareas. El sistema utiliza la indicación por colores:

- Verde: la tarea se ha completado con éxito.
- Amarillo: el proceso está en curso.
- Rojo - un error al realizar la tarea.

En caso de fallo (por ejemplo, si falta un archivo o se rompe una estructura de datos), el sistema inicia automáticamente el envío de una notificación.

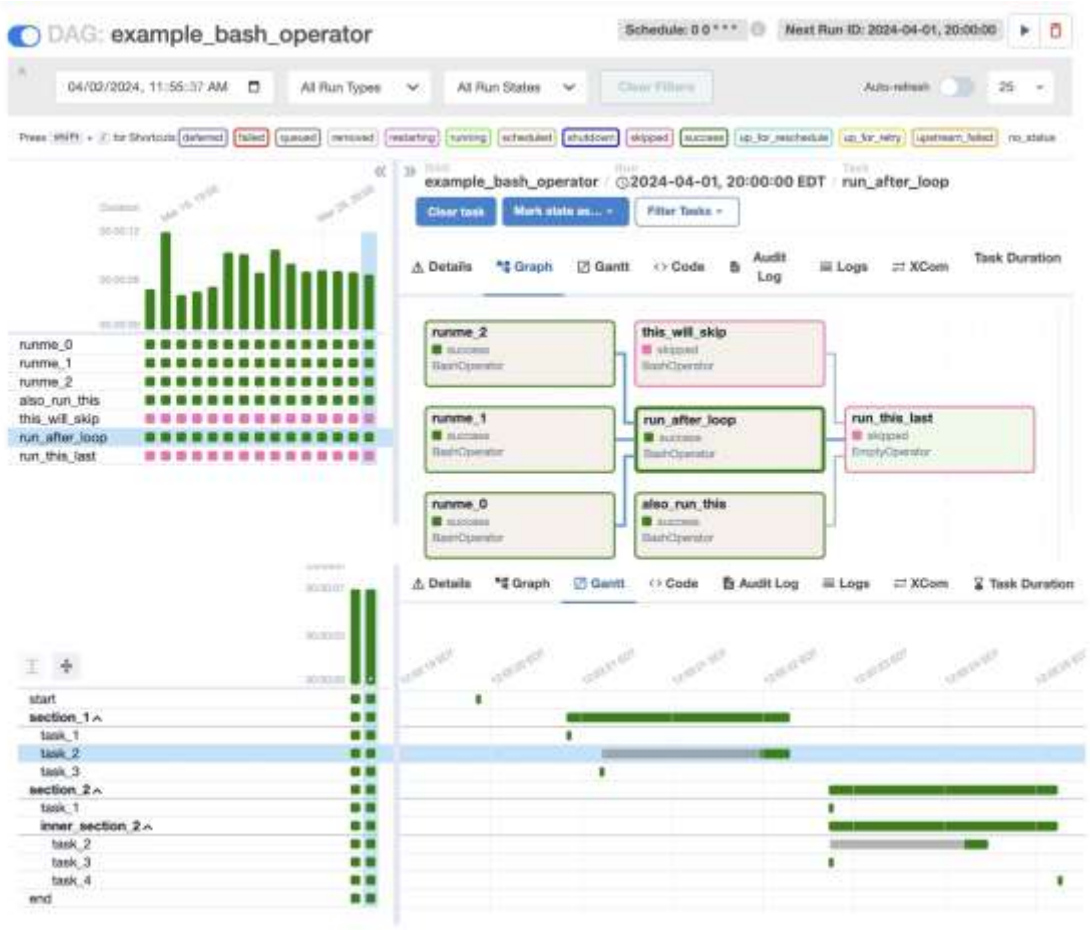


Fig. 7.4-4 Apache Airflow simplifica enormemente el diagnóstico de problemas, la optimización de procesos y la colaboración en equipo en complejas cadenas de procesamiento de datos.

> Apache Airflow es cómodo porque automatiza las tareas rutinarias, eliminando la necesidad de realizarlas manualmente. Garantiza la fiabilidad mediante la supervisión de la ejecución de los procesos y la notificación instantánea de errores. La flexibilidad del sistema facilita la incorporación de nuevas tareas o la modificación de las existentes, adaptando los flujos de trabajo a los requisitos cambiantes.

Además de Apache Airflow, existen herramientas similares para orquestar flujos de trabajo. Por ejemplo, Prefect, de código abierto y gratuito (Fig, 7.3-5), ofrece una sintaxis más sencilla y se integra mejor con Python. Luigi, desarrollado por Spotify, ofrece una funcionalidad similar y funciona bien con big data. También cabe destacar Kronos y Dagster, que ofrecen enfoques modernos para construir Pipeline centrándose en la modularidad y la escalabilidad. La elección de la herramienta de orquestación de tareas depende de las necesidades específicas del proyecto, pero todas ayudan a automatizar los complejos procesos de datos ETL

Cabe destacar Apache NiFi, una plataforma de código abierto, diseñada para el flujo y enrutamiento de datos. A diferencia de Airflow, que se centra en el procesamiento por lotes y la gestión de dependencias, NiFi se centra en la transformación de datos en tiempo real y sobre la marcha y en el enrutamiento flexible entre sistemas.

Apache NiFi para enrutamiento y conversión de datos

Apache NiFi es una potente plataforma de código abierto, diseñada para automatizar los flujos de datos entre distintos sistemas. Originalmente fue desarrollada en 2006 por la Agencia de Seguridad Nacional de Estados Unidos (NSA) con el nombre de "Niagara Files" para uso interno. En 2014, el proyecto se hizo de código abierto y se transfirió a la Apache Software Foundation, pasando a formar parte de sus iniciativas de transferencia de tecnología [141].

Apache NiFi está diseñado para recoger, procesar y transmitir datos en tiempo real. A diferencia de Airflow, que trabaja con tareas por lotes y requiere calendarios bien definidos, NiFi funciona en modo de procesamiento de flujos, lo que permite transferir datos continuamente entre distintos servicios.

> Apache NiFi es ideal para la integración con dispositivos IoT, sensores de construcción, sistemas de monitorización y, por ejemplo, la validación en streaming de formatos CAD en un servidor donde puede ser necesaria una respuesta inmediata a los cambios en los datos.

Con herramientas integradas de filtrado, transformación y enrutamiento, NiFi permite normalizar los datos (Transform) antes de transferirlos (Load) a sistemas de almacenamiento o análisis. Una de sus principales ventajas es el soporte de seguridad y el control de acceso incorporados, que lo convierten en una solución fiable para el tratamiento de información sensible.

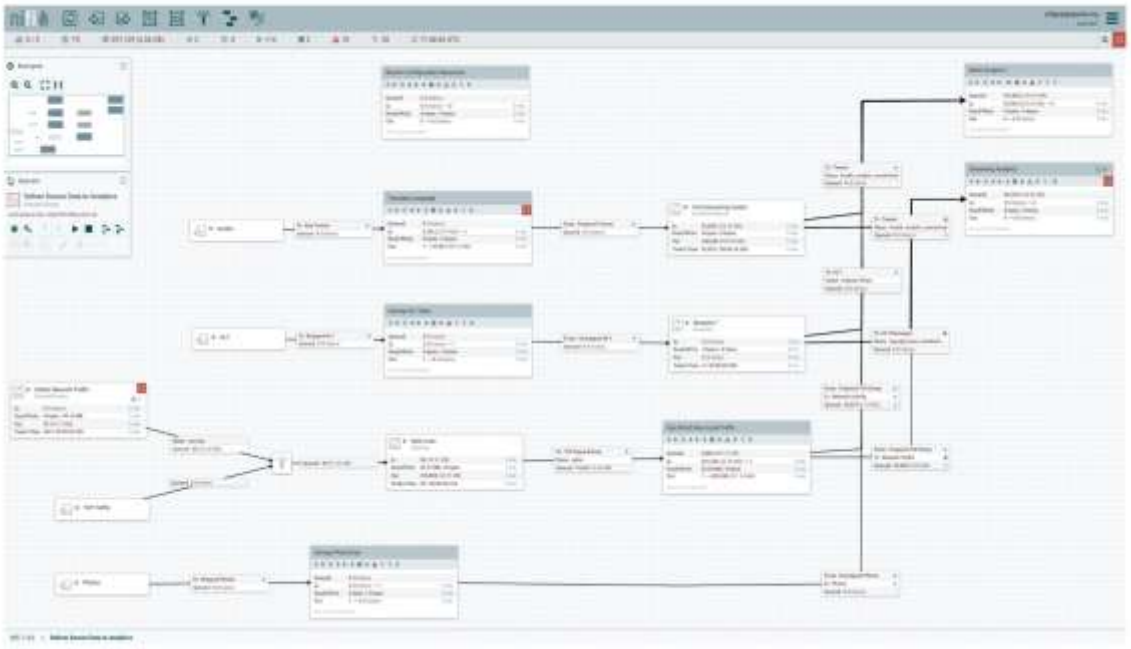


Fig. 7.4-5 Representación gráfica del flujo de datos en la interfaz Apache NiFi.

Apache NiFi gestiona eficazmente las tareas de flujo, filtrado y enrutamiento de datos en tiempo real. Es ideal para escenarios técnicamente intensivos en los que la transferencia estable de información entre sistemas y un alto rendimiento son importantes.

Sin embargo, cuando el objetivo principal es integrar varios servicios, automatizar operaciones rutinarias y establecer rápidamente flujos de trabajo sin profundos conocimientos de programación, se demandan soluciones con un umbral de entrada bajo y la máxima flexibilidad. Una de estas herramientas es n8n - una plataforma de clase Low-Code /No-Code centrada en la automatización empresarial y la orquestación visual de procesos.

n8n Low-Code, No-Code orquestación de procesos

n8n es una plataforma Open Source Low-Code / No-Code para la creación de flujos de trabajo automatizados, caracterizada por su facilidad de uso, flexibilidad y capacidad de integración rápida con una amplia gama de servicios externos.

> **No-Code** es un método para crear productos digitales sin escribir código. Todos los elementos del proceso -desde la lógica hasta la interfaz- se realizan exclusivamente con la ayuda de herramientas visuales. Las plataformas No-Code están dirigidas a usuarios sin conocimientos técnicos y permiten crear rápidamente automatizaciones, formularios, integraciones y aplicaciones web. Ejemplo: un usuario configura el envío automático de notificaciones o la integración con Google Sheets a través de una interfaz de arrastrar y soltar sin conocimientos de programación.

Con capacidades de código abierto y despliegue local, n8n en los procesos de automatización y creación de ETL Pipelines ofrece a las empresas un control total sobre sus datos al tiempo que garantiza la seguridad y la independencia de los proveedores de la nube.

A diferencia de Apache Airflow, que está orientado a tareas computacionales con orquestación rígida y requiere conocimientos de Python, n8n ofrece un editor visual que permite crear scripts sin necesidad de conocer lenguajes de programación (Fig, 7.4-6). Aunque su interfaz permite crear procesos automatizados sin escribir código (No-Code), en escenarios más complejos los usuarios pueden añadir sus propias funciones JavaScript y Python -para ampliar las capacidades (Low-Code).

> **Low-Code** es un enfoque del desarrollo de software en el que la lógica básica de una aplicación o proceso se crea utilizando una interfaz gráfica y elementos visuales, y el código del programa sólo se utiliza para personalizar o ampliar la funcionalidad. Las plataformas Low-Code permiten acelerar considerablemente el desarrollo de soluciones al implicar no sólo a programadores, sino también a usuarios empresariales con conocimientos técnicos básicos. Ejemplo: un usuario puede construir un proceso empresarial a partir de bloques ya creados y, si es necesario, añadir su propio script en JavaScript o Python.

Aunque n8n se posiciona como una plataforma con un umbral de entrada bajo, conocimientos básicos de programación, comprensión de las tecnologías web y habilidades para trabajar con API. La flexibilidad del sistema permite adaptarlo a una amplia gama de tareas, desde el procesamiento automatizado de datos hasta la integración con mensajeros, dispositivos IoT y servicios en la nube.

Principales características y ventajas del uso del n8n:

- Las opciones de **código abierto** e implantación local garantizan el control total de los datos, el cumplimiento de las normas de seguridad y la independencia de los proveedores de la nube.
- **Integración con más de 330 servicios**, incluidos CRM, ERP, comercio electrónico, plataformas en la nube, mensajería y bases de datos.
- **Flexibilidad de escenarios:** desde simples notificaciones hasta cadenas complejas con procesamiento API de -solicitudes, lógica de decisión y conexión de -servicios de IA.
- **Compatibilidad con JavaScript y Python**: los usuarios pueden incrustar código personalizado según sus necesidades, ampliando las capacidades de automatización.
- **Interfaz visual intuitiva:** permite configurar y visualizar rápidamente todos los pasos del proceso.

Plataformas de clase Low-Code proporcionan herramientas para crear soluciones digitales con un código mínimo, lo que las hace ideales para equipos que no tienen profundos conocimientos técnicos pero necesitan automatizar procesos.

En la construcción, n8n puede utilizarse para automatizar diversos procesos, como la integración con sistemas de gestión de proyectos, la comprobación de flujos, la redacción de informes y cartas listos para usar, la actualización automática de datos de inventario de materiales, el envío de notificaciones de estado de tareas a los equipos, etc. Un Pipeline personalizado en n8n puede multiplicar las operaciones manuales, reducir la probabilidad de errores y acelerar la toma de decisiones para la ejecución de proyectos.

> Puede elegir entre casi dos mil n8n Pipeline listos para usar, gratuitos y de código abierto, disponibles en: n8n.io/workflows, para automatizar tanto los flujos de trabajo de la construcción como las tareas personales, reduciendo las operaciones rutinarias.

Por ejemplo, una de las plantillas Pipeline listas para usar, disponible gratuitamente en n8n.io [142], que crea automáticamente borradores de respuestas en Gmail (Fig. 7.4-6), lo que ayuda a los usuarios que reciben un gran volumen de correos o tienen dificultades para redactar respuestas.

Esta plantilla n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Fig, 7.4-6) analiza los correos entrantes utilizando LLM de ChatGPT, determina si es necesaria una respuesta, genera un borrador desde ChatGPT y convierte el texto a HTML y lo añade a la cadena de mensajes en Gmail. Esto no envía automáticamente el correo electrónico, lo que le permite editar y aprobar manualmente la respuesta. La instalación lleva unos 10 minutos e incluye la configuración OAuth de la API de Gmail y la integración de la API OpenAI. El resultado es una solución cómoda y gratuita para automatizar la comunicación rutinaria por correo electrónico sin perder el control sobre el contenido de los mensajes.

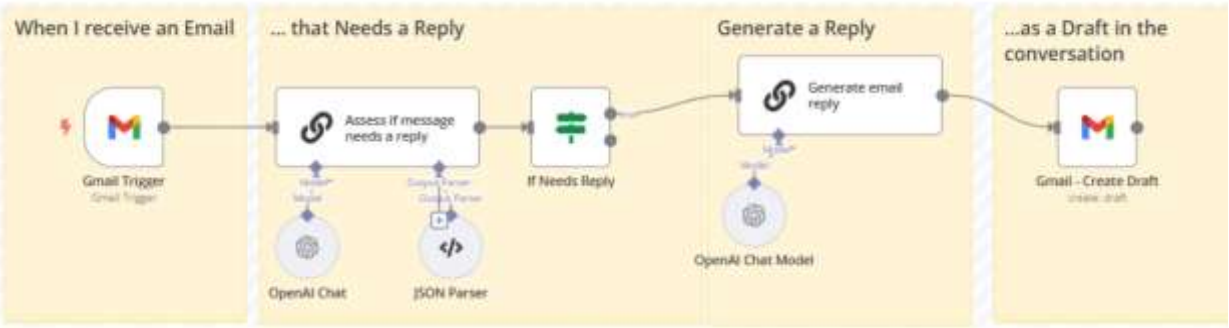


Fig. 7.4-6 Proceso automatizado de generación de respuestas de correo electrónico mediante n8n.

Otro ejemplo de automatización con n8n es la búsqueda de grandes ofertas en el mercado inmobiliario [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", recopila diariamente ofertas relevantes que coinciden con los criterios dados utilizando Zillow API. Calcula automáticamente las métricas clave de la inversión (Cash on Cash ROI, Monthly Cash Flow, Down Payment), actualiza Google Sheets y envía un informe resumido al correo electrónico (Fig, 7.4-7), lo que permite a los inversores ahorrar tiempo y responder rápidamente a las mejores ofertas.

Fig. 7.4-7 Proceso automatizado para evaluar el atractivo inversor de los bienes inmuebles.

Su flexibilidad y extensibilidad hacen del n8n una valiosa herramienta para las empresas que buscan trans-

formarse digitalmente y ser más competitivas en el mercado con herramientas de código abierto relativamente sencillas y gratuitas.

Herramientas como Apache NiFi, Airflow y n8n pueden considerarse tres capas de procesamiento de datos (Fig, 7.4-8). NiFi gestiona el flujo de datos, garantizando su entrega y transformación, Airflow orquesta la ejecución de tareas agregando datos en conductos de procesamiento, y n8n automatiza la integración con servicios externos y gestiona la lógica empresarial.

	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Fig. 7.4-8 Apache Airflow, Apache NiFi y n8n pueden considerarse tres capas complementarias de la arquitectura moderna de gestión de datos.

Juntas, estas herramientas gratuitas y de código abierto constituyen potencialmente un ejemplo de ecosistema eficaz para la gestión de datos y procesos en el sector de la construcción, que permite a las empresas aprovechar la información para la toma de decisiones y la automatización de procesos.

Próximos pasos: pasar de las operaciones manuales a las soluciones basadas en análisis

Las empresas de construcción de hoy en día operan en un entorno de gran incertidumbre: precios cambiantes de los materiales, retrasos en las entregas, escasez de mano de obra y plazos ajustados en los proyectos. El uso de cuadros de mando analíticos, ETL -conveyors y sistemas de BI ayuda a las empresas a identificar rápidamente las áreas problemáticas, evaluar la eficiencia de los recursos y predecir los cambios antes de que provoquen pérdidas financieras.

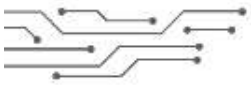
Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar las tecnologías comentadas en sus tareas diarias:

- Implementación de visualizaciones de datos y cuadros de mando analíticos
  - Dominar el proceso de creación de cuadros de mando para supervisar los indicadores clave de rendimiento (KPI)
  - Utilice herramientas de visualización para sus datos (Power BI, Tableau, Matplotlib, Plotly)

- Automatice el tratamiento de datos mediante ETL -processes
  - Establezca la recopilación automática de datos de diversas fuentes (documentación, tablas, CAD) mediante procesos ETL.
  - Organizar la transformación de datos (por ejemplo, comprobación o cálculo de expresiones regulares) mediante scripts de Python.
  - Pruebe a crear informes automáticos en PDF (o DOC) con la biblioteca FPDF, utilizando datos de archivos Excel o extrayendo información de otros documentos PDF.
- Utilización de modelos lingüísticos (LLM) para la automatización
  - Utilizar grandes modelos lingüísticos (LLM) para generar código que ayude a extraer y analizar datos de documentos no estructurados.
  - Familiarícese con la herramienta de automatización de n8n y explore plantillas ya preparadas y casos prácticos en su sitio web. Identifique qué procesos de su trabajo pueden automatizarse por completo utilizando el enfoque No-Code/Low-Code.

Un enfoque analítico de la automatización de datos y procesos no sólo reduce el tiempo dedicado a las operaciones rutinarias, sino que también mejora la calidad de la toma de decisiones. Las empresas que implantan herramientas de análisis visual y ETL -conveyors obtienen la oportunidad de reaccionar rápidamente a los cambios

Automatizar los procesos empresariales con herramientas como n8n, Airflow y NiFi es sólo el primer paso hacia la madurez digital. El siguiente paso es el almacenamiento y la gestión de calidad de los propios datos que sustentan la automatización. En la Parte 8, analizamos en profundidad cómo las empresas de construcción pueden crear una arquitectura de almacenamiento de datos sostenible, pasando de un caos de documentos y archivos multiformato a plataformas centralizadas de almacenamiento y análisis.



VIII PARTE

ALMACENAMIENTO Y GESTIÓN DE DATOS EN LA CONSTRUCCIÓN

La Parte 8 explora las modernas tecnologías de almacenamiento y gestión de datos para el sector de la construcción. Analiza formatos eficaces para manejar grandes volúmenes de información, desde los sencillos CSV y XLSX hasta los de mayor rendimiento Apache Parquet y ORC, con una comparación detallada de sus capacidades y limitaciones. Se analizan los conceptos de almacenes de datos (DWH), lagos de datos) y sus soluciones híbridas (Data Lakehouse), así como los principios de Gobernanza de Datos) y Minimalismo de Datos). Se abordan en detalle los problemas del pantano de datos (Data Swamp) y las estrategias para evitar el caos en los sistemas de información. Se presentan nuevos enfoques para trabajar con datos, entre ellos las bases de datos vectoriales y su aplicación en la construcción mediante el concepto de Bounding Box. En esta parte también se abordan las metodologías DataOps y VectorOps como nuevos estándares para organizar los flujos de trabajo con datos.

CAPÍTULO 8.1.
INFRAESTRUCTURA DE DATOS: DE LOS FORMATOS DE ALMACENAMIENTO A LOS REPOSITORIOS DIGITALES

Átomos de datos: la base de una gestión eficaz de la información

Todo en el Universo está formado por los más pequeños bloques de construcción: átomos y moléculas, y con el tiempo todas las cosas vivas y no vivas vuelven inevitablemente a este estado inicial. En la naturaleza, este proceso se produce con una rapidez asombrosa, que intentamos trasladar a los procesos controlados por el hombre.

> En el bosque, cualquier organismo vivo acaba transformándose en una sustancia nutritiva que sirve de base a nuevas plantas. Estas plantas, a su vez, se convierten en alimento de nuevos seres vivos formados por los mismos átomos que crearon el Universo hace millones de años.

En el mundo empresarial, también es importante descomponer las estructuras complejas y de múltiples capas en sus unidades más fundamentales y mínimamente procesadas, como los átomos y las moléculas en la naturaleza. Esto permite almacenar y gestionar eficazmente átomos de datos, convirtiéndolos en una base rica y fértil que se convierte en un recurso clave para el crecimiento de la analítica y la calidad de las decisiones.

Fig. 8.1-1 El análisis y la toma de decisiones se basan en datos reutilizados que en su día fueron procesados y almacenados.

Las composiciones musicales están formadas por notas que se combinan para crear piezas musicales complejas, mientras que las palabras se crean a partir de una unidad primitiva, la letra-sonido. Tanto en la naturaleza como en la ciencia, la economía, el arte o la tecnología, el mundo muestra una unidad y armonía notables en su búsqueda de la destrucción, la estructura, el ciclo y la creación. Del mismo modo, los procesos de los sistemas de cálculo de costes se desglosan en minúsculas unidades estructuradas (elementos de recursos) en el nivel de cálculo de costes y calendarios. Estas unidades, como notas, se utilizan después para formar cálculos y calendarios más complejos. El mismo principio utilizan los sistemas de diseño asistido por ordenador, en los que los proyectos complejos de arquitectura e ingeniería se construyen a partir de elementos básicos -elementos individuales y componentes de biblioteca-, a partir de los cuales se crea un -modelo 3D completo del proyecto de un edificio o estructura complejos.

> El concepto de ciclicidad y estructura inherente a la naturaleza y la ciencia también se refleja en el mundo moderno de los datos. Al igual que en la naturaleza todos los seres vivos vuelven a los átomos y las moléculas, en el mundo de las herramientas modernas de tratamiento de datos la información tiende a volver a su forma más primitiva.

Los elementos más pequeños, con su indivisibilidad finita, son los componentes básicos de los procesos empresariales. Es importante estudiar detenidamente desde el principio cómo recopilar, estructurar (descomponer en átomos) y almacenar estos minúsculos bloques de construcción procedentes de diversas fuentes. Organizar y almacenar los datos no es sólo cuestión de descomponerlos en sus partes constituyentes. Es igualmente importante garantizar que se integran y almacenan de forma estructurada para que los datos puedan recuperarse, analizarse y utilizarse fácilmente para la toma de decisiones siempre que sea necesario.

Para procesar la información con eficacia, hay que elegir cuidadosamente el formato y los métodos de almacenamiento de datos, igual que hay que preparar el suelo para el crecimiento de los árboles. Los almacenes de datos deben organizarse para garantizar una alta calidad y pertinencia de la información, eliminando los datos redundantes o irrelevantes. Cuanto mejor estructurado esté este "suelo de información", más rápido y con mayor precisión podrán los usuarios encontrar los datos adecuados y resolver problemas analíticos.

Almacenamiento de información: ficheros o datos

Los almacenes de datos permiten a las empresas recopilar y combinar información de distintos sistemas, creando un centro único para su posterior análisis. Los datos históricos recopilados no sólo permiten un análisis más profundo de los procesos, sino también la identificación de patrones que pueden afectar al rendimiento empresarial.

Supongamos que una empresa está trabajando en varios proyectos al mismo tiempo. Un ingeniero quiere saber cuánto hormigón se ha vertido y cuánto queda por comprar. Con un enfoque tradicional, tendría que buscar manualmente en el servidor y abrir varias tablas de estimaciones, compararlas con los certificados de obras terminadas y comprobar los saldos de existencias actuales. Esto lleva horas o incluso días. Incluso con procesos ETL y scripts automáticos, la tarea sigue siendo semimanual: el ingeniero tiene que especificar manualmente la ruta a carpetas o archivos concretos del servidor. Esto reduce el efecto global

de la automatización, ya que sigue ocupando un valioso tiempo de trabajo.

Al pasar a la gestión de datos, en lugar de trabajar con el sistema de archivos del servidor, el ingeniero obtiene acceso a una estructura de almacenamiento unificada donde la información se actualiza en tiempo real. Una sola consulta -en forma de código, SQL -query, o incluso una llamada al agente LLM -puede proporcionar al instante datos precisos sobre los saldos actuales, el trabajo realizado y las próximas entregas, si los datos se han preparado con antelación y se han combinado en un almacén de datos donde no es necesario deambular por carpetas, abrir decenas de archivos y comparar valores manualmente.

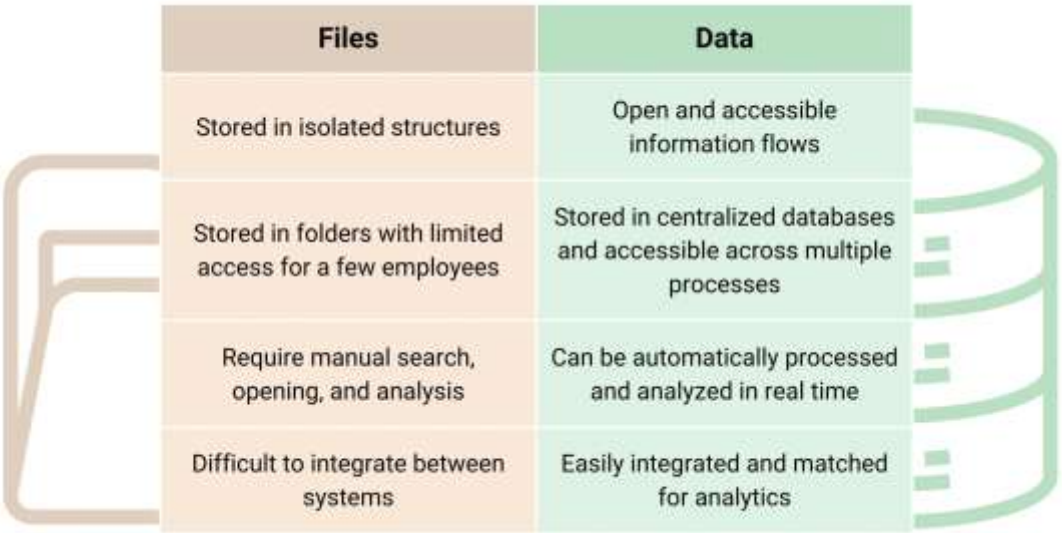
Durante mucho tiempo, las empresas de construcción han utilizado documentos PDF -documents, planos DWG -drawings, modelos RVT -models y cientos y miles de tablas Excel -tables and other disparate formats, que se almacenan en carpetas específicas de los servidores de la empresa, lo que dificulta la búsqueda de información, su comprobación y análisis. Como consecuencia, los archivos que quedan tras la finalización de los proyectos suelen trasladarse de nuevo al servidor a carpetas de almacenamiento de archivo, que prácticamente no se utilizan en el futuro. Este almacenamiento de datos tradicional basado en archivos pierde relevancia a medida que aumenta el flujo de datos, debido a su vulnerabilidad a los errores humanos.

> **Un archivo** no es más que un contenedor aislado en el que se almacenan datos. Los archivos se crean para las personas, no para los sistemas, por lo que requieren una apertura, lectura e interpretación manuales. Ejemplos de ello son una tabla de Excel, un documento PDF o un dibujo CAD que hay que abrir con una herramienta concreta para acceder a la información deseada. Sin una recuperación y un tratamiento estructurados, la información que contienen queda inutilizada.
>
> **Los datos**, a su vez, son información legible por máquina que se vincula, actualiza y analiza automáticamente. En un almacén de datos único (por ejemplo, base de datos, DWH o Data Lake), la información se representa en forma de tablas, registros y relaciones. Esto permite un almacenamiento uniforme, consultas automatizadas, análisis de valores e informes en tiempo real.

El uso de datos en lugar de ficheros (Fig. 8.1-1) elimina el proceso de búsqueda manual y unifica el procesamiento. Las empresas que ya han implantado este enfoque obtienen una ventaja competitiva gracias a la rapidez de acceso a la información y a la capacidad de integrarla rápidamente en los procesos empresariales.

> El paso de los archivos a los datos es un cambio inevitable que marcará el futuro del sector de la construcción.

Todas las empresas del sector de la construcción se enfrentarán a una elección clave: seguir almacenando información en archivos y silos dispares que deben ser leídos por humanos mediante programas especiales o transformarla en las primeras fases de procesamiento en datos estructurados, creando una única base digital integrada para la gestión automatizada de proyectos.



Files	Data
Stored in isolated structures	Open and accessible information flows
Stored in folders with limited access for a few employees	Stored in centralized databases and accessible across multiple processes
Require manual search, opening, and analysis	Can be automatically processed and analyzed in real time
Difficult to integrate between systems	Easily integrated and matched for analytics

Fig. 8.1-1 Evolución del flujo de información: de los ficheros aislados a los datos integrados.

Con la explosión de la información, los métodos tradicionales de almacenamiento y tratamiento de archivos son cada vez menos eficaces. En la industria de la construcción, como en otros sectores, ya no basta con confiar en carpetas de archivos dispares con distintos formatos de archivo o bases de datos inconexas.

Las empresas que quieran seguir siendo competitivas en la era digital cambiarán inevitablemente a plataformas digitales integradas, utilizarán tecnologías de macrodatos y sistemas de análisis automatizados.

Pasar del almacenamiento basado en archivos a los flujos de trabajo basados en datos exigirá un replanteamiento de los enfoques de gestión de la información y una elección consciente de los formatos adecuados para su posterior integración en repositorios centralizados. Esta elección determinará la eficiencia con la que se pueden procesar los datos, la rapidez con la que se puede acceder a ellos y la facilidad con la que se pueden integrar en los procesos digitales de una empresa.

Almacenamiento de Big Data: análisis de los formatos más populares y su eficacia

Los formatos de almacenamiento desempeñan un papel clave en la escalabilidad, fiabilidad y rendimiento de la infraestructura analítica. Para analizar y procesar datos -como filtrar, agrupar y agregar- nuestros ejemplos utilizaron Pandas DataFrame -una estructura popular para trabajar con datos en RAM-.

Sin embargo, Pandas DataFrame no tiene su propio formato de almacenamiento, por lo que una vez finalizado el procesamiento, los datos se exportan a uno de los formatos externos - la mayoría de las veces CSV o XLSX. Estos formatos tabulares son fáciles de intercambiar y compatibles con la mayoría de los sistemas externos, pero tienen una serie de limitaciones: baja eficiencia de almacenamiento, falta de compresión y escaso soporte de versiones:

- **CSV** (Comma-Separated Values): un formato de texto sencillo ampliamente soportado por diversas

plataformas y herramientas. Es fácil de usar, pero no admite tipos de datos complejos ni compresión.

- **XLSX** (Excel Open XML Spreadsheet): formato de archivo de Microsoft Excel que admite funciones sofisticadas como fórmulas, gráficos y estilos. Aunque es útil para el análisis y la visualización manual de datos, no está optimizado para el tratamiento de datos a gran escala.

Además de los populares formatos tabulares XLSX y CSV, existen varios formatos populares para almacenar datos estructurados de forma eficiente (Fig, 8.1-2), cada uno con ventajas únicas en función de los requisitos específicos de almacenamiento y análisis de datos:

- **Apache Parquet**: un formato de archivo de almacenamiento de datos en columnas optimizado para su uso en sistemas de análisis de datos. Ofrece esquemas eficientes de compresión y codificación de datos, por lo que es ideal para estructuras de datos complejas y procesamiento de big data.
- **Apache ORC** (Optimised Row Columnar): Similar a Parquet, ORC proporciona una alta compresión y un almacenamiento de datos eficiente. Está optimizado para operaciones de lectura pesadas y es muy adecuado para almacenar lagos de datos.
- **JSON** (JavaScript Object Notation): aunque JSON no es tan eficiente en términos de almacenamiento de datos en comparación con formatos binarios como Parquet u ORC, es muy accesible y fácil de trabajar, por lo que es ideal para scripts en los que la legibilidad y la compatibilidad web son importantes.
- **Feather**: un formato de almacenamiento de datos binarios en columnas orientado a la analítica, rápido, ligero y fácil de usar. Está diseñado para transferir datos de forma eficiente entre Python (Pandas) y R, lo que lo convierte en una excelente opción para proyectos que impliquen estos entornos de programación.
- **HDF5** (Hierarchical Data Format versión 5): diseñado para almacenar y organizar grandes cantidades de datos. Admite una amplia gama de tipos de datos y es muy adecuado para trabajar con colecciones complejas de datos. HDF5 es especialmente popular en la informática científica por su capacidad para almacenar y acceder eficazmente a grandes conjuntos de datos.

	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame
Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
Compression	Built-in	None	High	Built-in	None (in-memory)
Performance	Low	Medium	High	High	High (memory dependent)
Complexity	High (formatting, styles)	Low	Medium	Medium	Low
Data Type Support	Limited	Very limited	Extended	Extended	Extended
Scalability	Low	Low	High	High	Medium (memory limited)

Fig. 8.1-2 Comparación de formatos de datos que muestra las principales diferencias en aspectos de almacenamiento y procesamiento.

Para realizar un análisis comparativo de los formatos utilizados en la fase de carga del proceso ETL, se creó una tabla que mostraba el tamaño de los ficheros y los tiempos de lectura (Fig. 8.1-3). En el estudio se utilizaron ficheros con datos idénticos: la tabla contenía 10.000 filas y 10 columnas rellenas con valores aleatorios.

En el estudio se incluyen los siguientes formatos de almacenamiento: CSV, Parquet, XLSX y HDF5, así como sus versiones comprimidas en archivos ZIP. Los datos en bruto se generaron utilizando la biblioteca NumPy y se representaron como una estructura Pandas DataFrame. El proceso de prueba consistió en los siguientes pasos:

- Almacenamiento de archivos: el marco de datos se guarda en cuatro formatos diferentes: CSV, Parquet, XLSX, y HDF5. Cada formato tiene características únicas en la forma de almacenar los datos, lo que afecta al tamaño del archivo y a la velocidad de lectura.
- Compresión de archivos ZIP: para analizar la eficacia de la compresión estándar, cada archivo se comprimió aún más en un archivo ZIP.
- Lectura de archivos (ETL - Load): se midió el tiempo de lectura de cada archivo tras desempaquetarlo del ZIP. Esto permite estimar la velocidad de acceso a los datos tras extraerlos del archivo.

Es importante señalar que Pandas DataFrame no se utilizó directamente en el análisis del tamaño o del tiempo de lectura, ya que no representa un formato de almacenamiento por derecho propio. Sólo sirvió

como estructura intermedia para la generación y posterior almacenamiento de datos en diferentes formatos.

	Original Size (MB)	ZIP Size (MB)	Read Time (s)
CSV	18.82	8.73	0.48
Parquet	9.67	9.22	0.17
XLSX	12.59	12.59	41.73
HDF5	8.50	7.62	0.13

Fig. 8.1-3 Comparación de formatos de almacenamiento por tamaño y velocidad de lectura.

Los archivos CSV y HDF5 demuestran (Fig, 8.1-3) una alta eficiencia de compresión, reduciendo significativamente su tamaño cuando se empaquetan en ZIP, lo que puede ser especialmente útil en escenarios que requieran una optimización del almacenamiento. Los archivos XLSX, por otro lado, son prácticamente incompresibles y su tamaño en ZIP sigue siendo comparable al original, lo que los hace menos favorables para su uso en grandes volúmenes de datos o en entornos en los que la velocidad de acceso a los datos es importante. Además, el tiempo de lectura de XLSX es significativamente mayor en comparación con otros formatos, lo que lo hace menos favorable para operaciones rápidas de lectura de datos. Apache Parquet ha demostrado un alto rendimiento para tareas analíticas y grandes volúmenes de datos gracias a su estructura en columnas.

Optimizar el almacenamiento con Apache Parquet

Uno de los formatos más populares para almacenar y procesar big data es Apache Parquet. Este formato está diseñado específicamente para el almacenamiento en columnas (similar a Pandas), lo que permite reducir significativamente la huella de memoria y aumentar la velocidad de las consultas analíticas. A diferencia de formatos tradicionales como CSV y XLSX, Parquet soporta compresión nativa y está optimizado para sistemas de big data como Spark, Hadoop y almacenamiento en la nube.

Entre las principales características de Parquet figura la compatibilidad con la compresión y codificación de datos, que reduce significativamente el tamaño de almacenamiento y acelera las operaciones de lectura de datos al trabajar directamente sobre las columnas deseadas en lugar de sobre todas las filas de datos.

Para un ejemplo ilustrativo de lo fácil que es obtener el código necesario para convertir datos a Apache Parquet, utilicemos el LLM.

- Envíe una solicitud de texto al chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Escribir código para guardar datos de Pandas DataFrame a Apache Parquet. ⏎

- Respuesta de LLM:

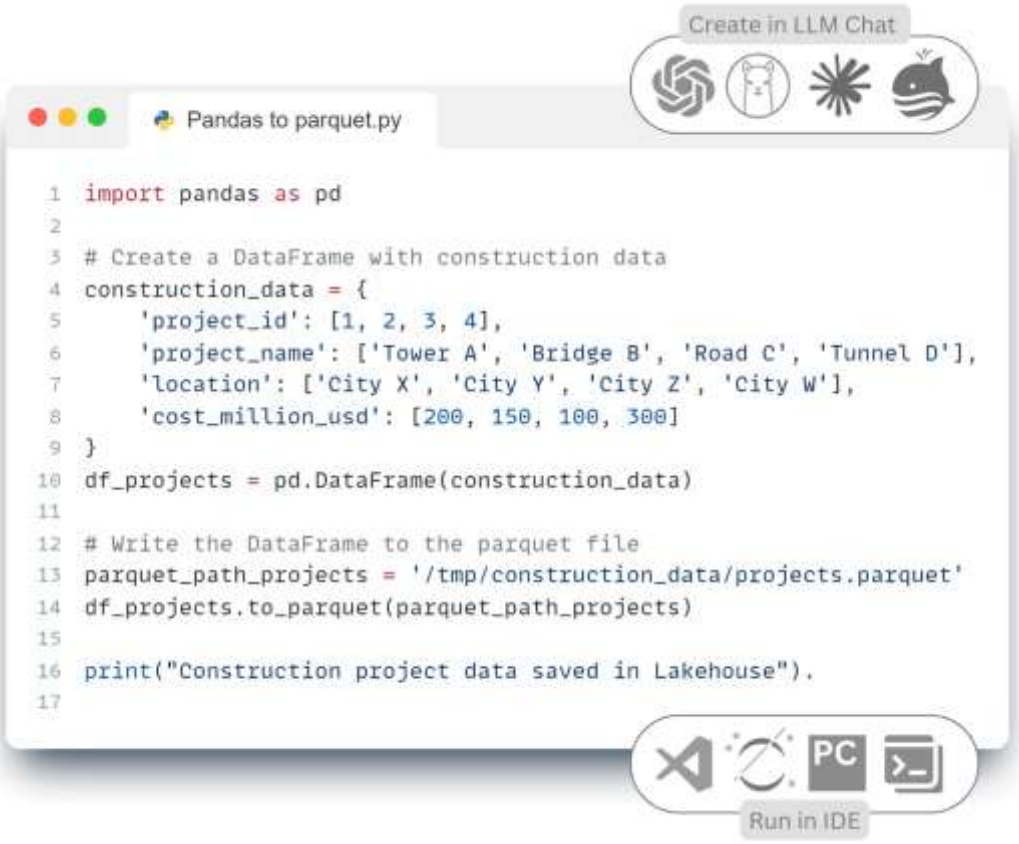


Fig. 8.1-4 Transferencia de datos Dataframe de RAM a un formato Apache Parquet de almacenamiento eficiente utilizando unas pocas líneas de Python.

Siguiente ejemplo: vamos a simular el proceso ETL con datos almacenados en formato Parquet para filtrar

proyectos por un determinado valor de uno de los atributos "cost_million_usd" (Fig. 8.1-4).

- Mientras el chat continúa, envía una solicitud de texto a LLM:

> Escribir código en el que queramos filtrar los datos de una tabla y guardar sólo aquellos proyectos (filas de la tabla) de los datos de Apache Parquet, cuyo coste (parámetro cost_million_usd) supere los 150 millones de dólares. ⏎

- Respuesta de LLM:

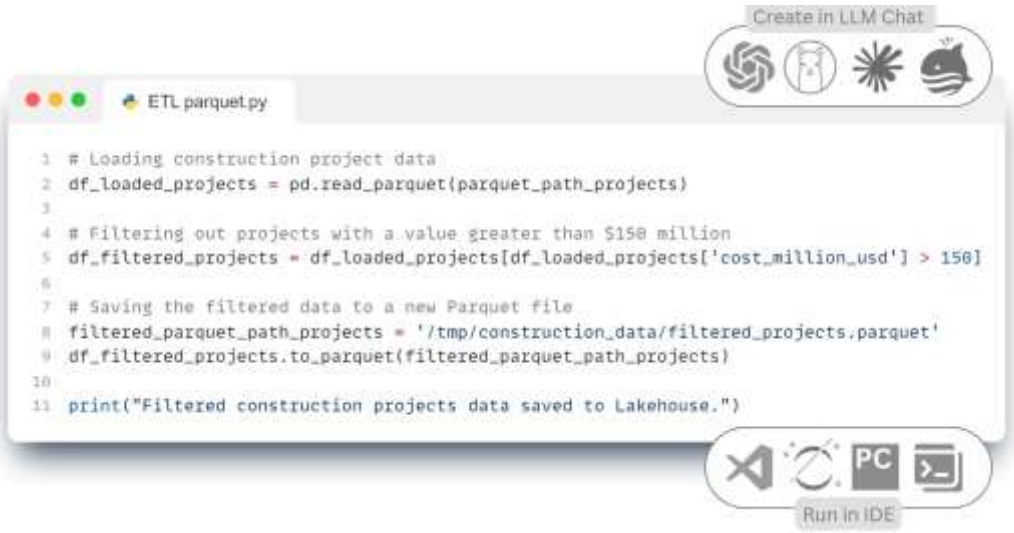


Fig. 8.1-5 El proceso ETL cuando se trabaja con datos en formato Apache Parquet tiene el mismo aspecto que con otros formatos estructurados.

Utilizando el formato Parquet (en relación con XLSX, CSV, etc.) reduce significativamente la cantidad de información almacenada y acelera las operaciones de búsqueda. Esto lo hace excelente tanto para almacenar como para analizar datos. Parquet se integra con diversos sistemas de procesamiento, proporcionando un acceso eficaz en arquitecturas híbridas.

Sin embargo, un formato de almacenamiento eficiente es sólo un elemento de una experiencia de datos completa. Para crear un entorno sostenible y escalable, se requiere una arquitectura de gestión de datos bien diseñada. Los sistemas de clase DWH (Data Warehouse) cumplen esta función. Proporcionan agregación de datos procedentes de fuentes heterogéneas, transparencia de los procesos empresariales y la posibilidad de realizar análisis complejos mediante herramientas de BI y algoritmos de aprendizaje automático.

DWH: Almacén de datos almacenes de datos

Al igual que el formato Parquet está optimizado para el almacenamiento eficiente de grandes cantidades de información, el Almacén de Datos está optimizado para integrar y estructurar datos con el fin de apoyar

la analítica, la previsión y la toma de decisiones de gestión.

En las empresas actuales, los datos proceden de muchas fuentes dispares: ERP, CAFM, CPM, sistemas CRM, contabilidad y gestión de almacenes, modelos digitales CAD de edificios, sensores IoT y otras soluciones. Para obtener una imagen holística, no basta con recopilar datos: es necesario organizarlos, estandarizarlos y centralizarlos en un único repositorio. Esto es exactamente lo que hace DWH: un sistema de almacenamiento centralizado que permite agregar información de diversas fuentes, estructurarla y ponerla a disposición de la analítica y la gestión estratégica.

> **DWH (Data Warehouse**) es un sistema de almacén de datos centralizado que agrega información procedente de múltiples fuentes, la estructura y la pone a disposición para su análisis y la elaboración de informes.

En muchas empresas, los datos están dispersos en distintos sistemas, de los que ya hablamos en las primeras partes del libro (Fig, 1.2-4). El DWH integra estas fuentes, garantizando una total transparencia y fiabilidad de la información. Un almacén de datos DWH es una base de datos especializada (una gran base de datos) que recopila, procesa y almacena datos de múltiples fuentes. Las principales características de un DWH son:

- **Utilizando ETL -procesos** (Extraer, Transformar, Cargar) - extraer datos de las fuentes, limpiarlos, transformarlos, cargarlos en el repositorio y automatizar estos procesos, que se trataron en la séptima parte del libro.
- **Granularidad de los datos**: los datos de DWH pueden almacenarse tanto de forma agregada (informes resumidos) como granular (datos sin procesar). A partir de 2024, son los proveedores de CAD- los que han empezado a hablar de datos granulares [125], lo que tal vez indica que el sector se está preparando para la transición a un almacenamiento especializado en la nube para manejar datos de modelos digitales de construcción.
- **Soporte de análisis y predictivo** - los almacenes de datos proporcionan la base para las herramientas de BI, Big Data -análisis y aprendizaje automático.

DWH sirve de base para la inteligencia empresarial, permitiendo el análisis de indicadores clave de rendimiento, la previsión de ventas, compras y costes, y la elaboración automatizada de informes y la visualización de datos (Fig, 8.1-6).

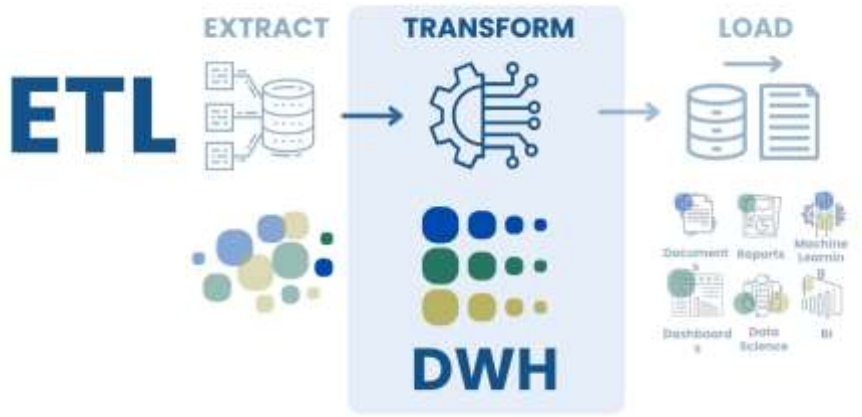


Fig. 8.1-6 En un proceso ETL, el DWH puede actuar como repositorio central donde los datos extraídos de varios sistemas se someten a etapas de transformación y descarga.

DWH desempeña un papel clave en la integración, limpieza y estructuración de la información, proporcionando una base sólida para la inteligencia empresarial y los procesos de toma de decisiones. Sin embargo, en el entorno actual, en el que los volúmenes de datos crecen rápidamente y las fuentes de datos son cada vez más diversas, el enfoque tradicional de DWH para el almacenamiento de información a menudo requiere una ampliación en forma de ELT y Data Lake

Data Lake - evolución de ETL a ELT: de la limpieza tradicional al tratamiento flexible

Los clásicos DWH - data warehouses, diseñados para almacenar datos estructurados en un formato optimizado para consultas analíticas, se han enfrentado a limitaciones en el manejo de datos no estructurados y la escalabilidad. En respuesta a estos retos, han surgido los Data Lakes), que ofrecen almacenamiento flexible para grandes cantidades de datos heterogéneos.

> **Data Lake** ofrece un enfoque DWH alternativo -que permite trabajar con datos no estructurados, semiestructurados y en bruto sin un esquema rígido previo. Este método de almacenamiento suele ser relevante para el procesamiento de datos en tiempo real, el aprendizaje automático y la analítica avanzada. A diferencia del DWH, que estructura y agrega los datos antes de cargarlos, el Data Lake permite almacenar la información en bruto, lo que aporta flexibilidad y escalabilidad

Fue la frustración con los almacenes de datos tradicionales (RDBMS, DWH) y el interés por los "grandes datos" lo que llevó a la aparición de los lagos de datos, en los que, en lugar de complejos ETL, ahora los datos simplemente se cargan en un repositorio poco estructurado, cuyo procesamiento tiene lugar en la fase de análisis:

- En los almacenes de datos tradicionales, los datos se suelen preprocesar, transformar y limpiar

(ETL - Extraer, Transformar, Cargar) antes de cargarlos en el almacén (Fig, 8.1-6). Esto significa que los datos se estructuran y optimizan para futuras tareas específicas de análisis y elaboración de informes. Se hace hincapié en mantener un alto rendimiento de las consultas y la integridad de los datos. Sin embargo, este enfoque puede ser costoso y menos flexible a la hora de integrar nuevos tipos de datos y esquemas de datos que cambian rápidamente.

- Los lagos de datos, por su parte, están diseñados para almacenar grandes cantidades de datos en bruto en su formato original (Fig, 8.1-7). El proceso ETL (Extract, Transform, Load), está siendo sustituido por ELT (Extract, Load, Transform), en el que los datos se cargan primero en el almacén "tal cual" y sólo entonces pueden transformarse y analizarse según las necesidades. Esto proporciona una mayor flexibilidad y la capacidad de almacenar datos heterogéneos, incluidos datos no estructurados como texto, imágenes y registros.

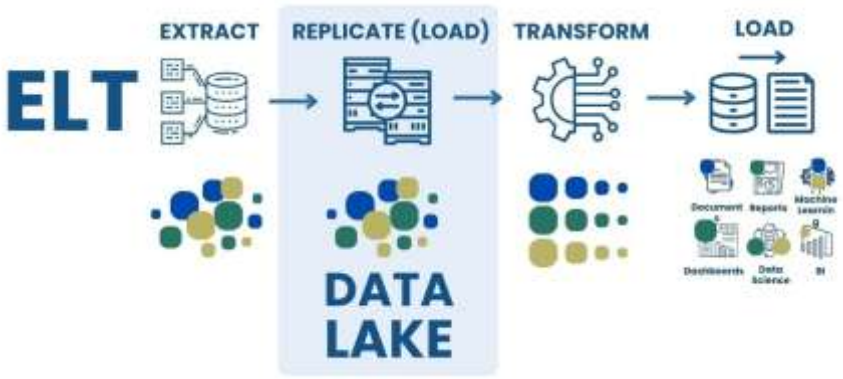


Fig. 8.1-7 A diferencia de ETL, Data Lake utiliza ELT, en el que la información se carga primero en "bruto" y la transformación se realiza en la fase de carga.

Los almacenes de datos tradicionales se centran en preprocesar los datos para garantizar un alto rendimiento de las consultas, mientras que los lagos de datos priorizan la flexibilidad: almacenan los datos en bruto y los transforman según sea necesario (Fig, 8.1-8).

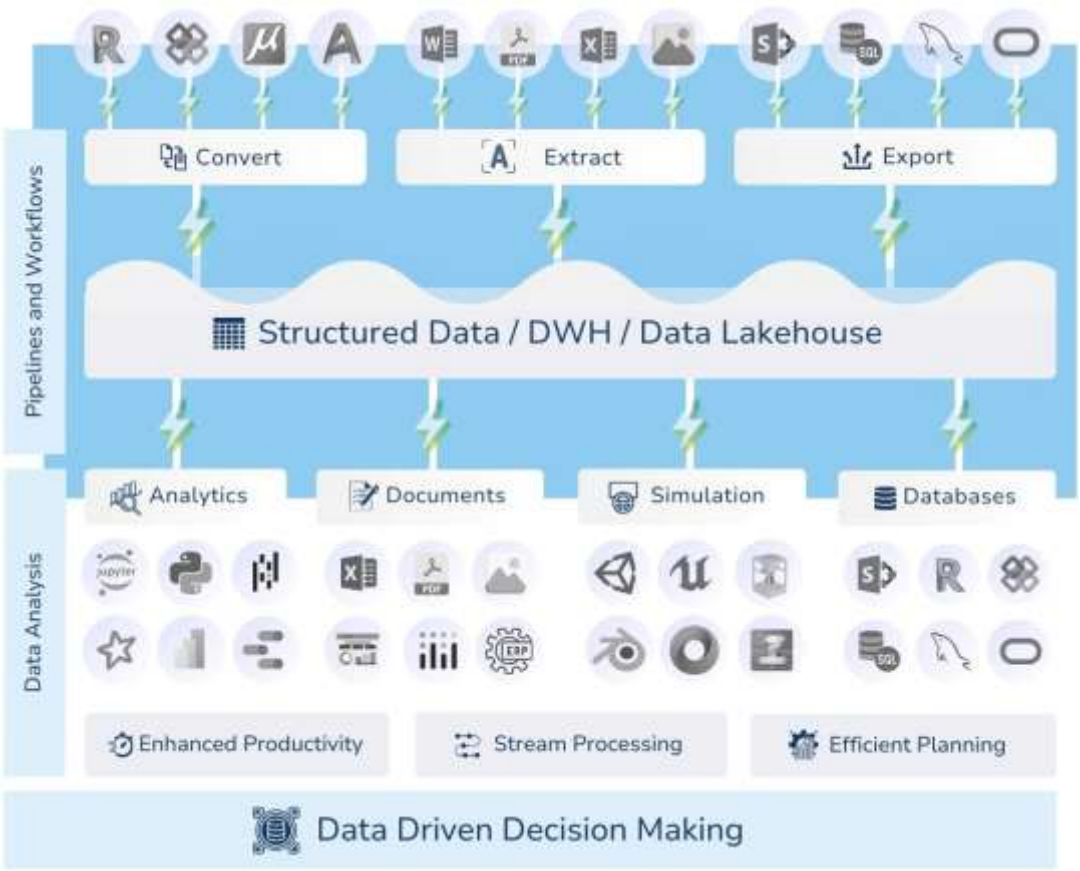


Fig. 8.1-8 Los conceptos modernos de almacenamiento tienen como objetivo almacenar y procesar todo tipo de datos para la toma de decisiones.

Sin embargo, a pesar de todas sus ventajas, los lagos de datos no están exentos de inconvenientes. La falta de una estructura estricta y la complejidad de la gestión de la información pueden llevar al caos, en el que los datos se duplican, se contradicen o pierden relevancia. Además, buscar y analizar datos en un repositorio de este tipo requiere un esfuerzo considerable, sobre todo cuando se trata de información heterogénea. Para superar estas limitaciones y combinar las mejores características de los almacenes de datos tradicionales y los lagos de datos, se desarrolló la arquitectura Data Lakehouse.

Arquitectura Data Lakehouse: sinergia de almacenes y lagos de datos

Para combinar las mejores características de DWH (estructurado, gestionable, analítica de alto rendimiento) y Data Lake (escalabilidad, manejo de datos heterogéneos), se desarrolló el enfoque Data Lakehouse. Esta arquitectura combina la flexibilidad de los lagos de datos con las potentes herramientas de procesamiento y gestión típicas de los almacenes tradicionales, logrando un equilibrio entre almacenamiento, analítica y aprendizaje automático. Data Lakehouse es una síntesis de los lagos de datos y los almacenes de datos, que combina la flexibilidad y escalabilidad de los primeros con la capacidad de gestión y optimización de consultas de los segundos.

Data Lakehouse es un enfoque arquitectónico que pretende combinar la flexibilidad y escalabilidad de los lagos de datos con la manejabilidad y el rendimiento de consulta de los almacenes de datos (Fig, 8.1-9).

Entre las principales características de Data Lakehouse se incluyen:

- **Formato abierto de almacenamiento de datos:** el uso de formatos abiertos para el almacenamiento de datos, como Apache Parquet, proporciona eficacia y optimización de las consultas.
- **Esquema** de sólo lectura**:** en contraste con el enfoque tradicional de un esquema de sólo escritura en DWH, Lakehouse admite un esquema de sólo lectura, que permite una mayor flexibilidad en la gestión de la estructura de datos.
- **Flexible y escalable**: admite el almacenamiento y análisis de datos estructurados y no estructurados, proporcionando un alto rendimiento de consulta gracias a la optimización a nivel de almacenamiento.

Data Lakehouse ofrece una solución de compromiso que combina las ventajas de ambos enfoques, por lo que resulta ideal para las cargas de trabajo analíticas modernas que requieren flexibilidad en el procesamiento de datos.

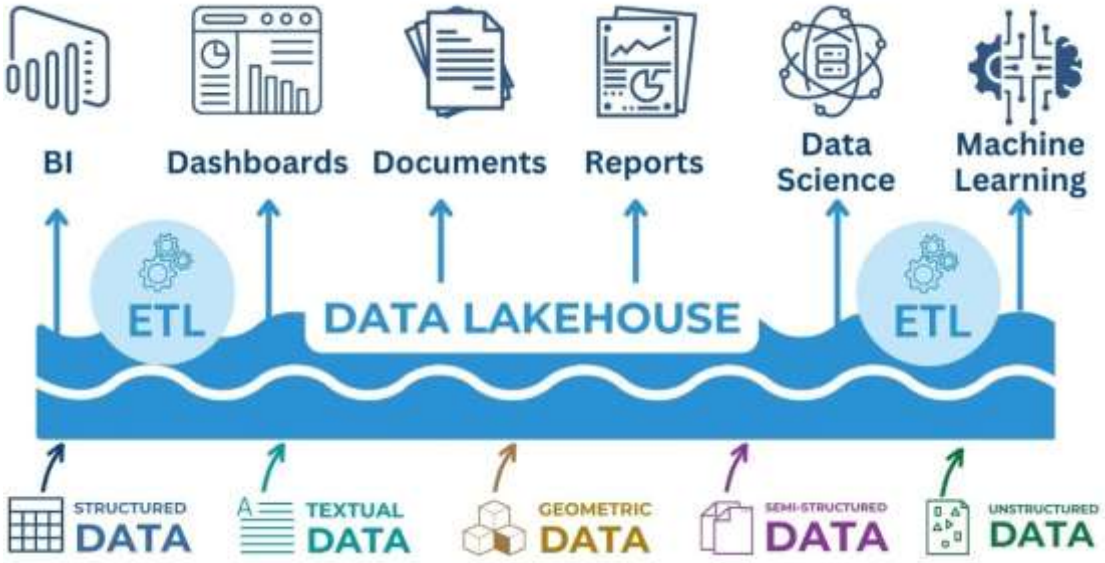


Fig. 8.1-9 Data Lakehouse es la nueva generación de sistemas de almacenamiento diseñados para satisfacer requisitos complejos y en constante cambio.

La idea que subyace a los almacenes de datos modernos parece sencilla: si todos los datos están en un mismo lugar, es más fácil analizarlos. Sin embargo, en la práctica no todo es tan sencillo. Imaginemos que una empresa decide abandonar por completo los sistemas contables y de gestión habituales (ERP, PMIS, CAFM u otros), sustituyéndolos por un enorme lago de datos al que todo el mundo tiene acceso. ¿Qué ocurrirá? Lo más probable es que se produzca un caos: los datos estarán duplicados, serán contradictorios y se perderá o corromperá información crítica. Incluso si el lago de datos sólo se utiliza para análisis, sin

una gestión adecuada, se verá gravemente comprometido:

- Los datos son difíciles de entender: en los sistemas convencionales, los datos tienen una estructura clara, pero en un lago no son más que una enorme acumulación de archivos y tablas. Para encontrar algo, especialista tiene que averiguar de qué es responsable cada fila y cada columna.
- Los datos pueden ser inexactos: si se almacenan muchas versiones de la misma información en un mismo lugar, es difícil saber qué versión está actualizada. Como consecuencia, se toman decisiones basadas en datos obsoletos o erróneos.
- Es difícil preparar los datos para el trabajo: no sólo hay que almacenarlos, sino también presentarlos de forma conveniente: en forma de informes, gráficos, tablas. En los sistemas tradicionales esto se hace automáticamente, pero en los lagos de datos requiere un procesamiento adicional.

En consecuencia, cada concepto de data warehousing tiene sus propias características, enfoques de procesamiento y aplicaciones empresariales. Las bases de datos tradicionales se centran en las operaciones transaccionales, los almacenes de datos (DWH) proporcionan una estructura para la analítica, los lagos de datos (Data Lake) almacenan la información en bruto, y los almacenes híbridos (Data Lakehouse) combinan las ventajas de los DWH y los Data Lake (Fig, 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Fig. 8.1-10 DWH, Data Lake y Data Lakehouse: diferencias clave en tipos de datos, escenarios de uso, métodos de procesamiento y enfoques de almacenamiento.

La elección de una arquitectura de almacenamiento es un proceso complejo, que depende de las necesidades de la empresa, el volumen de información y los requisitos analíticos. Cada solución tiene sus pros y sus contras: DWH aporta estructura, Data Lake aporta flexibilidad y Lakehouse ofrece un equilibrio entre ambas. Las organizaciones rara vez se limitan a una única arquitectura de datos.

> Independientemente de la arquitectura elegida, los sistemas automatizados de gestión de datos son muy superiores a los métodos manuales. Minimizan los errores humanos, aceleran el tratamiento de la información y garantizan la transparencia y la trazabilidad de los datos en todas las fases de los procesos empresariales.

Y mientras que los almacenes de datos centralizados ya se han convertido en una norma del sector en muchas áreas de la economía, la situación en la construcción sigue estando fragmentada. Los datos aquí se distribuyen a través de diferentes plataformas (CDE, PMIS, ERP, etc.), lo que dificulta la creación de una imagen unificada de lo que está sucediendo y requiere arquitecturas capaces de combinar estas fuentes en un entorno digital holístico y analíticamente utilizable.

CDE, PMIS, ERP o DWH y Data Lake

Algunas empresas de construcción e ingeniería ya utilizan el concepto de Entorno Común de Datos (CDE) según la norma ISO 19650. En esencia, el CDE desempeña las mismas funciones que un almacén de datos (DWH) en otras industrias: centralizar la información, proporcionar control de versiones, facilitar el acceso a información validada.

> **Un Entorno Común de Datos (CDE**) es un espacio digital centralizado utilizado para gestionar, almacenar, compartir y colaborar en la información del proyecto a lo largo de todas las fases del ciclo de vida de una instalación. El CDE suele implementarse utilizando tecnologías basadas en la nube e integradas con sistemas CAD (BIM).

Los sectores financiero, minorista, logístico e industrial llevan décadas utilizando sistemas centralizados de gestión de datos, combinando información de distintas fuentes, controlando su relevancia y proporcionando análisis. CDE lleva estos principios más allá adaptándolos a los retos del diseño de edificios y la gestión del ciclo de vida.

Al igual que DWH, CDE estructura los datos, captura los cambios y proporciona un único punto de acceso a la información verificada. Con el paso a la nube y la integración con herramientas analíticas, las diferencias entre ambos son cada vez menos evidentes. Si añadimos al CDE los datos granulares, cuyo concepto llevan debatiendo los proveedores de CAD desde 2023[93, 125], se pueden ver aún más paralelismos con el DWH clásico.

En el capítulo "Sistemas ERP y PMIS para la construcción" ya hemos hablado del PMIS (Sistema de Información para la Gestión de Proyectos) y del ERP (Planificación de Recursos Empresariales). En los proyectos de construcción, el CDE y el PMIS trabajan juntos: el CDE sirve como repositorio de datos, incluidos planos, modelos y documentación del proyecto, mientras que el PMIS gestiona procesos como el control de plazos, tareas, recursos y presupuestos.

ERP, responsable de la gestión de la empresa en su conjunto (finanzas, compras, personal, producción), puede integrarse con PMIS, proporcionando control de costes y presupuestos a nivel de empresa. Para la analítica y la elaboración de informes, DWH puede utilizarse para recopilar, estructurar y agregar datos de CDE, PMIS y ERP para evaluar KPI financieros (ROI) e identificar patrones. A su vez, Data Lake (DL) puede complementar a DWH almacenando datos en bruto y no estructurados (por ejemplo, registros, datos de sensores, imágenes). Estos datos pueden procesarse y cargarse en el DWH para su posterior análisis.

Así, CDE y PMIS se centran en la gestión de proyectos, ERP se centra en los procesos empresariales, y DWH y Data Lake se centran en la analítica y las operaciones de datos.

Al comparar los sistemas CDE, PMIS y ERP con los DWH y Data Lake, se observan diferencias significativas en términos de independencia del proveedor, coste, flexibilidad de integración, independencia de los datos, velocidad de adaptación al cambio y capacidades analíticas (Fig, 8.1-11). Los sistemas tradicionales, como CDE, PMIS y ERP, suelen estar vinculados a soluciones y estándares de proveedores específicos, lo que los hace menos flexibles y aumenta su coste debido a las licencias y el soporte. Además, los datos de estos sistemas suelen estar encapsulados en formatos patentados y cerrados, lo que limita su uso y análisis.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Fig. 8.1-11 DWH y Data Lake ofrecen mayor flexibilidad e independencia de datos que sistemas como CDE, PMIS y ERP.

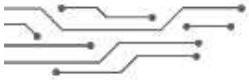
Por el contrario, DWH y Data Lake proporcionan una mayor flexibilidad en la integración con diferentes fuentes de datos, y su uso de tecnologías y plataformas abiertas ayuda a reducir el coste total de propiedad. Además, DWH y Data Lake admiten una amplia gama de herramientas analíticas, lo que mejora las capacidades de análisis y gestión.

> Con el desarrollo de herramientas de ingeniería inversa para formatos CAD y el acceso a bases de datos de aplicaciones CAD, la pregunta se hace cada vez más acuciante: ¿hasta qué punto está justificado seguir utilizando plataformas cerradas y aisladas si los datos de diseño deben estar a disposición de un amplio abanico de especialistas que trabajan en decenas de contratistas y organizaciones de diseño?

Esta dependencia tecnológica de un proveedor específico puede limitar considerablemente la flexibilidad de la gestión de datos, ralentizar las respuestas a los cambios del proyecto e inhibir la colaboración eficaz entre los participantes.

Los enfoques tradicionales de la gestión de datos -incluidos DWH, Data Lake, CDE y PMIS - se han centrado principalmente en almacenar, estructurar y procesar la información. Sin embargo, con el desarrollo de la inteligencia artificial y el aprendizaje automático, crece la necesidad de nuevas formas de organizar los datos que no solo agreguen, sino que también revelen relaciones complejas, encuentren patrones ocultos y proporcionen acceso instantáneo a la información más relevante.

Las bases de datos vectoriales -un nuevo tipo de almacenamiento optimizado para incrustaciones de alta dimensión- empiezan a desempeñar un papel especial en esta dirección.



CAPÍTULO 8.2.
GESTIÓN DE ALMACENES DE DATOS Y PREVENCIÓN DEL CAOS

Bases de datos vectoriales y el cuadro delimitador

Las bases de datos vectoriales son una nueva clase de repositorios que no se limitan a almacenar datos, sino que permiten buscar por significado, comparar objetos por proximidad semántica y crear sistemas inteligentes: desde recomendaciones hasta análisis automáticos y generación de contexto. A diferencia de las bases de datos tradicionales, que se centran en las coincidencias exactas, las bases de datos vectoriales encuentran objetos similares basándose en atributos, aunque no haya una coincidencia exacta

> **Una** base de datos **vectorial** es un tipo especializado de base de datos que almacena datos en forma de vectores multidimensionales, cada uno de los cuales representa determinadas características o cualidades. Estos vectores pueden tener diferentes números de dimensiones, dependiendo de la complejidad de los datos (en un caso pueden ser unas pocas dimensiones, y en otro — miles).

La principal ventaja de las bases de datos vectoriales es la búsqueda por relevancia semántica en lugar de por coincidencia exacta de valores. En lugar de SQL- y Pandas -consultas con filtros "iguala" o "contiene", se utiliza la búsqueda de vecinos más cercanos (k-NN) (hablaremos más sobre k-NN en la siguiente parte del libro) en el espacio de características.

Con el desarrollo de LLM (Large Language Models) y modelos generativos, la interacción con las bases de datos está empezando a cambiar. Ahora es posible consultar datos en lenguaje natural, realizar búsquedas semánticas en documentos, extraer automáticamente términos clave y construir relaciones contextuales entre objetos, todo ello sin necesidad de dominar el lenguaje SQL ni conocer la estructura de las tablas. Esto se trató con más detalle en la sección "Los LLM y su papel en el tratamiento de datos y los procesos empresariales".

Sin embargo, es importante tener en cuenta que los LLM no estructuran ni ordenan automáticamente la información. El modelo se limita a flotar entre los datos y encontrar los más relevantes en función del contexto de la consulta. Si los datos no se han limpiado o transformado previamente, la búsqueda profunda será como intentar encontrar una respuesta en la "basura" digital: puede que funcione, pero la calidad de los resultados será menor. Lo ideal sería poder estructurar los datos (por ejemplo, traducir los documentos a Markdown) y cargarlos en una base de datos vectorial. Esto aumenta significativamente la precisión y relevancia del resultado.

> Inicialmente, las bases de datos vectoriales se utilizaban en el aprendizaje automático, pero hoy en día encuentran cada vez más aplicaciones fuera de él: en motores de búsqueda, personalización de contenidos y analítica inteligente.

Uno de los ejemplos más evidentes del enfoque vectorial en la construcción es el Bounding Box (paralelepípedo delimitador). Se trata de una construcción geométrica que describe los límites de un objeto en el espacio tridimensional. El Bounding Box está definido por las coordenadas X, Y y Z mínimas y máximas, formando una "caja" alrededor del objeto. Este método permite estimar el tamaño y la colocación de un elemento sin tener que analizar toda la geometría.

Cada Bounding Box puede representarse como un vector en un espacio multidimensional: por ejemplo [x, y, z, anchura, altura, profundidad] - ya 6 dimensiones (Fig. 8.2-1).

Bounding Box

	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Fig. 8.2-1 Bounding Box -La información de coordenadas de los elementos y su ubicación en el modelo de proyecto es análoga a una base de datos vectorial.

Esta representación de datos facilita muchas tareas, como la comprobación de intersecciones entre objetos, la planificación de la distribución espacial de elementos de construcción y la realización de cálculos automatizados. Bounding Box puede servir de puente entre los modelos 3D complejos y las bases de datos vectoriales tradicionales, permitiendo aprovechar eficazmente las ventajas de ambos enfoques en el modelado arquitectónico y de ingeniería

> Bounding Box es "vectorización de la geometría", e incrustación (una forma de transformar algo abstracto) es "vectorización del significado". Ambos enfoques permiten pasar de la búsqueda manual a la búsqueda inteligente, ya sea en 3D -objetos en un modelo de proyecto o conceptos en un texto.

La búsqueda de objetos en el proyecto (por ejemplo, "encontrar todas las ventanas con anchura > 1,5 m") es similar a la búsqueda de vecinos más próximos (k-NN) en una base de datos vectorial, donde los criterios definen una "zona" en el espacio de características. (Hablaremos más sobre la búsqueda de vecinos más próximos k-NN en la siguiente parte sobre aprendizaje automático) (Fig. 8.2-2). Si añadimos parámetros

adicionales (material, peso, tiempo de producción) a los atributos del cuadro delimitador, la tabla se convierte en un vector de alta dimensión, donde cada atributo es una nueva dimensión. Esto se acerca más a las bases vectoriales modernas, donde las dimensiones se cuentan por cientos o miles (por ejemplo, incrustando de redes neuronales).

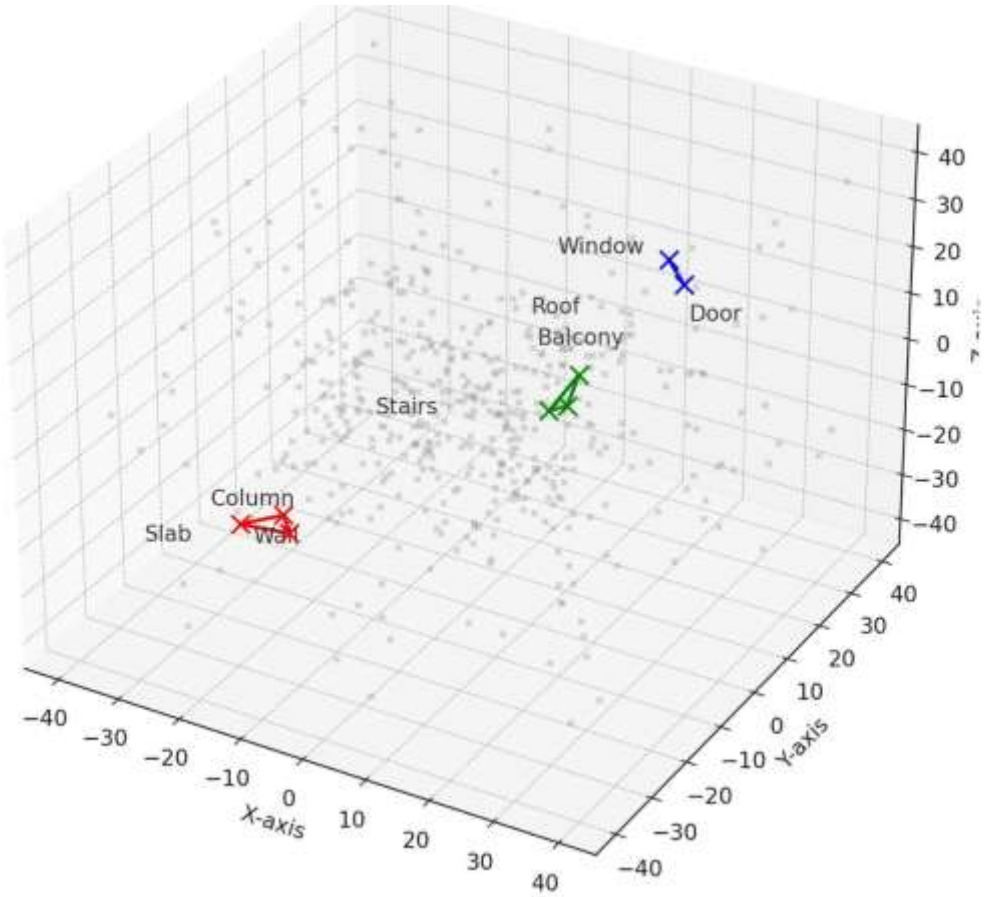


Fig. 8.2-2 Búsqueda de objetos en el proyecto mediante bases de datos vectoriales.

El planteamiento utilizado en Bounding Box, es aplicable no sólo a los objetos geométricos, sino también al análisis de texto y lenguaje. Las representaciones vectoriales de datos ya se utilizan activamente en el procesamiento del lenguaje natural (PLN). Al igual que los objetos de un proyecto de construcción pueden agruparse según su proximidad espacial (Fig, 8.2-2), las palabras de un texto pueden analizarse según su proximidad semántica y contextual.

Por ejemplo, las palabras "arquitecto", "construcción", "diseño" estarán una al lado de la otra en el espacio vectorial porque tienen un significado similar. En LLM este mecanismo permite una categorización automática, sin necesidad de manual:

- Identificar el tema de un texto
- Realizar búsquedas semánticas en el contenido de los documentos

- Generación automática de anotaciones y resúmenes de texto
- Buscar sinónimos y términos relacionados

Las bases de datos vectoriales permiten analizar textos y encontrar en ellos términos relacionados, del mismo modo que Bounding Box ayuda a analizar objetos espaciales en modelos 3D. El ejemplo de Bounding Box de los elementos de un proyecto ayuda a comprender que la representación vectorial no es un concepto puramente "artificial" de ML, sino una forma natural de estructurar los datos para resolver problemas aplicados, ya se trate de buscar columnas en un proyecto CAD o de acercar semánticamente imágenes en una base de datos.

> Los especialistas que trabajan con bases de datos deberían prestar atención a los almacenes vectoriales. Su proliferación indica una nueva etapa en el desarrollo de bases de datos, en la que los sistemas relacionales clásicos y las tecnologías orientadas a la IA empiezan a entrelazarse, formando las soluciones híbridas del futuro.

Los usuarios que desarrollen aplicaciones de IA complejas y a gran escala utilizarán bases de datos especializadas para la búsqueda vectorial. Al mismo tiempo, los que sólo necesiten funciones de IA independientes para integrarlas en aplicaciones existentes optarán más probablemente por las capacidades de búsqueda vectorial incorporadas en las bases de datos que ya utilizan (PostgreSQL, Redis).

Aunque sistemas como DWH, Data Lake, CDE, PMIS, bases de datos vectoriales y otros ofrecen diferentes enfoques para el almacenamiento y la gestión de datos, su eficacia viene determinada no solo por su arquitectura, sino también por lo bien que se organicen y gestionen los propios datos. Incluso cuando se utilizan soluciones modernas -ya sean bases de datos vectoriales, SGBD relacionales clásicos o almacenes tipo Data Lake-, la falta de reglas claras para gestionar, estructurar y actualizar los datos puede provocar las mismas dificultades a las que se enfrentan los usuarios que trabajan con archivos dispares y datos multiformato.

Sin Data Governance), incluso las soluciones más potentes pueden volverse caóticas y desestructuradas, convirtiendo los lagos de datos en pantanos de datos). Para evitarlo, las empresas no solo deben elegir la arquitectura de almacenamiento adecuada, sino también aplicar estrategias de minimización de datos), gestión de accesos y control de calidad para convertir los datos en una herramienta eficaz de toma de decisiones.

Gobernanza de datos), Minimalismo de datos) y Pantano de datos)

Comprender y aplicar los conceptos de gobernanza de datos (Data Governance)), minimalismo de datos (Data Minimalism)) y prevención del pantano de datos (Data Swamp)) son fundamentales para gestionar con éxito los almacenes de datos y aportar valor empresarial (Fig, 8.2-3).

> Según un estudio de Gartner (2017), el 85% de los proyectos de big data fracasan, y una de las razones clave es la insuficiente calidad de los datos y gobernanza de datos [144].

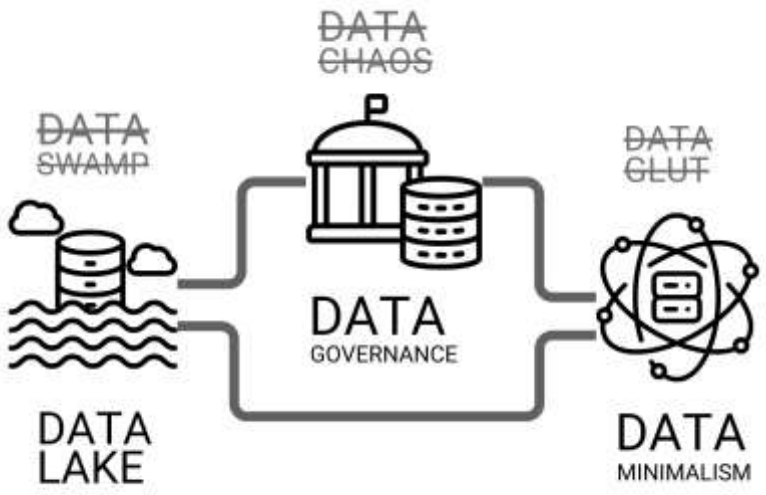


Fig. 8.2-3 Algunos de los aspectos clave de la gobernanza de datos son la Gobernanza de Datos y el Minimalismo de Datos.

La gobernanza de datos (Data Governance) es un componente fundamental de la gestión de datos, ya que garantiza que éstos se utilicen de forma adecuada y eficaz en todos los procesos empresariales. No se trata sólo de establecer normas y procedimientos, sino también de garantizar la disponibilidad, fiabilidad y seguridad de los datos:

- Definir y clasificar los datos de: definir y clasificar claramente las entidades permite a las organizaciones comprender qué entidades son necesarias en la empresa y determinar cómo deben utilizarse.
- Derechos de acceso y gestión: la elaboración de políticas y procedimientos de acceso y gestión de datos garantiza que sólo los usuarios autorizados puedan acceder a determinados datos.
- Proteger los datos de amenazas externas: Proteger los datos de amenazas externas es un aspecto clave de la gestión de datos. Esto incluye no solo medidas técnicas, sino también la formación de los empleados en los aspectos básicos de la seguridad de la información.

El Minimalismo de Datos (Data Minimalism) es un enfoque para reducir los datos a los atributos y entidades más valiosos y significativos de la formación (Fig, 8.2-4), reduciendo así los costes y mejorando la utilización de los datos:

- Simplificación de la toma de decisiones: reducir el número de objetos y sus atributos a los más relevantes simplifica la toma de decisiones al reducir el tiempo y los recursos necesarios para analizar y procesar los datos.
- Centrarse en lo importante: seleccionar las entidades y atributos más relevantes permite centrarse en la información que realmente importa a la empresa, eliminando el ruido y los datos innecesarios.
- Asignación eficaz de los recursos: la minimización de los datos permite una asignación más eficaz de los recursos, reduciendo los costes de almacenamiento y tratamiento de los datos y mejorando su calidad y seguridad.

La lógica de trabajo con los datos no debe comenzar con su creación como tal (Fig. 8.2-4), sino con la

comprensión de los escenarios futuros de utilización de estos datos incluso antes de que comience el proceso de generación. Este enfoque permite definir de antemano los requisitos mínimos necesarios para los atributos, sus tipos y valores límite. Estos requisitos constituyen la base para crear entidades correctas y estables en el modelo de información. La comprensión preliminar de los propósitos y usos de los datos contribuye a la formación de una estructura adecuada para el análisis. En el capítulo "Modelización de datos: modelo conceptual, lógico y físico" se trataron más detalles sobre los enfoques de la modelización de datos en los niveles conceptual, lógico y físico.

> En los procesos empresariales tradicionales de las empresas de construcción, el tratamiento de datos suele parecerse más al vertido de datos en un pantano, donde primero se crean los datos y luego los especialistas intentan integrarlos en otros sistemas y herramientas.

El pantano de datos (Data Swamp) es el resultado de la recopilación y almacenamiento incontrolados de datos sin una organización, estructuración y gestión adecuadas, lo que da lugar a datos desestructurados, difíciles de utilizar y de escaso valor.

Cómo evitar que el flujo de información se convierta en un atolladero:

- **Gestión de la estructura de los datos:** garantizar que los datos de estén estructurados y categorizados ayuda a evitar la saturación de datos al hacerlos ordenados y fácilmente accesibles.
- **Comprensión e interpretación de** los datos: una descripción clara de los orígenes, modificaciones y significados de los datos garantiza su correcta comprensión e interpretación.
- **Mantenimiento de la calidad de los datos**: el mantenimiento y la limpieza periódicos de los datos ayudan a mantener su calidad, pertinencia y valor para los análisis y los procesos empresariales.

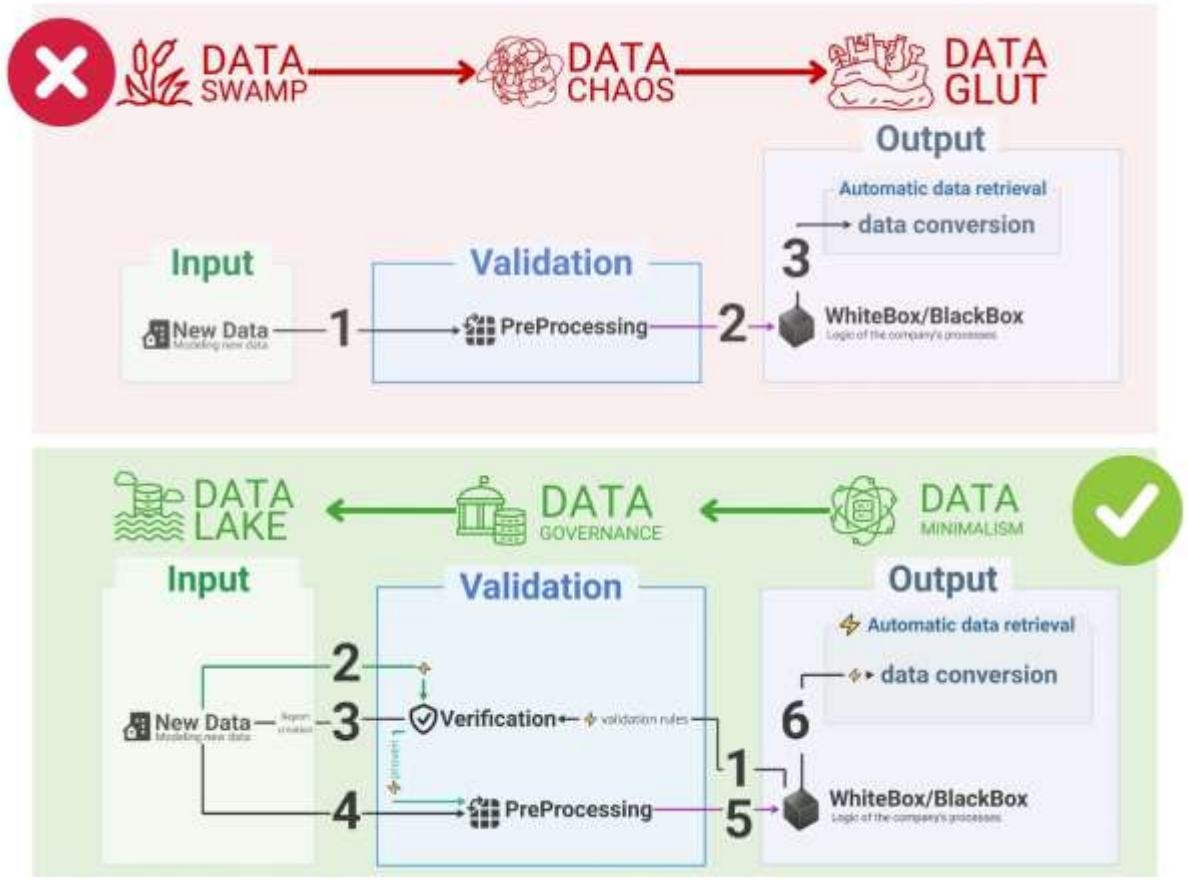


Fig. 8.2-4 Para evitar el desorden en el almacén de datos, debe iniciar el proceso de creación de datos recopilando los requisitos de atributos.

> Integrando los principios de gobernanza y minimalismo de datos en los procesos de gestión de datos, y evitando activamente que los almacenes de datos se conviertan en pantanos de datos, las organizaciones pueden maximizar el potencial de sus datos.

La siguiente etapa en la evolución del trabajo con datos, tras resolver las cuestiones de gestión y minimalismo, es la estandarización del procesamiento automático, el aseguramiento de la calidad y la implementación de métodos que hagan que los datos sean utilizables para el análisis, la transformación y la toma de decisiones. Esto es lo que están haciendo las metodologías DataOps y VectorOps, que se están convirtiendo en herramientas importantes para las empresas que trabajan con big data y aprendizaje automático.

DataOps y VectorOps: nuevos estándares de datos

Mientras que Data Governance se encarga de controlar y organizar los datos, DataOps ayuda a garantizar su exactitud, coherencia y fluidez dentro de la empresa. Esto es especialmente crítico para una serie de casos empresariales en la construcción, donde los datos se generan continuamente y requieren un procesamiento oportuno. Por ejemplo, en situaciones en las que los modelos de información del edificio, los

requisitos del proyecto y los informes analíticos deben sincronizarse entre distintos sistemas en una misma jornada laboral, el papel de DataOps puede ser clave. Permite construir procesos de procesamiento de datos estables y repetibles, reduciendo el riesgo de retrasos y pérdida de relevancia de la información.

La gobernanza de datos por sí sola no es suficiente: es esencial que los datos no sólo se almacenen, sino que se utilicen activamente en las operaciones diarias. Aquí es donde entra en juego DataOps, una metodología centrada en la automatización, la integración y el flujo continuo de datos.

> **DataOps** se centra en mejorar la colaboración, integración y automatización de los flujos de datos en las organizaciones. La adopción de prácticas de DataOps favorece la precisión, coherencia y disponibilidad de los datos, lo que es fundamental para las aplicaciones centradas en los datos.

Las herramientas clave del ecosistema DataOps son Apache Airflow (Fig, 7.4-4) para la orquestación de flujos de trabajo, y Apache NiFi (Fig, 7.4-5) para el enrutamiento y transformación de flujos de datos. Juntas, estas tecnologías permiten canalizaciones de datos flexibles, fiables y escalables para procesar, controlar e integrar automáticamente la información entre sistemas (más detalles en el capítulo "ETL automático - conveyor "). A la hora de aplicar el enfoque DataOps en los procesos de construcción, es importante tener en cuenta cuatro aspectos fundamentales:

1. **Las personas y las herramientas son más importantes que los datos**: depósitos de datos en silos puede considerarse un problema importante, pero la realidad es más compleja. Además de la fragmentación de los datos, el aislamiento de los equipos y las herramientas dispares que utilizan desempeñan un papel importante. En la construcción, especialistas de distintas disciplinas trabajan con datos: ingenieros y analistas de datos, equipos de BI y visualización, así como expertos en gestión de proyectos y calidad. Cada uno de ellos tiene formas distintas de trabajar, por lo que resulta importante crear un ecosistema en el que los datos fluyan libremente entre los participantes, proporcionando una versión única y coherente de la información.
2. **Automatice las pruebas y la detección de errores**: los datos de construcción siempre contienen errores, ya sean imprecisiones en los modelos, errores de cálculo o especificaciones obsoletas. Comprobar regularmente los datos y eliminar los errores recurrentes puede mejorar significativamente la calidad de los datos. Como parte de DataOps, es necesario implantar controles automatizados y mecanismos de validación que supervisen la corrección de los datos, analicen los errores e identifiquen patrones, y capturen y aborden los fallos del sistema en cada flujo de trabajo. Cuanto mayor sea el grado de validación automatizada, mayor será la calidad general de los datos y menor la probabilidad de errores en las fases finales.
3. **Los datos deben comprobarse del mismo modo que el código de los programas**: la mayoría de las aplicaciones de construcción se basan en el procesamiento de datos, pero su control suele dejarse en manos de funciones secundarias. Si los modelos de aprendizaje automático se entrenan con datos inexactos, se producen predicciones erróneas y pérdidas económicas. En DataOps, los datos deben someterse al mismo escrutinio riguroso que el código de software: comprobaciones lógicas, pruebas de estrés, evaluación del comportamiento de los modelos cuando cambian los valores de entrada. Sólo los datos validados y fiables pueden servir de base para las decisiones de gestión.
4. **Observabilidad de los datos sin comprometer el rendimiento**: la supervisión de los datos no es sólo una colección de métricas, sino una herramienta estratégica de gestión de la calidad. Para

que DataOps funcione eficazmente, la observabilidad debe integrarse en todas las fases del tratamiento de datos, desde el diseño hasta la explotación. Al mismo tiempo, es importante que la supervisión no ralentice el sistema. En los proyectos de construcción, es fundamental no sólo recopilar datos, sino hacerlo de manera que no se interrumpa en modo alguno el trabajo de los profesionales (por ejemplo, los diseñadores) que los crean. Este equilibrio permite controlar la calidad de los datos sin comprometer la productividad.

DataOps no es una carga adicional para los científicos de datos, sino la columna vertebral de su trabajo. Al implantar DataOps, las empresas de construcción pueden pasar de una gestión de datos caótica a un ecosistema eficiente en el que los datos trabajan para el negocio.

Por su parte, VectorOps representa la siguiente etapa en la evolución de DataOps, centrada en el procesamiento, almacenamiento y análisis de datos vectoriales multidimensionales (de los que se habló en el capítulo anterior). Esto es especialmente relevante en ámbitos como los gemelos digitales, los modelos de redes neuronales y la búsqueda semántica, que están empezando a introducirse en el sector de la construcción. VectorOps se basa en las bases de datos vectoriales para almacenar, indexar y buscar eficazmente representaciones multidimensionales de objetos.

> **VectorOps** es el paso siguiente a DataOps, centrado en el procesamiento, el análisis y el uso de datos vectoriales en la construcción. A diferencia de DataOps, que se centra en el flujo, la coherencia y la calidad de los datos, VectorOps se centra en la gestión de las representaciones de objetos multidimensionales necesarias para el aprendizaje automático.

A diferencia de los enfoques tradicionales, VectorOps permite lograr descripciones de objetos más precisas, lo que resulta fundamental para los gemelos digitales, los sistemas de diseño generativo y la detección automática de errores en datos CAD convertidos a formato vectorial. La implementación combinada de DataOps y VectorOps constituye una base sólida para el trabajo escalable y automatizado con grandes volúmenes de información, desde tablas clásicas hasta modelos espaciales semánticamente ricos

Los próximos pasos: del almacenamiento caótico al almacenamiento estructurado

Los enfoques tradicionales de almacenamiento de datos suelen dar lugar a la creación de "silos de información" dispares en los que no es posible acceder a información importante para el análisis y la toma de decisiones. Los conceptos modernos de almacenamiento, como Data Warehouse, Data Lake y sus híbridos, permiten unificar información dispar y ponerla a disposición de forma centralizada para el flujo de datos y la inteligencia empresarial. No sólo es importante elegir la arquitectura de almacenamiento adecuada, sino también aplicar la Gobernanza de Datos) y el Minimalismo de Datos) para evitar que las instalaciones de almacenamiento se conviertan en Pantanos de Datos incontrolables).

Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar los conceptos tratados a sus tareas cotidianas:

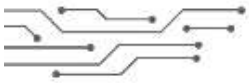
- Seleccionar formatos eficientes de almacenamiento de datos

- ☐ Pasar de CSV y XLSX a formatos más eficientes (Apache Parquet, ORC) para almacenar grandes cantidades de datos.
- ☐ Implantar un sistema de control de versiones de datos para rastrear los cambios
- ☐ Utilizar metadatos para describir la estructura y procedencia de la información

- Crear una arquitectura unificada de datos de la empresa
  - ☐ Compare diferentes arquitecturas de almacenamiento: RDBMS, DWH y Data Lake. Elija la que mejor se adapte a sus necesidades de escalabilidad, integración de fuentes y procesamiento analítico.
  - ☐ Diseñe un mapa de procesos para extraer, cargar y transformar datos (ETL) de varias fuentes para sus tareas. Utiliza herramientas de visualización como Miro, Lucidchart o Draw.io para visualizar los pasos clave y los puntos de integración.
- Implantar prácticas de Gobernanza de Datos y Minimalismo de Datos
  - ☐ Siga el enfoque del Minimalismo de Datos - sólo almacene y procese lo que sea realmente valioso.
  - ☐ Aplicar los principios de gobernanza de datos - definir la responsabilidad de los datos, garantizar la calidad y la transparencia.
  - ☐ Más información sobre políticas de gestión de datos y conceptos de DataOps, VectorOps
  - ☐ Definir los criterios de calidad de los datos y los procedimientos de validación de datos en DataOps.

Un almacenamiento de datos bien organizado crea la base para centralizar los procesos analíticos de una empresa. La transición de la acumulación caótica de archivos a los almacenamientos estructurados permite convertir la información en un activo estratégico que ayuda a tomar decisiones informadas y a mejorar la eficiencia de los procesos empresariales.

Una vez automatizados y normalizados los procesos de recopilación, transformación, análisis y almacenamiento estructurado de datos, la siguiente etapa de la transformación digital es el tratamiento integral de Big Data.



IX PARTE
BIG DATA, APRENDIZAJE AUTOMÁTICO Y PREDICCIONES

La novena parte se centra en los macrodatos, el aprendizaje automático y el análisis predictivo en el sector de la construcción. Explora la transición de la toma de decisiones intuitiva a los análisis objetivos basados en datos históricos. Se utilizan ejemplos prácticos para demostrar el análisis de big data en la construcción, desde el análisis sintáctico del conjunto de datos de permisos de construcción de San Francisco hasta el procesamiento de CAD, proyectos con millones de elementos. Se presta especial atención a los métodos de aprendizaje automático para predecir el coste y el calendario de los proyectos de construcción, con un análisis detallado de los algoritmos de regresión lineal y k-nearest neighbours. Se muestra cómo los datos estructurados se convierten en la base de modelos predictivos para evaluar riesgos, optimizar recursos y mejorar la eficiencia de la gestión de proyectos. La parte también ofrece recomendaciones sobre cómo seleccionar muestras de datos representativas y explica por qué no siempre se necesitan grandes conjuntos de datos para realizar análisis eficaces.

CAPÍTULO 9.1.
BIG DATA Y SU ANÁLISIS

Big data en la construcción: de la intuición a la previsibilidad

El término "big data" no tiene una definición estricta. El concepto apareció originalmente cuando el volumen de información empezó a superar las capacidades de los métodos tradicionales para su procesamiento. Hoy en día, el volumen y la complejidad de los datos en muchas industrias, incluida la de la construcción, han aumentado tanto que no caben en la memoria local de los ordenadores y requieren el uso de nuevas tecnologías para procesarlos.

La esencia del trabajo con Big Data no es sólo el almacenamiento y el procesamiento, sino también la capacidad de predicción. En el sector de la construcción, los Big Data abren el camino desde las decisiones intuitivas basadas en la interpretación subjetiva de tablas y visualizaciones (como se ha comentado antes) hasta las previsiones informadas respaldadas por observaciones y estadísticas reales.

Contrariamente a la creencia popular, el objetivo de trabajar con big data no es "hacer que una máquina piense como un humano", sino aplicar modelos matemáticos y algoritmos para analizar cantidades masivas de datos con el fin de identificar patrones, predecir acontecimientos y optimizar procesos.

> **Los macrodatos** no son un mundo frío de algoritmos desprovistos de influencia humana. Al contrario, los macrodatos funcionan en conjunción con nuestros instintos, errores y creatividad. Es la imperfección del pensamiento humano lo que nos permite encontrar soluciones no estándar y hacer grandes avances.

Con el desarrollo de la tecnología digital, la industria de la construcción ha empezado a utilizar activamente técnicas de tratamiento de datos procedentes del sector informático. Gracias a herramientas como Pandas y Apache Parquet, es posible combinar datos estructurados y no estructurados, lo que simplifica el acceso a la información y reduce las pérdidas hasta el análisis, mientras que los grandes conjuntos de datos procedentes de documentos o proyectos CAD (Fig. 9.2-10 - Fig. 9.2-12) permiten recopilar, analizar y predecir datos en todas las fases del ciclo de vida del proyecto.

El Big Data está teniendo un impacto transformador en el sector de la construcción, influyendo potencialmente en él de diversas maneras. La aplicación de las tecnologías de Big Data está dando resultados en una serie de áreas clave, entre las que se incluyen, por ejemplo, las siguientes:

- **Análisis del potencial de inversión**: previsión de la rentabilidad y los periodos de amortización de los proyectos a partir de datos de instalaciones anteriores.
- **Mantenimiento predictivo**: identificación de averías probables en los equipos antes de que se produzcan realmente, lo que reduce el tiempo de inactividad.
- **Optimización de la cadena de suministro**: previsión de interrupciones y mejora de la eficacia logística.

- **Análisis de eficiencia energética**: ayuda al diseño de edificios de bajo consumo energético.
- **Supervisión de la seguridad**: uso de sensores y dispositivos portátiles para controlar las condiciones del lugar.
- **Control de calidad**: supervisión en tiempo real del cumplimiento de las normas del proceso.
- **Gestión de recursos humanos**: análisis del rendimiento y previsión de las necesidades de personal.

Es difícil encontrar un sector de la construcción en el que no se demanden análisis de datos y predicciones. La principal ventaja de los algoritmos de predicción es su capacidad de autoaprendizaje y mejora constante a medida que se acumulan datos.

> En un futuro próximo, la inteligencia artificial no se limitará a ayudar a los constructores, sino que tomará decisiones clave, desde los procesos de diseño hasta las cuestiones de funcionamiento de los edificios.

En la siguiente parte del libro, "Aprendizaje automático y predicciones", se explica cómo se generan las predicciones y se utilizan los modelos de aprendizaje.

La transición al trabajo pleno con big data exige un cambio en el propio planteamiento de la analítica. Mientras que los sistemas clásicos que hemos considerado hasta ahora se centraban en las relaciones causa-efecto, la analítica de big data está cambiando su enfoque hacia la búsqueda de patrones estadísticos y correlaciones que nos permitan identificar relaciones ocultas y predecir el comportamiento de los objetos incluso sin una comprensión completa de todos los factores.

Cuestionar la viabilidad de los macrodatos: correlación, estadística y muestreo de datos

Tradicionalmente, la construcción se basaba en hipótesis subjetivas y en la experiencia personal. Los ingenieros suponían -con cierto grado de probabilidad- cómo se comportaría el material, qué cargas soportaría la estructura y cuánto duraría el proyecto. Estas hipótesis se comprobaban en la práctica, a menudo a costa de tiempo, recursos y riesgos futuros.

> Con la llegada de los macrodatos, el planteamiento está cambiando radicalmente: las decisiones ya no se toman sobre la base de corazonadas intuitivas, sino como resultado del análisis de conjuntos de datos a gran escala. La construcción está dejando poco a poco de ser un arte de la intuición para convertirse en una ciencia precisa de la predicción.

La transición a la idea de utilizar big data plantea inevitablemente una pregunta importante: ¿hasta qué punto es decisiva la cantidad de datos y cuánta información se necesita realmente para un análisis predictivo fiable? La creencia generalizada de que "cuantos más datos, mayor precisión" no siempre resulta estadísticamente válida en la práctica.

> Ya en 1934, el estadístico Jerzy Neumann demostró [145] que la clave de la precisión de la inferencia estadística no reside tanto en la cantidad de datos como en su representatividad y en la aleatoriedad del muestreo.

Esto es especialmente cierto en el sector de la construcción, donde se recopilan grandes masas de datos mediante sensores IoT, escáneres, cámaras de vigilancia, drones e incluso modelos CAD multiformato, lo que aumenta el riesgo de puntos ciegos, valores atípicos y distorsiones de los datos.

Consideremos un ejemplo de control del estado del firme de una carretera. Un conjunto de datos completo de todos los tramos de carretera puede ocupar X GB y tardar aproximadamente un día en procesarse. Al mismo tiempo, una muestra aleatoria que incluyera sólo cada 50 tramos de carretera ocuparía sólo X/50 GB y se procesaría en media hora, proporcionando al mismo tiempo una precisión similar de las estimaciones para determinados cálculos (Fig, 9.1-1).

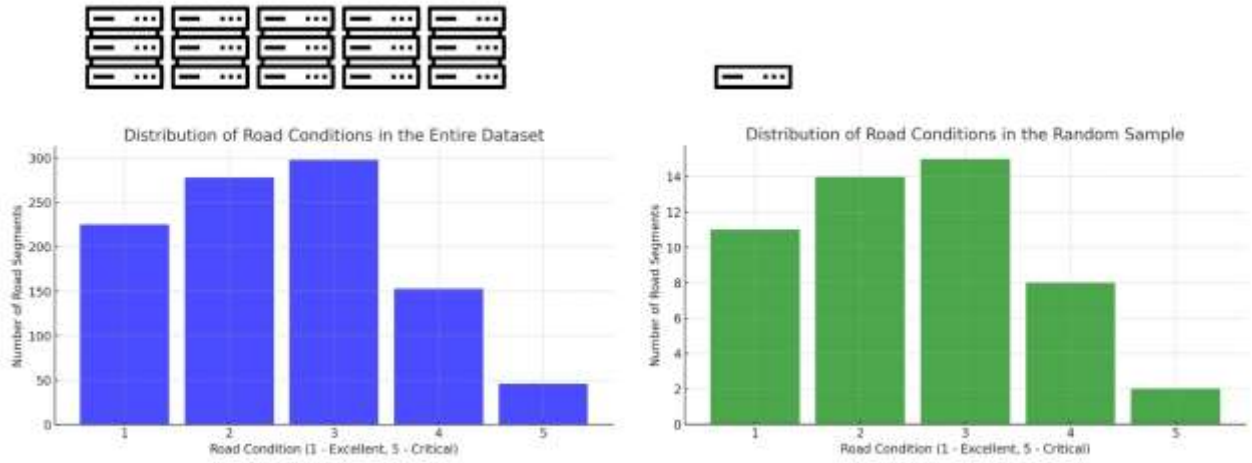


Fig. 9.1-1 Histogramas del estado del firme: el conjunto completo de datos y el muestreo aleatorio muestran resultados idénticos.

Así pues, la clave del éxito del análisis de datos puede no ser a menudo la cantidad de datos, sino la representatividad de la muestra y la calidad de los métodos de tratamiento utilizados. El paso al muestreo aleatorio y a un enfoque más selectivo exige un cambio de mentalidad en el sector de la construcción. Históricamente, las empresas han seguido la lógica de "cuantos más datos, mejor", creyendo que cubrir todos los indicadores posibles maximizaría la precisión.

Este planteamiento recuerda a un error popular de la gestión de proyectos: "cuantos más especialistas atraiga, más eficaz será el trabajo". Sin embargo, al igual que ocurre con los recursos humanos, lo más importante es la calidad y las herramientas, no la cantidad. Si no se tienen en cuenta las interrelaciones (correlaciones) entre los datos o los participantes en el proyecto, aumentar el volumen sólo puede generar ruido, distorsiones, duplicaciones y despilfarros innecesarios.

Al final, a menudo resulta que es mucho más productivo disponer de un conjunto de datos más reducido, pero cualitativamente preparado, capaz de producir previsiones estables y razonables, que basarse en una información masiva pero caótica que contenga muchas señales contradictorias.

> Un volumen excesivo de datos no sólo no garantiza una mayor precisión, sino que también puede dar lugar a conclusiones distorsionadas debido a la presencia de ruido, características redundantes, correlaciones ocultas e información irrelevante. En tales circunstancias, aumenta el riesgo de sobreajuste de los modelos y disminuye la fiabilidad de los resultados analíticos.

En el sector de la construcción, uno de los principales retos a la hora de tratar los macrodatos es determinar la cantidad y calidad óptimas de los datos. Por ejemplo, al supervisar el estado de las estructuras de hormigón, utilizar miles de sensores y recopilar información cada minuto puede saturar el sistema de almacenamiento y análisis. Sin embargo, si se realiza un análisis de correlación y se selecciona el 10% de los sensores más informativos, se puede obtener una precisión casi idéntica en las predicciones, gastando muchas veces, a veces decenas o cientos de veces, menos recursos.

El uso de un subconjunto más pequeño de datos reduce tanto la cantidad de almacenamiento necesario como el tiempo de procesamiento, lo que reduce significativamente el coste de almacenamiento y análisis de datos y a menudo convierte el muestreo aleatorio en una solución ideal para el análisis predictivo, especialmente en grandes proyectos de infraestructuras o cuando se trabaja en tiempo real. En última instancia, la eficiencia de los procesos de construcción no viene determinada por la cantidad de datos recopilados, sino por la calidad de su análisis. Sin un enfoque crítico y un análisis cuidadoso, los datos pueden llevar a conclusiones erróneas.

> A partir de cierta cantidad de datos, cada nueva unidad de información arroja resultados cada vez menos útiles. En lugar de recopilar información sin cesar, es importante centrarse en su representatividad y en los métodos de análisis (Fig, 9.2-2).

Este fenómeno está bien descrito por Allen Wallis [146], que ilustra el uso de métodos estadísticos con el ejemplo de la prueba de dos diseños alternativos de proyectiles de la US Navy.

> *La Marina probó dos diseños de proyectil alternativos (A y B) realizando una serie de rondas emparejadas. En cada ronda, A recibe un 1 o un 0 en función de si su rendimiento es mejor o peor que el de B, y viceversa. El enfoque estadístico estándar consiste en realizar un número fijo de ensayos (por ejemplo, 1000) y determinar el ganador basándose en una distribución porcentual (por ejemplo, si A obtiene un 1 más del 53% de las veces, se considera el mejor). Cuando Allen Wallis discutió este problema con el capitán Garrett L. Schuyler (de la Marina), el capitán objetó que tal prueba, citando la historia de Allen, podría ser inútil. Si un oficial de artillería sabio y experimentado como Schuyler hubiera estado en el lugar, habría visto después de los primeros cientos [de disparos] que el experimento no necesita terminarse, ya sea porque el nuevo método es claramente inferior o porque es claramente superior a lo que se esperaba [146].*
>
> - Grupo de Investigación Estadística del Gobierno de los Estados Unidos en la Universidad de Columbia, período de la Segunda Guerra Mundial.

Este principio se utiliza ampliamente en diversas industrias. En medicina, por ejemplo, los ensayos clínicos de nuevos fármacos se realizan en muestras aleatorias de pacientes, lo que permite obtener resultados estadísticamente significativos sin necesidad de probar el medicamento en toda la población de personas que viven en el planeta. En economía y sociología, se realizan encuestas representativas para reflejar la opinión de la sociedad sin necesidad de entrevistar a todos los habitantes del país.

Al igual que los gobiernos y las organizaciones de investigación encuestan a pequeñas poblaciones para comprender las tendencias sociales generales, las empresas del sector de la construcción pueden utilizar muestras aleatorias de datos para realizar un seguimiento eficaz y crear previsiones para la gestión de proyectos (Fig, 9.1-1).

> *Los macrodatos pueden cambiar el enfoque de las ciencias sociales, pero no sustituirán al sentido común estadístico [147].*
>
> - Thomas Landsall-Welfair, "Predecir el estado de ánimo actual de la nación", Significance v. 9(4), 2012 г.

Desde el punto de vista del ahorro de recursos, a la hora de recopilar datos para futuras predicciones y toma de decisiones, es importante responder a la siguiente pregunta: ¿tiene sentido dedicar importantes recursos a recopilar y procesar enormes conjuntos de datos cuando puede utilizarse un conjunto de datos de prueba mucho más pequeño y barato que puede ampliarse de forma incremental? La eficacia del muestreo aleatorio demuestra que las empresas pueden reducir decenas o incluso miles de veces los costes de recopilación y formación de modelos eligiendo métodos de recopilación de datos que no requieran una cobertura exhaustiva, pero que proporcionen suficiente precisión y representatividad. Este enfoque permite incluso a las pequeñas empresas lograr resultados equiparables a los de las grandes corporaciones utilizando muchos menos recursos y volúmenes de datos, lo que es importante para las empresas que buscan optimizar costes y acelerar la toma de decisiones informadas utilizando pocos recursos. En los siguientes capítulos, exploraremos ejemplos de analytics y predictive analytics basados en conjuntos de datos públicos utilizando herramientas de big data.

Big data: análisis de los datos del millón de permisos de construcción de San Francisco

Trabajar con conjuntos de datos abiertos brinda una oportunidad única para poner en práctica los principios analizados en capítulos anteriores: selección juiciosa de características, muestreo representativo, visualización y análisis crítico. En este capítulo, exploraremos cómo fenómenos complejos como la actividad de la construcción en una gran ciudad pueden investigarse utilizando datos abiertos, en concreto, más de un millón de registros de permisos de construcción en San Francisco.

Los datos disponibles públicamente sobre más de un millón de permisos de construcción (Fig, 9.1-2) (registros en dos conjuntos de datos en formato CSV) del "Departamento de Edificios de San Francisco" [148] nos permiten utilizar la tabla CSV en bruto para analizar no solo la actividad de la construcción en la ciudad,,

sino también para analizar de forma crítica las tendencias recientes y la historia de la industria de la construcción de San Francisco en los últimos 40 años, de 1980 a 2019.

Los ejemplos de código utilizados para crear las visualizaciones del conjunto de datos (Fig,s 9.1-3- Figuras 9.1-8), así como los gráficos visuales con código, explicaciones y comentarios, pueden encontrarse en la plataforma Kaggle buscando "San Francisco. Sector de la construcción 1980-2019". [149].

count 1.137695e+06

Building Permits on or after January 1, 2013 — June 13, 2020 — 2,237 Views

Building Permits before January 1, 2013 — June 13, 2020 — 860 Views

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992

	permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0	(37.79934687604 98, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.72925851600838 6, -122.4644245667462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.79506002552974, -122.3959322446180 5)

Fig. 9.1-2 Los conjuntos de datos contienen información sobre permisos de construcción expedidos con diferentes atributos de objeto.

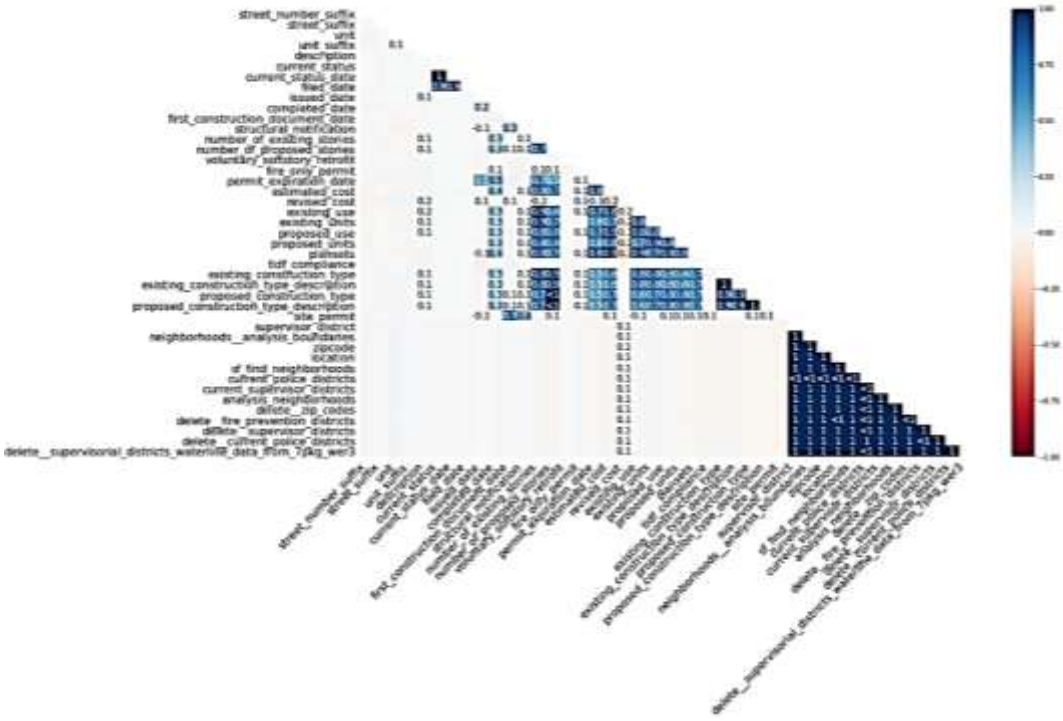


Fig. 9.1-3 Un mapa de calor (Pandas y Seaborn) que visualiza todos los atributos de un conjunto de datos y ayuda a identificar las relaciones entre pares de atributos.

De la tabla proporcionada por el Departamento de Edificios de San Francisco (Fig, 9.1-2) no se desprende ninguna tendencia ni conclusión. Las cifras secas en forma de tabla no sirven de base para la toma de decisiones. Para que los datos sean visualmente comprensibles, como se explica en detalle en los capítulos sobre visualización de datos, deben visualizarse utilizando las distintas bibliotecas que se tratan en la séptima parte del libro sobre el tema "ETL y visualización de resultados en forma de gráficos".

Analizando los datos, mediante Pandas DataFrame y las bibliotecas de visualización de Python, sobre el valor de 1.137.695 permisos [148], podemos concluir que la actividad de la construcción en San Francisco está estrechamente vinculada a los ciclos económicos, especialmente en la pujante industria tecnológica de Silicon Valley (Fig, 9.1-4).

Los auges y crisis económicos tienen un impacto significativo en el número y el valor de los proyectos de construcción. Por ejemplo, el primer pico de la actividad constructora coincidió con el auge de la electrónica a mediados de la década de 1980 (se utilizaron Pandas y Matplotlib), y los picos y descensos posteriores se asociaron a la burbuja de las puntocom y al auge tecnológico de los últimos años.

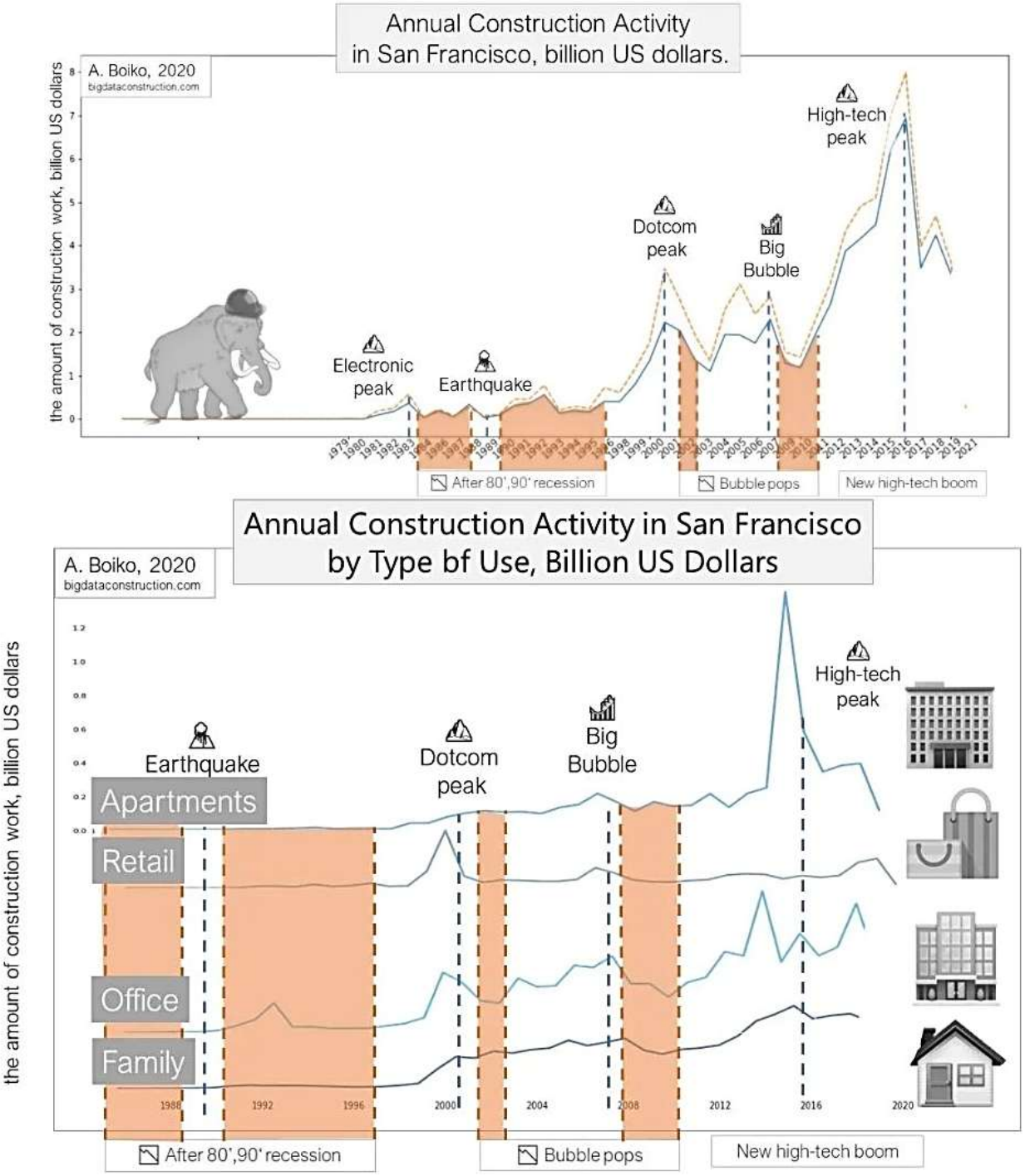


Fig. 9.1-4 En el sector inmobiliario de San Francisco, la inversión está correlacionada con el desarrollo tecnológico de Silicon Valley.

Los análisis de datos sugieren que, en San Francisco, la mayor parte de los 91.500 millones de dólares invertidos en construcción y reurbanización en la última década -casi el 75%- se concentra en el centro de la ciudad (Fig, 9.1-5 - uso de Pandas y la biblioteca de visualización Folium) y en un radio de 2 km del centro de la ciudad, lo que refleja la mayor densidad de inversión en estas zonas centrales.

El coste medio de los permisos de construcción varía considerablemente según el barrio: las solicitudes en el centro de la ciudad cuestan tres veces más que fuera de él, debido al mayor coste del suelo, la mano de obra y los materiales, y a la estricta normativa de construcción, que exige materiales más caros para mejorar la eficiencia energética.

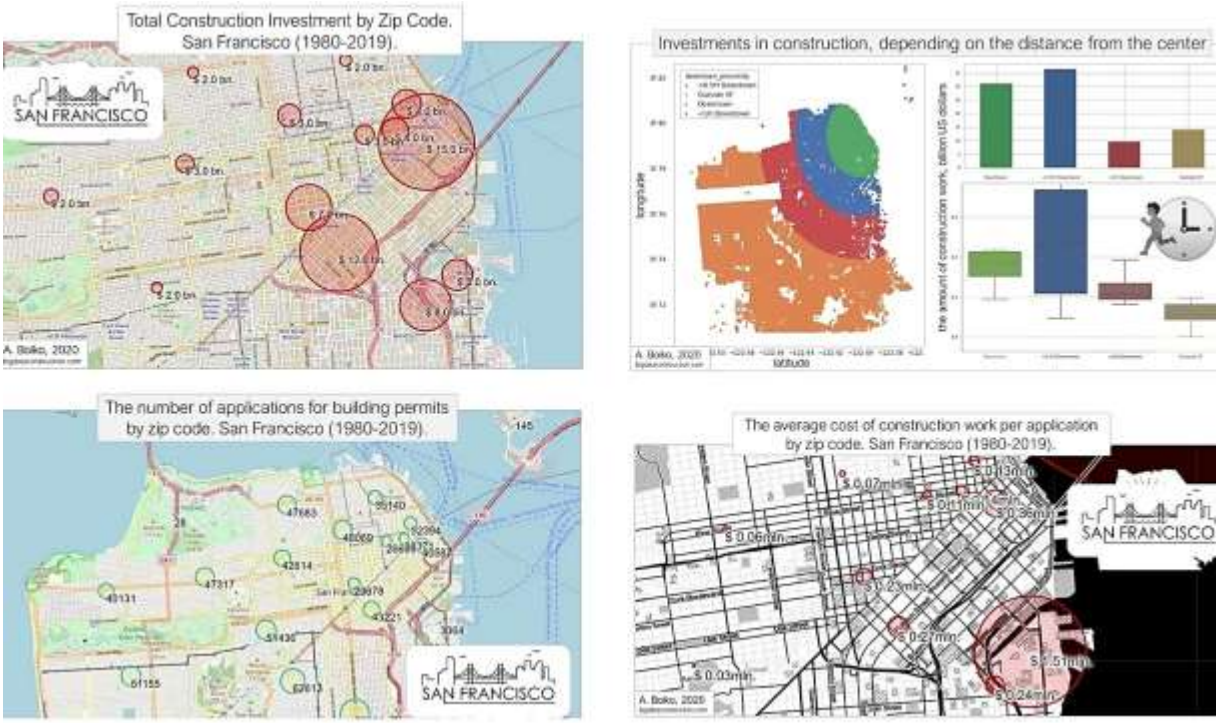


Fig. 9.1-5 En San Francisco, el 75% de la inversión en construcción (91.500 millones de dólares) se concentra en el centro de la ciudad.

El conjunto de datos también permite calcular los precios medios de reparación no sólo por tipo de vivienda, sino también por barrios de la ciudad y direcciones individuales (códigos postales). En San Francisco, la dinámica de los costes de renovación del hogar muestra tendencias distintas para los distintos tipos de renovación y vivienda (Fig, 9.1-6 - se ha utilizado Pandas y Matplotlib). Las reformas de cocinas son notablemente más caras que las de cuartos de baño: la reforma media de una cocina en una vivienda unifamiliar cuesta unos 28.000 $, frente a los 25.000 $ de una vivienda bifamiliar.

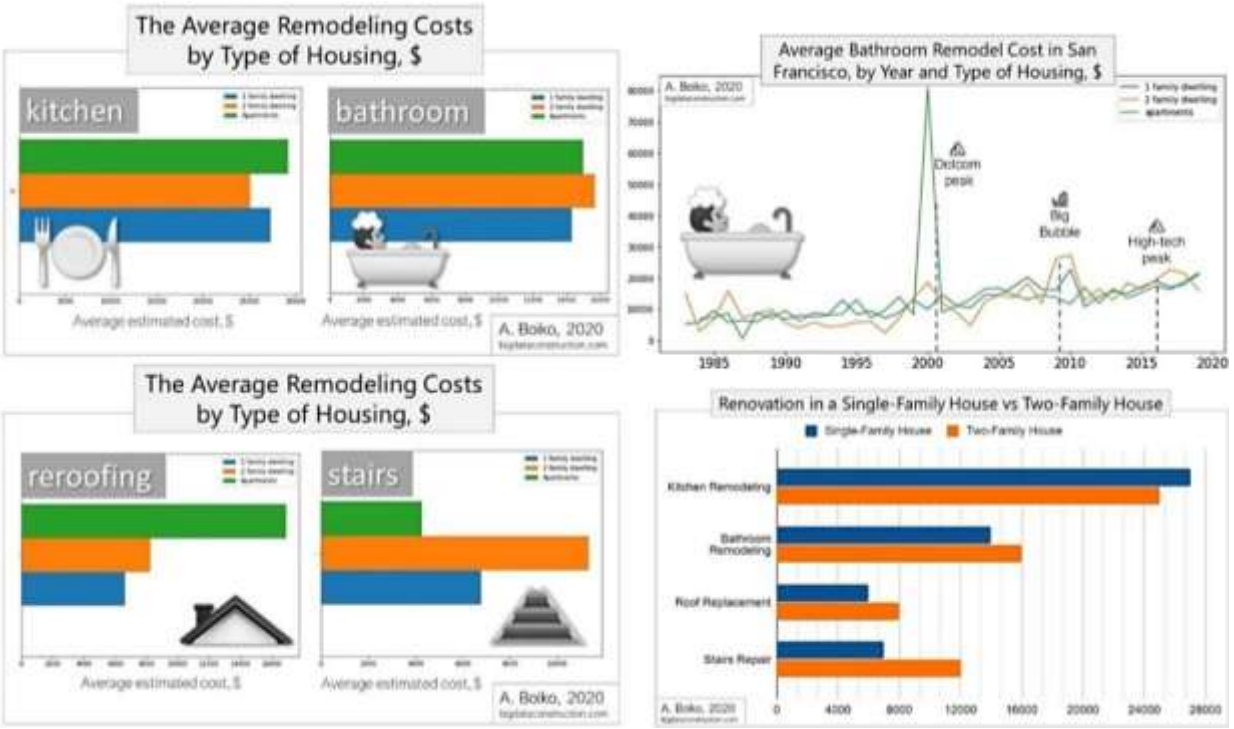


Fig. 9.1-6 En SF, las reformas de la cocina cuestan casi el doble que las del baño y los propietarios necesitan reservar$ 350 cada mes durante 15 años para cubrir el coste de las reparaciones importantes de la vivienda.

La inflación de los costes de construcción en San Francisco a lo largo de los años puede rastrearse analizando los datos agrupados por tipo de vivienda y año (Fig, 9.1-7 - utilizada por Pandas y Seaborn), que muestra un aumento constante de los costes medios de reparación desde 1990 y revela ciclos trienales a corto plazo en el coste de las reparaciones de edificios multifamiliares.

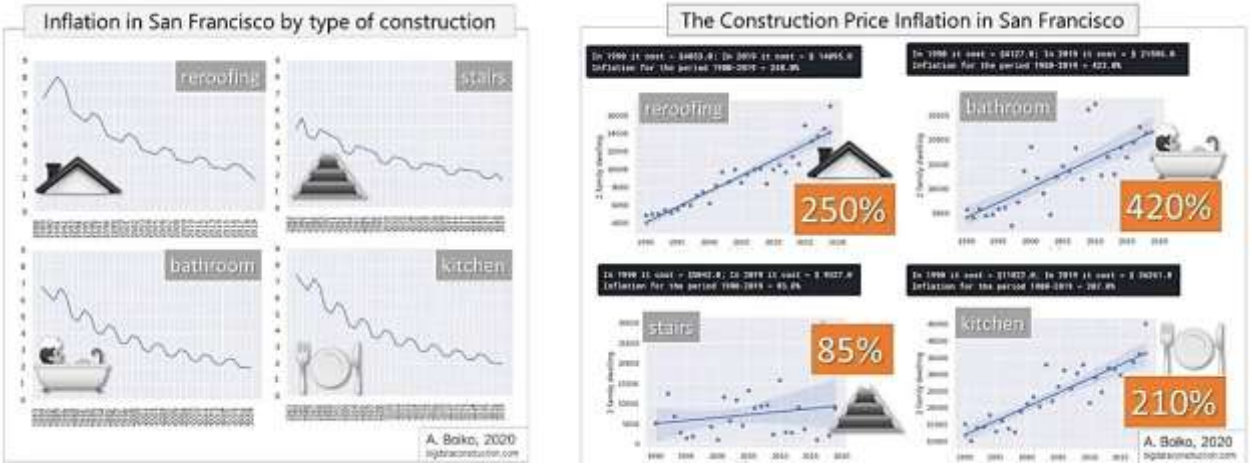


Fig. 9.1-7 De 1980 a 2019, el coste de las renovaciones de baños en SF se ha multiplicado por cinco, mientras que las renovaciones de tejados y cocinas han triplicado su precio y las renovaciones de escaleras solo han aumentado un 85%.

Un estudio de los datos públicos del Departamento de Edificación de San Francisco (Fig, 9.1-3) revela que los costes de construcción en la ciudad son extremadamente variables y a menudo impredecibles, influidos por diversos factores. Estos factores incluyen el crecimiento económico, la innovación tecnológica y los requisitos únicos de los distintos tipos de vivienda.

En el pasado, estos análisis requerían profundos conocimientos de programación y analítica. Sin embargo, con la llegada de las herramientas LLM-, el proceso se ha vuelto accesible y comprensible para un amplio abanico de profesionales del sector de la construcción, desde ingenieros de departamentos de diseño hasta altos directivos.

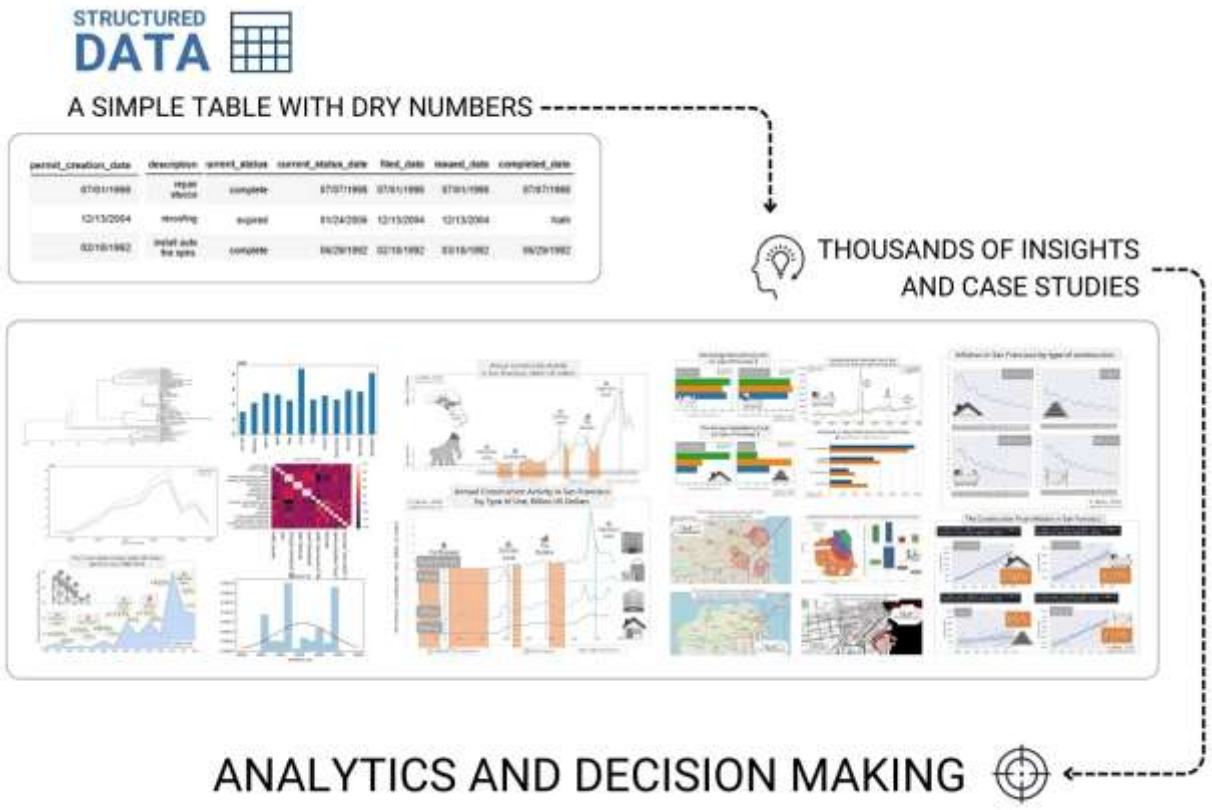


Fig. 9.1-8 El paso a datos visualmente comprensibles permite automatizar la toma de decisiones mediante el reconocimiento de patrones ocultos.

Del mismo modo que analizamos los datos del conjunto de datos tabulares "San Francisco Building Authority", podemos visualizar y analizar cualquier conjunto de datos, desde imágenes y documentos hasta datos de IoT, o datos de bases de datos CAD derivadas.

Ejemplo de big data basado en datos CAD (BIM)

En el siguiente ejemplo analizaremos un gran conjunto de datos procedentes de distintas herramientas CAD (BIM). Para recopilar y crear el gran conjunto de datos, se utilizó un rastreador web automatizado especializado (script), configurado para buscar y recopilar automáticamente archivos de diseño de sitios

web que ofrecen modelos arquitectónicos gratuitos en los formatos RVT e IFC. En pocos días, el rastreador encontró y descargó con éxito 4.596 archivos IFC, 6.471 archivos RVT y 156.024 archivos DWG[149].

Tras recopilar proyectos en formatos RVT e IFC de diferentes versiones y convertirlos a un formato CSV estructurado utilizando los SDK gratuitos de ingeniería inversa , se recopilaron casi 10.000 proyectos RVT e IFC en un gran archivo de tabla Apache Parquet y se cargaron en Pandas DataFrame para su análisis (Fig, 9.1-9).

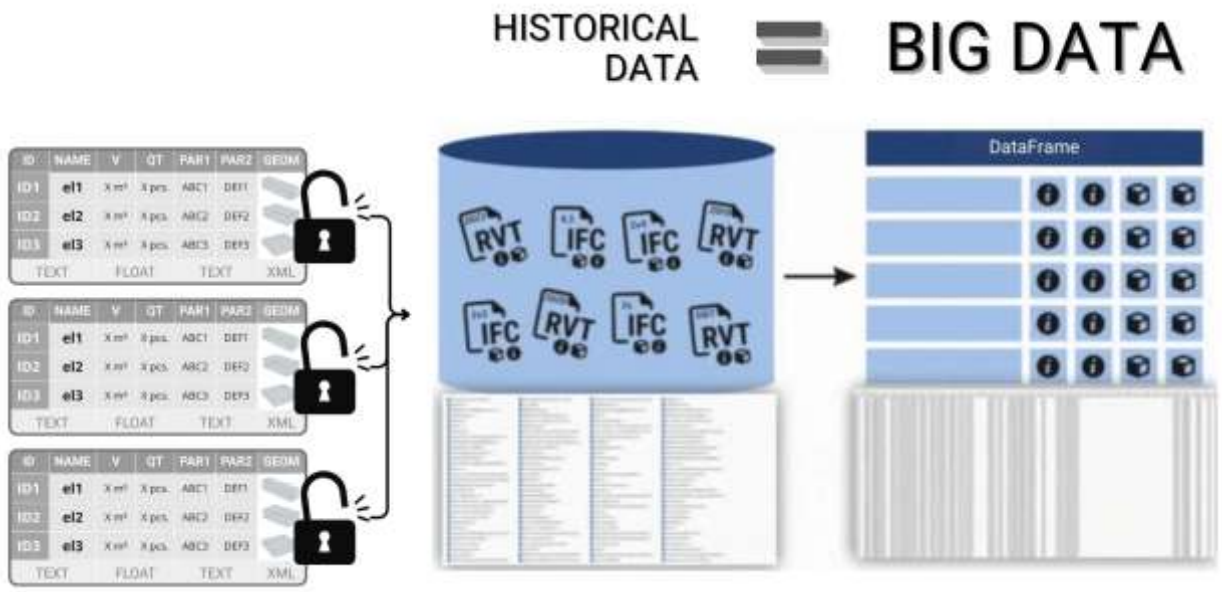


Fig. 9.1-9 Datos estructurados datos de proyecto le permite combinar cualquier número de proyectos en una única tabla bidimensional.

Los datos de esta colección a gran escala contienen la siguiente información: el conjunto de ficheros IFC contiene unos 4 millones de entidades (filas) y 24.962 atributos (columnas), y el conjunto de ficheros RVT, formado por unos 6 millones de entidades (filas), contiene 27.025 atributos diferentes (columnas).

Estos conjuntos de información (Fig, 9.1-10) abarcan millones de elementos, para cada uno de los cuales se obtuvieron además las coordenadas de la geometría Bounding Box (rectángulo que define los límites de un objeto en el proyecto) y se añadieron a una tabla común - se crearon las coordenadas de la geometría Bounding Box (rectángulo que define los límites de un objeto en el proyecto) e imágenes de cada elemento en formato PNG y la geometría en el formato XML abierto - DAE (Collada).

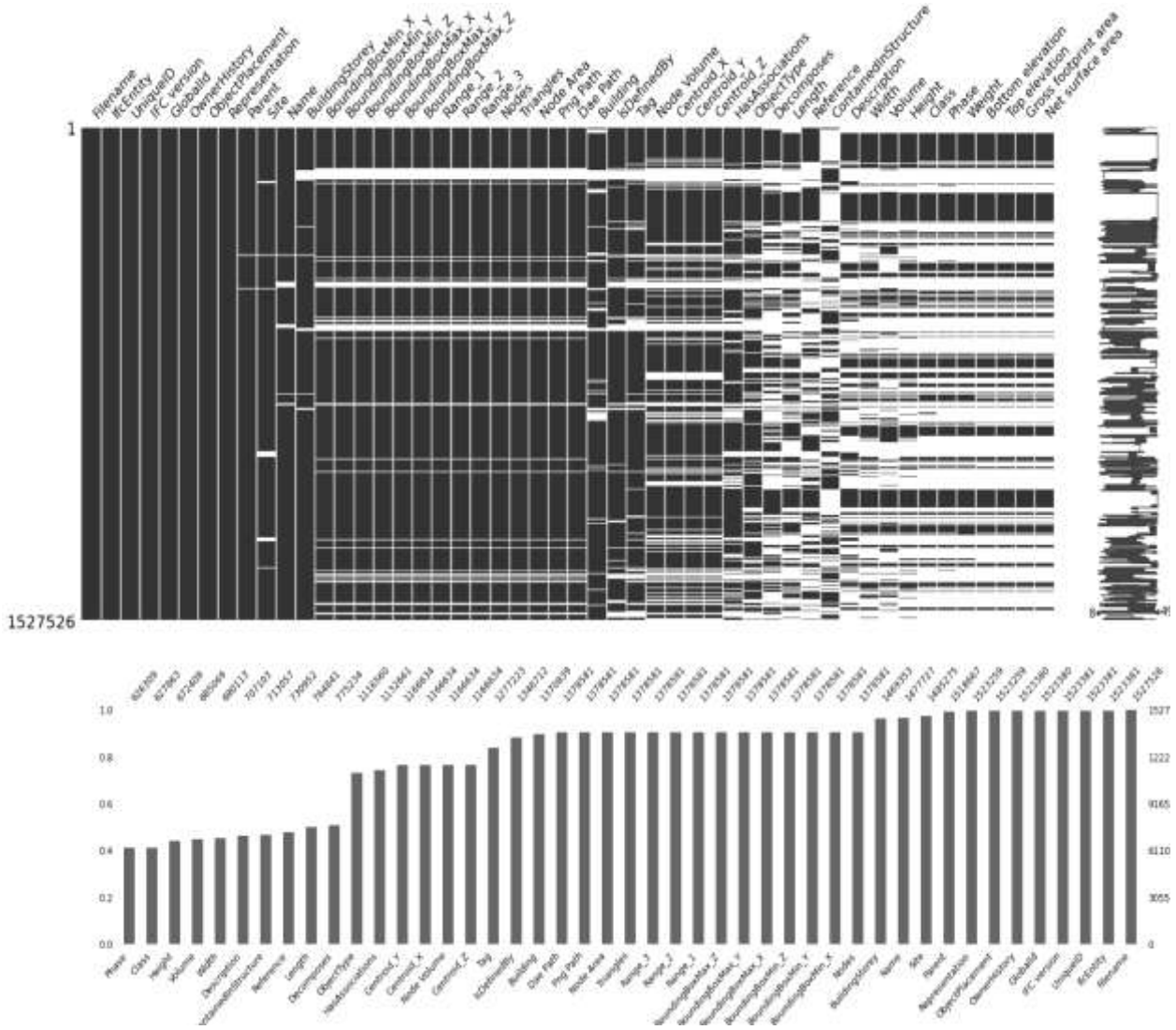


Fig. 9.1-10 Sabset de 1,5 millones de elementos y visualización (biblioteca missingno) de la ocupación de los 100 primeros atributos en forma de histograma.

Así, obtuvimos toda la información sobre decenas de millones de elementos de 4.596 proyectos IFC y 6.471 proyectos RVT, en los que todos los atributos-propiedades de todos los elementos de la entidad y su geometría (Bounding Box) se tradujeron a la forma estructurada de una única tabla (DataFrame) (Fig. 9.1-10 - los datos de las poblaciones del dataframe aparecen como histogramas).

> Los histogramas (Fig. 9.1-10, Fig. 9.2-6, Fig. 9.2-7) trazados durante el proceso de análisis permiten evaluar rápidamente la densidad de datos y la frecuencia de aparición de valores en las columnas. De este modo se obtiene una primera idea de la distribución de las características, la presencia de valores atípicos y la utilidad potencial de los atributos individuales en el análisis y la creación de modelos de aprendizaje automático.

Un ejemplo de uso práctico de este conjunto de datos (Fig. 9.1-10) es el proyecto "5000 IFC y RVT ". [149],

disponible en la plataforma Kaggle. Presenta el cuaderno Jupyter Notebook con una solución Pipeline completa: desde el preprocesamiento y análisis de los datos hasta la visualización de los resultados utilizando librerías Python - pandas, matplotlib, seaborn, folium y otras (Fig. 9.1-11).

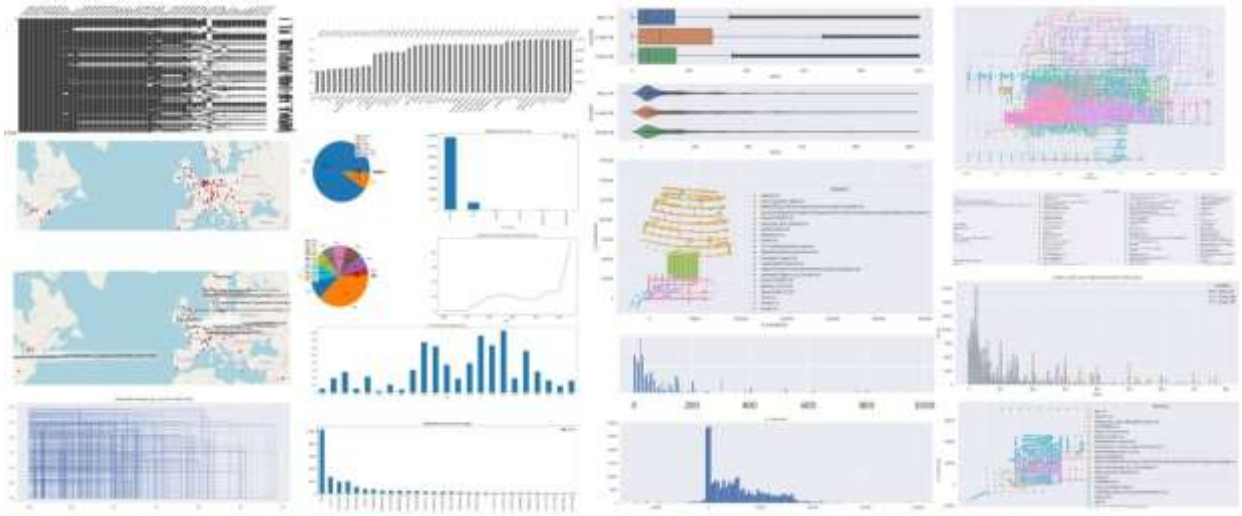


Fig. 9.1-11 Ejemplos de análisis de datos procedentes de formatos CAD (BIM) utilizando las bibliotecas de visualización de Python y la biblioteca pandas.

A partir de la metainformación, es posible determinar en qué ciudades se han desarrollado determinados proyectos y visualizarlo en un mapa (por ejemplo, utilizando la biblioteca folium). Además, las marcas de tiempo en los datos permiten explorar patrones sobre cuándo se guardaron o editaron los archivos: por día de la semana, hora del día y mes.

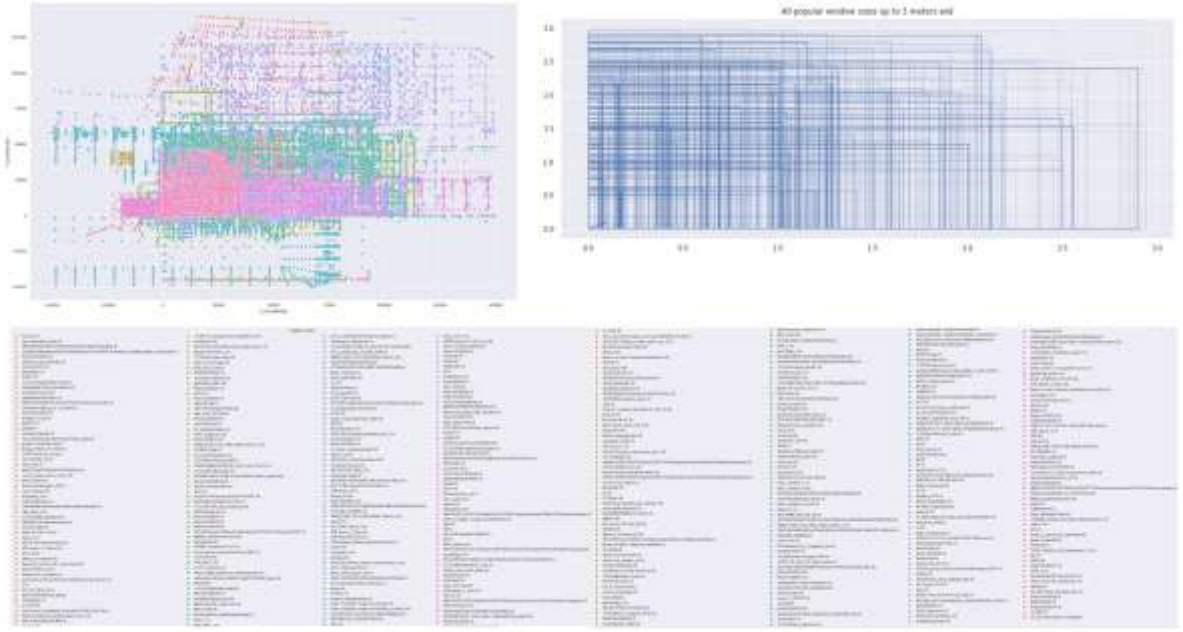


Fig. 9.1-12 Visualización de la posición geométrica de todas las columnas y las dimensiones de todas las ventanas de hasta 3 metros en proyectos de la lista situada en la parte inferior del gráfico.

Los parámetros geométricos en forma de Bounding Box extraídos de los modelos también se prestan a análisis agregados. Por ejemplo, la Fig. 9.1-12 muestra dos gráficos: el de la izquierda muestra la distribución de distancias entre columnas para todos los proyectos en relación con el punto cero, y el de la derecha muestra las dimensiones de todas las ventanas de hasta 3 metros de altura en una muestra de decenas de miles de elementos de ventana (tras agrupar todo el conjunto de datos por el parámetro "Categoría" con el valor "OST_Windows", "IfcWindows").

El código de análisis Pipeline para este ejemplo y el propio conjunto de datos están disponibles en el sitio web de Kaggle bajo el título "5000 IFC and RVT | DataDrivenCo-nstruction.io projects" [149]. Este Pipeline terminado junto con el conjunto de datos puede copiarse y ejecutarse gratuitamente en línea en Kaggle o sin conexión en uno de los IDE populares: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con el plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con el plugin Python, JupyterLab o las populares herramientas en línea Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Los conocimientos analíticos obtenidos del procesamiento y estudio de enormes cantidades de datos estructurados desempeñarán un papel crucial en los procesos de toma de decisiones del sector de la construcción.

Con este tipo de análisis de la información basado en proyectos anteriores, los especialistas pueden prever eficazmente, por ejemplo, las necesidades de material y mano de obra, y optimizar las soluciones de diseño antes de que empiece la construcción

Sin embargo, mientras que los datos de diseño o los permisos de construcción son información relativamente estática que cambia con relativa lentitud, el propio proceso de construcción se está saturando rápidamente de una gran variedad de sensores y dispositivos IoT: cámaras, sistemas de supervisión automatizados que transmiten datos en tiempo real... todo ello convierte la obra en un entorno digital dinámico en el que los datos deben analizarse en tiempo real.

IoT Internet de las cosas y contratos inteligentes

IoT El Internet de los objetos representa una nueva ola de transformación digital en la que cada dispositivo obtiene su propia dirección IP y pasa a formar parte de una red global. IoT es un concepto que consiste en conectar objetos físicos a Internet para recopilar, procesar y transmitir datos. En la construcción, esto significa la capacidad de controlar los procesos de construcción en tiempo real, minimizar el desperdicio de materiales, predecir el desgaste de los equipos y automatizar la toma de decisiones.

> Según el artículo de la CFMA "Prepararse para el futuro con la construcción conectada" [150], el sector de la construcción experimentará una gran transformación digital en la próxima década, que culminará en el concepto de Construcción Conectada: una obra totalmente integrada y automatizada.

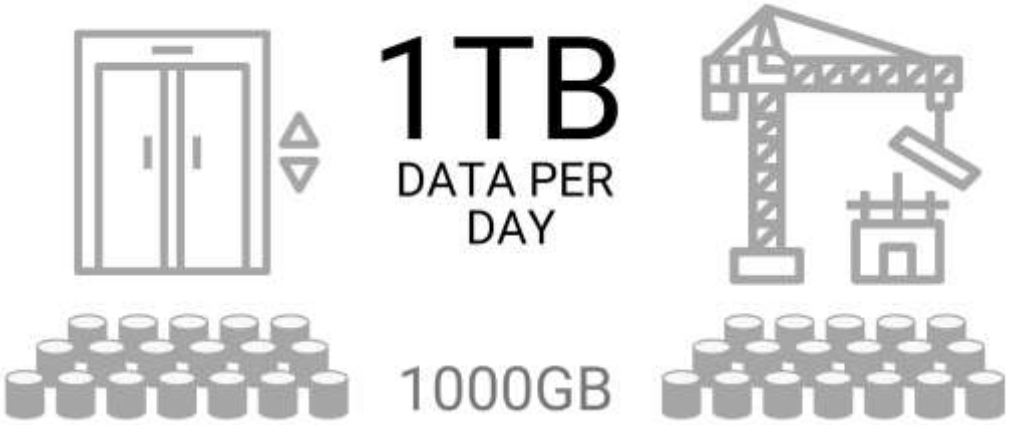


Fig. 9.1-13 Los dispositivos de datos de la IoT o de las obras de construcción pueden producir y transmitir terabytes de datos al día.

Una obra digital implica que todos los elementos de la construcción -desde la planificación y la logística hasta la ejecución de los trabajos y el control de calidad en la obra mediante cámaras fijas y cuadricópteros- se integrarán en un único ecosistema digital dinámico. Anteriormente, en la Parte 7 de este libro, ya hemos examinado las capacidades de Apache NiFi (Fig, 7.4-5), una herramienta gratuita y de código abierto que permite la transmisión de datos en tiempo real, desde su recogida de diversas fuentes hasta su transferencia a plataformas de almacenamiento o análisis.

Los datos sobre el avance de la construcción, el consumo de materiales, el estado de los equipos y la seguridad se transmitirán en tiempo real a sistemas analíticos (Fig. 9.1-13). Esto permite predecir posibles riesgos, responder con prontitud a las desviaciones y optimizar los procesos de la obra. Los componentes clave de una obra digital incluyen:

- IoT -sensores - seguimiento de parámetros medioambientales, supervisión de maquinaria de construcción y control de las condiciones laborales.
- Gemelos digitales: modelos virtuales de edificios e infraestructuras para predecir posibles desviaciones y evitar errores.
- Sistemas logísticos automatizados: gestión de la cadena de suministro en tiempo real para reducir los tiempos de inactividad y los costes.
- Sistemas de construcción robotizados: uso de máquinas autónomas para realizar tareas rutinarias y peligrosas.

Robotización, el uso de IoT y el concepto de obra digital Connected Site (Construction) no solo aumentarán la eficiencia y reducirán los costes, sino que también marcarán el comienzo de una nueva era de seguridad, construcción sostenible y gestión predictiva de proyectos.

> Las etiquetas RFID (identificación por radiofrecuencia) son también uno de los componentes más importantes del IoT. Se utilizan para identificar y rastrear materiales, maquinaria e incluso personal en una obra, aumentando la transparencia y el control de los recursos del proyecto.

RFID -technology is used to automatically recognise objects using radio signals. Consta de tres elementos clave:

- RFID -etiquetas (pasivas o activas)- contienen un identificador único y se fijan a materiales, herramientas o maquinaria.
- Los escáneres son dispositivos que leen la información de las etiquetas y la transmiten al sistema.
- Base de datos centralizada: almacena información sobre la ubicación, el estado y el movimiento de los objetos.

Aplicación de RFID en la construcción:

- Contabilidad automática de materiales: las etiquetas de los productos de hormigón premezclado, las armaduras o los paquetes de paneles sándwich permiten controlar las existencias y evitar robos.
- Control del trabajo del personal - RFID - las tarjetas identificativas de los empleados registran las horas de inicio y fin de turno, proporcionando un registro de las horas de trabajo.
- Monitorización de equipos - RFID - el sistema rastrea el movimiento de los equipos, evitando tiempos de inactividad y mejorando la eficiencia logística.

Complementan este conjunto de tecnologías los contratos inteligentes basados en blockchain que automatizan los pagos, el control de las entregas y el cumplimiento de los contratos sin necesidad de intermediarios, reduciendo el riesgo de fraude y los retrasos.

> Hoy en día, a falta de un modelo de datos común, los contratos inteligentes son simplemente código que los participantes acuerdan. Sin embargo, con un enfoque centrado en los datos, es posible crear un modelo común de parámetros contractuales, codificarlo en una blockchain y automatizar el cumplimiento de los términos.

Por ejemplo, en un sistema de gestión de la cadena de suministro, un contrato inteligente puede rastrear la entrega de un envío a partir de IoT -sensores y RFID -etiquetas y transferir automáticamente el pago cuando llegue. Del mismo modo, en una obra de construcción, un contrato inteligente podría registrar la finalización de una fase de trabajo -como la instalación de barras de refuerzo o el vertido de cimientos- a partir de datos de drones o sensores de construcción e iniciar automáticamente el siguiente pago al contratista sin necesidad de comprobaciones manuales y certificados en papel.

Pero a pesar de las nuevas tecnologías y de los esfuerzos de las organizaciones internacionales de normalización, una plétora de normas en competencia complican el panorama de la IO.

> Según un estudio de Cisco publicado en 2017 [151], casi el 60% de las iniciativas de Internet de las Cosas (IoT) se detienen en la fase de prueba de concepto, y solo el 26% de las empresas consideran que sus proyectos de IoT han tenido pleno éxito. Además, un tercio de los proyectos finalizados no alcanzan sus objetivos declarados y no se reconocen como exitosos ni siquiera después de su implementación.

Una de las principales razones es la falta de interoperabilidad entre plataformas que procesan datos de

distintos sensores. Como resultado, los datos permanecen aislados dentro de soluciones separadas. Una alternativa a este planteamiento, como en otros casos similares (que hemos tratado en este libro), es una arquitectura construida en torno a los propios datos como activo principal.

> Los sensores IoT desempeñan un papel clave no solo en la supervisión del estado técnico de los equipos, sino también en el análisis predictivo para reducir los riesgos en las obras y mejorar el rendimiento general de los procesos mediante la predicción de fallos y desviaciones.

Los datos recogidos por los sensores IoT y las etiquetas RFID pueden procesarse en tiempo real mediante algoritmos de aprendizaje automático capaces de detectar anomalías y alertar a los ingenieros de posibles averías con antelación. Esto puede abarcar desde microfisuras en estructuras de hormigón hasta pausas inusuales en el funcionamiento de grúas torre, lo que indica fallos técnicos o infracciones de la normativa. Además, los algoritmos avanzados de análisis del comportamiento pueden captar patrones de comportamiento que pueden indicar, por ejemplo, fatiga física del personal, mejorando la gestión proactiva de la seguridad y el bienestar de los empleados en las obras.

En el sector de la construcción, los accidentes y las averías -ya sean de maquinaria o de personas- rara vez ocurren de repente. Suelen ir precedidos de pequeñas desviaciones que pasan desapercibidas. El análisis predictivo y el aprendizaje automático permiten detectar estas señales en una fase temprana, incluso antes de que se produzcan consecuencias críticas.

Mientras que los documentos, los archivos del proyecto y los datos de los dispositivos IoT y las etiquetas RFID forman la huella digital de los proyectos de construcción, el aprendizaje automático puede ayudar a extraer información útil de ellos. Con el crecimiento de los datos y la democratización del acceso a los mismos, el sector de la construcción está obteniendo nuevas oportunidades en las aplicaciones de análisis, análisis predictivo e inteligencia artificial.

CAPÍTULO 9.2.
APRENDIZAJE AUTOMÁTICO Y PREDICCIONES

Aprendizaje automático e inteligencia artificial cambiarán nuestra forma de construir

Las bases de datos de los distintos sistemas del sector de la construcción -con su infraestructura inevitablemente decadente y cada vez más compleja- se están convirtiendo en un caldo de cultivo para futuras soluciones. Los servidores de las empresas, como un bosque, son ricos en biomasa de información importante, a menudo oculta bajo tierra, en las entrañas de carpetas y servidores. Las masas de datos de los distintos sistemas que se crean hoy -tras su uso, tras caer al fondo del servidor y tras años de fosilización- alimentarán en el futuro el aprendizaje automático y los modelos lingüísticos. Las salas de chat internas de las empresas (por ejemplo, una instancia independiente de ChatGPT configurada localmente, LlaMa, Mistral, DeepSeek) se construirán sobre estos modelos internos utilizando el almacenamiento centralizado para recuperar información de forma rápida y cómoda y generar los gráficos, cuadros de mando y documentos necesarios.

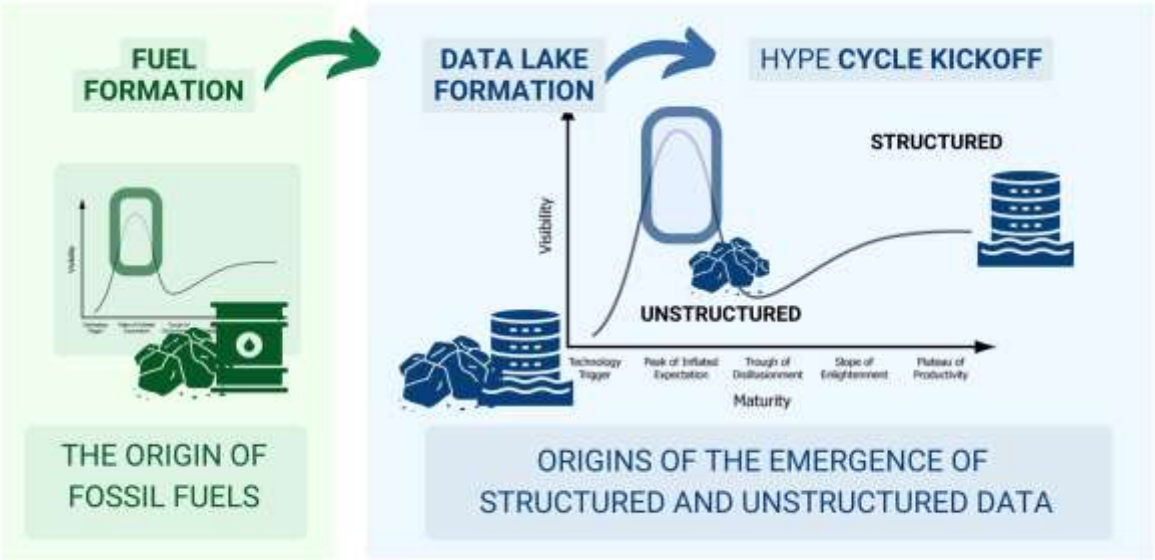


Fig. 9.2-1 Al igual que los árboles se convierten en carbón, la información también se transforma con el tiempo en valiosa energía empresarial bajo la presión del tiempo y la analítica.

La fosilización de la masa vegetal en combinación con la presión y la temperatura crea una masa homogénea y estructurada de forma única de árboles de diferentes especies que vivieron en diferentes épocas: el carbón vegetal [152]. Del mismo modo, la información grabada en discos duros en diferentes formatos y en diferentes momentos bajo la presión de los departamentos de análisis y la temperatura de la gestión de calidad acaba formando una masa estructurada homogénea de información valiosa (Fig. 9.2-1).

Estas capas (o, más a menudo, pepitas aisladas) de información se crean mediante una minuciosa organización de los datos por parte de analistas experimentados que empiezan a extraer gradualmente información valiosa de datos aparentemente irrelevantes durante mucho tiempo.

En el momento en que estas capas de datos maduros dejan de estar "quemadas" en los informes y empiezan a circular en los procesos empresariales, enriqueciendo las decisiones y mejorando los procesos, la empresa está preparada para el siguiente paso: la transición al aprendizaje automático y la inteligencia artificial (Fig. 9.2-2).

> El aprendizaje automático (ML - Machine learning) es una clase de métodos para resolver problemas de inteligencia artificial. Los algoritmos de aprendizaje automático reconocen patrones en grandes conjuntos de datos y los utilizan para aprender por sí mismos. Cada nuevo conjunto de datos permite a los algoritmos matemáticos mejorar y adaptarse en función de la información obtenida, lo que permite mejorar constantemente la precisión de las recomendaciones y predicciones.

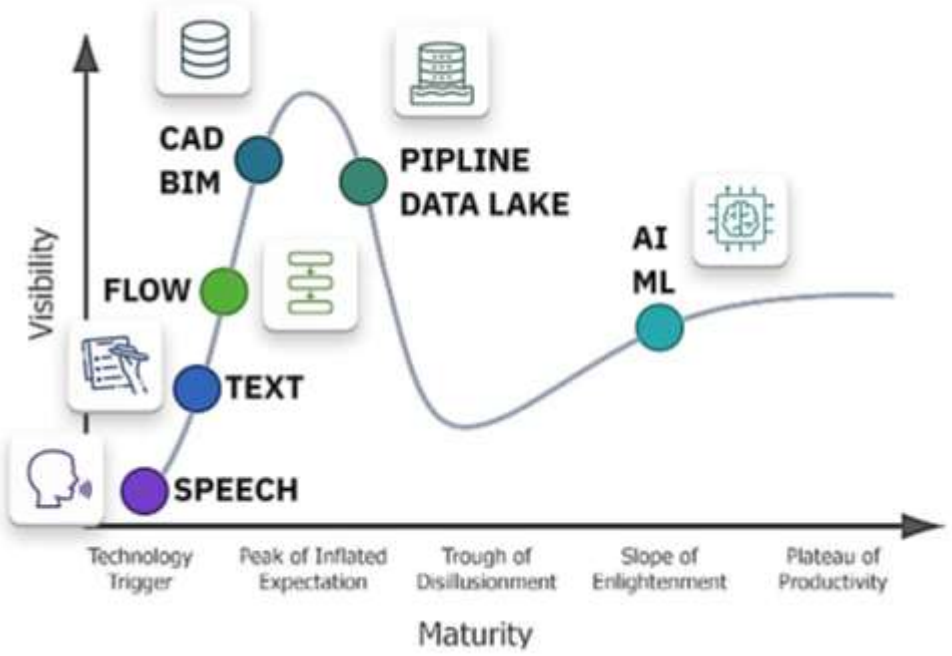


Fig. 9.2-2 El desvanecimiento de las tecnologías de creación de datos y la aplicación de herramientas de análisis abre la puerta al tema del aprendizaje automático.

Como dijo el influyente director general del mayor fondo de inversión del mundo (que posee participaciones clave en casi todas las mayores empresas de software para la construcción, así como en las empresas propietarias de la mayor cantidad de inmuebles del mundo [55]) en una entrevista de 2023 - el aprendizaje automático cambiará el mundo de la construcción.

> *AI tiene un enorme potencial. Cambiará nuestra forma de trabajar y de vivir. La IA y la robótica cambiarán nuestra forma de trabajar y de construir, y podremos utilizar la IA y la robótica como medios para crear una productividad mucho mayor [153].*
>
> - CEO del mayor fondo de inversión del mundo, entrevista, septiembre de 2023.

Aprendizaje automático (ML) funciona procesando grandes cantidades de datos, utilizando técnicas estadísticas para imitar aspectos del pensamiento humano. Sin embargo, la mayoría de las empresas no disponen de tales conjuntos de datos, y si los tienen, a menudo no están suficientemente etiquetados. Aquí es donde pueden ayudar las tecnologías semánticas y el aprendizaje por transferencia, una técnica que permite al ML ser más eficaz cuando se trata de pequeñas cantidades de datos, cuya viabilidad se ha tratado en capítulos anteriores de esta parte.

La esencia del aprendizaje por transferencia es que, en lugar de procesar cada tarea desde cero, se pueden utilizar los conocimientos adquiridos en campos relacionados. Hay que darse cuenta de que los patrones y descubrimientos de otras industrias pueden adaptarse y aplicarse en el sector de la construcción. Por ejemplo, los métodos de optimización de los procesos logísticos desarrollados en el comercio minorista ayudan a mejorar la eficiencia de la gestión de la cadena de suministro de la construcción. El análisis de macrodatos, que se utiliza activamente en las finanzas, puede aplicarse a la previsión de costes y la gestión de riesgos en los proyectos de construcción. Y las tecnologías de visión por ordenador y robótica que se están desarrollando en la industria ya están encontrando aplicaciones en el control de calidad automatizado, la supervisión de la seguridad y la gestión de las instalaciones de las obras de construcción.

> El aprendizaje por transferencia permite no sólo acelerar la introducción de innovaciones, sino también reducir el coste de su desarrollo, utilizando la experiencia ya acumulada de otras industrias.

Fig. 9.2-3 Inteligencia Artificial tecnologías y la robótica serán el principal motor del futuro para aumentar la productividad en el sector de la construcción.

El pensamiento humano se organiza según un principio similar: nos basamos en conocimientos adquiridos previamente para resolver nuevos problemas (Fig. 4.4-19, Fig. 4.4-20, Fig. 4.4-21). En el aprendizaje automático, este enfoque también funciona: al simplificar el modelo de datos y hacerlo más elegante, podemos reducir la complejidad del problema para los algoritmos de ML. Esto, a su vez, reduce la necesidad de grandes cantidades de datos y reduce el coste computacional.

De la evaluación subjetiva a la previsión estadística

La época en que las decisiones estratégicas dependían de la intuición de cada directivo (Fig, 9.2-4) es cosa del pasado. En un entorno económico cada vez más competitivo y desafiante, un enfoque subjetivo resulta demasiado arriesgado e ineficaz. Las empresas que siguen confiando en las opiniones personales en lugar de analizar objetivamente los datos de, pierden la capacidad de responder rápidamente a los cambios.

> El entorno competitivo exige precisión y repetibilidad basadas en datos, patrones estadísticos y probabilidad computable. Las decisiones ya no pueden basarse en sensaciones, sino en correlaciones, tendencias y modelos predictivos derivados de la analítica y el aprendizaje automático. No se trata solo de un cambio de herramientas, sino de un cambio en la lógica del pensamiento: de las suposiciones a las pruebas, de las probabilidades subjetivas a las desviaciones calculadas estadísticamente, de las sensaciones a los hechos.

Fig. 9.2-4 La era de las decisiones tomadas por HiPPO (la opinión del empleado mejor pagado) será cosa del pasado con la llegada de los macrodatos y el aprendizaje automático.

Los directivos que solían confiar únicamente en sus propios sentimientos se enfrentarán inevitablemente a una nueva realidad: la autoridad ya no determina las decisiones. En el centro de la gestión se encuentran ahora sistemas que analizan millones de parámetros y vectores, identifican patrones ocultos y sugieren estrategias óptimas.

La principal razón por la que las empresas actuales siguen evitando aplicar el ML es su falta de transparencia. La mayoría de los modelos funcionan como "cajas negras" para los directivos, sin explicar cómo

exactamente llegan a sus conclusiones. Esto acarrea problemas: los algoritmos pueden reforzar estereotipos e incluso crear situaciones cómicas, como en el caso del chatbot de Microsoft, que rápidamente se convirtió en una herramienta de comunicación tóxica [154].

> En Deep Thinking, Garry Kasparov, antiguo campeón del mundo de ajedrez, reflexiona sobre su derrota ante el ordenador Big Blue de IBM [155]. Sostiene que el verdadero valor de la IA no está en copiar la inteligencia humana, sino en complementar nuestras capacidades. La IA debe realizar tareas en las que los humanos son débiles, mientras que los humanos aportan creatividad. Los ordenadores han cambiado el enfoque tradicional del análisis del ajedrez. En lugar de crear historias fascinantes sobre las partidas, los programas informáticos de ajedrez evalúan cada jugada de forma imparcial, basándose únicamente en su fuerza o debilidad real. Kasparov señala que la tendencia humana a ver los acontecimientos como historias coherentes y no como acciones individuales suele llevar a conclusiones erróneas, no sólo en el ajedrez, sino en la vida en general.

Por lo tanto, si tiene previsto utilizar el aprendizaje automático para la predicción y el análisis, es importante que comprenda sus principios básicos -cómo funcionan los algoritmos y cómo se procesan los datos- antes de empezar a utilizar las herramientas de aprendizaje automático y AI en su trabajo. La mejor manera de empezar es a través de la experiencia práctica.

Una de las herramientas más convenientes para una introducción inicial al tema del aprendizaje automático y la predicción es el Jupyter Notebook y el popular conjunto de datos clásico Titanic, que proporcionará una introducción visual a los métodos clave de análisis de datos y construcción de modelos de ML.

Titanic dataset: Hello World en el mundo de la analítica data y big data

Uno de los ejemplos más famosos de uso del ML en el análisis de datos es el análisis de la tabla de datos del Titanic, que suele utilizarse para estudiar la probabilidad de supervivencia de los pasajeros. El aprendizaje de esta tabla es análogo al programa "Hello World" cuando se aprenden lenguajes de programación.

El hundimiento del RMS Titanic en 1912 causó la muerte de 1.502 de las 2.224 personas que lo integraban. El conjunto de datos del Titanic no sólo contiene información sobre si un pasajero sobrevivió, sino también atributos como: edad, sexo, clase de billete y otros parámetros. Este conjunto de datos está disponible gratuitamente y puede abrirse y analizarse en diversas plataformas en línea y fuera de línea.

```
Enlace al conjunto de datos sobre el Titanic:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Anteriormente, en el capítulo "IDEs habilitados para LLM y futuros cambios en la programación" ya hablamos de Jupyter Notebook - uno de los entornos de desarrollo más populares para el análisis de datos y el aprendizaje automático. Los análogos gratuitos en la nube de Jupyter Notebook son las plataformas Kaggle y Google Collab, que permiten ejecutar código Python sin instalar software y proporcionan acceso gratuito a recursos informáticos.

> Kaggle es la mayor plataforma de competición de análisis de datos y aprendizaje automático con un entorno de ejecución de código integrado. En octubre de 2023, Kaggle contaba con más de 15 millones de usuarios [156] de 194 países.

Descargue y utilice el conjunto de datos Titanic en la plataforma Kaggle (Fig, 9.2-5) para almacenar el conjunto de datos (una copia del mismo) y ejecutar el código Python con las bibliotecas preinstaladas directamente en un navegador, sin tener que instalar un IDE dedicado.

Fig. 9.2-5 Estadísticas de la tabla Titanic: el conjunto de datos de entrenamiento más popular para el aprendizaje de análisis de datos y aprendizaje automático.

El conjunto de datos Titanic incluye datos sobre los 2.224 pasajeros a bordo del *RMS Titanic* en el momento de su naufragio en 1912. El conjunto de datos se presenta en dos tablas separadas, una muestra de entrenamiento (train.csv) y otra de prueba (test.csv), lo que permite utilizarlo tanto para entrenar modelos como para evaluar su precisión con nuevos datos.

El conjunto de datos de entrenamiento contiene tanto atributos de los pasajeros (edad, sexo, clase de billete y otros) como información sobre quién ha sobrevivido (columna con valores binarios "Sobrevivido"). El conjunto de datos de entrenamiento (Fig, 9.2-6 - archivo train.csv) se utiliza para entrenar el modelo. El conjunto de datos de prueba (Fig, 9.2-7 - archivo test.csv) incluye sólo atributos de pasajeros sin información sobre supervivientes (sin una sola columna "Survivor"). El conjunto de datos de prueba está diseñado para probar el modelo con nuevos datos y evaluar su precisión.

Así, tenemos atributos de pasajeros casi idénticos en los conjuntos de datos de entrenamiento y de prueba. La única diferencia clave es que en el conjunto de datos de prueba tenemos una lista de pasajeros que no tienen la columna "Superviviente", la variable objetivo, que queremos aprender a predecir utilizando varios algoritmos matemáticos. Y después de construir el modelo, podremos comparar la salida de nuestro modelo con el parámetro real "Superviviente" del conjunto de datos de prueba, que tendremos en cuenta para

evaluar los resultados.

Las columnas principales de la tabla, parámetros de los pasajeros en el conjunto de datos de entrenamiento y de prueba:

- **PassengerId** - identificador único del pasajero
- **Sobrevivió** - 1 si el pasajero sobrevivió, 0 si murió (no disponible en el conjunto de pruebas)
- **Pclass** - clase de billete (1, 2 o 3)
- Nombre - nombre del pasajero
- **Sexo** - sexo del pasajero (hombre/mujer)
- Edad
- **SibSp** - número de hermanos/hermanas o cónyuges a bordo
- **Parch** - número de padres o hijos a bordo
- **Billete** - número de billete
- **Tarifa** - precio del billete
- Cabina - número de cabina (faltan muchos datos)
- **Embarcado** es el puerto de embarque (C = Cherbourg, Q = Queenstown, S = Southampton).

Para visualizar los datos que faltan en ambas tablas, puede utilizar la biblioteca missingno (Fig. 9.2-6, Fig. 9.2-7), que muestra los valores que faltan en forma de histograma, donde los campos blancos muestran los datos que faltan. Esta visualización permite evaluar rápidamente la calidad de los datos antes de procesarlos.

titanic_missingno.py

```
1  import pandas as pd
2  import missingno as msno
3  import matplotlib.pyplot as plt
4
5  # Load the dataset
6  train = pd.read_csv('train.csv')
7
8  # Visualize missing values
9  msno.matrix(train)
10 plt.show()
```

Run in IDE

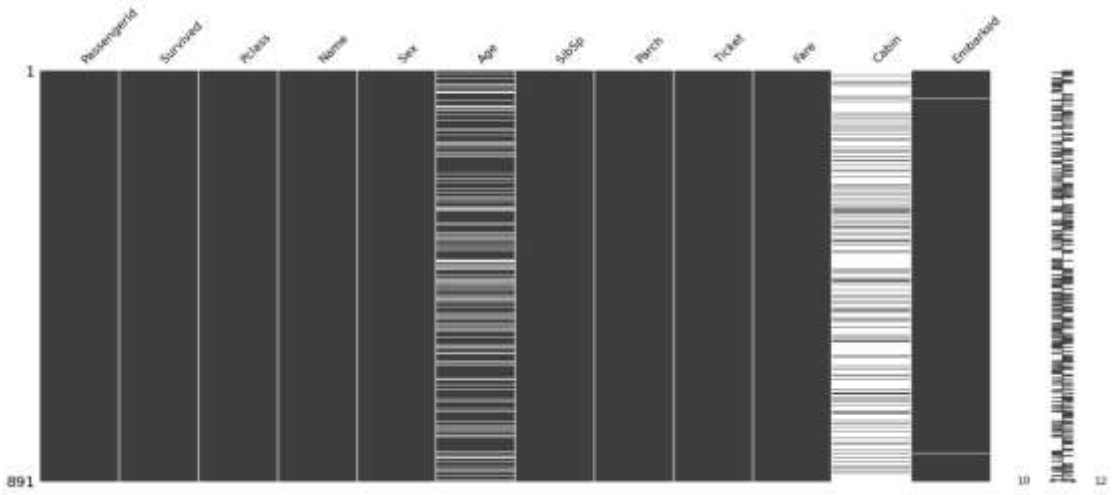


Fig. 9.2-6 Se utilizan unas pocas líneas de código para visualizar los datos que faltan en el conjunto de datos de entrenamiento Titanic, donde el parámetro clave para el entrenamiento es el parámetro "Survived".

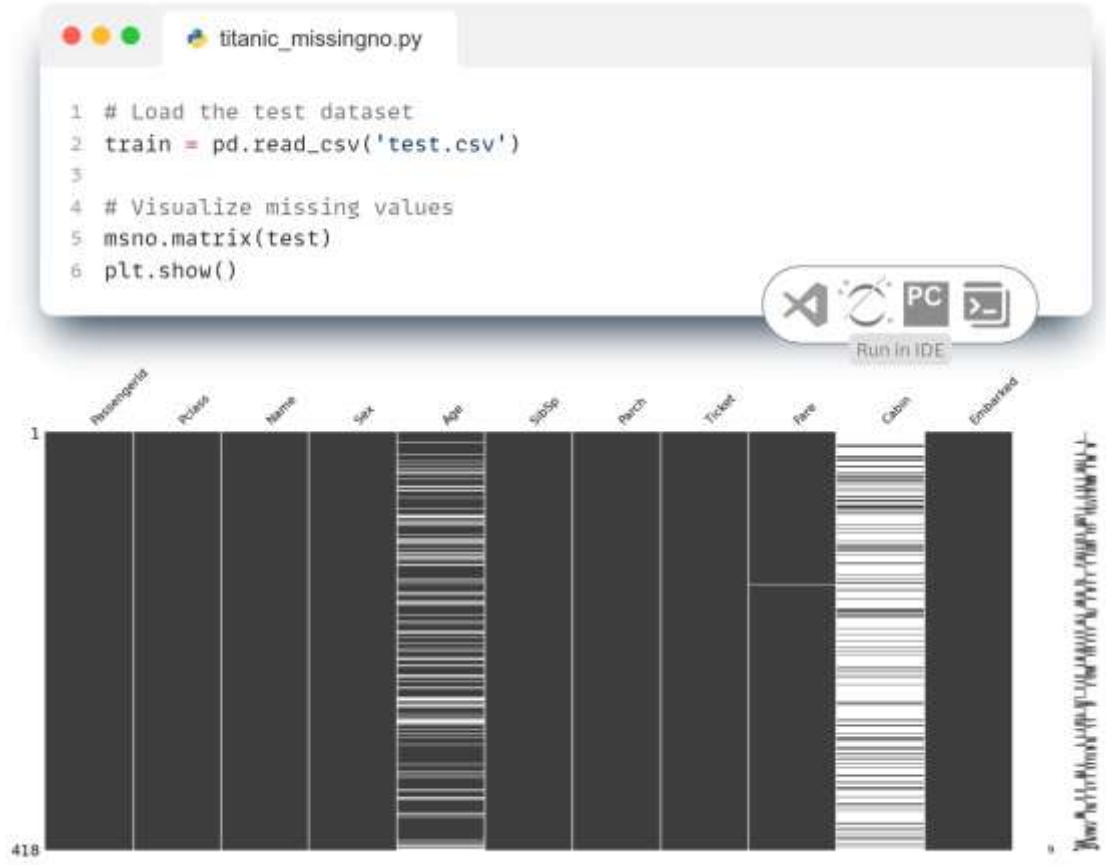


Fig. 9.2-7 Visualización de los datos que faltan en el conjunto de datos de prueba Titanic, que sólo contiene características de los pasajeros sin información.

Antes de formular hipótesis y hacer predicciones basadas en el conjunto de datos, los análisis visuales ayudan a identificar patrones clave en los datos, evaluar su calidad e identificar posibles dependencias. Existen muchas técnicas de visualización que pueden ayudarle a comprender mejor el conjunto de datos del Titanic. Puedes utilizar gráficos de distribución para analizar los grupos de edad de los pasajeros, gráficos de supervivencia por sexo y clase, y matrices de datos perdidos para evaluar la calidad de la información y comprender los datos.

- Pidamos a LLM que nos ayude a visualizar los datos del conjunto de datos Titanic enviando la siguiente petición de texto a cualquier modelo LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o cualquier otro):

> Por favor, muestre algunos gráficos sencillos para el conjunto de datos Titanic. Descargue usted mismo el conjunto de datos y muestre los ⏎

- Respuesta LLM en forma de código ya preparado y gráficos que visualizan los parámetros del conjunto de datos

Create in LLM Chat

titanic_plot.py

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = "/mnt/data/titanic.csv"
df = pd.read_csv(file_path)

# Convert column names to lowercase for consistency
df.columns = df.columns.str.lower()

# Basic Plots
plt.figure(figsize=(10, 5))

# Bar Chart: Survival Count
plt.subplot(1, 3, 1)
df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
plt.title("Survivors vs. Non-Survivors")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Count")

# Histogram: Age Distribution
plt.subplot(1, 3, 2)
df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")

# Scatter Plot: Fare vs. Age
plt.subplot(1, 3, 3)
plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
plt.title("Fare vs. Age")
plt.xlabel("Age")
plt.ylabel("Fare")

# Show the plots
plt.tight_layout()
plt.show()
```

Run in IDE

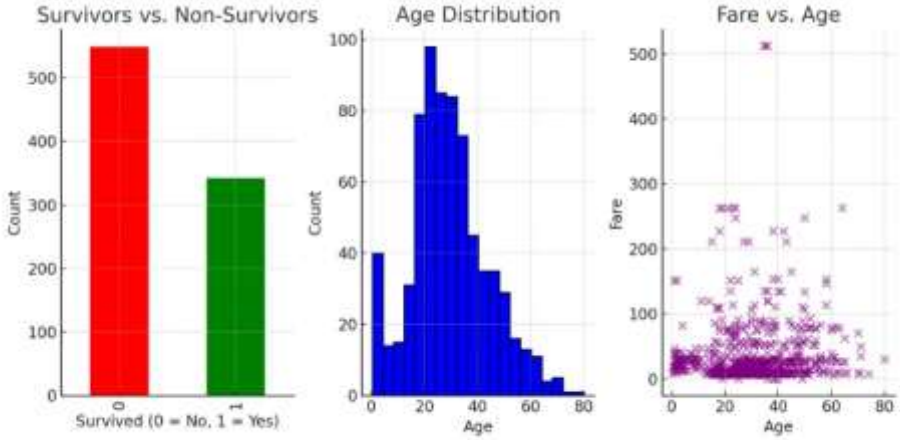


Fig. 9.2-8 LLM le ayuda a obtener una visualización instantánea de los datos de su conjunto de datos.

La visualización de datos es un paso importante para preparar el conjunto de datos para la posterior construcción de un modelo de aprendizaje automático, al que sólo se puede acceder mediante la comprensión de los datos.

Aprendizaje automático en acción: de los pasajeros del Titanic a la gestión de proyectos

La principal hipótesis utilizada para explorar el marco de aprendizaje automático a partir del conjunto de datos del Titanic es que determinados grupos de pasajeros tenían más probabilidades de sobrevivir.

> La tabla pequeña de los pasajeros del Titanic se ha hecho popular en todo el mundo, y millones de personas la utilizan para entrenar, experimentar y probar modelos para averiguar qué algoritmos e hipótesis construirán el modelo de predicción de supervivencia más preciso basado en el conjunto de datos de entrenamiento de los pasajeros del Titanic.

El atractivo del conjunto de datos Titanic se debe a su compacidad: con varios cientos de filas y doce columnas (Fig. 9.2-6), ofrece amplias posibilidades de análisis. El conjunto de datos es, de forma relativamente sencilla, un ejemplo clásico de una solución de clasificación binaria, donde el objetivo del problema -la supervivencia- se expresa en el cómodo formato 0 o 1.

> John Wheeler en "It from Bit" [7] sostiene que el universo se basa en elecciones binarias. Del mismo modo, una empresa dirigida por personas formadas por moléculas se basa en realidad en una serie de elecciones binarias.

Además, los datos se basan en un acontecimiento histórico real, lo que los hace valiosos para la investigación, a diferencia de los ejemplos creados artificialmente. Sólo en la plataforma Kaggle, una de las mayores de Data Pipeline y ETL, 1.355.998 personas participaron en los retos basados en el conjunto de datos del Titanic, desarrollando 53.963 soluciones únicas de Data Pipeline [157] (Fig, 9.2-9).

> Parece increíble, pero sólo 1.000 líneas de datos sobre los pasajeros del Titanic con 12 parámetros se han convertido en campo de millones de hipótesis, cadenas lógicas y Data-Pipelines únicos. De un pequeño conjunto de datos nacen infinitas percepciones, hipótesis e interpretaciones: desde simples modelos de supervivencia hasta complejos conjuntos que tienen en cuenta patrones ocultos y complejos laberintos de razonamiento.

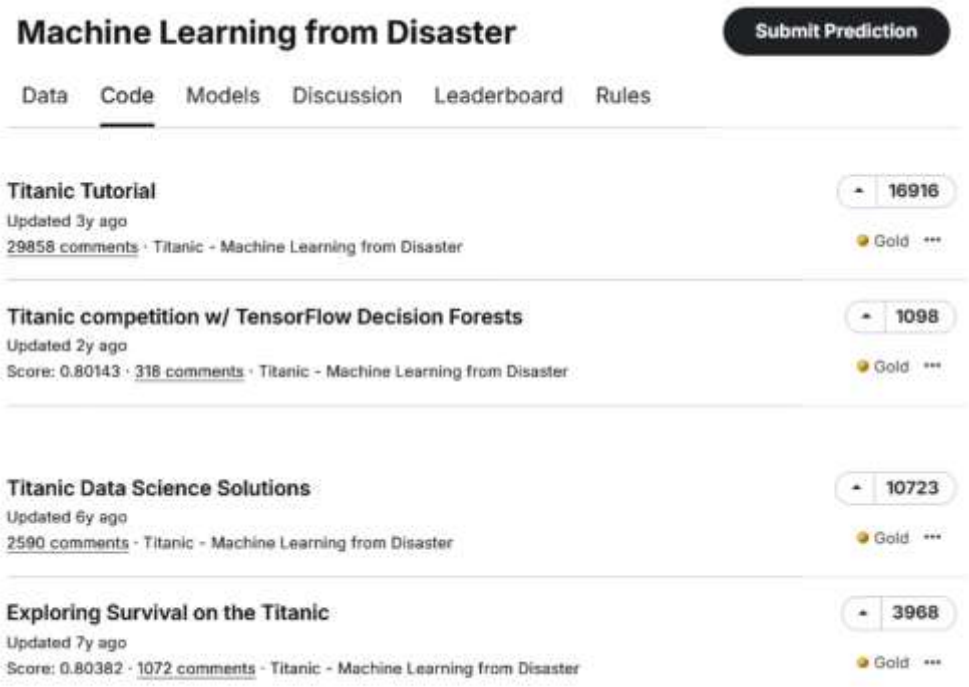


Fig. 9.2-9 Las cinco primeras soluciones de un total de 53.963 soluciones Pipeline listas y de código abierto. Casi 1,5 millones de personas ya han intentado resolver este problema sólo en Kaggle [157].

Si incluso una tabla tan pequeña puede generar millones de soluciones únicas (Fig. 9.2-9), ¿qué podemos decir de los conjuntos de datos reales de construcción industrial en los que los parámetros se miden en decenas de miles?

> Un diseño CAD estándar de un edificio relativamente pequeño contiene decenas de miles de entidades con miles de parámetros, desde características geométricas a atributos de coste y tiempo. Imagínese cuántas ideas, relaciones, predicciones e hipótesis de gestión potenciales se esconden en los datos de todos los proyectos de su empresa recopilados a lo largo de los años. Los datos históricos de los proyectos no son sólo un archivo: son la memoria viva de una organización, su huella digital que puede analizarse para construir un gran número de hipótesis únicas.

Y lo que es más importante, no tiene que esperar a que la comunidad de Kaggle se interese por su empresa o sus datos. Puede empezar a trabajar con lo que tiene hoy: ejecutar análisis con sus propios datos, entrenar modelos con sus propios datos, identificar repeticiones, anomalías y patrones de. Donde antes hacían falta años de experimentación y costosas consultorías, ahora basta con iniciativa, un LLM, un enfoque abierto de los datos y ganas de aprender.

- Para construir un algoritmo de aprendizaje automático, que prediga las tasas de supervivencia de los pasajeros basándose en el conjunto de datos de entrenamiento de pasajeros train.csv, pidamos a LLM que resuelva este problema por nosotros:

A partir del conjunto de datos de entrenamiento de los pasajeros del Titanic, construir un modelo de aprendizaje automático para predecir la capacidad de supervivencia. ⏎

Respuesta de LLM:

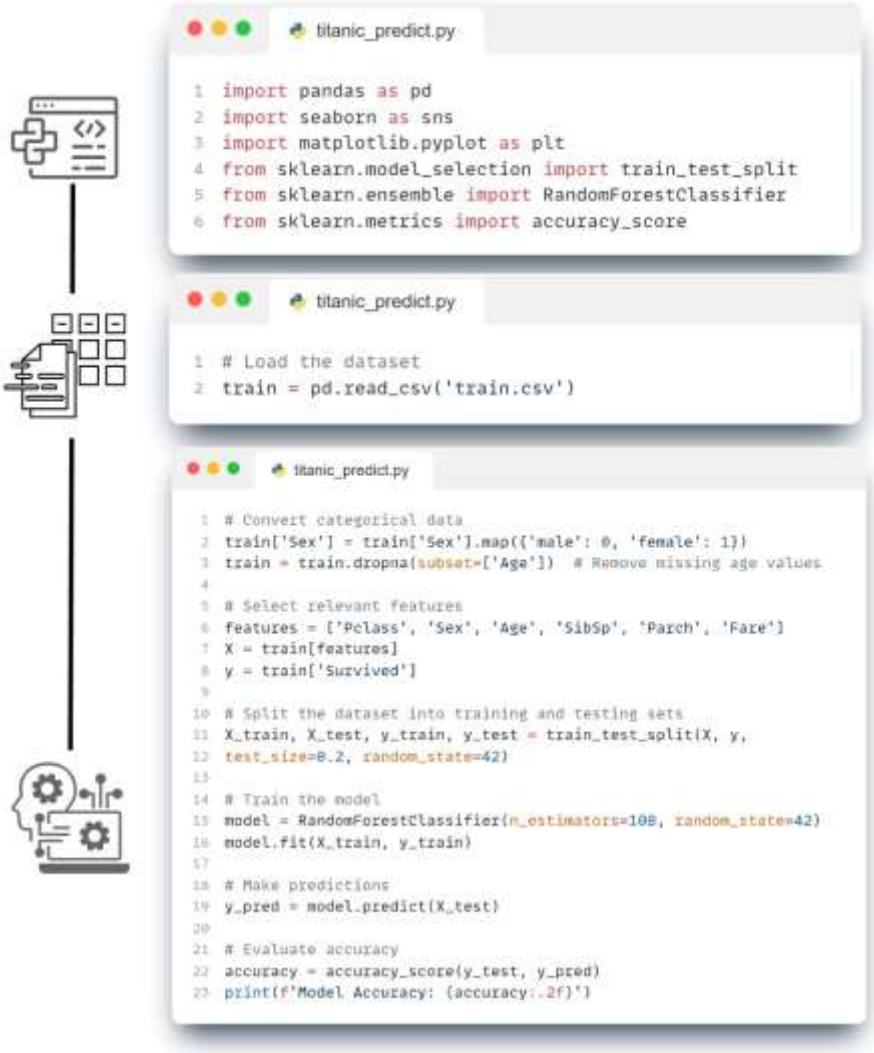


Fig. 9.2-10 LLM construyó una predicción de supervivientes del Titanic utilizando el algoritmo de aprendizaje automático Random Forest.

El código resultante de LLM (Fig. 9.2-10) carga los datos de los pasajeros del Titanic, los limpia, convierte las variables categóricas (por ejemplo, género a formato numérico), y entrena el modelo mediante el algoritmo RandomForestClassifier para predecir si un pasajero sobrevivió o no (hablaremos más sobre algoritmos populares en los siguientes capítulos).

El código separa los datos de entrenamiento en conjuntos de entrenamiento y de prueba (el sitio web de Kaggle ya ha creado test.csv (Fig. 9.2-7) y train.csv (Fig. 9.2-6) para el entrenamiento, después se entrena el modelo con los datos de entrenamiento y se prueba con los datos de prueba para ver lo bueno que es un modelo de predicción concreto. Tras el entrenamiento, se introducen en el modelo los datos de prueba de test.csv (con datos reales sobre los que sobrevivieron o no) y éste predice quién sobrevivió y quién no. En nuestro caso, la precisión del modelo de aprendizaje automático que obtuvimos es de aproximadamente el 80%, lo que demuestra que capta los patrones bastante bien.

> El aprendizaje automático puede compararse a un niño que intenta encajar un bloque rectangular en un agujero redondo. En las fases iniciales, el algoritmo prueba muchas aproximaciones, encontrando errores e incoherencias. Este proceso puede parecer ineficiente, pero proporciona un aprendizaje importante: analizando cada error, el modelo mejora sus predicciones y toma decisiones cada vez más precisas.

Ahora este modelo (Fig. 9.2-10) se puede utilizar para predecir la tasa de supervivencia de nuevos pasajeros y, por ejemplo, si lo alimentas con información de pasajeros utilizando la función model.predict los parámetros: "hombre", "3ª clase", "25 años", "sin familiares a bordo", el modelo producirá una predicción - que el pasajero con un 80% de probabilidad no sobrevivirá al desastre si estaba en el Titanic en 1912 (Fig. 9.2-11).

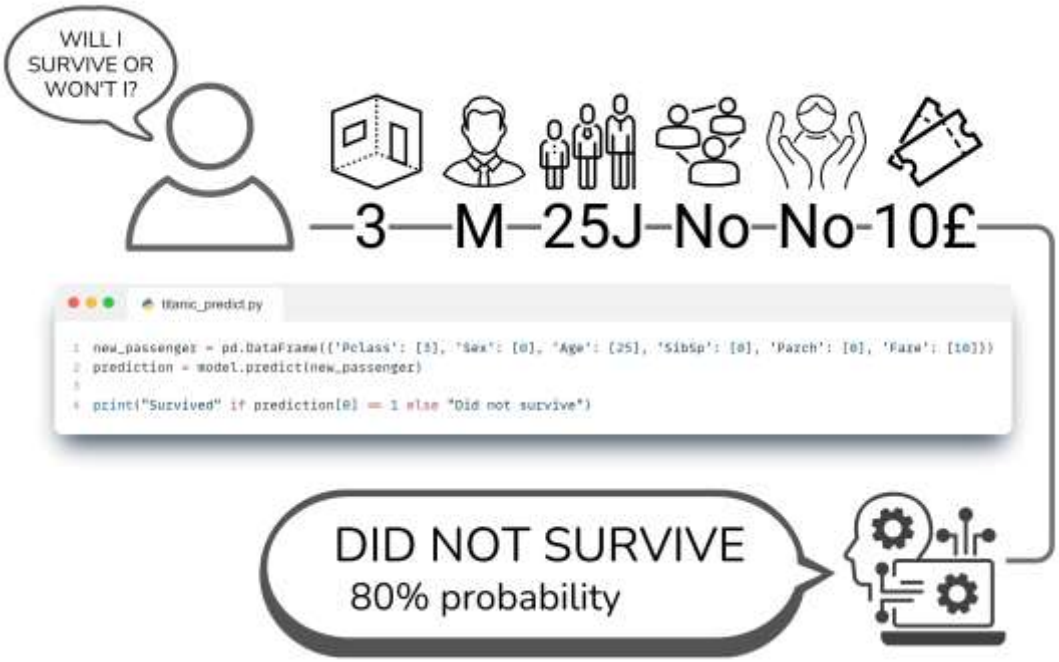


Fig. 9.2-11 El modelo que hemos creado anteriormente puede predecir ahora con un 80% de probabilidad si sobrevivirá o no algún nuevo pasajero del Titanic.

El modelo de predicción de la supervivencia de los pasajeros del Titanic ilustra un concepto mucho más amplio: miles de profesionales del sector de la construcción toman a diario decisiones "duales" similares: la vida o la muerte de una decisión, un proyecto, un presupuesto, una herramienta, beneficios o pérdidas, seguridad o riesgo. Como en el ejemplo del Titanic, donde el resultado dependía de factores (sexo, edad,

clase), en la construcción cada aspecto de la decisión está influido por muchos factores y variables propios (columnas de tablas): coste de los materiales, cualificación de los trabajadores, plazos, meteorología, logística, riesgos técnicos, comentarios y cientos de miles de parámetros más.

> En el sector de la construcción, el aprendizaje automático sigue los mismos principios que en otros campos: los modelos se entrenan con datos históricos -de proyectos, contratos, estimaciones- para probar diversas hipótesis y encontrar las soluciones más eficaces. Este proceso es muy parecido al de enseñar a un niño por ensayo y error: con cada ciclo, los modelos se adaptan y se vuelven más precisos.

El uso de datos acumulados abre nuevos horizontes a la construcción. En lugar de cálculos manuales que requieren mucha mano de obra, se pueden entrenar modelos capaces de predecir características clave de futuros proyectos con un alto grado de precisión. De este modo, el análisis predictivo convierte el sector de la construcción en un espacio en el que no sólo se puede planificar, sino también predecir con confianza los acontecimientos.

Predicciones y previsiones basadas en datos históricos

Los datos recopilados sobre los proyectos de la empresa abren la posibilidad de construir modelos capaces de predecir las características de coste y tiempo de futuros objetos aún no realizados, sin necesidad de realizar cálculos y comparaciones manuales que requieren mucho tiempo. Esto permite acelerar y simplificar considerablemente los procesos de valoración, sin basarse en suposiciones subjetivas, sino en previsiones matemáticas sólidas.

Anteriormente, en la cuarta parte del libro, hemos considerado en detalle los métodos tradicionales de estimación de costes de proyectos, incluido el método basado en recursos, y también hemos mencionado los enfoques paramétricos y de expertos. Estos métodos siguen siendo pertinentes, pero en la práctica moderna empiezan a enriquecerse con herramientas de análisis estadístico y aprendizaje automático, que pueden mejorar notablemente la precisión y reproducibilidad de las estimaciones.

Los procesos de cálculo manual y semiautomático de precios y atributos temporales se complementarán en el futuro con la opinión y las predicciones de modelos ML capaces de analizar datos históricos, encontrar patrones ocultos y proponer decisiones informadas.
Se generarán automáticamente nuevos datos y escenarios a partir de la información ya disponible, de forma similar a como los modelos lingüísticos (LLM) crean textos, imágenes y código a partir de los datos recogidos durante años de fuentes abiertas [158].

> Del mismo modo que hoy los seres humanos se basan en la experiencia, la intuición y las estadísticas internas para evaluar los acontecimientos futuros, en los próximos años el futuro de los proyectos de construcción vendrá determinado cada vez más por una combinación de conocimientos acumulados y modelos matemáticos de aprendizaje automático.

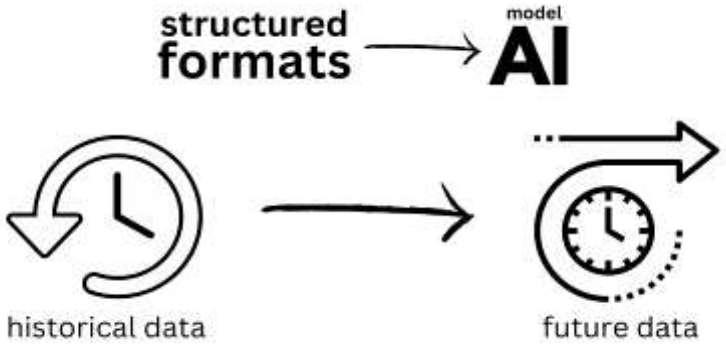


Fig. 9.2-12 Los datos históricos cualitativos y estructurados de las empresas son el material sobre el que se construyen los modelos de aprendizaje automático y las predicciones.

Consideremos un ejemplo sencillo: predecir el precio de una casa en función de su superficie, el tamaño de la parcela, el número de habitaciones y la ubicación geográfica. Un enfoque consiste en construir un modelo clásico que analice estos parámetros y calcule el precio esperado (Fig, 9.2-13). Este enfoque requiere una fórmula precisa y conocida de antemano, lo que es prácticamente imposible en la práctica real.

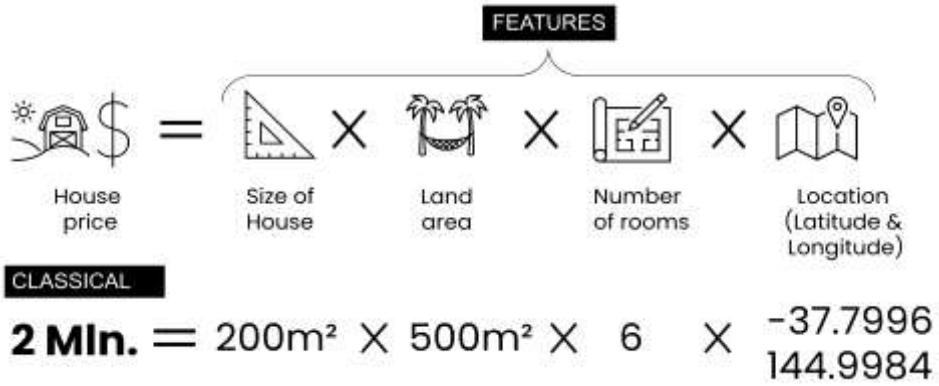


Fig. 9.2-13 Se puede utilizar un algoritmo clásico para estimar el valor de una casa con una fórmula fija que hay que encontrar.

El aprendizaje automático elimina la búsqueda manual de fórmulas y las sustituye por algoritmos entrenados que identifican de forma independiente dependencias que son muchas veces más precisas que cualquier ecuación predeterminada. Como alternativa, creemos un algoritmo de aprendizaje automático, que generará un modelo basado en una comprensión previa del problema y en datos históricos que pueden ser incompletos (Fig, 9.2-14).

> Utilizando la fijación de precios como ejemplo, el aprendizaje automático permite crear distintos tipos de modelos matemáticos que no requieren conocer el mecanismo exacto de formación de costes. El modelo "aprende" de los datos de proyectos anteriores, ajustándose a patrones reales entre los parámetros de construcción, su coste y los plazos.

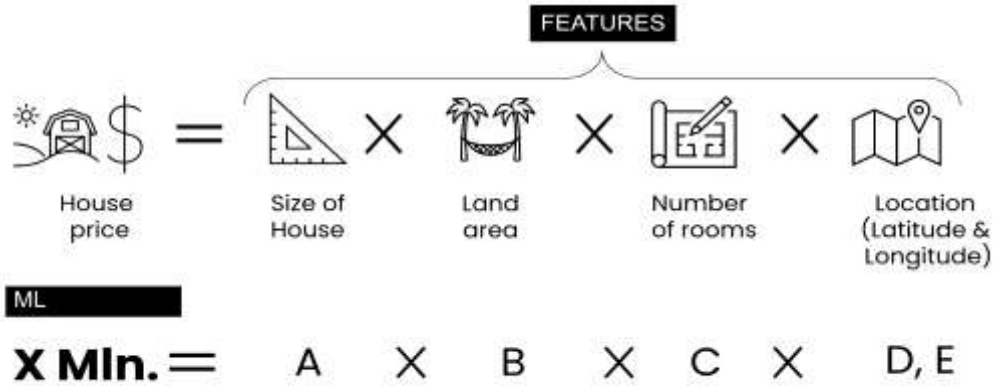


Fig. 9.2-14 A diferencia de la estimación clásica basada en fórmulas, el algoritmo de aprendizaje automático se entrena con datos históricos.

En el contexto del aprendizaje automático supervisado, cada proyecto del conjunto de datos de formación contiene atributos de entrada (por ejemplo, coste y tiempo de edificios similares) y valores de salida previstos (por ejemplo, coste o tiempo). Se utiliza un conjunto de datos similar para crear y personalizar el modelo de aprendizaje automático (Fig, 9.2-15). Cuanto mayor sea el conjunto de datos y mayor sea la calidad de los datos que contiene, más preciso será el modelo y más exactos serán los resultados de la predicción.

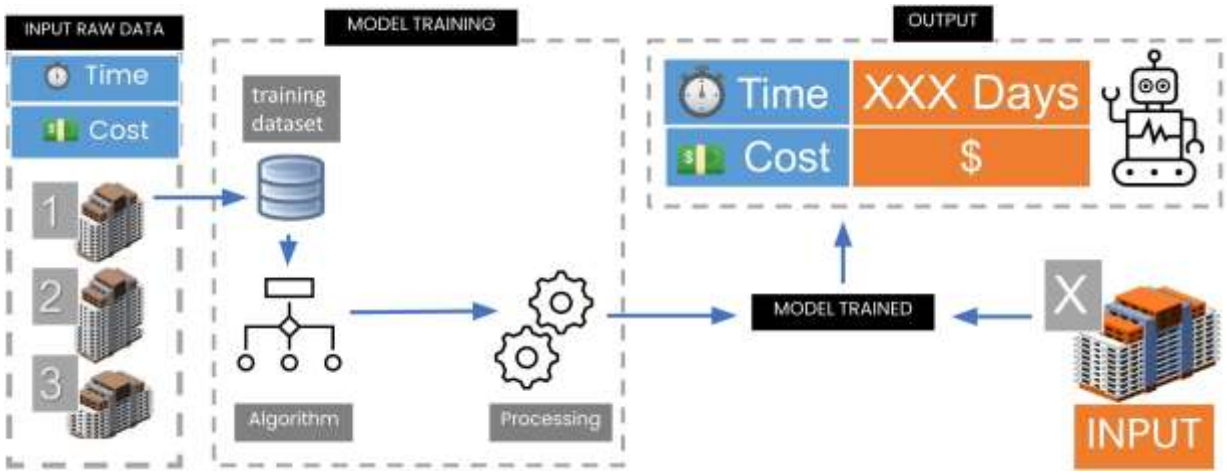


Fig. 9.2-15 Un modelo ML entrenado con datos de costes y plazos de proyectos anteriores determinará el coste y el plazo de un nuevo proyecto con una cierta probabilidad.

Una vez creado el modelo y entrenado para estimar la construcción de un nuevo proyecto, basta con proporcionar al modelo nuevos atributos para el nuevo proyecto y el modelo proporcionará resultados estimados basados en los patrones aprendidos previamente con cierta probabilidad.

Conceptos clave del aprendizaje automático

Aprendizaje automático no es magia, sólo matemáticas, datos y búsqueda de patrones. No tiene inteligencia real, sino que es un programa entrenado en datos para reconocer patrones y tomar decisiones sin la constante intervención humana.

El aprendizaje automático utiliza una serie de conceptos clave para describir su estructura (Fig, 9.2-16):

- **Las etiquetas** son variables objetivo o atributos (el parámetro "Superviviente" en el conjunto de datos Titanic) que el modelo debe predecir. Ejemplo: coste de construcción (en dólares), duración de las obras (en meses).
- **Las características** son variables independientes o atributos que sirven de entrada al modelo. En un modelo de previsión, se utilizan para predecir etiquetas. Ejemplos: tamaño de la parcela (en metros cuadrados), número de plantas de un edificio, superficie total de un edificio (en metros cuadrados), ubicación geográfica (latitud y longitud), tipo de materiales utilizados en la construcción. El número de características también determina la dimensionalidad de los datos.
- **Un modelo** es un conjunto de hipótesis diferentes, una de las cuales se aproxima a la función objetivo que debe predecirse o aproximarse. Ejemplo: modelo de aprendizaje automático, que utiliza técnicas de análisis de regresión para predecir el coste y el calendario de la construcción.
- **Algoritmo de aprendizaje El algoritmo de aprendizaje** es el proceso de encontrar la mejor hipótesis de un modelo que se ajuste exactamente a la función objetivo utilizando un conjunto de datos de entrenamiento. Ejemplo: un algoritmo de regresión lineal, KNN o bosque aleatorio que analiza datos de costes y tiempos de construcción para identificar relaciones y patrones.
- **Entrenamiento** - Durante el proceso de entrenamiento, el algoritmo analiza los datos de entrenamiento, encontrando patrones que corresponden a la relación entre los atributos de entrada y las etiquetas objetivo. El resultado de este proceso es un modelo de aprendizaje automático entrenado, listo para la predicción. Ejemplo: un proceso en el que un algoritmo analiza datos históricos de construcción (coste, tiempo, características de la obra) para crear un modelo predictivo.

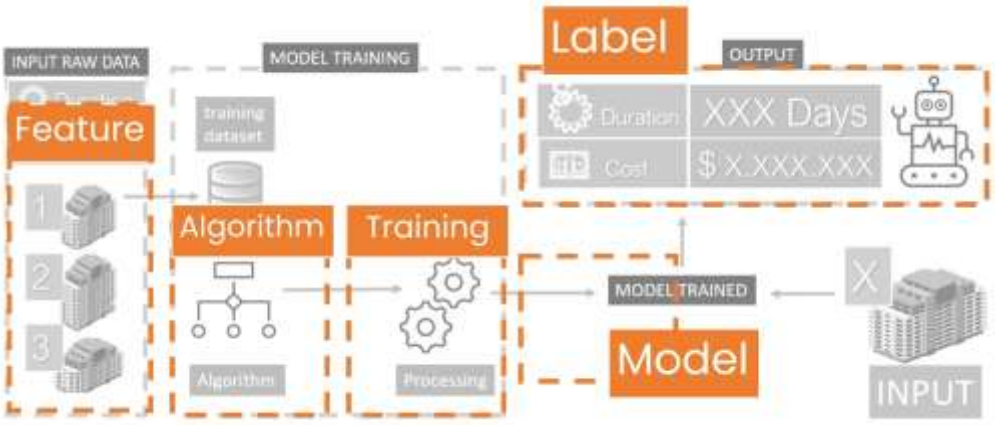


Fig. 9.2-16 ML utiliza etiquetas y atributos para crear modelos que se entrenan en datos utilizando algoritmos para predecir resultados.

El aprendizaje automático no existe de forma aislada, sino que forma parte de un ecosistema más amplio

de disciplinas analíticas que incluyen la estadística, las bases de datos, la minería de datos, el reconocimiento de patrones, el análisis de big data y la inteligencia artificial. La Fig. 9.2-17 muestra cómo estos campos se solapan y complementan entre sí, proporcionando un marco integral para los modernos sistemas de toma de decisiones y automatización.

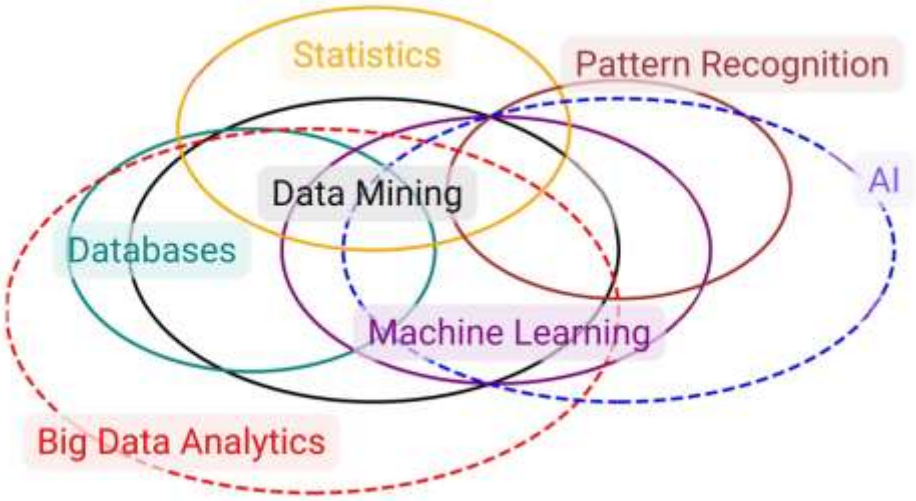


Fig. 9.2-17 La relación entre las diferentes áreas del análisis de datos: estadística, aprendizaje automático, inteligencia artificial, big data, reconocimiento de patrones y minería de datos.

> El principal objetivo del aprendizaje automático es dotar a los ordenadores de la capacidad de aprender automáticamente conocimientos sin intervención ni ayuda humana y ajustar sus acciones en consecuencia [159].

Así pues, en el futuro, el papel del ser humano consistirá únicamente en dotar a la máquina de capacidades cognitivas: establecerá las condiciones, ponderaciones y parámetros, y el modelo de aprendizaje automático hará el resto.

En el próximo capítulo estudiaremos ejemplos concretos de aplicación del algoritmo. Se utilizarán tablas reales y modelos simplificados para mostrar cómo se construye la previsión paso a paso.

CAPÍTULO 9.3.
PREDICCIÓN DE COSTES Y PLAZOS MEDIANTE APRENDIZAJE AUTOMÁTICO

Un ejemplo de uso del aprendizaje automático para encontrar el coste y los plazos de un proyecto

La estimación del tiempo y el coste de la construcción es uno de los procesos clave en las actividades de una empresa constructora. Tradicionalmente, dichas estimaciones han sido realizadas por expertos basándose en la experiencia, libros de referencia y bases de datos normativas. Sin embargo, con la transformación digital y el aumento de la disponibilidad de datos, ahora es posible utilizar modelos de aprendizaje automático (ML) para mejorar la precisión y la automatización de dichas estimaciones.

La introducción del aprendizaje automático en el proceso de cálculo del coste y el calendario de construcción no sólo permite una planificación más eficiente, sino que también se convierte en un punto de partida para integrar modelos inteligentes en otros procesos empresariales, desde la gestión de riesgos hasta la optimización de la logística y las adquisiciones.

> Es importante poder determinar rápidamente cuánto se tardará en construir un proyecto y cuál será su coste total. Estas cuestiones sobre el tiempo y el coste del proyecto han estado tradicionalmente en el punto de mira tanto de los clientes como de las empresas constructoras desde el nacimiento de la industria de la construcción.

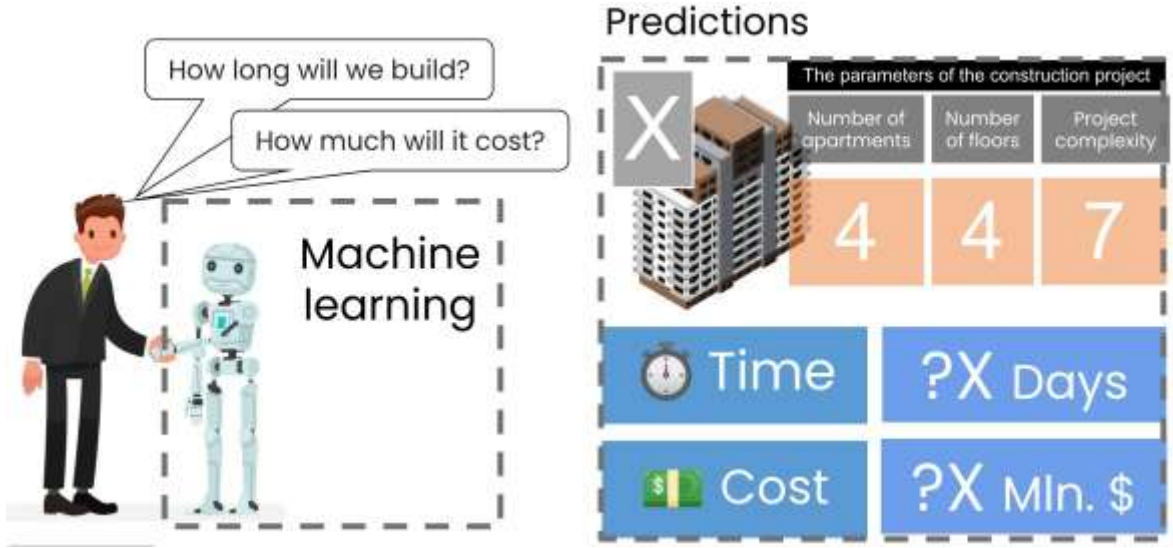


Fig. 9.3-1 En los proyectos de construcción, la rapidez y la calidad de la estimación del tiempo y el coste de construcción son factores clave para el éxito.

En el siguiente ejemplo, se extraerán datos clave de proyectos anteriores y se utilizarán para desarrollar un

modelo de aprendizaje automático, que nos permitirá utilizar el modelo para estimar el coste y el calendario de nuevos proyectos de construcción con nuevos parámetros (Fig, 9.3-1).

Consideremos tres proyectos con tres atributos clave: el número de pisos (donde 100 pisos equivale al número 10 para facilitar la visualización), el número de plantas y una medida condicional de la complejidad de la construcción en una escala de 1 a 10, donde 10 es la puntuación más alta de complejidad. En el aprendizaje automático, el proceso de convertir y simplificar valores como 100 a 10 o 50 a 5 se denomina "normalización".

> **La normalización** en el aprendizaje automático es el proceso de llevar datos numéricos diferentes a una escala común para facilitar su procesamiento y análisis. Este proceso es especialmente importante cuando los datos tienen diferentes escalas y unidades de medida.

Supongamos que el primer proyecto (Fig, 9.3-2) tenía 50 pisos (tras la normalización, 5), 7 plantas y una puntuación de complejidad de 2, lo que significaba una construcción relativamente sencilla. El segundo proyecto ya tenía 80 pisos, 9 plantas y un diseño relativamente complejo. En estas condiciones, la construcción del primer y segundo edificio de apartamentos duró 270 y 330 días, y el coste total del proyecto fue de 4,5 y 5,8 millones de dólares, respectivamente.

Fig. 9.3-2 Ejemplo de un conjunto de proyectos pasados que se utilizará para estimar el tiempo y el coste del futuro proyecto X.

Cuando se construye un modelo de aprendizaje automático para este tipo de datos, la tarea principal consiste en identificar los atributos (o etiquetas) críticos para la predicción, en este caso, el tiempo y el coste de construcción. Con un conjunto de datos pequeño, utilizaremos información sobre proyectos de construcción anteriores para planificar otros nuevos: utilizando algoritmos de aprendizaje automático, tenemos que predecir el coste y la duración de la construcción de un nuevo proyecto X basándonos en atributos dados del nuevo proyecto, como 40 pisos, 4 plantas y una complejidad de proyecto relativamente alta de 7 (Fig. 9.3-2). En el mundo real, el número de parámetros de entrada puede ser mucho mayor, oscilando entre

varias decenas y cientos de factores. Estos pueden incluir: tipo de materiales de construcción, zona climática, nivel de cualificación de los contratistas, disponibilidad de servicios públicos, tipo de cimentación, temporada de inicio de las obras, comentarios de los capataces, etc.

> Para crear un modelo de aprendizaje automático predictivo, necesitamos elegir un algoritmo para crearlo. Un algoritmo en aprendizaje automático es como una receta matemática que enseña al ordenador cómo hacer predicciones (mezclar en el orden correcto los parámetros) o tomar decisiones basadas en datos.

Para analizar datos sobre proyectos de construcción anteriores y predecir el tiempo y el coste de proyectos futuros (Fig, 9.3-2), puede utilizarse un conocido algoritmo de aprendizaje automático:

- **Regresión lineal (Linear regression):** este algoritmo trata de encontrar una relación directa entre atributos, por ejemplo entre el número de plantas y el coste de construcción. El objetivo del algoritmo es encontrar una ecuación lineal que describa lo mejor posible esta relación, lo que permite hacer predicciones.
- **Algoritmo k-nearest neighbours (k-NN):** este algoritmo compara un nuevo proyecto con proyectos anteriores similares en tamaño o complejidad. El k-NN clasifica los datos en función de cuál de los k (número) ejemplos de entrenamiento se le aproxima más. En el contexto de la regresión, el resultado es la media o la mediana de los k vecinos más cercanos.
- **Árboles de decisión**: es un modelo de modelización predictiva que divide los datos en subconjuntos basados en diferentes condiciones utilizando una estructura de árbol. Cada nodo del árbol representa una condición o pregunta que conduce a una división posterior de los datos, y cada hoja representa la predicción o resultado final. El algoritmo divide los datos en grupos más pequeños en función de distintas características, por ejemplo, primero por número de historias, luego por complejidad, etc., para hacer una predicción.

Echemos un vistazo a los algoritmos de aprendizaje automático para estimar el coste de un nuevo proyecto utilizando como ejemplo dos algoritmos populares: la regresión lineal y el algoritmo K-nearest neighbours.

Predicción de costes y plazos de proyectos mediante regresión lineal

Regresión lineal es un algoritmo fundamental de análisis de datos que predice el valor de una variable basándose en una relación lineal con otra u otras variables. Este modelo asume que existe una relación lineal directa entre la variable dependiente y una o más variables independientes, y el objetivo del algoritmo es encontrar esta relación.

> La sencillez y claridad de la regresión lineal la han convertido en una herramienta popular en diversos campos. Cuando se trata de una sola variable, la regresión lineal consiste en encontrar la línea que mejor se ajuste a los puntos de datos.

Regresión lineal encuentra la mejor línea recta (línea roja) que se aproxima a la dependencia entre la

variable de entrada X y la variable de salida Y. Esta línea permite predecir los valores de Y para nuevos valores de X basándose en la relación lineal identificada (Fig, 9.3-3).

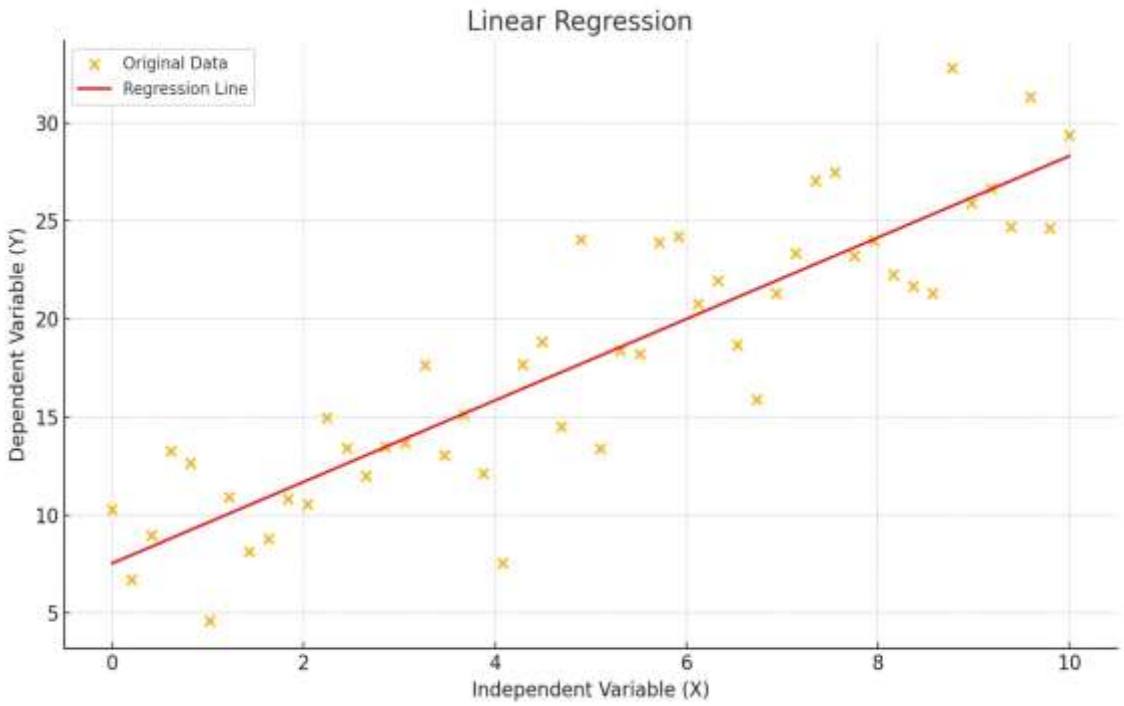


Fig. 9.3-3 El principio de la regresión lineal es encontrar la mejor línea recta a través de los valores de entrenamiento.

Esta línea está representada por una ecuación en la que la introducción de un valor de la variable independiente (X) produce un valor predicho de la variable dependiente (Y). Este proceso permite predecir eficazmente Y a partir de valores conocidos de X utilizando una relación lineal entre ellos. Un ejemplo de cómo encontrar una línea de este tipo con un promedio estadístico puede verse en la Evaluación de Datos de Permisos de Construcción de San Francisco (Fig, 9.1-7), donde se calculó la inflación mediante regresión lineal para diferentes tipos de instalaciones.

Carguemos la tabla de datos del proyecto (Fig, 9.3-2 del capítulo anterior) directamente en el LLM y pidámosle que construya un modelo sencillo de aprendizaje automático para nosotros.

- Envíe una solicitud de texto al chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Necesitamos mostrar la construcción de un modelo sencillo de aprendizaje automático para predecir el coste y el tiempo de realización de un nuevo proyecto X (Fig. 9.3-2 como imagen adjunta) ⏎

Respuesta de LLM:

Create in LLM Chat

Linear regression.py

```
1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 236])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost
```

Run in IDE

Result

```
(array([238.44444444]), array([3042337.77777778]))
```

Fig. 9.3-4 LLM eligió la regresión lineal para crear un modelo de aprendizaje automático para predecir el coste y el tiempo del proyecto.

El LLM reconoció automáticamente la tabla de la imagen adjunta y convirtió los datos de un formato visual a un array de tablas (Fig. 9.3-4 - fila 6). Esta matriz se utilizó como base para crear características y etiquetas a partir de las cuales se creó un modelo de aprendizaje automático (Fig. 9.3-4 - fila 17-22), que utilizó la regresión lineal.

Utilizando un modelo básico de regresión lineal que se entrenó en un conjunto de datos "extremadamente pequeño", se hicieron predicciones para un nuevo proyecto de construcción hipotético etiquetado Proyecto

X. En nuestro problema, este proyecto se caracteriza por 40 pisos, 4 plantas y un nivel de complejidad de 7 (Fig, 9.3-2).

Como predice un modelo de regresión lineal basado en un conjunto de datos limitado y pequeño para el nuevo Proyecto X (Fig, 9.3-4 - línea 24-29):

- **La duración de las obras** será de aproximadamente 238 días (238,4444444)
- **El gasto total** será de aproximadamente$ 3.042.338 (3042337,777)

Para profundizar en la hipótesis del coste del proyecto, resulta útil experimentar con distintos algoritmos y métodos de aprendizaje automático. Por lo tanto, vamos a predecir los mismos valores de coste y tiempo para un nuevo proyecto X basándonos en un pequeño conjunto de datos históricos utilizando el algoritmo K-Nearest Neighbours (k-NN).

Predicción de costes y plazos de un proyecto mediante el algoritmo K-nearest neighbour (k-NN)

Utilizamos el algoritmo k-Nearest Neighbours (k-NN) como predictor adicional para estimar el coste y la duración de un nuevo proyecto. El algoritmo k-Nearest Neighbours (k-NN) es un método de aprendizaje automático supervisado (aprendizaje automático supervisado) tanto para la clasificación como para la regresión. También hemos hablado anteriormente del algoritmo k-NN en el contexto de la búsqueda de bases de datos vectoriales (Fig, 8.2-2), donde se utiliza para encontrar los vectores más cercanos (por ejemplo, textos, imágenes o descripciones técnicas). En este enfoque, cada proyecto se representa como un punto en un espacio multidimensional, donde cada dimensión corresponde a un atributo específico del proyecto.

En nuestro caso, dados los tres atributos de cada proyecto, los representaremos como puntos en un espacio tridimensional (Fig, 9.3-5). Así, nuestro próximo proyecto X se localizará en este espacio con coordenadas (x=4, y=4, z=7). Hay que tener en cuenta que, en condiciones reales, el número de puntos y la dimensionalidad del espacio pueden ser órdenes de magnitud mayores.

> El algoritmo K-NN (k-nearest neighbours) funciona midiendo la distancia entre el proyecto deseado X y los proyectos de la base de datos de entrenamiento. Comparando estas distancias, el algoritmo determina los proyectos más cercanos al punto del nuevo proyecto X.

Por ejemplo, si el segundo proyecto (x=8, y=9, z=6) de nuestro conjunto de datos original está mucho más lejos de X (Fig, 9.3-5) que los demás proyectos, puede excluirse del análisis posterior. En consecuencia, sólo podrán utilizarse para los cálculos los dos (k=2) proyectos más cercanos, a partir de los cuales se determinará el valor medio.

Este método, mediante una búsqueda de vecindad, permite evaluar las similitudes entre proyectos, lo que a su vez ayuda a sacar conclusiones sobre el coste y el calendario probables de un nuevo proyecto basándose en proyectos similares realizados con anterioridad.

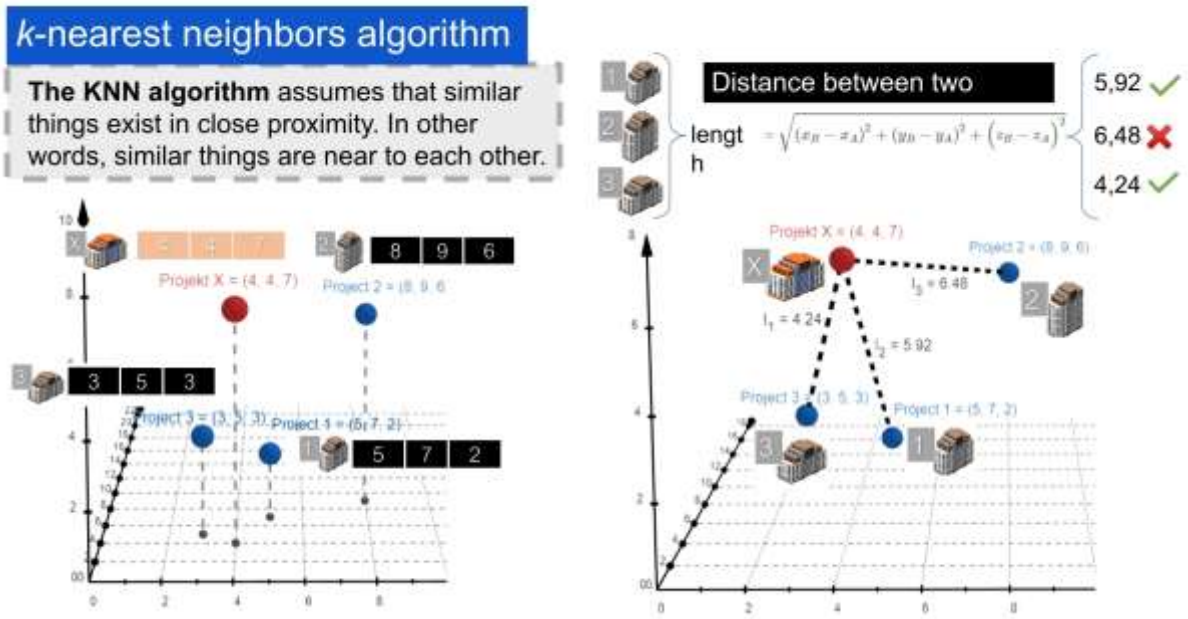


Fig. 9.3-5 En el algoritmo K-NN, los proyectos se representan como puntos en un espacio multidimensional, y los proyectos más cercanos se seleccionan en función de las distancias para la evaluación de similitudes y la predicción.

El trabajo de k-NN implica varios pasos clave:

- **Preparación de los datos:** primero se cargan los conjuntos de datos de entrenamiento y de prueba. Los datos de entrenamiento se utilizan para "entrenar" el algoritmo, y los de prueba para comprobar su eficacia.
- **Selección del parámetro K:** se selecciona un número K, que indica cuántos vecinos más próximos (puntos de datos) deben tenerse en cuenta en el algoritmo. El valor de "K" es muy importante porque afecta al resultado.
- Proceso de clasificación y regresión para datos de prueba:
  - **Cálculo de distancias:** para cada elemento de los datos de prueba, se calcula la distancia a cada elemento de los datos de entrenamiento (Fig, 9.3-5). Para ello pueden utilizarse distintos métodos de medición de distancias, como la distancia euclidiana (el método más común), la distancia Manhattan o la distancia Hamming.
  - **Ordenación y selección de los K vecinos** más cercanos**:** tras calcular las distancias, se ordenan y se seleccionan los K puntos más cercanos al punto de prueba.
  - **Determinación de la clase o el valor de un** punto de prueba**:** si se trata de una tarea de clasificación, la clase del punto de prueba se determina en función de la clase más frecuente entre K vecinos seleccionados. Si se trata de una tarea de regresión, se calcula la media (u otra medida de tendencia central) de los valores de los K vecinos.
- **Finalización del proceso:** una vez que se han clasificado todos los datos de la prueba o se han hecho predicciones para ellos, el proceso ha finalizado.

El algoritmo k-nearest neighbours (k-NN) es eficaz en muchas aplicaciones prácticas y constituye una de las principales herramientas del arsenal de los especialistas en aprendizaje automático. Este algoritmo es popular por su sencillez y eficacia, especialmente en tareas en las que las relaciones entre los datos son fáciles de interpretar.

En nuestro ejemplo, tras aplicar el algoritmo de K vecinos más próximos, se identificaron los dos proyectos (de nuestra pequeña muestra) con menor distancia al proyecto X (Fig, 9.3-5). A partir de estos proyectos, el algoritmo determina la media de su precio y duración de la construcción. Tras el análisis (Fig. 9.3-6), el algoritmo, haciendo la media de los vecinos más próximos, concluye que el proyecto X costará aproximadamente$ 3.800.000 $ y tardará unos 250 días en completarse.

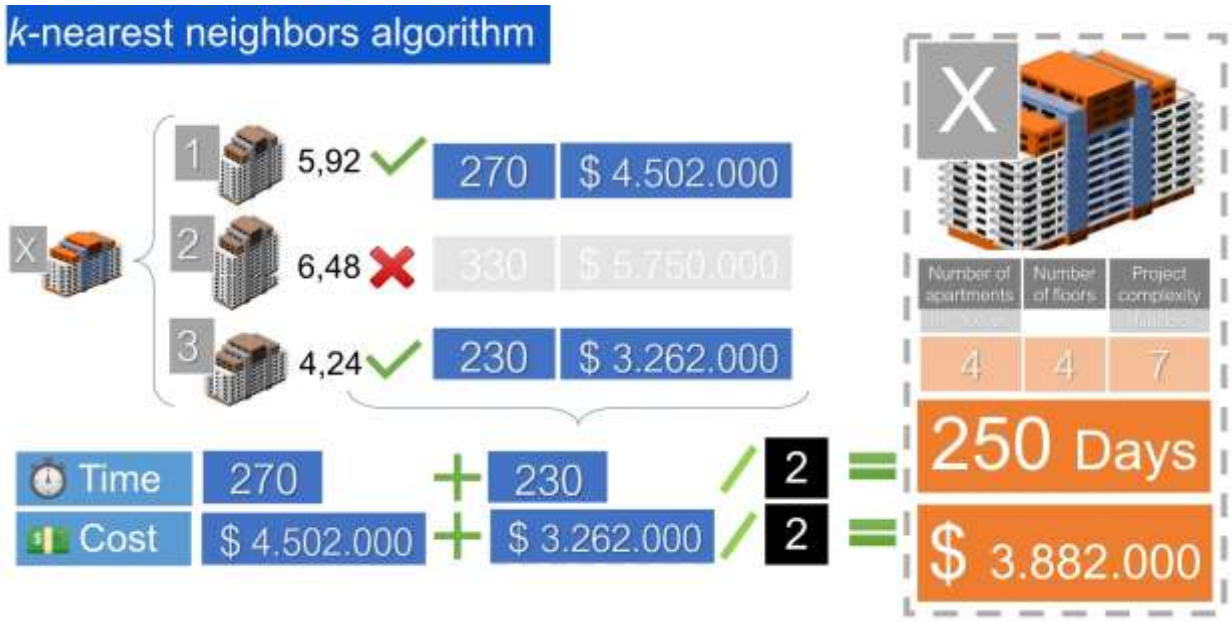


Fig. 9.3-6 El algoritmo K-nearest neighbours determina el coste y el calendario del proyecto X analizando los dos proyectos más cercanos de la muestra.

El algoritmo k-Nearest Neighbors (k-NN) es especialmente popular en tareas de clasificación y regresión, como los sistemas de recomendación, donde se utiliza para sugerir productos o contenidos basados en preferencias similares a los intereses de un usuario concreto. Además, k-NN se utiliza ampliamente en diagnósticos médicos para clasificar tipos de enfermedades basándose en los síntomas del paciente, en el reconocimiento de patrones y en el sector financiero para evaluar la solvencia de los clientes.

Incluso con datos limitados, los modelos de aprendizaje automático pueden proporcionar predicciones útiles y mejorar significativamente el componente analítico de la gestión de proyectos de construcción. A medida que se amplían y depuran los datos históricos, es posible pasar a modelos más sofisticados -por ejemplo, teniendo en cuenta el tipo de construcción, la ubicación, la temporada de inicio de las obras y otros factores-.

Nuestra tarea simplificada utilizaba tres atributos para la visualización en el espacio 3D, pero los proyectos reales, por término medio, incluyen cientos o miles de atributos (véase el conjunto de datos del capítulo "Un ejemplo de big data basado en CAD (BIM)"), lo que aumenta significativamente la dimensionalidad del espacio y la complejidad de representar los proyectos como vectores (Fig, 9.3-7).

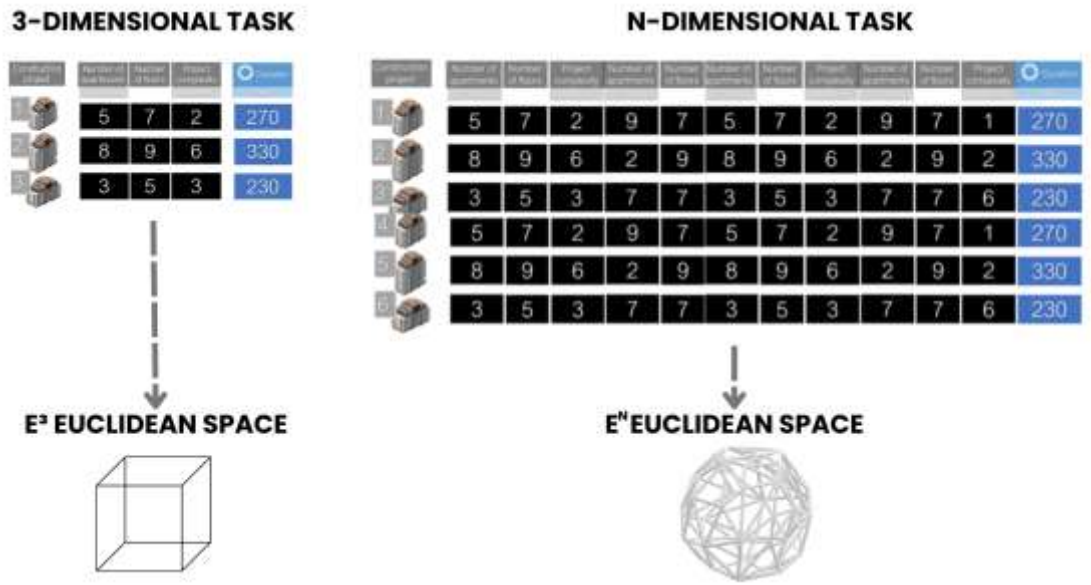


Fig. 9.3-7 En el ejemplo simplificado, se han utilizado tres atributos para la visualización 3D, mientras que los proyectos reales tienen más.

La aplicación de diferentes algoritmos al mismo conjunto de datos para el proyecto X, que tiene 40 pisos, 4 plantas y un nivel de complejidad 7, arrojó diferentes valores de predicción. El algoritmo de regresión lineal predijo un plazo de ejecución de 238 días y un coste de$ 3.042.338 (Fig, 9.3-4), mientras que el algoritmo k-NN predijo 250 días y$ 3.882.000 (Fig, 9.3-6).

La precisión de las predicciones obtenidas mediante modelos de aprendizaje automático, depende directamente del volumen y la calidad de los datos de entrada. Cuantos más proyectos participen en el entrenamiento, y cuanto más completas y precisas estén representadas sus características (atributos) y resultados (etiquetas), mayor será la probabilidad de obtener predicciones fiables con valores de error mínimos.

Las técnicas de preprocesamiento de datos desempeñan un papel importante en este proceso:

- Normalización para ajustar las características a una escala común;
- Detección y eliminación de valores atípicos, eliminando la distorsión del modelo;
- Codificación de atributos categóricos para permitir la manipulación de datos textuales;
- Rellenar los valores que faltan, aumentando la solidez del modelo.

Además, se utilizan métodos de validación cruzada para evaluar la generalizabilidad del modelo y su robustez ante nuevos conjuntos de datos para detectar el exceso de ajuste y mejorar la fiabilidad de la predicción.

> *El caos es un orden que hay que descifrar [160].*
>
> - José Saramago, "El doble"

Aunque le parezca que el caos de sus tareas no puede describirse formalmente, debe saber que cualquier acontecimiento del mundo y, en especial, los procesos de construcción están sujetos a leyes matemáticas, que pueden necesitar el apoyo del cálculo de valores no mediante fórmulas estrictas, sino con la ayuda de estadísticas y datos históricos.

Tanto las estimaciones tradicionales de los departamentos de costes como los modelos de aprendizaje automático se enfrentan inevitablemente a la incertidumbre y a posibles fuentes de error. Sin embargo, si se dispone de suficientes datos de calidad, los modelos de aprendizaje automático pueden demostrar una precisión de predicción comparable e incluso superior a la de los juicios de expertos.

> Es probable que el aprendizaje automático se convierta en una herramienta complementaria fiable para realizar análisis que puedan: afinar los cálculos, sugerir escenarios alternativos e identificar dependencias ocultas entre los parámetros del proyecto. Estos modelos no pretenden ser universales, pero pronto ocuparán un lugar importante en los cálculos y los procesos de toma de decisiones. Las tecnologías de aprendizaje automático no excluirán la participación de ingenieros, estimadores y analistas, sino que, por el contrario, ampliarán sus capacidades ofreciendo un punto de vista adicional basado en datos históricos.

Si se integra adecuadamente en los procesos de negocio de las empresas de construcción, el aprendizaje automático tiene potencial para convertirse en un elemento importante del sistema de apoyo a la toma de decisiones de gestión, no como sustituto de los humanos, sino como extensión de su intuición profesional y su lógica ingenieril.

Próximos pasos: del almacenamiento al análisis y la previsión

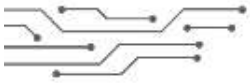
Los enfoques modernos del trabajo con datos están empezando a cambiar la forma de tomar decisiones en el sector de la construcción. Pasar de las evaluaciones intuitivas al análisis objetivo de los datos no sólo mejora la precisión, sino que también abre nuevas oportunidades para optimizar los procesos. Para resumir esta parte, merece la pena destacar los principales pasos prácticos que le ayudarán a aplicar los métodos comentados en sus tareas cotidianas:

- Creación de una infraestructura de almacenamiento sostenible
  - Intentar combinar documentos y datos de proyectos dispares en un único modelo tabular, agregando la información clave en un único marco de datos para su posterior análisis.
  - Utilizar formatos de almacenamiento de datos eficientes (por ejemplo, formatos columnares como Apache Parquet en lugar de CSV o XLSX), especialmente para aquellos conjuntos que podrían utilizarse para entrenar modelos de aprendizaje automático en el futuro.

- ☐ Establecer un sistema de versiones de los datos para realizar un seguimiento de los cambios a lo largo del proyecto.

- Aplicación de herramientas de análisis y automatización
  - ☐ Empezar a analizar los datos históricos de los proyectos -por documentación, modelos, estimaciones- para identificar patrones, tendencias y anomalías.
  - ☐ Dominar los procesos ETL (Extraer, Transformar, Cargar) para cargar y preparar los datos automáticamente.
  - ☐ Aprenda a visualizar métricas clave utilizando varias bibliotecas de visualización gratuitas de Python.
  - ☐ Empezar a aplicar métodos estadísticos y de muestreo aleatorio para obtener resultados analíticos representativos y reproducibles.
- Mayor madurez en el trabajo con datos
  - ☐ Aprenda algunos algoritmos básicos de aprendizaje automático con ejemplos sencillos y directos como el conjunto de datos Titanic.
  - ☐ Analice los procesos actuales e identifique dónde puede pasar de una lógica rígida de causa y efecto a métodos estadísticos de previsión y estimación.
  - ☐ Empezar a tratar los datos como un activo estratégico y no como un subproducto: construir procesos de toma de decisiones en torno a modelos de datos y no en torno a soluciones de software específicas.

Las empresas constructoras que se han dado cuenta del valor de los datos están entrando en una nueva fase de desarrollo en la que la ventaja competitiva no viene determinada por la cantidad de recursos, sino por la velocidad de la toma de decisiones basada en la analítica.



MÁXIMA COMODIDAD CON LA VERSIÓN IMPRESA

Tiene en sus manos una versión digital gratuita de **Data-Driven Construction**. Para trabajar con mayor comodidad y acceder rápidamente a los materiales, le recomendamos que preste atención a la **edición impresa**:

- **Siempre a mano**: el libro en formato impreso se convertirá en una herramienta de trabajo fiable, que le permitirá encontrar y utilizar rápidamente las visualizaciones y diagramas adecuados en cualquier situación de trabajo.
- **Ilustraciones de alta calidad**: todas las imágenes y gráficos de la edición impresa se presentan con la máxima calidad.
- **Acceso rápido a la información**: navegación sencilla, posibilidad de tomar notas, marcar favoritos y trabajar con el libro en cualquier lugar.

Al adquirir la versión impresa completa del libro, obtendrá una cómoda herramienta para trabajar con la información de forma cómoda y eficaz: la posibilidad de utilizar rápidamente materiales visuales en las tareas cotidianas, encontrar rápidamente los esquemas necesarios y tomar notas. Además, su compra apoya la difusión del conocimiento abierto.

Solicite la versión impresa del libro en: datadrivenconstruction.io/books

X PARTE

LA INDUSTRIA DE LA CONSTRUCCIÓN EN LA ERA DIGITAL. OPORTUNIDADES Y RETOS

La décima parte final es una mirada exhaustiva al futuro del sector de la construcción en la era de la transformación digital. Analiza el paso del análisis causal al trabajo con correlaciones de big data. Se establecen paralelismos entre la evolución de las bellas artes y el desarrollo del trabajo con datos en la construcción, demostrando cómo el sector está pasando de un control detallado a una comprensión holística de los procesos. Se explora el concepto de "uberización " del sector de la construcción, en el que la transparencia de los datos y los cálculos automatizados pueden cambiar radicalmente los modelos de negocio tradicionales, eliminando la necesidad de intermediarios y reduciendo las oportunidades de especulación. Las cuestiones pendientes, como la clasificación universal de los elementos de, se analizan en detalle, dando tiempo a las empresas de la construcción para adaptarse al nuevo entorno. La parte concluye con recomendaciones específicas para configurar una estrategia de transformación digital que incluya el análisis de las vulnerabilidades y la ampliación de los servicios para seguir siendo competitivos en un sector cambiante.

CAPÍTULO 10.1.
ESTRATEGIAS DE SUPERVIVENCIA: CREAR VENTAJAS COMPETITIVAS

Correlaciones en lugar de cálculos: el futuro de la analítica de la construcción

Debido a la rápida digitalización de la información (Fig, 1.1-5), la construcción moderna está experimentando una transformación fundamental en la que los datos se están convirtiendo no sólo en una herramienta, sino en un activo estratégico que puede cambiar radicalmente los enfoques tradicionales de la gestión de proyectos y empresas.

Durante miles de años, las actividades de construcción se han basado en métodos deterministas: cálculos precisos, detalles y un control estricto de los parámetros. En los primeros siglos de nuestra era, los ingenieros romanos aplicaban principios matemáticos a la construcción de acueductos y puentes. En la Edad Media, los arquitectos se esforzaban por conseguir las proporciones ideales de las catedrales góticas, y en la industrialización del siglo XX se formaron sistemas de normas y reglamentos estandarizados, que se convirtieron en la base de la construcción en masa

Hoy en día, el vector de desarrollo está pasando de la búsqueda de relaciones estrictamente de causa y efecto al análisis probabilístico, la búsqueda de correlaciones y patrones ocultos. La industria está entrando en una nueva fase: los datos se están convirtiendo en un recurso clave, y la analítica basada en ellos está sustituyendo a los enfoques intuitivos y optimizados localmente.

Fig. 10.1-1 Potencial oculto de los datos de construcción: los cálculos existentes en la empresa son sólo la punta del iceberg que la dirección puede analizar.

El sistema de información de una empresa es como un iceberg (Fig. 10.1-1): sólo una pequeña parte del potencial de los datos es visible para la dirección de la empresa, mientras que el valor principal se oculta en las profundidades. Es importante evaluar los datos no sólo por su uso actual, sino también por las oportunidades que desbloquearán en el futuro. Son las empresas que aprenden a extraer patrones ocultos y a crear nuevas perspectivas a partir de los datos las que crearán una ventaja competitiva sostenible

Encontrar patrones ocultos y dar sentido a los datos no es sólo trabajar con números, sino un proceso creativo que requiere un pensamiento abstracto y la capacidad de ver la imagen completa detrás de elementos dispares. En este sentido, la evolución del trabajo con datos puede compararse con la evolución de las artes visuales (Fig, 10.1-2).

El desarrollo de la construcción es notablemente similar al progreso de las bellas artes. En ambos casos, la humanidad ha pasado de métodos primitivos a sofisticadas tecnologías de visualización y análisis. En la prehistoria, la gente utilizaba dibujos rupestres y herramientas primitivas para resolver las tareas cotidianas. Durante la Edad Media y el Renacimiento, el nivel de sofisticación de la arquitectura y el arte aumentó espectacularmente. A principios de la Edad Media, las herramientas de construcción habían evolucionado desde un simple hacha hasta amplios juegos de herramientas que simbolizaban el crecimiento de los conocimientos técnicos.

La Era del Realismo supuso la primera revolución en las artes visuales: los artistas aprendieron a reproducir hasta el más mínimo detalle, logrando la máxima verosimilitud. En la construcción, la contrapartida de este periodo fueron las precisas técnicas de ingeniería, los dibujos detallados y los cálculos estrictamente regulados que se convirtieron en la base de la práctica del diseño durante siglos.

Más tarde, el Impresionismo cambió la percepción misma de la realidad artística: en lugar de transmitir literalmente la forma, los artistas empezaron a captar el estado de ánimo, la luz y la dinámica, con el objetivo de reflejar una impresión general en lugar de una precisión absoluta. Del mismo modo, el aprendizaje automático en el análisis de edificios se está alejando de los modelos lógicos rígidos para acercarse al reconocimiento de patrones y patrones probabilísticos que permiten "ver" dependencias ocultas en los datos que son inaccesibles en el análisis clásico. Este enfoque se hace eco de las ideas de minimalismo y funcionalidad de la Bauhaus, donde el significado (la función) es más importante que la forma. La Bauhaus pretendía eliminar lo superfluo, abandonar la ornamentación en aras de la claridad, el utilitarismo y el atractivo de masas. Las cosas tenían que ser comprensibles y útiles, sin excesos: la estética nacía de la lógica del diseño y la finalidad.

Con la llegada de la fotografía a finales del siglo XIX, el arte adquirió una nueva herramienta para captar la realidad con una precisión sin precedentes y dio un vuelco a las actitudes hacia las artes visuales. Del mismo modo, en la construcción, la revolución industrial del siglo XXI está dando lugar al uso de tecnologías robóticas, láseres, IoT, RFID y conceptos como el de construcción conectada, en el que la recopilación de parámetros individuales ha evolucionado hacia una captura inteligente escalable de toda la realidad de la obra.

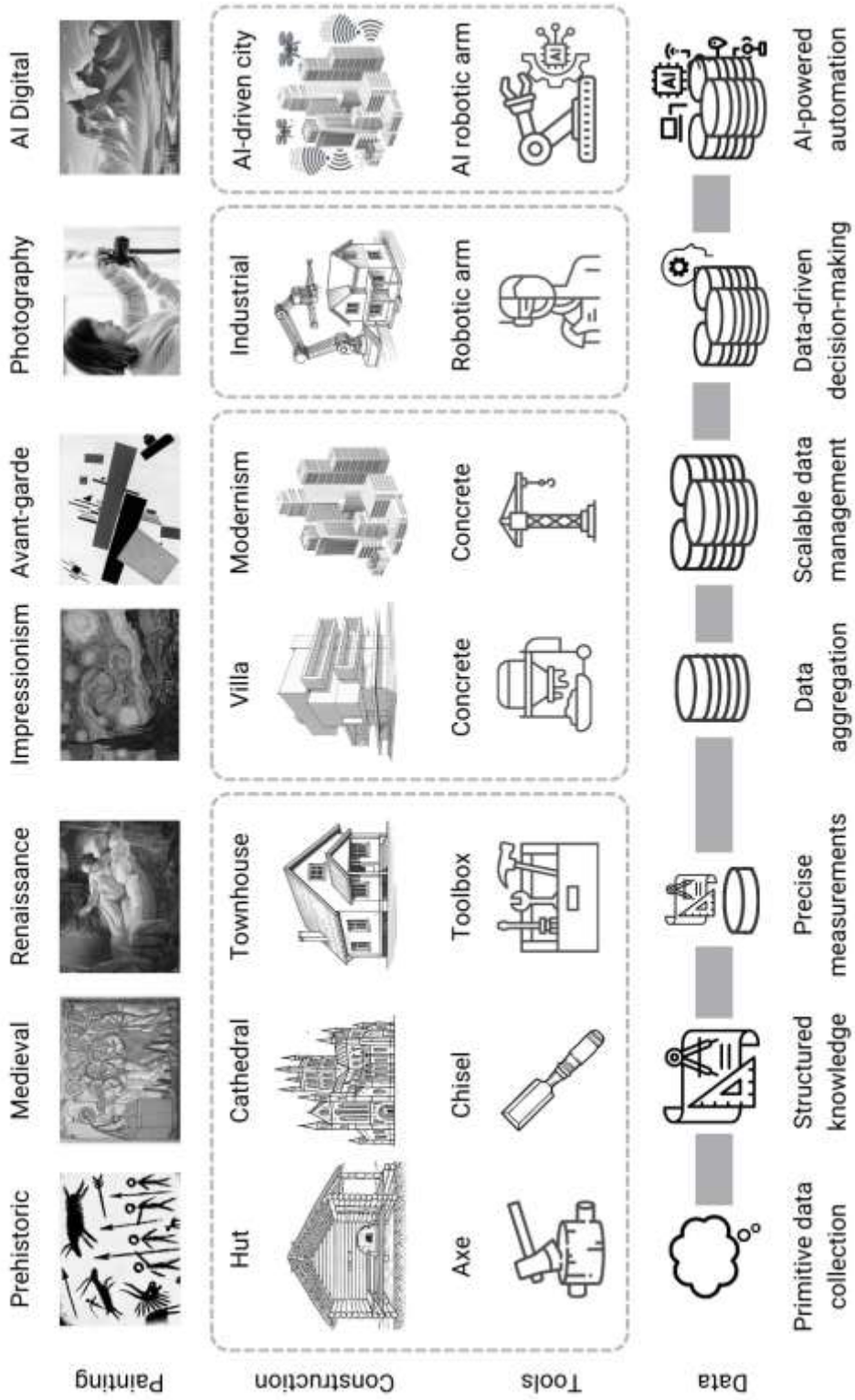


Fig. 10.1-2 Era de las Evoluciones de las Bellas Artes es coherente con la evolución enfoques del trabajo con datos en el sector de la construcción.

Hoy en día, al igual que las artes visuales están experimentando un replanteamiento con la llegada de las herramientas de IA y LLM, la industria de la construcción está experimentando otro salto cuántico: los sistemas inteligentes impulsados por la inteligencia artificial (AI), los chats LLM permiten predecir, optimizar y generar soluciones con una intervención humana mínima.

El papel de los datos en el diseño y la gestión ha cambiado radicalmente. Mientras que antes el conocimiento se transmitía verbalmente y de forma empírica -al igual que hasta el siglo XIX la realidad se plasmaba en cuadros pintados a mano-, hoy en día la atención se centra en la captura digital completa de la "imagen" de la construcción. Con la ayuda de algoritmos de aprendizaje automático, esta imagen digital se transforma en una representación impresionista de la realidad de la construcción, no una réplica exacta, sino una comprensión generalizada y probabilística de los procesos.

> Nos acercamos rápidamente a una era en la que los procesos de diseño, construcción y explotación de edificios no sólo se verán aumentados, sino que estarán dirigidos en gran medida por sistemas de inteligencia artificial. Del mismo modo que el arte digital moderno se crea sin pincel -mediante indicaciones textuales y modelos generativos-, las soluciones arquitectónicas y de ingeniería del futuro se conformarán a partir de consultas y parámetros clave establecidos por el usuario.

En el siglo XXI, el acceso a los datos, su interpretación y la calidad de los análisis se están volviendo indispensables para el éxito de un proyecto. Y el valor de los datos no viene determinado por su volumen, sino por la capacidad de los especialistas para analizarlos, verificarlos y convertirlos en acción.

Enfoque basado en datos en la construcción: un nuevo nivel de infraestructura

En la historia de la humanidad, cada salto tecnológico de este tipo ha traído consigo cambios fundamentales en la economía y la sociedad. Hoy asistimos a una nueva ola de transformación comparable en escala a la revolución industrial del siglo XIX. Sin embargo, mientras que hace cien años el principal motor del cambio eran las fuerzas mecánicas y las tecnologías energéticas, ahora son los datos y la inteligencia artificial.

Los agentes de aprendizaje automático, LLM y AI cambian la esencia misma de las aplicaciones, haciendo innecesarias las pilas de software tradicionales (analizadas en la segunda parte del libro) (Fig. 2.2-3). Toda la lógica de datos se centra en los agentes de IA en lugar de en reglas de negocio codificadas (Fig. 2.2-4).

> En la era de los datos, la visión tradicional de las aplicaciones se está transformando radicalmente. Avanzamos hacia un modelo en el que los sistemas empresariales voluminosos y modulares darán paso inevitablemente a soluciones abiertas, ligeras y personalizadas.

> *En el futuro, sólo permanecerá la estructura de datos subyacente, y toda interacción con ella se realizará a través de agentes que trabajen directamente con la base de datos. Realmente creo que toda la pila de aplicaciones desaparecerá, porque sencillamente no hay necesidad de ella cuando la inteligencia artificial interactúe directamente con la base de datos subyacente. He pasado toda mi carrera trabajando en SaaS - construyendo empresas, trabajando en ellas, y para ser honesto, probablemente no lanzaría un nuevo negocio SaaS ahora mismo. Y probablemente no invertiría en empresas SaaS ahora mismo. La situación es demasiado incierta. Eso no quiere decir que no vaya a haber empresas de software en el futuro, sólo que tendrán un aspecto muy diferente. Los sistemas del futuro serán bases de datos con lógica empresarial incorporada a agentes [AI]. Estos agentes trabajarán con múltiples repositorios de datos al mismo tiempo, sin limitarse a una única base de datos. Toda la lógica se trasladará a la capa de IA [46].*
>
> - Matthew Berman, CEO de Forward Future

La diferencia clave del nuevo paradigma es la minimización del lastre tecnológico. En lugar de monumentales sistemas de software complejos y cerrados, obtendremos módulos flexibles, abiertos y rápidamente personalizables que literalmente "vivirán" dentro del flujo de datos (Fig. 7.4-1 - Apache Airflow, NiFi). La arquitectura de la gestión de procesos del futuro prevé el uso de microaplicaciones, herramientas compactas y específicas, fundamentalmente diferentes de los sistemas masivos y cerrados ERP, PMIS, CDE, CAFM. Los nuevos agentes serán lo más adaptables, integrados y específicos del negocio que sea posible (por ejemplo, Low-Code/No-Code Fig. 7.4-6).

> *Toda la lógica de negocio irá a estos agentes [AI], y estos agentes realizarán operaciones CRUD [Crear, Leer, Actualizar y Eliminar] en múltiples repositorios, lo que significa que no distinguirán qué backend se está utilizando. Actualizarán múltiples bases de datos, y toda la lógica acabará en el llamado nivel de IA. Y una vez que la capa de IA sea donde esté toda la lógica, la gente empezará a sustituir los backends. Ya estamos viendo un porcentaje bastante alto de victorias de mercado en el uso de backends y agentes de Dynamics, y vamos a movernos agresivamente en esa dirección, tratando de unirlo todo. Ya sea en el servicio de atención al cliente o en otras áreas, por ejemplo, no solo en CRM, sino también en nuestras soluciones de finanzas y operaciones. Porque la gente quiere más aplicaciones empresariales impulsadas por IA en las que la capa lógica pueda ser impulsada por IA y agentes de IA. [...]. Una de las cosas más emocionantes para mí es Excel con Python, que es comparable a GitHub con Copilot. Es decir, lo que hemos hecho es: ahora que tienes Excel, deberías abrirlo, ejecutar Copilot y empezar a jugar con él. Ya no se trata sólo de entender los números disponibles: hará un plan por sí mismo. Al igual que el espacio de trabajo de GitHub Copilot crea un plan y luego lo ejecuta, es como un analista de datos, que utiliza Excel como herramienta para visualizar filas y columnas para el análisis. Así que Copilot utiliza Excel como herramienta con todas sus capacidades porque puede generar datos y tiene un intérprete de Python.*
>
> - Satya Nadella, consejero delegado de Microsoft, entrevista con el canal BG2 en diciembre de 2024. [28]

La transformación que estamos viendo en la lógica de las aplicaciones ofimáticas -pasar de sistemas modulares y cerrados a agentes de IA que trabajan directamente con datos abiertos- es sólo una parte de un proceso mucho más amplio. No se trata solo de cambiar las interfaces o la arquitectura del software: los cambios afectarán a los principios fundamentales de la organización del trabajo, la toma de decisiones y la gestión empresarial. En la construcción, esto conducirá a una lógica basada en los datos, en la que éstos se convertirán en la pieza central de los procesos, desde el diseño a la gestión de recursos y la supervisión de la construcción.

La oficina digital de nueva generación: cómo la IA está cambiando el espacio de trabajo

Hace casi un siglo, la humanidad ya experimentaba una revolución tecnológica similar. La transición de las máquinas de vapor a los motores eléctricos duró más de cuatro décadas, pero acabó catalizando un crecimiento de la productividad sin precedentes, gracias sobre todo a la descentralización de la capacidad energética y a la flexibilidad de las nuevas soluciones. Este cambio no sólo modificó el curso de la historia, desplazando al grueso de la población de las zonas rurales a las urbanas, sino que también sentó las bases de la economía moderna. La historia de la tecnología es un viaje del trabajo físico a la automatización y los sistemas inteligentes. Al igual que el tractor sustituyó a decenas de labradores de la tierra, la tecnología digital moderna está reemplazando a los métodos tradicionales de gestión de la construcción basados en oficinas (Fig, 10.1-3). A principios del siglo XX, la mayor parte de la población mundial trabajaba la tierra a

mano, hasta que en los años 30 comenzó la mecanización del trabajo con máquinas y tractores.

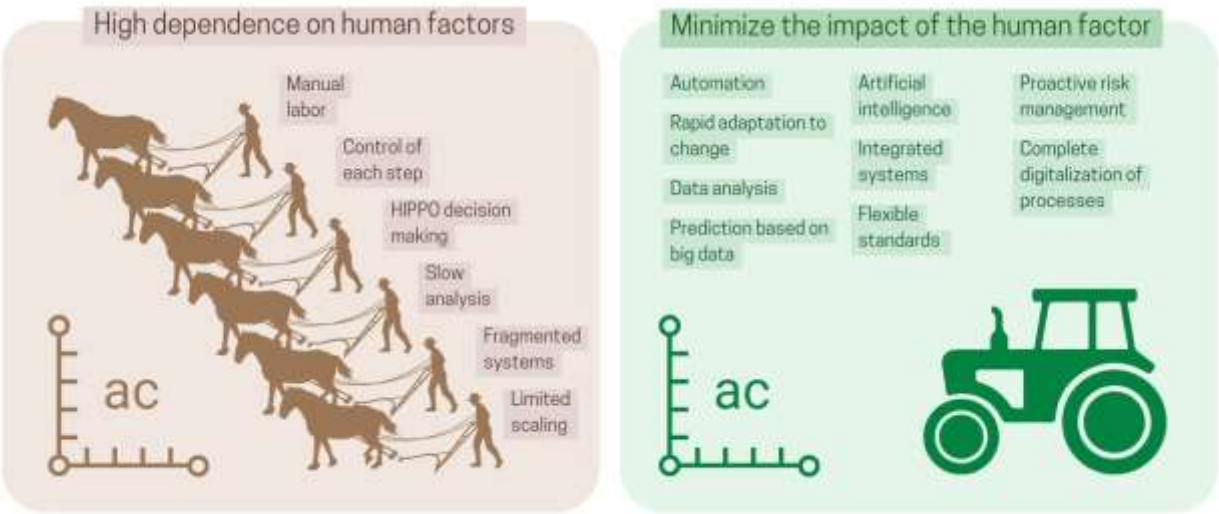


Fig. 10.1-3 Al igual que el tractor sustituyó a decenas de personas a principios del siglo XX, el aprendizaje automático sustituirá a los métodos tradicionales de gestión empresarial y de proyectos en el siglo XXI.

Al igual que hace cien años la humanidad pasó de cultivar parcelas individuales con herramientas primitivas a la agricultura a gran escala con el uso de maquinaria, hoy estamos haciendo la transición de procesar "silos" dispares de información a trabajar con matrices de datos con la ayuda de potentes "tractores": ETL-pipeline y algoritmos de inteligencia artificial.

Estamos a punto de dar un salto similar, pero en el plano digital: de la gestión empresarial tradicional y manual a los modelos basados en datos.

El camino hacia una arquitectura plenamente basada en datos requerirá tiempo, inversión y esfuerzos organizativos. Pero este camino abre la vía no solo a una mejora gradual, sino a un salto cualitativo hacia una mayor eficiencia, transparencia y gestionabilidad de los procesos de construcción. Todo ello está supeditado a la introducción sistemática de herramientas digitales y al abandono de prácticas empresariales obsoletas.

> La parametrización de tareas, el ETL, el LLM, los componentes de IoT, la RFID, la tokenización, los **macrodatos** y el aprendizaje automático transformarán la construcción tradicional en **una construcción basada en datos**, en la que cada detalle del proyecto y del negocio de la construcción se controlará y optimizará mediante datos.

Antes se necesitaban miles de horas de trabajo para analizar la información. Ahora estas tareas las realizan algoritmos y LLM que convierten conjuntos de datos dispares en fuentes estratégicas mediante indicaciones. En el mundo de la tecnología está ocurriendo lo mismo que en la agricultura: estamos pasando de la azada a la agroindustria automatizada. Lo mismo ocurre con el trabajo de oficina en la construcción: de los archivos Excel y el resumen manual se está pasando a un sistema inteligente en el que los datos se recopilan, se limpian, se estructuran y se convierten en ideas.

Ya hoy, las empresas deben empezar a "cultivar" los campos de información mediante la recogida de datos de calidad y la estructuración de la información, y "abonarlos" con herramientas de limpieza y normalización, para luego "cosecharlos" en forma de análisis predictivo y soluciones automatizadas. Si un agricultor moderno puede sustituir con una máquina a un centenar de labradores de la tierra, los algoritmos inteligentes podrán eliminar la rutina de los empleados y transferirlos al papel de gestores estratégicos de los flujos de información.

Sin embargo, es importante comprender que la creación de una organización verdaderamente basada en datos no es un proceso rápido. Se trata de una dirección estratégica a largo plazo, similar a la creación de un nuevo emplazamiento para cultivar un nuevo bosque (Fig. 1.2-5) de sistemas, donde cada "árbol" de este ecosistema es un proceso, competencia o herramienta independiente que tarda tiempo en crecer y desarrollarse. Como en el caso de un bosque real, el éxito no sólo depende de la calidad del material de plantación (tecnología), sino también del suelo (cultura empresarial), el clima (entorno empresarial) y los cuidados (enfoque de sistemas).

Las empresas ya no podrán confiar únicamente en soluciones cerradas y listas para usar. A diferencia de etapas anteriores del desarrollo tecnológico, es poco probable que la transición actual -hacia el acceso abierto a los datos, la inteligencia artificial y el código abierto- cuente con el apoyo de los grandes proveedores, ya que amenaza directamente sus modelos de negocio establecidos y sus principales fuentes de ingresos.

Como muestra el estudio de la Harvard Business School [40], del que ya se ha hablado en el capítulo dedicado a la cuarta y quinta revoluciones tecnológicas, el coste de crear desde cero las soluciones Open Source más utilizadas por todas las empresas sería de unos 4,15 billones de dólares. Sin embargo, si imaginamos que cada empresa desarrollara sus propias alternativas sin acceso a las herramientas Open Source existentes, que es lo que ha estado ocurriendo durante las últimas décadas, los costes empresariales totales podrían alcanzar la friolera de 8,8 billones de dólares - esto es el precio de la demanda irracional en el que se puede valorar el mercado del software.

El progreso tecnológico conducirá inevitablemente a un replanteamiento de los modelos de negocio establecidos. Mientras que antes las empresas podían ganar dinero con procesos complejos y opacos y datos cerrados, con el desarrollo de la IA y la analítica este enfoque es cada vez menos viable.

> Como consecuencia de la democratización del acceso a los datos y las herramientas, el mercado tradicional de venta de software puede reducirse considerablemente. Sin embargo, al mismo tiempo, crecerá un nuevo mercado: el de la experiencia digital, la personalización, la integración y el diseño de soluciones. Aquí, el valor no vendrá de la venta de licencias, sino de la capacidad de construir procesos digitales flexibles, abiertos y adaptables. Del mismo modo que la electrificación y la llegada de los tractores han dado lugar a nuevas industrias, la aplicación de big data, AI y LLM abrirá horizontes completamente nuevos para las empresas del sector de la construcción, lo que requerirá no solo inversión tecnológica, sino también una profunda transformación de mentalidades, procesos y estructuras organizativas. Y aquellas empresas y profesionales que se den cuenta de ello y empiecen a actuar hoy serán los líderes del mañana.

En un mundo en el que los datos abiertos se están convirtiendo en un activo importante, la disponibilidad

de información cambiará las reglas del juego. Inversores, clientes y reguladores exigirán cada vez más transparencia, y los algoritmos de aprendizaje automático podrán identificar automáticamente discrepancias en estimaciones, plazos y costes. Esto crea las condiciones para una nueva etapa de transformación digital, que nos conduce gradualmente a la "uberización " del sector de la construcción.

Los datos abiertos y la uberización son una amenaza para el negocio actual de la construcción

La construcción se está convirtiendo en un proceso de gestión de la información. Cuanto más precisos, completos y de calidad sean los datos, más eficaces serán el diseño, los cálculos, la estimación de costes, la construcción y el funcionamiento de los edificios. En el futuro, el recurso clave no será la grúa, el hormigón y las barras de refuerzo, sino la capacidad de recopilar, analizar y utilizar la información.

> Los clientes de las empresas de construcción -inversores y clientes que financian la construcción- utilizarán inevitablemente en el futuro el valor de los datos abiertos y la analítica de datos históricos. Esto abrirá oportunidades para automatizar el cálculo de los plazos y costes de los proyectos, sin implicar a las empresas constructoras en cuestiones de cálculo de costes, lo que ayudará a controlar los costes y a identificar más rápidamente los costes redundantes.

Imaginemos una obra en la que escáneres láser, cuadricópteros y sistemas de fotogrametría recopilan datos precisos en tiempo real sobre los volúmenes de hormigón utilizados. Esta información se convierte automáticamente en sencillos modelos planos MESH con metadatos, evitando los engorrosos sistemas CAD (BIM), sin dependencias de complejos núcleos geométricos, ERP o PMIS. Estos datos recogidos en la obra se transfieren de forma centralizada a un único repositorio estructurado a disposición del cliente para su análisis independiente, en el que se cargan precios reales de diferentes talleres de construcción y parámetros que van desde la tasa de financiación crediticia hasta factores que cambian dinámicamente, como las condiciones meteorológicas, las cotizaciones bursátiles de los materiales de construcción, las tarifas logísticas y las fluctuaciones estadísticas estacionales de los precios de la mano de obra. En tales condiciones, cualquier discrepancia entre el diseño y los volúmenes reales de material se hace evidente al instante, lo que hace imposible manipular las estimaciones tanto en la fase de diseño como durante la entrega del proyecto. Como resultado, la transparencia del proceso de construcción no se consigue mediante un ejército de supervisores y gestores, sino a través de datos digitales objetivos, que minimizarán el factor humano y la posibilidad de especulación.

En el futuro, este tipo de trabajo de control de datos lo realizarán más bien los gestores de datos del lado del cliente (Fig. 1.2-4 Gestor CQMS). Esto es especialmente cierto en el caso de los cálculos y las estimaciones de proyectos: donde antes había todo un departamento de estimadores, mañana ya habrá herramientas de aprendizaje automático y de previsión que fijarán los límites de precios a los que se ajustarán las empresas de construcción.

> *Dada la naturaleza fragmentada de la industria [de la construcción], donde la mayoría de los sistemas y subsistemas son suministrados por PYME, la estrategia digital debe venir del cliente. Los clientes deben crear las condiciones y los mecanismos para desbloquear las capacidades digitales de la cadena de suministro [20].*
>
> - Andrew Davis y Giuliano Denicol, Accenture "Crear más valor mediante proyectos de capital"

Esta apertura y transparencia de los datos supone una amenaza para las empresas constructoras, acostumbradas a ganar dinero con procesos opacos e informes confusos, en los que la especulación y los costes ocultos pueden esconderse tras formatos complejos y cerrados y plataformas modulares de datos propietarios. Por lo tanto, es poco probable que las empresas de construcción, al igual que los proveedores que promueven soluciones de código abierto, estén interesadas en implantar plenamente los datos abiertos en sus procesos empresariales. Si los datos están disponibles y son fáciles de procesar para el cliente, pueden comprobarse automáticamente, lo que eliminará la posibilidad de sobreestimar volúmenes y manipular estimaciones.

> Según el informe del Foro Económico Mundial "Shaping the Future of Construction" (2016) [5], uno de los principales retos del sector sigue siendo el papel pasivo del cliente. Sin embargo, son los clientes quienes deben asumir una mayor responsabilidad en el resultado de los proyectos: desde la planificación temprana, pasando por la selección de modelos sostenibles de interacción, hasta la supervisión del rendimiento. Sin la participación activa de los propietarios de los proyectos, la transformación sistémica de la industria de la construcción es imposible.

La pérdida de control sobre el volumen y el cálculo de costes ya ha transformado otras industrias en los últimos 20 años, al permitir a los clientes acceder directamente, sin intermediarios, a. La digitalización y la transparencia de los datos han transformado muchos modelos de negocio tradicionales, como el de los taxistas con la aparición de Uber (Fig, 10.1-4), el de los hoteleros con la llegada de Airbnb y el de los minoristas y tiendas con el auge de Amazon, y el de los bancos con el auge de los neobancos y los ecosistemas fintech descentralizados, en los que el acceso directo a la información y automatización de los cálculos de tiempo y costes han reducido considerablemente el papel de los intermediarios.

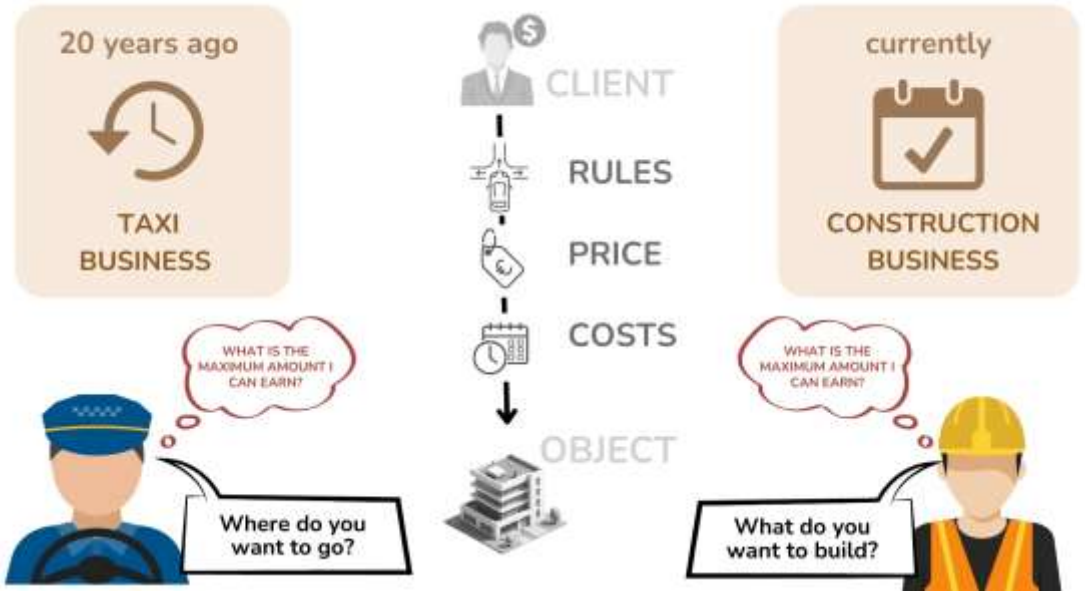


Fig. 10.1-4 El negocio de la construcción se enfrentará a la uberización a la que tuvieron que enfrentarse los taxistas, los hoteleros y los minoristas hace 10 años.

El proceso de democratización del acceso a los datos y a las herramientas para su tratamiento es inevitable y, con el tiempo, los datos abiertos sobre todos los componentes del proyecto se convertirán en un requisito de los clientes y en una nueva norma. Por lo tanto, la introducción de formatos abiertos y cálculos transparentes será promovida por los inversores, los clientes, los bancos y los fondos de capital riesgo (private equity), que son quienes acaban siendo los usuarios finales de las instalaciones construidas y luego las explotan durante décadas.

> Los grandes inversores, clientes y bancos ya exigen transparencia en el sector de la construcción. Según el estudio de Accenture "Creating more value through capital projects" (2020) [20], los datos transparentes y fiables se están convirtiendo en un factor decisivo para las decisiones de inversión en la construcción. Como señalan los expertos, una gestión fiable y eficaz de los proyectos es imposible sin transparencia, especialmente en tiempos de crisis. Además, los propietarios de activos y los contratistas se inclinan cada vez más por contratos que incentivan el intercambio de datos y el análisis colaborativo, lo que refleja las crecientes demandas de los inversores, los bancos y los reguladores en materia de rendición de cuentas y transparencia.

> El movimiento del inversor, el cliente, desde la idea hasta el edificio terminado, en el futuro será similar a viajar con el piloto automático: sin un conductor en forma de empresa constructora, promete independizarse de la especulación y la incertidumbre.

La era de los datos abiertos y la automatización cambiará inevitablemente el negocio de la construcción, como ya lo ha hecho en la banca, el comercio, la agricultura y la logística. En estos sectores, el papel de los intermediarios y las formas tradicionales de hacer negocios están dando paso a la automatización y la robotización, sin dejar lugar a márgenes injustificados ni a la especulación.

Los datos y procesos de todas las actividades económicas humanas no difieren de los que tienen que manejar los profesionales del sector de la construcción. A largo plazo, las empresas constructoras, que hoy dominan el mercado fijando las normas de precio y calidad del servicio, pueden perder su papel de intermediario clave entre el cliente y su proyecto de construcción.

Problemas sin resolver de la uberización como última oportunidad de aprovechar el tiempo para la transformación

Pero volvamos a las realidades del sector de la construcción. Aunque en algunos sectores de la economía están surgiendo coches que se conducen solos, sistemas financieros descentralizados y soluciones basadas en la inteligencia artificial, una proporción significativa de las empresas de la construcción siguen siendo organizaciones basadas en el papel, en las que las decisiones clave se toman más en función de la intuición y la experiencia de cada especialista.

En este paradigma, una empresa de construcción moderna puede compararse con una empresa de taxis de hace 20 años que controla los recursos, las rutas y los plazos de entrega, y es responsable de los plazos y los costes del "viaje": desde la idea del proyecto (proceso de logística e instalación) hasta la entrega del proyecto. Al igual que el GPS (en la construcción IoT, RFID) y los algoritmos de aprendizaje automático en los cálculos de tiempo/coste transformaron en su día el campo del transporte, los datos, los algoritmos y los agentes de IA tienen el potencial de transformar la gestión de la construcción: de evaluaciones intuitivas a modelos predictivos y guiados. En los últimos 20 años, muchos sectores -finanzas, agricultura, comercio minorista y logística- han visto cómo desaparecía gradualmente la capacidad de especular gracias a la opacidad de los datos. Los precios, los costes de entrega o las transacciones financieras se calculan de forma automática y estadísticamente sólida, en cuestión de segundos en plataformas digitales.

De cara al futuro, las empresas constructoras deben reconocer que la democratización del acceso a los datos y a las herramientas para analizarlos alterará el enfoque tradicional de estimación de costes y plazos de los proyectos y eliminará la posibilidad de especular con datos opacos sobre volúmenes y precios.

> Al igual que conducir por una carretera regulada sin la intervención del conductor, los procesos de construcción del futuro se parecerán cada vez más a un sistema "Uberizado", con una estimación automatizada del tiempo y los costes, un enrutamiento transparente de las tareas y una dependencia mínima de los factores humanos. Esto cambiará la naturaleza misma del "viaje" desde la idea hasta la realización, haciéndolo más predecible, gestionable y basado en datos.

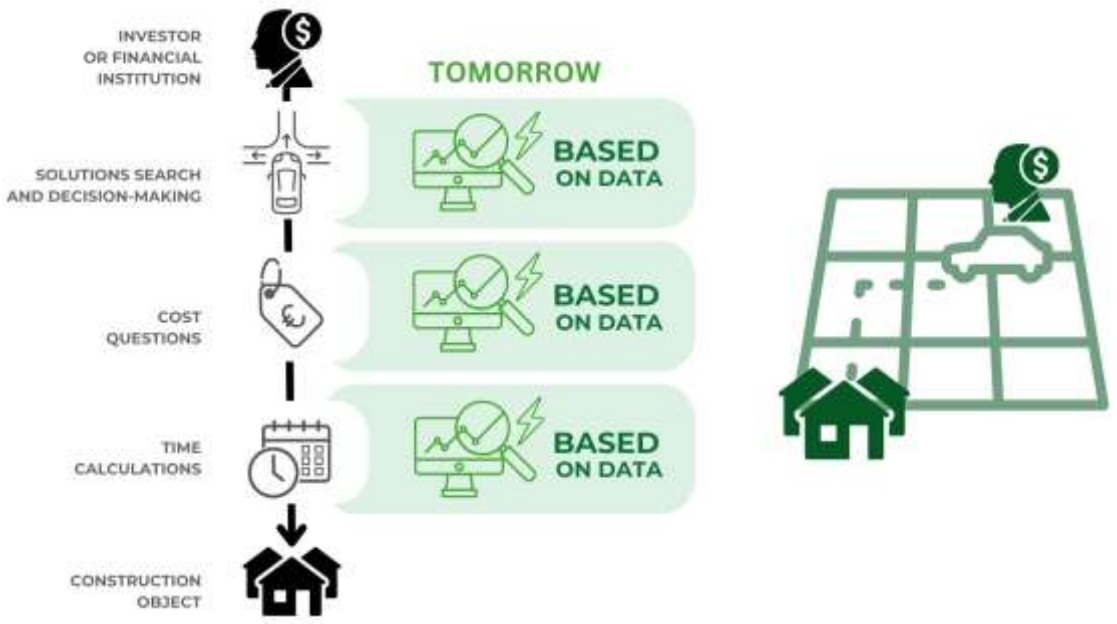


Fig. 10.1-5 El coste y el tiempo de "viaje" durante la construcción se determinarán utilizando el aprendizaje automático y herramientas estadísticas.

Con la introducción gradual de nuevas normativas y requisitos en casi todos los países del mundo, que obligan a transferir modelos CAD- (BIM)- a los clientes o bancos que financian proyectos de construcción, el cliente y el cliente tienen la oportunidad de garantizar de forma independiente la transparencia de los cálculos de costes y alcance. Esto es especialmente relevante para los grandes clientes e inversores, que disponen de competencias y herramientas suficientes para analizar rápidamente los volúmenes y controlar los precios del mercado. Para las empresas que ejecutan proyectos estándar a gran escala -tiendas, edificios de oficinas, complejos residenciales- estas prácticas se están convirtiendo en habituales.

A medida que el contenido informativo de los modelos se hace más completo y estandarizado, la posibilidad de manipulación y especulación prácticamente desaparece. La transformación digital está cambiando gradualmente las reglas del juego en el sector de la construcción, y las empresas que no se adapten a estos cambios pueden enfrentarse a graves problemas.

> El aumento de la competencia, la disrupción tecnológica y la reducción de los márgenes pueden afectar a la sostenibilidad de las empresas. Con una liquidez limitada, cada vez más actores del sector recurren a la automatización, el análisis y las tecnologías de datos como forma de mejorar la eficiencia y la transparencia de los procesos. Estas herramientas se están convirtiendo en un recurso importante para seguir siendo competitivos en un entorno económico cambiante.

Quizá no debamos esperar a que las circunstancias externas obliguen a actuar con urgencia: es mucho más eficaz empezar a prepararse hoy mismo reforzando las competencias digitales, implantando soluciones modernas y creando una cultura centrada en los datos.

Una de las últimas barreras tecnológicas clave para la transformación digital a gran escala del sector de la

construcción, que afectará a todas las empresas en los próximos años, es el problema de la clasificación automática de elementos de los proyectos de construcción.

> Sin una clasificación fiable, precisa y escalable, es imposible crear la base para un análisis completo, la automatización de procesos y la gestión del ciclo de vida mediante IA y modelos predictivos. Mientras la clasificación de objetos siga dependiendo de la interpretación manual por parte de especialistas experimentados -capataces, diseñadores, estimadores-, el sector de la construcción aún tiene una ventana de oportunidad. Este tiempo puede aprovecharse para prepararse para los cambios inevitables: mayores exigencias de transparencia, democratización de herramientas y datos, y aparición de sistemas de clasificación automática que cambiarán radicalmente las reglas del juego.

La tarea de clasificación automática de los elementos del mundo de la construcción es comparable en su complejidad al reconocimiento de objetos en los sistemas de conducción no tripulados, que constituye uno de los principales retos. Imaginemos un coche no tripulado que se desplaza del punto A al punto B (Fig. 10.1-5). Los actuales sistemas de conducción automática se ven atascados por el problema de la clasificación de los objetos reconocidos por los lidares y las cámaras. No basta con que un coche "vea" un obstáculo o un punto de referencia, sino que debe ser capaz de reconocer si se trata de un peatón, una señal de tráfico o un cubo de basura.

Todo el sector de la construcción se enfrenta a un reto fundamental similar. Los elementos del proyecto -como ventanas, puertas o columnas- pueden registrarse en la documentación, representarse en modelos CAD, fotografiarse en la obra o reconocerse en nubes de puntos procedentes de escaneado láser. Sin embargo, su reconocimiento visual o geométrico aproximado no basta para construir un sistema de gestión de proyectos verdaderamente automatizado. Es necesario garantizar que cada elemento se categoriza de forma precisa y coherente en un tipo que será identificable de forma única en todos los procesos posteriores, desde los presupuestos y las especificaciones hasta la logística, el control de inventario y, lo que es más importante, las operaciones (Fig. 4.2-6).

Es en esta fase -la transición del reconocimiento a una clasificación significativa- donde surge uno de los principales obstáculos. Aunque los sistemas digitales sean técnicamente capaces de distinguir e identificar objetos en modelos y en la obra, la principal dificultad reside en la definición correcta y contextualmente estable del tipo de elemento para distintos entornos de software.. Por ejemplo, una puerta puede ser etiquetada por el diseñador en un modelo CAD como elemento de la categoría "puerta", pero al transferirla a un sistema ERP o PMIS puede estar mal tipificada, ya sea por un error del diseñador o por incoherencias entre los sistemas. Además, a menudo el elemento pierde algunos atributos importantes o desaparece de la contabilidad del sistema durante las exportaciones e importaciones de datos. Esto provoca una brecha en el flujo de datos y socava el principio de digitalización de extremo a extremo de los procesos de construcción. Se crea así una brecha crítica entre el significado semántico "visible" y el "comprensible", que socava la integridad de los datos y complica considerablemente la automatización de los procesos a lo largo de todo el ciclo de vida de un proyecto de construcción.

Resolver el problema de la clasificación universal de elementos de construcción utilizando tecnologías de big data y aprendizaje automático (Fig. 10.1-6) será un catalizador para la transformación de todo el sector, y quizá un descubrimiento inesperado para muchas empresas de construcción. Un sistema de clasificación

unificado y con capacidad de aprendizaje será la base para una analítica escalable, la gestión digital y la adopción de la IA en las prácticas diarias de las organizaciones de la construcción.

NVIDIA y otros líderes tecnológicos ya están proporcionando soluciones en otros sectores capaces de categorizar y estructurar automáticamente ingentes cantidades de información textual y visual.

> El modelo NeMo Curator de NVIDIA [161], por ejemplo, se especializa en la clasificación y categorización automática de datos en categorías predefinidas, desempeñando un papel clave en la optimización de los canales de procesamiento de información para el ajuste fino y el preentrenamiento de modelos generativos de IA. La plataforma Cosmos se entrena con escenas de vídeo del mundo real y 3D [162], lo que proporciona una base para los sistemas autónomos y los gemelos digitales que ya se están construyendo en el ecosistema NVIDIA. NVIDIA Omniverse, que en 2025 se ha convertido en la herramienta líder para trabajar con el formato USD, una descripción universal de escenas que podría acabar sustituyendo al formato IFC en los procesos de transferencia de información de diseño. Junto con Isaac Sim, un simulador de procesos robóticos [163], soluciones como NeMo Curator, Cosmos y Omniverse representan un nuevo nivel de automatización: desde la limpieza y el filtrado de datos hasta la generación de conjuntos de entrenamiento, el modelado de propiedades de objetos y el entrenamiento de robots en la obra. Y todas estas herramientas son gratuitas y de código abierto, lo que reduce significativamente las barreras para su adopción en las prácticas de ingeniería y construcción.

La clasificación automática de datos a nivel de tablas estructuradas no es tan difícil como puede parecer a primera vista. Como hemos mostrado en el capítulo anterior (Fig. 9.1-10), es posible compensar valores de clase ausentes o incorrectos basándose en parámetros similares de otros elementos si se dispone de datos históricos acumulados. Si elementos con características similares ya se han clasificado correctamente en varios proyectos finalizados, el sistema puede sugerir un valor adecuado para un elemento nuevo o incompleto con una alta probabilidad (Fig. 10.1-6). Este tipo de lógica, basada en valores promediados y en el análisis del contexto, puede resultar especialmente eficaz en el tratamiento masivo de datos tabulares procedentes de presupuestos, pliegos de condiciones o modelos CAD.

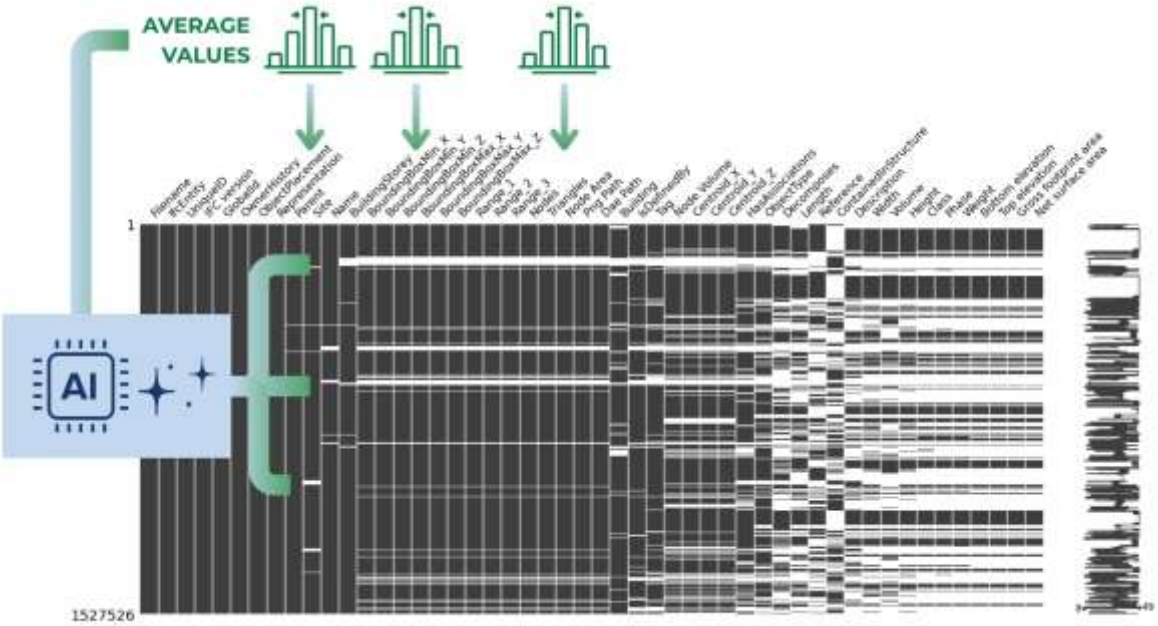


Fig. 10.1-6 El aprendizaje automático puede ayudarle a encontrar automáticamente valores medios para los parámetros de tabla sin rellenar (casillas blancas) basándose en proyectos anteriores.

> Con el telón de fondo de un progreso tan rápido en el aprendizaje automático, está claro: en 2025, es ingenuo creer que el problema de la clasificación automática de elementos de construcción seguirá sin resolverse durante mucho tiempo. Sí, los algoritmos actuales aún no han alcanzado su plena madurez, especialmente con datos incompletos o heterogéneos, pero la ventana de oportunidad para la adaptación se está cerrando rápidamente.

Las empresas que ya están invirtiendo en recopilar, limpiar y organizar sus datos, así como en adoptar herramientas de automatización ETL, tendrán ventaja. El resto corre el riesgo de quedarse rezagado, al igual que en su día las empresas no consiguieron superar los retos de la transformación digital en los sectores del transporte y las finanzas

Quienes sigan confiando en la gestión manual de datos y en los métodos tradicionales de estimación de costes y plazos corren el riesgo de encontrarse en la situación de las flotas de taxis de la década de 2000, incapaces de adaptarse a la era de las aplicaciones móviles y los cálculos automatizados de rutas a principios de la década de 2020.

CAPÍTULO 10.2.
GUÍA PRÁCTICA PARA APLICAR UN ENFOQUE BASADO EN DATOS

De la teoría a la práctica: hoja de ruta para la transformación digital en la construcción

El sector de la construcción está entrando gradualmente en una nueva fase de desarrollo, en la que los procesos conocidos se ven cada vez más complementados -y a veces incluso sustituidos- por plataformas digitales y modelos de interacción transparentes. Esto plantea a las empresas no solo retos, sino también importantes oportunidades. Las organizaciones que ya están construyendo hoy una estrategia digital a largo plazo no solo podrán mantener su posición en el mercado, sino también ampliarla ofreciendo a los clientes enfoques modernos y soluciones fiables con respaldo tecnológico.

> Es importante darse cuenta de que el conocimiento de conceptos y tecnologías es sólo un punto de partida. Los directivos y especialistas se enfrentan a una cuestión práctica: por dónde empezar la aplicación y cómo convertir las ideas teóricas en valor real. Además, cada vez se plantea más la cuestión de en qué se basará el negocio, si los métodos tradicionales de cálculo de costes y plazos pueden ser revisados por el cliente en cualquier momento.

La respuesta probablemente no esté tanto en la tecnología como en la formación de una nueva cultura profesional en la que el trabajo con datos se perciba como parte integrante de la práctica cotidiana. Es la falta de atención a la tecnología digital y a la innovación lo que ha inoculado al sector de la construcción el grave retraso que se ha observado en las últimas décadas [43].

> Según McKinsey, el gasto en I+D en la industria de la construcción es inferior al 1% de los ingresos, mientras que en las industrias automovilística y aeroespacial esta cifra alcanza el 3,5-4,5%. Del mismo modo, los costes de TI en la construcción siguen siendo inferiores al 1% de los ingresos totales [107].

Como consecuencia, en la construcción no sólo está disminuyendo el nivel de automatización, sino también la productividad laboral, y en 2020 el trabajador de la construcción producirá menos que hace medio siglo (Fig. 10.2-1)

Estos problemas de productividad en el sector de la construcción son comunes a la mayoría de los países desarrollados y en desarrollo (la productividad de la construcción ha caído en 16 de los 29 países de la OCDE (Fig, 2.2-1)), y apuntan no sólo a la falta de tecnología, sino también a la necesidad de cambios sistémicos en los propios planteamientos de gestión, formación e innovación.

El éxito de la transformación digital no depende tanto del número y la disponibilidad de herramientas, sino de la capacidad de las organizaciones para revisar sus procesos y desarrollar una cultura abierta al cambio. La clave no está en la tecnología en sí, sino en las personas y los procesos que garantizan su uso eficaz, apoyan el aprendizaje continuo y fomentan la aceptación de nuevas ideas.

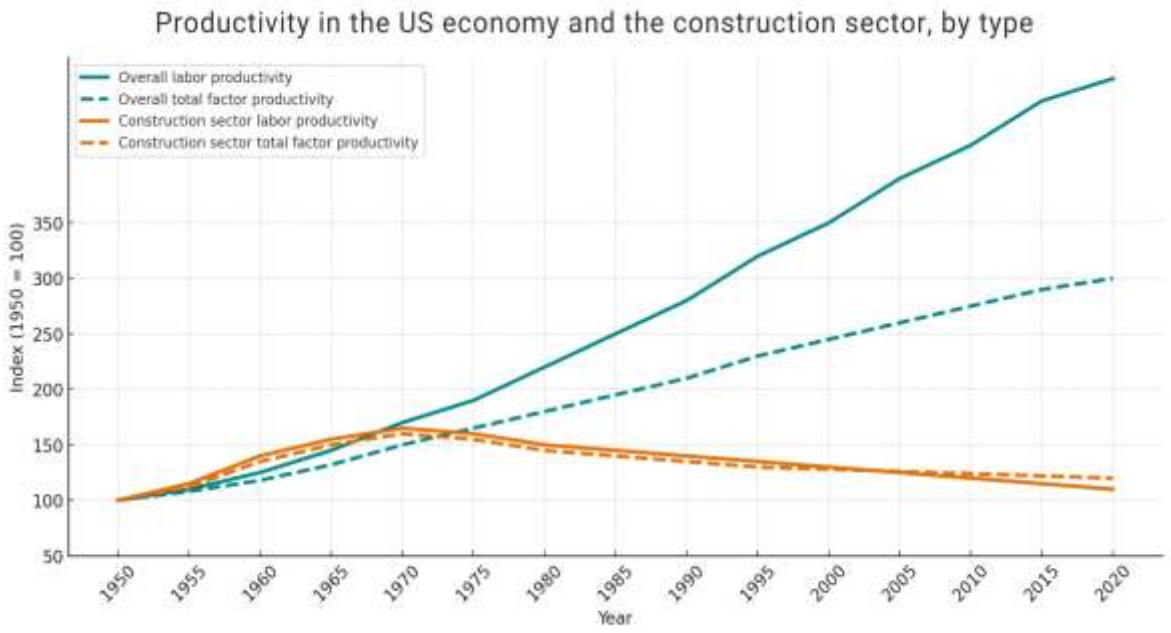


Fig. 10.2-1 La paradoja de la productividad del trabajo y la productividad total de los recursos en la economía y el sector de la construcción de EE.UU. (1950-2020) (basado en [43]).

En las primeras partes del libro, el modelo de entorno empresarial se comparaba con un ecosistema forestal (Fig, 2.1-2, Fig. 1.2-4, Fig. 1.3-2). En un bosque sano, los incendios periódicos, pese a todo su poder destructivo, desempeñan un papel clave en la renovación a largo plazo. Limpian el suelo de vegetación vieja, devuelven los nutrientes almacenados y crean espacio para la vida nueva. Algunas especies vegetales incluso han evolucionado para que sus semillas sólo se abran cuando se exponen a altas temperaturas de fuego, un mecanismo natural que proporciona el momento ideal para la germinación.

Del mismo modo, en las empresas, las crisis pueden actuar como un "agotamiento controlado", fomentando la aparición de nuevos enfoques y empresas que no estén atadas a sistemas anticuados. Estos periodos obligan a abandonar prácticas ineficaces, liberando recursos para la innovación. Al igual que un bosque después de un incendio comienza con plantas pioneras, también una empresa después de una crisis forma procesos nuevos y flexibles que se convierten en la base de un entorno de información maduro.

Las empresas que consigan interpretar correctamente estas "señales de fuego" y transformar su energía disruptiva en un cambio constructivo alcanzarán un nuevo nivel de rendimiento, con procesos de datos más transparentes y adaptables que potencien la capacidad natural de una organización para renovarse y

crecer.

> Ya no cabe duda de la creciente influencia de la inteligencia artificial y el aprendizaje automático en el entorno empresarial. No se trata de una tendencia pasajera, sino de una necesidad estratégica. Las empresas que ignoren la IA, corren el riesgo de perder competitividad en un mercado que fomenta cada vez más la innovación y la flexibilidad.

El futuro pertenece a quienes ven la IA no sólo como una herramienta, sino como una oportunidad para replantearse todos los aspectos de su negocio, desde la optimización de los procesos hasta la toma de decisiones de gestión.

Sentar las bases digitales: 1-5 pasos hacia la madurez digital

En este capítulo, examinamos la hoja de ruta de la transformación digital e identificamos los pasos clave necesarios para implantar un enfoque basado en los datos que pueda ayudar a transformar tanto la cultura corporativa como el ecosistema de información de la empresa.

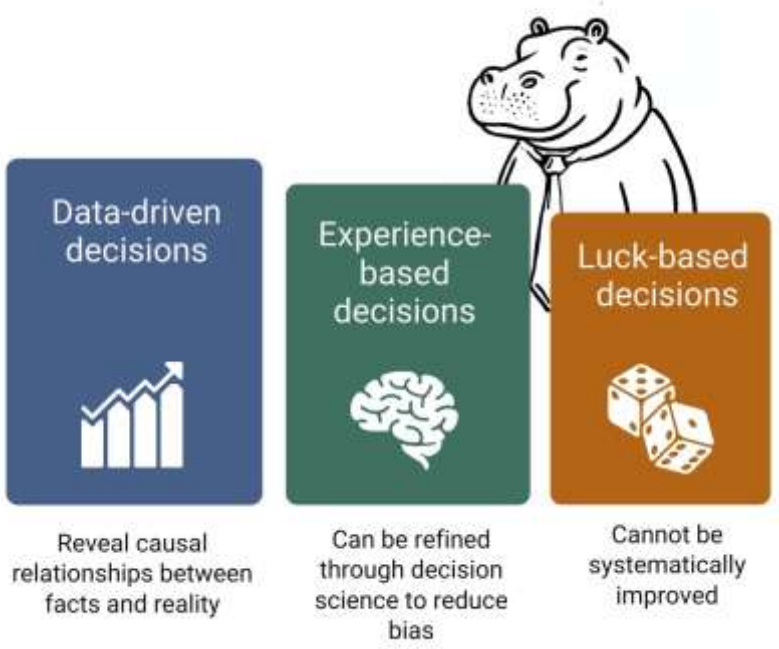


Fig. 10.2-2 Actualización controlada y selección de estrategias: caso, experiencia o datos.

Según el estudio de McKinsey "Por qué fracasan las estrategias digitales" (2018), existen al menos cinco razones[164] por las que las empresas no logran alcanzar los objetivos de transformación digital

- **Definiciones confusas**: los ejecutivos y directivos tienen diferentes concepciones de lo que significa "digital", lo que da lugar a malentendidos e incoherencias.
- **Incomprensión de la economía digital**: muchas empresas subestiman la magnitud de los

cambios que la digitalización está introduciendo en los modelos de negocio y en la dinámica del sector (Fig, 10.1-6).

- **Ignorar los ecosistemas**: las empresas se centran en soluciones tecnológicas individuales (silos de datos), pasando por alto la necesidad de integrarse en ecosistemas digitales más amplios (Fig, 2.2-2, Fig. 4.1-12).
- **Subestimación de la digitalización por parte de los competidores**: los directivos no tienen en cuenta que los competidores también están adoptando activamente tecnologías digitales, lo que puede suponer una pérdida de ventaja competitiva.
- **Pasar por alto la dualidad de la digitalización**: los CEO delegan la responsabilidad de la transformación digital en otros ejecutivos, lo que burocratiza el control y ralentiza el proceso de cambio.

Para hacer frente a estos retos es necesario comprender claramente y alinear las estrategias digitales en todos los niveles de la organización. Antes de elaborar una estrategia digital, es importante comprender el punto de partida. Muchas organizaciones tienden a adoptar nuevas herramientas y plataformas sin tener una visión completa del estado actual.

Paso 1: Realice una auditoría de sus sistemas y datos actuales.

Antes de cambiar los procesos, es importante entender lo que ya está en marcha. Realizar una auditoría permite detectar puntos débiles en la gestión de datos y saber qué recursos pueden utilizarse. Una auditoría es una especie de "radiografía" de sus procesos empresariales. Le permite identificar áreas de riesgo y determinar qué datos son críticos para su proyecto o negocio, y cuáles son secundarios.

Acciones principales:

- Haga un mapa de su entorno informático (en Draw.io, Lucidchart, Miro, Visio o Canva). Enumere los sistemas utilizados (ERP, CAD, CAFM, CPM, SCM y otros) que intervienen en sus procesos y de los que hablamos en el capítulo "Tecnologías y sistemas de gestión en la construcción moderna" (Fig. 1.2-4)
- Evaluar los problemas de calidad de los datos de cada sistema en cuanto a la frecuencia de duplicados, posibles valores omitidos e incoherencias de formato en cada sistema.
- Identifique los "puntos conflictivos", es decir, los lugares donde los procesos pueden fallar o suelen requerir intervención manual: importaciones, exportaciones y procesos de validación adicionales.

> Si quiere que el equipo confíe en los informes, debe asegurarse de que los datos son correctos desde el principio.

Una auditoría de datos de calidad le mostrará qué datos:

- Necesita más desarrollo (hay que establecer procesos de limpieza automáticos o transformaciones adicionales).

- Son "basura" que sólo atasca los sistemas y de la que se puede prescindir no utilizándolos más en los procesos.

Es posible realizar una auditoría de este tipo por cuenta propia. Pero a veces es útil contratar a un consultor externo, sobre todo de otros sectores: una perspectiva fresca y la independencia de las "peculiaridades" de la construcción ayudarán a evaluar con sobriedad el statu quo y evitar los típicos escollos de la parcialidad hacia determinadas soluciones y tecnologías.

Paso 2: Identificar las normas clave para la armonización de datos.

Tras la auditoría, es necesario crear normas comunes para trabajar con los datos. Como comentamos en el capítulo "Normas: de los archivos aleatorios a un modelo de datos inteligente", esto ayudará a eliminar los flujos de datos en silos.

> Sin una norma única, cada equipo seguirá trabajando "a su manera" y usted mantendrá un "zoo" de integraciones en el que se pierden datos con cada conversión.

Acciones principales:

- Seleccione las normas de datos para intercambiar información entre sistemas:
  - Para los datos tabulares, puede tratarse de formatos estructurados como CSV, XLSX o formatos más eficientes como Parquet
  - Para intercambiar datos y documentos poco estructurados: JSON o XML
- Dominar el trabajo con modelos de datos:
  - Empezar por parametrizar las tareas a nivel del modelo conceptual de datos - como se describe en el capítulo "Modelización de datos: modelo conceptual, lógico y físico" (Fig. 4.3-2)
  - A medida que profundice en la lógica del proceso de negocio, pase a formalizar los requisitos mediante parámetros en los modelos lógico y físico (Fig, 4.3-6)
  - Identificar las entidades clave, sus atributos y relaciones dentro de los procesos, y visualizar estas relaciones, tanto entre entidades como entre parámetros (Fig, 4.3-7).
- Utilizar expresiones regulares (RegEx) para validar y normalizar los datos (Fig, 4.4-7), tal y como comentamos en el capítulo "Requisitos estructurados y expresiones regulares RegEx". RegEx no es un tema complejo, pero sí extremadamente importante en el trabajo de creación de requisitos a nivel de modelos físicos de datos.

Sin normas a nivel de datos y visualización de procesos, es imposible ofrecer un entorno digital coherente y escalable. Recuerde que "los datos malos salen caros". Y el coste del error aumenta a medida que un proyecto u organización se hace más complejo. Unificar formatos, definir nombres, estructuras y reglas de validación es una inversión en la estabilidad y escalabilidad de futuras soluciones.

Paso 3. Implantar DataOps y automatizar procesos.

Sin una arquitectura bien definida, las empresas se enfrentarán inevitablemente a datos dispares contenidos en sistemas de información aislados. Los datos estarán desintegrados, duplicados en múltiples ubicaciones y su mantenimiento será costoso.

> Imaginemos que los datos son agua, y la arquitectura de datos es el complejo sistema de tuberías que transporta esa agua desde su fuente de almacenamiento hasta su punto de uso. Es la arquitectura de datos la que determina cómo se recopila, almacena, transforma, analiza y entrega la información a los usuarios finales o a las aplicaciones.

DataOps (Data Operations) es una metodología que integra la recopilación, limpieza, validación y uso de datos en un único flujo de proceso automatizado, tal y como comentamos en detalle en la Parte 8 del libro.

Acciones principales:

- Crear y personalizar ETL -transportadores para automatizar procesos:
  - Extraer: organizar la recopilación automática de datos de documentos PDF (Fig. 4.1-2, Fig. 4.1-5, Fig. 4.1-7), hojas de cálculo Excel, modelos CAD (Fig. 7.2-4), sistemas ERP y otras fuentes con las que trabaje.
  - Transformar: establecer procesos automáticos para transformar los datos a un formato estructurado único y automatizar los cálculos que tendrán lugar fuera de las aplicaciones cerradas (Fig. 7.2-8)
  - Cargar: intente crear una carga automática de datos a tablas resumen, documentos o repositorios centralizados (Fig. 7.2-9, Fig. 7.2-13, Fig. 7.2-16)
- Automatice los procesos de cálculo y QTO (Quantity Take-Off) tal y como comentamos en el capítulo "QTO Quantity Take-Off: Agrupación de datos de proyecto por atributos":
  - Configurar la extracción automática de volúmenes a partir de modelos CAD, utilizando API, plug-ins o herramientas de ingeniería inversa (Fig. 5.2-5).
  - Crear reglas para agrupar elementos de diferentes clases por atributos en forma de tablas (Fig, 5.2-12)
  - Intentar automatizar los cálculos de volumen y costes que se repiten con frecuencia fuera de los sistemas modulares cerrados (Fig. 5.2-15).
- Empieza a utilizar Python y Pandas para procesar datos, tal y como comentamos en el capítulo "Python Pandas: una herramienta indispensable para trabajar con datos":
  - Aplique DataFrame para trabajar con archivos XLSX y automatizar el procesamiento de datos tabulares (Fig. 3.4-6)
  - Automatizar la agregación y transformación de la información mediante diversas bibliotecas de Python
  - Utilice el LLM para simplificar la escritura de bloques de código ya creados y Pipelines completos (Fig, 7.2-18)

- Pruebe a construir un Pipeline en Python, que encuentre errores o detecte anomalías y envíe una notificación a la persona responsable (por ejemplo, el director del proyecto) (Fig, 7.4-2)

La automatización basada en los principios de DataOps permite pasar de una gestión manual y fragmentada de los datos a procesos sostenibles y repetibles. Esto no solo reduce la carga de los empleados que se ocupan de las mismas transformaciones todos los días, sino que también aumenta drásticamente la fiabilidad, escalabilidad y transparencia de todo el sistema de información.

Paso 4: Crear un ecosistema de gobernanza de datos abiertos.

A pesar del desarrollo de sistemas modulares cerrados y de su integración con nuevas herramientas, las empresas se enfrentan a un grave problema: la creciente complejidad de estos sistemas supera su utilidad. La idea inicial de crear una única plataforma propietaria que cubriera todos los procesos empresariales ha conducido a una centralización excesiva, en la que cualquier cambio requiere importantes recursos y tiempo de adaptación.

Como comentamos en el capítulo "Micelio corporativo: cómo los datos conectan los procesos empresariales", la gestión eficaz de los datos requiere un ecosistema abierto y unificado que conecte todas las fuentes de información.

Elementos clave del ecosistema:

- Seleccione un almacén de datos adecuado:
  - Para las tablas y los cálculos, utilice bases de datos, por ejemplo, PostgreSQL o MySQL (Fig. 3.1-7).
  - Para documentos e informes, el almacenamiento en la nube (Google Drive, OneDrive) o los sistemas compatibles con el formato JSON pueden ser adecuados.
  - Compruebe las capacidades de los almacenes de datos, los lagos de datos y otras herramientas para el almacenamiento centralizado y el análisis de grandes cantidades de información (Fig, 8.1-8)
- Implantar soluciones para acceder a datos protegidos:
  - Si utiliza sistemas propietarios, configure el acceso a ellos mediante API o SDK para obtener datos para procesamiento externo (Fig. 4.1-2)
  - Familiarícese con el potencial de las herramientas de ingeniería inversa para formatos CAD (Fig, 4.1-13)
  - Establecer ETL-Pipelines que recojan periódicamente datos de aplicaciones o servidores, los conviertan en formatos estructurados abiertos y los guarden en repositorios (Fig, 7.2-3).
  - Debatir en el equipo cómo facilitar el acceso a los datos sin necesidad de software propietario.
  - Recuerde: los datos son más importantes que las interfaces. Lo que aporta valor a largo

plazo es la estructura y la disponibilidad de la información, no las herramientas específicas de la interfaz de usuario.

- Piense en la posibilidad de crear un centro de excelencia (CoE) para los datos, como se explica en el capítulo "Centro de excelencia (CoE) para el modelado de datos", o en cómo puede aportar su experiencia en datos de otras formas (Fig, 4.3-9).

El ecosistema de gestión de datos crea un espacio de información unificado en el que todos los participantes en el proyecto trabajan con información coherente, actualizada y verificada. Es la base de procesos digitales escalables, flexibles y fiables

Liberar el potencial de los datos: 5-10 pasos hacia la madurez digital

Además de la integración técnica, un factor importante para el éxito de la implantación de soluciones digitales es su adopción por parte de los usuarios finales. Implicar a los clientes o usuarios en la medición del rendimiento es tanto un reto de mejora de la experiencia del usuario como de gestión del cambio en la empresa. Si una solución no encaja en un flujo de trabajo conocido o no resuelve problemas reales de usuarios o clientes, no se utilizará, y ninguna cantidad de medidas e incentivos adicionales lo arreglará.

La transformación es un proceso iterativo basado en el análisis de los datos de interacción de los usuarios con los nuevos procesos, con frecuentes ciclos de pruebas, retroalimentación constante y refinamientos.

Paso 5: Crear una cultura de datos, formar al personal y recabar opiniones

Ni siquiera el sistema más avanzado funcionará sin el compromiso de los empleados. Hay que crear un entorno en el que los datos se utilicen a diario y el equipo comprenda su valor.

> El informe 2024 publicado por el gobierno británico "Data Analytics and AI in Government Project Delivery" señala [83] que la formación de profesionales con las competencias necesarias en el tratamiento y la interpretación de datos es fundamental para implantar con éxito data analytics and AI.

La falta de experiencia en análisis de datos es uno de los principales problemas que limitan la transformación digital. Los líderes están acostumbrados a rutinas establecidas: ciclos trimestrales, iniciativas priorizadas y formas tradicionales de sacar adelante los proyectos. El cambio requiere un líder especial, de rango suficiente para tener influencia, pero no tan alto como para disponer del tiempo y la motivación necesarios para dirigir un proyecto de transformación a largo plazo.

Acciones principales:

- Reconocer la necesidad de pasar de decisiones subjetivas basadas en la opinión de un empleado muy bien pagado (HiPPO) a una cultura de toma de decisiones basada en hechos y datos, como se expone en el capítulo "HiPPO o el peligro de la opinión en la toma de decisiones" (Fig, 2.1-9).
- Organizar una formación sistemática:
    - Organizar sesiones de formación sobre cómo utilizar los datos estructurados e invitar a expertos de otros sectores que no tengan prejuicios hacia los productos y conceptos

populares en el sector de la construcción en la actualidad.

- ☐ Discutir enfoques y herramientas de análisis de datos con colegas, y aprender de forma independiente el trabajo práctico con herramientas como Python, pandas y LLM (Fig. 4.1-3, Fig. 4.1-6)
- ☐ Crear una biblioteca de tutoriales (preferiblemente con vídeos cortos) sobre el tema de la estructuración de datos (Fig. 3.2-15) y la creación de modelos de datos (Fig. 4.3-6, Fig. 4.3-7).

- Utilizar tecnologías modernas de aprendizaje:
  - ☐ Utilizar modelos de lenguaje (LLMs) para apoyar la manipulación de código y datos, incluyendo la generación de código, refactorización y análisis, así como el procesamiento e interpretación de información tabular (Fig, 3.4-1)
  - ☐ Explorar cómo el código generado por LLM puede adaptarse e integrarse en una solución Pipeline completa cuando se trabaja en un entorno de desarrollo offline (IDE) (Fig. 4.4-14, Fig. 5.2-13)

> Cuando un directivo sigue tomando decisiones a la "antigua usanza", ninguna formación convencerá a la gente de que se tome en serio la analítica.

Crear una cultura de datos es imposible sin una retroalimentación continua. La retroalimentación permite identificar lagunas en procesos, herramientas y estrategias que no pueden descubrirse a través de informes internos o KPI formales metrics. Los comentarios elogiosos de los usuarios de sus soluciones no aportarán valor práctico. Lo valioso son los comentarios críticos, sobre todo si se basan en observaciones y hechos concretos. Pero obtener esa información requiere un esfuerzo: hay que crear procesos en los que los participantes -tanto internos como externos- puedan compartir comentarios (puede tener sentido hacerlo de forma anónima) sin distorsiones y sin miedo a que sus opiniones puedan afectar a su propio trabajo. Es importante que lo hagan sin distorsiones y sin temor a consecuencias negativas para ellos mismos.

> *En última instancia, todo aprendizaje es autodidacta [165].*
>
> - Milton Friedman, economista y estadístico estadounidense

La implantación de herramientas analíticas debe ir acompañada de una verificación periódica de su eficacia en la práctica (ROI, KPI), que sólo puede lograrse mediante una retroalimentación estructurada de empleados, clientes y socios. Esto permite a las empresas no sólo evitar repetir errores, sino también adaptarse más rápidamente a los cambios del entorno. Disponer de un mecanismo de recogida y análisis de opiniones es uno de los signos de madurez de una organización que pasa de iniciativas digitales ocasionales a un modelo sostenible de mejora continua (Fig, 2.2-5).

Paso 6: De los proyectos piloto a la ampliación

> *Elige batallas lo bastante grandes para que importen y lo bastante pequeñas para ganarlas.*
>
> - Jonathan Kozol

Lanzar la transformación digital "de golpe y por todas partes" es extremadamente arriesgado. Un enfoque más eficaz es empezar con proyectos piloto e ir ampliando gradualmente las prácticas exitosas.

Acciones principales:

- Elija el proyecto adecuado para el piloto:
  - Definir un objetivo o proceso empresarial específico con resultados mensurables (KPI, ROI) (Fig. 7.1-5)
  - Seleccione un proceso de automatización ETL, como la validación automática de datos o el cálculo de la carga de trabajo (QTO) utilizando Python y Pandas (Fig, 5.2-10
  - Establecer parámetros claros para medir el éxito (por ejemplo, reducir de una semana a un día el tiempo necesario para elaborar especificaciones de inspección o informes de validación de datos).
- Adopte enfoques iterativos:
  - Empiece con procesos sencillos de conversión de datos y cree conversiones en flujo de datos multiformato a los formatos que necesite para sus procesos (Fig. 4.1-2, Fig. 4.1-5).
  - Aumente gradualmente la complejidad de las tareas y amplíe la automatización de los procesos formando un Pipeline completo en el IDE basado en bloques de código documentados (Fig. 4.1-7, Fig. 7.2-18).
  - Documentar y grabar (preferiblemente con vídeos cortos) las soluciones exitosas y compartirlas con colegas o en comunidades profesionales.
- Desarrollar plantillas y documentación complementaria para reproducir dichas soluciones, de modo que puedan ser utilizadas eficazmente por sus colegas (o miembros de la comunidad profesional, incluidos los usuarios de los medios sociales).

Un "roll-up" paso a paso permite mantener la alta calidad de los cambios y no caer en el caos de las implantaciones paralelas. La estrategia "de lo pequeño a lo grande" minimiza los riesgos y permite aprender de los pequeños errores sin dejar que se conviertan en problemas críticos.

Pasar de un enfoque por proyectos, en el que los empleados sólo participan parcialmente, a la formación de equipos permanentes (por ejemplo, centros de competencia - CoE) ayuda a garantizar el desarrollo sostenible del producto incluso después de que salga al mercado su primera versión. Estos equipos no sólo apoyan las soluciones existentes, sino que siguen mejorándolas.

Esto reduce la dependencia de largas aprobaciones: los miembros del equipo están facultados para tomar decisiones dentro de su área de responsabilidad. Como resultado, los directivos se liberan de la necesidad de microgestionar, y los equipos pueden centrarse en crear valor real.

> Desarrollar nuevas soluciones no es un sprint, sino un maratón. Los que lo consiguen son los que inicialmente se proponen un trabajo a largo plazo y constante.

Es importante darse cuenta de que la tecnología requiere un desarrollo constante. Invertir en el desarrollo a largo plazo de soluciones tecnológicas es la base del éxito de las operaciones.

Paso 7: Utilizar formatos y soluciones de datos abiertos

Como ya hemos comentado en los capítulos dedicados a las plataformas modulares (ERP, PMIS, CAFM, CDE, etc.), es importante centrarse en formatos de datos abiertos y universales que garanticen la independencia de las soluciones de los proveedores y aumenten la disponibilidad de la información para todas las partes interesadas.

Acciones principales:

- Pasar de formatos cerrados a formatos abiertos:
  - Utilice formatos abiertos en lugar de formatos propietarios, o encuentre la forma de configurar la carga automática o la conversión de formatos cerrados a formatos abiertos (Fig, 3.2-15).
  - Implementar herramientas para trabajar con Parquet, CSV, JSON, XLSX, que son estándares de intercambio entre la mayoría de los sistemas modernos (Fig. 8.1-2)
  - Si el trabajo con geometría 3D desempeña un papel importante en sus procesos, considere la posibilidad de utilizar formatos abiertos como USD, glTF, DAE u OBJ (Fig, 3.1-14).
- Utilizar bases de datos vectoriales bases de datos para analizar y buscar información de forma eficaz:
  - Utilice Bounding Box y otros métodos para simplificar la geometría 3D (Fig, 8.2-1)
  - Piense dónde puede aplicar la vectorización de datos, convirtiendo textos, objetos o documentos en representaciones numéricas (Fig, 8.2-2).
- Aplicar herramientas de análisis de macrodatos:
  - Organizar el almacenamiento de los datos históricos acumulados (por ejemplo, PDF, XLSX, CAD) en formatos adecuados para su análisis (Apache Parquet, CSV, ORC) (Fig. 8.1-2).
  - Empezar a aplicar métodos estadísticos básicos y trabajar con muestras representativas o, como mínimo, familiarizarse con los principios fundamentales de la estadística (Fig, 9.2-5).

- Implementar y aprender herramientas para visualizar los datos y las relaciones entre ellos para visualizar los resultados del análisis. Sin una buena visualización, es imposible comprender plenamente los propios datos o los procesos basados en ellos (Fig, 7.1-4).

El paso a formatos de datos abiertos y la introducción de herramientas de análisis, almacenamiento y visualización de la información sientan las bases de una gobernanza digital sostenible e independiente. Esto no solo reduce la dependencia de los proveedores, sino que garantiza la igualdad de acceso a los datos para todas las partes interesadas.

Paso 8. Empezar a aplicar el aprendizaje automático para la predicción

Muchas empresas han acumulado ingentes cantidades de datos, una especie de "géiseres de información" que siguen sin utilizarse. Estos datos se han recopilado en cientos o miles de proyectos, pero a menudo sólo se han utilizado una vez o no se han utilizado en absoluto en procesos posteriores. Los documentos y modelos almacenados en formatos y sistemas cerrados suelen percibirse como obsoletos e inútiles. En realidad, sin embargo, son un recurso valioso: la base para analizar los errores cometidos, automatizar operaciones rutinarias y desarrollar soluciones innovadoras de autoclasificación y reconocimiento de características en proyectos futuros.

El principal reto es aprender a extraer estos datos y transformarlos en información práctica. Como se explica en el capítulo sobre aprendizaje automático y predicciones, las técnicas de aprendizaje automático tienen el potencial de mejorar significativamente la precisión de las estimaciones y predicciones en diversos procesos relacionados con la construcción. La plena utilización de los datos acumulados abre el camino a la mejora de la eficiencia, la reducción del riesgo y la construcción de procesos digitales sostenibles.

Acciones principales:

- Empiece con algoritmos sencillos:
  - Pruebe a aplicar la regresión lineal -utilizando pistas del LLM- para predecir el rendimiento recurrente en conjuntos de datos en los que las dependencias de un gran número de factores son nulas o mínimas (Fig, 9.3-4)
  - Considere en qué etapas de sus procesos podría aplicarse teóricamente el algoritmo k-nearest neighbours (k-NN) - por ejemplo, para tareas de clasificación, evaluación de similitud de objetos o predicción basada en análogos históricos (Fig. 9.3-5).
- Recopilar y estructurar datos para entrenar modelos:
  - Recopilar los datos históricos del proyecto en un solo lugar y en un único formato (Fig, 9.1-10)
  - Trabajar en la calidad y representatividad de las muestras de entrenamiento, a través del ETL automatizado (Fig, 9.2-8)
  - Aprender a separar los datos en conjuntos de entrenamiento y de prueba, como hicimos en el ejemplo del conjunto de datos Titanic (Fig, 9.2-6, Fig. 9.2-7).
- Considerar la posibilidad de ampliar la aplicación de técnicas de aprendizaje automático a una

amplia gama de tareas, desde la predicción de los plazos de los proyectos hasta la optimización de la logística, la gestión de recursos y la identificación temprana de posibles problemas.

El aprendizaje automático es una herramienta para convertir los datos archivados en un valioso activo para la predicción, la optimización y la toma de decisiones informadas. Empiece con conjuntos de datos pequeños (Fig, 9.2-5) y modelos sencillos, aumentando gradualmente su complejidad.

Paso 9. Integrar IoT y tecnologías modernas de recopilación de datos

El mundo de la construcción se está digitalizando rápidamente: cada foto de construcción, cada mensaje de los equipos ya forma parte de un proceso más amplio de parametrización y tokenización de la realidad. Al igual que el GPS transformó en su día la logística, las tecnologías IoT, RFID y de recopilación automática de datos están cambiando el sector de la construcción. Como se expone en el capítulo "IoT Internet de las cosas y contratos inteligentes", la obra digital con sensores y supervisión automatizada es el futuro del sector.

Acciones principales:

- Implantar IoT -dispositivos, RFID -etiquetas y detallar los procesos asociados a ellos:
  - Evaluar qué áreas o fases de un proyecto pueden beneficiarse del mayor retorno de la inversión (ROI); por ejemplo, para controlar la temperatura, la vibración, la humedad o el movimiento de
  - Considere la posibilidad de utilizar RFID para realizar el seguimiento de materiales, herramientas y equipos a lo largo de la cadena de suministro.
  - Considere cómo pueden integrarse los datos recogidos en un único sistema de información, como Apache NiFi, para su procesamiento y análisis automatizados en tiempo real (Fig, 7.4-5)
- Establecer un sistema de supervisión en tiempo real:
  - Desarrollar cuadros de mando para realizar un seguimiento de los procesos clave o métricas del proyecto utilizando herramientas de visualización como Streamlit, Flask o Power BI)
  - Establezca notificaciones automáticas para señalar las desviaciones críticas del plan o las normas (Fig. 7.4-2).
  - Evaluar el potencial de mantenimiento predictivo de los equipos basándose en los datos recogidos y los patrones identificados (Fig, 9.3-6).
- Combinar datos de distintas fuentes:
  - Empiece por visualizar el modelo de datos a nivel físico: refleje la estructura de los flujos de información y los parámetros clave procedentes de sistemas CAD, dispositivos IoT y plataformas ERP (Fig. 4.3-1).
  - Empiece por crear un esbozo de una plataforma unificada para el análisis de datos y el apoyo a la toma de decisiones de gestión. Capture las funciones clave, las fuentes de

datos, los usuarios y los escenarios de aplicación previstos (Fig, 4.3-7).

Cuanto antes empiece a conectar los procesos reales al mundo digital, antes podrá gestionarlos con datos, de forma eficaz, transparente y en tiempo real.

Paso 10. Prepararse para el futuro de los cambios en la industria

Las empresas de la construcción están sometidas constantemente a la presión del entorno exterior: crisis económicas, saltos tecnológicos, cambios normativos. Como un bosque que tiene que soportar la lluvia, la nieve, la sequía y el sol abrasador, las empresas viven en condiciones de adaptación continua. Y del mismo modo que los árboles se hacen resistentes a las heladas y la sequía gracias a sus profundos sistemas radiculares, sólo las organizaciones que cuentan con una sólida base de procesos automatizados, capacidad para anticiparse a los cambios y estrategias de adaptación flexible siguen siendo viables y competitivas.

Como se menciona en el capítulo "Estrategias de supervivencia: creación de ventajas competitivas", el sector de la construcción está entrando en una fase de transformación radical. La interacción entre cliente y contratista está evolucionando hacia un modelo de persuasión, en el que la transparencia, la previsibilidad y las herramientas digitales están sustituyendo a los enfoques tradicionales. En esta nueva realidad, no ganan los más grandes, sino los más flexibles y tecnológicamente maduros.

Acciones principales:

- Analizar las vulnerabilidades empresariales en el contexto de los datos abiertos:
  - ☐ Evalúe cómo la democratización del acceso a los datos como parte de la Uberización podría tener un impacto devastador en su ventaja competitiva y su negocio (Fig, 10.1-5)
  - ☐ Piense en una estrategia para pasar de procesos opacos y aislados a modelos empresariales basados en soluciones abiertas, interoperabilidad de sistemas y transparencia de datos (Fig. 2.2-5).
- Desarrollar una estrategia digital a largo plazo:
  - ☐ Determine si aspira a ser un líder en innovación o prefiere un escenario de "recuperación" en el que conservará sus recursos.
  - ☐ Describir las etapas: a corto plazo (automatización de procesos, estructuración de datos), a medio plazo (implantación de LLM y ETL), a largo plazo (ecosistemas digitales, repositorios centralizados).
- Piense en ampliar su cartera de servicios:
  - ☐ Considerar la posibilidad de ofrecer nuevos servicios (centrados en la eficiencia energética, ESG, servicios de datos). Hablaremos más sobre nuevos modelos de negocio en el próximo capítulo
  - ☐ Intente posicionarse como un socio tecnológico fiable que respalda todo el ciclo de vida de una instalación, desde el diseño hasta la explotación. La confianza en usted debe basarse en un enfoque sistemático, procesos transparentes y la capacidad de ofrecer soluciones tecnológicas sostenibles.

En un entorno de transformación, no ganan los que se limitan a reaccionar ante el cambio, sino los que son proactivos. La flexibilidad, la apertura y la madurez digital son las bases de la sostenibilidad en la construcción del mañana.

Hoja de ruta de la transformación: del caos a la empresa basada en datos

El siguiente plan puede servir como punto de referencia inicial, un punto de partida para dar forma a su propia estrategia de transformación digital basada en datos:

- **Auditoría y normas**: analizar el estado actual, unificar datos
- **Estructuración y clasificación de datos**: automatización de la transformación de datos no estructurados y poco estructurados.
- **Automatizar agrupaciones**, cálculos y cómputos: utilizar herramientas y bibliotecas de código abierto para la automatización
- **Ecosistema y COE**: crear un equipo interno para formar un ecosistema de datos unificado en la empresa.
- **Cultura y aprendizaje**: pasar de las soluciones HiPPO a las soluciones basadas en datos
- **Pilotos, reacciones y ampliación**: actuar de forma iterativa: probar nuevos métodos a escala limitada, recoger reacciones válidas y ampliar gradualmente las soluciones.
- **Formatos abiertos**: utilice formatos universales y abiertos para independizarse de los proveedores de software.
- **Aprendizaje automático**: integración de algoritmos de ML en procesos de predicción y optimización
- **IoT y la obra digital**: integrar las modernas tecnologías de recopilación de datos en los procesos
- **Adaptación estratégica**: prepararse para los futuros cambios del sector

> Y lo que es más importante, recuerde que "los datos por sí solos no cambian una empresa: son las personas que saben trabajar con esos datos las que la cambian". Hacer hincapié en la cultura, los procesos transparentes y el compromiso con la mejora continua.

Un enfoque sistémico permite pasar de iniciativas digitales aisladas a un modelo de gestión plenamente basado en datos en el que las decisiones no se basan en intuiciones o suposiciones, sino en datos, hechos y probabilidades calculadas matemáticamente. La transformación digital del sector de la construcción no consiste solo en adoptar tecnología, sino en dar forma a un ecosistema empresarial en el que la información del proyecto se transfiera de forma fluida e iterativa entre diferentes sistemas. Al hacerlo, los algoritmos de aprendizaje automático proporcionan análisis, previsiones y optimización de procesos automáticos y continuos. En un entorno así, la especulación y los datos ocultos dejan de tener importancia: solo quedan los modelos probados, los cálculos transparentes y los resultados predecibles.

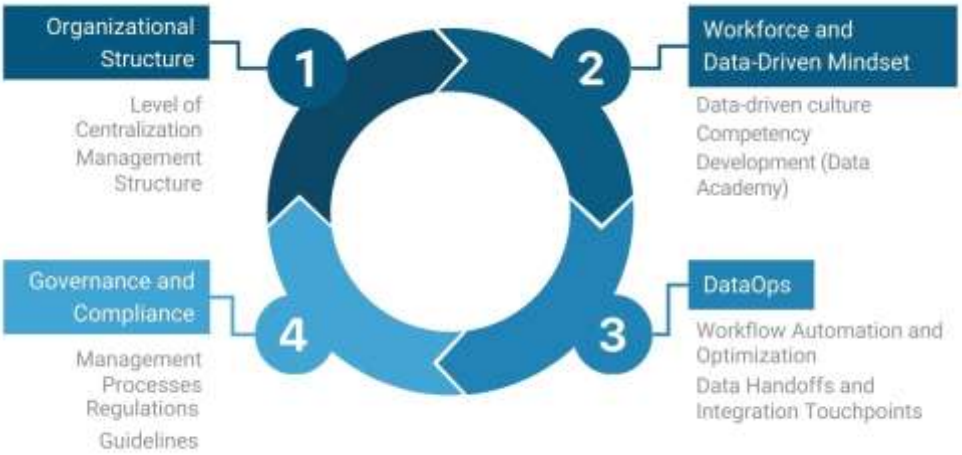


Fig. 10.2-3 Elementos clave para el éxito de la gestión de datos a nivel de empresa.

Cada parte del libro corresponde a una etapa específica del procesamiento y análisis de datos en los proyectos de construcción (Fig. 2.2-5). Si desea volver a uno de los temas tratados anteriormente y analizarlo desde una perspectiva global del flujo de utilización de datos, puede consultar los títulos de las partes en la Fig. 10.2-4.

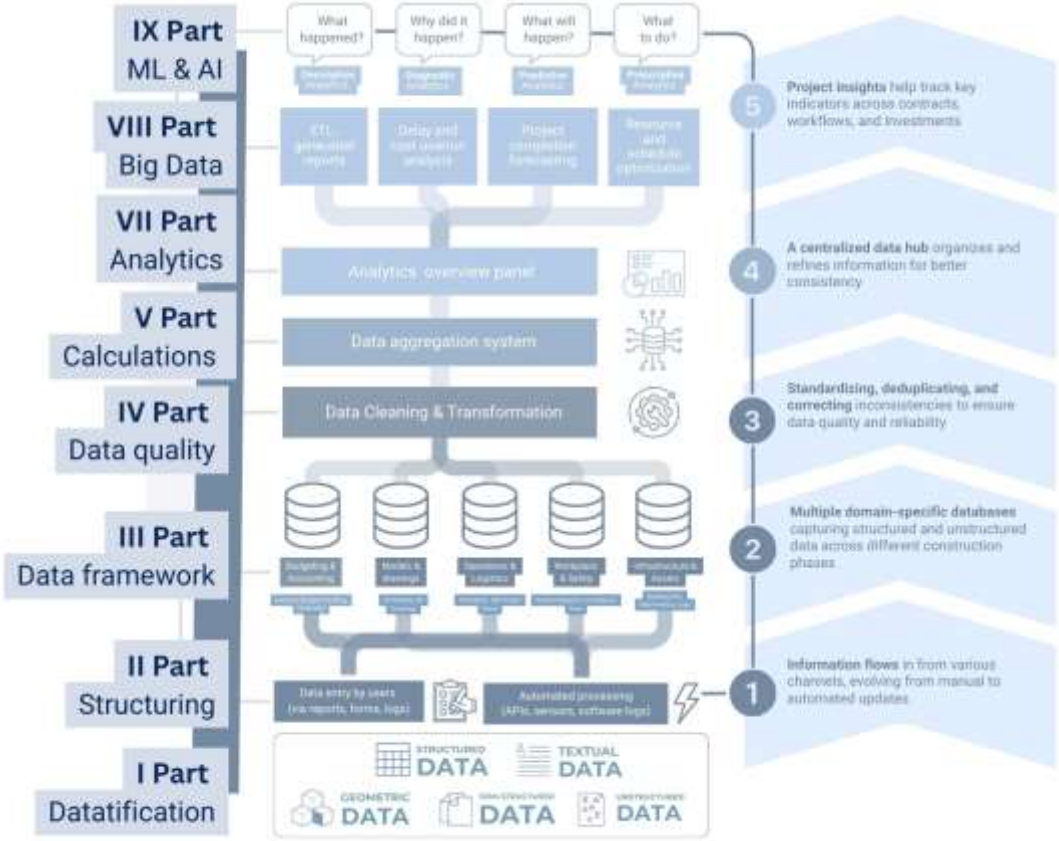


Fig. 10.2-4 Partes del libro en el contexto de la cadena de procesamiento de datos (Fig. 2.2-5): de la digitalización de la información a la analítica y la inteligencia artificial.

Independientemente del tamaño de su organización, su nivel de madurez tecnológica o su presupuesto, puede empezar hoy mismo a adoptar un enfoque basado en los datos en. Incluso los pequeños pasos en la dirección correcta producirán resultados con el tiempo.

> La transformación basada en datos no es un proyecto puntual, sino un proceso de mejora continuo e iterativo que incluye la introducción de nuevas herramientas, el rediseño de procesos y el desarrollo de una cultura de toma de decisiones basada en datos.

Construir en la Industria 5.0: cómo ganar dinero cuando ya no puedes esconderte

Durante mucho tiempo, las empresas de construcción han estado ganando dinero con la opacidad de los procesos. El principal modelo de negocio era la especulación -sobreestimar el coste de los materiales, el alcance de los trabajos y los porcentajes de margen en sistemas cerrados ERP - y PMIS - inaccesibles a auditorías externas. El acceso limitado de los clientes y sus personas de confianza a los datos originales del proyecto creaba el terreno para esquemas en los que resultaba casi imposible verificar la fiabilidad de los cálculos.

Sin embargo, este modelo está perdiendo relevancia rápidamente. Con la democratización del acceso a los datos, la aparición de los LLM, la llegada de los datos abiertos y las herramientas de automatización ETL, el sector está pasando a un nuevo estándar de trabajo.

> Como resultado, la opacidad ya no es una ventaja competitiva: pronto se convertirá en una carga de la que será difícil desprenderse. La transparencia pasa de ser una opción a convertirse en un requisito indispensable para permanecer en el mercado.

¿Con quién trabajarán los clientes -bancos, inversores, clientes físicos, capital riesgo, clientes gubernamentales- en la nueva realidad digital? La respuesta es obvia: con quienes sean capaces de proporcionar no sólo el resultado, sino también la justificación de cada paso en el camino hacia él. A medida que aumente el volumen de datos abiertos, los socios y clientes elegirán empresas que garanticen la transparencia, la precisión y la previsibilidad de los resultados.

En este contexto, están surgiendo nuevos modelos de negocio basados en la gestión de datos y la confianza, más que en la especulación:

- **Vender procesos en lugar de metros cuadrados**: el activo clave pasa a ser la confianza y la eficacia en lugar de los acuerdos concretos descontados. El valor principal será la previsibilidad del resultado basada en datos fiables y verificados. Las empresas modernas no venderán el objeto de construcción como tal, sino:
  - plazos precisos y calendarios de trabajo transparentes;
  - estimaciones razonables, respaldadas por cálculos;
  - trazabilidad y control digitales completos en todas las fases del proyecto.
- **Ingeniería y analítica como servicio**: el modelo "Data-as-a-Service " (una forma de ofrecer datos ya preparados a los usuarios a través de Internet, como un servicio), en el que cada proyecto pasa a formar parte de una cadena digital de datos, y el valor empresarial está en la capacidad de gestionar esta cadena. Las empresas se están transformando en plataformas inteligentes que ofrecen soluciones basadas en la automatización y la analítica:
  - Elaboración automatizada y transparente de presupuestos y planes;
  - evaluación de riesgos y plazos basada en algoritmos de aprendizaje automático;
  - cálculo de indicadores medioambientales (ESG, CO_2, eficiencia energética);

    - generación de informes a partir de fuentes abiertas auditadas.

- **Productización de los conocimientos de ingeniería**: los desarrollos de la empresa pueden utilizarse repetidamente dentro de la empresa y distribuirse como un producto independiente, lo que constituye una fuente adicional de ingresos a través de los servicios digitales. En el nuevo entorno, las empresas no solo crean proyectos, sino también activos digitales:
    - bibliotecas de componentes y plantillas de presupuestos;
    - módulos de verificación automatizada;
    - Plugins y scripts de código abierto (venta de consultoría) para trabajar con datos.
- Un **nuevo tipo de empresa: el integrador basado en datos**: un agente del mercado que no depende de proveedores de software específicos ni de sistemas modulares y que no está "encerrado" en una única interfaz de software. Opera libremente con datos y basa en ellos su competitividad. La empresa de construcción del futuro no es sólo un contratista, sino un integrador de información capaz de realizar las siguientes funciones para el cliente:
    - Combine datos de fuentes dispares y realice análisis;
    - Garantizar la transparencia y credibilidad de los procesos;
    - asesorar sobre la optimización de los procesos empresariales;
    - desarrollar herramientas que funcionen en el ecosistema de datos abiertos, LLM, ETL y Pipelines.

La Industria 5.0 (Fig. 2.1-12) marca el final de la "era de los promedios manuales" y de las reuniones nocturnas entre los directores generales y el departamento de estimaciones y contabilidad. Todo lo que antes estaba oculto -cálculos, estimaciones, volúmenes- pasa a ser abierto, verificable y comprensible incluso para los no expertos. Los primeros en reorientarse serán los ganadores. Todos los demás quedarán fuera de la nueva economía digital del sector de la construcción.

CONCLUSIÓN

El sector de la construcción está entrando en una era de cambios fundamentales. Desde los primeros registros en tablillas de arcilla hasta las ingentes cantidades de datos digitales que fluyen de los servidores de proyectos y las obras, la historia de la gestión de la información en la construcción siempre ha reflejado la madurez de la tecnología de su época. Hoy, con la llegada de la automatización, los formatos abiertos y los sistemas de análisis inteligentes, el sector se enfrenta no a una evolución gradual, sino a una rápida transformación digital.

Al igual que en otros sectores de la economía, la construcción tendrá que replantearse no sólo las herramientas, sino también los principios de trabajo. Las empresas que solían dictar el mercado y servir de intermediario principal entre el cliente y el proyecto están perdiendo su posición única. La confianza y la capacidad de trabajar con datos -desde su recogida y estructuración hasta el análisis, la previsión y la automatización de las decisiones- están pasando a un primer plano.

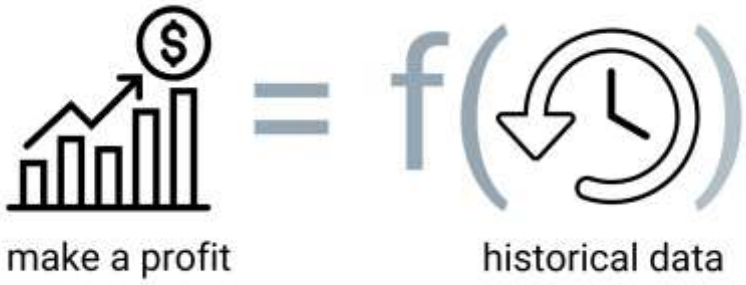


Fig. 10.2-1 Los datos históricos estructurados son el combustible de una empresa eficiente y gestionable.

Este libro detalla los principios clave de la gestión de datos en el sector de la construcción: desde la auditoría y la normalización hasta la automatización de procesos, el uso de herramientas de visualización y la implantación de algoritmos inteligentes. Examinamos cómo, incluso con recursos limitados, se puede construir una arquitectura de datos que funcione y empezar a tomar decisiones basadas en hechos verificables y no en la intuición. Trabajar con datos ya no es sólo tarea del departamento informático: se convierte en la base de la cultura de gestión, que determina la flexibilidad, adaptabilidad y sostenibilidad a largo plazo de la empresa.

La aplicación de tecnologías de aprendizaje automático, sistemas de procesamiento automático, gemelos digitales y formatos abiertos permite ya hoy eliminar el factor humano allí donde antes era crítico. La construcción avanza hacia la autonomía y la controlabilidad, donde el paso de la idea a la realización del proyecto puede compararse a la navegación en modo piloto automático: sin dependencia de decisiones subjetivas, sin necesidad de intervención manual en cada fase, pero con total trazabilidad y control digitales (Fig. 10.2-2).

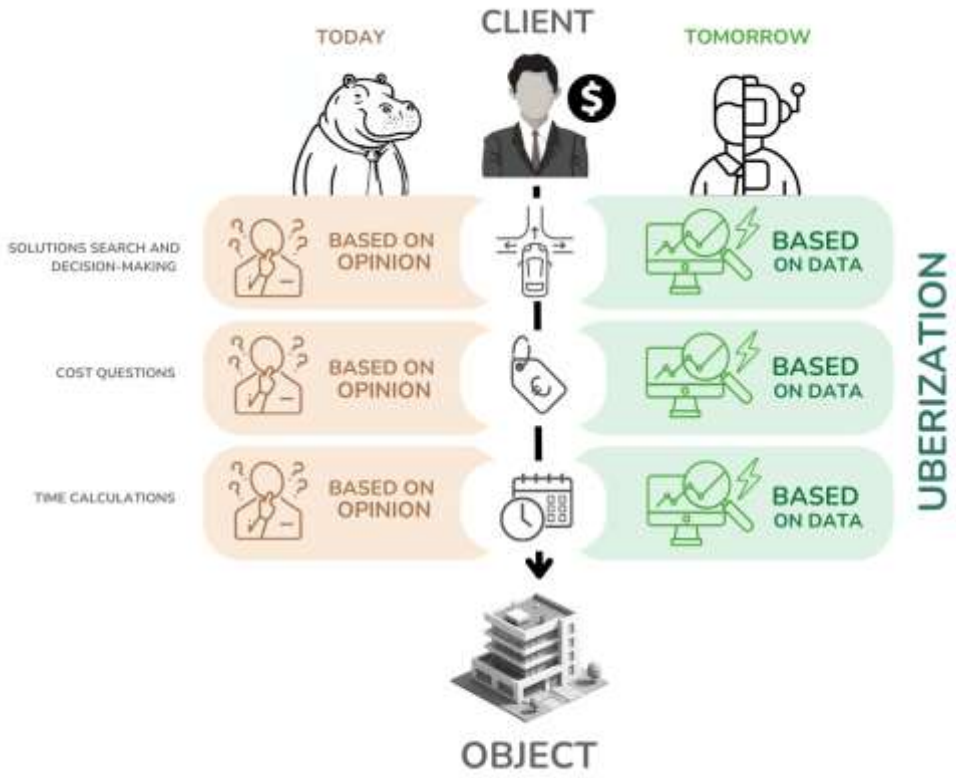


Fig. 10.2-2 El cambio de la toma de decisiones basada en las opiniones de importantes expertos (HiPPO) al análisis de datos será promovido principalmente por el cliente.

Al aprender los métodos, principios y herramientas que se presentan en este libro, podrá empezar a tomar decisiones en su empresa basadas en datos, en lugar de intuitivas. También podrá ejecutar cadenas de módulos en LLM, copiar ETL Pipelines listas para usar en su entorno de desarrollo (IDE) y procesar datos automáticamente para obtener la información que necesita en la forma que desea. Basándose en los capítulos del libro sobre big data y aprendizaje automático, podrá implementar escenarios más complejos, extrayendo nuevas perspectivas de datos históricos y aplicando algoritmos de aprendizaje automático para predecir y optimizar sus procesos.

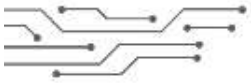
Los datos y procesos abiertos sentarán las bases para realizar estimaciones más precisas de los costes y plazos de los proyectos, evitando que las empresas constructoras especulen con datos opacos. Se trata tanto de un reto como de una oportunidad para que el sector se replantee su papel y se adapte a un nuevo entorno en el que la transparencia y la eficiencia se convertirán en factores clave del éxito.

> La voluntad de tomar los conocimientos y ponerlos en práctica es clave para el éxito en la era de la transformación digital.

Las empresas que se den cuenta antes tendrán ventaja en la nueva competencia digital. Pero es importante darse cuenta de que los datos por sí solos no cambiarán nada. Muchas personas tendrán que cambiar su forma de pensar, y eso requiere un incentivo. Su empresa debe replantearse la forma en que comparte los datos

Las personas que cambian la empresa son las que saben trabajar con estos datos, interpretarlos, utilizarlos para optimizarlos y crear una nueva arquitectura de procesos basada en ellos.

Si estás leyendo estas líneas, estás preparado para el cambio y ya vas un paso por delante. Gracias por elegir este camino. ¡Bienvenido a la era de la transformación digital!



SOBRE EL AUTOR

Me llamo Artem Boiko. Mi andadura en la construcción empezó en 2007, con un trabajo como minero en una mina de esquisto bituminoso, en mi ciudad natal, mientras estudiaba en la Universidad de Minas de San Petersburgo, especializándome en construcción de minas y subterráneos. En la contraportada de este libro se puede ver un explosivo en el frente donde extraíamos y volábamos cientos de cubos de pizarra bituminosa. Mi carrera me ha llevado en muchas direcciones diferentes, desde trabajar como obrero de la construcción minera y subterránea hasta escalador industrial, techador e instalador de equipos de elevación. He tenido el honor de participar en proyectos de diversa envergadura, desde la construcción de casas particulares hasta grandes instalaciones industriales en distintas regiones del mundo.

Con el tiempo, mi trabajo ha pasado de la construcción física a la gestión de la información y los procesos digitales. Desde 2013, he trabajado en diversos puestos en pequeñas, medianas y grandes empresas de construcción de varias regiones de Alemania, desde diseñador hasta gestor de gestión de datos. En cuanto a la gestión de datos, mi experiencia consiste en trabajar con datos en varios sistemas ERP, CAD (BIM), MEP, FEM, CMS. He participado en la optimización, la automatización de procesos, así como en el análisis, el aprendizaje automático y el procesamiento de datos en las fases de planificación, cálculo y ejecución de obras de construcción en empresas de construcción industrial, residencial, de infraestructuras y de servicios públicos.

Llevo trabajando con software de código abierto y datos abiertos desde 2003. Durante este tiempo, he implementado muchos proyectos web, desde sitios web y tiendas online hasta aplicaciones web completas utilizando soluciones de código abierto y CMS abiertos. Estas plataformas, similares en muchos aspectos a los modernos ERP de construcción, tienen una arquitectura modular, gran adaptabilidad y accesibilidad. Esta experiencia sentó las bases de mi enfoque profesional: centrarme en las tecnologías de código abierto y en una cultura de desarrollo colaborativo. El respeto por el código abierto y el libre intercambio de conocimientos es algo que me esfuerzo por promover en el sector de la construcción. Mi trabajo para mejorar la accesibilidad de los datos en el sector de la construcción se ha traducido en la creación de varias comunidades en las redes sociales para debatir sobre la apertura de los datos y el uso del código abierto en la construcción, así como en el lanzamiento de varias startups que desarrollan soluciones para facilitar el acceso a los datos desde diversos sistemas y plataformas cerrados.

Mi contribución a la comunidad profesional se concreta en la participación como ponente en conferencias sobre interoperabilidad CAD (BIM), ERP, 4D-5D, LLM Machine Learning e Inteligencia Artificial, así como en artículos publicados en publicaciones europeas del sector de la construcción. Uno de mis logros notables es la creación de "BIM History" [111], un mapa exhaustivo de importantes soluciones de software para la gestión de datos en el sector de la construcción. Mi serie de artículos en 7 partes "BIM Development and Lobbying Games", traducida a varios idiomas, ha sido ampliamente reconocida como un intento de iluminar la dinámica oculta de los estándares digitales.

Así es como pasé de extraer rocas a extraer y sistematizar datos de construcción. Siempre estoy abierto al diálogo profesional, a nuevas ideas y a proyectos conjuntos. Aceptaré con gratitud cualquier comentario y espero recibir sus mensajes o verle entre mis seguidores de las redes sociales. Muchas gracias por leer este libro hasta el final. Me alegraría que este libro le ayudara a comprender mejor el tema de los datos en el sector de la construcción.

MAYOR RELACIÓN

Las opiniones de los lectores desempeñan un papel importante en el desarrollo de las publicaciones y la selección de temas prioritarios. Los comentarios sobre qué ideas han resultado útiles y cuáles han suscitado dudas y requieren más aclaraciones o la cita de fuentes son especialmente valiosos. El libro incluye una amplia gama de material y análisis, algunos de los cuales pueden parecer controvertidos o subjetivos. Si en el transcurso de la lectura encuentra imprecisiones, fuentes citadas incorrectamente, incoherencias lógicas o erratas, le agradecería sus comentarios, reflexiones o críticas, que puede enviar a boikoartem@gmail.com. O a través de mensajes en: linkedin.com/in/boikoartem.

Agradecería enormemente cualquier mención del libro Data-Driven Construction en las redes sociales: compartir la experiencia de lectura ayuda a difundir los datos y herramientas abiertos y apoya mi trabajo.

COMENTARIO DE TRADUCCIÓN

Este libro se ha traducido utilizando tecnología de inteligencia artificial. Esto ha agilizado mucho el proceso de traducción. Sin embargo, como ocurre con cualquier operación tecnológica, pueden producirse errores o imprecisiones. Si observa algo que parezca incorrecto o mal traducido, envíeme un correo electrónico. Sus comentarios ayudarán a mejorar la calidad de la traducción.

COMMUNITY

Es un lugar donde puede plantear libremente sus preguntas y compartir sus problemas y soluciones:

DataDrivenConstruction.io: https://datadrivenconstruction.io

LinkedIn: https://www.linkedin.com/company/datadrivenconstruction/

Twitter: https://twitter.com/datadrivenconst

Telegram: https://t.me/datadrivenconstruction

YouTube: https://www.youtube.com/@datadrivenconstruction

OTRAS COMPETENCIAS Y CONCEPTOS

Además de los principios clave del trabajo con datos en el sector de la construcción, DataDrivenConstruction aborda una amplia gama de conceptos, programas y habilidades adicionales que son esenciales para el profesional orientado a los datos. Algunos de ellos se presentan sólo a modo de resumen, pero son fundamentales para la práctica.

El lector interesado puede visitar el sitio web DataDrivenConstruction.io para obtener enlaces a materiales adicionales sobre competencias clave. Estos materiales incluyen el trabajo con Python y Pandas, la construcción de procesos ETL, ejemplos de procesamiento de datos en proyectos CAD de construcción, sistemas de big data y enfoques modernos de visualización y análisis de datos de construcción.

En la elaboración del libro "DataDrivenConstruction" y de todos los casos prácticos se han utilizado numerosas herramientas y programas de código abierto. El autor desea dar las gracias a los desarrolladores y coautores de las siguientes soluciones:

- Python y Pandas - la columna vertebral de la manipulación y automatización de datos
- Scipy, NumPy, Matplotlib y Scikit-Learn: bibliotecas para el análisis de datos y el aprendizaje automático.
- SQL y Apache Parquet - herramientas para almacenar y procesar grandes cantidades de datos de construcción.
- Fuente abierta CAD (BIM) herramientas de datos abiertos en formatos abiertos
- N8n, Apache Airflow, Apache NiFi - sistemas de orquestación y automatización de flujos de trabajo
- DeepSeek, LlaMa, Mistral - Código abierto LLM

Un agradecimiento especial a todos los participantes en los debates sobre el tema de los datos y herramientas abiertos en comunidades profesionales y redes sociales, cuyas críticas, comentarios e ideas han contribuido a mejorar el contenido y la estructura de este libro.

Siga el desarrollo del proyecto en el sitio web DataDrivenConstruction.io, donde no sólo se publican actualizaciones y correcciones del libro, sino también nuevos capítulos, tutoriales y ejemplos prácticos de aplicación de las técnicas descritas.

MÁXIMA COMODIDAD CON LA VERSIÓN IMPRESA

Tiene en sus manos una versión digital gratuita de **Data-Driven Construction**. Para trabajar con mayor comodidad y acceder rápidamente a los materiales, le recomendamos que preste atención a la **edición impresa**:

- **Siempre a mano**: el libro en formato impreso se convertirá en una herramienta de trabajo fiable, que le permitirá encontrar y utilizar rápidamente las visualizaciones y diagramas adecuados en cualquier situación de trabajo.
- **Ilustraciones de alta calidad**: todas las imágenes y gráficos de la edición impresa se presentan con la máxima calidad.
- **Acceso rápido a la información**: navegación sencilla, posibilidad de tomar notas, marcar favoritos y trabajar con el libro en cualquier lugar.

Al adquirir la versión impresa completa del libro, obtendrá una cómoda herramienta para trabajar con la información de forma cómoda y eficaz: la posibilidad de utilizar rápidamente materiales visuales en las tareas cotidianas, encontrar rápidamente los esquemas necesarios y tomar notas. Además, su compra apoya la difusión del conocimiento abierto.

Solicite la versión impresa del libro en: datadrivenconstruction.io/books

OPORTUNIDAD ÚNICA PARA POSICIONAMIENTO ESTRATÉGICO

Le ofrecemos colocar material publicitario en la versión gratuita de DataDriven-Construction. La versión de pago de la publicación ha atraído la atención de especialistas de más de 50 países del mundo -desde América Latina hasta la región Asia-Pacífico- durante el primer año tras su publicación. Para discutir las condiciones individuales de cooperación y obtener información detallada sobre las oportunidades de colocación, rellene el formulario de comentarios en el portal oficial datadrivenconstruction.io o escriba a los contactos que figuran al final del libro.

LOS CAPÍTULOS DEL LIBRO ESTÁN DISPONIBLES EN DATADRIVENCONSTRUCTION.IO

Puede leer capítulos de Data-Driven Construction en el sitio web de Data-Driven Construction, donde se van publicando secciones del libro para que pueda encontrar rápidamente la información que necesita y utilizarla en su trabajo. También encontrará muchas otras publicaciones sobre temas similares, así como ejemplos de aplicaciones y soluciones que le ayudarán a desarrollar sus habilidades y aplicar los datos a la construcción.

ÚLTIMAS VERSIONES DEL LIBRO DESCARGAR DESDE EL SITIO WEB OFICIAL

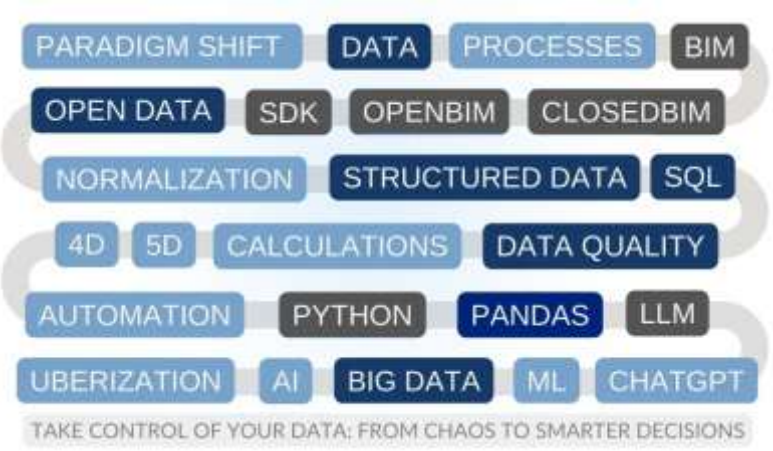
Las versiones actuales y más recientes del libro DataDrivenConstruction están disponibles para su descarga en datadrivenconstruction.io. Si desea recibir actualizaciones con nuevos capítulos del libro, consejos prácticos o reseñas de nuevas aplicaciones, suscríbase al boletín:

- Serás el primero en conocer las nuevas secciones del libro
- Reciba casos prácticos y asesoramiento sobre analítica y automatización en el sector de la construcción
- Manténgase al día de las tendencias, publicaciones y ejemplos de aplicación

Visite datadrivenconstruction.io para suscribirse.

CONSTRUCCIÓN BASADA EN DATOS: CONSULTORÍA, TALLERES Y FORMACIÓN

Los programas de formación y consultoría de DataDrivenConstruction han ayudado a docenas de empresas líderes del sector de la construcción de todo el mundo a aumentar su eficacia, reducir costes y mejorar la calidad de las soluciones. Entre los clientes de DataDrivenConstruction se encuentran algunas de las empresas más importantes de este mercado de miles de millones de euros, como constructoras, consultoras y empresas de TI.



¿Por qué elegirnos?

- **Relevancia:** hablar de las principales tendencias y perspectivas del sector
- **Práctica:** ayudar a los profesionales a resolver eficazmente las tareas cotidianas mediante la PdC.
- **Enfoque personalizado:** adaptado a su empresa, maximizando los beneficios de la formación y el asesoramiento

Las principales áreas de interés del equipo de DataDrivenConstruction:

- **Gestión de la calidad de los datos:** ayuda a parametrizar las tareas, recopilar los requisitos, validar y preparar los datos para el tratamiento automatizado.
- **Minería de datos - extraer y estructurar datos:** configurar procesos ETL y extraer datos de correos electrónicos, PDF, Excel, imágenes y otras fuentes.
- **BIM y CAD analítica**: recopilar, estructurar y analizar información de archivos RVT, IFC, DWG y otros formatos CAD (BIM).
- **Análisis y transformación de datos:** convertir información dispar en datos estructurados, análisis, perspectivas y decisiones.
- **Integración de datos y automatización de procesos:** desde la creación automatizada de documentos hasta la integración con sistemas internos y bases de datos externas.

Póngase en contacto con DataDrivenConstruction.io para saber cómo el uso de la automatización puede ayudar a su empresa a conseguir resultados empresariales tangibles.

GLOSARIO

IA (Inteligencia Artificial) - Inteligencia artificial; capacidad de los sistemas informáticos para realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de patrones, el aprendizaje y la toma de decisiones.

Apache Airflow es una plataforma de orquestación de flujos de trabajo de código abierto que permite crear, programar y realizar el seguimiento de flujos de trabajo y ETL mediante DAG (grafos acíclicos dirigidos).

Apache NiFi es una herramienta para automatizar los flujos de datos entre sistemas, especializada en el enrutamiento y la transformación de datos.

Apache Parquet es un formato de archivo eficiente para el almacenamiento de datos en columnas, optimizado para su uso en sistemas de análisis de big data. Ofrece una compresión significativa y un procesamiento rápido.

API (Interfaz de Programación de Aplicaciones): interfaz formalizada que permite a un programa interactuar con otro sin acceder al código fuente, intercambiando datos y funcionalidades mediante peticiones y respuestas normalizadas.

Atributo: característica o propiedad de un objeto que describe sus rasgos (por ejemplo, superficie, volumen, coste, material).

Las bases de datos son estructuras organizadas de almacenamiento, gestión y acceso a la información, que se utilizan para recuperar y procesar datos de forma eficaz.

BEP (BIM Execution Plan) - Plan de implantación del modelado de información de edificios que define los objetivos, métodos y procesos para implantar BIM en un proyecto.

Big Data: conjuntos de información de gran volumen, variedad y velocidad de actualización, que requieren tecnologías especiales para su procesamiento y análisis.

BI (Business Intelligence) - **Inteligencia** empresarial; los procesos, tecnologías y herramientas para transformar los datos en información significativa para la toma de decisiones.

BIM (Building Information Modeling) - Modelado de información para la construcción; proceso de creación y gestión de representaciones digitales de las características físicas y funcionales de los proyectos de construcción, incluyendo no sólo modelos 3D sino también información sobre rendimiento, materiales, plazos y costes.

BlackBox/WhiteBox - enfoques para comprender el sistema: en el primer caso, la lógica interna está oculta, sólo son visibles las entradas y salidas; en el segundo, el procesamiento es transparente y está disponible para el análisis.

Bounding Box es una construcción geométrica que describe los límites de un objeto en el espacio tridimensional mediante coordenadas X, Y y Z mínimas y máximas, creando una "caja" alrededor del objeto.

BREP (Boundary Representation) es una representación geométrica de objetos que los define a través de los límites de las superficies.

CAD (diseño asistido por ordenador) es un sistema de diseño asistido por ordenador que se utiliza para crear, editar y analizar dibujos precisos y modelos 3D en arquitectura, construcción, ingeniería y otras industrias.

CAFM (Computer-Aided Facility Management) es un software de gestión de propiedades e infraestructuras que incluye planificación de espacios, gestión de activos, mantenimiento y control de costes.

CDE (Entorno Común de Datos): un espacio digital centralizado para gestionar, almacenar, compartir y colaborar con la información de los proyectos en todas las fases del ciclo de vida de las instalaciones.

Un Centro de Excelencia (CdE) es una estructura especializada dentro de una organización encargada de desarrollar un área específica de conocimiento, elaborar normas y buenas prácticas, formar al personal y apoyar la introducción de innovaciones.

CoClass es un moderno sistema de clasificación de elementos de construcción de tercera generación.

Un modelo conceptual de datos es una representación de alto nivel de las entidades básicas y sus relaciones, sin detalle de atributos, que se utiliza en las fases iniciales del diseño de la base de datos.

CRM (Customer Relationship Management) es un sistema de gestión de las relaciones con los clientes que se utiliza para automatizar los procesos de venta y servicio.

DAG (Directed Acyclic Graph) es un grafo acíclico dirigido utilizado en sistemas de orquestación de datos (Airflow, NiFi) para determinar secuencias de tareas y dependencias.

Dash es un framework de Python para crear visualizaciones interactivas de datos basadas en la web.

Cuadro de mandos: un **cuadro de mandos** que presenta visualmente indicadores clave de rendimiento y métricas en tiempo real.

El enfoque centrado en los datos es una metodología que da prioridad a los datos sobre las aplicaciones o el código de software, convirtiendo los datos en el activo central de la organización.

Gobernanza de datos: conjunto de prácticas, procesos y políticas que garantizan el uso adecuado y eficaz de los datos en una organización, incluidos los controles de acceso, calidad y seguridad.

Data Lake es una instalación de almacenamiento diseñada para guardar grandes cantidades de datos en bruto en su formato original hasta que se utilicen.

Data Lakehouse es un enfoque arquitectónico que combina la flexibilidad y escalabilidad de los lagos de datos (Data **Lake**) con la capacidad de gestión y rendimiento de los almacenes de datos (DWH).

La construcción basada en datos es un enfoque estratégico en el que cada etapa del ciclo de vida de las instalaciones -desde el diseño hasta las operaciones- se apoya en sistemas automatizados e interconectados. Este enfoque proporciona un aprendizaje continuo basado en hechos, reduce la incertidumbre y permite a las empresas alcanzar un liderazgo sostenible en el sector.

Integrador de datos: empresa especializada en combinar datos de fuentes dispares y analizarlos para tomar decisiones de gestión.

Enfoque basado en los datos: una metodología en la que los datos se consideran un activo estratégico y las decisiones se toman basándose en un análisis objetivo de la información y no en opiniones subjetivas.

Minimalismo de datos: un enfoque para reducir los datos a los más valiosos y significativos, lo que permite simplificar el tratamiento y análisis de la información.

Pantano de datos - Masa dispersa de datos no estructurados que se produce cuando la información se recopila y almacena de forma incontrolada sin una organización adecuada.

DataOps es una metodología que combina principios DevOps, datos y analítica, centrada en mejorar la colaboración, integración y automatización de los flujos de datos.

La digitalización de **la información** es el proceso de convertir todos los aspectos de la actividad de la construcción en un formato digital apto para el análisis, la interpretación y la automatización.

DataFrame - Una estructura de datos tabular bidimensional en la biblioteca Pandas, donde las filas representan registros individuales u objetos y las columnas representan sus características o atributos.

Análisis **descriptivo**: análisis de datos históricos para comprender lo que ocurrió en el pasado.

Análisis de **diagnóstico**: análisis de datos para determinar por qué ha ocurrido algo.

Un diagrama de Gantt es una herramienta de planificación de proyectos que representa las tareas como barras horizontales en una línea de tiempo, lo que permite visualizar la secuencia y duración del trabajo.

DWH (Data Warehouse) es un sistema de almacén de datos centralizado que agrega información de múltiples fuentes, la estructura y la pone a disposición para análisis e informes.

ESG (Environmental, Social, Governance): conjunto de criterios para evaluar el impacto medioambiental, social y de gobernanza de una empresa o proyecto.

ELT (Extraer, Cargar, Transformar) es un proceso en el que los datos se extraen primero de las fuentes, se cargan en un repositorio y luego se transforman con fines analíticos.

ETL (Extract, Transform, Load) es el proceso de extraer datos de diversas fuentes, transformarlos al formato deseado y cargarlos en el almacenamiento de destino para su análisis.

Diagrama ER (Entidad-Relación): diagrama visual que muestra las entidades, sus atributos y las relaciones entre ellas, utilizado en la modelización de datos.

ERP (Enterprise Resource Planning) es un completo sistema modular de planificación de recursos empresariales utilizado para gestionar y optimizar diversos aspectos del proceso de construcción.

Características - En el aprendizaje automático, variables independientes o atributos utilizados como entradas de un modelo.

Modelo físico de datos: representación detallada de la estructura de la base de datos, con tablas, columnas, tipos de datos, claves e índices, optimizada para un SGBD concreto.

FPDF es una biblioteca de Python para crear documentos PDF.

Geometric Core es un componente de software que proporciona algoritmos básicos para crear, editar y analizar objetos geométricos en CAD, BIM y otras aplicaciones de ingeniería.

HiPPO (Highest Paid Person's Opinion): enfoque de la toma de decisiones basado en la opinión de la persona mejor pagada de la organización, en lugar de en datos objetivos.

IDE (entorno de desarrollo integrado): entorno de desarrollo integrado, una herramienta completa para escribir, probar y depurar código (por ejemplo, PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) es una especificación de entrega de información que define los requisitos de datos en las distintas fases de un proyecto.

IFC (Industry Foundation Classes) es un formato de intercambio de datos BIM que garantiza la interoperabilidad entre distintas soluciones de software.

La Industria 5.0 es un concepto de desarrollo industrial que combina las capacidades de la digitalización, la automatización y la inteligencia artificial con el potencial humano y la sostenibilidad medioambiental.

La integración de datos es el proceso de combinar datos de distintas fuentes en un sistema único y coherente para ofrecer una visión unificada de la información.

Los silos de información son sistemas de almacenamiento de datos aislados que no comparten información con otros sistemas, lo que crea barreras para el uso eficiente de los datos.

IoT (Internet de los objetos) es el concepto de conectar objetos físicos a Internet para recoger, procesar y transmitir datos.

k-NN (k-Nearest Neighbors) es un algoritmo de aprendizaje automático que clasifica objetos basándose en la similitud con los vecinos más próximos de la muestra de entrenamiento.

Kaggle es una plataforma de concursos de análisis de datos y aprendizaje automático.

Cálculo: cálculo del coste de las obras o procesos de construcción para una determinada unidad de medida (por ejemplo, 1 m^2 de pared de cartón yeso, 1 m^3 de hormigón).

Los KPI (Key Performance Indicators) son indicadores clave de rendimiento, métricas cuantificables utilizadas para evaluar el éxito de una empresa o un proyecto concreto.

Etiquetas - En el aprendizaje automático, las variables o atributos objetivo que el modelo debe predecir.

Algoritmo de aprendizaje - Proceso de búsqueda de la mejor hipótesis de un modelo correspondiente a una función objetivo a partir de un conjunto de datos de entrenamiento.

Regresión lineal - Método estadístico de modelización de la relación entre una variable dependiente y una o más variables independientes.

LLM (Large Language Model) - Gran Modelo Lingüístico, una inteligencia artificial entrenada para comprender y generar texto a partir de grandes cantidades de datos, capaz de analizar el contexto y escribir código de programación.

LOD (Level of Detail/Development) - nivel de detalle del modelo que determina el grado de precisión geométrica y el contenido de información.

Un modelo lógico de datos es una descripción detallada de entidades, atributos, claves y relaciones que refleja la información y las reglas de negocio, una etapa intermedia entre los modelos conceptual y físico.

Aprendizaje automático - Clase de técnicas de inteligencia artificial que permiten a los sistemas informáticos aprender y hacer predicciones a partir de datos sin programación explícita.

Masterformat es un sistema de clasificación de primera generación utilizado para estructurar las especificaciones de construcción en secciones y disciplinas.

MEP (Mechanical, Electrical, Plumbing) - Sistemas de ingeniería de edificios que incluyen componentes mecánicos, eléctricos y de fontanería.

La malla es una representación en malla de objetos 3D formada por vértices, aristas y caras.

Modelo - En el aprendizaje automático, conjunto de hipótesis diferentes, una de las cuales se aproxima a la función objetivo que debe predecirse o aproximarse.

El modelado de datos es el proceso de creación de una representación estructurada de los datos y sus relaciones para su aplicación en sistemas de información, incluidos los niveles conceptual, lógico y físico.

n8n es una herramienta de código abierto para automatizar flujos de trabajo e integrar aplicaciones mediante un enfoque de bajo código.

Normalización: en el aprendizaje automático, el proceso de llevar datos numéricos diferentes a una escala común para facilitar su procesamiento y análisis.

Ingeniería inversa - proceso de estudio del dispositivo, funcionamiento y tecnología de fabricación de un objeto mediante el análisis de su estructura, funciones y funcionamiento. En el contexto de los datos: extraer información de formatos propietarios para utilizarla en sistemas abiertos.

OCR (Optical Character Recognition) es una tecnología de reconocimiento **óptico** de **caracteres** que convierte imágenes de texto (documentos escaneados, fotos) en un formato de texto legible por máquina.

OmniClass es una norma internacional de clasificación de segunda generación para la gestión de la información sobre la construcción.

Ontología - Sistema de interrelaciones de conceptos que formaliza un determinado campo del conocimiento.

Código abierto: modelo de desarrollo y distribución de software de código abierto disponible para su libre uso, estudio y modificación.

El concepto de BIM abierto implica el uso de normas y formatos abiertos para el intercambio de datos entre distintas soluciones informáticas.

Estándares abiertos: especificaciones disponibles públicamente para realizar una tarea específica que permiten a diferentes sistemas interactuar e intercambiar datos.

Pandas es una biblioteca de código abierto de Python para el procesamiento y análisis de datos, que proporciona estructuras de datos DataFrame y Series para el manejo eficiente de información tabular.

El paradigma de los datos abiertos es un enfoque del tratamiento de datos en el que la información se pone a libre disposición de cualquiera para su uso, reutilización y difusión.

El método **paramétrico** es un método de estimación de proyectos de construcción que utiliza modelos estadísticos para estimar el valor en función de los parámetros del proyecto.

PIMS (Project Information Model) es un sistema digital diseñado para organizar, almacenar y compartir toda la información del proyecto.

Pipeline - Secuencia de procesos de tratamiento de datos, desde la extracción y transformación hasta el análisis y la visualización.

PMIS (Project Information Management System) es un sistema de gestión de proyectos diseñado para el control detallado de tareas a nivel de un proyecto de construcción individual.

El análisis **predictivo** es una sección del análisis que utiliza métodos estadísticos y aprendizaje automático para predecir resultados futuros basándose en datos históricos.

Análisis **prescriptivo**: una sección de análisis que no sólo predice resultados futuros, sino que también sugiere acciones óptimas para lograr los resultados deseados.

Los formatos **propietarios** son formatos de datos cerrados controlados por una empresa concreta que limitan la capacidad de compartir información y aumentan la dependencia de un software específico.

QTO (Quantity Take-Off) es el proceso de extraer las cantidades de elementos de los documentos de diseño para calcular las cantidades de materiales necesarias para realizar el proyecto.

Sistema de gestión de la calidad: sistema de gestión de la calidad que garantiza que los procesos y los resultados cumplen los requisitos establecidos.

RAG (Retrieval-Augmented Generation) es un método que combina las capacidades generativas de los modelos lingüísticos con la extracción de información relevante de bases de datos corporativas, mejorando la precisión y relevancia de las respuestas.

RDBMS (Relational Database Management System) es un sistema de gestión de bases de datos relacionales que organiza la información en forma de tablas interrelacionadas.

RegEx (expresiones regulares) es un lenguaje formalizado para buscar y procesar cadenas, que permite especificar plantillas para comprobar si los datos de texto cumplen determinados criterios.

La regresión es un método estadístico de análisis de la relación entre variables.

Los cálculos de CO_2 son un método para estimar las emisiones de dióxido de carbono asociadas a la producción y el uso de materiales y procesos de construcción.

El método de los recursos es un método de elaboración de estimaciones basado en un análisis detallado de todos los recursos (materiales, mano de obra, equipos) necesarios para realizar los trabajos de construcción.

La RFID (identificación por radiofrecuencia) es una tecnología de identificación automática de objetos mediante señales de radio, utilizada para el seguimiento de materiales, maquinaria y personal.

ROI (Return on Investment) es un indicador que refleja la relación entre el beneficio y los fondos invertidos, utilizado para evaluar la eficacia de las inversiones.

SaaS (Software as a Service**)** es un modelo de software como servicio en el que las aplicaciones son alojadas por un proveedor y puestas a disposición de los usuarios a través de Internet.

SCM (Supply Chain Management): gestión de la cadena de suministro, que incluye la coordinación y optimización de todos los procesos, desde la adquisición de materiales hasta la entrega de productos acabados.

Los silos de datos son almacenes aislados de información dentro de una organización que no están integrados con otros sistemas, lo que dificulta el intercambio de datos y los hace ineficaces.

SQL (Structured Query **Language)** es un lenguaje de consulta estructurado que se utiliza para trabajar con bases de datos relacionales.

SQLite es un SGBD ligero, integrable y multiplataforma que no requiere un servidor independiente y admite funciones SQL básicas, muy utilizado en aplicaciones móviles y sistemas integrados.

Datos estructurados: información organizada en un formato específico con una estructura clara, como en bases de datos relacionales o tablas.

Datos poco estructurados: información con una organización parcial y una estructura flexible, como JSON o XML, en la que distintos elementos pueden contener diferentes conjuntos de atributos.

Una entidad es un objeto concreto o abstracto del mundo real que puede identificarse, describirse y representarse unívocamente en forma de datos.

Aprendizaje **supervisado** - Tipo de aprendizaje automático en el que un algoritmo se entrena con datos divididos en los que se conoce el resultado deseado para cada ejemplo.

La taxonomía es un sistema de clasificación jerárquica utilizado para categorizar sistemáticamente elementos basándose en características comunes.

Titanic Dataset es un popular conjunto de datos para entrenar y probar modelos de aprendizaje automático.

Entrenamiento - Proceso en el que un algoritmo de aprendizaje automático analiza los datos para identificar patrones y formar un modelo.

El aprendizaje por transferencia es una técnica de aprendizaje automático en la que un modelo entrenado para una tarea se utiliza como punto de partida para otra tarea.

Transformación - Proceso de cambiar el formato, la estructura o el contenido de los datos para su uso posterior.

Requisitos de datos: criterios formalizados que definen la estructura, el formato, la integridad y la calidad de la información necesaria para apoyar los procesos empresariales.

La Uberización del sector de la construcción es el proceso de transformación de los modelos de negocio tradicionales en la construcción bajo la influencia de plataformas digitales que proporcionan una interacción directa entre clientes y contratistas sin intermediarios.

Uniclass es un sistema de clasificación de elementos de construcción de segunda y tercera generación ampliamente utilizado en el Reino Unido.

USD (Universal Scene Description) es un formato de datos desarrollado para gráficos por ordenador, pero que ha ganado aplicación en sistemas de ingeniería debido a su estructura simple e independencia de núcleos geométricos.

La validación de datos es el proceso de cotejar la información con los criterios y requisitos establecidos para garantizar la exactitud, integridad y coherencia de los datos.

Base de datos **vectorial** - Tipo especializado de base de datos que almacena datos como vectores multidimensionales para una búsqueda semántica y una comparación de objetos eficaces.

La representación vectorial (embedding) es un método de transformación de datos en vectores numéricos multidimensionales que permite a los algoritmos de las máquinas procesar y analizar eficazmente la información.

VectorOps es una metodología centrada en el procesamiento, almacenamiento y análisis de datos vectoriales multidimensionales, especialmente relevante en áreas como los gemelos digitales y la búsqueda semántica.

Visualización - Representación gráfica de datos para una mejor percepción y análisis de la información.

La categorización alfabética de los términos se hizo por sus nombres en inglés.

LISTA DE REFERENCIAS Y MATERIAL EN LÍNEA

[1] Gartner, "Datos de métricas clave de TI 2017: Índice de documentos y métricas publicados", 12 de diciembre de 2016. [En línea]. Disponible: https://www.gartner.com/en/documents/3530919. [Fecha de acceso: 1 de marzo de 2025].

[2] KPMG, "Desafíos conocidos - nuevos enfoques. 2023 Global Construction Survey", 1 de enero de 2023. [En línea]. Disponible: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [Consultado: 5 de marzo de 2025].

[3] F. R. Barnard, "Una imagen vale más que mil palabras", 10 Mapi 1927. [en línea]. Disponible: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Fecha del discurso: 15 de marzo de 2025].

[4] M. Bastian, "Satya Nadella, consejero delegado de Microsoft, dice que autoproclamarse AGI es una 'piratería sin sentido'", 21 feb. 2025. [En línea]. Disponible: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Fecha de acceso: 15 de marzo de 2025.]

[5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1 de enero de 2016. [En línea]. Disponible: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Fecha de acceso: 2 de marzo de 2025].

[6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [En línea]. Disponible: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papiro, siglo III a.C. La lengua es el griego", 2024. [en línea]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Vigilancia: utilizar las herramientas disponibles", 1980. [En línea]. Disponible: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Fecha de la ponencia: 15 de marzo de 2025].

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15 de febrero de 2015. [En línea]. Disponible: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Fecha de acceso: 15 de marzo de 2025].

[10] Skanska USA, "Tendencias del mercado de la construcción en otoño", 2 de noviembre de 2023. [En línea]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Oxford Essential Quotations (4ª ed.)", Oxford University Press, 2016. [En línea]. Available:

https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236. [Fecha de acceso: 1 de marzo de 2025].

[12] "Cita: Sondergaard sobre el análisis de datos". [en línea]. Disponible: https://www.causeweb.org/cause/resources/library/r2493. [Fecha de acceso: 15 de marzo de 2025].

[13] "Cómo el interés mundial por la IA está impulsando el mercado de la gestión de datos", 28 de mayo de 2024. [En línea]. Disponible: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Fecha de acceso: 15 de marzo de 2025].

[14] И. McCue, "Historia de ERP", 2024. [En línea]. Disponible: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "ERP Price: ¿Cuánto cuesta un sistema ERP?"," [en línea]. Disponible: https://erpscout.de/en/erp-costs/. [Fecha de acceso: 15 de marzo de 2025].

[16] softwarepath, "What 1,384 ERP projects tell us about selecting ERP (2022 ERP report)", 18 de enero de 2022. [en línea]. Disponible: https://softwarepath.com/guides/erp-report. [Fecha de acceso: 15 de marzo de 2025].

[17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16 de diciembre de 2016. [En línea]. Disponible: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Fecha de la dirección: 1 de marzo de 2025].

[18] Mckinsey, "The data-driven enterprise of 2025", 28 de enero de 2022. [En línea]. Disponible: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Fecha de consulta: 22 de mayo de 2024].

[19] Wikipedia, "Ley de Moore". [en línea]. Disponible: https://en.wikipedia.org/wiki/Moore%27s_law. [Fecha de acceso: 15 de marzo de 2025].

[20] Accenture, "Building More Value With Capital Projects", 1 de enero de 2020. [En línea]. Disponible: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Fecha de acceso: 3 de marzo de 2024].

[21] Б. Marr, "¿Cuántos datos creamos cada día? Las alucinantes estadísticas que todo el mundo debería leer", 2018. [En línea]. Disponible: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read.

[22] "¿Cuántos datos se producen cada día?", 2024. [En línea]. Disponible:

https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T. Sullivan, "La IA y la 'datasfera' mundial: ¿cuánta información tendrá la humanidad en 2025? [En línea]. Disponible: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Número total de libros impresos producidos en diversas regiones de Europa Occidental en cada medio siglo entre 1454 y 1800", [Internet]. [en línea]. Disponible: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Fecha de consulta: 1 de marzo de 2025].

[25] "Ejemplos de fijación de precios", 2024. [En línea]. Disponible: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Enterprises outsource data storage as complexity rises", 10 de mayo de 2024. [En línea]. Disponible: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Fecha de acceso: 15 de marzo de 2025.]

[27] JETSOFTPRO, "¿El SaaS ha muerto? Microsoft CEO's Shocking Prediction Explained", 13 de enero de 2025. [En línea]. Disponible: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12 de diciembre de 2024. [En línea]. Disponible: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Fecha de acceso: 15 de marzo de 2025].

[29] GoodReads, "Tim Berners-Lee", [en línea]. Disponible: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Fecha de acceso: 15 de marzo de 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1 de enero de 2023. [En línea]. Disponible: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Consultado: 5 de marzo de 2025].

[31] И. Deininger, B. Koch, R. Bauknecht y M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model", 2024. [En línea]. Disponible: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.

[32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1 de febrero de 2017. [En línea]. Disponible: https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-

construction-a-route-to-higher-productivity-full-report.pdf.

[33] Construction Task Force to the Deputy Prime Minister, "Rethinking Construction", 1 de octubre de 2014. [En línea]. Disponible: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

[34] Forbes, "Sin opinión, no eres más que otra persona con datos", 15 mar. 2016. [En línea]. Disponible: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/. [Fecha de acceso: 15 de marzo de 2025].

[35] Wikiquote, "Charles Babbage", [en línea]. Disponible: https://en.wikiquote.org/wiki/Charles_Babbage. [Fecha de acceso: 15 de marzo de 2025].

[36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves," 12 mar. 2025. [en línea]. Disponible: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Fecha de acceso: 15 de marzo de 2025].

[37] Asociación Canadiense de la Construcción y KPMG en Canadá, 2021, "La construcción en un mundo digital", 1 de mayo de 2021. [En línea]. Disponible: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Fecha de acceso: 5 de marzo de 2025.]

[38] ZCS, "Descifrando la Quinta Revolución Industrial". [en línea]. Disponible: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Fecha de acceso: 15 de marzo de 2025].

[39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.

[40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "El valor del software de código abierto", 24 ene. 2024. [En línea]. Disponible: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Fecha de acceso: 15 de marzo de 2025].

[41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1 de septiembre de 2008. [en línea]. Disponible: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Mejorar la productividad de la construcción", [en línea]. Disponible: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Fecha de la dirección: 15 de marzo de 2025].

[43] A. G. a. C. Syverson, "La extraña y horrible trayectoria de la productividad en el sector de la construcción de EE. [En línea]. Disponible: https://bfi.uchicago.edu/insight/research-summary/the-

strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Fecha de acceso: 1 de marzo de 2025].

[44] McKinsey, "Cumplir con la productividad de la construcción ya no es opcional", 9 de agosto de 2024. [En línea]. Disponible: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Fecha de acceso: 5 de marzo de 2025.]

[45] ING Group, "Lagging productivity in construction is driving up construction costs", 12 de diciembre de 2022. [En línea]. Disponible: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/. [Fecha de la dirección: 15 de marzo de 2025].

[46] M. Berman, "Impactante predicción del CEO de Microsoft: 'Los agentes sustituirán a TODO el software'", 19 de diciembre de 2024. [En línea]. Disponible: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Fecha de acceso: 15 de marzo de 2025.]

[47] Business Insider, "Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were in charge of", 15 mar 2025. [En línea]. Disponible: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Fecha de acceso: 30 de marzo de 2025.]

[48] Statista, "Comparación de la popularidad de los sistemas de gestión de bases de datos (SGBD) en todo el mundo a junio de 2024, por categoría", junio de 2024. [En línea]. Disponible: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Fecha de la dirección: 15 de marzo de 2025].

[49] DB-Engines, "Clasificación DB-Engines", [en línea]. Disponible: https://db-engines.com/en/ranking. [Fecha de consulta: 15 de marzo de 2025.]

[50] "Stack Overflow Developer Survey 2023", 2024. [En línea]. Disponible: https://survey.stackoverflow.co/2023/.

[51] "SQL", 2024. [En línea]. Disponible: https://en.wikipedia.org/wiki/SQL.

[52] "Datos estructurados y no estructurados: ¿cuál es la diferencia? [En línea]. Disponible: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF COMPARACIÓN DE FORMATOS DE DATOS PARA PROYECTOS DE CONSTRUCCIÓN", 23 abr 2024. [en línea]. Disponible: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] "Building Information Modeling Whitepaper site", 2003. [En línea]. Disponible: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boiko, "Las guerras de lobbies y el desarrollo de BIM. Parte 5: BlackRock es el amo de todas las tecnologías. Cómo controlan las corporaciones el código fuente abierto", 2024. [En línea]. Disponible: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14 11 2011. [En línea]. Disponible: https://isicad.net/articles.php?article_num=14805. [Fecha de la dirección: 02 2025].

[57] C. Eastman y A. Cthers, "Eastman, Charles; And Cthers", septiembre de 1974. [en línea]. Disponible: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Fecha de acceso: 15 de marzo de 2025].

[58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11 de noviembre de 2011. [En línea]. Disponible: https://isicad.net/articles.php?article_num=14805. [Fecha de consulta: 15 de marzo de 2025].

[59] D. Weisberg, "Historia de CAD", 12 de diciembre de 2022. [en línea]. Disponible: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Fecha de consulta: 15 de marzo de 2025].

[60] ADSK, "Libro Blanco Building Information Modeling", 2002. [En línea]. Disponible: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Fecha de consulta: 15 de marzo de 2025].

[61] ADSK, "Libro Blanco Building Information Modeling in Practice", [en línea]. Disponible: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Fecha de acceso: 15 de marzo de 2025].

[62] A. Boiko, "Lobbying wars and BIM development. Parte 2: BIM abierto VS BIM cerrado. Europa VS el resto del mundo", 2024. [En línea]. Disponible: https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/.

[63] A. Boiko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs", 2024. [En línea]. Disponible: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Libro Blanco BIM", 2002. [En línea]. Disponible: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Fecha de la dirección: 15 de marzo de 2025].

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale", [en línea]. Disponible: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Fecha de acceso: 15 de marzo de 2025.]

[66] M. Shacklett, "Datos estructurados y no estructurados: Diferencias clave", 2024. [En línea].

Disponible: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Dar sentido al crecimiento de los datos no estructurados", 2024. [En línea]. Disponible: https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/.

[68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [En línea]. Disponible: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf. [Fecha de acceso: 02 2025].

[69] CrowdFlower, "Informe sobre ciencia de datos 2016", 2016. [En línea]. Disponible: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Fecha de acceso: 15 de marzo de 2025].

[70] Analyticsindiamag, "Las 6 tareas que más tiempo consumen a los científicos de datos", 15 de mayo de 2019. [En línea]. Disponible: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Informe: los científicos de datos pasan la mayor parte del tiempo limpiando", 06 de julio de 2015. [En línea]. Disponible: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html. [Fecha de acceso: 5 de marzo de 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 de julio de 2015. [En línea]. Disponible: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Fecha de acceso: 15 de marzo de 2025].

[73] Б. Cyphers y K. Doctorow, "Privacidad sin monopolio: protección de datos e interoperabilidad", 2024. [En línea]. Disponible: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1 de octubre de 2013. [En línea]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Fecha de acceso: 15 de marzo de 2025].

[75] A. Boiko, "La lucha por los datos abiertos en la industria de la construcción. La historia de AUTOLISP, intelliCAD, openDWG, ODA y openCASCADE", 15 05 2024. [En línea]. Disponible: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Fecha de la ponencia: 16 02 2025].

[76] Wikipedia, "Microsoft y el código abierto", [en línea]. Disponible: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Fecha de acceso: 15 de marzo de 2025].

[77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking. Here's Why That Matters",

5 de noviembre de 2024. [En línea]. Disponible: https://time.com/7171962/open-closed-ai-models-epoch/. [Fecha de acceso: 15 de marzo de 2025].

[78] The Verge, "More than a quarter of new code at Google is generated by AI", 29 oct. 2024. [En línea]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Fecha de acceso: 15 de marzo de 2025].

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15 de diciembre de 2020. [En línea]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing. [Consultado: 15 de marzo de 2025].

[80] PWC, "PwC Open Source Monitor 2019", 2019. [En línea]. Disponible: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [Consultado el 15 de mar. de 2025].

[81] Travers Smith, "El secreto a voces: el software de código abierto", 2024. [En línea]. Disponible: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Fecha de acceso: 15 de marzo de 2025].

[82] Deloitte, "El proceso de transferencia de datos en las transformaciones empresariales", 2021. [En línea]. Disponible: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf. [Fecha de consulta: 15 de marzo de 2025].

[83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20 mar. 2024. [En línea]. Disponible: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Fecha de acceso: 5 de marzo de 2025].

[84] "Cita de Origen: Todo debe simplificarse al máximo, pero no simplificarse", 13 de mayo de 2011. [En línea]. Disponible: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Fecha de acceso: 15 de marzo de 2025].

[85] "Transformer (arquitectura de aprendizaje profundo)". [en línea]. Disponible: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Fecha de acceso: 15 de marzo de 2025].

[86] "Estadísticas de descarga de paquetes Python", 2024. [En línea]. Disponible: https://www.pepy.tech/projects/pandas.

[87] Interview Bit, "Las 10 mejores bibliotecas de Python", 2023. [En línea]. Disponible: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Fecha de acceso: 30 de marzo de 2025].

[88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations," 7 mar. 2025. [En línea]. Disponible: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Fecha de acceso: 15 de marzo de 2025].

[89] P. Orac, "Cómo procesar un DataFrame con millones de filas en segundos", 2024. [En línea]. Disponible: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "¿Qué es Kaggle?", 2024. [en línea]. Disponible: https://www.datacamp.com/blog/what-is-kaggle.

[91] "Conferencia de Jensen Huang, CEO de NVIDIA, en COMPUTEX 2024", 2 de junio de 2024. [En línea]. Disponible: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [Consultado el 15 de marzo de 2025].

[92] "Miembros: Fundadores y miembros corporativos", 2024. [En línea]. Disponible: https://www.opendesign.com/member-showcase.

[93] A. Boiko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 de noviembre de 2024. [En línea]. Disponible: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Consultado: 23 de febrero de 2025].

[94] A. Boiko, "El mundo post-BIM. Transición a datos y procesos y si el sector de la construcción necesita semántica, formatos e interoperabilidad", 20 dic. 2024. [En línea]. Disponible: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Fecha de la ponencia: 23 de febrero de 2025].

[95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE," 2016. [En línea]. Disponible: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Fecha de acceso: 23 de febrero de 2025].

[96] Harvard Business Review, "Bad Data Costs the U.S. $3 Trillion Per Year", 22 de septiembre de 2016. [En línea]. Disponible: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Impacts of Data Quality", 1 de enero de 2025. [En línea]. Disponible: https://delpha.io/impacts-of-data-quality/.

[98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects", [Diseño para el mantenimiento: la importancia de las consideraciones sobre operaciones y mantenimiento durante la fase de diseño de los proyectos de construcción]. [en línea]. Disponible: https://www.wbdg.org/resources/design-for-maintainability. [Fecha de acceso: 15 de marzo de 2025].

[99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", abril de 2014. [En línea]. Disponible: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Fecha de acceso: 15 de marzo de 2025].

[100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1 de enero de 2025. [En línea]. Disponible: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "A falta de un clavo". [en línea]. Disponible: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Fecha de acceso: 15 de marzo de 2025].

[102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", octubre de 2013. [En línea]. Disponible: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Fecha de acceso: 15 de marzo de 2025].

[103] EY, "El camino hacia la neutralidad del carbono", 10 mar. 2023. [En línea]. Disponible: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Fecha de acceso: 15 de marzo de 2025].

[104] PWC, "ESG Awareness", 1 de julio de 2024. [En línea]. Disponible: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Fecha de acceso: 15 de marzo de 2025].

[105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [En línea]. Disponible: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_calcular el carbono incorporado", 2024. [en línea]. Disponible: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Imagining Construction's Digital Future", 24 de junio de 2016. [En línea]. Disponible: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Fecha de acceso: 25 de febrero de 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 de octubre de 2024. [en línea]. Disponible: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Fecha de la dirección: 15 de marzo de 2025].

[109] SAS, "Data lake and data warehouse - know the difference", [en línea]. Disponible: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Fecha de consulta: 15 de marzo de 2025].

[110] ADSK, "Modelado de información para la construcción", 2002. [En línea]. Disponible: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Fecha de la dirección: 15 de marzo de 2025].

[111] A. Boiko, "Mapa histórico de BIM", 2024. [En línea]. Disponible: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definiciones de BIM por organizaciones y normas", 27 de diciembre de 2023. [En línea]. Disponible: https://encyclopedia.pub/entry/53149. [Fecha de la dirección: 5 de marzo de 2025].

[113] CAD Vendor, "Programa OPEN BIM", 2012. [En línea]. Disponible: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Fecha de consulta: 30 de marzo de 2025].

[114] Wikipedia, "Clases de fundamentos de la industria". [en línea]. Disponible: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Fecha de acceso: 15 de marzo de 2025].

[115] Wikipedia, "IGES". [en línea]. Disponible: https://en.wikipedia.org/wiki/IGES. [Fecha de consulta: 30 de marzo de 2025].

[116] A. Boiko, "Historia del CAD (BIM)", 15 de diciembre de 2021. [En línea]. Disponible: https://miro.com/app/board/o9J_laML2cs=/. [Fecha de la dirección: 24 de febrero de 2025].

[117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I", 8 ene. 2021. [En línea]. Disponible: https://arxiv.org/pdf/2007.10951. [Fecha de acceso: 5 de marzo de 2025].

[118] И. Rogachev, "Hablemos de BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 13 de abril de 2021. [En línea]. Disponible: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be. [Fecha de la ponencia: 5 de marzo de 2025.]

[119] D. Ares, "El RETS en el sector inmobiliario: por qué es crucial para la eficiencia y el crecimiento", 17 de diciembre de 2024. [En línea]. Disponible: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Fecha de acceso: 5 de marzo de 2025].

[120] "Coste de las fichas flexibles", 2024. [En línea]. Disponible: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boiko, "Olvidar el BIM y democratizar el acceso a los datos (17. Kolloquium Investor - Hochschule - Bauindustrie)", 2024. [En línea]. Disponible: https://www.bim.bayern.de/wp-

content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagment-und-Bay-Bauindustrie.pdf.

[122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh y F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [En línea]. Disponible: https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle.

[123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle", 1 de enero de 2018. [En línea]. Disponible: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Fecha de acceso: 2 mar 2025].

[124] Apple.com, "Pixar, Adobe, Apple y NVIDIA forman una alianza para OpenUSD con el fin de impulsar estándares abiertos para contenidos 3D", 1 de agosto de 2023. [En línea]. Disponible: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [Consultado: 2 de marzo de 2025].

[125] AECmag, "La estrategia de datos granulares de ADSK", 25 jul 2024. [En línea]. Disponible: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Fecha de acceso: 15 de marzo de 2025].

[126] A. Boiko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 11 2024. [En línea]. Disponible: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Fecha de publicación: 23 de febrero de 2025].

[127] A. Boiko, "ENG BIM Cluster 2024 | La batalla por los datos y la aplicación de LLM y ChatGPT en la construcción", 7 de agosto de 2024. [En línea]. Disponible: ENG BIM Cluster 2024 | La batalla por los datos y la aplicación de LLM y ChatGPT en la construcción. [Consultado: 15 de marzo de 2025].

[128] "Jeffrey Zeldman presenta", 6 de mayo de 2008. [En línea]. Disponible: https://zeldman.com/2008/05/06/content-precedes-design/. [Fecha de acceso: 15 de marzo de 2025].

[129] A. Boiko, "DWG Analyse with ChatGPT | DataDrivenConstruction", 5 mar. 2024. [En línea]. Disponible: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Fecha de consulta: 15 de marzo de 2025].

[130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [En línea]. Disponible: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Fecha de la dirección: 30 de marzo de 2025].

[131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31 de marzo de 2016. [En línea]. Disponible: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Fecha de acceso: 15 de marzo de 2025].

[132] J. Bertin, "Gráficos y tratamiento gráfico de la información", 8 de septiembre de 2011. [En línea]. Disponible: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Fecha de acceso: 15 de marzo de 2025].

[133] CauseWeb, "Wells/Wilks on Statistical Thinking", [en línea]. Disponible: https://www.causeweb.org/cause/resources/library/r1266. [Fecha de acceso: 15 de marzo de 2025].

[134] Ministrymagazine, "Cómo la ciencia descubrió la Creación", enero de 1986. [en línea]. Disponible: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Fecha de consulta: 15 de marzo de 2025].

[135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23 de mayo de 2017. [En línea]. Disponible: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [Consultado el 15 de mayo de 2024].

[136] "How to build a data architecture to drive innovation-today and tomorrow", 3 de junio de 2020. [En línea]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Fecha del discurso: 15 de marzo de 2025].

[137] Oxford, "Woodrow Wilson 1856-1924". [en línea]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Fecha del discurso: 15 de marzo de 2025].

[138] "Convertidores", 2024. [en línea]. Disponible: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1 de enero de 2017. [En línea]. Disponible: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Fecha de acceso: 18 de febrero de 2025].

[140] "Tuberías en construcción", 2024. [en línea]. Disponible: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, "Apache NiFi", 1 de enero de 2025. [En línea]. Disponible: https://de.wikipedia.org/wiki/Apache_NiFi. [Fecha de acceso: 5 de marzo de 2025].

[142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails," 1 de mayo de 2024. [En línea]. Disponible: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/. [Consultado el 15 de marzo de 2025].

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 1 mar 2025. [En

línea]. Disponible: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Fecha de acceso: 15 de marzo de 2025.]

[144] B. T. O'Neill, "Failure rates for analytics, AI, and big data projects = 85% - yikes!", 1 de enero de 2025. [En línea]. Disponible: https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/.

[145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.

[146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1 de enero de 2025. [En línea]. Disponible: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Fecha de acceso: 1 de mayo de 2024].

[147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.

[148] A. Boiko, "San Francisco. Sector de la construcción 1980-2019", 2024. [En línea]. Disponible: https://www.kaggle.com/search?q=San+Francisco.+Sector+de+la+construcción+1980-2019.

[149] A. Boiko, "Kaggle: RVT IFC presenta 5000 proyectos", 2024. [En línea]. Disponible: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Prepararse para el futuro con la construcción conectada". [en línea]. Disponible: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Consultado el 15 de marzo de 2025].

[151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22 de mayo de 2017. [En línea]. Disponible: https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html.

[152] "Condiciones necesarias para la conservación de fósiles vegetales", 2024. [en línea]. Disponible: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "BlackRock's Fink on bonds, mergers and acquisitions, the U.S. recession, and the election: Full Interview," 2023. [En línea]. Disponible: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 famosos desastres de la IA", 02 de octubre de 2024. [en línea]. Disponible: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Fecha de acceso: 15 de marzo de 2025].

[155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.

[156] Wikipedia, "Kaggle", 1 de enero de 2025. [En línea]. Disponible: https://en.wikipedia.org/wiki/Kaggle. [Fecha de acceso: 15 de marzo de 2025].

[157] Kaggle, "Titanic - Aprendizaje automático del desastre", 1 de enero de 2025. [En línea]. Disponible: https://www.kaggle.com/competitions/titanic/overview. [Consultado el 10 de marzo de 2025].

[158] Ш. Johri, "Creación de ChatGPT: de los datos al diálogo", 2024. [En línea]. Disponible: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Algunas cosas útiles que hay que saber sobre el aprendizaje automático", 2024. [En línea]. Disponible: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Cita citable", [en línea]. Disponible: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Fecha de acceso: 17 de marzo de 2025].

[161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models," 19 de diciembre de 2024. [En línea]. Disponible: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [Consultado el 25 de marzo de 2025].

[162] "NVIDIA anuncia una nueva versión de los modelos Cosmos World Foundation y las herramientas de datos físicos de IA", 18 mar 2025. [En línea]. Disponible: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Fecha de acceso: 25 de marzo de 2025].

[163] NVIDIA, "NVIDIA Isaac Sim" [En Internet]. Disponible: https://developer.nvidia.com/isaac/sim. [Consultado el 25 de marzo de 2025].

[164] M. Quarterly, "Por qué fracasan las estrategias digitales", 25 ene. 2018. [En línea]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail. [Fecha de acceso: 15 de marzo de 2025].

[165] M. J. Perry, "Mis citas favoritas de Milton Friedman", 17 de noviembre de 2006. [En línea]. Disponible: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Fecha de acceso: 1 de marzo de 2025].

[166] J. A. Wheeler, "Información, física, cuántica: la búsqueda de vínculos", 1990.

[169] A. Boiko, "Las guerras de lobbies y el desarrollo de BIM. Parte 5: BlackRock es el amo de todas las tecnologías. Cómo controlan las corporaciones el código fuente abierto", 2024. [En línea]. Disponible: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen y J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced*

Engineering Informatics, 2024.

[171] "Número de empresas del sector de la construcción en el Reino Unido en 2021, por tamaño de empresa". [En línea]. Disponible: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5.000 proyectos IFC&RVT", 2024. [En línea]. Disponible: https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io.

[173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe," 2008. [En línea]. Disponible: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Fecha de consulta: febrero de 2025].

[174] *Lobbying Wars Over Data in Construction | Tecnofeudalismo y la historia del pasado oculto de BIM.* [Película]. Alemania: Artem Boiko, 2023.

[175] A. Boiko, "CHATGPT CON REVIT E IFC | Recuperación automática de documentos y datos de proyectos", 16 nov. 2023. [En línea]. Disponible: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Fecha de acceso: 2 de marzo de 2025].

[176] M. &. Company, "Three new mandates for capturing a digital transformation's full value", 22 de enero de 2022. [En línea]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Fecha de acceso: 15 de febrero de 2025].

[177] KPMG, "La construcción en un mundo digital", 1 de mayo de 2021. [En línea]. Disponible: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Fecha de acceso: 5 de abril de 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17 mar. 2023. [En línea]. Disponible: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Fecha de acceso: 15 de febrero de 2025].

[179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models,'" 10 ene. 2025. [En línea]. Disponible: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Fecha de acceso: 15 de marzo de 2025].

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", 23 mar. 2016. [En línea]. Disponible: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Fecha de acceso: 15 de marzo de 2025].

[181] Oficina de Asuntos Exteriores, Commonwealth y Desarrollo del Reino Unido, "Digital development

strategy 2024 to 2030", 18 mar. 2024. [En línea]. Disponible: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Fecha de acceso: 15 de marzo de 2025.]

[182] "Visión y estrategia en el sector del diseño de edificios", 7 de noviembre de 2003. [En línea]. Disponible: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Fecha del discurso: 5 de marzo de 2025].

[183] M. Bocharov, "Modelización de la información", marzo de 2025. [En línea]. Disponible: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Fecha de la ponencia: 15 de marzo de 2025].

[184] "Integrated Design-Through-Manufacturing: Benefits and Rationale" (Diseño integrado a través de la fabricación: ventajas y justificación), 2000. [en línea]. Disponible: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Fecha de consulta: 25 de marzo de 2025.]

[185] CAD Vendor, "El programa Open BIM es una campaña de marketing", 12 mar. 2012. [En línea]. Disponible: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Fecha de la dirección: 30 de marzo de 2025].

ÍNDICE DE MATERIAS

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Apache Parquet, 62, 67, 380, 381, 494
API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

SDE, 258, 259
Big Data, 9, 67, 245, 383
BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497
BlackBox, 240, 242, 243.
BMS, 8
Bokeh, 320, 337
BOM, 76, 77, 79, 257, 263
Caja delimitadora, 234, 373, 392, 393, 394, 414, 415, 480
BREP, 142, 234, 263, 264, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497
CAE, 16, 283
CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
CAM, 16, 78
CAPEX, 14, 82
CDE, 84, 175, 388, 389, 390
ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421
Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443
CO_2, 72
CO_2, 229, 234, 235, 236, 237, 238, 239, 243
COBie, 156, 289, 292.
CdE, 56, 168, 169, 170, 477
Copiloto, 114, 116, 459
CPIXML, 143, 272, 273, 276, 277, 279, 285, 296
CPM, 14, 17, 62, 166, 175, 233, 326, 473
CQMS, 14, 84, 177, 178, 462
CRM, 109, 369, 459.
CRUD, 51, 459

CSG, 263
CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
DAG, 362, 363, 365, 366
Dash, 320, 336, 337.
Gobernanza de datos, 373, 395, 396, 398, 400, 401
Lago de datos, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Data Lakehouse, 67, 373, 386, 387, 388
Minimalismo de datos, 373, 395, 396, 400, 401
Pantano de datos, 373, 395, 397, 401
Almacén de datos, 382, 383, 400
Datos como servicio, 487
basado en datos, 50, 170, 460, 461, 484, 486
DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
DataOps, 170, 373, 398, 399, 400, 401, 475
DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
DGN, 8, 140, 186, 227, 357
DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
DXF, 8, 73, 277

E

ECM, 58, 175
ECS, 142
RIE, 289
eLOD, 289
ELT, 384, 385
EPM, 14, 166, 198
ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
ESG, 196, 235, 236, 238
ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
Extracto, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Pluma, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
SIG, 58
GLTF, 143, 278
Hojas de cálculo de Google, 368, 370
Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
HiPPO, 29, 37, 95, 424, 477, 484, 490
HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
IGES, 262, 263, 276.
iLOD, 289
IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
ISO 19650, 388

J

JavaScript, 320, 369, 378
JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
k-NN, 392, 393, 442, 445, 446, 447, 448
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329, 343, 349, 381, 421, 429, 443, 494
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494
Carga, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475
LOD, 287, 289
LOI, 287
LOMD, 287
Bajo código, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
MCAD, 77, 78, 257, 284
MEP, 14, 175, 492
MALLA, 234, 283, 284, 285, 296, 462
Microsoft SQL, 65
Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
MRP, 11, 12
MS Project, 70
MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
PNL, 69, 394
Sin código, 368, 369
NURBS, 142, 282, 283, 284, 285
NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
OCR, 69, 128, 132, 134
OmniClass, 154, 155, 156.
BIM abierto, 142, 216, 219, 256, 261, 278, 291
Código abierto, 45, 55, 97, 98, 107, 108, 273, 275, 494
BÚHO, 267, 268, 269.

P

Pandas, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494
Parquet, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480
PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
PDM, 16
PHP, 63, 339
Tuberías, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476
PLM, 16, 246
PLN, 8, 272, 296
Plotly, 320, 336, 337, 372
PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487
PMS, 84, 151, 233
PostgreSQL, 63, 64, 65, 329, 395
Power BI, 320, 372, 482
capital riesgo, 464, 487
Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479

QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
RDBMS, 63, 64, 65, 82, 89.
RDF, 267, 268, 269
RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
ROI, 311, 317, 319, 321, 351, 370, 389, 479
RPM, 14, 84, 151, 331, 333, 334
RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
ALCANCE, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
PASO, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transformar, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclase, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474

DESCUBRA DE QUÉ MANERA LOS DATOS ESTÁN REVOLUCIONANDO LA CONSTRUCCIÓN.

¿Qué contiene?

- Más de 100 temas fundamentales relacionados con datos en AEC
- Más de 300 visualizaciones y gráficos exclusivos
- Más de 50 casos de empresas del mundo real
- LLM práctico y aplicaciones de inteligencia artificial
- Ejemplos de código y flujos de trabajo preconfigurados

Temas

- Construcción fundamentada en datos
- Digitalización en AEC
- Análisis de datos y automatización
- Calidad y gestión de datos
- CAD, BIM e interoperabilidad de datos
- LLM, IA y aprendizaje automático en la edificación
- Estimación de costos y plazos del proyecto

Audiencia

- Directores de proyectos de construcción
- Arquitectos e ingenieros estructurales
- Coordinadores BIM y Administradores de Datos
- Pioneros en la transformación digital
- Especialistas en TI e IA en AEC
- Urbanistas y consultores de sostenibilidad
- Estudiantes de Arquitectura e Ingeniería
- Educadores e investigadores

- **Sitio web**
- **www.datadrivenconstruction.io**

Comentarios sobre la edición inaugural:

«Boiko es el James Carville de las TI. En su célebre "Es la economía, estúpido", solo se requiere una palabra para este famoso libro: "Son los datos, estúpido" (no el software). Para orientarse en el vasto universo de los datos, un proverbio de los antiguos romanos, que se remonta a la época griega, sigue siendo pertinente: "Navigare necesse est". El autor guía a sus lectores a través de los abismos del océano de datos con un pulso firme y una brújula inquebrantable, además de un exhaustivo enfoque histórico y, por último, pero no menos importante, gráficos de gran originalidad...»

— Dr. Burkhard Talebitari

El libro de Artem Boiko representa un hito en la democratización de la digitalización en el sector de la construcción y constituye un verdadero punto de inflexión para las pequeñas y medianas empresas (pymes). Este libro es un llamado a la acción. Se presenta como una guía valiosa para quienes no solo buscan comprender la transformación digital en el sector de la construcción, sino que también desean contribuir de manera activa, pragmática, eficiente y con una visión de futuro. Es el momento propicio para colaborar en la difusión de este conocimiento y aumentar de forma sostenible la productividad del sector de la construcción.

— Dr. Michael Max Buehler

Esta guía práctica capacita a profesionales y principiantes para navegar en el dinámico mundo de la construcción basada en datos. Desde los principios de la gestión de datos hasta flujos de trabajo digitales avanzados, herramientas de inteligencia artificial y aplicaciones prácticas, este libro es su referencia hacia procesos de construcción más inteligentes, rápidos y eficientes.

ISBN: 978-3-98262-551-5

9 783982 625515