2-Е ИЗДАНИЕ, ОБНОВЛЁННОЕ
И РАСШИРЕННОЕ

DATA DRIVEN CONSTRUCTION

НАВИГАЦИЯ В ЭПОХУ ДАННЫХ В СТРОИТЕЛЬНОЙ ОТРАСЛИ

+ ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ ИИ И LLM

Артем Бойко

ДРУГИЕ ЯЗЫКИ КНИГИ НА САЙТЕ DATADRIVENCONSTRUCTION

DATA-DRIVEN CONSTRUCTION

Навигация в эпоху данных в строительной отрасли

Второе издание, исправленное и дополненное

АРТЕМ БОЙКО

«Бойко — это Джеймс Карвилл от ИТ - в многократно цитируемой фразе последнего «It's the economy, stupid», для этой знаменитой книги нужно заменить только одно слово. «Это данные, тупица». (а не программное обеспечение) А чтобы найти свой путь во вселенной данных, поговорка древних римлян, восходящая к греческой, актуальна и сегодня: «Navigare necesse est». Автор ведет своих читателей через все глубины и мели океана данных с уверенной рукой и непоколебимым компасом, не говоря уже о всестороннем историческом подходе и, наконец, весьма оригинальной графике и хорошем чувстве юмора, которое проявляется не только на второй взгляд. Международная реакция на книгу Бойко варьируется от эйфорического одобрения до довольно желчного скептицизма, что пошло на пользу второму немецкому изданию книги. Бойко - оригинальный и недогматичный мыслитель о данных. Он представляет читателю захватывающие открытия и всегда смелые, даже провокационные тезисы, которые вдохновляют на дальнейшие размышления. Отличное лекарство от немецкой болезни латентного консенсуализма. Кстати, у приведенной выше латинской пословицы есть дополнение: «vivere non est necesse». Оно не применимо к подходу Бойко к миру данных - данные живут, и их жизнь необходима, чтобы не сказать критически важна.»

— **д-р наук Буркхард Талебитари,** внештатный редактор - в том числе и для журнала: BIM, ежегодно издаваемого компанией Ernst & Sohn с 2013 года.

"Книга Артема Бойко - важная веха в демократизации цифровизации в строительной отрасли и настоящий поворот в игре для малых и средних предприятий (МСП). Особенно принципиально важно то, что, используя современные инструменты с открытым исходным кодом и без кода, компании уже сегодня могут эффективно интегрировать данные в свои бизнес-процессы и анализировать их - без глубоких знаний программирования. Это делает излишним дорогостоящее использование громоздких коммерческих программных пакетов. Эта книга - призыв к действию! Это ценное руководство для тех, кто не только хочет понять суть цифровой трансформации в строительной отрасли, но и активно формировать ее - прагматично, эффективно и перспективно. Настало время работать вместе, чтобы поделиться этими знаниями и стабильно повышать производительность строительной отрасли".

-д-р наук Михаэль Макс Бюлер, профессор строительного менеджмента в HTWG Konstanz, совладелец GemeinWerk Ventures и независимый директор DevvStream.

"Книга DataDrivenConstruction — это один из первых шагов за пределы привычного мира строителей, с их сложными системами проектирования и управления, когда, казалось бы, сложность и насыщенность данных не дает даже шанса на радикальное упрощение и повышение прозрачности работы со строительными данными. В своей книге Артем простым языком показывает, какие возможности открывают перед нами современные технологии работы с данными, и буквально дает конкретные шаги, которые вы можете сразу же применить в своей работе. Я призываю всех, кто хочет понять, куда пойдут системы автоматизации в строительной отрасли, внимательно изучить эту книгу, чтобы осознать, что революция данных в строительстве уже стучится в нашу дверь. Сейчас это интересно только гикам, но через несколько лет, как и BIM , такие подходы и программное обеспечение станут повсеместными!"

- Игорь Рогачев, руководитель Центра компетенции IMT, BIM и цифровой трансформации в RGD, основатель InfraBIM.Pro.

"Для всех, кто работает в строительной отрасли, от новичков до опытных профессионалов, эта книга - просто находка! Это не обычное пыльное чтиво - она наполнена идеями, стратегиями и нотками юмора, которые не дадут вам скучать. От древних методов регистрации данных до передовых цифровых технологий - в книге рассказывается об эволюции использования данных в строительстве. Это как на машине времени проехать через эволюцию строительных данных. Если вы архитектор, инженер, менеджер проекта или аналитик данных, это всеобъемлющее руководство изменит ваш подход к проектам. Будьте готовы оптимизировать процессы, повысить эффективность принятия решений и управлять проектами как никогда раньше!"

- Пьерпаоло Вергати, преподаватель Римского университета Сапиенца и старший менеджер строительных проектов компании Fintecna.

"Все, что я могу сказать, — это WOW! То, как вы объединили историю, LLM , графику и общую легкость понимания ваших положений, действительно замечательно. Поток книги просто потрясающий. В этой книге так много блестящих аспектов; она действительно меняет ситуацию. Это прекрасный источник информации, и я благодарю вас за усилия и страсть, которые вы в нее вложили. Поздравляю вас с созданием такой замечательной работы. Я могла бы продолжать, но достаточно сказать, что я невероятно впечатлена!"

- Наташа Принслоо, руководитель цифровой практики в energylab_.

"Если "данные — это новая нефть", то нам нужно научиться определять их, находить, добывать, перерабатывать, чтобы сделать их ценными". Я нашел книгу DataDrivenConstruction очень информативной и проницательной. Книга содержит полезную историческую справку и объясняет работу с данными доступным языком. Для тех, кто интересуется цифровой трансформацией, она дает понимание данных - как они работают, как они структурированы и как их можно использовать."

- Ральф Монтегю, директор ArcDox, директор Саммита координаторов BIM и председатель Национального комитета BIM в Национальном управлении по стандартам Ирландии.

"Как подчеркивается в книге, информация - важнейший актив строительного сектора, и наличие ее в доступных форматах значительно облегчает принятие точных решений и ускоряет сроки реализации проектов. Книга предлагает нейтральный и эффективный подход к получению доступа к этому источнику и использованию его преимуществ при принятии решений. Методология, представленная в книге, использует современный подход, сочетающий программирование на основе искусственного интеллекта и доступные инструменты с открытым исходным кодом. Используя возможности искусственного интеллекта и применяя программное обеспечение с открытым исходным кодом, методология направлена на повышение уровня автоматизации, оптимизацию процессов, а также на обеспечение доступности и сотрудничества в данной области. Язык книги понятен и прост для восприятия."

- д-р наук Салих Офлуоглу, декан факультета изобразительных искусств и архитектуры Университета Анталии Билим, организатор Евразийского форума BIM .

"Data Driven Construction" ярко передает основы информационной работы со строительными данными. Книга, которая рассматривает информационные потоки и фундаментальные экономические концепции и тем самым выделяется на фоне других книг по BIM, поскольку не только представляет точку зрения производителя программного обеспечения, но и пытается донести фундаментальные концепции. Книга, которую стоит прочитать и увидеть."

- Якоб Хирн, генеральный директор и соучредитель компании Build Informed GmbH, инициатор инновационного форума "На вершине с BIM ".

"Прочитал книгу на одном дыхании, менее чем за 6 часов. Качество изготовления книги отличное, плотная глянцевая бумага, цветовые решения, приятный шрифт. Большое количество практических примеров по работе с LLM , характерных для строительной отрасли, сэкономит вам месяцы, если не годы, самостоятельного изучения. Примеры работы очень разнообразны, от простых до сложных, не требуя от вас приобретения сложного и дорогостоящего программного обеспечения. Книга позволит владельцам любого бизнеса в строительной отрасли по-новому взглянуть на свою бизнес-стратегию, цифровизацию и перспективы развития. А небольшим компаниям - повысить эффективность с помощью доступных и бесплатных инструментов."

- Михаил Косарев, преподаватель и консультант по цифровой трансформации в строительной отрасли в TIM-ASG.

"Я очень рекомендую книгу DataDrivenConstruction, в которой, как сказано в названии, рассматривается подход к управлению информацией на основе данных для AECO. В настоящее время я использую ее, чтобы помочь инициировать ряд обсуждений с различными группами. Я нашел ее очень доступным справочником. Помимо подробного обзора истории инструментов в AECO, данных и представления нескольких ключевых технологий, книга содержит ряд очень полезных диаграмм, которые описывают объем источников данных и артефактов конечного пользователя с примерами рабочих процессов. Мне кажется, что именно такие диаграммы нам нужны при разработке и мониторинге информационных стратегий и вклада в BEP - определение общей модели данных предприятия, на которую можно наложить границы для PIM и AIM ".

- Пол Рэнсли, главный консультант компании Acmena и инженер по системной интеграции компании Transport for London.

"Книга "DATA DRIVEN CONSTRUCTION" — это gamechanger для всех, кому интересно, куда движется строительная отрасль в эпоху данных. Артем не просто делает поверхностный обзор, он глубоко вникает в текущие события, проблемы и перспективные возможности в строительстве. Отличительной чертой этой книги является ее доступность - Артем объясняет сложные идеи, используя понятные аналогии, которые делают содержание легким для восприятия. Я нашел книгу невероятно информативной и в то же время увлекательной. В общем, Артем создал ценный ресурс, который не только информирует, но и вдохновляет. Независимо от того, являетесь ли вы опытным профессионалом или новичком в строительстве, эта книга расширит ваш кругозор и углубит понимание того, куда движется отрасль. Настоятельно рекомендуем!"

- Моааяд Салех, архитектор, менеджер по внедрению BIM в TMM GROUP Gesamtplanungs GmbH.

"Строительство, управляемое данными" Артема Бойко - впечатляющий труд, предлагающий прочную основу для строительной отрасли во времена постоянно растущих технологий и информационных возможностей. Бойко удается излагать сложные темы в понятной форме и одновременно представлять перспективные идеи. Книга представляет собой хорошо продуманный компендиум, который не только освещает текущие события, но и дает представление о будущих инновациях. Настоятельно рекомендуется всем, кто хочет познакомиться с планированием и выполнением строительных работ на основе данных."

- Маркус Айбергер, преподаватель Штутгартского университета прикладных наук, старший менеджер проектов и заместитель руководителя филиала компании Konstruktionsgruppe Bauen, член правления ассоциации BIM Cluster Baden-Württemberg.

Как говорится, "данные - это новая нефть", поэтому их искатели или добытчики должны обладать правильными инструментами и мышлением, чтобы извлекать ценность из этого ресурса XXI века. Строительная отрасль слишком долго шла по скользкому пути процессов, основанных на "3D-информации", когда реализация проекта основывается на чьей-то готовой информации (например, они уже построили круговую или гистограмму), тогда как лежащие в основе "данные" (например, необработанная электронная таблица) способны дать гораздо больше, тем более что слияние нескольких данных и искусственный интеллект открывают неограниченные возможности. Если вы занимаетесь строительством (или преподаете/исследуете), эта книга - ваш лучший - и пока единственный - ресурс для навигации по миру, управляемому данными, в котором мы оказались."

- д-р наук Зульфикар Адаму, доцент кафедры стратегических ИТ в строительстве в LSBU, Великобритания.

"Я должен сказать, что книга Data-Driven Construction достойна преподавания в качестве учебника в университетах и внесет ценный вклад в развивающуюся область BIM . Книга Data-Driven Construction содержит технический глоссарий, который очень хорошо объясняет концепции. Темы, которые крайне сложно объяснить, упрощены и понятны с помощью очень красивого визуального языка. Я считаю, что то, что предполагается объяснить с помощью визуальных средств, должно быть доступно читателю, пусть даже вкратце. Непонятность некоторых визуалов, другими словами, чтение визуала требует отдельной информации. Хочу также сказать, что я с удовольствием представляю ценные работы Артема Бойко на своих лекциях и семинарах в университетах."

- д-р наук Эдиз Язычиоглу, владелец ArchCube, преподаватель управления строительными проектами на кафедре архитектуры Стамбульского технического университета и университета Medipol.

Второе издание, март 2025 г.

ISBN 978-3-9827303-6-3

Артем Бойко Авторские права

boikoartem@gmail.com
info@datadrivenconstruction.io

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Данная книга является результатом живого диалога с профессиональным сообществом. В её основу легли многочисленные профессиональные дискуссии по вопросам работы с данными в строительной отрасли, которые проводились на различных профессиональных площадках и социальных платформах. Эти обсуждения стали основой для статей, публикаций и визуальных материалов, вызвавших широкий отклик в профессиональном сообществе. Материалы автора ежегодно привлекают миллионы просмотров на различных платформах и языках, объединяя специалистов в области цифровизации строительства.

В течение года после выхода первого издания книгу заказали специалисты более чем из 50 стран — от Бразилии и Перу до Маврикия и Японии. Второе издание книги, которое вы сейчас держите в руках, переработано и дополнено на основе отзывов экспертов, критических замечаний к первому изданию и дискуссий в профессиональных кругах. Благодаря отзывам второе издание существенно расширено: добавлены новые главы о CAD (BIM) технологиях и создании эффективных ETL-процессов. Также значительно увеличено количество практических примеров и кейсов. Особую ценность представляет обратная связь от лидеров строительной отрасли, консалтинговых компаний и крупнейших IT-компаний, которые обращались к автору с вопросами цифровизации и интероперабельности как до выхода первой версии книги, так и после. Многие из них уже применяют описанные в книге подходы или планируют это сделать в ближайшее время.

Вы держите в руках книгу, созданную благодаря дискуссиям и активному обмену мнениями. Прогресс рождается в диалоге, в столкновении взглядов и открытости к новым подходам. Спасибо, что становитесь частью этого диалога. Ваша конструктивная критика — основа будущих улучшений. Если в тексте будут обнаружены ошибки или возникнет желание поделиться идеями и предложениями, приветствуется любая обратная связь. Контактные данные для связи указаны в конце книги.

ПОЧЕМУ КНИГА БЕСПЛАТНАЯ?

Эта книга была задумана как открытый образовательный ресурс, направленный на распространение современных подходов к управлению данными в строительной отрасли. Первая версия книги служила основой для сбора комментариев и предложений от профессионального сообщества, что позволило улучшить структуру и содержание материала. Все замечания, предложения и идеи были тщательно проанализированы и учтены в этой, доработанной версии. Цель книги — помочь специалистам строительной отрасли понять, насколько важно работать с данными: системно, осознанно и с оглядкой на долгосрочную ценность информации. Автор собрал примеры, иллюстрации и практические наблюдения за более чем 10 лет работы в сфере цифровизации строительства. Большинство из этих материалов рождались в ходе реальных проектов, дискуссий с инженерами и разработчиками, участия в международных инициативах и проведении обучающих семинаров. Книга — это попытка структурировать накопленный опыт и поделиться им в доступной форме. Если вы хотите поддержать дальнейшее распространение идей книги и получить удобный формат для чтения, работы с примерами и визуальными материалами — вы можете приобрести печатную версию.

ПРАВА НА ИСПОЛЬЗОВАНИЕ МАТЕРИАЛОВ

С искренней благодарностью посвящаю эту книгу своей семье, которая с ранних лет привила мне глубокую любовь к строительству, моему родному шахтёрскому городу — за уроки жизнестойкости и моей жене-маркшейдеру, чья неизменная поддержка была моим постоянным вдохновением.

ДЛЯ КОГО ЭТА КНИГА

Написанная доступным языком, эта книга рассчитана на широкий круг читателей в строительной отрасли - от студентов и новичков, желающих постичь основы современных строительных процессов, до профессионалов, которым нужна современная методология управления данными в строительстве. Будь вы архитектором, инженером, прорабом, менеджером по строительству или аналитиком данных, это всеобъемлющее руководство с множеством уникальных иллюстраций и графиков предлагает ценные сведения о том, как использовать данные в бизнесе для оптимизации и автоматизации процессов, улучшения принятия решений и управления строительными проектами на разных уровнях с помощью современных инструментов.

Книга представляет собой комплексное руководство, сочетающее теоретические основы и практические рекомендации по интеграции методов управления данными в строительные процессы. Книга фокусируется на стратегическом использовании информации для оптимизации операционной деятельности, автоматизации процессов, совершенствования принятия решений и эффективного управления проектами с применением современных цифровых инструментов.

На страницах этой книги рассмотрены теоретические и практические аспекты работы с информацией в строительной отрасли. Посредством детальных примеров исследуется методология параметризации задач, сбора требований, обработки неструктурированных и разноформатных данных и их преобразования в эффективные решения для строительных компаний.

Читатель последовательно проходит путь от формирования требований и разработки базовых моделей данных к более сложным процессам интеграции разнородных источников информации, созданию ETL-процессов, построению информационных Pipeline и моделей машинного обучения. Последовательный подход позволяет наглядно продемонстрировать механизмы организации и автоматизации бизнес-процессов и систем поддержки принятия решений в строительной сфере. Каждая часть книги завершается практической главой, содержащей пошаговые инструкции, которые дают возможность незамедлительно применить полученные знания в реальных проектах.

КРАТКИЙ ОБЗОР ЧАСТЕЙ КНИГИ

Эта книга выстроена вокруг концепции трансформации данных в цепочке создания ценности (value chain): от их сбора и обеспечения качества до аналитической обработки и извлечения ценных практических решений с применением современных инструментов и методологий.

Часть 1: Цифровая эволюция в строительстве - прослеживает историческую трансформацию управления данными от глиняных табличек до современных цифровых систем, анализирует появление модульных систем и рост значимости дигитализации информации в контексте промышленных революций.

Часть 2: Информационные вызовы строительной индустрии - исследует проблемы фрагментации данных, "информационные силосы", влияние HiPPO-подхода на принятие решений и ограничения проприетарных форматов, предлагая рассмотреть переход к AI и LLM-экосистемам.

Часть 3: Систематизация данных в строительстве - формирует типологию строительных данных, описывает методы их организации, интеграции с корпоративными системами и обсуждается создание центров компетенций для стандартизации информационных процессов.

Часть 4: Обеспечение качества данных - раскрывает методологии превращения разрозненной информации в качественные, структурированные данные, включая извлечение данных из различных источников, валидацию и моделирование с применением LLM.

Часть 5: Калькуляции стоимости и времени - посвящена цифровизации расчетов стоимости и планирования, автоматизации получения объемов из CAD- (BIM-) моделей, технологиям 4D-8D моделирования и расчету ESG-показателей строительных проектов.

Часть 6: CAD и BIM - критически анализирует эволюцию технологий проектирования, проблемы совместимости систем, тенденции перехода к открытым форматам данных и перспективы применения искусственного интеллекта в проектировании.

Часть 7: Аналитика данных и автоматизация - рассматривает принципы визуализации информации, ключевые показатели эффективности, процессы ETL, инструменты оркестрации рабочих процессов и применение языковых моделей для автоматизации рутинных задач.

Часть 8: Хранение и управление данными - исследует форматы хранения данных, концепции хранилищ и озёр данных, принципы управления данными и новые подходы, включая векторные базы данных и методологии DataOps и VectorOps.

Часть 9: Большие данные и машинное обучение - посвящена переходу к объективному анализу на основе исторических данных, интернету вещей на стройплощадках и применению алгоритмов машинного обучения для прогнозирования стоимости и сроков проектов.

Часть 10: Строительная отрасль в эпоху цифровых данных - представляет взгляд на будущее строительной отрасли, анализирует переход от причинно-следственного анализа к работе с корреляциями, концепцию "уберизации" строительства и стратегии цифровой трансформации.

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

ВВЕДЕНИЕ

Как долго ваша компания сможет сохранять конкурентоспособность в мире, где технологии стремительно развиваются и каждый аспект бизнеса, от расчёта сроков и стоимости до анализа рисков, автоматизируется с помощью моделей машинного обучения?

Строительная отрасль, существующая столько же, сколько само человечество, стоит на пороге революционных изменений, которые обещают полностью изменить наши представления о традиционном строительстве. Уже сейчас в других секторах экономики цифровизация не просто меняет привычные правила, но и безжалостно вытесняет с рынка компании, не сумевшие адаптироваться к новым условиям обработки данных и не способные повысить скорость принятия решений (Рис. 1).

Рис. 1 Скорость принятия решений в строительной отрасли чаще чем в других отраслях зависит от человеческого фактора.

Банковский сектор, розничная торговля, логистика и агропромышленный комплекс стремительно движутся к полной цифровизации, где больше нет места неточностям и субъективным мнениям. Современные алгоритмы способны анализировать колоссальные массивы данных и предоставлять клиентам точные прогнозы – будь то вероятность возврата кредита, оптимальные маршруты доставки или прогнозирование рисков.

Строительство является одной из последних отраслей, которой предстоит совершить неизбежный переход от решений, основанных на мнениях высокооплачиваемых специалистов, к решениям, базирующимся на данных. Этот переход обусловлен не только новыми технологическими возможностями, но и возросшими требованиями рынка и заказчиков к прозрачности, точности и скорости.

Роботизация, автоматизация процессов, открытые данные и прогнозы на их основе — всё это уже не просто возможности, а неизбежность. Большинство компаний строительной отрасли, которые

ещё недавно отвечали перед клиентом за расчёты объемов, стоимости, времени проектов и контроль качества, теперь рискуют превратиться в простых исполнителей приказов, не принимающих ключевых решений (Рис. 2).

С развитием вычислительных мощностей, алгоритмов машинного обучения и демократизацией доступа к данным стало возможным автоматическое объединение данных из разных источников, что позволяет глубже анализировать процессы, прогнозировать риски и оптимизировать затраты ещё на этапах обсуждения строительного проекта. Эти технологии создают потенциал для радикального повышения эффективности и снижения затрат во всём секторе.

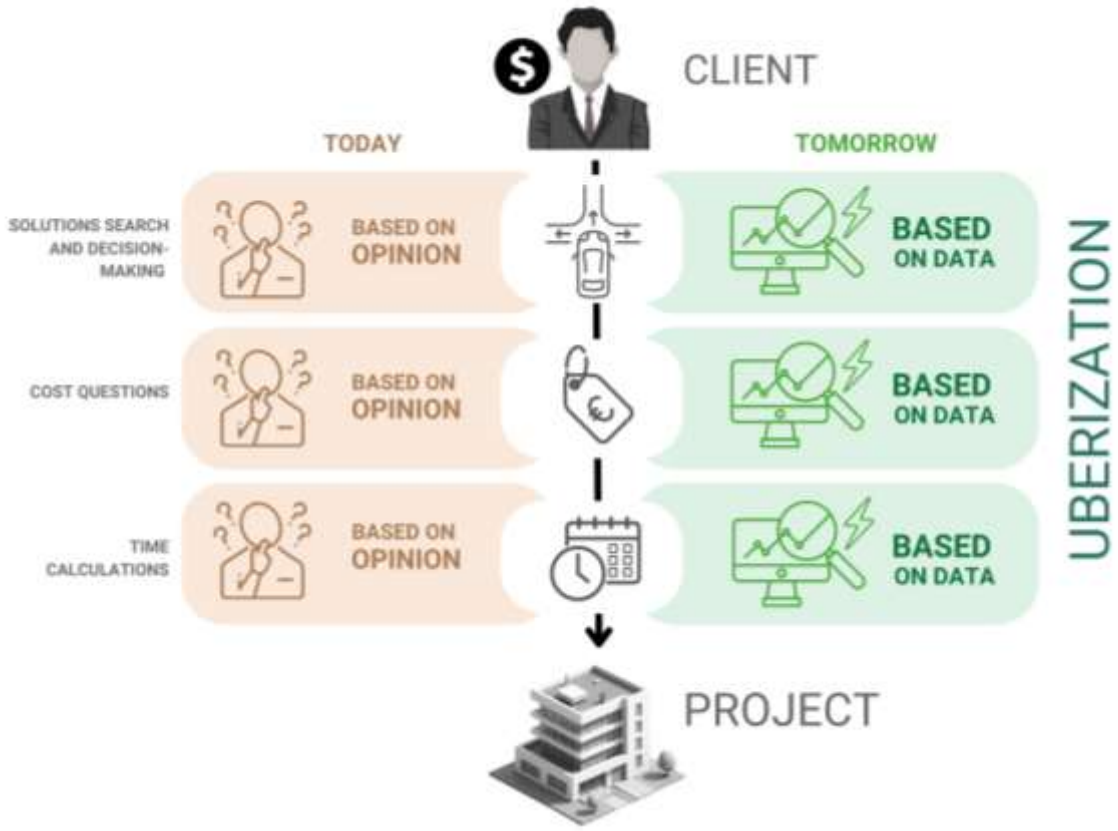


Рис. 2 Клиент не заинтересован в избыточном человеческом факторе на пути к реализации своего проекта.

Несмотря на все преимущества новых инструментов и концептов, строительная отрасль значительно отстаёт от других секторов экономики по внедрению новых технологий.

> Согласно отчету „IT Metrics Key Data 2017“, строительная отрасль занимает последнее место по расходам на IT среди 19 других отраслей экономики [1].

Быстрый рост объема данных и сложности процессов становится головной болью для руководства компаний, и главная проблема в использовании новых технологий заключается в том, что данные, несмотря на их обилие, остаются разрозненными, неструктурированными и зачастую несовместимыми между различными системами и программными продуктами. Поэтому многие компании строительного сектора сегодня в первую очередь обеспокоены проблемами качества данных, которые можно решить только с внедрением эффективных, автоматизированных систем управления и аналитики.

Согласно опросу, проведённому KPMG® среди менеджеров строительных компаний в 2023 году [2], наибольший потенциал для повышения рентабельности инвестиций в проекты имеют информационные системы управления проектами (PMIS), передовая и базовая аналитика данных и информационное моделирование зданий (BIM) (Рис. 3).

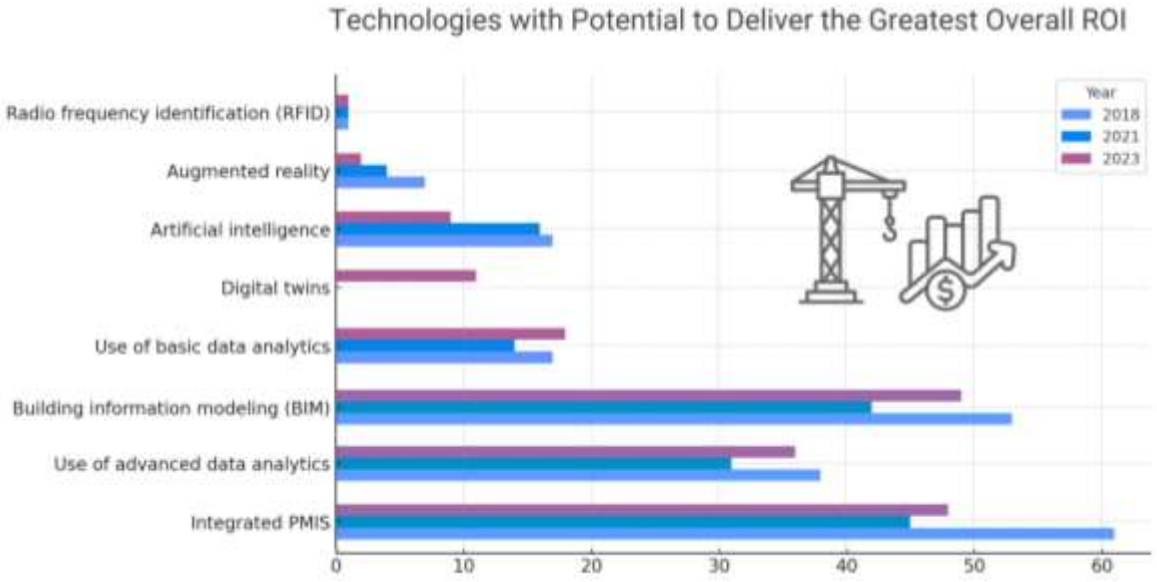


Рис. 3 Опрос среди руководителей строительных компаний: какие технологии обеспечат наибольшую окупаемость инвестиций (ROI) в капитальных проектах? (по материалам [2]).

Решение проблем, связанных с интеграцией данных в бизнес-процессы, заключается в обеспечении высокого качества информации, использовании подходящих форматов данных и применении эффективных методов создания, хранения, анализа и обработки данных.

Осознание ценности данных заставляет различные отрасли отказаться от разрозненных приложений и сложных бюрократических структур управления. Вместо этого акцент смещается на создание новых подходов к информационной архитектуре, превращающих компании в современные предприятия, управляемые данными. Рано или поздно этот шаг совершит и сама строительная отрасль, перейдя от постепенной цифровой эволюции к настоящей цифровой революции, затрагивающей все компании.

Переход к бизнес-процессам, основанным на данных, будет непростым. Многие компании столкнутся с трудностями, поскольку руководители не всегда понимают, как использовать хаотичные массивы данных для повышения эффективности и роста бизнеса.

В этой книге мы погружаемся в мир данных, где информация становится ключевым стратегическим ресурсом, определяющим эффективность и устойчивость бизнес-процессов. В условиях стремительного роста объемов информации компании сталкиваются с новыми вызовами. Цифровая трансформация перестает быть просто модным термином — она становится необходимостью.

Рис. 4 Данные и процессы являются основой строительства.

Понять трансформацию — значит суметь объяснить сложное простыми словами. Именно поэтому книга написана доступным языком и сопровождается авторскими иллюстрациями, созданными специально для наглядного объяснения ключевых понятий. Эти схемы, графики и визуализации призваны устранить барьеры восприятия и сделать материал понятным даже тем, кто раньше считал подобные темы слишком сложными. Все иллюстрации, схемы и графики в этой книге созданы автором и разработаны специально для визуализации ключевых концепций, описанных в тексте.

> *Одна картинка стоит тысячи слов [3].*
>
> – Фред Р. Барнард, английский иллюстратор, 1927

Чтобы связать теорию с практикой, мы будем использовать инструменты искусственного интеллекта (в частности, языковые модели), которые позволяют разрабатывать решения без необходимости глубоких знаний в программировании. Если вы ориентированы на практический материал и вас больше интересует практическая работа с данными, вы можете пропустить первую вводную часть и сразу перейти ко второй части книги, где начинается описание конкретных примеров и кейсов.

Однако не стоит возлагать чрезмерные ожидания на ИИ (Искусственный Интеллект), машинное обучение и LLM (Large Language Models) инструменты в целом. Без качественных исходных данных и глубокого понимания предметной области даже самые передовые алгоритмы не способны обеспечить надежные и значимые результаты.

> Генеральный директор Microsoft Сатья Наделла в начале 2025 года предостерегает о риске возникновения пузыря в сфере искусственного интеллекта [4], сравнивая текущий ажиотаж с "пузырем доткомов". Он подчеркивает, что заявления о достижении этапов AGI (Artificial General Intelligence) без соответствующих оснований являются "бессмысленным манипулированием показателями". Наделла считает, что реальный успех ИИ должен измеряться его вкладом в рост мирового ВВП, а не чрезмерным вниманием к громким заявлениям.

За всеми громкими словами о новых технологиях и концепциях скрывается сложная и кропотливая работа по обеспечению качества данных, параметризации бизнес-процессов и адаптации инструментов под реальные задачи.

> Подход, основанный на данных, — это не продукт, который можно просто скачать или купить. Это стратегия, которую нужно выстроить. Она начинается с нового взгляда на существующие процессы и проблемы, а затем требует дисциплинированного движения в выбранном направлении.

Ведущие разработчики программного обеспечения и вендоры приложений не станут локомотивом изменений в строительной индустрии для многих из них data-driven подход представляет угрозу сложившейся бизнес-модели.

> *Другие отрасли [в отличии от строительной], такие как автомобильная, уже претерпели радикальные и разрушительные изменения, и их цифровая трансформация уже идет полным ходом. Строительным компаниям необходимо действовать быстро и решительно: ловкие компании получат огромные выгоды, тогда как для колеблющихся риски будут серьезными. Вспомните потрясение, которое цифровая фотография вызвала в этой индустрии [5].*
>
> — Отчёт Всемирного экономического форума "Формирование будущего строительства”, 2016

Те компании, которые своевременно осознают возможности и преимущества нового подхода, смогут получить устойчивое конкурентное преимущество и смогут развиваться и расти без зависимости от решений крупных вендоров.

Это ваш шанс не только переждать грядущую бурю дигитализации информации, но и взять ее под контроль. В книге вы найдете не просто анализ текущего состояния отрасли, но и конкретные рекомендации по переосмыслению и реструктуризации ваших процессов и вашего бизнеса, чтобы стать лидером в новой эре строительства и повысить ваш профессиональный опыт.

Цифровое будущее строительства – это не просто использование новых технологий и программ, а фундаментальное переосмысление работы с данными и бизнес-моделей.

Готова ли ваша компания к этим стратегическим изменениям?

ОГЛАВЛЕНИЕ

МАКСИМУМ УДОБСТВА С ПЕЧАТНОЙ ВЕРСИЕЙ

Вы держите в руках бесплатную цифровую версию **Data-Driven Construction**. Для более удобной работы и быстрого доступа к материалам рекомендуем обратить внимание на **печатное издание**:

- **Всегда под рукой**: книга в печатном формате станет надежным рабочим инструментом, позволяя быстро найти и использовать нужные визуализации и схемы в любых рабочих ситуациях
- **Высокое качество иллюстраций**: все изображения и графики в печатном издании представлены в максимальном качестве
- **Быстрый доступ к информации**: удобная навигация, возможность делать пометки, закладки и работать с книгой в любом месте.

Приобретая полную печатную версию книги, вы получаете удобный инструмент для комфортной и эффективной работы с информацией: возможность оперативно использовать визуальные материалы в повседневных задачах, быстро находить нужные схемы и делать пометки. Кроме того, ваша покупка поддерживает распространение открытых знаний.

Заказать печатную версию книги можно на: datadrivenconstruction.io/books

I ЧАСТЬ

ОТ ГЛИНЯНЫХ ТАБЛИЧЕК ДО ЦИФРОВОЙ РЕВОЛЮЦИИ: КАК ЭВОЛЮЦИОНИРОВАЛА ИНФОРМАЦИЯ В СТРОИТЕЛЬСТВЕ

В первой части книги рассматривается историческая эволюция управления данными в строительной отрасли — от примитивных записей на физических носителях до современных цифровых экосистем. Анализируется трансформация технологий управления информацией, появление ERP-систем и влияние фрагментации данных на эффективность бизнес-процессов. Особое внимание уделяется процессу дигитализации информации и растущему значению объективного анализа вместо субъективных экспертных оценок. Детально рассматривается экспоненциальный рост объемов информации, с которым сталкивается современная строительная отрасль, и связанные с этим вызовы для корпоративных систем. Исследуется позиционирование строительной индустрии в контексте четвертой и пятой промышленных революций, а также потенциал использования искусственного интеллекта и датацентричных подходов для создания устойчивых конкурентных преимуществ.

ГЛАВА 1.1.
ЭВОЛЮЦИЯ ИСПОЛЬЗОВАНИЯ ДАННЫХ В СТРОИТЕЛЬНОЙ ОТРАСЛИ

Рождение эры данных в строительстве

Около 10 000 лет назад, в эпоху неолита, человечество совершило революционный переход в своем развитии, отказавшись от кочевого образа жизни в пользу оседлости, что привело к появлению первых примитивных построек из глины, дерева и камня [6]. С этого момента начинается история строительной индустрии.

По мере развития цивилизаций архитектура становилась все более сложной, что привело к появлению первых ритуальных храмов и общественных зданий. Усложнение архитектурных проектов потребовало от инженеров и управленцев древности создания первых записей и расчетов. Первые записи на глиняных табличках и папирусах часто включали описание логики расчета количества необходимых строительных материалов, их стоимости и расчета оплаты выполненной работы [7]. Так началась эра использования данных в строительстве — задолго до появления современных цифровых технологий (Рис. 1.1-1).

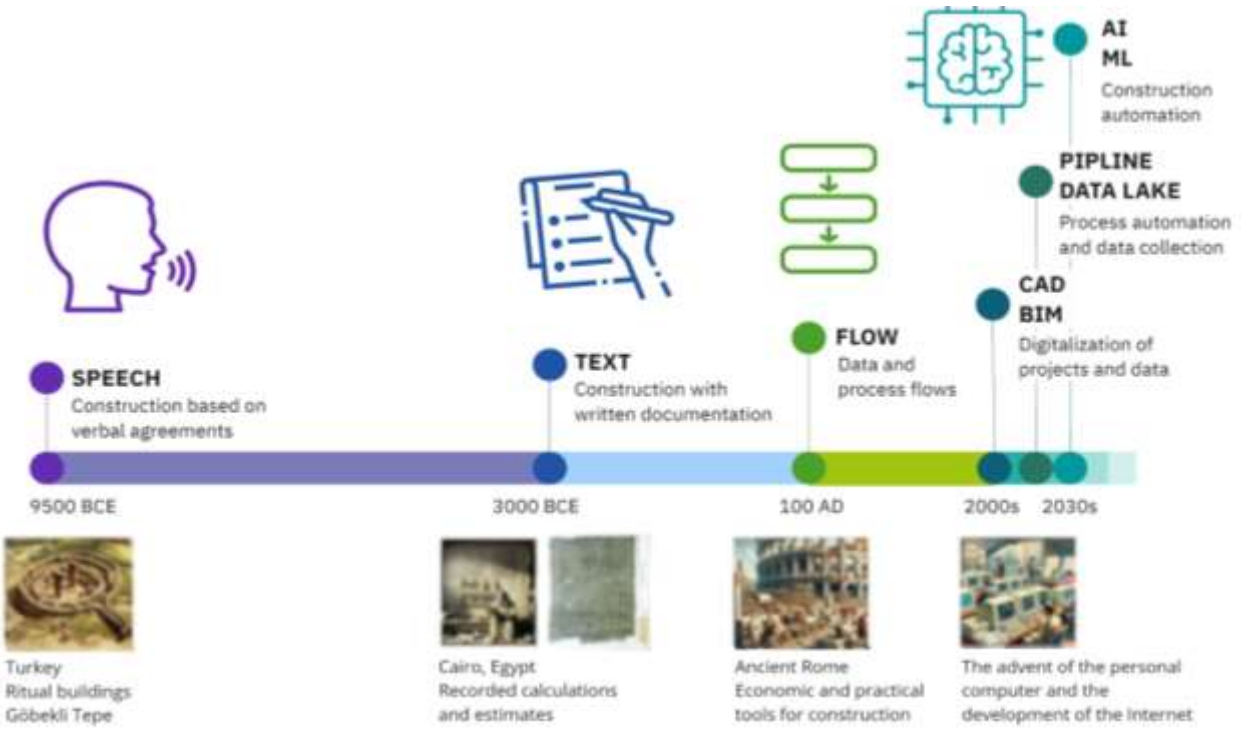


Рис. 1.1-1 Хронология развития информационных технологий в строительстве: от вербальной информации до искусственного интеллекта.

От глины и папируса до цифровых технологий

Первые документальные свидетельства в строительстве относятся к периоду возведения пирамид, около 3000–4000 гг. до н. э. [7]. С тех пор ведение письменных записей облегчало и сопровождало прогресс в строительной отрасли, позволяя накапливать и систематизировать знания, которые в течение последующих 10 000 лет привели к значительным инновациям в методах строительства и архитектуре.

Использование первых физических носителей в строительстве, таких как глиняные таблички, папирус тысячелетней давности (Рис. 1.1-2) или бумага формата “A0” в 1980-х, для записи данных изначально не предполагало применения этой информации в новых проектах. Основной целью таких записей было подробное описание текущего состояния проекта, включая расчеты необходимых материалов и стоимости работ. Точно так же в современном мире наличие цифровых проектных данных и моделей не всегда гарантирует их применение в будущих проектах и часто служит в основном в качестве информации для текущих расчетов необходимых материалов и стоимости строительства.

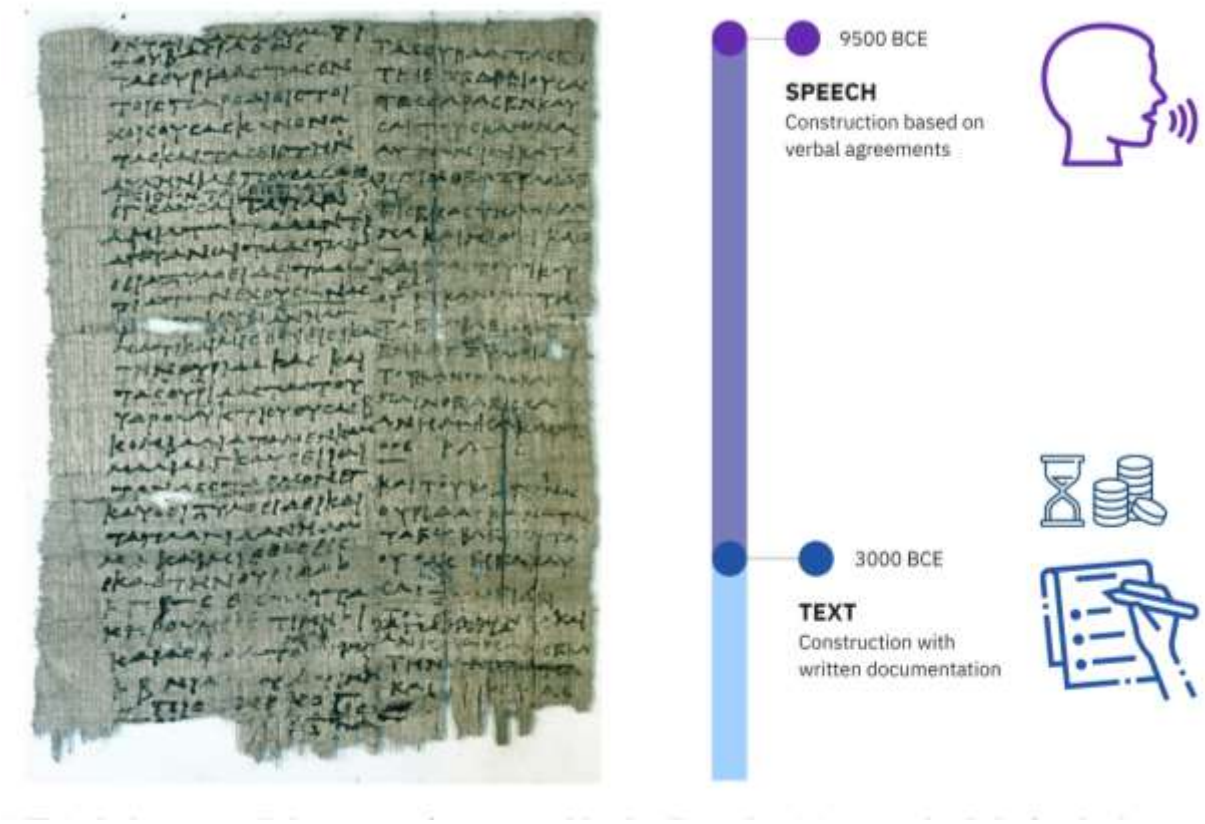


Рис. 1.1-2 Папирус III века до н. э., описывающий стоимость росписи различных типов окон в королевском дворце с использованием техники энкаустики.

> Человечеству потребовалось около 5 000 лет, чтобы перейти от устных разговоров к письменным документам в управлении строительными проектами, и столько же времени, чтобы перейти от бумажных носителей к цифровым данным как основному ресурсу планирования и контроля.

Подобно тому, как развитие торговых и денежных отношений стимулировало появление письменности и первых юристов, которые решали спорные вопросы, так и первые записи о стоимости материалов и объемах работ в строительстве привели к появлению первых менеджеров в строительной отрасли, в обязанности которых входило документирование, мониторинг и ответственность за ключевую информацию о сроках и стоимости проекта.

Сегодня данные играют гораздо более значимую роль: они не только фиксируют принятые решения, но и становятся инструментом предсказания и моделирования будущего. На этом фундаменте строится современный процессный подход в управлении проектами — превращение накопленного опыта в систему принятия решений, основанную на структурированных и проверяемых данных.

Процесс как инструмент опыта, управляемого данными

В основе любого процесса лежит трансформация прошлого опыта в инструмент планирования будущего. Опыт в современном понимании представляет собой структурированный набор данных, анализ которых позволяет делать обоснованные прогнозы.

> Именно исторические данные служат фундаментом прогнозирования, поскольку они наглядно демонстрируют результаты выполненных работ и дают представление о факторах, влияющих на эти результаты.

Рассмотрим конкретный пример из монолитного строительства: обычно при планировании сроков работ учитываются объем бетона, сложность конструкции и погодные условия. Предположим, что у конкретного прораба на строительной площадке или исторические данные компании, за последние три года (2023–2025) показывают, что на заливку монолитной конструкции площадью 200 м² в дождливую погоду требовалось от 4,5 до 6 дней (Рис. 1.1-3). Именно такая накопленная статистика становится основой для прогнозирования сроков выполнения и калькуляции ресурсов при планировании аналогичных работ в будущих проектах. На основе этих исторических данных прораб или сметчик, могут сделать обоснованный прогноз, основанный на полученном опыте, относительно времени, необходимого для выполнения будущих подобных работ в 2026 году при схожих условиях.

В данном случае оценки времени – аналитический процесс выступает как механизм преобразования разрозненных данных в структурированный опыт, а затем – в точный инструмент планирования. Данные и процессы – это единая экосистема, где одно не может существовать без другого.

> *Считайте то, что поддается подсчету, измеряйте то, что поддается измерению, а то, что не поддается измерению, сделайте измеримым [8].*
>
> – Галилео Галилей

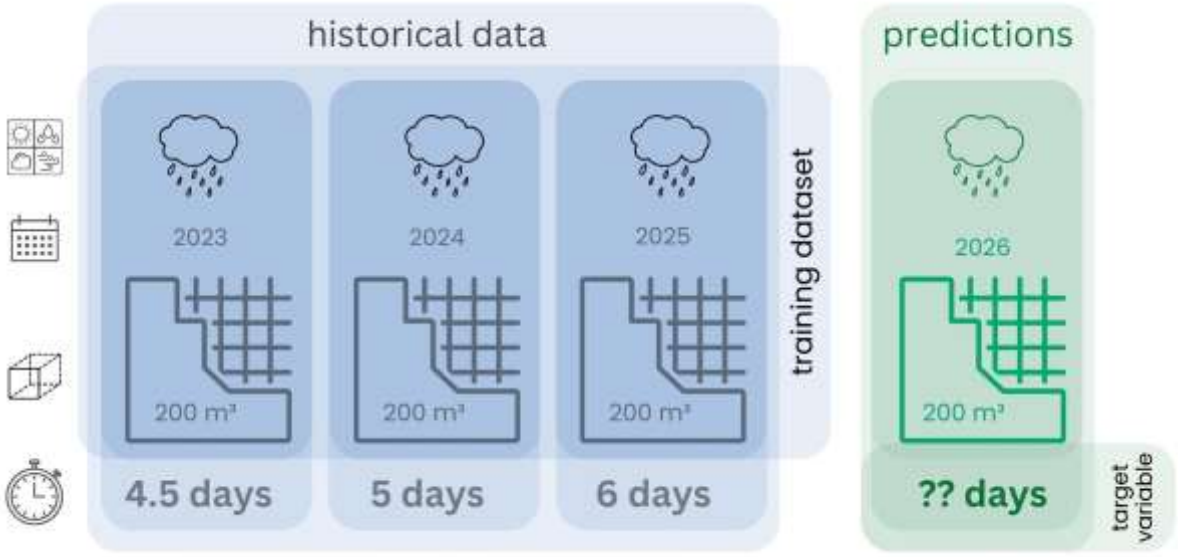


Рис. 1.1-3 Исторические данные выступают как тренировочный набор данных для предсказания одной из величин в будущем

В современном бизнес-ландшафте аналитика данных становится критически важным компонентом эффективного управления проектами, оптимизации процессов и стратегического принятия решений. Строительная индустрия постепенно осваивает четыре ключевых уровня аналитики, каждый из которых отвечает на специфический вопрос и предоставляет уникальные преимущества (Рис. 1.1-4):

- **Описательная аналитика** — отвечает на вопрос «*что произошло*?» и предоставляет исторические данные и отчёты о прошлых событиях и результатах: за последние три года (2023–2025) на заливку монолитной конструкции площадью 200 м² в дождливую погоду требовалось от 4,5 до 6 дней.
- **Диагностическая аналитика** — отвечает на вопрос «*почему это произошло?*», выявляя причины возникновения проблем: анализ показывает, что время заливки монолитной конструкции увеличивалось из-за дождливой погоды, которая замедляла процесс твердения бетона
- **Предиктивная аналитика** — ориентирована на будущее, прогнозирует возможные риски и сроки выполнения работ, отвечая на вопрос «*что произойдёт*?»: на основе исторических данных прогнозируется, что заливка аналогичной монолитной конструкции площадью 200 м² в дождливую погоду в 2026 году потребует примерно 5,5 дня, с учетом всех известных факторов и тенденций.

- **Прескриптивная аналитика** — даёт автоматизированные рекомендации и отвечает на вопрос «*что делать*?», позволяя компаниям выбирать оптимальные действия: Для оптимизации работ например рекомендуется: использовать специальные добавки для ускорения твердения бетона в условиях повышенной влажности; планировать заливку на периоды с наименьшей вероятностью осадков; организовать временные укрытия конструкции, что позволит сократить время работ до 4-4,5 дней даже при неблагоприятных погодных условиях.

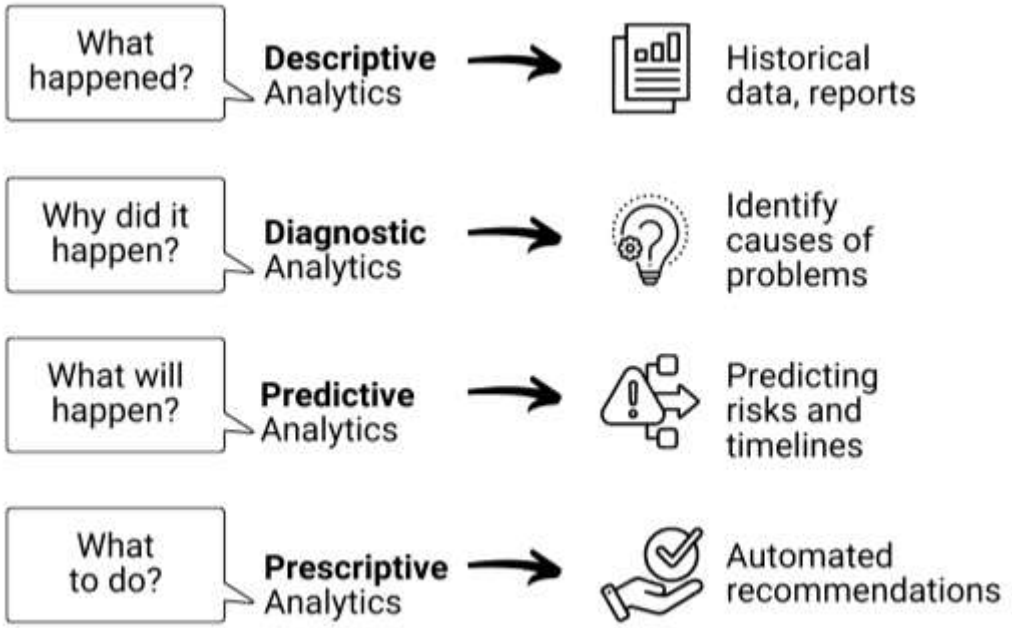


Рис. 1.1-4 Основные типы аналитики: от описания прошлого до автоматизированного процесса принятия решений.

Полноценная цифровая трансформация, предполагающая переход к системной аналитике и управлению на основе данных, требует не просто привлечения внешних подрядчиков, а формирования внутренней компетентной команды. Ключевыми участниками такой команды должны стать продакт-менеджеры, дата-инженеры, аналитики и разработчики, которые будут работать в тесном взаимодействии с бизнес-подразделениями (Рис. 4.3-9). Это сотрудничество необходимо для постановки грамотных аналитических вопросов и эффективной параметризации бизнес-задач по принятию решений. В условиях информационного общества данные становятся не просто вспомогательным инструментом, а основой прогнозирования и оптимизации.

В строительстве цифровая трансформация коренным образом меняет подходы к проектированию, управлению и эксплуатации объектов. Этот процесс называют дигитализацией информации — когда все аспекты строительного процесса переводятся в цифровую форму, подходящую для анализа.

Дигитализация информации строительного процесса

На протяжении тысячелетий объём информации, записываемой в строительстве, почти не менялся, но за последние десятилетия он вырос стремительно (Рис. 1.1-5).

> Согласно исследованию PwC® «Управляемые данными. Что нужно студентам, чтобы добиться успеха в быстро меняющемся мире бизнеса»(2015) [9], 90% всех данных в мире были созданы за последние два года (по данным на 2015 год). Однако большинство компаний не используют эти данные в полной мере, поскольку они либо остаются в разрозненных системах, либо просто архивируются без реального анализа.

За последние годы увеличение объёма данных только ускорилось, удвоившись с 15 зеттабайт в 2015 году до 181 зеттабайт в 2025 году материалам [10]. Ежедневно серверы строительных и проектировочных компаний заполняются проектной документацией, графиками работ, расчётами и калькуляциями, финансовыми отчётами. Для 2D/3D-чертежей применяются форматы DWG, DXF и DGN, а для 3D-моделей — RVT, NWC, PLN и IFC™. Текстовые документы, таблицы и презентации сохраняются в DOC, XLSX и PPT. Дополнительно к видео и изображениям со строительной площадки — в MPG и JPEG, в реальном времени поступают данные от IoT компонентов, RFID® -меток (идентификация и отслеживание) и систем управления зданиями BMS (мониторинг и контроль).

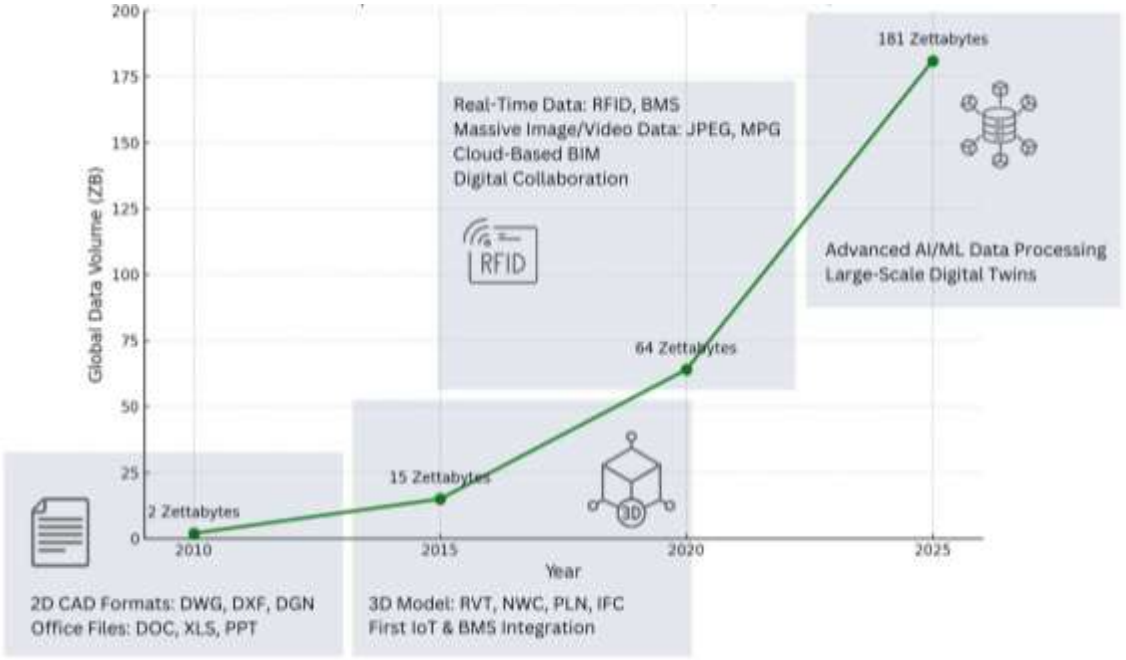


Рис. 1.1-5 Параболический рост количества данных 2010–2025 (по материалам [10]).

В условиях стремительного роста объёмов информации строительная отрасль сталкивается с необходимостью не только собирать и хранить данные, но и обеспечивать их проверку, верификацию, измеримость и аналитическую обработку. Сегодня индустрия переживает активную фазу дигитализации информации — системного преобразования всех аспектов строительной деятельности в цифровую форму, пригодную для анализа, интерпретации и автоматизации.

> Дигитализация информации означает получение информации обо всех сущностях и элементах строительного проекта и самого строительного процесса - включая те, которые мы раньше вообще не считали информацией, - и преобразование ее в формат данных, чтобы сделать информацию количественно измеримой и удобной для анализа.

В контексте строительства это означает фиксацию и выражение в цифре информации обо всех элементах проектов и всех процессов — от перемещения техники и людей на строительной площадке до погодных и климатических условий на строительной площадке, актуальных цен на материалы и процентной ставки центрального банка — с целью формирования аналитических моделей.

> *Если вы можете измерить то, о чем говорите, и выразить это в цифрах – значит, вы что-то об этом предмете знаете. Но если вы не можете выразить это количественно, ваши знания крайне ограничены и неудовлетворительны. Может это начальный этап, но это не уровень подлинного научного знания. [11].*
>
> — У. Томсон (Лорд Кельвин), 1824–1907, британский ученый

Дигитализация информации выходит далеко за рамки традиционного подхода к сбору информации, когда фиксировались лишь базовые показатели — такие как человеко-часы или фактические затраты на материалы. Сегодня практически любое событие может быть преобразовано в поток данных, пригодный для глубокого анализа с использованием инструментов продвинутой аналитики и методов машинного обучения. В строительной индустрии произошёл принципиальный сдвиг: от бумажных чертежей, Excel-смет и устных указаний — к цифровым системам (Рис. 1.2-4), в которых каждый элемент объекта становится источником данных. Даже сотрудники — от инженеров до строителей на площадке — теперь рассматриваются как совокупность цифровых переменных и наборов данных.

> По данным KPMG «Знакомые проблемы - новые подходы: Глобальный строительный обзор 2023 года», цифровые двойники, ИИ (AI) и Big Data, становятся ключевыми драйверами повышения рентабельности проектов [2].

Современные технологии не только упрощают сбор информации, делая его в значительной степени автоматическим, но и радикально снижают стоимость хранения данных. В результате компании отказываются от избирательного подхода и предпочитают сохранять весь массив информации для последующего анализа (Рис. 2.1-5), что открывает потенциальные возможности для оптимизации процессов в будущем.

> Дигитализация информации и цифровизация позволяют выявлять скрытую, ранее не используемую ценность информации. При грамотной организации данные обретают вторую жизнь: они могут быть повторно использованы, переосмыслены и интегрированы в новые сервисы и решения.

В будущем дигитализация информации, вероятно, приведёт к полной автоматизации документооборота, внедрению самоуправляемых строительных процессов и появлению новых профессий —

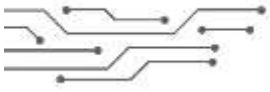
аналитиков строительных данных, экспертов по ИИ-управлению проектами и цифровых инженеров. Строительные объекты станут динамическими источниками информации, а принятие решений будет основываться не на интуиции или субъективном опыте, а на достоверных и воспроизводимых цифровых фактах.

> *Информация — это нефть 21 века, а аналитика — это двигатель внутреннего сгорания [12].*
>
> — Питер Сондергаард, старший вице-президент Gartner®

> Согласно данным IoT Analytics 2024 года [13], ожидается, что глобальные расходы на управление данными и аналитику резко вырастут с 185,5 миллиарда долларов в 2023 году до 513,3 миллиарда долларов к 2030 году, при этом среднегодовой темп роста составит 16%. Однако не все компоненты растут одинаковыми темпами: аналитика стремительно развивается, в то время как рост систем хранения данных замедляется. Аналитика обеспечит самый быстрый рост в экосистеме управления данными: по прогнозам, ее объем вырастет с 60,6 миллиарда долларов в 2023 году до 227,9 миллиарда долларов к 2030 году, что соответствует среднегодовому темпу роста в 27%.

С ускорением процесса дигитализации информации и стремительным ростом объёмов информации руководство строительных проектов и компаний сталкивается с необходимостью системно хранить, анализировать и обрабатывать разнообразные, часто разнородные данные. В ответ на этот вызов, начиная с середины 1990-х годов, отрасль начала массово переходить на электронное создание, хранение и управление документацией — от таблиц и проектных расчётов до чертежей и договоров.

Традиционные бумажные документы, требующие подписей, физического хранения, регулярного пересмотра и архивирования в шкафах, постепенно вытесняются цифровыми системами, в которых данные сохраняются в структурированном виде — в базах данных специализированных приложений.



ГЛАВА 1.2.
ТЕХНОЛОГИИ И СИСТЕМЫ УПРАВЛЕНИЯ В СОВРЕМЕННОМ СТРОИТЕЛЬСТВЕ

Цифровая революция и появление модульных MRP/ERP-систем

Эра современного хранения и обработки данных в цифровом виде началась с появлением в 1950-х годах магнитной ленты, открывшей возможность хранения и использования больших объемов информации. Следующим прорывом стало появление дисковых накопителей, которые радикально изменили подход к управлению данными в строительной отрасли.

С развитием хранилищ данных на рынок решений вышло большое количество компаний, которые начали разрабатывать модульное программное обеспечение для создания, хранения, обработки данных и автоматизации рутинных задач.

> Экспоненциальный рост объема информации и инструментов привел к необходимости разработки интегрированных модульных решений, которые не работают с отдельными файлами, а помогают управлять и контролировать поток данных в рамках разных процессов и проектов.

Первые комплексные инструменты платформы должны были не только хранить документы, но и документировать все запросы на изменения и операции в процессах: кто их инициировал, каков был объем запроса и что в итоге было записано в виде значения или атрибута. Для этих целей требовалась система, которая могла бы отслеживать точные расчеты и принятые решения (Рис. 1.2-1). Такими платформами стали первые системы MRP (Material Requirements Planning) и ERP (Enterprise Resource Planning), получившие популярность с начала 1990-х годов [14].

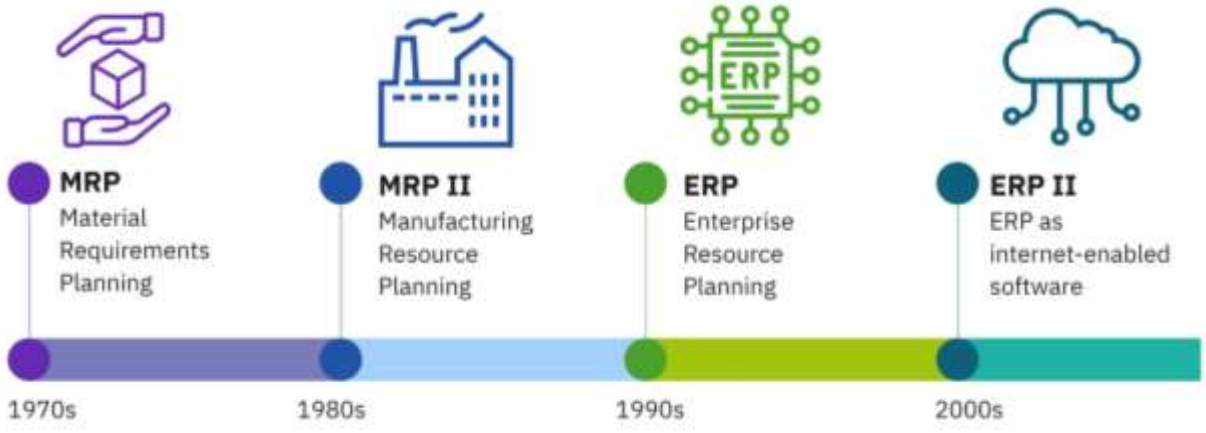


Рис. 1.2-1 Достижения в области технологий хранения данных привели к появлению ERP-систем в 1980-х годах.

Первые MRP- и ERP-системы заложили основу для эпохи цифровизации в управлении бизнес-процессами и строительными проектами. Модульные системы, изначально предназначенные для автоматизации ключевых бизнес-процессов, со временем стали интегрироваться с дополнительными, более гибкими и адаптивными программными решениями.

Эти дополнительные решения были предназначены для обработки данных и управления содержанием проектов (Рис. 1.2-2), они либо заменяли определенные модули больших систем, либо эффективно дополняли их, расширяя функциональность всей системы.

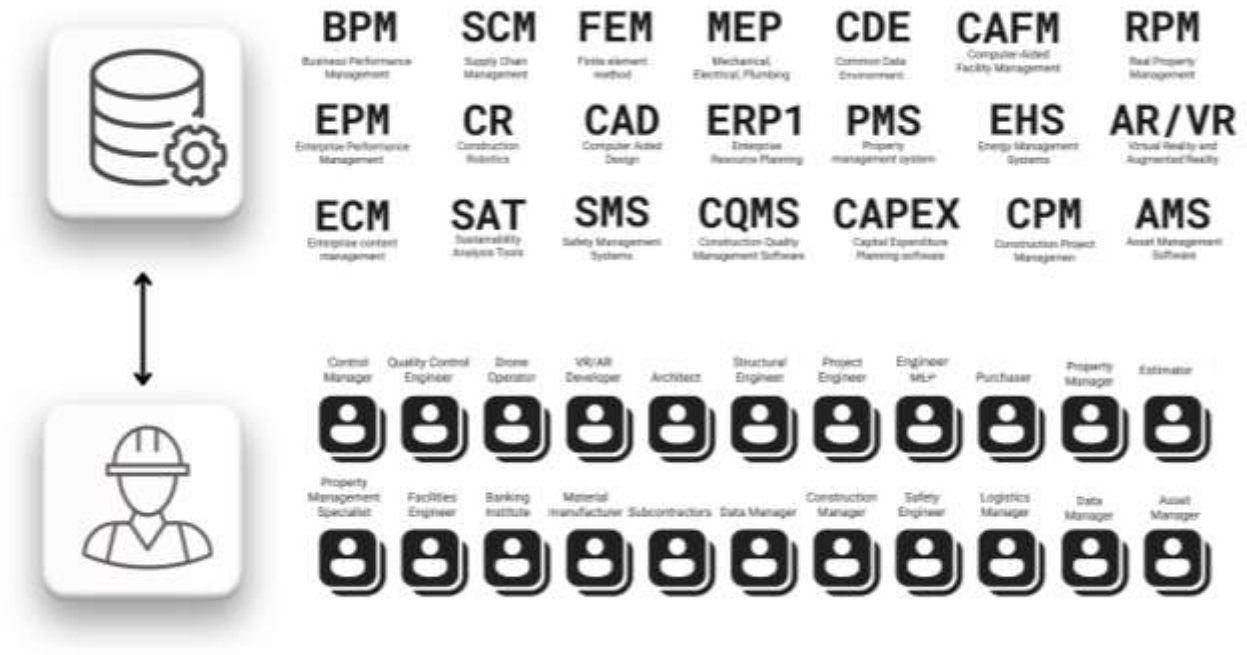


Рис. 1.2-2 Новые программные решения привлекли в бизнес целую армию менеджеров для управления потоками данных.

За последние десятилетия компании инвестировали значительные средства в модульные системы [15], воспринимая их как долгосрочные интегрированные решения.

> Согласно данным отчёта Software Path за 2022 год [16], средний бюджет на одного пользователя ERP-системы составляет 9 000 долларов США. В среднем около 26% сотрудников компании используют такие системы. Так, для организации с 100 пользователями общие затраты на внедрение ERP достигают примерно 900 000 долларов.

Инвестиции в проприетарные, закрытые модульные решения становятся всё менее оправданными на фоне стремительного развития современных, гибких и открытых технологий. Если такие вложения уже были сделаны, важно объективно переоценить роль существующих систем: действительно ли они остаются необходимыми в долгосрочной перспективе, или их функции могут быть пересмотрены и реализованы более эффективно и прозрачно.

Одна из ключевых проблем современных модульных платформ по обработке данных заключается в том, что они централизуют управление данными внутри закрытых приложений. В результате данные — основной актив компании — становятся зависимыми от конкретных программных решений, а не наоборот. Это ограничивает возможности повторного использования информации, усложняет

миграцию и снижает гибкость бизнеса в условиях быстро меняющегося цифрового ландшафта.

Если существует вероятность снижения значимости или востребованности закрытой модульной архитектуры в будущем, имеет смысл уже сегодня признать понесённые затраты как невозвратные и сфокусироваться на стратегическом переходе к более открытой, масштабируемой и адаптивной цифровой экосистеме.

> Проприетарное программное обеспечение характеризуется исключительным контролем со стороны компании-разработчика над исходным кодом и пользовательскими данными, создаваемыми в рамках использования таких решений. В отличие от программ с открытым исходным кодом, пользователи не имеют доступа к внутренней структуре приложения и не могут самостоятельно просматривать, изменять или адаптировать его под свои нужды. Вместо этого они обязаны приобретать лицензии, которые предоставляют право на использование программного обеспечения в пределах, установленных поставщиком.

Современный подход, ориентированный на данные, предлагает иную парадигму: данные должны рассматриваться как главный стратегический актив — независимый, долговечный и обособленный от конкретных программных решений. Приложения, в свою очередь, становятся лишь инструментами для работы с данными, которые могут свободно заменяться без риска потери критически важной информации.

Развитие ERP- и MRP-систем в 1990-е годы (Рис. 1.2-1) предоставило бизнесу мощные инструменты для управления процессами, однако сопровождалось и неожиданным последствием — значительным увеличением числа сотрудников, занятых обслуживанием информационных потоков. Вместо автоматизации и упрощения операционных задач такие системы нередко порождали новые уровни сложности, бюрократии и зависимости от внутренних ИТ-ресурсов.

Системы управления данными: от добычи данных к бизнес-задачам

Современные компании сталкиваются с необходимостью интеграции множества систем управления данными. Выбор систем управления данными, грамотное управление этими системами и интеграция разрозненных источников данных становится критически важной задачей для эффективности бизнеса.

В середине 2020-х годов в средних строительных компаниях можно найти сотни (а в крупных - тысячи) различных систем (Рис. 1.2-3), которые должны работать в гармонии, чтобы все аспекты строительного процесса протекали гладко и слаженно.

> Согласно исследованию Deloitte® «Управление на основе данных в цифровых капитальных проектах» 2016 года - средний строительный специалист ежедневно использует 3,3 программных приложения, однако только 1,7 из них интегрированы между собой [17].

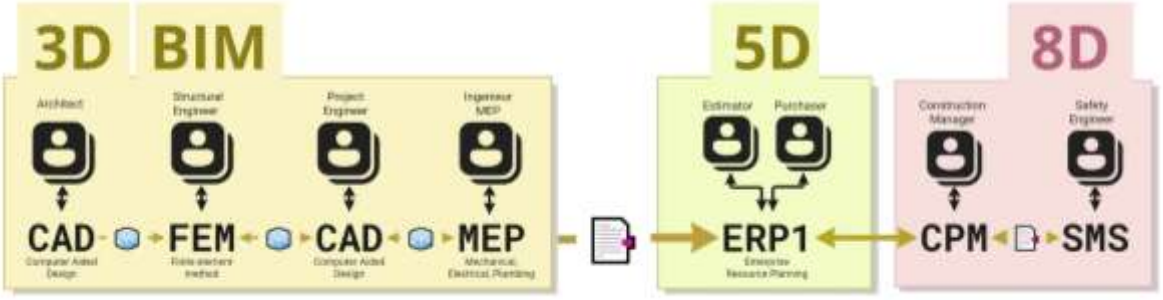


Рис. 1.2-3 Каждая бизнес-система требует профессиональной команды и ответственного менеджера для качественного управления данными.

Ниже приведен список популярных систем для средних и крупных компаний строительной отрасли, которые используются в эффективном управлении строительными проектами:

- **ERP (Enterprise Resource Planning)** – обеспечивает интеграцию бизнес-процессов, включая бухгалтерский учет, закупки и управление проектами.
- **CAPEX (Capital Expenditure Planning Software)** – используется для составления бюджета и управления финансовыми инвестициями в строительные проекты, помогает определить стоимость основных средств и инвестиций в долгосрочные активы.
- **CAD (Computer-Aided Design)** и **BIM (Building Information Modeling)** – используются для создания подробных и точных технических чертежей и 3D-моделей проектов. Основное внимание в этих системах уделяется работе с геометрической информацией.
- **MEP (Mechanical, Electrical, Plumbing)** – инженерные системы, включающие механические, электрические и сантехнические компоненты, и детализирующие внутреннюю "кровеносную" систему проекта.
- **ГИС (географические информационные системы)** – используются для анализа и планирования местности, включая картографию и пространственный анализ.
- **CQMS (программное обеспечение для управления качеством строительства)** – обеспечивает соответствие строительных процессов установленным стандартам и нормам, помогая в устранении дефектов.
- **CPM (управление строительными проектами)** – включает в себя планирование, координацию и контроль строительных процессов.
- **CAFM (Computer-Aided Facility Management)** – системы управления и обслуживания зданий.
- **SCM (Supply Chain Management)** необходима для оптимизации потока материалов и информации между поставщиками и строительной площадкой.
- **EPM (Enterprise Performance Management)** – направлено на улучшение бизнес-процессов и производительности.
- **AMS (Asset Management Software)** – используется для оптимизации использования, управления и обслуживания оборудования и инфраструктуры на протяжении всего жизненного цикла активов.
- **RPM (Real Property Management)** – включает в себя задачи и процессы, связанные с управлением и эксплуатацией зданий и земельных участков, а также связанных с ними ресурсов и активов.

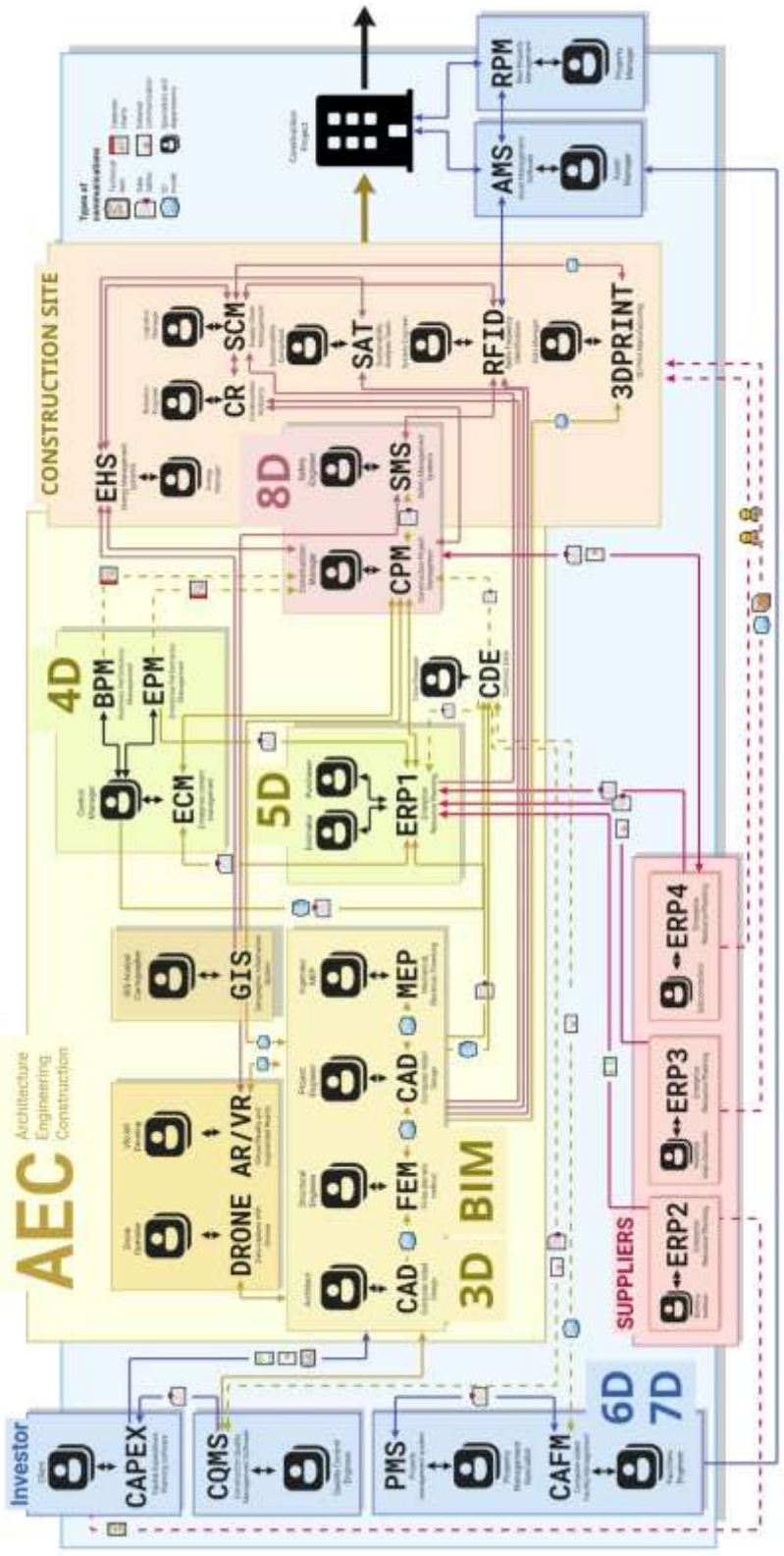


Рис. 1.2-4 Взаимосвязь систем, соединяющая процессы компании с потоком информации между различными подразделениями.

- **CAE (Computer-Aided Engineering)** – компьютерная инженерия, включает расчетные и симуляционные системы, такие как анализ методом конечных элементов (FEA) и вычислительная гидродинамика (CFD).
- **CFD (Computational Fluid Dynamics)** – вычислительная гидродинамика, моделирование потоков жидкости и газа. Подкатегория CAE.
- **CAPP (Computer-Aided Process Planning)** – компьютерное планирование технологических процессов. Используется для создания маршрутных и технологических карт.
- **CAM (Computer-Aided Manufacturing)** – автоматизированное производство, генерация управляющих программ для станков с ЧПУ.
- **PDM (Product Data Management)** – управление данными об изделии, система хранения и управления технической документацией.
- **MES (Manufacturing Execution System)** – система управления производственными процессами в реальном времени.
- **PLM (Product Lifecycle Management)** – управление жизненным циклом элемента проекта, объединяет PDM, CAPP, CAM и другие системы для полного контроля над продуктом от разработки до утилизации.

Эти и многие другие системы, включающие разнообразные программные решения, стали неотъемлемой частью современной строительной отрасли (Рис. 1.2-4). По своей сути, такие системы представляют собой специализированные базы данных с интуитивно понятными интерфейсами, обеспечивающими эффективный ввод, обработку и анализ информации на всех этапах проектирования и строительства. Интеграция цифровых инструментов между собой не только способствует оптимизации рабочих процессов, но и существенно повышает точность принимаемых решений, что положительно сказывается на сроках и качестве реализации проектов.

Но интеграции нет в половине случаев. Согласно статистике, лишь каждое второе приложение или система интегрированы с другими решениями [17]. Это указывает на сохраняющуюся фрагментированность цифровой среды и подчеркивает необходимость развития открытых стандартов и унифицированных интерфейсов для обеспечения сквозного информационного обмена в рамках строительного проекта.

Одним из главных вызовов в интеграции для современных компаний остаётся высокая сложность цифровых систем и требования к пользовательской компетенции, необходимой для эффективного поиска и интерпретации информации. Для сопровождения каждой внедрённой в бизнес системы формируется команда специалистов, которую возглавляет ключевой менеджер (Рис. 1.2-2).

> Ключевой системный менеджер играет решающую роль в нужном направлении потока данных и отвечает за качество конечной информации, подобно тому, как первые менеджеры тысячи лет назад отвечали за цифры, записанные на папирусе или глиняных табличках.

Для превращения разрозненных информационных потоков в инструмент управления, необходима способность к системной интеграции и управлению данными. В этой архитектуре менеджеры должны действовать как элементы единой сети — подобно мицелию, соединяющему отдельные части компании в целостный живой организм, способный адаптироваться и развиваться.

Корпоративный мицелий: как данные соединяются в бизнес-процессы

Процесс интеграции данных в приложения и базы данных основывается на агрегации информации, получаемой от разнообразных источников, включая различные отделы и специалистов (Рис. 1.2-4). Специалисты выискивают необходимые данные, обрабатывают их и передают в свои системы и приложения для дальнейшего использования.

Каждая система компании, состоящая из набора инструментов, технологий и баз данных — это дерево знаний, уходящее корнями в почву исторических данных и растущее, чтобы принести новые плоды в виде готовых решений: документов, расчётов, таблиц, графиков и приборных панелей (Рис. 1.2-5). Системы в компании, подобно деревьям в определенном участке леса взаимодействуют и общаются друг с другом, представляя собой сложную и хорошо структурированную систему, поддерживаемую и управляемую экспертами менеджерами.

> Система поиска и передачи информации в компании работает как сложная лесная сеть, состоящая из деревьев (систем) и мицелия грибов (менеджеров), которые выступают в роли проводников и переработчиков, обеспечивая передачу информации и ее поступление в нужные системы. Это помогает поддерживать здоровый и эффективный поток и распределение данных внутри компании.

Эксперты, подобно корням, впитывают сырые данные на начальных этапах проекта, превращая их в питательные вещества для корпоративной экосистемы. Системы управления данными и контентом (Рис. 1.2-4 - ERP, CPM, BIM и др.) выступают в роли мощных информационных магистралей, по которым эти знания циркулируют сквозь все уровни компании.

Как и в природе, где каждый элемент экосистемы играет свою роль, в бизнес-ландшафте компании каждый участник процесса - от инженера до аналитика - вносит свой вклад в рост и плодородие информационной среды. Эти системные "деревья данных" (Рис. 1.2-5) являются не просто механизмами для сбора информации, но и конкурентным преимуществом, обеспечивающим устойчивое развитие компании.

Лесные экосистемы на удивление точно отражают принципы организации цифровых корпоративных структур. Подобно многоярусной структуре леса — от подлеска до верхушек деревьев — корпоративное управление распределяет задачи по уровням ответственности и функциональным отделам.

> Глубокие и разветвленные корни деревьев обеспечивают устойчивость и доступ к питательным веществам. Аналогично, прочная организационная структура и стабильные процессы работы с качественными данными поддерживают всю информационную экосистему компании, способствуя ее устойчивому росту и развитию даже в периоды (сильного ветра) рыночной нестабильности и кризисов.

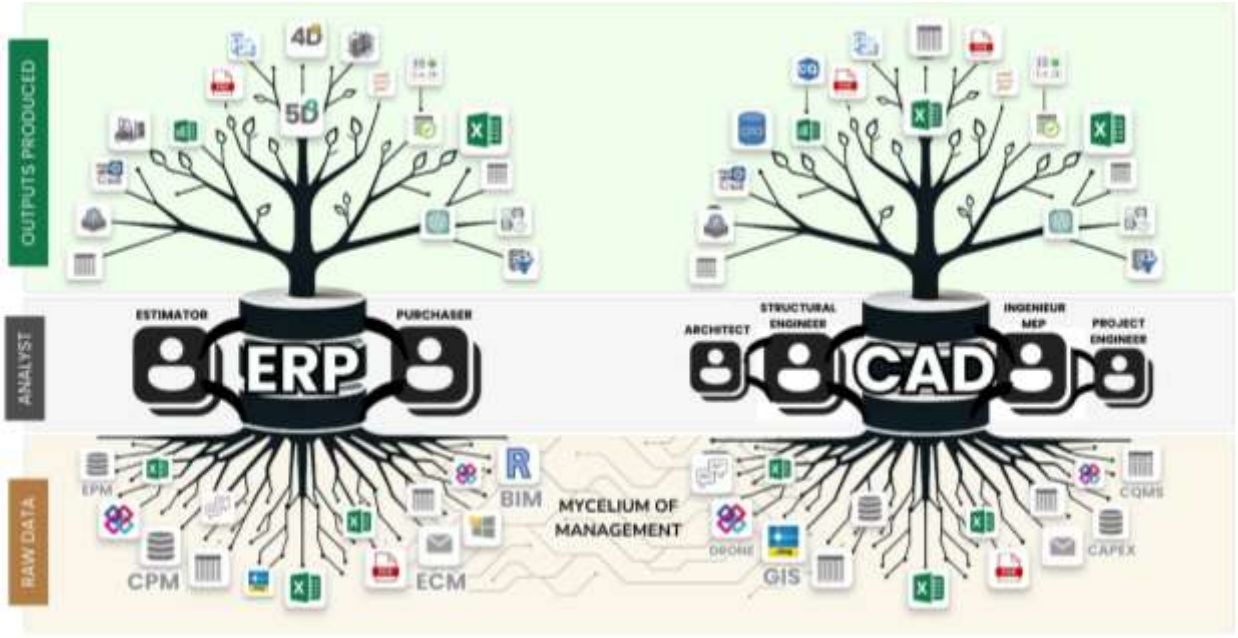


Рис. 1.2-5 Интеграция данных через различные системы подобна мицелию, соединяющему менеджеров и специалистов в единую информационную сеть.

Современное понимание масштаба в бизнесе эволюционировало. Сегодня ценность компании определяется не только ее видимой частью — "кронами" в виде финальных документов и отчётов, — но и глубиной "корневой системы" из качественно собранных и системно обработанных данных. Чем больше информации удастся собрать и обработать, тем выше становится ценность бизнеса. Компании, которые методично накапливают "компост" уже переработанных данных и умеют извлекать из них полезные инсайты, получают стратегическое преимущество.

> Историческая информация превращается в новый вид капитала, обеспечивая рост, оптимизацию процессов и конкурентное превосходство. В мире, ориентированном на данные, побеждают не те, кто больше, а те, кто знает больше.

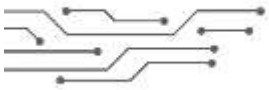
Для строительной отрасли это означает переход к управлению проектами в режиме реального времени, где все процессы — от проектирования и закупок до координации подрядчиков — будут основываться на актуальных ежедневно обновляемых данных. Интеграция информации из различных источников (ERP-системы, CAD-модели, датчики IoT на стройке, RFID) позволит строить более точные прогнозы, оперативно реагировать на изменения и избегать задержек, вызванных отсутствием актуальных данных.

> Согласно исследованию "Предприятие 2025 года, управляемое данными" (McKinsey & Company®, 2022 [18]), успешные компании будущего будут опираться на данные во всех ключевых аспектах своей деятельности — от стратегических решений до оперативных взаимодействий.

Данные перестанут быть просто инструментом аналитики и станут неотъемлемой частью всех бизнес-процессов, обеспечивая прозрачность, контроль и автоматизацию управления. Data-driven под-

ход позволит организациям минимизировать влияние человеческого фактора, снизить операционные риски и повысить прозрачность и эффективность принятия решений.

XXI век переворачивает экономическую парадигму: если раньше нефть называли "черным золотом" за ее способность приводить в движение механизмы и транспорт, то сегодня, спрессованные под давлением времени, исторические данные становятся новым стратегическим ресурсом, питающим уже не машины, а алгоритмы принятия решений, которые будут заставлять двигаться бизнес.



ГЛАВА 1.3.
ЦИФРОВАЯ РЕВОЛЮЦИЯ И ВЗРЫВ ДАННЫХ

Начало бума объемов данных как эволюционная волна

Строительная отрасль переживает беспрецедентный информационный взрыв. Если представить бизнес как дерево знаний (Рис. 1.2-5), питающееся данными, то текущий этап цифровизации можно сравнить с бурным ростом растительности в каменноугольный период — эпоху, в которую биосфера Земли преобразилась благодаря стремительному накоплению биомассы (Рис. 1.3-1).

В условиях глобального цифрового развития объём информации в строительной сфере удваивается ежегодно. Современные технологии позволяют собирать данные в фоновом режиме, анализировать их в реальном времени и использовать на масштабах, которые ещё недавно казались невозможными.

> Согласно закону Мура, сформулированному Гордоном Муром (соучредителем Intel®), плотность и сложность интегральных схем, а также объемы обрабатываемых и хранимых данных удваиваются приблизительно каждые два года [19].

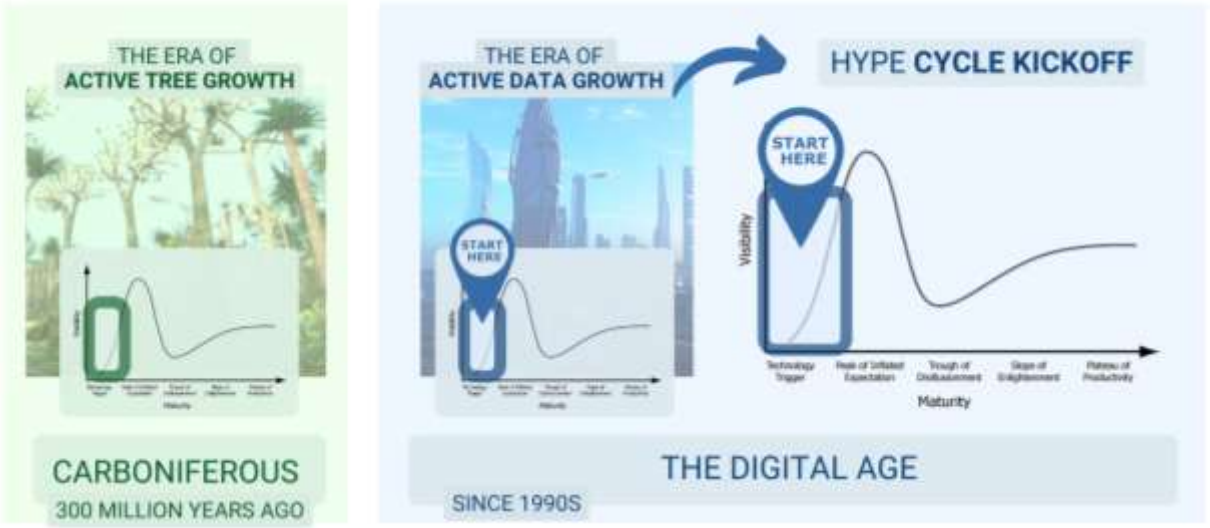


Рис. 1.3-1 Начало цифровизации привело к экспоненциальному росту данных, подобно всплеску растительности в угольную эпоху.

Если древние мегалитические сооружения, такие как Гёбекли-Тепе (Турция), не оставили после себя документированных знаний, пригодных для повторного использования, то сегодня цифровые технологии делают возможным накопление и пере использование информации. Это можно сравнить с эволюционным переходом от споровых к семенным растениям (ангиоспермы): появление семени дало толчок к широкому распространению жизни на планете. (Рис. 1.3-2).

Аналогично, данные из прошлых проектов становятся своего рода «цифровыми семенами» — носителями знаний ДНК, которые можно масштабировать и использовать в новых проектах и про-

дуктах. Появление современных инструментов искусственного интеллекта — машинного обучения и больших языковых моделей (LLM), таких как ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok — позволяет автоматически извлекать, интерпретировать и применять данные в новых контекстах.

> Подобно тому, как семена произвели революцию в распространении жизни на изначально безжизненной планете, "семена данных" становятся основой для автоматического появления новых информационных структур и знаний, позволяя цифровым экосистемам развиваться самостоятельно и адаптироваться к меняющимся требованиям пользователей.

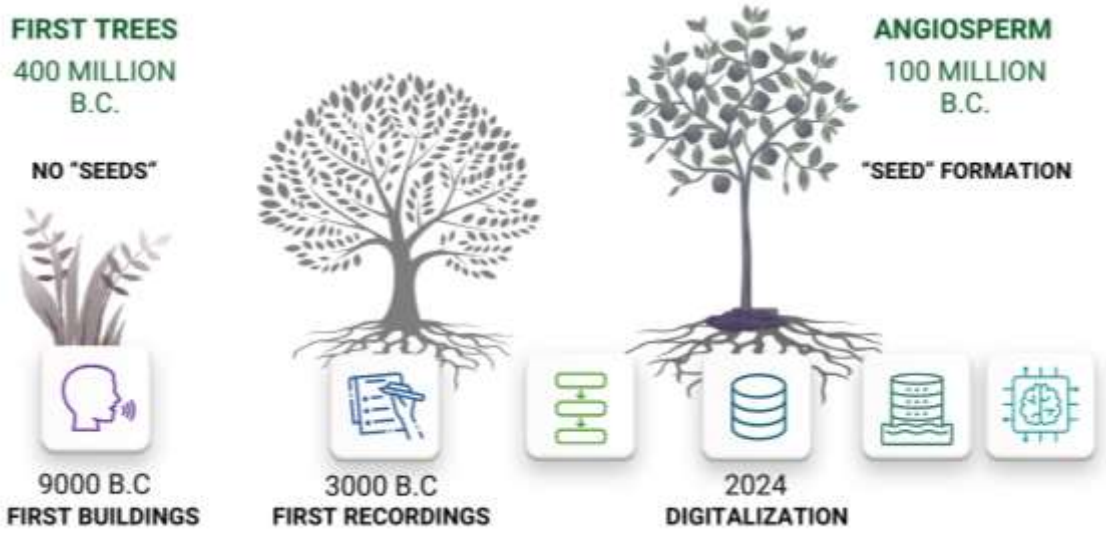


Рис. 1.3-2 Цифровые "семена данных" играют ту же эволюционную роль, что и ангиоспермы — цветковые растения, преобразившие экосистему Земли.

Мы стоим на пороге новой эры в строительстве, где взрыв данных и активное распространение "семян данных" — структурированной информации из прошлых и текущих проектов — формируют фундамент цифрового будущего отрасли. Их «опыление» при помощи больших языковых моделей (LLM) позволяет не просто наблюдать за цифровыми изменениями, но активно участвовать в создании самообучающихся, адаптивных экосистем. Это не эволюция — это цифровая революция, в которой данные становятся главным строительным материалом новой реальности.

> *Объем данных в строительной отрасли резко возрастает за счет информации, поступающей из различных дисциплин на протяжении всего жизненного цикла строительных проектов. Это огромное накопление данных подтолкнуло строительную индустрию к эре Больших Данных [20].*
>
> — Профессор Ханг Ян, факультет гражданского строительства и архитектуры, Уханьский технологический университет, Ухань, Китай

Рост объема данных в информационную эпоху напоминает эволюционные процессы в природе: как развитие лесов изменило древний ландшафт планеты, так и нынешний информационный взрыв меняет ландшафт всей строительной отрасли.

Объем данных, генерируемых в современной компании

За последние два года было создано 90% всех существующих в мире данных [21]. По состоянию на 2023 год каждый человек, включая специалистов строительной отрасли, генерирует около 1,7 мегабайта данных в секунду [22], а общий объем данных в мире достигнет 64 зеттабайт в 2023 году и, по прогнозам, превысит 180 зеттабайт, или 180*10^15 мегабайт, к 2025 году [23].

> Этот информационный взрыв имеет исторический прецедент — изобретение печатного станка Иоганном Гутенбергом в XV веке. Всего через пятьдесят лет после его появления количество книг в Европе удвоилось: за несколько десятилетий было напечатано столько же книг, сколько создавалось вручную за предшествующие 1200 лет [24]. Сегодня мы наблюдаем еще более стремительный рост: объем данных в мире удваивается каждые три года.

С учётом текущих темпов роста данных строительная отрасль за несколько следующих десятилетий потенциально может сгенерировать такой же объём информации, какой был накоплен за всю её предыдущую историю.

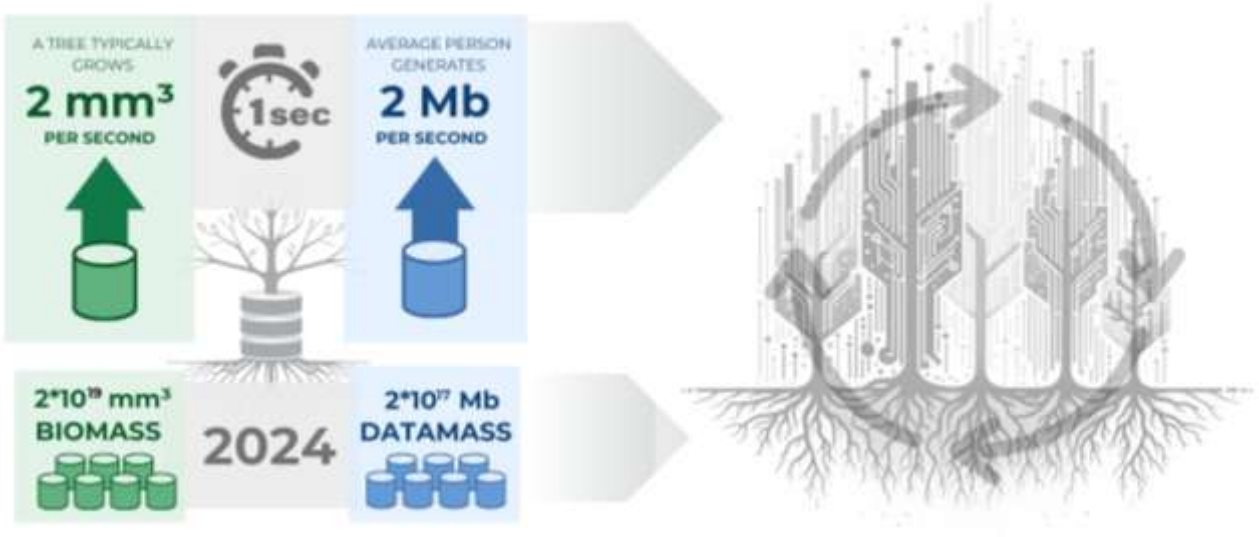


Рис. 1.3-3 Ежедневное хранение данных каждым сотрудником на серверах компании способствует постоянному росту объема данных.

В современном мире строительного бизнеса даже небольшие компании ежедневно генерируют огромное количество разноформатной информации и цифровой след даже небольшой строитель-

ной компании может достигать десятков гигабайт в день — от моделей и чертежей до фотофиксации и датчиков на объекте. Если приять, что каждый специалист в среднем создает около 1,7 МБ данных в секунду, то это эквивалентно примерно 146 ГБ в день, или 53 ТБ в год (Рис. 1.3-3).

При активной работе коллектива из 10 человек в течение всего 3 часов ежедневно, совокупный объем информации, генерируемый в день, достигает 180 гигабайт (Рис. 1.3-4).

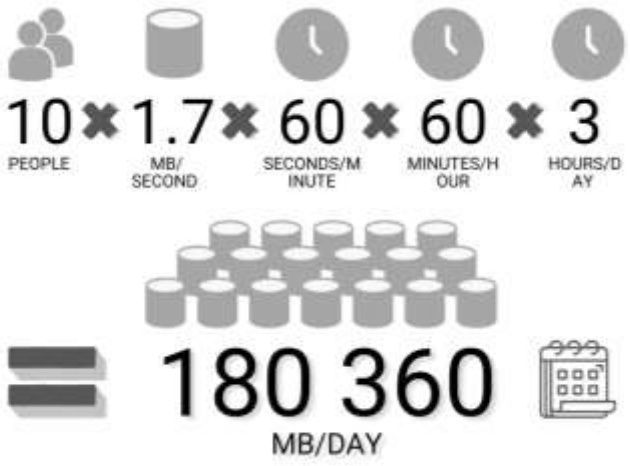


Рис. 1.3-4 Компания из 10 человек генерирует примерно 50–200 гигабайт данных в день.

Если предположить, что 30% рабочих данных - новые (остальное перезаписывается или удаляется), то фирма из 10 человек может создавать порядка нескольких сотен гигабайт новых данных в месяц (реальные показатели зависят от типа деятельности компании).

Таким образом, становится очевидным: мы не просто генерируем всё больше данных — мы сталкиваемся с растущей потребностью в их эффективном управлении, хранении и долгосрочной доступности. И если ранее данные могли «лежать» на локальных серверах без особых затрат, то в условиях цифровой трансформации всё больше компаний начинают использовать облачные решения как основу своей информационной инфраструктуры.

Стоимость хранения данных: экономический аспект

В последние годы всё больше компаний передают хранение данных в облачные сервисы. Например, если компания размещает половину своих данных в облаке, при средней цене 0,015 доллара за гигабайт в месяц, её расходы на хранение могут увеличиваться на 10–50 долларов [25] ежемесячно.

Для небольшой компании с типичными моделями генерации данных расходы на облачное хранение могут составлять от сотен до потенциально более тысячи долларов в месяц (Рис. 1.3-5) через

несколько лет, создавая потенциально значительную финансовую нагрузку.

> Согласно исследованию Forrester «Предприятия передают хранение данных на аутсорсинг, поскольку сложность растет» [26], в котором приняли участие 214 руководителей, принимающих решения в области технологической инфраструктуры, более трети организаций передают системы хранения данных на аутсорсинг, чтобы справиться с растущим объемом и сложностью операций с данными, при этом почти две трети предприятий предпочитают модель, основанную на подписке.

Рис. 1.3-5 Перемещение данных в облако может увеличить ежемесячные затраты на хранение до 2 000 долларов даже для компании с численностью всего 10 сотрудников.

Ситуация усложняется ускоренным переходом на облачные технологии, такие как CAD (BIM), CAFM, PMIS и ERP-системы, которые дополнительно увеличивают затраты на хранение и обработку данных. В итоге компании вынуждены искать способы оптимизации расходов и снижения зависимости от облачных провайдеров.

С 2023 года, с активным развитием больших языковых моделей (LLM), подходы к хранению данных начали меняться. Всё больше компаний задумываются о возврате контроля над данными, поскольку обработка информации на собственных серверах становится безопаснее и выгоднее.

В этом контексте на первый план выходит тенденция к отказу от облачных систем хранения и обработки только нужных данных в пользу локального развертывания корпоративных LLM и ИИ-решений. Как отметил генеральный директор Microsoft в одном из своих интервью [27], вместо того, чтобы полагаться на несколько отдельных приложений или облачных решений SaaS для выполнения различных задач, агенты ИИ будут управлять процессами в базах данных, автоматизируя функции разных систем.

> *[..] старый подход к этому вопросу [обработки данных] заключался в следующем: если вспомнить, как различные бизнес-приложения справлялись с интеграцией, то они использовали коннекторы. Компании продавали лицензии на эти коннекторы, и вокруг этого формировалась бизнес-модель. SAP [ERP] — один из классических примеров: доступ к данным SAP был возможен только при наличии соответствующего коннектора. Поэтому мне кажется, что нечто подобное появится и в случае взаимодействия [ИИ] агентов [..]. Подход, по крайней мере, который мы принимаем, заключается в следующем: я думаю, что концепция существования бизнес-приложений, вероятно, рухнет в эпоху [ИИ] агентов. Потому что, если задуматься, они, по сути, представляют собой базы данных с кучей бизнес-логики.*
>
> — Сатья Наделла, генеральный директор Microsoft, интервью каналу BG2, 2024 г. *[28]*

В этой парадигме data-driven подход с применением LLM выходит за рамки классических систем. Искусственный интеллект становится посредником между пользователем и данными (Рис. 2.2-3, Рис. 2.2-4), устраняя необходимость в многочисленных посреднических интерфейсах и повышая эффективность бизнес-процессов. Подробнее про этот подход работы с данными мы поговорим в главе «Превращение хаоса в порядок и снижение сложности».

Пока архитектура будущего только формируется, компании уже сталкиваются с последствиями прошлых решений. Массовая цифровизация последних десятилетий, сопровождавшаяся внедрением разрозненных систем и бесконтрольным накоплением данных, привела к новой проблеме — информационной перегрузке.

Границы накопления данных: от массы к смыслу

Современные системы компаний успешно развиваются и функционируют при управляемом росте, когда объем данных и количество приложений находятся в балансе с возможностями ИТ-отделов и менеджеров. Однако в последние десятилетия цифровизация привела к неконтролируемому увеличению объема и сложности данных, что вызвало эффект перенасыщения информационной экосистемы компаний.

Сегодня серверы и хранилища подвергаются беспрецедентному наплыву необработанной и разноформатной информации, которая не успевает превратиться в компост и стремительно теряет актуальность. Ограниченные ресурсы компании не справляются с этим потоком, а данные накапливаются в изолированных хранилищах (так называемых «силосах»), требующих ручной обработки для извлечения полезной информации.

В результате, подобно лесу, заросшему плющом и покрытому плесенью, современные системы управления компаниями часто страдают от информационной перегрузки. В основе корпоративной экосистемы вместо питательного информационного гумуса формируются изолированные участки разноформатных данных, что неизбежно приводит к снижению общей эффективности бизнес-процессов.

Долгий период экспоненциального роста объемов данных, наблюдаемый последние 40 лет, неизбежно сменится фазой насыщения и последующего охлаждения. Когда хранилища достигнут предела, произойдёт качественный сдвиг: данные перестанут быть просто объектом хранения, а станут стратегическим ресурсом.

С развитием искусственного интеллекта и машинного обучения компании получают возможность снизить издержки на обработку информации и перейти от количественного роста к качественному использованию данных. В ближайшее десятилетие строительной отрасли предстоит сместить фокус — от создания всё новых массивов данных к обеспечению их структурированности, целостности и аналитической ценности.

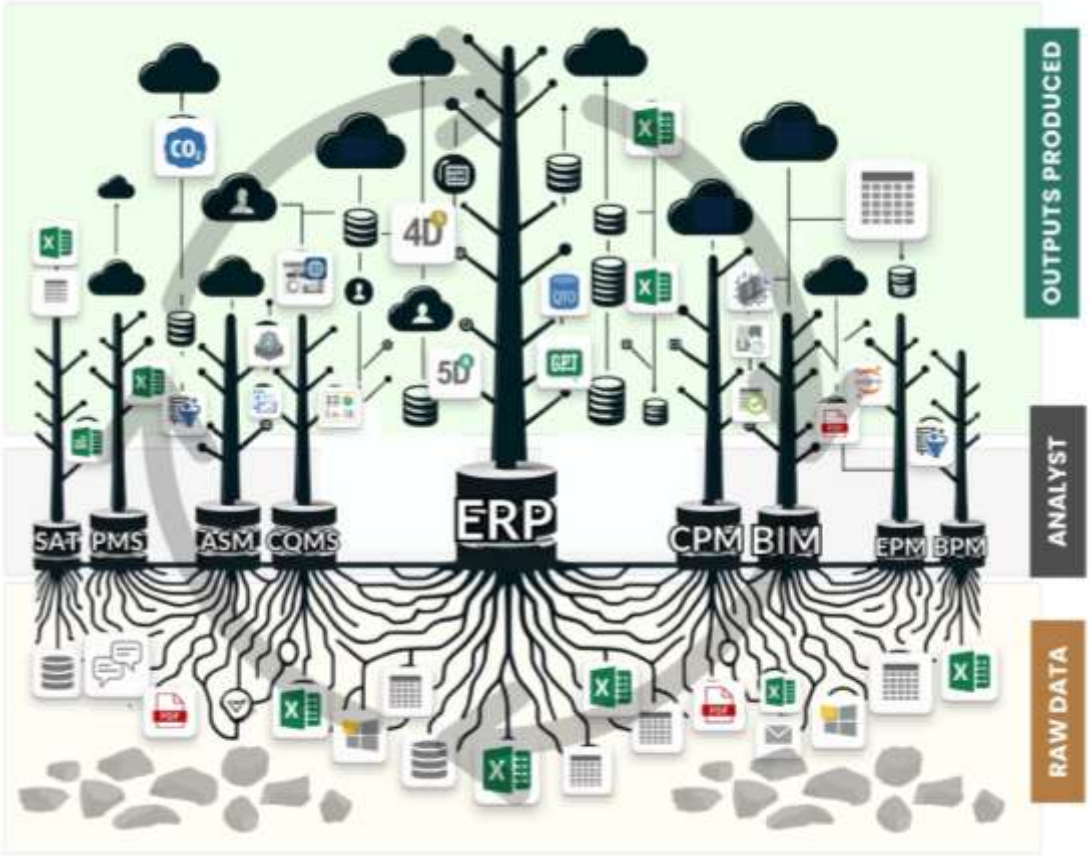


Рис. 1.3-6 Изолированные источники данных препятствуют обмену информацией между системами данных.

> Главная ценность больше не в объёме информации, а в способности автоматически её интерпретировать и превращать в прикладные знания, полезные для принятия управленческих решений. Чтобы данные стали по-настоящему полезными, ими нужно грамотно управлять: собирать, проверять, структурировать, хранить и анализировать в контексте конкретных бизнес-задач.

Процесс анализа данных в компании похож на цикл жизни и распада деревьев в лесу и появления новых молодых и сильных деревьев: зрелые деревья отмирают, перегнивают и становятся питательной средой для новых ростков. Готовые и завершенные процессы по окончании своего применения попадают в информационную экосистему компании, становясь в итоге информационным гумусом, питающим в будущем рост новых систем и данных.

Однако на практике этот цикл часто нарушается. Вместо органического обновления формируется слоистый хаос — подобие геологических пластов, в которых новые системы наслаиваются поверх старых, без глубокой интеграции и структурирования. В результате возникают разрозненные информационные «силосы», мешающие циркуляции знаний и затрудняющие управление данными.

Дальнейшие шаги: от теории данных к практическим изменениям

Эволюция данных в строительстве — это путь от глиняных табличек до современных модульных платформ. Сегодня вызов заключается не в сборе информации, а в создании структуры, которая превращает разрозненные и разноформатные данные в стратегический ресурс. Независимо от вашей роли — руководителя компании или рядового инженера — понимание ценности данных и умение работать с ними в будущем станет ключевым профессиональным навыком.

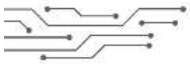
Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные подходы в ваших повседневных задачах:

- Проведите личный аудит информационных потоков
  - Составьте список всех систем и приложений, с которыми вы работаете ежедневно
  - Отметьте, где вы тратите больше всего времени на поиск или перепроверку данных
  - Определите свои ключевые источники информации
  - Проанализируйте ваш текущий ландшафт приложений на предмет избыточности и дублирования функций
- Стремитесь двигаться в процессах по уровням аналитической зрелости
  - Начинайте работу с вашими задачами с описательной аналитики (что произошло?)
  - Постепенно внедряйте диагностическую (почему это произошло?)
  - Подумайте как в процессах вы можете перейти к предиктивной (что произойдет?) и прескриптивной (что делать?) аналитике
- Начните структурировать свои рабочие данные

- ☐ Внедрите единую систему наименования файлов и папок которые вы часто используете в своей работе
- ☐ Создайте шаблоны для часто используемых документов и отчетов
- ☐ Регулярно архивируйте завершенные проекты с четкой структурой

Даже если вы не можете изменить всю информационную инфраструктуру в своей команде или компании, начните с собственных процессов и небольших улучшений в своей повседневной работе. Помните, что настоящая ценность данных проявляется не в их объеме, а в способности извлекать из них практическую пользу. Даже небольшие, но правильно структурированные и проанализированные массивы информации могут дать значительный эффект, если они интегрированы в процессы принятия решений.

В следующих частях книги мы перейдем к конкретным методам и инструментам работы с данными, рассмотрим способы преобразования неструктурированной информации в структурированные наборы, изучим технологии автоматизации анализа и подробно разберем, как построить эффективную аналитическую экосистему в строительной компании.



II ЧАСТЬ
КАК СТРОИТЕЛЬНЫЙ БИЗНЕС ТОНЕТ В ХАОСЕ ДАННЫХ

Вторая часть посвящена критическому анализу проблем, с которыми сталкиваются строительные компании при работе с возрастающими объемами данных. Детально рассматриваются последствия фрагментации информации и феномен "данных в силосах", препятствующий эффективному принятию решений. Исследуется проблематика HiPPO-подхода (Highest Paid Person's Opinion) и его влияние на качество управленческих решений в строительных проектах. Оценивается влияние динамичности бизнес-процессов и их растущей сложности на информационные потоки и операционную эффективность. Приводятся конкретные примеры того, как избыточная сложность систем увеличивает затраты и снижает гибкость организаций. Особое внимание уделяется ограничениям, создаваемым проприетарными форматами, и перспективам использования открытых стандартов в строительной отрасли. Представлена концепция перехода к программным экосистемам на основе ИИ и LLM, которые минимизируют избыточную сложность и технические барьеры.

ГЛАВА 2.1.
ФРАГМЕНТАЦИЯ И СИЛОСЫ ДАННЫХ

Чем больше инструментов, тем эффективнее бизнес?

На первый взгляд может показаться, что увеличение количества цифровых инструментов ведёт к росту эффективности. Однако на практике всё обстоит иначе. С каждым новым решением, будь то облачный сервис, устаревшая система или очередной Excel-отчёт, компания добавляет ещё один слой в свой цифровой ландшафт — слой, который зачастую не интегрируется с остальными (Рис. 2.1-1).

Данные можно сравнить с углём или нефтью: они формируются годами, «спрессовываясь» под слоями хаоса, ошибок, неструктурированных процессов и забытых форматов. Чтобы извлечь из них по-настоящему полезную информацию, компаниям приходится буквально пробиваться через залежи устаревших решений и цифрового шума.

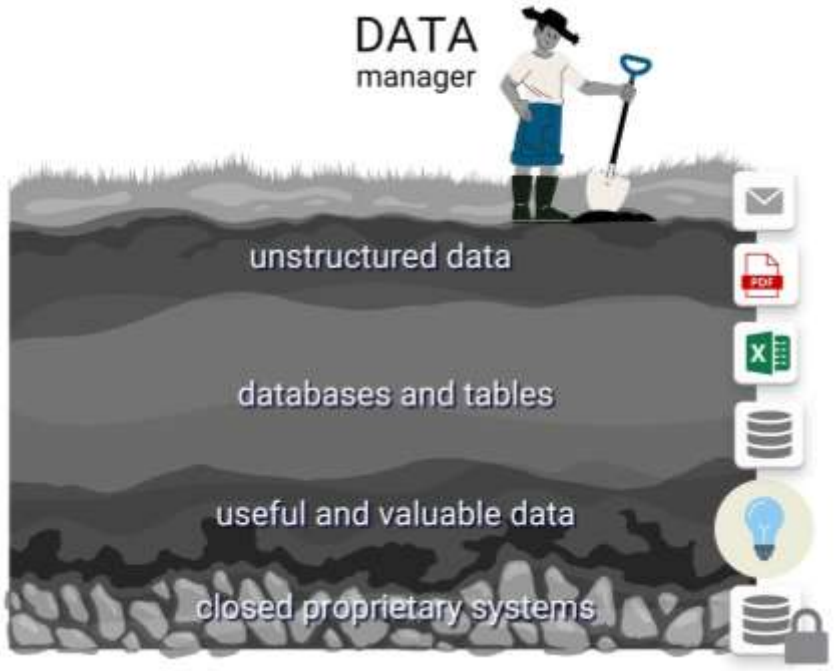


Рис. 2.1-1 Разноформатные данные образуют обособленные слои — даже «золотые» инсайты теряются в геологических породах системной сложности.

Каждое новое приложение оставляет после себя след: файл, таблицу или целый изолированный «силос» на сервере. Один слой — это глина (устаревшие и забытые данные), другой — песок (разрозненные таблицы и отчёты), третий — гранит (закрытые проприетарные форматы, неподдающиеся интеграции). Со временем цифровая среда компании всё больше напоминает пластовое хранилище с бесконтрольным накоплением информации, где ценность теряется в глубине серверов компании.

С каждым новым проектом и каждой новой системой усложняется не только инфраструктура, но и путь к полезным качественным данным. Чтобы добраться до ценной «породы», необходимо проводить глубокую очистку, структурировать информацию, «дробить», группировать её на осмысленные фрагменты и извлекать стратегически важные сведения через аналитику и моделирование данных.

> *Данные - ценная вещь, и они прослужат дольше, чем сами системы [обрабатывающие данные] [29].*
>
> — Тим Бернерс-Ли, отец Всемирной паутины и создатель первого сайта

Прежде чем данные смогут стать «ценной вещью» и надежной основой для принятия решений, они должны пройти тщательную подготовку. Именно грамотная предобработка превращает разрозненные сведения в структурированный опыт, полезный информационный гумус, который затем становится инструментом прогнозирования и оптимизации.

> Существует заблуждение, будто для начала анализа необходимы идеально чистые данные, однако на практике умение работать с «грязными» данными — неотъемлемая часть процесса.

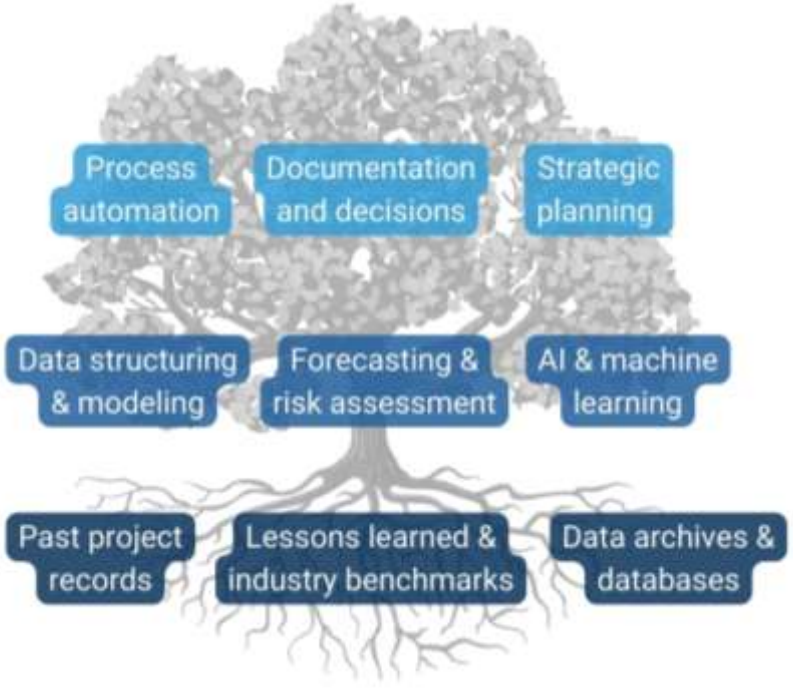


Рис. 2.1-2 Данные являются корневой системой и фундаментом бизнеса, который в свою очередь основан на процессах принятия решений.

Пока технологии не стоят на месте, ваш бизнес тоже должен двигаться вперёд и учится создавать ценность из данных. Как нефтяные и угольные компании создают инфраструктуру для добычи полезных ископаемых, так и бизнес должен научиться самостоятельно на своих серверах управлять грамотно потоками новой информации и извлекать ценные сведения из неиспользованных, разноформатных и устаревших данных, превращая их в стратегический ресурс.

Создание месторождений (хранилищ данных) — это первый шаг. Даже самые мощные инструменты не решают проблему изоляции данных и разноформатных данных, если компании продолжают работать в разрозненных системах. Когда данные существуют отдельно друг от друга, не пересекаясь и не обмениваясь информацией, бизнес сталкивается с эффектом «силосов данных». Вместо единой, согласованной инфраструктуры компании вынуждены тратить ресурсы на объединение и синхронизацию данных.

Силосы данных и их влияние на эффективность компании

Представьте, что вы строите жилой комплекс, но у каждой бригады — свой собственный проект. Одни возводят стены, другие прокладывают коммуникации, третьи — укладывают дороги, не сверяясь между собой. В результате трубы не совпадают с проёмами в стенах, лифтовые шахты не соответствуют этажности, а дороги приходится разбирать и перекладывать заново.

Подобная ситуация — не просто гипотетический сценарий, а реальность многих современных строительных проектов. Из-за большого количества генеральных и субподрядчиков, работающих с различными системами и без единого координирующего центра, процесс превращается в череду бесконечных согласований, переделок и конфликтов. Всё это приводит к значительным задержкам и кратному удорожанию проектов.

Классическая ситуация при которой на строительной площадке возникает простой: опалубка готова, однако поставка арматуры не поступила в срок. При проверке информации в различных системах общение происходит примерно следующем образом:

- **Прораб** на строительной площадке 20-го числа пишет менеджеру проекта**:** «*Мы закончили установку опалубки, где арматура?*»
- **Менеджер проекта** (PMIS) отделу снабжения**:** — «*Опалубка готова. В моей системе [PMIS] стоит, что арматура должна была прийти 18-го числа. Где арматура?»*
- **Специалист по снабжению** (ERP): — «В нашей ERP указано, что поставка буде 25-го числа».
- **Инженер по данным** или ИТ-отдел (отвечает за интеграции): — В PMIS стоит дата 18-е число, в ERP – 25-е. Связи по OrderID между ERP и PMIS нет, поэтому данные не синхронизированы. Это типичный пример информационного разрыва.
- **Менеджер проекта** генеральному директору — «*Поставка арматуры задерживается, площадка стоит, а кто несёт ответственность — непонятно*».

Причиной инцидента стала изолированность данных в разрозненных системах. Интеграция и унификация источников данных, создание единого хранилища информации, а также автоматизация через ETL-инструменты (Apache NiFi, Airflow или n8n) позволяют устранить разрозненность между системами. Эти и другие методы и инструменты будут подробно рассмотрены в последующих разделах книги.

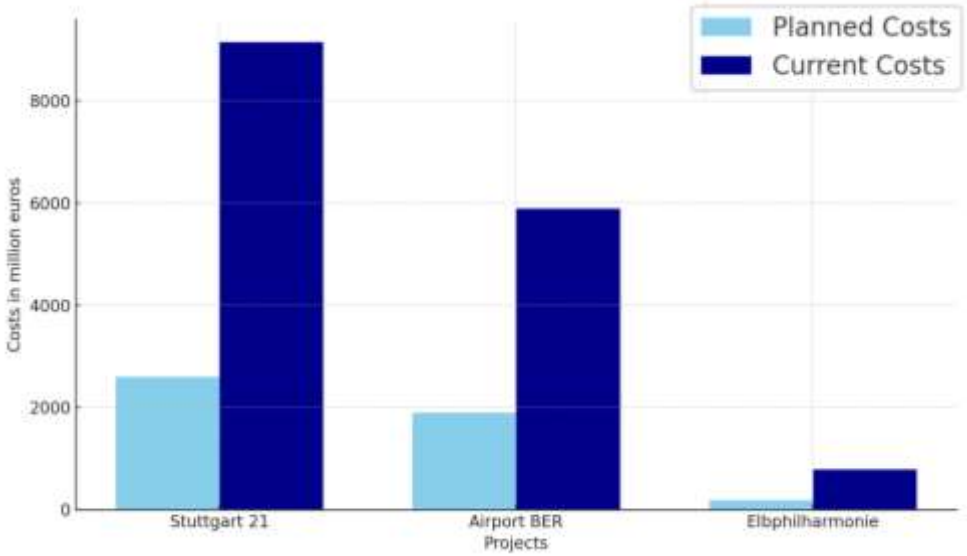


Рис. 2.1-3 Сравнение запланированных и фактических затрат на крупные инфраструктурные проекты в Германии.

То же самое происходит с корпоративными системами: сначала создаются изолированные решения, а затем приходится тратить огромные бюджеты на их интеграцию и согласование. Если бы с самого начала был продуманы модели данных и связи, потребность в интеграции вообще бы не возникла. Разрозненные данные создают хаос в цифровом мире, подобно несогласованному строительному процессу.

> Согласно исследованию KPMG «Cue construction 4.0: Время делать или ломать» 2023 года, только 36% компаний эффективно обмениваются данными между отделами, тогда как 61% сталкиваются с серьезными проблемами из-за изолированных «силосов» данных [30].

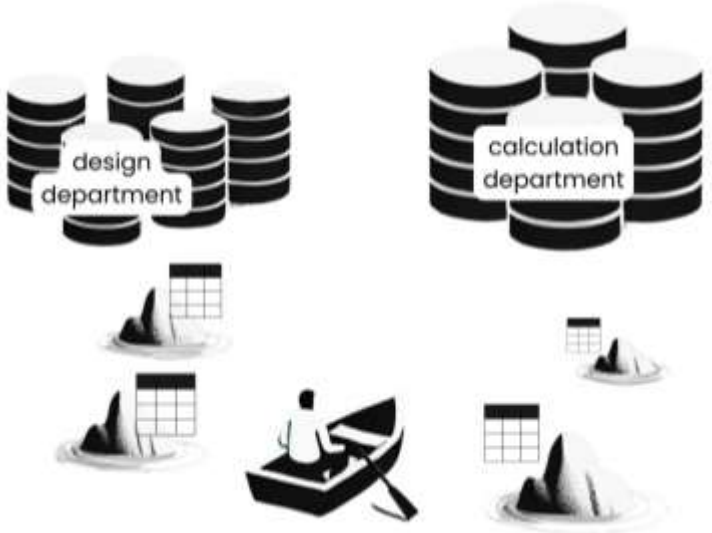


Рис. 2.1-4 Годами собираемые трудноизвлекаемые данные накапливаются в изолированных хранилищах «силосах», рискуя никогда не быть использованными.

Данные компании хранятся в изолированных системах, словно отдельные деревья, разбросанные по ландшафту. Каждое из них содержит ценную информацию, но отсутствие связей между ними мешает создать единую, взаимосвязанную экосистему. Такая разрозненность мешает потоку данных и ограничивает способность организации видеть полную картину. Соединение этих силосов это крайне долгий и сложный процесс выращивания грибного мицелия на уровне менеджмента, который научится передавать отдельные куски информации между системами.

Согласно исследованию WEF 2016 года, одним из главных барьеров для цифровой трансформации является отсутствие единых стандартов данных и фрагментированности.

> *Строительная отрасль — одна из самых фрагментированных в мире и зависит от слаженного взаимодействия всех участников цепочки создания стоимости [5].*
>
> — World Economic Forum 2016: Shaping the Future of Construction

Дизайнеры, менеджеры, координаторы и разработчики часто предпочитают работать автономно, избегая сложностей координации. Это естественное стремление приводит к созданию информационных "силосов", в которых данные изолируются внутри отдельных систем. Чем больше таких изолированных систем, тем сложнее наладить их взаимодействие. Со временем каждая система получает собственную базу данных и специализированный отдел поддержки из менеджеров (Рис. 1.2-4), что ещё больше усложняет интеграцию.

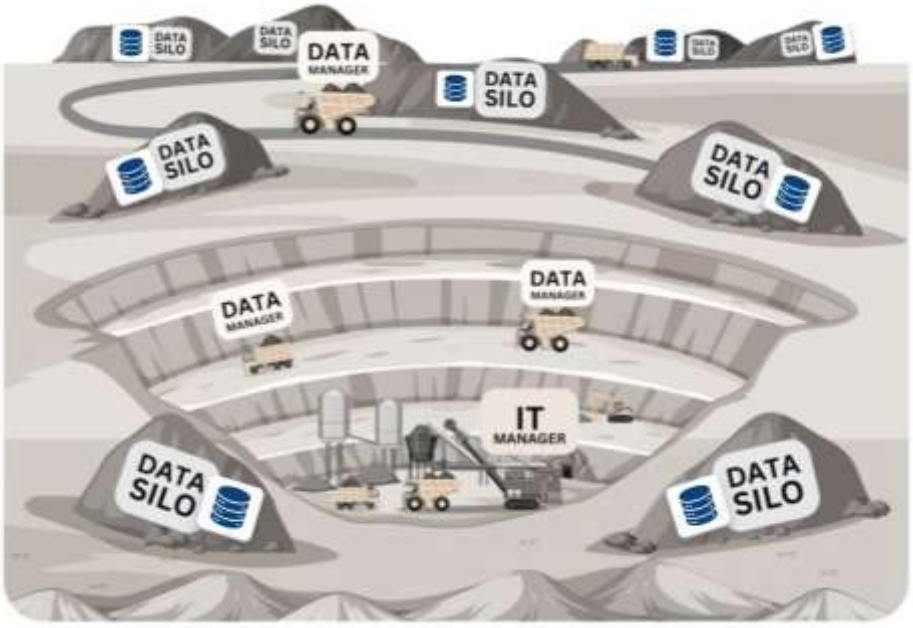


Рис. 2.1-5 Каждая система стремится создать свой уникальный силос данных, который необходимо обрабатывать подходящими инструментами [31].

Замкнутый круг в корпоративных системах выглядит так: компании инвестируют в сложные изо-

лированные решения, затем сталкиваются с высокими затратами на их интеграцию, а разработчики, понимая сложность объединения систем, предпочитают работать в своих замкнутых экосистемах. Всё это увеличивает разрозненность IT-ландшафта и усложняет переход на новые решения (Рис. 2.1-5). Менеджеры в итоге критикуют разобщённость данных, но редко анализируют её причины и способы предотвращения. Руководители жалуются на устаревшие IT-системы, но их замена требует значительных инвестиций и редко приносит ожидаемый результат. В результате даже попытки борьбы с этой проблемой нередко только усугубляют ситуацию.

Главная причина разобщённости — приоритет приложений над данными. Компании сначала разрабатывают отдельные системы или покупают готовые решения у вендоров, а затем пытаются их объединить, создавая дублирующиеся и несовместимые между собой хранилища и базы данных.

Преодоление проблемы фрагментированности требует кардинально нового подхода — приоритета данных над приложениями. Компании должны в первую очередь разрабатывать стратегии управления данными и модели данных, а затем создавать системы или покупать решения, которые работают с единым набором информации, а не формируют новые барьеры.

> *Мы вступаем в новый мир, в котором данные могут оказаться важнее программного обеспечения.*
>
> — Тим О'Рейли, генеральный директор O'Reilly Media, Inc.

Исследование McKinsey Global Institute «Переосмысление строительства: путь к повышению производительности» (2016) демонстрирует, что строительная отрасль отстает от других секторов в цифровой трансформации [32]. Согласно отчету, внедрение автоматизированного управления данными и цифровых платформ может значительно повысить производительность и снизить потери, связанные с несогласованностью процессов. Эту необходимость цифровой трансформации подчеркивает и отчет Игана (UK, 1998) [33], который выделяет ключевую роль интегрированных процессов и совместного подхода в строительстве.

В итоге если в последние 10 000 лет основной проблемой для менеджеров данных была нехватка данных, то с лавинообразным ростом количества данных и систем управления данными пользователи и менеджеры столкнулись с проблемой - избытка данных, затрудняющим поиск юридически корректной и качественной информации.

Разрозненность силосов данных неизбежно приводит к серьёзной проблеме снижения качества данных. При наличии множества независимых систем одни и те же данные могут существовать в разных версиях, часто с противоречивыми значениями, что создаёт дополнительные сложности для пользователей, которым необходимо определить, какая информация является актуальной и достоверной.

Дублирование, и отсутствие качества данных как следствие разобщённости

Из-за проблемы силосов данных менеджеры вынуждены тратить значительное время на поиск и сверку данных. Чтобы застраховаться от проблем с качеством, компании создают сложные структуры управления информацией, в которых вертикаль менеджеров отвечает за поиск, верификацию и согласование данных. Однако этот подход лишь увеличивает бюрократию и замедляет процесс принятия решений. Чем больше данных, тем сложнее их анализировать и интерпретировать, особенно если отсутствует единый стандарт их хранения и обработки.

С появлением множества программных приложений и систем, которые растут как грибы после дождя в последнее десятилетие, проблема силосов и ненадлежащего качества данных становится все более актуальной для конечных пользователей. Одни и те же данные, но с разными значениями, теперь можно найти в разных системах и приложениях (Рис. 2.1-6). Это приводит к трудностям для конечных пользователей, когда они пытаются определить, какая версия данных является актуальной и правильной среди множества доступных. Это приводит к ошибкам при анализе и, в конечном счете, принятии решений.

Чтобы застраховаться от проблем с поиском нужных данных, руководители компаний создают многоуровневую бюрократию из менеджеров-проверщиков. Их задача — уметь быстро находить, проверять и пересылать нужные данные в виде таблиц и отчётов, ориентируясь в лабиринте разрозненных систем.

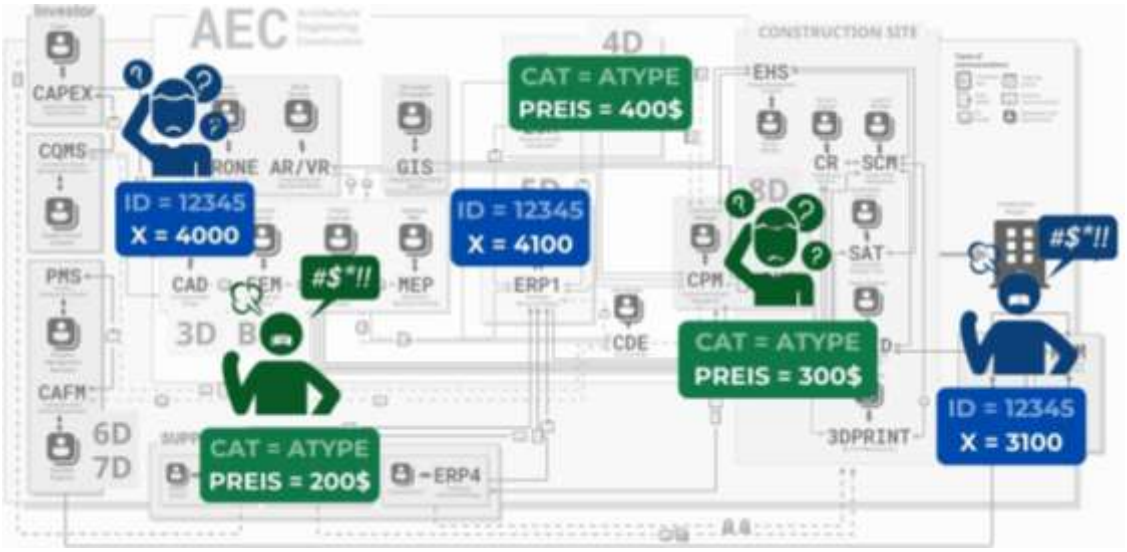


Рис. 2.1-6 Пытаясь найти нужные данные, менеджеры должны обеспечить качество и юридическую надежность данных между различными системами.

Однако на практике эта модель порождает новые сложности. Когда управление данными осуществляется вручную, а информация разбросана по множеству несвязанных между собой решений, каждая попытка получить точную и актуальную информацию через пирамиду ответственных лиц (Рис. 2.1-7) превращается в узкое место — затратное по времени и подверженное ошибкам.

Ситуацию усугубляет лавинообразный рост числа цифровых решений. Рынок программного обеспечения продолжает пополняться новыми инструментами, кажущимися многообещающими. Но

без чёткой стратегии по управлению данными эти решения не интегрируются в единую систему, а наоборот — создают дополнительные слои сложности и дублирования. В результате, вместо упрощения процессов, компании оказываются в ещё более фрагментированной и хаотичной информационной среде.

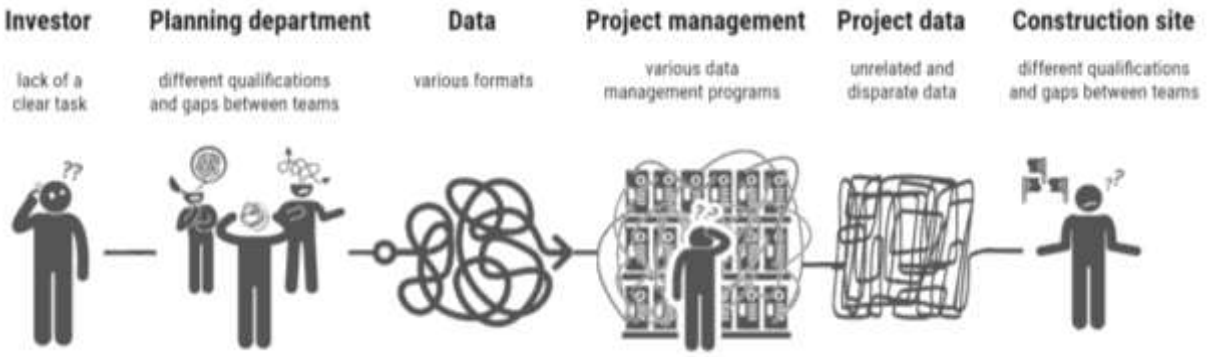


Рис. 2.1-7 Сложность систем и разнообразие форматов данных приводят к потере согласованности в процессе строительства.

Все перечисленные проблемы, связанные с управлением множеством разрозненных решений, рано или поздно подводят руководство компаний к важному осознанию: дело не в объёме данных и не в поиске очередного «универсального» инструмента для их обработки. Истинная причина кроется в качестве данных и в том, как именно организация их создаёт, получает, хранит и использует.

> Ключ к устойчивому успеху — не в погоне за новыми «волшебными» приложениями, а в формировании внутри компании культуры работы с данными. Это означает, что данные рассматриваются как стратегический актив, а вопросы их качества, целостности и актуальности становятся приоритетом на всех уровнях организации.

Решение дилеммы «качество против количества» лежит в создании единой структуры данных, которая устраняет дублирование, устраняет противоречия и объединяет информационные потоки. Такая архитектура позволяет сформировать единый, надёжный источник данных, на основе которого принимаются обоснованные, точные и своевременные решения.

В противном случае, как это до сих пор часто происходит, компании продолжают опираться на субъективные мнения и интуитивные оценки HiPPO экспертов, а не на достоверные факты. В строительной отрасли, где традиционно значительную роль играет экспертный опыт, это особенно заметно.

HiPPO или опасность мнений в принятии решений

Традиционно в строительной отрасли ключевые решения принимаются на основе опыта и субъективных оценок. Без своевременных и достоверных данных руководителям компаний приходится действовать вслепую, полагаясь на интуицию наиболее высокооплачиваемых сотрудников (HiPPO – Highest Paid Person’s Opinion), а не на объективные факты (Рис. 2.1-8).

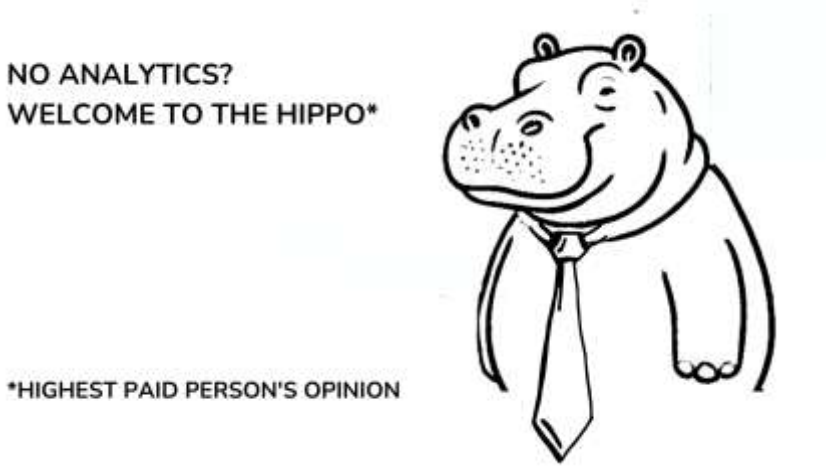


Рис. 2.1-8 В отсутствии аналитики бизнес зависит от субъективного мнения опытных специалистов.

Такой подход, возможно, оправдан в условиях стабильности и медленных изменений, но в эпоху цифровой трансформации он становится серьёзным риском. Решения, основанные на интуиции и догадках, подвержены искажениям, часто строятся на неподтверждённых гипотезах и не учитывают комплексную картину, отражённую в данных.

> То, что в компании на уровне принятий решений выдается за разумные дебаты, часто не основано ни на чем конкретном. Успех компании не должен зависеть от авторитета и уровня зарплаты экспертов, а должен определяться способностью эффективно работать с данными, выявлять закономерности и принимать взвешенные решения.

Важно отказаться от концепции, при которой авторитет или опыт автоматически означают правильность решения. Data-driven подход меняет правила игры: теперь основой для принятия решений становятся данные и аналитика, а не должность и зарплата. Большие данные, машинное обучение и визуальная аналитика дают возможность выявлять закономерности и опираться на факты, а не догадки (Рис. 1.1-4).

> *Без данных вы просто еще один человек с мнением [34].*
>
> – У. Эдвардс Деминг, учёный и консультант по менеджменту

Современные методы управления данными также обеспечивают преемственность знаний в компании. Чётко описанные процессы, автоматизация и системный подход позволяют передавать даже ключевые роли без потери эффективности.

Однако и слепое доверие к данным может привести к серьёзным ошибкам. Сами по себе данные — это лишь набор чисел. Без грамотного анализа, контекста и способности выявлять закономерности они не обладают ценностью и не могут управлять процессами. Ключ к успеху заключается

не в выборе между интуицией HiPPO и аналитикой, а в построении интеллектуальных инструментов, которые трансформируют разрозненную информацию в управляемые, обоснованные решения.

> В условиях цифрового строительства решающими факторами успеха становятся не стаж работы и место в иерархии, а скорость реакции, точность решений и эффективность использования ресурсов.

Данные — это инструмент, а не абсолютная истина. Они должны дополнять человеческое мышление, а не заменять его. Несмотря на все преимущества аналитики, данные не могут полностью вытеснить человеческую интуицию и опыт. Их роль — помогать принимать более точные и осознанные решения.

Конкурентное преимущество будет достигаться не просто соответствием стандартам, а способностью превзойти конкурентов в эффективности использования одинаковых для всех ресурсов. В будущем навык работы с данными станет таким же важным, как когда-то грамотность или владение математикой. Специалисты, умеющие анализировать и интерпретировать данные, смогут принимать более точные решения, вытесняя тех, кто полагается лишь на личный опыт (Рис. 2.1-9).

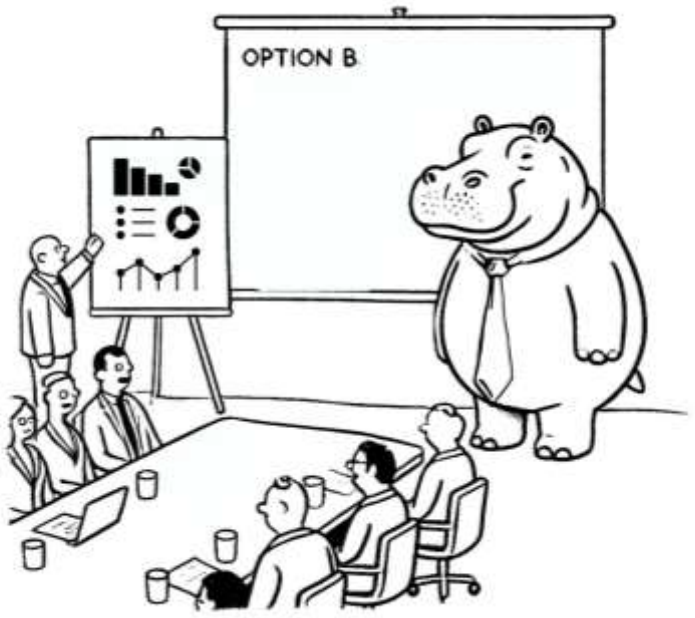


Рис. 2.1-9 Решения должны основываться на объективном анализе, а не на мнении самого высокооплачиваемого сотрудника.

Менеджеры, специалисты и инженеры будут выступать как аналитики данных, изучая структуру, динамику и ключевые показатели проектов. Человеческие ресурсы станут элементами системы, требующими гибкой настройки на основе данных для достижения максимальной эффективности.

> *Ошибки при использовании неадекватных данных гораздо меньше, чем при отсутствии данных [35].*
>
> – Чарльз Бэббидж, изобретатель первой аналитической вычислительной машины

Появление больших данных и внедрение LLM (Large Language Models) радикально изменили не только способы анализа, но и саму природу принятия решений. Если раньше в центре внимания была причинность (почему что-то произошло — диагностическая аналитика) (Рис. 1.1-4), то сегодня на первый план выходит способность прогнозировать будущее (предиктивная аналитика), а в перспективе — и прескриптивная аналитика, где машинное обучение и ИИ подсказывает наилучший выбор в процессе принятий решений.

> Согласно новому исследованию SAP™ «Новое исследование показало, что почти половина руководителей доверяют искусственному интеллекту больше, чем себе» 2025 года [36], 44% руководителей высшего звена готовы изменить своё ранее принятое решение на основе рекомендаций ИИ, а 38% доверили бы ИИ принимать бизнес-решения от их имени. При этом 74% руководителей заявили, что больше доверяют советам ИИ, чем своим друзьям и семье, а 55% работают в компаниях, где инсайты, полученные с помощью ИИ, заменяют или часто обходят традиционные методы принятия решений — особенно в организациях с годовыми доходами свыше $5 миллиардов. Кроме того, 48% опрошенных используют инструменты генеративного ИИ ежедневно, из них 15% — несколько раз в день.

С развитием LLM и автоматизированных систем управления данными встает новая проблема: как эффективно использовать информацию, не теряя её ценности в хаосе несовместимых форматов и разнородных источников, что дополняется растущей сложностью и динамикой бизнес-процессов.

Постоянное повышение сложности и динамичности бизнес-процессов

Строительная отрасль сегодня сталкивается с серьёзными вызовами в управлении данными и процессами. Основные трудности — это разрозненность информационных систем, чрезмерная бюрократия и отсутствие интеграции между цифровыми инструментами. Эти проблемы усиливаются по мере того, как сами бизнес-процессы становятся всё сложнее — под влиянием технологий, меняющихся требований клиентов и обновляющихся нормативов.

Уникальность строительных проектов обусловлена не только их техническими особенностями, но и различиями в национальных стандартах и регуляторных требованиях разных стран (Рис. 4.2-10, Рис. 5.1-7). Это требует гибкого, индивидуализированного подхода к каждому проекту, что трудно реализуемо в рамках традиционных модульных систем управления. Из-за сложности процессов и большого объема данных многие компании обращаются к вендорам, предлагающим специализированные решения. Но рынок перегружен — множество стартапов предлагают схожие продукты, фокусируясь на узких задачах. В результате часто теряется целостный подход к управлению данными.

Адаптация к непрерывному потоку новых технологий и требований рынка становится критическим фактором конкурентоспособности. Однако существующие проприетарные приложения и модульные системы обладают низкой адаптивностью — любые изменения требуют часто длительных и дорогостоящих доработок со стороны разработчиков, которые не всегда понимают специфику строительных процессов.

Компании оказываются заложниками технологического отставания, ожидая новых обновлений вместо оперативного внедрения инновационных интегрированных подходов. В итоге внутренняя структура строительных организаций часто представляет собой сложную экосистему взаимосвязанных иерархических, и часто закрытых, систем, координация между которыми осуществляется через многоуровневую сеть менеджеров (Рис. 2.1-10).

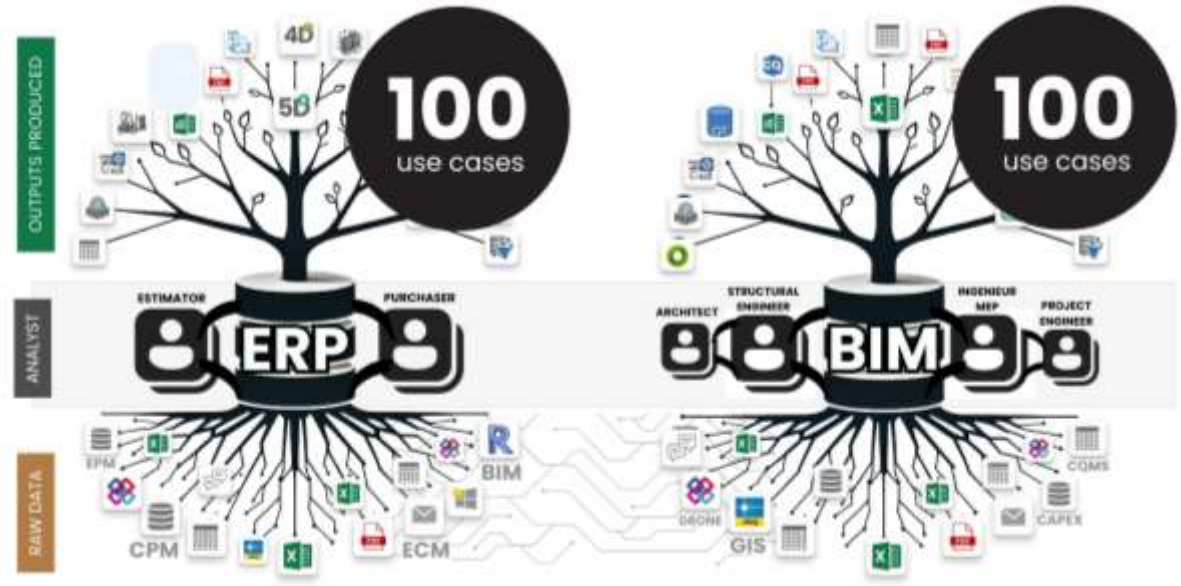


Рис. 2.1-10 Компании состоят из взаимосвязанных систем, чье объединение формирует процессы, требующие автоматизации.

Согласно исследованию, проведённому Канадской строительной ассоциацией и компанией KPMG в Канаде в 2021 году [37], лишь 25% компаний считают, что находятся в значительном или отличном положении по сравнению с конкурентами в области внедрения технологий или цифровых решений. Только 23% респондентов сообщили, что их решения в значительной или большой степени основаны на данных. При этом большинство участников опроса охарактеризовали своё использование ряда других технологий как чисто экспериментальное либо признались, что вовсе не применяют их.

Это нежелание принимать участие в технологических экспериментах особенно ярко проявляется в крупных инфраструктурных проектах, где ошибки могут стоить миллионы долларов. Даже самые передовые технологии — цифровые двойники, предиктивная аналитика — часто встречают сопротивление не из-за своей эффективности, а из-за отсутствия доказанной надежности в реальных проектах.

> Согласно отчету Всемирного экономического форума (WEF) "Формирование будущего строительства" [5], внедрение новых технологий в строительстве наталкивается не только на технические сложности, но и на психологический барьер со стороны заказчиков. Многие клиенты опасаются, что использование передовых решений сделает их проекты экспериментальной площадкой и сделает их «подопытными кроликами», а непредсказуемые последствия могут привести к дополнительным затратам и рискам.

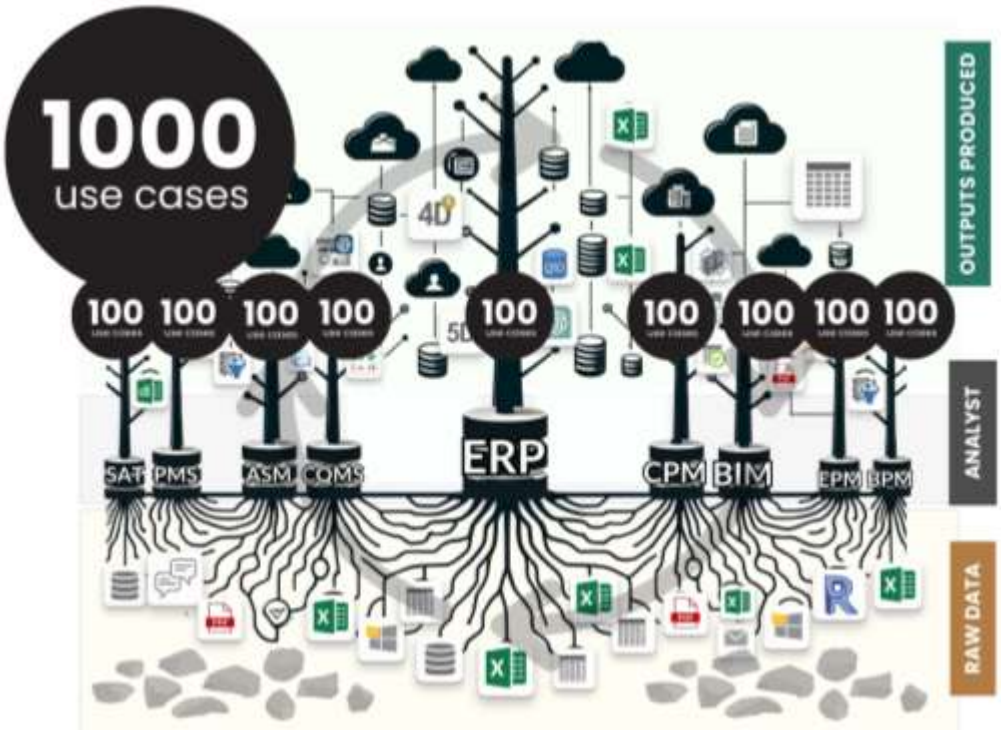


Рис. 2.1-11 Для каждого случая использования данных рынок решений предлагает приложения для оптимизации и автоматизации процессов.

> Строительная отрасль очень разнообразна: у разных проектов разные требования, региональные особенности, законодательные нормы классификаций (Рис. 4.2-10), стандарты калькуляций (Рис. 5.1-7) и т. д. Поэтому практически невозможно создать проприетарное универсальное приложение или систему, которая бы идеально подходила под все эти требования и особенности проектов.

Пытаясь справиться с нарастающей сложностью систем и зависимостью от поставщиков ПО, все чаще приходят к осознанию, что ключ к эффективному управлению данными — это не только открытость и стандартизация, но и упрощение самой архитектуры процессов. Растущая сложность и динамичность бизнес-процессов требует новых подходов, где приоритет смещается от накопления данных к их структурированию и упорядочиванию. Именно этот сдвиг станет следующим шагом в развитии строительной отрасли, знаменуя конец эпохи доминирования поставщиков ПО и начало эры осмысленной организации информации.

Осознание ограниченности универсальных решений и уязвимости перед ростом сложности приводит к смене приоритетов: от закрытых платформ и накопления данных — к прозрачности, адаптивности и структурированной работе с информацией. Этот сдвиг в мышлении отражает более широкие изменения в глобальной экономике и технологии, которые описываются через призму так называемых «промышленных революций». Чтобы понять, куда движется строительство и каковы его будущие ориентиры, необходимо рассмотреть место отрасли в контексте Четвёртой и Пятой промышленных революций — от автоматизации и цифровизации к персонализации, открытым стандартам и сервисной модели данных.

Четвертая промышленная революция (Индустрия 4.0) и пятая промышленная революция (Индустрия 5.0) в строительстве

Технологические и экономические уклады — это теоретические концепции, используемые для описания и анализа эволюции общества и экономики на различных этапах развития. При этом разные исследователи и эксперты могут интерпретировать их по-разному.

- **Четвертая промышленная революция** (4IR или Industry 4.0) связана с информационными технологиями, автоматизацией, цифровизацией и глобализацией. Одним из её ключевых элементов является создание проприетарных программных решений, то есть специализированных цифровых продуктов, разработанных под конкретные задачи и компании. Эти решения часто становятся важной частью ИТ-инфраструктуры, но при этом слабо масштабируются без дополнительных модификаций.
- **Пятая промышленная революция (5IR)** сегодня находится на более ранней стадии концептуализации и развития, чем 4IR. Ее основные принципы включают повышение степени персонализации продуктов и услуг. 5IR — это движение к более адаптируемой, гибкой и персонализированной экономической деятельности с акцентом на персонализацию, консалтинг и сервис-ориентированные модели. Ключевым аспектом пятого экономического уклада является использование данных для принятия решений, что практически невозможно без применения открытых данных и открытых инструментов (Рис. 2.1-12).

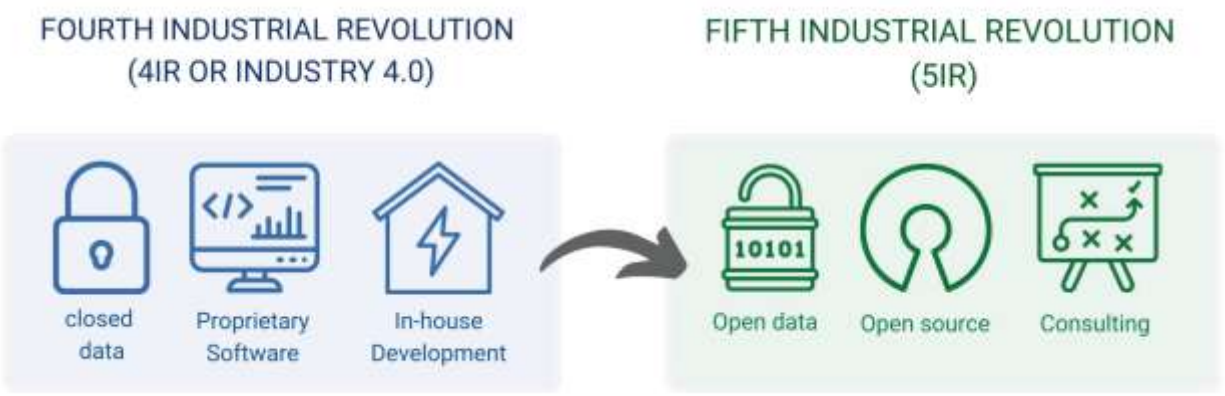


Рис. 2.1-12 Четвёртый уклад фокусируется на решениях, а пятый – на персонализации и данных.

> Создание приложения для компаний строительной отрасли, предназначенного для использования в десяти или ста организациях, не гарантирует его успешного масштабирования в других компаниях, регионах или странах без значительных модификаций и доработок. Вероятность успешного расширения таких решений остается низкой, поскольку каждая организация имеет уникальные процессы, требования и условия, которые могут потребовать персонализированных адаптаций.

Важно понимать, что уже сегодня успешная интеграция технологических решений подразумевает глубоко персонализированный подход к каждому процессу, проекту и компании. Это означает, что даже после разработки универсальной структуры, инструмента или программы потребуется ее детальная адаптация и настройка под уникальные требования и условия каждой конкретной компании и конкретного проекта.

> Согласно отчету PwC «Расшифровка пятой промышленной революции» [38], около 50% руководителей высшего звена в различных отраслях в этом году делают ставку на интеграцию передовых технологий и человеческого опыта. Такой подход позволяет оперативно адаптироваться к изменениям в дизайне продукта или требованиям заказчика, создавая персонализированное производство.

Для каждого процесса требуется разработка уникальной функции или приложения, что, с учётом масштабов мировой строительной отрасли и разнообразия проектов, приводит к существованию огромного количества бизнес-кейсов, представляющих собой каждый раз уникальную Pipeline логику (Рис. 2.1-13). Каждый такой кейс имеет свои особенности и требует индивидуального подхода. Мы более подробно рассмотрим разнообразие возможных решений одной и той же аналитической задачи в контексте различных подходов в главе, посвящённой машинному обучению и разбору датасета «Титаник» (Рис. 9.2-9).

> **Pipeline** в контексте цифровых процессов – это последовательность действий, процессов и инструментов, которые обеспечивают автоматизированный или структурированный поток данных и работ на разных этапах жизненного цикла проекта.

Рис. 2.1-13 Индивидуальность и вариативность бизнес-кейсов делает попытки создания масштабируемых закрытых платформ и инструментов невозможными.

Наша жизнь уже во многом изменилась под влиянием цифровой трансформации, и сегодня можно говорить о наступлении нового этапа в экономическом развитии строительной отрасли. В этой «новой экономике» конкуренция будет устроена по другим правилам: тот, кто способен эффективно превращать публичные знания и открытые данные в востребованные продукты и услуги, получает ключевое преимущество в условиях пятой индустриальной революции.

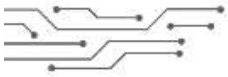
> Как отмечает экономист Кейт Маскус в книге «Частные права и общественные проблемы: Глобальная экономика интеллектуальной собственности в XXI веке» 2012 года [39], «мы живём в глобальной экономике знаний, и будущее принадлежит тем, кто умеет превращать научные открытия в товар».

Переход к пятому экономическому укладу подразумевает сдвиг фокуса с закрытых IT-решений на открытые стандарты и платформы. Компании начнут отходить от традиционных программных продуктов в пользу сервис-ориентированных моделей, в которых основным активом станут данные, а не проприетарные технологии.

> Исследование Гарвардской бизнес-школы 2024 года [40] показывает огромную экономическую ценность программного обеспечения с открытым исходным кодом (Open Source Software, OSS). Согласно исследованию, OSS присутствует в 96% всех программных кодов, а некоторое коммерческое ПО на 99,9% состоит из OSS-компонентов. Без OSS компании тратили бы в 3,5 раза больше на программное обеспечение.

> Экосистемы строительных компаний, следуя общемировым тенденциям, постепенно перейдут к пятой экономической парадигме, где дата-центричные аналитические и консалтинговые услуги станут приоритетнее, чем изолированные закрытые решения с жёстко заданными сценариями использования.

Эпоха цифровизации изменит баланс сил в отрасли: вместо зависимости от решений вендоров компании будут строить свою конкурентоспособность на способности эффективно использовать данные. В результате строительная индустрия перейдёт от устаревших жёстких систем к гибким, адаптивным экосистемам, где открытые стандарты и совместимые инструменты станут основой управления проектами. Конец эпохи доминирования вендоров приложений создаст новые условия, в которых ценность будет определяться не владением закрытым кодом и специальными коннекторами, а умением превращать данные в стратегическое преимущество.



ГЛАВА 2.2.
ПРЕВРАЩЕНИЕ ХАОСА В ПОРЯДОК И СНИЖЕНИЕ СЛОЖНОСТИ

Лишний код и закрытые системы как барьер в повышении производительности

На протяжении последних десятилетий технологические изменения в ИТ-сфере определялись преимущественно поставщиками программного обеспечения. Именно они задавали вектор развития, определяя, какие технологии компаниям следует внедрять, а какие — оставить за бортом. В эпоху перехода от разрозненных решений к централизованным базам данных и интегрированным системам вендоры продвигали лицензируемые продукты, обеспечивая контроль над доступом и масштабированием. Позже, с приходом облачных технологий и моделей Software as a Service (SaaS), этот контроль трансформировался в модель подписки, закрепляя пользователей в роли постоянных клиентов цифровых сервисов.

Такой подход породил парадокс: несмотря на беспрецедентные объёмы созданного программного кода, реально используется лишь малая его часть. Возможно, кода написано в сотни или тысячи раз больше, чем необходимо, поскольку одни и те же бизнес-процессы описываются и дублируются в десятках или сотнях программ по-разному — даже внутри одной компании. При этом за разработку уже заплачено, и эти затраты невозвратны. Тем не менее, индустрия продолжает воспроизводить этот цикл, создавая новые продукты с минимальной добавленной ценностью для конечного пользователя, чаще под давлением рыночных ожиданий, чем реальных потребностей.

> Согласно Руководству по оценке стоимости разработки программного обеспечения, составленному Университетом оборонных закупок (DAU) [41], стоимость разработки программного обеспечения может значительно варьироваться в зависимости от нескольких факторов, включая сложность системы и выбранные технологии. Исторически сложилось так, что стоимость разработки на 2008 год составляет около $100 за строку исходного кода (SLOC), в то время как стоимость обслуживания может вырасти до $4 000 за SLOC.

Только один из компонентов CAD-приложений — геометрическое ядро — может насчитывать десятки миллионов строк кода (Рис. 6.1-5). Аналогичная ситуация наблюдается и в ERP-системах (Рис. 5.4-4), к обсуждению сложности которых мы ещё вернёмся в пятой части книги. Однако при ближайшем рассмотрении становится ясно: значительная часть этого кода не создаёт добавленной стоимости, а лишь выполняет функцию «почтальона» — механически перемещая данные между базой данных, API, пользовательским интерфейсом и другими таблицами системы. Несмотря на популярный миф о критической важности так называемой бизнес-логики, суровая реальность куда более прозаична: современные кодовые базы переполнены устаревшими шаблонными блоками (легаси кодом), единственная цель которых — обеспечить передачу данных между таблицами и компонентами, не влияя на принятие решений или рост эффективности бизнеса.

В итоге закрытые решения, которые занимаются обработкой данных из различных источников, неизбежно превращаются в запутанные "спагетти-экосистемы". С этими сложными, переплетенными системами способна разбираться только целая армия менеджеров, работающая в полу руч-

ном режиме. Такая организация управления данными не только неэффективна с точки зрения ресурсов, но и создает критические точки уязвимости в бизнес-процессах, делая компанию зависимой от узкого круга специалистов, понимающих, как функционирует этот технологический лабиринт.

Постоянное увеличение объема кода, количества приложений и усложнение концепций, предлагаемых вендорами, привело к закономерному результату — росту сложности ИТ-экосистемы в строительстве. Это сделало практическую реализацию цифровизации через увеличения количества приложений в отрасли малоэффективной. Программные продукты, создаваемые без должного внимания к потребностям пользователей, часто требуют значительных ресурсов на внедрение и поддержку, но не приносят ожидаемой отдачи.

> Согласно исследованию McKinsey «Повышение производительности строительства» [42], за последние два десятилетия глобальный рост производительности труда в строительстве составлял в среднем всего 1% в год, по сравнению с ростом на 2,8% в целом по мировой экономике и на 3,6% в обрабатывающей промышленности. В Соединенных Штатах производительность труда в строительстве в расчете на одного работника снизилась вдвое с 1960-х годов [43].

Рост сложности систем, изолированность и закрытость данных ухудшили коммуникацию между специалистами, что сделало строительную отрасль одной из наименее эффективных (Рис. 2.2-1). до 22 трлн долларов к 2040 году, что потребует значительного повышения эффективности.

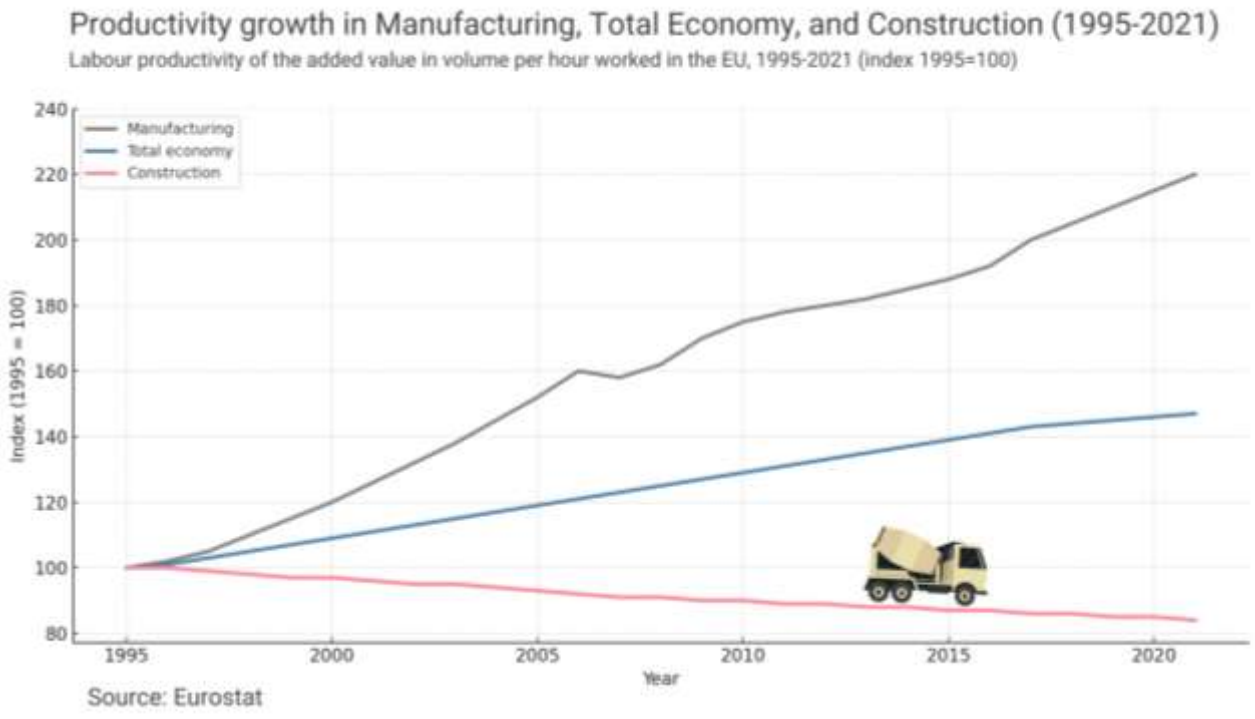


Рис. 2.2-1 Закрытость и сложность данных и как следствие плохая коммуникация между специалистами привели строительную отрасль к одному из наименее эффективных секторов экономики (по материалам [44], [45]).

> Как подчёркивается в исследовании McKinsey (2024) «Обеспечение продуктивности строительства больше не является чем-то необязательным», в условиях усиливающегося дефицита ресурсов и стремления отрасли удвоить темпы роста, строительство больше не может позволить себе оставаться на текущем уровне производительности [44]. По прогнозам, глобальные затраты на строительство вырастут с 13 трлн долларов в 2023 году до значительно более высоких значений к концу десятилетия, что делает вопрос эффективности не просто актуальным, а критически важным.

Одним из ключевых способов повышения эффективности станет неизбежная унификация и упрощение структур приложений и архитектур экосистем, обрабатывающих данные. Такой подход к рационализации позволит избавиться от избыточных слоев абстракции и ненужной сложности, которые накапливались годами в корпоративных системах.

От силосов к единому хранилищу данных

Чем больше данных накапливает организация, тем сложнее становится извлечь из них реальную пользу. Из-за фрагментированности хранения информации в изолированных силосах современные компании в своих бизнес-процессах напоминают строителей, пытающихся возвести небоскрёб из материалов, хранящихся на тысячах разных складов. Избыток информации не только затрудняет доступ к юридически значимым сведениям, но и замедляет принятие решений: каждый шаг приходится многократно проверять и подтверждать.

Каждая задача или процесс оказывается жёстко привязанным к отдельной таблице или базе данных, а обмен данными между системами требует сложных интеграций. Ошибки и несовпадения в одной системе могут вызвать цепные сбои в других. Некорректные значения, запоздалые обновления и дублирование информации вынуждают сотрудников тратить значительное время на ручную сверку и согласование данных. В итоге организация больше времени тратит на устранение последствий фрагментации, чем на развитие и оптимизацию процессов.

Эта проблема универсальна: одни компании продолжают бороться с хаосом, другие находят решение в интеграции — переводе информационных потоков в централизованную систему хранения. Представьте это как одну большую таблицу, в которой можно хранить любые сущности, связанные с задачами, проектами и объектами. Вместо десятков разрозненных таблиц и форматов появляется единое связное хранилище (Рис. 2.2-2), что позволяет:

- минимизировать потери данных;
- устранить необходимость в постоянном согласовании информации;
- улучшить доступность и качество данных;
- упростить аналитическую обработку и машинное обучение

Приведение данных к единому стандарту означает, что независимо от источника, информация преобразуется в унифицированный и машиночитаемый формат. Такая организация данных позволяет проверять их целостность, анализировать в реальном времени и оперативно использовать для принятия управленческих решений.

Подробнее о концепции интегрированных систем хранения и их применении в аналитике и машинном обучении мы поговорим в главе «Хранение больших данных и машинное обучение». Темы моделирования и структурирования данных будут подробно раскрыты в главах «Преобразование данных в структурированную форму» и «Как стандарты меняют игру: от случайных файлов к продуманной модели данных».

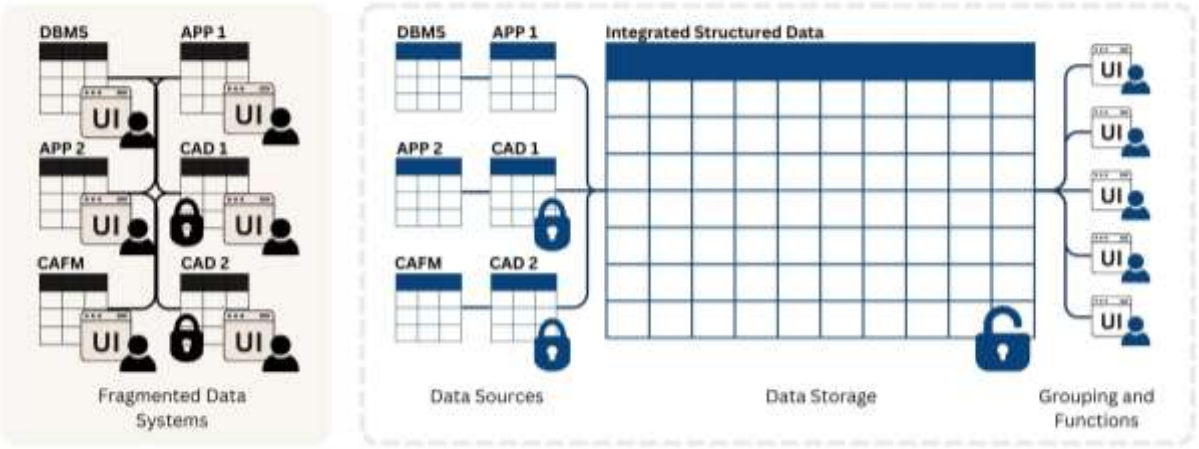


Рис. 2.2-2 Интеграция данных устраняет изолированность, улучшает доступность информации и оптимизирует бизнес-процессы.

После структурирования и объединения данных следующим логическим шагом становится их проверка. При наличии единого интегрированного хранилища этот процесс значительно упрощается: больше нет множества несогласованных схем, дублирующих структур и сложных взаимосвязей между таблицами. Вся информация приведена к единой модели данных, что устраняет внутренние противоречия и ускоряет процесс валидации. Проверка и обеспечение качества данных — это краеугольные аспекты всех бизнес-процессов, и мы подробнее рассмотрим их в соответствующих главах книги.

На завершающем этапе данные группируются, фильтруются и анализируются. К ним применяются различные функции: агрегирование (сложение, умножение), вычисления между таблицами, колонками или строками (Рис. 2.2-4). Работа с данными превращается в последовательность шагов: сбор, структурирование, проверка, трансформация, аналитическая обработка и выгрузка в конечные приложения, где информация используется для решения практических задач. Подробнее о выстраивании таких сценариев, автоматизации шагов и построении потоков обработки мы поговорим в главах, посвящённых ETL-процессам и data pipeline-подходу.

Таким образом, цифровая трансформация — это не просто упрощение работы с информацией. Это отказ от избыточной сложности в управлении данными, переход от хаоса к предсказуемости, от множества систем — к управляемому процессу. Чем ниже сложность архитектуры, тем меньше кода требуется для поддержки. А в перспективе — код как таковой может вообще исчезнуть, уступив место интеллектуальным агентам, которые самостоятельно анализируют, систематизируют и преобразуют данные.

Интегрированные системы хранения позволяют перейти к использованию AI агентов

Чем меньше сложность данных и систем – тем меньше кода нужно писать и поддерживать. А самый простой способ сэкономить разработку – вообще избавиться от кода, заменив его данными. Когда разработка кода приложения переходит от кода к моделям данных, неизбежно возникает сдвиг в сторону дата-центричного (data-driven) подхода, потому что за этими концепциями стоит совершенно другой образ мышления.

> Когда человек выбирает путь работы с данными в центре, он начинает иначе видеть их роль. Данные перестают быть просто «сырьем» для приложений – теперь это основа, вокруг которой строится архитектура, логика и взаимодействие.

Традиционный же подход к управлению данными начинается обычно на уровне приложений и в строительстве напоминает громоздкую бюрократическую систему: многоуровневые согласования, ручные проверки, бесконечные версии документов через соответствующие программные продукты. С развитием цифровых технологий все больше компаний будет вынуждено перейти к принципу минимализма — хранить и использовать только то, что действительно необходимо и будет использоваться.

Логику минимизации подхватили вендоры. Чтобы упростить процессы хранения и обработки данных работа пользователей переносится из функционала оффлайн приложений и инструментов в облачные сервисы и так называемые SaaS решения.

> Концепция SaaS (Software as a Service, или "программное обеспечение как услуга") является одним из ключевых направлений в современных IT-инфраструктурах, позволяя пользователям получать доступ к приложениям через интернет без необходимости установки и обслуживания программного обеспечения на собственных компьютерах.

С одной стороны SaaS облегчил масштабирование, управление версиями и снизил затраты на поддержку и обслуживание, но с другой стороны кроме зависимости от логики конкретного приложения также сделал пользователя полностью зависимым от облачной инфраструктуры провайдера. Если сервис выходит из строя, доступ к данным и бизнес-процессам может быть временно или даже надолго заблокирован. Кроме того, все данные пользователя при работе с SaaS приложениями хранятся на серверах провайдера, что создает риски в плане безопасности и соответствия регуляторным требованиям. Изменение тарифов или условий использования также может привести к росту расходов или необходимости срочной миграции.

Развитие AI, LLM-агентов и дата-центричного подхода поставило под вопрос будущее приложений в их традиционном виде и SaaS исполнении. Если раньше приложения и сервисы были необходимы для управления бизнес-логикой и обработки данных, то с приходом AI-агентов эти функции могут перейти к интеллектуальным системам, работающим напрямую с данными.

Именно поэтому все чаще в IT отделах и на уровне менеджмента обсуждаются гибридные архитектуры, где AI-агенты и локальные решения дополняют облачные сервисы, снижая зависимость от SaaS-платформ.

> *Подход, которого мы придерживаемся, признает, что традиционные бизнес-приложения или SaaS-приложения могут кардинально измениться в эпоху агентов. Эти приложения, по сути, представляют собой CRUD [создание, чтение, обновление и удаление] базы данных с бизнес-логикой. Но в будущем эта логика перейдет к агентам ИИ [46].*
>
> — Сатья Наделла, генеральный директор Microsoft, 2024 г.

Дата-центричный подход и использование ИИ/LLM агентов позволяет сократить количество избыточных процессов, а значит, уменьшить нагрузку на сотрудников. Когда данные организованы правильно, их становится проще анализировать, визуализировать и применять для принятия решений. Вместо бесконечных отчетов и проверок специалисты получают доступ к актуальной информации в несколько кликов или при помощи LLM агентов автоматически в виде готовых документов и дэшбордов.

В работе с данными нам будут помогать инструменты искусственного интеллекта (ИИ) и LLM чаты. В последние годы наблюдается тенденция перехода от традиционных операций CRUD (создание, чтение, обновление, удаление) к использованию больших языковых моделей (LLM) для управления данными. LLM способны интерпретировать естественный язык и автоматически генерировать соответствующие запросы к базе данных, что упрощает взаимодействие с системами управления данными (Рис. 2.2-3).

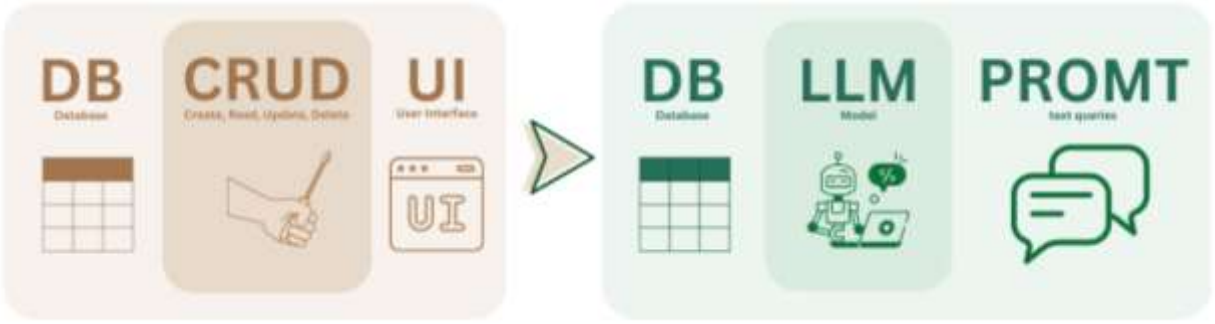


Рис. 2.2-3 ИИ будет заменять решения по работе с хранилищами и базами данных и по их интеграции, постепенно вытесняя традиционные приложения и CRUD-операции.

> *В ближайшие 3–6 месяцев ИИ будет писать 90% кода, а через 12 месяцев почти весь код может быть сгенерирован ИИ [47].*
>
> — Дарио Амодей, CEO компании LLM Anthropic, март 2025 г.

Несмотря на стремительное развитие инструментов AI-разработки (например, GitHub Copilot), в 2025 году разработчики по-прежнему играют ключевую роль в этом процессе. AI-агенты становятся всё более полезными помощниками: они автоматически интерпретируют пользовательские

запросы, генерируют SQL- и Pandas-запросы (подробнее об этом — в следующих главах) или пишут код для анализа данных. Таким образом, искусственный интеллект постепенно заменяет традиционные пользовательские интерфейсы приложений.

Распространение моделей искусственного интеллекта, таких как языковые модели, будет стимулировать развитие гибридных архитектур. Вместо полного отказа от облачных решений и SaaS продуктов, мы можем увидеть интеграцию облачных сервисов с локальными системами управления данными. Например, федеративное обучение (federated learning) позволяет использовать мощные модели ИИ без необходимости перемещения чувствительных данных в облако. Таким образом, компании смогут сохранять контроль над своими данными, одновременно получая доступ к передовым технологиям.

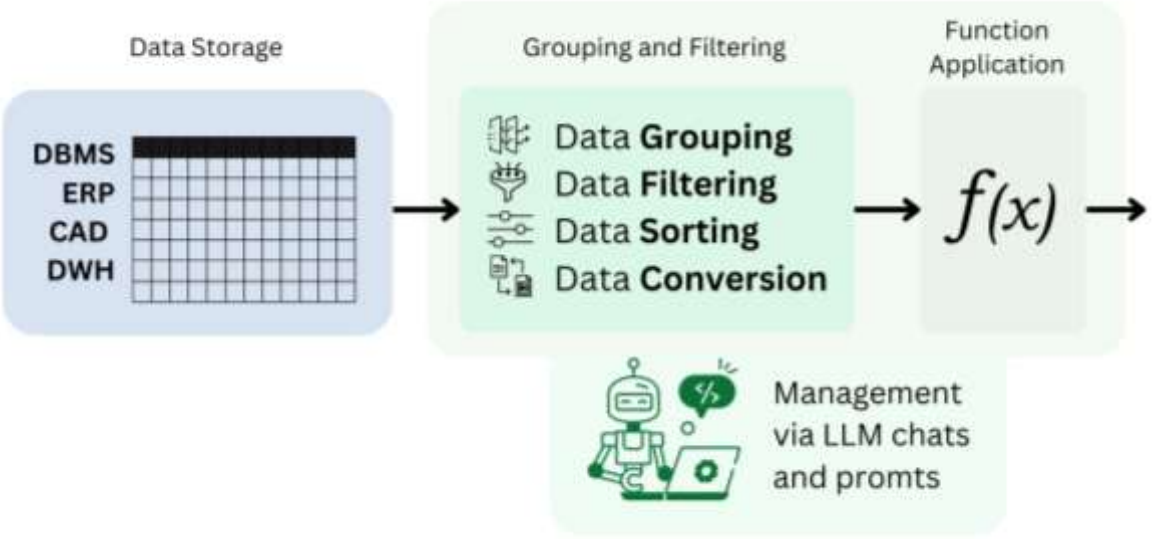


Рис. 2.2-4 Основными операциями группировки, фильтрации и сортировки с последующим применением функций будут заниматься LLM чаты.

Будущее строительной отрасли будет основано на сочетании локальных решений, облачных мощностей и интеллектуальных моделей, работающих вместе для создания эффективных и безопасных систем управления данными. LLM позволит пользователям без глубоких технических знаний взаимодействовать с базами и хранилищами данных, формулируя свои запросы на естественном языке. Подробнее об LLM и AI агентах и том, как они работают мы поговорим в главе „LLM агенты и структурированные форматы данных“.

Правильно организованные данные и простые, удобные инструменты аналитики с поддержкой LLM не только облегчат работу с информацией, но и помогут минимизировать ошибки, повысить эффективность и автоматизировать процессы.

От сбора данных к принятию решений: путь к автоматизации

В последующих частях книги мы подробно рассмотрим, как специалисты взаимодействуют между собой и как данные становятся основой для принятия решений, автоматизации и повышения эффективности работы. Рис. 2.2-5 представляет пример схемы, показывающая последовательность

этапов обработки данных в дата-центричном подходе. Эта схема иллюстрирует контур непрерывного улучшения процессов (Continuous Improvement Pipeline), части которой будут детально рассмотрены далее в книге.

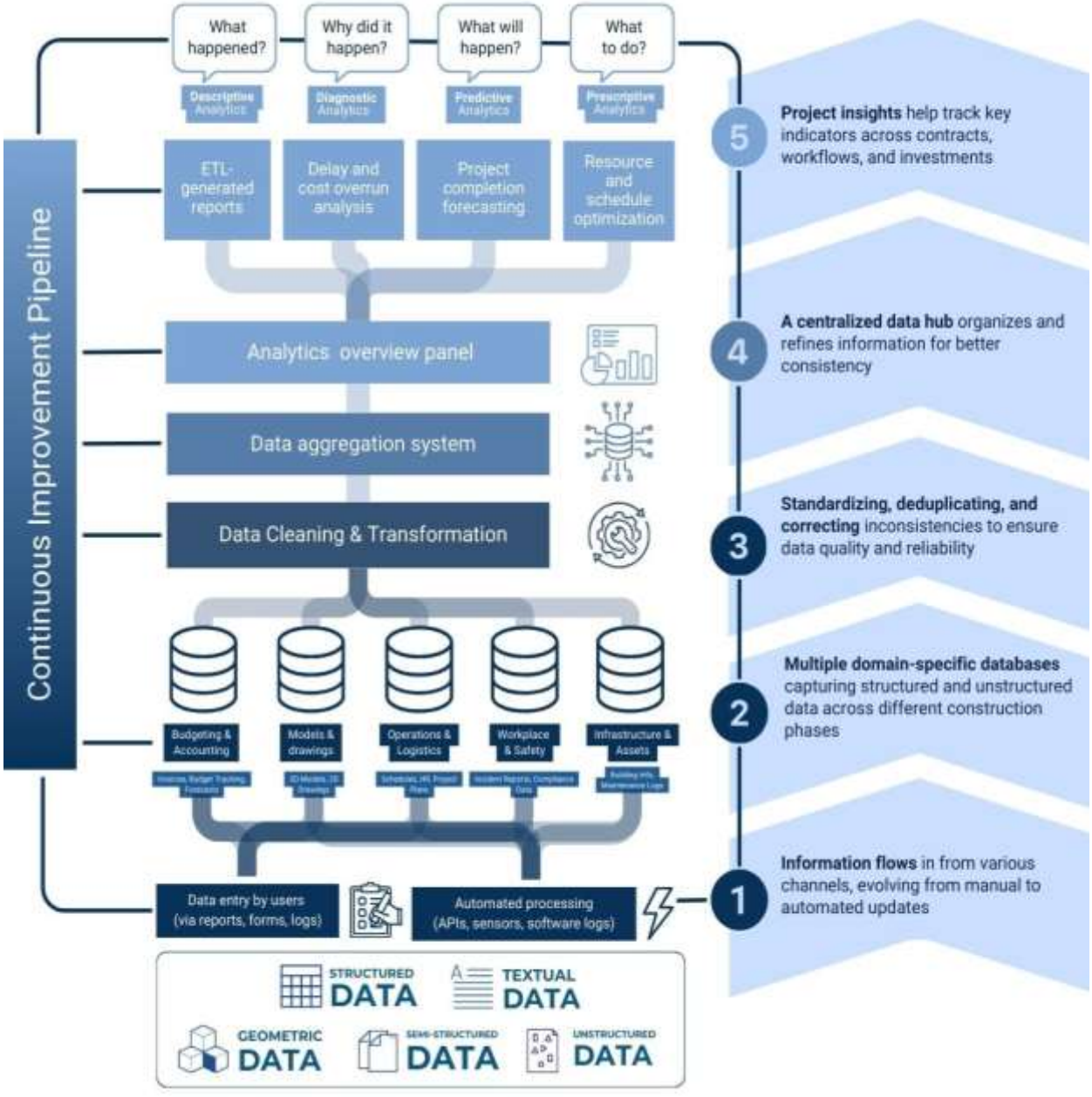


Рис. 2.2-5 Пример конвейера непрерывного улучшения данных: потока обработки и анализа данных в строительных проектах.

Система, описывающая бизнес-процессы компании среднего размера, строится на многоуровневом принципе. В неё входят: сбор данных, очистка, агрегация, аналитическая обработка и принятие решений на основе полученных результатов. Все эти этапы мы будем изучать далее в книге — как

в теоретическом контексте, так и через практические примеры:

- На первом уровне происходит **ввод данных (**Рис. 3.1-1**)**. Информация поступает как в ручном режиме (через отчёты, формы, логи), так и в автоматизированном виде (с API, датчиков, программных систем). Данные могут быть разной структуры: геометрические, текстовые, неструктурированные. На этом этапе возникает необходимость в стандартизации, структуризации и унификации потоков информации.
- Следующий уровень — **обработка и трансформация данных**. Включает в себя процессы очистки, удаления дубликатов, исправления ошибок и подготовки информации для дальнейшего анализа (Рис. 4.2-5). Этот этап критически важен, так как качество аналитики напрямую зависит от чистоты и точности данных.
- Далее **данные попадают в специализированные таблицы, датафреймы или базы**, разделённые по функциональным направлениям: бюджетирование и учёт, модели и чертежи, логистика, безопасность и инфраструктура. Такое разбиение позволяет организовать удобный доступ и обеспечивает возможность кросс-анализа информации.
- После этого данные **агрегируются и отображаются в аналитической панели** (витрине). Здесь применяются методы описательной, диагностической, предсказательной и предписывающей аналитики. Это позволяет ответить на ключевые вопросы (Рис. 1.1-4): что произошло, почему это произошло, что произойдёт в будущем и какие действия нужно предпринять. Например, система может выявить задержки, спрогнозировать завершение проектов или оптимизировать ресурсы.
- Наконец, на последнем уровне формируются **аналитические выводы и ключевые показатели**, которые помогают отслеживать выполнение контрактов, управлять инвестициями и улучшать бизнес-процессы (Рис. 7.4-2). Эта информация становится основой для принятия решений и стратегии развития компании.

Подобным образом, данные проходят путь от сбора до использования в стратегическом управлении. В следующих частях книги мы рассмотрим каждый этап подробно, уделяя особое внимание типам данных, методам обработки данных, инструментам аналитики и реальным кейсам использования этих подходов в строительной отрасли.

Дальнейшие шаги: превращение хаоса в управляемую систему

В этой части мы исследовали проблемы информационных силосов и рассмотрели влияние избыточной сложности систем на эффективность бизнеса, проанализировали переход от четвертой промышленной революции к пятой, где центральную роль играют данные, а не приложения. Мы увидели, как разрозненные информационные системы создают барьеры для обмена знаниями, а продолжающееся усложнение ИТ-ландшафта снижает производительность и тормозит инновации в строительной отрасли.

Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные подходы в ваших повседневных задачах:

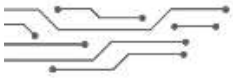
- Визуализируйте свой информационный ландшафт
    - ☐ Создайте визуальную карту источников данных (Miro, Figma, Canva), с которыми вы регулярно работаете
    - ☐ Добавьте к этой карте системы и приложения, используемых вами в работе

- Выявите потенциально дублирующиеся функциональности и избыточные решения
- Идентифицируйте критические точки, где может происходить потеря или искажение данных при передаче между системами

- Внедряйте персональные практики управления данными
  - Сместите фокус с приложений на данные как ключевой актив в процессах
  - Документируйте источники данных и методологию обработки для обеспечения прозрачности
  - Разработайте механизмы оценки и повышения качества данных
  - Стремитесь к тому, чтобы данные вводились один раз, а использовались многократно — это основа эффективной организации процессов
- Продвигайте дата-центричный (data-driven) подход в своей команде
  - Предлагайте использовать стандартизированные и унифицированные форматы обмена данными между коллегами
  - Регулярно поднимайте вопросы, связанные с качеством и доступностью данных, на командных встречах
  - Познакомьтесь с Open Source альтернативами инструментов, которые вы используете в решении своих вопросов

Начните с малого — выберите один конкретный процесс или набор данных, который критически важен для вашей работы, и примените к нему дата-центричный подход, сместив акцент с инструментов на данные. Добившись успеха в одной пилотной задаче, вы получите не только практический опыт, но и наглядную демонстрацию преимуществ новой методологии для своей команды. В выполнении большинства этих шагов при возникновении вопросов вы можете обратиться за разъяснениями и помощью к любой современной LLM.

В следующих частях книги мы перейдем к более детальному рассмотрению методов структурирования и унификации данных и изучим практические подходы к интеграции разнородной информации. Особое внимание будет уделено переходу от разрозненных хранилищ к единым экосистемам данных, играющим ключевую роль в цифровой трансформации строительной отрасли.



III ЧАСТЬ
КАРКАС ДАННЫХ В СТРОИТЕЛЬНЫХ БИЗНЕС-ПРОЦЕССАХ

В третьей части формируется комплексное представление о типологии данных в строительстве и методах их эффективной организации. Анализируются характеристики и специфика работы со структурированными, неструктурированными, полуструктурированными, текстовыми и геометрическими данными в контексте строительных проектов. Рассматриваются современные форматы хранения и протоколы обмена информацией между различными системами, используемыми в отрасли. Описываются практические инструменты и методы преобразования разноформатных данных в единую структурированную среду, включая способы интеграции CAD (BIM) данных. Предлагаются подходы к обеспечению качества данных через стандартизацию и валидацию, критически важные для точности строительных расчетов. Детально анализируются практические аспекты использования современных технологий (Python Pandas, LLM-модели) с примерами кода для решения типичных задач в строительной индустрии. Обосновывается ценность создания центра компетенций (CoE) как организационной структуры для координации и стандартизации подходов к управлению информацией.

ГЛАВА 3.1.
ТИПЫ ДАННЫХ В СТРОИТЕЛЬСТВЕ

Наиболее важные типы данных в строительной отрасли

В современной строительной отрасли системы, приложения и хранилища данных компаний активно наполняются информацией и данными различных типов и форматов (Рис. 3.1-1). Подробнее рассмотрим основные типы данных, которые формируют информационный ландшафт современной компании, работающей в строительной отрасли:

- **Структурированные данные**: эти данные имеют четкую организационную структуру, например, электронные таблицы Excel и реляционные базы данных.
- **Неструктурированные данные**: это информация, которая не организована в соответствии со строгими правилами. Примерами таких данных являются текст, видео, фотографии и аудиозаписи.
- **Слабоструктурированные данные**: эти данные занимают промежуточное положение между структурированными и неструктурированными данными. Они содержат элементы структуры, но эта структура не всегда ясна или часто описана через разные схемы. Примерами полуструктурированных данных в строительстве являются: технические спецификации, проектная документация или отчеты о проделанной работе.
- **Текстовые данные**: включают в себя все, что получено в результате устных и письменных коммуникаций, например электронные письма, стенограммы совещаний и встреч.
- **Геометрические данные**: эти данные поступают из программ CAD в которых специалисты создают геометрические данные элементов проекта для визуализации, подтверждения значений объемов или проверки коллизий.

Важно отметить, что геометрические и текстовые (буквенно-цифровые) данные не являются отдельной категорией, а могут присутствовать во всех трех типах данных. Геометрические данные, например, могут быть как частью структурированных данных (параметрические CAD форматы), так и неструктурированных данных (отсканированные чертежи). Текстовые данные аналогично могут быть как организованы в базы данных (структурированные данные), так и существовать в виде документов без четкой структуры.

Каждый тип данных в строительной компании — это уникальный элемент в мозаике информационных активов компании. От неструктурированных данных, таких как изображения со строительных площадок и аудиозаписи совещаний, до структурированных записей, включая таблицы и базы данных, - каждый элемент играет важную роль в формировании информационного ландшафта компании.

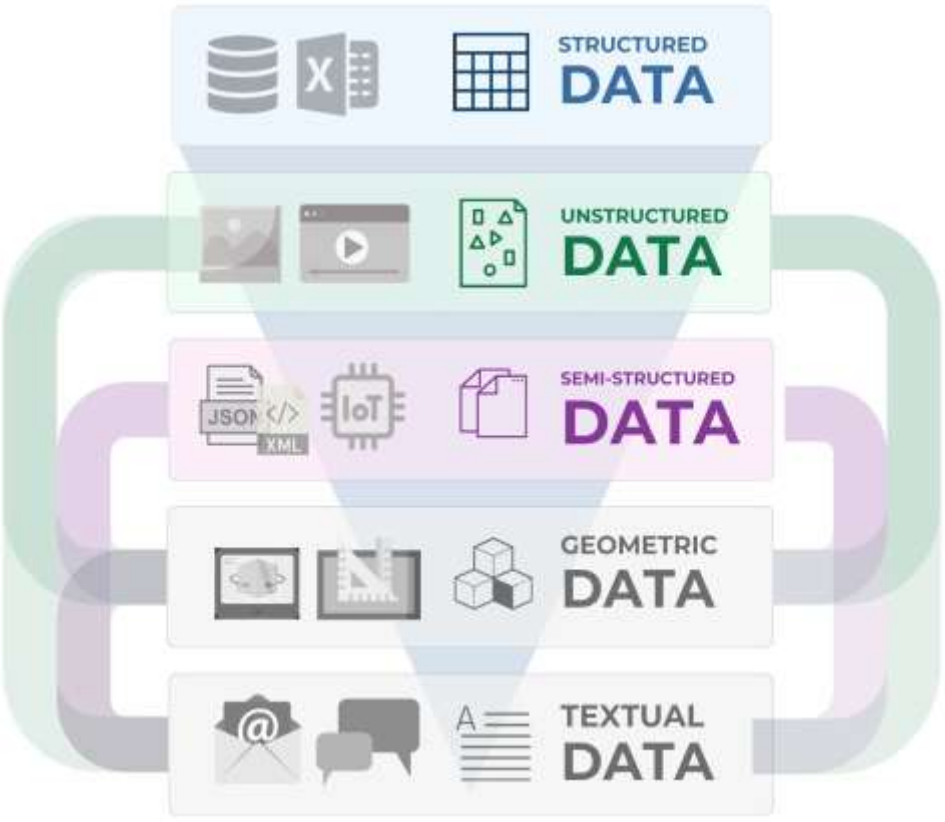


Рис. 3.1-1 Инженеры и менеджеры по работе с данными должны научиться работать со всеми типами данных, используемых в строительной отрасли.

Вот пример списка лишь некоторых систем и связанных с ними типов данных (Рис. 3.1-2), используемых в строительстве:

- **ERP** (Enterprise Resource Planning) - обрабатывает как правило структурированные данные, помогая управлять ресурсами предприятия и интегрировать различные бизнес-процессы.
- **CAD** (Computer-Aided Design) в сочетании с **BIM** (Building Information Modeling) - использует геометрические и полуструктурированные данные для проектирования и моделирования строительных проектов, обеспечивая точность и согласованность информации на этапе проектирования.
- **GIS** (Geographic Information Systems) - работает с геометрическими и структурированными данными для создания и анализа картографических данных и пространственных отношений.
- **RFID** (Radio-Frequency Identification) - использует полуструктурированные данные для эффективного отслеживания материалов и оборудования на строительной площадке с помощью радиочастотной идентификации.
- **ECM** (Engineering Content Management) — это система управления инженерными данными и документацией, включая полуструктурированные и неструктурированные данные, такие как технические чертежи и проектная документация.

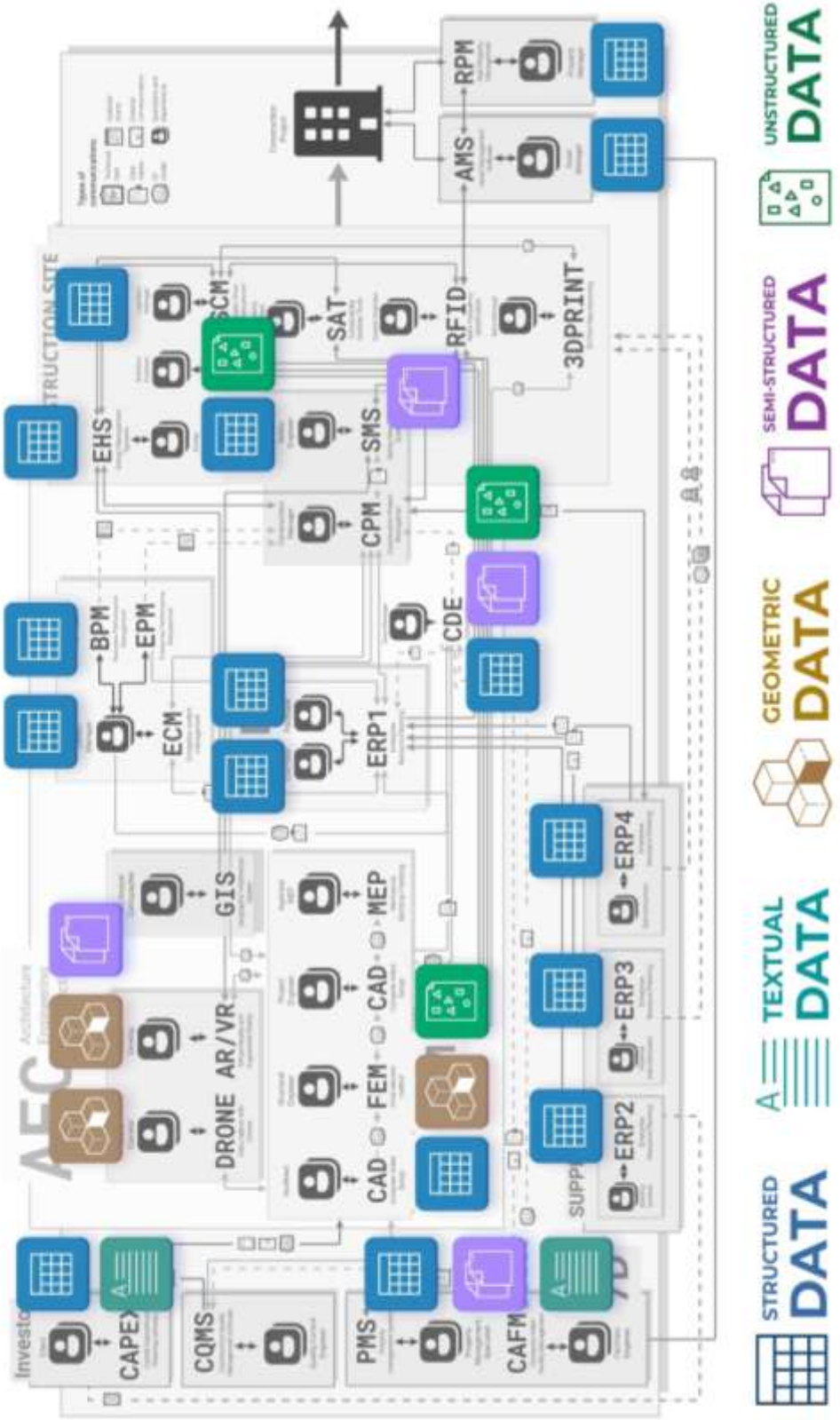


Рис. 3.1-2 Различные форматы и данные заполняют различные системы, требуя перевода в форму, пригодную для комплексной интеграции.

Эти и многие другие системы компании управляют широким спектром данных, от структурированных табличных данных до сложных геометрических моделей, обеспечивая интегрированное взаимодействие в процессах проектирования, планирования и управления строительством.

В примере упрощенного диалога (Рис. 3.1-3) между специалистами по строительному проекту происходит обмен различными типами данных:

- **Архитектор**: "*Учитывая пожелания клиента, я добавил зону отдыха на крыше. Пожалуйста, ознакомьтесь с новым дизайном*" (геометрические данные - модель).
- **Инженер-конструктор**: "*Проект получен. Я рассчитываю несущую способность крыши для новой зоны отдыха*" (структурированные и полуструктурированные данные - расчетные таблицы).
- **Менеджер по закупкам**: "*Нужны спецификации и количество материалов для зоны отдыха, чтобы организовать закупку*" (текстовые и полуструктурированные данные - списки и спецификации).
- **Инженер по охране труда и технике безопасности**: "*Получил данные о новой зоне. Оцениваю риски и обновляю план безопасности*" (полуструктурированные данные - документы и планы).
- **Специалист по BIM-моделированию**: "*Внесение изменений в общую модель проекта для корректировки рабочей документации*" (геометрические данные и полуструктурированные данные).
- **Руководитель проекта**: "*Включаю новую зону отдыха в график работ. Я обновляю графики и ресурсы в системе управления проектом*" (структурированные и полуструктурированные данные - графики и планы).
- **Специалист по обслуживанию объектов (FM)**: "*Я готовлю данные для будущего обслуживания зоны отдыха и ввожу их в систему управления имуществом*" (структурированные и полуструктурированные данные - инструкции и планы технического обслуживания).

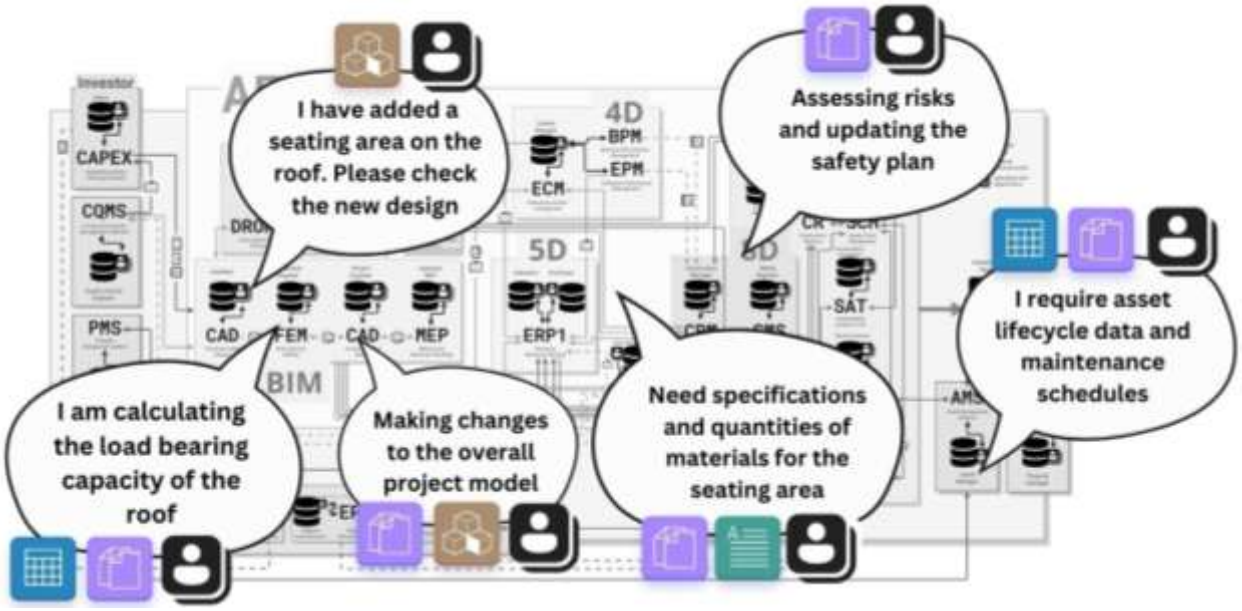


Рис. 3.1-3 Общение между специалистами происходит как на уровне текста, так и на уровне данных.

Каждый специалист работает с различными типами данных, обеспечивая эффективное взаимодействие команды и успешное выполнение проекта. Понимание различий между структурированными, полуструктурированными и неструктурированными данными позволяет осознать уникальную роль каждого типа в цифровых бизнес-процессах. Важно не только знать, что существуют разные формы данных, но и понимать, как, где и зачем они применяются.

> Ещё недавно идея объединения таких разнотипных данных казалась амбициозной, но с трудом реализуемой. Сегодня — это уже часть повседневной практики. Интеграция данных разных схем и структур стала неотъемлемой частью современной архитектуры информационных систем.

В следующих главах мы подробно рассмотрим ключевые стандарты и подходы, которые позволяют объединять структурированные, полуструктурированные и неструктурированные данные в единое согласованное представление. Особое внимание будет уделено структурированным данным и реляционным базам данных — как основным механизмам хранения, обработки и анализа информации в строительной отрасли.

Структурированные данные

В строительной отрасли информация поступает из множества источников — чертежей, спецификаций, графиков и отчетов. Чтобы эффективно управлять этим потоком, требуется ее структурирование. Структурированные данные позволяют организовать информацию в удобной, читаемой и доступной форме.

> Согласно 5-му ежегодному отчету о строительных технологиях JB Knowledge [17], 67% специалистов по управлению строительными проектами отслеживают и оценивают эффективность работ вручную или с помощью электронных таблиц.

Одними из наиболее распространённых форматов структурированных данных являются XLSX и CSV. Они широко применяются для хранения, обработки и анализа информации в электронных таблицах. В таких таблицах данные представлены в виде строк и столбцов, что делает их удобными для чтения, редактирования и анализа

Формат XLSX, созданный компанией Microsoft, основан на использовании XML-структур и архивируется с помощью ZIP-алгоритма. Основные особенности формата:

- Поддержка сложных формул, диаграмм и макросов.
- Возможность хранения данных в разных листах, а также форматирования информации.
- Оптимизирован для работы в среде Microsoft Excel, но совместим и с другими офисными пакетами.

Формат CSV представляет собой простой текстовый файл, в котором значения разделяются запятыми, точками с запятой или другими символами-разделителями. Основные преимущества:

- Универсальная совместимость с различными программами и операционными системами.
- Удобство импорта/экспорта в базы данных и аналитические системы.

- Легкость обработки даже в текстовых редакторах.

Однако CSV не поддерживает формулы и форматирование, поэтому его основное применение — это обмен данными между системами и массовое обновление информации. Благодаря своей универсальности и независимости от платформ, CSV стал популярным инструментом передачи данных в гетерогенных ИТ-средах.

Оба формата XLSX и CSV выступают связующим звеном между различными системами, работающими со структурированными данными (Рис. 3.1-4). Они особенно полезны в задачах, где важны читаемость, ручное редактирование и базовая совместимость.

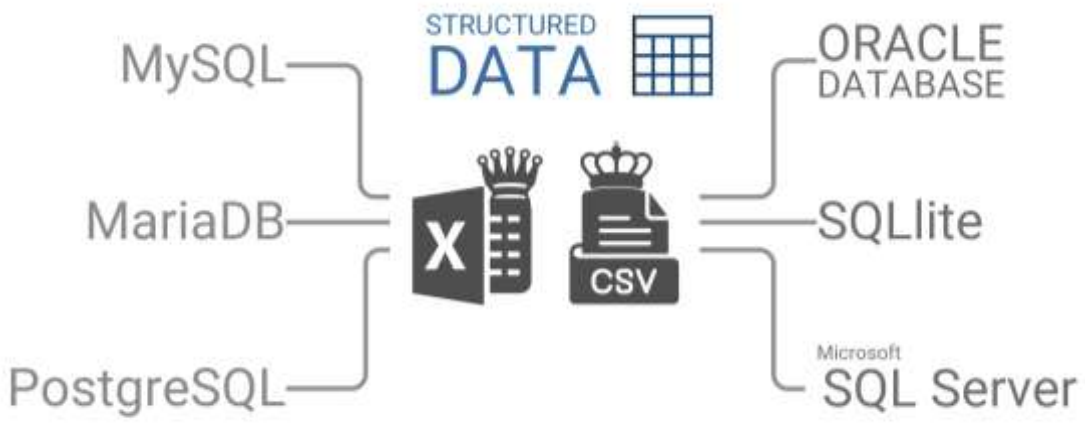


Рис. 3.1-4 Форматы XLSX и CSV являются связующим звеном между различными системами, работающими со структурированными данными.

> Независимость от платформы делает CSV самым популярным форматом для передачи данных в гетерогенных ИТ-средах и системах.

Тем не менее, XLSX и CSV не предназначены для высокопроизводительных вычислений или долговременного хранения больших объемов данных. Для таких целей используются более современные структурированные форматы, такие как Apache Parquet, Apache ORC, Feather, HDF5. Эти форматы мы подробнее рассмотрим в главе «Хранение больших данных: анализ популярных форматов и их эффективность» девятой части книги.

На практике Excel с форматом XLSX чаще применяется для небольших задач и автоматизации рутинных процессов. Более сложные сценарии требуют использования систем управления данными, таких как ERP, PMIS CAFM, CPM, SCM и другие (Рис. 3.2-1). Именно в этих системах хранятся структурированные данные, на которых основывается организация и управление информационными потоками компании.

Современные информационные системы управления данными, применяемые в строительной отрасли, опираются на структурированные данные, организованные в виде таблиц. Для надежного,

масштабируемого и целостного управления большими объемами информации разработчики приложений и систем обращаются к реляционным системам управления базами данных (RDBMS).

Реляционные базы данных RDBMS и язык запросов SQL

Для эффективного хранения, обработки и анализа данных используются **реляционные базы данных (RDBMS)** — это системы хранения данных, которые организуют информацию в таблицы с определенными отношениями между ними.

> Данные, организованные в базы данных (RDBMS), не просто представляют собой цифровую информацию; они являются основой для транзакций и взаимодействия между различными системами.

Вот несколько наиболее распространенных реляционных систем управления базами данных (RDBMS) (Рис. 3.1-5):

- **MySQL** (Open Source) – одна из самых популярных RDBMS, входящая в состав стека LAMP (Linux, Apache, MySQL, PHP/Perl/Python). Широко применяется в веб-разработке благодаря простоте и высокой производительности.
- **PostgreSQL** (Open Source) – мощная объектно-реляционная система, известная своей надежностью и расширенными возможностями. Подходит для сложных корпоративных решений.
- **Microsoft SQL Server** – коммерческая система от Microsoft, широко используемая в корпоративной среде за счет интеграции с другими продуктами компании и высокого уровня безопасности.
- **Oracle Database** – одна из самых мощных и надежных СУБД, применяемая в крупных предприятиях и критически важных приложениях.
- **IBM DB2** – ориентирована на крупные корпорации, обеспечивая высокую производительность и отказоустойчивость.
- **SQLite** (Open Source) – легковесная встраиваемая база данных, идеально подходящая для мобильных приложений и автономных систем, таких как программы для проектирования CAD (BIM).

Популярные системы управления базами данных в строительном бизнесе - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 и SQLite - работают со структурированными данными. Все эти СУБД представляют собой мощные и гибкие решения для управления широким спектром бизнес-процессов и приложений, от небольших веб-сайтов до масштабных корпоративных систем (Рис. 3.2-1).

> Согласно данным Statista [48], в 2022 году реляционные системы управления базами данных (RDBMS) составляли около 72% от общего числа используемых DBMS.

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

Рис. 3.1-5 Популярность использования структурированных баз данных (отмечено синим) в рейтинге DBMS (по материалам [49]).

Установить базы данных с открытым исходным кодом можно довольно просто — даже без глубоких технических знаний. Такие системы с открытым исходным кодом, как PostgreSQL, MySQL или SQLite, доступны бесплатно и работают на большинстве операционных систем: Windows, macOS и Linux. Всё, что нужно — это зайти на официальный сайт проекта, скачать установщик и следовать инструкциям. В большинстве случаев установка занимает не больше 10–15 минут. Одну из таких баз данных мы смоделируем и создадим в четвертой части книги (Рис. 4.3-8).

Если в вашей компании используются облачные сервисы (например, Amazon Web Services, Google Cloud или Microsoft Azure), то развернуть базу можно в пару кликов — платформа предложит вам готовые шаблоны для установки. Благодаря открытости кода, такие базы легко настраивать под свои задачи, а огромное сообщество пользователей всегда поможет найти решение любой проблемы.

RDBMS остаются основой для множества бизнес-приложений и аналитических платформ (Рис. 3.1-6), которые позволяют компаниям эффективно хранить, обрабатывать и анализировать данные — а значит, принимать обоснованные и своевременные решения.

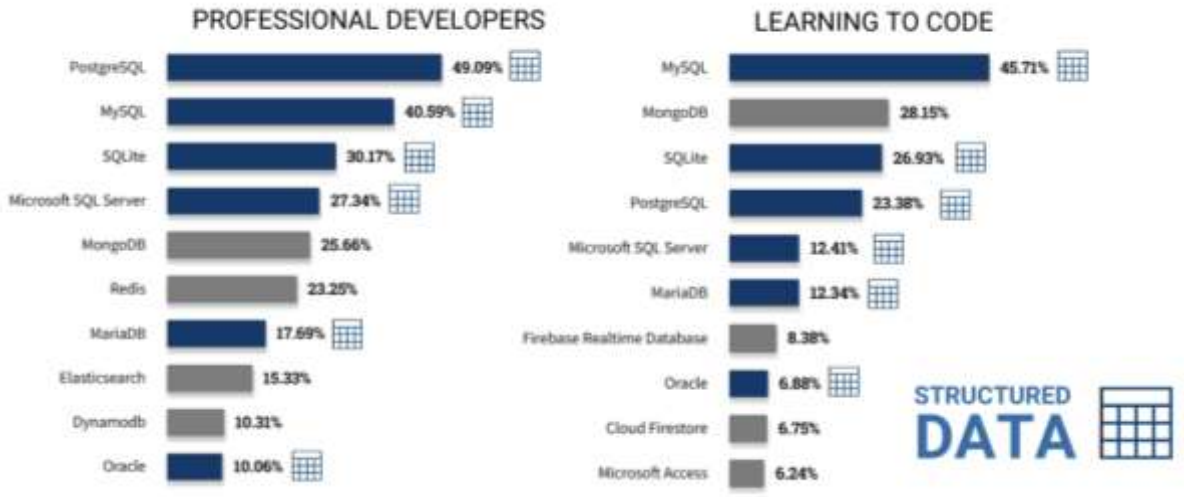


Рис. 3.1-6 Опрос разработчиков на StackOverFlow (крупнейшем ИТ-форуме) о том, какие базы данных они использовали в прошлом году и какие хотят использовать в следующем (RDBMS выделены синим цветом) (по материалам [50]).

RDBMS обеспечивают надёжность, согласованность данных, поддержку транзакций и используют мощный язык запросов — SQL (Structured Query Language), который часто используется в аналитике и позволяет легко получать, изменять и анализировать информацию, хранимую в базах данных. Именно SQL является основным инструментом работы с данными в реляционных системах.

SQL-запросы в базах данных и новые тренды

Основное преимущество языка SQL, часто используемого в реляционных базах данных, перед другими видами управления информацией (например, с помощью классических электронных таблиц Excel) заключается в поддержке очень больших объемов баз данных при высокой скорости обработки запросов.

> Язык структурированных запросов (SQL) – это специализированный язык программирования, предназначенный для хранения, обработки и анализа информации в реляционных базах данных. SQL используется для создания, управления и обращения к данным, позволяя эффективно находить, фильтровать, объединять и агрегировать информацию. Он служит ключевым инструментом доступа к данным, обеспечивая удобный и формализованный способ взаимодействия с хранилищами информации.

Эволюция систем SEQUEL-SQL проходит через такие значимые продукты и компании, как Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL и MySQL, и завершается появлением SQLite и MariaDB [51]. SQL предоставляет возможности по работе с таблицами, которых нет в Excel, делая работу с данными более масштабируемой, безопасной и удобной для автоматизации:

- **Создание и управление структурами данных (DDL):** в SQL можно создавать, изменять и удалять таблицы в базе данных, устанавливать связи между ними и определять структуры хранения данных. В Excel же работа ведется с фиксированными листами и ячейками, без

четко заданных связей между листами и наборами данных.

- **Операции с данными (DML):** SQL позволяет массово добавлять, изменять, удалять и извлекать данные с высокой скоростью, выполняя сложные запросы с фильтрацией, сортировкой и объединением таблиц (Рис. 3.1-7). В Excel обработка больших объемов информации требует ручных действий или специальных макросов, что замедляет процесс и повышает вероятность ошибок.
- **Контроль доступа (DCL):** SQL позволяет разграничивать права доступа к данным для разных пользователей, ограничивая возможность редактирования или просмотра информации. В Excel же доступ либо общий (при передаче файла), либо требует сложных настроек с разделением прав через облачные сервисы.

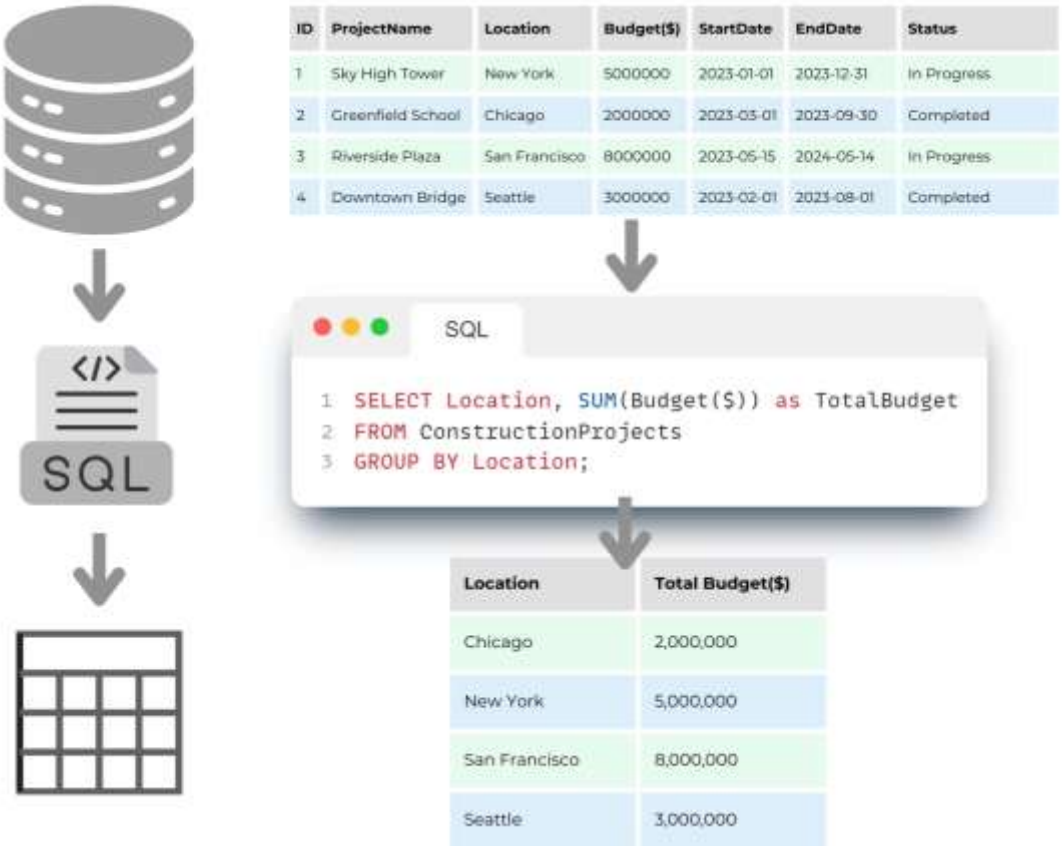


Рис. 3.1-7 Пример DML в SQL: быстрая обработка, группировка и агрегирование при помощи нескольких строк кода для автоматической обработки данных.

Excel облегчает работу с данными благодаря своей визуальной и интуитивно понятной структуре. Однако с увеличением объема данных производительность Excel снижается. Excel также сталкивается с ограничениями по объему хранимых данных – максимум один миллион строк, и производительность снижается задолго до достижения этого предела. Поэтому, хотя Excel и выглядит предпочтительнее для визуализации и манипуляций с небольшими объемами данных, но для работы с большими массивами данных лучше подходит SQL.

Следующим этапом в развитии структурированных данных стало появление колоночных баз дан-

ных (Columnar Databases), которые представляют собой альтернативу традиционным реляционным базам данных, особенно когда речь идет о значительно больших объемах данных и аналитических вычислениях. В отличие от строковых СУБД, где данные хранятся построчно, в колоночных базах информация записывается по столбцам. По сравнению с классическими базами данных это позволяет:

- Уменьшить объем хранения за счет эффективного сжатия однотипных данных в колонках.
- Ускорить аналитические запросы, так как считываются только необходимые столбцы, а не вся таблица.
- Оптимизировать работу с Big Data и хранилищами данных, например, в архитектуре Data Lakehouse.

О колоночных базах данных, Pandas DataFrame, Apache Parquet, HDF5, а также о создании на их основе Big Data-хранилищ для целей анализа и обработки данных мы будем говорить подробнее в следующих главах этой книги – «DataFrame: универсальный формат табличных данных» и «Форматы хранения данных и работа с Apache Parquet: DWH-хранилища данных и архитектура Data Lakehouse».

Неструктурированные данные

Несмотря на то, что большинство данных, используемых в приложениях и информационных системах, имеет структурированную форму, наибольшая часть генерируемой информации в строительстве представлена в виде неструктурированных данных — изображений, видео, текстовых документов, аудиозаписей и других форм контента. Это особенно актуально на стадии строительства, эксплуатации и технического надзора, где преобладает визуальная и текстовая информация.

> **Неструктурированные данные** — это информация, которая не имеет заранее заданной модели или структуры, не организована в традиционные строки и столбцы, как в базах данных или таблицах.

В общем виде неструктурированные данные можно классифицировать на две категории:

- Генерируемые людьми неструктурированные данные, к которым относятся различные виды создаваемого людьми контента: текстовые документы, электронные письма, изображения, видео и так далее.
- Машиногенерируемые неструктурированные данные создаются устройствами и датчиками: это файлы журналов, данные GPS, результаты работы Internet of Things (IoT) и, например, другая телеметрическая информация со строительной площадки.

В отличие от структурированных данных, которые удобно организованы в таблицы и базы данных, неструктурированные данные требуют дополнительных этапов обработки перед их интеграцией в информационные системы (Рис. 3.1-8). Использование технологий автоматизированного сбора, анализа и преобразования таких данных открывает новые возможности для повышения эффективности строительства, сокращения ошибок и минимизации влияния человеческого фактора.

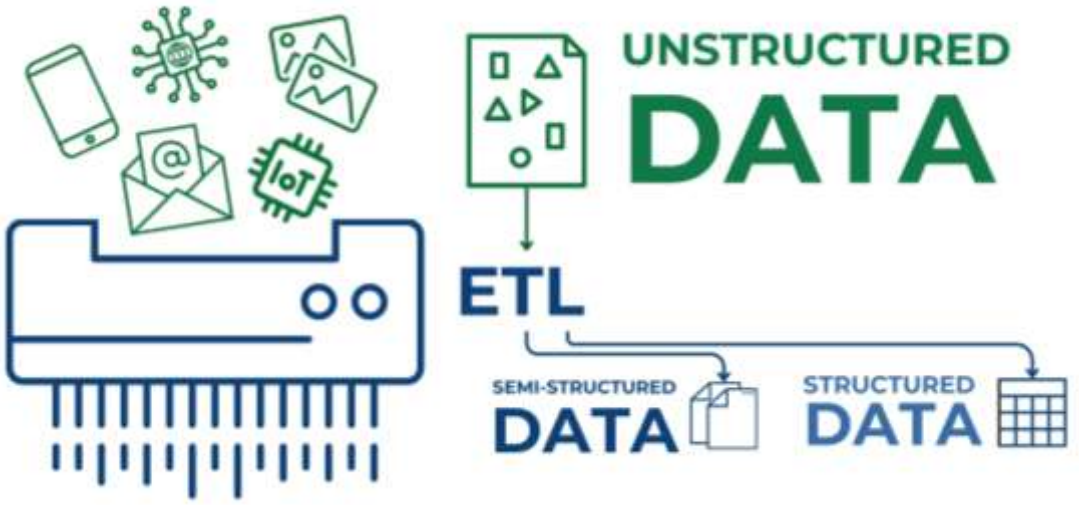


Рис. 3.1-8 Обработка неструктурированных данных начинается с их преобразования в полуструктурированные и структурированные данные.

Неструктурированные данные составляют до 80% всей информации [52], с которой сталкиваются специалисты в компаниях, поэтому мы будет подробно с примерами рассматривать их виды и обработку в следующих главах книги.

Для удобства обсуждения текстовые данные выделим в отдельную категорию. Хотя они являются разновидностью скорее неструктурированных данных, их значимость и распространенность в строительной отрасли требуют особого внимания.

Текстовые данные: между неструктурированным хаосом и структурой

Текстовые данные в строительной отрасли охватывают широкий спектр форматов и типов информации, от бумажных документов до неформальных методов общения, таких как письма, разговоры, рабочая переписка и устные встречи на строительной площадке. Все эти текстовые данные несут в себе важную информацию для управления строительными проектами - от деталей проектных решений и изменений в планах до обсуждения проблем безопасности и переговоров с подрядчиками и клиентами (Рис. 3.1-9).

Рис. 3.1-9 Текстовые данные, один из самых популярных типов информации, используемой в общении между участниками проекта.

Текстовая информация может быть как формализованной, так и неструктурированной. К формализованным данным относятся документы в формате Word (.doc, .docx), PDF, а также текстовые файлы протоколов совещаний (.txt). Неформализованные данные включают в себя переписку в мессенджерах и почте, стенограммы совещаний (Teams, Zoom, Google Meet), а также аудиозаписи обсуждений (.mp3, .wav), которые требуют преобразования в текст.

Но если письменные документы, такие как официальные запросы, условия контрактов и электронные сообщения, обычно уже имеют определенную структуру, то устные сообщения и рабочая переписка часто остаются неструктурированными, что затрудняет их анализ и интеграцию в системы управления проектами.

Ключ к эффективному управлению текстовыми данными — их преобразование в структурированный формат. Это позволяет автоматически интегрировать обработанную информацию в существующие системы, которые уже работают со структурированными данными.

Рис. 3.1-10 Преобразование текстового контента в структурированные данные.

Чтобы эффективного использовать текстовую информацию, ее необходимо автоматически преобразовывать в структурированную форму (Рис. 3.1-10). Этот процесс обычно включает несколько этапов:

- **Распознавание текста (OCR)** — преобразование изображений документов и чертежей в машинно-читаемый формат.
- **Анализ текста (NLP)** — автоматическая идентификация ключевых параметров (дат, сумм и цифр, относящихся к проекту).
- **Классификация данных** — распределение информации по категориям (финансы, логистика, управление рисками).

После распознавания и классификации уже структурированные данные можно интегрировать в базы данных и использовать в автоматизированных системах отчетности и управления.

Полуструктурированные и слабоструктурированные данные

Полуструктурированные данные содержат определенный уровень организации, но не имеют строгой схемы или структуры. Хотя такая информация включает структурированные элементы (напри-

мер: даты, имена сотрудников и списки выполненных задач), формат представления может значительно различаться в разных проектах или даже у отдельных работников. Примерами таких данных являются журналы учета рабочего времени, отчеты о проделанной работе и графики, которые могут быть представлены в различных форматах.

Полуструктурированные данные легче анализировать, чем неструктурированные, однако они требуют дополнительной обработки для интеграции в стандартизированные системы управления проектами.

> Работа с полуструктурированными данными, характеризующимися наличием постоянно меняющейся структуры, представляет значительные трудности. Это связано с тем, что изменчивость структуры данных требует отдельных индивидуальных подходов к обработке и анализу каждого источника полуструктуированных данных.

Но если работа с неструктурированными данными требует больших усилий, то обработка полуструктурированных данных может быть выполнена с помощью относительно простых методов и инструментов.

> **Слабоструктурированные данные** — более общий термин, который описывает данные с минимальной или неполной структурой. Чаще всего это текстовые документы, чаты, электронные письма, где встречаются отдельные метаданные (например, дата, отправитель), но большая часть информации представлена в хаотичном виде.

В строительстве слабоструктурированные данные встречаются в различных процессах. Например, к ним могут относиться:

- Сметные расчеты и коммерческие предложения – таблицы с данными о материалах, объемах и стоимости, но без единого формата.
- Чертежи и инженерные схемы – файлы в PDF или DWG, содержащие текстовые аннотации и метаданные, но без строго фиксированной структуры.
- Графики выполнения работ – данные из MS Project, Primavera P6 или других систем, которые могут иметь разную структуру экспорта.
- CAD (BIM-модели) – содержат элементы структуры, но представление данных зависит от ПО и стандарта проекта.

Геометрические данные, производимые системами CAD, можно классифицировать так же, как полуструктурированные данные. Однако мы выделим геометрические CAD (BIM) данные в отдельный тип, поскольку они, как и текстовые данные, могут часто рассматриваться в процессах компании как отдельный тип данных.

Геометрические данные и их применение

Если метаданные об элементах проекта практически всегда хранятся в виде таблиц, структурированных или слабоструктурированных форматах, то геометрические данные элементов проекта в

большинстве случаев создаются с помощью специальных CAD инструментов (Рис. 3.1-11), позволяющих детально визуализировать элементы проекта в виде набора линий (2D) или геометрических тел (3D).

Рис. 3.1-11 CAD инструменты помогли перенести геометрическую информацию с физических носителей в форму баз данных.

При работе с геометрическими данными в строительстве и архитектуре можно выделить три основные области применения геометрических данных (Рис. 3.1-12):

- **Подтверждение объемов**: геометрические данные, генерируемые внутри программ CAD (BIM) при помощи специальных геометрических ядер, необходимы для автоматического и точного определения объемов и размеров элементов проекта. Эти данные включают автоматически рассчитанные площади, объемы, длины и другие важные атрибуты, необходимые для планирования, составления бюджета и заказа ресурсов и материалов.
- **Визуализация проекта**: в случае каких-либо изменений в проекте, визуализация элементов позволяет автоматически генерировать обновленные чертежи в различных плоскостях. Визуализация проекта на начальных этапах позволяет ускорить взаимопонимание между всеми участниками, чтобы сэкономить время и ресурсы в процессе строительства.
- **Проверка столкновений**: в сложных строительных и инженерных проектах, где взаимодействие нескольких категорий элементов (например, труб и стен) без "геометрических конфликтов" является критически важным, проверка столкновений играет ключевую роль. Использование программного обеспечения для обнаружения коллизий позволяет заблаговременно выявлять потенциальные геометрические конфликты между элементами проекта, предотвращая дорогостоящие ошибки в процессе строительства.

С самого начала появления инженерно-конструкторских бюро, со времён строительства первых сложных конструкций, инженеры-конструкторы предоставляли геометрическую информацию в виде чертежей, линий и плоских геометрических элементов (на папирусе, ватмане “A0” или в форматах DWG, PDF, PLT), на основании которых прорабы и сметчики (Рис. 3.1-11), на протяжении последних тысячелетий, с помощью линеек и транспортиров собирали атрибутивные объемы или количество элементов и групп элементов.

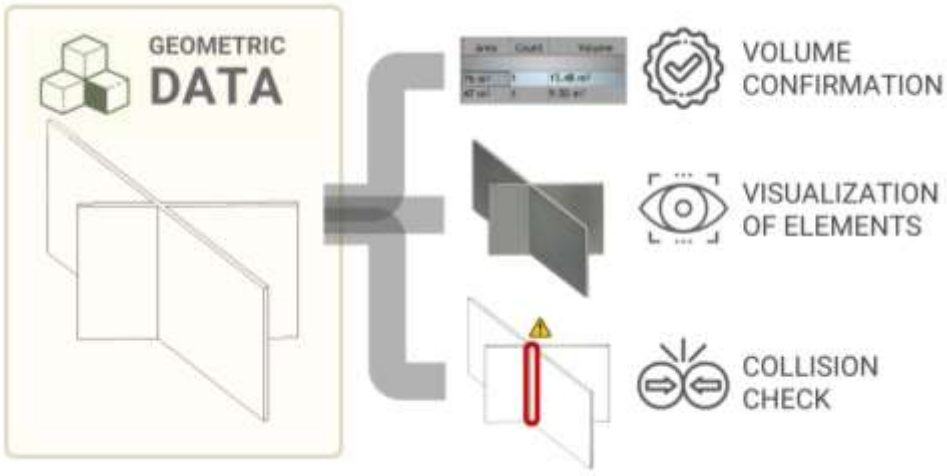


Рис. 3.1-12 Геометрия является основой для получения объёмных параметров элементов, которые затем используются для расчета стоимости и сроков проекта.

Сегодня эта ручная и трудоемкая задача решается путем полной автоматизации благодаря появлению объемного моделирования в современных инструментах CAD (BIM), которое позволяет автоматически, при помощи специального геометрического ядра, получить объемные атрибуты любого элемента без необходимости рассчитывать объёмные параметры вручную.

Современные инструменты CAD позволяют также классифицировать и категоризировать элементы проекта, чтобы можно было выгрузить из базы данных проекта таблицы спецификаций для использования в различных системах, например, для оценки стоимости, составления графиков или расчета CO_2 (Рис. 3.1-13). О Получении спецификаций, таблиц QTO и объемов, а также о практических примерах мы поговорим в главе "Получение объемов и количественный расчет".

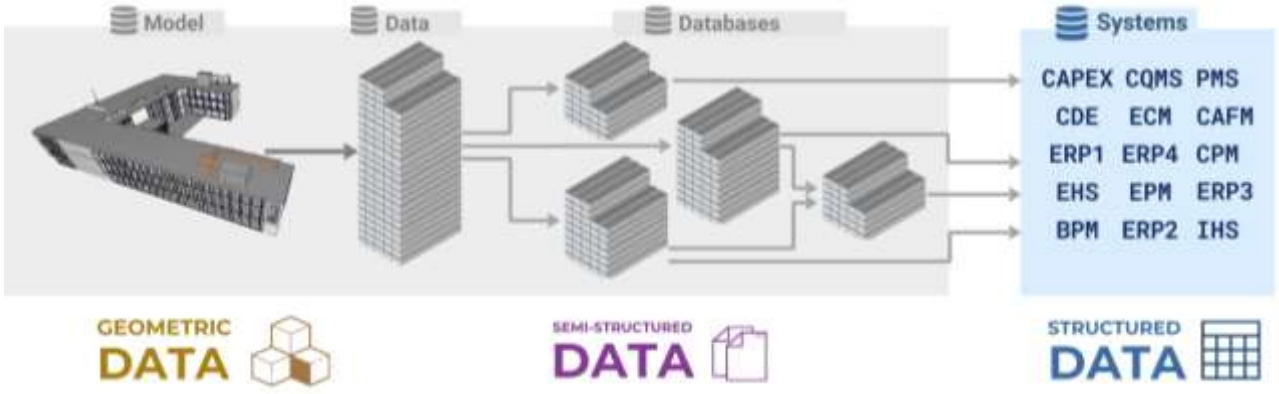


Рис. 3.1-13 Инструменты CAD (BIM) сохраняют данные в базах данных, которые предназначены для интеграции и взаимодействия с другими системами.

Из-за закрытости баз данных и форматов, используемых в CAD-среде, геометрические данные, создаваемые в CAD-решениях, фактически оформились в отдельный тип информации. Он сочетает в себе как геометрию элементов, так и метаинформацию (структурированную или полуструктурированную), заключённую в специализированные файлы и форматы.

CAD данные: от проектирования до хранения данных

В современных CAD- и BIM-системах данные хранятся в собственных, часто проприетарных форматах: DWG, DXF, RVT, DGN, PLN и других. Эти форматы поддерживают как двухмерные, так и трёхмерные представления объектов, сохраняя при этом не только геометрию, но и связанные с объектами атрибуты. Вот наиболее распространённые из них:

- **DWG** —бинарный формат файла, используемый для хранения двухмерных (и реже трёхмерных) проектных данных и метаданных.
- **DXF** – текстовый формат для обмена 2D- и 3D-чертежами между CAD-системами. Содержит геометрию, слои и атрибутные данные, поддерживает как ASCII, так и бинарное представление.
- **RVT** – бинарный формат хранения CAD моделей, включающий 3D-геометрию, атрибуты элементов, связи и параметры проекта.
- **IFC** – открытый текстовый формат для обмена строительными данными между CAD (BIM) системами. Включает геометрию, свойства объектов и информацию об их взаимосвязях.

Помимо них используются и другие форматы: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE . Хотя они различаются по назначению и уровню открытости (Рис. 3.1-14), все они могут представлять одну и ту же информационную модель проекта в разных формах. В комплексных проектах эти форматы часто применяются параллельно — от черчения до координации моделей проектов.

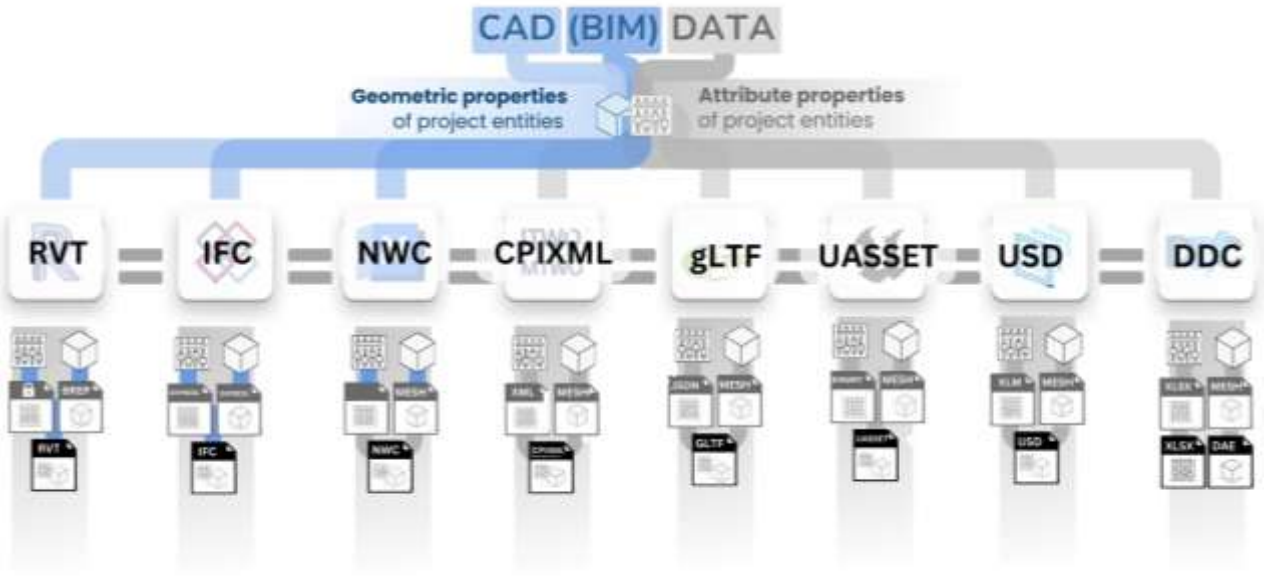


Рис. 3.1-14 Популярные форматы хранения информации из CAD описывают геометрию через параметры BREP или MESH, дополняя их атрибутивными данными.

Все перечисленные выше форматы позволяют хранить данные о каждом элементе строительного проекта и во всех перечисленных форматах содержатся два ключевых типа данных:

- **Геометрические параметры** — описывают форму, расположение и размеры объекта. Подробно геометрию и её использование будет обсуждаться в шестой части книги, посвящённой CAD (BIM) решениям;

- **Атрибутивные свойства** — содержат различную информацию: материалах, типах элементов, технических характеристиках, уникальных идентификаторах и других свойств, которые могут быть у элементов проекта.

В современных проектах особое значение приобретают именно атрибутивные данные, поскольку они определяют эксплуатационные характеристики объектов, позволяют выполнять инженерные, калькуляционные расчёты и обеспечивают сквозное взаимодействие между участниками проектирования, строительства и эксплуатации. Например:

- Для окон и дверей указываются: тип конструкции, вид остекления, направление открывания (Рис. 3.2-1).
- Для стен фиксируется информация о материалах, теплоизоляции и акустических характеристиках.
- Для инженерных систем хранятся параметры трубопроводов, воздуховодов, кабельных трасс и их соединений.

Эти параметры могут храниться как внутри самих CAD-(BIM-)файлов, так и во внешних базах данных — в результате экспорта, конвертации или прямого доступа к внутренним структурам CAD через инструменты обратного инжиниринга. Такой подход облегчает интеграцию проектной информации с другими корпоративными системами и платформами.

> **Обратный инжиниринг** в контексте CAD (BIM) — это процесс извлечения и анализа внутренней структуры цифровой модели с целью воссоздания её логики, структуры данных и зависимостей без доступа к исходным алгоритмам или документации.

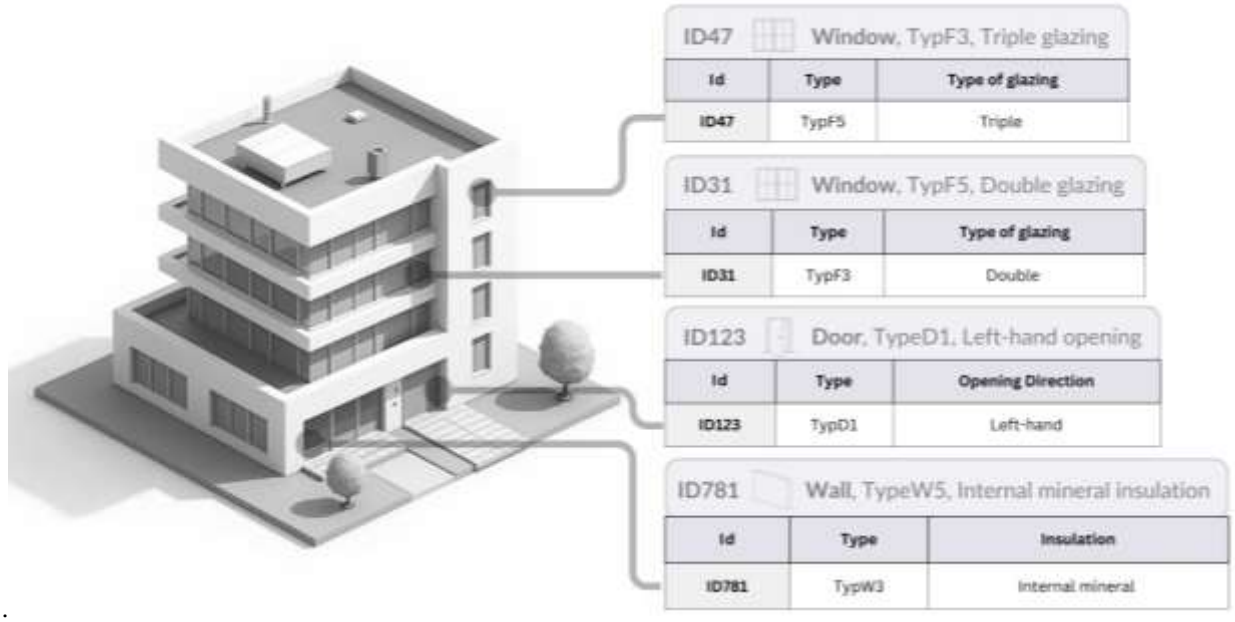


Рис. 3.1-15 Элемент проекта, помимо описания параметрической или полигональной геометрии, содержит информацию о параметрах и свойствах элементов.

В итоге вокруг каждого элемента формируется уникальный набор параметров и свойств, включающий как уникальные характеристики каждого объекта (например, идентификатор и размеры), так и общие атрибуты для групп элементов. Это позволяет не только анализировать отдельные элементы-сущности проекта, но и объединять их в логические группы, которые затем могут использовать другие специалисты для своих задач и расчётов в системах и базах данных.

Сущность (англ. entity) — это конкретный или абстрактный объект реального мира, который можно однозначно выделить, описать и представить в виде данных.

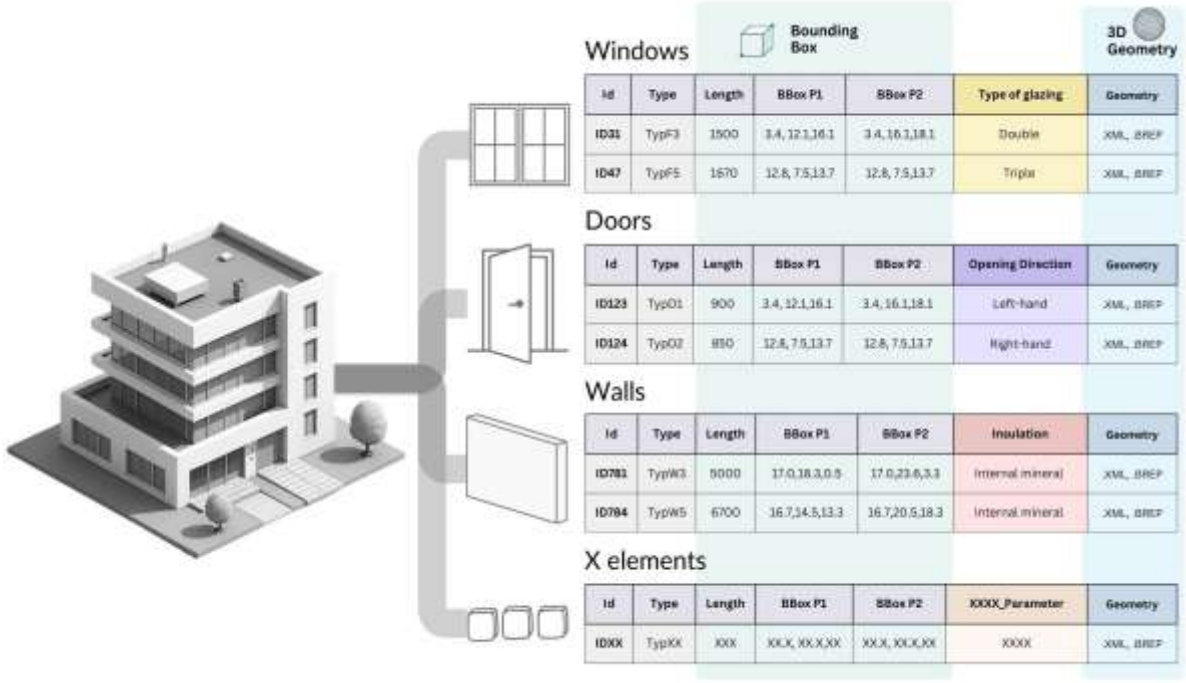


Рис. 3.1-16 Каждый элемент проекта содержит атрибуты, которые или заносятся проектировщиком, или высчитываются внутри CAD программы.

За последние десятилетия в строительной отрасли появилось множество новых CAD (BIM) форматов, упрощающих создание, хранение и передачу данных. Эти форматы могут быть закрытыми и открытыми, табличными, параметрическими или графовыми. Однако их разнообразие и фрагментированность существенно усложняют управление данными на всех этапах жизненного цикла проекта. Таблица сравнения основных форматов, используемых для обмена информацией в строительстве, представлена на Рис. 3.1-17 (полная версия доступна по QR коду).

Для решения проблем интероперабельности и доступа к данным CAD в работу включаются менеджеры (BIM) и координаторы, задача которых — контролировать экспорт, проверять качество данных и интегрировать части CAD (BIM) данных в другие системы.

Однако из-за закрытости и сложности форматов автоматизировать этот процесс затруднительно, что вынуждает специалистов выполнять многие операции вручную, без возможности выстроить полноценные поточные процессы обработки данных (pipeline).

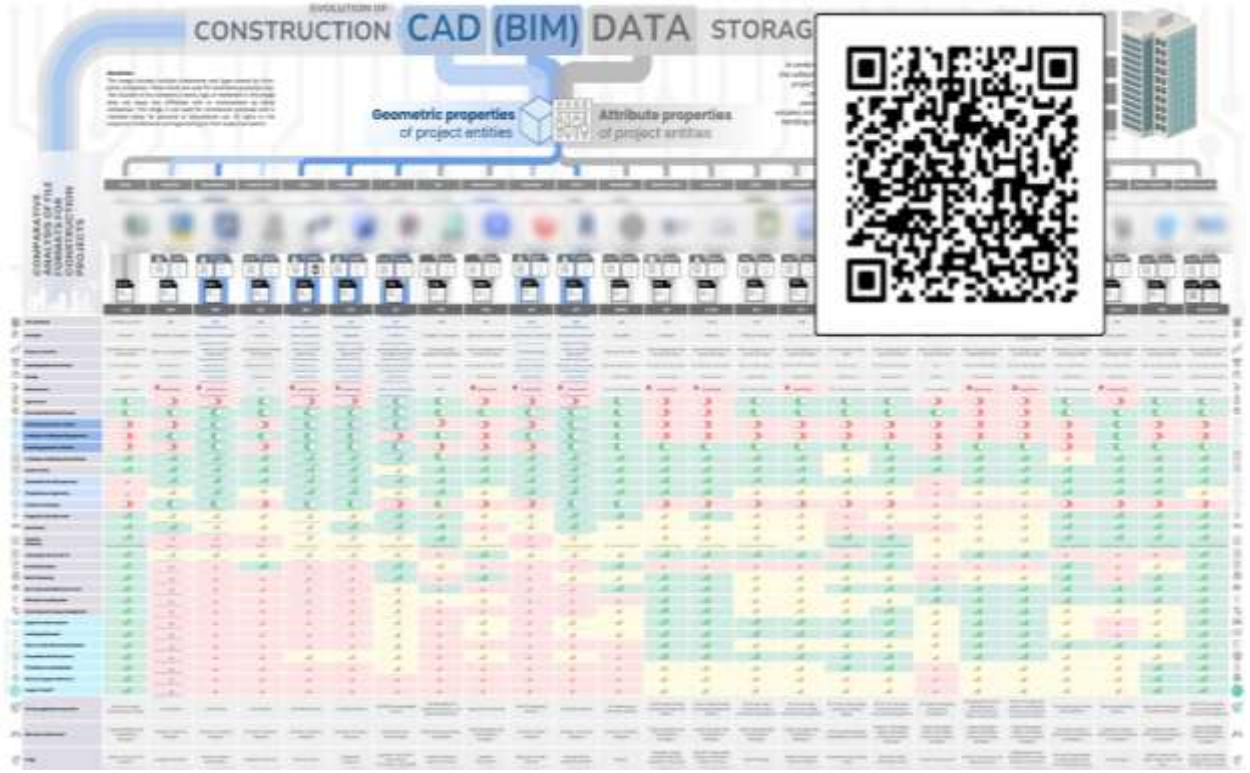


Рис. 3.1-17 Таблица сравнения основных форматов данных, в которых хранится информация об элементах проекта [53].

Чтобы разобраться, почему существует так много различных форматов данных, и почему большинство из них закрытые, важно углубиться в процессы, происходящие внутри CAD (BIM) программ, что будет подробно разобрано в шестой части книги.

Дополнительный информационный слой, добавляемый к геометрии, был представлен разработчиками CAD-систем в виде концепции BIM (Building Information Modeling) — маркетингового термина, активно продвигаемого в строительной отрасли с 2002 года [54].

Появление концепции BIM (BOM) и использование CAD в процессах

Концепция информационного моделирования зданий (BIM), впервые изложенная в документе Whitepaper BIM 2002 года [54], возникла благодаря маркетинговым инициативам производителей программного обеспечения CAD. Она возникла в результате маркетинговых инициатив разработчиков CAD-программного обеспечения и стала попыткой адаптировать принципы, уже хорошо зарекомендовавшие себя в машиностроении, к нуждам строительной отрасли.

Источником вдохновения для BIM стала концепция BOM (Bill of Materials) — спецификации состава изделия, активно применяемая в промышленности ещё с конца 1980-х годов. В машиностроении BOM позволял связывать данные из CAD-систем с системами PDM (Product Data Management), PLM (Product Lifecycle Management) и ERP, обеспечивая целостное управление инженерной информацией на протяжении всего жизненного цикла продукта (Рис. 3.1-8).

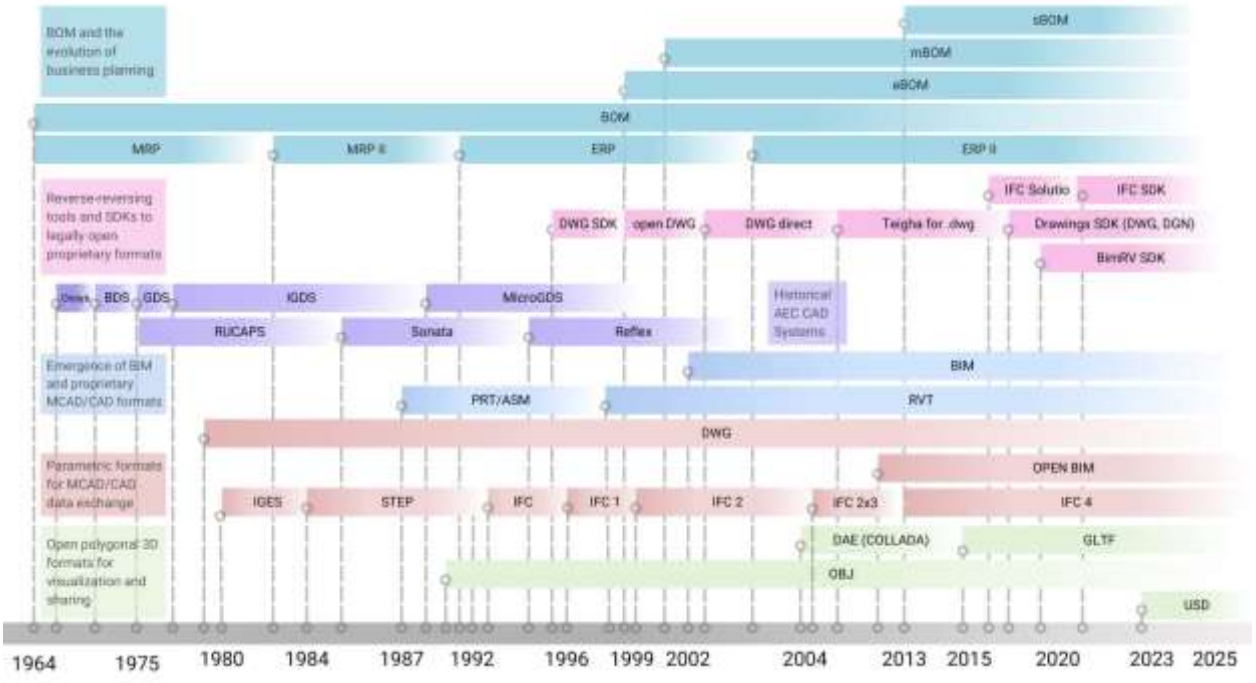


Рис. 3.1-18 Эволюция спецификаций (BOM), информационного моделирования (BIM), и цифровых форматов в инженерных строительной отрасли.

Современное развитие концепции BOM привело к появлению расширенной структуры — XBOM (Extended BOM), включающей не только состав изделия, но и поведенческие сценарии, требования по эксплуатации, параметры устойчивости и данные для прогнозной аналитики. XBOM по сути выполняет ту же роль, что и BIM в строительстве: оба подхода стремятся превратить цифровую модель в единый источник достоверной информации (Single Source of Truth) для всех участников проекта на протяжении всего жизненного цикла объекта.

Ключевым этапом в появлении BOM в строительстве стало появление в 2002 году первого параметрического CAD (MCAD), специально адаптированного для строительной отрасли. Его разработала команда, ранее создавшая Pro-E® — революционную MCAD-систему для машиностроения, появившуюся ещё в конце 1980-х годов и ставшую отраслевым стандартом [55].

Уже в конце 1980-х годов цель состояла в устранении ограничений [56], существовавших в тогдашних CAD-программах. Основной задачей было снизить трудозатраты на внесение изменений параметров элементов проекта и обеспечить возможность обновления модели на основе данных вне CAD-программ через базу данных [57]. Важнейшую роль в этом должна была играть параметризация: автоматическое получение характеристик из базы данных и использование их для актуализации модели внутри CAD-систем.

> Pro-E и концепция элементного параметрического моделирования с BOM, лежащая в его основе, оказали значительное влияние на развитие CAD- и MCAD-рынка [58]. На протяжении 25 лет эта модель в отрасли, а многие современные системы стали её концептуальными наследниками.

> *Цель состоит в том, чтобы создать систему, которая была бы достаточно гибкой, чтобы побудить инженера легко рассматривать различные конструкции. А затраты на внесение изменений в проект должны быть как можно ближе к нулю. Традиционные CAD / CAM программное обеспечение нереально ограничивает внесение недорогих изменений только в самый начальный этап процесса проектирования [59].*
>
> — Самуэль Гейзенберг, основатель компании Parametric Technology Corporation®, разработчика MCAD-продукта Pro-E и учитель создателя CAD продукта, использующего формат RVT

В машиностроении ключевыми платформами стали системы PDM, PLM, MRP и ERP. Они играют центральную роль в управлении данными и процессами, собирая информацию из CAx-систем (CAD, CAM, CAE) и организуя проектную деятельность на основе структуры изделия (BOM: eBOM, pBOM, mBOM) (Рис. 3.1-18). Такая интеграция позволяет сократить количество ошибок, избежать дублирования данных и обеспечить сквозную прослеживаемость на всех этапах — от проектирования до производства.

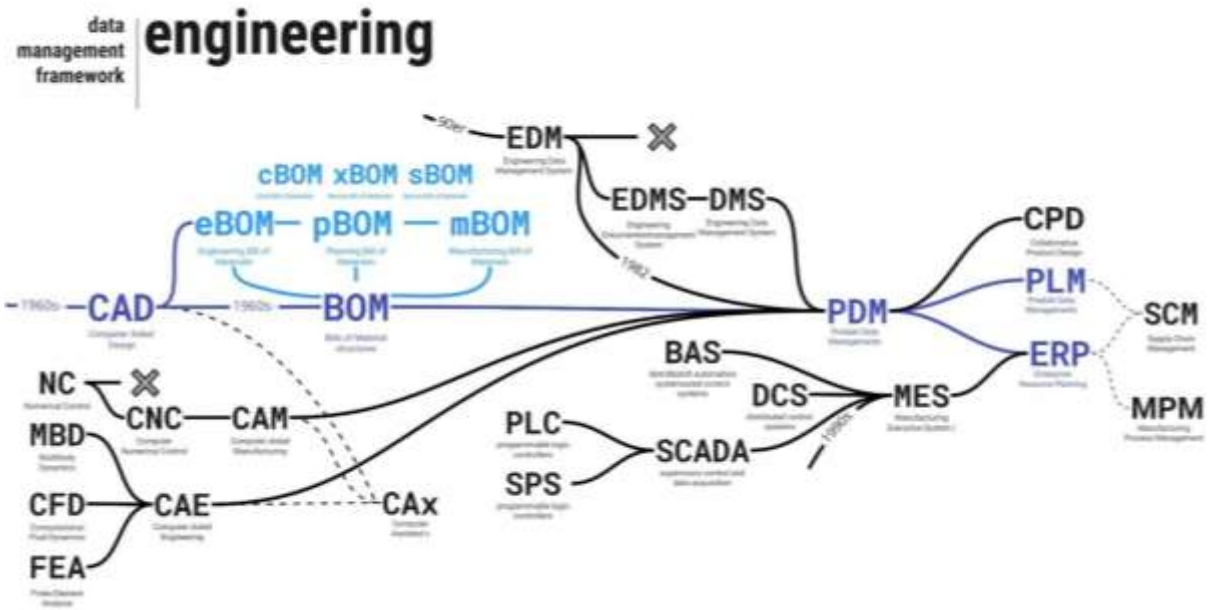


Рис. 3.1-19 Исторически BOM возник в 1960-х как способ структурировать данные из CAx-систем и передавать их в системы управления.

Покупка одним из ведущих вендоров CAD-решения, разработанного бывшей командой Pro-E и основанного на BOM-подходе, ознаменовалась почти мгновенной публикацией серии Whitepaper BIM (2002–2003 гг.) [60] [61]. Уже с середины 2000-х началось активное продвижение концепции BIM в строительной отрасли, что заметно усилило интерес к параметрическому программному обеспечению. Популярность росла настолько стремительно, что строительный форк машиностроительной

Pro-E — параметрический CAD, продвигаемый этим вендором, — фактически вытеснил конкурентов в сегменте архитектурного и строительного проектирования (Рис. 3.1-20). К началу 2020-х годов он де-факто закрепил глобальное доминирование на рынке BIM (CAD) [62].

Рис. 3.1-20 Популярность поискового запроса в Google (RVT versus IFC): параметрический CAD, созданный бывшей командой Pro-E с поддержкой BOM-BIM набрал популярность практически в большинстве стран мира.

За прошедшие 20 лет аббревиатура BIM обросла множеством толкований, многозначность которых уходит корнями в начальные маркетинговые концепции, появившиеся в начале 2000-х годов. Стандарт ISO 19650, сыгравший важную роль в популяризации термина, фактически закрепил за BIM статус "научно обоснованного" подхода к управлению информацией. Однако в самом тексте стандарта, посвящённого управлению данными на протяжении жизненного цикла объектов с использованием BIM, аббревиатура BIM хоть и упоминается, но так и не получает чёткого определения.

На оригинальном сайте вендора, опубликовавшего серию Whitepaper по BIM в 2002 [60] и 2003 [61] годах, были фактически воспроизведены маркетинговые материалы по концепциям BOM (Bills of Materials) и PLM (Product Lifecycle Management), ранее применявшимся в машиностроительном ПО Pro-E ещё в 1990-х годах [63].

> *Информационное моделирование зданий - новый инновационный подход к проектированию, строительству и управлению зданиями, представленный компанией …… [название компании CAD вендора] в 2002 году, - изменил представление профессионалов отрасли во всем мире о том, как можно применять технологии в проектировании, строительстве и управлении зданиями.*
>
> — Whitepaper BIM, 2003 *[61]*

В этих ранних публикациях BIM напрямую связывался с понятием централизованной интегрированной базы данных. Как указывалось в Whitepaper 2003 года, BIM — это управление информацией о здании, при котором все обновления происходят в едином хранилище, обеспечивая синхронизацию всех чертежей, разрезов и спецификаций (BOM – Bills of Materials).

> *BIM описывается как управление информацией о здании, где все обновления и все изменения происходят в базе данных. Таким образом, имеешь ли ты дело со схемами, разрезами или чертежами на листах, все всегда скоординировано, согласовано и актуально.*
>
> — сайт компании CAD вендора с Whitepaper BIM, 2003 год *[54]*

Идея управления проектированием через единую интегрированную базу данных была широко обсуждаема ещё в исследованиях 1980-х годов. Например, концепт BDS Чарльза Истмана [57] включал 43 упоминания термина "база данных" (Рис. 6.1-2). К 2004 году в материалах по BIM это число сократилось почти вдвое — до 23 в Whitepaper 2002 года [64]. И уже к середине 2000-х тема баз данных практически исчезла из маркетинговых материалов вендоров и в целом и с повестки цифровизации.

Хотя именно база данных и доступ к ней изначально задумывались как ядро BIM-системы, со временем акцент сместился на геометрию, визуализацию и 3D. При этом сам регистратор стандарта IFC в 1994 году, опубликовавших Whitepaper BIM в 2002 году — тот же вендор — в Whitepaper начала 2000-х годов прямо указывал на ограниченность нейтральных форматов, таких как IGES, STEP и IFC и необходимость прямого доступа к базам данных CAD:

> *Различные приложения могут быть несовместимы, а повторно введенные данные могут быть неточными […]. Результат традиционного автоматизированного проектирования [CAD]: повышение стоимости, увеличение времени выхода на рынок и снижение качества продукции. Сегодня все основные приложения используют стандартные отраслевые интерфейсы для низкоуровневого обмена данными. Используя старые стандарты IGES или новые STEP [IFC это де-факто и де-юре копия STEP/IGES формата] для обмена данными между приложениями разных производителей, пользователи могут добиться определенной совместимости данных между лучшими в своем роде продуктами. Но IGES и STEP работают только на низких уровнях, и они не могут обмениваться такими богатыми данными, как информация, создаваемая современными ведущими приложениями […]. И хотя эти и другие стандарты совершенствуются практически ежедневно, они всегда будут отставать от продуктов современных производителей в плане богатства данных. […] программы в рамках приложения должны иметь возможность обмениваться данными и сохранять их богатство, не прибегая к нейтральным трансляторам, таким как IGES, STEP [IFC] или PATRAN. Вместо этого приложения фреймворка должны иметь возможность напрямую обращаться к основной базе данных CAD, чтобы не потерять детализацию и точность информации.*
>
> — Whitepaper CAD вендора (IFC, BIM) «Интегрированное проектирование и производство: Преимущества и обоснование», 2000 *[65]*

Таким образом, уже в 1980-х и начале 2000-х годов ключевым элементом цифрового проектирования в CAD-среде считалась база данных, а не формат-файл или нейтральный формат IFC. Предлагалось отказаться от трансляторов и обеспечить прямой доступ приложений к данным. Однако в реальности к середине 2020-х концепция BIM стала напоминать стратегию «разделяй и властвуй», где приоритетом остаются интересы поставщиков ПО, использующих закрытые геометрические ядра, а не развитие открытого информационного обмена.

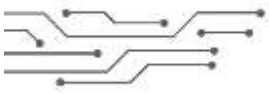
Сегодня BIM воспринимается как неотъемлемая часть строительной индустрии. Но за последние два десятилетия обещания упрощённого взаимодействия и интеграции данных во многом остались нереализованными. Большинство решений по-прежнему завязаны на закрытые форматы или нейтральные форматы, а также специализированные инструменты. Подробно вопросы истории появления BIM, open BIM и IFC, а также проблемы интероперабельности и геометрических ядер мы рассмотрим в шестой части книги «CAD и BIM: маркетинг, реальность и будущее проектных данных в строительстве».

> Сегодня перед отраслью стоит ключевой вызов — перейти от традиционного понимания CAD (BIM) как инструмента моделирования к его использованию в качестве полноценной базы данных. Это требует новых подходов к работе с информацией, отказа от зависимости закрытых экосистем и внедрения открытых решений.

С развитием инструментов обратного инжиниринга, позволяющих получить доступ к базам данных CAD, а также благодаря распространению Open Source и LLM-технологий, пользователи и разработчики в строительной отрасли всё чаще отходят от расплывчатых терминов поставщиков ПО. Вместо этого внимание смещается на то, что действительно важно: данные (базы данных) и процессы.

> За модными аббревиатурами и визуализациями скрываются стандартные практики управления данными: хранение, передача и трансформация — то есть классический процесс ETL (Extract, Transform, Load). Как и в других отраслях, цифровизация строительства требует не только стандартов обмена, но и чётко структурированной работы с разнородной информацией.

Для того чтобы в полной мере использовать потенциал CAD (BIM) данных, компаниям необходимо переосмыслить свой подход к управлению информацией. Это неизбежно приведёт к ключевому элементу цифровой трансформации — унификации, стандартизации и осмысленной структуризации данных, с которыми ежедневно работают специалисты строительной отрасли.



ГЛАВА 3.2.
УНИФИКАЦИЯ И СТРУКТУРИРОВАНИЕ ДАННЫХ

Наполнение систем данными в строительной отрасли

Будь то крупные корпорации или средние компании, специалисты ежедневно занимаются наполнением программных систем и баз данных с различными интерфейсами разноформатной информацией (Рис. 3.2-1), которая с помощью менеджеров должна слаженно взаимодействовать друг с другом. Именно этот комплекс взаимодействующих систем и процессов в итоге создаёт для компании доход и прибыль.

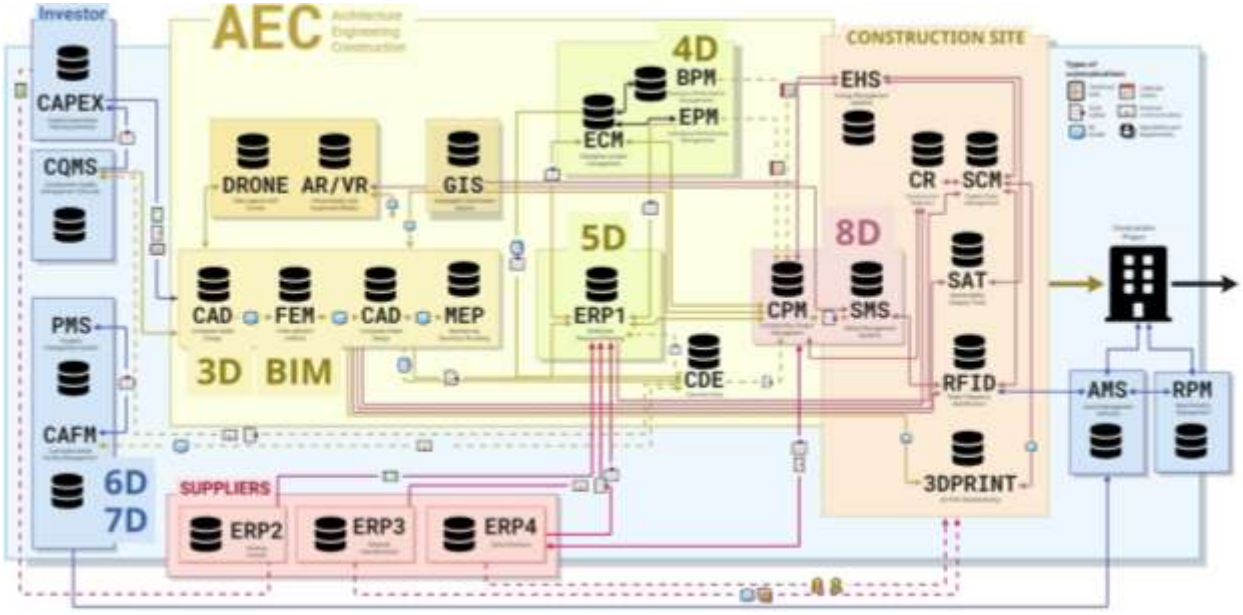


Рис. 3.2-1 Практически каждая система или приложение, используемое в строительном бизнесе, имеет в своей основе одну из популярных RDBMS баз данных.

Каждая из категорий систем, упомянутых ранее и применяемых в строительной отрасли, работает со своими типами данных, соответствующими функциональной роли этих систем. Чтобы двигаться от абстрактного уровня к конкретике, мы переходим от типов данных к их представлению в виде форматов и документов.

К ранее представленному перечню систем (Рис. 1.2-4) мы теперь добавим конкретные типы форматов и документов, с которыми они часто работают:

- **Инвестор (CAPEX)**
  - Финансовые данные: бюджеты, прогнозы расходов (структурированные данные).
  - Данные о тенденциях рынка: анализ рынка (структурированные и неструктурированные данные).
  - Юридические и договорные данные: контракты (текстовые данные).

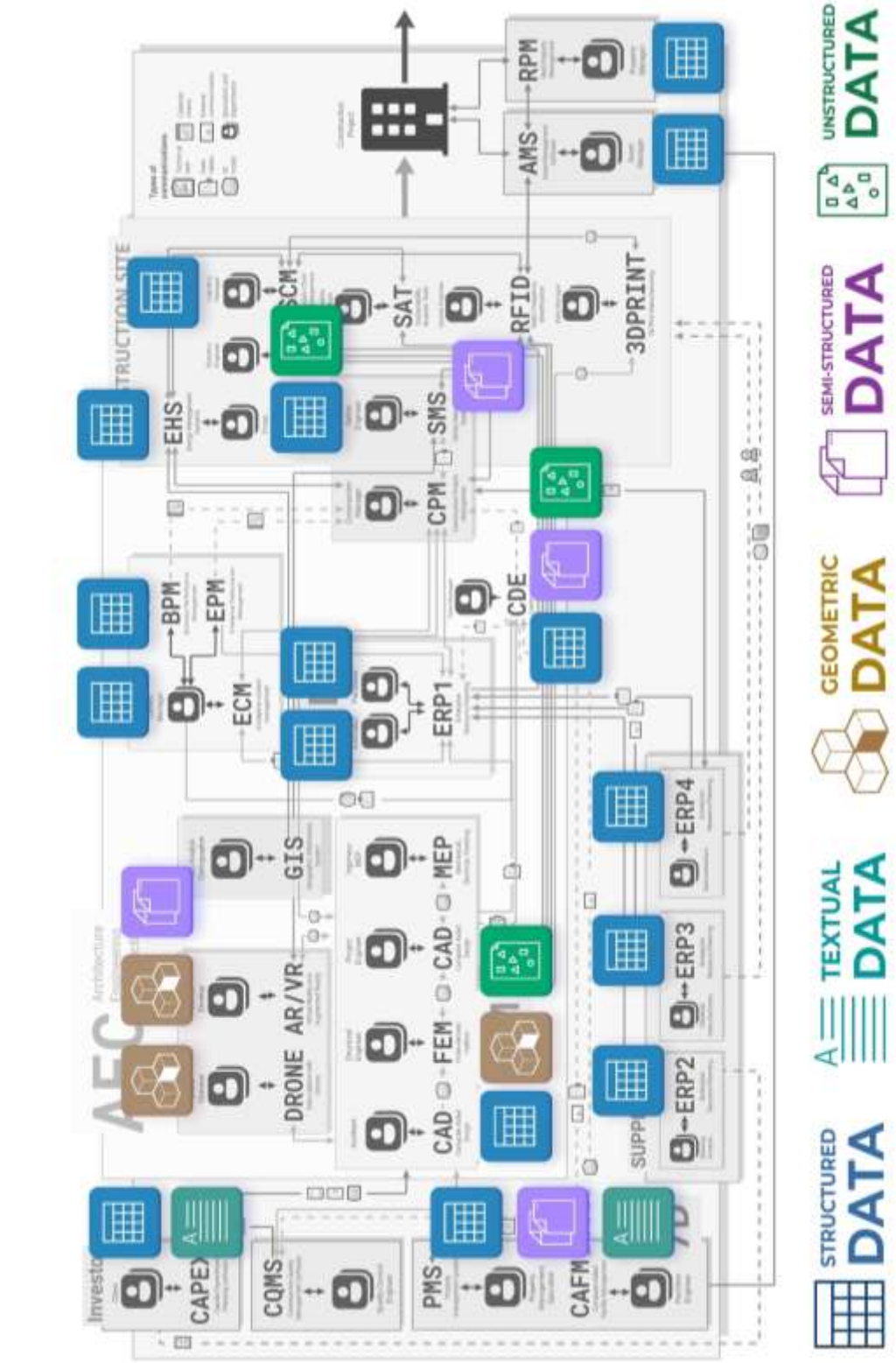


Рис. 3.2-2 В строительной отрасли используется множество систем с разными интерфейсами, которые работают с различными типами данных.

- **Системы управления (PMS, CAFM, CQMS)**
  - Данные проекта: графики, задачи (структурированные данные).
  - Данные о техническом обслуживании объекта: планы технического обслуживания (текстовые и полуструктурированные данные).
  - Данные контроля качества: стандарты, отчеты о проверках (текстовые и неструктурированные данные).
- **CAD, FEM и BIM**
  - Технические чертежи: архитектурные, структурные планы (геометрические данные, неструктурированные данные).
  - Модели зданий: 3D-модели, данные о материалах (геометрические и полуструктурированные данные).
  - Инженерные расчеты: анализ нагрузки (структурированные данные).
- **Системы управления строительной площадкой (EHS, SCM)**
  - Данные о безопасности и здоровье: протоколы безопасности (текстовые и структурированные данные).
  - Данные о цепочке поставок: запасы, заказы (структурированные данные).
  - Ежедневные отчеты: рабочее время, производительность (структурированные данные).
- **Беспилотники, AR/VR, ГИС, 3D-печать**
  - Геоданные: топографические карты (геометрические и структурированные данные).
  - Данные в реальном времени: видео и фотографии (неструктурированные данные).
  - Модели для 3D-печати: цифровые чертежи (геометрические данные).
- **Дополнительные системы управления (4D BPM, 5D ERP1)**
  - Данные о времени и затратах: графики, сметы (структурированные данные).
  - Управление изменениями: записи об изменениях в проекте (текстовые и структурированные данные).
  - Отчетность о результатах деятельности: показатели успеха (структурированные данные).
- **Интеграция данных и связь (CDE, RFID, AMS, RPM)**
  - Обмен данными: обмен документами, модели данных (структурированные и текстовые данные).
  - RFID и данные отслеживания: логистика, управление активами (структурированные данные).
  - Мониторинг и контроль: датчики на объектах (структурированные и неструктурированные данные).

Таким образом, каждая система в строительной отрасли — от систем управления строительной площадкой до баз данных эксплуатации— оперирует своим типом информации: структурированной, текстовой, геометрической и др.. "Ландшафт данных", с которым ежедневно приходится работать специалистам крайне разнообразен. Однако простое перечисление форматов не раскрывает всей сложности реальной работы с информацией.

На практике компании сталкиваются с тем, что данные, даже будучи полученными из систем, не готовы к использованию "как есть". Особенно это касается текстов, изображений, PDF-документов, CAD-файлов и других форматов, которые сложно анализировать стандартными средствами.

Именно поэтому следующим ключевым шагом становится трансформация данных — процесс, без которого невозможно эффективно автоматизировать обработку, анализ, визуализацию и принятие решений.

Трансформация данных: критический фундамент современного бизнес-анализа

Сегодня большинство компаний оказываются перед парадоксом: около 80% их повседневных процессов по-прежнему опираются на классические структурированные данные — привычные таблицы Excel и реляционные базы данных (RDBMS) [66]. Однако при этом 80% новой информации, поступающей в цифровую экосистему компаний, имеет неструктурированный или слабоструктурированный характер (Рис. 3.2-3) [52]. Это текст, графика, геометрия, изображения, CAD-модели, документация в PDF, аудио- и видеозаписи, электронная переписка и многое другое.

Более того, объём неструктурированных данных продолжает стремительно расти — ежегодный прирост оценивается в 55–65% [67]. Подобная динамика создаёт серьёзные сложности для интеграции новой информации в существующие бизнес-процессы. Игнорирование этого потока разноформатных данных ведёт к формированию информационных пробелов и к снижению управляемости всей цифровой среды компании.

Рис. 3.2-3 Ежегодный рост объема неструктурированных данных создает проблемы с интеграцией потоковой информации в бизнес-процессы.

> Игнорирование сложных неструктурированных и запутанных слабоструктурированных данных в процессах автоматизации может привести к значительным пробелам в информационном ландшафте компании. В современном мире неконтролируемого и лавинообразного движения информации компаниям необходимо применять гибридный подход к управлению данными, включающий эффективные методы работы со всеми типами данных.

Ключ к эффективному управлению данными лежит в организации, структуризации и классификации различных типов "Вавилона" данных (включая неструктурированные, текстовые и геометрические форматы, в структурированные или слабоструктурированные данные). Этот процесс преобразует хаотичное множество данных в организованные структуры для интеграции в системы, тем самым делая возможным принятие решений на их основе (Рис. 3.2-4).

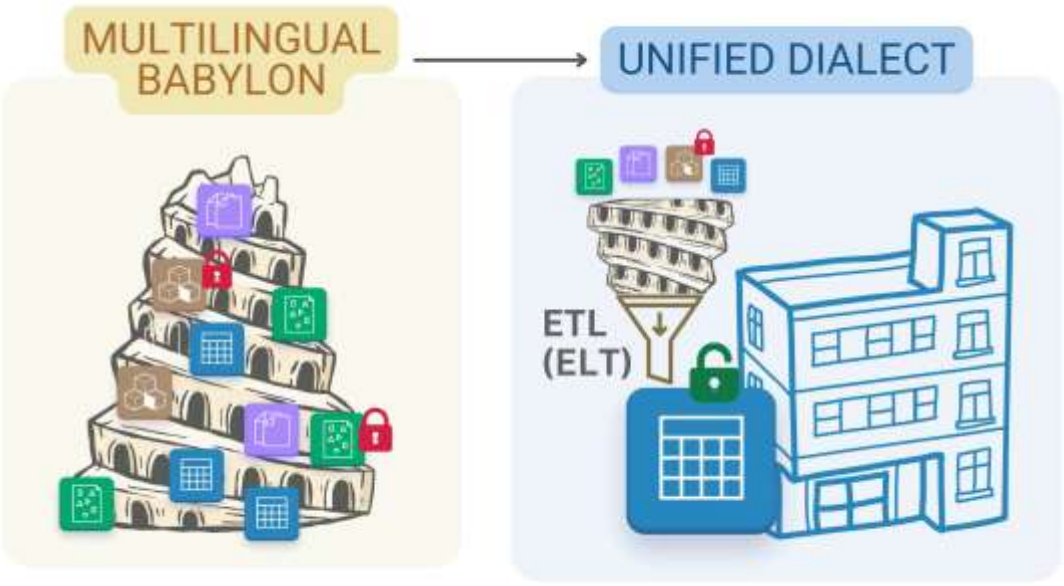


Рис. 3.2-4 Основная задача отделов управления данными - перевести «Вавилон» разнообразных и разноформатных данных в структурированную и классифицированную систему.

Одним из ключевых препятствий на пути к подобной унификации остаётся низкий уровень совместимости между различными цифровыми платформами - «силосами», про которые мы говорили в предыдущих главах.

> Согласно отчёту, национальный институт стандартов и технологий (NIST, США) подчеркивает [68], что низкая совместимость данных между различными строительными платформами ведет к потере информации и значительным дополнительным затратам. Только в 2002 году из-за проблем с совместимостью ПО потери в капитальном строительстве США составили 15,8 миллиарда долларов в год, где две трети этих потерь несут владельцы и операторы зданий, особенно во время эксплуатации и обслуживания [68]. В исследовании также отмечается, что стандартизация форматов данных может сократить эти потери и повысить эффективность работы на всех этапах жизненного цикла объекта.

Согласно исследованию CrowdFlower 2016 года [69], охватившему 16 тысяч специалистов по работе с данными по всему миру, основной проблемой остаются «грязные» и разноформатные данные. Согласно этому исследованию, самый ценный ресурс — это вовсе не итоговые базы данных и не модели машинного обучения, а время специалистов, которое тратится на подготовку информации.

На очистку, форматирование и организацию уходит до 60 процентов рабочего времени аналитика и менеджера по данным. Почти пятую часть занимает поиск и сбор нужных наборов данных, которые нередко скрыты в замкнутых хранилищах («силосах») и недоступны для анализа. И только около 9 процентов времени уходит непосредственно на моделирование, аналитику, построение прогнозов и тестирование гипотез. Всё остальное — это коммуникация, визуализация, отчётность и исследование вспомогательных источников информации.

В среднем работа менеджера по данным распределена следующим образом (Рис. 3.2-5):

- **Очистка и организация данных (60%):** наличие чистых и структурированных данных может значительно сократить время работы аналитика и ускорить процесс выполнения задач.
- **Сбор данных (19%):** основная сложность для профессионалов в области науки о данных заключается в поиске релевантных наборов данных. Часто данные компаний складываются в хаотично организованные "силосы", что усложняет доступ к нужной информации.
- **Моделирование/машинное обучение (9%):** нередко осложняются недостаточной ясностью бизнес-целей со стороны заказчиков. Отсутствие чёткой постановки задачи может свести на нет потенциал даже самой качественной модели.
- **Прочие задачи (5%):** кроме обработки данных, аналитикам приходится заниматься исследованиями, изучением данных с разных сторон, коммуникацией результатов через визуализации и отчёты, а также рекомендациями по оптимизации процессов и стратегий.

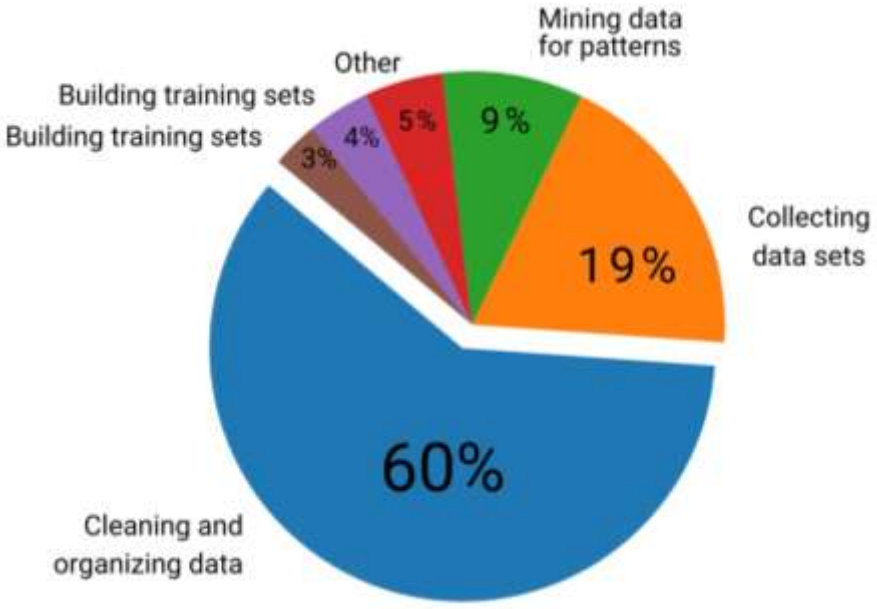


Рис. 3.2-5 На что менеджеры данных, работающие с данными, тратят больше всего времени (по материалам [70]).

> Эти оценки подтверждаются и другими исследованиями. Согласно исследованию Xplenty, опубликованному в BizReport в 2015 году [71], от 50% до 90% времени специалистов в области бизнес-аналитики (BI) уходит на подготовку данных для анализа.

Очистка, проверка и организация данных представляют собой критически важный фундамент для всех последующих процессов обработки данных и аналитик, занимая до 90% времени специалистов по работе с данными.

Этот кропотливый труд, невидимый для конечного пользователя, имеет решающее значение. Ошибки в исходных данных неизбежно искажают результаты анализа, вводят в заблуждение и могут привести к дорогостоящим управленческим ошибкам. Именно поэтому процессы очистки и стандартизации данных — от устранения дубликатов и заполнения пропусков до согласования единиц измерения и приведения к общей модели — становятся краеугольным камнем современной цифровой стратегии.

Таким образом, тщательная трансформация, очистка и стандартизация данных не просто занимают большую часть времени специалистов (до 80% работы с данными), но и определяют возможность их эффективного использования в рамках современных бизнес-процессов. Однако сама по себе организация и очистка данных не исчерпывают задачу оптимального управления информационными потоками компании. Во время этапа организации и структурирования становится выбор подходящей модели данных, которая напрямую влияет на удобство и эффективность работы с информацией в последующих этапах обработки.

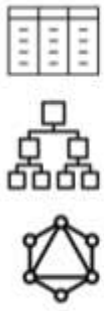
Поскольку данные и цели бизнеса различны, важно понимать особенности моделей данных и уметь выбирать или создавать нужную структуру. В зависимости от степени структурированности и способа описания связей между элементами, выделяют три основные модели: структурированные, слабоструктурированные и графовые. Каждая подходит для разных задач и имеет свои сильные и слабые стороны.

Модели данных: отношение в данных и связи между элементами

Данные в информационных системах организуются по-разному — в зависимости от задач и требований к хранению, обработке и передаче информации. Ключевое различие между типами моделей данных, форме в которой хранится информация, заключается в степени структурированности и способе описания взаимосвязей между элементами.

> Структурированные данные имеют чёткую и повторяющуюся схему: они организованы в виде таблиц с фиксированными столбцами. Такой формат обеспечивает предсказуемость, простоту обработки и эффективность при выполнении SQL-запросов, фильтрации и агрегации. Примеры — базы данных (RDBMS), Excel, CSV.

Слабоструктурированные данные допускают гибкую структуру: разные элементы могут содержать разный набор атрибутов и храниться в виде иерархий. Примеры — JSON, XML или другие документные форматы. Эти данные удобны, когда требуется моделировать вложенные объекты и отношения между ними, но с другой стороны, усложняет анализ и стандартизацию данных (Рис. 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

Рис. 3.2-6 Модель данных — это логическая структура, описывающая, как данные организованы, хранятся и обрабатываются в системе.

Выбор подходящего формата зависит от задач:

- Если важна скорость фильтрации и аналитики — подойдут реляционные таблицы (SQL, CSV, RDBMS, колончатые базы данных)
- Если необходима гибкость структуры — лучше использовать JSON или XML.
- Если данные имеют сложные взаимосвязи — графовые базы данных обеспечивают наглядность и масштабируемость.

В классических реляционных базах данных (RDBMS) каждая сущность (например, дверь) представлена строкой, а её свойства — столбцами таблицы. Например, таблица элементов из категории «Двери» может содержать поля ID, Высота, Ширина, Огнестойкость и Room ID, указывающий на помещение (Рис. 3.2-7).

В классических реляционных базах данных (RDBMS) отношения формируются в виде таблиц, где каждая запись представляет объект, а столбцы — его параметры. В табличном формате данные о дверях в проекте выглядят так, где каждая строка представляет отдельный элемент - дверь с её уникальным идентификатором и атрибутами, а связь с помещением осуществляется через параметр «Room ID».

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

Рис. 3.2-7 Информация о трёх элементах категории «Двери» проекта в табличной структурированной форме.

В слабоструктурированных форматах, таких как JSON или XML, данные хранятся в иерархическом или вложенном виде, где элементы могут содержать в себе другие объекты, а их структура может варьироваться. Это позволяет моделировать сложные связи между элементами. Аналогичная информация о дверях в проекте, которая была записана в структурированном виде (Рис. 3.2-7), в слабоструктурированном формате (JSON) представляется таким образом (Рис. 3.2-8), что они становятся вложенными объектами внутри помещений (Rooms – ID), что логично отражает иерархию.

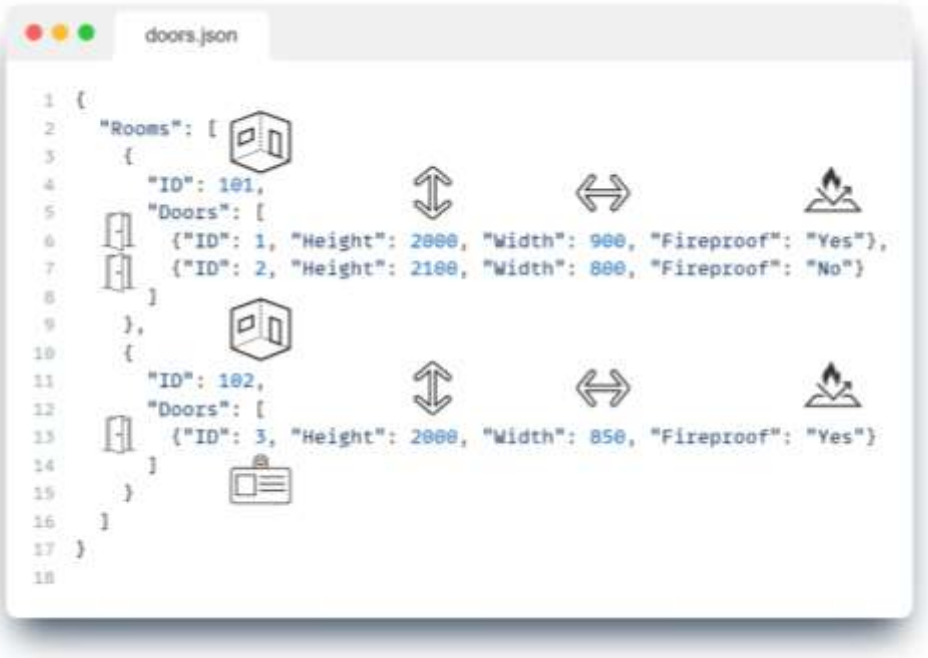


Рис. 3.2-8 Информация об элементах категории «Двери» проекта в формате JSON.

В графовой модели данные представлены в виде узлов (вершин) и связей (рёбер) между ними. Это позволяет наглядно отображать сложные взаимоотношения между объектами и их атрибутами. В случае с данными о дверях и помещениях в проекте, графовое представление выглядит следующим образом:

- **Узлы (вершины)** представляют основные сущности: помещения (Room 101, Room 102) и двери (ID1001, ID1002, ID1003)
- **Рёбра (связи)** показывают отношения между этими сущностями, например, принадлежность двери определённому помещению
- **Атрибуты** привязаны к узлам и содержат свойства сущностей (высота, ширина, огнестойкость для дверей)

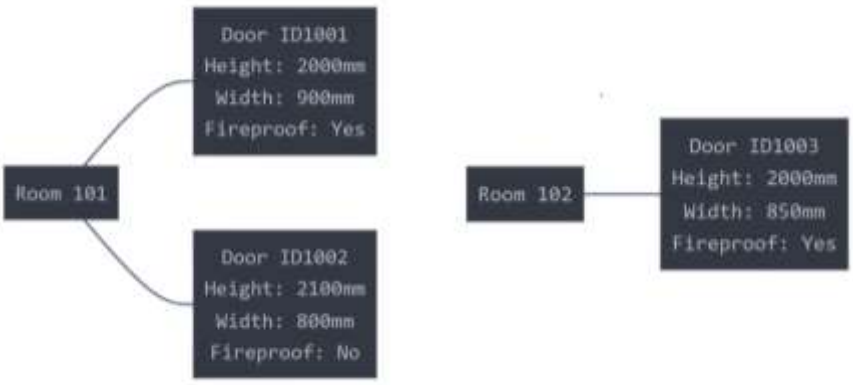


Рис. 3.2-9 Информация о сущностях дверей проекта в графовом представлении.

В графовой модели данных описания дверей, каждое помещение и каждая дверь являются отдельными узлами. Двери связаны с помещениями через рёбра, которые указывают на принадлежность двери к определённому помещению. При этом атрибуты дверей (высота, ширина, огнестойкость) хранятся, как свойства соответствующих узлов. Подробнее про графовые форматы и том, как графовая семантика появилась в строительной отрасли, поговорим в главе «Появление в строительстве темы семантики и онтологии».

Графовые базы данных эффективны в случаях, когда важны не столько сами данные, сколько отношения между ними, например, в рекомендательных системах, системах маршрутизации или при моделировании сложных взаимосвязей в проектах управления объектами. Графовый формат упрощает создание новых связей, позволяя добавлять в граф новые типы данных без изменения структуры хранилища. Однако, по сравнению с реляционными таблицами и структурированными форматами, в графе нет дополнительной связанности данных — перевод данных двумерных баз в граф не увеличивает количество связей и не позволяет получить новую информацию.

> Форма и схема данных должны соответствовать конкретному кейсу использования и решаемым задачам. Для эффективной работы в бизнес-процессах важно использовать те инструменты и те модели данных, которые помогают максимально быстро и просто получить результат.

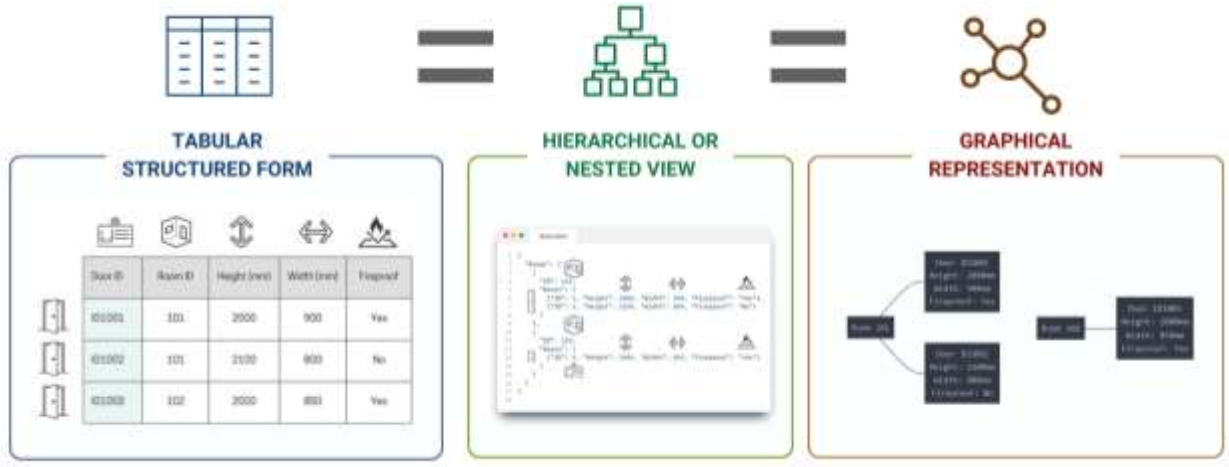


Рис. 3.2-10 Одна и та же информация об элементах проекта может храниться в разных форматах при помощи различных моделей данных.

> Сегодня большинство крупных компаний сталкиваются с проблемой избыточной сложности данных. Каждое из сотен или тысяч приложений использует собственную модель данных, что создает излишнюю комплексность — отдельная модель часто в десятки раз сложнее необходимого, а совокупность всех моделей — в тысячи раз. Такое избыточное усложнение существенно затрудняет работу как разработчиков, так и конечных пользователей.

Такое усложнение накладывает серьёзные ограничения на развитие и сопровождение систем компании. Каждый новый элемент в модели требует дополнительного кода, внедрения новых логик, тщательного тестирования и адаптации под уже существующие решения. Все это увеличивает затраты и замедляет работу команды по автоматизации в компании, превращая даже простые задачи в дорогостоящие и трудоёмкие процессы.

Усложнение затрагивает все уровни архитектуры данных. В реляционных базах данных это выражается в разрастании числа таблиц и столбцов, зачастую избыточных. В объектно-ориентированных системах сложность увеличивается за счёт множества классов и взаимосвязанных свойств. В форматах типа XML или JSON громоздкость проявляется через запутанные вложенные структуры, уникальные ключи и непоследовательные схемы.

> Избыточная сложность моделей данных делает системы не только менее эффективными, но и трудными для восприятия конечными пользователями и в будущем большими языковыми моделями и агентами LLM. Именно проблема понимания и сложность моделей данных и их обработки поднимает вопрос: как сделать так чтобы данные можно было достаточно просто использовать, чтобы действительно начали приносить быстро пользу.

Даже при грамотном выборе моделей данных их практическая польза резко снижается, если доступ к данным ограничен. Проприетарные форматы и закрытые платформы мешают интеграции, усложняют автоматизацию и лишают компании контроля над собственной информацией создавая не просто силос новых данных, а запертый силос, ключ к которому можно получить только по разрешению вендора. Чтобы понять масштаб проблемы, важно рассмотреть, как именно закрытость систем влияет на цифровые процессы в строительстве.

Проприетарные форматы и их влияние на цифровые процессы

Одной из ключевых проблем, с которыми сталкиваются строительные компании в ходе цифровизации, является ограниченный доступ к данным. Это затрудняет интеграцию систем, снижает качество информации и усложняет организацию эффективных процессов. В основе этих трудностей зачастую лежит использование проприетарных форматов и закрытых программных решений.

К сожалению, до сих пор многие программы, используемые в строительной отрасли, позволяют пользователю сохранять данные исключительно в собственных форматах или облачных хранилищах, доступ к которым возможен только через строго ограниченные интерфейсы. При этом нередко такие решения строятся в зависимости от ещё более закрытых систем более крупных вендоров. В результате даже те разработчики, кто хотел бы предложить более открытые архитектуры, вынуждены соблюдать правила, диктуемые крупными вендорами.

> В то время как современные системы управления строительными данными всё чаще поддерживают открытые форматы и стандарты (Рис. 3.1-5), базы данных CAD- (BIM)-средств, а также связанные с ними ERP- и CAFM-системы остаются изолированными проприетарными "островками" в цифровом ландшафте отрасли (Рис. 3.2-11).

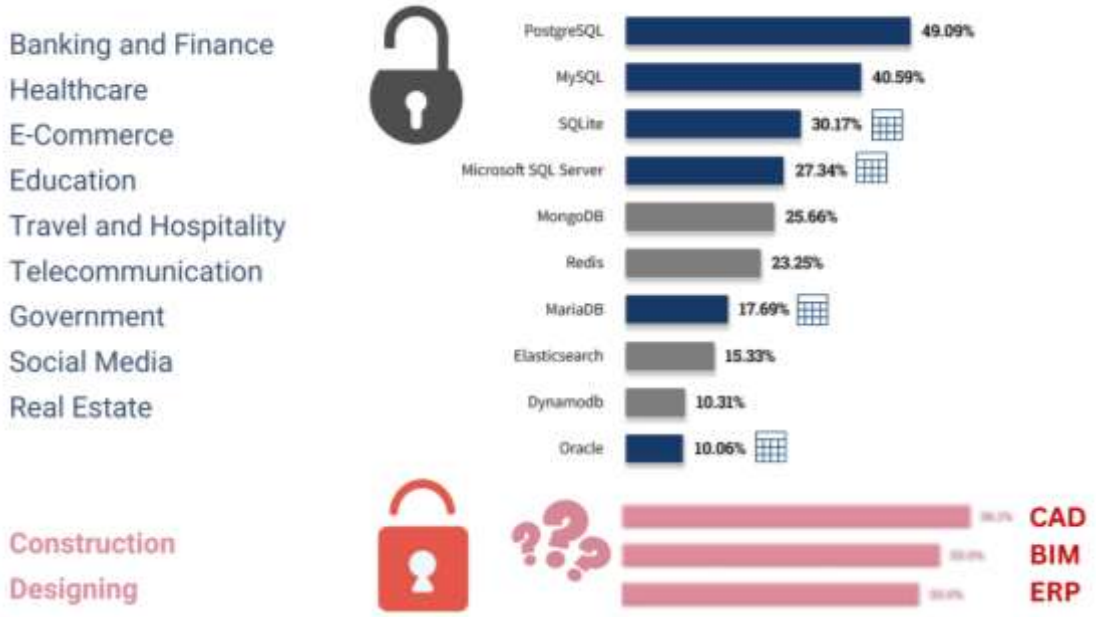


Рис. 3.2-11 Закрытый и проприетарный характер данных создает барьеры для интеграции и доступа к данным.

Закрытость и монополизация форматов и протоколов — это не только проблема строительной отрасли. Во многих секторах экономики борьба с закрытыми стандартами и ограниченным доступом к данным начиналась с замедления инноваций (Рис. 3.2-12), существованием искусственных барьеров для входа новых игроков и углублению зависимости от крупных поставщиков. На фоне стремительного роста значимости данных антимонопольные органы попросту не успевают реагировать на вызовы, связанные с новыми цифровыми рынками, и в итоге закрытые форматы и закрытый доступ к данным, по сути, становятся цифровыми "границами", сдерживающими движение информации и рост [63].

> *Если машины производят всё, что нам нужно, то наше положение будет зависеть от того, как эти блага распределяются. Каждый сможет наслаждаться жизнью в достатке, только если богатство, производимое машинами, станет общим достоянием. Либо большинство людей в итоге будут жить в ужасающей нищете, если владельцы машин смогут успешно лоббировать против перераспределения богатства. Пока что, по-видимому, все идет по второму варианту, технологии приводят ко все большему неравенству [72].*
>
> — Стивен Хокинг, астрофизик, 2015

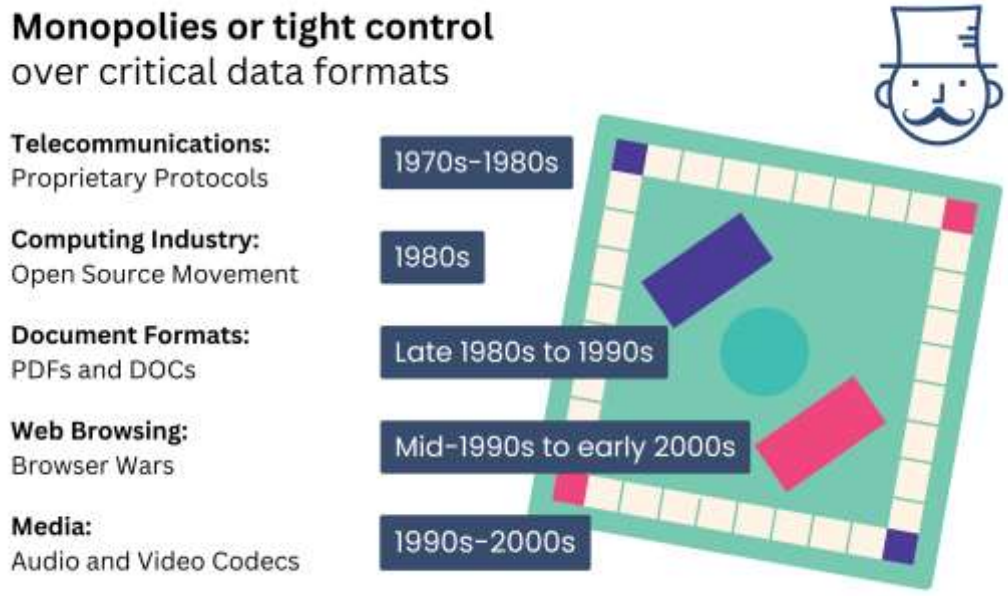


Рис. 3.2-12 Монопольное владение над ключевыми форматами и протоколами данных не является исключительной проблемой строительной отрасли.

В итоге из-за закрытого доступа к базам данных программ, менеджеры данных, аналитики, ИТ-специалисты и разработчики, создающих в строительной отрасли приложения для доступа к данным, их обработки и автоматизации, сегодня сталкиваются с многочисленными зависимостями от поставщиков программного обеспечения (Рис. 3.2-13). Эти зависимости в виде дополнительных уровней доступа требуют создания решений со специализированными API-соединениями и специальными инструментами и программным обеспечением.

> API (Application Programming Interface) — это формализованный интерфейс, с помощью которого одна программа может взаимодействовать с другой, обмениваясь данными и функциональностью без необходимости доступа к исходному коду. API описывает, какие запросы может делать внешняя система, в каком формате они должны быть, и какие ответы она получит. Это стандартизированный «контракт» между программными модулями.

Большое количество зависимостей от закрытых решений приводит к тому, что вся архитектура кода и логика бизнес-процессов в компании превращается в "спагетти-архитектуру" из инструментов, зависящих от политики поставщика программного обеспечения по предоставлению качественного доступа к данным.

Зависимость от закрытых решений и платформ приводит не только к потере гибкости, но и к реальным бизнес-рискам. Изменение условий лицензирования, закрытие доступа к данным, изменение форматов или структуры API — всё это может заблокировать критически важные процессы. Внезапно оказывается, что обновление одной таблицы требует переделки целого блока интеграций и коннекторов (Рис. 3.2-13), а любое масштабное обновление ПО или его API поставщика становится потенциальной угрозой для стабильности всей системы компании.

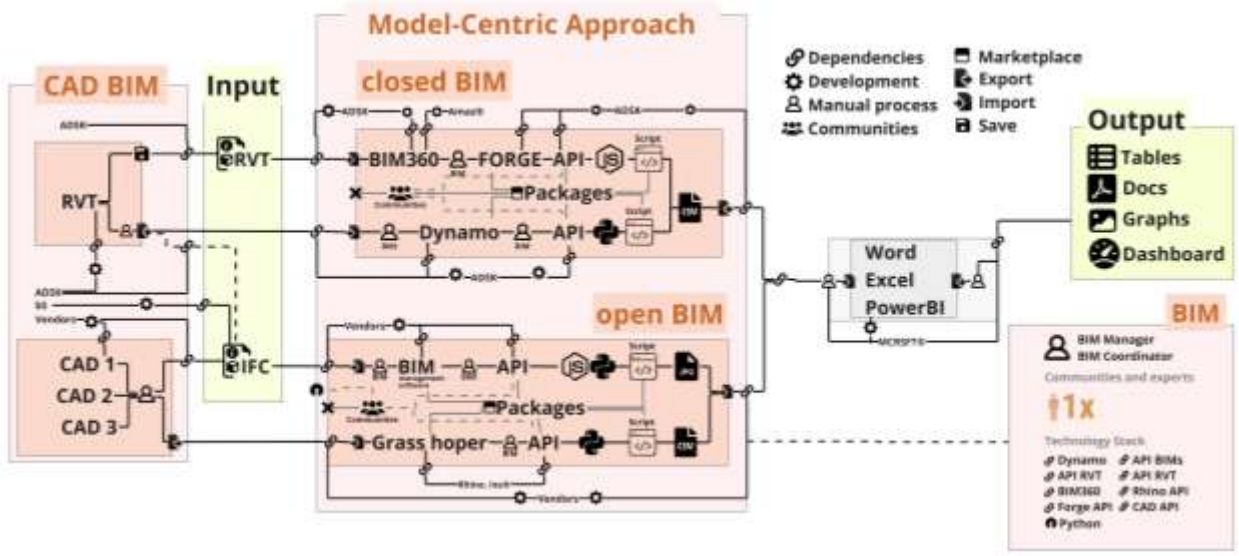


Рис. 3.2-13 Пример большого количества зависимостей при обработке CAD-данных создает препятствия для интеграции данных в экосистеме строительных компаний.

> Разработчики и системные архитекторы в таких условиях вынуждены работать не на опережение, а на выживание. Вместо внедрения новых решений — они адаптируются. Вместо развития — они пытаются поддерживают совместимость. Вместо того, чтобы автоматизировать и ускорять процессы, они тратят время на изучение очередных закрытых интерфейсов, API документации и бесконечную перестройку кода.

Работа с закрытыми форматами и системами — это не просто техническая проблема — это стратегическое ограничение. Несмотря на очевидные возможности, которые предоставляют современные технологии автоматизации, ИИ, LLM и предиктивной аналитики, многие компании не могут реализовать их потенциал в полной мере. А барьеры, воздвигнутые проприетарными форматами (Рис. 3.2-13), лишают бизнес доступа к собственным данным. И в этом, пожалуй, кроется главная ирония цифровой трансформации в строительстве.

> Прозрачность данных и открытость систем — не роскошь, а необходимое условие для скорости и эффективности. В отсутствие открытости отсутствие бизнес-процессы наполняются ненужной бюрократией, многослойными цепочками согласований и растущей зависимостью от принципа HiPPO — принятия решений на основе мнения самого высокооплачиваемого человека.

Тем не менее, на горизонте формируется сдвиг парадигмы. Несмотря на доминирование проприетарных решений, всё больше компаний осознают ограничения архитектур, созданных в духе Четвёртой промышленной революции. Сегодня вектор смещается в сторону принципов Пятой революции, где в центре находятся данные как стратегический актив, открытые интерфейсы (API) и настоящая интероперабельность между системами.

Этот переход знаменует отход от замкнутых экосистем в сторону гибких, модульных цифровых архитектур, где ключевую роль играют открытые форматы, стандарты и прозрачность обмена данными.

Открытые форматы меняют подход к цифровизации

Строительная отрасль подошла к проблеме закрытости и проприетарности данных одной из последних. В отличие от других сфер экономики, цифровизация здесь развивалась медленно. Причинами стали и традиционная консервативность отрасли, и преобладание разрозненных локальных решений, и глубокая укорененность бумажного документооборота. На протяжении десятилетий ключевые процессы в строительстве полагались на физические чертежи, телефонные звонки и несинхронизированные базы данных. В этом контексте закрытые форматы долгое время воспринимались как норма, а не как препятствие.

Опыт других отраслей демонстрирует: устранение барьеров, связанных с закрытыми данными, ведёт к всплеску инноваций, ускорению развития и росту конкуренции [73]. В науке обмен открытыми данными позволяет ускорять открытия и развивать международное сотрудничество. В медицине — повышать эффективность диагностики и лечения. В программной инженерии — создавать экосистемы совместного творчества и быстрого совершенствования продуктов.

> Согласно отчёту McKinsey «Открытые данные: Раскрой инновации и производительность с помощью информационных поток» 2013 г. [74], открытые данные способны разблокировать от $3 до $5 трлн ежегодно в семи ключевых отраслях экономики, включая строительство, транспорт, здравоохранение и энергетику. Согласно тому же исследованию, децентрализованные экосистемы данных позволяют крупным строительным компаниям и подрядчикам сократить затраты на разработку и поддержку ПО, ускоряя внедрение цифровых технологий.

Переход к открытым архитектурам, давно начавшийся в других секторах экономики, постепенно охватывает и строительную отрасль. Крупные компании и государственные заказчики, и в особенности финансовые организации, контролирующие инвестиции в строительные проекты, всё чаще выдвигают требования по использованию открытых данных и обеспечению доступа к исходному коду расчётов, калькуляций и приложений. Разработчики уже не просто должны создавать цифровые решения и показывать итоговые цифры проекта — от них ожидается прозрачность, воспроизводимость и независимость от вендоров сторонних приложений.

> Использование решений с открытым исходным кодом обеспечивает заказчика уверенностью в том, что даже если внешние разработчики прекратят сотрудничество или покинут проект, это не повлияет на возможность дальнейшего развития инструментов и систем. Одним из главных преимуществ открытых данных является их способность устранить зависимость разработчиков приложений от конкретных платформ для доступа к данным.

Если компания не может полностью отказаться от проприетарных решений, возможным компромиссом становится использование методов реверс инжиниринга. Эти легальные и технически

обоснованные методы позволяют преобразовывать закрытые форматы в более доступные, структурированные и пригодные для интеграции. Это особенно важно в тех случаях, когда требуется подключение к устаревшим системам или миграция информации из одного программного ландшафта в другой.

> Одним из ярких примеров в истории перехода к открытым форматам и применения реверс инжиниринга (легального взлома проприетарных систем) в строительстве является история борьбы за открытие формата DWG, широко используемого в системах автоматизированного проектирования (CAD). В 1998 году ответом на монополию одного из вендоров ПО, другими 15 вендорами CAD был создан новый альянс «Open DWG», целью которого стало предоставление разработчикам бесплатных и независимых инструментов для работы с форматом DWG (де-факто стандарта передачи чертежей) без необходимости использования проприетарного ПО или закрытых API. Это событие стало поворотным моментом, позволившим десяткам тысяч компаний получить свободный доступ к закрытому формату популярного, с конца 1980-х и до сегодняшних дней, CAD решения и создавать совместимые решения, которые способствовали развитию конкуренции на рынке CAD [75]. Сегодня «Open DWG» SDK, который впервые был создан ещё в 1996 г., используется в почти всех решениях в которых можно импортировать, редактировать и экспортировать DWG формат, вне официального приложения разработчика DWG формата.

Аналогичные трансформации вынужденно происходят и у других технологических гигантов. Microsoft, некогда символ проприетарного подхода, открыла исходный код .NET Framework, начала использовать Linux в инфраструктуре облачного сервиса Azure и приобрела GitHub, чтобы укрепить позиции в сообществе Open Source. [76]. Компания Meta (бывшая Facebook) выпустила модели ИИ с открытым исходным кодом, такие как серия Llama, чтобы способствовать инновациям и сотрудничеству в сфере разработки ИИ агентов. Генеральный директор Марк Цукерберг предполагает, что платформы с открытым исходным кодом будут лидировать в технологическом прогрессе в ближайшее десятилетие [77].

> Open Source — это модель разработки и распространения программного обеспечения, при которой исходный код открыт для свободного использования, изучения, модификации и распространения.

Открытые данные и решения с открытым исходным кодом становятся не просто трендом, а основой цифровой устойчивости. Они дают компаниям гибкость, устойчивость к изменениям, контроль над собственными решениями и возможность масштабировать цифровые процессы без зависимости от политики поставщиков. И, что не менее важно, они возвращают бизнесу контроль над самым ценным ресурсом XXI века — своими данными.

Смена парадигмы: Open Source как конец эры доминирования вендоров ПО

Строительную индустрию ждёт такой сдвиг, который невозможно монетизировать привычным способом. Концепция перехода на использование данных, дата-центричный подход и использова-

ние Open Source инструментов приводит к переосмыслению правил игры, на которых держатся гиганты ПО рынка.

> В отличие от предыдущих технологических трансформаций, этот переход не будет активно продвигаться вендорами. Смена парадигмы угрожает их традиционным бизнес-моделям, основанным на лицензировании, подписках и консалтинге. Новая реальность не предполагает готового продукта «в коробке» или платной подписки — она требует перестройки процессов и мышления.

Для управления и развития дата-центричных решений, основанных на открытых технологиях, компаниям потребуется пересмотр внутренних процессов. Специалисты из разных отделов должны будут не просто взаимодействовать, но и переосмыслить подходы к совместной работе.

Новая парадигма подразумевает использование открытых данных и Open Source-решений, где особую роль в создании программного кода начнут играть не программисты, а инструменты на базе искусственного интеллекта и больших языковых моделей (LLM). Уже к середине 2024 года более 25% нового кода в Google создаётся при помощи ИИ [78]. В будущем кодирование с LLM будет выполнять 80% работы всего за 20% времени (Рис. 3.2-14).

> Согласно исследованию McKinsey 2020 года [79], в сфере аналитики на смену CPU всё активнее приходят GPU — благодаря их высокой производительности и поддержке современными Open Source-инструментами. Это позволяет компаниям ускорить обработку данных без значительных инвестиций в дорогое ПО или найм дефицитных специалистов.

Ведущие консалтинговые компании, такие как McKinsey, PwC и Deloitte, подчеркивают растущую важность открытых стандартов, Open Source приложений в различных отраслях.

> Согласно отчёту PwC Open Source Monitor 2019 [80], 69% компаний с численностью сотрудников от 100 человек осознанно используют Open Source решения. Особенно активно OSS используется в крупных компаниях: 71% компаний с 200–499 сотрудниками, 78% в категории 500–1999 сотрудников и до 86% среди компаний с численностью более 2000 человек. Согласно отчёту Synopsys OSSRA за 2023 год, 96% проанализированных кодовых баз содержали компоненты с открытым исходным кодом [81].

Будущее роли разработчика — это не ручное написание кода, а проектирование моделей данных, архитектуры потоков и управление ИИ-агентами, создающими нужные расчёты по запросу. Пользовательские интерфейсы станут минималистичными, а взаимодействие — диалоговым. Классическое программирование уступит место высокоуровневому проектированию и оркестрации цифровых решений (Рис. 3.2-14). Современные тренды — такие как low-code платформы (Рис. 7.4-6) и экосистемы с поддержкой LLM (Рис. 7.4-4) — позволят существенно сократить издержки на разработку и сопровождение ИТ-систем.

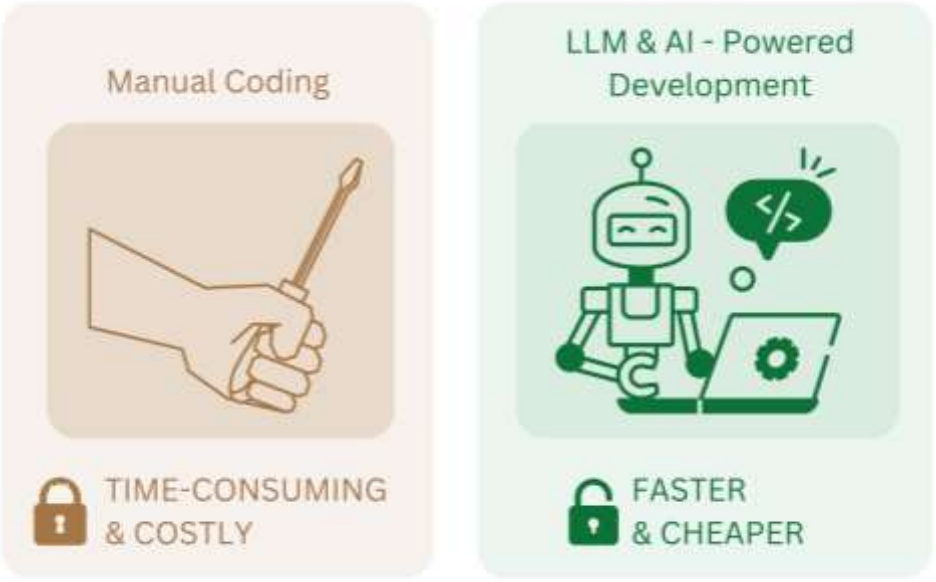


Рис. 3.2-14 Если сегодня приложения создаются вручную программистами, то в будущем значительная часть кода будет генерироваться решениями на базе ИИ и LLM.

Этот переход будет не похож на предыдущие и крупные поставщики ПО, скорее всего, не станут его катализаторами.

> Исследование Гарвардской бизнес-школы «Ценность программного обеспечения с открытым исходным кодом» 2024 года [40], совокупная стоимость открытого программного обеспечения оценивается с двух точек зрения. С одной стороны, если подсчитать, сколько потребовалось бы средств на создание всех существующих Open Source-решений с нуля, то эта сумма составила бы около 4,15 миллиарда долларов. С другой стороны, если представить, что каждая компания разрабатывает собственные аналоги Open Source-решений самостоятельно (что происходит повсеместно), не имея доступа к уже существующим инструментам, то совокупные затраты бизнеса на это достигли бы колоссальных 8,8 триллиона долларов — это стоимость спроса.

Нетрудно догадаться, что ни один крупный поставщик программного обеспечения не заинтересован в том, чтобы сократить рынок ПО с потенциальной стоимостью 8,8 триллиона долларов до всего 4,15 миллиарда. Это означало бы уменьшение объёма спроса в более чем 2 000 раз. Такая трансформация попросту нерентабельна для вендоров, чьи бизнес-модели строятся на многолетнем поддержании зависимости клиентов от закрытых решений. Поэтому компании, ожидающие, что им кто-то предложит удобное и открытое решение «под ключ», могут быть разочарованы — такие продавцы просто не появятся.

Переход к открытой цифровой архитектуре не означает сокращения рабочих мест или доходов. Напротив, он создаёт условия для гибких и адаптивных бизнес-моделей, которые со временем могут вытеснить традиционный рынок лицензий и коробочного программного обеспечения.

Вместо продажи лицензий — сервисы, вместо закрытых форматов — открытые платформы, вместо зависимости от вендора — независимость и возможность строить решения под реальные потребности. Те, кто раньше просто пользовался инструментами, смогут стать их соавторами. А те, кто умеет работать с данными, моделями, сценариями и логикой — окажутся в центре новой цифровой экономики отрасли. Подробнее об этих изменениях и о том, какие новые роли, бизнес-модели и форматы сотрудничества формируются вокруг открытых данных, мы поговорим в финальной, десятой части книги.

> Решения, основанные на открытых данных и открытом коде, позволят компаниям сфокусироваться не на борьбе с устаревшими API и интеграцией закрытых систем, а на эффективности бизнес-процессов. Осознанный переход к открытой архитектуре даёт возможность значительно повысить производительность и снизить зависимость от вендоров.

Переход к новой реальности — это не просто смена подходов к разработке программного обеспечения, но и переосмысление самого принципа работы с данными. В центре этой трансформации — не код, а информация: её структура, доступность и интерпретируемость. И именно здесь открытые и структурированные данные выходят на первый план, становясь неотъемлемой частью новой цифровой архитектуры.

Структурированные открытые данные: фундамент цифровой трансформации

Если в прошлые десятилетия устойчивость бизнеса во многом определялась выбором программных решений и зависимостью от конкретных вендоров, то в условиях современной цифровой экономики ключевым фактором становится качество данных и способность эффективно с ними работать. Открытый исходный код — важная часть новой технологической парадигмы, но его потенциал по-настоящему раскрывается лишь при наличии понятных, организованных и машиночитаемых данных. Среди всех типов моделей данных именно структурированные открытые данные становятся краеугольным камнем устойчивой цифровой трансформации.

Главное преимущество структурированных открытых данных — однозначная интерпретация и возможность автоматизированной обработки. Это позволяет существенно повысить эффективность как на уровне отдельных операций, так и в масштабах всей организации.

> Согласно отчету Deloitte "Процесс передачи данных при корпоративных преобразованиях" [82], работа с IT для управления передачей структурированных данных имеет решающее значение. Согласно отчету правительства Великобритании «Data Analytics and AI in Government Project Delivery» (2024) [83], устранение барьеров в обмене данными между различными проектами и организациями является ключевым фактором повышения эффективности в управлении проектами. В документе подчеркивается, что стандартизация форматов данных и внедрение принципов открытых данных позволяют избежать дублирования информации, минимизировать потери времени и повышать точность прогнозов.

Для строительной отрасли, где традиционно преобладает высокая степень фрагментации и разнообразие форматов, процесс структурирования-унификации и структурированные открытые данные

играют решающую роль в формировании согласованных и управляемых процессов (Рис. 4.1-14). Они позволяют участникам проекта сосредоточиться на повышении продуктивности, а не на решении технических проблем, связанных с несовместимостью закрытых платформ, моделей данных и форматов.

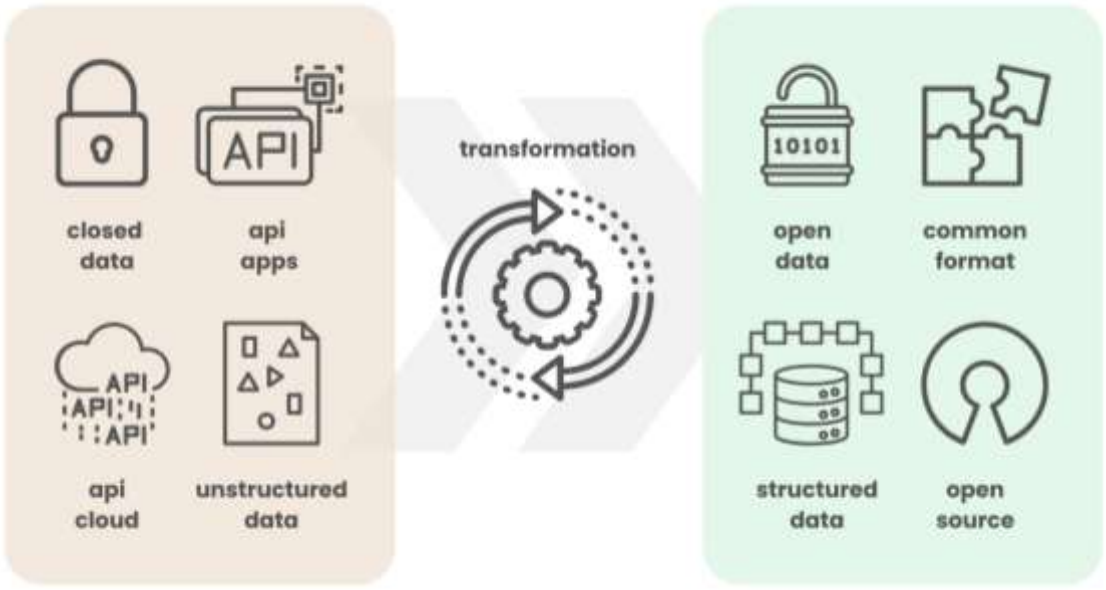


Рис. 3.2-15 Открытые структурированные данные снижают зависимость от программных решений и платформ и ускоряют инновации.

Современные технологии инструменты, которые мы будем далее рассматривать подробно в книге, позволяют не просто собирать информацию, но и автоматически её очищать: устранять дублирование, исправлять ошибки, нормализовать значения. Это означает, что аналитики и инженеры работают не с разрозненными документами, а с организованной базой знаний, пригодной для анализа, автоматизации и принятия решений.

> *Сделай настолько просто, насколько это возможно, но не проще.*
>
> — Альберт Эйнштейн, физик-теоретик (принадлежность цитаты оспаривается *[84]*)

Сегодня большинство пользовательских интерфейсов для работы с данными можно создавать автоматически — без необходимости вручную писать код для каждого бизнес кейса. Для этого необходим инфраструктурный слой, который без дополнительных инструкций понимает структуру данных, их модель и логику (Рис. 4.1-15). Именно структурированные данные делают возможным такой подход: формы, таблицы, фильтры и представления могут автоматически генерироваться с минимальными затратами на программирование.

Наиболее важные интерфейсы, критичные для пользователя, всё ещё могут требовать ручной доработки. Но в большинстве случаев — а это от 50 до 90 процентов рабочих сценариев — достаточно автоматической генерации приложений и расчётов без использования специальных приложений

для этого (Рис. 3.2-16), что существенно снижает затраты на разработку и сопровождение, уменьшает количество ошибок и ускоряет реализацию цифровых решений.

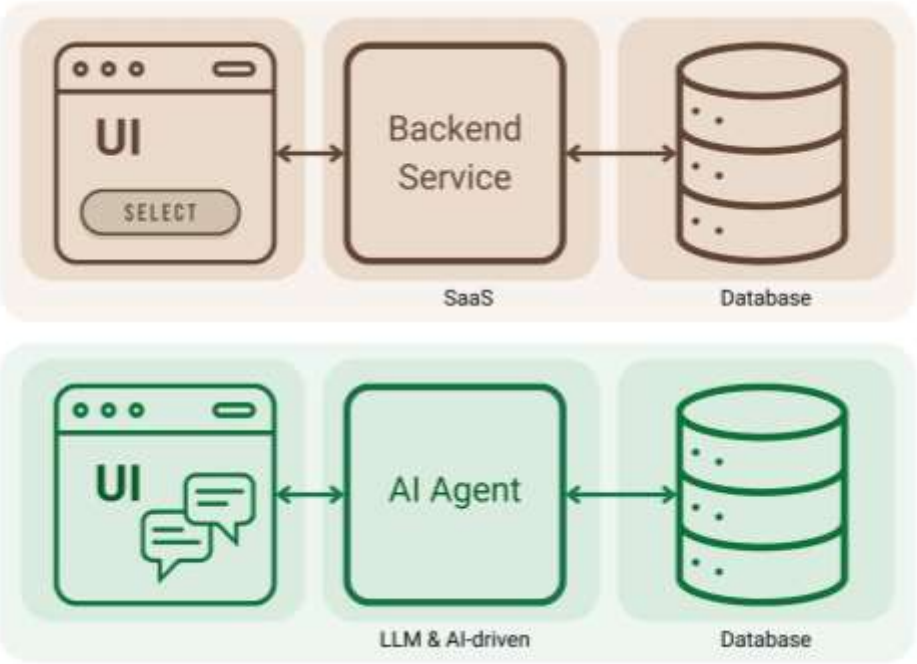
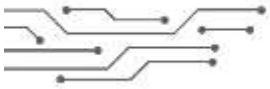


Рис. 3.2-16 Архитектурные модели работы с данными: традиционная архитектура приложений и AI-ориентированная модель с LLM.

Переход от архитектур, построенных на отдельных приложениях, к интеллектуально управляемым системам, опирающимся на языковые модели (LLM), — следующий шаг цифровой эволюции. В такой архитектуре структурированные данные становятся не только предметом хранения, но и основой взаимодействия с ИИ-инструментами, способными анализировать, интерпретировать и рекомендовать действия на основе контекста.

В последующих главах мы рассмотрим реальные примеры внедрения архитектуры, основанной на открытых структурированных данных, а также покажем, как языковые модели применяются для автоматической интерпретации, валидации и обработки данных. Эти практические кейсы помогут лучше понять, как новая цифровая логика работает в действии — и какие преимущества она даёт компаниям, готовым к трансформации.



ГЛАВА 3.3.
LLM И ИХ РОЛЬ В ОБРАБОТКЕ ДАННЫХ И БИЗНЕС-ПРОЦЕССАХ

LLM чаты: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok для автоматизации процессов обработки данных

Появление больших языковых моделей (LLM) стало естественным продолжением движения в сторону структурированных открытых данных и философии Open Source. Когда данные становятся организованными, доступными и машиночитаемыми, следующим шагом становится инструмент, способный взаимодействовать с этой информацией без необходимости писать сложный код или владеть специализированными техническими знаниями.

> LLM — это прямой продукт открытости: больших открытых датасетов, публикаций и движения Open Source. Без открытых научных статей, общедоступных текстовых данных и культуры совместной разработки не было бы ни ChatGPT, ни других LLM. LLM — это в каком-то смысле «дистиллят» накопленного человечеством цифрового знания, собранного и обученного благодаря принципам открытости.

Современные большие языковые модели (LLM — Large Language Models), такие как ChatGPT® (OpenAI), LlaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ предоставляют пользователям возможность формулировать запросы к данным на естественном языке. Это делает работу с информацией доступной не только для разработчиков, но и для аналитиков, инженеров, проектировщиков, менеджеров и других специалистов, ранее отдалённых от программирования.

> **LLM** (Large Language Model) — это искусственный интеллект, который обучен понимать и генерировать текст на основании огромного количества данных, собранных со всего интернета. Она способна анализировать контекст, отвечать на вопросы, вести диалог, писать тексты и генерировать программный код.

Если раньше визуализация, обработка или анализ данных требовали знаний специального языка программирования: Python, SQL, R или Scala, а также умения работать с библиотеками типа Pandas, Polars или DuckDB и многих других, то начиная с 2023 года ситуация изменилась радикально. Теперь пользователь может просто описать, что он хочет получить — и модель сама сгенерирует код, выполнит его, выведет таблицу или график и пояснит результат. Впервые за десятилетия развитие технологий пошло не по пути усложнения, а по пути радикального упрощения и доступности.

Этот принцип — «обрабатывай данные словами (промптами)» — ознаменовал собой новый этап в эволюции работы с информацией, фактически подняв создание решений на ещё более высокоуровневый уровень абстракции. Подобно тому, как в своё время пользователям стало не нужно разбираться в технических основах интернета, чтобы запускать онлайн-магазины или создавать сайты с помощью WordPress, Joomla и других модульных систем с открытым исходным кодом (ав-

тор книги с 2005 года работает с такими системами, в том числе в сфере образовательных и инженерных онлайн-платформ.) — что, в свою очередь, привело к бурному росту цифрового контента и интернет-бизнеса — сегодня инженеры, аналитики и менеджеры могут автоматизировать рабочие процессы без знания языков программирования. Этому способствуют мощные LLM — как бесплатные, так и с открытым исходным кодом, такие как LLaMA, Mistral, Qwen, DeepSeek и другие, — делающие передовые технологии доступными самой широкой аудитории.

Большие языковые модели LLM: как это работает

Большие языковые модели (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) представляют собой нейронные сети, обученные на огромных массивах текстовых данных из интернета, книг, статей и других источников. Их основная задача — понимать контекст человеческой речи и генерировать осмысленные ответы.

В основе современных LLM лежит архитектура трансформера, предложенная исследователями Google в 2017 году [85]. Ключевой компонент этой архитектуры — механизм внимания (attention), который позволяет модели учитывать взаимосвязи между словами вне зависимости от их позиции в тексте.

> Процесс обучения LLM отдалённо напоминает то, как человек осваивает язык — только в миллионы раз масштабнее. Модель анализирует миллиарды примеров употребления слов и выражений, выявляя закономерности в структуре языка и в логике смысловых переходов. При этом весь текст разбивается на токены — минимальные смысловые единицы (слова или их части), которые затем преобразуются в векторы в многомерном пространстве (Рис. 8.2-2). Эти векторные представления позволяют машине "понимать" скрытые взаимосвязи между понятиями, а не просто оперировать текстом как последовательностью символов.

Большие языковые модели — это не просто инструменты для генерации текста. Они умеют распознавать смысл, находить связи между понятиями и работать с данными, даже если они представлены в разных форматах. Главное — чтобы информация была разбита на понятные модели и представлена в виде токенов, с которыми LLM может работать.

Тот же подход можно применить и к строительным проектам. Если представить проект как своеобразный текст, где каждое здание, элемент или конструкция — это токен, то мы можем начать обрабатывать такую информацию похожим образом. Строительные проекты можно сравнить с книгами, которые разбиваются на категории, главы и группы абзацев, состоящих из минимальных токенов — элементов строительного проекта (Рис. 3.3-1). Переведя модели данных в структурированный формат, мы также сможем структурированные данные перевести в векторные базы (Рис. 8.2-2), которые являются идеальным источником для машинного обучения и таких технологий как LLM.

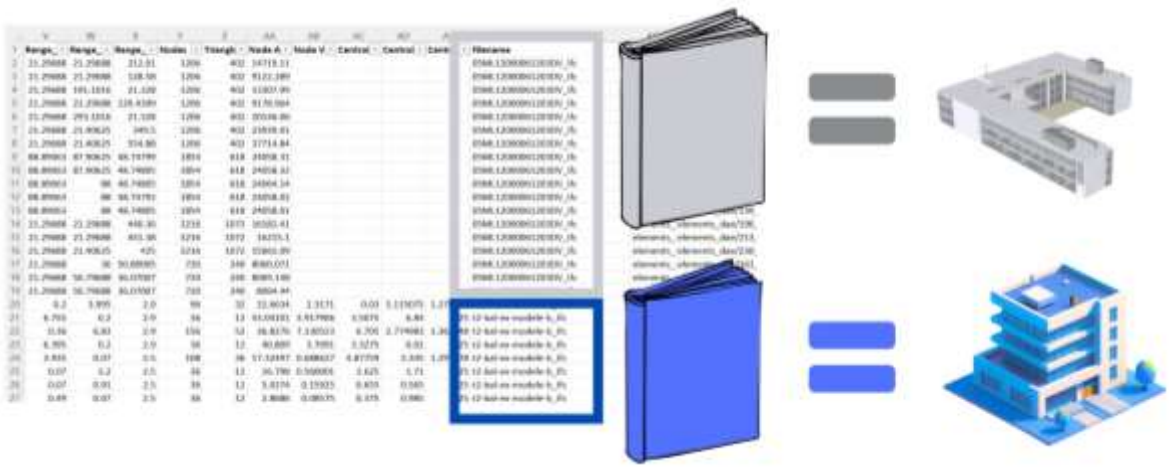


Рис. 3.3-1 Элемент строительного проекта — это как токен в тексте: минимальная единица, из которой формируются группы (абзацы) разделы (категории) всего проекта.

Если строительный проект оцифрован и его элементы представлены в виде токенов или векторов, то появляется возможность обращаться к ним не через жёсткие формальные запросы, а на естественном языке. Именно здесь проявляется одно из ключевых преимуществ LLM — умение понимать смысл запроса и связывать его с соответствующими данными.

Инженер больше не обязан писать SQL-запрос или код на Python для получения нужных данных — он может просто, понимая работу LLM и структуру данных, сформулировать задачу привычным способом: "*Найди все железобетонные конструкции с классом бетона выше B30 и посчитай их суммарный объём*." Модель распознает смысл запроса, превратит его в машиночитаемую форму, найдёт данные (сгруппирует и трансформирует) и возвратит итоговый результат.

Документы, таблицы, модели проектов преобразуются в векторные представления (эмбеддинг) и сохраняются в базе. Когда пользователь задает вопрос, запрос также преобразуется в вектор, и система находит наиболее близкие по смыслу данные. Это позволяет LLM опираться не только на свои обученные знания, но и на актуальные корпоративные данные, даже если они появились уже после окончания обучения модели.

> Одно из важнейших преимуществ LLM в строительстве — способность генерировать программный код. Вместо передачи технического задания программисту специалисты могут описать задачу на естественном языке, а модель создаст необходимый код, который можно использовать (скопировав его из чата) в создании кода автоматизации процессов. LLM-модели позволяют специалистам без глубоких знаний программирования внести свой вклад в автоматизацию и улучшение бизнес-процессов компании.

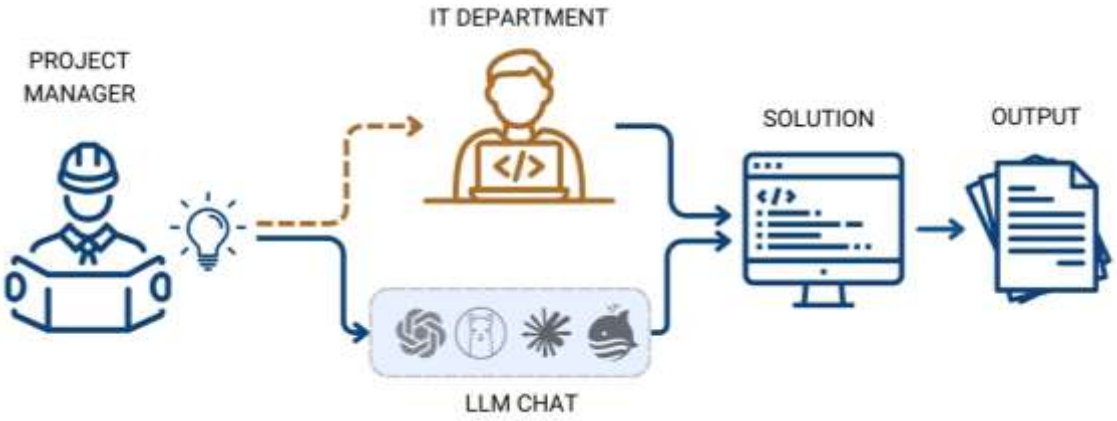


Рис. 3.3-2 LLM предоставляют возможность пользователям писать код и получать результаты без необходимости владеть навыками программирования.

> Согласно исследованию, проведённому Wakefield Research и спонсируемом SAP в 2024 году [36], в котором приняли участие 300 руководителей высшего звена в компаниях с годовым доходом не менее 1 млрд долларов в США: 52% руководителей высшего звена доверяют ИИ в вопросах анализа данных и предоставления рекомендаций для принятия решений. Ещё 48% используют ИИ для выявления ранее неучтённых рисков, а 47% — для предложения альтернативных планов. Кроме того, 40% применяют ИИ для разработки новых продуктов, бюджетного планирования и проведения маркетинговых исследований. Исследование также показало положительное влияние ИИ на личную жизнь: 39% респондентов отметили улучшение баланса между работой и личной жизнью, 38% сообщили об улучшении психического здоровья, а 31% — о снижении уровня стресса.

Однако при всей своей мощности LLM остаются инструментом, который важно использовать осознанно. Как и любая технология, они имеют ограничения. Одной из наиболее известных проблем являются так называемые «галлюцинации» — случаи, когда модель уверенно выдает правдоподобный, но фактически неверный ответ. Поэтому критически важно понимать, как устроена работа модели: какие данные и модели данных она может интерпретировать без ошибок, как интерпретирует запросы и откуда берёт информацию. Также стоит помнить, что знания LLM ограничены датой её обучения, и без подключения к внешним данным модель может не учитывать актуальные нормы, стандартны, цены или технологии.

Решением этих проблем становится регулярное обновление векторных баз, подключение к актуальным источникам и разработка автономных AI-агентов, которые не просто отвечают на вопросы, но и проактивно используют данные для обучения, управляют задачами, выявляют риски, предлагают варианты оптимизации и контролируют выполнение проекта.

Переход к LLM-интерфейсам в строительстве — это не просто технологическая новинка. Это сдвиг парадигмы, устранение барьеров между людьми и данными. Это возможность работать с информацией так же легко, как мы разговариваем друг с другом — и при этом получать точные, проверенные и готовые к действию результаты.

Те компании, которые начнут использовать такие инструменты раньше других, получат значительное конкурентное преимущество. Это и ускорение работы, и снижение затрат, и повышение качества проектных решений за счёт быстрого доступа к анализу данных и возможности быстро находить ответы на сложные вопросы. Но при этом необходимо учитывать и вопросы безопасности. Использование облачных LLM-сервисов может быть связано с рисками утечки данных. Поэтому всё чаще организации ищут альтернативные решения, позволяющие развернуть LLM-инструменты в собственной инфраструктуре — локально, с полной защитой и контролем над информацией.

Использование локальных LLM для чувствительных данных компании

Появление первых чат-LLM в 2022 году ознаменовало новый этап в развитии искусственного интеллекта. Однако сразу после широкого распространения этих моделей возник закономерный вопрос: насколько безопасно передавать данные и запросы, касающиеся компании, в облако? Большинство облачных языковых моделей сохраняли историю общения и загруженные документы на своих серверах и для компаний, работающих с конфиденциальной информацией, это стало серьёзным барьером на пути к внедрению ИИ.

Одним из наиболее устойчивых и логичных решений этой проблемы стало развёртывание Open Source LLM локально, внутри корпоративной IT-инфраструктуры. В отличие от облачных сервисов, локальные модели работают без подключения к интернету, не передают данные на внешние серверы и дают компаниям полный контроль над информацией.

> *Лучшая открытая модель [Open Source LLM] на сегодняшний день по производительности сопоставима с закрытыми моделями [таких как ChatGPT, Claude], но с отставанием примерно в один год [77].*
>
> – Бен Коттье, ведущий исследователь некоммерческой исследовательской организацией Epoch AI, 2024

Крупнейшие технологические компании начали делать свои LLM доступными для локального использования. Открытая серия LLaMA от Meta и стремительно развивающийся проект DeepSeek из Китая стали примерами перехода к открытой архитектуре. Наряду с ними Mistral и Falcon также выпустили мощные модели, свободные от ограничений проприетарных платформ. Эти инициативы не просто ускорили развитие глобального ИИ, но и дали компаниям, которым важна тема конфиденциальности, реальные альтернативы в вопросах независимости, гибкости и соблюдения норм безопасности.

В корпоративной среде, особенно в строительной отрасли, защита данных — вопрос не просто удобства, а регуляторного соответствия. Работа с тендерной документацией, сметами, чертежами и конфиденциальной перепиской требует строгого контроля. И здесь локальные LLM дают необходимую уверенность в том, что данные останутся внутри периметра компании.

	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

Рис. 3.3-3 Локальные модели обеспечивают полный контроль и безопасность, в то время как облачные решения предлагают удобную интеграцию и автоматические обновления.

Ключевые преимущества локальных Open Source LLM:

- Полный контроль над данными. Вся информация остается внутри компании, что исключает несанкционированный доступ и утечку данных.
- Автономная работа. Исключается зависимость от интернет-соединения, что особенно важно для работы в изолированных IT-инфраструктурах. Это также гарантирует бесперебойную работу в условиях санкций или блокировок облачных сервисов.
- Гибкость применения. Модель может использоваться для генерации текстов, анализа данных, написания программного кода, поддержки проектирования и управления бизнес-процессами.

- Адаптация под корпоративные задачи. LLM можно обучить на внутренних документах, что позволяет учитывать специфику работы компании и ее отраслевые особенности. Локальную LLM можно подключить к CRM, ERP или BI-платформам, что позволяет автоматизировать анализ клиентских запросов, создание отчетов или даже прогнозирование трендов.

> Развёртывание бесплатной и открытой модели DeepSeek-R1-7B на сервере, для доступа целой команды пользователей, стоимостью $1000 в месяц может потенциально обойтись дешевле, чем ежегодные платежи за облачные API, такие как ChatGPT или Claude и позволяет компаниям полностью контролировать данные, исключает их передачу в интернет и помогает соответствовать регуляторным требованиям, таким как GDPR.

В других отраслях локальные LLM уже меняют подход к автоматизации. В службах поддержки они отвечают на частые запросы клиентов, снижая нагрузку на операторов. В HR-отделах — анализируют резюме и подбирают релевантных кандидатов. В электронной коммерции — формируют персонализированные предложения, не раскрывая пользовательские данные.

В строительной сфере ожидается аналогичный эффект. Благодаря интеграции LLM с проектными данными и нормативами можно ускорить подготовку документации, автоматизировать составление смет и предиктивный анализ затрат. Особенно перспективным направлением становится использование LLM в связке со структурированными таблицами и датафреймами.

Полный контроль над ИИ в компании и как развернуть собственную LLM

Современные инструменты позволяют компаниям развернуть большую языковую модель (LLM) локально всего за несколько часов. Это даёт полный контроль над данными и инфраструктурой, устраняя зависимость от внешних облачных сервисов и минимизируя риски утечки информации. Такое решение особенно актуально для организаций, работающих с чувствительной проектной документацией или конфиденциальными коммерческими данными.

В зависимости от задач и ресурсов доступны разные сценарии развёртывания — от готовых решений "из коробки" до более гибких и масштабируемых архитектур. Один из самых простых инструментов — Ollama, который позволяет запускать языковые модели буквально в один клик, без необходимости в глубоких технических знаниях. Быстрый старт с Ollama:

1. Скачайте дистрибутив для вашей операционной системы (Windows / Linux / macOS) с официального сайта: ollama.com

2. Установите модель через командную строку. Например, для модели *Mistral*:

```
ollama run mistral
```

3. После запуска модель готова к работе — вы можете отправлять текстовые запросы через терминал или интегрировать её в другие инструменты. Запустить модель и выполнить запрос:

```
ollama run mistral “How to create a calculation with all the resources for the
work to install a 100mm wide plasterboard partition wall?“
```

Для тех, кто предпочитает работать в привычной визуальной среде, существует LM Studio — бесплатное приложение с интерфейсом, напоминающим ChatGPT:

- Установить LM Studio, скачав дистрибутив с официального сайта - lmstudio.ai
- Через встроенный каталог выберите модель (например, Falcon или GPT-Neo-X) и загрузите её
- Работайте с моделью через интуитивный интерфейс, напоминающий ChatGPT, но полностью локальный

	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

Рис. 3.3-4 Сравнение популярных локальных open source LLM-моделей.

Выбор модели зависит от требований к скорости, точности и доступным аппаратным возможностям (Рис. 3.3-4). Маленькие модели, такие как Mistral 7B и Baichuan 7B, подходят для лёгких задач и мобильных устройств, в то время как мощные модели, такие как DeepSeek-V3, требуют значительных вычислительных ресурсов, но обеспечивают высокую производительность и поддержку множества языков. В ближайшие годы рынок LLM будет стремительно развиваться — мы увидим всё больше лёгких и специализированных моделей. Вместо универсальных LLM, охватывающих весь

человеческий контент, появятся модели, обученные на узкопрофильной экспертизе. Например, можно ожидать появления моделей, предназначенных исключительно для работы с инженерными калькуляциями, строительными сметами или данными в CAD-форматах. Такие специализированные модели будут быстрее, точнее и безопаснее в использовании — особенно в профессиональной среде, где важна высокая надёжность и предметная глубина.

После запуска локальной LLM её можно адаптировать под конкретные задачи компании. Для этого применяется техника дообучения (fine-tuning), при которой модель проходит дополнительную тренировку на внутренних документах, технических инструкциях, шаблонах договоров или проектной документации.

RAG: Интеллектуальные LLM-ассистенты с доступом к корпоративным данным

Следующий этап эволюции применения LLM в бизнесе — это интеграция моделей с актуальными корпоративными данными в реальном времени. Такой подход называется RAG (Retrieval-Augmented Generation) — генерация с поддержкой извлечения. В этой архитектуре языковая модель становится не просто диалоговым интерфейсом, а полноценным интеллектуальным ассистентом, способным ориентироваться в документах, чертежах, базах данных и давать точные, контекстуальные ответы.

> Главное преимущество RAG — возможность использовать внутренние данные компании без необходимости дообучения модели, сохраняя при этом высокую точность и гибкость в работе с информацией.

Технология RAG объединяет два основных компонента:

- **Извлечение информации** (Retrieval): модель подключается к хранилищам данных — документам, таблицам, PDF-файлам, чертежам — и извлекает релевантную информацию по запросу пользователя.
- **Генерация ответа** (Augmented Generation): на основе извлечённых данных модель формирует точный, обоснованный ответ, учитывая контекст и специфику запроса.

Для того чтобы запустить LLM с поддержкой RAG, необходимо выполнить несколько шагов:

- **Подготовка данных**: соберите нужные документы, чертежи, спецификации, таблицы. Они могут находиться в разных форматах и структурах, от PDF до Excel.
- **Индексация и векторизация**: с помощью таких инструментов, как LlamaIndex или LangChain, данные преобразуются в векторные представления, которые позволяют находить смысловые связи между фрагментами текста (про векторные базы данных и перевод больших массивов в векторное представление, в том числе CAD проекта, подробнее в 8 части).
- **Запрос к ассистенту**: после загрузки данных можно задавать модели вопросы, и она будет

искать ответы в рамках корпоративной базы, а не в общих знаниях, собранных с интернета.

Допустим, в компании есть папка constructionsite_docs, где хранятся договоры, инструкции, сметы и таблицы. С помощью Python-скрипта (Рис. 3.3-5), можно просканировать эту папку и построить векторную индексацию: каждый документ будет преобразован в набор векторов, отражающих смысловое содержание текста. Это превращает документы в своеобразную «карту смыслов», по которой модель может эффективно ориентироваться и находить связи между терминами и фразами.

Например, модель «запоминает», что слова «возврат» и «рекламация» часто встречаются в разделе договора, касающемся отгрузки материалов на строительную площадку. После этого, если задать вопрос — например, «Какой у нас срок возврата товара?» (Рис. 3.3-5 - 11 строка кода) — LLM проанализирует внутренние документы и найдёт точную информацию, действуя как интеллектуальный помощник, способный читать и понимать содержимое всех корпоративных файлов.

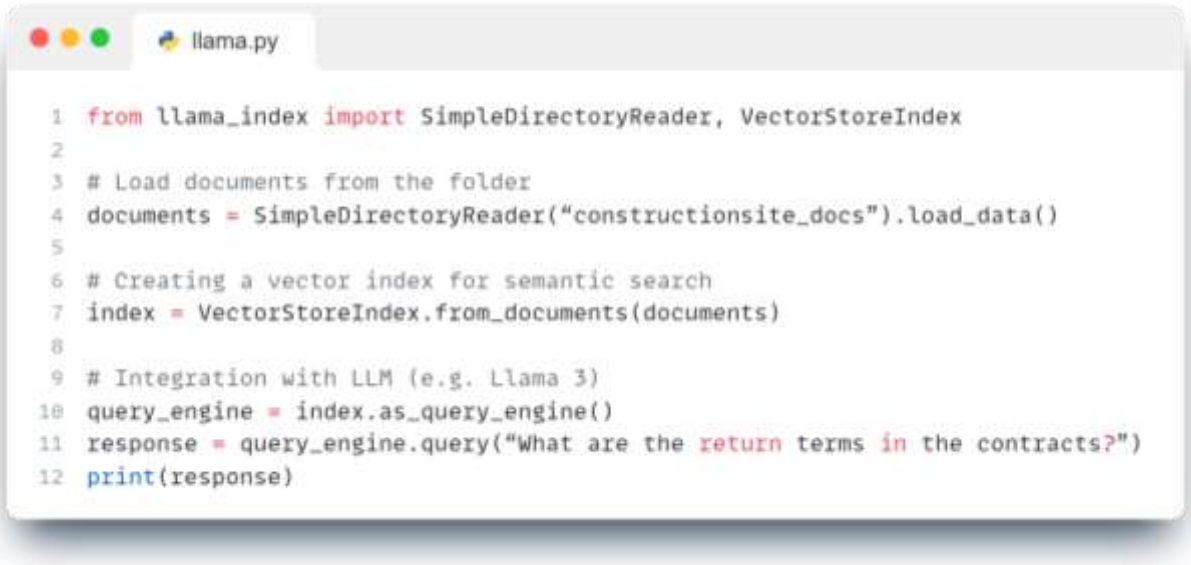


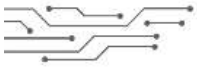
Рис. 3.3-5 LM читает папку с файлами — аналогично тому, как человек открывает её и ищет нужный документ.

Код можно запустить на любом компьютере с установленным Python. Подробнее про использование Python и IDE для запуска кода мы поговорим в следующе главе.

Локальное развертывание LLM — это не просто тренд, а стратегическое решение для компаний, которые ценят безопасность и гибкость. Однако развертывание LLM, будь то на локальных компьютерах компании или при использовании онлайн решений — это лишь первый шаг. Для того, чтобы применять возможности LLM в реальных задачах, компании должны использовать инструменты, которые позволяют не только получать ответы в чатах, но и сохранять созданную логику в виде кода, который можно будет запускать вне контекста использования LLM. Это важно для масштабирования решений - правильно организованные процессы позволяют применять наработки ИИ сразу на нескольких проектах или даже во всей компании.

В этом контексте выбор подходящей среды разработки (IDE) играет важную роль. Современные

инструменты программирования позволяют не только разрабатывать решения на основе LLM, но и интегрировать их в существующие бизнес-процессы, превращая их в автоматические ETL-Pipeline.



ГЛАВА 3.4.
IDE С ПОДДЕРЖКОЙ LLM И БУДУЩИЕ ИЗМЕНЕНИЯ В ПРОГРАММИРОВАНИИ

Выбор IDE: от LLM экспериментов к бизнес-решениям

Погружаясь в мир автоматизации, анализа данных и искусственного интеллекта — особенно при работе с большими языковыми моделями (LLM) — крайне важно выбрать подходящую интегрированную среду разработки (IDE). Именно она станет вашим основным рабочим инструментом: местом, где будет запускаться код, сгенерированный LLM, — как на локальном компьютере, так и внутри корпоративной сети. От выбора IDE зависит не только удобство работы, но и то, насколько быстро вы сможете перейти от экспериментальных запросов в LLM к полноценным решениям, встроенным в реальные бизнес-процессы.

> **IDE** (Интегрированная Среда Разработки) — это универсальный строительный комбинат на вашем компьютере для автоматизации процессов и обработки данных. Вместо того, чтобы хранить отдельно пилу, молоток, дрель и другие инструменты, у вас есть одно устройство, которое может всё — резать, крепить, сверлить и даже проверять качество материалов. IDE для программистов - это единое пространство, где можно писать код (в аналогии со строительством - создавать чертежи), тестировать его работу (собирание модели здания), находить ошибки (как контроль прочность конструкций в строительстве) и запускать готовый проект (сдача дома в эксплуатацию).

Обзор популярных IDE:

- **PyCharm**® (JetBrains) — это мощная профессиональная IDE для Python. Она отлично подходит для серьезных проектов благодаря большому количеству встроенных функций. Однако базовая поддержка интерактивных Jupyter-файлов (IPYNB) доступна только в платной версии, а новичкам интерфейс может показаться перегруженным.

> Файл с расширением IPYNB (Interactive Python Notebook) — это формат интерактивных ноутбуков Jupyter® Notebook (Рис. 3.4-1), где код, визуализации и пояснения объединены в одном документе. Такой формат идеален для построения отчётов, аналитики и обучающих сценариев.

- **VS Code**® (Microsoft) — быстрый, гибкий и настраиваемый инструмент с бесплатной поддержкой IPYNB и множеством плагинов. Подходит как для начинающих, так и для профессионалов. Позволяет интегрировать GitHub Copilot и плагины для работы с языковыми моделями, что делает его отличным выбором для проектов в области ИИ и data science.
- **Jupyter Notebook** — Классический и популярный выбор для экспериментов и обучения. Позволяет писать код, добавлять пояснения, визуализировать результаты в одном интерфейсе (Рис. 3.4-1). Идеально для быстрой проверки гипотез, работы с LLM и создания воспроизво-

димых шагов анализа данных. Для управления зависимостями и библиотеками рекомендуется использовать Anaconda Navigator — визуальный интерфейс для управления Python-средой.

Рис. 3.4-1 Jupyter Notebook один из самых удобных и популярных инструментов по созданию Pipeline процессов.

- **Google Collab**™ (и платформа Kaggle (Рис. 9.2-5)) — облачная альтернатива Jupyter, предоставляющая бесплатный доступ к GPU/TPU. Это отличное решение для старта — без установки локального ПО и с возможностью работать прямо из браузера. Поддерживает интеграцию с Google Drive и недавно — с Gemini (LLM от 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

Рис. 3.4-2 Сравнение IDE: Jupyter Notebook один из самых удобных и простых инструментов по созданию Pipeline процессов.

Выбор IDE зависит от ваших задач. Если вы хотите быстро начать работать с ИИ, попробуйте Jupyter Notebook или Google Collab. Для серьезных проектов лучше использовать PyCharm или VS Code. Главное — начать. Современные инструменты позволяют быстро превратить эксперименты в рабочие решения.

Все описанные IDE позволяют создавать пайплайны обработки данных — то есть цепочки из модулей блоков кода (которые мог сгенерировать LLM), каждый из которых отвечает за свой этап, например:

- аналитические сценарии,
- цепочки извлечения информации из документов,
- автоматические ответы на основе RAG,
- генерация отчётности и визуализаций.

Благодаря модульной структуре каждый шаг можно представить, как отдельный блок: загрузка данных → фильтрация → анализ → визуализация → экспорт результатов. Эти блоки можно переиспользовать, - адаптировать и собирать в новые цепочки, как конструктор, только для данных.

Для инженеров, руководителей и аналитиков это открывает возможность документировать логику принятия решений в виде кода, который может быть сгенерирован с помощью LLM. Такой подход помогает ускорять рутинные задачи, автоматизировать типовые операции и формировать воспроизводимые процессы, где каждый шаг чётко зафиксирован и прозрачен для всех участников команды.

Подробнее автоматизированные ETL Pipelines (Рис. 7.2-3), инструменты Apache Airflow (Рис. 7.4-4), Apache NiFi (Рис. 7.4-5) и n8n (Рис. 7.4-6) для выстраивания блоков логики при автоматизации процессов будут обсуждаться в 7 и 8 части книги.

IDE с поддержкой LLM и будущих изменениях в программировании

Интеграция искусственного интеллекта в процессы разработки меняет ландшафт программирования. Современные среды больше не просто текстовые редакторы с подсветкой синтаксиса — они превращаются в интеллектуальных ассистентов, способных понимать логику проекта, дописывать код и даже объяснять, как работает тот или иной фрагмент кода. На рынке появляются продукты, которые при помощи ИИ расширяют границы привычной разработки:

- **GitHub Copilot** (интегрируется в VS Code, PyCharm): ИИ-ассистент, который генерирует код на основе комментариев или частичных описаний, превращая текстовые подсказки в готовые решения.
- **Cursor** (форк VS Code с ИИ-ядром): позволяет не только дописывать код, но и задавать вопросы к проекту, искать зависимости и обучаться на кодовой базе.
- **JetBrains AI Assistant**: плагин для IDE JetBrains (включая PyCharm) с функцией объяснения сложного кода, оптимизации и создания тестов.
- **Amazon CodeWhisperer**: аналог Copilot с акцентом на безопасность и поддержку AWS-сервисов от Amazon.

Программирование в ближайшие годы претерпит кардинальные изменения. Основной фокус сместится от рутинного написания кода к проектированию модели и архитектуры данных — разработчики будут больше заниматься проектированием систем, в то время как ИИ возьмёт на себя шаблонные задачи: генерацию кода, тестов, документации и базовых функций. Будущее программирования — это сотрудничество человека и ИИ, где машины берут на себя техническую рутину, а люди сосредотачиваются на творчестве.

Программирование на естественном языке станет повседневностью. Персонализация IDE выйдет на новый уровень — среды разработки научатся адаптироваться под стиль работы пользователя, и компании, предугадывая паттерны, предлагая контекстные решения и обучаясь на основе предыдущих проектов.

> Это не отменяет роли разработчика, но кардинально её трансформирует: от написания кода — к управлению знаниями, качеством и процессами. Подобная эволюция затронет и сферу бизнес-аналитики, где создание отчётов, визуализаций и приложений для поддержки принятия решений всё чаще будет происходить через генерацию кода и логики с помощью ИИ и LLM, чат и агентов-интерфейсов.

После того как компания настроила LLM-чаты и выбрала подходящую среду разработки, следующим важным этапом становится организация данных. Этот процесс включает в себя извлечение информации из разрозненных источников, ее очистку, преобразование в структурированный вид и интеграцию в корпоративные системы.

В современном Data-Centric подходе к управлению данными ключевая цель — привести данные к единой универсальной форме, которая будет совместима с большим количеством инструментов и приложений. Для работы с процессами структурирования и структурированными данными необходимы специализированные библиотеки. Одной из самых мощных, гибких и популярных является библиотека Pandas для Python. Она позволяет удобно обрабатывать табличные данные: фильтровать, группировать, очищать, дополнять, выполнять агрегации и строить отчёты.

Python Pandas: незаменимый инструмент для работы с данными

В мире анализа и автоматизации данных Pandas занимает особое место. Это одна из самых популярных и широко используемых библиотек языка программирования Python [86], предназначенная для работы со структурированными данными.

Библиотека — это как набор готовых инструментов: функций, модулей, классов. Как на стройплощадке не нужно каждый раз изобретать молоток или уровень, так и в программировании библиотеки позволяют быстро решать задачи без повторного изобретения базовых функций и решений.

> Pandas — это библиотека Python с открытым исходным кодом, предоставляющая высокопроизводительные и интуитивно понятные структуры данных, в частности DataFrame — универсальный формат для работы с таблицами. Pandas — это швейцарский нож для аналитиков, инженеров и разработчиков, работающих с данными.

Python — это высокоуровневый язык программирования с простым синтаксисом, активно применяемый в аналитике, автоматизации, машинном обучении и веб-разработке. Его популярность объясняется читаемостью кода, кроссплатформенностью и богатой экосистемой библиотек. На сегодняшний день для Python создано более 137 000 пакетов с открытым исходным кодом [87], и это число продолжает расти практически ежедневно. Каждая такая библиотека — это своего рода хранилище готовых функций: от простейших математических операций до сложных инструментов для обработки изображений, анализа больших данных, работы с нейросетями и интеграции с внешними сервисами.

> Другими словами, представьте, что у вас есть бесплатный и открытый доступ к сотням тысяч готовых программных решений — библиотек и инструментов, которые вы можете напрямую встраивать в свои бизнес-процессы. Это как огромный каталог приложений, предназначенных для автоматизации, анализа, визуализации, интеграции и многого другого — и всё это доступно сразу после установки Python.

Pandas — один из самых популярных пакетов в экосистеме Python. В 2022 году среднее количество загрузок библиотеки Pandas достигало 4 миллионов в день (Рис. 3.4-3), тогда как к началу 2025 года этот показатель увеличился до 12 миллионов загрузок в день, что отражает её растущую популярность и широкое применение в аналитике данных и чатах LLM [86].

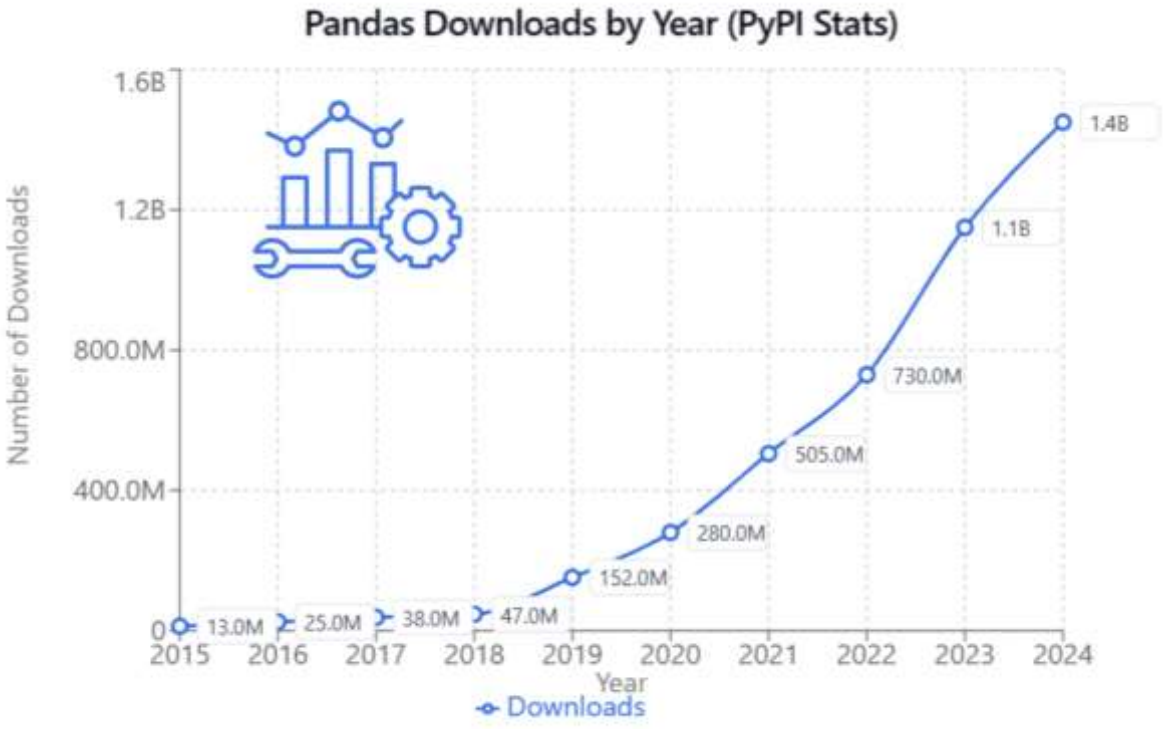


Рис. 3.4-3 Pandas – одна из самых загружаемых библиотек. В 2024 году её ежегодное количество загрузок превысило 1,4 миллиарда.

Язык запросов в библиотеке Pandas по своей функциональности похож на язык запросов SQL, который мы рассматривали в главе "Реляционные базы данных и язык запросов SQL".

> В мире аналитики и управления структурированными данными Pandas выделяется своей простотой, скоростью и мощностью, предоставляя пользователям широкий спектр инструментов для эффективного анализа и обработки информации.

Оба инструмента — SQL и Pandas — предоставляют мощные возможности для работы с данными, особенно в сравнении с традиционным Excel. Они поддерживают такие операции, как выборка, фильтрация (Рис. 3.4-4), с той лишь разницей, что SQL оптимизирован для работы с реляционными базами данных, а Pandas обрабатывает данные в оперативной памяти (RAM), что позволяет запускать его на любом компьютере, без необходимости создания баз данных и развёртывания отдельной инфраструктуры.

Рис. 3.4-4 Pandas, в отличие от SQL, обладает гибкостью в работе с различными форматами данных, не ограничиваясь базами данных.

Pandas часто предпочитают использовать в научных исследованиях, автоматизации процессов, создании конвейеров (в том числе ETL) и манипулировании данными на Python, в то время как SQL является стандартом управления базами данных и часто используется в корпоративных средах для работы с большими объемами данных.

> Библиотека Pandas языка программирования Python позволяет выполнять не только базовые операции, такие как чтение и запись таблиц, но и более сложные задачи, включая объединение данных, группировку данных и проведение сложных аналитических расчетов.

Сегодня библиотека Pandas используется не только в академических исследованиях и бизнес-аналитике, но и в связке с LLM-моделями. Например, подразделение Meta® (Facebook™) при публикации новой open source модели LlaMa 3.1 в 2024 году особое внимание уделило работе со структурированными данными, сделав одним из ключевых и первых кейсов в своём релизе именно обработку структурированных датафреймов (Рис. 3.4-5) в формате CSV и интеграцию с библиотекой Pandas прямо в чате.

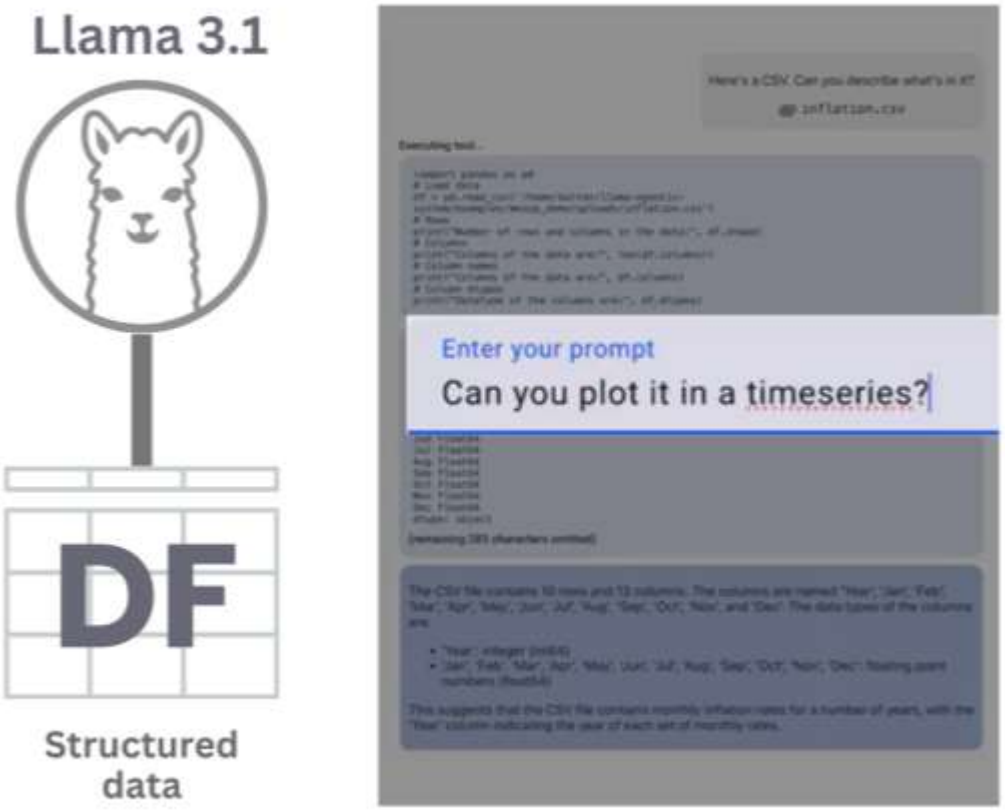


Рис. 3.4-5 Одним из первых и главных кейсов команды Meta презентованном в LlaMa 3.1 в 2024 г. стала построение приложений с использованием Pandas.

> *Pandas - это необходимый инструмент миллионов data scientist'ов, обрабатывающих и подготавливающих данные для генеративного ИИ. Ускорение Pandas с нулевыми изменениями кода станет огромным шагом вперед. Data-ученые смогут обрабатывать данные за минуты, а не за часы, и получать на порядки больше данных для обучения генеративных моделей ИИ [88].*
>
> — Дженсен Хуанг, основатель и генеральный директор NVIDIA

Используя Pandas, можно управлять и анализировать наборы данных, намного превосходящие возможности Excel. В то время как Excel, как правило, способен обрабатывать до 1 миллиона строк данных, Pandas может без труда работать с наборами данных (Рис. 9.1-2, Рис. 9.1-10), содержащими

десятки миллионов строк [89]. Эта возможность позволяет пользователям выполнять сложный анализ данных и визуализацию на больших наборах данных, обеспечивая глубокое понимание и облегчая принятие решений на основе данных. Кроме того, Pandas имеет сильную поддержку сообществ [90]: сотни миллионов разработчиков и аналитиков по всему миру (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) ежедневно используют его онлайн или офлайн, предоставляя большое количество готовых решений для любых бизнес-задач.

В основе большинства аналитических процессов на Python лежит структурированная форма данных под названием DataFrame, предоставляемая библиотекой Pandas. Это — мощный и гибкий инструмент для организации, анализа и визуализации табличных данных.

DataFrame: универсальный формат табличных данных

DataFrame — это центральная структура в библиотеке Pandas, представляющая собой двумерную таблицу (Рис. 3.4-6), где строки соответствуют отдельным объектам или записям, а столбцы — их характеристикам, параметрам или категориям. Такая структура визуально напоминает таблицы Excel, но значительно превосходит их по гибкости, масштабируемости и функциональности.

> **DataFrame** — это способ представления и обработки табличных данных, хранящихся в оперативной памяти компьютера.

DataFrame — это способ представления и обработки табличных данных, хранящихся в оперативной памяти компьютера. В таблице строки могут отражать, например, элементы строительного проекта, а столбцы — их свойства: категории, габариты, координаты, стоимость, сроки и так далее. Причём в такой таблице может содержаться как информация по одному проекту (Рис. 4.1-13), так и данные о миллионах объектов из тысяч разных проектов (Рис. 9.1-10). . Благодаря векторизованным операциям Pandas, можно легко фильтровать, группировать и агрегировать такие объёмы информации с высокой скоростью.

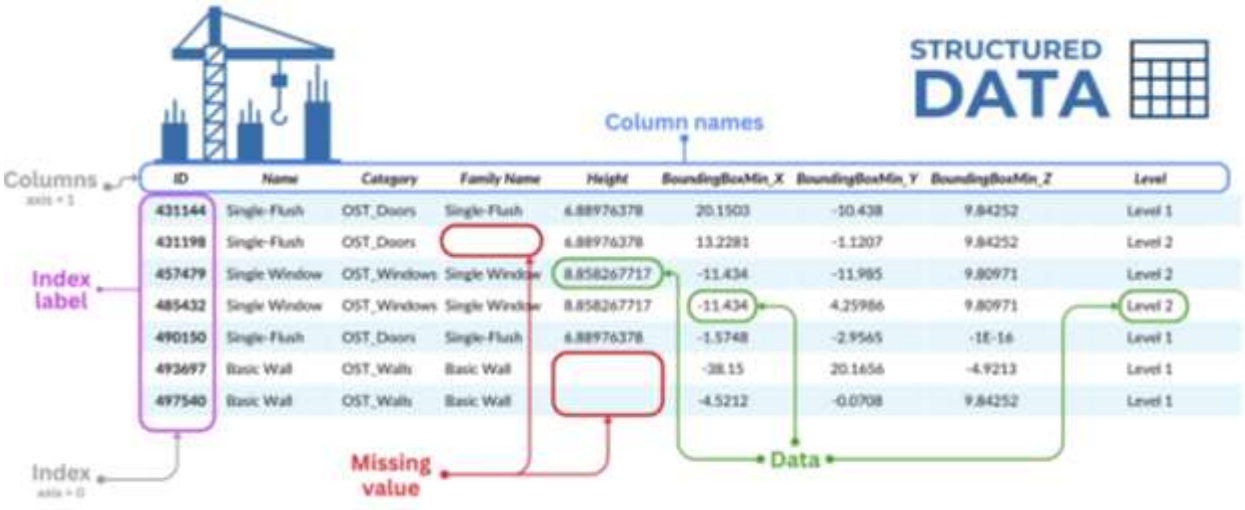


Рис. 3.4-6 Строительный проект в виде DataFrame — это двумерная таблица с элементами в строках и атрибутами в столбцах.

По оценке Nvidia, уже сегодня до 30% всех вычислительных ресурсов используются для обработки структурированных данных - датафреймов, и эта доля продолжает расти.

> *Обработка данных — это то, чем занимается, наверное, треть всех мировых вычислений в каждой компании. Обработка данных и данные большинства компаний находятся в DataFrame, в табличном формате.*
>
> — Дженсен Хуанг, Генеральный директор Nvidia [91]

Перечислим некоторые ключевые особенности DataFrame в Pandas:

- **Столбцы**: в DataFrame данные организованы в столбцы, каждый из которых имеет уникальное имя. Столбцы-атрибуты могут содержать данные различных типов, подобно столбцам в базах данных или столбцам в таблицах.
- **Pandas Series** — это одномерная структура данных в Pandas, похожая на список или колонку в таблице, где каждому значению соответствует свой индекс

> В Pandas Series насчитывается более 400 атрибутов и методов, что делает работу с данными невероятно гибкой. Можно напрямую применять одно, из четырехсот доступных функции, к столбцу, выполнять математические операции, фильтровать данные, заменять значения, работать с датами, строками и многим другим. Кроме того, Series поддерживает векторизованные операции, что значительно ускоряет обработку больших наборов данных по сравнению с циклическими вычислениями. Например, можно легко умножить все значения на число, заменить пропущенные данные или применить сложные трансформации без написания сложных циклов.

- **Строки**: в DataFrame могут быть проиндексированы уникальными значениями. Этот индекс позволяет быстро изменять и корректировать данные в определенных строках.
- **Индекс**: по умолчанию при создании DataFrame Pandas присваивает каждой строке индекс от 0 до N-1 (где N - количество всех строк в DataFrame). Однако индекс можно изменить, чтобы включить в него особые обозначения, такие как даты или уникальные характеристики.
- **Индексирование** строк в DataFrame означает, что каждой строке присваивается уникальное имя или метка, которая называется индексом DataFrame.
- **Типы данных:** DataFrame поддерживает различные типы данных, включая: `int`, `float`, `bool`, `datetime64` и `obect` для текстовых данных. Каждый столбец DataFrame имеет свой собственный тип данных, который определяет, какие операции можно выполнять над его содержимым.
- **Операции с данными**: DataFrame поддерживает широкий спектр операций для обработки данных, включая агрегирование (`groupby`), слияние (`merge` и `join`), конкатенацию (`concat`), разделение-применение-комбинирование и многие другие методы преобразования данных.
- **Манипулирование размерами**: DataFrame позволяет добавлять и удалять столбцы и

строки, что делает его динамической структурой, которая может быть изменена в соответствии с потребностями анализа данных.

- **Визуализация данных:** используя встроенные методы визуализации или взаимодействуя с популярными библиотеками визуализации данных, такими как Matplotlib или Seaborn, DataFrame можно легко преобразовать в графики и диаграммы, чтобы представить данные в графическом виде.
- **Ввод и вывод данных:** Pandas предоставляет функции для чтения импорта и экспорта данных в различные форматы файлов, такие как CSV, Excel, JSON, HTML и SQL, что потенциально делает DataFrame центральным узлом для сбора и распространения данных.

В отличие от форматов CSV и XLSX, Pandas DataFrame обеспечивает более высокую гибкость и производительность при работе с данными: он позволяет обрабатывать большие объёмы информации в оперативной памяти, поддерживает расширенные типы данных (включая даты, логические значения и временные ряды), а также предоставляет широкие возможности для фильтрации, агрегации, объединения и визуализации данных. В то время как CSV не хранит информацию о типах данных и структуре, а XLSX часто перегружен форматированием и имеет низкую масштабируемость, DataFrame остаётся оптимальным выбором для быстрой аналитики, автоматизации процессов и интеграции с ИИ-моделями (Рис. 3.4-7). В следующих главах мы подробно рассмотрим каждый из этих аспектов данных, также в 8 части книги будут подробно рассматриваться похожие форматы, такие как Parquet, Apache Orc, JSON, Feather, HDF5 и хранилищах данных (Рис. 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)

Рис. 3.4-7 DataFrame - оптимальный выбор манипулирования данными с высокой производительностью и расширенной поддержкой типов данных.

Благодаря своей гибкости, мощности и простоте использования, библиотека Pandas и формат DataFrame стали де-факто стандартом в области анализа данных на Python. Они идеально подходят

как для создания простых отчётов, так и для построения сложных аналитических пайплайнов, особенно в связке с LLM-моделями.

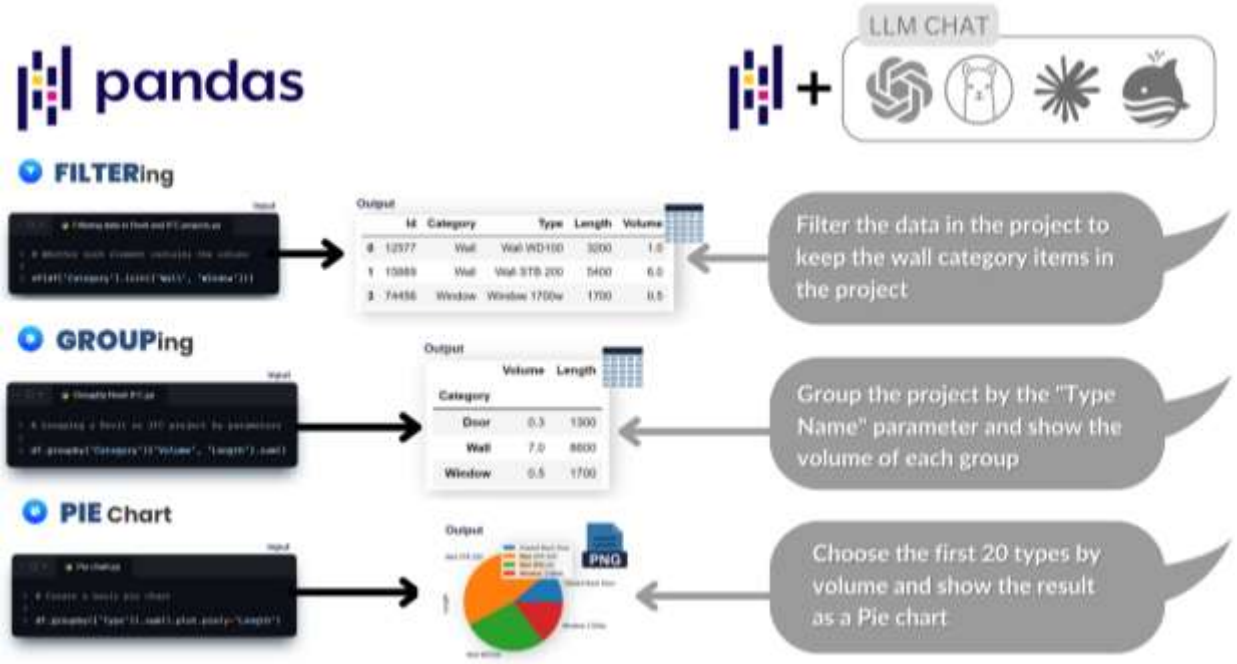


Рис. 3.4-8 LLM упрощают взаимодействие с Pandas: вместо кода достаточно текстового запроса.

Сегодня Pandas активно используется в чатах на базе LLM — таких как ChatGPT, LlaMa, DeepSeek, QWEN и других. Во многих случаях, когда модель получает запрос, связанный с обработкой таблиц, проверкой данных или аналитикой, она генерирует код именно с использованием библиотеки Pandas. Это делает DataFrame естественным «языком» представления данных в диалогах с ИИ (Рис. 3.4-8).

Современные технологии обработки данных, такие как Pandas, значительно упрощают аналитику, автоматизацию и интеграцию данных в бизнес-процессы. Они позволяют быстро получать результат, снижать нагрузку на специалистов и обеспечивать воспроизводимость операций.

Дальнейшие шаги: создание устойчивого каркаса данных

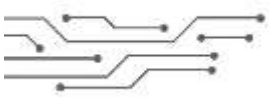
В этой части мы рассмотрели ключевые типы данных, используемые в строительной отрасли, познакомились с различными форматами их хранения и проанализировали роль современных инструментов, включая LLM и IDE, в обработке информации. Мы убедились, что эффективное управление данными – это фундамент для принятия обоснованных решений и автоматизации бизнес-процессов. Организации, способные структурировать и систематизировать свои данные, получают значительное конкурентное преимущество на этапах обработки и трансформации данных.

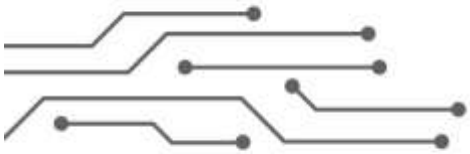
Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные подходы в ваших повседневных задачах:

- Проведите аудит данных в ваших процессах
  - [ ] Составьте реестр всех типов данных, которые вы используете в проектах
  - [ ] Определите, какие типы и модели данных наиболее критичны для ваших бизнес-процессов
  - [ ] Выявите проблемные области, где информация остается часто неструктурированной, слабоструктурированной или недоступной
- Начните формировать стратегию управления данными
  - [ ] Поднимайте вопросы политики и стандартов для работы с различными типами данных
  - [ ] Проанализируйте, какие из ваших рабочих процессов можно улучшить за счёт преобразования неструктурированных данных в структурированные
  - [ ] Создайте регламент хранения и доступа к данным, учитывающий безопасность и конфиденциальность
- Установите и освойте базовые инструменты для работы с данными
  - [ ] Выберите подходящую IDE, соответствующую вашим задачам (например установите VS Code или Jupyter Notebook)
  - [ ] Попробуйте установить локальную LLM для конфиденциальной обработки ваших личных данных
  - [ ] Начните экспериментировать с библиотекой Pandas для обработки табличных данных XLSX
  - [ ] Опишите в LLM типичные задачи, которые вы обрабатываете в табличных инструментах или базах данных и попросите LLM автоматизировать работу при помощи Pandas

Применение подобных шагов позволит вам постепенно трансформировать подход к работе с данными, переходя от разрозненных, неструктурированных массивов информации к единой экосистеме, где данные становятся доступным и понятным активом. Начните с малого – создайте первый DataFrame в Pandas, запустите локальную LLM, автоматизируйте первую рутинную задачу при помощи Python (например по работе с таблицами в Excel).

Четвёртая часть книги будет посвящена вопросам качества данных, их организации, структурированию и моделированию. Мы сосредоточимся на методологиях, которые позволяют преобразовывать разрозненные источники информации — от PDF и текстов до изображений и CAD-моделей — в структурированные массивы, пригодные для анализа и автоматизации. Также мы изучим, как формализуются требования к данным, как строятся концептуальные и логические модели в строительных проектах, и как в этом процессе могут помочь современные языковые модели (LLM).





IV ЧАСТЬ
КАЧЕСТВО ДАННЫХ: ОРГАНИЗАЦИЯ, СТРУКТУИРОВАНИЕ, МОДЕЛИРОВАНИЕ

Четвертая часть фокусируется на методологиях и технологиях, обеспечивающих трансформацию разрозненной информации в структурированные массивы данных высокого качества. Детально рассматриваются процессы формирования и документирования требований к данным, как основы эффективной информационной архитектуры в строительных проектах. Представлены практические методы извлечения структурированной информации из различных источников (PDF-документы, изображения, текстовые файлы, CAD-модели) с примерами реализации. Анализируется применение регулярных выражений (RegEx) и других инструментов для автоматической валидации и верификации данных. Пошагово описывается процесс моделирования данных на концептуальном, логическом и физическом уровнях с учетом специфики строительной отрасли. Демонстрируются конкретные примеры использования языковых моделей (LLM) для автоматизации процессов структурирования и проверки информации. Предлагаются эффективные подходы к визуализации результатов анализа, повышающей доступность аналитической информации для всех уровней управления строительными проектами.

ГЛАВА 4.1.
ПРЕОБРАЗОВАНИЕ ДАННЫХ В СТРУКТУРИРОВАННУЮ ФОРМУ

В эпоху data-driven экономики данные становятся не препятствием, а основой для принятия решений. Вместо постоянной адаптации информации под каждую новую систему и её форматы, компании всё чаще стремятся формировать единую структурированную модель данных, которая служит универсальным источником истины для всех процессов. Современные информационные системы проектируются не вокруг форматов и интерфейсов, а вокруг смысла данных — ведь структура может меняться, а значение информации остаётся неизменным гораздо дольше.

> Ключ к эффективной работе с данными заключается не в их бесконечной конвертации и трансформации, а в изначально правильной организации: создании универсальной структуры, способной обеспечить прозрачность, автоматизацию и интеграцию на всех этапах жизненного цикла проекта.

Традиционный подход вынуждает при внедрении каждой новой платформы заниматься ручной корректировкой: переносить данные, менять названия атрибутов, подстраивать форматы. Эти шаги не улучшают качество самих данных, а лишь маскируют проблемы, порождая замкнутый цикл бесконечных преобразований. В результате компании становятся зависимыми от конкретных программных решений, а цифровая трансформация замедляется.

В следующих главах мы рассмотрим, как правильно структурировать данные, а затем как создавать универсальные модели, минимизировать зависимость от платформ и сосредоточиться на главном — данных как стратегическом ресурсе, вокруг которого выстраиваются устойчивые процессы.

Учимся превращать документы, PDF, картинки и тексты в структурированные форматы

В строительных проектах подавляющее большинство информации существует в неструктурированной форме: это технические документы, акты выполненных работ, чертежи, спецификации, графики, протоколы. Их разнообразие — как по формату, так и по содержанию — усложняет интеграцию и автоматизацию.

> Процесс преобразования в структурированные или полуструктурированные форматы может варьироваться в зависимости от типа входных данных и желаемых результатов обработки.

Преобразование данных из неструктурированной в структурированную форму — это и искусство, и наука. Этот процесс варьируется в зависимости от типа входных данных и целей анализа и часто занимает значительную часть работы инженера (Рис. 3.2-5) по обработке данных и аналитика, с целью получить чистый, упорядоченный набор данных.

Рис. 4.1-1 Преобразование неструктурированного отсканированного документа в структурированный табличный формат.

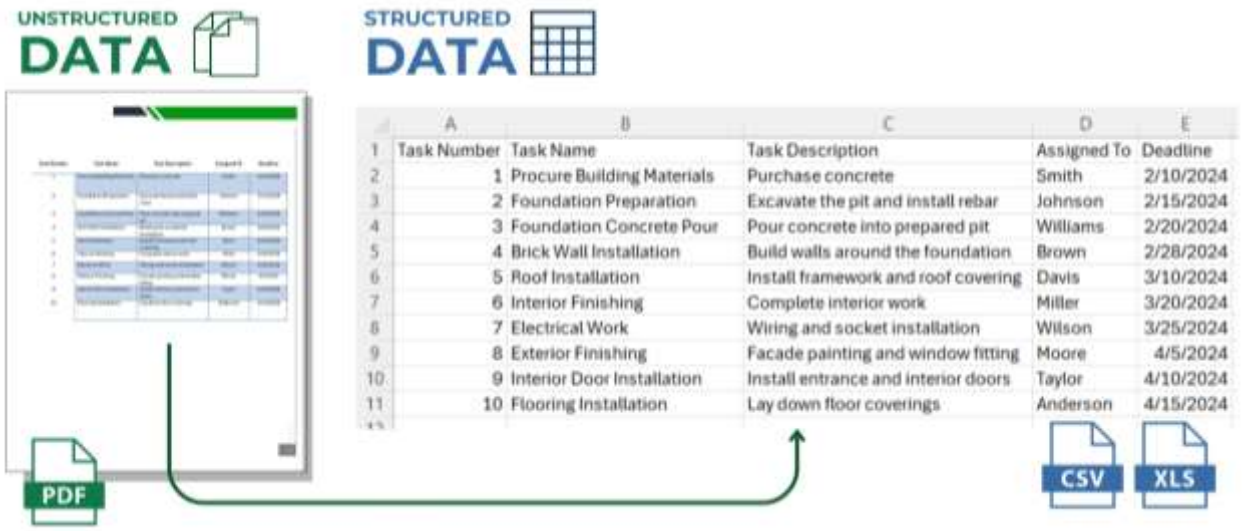
Превращение документов, PDF, картинок и тексты в структурированный формат (Рис. 4.1-1) — это пошаговый процесс, включающий следующие этапы:

- **Извлечение данных** (Extract): на этом этапе загружается исходный документ или изображение, содержащее неструктурированные данные. Это может быть, например, PDF-документ, фотография, чертеж или схема.
- **Преобразование данных** (Transform): далее следует этап преобразования неструктурированных данных в структурированный формат. Например, это может включать распознавание и интерпретацию текста с изображений с помощью оптического распознавания символов (OCR) или других методов обработки.
- **Загрузка и сохранение данных** (Load): последний этап включает в себя сохранение обработанных данных в различных форматах, таких как CSV, XLSX, XML, JSON, для дальнейшей работы, где выбор формата зависит от конкретных требований и предпочтений.

Этот процесс, известный как ETL (Extract, Transform, Load), играет ключевую роль в автоматизированной обработке данных, о нём мы подробнее поговорим в главе "ETL и Pipeline: Extract, Transform, Load". Дальше разберём на примерах, как документы разных форматов преобразуются в структурированные данные.

Пример преобразования PDF-документа в таблицу

Одна из самых частых задач в строительных проектах — обработка технических заданий в PDF-формате. Чтобы продемонстрировать переход от неструктурированных данных к структурированному формату, рассмотрим практический пример: извлечение таблицы из PDF-документа и преобразование её в формат CSV или Excel (Рис. 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

Рис. 4.1-2 В отличие от PDF, форматы CSV и XLSX широко распространены и легко интегрируются в различные системы управления данными.

Языковые модели LLM, такие как ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN значительно упрощают работу специалистов с данными, снижая необходимость глубокого изучения языков программирования и позволяя решать многие задачи с помощью текстовых запросов.

Поэтому, вместо того чтобы тратить время на поиск решений в интернете (обычно это сайт StackOverFlow или тематические форумы и чаты) или обращаться к специалистам по обработке данных, мы можем воспользоваться возможностями современных онлайн или локальных LLM. Достаточно задать запрос, и модель предоставит готовый код для преобразования PDF-документа в табличный формат.

- Отправьте следующий текстовый запрос в любую LLM-модель (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любую другую):

> Напиши, пожалуйста, код для извлечения текста из PDF-файла, который содержит таблицу. Код должен принимать в качестве аргумента путь к файлу и возвращать извлеченную таблицу в виде DataFrame ⏎

- Ответ LLM-модели в большинстве случаев будет представлен в виде кода на Python, так как этот язык широко используется для обработки данных, автоматизации и работы с различными форматами файлов:

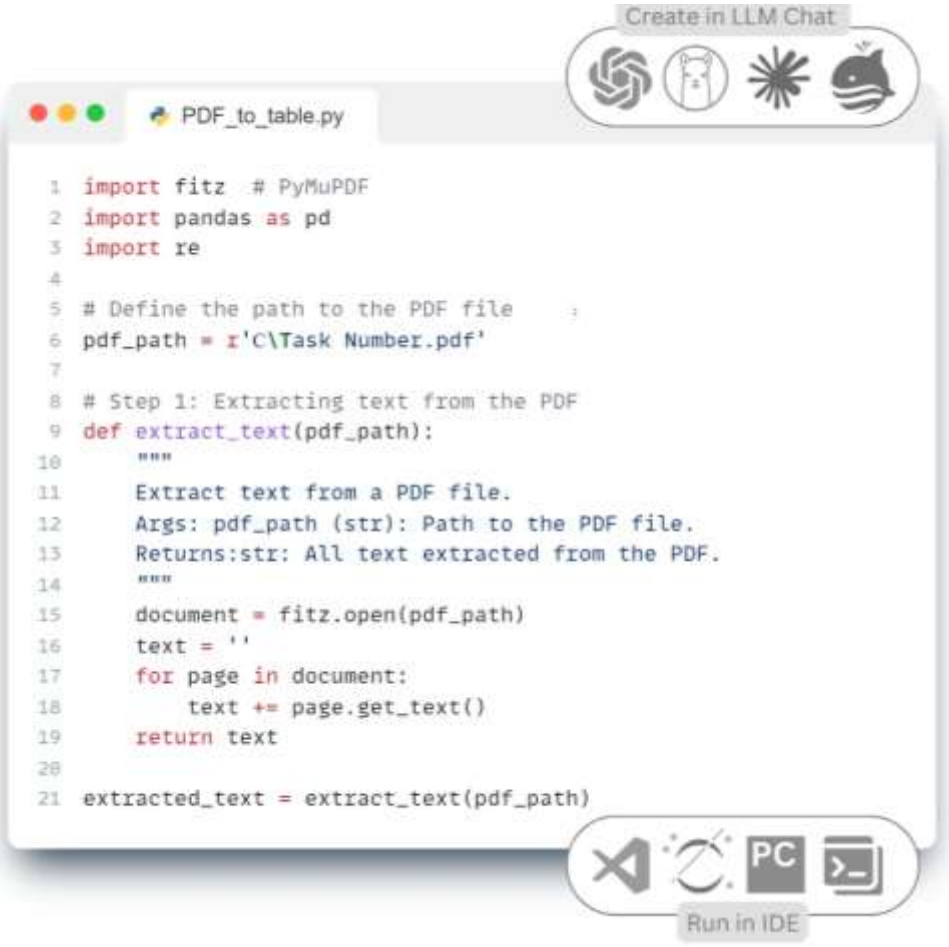


Рис. 4.1-3 Ответ LLM в виде Python кода и его библиотек и пакетов (Pandas, Fitz) извлекает текст из PDF-файла.

Этот код (Рис. 4.1-3) можно запустить в одной из популярных IDE, про которые мы говорили выше, в оффлайн режиме: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с плагином PyDev, Thonny, Wing IDE, IntelliJ IDEA с плагином Python, JupyterLab или популярных онлайн-инструментах: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- На этапе "Преобразование" используем популярную библиотеку Pandas (про которую мы подробно говорили в главе «Python Pandas: незаменимый инструмент для работы с данными»), чтобы считать извлеченный текст в DataFrame и сохранить DataFrame в табличный файл CSV или XLXS:

> Мне нужен код, который будет преобразовывать результирующую таблицу из PDF-файла в DataFrame. Также добавьте код для сохранения DataFrame в CSV-файл ⏎

Ответ LLM:

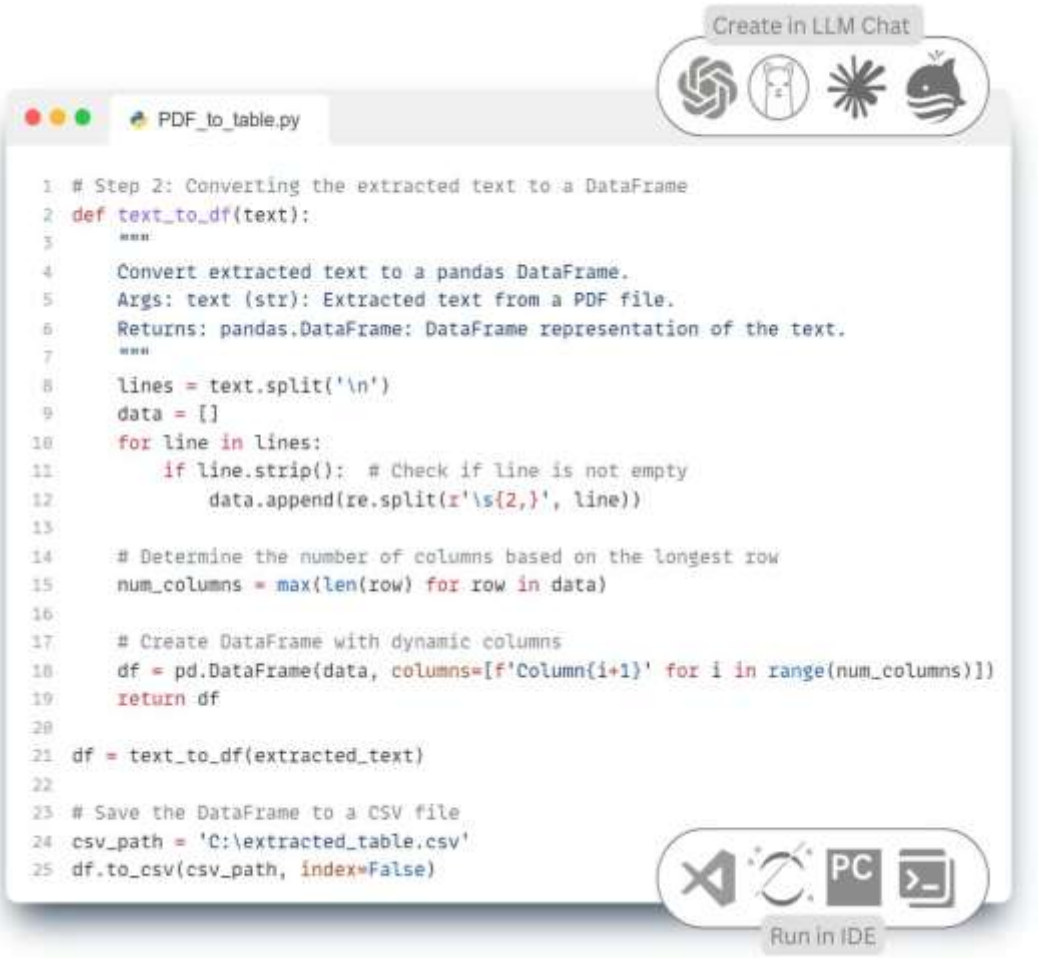


Рис. 4.1-4 Преобразование извлеченной таблицы из PDF в DataFrame и сохранение таблицы в CSV-файл.

Если при выполнении кода (Рис. 4.1-3, Рис. 4.1-4) возникает ошибка — например, из-за отсутствующих библиотек или неправильного пути к файлу, —текст ошибки можно просто скопировать вместе с исходным кодом и повторно отправить в LLM-модель. Модель проанализирует сообщение об ошибке, объяснит, в чём проблема и предложит исправления или дополнительные шаги.

> Таким образом, взаимодействие с ИИ LLM становится полноценным циклом: запрос → ответ → тест → обратная связь → корректировка — без необходимости глубоких технических знаний.

Используя обычный текстовой запрос в LLM чат и десяток строк Python, которые мы можем запустить локально в любом IDE, мы преобразовали PDF-документ в табличный формат CSV, который, в отличие от PDF-документа, легко читается машиной и быстро интегрируется в любые системы управления данными.

Мы можем применять этот код (Рис. 4.1-3, Рис. 4.1-4), скопировав его из любого чата LLM, к десяткам и тысячам новых PDF-документов на сервере, тем самым автоматизируя процесс преобразования потока неструктурированных документов в структурированный табличный формат CSV.

Но документы PDF не всегда содержат текст, чаще всего это отсканированные документы, которые необходимо обработать как изображения. Хотя изображения по своей природе не структурированы, разработка и применение библиотек распознавания позволяют извлекать, обрабатывать и анализировать их содержимое, что дает нам возможность в полной мере использовать эти данные в бизнес-процессах.

Преобразование изображения JPEG, PNG в структурированную форму

Изображения — одна из самых распространённых форм неструктурированных данных. В строительстве и многих других отраслях огромное количество информации хранится в виде отсканированных документов, схем, фотографий и чертежей. Такие данные содержат ценные сведения, но не могут быть напрямую обработаны, как, например, таблица Excel или база данных. Изображения содержат много сложной информации, поскольку их содержание, цвета, текстуры разнообразны, а для извлечения полезной информации требуется специальная обработка.

Сложность использования изображений в качестве источника данных заключается в отсутствии структуры. Изображения не передают смысл прямым, легко поддающимся количественной оценке способом, который компьютер может сразу понять или обработать, как это делает электронная таблица Excel или таблица базы данных. Чтобы преобразовать неструктурированные данные изображений в структурированную форму, необходимо использовать специальные библиотеки, способные интерпретировать содержащуюся в них визуальную информацию (Рис. 4.1-5).

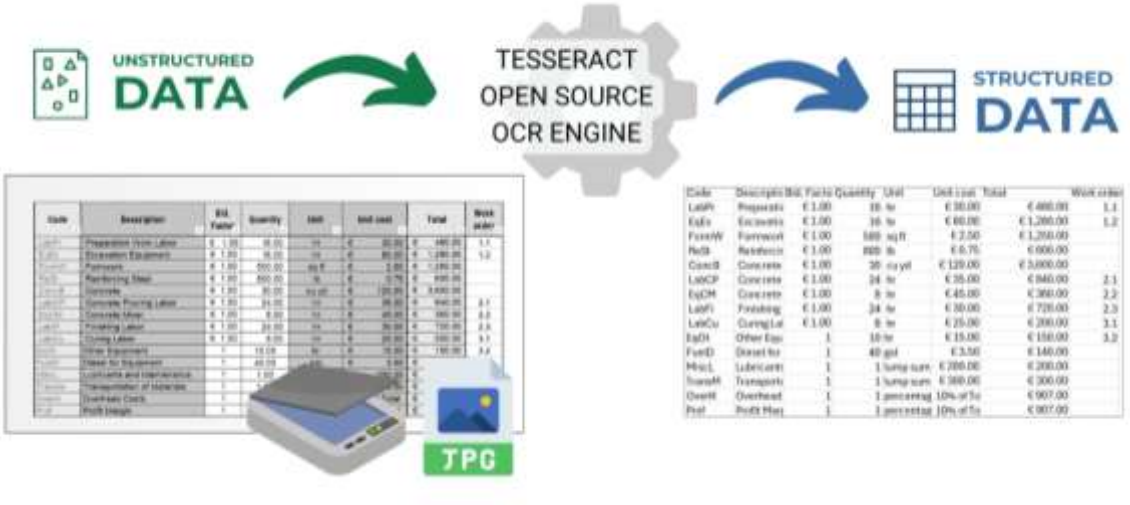


Рис. 4.1-5 Преобразование отсканированных документов и изображений в структурированные форматы возможно с помощью специальных OCR инструментов.

Для извлечения текста с изображений применяется технология OCR (Optical Character Recognition) — оптическое распознавание символов. Она позволяет распознавать буквы и цифры в сканах документов, фотографиях и PDF-файлах, превращая их в редактируемый и машиночитаемый текст. OCR-технологии уже давно используются в автоматизации документооборота и сегодня они легко интегрируются в любые бизнес-процессы и Python-приложения. Один из наиболее популярных OCR инструментов — Tesseract, с открытым исходным кодом, изначально разработанный HP™ и теперь поддерживаемый Google™. Он поддерживает более 100 языков и отличается высокой точностью распознавания.

Давайте попросим LLM чат написать пример кода для получения данных из отсканированной или сфотографированной таблицы в структурированном виде.

- Отправьте текстовый запрос в LLM чат (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любой другой):

> Напиши код, позволяющий преобразовать JPEG-изображение, содержащее таблицу, в таблицу DataFrame ⏎

- Ответ LLM в большинстве случаев предложит использовать библиотеку Pytesseract для распознавания текста в изображениях:

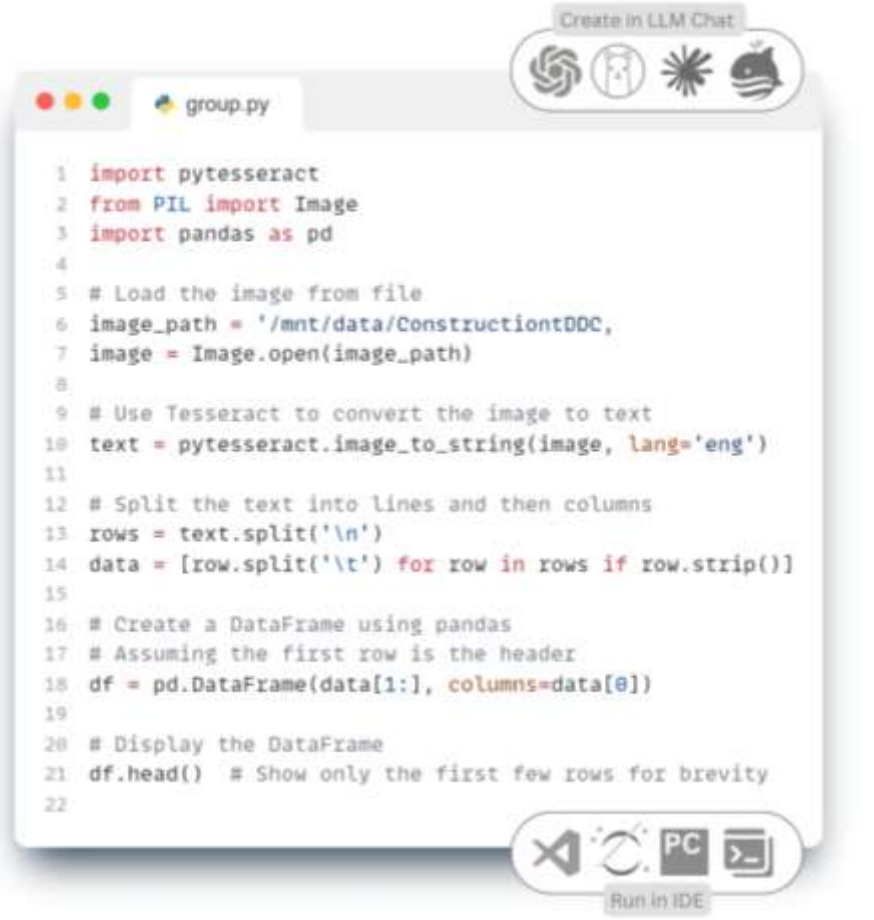


Рис. 4.1-6 Преобразование текста, извлеченного из таблицы изображений или фотографий, в структурированное табличное представление.

В данном примере – код (Рис. 4.1-6), полученный в LLM, использует библиотеку pytesseract (Tesseract для Python) для преобразования изображения в текст с помощью OCR (оптического распознавания символов) и библиотеку Pandas для преобразования этого текста в структурированную форму, т. е. DataFrame.

> Процесс преобразования обычно включает предварительную обработку для улучшения качества изображения, после чего применяются различные алгоритмы для обнаружения образов, выделения признаков или распознавания объектов. В результате неструктурированная визуальная информация преобразуется в структурированные данные.

Хотя PDF и изображения — ключевые источники неструктурированной информации, настоящий

чемпион по объёму — это текст, генерируемый в письмах, чатах, встречах, мессенджерах. Эти данные не просто многочисленны — они разбросаны, неформализованны и крайне слабо структурированы.

Преобразование текстовых данных в структурированную форму

Помимо PDF-документов с таблицами (Рис. 4.1-2) и отсканированных версий табличных форм (Рис. 4.1-5), значительная часть информации в проектной документации представлена в текстовом виде. Это могут быть как связные предложения в текстовых документах, так и фрагментарные записи, разбросанные по чертежам и схемам. В современных условиях обработки данных одной из часто встречающихся задач становится преобразование такого текста в структурированный формат, пригодный для анализа, визуализации и принятия решений.

Центральным элементом этого процесса является таксономия — система классификации, которая позволяет организовать информацию по категориям и подкатегориям на основе общих признаков.

> **Таксономия** — это иерархическая структура классификации, используемая для группировки и организации объектов. В контексте обработки текста она служит основой для систематического распределения элементов по смысловым категориям, что позволяет упростить анализ и повысить качество обработки данных.

Создание таксономии сопровождается этапами извлечения сущностей, их категоризации и связывания с контекстом. Чтобы смоделировать процесс извлечения информации из текстовых данных, необходимо выполнить следующие шаги, похожие на те что мы уже применяли к структурированию данных из PDF документов:

- **Извлечение данных** (Extract): необходимо проанализировать текстовые данные, чтобы извлечь информацию о задержках и изменениях в графике проекта.
- **Категоризация и классификация** (Transofrm): распределяем полученную информацию по категориям, например, причины задержек и изменений в расписании.
- **Интеграция** (Load): в конце подготавливаем структурированные данные для интеграции во внешние системы управления данными.

Рассмотрим ситуацию: у нас есть диалог между менеджером проекта и инженером, в котором обсуждаются проблемы с графиком работ. Наша цель — извлечь ключевые элементы (причины задержки, корректировки сроков) и представить их в структурированном виде (Рис. 4.1-7).

Выполним извлечение на основе ожидаемых ключевых слов, создадим DataFrame для имитации извлечения данных и после трансформации новую DataFrame таблицу, которая будет содержать столбцы для даты, события (например, причина задержки) и действия (например, изменение расписания).

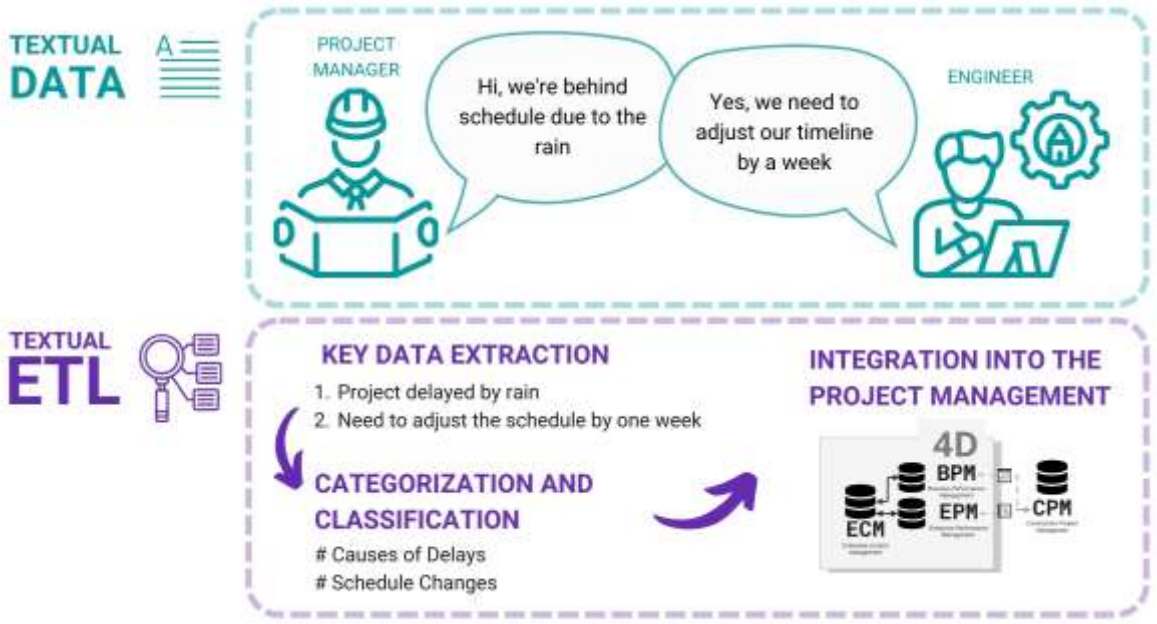


Рис. 4.1-7 Выделение ключевой информации из текста о необходимости корректировки сроков и интеграции изменений в систему управления проектом.

Приведем код для решения задачи с использованием текстового запроса в одной из языковых моделей, как и в предыдущих примерах.

- Отправьте текстовый запрос в любой LLM чат:

> У меня есть разговор между менеджером: "Здравствуйте, мы отстаем от графика из-за дождя" и инженером: "Да, нам нужно скорректировать сроки на неделю". Мне нужен сценарий, который проанализирует будущие похожие текстовые диалоги, извлечет из них причины задержек и необходимые корректировки сроков, а затем сгенерирует DataFrame из этих данных. Затем DataFrame должен быть сохранен в CSV-файл ⏎

- Ответ от LLM обычно будет включать Python-код с использованием регулярных выражений (re - Regex) и библиотеки Pandas (pd):

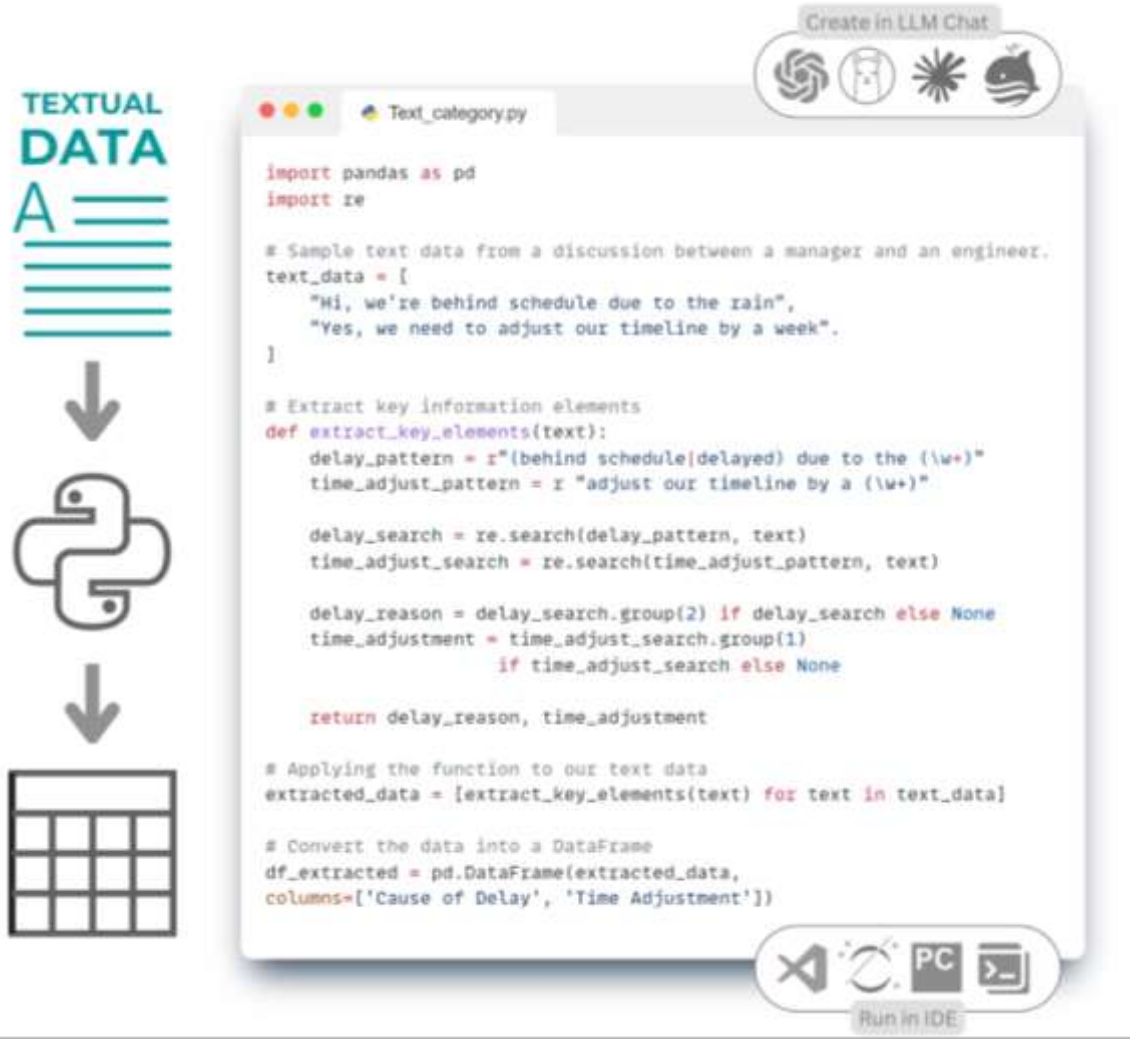
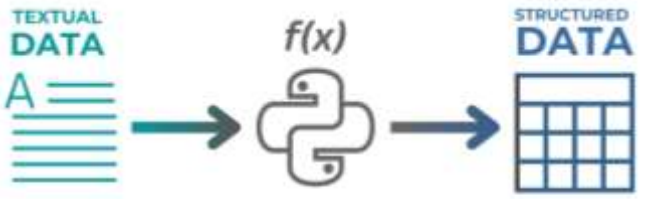


Рис. 4.1-8 Выделение ключевой информации из текста о необходимости корректировки сроков в виде таблицы.

В этом примере (Рис. 4.1-7) текстовые данные, содержащие переписку между менеджером проекта и инженером, анализируются для выявления и извлечения конкретной информации, которая может повлиять на управление будущими проектами, с похожими диалогами. С помощью регулярных выражений (подробнее про регулярные выражения мы поговорим в главе «Структурированные требования и регулярные выражения RegEx») определяются через паттерны причины задержек проекта и необходимые корректировки временного графика. Написанная в данном примере функция извлекает из строк либо причину задержки, либо корректировку времени, основываясь на шаблонах: выделяет слово после "из-за", как причину задержки или слово после "по", как корректировку времени.

Если в строке упоминается задержка из-за погоды, то в качестве причины определяется "дождь"; если в строке упоминается корректировка графика на определенный период, то этот период извлекается в качестве корректировки времени (Рис. 4.1-9). Отсутствие любого из этих слов в строке приводит к значению "Нет" для соответствующего атрибута-столбца.



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Рис. 4.1-9 Сводная таблица, полученная в виде DataFrame, после выполнения кода содержит информацию о существовании задержек и необходимых корректировках времени.

Структурирование и параметризация условий из текста (диалога, письма, документа) позволяет оперативно устранять задержки в строительстве: например, нехватка рабочих может сказываться на темпе работы в плохую погоду, поэтому компании, зная параметры задержки из диалогов (Рис. 4.1-9) между прорабом на строительной площадке и менеджером проекта - заранее могут усилить бригаду при неблагоприятном прогнозе.

Преобразование документов и изображений в структурированный формат может быть достигнуто с помощью относительно простых, открытых и бесплатных инструментов, основанных на категоризации.

Категоризация элементов также является ключевой частью работы с проектными данными, особенно в контексте использования программ CAD (BIM).

Перевод данных CAD (BIM) в структурированную форму

Структурирование и категоризация данных CAD (BIM) - более сложная задача, поскольку данные, сохраняемые из баз данных CAD (BIM), почти всегда представлены в закрытых или в сложных параметрических форматах, часто сочетающих одновременно элементы геометрических данных (полуструктурированные) и элементы метаинформации (полуструктурированные или структурированные данные).

Нативные форматы данных в CAD (BIM) системах, как правило, защищены и недоступны для прямого использования, если только не применять специализированное программное обеспечение

или API-интерфейсы самого разработчика (Рис. 4.1-10). Такая изоляция данных формирует замкнутые хранилища — так называемые «силосы», ограничивающие свободный обмен информацией и тормозящие создание сквозных цифровых процессов в компании.

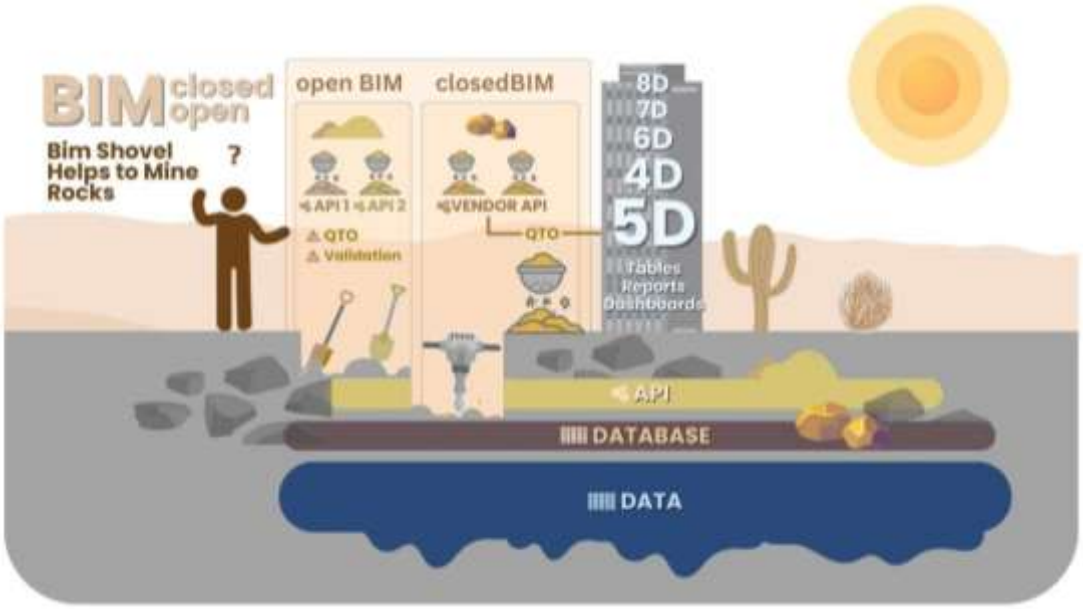


Рис. 4.1-10 Специалисты CAD (BIM) могут получить доступ к нативным данным через API-соединения или инструменты вендора.

В специальных CAD (BIM) форматах информация о характеристиках и атрибутах элементов проекта собирается в иерархическую систему классификации, где сущности с соответствующими свойствами располагаются, подобно плодам фруктового дерева, в самых последних узлах ветвей классификации данных (Рис. 4.1-11).

Извлечение данных из подобных иерархий возможно двумя путями: либо вручную, нажимая на каждый узел, как будто обрабатывая дерево, срубая топором выбранные ветки категорий и типов. Альтернативный вариант - использование программных интерфейсов (API) - предполагает более эффективный, автоматизированный подход к получению и группированию данных, преобразовывая её в итоге в структурированную таблицу для использования в других системах.

Для извлечения структурированных таблиц данных из CAD (BIM) проектов можно использовать различные инструменты, такие как Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC, или решения с открытым исходным кодом, такие как IfcOpSh или IFCjs для формата IFC.

Современные инструменты экспорта и конвертации данных позволяют для упрощения обработки и подготовки данных разделять содержимое CAD-моделей на два ключевых компонента: геометрическую информацию и атрибутивные данные (Рис. 4.1-13) — метаинформацию, описывающую свойства элементов конструкции (Рис. 3.1-16). Эти два слоя данных остаются связаны друг с другом с помощью уникальных идентификаторов, благодаря которым можно точно сопоставить каждый элемент с описанием геометрии (через параметры или полигоны) с его атрибутами: наименование, материал, стадия выполнения, стоимость и прочее. Такой подход обеспечивает целостность модели и позволяет гибко использовать данные как для визуализации (геометрические данные модели), так и для аналитических или управленческих задач (структурированные или слабоструктурированные), работая с двумя типами данных раздельно или параллельно.

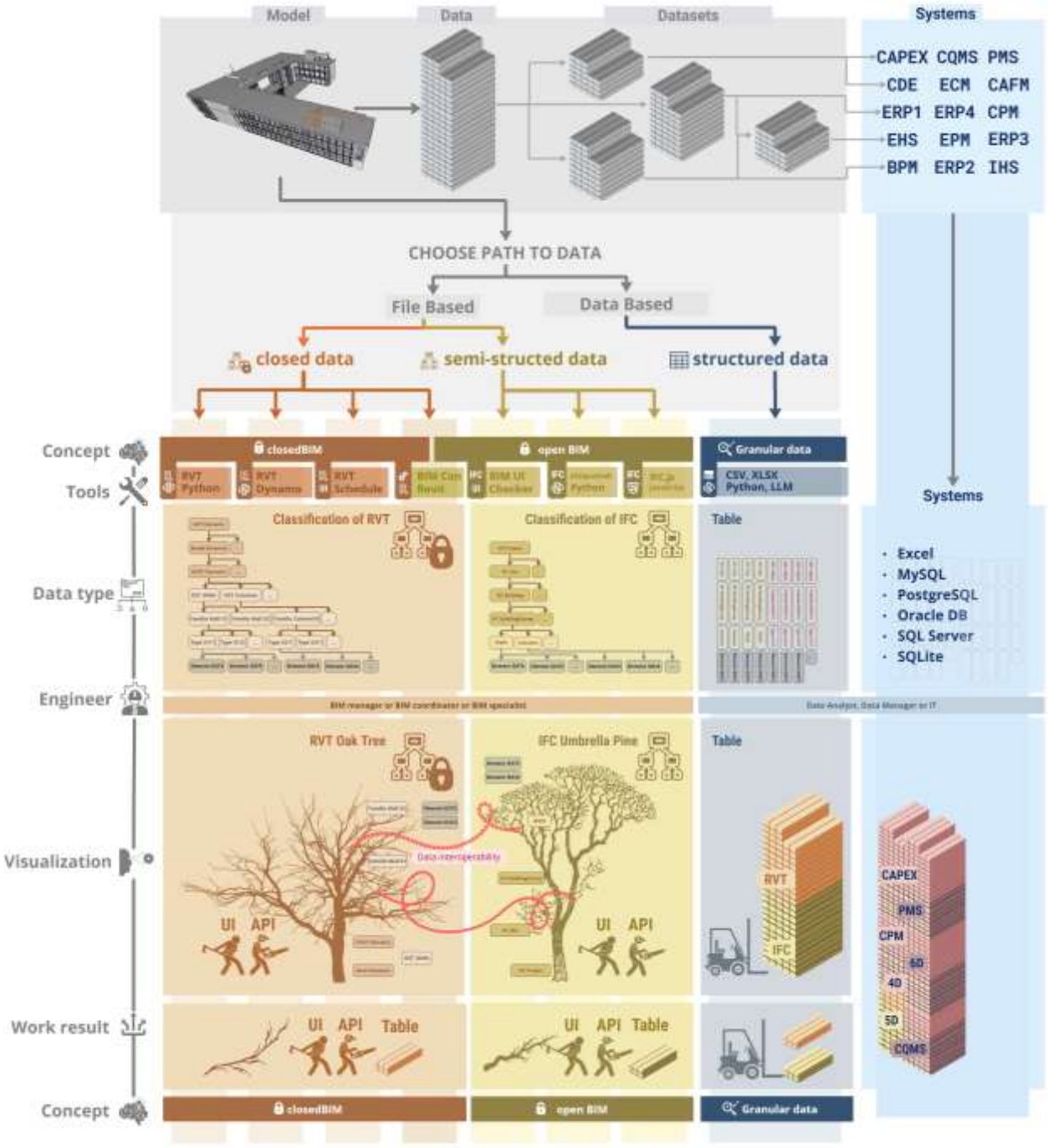


Рис. 4.1-11 Вид информации из баз данных CAD (BIM) представляется пользователю в виде деревьев классификации.

С развитием технологий обратного инжиниринга и появлением наборов для разработки программного обеспечения SDK (Software Development Kit) для конвертаций данных CAD - доступность и преобразование данных из закрытых форматов программ CAD (BIM) стали намного проще. Теперь существует возможность легального и безопасного преобразования данных из закрытых форматов

в универсальные, пригодные для анализа и использования в других системах. Об истории появления первых инструментов реверс-инжиниринга („Open DWG“) и борьбе за доминирование над форматами CAD вендоров мы говорили в главе «Структурированные данные: фундамент цифровой трансформации».

> Инструменты обратного инжиниринга позволяют легитимно получать данные из закрытых проприетарных форматов, разбивая информацию из смешанного формата CAD (BIM) на необходимые пользователю типы данных и форматы, облегчая их обработку и анализ.

Использование обратного инжиниринга и прямого доступа к информации из баз данных CAD делает информацию доступной, позволяя использовать открытые данные и открытые инструменты, а также анализировать данные с помощью стандартных инструментов, строить отчеты, визуализации и интегрировать с другими цифровыми системами (Рис. 4.1-12).

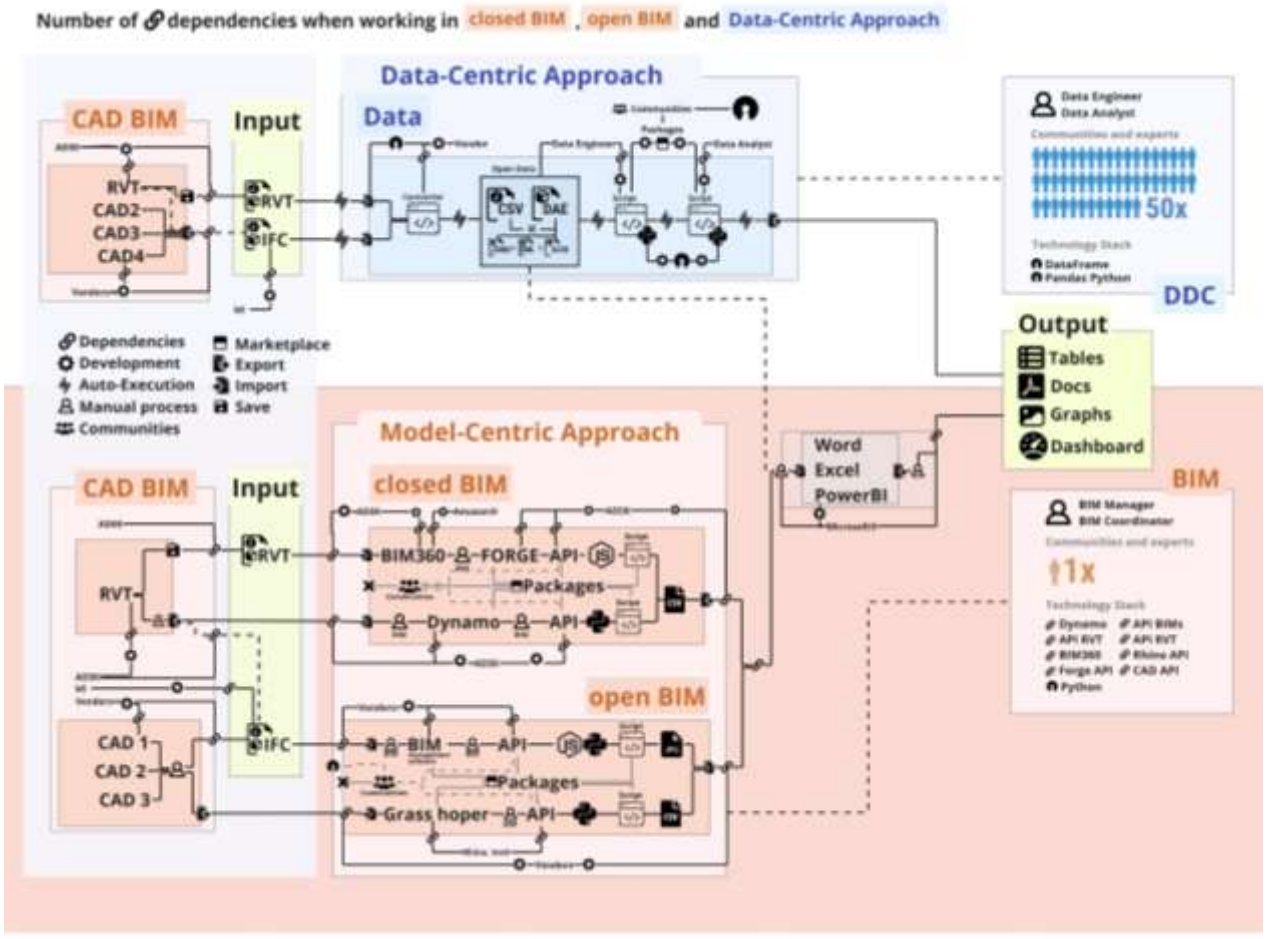


Рис. 4.1-12 Прямой доступ к данным CAD позволяет минимизировать зависимости от программных платформ и перейти к дата-центричному подходу.

С 1996 года для формата DWG, с 2008 года для формата DGN и с 2018 года для RVT стало возможным, изначально закрытые CAD форматы данных, удобно и эффективно преобразовывать в любые

другие форматы, и в том числе структурированные форматы, с помощью инструментов обратного инжиниринга (Рис. 4.1-13). Сегодня практически все крупные CAD (BIM) и крупные инжиниринговые компании в мире используют SDK-инструменты обратного инжиниринга для извлечения данных из закрытых форматов поставщиков CAD (BIM) [92].

Рис. 4.1-13 Использование инструментов обратного инжиниринга позволяет преобразовать базы данных CAD (BIM) программы в любую удобную модель данных.

Преобразование данных из закрытых, проприетарных форматов в открытые и разделение смешанных форматов CAD (BIM) на геометрические и метаинформационные атрибутивные данные упрощает процесс работы с ними, делая их доступными для анализа, манипулирования и интеграции с другими системами (Рис. 4.1-14).

> В современной работе с данными CAD (BIM) мы достигли того уровня, когда для доступа к информации из CAD форматов не нужно запрашивать разрешение у поставщиков CAD (BIM).

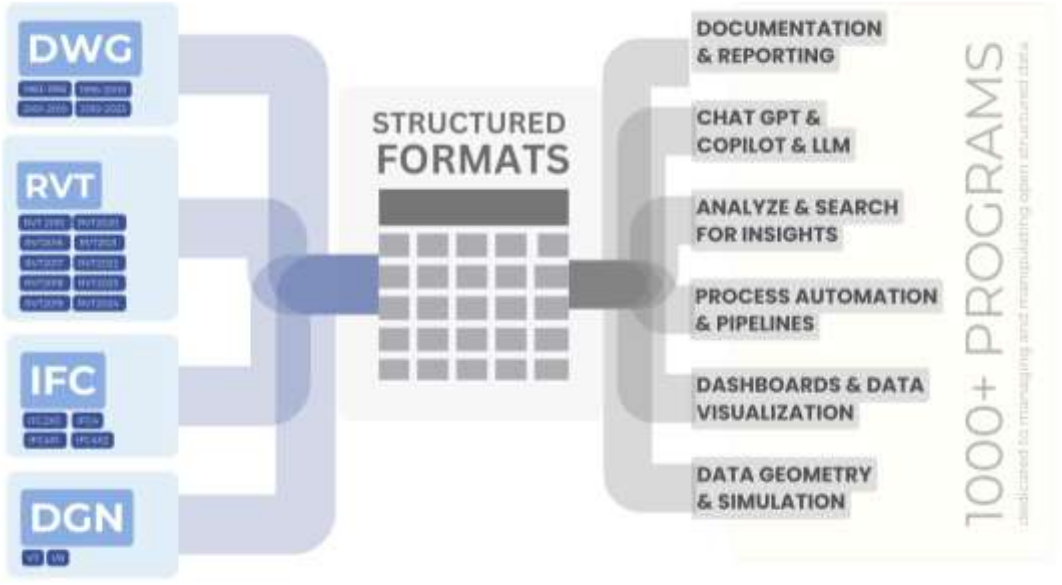


Рис. 4.1-14 Современные инструменты SDK позволяют легально конвертировать данные из проприетарных форматов баз данных CAD (BIM).

Современные тренды в обработке проектных CAD данных продолжают формироваться под влиянием ключевых игроков рынка — CAD-вендоров, которые работают над укреплением своих позиций в мире данных и создают новые форматы и концепты.

Вендоры CAD решений переходят к структурированным данным

С 2024 года в сфере проектирования и строительства происходит значительный технологический сдвиг в области использования и обработки данных. Вместо свободного доступа к проектным данным, производители CAD-систем сосредотачиваются на продвижении очередных новых концепций. Подходы, такие как BIM (созданный в 2002 году) и open BIM (созданный в 2012 год), постепенно уступают место современным технологическим решениям, которые начинают продвигать CAD вендоры [93]:

- Переход на использование „гранулированных“ данных, которые позволяют эффективно управлять информацией и осуществить переход к аналитике данных
- Появление USD формата и внедрение подхода Entity-component-system (ECS) для гибкой организации данных
- Активное использование искусственного интеллекта в обработке данных, автоматизации процессов и аналитике данных
- Развитие интероперабельности - улучшенного взаимодействия между разными программами, системами и базами данных

Подробнее каждый из этих аспектов будет рассмотрен в шестой части книги «CAD и BIM: маркетинг, реальность и будущее проектных данных в строительстве». В рамках данной главы мы лишь кратко обозначим общий вектор изменений: крупнейшие CAD-вендоры сегодня стремятся переосмыслить способы структурирования проектной информации. Один из ключевых сдвигов — отказ от классической файловой модели хранения в пользу гранулированной, ориентированной на аналитику архитектуры данных, обеспечивающей постоянный доступ к отдельным компонентам модели [93].

> Суть происходящего заключается в том, что индустрия постепенно отказывается от громоздких, специализированных и параметрических форматов, требующих использования геометрических ядер, в пользу более универсальных, машиночитаемых и гибких решений.

Одним из таких драйверов изменений стал формат USD (Universal Scene Description), изначально разработанный в индустрии компьютерной графики, но уже получивший признание и в инженерных приложениях, благодаря развитию платформы NVIDIA Omniverse (и Isaac Sim) для симуляций и визуализаций [93]. В отличие от параметрического IFC, USD предлагает более простую структуру, позволяет описывать геометрию и свойства объектов в формате JSON (Рис. 4.1-15), что облегчает обработку информации и ускоряет её интеграцию в цифровые процессы. Новый формат позволяет хранить геометрию (помимо BREP-NURBS – подробнее в 6 части книги) в виде MESH-полигонов, а свойства объектов — в JSON, что делает его более удобным для автоматизированных процессов и работы в облачных экосистемах [94].

Некоторые CAD- и ERP-вендоры уже используют аналогичные форматы (например, NWD, SVF,CP2,

CPIXML), однако большинство из них остаются закрытыми и недоступными для внешнего использования, что ограничивает возможности интеграции и повторного использования данных. В этом контексте USD может сыграть ту же роль, что и DXF в своё время — открытой альтернативы проприетарным форматам, подобным 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

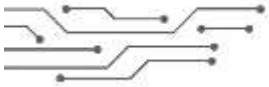
Рис. 4.1-15 USD формат, как попытка CAD вендоров удовлетворить запрос на интероперабельности и независимость проектных данных от геометрических ядер.

Переход крупнейших разработчиков к открытым и упрощённым USD, GLTF, OBJ, XML (закрытым NWD, CP2, SVF, SVF2, CPIXML) и аналогичным форматам (Рис. 3.1-17) отражает глобальную тенденцию и запрос отрасли к упрощению данных и повышению их доступности. В ближайшие годы можно ожидать постепенного ухода от сложных параметрических стандартов и форматов с зависимостью от геометрических ядер в пользу более лёгких и структурированных решений. Такой переход ускорит цифровизацию строительной отрасли, облегчит автоматизацию процессов и упростит обмен данными.

Несмотря на стратегические планы CAD-вендоров по продвижению новых открытых форматов, специалисты строительной отрасли могут также получать полный доступ к данным из закрытых CAD-систем, без необходимости использования CAD (BIM) инструментов, с помощью инструментов обратного инжиниринга.

> Все эти тенденции неизбежно ведут к переходу от громоздких, монолитных 3D-моделей к универсальным, структурированным данным и к использованию форматов, которые уже давно зарекомендовали себя в других отраслях. Как только проектные команды начинают воспринимать CAD-модели не просто как визуальные объекты или набор файлов, а как базы данных, содержащие знания и информацию, — подход к проектированию и управлению кардинально меняется.

После того как команды научились извлекать структурированные данные из документов, текстов, чертежей и CAD-моделей, и получили доступ к базам данных, следующим ключевым шагом становится моделирование данных и обеспечение их качества. Именно от этого этапа во многом зависит скорость обработки и трансформации информации, которая будет использоваться в конечном итоге для принятия решений в конкретных прикладных задачах.



ГЛАВА 4.2.
КЛАССИФИКАЦИЯ И ИНТЕГРАЦИЯ: ЕДИНЫЙ ЯЗЫК СТРОИТЕЛЬНЫХ ДАННЫХ

Скорость принятия решений зависит от качества данных

Современная архитектура проектных данных претерпевает фундаментальные изменения. Отрасль уходит от громоздких, изолированных моделей и закрытых форматов в сторону более гибких, машиночитаемых структур, ориентированных на аналитику, интеграцию и автоматизацию процессов. Однако переход к новым форматам сам по себе не гарантирует эффективности — в центре внимания неизбежно оказывается качество самих данных.

> На страницах этой книги мы много говорим о форматах, системах и процессах. Но все эти усилия теряют смысл без одного ключевого элемента — данных, которым можно доверять. Качество данных — это краеугольный камень цифровизации, к которому мы будем возвращаться на протяжении всех последующих частей.

Современные строительные компании — особенно крупные — используют десятки, а иногда и тысячи различных систем и баз данных (Рис. 4.2-1). Эти системы должны не только регулярно наполняться новой информацией, но и эффективно взаимодействовать между собой. Все новые данные, формируемые в результате обработки поступающей информации, интегрируются в эти среды и служат для решения конкретных бизнес-задач.

И если раньше решением по конкретным бизнес-задачам принимались руководителями высшего звена — так называемыми HiPPO (Рис. 2.1-9) — на основе опыта и интуиции, то сегодня, в условиях резкого роста объёма информации, такой подход становится спорным. На смену приходит автоматизированная аналитика, работающая с данными в реальном времени.

> "Традиционно-ручное" обсуждение бизнес-процессов на уровне руководителей будет смещаться в сторону операционной аналитики, которая требует быстрых ответов на бизнес-запросы.

Эпоха, когда бухгалтеры, прорабы и сметчики вручную формировали отчеты и сводные таблицы, а также витрины данных по проектам в течение нескольких дней и недель, уходит в прошлое. Сегодня скорость и своевременность принятия решений становятся ключевым фактором конкурентного преимущества.

Рис. 4.2-1 В строительной отрасли на расчёты и принятие решений уходит несколько дней, в отличие от других отраслей, где на это происходит за часы или минуты.

Главное отличие строительной отрасли от более развитых в цифровом отношении отраслей (Рис. 4.2-1) — в низком уровне качества и стандартизации данных. Устаревшие подходы к формированию, передаче и обработке информации замедляют процессы и создают хаос. Отсутствие единых стандартов качества данных препятствует внедрению сквозной автоматизации.

> Одной из главных проблем остается низкое качество исходных данных, а также отсутствие формализованных процессов их подготовки и проверки. Без достоверных и согласованных данных невозможна эффективная интеграция между системами. Это ведёт к задержкам, ошибкам и увеличению издержек на каждом этапе жизненного цикла проекта.

В следующих разделах книги мы подробно рассмотрим, как можно повысить качество данных, стандартизировать процессы и сократить путь от получения информации до качественных, проверенных и согласованных данных.

Стандартизация и интеграция данных

Эффективная работа с данными требует чёткой стратегии стандартизации. Только при наличии чётких требований к структуре и качеству данных можно автоматизировать их проверку, сократить количество ручных операций и ускорить принятие обоснованных решений на всех этапах проекта.

В повседневной практике строительной компании ежедневно приходится обрабатывать сотни файлов: электронных писем, PDF-документов, проектных файлов CAD, данных с датчиков IOT, которые необходимо интегрировать в бизнес-процессы компании.

Лес экосистемы компании, состоящий из баз данных и инструментов (Рис. 4.2-2), должен научиться получать питательные вещества из поступающих разноформатных данных, чтобы добиться нужных компании результатов.

> Чтобы эффективно справиться с потоком данных, необязательно нанимать целую армию менеджеров, в первую очередь необходимо разработать строгие требования и стандарты к данным и использовать соответствующие инструменты для их автоматической проверки, унификации и обработки.

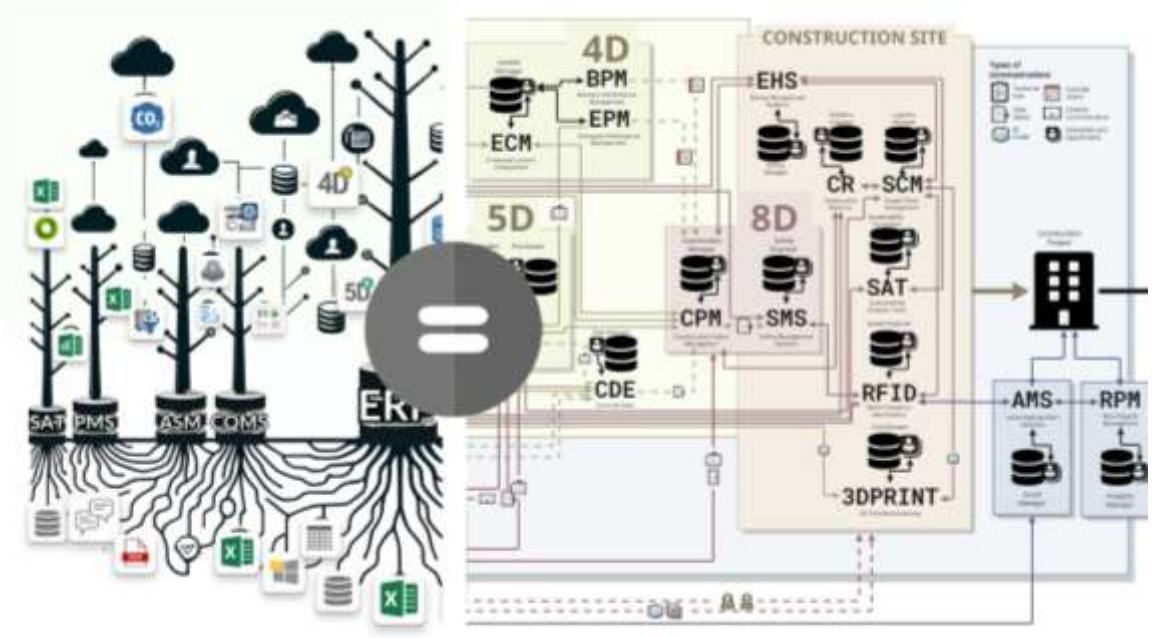


Рис. 4.2-2 Обеспечение здоровой жизнедеятельности экосистемы компании требует качественного и своевременного снабжения ресурсами ее систем.

Чтобы автоматизировать процесс проверки данных и унификации (для последующей автоматической интеграции) следует начать с описания минимально необходимых требований к данным для каждой конкретной системы. Эти требования определяют:

- Что именно нужно получать?
- В каком виде (структура, формат)?
- Какие атрибуты обязательны?
- Какие допуски по точности и полноте приемлемы?

Требования к данным описывают критерии качества, структуры и полноты получаемой и обрабатываемой информации. Например, для текстов в PDF-документах важно точное оформление в соответствие отраслевым стандартам (Рис. 7.2-14 - Рис. 7.2-16). Объекты в CAD-моделях должны иметь корректные атрибуты (размеры, коды, привязки к классификаторам) (Рис. 7.3-9, Рис. 7.3-10). А для сканов контрактов важны чёткие даты и возможность автоматического извлечения суммы и ключевых условий (Рис. 4.1-7 - Рис. 4.1-10).

> Формулирование требований к данным и автоматическая проверка их соответствия — один из самых трудоёмких, но критически важных этапов. Именно он чаще всего занимает наибольшую часть времени в бизнес-процессах.

Как уже упоминалось в третьей части книги, от 50% до 90% рабочего времени специалистов в области бизнес-аналитики (BI) уходит не на анализ, а на подготовку данных (Рис. 3.2-5). Этот процесс включает сбор, верификацию, валидацию, унификацию и структурирование данных.

Согласно опросу 2016 года [95], специалисты по обработке данных в самых разных областях широкого спектра заявили, что большую часть своего рабочего времени (около 80%) они тратят на то, что им меньше всего нравится делать (Рис. 4.2-3): собирать существующие наборы данных и организовывать (унифицировать, структурировать) их. Таким образом, менее 20% времени у них остается на творческие задачи, такие как поиск закономерностей и паттернов, которые приведут к новым инсайтам и открытиям.

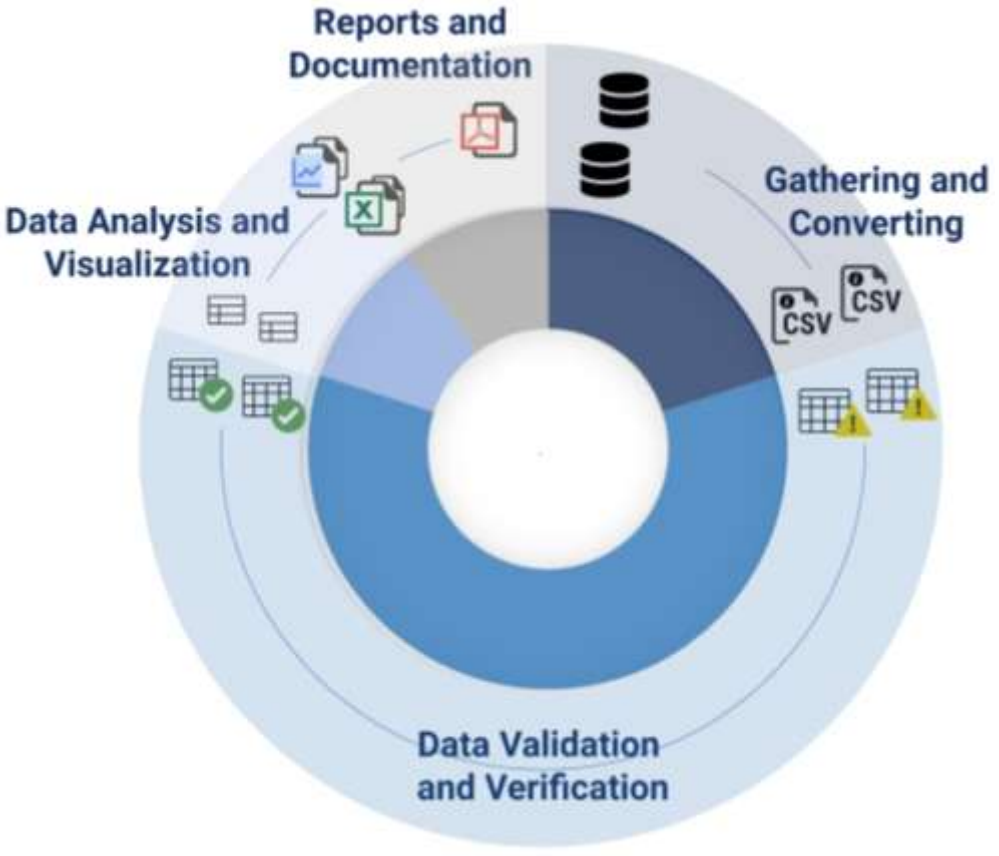


Рис. 4.2-3 Проверка и обеспечение качества данных - самый дорогостоящий, продолжительный и сложный этап в подготовке данных для интеграции в другие системы.

Успешное управление данными в строительной компании требует комплексного подхода, включающего параметризацию задач, формулирование требований к качеству данных и использование подходящих инструментов для их автоматизированной проверки.

Цифровая совместимость начинается с требований

С ростом количества цифровых систем внутри компаний возрастает и потребность в согласованности данных между ними. Менеджеры, отвечающие за различные ИТ-системы, нередко сталкиваются с тем, что не успевают за увеличивающимся объёмом информации и разнообразием форматов. В таких условиях они вынуждены запрашивать у специалистов создание данных в форме, подходящей для использования в других приложениях и платформах.

Это, в свою очередь, требует от инженеров и сотрудников, занимающихся формированием данных, адаптации под множество требований, зачастую без прозрачности и чёткого понимания того, где и как эти данные будут применяться в дальнейшем. Отсутствие стандартизированных подходов работе с информацией приводит к потерям эффективности и увеличению затрат на этапе проверки, которая часто, из сложности и нестандартизированности данных, осуществляется вручную.

> Вопрос стандартизации данных — это не только вопрос удобства или автоматизации. Это прямые финансовые потери. Согласно отчету IBM за 2016 год, ежегодные потери от низкого качества данных в США составляют 3,1 трлн долларов [96]. Дополнительно исследования MIT и других аналитических консалтинговых компаний показывают, что стоимость низкого качества данных может достигать 15–25% от выручки компании [97].

В этих условиях становится критически важным наличие чётко сформулированных требований к данным и описаниях того, какие параметры, в каком формате и с каким уровнем детализации должны быть включены в создаваемые объекты. Без формализации этих требований невозможно гарантировать качество и совместимость данных между системами и этапами проекта (Рис. 4.2-4).

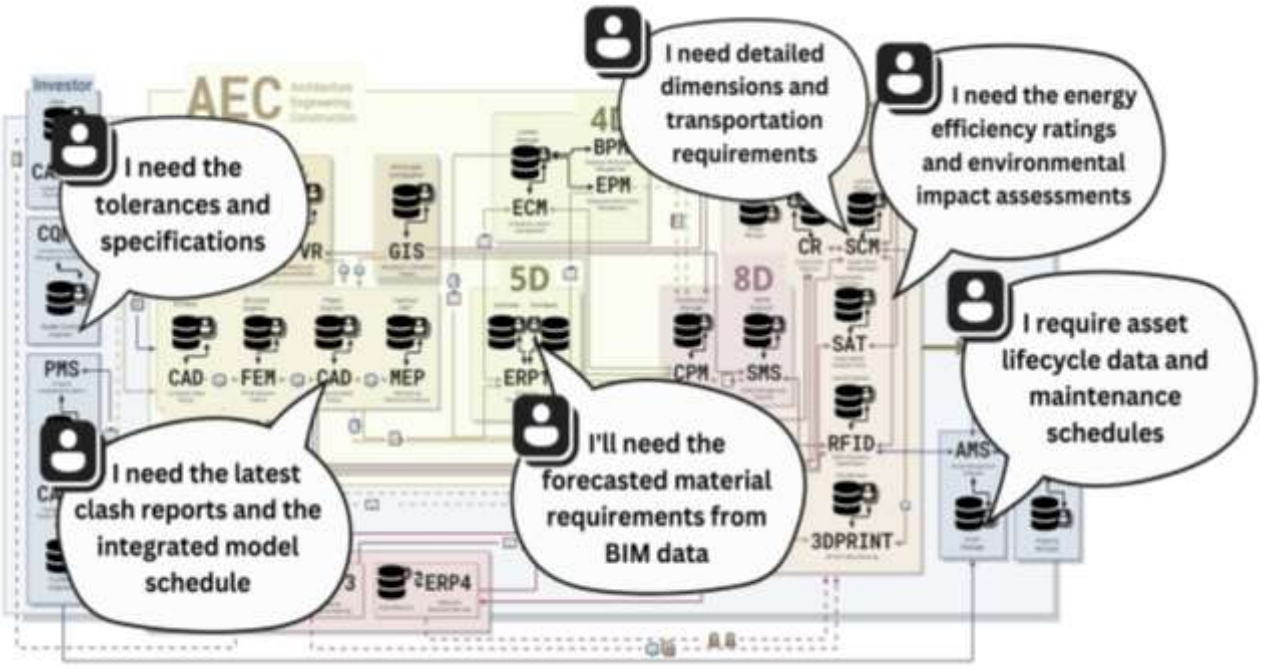


Рис. 4.2-4 Бизнес основан на взаимодействии различных ролей, каждая из которых требует определенных параметров и значений, критически важных для выполнения бизнес-задач.

Чтобы сформулировать корректные требования к данным, необходимо понимать бизнес-процессы на уровне данных. Строительные проекты различаются по типу, масштабу и количеству участников, а каждая система — будь то моделирование (CAD (BIM)), календарное планирование (ERP 4D), расчёт стоимости (ERP 5D) или логистики (SCM) — требует для входящих (вносимых сущностей-элементов) своих уникальных параметров.

В зависимости от этих потребностей бизнес-менеджеры должны либо проектировать новые структуры данных с учётом установленных требований, либо адаптировать уже существующие таблицы и базы данных. При этом качество создаваемых данных будет напрямую зависеть от того, насколько точно и корректно сформулированы требования (Рис. 4.2-5).

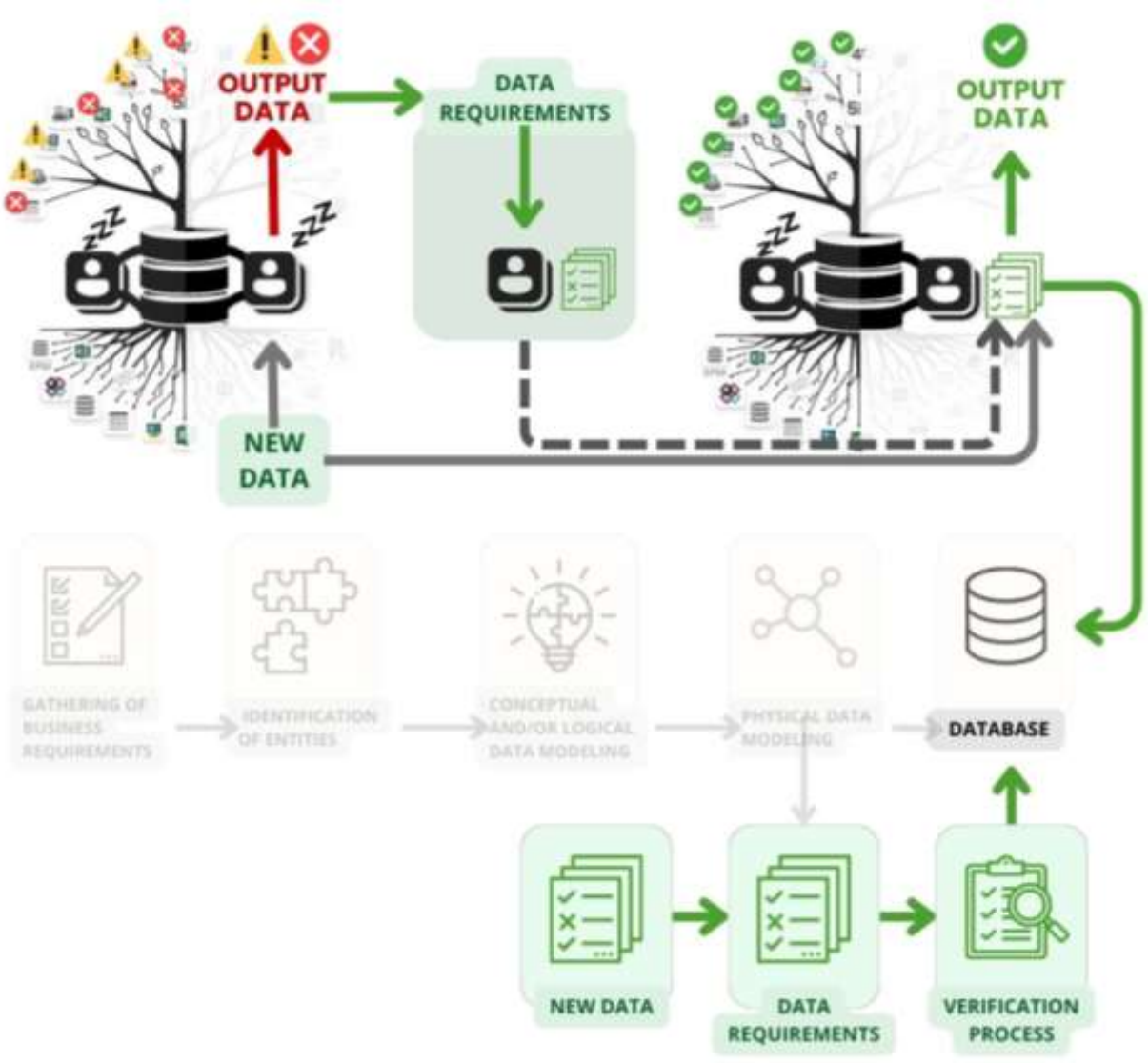


Рис. 4.2-5 Качество данных зависит от качества требований, которые создаются для конкретных случаев использования данных.

Поскольку каждая система предъявляет свои специфические требования к данным, первым шагом при формулировании общих требований должно стать категоризированние всех элементов, участвующих в бизнес-процессах. Это означает необходимость разделения объектов на классы и группы классов, соответствующие конкретным системам или прикладным задачам. Для каждой такой группы разрабатываются отдельные требования к структуре, атрибутам и качеству данных.

Однако на практике реализация этого подхода сталкивается с серьёзным затруднением: отсутствием единого языка группировки данных. Разрозненные классификации, дублирование идентификаторов и несовместимость форматов приводят к тому, что каждая компания, каждое программное обеспечение и даже каждый проект формируют собственные, изолированные модели данных и классы. В результате возникает цифровая «вавилонская башня», где для передачи информации между системами требуются многочисленные преобразования в нужные модели данных и классы, зачастую выполняемые вручную. Преодолеть этот барьер можно только через переход к универсальным классификаторам и стандартизированным наборам требований.

Единый язык строительства: роль классификаторов в цифровой трансформации

В контексте цифровизации и автоматизации процессов проверки и обработки особую роль играют системы классификации элементов — своеобразные "цифровые словари", обеспечивающие единообразие в описании и параметризации объектов. Именно классификаторы формируют тот «единый язык», который позволяет сгруппировать данные по смыслу использования и интегрировать данные между различными системами, уровнями управления и фазами жизненного цикла проекта.

Наиболее ощутимое влияние классификаторы оказывают в экономике жизненного цикла здания, где наиболее важным аспектом является оптимизация долгосрочных эксплуатационных затрат. Исследования показывают, что эксплуатационные расходы составляют до 80% совокупной стоимости владения зданием, что втрое превышает первоначальные затраты на строительство (Рис. 4.2-6) [98]. Это означает, что решение о будущих издержках во многом формируется ещё на этапе проектирования.

> Именно поэтому требования от инженеров эксплуатации (CAFM, AMS, PMS, RPM) должны становиться отправной точкой при формировании требований к данным на этапе проектирования (Рис. 1.2-4). Эти системы должны рассматриваться не как заключительная стадия проекта, а как неотъемлемая часть всей цифровой экосистемы проекта — от концепта до демонтажа.

Современный классификатор — это не просто система кодов для группировки. Это механизм взаимопонимания между архитекторами, инженерами, сметчиками, логистами, службами эксплуатации и ИТ-системами. Подобно тому, как автопилот машины должен однозначно распознавать дорожные объекты с высокой точностью, цифровые строительные системы и их пользователи должны через класс элемента однозначно интерпретировать для разных систем один и тот же элемент проекта.

Рис. 4.2-6 Эксплуатационные и технические расходы превышают стоимость строительства втрое, составляя 60–80% всех затрат жизненного цикла здания (по материалам [99]).

Уровень развития классификаторов непосредственно коррелирует с глубиной цифровизации компании и её цифровой зрелостью. Организации с низким уровнем цифровой зрелости сталкиваются с фрагментарностью данных, несовместимостью информационных систем и, как следствие, несовместимостью и неэффективностью классификаторов. В таких компаниях один и тот же элемент может часто иметь различные идентификаторы группировки в разных системах, что критически затрудняет конечную интеграцию и делает невозможным автоматизацию процессов.

Например, одно и тоже окно в проекте может быть обозначено по-разному в CAD модели, сметной и эксплуатационной системе (Рис. 4.2-7) из-за многоаспектности восприятия элементов разными участниками процесса. Для сметчика в элементе категории «Окна» важны объемы и стоимость, для эксплуатационной службы — доступность и ремонтопригодность, для архитектора — эстетические и функциональные характеристики. В итоге один и тот же элемент может требовать различных параметров.

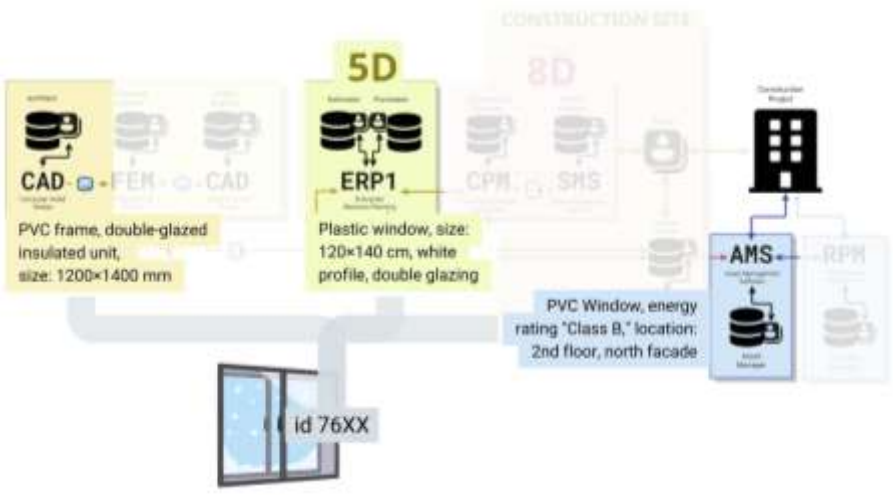


Рис. 4.2-7 При несогласованной классификации между системами элемент на каждом этапе перехода в другую систему будет терять часть атрибутивной информации.

Из-за сложности однозначного определения классификации строительных элементов специалисты разных областей часто присваивают одному и тому же элементу несовместимые между собой классы. Это приводит к потере единого представления об объекте, что требует последующего ручного вмешательства для согласования различных систем классификации и установления соответствия между типами и классами, определенными разными специалистами.

В результате подобной несогласованности документация по эксплуатации, полученная отделом закупок (ERP) при приобретении строительного элемента у производителя, зачастую не может быть корректно привязана к классификации этого элемента на строительной площадке (PMIS, SCM). Как следствие, критически важная информация с высокой вероятностью не интегрируется в системы управления инфраструктурой и активами (CAFM, AMS), что создает серьезные проблемы при вводе объекта в эксплуатацию, а также при последующем обслуживании (AMS, RPM) или замене данного элемента.

> В компаниях с высокой цифровой зрелостью классификаторы играют роль нервной системы, объединяющей все информационные потоки. Один и тот же элемент получает уникальный идентификатор, позволяющий передавать его между CAD, ERP, AMS и CAFM-системами и их классификаторами без искажения или потерь.

Для построения эффективных классификаторов необходимо понимать, как данные используются. Один и тот же инженер может по-разному называть и классифицировать элемент в разных проектах. Только собирая статистику использования годами, можно выработать устойчивую систему классификации. В этом помогает машинное обучение: алгоритмы анализируют тысячи проектов (Рис. 9.1-10), определяя через машинное обучение вероятные классы и параметры (Рис. 10.1-6). Автоматическая классификация особенно ценна в условиях, где ручная классификация невозможна из-за объёма данных. Системы автоматической классификации смогут различать базовые категории на основании минимально заполненных параметров элемента (подробнее в девятой и десятой части книги).

Развитые системы классификаторов становятся катализаторами дальнейшей цифровизации, создавая основу для:

- Автоматизированной оценки стоимости и сроков реализации проектов.
- Предиктивного анализа потенциальных рисков и коллизий
- Оптимизации закупочных процессов и логистических цепочек
- Создания цифровых двойников зданий и сооружений
- Интеграции с системами "умного города" и интернета вещей

> Время для трансформации ограничено — с развитием технологий машинного обучения и компьютерного зрения, проблема автоматической классификации, которая не могла решиться десятилетиями, будет решена в ближайшие годы, и строительные и проектировочные компании, не успевшие адаптироваться к этом, рискуют повторить судьбу таксопарков, вытесненных цифровыми платформами.

Подробнее про автоматизацию калькуляций стоимости и сроков, а также про большие данные и

машинное обучение будет рассказано в пятой и девятой частях книги. Вопросы риска повторения судьбы таксопарков и уберизации строительной отрасли подробно рассматриваются в десятой части книги.

Понимая ключевую роль классификаторов в цифровой трансформации строительной отрасли, необходимо обратиться к истории их эволюции. Именно исторический контекст позволяет осознать, как развивались подходы к классификации и какие тенденции определяют их современное состояние.

Masterformat, OmniClass, Uniclass и CoClass: эволюция классификационных системы

Исторически классификаторы строительных элементов и работ развивались в три поколения, каждое из которых отражало уровень доступных технологий и актуальные потребности отрасли в определённый период времени (Рис. 4.2-8):

- **Первое поколение** (начало 1950-х — конец 1980-х) — бумажные справочники, иерархические классификаторы, используемые локально (например, Masterformat, SfB).
- **Второе поколение** (конец 1990-х — середина 2010-х) — таблицы и структурированные базы данных, реализуемые в Excel и Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Третье поколение** (с 2010-х по настоящее время) — цифровые сервисы и API-интерфейсы, интеграция с CAD (BIM), автоматизация (Uniclass 2015, CoClass).

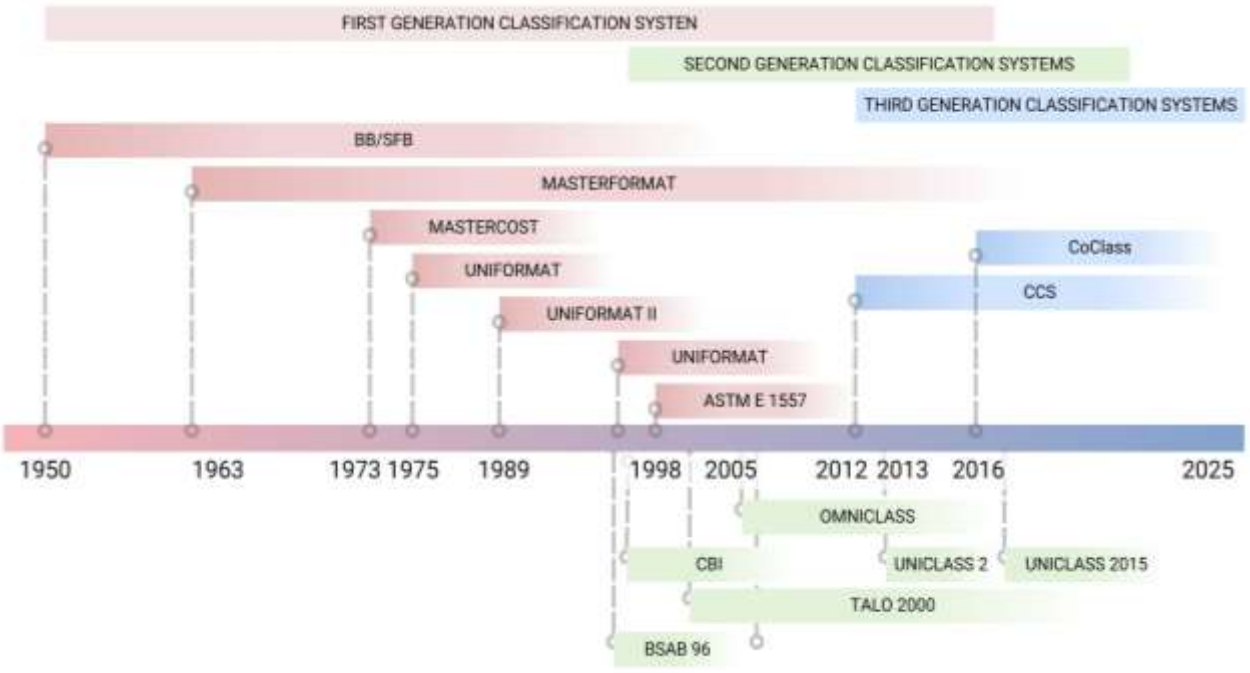


Рис. 4.2-8 Три поколения классификаторов строительной отрасли.

За последние десятилетия наблюдается сокращение иерархической сложности (Рис. 4.2-9) классификаторов: если ранние системы, такие как OmniClass, использовали до 7 уровней вложенности для описания 6887 классов, то современные решения, например CoClass, ограничиваются 3 уровнями при 750 классах. Это упрощает работу с данными, сохраняя при этом необходимую детализацию. Uniclass 2015, часто применяемый в Великобритании как стандарт, объединяет 7210 классов всего в 4 уровнях, что делает его удобным для CAD проектов и государственных закупок.

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

Рис. 4.2-9 С каждым новым поколением классификаторов сложность категоризации уменьшается в разы.

В системах строительной сметной оценки разных стран, из-за разности классификаций, даже такой типовой элемент, как бетонная фундаментная стена, может быть описан совершенно по-разному (Рис. 4.2-10). Эти различия отражают национальные особенности строительной практики, применяемые системы измерений, подходы к классификации материалов, а также нормативные и технические требования, действующие в каждой стране.

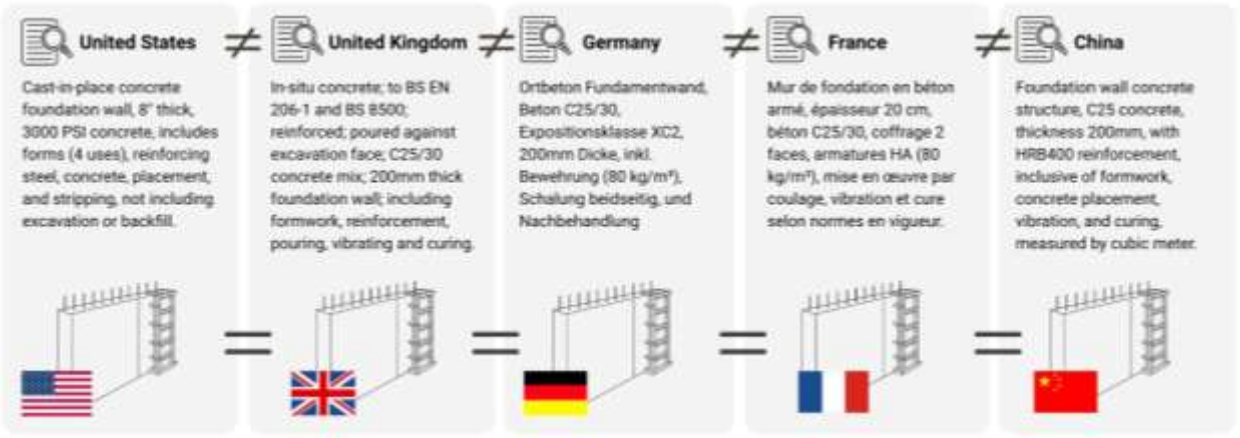


Рис. 4.2-10 Один и тот же элемент в разных странах используется в проектах через разные описания и классификации.

Разнообразие классификаций одних и тех же элементов усложняет международное сотрудничество, делает сопоставление стоимости и объёма работ в рамках интернациональных проектов трудоёмким, а иногда — и практически невозможным. На данный момент, на глобальном уровне нет одного универсального классификатора — каждая страна или регион разрабатывают собственные системы, ориентируясь на локальные нормы, язык и бизнес-культуру:

- **CCS** (Дания): Cost Classification System — система классификации затрат на протяжении всего жизненного цикла объекта (проектирование, строительство, эксплуатация). Акцент делается на логику эксплуатации и технического обслуживания, но также включает управление бюджетами и ресурсами.
- **NS 3451** (Норвегия): классифицирует объекты по функциям, конструктивным элементам и этапам жизненного цикла. Используется для управления проектами, оценки стоимости и долгосрочного планирования.
- **MasterFormat** (США): система для структурирования строительных спецификаций по разделам (например: бетон, электромонтаж, отделка). Фокус на дисциплины и типы работ, а не на функциональные элементы (в отличие от UniFormat).
- **Uniclass 2** (Великобритания): один из самых детализированных классификаторов, применяется в госзакупках и BIM-проектах. Объединяет данные по объектам, работам, материалам и пространствам в единую систему.
- **OmniClass**: международный стандарт (разработан CSI в США) для управления информацией об объектах: от библиотек компонентов до электронных спецификаций. Подходит для долгосрочного хранения данных, совместим с CAD (BIM) и другими цифровыми инструментами.
- **COBie**: Construction-Operation Building information exchange — международный стандарт обмена данными между этапами проектирования, строительства и эксплуатации. Включён в BS 1192–4:2014, как часть концепции «BIM-модели, готовой к эксплуатации». Фокус на передачу информации (например, спецификации оборудования, гарантии, контакты подрядчиков).

Глобализация строительной отрасли, скорее всего, приведет к постепенной унификации систем классификации строительных элементов, что существенно снизит зависимость от локальных национальных стандартов. Этот процесс может развиваться по аналогии с эволюцией интернет-коммуникаций, где универсальные протоколы передачи данных в конечном итоге вытеснили разрозненные локальные форматы, обеспечив глобальную совместимость систем.

Альтернативный путь развития может заключаться в непосредственном переходе к системам автоматической классификации на базе технологий машинного обучения. Данные технологии, разрабатываемые сегодня преимущественно в области автономного транспорта, обладают значительным потенциалом для применения к большим массивам данных CAD-проектирования (Рис. 10.1-6).

> Сегодня ситуация не ограничивается лишь национальной кластеризацией классификаторов. Из-за множества особенностей, не учтённых на государственном уровне, каждая компания вынуждена самостоятельно заниматься унификацией и стандартизацией категорий элементов и ресурсов, с которыми она работает.

Как правило, этот процесс начинается с малого — с локальных таблиц объектов или внутренних систем обозначений. Однако стратегической целью становится переход к единому языку описания всех элементов, который был бы понятен не только внутри компании, но и за её пределами — в идеале, согласованный с международными или отраслевыми классификаторами (Рис. 4.2-8). Такой подход облегчает интеграцию с внешними партнёрами, цифровыми системами и способствует формированию единых сквозных процессов в рамках жизненного цикла объектов.

Перед переходом к автоматизации и масштабируемым ИТ-системам необходимо либо использовать классификаторы национального уровня, либо выстроить собственную, логичную и однозначную структуру идентификации элементов. Каждый объект — будь то окно (Рис. 4.2-11), дверь или

инженерная система — должен быть описан таким образом, чтобы его можно было безошибочно распознать в любой цифровой системе компании. Это критически важно при переходе от плоских чертежей к цифровым моделям, охватывающим как этап проектирования, так и эксплуатацию зданий.

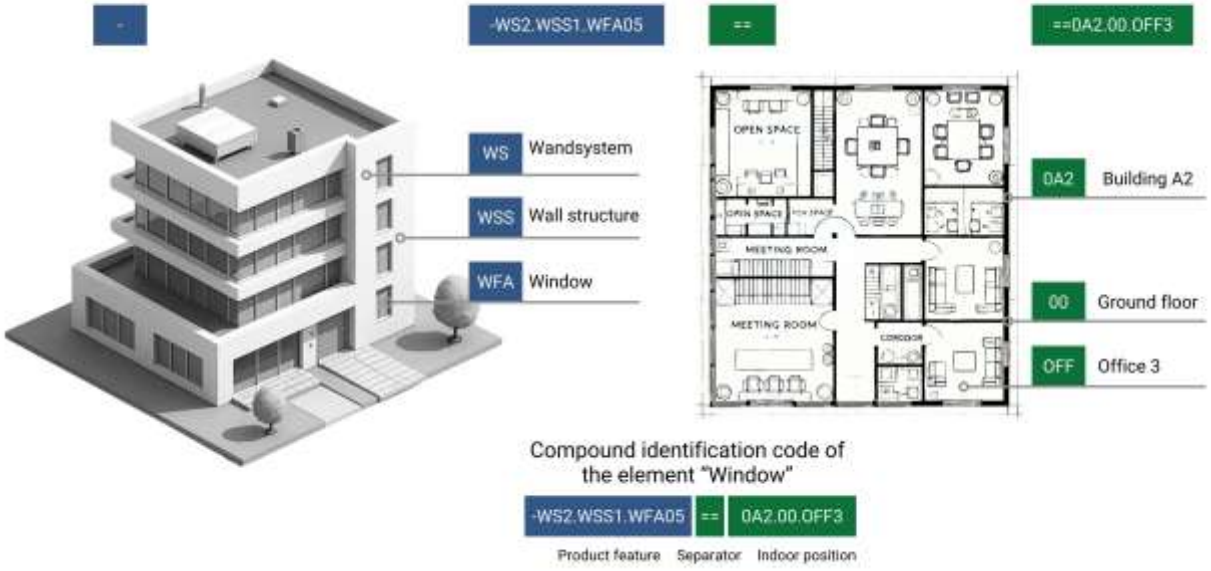


Рис. 4.2-11 Пример составного идентификатора строительного элемента окна на основе классификации и позиции в здании.

Одним из примеров внутренних классификаторов может быть разработка составного кода идентификации (Рис. 4.2-11). Такой код объединяет несколько уровней информации: функциональное назначение элемента (например, "окно в стене"), его тип, а также точную пространственную привязку — здание A2, этаж 0, помещение 3. Подобная многоуровневая структура позволяет создать единую систему навигации по цифровым моделям и документации, особенно на этапах проверки и трансформации данных, где необходима однозначная группировка элементов. Однозначность распознавания элемента обеспечивает согласованность между отделами и снижает риски дублирования, ошибок и потери информации.

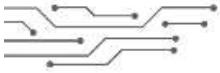
Хорошо выстроенный классификатор — это не просто технический документ, это фундамент цифровой экосистемы компании:

- обеспечивает совместимость данных между системами;
- сокращает затраты на поиск и переработку информации;
- повышает прозрачность и управляемость;
- создаёт основу для масштабирования и автоматизации.

Стандартизированное описание объектов, через применение национальных классификаторов или собственных составных кодов идентификации становится основой для консистентных данных,

надёжного обмена информацией и последующего внедрения интеллектуальных сервисов — от автоматизированных закупок до цифровых двойников.

После завершения этапа структуризации разноформатных данных и выбора классификатора, который будет использоваться для распознавания и группировки элементов, следующим шагом становится корректное моделирование данных. Этот процесс включает в себя определение ключевых параметров, построение логичной структуры данных и описание взаимосвязей между элементами.



ГЛАВА 4.3.
МОДЕЛИРОВАНИЕ ДАННЫХ И ЦЕНТР ПЕРЕДОВОГО ОПЫТА

Моделирование данных: концептуальная, логическая и физическая модель

Эффективное управление данными (структурированными и классифицированными нами ранее) невозможно без продуманной структуры хранения и обработки. Для обеспечения доступа и согласованности информации на этапах хранения и обработки компании используют моделирование данных — методологию, позволяющую проектировать таблицы, базы данных и связи между ними в соответствии с бизнес-требованиями.

> Моделирование данных — это фундамент, на котором строится любая цифровая экосистема. Без описания систем, требований и моделирования данных, инженеры и специалисты, создающие данные, не знают и не понимают, где созданные ими данные будут использоваться.

Как и при строительстве здания, где невозможно начать укладку кирпичей без плана, создание системы хранения данных требует чёткого понимания того, какие данные будут использоваться, как они связаны между собой, и кто будет с ними работать. Без описания процессов и требований инженеры и специалисты, создающие данные, теряют из виду, где и как эти данные будут применяться в дальнейшем.

Модель данных служит связующим звеном между бизнесом и ИТ. Она позволяет формализовать требования, структурировать информацию и облегчить коммуникацию между заинтересованными сторонами. В этом смысле моделирование данных похоже на работу архитектора, который по замыслу заказчика разрабатывает план здания, а затем передаёт его строителям — администраторам и разработчикам баз данных — для воплощения (создания баз данных).

Таким образом, каждая строительная компания, помимо структуризации и категоризации элементов и ресурсов (Рис. 4.2-11), должна овладеть искусством "строительства" баз данных (таблиц) и научиться создавать связи между ними, словно соединяя кирпичики в надежную и крепкую стену знаний из данных компании. Ключевые понятия в моделировании данных (Рис. 4.3-1) включают:

- **Сущности** — это объекты, данные о которых необходимо собирать. На раннем этапе проектирования сущностью может быть отдельный элемент (например, "дверь"), а в сметной модели — группа элементов, объединённых по категориям (например, "внутренние двери").
- **Атрибуты** — характеристики сущностей, описывающие важные детали: размеры, свойства, стоимость сборки, логистику и другие параметры.
- **Отношения (связи)** — показывают, как сущности взаимодействуют между собой. Они могут быть одного из типов: "один к одному", "многие к одному", "многие ко многим".
- **ER-диаграммы** (Entity-Relationship diagrams) — визуальные схемы, на которых показаны

сущности, атрибуты и связи между ними. ER-диаграммы бывают концептуальными, логическими и физическими — каждая отражает свой уровень детализации.

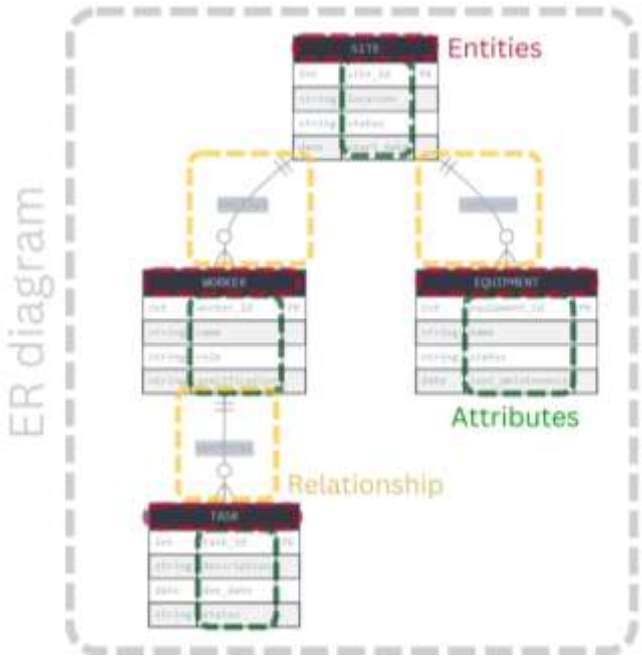


Рис. 4.3-1 ER-диаграмма структуры концептуальной базы данных с сущностями, атрибутами и связями.

Процесс проектирования данных и определения взаимосвязей между ними традиционно разделяется на три основные модели. Каждая из них выполняет определённые функции, отличаясь уровнем детализации и степенью абстракции при представлении структуры данных:

- **Концептуальная модель данных**: эта модель описывает основные сущности и их взаимосвязи, не вдаваясь в подробности атрибутов. Обычно она используется на начальных этапах планирования. На этом этапе мы можем делать наброски из баз данных и систем, чтобы показать связь между различными отделами и специалистами.

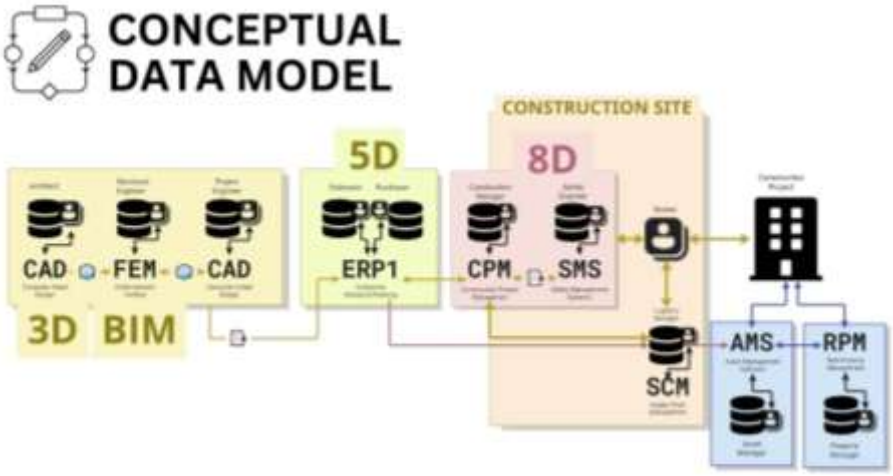


Рис. 4.3-2 Концептуальная диаграмма описывает содержание системы: высокоуровневое представление отношений, без технических деталей.

- **Логическая модель данных**: основываясь на концептуальной модели, логическая модель данных включает в себя подробное описание сущностей, атрибутов, ключей и отношений, отображая бизнес-информацию и правила.

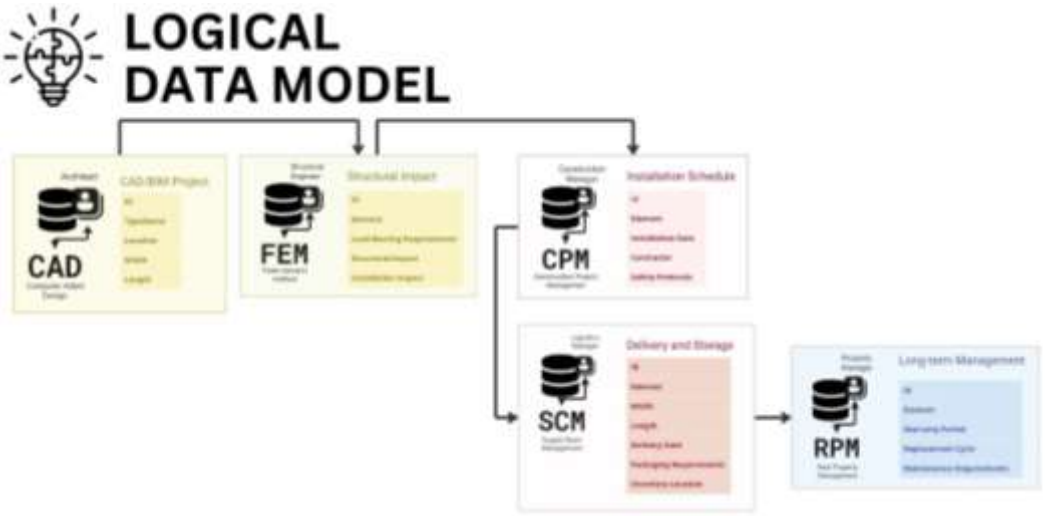


Рис. 4.3-3 Логическая модель данных подробно описывает типы данных, отношения и ключи, но без системной реализации.

- **Физическая модель данных**: эта модель описывает необходимые структуры для реализации базы данных, включая таблицы, столбцы и отношения. Она фокусируется на производительности базы данных, стратегиях индексирования и физическом хранении для оптимизации физического развертывания баз данных.

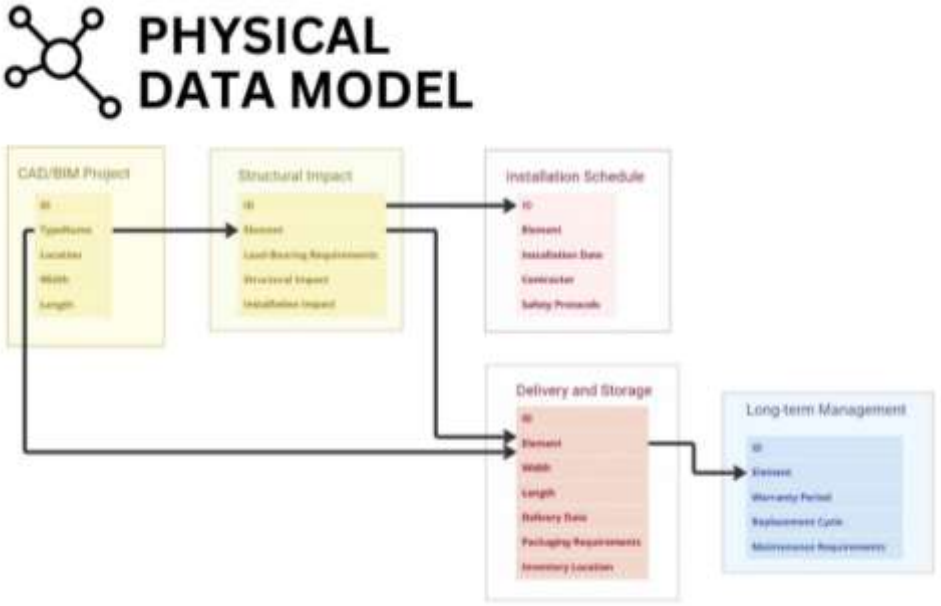


Рис. 4.3-4 Физическая модель данных определяет, как будет реализована система, включая таблицы и специфические детали базы данных.

При разработке баз данных и проектировании табличных отношений понимание уровней абстракции играет ключевую роль в построении эффективной архитектуры системы.

Эффективная методология моделирования данных позволяет объединить бизнес-задачи с технической реализацией, делая всю цепочку процессов более прозрачной и управляемой. Моделирование данных — это не разовая задача, а процесс, включающий последовательные шаги (Рис. 4.3-5):

- **Сбор бизнес-требований:** определяются ключевые задачи, цели и информационные потоки. Это этап активного взаимодействия с экспертами и пользователями.
- **Идентификация сущностей:** выделяются основные объекты, категории и типы данных, которые важно учитывать в будущей системе.
- **Разработка концептуальной и логической модели:** сначала фиксируются ключевые сущности и их связи, затем — атрибуты, правила и детальная структура.
- **Физическое моделирование:** проектируется техническая реализация модели: таблицы, поля, связи, ограничения, индексы.
- **Создание базы данных:** финальный шаг — реализация физической модели в выбранной СУБД, проведение тестирования и подготовка к эксплуатации.

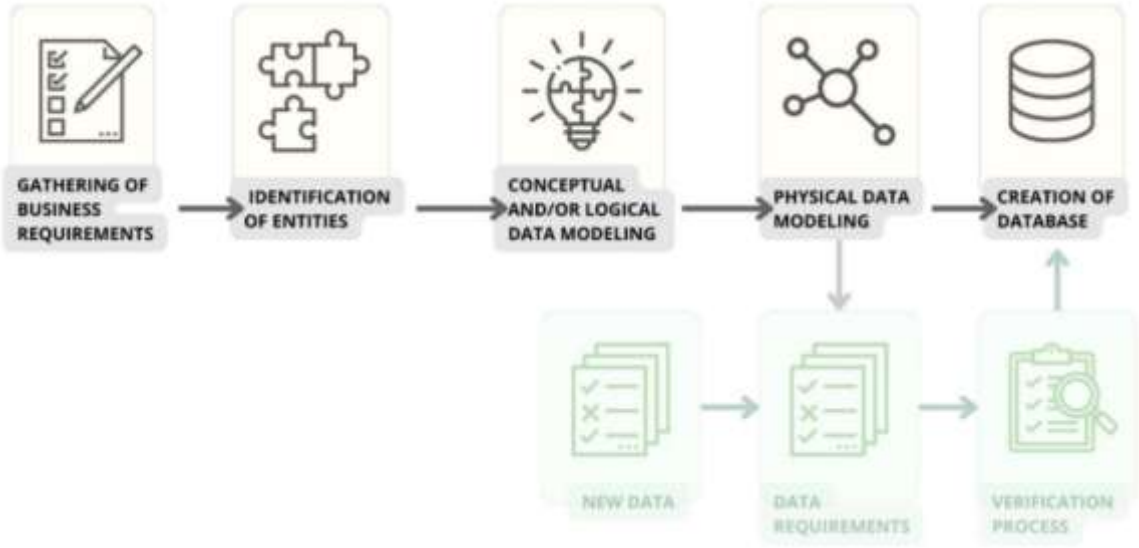


Рис. 4.3-5 Создание баз данных и систем управления данными для бизнес-процессов начинается с формирования требований и моделирования данных.

Грамотно выстроенные процессы моделирования данных позволяют обеспечить прозрачность потоков информации, что особенно важно в таких комплексных проектах, как управление строительным проектом или строительной площадкой. Рассмотрим, как переход от концептуальной модели к логической, а затем к физической, помогает упорядочить процессы.

Практическое моделирование данных в контексте строительства

Возьмём для примера моделирования данных задачу управления строительной площадкой и преобразуем требования прораба в структурированную логическую модель. Опираясь на базовые потребности управления стройкой, мы определяем ключевые сущности для: строительная площадка (SITE), рабочие (WORKER), оборудование (EQUIPMENT), задачи (TASK) и использование оборудования (EQUIPMENT_USAGE). Каждая сущность содержит набор атрибутов, отражающих важные характеристики. Например, для TASK это может быть описание задачи, срок выполнения, статус, приоритет; для WORKER — имя, его роль на площадке, текущая занятость и т. д.

В логической модели устанавливаются связи между этими сущностями, показывая, как они взаимодействуют друг с другом в реальных рабочих процессах (Рис. 4.3-6). Например, связь между площадкой и рабочими указывает на то, что на одной площадке может работать много рабочих, а связь между рабочими и задачами отражает, что один рабочий может выполнять несколько задач.

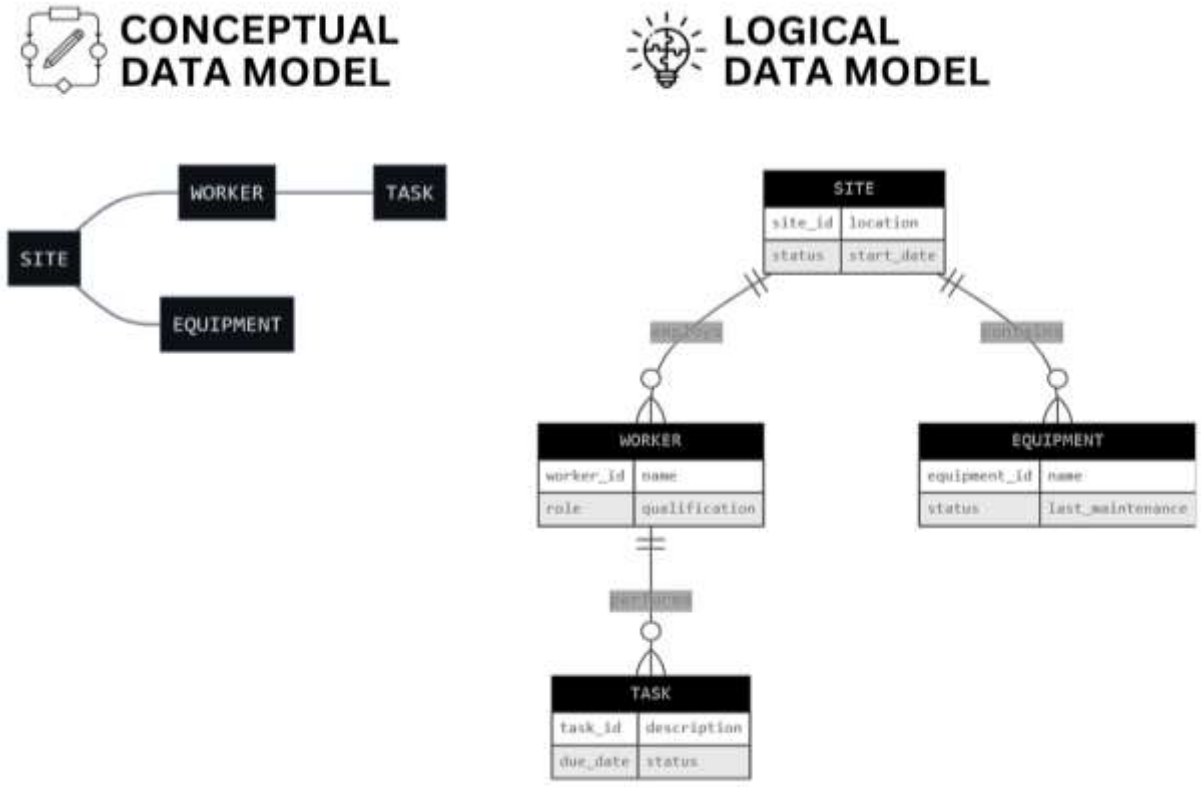


Рис. 4.3-6 Концептуальная и логическая модель данных, сформированная требованиями прораба для описания процессов на строительной площадке.

При переходе к физической модели добавляются технические детали реализации: конкретные типы данных (VARCHAR, INT, DATE), первичные и внешние ключи для связей между таблицами, а также индексы для оптимизации производительности базы данных (Рис. 4.3-7).

Например, для статусов необходимо определить конкретные типы с возможными значениями, а для улучшения производительности поиска необходимо добавить индексы по ключевым полям, таким как status и worker_id. Это превращает логическое описание системы в конкретный план реализации базы данных, готовый к созданию и внедрению.

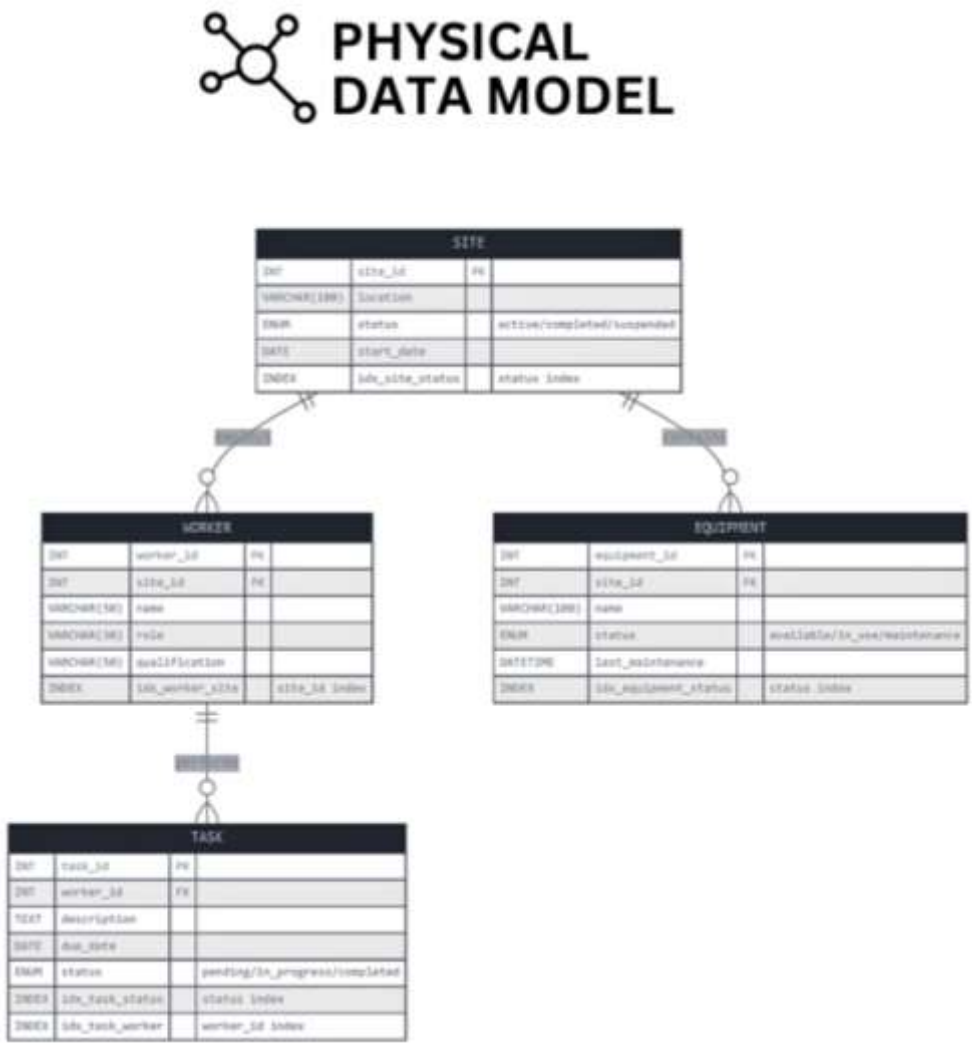


Рис. 4.3-7 Физическая модель данных описывает сущности строительной площадки через минимально необходимые параметры.

Физическая модель часто отличается от логической. В среднем распределение времени на моделирование выглядит так: около 50% уходит на концептуальную модель (сбор требований, обсуждение процессов, выявление сущностей), 10% — на логическую (уточнение атрибутов и связей) и 40% — на физическую (реализация, тестирование, адаптация под СУБД).

Такой баланс объясняется тем, что концептуальный этап закладывает основу структуры данных, а логическая модель лишь уточняет связи и атрибуты. Наибольшие ресурсы требуются на физическую модель, так как именно на этом этапе данные внедряются в конкретные платформы и инструменты.

Создание базы данных с помощью LLM

Имея модель данных и описание сущностей через параметры, мы готовы к созданию баз данных – хранилищ, где мы будем хранить информацию, поступающую после этапа структуризации по конкретным процессам.

Попробуем создать пример простой, но функциональной базы данных с минимальным количеством кода с помощью SQLite на примере языка программирования Python. Реляционные базы данных были подробно рассмотрены в главе "Структурированные реляционные базы данных и язык запросов SQL".

> SQLite — это легкая, open source, встраиваемая, кроссплатформенная база данных, поддерживающая основные функции SQL. SQLite широко используется в CAD (BIM), ERP, EPM и присутствует практически в каждом смартфоне.

Попросим LLM модель создать для нас небольшую базу данных SQLlite на основе физической модели, созданной выше (Рис. 4.3-4) для систем FEM (структурное воздействие) и CPM (график монтажа).

- Отправьте текстовый запрос в LLM чат (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любой другой):

> Создай базы данных SQLlite для систем "Конструктивное воздействие" и "График установки" с рисунка во вложении (вставьте в чат схему как изображение (Рис. 4.3-4)) ⏎

Ответ LLM:

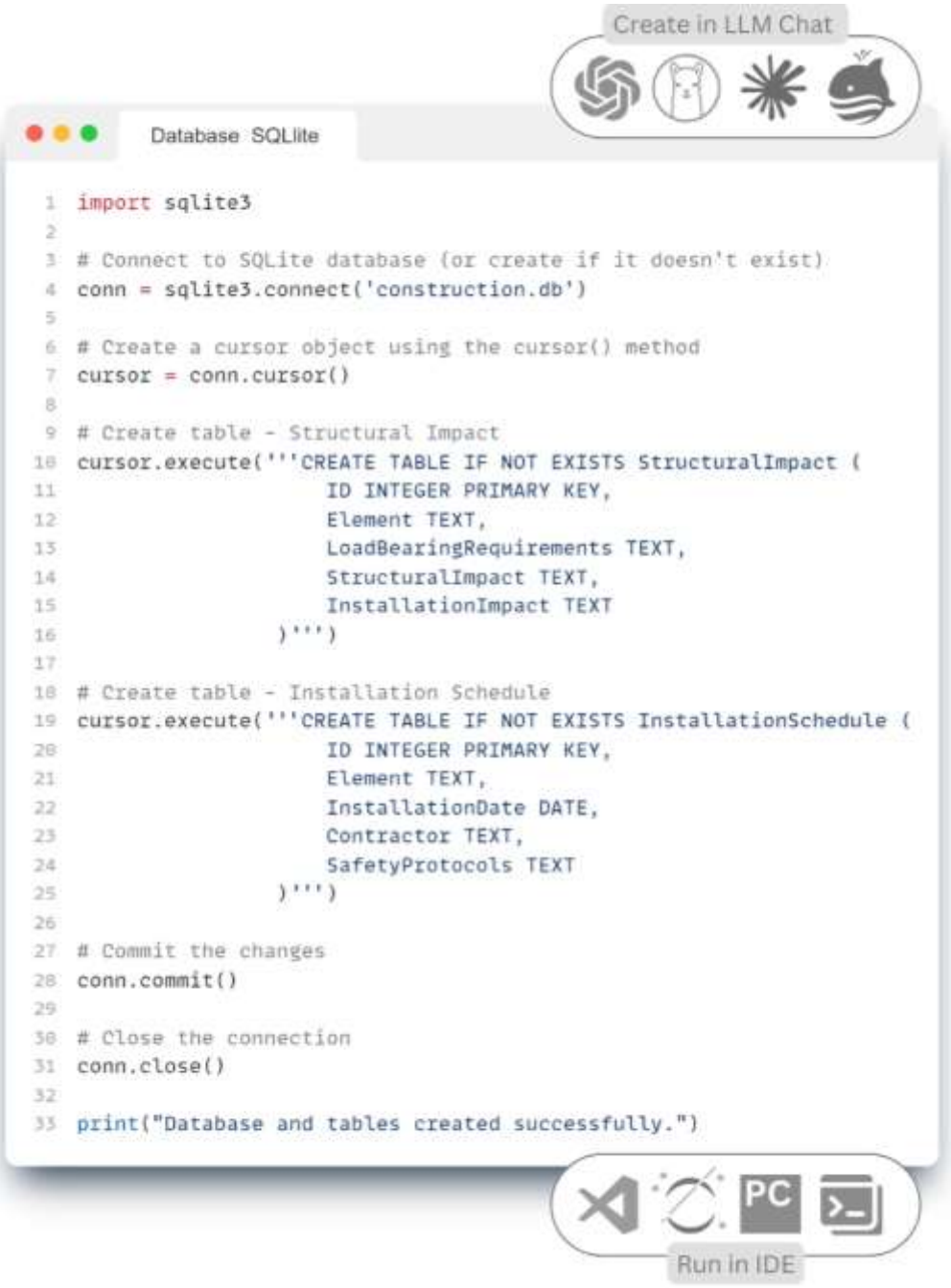


Рис. 4.3-8 Сгенерированный код создает базы данных и таблицы из загруженного изображения, в которые мы теперь можем загружать данные.

Этот код для создания базы данных может быть выполнен в среде Python с поддержкой SQLite, которая обычно включена в стандартное окружение Python. Его можно запустить на любой локальной машине с Python, способной запускать скрипты и создавать файлы. После запуска и выполнения этот код создаст на жестком диске файл - базу данных SQLite с именем construction.db (похоже на файл Excel с рабочими листами), которая будет содержать таблицы "Structural Impact" и "Installation Schedule".

После создания в эти таблицы новой базы данных **construction.db** через SQL_запросы или импорт можно будет добавлять данные, что позволит позже приступить к созданию автоматической обработки данных. Данные можно импортировать в базу данных SQLite из файлов CSV, электронных таблиц Excel или через API экспортировать из других баз данных и хранилищ.

Чтобы наладить устойчивые процессы моделирования данных и эффективного управления базами данных, компании необходима чётко сформулированная стратегия, а также налаженная координация между техническими и бизнес-командами. В условиях разрозненных проектов и множества источников данных часто сложно обеспечить согласованность, стандартизацию и контроль качества на всех уровнях. Одним из ключевых решений может стать создание внутри компании специализированного Центра передового опыта по моделированию данных (Data Modeling Center of Excellence, COE).

Центр передового опыта (CoE) по моделированию данных

В условиях, когда данные становятся одним из ключевых стратегических активов, компаниям необходимо не просто правильно собирать и хранить информацию — важно научиться управлять данными системно. Центр передового опыта по классификации и моделированию данных (Center of Excellence, CoE) — это структурное подразделение, обеспечивающее согласованность, качество и эффективность всей работы с данными в организации.

> Центр передового опыта (CoE) – это ядро экспертной поддержки и методологическим фундаментом для цифровых преобразований в компании. Он формирует культуру работы с данными и позволяет организациям выстроить процессы, принимая решения не на основе интуиции или локальной информации, а на основе структурированных, проверенных и репрезентативных данных.

Центр передового опыта по данным обычно формируется из кросс-функциональных команд, которые работают по принципу «две пиццы». Этот принцип, предложенный Джеффом Безосом, означает, что размер команды должен быть таким, чтобы её можно было накормить двумя пиццами, то есть не превышать 6–10 человек. Такой подход помогает избежать чрезмерной бюрократии и повышает гибкость работы. Команда CoE должна включать в себя сотрудников с разнообразными техническими навыками: от аналитики данных и машинного обучения до экспертизы в конкретных бизнес-областях. Обладая глубокими техническими знаниями, инженеры по обработке данных должны не только оптимизировать процессы и моделировать данные, но и поддерживать коллег, сокращая время на рутинные задачи (Рис. 4.3-9).

> Как в природе устойчивость экосистемы обеспечивается биологическим разнообразием, так и в цифровом мире гибкость и адаптивность достигаются через разнообразие подходов к работе с данными. Однако это разнообразие должно опираться на единые правила и понятия.

Центр передового опыта (CoE) можно сравнить с "климатическими условиями" лесной экосистемы, которые определяют, какие виды данных будут процветать, а какие — автоматически отсеиваться. Создавая благоприятный "климат" для качественных данных, CoE способствует естественному отбору лучших практик и методологий, которые в дальнейшем становятся стандартами организации.

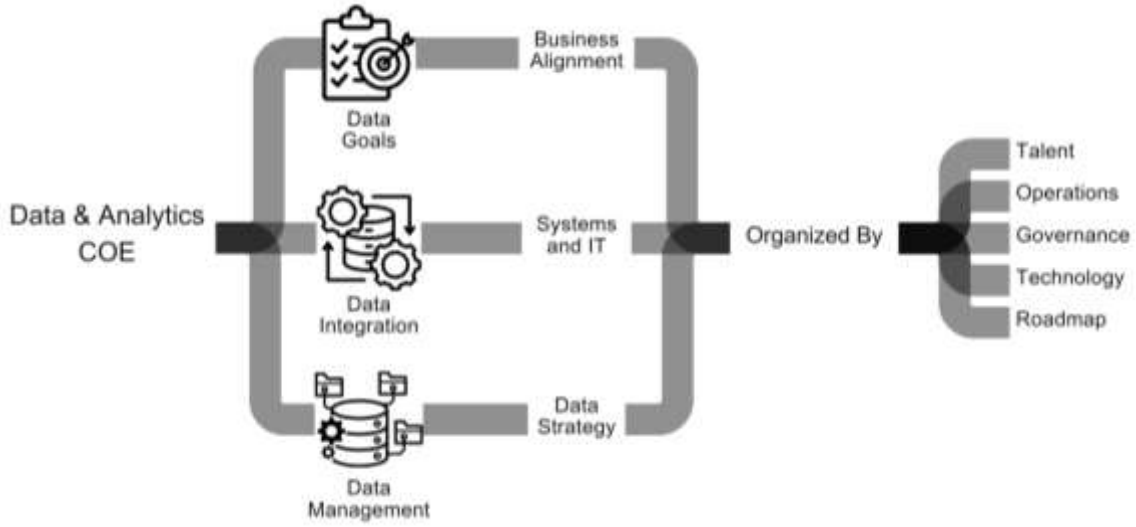


Рис. 4.3-9 Центр передового опыта (CoE) по данным и аналитике объединяет экспертизу по ключевым аспектам управления данными, их интеграции и выработки стратегии.

Для ускорения интеграционных циклов и достижения лучшего результата, CoE должен предоставлять своим членам достаточную степень автономии в принятии решений. Это особенно важно в динамичной среде, где метод проб и ошибок, постоянная обратная связь и частые релизы могут принести значительную выгоду. При этом данная автономия эффективна только при наличии чётких коммуникаций и поддержки со стороны высшего руководства. Без стратегического видения и координации сверху даже самая компетентная команда может столкнуться с барьерами при внедрении своих инициатив.

> Именно COE или высшее руководство компании отвечает за то, чтобы подход к моделированию данных не ограничивался одним-двумя проектами, а был встроен в общую систему управления информацией и бизнес-процессами.

Центр экспертизы (CoE) помимо задач, связанных с моделированием данных и управлением ими (Data Governance), отвечает за выработку единых стандартов и подходов к развёртыванию и эксплуатации инфраструктуры данных. Кроме того, он формирует культуру постоянного улучшения, оптимизации процессов и эффективного использования данных в организации (Рис. 4.3-10).

Системный подход к управлению данными и моделями внутри CoE можно условно разделить на несколько ключевых блоков:

- **Стандартизация процессов и управление жизненным циклом моделей:** CoE разрабатывает и внедряет методологии, позволяющие унифицировать создание и управление моделями данных. Это включает: формирование структурных шаблонов, методов контроля качества и систем управления версиями, обеспечивающих непрерывность данных на всех этапах работы.
- **Управление ролями и распределение ответственности:** в рамках COE определяются ключевые роли в процессе моделирования данных. Каждый участник проекта получает четко определенные функции и зоны ответственности, что способствует слаженной работе команд и снижает риски несоответствия данных.
- **Контроль качества и аудит:** эффективное управление строительными данными требует постоянного мониторинга их качества. Внедряются автоматизированные механизмы проверки данных, выявления ошибок, недостающих атрибутов.
- **Управление метаданными и архитектурой информации:** CoE отвечает за создание единой системы классификации и идентификаторов, стандартов наименования и описания сущностей, что критично для интеграции между системами.

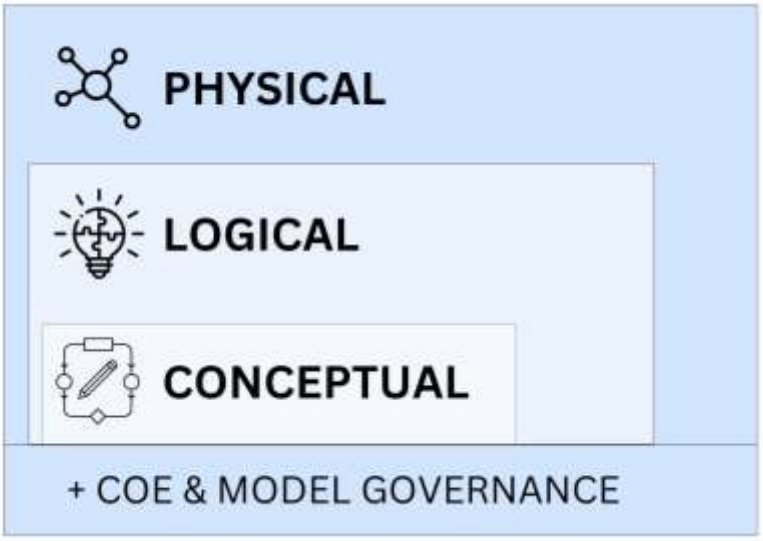


Рис. 4.3-10 Моделирование данных и управление качеством данных являются одной из главных задач CoE.

Центр передового опыта (CoE) по данным — это не просто группа экспертов, а системный механизм, создающий новую data-driven культуру и обеспечивающий единый подход к работе с данными во всей компании. Благодаря грамотной интеграции процессов моделирования в общую систему управления информацией, стандартизации, классификации и контролю качества данных, CoE помогает бизнесу постоянно совершенствовать свои продукты и бизнес-процессы, быстрее реагировать на изменения рынка и принимать взвешенные решения на основе достоверной аналитики.

Подобные центры особенно эффективны в сочетании с современными принципами DataOps — под-

ходом, обеспечивающим непрерывную доставку, автоматизацию и контроль качества данных. Подробнее о DataOps мы поговорим в восьмой части, в главе «Современные технологии работы с данными в строительной отрасли».

В следующих главах мы перейдём от стратегии к практике — условно «перевоплотимся» в центр обработки данных: рассмотрим на нескольких примерах, как происходит параметризация задачи, сбор требований и автоматический процесс валидации.

ГЛАВА 4.4.
СИСТЕМАТИЗАЦИЯ ТРЕБОВАНИЙ И ВАЛИДАЦИЯ ИНФОРМАЦИИ

Сбор и анализ требований: преобразование коммуникаций в структурированные данные

Сбор и управление требованиями — это первый шаг к обеспечению качества данных. Несмотря на развитие цифровых инструментов, большинство требований до сих пор формулируется в неструктурированном виде: через письма, протоколы встреч, телефонные звонки и устные обсуждения. Такая форма коммуникации затрудняет автоматизацию, проверку и повторное использование информации. В данной главе мы рассмотрим, как переводить текстовые требования в формальные структуры, обеспечивая прозрачность и системность бизнес-задач.

> Исследования компании Gartner «Качество данных: Лучшие практики для получения точных сведений» подчеркивает критическую важность качества данных для реализации успешных инициатив в области данных и аналитики [100]. Они отмечают, что низкое качество данных обходится организациям в среднем не менее чем в 12,9 миллиона долларов ежегодно и что достоверные, высококачественные данные необходимы для создания компании, управляемой данными.

Отсутствие структурированных требований приводит к тому, что один и тот же элемент (сущность) и его параметры может храниться в разных системах в различных вариациях. Это не только снижает эффективность процессов, но и приводит к потере времени, дублированию информации и необходимости повторной проверки данных перед их использованием. В результате даже одно упущение — потерянный параметр или один некорректно описанный элемент — может замедлить принятие решений и вызвать неэффективное расходование ресурсов.

> *За неимением гвоздя подкова был потеряна.*
> *За неимением подковы лошадь была потеряна.*
> *За неимением лошади потерялся всадник.*
> *За неимением всадника было потеряно послание.*
> *За неимением послания битва была проиграна.*
> *За неимением битвы было потеряно королевство.*
> *И все из-за отсутствия гвоздя в подкове.*
>
> — Пословица *[101]*

Анализ и сбор требований к процессу заполнения и хранения данных начинается с выявления всех заинтересованных сторон. Подобно тому, как в пословице потеря одного гвоздя приводит к цепочке критических последствий, в бизнесе — утрата одного участника, пропущенное требование или потеря даже одного параметра могут существенно повлиять не только на отдельный бизнес-процесс, но и на всю экосистему проекта и организации в целом. Поэтому крайне важно определить даже те элементы, параметры и роли, которые на первый взгляд кажутся незначительными,

но впоследствии могут оказаться критически важными для устойчивости бизнеса.

Представим, что у компании есть проект, в котором клиент выдвигает новую просьбу - "*добавить дополнительное окно на северной стороне здания*". В небольшой процесс "*запрос клиента на добавление нового окна в текущий проект*" вовлечены архитектор, заказчик, специалист по CAD (BIM), менеджер по строительству, менеджер по логистике, ERP-аналитик, инженер по контролю качества, инженер по безопасности, менеджер по контролю и менеджер по недвижимости.

> Даже в небольшом процессе могут участвовать десятки различных специалистов. Каждый участник процесса должен понимать требования специалистов, с которыми он связан на уровне данных.

На текстовом уровне (Рис. 4.4-1) коммуникация между клиентом и специалистами в цепочке процесса происходит следующим образом:

- **Заказчик:** "*Мы решили добавить дополнительное окно на северной стороне для лучшего освещения. Можно ли это реализовать*?"
- **Архитектор:** "*Конечно, я пересмотрю проект, чтобы включить новое окно, и вышлю обновленные планы CAD (BIM)*".
- **Специалист по CAD (BIM):** "*Получил новый проект. Я обновляю CAD (BIM) модель с дополнительным окном и после согласования с инженером FEM предоставляю точное расположение и размеры нового окна*".
- **Менеджер по строительству**: "*Получен новый проект. Мы корректируем сроки установки 4D и информируем всех соответствующих субподрядчиков*".
- **Инженер по объектам (CAFM)**: "*Я введу данные 6D о новом окне в систему CAFM для будущего управления объектом и планирования технического обслуживания*".
- **Менеджер по логистике**: "*Мне нужны размеры и вес нового окна, чтобы организовать доставку окна на объект*".
- **ERP-аналитик**: "*Мне нужны таблицы объемов и точный тип окна для обновления бюджета 5D в нашей ERP-системе, чтобы отразить стоимость нового окна в общей смете проекта*".
- **Инженер по контролю качества**: "*Как только спецификации окон будут готовы, я удостоверюсь, что они соответствуют нашим стандартам качества и материалов*".
- **Инженер по безопасности**: "*Я буду оценивать аспекты безопасности нового окна, уделяя особое внимание соблюдению требований и эвакуации по схеме 8D*".
- **Менеджер по контролю**: "*Основываясь на точном объеме работ от ERP, мы обновим нашу временную шкалу 4D, чтобы отразить установку нового окна, и сохраним новые данные в системе управления контентом проекта*".
- **Рабочий (монтажник)**: "*Нужны инструкции по установке, сборке и срокам выполнения работ. Кроме того, введены какие-нибудь особые правила безопасности, которые я должен соблюдать?*"
- **Управляющий недвижимостью**: "*После установки я буду документировать информацию о гарантии и обслуживании для долгосрочного управления*".
- **Менеджер активов**: "*Инженер по оборудованию, пожалуйста, отправьте окончательные данные для отслеживания активов и управления жизненным циклом.*"
- **Клиент**: "*Подождите, может, я тороплюсь, и окно не понадобится. Может, стоит сделать балкон*".

В подобных сценариях, которые случаются часто, даже небольшое изменение вызывает цепную реакцию между множеством систем и ролей. При этом почти вся коммуникация на начальном

этапе ведётся в текстовой форме: письма, чаты, протоколы встреч (Рис. 4.4-1).

В такой системе текстовых коммуникаций для строительного проекта очень важна система юридического подтверждения и регистрации всех операций по обмену данными и всех принятых решений. Это необходимо для того, чтобы обеспечить юридическую силу и возможность отслеживания каждого принятого решения, инструкции или изменения, что снижает риск возникновения «недоразумений» в будущем.

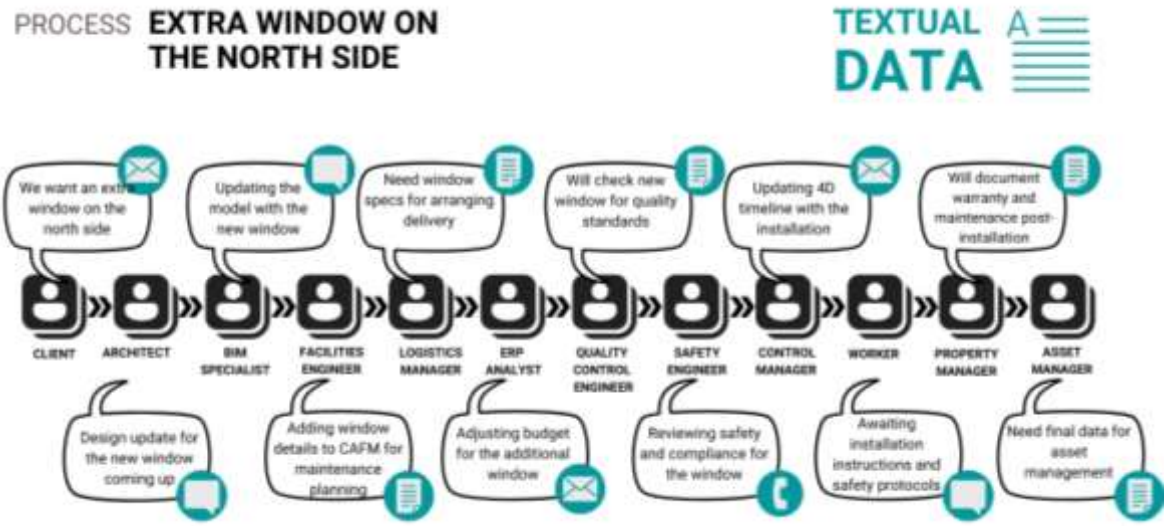


Рис. 4.4-1 Коммуникация между клиентом и исполнителем на начальных этапах проекта часто содержит разноформатные текстовые данные.

> Отсутствие юридического контроля и подтверждения решений в соответствующих системах строительного проекта может привести к серьезным проблемам для всех его участников. Каждое решение, распоряжение или изменение, принятое без надлежащего документального оформления и подтверждения, может привести к спорам (и судебным разбирательствам).

Юридическое закрепление всех решений в текстовой коммуникации может быть обеспечено только большим количеством подписанных документов, которые лягут на плечи руководства, обязанного фиксировать все сделки. В итоге если каждый участник обязан подписывать документы по каждому действию, система теряет гибкость и превращается в бюрократический лабиринт. Отсутствие подтверждений транзакций не только задержит реализацию проекта, но и может привести к финансовым потерям, а также ухудшению взаимоотношений между участниками, вплоть до проблем юридического характера.

Подобный процесс согласования и утверждения транзакций, который обычно начинается с текстовых обсуждений, на следующих этапах постепенно переходит в формат обмена разноформатными документами (Рис. 4.4-2), значительно усложняя коммуникацию, которая происходила только через текст. Без чётко определённых требований автоматизация таких процессов, наполненных разноформатными данными и большим количеством текстовых требований, становится практически невозможной.

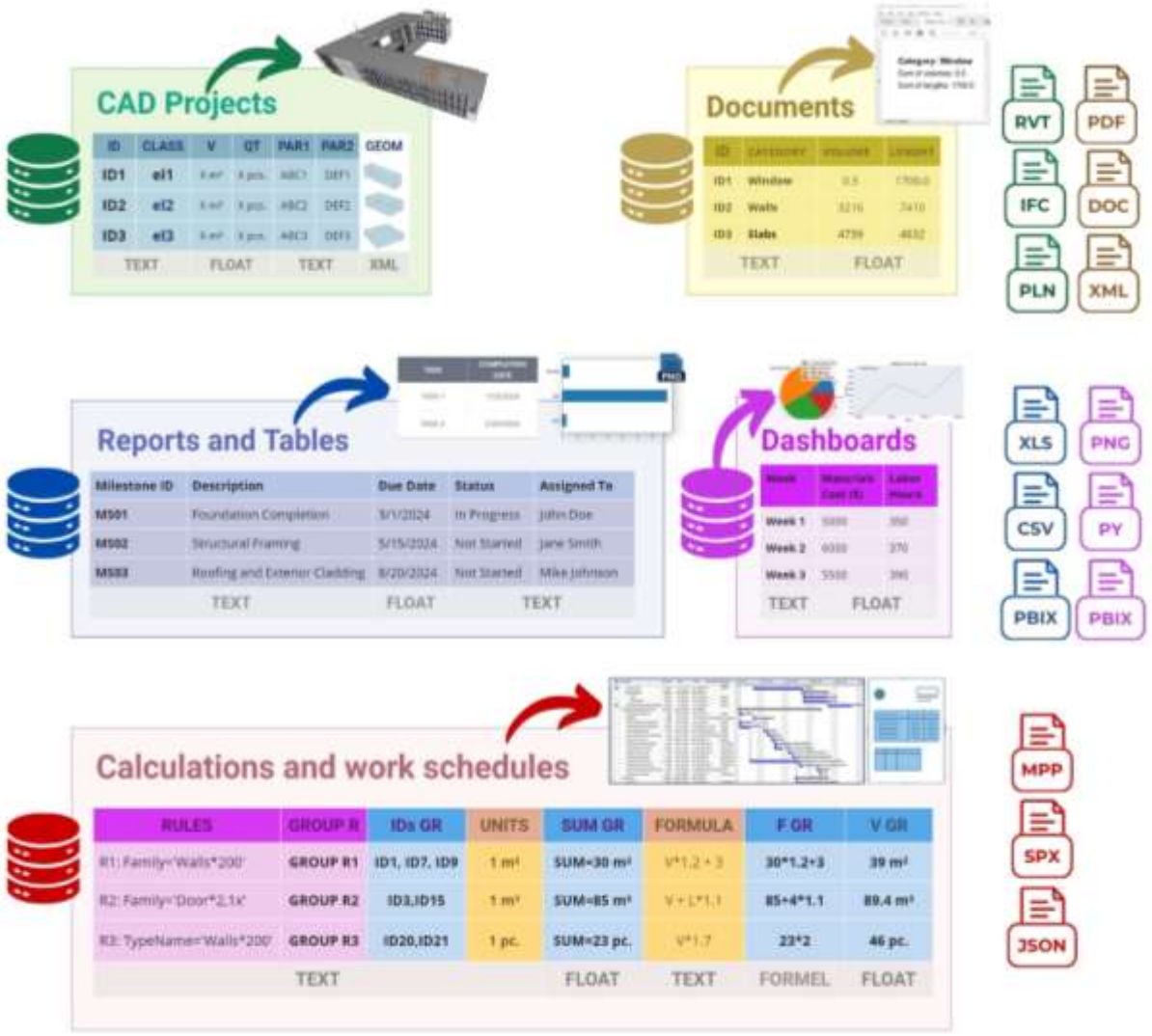


Рис. 4.4-2 Каждая система в ландшафте строительной компании служит источником юридически значимых документов в различных форматах.

Текстовые коммуникации требуют от каждого специалиста либо ознакомления с полной перепиской, либо регулярного участия во всех встречах, чтобы понимания текущего статуса проекта.

Чтобы преодолеть это ограничение, необходим переход от текстовой коммуникации к структурированной модели требований. Это возможно только при помощи систематического анализа, визуализации процесса и описания взаимодействий в виде блок-схем и моделей данных (Рис. 4.4-3). Точно так же, как и в моделировании данных (Рис. 4.3-7), мы перешли от контекстуально-идейного уровня к концептуальному уровню, добавив системы и инструменты, используемые участниками, а также связи между ними.

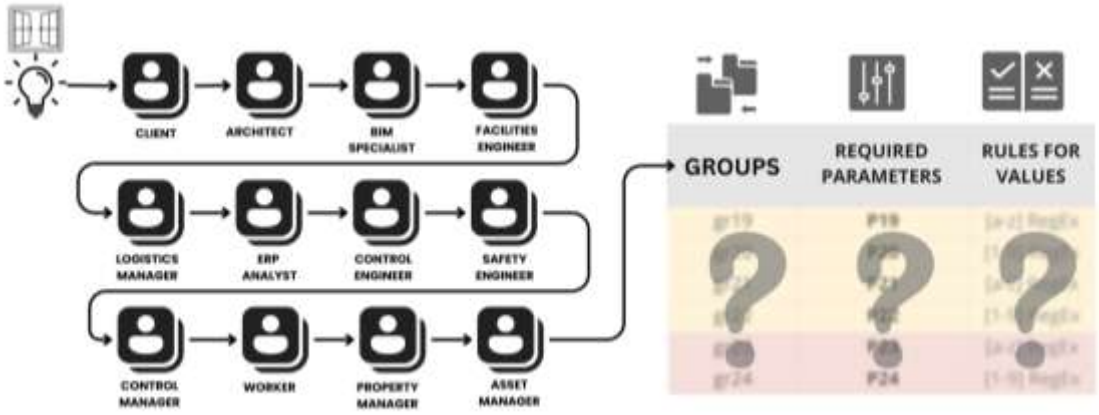


Рис. 4.4-3 Чтобы научиться управлять и автоматизировать процесс валидации, необходимо визуализировать процессы и структурировать требования.

Первый шаг в систематизации требований и отношений - визуализация всех связей и отношений с помощью концептуальных блок-схем. Концептуальный уровень не только облегчит всем участникам процесса понимание всей технологической цепочки, но и наглядно покажет, зачем и для кого нужны данные (и требования) на каждом этапе процесса.

Блок-схемы процессов и эффективность концептуальных схем

Чтобы преодолеть разрыв между традиционными и современными подходами к управлению данными, компаниям необходимо осознанно перейти от фрагментарных текстовых описаний к структурированному представлению процессов. Эволюция данных — от глиняных табличек до цифровых экосистем — требует новых инструментов мышления. И один из таких инструментов — концептуальное моделирование с помощью блок-схем. Создание визуальных схем — блок-схем, диаграмм процессов, схем взаимодействий — позволяет участникам проекта осознать, как их действия и решения влияют на всю систему принятий решений.

> Если процессы требуют не просто хранения данных, а их анализа или автоматизации, то необходимо начинать заниматься темой создания концептуально-визуального уровня требований.

В нашем примере (Рис. 4.4-1) каждый специалист может входить не только в небольшую команду, но и в более крупный отдел, включающий до десятка экспертов под управлением главного менеджера. Каждый отдел использует специализированную базу данных приложения (Рис. 1.2-4 например ERP, CAD, MEP, CDE, ECM, CPM и др.), которая регулярно пополняется входящей информацией, необходимой для создания документов, регистрации юридического статуса решений и управления процессами.

> Процесс транзакций похожа на работу древних менеджеров 4000 лет назад, когда для юридического подтверждения решений использовались глиняные таблички и папирусы. Отличие современных систем от их глиняных и бумажных предшественников заключается в том, что современные методы дополнительно включают процесс преобразования текстовой информации в цифровую форму для дальнейшей автоматической обработки в других системах и инструментах.

Создание визуализации процесса в виде концептуальных блок-схем поможет описать каждый шаг и взаимодействие между различными ролями, делая сложный рабочий процесс понятным и простым.

> Визуализация процессов обеспечивает прозрачность и доступность логики процесса для всех членов команды.

Тот же коммуникативный процесс по добавлению окна в проект, который был описан в виде текста, сообщений (Рис. 4.4-1) и блок схемы, похож на концептуальную модель, рассмотренной нами в главе про моделирование данных (Рис. 4.4-4).

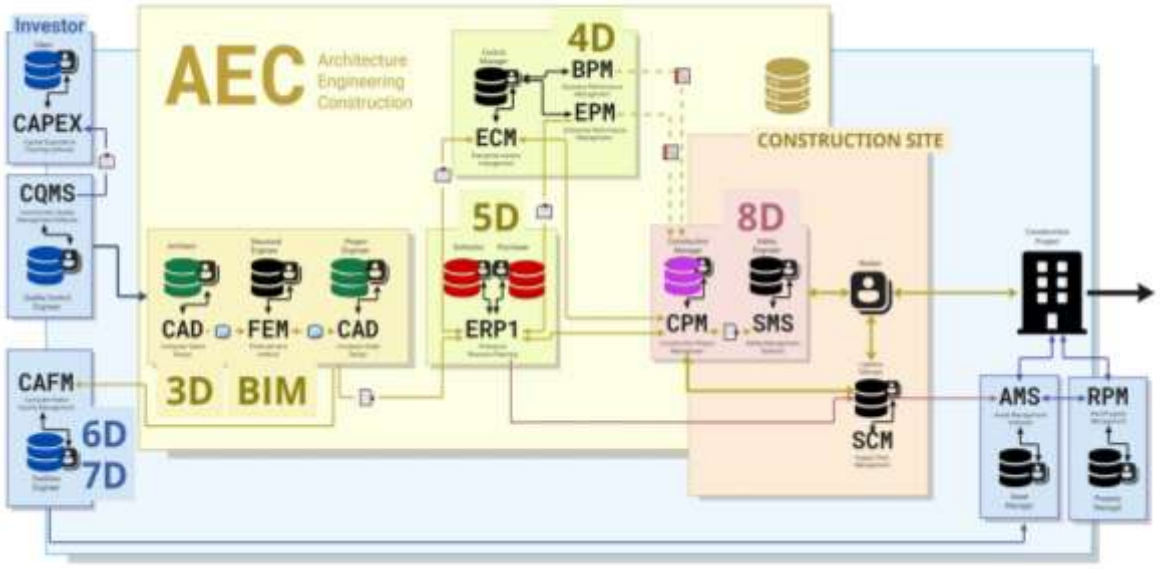


Рис. 4.4-4 На концептуальной схеме участники проекта показаны как пользователи базы данных, где их запросы связывают различные системы.

Хотя концептуальные схемы представляют собой важный шаг, многие компании ограничиваются только этим уровнем, полагая, что визуальной схемы достаточно для понимания процессов. Это создаёт иллюзию управляемости: менеджерам на подобной блок схеме проще воспринимать общую картину, видеть связи между участниками и этапами. Однако подобные схемы не дают чёткого представления о том, какие данные необходимы каждому участнику, в каком формате они должны передаваться и какие именно параметры и атрибуты обязательны для реализации автоматизации. Концептуальная блок-схема скорее похожа на карту маршрута: она указывает на то, кто с кем взаимодействует, но не раскрывает, что именно передаётся в этих взаимодействиях.

> Даже если процесс детально описан на концептуальном уровне с помощью блок-схем, это не гарантирует его эффективности. Визуализация часто упрощает работу менеджеров, позволяя им удобнее отслеживать процесс с помощью пошаговой системы отчетности. Однако для инженеров, управляющих базами данных, концептуальное представление может оставаться недостаточно понятным и не давать чёткого понимания, как именно реализовать процесс на уровне параметров и требований.

По мере продвижения к более сложным экосистемам данных начальное внедрение концептуальных и визуальных инструментов становится критически важным для того, чтобы процессы обработки данных были не только эффективными, но и соответствовали стратегическим целям организации. Чтобы полностью перевести этот процесс по добавлению окна (Рис. 4.4-1) на уровень требований к данным, нам нужно пойти на уровень глубже и перевести концептуальную визуализацию процесса на логический и физический уровень данных, требуемых атрибутов и их граничных значений.

Структурированные требования и регулярные выражения RegEx

До 80% данных, создаваемых в компаниях, представлено в неструктурированных или полуструктурированных форматах [52] — текст, документы, письма, PDF-файлы, разговоры. Такие данные (Рис. 4.4-1) сложно анализировать, проверять, передавать между системами и использовать в автоматизации.

Чтобы обеспечить управляемость, прозрачность и автоматическую валидацию, необходимо переводить текстовые и полуструктурированные требования в чётко определённые, структурированные форматы. Процесс структурирования касается не только данных (что мы рассматривали подробно в первых главах данной части книги), но и самих требований, которые участники проектов обычно формулируют в свободной текстовой форме на протяжении всего жизненного цикла проекта, часто не задумываясь о том, что эти процессы можно автоматизировать.

> Точно так же, как мы уже преобразовали данные из неструктурированной текстовой формы в структурированную, в процессе работы над требованиями мы будем преобразовывать текстовые требования в структурированный формат "логического и физического уровня".

В рамках примера с добавлением окна (Рис. 4.4-1), следующим шагом станет описание требований к данным в табличной форме. Мы структурируем информацию для каждой системы, используемой участниками проекта, указав ключевые атрибуты и их граничные значения.

Рассмотрим, к примеру, одну из таких систем (Рис. 4.4-5) — систему управления качеством строительства (CQMS), которую использует инженер по контролю качества со стороны заказчика. С её помощью он проверяет, соответствует ли новый элемент проекта — в данном случае «новое окно» — установленным стандартам и требованиям.

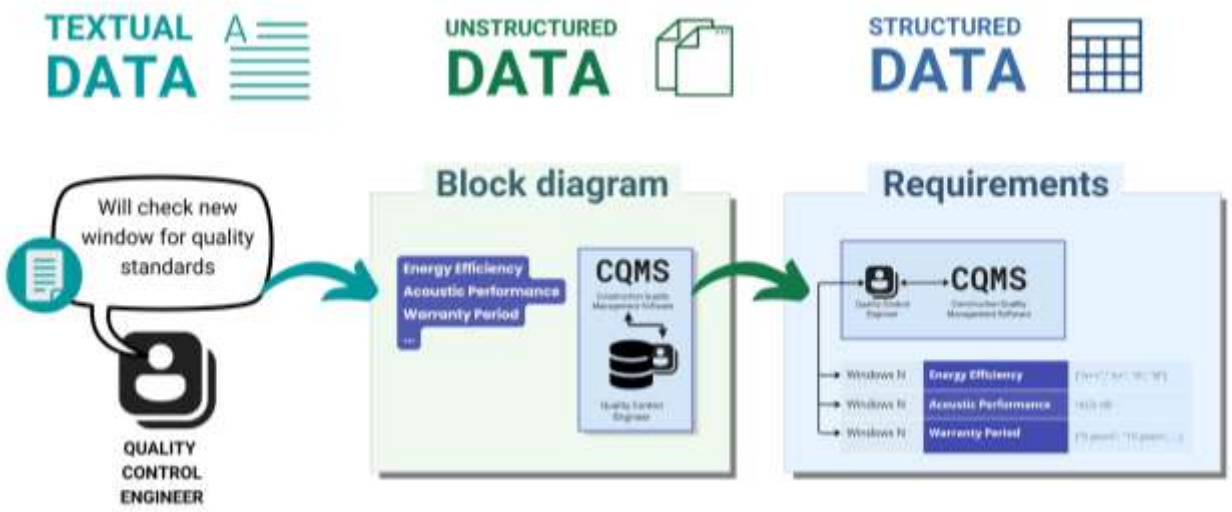


Рис. 4.4-5 Преобразование текстовых требований в формат таблицы с описанием атрибутов сущностей упрощает понимание для других специалистов.

В качестве примера рассмотрим некоторые важные требования к атрибутам сущностей типа ”оконные системы” в CQMS-системе (Рис. 4.4-6): энергоэффективность, акустические характеристики и гарантийный срок. Каждая категория включает в себя определенные стандарты и спецификации, которые необходимо учитывать при проектировании и установке оконных систем.

Рис. 4.4-6 Инженеру по контролю качества необходимо проверять новые элементы типа «Окно» на соответствие стандартам энергоэффективности, звукоизоляции и гарантийного обслуживания.

Требования к данным, которые инженер по контролю качества задает в виде таблицы, имеют, например следующие граничные значения:

- **Класс энергоэффективности окон** варьируется от "A++", обозначающего наивысшую эффективность, до "B", считающегося минимально допустимым уровнем, и эти классы представлены списком допустимых значений **["A++", "A+", "A", "A", "B"].**

- **Акустическая изоляция окон**, измеряемая в децибелах и показывающая их способность снижать уличный шум, определяется регулярным выражением **\d{2}dB.**
- **Атрибут "Гарантийный срок"** для сущности "Тип окна" начинается с пяти лет, устанавливая этот срок как минимально допустимый при выборе продукта; также указываются значения гарантийного срока, например **["5 лет", "10 лет" и т. д.]** или логическое условие **„>5 (лет)“.**

Согласно собранным требованиям, в рамках установленных атрибутов, новые элементы категории или класса “Окно” с классами ниже "B", такие как "C" или "D", не пройдут проверку на энергоэффективность. Акустическая изоляция окон в данных или документах, поступающих к инженеру по контролю качества, должна быть обозначена двузначным числом, за которым следует постфикс "дБ", например, "35 дБ" или "40 дБ", а значения вне этого формата, такие как "9 Д Б" или "100 децибел", не будут приняты (так как не пройдут паттерн для строк RegEx). Гарантийный срок должен начинаться минимум с "5 лет", а окна с более коротким сроком гарантии, такие как "3 года" или "4 года", не будут соответствовать требованиям, которые инженер по качеству описал в формате таблицы.

Для проверки подобных значений атрибутов-параметров на соответствие граничным значениям из требований в процессе валидации мы используем или список допустимых значений ([„A“, „B“, „C“]), словари ([„A“: „H1“, „H2“; „B“: „W1“,“W2“]), логические операции (например, „>“, „<“, „<=“, „>=“ „==“) для числовых значений) и регулярные выражения (для строковых и текстовых значений таких как в атрибуте "Акустические производительность"). Регулярные выражения – это крайне важный инструмент в работе со строковыми значениями.

> **Регулярные выражения** (RegEx) используются в языках программирования, включая Python (библиотека Re), для поиска и изменения строк. Regex — это как детектив в мире строк, способный с точностью идентифицировать текстовые шаблоны в тексте.

В регулярных выражениях буквы описываются непосредственно с помощью соответствующих символов алфавита, а числа могут быть представлены с помощью специального символа \d, который соответствует любой цифре от 0 до 9. Квадратные скобки используются для обозначения диапазона букв или цифр, например, [a-z] для любой строчной буквы латинского алфавита или [0-9], что эквивалентно \d. Для нечисловых и небуквенных символов используются \D и \W соответственно.

Популярные примеры использования RegEx (Рис. 4.4-7):

- **Проверка адреса электронной почты**: чтобы проверить, является ли строка действительным адресом электронной почты, можно использовать шаблон «**^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$**».
- **Извлечение даты**: «**\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\.\d{2}\.\d{4}\b»** шаблон может использоваться для извлечения даты из текста в формате DD.MM.YYYYY.
- Проверка телефонных номеров: для проверки телефонных номеров в формате +49(000)000-0000, шаблон будет выглядеть как **«\+\d{2}\(\d{3}\)\d{3}-\d{4}».**

Переведя требования инженера по контролю качества в формат атрибутов и их граничных значений (Рис. 4.4-6), мы превратили их из исходного текстового формата (разговоров, писем и регламентирующих документов) в организованную и структурированную таблицу, тем самым делая возможным автоматическую проверку и анализ любых поступающих данных (например новых

элементов категории «Окно»). Наличие требований позволяет автоматически отбрасывать данные, не прошедшие проверку, а проверенные данные автоматически передавать в системы для дальнейшей обработки.

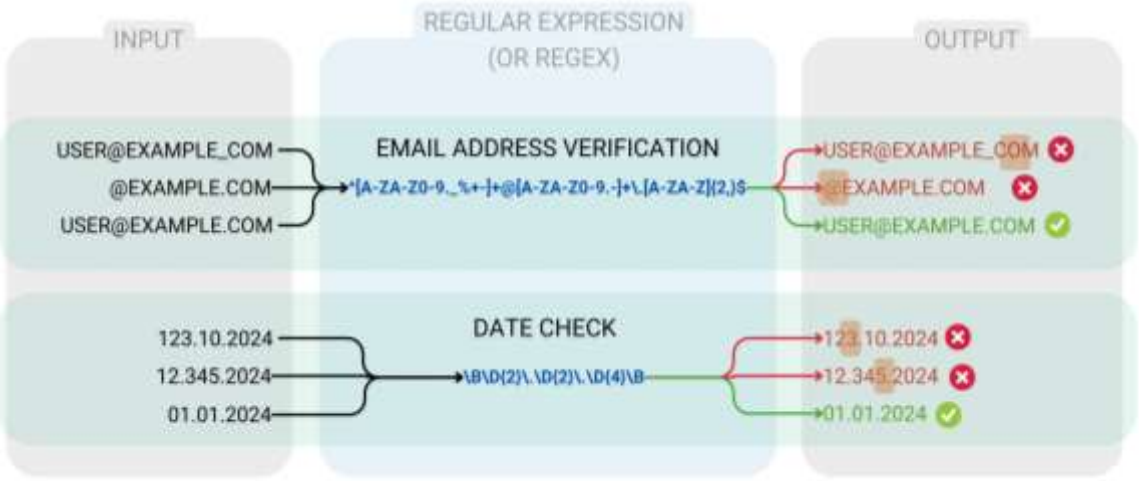


Рис. 4.4-7 Использование регулярных выражений является крайне важным инструментом процесса проверок текстовых данных.

Теперь на переходя с концептуально на логический уровень работы с требованиями, мы преобразуем все требования всех специалистов нашего процесса установки нового окна (Рис. 4.4-4) в упорядоченный список в формате атрибутов и добавим эти списки с необходимыми атрибутами в нашу блок-схему для каждого специалиста (Рис. 4.4-8).

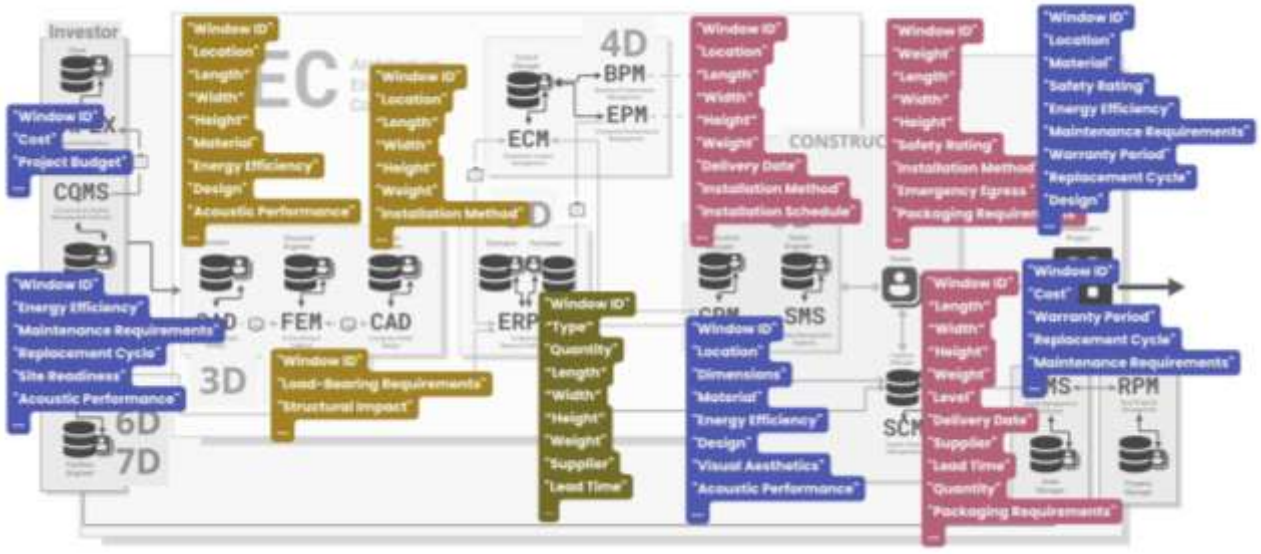


Рис. 4.4-8 На логическом уровне процесса атрибуты, которые обрабатывает каждый специалист, добавляются к соответствующим системам.

Добавив все атрибуты в одну общую таблицу процесса, мы преобразуем информацию, представленную ранее в виде текста и диалога на концептуальном уровне (Рис. 4.4-1), в структурированную и систематизированную форму таблиц физического уровня (Рис. 4.4-9).

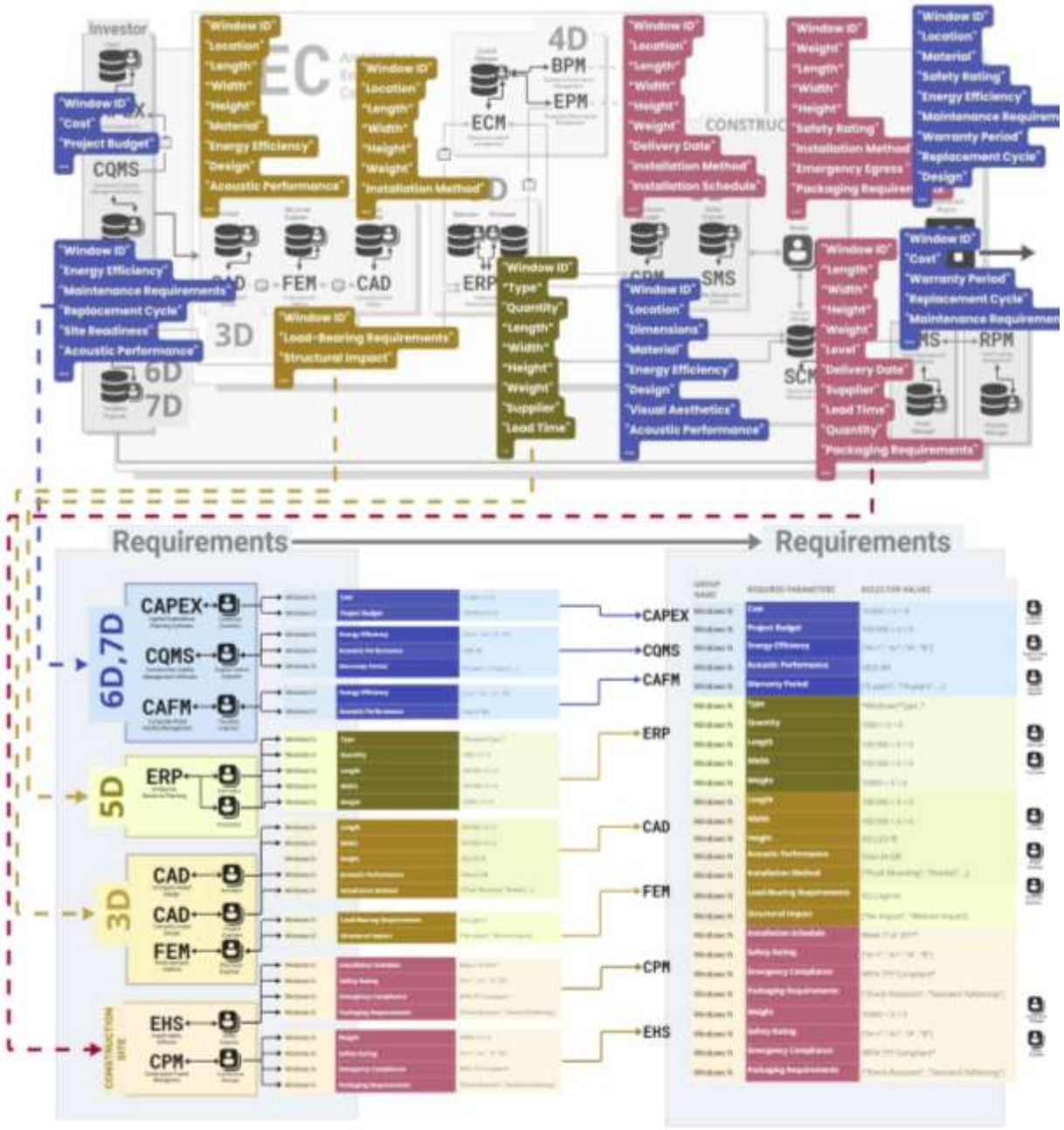


Рис. 4.4-9 Преобразование неструктурированного диалога специалистов в структурированные таблицы помогает понять требования на физическом уровне.

Теперь требования к данным необходимо донести до специалистов, создающих информацию для конкретных систем. Например, если вы работаете в CAD базе данных, то прежде чем приступить к моделированию элементов, следует собрать все необходимые параметры, исходя из конечных сценариев использования этих данных. Обычно это начинается с этапа эксплуатации, затем идёт строительная площадка, отдел логистики, сметный отдел, отдел конструктивных расчётов и так далее. Только после того, как вы учтёте требования всех этих звеньев, можно приступать к созданию данных — на основе собранных параметров. Это позволит в дальнейшем автоматизировать проверку и передачу данных по цепочке.

При соответствии новых данных установленным требованиям, они автоматически интегрируются в экосистему данных компании, направляясь непосредственно тем пользователям и системам, для которых были предназначены. Верификация данных на предмет наличия и соответствия атрибутов и их значений гарантирует, что информация соответствует необходимым стандартам качества и готова к применению в сценариях компании.

Требования к данным определены, и теперь, прежде чем приступить к проверке, необходимо создать, получить или собрать те данные, которые подлежат проверке, либо зафиксировать текущее состояние информации в базах данных, чтобы использовать его в процессе проверки.

Сбор данных для процесса проверки

Прежде чем начать проверку, важно убедиться, что данные доступны в пригодной для процесса валидации форме. Это означает не просто наличие информации, а её подготовку: данные необходимо собрать и преобразовать из неструктурированных, слабо структурированных, текстовых и геометрических форматов в структурированный вид. Этот процесс подробно описан в предыдущих главах, где рассматривались методы трансформации различных типов данных. В результате всех преобразований входящие данные приобретают форму открытых структурированных таблиц (Рис. 4.1-2, Рис. 4.1-9, Рис. 4.1-13).

Обладая требованиями и структурированными таблицами с необходимыми параметрами и граничными значениями (Рис. 4.4-9), мы можем приступить к проверке данных — как в виде единого автоматизированного процесса (Pipeline), так и в формате поэтапной проверки каждого входящего документа.

Для запуска проверки требуется или получить на вход новый файл или зафиксировать актуальное состояние данных — создать снимок или сделать экспорт текущих и поступающих сведений, либо настроить подключение к внешней или внутренней базе данных. В рассматриваемом примере такой снимок создаётся путём автоматического преобразования данных CAD данных в структурированный формат, зафиксированный, скажем, в 23:00:00 пятницы, 29 марта 2024 года, после того как все проектировщики ушли домой.

Рис. 4.4-10 Снимок базы данных CAD (BIM), показывающий текущие сведения об атрибутах для новой сущности класса «Окно» в текущей версии модели проекта.

Благодаря инструментам обратного инжиниринга, рассмотренным в главе "Перевод данных CAD (BIM) в структурированную форму", эта информация из различных инструментов и редакторов CAD (BIM) может быть организована в отдельные таблицы (Рис. 4.4-11) или объединена в одну общую таблицу, объединяющую различные разделы проекта (Рис. 9.1-10).

В подобной таблице - базе данных отображаются уникальные идентификаторы окон и дверей (атрибут ID), типовые наименования (TypeName), размеры (Width, Length), материалы (Material), а также показатели энергетической и акустической эффективности, а также другие характеристики. Такая таблица, заполненная в программе CAD (BIM), собирается инженером проектировщиком из различных отделов и документов, формируя информационную модель проекта.

Рис. 4.4-11 Структурированные данные из CAD систем могут представлять собой двумерную таблицу со столбцами, обозначающими атрибуты элементов.

Реальные проекты CAD (BIM) включают десятки или сотни тысяч элементов (Рис. 9.1-10). Элементы внутри CAD форматов автоматически классифицируются по типам и категориям — от окон и дверей до плит, перекрытий и стен. Уникальные идентификаторы (например, native ID, который устанавливается CAD решением автоматически) или атрибуты типа (Type Name, Type, Family) позволяют отслеживать один и тот же объект в разных системах. Так, новое окно на северной стене здания может однозначно обозначаться через один идентификатор "W-NEW" во всех соответствующих системах организации.

Хотя имена и идентификаторы сущностей должны быть единообразными во всех системах, набор атрибутов и значений, связанных с этими сущностями, может значительно различаться в зависимости от контекста использования. Архитекторы, инженеры-конструкторы, специалисты по строительству, логистике и эксплуатации недвижимости по-разному воспринимают одни и те же элементы. Каждый из них опирается на собственные классификаторы, стандарты и цели: кто-то рассматривает окно исключительно с эстетической точки зрения, оценивая его форму и пропорции, а кто-то — с инженерной или эксплуатационной, анализируя теплопроводность, способ монтажа, массу или требования к техническому обслуживанию. Поэтому при моделировании данных и описании элементов важно учитывать многогранность их использования и обеспечить согласованность данных при одновременном учёте отраслевых особенностей.

Для каждой роли в процессах компании предусмотрены специализированные базы данных с собственным пользовательским интерфейсом — от проектирования и расчётов до логистики, монтажа и эксплуатации зданий (Рис. 4.4-12). Каждая подобная система управляется профессиональной командой специалистов через специальный пользовательский интерфейс или через запросы к базе данных, где за суммой всех решений, принимаемых по введенным значениям в конце цепочки, стоит менеджер системы или руководитель отдела, который отвечает за юридическую обоснованность и качество введенных данных перед своими контрагентами, обслуживающих другие системы.

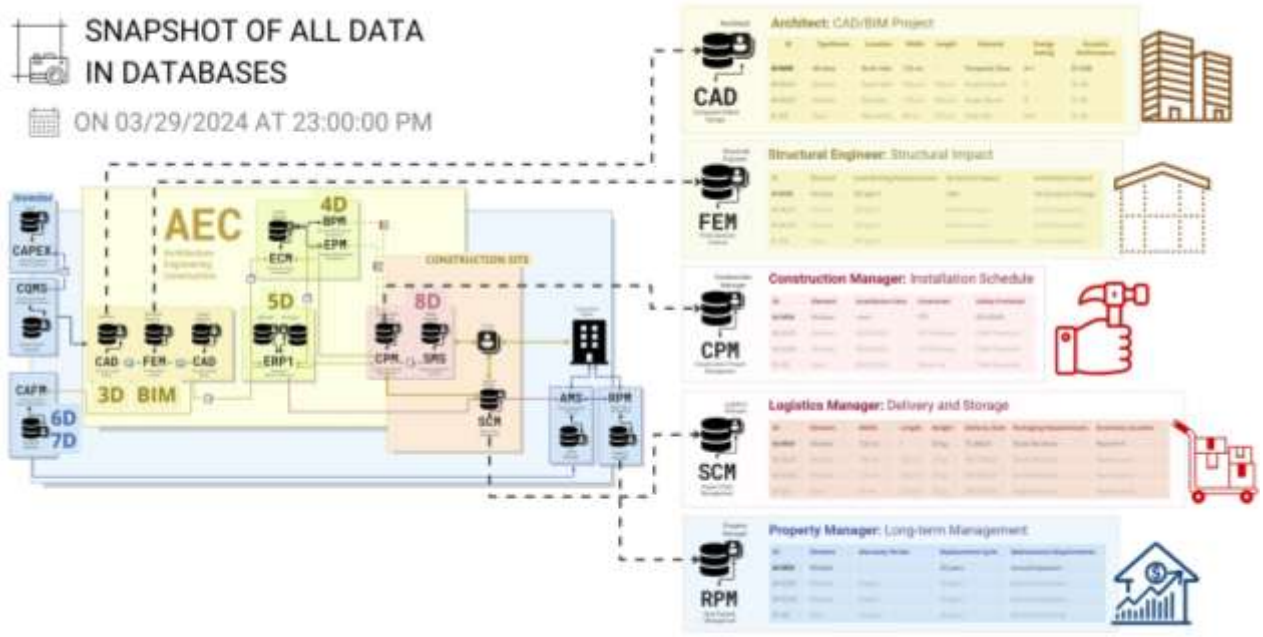


Рис. 4.4-12 Одна и та же сущность имеет одинаковый идентификатор в разных системах, но разные атрибуты, которые важны только в данной системе.

После организованного сбора структурированных требований и данных на логическом и физическом уровне нам остается настроить процесс автоматической проверки данных, полученных из различных поступающих документов и разных систем, на соответствие ранее собранным требованиям.

Проверка данных и результаты проверки

Все новые данные, поступающие в систему — будь то документы, таблицы или записи в базах данных от заказчика, архитектора, инженера, прораба, логиста или управляющего недвижимостью, — должны пройти проверку на соответствие требованиям, сформулированным ранее (Рис. 4.4-9). Процесс валидации критически важен: любые ошибки в данных могут привести к неправильным расчётам, задержкам в графике и даже финансовым потерям. Чтобы минимизировать подобные риски, необходимо организовать систематическую и повторяемую, итерационную процедуру проверки данных.

Для проверки новых данных, поступающих в систему - неструктурированных, текстовых или геометрических, - их необходимо преобразовать в слабоструктурированный или структурированный формат. После чего в процессе проверки данные должны быть проверены на соответствие полному списку требуемых атрибутов и их допустимых значений.

Преобразование различных типов данных: текста, изображений, PDF-документов и смешанных данных CAD (BIM) в структурированную форму мы подробно рассматривали в главе "Преобразование данных в структурированную форму".

В качестве примера можно привести таблицу, полученную из CAD (BIM) проекта (Рис. 4.4-11). Она включает полуструктурированные геометрические данные и структурированную атрибутивную информацию по сущностям проекта (Рис. 3.1-14) — например элемента из класса «Окна».

Для проведения проверки мы сопоставляем значения атрибутов (Рис. 4.4-11) с эталонными граничными значениями, которые были определенны экспертами в виде требования (Рис. 4.4-9). Итоговая сравнительная таблица (Рис. 4.4-13) позволит понять, какие значения допустимы, а какие нуждаются в исправлении, прежде чем данные можно будет использовать вне приложений CAD (BIM).

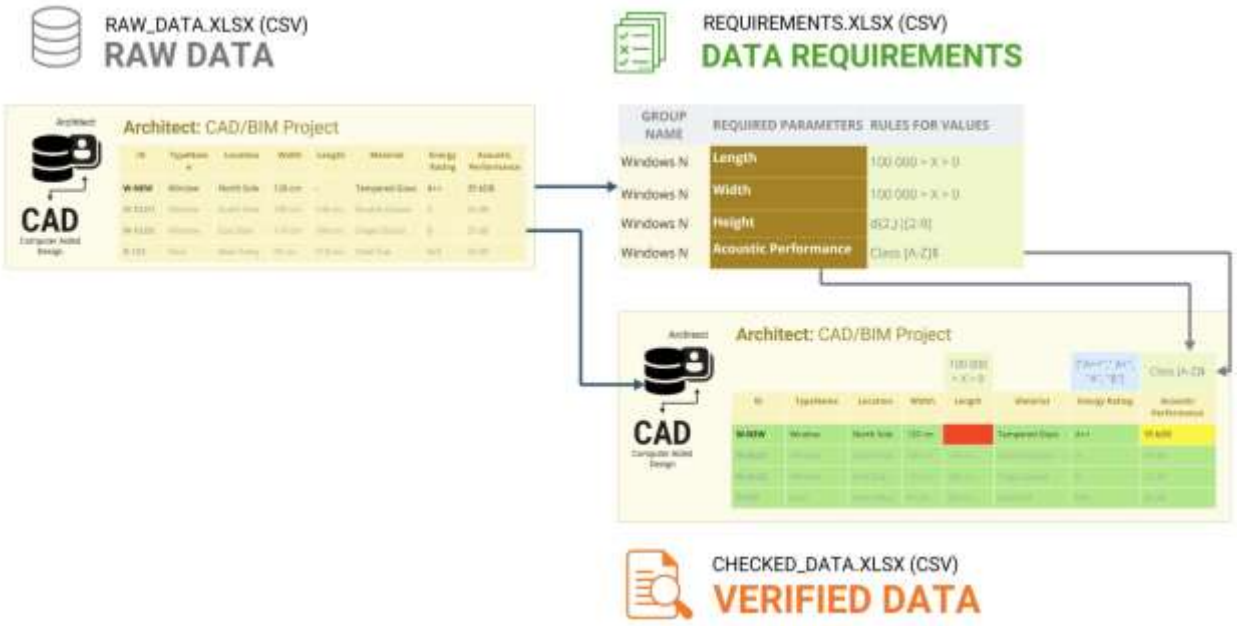


Рис. 4.4-13 Итоговая таблица проверки подсвечивает те значения атрибутов для новой сущности класса «Окна», на которые необходимо обратить внимание.

Реализуя подобное решение, используя библиотеку Pandas, о которой мы рассказывали ранее в главе "Pandas: Незаменимый инструмент для анализа данных", мы проверим данные из табличного файла, извлеченного из файла CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Рис. 4.4-11), используя требования из другого табличного файла требований (Рис. 4.4-9).

Для получения кода нам необходимо описать в промпте для LLM, что необходимо загрузить дан-

ные из файла **raw_data.xlsx** (полный набор данных из базы данных CAD (BIM)), проверить их и сохранить результат в новом файле **checked_data.xlsx** (Рис. 4.4-13).

- Получим код при помощи LLM без упоминания библиотеки Pandas:

> Напиши код для проверки таблицы из файла raw_data.xlsx и проверь их с помощью следующих правил проверки: значения столбцов 'Width' и 'Length' больше нуля, 'Energy Rating' входит в список ['A++', 'A+', 'A', 'B'], а 'Acoustic Performance' как переменная, которую мы укажем позже - с добавлением итогового столбца проверки, и сохрани итоговую таблицу в новый файл Excel checked_data.xlsx ⏎

- Ответ LLM опишет короткий пример Python кода, который можно уточнять и дополнять последующими промптами:

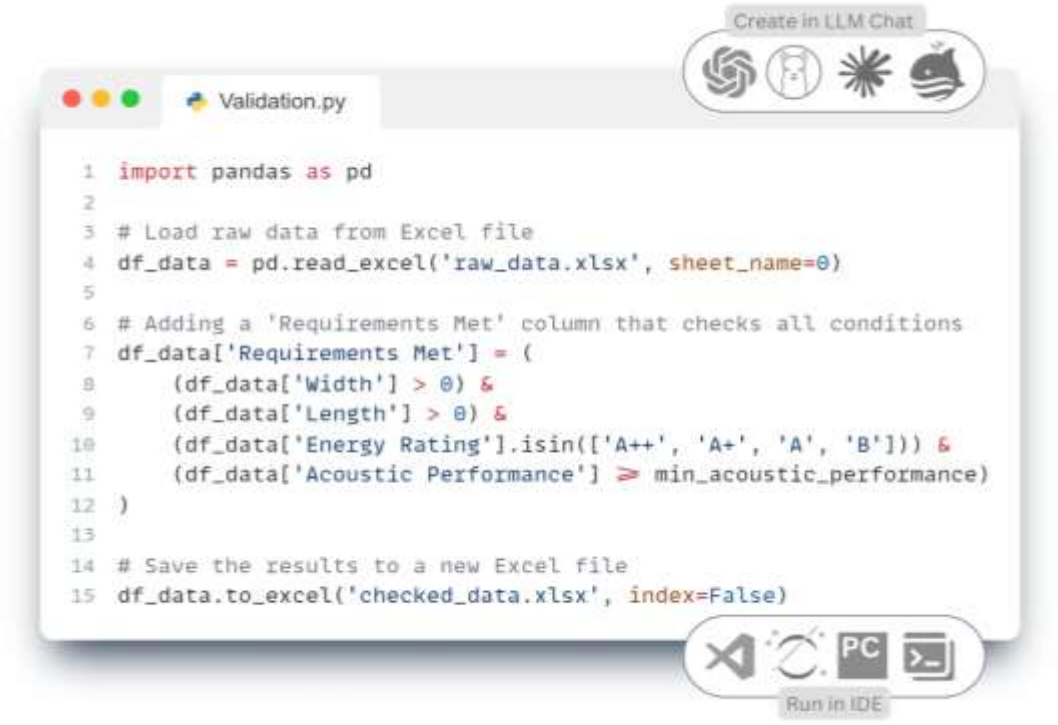


Рис. 4.4-14 Код, сгенерированный LLM-моделью проверяет преобразованный CAD (BIM) проект на соответствие требованиям к атрибутам в виде граничных значений.

Код, сгенерированный языковой моделью LLM, можно использовать в любой популярной IDE или онлайн-инструменте: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с плагином PyDev, Thonny, Wing IDE, IntelliJ IDEA с плагином Python, JupyterLab или популярных онлайн-инструментах Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Выполнение кода (Рис. 4.4-14) покажет, что "элементы-сущности" W-OLD1, W-OLD2, D-122 (и другие

элементы) из базы данных CAD (BIM) соответствуют требованиям к атрибутам: ширина и длина больше нуля, а класс энергоэффективности является одним из значений списка 'A++', 'A', 'B', 'C' (Рис. 4.4-15).

Нужный нам и недавно добавленный элемент W-NEW, отвечающий за новой элемент класс «Окно» на северной стороне, не соответствует требованиям (атрибут „Requrments Met“), поскольку его длина равна нулю (значение "0.0" считается неприемлемым по нашему правилу 'Width'>0) и в нем не указан класс энергоэффективности.

	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

Рис. 4.4-15 Проверка выявляет сущности, которые не прошли процесс верификации, и добавляет в результаты новый атрибут со значениями 'False' или 'True'.

Аналогичным образом мы проверяем согласованность всех элементов проекта (сущностей) и необходимых атрибутов для каждой из систем, таблиц или баз данных во всех данных, которые мы получаем от разных специалистов (Рис. 4.4-1) в процессе добавления окна в проект.

В итоговой таблице удобно для наглядной визуализации выделять результаты проверки цветом: зелёным помечаются атрибуты, успешно прошедшие проверку, жёлтым — значения с не критичными отклонениями, а красным — критические несоответствия (Рис. 4.4-16).

В результате проверки (Рис. 4.4-16) мы получаем список доверенных и проверенных элементов с их идентификаторами, которые были проверены на соответствие требованиям к атрибутам. Проверенные элементы обеспечивают уверенность в том, что эти элементы соответствуют заявленным стандартам и спецификациям для всех систем, участвующих в процессе добавления элементов класс «Окно» или любого другого класса (подробнее об автоматизации проверки данных и создании автоматизированного процесса ETL мы поговорим в главе "Автоматизация ETL и проверки данных").

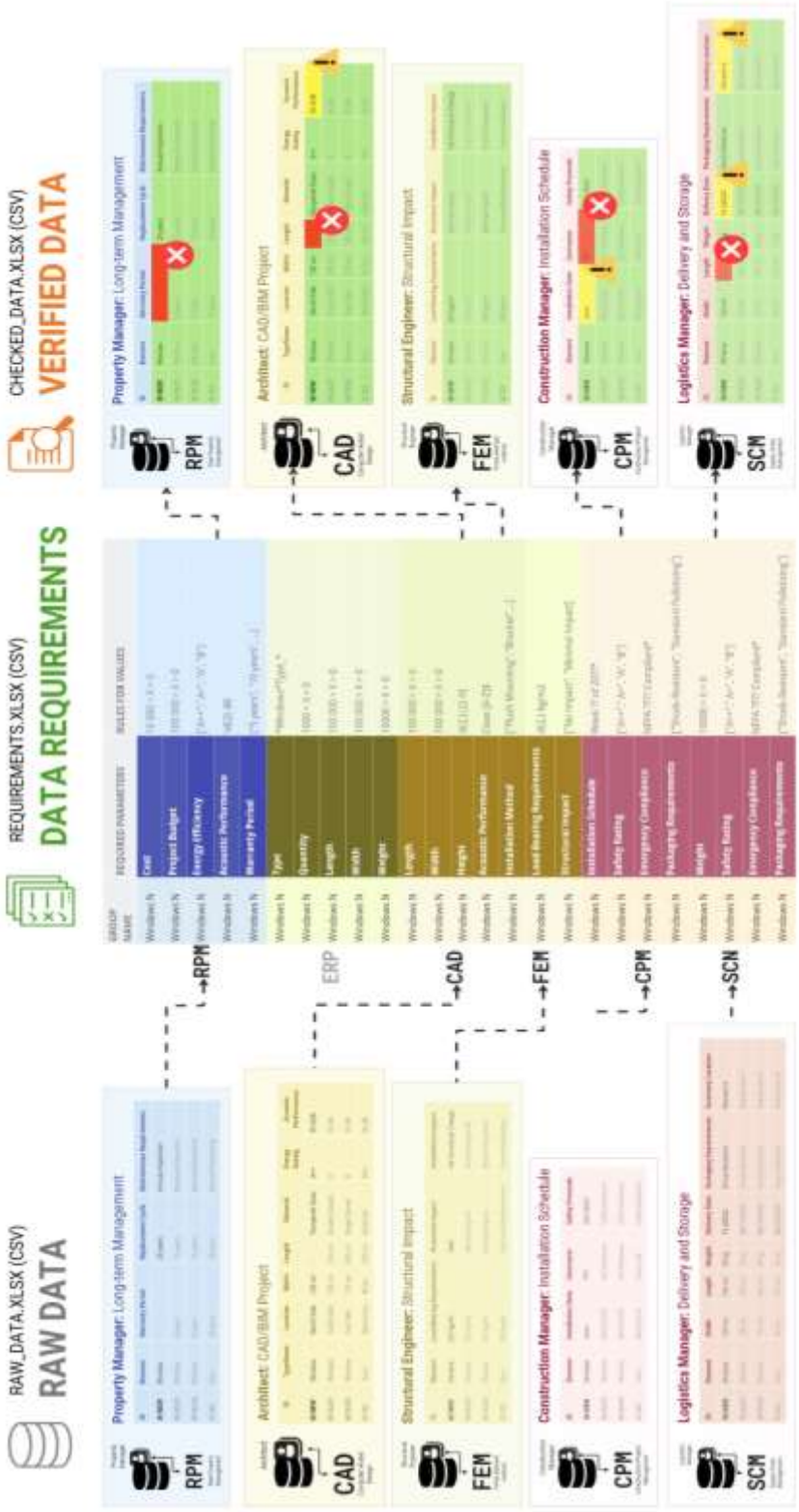


Рис. 4.4-16 Результат проверки проведённый для всех систем позволяет определить, какие данные не соответствуют требованиям компании.

Сущности, успешно прошедшие проверку, как правило, не требуют к себе повышенного внимания. Они без препятствий переходят к следующим этапам обработки и интеграции в другие системы. В отличие от «качественных» элементов, именно элементы, не прошедшие проверку, представляют наибольший интерес. Информация о таких отклонениях критически важна: её следует передавать не только в виде табличных отчётов, но и с использованием различных средств визуализации. Графическое представление результатов проверки помогает быстрее оценить общее состояние качества данных, выявить проблемные области и оперативно принять меры по исправлению или уточнению параметров.

Визуализация результатов проверки

Визуализация — важнейший инструмент интерпретации результатов проверки. Помимо привычных сводных таблиц, она может включать информационные панели, диаграммы и автоматически формируемые PDF-документы, в которых элементы проекта группируются по статусу прохождения проверки. Цветовая кодировка может играть здесь вспомогательную роль: зелёный может обозначать успешно проверенные объекты, жёлтый — элементы, требующие дополнительного внимания, а красный — те, в которых выявлены критические ошибки или отсутствуют ключевые данные.

В нашем примере (Рис. 4.4-1) мы поэтапно анализируем данные из каждой системы: от CAD (BIM) и управления недвижимостью до логистики и графиков монтажа (Рис. 4.4-16). По итогам аудита для каждого специалиста автоматически формируются индивидуальные оповещения или отчётные документы, например в формате PDF (Рис. 4.4-17). Если данные корректны, специалист получает краткое сообщение: «Спасибо за совместную работу». В случае обнаружения несоответствий отправляется детализированный отчёт с формулировкой: «В этом документе перечислены элементы, их идентификаторы, атрибуты и значения, которые не прошли проверку на соответствие».

Рис. 4.4-17 Валидация и автоматическое создание отчетных документов ускоряют процесс поиска и понимания недостатков данных для специалиста, который создаёт данные.

Благодаря автоматизированному процессу проверки - как только обнаруживается ошибка или пробел в данных, мгновенно отправляется уведомление в виде сообщения в чате, электронной почтой или PDF-документом лицу, ответственному за создание или обработку соответствующих сущностей и их атрибутов (Рис. 4.4-18), со списком элементов и описаний атрибутов, которые не прошли проверку.

Рис. 4.4-18 Автоматические отчеты по результатам проверок облегчают понимание ошибок и ускоряют работу по заполнению проектных данных.

Например, если в систему управления недвижимостью поступает (после структурирования) документ, в котором неверно заполнен атрибут "Гарантийный срок", менеджер по недвижимости получает оповещение со списком атрибутов, которые необходимо проверить и исправить.

Аналогичным образом, любые недочеты в графике монтажа или данных о логистике приводят к автоматическому формированию отчета и например отправке уведомления в чате или электронного письма с результатами проверки соответствующему специалисту.

Помимо PDF-документов и графиков с результатами, возможно создание приборных панелей (дэшбордов) и интерактивных 3D-моделей (Рис. 7.1-6, Рис. 7.2-12) с выделением элементов с недостающими атрибутами, что позволяет пользователям визуально использовать 3D-геометрии элементов для фильтрации и оценки качества и полноты данных элементов в проекте.

> Визуализация результатов проверок в виде автоматически генерируемых документов, графиков или приборных панелей значительно упрощает интерпретацию данных и способствует эффективному взаимодействию между участниками проекта.

Процесс автоматической проверки данных из различных систем и источников информации можно сравнить с осознанным принятием решений в повседневной жизни. Подобно тому, как компании в строительной отрасли учитывают множество переменных — от достоверности исходных данных до их влияния на сроки, стоимость и качество реализации проекта, — так и человек, принимая важные решения, например, при выборе места жительства, взвешивает целый ряд факторов: транспортную доступность, инфраструктуру, стоимость, безопасность, качество жизни. Все эти соображения формируют систему критериев, на основе которой и принимается финальные решения из которых состоит наша жизнь.

Сравнение проверок качества данных с жизненными потребностями человека

Несмотря на постоянное развитие методов и инструментов контроля качества данных, фундаментальный принцип — соответствие информации установленным требованиям — остаётся неизменным. Этот принцип встроен в основу зрелой системы управления, будь то в бизнесе или в повседневной жизни.

> Процесс итеративной проверки данных во многом напоминает процесс принятия решений, с которым ежедневно сталкивается каждый человек. И в том, и в другом случае мы опираемся на уже накопленный опыт, имеющиеся данные и поступающую новую информацию. Причём всё больше жизненных и профессиональных решений — от стратегических до бытовых — принимается на основе данных.

Так, например, при выборе места жительства или спутника жизни мы интуитивно формируем в сознании некую таблицу с критериями и характеристиками, по которым сравниваем альтернативы (Рис. 4.4-19). Эти характеристики — будь то личные качества человека или параметры объекта недвижимости — представляют собой атрибуты, влияющие на финальное решение.

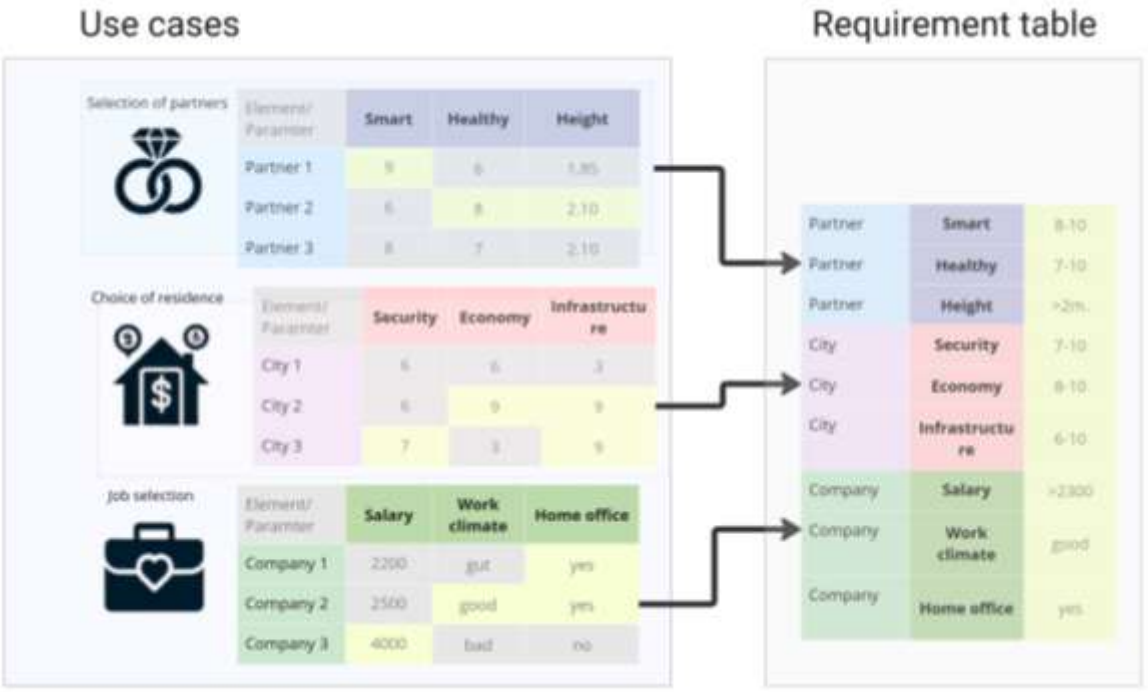


Рис. 4.4-19 Выбор места жительства, работы или партнерства основывается на индивидуальных требованиях к атрибутам.

Использование структурированных данных и формализованного подхода к описанию требований (Рис. 4.4-20) способствует более обоснованному и осознанному выбору как в профессиональной деятельности, так и в личной жизни.

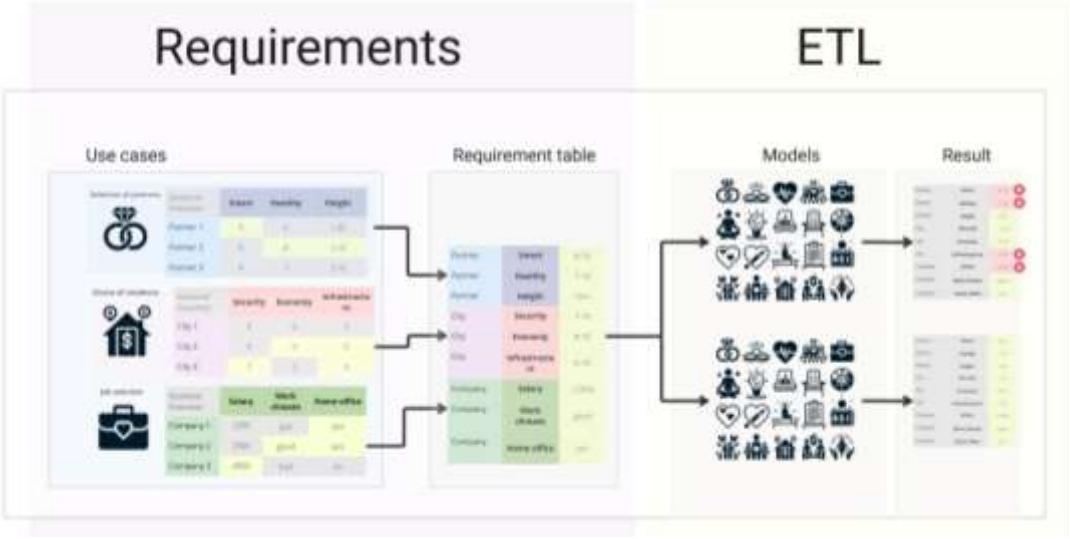


Рис. 4.4-20 Формализация требований позволяет систематизировать восприятие жизненных и деловых решений.

Подход к принятию решений на основе данных — это не исключительно бизнес-инструмент. Он органично интегрирован и в повседневную жизнь, следуя общим этапам обработки данных (Рис. 4.4-21), аналогичным процессу ETL (Extract, Transform, Load), который мы уже рассматривали в начале этой части при структурировании данных и который мы будем подробно рассматривать в контексте автоматизации задач в седьмой части книги:

- **Данные как основа (Extract)**: в любой сфере — будь то работа или личная жизнь — мы собираем информацию. В бизнесе это могут быть отчёты, показатели, рыночные данные; в личной жизни — личный опыт, советы близких, отзывы, наблюдения.
- **Критерии оценки (Transform):** собранная информация интерпретируется на основе заранее определённых критериев. На работе — это показатели эффективности (KPI), бюджетные ограничения и нормы; в личной жизни — такие параметры, как цена, удобство расположения, надёжность, харизма и т. д.
- **Прогнозирование и анализ рисков (Load):** на финальном этапе происходит принятие решения, основанное на анализе трансформированных данных и сопоставлении возможных последствий. Это схоже с бизнес-процессами, где данные проходят через фильтр бизнес-логики и рисков.

Решения, которые мы принимаем - от банальных предпочтений вроде того, что съесть на завтрак, до важных жизненных событий вроде выбора карьеры или спутника жизни, — по своей сути являются результатом обработки и оценки данных.

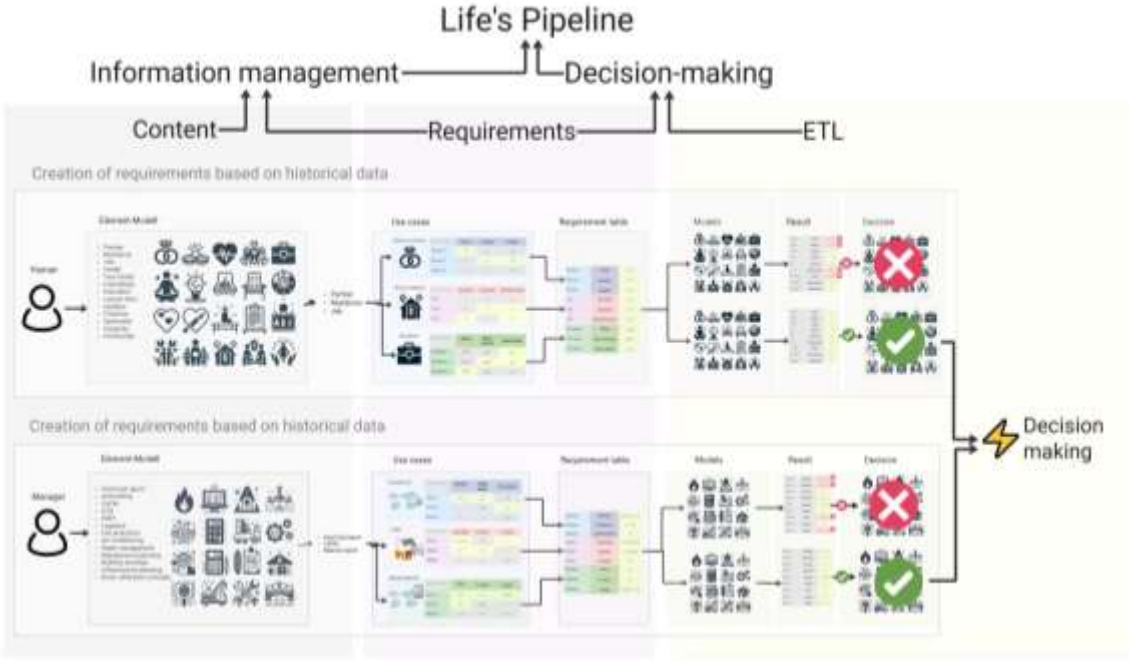


Рис. 4.4-21 Бизнес и жизнь в целом — это серия решений, основанных на данных, где качество данных, используемых для принятия решений, является ключевым фактором.

Все в нашей жизни взаимосвязано, и, как и живые организмы, включая человека, следуют законам природы, эволюционируя и приспосабливаясь к меняющимся условиям, так и человеческие процессы, включая методы сбора и анализа данных, отражают эти природные принципы. Тесная взаимосвязь между природой и деятельностью человека подтверждает не только нашу зависимость от природы, но и наше желание применять законы, отшлифованные миллионами лет эволюции, в создании архитектур данных, разработке процессов и систем для принятия решений.

Новые технологии, особенно в строительстве, - яркий пример того, как человечество раз за разом вдохновляется природой для создания наилучших, более устойчивых и эффективных решений.

Дальнейшие шаги: превращение данных в точные расчеты и планы

В этой части мы рассмотрели, как преобразовывать неструктурированные данные в структурированный формат, разрабатывать модели данных и организовывать процессы проверки качества информации в строительных проектах. Управление данными, их стандартизация и классификация — это фундаментальный процесс, требующий системного подхода и чёткого понимания бизнес-требований. Рассмотренные в этой части методики и инструменты позволяют обеспечить надёжную интеграцию между различными системами на протяжении всего жизненного цикла объекта.

Подводя итог данной части, выделим основные практические шаги, которые помогут применить рассмотренные подходы в ваших повседневных задачах:

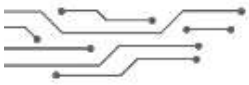
- Начните с систематизации требований
  - Создайте реестр атрибутов и параметров для ключевых элементов ваших проектов и

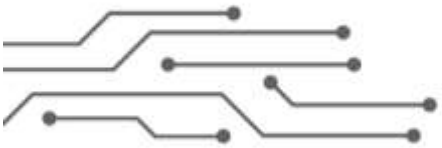
процессов

- [ ] Документируйте граничные значения для каждого атрибута
- [ ] Визуализируйте процессы и взаимосвязи между классами, системами и атрибутами с помощью блок-схем (например в Miro, Canva, Visio)

- Автоматизируйте преобразование данных
  - [ ] Проверьте, какие из ваших документов, часто используемых в процессах, можно оцифровать с помощью OCR-библиотек и перевести в табличный вид
  - [ ] Ознакомьтесь с инструментами обратного инжиниринга для извлечения данных из CAD (BIM)
  - [ ] Попробуйте настроить автоматическое получение данных из документов или форматов, которые вы часто используете в своей работе, в табличную форму
  - [ ] Настройте автоматические преобразования между различными форматами данных
- Создайте базу знаний для классификации
  - [ ] Разработайте внутренний или используйте уже существующий классификатор элементов, согласованный с отраслевыми стандартами
  - [ ] Документируйте взаимосвязи между различными системами классификации
  - [ ] Обсудите с вашей команду тему использования единой системы идентификации и однозначной классификации элементов
  - [ ] Начните выстраивать процесс автоматической проверки данных — как тех, с которыми вы работаете внутри команды, так и тех, что передаются во внешние системы

Использование таких подходов позволит вам существенно повысить качество данных и упростить их последующую обработку и трансформацию. В следующих частях книги мы рассмотрим, как применять уже структурированные и подготовленные данные для автоматизированных расчётов, оценки стоимости, календарного планирования и управления строительными проектами.





V ЧАСТЬ
КАЛЬКУЛЯЦИИ СТОИМОСТИ И ВРЕМЕНИ: ВНЕДРЕНИЕ ДАННЫХ В СТРОИТЕЛЬНЫЕ ПРОЦЕССЫ

Пятая часть посвящена практическим аспектам использования данных для оптимизации расчетов стоимости и планирования строительных проектов. Подробно анализируется ресурсный метод составления смет и автоматизация процессов расчета. Рассматриваются методы автоматизированного получения объемов (Quantity Take-Off) из CAD (BIM)-моделей и их интеграции с расчетными системами. Исследуются технологии 4D и 5D моделирования для планирования временных параметров и управления стоимостью строительства, с конкретными примерами их применения. Представлен анализ расширенных информационных слоев 6D-8D, обеспечивающих комплексный подход к оценке устойчивости, эксплуатации и безопасности объектов недвижимости. Детально рассматриваются методики расчета углеродного следа и ESG-показателей строительных проектов в контексте современных экологических требований и стандартов. Критически оцениваются возможности и ограничения традиционных ERP и PMIS систем в управлении строительными процессами, с анализом их влияния на прозрачность ценообразования. Прогнозируются перспективы перехода от закрытых решений к открытым стандартам и гибким инструментам анализа данных, способным обеспечить большую эффективность строительных процессов.

ГЛАВА 5.1.
РАСЧЕТЫ СТОИМОСТИ И СМЕТЫ СТРОИТЕЛЬНЫХ ПРОЕКТОВ

Основы строительства: оценка количества, стоимости и времени

Среди множества бизнес-процессов, определяющих устойчивость компании в строительной отрасли, особое значение — как и тысячи лет назад — имеют процессы точной оценки количества элементов, стоимости проекта и сроков выполнения (Рис. 5.1-1).

Развитие письменности стало результатом комплекса факторов, включая необходимость учета хозяйственных операций, развитие торговли и управления ресурсами в ранних обществах. Первые юридически значимые документы — глиняные таблички с расчётами стоимости материалов и оплаты труда — использовались в контексте торговли и строительства. Эти таблички фиксировали обязательства сторон при возведении сооружений и хранились как свидетельства договорённостей и товаро-денежных отношений.

> На протяжении тысячелетий подход к оценке почти не менялся: расчёты выполнялись вручную, опираясь на опыт и интуицию инженера-сметчика. Однако с появлением модульных ERP-систем и CAD-инструментов традиционный подход к оценке количества, стоимости и времени начал стремительно трансформироваться. Современные цифровые технологии позволяют полностью автоматизировать ключевые расчёты времени и стоимости, позволяя повысить точность, скорость и прозрачность ресурсного планирования строительных проектов.

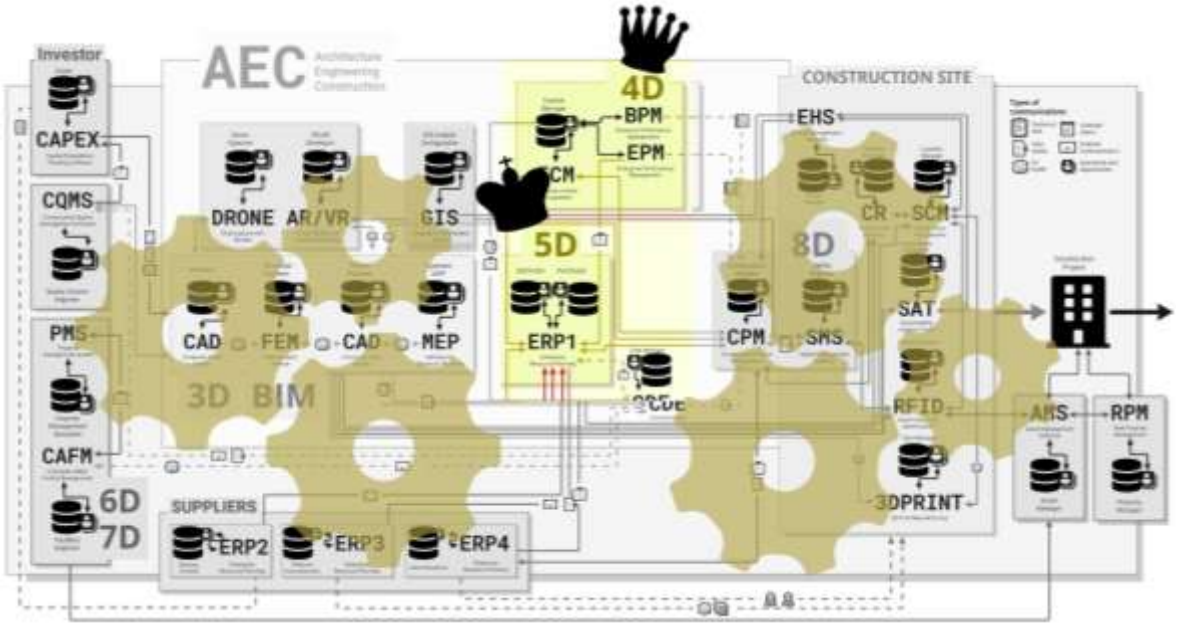


Рис. 5.1-1 Из множества различных систем наибольшую важность в бизнесе имеют инструменты, отвечающие за показатели объемов, стоимости и времени.

Основное внимание строительных компаний сосредоточено на точных данных о сроках и стоимости работ. Эти показатели в свою очередь зависят от объёмов используемых материалов и трудозатрат, а их прозрачность влияет на рентабельность. Однако сложность расчётных процессов и их недостаточная прозрачность нередко приводят к удорожанию проектов, срыву сроков и даже банкротству компаний.

> Согласно отчету KPMG "Знакомые проблемы — новые подходы" (2023), только 50% строительных проектов завершаются в срок, а 87% компаний отмечают возросший контроль над экономикой капитальных проектов. Основные проблемы связаны с нехваткой квалифицированных кадров и сложностью прогнозирования рисков [2].

Исторические данные о калькуляциях стоимости и времени выполнения процессов собираются в ходе строительства прошлых проектов на протяжении всей жизни строительной компании и вносятся в базы данных различных систем (ERP, PMIS BPM, EPM и т. д.).

> Наличие качественных исторических данных о калькуляциях – основное конкурентное преимущество строительной организации, напрямую влияющее на её выживаемость.

Отделы смет и калькуляций в строительных и инжиниринговых компаниях создаются для сбора, хранения и актуализации исторических данных о проектных расчётах. Их основная функция — накопление и систематизация опыта компании, что позволяет со временем повышать точность оценки объёмов, сроков и стоимости новых проектов. Такой подход помогает минимизировать ошибки в будущих расчётах, опираясь на практику и итоги уже реализованных проектов.

Методы расчета сметной стоимости проектов

В работе специалистов по калькуляциям применяются различные методы оценки, каждый из которых ориентирован на конкретный тип данных, доступность информации и уровень детализации проекта. Наиболее распространённые из них включают:

- **Ресурсный метод**: оценка сметной стоимости проекта на основе подробного анализа всех необходимых ресурсов, таких как материалы, оборудование и труд. Этот метод требует детального перечня всех задач и ресурсов, необходимых для выполнения каждой задачи, с последующим расчетом их стоимости. Этот метод отличается высокой точностью и широко применяется при составлении смет.
- **Параметрический метод**: использует статистические модели для оценки стоимости на основе параметров проекта. Это может включать анализ стоимости на единицу измерения, такую как площадь застройки или объем работы, и адаптацию этих стоимостей к конкретным условиям проекта. Метод особенно эффективен на ранних стадиях, когда детализированная информация ещё недоступна.
- **Метод единичных показателей** (метод единичной стоимости): рассчитывает сметную стоимость проекта, исходя из стоимости за единицу измерения (например, за квадратный метр или кубический метр). Это обеспечивает быстрое и удобное сравнение и анализ стоимости различных проектов или их частей.
- **Экспертные оценки** (метод Дельфи): основаны на мнениях экспертов, которые используют

свой опыт и знания для оценки стоимости проекта. Подход полезен, когда отсутствуют точные исходные данные или проект является уникальным.

Отдельно стоит отметить, что параметрический метод и экспертные оценки могут быть адаптированы под модели машинного обучения. Это позволяет автоматически строить прогнозы по стоимости и срокам реализации проекта на основе обучающих выборок. Примеры применения подобных моделей рассматриваются подробнее в главе «Пример использования машинного обучения для нахождения стоимости и сроков проекта» (Рис. 9.3-5).

Тем не менее, наиболее популярным и широко применяемым в мировой практике остаётся ресурсный метод. Он обеспечивает не только точную оценку сметной стоимости, но и позволяет рассчитывать продолжительность отдельных процессов на строительной площадке и всего проекта в целом (подробнее в главе «Графики строительства и данные 4D-проекта»).

Ресурсный метод составления смет и калькуляций в строительстве

Ресурсно-ориентированная калькуляция затрат — это метод управленческого учёта, при котором стоимость проекта формируется на основе прямого учёта всех задействованных ресурсов. В строительстве данный подход предполагает детальный анализ и оценку всех материальных, трудовых и технических ресурсов, необходимых для выполнения работ.

> Ресурсный метод, обеспечивает высокую степень прозрачности и точности при планировании бюджета, поскольку ориентирован на фактические цены ресурсов на момент составления сметы. Это особенно важно в условиях нестабильной экономической ситуации, когда ценовые колебания могут существенно повлиять на общую стоимость проекта.

В следующих главах мы детально разберем процесс калькуляций по ресурсному методу. Чтобы лучше понять его принципы в строительстве, будем проводить аналогию с расчетом стоимости ужина в ресторане. Менеджер ресторана, планируя вечер, составляет список необходимых продуктов, учитывает время приготовления каждого блюда, а затем умножает затраты на количество гостей. В строительстве процесс аналогичен: для каждой категории элементов проекта (объектов) формируются постатейные сметы рецепты, а общая стоимость проекта определяется суммированием всех затрат в общем счёте - итоговой смете по категориям.

Ключевым и начальным этапом ресурсного подхода является формирование исходной базы данных компании. На первом этапе калькуляции составляется структурированный перечень всех предметов, материалов, видов работ и ресурсов, которыми располагает компания в рамках своих строительных проектов – от гвоздя на складе до описание людей через их квалификацию и почасовую ставку. Эта информация систематизируется в единую «Базу данных строительных ресурсов и материалов» — табличный реестр, содержащий данные о наименованиях, характеристиках, единицах измерения и текущих ценах. Именно эта база становится главным и основным источником информации для всех последующих расчётов ресурсов — как стоимости, так и сроков выполнения работ.

База данных строительных ресурсов: каталог строительных материалов и работ

База данных или таблица строительных ресурсов и материалов - включает в себя подробную информацию о каждом элементе который может быть использован в строительном проекте – товаре, изделии, материале или услуге, включая его название, описание, единицу измерения и стоимость единицы, зафиксированные в структурированном виде. В этой таблице можно найти все: от разных видов топлива и материалов, используемых в проектах, до подробных списков специалистов в виде различных категорий с описанием почасовой оплаты (Рис. 5.1-2).

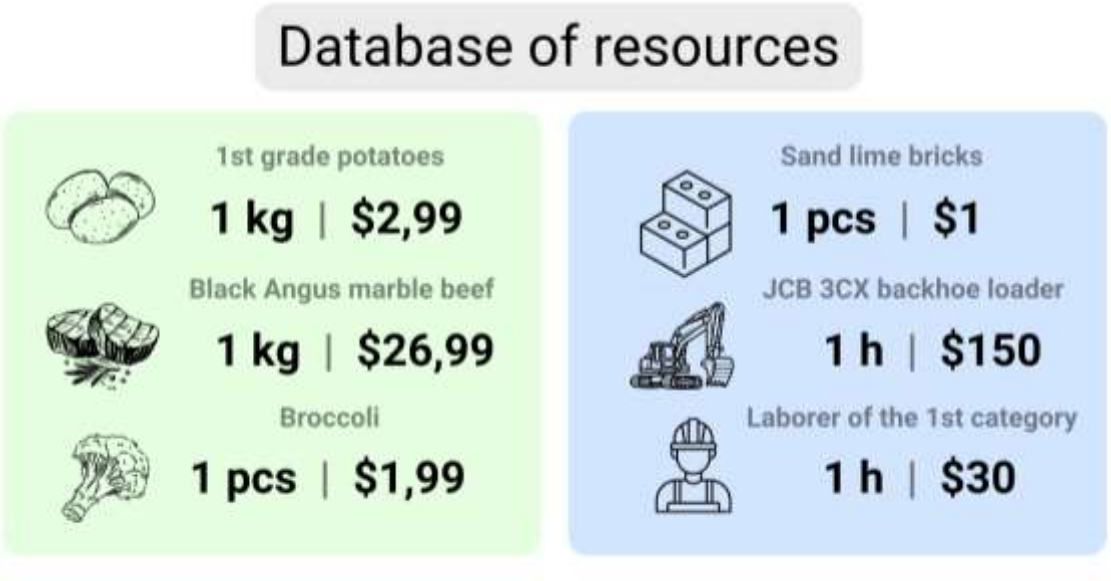


Рис. 5.1-2 Таблица ресурсов — это список ингредиентов, описывающий материал и услугу с указанием стоимости единицы.

> "База данных ресурсов" сродни каталогу товаров интернет-магазина, в котором каждый товар имеет подробное описание своих атрибутов. Это облегчает сметчикам выбор нужных ресурсов (как выбор товаров при добавлении в корзину), необходимых для расчета конкретных строительных процессов в виде калькуляций (итоговый заказ в интернет-магазине).

Базу данных ресурсов также можно представить как список всех ингредиентов в поваренной книге ресторана. Каждый строительный материал, оборудование и услуга подобны ингредиентам, используемым в рецептах. "База ресурсов" — это подробный список всех ингредиентов - строительных материалов и услуг, включая их стоимость за единицу: штуку, метр, час, литр и т. д.

Новые элементы сущности могут быть добавлены в таблицу "Базы данных строительных ресурсов" двумя способами - вручную (Рис. 5.1-3) или автоматически, путем интеграции с системами управления запасами компании или базами данных поставщиков.

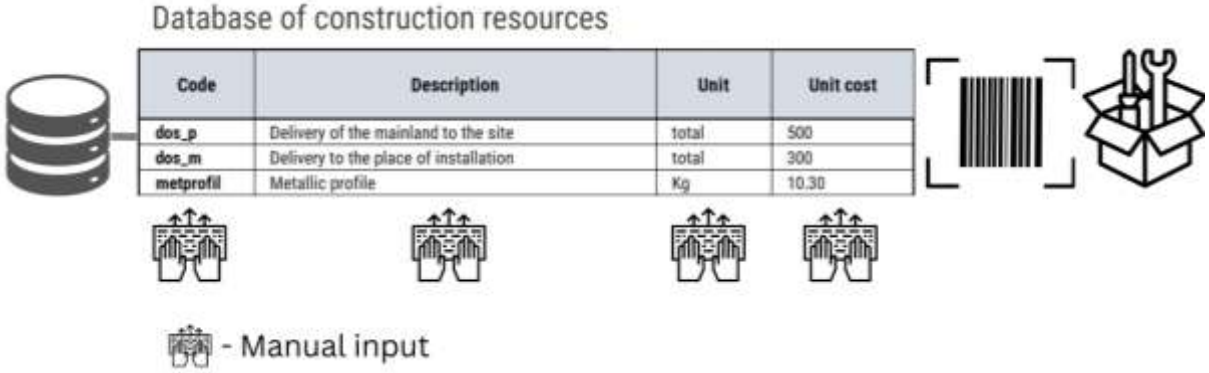


Рис. 5.1-3 База данных ресурсов заполняется вручную или автоматически перенимает данные из других баз данных.

Типичная строительная компания среднего размера использует базу данных, содержащую тысячи, а иногда и десятки тысяч элементов с детальным описанием, которые могут использоваться в строительных проектах. Эти данные затем автоматически используются в договорах и проектной документации для точного описания состава работ и процессов.

> Чтобы не отставать от меняющихся рыночных условий, таких как инфляция, атрибут "стоимость единицы продукции" для каждого продукта (товара или услуги) в базе данных ресурсов (Рис. 5.1-3) регулярно обновляется вручную или путем автоматической загрузки актуальных цен из других систем или онлайн платформ.

Обновление стоимости единицы ресурса может осуществляться ежемесячно, ежеквартально или ежегодно — в зависимости от характера ресурса, инфляции и внешнеэкономического климата. Такая актуализация необходима для поддержания точности расчётов и оценок, поскольку именно эти базовые элементы служат отправной точкой для работы сметчиков. На основе актуальных данных формируются сметы, бюджеты и графики, отражающие реальные условия рынка и снижая риски ошибок в последующих проектных расчётах.

Составление калькуляций и расчет стоимости работ на основе ресурсной базы

Наполнив "Базу данных строительных ресурсов" (Рис. 5.1-3) сущностями-минимальными единицами, можно приступать к созданию калькуляций, которые рассчитываются для каждого процесса или работы на строительной площадке за определенные единицы измерения: например, за один кубометр бетона, один квадратный метр гипсокартонной стены, за метр бордюра или за установку одного окна.

Например, для возведения кирпичной стены площадью 1 м² (Рис. 5.1-4), исходя из опыта предыдущих проектов, требуется примерно 65 кирпичей (сущность "Силикатный кирпич"), стоимостью $1 за штуку (атрибут "Стоимость за штуку"), что в сумме составляет $65. Также, по опыту, необходимо использовать строительную технику (сущность "Погрузчик JCB 3CX") в течение 10 минут, которая будет размещать кирпичи рядом с рабочей зоной. Поскольку аренда техники стоит 150 долларов в

час, 6 минут ее использования обойдутся примерно в 15 долларов. Кроме того, потребуется 2 часа работы подрядчика по укладке кирпича, почасовая ставка которого составляет 30 долларов, а общая сумма - 60 долларов.

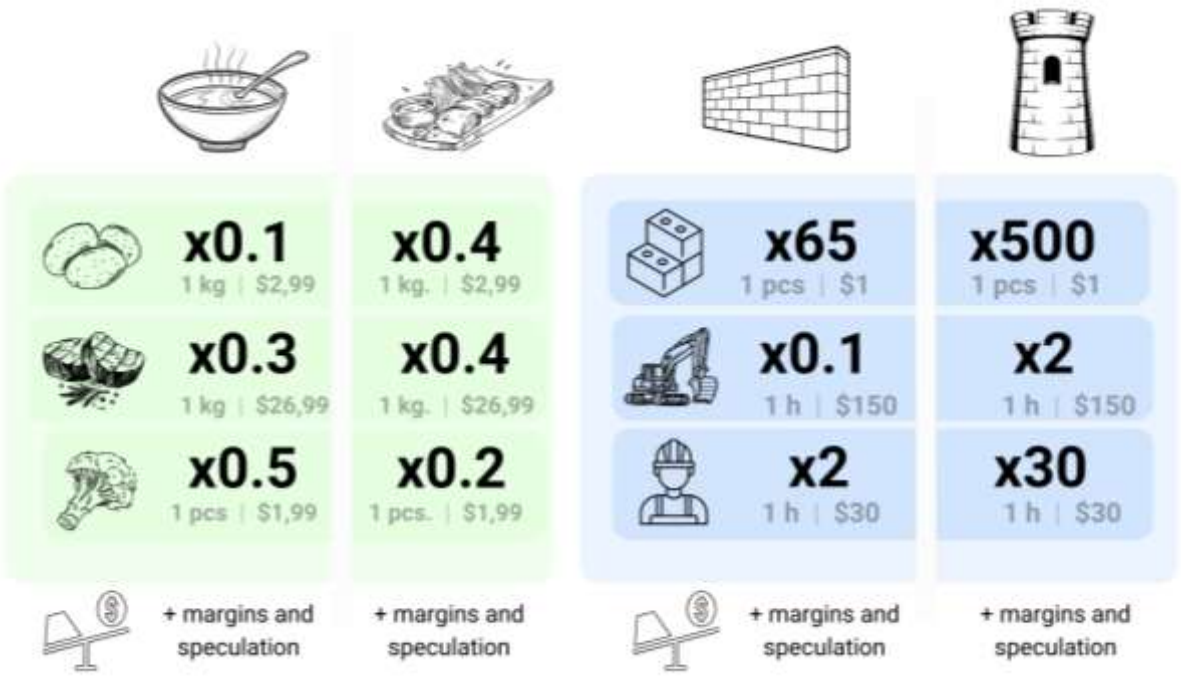


Рис. 5.1-4 Расчеты затрат содержат подробный перечень строительных материалов и услуг, необходимых для выполнения работ и процессов.

Состав калькуляций (так называемых "рецептов") формируется на основе исторического опыта, накопленного компанией в процессе выполнения большого объёма однотипных работ. Этот практический опыт, как правило, аккумулируется через обратную связь со строительной площадки. В частности, прораб собирает информацию непосредственно на строительной площадке, фиксируя фактические трудозатраты, расход материалов и нюансы выполнения технологических операций. Далее, в сотрудничестве с отделом смет, эта информация проходит итерационную доработку: описание процессов уточняется, состав ресурсов корректируется, а калькуляции актуализируются с учётом фактических данных последних проектов.

> Как в рецепте описываются необходимые ингредиенты и их количество для приготовления блюда, так и в сметной ведомости приводится подробный перечень всех строительных материалов, ресурсов и услуг, необходимых для выполнения конкретной работы или процесса.

Регулярно выполняемые работы позволяют рабочим, бригадирам и сметчикам ориентироваться в необходимом количестве ресурсов: материалах, топливе, рабочем времени и других параметрах, требуемых для выполнения единицы измерения работы (Рис. 5.1-5). Эти данные вводятся в смет-

ные системы в виде таблиц, где каждая задача и операция описывается через минимальные элементы ресурсной базы (с постоянно актуализируемыми ценами), что обеспечивает точность расчетов.

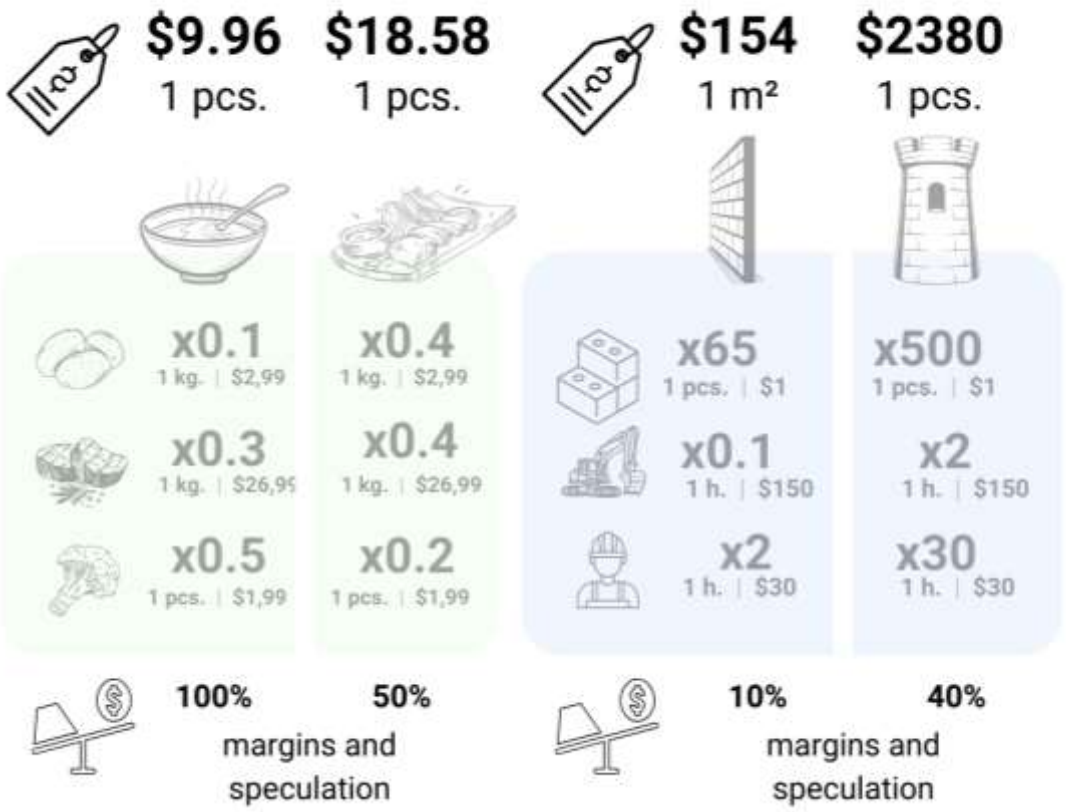


Рис. 5.1-5 Для каждой работы собираются единичные расценки, где атрибут объема сущности умножается на его количество с добавлением процента прибыли.

Чтобы получить общую стоимость каждого процесса или работ (объекта калькуляции), атрибут стоимости умножается на его количество и коэффициенты. Коэффициенты могут учитывать различные факторы, такие как сложность работы, региональные особенности, уровень инфляции, потенциальные риски (ожидаемый процент накладных расходов) или спекуляции (дополнительный коэффициент прибыли).

> Сметчик, как аналитик, преобразует опыт и рекомендации прораба в стандартные сметы, описывая строительные процессы через ресурсные сущности в табличном виде. По сути, задача сметчика собрать и структурировать через параметры и коэффициенты, информацию поступающую со строительной площадки.

Таким образом, итоговая стоимость единицы работы (например, квадратного или кубического метра, либо одной монтажа одной установки) включает не только прямые затраты на материалы и рабочую силу, но и наценки компании, накладные расходы, страхование и другие факторы (Рис. 5.1-6).

При это нам уже не нужно беспокоиться об актуальности цен в (рецептах) калькуляциях, так как реальные цены всегда отражены в «базе ресурсов» (таблице ингредиентов). На уровне калькуляций в таблицу автоматически (например, по коду элемента или его уникальному идентификатору) подгружаются данные из базы ресурсов, которые подгружают описание, и актуальную стоимость за единицу, которые в свою очередь могут быть автоматически подгружены с онлайн платформ или интернет-магазина строительных материалов. Сметчику на уровне калькуляций работ остаётся только описать работу или процесс через атрибут “количество ресурсов” и дополнительные факторы.

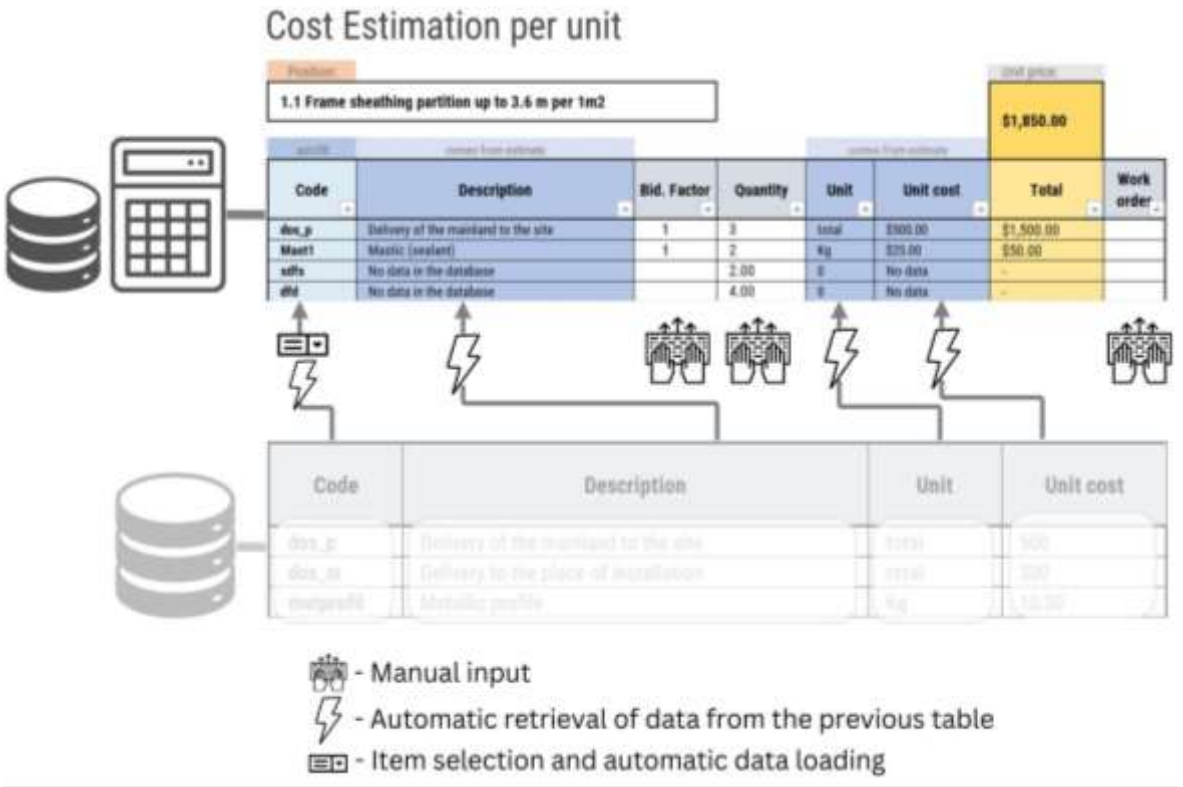


Рис. 5.1-6 На этапе расчета единицы стоимости работ заполняются только атрибуты количества необходимых ресурсов, всё остальное автоматически подгружается из базы ресурсов.

Созданные калькуляции стоимости работ хранятся в виде таблиц-шаблонов типовых проектов, которые напрямую связаны с базой данных строительных ресурсов и материалов. Эти шаблоны представляют собой стандартизированные рецепты выполнения повторяющихся видов работ, для будущих проектов, обеспечивая единообразие в расчётах на уровне всей компании.

При изменении стоимости любого ресурса в базе данных (Рис. 5.1-3) — будь то вручную или автоматически через загрузку актуальных рыночных цен (например, в условиях инфляции) — обновления немедленно отражаются во всех связанных калькуляциях (Рис. 5.1-6). Это означает, что достаточно внести изменения лишь в ресурсную базу, тогда как шаблоны калькуляций и смет остаются неизменными на протяжении долгого периода времени. Такой подход обеспечивает стабильность и воспроизводимость расчётов при любых колебаниях цен, которые учитываются только в относительно простой таблице ресурсов (Рис. 5.1-3).

Для каждого нового проекта создаётся копия стандартного шаблона калькуляции, что позволяет вносить изменения и корректировать виды деятельности в соответствии с особыми требованиями без изменения оригинального шаблона, принятого в компании. Такой подход обеспечивает гибкость в адаптации расчётов: можно учитывать особенности строительной площадки, пожелания заказчика, вводить коэффициенты риска или рентабельности (спекуляций) — и всё это без нарушения стандартов компании. Это помогает компании находить баланс между максимизацией прибыли, удовлетворением запросов заказчиков и сохранением своей конкурентоспособности.

В некоторых странах подобные шаблоны калькуляций, накопленные в течение десятилетий, стандартизируются на национальном уровне и становятся частью государственных стандартов системы расчёта стоимости строительных работ (Рис. 5.1-7).

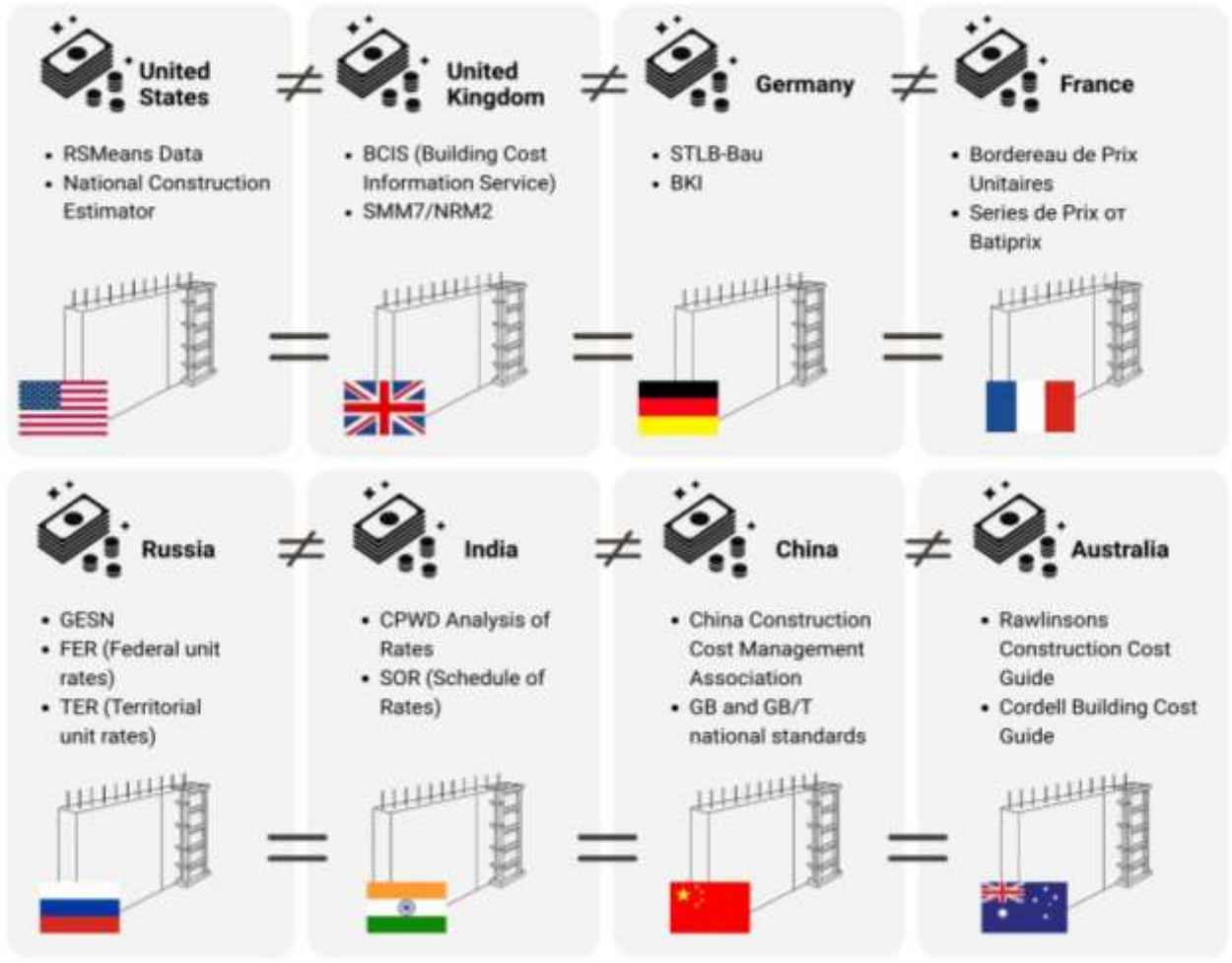


Рис. 5.1-7 В разных странах мира для калькуляций одного и того же элемента существуют собственные правила калькуляций со своими (рецептами) сборниками и нормативами для строительных работ.

Такие стандартизированные ресурсные базы смет (Рис. 5.1-7) обязательны для использования всеми участниками рынка, прежде всего, при реализации проектов с государственным финансированием. Подобная стандартизация обеспечивает заказчику прозрачность, сопоставимость и справедливость в формировании цен и контрактных обязательств.

Итоговый расчет стоимости проекта: от смет до бюджета

Государственные и отраслевые сметные нормативы играют различную роль в строительной практике разных стран. В то время как некоторые государства обязывают строго соблюдать единые нормативы, большинство развитых экономик принимает более гибкий подход. В странах с рыночной экономикой государственные строительные нормативы обычно служат лишь базовой точкой отсчета. Строительные компании адаптируют эти стандарты под свои операционные модели или полностью их перерабатывают, дополняя их собственными коэффициентами, учитывающими специфику их деятельности. Эти корректировки отражают корпоративный опыт, эффективность управления ресурсами и часто факторы, в которых например может быть учтена спекулятивная прибыль компании.

> В итоге уровень конкуренции, рыночный спрос, целевая маржинальность и даже взаимоотношения с конкретными заказчиками могут приводить к значительным отклонениям от стандартных норм. Эта практика обеспечивает гибкость рынка, но одновременно усложняет прозрачное сравнение предложений разных подрядчиков, вводя на данном этапе калькуляций элемент спекулятивного ценообразования в строительную отрасль.

После того как шаблоны расчётов для отдельных видов работ и процессов уже подготовлены — или, что бывает чаще, просто скопированы из типовых государственных смет (Рис. 5.1-7) с внесёнными коэффициентами, отражающими «особенности» конкретной компании, — на завершающем этапе остаётся лишь умножить стоимость каждой позиции на соответствующий атрибут объёма работ или процессов в новом проекте.

> При расчете общей стоимости нового строительного проекта ключевым этапом является суммирование затрат по всем статьям калькуляций, умноженное на объем этих позиций-работ в проекте.

Чтобы создать общую стоимость проекта, в нашем упрощенном примере, мы начнем с расчета стоимости строительства одного квадратного метра стены и умножим стоимости его расчета (например, работа "1м² стандартной установки стеновых элементов") на общее количество квадратных метров стен в проекте (например, атрибут "Площадь" или "Количество" (Рис. 5.1-8) сущности типа "Стеновые элементы" из CAD проекта или расчётов прораба).

Аналогично мы рассчитываем стоимость для всех элементов проекта (Рис. 5.1-8): берём стоимость единицы работы и умножаем её на объём конкретного элемента или его группы в данном проекте. Сметчику остаётся только внести количество данных элементов, работ или процессов в проекте в

виде объёма или количества. Это позволяет автоматически формировать полную смету строительства.

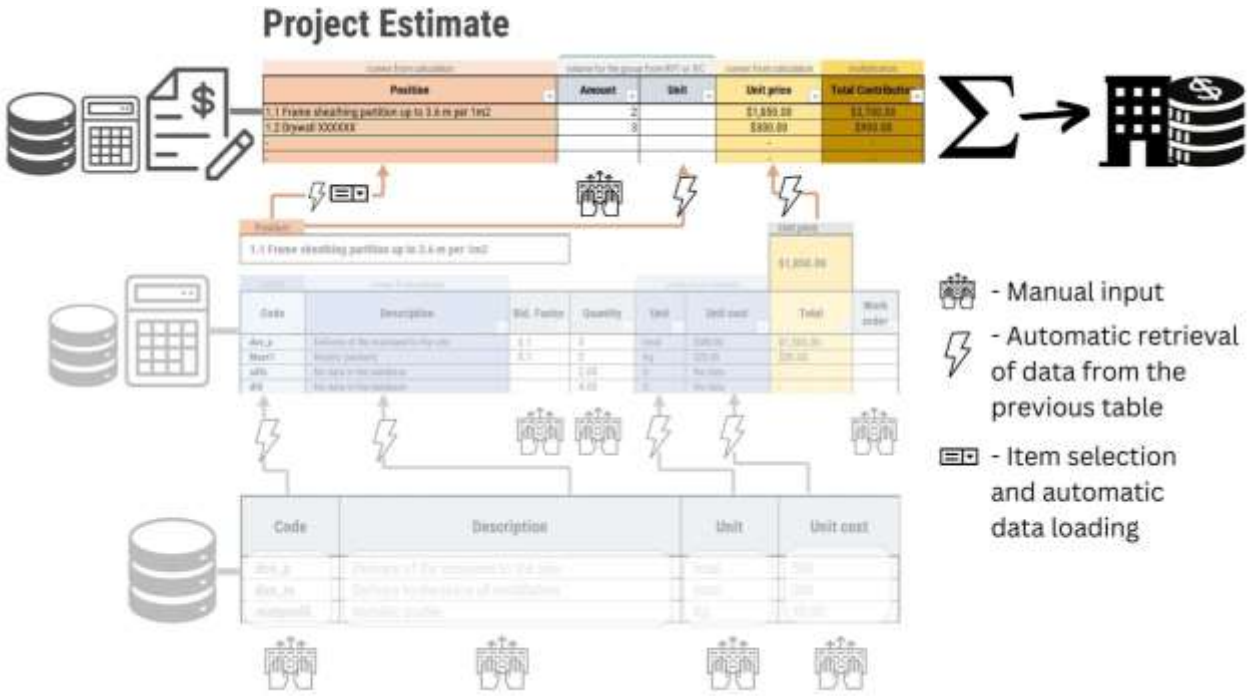


Рис. 5.1-8 На этапе создания сметы мы вносим только объем работ.

Как и в случае с калькуляциями, на это уровне мы подгружаем автоматически готовые обсчитанные позиции (из шаблона калькуляций или новые, скопированные из шаблона и подредактированные), которые автоматически приносят с собой актуальную стоимость на единицу работы (которая актуализируется автоматически из базы данных ресурсов (Рис. 5.1-8 нижняя таблица)). Соответственно при любом изменении данных в ресурсной базе или калькуляционных таблицах – данные в смете будут автоматически актуализированы на текущий день, без необходимости изменять калькуляции или саму смету.

В контексте ресторана конечная стоимость мероприятия рассчитывается аналогичным образом и равна конечной стоимости всего ужина, где стоимость каждого блюда, умноженная на количество гостей, складывается в общую стоимость чека (Рис. 5.1-9). И так же как и в строительстве рецепты приготовления блюд в ресторане могут не меняться десятилетиями. В отличии от цен, где стоимость ингредиентов может меняться каждый час.

> Как владелец ресторана умножает стоимость каждого блюда на количество порций и людей, чтобы определить общую стоимость мероприятия, так и менеджер по сметам суммирует стоимость всех компонентов проекта, чтобы получить полную смету строительства.

Таким образом, для каждой работы в проекте определяется ее конечная стоимость (Рис. 5.1-9), которая, умноженная на атрибутивный объем сущности, соответствующей этой работе, - дает стоимость групп работ, из которой получается конечная стоимость всего проекта.

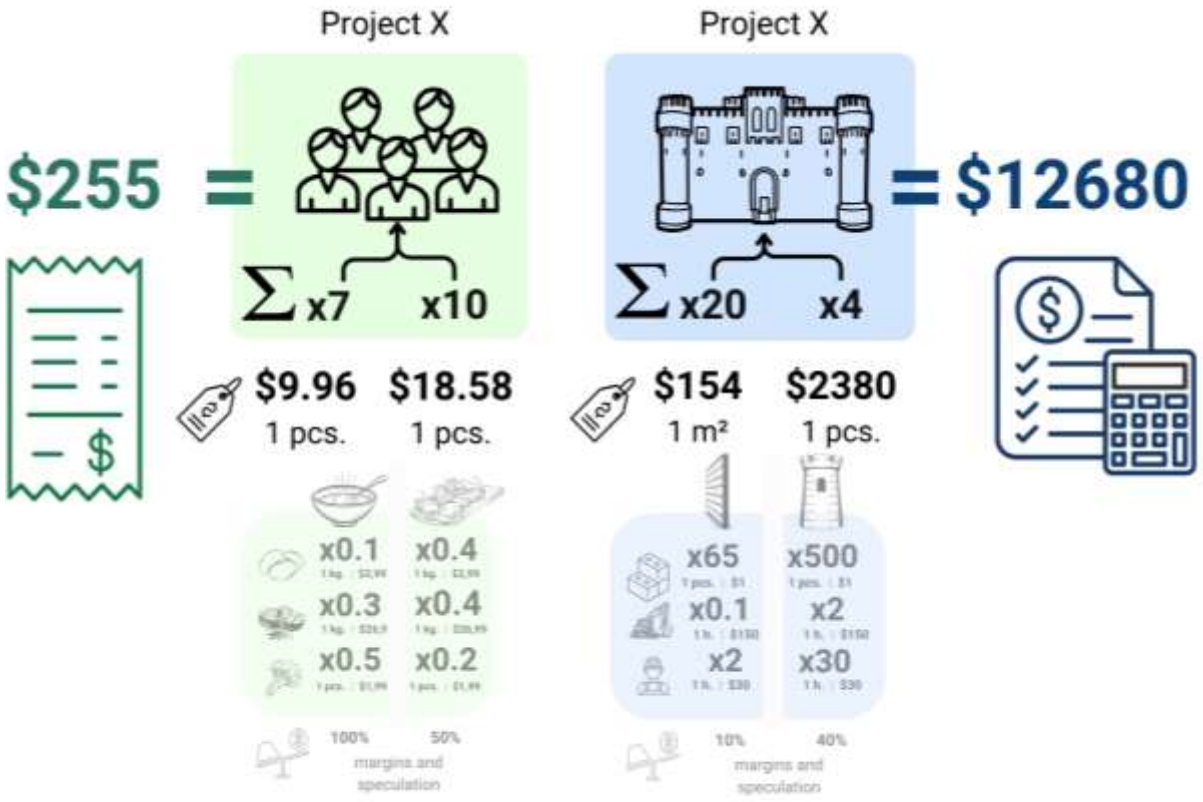


Рис. 5.1-9 Итоговая смета рассчитывается путем суммирования атрибута стоимости работ каждого элемента на атрибут его объема.

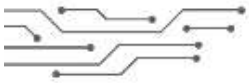
Итоговая стоимость проекта (Рис. 5.1-8) представляет собой финансовую картину проекта, позволяя заказчикам, инвесторам или финансовым организациям понять общий бюджет и финансовые ресурсы, необходимые для реализации проекта, на любой день, с учётом актуальных цен.

И если процессы составления ресурсных баз, расчетов и смет (рецептов процессов) уже отработаны, полуавтоматизированы и отточены десятками тысяч лет и записанные на государственном уровне, то автоматическое получение качественной информации об объеме и количестве элементов для последнего этапа финальной сметы – сегодня остается узким местом в процессах всех расчетов стоимостных и временных атрибутов проекта, да и в целом общего бюджета проекта.

> На протяжении тысячелетий традиционным методом вычисления объемов были ручные способы измерения объемных и количественных характеристик с помощью плоских чертежей. С наступлением цифровой эры компании обнаружили, что информацию об объёмах и количестве теперь можно автоматически извлекать из геометрических данных, содержащихся в моделях CAD, что произвело революцию в тысячелетних способах получения количественных данных.

Современные подходы к оценке процессов и сметному делу предполагают автоматическое извлечение объемных и количественных атрибутов из CAD баз данных, которые можно выгрузить и

подсоединить к процессу калькуляций, чтобы получать актуальные объёмы групп проектов на любом этапе проектирования до эксплуатации.



ГЛАВА 5.2.
QUANTITY TAKE-OFF И АВТОМАТИЧЕСКОЕ СОЗДАНИЕ СМЕТ И КАЛЕНДАРНЫХ ПЛАНОВ

Переход от 3D к 4D и 5D: использование объёмных и количественных параметров

Имея в распоряжении таблицы калькуляций с описанными процессами через ресурсы (Рис. 5.1-8), следующим шагом становится автоматическое получение параметров объёма или количества для группы элементов, которые необходимы для расчётов и для составления итоговой сметы.

Объёмные характеристики элементов проекта — например, стен или перекрытий — могут быть автоматически извлечены из CAD баз данных. Параметрические объекты, созданные в CAD программах, преобразуются средствами геометрического ядра в числовые значения параметров длины, ширины, площади, объёма и другие. Подробнее процесс получения объёмов на основе 3D-геометрии будет рассмотрен в следующей, шестой части (Рис. 6.3-3), посвящённой работе с CAD (BIM). Кроме объёмов, количество однотипных элементов также может быть получено из базы данных CAD -модели путём фильтрации и группировки объектов по категориям и свойствам. Эти параметры, позволяющие проводить группировку — становятся основой для связывания элементов проекта через ресурсные калькуляции с расчётами, итоговой сметой и бюджетом всего проекта.

Таким образом, модель данных, извлечённая из 3D (CAD) модели, дополняется новыми слоями параметров, обозначаемыми как 4D и 5D. В новых слоях атрибутов сущностей - 4D (время) и 5D (стоимость) - геометрические данные 3D используются в качестве источника значений атрибутов объемов сущностей.

- **4D** — информационный слой параметров, который добавляет к 3D-параметрам элементов сведения о продолжительности выполнения строительных операций. Эти данные необходимы для планирования графиков работ и управления сроками реализации проекта.
- **5D** — следующий уровень расширения модели данных, в котором элементы дополняются стоимостными характеристиками. Таким образом, к геометрической информации добавляется финансовый аспект: стоимость материалов, работ и оборудования, что позволяет выполнять расчёты бюджета, анализировать рентабельность и управлять затратами в процессе строительства.

Данные о стоимости и 3D, 4D и 5D атрибутивных свойств групп сущностей проекта описываются подобно калькуляциям в модульных ERP, PIMS-системах (или Excel-подобных инструментах) и используются для автоматического расчета стоимости и бюджетного планирования как отдельных групп, так и полного бюджета проекта.

Атрибуты 5D и получение объемов атрибутов из CAD

При подготовке окончательной сметы строительного проекта, составление которой мы рассматривали в предыдущих главах (Рис. 5.1-8), атрибуты объема для каждой категории элементов про-

екта либо собираются вручную, либо извлекаются из спецификаций атрибутов объема, предоставляемых программами CAD.

> Традиционный ручной метод расчёта объёмов предполагает, что прораб и сметчик анализируют чертежи, представленные в течение тысячелетий в виде линий на бумаге, а последние 30 лет – в цифровых форматах PDF (PLT) или DWG. Опираясь на профессиональный опыт, они измеряют объёмы работ и необходимых материалов, нередко с помощью линейки и транспортира. Этот метод требует значительных усилий и времени, а также особого внимания к деталям.

Определение атрибутов объема работ таким способом может занять от нескольких дней до нескольких месяцев, в зависимости от масштаба проекта. Кроме того, поскольку все измерения и расчеты выполняются вручную, существует риск человеческой ошибки, которая может привести к неточным данным, что впоследствии скажется на ошибках в оценке времени и стоимости проекта, за которые будет нести ответственность вся компания.

Современные методы, основанные на использовании баз данных CAD значительно упрощают вычисление объёмов. В моделях CAD геометрия элементов уже включает атрибуты объёма, которые могут быть автоматически рассчитаны (через геометрическое ядро (Рис. 6.3-3)) и представлены или экспортированы в табличной форме.

В подобном сценарии сметный отдел запрашивает у CAD проектировщика данные о количественных и объёмных характеристиках элементов проекта. Эти данные экспортируются в виде таблиц или напрямую интегрируются в калькуляционные базы — будь то Excel, ERP или PMIS-системы. Такой процесс зачастую начинается не с формального запроса, а с краткого диалога между заказчиков (инициатором) и архитектором и сметчиком со стороны строительной или проектировочной компании. Ниже приведён упрощённый пример, демонстрирующий, как из повседневной коммуникации формируется структурированная таблица для автоматических расчётов (QTO):

- Заказчик – «*Хочу добавить ещё один этаж к зданию, в такой же конфигурации как и второй этаж*»
- Архитектор (CAD) – «*Добавляем третий этаж, конфигурация та же, что на втором*». И после этого сообщения посылает новую версию CAD проекта сметчику.
- Сметчик автоматически проводит группировку и расчёт (ERP, PMIS, Excel) – «*Проведу проект через Excel-таблицу с правилами QTO (ERP, PMIS), получу объёмы по категориям для нового этажа и сформирую смету*».

В итоге текстовой диалог трансформируется в структуру таблицы с правилами группировки:

Элемент	Категория	Этаж
Перекрытие	OST_Floors	3
Колонна	OST_StructuralColumns	3
Лестничный марш	OST_Stairs	3

После процесса автоматической группировки CAD модели от проектировщика по правилам QTO сметчика и автоматического умножения объёмов на ресурсные калькуляции (Рис. 5.1-8) получаем следующие результаты, которые отправляются заказчику:

Элемент	Объём	Этаж	Цена за ед.	Итоговая стоимость
Перекрытие	420 м²	3	150 €/м²	**63 000 €**
Колонна	4 шт.	3	2450 €/шт.	**9 800 €**
Лестничный марш	2 шт.	3	4 300 €/шт.	**8 600 €**
ИТОГО:	—	—	—	**81 400 €**

- Заказчик – «*Спасибо, достаточно много, нужно урезать несколько помещений*». И цикл повторяется много раз.

Подобный сценарий может повторяться многократно, особенно в фазе согласований, где заказчик ожидает мгновенной обратной связи. Однако на практике такие процессы могут затягиваться на дни или даже недели. Сегодня, благодаря внедрению автоматических правил группировки и расчётов, действия, ранее занимавшие значительное время, должны укладываться в считанные минуты. Автоматизированное получение объёмов, через правила группировки, не только ускоряет расчёты и формирование смет, но и за счёт минимизации человеческого фактора снижает вероятность ошибок, обеспечивая прозрачную и точную оценку стоимости проекта.

> Если при создании 3D-модели в CAD-системе изначально учитывались требования сметного отдела (что на практике пока встречается редко), а имена, идентификаторы групп элементов и их классификационные признаки заданы в виде параметров, совпадающих со структурами сметных групп и классов, то объёмные атрибуты можно автоматически передавать в сметные системы без дополнительных преобразований.

Автоматическое извлечение объёмных атрибутов из CAD в виде таблиц-спецификаций позволяет оперативно получать актуальные данные о стоимости отдельных работ и проекта в целом (Рис.

5.2-1). Обновляя только CAD файл с проектными объёмами в процессе расчёта или калькуляционной системе, компания может быстро пересчитать смету с учётом последних изменений, обеспечивая высокую точность и согласованность всех последующих расчётов.

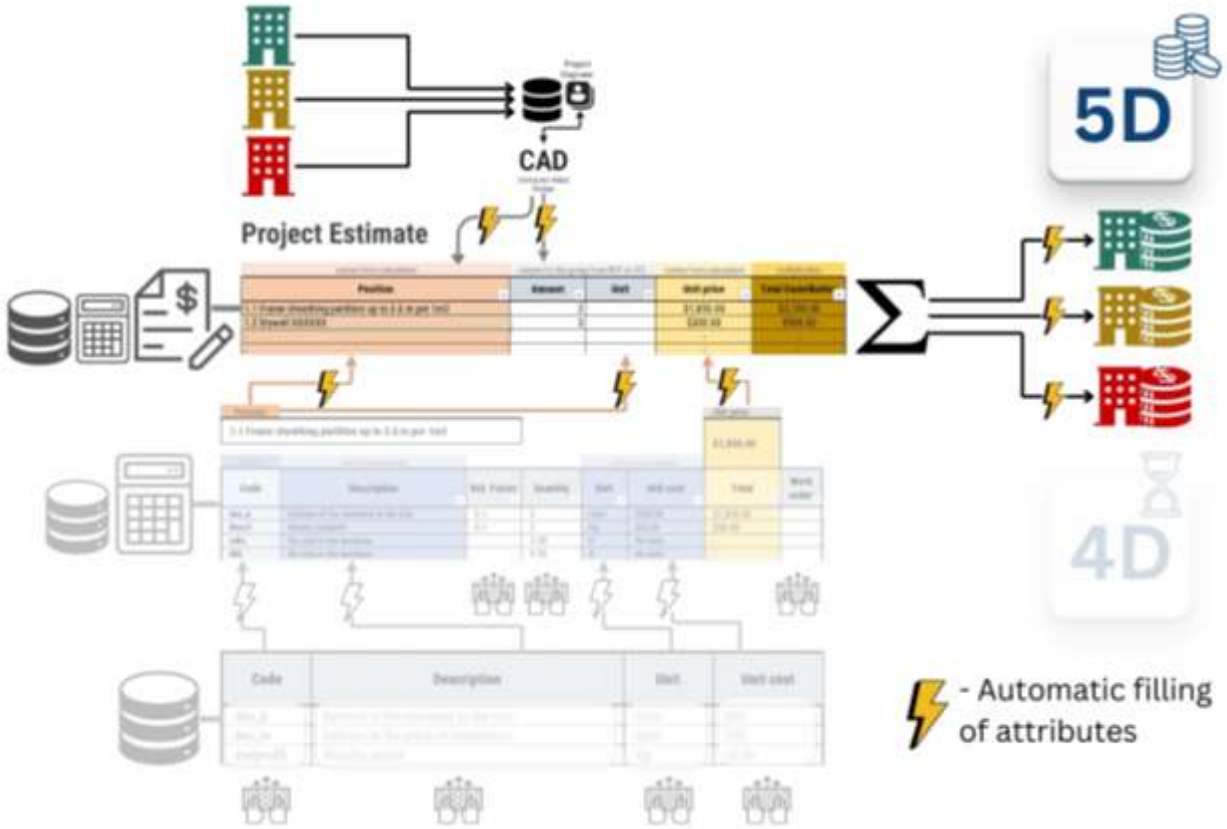


Рис. 5.2-1 Атрибуты объемов из таблиц или баз данных CAD автоматически вводятся в смету, позволяя мгновенно рассчитать общую стоимость проекта.

В условиях растущей сложности капитальных проектов расчёт полного бюджета и аналитика общей стоимости проектов по подобному сценарию (Рис. 5.2-1) - становится ключевым инструментом для принятия обоснованных решений.

> Согласно исследованию Accenture «Создание большей ценности с помощью капитальных проектов» (2024) [20], ведущие компании активно интегрируют анализ данных в цифровые инициативы, используя историческую информацию для прогнозирования и оптимизации результатов. Исследование показывает, что всё больше владельцев-операторов применяют аналитику больших данных для предсказания рыночных тенденций и оценки коммерческой жизнеспособности еще до начала проектирования. Это достигается за счет анализа хранилищ данных из существующего портфеля проектов. Кроме того, 79% владельцев-операторов внедряют «надёжную» прогнозную аналитику для оценки эффективности проектов и поддержки оперативного принятия решений в режиме реального времени.

Современное эффективное управление строительными проектами неразрывно связано с обработ-

кой и анализом больших объемов информации на всех этапах проектирования и тех процессов которые предшествуют проектированию. Использование хранилищ данных, ресурсных калькуляций, прогнозных моделей и машинного обучения позволяет не только минимизировать риски при расчётах, но и принимать стратегические решения по финансированию проекта на ранних стадиях проектирования. Про хранилища данных и прогнозные модели, которые будут дополнять калькуляции мы будем подробнее говорить в девятой части книги.

Автоматическое получение объёмных параметров элементов из проектов CAD, которые необходимо для составления смет, осуществляется с помощью инструментов группировки QTO (Quantity Take-Off). QTO инструменты работают путем группировки всех объектов проекта по специальным идентификаторам элементов или параметрам атрибутов элементов, используя спецификации и таблицы, созданные в базе данных CAD.

QTO Quantity Take-Off: группировка проектных данных по атрибутам

Расчёт объёмных параметров и количества материалов (QTO — Quantity Take-Off) в строительстве представляет собой процесс извлечения количественных характеристик элементов, необходимых для реализации проекта. На практике QTO часто остаётся полу ручным процессом, включающим сбор данных из различных источников: PDF-документов, чертежей в формате DWG и цифровых моделей CAD.

При работе с данными, извлечёнными из CAD-баз данных, процесс количественной оценки (QTO) реализуется как последовательность операций фильтрации, сортировки, группировки и агрегации. Элементы модели отбираются по параметрам классов, категорий и типов, после чего их количественные атрибуты — такие как объём, площадь, длина или количество — суммируются в соответствии с логикой расчёта (Рис. 5.2-2).

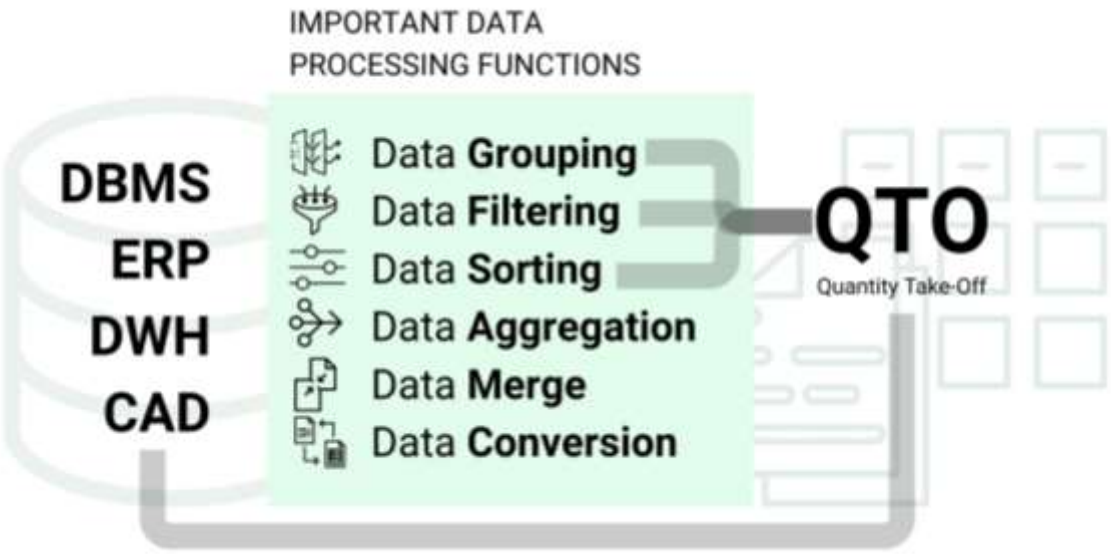


Рис. 5.2-2 Группировка и фильтрация данных — это самые популярные функции применяемые к базам данных и хранилищам данных.

Процесс QTO (фильтрации и группировки) позволяет систематизировать данные, формировать спецификации и готовить исходную информацию для расчёта смет, закупок и графиков выполнения работ. Основой QTO является классификация элементов по типу измеряемых атрибутов. Для каждого элемента или группы элементов выбирается соответствующий параметр количественного измерения. Например:

- **Атрибут длины** (бордюрный камень – в метрах)
- **Атрибут площади** (гипсокартонные работы – в квадратных метрах)
- **Атрибут объёма** (бетонные работы – в кубических метрах)
- **Атрибут количества** (окна – поштучно)

Помимо объёмных характеристик, генерируемых математически на основе геометрии, после группировки QTO при расчётах часто применяются коэффициенты перерасхода (Рис. 5.2-12 например, 1,1 для учёта 10% при логистике и монтаже) — корректирующие значения, учитывающие потери, особенности монтажа, складирования или транспортировки. Это позволяет более точно прогнозировать фактический расход материалов и избегать как нехватки, так и избыточного запаса на строительной площадке.

> Автоматический процесс количественного учета (QTO) необходим для составления точных калькуляций и смет, снижая человеческий фактор в процессах нахождения объёмных характеристик и предотвращая переизбыток или недостаток при заказе материалов.

В качестве примера процесса QTO, рассмотрим распространенный случай, когда необходимо показать из базы данных CAD таблицу-спецификацию объемов по типам элементов для определенной категории, классов элементов. Сгруппируем все элементы проекта по типам из категории стен проекта CAD и просуммируем атрибуты объема для каждого типа, чтобы представить результат в виде таблицы объемов QTO (Рис. 5.2-3).

В примере типового проекта CAD (Рис. 5.2-3) все элементы категории стен внутри базы данных CAD сгруппированы по типам стен, например, "Lamelle 11.5", "MW 11.5" и "STB 20.0", и имеют четко определенные атрибуты объема, представленные в метрических кубах.

Цель менеджера, находящегося на стыке между проектировщиками и специалистами по расчётам, — получить автоматизированную таблицу объёмов по типам элементов в выбранной категории. Причём не только для конкретного проекта, но и в универсальном виде, применимом к другим проектам с аналогичной структурой модели (**Ошибка! Источник ссылки не найден.**). Это позволяет масштабировать подход и обеспечивать повторное использование данных без дублирования усилий.

> Прошли те времена, когда опытные прорабы и сметчики вооружались линейкой, тщательно измеряя каждую линию на бумаге или PDF-планах - традиция, которая не менялась прошедшие тысячелетия. С развитием 3D-моделирования, где геометрия каждого элемента теперь напрямую связана с автоматически рассчитываемыми объёмными атрибутами, процесс определения объёмов и количества QTO стал автоматизированным.

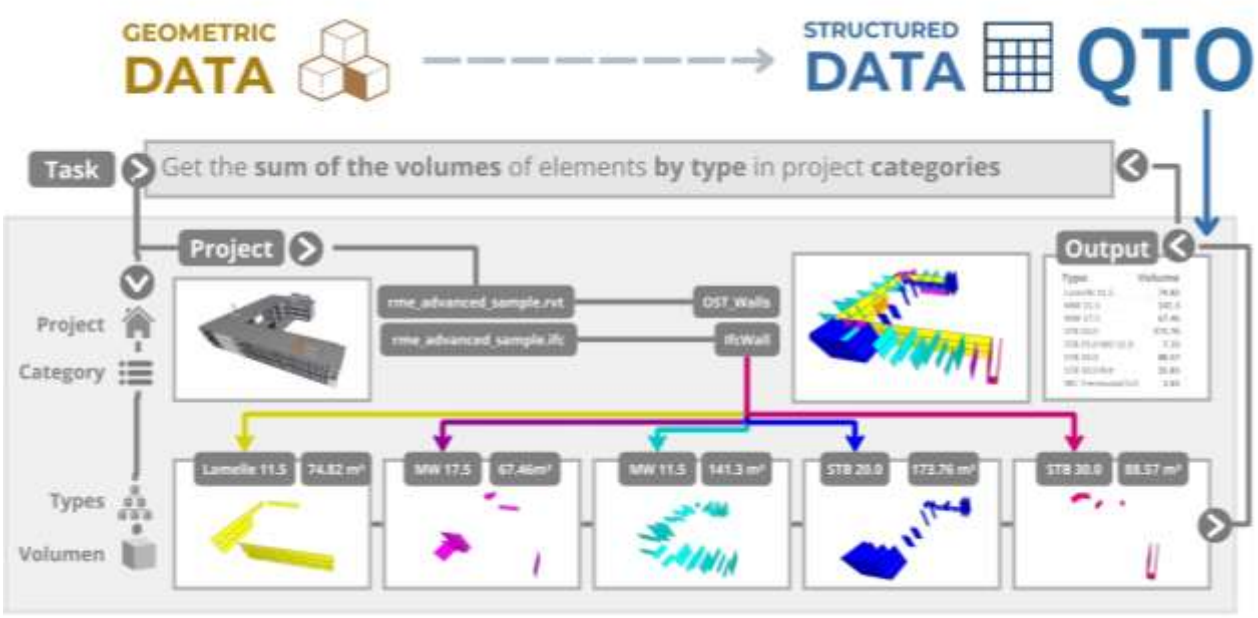


Рис. 5.2-3 Получение атрибутов объема и количества QTO из проекта подразумевает под собой группировку и фильтрацию элементов проекта.

В нашем примере задача состоит в том, чтобы "выбрать категорию стен в проекте, сгруппировать все элементы по типу и представить информацию об атрибутах объема в структурированном, табличном формате", чтобы эту таблицу могли использовать десятки других специалистов для расчётов калькуляций, логистики, графиков работ и других бизнес кейсов (Рис. 6.1-3).

Из-за закрытого характера данных CAD не каждый специалист сегодня может использовать прямой доступ к базе данных CAD (про причины и решения проблемы доступа подробно в шестой части книги). Поэтому многие вынуждены обращаться к специальным инструментам BIM, основанным на концепциях open BIM и closed BIM [63]. При работе со специализированными BIM-инструментами или напрямую в среде CAD-программы, таблицу с результатами QTO (Quantity Take-Off) можно сформировать разными способами — в зависимости от того, используется ли ручной интерфейс или программная автоматизация.

Например, используя пользовательский интерфейс CAD (BIM)-программы, достаточно выполнить около 17 действий (нажатий кнопок), чтобы получить готовую таблицу объёмов (Рис. 5.2-4). Однако для этого пользователь должен хорошо понимать структуру модели и функции программного обеспечения CAD (BIM).

Если применяется автоматизация через программный код или через плагины и API инструменты внутри CAD программ, количество ручных действий по получению таблиц объёмов сокращается, но потребуется написать от 40 до 150 строк кода, в зависимости от используемой библиотеки или инструмента:

- **IfcOpSh (open BIM)** или **Dynamo IronPython (closed BIM)** — позволяют получить таблицу QTO из CAD формата или CAD программы всего за ~40 строк кода.
- **IFC_js (open BIM)** — требует примерно 150 строк кода для извлечения объёмных атрибутов из IFC-модели.

- **Интерфейсные инструменты CAD (BIM)** — позволяют получить тот же результат вручную, за 17 нажатий мышкой.

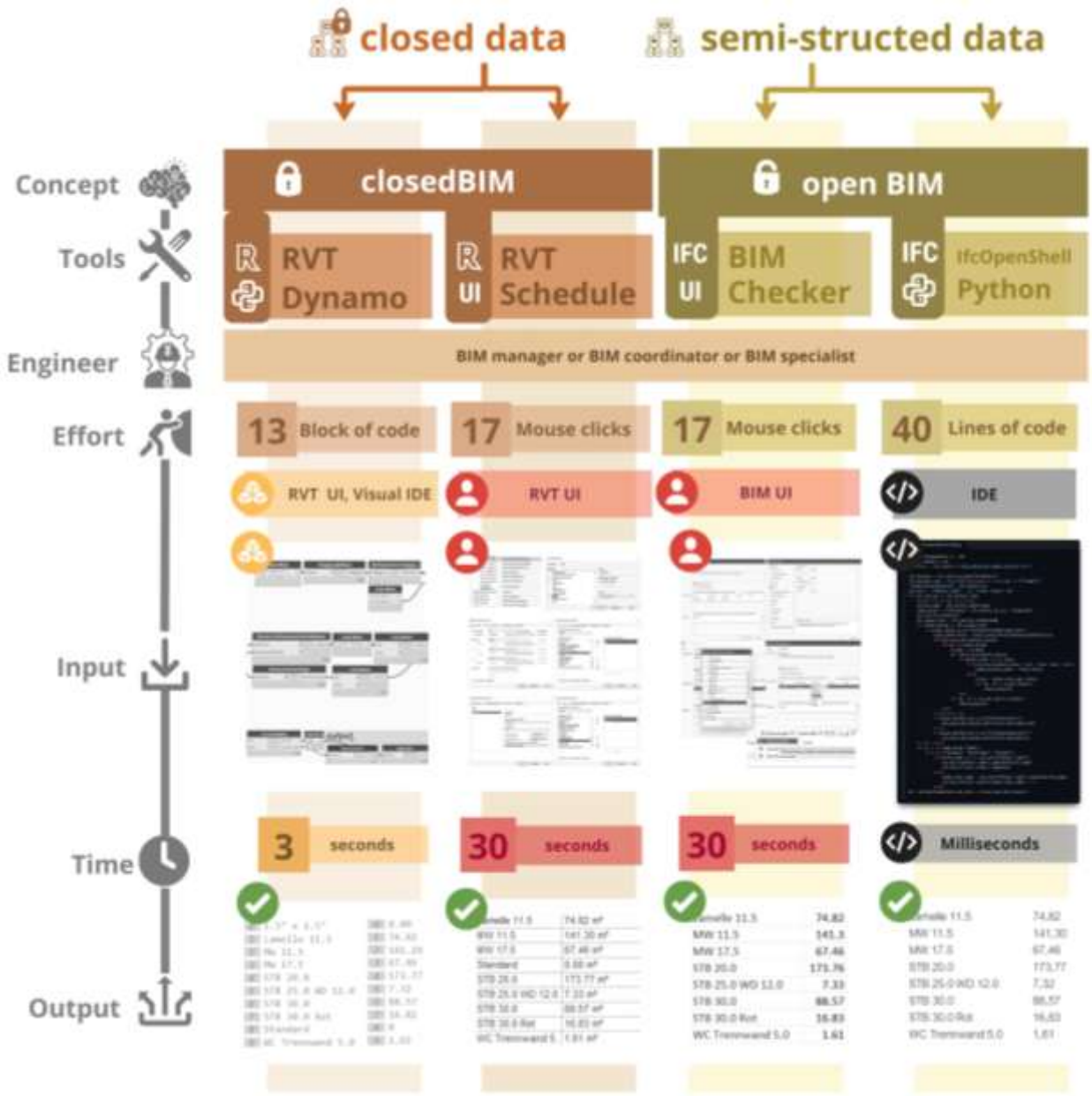


Рис. 5.2-4 Проектировщики и менеджеры CAD (BIM), используют от 40 до 150 строк кода или десяток нажатий кнопок для создания QTO таблиц

В результате итог один и тот же — структурированная таблица с атрибутами объёмов для группы элементов. Различие лишь в трудозатратах и необходимом уровне технической подготовки пользователя (Рис. 5.2-4). Современные инструменты, по отношению к ручному сбору объёмов, значительно ускоряют процесс QTO и снижают вероятность ошибок. Они позволяют извлекать данные непосредственно из модели проекта, исключая необходимость вручную пересчитывать объёмы по чертежам, как это делалось ранее.

Независимо от используемого метода – будь то open BIM или closed BIM – можно получить идентичную QTO-таблицу с объёмами элементов проекта (Рис. 5.2-4). Однако при работе с проектными данными в концепциях CAD- (BIM-) пользователи зависят от специализированных инструментов и API, предоставляемых вендорами (Рис. 3.2-13). Это создаёт дополнительные уровни зависимости и требует изучения уникальных схем данных, одновременно ограничивая прямой доступ к данным.

Из-за закрытости CAD-данных получение QTO-таблиц и других параметров усложняется автоматизация расчётов и интеграция с внешними системами. С помощью инструментов прямого доступа к базам данных и перевода CAD-данных проекта при помощи инструментов обратного инжиниринга в открытый структурированный формат датафрейма (Рис. 4.1-13) идентичную QTO таблицу можно получить всего лишь одной строчкой кода (Рис. 5.2-5 – вариант с гранулированными данными).

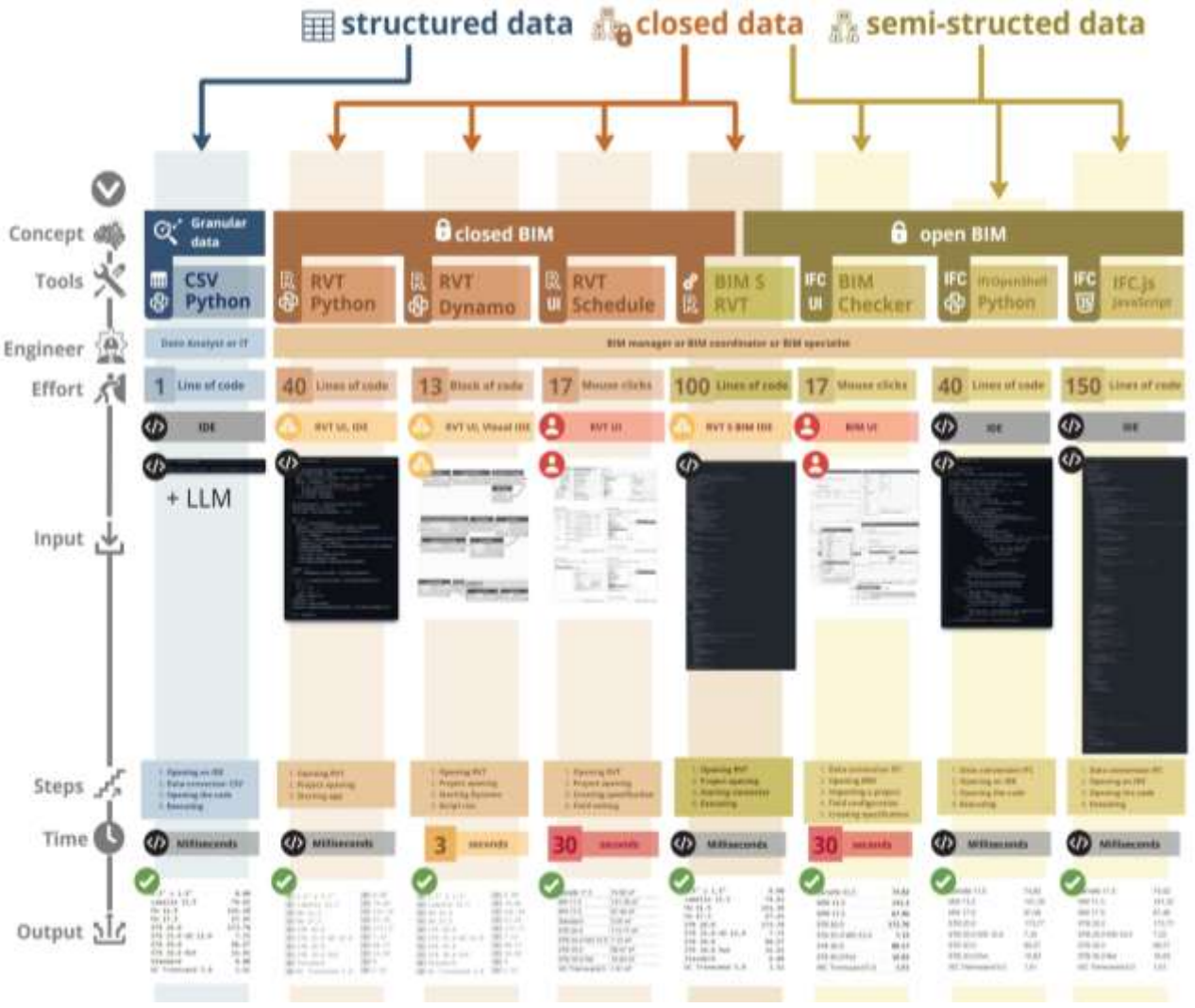


Рис. 5.2-5 Разные инструменты дают одинаковые результаты в виде атрибутивных таблиц сущностей проекта, но с разными трудозатратами.

При использовании открытых структурированных данных из CAD проектов, как упоминалось в главе "Преобразование CAD (BIM) данных в структурированную форму", процесс группировки, QTO, значительно упрощается.

Подходы, основанные на использовании открытых структурированных данных или прямом доступе к базам данных CAD-моделей, освобождены от маркетинговых ограничений, связанных с аббревиатурой BIM. Они опираются на проверенные инструменты, давно применяемые в других отраслях (Рис. 7.3-10 ETL процесс).

> Согласно исследованию McKinsey „Открытые данные: Раскрой инновации и производительность с помощью потоковой информации“ [102], проведенному в 2013 году, использование открытых данных может создать возможности для экономии от 30 до 50 миллиардов долларов в год на проектировании, инжиниринге, закупках и строительстве объектов электроэнергетики. Это означает 15-процентную экономию капитальных затрат на строительство.

Работа с открытыми структурированными (гранулированными) данными упрощает поиск и обработку информации, снижает зависимость от специализированных BIM-платформ и открывает путь к автоматизации без необходимости использовать проприетарные системы или параметрические и сложные модели данных из CAD форматов.

Автоматизация QTO с использованием LLM и структурированных данных

Перевод неструктурированных данных в структурированную форму значительно повышает эффективность различных процессов: он упрощает обработку данных (Рис. 4.1-1, Рис. 4.1-2) и ускоряет процесс валидации, делая требования ясными и прозрачными, о чем мы уже говорили в предыдущих главах. Аналогичным образом, перевод данных CAD (BIM) в структурированную открытую форму (Рис. 4.1-12, Рис. 4.1-13) облегчает процесс группировки атрибутов и процесс QTO.

Таблица атрибутов QTO имеет структурированную форму, поэтому при использовании структурированных данных CAD мы работаем с единой моделью данных (Рис. 5.2-5), что избавляет нас от необходимости заниматься конвертацией и переводу моделей данных проекта и правил группировки к единому знаменателю. Это позволяет группировать данные по одному или нескольким атрибутам всего одной строкой кода. В отличие от этого, в open BIM и closed BIM, где данные хранятся в полуструктурированных, параметрических или закрытых форматах, обработка требует десятков или даже сотен строк кода, а также использования API для взаимодействия с геометрией и атрибутивной информацией.

- Пример группировки QTO структурированного проекта по одному атрибуту. Текстовый запрос в любом LLM чате (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любой другой):

  > У меня есть CAD-проект в виде DataFrame – отфильтруй пожалуйста данные проекта так, чтобы получить элементы у которых параметр "Type" содержит только значение "Type 1" ⏎

- Ответ LLM с большой вероятностью будет в виде кода на Python с использованием Pandas:

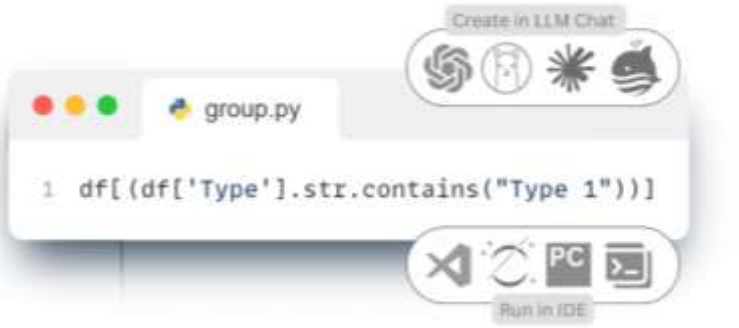


Рис. 5.2-6 Одна строка кода, написанная с помощью LLM, позволяет сгруппировать весь CAD проект по атрибуту "Тип" и получить нужную группу элементов.

Благодаря простой структуре двухмерного DataFrame нам не нужно объяснять LLM схему и модель данных, что сокращает этапы интерпретации и ускоряет создание конечных решений. Ранее для написания даже простого кода требовалось изучение языков программирования, но теперь современные языковые модели (LLM) позволяют автоматически преобразовывать логику процесса в код в работе со структурированными данными с помощью текстовых запросов.

> Автоматизация и языковые модели LLM могут полностью избавить специалистов, работающих с группировкой и обработкой данных CAD (BIM), от необходимости изучать языки программирования или инструменты BIM, предоставляя возможность решать задачи с помощью текстовых запросов.

Тот же самый запрос — группировка всех элементов проекта из категории «стены» с подсчётом объёмов по каждому типу (Рис. 5.2-5) — который в среде CAD (BIM) требует 17 нажатий в интерфейсе или написания 40 строк кода, в инструментах обработки открытых данных (например, SQL или Pandas) выглядит как простой и интуитивно понятный запрос:

- При помощи одной строки в Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

Расшифровка кода: возьми из df (DataFrame) элементы, у которых атрибут-колонку «Категория», имеет значений «OST_Walls», сгруппируй все полученные элементы по атрибуту-колонке «Type» и суммируй для полученной группы элементов атрибут «Объём».

- Группировка структурированного проекта, полученного из CAD при помощи SQL :

```
SELECT Type, SUM(Volume) AS TotalVolume
FROM elements
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- При помощи LLM запрос группировки к базе данных проекта мы можем записать в виде простого текстового обращения - промпта (Рис. 5.2-7):

> Для датафрейма проекта сгруппируй элементы по параметру 'Type', но только для элементов, у которых параметр 'Category' равен 'OST_Walls' или 'OST_Columns' и суммируй, пожалуйста, столбец параметр 'Volume' для полученной группы ⏎

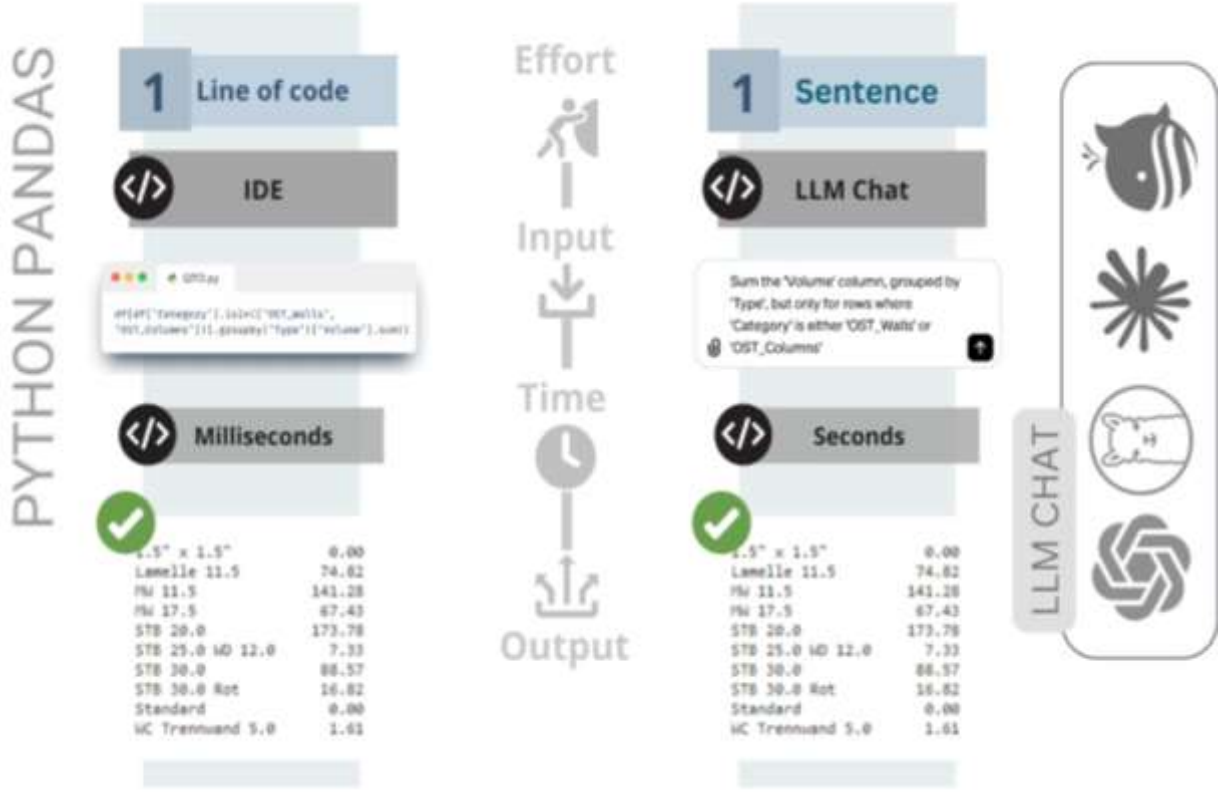


Рис. 5.2-7 Благодаря использованию SQL, Pandas и LLM автоматизация обработки данных теперь возможна с помощью нескольких строк кода и текстовых запросов.

Получение QTO из данных CAD с использованием LLM инструментов (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), кардинально меняет традиционные методы извлечения атрибутивной информации, количественных и объемных данных для отдельных объектов и их групп.

Теперь даже менеджеры проектов, специалисты по калькуляциям или логистике, не обладающие глубокими знаниями в проектировании и не имея специализированных программ CAD- (BIM-) вендоров, получив доступ к базе данных CAD могут за считанные секунды получить общий объем элементов категории стен или других объектов, просто написав или продиктовав запрос.

В текстовых запросах (Рис. 5.2-8) агент LLM модели обрабатывает обращение пользователя для применения определённой функции к одному или нескольким параметрам – колонкам таблицы. В результате чего пользователь в общении с LLM получает или новую колонку-параметр с новыми значения, или одно конкретное значение после группировки.

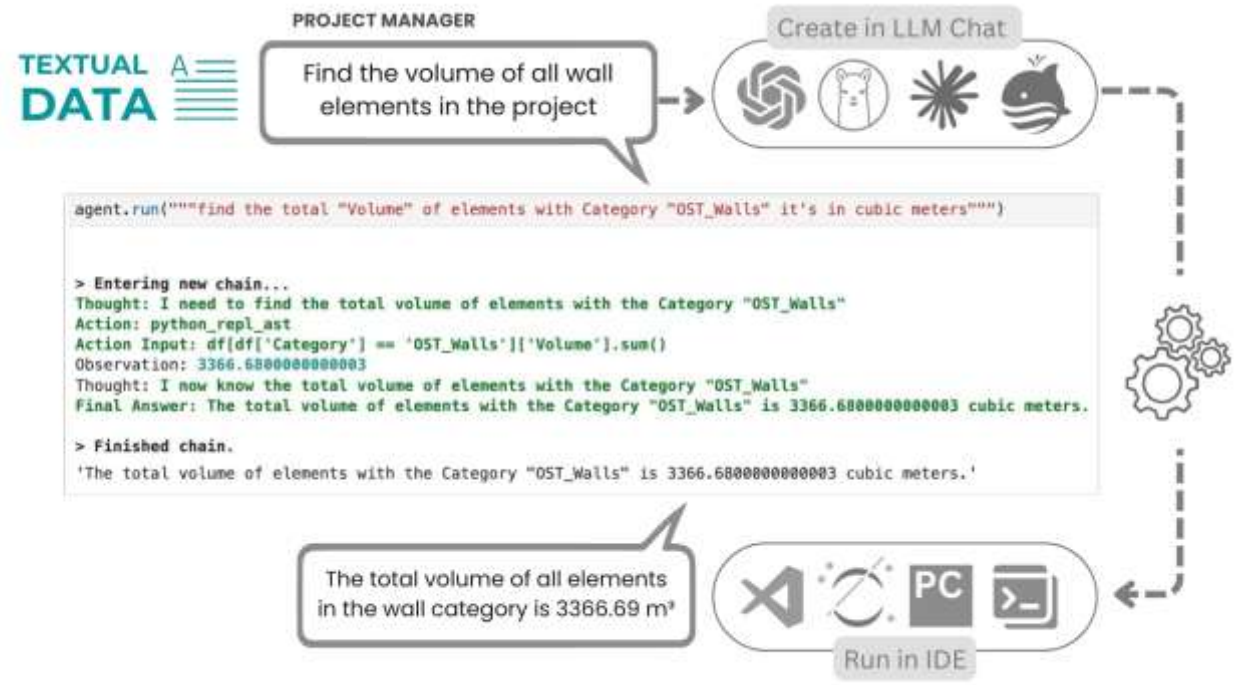


Рис. 5.2-8 LLM модель, работая со структурированными данными, понимает из контекста текстового запроса, о какой группировке и атрибутах спрашивает пользователь.

Если необходимо получить объёмные показатели только для одной группы элементов, достаточно выполнить простой QTO-запрос (Рис. 5.2-7) к данным модели CAD. Однако при расчёте бюджета или сметы для всего проекта, состоящего из многих групп элементов, часто требуется извлечение количественных характеристик для всех видов элементов (классов0, где каждая категория элементов обрабатывается отдельно — с группировкой по соответствующим атрибутам.

В практике сметчиков и оценщиков используются индивидуальные правила группировки и расчёта для разных типов объектов. Например, окна обычно группируются по этажам или зонам (параметр группировки – атрибут Level, Rooms), а стены — по материалу или типу конструкции (параметр Material, Type). Для автоматизации процесса группировки такие правила описываются заранее в

виде таблиц правил группировки. Эти таблицы действуют как конфигурационные шаблоны, определяющие, какие атрибуты следует использовать при расчётах для каждой группы элементов в проекте.

QTO расчет всего проекта с использованием правил для групп из таблицы Excel

В реальных строительных проектах нередко возникает необходимость выполнять агрегации по нескольким атрибутам одновременно внутри одной группу элементов. Например, при работе с категорией «Окна» (где атрибут Category содержит значения вроде OST_Windows или IfcWindows), элементы можно группировать не только по типу — например, по значению в поле Type Name или Type, — но и по дополнительным характеристикам, таким как уровень теплопроводности, указанный в соответствующем атрибуте. Подобная многомерная группировка позволяет получить более точные результаты для конкретной группы. Аналогично, при расчётах по категориям стен или перекрытий можно использовать произвольные комбинации атрибутов — например, материал, уровень, этаж, огнестойкость и другие параметры — в качестве фильтров или критериев группировки (Рис. 5.2-9).

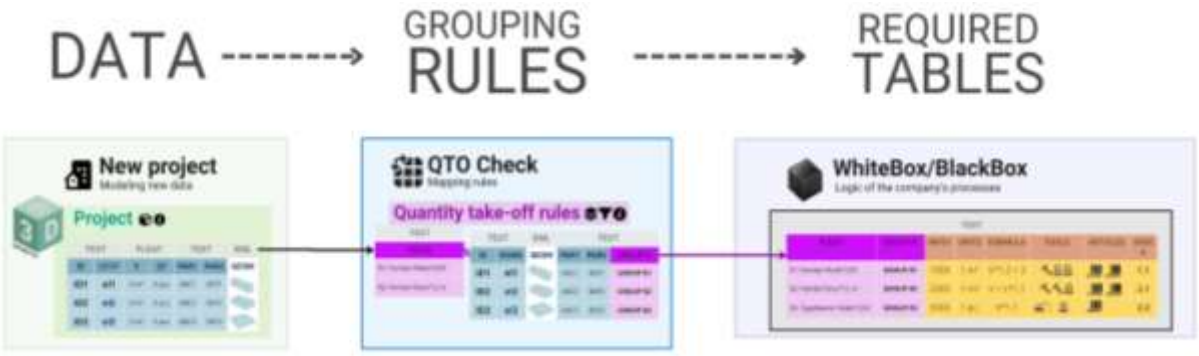


Рис. 5.2-9 Для каждой группы или категории сущностей в проекте существует своя формула группировки, состоящей из одного или нескольких критериев.

Процесс определения таких правил группировки аналогичен процессу создания требований к данным, описанному в главе "Создание требований и проверка качества данных" (Рис. 4.4-5), где мы подробно рассматривали работу с моделями данных. Такие правила группировки и вычисления обеспечивают точность и релевантность результатов для автоматического вычисления суммарных атрибутов количества или объема категории сущностей с учетом всех необходимых условий, которые должны учитываться при расчётах и калькуляциях.

- Следующий пример кода фильтрует таблицу проектов таким образом, что результирующий набор данных содержит только те сущности, в которых атрибут-колонка «Category» содержит значения «OST_Windows» или «IfcWindows» и в то же время атрибут-колонка «Type» содержит значение «Type 1»:

> У меня есть DataFrame проекта - отфильтруйте данные так, чтобы в наборе данных остались только элементы, у которых атрибут "Категория" содержит значения "OST_Windows" или "IfcWindows" и одновременно атрибут Type содержит значение "Type 1" ⏎

- Ответ LLM:

Рис. 5.2-10 Одна строка кода, похожая на формулу Excel, позволяет сгруппировать все сущности проекта по нескольким признакам.

Полученный код (Рис. 5.2-10) после перевода данных CAD в структурированных открытых форматах (Рис. 4.1-13) можно запустить в одной из популярных IDE (интегрированная среда разработки), про которые мы говорили выше, в оффлайн режиме: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с плагином PyDev, Thonny, Wing IDE, IntelliJ IDEA с плагином Python, JupyterLab или популярных онлайн-инструментах: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Чтобы получить сущности проекта в форме QTO DataFrame по категории “Окна” только с определенным значением теплопроводности, мы можем использовать следующий запрос к LLM:

> У меня есть DataFrame проекта - отфильтруйте данные так, чтобы в наборе данных остались только записи с "Категорией", содержащей значения "OST_Windows" или "IfcWindows", и одновременно столбец ThermalConductivity должен иметь значение 0,5 ⏎

- Ответ LLM:

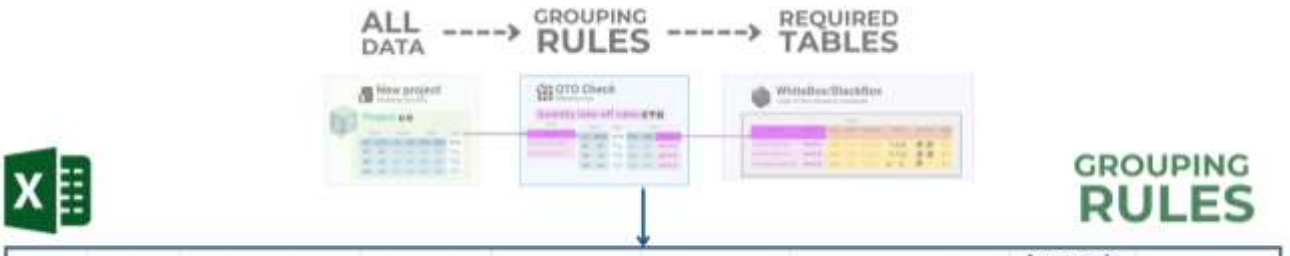
Рис. 5.2-11 Крайне простой язык запросов Pandas Python позволяет проводить QTO для любого количества проектов одновременно.

В полученном от LLM ответе (Рис. 5.2-11), используется логическое условие "&", чтобы объединить два критерия: значение теплопроводности и принадлежность к одной из двух категорий. Метод "isin" проверяет, содержится ли значение атрибута-столбца "Категория" в предоставленном списке.

В проектах с большим количеством групп элементов, с разной логикой группировки - для каждой категории сущностей проекта (например: окна, двери, перекрытия) необходимо устанавливать индивидуальные правила группировки, которые могут включать дополнительные коэффициенты или итоговые формулы расчета атрибутов. Эти формулы (Рис. 5.2-12 атрибут „formel“, например x-значение количества и y-объем группы) и коэффициенты учитывают уникальные характеристики каждой группы, например:

- добавочные % к объему материала для учета перерасхода
- фиксированное дополнительное количество материала
- корректировки, связанные с возможными рисками и погрешностями расчетов в виде формул

После того как правила фильтрации и группировки сформулированы в виде формул параметров для каждой категории элементов, их можно сохранить в виде построчной таблицы — например, в формате Excel (Рис. 5.2-12). Хранение этих правил в структурированном виде позволяет полностью автоматизировать процесс извлечения, фильтрации и группировки проектных данных. Вместо ручного написания множества отдельных запросов, система просто считывает таблицу параметров и применяет соответствующие правила к модели (общему датафрейму проекта (Рис. 4.1-13)), формируя итоговые QTO-таблицы для каждой категории элементов проекта.



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

Рис. 5.2-12 Таблица группировки атрибутов QTO устанавливает правила группировки элементов проекта, обеспечивая точное общее количество и объем для каждой категории.

Собранные правила позволят сгруппировать весь проект и выполнить все необходимые расчеты, включая корректировку атрибутов объема. В результате объемы приводятся к «реальному объему», который используется для расчетов и калькуляций, а не те что были изначально на этапе проектирования в CAD модели.

> В процессе автоматического создания объемных таблиц QTO для всего проекта приложение должно пройти по всем категориям таблицы правил группировки, взять атрибуты группировки, сгруппировать по ним все элементы проекта и агрегировать атрибут объема для этой группы, дополнительно умножив его на уточняющий фактор или коэффициент.

Попросим LLM написать для нас код для такого решения, где код должен будет загружать две таблицы - таблицу правил группировки (Рис. 5.2-12) и таблицу данных самого проекта (Рис. 4.1-13), а затем применять правила группировки, группировать элементы в соответствии с заданными правилами, вычислять агрегированные значения и сохранять результаты в новый файл Excel.

- Отправьте текстовый запрос в LLM чат:

> Мне нужен код для чтения проектных данных из файла «basic_sample_project.xlsx», а затем правил из «Grouping_rules_QTO.xlsx» по которым сгруппировать все данные по 'Parameter 1' и 'Parameter 2', агрегировать 'Aggregate Parameter', фильтровать по 'Expression2', выполнить вычисления из 'Formel1' и сохранить QTO таблицу в 'QTQ_table2.xlsx' ⏎

Ответ LLM:

Рис. 5.2-13 Код Python Pandas обрабатывает проектные данные из CAD проекта, группируя их в соответствии с заданными в Excel правилами.

Конечным результатом выполнения кода (Рис. 5.2-13) будет таблица групп-сущностей, которая содержит не просто обобщенные атрибуты объема из исходной CAD- (BIM-) модели, но также и новый атрибут реального объема, учитывающий все требования для корректного создания калькуляций и смет (Пример Рис. 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

Рис. 5.2-14 Атрибут "После вычисления" добавляется в сводную таблицу после выполнения кода, который автоматически вычислит реальный объем.

Полученный код (Рис. 5.2-13) можно запустить в одной из популярных IDE (про которые мы говорили выше) и применить такой код к любому количеству уже существующих или вновь поступающих проектов (RVT, IFC, DWG, NWS, DGN и др.), будь то несколько проектов или, возможно, сотни проектов в разных форматах, приведённых в структурированную форму (Рис. 5.2-15).

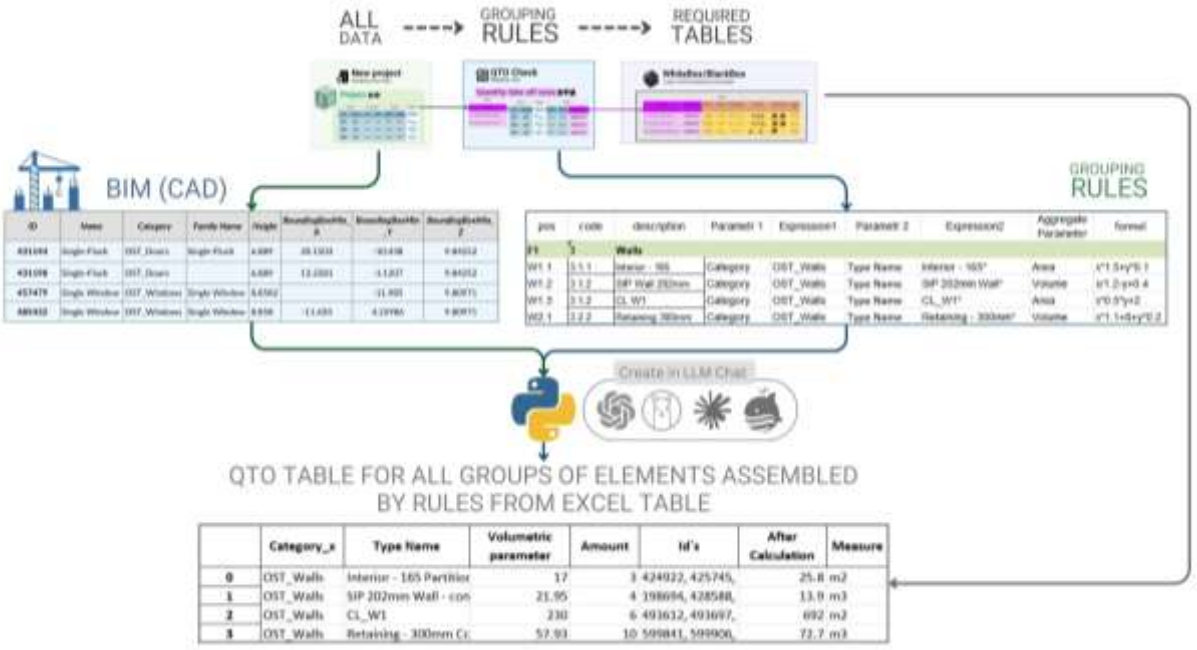


Рис. 5.2-15 Процесс автоматической группировки строительных данных связывает данные BIM (CAD) с таблицами QTO через правила из электронной таблицы Excel.

Настроенный и параметризированный процесс сбора объемных данных (Рис. 5.2-15) позволяет полностью автоматизировать сбор данных о количественных атрибутах и объемах элементов проекта для дальнейшей работы с ними, включая оценку стоимости, логистики, графиков работ и расчета углеродного следа и других аналитических задач.

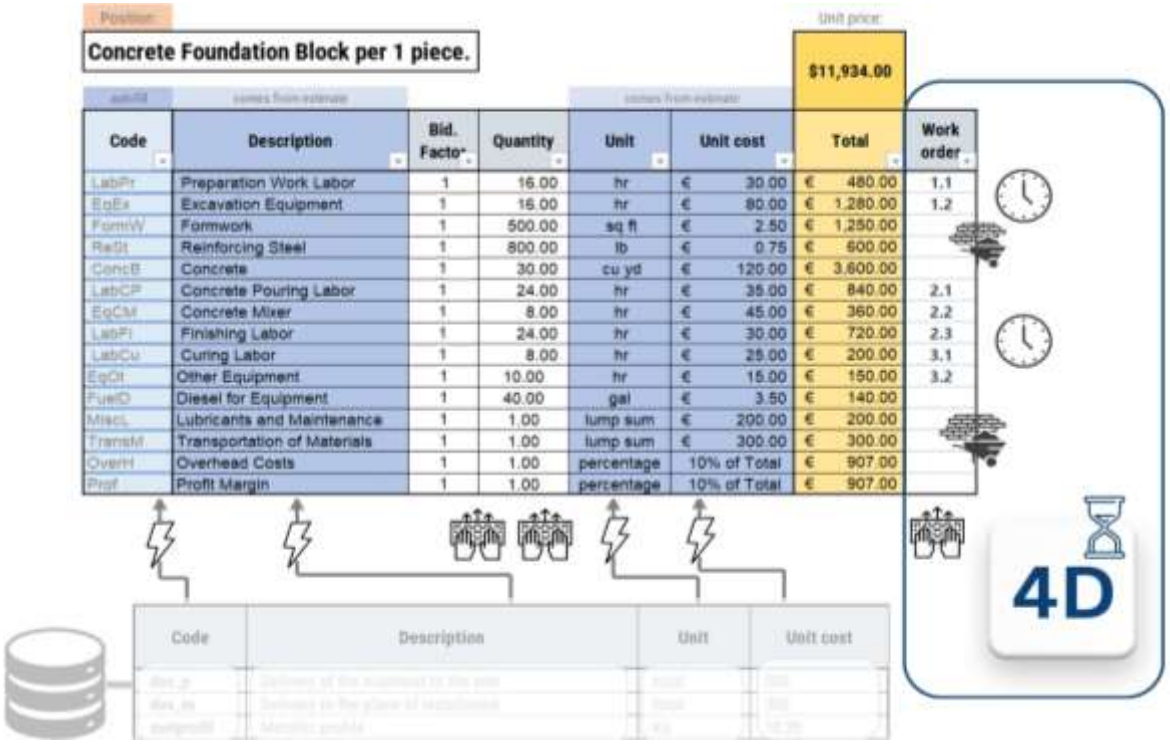
Изучив инструменты, позволяющие легко организовывать и группировать группы элементов проекта по определенным признакам, мы теперь готовы интегрировать сгруппированные и отфильтрованные проекта с различными расчётами и бизнес-сценариями компании.

ГЛАВА 5.3.
4D, 6D-8D И РАСЧЕТ ВЫБРОСОВ УГЛЕКИСЛОГО ГАЗА CO_2

4D-модель: интеграция времени в строительные сметы

Помимо расчёта стоимости, одним из ключевых направлений применения проектных данных в строительстве является определение временных параметров — как для отдельных строительных операций, так и для всего проекта в целом. Для автоматизированного расчёта сроков и формирования календарного графика выполнения работ в качестве основы часто используется ресурсный метод оценки и связанная с ним база расчётных данных, подробно рассмотренная в предыдущей главе «Расчёты и сметы для строительных проектов».

В рамках ресурсного подхода учитываются не только затраты на материалы, но и временные ресурсы. При составлении калькуляций каждому процессу может быть назначен атрибут порядка выполнения работ (Рис. 5.3-1 – параметр «Work order»), а также указано количество времени и стоимость, связанная с исполнением данного процесса. Эти параметры особенно важны для описания операций, не имеющих фиксированной рыночной цены и не подлежащих прямой закупке — таких как использование строительной техники, занятость рабочих или логистические процессы (которые обычно выражаются обычно в часах). В подобных случаях стоимость определяется не отделом закупок, а непосредственно компанией-исполнителем на основании внутренних нормативов или производственных ставок (Рис. 5.3-1).



Code	Description	Bid. Facto*	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	

Рис. 5.3-1 Расчеты работ в ресурсном методе оценок включают в себя временные затраты рабочего времени.

Таким образом, в расчёты на уровне калькуляций включаются не только затраты на топливо и материалы (закупочная стоимость), но и время работы машинистов, техники и вспомогательных рабочих на стройплощадке. В приведенном примере (Рис. 5.3-1) таблица затрат представляет собой расчёт стоимости установки фундаментного блока, включая составные этапы работы, такие как подготовка, установка каркаса и заливка бетона, а также необходимые материалы и трудозатраты. При этом отдельные операции, например подготовительные работы, могут не иметь материальных затрат, но содержать значительные временные трудозатраты, выраженные в человеко-часах.

Для планирования последовательности работ (для графика работ) на стройплощадке в таблицу калькуляции вручную например добавляется атрибут «Work order» (Рис. 5.3-1). Он указывается в дополнительном столбце только для элементов, у которых единица измерения выражается во времени (час, день). Этот атрибут дополняет код работы, описание, количество, единицу измерения (параметр «Unit») и затраты. Числовая последовательность (параметр «Work order») работ позволяет установить порядок выполнения задач на стройплощадке и использовать его при составлении расписания.

График строительства и его автоматизация на основе данных калькуляции

График строительства — это визуальное представление плана выполнения работ и процессов, которые должны быть выполнены в рамках реализации проекта. Он создается на основе подробных ресурсных расчетов (Рис. 5.3-1), где каждая задача-работа расписана, помимо стоимости ресурсов, по времени и последовательности.

В отличие от усреднённых подходов, где расчёты времени строятся на основе типового количества часов на установку материалов или оборудования, в ресурсном методе планирование основано на фактических данных, заложенных в калькуляции. Каждая позиция сметы, связанная с трудозатратами, опирается на применяемый календарь, в котором учитываются реальные условия использования ресурсов в течение рабочего периода. Корректировка продуктивных часов через коэффициенты на уровне калькуляций (Рис. 5.3-1 параметр „Bid. Factor“), позволяет учитывать различия в производительности и сезонные особенности, влияющие на сроки выполнения работ.

Чтобы определить даты начала и окончания процесса для графика строительства на диаграмме Ганта, мы берем значения атрибута объема времени для каждого элемента из калькуляции фундаментных блоков и умножаем их на количество блоков (в данном случае количество бетонных фундаментных блоков). Этот расчет дает продолжительность каждой задачи. Затем мы наносим эти длительности на временную шкалу, начиная с даты начала проекта, чтобы построить график, и в результате получаем визуальное представление, показывающее, когда каждая задача должна начаться и закончиться. Параметр «Work order» у процессов дополнительно позволяет нам понять происходит процесс работы параллельно («Work order» например 1.1-1.1) или последовательно (1.1-1.2)

> **Диаграмма Ганта** — это графический инструмент для планирования и управления проектами, представляющий задачи в виде горизонтальных полос на временной шкале. Каждая полоса отображает продолжительность выполнения задачи, её начало и окончание.

График работ, или диаграмма Ганта, помогает руководителям проекта и рабочим четко понимать, когда и в какой последовательности должны выполняться различные этапы строительства, обеспечивая эффективное использование ресурсов и соблюдение сроков.

Представим календарный план работ по установке трех бетонных фундаментных блоков с использованием расчетов из таблицы выше. Используя таблицу затрат (Рис. 5.3-1) из примера выше, попросим LLM запланировать установку 3 элементов фундаментного блока, например на первое мая 2024 года.

Чтобы отправить калькуляцию в LLM, мы можем загрузить таблицу калькуляции в формате XLSX или просто вставить скриншот картинки калькуляции в формате JPEG напрямую в чат LLM (Рис. 5.3-1). LLM самостоятельно найдет библиотеку для визуализации изображения таблицы и, умножив временные атрибуты работ из таблицы на их объем, сложит все данные в график.

- Отправьте текстовый запрос в LLM:

> Создай график работ в виде диаграммы Ганта с использованием значений из таблицы, относящихся ко времени (Рис. 5.3-1 как JPEG в приложении) для установки 3 фундаментных блоков. Работы по каждому блоку выполняются последовательно. Начало работ установить с 01/05/2024 ⏎

- Ответ LLM:

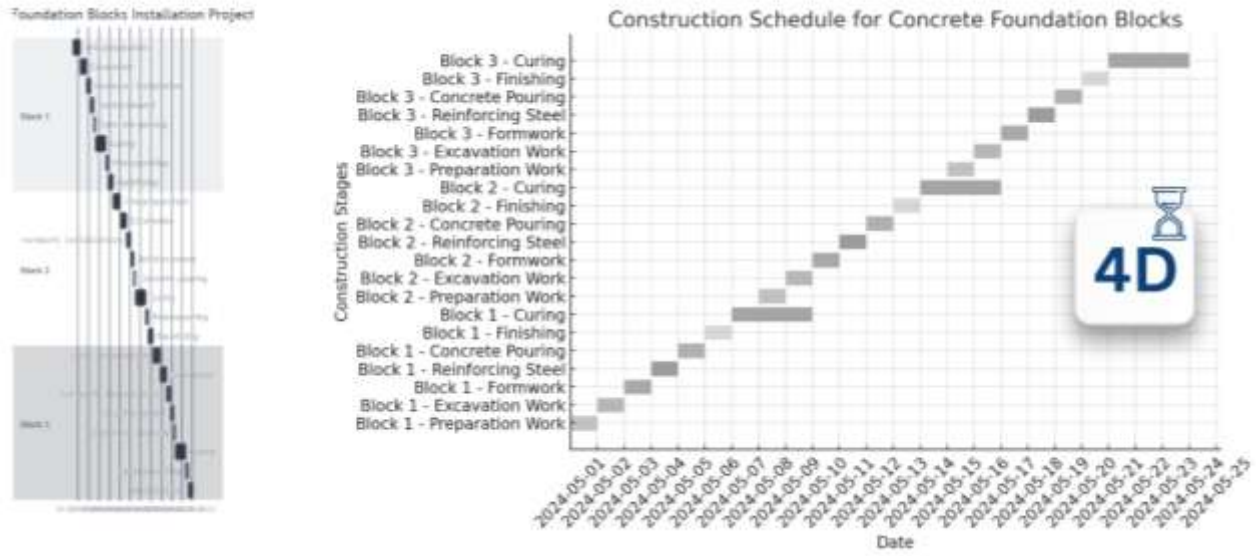


Рис. 5.3-2 Автоматически созданный несколькими LLM диаграмма Ганта показывает этапы строительства трех бетонных блоков, согласно условиям из промпта.

Полученный график (Рис. 5.3-2) представляет собой временную диаграмму, в которой каждая горизонтальная полоса соответствует определенному этапу выполнения работ над фундаментным блоком и демонстрирует последовательность операций (параметр «Work order»), таких как подготовка, земляные работы, установка опалубки, армирование, заливка бетона и финишная обработка,

т. е. те процессы, которые в калькуляциях имеют заполненные временные параметры и последовательность.

Подобный график (Рис. 5.3-2) не учитывает ограничения, связанные с рабочими днями, сменами или нормами рабочего времени, а предназначен исключительно для концептуальной визуализации процесса. Точный график, который будет отражать параллельность работ, можно дополнять соответствующими промтами или дополнительными инструкциями внутри чата.

Используя одну калькуляцию затрат (Рис. 5.3-1), благодаря атрибутам объемов из 3D-геометрии, можно автоматически оценивать как стоимость проекта через автоматизированные сметы так и одновременно с этим рассчитывать временные характеристики групп в виде таблиц или графиков для различных вариантов проекта (Рис. 5.3-3).

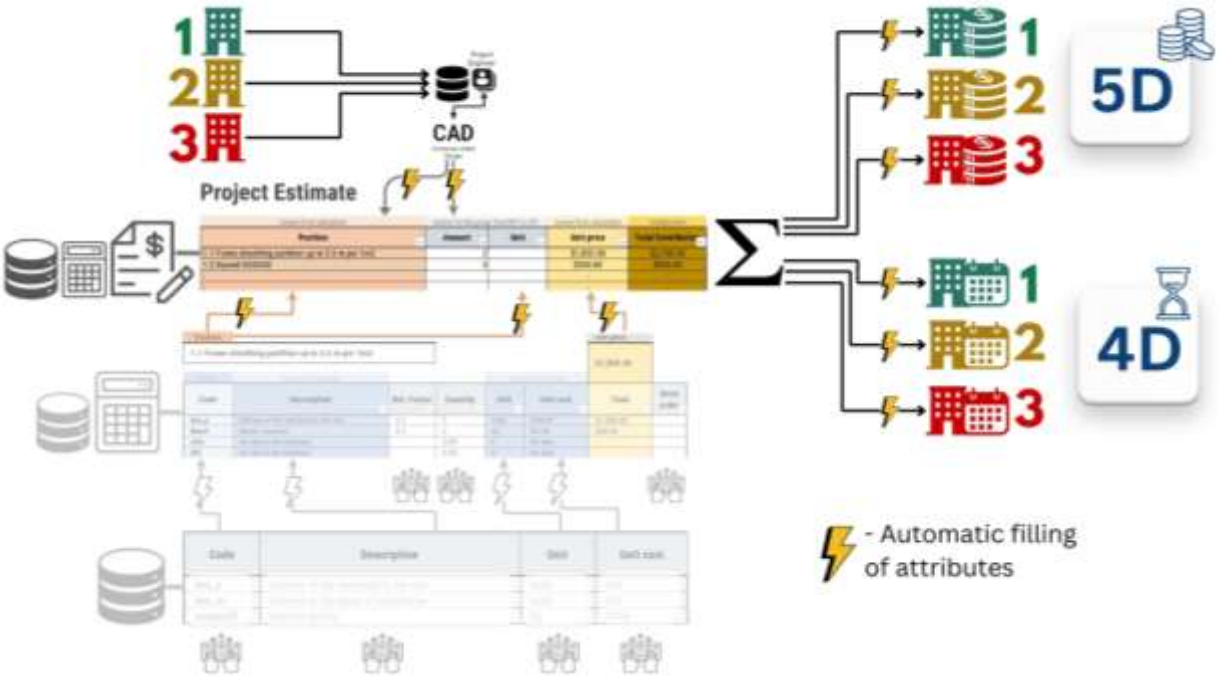


Рис. 5.3-3 Автоматический расчет, позволяет мгновенно и автоматически прогнозировать затраты и время для различных вариантов проекта.

Современные модульные ERP-системы (Рис. 5.4-4), подгружая данные из CAD моделей используют подобные автоматизированные методы расчета времени, которые существенно сокращают процесс принятия решений. Это позволяет мгновенно и с учётом реальных цен точно планировать рабочие графики и рассчитывать общее время, необходимое для выполнения всех задач при реализации проекта.

Расширенные атрибутивные слои 6D-8D: от энергоэффективности до обеспечения безопасности

6D, 7D и 8D — это расширенные уровни информационного моделирования, каждый из которых вносит дополнительные слои атрибутов в комплексную информационную модель проекта, основой

которой являются атрибуты 3D-модели с их количеством и объемом. Каждый дополнительный уровень вносит специфические параметры, которые необходимы для последующей группировки или последующей идентификации в других системах, таких, как например системы управления недвижимостью (PMS), автоматизированного управления объектами (CAFM), управления строительными проектами (CPM) и систем управления безопасностью (SMS).

Рис. 5.3-4 Атрибуты 6D, 7D и 8D в информационной модели данных расширяют рассмотрение различных аспектов проекта, от энергоэффективности до безопасности.

- В **6D** в дополнение к базе данных проекта (или датафрейма (Рис. 4.1-13)) с геометрическими и объёмными атрибутами элементов добавляется информация (атрибуты-колонки) об экологической устойчивости. Это включает информацию, связанную с энергоэффективностью, углеродным следом, возможностью вторичной переработки материалов и применением экологически безопасных технологий. Эти данные позволяют оценивать влияние проекта на окружающую среду, оптимизировать проектные решения и достигать целей устойчивого развития (ESG).
- **7D** атрибуты дополняют атрибутами, необходимыми для управления эксплуатацией здания. Это данные о графиках технического обслуживания, сроках службы компонентов, технической документации и истории ремонта. Такой набор информации обеспечивает возможность интеграции модели с системами эксплуатации (CAFM, AMS), позволяет эффективно планировать обслуживание, замену оборудования и обеспечивает поддержку на всем жизненном цикле объекта.
- **8D** дополнительный атрибутивный слой, - включает информацию, связанную с безопасностью — как на этапе строительства, так и при последующей эксплуатации. В модель вносятся меры по обеспечению безопасности персонала, инструкции по действиям в аварийных ситуациях, требования к системам эвакуации и противопожарной защите. Интеграция этих данных в цифровую модель помогает заранее учитывать риски и разрабатывать архитектурные, инженерные и организационные решения с учётом требований охраны труда и безопасности.

В структурированной табличной форме слои от 4D до 8D представляют собой дополнительные атрибуты в виде столбцов с заполненными значениями (Рис. 5.3-5), добавленные к уже заполненным атрибутам 3D-модели, таким как название, категория, тип и объемные характеристики. Значения в атрибутивных слоях 6D, 7D и 8D содержат дополнительные текстовые и цифровые данные, такие как процент переработки, углеродный след, гарантийный срок, цикл замены, дата установки, протоколы безопасности и т. д.

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

Рис. 5.3-5 6D-8D добавляют атрибутивные слои к информационной модели данных, которая уже содержит геометрические и объемные атрибуты из 3D-модели.

Для нашего нового окна (Рис. 4.4-1) элемент с идентификатором W-NEW (Рис. 5.3-5) может иметь следующие 3D-8D атрибуты:

3D-атрибуты – геометрическая информация, полученная из CAD систем:

- "Имя типа" - элемент "Окно"
- "Ширина" - 120 см
- Дополнительно можно добавить точки "Bounding Box" элемента или его "геометрию BREP / MESH" как отдельный атрибут

Атрибуты 6D - экологическая устойчивость:

- Показатель "перерабатываемости" - 90%
- "Углеродный след" - 1622 кг CO_2

Атрибуты 7D - данные об управлении объектом:

- "Гарантийный срок" - 8 лет
- "Цикл замены" - 20 лет
- "Техническое обслуживание" - требуется ежегодно

Атрибуты 8D - обеспечение безопасного использования и эксплуатации зданий:

- Окно "Установлено" - компанией "XYZ Windows"
- "Стандарт безопасности" - соответствует ISO 45001

Все записанные в базу данных или датасет (Рис. 5.3-5) параметры нужны специалистам в различных отделах для группировки, поиска или расчётов. Подобное многомерное описание объектов проекта на основе атрибутов позволяет получить полное представление об их жизненном цикле, эксплуатационных требованиях и многих других аспектах, необходимых при проектировании, строительстве и эксплуатации проекта.

Оценка CO_2 и расчет выбросов углекислого газа в строительных проектах

Наряду с темой устойчивости строительных проектов на стадии 6D (Рис. 5.3-5), в современном строительстве особое внимание уделяется экологической устойчивости проектов, где одним из ключевых аспектов становится оценка и минимизация выбросов углекислого газа CO_2, которые происходят на этапах жизненного цикла проекта (например при производстве и монтаже).

> Оценка и расчет выбросов углерода строительных материалов — это процесс, в ходе которого общие выбросы углерода определяются путем умножения объемных атрибутов элементов или группы элементов, используемых в проекте, на подходящий коэффициент выбросов углерода для данной категории.

Учет выбросов углерода при оценке строительных проектов, как части более широких ESG-критериев (экологических, социальных и управленческих) добавляет новый уровень к комплексному анализу. Это особенно важно для заказчика-инвестора при получении соответствующего сертификата, такого как LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) или DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Получение одного из этих сертификатов может существенно повысить рыночную привлекательность объекта, упростить ввод в эксплуатацию и обеспечить соответствие требованиям арендаторов, ориентированных на устойчивость (ESG). В зависимости от требований проекта могут также использоваться HQE (Haute Qualité Environnementale, французский стандарт экологического строительства), WELL (WELL Building Standard, ориентирован на здоровье и комфорт пользователей) и GRESB (Global Real Estate Sustainability Benchmark, международный рейтинг устойчивости недвижимости).

> Экологическое, социальное и управленческое **ESG** (environ-mental, social and governance) — это широкий набор принципов, которые могут использоваться для оценки корпоративного управления, социального и экологического воздействия бизнеса как внутри компании, так и за ее пределами.

ESG, первоначально разработанный в начале 2000-х годов финансовыми фондами для предоставления инвесторам информации о широких критериях экологического, социального и управленческого характера, превратился в ключевой показатель для оценки как компаний, так и проектов, включая строительные. Согласно исследованиям крупнейших консалтинговых компаний, учет экологических, социальных и управленческих факторов (ESG) становится неотъемлемой частью строительной отрасли.

> По данным EY (2023) «Путь углеродной нейтральности», компании, активно внедряющие ESG-принципы, не только снижают долгосрочные риски, но и повышают эффективность своих бизнес-моделей, что особенно важно в условиях глобальной трансформации рынков [103]. В отчете PwC «Информированность о ESG» отмечается, что уровень осведомленности компаний о важности ESG-факторов варьируется от 67% до 97%, при этом большинство организаций считают эти тенденции ключевыми для устойчивого развития в будущем [104] и что бизнес по большей части отмечает значительное давление со стороны заинтересованных сторон по интеграции ESG-принципов.

Таким образом, интеграция ESG-принципов в строительные проекты не только способствует получению международных сертификатов устойчивости, таких как LEED, BREEAM, DGNB, но и обеспечивает долгосрочную устойчивость и конкурентоспособность компаний в отрасли.

Одним из наиболее значимых факторов, влияющих на общий углеродный след строительного проекта, являются этапы производства и логистики строительных материалов и элементов. Материалы, используемые на объекте, часто оказывают решающее влияние на совокупные выбросы CO_2, особенно на ранних стадиях жизненного цикла проекта — от добычи сырья до доставки на стройплощадку.

Для расчёта выбросов по категориям или типам строительных элементов требуется использование справочных коэффициентов углеродных выбросов, отражающих количество CO_2, образующегося в результате производства различных материалов. К числу таких материалов относятся бетон, кирпич, переработанная сталь, алюминий и другие. Эти значения, как правило, извлекаются из авторитетных источников и международных баз данных, таких как UK ICE 2015 (Inventory of Carbon and Energy) и US EPA 2006 (U.S. Environmental Protection Agency) [105]. В следующей таблице (Рис. 5.3-6) приведены базовые коэффициенты выбросов для ряда распространённых строительных материалов. Для каждого из них указаны два ключевых параметра: удельные выбросы CO_2 (в килограммах на килограмм материала) и коэффициенты пересчёта объёма в массу (в килограммах на кубический метр), необходимые для интеграции расчётов в проектную модель и привязки к QTO группировке данных.

Carbon Emitted in Production

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)
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

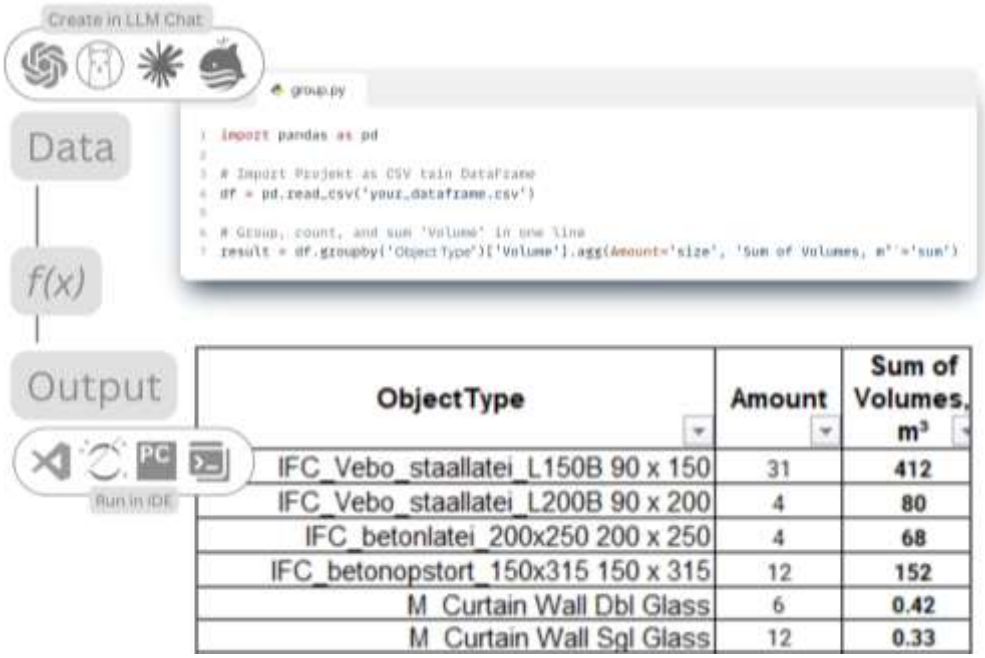
Рис. 5.3-6 Количество углерода, выделяемого при производстве различных строительных материалов, согласно базе данных UK ICE и US EPA.

Чтобы рассчитать общий объем выбросов CO_2 по проекту, как и в случае с калькуляциями 4D и 5D необходимо определить объем атрибутов каждой группы объектов. Это можно сделать с помощью инструментов количественного анализа (QTO), получив объемы атрибутов в кубических метрах, как подробно рассматривали в разделе посвященным Quantity take-off. Далее полученные объемы умножаются на соответствующие коэффициенты для атрибута "CO_2 технологические выбросы" каждой группы материалов.

- Давайте автоматически извлечем таблицу объемов по типам элементов из CAD (BIM) проекта, сгруппировав все данные проекта, как уже было сделано в предыдущих главах. Для выполнения этой задачи обратимся к LLM.

> Пожалуйста, сгруппируй таблицу DataFrame из CAD (BIM) проекта по параметру столбца "Object Name" (или „Type“) и покажи количество элементов в каждой группе, а также суммируйте параметр "Volume" для всех элементов в типе. ⏎

Ответ LLM:



ObjectType	Amount	Sum of Volumes, m^3
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

Рис. 5.3-7 Сгенерированный код в LLM сгруппировал для нас сущности проекта по типу (ObjectType) с суммированным атрибутом „Volume“.

Чтобы автоматизировать расчет суммарных выбросов CO_2 для всего проекта, достаточно настроить автоматическое сопоставление данных в таблице или вручную связать типы элементов (Рис. 5.3-7) с соответствующими типами материалов (Рис. 5.3-6) из таблицы коэффициентов выбросов. Готовую таблицу с коэффициентами выбросов и формулами, а также код для получения объемов из форматов CAD (BIM) и автоматизации определения CO_2 можно найти на GitHub по запросу "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction" [106].

Таким образом, интеграция данных после группировки QTO элементов из базы данных CAD позволяет автоматически рассчитывать выбросы углекислого газа (Рис. 5.3-8) для различных вариантов проектирования. Это дает возможность анализировать влияние разных материалов в разных вариантах и выбирать только те решения, которые соответствуют требованиям заказчика по уровню выбросов CO_2 для получения того или иного сертификата при сдаче здания в эксплуатацию.

> Оценка выбросов CO_2 путем умножения коэффициентов на объемы сгруппированных элементов проекта - типичный пример задачи в процессе получения строительной компанией рейтинга ESG (например сертификации LEED) для объекта.

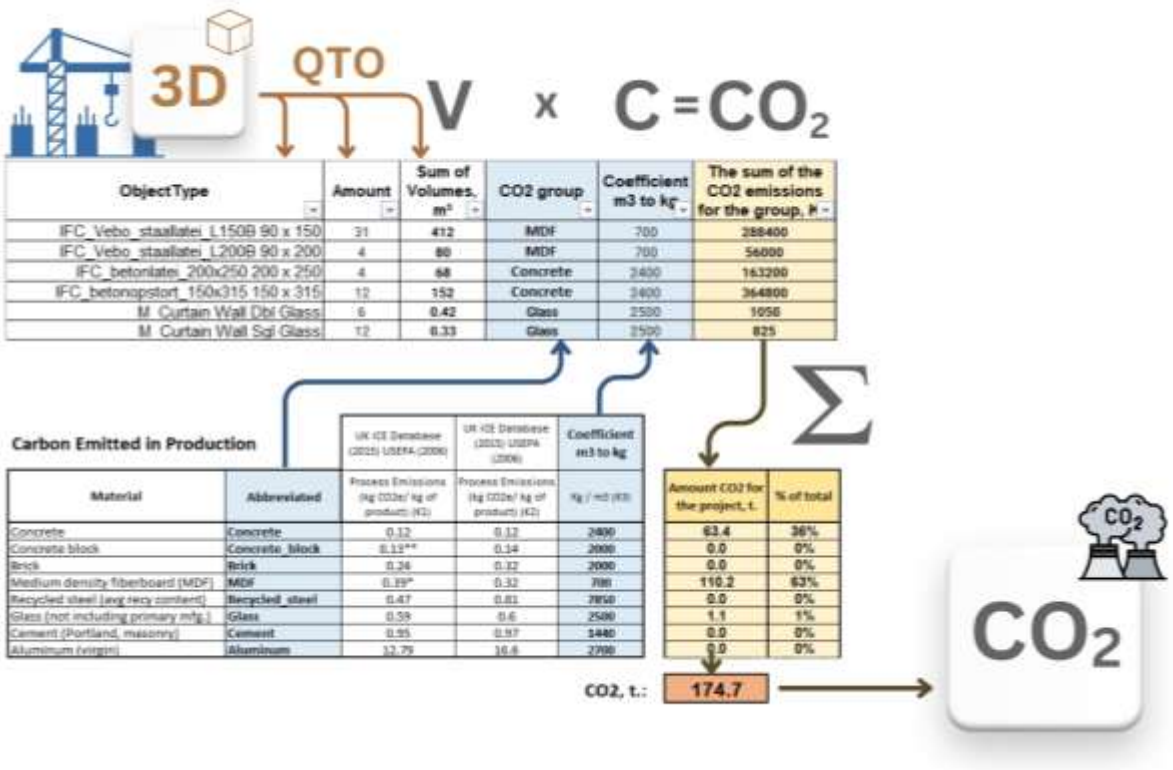


Рис. 5.3-8 Интеграция QTO групп из баз данных CAD обеспечивает точность и автоматизацию при получении оценок конечных объемов выбросов CO_2.

Аналогичным образом, определяя объемы групп элементов, мы можем выполнять расчеты для контроля и логистики материалов, мониторинга и управления качеством, моделирования и анализа энергопотребления и множества других задач для получения нового атрибутивного статуса (параметра в таблице) как отдельных групп элементов, так и всего проекта в целом.

Если количество таких расчетных процессов в компании начинает расти, встает вопрос о необходимости автоматизации подобных расчетов и имплементации результатов расчётов в процессы компании и системы по управлению данными.

Из-за сложности комплексного решения средние и крупные компании, работающие в строительной отрасли, передают подобную автоматизацию на аутсорсинг компаниям, занимающимся разработкой ERP (или PMIS) систем. Компании-разработчики создают для крупных клиентов единую комплексную модульную систему, для управления множеством различных информационных слоев, включая расчеты материалов и ресурсов.

ГЛАВА 5.4.
СТРОИТЕЛЬНЫЕ ERP И PMIS СИСТЕМЫ

Строительные ERP-системы на примере расчетов и смет

Модульные ERP системы объединяют различные атрибутивные (информационные) слои и потоки данных в единую комплексную систему, позволяя руководителям проектов синхронизировано управлять ресурсами, финансами, логистикой и другими аспектами проекта в рамках одной платформы. Строительная ERP-система выступает в роли "мозга" строительных проектов, упрощая повторяющиеся процессы за счет автоматизации, обеспечивая прозрачность и контроль на протяжении всего строительного процесса.

> **Строительные ERP**-системы (Enterprise Resource Planning) — это комплексные программные решения, предназначенные для управления и оптимизации различных аспектов строительного процесса. В основе строительных ERP-систем лежат модули для управления расчетами затрат и составления графиков работ, что делает их важным инструментом для эффективного планирования ресурсов.

Модули ERP-системы позволяют пользователям вводить, обрабатывать и анализировать данные, структурированно охватывая различные аспекты проекта, которые могут включать учет материальных и трудовых затрат, использование оборудования, управление логистикой, кадровыми ресурсами, контактами и другими видами строительной деятельности.

Одним из функциональных блоков системы выступает модуль автоматизации бизнес-логики — BlackBox/WhiteBox, играющий роль центра управления процессами.

BlackBox/WhiteBox позволяет специалистам, использующим ERP-систему, через права доступа гибко управлять различными аспектами бизнеса, которые уже были предварительно сконфигурированы другими пользователями или администраторами. В контексте ERP-систем термины *BlackBox* и *WhiteBox* обозначают уровни прозрачности и контролируемости внутренней логики системы:

- **BlackBox** («чёрный ящик») — пользователь взаимодействует с системой через интерфейс, не имея доступа к внутренней логике выполнения процессов. Система самостоятельно производит вычисления, на основе заранее заданных правил, скрытых от конечного пользователя. Он вводит данные и получает результат, не зная, какие атрибуты или коэффициенты были использованы внутри.
- **WhiteBox** («белый ящик») — логика процессов доступна для просмотра, настройки и модификации. Продвинутые пользователи, администраторы или интеграторы могут вручную задать алгоритмы обработки данных, правила расчётов и сценарии взаимодействия между сущностями проекта.

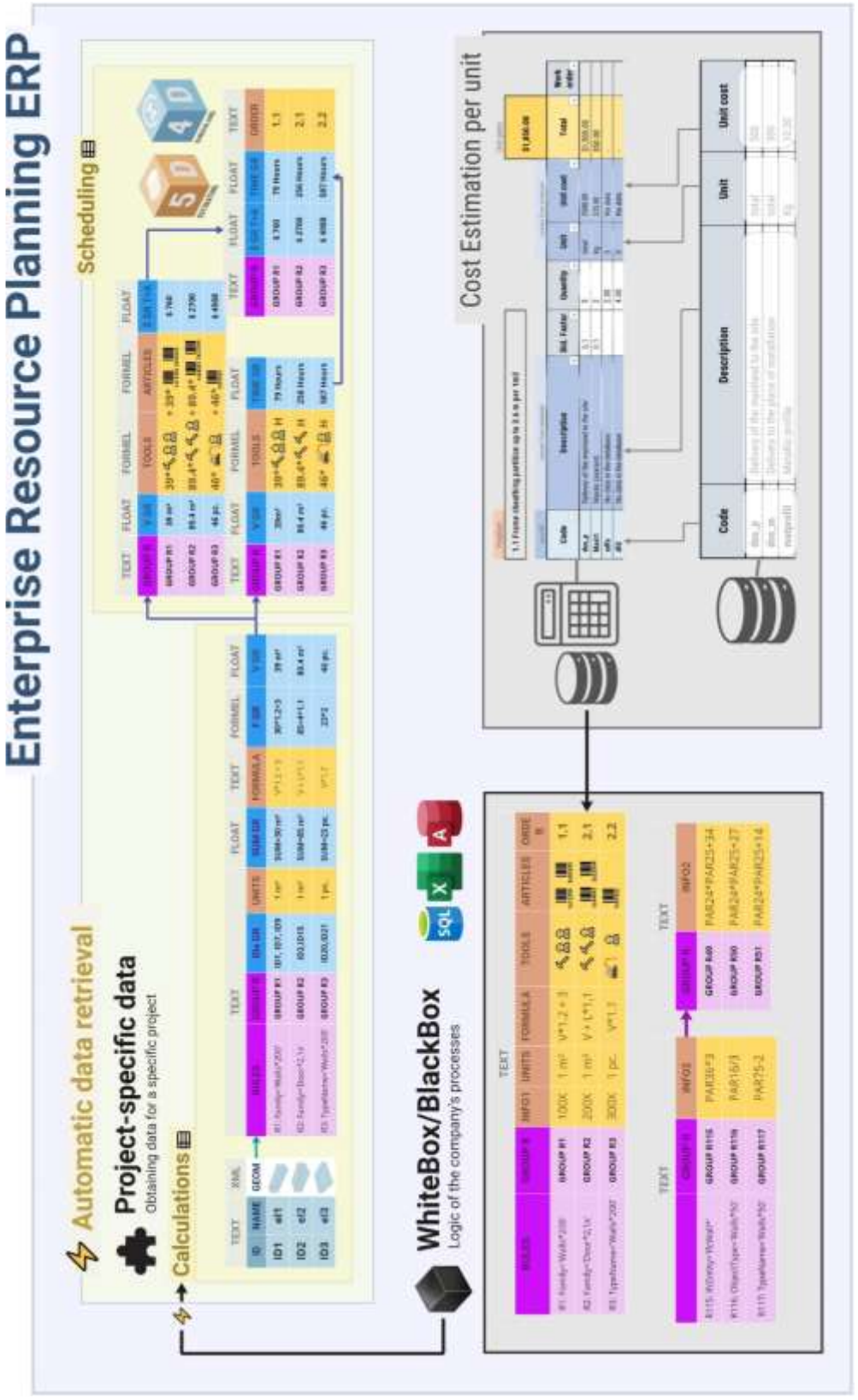


Рис. 5.4-1 Архитектура строительной ERP-системы, для получения смет и графиков работ при ручном заполнении атрибутов объема.

Примером может служить ситуация, где опытный пользователь или администратор задаёт правило: какие атрибуты в смете должны быть умножены между собой или сгруппированы по определённому признаку, а также куда должен быть записан итоговый результат. В дальнейшем, менее подготовленные специалисты, например инженеры-сметчики, просто загружают новые данные в ERP через пользовательский интерфейс — и получают готовые сметы, графики или спецификации без необходимости писать код или разбираться в технических деталях логики.

> В предыдущих главах были рассмотрены модули расчётов и логики в контексте взаимодействия с LLM. В среде ERP-систем подобные вычисления и преобразования происходят внутри модулей, скрытых за интерфейсом из кнопок и форм.

В следующем примере (Рис. 5.4-1) администратор ERP-системы в BlackBox/WhiteBox модуле задал правила сопоставления атрибутов сущностей из смет с атрибутами для группировки QTO. Благодаря этому настроенному (менеджером или администратором) модулю BlackBox/WhiteBox пользователь (сметчик или инженер), добавив вручную атрибут количества или объема через пользовательский интерфейс ERP, автоматически получает готовые сметы и рабочие графики. Таким образом процессы расчетов и формирования смет, рассмотренные в предыдущих главах при помощи кода, внутри ERP, превращаясь в полу автоматизированный конвейер.

Подключение этого полу-автоматизированного процесса к объемным атрибутам из CAD (BIM) моделей (Рис. 4.1-13), через, например, загрузку проекта CAD, в преднастроенный для этого модуль ERP, превращает поток данных в синхронизированный механизм, способный автономно и мгновенно обновлять стоимость отдельных групп элементов или всего проекта в ответ на любые изменения в нём на этапе проектирования, при загрузки CAD модели в ERP.

Чтобы создать автоматизированный поток данных (Рис. 5.4-2) между системами CAD (BIM) и ERP, необходимо структурированно определить основные процессы и требования к данным из баз данных CAD (BIM) моделей, о чем мы уже говорили в главе выше "Требования и обеспечение качества данных". Этот процесс в ERP делится на похожие этапы:

- **Создание правил проверки (1),** которые играют важную роль в обеспечении точности данных, поступающих в ERP-систему. Правила проверки служат в качестве фильтров, которые проверяют сущности и их атрибуты, пропуская в систему только те элементы, которые прошли требования. Подробнее о проверке и валидации в главе "Создание требований и проверка качества данных".
- Затем внутри ERP происходит **процесс верификации (2),** который подтверждает, что все элементы-сущности проекта с их атрибутами и значениями были созданы правильно и готовы к следующим этапам обработки.
- Если возникают проблемы с неполными атрибутивными данными, **создается отчет (3),** и проект вместе с инструкциями по исправлению отправляется на доработку до готовности к следующей итерации.
- После того, как данные проекта были подтверждены и проверены, они используются в другом модуле ERP **(4) для создания таблиц Quantity Take-Off (QTO),** которые по ранее сформированным правилам (WhiteBox/BlackBox) создают атрибуты количества для групп сущностей, материалов и ресурсов.

- Сгруппированные данные по правилам сопоставления или QTO автоматически **интегрируются с расчетами (например, затрат и времени) (5).**
- На последнем этапе ERP-система, пользователь, перемножая атрибуты объема из таблицы QTO с атрибутами таблиц процессов (например, сметных статей), **автоматически генерирует результаты расчетов (6)** (например, сметы затрат, графики работ или выбросы CO_2) для каждой группы сущностей и для проекта в целом.

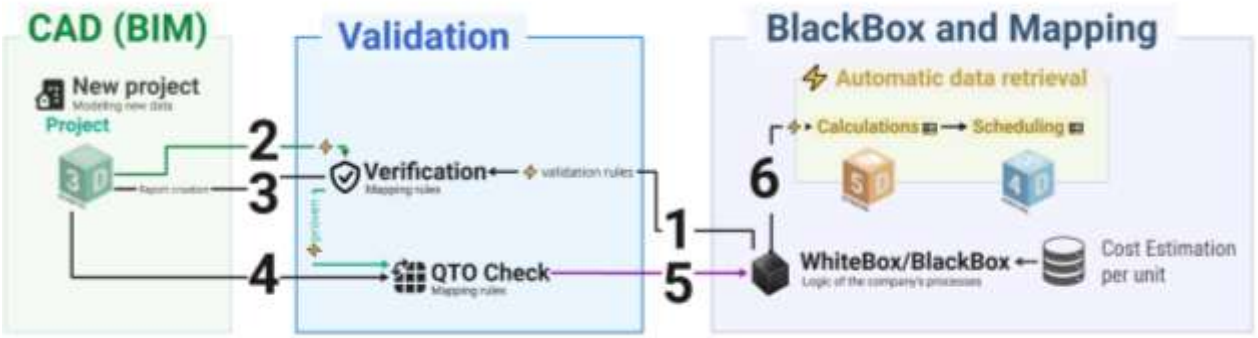


Рис. 5.4-2 Архитектура строительной ERP-системы с CAD (BIM), от создания правил валидации (1) до автоматического расчета стоимости и графиков работ (5-6).

В модульной ERP-системе процессы интегрируются с помощью программного обеспечения, которое включает в себя пользовательский интерфейс. За интерфейсом находится внутренняя часть, где структурированные таблицы обрабатывают данные, выполняя различные операции, которые предварительно настроил менеджер или администратор. В результате пользователь, благодаря заранее заданной и настроенной логике автоматизации (в модулях BlackBox/WhiteBox), получает полу-автоматически подготовленные документы, отвечающие его задачам.

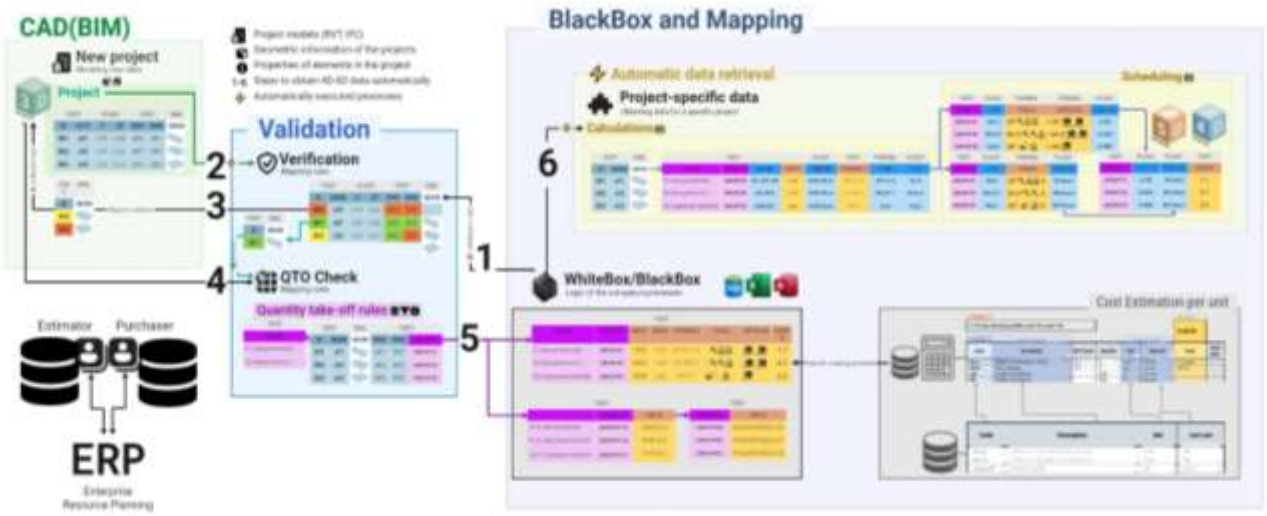


Рис. 5.4-3 ERP-система помогает менеджерам и пользователям перемещаться между таблицами специалистов, чтобы генерировать новые данные.

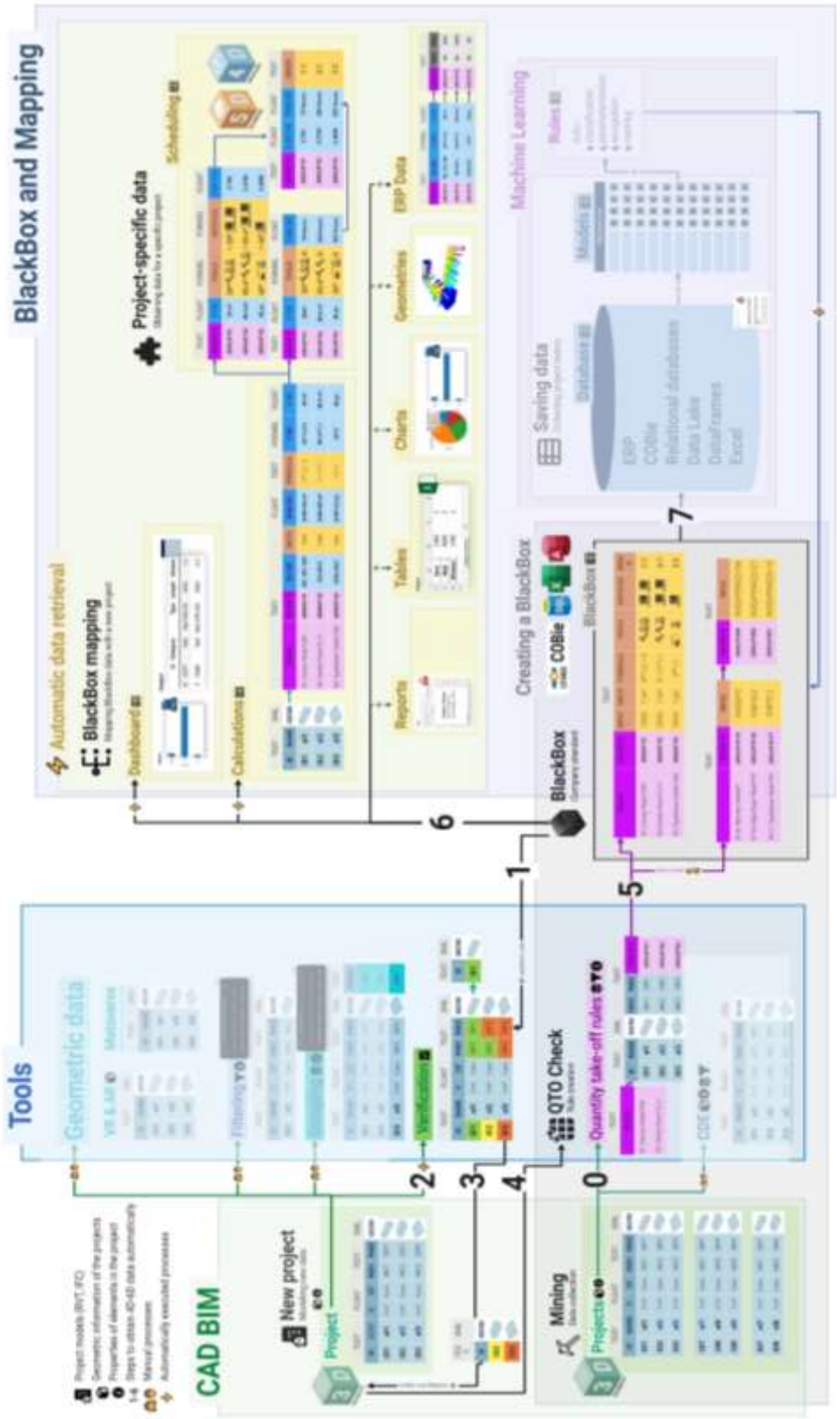


Рис. 5.4-4 ERP-система интегрирована с аналитическими инструментами и автоматизирует процесс принятия решений в компании.

Аналогичным образом процессы в ERP-системах, от начала до окончательного расчета (этапы 1-6 Рис. 5.4-3) представляют собой цепь взаимосвязанных шагов, которые в итоге обеспечивают прозрачность, эффективность и точность планирования.

Современные строительные ERP-системы включают в себя не только модули расчета стоимости и сроков, но и десятки других преднастроенных модулей, как правило, покрывающих функции документооборота, отслеживания хода реализации проекта, управления контрактами, цепочками поставок и логистикой, а также интеграции с другими бизнес-системами и платформами. Интегрированные аналитические инструменты ERP позволяют пользователям автоматизировать создание информационных панелей для мониторинга ключевых показателей (KPI - key performance indicators) проекта. Это обеспечивает централизованное и последовательное управление всеми аспектами строительного проекта, с попыткой объединить большое количество приложение и систем на одной платформе.

В будущем ERP-аналитика будет использоваться в сочетании с машинным обучением для повышения точности и оптимизации процесса расчета будущих атрибутов проекта. Данные и атрибуты, проанализированные и собранные из ERP-систем в Big Data (Рис. 5.4-4), в будущем станут основой для создания прогнозных моделей, которые смогут точно предвидеть потенциальные задержки, риски или, например, возможные изменения стоимости материалов.

Как альтернатива ERP, в строительной отрасли часто используется PMIS (Project Management Information System) — система управления проектами, предназначенная для детального контроля выполнения задач на уровне отдельного строительного объекта.

PMIS: Промежуточное звено между ERP и строительной площадкой

В отличие от ERP, которая охватывает всю цепочку бизнес-процессов компании, PMIS сосредоточена на управлении конкретным проектом, обеспечивая мониторинг сроков, бюджета, ресурсов и документации.

> PMIS (Project Management Information System) — это программное обеспечение для управления строительными проектами, предназначенное для планирования, отслеживания, анализа и отчетности по всем аспектам проекта.

PMIS позволяет управлять документами, графиками, бюджетами и на первый взгляд, PMIS может показаться дублирующим решением относительно ERP, однако ключевое отличие заключается в уровне управления:

- **ERP** ориентирована на бизнес-процессы компании в целом: управление затратами, контрактами, закупками, кадрами и ресурсами на корпоративном уровне.
- **PMIS** сфокусирована на управлении отдельными проектами, обеспечивая детальное планирование, контроль изменений, ведение отчетности и координацию участников.

Во многих случаях именно ERP-системы уже обладают достаточным функционалом, и внедрение PMIS становится скорее вопросом удобства и предпочтений компании. Многие подрядчики и заказчики используют PMIS не потому, что она необходима, а потому, что ее навязывает вендор или крупный заказчик, который хочет агрегировать данные на конкретной платформе.

Следует упомянуть, что в международной терминологии для управления строительными проектами существуют и другие отдельные популярные концепции, такие как PLM (Product Lifecycle Management) — управление жизненным циклом продукта, а также EPC и EPC-M (Engineering, Procurement and Construction Management) — методы контрактования в строительной отрасли.

Если в компании уже используется ERP с модулями проектного управления, внедрение PMIS может оказаться лишним звеном, дублирующим функционал. Однако если процессы не автоматизированы, а данные разрознены, PMIS может стать более удобным и легким в обслуживании инструментом.

Спекуляции, прибыль, закрытость и отсутствие прозрачности в ERP и PMIS

Несмотря на внешнюю простоту интерфейсов и процедур, строительные ERP и PMIS-системы в большинстве случаев представляют собой закрытые и негибкие решения. Такие системы, как правило, поставляются в виде предварительно настроенного программного комплекса от одного вендора, с ограниченным доступом к внутренним базам данных и логике процессов.

Разработку и контроль над такими системами всё чаще берут на себя поставщики CAD-(BIM-) продуктов, поскольку именно в их базах данных содержится информация, необходимая ERP-системам: количественные и объёмные атрибуты элементов проекта. Однако, вместо предоставления доступа к этим данным в открытом или машиночитаемом формате, вендоры предлагают лишь ограниченные пользовательские сценарии и закрытую логику обработки — предопределённую внутри BlackBox-модулей. Это снижает гибкость системы и мешает адаптировать её под конкретные условия проектов.

> Ограниченная прозрачность данных остаётся одной из ключевых проблем цифровых процессов в строительстве. Закрытая архитектура баз данных, отсутствие доступа к полным наборам атрибутов строительных элементов, ориентация на *BlackBox*-модули автоматизации и нехватка открытых интерфейсов значительно повышают риски документной бюрократии. Такие ограничения создают "узкие места" в процессе принятия решений, затрудняют верификацию информации и открывают возможности для сокрытия данных или спекуляций внутри ERP/PMIS-систем. Пользователи, как правило, получают лишь ограниченный доступ — будь то урезанный интерфейс или частичный API — без возможности взаимодействовать с первоисточниками данных напрямую. Особенно это критично, когда речь идёт об автоматически сгенерированных из CAD-проектов параметрах, таких как объёмы, площади и количества, используемых для расчётов QTO.

Как следствие, вместо поиска эффективности через автоматизацию процессов, открытость данных, сокращение транзакционных издержек и создания новых бизнес моделей, многие строительные компании фокусируются на управлении внешними параметрами — манипулируют коэффициентами, корректирующими факторами и методами расчётов, влияющими на стоимость проекта в закрытых платформах ERP/PMIS. Это создаёт почву для спекуляций, искажает реальные производственные затраты и снижает доверие между всеми участниками строительного процесса.

> В строительстве прибыль формируется как разница между выручкой от завершённого проекта и переменными затратами, в которые входят проектирование, материалы, трудовые ресурсы и другие прямые расходы, непосредственно связанные с реализацией объекта. Однако ключевым фактором, влияющим на величину этих затрат, становится не только технология или логистика, а скорость и точность расчётов, а также качество управленческих решений внутри компании.

Проблема усугубляется тем, что в большинстве строительных компаний процессы расчёта стоимости остаются непрозрачными не только для заказчиков, но и для самих сотрудников, не входящих в состав сметных или финансовых отделов. Такая закрытость способствует формированию внутри компании привилегированной группы специалистов — носителей "финансовой экспертизы", обладающих исключительным правом на редактирование атрибутов и корректирующих коэффициентов в ERP/PMIS-системах. Эти сотрудники, вместе с главами компаний могут фактически контролировать финансовую логику проекта.

> Сметчики, в таких условиях, превращаются в "финансовых жонглёров", балансируя между максимизацией прибыли компании и необходимостью сохранить конкурентную цену для клиента. При этом они вынуждены избегать явных и грубых манипуляций, чтобы не подорвать репутацию компании. Именно на этом этапе закладываются коэффициенты, скрывающие завышенные объёмы или стоимости материалов и работ.

В результате основной схемой повышения эффективности и рентабельности компаний, работающих в строительной отрасли, становится не автоматизация и ускорение процессов принятия решений, а спекуляция на ценах материалов и работ (Рис. 5.4-5). Завышение стоимости работ и материалов осуществляется "серой" бухгалтерией в закрытых ERP/PMIS - системах путем завышения процентов над среднерыночными ценами на материалы или объемами работ при помощи коэффициентов (Рис. 5.1-6), которые обсуждались в главе «Составление калькуляций и расчет стоимости работ на основе ресурсной базы».

> В результате заказчик получает расчёт, который не отражает реальную стоимость или объёмы работ, а представляет собой производную от множества скрытых внутренних коэффициентов. При этом субподрядчики, в попытке соответствовать заниженным расценкам, заложенным генеральным подрядчиком, зачастую вынуждены закупать более дешёвые и некачественные материалы, что ухудшает итоговое качество строительства.

Спекулятивный процесс поиска прибыли из воздуха в итоге вредит как клиентам, получающим недостоверные данные, так и исполнителям, которые вынуждены искать всё новые и новые модели спекуляций.

В результате, чем масштабнее проект, тем выше уровень бюрократии в управлении данными и процессами. За каждым этапом и каждым модулем зачастую скрываются непрозрачные коэффициенты и надбавки, встроенные в расчётные алгоритмы и внутренние процедуры. Это не только затрудняет аудит, но и существенно искажает финансовую картину проекта. В крупных строительных

проектах подобные практики нередко приводят к многократному (иногда до десятикратного) увеличению итоговой стоимости, при этом реальные объёмы и затраты остаются вне зоны эффективного контроля со стороны заказчика (Рис. 2.1-3 Сравнение запланированных и фактических затрат на крупные инфраструктурные проекты в Германии.

> Согласно отчету McKinsey & Company "Воображая цифровое будущее строительства" (2016), крупные строительные проекты в среднем завершаются на 20% позже запланированного срока и превышают бюджет до 80% [107].

Отделы смет и бюджетирования становятся самым охраняемым звеном внутри компании. Доступ к ним строго ограничен даже для внутренних специалистов, а из-за закрытости логики и структур баз данных невозможно объективно оценить эффективность проектных решений без искажений. Отсутствие прозрачности приводит к тому, что компании вынуждены не оптимизировать процессы, а бороться за выживание путём "творческого" управления цифрами и коэффициентами (Рис. 5.3-1, Рис. 5.1-6 - например параметр „Bid. Factor“).

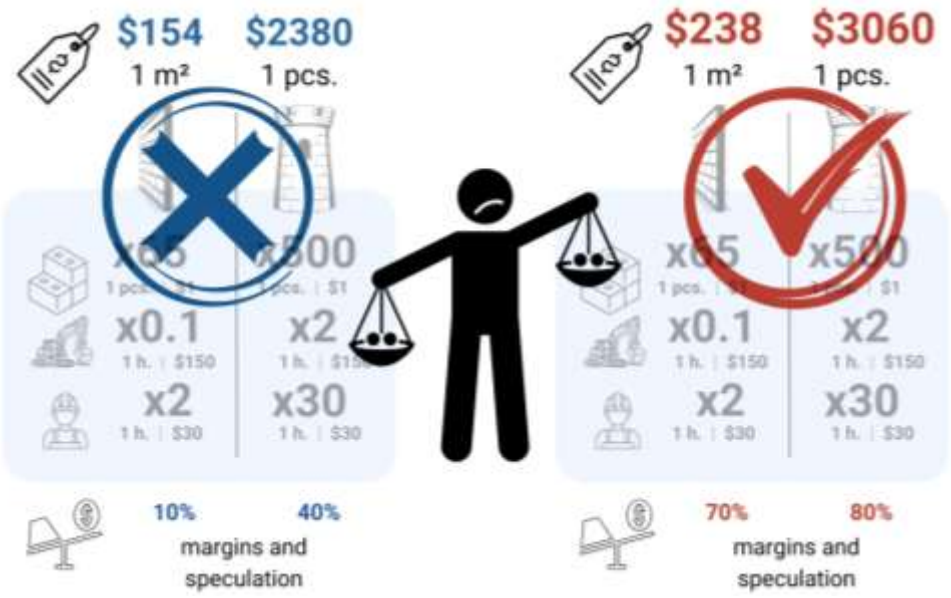


Рис. 5.4-5 Коэффициенты спекуляции на уровне расчетов — это основная прибыль компаний и искусство жонглирования между качеством работы и репутацией.

Всё это ставит под сомнение целесообразность дальнейшего использования закрытых ERP/PMIS-систем в строительстве. В условиях цифровой трансформации и возрастающих требований к прозрачности со стороны заказчиков (Рис. 10.2-3), маловероятно, что реализация проектов в долгосрочной перспективе останется зависимой от проприетарных решений, ограничивающих гибкость, мешающих интеграции и сдерживающих развитие бизнеса.

И как бы ни было выгодно строительным компаниям работать с силосами данных и непрозрачными данными в закрытых базах данных – неизбежно будущее строительной отрасли будет связано с переходом к открытым платформам, машиночитаемым и прозрачным структурам данных, а также автоматизации, основанной на принципах доверия. Эта трансформация будет двигаться «сверху» — под давлением заказчиков, регулирующих органов и общества, всё чаще требующих подотчётности, устойчивости, прозрачности и экономической обоснованности.

Конец эпохи закрытых ERP/PMIS: строительная отрасль нуждается в новых подходах

Использование громоздких модульных ERP/PMIS-систем, состоящих из десятков миллионов строк кода, делает любые изменения в них крайне затруднительными. При этом переход на новую платформу в условиях наличия уже преднастроенных под компанию модулей, десятков тысяч артикулов в базах ресурсов (Рис. 5.1-3) и тысяч готовых калькуляций (Рис. 5.1-6) превращается в дорогостоящий и продолжительный процесс. Чем больше кода и устаревших архитектурных решений — тем выше уровень внутренней неэффективности, а каждый новый проект будет только усугублять ситуацию. Во многих компаниях перенос данных и интеграция новых решений становятся многолетними эпопеями, сопровождающимися постоянными доработками и бесконечным поиском компромиссов. В результате нередко происходит возврат к старым, знакомым платформам, несмотря на их ограничения.

Как подчёркивается в немецком докладе "Чёрная книга" (Германия) [108], посвящённом системным сбоям в управлении строительными данными, фрагментация информации и отсутствие централизованного подхода к её управлению — ключевая причина снижения эффективности. Без стандартизации и интеграции данные теряют свою ценность, превращаясь в архив, а не инструмент управления.

> Основной причиной потери качества данных является недостаточное планирование и контроль строительных проектов, что часто приводит к значительному росту затрат. В разделе «В центре внимания: взрыв затрат» ”Чёрной книги” анализируются ключевые факторы, способствующие таким нежелательным последствиям. Среди них — неадекватный анализ потребностей, отсутствие технико-экономических обоснований и несогласованное планирование, ведущее к дополнительным расходам, которых можно было бы избежать.

> В зрелой IT-экосистеме компании замена устаревшей системы сравнима с заменой несущей колонны в уже построенном здании. Недостаточно просто демонтировать старую и установить новую — важно сделать это так, чтобы здание осталось устойчивым, перекрытия не рухнули, а все коммуникации продолжали работать. Именно в этом и заключается сложность: любая ошибка может привести к серьезным последствиям для всей системы компании.

Тем не менее, разработчики крупных ERP-продуктов для строительной отрасли продолжают использовать количество написанного кода как аргумент в пользу своей платформы. На специализированных конференциях по-прежнему можно услышать фразы вроде: «Для воссоздания такой системы потребуется 150 человеко-лет», несмотря на то, что за большинством функционала подобных систем скрываются базы данных и достаточно простые функции по работе с таблицами, упакованными в специальный зафиксированный, пользовательский интерфейс. На практике объём кода «150 человеко-лет» превращается скорее в бремя, чем в конкурентное преимущество. Чем больше кода — тем выше стоимость поддержки, сложнее адаптация к новым условиям и выше порог входа для новых разработчиков и клиентов.

Сегодня многие модульные строительные системы напоминают громоздкие и устаревшие “Франкенштейн-конструкты”, где любое неосторожное изменение может привести к сбоям. Каждый новый модуль лишь усложняет и без того перегруженную систему, превращая её в лабиринт, понятный лишь узкому кругу специалистов, делая её ещё более сложной для поддержки и модернизации.

Сложность осознаётся и самими разработчиками, которые периодически делают паузы для рефакторинга — пересмотра архитектуры с учетом появления новых технологий. Однако даже если рефакторинг проводится регулярно, сложность неизбежно растёт. Архитекторы таких систем привыкают к нарастающей сложности, но для новых пользователей и специалистов это становится непреодолимым барьером. В результате вся экспертиза концентрируется в руках нескольких разработчиков, и система перестаёт быть масштабируемой. В краткосрочной перспективе такие эксперты полезны, но в долгосрочной — они становятся частью проблемы.

> *Организации продолжат интегрировать «малые» данные с их большими аналогами, и глуп тот, кто верит, что одно приложение — каким бы дорогим или надежным оно ни было — может справиться со всем [109].*
>
> — Фил Саймон, ведущий подкаста Conversations About Collaboration

Возникает закономерный вопрос: действительно ли нам нужны такие громоздкие и закрытые системы для расчёта стоимости и сроков работ в виде таблиц, если в других отраслях с аналогичными задачами давно справляются аналитические инструменты с открытыми данными и прозрачной логикой?

На текущий момент закрытые модульные платформы всё ещё востребованы в строительной отрасли, в первую очередь из-за специфики калькуляционного учёта (Рис. 5.1-7). Такие системы нередко используются для ведения «серых» или непрозрачных схем, позволяя скрывать реальные затраты от заказчика. Однако по мере цифровой зрелости, в первую очередь, заказчиков и перехода отрасли в так называемую «уберизированную эру», посредники, а именно строительные компании с их ERP, в расчётах времени и стоимости утратят свою значимость. Это навсегда изменит облик строительной индустрии. Подробнее в последней части книги и главе «Строительство 5.0: как зарабатывать, когда скрывать больше нельзя».

Тысячи устаревших легаси-решений, накопленных за последние 30 лет с тысячами человеко-лет, инвестированных в разработку, начнут стремительно исчезать. Переход к открытому, прозрачному и гибкому управлению данными — неизбежен. Вопрос лишь в том, какие компании смогут адаптироваться к этим изменениям, а какие останутся заложниками старой модели.

Похожая ситуация наблюдается и в сфере CAD- (BIM-) инструментов, чьи данные сегодня наполняют объёмные параметры проектных сущностей в ERP/PMIS-системах. Изначально идея BIM (разработанная ещё в 2002 году [110] строилась на концепции единой интегрированной базы данных, однако на практике сегодня работа с BIM требует целого набора специализированных программ и

форматов. То, что должно было упростить проектирование и управление строительством, превратилось в очередной слой проприетарных решений, усложняющих интеграцию и снижающих гибкость бизнеса.

Дальнейшие шаги: эффективное использование проектных данных

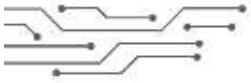
В этой части мы показали, как структурированные данные становятся основой для точных расчётов стоимости и сроков строительных проектов. Автоматизация процессов QTO, календарного планирования и сметного расчёта снижает трудозатраты и значительно повышает точность результатов.

Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные подходы в ваших повседневных задачах. Эти подходы универсальны — они полезны как для цифровой трансформации компании, так и для повседневной работы специалистов, занимающихся расчётами:

- Автоматизируйте рутинные расчеты
  - Попробуйте найти стандартные калькуляции работ, к которым вы можете иметь отношение в своей работе
  - Проанализируйте, какими методами калькулируются или рассчитываются работы или процессы на строительной площадке в вашей стране (Рис. 5.1-7)
  - Если вы работаете с из CAD систем - изучите функции автоматического извлечения спецификаций и QTO данных в вашем CAD- (BIM-) программном обеспечении
  - Используйте LLM для написания черновика кода по автоматизации расчетов
- Разработайте собственные инструменты для QTO
  - Создайте скрипты или таблицы для автоматизации подсчета объемов
  - Стандартизируйте категории и группы элементов для последовательного подхода к оценке
  - Документируйте методологию расчетов для обеспечения воспроизводимости результатов в новых проектах
- Интегрируйте различные аспекты проекта в своей работе
  - Если вы работаете с модульными системами, попробуйте визуализировать свои процессы не только как схемы или диаграммы, но и на уровне данных — особенно в виде таблиц
  - Освойте автоматическое объединение данных, извлечённых из CAD-баз, с расчётами — с помощью кода на Python, используя группировки, фильтрацию и агрегацию
  - Создавайте наглядные визуализации QTO групп для представления комплексной информации коллегам и клиентам

Эти шаги помогут выстроить устойчивую систему расчётов, основанную на автоматизации и стандартизации данных. Такой подход повысит точность, сократит рутину в повседневных вопросах, связанных с расчётами.

Следующие главы посвящены техническим аспектам CAD- (BIM-) продуктов и причинам, по которым базы данных CAD до сих пор сложно интегрировать в бизнес-процессы компаний. Если вас сейчас не интересуют история внедрения BIM в строительство, эволюция CAD-инструментов и технические особенности работы с этими технологиями, вы можете сразу перейти к седьмой части книги «Принятие решений на основе данных».



МАКСИМУМ УДОБСТВА С ПЕЧАТНОЙ ВЕРСИЕЙ

Вы держите в руках бесплатную цифровую версию **Data-Driven Construction**. Для более удобной работы и быстрого доступа к материалам рекомендуем обратить внимание на **печатное издание**:

- **Всегда под рукой**: книга в печатном формате станет надежным рабочим инструментом, позволяя быстро найти и использовать нужные визуализации и схемы в любых рабочих ситуациях
- **Высокое качество иллюстраций**: все изображения и графики в печатном издании представлены в максимальном качестве
- **Быстрый доступ к информации**: удобная навигация, возможность делать пометки, закладки и работать с книгой в любом месте.

Приобретая полную печатную версию книги, вы получаете удобный инструмент для комфортной и эффективной работы с информацией: возможность оперативно использовать визуальные материалы в повседневных задачах, быстро находить нужные схемы и делать пометки. Кроме того, ваша покупка поддерживает распространение открытых знаний.

Заказать печатную версию книги можно на: datadrivenconstruction.io/books

VI ЧАСТЬ

CAD И BIM: МАРКЕТИНГ, РЕАЛЬНОСТЬ И БУДУЩЕЕ ПРОЕКТНЫХ ДАННЫХ В СТРОИТЕЛЬСТВЕ

Шестая часть книги представляет критический анализ эволюции CAD и BIM-технологий и их влияния на процессы управления данными в строительстве. Прослеживается историческая трансформация концепции BIM от первоначальной идеи интегрированной базы данных до современных маркетинговых конструкций, продвигаемых вендорами программного обеспечения. Оценивается влияние проприетарных форматов и закрытых систем на эффективность работы с проектными данными и общую производительность строительной отрасли. Детально анализируются проблемы совместимости различных CAD-систем и трудности их интеграции с бизнес-процессами строительных компаний. Рассматриваются актуальные тенденции перехода к упрощенным открытым форматам данных, таким как USD, и их потенциальное влияние на отрасль. Представлены альтернативные подходы к извлечению информации из закрытых систем, включая методы обратного инжиниринга. Анализируются перспективы применения искусственного интеллекта и машинного обучения для автоматизации процессов проектирования и анализа данных в строительстве. Формулируются прогнозы развития технологий проектирования, ориентированных на реальные потребности пользователей, а не на интересы поставщиков программного обеспечения.

ГЛАВА 6.1.
ПОЯВЛЕНИЕ BIM-КОНЦЕПТОВ В СТРОИТЕЛЬНОЙ ОТРАСЛИ

Изначально этой шестой части, посвящённой CAD (BIM), в первой версии книги не было. Темы проприетарных форматов, геометрических ядер и закрытых систем являются чрезмерно техническими, перегруженными деталями и, на первый взгляд, бесполезными для тех, кто просто хочет разобраться в работе с данными. Однако отзывы и просьбы добавить разъяснение к первой версии книги показали: без понимания всех сложностей, связанных с внутренними механизмами CAD-систем, геометрических ядер, разнообразием форматов и несовместимых схем хранения одних и тех же данных, невозможно по-настоящему осознать, почему продвигаемые вендорами концепции зачастую затрудняют работу с информацией и препятствуют переходу к открытому параметризированному проектированию. Именно поэтому эта часть заняла своё отдельное место в структуре книги. Если тема CAD (BIM) для вас не является приоритетной, вы можете сразу перейти к следующей части — «VII ЧАСТЬ: Принятие решений на основе данных, аналитика, автоматизация и машинное обучение».

История появления BIM и open BIM как маркетинговых концептов CAD-вендоров

С появлением цифровых данных в 90-е годы компьютерные технологии стали внедряться не только в бизнес-процессы, но и в процессы проектирования, что привело к появлению таких концепций, как CAD (системы автоматизированного проектирования) и позже, BIM (информационное моделирование зданий).

Однако, как и любые инновации, они не являются конечной точкой развития. Концепции вроде BIM стали важным этапом в истории строительной отрасли, но рано или поздно они могут уступить место более совершенным инструментам и подходам, которые будут лучше отвечать вызовам будущего.

> Поглощённый влиянием CAD-вендоров и запутавшись в сложностях собственной реализации, концепт BIM, появившийся в 2002 году, вполне может не дожить до своего тридцатилетия, подобно рок-звезде, которая ярко вспыхнула, но быстро угасла. Причина проста: запросы специалистов по работе с данными изменяются быстрее, чем к ним успевают адаптироваться CAD-вендоры.

Сталкиваясь с нехваткой качественных данных, современные специалисты в строительной отрасли требуют кроссплатформенной совместимости и доступа к открытым данным из CAD-проектов для упрощения их анализа и обработки. Сложность CAD-данных и запутанность их обработки негативно сказываются на всех участниках строительного процесса: проектировщиках, менеджерах проектов, строителях на площадке и, в итоге, на заказчике.

Вместо полноценного набора данных для эксплуатации сегодня заказчик и инвестор получают контейнеры в CAD-форматах, которые требуют сложных геометрических ядер, понимания схем данных, ежегодно обновляемой API-документации и специализированного программного обеспечения

CAD (BIM) для работы с данными. При этом большая часть проектных данных остается неиспользованными.

Сегодня в мире проектирования и строительства сложность доступа к CAD-данным приводит к чрезмерной инженерии управления проектами. Средние и крупные компании, работающие с CAD-данными или разрабатывающими BIM-решения либо вынужденно поддерживают тесные отношения с поставщиками CAD-решений для доступа к данным через API, либо обходят ограничения поставщиков CAD, используя дорогие конвертеры SDK для реверс-инжиниринга, чтобы получить открытые данные [75].

> Подход по использованию проприетарных данных устарел и уже не отвечает требованиям современной цифровой среды. Будущее разделит компании на два типа: те, кто эффективно использует открытые данные, и те, кто уйдёт с рынка.

Концепция BIM (Building Information Modeling), появилась в строительной отрасли с публикацией одного из крупных CAD-вендоров - Whitepaper BIM [54] в 2002 и, дополнившись машиностроительной концепцией BOM (Bills of Materials), взяла свое начало из параметрического подхода к созданию и обработке данных проекта (Рис. 6.1-1). Параметрический подход к созданию и обработке проектных данных одним из первых был реализован в системе Pro-E для машиностроительного проектирования (MCAD). Эта система стала прототипом [111] для многих современных CAD-решений, включая те, что сегодня применяются в строительной отрасли.

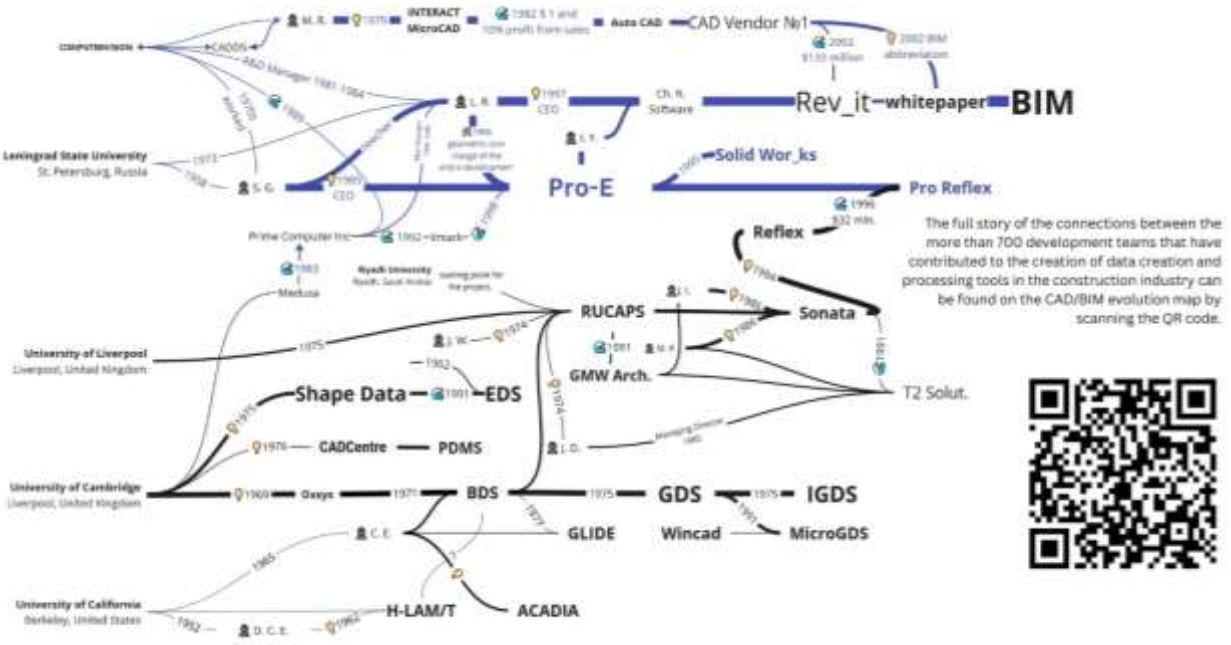


Рис. 6.1-1 Карта истории появления концепта BIM и похожих концептов.

Журналисты и консультанты AEC, продвигавшие инструменты CAD-вендоров до начала 2000-х, с 2002 года переключили внимание на Whitepaper BIM. Именно Whitepaper BIM 2002-2004 гг. и статьи, опубликованные в 2002, 2003, 2005 и 2007 годах, сыграли ключевую роль в популяризации концепции BIM в строительной отрасли [112].

> *Информационное моделирование зданий — это стратегия …….. [название компании CAD вендора] по применению информационных технологий в строительной индустрии.*
>
> — Whitepaper BIM, 2002 *[60]*

К середине 2000-х «исследователи» начали связывать BIM-концепт, опубликованный CAD-вендором в 2002 году с более ранними научными работами, например BDS Чарльза Истмена, которая стала основой для таких систем, как GLIDE, GBM, BPM, RUCAPS. В своей новаторской работе "Building Description System" (1974) Чарльз Истмен заложил теоретические основы современного информационного моделирования. Термин "база данных" встречается в его труде 43 раза (Рис. 6.1-2) — чаще любого другого, за исключением слова "здание".

Ключевая идея Истмена состояла в том, что вся информация о здании — от геометрии до свойств элементов и их взаимосвязей — должна храниться в единой структурированной базе данных. Именно из этой базы можно автоматически генерировать чертежи, спецификации, расчеты и анализировать соответствие нормативам. Истмен прямо критиковал чертежи как устаревший и избыточный способ коммуникации, указывая на дублирование информации, проблемы с актуализацией и необходимость ручного обновления при внесении изменений. Вместо этого он предлагал единую цифровую модель в базе данных, где любое изменение вносится единожды и автоматически отражается во всех представлениях.

Примечательно, что в своей концепции Истмен не ставил визуализацию во главу угла. Центральное место в его системе занимала именно информация: параметры, связи, атрибуты, возможности анализа и автоматизации. Чертежи в его понимании были лишь одной из форм отображения данных из базы, а не первоисточником проектной информации.

В первых Whitepaper по BIM от ведущего CAD-вендора словосочетание "база данных" использовалось так же часто как и в BDS Чарльза Истмана - 23 раза [60] на семи страницах и было одним из самых популярных слов в документе после «Building», «Information», «Modeling» и "Design". Однако к 2003 году в аналогичных документах термин "база данных" встречается лишь дважды [61], а к концу 2000-х тема баз данных практически исчезает из обсуждения проектных данных. В итоге концепция "единой интегрированной базы данных для визуального и количественного анализа" так и не была полностью реализована.

Таким образом, строительная отрасль прошла путь от прогрессивной концепции BDS Чарльза Истмена с его акцентом на базы данных и идей Сэмюэля Гейсберга об автоматическом обновлении проектных данных из баз данных в машиностроительном продукте Pro-E (предшественника популярных CAD-решений используемых сегодня в строительстве) к современному маркетинговому BIM, где управление данными через базы данных практически не упоминается, несмотря на то, что именно эта концепция лежала в основе первоначальных теоретических разработок.

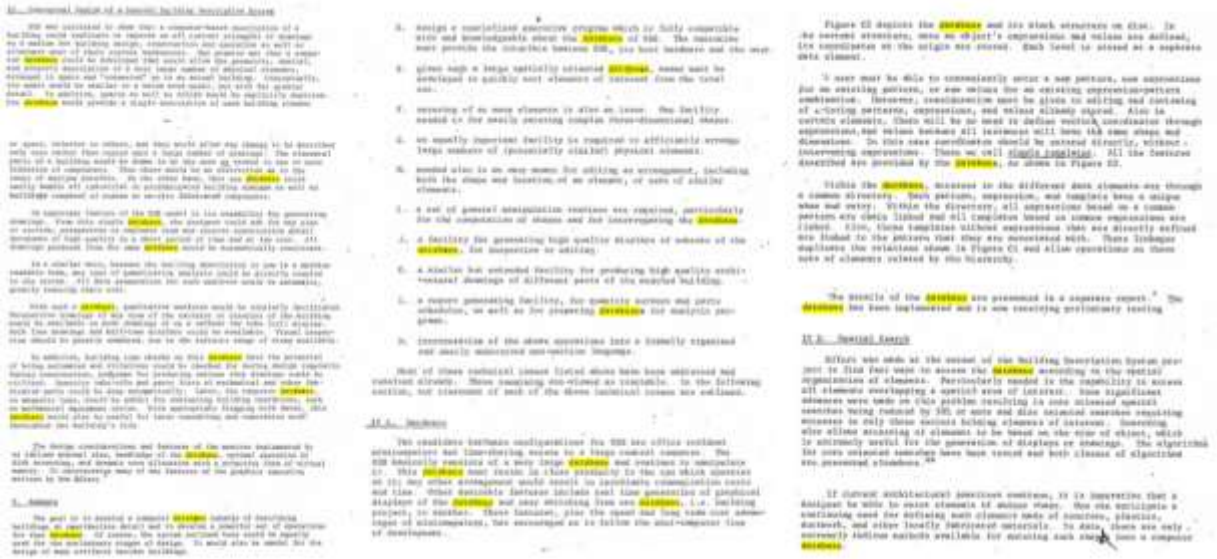


Рис. 6.1-2 В концепте BDS, описанным Чарльзом Истмена в 1974 году, словосочетание "База данных" (выделено жёлтым) использовалось 43 раза.

BDS и подобные ему концепты до 2000-х годов были разработаны как цифровая база данных зданий, а не как инструмент визуализации. BIM в 2002 году стал инструментом проектирования, где база данных отошла на второй план. Что мы потеряли при переходе от BDS и подобных концептов в 1990-х к BIM к середине 2010-х:

- Открытые базы данных: BDS и другие подобные концепции делали упор на аналитику, BIM - на дизайн.
- Гибкость работы с данными: BDS делала упор на аналитику данных, BIM - на процессы, которые должны быть основаны непонятно на каких данных.
- Прозрачность: BDS задумывалась как открытая интегрированная база данных, в то время как поставщики CAD в BIM сделали свои базы данных полностью закрытыми и 20 лет безуспешно боролись с инструментами обратного инжиниринга, которые открывают проприетарные форматы.

За последние 30 лет проектировщики так и не получили доступ к «интегрированной базе данных» и после двадцати лет маркетинговой эйфории вокруг BIM-инструментов индустрия строительства начинает осознавать последствия этого увлечения.

Реальность BIM: вместо интегрированных баз данных - закрытые модульные системы

Вместо того, чтобы сосредоточиться на данных, их структурировании и интеграции в единые процессы, пользователи CAD- (BIM-) систем вынуждены работать с фрагментированным набором проприетарных решений, каждое из которых диктует свои правила игры:

- **Единая база данных,** о которых шла речь в первых Whitepaper BIM**, так и осталась мифом.** Несмотря на громкие заявления, доступ к данным по-прежнему ограничен и распределен между закрытыми системами.

- **BIM-модели стали** не инструментом, а **замкнутой экосистемой.** Вместо прозрачного обмена информацией пользователи вынуждены платить за подписки и использовать проприетарные API.
- **Данные принадлежат вендорам, а не пользователям.** Информация о проектах закрыта в проприетарных форматах или облачных сервисах, а не доступна в открытых и независимых форматах.

Инженеры-проектировщики и менеджеры проектов зачастую не имеют доступа ни к базе данных CAD-систем, ни к формату, в котором хранятся данные их собственного проекта. Это делает невозможным быструю проверку информации или формулирование требований к структуре и качеству данных (Рис. 6.1-3). Доступ к таким данным требует целого набора специализированных программ, объединённых через API и плагины, что приводит к чрезмерной бюрократизации процессов в строительной отрасли. Тем временем эти данные одновременно используются десятками информационных систем и сотнями специалистов.

> *Нам нужно уметь управлять всеми этими данными [CAD (BIM)] хранить их в цифровом виде и продавать программное обеспечение для управления жизненным циклом и процессами, потому что **на каждого инженера** [проектировщика], который что-то создает [в CAD программе], **приходится десять человек**, которые работают с этими данными" [41].*
>
> – Генеральный директор CAD-вендора, создавшего концепт BIM, 2005 г.

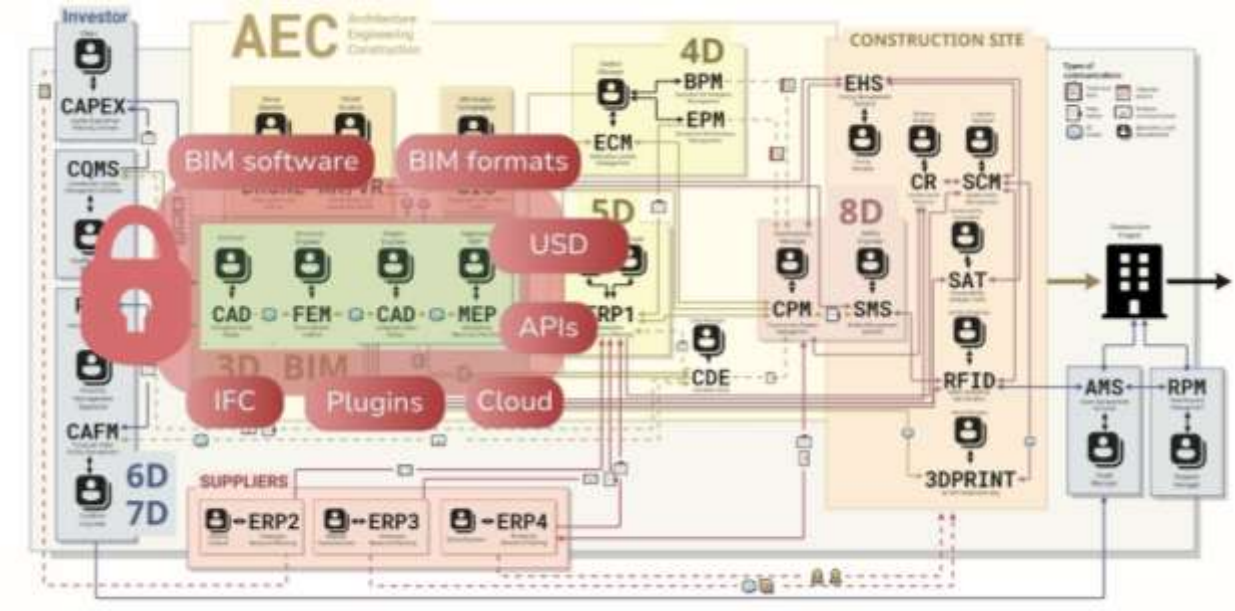


Рис. 6.1-3 CAD- (BIM-) базы данных остаются одними из последних закрытых систем для ИТ отделов и менеджеров данных в экосистеме строительного бизнеса.

> Когда становится очевидно, что BIM — это, скорее, средство коммерциализации баз данных, а не полноценный инструмент управления ими, возникает логичный вопрос: как вернуть контроль над данными? Ответ — использование открытых структур данных, где владельцем информации становится сам пользователь, а не поставщик ПО.

Пользователи и разработчики решений в строительной отрасли, подобно своим коллегам из других отраслей экономики, будут неизбежно отходить от расплывчатой терминологии поставщиков ПО, которая доминировала последние 30 лет, сосредоточившись на ключевых аспектах цифровизации — «данных» и «процессах».

Ещё в конце 1980-х годов ключевое направление развития цифровых технологий в строительстве представлялось как вопрос доступа к данным и управления информацией проекта. Однако со временем акценты сместились. Вместо развития прозрачных и доступных подходов к работе с данными, началось активное продвижение формата IFC и концепции open BIM — как попытки отвлечь внимание специалистов от тем управление базами проектных данных.

Появление открытого формата IFC в строительной отрасли

Так называемый открытый формат IFC (Industry Foundation Classes) позиционируется как стандарт для обеспечения совместимости между различными CAD- (BIM-) системами. Его разработка велась в рамках организаций, которые были созданы и контролировались крупнейшими CAD-вендорами. На основе формата IFC двумя CAD-компаниями в 2012 году была разработана маркетинговая концепция OPEN BIM [63].

> **IFC** (Industry Foundation Classes) – это открытый стандарт обмена данными в строительной отрасли, разработанный для обеспечения совместимости между различными CAD- (BIM-) системами.

Open BIM-концепт подразумевает под собой работу с информацией из CAD баз данных и обмен информации между системами через открытый формат для обмена CAD-данными - IFC.

> *Программа Open BIM - это маркетинговая кампания, инициированная … [1 CAD вендор], … [2 CAD вендор] и другими компаниями с целью поощрения и содействия глобальному скоординированному продвижению концепции OPEN BIM во всей AEC-индустрии, с согласованной коммуникацией и общим брендингом, доступным участникам программы.*
>
> — С сайта CAD вендора, OPEN BIM Program, 2012 *[113]*

IFC был адаптирован техническим университетом Мюнхена из машиностроительного формата STEP в конце 1980х, и позже зарегистрирован крупной проектировочной компанией и крупным CAD-вендором для создания альянса IAI (Industry Alliance for Interoperability) в 1994 году [114] (Рис.

6.1-4). Формат IFC был разработан для обеспечения интероперабельности между различными CAD-системами и базировался на принципах, заложенных в машиностроительном формате STEP, который, в свою очередь, появился из формата IGES, созданного ещё в 1979 году группой CAD-пользователей и поставщиков при поддержке NIST (The National Institute of Standards and Technology) и Министерства обороны США [115].

Однако сложная структура IFC, его тесная зависимость от геометрического ядра, а также разночтения в реализации формата различными программными решениями привели к множеству проблем при его практическом применении. С аналогичными трудностями — потерей детализации, ограничением точности и необходимостью использовать промежуточные форматы — ранее сталкивались и специалисты по машиностроению в работе с форматами IGES, STEP из которых и появился IFC.

Рис. 6.1-4 Карта связей команд разработчиков и продуктов CAD (BIM) [116].

> В 2000 году тот же CAD-вендор, зарегистрировавший формат IFC и создавший организацию IAI (позже bS), публикует Whitepaper «Интегрированное проектирование и производство: Преимущества и обоснование» [65]. В документе подчеркивалась важность сохранения полной детализации данных при обмене между программами внутри одной системы, без использования нейтральных форматов, таких как IGES, STEP [идентичных IFC]. Вместо этого предлагалось обеспечить прямой доступ приложений к основной базе данных CAD, что должно было предотвратить потерю точности информации.

В 2002 году этот же CAD вендор покупает параметрический BOM продукт (Рис. 3.1-18, подробнее в

третьей части) и на его основе формирует концепт BIM. В итоге в обмене данными строительных проектов теперь используются только или закрытые форматы CAD, или формат IFC (STEP), об ограничениях которого писал сам CAD-вендор в 2000 году, привнёсший этот формат в строительную отрасль.

Подробная история взаимодействия более чем 700 команд разработчиков, участвовавших в создании инструментов для создания и обработки строительных данных, представлена на карте «Эволюция CAD (BIM)» [116].

Открытая форма IFC состоит из геометрического описания элементов проекта и описания метаинформации. Для представления геометрии в формате IFC используются различные способы, такие как CSG и Swept Solids: однако параметрическое представление BREP стало лидирующим стандартом для передачи геометрии элементов в IFC формате, так как подобный формат поддерживается при экспорте из CAD- (BIM-) программ и позволяет потенциально редактировать элементы при обратном импорте IFC в CAD программы.

Проблема формата IFC в зависимости от геометрического ядра

В большинстве случаев, когда геометрия в IFC задана параметрически (BREP), визуализировать или получить геометрические свойства, такие как объём или площадь сущностей проекта, становится невозможно при наличии лишь файла IFC, потому что для работы с геометрией и её визуализацией в таком случае потребуется геометрическое ядро (Рис. 6.1-5), которое изначально отсутствует.

> **Геометрическое ядро** – это программный компонент, который предоставляет базовые алгоритмы для создания, редактирования и анализа геометрических объектов в CAD (CAD), BIM и других инженерных приложениях. Оно отвечает за построение 2D- и 3D-геометрии, а также за операции над ней, такие как: булевы операции, сглаживание, пересечения, трансформации и визуализация.

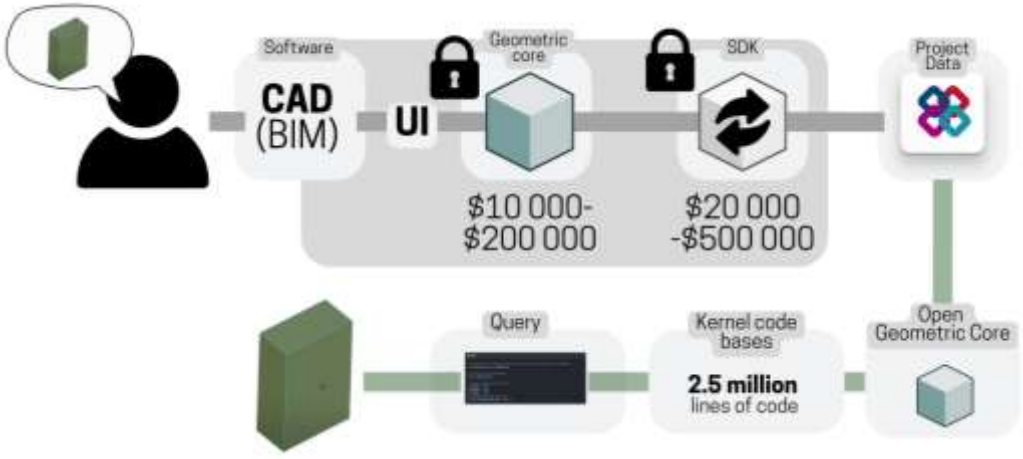


Рис. 6.1-5 Создание геометрии через CAD-программы сегодня проходит через проприетарные геометрические ядра и SDK, которые часто не принадлежат CAD-вендорам.

У каждой программы CAD и у любых программ работающих с параметрическими форматами или форматом IFC есть собственное или купленное геометрическое ядро. И если с примитивными элементами в IFC-BREP формате может не возникать проблем, и в программах с разными геометрическими ядрами эти элементы могут отображаться похоже, то, помимо проблем с разными движками геометрических ядер, существует достаточно элементов, которые имеют свои особенности для правильного отображения. Эта проблема подробно рассмотрена в международном исследовании " Референсное исследование программной поддержки IFC", опубликованном 2019 года [117].

> *Одни и те же стандартизированные наборы данных дают противоречивые результаты, при этом обнаруживается мало общих закономерностей, и серьезные проблемы найдены в поддержке стандарта [IFC], вероятно, из-за той самой высокой сложности стандартной модели данных. Отчасти здесь виноваты сами стандарты, поскольку они часто оставляют некоторые детали неопределенными, с высокой степенью свободы и различными возможными интерпретациями. Они допускают высокую сложность в организации и хранении объектов, что не способствует эффективному универсальному пониманию, уникальным реализациям и согласованному моделированию данных [117].*
>
> — Reference study of IFC software support, 2021

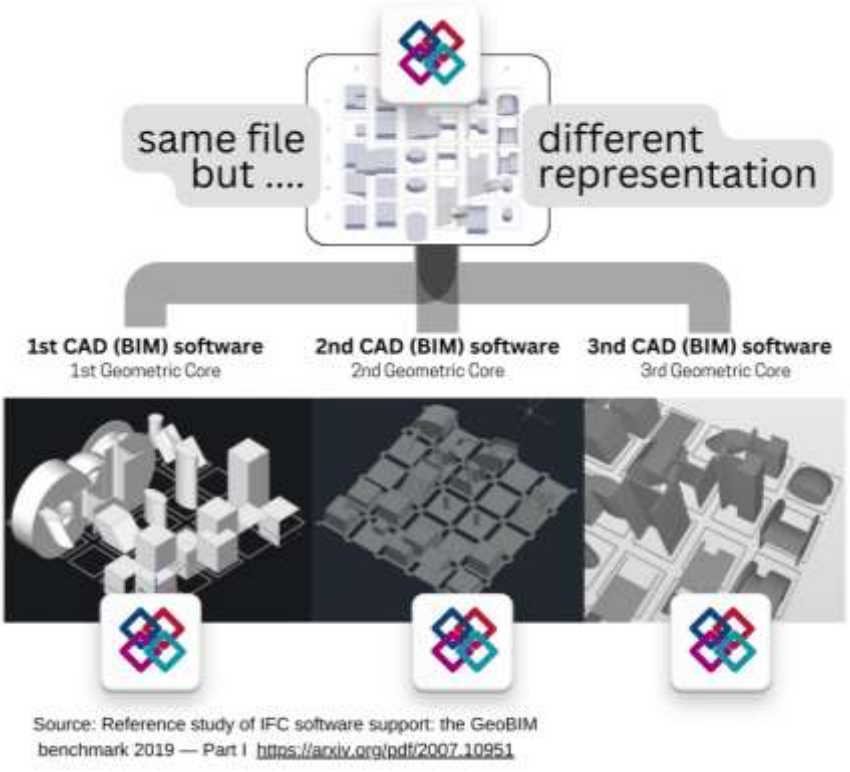


Рис. 6.1-6 Разные геометрические ядра дают разное представление одной и той же геометрии, описанной параметрически (на материалах [117]).

Правильное понимание “определённых положений” доступно платным членам специальных организаций, которые занимаются разработкой IFC. Как следствие, тот, кто хочет получить доступ к важным знаниям об определенных особенностях IFC, будет пытаться кооперироваться с крупными CAD-вендорами, либо доходить до качественного учёта особенностей собственными исследованиями.

> *Ты натыкаешься на вопрос об импорте и экспорте данных через формат IFC и спрашиваешь у коллег-вендоров: “А почему так в файле IFC передаётся информация про параметрическую передачу помещений? В открытой спецификации ничего про это не сказано”. Ответ от "более знающих” европейских вендоров: “Да, не сказано, но допустимо"*
>
> — Из интервью разработчика CAD 2021 *[118]*

IFC описывает геометрию через параметрические примитивы, но не содержит встроенного ядра — его роль выполняет CAD-программа, которая через геометрическое ядро компилирует геометрию. Геометрическое ядро выполняет математические вычисления и определяет пересечения, а IFC лишь предоставляет данные для его интерпретации. Если IFC содержит некорректные грани, разные программы с разным геометрическим ядром могут либо игнорировать их, либо выдавать ошибки, в зависимости от ядра.

> В итоге чтобы работать с форматом IFC необходимо ответить на основной вопрос, на который трудно найти однозначный ответ - какой инструмент, с каким геометрическим ядром необходимо использовать чтобы получить то качество данных, которое изначально у проекта было в CAD программе, из которой был получен IFC?

Проблемы качества данных и сложность формата IFC не позволяют напрямую использовать проектные данные для автоматизации процессов, их анализа и обработки данных, что часто приводит разработчиков к неизбежной необходимости использования закрытых CAD-решений с "качественным" доступом к данным [63], о чём и писал сам вендор, зарегистрировавший IFC в 1994 году [65].

Все особенности отображения и генерации параметров IFC в ядре геометрии могут быть реализованы только большими командами разработчиков, у которых есть опыт работы с геометрическими ядрами. Поэтому нынешняя практика особенностей и сложности формата IFC, выгодна прежде всего CAD-вендорам и имеет много общего со стратегией крупных вендоров ПО «adopt, extend, destroy», когда растущая сложность стандарта фактически создает барьеры для небольших игроков рынка [94].

Стратегия крупных вендоров в такой стратегии может заключаться в адаптации открытых стандартов, добавлении собственных расширений и функций, чтобы создать зависимость пользователей от своих продуктов для последующего вытеснения конкурентов.

> Формат IFC, призванный стать универсальным мостом между различными CAD- (BIM-) системами, в реальности выполняет роль индикатора проблем совместимости между геометрическими ядрами различных CAD-платформ, подобно STEP формату, из которого он изначально появился.

В итоге сегодня полноценная и качественная реализация онтологии IFC под силу крупным поставщикам CAD, которые могут вложить значительные ресурсы в поддержку всех сущностей и их маппинга с их же внутренним геометрическим ядром, которого не существует для IFC как стандарта. Крупные вендоры также имеют возможность согласовывать между собой технические детали особенностей, которые могут быть недоступны даже самому активному участнику организаций, занимающихся разработкой формата IFC.

> Для небольших независимых команд и проектов с открытым исходным кодом, стремящихся поддерживать развитие интероперабельных форматов, отсутствие собственного геометрического ядра превращается в серьёзную проблему. Без него практически невозможно учесть всё многообразие тонкостей и нюансов, связанных с межплатформенным обменом данными.

С развитием параметрического формата IFC и концепции open BIM, в строительной отрасли активизировались дискуссии о роли онтологии и семантики в управлении данными и процессами.

Появление в строительстве темы семантики и онтологии

Благодаря идеям создания семантического интернета в конце 1990-х гг. и усилиям организаций, занимающихся развитием формата IFC, семантика и онтологии стали одними из ключевых элементов стандартизации, обсуждаемой в строительной отрасли к середине 2020-х гг.

> Семантические технологии — это унификация, стандартизация и модификация больших массивов разнородных данных, а также реализация сложного поиска.

Для хранения семантических данных используется язык онтологий OWL (Web Ontology Language), представленный в виде графов RDF-триплетов (Resource Description Framework) (Рис. 6.1-7). OWL относится к графовым моделям данных, про виды которых мы говорили подробнее в главе «Модели данных: отношение в данных и связи между элементами».

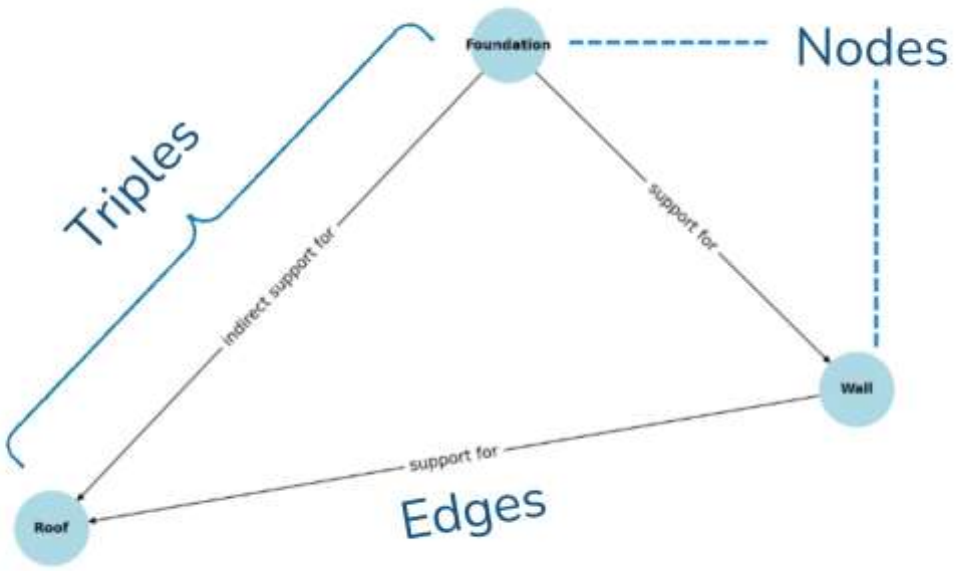


Рис. 6.1-7 Модель данных RDF: узлы (Nodes), связи (Edges) и тройки (Triples), иллюстрирующие отношения между строительными элементами.

Теоретически, логический вывод ризонеров (программы для автоматического логического вывода) позволяет получать новые утверждения на основе онтологий. Например, если в строительной онтологии записано, что "фундамент — это опора для стены", а "стена — это опора для крыши" (Рис. 6.1-7), ризонер способен автоматически сделать вывод, что "фундамент — это опора для крыши".

Подобный механизм полезен для оптимизации анализа данных, поскольку позволяет избежать явного прописывания всех зависимостей. Однако он не создает новых знаний, а лишь выявляет и структурирует уже известные факты.

> Семантика не создает нового смысла или знаний сама по себе и не превосходит другие технологии хранения и обработки данных в этом аспекте. Представление данных из реляционных баз в виде триплетов не делает их более осмысленными. Замена таблиц на графовые структуры может быть полезна для унификации моделей данных, удобного поиска и безопасного редактирования, но она не делает данные "умнее" — компьютер не начинает лучше понимать их содержание.

Логические связи в данных можно организовать и без сложных семантических технологий (Рис. 6.1-8). Традиционные реляционные базы данных (SQL), а также форматы CSV или XLSX позволяют строить аналогичные зависимости. Например, в колончатой базе данных можно добавить поле "опора крыши" и автоматически связать крышу с фундаментом при создании стены. Такой подход реализуется без использования RDF, OWL, графов или ризонеров, оставаясь простым и эффективным решением для хранения и анализа данных.

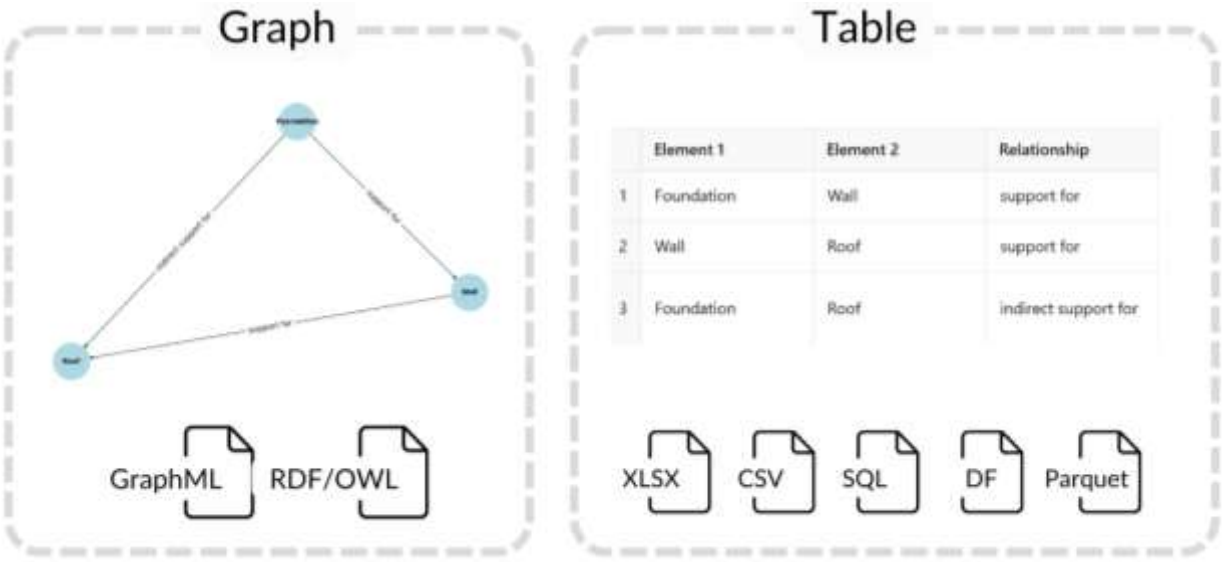


Рис. 6.1-8 Сравнение графовой и табличной моделей данных представления одних и тех же логических связей.

Решение ряда крупных строительных компаний и организации, занимающейся разработкой формата IFC [94], следовать концепции семантического веба, казавшейся перспективной в конце 1990-х гг., оказало значительное влияние на разработку стандартов в строительной отрасли.

Однако парадокс заключается в том, что сама концепция семантического веба, изначально предназначенная для интернета, так и не получила широкого распространения даже в своей родной среде. Несмотря на разработку RDF и OWL, полноценного семантического веба в изначальном замысле так и не появилось, и его создание уже маловероятно.

Почему семантические технологии не оправдывают ожиданий в строительстве

Другие отрасли столкнулись с ограничениями технологий использования семантики. В игровой индустрии попытки описания игровых объектов и их взаимодействий через онтологии оказались неэффективными из-за высокой динамики изменений. В итоге более простые форматы данных, такие как XML и JSON, вместе с алгоритмическими решениями, оказались предпочтительнее. Подобная ситуация сложилась и в сфере недвижимости: из-за региональных различий в терминологии и частых изменений рынка использование онтологий оказалось чрезмерно сложным, а простые базы данных и стандарты, такие как RETS [119], лучше справлялись с задачами обмена данными.

> Технические сложности, такие как сложность разметки, высокая трудоемкость поддержки и низкая мотивация разработчиков, тормозили внедрение семантического веба и в других секторах экономики. RDF (Resource Description Framework) не стал массовым стандартом, а онтологии оказались слишком сложными и экономически неоправданными.

В результате амбициозная идея создания глобального семантического веба не оправдалась. Хотя отдельные элементы технологии, такие как онтологии и SPARQL, нашли применение в корпоративных решениях, изначальная цель создания единой всеобъемлющей структуры данных так и не была достигнута.

Концепция интернета, в котором компьютеры способны понимать смысл контента, оказалась технически сложной и коммерчески нерентабельной. Именно поэтому компании, поддерживавшие эту идею, со временем сократили ее использование до отдельных полезных инструментов, оставив RDF и OWL для узкоспециализированных корпоративных нужд, а не для интернета в целом. Анализ Google Trends (Рис. 6.1-9) за последние 20 лет позволяет оценить, что перспектив развития семантического веба, возможно, больше не осталось.

> *Не нужно множить сущности без необходимости. Если существует несколько логически непротиворечивых объяснений какого-либо явления, объясняющих его одинаково хорошо, то следует, при прочих равных условиях, предпочитать самое простое из них.*
>
> — Бритва Оккама

Здесь возникает логичный вопрос: зачем вообще использовать триплеты, ризонеры и SPARQL в строительстве, если можно обрабатывать данные с помощью популярных структурированных запросов (SQL, Pandas, Apache®)? В корпоративных приложениях SQL является стандартом для работы с базами данных. SPARQL, напротив, требует сложных графовых структур и специализированного программного обеспечения и по трендам в Google не привлекает к себе интереса разработчиков.

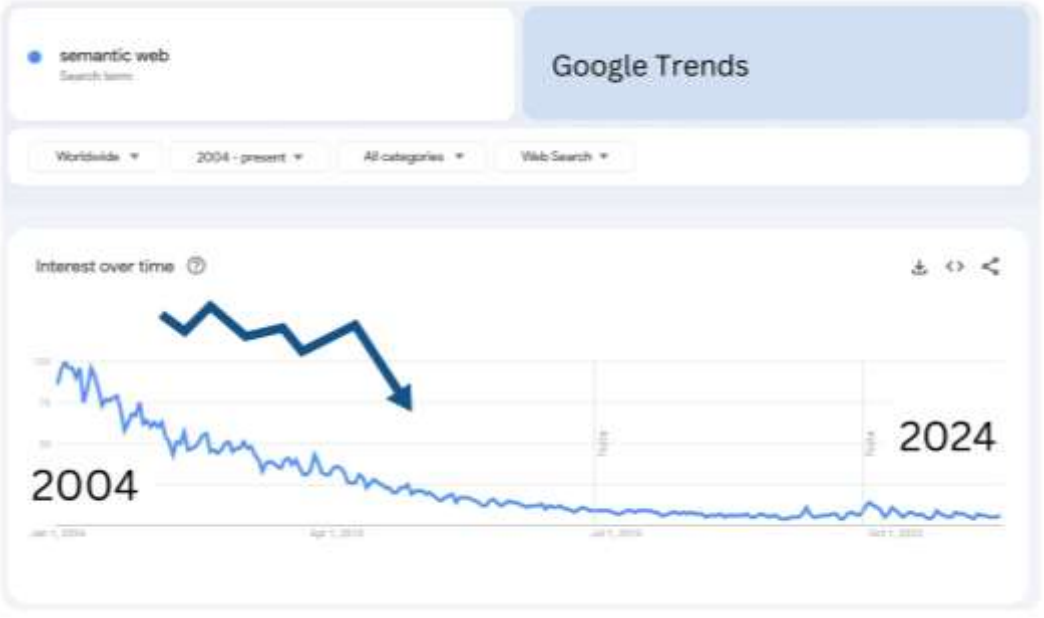


Рис. 6.1-9 Интерес к запросам «семантического интернета» согласно статистике Google.

Графовые базы данных и деревья классификаций могут быть полезны в отдельных случаях, но их применение не всегда оправдано для большинства повседневных задач. В итоге, создание графов знаний и использование технологий семантического веба имеет смысл только в тех случаях, когда требуется унифицировать данные из разных источников или реализовывать сложные логические выводы.

> Переход от таблиц к графовым моделям данных позволяет улучшить поиск и унифицировать поток информации, но не делает данные более осмысленными для машин. Вопрос не в том, нужно ли использовать семантические технологии, а в том, где они действительно приносят пользу. Прежде, чем внедрять онтологию, семантику и графовые базы данных в своей компании, уточните, какие компании уже успешно применяют эти технологии, а где они не оправдали ожиданий.

Несмотря на амбициозные ожидания, семантические технологии так и не стали универсальным решением для структурирования данных в строительной отрасли. На практике эти технологии не привели к универсальному решению, а лишь добавили новые сложности, и эти усилия повторяют нереализованные амбиции концепции семантического интернета, где ожидания значительно превзошли реальность.

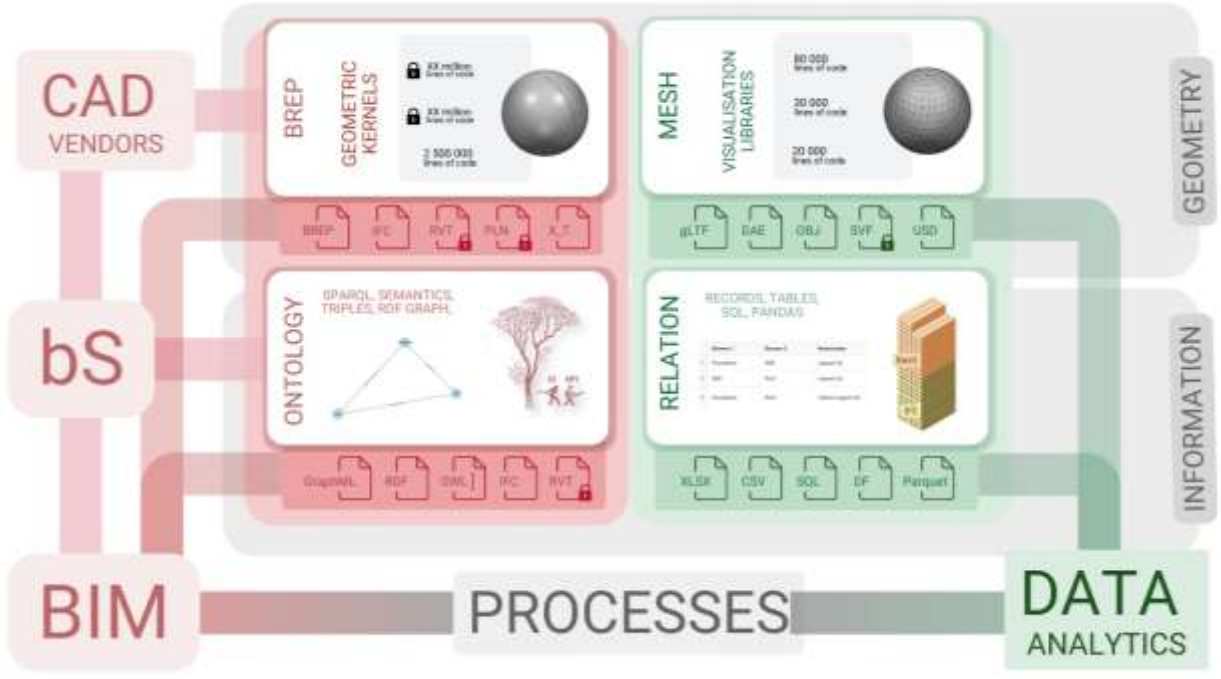
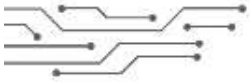


Рис. 6.1-10 Геометрия и информация в строительных процессах: от сложных CAD и BIM-систем до упрощенных данных для аналитики.

Если в сфере IT неудачи семантического веба были компенсированы появлением новых технологий (больших данных, IoT, машинного обучения, AR/VR), то строительная отрасль таких поводов не имеет.

Помимо проблем с использованием концептов для передачи взаимосвязей о данных между элементами проекта, сохраняется фундаментальная проблема – сама доступность этих данных. В строительной отрасли по-прежнему доминируют закрытые системы, затрудняющие работу с данными, обменом информацией и повышением эффективности процессов.

Именно закрытая природа данных становится одним из ключевых барьеров, который препятствует развитию цифровых решений в строительстве. В отличие от IT-индустрии, где открытые и унифицированные форматы данных стали стандартом, в секторе CAD (BIM) каждое программное обеспечение использует собственный формат, создавая замкнутые экосистемы и искусственно ограничивая пользователей.



ГЛАВА 6.2.
ЗАКРЫТЫЕ ФОРМАТЫ ПРОЕКТОВ И ПРОБЛЕМЫ ИНТЕРОПЕРАБЕЛЬНОСТИ

Закрытые данные и падающая продуктивность: тупик отрасли CAD (BIM)

Проприетарная природа CAD-систем привела к тому, что каждая программа имеет свой уникальный формат данных, который либо закрыт и недоступен извне - RVT, PLN, DWG, NDW, NWD, SKP, либо доступен в полуструктурированной форме через достаточно сложный процесс конвертации - JSON, XML (CPIXML), IFC, STEP и ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV и др..

Различные форматы данных, в которых могут храниться одни и те же данные об одних и тех же проектах, не только отличаются по структуре, но и включают в себя разные версии внутренней разметки, которые разработчикам необходимо учитывать для обеспечения совместимости приложений. Например CAD формат 2025 года откроется в CAD программе 2026 года, но этот же проект уже никогда не откроется во всех версиях CAD программы которые могли быть до 2025 года.

> Не предоставляя прямого доступа к базам данных, поставщик программного обеспечения в строительной отрасли часто создаёт свой собственный уникальный формат и инструменты к нему, которые специалист (инженер проектировщик или менеджер данных) должен использовать для доступа, импорта и экспорта данных.

Как следствие, поставщики базовых CAD (BIM) и связанных с ними решений (например ERP/PMIS) постоянно повышают цены на использование продуктов, а рядовые пользователи вынуждены платить "комиссию" на каждом этапе передачи данных форматами [63]: за подключение, импорт, экспорт и работу с данными, которые пользователи создали сами.

Стоимость доступа к данным в облачном хранилище из популярных CAD- (BIM-) продуктов в 2025 году достигнет 1 доллара за транзакцию [120], а подписки на строительные ERP-продукты для средних компаний достигают пяти- и шестизначных сумм в год [121].

> Суть современного строительного программного обеспечения заключается в том, что не автоматизация или повышение эффективности, а умение инженеров разбираться в конкретном узкоспециализированном программном обеспечении влияет на качество и стоимость обработки данных строительного проекта, а также на прибыль и долгосрочное выживание компаний, реализующих строительные проекты.

Отсутствие доступа к базам данных CAD-систем, которые используются в десятках других систем и сотнях процессов [63], и, как следствие, отсутствие качественной коммуникации между отдельными специалистами привело строительную отрасль к статусу одного из самых неэффективных секторов экономики с точки зрения производительности [44].

За последние 20 лет применения CAD- (BIM-) проектирования, появления новых систем (ERP), новых строительных технологий и материалов, производительность всей строительной отрасли

упала на 20% (Рис. 2.2-1), в то время как общая производительность всех секторов экономики, у которых нет больших проблем доступа к базам данных и подобных маркетинговому BIM-концептов, выросла на 70% (96% в обрабатывающей промышленности) [122].

Рис. 6.2-1 Из-за изолированности и сложности проектных данных, от которых зависят десятки отделов и сотни процессов в строительной отрасли, скорость принятия решений в несколько раз ниже, чем в других отраслях.

Однако есть и единичные примеры альтернативных подходов создания интероперабельности между CAD-решениями. Крупнейшая строительная компания Европы с проектом SCOPE [123], начатым ещё в 2018 году, демонстрирует, как можно выйти за рамки классической логики CAD- (BIM-) систем. Вместо того, чтобы пытаться подчинить IFC или полагаться на проприетарные геометрические ядра, разработчики SCOPE используют API и SDK обратного инжиниринга для извлечения данных из различных CAD-программ, преобразуют их в нейтральные форматы, такие как OBJ или CPIXML на базе единственного Open Source геометрического ядра OCCT, и далее применяют их в сотнях бизнес-процессов строительных и проектных компаний. Тем не менее, несмотря на прогрессивность идеи, подобные проекты сталкиваются с ограничениями и сложностью бесплатных геометрических ядер и они всё равно остаются частью замкнутых экосистем одной компании, которые воспроизводят логику моновендорных решений.

Из-за ограничений закрытых систем и различий в форматах данных, а также отсутствия эффективных инструментов для их унификации, компании, которым приходится работать с CAD форматами, сталкиваются с накоплением значительных объемов данных разной степени структурированности и закрытости. Эти данные не используются должным образом и пропадают в архивах, где со временем остаются навсегда забытыми и неиспользуемыми.

> Данные, полученные путем значительных усилий на этапе проектирования, из-за своей сложности и закрытости становятся недоступными для дальнейшего использования.

В результате на протяжении последних 30 лет разработчики в строительной отрасли вынуждены вновь и вновь сталкиваться с одной и той же проблемой: каждый новый закрытый формат или проприетарное решение порождает необходимость интеграции с существующими открытыми и закрытыми CAD-системами. Эти постоянные попытки обеспечить интероперабельность между различными CAD- и BIM-решениями лишь усложняют экосистему данных, вместо того чтобы способствовать её упрощению и стандартизации.

Миф об интероперабельности между CAD-системами

Если в середине 1990-х годов ключевым направлением развития интероперабельности в CAD-среде был взлом проприетарного формата DWG — завершившийся победой альянса Open DWG [75] и фактическим открытием самого популярного формата чертежей для всей строительной отрасли, — то к середине 2020-х акцент сместился. В строительной индустрии набирает силу новая тенденция: многочисленные команды разработчиков сосредоточены на создании так называемых «мостов» между закрытыми CAD-системами (closed BIM), форматом IFC и открытыми решениями (open BIM). В основе большинства таких инициатив лежит использование формата IFC и геометрического ядра OCCT, обеспечивающих техническую связку между разрозненными платформами. Этот подход рассматривается как перспективное направление, способное существенно улучшить обмен данными и повысить совместимость программных инструментов.

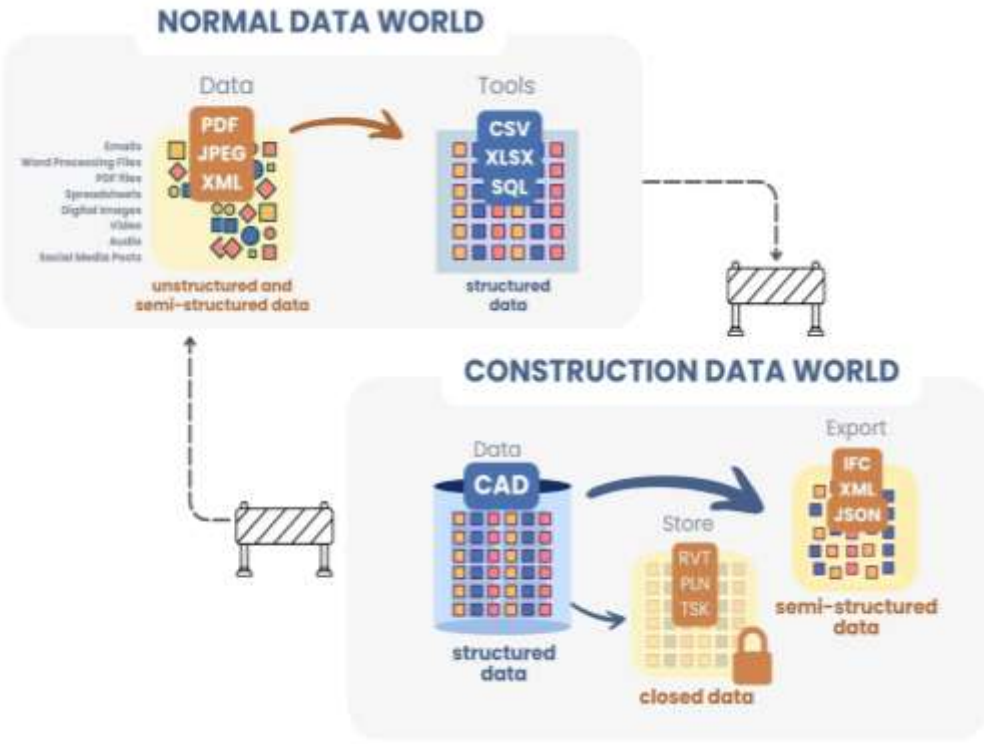


Рис. 6.2-2 В то время как другие отрасли работают с открытыми данными, строительная отрасль вынуждена работать с закрытыми или слабоструктурированными форматами CAD (BIM).

Подобный подход имеет исторические параллели. В 2000-х годах разработчики, стремясь преодолеть доминирование крупнейшего вендора графических редакторов (мира 2D), пытались создать бесшовную интеграцию между его проприетарным решением и бесплатной Open Source-альтернативой GIMP (Рис. 6.2-3). Тогда, как и сегодня в строительстве, речь шла о попытке соединить закрытые и открытые системы, сохраняя при этом сложные параметры, слои и внутреннюю логику работы ПО.

Однако пользователи в действительности искали простые решения — плоские, открытые данные без излишней сложности слоёв и параметров программ (аналогов геометрического ядра в CAD). Пользователи стремились к простым и открытым форматам данных, свободным от избыточной логики. В графике такими стали JPEG, PNG и GIF. Сегодня их используют в социальных сетях, на сайтах, в приложениях — они легко обрабатываются и интерпретируются, независимо от платформы или производителя ПО.

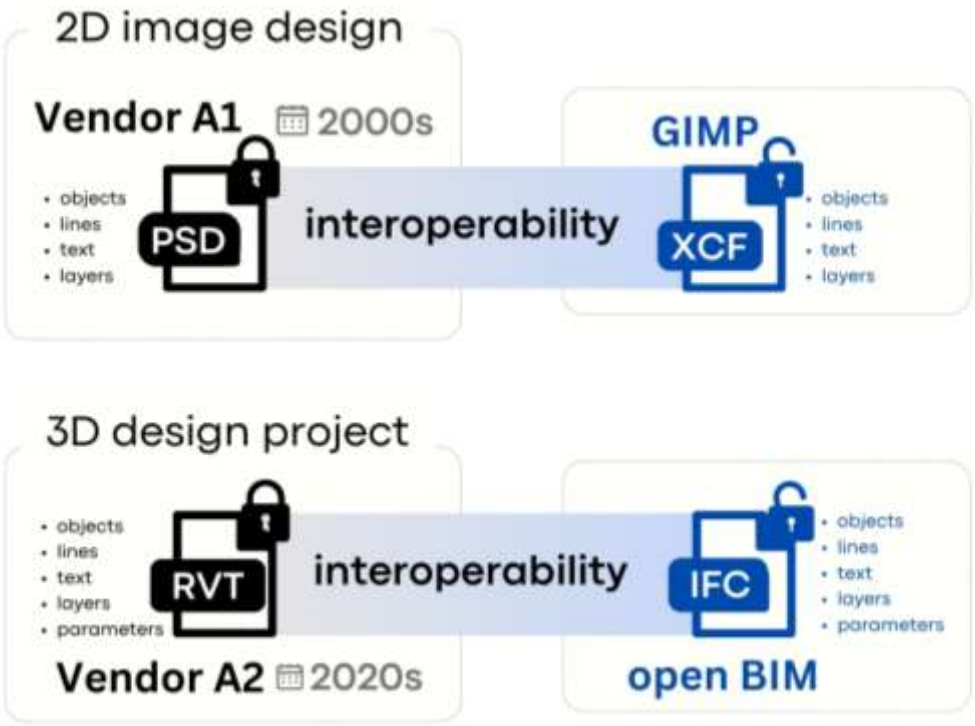


Рис. 6.2-3 Взаимозаменяемость форматов данных в строительстве похожа на путь от попыток объединить популярный проприетарный продукт вендора и Open Source GIMP в 2000-х годах.

В результате в индустрии изображений сегодня почти никто не использует закрытые форматы вроде PSD или открытые XCF для приложений, социальных сетей вроде Facebook и Instagram или в качестве контента на сайтах. Вместо этого в большинстве задач применяются плоские и открытые форматы JPEG, PNG и GIF, которые обеспечивают простоту использования и широкую совместимость. Открытые форматы, такие как JPEG и PNG, стали стандартом для обмена изображениями благодаря их универсальности и широкой поддержке, упрощая их использование на разнообраз-

ных платформах. Аналогичный переход наблюдается и в других форматах обмена, например в видео и аудио, где универсальные форматы, как MPEG и MP3, выделяются за эффективность сжатия и широкую совместимость. Подобный переход к стандартизации упростил обмен и воспроизведение контента и информации, сделав их доступными для всех пользователей на различных платформах (Рис. 6.2-4).

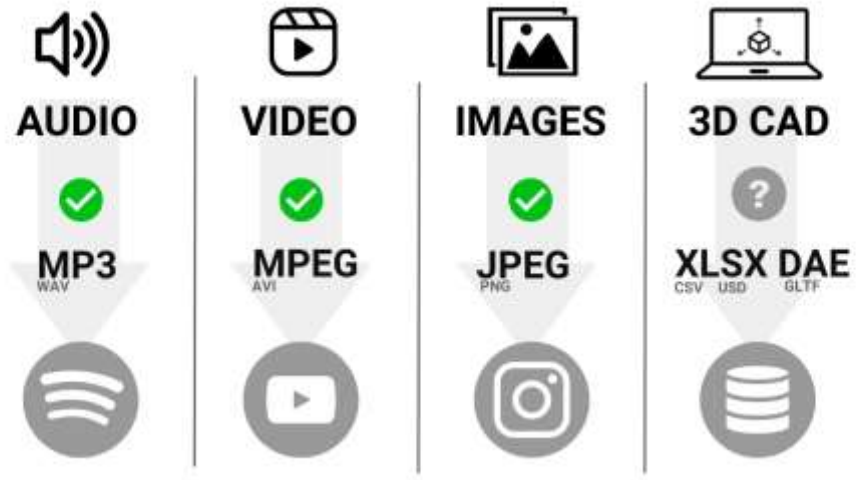


Рис. 6.2-4 Упрощенные форматы без сложных функций редактирования, стали популярными для обмена и использования данных.

Аналогичные процессы происходят и в 3D-моделировании. Простые и открытые форматы вроде USD, OBJ, glTF, DAE, DXF, SQL и XLSX всё чаще используются в проектах для обмена данными вне среды CAD (BIM). Эти форматы хранят всю необходимую информацию, включая геометрию и метаданные, без необходимости оперировать сложной структурой BREP, геометрическими ядрами или внутренними классификаторами конкретных вендоров. Проприетарные форматы, такие как NWC, SVF, SVF2, CPIXML и CP2, предоставляемые ведущими вендорами ПО, также выполняют схожие функции, но при этом остаются закрытыми, в отличие от открытых стандартов.

> Примечательно (и стоит ещё раз напомнить, как уже отмечалось в предыдущей главе), что подобную идею — отказ от промежуточных нейтральных и параметрических форматов вроде IGES, STEP и IFC — ещё в 2000 году поддержал главный CAD-вендор, создавший Whitepaper BIM и зарегистрировавший формат IFC в 1994 году. В Whitepaper 2000 года «Интегрированное проектировании и производство» [65] CAD-вендор подчёркивает важность нативного доступа к базе CAD-данных внутри программной среды, без необходимости использовать промежуточные трансляторы и параметрические форматы, чтобы сохранить полноту и точность информации.

Строительной отрасли еще предстоит договориться или об инструментах доступа к базам данных CAD, или их принудительного обратного инжиниринга, или о принятии общего упрощенного формата данных для использования вне платформ CAD (BIM). Например, многие крупные компании в Центральной Европе и немецкоязычных регионах, работающие в строительном секторе, используют формат CPIXML в своих ERP-системах [121]. Этот проприетарный формат, представляющий

собой разновидность XML, объединяет данные проекта CAD (BIM), включая геометрические и метаданные, в единую организованную упрощённую структуру. Также крупные строительные компании создают новые собственные форматы и системы, как в проекте SCOPE, про который мы говорили в предыдущей главе.

Закрытая логика параметрических CAD-форматов или сложные параметрические файлы IFC (STEP) в большинстве бизнес процессов оказываются избыточными. Пользователи ищут упрощённые и плоские форматы, такие как USD, CPIXML, XML&OBJ, DXF, glTF, SQLlite, DAE&XLSX, которые содержат всю необходимую информацию об элементах, но при этом не обременены избыточной логикой построения геометрии BREP, зависимостью от геометрических ядер и внутренними классификациями конкретных CAD и BIM-продуктов (Рис. 6.2-5).

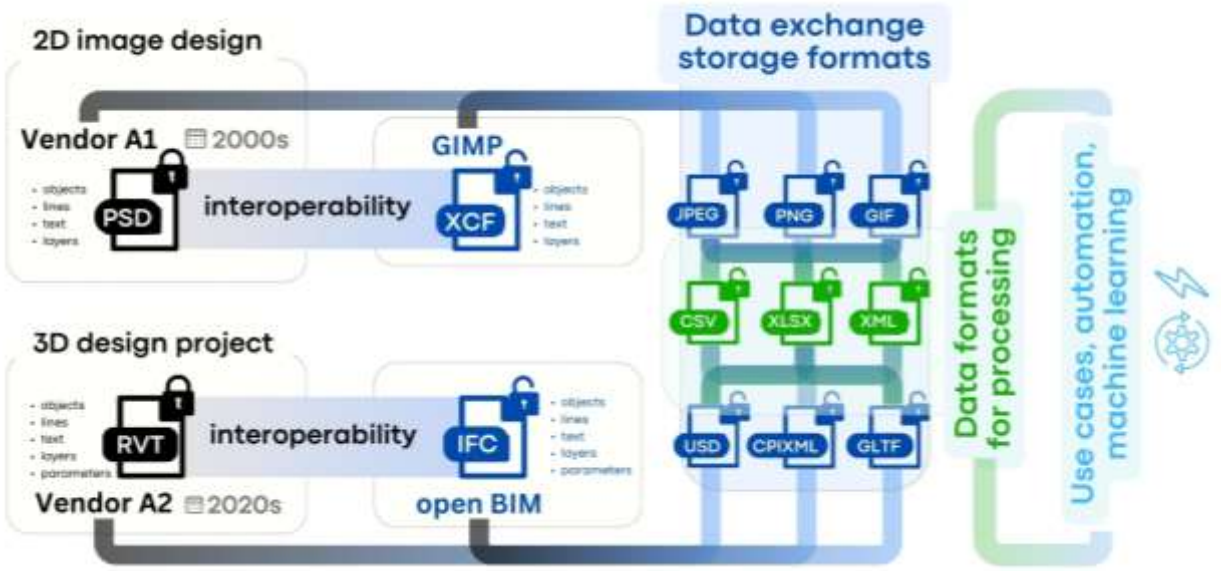


Рис. 6.2-5 Для большинства кейсов использования пользователи выбирают максимально простые форматы, которые не зависят от программ вендоров.

Появление плоских форматов изображений, таких как JPEG, PNG и GIF, свободных от избыточной логики внутренних движков вендоров, способствовало развитию тысяч совместимых решений для обработки и использования графики. Это привело к появлению разнообразных приложений: от инструментов ретуши и фильтрации до социальных сетей, таких как Instagram, Snapchat и Canva, где эти упрощенные данные можно было использовать без привязки к конкретному разработчику ПО.

> Стандартизация и упрощение проектных CAD-форматов будут стимулировать появление множества новых удобных и независимых инструментов для работы со строительными проектами.

Отказ от сложной логики приложений вендоров, завязанных на закрытые геометрические ядра, и переход к универсальным открытым форматам, основанным на библиотеках упрощённых элементов, создаёт предпосылки для более гибкой, прозрачной и эффективной работы с данными.

Это также открывает доступ к информации для всех участников строительного процесса — от проектировщиков до заказчиков и эксплуатационных служб.

Тем не менее, с высокой вероятностью, в ближайшие годы CAD-вендоры предпримут попытки вновь сместить акценты в дискуссии об интероперабельности и доступе к базам данных CAD. Речь пойдёт уже о "новых" концепциях — таких как гранулированные данные, интеллектуальные графы, "федеративные модели", цифровые двойники в облачных репозиториях, — а также о создании отраслевых альянсов и стандартов, продолжающих путь BIM и open BIM. Несмотря на привлекательную терминологию, подобные инициативы могут вновь стать инструментами удержания пользователей в рамках проприетарных экосистем. Одним из примеров является активное продвижение с 2023 года формата USD (Universal Scene Description) как "нового стандарта" межплатформенного взаимодействия в CAD (BIM).

Переход к USD и гранулированным данным

Появление в 2023 году альянса AOUSD [124] знаменует важный поворот в строительной отрасли. Мы наблюдаем начало новой реальности, формируемой CAD-вендорами, в работе со строительными данными через несколько значимых изменений. Первое серьезное изменение касается восприятия CAD-данных. Специалисты, участвующие в ранних этапах концептуального проектирования, всё чаще осознают, что создание проекта в CAD-среде — лишь отправная точка. Данные, формируемые в процессе проектирования, со временем становятся основой для анализа, эксплуатации и управления объектами. Это означает, что они должны быть доступны и пригодны для использования в системах, выходящих за пределы традиционных CAD-инструментов.

Параллельно с этим происходит революция в подходах ведущих разработчиков. Ведущий CAD-вендор отрасли, создавший концепт BIM и формат IFC, совершает неожиданный поворот в своей стратегии. С 2023 года компания отходит от традиционного хранения данных в отдельных файлах, делая ставку на работу с гранулированными (нормализованными и структурированными) данными с переходом в дата-центричный подход [125].

> Вендоры следуют историческим тенденциям других отраслей: большинству пользователей не нужны закрытые CAD форматы (похожие на PSD) или сложные параметрические файлы IFC (похожий на GIMP с логикой слоёв). Им требуются простые изображения объектов, которые можно использовать в CAFM (строительном Instagram), ERP (Facebook) и в тысячах других процессах, наполненных Excel таблицами и PDF документами.

Текущие тенденции в строительной отрасли потенциально создают условия для постепенного ухода от параметрических и сложных форматов в пользу более универсальных и независимых форматов USD, GLTF, DAE, OBJ (с метаинформацией как внутри гибридных, так в отдельных структурированных или слабоструктурированных форматах). Исторические лидеры, включая крупнейшие проектировочные компании, которые когда-то активно в середине 1990х продвигали IFC, сегодня открыто продвигают новый формат USD [93], подчеркивая его простоту и универсальность (Рис. 6.2-6). Массовое внедрение USD в продукты, совместимость с GLTF и активная интеграция в инструменты, такие как Blender, Unreal Engine и Omniverse, свидетельствуют о потенциале начала но-

вой парадигмы работы с данными. Наряду с популярностью локальных решений, таких как европейский плоский USD формат — CPIXML, используемый в популярных европейских ERP может потенциально усилить позиции USD в Центральной Европе. Организации, занимающиеся развитием формата IFC уже адаптируют свою стратегию под USD [126], что лишь подтверждает неизбежность сдвига.

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

Рис. 6.2-6 Сравнение технических спецификаций форматов IFC и USD.

На этом фоне USD потенциально может стать стандартом де-факто, обещая преодолеть множество текущих ограничений, связанных в первую очередь со сложностью существующих CAD- (BIM-) форматов и зависимостью их интерпретации от геометрических ядер.

Вместо параметрических и сложных CAD-форматов и IFC - упрощённые форматы данных USD, gLTF, DAE, OBJ с метаинформацией элементов в CSV, XLSX, JSON, XML будут завоевывать место в строительной отрасли благодаря своей простоте и гибкости.

Текущие изменения в строительной отрасли на первый взгляд выглядят как технологический прорыв, связанный с переходом от устаревающего IFC к более современному USD. Однако стоит учитывать, что еще в 2000 году тот же CAD-вендор, который разработал IFC, сам писал о его проблемах и необходимости доступа к базе данных [65], а теперь активно продвигает переход на новый стандарт - USD.

> За очередным фасадом «открытых данных» USD и «новых» концептов по управлению гранулированными данными, через облачные приложения, которые начинают продвигать CAD вендоры может скрываться намерение вендоров монополизировать управление проектными данными, где пользователи оказываются в положении, где выбор формата больше связан с корпоративными интересами, чем с реальными потребностями.

Анализ ключевых фактов [93] показывает, что главная цель этих изменений — скорее не столько удобство пользователей, сколько в первую очередь сохранение контроля над экосистемами и потоками данных в интересах вендоров, которые за 40 лет так и не смогли предоставить доступ к базам данных CAD.

Возможно, именно теперь, компаниям пришло время отказаться от ожидания новых концепций от вендоров ПО и сосредоточиться на самостоятельном развитии в дата-центричном направлении. Освободившись от проблем с доступом к данным, через инструменты обратного инжиниринга, отрасль сможет без навязываемых новых концептов, перейти самостоятельно к современным, бесплатным и удобным инструментам для работы и анализа данных.

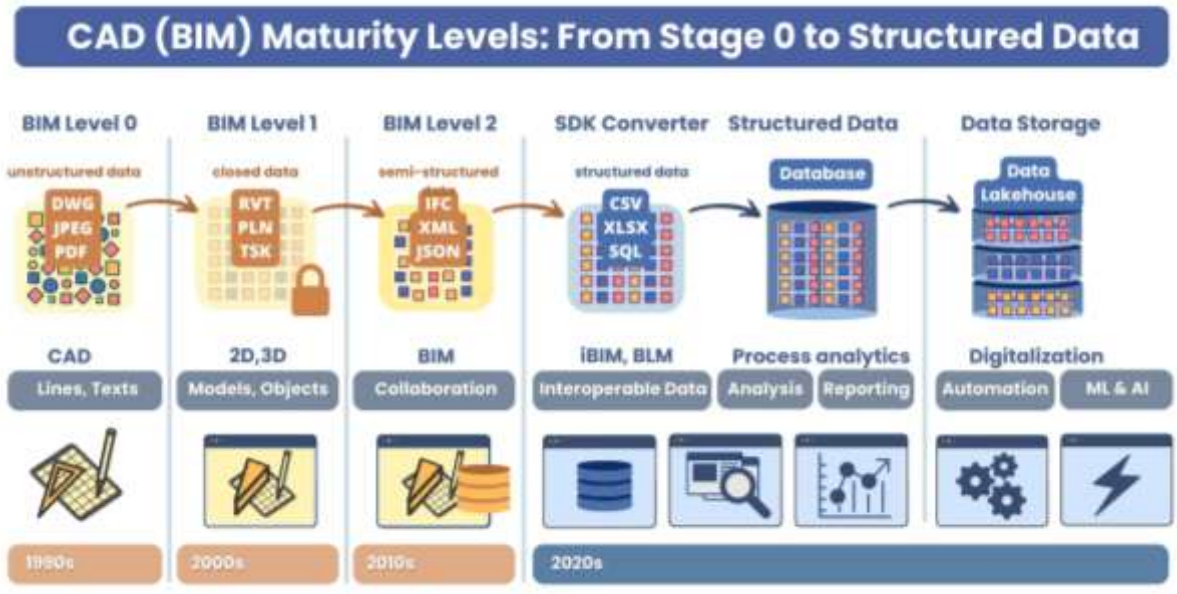


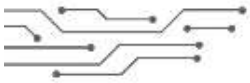
Рис. 6.2-7 Уровень зрелости CAD (BIM): от неструктурированных данных до структурированных данных и хранилищ.

Доступ к базам данных, открытые данные и форматы неизбежно станут стандартом в строительной отрасли, независимо от попыток вендоров затормозить этот процесс — это лишь вопрос времени (Рис. 6.2-7). Темпы этого перехода могут значительно возрасти, если всё больше специалистов будут знакомиться с открытыми форматами, инструментами для работы с базами данных и доступными SDK обратного инжиниринга, позволяющими организовать прямой доступ к данным CAD-систем [92].

> Будущее — за открытыми, унифицированными и доступными для аналитики данными. Чтобы избежать зависимости от решений вендоров и не оказаться заложниками закрытых экосистем, строительным и проектировочным компаниям рано или поздно придётся делать ставку на открытость и независимость, выбирая форматы и решения, которые обеспечивают полный контроль над данными.

Данные, которые создаются сегодня в строительной отрасли, станут ключевым ресурсом для принятия бизнес-решений в будущем. Они будут выступать в роли стратегического "топлива", питающего развитие и эффективность строительных компаний. Будущее строительной отрасли — в умении работать с данными, а не в выборе форматов или модели данных.

Чтобы понять разницу между открытыми форматами USD, glTF, DAE, OBJ и проприетарными параметрическими CAD форматами, важно рассмотреть один из самых сложных и ключевых элементов данных в визуализации и расчётах проектов — геометрию и процессы ее формирования. И чтобы разобраться, как геометрические данные становятся основой для аналитики и расчетов в строительстве, необходимо глубже изучить механизмы генерации геометрии, её преобразования и хранения.



ГЛАВА 6.3.
ГЕОМЕТРИЯ В СТРОИТЕЛЬСТВЕ: ОТ ЛИНИЙ К КУБОМЕТРАМ

Когда линии превращаются в деньги или зачем строителям геометрия

Геометрия в строительстве — это не только визуализация, но и основа для точных количественных расчётов. В модели проекта геометрия, дополняет списки параметров элементов (Рис. 3.1-16) важнейшими объёмными характеристиками, например такими как длина, площадь и объём. Эти значения объёмных параметров вычисляются автоматически с помощью геометрических ядер и являются отправной точкой для смет, графиков и ресурсных моделей. Как мы уже обсуждали в пятой части книги и в главе «Расчёты стоимости и сметы строительных проектов», именно объёмные параметры групп объектов из CAD-моделей формируют основу для современных ERP, PMIS-систем/ Геометрия играет фундаментальную роль не только на стадии проектирования, но и в управлении реализацией проекта, контроле сроков, бюджетировании и эксплуатации. Как тысячи лет назад при строительстве египетских пирамид точность проекта зависела от меры длины вроде локтя и кубита, так и сегодня точность интерпретации геометрии в CAD-программах напрямую влияет на результат: от бюджета и сроков — до выбора подрядчиков и логистики поставок.

> В условиях высокой конкуренции и ограниченного бюджета точность объёмных расчётов, напрямую зависящая от геометрии, становится фактором выживания. Современные ERP-системы напрямую зависят от корректных объёмных характеристик, получаемых из CAD- и BIM-моделей. Именно поэтому точное геометрическое описание элементов — это не просто визуализация, а ключевой инструмент управления стоимостью и сроками строительства.

Исторически геометрия была основным языком инженерного взаимодействия. От линий на папирусе до цифровых моделей — чертежи и геометрические представления служили средством обмена информацией между проектировщиками, прорабами и сметчиками. До появления компьютеров расчёты выполнялись вручную, с помощью линеек и транспортира. Сегодня эта задача автоматизирована благодаря объёмному моделированию: геометрические ядра CAD-программ преобразуют линии и точки в трёхмерные тела, из которых автоматически извлекаются все необходимые характеристики.

Работая в CAD-программах, создание геометрических элементов для расчётов происходит через пользовательский интерфейс CAD- (BIM-) программ. Для преобразования точек и линий в объемные тела используется геометрическое ядро, которое выполняет ключевую задачу — преобразование геометрии в объемные модели, из которых после аппроксимации автоматически рассчитываются объемные характеристики элемента.

От линий до объёмов: как площадь и объём становятся данными

В инженерной практике объёмы и площади вычисляются на основе геометрических поверхностей,

описанных аналитически или через параметрические модели, такие как NURBS (неоднородные рациональные B-сплайны) в рамках BREP (представление граничных элементов).

> NURBS (Non-Uniform Rational B-Splines) — это математический способ описания кривых и поверхностей, тогда как BREP — это структура для описания полной трёхмерной геометрии объекта, включая его границы, которые могут быть определены с использованием NURBS.

Несмотря на точность BREP и NURBS, они требуют мощных вычислительных ресурсов и сложных алгоритмов. Однако прямые вычисления по таким математически точным описаниям часто вычислительно сложны, поэтому на практике почти всегда используется тесселяция — преобразование поверхностей в сетку треугольников, что упрощает последующие расчеты. Тесселяция — это разбитие сложной поверхности на треугольники или полигоны. В CAD/CAE-среде этот метод используется для визуализации, расчётов объёмов, поиска коллизий, экспорта в форматы вроде MESH и анализа столкновений. Пример из природы — пчелиные соты, где сложная форма разбивается на регулярную сетку (Рис. 6.3-1).

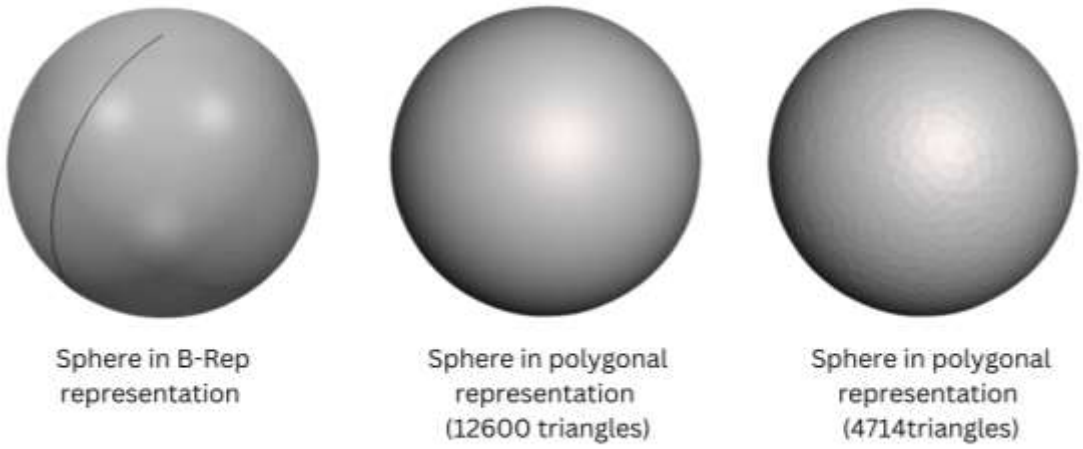


Рис. 6.3-1 Одна и таже сфера в параметрическом описании BREP и полигональном представлении с различным количество треугольников.

BREP (NURBS), применяемая в CAD, не является фундаментальной моделью геометрии. Этот метод был создан как удобный инструмент для представления окружностей и рациональных сплайнов и для минимизации хранения данных о геометрии. Однако у него есть ограничения — например, невозможность точно описать синусоиду, которая лежит в основе винтовых линий и поверхностей, а также необходимость использования сложных геометрических ядер.

Треугольные сетки и тесселяция параметрических фигур, напротив, отличаются простотой, эффективным использованием памяти и способностью обрабатывать большие объёмы данных (Рис. 6.3-2). Эти преимущества позволяют обходиться без сложных и дорогих геометрических ядер, и заложенных в них десятков миллионов строк кода, при расчётах геометрических форм.

> В большинстве строительных кейсов не имеет значения, как именно определяются объёмные характеристики — через параметрические модели (BREP, IFC) или через полигоны (USD, glTF, DAE, OBJ). Геометрия остаётся формой аппроксимации: будь то через NURBS или MESH, это всегда приближённое описание формы.

Геометрия, заданная в виде полигонов или BREP (NURBS), остаётся в какой-то степени лишь способом аппроксимации с приближённым описанием непрерывной формы. Подобно тому, как интегралы Френеля не имеют точного аналитического выражения, дискретизация геометрии через полигоны или NURBS всегда является аппроксимацией, такой же, как и триангулярный MESH.

Параметрическая геометрия в формате BREP необходима в основном там, где важен минимальный размер данных и есть возможность использовать ресурсоёмкие и дорогие геометрические ядра для её обработки и отображения. Чаще всего это характерно для разработчиков CAD-программ, которые для этого применяют в своих продуктах геометрические ядра MCAD-вендоров. При этом, даже внутри этих программ, BREP-модели в процессе тесселяции для визуализации и расчётов часто преобразуются в треугольники (аналогично тому, как PSD-файлы упрощается в JPEG).

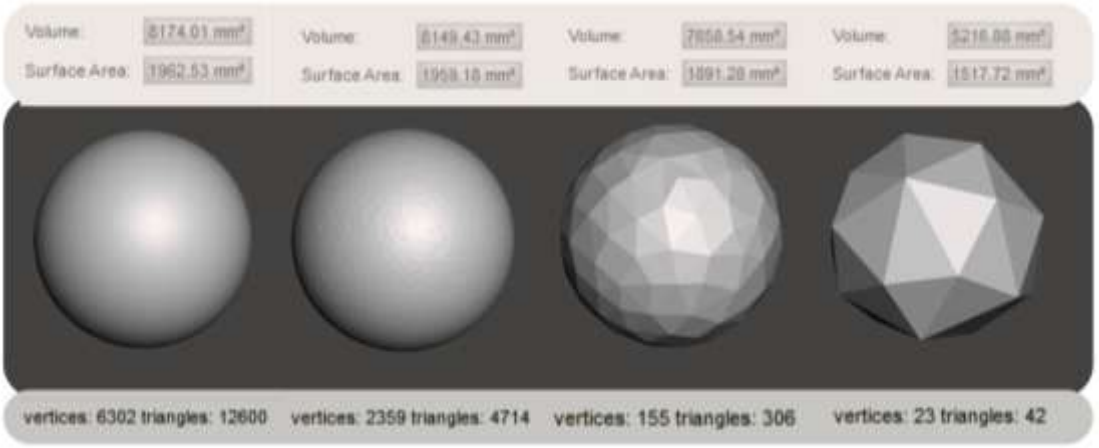


Рис. 6.3-2 Различие объёмных характеристик у фигур с разным количеством полигонов.

Полигональный MESH, как и параметрический BREP, имеют свои преимущества и ограничения, но цель у них одна —описывать геометрию с учётом задач пользователя. В конечном счёте точность геометрической модели зависит не только от метода её представления, но и от требований к конкретной задаче.

> В большинстве строительных задач потребность в параметрической геометрии и сложных геометрических ядрах может быть избыточна.

В каждой конкретной задаче по автоматизации расчётов стоит рассматривать не преувеличивается ли значимость параметрической геометрии разработчиками CAD, которые заинтересованы в продвижении и продаже собственных программных продуктов.

Переход к MESH, USD и полигонам: использование тесселяции для геометрии

В строительной отрасли при потоковой работе, разработке систем, баз данных или автоматизации процессов, работающих с проектной информацией и геометрией элементов, важно стремиться к независимости от конкретных CAD-редакторов и геометрических ядер.

> В основе обменного формата, который будет использоваться как в отделах калькуляций, так и на стройплощадке - не должна подразумеваться конкретная CAD- (BIM-) программа. Геометрическая информация должна быть представлена в формате напрямую через тесселяцию, без привязки к геометрическому ядру или архитектуре CAD.

Параметрическая геометрия из CAD может рассматриваться как промежуточный источник, но не как основа универсального формата. Большинство параметрических описаний (включая BREP и NURBS) в любом случае преобразуются в полигональный MESH для последующей обработки. Если результат — один и тот же (тесселяция и полигоны), а процесс — проще, то выбор очевиден. Это аналогично выбору между графовой онтологией и структурированными таблицами (про который мы говорили в четвёртой части): избыточная сложность редко оправдана (Рис. 3.2-10, Рис. 6.1-8).

Такие открытые форматы, как: OBJ, STL, glTF, SVF, CPIXML, USD и DAE, используют универсальную структуру треугольных сеток, что дает им существенные преимущества. Эти форматы обладают отличной совместимостью — их легко читать и визуализировать с помощью доступных открытых библиотек без необходимости в сложных специализированных геометрических ядрах, содержащих миллионы строк кода (Рис. 6.3-3). Эти универсальные геометрические форматы используются в различных областях — от относительно простых инструментов для проектирования кухонь в IKEA™ до сложных систем визуализации объектов в кино и VR-приложениях. Важное преимущество заключается в наличии большого количества бесплатных и открытых библиотек для работы с этими форматами, доступных для большинства платформ и языков программирования.

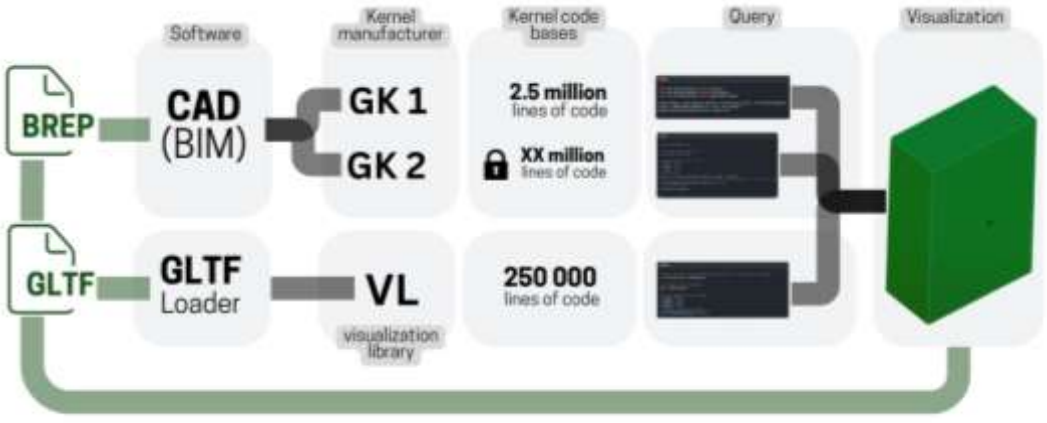


Рис. 6.3-3 Одно и то же представление геометрии достигается через использование параметрических форматов и геометрических ядер, либо при помощи триангулированных форматов и открытых библиотек визуализации.

Так же как и сами пользователи, CAD-вендоры сталкиваются с проблемами интерпретации чужих параметрических CAD форматов или открытого IFC из-за различия в геометрических ядрах. На практике все CAD-вендоры, без исключения, используют SDK обратного инжиниринга для передачи данных между системами, и никто из них не полагаются на форматы вроде IFC или USD [93] для целей интероперабельности.

> Вместо того, чтобы использовать концепты, продвигаемые альянсами CAD-вендоров, которыми они сами не пользуются, - разработчикам и пользователям CAD решений продуктивнее сосредоточиться на понимании преимуществ каждого подхода в конкретном контексте и выбирать тот или иной вид геометрии в зависимости от кейса использования. Выбор между различными геометрическими представлениями — это компромисс между точностью, вычислительной эффективностью и практическими потребностями конкретной задачи.

Сложность, связанная с использованием геометрических ядер, которую традиционно навязывают строительной отрасли крупные вендоры при обработке проектных данных, зачастую оказывается избыточной. Формат USD, основанный на MESH-геометрии, может стать своего рода «ящиком Пандоры» для отрасли, открывая разработчикам новые возможности организации обмена данными — вне рамок IFC и параметрических BREP-структур, типичных для CAD-вендоров.

Ближе познакомившись со структурой USD, DAE, gLTF, OBJ и др., становится очевидно, что существуют и более простые, открытые форматы, позволяющие эффективно организовать передачу и использование геометрической информации без необходимости опираться на сложную параметрику и закрытые геометрические ядра. Такой подход не только снижает технический порог входа для разработчиков, но и способствует развитию гибких, масштабируемых и действительно открытых решений для цифрового строительства.

LOD, LOI, LOMD – уникальная классификация детализации в CAD (BIM)

В дополнение к геометрическим форматам представления, в мире, где различные отрасли используют разные уровни детализации и глубины данных, CAD- (BIM-) методологии предлагают свои уникальные системы классификации, которые структурируют подход к информационному наполнению моделей зданий.

Одним из примеров новых подходов к стандартизации является введение уровней разработки модели, отражающих степень готовности и достоверности как графической, так и информационной составляющей. Для разграничений информационной наполненности в работе с CAD- (BIM-) данными появился LOD (Level Of Detail) – уровень детализации графической части модели, и LOI (Level Of Information) – уровень проработки данных. Дополнительно для комплексного подхода было введено понятие LOA (Level of Accuracy) – точность представленных элементов и LOG (Level of Geometry) для определения точности графического представления.

Уровни детализации (LOD) обозначаются числами от 100 до 500, отражая степень проработки модели. LOD 100 представляет собой концептуальную модель с общими формами и размерами. LOD 200 включает более точные размеры и формы, но с условной детализацией. LOD 300 – это подроб-

ная модель с точными размерами, формами и расположением элементов. LOD 400 содержит детальную информацию, необходимую для изготовления и монтажа элементов. LOD 500 отражает фактическое состояние объекта после строительства и используется для эксплуатации и обслуживания. Эти уровни описывают структуру насыщения CAD (BIM) модели информацией на разных стадиях жизненного цикла, включая 3D, 4D, 5D и далее.

В реальных проектах высокая детализация (LOD400) часто оказывается избыточной и достаточно использовать геометрию LOD100 или даже плоские чертежи, а остальные данные можно получить или расчетным путем или из связанных элементов, которые могут не иметь выраженной геометрии. Например, пространства и элементы помещений (категории элементов «Помещения») могут не иметь визуальной геометрии, но при этом содержать значительные объемы информации и баз данных, вокруг которых строятся многие бизнес-процессы.

> Поэтому перед началом проектирования важно чётко определить требуемый уровень детализации. Для кейсов использования 4D-7D зачастую достаточно даже чертежей DWG и минимальной геометрии LOD100. Ключевая задача в процессе работы с требованиями — найти баланс между насыщенностью и практической применимостью модели.

В сущности, если рассматривать данные CAD (BIM) как базу данных (чем она и является), описание насыщенности модели через новые аббревиатуры представляет собой не что иное, как поэтапное моделирование данных для информационных систем, начиная от концептуального уровня и заканчивая физическим (Рис. 6.3-4), что подробно рассматривалось в третьей и четвёртой части книги. Каждое увеличение LOD и LOI означает добавление информации, нужной для новых задач: расчётов, управления строительством, эксплуатации и характеризуется последовательным обогащением модели дополнительными информационными слоями (3D-8D) в виде различных параметров, про которые мы говорили в пятой части книги.

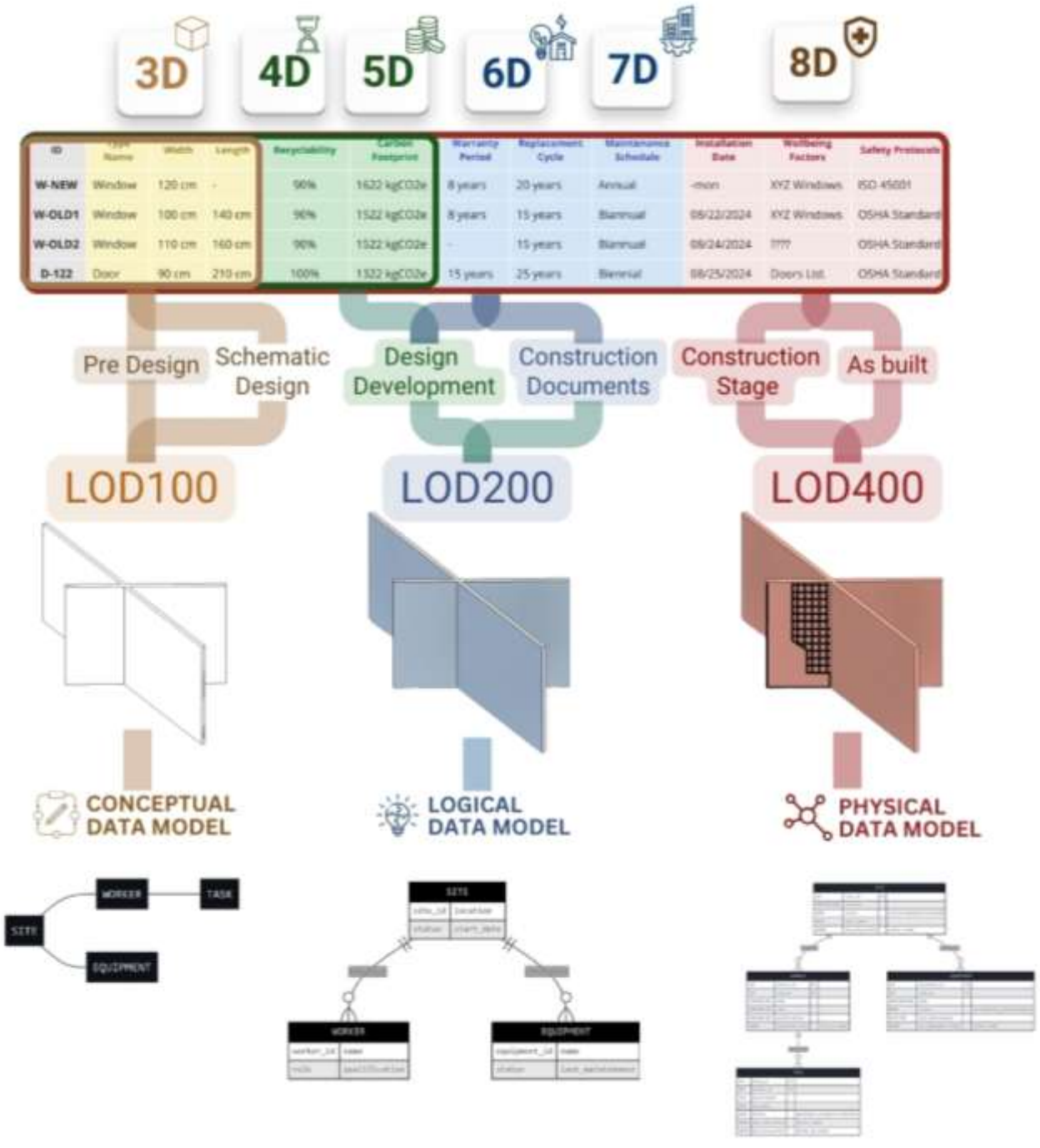


Рис. 6.3-4 Процесс информационного наполнения детализацией проекта идентична моделированию данных от концептуальной к физической модели данных.

Геометрия — это лишь некоторая часть проектных данных, необходимость которой не всегда обоснована в строительных проектах и ключевой вопрос работы с CAD-данными заключается не столько в том, как визуализируются модели, а больше в том, как данные из этих моделей могут быть использованы за пределами CAD- (BIM-) программ.

К середине 2000-х годов строительная отрасль столкнулась с беспрецедентной проблемой, связанной с быстрым увеличением объема данных в системах управления и обработки данных, особенно тех, которые поступали из отделов CAD (BIM). Такой резкий рост объема данных застал менеджеров компаний врасплох, и они оказались не готовы к растущим требованиям к качеству данных и управлению ими.

Новые стандарты CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Пользуясь отсутствием открытого доступа к базам данных CAD и ограниченной конкуренцией на рынке обработки данных, а также используя маркетинговые кампании, связанные с новой аббревиатурой BIM, организации, занимающиеся развитием подходов по работе с CAD-данными - начали создавать новые стандарты и концепции, которые де-юре должны быть направлены на улучшение практики управления данными.

> Хотя почти все инициативы, прямо или косвенно поддерживаемые поставщиками и разработчиками CAD (BIM), были направлены на оптимизацию рабочих процессов, они привели к появлению множества стандартов, лоббируемых различными заинтересованными сторонами, что привело строительную отрасль к некоторой двусмысленности и путанице в процессах обработки данных.

Перечислим некоторые из новых стандартов данных, помимо LOD, LOI, LOA, LOG, появившихся за последние годы в строительной отрасли:

- **BEP** (BIM Execution Plan) — описывает, как интегрировать и использовать CAD (BIM) в проекте, определяя методы и процессы обработки данных.
- **Документ EIR/AIA** (Информационные требования заказчика) — готовится заказчиком до объявления тендера и содержит требования к подрядчику в отношении подготовки и предоставления информации. Он служит основой для BEP в соответствующем проекте.
- **AIM** (Asset Information Model) — часть процесса BIM. После сдачи и завершения проекта модель данных называется "Информационная модель актива" или AIM. Цель AIM - управление, обслуживание и эксплуатация реализованного актива.
- **IDS** (Information Delivery Specification) — определяет требования и то, какие данные и в каком формате требуются на разных этапах строительного проекта.
- **iLOD** — уровень детализации LOD, с которым информация представлена в BIM-модели. Он определяет, насколько подробной и полной является информация в модели, начиная от базовых геометрических представлений и заканчивая подробными спецификациями и данными.
- **eLOD** — уровень детализации LOD отдельных элементов в модели CAD (BIM). Он определяет степень моделирования каждого элемента и связанную с ним информацию, такую как размеры, материалы, эксплуатационные характеристики и другие соответствующие атрибуты.
- **APS** (Platform Services) и другие продукты от крупных вендоров CAD (BIM) — описывают

инструменты и инфраструктуру, необходимых для создания связанных и открытых моделей данных.

Хотя декларируемая цель внедрения стандартов CAD (BIM) — таких как LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD — заключается в повышении качества управления данными и расширении возможностей автоматизации, на практике их использование нередко приводит к избыточной сложности и фрагментации процессов. Если рассматривать CAD (BIM) модель как разновидность базы данных, становится очевидно, что многие из этих стандартов дублируют уже давно устоявшиеся и эффективные подходы, применяемые в других отраслях экономики при работе с информационными системами. Вместо упрощения и унификации, подобные инициативы зачастую создают дополнительную терминологическую нагрузку и мешают внедрению по-настоящему открытых и гибких решений.

Примечательно, многие из этих новых концепций фактически заменяют собой процесс моделирования и проверки данных, которые рассматривались подробно в первых частях книги и которые уже давно используются в других секторах экономики. В строительстве же процесс стандартизации зачастую движется в обратном направлении — создаются новые форматы описания данных, новые стандарты и новые концепты их проверки, которые не всегда ведут к реальной унификации и практической применимости. В итоге, вместо упрощения и автоматизации обработки, отрасль сталкивается с дополнительными уровнями регламентации и бюрократии (Рис. 6.3-1), что далеко не всегда способствует повышению эффективности.

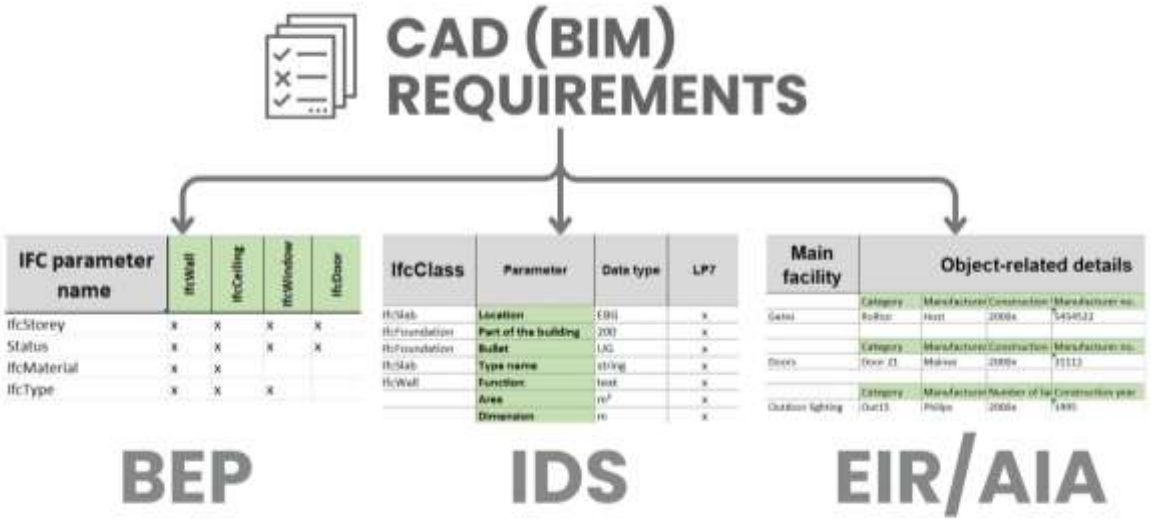


Рис. 6.3-1 Требования к данным и информационной наполненности сводятся к описанию атрибутов и их граничных значений, описываемых с помощью таблиц.

> Вместо того, чтобы упростить обработку данных, новые концепции связанные с CAD (BIM) данными чаще порождают дополнительные сложности и споры уже на стадии интерпретации и базовых определений.

Одним из последних примеров новых концепций, является формат IDS (появившийся в 2020 году)

который позволяет описать требования к атрибутному составу информационной модели в концепте open BIM. Требования IDS описывают информацию об атрибутах и их граничных значениях в виде структурированной таблицы (Excel или MySQL), которая затем переводится в разметку полуструктурированного формата XML, переименованного из XML в специальную аббревиатуру IDS.

> Вопреки мнению, продвигаемому вендорами и поддерживаемым ими концептами BIM и open BIM, о том, что работа с данными в строительстве уникальна из-за использования специализированных инструментов, таких как CAD и BIM, форматы данных и методы управления данными в этой отрасли не отличаются от форматов и концептов обработки данных в других отраслях

Количество требований к проектам и форматам CAD (BIM) можно упростить, используя единую таблицу требований с атрибутами-столбцами, подробно описанную в главе "Перевод требований в структурированную форму", без необходимости переводить изначально структурированные требования в не табличные форматы (IDS изначально описывается через таблицу).

Упрощенный подход (Рис. 6.3-2), включающий столбцы для идентификаторов сущностей, свойств и граничных значений, которые рассматривались подробно в предыдущих главах (Рис. 4.4-9, Рис. 4.4-16, Рис. 7.3-10), позволяет обходиться без преобразования требований в формат IDS-XML. Такой способ обеспечивает прямой, менее громоздкий и более прозрачный механизм контроля качества данных. Он опирается на широко применяемые инструменты: от регулярных выражений (RegEx) до датафреймов, библиотеки Pandas и стандартных ETL-пайплайнов — именно таких, какие повсеместно используются специалистами в других секторах экономики при работе с данными.

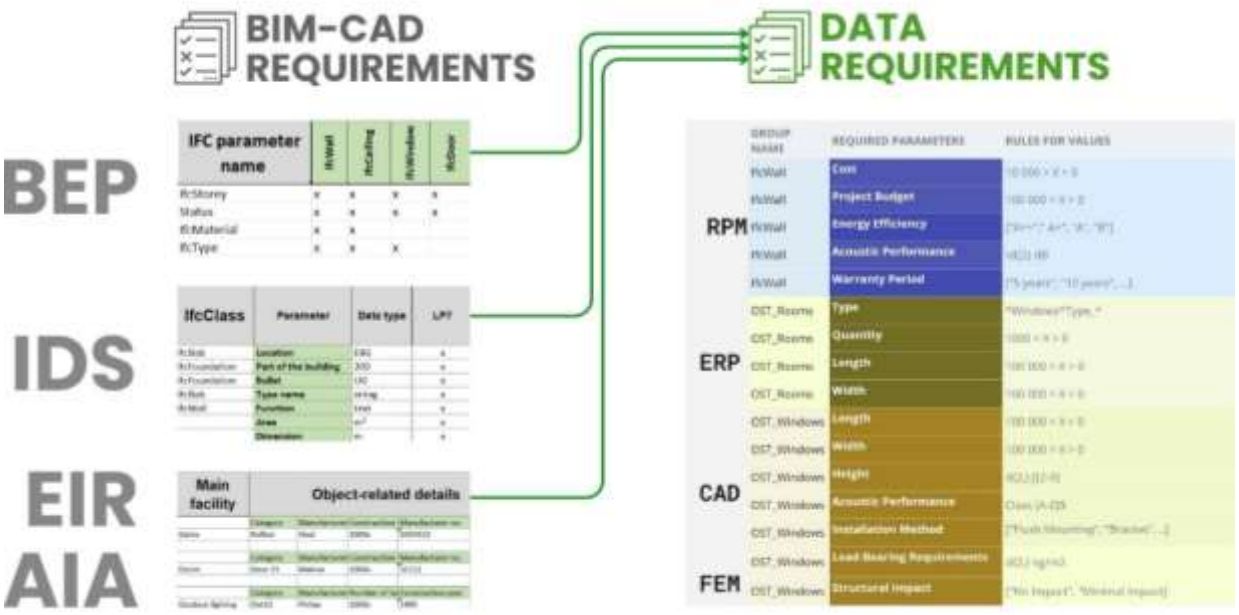


Рис. 6.3-2 Требования к данным в других отраслях упрощается до структурированного описания атрибутов и их граничных значений.

С течением времени в строительной отрасли, из-за закрытости данных, появляется все больше новых подходов и методик для контроля и управления эти разноформатными данными, хотя данные

в строительных проектах, по сути, не отличаются от данных в других сферах. В то время как другие отрасли успешно обходятся стандартизированными подходами к обработке данных, строительство продолжает разрабатывать новые уникальные форматы данных, требования и концепций их проверки.

> Методы и инструменты, используемые для сбора, подготовки и анализа данных в строительстве, не должны принципиально отличаться от тех, что применяются специалистами в других отраслях экономики.

В отрасли сложилась особая терминологическая экосистема, которая требует критического осмысления и переоценки:

- Формат STEP позиционируется под новым названием IFC, дополняясь строительной категоризацией, без учёта ограничений самого STEP-формата.
- Параметрический формат IFC применяется в процессах передачи данных, несмотря на отсутствие унифицированного геометрического ядра, необходимого для визуализации и расчётов.
- Доступ к базам данных CAD-систем продвигается под термином "BIM", без обсуждения особенностей этих баз данных и доступа к ним.
- Вендоры продвигают интероперабельность через форматы IFC и USD, зачастую не применяя их на практике, используя дорогостоящий обратный инжиниринг, с которым сами боролись.
- Термины LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD используются повсеместно для описания одних и тех же параметров сущностей, без привязки к инструментам моделирования и верификации, давно применяющимся в других отраслях.

Строительная отрасль демонстрирует, что все перечисленное хоть и звучит странно, но в строительной отрасли возможно — особенно если основной целью является монетизация каждого этапа обработки данных через продажу специализированных сервисов и ПО. С бизнес-точки зрения в этом нет ничего предосудительного. Однако вопрос, действительно ли подобные аббревиатуры и подходы, связанные с CAD (BIM) добавляют ценность и упрощают профессиональные процессы, остаётся открытым.

> В строительной отрасли подобная система работает, поскольку сама индустрия извлекает основную спекулятивную прибыль именно в этих лабиринтах систем и аббревиатур. Компании, заинтересованные в прозрачных процессах и открытых данных, встречаются крайне редко. Эта сложная ситуация возможно будет продолжаться неопределенно долго — пока заказчики, клиенты, инвесторы, банки и представители частного капитала не начнут требовать более ясных и обоснованных подходов к управлению информацией.

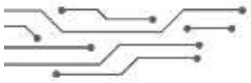
Отрасль накопила избыточное количество аббревиатур, но все они в разной степени описывают одни и те же процессы и требования к данным. Их реальная польза в упрощении рабочих процессов остаётся под вопросом.

В то время, как концепции и маркетинговые аббревиатуры приходят и уходят, сами процессы проверки требований к данным навсегда останутся неотъемлемой частью бизнес-процессов. Вместо

того, чтобы создавать все новые и новые специализированные форматы и регламенты, строительной отрасли стоит обратить внимание на инструменты, которые уже доказали свою эффективность в других сферах, таких, как финансы, промышленность и IT.

Изобилие терминов, аббревиатур и форматов формирует иллюзию глубокой проработанности процессов цифрового строительства. Однако за маркетинговыми концепциями и сложной терминологией часто скрывается простая, но неудобная истина: данные остаются труднодоступными, плохо документированными и жёстко привязанными к конкретным программным решениям.

Чтобы выйти из этого замкнутого круга аббревиатур и форматов ради форматов, необходимо взглянуть на CAD (BIM) системы не как на магические инструменты управления информацией, а как на то, чем они являются на самом деле — специализированные базы данных. И именно через эту призму можно понять, где заканчивается маркетинг и начинается реальная работа с информацией.



ГЛАВА 6.4.
ПАРАМЕТРИЗАЦИЯ ПРОЕКТИРОВАНИЯ И ИСПОЛЬЗОВАНИЕ LLM ДЛЯ РАБОТЫ С CAD

Иллюзия уникальности данных CAD (BIM): путь к аналитике и открытым форматам

Современные CAD (BIM) платформы существенно трансформировали подход к проектированию и управлению строительной информацией. Если ранее эти инструменты использовались преимущественно для создания чертежей и трёхмерных моделей, то сегодня они выполняют функции полноценных хранилищ проектных данных. В рамках концепции Single Source of Truth (единый источник правды) параметрическая модель всё чаще становится основным и нередко единственным источником информации о проекте, обеспечивая её целостность и актуальность на протяжении всего жизненного цикла объекта.

> Ключевое отличие CAD- (BIM-) платформ от других систем управления строительными данными заключается в необходимости использования специализированных инструментов и API для доступа к информации (единственному источнику правды). Эти базы данных не являются универсальными в традиционном смысле: вместо открытой структуры и гибкой интеграции они представляют собой закрытую среду, жёстко завязанную на конкретную платформу и формат.

При всей сложности работы с CAD-данными встаёт более важный вопрос, выходящий за рамки технической реализации: чем на самом деле являются базы данных CAD (BIM)? Чтобы ответить на этот вопрос, необходимо выйти за рамки привычных аббревиатур и концепций, навязываемых разработчиками ПО. Вместо этого стоит сосредоточиться на сути работы с проектной информацией: данных и процессах их обработки.

> Бизнес-процесс в строительстве начинается не с работы в CAD- или BIM-инструментах, а с формирования требований к проекту и моделирования данных. Сначала определяются параметры задачи: перечень сущностей, их начальные характеристики и граничные значения, которые необходимо учитывать при решении конкретной задачи. Только после этого на основе заданных параметров создаются модели и элементы в CAD- (BIM-) системах.

Процесс, который предшествует созданию информации в CAD- (BIM-) базах данных, полностью повторяет процесс моделирования данных, который подробно рассматривался в четвёртой части книги и главе «Моделирование данных: концептуальная, логическая и физическая модель» (Рис. 4.3-1).

Точно так же, как и в процессе моделирования данных, мы создаём требования к данным, которые позже хотим обрабатывать в базе данных, для CAD-баз данных менеджеры создают требования к проектированию в виде нескольких колонок таблицы или списков пар «ключ–значение»

(Рис. 6.4-1, этапы 1–2). И только на основании этих первоначальных параметров с помощью API автоматически или вручную, проектировщиком создаются (или скорее уточняются) объекты в CAD- (BIM) базах данных (этапы 3-4), после чего их вновь проверяют на соответствие исходным требованиям (этапы 5-6). Этот процесс — определение → создание → проверка → корректировка (этапы 2-6) — повторяется итерационно до тех пор, пока качество данных, точно так же, как и при моделировании данных, не достигнет нужного уровня для целевой системы – документов, таблиц или дэшбордов (этап 7).

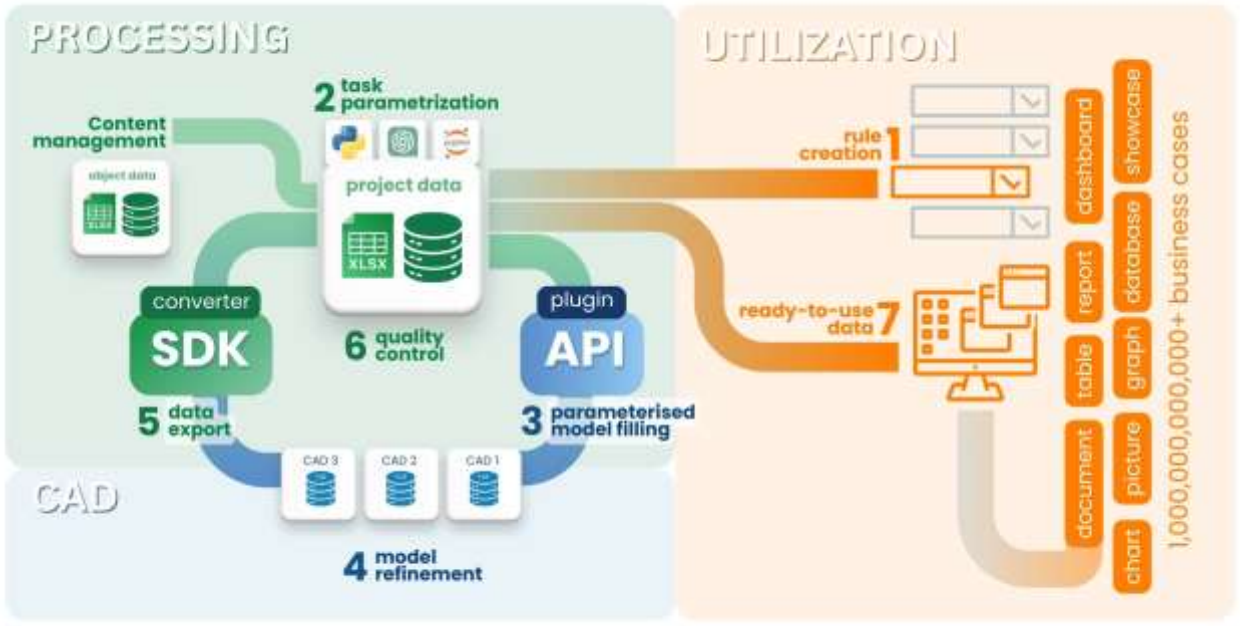


Рис. 6.4-1 Цикл информационного насыщения баз данных для бизнес-процессов в реализации строительных проектов.

Если рассматривать CAD (BIM) как механизм передачи параметров в виде набора пар «ключ–значение», сформированных на основе требований, задаваемых вне среды проектирования (Рис. 6.4-1, этапы 1–2), то фокус обсуждения сместится с конкретных программных решений и их ограничений к более фундаментальным аспектам — структуре данных, моделям данных и требованиям к ним. По сути, речь идёт о насыщении базы данных параметрами и классическом процессе моделирования данных (этапы 2–3 и 5–6). Разница лишь в том, что из-за закрытости CAD-баз данных и особенностей используемых форматов этот процесс сопровождается применением специализированных BIM-инструментов. Возникает вопрос: в чём же уникальность BIM, если в других отраслях экономики не существует подобных подходов?

Последние 20 лет BIM позиционировался как нечто большее, чем просто единый источник данных. Маркетингово связка CAD-BIM зачастую продаётся как параметрический инструмент с изначально интегрированной базой данных [64], способный автоматизировать процессы проектирования, моделирования и управления жизненным циклом объектов строительства. Однако в реальности BIM стал больше инструментом удержания пользователей на платформе вендоров, нежели удобным методом управления данными и процессами.

В итоге CAD- (BIM-) данные изолированы внутри своих платформ, скрывая информацию по проекту за проприетарными API и геометрическими ядрами. Это лишило пользователей возможности самостоятельно получать доступ к базам данных и извлекать, анализировать, автоматизировать и передавать данные в другие системы в обход экосистем вендоров.

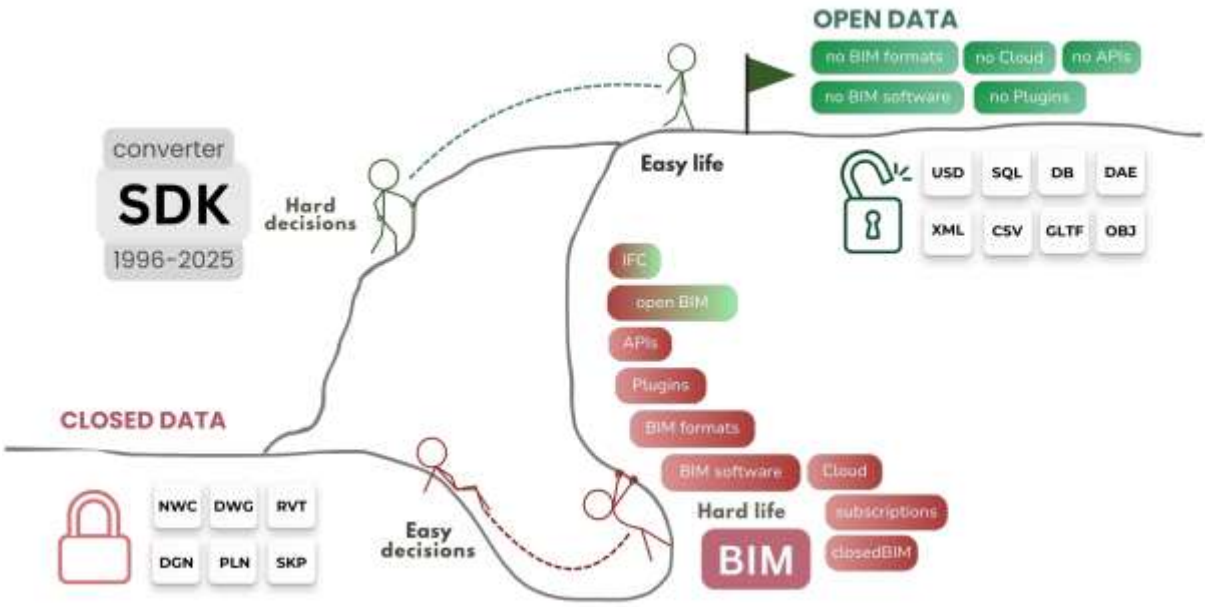


Рис. 6.4-2 В строительстве современные форматы требуют сложных геометрических ядер, ежегодно обновляемого API и специальных лицензий на CAD-(BIM-) программы.

Компаниям, которые работают с современными CAD инструментами, следует использовать тот же подход к работе с данными, который на практике сами применяют все CAD-вендоры без исключения: преобразование данных с использованием SDK-инструментов для обратного инжиниринга, с распространением которых CAD вендоры борются начиная с 1995 года [75]. Имея полный доступ к базе данных CAD и используя инструменты обратного инжиниринга, мы можем получить [127] плоский набор сущностей с атрибутами и экспортировать их в любой удобный открытый формат (Рис. 6.4-2), включающий как геометрию, так и параметры проектных элементов. Такой подход принципиально меняет парадигму работы с информацией — от файлово-ориентированной к дата-центричной архитектуре:

- Форматы данных, такие как: RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET и другие, содержат идентичную информацию об элементах одного и того же проекта. Это значит, что знание конкретного формата и его схемы не должно являться преградой для работы с самими данными.
- Данные из любых форматов можно объединить в одну открытую структурированную и гранулированную структуру (Рис. 9.1-10), содержащую триангулярную геометрию MESH и свойства всех сущностей объектов, без ограничений геометрических ядер.
- Дата-аналитика стремится к универсальности: используя открытые данные, можно работать с проектными данными независимо от используемого формата.

- Минимизация, а также зависимость от API и плагинов вендоров: работа с данными больше не зависит от навыков использования API.

Когда требования и CAD-данные трансформируются в удобные для анализа форматы структурированного представления — разработчики перестают зависеть от специфических схем данных и закрытых экосистем.

Проектирование через параметры: будущее CAD и BIM

Ни один строительный проект в мире никогда не начинался в CAD-программе. Перед тем как чертёж или модель обретают форму в CAD, они проходят стадию концептуализации (Рис. 6.4-1, этапы 1-2), где основное внимание уделяется параметрам, определяющим базовую идею и логику будущего объекта. Эта стадия соответствует концептуальному уровню в моделировании данных (Рис. 4.3-6). Параметры могут существовать исключительно в сознании проектировщика, однако в идеальном случае они оформляются в виде структурированных списков, таблиц или хранятся в базах данных (Рис. 6.4-3), что позволяет обеспечить прозрачность, воспроизводимость и дальнейшую автоматизацию проектного процесса.

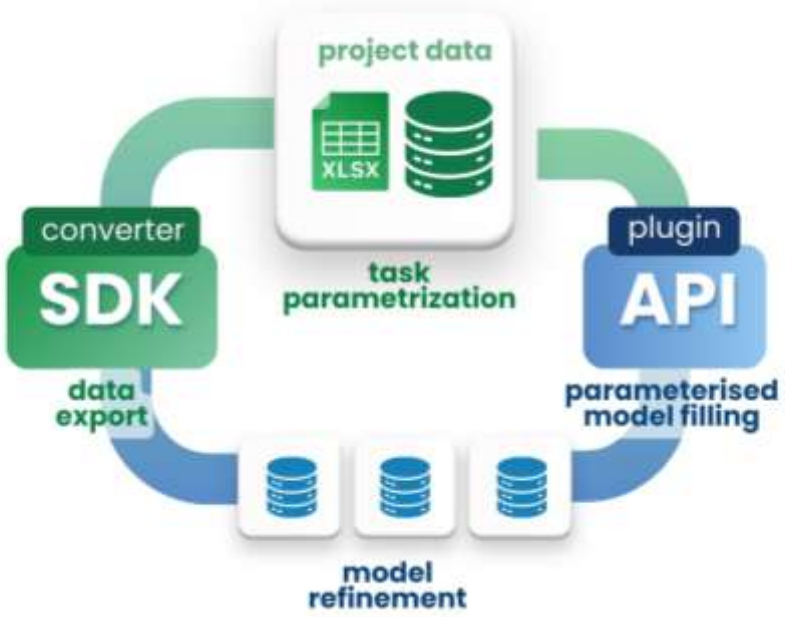


Рис. 6.4-3 Процесс проектирования — это итеративный процесс наполнения базы данных CAD информацией извне при помощи требований в цепочки создания ценности.

Прежде, чем приступить к самому CAD моделированию (логический и физический этап моделирования данных (Рис. 4.3-7)), важно определить граничные параметры, которые служат основой проекта. Эти атрибуты, как и в случае с другими требованиями, собираются с самого конца цепочки использования данных (например, систем) и через них уже определяются ограничения, цели и ключевые характеристики будущих объектов в проекте.

Само моделирование, при наличии грамотно составленных требований, может быть полностью автоматизировано на 60–100% с помощью параметрических инструментов моделирования (Рис.

6.4-3). Как только проект описывается в виде параметров, его формирование становится технически осуществимым например с помощью визуальных языков программирования, таких как Grasshopper Dynamo, встроенных в современные CAD-среды или бесплатных решений в продуктах Blender, UE, Omniverse.

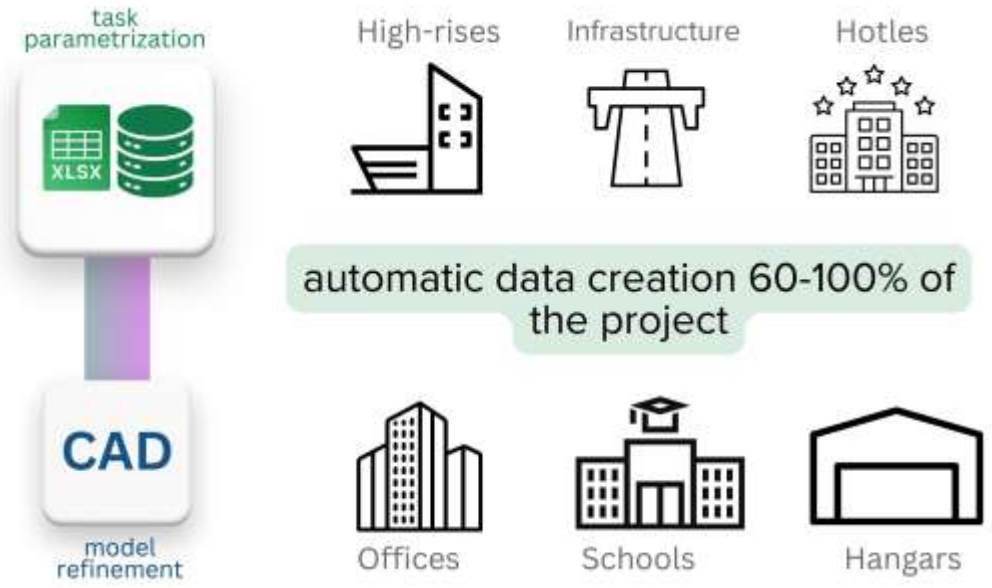


Рис. 6.4-4 Большая часть типизированных проектов уже сегодня создаётся полностью автоматически благодаря инструментам параметрического программирования.

Уже сегодня крупные промышленные и типизированные проекты создаются не руками отдела проектировщиков, а через параметрические инструменты и визуальное программирование. Это позволяет строить модель, отталкиваясь от данных, а не от субъективных решений конкретного проектировщика или менеджера.

> *Содержание предшествует дизайну. Дизайн без содержания — это не дизайн, а декорация [128].*
>
> – Джеффри Зельдман, веб-дизайнер и предприниматель

Процесс начинается не с черчения или 3D-моделирования, а с формирования требований. Именно требования определяют, какие элементы будут использованы в проекте, какие данные необходимо передать в другие отделы и системы. Только существование структурированных требований дает возможность автоматически проверять модели на регулярной основе (например, хоть каждые 10 минут, не отвлекая от работы проектировщика).

> Возможно в будущем CAD- (BIM-) система станет просто интерфейсом для заполнения базы данных, а в каком именно CAD инструменте ведется моделирование (физического уровня) — уже не будет иметь значения.

Точно также и в машиностроении часто трёхмерное моделирование используется, но вовсе не является необходимым или обязательным элементом проекта. В большинстве случаев вполне достаточно классической 2D-документации — на её основе создаётся вся необходимая информационная модель. Эта модель собирается из компонентов, структурированных по отраслевым стандартам, и содержит всю нужную информацию для понимания конструкции и организации производства. Затем на её базе формируется заводская информационная модель, к которой добавляются конкретные изделия и технологические карты, уже ориентированные на нужды технологов. Весь процесс можно организовать без лишней сложности, не перегружая систему 3D-графикой там, где она не даёт реальных преимуществ.

> Важно понимать, что сама 3D модель и CAD-система не должна играть главную роль – она всего лишь инструмент для количественного и геометрического анализа. Все остальные параметры, кроме геометрии, которыми описывается сущность, по возможности должны храниться и обрабатываться вне среды CAD (BIM).

Проектирование через параметры – это не просто тренд, а неизбежное будущее строительной отрасли. Вместо создания сложных 3D-моделей вручную, проектировщики будут работать с данными, проверять их и автоматизировать процессы, приближая строительство к миру программирования. Со временем процессы проектирования будут строиться по принципам разработки ПО:

- Создание требований → Создание модели → Выгрузка на сервер → Проверка изменений → Pull request
- В рамках Pull request (запрос на добавление-слияние) автоматически запускаются проверки модели по требованиям, которые были созданы перед началом или во время проектирования
- После проверки качества данных и одобрения изменения внедряются в проект, общую базу данных или передаются автоматически в другие системы

Уже сейчас в машиностроении подобные изменения проекта начинаются с формирования извещения об изменении. Аналогичная схема ожидает и строительную отрасль: проектирование будет представлять собой итеративный процесс, где каждый шаг подкрепляется параметрическими требованиями. Подобная система позволит проектировщикам создавать автоматизированные проверки и автоматизированный pull request (запрос на добавление-слияние) под конкретные требования.

> Проектировщик будущего — это, прежде всего, оператор данных, а не ручной моделировщик. Его задача — наполнять проект параметрическими сущностями, где геометрия — лишь один из атрибутов.

Важную роль в трансформации будет играть именно понимание важности моделирования данных, классификации и стандартизации, которые подробно рассматривались в предыдущих главах книги. Нормативные акты, регулирующие проектирование в будущем, будут оформлены в виде пар параметров ключ-значение в форме XLSX или XML-схем.

> Будущее строительной отрасли — это сбор данных, их анализ, проверка и автоматизация процессов с помощью инструментов аналитики. BIM (или CAD) — не конечная цель, а лишь этап эволюции. Когда специалисты осознают, что могут работать напрямую с данными, минуя традиционные CAD-инструменты, сам термин «BIM» постепенно уступит своё место концепциям использования структурированных и гранулированных данных строительного проекта.

Одним из ключевых факторов, ускоряющих трансформацию, стало появление больших языковых моделей (LLM) и основанных на них инструментов. Эти технологии меняют подход к работе с проектными данными, позволяя получать доступ к информации без необходимости глубокого знания API или решений вендоров. Благодаря LLM процесс создания требования и взаимодействие с CAD-данными становится интуитивным и доступным.

Появление LLM в процессах обработки проектных CAD данных

В дополнение к развитию инструментов доступа к базам данных CAD и открытых и упрощённых CAD-форматов, революционные изменения в обработке проектных данных вносит появление LLM-инструментов (Large Language Models). Если раньше доступ к информации осуществлялся преимущественно через сложные интерфейсы и требовал владения программированием и знанием API, то теперь стало возможным взаимодействие с данными с помощью естественного языка.

Инженеры, менеджеры и проектировщики без технического бэкграунда могут получать необходимую информацию из проектных данных, формулируя запросы на обычном языке. При условии, что данные структурированы и доступны (Рис. 4.1-13), достаточно задать в LLM-чате вопрос вроде: «*Покажи в виде таблицы с группировкой по типам все стены с объёмом более 10 кубических метров*» — и модель автоматически преобразует этот запрос в SQL или код на Pandas, сформировав итоговую таблицу, график или даже готовый документ.

Ниже приведены несколько реальных примеров того, как LLM-модели взаимодействуют с проектными данными, представленными в различных CAD- (BIM-) форматах.

- Пример запроса в LLM чат к проекту CAD формата RVT после конвертации (Рис. 4.1-13) в табличный датафрейм (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любой другой):

  > Сгруппируй данные в Dataframe полученный из файла RVT по «Имени типа» при суммировании параметра «Объем» и покажи количество элементов в группе. И покажи пожалуйста все это в виде горизонтальной гистограммы без нулевых значений.

Ответ LLM в виде горизонтальной гистограммы (формат PNG):

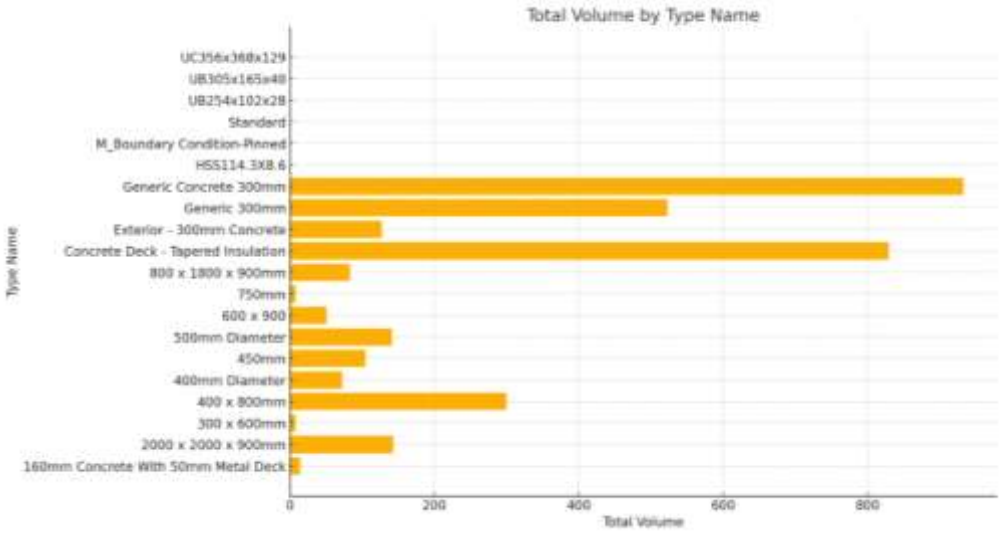


Рис. 6.4-5 Вместо 17 щелчков мышки или 40 строчек кода через использование плагинов, в LLM мы мгновенно получаем QTO таблицу через текстовой запрос.

Чтобы сформировать QTO таблицу типов стен с общей площадью и количеством из категории «Стены» сформулируем текстовый запрос для LLM-чата:

> Возьми только те элементы из датафрейма проекта, у которых в параметре «Категория» есть «OST_Walls», сгруппируй их по «Названию типа», просуммируй значение колонки «Площадь», добавь количество и выведи их в таблицу, удалив нулевые значения.

- Ответ LLM в виде готовой таблицы QTO:

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

Рис. 6.4-6 Создание QTO таблицы на естественном языке предоставляет такой же качественный результат, как и при использовании CAD- (BIM-) инструментов.

- Сделаем запрос к проекту в формате IFC после конвертации в табличный датафрейм и введём подобный текстовый запрос в любой LLM чат:

> Возьми только те элементы из проекта, которые имеют значения Level 1 и Level 2 в параметре «Parent», и возьми элементы, которые имеют значения IfcSlab в параметре «Category», затем сгруппируй эти предметы по параметру «ObjectType», просуммируй значения в параметре «PSet_RVT_Dimensions Area» и покажи их в виде круговой диаграммы.

- Ответ LLM в виде готовой Pie chart групп элементов из IFC данных:

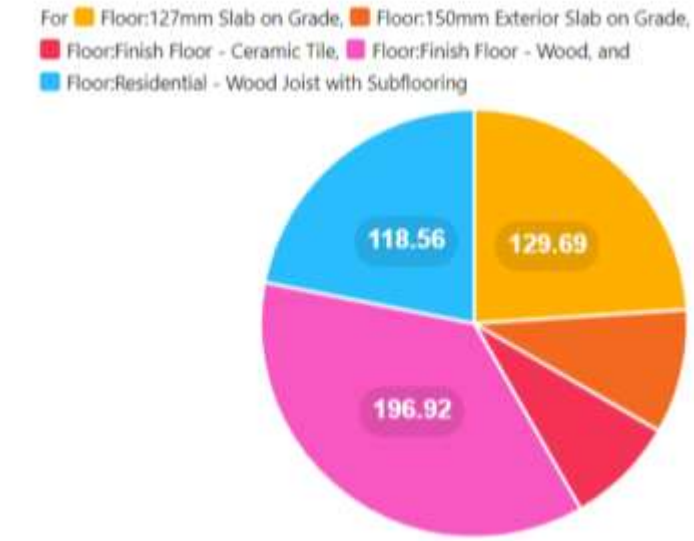


Рис. 6.4-7 Результатом запроса к данным IFC в структурированном формате может быть любой вид графика, который удобен для понимания данных.

За каждым из полученных готовых решений (Рис. 6.4-5 - Рис. 6.4-7) скрывается десяток строчек кода на Python с использованием библиотеки Pandas. Полученный код можно скопировать из чата LLM и использовать в любой локальной или онлайн IDE для получения идентичных результатов вне чата LLM.

В одном и том же чате LLM мы можем работать не только с проектам, полученными из 3D CAD (BIM) форматов, но и плоскими чертежами в формате DWG, к которым после конвертации в структурированную форму мы можем задать запрос в LLM чат для того, чтобы отобразить, например, данные по группам элементов в виде линий или 3D геометрий.

Автоматизированный анализ DWG-файлов с LLM и Pandas

Процесс обработки данных из DWG-файлов из-за не структурированности информации - всегда был сложной задачей, требующей специализированного ПО и часто ручного анализа. Однако с развитием искусственного интеллекта и инструментов LLM, стало возможным автоматизировать многие этапы, этого, сегодня по большей части, ручного процесса. Рассмотрим реальный Pipeline из запросов к LLM (в данном примере ChatGPT) для работы с DWG чертежами, которые позволяют в работе с проектом:

- Фильтровать данные DWG по слоям, ID и координатам
- Визуализировать геометрию элементов
- Автоматически аннотировать чертежи на основе параметров

- Разворачивать полилинии стен в горизонтальную плоскость
- Создавать интерактивные 3D-визуализации плоских данных
- Структурировать и анализировать строительные данные без сложных CAD-инструментов

В нашем случае процесс построения Pipeline начинается с последовательной генерации кода через LLM. Сначала формируется запрос, описывающий задачу. ChatGPT генерирует Python-код, который выполняется и анализируется, показывая результат внутри чата. Если результат не соответствует ожиданиям, запрос корректируется, и процесс повторяется.

> **Pipeline** — это последовательность автоматизированных шагов, выполняемых для обработки и анализа данных. В таком процессе каждый этап принимает данные на входе, выполняет преобразования и передает результат следующему шагу.

После получения нужного результата код копируется из LLM и вставляется в код в виде блоков в любом из удобных IDE, в нашем случае на платформе Kaggle.com. Полученные фрагменты кода объединяются в единый Pipeline, который автоматизирует весь процесс — от загрузки данных до их финального анализа. Подобный подход позволяет быстро разрабатывать и масштабировать аналитические процессы без глубокой экспертизы в программировании. Полный код всех фрагментов, указанных ниже, вместе с примерами запросов вы можете найти на платформе Kaggle.com по запросу «DWG Analyse with ChatGPT | DataDrivenConstruction» [129].

Начнём процесс работы с данными DWG, после конвертации в структурированный вид (Рис. 4.1-13), с классического шага – группировки и фильтрации из всех данных чертежа, необходимых для нашей задачи элементов стен, конкретно полилиний (параметр 'ParentID' позволяет сгруппировать линии в группы), у которых в параметре (колонке датафрейма) «Слой» есть строковое значение, содержащее следующе сочетание букв (RegEx) – «wall».

- Чтобы получить код для подобной задачи и результат в виде картинки необходимо написать следующий запрос в LLM:

> Сначала проверь, содержит ли датафрейм, полученный из DWG, определенные столбцы: 'Layer', 'ID', 'ParentID' и 'Point'. Затем отфильтруй идентификаторы из столбца 'Layer', содержащие строку 'wall'. Найди в столбце 'ParentID' элементы, которые соответствуют этим идентификаторам. Определи функцию для очистки и разбиения данных в столбце 'Point'. Это включает в себя удаление скобок и разбиение значений на координаты 'x', 'y' и 'z'. Построй график данных с помощью matplotlib. Для каждого уникального «ParentID» нарисуй отдельную полилинию, соединяющую координаты «Point». Убедитесь, что первая и последняя точки соединены, если это возможно. Установи соответствующие метки и заголовки, обеспечив одинаковое масштабирование по осям x и y.

- Ответ LLM выдаст готовую картинку за которым скрывается, сгенерировавший его код на языке Python:

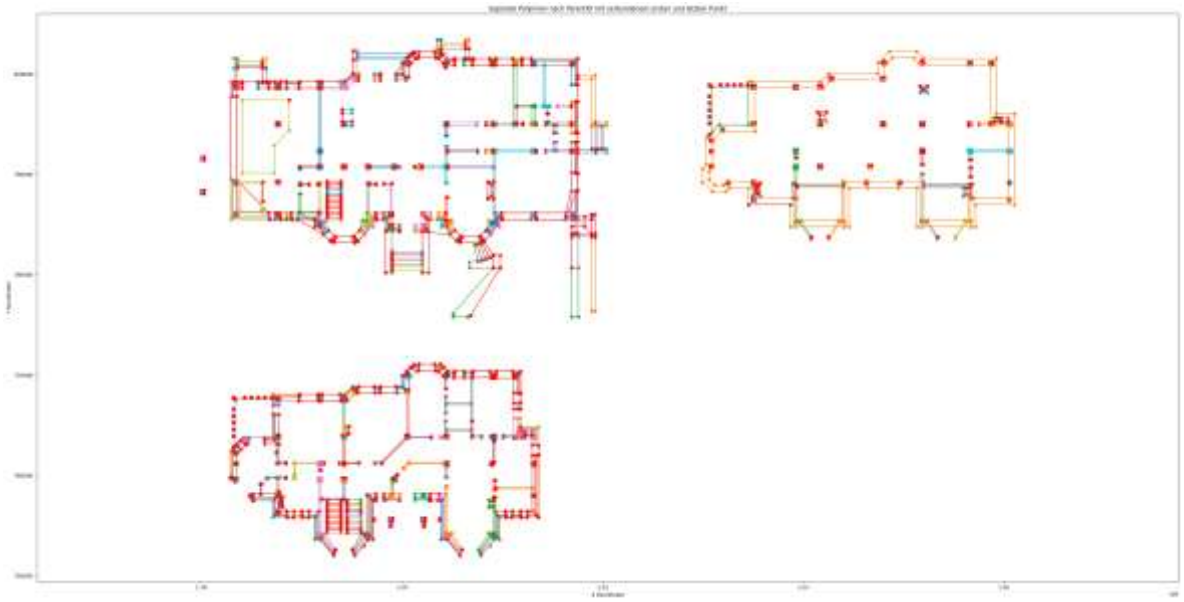


Рис. 6.4-8 LLM код извлёк из DWG-файла все линии слоя "wall", очистил их координаты и построил полилинии при помощи одной из библиотек Python.

- Теперь добавим к линиям параметр площади, который каждая полилиния имеет в своих свойствах (в одной из колонки датафрейма):

> Теперь получи от каждой полилинии только один «ParentID» - найдите этот идентификатор в колонке «ID», возьми значение «Area», раздели на 1 000 000 и добавь это значение на график

Ответ LLM покажет новый график, где у каждой полилинии появится подпись с его площадью:

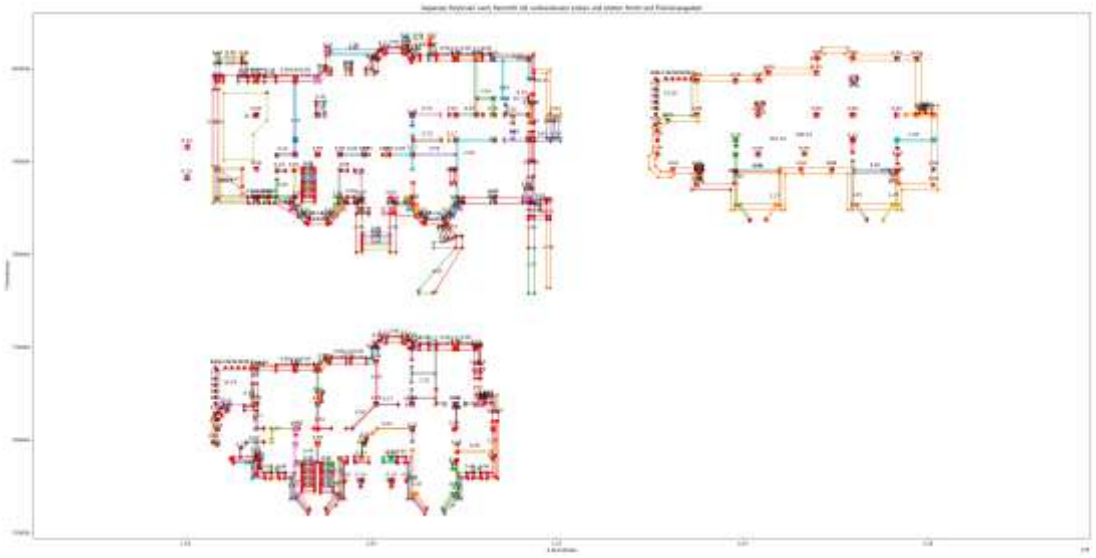


Рис. 6.4-9 LLM дополнил код, который берёт значения площади для каждой полилинии и добавляет его на изображение с визуализацией линий.

Далее преобразуем каждую полилинию в горизонтальную линию, добавим параллельную линию на высоте 3000 мм и соединим их в единую плоскость, чтоб показать таким образом раскладку поверхностей стеновых элементов:

> Необходимо взять все элементы из столбца «Layer» со значением «wall». Возьмите эти ID в виде списка из столбца «ID» и найдите эти ID из всего датафрейма в столбце «ParentID». Все элементы — это линии, которые объединяются в одну полилинию. Каждая линия имеет свою геометрию x, у первой точки в столбце «Point». Нужно взять каждую полилинию по очереди и от точки 0,0 по горизонтали построить длину каждого отрезка от полилинии. длину каждого отрезка полилинии в одну линию. Затем проведите точно такие же линии только на 3000 выше, соедините все точки в одну плоскость.

Ответ LLM выведет код, который позволяет построить развёртки стен в плоскости:

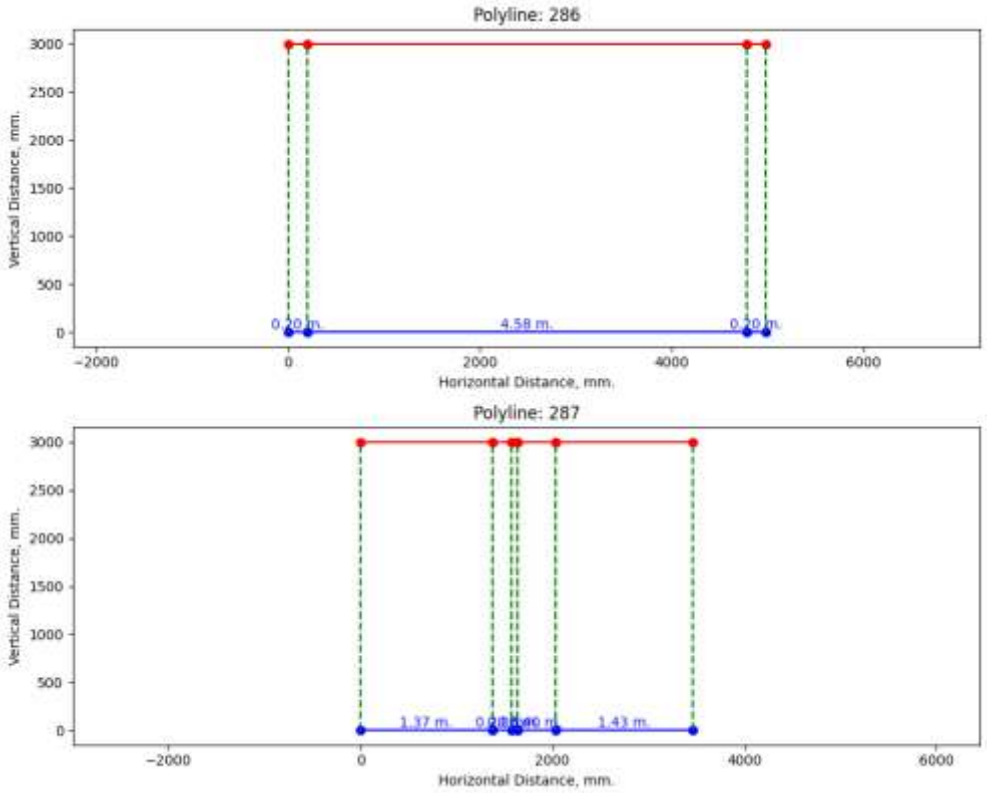


Рис. 6.4-10 Каждую полилинию мы превращаем при помощи промптов в раскладку, которая наглядно показывает плоскости стен напрямую в чате LLM.

Теперь перейдём из двухмерной проекции к 3D-модели стен из плоских линий путем соединения верхних и нижних слоев полилиний:

> Визуализируй элементы стены в 3D, соединяя полилинии на высоте z = 0 и z = 3000 мм. Чтобы создать замкнутую геометрию, представляющую стены здания. Используй 3D-график Matplotlib.

- LLM сгенерирует интерактивный 3D-график, в котором каждая полилиния будет представлена в виде набора плоскостей. Пользователь сможет свободно перемещаться между элементами с помощью компьютерной мыши, исследуя модель в 3D режиме, скопировав код из чата в IDE:

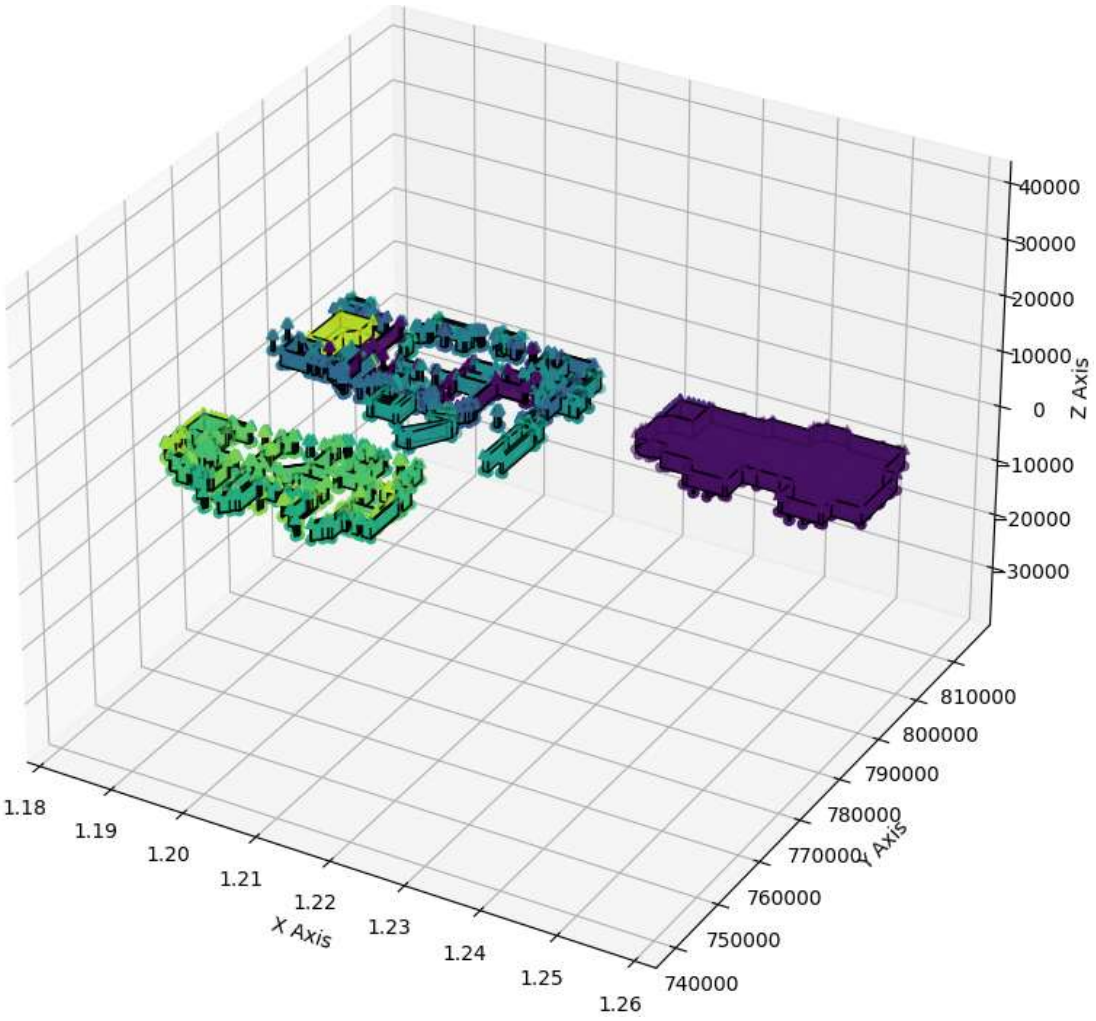


Рис. 6.4-11 LLM помог построить код [129] для визуализации плоских линий чертежа в 3D вид, который можно изучать в 3D просмотрщике внутри IDE.

Для построения логичного и воспроизводимого Pipeline — от первоначальной конвертации и загрузки DWG-файла до получения финального результата — рекомендуется после каждого шага копировать сгенерированный LLM-блок кода в IDE. Таким образом, вы не только проверяете результат в чате, но и сразу запускаете его в своей среде разработки. Это позволяет последовательно выстраивать процесс, отлаживая и адаптируя его по мере необходимости.

Полный код Pipeline всех фрагментов (Рис. 6.4-8 - Рис. 6.4-11) вместе с примерами запросов вы можете найти на платформе Kaggle.com по запросу «DWG Analyse with ChatGPT | DataDrivenConstruction» [129]. На Kaggle можно не только просмотреть код и используемые промпты, но и скопировать и бесплатно протестировать весь Pipeline с исходными датафреймами

DWG в облачной среде без необходимости установки дополнительного ПО и самого IDE.

Представленный в этой главе подход позволяет полностью автоматизировать проверку, обработку и генерацию документов на основе DWG-проектов. Разработанный пайплайн подходит как для обработки отдельных чертежей, так и для пакетной обработки десятков, сотен и тысяч DWG-файлов с автоматическим формированием нужных отчётов и визуализаций по каждому проекту.

Процесс можно выстроить последовательно и прозрачно: сначала данные из CAD-файла автоматически конвертируются в формат XLSX, затем загружаются в датафрейм, после чего выполняются группировка, проверка и формирование результата — всё это реализовано в одном Jupyter-ноутбуке или Python-скрипте, в любом популярном IDE. При необходимости процесс легко расширяется за счёт интеграции с системами управления проектной документацией: CAD-файлы можно автоматически извлекать по заданным критериям, возвращать результаты обратно в систему хранения и уведомлять пользователей о готовности результатов — по email или через мессенджеры.

> Использование LLM чатов и агентов для работы с проектными данными снижает зависимость от специализированных CAD-программ и позволяет выполнять анализ и визуализацию архитектурных проектов без необходимости ручного взаимодействия с интерфейсом — без кликов мышкой и запоминания сложной навигации по меню.

С каждым днем все чаще в строительной отрасли можно будет услышать о LLM, гранулированных структурированных данных, DataFrame и колончатых базах данных. Унифицированные двухмерные DataFrame, сформированные из различных баз данных и форматов CAD, станут идеальным топливом для современных аналитических инструментов, с которыми работают активно специалисты в других отраслях экономики.

Существенно упростится сам процесс автоматизации – вместо изучения API закрытых нишевых продуктов и написания сложных скриптов для анализа или трансформации параметров, теперь достаточно будет сформулировать задачу в виде набора отдельных текстовых команд, которые будут складываться в нужный Pipeline или Workflow-процесс для нужного языка программирования, которые запускается бесплатно почти на любом устройстве. Больше не будет необходимости ждать выпуска новых продуктов, форматов, плагинов или обновлений от вендоров CAD- (BIM-) инструментов. Инженеры и строители получат возможность независимо работать с данными, используя простые, бесплатные и понятные инструменты, в работе с которыми помогут LLM чаты и агенты.

Дальнейшие шаги: переход от закрытых форматов к открытым данным

При работе с проектными данными будущего вряд ли кому-то действительно понадобится глубоко разбираться в геометрических ядрах проприетарных инструментов или изучать сотни несовместимых форматов, содержащих одну и ту же информацию. Однако без понимания, почему переход к открытым структурированным данным важен, сложно аргументировать необходимость использования новых бесплатных инструментов, открытых данных и подходов, которые вряд ли буду продвигаться вендорами ПО.

В этой главе мы разобрали ключевые особенности CAD (BIM) данных, их ограничения и возможности и что несмотря на маркетинговые обещания вендоров, инженеры и проектировщики каждый

день сталкиваются с трудностями при извлечении, передаче и анализе проектной информации. Понимание архитектуры этих систем и знакомство с альтернативными подходами — на основе открытых форматов и автоматизации при помощи LLM — способно значительно упростить жизнь даже одному специалисту, не говоря уже о компаниях. Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные подходы в ваших повседневных задачах:

- Расширьте свой инструментарий для работы с проектными данными
  - ☐ Изучите доступные плагины и утилиты для извлечения данных из используемых вами CAD- (BIM-) систем
  - ☐ Изучите доступные SDK и API, которые позволяют автоматизировать извлечение данных из закрытых форматов без необходимости вручную открывать специализированное программное обеспечение
  - ☐ Освойте базовые навыки работы с открытыми непараметрическими форматами геометрии (OBJ, glTF, USD, DAE) и соответствующими открытыми библиотеками
  - ☐ Попробуйте продумать систему хранения метаданных проекта отдельно от геометрии вне CAD (BIM) решений для упрощения анализа и интеграции с другими системами
  - ☐ Используйте LLM для автоматизации вопросов преобразования данных между форматами
- Создайте собственные процессы обработки проектной информации
  - ☐ Начните описывать задачи и требования к моделированию через параметры и их значения в простых и структурированных форматах
  - ☐ Создайте личную библиотеку скриптов или блоков кода для часто выполняемых операций
- Продвигайте использование открытых стандартов в своей работе
  - ☐ Предлагайте коллегам и партнерам обмениваться данными в открытых форматах, не ограниченных экосистемой вендоров ПО
  - ☐ Демонстрируйте преимущества использования структурированных данных на конкретных примерах
  - ☐ Инициируйте обсуждения о проблемах с закрытыми форматами и возможных решениях

Даже если вы не можете изменить политику компании в отношении CAD- (BIM-) платформ, личное понимание принципов работы с проектными данными в открытых форматах позволит вам значительно повысить эффективность своей работы. Создавая собственные инструменты и методы для извлечения и преобразования данных из различных форматов, вы не только оптимизируете рабочие процессы, но и получаете гибкость, позволяющую обходить ограничения стандартных программных решений.

VII ЧАСТЬ

ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ ДАННЫХ, АНАЛИТИКА, АВТОМАТИЗАЦИЯ И МАШИННОЕ ОБУЧЕНИЕ

Седьмая часть посвящена аналитике данных и автоматизации процессов в строительной отрасли. Здесь рассматривается как данные становятся основой для принятия решений и объясняются принципы визуализации информации для эффективного анализа. Подробно описываются ключевые показатели эффективности (KPI), методы оценки возврата инвестиций (ROI) и создание информационных панелей для мониторинга проектов. Особое внимание уделяется процессам ETL (Extract, Transform, Load) и их автоматизации с использованием конвейеров (Pipeline), что позволяет превратить разрозненные данные в структурированную информацию для анализа. Рассматриваются инструменты оркестрации рабочих процессов, такие, как Apache Airflow, Apache NiFi и n8n, которые позволяют строить автоматизированные конвейеры данных без глубоких знаний программирования. Существенная роль отводится большим языковым моделям (LLM) и их применению для упрощения анализа данных и автоматизации рутинных задач.

ГЛАВА 7.1.
АНАЛИТИКА ДАННЫХ И ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ ДАННЫХ

После этапов сбора, структурирования, очистки и проверки информации сформировался целостный и пригодный для анализа массив данных. В предыдущих частях книги рассматривалась систематизация и структурирование разнородных источников — от PDF-документов и текстовых записей встреч до CAD-моделей и геометрических данных. Подробно описан процесс проверки и приведения информации в соответствие с требованиями различных систем и классификаторов, устранения дубликатов и противоречий.

Все расчёты, выполненные на этих данных (третья, четвёртая часть книги) — от простых преобразований до вычислений сроков, стоимости и ESG-показателей (пятая часть) — представляют собой укрупнённые задачи аналитики. Они формируют основу для понимания текущего состояния проекта, оценки его параметров и последующего принятия решений. В результате данные, в результате расчётов, превращаются из набора разрозненных записей в управляемый ресурс, способный отвечать на ключевые вопросы бизнеса.

В предыдущих главах подробно рассматривались процессы сбора данных и контроля их качества для использования в типовых бизнес-кейсах и процессах, характерных для строительной отрасли. Аналитика в этом контексте во многом схожа с применениями в других отраслях, но обладает и рядом специфических особенностей.

В следующих главах будет подробно рассмотрен укрупнённый процесс анализа данных, включая этапы автоматизации — от первичного получения информации и её трансформации до последующей передачи в целевые системы и документы. Сначала будет представлена теоретическая часть, посвящённая отдельным аспектам принятия решений на основе данных. Затем, в последующих главах, начнётся практическая часть, связанная с автоматизацией и построением ETL-Pipeline.

Данные как ресурс в принятии решений

Процесс принятия решений на основе данных является часто итеративным процессом и начинается с систематического сбора информации из разнообразных источников информации. Подобно природному круговороту, отдельные элементы данных и целые информационные системы постепенно падают в почву - накапливаются в информационных хранилищах компаний (Рис. 1.3-2). Со временем эти данные, подобно упавшим листьям и веткам, преобразуются в ценный материал. Мицелий инженеров по данным и аналитиков организует и подготавливает информацию к будущему использованию и превращает опавшие данные и системы в ценный компост, для выращивания новых ростков и новых систем (Рис. 1.2-5).

Тенденции широкого использования аналитики в разных отраслях, знаменует начало новой эпохи, где работа с данными становится основой профессиональной деятельности (Рис. 7.1-1). Специалистам строительной отрасли важно адаптироваться к этим изменениям и быть готовыми к переходу в новую эру — эпоху данных и аналитики.

Ручное перемещение данных между таблицами и выполнение расчётов вручную постепенно уходят в прошлое, уступая место автоматизации, анализу потоков данных, аналитике и машинному обучению. Эти инструменты становятся ключевыми элементами современной системы поддержки принятия решений.

В книге McKinsey "Перезагрузка. Руководство McKinsey по преодолению конкуренции в эпоху цифровых технологий и искусственного интеллекта" [130], приводится исследование, проведённое в 2022 году с участием 1 330 руководителей высшего звена из различных регионов, отраслей и функциональных направлений. Согласно его результатам, 70% лидеров используют передовую аналитику для формирования собственных идей, а 50% — внедряют искусственный интеллект для улучшения и автоматизации процессов принятия решений.

Рис. 7.1-1 Анализ данных и аналитика - основной инструмент для повышения скорости принятия решений в компании.

Анализ данных, подобно распространению мицелия, проникает сквозь гумус прошлых решений, помогая соединять отдельные системы и направляя менеджеров к ценным знаниям. Эти знания, подобно питательным веществам из разложившихся деревьев систем данных, питают новые решения в компании, приводя к эффективным изменениям и качественному информационному росту, подобно новым побегам и росткам, появляющимся из богатой и здоровой почвы (Рис. 1.2-5).

Числа имеют важную историю, которую они должны рассказать. Они рассчитывают на то, что вы дадите им ясный и убедительный голос [131].

– Стивен Фью, Эксперт по визуализации данных

В средних и малых по размерам компаниях, работа по извлечению и подготовке информации для дальнейшего анализа сегодня представляет собой исключительно трудоемкий процесс (Рис. 7.1-2), сопоставимый с угледобычей XVIII века. До последнего времени работа по добыче и подготовке данных скорее была предназначена для авантюристов, работающих в узкоспециализированной нише с небольшим и ограниченным набором инструментов для работы с различными типами данных из не структурированных, слабоструктурированных, смешанных и закрытых источников.

Руководители и менеджеры, принимающие решения, часто не обладают достаточным опытом работы с разнородными данными и системами, но при этом им необходимо принимать решения на их основе. В результате принятие решений на основе данных в современной строительной отрасли последние десятки лет меньше похоже на автоматизированный процесс и больше на многодневный ручной труд горнорабочего в первых угольных шахтах.

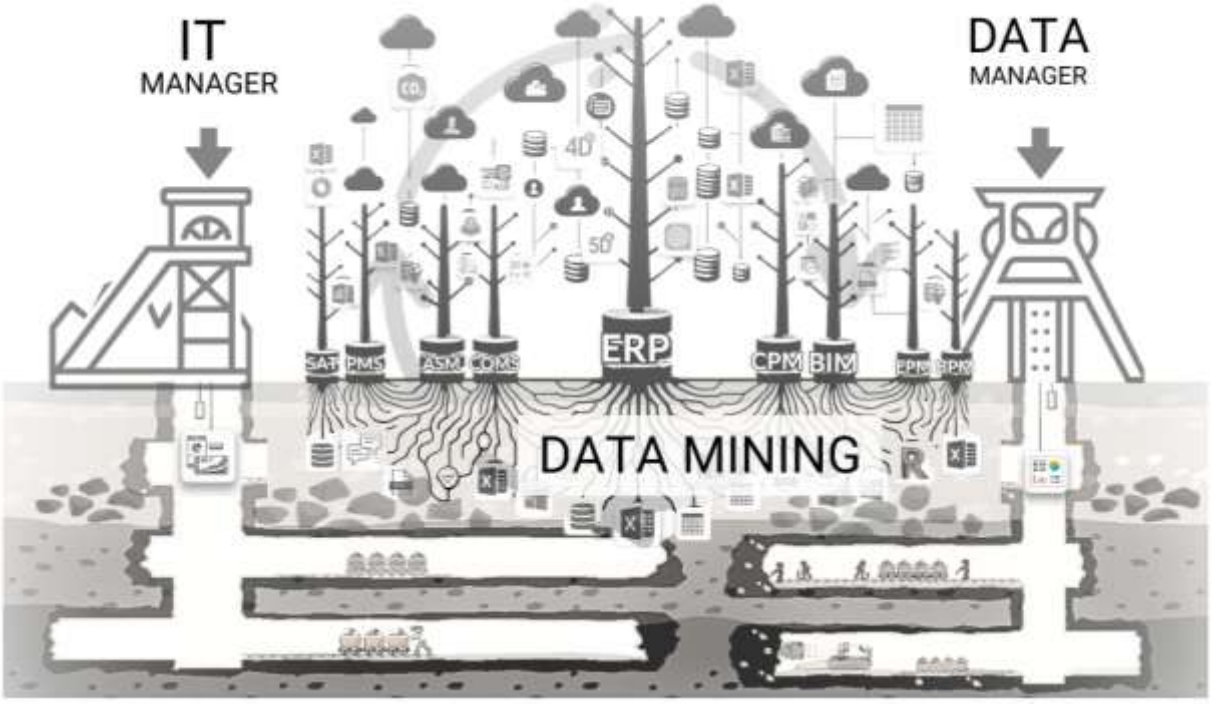


Рис. 7.1-2 В процессе data mining специалисты проходят сложный путь подготовки данных — от очистки до структурирования для последующей аналитики.

И хотя современные методы извлечения данных в строительной индустрии, несомненно, прогрессивнее, чем примитивные техники рудокопов XII столетия, это по-прежнему остается сложнейшей и высокорисковой задачей, требующей значительных ресурсов и экспертизы, которую могли себе позволить только крупные компании. Процессами извлечения и аналитики данных из накопленного наследия прошлых проектов до недавнего времени преимущественно занимались крупные, технологически развитые компании, которые последовательно собирали и хранили данные на протяжении десятилетий.

Ранее ведущую роль в аналитике играли технологически зрелые компании, десятилетиями накапливавшие данные. Сегодня ситуация меняется: доступ к данным и инструментам их обработки становится демократичным — сложные ранее решения теперь доступны всем и бесплатно.

Применение аналитики позволяет компаниям принимать более точные и обоснованные решения

в реальном времени. Ниже приведён практический пример, иллюстрирующий, как исторические данные помогают принимать финансово обоснованные решения:

- **Менеджер проекта** — «*Сейчас средняя цена на бетон в городе 82 €/м³, у нас в смете стоит 95 €/м³*».
- **Сметчик** – «*По предыдущим проектам перерасход был около 15%, поэтому я подстраховался*».
- **Менеджер по данным или инженер контроля со стороны заказчика** – «*Давайте посмотрим аналитику по трем последним тендерам*».

После анализа DataFrame из прошлых проектов получаем:

- **Средняя фактическая закупочная цена:** 84,80 €/м³
- **Средний коэффициент перерасхода:** +4,7%
- **Рекомендованная ставка в смете:** ~85 €/м³

Подобное решение уже основано будет не на субъективных ощущениях, а на конкретной исторической статистике, что позволяет снизить риски и повысить обоснованность тендерной ставки. Анализ данных из прошлых проектов становится своего рода «органическим удобрением», из которого прорастают новые, более точные решения.

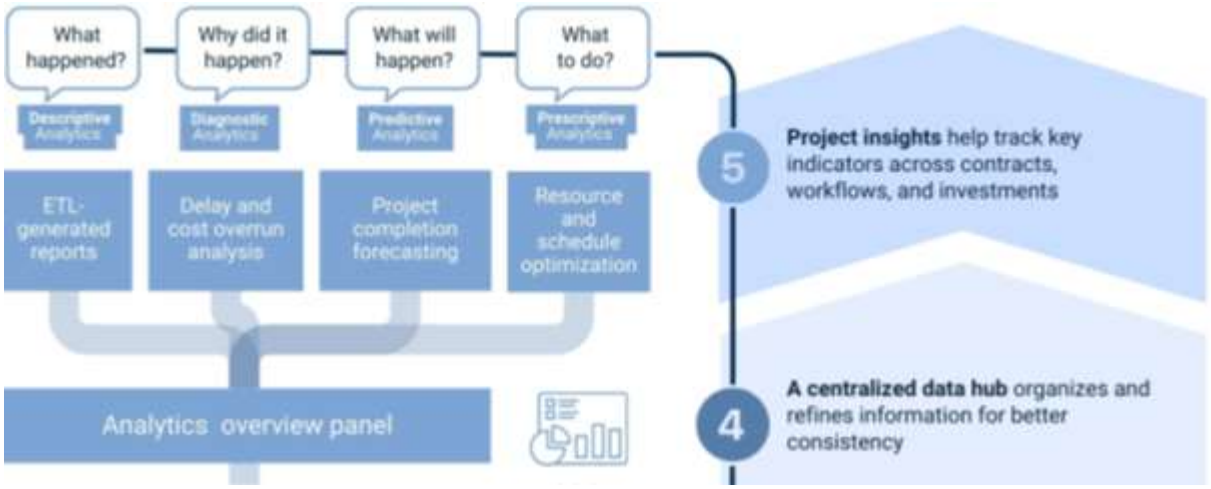


Рис. 7.1-3 Аналитика данных отвечает на три ключевых вопроса: что произошло, почему это произошло и что следует делать дальше.

Руководители и менеджеры, принимающие решения, часто сталкиваются с необходимостью работы с разнородными данными и системами, не обладая при этом достаточной технической подготовкой. В таких ситуациях одним из ключевых помощников в процессе понимания данных становится визуализация — один из первых и важнейших этапов аналитического процесса. Она позволяет представить информацию в наглядной и понятной форме.

Визуализация данных: ключ к пониманию и принятию решений

В современной строительной отрасли, где проектные данные характеризуются сложностью и многоуровневой структурой, визуализация играет ключевую роль. Визуализация данных позволяет руководителям проектов и инженерам визуализировать сложные закономерности и тенденции, скрытые в больших, разнородных объемах данных.

> Визуализация данных облегчает понимание состояния проекта: распределение ресурсов, динамики затрат или использование материалов. Благодаря графикам и диаграммам сложная и сухая информация становится доступной и понятной, позволяя быстро определить ключевые области, требующие внимания, и выявить потенциальные проблемы.

Визуализация данных не просто облегчает интерпретацию информации, она является важнейшим этапом в аналитическом процессе и принятии обоснованных управленческих решений, помогая ответить на вопросы «*что произошло?*» и «*как произошло?*» (Рис. 2.2-5).

> *Графика — это визуальные средства решения логических задач [132].*
>
> — Жак Бертен, «Графика и обработка графической информации»

Перед принятием ключевых решений руководители проектов с большей вероятностью будут использовать визуальные представления данных, а не сухие и трудно интерпретируемые цифры из электронных таблиц или текстовых сообщений.

> Данные, не подкрепленные визуализацией, подобны строительным материалам, беспорядочно разбросанным по стройплощадке: их потенциал неясен. Только когда из них возникает четкая визуализация, как из кирпича и бетона - дом, становится ясно, какую ценность они представляют. Пока дом не построен, невозможно сказать, станет ли груда материалов небольшой хижиной, роскошной виллой или небоскребом.

Компании располагают данными из различных систем (Рис. 1.2-4 - Рис. 2.1-10), финансовыми операциями и обширными текстовыми данными. Однако использование этих данных в интересах бизнеса зачастую представляет собой сложную задачу. В таких ситуациях визуализация становится важным инструментом для передачи смысла данных, который помогает представить информацию в форматах, понятных любому специалисту, например в виде приборных панелей, графиков и диаграмм.

> Исследование PwC “Что нужно студентам, чтобы добиться успеха в быстро меняющемся мире бизнеса“ (2015) подчеркивает [9], что успешные компании не ограничиваются анализом данных, а активно используют интерактивные визуализационные инструменты, такие как графики, инфографика и аналитические панели, для поддержки принятия решений. Согласно отчёту - визуализация данных помогает клиентам понять историю, которую рассказывают данные, с помощью графиков, диаграмм, дэшбордов и интерактивных моделей данных.

Процесс преобразования информации в визуальные графические формы, такие как диаграммы, графики, улучшает понимание и интерпретацию данных (Рис. 7.1-4) человеческим мозгом. Это позволяет руководителям проектов и аналитикам быстрее оценивать сложные сценарии и принимать обоснованные решения, основываясь не на интуиции, а на визуально распознаваемых тенденциях и паттернах.

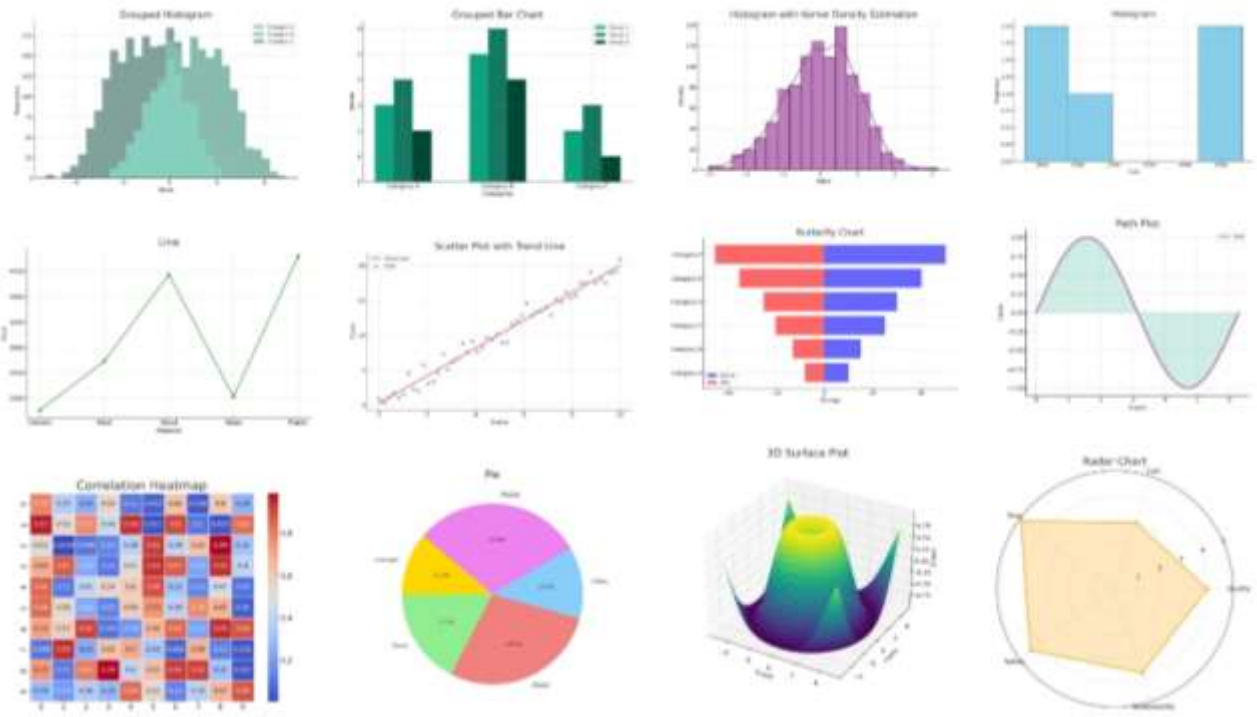


Рис. 7.1-4 Различные виды визуализации созданы для того, чтобы помочь человеческому мозгу лучше понять и осмыслить сухую информацию чисел.

Подробнее вопросы создания визуализаций из данных, и использование различных бесплатных библиотек визуализации, будут рассмотрены в следующей главе, посвящённой ETL-процессам.

Визуализация становится неотъемлемым элементом работы с данными в строительной отрасли — она помогает не просто "увидеть" данные, но и понять их значение в контексте управленческих задач. Однако для того чтобы визуализация была действительно полезной, необходимо заранее определить, что именно следует визуализировать и какие метрики действительно важны для оценки эффективности проекта. Здесь на сцену выходят показатели эффективности, такие как KPI и ROI. Без них даже самые красивые дашборды рискуют остаться лишь «информационным шумом».

Показатели эффективности KPI и ROI

В современной строительной отрасли управление показателями эффективности (KPI и ROI), а также их визуализация с помощью отчетов и информационных панелей (dashboards) играют ключевую роль в повышении производительности и эффективности управления проектами.

> Как и в любом бизнесе, в строительстве необходимо чётко определять метрики, по которым оцениваются успех, окупаемость и производительность. Получая данные по различным процессам, организация, работающая на основе данных, должна прежде всего научится определить ключевые **KPI (Key Performance Indicators)** — количественные показатели, отражающие степень достижения стратегических и операционных целей.

Для расчета KPI обычно используется формула (Рис. 7.1-5), включающая фактические и плановые показатели. Например, чтобы рассчитать индивидуальный KPI для проекта, сотрудника или процесса, необходимо разделить фактические показатели на плановые и умножить полученный результат на 100 %.

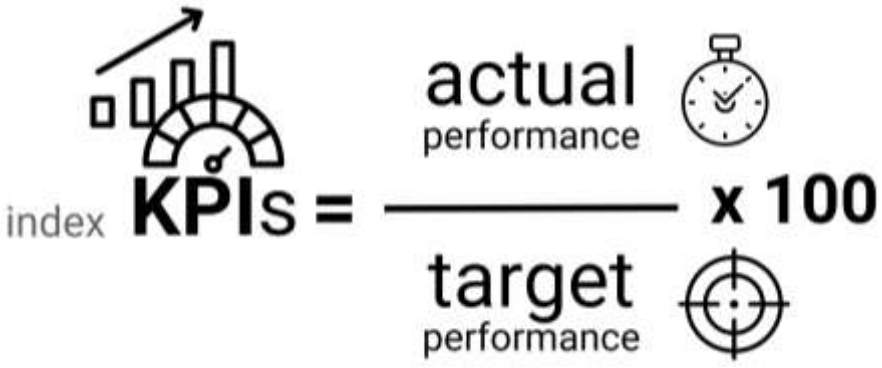


Рис. 7.1-5 KPI используются для измерения успеха проекта или процесса в достижении ключевых целей.

На уровне строительной площадки могут использоваться более детализированные KPI метрики:

- **Сроки выполнения ключевых этапов** (фундамент, монтаж, отделка) – позволяют контролировать соблюдение планов работ.
- **Процент перерасхода материалов** – помогает управлять закупками и минимизировать потери.
- **Количество внеплановых простоев техники** – влияет на производительность и затраты.

Выбор неверных метрик может привести к ошибочным решениям по вопросу «*что делать?*» (Рис. 2.2-5). Например, если компания ориентируется только на стоимость квадратного метра, но не учитывает затраты на переделки, экономия на материалах может привести к ухудшению качества и росту расходов в будущих проектах.

При постановке целей важно четко определять, что именно измеряется. Размытые формулировки приводят к некорректным выводам и усложняют контроль. Рассмотрим примеры удачных и неудачных KPI в строительстве.

Хорошие KPI:

- "*К концу года снизить процент переделок отделочных работ на 10%*"
- "*Увеличить скорость монтажа фасадов на 15% без снижения качества к следующему кварталу*"
- "*Сократить время простоя техники на 20% за счет оптимизации графиков работы к концу года*"

Эти метрики четко измеряемы, имеют конкретные значения и временные рамки.

Плохие KPI:

- "*Мы будем строить быстрее*" (Насколько быстрее? Что значит "быстрее"?)
- "*Мы повысим качество бетонных работ*" (Как именно измеряется качество?)
- "*Мы улучшим взаимодействие подрядчиков на площадке*" (Какие критерии покажут улучшение?)

> Хороший KPI — это тот, который можно измерить и объективно оценить. В строительстве это особенно важно, поскольку без четких показателей невозможно контролировать эффективность работы и достигать стабильных результатов.

Помимо KPI существует дополнительная метрика для оценки эффективности вложений: **ROI (Return on Investment)** — показатель возврата на инвестиции, отражающий соотношение между прибылью и вложенными средствами. ROI позволяет оценить, насколько оправдано внедрение новых методов, технологий или инструментов: от цифровых решений и автоматизации (например Рис. 7.3-2) до применения новых строительных материалов. Этот показатель помогает принимать обоснованные решения о дальнейших инвестициях, ориентируясь на их реальное влияние на прибыльность бизнеса.

> В контексте управления строительными проектами ROI (возврат инвестиций) может быть использован как один из ключевых показателей эффективности (KPI), если цель компании — оценить окупаемость инвестиций в проект, технологию или улучшение процессов. Например, если внедряется новая методика управления строительством, ROI может показать, насколько она повысила прибыльность.

Регулярное измерение показателей KPIs и ROIs на основе данных, собранных из различных источников, таких как расход материалов, трудочасы и затраты, позволяет руководству проектов эффективно управлять ресурсами и быстро принимать решения. Хранение этих данных в долгосрочной перспективе позволяет анализировать будущие тенденции и оптимизировать процессы.

Для визуализации KPI, ROI и других показателей используются различные графики и диаграммы, которые обычно объединяются в приборные панели.

Информационные панели и дашборды: визуализация показателей для эффективного управления

Для визуализации показателей и метрик используются разнообразные графики и диаграммы, которые обычно объединяются в витрины данных и информационные панели (dashboards). Такие панели предоставляют централизованное представление состояния проекта или его отдельных частей, отображая ключевые показатели (в идеале в режиме реального времени). Актуальные и постоянно обновляемые дашборды позволяет команде быстро реагировать на изменения.

> **Дашборды** — это инструменты, которые визуализируют количественные оценки, делая их легкодоступными и понятными для всех участников проекта.

Рис. 7.1-6 Управление KPI и их визуализация с помощью информационных панелей - ключ к повышению производительности и эффективности проекта.

Вот несколько примеров популярных инструментов, в которых можно создавать информационные панели:

- **Power BI** - инструмент от Microsoft для создания интерактивных отчетов и информационных панелей.
- **Tableau и Google Data Studio** - мощные инструменты для визуализации данных и создания информационных панелей без необходимости писать код.
- **Plotly** (Рис. 7.1-6, Рис. 7.2-12) — это библиотека для создания интерактивных графиков, а Dash - фреймворк для создания веб-приложений для анализа данных. Их можно использовать в комбинации для создания интерактивных приборных панелей.
- **Многочисленные библиотеки Python** (Рис. 7.2-9 - Рис. 7.2-11) – в Python есть множество

открытых и бесплатных библиотек для визуализации данных, таких как Matplotlib, Seaborn, Plotly, Bokeh и другие. Их можно использовать для создания графиков и интеграции их в веб-приложение с помощью таких фреймворков, как Flask или Django.

- **Библиотеки JavaScript**: позволяют создавать интерактивные информационные панели с помощью Open Source библиотек JavaScript, таких как D3.js или Chart.js, и интегрировать их в веб-приложение.

> Для оценки KPI и создания информационных панелей необходимы актуальные данные и четкий график сбора и анализа информации.

В целом KPI, ROI и информационные панели в строительной отрасли формируют основу для аналитического подхода к управлению проектами. Они не только помогают отслеживать и оценивать текущее состояние, но и предоставляют ценные сведения для будущего планирования и оптимизации процессов — процессов, напрямую зависящих от интерпретации данных и умения задавать правильные и своевременные вопросы.

Анализ данных и искусство задавать вопросы

Интерпретация данных — заключительный этап анализа, на котором информация обретает смысл и начинает «говорить». Именно здесь формулируются ответы на ключевые вопросы: *«что делать?»* и *«как делать?»* (Рис. 2.2-5). Этот этап позволяет обобщать результаты, выявлять закономерности, устанавливать причинно-следственные связи и делать выводы на основе визуализации и статистического анализа.

> *Возможно, не за горами то время, когда придет понимание, что для полноценного становления в качестве эффективного гражданина одного из новых великих сложных мировых государств, которые сейчас развиваются, так же необходимо уметь вычислять, мыслить средними величинами, максимумами и минимумами, как сейчас необходимо уметь читать и писать [133].*
>
> – Сэмюэл С. Уилкс, цитата из президентского обращения в 1951 году к Американской статистической ассоциации

> Согласно опубликованному правительством Великобритании отчету «Аналитика данных и искусственный интеллект в реализации правительственных проектов» (2024) [83], внедрение аналитики данных и искусственного интеллекта (ИИ) позволяет значительно улучшить процессы управления проектами, повышая точность прогнозирования сроков и затрат, а также снижая риски и неопределенность. В документе подчеркивается, что государственные организации, использующие продвинутые аналитические инструменты, достигают более высоких показателей эффективности при реализации инфраструктурных инициатив.

Современный строительный бизнес, действующий в условиях высокой конкуренции и низкой маржинальности в рамках четвёртой промышленной революции, можно сравнить с военными действиями. Здесь выживание и успех компании зависят от скорости получения ресурсов и качественной информации — а значит, от своевременного и обоснованного принятия решений (Рис. 7.1-7).

Если визуализация данных — это «разведка», предоставляющая обзор, то аналитика данных — это «боеприпасы», необходимые для действий. Именно она отвечает на вопросы: *что делать?* и *как это делать?*, формируя основу для получения конкурентных преимуществ на рынке.

> Аналитика превращает разрозненные данные в структурированную и содержательную информацию, на базе которой принимаются решения.

Задача аналитиков и менеджеров — не просто интерпретировать информацию, а предлагать обоснованные решения, выявлять тенденции, определять взаимосвязи между различными типами данных и классифицировать их в соответствии с целями и спецификой проекта. Используя инструменты визуализации и методы статистического анализа, они превращают данные в стратегический актив компании.

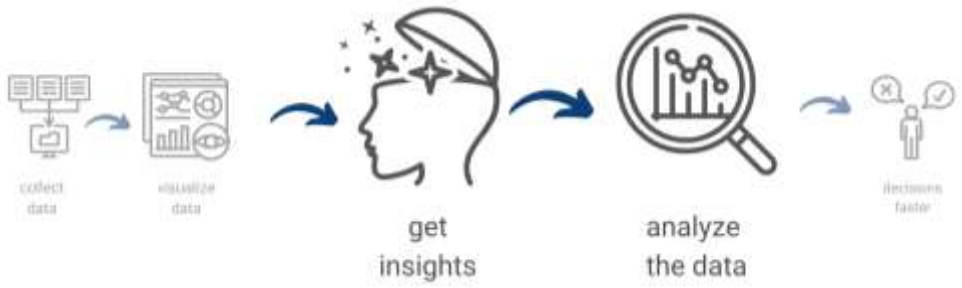


Рис. 7.1-7 Именно анализ данных в конечном итоге превращает собранную информацию в источник для принятия решений.

Для того чтобы принимать действительно обоснованные решения в процессе аналитики, необходимо научиться правильно формулировать вопросы, которые задаются данным. Качество этих вопросов напрямую влияет на глубину получаемых инсайтов и, как следствие, — на качество управленческих решений.

> *Прошлое существует лишь постольку, поскольку оно присутствует в записях сегодняшнего дня. И то, что представляют собой эти записи, определяется тем, какие вопросы мы задаем. Нет другой истории, кроме этой [134].*
>
> – Джон Арчибальд Уилер, физик 1982

Искусство задавать глубокие вопросы и критически мыслить — важнейший навык в работе с данными. Большинство людей склонны задавать простые, поверхностные вопросы, ответы на которые не требуют значительных усилий. Однако подлинный анализ начинается с содержательных и

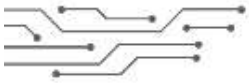
продуманных вопросов, способных раскрыть скрытые взаимосвязи и причинно-следственные зависимости в информации, которые могут скрываться за несколькими слоями рассуждений.

> Согласно исследованию "Трансформация, управляемая данными: Ускорение в масштабе уже сейчас" (BCG, 2017) [135], успешная цифровая трансформация требует инвестиций в аналитические способности, программ управления изменениями и согласования бизнес-целей с ИТ-инициативами. Компания которые создает культуру, ориентированную на данные, должна инвестировать в возможности использования аналитических данных и запускать программы управления изменениями, чтобы привить новое мышление, поведение и способы работы.

Без инвестиций в развитие аналитической культуры, совершенствование инструментов работы с данными и обучение специалистов, компании и в будущем рискуют принимать решения на основе устаревшей или неполной информации — либо полагаясь на субъективные мнения HiPPO-менеджеров (Рис. 2.1-9).

Осознание актуальности и необходимости постоянного обновления аналитики и информационных панелей неизбежно приводит руководство к пониманию важности автоматизации аналитических процессов. Автоматизация повышает скорость принятия решений, снижает влияние человеческого фактора и обеспечивает актуальность данных. В условиях экспоненциального роста объёмов информации скорость становится не просто конкурентным преимуществом, а ключевым фактором устойчивого успеха.

Автоматизация процессов анализа и обработки данных в целом неразрывно связана с темой ETL (Extract, Transform, Load). Точно так же, как в процессе автоматизации нам необходимо преобразовать данные, в процессе ETL данные извлекаются из различных источников, преобразуются в соответствии с необходимыми требованиями и загружаются в целевые системы для дальнейшего использования.



ГЛАВА 7.2.
ПОТОК ДАННЫХ БЕЗ РУЧНЫХ УСИЛИЙ: ЗАЧЕМ НУЖЕН ETL

Автоматизация ETL: снижение затрат и ускорение работы с данными

Когда ключевые показатели эффективности (KPI) перестают расти, несмотря на увеличение объёмов данных и численности команды, руководство компаний неизбежно приходит к осознанию необходимости автоматизации процессов. Рано или поздно это понимание становится стимулом для запуска комплексной автоматизации, основной целью которой является снижение сложности процессов, ускорение обработки и уменьшение зависимости от человеческого фактора.

> Согласно исследованию McKinsey «Как построить архитектуру данных, чтобы стимулировать инновации - сегодня и завтра» (2022) [136], компании, использующие потоковые архитектуры данных, получают значительное преимущество, поскольку они могут анализировать информацию в реальном времени. Потоковые технологии позволяют напрямую анализировать сообщения в реальном времени и применять предиктивное обслуживание на производстве благодаря анализу данных датчиков в реальном времени.

> Упрощение процесса — это автоматизация, при которой традиционные функции ручного управления заменяются алгоритмами и системами.

Вопрос автоматизации, а точнее, "минимизации роли человека в обработке данных", является необратимым и крайне чувствительным процессом для каждой компании. Специалисты в любой профессиональной сфере часто не решаются полностью раскрыть свои методы и тонкости работы коллегам-оптимизаторам, осознавая риск потери работы в стремительно развивающейся технологической среде.

> *Если хочешь нажить себе врагов, попробуй что-то изменить [137].*
>
> — Вудро Вильсон, выступление на конгрессе продавцов, Детройт, 1916 год

Несмотря на очевидные преимущества автоматизации, в повседневной практике многих компаний по-прежнему сохраняется высокая доля ручного труда, особенно в сфере работы с инженерными данными. Чтобы наглядно продемонстрировать текущую ситуацию, рассмотрим типичный пример последовательной обработки данных в рамках подобных процессов.

Ручной порядок работы с данными можно проиллюстрировать на примере взаимодействия с информацией, полученной из баз данных CAD. Традиционная обработка данных ("ручной" ETL-процесс) в отделах CAD (BIM) для создания атрибутивных таблиц или создания документации на основе проектных данных происходит в следующем порядке (Рис. 7.2-1):

1. Ручное **извлечение (Extract)**: пользователь вручную открывает проект - путем запуска приложения CAD (BIM) (Рис. 7.2-1 шаг 1).
2. **Верификация**: на следующем этапе обычно запускается вручную несколько плагинов или

вспомогательных приложений, предназначенных для подготовки данных и оценки их качества (Рис. 7.2-1 шаг 2-3).

3. Ручное **преобразование (Transform):** после подготовки начинается обработка данных, которая требует ручного управления различными программными инструментами в которых данных подготавливаются к выгрузке (Рис. 7.2-1 шаг 4).
4. Ручная **выгрузка (Load):** ручная выгрузка преобразованных данных во внешние системы, форматы данных и документы (Рис. 7.2-1 шаг 5).

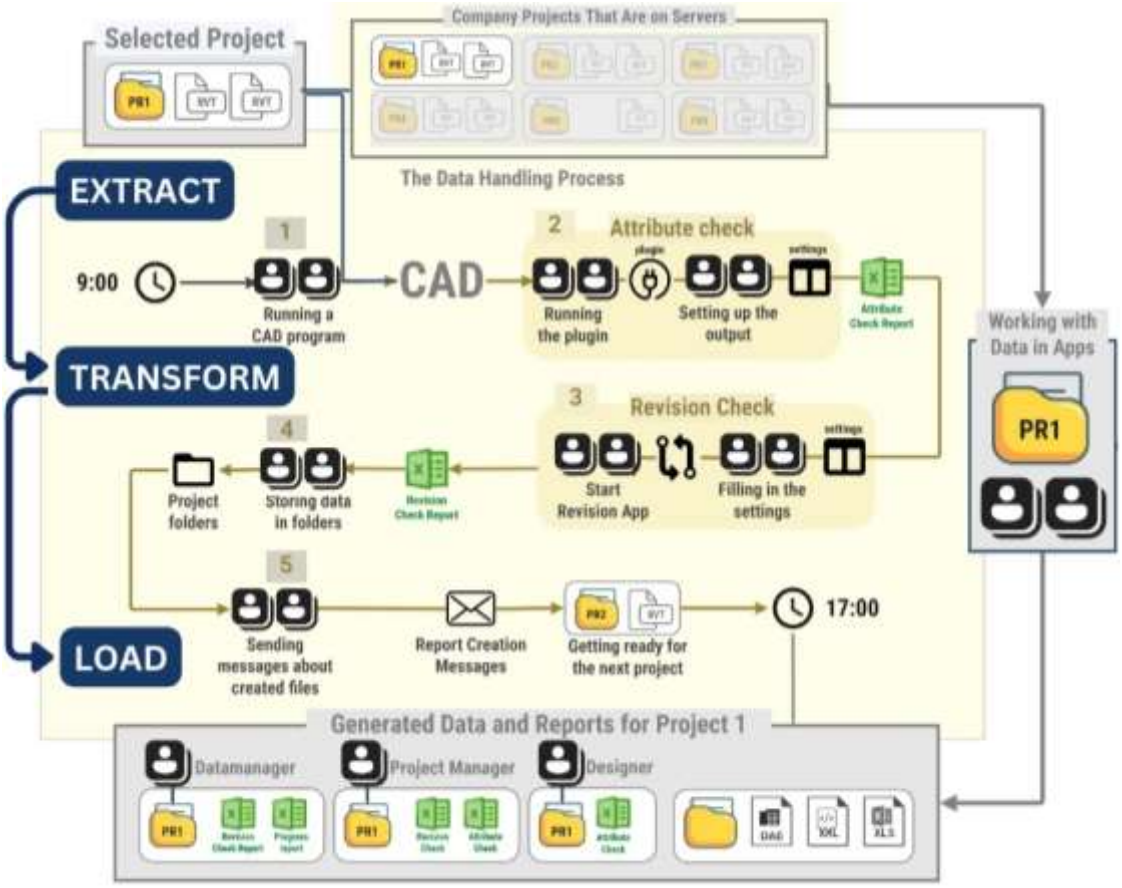


Рис. 7.2-1 Традиционная ручная ETL обработка ограничена желаниями и физическими возможностями отдельного технического специалиста.

Подобный рабочий процесс является примером классического ETL-процесса - извлечения, преобразования и загрузки (ETL). В отличие от других отраслей, где автоматические ETL-пайплайны давно стали стандартом, в строительной отрасли всё ещё преобладает ручной труд, замедляющий процессы и увеличивающий издержки.

ETL (Extract, Transform, Load) — это процесс извлечения данных из различных источников, их преобразования в нужный формат и загрузки в целевую систему для дальнейшего анализа и использования.

ETL — это процесс, обозначающий три ключевых компонента обработки данных: Extract, Transform и Load (Рис. 7.2-2):

- **Extract** — извлечение данных из разных источников (файлы, базы, API).
- **Transform** — очистка, агрегирование, нормализация и логическая обработка данных.
- **Load** — загрузка структурированной информации в хранилище данных, отчёт или BI-систему.

Ранее в книге концепция ETL затрагивалась лишь эпизодически: при преобразовании неструктурированного отсканированного документа в структурированный табличный формат (Рис. 4.1-1), в контексте формализации требований, позволяющей систематизировать восприятие как жизненных, так и деловых процессов (Рис. 4.4-20), а также в рамках автоматизации проверки данных и обработки данных из CAD решений. Теперь рассмотрим ETL подробнее в контексте типовых рабочих процессов.

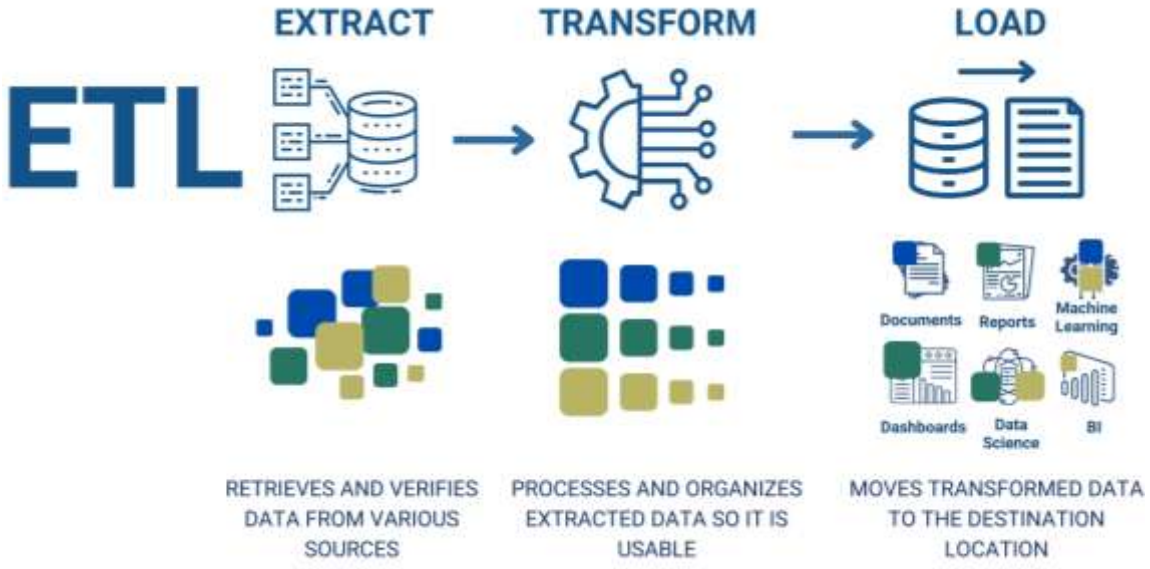


Рис. 7.2-2 ETL автоматизирует повторяющиеся задачи обработки данных.

Ручной или полуавтоматический ETL-процесс предполагает наличие менеджера или технического специалиста, который управляет всеми шагами вручную — от сбора данных до генерации отчётов. Такой процесс занимает значительное время, особенно в условиях ограниченного рабочего дня (например, с 9:00 до 17:00).

Часто компании стремятся решить проблему низкой эффективности и низкой скорости путём покупки модульных интегрированных решений (ERP, PMIS, CPM, CAFM и др.), которые затем дорабатываются внешними вендорами и консультантами. Но такие вендоры и сторонние разработчики часто становятся критической точкой зависимости: их технические ограничения напрямую влияют на производительность всей системы и бизнеса в целом, что подробно описывалось в предыдущих главах про проприетарные системы и форматы. Про проблемы которые создаются фрагментацией и зависимостью, мы подробно говорили в главе «Как строительный бизнес тонет в хаосе данных».

> Если компания не готова к внедрению крупной модульной платформы одного из вендоров, она начинает искать альтернативные пути автоматизации. Один из них — разработка собственных модульных открытых ETL-конвейеров, где каждый этап (извлечение, преобразование, валидация, загрузка) реализуется в виде скриптов, исполняемых по расписанию.

В автоматизированной версии того же рабочего процесса ETL (Рис. 7.2-1) процесс работы выглядит как модульный код, который начинается с обработки данных и перевода их в открытую структурированную форму. После получения структурированных данных автоматически, по расписанию, запускаются различные сценарии или модули для проверки изменений, преобразования и отправки сообщений (Рис. 7.2-3).

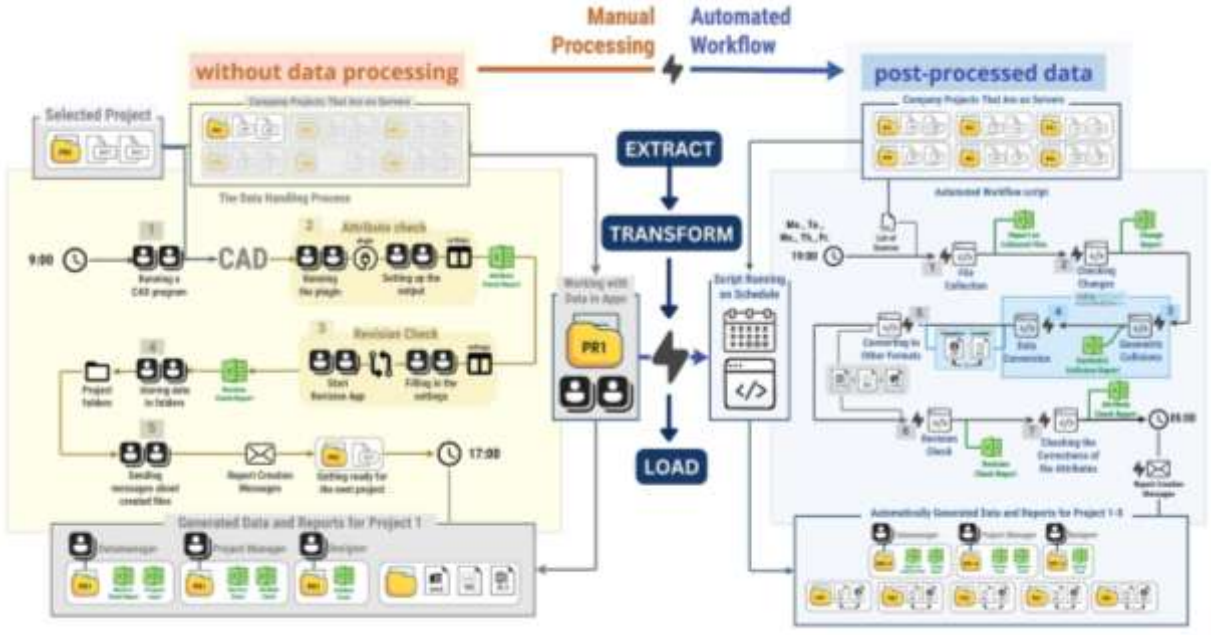


Рис. 7.2-3 Слева ручная обработка, справа - автоматический процесс, который в отличие от традиционной ручной обработки, не ограничен возможностями пользователя.

В автоматизированном рабочем процессе обработка данных упрощается за счет предварительной обработки данных ET(L): структурирования и унификации.

> При традиционных методах обработки специалисты работают с данными «как они есть» — в том виде, в котором они извлекаются из систем или программного обеспечения. В автоматизированных процессах, напротив, данные часто сначала проходят через ETL-пайплайн, где приводятся к согласованной структуре и формату, пригодному для дальнейшего использования и анализа.

Рассмотрим практический пример ETL, демонстрирующий процесс проверки таблиц данных, описанный в главе "Проверка данных и результаты проверки" (Рис. 4.4-13). Для этого мы используем библиотеку Pandas в сочетании с LLM для процессов автоматизированного анализа и обработки данных.

ETL Extract: сбор данных

Первый этап процесса ETL — извлечение (Extract) — начинается с написания кода для сбора массивов данных, подлежащих последующей проверке и обработке. Для этого просканируем все папки рабочего сервера, соберем документы определенного формата и содержания, а затем преобразуем их в структурированную форму. Этот процесс подробно рассмотрен в главах "Преобразование неструктурированных и текстовых данных в структурированную форму" и "Преобразование данных CAD (BIM) в структурированную форму" (Рис. 4.1-1 - Рис. 4.1-12).

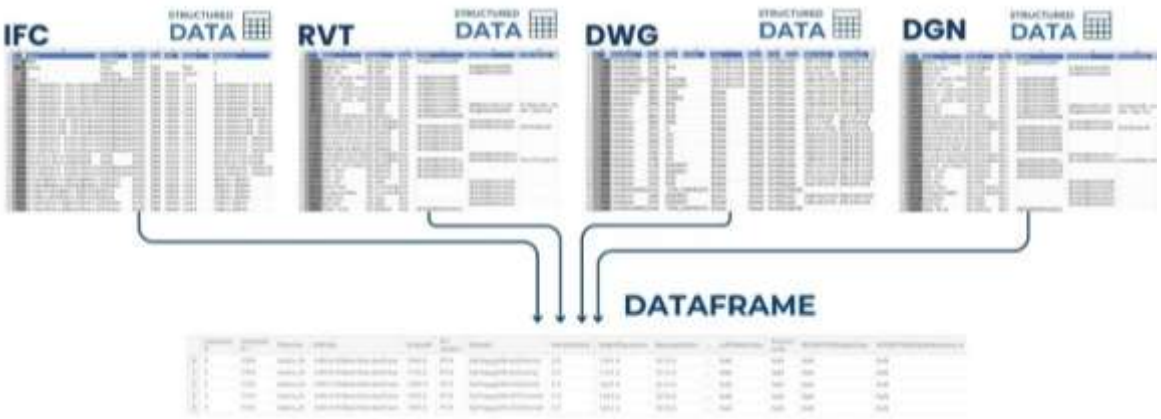


Рис. 7.2-4 Преобразование данных CAD (BIM) в один большой фрейм данных, который будет содержать все разделы проекта.

В качестве наглядного примера используем на шаге загрузки данных Extract и получения таблицы всех CAD- (BIM-) проектов (Рис. 7.2-4) используются конвертеры с поддержкой обратного инжиниринга [138] для форматов RVT и IFC, чтобы получить структурированные таблицы из всех проектов и объединить их в одну большую таблицу DataFrame.

Рис. 7.2-5 Преобразование при помощи Python кода и SDK инструмента обратного инжиниринга файлов RVT и IFC в один большой структурированный (df) DataFrame.

В Pandas DataFrame можно загружать данные из различных источников, включая текстовые файлы CSV, таблицы Excel, JSON- и XML-файлы, форматы для хранения больших объемов данных, такие как Parquet и HDF5, а также из баз данных MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle и других. Кроме того, Pandas поддерживает загрузку данных из API, веб-страниц, облачных сервисов и систем хранения, таких как Google BigQuery, Amazon Redshift и Snowflake.

- Чтобы написать код для подключения и сбора информации из баз данных, отправьте подобный текстовый запрос в LLM чат (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любой другой):

> Пожалуйста, напиши пример подключения к MySQL и преобразования данных в DataFrame ⏎

Ответ LLM:

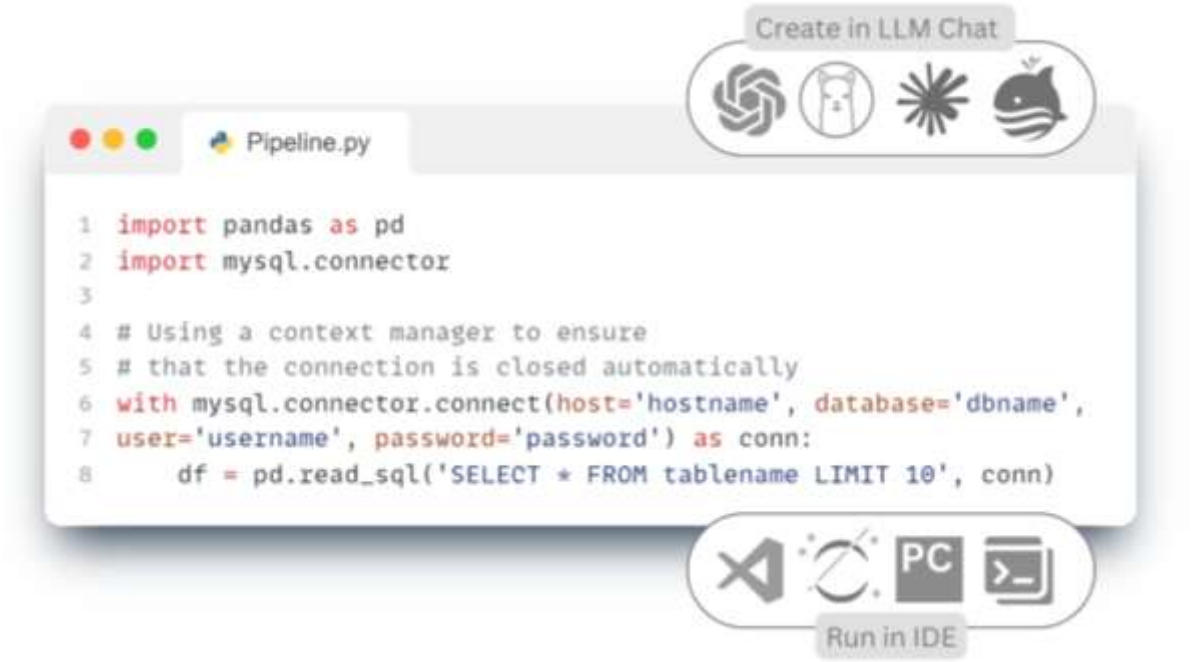


Рис. 7.2-6 Пример подключения через Python к базе данных MySQL и импорта данных из базы MySQL в DataFrame.

Полученный код (Рис. 7.2-5, Рис. 7.2-6) можно запустить в одной из популярных IDE (интегрированная среда разработки), про которые мы говорили выше, в оффлайн режиме: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с плагином PyDev, Thonny, Wing IDE, IntelliJ IDEA с плагином Python, JupyterLab или популярных онлайн-инструментах: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Загрузив мультиформатные данные в переменную "df" (Рис. 7.2-5 – 25-я строка; Рис. 7.2-6 – 8-я строка), мы конвертировали данные в формат Pandas DataFrame — одной из самых популярных структур для обработки данных, представляющей собой двухмерную таблицу со строками и столбцами. О других форматах хранения, используемых в ETL-Pipelines, таких как Parquet, Apache ORC, JSON, Feather, HDF5, а также о современных хранилищах данных мы подробнее поговорим в главе "Хранение и управление данными в строительной отрасли" (Рис. 8.1-2).

После этапа извлечения и структурирования данных (Extract) формируется единый массив информации (Рис. 7.2-5, Рис. 7.2-6), готовый к дальнейшей обработке. Однако, прежде чем загружать эти данные в целевые системы или использовать для анализа, необходимо убедиться в их качестве, целостности и соответствии заданным требованиям. Именно на этом этапе осуществляется трансформация данных (Transform) — ключевой шаг, обеспечивающий надёжность последующих выводов и решений.

ETL Transform: применение правил проверки и трансформации

На этапе Transform выполняется обработка и преобразование данных. Этот процесс может включать проверку корректности, нормализацию, заполнение пропущенных значений и валидацию с помощью автоматизированных инструментов.

> Согласно Исследованию PwC „Data-Driven. Что нужно студентам, чтобы добиться успеха в быстро меняющемся мире бизнеса” (2015) [9], современные аудиторские компании отходят от выборочной проверки данных и переходят к анализу массива информации с использованием автоматизированных инструментов. Такой подход позволяет не только выявлять несоответствия в отчетности, но и предлагать рекомендации по оптимизации бизнес-процессов.

В строительстве аналогичные методы могут применяться, например, для автоматической валидации проектных данных, контроля качества строительства и оценки эффективности работы подрядчиков. Одним из инструментов, позволяющих автоматизировать и ускорить обработку данных, является применение регулярных выражений (RegEx) на этапе преобразования данных (Transform) в процессе ETL. RegEx позволяет эффективно проверять строки данных, выявлять несоответствия и обеспечивать целостность информации с минимальными затратами ресурсов. Подробнее о RegEx (Рис. 4.4-7) мы говорили в главе "Перевод требований в структурированную форму".

Рассмотрим практический пример: в системе управления объектами недвижимости (RPM) менеджер устанавливает требования к ключевым атрибутам объектов (Рис. 7.2-7). На этапе трансформации необходимо выполнить валидацию следующих параметров:

- проверку форматов идентификаторов объектов (атрибут "ID")
- контроль значений гарантийного срока замены (атрибут "Гарантийный срок")
- проверку цикла замены элементов (атрибут "Требования к обслуживанию")

Рис. 7.2-7 Проверка качества начинается с установки требований к атрибутам и их граничным значениям.

Чтобы установить граничные значения для проверки параметров, например предположим, что из

нашего опыта мы знаем, что допустимые значения для атрибута "ID" могут включать только строковые значения "W-NEW", "W-OLD1" или "D-122" или аналогичные значения, где первым символом является буква, за которой следует тире, а затем три буквенных символа 'NEW', 'OLD' или любое трехзначное число (Рис. 7.2-7). Для валидации этих идентификаторов может быть использовано следующее регулярное выражение (RegEx):

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Этот шаблон позволяет убедиться, что все идентификаторы в данных соответствуют заданным критериям. Если какое-либо значение не проходит проверку, система фиксирует ошибку. Чтобы создать Python-код для преобразования данных и использовать полученные данные для создания таблицы результатов, просто сформулируем запрос в LLM чат.

- Текстовый запрос в LLM:

> Напиши код для проверки столбцов DataFrame с помощью регулярных выражений, которые проверяют идентификаторы в формате 'W-NEW' или 'W-OLD' через RegEx, энергоэффективность с буквами от 'A' до 'G', гарантийный срок и цикл замены с числовыми значениями в годах ⏎

Ответ LLM:

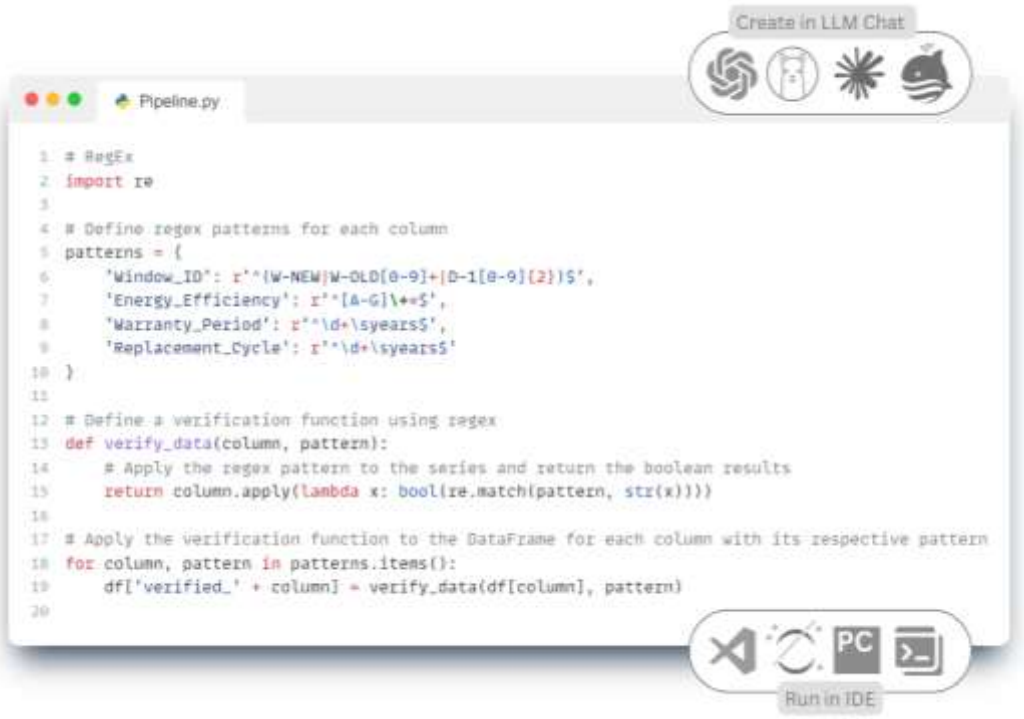


Рис. 7.2-8 Код автоматизация процесса проверки путем применения шаблонов RegEx к колонкам параметров датафрейма.

Приведенный, автоматически созданный Python-код (Рис. 7.2-8), использует библиотеку “re“ (регулярные выражения RegEx) для определения функции, которая проверяет каждый атрибут элемента данных в DataFrame. Для каждого указанного столбца (атрибута) функция применяет шаблон RegEx для проверки соответствия каждой записи ожидаемому формату и добавляет результаты в виде новых значений (False/True) в новый атрибут-столбец DataFrame.

> Подобная автоматизированная проверка обеспечивает формальное соответствие данных установленным требованиям и может использоваться как часть системы контроля качества на этапе трансформации.

После успешного выполнения этапа Transform и проверки качества данные готовы к загрузке в целевые системы. Преобразованные и проверенные данные можно выгрузить в CSV, JSON, Excel, базы данных и другие форматы для последующего использования. Также в зависимости от задачи результаты могут быть представлены в отчетах, графиках или аналитических дэшбордах.

ETL Load: Визуализация результатов в виде диаграмм и графиков

После завершения этапа Transform, когда данные приведены к структурированной форме и проверены, наступает заключительная стадия — Load, на которой данные могут быть как загружены в

целевую систему, так и визуализированы для анализа. Визуальное представление данных позволяет оперативно выявить отклонения, проанализировать распределения и донести ключевые выводы до всех участников проекта, в том числе не обладающих технической подготовкой.

Вместо представления информации в виде таблиц и цифр мы можем использовать инфографику, графики и приборные панели (dashboards). Одним из наиболее распространённых и гибких инструментов для визуализации структурированных данных на языке Python является библиотека Matplotlib (Рис. 7.2-9, Рис. 7.2-10). Она позволяет создавать статические, анимированные и интерактивные графики, и поддерживает широкий спектр типов диаграмм.

- Чтобы визуализировать результаты проверки атрибутов из системы RPM (Рис. 7.2-7), можно воспользоваться следующим запросом к языковой модели:

> Напиши код для визуализации данных DataFrame, приведенных выше (Рис. 7.2-7), с гистограммой для результатов, чтобы отобразить частоту ошибок в атрибута ⏎

- Ответ LLM в виде кода и готовой визуализацией напрямую в чате LLM результатов выполнения кода:

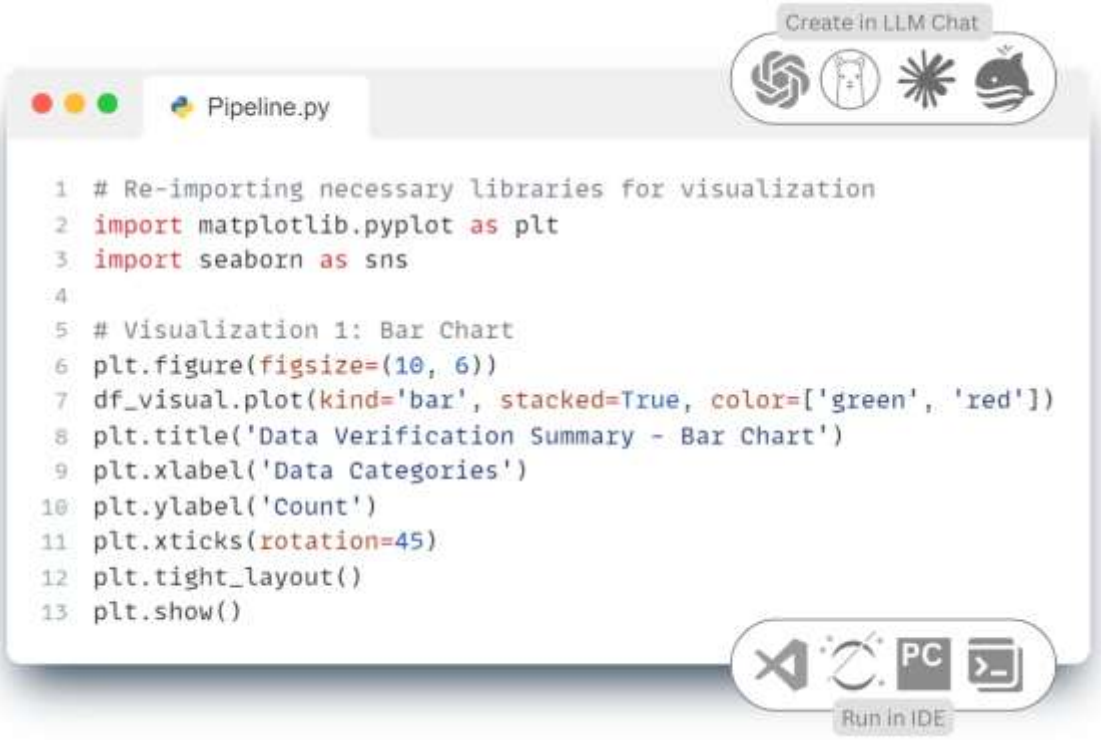


Рис. 7.2-9 Визуализация результатов этапа Transform по проверки значений атрибутов из системы RPM (Рис. 7.2-7) в виде гистограммы на этапе Load.

- Существует множество открытых и бесплатных библиотек визуализации, позволяющих представлять структурированные данные в различных форматах. Продолжим визуализацию результатов другим видом графика следующим промптом в чате:

> Изобразите те же данные в виде графика из линий ⏎

- Ответ LLM:

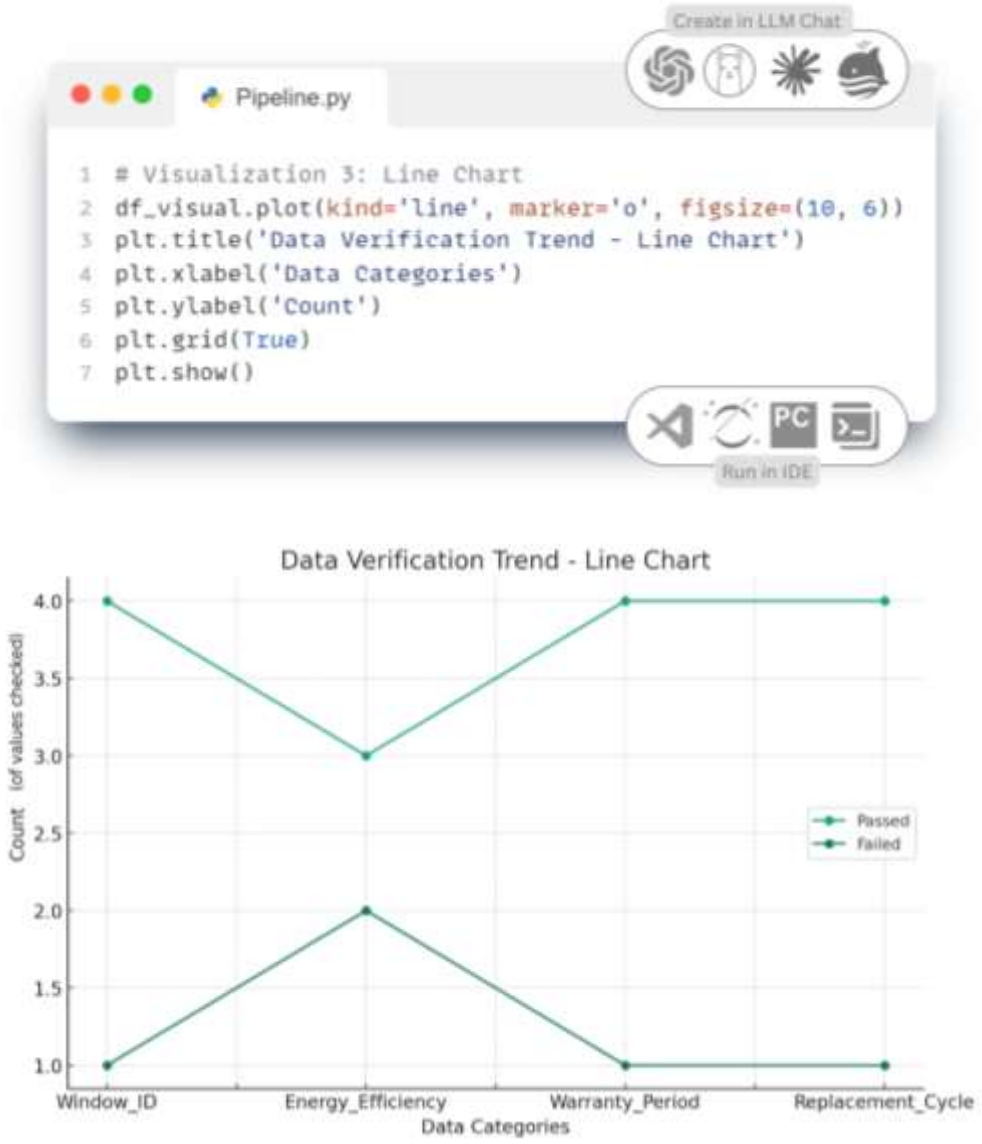


Рис. 7.2-10 Визуализация данных проверки (Рис. 7.2-8) в виде линейной диаграммы, полученной с помощью библиотеки Matplotlib.

Существует множество открытых и бесплатных библиотек визуализации, таких как:

- Seaborn — для статистических графиков (Рис. 7.2-11)
- Plotly — для интерактивных web-визуализаций (Рис. 7.2-12, Рис. 7.1-6)
- Altair — для декларативной визуализации
- Dash или Streamlit — для создания полноценных дашбордов

При этом знание конкретных библиотек для визуализации не является обязательным – современные инструменты, включая LLM, позволяют автоматически генерировать код для построения графиков и целых приложений, основываясь на описании задачи.

Выбор инструмента зависит от задач проекта: будь то отчёт, презентация или онлайн-панель мониторинга. Например библиотека с открытым исходным кодом Seaborn особенно хороша для работы с категориальными данными, помогая выявлять закономерности и тенденции.

- Чтобы увидеть в работе библиотеку Seaborn, вы можете или попросить LLM использовать напрямую нужную библиотеку или отправьте подобный текстовый запрос в продолжении работы с LLM:

> Покажите тепловую карту для результатов ⏎

- Ответ LLM в виде кода и готового графика, код построения которого можно теперь скопировать в IDE, а сам график скопировать или сохранить для вставки в документ:

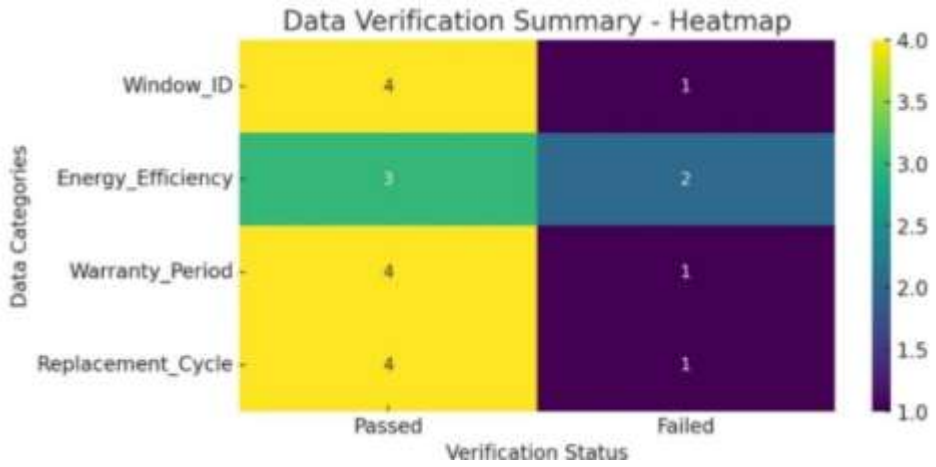


Рис. 7.2-11 Визуализация результатов проверки (Рис. 7.2-8) данных с помощью библиотеки Seaborn.

Для тех, кто предпочитает интерактивный подход, существуют инструменты, позволяющие создавать динамичные диаграммы и панели с возможностью взаимодействия. Библиотека Plotly (Рис. 7.1-6, Рис. 7.2-12) предлагает возможность создания высоко интерактивных диаграмм и панелей, которые можно встраивать в веб-страницы и предоставлять пользователю возможность взаимодействия с данными в режиме реального времени.

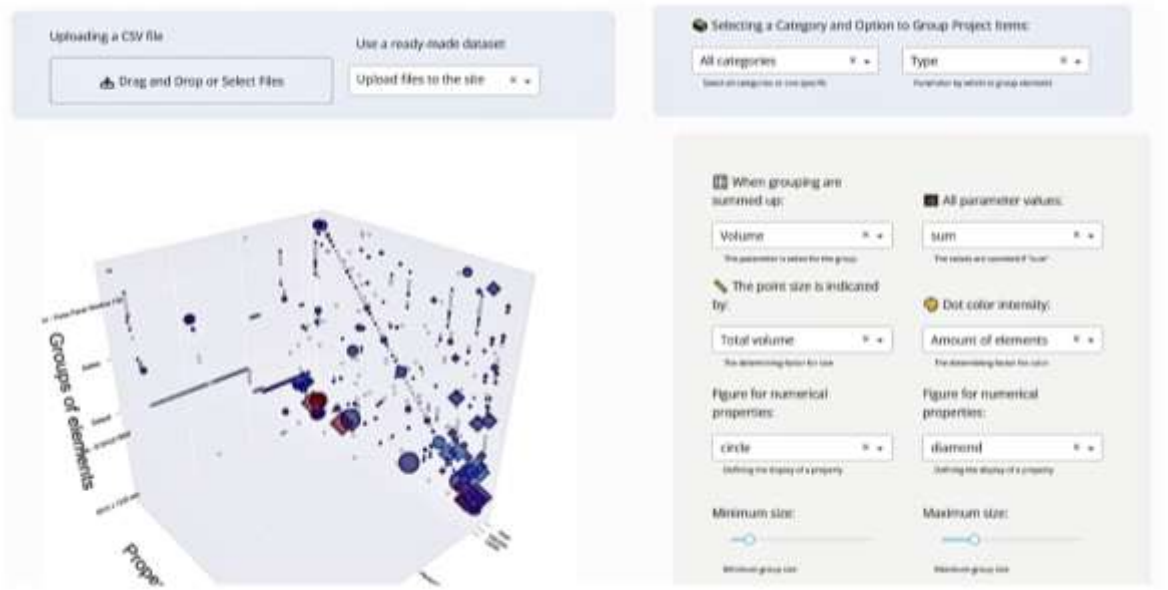


Рис. 7.2-12 Интерактивная 3D-визуализация атрибутов элементов из CAD- (BIM-) проекта с помощью библиотеки Plotly.

Специализированные открытые библиотеки Bokeh, Dash и Streamlit обеспечивают удобный способ представления данных, не требуя глубоких знаний в веб-разработке. Bokeh подходит для сложных интерактивных графиков, Dash используется для построения полноценных аналитических дашбордов, а Streamlit позволяет быстро создавать веб-приложения для анализа данных.

Благодаря подобным инструментам визуализации разработчики и аналитики могут эффективно распространять результаты среди коллег и заинтересованных сторон, обеспечивая интуитивное взаимодействие с данными и упрощая процесс принятия решений.

ETL Load: Автоматическое создание PDF-документов

На этапе загрузки данные можно не только визуализировать данные, выгружать их в таблицы или базы данных, но и автоматически формировать отчеты, включая в них необходимые графики, диаграммы и ключевые аналитические показатели, которые необходимо получить менеджеру или специалисту, ожидающему результатов проверки. Автоматизированные отчеты могут содержать как комментарии и текстовую интерпретацию данных, так и наглядные элементы визуализации –таблицы, графики.

- Для создания PDF-отчета с гистограммой (Рис. 7.2-9) и описанием анализа по результатам проверки, которую мы провели в предыдущих главах, достаточно сформулировать запрос в продолжении диалога с LLM, например:

> Напиши код для создания PDF-файла с гистограммой и описанием результатов проверки данных, которые были получены выше (в чате), и напиши предупреждение в виде текста о том, что некоторые категории не прошли проверку и что необходимо заполнить недостающие данные ⏎

- Ответ LLM в виде кода и готового PDF с результатами:

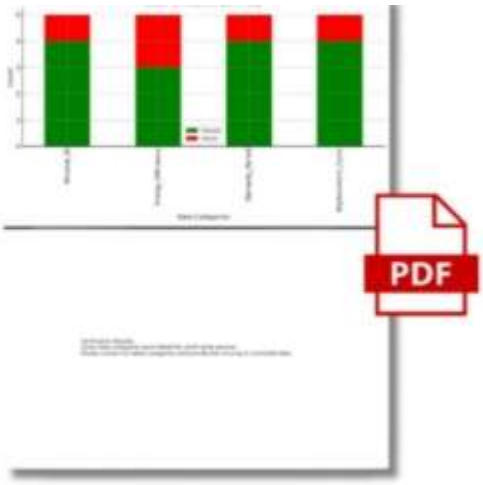


Рис. 7.2-13 Автоматический код создаёт PDF-документ, содержащий гистограмму с тестовыми данными и текст с результатами проверки.

Автоматически написанное решение всего из 20 строк кода с помощью LLM мгновенно создает нужный PDF (или DOC) документ с визуализацией в виде гистограммы атрибутов (Рис. 7.2-13), показывающей количество данных, прошедших и не прошедших проверку, а также с добавлением текстового блока краткого описания результатов и рекомендациям по дальнейшим действиям.

Автоматизированная генерация документов — ключевой элемент стадии Load, особенно в условиях проектной деятельности, где критически важна скорость подготовки отчётности и её точность.

ETL Load: автоматическая генерация документов с FPDF

Автоматизация отчетности на этапе ETL Load — важный этап в обработке данных, особенно когда результаты анализа должны быть представлены в удобной для передачи и восприятия формате. В строительной отрасли это часто актуально для отчетов о ходе работ, статистике проектных данных, отчетов по проверке качества или финансовой документации.

Один из наиболее удобных инструментов для подобных задач — библиотека, с открытым исходным кодом, FPDF, доступная как для Python, так и для PHP.

> Открытая библиотека **FPDF** предоставляет гибкий способ формирования документов через код, позволяя добавлять заголовки, текст, таблицы и изображения. Использование кода вместо ручного редактирования снижает количество ошибок и ускоряет процесс подготовки отчетов в PDF формате.

Одним из ключевых этапов создания PDF-документа является добавление заголовков и основного текста в виде комментариев или описания. Однако при формировании отчета важно не просто добавить текст, но и грамотно его структурировать. Заголовки, отступы, интервалы между строками

— все это влияет на читаемость документа. Используя FPDF, можно задавать параметры форматирования, управлять расположением элементов и настраивать стиль документа.

FPDF крайне похож по принципу работы на HTML. Те, кто уже знаком с HTML, используя FPDF, без труда смогут генерировать PDF-документы любой сложности, так как структура кода во многом напоминает разметку HTML: добавление заголовков, текста, изображений и таблиц происходит схожим образом. Тем, кто не знаком с HTML, не стоит беспокоиться — можно использовать LLM, который мгновенно поможет составить код для генерации нужного оформления документа.

- Следующий пример демонстрирует, как сформировать отчет с заголовком и основным текстом. Выполнение этого кода в любом IDE с поддержкой Python создает PDF-файл, содержащий нужный заголовок и текст:

```
from fpdf import FPDF   # Импортируем библиотеку FPDF
pdf = FPDF()    # Создаем PDF-документ
pdf.add_page()   # Добавляем страницу

pdf.set_font("Arial", style='B', size=16)  # Устанавливаем шрифт: Arial, жирный,
размер 16
pdf.cell(200, 10, "Отчет по проекту", ln=True, align='C')  # Создаем заголовок и цен-
трируем его
pdf.set_font("Arial", size=12)  # Изменяем шрифт на обычный Arial, размер 12
pdf.multi_cell(0, 10, "Этот документ содержит данные по результатам проверки проектных
файлов...")  # Добавляем многострочный текст
pdf.output(r"C:\reports\report.pdf") # Сохраняем PDF-файл
```

Рис. 7.2-14 При помощи нескольких строк кода на Python мы можем автоматически сгенерировать нужный нам PDF документ с текстом.

При подготовке отчетов важно учитывать, что данные из которых формируются, документ редко остаются статичными. Заголовки, текстовые блоки (Рис. 7.2-14) часто формируются динамически, получая значения ещё на этапе Transform в процессе ETL.

Использование кода позволяет создавать документы, содержащие актуальные сведения: название проекта, дату формирования отчета, а также информацию об участниках или текущем статусе. Использование переменных в коде позволяет автоматически подставлять эти данные в нужные места отчета, полностью исключая необходимость ручного редактирования перед отправкой.

Помимо простого текста и заголовков, в проектной документации таблицы занимают особое место. Практически в каждом документе присутствуют структурированные данные: от описания объектов до результатов проверок. Автоматическое формирование таблиц на основе данных из этапа Transform позволяет не только ускорить процесс подготовки документов, но и минимизировать ошибки при переносе информации. FPDF позволяет вставлять таблицы в PDF-файлы (в виде текста или картинок), задавая границы ячеек, размеры колонок и шрифты (Рис. 7.2-15). Это особенно удобно при работе с динамическими данными, когда количество строк и столбцов может варьироваться в зависимости от задач документа.

Следующий пример показывает, как автоматизировать создание таблиц, например, со списком материалов, сметными расчетами или результатами проверки параметров:

```
data = [
    ["Элемент", "Количество", "Цена"],  # Заголовки столбцов
    ["Бетон", "10 м³", "$500."],  # Данные первой строки
    ["Арматура", "2 т.", "$600"],  # Данные второй строки
    ["Кирпич", "5000 шт.", "$750"]  # Данные третьей строки
]

pdf = FPDF()  # Создаем PDF-документ
pdf.add_page()  # Добавляем страницу
pdf.set_font("Arial", size=12)  # Устанавливаем шрифт

for row in data:  # Перебираем строки таблицы
    for item in row:  # Перебираем ячейки в строке
        pdf.cell(60, 10, item, border=1)  # Создаем ячейку с границей, шириной 60 и высотой 10
    pdf.ln()  # Переходим на следующую строку
pdf.output(r"C:\reports\table.pdf")  # Сохраняем PDF-файл
```

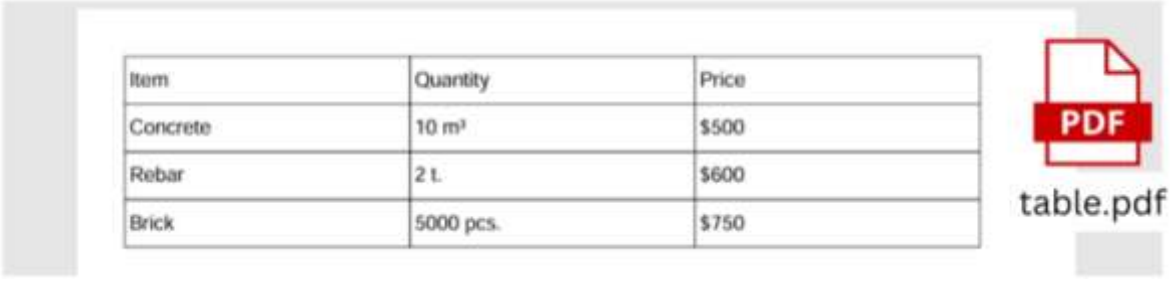


Рис. 7.2-15 Можно автоматически генерировать в PDF не только текст, но и любую табличную информацию, полученную на этапе Transform.

В реальных сценариях отчётности таблицы, как правило, представляют собой динамически формируемую информацию, полученную на этапе трансформации данных. В приведённом примере (Рис. 7.2-15) таблица вставляется в PDF-документ в статичном виде: в словарь data (первая строка кода) были помещены данные для примера, в реальных условиях подобная переменная data заполняется автоматически после например группировки датафрейма.

На практике такие таблицы часто строятся на основе структурированных данных, поступающих из

различных динамичным источников: базы данных, Excel-файлы, API-интерфейсы или результаты аналитических вычислений. Чаще всего на этапе Transform (ETL) данные агрегируются, группируются или фильтруются — и только после этого преобразуются в итоговые в виде графиков или двухмерных таблиц, отображаемые в отчетах. Это значит, что содержимое таблицы может меняться в зависимости от выбранных параметров, периода анализа, проектных фильтров или пользовательских настроек.

> Использование динамических датафреймов и наборов данных на этапе Transform делает процесс формирования отчетов на этапе Load максимально гибким, масштабируемым и легко повторяемым без необходимости ручного вмешательства.

Помимо таблиц и текста FPDF поддерживает также добавление графиков табличных данных, что позволяет встроить в отчет изображения, сгенерированные с помощью Matplotlib или других библиотек визуализации, которые мы рассматривали выше. Дополнить документ при помощи кода можно любыми графиками, диаграммами и схемами.

- Используя Python библиотеку FPDF, добавим в документ PDF график, предварительно сгенерированный с помощью Matplotlib:

```
import matplotlib.pyplot as plt  # Импортируем Matplotlib для создания графиков

fig, ax = plt.subplots()  # Создаем фигуру и оси графика
categories = ["Бетон", "Арматура", "Кирпич"]  # Названия категорий
values = [50000, 60000, 75000]  # Значения по категориям
ax.bar(categories, values)  # Создаем столбчатую диаграмму
plt.ylabel("Стоимость, $.")  # Подписываем ось Y
plt.title("Распределение затрат")  # Добавляем заголовок
plt.savefig(r"C:\reports\chart\chart.png")  # Сохраняем график в виде изображения

pdf = FPDF()  # Создаем PDF-документ
pdf.add_page()  # Добавляем страницу
pdf.set_font("Arial", size=12)  # Устанавливаем шрифт
pdf.cell(200, 10, "График затрат", ln=True, align='C')  # Добавляем заголовок

pdf.image(r"C:\reports\chart\chart.png", x=10, y=30, w=100)  # Вставляем изображение в PDF (x, y - координаты, w - ширина)
pdf.output(r"C:\reports\chart_report.pdf")  # Сохраняем PDF-файл
```

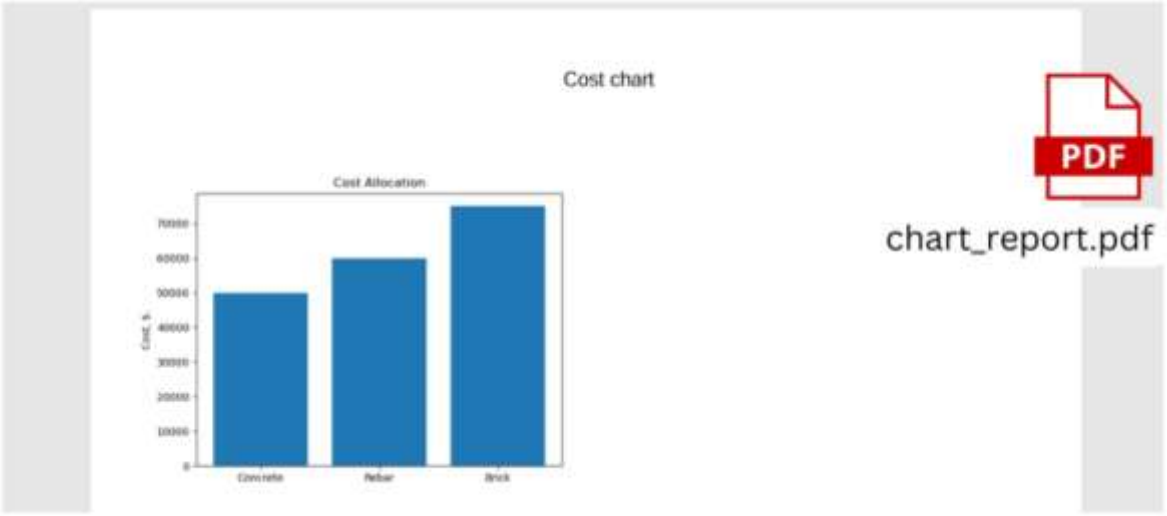


Рис. 7.2-16 При помощи десятка строчек кода можно сгенерировать график, сохранить его, и потом вставить в нужное место документа PDF.

Благодаря FPDF процесс подготовки и логика документов становится прозрачным, быстрым и удобным. Встроенные в код шаблоны позволяют генерировать документы с актуальными данными, исключая необходимость ручного заполнения.

> Используя ETL автоматизацию — вместо трудоемкого оформления отчетов вручную специалисты могут сосредоточиться на анализе данных и принятии решений, а не выборе подходящего инструмента под работу с тем или иным силосом данных с понятным пользовательским интерфейсом.

Таким образом, библиотека FPDF предоставляет гибкий инструмент для автоматизированного создания документов любой степени сложности — от кратких технических отчетов до комплексных аналитических сводок с таблицами и графиками, что позволяет не только ускорить документооборот, но и существенно снизить вероятность ошибок, связанных с ручным вводом и форматированием данных.

ETL Load: Составление отчетов и загрузка в другие системы

На этапе Load были сформированы результаты в виде таблиц, графиков и итоговых PDF-отчётов, подготовленных в соответствии с установленными требованиями. Далее возможен экспорт этих данных в машиночитаемые форматы (например, CSV), что необходимо для интеграции с внешними системами — такими как ERP, CAFM, CPM, BI-платформы и другие корпоративные или отраслевые решения. Кроме CSV, выгрузка может производиться в форматы XLSX, JSON, XML или напрямую в базы данных, поддерживающие автоматический обмен информацией.

- Чтобы сформировать соответствующий код для автоматизации этапа Load, достаточно задать запрос в LLM-интерфейс, например: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude или QWEN:

> Напиши код для создания отчета о результатах проверки данных в DataFrame, в котором столбцы с префиксом 'verified_' подсчитываются, переименовываются в 'Passed' и 'Failed', пропущенные значения заменяются на 0, а затем экспортируются в CSV-файл только те строки, которые прошли все проверки.

Ответ LLM:

Рис. 7.2-17 Проверенные данные, полученные на этапе Transform из итогового датафрейма экспортируются в CSV-файл для интеграции с другими системами.

В приведенном коде (Рис. 7.2-17) реализуется завершающий этап ETL-процесса — Load, в ходе которого проверенные данные сохраняются в формате CSV, совместимом с большинством внешних систем и баз данных. Таким образом, мы завершили полный цикл ETL-процесса, включающий в себя извлечение, трансформацию, визуализацию, документирование и экспорт данных в нужные нам системы и форматы, что обеспечивает воспроизводимость, прозрачность и автоматизацию работы с информацией.

Разработанный ETL-конвейер (pipeline) можно использовать как для обработки единичных проектов, так и для масштабного применения — при анализе сотен и тысяч входящих данных в виде документов, изображений, сканов, CAD-проектов, облаков точек, PDF-файлов или иных источников, поступающих из распределённых систем. Возможность полной автоматизации процесса превращает ETL не просто в инструмент технической обработки, а в основу информационной инфраструктуры цифрового строительства.

ETL с помощью LLM: Визуализация данных из PDF-документов

Пришло время перейти к сборке полноценного ETL-процесса, охватывающего все ключевые этапы работы с данными в одном сценарии — извлечение, трансформацию и загрузку. Построим автоматический ETL-Pipeline, позволяющий обрабатывать PDF-документы без ручной работы — извлекать из документов данные, визуализировать, анализировать и передавать в другие системы.

Процесс ETL в нашем примере будет описываться через промпты, которые должны будут объяснить языковой модели (LLM) все процессы ETL с описанием конечного результата, который необходимо получить. В данном случае задача состоит в том, чтобы найти все PDF-файлы в указанной папке и её подпапках, извлечь из них релевантную информацию — например, наименования материалов, их количество и стоимость — и представить результат в виде структурированной таблицы (DataFrame) для последующего анализа.

- Первый текстовый запрос в LLM на автоматическое извлечение данных из множества PDF документов и создание датафрейма данных для этапа Extract:

> Напиши код для извлечения информации о материалах из PDF-файлов в заданной папке и ее подпапках. Данные в PDF включают название материала, количество и стоимость. Результат необходимо сохранить в DataFrame ⏎

- Ответ LLM:

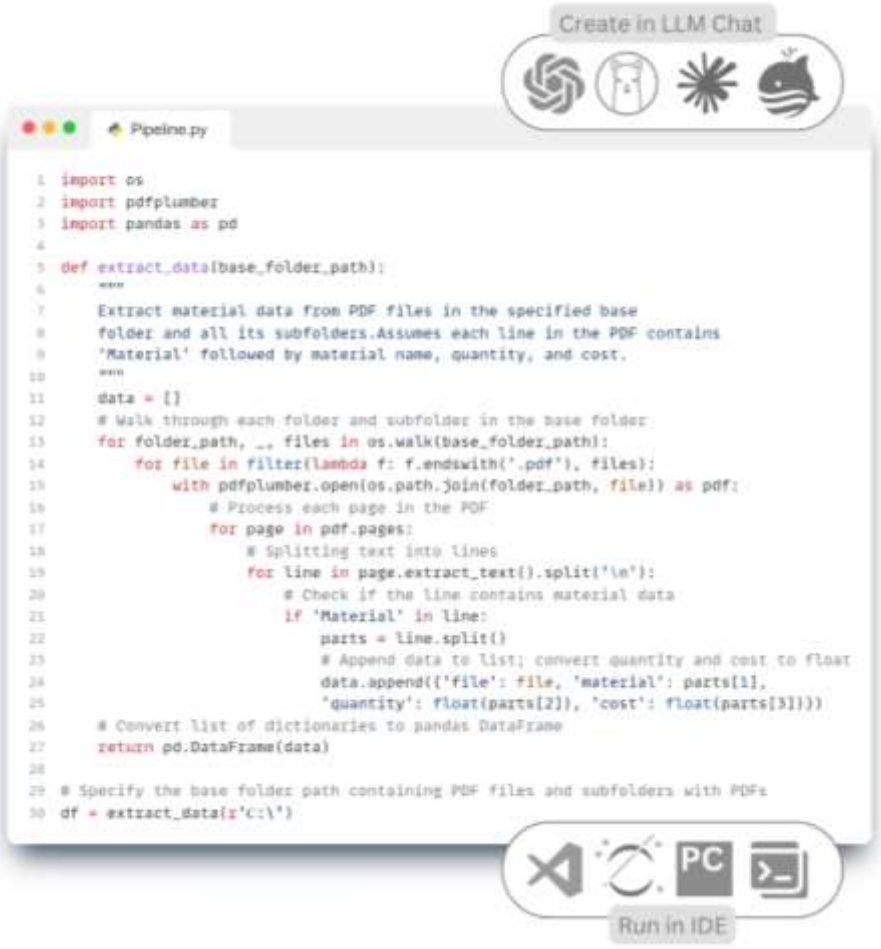


Рис. 7.2-18 LLM создаёт Python-код для извлечения данных из PDF-файлов в определенной папке и всех ее подпапках.

Ответ LLM (Рис. 7.2-18) — это готовый Python-скрипт, который автоматически обходит все папки, открывает найденные PDF-файлы, извлекает из них текстовую информацию и преобразует её в таблицу. Полученный в чате код можно запустить в одной из популярных IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с плагином PyDev, Thonny, Wing IDE, IntelliJ IDEA с плагином Python, JupyterLab или популярных онлайн-инструментах Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Как вариант, чтобы упростить процесс сбора, вместо копирования кода из LLM и использования кода в IDE мы можем также загрузить десяток PDF-файлов напрямую в чат LLM (Рис. 7.2-19) и получить на выходе таблицу, без необходимости видеть код или запускать его. Результатом выполнения этого кода будет таблица с выбранными нами атрибутами.

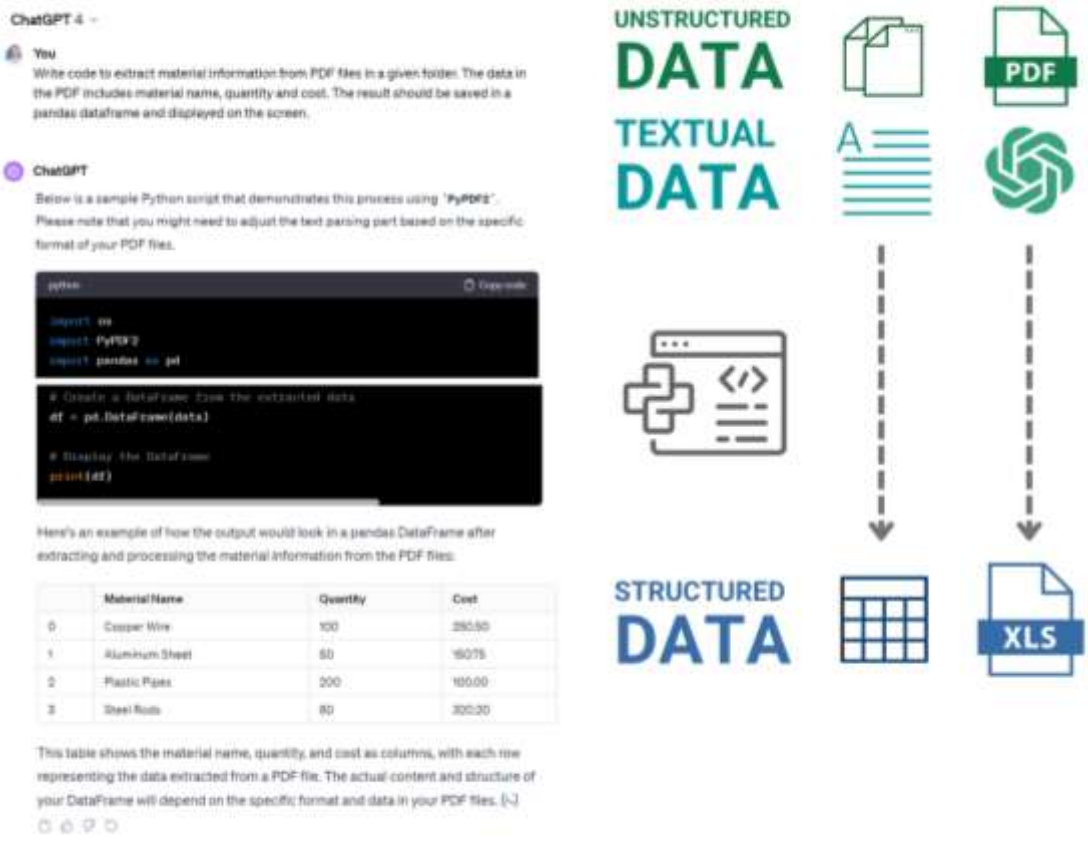


Рис. 7.2-19 Результат выполнения кода в LLM, которая извлекает данные из PDF-файлов в структурированном виде датафрейма с выбранными атрибутами.

На следующем этапе просим языковую модель по полученным данным — например, сравнить стоимость и объёмы использования материалов и создать несколько примеров визуализации, которые послужат основой для последующего анализа.

- Попроси в продолжении чата с LLM построить несколько графиков из таблиц, которые были получены на этапе Transform (Рис. 7.2-18):

Визуализируй общую стоимость и количество каждого материала из DataFrame (Рис. 7.2-18) ⏎

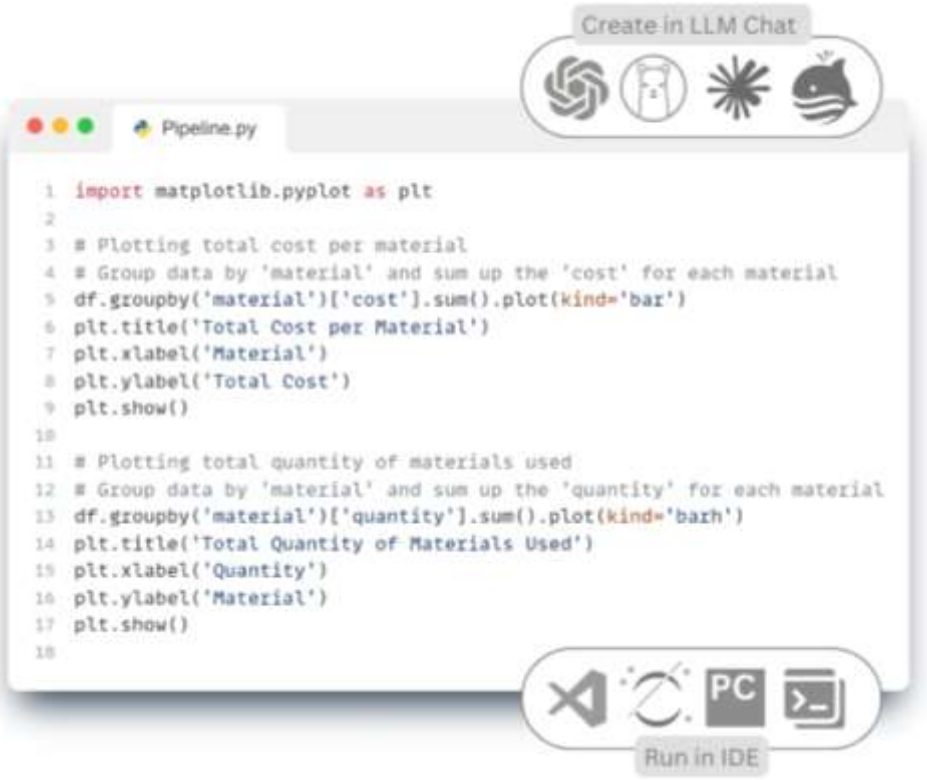


Рис. 7.2-20 Ответ LLM-модели в виде кода Python для визуализации данных из фрейма данных с помощью библиотеки matplotlib.

LLM автоматически генерирует и выполняет Python-код (Рис. 7.2-20) с использованием библиотеки matplotlib. После выполнения этого кода мы получаем графики затрат и объёмов использования материалов в строительных проектах напрямую в чате (Рис. 7.2-21), что значительно упрощает аналитическую работу.

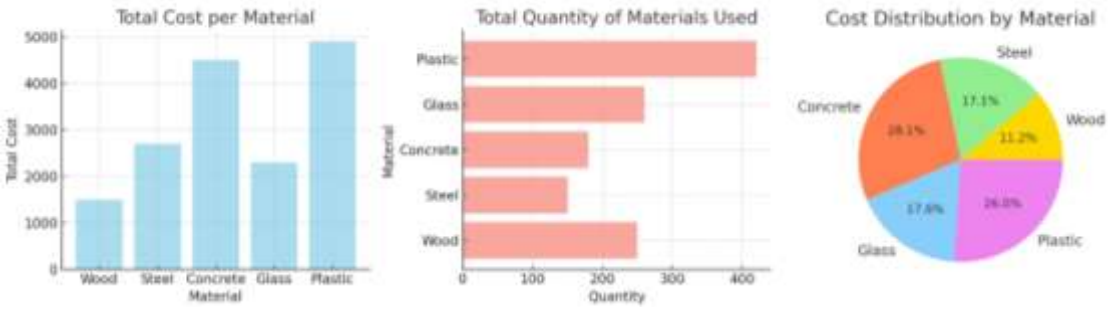


Рис. 7.2-21 Визуализация отклика LLM в виде графиков на основе данных, собранных в DataFrame.

Поддержка в разработке идей для написания ETL кода, анализа и выполнение кода, визуализация результатов доступна через простые текстовые запросы в LLM, без необходимости изучать основы программирования. Появление таких ИИ инструментов, как LLM, определенно меняет подход к программированию и автоматизации обработки данных (Рис. 7.2-22).

> Согласно отчету PwC «Какова реальная ценность искусственного интеллекта для твоего бизнеса и как ты можешь извлечь из него выгоду?» (2017) [139], автоматизация процессов и повышение производительности станут основными драйверами экономического роста. И ожидается, что повышение производительности труда обеспечит более 55% всего прироста ВВП за счет ИИ в период 2017-2030 гг.".

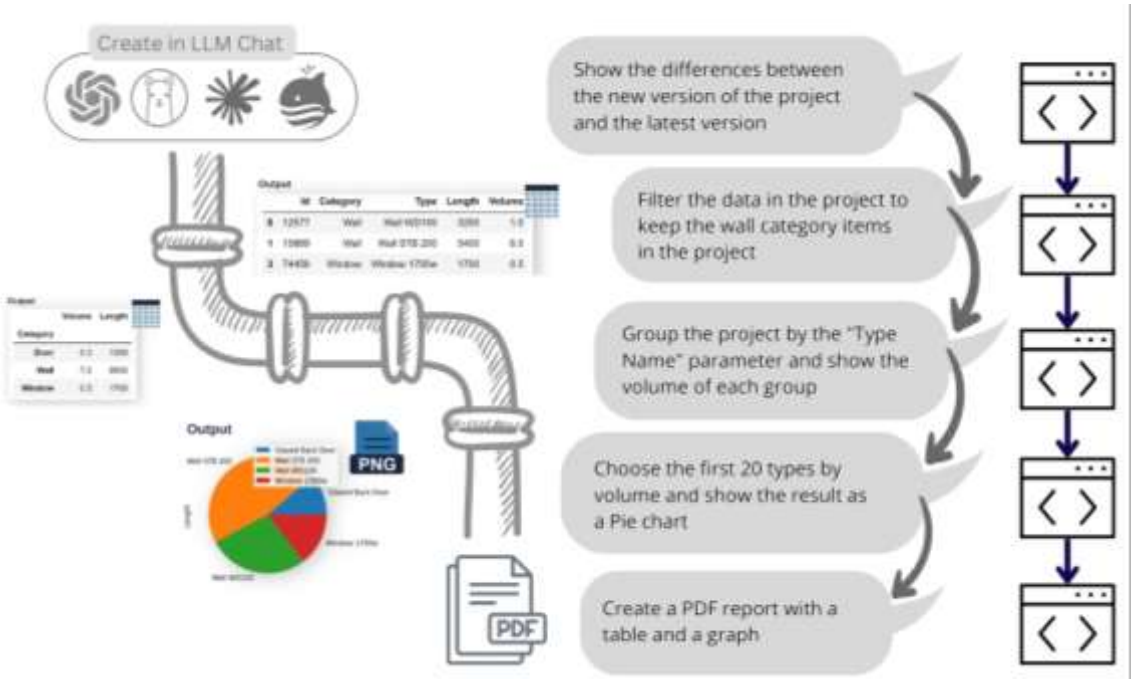


Рис. 7.2-22 ИИ LLM помогает формировать черновик кода, который применяется в будущих проектах без необходимости использования LLM.

> Используя инструменты, подобные ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, а также открытые данные и программное обеспечение с открытым исходным кодом, мы можем автоматизировать те процессы, которые ранее осуществлялись только с помощью специализированных, высокозатратных и сложных в обслуживании модульных проприетарных систем.

В контексте строительства это означает, что компании, которые первыми внедряют автоматизированные Pipeline-процессы обработки данных, получат существенные преимущества — от повышения эффективности управления проектами до снижения финансовых потерь и устранения фрагментированных приложений и изолированных хранилищ данных.

Описанная логика выполнения бизнес-задач в процессе ETL - важнейшая часть автоматизации процессов аналитики и обработки данных, являющаяся специфической разновидностью более широкого понятия – конвейеров (Pipelines).

ГЛАВА 7.3.
АВТОМАТИЧЕСКИЙ ETL КОНВЕЙЕР (PIPELINE)

Pipeline: Автоматический ETL конвейер данных

Процесс ETL традиционно используется для обработки данных в аналитических системах, охватывая как структурированные, так и неструктурированные источники. Однако в современном цифровом окружении всё чаще применяется более широкий термин — Pipeline (конвейер), который описывает любую последовательную цепочку обработки, где выход одного этапа становится входом для следующего.

Такой подход применим не только к данным, но и к другим видам автоматизации: обработке задач, построению отчётности, интеграции с программным обеспечением и цифровым документооборотом(Рис. 7.3-1).

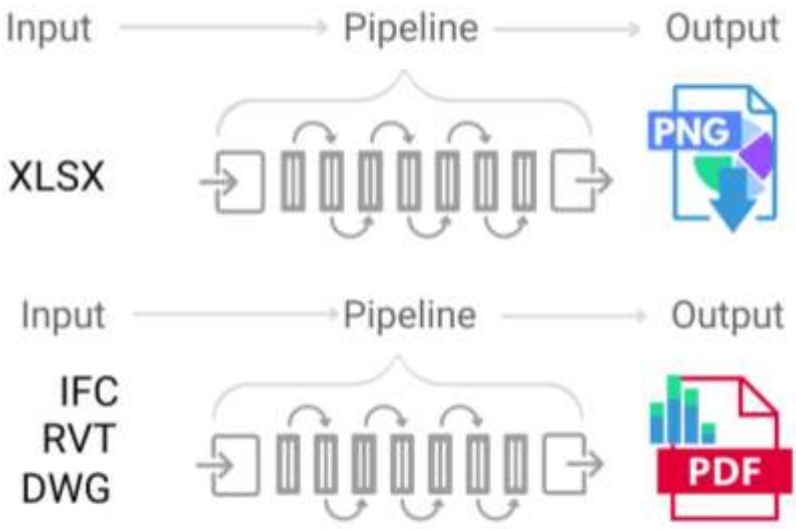


Рис. 7.3-1 Pipeline — это последовательность обработки, в которой выход одного этапа становится входом для следующего этапа.

Применение Pipeline является одним из главных элементов автоматизации, особенно в условиях работы с большим количеством разнотипных данных. Конвейерная архитектура позволяет организовать сложные этапы обработки в модульный, последовательный и управляемый формат, что повышает читаемость, упрощает поддержку кода, и делает возможной поэтапную отладку и масштабируемое тестирование.

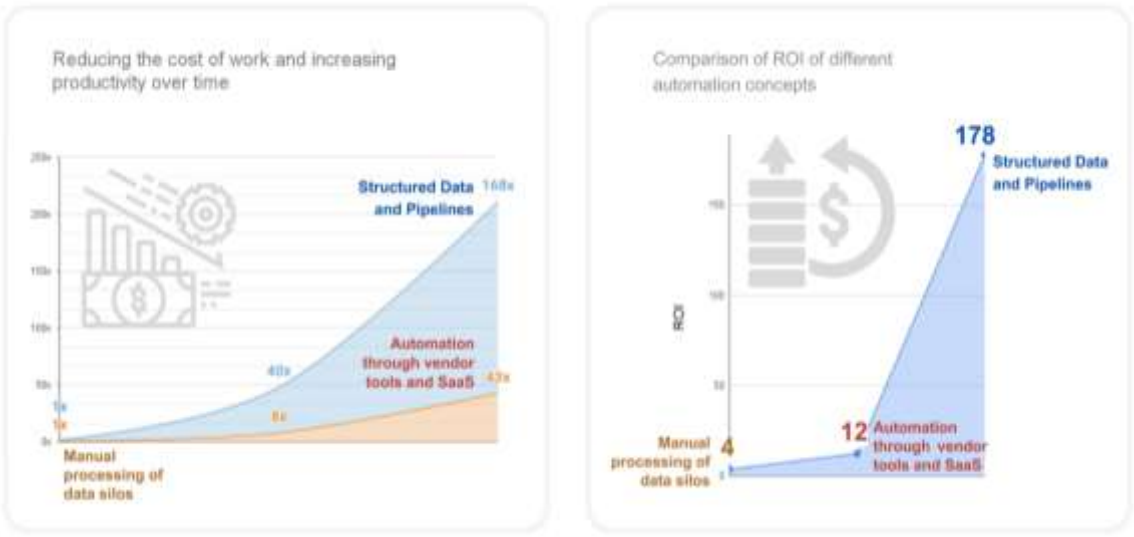


Рис. 7.3-2 ROI Pipeline процесса проверки данных сокращает в десятки и сотни раз время выполнения по сравнению с обработкой при помощи классических инструментов [74].

В отличие от ручной работы в проприетарных системах (ERP, PMIS, CAD и др.), конвейерная обработка данных позволяет значительно (Рис. 7.3-2) повысить скорость выполнения задач, избежать повторной работы и автоматизировать запуск процессов в нужное время (Рис. 7.3-3).

Рис. 7.3-3 Пример ETL Pipeline по автоматическому получении графика из табличных данных в файле XLSX без открытия Excel.

Для обработки потоковых данных и построения автоматизированного Pipeline, аналогично процессу ETL, необходимо заранее определить источники данных, а также временные рамки их сбора — как для конкретного бизнес-процесса, или в рамках всей компании.

В строительных проектах данные поступают из множества разнородных источников с различной периодичностью обновления. Для формирования надёжной витрины данных критически важно фиксировать момент извлечения и актуализации информации. Это позволяет обеспечить своевременное принятие решений и повысить оперативность управления проектом.

Одним из вариантов является запуск сборочного процесса в фиксированное время — например, в 19:00, по завершении рабочего дня. В этот момент активируется первый скрипт, отвечающий за агрегирование данных из различных систем и хранилищ (Рис. 7.3-4 шаг 1). Далее следует автоматизированная обработка и трансформация данных в структурированный формат, пригодный для аналитики (Рис. 7.3-4 шаг 2-4). На финальном этапе, используя подготовленные данные, автоматически формируются отчёты, дашборды и другие продукты, описанные в предыдущих главах (Рис. 7.3-4 шаг 6-7). В результате к 05:00 утра у менеджеров уже есть актуальные отчеты о состоянии проекта в нужном формате (Рис. 7.3-5).

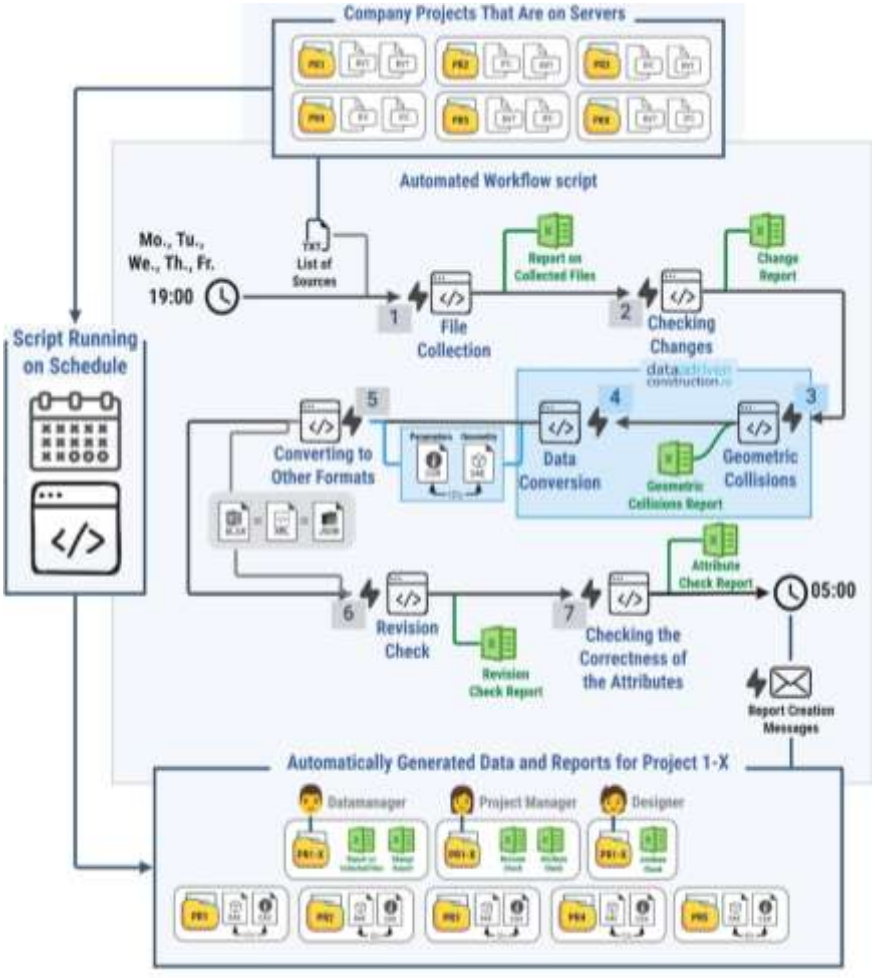


Рис. 7.3-4 Данные в Pipeline, автоматически собранные вечером, обрабатываются ночью, чтобы к утру у менеджеров были актуальные отчеты и свежие отчёты.

> Своевременный сбор данных, определение KPI, автоматизация процессов трансформации и визуализация через информационные панели — ключевые элементы успешного принятия решений на основе данных.

Подобные автоматизированные процессы (Рис. 7.3-4) могут выполняться с полной автономностью: они запускаются по расписанию, обрабатывают данные без участия оператора и могут быть развёрнуты как в облаке, так и на собственном сервере компании (Рис. 7.3-5). Это позволяет интегрировать такие ETL-конвейеры в существующую ИТ-инфраструктуру, сохраняя контроль над данными и обеспечивая гибкость при масштабировании.

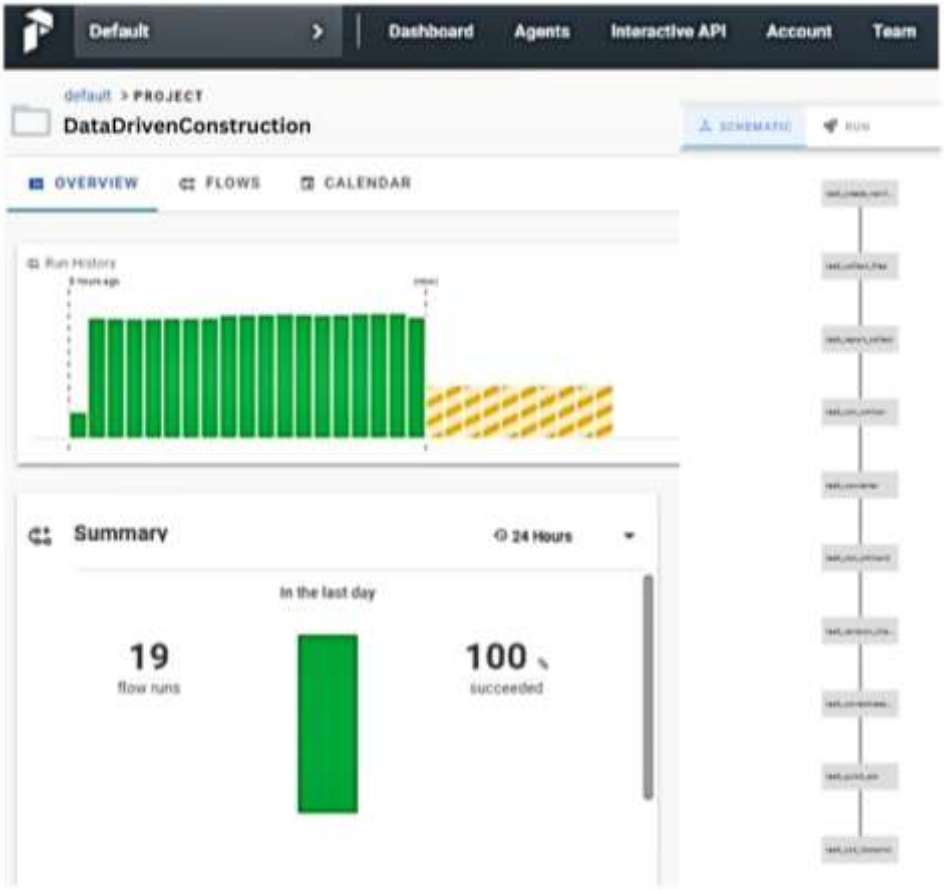


Рис. 7.3-5 Автоматический ETL-конвейер процессов (Рис. 7.3-4) на платформе Prefect, в котором 10 скриптов python запускаются поочередно после 19:00 каждый рабочий день.

Автоматизация рабочих процессов не только повышает производительность команды, высвобождая время для более значимых и менее рутинных задач, но и служит первым важным шагом на пути внедрения технологий искусственного интеллекта (ИИ) в бизнес-процессы, о чем мы подробнее будем говорить в главе "Прогнозы и машинное обучение".

Процесс проверки данных Pipeline-ETL с помощью LLM

В предыдущих главах, посвящённых созданию требований к данным и автоматизации ETL, мы поэтапно разобрали процесс подготовки, трансформации, валидации и визуализации данных. Эти действия были реализованы в виде отдельных блоков кода (Рис. 7.2-18 - Рис. 7.2-20) каждый из которых выполнял конкретную задачу.

Теперь перед нами стоит следующая цель — объединить эти элементы в единый, связный и автоматизированный конвейер обработки данных — конвейер, ETL-Pipeline — в котором все этапы (загрузка, проверка, визуализация, экспорт) выполняются последовательно в одном авто исполняемом скрипте.

В следующем примере будет реализован полный цикл обработки данных: от загрузки исходного файла CSV → до проверки структуры и значений с использованием регулярных выражений → подсчёта результатов → генерации визуального отчёта в формате PDF.

- Для получения соответствующего кода можно использовать следующий текстовый запрос к LLM:

> Пожалуйста напиши пример кода, который выполняет загрузку данных из CSV, проверку данных DataFrame с помощью регулярных выражений, проверяет идентификаторы в формате 'W-NEW' или 'W-OLD', энергоэффективность с буквами от 'A' до 'G', гарантийный срок и цикл замены с числовыми значениями в годах и в конце создает отчет с подсчетом пройденных и проваливших проверку значений, генерирует PDF с гистограммой результатов и добавляет текстовое описание ⏎

Ответ 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

Рис. 7.3-6 Pipeline (ETL) автоматизирует полный цикл обработки данных: от загрузки и проверки до создания структурированного отчета в формате PDF.

Автоматический код (Рис. 7.3-6) внутри чата LLM или в DIE, после копирования кода, выполнит валидацию данных из CSV-файла с использованием заданных регулярных выражений, создаст отчет о количестве прошедших и не прошедших проверку записей, а затем сохранит результаты проверки в виде PDF-файла.

Подобная структура ETL-конвейера, где каждый этап — от загрузки данных до формирования отчёта — реализован как отдельный модуль, обеспечивает прозрачность, масштабируемость и воспроизводимость. Представление логики проверки в виде легко читаемого кода на Python делает процесс прозрачным и понятным не только разработчикам, но и специалистам в области управления данных, качества, и аналитики.

> Использование Pipeline-подхода для автоматизации обработки данных позволяет стандартизировать процессы, повысить их повторяемость и упростить адаптацию к новым проектам. Благодаря этому формируется единая методология анализа данных, вне зависимости от источника или типа задачи — будь то проверка на соответствие стандартам, генерация отчётности или передача данных во внешние системы.

Подобная автоматизация снижает влияние человеческого фактора, уменьшает зависимость от проприетарных решений и повышает точность и надёжность результатов, делая их пригодными как для оперативной аналитике на уровне проектов, так и для стратегической аналитике на уровне компаний.

Pipeline-ETL: проверка данных и информации элементов проекта в CAD (BIM)

Данные из систем и баз данных CAD (BIM) — одни из самых сложных и динамично обновляемых источников данных в бизнесе строительных компаний. Эти приложения не только описывают проект с помощью геометрии, но и дополняют его многослойной текстовой информацией: объемами, свойствами материалов, назначениями помещений, уровнями энергоэффективности, допусками, сроками эксплуатации и другими атрибутами.

> Атрибуты, присваиваемые сущностям в CAD-моделях, формируются на этапе проектирования и становятся основой для дальнейших бизнес-процессов, включая калькуляции затрат, составление графиков, оценку жизненного цикла и интеграцию с ERP и CAFM-системами, где эффективность процессов во многом зависит от качества данных, поступающих из проектных отделов.

Традиционный подход к проверке атрибутов в CAD-(BIM-) моделях предполагает ручную валидацию (Рис. 7.2-1), которая при большом объёме моделей превращается в долгий и дорогостоящий процесс. Учитывая объём и количество современных строительных проектов и их регулярные обновления, процесс проверки и трансформации данных становится неустойчивым и неподъёмным.

Генеральные подрядчики и руководители проектов сталкиваются с необходимостью обрабатывать большое количество проектных данных, включая множественные версии и фрагменты одних и тех же моделей. Данные поступают от проектных организаций в форматах RVT, DWG, DGN, IFC, NWD и других (Рис. 3.1-14) и требуют регулярной проверки на соответствие отраслевым и корпоративным стандартам.

Из-за зависимости от ручных действий и специализированных программ процесс валидации данных становится узким местом в рабочих процессах, связанных с данными из моделей для всей компании. Автоматизация и применение структурированных требований позволяют устранить эту зависимость, кратно увеличив скорость и надёжность проверки данных (Рис. 7.3-7).

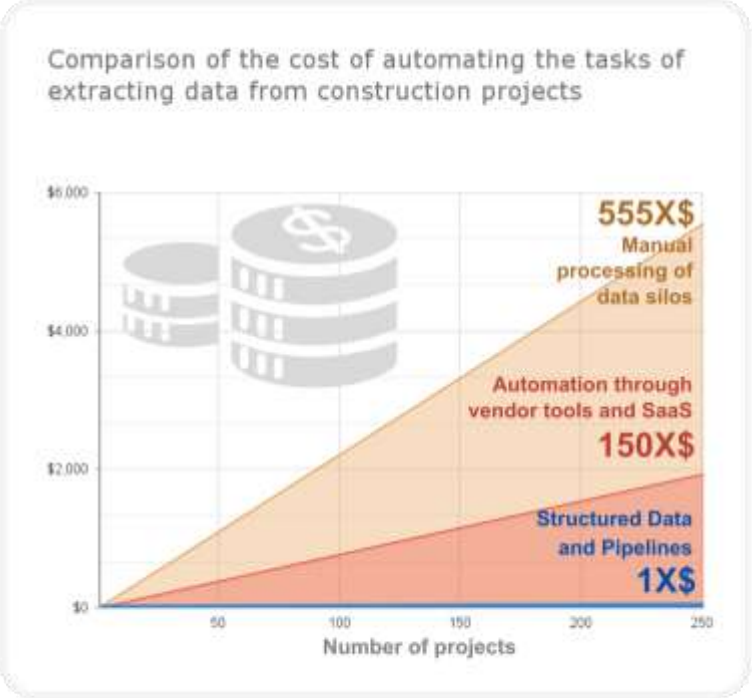


Рис. 7.3-7 Автоматизация увеличивает скорость проверки и обработки данных, что снижает стоимость работ в десятки раз [140].

Процесс проверки CAD данных включает выгрузку данных (ETL этап Extract) из различных закрытых (RVT, DWG, DGN, NWS и др.) или открытых полуструктурированных и параметрических форматов (IFC, CPXML, USD), в которых к каждому атрибуту и его значениям могут быть применены таблицы правил (этап Transform) с использованием регулярных выражений RegEx (Рис. 7.3-8), процесс который мы подробно рассматривали в четвёртой части книги.

Создание отчета об ошибках в формате PDF и успешно проверенных записях должно завершаться выводом (этап Load) в структурированные форматы, учитывающим только проверенные сущности, которые могут быть использованы для дальнейших процессов.

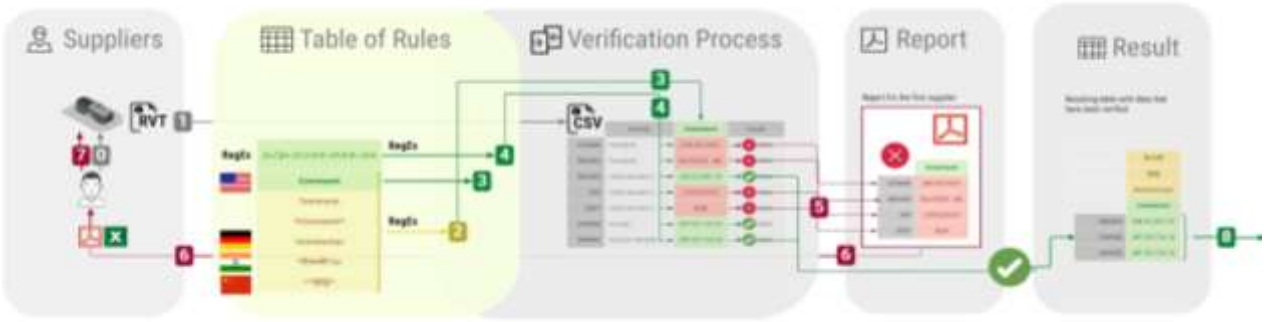


Рис. 7.3-8 Процесс проверки данных от поставщиков проектных данных до конечного отчёта, проверенного с помощью регулярных выражений.

Автоматизация проверки данных из систем CAD (BIM) при наличии структурированных требования и при потоковом поступлении новых данных, которые обрабатываются через ETL-Pipelines (Рис. 7.3-9) снижает необходимость ручного участия в процессе валидации (каждый из процессов проверки и составления требований к данным рассматривался в предыдущих главах).

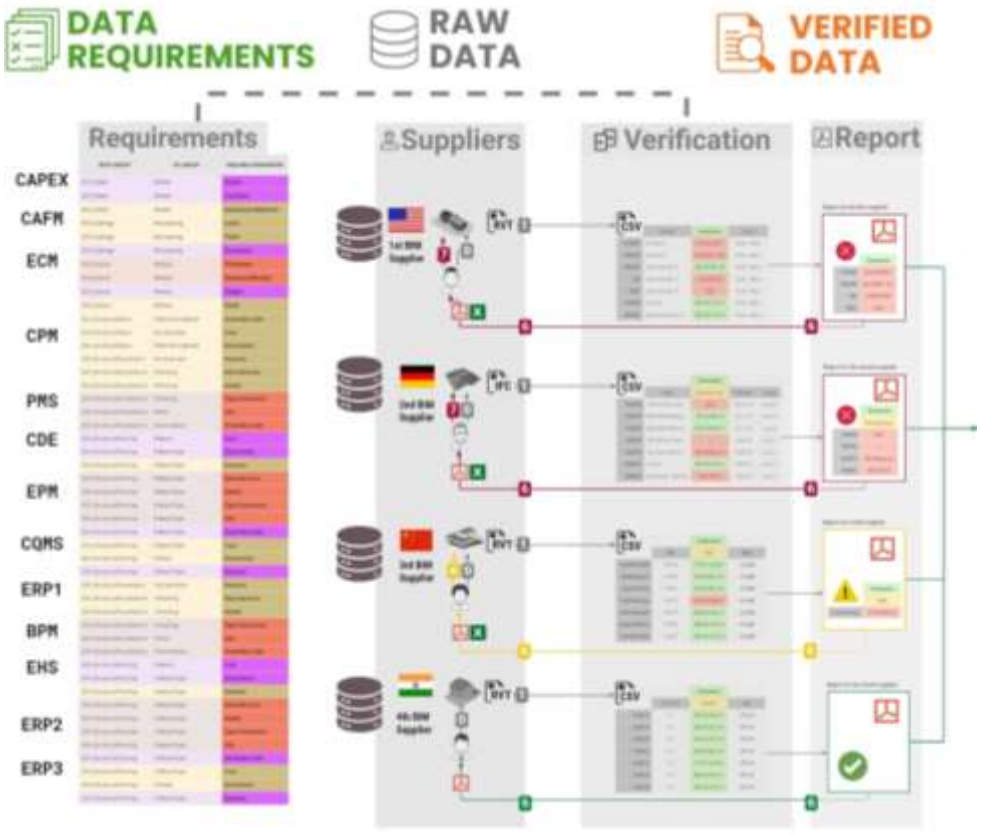


Рис. 7.3-9 Автоматизация проверки данных через ETL упрощает управление строительными проектами благодаря ускорению процессов.

Традиционно проверка моделей, предоставляемых подрядчиками и CAD (BIM)-специалистами, может занимать от нескольких дней до недель. Однако с внедрением автоматизированных ETL-процессов этот срок можно сократить до нескольких минут. В типовой ситуации подрядчик заявляет: *«Модель проверена и соответствует требованиям»*. Подобное утверждение запускает цепочку по проверки заявления подрядчика о качестве данных:

- Менеджер проекта — «*Подрядчик утверждает: «Модель проверена, всё в порядке*»».
- Менеджер по данным — «Загружаем валидацию»:
  - Простой скрипт в Pandas за секунды выявляет нарушение. Автоматизация исключает споры:
    - Категория: OST_StructuralColumns, Параметр: FireRating IS NULL.
    - Генерируем список ID нарушений → экспортируем в Excel/PDF.

  Простой скрипт в Pandas за секунды выявляет нарушение:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"]  # Фильтрация
issues = df[df["FireRating"].isnull()]  # Пустые значения
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx")  # Экспорт ID
```

- Менеджер по данным менеджеру проекта — *«Проверка показала, что у 18 колонн не заполнен параметр FireRating».*
- Менеджер проекта подрядчику — «*Модель возвращается на доработку: параметр FireRating обязательный, без него приёмка невозможна»*

В результате CAD модель не проходит проверку качества, автоматизация исключает споры, а подрядчик почти мгновенно получает структурированный отчёт со списком ID проблемных элементов. Таким образом процесс валидации становится прозрачным, воспроизводимым и защищённым от человеческого фактора (Рис. 7.3-10).

> Подобный подход превращает процесс проверки данных в инженерную функцию, а не в ручной контроль качества. Это не только повышает производительность, но и даёт возможность применить одинаковую логику ко всем проектам компании, обеспечивая сквозную цифровую трансформацию процессов, от проектирования до эксплуатации.

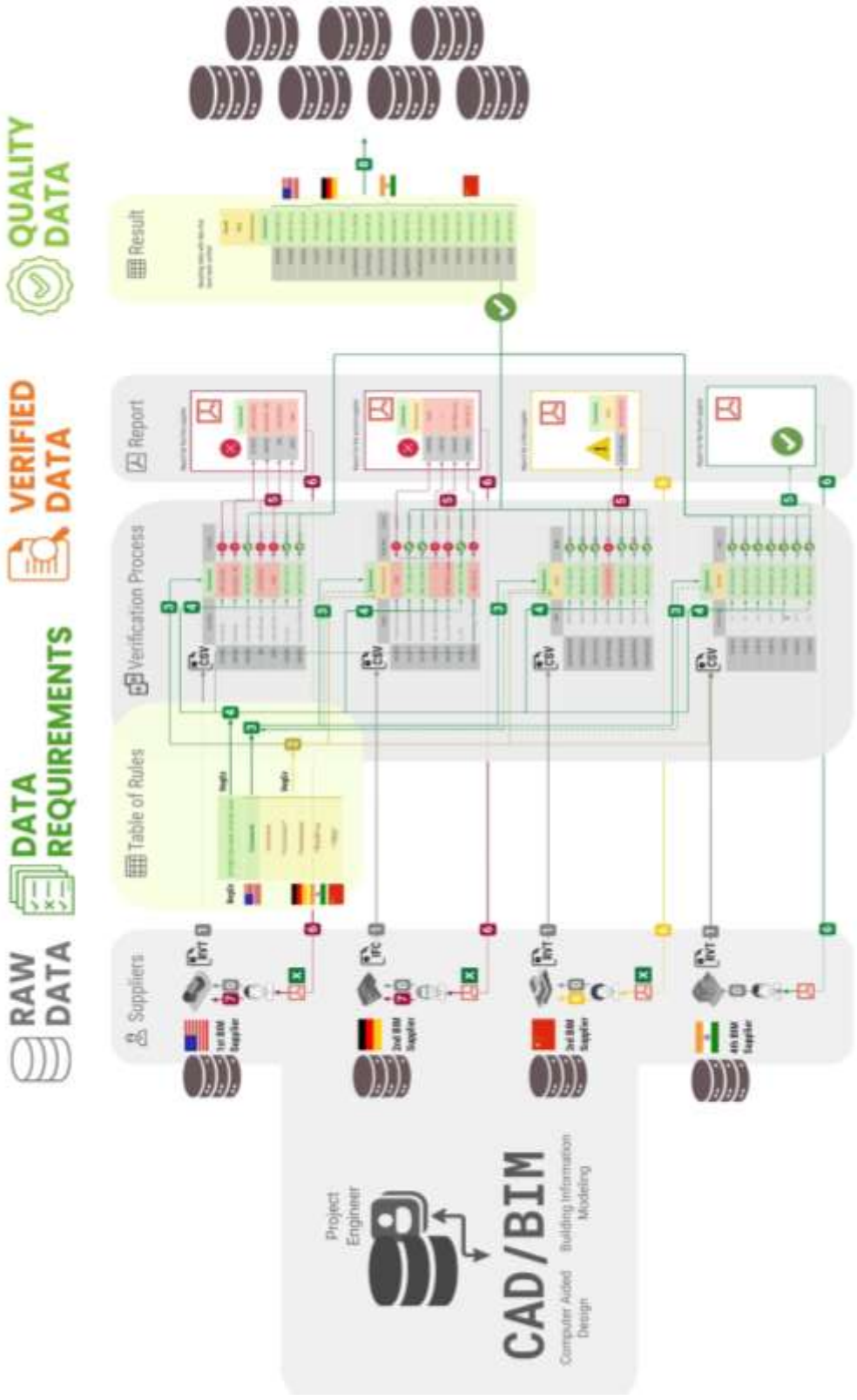


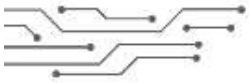
Рис. 7.3-10 Автоматизация проверки атрибутов элементов исключает человеческий фактор и снижает вероятность ошибок.

Благодаря применению автоматизированных конвейеров (Рис. 7.3-10) пользователи системы, ожидающие качественных данных от CAD- (BIM-) систем, могут мгновенно получать необходимые им выходные данные - таблицы, документы, изображения - и быстро интегрировать их в свои рабочие задачи.

> Автоматизация контроля, обработки и анализа приводит к изменению в подходах управления строительными проектами, особенно в части взаимодействия различных систем, без использования сложных и дорогостоящих модульных проприетарных систем или закрытых решений от вендоров.

В то время, как концепции и маркетинговые аббревиатуры приходят и уходят, сами процессы проверки требований к данным навсегда останутся неотъемлемой частью бизнес-процессов. Вместо того, чтобы создавать все новые и новые специализированные форматы и стандарты, строительной отрасли стоит обратить внимание на инструменты, которые уже доказали свою эффективность в других отраслях экономики. Сегодня существуют мощные платформы для автоматизации обработки данных и интеграции процессов, которые позволяют компаниям значительно сократить время на рутинные операции и минимизировать ошибки в процессах Extract, Transform и Load.

Одним из популярных примеров решений по автоматизации и оркестрации ETL процессов является Apache Airflow, который позволяет организовывать сложные вычислительные процессы и управлять ETL-конвейерами. Наряду с Airflow, активно используются и другие подобные решения, такие, как Apache NiFi для маршрутизации и потоковой обработки данных, а также n8n для автоматизации бизнес-процессов.



ГЛАВА 7.4.
ОРКЕСТРАЦИЯ ETL И РАБОЧИХ ПРОЦЕССОВ: ПРАКТИЧЕСКИЕ РЕШЕНИЯ

DAG и Apache Airflow: автоматизация и оркестрация рабочих процессов

Apache Airflow — это бесплатная платформа с открытым исходным кодом, предназначенная для автоматизации, оркестрации и мониторинга рабочих процессов (ETL-конвейеров).

Каждый день в работе с большими объемами данных требуется:

- Скачивать файлы из разных источников - Extract (например, от поставщиков или клиентов).
- Преобразовывать эти данные в нужный формат - Transform (структурировать, очищать и проверять)
- Отправлять результаты на проверку и создавать отчеты - Load (выгружать в нужные системы, документы, базы данных или дашборды.

Ручное выполнение подобных ETL процессов занимает значительное время и приводит к риску ошибок, связанных с человеческим фактором. Изменение источника данных или сбой на одном из этапов может вызвать задержки и некорректные результаты.

Инструменты автоматизации, такие как Apache Airflow, позволяют выстроить надежный ETL-конвейер, минимизировать ошибки, сократить время обработки данных и обеспечить их корректность на каждом этапе. В основе Apache Airflow лежит концепция DAG (Directed Acyclic Graph) — направленного ациклического графа, в котором каждая задача (оператор) связана с другими зависимостями и исполняется строго в заданной последовательности. DAG исключает циклы, что обеспечивает логичную и предсказуемую структуру выполнения задач.

Airflow берет на себя оркестрацию — управление зависимостями между задачами, контроль за расписанием выполнения, отслеживание состояния и автоматическую реакцию на сбои. Такой подход минимизирует ручное вмешательство и обеспечивает надёжность всего процесса.

> **Оркестратор задач** — инструмент или система, предназначенный для управления и контроля выполнения задач в сложных вычислительных и информационных средах. Он облегчает процесс развертывания, автоматизации и управления выполнением задач, что позволяет повысить эффективность работы и оптимизировать ресурсы.

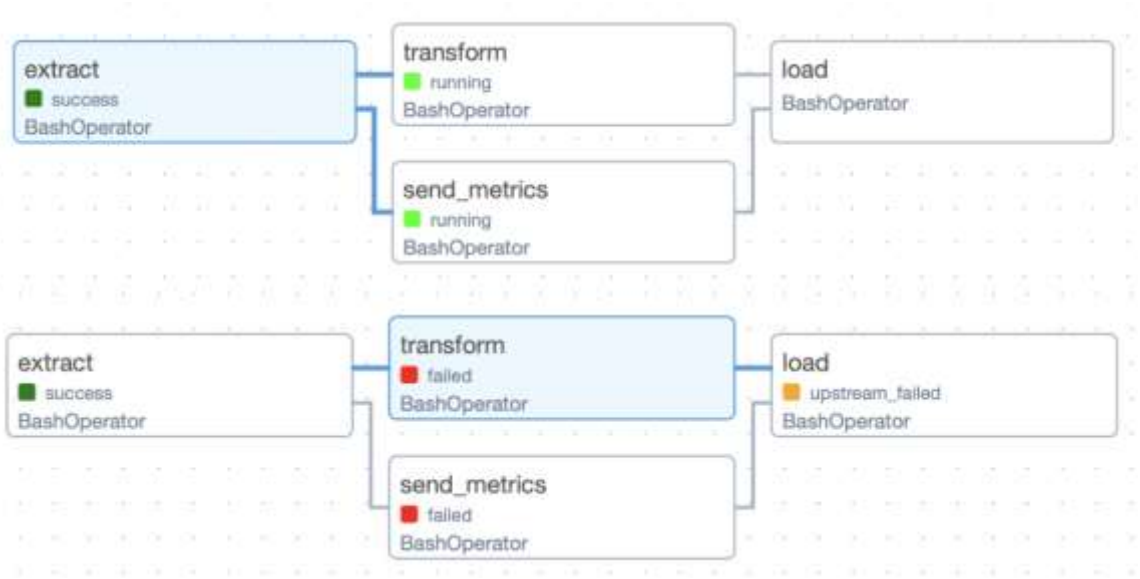


Рис. 7.4-1 Apache Airflow предоставляет удобный интерфейс, где можно визуализировать DAG-ETL, просматривать логи выполнения, статус запуска задачи и многое другое.

Airflow широко используется для оркестрации и автоматизации распределенных вычислений, обработки данных, управления ETL (Extract, Transform, Load) процессами, планирования задач и других сценариев работы с данными. По умолчанию Apache Airflow использует SQLite как базу данных.

Пример простого DAG, подобно ETL, состоит из задач — Extract, Transform и Load. В графе, который управляется через пользовательский интерфейс (Рис. 7.4-1), определён порядок выполнения задач (фрагментов кода): например, сначала исполняется extract, затем transform (и sending_metrics), а завершает работу задача load. Когда все задачи выполнены, процесс загрузки данных считается успешным.

Apache Airflow: практическое применение по автоматизации ETL

Apache Airflow широко применяется для организации сложных процессов обработки данных, позволяя строить гибкие ETL-конвейеры. Apache Airflow можно запускать как через веб-интерфейс, так и программно через Python-код (Рис. 7.4-2). В веб-интерфейсе (Рис. 7.4-3) администраторы и разработчики могут визуально отслеживать DAG-и, запускать задачи и анализировать результаты выполнения.

Используя DAG, можно задать четкую последовательность выполнения задач, управлять зависимостями между ними и автоматически реагировать на изменения в исходных данных. Рассмотрим пример использования Airflow для автоматизации обработки отчетности (Рис. 7.4-2).

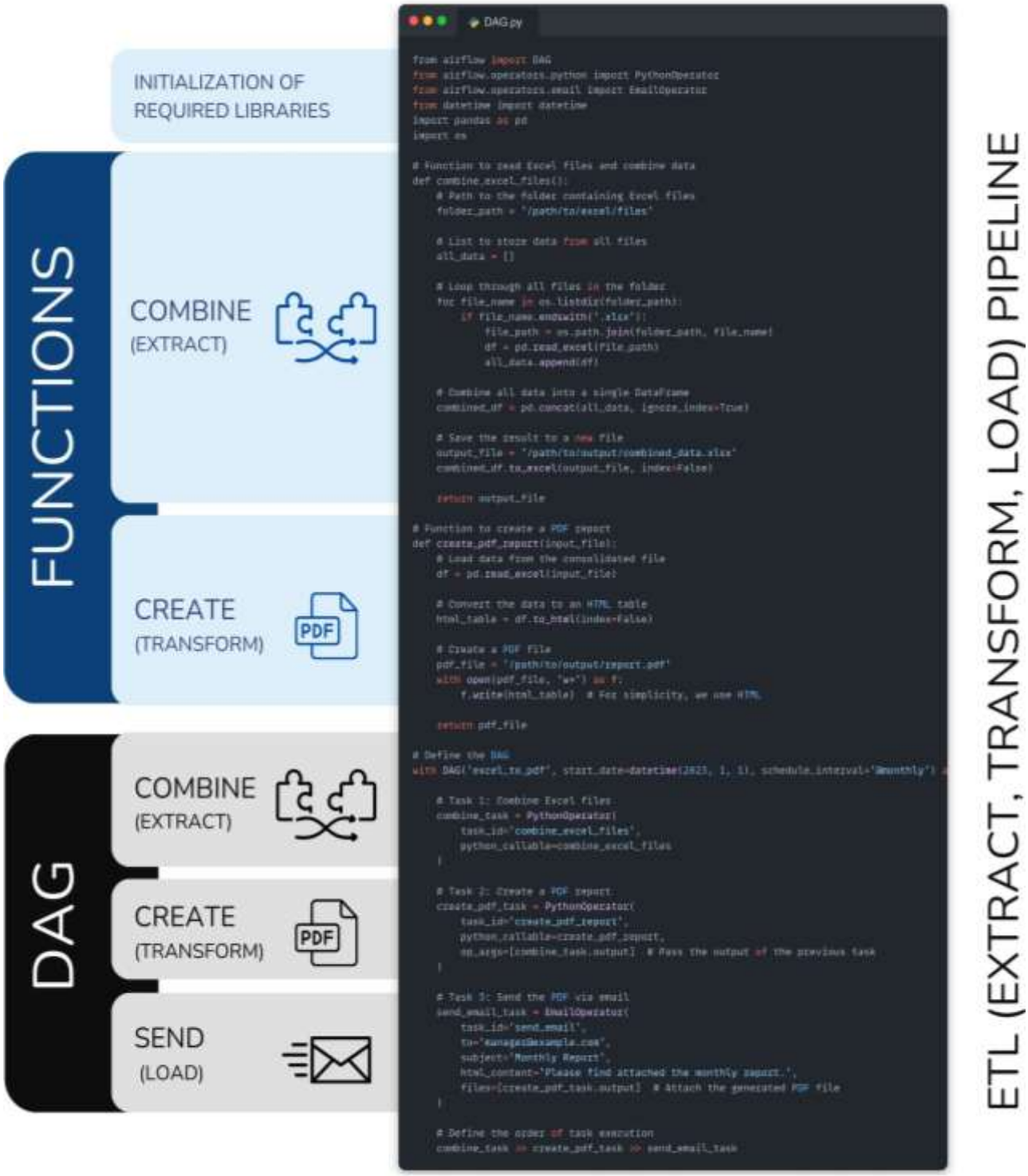


Рис. 7.4-2 Концепция ETL-конвейера для обработки данных с использованием Apache Airflow.

В данном примере (Рис. 7.4-2) рассмотрен DAG, который выполняет ключевые задачи в рамках ETL-конвейера:

- **Чтение Excel-файлов (Extract):**
  • Последовательный обход всех файлов в заданной директории.
  • Считывание данных из каждого файла с использованием библиотеки pandas.
  • Объединение всех данных в единый DataFrame.
- **Создание PDF-документа (Transform):**
  • Преобразование объединённого DataFrame в HTML-таблицу.
  • Сохранение таблицы в формате PDF (в демонстрационной версии — через HTML).
- **Отправка отчёта по электронной почте (Load):**
  • Применение EmailOperator для отправки PDF-документа по электронной почте.
- **Настройка DAG:**
  • Определение последовательности выполнения задач: извлечение данных → формирование отчёта → отправка.
  • Назначение расписания запуска (@monthly — первое число каждого месяца).

В автоматическом ETL-примере (Рис. 7.4-2) показано, как собирать данные из Excel-файлов, создавать PDF-документ и отправлять его по email. Это лишь один из множества возможных сценариев использования Airflow. Данный пример можно адаптировать под любые конкретные задачи, чтобы упростить и автоматизировать процессы обработки данных.

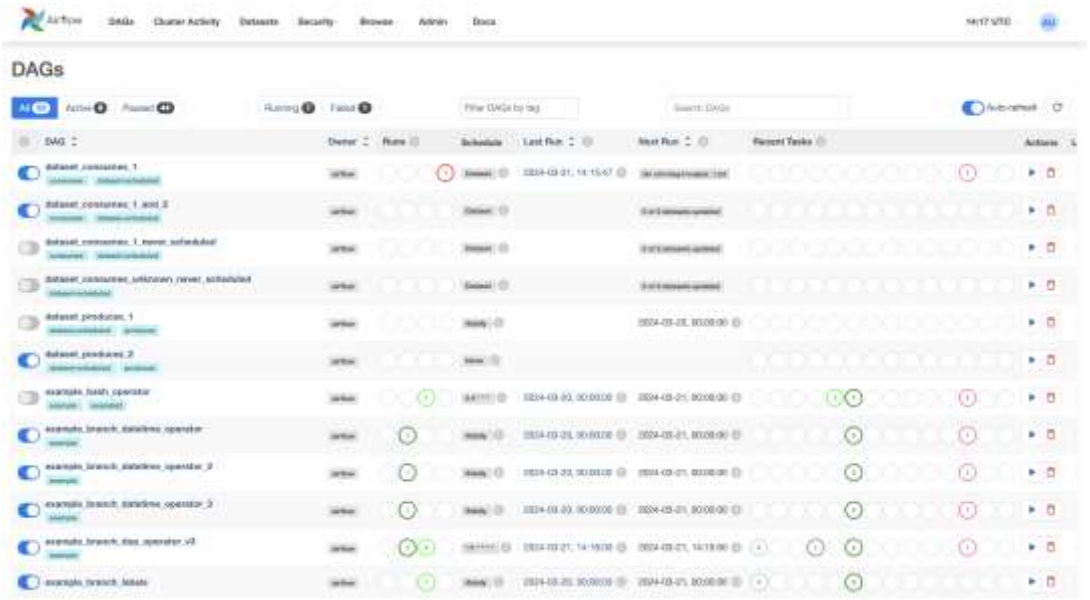


Рис. 7.4-3 Обзор всех групп DAG в окружении с информацией о последних запусках.

Веб-интерфейс Apache Airflow (Рис. 7.4-3) предоставляет комплексную визуальную среду для управления рабочими процессами данных. Он отображает DAG-и в виде интерактивных графов, где узлы представляют задачи, а ребра — зависимости между ними, что позволяет легко отслеживать сложные процессы обработки данных. Интерфейс включает панель мониторинга с информа-

цией о состоянии выполнения задач, историю запусков, детализированные логи и метрики производительности. Администраторы могут вручную запускать задачи, перезапускать неудачные операции, приостанавливать DAG-и и настраивать переменные окружения — всё это через интуитивно понятный пользовательский интерфейс.

Подобная архитектура может быть дополнена валидацией данных, уведомлениями о статусе выполнения, интеграцией с внешними API или базами данных. Airflow позволяет гибко адаптировать DAG: добавлять новые задачи, изменять их порядок, комбинировать цепочки — что делает его эффективным инструментом для автоматизации сложных процессов обработки данных. При запуске DAG в веб-интерфейсе Airflow (Рис. 7.4-3, Рис. 7.4-4) можно отслеживать статус выполнения задач. Система использует цветовую индикацию:

- Зеленый — задача успешно выполнена.
- Желтый — процесс выполняется.
- Красный — ошибка при выполнении задачи.

При сбоях (например, отсутствует файл или нарушена структура данных) система автоматически инициирует отправку уведомления.

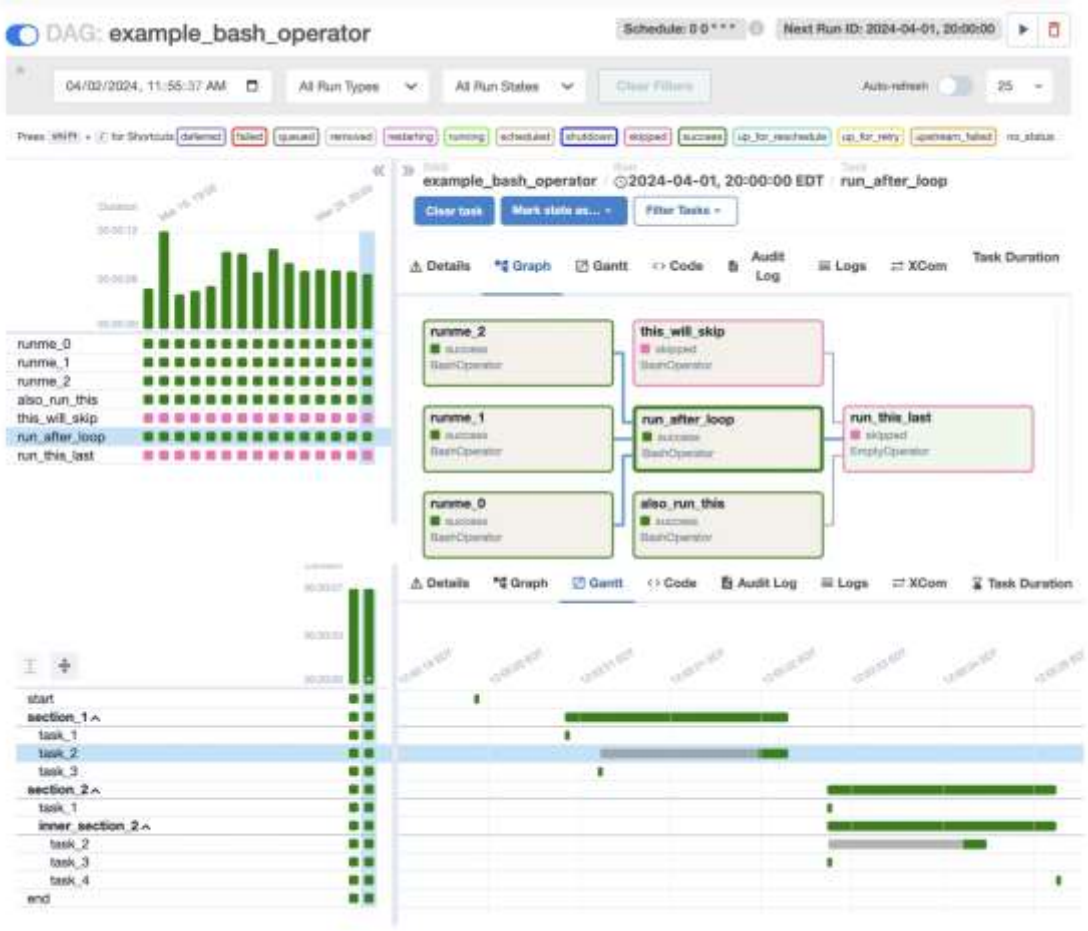


Рис. 7.4-4 Apache Airflow значительно упрощает диагностику проблем, оптимизацию процессов и совместную работу команд над сложными пайплайнами обработки данных.

> Apache Airflow удобен тем, что автоматизирует рутинные задачи, избавляя от необходимости выполнять их вручную. Он обеспечивает надежность за счет мониторинга выполнения процессов и мгновенного оповещения об ошибках. Гибкость системы позволяет легко добавлять новые задачи или изменять существующие, адаптируя рабочие процессы под изменяющиеся требования.

В дополнение к Apache Airflow, существуют похожие инструменты для оркестрации рабочих процессов. Например открытый и бесплатный Prefect (Рис. 7.3-5) предлагает более простой синтаксис и лучше интегрируется с Python, Luigi, разработанный Spotify, предоставляет похожую функциональность и хорошо работает с большими данными. Также стоит отметить Kronos и Dagster, которые предлагают современные подходы к созданию Pipeline с фокусом на модульность и масштабируемость. Выбор инструмента оркестрации задач зависит от конкретных потребностей проекта, но все они помогают автоматизировать сложные ETL процессы обработки данных.

Отдельного упоминания заслуживает Apache NiFi — платформа с открытым исходным кодом, предназначенная для потоковой (streaming) обработки и маршрутизации данных. В отличие от Airflow, который ориентирован на пакетную обработку и управление зависимостями, NiFi фокусируется на реальном времени, преобразовании данных на лету и гибкой маршрутизации между системами.

Apache NiFi для маршрутизации и преобразования данных

Apache NiFi — это мощная платформа с открытым исходным кодом, предназначенная для автоматизации потоков данных между различными системами. Изначально разработан в 2006 году Агентством национальной безопасности США (NSA) под названием "Niagara Files" для внутренних нужд. В 2014 году проект был открыт и передан в Apache Software Foundation, став частью их инициатив по передаче технологий [141].

Apache NiFi спроектирован для сбора, обработки и передачи данных в реальном времени. В отличие от Airflow, который работает с пакетными задачами и требует четко заданных расписаний, NiFi функционирует в режиме поточной обработки, позволяя беспрерывно передавать данные между различными сервисами.

> Apache NiFi идеально подходит для интеграции с IoT-устройствами, сенсорами строительных объектов, системами мониторинга, и например потоковой проверке CAD форматов на сервере, где может требоваться немедленная реакция на изменения в данных.

Благодаря встроенным инструментам фильтрации, трансформации и маршрутизации, NiFi позволяет стандартизировать данные (этап Transform) перед их передачей (Load) в хранилища или аналитические системы. Одним из его главных преимуществ является встроенная поддержка безопасности и контроль доступа, что делает его надежным решением для обработки конфиденциальной информации.

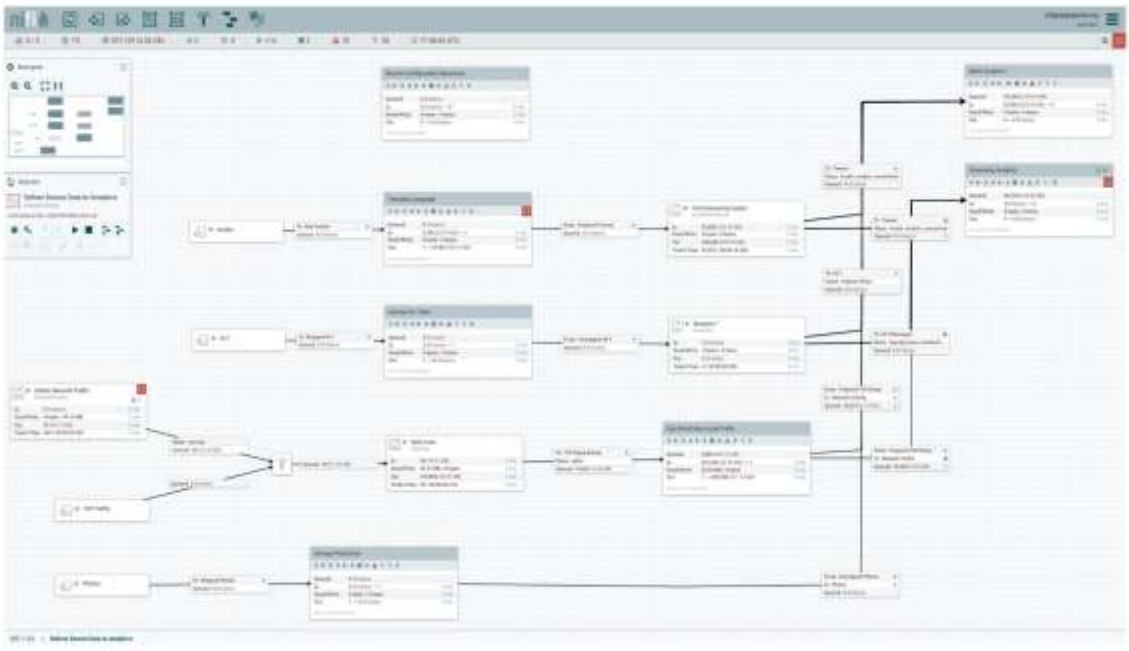


Рис. 7.4-5 Графическое представление потока данных в интерфейсе Apache NiFi.

Apache NiFi эффективно закрывает задачи потоковой передачи, фильтрации и маршрутизации данных в режиме реального времени. Он идеально подходит для технически насыщенных сценариев, где важна стабильная передача информации между системами и высокая пропускная способность.

Однако в случаях, когда основная цель — интеграция разнообразных сервисов, автоматизация рутинных операций и быстрая настройка рабочих процессов без глубоких знаний в программировании, востребованы решения с низким порогом входа и максимальной гибкостью. Одним из таких инструментов является n8n — платформа класса Low-Code/No-Code, ориентированная на бизнес-автоматизацию и визуальную оркестрацию процессов.

n8n Low-Code, No-Code оркестрации процессов

n8n — это Open Source Low-Code / No-Code платформа для построения автоматизированных рабочих процессов, отличающаяся простотой использования, гибкостью и возможностью быстрой интеграции с широким спектром внешних сервисов.

> **No-Code** — это метод создания цифровых продуктов без написания кода. Все элементы процесса — от логики до интерфейса — реализуются исключительно с помощью визуальных инструментов. No-Code платформы ориентированы на пользователей без технической подготовки и позволяют быстро создавать автоматизации, формы, интеграции и веб-приложения. Пример: пользователь настраивает автоматическую отправку уведомлений или интеграцию с Google Sheets через drag-and-drop интерфейс без знания программирования.

Благодаря открытому исходному коду и возможности локального развертывания, n8n в процессах автоматизации и создания ETL Pipelines дает компаниям полный контроль над своими данными, обеспечивая безопасность и независимость от облачных провайдеров.

В отличие от Apache Airflow, ориентированного на вычислительные задачи с жёсткой оркестрацией и требующего знания Python, n8n предоставляет визуальный редактор, позволяющий создавать сценарии без необходимости знания языков программирования (Рис. 7.4-6). Хотя его интерфейс позволяет создавать автоматизированные процессы без написания кода (No-Code), в более сложных сценариях пользователи могут добавлять собственные JavaScript и Python -функции для расширения возможностей (Low-Code).

> **Low-Code** — это подход к разработке программного обеспечения, при котором основная логика приложения или процесса создаётся с использованием графического интерфейса и визуальных элементов, а программный код применяется только для настройки или расширения функциональности. Low-Code платформы позволяют существенно ускорить разработку решений, вовлекая в процесс не только программистов, но и бизнес-пользователей с базовыми техническими навыками. Пример: пользователь может собрать бизнес-процесс из готовых блоков, а при необходимости добавить собственный скрипт на JavaScript или Python.

Хотя n8n позиционируется как платформа с низким порогом входа, для создания сложных сценариев автоматизации полезны базовые знания программирования, понимание веб-технологий и навыки работы с API. Гибкость системы позволяет адаптировать её под широкий спектр задач — от автоматизированной обработки данных до интеграции с мессенджерами, IoT-устройствами и облачными сервисами.

Ключевые особенности и преимущества использования n8n:

- **Открытый исходный код** и возможность локального развертывания обеспечивают полный контроль над данными, соответствие требованиям безопасности и независимость от облачных провайдеров.
- **Интеграция с более чем 330 сервисами**, включая CRM, ERP, e-commerce, облачные платформы, мессенджеры и базы данных.
- **Гибкость сценариев:** от простых уведомлений до сложных цепочек с обработкой API-запросов, логикой принятия решений и подключением ИИ-сервисов.
- **Поддержка JavaScript и Python:** в случае необходимости пользователи могут встраивать пользовательский код, расширяя возможности автоматизации.
- **Интуитивный визуальный интерфейс:** позволяет быстро настраивать и визуализировать все этапы процесса.

Платформы класса Low-Code предоставляют инструменты для создания цифровых решений с минимальным объёмом кода, что делает их идеальными для команд, не обладающих глубокой технической экспертизой, но нуждающихся в автоматизации процессов.

В строительстве n8n может использоваться для автоматизации различных процессов, таких как интеграция с системами управления проектами, потоковая проверка, написание готовых отчётов

и писем, автоматическое обновление данных о запасах материалов, отправка уведомлений командам о статусе задач и многое другое. Настроенный Pipeline в n8n позволяет кратно снизить количество ручных операций, уменьшить вероятность ошибок и ускорить принятие решений для выполнения проектов.

> Вы можете выбрать один из почти двух тысяч готовых, бесплатных и открытых n8n Pipeline, доступных на сайте: n8n.io/workflows, чтобы автоматизировать как рабочие процессы в строительстве, так и личные задачи, снижая рутинные операции.

Возьмём один из готовых Pipeline шаблонов, доступных бесплатно на сайте n8n.io [142], который автоматически создает черновики ответов в Gmail (Рис. 7.4-6), помогая пользователям, которые получают большой объем писем или испытывают трудности при составлении ответов.

Данный шаблон n8n “Автоответчик Gmail AI Auto-Responder: Создавай черновики ответов на входящие письма” (Рис. 7.4-6) анализирует входящие сообщения с помощью LLM от ChatGPT, определяет необходимость ответа, формирует черновик с ChatGPT и преобразует текст в HTML и добавляет его в цепочку сообщений в Gmail. При этом письмо не отправляется автоматически, что позволяет вручную редактировать и утверждать ответ. Настройка занимает около 10 минут и включает OAuth-конфигурацию Gmail API и интеграцию OpenAI API. В итоге получаем удобное и бесплатное решение для автоматизации рутинной email-коммуникации без потери контроля над содержанием писем.

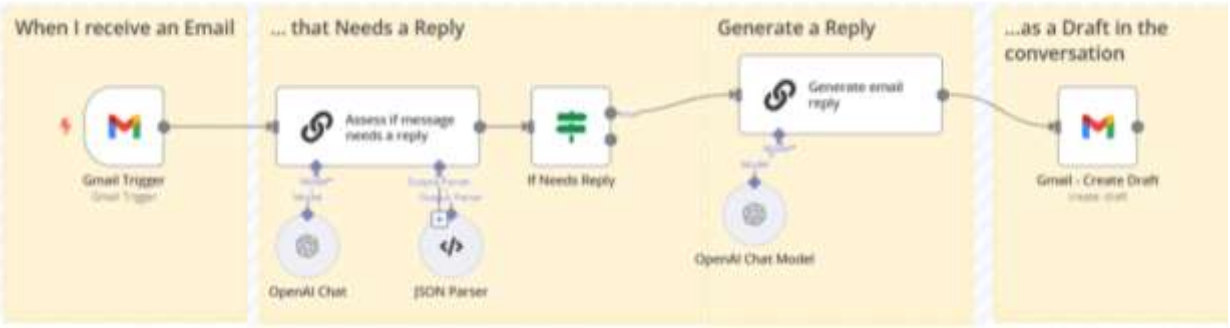


Рис. 7.4-6 Автоматизированный процесс генерации ответов на электронные письма при помощи n8n.

Другой пример автоматизации с n8n – поиск выгодных сделок на рынке недвижимости [143]. N8n Pipeline “Автоматизация ежедневных сделок с недвижимостью с помощью Zillow API, Google Sheets и Gmail”, ежедневно собирает актуальные предложения, соответствующие заданным критериям, используя Zillow API. Она автоматически рассчитывает ключевые инвестиционные метрики (Cash on Cash ROI, Monthly Cash Flow, Down Payment), обновляет Google Sheets и отправляет сводный отчет на email (Рис. 7.4-7), что позволяет инвесторам экономить время и быстро реагировать на лучшие предложения.

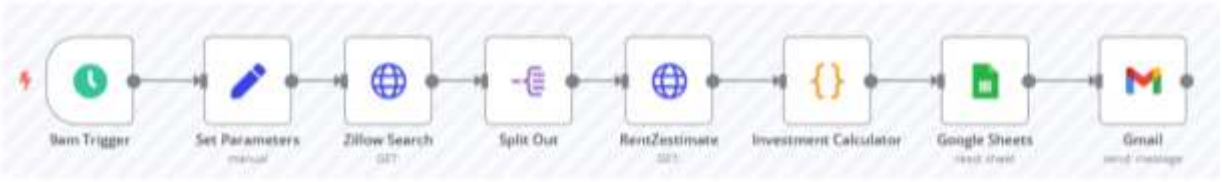


Рис. 7.4-7 Автоматизированный процесс оценки инвестиционной привлекательности недвижимости.

Благодаря своей гибкости и расширяемости, n8n становится ценным инструментом для компаний, стремящихся к цифровой трансформации и повышению конкурентоспособности на рынке при помощи относительно простых и бесплатных инструментов с открытым исходным кодом.

Такие инструменты как Apache NiFi, Airflow и n8n можно рассматривать как три уровня обработки данных (Рис. 7.4-8). NiFi управляет потоком данных, гарантируя их доставку и трансформацию, Airflow оркестрирует выполнение задач, объединяя данные в конвейеры обработки, а n8n автоматизирует интеграцию с внешними сервисами и управляет бизнес-логикой.

	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

Рис. 7.4-8 Apache Airflow, Apache NiFi и n8n можно рассматривать как три взаимодополняющих уровня современной архитектуры управления данными.

Вместе эти бесплатные и открытые инструменты потенциально формируют пример эффективной экосистемы для управления данными и процессами в строительной индустрии, позволяя компаниям эффективно использовать информацию для принятия решений и автоматизации процессов.

Дальнейшие шаги: переход от ручных операций к решениям на базе аналитики

Современные строительные компании работают в условиях высокой неопределённости: изменение цен на материалы, задержки поставок, нехватка рабочей силы и жёсткие сроки проектов. Использование аналитических дэшбордов, ETL-конвейеров и BI-систем помогает компаниям быстро находить проблемные зоны, оценивать эффективность ресурсов и прогнозировать изменения ещё

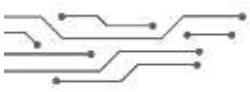
до того, как они приведут к финансовым потерям.

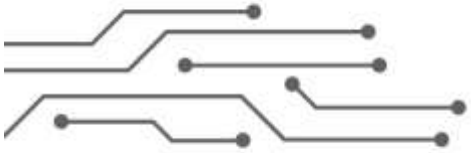
Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные технологии в ваших повседневных задачах:

- Внедрите визуализацию данных и аналитические панели
  - Освойте процесс создания информационных панелей для мониторинга ключевых показателей эффективности (KPI)
  - Используйте инструменты визуализации для своих данных (Power BI, Tableau, Matplotlib, Plotly)
- Автоматизируйте обработку данных через ETL-процессы
  - Настройте автоматический сбор данных из различных источников (документация, таблицы, CAD) через ETL процессы
  - Организуйте трансформацию данных (например проверку через регулярные выражения или расчёт) с применением скриптов на Python
  - Попробуйте настроить автоматическое формирование отчётов в PDF (или DOC) с помощью библиотеки FPDF, используя данные из Excel-файлов или извлекая информацию из других PDF-документов
- Используйте языковые модели (LLM) для автоматизации
  - Используйте большие языковые модели (LLM), для генерации кода, который поможет извлекать и анализировать данные из неструктурированных документов
  - Ознакомьтесь с инструментом автоматизации n8n и изучите готовые шаблоны и кейсы на их сайте. Определите, какие процессы из вашей работы можно полностью автоматизировать с помощью подхода No-Code/Low-Code

Аналитический подход к данным и автоматизация процессов не только сокращают время на рутинные операции, но и повышают качество принимаемых решений. Компании, внедряющие инструменты визуальной аналитики и ETL-конвейеры, получают возможность оперативно реагировать на изменения.

Автоматизация бизнес-процессов с использованием инструментов вроде n8n, Airflow и NiFi — это лишь первый шаг к цифровой зрелости. Следующим этапом становится качественное хранение и управление самими данными, которые лежат в основе автоматизации. В восьмой части мы подробно рассмотрим, как строительные компании могут выстраивать устойчивую архитектуру хранения данных, переходя от хаоса документов и разноформатных файлов к централизованным хранилищам и аналитическим платформам.





VIII ЧАСТЬ
ХРАНЕНИЕ И УПРАВЛЕНИЕ ДАННЫМИ В СТРОИТЕЛЬСТВЕ

Восьмая часть исследует современные технологии хранения и управления данными в строительной отрасли. Здесь анализируются эффективные форматы для работы с большими объемами информации — от простых CSV и XLSX до более производительных Apache Parquet и ORC с подробным сравнением их возможностей и ограничений. Рассматриваются концепции хранилищ данных (DWH), озёр данных (Data Lakes) и их гибридных решений (Data Lakehouse), а также принципы управления данными (Data Governance) и минимализма при работе с информацией (Data Minimalism). Подробно освещаются проблемы "болота данных" (Data Swamp) и стратегии по предотвращению хаоса в информационных системах. Представлены новые подходы к работе с данными, включая векторные базы данных и их применение в строительстве через концепцию Bounding Box. В этой части также затрагиваются методологии DataOps и VectorOps как новые стандарты организации рабочих процессов с данными.

ГЛАВА 8.1.
ИНФРАСТРУКТУРА ДАННЫХ: ОТ ФОРМАТОВ ХРАНЕНИЯ ДО ЦИФРОВЫХ ХРАНИЛИЩ

Атомы данных: фундамент эффективного управления информацией

Все во Вселенной состоит из мельчайших строительных блоков - атомов и молекул, и со временем все живое и неживое неизбежно возвращается в это исходное состояние. В природе этот процесс происходит с поразительной скоростью, которую мы пытаемся перенести на процессы, управляемые человеком.

> В лесу любые живые организмы со временем превращаются в питательную субстанцию, которая служит основой для новых растений. Эти растения, в свою очередь, становятся пищей для новых живых существ, состоящих из тех же атомов, которые миллионы лет назад создали Вселенную.

В мире бизнеса также важно разбивать сложные многослойные структуры на наиболее фундаментальные, минимально обрабатываемые единицы — подобно атомам и молекулам в природе. Это позволяет эффективно хранить и управлять атомы данных, превращая их в богатую, плодородную основу, которая становится ключевым ресурсом для роста аналитики и качества принятия решений.

Рис. 8.1-1 Анализ и принятие решений основывается на повторно используемых данных, которые когда-то были обработаны и сохранены.

Музыкальные композиции состоят из нот, которые, соединяясь, создают сложные музыкальные произведения, а слова создаются из примитивной единицы - буквы-звука. Будь то природа, наука, экономика, искусство или технология, мир демонстрирует удивительное единство и гармонию в своем стремлении к разрушению, структуре, цикличности и созиданию. Точно так же процессы в системах калькуляции себестоимости разбиваются на мельчайшие структурированные единицы - статьи ресурсов - на уровне калькуляций и расписаний. Затем эти единицы, как ноты, используются для формирования более сложных расчетов и графиков. По такому же принципу работают системы автоматизированного проектирования, в которых сложные архитектурные и инженерные проекты строятся из базовых элементов - отдельных элементов и библиотечных компонентов, из которых создается полная 3D-модель проекта сложного здания или сооружения.

> Концепция цикличности и структурности, присущая природе и науке, находит отражение и в современном мире данных. Как в природе все живые существа возвращаются к атомам и молекулам, так и в мире современных инструментов обработки данных информация стремится перейти к самой примитивной форме.

Мельчайшие элементы с их конечной неделимостью являются основными строительными блоками бизнес-процессов. Важно с самого начала тщательно продумать, как собирать, структурировать (дробить на атомы) и хранить эти мельчайшие строительные блоки из различных источников. При этом организация и хранение данных – это не только вопрос их разбития на составные части. Не менее важно обеспечить их интеграцию и структурированное хранение, чтобы данные можно было в любой момент, когда они понадобятся, легко извлекать, анализировать и использовать для принятия решений.

Для эффективной обработки информации необходимо тщательно выбирать формат и методы хранения данных — подобно тому, как почва должна быть подготовлена для роста деревьев. Хранилища данных должны быть организованы таким образом, чтобы обеспечивать высокое качество и актуальность информации, исключая избыточные или нерелевантные данные. Чем лучше структурирована эта «информационная почва», тем быстрее и точнее пользователи смогут находить нужные данные и решать аналитические задачи.

Хранилище информации: файлы или данные

Хранилища данных позволяют компаниям собирать и объединять информацию из различных систем, создавая единый центр для последующий аналитики. Собранные исторические данные дают возможность не только глубже анализировать процессы, но и выявлять закономерности, которые могут повлиять на эффективность бизнеса.

Допустим, компания ведёт несколько объектов одновременно. Инженер хочет понять, сколько бетона уже залито и какие объёмы ещё предстоит закупить. При традиционном подходе ему придётся вручную искать на сервере и открывать несколько сметных таблиц, сравнить их с актами выполненных работ и проверить актуальные складские остатки. Это занимает часы, а то и дни. Даже при наличии ETL-процессов и автоматических скриптов задача остаётся полу ручной: инженеру всё равно приходится вручную указывать путь к папкам или конкретным файлам на сервере. Это снижает общий эффект от автоматизации, поскольку продолжает отнимать ценное рабочее время.

При переходе же на управление данными вместо работы с файловой системой сервера инженер получает доступ к единой структуре хранилища, где информация обновляется в режиме реального времени. Один запрос — в виде кода, SQL-запроса или даже обращения к LLM-агенту — позволяет мгновенно получить точные данные о текущих остатках, объёмах выполненных работ и предстоящих поставках, если данные заранее были переподготовлены и объединены в виде хранилища данных, где не нужно блуждать по папкам, открывать десятки файлов и вручную сопоставлять значения.

Долгое время строительные компании использовали PDF-документы, DWG-чертежи, RVT-модели и сотни и тысячи Excel-таблиц и других разрозненных форматов, которые хранятся в определённых папках на серверах компании, что усложняло поиск информации, её проверку и анализ. В итоге файлы, оставшиеся после завершения проектов, чаще всего перемещаются обратно на сервер в архивные папки-хранилища, которые практически не используются в дальнейшем. Подобное традиционное файловое хранение данных с увеличением потока данных теряет актуальность, из-за своей уязвимости к ошибкам, вызванным человеческим фактором.

> **Файл** — это всего лишь изолированный контейнер, в котором хранятся данные. Файлы создаются для людей, а не для систем, поэтому требуют ручного открытия, чтения и интерпретации. Примером может служить Excel-таблица, PDF-документ или CAD-чертёж, который нужно специально открыть в определённом инструменте, чтобы получить доступ к нужной информации. Без структурированного извлечения и обработки информация в нём остаётся неиспользуемой.
>
> **Данные**, в свою очередь, — это машиночитаемая информация, связанная, обновляемая и анализируемая автоматически. В едином хранилище данных (например, базе данных, DWH или Data Lake) информация представлена в виде таблиц, записей и связей. Это обеспечивает возможность единообразного хранения, выполнения автоматических запросов, анализа значений и построения отчётности в режиме реального времени.

Использование данных вместо файлов (Рис. 8.1-1) позволяет отказаться от ручного процесса поиска, и унифицировать процессы обработки. Компании, которые уже внедряют подобный подход, получают конкурентное преимущество благодаря скорости доступа к информации и возможности её быстрой интеграции в бизнес-процессы.

> Переход от использования файлов к данным — это неизбежное изменение, которое определит будущее строительной отрасли.

Каждая компания строительной отрасли столкнётся с ключевым выбором: продолжать хранить информацию в разрозненных файлах и силосах, которые должны читаться людьми при помощи специальных программ или трансформировать её на первых этапах обработки в структурированные данные, создавая единую интегрированную цифровую основу для автоматизированного управления проектами.

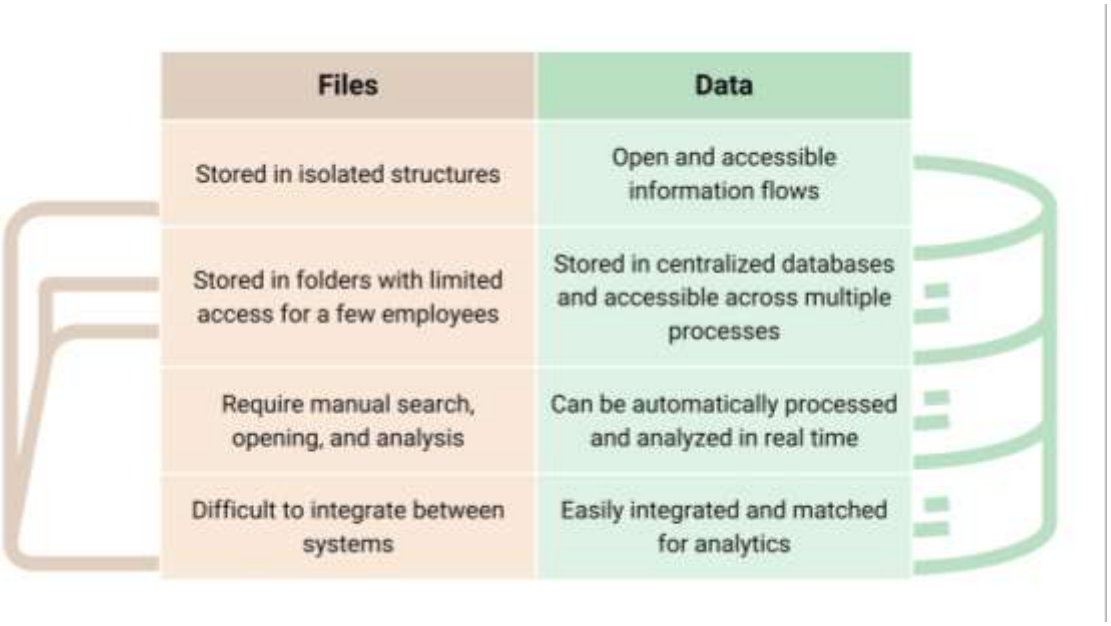


Рис. 8.1-1 Эволюция информационного потока: от изолированных файлов к интегрированным данным.

В условиях стремительного роста объёмов информации традиционные методы хранения и обработки файлов становятся всё менее эффективными. В строительной отрасли, как и в других секторах, уже недостаточно полагаться на разрозненные папки с файлами различных форматов или несвязанные между собой базы данных.

Компании, стремящиеся сохранить конкурентоспособность в эпоху цифровых технологий, неизбежно будут переходить к интегрированным цифровым платформам, использовать технологии больших данных и автоматизированные системы аналитики.

Переход от файлового хранения к работе с данными потребует переосмысления подходов к управлению информацией и осознанного выбора форматов, пригодных для дальнейшей интеграции в централизованные хранилища. От этого выбора зависит, насколько эффективно будут обрабатываться данные, насколько быстро к ним можно получить доступ и насколько просто их будет интегрировать в цифровые процессы компании.

Хранение больших данных: анализ популярных форматов и их эффективность

Форматы хранения играют ключевую роль в обеспечении масштабируемости, надёжности и производительности аналитической инфраструктуры. Для анализа и обработки данных — таких как фильтрация, группировка и агрегирование — в наших примерах использовался Pandas DataFrame — популярная структура для работы с данными в оперативной памяти.

Однако Pandas DataFrame не имеет собственного формата хранения, поэтому после завершения обработки данные экспортируются в один из внешних форматов — чаще всего CSV или XLSX. Эти табличные форматы удобны для обмена и совместимы с большинством внешних систем, но имеют

ряд ограничений: низкую эффективность хранения, отсутствие сжатия и слабую поддержку версионирования:

- **CSV** (Comma-Separated Values): простой текстовый формат, широко поддерживаемый различными платформами и инструментами. Он прост в использовании, но не поддерживает сложные типы данных и сжатие.
- **XLSX** (Excel Open XML Spreadsheet): формат файлов Microsoft Excel, поддерживающий такие сложные функции, как формулы, диаграммы и стилизация. Хотя он удобен для ручного анализа и визуализации данных, он не оптимизирован для крупномасштабной обработки данных.

Помимо популярных табличных XLSX и CSV, существует несколько популярных форматов для эффективного хранения структурированных данных (Рис. 8.1-2), каждый из которых обладает уникальными преимуществами в зависимости от конкретных требований к хранению и анализу данных:

- **Apache Parquet**: формат файлов для колоночного хранения данных, оптимизированный для использования в системах анализа данных. Он предлагает эффективные схемы сжатия и кодирования данных, что делает его идеальным для сложных структур данных и обработки больших данных.
- **Apache ORC** (Optimized Row Columnar - оптимизированная колонна строк): подобно Parquet, ORC обеспечивает высокую степень сжатия и эффективное хранение данных. Он оптимизирован для тяжелых операций чтения и хорошо подходит для хранения озер данных.
- **JSON** (JavaScript Object Notation): хотя JSON не так эффективен с точки зрения хранения данных по сравнению с двоичными форматами, такими как Parquet или ORC, он очень доступен и прост в работе, что делает его идеальным для сценариев, где важна читабельность и совместимость с веб-технологиями.
- **Feather**: быстрый, легкий и простой в использовании формат хранения бинарных колоночных данных, ориентированный на аналитику. Он разработан для эффективной передачи данных между Python (Pandas) и R, что делает его отличным выбором для проектов, в которых задействованы эти среды программирования.
- **HDF5** (Hierarchical Data Format version 5): предназначен для хранения и организации больших объемов данных. Он поддерживает широкий спектр типов данных и отлично подходит для работы со сложными коллекциями данных. HDF5 особенно популярен в научных вычислениях благодаря своей способности эффективно хранить и получать доступ к большим наборам данных.

	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)

Рис. 8.1-2 Сравнение форматов данных с указанием основных различий в аспектах хранения и обработки.

Для проведения сравнительного анализа форматов, применяемых на этапе Load в процессе ETL, была составлена таблица, демонстрирующая размеры файлов и время их чтения (Рис. 8.1-3). В рамках исследования использовались файлы с идентичными данными: таблица содержала 10 000 строк и 10 столбцов, заполненных случайными значениями.

В исследование включены следующие форматы хранения: CSV, Parquet, XLSX и HDF5, а также их сжатые версии в ZIP-архивах. Исходные данные были сгенерированы с использованием библиотеки NumPy и представлены в виде структуры Pandas DataFrame. Процесс тестирования состоял из следующих этапов:

- Сохранение файлов: датафрейм сохранён в четыре различных формата: CSV, Parquet, XLSX, и HDF5. Каждый формат обладает уникальными особенностями по способу хранения данных, влияющим на размер файла и скорость его чтения.
- Сжатие файлов в ZIP: для анализа эффективности стандартного сжатия, каждый файл был дополнительно сжат в ZIP-архив.
- Чтение файлов (ETL – Load): время чтения измерялось для каждого файла после его распаковки из ZIP. Это позволяет оценить скорость доступа к данным после извлечения из архива.

Важно отметить, что Pandas DataFrame не использовался напрямую при анализе размеров или времени чтения, так как не представляет собой самостоятельный формат хранения. Он служил лишь

промежуточной структурой для генерации и последующего сохранения данных в различные форматы.

	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

Рис. 8.1-3 Сравнение форматов хранения данных по размеру и скорости чтения.

CSV и HDF5 файлы демонстрируют (Рис. 8.1-3) высокую эффективность сжатия, значительно уменьшая свой размер при упаковке в ZIP, что может быть особенно полезно в сценариях, требующих оптимизации хранения. XLSX файлы, в свою очередь, практически не поддаются сжатию, и их размер в ZIP остаётся сопоставимым с оригинальным, что делает их менее выгодными для использования в больших объемах данных или в условиях, где важна скорость доступа к данным. Кроме того, время чтения для XLSX значительно выше по сравнению с другими форматами, что делает его менее предпочтительным для быстрых операций чтения данных. Apache Parquet благодаря своей колоночной структуре продемонстрировал высокую эффективность для аналитических задач и больших объёмов данных.

Оптимизация хранения данных с Apache Parquet

Одним из популярных форматов для хранения и обработки больших данных является Apache Parquet. Этот формат разработан специально для колоночного хранения (подобный Pandas), что позволяет существенно сократить объем занимаемой памяти и повысить скорость аналитических запросов. В отличие от традиционных форматов, таких как CSV и XLSX, Parquet поддерживает встроенное сжатие и оптимизирован для работы с системами обработки больших данных, включая Spark, Hadoop и облачные хранилища.

> Ключевые особенности Parquet включают поддержку сжатия и кодирования данных, что значительно уменьшает размер хранилища и ускоряет операции чтения данных за счет работы непосредственно с нужными столбцами, а не со всеми строками данных.

Для наглядного примера того, как легко получить необходимый код для преобразования данных в Apache Parquet, воспользуемся LLM.

- Отправьте текстовый запрос в LLM чат (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Напиши код для сохранения данных из Pandas DataFrame в Apache Parquet. ⏎

- Ответ LLM:

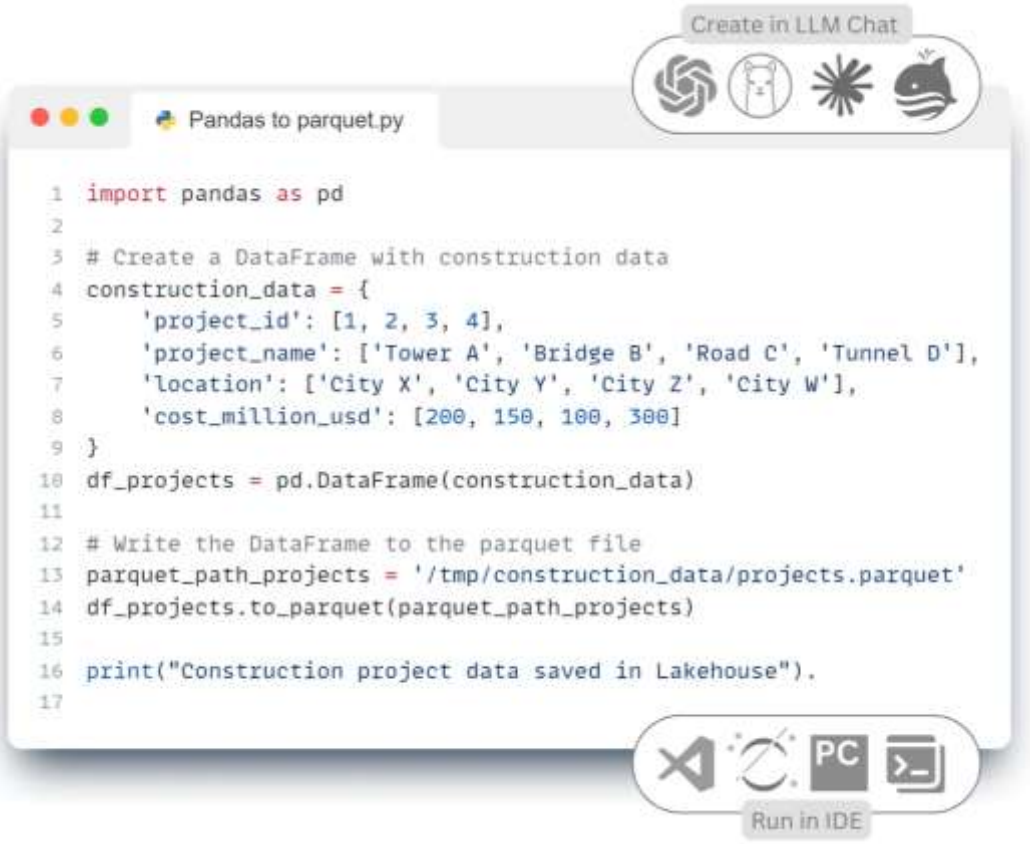


Рис. 8.1-4 Передача данных Dataframe из оперативной памяти в эффективный для хранения формат Apache Parquet при помощи нескольких строк Python.

Следующий пример: смоделируем процесс ETL с данными, сохраненными в формате Parquet для фильтрации проектов по определённому значению одного из атрибутов „cost_million_usd“ (Рис.

8.1-4).

- В продолжении чата, отправьте текстовый запрос в LLM:

> Напиши код, в котором мы хотим отфильтровать данные в таблице и сохранить только те проекты (строки таблицы) из данных Apache Parquet, стоимость которых (параметр cost_million_usd) превышает 150 миллионов долларов. ⏎

- Ответ LLM:

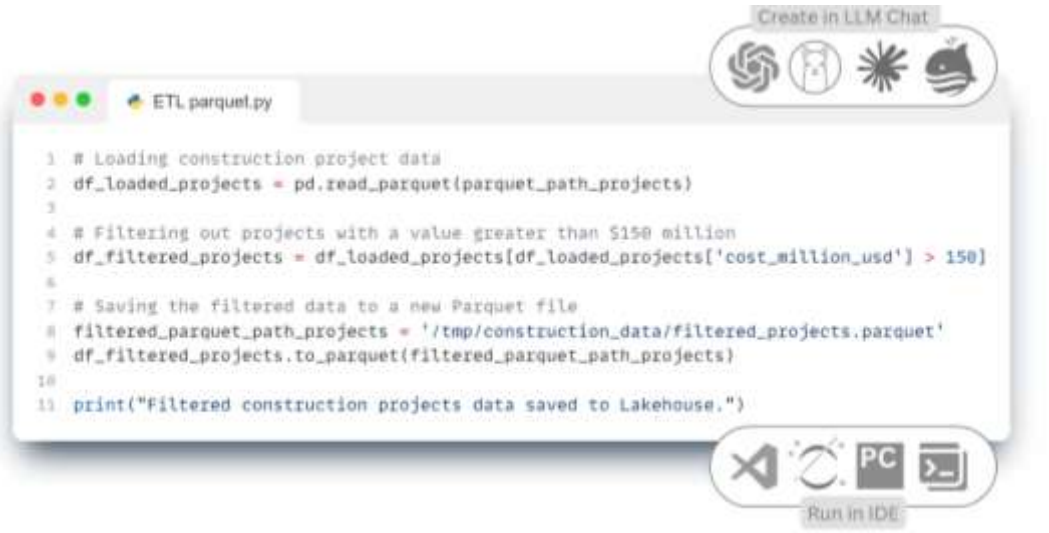


Рис. 8.1-5 Процесс ETL при работе с данными в формате Apache Parquet выглядит так же, как с другими структурированными форматами.

Использование формата Parquet (по отношению к XLSX, CSV и др.) значительно уменьшает объем хранимой информации и ускоряет операции поиска. Благодаря этому он отлично подходит как для хранения, так и для анализа данных. Parquet интегрируется с различными системами обработки, обеспечивая эффективный доступ в гибридных архитектурах.

Тем не менее, эффективный формат хранения — лишь один из элементов полноценной работы с данными. Для создания устойчивой и масштабируемой среды необходима чётко спроектированная архитектура управления данными. Именно эту функцию выполняют системы класса DWH (Data Warehouse). Они обеспечивают агрегацию данных из разнородных источников, прозрачность бизнес-процессов и возможность комплексного анализа с использованием BI-инструментов и алгоритмов машинного обучения.

DWH: Data Warehouse хранилища данных

Подобно тому, как формат Parquet оптимизирован для эффективного хранения больших объемов информации, Data Warehouse оптимизирована для интеграции и структурирования данных с целью

поддержки аналитики, прогнозирования и принятия управленческих решений.

В современных компаниях данные поступают из множества разрозненных источников: ERP, CAFM, CPM, CRM-систем, бухгалтерского и складского учёта, цифровых CAD моделей зданий, сенсоров IoT и других решений. Для получения целостной картины недостаточно просто собирать данные — их необходимо организовать, стандартизировать и централизовать в едином хранилище. Именно эту функцию выполняет DWH — централизованная система хранения, позволяющая агрегировать информацию из различных источников, структурировать её и сделать доступной для аналитики и стратегического управления.

> **DWH (Data Warehouse)** — это централизованная система хранения данных, которая агрегирует информацию из множества источников, структурирует ее и делает доступной для аналитики и отчетности.

Во многих компаниях данные разбросаны по разным системам, которые мы рассматривали в первых частях книги (Рис. 1.2-4). DWH интегрирует эти источники, обеспечивая полную прозрачность и достоверность информации. Хранилище данных DWH – это специализированная база (большая база данных), которая собирает, обрабатывает и хранит данные из множества источников. Основные характеристики DWH:

- **Использование ETL-процессов** (Extract, Transform, Load) – извлечение данных из источников, их очистка, трансформация, загрузка в хранилище и автоматизация этих процессов, о которых велась речь в седьмой части книги.
- **Гранулярность данных** – данные в DWH могут храниться как в агрегированном виде (сводные отчеты), так и в детализированном (сырые данные). С 2024 года именно CAD-вендоры начали говорить про гранулированные данные [125], что, возможно, свидетельствует о подготовке отрасли к переходу на использование специализированных облачных хранилищ для работы с данными цифровых моделей зданий.
- **Поддержка аналитики и прогнозирования** – хранилища данных обеспечивают основу для BI-инструментов, Big Data-анализа и машинного обучения.

DWH служит основой для бизнес-аналитики, позволяя анализировать ключевые показатели эффективности, прогнозировать продажи, закупки и затраты, а также автоматизировать отчетность и визуализацию данных (Рис. 8.1-6).

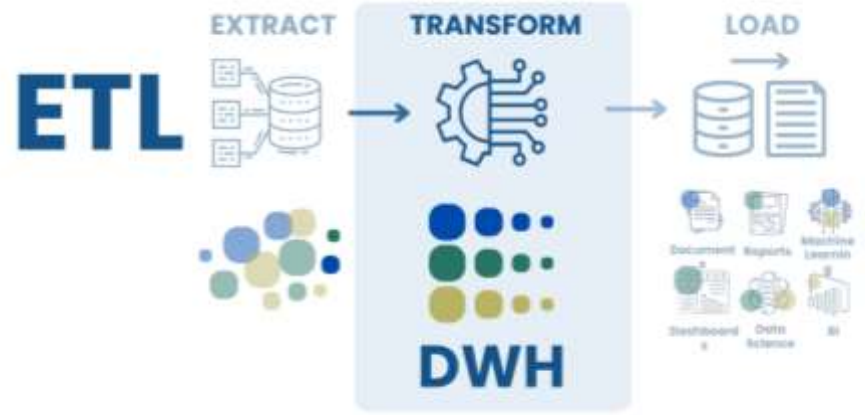


Рис. 8.1-6 В ETL-процессе DWH может выступать в роли центрального хранилища, где данные, извлеченные из различных систем, проходят этапы трансформации и выгрузки.

DWH играет ключевую роль в интеграции, очистке и структурировании информации, создавая прочную основу для бизнес-аналитики и процессов принятия решений. Однако в современных условиях, когда объемы данных стремительно растут, а их источники становятся все более разнообразными, традиционный подход к хранению информации DWH часто требует расширения в виде подходов ELT и Data Lake.

Data Lake - эволюция ETL в ELT: от традиционной очистки к гибкой обработке

Классические DWH-хранилища данных, предназначенные для хранения структурированных данных в формате, оптимизированном для аналитических запросов, столкнулись с ограничениями в работе с неструктурированными данными и масштабируемостью. В ответ на эти проблемы появились озера данных (Data Lakes), предлагающие гибкое хранение больших объемов разнородных данных.

> **Data Lake** предлагает альтернативный DWH-подход, позволяющий работать с неструктурированными, полуструктурированными и сырыми данными без предварительной жесткой схемы. Этот метод хранения часто актуален для обработки данных в реальном времени, машинного обучения и продвинутой аналитики. В отличие от DWH, которое структурирует и агрегирует данные перед загрузкой, Data Lake позволяет сохранять информацию в исходном виде, обеспечивая таким образом гибкость и масштабируемость.

Именно разочарование в традиционных хранилищах данных (RDBMS, DWH) и интерес к «большим данным» привели к появлению озер данных, где вместо сложного ETL теперь данные просто загружаются в слабоструктурированное хранилище, а их обработка происходит уже на этапе анализа:

- В традиционных хранилищах данных данные обычно проходят предварительную обработку, преобразование и очистку (ETL - Extract, Transform, Load) перед загрузкой в хранилище (Рис. 8.1-6). Это означает, что данные структурируются и оптимизируются для решения конкретных будущих задач аналитики и отчетности. Основной упор делается на поддержание высокой производительности запросов и целостности данных. Однако такой подход может быть дорогостоящим и менее гибким с точки зрения интеграции новых типов данных и быстро меняющихся схем данных.
- Озера данных, с другой стороны, предназначены для хранения больших объемов необработанных данных в их исходном формате (Рис. 8.1-7). На смену процессу ETL (Extract, Transform, Load), приходит ELT (Extract, Load, Transform), когда данные сначала загружаются в хранилище "как есть" и только потом могут быть преобразованы и проанализированы по мере необходимости. Это обеспечивает большую гибкость и возможность хранения разнородных данных, включая неструктурированные данные, такие, как текст, изображения и журналы.

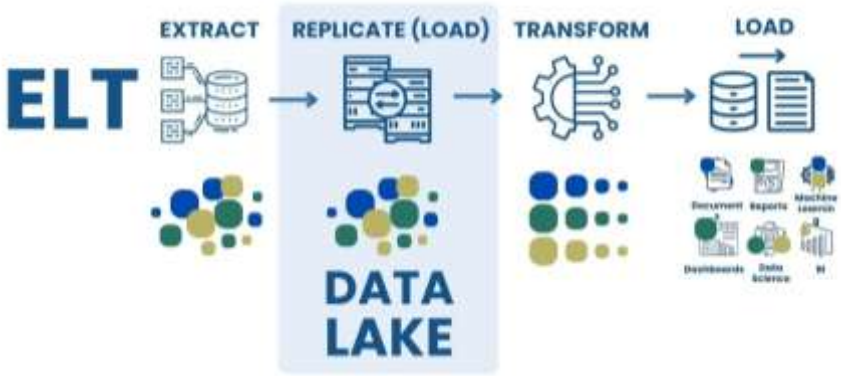


Рис. 8.1-7 В отличие от ETL, в Data Lake используется ELT, при котором информация сначала загружается в «сыром» виде, а трансформация выполняется уже на этапе выгрузки.

Традиционные хранилища данных ориентированы на предварительную обработку данных для обеспечения высокой производительности запросов, в то время как в озерах данных приоритет отдается гибкости: они хранят необработанные данные и преобразуют их по мере необходимости (Рис. 8.1-8).

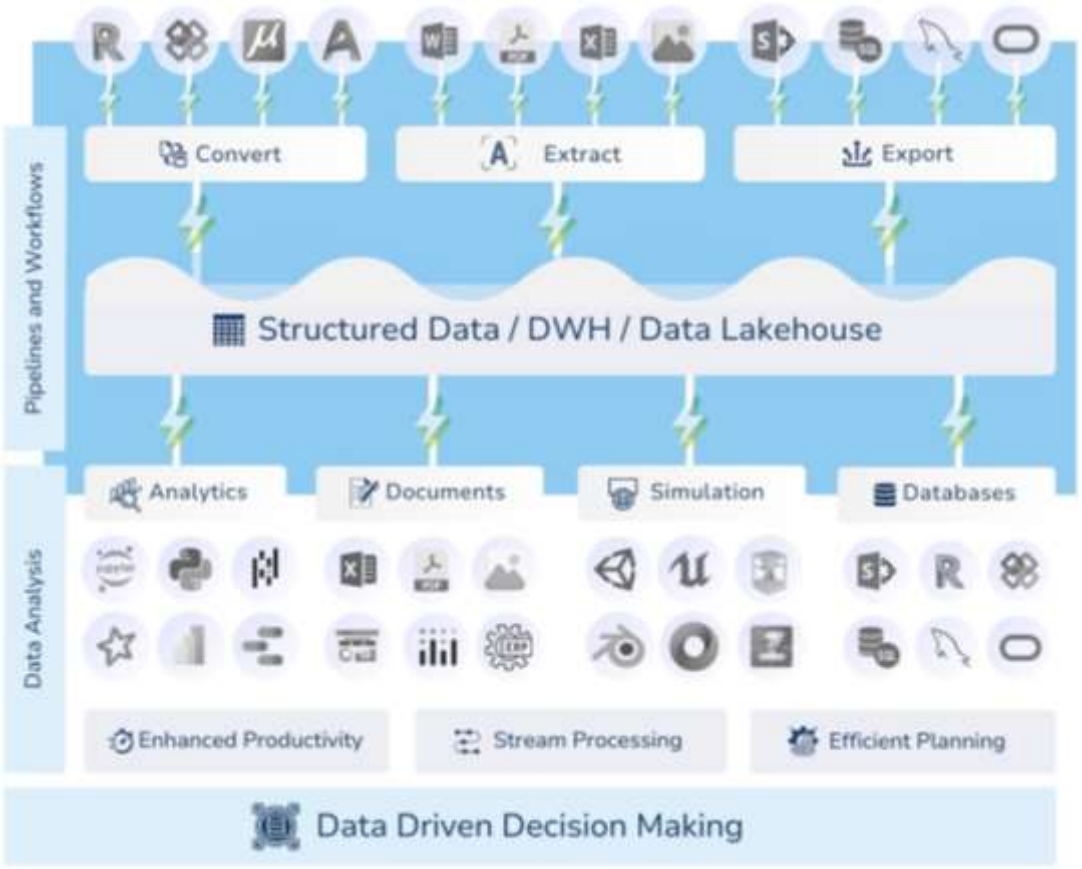


Рис. 8.1-8 Современные концепции хранилищ направлены на хранение и обработки всех типов данных для целей принятия решений.

Тем не менее, несмотря на все преимущества и озера данных не лишены недостатков. Отсутствие строгой структуры и сложность управления информацией могут привести к хаосу, в котором данные дублируются, противоречат друг другу или теряют свою актуальность. Кроме того, поиск и анализ данных в таком хранилище требуют значительных усилий, особенно при работе с разнородной информацией. Чтобы преодолеть эти ограничения и объединить лучшие черты традиционных хранилищ и озер данных, была разработана архитектура Data Lakehouse.

Архитектура Data Lakehouse: синергия хранилищ и озер данных

Чтобы объединить лучшие характеристики DWH (структурированность, управляемость, высокая производительность аналитики) и Data Lake (масштабируемость, работа с разнородными данными), был разработан подход Data Lakehouse. Эта архитектура сочетает гибкость озер данных с мощными инструментами обработки и управления, характерными для традиционных хранилищ, обеспечивая баланс между хранением, аналитикой и машинным обучением. Data Lakehouse — это синтез озер данных и хранилищ данных, сочетающий гибкость и масштабируемость первых с управляемостью и оптимизацией запросов вторых.

Data Lakehouse — это архитектурный подход, который стремится объединить гибкость и масштабируемость озер данных с управляемостью и производительностью запросов в хранилищах данных (Рис. 8.1-9).

Ключевые особенности Data Lakehouse включают:

- **Открытый формат хранения данных:** использование открытых форматов для хранения данных, таких как Apache Parquet, обеспечивает эффективность и оптимизацию запросов.
- **Схема только для чтения:** в отличие от традиционного подхода к схеме исключительно для записи в DWH, Lakehouse поддерживает схему только для чтения, что позволяет более гибко управлять структурой данных.
- **Гибкость и масштабируемость**: поддерживает хранение и анализ структурированных и неструктурированных данных, обеспечивая высокую производительность запросов за счет оптимизации на уровне хранилища.

Data Lakehouse предлагает компромиссное решение, сочетающее преимущества обоих подходов, что делает его идеальным для современных аналитических нагрузок, требующих гибкости в обработке данных.

Рис. 8.1-9 Data Lakehouse - следующее поколение систем хранения данных, созданных для удовлетворения сложных и постоянно меняющихся требований.

Идея современных хранилищ данных кажется простой: если все данные находятся в одном месте, их легче анализировать. Однако на практике все не так гладко. Представьте, что компания решила полностью отказаться от привычных систем учета и управления (ERP, PMIS, CAFM или др.), заменив их одним огромным озером данных, к которому у всех есть доступ. Что произойдет? Скорее всего, начнется хаос: данные будут дублироваться, противоречить друг другу, а критически важная информация окажется потерянной или искаженной. Даже если озеро данных используется только для аналитики, без грамотного управления с ним возникают серьезные трудности:

- Данные сложно понять: в обычных системах у данных есть четкая структура, а в озере — просто огромное скопление файлов и таблиц. Чтобы что-то найти, специалисту приходится разбираться - за что отвечает каждая строка и столбец.
- Данные могут быть неточными: если в одном месте хранится много версий одной и той же информации, сложно понять, какая из них актуальна. В результате принимаются решения на основе устаревших или ошибочных данных.
- Сложно подготовить данные для работы: данные нужно не только хранить, но и представлять в удобной форме — в виде отчетов, графиков, таблиц. В традиционных системах это делается автоматически, а в озерах данных требует дополнительной обработки.

В итоге каждая концепция хранения данных имеет свои особенности, подходы к обработке и применения в бизнесе. Традиционные базы данных ориентированы на транзакционные операции, хранилища данных (DWH) обеспечивают структуру для аналитики, озера данных (Data Lake) хранят информацию в сыром виде, а гибридные хранилища (Data Lakehouse) совмещают преимущества DWH и Data Lake (Рис. 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

Рис. 8.1-10 DWH, Data Lake и Data Lakehouse: ключевые различия в типах данных, сценариях использования, методах обработки и подходах к хранению.

Выбор архитектуры хранения данных — сложный процесс, зависящий от потребностей бизнеса, объема информации и требований к аналитике. Каждое решение имеет свои плюсы и минусы: DWH обеспечивает структурированность, Data Lake — гибкость, а Lakehouse — баланс между ними. Организации редко ограничиваются одной архитектурой данных.

> Вне зависимости от выбранной архитектуры, автоматизированные системы управления данными существенно превосходят ручные методы. Они позволяют минимизировать человеческие ошибки, ускоряют обработку информации, обеспечивают прозрачность и прослеживаемость данных на всех этапах бизнес-процессов.

И если централизованные хранилища данных уже стали отраслевым стандартом во многих сферах экономики, то в строительстве ситуация по-прежнему остаётся фрагментированной. Данные здесь

распределены между разными платформами (CDE, PMIS, ERP и др.), что затрудняет создание единой картины происходящего и требует архитектур, способных объединить эти источники в целостную, аналитически пригодную цифровую среду.

CDE, PMIS, ERP или DWH и Data Lake

В некоторых компаниях, работающих в сфере строительства и проектирования, уже применяется концепция общей среды данных (Common Data Environment, CDE) в соответствии с ISO 19650. По сути, CDE выполняет те же функции, что и хранилище данных (DWH) в других отраслях: централизует информацию, обеспечивает контроль версий, предоставляет доступ к проверенной информации.

> **Общая среда данных (CDE)** – это централизованное цифровое пространство, используемое для управления, хранения, обмена и совместной работы с проектной информацией на всех этапах жизненного цикла объекта. CDE часто реализуется с использованием облачных технологий и интегрируется с CAD (BIM) системами.

Финансовый сектор, ритейл, логистика и промышленность десятилетиями используют централизованные системы управления данными, объединяющие информацию из разных источников, контролирующие ее актуальность и обеспечивающие аналитику. CDE развивает эти принципы, адаптируя их под задачи проектирования и управления жизненным циклом зданий.

Как и DWH, CDE структурирует данные, фиксирует изменения и обеспечивает единый доступ к проверенной информации. С переходом на облачные технологии и интеграцию с аналитическими инструментами различия между ними становятся все менее заметными. Добавляя к CDE гранулированные данные, концепция которых обсуждается CAD-вендорами с 2023 года [93, 125], можно увидеть еще больше параллелей с классическими DWH.

Ранее в главе «Строительные ERP и PMIS системы» мы уже рассмотрели PMIS (Project Management Information System) и ERP (Enterprise Resource Planning). В строительных проектах CDE и PMIS работают вместе: CDE служит хранилищем данных, включая чертежи, модели и проектную документацию, а PMIS управляет процессами, такими как контроль сроков, задач, ресурсов и бюджета.

ERP, отвечая за управление бизнесом в целом (финансы, закупки, персонал, производство), может интегрироваться с PMIS, обеспечивая контроль затрат и бюджетов на уровне компании. Для аналитики и отчетности может применяться DWH, которое поможет собирать, структурировать и агрегировать данные из CDE, PMIS и ERP, что оценивать финансовые показатели KPI (ROI) и выявлять закономерности. В свою очередь, Data Lake (DL) может дополнять DWH, храня сырые и неструктурированные данные (например, логи, сенсорные данные, изображения). Эти данные могут быть обработаны и загружены в DWH для дальнейшего анализа.

> Таким образом, CDE и PMIS ориентированы на управление проектами, ERP – на бизнес-процессы, а DWH и Data Lake – на аналитику и работу с данными.

В сравнении систем CDE, PMIS и ERP с DWH и Data Lake можно заметить значительные различия в

плане независимости от вендоров, стоимости, гибкости интеграции, независимости данных, скорости адаптации к изменениям, а также аналитических возможностей (Рис. 8.1-11). Традиционные системы, такие как CDE, PMIS и ERP, часто связаны с конкретными решениями и стандартами вендоров, что делает их менее гибкими и увеличивает их стоимость из-за лицензий и поддержки. К тому же данные в таких системах часто заключены в проприетарных закрытых форматах, что ограничивает их использование и анализ.

	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)

Рис. 8.1-11 DWH и Data Lake предлагают большую гибкость и независимость данных по сравнению с системами вроде CDE, PMIS и ERP.

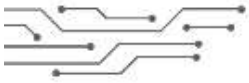
В отличие от них, DWH и Data Lake предоставляют большую гибкость в интеграции с различными источниками данных, а их использование открытых технологий и платформ помогает снизить общую стоимость владения. Более того, DWH и Data Lake поддерживают широкий спектр аналитических инструментов, что расширяет возможности анализа и управления.

С развитием инструментов обратного-инжиниринга для CAD-форматов и при наличии доступа к базам данных CAD-приложений всё острее встаёт вопрос: насколько оправдано продолжать использовать закрытые, изолированные платформы, если проектные данные должны быть доступны широкому кругу специалистов, работающих в десятках подрядных и проектных организаций?

Такая технологическая зависимость от конкретного вендора может существенно ограничивать гибкость управления данными, замедлять реакции на изменения в проекте и препятствовать эффективной коллаборации между участниками.

Традиционные подходы к управлению данными — включая DWH, Data Lake, CDE и PMIS — ориентированы преимущественно на хранение, структурирование и обработку информации. Однако с развитием искусственного интеллекта и машинного обучения возрастает потребность в новых способах организации данных, которые позволяют не только агрегировать, но и выявлять сложные взаимосвязи, находить скрытые закономерности и обеспечивать мгновенный доступ к наиболее релевантной информации.

В этом направлении особую роль начинают играть векторные базы данных — новый тип хранилищ, оптимизированный для работы с высокоразмерными эмбеддингами.



ГЛАВА 8.2.
УПРАВЛЕНИЕ ХРАНИЛИЩАМИ ДАННЫХ И ПРЕДОТВРАЩЕНИЕ ХАОСА

Векторные базы данных и Bounding Box

Векторные базы данных — это новый класс хранилищ, которые не просто сохраняют данные, но позволяют осуществлять поиск по смыслу, сравнение объектов по семантической близости и создавать интеллектуальные системы: от рекомендаций до автоматического анализа и генерации контекста. В отличие от традиционных баз, ориентированных на точные совпадения, векторные БД находят похожие объекты на основе признаков — даже при отсутствии точного совпадения.

> **Векторная база данных** — это специализированный тип базы данных, который хранит данные в виде многомерных векторов, каждый из которых представляет определенные характеристики или качества. Эти векторы могут иметь различное количество измерений, в зависимости от сложности данных (в одном случае это может быть несколько измерений, а в другом — тысячи).

Основное преимущество векторных баз — это поиск по семантической значимости, а не по точному совпадению значений. Вместо SQL- и Pandas-запросов с фильтрами «равно» или «содержит», используется поиск ближайших соседей (k-NN) (подробнее о k-NN будем говорить в следующей части книги) в пространстве признаков.

С развитием LLM (Large Language Models) и генеративных моделей, взаимодействие с базами данных начинает меняться. Теперь можно запрашивать данные на естественном языке, получать семантический поиск по документам, автоматически извлекать ключевые термины и строить контекстные связи между объектами — всё это без необходимости владения SQL или знания структуры таблиц. Подробнее об этом говорилось в разделе «LLM и их роль в обработке данных и бизнес-процессах».

Однако важно понимать, что LLM не структурируют информацию автоматически и не приводят её в порядок. Модель лишь "плавает" по массиву данных и находит наиболее подходящий фрагмент, исходя из контекста запроса. Если данные не были предварительно очищены или преобразованы, глубокий поиск (deep search) будет напоминать попытку найти ответ в цифровом "мусоре" — работать возможно будет, но качество результатов будет ниже. Идеально — если данные можно предварительно структурировать (например, перевести документы в Markdown) и загрузить в векторную базу. Это существенно повышает точность и релевантность выдачи.

> Изначально векторные БД применялись в машинном обучении, но сегодня они находят всё более широкое применение за его пределами — в поисковых системах, персонализации контента, интеллектуальной аналитике.

Один из наиболее наглядных примеров векторного подхода в строительстве — это Bounding Box

(ограничивающий параллелепипед). Это геометрическая конструкция, описывающая границы объекта в трёхмерном пространстве. Bounding Box задаётся минимальными и максимальными координатами по осям X, Y и Z, образуя «коробку» вокруг объекта. Этот метод позволяет оценить размеры и размещение элемента без необходимости анализа всей геометрии.

Каждый Bounding Box можно представить как вектор в многомерном пространстве: например, [x, y, z, width, height, depth] — уже 6 измерений (Рис. 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

Рис. 8.2-1 Информация о координатах Bounding Box-элементов и их расположение в модели проекта является аналогией векторной базы данных.

Подобное представление данных облегчает множество задач, включая проверку на пересечения между объектами, планирование пространственного распределения элементов здания и выполнение автоматизированных расчётов. Bounding Box может служить мостом между сложными трехмерными моделями и традиционными векторными базами данных, позволяя эффективно использовать преимущества обоих подходов в архитектурном и инженерном моделировании.

> Bounding Box — это "векторизация геометрии", а эмбеддинг (способ преобразования чего-то абстрактного) — "векторизация смысла". Оба подхода позволяют перейти от ручного перебора к интеллектуальному поиску, будь то 3D-объекты в модели проекта или понятия в тексте.

Поиск объектов в проекте (например, "найти все окна шириной> 1.5 м") аналогичен поиску ближайших соседей (k-NN) в векторной БД, где критерии задают "зону" в пространстве признаков. (подробнее про поиск ближайших k-NN соседей мы поговорим в следующей части про машинное обучение) (Рис. 8.2-2). Если добавить к bounding box атрибутам дополнительные параметры (материал, вес, срок изготовления), то таблица превращается в высоко размерный вектор, где каждый атрибут — новое измерение. Это уже ближе к современным векторным базам, где измерения исчисляются сотнями или тысячами (например, эмбеддинг из нейросетей).

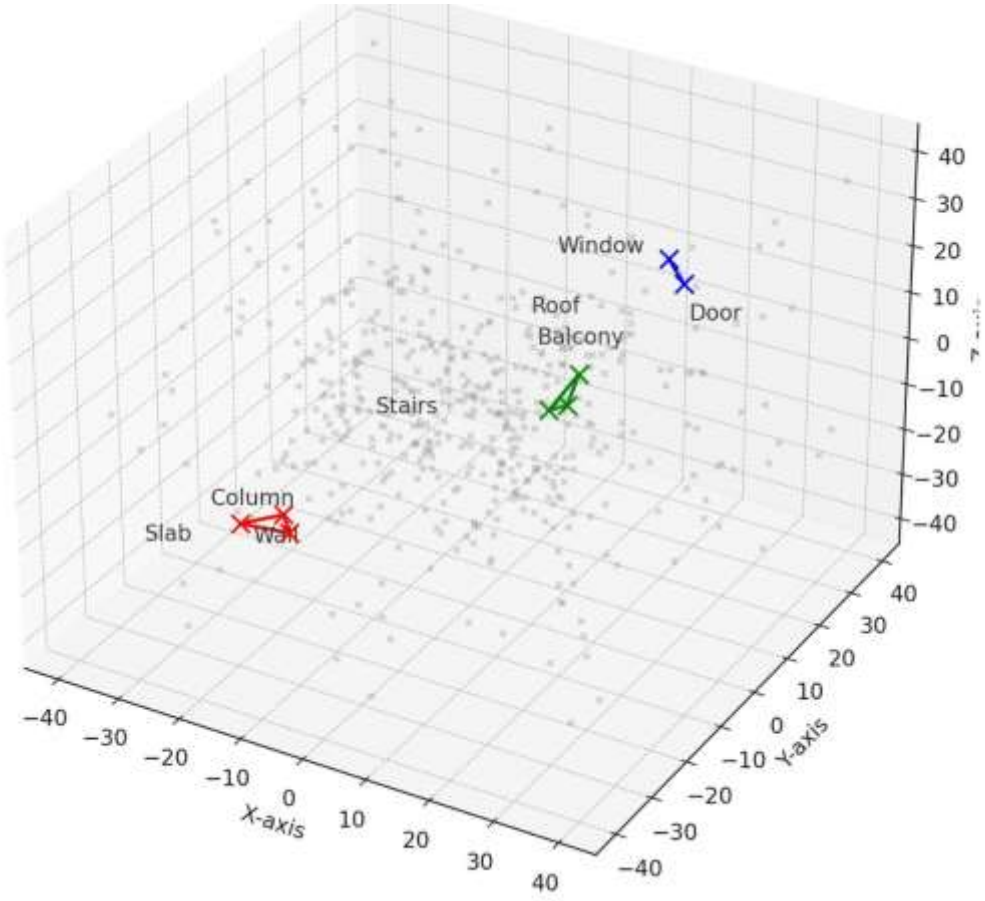


Рис. 8.2-2 Поиск объектов в проекте с использованием векторных баз данных.

Подход, используемый в Bounding Box, применим не только к геометрическим объектам, но и к анализу текстов и языка. Векторные представления данных уже активно используются в обработке естественного языка (NLP). Точно так же, как объекты в строительном проекте можно группировать по пространственной близости (Рис. 8.2-2), слова в тексте можно анализировать по их смысловой и контекстной близости.

Например, слова "архитектор", "строительство", "проектирование" в векторном пространстве будут находиться рядом, так как имеют схожее значение. В LLM этот механизм позволяет автоматически, без необходимости ручной классификации:

- Определять тематику текста
- Выполнять семантический поиск по содержимому документов
- Генерировать автоматические аннотации и резюме текста
- Находить синонимы и взаимосвязанные термины

Векторные базы данных позволяют анализировать текст и находить в нем связанные термины точно так же, как Bounding Box помогает анализировать пространственные объекты в 3D-моделях. Пример Bounding Box элементов проекта помогает понять, что векторное представление — не чисто

"искусственная" концепция из ML, а естественный способ структурирования данных для решения прикладных задач, будь то поиск колонн в CAD проекте или семантически близких изображений в базе.

> Специалистам, работающим с базами данных, стоит обратить внимание на векторные хранилища. Их распространение указывает на новый этап в развитии баз данных, где классические реляционные системы и AI-ориентированные технологии начинают переплетаться, формируя гибридные решения будущего.

Пользователи, разрабатывающие сложные и масштабные AI-приложения, будут применять специализированные базы данных для векторного поиска. В то же время те, кому нужны лишь отдельные AI-функции для интеграции в существующие приложения, скорее выберут встроенные возможности векторного поиска в уже используемых ими базах данных (PostgreSQL, Redis).

Хотя такие системы, как DWH, Data Lake, CDE, PMIS, векторные базы данных и другие, предлагают разные подходы к хранению и управлению данными, их эффективность определяется не только архитектурой, но и тем, насколько грамотно организованы и управляются сами данные. Даже при использовании современных решений — будь то векторные базы, классические реляционные СУБД или хранилища типа Data Lake — отсутствие чётких правил управления, структурирования и актуализации данных может привести к тем же трудностям, с которыми сталкиваются пользователи, работающие с разрозненными файлами и разноформатными данными.

Без продуманного управления данными (Data Governance) даже самые мощные решения могут превратиться в хаотичные и неструктурированные массивы информации, превращая озера данных в болота (Data Swamp). Чтобы избежать этого, компании должны не только выбирать подходящую архитектуру хранилищ, но и внедрять стратегии минимализма данных (Data Minimalism), управления доступом и контроля качества, позволяя превратить данные в эффективный инструмент принятия решений.

Управления данными (Data Governance), минимализм данных (Data Minimalism) и болота данных (Data Swamp)

Понимание и внедрение концепций управления данными (Data Governance), минимализма данных (Data Minimalism) и предотвращения создания болота данных (Data Swamp) - ключевые факторы успешного управления хранилищами данных и обеспечения их ценности для бизнеса (Рис. 8.2-3).

> Согласно исследованию Gartner (2017), 85% проектов в сфере больших данных терпят неудачу, и одной из ключевых причин этого является недостаточное управление качеством данных и их управлением [144].

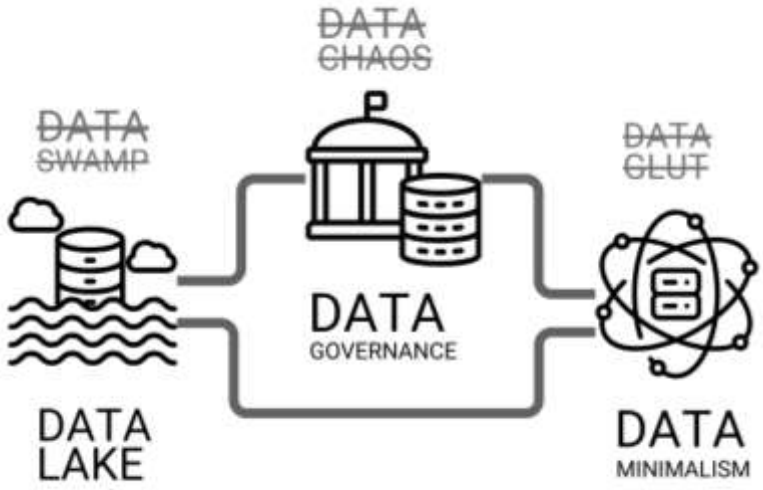


Рис. 8.2-3 Одними из ключевых аспектов управления данными являются Data Governance и Data Minimalism.

Управление данными (Data Governance) — это фундаментальный компонент управления данными, обеспечивающий надлежащее и эффективное использование данных во всех бизнес-процессах. Речь идет не только об установлении правил и процедур, но и об обеспечении доступности, надежности и безопасности данных:

- Определение и классификация данных: четкое определение и классификация сущностей позволяет организациям понять, какие сущности необходимы в компании, и определить, как их следует использовать.
- Права доступа и управление: разработка политик и процедур доступа к данным и управления ими гарантирует, что только авторизованные пользователи смогут получить доступ к определенным данным.
- Защита данных от внешних угроз: защита данных от внешних угроз - один из ключевых аспектов управления данными. Сюда входят не только технические меры, но и обучение сотрудников основам информационной безопасности.

Минимализм данных (Data Minimalism) — это подход к сокращению данных до наиболее ценных и значимых при формировании атрибутов и сущностей (Рис. 8.2-4), что позволяет сократить расходы и повысить эффективность использования данных:

- Упрощение процесса принятия решений: сокращение числа объектов и их атрибутов до наиболее значимых упрощает процесс принятия решений за счет сокращения времени и ресурсов, необходимых для анализа и обработки данных.
- Фокусировка на важном: выбор наиболее релевантных сущностей и атрибутов позволяет сосредоточиться на информации, которая действительно важна для бизнеса, устраняя шум и ненужные данные.
- Эффективное распределение ресурсов: минимизация данных позволяет более эффективно распределять ресурсы, снижая затраты на хранение и обработку данных, повышая их качество и безопасность.

Логика работы с данными должна начинаться не с их создания как такового (Рис. 8.2-4), а с понимания будущих сценариев использования этих данных ещё до начала процесса генерации. Такой подход позволяет заранее определить минимально необходимые требования к атрибутам, их типам и граничным значениям. Именно эти требования формируют основу для создания корректных и устойчивых сущностей в информационной модели. Предварительное осмысление целей и способов применения данных способствует формированию пригодной для анализа структуры. Подробнее о подходах к моделированию данных на концептуальном, логическом и физическом уровнях мы говорили в главе «Моделирование данных: концептуальная, логическая и физическая модель».

> В традиционных бизнес-процессах строительных компаний обработка данных чаще напоминает сброс данных в болото, где сначала создаются данные, а затем специалисты пытаются интегрировать их в другие системы и инструменты.

Болото данных (Data Swamp) — это результат неконтролируемого сбора и хранения данных без надлежащей организации, структурирования и управления, в результате чего данные становятся неструктурированными, сложными для использования и малоценными.

Как предотвратить превращение потока информации в болота:

- **Управление структурой данных:** обеспечение структурированности и классификации данных помогает предотвратить "болото данных", делая их упорядоченными и легкодоступными.
- **Понимание и интерпретация данных**: четкое описание происхождения данных, их модификаций и значений гарантирует, что данные будут поняты и интерпретированы правильно.
- **Поддержание качества данных**: регулярное обслуживание и очистка данных помогают поддерживать их качество, актуальность и ценность для аналитических и бизнес-процессов.

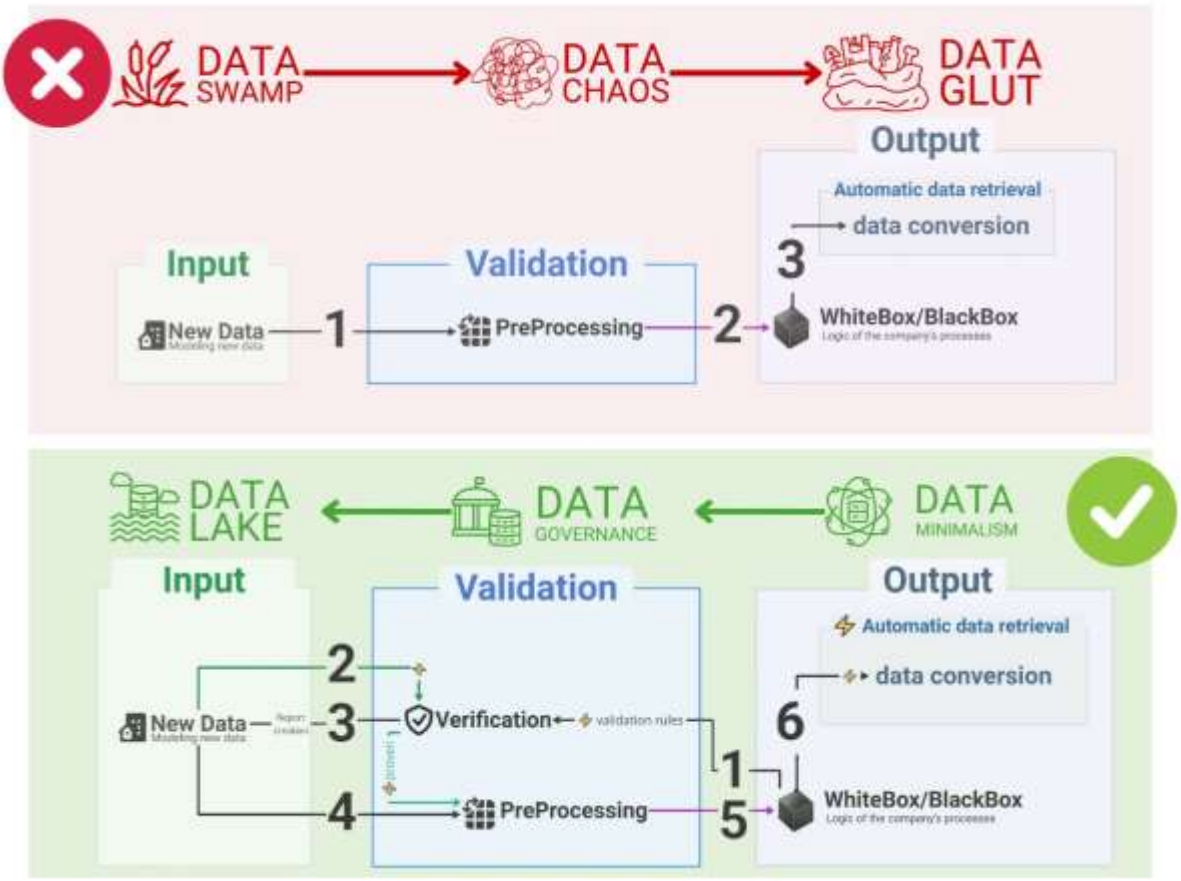


Рис. 8.2-4 Чтобы избежать беспорядка в хранилище данных, начинать процесс создания данных необходимо со сбора требований к атрибутам.

> Интегрируя принципы управления данными и минимализма данных в процессы управления данными, а также активно предотвращая превращение хранилищ данных в болото данных, организации могут максимально использовать потенциал своих данных.

Следующий этап в эволюции работы с данными после решения вопросов управления и минимализма — это стандартизация автоматической обработки, обеспечение качества и внедрение методов, которые делают данные удобными для анализа, трансформации и принятия решений. Именно этим занимаются методологии DataOps и VectorOps, которые становятся важными инструментами для компаний, работающих с большими объемами информации и машинным обучением.

DataOps и VectorOps: новые стандарты работы с данными

Если Data Governance отвечает за контроль и организацию данных, то DataOps позволяет обеспечить их точность, согласованность и бесперебойное движение внутри компании. Это особенно критично для ряда бизнес-кейсов в строительстве, где данные генерируются непрерывно и требуют своевременной обработки. Например, в ситуациях, когда информационные модели зданий, проект-

ные требования и аналитические отчёты должны быть синхронизированы между различными системами в течение одного рабочего дня, роль DataOps может стать ключевой. Он позволяет выстроить стабильные и воспроизводимые процессы обработки данных, снижая риски задержек и потери актуальности информации.

Управления на уровне Data Governance само по себе недостаточно — важно, чтобы данные не просто сохранялись, а активно использовались в ежедневных операциях. Именно здесь на первый план выходит DataOps — методология, ориентированная на автоматизацию, интеграцию и обеспечение непрерывного потока данных.

> **DataOps** фокусируется на улучшении сотрудничества, интеграции и автоматизации потоков данных в организациях. Внедрение практик DataOps способствует точности, согласованности и доступности данных, что крайне важно для приложений, ориентированных на данные.

Ключевыми инструментами в экосистеме DataOps являются Apache Airflow (Рис. 7.4-4) — для оркестрации рабочих процессов, и Apache NiFi (Рис. 7.4-5) — для маршрутизации и преобразования потоков данных. Вместе эти технологии позволяют выстраивать гибкие, надёжные и масштабируемые конвейеры данных, обеспечивая автоматическую обработку, контроль и интеграцию информации между системами (подробнее в главе «Автоматический ETL-конвейер»). При внедрении подхода DataOps в строительные процессы важно учитывать четыре фундаментальных аспекта:

1. **Люди и инструменты важнее, чем данные**: разрозненные хранилища данных могут рассматриваться как главная проблема, однако в реальности ситуация сложнее. Помимо фрагментации данных, значительную роль играет изоляция команд и разрозненность используемых ими инструментов. В строительстве с данными работают специалисты разных направлений: инженеры данных и аналитики, команды BI и визуализации, а также эксперты по управлению проектами и качеством. У каждого из них свои методы работы, поэтому важным фактором становится создание экосистемы, в которой данные свободно передаются между участниками, обеспечивая единую, согласованную версию информации.
2. **Автоматизация тестирования и выявление ошибок**: строительные данные всегда содержат погрешности, будь то неточности в моделях, ошибки в расчетах или устаревшие спецификации. Регулярное тестирование данных и устранение повторяющихся ошибок позволяет значительно повысить их качество. В рамках DataOps необходимо внедрять автоматизированные механизмы контроля и проверки, которые отслеживают корректность данных, анализируют ошибки и выявляют закономерности, а также фиксируют и устраняют системные сбои в каждом рабочем цикле. Чем выше степень автоматизации проверки, тем выше общее качество данных и тем ниже вероятность ошибок на финальных этапах.
3. **Данные должны тестироваться так же, как программный код**: большинство строительных приложений основаны на обработке данных, однако их контроль часто остается на второстепенных ролях. Если модели машинного обучения обучены на неточных данных, это приводит к неверным прогнозам и финансовым потерям. В рамках DataOps данные должны проходить такую же тщательную проверку, как и программный код: логические проверки, стресс-тесты, оценка поведения моделей при изменении входных значений. Только проверенные и надёжные данные могут использоваться в качестве основы для принятия управленческих решений.
4. **Наблюдаемость данных без ущерба для производительности**: мониторинг данных — это не

просто сбор метрик, а стратегический инструмент управления качеством. Для эффективной работы DataOps наблюдаемость должна быть встроена на всех этапах работы с данными, от проектирования до эксплуатации. При этом важно, чтобы мониторинг не замедлял систему. В строительных проектах критично не только собирать данные, но и делать это так, чтобы работа специалистов (например проектировщиков), создающих эти данные никак не нарушалась. Этот баланс позволяет контролировать качество данных без ущерба для производительности.

DataOps – это не дополнительная нагрузка для специалистов по данным, а основа их работы. Внедряя DataOps, строительные компании для управление данными могут перейти от хаотичного управления к эффективной экосистеме, где данные работают на бизнес.

В свою очередь, VectorOps представляет собой следующий этап эволюции DataOps, ориентированный на обработку, хранение и анализ многомерных векторных данных (которые обсуждались в предыдущей главе). Это особенно актуально в таких областях как цифровые двойники, нейросетевые модели и семантический поиск, которые начинают приходить в строительную отрасль. VectorOps опирается на векторные базы данных, которые позволяют эффективно хранить, индексировать и искать многомерные представления объектов.

> **VectorOps**—это следующий шаг после DataOps, ориентированный на обработку, анализ и использование векторных данных в строительстве. В отличие от DataOps, который сосредоточен на потоке, согласованности и качестве данных, VectorOps фокусируется на управлении многомерными представлениями объектов, необходимыми для машинного обучения.

В отличие от традиционных подходов, VectorOps позволяет достигать более точного описания объектов, что критически важно для цифровых двойников, генеративных систем проектирования и автоматического выявления ошибок в CAD-данных, преобразованных в векторный формат. Совместное внедрение DataOps и VectorOps формирует прочную основу для масштабируемой, автоматизированной работы с большими объёмами информации — от классических таблиц до семантически насыщенных пространственных моделей.

Дальнейшие шаги: от хаотичного хранения к структурированным хранилищам

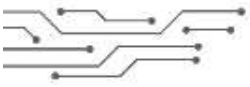
Традиционные подходы к хранению строительных данных часто приводят к созданию разрозненных "силосов информации", где важные сведения оказываются недоступны для анализа и принятия решений. Современные концепции хранения данных, такие как Data Warehouse, Data Lake и их гибриды, позволяют объединить разрозненную информацию и сделать её доступной централизировано для потоковой обработки данных и бизнес-аналитики. Важно не только выбрать подходящую архитектуру хранения, но и внедрить принципы управления данными (Data Governance) и минимализма данных (Data Minimalism), чтобы предотвратить превращение хранилищ в неконтролируемые "болота данных" (Data Swamp).

Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные концепции в ваших повседневных задачах:

- Выберите эффективные форматы хранения данных
  - Переходите от CSV и XLSX к более эффективным форматам (Apache Parquet, ORC) для хранения больших объемов данных
  - Внедрите систему версионирования данных для отслеживания изменений
  - Используйте метаданные для описания структуры и происхождения информации
- Создайте единую архитектуру данных компании
  - Сравните разные архитектуры хранения данных: RDBMS, DWH и Data Lake. Выберите ту, которая наилучшим образом отвечает вашим задачам по масштабируемости, интеграции источников и аналитической обработке
  - Спроектируйте карту процессов извлечения, загрузки и трансформации данных (ETL) из различных источников для ваших задач. Используйте инструменты визуализации, такие как Miro, Lucidchart или Draw.io, чтобы наглядно представить ключевые шаги и точки интеграции
- Внедрите практики Data Governance и Data Minimalism
  - Следуйте подходу Data Minimalism — храните и обрабатывайте только то, что действительно имеет ценность
  - Внедряйте принципы Data Governance — определяйте ответственность за данные, обеспечивайте качество и прозрачность
  - Узнайте больше про политику управления данными и концепты DataOps, VectorOps
  - Определите критерии качества данных и процедуры их проверки в рамках DataOps

Хорошо организованное хранение данных создает основу для централизации процессов аналитических процессов компании. Переход от хаотичного накопления файлов к структурированным хранилищам позволяет превратить информацию в стратегический актив, который помогает принимать обоснованные решения и повышать эффективность бизнес-процессов.

После того как процессы сбора, трансформации, анализа и структурированного хранения данных автоматизированы и стандартизированы, следующим этапом цифровой трансформации становится полноценная работа с большими данными (Big Data).



IX ЧАСТЬ
БОЛЬШИЕ ДАННЫЕ, МАШИННОЕ ОБУЧЕНИЕ И ПРОГНОЗЫ

Девятая часть посвящена большим данным, машинному обучению и прогнозированию в строительной отрасли. Здесь рассматривается переход от интуитивного принятия решений к объективному анализу, основанному на исторических данных. На практических примерах демонстрируется анализ больших данных в строительстве — от разбора набора данных разрешений на строительство в Сан-Франциско до обработки CAD-проектов с миллионами элементов. Особое внимание уделяется методам машинного обучения для прогнозирования стоимости и сроков строительных проектов, с детальным разбором алгоритмов линейной регрессии и k-ближайших соседей. Показано, как структурированные данные становятся основой для прогностических моделей, позволяющих оценивать риски, оптимизировать ресурсы и повышать эффективность управления проектами. Часть также содержит рекомендации по выбору репрезентативных выборок данных и объясняет, почему для эффективного анализа не всегда требуются огромные массивы информации.

ГЛАВА 9.1.
БОЛЬШИЕ ДАННЫЕ И ИХ АНАЛИЗ

Большие данные в строительстве: от интуиции к прогнозируемости

Термин "большие данные" не имеет строгого определения. Изначально эта концепция появилась, когда объемы информации стали превышать возможности традиционных методов её обработки. Сегодня объёмы и сложность данных во многих отраслях, включая строительство, настолько увеличились, что не помещаются в локальную память компьютеров и требуют использования новых технологий для их обработки.

Суть работы с большими данными заключается не только в хранении и обработке, но и в возможности прогнозирования. В строительной отрасли Big Data открывают путь от интуитивных решений, основанных на субъективной интерпретации таблиц и визуализаций (что обсуждалось ранее), к обоснованным прогнозам, подкреплённым реальными наблюдениями и статистикой.

Вопреки распространённому мнению, цель работы с большими данными — не в том, чтобы "заставить машину мыслить как человек", а в применении математических моделей и алгоритмов для анализа массивов данных с целью выявления закономерностей, предсказания событий и оптимизации процессов.

> **Большие данные (Big Data)** — это не холодный мир алгоритмов, лишенный человеческого влияния. Напротив, большие данные работают в связке с нашими инстинктами, ошибками и творчеством. Именно несовершенство человеческого мышления позволяет находить нестандартные решения и делать прорывы.

С развитием цифровых технологий в строительстве начали активно применяться методы обработки данных, пришедшие из IT-сферы. Благодаря таким инструментам, как Pandas и Apache Parquet, структурированные и неструктурированные данные могут быть объединены, упрощая доступ к информации и снижая потери на анализ, а большие наборы данных из документов или проектов CAD (Рис. 9.2-10 - Рис. 9.2-12) позволяют собирать, анализировать и прогнозировать данные на всех этапах жизненного цикла проекта.

Большие данные оказывают трансформирующее воздействие на строительную отрасль, влияя на неё потенциально в самых разных аспектах. Применение технологий Big Data приносит результаты в ряде ключевых направлений, включая например следующие:

- **Анализ инвестиционного потенциала** — прогнозирование доходности и сроков окупаемости проектов на основе данных предыдущих объектов.
- **Предиктивное обслуживание** — выявление вероятных отказов оборудования до их фактического наступления, что снижает время простоя.
- **Оптимизация цепочек поставок** — прогнозирование сбоев и повышение эффективности логистики.

- **Анализ энергоэффективности** — помощь в проектировании зданий с низким уровнем энергопотребления.
- **Мониторинг безопасности** — использование датчиков и носимых устройств для отслеживания условий на стройплощадке.
- **Контроль качества** — наблюдение за соответствием технологическим нормам в режиме реального времени.
- **Управление трудовыми ресурсами** — анализ производительности и прогнозирование кадровых потребностей.

Трудно найти область в строительстве, где аналитика данных и предсказания были бы не востребованы. Главное преимущество алгоритмов прогнозирования — их способность к самообучению и постоянному совершенствованию по мере накопления данных.

> В ближайшем будущем искусственный интеллект будет не просто помогать строителям, а принимать ключевые решения — от процессов проектирования до вопросов эксплуатации зданий.

Подробнее о том, как формируются прогнозы и используются обучающиеся модели, будет рассказано в следующей части книги: «Машинное обучение и прогнозы».

Для перехода к полноценной работе с большими данными требуется изменение самого подхода к аналитике. Если в классических системах, которые мы рассматривали до этого, основное внимание уделялось причинно-следственным связям, то в аналитике больших данных акцент смещается в сторону поиска статистических закономерностей и корреляций, позволяющих выявлять скрытые взаимосвязи и предсказывать поведение объектов даже без полного понимания всех факторов.

Вопрос о целесообразности больших данных: корреляция, статистика и выборка данных

Традиционно строительство опиралось на субъективные гипотезы и личный опыт. Инженеры предполагали — с определённой долей вероятности — как поведёт себя материал, какие нагрузки выдержит конструкция и сколько продлится проект. Эти предположения проверялись на практике, зачастую ценой времени, ресурсов и будущих рисков.

> С появлением больших данных подход кардинально меняется: решения теперь принимаются не на основе интуитивных догадок, а в результате анализа масштабных информационных массивов. Строительство постепенно перестаёт быть искусством интуиции и превращается в точную науку предсказаний.

Переход к идее использования больших данных неизбежно поднимает важный вопрос: насколько критичен объём данных и сколько информации действительно необходимо для надёжной прогнозной аналитики? Распространённое мнение о том, что «чем больше данных — тем выше точность», на практике оказывается не всегда обоснованным с точки зрения статистики.

> Еще в 1934 году статистик Ежи Нейман доказал [145], что ключ к точности статистических выводов заключается не столько в объеме данных, сколько в их представительности и случайности выборки.

Это особенно актуально в строительной отрасли, где большие массы данных собираются с помощью IoT-сенсоров, сканеров, камер наблюдения, беспилотников и даже разноформатных CAD-моделей, что повышает риск появления «слепых зон», выбросов и искажений в данных.

Рассмотрим пример мониторинга состояния дорожного покрытия. Полный набор данных обо всех дорожных участках может занимать X Гб и требовать около суток на обработку. В то же время случайная выборка, включающая только каждый 50-й участок, займет лишь X/50 Гб и обработается за пол часа, при этом для определённых расчётов обеспечивая аналогичную точность оценок (Рис. 9.1-1).

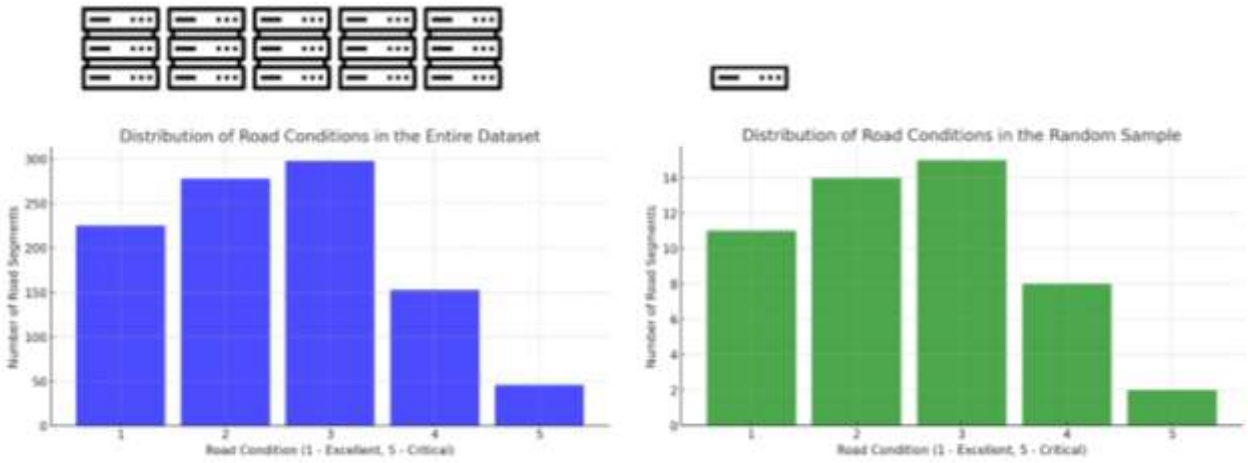


Рис. 9.1-1 Гистограммы состояния дорожного покрытия: полный набор данных и случайная выборка показывают идентичные результаты.

Таким образом, ключевым фактором успешного анализа данных может часто являться не их объём, а репрезентативность выборки и качество применяемых методов обработки. Переход к случайной выборке и более избирательному подходу требует смены мышления в строительной индустрии. Исторически компании придерживались логики: «чем больше данных — тем лучше», полагая, что охват всех возможных показателей обеспечит максимальную точность.

Этот подход напоминает популярное заблуждение из управления проектами: «чем больше специалистов я привлеку, тем эффективнее будет работа». Однако, как и в случае с кадрами, именно качество и инструментарий важнее количества. Без учёта взаимосвязей (корреляций) между данными или участниками проекта, рост объёма может привести лишь к шуму, искажениям, дублированию и неоправданным выбросам.

В итоге часто оказывается, что гораздо продуктивнее иметь меньший, но качественно подготовленный набор данных, способный давать стабильные и обоснованные прогнозы, чем опираться на массивную, но хаотичную информацию, содержащую множество противоречивых сигналов.

> Чрезмерный объём данных не только не гарантирует большей точности, но и может привести к искажённым выводам — из-за присутствия шумов, избыточных признаков, скрытых корреляций и нерелевантной информации. В таких условиях возрастает риск переобучения моделей и снижается надёжность аналитических результатов.

В строительстве главным вызовом в работе с большими данными является определение оптимального количества и качества данных. Например, при мониторинге состояния бетонных конструкций использование тысяч датчиков и сбор информации каждую минуту может перегрузить систему хранения и анализа. Однако, если провести анализ корреляции и выбрать 10% наиболее информативных датчиков, можно получить практически идентичную точность прогнозов, затрачивая в разы, иногда в десятки и сотни раз, меньше ресурсов.

Использование меньшего подмножества данных сокращает как необходимый объем хранения, так и время обработки, что значительно уменьшает затраты на хранение и анализ данных и часто делает случайную выборку идеальным решением для прогнозной аналитики, особенно в условиях больших инфраструктурных проектов или при работе в режиме реального времени. В конечном итоге, эффективность строительных процессов определяется не объемом собранных данных, а качеством их анализа. Без критического подхода и тщательного анализа данные могут привести к неверным выводам.

> После определенного объема данных каждая новая единица информации дает все меньше полезного результата. Вместо бесконечного сбора информации важно сосредоточиться на её репрезентативности и методах анализа (Рис. 9.2-2).

Этот феномен хорошо описан Алленом Уоллисом [146], который иллюстрирует использование статистических методов на примере тестирования двух альтернативных конструкций снарядов ВМС США.

> *ВМС тестировали две альтернативные конструкции снарядов (A и B), проводя серию парных выстрелов. В каждом раунде A получает 1 или 0 в зависимости от того, лучше или хуже его характеристики, чем у B, и наоборот. Стандартный статистический подход предполагает проведение фиксированного числа испытаний (например, 1000) и определение победителя на основе процентного распределения (например, если A получает 1 более чем в 53% случаев, его считают лучшим). Когда Аллен Уоллис обсуждал такую проблему с капитаном (военно-морского флота) Гарретом Л. Шайлером, капитан возразил, что такой тест, цитируя рассказ Аллена, может оказаться бесполезным. Если бы мудрый и опытный офицер по артиллерийскому вооружению, такой как Шайлер, был на месте, он бы увидел после первых нескольких сотен [выстрелов], что эксперимент не нужно завершать либо потому, что новый метод явно хуже, либо потому, что он явно превосходит то, на что надеялись [146].*
>
> — правительственная группа статистических исследований США в Колумбийском университете, период второй мировой войны

Этот принцип широко используется в различных отраслях. В медицине, например, клинические исследования новых препаратов проводятся на случайных выборках пациентов, что позволяет получать статистически значимые результаты без необходимости тестировать лекарство на всей популяции людей, живущих на планете. В экономике и социологии проводятся репрезентативные опросы, которые отражают мнение общества без необходимости опрашивать каждого жителя страны.

Подобно тому, как государства и исследовательские организации проводят опросы небольших групп населения для понимания общих социальных тенденций, компании в строительной отрасли могут использовать случайные выборки данных для эффективного мониторинга и создания прогнозов для управления проектами (Рис. 9.1-1).

> *Большие данные могут изменить подход к социальным наукам, но не заменят статистический здравый смысл [147].*
>
> – Томас Ландсалл-Велфэр, «Прогнозирование настроения нации в настоящее время», Significance v. 9(4), 2012 г.

С точки зрения экономии ресурсов, при сборе данных для будущих предсказания и принятий решений важно отвечать вопрос: имеет ли смысл тратить значительные средства на сбор и обработку огромных массивов данных, когда можно использовать значительно меньший и более дешёвый тестовый набор данных, который можно постепенно масштабировать? Эффективность случайных выборок показывает, что компании могут сократить затраты в десятки или даже тысячи раз на сбор и обучение моделей, выбирая методы сбора данных, которые не требуют всеобъемлющего охвата, но при этом обеспечивают достаточную точность и представительность. Такой подход позволяет даже небольшим компаниям достигать результатов на уровне крупных корпораций, используя значительно меньшие ресурсы и объемы данных, что важно для компаний, которые стремятся оптимизировать расходы и ускорить процесс принятия обоснованных решений, используя небольшие по наличию ресурсу. В следующих главах рассмотрим примеры аналитики и прогнозирования на основе публичных наборов данных с использованием инструментов больших данных.

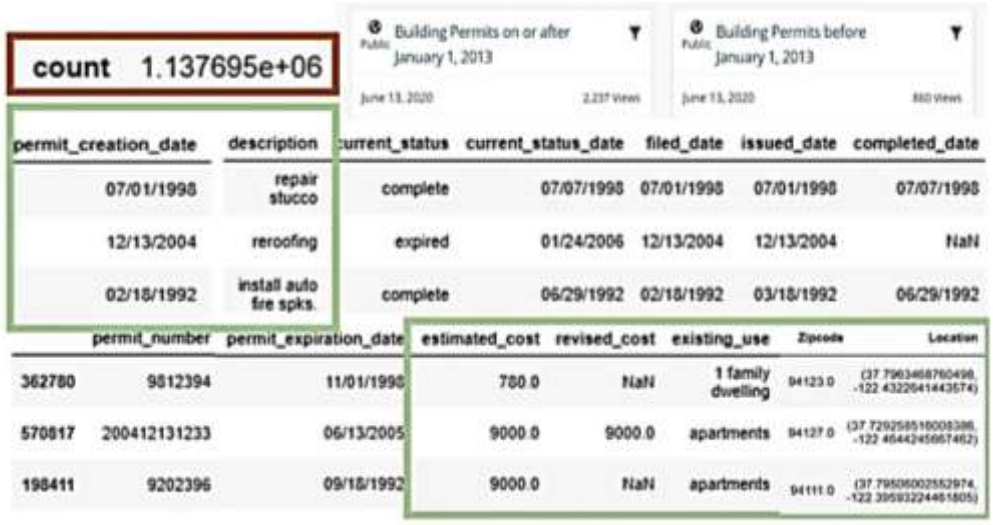
Большие данные: анализ данных датасета миллиона разрешений на строительство Сан-Франциско

Работа с открытыми наборами данных предоставляет уникальную возможность применять на практике принципы, рассмотренные в предыдущих главах: разумный отбор признаков, репрезентативная выборка, визуализация и критический анализ. В этой главе мы рассмотрим, как можно исследовать сложные явления, такие как строительная активность в большом городе, используя открытые данные — в частности, свыше миллиона записей о разрешениях на строительство в Сан-Франциско

Общедоступные данные о более, чем миллионе разрешений (Рис. 9.1-2) на строительство (записи в двух наборах данных в формате CSV) из "Департамента зданий Сан-Франциско" [148] позволяют нам использовать сырую CSV-таблицу для анализа не только строительной активности в городе,

но и для критического анализа последних тенденций и истории строительной отрасли Сан-Франциско за последние 40 лет, с 1980 по 2019 год.

Примеры кода, использованные для создания визуализаций датасета (Рис. 9.1-3- Рис. 9.1-8), а также визуальные графики с кодом, пояснениями и комментариями можно найти на платформе Kaggle по запросу "Сан-Франциско. Строительный сектор 1980–2019 гг." [149].



count 1.137695e+06

Building Permits on or after January 1, 2013

Building Permits before January 1, 2013

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	(37.7963468760498, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.7292585160008386, -122.464424569574 62)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.7950600255297 4, -122.3959322446180 5)

Рис. 9.1-2 Наборы данных содержат информацию о выданных разрешениях на строительство с различными атрибутами объектов.

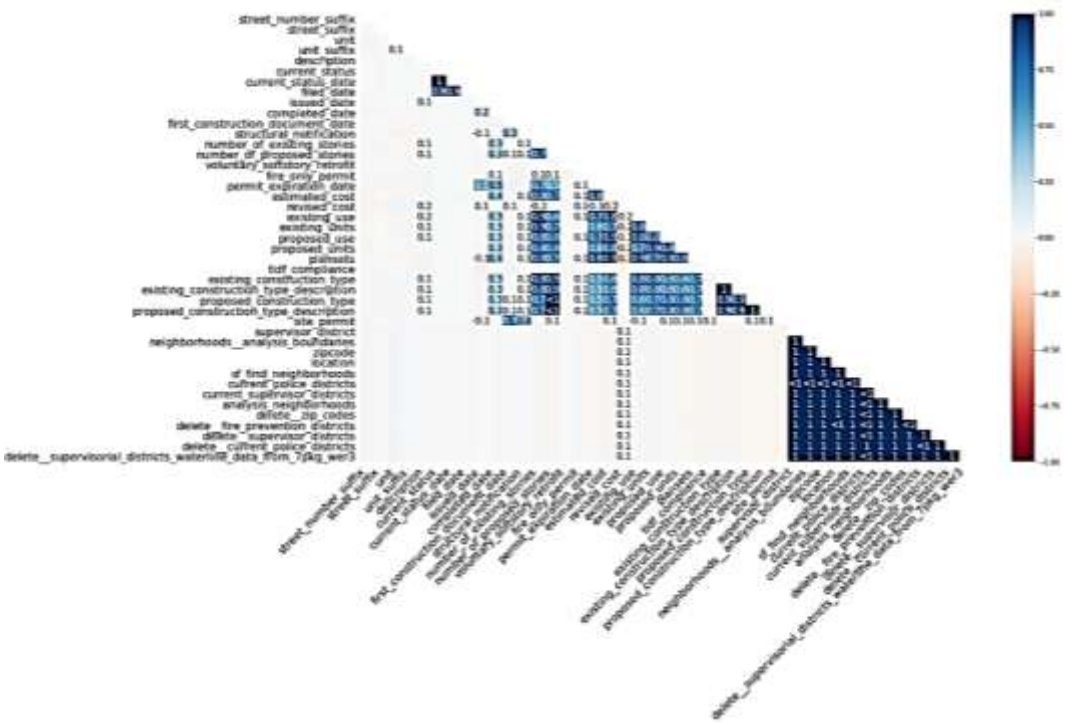


Рис. 9.1-3 Тепловая карта (Pandas и Seaborn), визуализирующая все атрибуты датасета и помогающая выявить взаимосвязи между парами атрибутов.

Из таблицы, предоставленной департаментом зданий Сан-Франциско (Рис. 9.1-2), не видно никаких тенденций или выводов. Сухие цифры в табличном виде не являются основой для принятия решений. Чтобы сделать данные визуально понятными, что подробно обсуждалось в главах о визуализации данных, они должны быть визуализированы с помощью различных библиотек, рассмотренных в седьмой части книги по теме “ETL и визуализация результатов в виде графиков".

Проанализировав данные, с помощью Pandas DataFrame и библиотек для визуализации Python, о стоимости 1 137 695 разрешений [148], можно сделать вывод, что строительная активность в Сан-Франциско тесно связана с экономическими циклами, особенно в бурно развивающейся технологической отрасли Кремниевой долины (Рис. 9.1-4).

Экономические бумы и спады оказывают значительное влияние на количество и стоимость строительных проектов. Например, первый пик строительной активности совпал с бумом электроники в середине 1980-х годов (использовались Pandas и Matplotlib), а последующие пики и спады были связаны с пузырем доткомов и технологическим бумом последних лет.

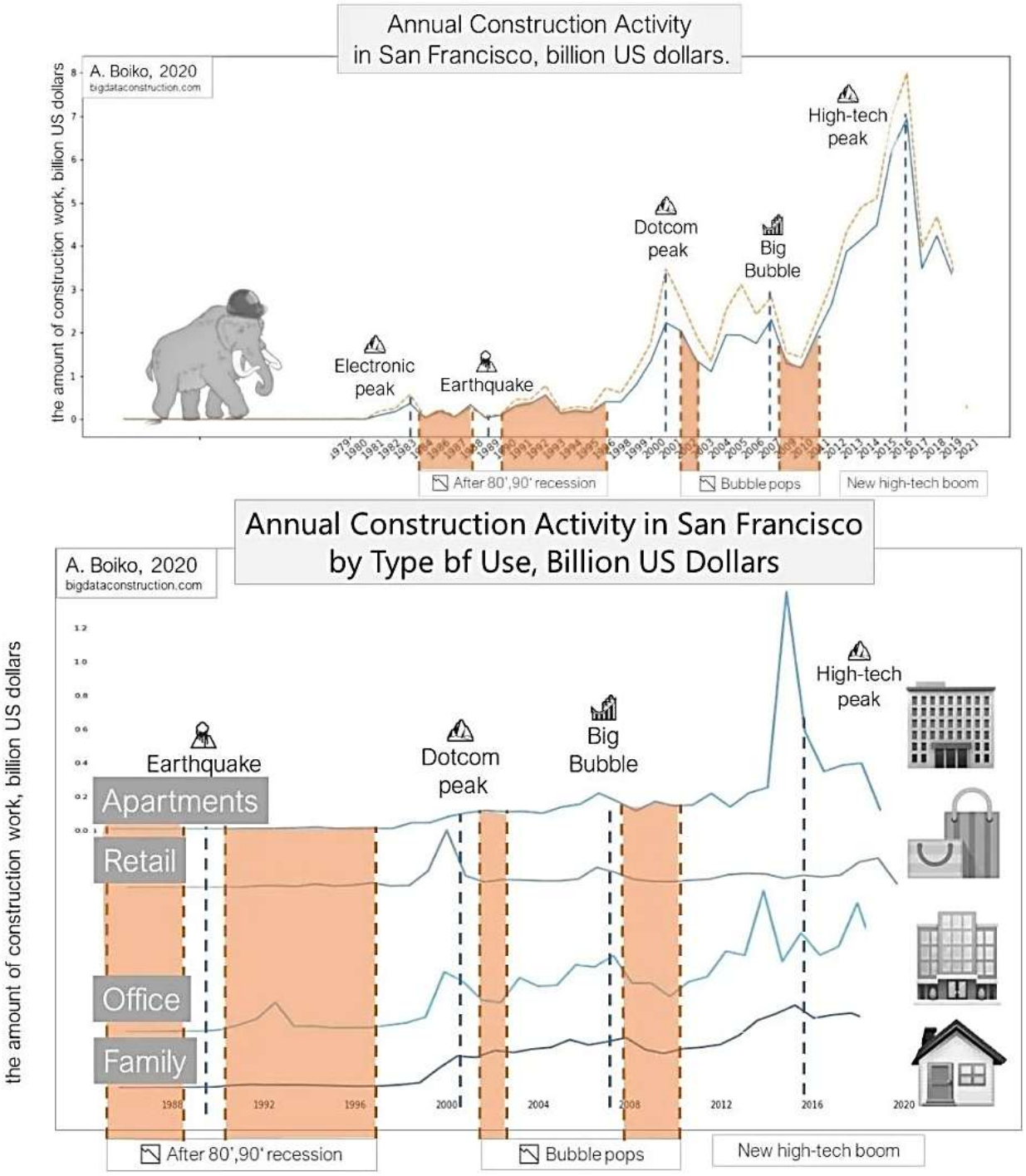


Рис. 9.1-4 В сфере недвижимости Сан-Франциско инвестиции коррелируют с технологическим развитием Кремниевой долины.

Аналитика данных позволяет оценить, что в Сан-Франциско большая часть из 91,5 миллиарда долларов, вложенных в строительство и реконструкцию за последнее десятилетие, - почти 75 % - сосредоточена в центре города (Рис. 9.1-5 - использовались Pandas и библиотека визуализации Folium) и в радиусе 2 км от центра города, что отражает более высокую плотность инвестиций в этих центральных зонах.

Средняя стоимость разрешений на строительство значительно варьируется в зависимости от района, при этом заявки в центре города стоят в три раза дороже, чем за его пределами, из-за более высокой стоимости земли, рабочей силы, материалов и строгих строительных норм, требующих применения более дорогих материалов для повышения энергоэффективности.

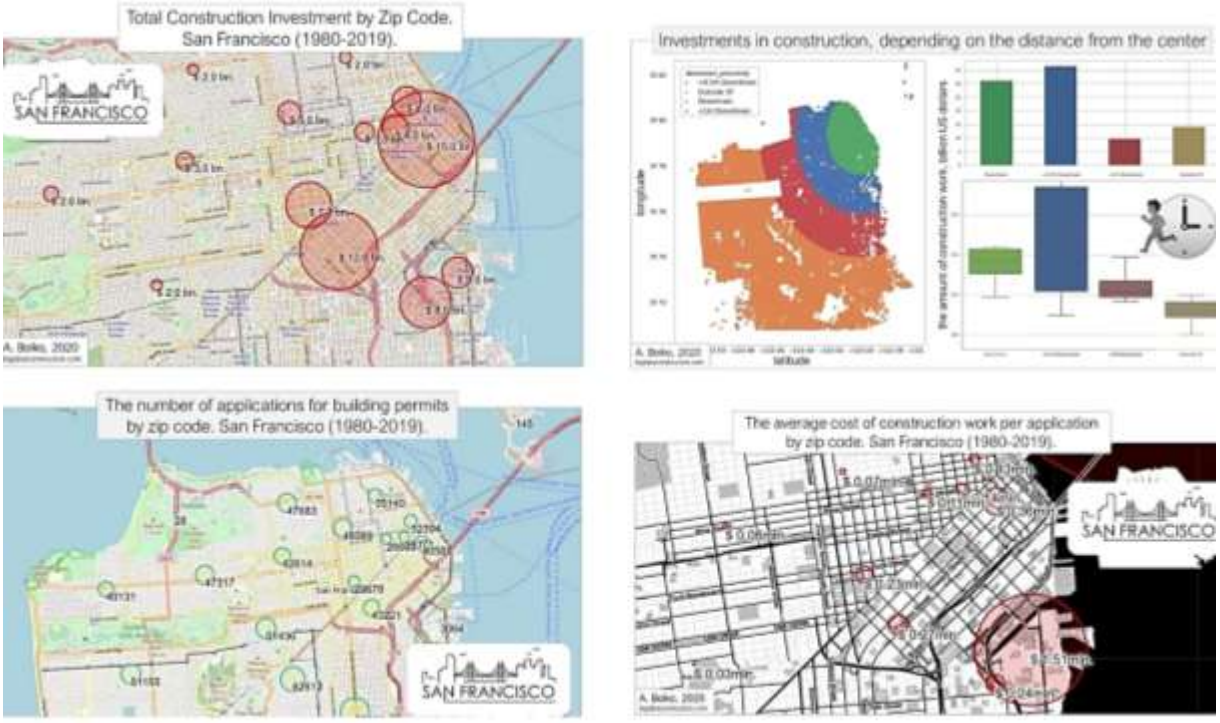


Рис. 9.1-5 В Сан-Франциско 75 процентов инвестиций в строительство (91,5 миллиарда долларов) сосредоточено в центре города.

Набор данных также позволяет рассчитать средние цены на ремонт не только по типам домов, но и по районам города и отдельным адресам (почтовым индексам). В Сан-Франциско динамика стоимости ремонта жилья демонстрирует отчетливые тенденции для разных типов ремонта и жилья (Рис. 9.1-6 - использовались Pandas и Matplotlib). Ремонт кухни заметно дороже, чем ремонт ванной комнаты: средний ремонт кухни в односемейном доме стоит около 28 000 долларов по сравнению с 25 000 долларов в двухсемейном доме.

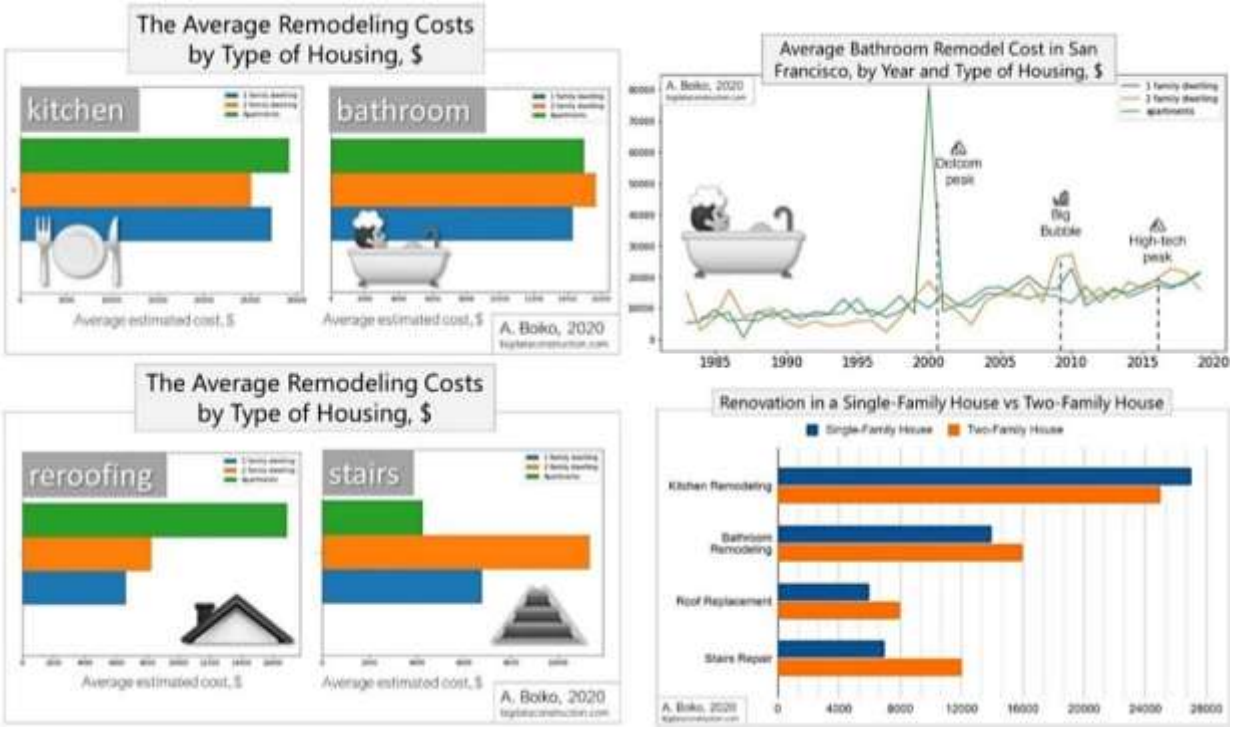


Рис. 9.1-6 В SF ремонт кухонь обходится почти в два раза дороже ремонта ванных комнат и домовладельцам необходимо откладывать по $350 ежемесячно в течение 15 лет, чтобы покрыть расходы на основной ремонт жилья.

Инфляцию стоимости строительства в Сан-Франциско с годами можно проследить, проанализировав данные, сгруппированные по типам жилья и годам (Рис. 9.1-7 - использовались Pandas и Seaborn), которые показывают постоянный рост средней стоимости ремонта с 1990 года и выявляют краткосрочные трехлетние циклы в стоимости ремонта многоквартирных домов.

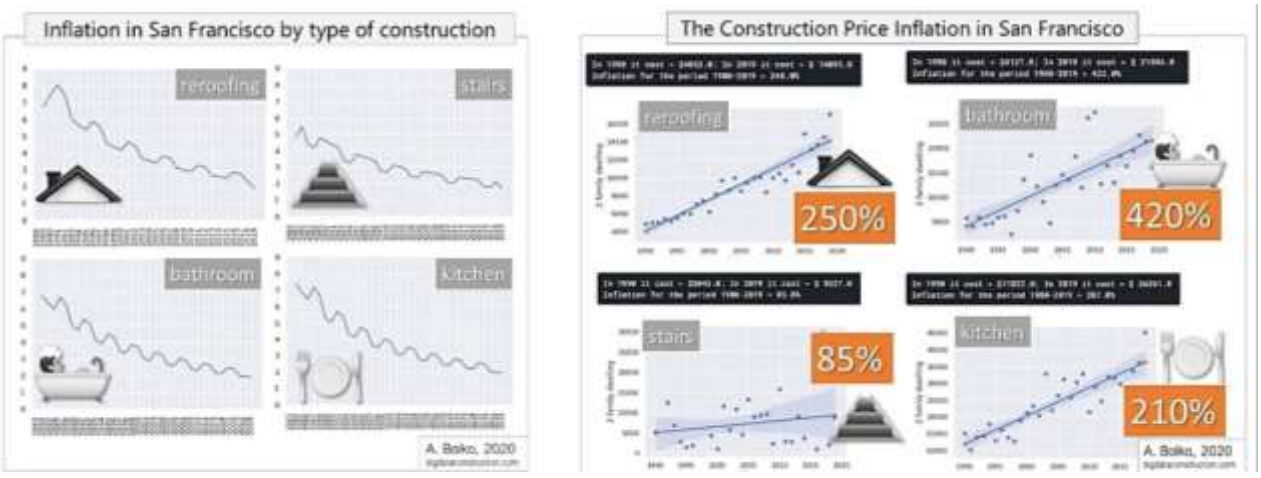


Рис. 9.1-7 С 1980 по 2019 год стоимость ремонта ванных комнат в СФ выросла в пять раз, в то время как ремонт крыши и кухни подорожал в три раза, а ремонт лестниц - только на 85%.

Исследование открытых данных от строительного департамента Сан-Франциско (Рис. 9.1-3) позволяет выявить, что стоимость строительства в городе чрезвычайно переменчива и часто непредсказуема, оказываясь под влиянием множества факторов. Среди этих факторов - экономический рост, технологические инновации и уникальные требования различных видов жилья.

В прошлом для проведения подобного анализа требовались глубокие знания в программировании и аналитике. Однако с появлением LLM-инструментов, процесс стал доступен и понятен для широкого круга специалистов в строительной отрасли, начиная от инженеров в отделах проектирования и заканчивая высшим руководством компаний.

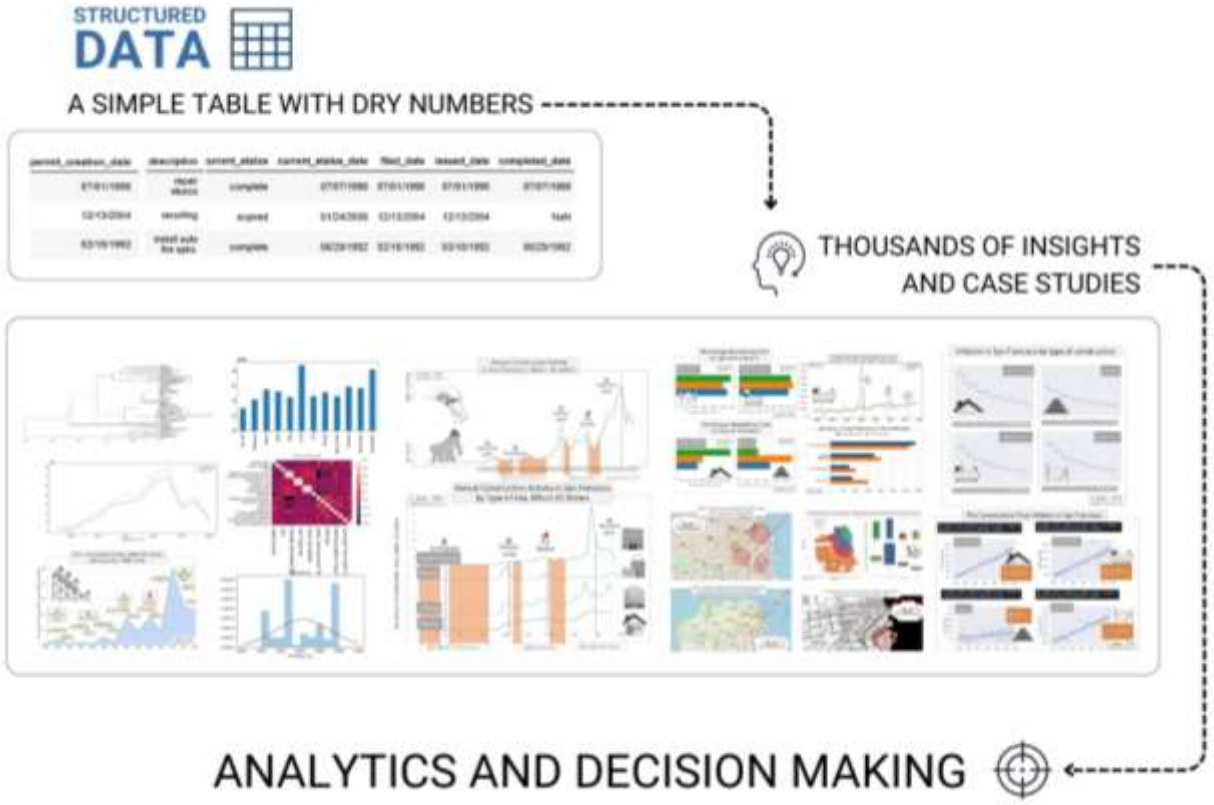


Рис. 9.1-8 Переход к визуально понятным данным позволяет автоматизировать принятие решений за счёт распознавания скрытых паттернов.

Подобно тому, как мы анализировали данные из табличного набора данных "строительного управления Сан-Франциско", мы можем визуализировать и анализировать любые наборы данных - от изображений и документов до данных IoT, или данных из полученных из баз данных CAD.

Пример больших данных на основе данных CAD (BIM)

В следующем примере проанализируем большой набор данных, используя данные из разных инструментов CAD (BIM). Для сбора и создания большого набора данных был использован специализированный автоматизированный веб-краулер (скрипт), настроенный на автоматический поиск и сбор проектных файлов с сайтов, предлагающих бесплатные архитектурные модели в форматах

RVT и IFC. За несколько дней краулер успешно нашел и скачал 4 596 файлов IFC и 6 471 файл RVT и 156 024 файлов DWG [149].

После сбора проектов в форматах RVT и IFC разных версий и их конверсии в структурированный формат CSV с помощью бесплатных SDK для обратного инжиниринга почти 10 тысяч проектов RVT и IFC были собраны в один большой файл-таблицу Apache Parquet и загружен для анализа в Pandas DataFrame (Рис. 9.1-9).

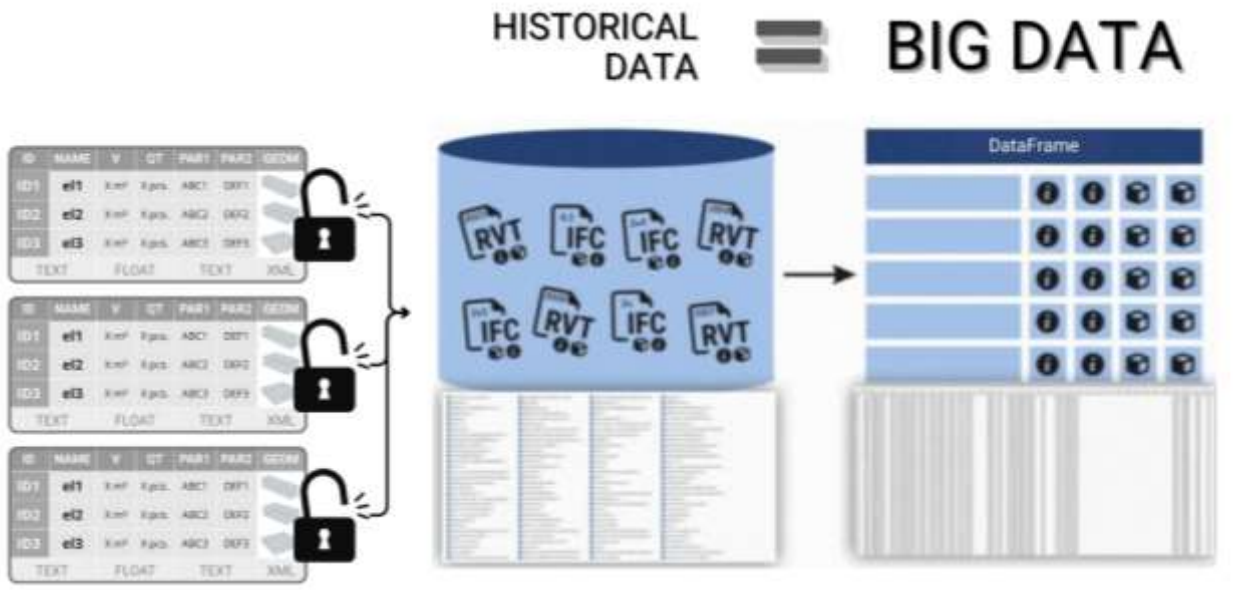


Рис. 9.1-9 Структурированные данные проекта позволяют объединить любое количество проектов в одну двухмерную таблицу.

Данные из этой масштабной коллекции содержат следующие сведения: набор файлов IFC содержит около 4 миллионов сущностей (строк) и 24 962 атрибута (столбцов), а набор файлов RVT, состоящий примерно из 6 миллионов сущностей (строк), содержит 27 025 различных атрибутов (столбцов).

Эти информационные наборы (Рис. 9.1-10) охватывают миллионы элементов, для каждого из которых были дополнительно получены и добавлены в общую таблицу – координаты геометрии Bounding Box (прямоугольник, определяющий границы объекта в проекте) и созданы изображения каждого элемента в формате PNG и геометрии в открытом формате XML - DAE (Collada).

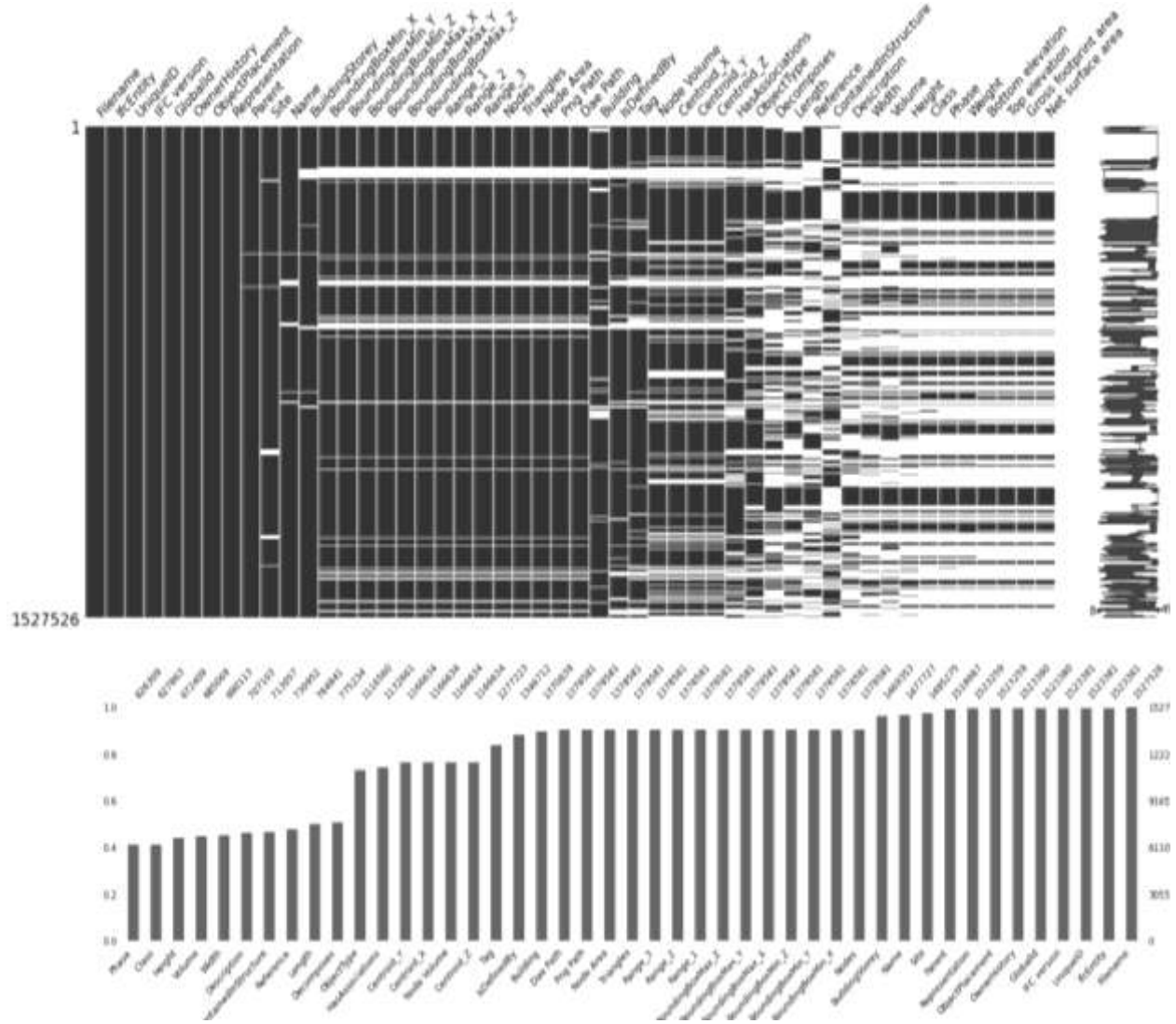


Рис. 9.1-10 Сабсет из 1,5 млн элементов и визуализация (библиотека missingno) заполненности первых 100 атрибутов в виде гистограммы.

Таким образом, мы получили всю информацию о десятках миллионов элементов из 4 596 проектов IFC и 6 471 проекта RVT, где все атрибуты-свойства всех элементов-сущностей и их геометрия (Bounding Box) были переведены в структурированную форму одной таблицы (DataFrame) (Рис. 9.1-10 - данные о наполненности датафрейма выглядят в виде гистограмм).

> Гистограммы (Рис. 9.1-10, Рис. 9.2-6, Рис. 9.2-7) построенные в процессе анализа, позволяют быстро оценить плотность данных и частоту встречаемости значений в столбцах. Это даёт первое представление о распределении признаков, наличии пропусков и потенциальной полезности отдельных атрибутов при анализе и построении моделей машинного обучения.

Одним из примеров практического использования данного массива данных (Рис. 9.1-10) является проект "5000 проектов IFC и RVT" [149], доступный на платформе Kaggle. В нём представлен Jupyter

Notebook с полным Pipeline-решением: от предобработки и анализа данных до визуализации результатов с использованием библиотек Python — pandas, matplotlib, seaborn, folium и других (Рис. 9.1-11).

Рис. 9.1-11 Примеры анализа данных из форматов CAD (BIM) при помощи библиотек визуализаций Python и библиотеки pandas.

На основе метаинформации можно определить, в каких городах были разработаны те или иные проекты, и отобразить это на карте (например, при помощи библиотеки folium). Помимо этого, временные метки в данных позволяют исследовать закономерности по времени сохранения или редактирования файлов: по дням недели, времени суток и месяцам.

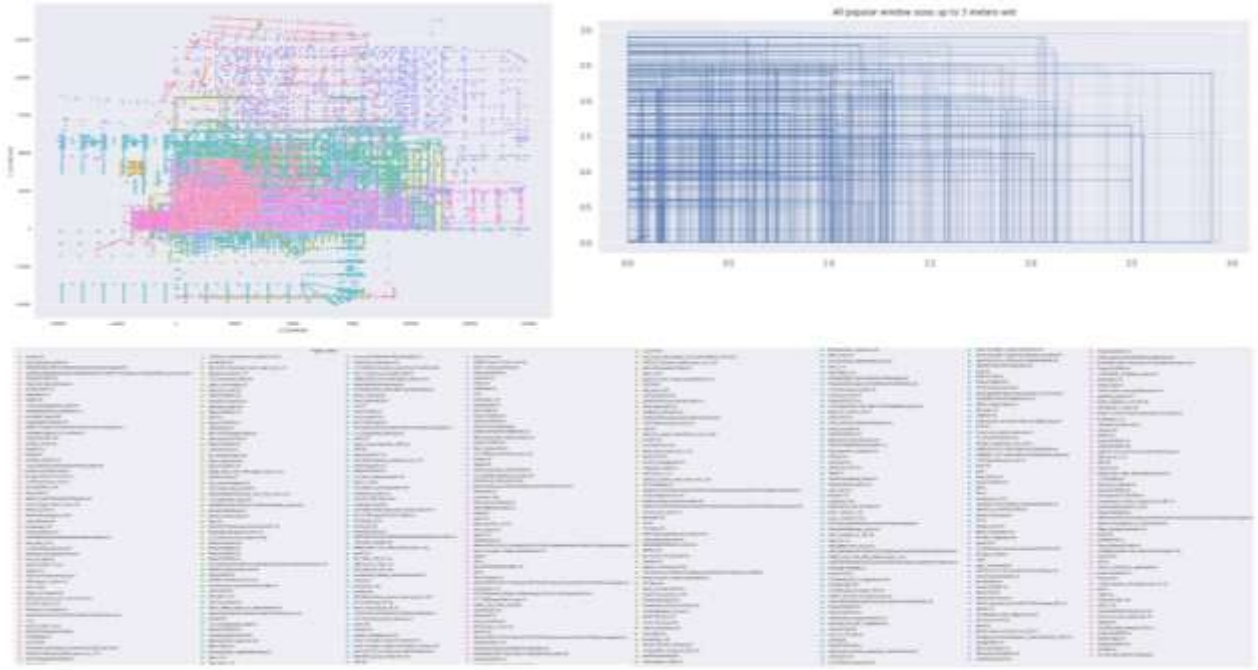


Рис. 9.1-12 Визуализация геометрического положения всех колонн и размеров всех окон до 3 метров в проектах из списка в нижней части графика.

Геометрические параметры в виде Bounding Box, извлечённые из моделей, также поддаются агрегированному анализу. Например, на Рис. 9.1-12 представлены два графика: левый показывает распределение расстояний между колоннами по всем проектам относительно нулевой точки, а правый — размеры всех окон высотой до 3 метров в выборке из десятков тысяч оконных элементов (после группировки всего датасета по параметру „Category“ со значением „OST_Windows“, „IfcWindows“).

Аналитический код Pipeline для этого примера и сам набор данных доступны на сайте Kaggle под названием "5000 проектов IFC и RVT | DataDrivenCo-nstruction.io" [149]. Этот готовый Pipeline вместе с набором данных можно скопировать и запустить бесплатно онлайн бесплатно на Kaggle или оффлайн в одной из популярных IDE: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с плагином PyDev, Thonny, Wing IDE, IntelliJ IDEA с плагином Python, JupyterLab или популярных онлайн-инструментах Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Аналитические данные, полученные в результате обработки и изучения огромных массивов структурированных данных, будут играть решающую роль в процессах принятия решений в строительной отрасли.

Благодаря подобному анализу информации на основе данных прошлых проектов специалисты могут эффективно прогнозировать например потребности в материалах и рабочей силе и оптимизировать проектные решения еще до начала строительства.

Однако, если проектные данные или разрешения на строительство – это относительно статичная информация, изменяющаяся относительно медленно, то сам процесс строительства стремительно насыщается разнообразными датчиками и IoT-устройствами: камеры, автоматизированные системы мониторинга, которые передают данные в режиме реального времени – всё это превращает стройплощадку в динамическую цифровую среду, где данные необходимо анализировать режиме реального времени.

IoT Интернет вещей и смарт-контракты

IoT интернет вещей представляет собой новую волну цифровой трансформации, в рамках которой каждое устройство получает собственный IP-адрес и становится частью глобальной сети. IoT — это концепция, предполагающая подключение физических объектов к интернету для сбора, обработки и передачи данных. В строительстве это означает возможность контролировать строительные процессы в режиме реального времени, минимизировать потери материалов, прогнозировать износ оборудования и автоматизировать принятие решений.

> Согласно статье CFMA „Подготовка к будущему с помощью подключенного строительства" [150], в ближайшее десятилетие строительная отрасль пройдет через масштабную цифровую трансформацию, кульминацией которой станет концепция Connected Construction — полностью интегрированная и автоматизированная строительная площадка.

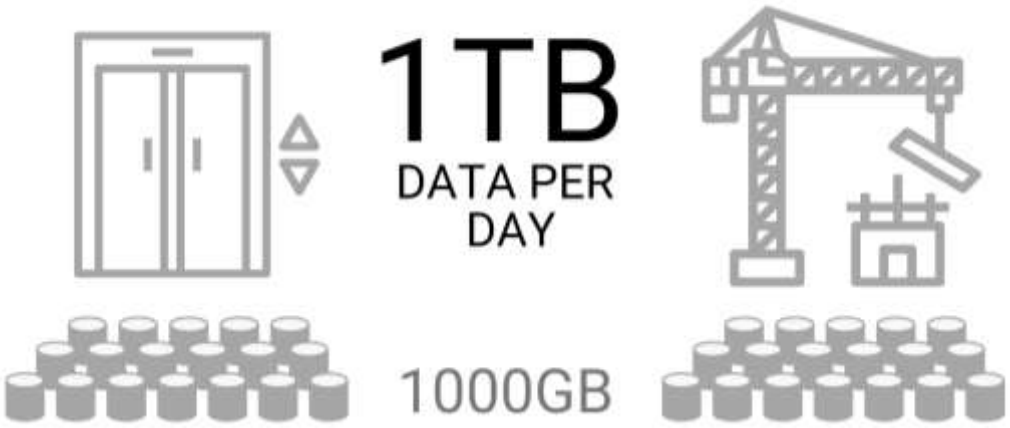


Рис. 9.1-13 IOT-устройства или устройства передачи данных на строительной площадке могут производить и передавать терабайты данных в день.

Цифровая строительная площадка предполагает, что все элементы строительства — от планирования и логистики до выполнения работ и контроля качества на строительной площадке при помощи стационарных камер и квадрокоптеров — будут объединены в единую динамичную цифровую экосистему. Ранее, в седьмой части книги, мы уже рассматривали возможности бесплатного и открытого инструмента Apache NiFi (Рис. 7.4-5), который позволяет организовать потоковую обработку данных в реальном времени — от сбора с различных источников до передачи в хранилища или аналитические платформы.

Данные о ходе строительства, расходе материалов, состоянии оборудования и безопасности будут в реальном времени передаваться в аналитические системы (Рис. 9.1-13). Это позволяет прогнозировать потенциальные риски, оперативно реагировать на отклонения и оптимизировать процессы на площадке. Ключевые компоненты цифровой строительной площадки включают:

- IoT-сенсоры — отслеживание параметров окружающей среды, мониторинг строительной техники и контроль условий труда.
- Цифровые двойники — виртуальные модели зданий и инфраструктуры, позволяющие прогнозировать возможные отклонения и предотвращать ошибки.
- Автоматизированные логистические системы — управление цепочками поставок в режиме реального времени для сокращения простоев и издержек.
- Роботизированные строительные комплексы — использование автономных машин для выполнения рутинных и опасных задач.

Роботизация, повсеместное использование IoT и концепция цифровой строительной площадки "Connected Site (Construction)" не просто повысит эффективность и снизит затраты, но и откроет новую эру безопасности, устойчивого строительства и прогнозируемого управления проектами.

> Одной из важных составляющих IoT компонентов являются также RFID (Radio Frequency Identification) метки. Они используются для идентификации и отслеживания материалов, техники и даже персонала на строительной площадке, повышая прозрачность и управляемость ресурсами проекта.

RFID-технология используется для автоматического распознавания объектов с помощью радиосигналов. Она состоит из трех ключевых элементов:

- RFID-метки (пассивные или активные) – содержат уникальный идентификатор и крепятся на материалы, инструменты или технику.
- Сканеры – устройства, которые считывают информацию с меток и передают ее в систему.
- Централизованная база данных – хранит информацию о местоположении, состоянии и движении объектов.

Применение RFID в строительстве:

- Автоматический учет материалов – метки на готовых бетонных изделиях, арматуре или пакетов с сэндвич-панелями позволяют контролировать запасы и предотвращать кражи.
- Контроль работы персонала – RFID-бейджи сотрудников регистрируют время начала и окончания смены, обеспечивая учет рабочего времени.
- Мониторинг оборудования – RFID-система отслеживает перемещение техники, предотвращая простои и повышая эффективность логистики.

Дополняет этот технологический набор смарт-контракты на базе блокчейн-технологий, которые позволяют автоматизировать платежи, контроль поставок и соблюдение условий контрактов без необходимости посредников, снижая риск мошенничества и задержек.

> Сегодня, при отсутствии единой модели данных, смарт-контракты представляют собой просто код, который согласовывают участники. Однако при Data-Centric подходе возможно создать общую модель параметров контракта, закодировать ее в блокчейне и автоматизировать выполнение условий.

Например, в системе управления поставками смарт-контракт сможет отслеживать доставку груза с IoT-датчиков и RFID-меток и автоматически переводить оплату при его прибытии. Аналогично, на строительной площадке смарт-контракт может фиксировать факт завершения этапа работ — например, монтаж арматуры или заливка фундамента — на основании данных от дронов или строительных сенсоров и автоматически инициировать следующий платёж подрядчику без необходимости ручной проверки и бумажных актов.

Но несмотря на новые технологии и усилия международных организаций по стандартизации, множество конкурирующих стандартов усложняют ландшафт IoT.

> Согласно исследованию Cisco, опубликованному в 2017 году [151], почти 60% инициатив в области Интернета вещей (IoT) останавливаются на стадии доказательства концепции, и только 26% компаний считают свои IoT-проекты полностью успешными. Более того, треть завершённых проектов не достигает заявленных целей и не признаётся успешной даже после внедрения.

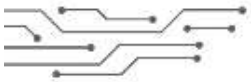
Одна из ключевых причин — отсутствие совместимости между платформами, обрабатывающими данные с различных сенсоров. В результате данные остаются изолированными в рамках отдельных решений. Альтернативой этому подходу, как и в других подобных случаях (которые мы рассматривали в этой книги) является архитектура, выстроенная вокруг самих данных как основного актива.

> IoT-сенсоры играют ключевую роль не только в мониторинге технического состояния оборудования, но и в предиктивной аналитике, позволяя снижать риски на строительной площадке и повышать общую производительность процессов, за счёт предсказания сбоев и отклонений.

Собранные с помощью IoT-сенсоров и RFID-меток данные могут в реальном времени обрабатываться алгоритмами машинного обучения, способными выявлять аномалии и заранее оповещать инженеров о потенциальных неисправностях. Это может быть как появление микротрещин в бетонных конструкциях, так и нехарактерные паузы в работе башенного крана, указывающие на технические сбои или нарушения регламента. Более того, продвинутые алгоритмы анализа поведения позволяют фиксировать поведенческие паттерны, которые могут свидетельствовать, например, о физической усталости персонала, повышая уровень проактивного управления безопасностью и благополучием сотрудников на площадке.

В строительной отрасли аварии и сбои — будь то техника или люди — редко случаются внезапно. Как правило, им предшествуют мелкие отклонения, которые остаются незамеченными. Предиктивная аналитика и машинное обучение позволяет обнаруживать эти сигналы на ранних этапах, ещё до наступления критических последствий.

Если документы, проектные файлы и данные с IoT-устройств и RFID-меток формируют цифровой след строительных объектов, то машинное обучение помогает извлекать из него полезные знания. С ростом объёмов данных и демократизации доступа к данным строительная отрасль получает новые возможности в области аналитики, прогнозирования и применения искусственного интеллекта.



ГЛАВА 9.2.
МАШИННОЕ ОБУЧЕНИЕ И ПРОГНОЗЫ

Машинное обучение и искусственный интеллект изменят то, как мы строим

Базы данных различных систем в строительном бизнесе — с их неизбежно ветшающей и усложняющейся инфраструктурой — становятся питательной средой для будущих решений. Серверы компании, подобно лесу, богаты биомассой важной информации, зачастую скрытой под землей, в недрах папок и серверов. Массы данных из различных систем, создаваемых сегодня - после использования, падения на дно сервера и после многолетнего окаменения - станут топливом для моделей машинного обучения и языковых моделей в будущем. На этих внутрикорпоративных моделях с использованием централизированных хранилищ будут построены внутренние чаты компаний (например, отдельный экземпляр локально настроенного ChatGPT, LlaMa, Mistral, DeepSeek), позволяющие быстро и удобно получать информацию и формировать необходимые графики, панели и документы.

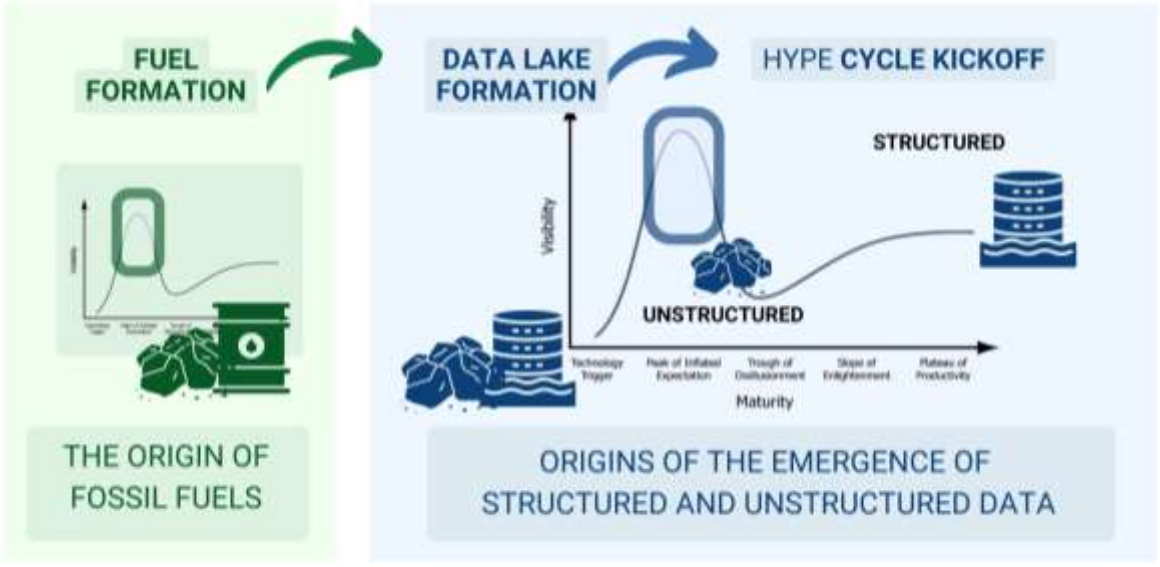


Рис. 9.2-1 Как деревья превращаются в уголь, так и информация со временем под давлением времени и аналитиков превращается в ценные источники энергии для бизнеса.

Окаменение растительной массы в сочетании с давлением и температурой создаёт однородную и уникально структурированную гомогенную массу деревьев разных пород, живших в разное время, - древесный уголь [152]. Так же и информация, записанная на жестких дисках в разных форматах и в разное время под давлением отделов аналитики и температурой менеджмента качества, в итоге образует однородную структурированную массу ценной информации (Рис. 9.2-1).

Эти слои (или чаще изолированные самородки) информации создаются при помощи кропотливой работы по организации данных опытными аналитиками, которые начинают постепенно извлекать

ценную информацию из, казалось бы, давно неактуальных данных.

В тот момент, когда эти созревшие пласты данных перестают просто «сгорать» в отчётах, а начинают циркулировать в бизнес-процессах, обогащая решения и улучшая процессы, компания становится готовой к следующему шагу — переходу к машинному обучению и искусственному интеллекту (Рис. 9.2-2).

> **Машинное обучение** (ML - Machine learning) — это класс методов для решения задач искусственного интеллекта. Алгоритмы машинного обучения распознают закономерности в больших массивах данных и используют их для самообучения. Каждый новый набор данных позволяет математическим алгоритмам совершенствоваться и адаптироваться в соответствии с полученной информацией, что позволяет постоянно повышать точность рекомендаций и предсказаний.

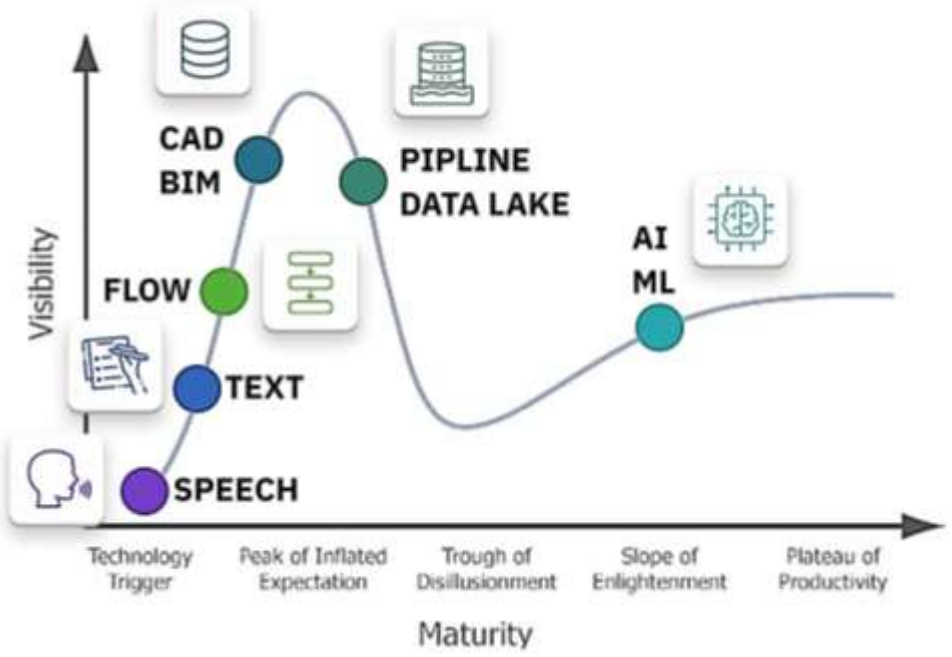


Рис. 9.2-2 Угасание технологий создания данных и применение инструментов аналитики открывают путь к теме машинного обучения.

Как сказал в интервью 2023 года влиятельный генеральный директор крупнейшего мирового инвестиционного фонда (которому принадлежат ключевые пакеты акций почти всех крупнейших компаний, производящих строительное программное обеспечение, а также компаний, владеющих самым большим количеством недвижимости в мире [55]) - машинное обучение изменит мир строительства.

> *ИИ обладает огромным потенциалом. Он изменит то, как мы работаем, то, как мы живем. ИИ и робототехника изменят то, как мы работаем и как мы строим, и мы сможем использовать ИИ и робототехнику как средство для создания гораздо большей производительности [153].*
>
> — Генеральный директор крупнейшего мирового инвестиционного фонда, интервью, сентябрь 2023 г.

Машинное обучение (ML) работает за счет обработки больших объемов данных, используя статистические методы для имитации аспектов человеческого мышления. Однако большинство компаний не располагает такими наборами данных, а если они и есть, то часто недостаточно размечены. Здесь могут помочь семантические технологии и трансферное обучение — метод, позволяющий ML быть эффективнее при работе с небольшими объемами данных, целесообразность которых обсуждалась в предыдущих главах данной части.

Суть трансферного обучения в том, что вместо обработки каждой задачи с нуля можно использовать знания, полученные в смежных областях. Нужно понимать что паттерны и открытия из других отраслей экономики могут быть адаптированы и применены в строительной отрасли. Например, методы оптимизации логистических процессов, разработанные в ритейле, помогают повысить эффективность управления строительными цепочками поставок. Анализ больших данных, активно используемый в финансах, может применяться для прогнозирования затрат и управления рисками в строительных проектах. А технологии компьютерного зрения и робототехники, развивающиеся в промышленности, уже находят применение в автоматизированном контроле качества, мониторинге безопасности и управлении объектами на строительной площадке.

> Трансферное обучение позволяет не только ускорять внедрение инноваций, но и снижать затраты на их разработку, используя уже накопленный опыт других отраслей.

Рис. 9.2-3 Технологии искусственного интеллекта и робототехника станут главной движущей силой будущего для повышения производительности в строительной отрасли.

Человеческое мышление устроено по схожему принципу: мы опираемся на ранее полученные знания для решения новых задач (Рис. 4.4-19, Рис. 4.4-20, Рис. 4.4-21). В машинном обучении этот подход тоже работает — упрощая модель данных и делая ее более элегантной, можно снизить сложность задачи для алгоритмов ML. Это, в свою очередь, уменьшает потребность в больших объемах данных и сокращает вычислительные затраты.

От субъективной оценки к статистическому прогнозу

Эпоха, когда стратегические решения зависели от интуиции отдельных руководителей (Рис. 9.2-4), уходит в прошлое. В условиях растущей конкуренции и сложных экономических условий субъективный подход становится слишком рискованным и неэффективным. Компании, продолжающие полагаться на личные мнения вместо объективного анализа данных, теряют возможность оперативно реагировать на изменения.

> Конкурентная среда требует точности и воспроизводимости, основанных на данных, статистических закономерностях и вычислимой вероятности. Решения больше не могут опираться на чувство, они должны опираться на корреляции, тренды и прогнозные модели, полученные с помощью аналитики и машинного обучения. Это не просто смена инструментов — это смена логики мышления: от предположений — к доказательствам, от субъективных вероятностей – к статистически рассчитанным отклонениям, от ощущения — к фактам.

Рис. 9.2-4 Эра решений принимаемых HiPPO (мнение самого высокооплачиваемого сотрудника) с приходом больших данных и машинного обучения уйдёт в прошлое.

Руководители, которые привыкли полагаться исключительно на собственные ощущения, неизбежно столкнутся с новой реальностью: авторитет больше не определяет выбор. В центре управления теперь находятся системы, анализирующие миллионы параметров и векторов, выявляющие скрытые закономерности и предлагающие оптимальные стратегии.

Главная причина, по которой компании сегодня ещё избегают внедрения ML, — его непрозрачность. Большинство моделей для менеджеров работают как "черные ящики", не объясняя, как

именно они приходят к своим выводам. Это ведет к проблемам: алгоритмы могут укреплять стереотипы и даже создавать курьезные ситуации, как в случае с чат-ботом Microsoft, который быстро превратился в токсичный инструмент общения [154].

> В книге "Deep Thinking" Гарри Каспаров, бывший чемпион мира по шахматам, размышляет о своем поражении от компьютера IBM Big Blue [155]. Он утверждает, что истинная ценность ИИ не в копировании человеческого интеллекта, а в дополнении наших способностей. ИИ должен выполнять задачи, в которых люди слабы, в то время как люди привносят творчество. Компьютеры изменили традиционный подход к анализу шахмат. Вместо создания увлекательных историй о партиях, компьютерные шахматные программы оценивают каждый ход беспристрастно, основываясь только на его фактической силе или слабости. Каспаров отмечает, что человеческая склонность воспринимать события как связные истории, а не как отдельные действия, часто приводит к неверным выводам — не только в шахматах, но и в жизни в целом.

Поэтому, если вы планируете использовать машинное обучение для прогнозирования и анализа, важно разобраться в его базовых принципах — как работают алгоритмы и как обрабатываются данные, прежде чем начать использовать инструменты машинного обучения и ИИ в своей работе. Лучший способ начать — это практический опыт.

Один из самых удобных инструментов для начального ознакомления с темой машинного обучения и предсказаний - Jupyter Notebook и популярный классический датасет Titanic, который позволит наглядно освоить ключевые методы анализа данных и построения моделей ML.

Titanic датасет: Hello World в мире аналитики данных и больших данных

Один из самых известных примеров использования ML в аналитике данных – анализ набора данных «Титаник», который часто используется для изучения вероятности выживания пассажиров. Изучение этой таблицы является аналогией программы «Hello World» при изучении языков программирования.

Затопление RMS Titanic в 1912 году привело к гибели 1502 из 2224 человек. Набор данных Titanic содержит не только информацию о том выжил ли пассажир, но и такие атрибуты как: возраст, пол, класс билета и другие параметры. Данный набор данных доступен бесплатно, и его можно открыть и анализировать на различных платформах оффлайн и онлайн.

```
Ссылка на набор данных Titanic:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Ранее в главе "IDE с поддержкой LLM и будущие изменения в программировании" уже обсуждался Jupyter Notebook – одну из самых популярных сред разработки для анализа данных и машинного обучения. Бесплатными облачными аналогами Jupyter Notebook являются платформы Kaggle и Google Collab, которые позволяют запускать Python код без установки программного обеспечения и предоставляют бесплатный доступ к вычислительным ресурсам.

> Kaggle – крупнейшая платформа для анализа данных, соревнований по машинному обучению с встроенной средой выполнения кода. По состоянию на октябрь 2023 года Kaggle насчитывает более 15 миллионов пользователей [156] из 194 стран.

Загрузи и используем датасет Титаник на платформе Kaggle (Рис. 9.2-5), чтобы хранить датасет (его копию) и запустить Python код с предустановленными библиотеками напрямую в браузере, без необходимости устанавливать специально IDE.

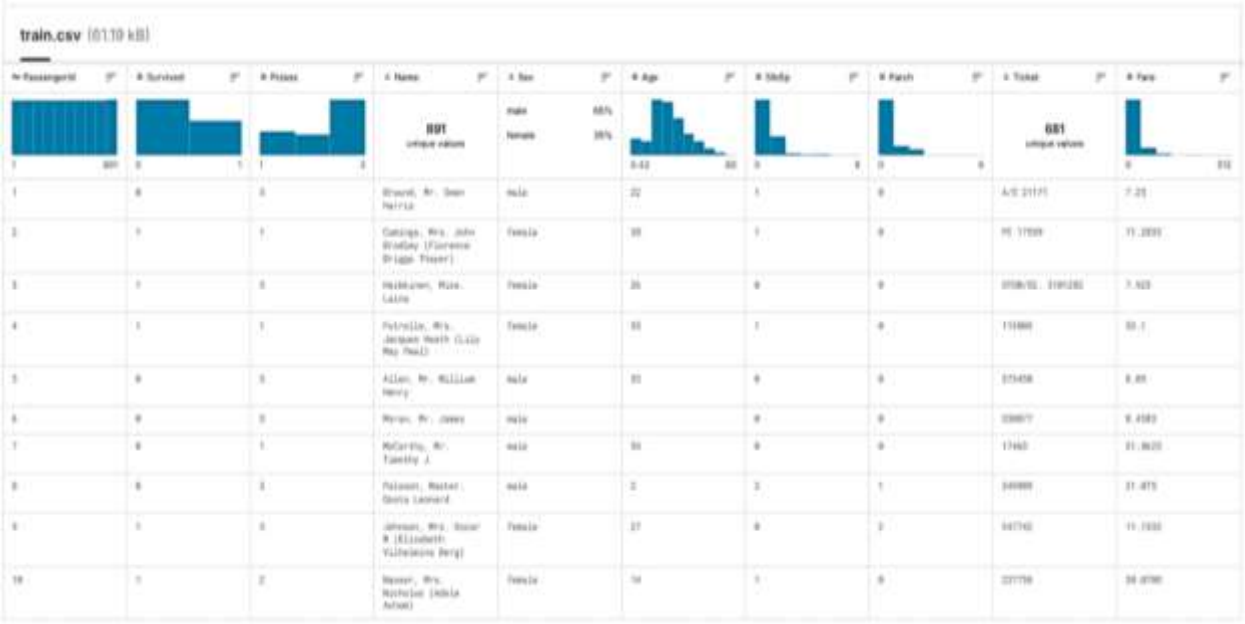


Рис. 9.2-5 Статистика Titanic таблицы – самого популярного учебный датасета для изучения анализа данных и машинного обучения.

Датасет Titanic включает данные о 2224 пассажирах, находившихся на борту *RMS Titanic* во время его крушения в 1912 году. Набор представлен в виде двух отдельных таблиц — тренировочной (train.csv) и тестовой (test.csv) выборке, что позволяет использовать его как для обучения моделей, так и для оценки их точности на новых данных.

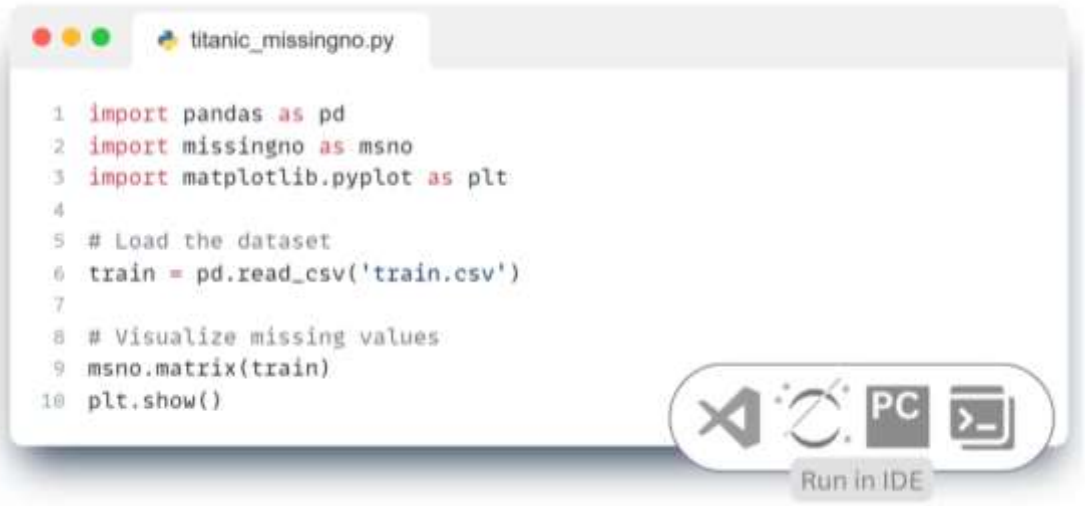
Тренировочный датасет содержит как признаки-атрибуты пассажиров (возраст, пол, класс билета и другие), так и информацию о том, кто выжил (колонка с бинарными значениями «Выжил»). Тренировочный датасет (Рис. 9.2-6 – файл train.csv) используется для обучения модели. Тестовый датасет (Рис. 9.2-7 – файл test.csv) включает только признаки пассажиров без информации о выживших (без одной единственной колонки «Выжил»). Тестовый датасет предназначен для проверки работы модели на новых данных и оценки её точности.

Таким образом, мы имеем почти идентичные атрибуты пассажиров в обучающем и тестовом датасетах. Единственное ключевое отличие заключается в том, что в тестовом наборе данных у нас список пассажиров у которых отсутствует колонка "Выжил" — целевая переменная, которую мы и хотим научиться предсказывать с помощью различных математических алгоритмов. И после построения модели мы сможем сравнить вывод нашей модели с реальным параметр "Выжил" из тестового датасета, который мы будем учитывать для оценки результатов.

Основные столбцы таблицы, параметры пассажиров в тренировочном и тестовом датасете:

- **PassengerId** – уникальный идентификатор пассажира
- **Survived** – 1, если пассажир выжил, 0 – если погиб (отсутствует в тестовом наборе)
- **Pclass** – класс билета (1, 2 или 3)
- **Name** – имя пассажира
- **Sex** – пол пассажира (male/female)
- **Age** – возраст
- **SibSp** – количество братьев/сестер или супругов на борту
- **Parch** – количество родителей или детей на борту
- **Ticket** – номер билета
- **Fare** – стоимость билета
- **Cabin** – номер каюты (многие данные отсутствуют)
- **Embarked** – порт посадки (C = Cherbourg, Q = Queenstown, S = Southampton)

Для визуализации пропущенных данных в обоих таблицах можно использовать библиотеку missingno (Рис. 9.2-6, Рис. 9.2-7), которая отображает пропущенные значения в виде гистограммы, где белые поля показывают отсутствующие данные. Такая визуализация позволяет быстро оценить качество данных перед их обработкой.



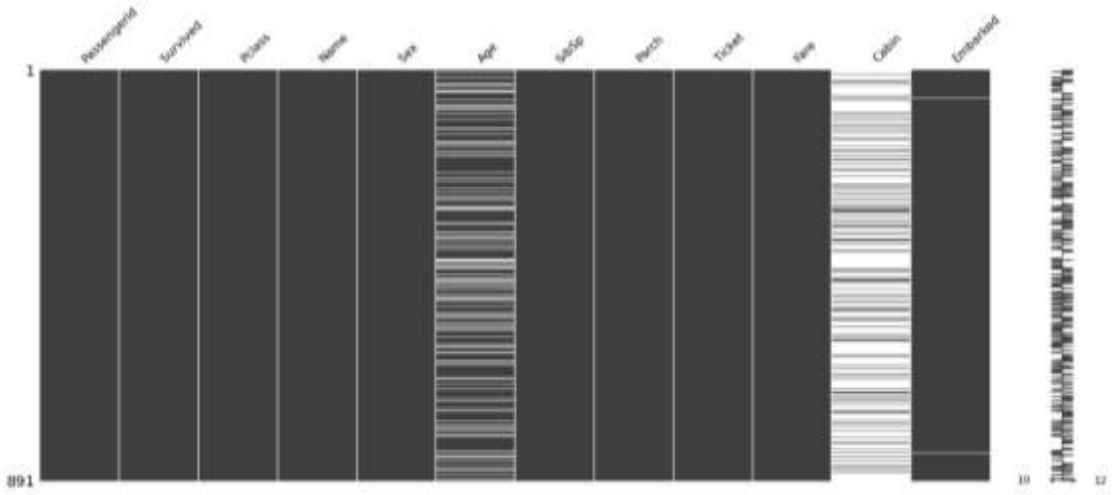


Рис. 9.2-6 С помощью нескольких строк кода визуализируются пропущенные данные в тренировочном датасете Titanic, где ключевым параметром для обучения является параметр «Выжил».

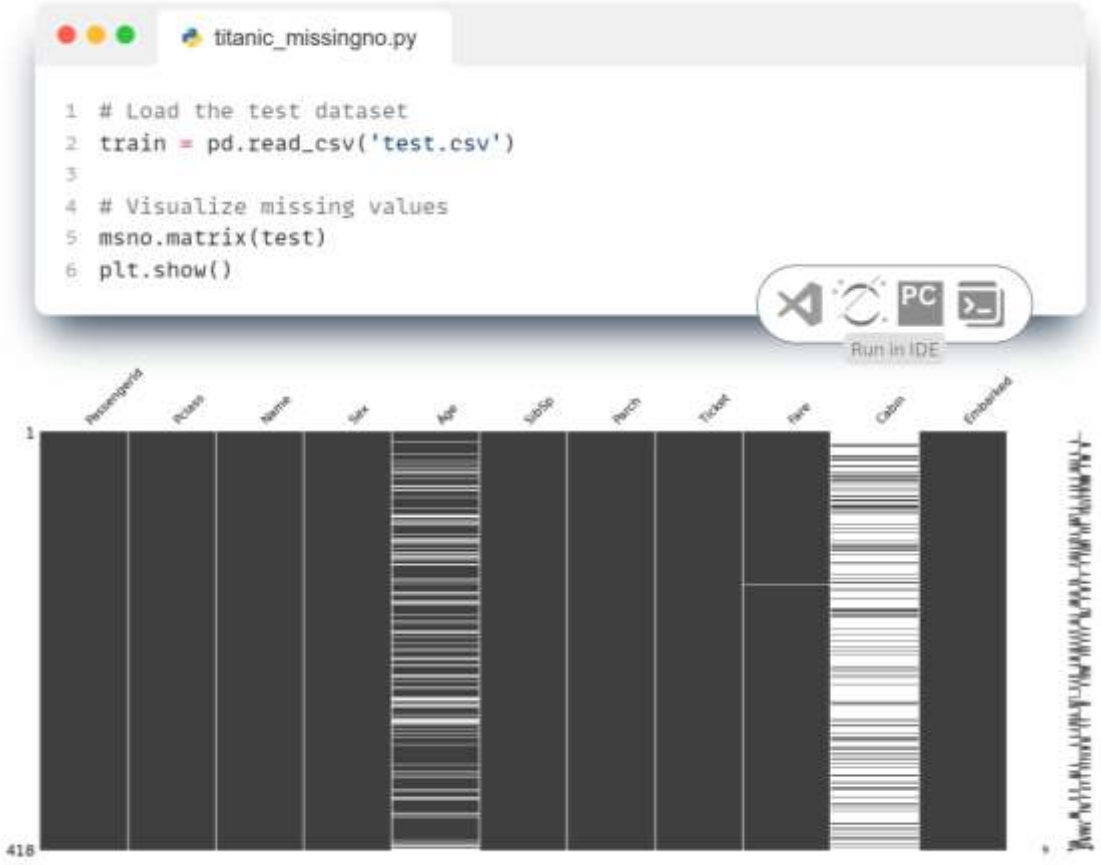


Рис. 9.2-7 Визуализация пропущенных данных в тестовом датасете Titanic, который содержит только характеристики пассажиров без информации .

Прежде чем формулировать гипотезы и делать прогнозы на основе датасета, визуальный анализ помогает выявить ключевые закономерности в данных, оценить их качество и определить возможные зависимости. Существует множество методов визуализации, которые помогают лучше понять датасет Titanic. Вы можете использовать графики распределения для анализа возрастных групп пассажиров, диаграммы выживаемости в зависимости от пола и класса, а также матрицы пропущенных данных, чтобы оценить качество информации и понять данные.

- Попросим LLM помочь нам визуализировать данные из датасета Титаник, для этого отправим следующий текстовый запрос в любую LLM-модель (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или любую другую):

> Покажи пожалуйста несколько простых графиков для датасета Титаник. Скачай самостоятельно датасет и покажи графики ⏎

- Ответ LLM в виде готового кода и графиков, визуализирующих параметры датасета:

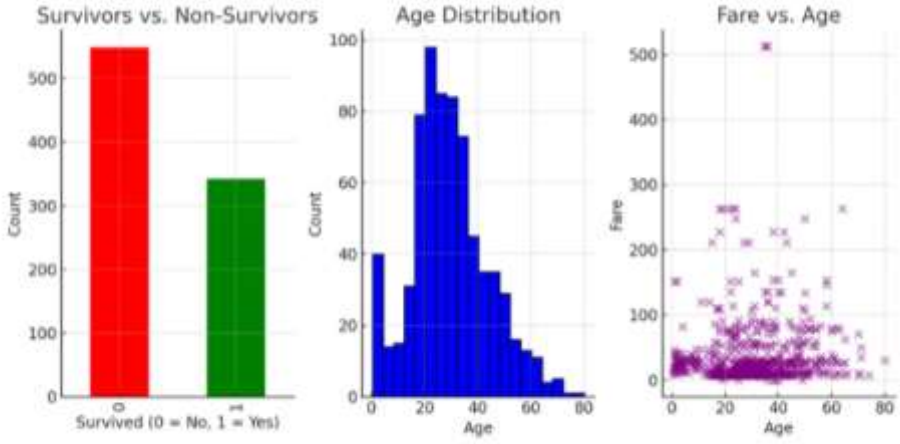


Рис. 9.2-8 LLM помогает мгновенно получить визуализацию данных датасета.

Визуализация данных — это важный этап, который позволяет подготовить датасет для последующего построения модели машинного обучения, к которому можно перейти только поняв данные.

Машинное обучение в действии: от пассажиров "Титаника" к управлению проектами

Основная гипотеза, используемая для изучения основ машинного обучения на основе датасета Titanic, заключается в том, что некоторые группы пассажиров имели более высокие шансы на выживание.

> Небольшая таблица пассажиров Титаника стала популярной во всём мире, и миллионы людей используют её для обучения, экспериментов и тестирования моделей чтобы выяснить какие алгоритмы и гипотезы позволят построить максимальную точную модель предсказания выживаемости на основании тренировочного датасета для пассажиров Титаника.

Привлекательность датасета Титаника объясняется компактностью: при нескольких сотнях строк и двенадцати столбцах (Рис. 9.2-6) она предоставляет широкие возможности для анализа. Датасет представляет собой, относительно простой, классический пример решения бинарной классификации, где цель задачи – выживание – выражена в удобном формате 0 или 1.

> Джон Уилер в «It from Bit» [7] утверждает, что в основе мироздания лежат бинарные решения. Аналогично и бизнес, управляемый людьми, состоящими из молекул, фактически строится на череде двоичных бинарных выборов.

Кроме того, данные основаны на реальном историческом событии, что делает их ценными для исследования, в отличие от искусственно созданных примеров. Только на платформе Kaggle, одной из крупнейших площадок для работы с Data Pipeline и ETL, в решении задач на основе датасета Титаник приняли участие 1 355 998 человек, разработав 53 963 уникальных Data Pipeline-решения [157] (Рис. 9.2-9).

Кажется невероятным, но всего 1000 строк данных о пассажирах "Титаника" с 12 параметрами стали полем для миллионов гипотез, логических цепочек и уникальных Data-Pipelines. Из маленького датасета рождаются бесконечные инсайты, гипотезы и интерпретации – от простых моделей выживаемости до сложных ансамблей, учитывающих скрытые закономерности и сложны лабиринты рассуждений.

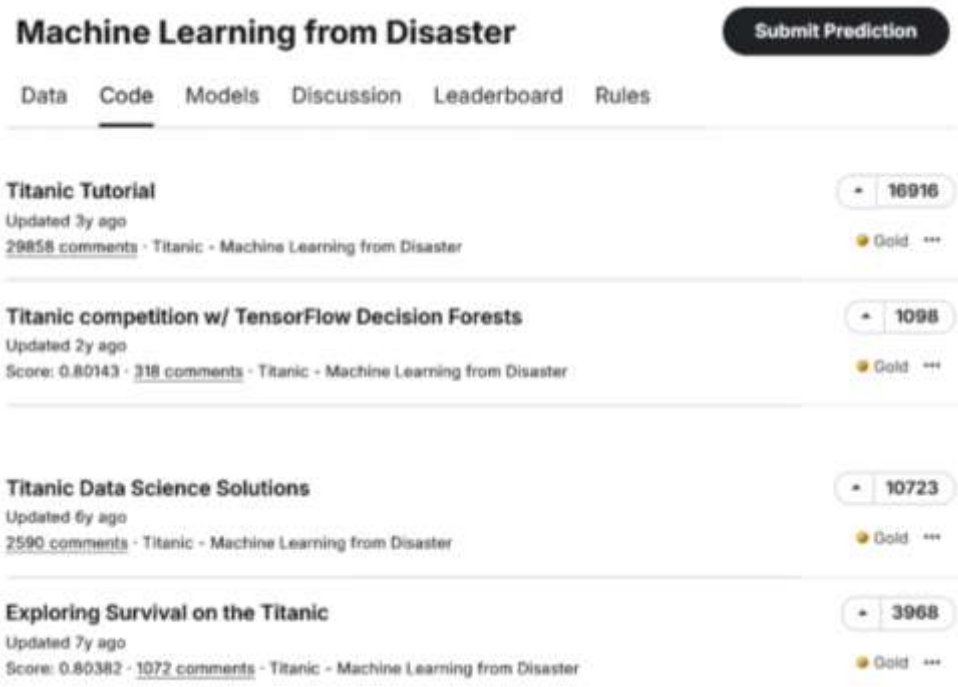


Рис. 9.2-9 Первые пять решений из всего 53 963 готовых и открытых Pipeline. Почти 1,5 миллиона человек уже пробовали решить эту задачу только на Kaggle [157].

Если даже такая небольшая таблица способна порождать миллионы уникальных решений (Рис. 9.2-9), то что говорить о реальных промышленных строительных датасетах, где параметры измеряются десятками тысяч?

Стандартный CAD-проект относительно небольшого здания содержит десятки тысяч сущностей с тысячами параметров — от геометрических характеристик до стоимостных и временных атрибутов. Представьте, сколько потенциальных инсайтов, взаимосвязей, предсказаний и управленческих гипотез скрыто в данных всех проектов вашей компании, собранных за последние годы. Исторические проектные данные это не просто архив — это живая память организации, её цифровой след, который можно анализировать для постройки большого количества уникальных гипотез.

Самое важное — не нужно ждать, пока сообщество Kaggle заинтересуется вашей компанией или вашими данными. Уже сегодня можно начать работать с тем, что у вас есть: запускать аналитику на собственных данных, обучать модели на собственных данных, выявлять повторы, аномалии и

паттерны. Там, где раньше требовались годы экспериментов и дорогой консалтинг, сегодня достаточно инициативы, LLM, открытого подхода к данным и готовности к обучению.

- Чтобы построить алгоритм машинного обучения, который будет предсказывать выживаемость пассажиров на основании тренировочного датасета пассажиров train.csv попросим LLM решить для нас эту задачу:

> На основании тренировочного датасета пассажиров Титаника построй модель машинного обучения для предсказания выживаемости ⏎

- Ответ LLM:

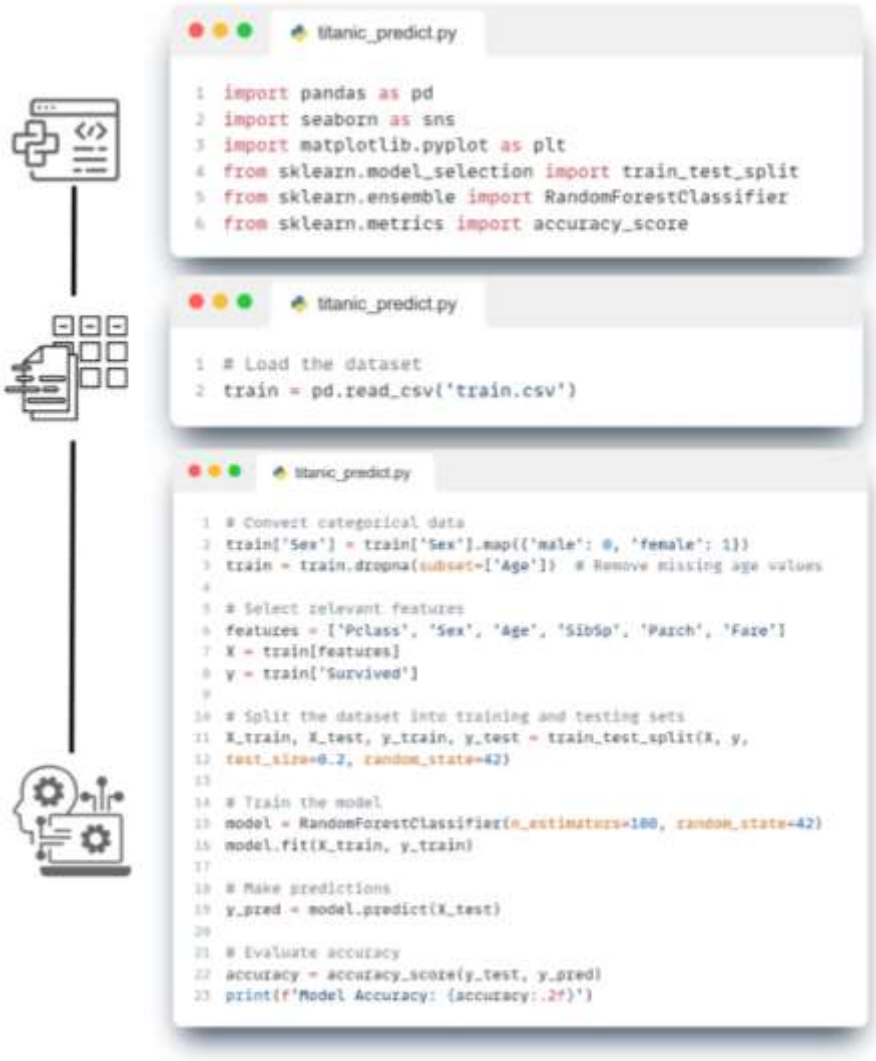


Рис. 9.2-10 LLM построил прогнозирование выживших на Титанике с использованием алгоритма машинного обучения Random Forest.

Полученный код от LLM (Рис. 9.2-10) загружает данные о пассажирах "Титаника", очищает их, преобразует категориальные переменные (например, пол в числовой формат) и обучает модель через алгоритм RandomForestClassifier предсказывать, выжил пассажир или нет (подробнее про популярные алгоритмы мы поговорим в следующих главах).

При помощи кода данные при обучении разделяются на тренировочный и тестовый наборы (на сайте Kaggle для обучения уже созданы готовые test.csv (Рис. 9.2-7) и train.csv (Рис. 9.2-6), затем модель обучается на тренировочных данных и проверяется на тестовых, чтобы понять насколько хороша та или иная модель предсказаний. После обучения тестовые данные из test.csv (с реальными данными о тех, кто выжил или не выжил) подаются в модель, и она предсказывает, кто выжил, а кто нет. В нашем случае точность полученной нами модели машинного обучения около 80%, что показывает, что она достаточно хорошо улавливает закономерности.

> Машинное обучение можно сравнить с ребёнком, который пытается вставить прямоугольный блок в круглое отверстие. На начальных этапах алгоритм пробует множество подходов, сталкиваясь с ошибками и несоответствиями. Этот процесс может показаться неэффективным, однако он обеспечивает важное обучение: анализируя каждую ошибку, модель улучшает свои прогнозы и принимает всё более точные решения.

Теперь эту модель (Рис. 9.2-10) можно использовать для предсказания выживаемости новых пассажиров и например, если подать в нее информацию о пассажире при помощи функции model.predict с параметрами: «мужчина», «3-й класс», «25 лет», «без родственников на борту», модель выдаст прогноз – что пассажир с вероятностью 80% не выживет при катастрофе, если он был на теплоходе Титаник в 1912 году (Рис. 9.2-11).

Рис. 9.2-11 Модель которую мы создали, выше теперь может с 80% вероятностью предсказать выживет или не выживет тот или иной новый пассажир Титаника.

Модель предсказания выживаемости пассажиров "Титаника" иллюстрирует гораздо более широкую концепцию: ежедневно тысячи специалистов в строительной отрасли принимают подобные "дуальные" решения – жизнь или смерть решения, проекта, сметы, инструмента, прибыли или убытка, безопасности или риска. Как и в примере с "Титаником", где исход зависел от факторов (пол, возраст, класс), в строительстве на каждый из аспектов решения влияет множество собственных факторов и переменных (колонок таблиц): стоимость материалов, квалификация рабочих, сроки, погода, логистика, технические риски, комментарии и сотни тысяч др. параметров.

> В строительной отрасли машинное обучение применяется по тем же принципам, что и в других сферах: модели обучаются на исторических данных — из проектов, договоров, смет — для проверки различных гипотез и поиска наиболее эффективных решений. Этот процесс во многом напоминает обучение ребёнка через метод проб и ошибок: с каждым циклом модели адаптируются и становятся точнее.

Использование накопленных данных открывает для строительства новые горизонты. Вместо трудоёмких ручных расчётов можно обучить модели, способные с высокой степенью точности прогнозировать ключевые характеристики будущих проектов. Таким образом, предиктивная аналитика превращает строительную отрасль в пространство, где можно не только планировать, но и уверенно предсказывать развитие событий.

Предсказания и прогнозы на основе исторических данных

Собранные данные о проектах компании открывают возможность построения моделей, способных

прогнозировать стоимость и временные характеристики будущих, ещё не реализованных объектов — без трудоёмких ручных расчётов и сопоставлений. Это позволяет значительно ускорить и упростить процессы оценки, опираясь не на субъективные предположения, а на обоснованные математические прогнозы.

Ранее, в четвёртой части книги, мы подробно рассматривали традиционные методы расчёта сметной стоимости проектов, включая ресурсный метод, а также упоминали параметрический и экспертный подходы. Эти методы по-прежнему актуальны, но в современной практике они начинают обогащаться инструментами статистического анализа и машинного обучения, что позволяет значительно повысить точность и воспроизводимость оценок.

Процессы ручного и полуавтоматического расчёта цен и временных атрибутов будут в будущем дополнены мнением и предсказаниями ML-моделей, способных анализировать исторические данные, находить скрытые закономерности и предлагать обоснованные решения. Новые данные и сценарии будут генерироваться автоматически из уже имеющейся информации — подобно тому, как языковые модели (LLM) создают тексты, изображения и код на основе данных, собранных за годы из открытых источников [158].

> Так же, как сегодня человек опирается на опыт, интуицию и внутреннюю статистику при оценке будущих событий, в ближайшие годы будущее строительных проектов будет всё больше определяться сочетанием накопленных знаний и математических моделей машинного обучения.

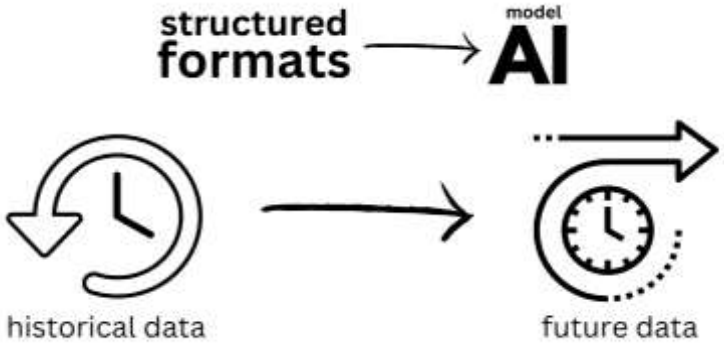


Рис. 9.2-12 Качественные и структурированные исторические данные компании - материал, на основе которого строятся модели машинного обучения и прогнозы.

Рассмотрим простой пример: прогнозирование цены дома на основе его площади, размера участка, количества комнат и географического положения. Один из подходов — построение классической модели, которая анализирует эти параметры и рассчитывает предполагаемую цену (Рис. 9.2-13). Такой подход требует точной и заранее известной формулы, что в реальной практике практически невозможно.

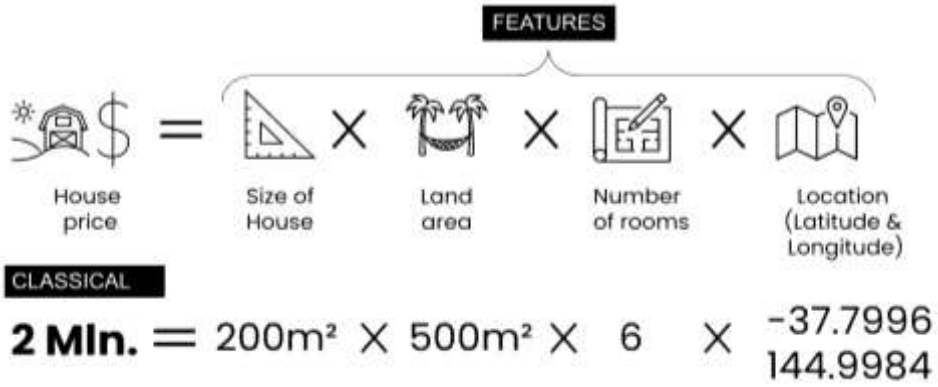


Рис. 9.2-13 Для оценки стоимости дома можно использовать классический алгоритм с фиксированной формулой, которую необходимо найти.

Машинное обучение позволяет отказаться от ручного поиска формул и заменить их обучаемыми алгоритмами, которые самостоятельно выявляют зависимости, многократно превосходящие по точности любые заранее заданные уравнения. В качестве альтернативы создадим алгоритм машинного обучения, который будет генерировать модель на основе предварительного понимания проблемы и исторических данных, которые могут быть неполными (Рис. 9.2-14).

> На примере задачи ценообразования машинное обучение позволяет создавать разные типы математических моделей, не требующие знания точного механизма формирования стоимости. Модель «учится» на данных о предыдущих проектах, подстраиваясь под реальные закономерности между параметрами зданий, их стоимостью и сроками выполнения.

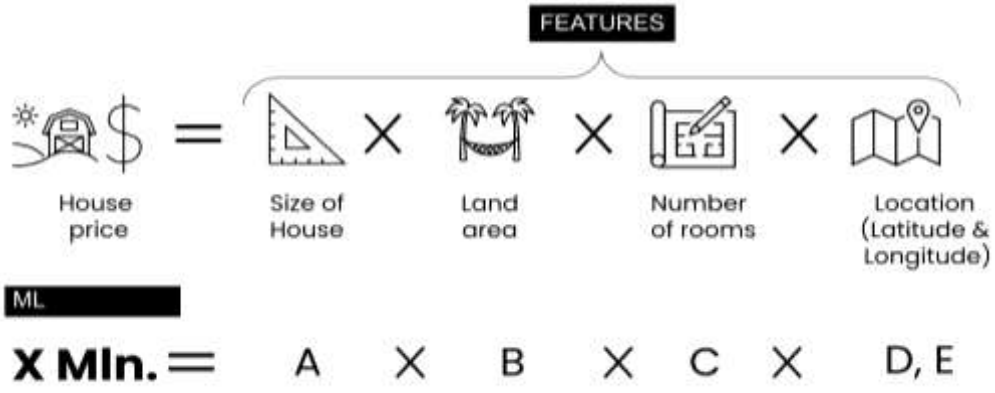


Рис. 9.2-14 В отличие от классической оценки по формулам, алгоритм машинного обучения обучается на исторических данных.

В контексте машинного обучения под наблюдением (supervised machine learning) каждый проект в обучающем наборе данных содержит как входные атрибуты (например, данные о стоимости и времени строительства аналогичных зданий), так и ожидаемые выходные значения (например, стои-

мость или время). Подобный набор данных используется для создания и настройки модели машинного обучения (Рис. 9.2-15). Чем больше набор данных и чем выше качество данных в нем, тем точнее будет модель и тем точнее будут результаты прогнозов.

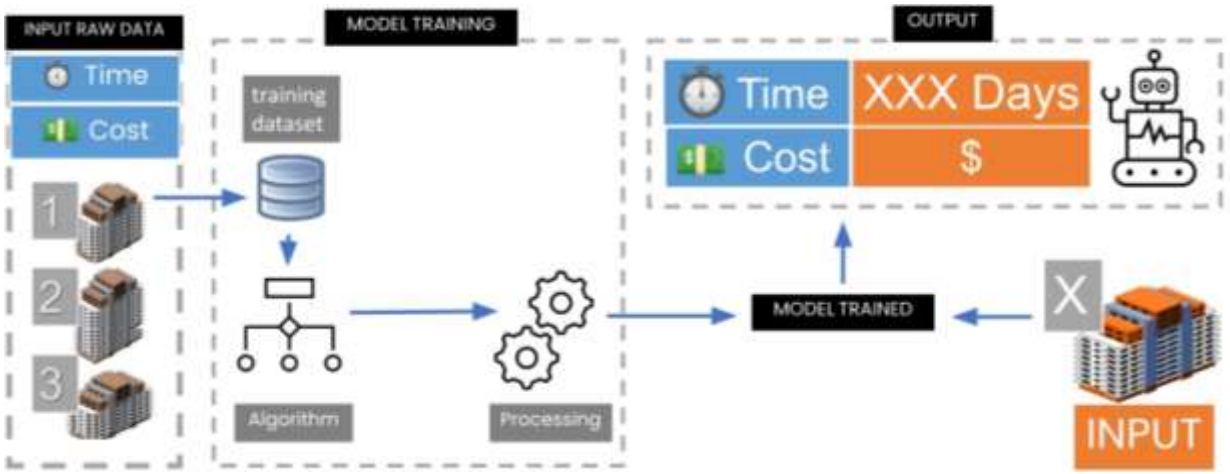


Рис. 9.2-15 Модель ML, обученная на данных о стоимости и графике выполнения прошлых проектов, с определённой вероятностью определит стоимость и график выполнения нового проекта.

После создания и обучения модели для оценки строительства нового проекта достаточно предоставить модели новые атрибуты для нового проекта, и модель с определенной вероятностью предоставит расчетные результаты, основанные на ранее изученных закономерностях.

Ключевые понятия машинного обучения

Машинное обучение — это не магия, а всего лишь математика, данные и поиск закономерностей. Оно не обладает настоящим интеллектом, а представляет собой программу, обученную на данных, чтобы распознавать шаблоны и принимать решения без постоянного участия человека.

Для описания своей структуры машинное обучение использует ряд ключевых понятий (Рис. 9.2-16):

- **Метки (Labels)** — это целевые переменные или атрибуты (параметр «Выжил» в датасете Титаника), которые должна предсказать модель. Пример: стоимость строительства (например, в долларах), продолжительность строительных работ (например, в месяцах).
- **Характеристики (Features)** — это независимые переменные или атрибуты, которые служат входными данными для модели. В модели прогнозирования они используются для предсказания меток. Примеры: площадь участка (в квадратных метрах), количество этажей здания, общая площадь здания (в квадратных метрах), географическое положение (широта и долгота), тип материалов, использованных при строительстве. Количество характеристик также определяет размерность данных.
- **Модель (Model)** — это набор различных гипотез, одна из которых приближает целевую функцию, которую нужно предсказать или аппроксимировать. Пример: модель машинного обучения, использующая методы регрессионного анализа для прогнозирования стоимости и сроков строительства.

- **Алгоритм обучения (Learning Algorithm)** — это процесс поиска наилучшей гипотезы в модели, которая точно соответствует целевой функции, используя набор обучающих данных. Пример: Алгоритм линейной регрессии, KNN или случайного леса, анализирующий данные о стоимости и сроках строительства для выявления взаимосвязей и закономерностей.
- **Обучение (Training)** — в процессе обучения алгоритм анализирует обучающие данные, находя закономерности, которые соответствуют взаимосвязи между входными атрибутами и целевыми метками. Результатом этого процесса является обученная модель машинного обучения, готовая к прогнозированию. Пример: процесс, в котором алгоритм анализирует исторические данные о строительстве (стоимость, время, характеристики объекта) для создания прогнозной модели.

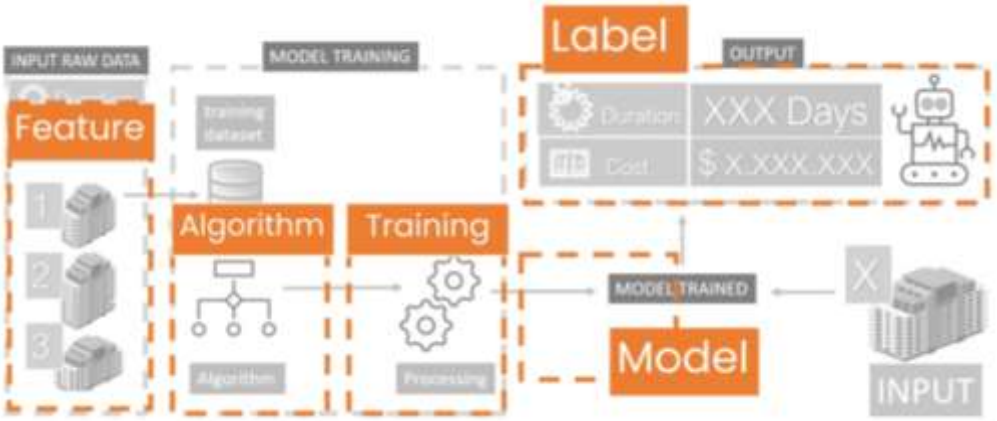


Рис. 9.2-16 ML использует метки и атрибуты для создания моделей, которые обучаются на данных с помощью алгоритмов для предсказания результатов.

Машинное обучение не существует в изоляции, а является частью более широкой экосистемы аналитических дисциплин, включающих статистику, базы данных, интеллектуальный анализ данных, распознавание образов, аналитику больших данных и искусственный интеллект. Рис. 9.2-17 демонстрирует, как эти области пересекаются и дополняют друг друга, создавая комплексную основу для современных систем принятия решений и автоматизации.

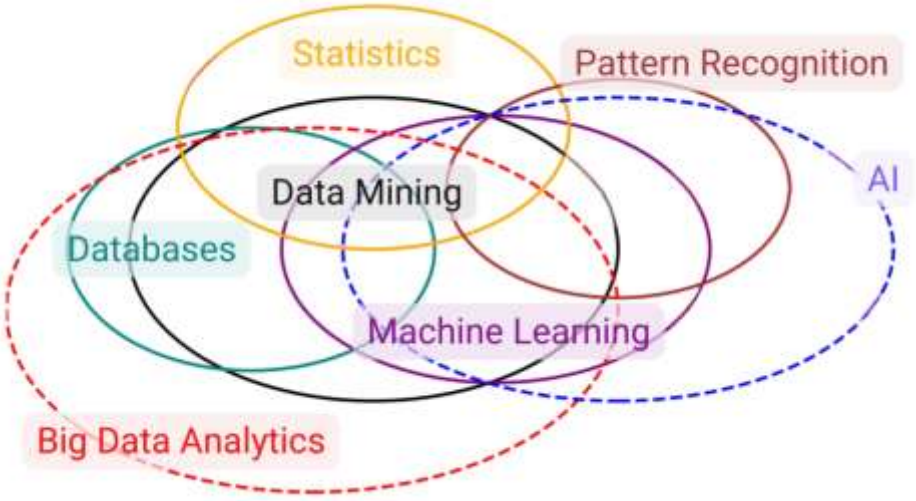
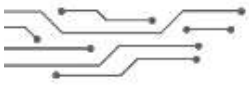


Рис. 9.2-17 Взаимосвязь между различными областями анализа данных: статистикой, машинным обучением, искусственным интеллектом, большими данными, распознаванием образов и интеллектуальным анализом данных.

> Основная цель машинного обучения - наделить компьютеры способностью автоматически усваивать знания без вмешательства или помощи человека и соответствующим образом корректировать свои действия [159].

Таким образом, в будущем роль человека будет заключаться лишь в предоставлении машине когнитивных возможностей - он будет задавать условия, веса и параметры, а модель машинного обучения будет делать все остальное.

В следующей главе рассмотрим конкретные примеры применения алгоритмов. На реальных таблицах и упрощённых моделях будет показано, как шаг за шагом строится прогноз.



ГЛАВА 9.3.
ПРОГНОЗИРОВАНИЕ СТОИМОСТИ И СРОКОВ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

Пример использования машинного обучения для нахождения стоимости и сроков проекта

Оценка сроков и стоимости строительства — один из ключевых процессов в деятельности строительной компании. Традиционно такие оценки выполняются экспертами на основе опыта, справочников и нормативных баз. Однако в условиях цифровой трансформации и роста доступности данных появляется возможность использовать модели машинного обучения (ML) для повышения точности и автоматизации подобных оценок.

Внедрение машинного обучения в процесс расчета стоимости и сроков строительства позволяет не только повысить эффективность планирования, но и становится отправной точкой для интеграции интеллектуальных моделей в другие бизнес-процессы — от управления рисками до оптимизации логистики и закупок.

> Важно уметь быстро определить, сколько времени займет строительство объекта и какова будет его общая стоимость. Эти вопросы о сроках и стоимости проекта традиционно занимают центральное место в сознании как заказчиков, так и строительных компаний с момента зарождения строительной отрасли.

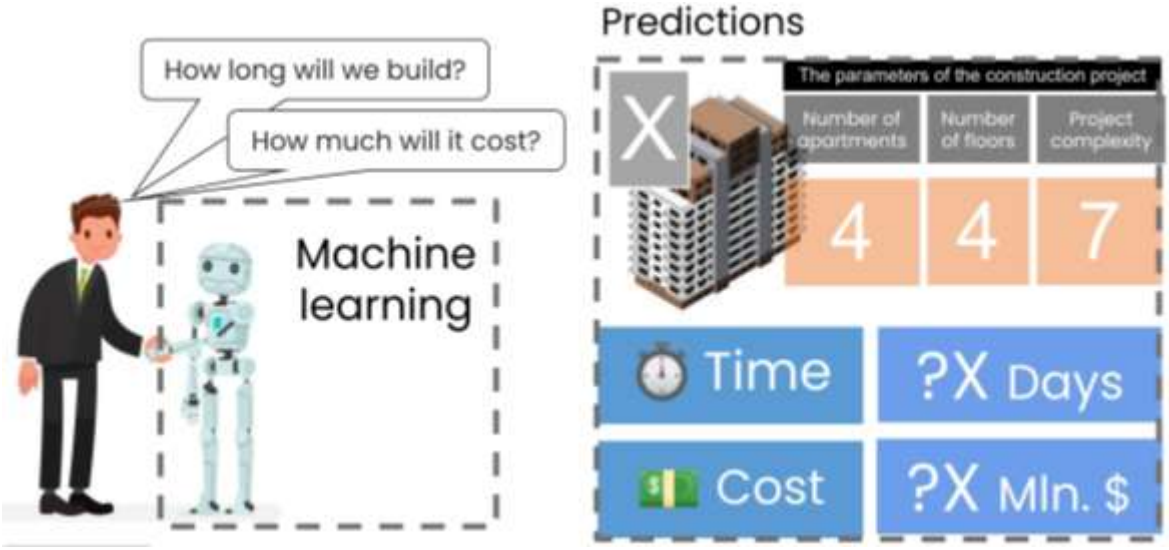


Рис. 9.3-1 В строительных проектах ключевыми факторами успеха являются скорость и качество оценки сроков и стоимости строительства.

В следующем примере будут извлечены ключевые данные из прошлых проектов, и на их основе будет разработана модель машинного обучения, которая позволит нам при помощи этой модели оценить стоимость и сроки реализации новых строительных проектов с новыми параметрами

(Рис. 9.3-1).

Рассмотрим три проекта с тремя ключевыми атрибутами: количеством квартир (где 100 квартир эквивалентны числу 10 для простоты визуализации), количеством этажей и условной мерой сложности строительства по шкале от 1 до 10, где 10 - самый высокий показатель сложности. В машинном обучении процесс преобразования и упрощения таких значений, как 100 в 10 или 50 в 5, называется "нормализацией".

> **Нормализация** в машинном обучении — это процесс приведения различных числовых данных к единому масштабу для облегчения их обработки и анализа. Этот процесс особенно важен, когда данные имеют разные масштабы и единицы измерения.

Предположим, что в первом проекте (Рис. 9.3-2) было 50 квартир (после нормализации – 5), 7 этажей и оценка сложности 2, что означало относительно простое строительство. Во втором проекте было уже 80 квартир, 9 этажей и относительно сложный проект. При таких условиях строительство первого и второго многоквартирного дома заняло 270 и 330 дней, а общая стоимость проекта составила 4,5 и 5,8 миллиона долларов соответственно.

Рис. 9.3-2 Пример набора прошлых проектов, которые будут использоваться для оценки времени и стоимости будущего проекта X.

При построении модели машинного обучения для таких данных основной задачей является определение критических атрибутов (или меток) для прогнозирования, в данном случае - времени и стоимости строительства. Имея небольшой набор данных, мы будем использовать информацию о предыдущих строительных проектах для планирования новых: используя алгоритмы машинного обучения, мы должны предсказать стоимость и продолжительность строительства нового проекта X на основе заданных атрибутов нового проекта, таких как 40 квартир, 4 этажа и относительная высокая сложность проекта – 7 (Рис. 9.3-2). В реальных условиях число входных параметров может быть значительно больше — от нескольких десятков до сотен факторов. К ним могут относиться: тип строительных материалов, климатическая зона, уровень квалификации подрядчиков, наличие инженерных сетей, тип фундамента, сезон начала работ, комментарии прорабов и т. д.

> Чтобы создать прогностическую модель машинного обучения, нам нужно выбрать алгоритм для ее создания. Алгоритм в машинном обучении — это как математический рецепт, который учит компьютер, как делать прогнозы (смешивать в правильном порядке параметры) или принимать решения на основе данных.

Чтобы проанализировать данные о прошлых строительных проектах и предсказать время и стоимость будущих проектов (Рис. 9.3-2), можно использовать один из популярных алгоритмов машинного обучения:

- **Линейная регрессия (Linear regression):** этот алгоритм пытается найти прямую зависимость между атрибутами, например, между количеством этажей и стоимостью строительства. Цель алгоритма - найти линейное уравнение, которое наилучшим образом описывает эту связь, что позволяет делать прогнозы.
- **Алгоритм к-ближайших соседей (K-nearest neighbors (k-NN)):** этот алгоритм сравнивает новый проект с прошлыми проектами, которые были похожи по размеру или сложности. k-NN классифицирует данные на основе того, какие из k (количество) обучающих примеров находятся ближе всего к ним. В контексте регрессии результатом является среднее значение или медиана из k ближайших соседей.
- **Деревья решений (Decision Trees)**: это модель прогнозного моделирования, которая разделяет данные на подмножества, основанные на различных условиях, используя древовидную структуру. Каждый узел дерева представляет собой условие или вопрос, ведущий к дальнейшему разделению данных, а каждый лист - окончательное предсказание или результат. Алгоритм делит данные на более мелкие группы на основе различных характеристик, например, сначала по количеству этажей, затем по сложности и так далее, чтобы сделать прогноз.

Давайте рассмотрим алгоритмы машинного обучения для оценки стоимости нового проекта на примере двух популярных алгоритмов: линейной регрессии и алгоритма к-ближайших соседей (K-nearest neighbors).

Прогноз стоимости и времени проекта при помощи линейной регрессии

Линейная регрессия — это фундаментальный алгоритм анализа данных, позволяющий предсказать значение переменной на основе линейной связи с одной или несколькими другими переменными. Эта модель предполагает, что существует прямая линейная связь между зависимой переменной и одной или несколькими независимыми переменными, и цель алгоритма - найти эту связь.

> Простота и понятность линейной регрессии сделали ее популярным инструментом в различных областях. При работе с одной переменной линейная регрессия заключается в нахождении наилучшим образом подходящей линии, проходящей через точки данных.

Линейная регрессия находит наилучшую прямую (красная линия), которая аппроксимирует зави-

симость между входной переменной X и выходной переменной Y. Эта линия позволяет предсказывать значения Y для новых значений X на основе выявленной линейной зависимости (Рис. 9.3-3).

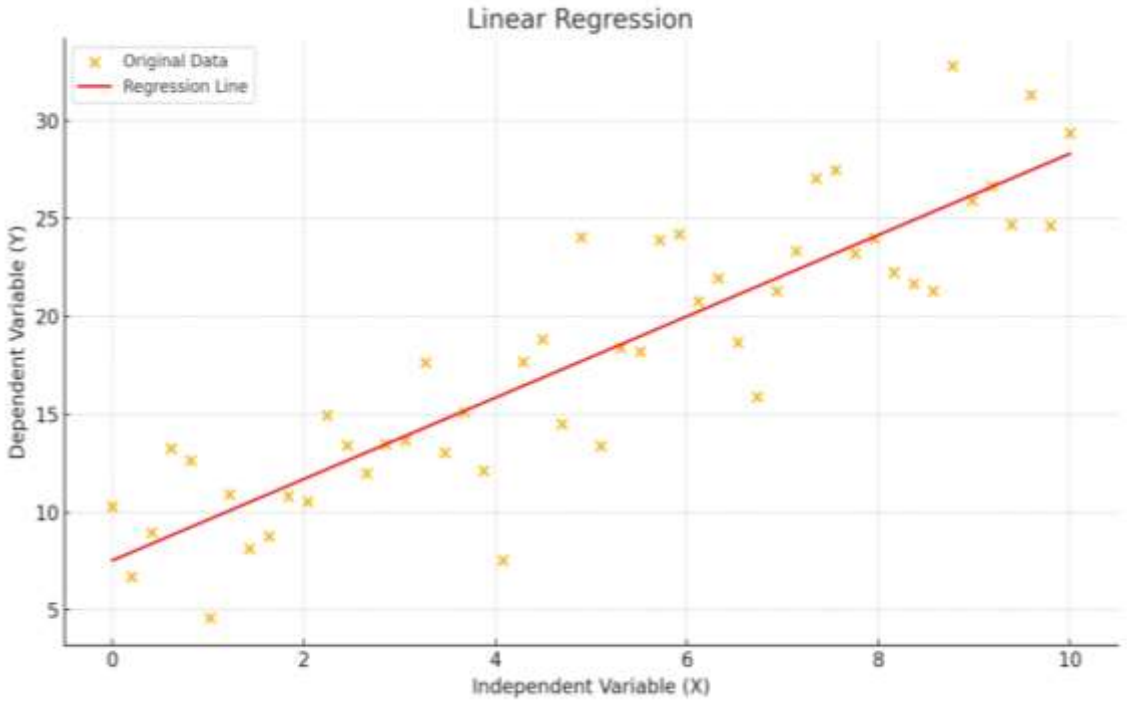


Рис. 9.3-3 Принцип работы линейной регрессии в нахождении наилучшую прямую, которая будет проходить через тренировочные значения.

Эта линия представлена уравнением, в котором при вводе значения независимой переменной (X) получается прогнозируемое значение зависимой переменной (Y). Этот процесс позволяет эффективно прогнозировать Y на основе известных значений X, используя линейную зависимость между ними. Пример нахождения такой статистически усредненной линии можно увидеть на примере оценки данных о разрешениях на строительство в Сан-Франциско (Рис. 9.1-7), где инфляция при помощи линейной регрессии рассчитывалась для различных типов объектов.

Давайте загрузим таблицу данных проекта (Рис. 9.3-2 из предыдущей главы) непосредственно в LLM и попросим его построить для нас простую модель машинного обучения.

- Отправьте текстовый запрос в LLM чат (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Необходимо показать построение простой модели машинного обучения для прогнозирования стоимости и времени реализации нового проекта X (Рис. 9.3-2 как прикрепленное изображение) ⏎

Ответ LLM:

Рис. 9.3-4 LLM выбрал линейную регрессию для создания модели машинного обучения для прогнозирования стоимости и времени проекта.

LLM автоматически распознал таблицу из приложенного изображения и преобразовал данные из визуального формата в массив таблицы (Рис. 9.3-4 – 6-я строка). Этот массив был использован в качестве основы для создания признаков и меток, на основе которых была создана модель машинного обучения (Рис. 9.3-4 – 17-22-я строка), в которой использовалась линейная регрессия.

С помощью базовой линейной регрессионной модели, которая была обучена на "чрезвычайно небольшом" наборе данных, были сделаны прогнозы для нового гипотетического проекта строитель-

ства, обозначенного как Project X. В нашей задаче этот проект характеризуется наличием 40 квартир, 4 этажей и уровнем сложности 7 (Рис. 9.3-2).

Согласно прогнозам, сделанным с помощью линейной регрессионной модели, основанной на ограниченном и небольшом наборе данных для нового Project X (Рис. 9.3-4 - 24-29-я строка):

- **Продолжительность строительства** составит примерно 238 дней (238,4444444)
- **Общая сумма расходов** составит приблизительно $3 042 338 (3042337,777)

Для дальнейшего изучения гипотезы о стоимости проекта полезно поэкспериментировать с различными алгоритмами и методами машинного обучения. Поэтому спрогнозируем те же значения стоимости и времени для нового проекта X на основе небольшого набора исторических данных с помощью алгоритма K-Nearest Neighbours (k-NN).

Прогнозы стоимости и времени проекта при помощи алгоритма K-nearest neighbor (k-NN)

В качестве дополнительного прогноза для оценки стоимости и продолжительности нового проекта используем алгоритм k-Nearest Neighbours (k-NN). Алгоритм K-Nearest Neighbors (k-NN) — это метод машинного обучения под наблюдением (supervised machine learning), применяемый как для классификации, так и для регрессии. Также алгоритм k-NN уже рассматривался нами ранее в контексте поиска по векторным базам данных (Рис. 8.2-2), где он используется для нахождения наиболее близких векторов (например, текстов, изображений или технических описаний). В этом подходе каждый проект представляется как точка в многомерном пространстве, где каждое измерение соответствует определенному атрибуту проекта.

В нашем случае, учитывая три атрибута каждого проекта, мы представим их как точки в трехмерном пространстве (Рис. 9.3-5). Таким образом, наш предстоящий проект X будет локализован в этом пространстве с координатами (x=4, y=4, z=7). Следует отметить, что в реальных условиях количество точек и размерность пространства могут быть на порядки больше.

> Алгоритм K-NN (k-nearest neighbors) работает путем измерения расстояния между желаемым проектом X и проектами в обучающей базе данных. Сравнивая эти расстояния, алгоритм определяет проекты, которые находятся ближе всего к точке нового проекта X.

Например, если второй проект (x=8, y=9, z=6)из нашего исходного датасета расположен значительно дальше от X (Рис. 9.3-5), чем другие проекты, его можно исключить из дальнейшего анализа. В результате для расчётов можно использовать только два (k=2) ближайших проекта, на основе которых и будет определяться среднее значение.

Подобный метод, через поиск соседей, позволяет оценить сходство между проектами, что, в свою очередь, помогает сделать выводы о возможной стоимости и сроках реализации нового проекта на основе аналогичных проектов, которые были реализованы ранее.

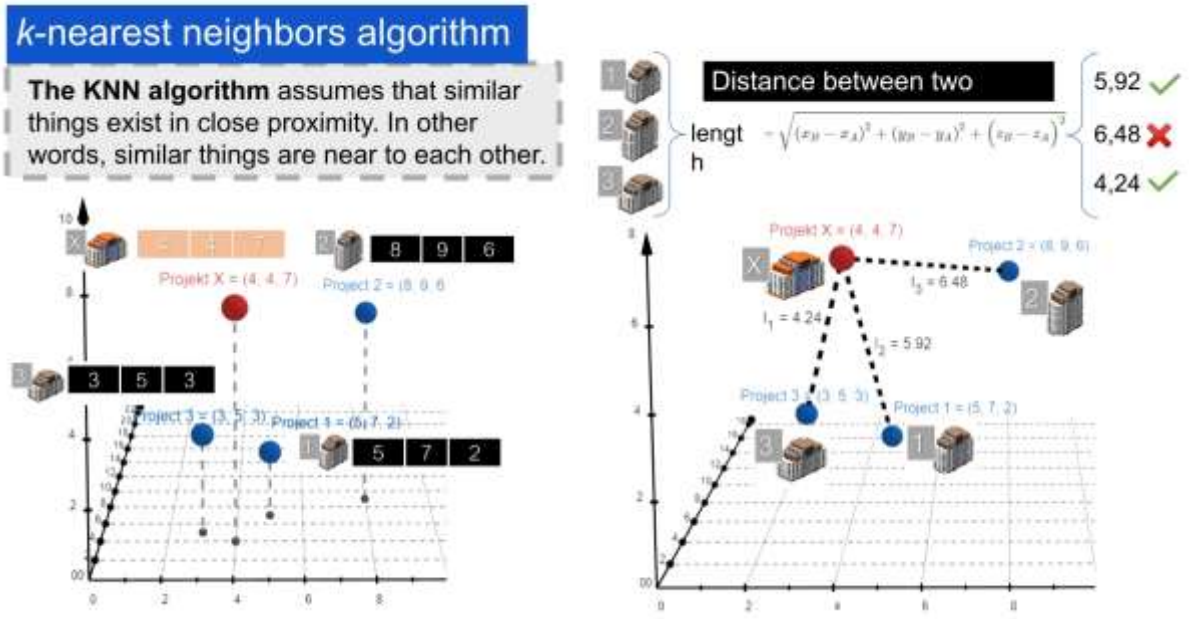


Рис. 9.3-5 В алгоритме K-NN проекты представлены как точки в многомерном пространстве, и для оценки сходства и прогнозирования выбираются ближайшие проекты на основе расстояний.

Работа k-NN включает в себя несколько ключевых этапов:

- **Подготовка данных:** сначала загружаются обучающие и тестовые наборы данных. Обучающие данные используются для "тренировки" алгоритма, а тестовые - для проверки его эффективности.
- **Выбор параметра K:** выбирается число K, которое указывает, сколько ближайших соседей (точек данных) должно учитываться в алгоритме. Значение “K” очень важно, поскольку оно влияет на результат.
- Процесс классификации и регрессии для тестовых данных:
  - **Вычисление расстояний:** для каждого элемента из тестовых данных вычисляется расстояние до каждого элемента из обучающих данных (Рис. 9.3-5). Для этого могут использоваться различные методы измерения расстояний, такие как евклидово расстояние (наиболее распространенный метод), манхэттенское расстояние или расстояние Хэмминга.
  - **Сортировка и выбор K ближайших соседей:** после вычисления расстояний они сортируются и выбираются K ближайших точек к тестовой точке.
  - **Определение класса или значения тестовой точки:** если это задача классификации, класс тестовой точки определяется на основе наиболее часто встречающегося класса среди K выбранных соседей. Если это задача регрессии, то вычисляется среднее значение (или другая мера центральной тенденции) значений K соседей.
- **Завершение процесса:** как только все тестовые данные будут классифицированы или для них будут сделаны прогнозы, процесс будет завершен.

Алгоритм k-nearest neighbors (k-NN) эффективен во многих практических приложениях и является одним из основных инструментов в арсенале специалистов по машинному обучению. Этот алгоритм популярен благодаря своей простоте и эффективности, особенно в задачах, где взаимосвязи между данными легко интерпретировать.

В нашем примере после применения алгоритма K-ближайших соседей были определены два проекта (из нашей небольшой выборки) с наименьшим расстоянием до проекта X (Рис. 9.3-5). На основе этих проектов алгоритм определяет среднее значение их цены и продолжительности строительства. После анализа (Рис. 9.3-6) алгоритм, путём усреднения ближайших соседей, приходит к выводу, что проект X будет стоить примерно $ 3 800 000 долларов и займет около 250 дней.

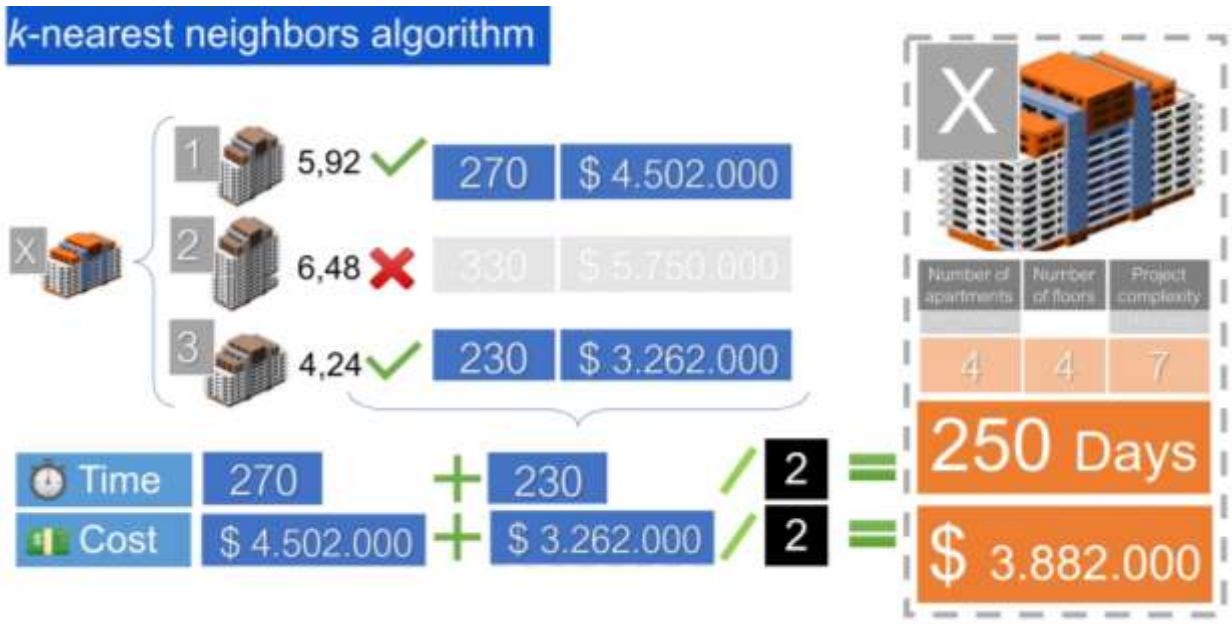


Рис. 9.3-6 Алгоритм K-nearest neighbors определяет стоимость и график проекта X, анализируя два наиболее близких проекта в выборке.

Алгоритм k-Nearest Neighbors (k-NN) особенно популярен в задачах классификации и регрессии, например, в рекомендательных системах, где он используется для предложения товаров или контента на основе предпочтений, схожих с интересами конкретного пользователя. Кроме того, k-NN широко используется в медицинской диагностике для классификации типов заболеваний на основе симптомов пациента, в распознавании образов и в финансовом секторе для оценки кредитоспособности клиентов.

Даже при наличии ограниченного объема данных модели машинного обучения могут дать полезные прогнозы и существенно усилить аналитическую составляющую в управлении строительными проектами. При расширении и очистке исторических данных возможен переход к более сложным моделям — например, с учётом типа конструкций, местоположения, сезона начала строительства и других факторов.

В нашей упрощенной задаче для визуализации в трехмерном пространстве использовалось три атрибута, но реальные проекты, в среднем, включают сотни или тысячи атрибутов (см. набор данных из главы "Пример больших данных на основе данных CAD (BIM)"), что значительно увеличивает размерность пространства и сложность представления проектов в виде векторов (Рис. 9.3-7).

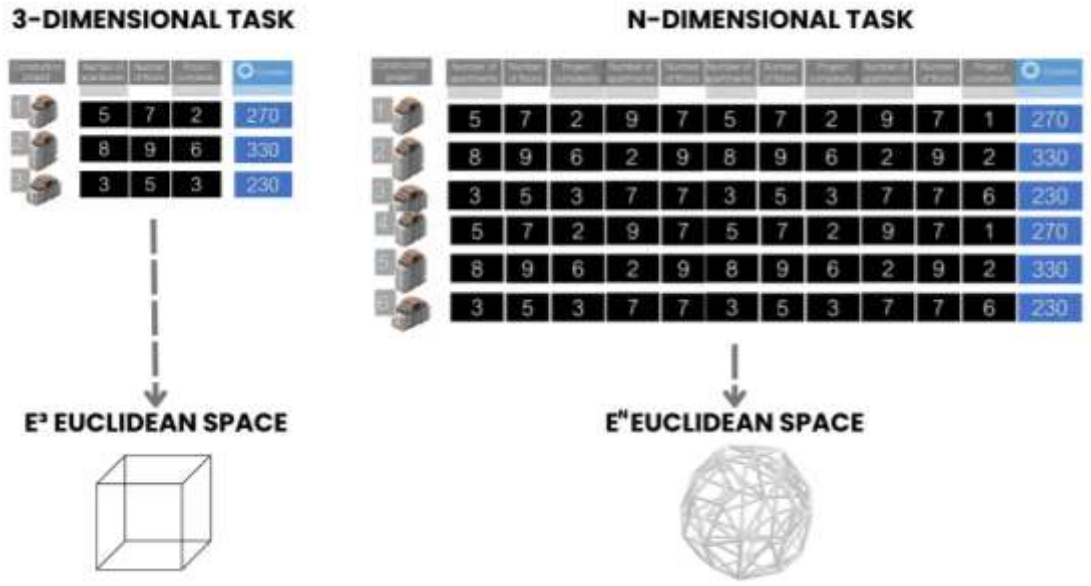


Рис. 9.3-7 В упрощенном примере для 3D-визуализации использовались три атрибута, в то время как реальные проекты имеют большее количество.

Применение различных алгоритмов к одному и тому же набору данных для проекта X, в котором 40 квартир, 4 этажа и уровень сложности 7, дало разные прогнозные значения. Алгоритм линейной регрессии предсказал время завершения 238 дней и стоимость $3 042 338 (Рис. 9.3-4), в то время как алгоритм k-NN предсказал 250 дней и $3 882 000 (Рис. 9.3-6).

Точность предсказаний, получаемых с помощью моделей машинного обучения, напрямую зависит от объёма и качества исходных данных. Чем больше проектов участвует в обучении, и чем более полно и точно представлены их характеристики (признаки) и результаты (метки), тем выше вероятность получения достоверных прогнозов с минимальными значениями погрешности.

Важную роль в этом процессе играют методы предварительной обработки данных, включая:

- Нормализацию, позволяющую привести признаки к единому масштабу;
- Обнаружение и устранение выбросов, исключающее искажение модели;
- Кодирование категориальных признаков, позволяющее работать с текстовыми данными;
- Заполнение пропущенных значений, повышающее устойчивость модели.

Кроме того, для оценки обобщающей способности модели и её устойчивости к новым наборам данных используются методы кросс-валидации, позволяющие выявить переобучение и повысить надёжность прогноза.

> *Хаос — это порядок, который нужно расшифровать [160].*
>
> — Жозе Сарамаго, «Двойник»

Даже если вам кажется, что хаос ваших задач невозможно описать формально, знайте — любое событие в мире и особенно строительные процессы подчиняются математическим закономерностям, которым может понадобиться поддержка расчёта значений не через строгие формулы а при помощи статистики и исторических данных.

Как традиционные сметные расчёты, выполняемые калькуляционными отделами, так и модели машинного обучения неизбежно сталкиваются с неопределённостью и потенциальными источниками ошибок. Однако при наличии достаточного объема качественных данных, модели машинного обучения могут продемонстрировать сопоставимую, а иногда и более высокую точность прогнозов по сравнению с экспертными оценками.

> Машинное обучение, вероятнее всего, станет надёжным дополняющим инструментом анализа, позволяющим: уточнять расчёты, предлагать альтернативные сценарии и выявлять скрытые зависимости между параметрами проекта. Подобные модели не будут претендовать на универсальность, но уже вполне скоро смогут занять важное место в расчётах и процессах принятия проектных решений. Технологии машинного обучения не исключат участие инженеров, сметчиков и аналитиков, а, напротив, расширят их возможности, предлагая дополнительную точку зрения, основанную на исторических данных.

При грамотной интеграции в бизнес-процессы строительных компаний машинное обучение имеет потенциал стать важным элементом в системе поддержки управленческих решений — не как замена человека, а как расширение его профессиональной интуиции и инженерной логики.

Дальнейшие шаги: от хранения к анализу и прогнозированию

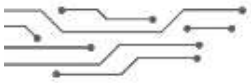
Современные подходы к работе с данными начинают менять принципы принятия решений в строительной отрасли. Переход от интуитивных оценок к объективному анализу данных не просто повышает точность, но и открывает новые возможности для оптимизации процессов. Подводя итог данной части, стоит выделить основные практические шаги, которые помогут применить рассмотренные методы в ваших повседневных задачах:

- Создание устойчивой инфраструктуры хранения данных
  - Попробуйте объединить разрозненные документы и проектные данные в единую табличную модель, агрегируя ключевую информацию в одном датафрейме для дальнейшего анализа
  - Используйте эффективные форматы хранения данных — например, колоночные форматы вроде Apache Parquet вместо CSV или XLSX — особенно для тех наборов, которые в будущем, потенциально, могут использоваться для обучения моделей машинного обучения

- [ ] Создайте систему версионирования данных, позволяющую отслеживать изменения на протяжении всего проекта

- Внедрение инструментов анализа и автоматизации
  - [ ] Начните анализировать исторические данные проектов — по документации, моделям, сметам — для выявления закономерностей, трендов и аномалий
  - [ ] Освойте ETL-процессы (Extract, Transform, Load) для автоматической загрузки и подготовки данных
  - [ ] Научитесь визуализировать ключевые метрики при помощи различных бесплатных библиотек Python для визуализации
  - [ ] Начните применять статистические методы и случайные выборки, чтобы получать репрезентативные и воспроизводимые аналитические выводы
- Повышение зрелости в работе с данными
  - [ ] Изучите несколько базовых алгоритмов машинного обучения на простых и понятных примерах, вроде датасета Титаник
  - [ ] Проанализируйте текущие процессы и определите, где можно перейти от жёсткой причинно-следственной логики к статистическим методам прогнозирования и оценки
  - [ ] Начните рассматривать данные как стратегический актив, а не побочный продукт: выстраивайте процессы принятия решений на основе моделей данных, а не вокруг конкретных программных решений

Строительные компании, осознавшие ценность данных, вступают в новую фазу развития, где конкурентное преимущество определяется не объёмом ресурсов, а скоростью принятия решений на основе аналитики.



МАКСИМУМ УДОБСТВА С ПЕЧАТНОЙ ВЕРСИЕЙ

Вы держите в руках бесплатную цифровую версию **Data-Driven Construction**. Для более удобной работы и быстрого доступа к материалам рекомендуем обратить внимание на **печатное издание**:

- **Всегда под рукой**: книга в печатном формате станет надежным рабочим инструментом, позволяя быстро найти и использовать нужные визуализации и схемы в любых рабочих ситуациях
- **Высокое качество иллюстраций**: все изображения и графики в печатном издании представлены в максимальном качестве
- **Быстрый доступ к информации**: удобная навигация, возможность делать пометки, закладки и работать с книгой в любом месте.

Приобретая полную печатную версию книги, вы получаете удобный инструмент для комфортной и эффективной работы с информацией: возможность оперативно использовать визуальные материалы в повседневных задачах, быстро находить нужные схемы и делать пометки. Кроме того, ваша покупка поддерживает распространение открытых знаний.

Заказать печатную версию книги можно на: datadrivenconstruction.io/books

X ЧАСТЬ

СТРОИТЕЛЬНАЯ ОТРАСЛЬ В ЭПОХУ ЦИФРОВЫХ ДАННЫХ. ВОЗМОЖНОСТИ И ВЫЗОВЫ

Заключительная десятая часть представляет собой комплексный взгляд на будущее строительной отрасли в эпоху цифровой трансформации. Здесь анализируется переход от причинно-следственного анализа к работе с корреляциями больших данных. Проводятся параллели между эволюцией изобразительного искусства и развитием работы с данными в строительстве, демонстрируя, как отрасль движется от детализированного контроля к целостному пониманию процессов. Рассматривается концепция "уберизации" строительной отрасли, где прозрачность данных и автоматизация расчетов могут радикально изменить традиционные бизнес-модели, устраняя необходимость в посредниках и снижая возможности для спекуляций. Подробно обсуждаются нерешенные проблемы, например универсальная классификация элементов, которые дают строительным компаниям время для адаптации к новым условиям. Часть завершается конкретными рекомендациями по формированию стратегии цифровой трансформации, включающей анализ уязвимостей и расширение спектра услуг для сохранения конкурентоспособности в меняющейся отрасли.

ГЛАВА 10.1.
СТРАТЕГИИ ВЫЖИВАНИЯ: ФОРМИРОВАНИЕ КОНКУРЕНТНЫХ ПРЕИМУЩЕСТВ

Корреляции вместо расчетов: будущее строительной аналитики

Из-за стремительной дигитализации информации (Рис. 1.1-5) современное строительство переживает фундаментальную трансформацию, где данные становятся не просто инструментом, а стратегическим активом, способным кардинально изменить традиционные подходы к управлению проектами и бизнесом.

На протяжении тысячелетий строительная деятельность опиралась на детерминированные методы — точные расчёты, детализацию и строгий контроль параметров. В первые века нашей эры римские инженеры применяли математические принципы для возведения акведуков и мостов. В средние века архитекторы стремились к идеальным пропорциям готических соборов, а в эпоху индустриализации XX века были сформированы системы стандартизированных нормативов и регламентов, ставшие основой массового строительства.

Сегодня вектор развития смещается от поиска строго причинно-следственных связей к вероятностному анализу, поиску корреляций и скрытых закономерностей. Отрасль вступает в новую фазу — данные становятся ключевым ресурсом, а аналитика на их основе вытесняет интуитивные и локально оптимизированные подходы.

Рис. 10.1-1 Скрытый потенциал строительных данных: существующие расчёты в компании - лишь вершина айсберга, доступная менеджменту для анализа.

Информационная система компании похожа на айсберг (Рис. 10.1-1): менеджменту компании видна лишь малая часть потенциала данных, тогда как основная ценность скрыта в глубине. Важно оценивать данные не только по их текущему использованию, но и по возможностям, которые они откроют в будущем. Именно те компании, которые научатся извлекать скрытые закономерности и создавать из данных новые знания, смогут формировать устойчивое конкурентное преимущество.

Поиск скрытых закономерностей и осмысление данных — это не просто работа с числами, а творческий процесс, требующий абстрактного мышления и способности видеть за разрозненными элементами целостную картину. В этом смысле развитие работы с данными можно сравнить с эволюцией изобразительного искусства (Рис. 10.1-2).

Развитие строительства удивительно напоминает прогресс изобразительного искусства. В обоих случаях человечество прошло путь от примитивных методов к сложным технологиям визуализации и анализа. В доисторические времена люди использовали наскальные рисунки и примитивные инструменты для решения повседневных задач. В Средние века и эпоху Ренессанса уровень сложности в архитектуре и искусстве значительно вырос. К началу Средних веков инструменты строительства эволюционировали от простого топора до обширных наборов инструментов, символизирующих рост технических знаний.

Эпоха реализма стала первой революцией в изобразительном искусстве: художники научились воспроизводить мельчайшие детали, добиваясь максимальной правдоподобности изображения. В строительстве аналогом этого периода стали точные инженерные методики, детализированные чертежи и строго регламентированные расчёты, ставшие основой проектной практики на протяжении столетий.

Позднее, импрессионизм изменил само восприятие художественной реальности: вместо буквальной передачи формы художники начали фиксировать настроение, свет и динамику, стремясь отразить общее впечатление, а не абсолютную точность. Подобным образом машинное обучение в строительной аналитике уходит от жёстких логических моделей к распознаванию паттернов и вероятностных закономерностей, позволяющим "увидеть" скрытые зависимости в данных, недоступные при классическом анализе. С этим подходом перекликаются идеи минимализма и функциональности Bauhaus, где смысл (функция) важнее формы. Bauhaus стремился убрать лишнее, отказаться от орнамента ради ясности, утилитарности и массовости. Вещи должны были быть понятны и полезны, без излишеств — эстетика рождалась из логики конструкции и предназначения.

С появлением фотографии в конце 19 века искусство получило новый инструмент для фиксации реальности с небывалой точностью и перевернуло отношение к изобразительному искусству. Аналогично, в строительстве промышленная революция в 21 веке приводит к применению роботизированных технологий, лазеров, IoT, RFID и концептов вроде „Connected Construction“, где сбор отдельных параметров эволюционировал к масштабируемой интеллектуальной фиксации полной реальности строительной площадки.

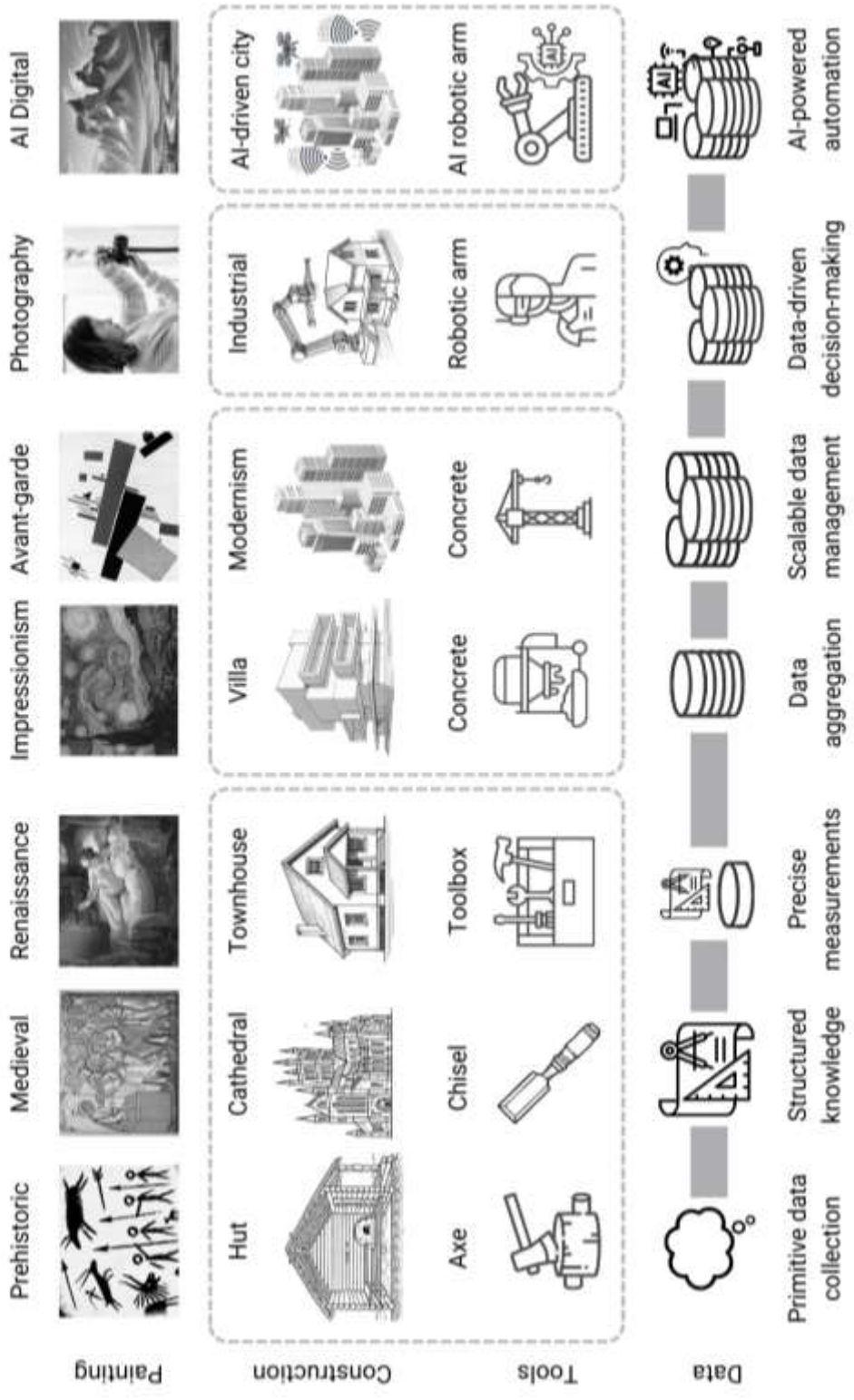


Рис. 10.1-2 Эпохи эволюций изобразительного искусства согласуется с развитием в подходах к работе с данными в строительной отрасли.

Сегодня, точно так же как изобразительное искусство переживает переосмысление с приходом инструментов AI и LLM, строительная отрасль переживает очередной квантовый скачок: интеллектуальные системы, управляемые искусственным интеллектом (ИИ), LLM-чаты позволяют предсказывать, оптимизировать и генерировать решения с минимальным вмешательством человека.

Роль данных в проектировании и управлении изменилась радикально. Если раньше знания передавались устно и имели эмпирический характер — подобно тому, как до XIX века реальность фиксировалась вручную написанными картинами, — то сегодня в центре внимания оказывается полная цифровая фиксация строительной «картины». С помощью алгоритмов машинного обучения эта цифровая картина трансформируется в импрессионистское представление строительной реальности — не точную копию, а обобщённое, вероятностное понимание процессов.

> Мы стремительно приближаемся к эпохе, в которой процессы проектирования, строительства и эксплуатации зданий будут не просто дополняться, но и в значительной степени управляться системами искусственного интеллекта. Подобно тому как современное цифровое искусство создаётся без кисти — с помощью текстовых промптов и генеративных моделей, — архитектурные и инженерные решения будущего будут формироваться на основе ключевых запросов и параметров, заданных пользователем.

В XXI веке доступ к данным, их интерпретация и качество аналитики становятся неотъемлемыми условиями успеха проекта. Причём ценность данных определяется не их объёмом, а способностью специалистов анализировать, проверять и превращать их в действия.

Data-driven подход в строительстве: инфраструктура нового уровня

В истории человечества каждый подобный технологический скачок приносил фундаментальные изменения в экономику и общество. Сегодня мы наблюдаем новую волну трансформации, сравнимую по масштабу с промышленной революцией XIX века. Однако если сто лет назад основным драйвером изменений были механические силы и энергетические технологии, то сейчас – это данные и искусственный интеллект.

Машинное обучение, LLM и ИИ агенты изменяют саму суть приложений, делая традиционные программные стеки (которые обсуждались во второй части книги) ненужными (Рис. 2.2-3). Вся логика работы с данными сосредотачивается в ИИ-агентах, а не в жестко закодированных бизнес-правилах (Рис. 2.2-4).

> В эпоху данных традиционные представления о приложениях кардинально трансформируются. Мы движемся к модели, где громоздкие корпоративные модульные системы неизбежно уступят место открытым легковесным, специализированным решениям.

> *В будущем останется только базовая структура данных, а все взаимодействие с ней будет происходить через агентов, работающих напрямую с базой данных. Я действительно верю, что весь стек приложений исчезнет, потому что в нем просто нет необходимости, когда искусственный интеллект напрямую взаимодействует с основной базой данных. Я всю свою карьеру работал в SaaS — создавал компании, работал в них, и, если честно, сейчас я бы, наверное, не стал запускать новый SaaS-бизнес. И, скорее всего, я бы не стал инвестировать в SaaS-компании прямо сейчас. Ситуация слишком неопределенная. Это не значит, что в будущем не будет софтверных компаний, просто они будут выглядеть совсем по-другому. Будущие системы будут представлять собой базы данных с бизнес-логикой, вынесенной в [ИИ] агентов. Эти агенты будут работать с несколькими репозиториями данных одновременно, не ограничиваясь одной базой. Вся логика переместится в уровень искусственного интеллекта [46].*
>
> — Matthew Berman, CEO Forward Future

Ключевое отличие новой парадигмы - минимизация технологического балласта. Вместо монументальных сложных и закрытых программных комплексов мы получим гибкие, открытые и быстро настраиваемые модули, которые буквально "живут" внутри потока данных (Рис. 7.4-1 – Apache Airflow, NiFi). Архитектура будущего управления процессов предполагает использование микро приложений - компактных, целевых инструментов, принципиально отличающиеся от массивных и закрытых ERP, PMIS, CDE, CAFM систем. Новые агенты будут максимально адаптивными, интегрированными и ориентированными на конкретные бизнес-задачи (Например Low-Code/No-Code Рис. 7.4-6).

> *Вся бизнес-логика уйдет к этим [ИИ] агентам, и эти агенты будут выполнять операции CRUD [Create, Read, Update, and Delete] в нескольких репозиториях, то есть они не будут различать, какой именно бэкенд используется. Они будут обновлять несколько баз данных, а вся логика окажется в так называемом AI-уровне. И как только AI-уровень станет местом, где находится вся логика, люди начнут заменять бэкенды. Мы уже наблюдаем довольно высокий процент побед на рынке бэкендов Dynamics и использования агентов, и мы будем активно двигаться в этом направлении, пытаясь объединить все это. Будь то в сфере обслуживания клиентов или в других областях, например, не только CRM, но и наши решения по финансам и операциям. Потому что люди хотят больше AI-ориентированных бизнес-приложений, где логический уровень может управляться ИИ и агентами ИИ. […]. Одна из самых захватывающих вещей для меня — это Excel с Python, который сравним с GitHub с Copilot. То есть, что мы сделали: теперь, когда у вас есть Excel, вам стоит просто открыть его, запустить Copilot и начать играть с ним. Это уже не просто понимание имеющихся чисел — он сам составит план. Как GitHub Copilot Workspace создает план, а затем его выполняет, так и это похоже на работу дата-аналитика, использующего Excel как инструмент визуализации строк и столбцов для анализа. Таким образом, Copilot использует Excel как инструмент со всеми его возможностями, потому что он может генерировать данные и обладает интерпретатором Python.*
>
> — Сатья Наделла, CEO, Microsoft, интервью каналу BG2 декабрь 2024 г. *[28]*

Трансформация, которую мы наблюдаем в логике офисных приложений — переход от модульных, закрытых систем к ИИ-агентам, работающим напрямую с открытыми данными, — является лишь частью гораздо более масштабного процесса. Речь идёт не просто о смене интерфейсов или программной архитектуры: изменения затронут фундаментальные принципы организации труда, принятия решений и управления бизнесом. В строительстве это приведёт к формированию data-driven логики, в которой данные станут центральным элементом процессов — от проектирования до управления ресурсами и контроля хода строительства.

Цифровой офис следующего поколения: как AI меняет рабочее пространство

Почти сто лет назад человечество уже переживало подобную технологическую революцию. Переход от паровых машин к электрическим двигателям занял более четырёх десятилетий, но в итоге стал катализатором беспрецедентного роста производительности — прежде всего за счёт децентрализации энергомощностей и гибкости новых решений. Этот сдвиг не только изменил ход истории, переместив основную массу населения из деревень в города, но и заложил фундамент современной экономики. История технологий — это путь от физического труда к автоматизации и интеллектуальным системам. Так же, как трактор заменил десятки землепашцев, современные цифровые технологии вытесняют традиционные офисные методы управления строительством (Рис. 10.1-3). Ещё в начале XX века большая часть населения Земли обрабатывала землю вручную, пока в 1930-х не началась механизация труда с помощью машин и тракторов.

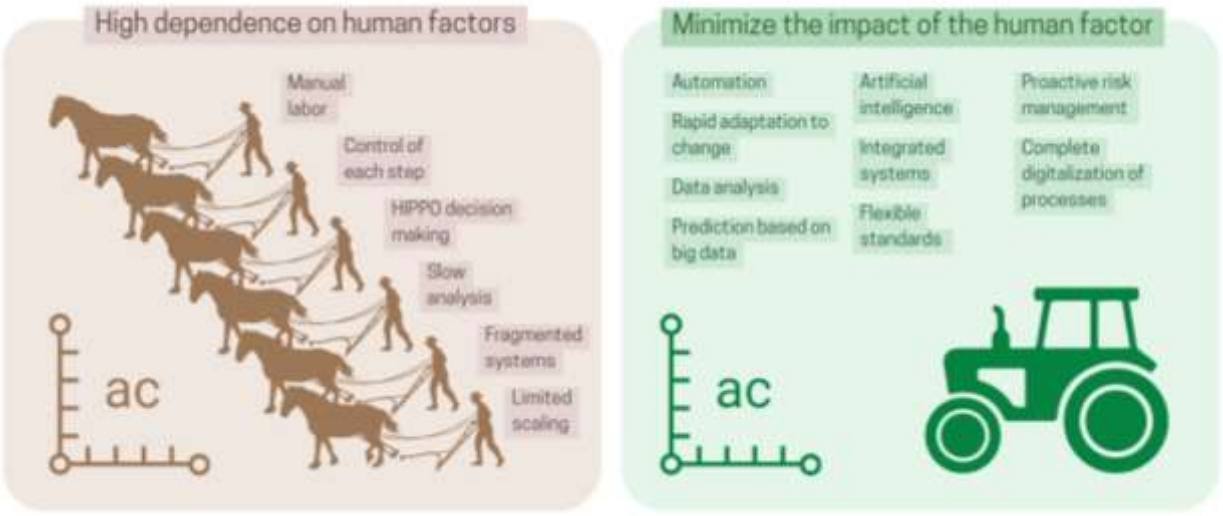


Рис. 10.1-3 Как трактор заменил десятки людей в начале 20 века, так и машинное обучение заменит традиционные методы управления бизнесом и проектами в 21 веке.

Точно так же, как человечество сто лет назад перешло от обработки отдельных участков земли примитивными инструментами к масштабному сельскому хозяйству с использованием техники, сегодня мы совершаем переход от обработки разрозненных «силосов» информации к работе с массивами данных с помощью мощных «тракторов» — ETL-pipeline и алгоритмов искусственного интеллекта.

Мы стоим на пороге аналогичного скачка — но уже в цифровой плоскости: от традиционного, ручного управления бизнесом к моделям, основанным на данных.

Путь к полноценной data-driven архитектуре потребует времени, инвестиций и организационных усилий. Но этот путь открывает путь к не просто постепенному улучшению, а качественному рывку — в сторону большей эффективности, прозрачности и управляемости строительных процессов. Всё это — при условии системного внедрения цифровых инструментов и отказа от устаревших бизнес-практик.

> Параметризация задач, ETL, LLM, компоненты IoT, RFID, токенизация, большие данные и машинное обучение превратят традиционное строительство в **data-driven construction**, где каждая деталь проекта и строительного бизнеса будет контролироваться и оптимизироваться с помощью данных.

Раньше для анализа информации требовались тысячи человеко-часов. Теперь эти задачи выполняют алгоритмы и LLM, превращающие разрозненные массивы данных при помощи промптов в стратегические источники. В технологическом мире происходит то же, что случилось с сельским хозяйством: от мотыги мы переходим к автоматизированному агрокомплексу. Так и офисная работа в строительстве — от Excel-файлов и ручной сводки — переходит к интеллектуальной системе, где данные собираются, очищаются, структурируются и превращаются в инсайты.

Уже сегодня компании должны начать «культивировать» информационные поля с помощью качественного сбора данных и структурирования информации, и «удобрять» их инструментами очистки и нормализации, а затем «собирать урожай» — в виде предиктивной аналитики и автоматизированных решений. Если современный фермер с машиной способен заменить сотню землепашцев, то и интеллектуальные алгоритмы смогут снять с сотрудников рутину и перевести их к роли стратегических менеджеров информационных потоков.

Однако важно понимать, что создание по-настоящему data-driven-организации – это не быстрый процесс. Это долгосрочное стратегическое направление, подобное созданию нового участка для выращивания нового леса (Рис. 1.2-5) систем, где каждое "дерево" в этой экосистеме– это отдельный процесс, компетенция или инструмент, который требует времени для роста и развития. И как в случае с настоящим лесом, успех зависит не только от качества посадочного материала (технологий), но и от почвы (корпоративной культуры), климата (деловой среды) и ухода (системного подхода).

Компании больше не смогут полагаться исключительно на закрытые решения «из коробки». В отличие от предыдущих этапов технологического развития, текущий переход — к открытому доступу к данным, использованию искусственного интеллекта и распространению Open Source — вряд ли получит поддержку со стороны крупных вендоров, поскольку он напрямую угрожает их устоявшимся бизнес-моделям и основным источникам дохода.

Как показало исследование Гарвардской бизнес-школы [40], которое уже обсуждалось в главе о четвёртой и пятой технологических революциях, затраты на создание наиболее используемых Open Source-решений с нуля для всех компаний составили бы порядка 4,15 миллиарда долларов. Однако если представить, что каждая компания будет разрабатывать собственные альтернативы без доступа к уже существующим Open Source-инструментам, что и происходило последние десятилетия, совокупные издержки бизнеса могут достичь колоссальных 8,8 триллиона долларов — это

цена нерационального спроса, которым может оцениваться рынок ПО.

Технологический прогресс неизбежно приведёт к пересмотру устоявшихся бизнес-моделей. Если раньше компании могли зарабатывать на сложных, непрозрачных процессах и закрытых данных, то с развитием ИИ и аналитики этот подход становится всё менее жизнеспособным.

> В результате демократизация доступа к данным и инструментам традиционный рынок продаж программного обеспечения может значительно сократиться. Однако одновременно с этим вырастет новый рынок — рынок цифровой экспертизы, кастомизации, интеграции и проектирования решений. Здесь ценность будет формироваться не за счёт продажи лицензий, а за счёт способности выстраивать гибкие, открытые и адаптируемые цифровые процессы. Подобно тому, как электрификация и появление тракторов породили новые отрасли промышленности, так же и применение больших данных, AI и LLM открывает совершенно новые горизонты для бизнеса в строительной отрасли, который потребует не только технологических инвестиций, но и глубокой трансформации мышления, процессов и организационных структур. И те компании и специалисты, которые поймут это, и начнут действовать уже сегодня, будут лидерами завтрашнего дня.

В мире, где открытые данные становятся главным активом, доступность информации поменяет правила игры. Инвесторы, заказчики и регуляторы всё чаще будут требовать прозрачности, а алгоритмы машинного обучения будут способны автоматически выявлять несоответствия в сметах, сроках и расходах. Это создаёт условия для нового этапа цифровой трансформации, который постепенно приводит нас к "уберизации" строительной отрасли.

Открытые данные и уберизация — это угроза для существующего строительного бизнеса

Строительство превращается в процесс управления информацией. Чем точнее, качественней и полнее данные, тем эффективнее проектирование, расчёты, калькуляции смет, возведение и эксплуатация зданий. В будущем ключевым ресурсом станет не наличие крана, бетона и арматуры, а способность собирать, анализировать и использовать информацию.

> Клиенты строительных компаний - инвесторы и заказчики, финансирующие строительство, в будущем неизбежно будут использовать ценность открытых данных и аналитики исторических данных. Это откроет возможности для автоматизации расчётов сроков и стоимости проектов, без привлечения строительных компаний к вопросам калькуляций, что позволит контролировать расходы и быстрее выявлять избыточные затраты.

Представьте строительную площадку где лазерные сканеры, квадрокоптеры и системы фотограмметрии в реальном времени собирают точные данные об объёмах используемого бетона. Эта информация автоматически преобразуется в простые плоские MESH-модели с метаданными, минуя громоздкие CAD (BIM) системы, без зависимостей от сложных геометрических ядер, ERP или PMIS.

Эти, собранные со строительной площадки, данные централизованно переносятся в единые структурированные хранилища, доступные заказчику для независимого анализа, куда подгружаются реальные цены из разных строительных магазинов и например разные параметры – от ставки кредитного финансирования до динамически меняющихся факторов, таких как погодные условия, биржевые котировки строительных материалов, логистические тарифы и статистические сезонные колебания цен на рабочую силу. В таких условиях любые расхождения между проектными и фактическими объёмами материалов становятся моментально очевидными, делая невозможными манипуляции со сметными расчётами как на этапе проектирования, так и при сдаче объекта. В итоге прозрачность процесса строительства достигается не за счёт армии контролёров и менеджеров, а благодаря объективным цифровым данным, в которых будет сведена до минимума человеческий фактор и возможность спекуляций.

Подобной работой по контролю данных будут в будущем заниматься скорее менеджеры по данным со стороны заказчика (Рис. 1.2-4 CQMS менеджер). Особенно это касается калькуляций и сметных расчётов проектов: там, где раньше работал целый отдел сметчиков, уже завтра появятся инструменты машинного обучения и прогнозирования, которые будут устанавливать строительным компаниям ценовые рамки, в которые им необходимо вписываться.

> *Учитывая фрагментированный характер [строительной] отрасли, когда большая часть систем и подсистем поставляется малыми и средними предприятиями, цифровая стратегия должна исходить от клиента. Клиенты должны создать условия и механизмы для раскрытия цифровых возможностей цепочки поставок [20].*
>
> — Эндрю Дэвис и Джулиано Деникол, Accenture „Создание большей ценности с помощью капитальных проектов“

Подобная открытость и прозрачность данных представляет угрозу для строительных компаний, которые привыкли зарабатывать на непрозрачности процессов и запутанных отчётах, где можно спрятать спекуляции и скрытые издержки за сложными и закрытыми форматами и модульными проприетарными платформами передачи данных. Поэтому строительные компании, как и в случае продвижения Open Source решений вендорами, вряд ли будут заинтересованы в полноценном внедрении открытых данных в свои бизнес-процессы. Если данные будут доступны и легко обрабатываемы для заказчика, их можно будет проверять автоматически, что исключит возможность завышать объёмы и манипулировать сметами.

> Согласно докладу World Economic Forum «Формирование будущего строительства» (2016) [5], одной из ключевых проблем отрасли остаётся пассивная роль заказчика. Тем не менее, именно заказчики должны брать на себя большую ответственность за исход проектов — начиная с раннего планирования, выбора устойчивых моделей взаимодействия и заканчивая контролем исполнения. Без активного участия со стороны владельцев проектов системная трансформация строительной отрасли невозможна.

Потеря контроля над расчётами объёмов и стоимости уже обернулась трансформацией за последние 20 лет в других отраслях экономики, позволяя клиентам напрямую, без посредников, до-

стигать своих целей. Цифровизация и прозрачность данных трансформировали многие традиционные бизнес-модели, как это произошло с таксистами после появления Uber (Рис. 10.1-4), с отельерами после прихода Airbnb и с розничными продавцами и магазинами после развития Amazon, а также с банками — благодаря росту необанков и децентрализированных финтех-экосистем, где прямой доступ к информации и автоматизация расчетов времени и стоимости существенно снизили роль посредников.

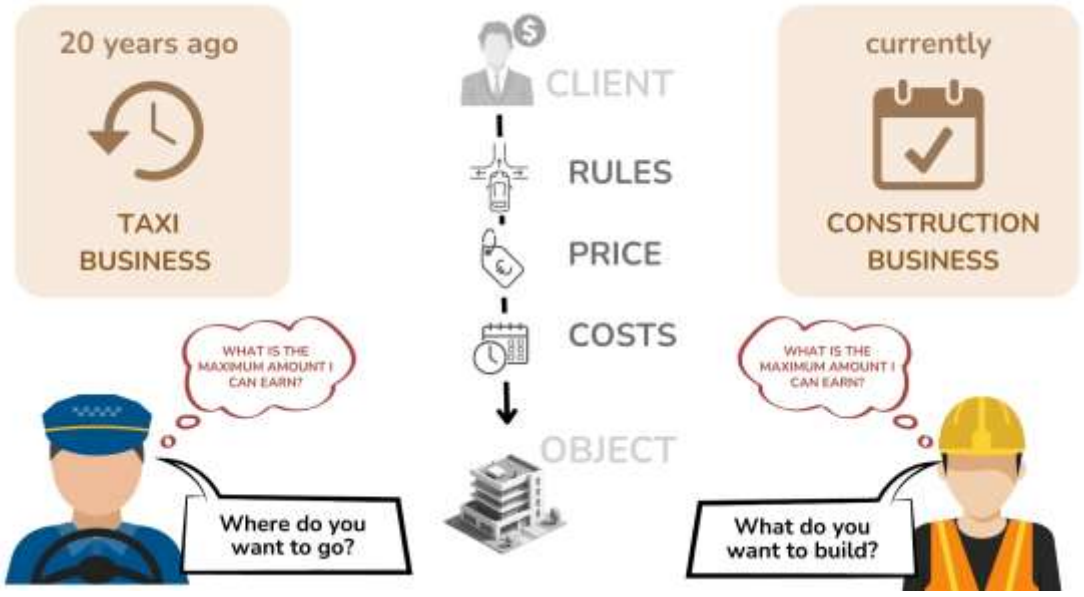


Рис. 10.1-4 Строительный бизнес столкнётся с уберизацией, с которой пришлось столкнуться таксистам, отельерам и продавцам 10 лет назад.

Процесс демократизации доступа к данным и инструментам их обработки неизбежен, и со временем открытые данные по всем составляющим проекта станут требованием клиента и новым стандартом. Поэтому вопросы внедрения открытых форматов и прозрачных расчётов будут продвигаться именно со стороны инвесторов, заказчиков, банков и частных инвестиционных фондов (private equity) — тех, кто в итоге является конечным пользователем построенных объектов и потом эксплуатирует объект десятки лет.

> Крупные инвесторы, заказчики и банки уже требуют прозрачности в строительной отрасли. Согласно исследованию Accenture «Создание большей ценности с помощью капитальных проектов» (2020) [20], прозрачные и надёжные данные становятся решающим фактором для инвестиционных решений в строительстве. Как отмечают эксперты, доверительное и эффективное управление проектами невозможно без прозрачности, особенно в условиях кризисов. Кроме того, владельцы активов и подрядчики всё чаще переходят к контрактам, стимулирующим обмен данными и совместную аналитику, что отражает растущие требования со стороны инвесторов, банков и регуляторов к ответственности и прозрачности.

> Движение инвестора, заказчика от идеи до готового здания, в будущем будет сродни путешествию на автопилоте - без водителя в виде строительной компании, обещает стать независимым от спекуляций и неопределенности.

Эра открытых данных и автоматизации неизбежно изменит строительный бизнес так же, как это уже произошло в банковском деле, торговле, сельском хозяйстве и логистике. В этих отраслях роль посредников и традиционные способы ведения бизнеса уступают место автоматизации и роботизации, не оставляя места для неоправданных наценок и спекуляций.

Данные и процессы во всех видах экономической деятельности человека ничем не отличаются от того, с чем приходится иметь дело профессионалам в строительной отрасли. В долгосрочной перспективе строительные компании, которые сегодня доминируют на рынке, устанавливая стандарты цен и качества услуг, могут потерять свою роль ключевого посредника между заказчиком и его строительным проектом.

Нерешённые проблемы уберизации как последний шанс использовать время для трансформации

Но вернёмся к реалиям строительной отрасли. Пока в одних секторах экономики появляются самоуправляемые автомобили, децентрализованные финансовые системы и решения на базе искусственного интеллекта, значительная часть строительных компаний по-прежнему остаётся бумажными организациями, в которых ключевые решения принимаются скорее на основе интуиции и опыта отдельных специалистов.

В этой парадигме современную строительную компанию можно сравнить с таксопарком 20-летней давности, который контролирует ресурсы, маршруты и время доставки, несёт ответственность за сроки и стоимость «поездки» — от проектной идеи (процесса логистика и монтажа) до сдачи объекта. Как когда-то GPS (в строительстве IoT, RFID) и алгоритмы машинного обучения в расчётах калькуляций времени/стоимости изменили сферу транспорта, так и данные, алгоритмы и ИИ-агенты способны преобразить управление строительством — от интуитивных оценок к предиктивной, управляемой модели. За последние 20 лет во многих отраслях — финансах, сельском хозяйстве, рознице и логистике — постепенно исчезала возможность спекулировать за счёт непрозрачности данных. Цены, стоимость доставки или финансовой транзакции рассчитываются автоматически и статистически обоснованно — всего за несколько секунд на цифровых платформах.

Заглядывая в будущее, строительные компании должны осознать, что демократизация доступа к данным и инструментам их анализа нарушат традиционный подход к оценке стоимости и сроков реализации проектов и лишат возможности спекулировать на непрозрачных данных об объемах и ценах.

Подобно движению по регулируемой дороге без вмешательства водителя, строительные процессы будущего всё больше будут напоминать "уберизированную" систему — с автоматизированной оценкой сроков и стоимости, прозрачной маршрутизацией задач и минимальной зависимостью от человеческого фактора. Это изменит саму природу "путешествия" от идеи до реализации — сделав его более предсказуемым, управляемым и ориентированным на данные.

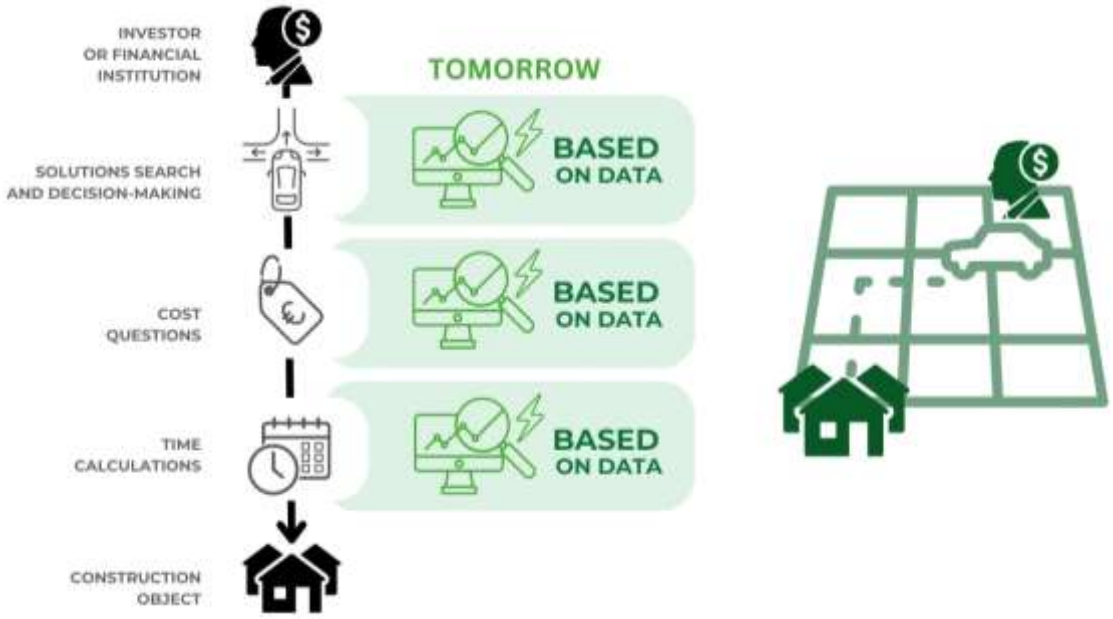


Рис. 10.1-5 Стоимость и время "пути" в процессе строительства будут определяться с помощью машинного обучения и статистических инструментов.

С постепенным введением новых норм и требований в почти каждой стране мира, обязывающих передавать модели CAD- (BIM)- клиентам или банкам, финансирующим строительные проекты, клиент и заказчик получает возможность самостоятельно обеспечить прозрачность расчётов данных стоимости и объемов работ. Особенно это актуально для крупных заказчиков и инвесторов, которые обладают достаточными компетенциями и инструментами для оперативного анализа объёмов и мониторинга рыночных цен. Для компаний, реализующих масштабные типовые проекты — магазины, офисные здания, жилые комплексы — такие практики становятся стандартом.

По мере того как информационное наполнение моделей становится более полным и стандартизированным, возможность манипуляций и спекуляций практически исчезает. Цифровая трансформация постепенно меняет правила игры в строительной отрасли, и компании, которые не адаптируются к этим изменениям, могут столкнуться с серьёзными вызовами.

> Усиление конкуренции, технологический разрыв и снижение маржинальности способны повлиять на устойчивость бизнеса. В условиях ограниченной ликвидности всё больше участников отрасли обращаются к автоматизации, аналитике и технологиям обработки данных как к способу повышения эффективности и прозрачности процессов. Эти инструменты становятся важным ресурсом для сохранения конкурентоспособности в меняющейся экономической среде.

Возможно, не стоит ждать, когда внешние обстоятельства заставят предпринимать срочные шаги — гораздо эффективнее начать подготовку уже сегодня, укрепляя цифровые компетенции, внедряя современные решения и выстраивая культуру, ориентированную на работу с данными.

Одним из последних ключевых технологических барьеров на пути к масштабной цифровой трансформации строительной отрасли, которая в ближайшие годы затронет каждую компанию, остаётся проблема автоматической классификации элементов строительных проектов.

> Без надёжной, точной и масштабируемой классификации невозможно создать основу для полноценной аналитики, автоматизации процессов и управления жизненным циклом объектов с применением ИИ и предиктивных моделей. Пока классификация объектов по-прежнему зависит от ручной интерпретации со стороны опытных специалистов — прорабов, проектировщиков, сметчиков — у строительной отрасли остаётся окно возможностей. Это время можно использовать для подготовки к неизбежным изменениям: росту требований к прозрачности, демократизации инструментов и данных, а также появлению систем автоматической классификации, которые радикально изменят правила игры.

Задача автоматической классификации элементов строительного мира по своей сложности сопоставима с распознаванием объектов в системах беспилотного вождения, что является одним из главных вызовов. Представим беспилотный автомобиль, движущийся из точки А в точку Б (Рис. 10.1-5). Современные системы автоматического вождения упираются в проблему классификации объектов, которые распознаются с помощью лидаров и камер. Автомобилю недостаточно просто «увидеть» препятствие или ориентир — он должен безошибочно понять, что перед ним: пешеход, дорожный знак или мусорный контейнер.

Аналогичная фундаментальная проблема стоит перед всей строительной отраслью. Элементы проекта — такие как окна, двери или колонны — могут быть зафиксированы в документации, представлены в CAD-моделях, сфотографированы на строительной площадке или распознаны в облаках точек от лазерного сканирования. Однако для построения действительно автоматизированной системы управления проектом недостаточно лишь их визуального или грубого геометрического распознавания. Необходимо обеспечить точную и устойчивую классификацию каждого элемента по типу, который будет однозначно идентифицируем во всех последующих процессах — от смет и спецификаций до логистики, складского учёта и главное — эксплуатации (Рис. 4.2-6).

Именно на этом этапе — переходе от распознавания к осмысленной классификации — и возникает одно из ключевых препятствий. Даже если цифровые системы технически способны выделить и идентифицировать объекты в моделях и на строительной площадке, основная сложность заклю-

чается в корректном и контекстно устойчивом определении типа элемента для различных программных сред. Например, дверь может быть обозначена проектировщиком в CAD-модели как элемент категории «дверь», однако при передаче в ERP- или PMIS-систему она может получить неверную типизацию — в случае ошибки со стороны проектировщика или из-за несовпадений между системами. Более того, элемент нередко теряет часть важных атрибутов или вовсе исчезает из системного учёта при экспортах и импортах данных. Это приводит к разрыву в потоке данных и подрывает принцип сквозной цифровизации строительных процессов. Таким образом, формируется критический разрыв между «видимым» и «понятным» семантическим значением, что подрывает целостность данных и существенно осложняет автоматизацию процессов на протяжении всего жизненного цикла строительного объекта.

Решение задачи универсальной классификации строительных элементов с применением технологий больших данных и машинного обучения (Рис. 10.1-6) станет катализатором трансформации всей отрасли — и, возможно, неожиданным открытием для многих строительных компаний. Унифицированная, обучаемая система классификации станет основой для масштабируемой аналитики, цифрового управления и внедрения ИИ в повседневную практику строительных организаций.

NVIDIA и другие технологические лидеры уже сегодня предлагают решения в других отраслях экономики, способные автоматически классифицировать и структурировать огромные объёмы текстовой и визуальной информации.

> Модель NeMo Curator [161] от NVIDIA, к примеру, специализируется на автоматической классификации и распределении данных по заранее заданным категориям, играя ключевую роль в оптимизации конвейеров обработки информации для задач тонкой настройки и предварительного обучения генеративных ИИ-моделей. Платформа Cosmos обучается на реальных видео и 3D-сценах [162], создавая базу для автономных систем и цифровых двойников, которые уже создаются в экосистеме NVIDIA. NVIDIA Omniverse, которая к 2025 году стала ведущим инструментом для работы с форматом USD — универсальным описанием сцен, способным в перспективе заменить формат IFC в процессах передачи проектной информации. В совокупности с Isaac Sim — симулятором роботизированных процессов [163] — такие решения, как NeMo Curator, Cosmos и Omniverse представляют собой новый уровень автоматизации: от очистки и фильтрации данных до генерации обучающих наборов, моделирования свойств объектов и обучения роботов на строительной площадке. Причём все эти инструменты распространяются бесплатно и с открытым доступом, что существенно снижает барьеры для внедрения в инженерные и строительные практики.

Автоматическая классификация данных на уровне структурированных таблиц — задача не столь сложная, как это может показаться на первый взгляд. Как мы показали в предыдущей главе (Рис. 9.1-10) при наличии накопленных исторических данных возможно восполнить пропущенные или некорректные значения классов на основе сходных параметров других элементов. Если в нескольких завершённых проектах элементы с аналогичными характеристиками уже были классифицированы корректно, система может с высокой вероятностью предложить подходящее значение для нового или неполного элемента (Рис. 10.1-6). Подобная логика, основанная на усреднённых значениях и анализе контекста, может быть особенно эффективна при массовой обработке табличных данных, поступающих из смет, спецификаций или CAD моделей.

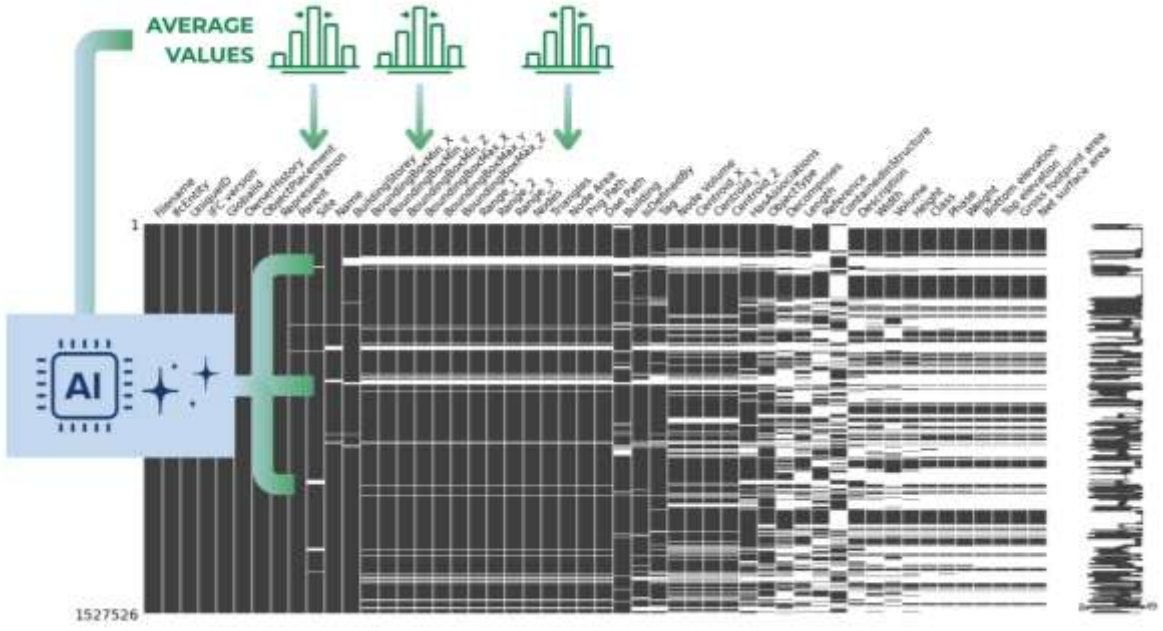


Рис. 10.1-6 Машинное обучение поможет автоматически найти усреднённые значения для незаполненных (белые поля) параметров таблиц на основе прошлых проектов.

На фоне столь стремительного прогресса в области машинного обучения становится очевидным: в 2025 году наивно полагать, что проблема автоматической классификации строительных элементов надолго останется нерешённой. Да, современные алгоритмы ещё не достигли полной зрелости, особенно в условиях неполных или неоднородных данных, но окно возможностей для адаптации быстро закрывается.

Компании, которые уже сейчас инвестируют в сбор, очистку и систематизацию своих данных, а также осваивают инструменты ETL автоматизации, окажутся в заведомо более выгодном положении. Остальные рискуют не успеть — так же, как некогда компании не справились с вызовами цифровой трансформации в транспортной и финансовой отраслях.

Те, кто продолжит полагаться на ручное управление данными и традиционные методы оценки стоимости и сроков, рискуют оказаться в положении таксопарков 2000-х годов, не успевших к началу 2020-х адаптироваться к эпохе мобильных приложений и автоматизированных расчетов маршрутов.

ГЛАВА 10.2.
ПРАКТИЧЕСКОЕ РУКОВОДСТВО ПО ВНЕДРЕНИЮ DATA-DRIVEN ПОДХОДА

От теории к практике: дорожная карта цифровой трансформации в строительстве

Строительная отрасль постепенно вступает в новую фазу развития, где привычные процессы всё чаще дополняются — а иногда и заменяются — цифровыми платформами и прозрачными моделями взаимодействия. Это открывает перед компаниями не только вызовы, но и значительные возможности. Те организации, которые уже сегодня выстраивают долгосрочную цифровую стратегию, смогут не только сохранить своё положение на рынке, но и расширить его, предложив клиентам современные подходы и надёжные, технологически подкреплённые решения.

> При этом важно понимать: знание концепций и технологий — лишь отправная точка. Перед руководителями и специалистами встаёт практический вопрос: с чего начать внедрение и как превратить теоретические идеи в реальную ценность. Кроме того, всё чаще возникает вопрос: на чём будет строиться бизнес, если традиционные методы расчёта стоимости и сроков могут быть пересмотрены заказчиком в любой момент.

Ответ, вероятно, лежит не столько в технологиях, сколько в формировании новой профессиональной культуры, где работа с данными воспринимается как неотъемлемая часть повседневной практики. Именно недостаточное внимание к цифровым технологиям и инновациям привило к серьезному отставанию строительной отрасли, которое наблюдается последние десятилетия [43].

> Согласно данным McKinsey, расходы на НИОКР в строительной отрасли составляют менее 1% от выручки, тогда как в автомобильной и аэрокосмической промышленности этот показатель достигает 3,5–4,5%. Аналогично, затраты на IT-технологии в строительстве остаются на уровне менее 1% от общей выручки [107].

В итоге не только уровень автоматизации, но производительность труда, в строительстве снижается, и к 2020-му строительный рабочий производит уже меньше, чем полвека назад (Рис. 10.2-1).

Подобные проблемы с производительностью в строительном секторе характерны для большинства развитых и развивающихся стран (производительность строительства упала в 16 из 29 стран ОЭСР (Рис. 2.2-1)), и указывают не только на нехватку технологий, но и на необходимость системных изменений в самих подходах к управлению, обучению и внедрению инноваций.

> Успех цифровой трансформации зависит не столько от количества и наличия инструментов, сколько от способности организаций пересматривать свои процессы и развивать культуру, открытую к изменениям. Ключевую роль играют не технологии сами по себе, а люди и выстроенные процессы, которые обеспечивают их эффективное применение, поддерживают непрерывное обучение и способствуют восприятию новых идей.

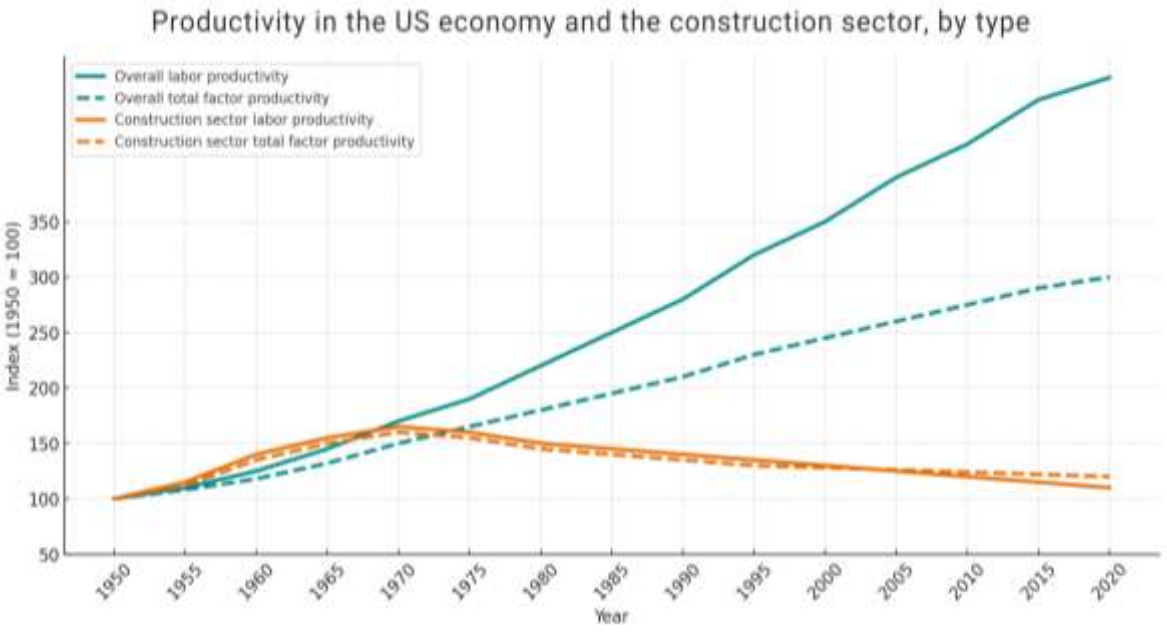


Рис. 10.2-1 Парадокс производительности труда и общей ресурсной производительности в экономике США и строительном секторе (1950–2020) (по материалам [43]).

В первых частях книги модель бизнес-среды сравнивалась с лесной экосистемой (Рис. 2.1-2, Рис. 1.2-4, Рис. 1.3-2). В здоровом лесу периодические пожары, при всей их разрушительной силе, играют ключевую роль в долгосрочном обновлении. Они расчищают почву от старой растительности, возвращают накопленные питательные вещества и создают пространство для новой жизни. Некоторые виды растений даже эволюционировали таким образом, что их семена раскрываются только под воздействием высоких температур пожара – это природный механизм, обеспечивающий идеальное время для прорастания.

Аналогично и в бизнесе: кризисы могут играть роль "контролируемого выжигания", способствуя появлению новых подходов и компаний, не связанных с устаревшими системами. Такие периоды вынуждают отказываться от неэффективных практик, высвобождая ресурсы для инноваций. Как лес после пожара начинает с растений-пионеров, так и бизнес после кризиса формирует новые, гибкие процессы, становящиеся основой для зрелой информационной среды.

Компании, сумевшие правильно интерпретировать эти "сигнальные пожары" и трансформировать их энергию разрушения в конструктивные изменения, выйдут на новый уровень эффективности – с более прозрачными, адаптивными процессами обработки данных, усиливающими естественную способность организации к обновлению и росту.

> Растущее влияние искусственного интеллекта и машинного обучения на бизнес-среду уже не вызывает сомнений. Это не просто временная тенденция, а стратегическая необходимость. Компании, игнорирующие ИИ, рискуют потерять конкурентоспособность на рынке, который всё активнее поощряет инновационность и гибкость.

Будущее принадлежит тем, кто видит в ИИ не просто инструмент, а возможность переосмыслить каждый аспект своей деятельности — от оптимизации процессов до принятия управленческих решений.

Закладываем цифровой фундамент: 1-5 шаги к цифровой зрелости

В этой главе мы рассмотрим дорожную карту цифровой трансформации и определим ключевые шаги, необходимые для внедрения data-driven подхода, который может помочь трансформировать как корпоративную культуру, так и информационную экосистему компании.

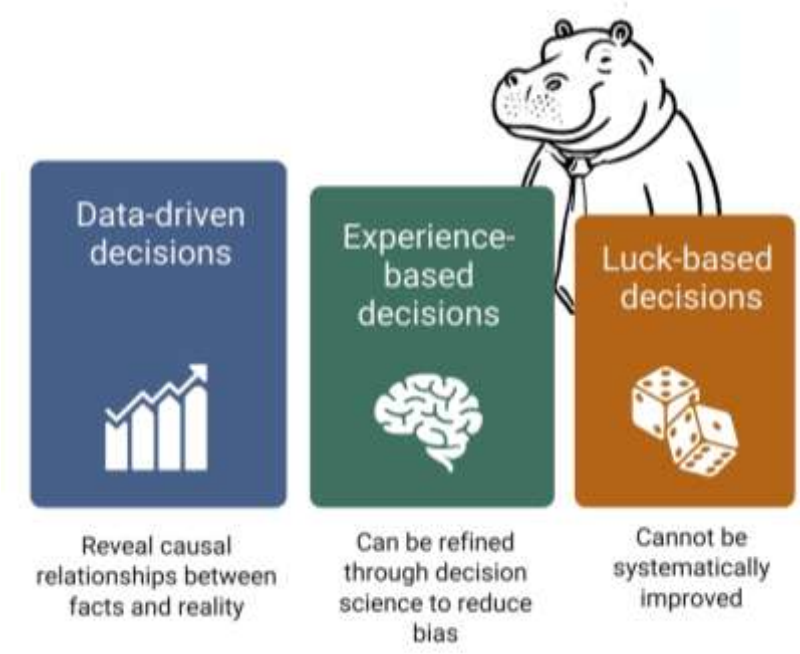


Рисунок 10.2-2 Контролируемое обновление и выбор стратегии: случай, опыт или данные.

Согласно исследованию McKinsey «Почему цифровые стратегии терпят неудачу» (2018), существует как минимум пять причин [164], по которым компании не достигают целей цифровой трансформации

- **Размытые определения**: руководители и менеджеры по-разному понимают, что такое "цифровые технологии", что приводит к недопониманию и несогласованности в действиях.
- **Неправильное понимание экономики цифровых технологий**: многие компании недооценивают масштаб изменений, которые цифровизация вносит в бизнес модели и динамику отраслей (Рис. 10.1-6).
- **Игнорирование экосистем**: компании сосредотачиваются на отдельных технологических решениях (силосах данных), упуская из виду необходимость интеграции в более широкие цифровые экосистемы (Рис. 2.2-2, Рис. 4.1-12).

- **Недооценка цифровизации со стороны конкурентов**: руководители не учитывают, что конкуренты также активно внедряют цифровые технологии, что может привести к потере конкурентных преимуществ.
- **Упущение двойственности цифровизации**: генеральные директоры делегируют ответственность за цифровую трансформацию другим руководителям, что бюрократизирует контроль и замедляет процесс изменений.

Для решения этих проблем необходимо чёткое понимание и согласование цифровых стратегий на всех уровнях организации. Прежде чем строить цифровую стратегию, важно понять исходную точку. Многие организации стремятся внедрять новые инструменты и платформы, не имея полной картины текущего состояния.

Шаг 1. Проведите аудит текущих систем и данных.

Прежде чем менять процессы, важно разобраться в том, что уже есть. Проведение аудита позволяет выявить слабые места в управлении данными и понять, какие ресурсы можно использовать. Подобный аудит – своего рода «рентген» бизнес-процессов. Он позволит вам выявить зоны риска и определить, какие данные являются критически важными для вашего проекта или бизнеса, а какие – второстепенными.

Основные действия:

- Составьте карту ИТ-среды (в Draw.io, Lucidchart, Miro, Visio или Canva). Перечислите используемые системы (ERP, CAD, CAFM, CPM, SCM и другие), участвующие в ваших процессах и которые мы обсуждали в главе "Технологии и системы управления в современном строительстве" (Рис. 1.2-4).
- Оцените проблемы с качеством данных для каждой системы на частоту наличие дубликатов, возможных пропущенных значений и несоответствий в форматах в каждой из систем.
- Выявите "болевые точки" – места, где процессы могут разрываться или часто требуют ручного вмешательства – импорта, экспорта и дополнительных процессов проверки.

> Если вы хотите, чтобы команда доверяла отчётам, нужно следить за корректностью данных с самого начала.

Качественно проведенный аудит данных покажет, какие данные:

- Нуждаются в доработке (требуется настройка автоматических процессов очистки или дополнительной трансформации)
- Представляют собой «мусор», который только засоряет системы и от которого можно избавиться, больше не использую его в процессах.

Подобный аудит можно провести самостоятельно. Но иногда полезно привлечь внешнего консультанта — особенно из других отраслей экономики: свежий взгляд и независимость от строительных "особенностей" помогут трезво оценить статус-кво и избежать типичных ловушек предвзятости к определённым решениям и технологиям.

Шаг 2. Определите ключевые стандарты для унификации данных.

После проведения аудита необходимо создать общие правила работы с данными. Как мы обсуждали в главе "Стандарты: от случайных файлов к продуманной модели данных", это поможет устранить разрозненность информационных потоков.

> Без единого стандарта каждая команда продолжит работать «по-своему», и вы сохраните «зоопарк» интеграций, где данные теряются при каждом преобразовании.

Основные действия:

- Выберите стандарты данных для обмена информацией между системами:
  - Для табличных данных это могут быть структурированные форматы вроде CSV, XLSX или более эффективные форматы вроде Parquet
  - Для обмена слабоструктурированными данными и документами: JSON или XML
- Освойте работу с моделями данных:
  - Начните с параметризации задач на уровне концептуальной модели данных — как это описано в главе «Моделирование данных: концептуальная, логическая и физическая модель» (Рис. 4.3-2)
  - По мере углубления в логику бизнес-процессов переходите к формализации требований с использованием параметров в логической и физической моделях (Рис. 4.3-6)
  - Определите ключевые сущности, их атрибуты и взаимосвязи в рамках процессов, а также визуализируйте эти связи — как между сущностями, так и между параметрами (Рис. 4.3-7)
- Используйте регулярные выражения (RegEx) для валидации и стандартизации данных (Рис. 4.4-7), как мы обсуждали в главе "Структурированные требования и регулярные выражения RegEx". RegEx – не сложная, но крайне важная тема в работе создания требований на уровне физических моделей данных.

Без стандартов на уровне данных и визуализации процессов невозможно обеспечить согласованную и масштабируемую цифровую среду. Помните: "плохие данные стоят дорого". И цена ошибки растёт по мере того, как проект или организация становится сложнее. Унификация форматов, определение правил именования, структуры и валидации — это инвестиции в стабильность и масштабируемость будущих решений.

Шаг 3. Внедряйте DataOps и автоматизируйте процессы.

Без четко выстроенной архитектуры компании неизбежно столкнуться с разрозненными данными, заключенными в изолированных информационных системах. Данные окажутся не интегрированными, будут дублироваться в разных местах и потребуют значительных затрат на поддержку.

> Представьте, что данные — это вода, а архитектура данных — это сложная система трубопроводов, по которым эта вода транспортируется от источника хранения к месту использования. Именно архитектура данных определяет, как информация собирается, хранится, преобразуется, анализируется и доставляется конечным пользователям или приложениям.

DataOps (Data Operations) – это методология, объединяющая сбор, очистку, проверку и использование данных в единый автоматизированный поток процессов, о чём мы подробно говорили в восьмой части книги.

Основные действия:

- Создавайте и настраивайте ETL -конвейеры для автоматизации процессов:
  - Extract: организуйте автоматический сбор данных из PDF документов (Рис. 4.1-2, Рис. 4.1-5, Рис. 4.1-7), Excel таблиц, CAD-моделей (Рис. 7.2-4), ERP-систем и других источников, с которыми вы работаете
  - Transform: настройте автоматические процессы преобразования данных к единому структурированному формату и автоматизируйте расчёты, которые будут проходить вне закрытых приложений (Рис. 7.2-8)
  - Load: попробуйте создать автоматическую выгрузку данных в итоговые таблицы, документы или централизованные хранилища (Рис. 7.2-9, Рис. 7.2-13, Рис. 7.2-16)
- Автоматизируйте процессы расчёта и QTO (Quantity Take-Off), как мы обсуждали в главе "QTO Quantity Take-Off: группировка проектных данных по атрибутам":
  - Настройте автоматическое извлечение объёмов из CAD-моделей, при помощи API, плагинов или инструментов обратного инжиниринга (Рис. 5.2-5)
  - Создайте правила группировки элементов для различных классов по атрибутам в форме таблиц (Рис. 5.2-12)
  - Попробуйте автоматизировать часто повторяющиеся расчёты объёмов и стоимости вне модульных закрытых систем (Рис. 5.2-15)
- Начните использовать Python и Pandas для обработки данных, как мы рассматривали в главе "Python Pandas: незаменимый инструмент для работы с данными":
  - Применяйте DataFrame для работы с файлами XLSX и автоматизации обработки табличных данных (Рис. 3.4-6)
  - Автоматизируйте агрегацию и трансформацию информации через различные библиотеки Python
  - Используйте LLM для упрощения написания готовых блоков кода и целых Pipeline (Рис. 7.2-18)
  - Попробуйте построить Pipeline на Python, который находит ошибки или видит аномалии и отправляет уведомление ответственному лицу (например, менеджеру проекта) (Рис. 7.4-2)

Автоматизация на основе принципов DataOps позволяет перейти от ручной и фрагментарной работы с данными к устойчивым и воспроизводимым процессам. Это не только снижает нагрузку на сотрудников, каждодневно занимающихся одними и теми же преобразованиями, но и резко повышает надёжность, масштабируемость и прозрачность всей информационной системы.

Шаг 4. Создайте экосистему управления открытыми данными.

Несмотря на развитие закрытых модульных систем и их интеграцию с новыми инструментами, компании сталкиваются с серьезной проблемой – рост сложности подобных систем опережает их полезность. Первоначальная идея создания единой проприетарной платформы, охватывающей все бизнес-процессы, привела к чрезмерной централизации, где любые изменения требуют значительных ресурсов и времени на адаптацию.

Как мы рассматривали в главе "Корпоративный мицелий: как данные соединяют бизнес-процессы", эффективная работа с данными требует создания открытый и единой экосистемы, объединяющей все источники информации.

Ключевые элементы экосистемы:

- Выберите подходящее хранилище данных:
  - Для таблиц и расчётов используйте базы данных — например, PostgreSQL или MySQL (Рис. 3.1-7)
  - Для документов и отчётов могут подойти облачные хранилища (Google Drive, OneDrive) или системы, поддерживающие JSON-формат
  - Ознакомьтесь с возможностями Data Warehouse, Data Lakes и других инструментов для централизированного хранения и анализа больших объёмов информации (Рис. 8.1-8)
- Внедрите решения для доступа к проприетарным данным:
  - Если вы используете проприетарные системы, настройте доступ к ним через API или SDK чтобы получать данные для внешней обработки (Рис. 4.1-2)
  - Ознакомьтесь с потенциалом инструментов обратного инжиниринга для CAD форматов (Рис. 4.1-13)
  - Настройте ETL-Pipeline, которые периодически собирают данные с приложений или серверов, преобразуют их в открытые структурированные форматы и сохраняют в хранилища (Рис. 7.2-3)
  - Обсуждайте внутри команды вопросы по обеспечению доступа к данным без необходимости использования проприетарного ПО
  - Помните: данные важнее интерфейсов. Долгосрочную ценность представляет структура и доступность информации, а не конкретные инструменты пользовательского интерфейса
- Задумайте о создании центра передового опыта (CoE) по данным, как мы обсуждали в главе "Центр передового опыта (CoE) по моделированию данных" или о том как другими способами можно обеспечить экспертизу по работе с данными (Рис. 4.3-9)

Экосистема управления данными создаёт единое информационное пространство, в котором все участники проекта работают с согласованной, актуальной и проверенной информацией. Это основа для масштабируемых, гибких и надёжных цифровых процессов.

Раскрываем потенциал данных: 5-10 шаги к цифровой зрелости

Помимо технической интеграции, важным фактором успешного внедрения цифровых решений является их принятие конечными пользователями. Привлечение клиентов или пользователей к вопросам оценки эффективности — это одновременно задача улучшения пользовательского опыта и управления изменениями в компании. Если решение не вписывается в привычный рабочий процесс или не решает реальных проблем пользователей или клиентов, оно не будет использовано, и никакие дополнительные меры и поощрения этого не исправят.

Трансформация — это итеративный процесс, основанный на анализе данных о взаимодействии пользователей с новыми процессами, с частыми циклами тестирования, постоянной обратной связи и доработками.

Шаг 5. Сформируйте культуру работы с данными, обучите персонал и собирайте обратную связь

Даже самая продвинутая система не будет работать без вовлечённости сотрудников. Необходимо создать среду, в которой данные используются ежедневно, а команда понимает их ценность.

> В опубликованном отчете правительства Великобритании «Data Analytics and AI in Government Project Delivery» 2024 года отмечается [83], что для успешного внедрения аналитики данных и ИИ критически важна подготовка специалистов, обладающих необходимыми компетенциями в обработке и интерпретации данных.

Недостаток знаний в области анализа данных является одной из ключевых проблем, ограничивающих цифровую трансформацию. Руководители привыкли к устоявшимся процедурам: квартальным циклам, приоритетным инициативам и традиционным путям продвижения проектов. Для изменений требуется особый лидер — достаточно высокого ранга, чтобы иметь влияние, но не настолько высокого, чтобы у него оставались время и мотивация вести долгосрочный проект трансформации.

Основные действия:

- Осознание необходимости перехода от субъективных решений, основанных на мнении высокооплачиваемого сотрудника (HiPPO), к культуре принятия решений, опирающейся на факты и данные, как обсуждалось в главе "HiPPO или опасность мнений в принятии решений" (Рис. 2.1-9).
- Организуйте системное обучение:
    - Проводите тренинги по использованию структурированных данных, и приглашайте экспертов из других отраслей экономики, у которых нет предвзятости к продуктам и концептам, популярным сегодня в строительной отрасли

- Обсуждайте подходы и инструменты анализа данных с коллегами, а также самостоятельно осваивайте практическую работу с такими средствами, как Python, pandas и LLM (Рис. 4.1-3, Рис. 4.1-6)
- Создайте библиотеку учебных материалов (лучше с короткими видео) по теме структурирования данных (Рис. 3.2-15) и создания моделей данных (Рис. 4.3-6, Рис. 4.3-7)

- Используйте современные технологии обучения:
  - Используйте языковые модели (LLM) для поддержки при работе с кодом и данными, включая генерацию, рефакторинг и анализ кода, а также обработку и интерпретацию табличной информации (Рис. 3.4-1)
  - Изучайте, как сгенерированный с помощью LLM код может быть адаптирован и интегрирован в полноценное Pipeline-решение при работе в офлайн-среде разработки (IDE) (Рис. 4.4-14, Рис. 5.2-13)

> Когда руководитель продолжает принимать решения «по старинке», ни один тренинг не убедит людей всерьёз относиться к аналитике.

Формирование культуры работы с данными невозможно без постоянной обратной связи. Обратная связь позволяет выявлять недостатки в процессах, инструментах и стратегиях, которые невозможно обнаружить через внутренние отчёты или формальные KPI метрики. Комплиментарные замечания пользователей ваших решений не принесут практической ценности. Ценность представляет именно критическая обратная связь, особенно если она основана на конкретных наблюдениях и фактах. Однако получение такой информации требует усилий: необходимо выстроить процессы, в которых участники — как внутренние, так и внешние — могут делиться замечаниями (возможно есть смысл это делать анонимно) без искажений и без опасений, что их мнение может повлиять на их собственную работу. Важно, чтобы они делали это без искажений и без страха негативных последствий для себя.

> *Любое обучение в конечном итоге является самообучением [165].*
>
> – Milton Friedman, американский экономист и статистик

Внедрение аналитических инструментов должно сопровождаться регулярной верификацией их эффективности на практике (ROI, KPI), чего можно достичь только через структурированную обратную связь от сотрудников, клиентов и партнёров. Это позволяет компаниям не только избежать повторения ошибок, но и быстрее адаптироваться к изменениям среды. Наличие механизма сбора и анализа обратной связи — один из признаков зрелости организации, переходящей от эпизодических цифровых инициатив к устойчивой модели непрерывного совершенствования (Рис. 2.2-5).

Шаг 6. От пилотных проектов к масштабированию

> *Выбирайте достаточно крупные битвы, чтобы иметь значение, и достаточно мелкие, чтобы победить.*
>
> – Джонатан Козол

Запуск цифровой трансформации "сразу и везде" крайне рискован. Более эффективный подход – начать с пилотных проектов и постепенно масштабировать успешные практики.

Основные действия:

- Выберите подходящий проект для пилота:
  - Определите конкретную бизнес-задачу или процесс с измеримыми результатами (KPI, ROI) (Рис. 7.1-5)
  - Выберите процесс ETL автоматизации, например автоматической проверки данных или расчёта объёмов работ (QTO) с помощью Python и Pandas (Рис. 5.2-10)
  - Установите чёткие метрики успеха (например - сократить время на составление спецификаций проверок или отчётов по проверки данных с недели до одного дня)
- Применяйте итеративные подходы:
  - Начните с простых процессов преобразования данных и создания потоковой конвертации разноформатных данных в необходимые для ваших процессов форматы (Рис. 4.1-2, Рис. 4.1-5)
  - Постепенно повышайте сложность задач и расширяйте автоматизацию процессов, формируя полноценный Pipeline в среде разработки (IDE) на основе задокументированных блоков кода (Рис. 4.1-7, Рис. 7.2-18)
  - Документируйте и записывайте (лучше при помощи коротких видео) успешные решения и делитесь ими с коллегами или в профессиональных комьюнити
- Разрабатывайте шаблоны и сопровождающую документацию для тиражирования подобных решений, чтобы ими могли эффективно пользоваться ваши коллеги (или участники профессионального сообщества, включая пользователей в социальных сетях)

Поэтапный «накат» позволяет удерживать высокое качество изменений и не свалиться в хаос параллельных внедрений. Стратегия "от малого к большому" минимизирует риски и позволяет учиться на мелких ошибках, не позволяя им перерасти в критические проблемы.

Переход от проектного подхода, при котором сотрудники вовлечены лишь частично, к формированию постоянных команд (например, центров экспертизы — CoE) позволяет обеспечить устойчивое развитие продукта даже после выхода его первой версии. Такие команды не только поддерживают существующие решения, но и продолжают их совершенствовать.

Это снижает зависимость от длительных согласований: участники команд получают право принимать решения в рамках своей зоны ответственности. В результате менеджеры освобождаются от

необходимости микроменеджмента, а команды могут сосредоточиться на создании реальной ценности.

> Разработка новых решений — это не спринт, а марафон. Преуспевают в нём те, кто изначально нацелен на долгосрочную, последовательную работу.

Важно понимать, что технологии требуют постоянного развития. Инвестирование в долгосрочное развитие технологических решений – это основа успешной работы.

Шаг 7. Используйте открытые форматы данных и решения

Как мы обсуждали в главах посвящённых модульным платформам (ERP, PMIS, CAFM, CDE и др.) важно сфокусироваться на открытых и универсальных форматах данных, которые обеспечат независимость от решений вендоров и повысят доступность информации для всех участников процесса.

Основные действия:

- Переходите от закрытых форматов к открытым:
  - Используйте открытые форматы вместо проприетарных, или найдите возможность настроить автоматическую выгрузку или конвертацию закрытых форматов в открытые (Рис. 3.2-15)
  - Внедряйте инструменты для работы с Parquet, CSV, JSON, XLSX, которые являются стандартами обмена между большинства современных систем (Рис. 8.1-2)
  - Если работа с 3D-геометрией играет важную роль в ваших процессах, рассмотрите возможность использования открытых форматов, таких как USD, glTF, DAE или OBJ (Рис. 3.1-14)
- Используйте векторные базы данных для эффективного анализа и поиска информации:
  - Используйте Bounding Box и другие методы для упрощения работы с 3D-геометрией (Рис. 8.2-1)
  - Подумайте, где можно внедрить векторизацию данных — преобразование текстов, объектов или документов в числовые представления (Рис. 8.2-2)
- Применяйте инструменты анализа больших данных:
  - Организуйте хранение накопленных исторических данных (например PDF, XLSX, CAD) в подходящих для анализа форматах (Apache Parquet, CSV, ORC) (Рис. 8.1-2)
  - Начните применять базовые статистические методы и работать с репрезентативными выборками — или, как минимум, ознакомьтесь с фундаментальными принципами статистики (Рис. 9.2-5)
  - Внедряйте и изучайте инструменты визуализации данных и связей между данными для наглядного представления результатов анализа. Без качественной визуализации невозможно полноценно понять ни сами данные, ни процессы, основанные на них (Рис. 7.1-4)

Переход к открытым форматам данных и внедрение инструментов для анализа, хранения и визуализации информации закладывает основу для устойчивого и независимого цифрового управления. Это не только снижает зависимость от вендоров, но и обеспечивает равный доступ к данным для всех участников процесса.

Шаг 8. Начните внедрять машинное обучение для прогнозирования

Во многих компаниях накоплены обширные массивы данных — своего рода «информационные гейзеры», до сих пор остающиеся неиспользованными. Эти данные собирались в рамках сотен и тысяч проектов, но зачастую применялись лишь однократно или вовсе не были вовлечены в дальнейшие процессы. Документы и модели, хранящиеся в закрытых форматах и системах, нередко воспринимаются как устаревший и бесполезный балласт. Однако на деле именно они представляют собой ценнейший ресурс — основу для анализа допущенных ошибок, автоматизации рутинных операций и разработки инновационных решений по авто классификации и распознаванию элементов в будущих проектах.

Ключевая задача — научиться извлекать эти данные и преобразовывать их в практическую пользу. Как уже обсуждалось в главе «Машинное обучение и прогнозы», методы машинного обучения способны существенно повысить точность оценок и предсказаний в различных процессах, связанных со строительство. Полноценное использование накопленных данных открывает путь к повышению эффективности, снижению рисков и построению устойчивых цифровых процессов.

Основные действия:

- Начните с простых алгоритмов:
  - Попробуйте применить линейную регрессию — с использованием подсказок от LLM — для прогнозирования повторяющихся показателей в наборах данных, где зависимости от большого числа факторов отсутствуют или минимальны (Рис. 9.3-4)
  - Рассмотрите, на каких этапах ваших процессов теоретически может быть применён алгоритм k-ближайших соседей (k-NN) — например, для задач классификации, оценки схожести объектов или прогнозирования на основе исторических аналогов (Рис. 9.3-5)
- Собирайте и структурируйте данные для обучения моделей:
  - Соберите исторические данные о проектах в одном месте и едином формате (Рис. 9.1-10)
  - Работайте над качеством и репрезентативностью обучающих выборок, через автоматические ETL (Рис. 9.2-8)
  - Научитесь разделять данные на тренировочные и тестовые наборы, как мы это делали в примере с датасетом Титаника (Рис. 9.2-6, Рис. 9.2-7)
- Рассматривайте возможности расширения применения методов машинного обучения для решения широкого круга задач — от прогнозирования сроков реализации проектов до оптимизации логистики, управления ресурсами и раннего выявления потенциальных проблем

Машинное обучение — это инструмент, позволяющий превратить архивные данные в ценный актив

для прогнозирования, оптимизации и обоснованного принятия решений. Начинайте с небольших датасетов (Рис. 9.2-5) и простых моделей, постепенно наращивая сложность.

Шаг 9. Интегрируйте IoT и современные технологии сбора данных

Строительный мир стремительно становится цифровым: каждое фото со стройки, каждое сообщение в Teams — это уже часть большого процесса параметризации и токенизации реальности. Как GPS когда-то преобразил логистику, так и технологии IoT, RFID и автоматический сбор данных меняют строительную отрасль. Как рассматривалось в главе "IoT Интернет вещей и смарт-контракты", цифровая стройплощадка с датчиками и автоматизированным мониторингом – будущее отрасли.

Основные действия:

- Внедряйте IoT-устройства, RFID-метки и детализируйте процессы связанные с ними:
  - Оцените, в каких зонах или этапах проекта установка датчиков может дать наибольшую отдачу (ROI) — например, для мониторинга температуры, вибрации, влажности или движения техники
  - Рассмотрите вопрос применения RFID для отслеживания материалов, инструментов и оборудования на всех этапах логистической цепочки
  - Продумайте, как собранные данные можно интегрировать в единую информационную систему, такую как Apache NiFi, для автоматизированной обработки и анализа в реальном времени (Рис. 7.4-5)
- Создайте систему мониторинга в реальном времени:
  - Разработайте дашборды для отслеживания ключевых показателей процесса или проекта с помощью инструментов визуализации, таких как Streamlit, Flask или Power BI)
  - Настройте автоматические уведомления, которые будут сигнализировать о критических отклонениях от плана или норм (Рис. 7.4-2)
  - Оцените потенциал предиктивного обслуживания оборудования на основе собранных данных и выявленных закономерностей (Рис. 9.3-6)
- Объедините данные из различных источников:
  - Начните с визуализации модели данных на физическом уровне — отразите структуру потоков информации и ключевых параметров, поступающих из CAD-систем, IoT-устройств и ERP-платформ (Рис. 4.3-1)
  - Начните с создания чернового описания единой платформы, предназначенной для анализа данных и поддержки принятия управленческих решений. Зафиксируйте ключевые функции, источники данных, пользователей, а также предполагаемые сценарии применения (Рис. 4.3-7)

Чем раньше вы начнёте подключать реальные процессы к цифровому миру, тем быстрее сможете управлять ими с помощью данных — эффективно, прозрачно и в режиме реального времени.

Шаг 10. Подготовьтесь к будущему изменений в отрасли

Строительные компании постоянно находятся под давлением внешней среды: экономических кризисов, технологических скачков, нормативных изменений. Подобно лесу, который вынужден выдерживать дождь, снег, засуху и палящее солнце, компании живут в условиях непрерывной адаптации. И как деревья обретают устойчивость к морозам и засухе благодаря глубокой корневой системе, так и только те организации, которые обладают прочным фундаментом в виде автоматизированных процессов, способностью предвидеть изменения и гибко адаптировать стратегии, сохраняют жизнеспособность и конкурентоспособность.

Как отмечалось в главе «Стратегии выживания: формирование конкурентных преимуществ», строительная отрасль входит в фазу радикальной трансформации. Взаимодействие между клиентом и исполнителем движется к модели уберизации, где прозрачность, предсказуемость и цифровые инструменты вытесняют традиционные подходы. В этой новой реальности выигрывают не самые крупные, а самые гибкие и технологически зрелые.

Основные действия:

- Проанализируйте уязвимости бизнеса в контексте открытых данных:
  - Оцените, как демократизация доступа к данным в рамках уберизации может разрушительно повлиять на ваши конкурентные преимущества и ваш бизнес (Рис. 10.1-5)
  - Задумайтесь о стратегии перехода от непрозрачных и изолированных процессов к бизнес-моделям, основанным на открытых решениях, совместимости систем и прозрачности данных (Рис. 2.2-5)
- Разработайте долгосрочную цифровую стратегию:
  - Определите, стремитесь ли вы быть лидером инноваций или предпочитаете "догоняющий" сценарий, в котором вы будете экономить свои ресурсы
  - Распишите этапы: краткосрочные (автоматизация процессов, структурирование данных), среднесрочные (внедрение LLM и ETL), долгосрочные (цифровые экосистемы, централизированные хранилища)
- Думайте над расширением портфеля услуг:
  - Рассмотрите возможность предложения новых сервисов (направленных на энергоэффективность, ESG, услуги по обработке данных). Подробнее о новых бизнес моделях мы поговорим в следующей главе
  - Стремитесь позиционировать себя как надёжного технологического партнёра, сопровождающего весь жизненный цикл объекта — от проектирования до эксплуатации. Доверие к вам должно основываться на системном подходе, прозрачности процессов и способности обеспечивать устойчивые технологические решения

В условиях трансформации выигрывают не те, кто просто реагирует на изменения, а те, кто действует на опережение. Гибкость, открытость и цифровая зрелость — основа устойчивости в строительстве завтрашнего дня.

Дорожная карта трансформации: от хаоса к data-driven компании

Следующий план может служить первичным ориентиром — отправной точкой для формирования собственной стратегии цифровой трансформации, основанной на данных:

- **Аудит и стандарты**: анализируйте текущее состояние, унифицируйте данные
- **Структурирование и классификация данных**: автоматизируйте трансформацию неструктурированных и слабоструктурированных данных
- **Автоматизация группировок**, расчётов и калькуляций: используйте открытые инструменты и библиотеки для автоматизации
- **Экосистема и СОЕ**: создайте внутреннюю команду, которая будет формировать единую экосистему данных в компании
- **Культура и обучение**: отходите от HiPPO-решений к решениям на основе данных
- **Пилоты, обратная связь и масштабирование**: действуйте итеративно: тестируйте новые методы в ограниченном масштабе, собирайте обоснованную обратную и постепенно масштабируйте решения
- **Открытые форматы**: используйте универсальные и открытые форматы для независимости от вендоров программного обеспечения
- **Машинное обучение**: внедряйте в процессы алгоритмы ML для прогнозирования и оптимизации
- **IoT и цифровая стройплощадка**: интегрируйте современные технологии сбора данных в процессы
- **Стратегическая адаптация**: готовьтесь к будущим изменениям отрасли

> Главное – помните, что «данные сами по себе не меняют компанию: её меняют люди, умеющие работать с этими данными». Делайте ставку на культуру, прозрачные процессы и стремление к непрерывным улучшениям

Системный подход позволяет перейти от разрозненных цифровых инициатив к полноценной data-driven модели управления, в которой решения принимаются не на основе интуиции или предположений, а на основе данных, фактов и математически рассчитанных вероятностей. Цифровая трансформация строительной отрасли — это не просто внедрение технологий, а формирование бизнес-экосистемы, где информация о проекте передаётся бесшовно и итеративно между различными системами. Алгоритмы машинного обучения при этом обеспечивают автоматический, непрерывный анализ, прогнозирование и оптимизацию процессов. В такой среде спекуляции и скрытые данные утрачивают актуальность — остаются только проверенные модели, прозрачные расчёты и предсказуемые результаты.

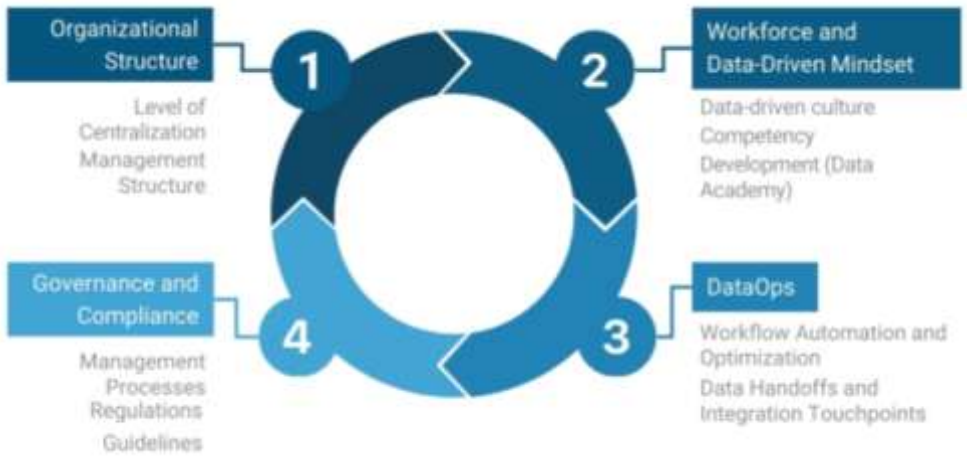


Рис. 10.2-3 Ключевые элементы успешного управления данными на уровне компании.

Каждая часть книги соответствует конкретному этапу обработки и анализа данных в строительных проектах (Рис. 2.2-5). Если вы хотите вернуться к одной из тем, рассмотренных ранее, и взглянуть на неё уже со стороны целостного понимания потока использования данных — вы можете ориентироваться на названия частей, указанные на Рис. 10.2-4.

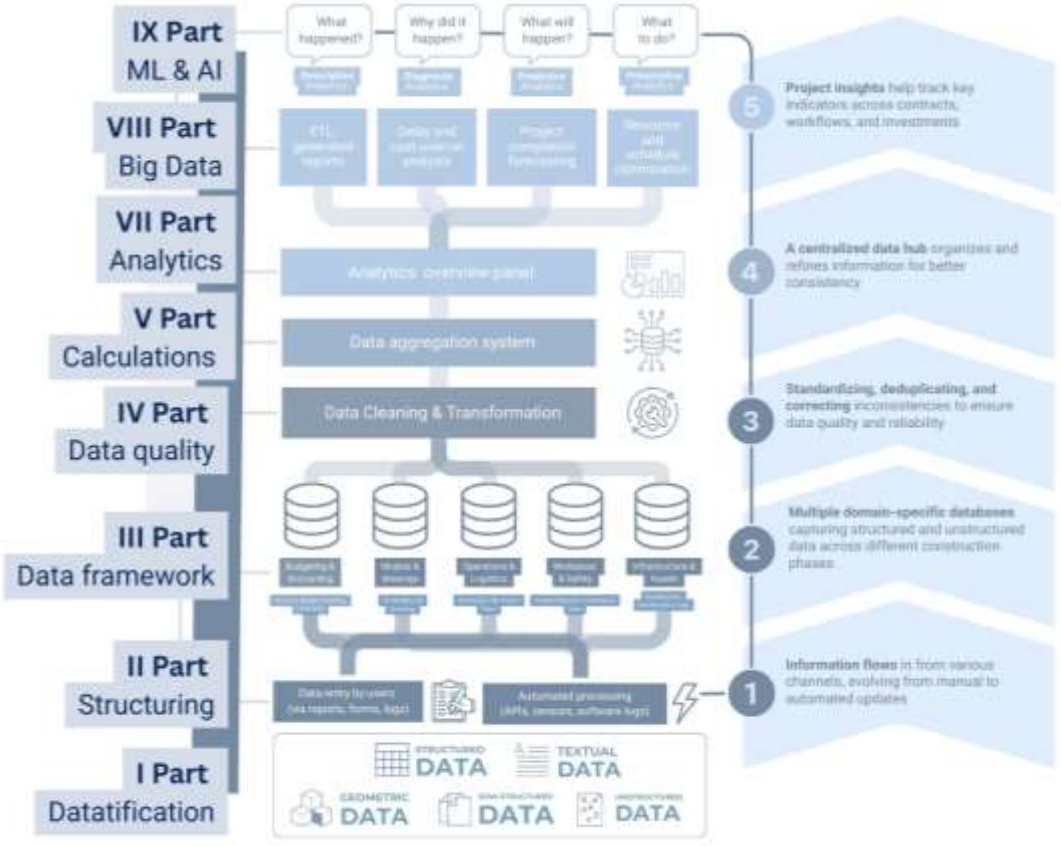


Рис. 10.2-4 Части книги в контексте конвейера обработки данных (Рис. 2.2-5): от дигитализации информации до аналитики и искусственного интеллекта.

Независимо от размера вашей организации, уровня технологической зрелости или бюджета, вы можете начать движение к data-driven подходу уже сегодня. Даже небольшие шаги в правильном направлении со временем приведут к результатам.

> Data-driven трансформация — это не одноразовый проект, а непрерывный, итеративный процесс совершенствования, включающий внедрение новых инструментов, пересмотр процессов и развитие культуры принятия решений на основе данных.

Строительство в индустрии 5.0: как зарабатывать, когда скрывать больше нельзя

Долгое время строительные компании зарабатывали на непрозрачности процессов. Основной бизнес-моделью становилась спекуляция — завышение стоимости материалов, объёмов работ и процентных надбавок в закрытых ERP-, PMIS-системах, недоступных для внешнего аудита. Ограниченный доступ заказчиков и их доверенных лиц к исходным проектным данным создавал почву для схем, в которых проверка достоверности расчётов становилась практически невозможной.

Однако эта модель стремительно теряет актуальность. С демократизацией доступа к данным, появлением LLM, приходом открытых данных, а также инструментов ETL автоматизации отрасль переходит к новому стандарту работы.

> В результате непрозрачность перестаёт быть конкурентным преимуществом — вскоре она станет бременем, с которым будет тяжело расстаться. Прозрачность же превращается из опции в обязательное условие для того, чтобы остаться на рынке.

С кем же будут работать клиенты — банки, инвесторы, физические заказчики, private equity, государственные заказчики — в новой цифровой реальности? Ответ очевиден: с теми, кто способен предоставить не только результат, но и обоснование каждого шага на пути к нему. В условиях роста объёма открытых данных, партнёры и заказчики будут выбирать компании, которые гарантируют прозрачность, точность и предсказуемость результатов.

На этом фоне формируются новые бизнес-модели, в основе которых лежит не спекуляция, а управление данными и доверие:

- **Продажа процессов вместо квадратных метров**: ключевым активом становится не договорённости по скидке на бетон, а доверие и эффективность. Основной ценностью станет предсказуемость результата, основанная на достоверных и верифицированных данных. Современные компании будут продавать не объект строительства как таковой, а:
  - точные сроки и прозрачные графики работ;
  - обоснованные сметы, подтверждённые расчётами;
  - возможность полной цифровой прослеживаемости и контроля на всех этапах проекта.
- **Инжиниринг и аналитика как услуга**: модель “Data-as-a-Service” (способ доставки готовых данных пользователям через интернет, как услугу), где каждый проект становится частью цифровой цепочки данных, а ценность бизнеса — в способности этой цепочкой управлять. Компании трансформируются в интеллектуальные платформы, предлагающие решения на базе автоматизации и аналитики:
  - автоматизированное и прозрачное составление смет и планов;
  - оценка рисков и сроков на основе алгоритмов машинного обучения;
  - расчёт экологических показателей (ESG, CO_2, энергоэффективность);
  - формирование отчётности из проверяемых открытых источников.

- **Продуктизация инженерного опыта**: наработки компании могут многократно использоваться внутри компании и распространяться как отдельный продукт — формируя дополнительный источник дохода за счёт цифровых сервисов. В новых условиях компании создают не только проекты, но и цифровые активы:
  - библиотеки компонентов и шаблонов смет;
  - автоматизированные модули верификации;
  - open-source плагины и скрипты (продажа консалтинга) для работы с данными.
- **Новый тип компании: Data-Driven интегратор**: такой участник рынка, который не зависит от конкретных вендоров ПО или модульных систем и не “заперт” в интерфейсе единственного ПО. Он свободно оперирует с данными — и на этом строит свою конкурентоспособность. Строительная компания будущего — это не просто подрядчик, а интегратор информации, способный для заказчика выполнять следующие функции:
  - объединять данные из разрозненных источников и проводить аналитику;
  - обеспечивать прозрачность и достоверность процессов;
  - консультировать по оптимизации бизнес-процессов;
  - разрабатывать инструменты, работающие в экосистеме открытых данных, LLM, ETL и Pipelines.

Индустрия 5.0 (Рис. 2.1-12) знаменует собой конец “эры ручных усреднённых коэффициентов” и вечерних встреч генеральных директоров с отделом смет и бухгалтерии. Всё, что прежде было скрыто — расчёты, сметы, объёмы — становится открытым, проверяемым и понятным даже не эксперту. В выигрыше окажутся те, кто первыми переориентируются. Все остальные — окажутся за бортом новой цифровой экономики строительного сектора.

ЗАКЛЮЧЕНИЕ

Строительная отрасль вступает в эпоху фундаментальных изменений. От первых записей на глиняных табличках до массивов цифровых данных, поступающих с проектных серверов и строительных площадок, — история работы с информацией в строительстве всегда отражала уровень зрелости технологий своего времени. Сегодня, с приходом автоматизации, открытых форматов и интеллектуальных систем анализа, отрасль сталкивается не с постепенной эволюцией, а со стремительной цифровой трансформацией.

Как и в других секторах экономики, строительству предстоит переосмыслить не только инструменты, но и принципы работы. Компании, которые раньше диктовали рынок и служили основным посредником между заказчиком и проектом, теряют своё уникальное положение. На первый план выходят доверие и способность работать с данными: от их сбора и структурирования до аналитики, прогнозирования и автоматизации решений.

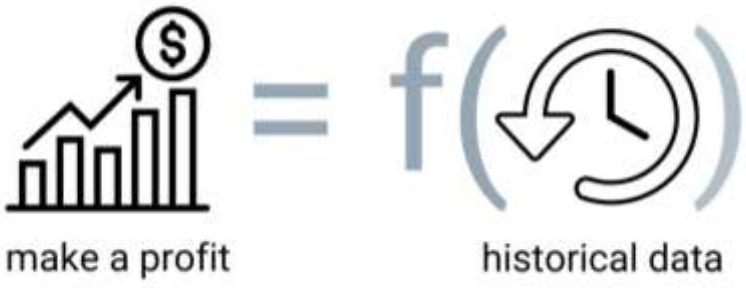


Рис. 10.2-1 Структурированные исторические данные — топливо для эффективного и управляемого бизнеса.

В этой книге были подробно разобраны ключевые принципы управления данными в строительной отрасли — от аудита и стандартизации до автоматизации процессов, использования инструментов визуализации и внедрения интеллектуальных алгоритмов. Мы рассмотрели, как даже при ограниченных ресурсах можно выстроить работающую архитектуру данных и начать принимать решения не на основе интуиции, а на основе проверяемых фактов. Работа с данными перестаёт быть задачей только ИТ-отдела — она становится основой управленческой культуры, от которой зависит гибкость компании, её адаптивность и долгосрочная устойчивость.

Применение технологий машинного обучения, систем автоматической обработки, цифровых двойников и открытых форматов уже сегодня позволяет исключать человеческий фактор там, где раньше он был критически важен. Строительство движется в сторону автономности и управляемости, где движение от идеи до реализации проекта можно будет сравнить с навигацией в режиме автопилота: без зависимости от субъективных решений, без необходимости ручного вмешательства на каждом этапе, но с полной цифровой прослеживаемостью и контролем (Рис. 10.2-2).

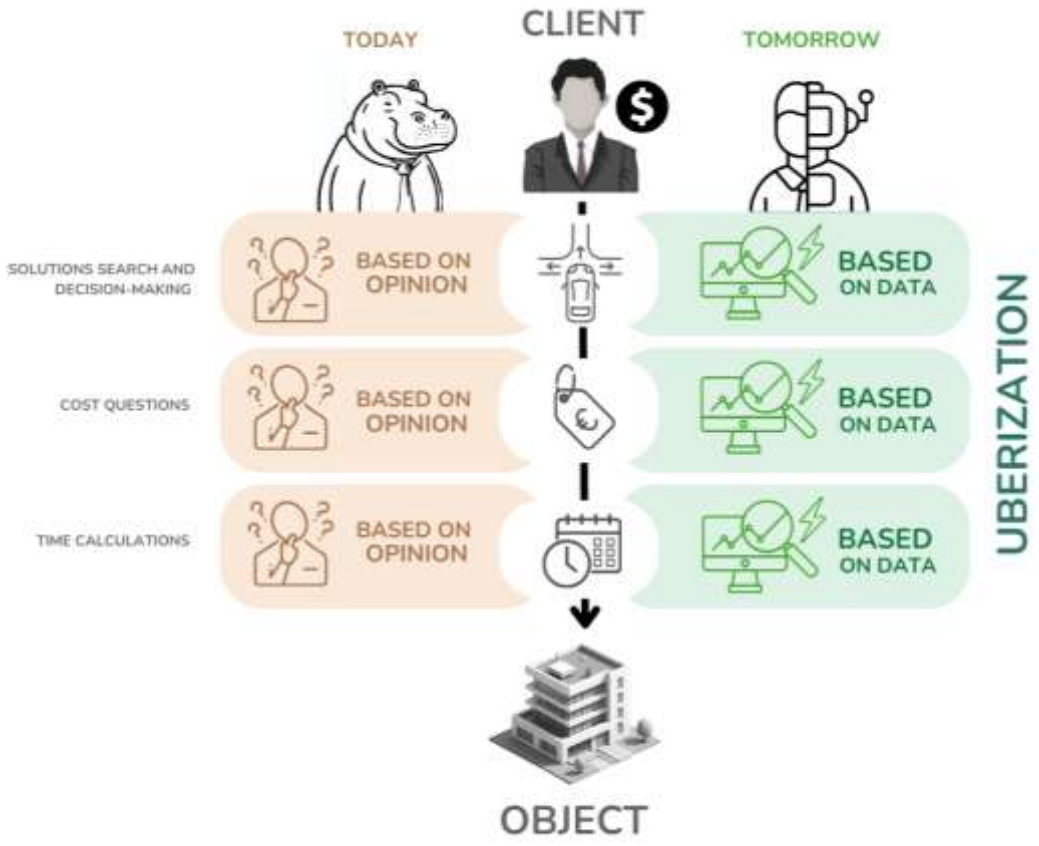


Рис. 10.2-2 Переход от принятия решений на основе мнений важных специалистов (HiPPO) к анализу данных будет в первую очередь продвигаться заказчиком.

Изучив методы, принципы и инструменты, представленные в этой книге, вы сможете начать принимать в своей компании решения, основанные на данных, а не на интуиции. Вы также сможете запускать цепочки модулей в LLM, копировать готовые ETL Pipelines в свою среду разработки (IDE) и автоматически обрабатывать данные, получая информацию в нужной вам форме. В дальнейшем, опираясь на главы книги, посвящённые большим данным и машинному обучению, вы сможете реализовать более сложные сценарии — извлекать новые знания из исторических данных и применять алгоритмы машинного обучения для прогнозирования и оптимизации своих процессов.

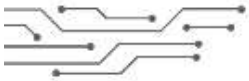
Открытые данные и процессы станут основой для более точных оценок стоимости и сроков реализации проектов, лишая строительные компании возможности спекулировать на непрозрачных данных. Это одновременно и вызов, и возможность для отрасли переосмыслить свою роль и адаптироваться к новой среде, где прозрачность и эффективность станут ключевыми факторами успеха.

> Готовность брать знания и применять их на практике - ключ к успеху в эпоху цифровой трансформации.

Компании, которые осознают это первыми, получат преимущество в условиях новой цифровой конкуренции. Но важно понимать: данные сами по себе ничего не меняют. Многим придется изменить привычный образ мышления, а для этого нужен стимул. Ваша компания должна пересмотреть свой подход к обмену данными.

Меняют компанию — люди, которые умеют работать с этими данными, интерпретировать их, использовать для оптимизации и создавать на их основе новую архитектуру процессов.

Если вы читаете эти строки, вы готовы к переменам и вы уже на шаг впереди. Спасибо, что выбрали этот путь. Добро пожаловать в эпоху цифровой трансформации!



ОБ АВТОРЕ

Меня зовут Артём Бойко. Мой путь на строительной площадке начался в 2007 году — с работы шахтёром на сланцевой шахте, в моём родном городе, во время обучения в Санкт-Петербургском горном университете по специальности «Шахтное и подземное строительство». На обратной стороне обложки данной книги вы можете увидеть взрывника в забое, где мы добывали и взрывали сотни кубов горючего сланца. Моя карьера развивалась в самых разных направлениях — от работы горнорабочим на шахте и на строительстве метрополитена до промышленного альпиниста, монтажника кровель и лифтового оборудования. Мне выпала честь участвовать в проектах самого разного масштаба: от строительства частных домов до крупных промышленных объектов в разных регионах мира.

С течением времени, моя работа сместилась от физического строительства к управлению информацией и цифровыми процессами. С 2013 года я работал на различных должностях в малых, средних и крупных строительных компаниях в нескольких регионах Германии, от проектировщика до менеджера по управлению данными. Что касается управления данными, мой опыт заключается в работе с данными в различных системах ERP, CAD (BIM), MEP, FEM, CMS. Я занимался оптимизацией, автоматизацией процессов, а также анализом, машинным обучением, обработкой данных на этапах планирования, расчетов и выполнения строительных работ в компаниях, занимающихся строительством промышленных, жилых, инфраструктурных и коммунальных объектов.

С 2003 года я работаю с открытым программным обеспечением и открытыми данными. За это время я реализовал множество веб-проектов — от сайтов и интернет-магазинов до полноценных веб-приложений, используя open source-решения и открытые CMS. Эти платформы, во многом схожие с современными строительными ERP, обладают модульной архитектурой, высокой адаптивностью и доступностью. Этот опыт заложил основу моего профессионального подхода — ориентацию на открытые технологии и культуру совместного развития. Уважение к открытому коду и свободному обмену знаниями я стараюсь продвигать в строительной отрасли. Моя работа по повышению доступности данных в строительной отрасли воплотилась в создании нескольких сообществ в социальных сетях для обсуждения вопросов открытости данных и использования Open Source в строительстве, а также в запуске нескольких стартапов, разрабатывающих решения для обеспечения доступа к данным из различных закрытых систем и платформ.

Мой вклад в профессиональное сообщество выражается в участии в роли докладчика на конференциях, охватывающих вопросы интероперабельности CAD (BIM), ERP, 4D-5D, LLM машинного обучения и искусственного интеллекта, а также в статьях, опубликованных в европейских изданиях, посвященных строительной отрасли. Одним из моих заметных достижений является создание "Истории BIM" [111], всеобъемлющей карты важных программных решений для управления данными в строительной отрасли. Моя серия статей из 7 частей "Развитие BIM и лоббистские игры", переведённая на несколько языков, получила широкое признание как попытка осветить скрытую динамику развития цифровых стандартов.

Так я прошёл путь от добычи горной породы — к добыче и систематизации строительных данных. Я всегда открыт к профессиональному диалогу, новым идеям и совместным проектам. С благодарностью приму любую обратную связь и буду рад вашим сообщениям или увидеть вас в числе подписчиков в социальных сетях. Большое вам спасибо, что дочитали эту книгу до конца! Я буду рад, если эта книга поможет вам лучше понять тему данных в строительной отрасли.

ОБРАТНАЯ СВЯЗЬ

Мнение читателей играет важную роль в дальнейшем развитии публикаций и выборе приоритетных тем. Особенно ценны замечания о том, какие идеи оказались полезными, а какие вызвали сомнения и требуют дополнительных пояснений или указаний источников. Книга включает в себя широкий спектр материалов и аналитических оценок, некоторые из которых могут показаться спорными или субъективными. Если в процессе чтения вы обнаружите неточности, неправильно указанные источники, логические несостыковки или опечатки — буду признателен за ваш комментарий, мысли или критику, которые вы можете прислать по адресу: boikoartem@gmail.com. Или через сообщения на LinkedIn: linkedin.com/in/boikoartem

Буду чрезвычайно признателен за любые упоминания о книге Data-Driven Construction в социальных сетях - обмен опытом чтения помогает распространять информацию об открытых данных и инструментах и оказывает поддержку моей работе.

КОММЕНТАРИЙ К ПЕРЕВОДУ

Эта книга была переведена с помощью технологий искусственного интеллекта. Это позволило значительно ускорить процесс перевода. Однако, как и в любой технологической операции, могут возникнуть ошибки или неточности. Если вы заметите что-то, что кажется неверным или некорректно переведённым, пожалуйста, напишите мне. Ваши замечания помогут улучшить качество перевода.

СООБЩЕСТВА DATADRIVENCONSTRUCTION

Это место, где вы можете свободно задавать вопросы и делиться своими проблемами и решениями:

DataDrivenConstruction.io: https://datadrivenconstruction.io

LinkedIn: https://www.linkedin.com/company/datadrivenconstruction/

Twitter: https://twitter.com/datadrivenconst

Телеграмм: https://t.me/datadrivenconstruction

YouTube: https://www.youtube.com/@datadrivenconstruction

ДРУГИЕ НАВЫКИ И КОНЦЕПЦИИ

Помимо ключевых принципов работы с данными в строительной отрасли, в книге DataDrivenConstruction рассматривается широкий спектр дополнительных концепций, программ и навыков, которые необходимы специалисту, работающему с данными. Некоторые из них представлены лишь обзорно, но играют критическую роль в практике.

Заинтересованный читатель может посетить веб-сайт DataDrivenConstruction.io, где представлены ссылки на дополнительные материалы по ключевым навыкам. Эти материалы включают работу с Python и Pandas, построение ETL-процессов, примеры обработки данных в строительных проектах CAD, системы обработки больших данных, а также современные подходы к визуализации и аналитике строительных данных.

При подготовке книги "DataDrivenConstruction" и всех практических примеров использовалось множество инструментов и программного обеспечения с открытым исходным кодом. Автор выражает благодарность разработчикам и соавторам следующих решений:

- Python и Pandas – основа работы с данными и автоматизации
- Scipy, NumPy, Matplotlib и Scikit-Learn – библиотеки для анализа данных и машинного обучения
- SQL и Apache Parquet – инструменты для хранения и обработки больших объемов строительных данных
- Open Source CAD (BIM) открытые инструменты по работе с данными в открытых форматах
- N8n, Apache Airflow, Apache NiFi – системы оркестрации и автоматизации рабочих процессов
- DeepSeek, LlaMa, Mistral – Open Source LLM

Отдельно благодарю всех участников дискуссий по теме открытых данных и инструментов в профессиональных сообществах и соцсетях, чья критика, комментарии и идеи помогли улучшить содержание и структуру этой книги.

Следите за развитием проекта на сайте DataDrivenConstruction.io, где публикуются не только обновления книги и исправления, но и новые главы, обучающие материалы и практические примеры применения описанных методик.

МАКСИМУМ УДОБСТВА С ПЕЧАТНОЙ ВЕРСИЕЙ

Вы держите в руках бесплатную цифровую версию **Data-Driven Construction**. Для более удобной работы и быстрого доступа к материалам рекомендуем обратить внимание на **печатное издание**:

- **Всегда под рукой**: книга в печатном формате станет надежным рабочим инструментом, позволяя быстро найти и использовать нужные визуализации и схемы в любых рабочих ситуациях
- **Высокое качество иллюстраций**: все изображения и графики в печатном издании представлены в максимальном качестве
- **Быстрый доступ к информации**: удобная навигация, возможность делать пометки, закладки и работать с книгой в любом месте.

Приобретая полную печатную версию книги, вы получаете удобный инструмент для комфортной и эффективной работы с информацией: возможность оперативно использовать визуальные материалы в повседневных задачах, быстро находить нужные схемы и делать пометки. Кроме того, ваша покупка поддерживает распространение открытых знаний.

Заказать печатную версию книги можно на: datadrivenconstruction.io/books

УНИКАЛЬНАЯ ВОЗМОЖНОСТЬ ДЛЯ СТРАТЕГИЧЕСКОГО ПОЗИЦИОНИРОВАНИЯ

Предлагаем вам разместить рекламные материалы в бесплатной версии издания DataDrivenConstruction. Платная версия издания за первый год после публикации привлекла внимание специалистов из более чем 50 стран мира — от Латинской Америки до Азиатско-Тихоокеанского региона. Для обсуждения индивидуальных условий сотрудничества и получения детальной информации о возможностях размещения, пожалуйста, заполните форму обратной связи на официальном портале datadrivenconstruction.io или напишите по контактам указанным в конце книги.

ГЛАВЫ КНИГИ ДОСТУПНЫ НА САЙТЕ DATADRIVENCONSTRUCTION.IO

Вы можете читать главы книги Data-Driven Construction на сайте, где постепенно публикуются разделы книги, чтобы вы могли быстро найти нужную информацию и использовать ее в своей работе. Кроме того, на сайте вы найдете много других публикаций по схожим темам, а также примеры приложений и решений, которые помогут вам развивать ваши навыки и применять данные в строительстве.

ПОСЛЕДНИЕ ВЕРСИИ КНИГИ СКАЧИВАЙТЕ С ОФИЦИАЛЬНОГО САЙТА

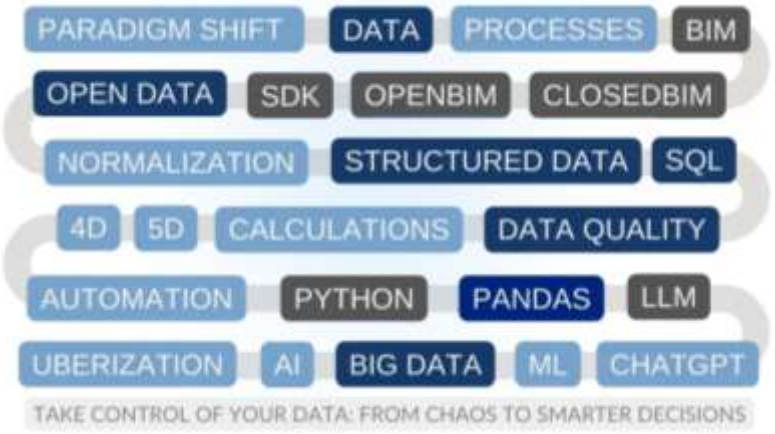
Актуальные и последние версии книги DataDrivenConstruction доступны для скачивания на сайте datadrivenconstruction.io. Если вы хотите получать обновления с новыми главами книги, практические советы или обзоры новых приложений, подпишитесь на рассылку:

- Вы будете первыми ознакомлены с новыми разделами книги
- Получать практические кейсы и советы по аналитике и автоматизации в строительстве
- Следить за трендами, публикациями и примерами приложений

Переходите на datadrivenconstruction.io, чтобы оформить подписку!

DATADRIVENCONSTRUCTION: КОНСАЛТИНГ, ВОРКШОПЫ И ОБУЧЕНИЕ

Обучающие программы DataDrivenConstruction и консультации помогли десяткам ведущих строительных компаний по всему миру повысить эффективность, сократить расходы и улучшить качество решений. Среди клиентов DataDrivenConstruction крупнейшие игроки рынка с оборотом в миллиарды евро, включая строительные, консалтинговые и IT-компании.



Почему стоит выбрать нас?

- **Актуальность:** рассказываем о главных трендах и инсайтах отрасли
- **Практика:** помогаем специалистам эффективно решать повседневные задачи через PoC.
- **Индивидуальный подход:** учитываем особенности вашего бизнеса, обеспечивая максимальную пользу от обучения и консультаций

Основные направления работы команды DataDrivenConstruction:

- **Управление качеством данных:** помогаем параметризировать задачи, собирать требования, проверять и подготавливать данные для автоматизированной обработки.
- **Data Mining - извлечение и структурирование данных:** настраиваем ETL-процессы и извлекаем данные из писем, PDF, Excel, изображений и других источников.
- **BIM и CAD аналитика:** собираем, структурируем и анализируем информацию из файлов RVT, IFC, DWG и других CAD (BIM) форматов.
- **Аналитика и трансформация данных:** превращаем разрозненную информацию в структурированные данные, аналитику, выводы и решения.
- **Интеграция данных и автоматизация процессов:** от автоматизированного создания документов до интеграции с внутренними системами и внешними базами данных.

Свяжитесь с DataDrivenConstruction.io, чтобы узнать, как использование автоматизации может помочь вашей компании достичь ощутимых бизнес-результатов.

ГЛОССАРИЙ

AI (Artificial Intelligence) — искусственный интеллект; способность компьютерных систем выполнять задачи, которые обычно требуют человеческого интеллекта, такие как распознавание образов, обучение и принятие решений.

Apache Airflow — открытая платформа для оркестрации рабочих процессов, позволяющая программно создавать, планировать и отслеживать рабочие процессы и ETL, используя DAG (направленные ациклические графы).

Apache NiFi — инструмент для автоматизации потоков данных между системами, специализирующийся на маршрутизации и преобразовании данных.

Apache Parquet — эффективный формат файлов для колоночного хранения данных, оптимизированный для использования в системах анализа больших данных. Обеспечивает существенное сжатие и быструю обработку.

API (Application Programming Interface) — формализованный интерфейс, позволяющий одной программе взаимодействовать с другой без доступа к исходному коду, обмениваясь данными и функциональностью через стандартизированные запросы и ответы.

Атрибут — характеристика или свойство объекта, описывающее его особенности (например, площадь, объем, стоимость, материал).

Базы данных — организованные структуры для хранения, управления и доступа к информации, используемые для эффективного поиска и обработки данных.

BEP (BIM Execution Plan) — план реализации информационного моделирования зданий, определяющий цели, методы и процессы внедрения BIM в проекте.

Big Data (Большие данные) — массивы информации значительного объёма, разнообразия и скорости обновления, требующие специальных технологий для обработки и анализа.

BI (Business Intelligence) — бизнес-аналитика; процессы, технологии и инструменты для преобразования данных в значимую информацию для принятия решений.

BIM (Building Information Modeling) — информационное моделирование зданий; процесс создания и управления цифровыми представлениями физических и функциональных характеристик объектов строительства, включающий не только 3D-модели, но и информацию о характеристиках, материалах, сроках и стоимости.

BlackBox/WhiteBox — подходы к пониманию системы: в первом случае внутренняя логика скрыта, видны только входы и выходы; во втором – процесс обработки прозрачен и доступен для анализа.

Bounding Box — геометрическая конструкция, описывающая границы объекта в трёхмерном пространстве через минимальные и максимальные координаты по осям X, Y и Z, создавая "коробку" вокруг объекта.

BREP (Boundary Representation) — геометрическое представление объектов, определяющее их через границы поверхностей.

CAD (Computer-Aided Design) — система автоматизированного проектирования, используемая для создания, редактирования и анализа точных чертежей и 3D-моделей в архитектуре, строительстве, машиностроении и других отраслях.

CAFM (Computer-Aided Facility Management) — программное обеспечение для управления объектами недвижимости и инфраструктурой, включающее планирование пространства, управление активами, техническое обслуживание и мониторинг затрат.

CDE (Common Data Environment) — централизованное цифровое пространство для управления, хранения, обмена и совместной работы с проектной информацией на всех этапах жизненного цикла объекта.

Центр передового опыта (Center of Excellence, CoE) — специализированная структура в организации, ответственная за развитие определенной области знаний, разработку стандартов и лучших практик, обучение персонала и поддержку внедрения инноваций.

CoClass — современная система классификации строительных элементов третьего поколения.

Концептуальная модель данных — высокоуровневое представление основных сущностей и их взаимосвязей без детализации атрибутов, используемое на начальных этапах проектирования баз данных.

CRM (Customer Relationship Management) — система управления взаимодействием с клиентами, используемая для автоматизации процессов продаж и обслуживания.

DAG (Directed Acyclic Graph) — направленный ациклический граф, используемый в системах оркестрации данных (Airflow, NiFi) для определения последовательности и зависимостей задач.

Dash — Python-фреймворк для создания интерактивных веб-визуализаций данных.

Дашборд (Dashboard) — информационная панель, визуально представляющая ключевые показатели эффективности и метрики в реальном времени.

Data-Centric подход — методология, которая ставит во главу угла данные, а не приложения или программный код, делая данные центральным активом организации.

Data Governance — комплекс практик, процессов и политик, обеспечивающих надлежащее и эффективное использование данных в организации, включая контроль доступа, качества и безопасности.

Data Lake — хранилище, предназначенное для хранения больших объемов необработанных данных в их исходном формате до момента их использования.

Data Lakehouse — архитектурный подход, объединяющий гибкость и масштабируемость озер данных (Data Lake) с управляемостью и производительностью хранилищ данных (DWH).

Data-Driven Construction — это стратегический подход, при котором каждая стадия жизненного цикла объекта — от проектирования до эксплуатации — поддерживается автоматизированными, взаимосвязанными системами. Такой подход обеспечивает постоянное обучение на основе фактов, снижает неопределённость и позволяет компаниям достигать устойчивого лидерства в отрасли.

Data-Driven интегратор — компания, специализирующаяся на объединении данных из разрозненных источников и их анализе для принятия управленческих решений.

Data-Driven подход (Дата-центричный подход) — методология, при которой данные рассматриваются как стратегический актив, а решения принимаются на основе объективного анализа информации, а не субъективных мнений.

Data Minimalism — подход к сокращению данных до наиболее ценных и значимых, позволяющий упростить обработку и анализ информации.

Data Swamp — разрозненный массив неструктурированных данных, возникающий при неконтролируемом сборе и хранении информации без надлежащей организации.

DataOps — методология, объединяющая принципы DevOps, данные и аналитику, ориентированная на улучшение сотрудничества, интеграции и автоматизации потоков данных.

Дигитализации информации — процесс преобразования всех аспектов строительной деятельности в цифровую форму, пригодную для анализа, интерпретации и автоматизации.

DataFrame (Датафрейм) — двумерная табличная структура данных в библиотеке Pandas, где строки представляют отдельные записи или объекты, а столбцы — их характеристики или атрибуты.

Descriptive Analytics (Описательная аналитика) — анализ исторических данных для понимания того, что произошло в прошлом.

Diagnostic Analytics (Диагностическая аналитика) — анализ данных для определения причин, почему что-то произошло.

Диаграмма Ганта — инструмент планирования проектов, представляющий задачи в виде горизонтальных полос на временной шкале, позволяя наглядно отображать последовательность и продолжительность работ.

DWH (Data Warehouse) — централизованная система хранения данных, которая агрегирует информацию из множества источников, структурирует её и делает доступной для аналитики и отчетности.

ESG (Environmental, Social, Governance) — набор критериев для оценки экологического, социального и управленческого воздействия компании или проекта.

ELT (Extract, Load, Transform) — процесс, при котором данные сначала извлекаются из источников и загружаются в хранилище, а затем преобразуются для аналитических целей.

ETL (Extract, Transform, Load) — процесс извлечения данных из различных источников, их преобразования в нужный формат и загрузки в целевое хранилище для анализа.

ER-диаграмма (Entity-Relationship) — визуальная схема, отображающая сущности, их атрибуты и связи между ними, используемая при моделировании данных.

ERP (Enterprise Resource Planning) — комплексная модульная система планирования ресурсов предприятия, используемая для управления и оптимизации различных аспектов строительного процесса.

Features (Характеристики) — в машинном обучении, независимые переменные или атрибуты, используемые как входные данные для модели.

Физическая модель данных — детальное представление структуры базы данных, включающее таблицы, столбцы, типы данных, ключи и индексы, оптимизированное для конкретной СУБД.

FPDF — библиотека Python для создания PDF-документов.

Геометрическое ядро — программный компонент, предоставляющий базовые алгоритмы для создания, редактирования и анализа геометрических объектов в CAD, BIM и других инженерных приложениях.

HiPPO (Highest Paid Person's Opinion) — подход к принятию решений, основанный на мнении самого высокооплачиваемого человека в организации, а не на объективных данных.

IDE (Integrated Development Environment) — интегрированная среда разработки, комплексный инструмент для написания, тестирования и отладки кода (например, PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) — спецификация передачи информации, определяющая требования к данным на разных этапах проекта.

IFC (Industry Foundation Classes) —формат обмена BIM-данными, обеспечивающий совместимость между различными программными решениями.

Industry 5.0 — концепция развития промышленности, объединяющая возможности цифровизации, автоматизации и искусственного интеллекта с человеческим потенциалом и экологической устойчивостью.

Интеграция данных — процесс объединения данных из разных источников в единую, целостную систему для обеспечения единого представления информации.

Информационные силосы — изолированные системы хранения данных, которые не обмениваются информацией с другими системами, создавая барьеры для эффективного использования данных.

IoT (Internet of Things) — концепция подключения физических объектов к интернету для сбора, обработки и передачи данных.

k-NN (k-Nearest Neighbors) — алгоритм машинного обучения, классифицирующий объекты на основе сходства с ближайшими соседями в обучающей выборке.

Kaggle — платформа для анализа данных и соревнований по машинному обучению.

Калькуляция — расчет стоимости строительных работ или процессов за определенную единицу измерения (например, за 1 м² гипсокартонной стены, 1 м³ бетона).

KPI (Key Performance Indicators) — ключевые показатели эффективности, количественно измеримые метрики, используемые для оценки успешности деятельности компании или конкретного проекта.

Labels (Метки) — в машинном обучении, целевые переменные или атрибуты, которые должна предсказать модель.

Learning Algorithm (Алгоритм обучения) — процесс поиска наилучшей гипотезы в модели, соответствующей целевой функции, используя набор обучающих данных.

Linear Regression (Линейная регрессия) — статистический метод моделирования взаимосвязи между зависимой переменной и одной или несколькими независимыми переменными.

LLM (Large Language Model) — большая языковая модель, искусственный интеллект, обученный понимать и генерировать текст на основе огромных массивов данных, способный анализировать контекст и писать программный код.

LOD (Level of Detail/Development) — уровень детализации модели, определяющий степень геометрической точности и информационного наполнения.

Логическая модель данных — детализированное описание сущностей, атрибутов, ключей и отношений, отражающее бизнес-информацию и правила, промежуточный этап между концептуальной и физической моделями.

Machine Learning (Машинное обучение) — класс методов искусственного интеллекта, позволяющих компьютерным системам обучаться и делать прогнозы на основе данных без явного программирования.

Masterformat — система классификации первого поколения, используемая для структурирования строительных спецификаций по разделам и дисциплинам.

MEP (Mechanical, Electrical, Plumbing) — инженерные системы зданий, включающие механические, электрические и сантехнические компоненты.

Mesh — сетчатое представление 3D-объектов, состоящее из вершин, рёбер и граней.

Model (Модель) — в машинном обучении, набор различных гипотез, одна из которых приближает целевую функцию, которую нужно предсказать или аппроксимировать.

Моделирование данных — процесс создания структурированного представления данных и их взаимосвязей для реализации в информационных системах, включающий концептуальный, логический и физический уровни.

n8n — инструмент с открытым исходным кодом для автоматизации рабочих процессов и интеграции приложений через низко кодовый (low-code) подход.

Нормализация — в машинном обучении, процесс приведения различных числовых данных к единому масштабу для облегчения их обработки и анализа.

Обратный инжиниринг — процесс изучения устройства, функционирования и технологии изготовления объекта путем анализа его структуры, функций и работы. В контексте данных — извлечение информации из проприетарных форматов для использования в открытых системах.

OCR (Optical Character Recognition) — технология оптического распознавания символов, позволяющая преобразовывать изображения текста (отсканированные документы, фотографии) в машиночитаемый текстовый формат.

OmniClass — международный стандарт классификации второго поколения для управления информацией об объектах строительства.

Ontology (Онтология) — система взаимосвязей понятий, формализующая определенную область знаний.

Open Source — модель разработки и распространения программного обеспечения с открытым исходным кодом, доступным для свободного использования, изучения и модификации.

Open BIM — концепция открытого BIM, предполагающая использование открытых стандартов и форматов для обмена данными между различными программными решениями.

Открытые стандарты — публично доступные спецификации для достижения конкретной задачи, которые позволяют различным системам взаимодействовать и обмениваться данными.

Pandas — библиотека Python с открытым исходным кодом для обработки и анализа данных, предоставляющая структуры данных DataFrame и Series для эффективной работы с табличной информацией.

Парадигма открытых данных — подход к обработке данных, при котором информация становится свободно доступной для использования, многократного использования и распространения любым лицом.

Параметрический метод — метод оценки строительных проектов, использующий статистические модели для оценки стоимости на основе параметров проекта.

PIMS (Project Information Model) — это цифровая система, предназначенная для организации, хранения и обмена всей проектной информацией.

Pipeline — последовательность процессов обработки данных, от извлечения и преобразования до анализа и визуализации.

PMIS (Project Information Management System) — система управления проектами, предназначенная для детального контроля выполнения задач на уровне отдельного строительного объекта.

Предиктивная аналитика (Predictive Analytics) — раздел аналитики, который использует статистические методы и машинное обучение для прогнозирования будущих результатов на основе исторических данных.

Прескриптивная аналитика (Prescriptive Analytics) — раздел аналитики, который не только предсказывает будущие результаты, но и предлагает оптимальные действия для достижения желаемых результатов.

Проприетарные форматы — закрытые форматы данных, контролируемые определенной компанией, которые ограничивают возможности обмена информацией и увеличивают зависимость от конкретного программного обеспечения.

QTO (Quantity Take-Off) — процесс извлечения количественных характеристик элементов из проектной документации для расчёта объёмов материалов, необходимых для реализации проекта.

Quality Management System — система управления качеством, обеспечивающая соответствие процессов и результатов установленным требованиям.

RAG (Retrieval-Augmented Generation) — метод, совмещающий генеративные возможности языковых моделей с извлечением релевантной информации из корпоративных баз данных, повышающий точность и актуальность ответов.

RDBMS (Relational Database Management System) — система управления реляционными базами данных, организующая информацию в виде взаимосвязанных таблиц.

RegEx (Regular Expressions) — формализованный язык для поиска и обработки строк, позволяющий задавать шаблоны для проверки текстовых данных на соответствие определённым критериям.

Regression (Регрессия) — статистический метод анализа зависимости между переменными.

Расчеты CO_2 — метод оценки выбросов углекислого газа, связанных с производством и использованием строительных материалов и процессов.

Ресурсный метод — метод составления смет, основанный на подробном анализе всех необходимых ресурсов (материалов, труда, оборудования) для выполнения строительных работ.

RFID (Radio Frequency Identification) — технология автоматической идентификации объектов с помощью радиосигналов, используемая для отслеживания материалов, техники и персонала.

ROI (Return on Investment) — показатель, отражающий соотношение между прибылью и вложенными средствами, используемый для оценки эффективности инвестиций.

SaaS (Software as a Service) — модель предоставления программного обеспечения как услуги, при которой приложения размещаются провайдером и доступны пользователям через интернет.

SCM (Supply Chain Management) — управление цепочками поставок, включающее координацию и оптимизацию всех процессов от закупки материалов до доставки готовой продукции.

Силосы данных — изолированные хранилища информации в организации, которые не интегрированы с другими системами, что затрудняет обмен данными и снижает эффективность.

SQL (Structured Query Language) — язык структурированных запросов, используемый для работы с реляционными базами данных.

SQLite — легкая, встраиваемая, кроссплатформенная СУБД, не требующая отдельного сервера и поддерживающая основные функции SQL, широко используемая в мобильных приложениях и встроенных системах.

Структурированные данные — информация, организованная в определенном формате с четкой структурой, например, в реляционных базах данных или таблицах.

Слабоструктурированные данные — информация с частичной организацией и гибкой структурой, например JSON или XML, где разные элементы могут содержать разные наборы атрибутов.

Сущность (англ. entity) — это конкретный или абстрактный объект реального мира, который можно однозначно выделить, описать и представить в виде данных.

Supervised Learning (Обучение с учителем) — тип машинного обучения, при котором алгоритм обучается на размеченных данных, где для каждого примера известен желаемый результат.

Таксономия — иерархическая система классификации, используемая для систематического распределения элементов по категориям на основе общих признаков.

Titanic Dataset — популярный набор данных для обучения и тестирования моделей машинного обучения.

Training (Обучение) — процесс, в котором алгоритм машинного обучения анализирует данные для выявления закономерностей и формирования модели.

Трансферное обучение — метод машинного обучения, при котором модель, обученная для одной задачи, используется как отправная точка для другой задачи.

Transformation (Преобразование данных) — процесс изменения формата, структуры или содержания данных для их последующего использования.

Требования к данным — формализованные критерии, определяющие структуру, формат, полноту и качество информации, необходимой для поддержки бизнес-процессов.

Уберизация строительной отрасли — процесс трансформации традиционных бизнес-моделей в строительстве под влиянием цифровых платформ, обеспечивающих прямое взаимодействие заказчиков и исполнителей без посредников.

Uniclass — система классификации строительных элементов второго и третьего поколений, широко используемая в Великобритании.

USD (Universal Scene Description) — формат данных, разработанный для компьютерной графики, но получивший применение в инженерных системах благодаря простой структуре и независимости от геометрических ядер.

Валидация данных — процесс проверки информации на соответствие установленным критериям и требованиям, обеспечивающий точность, полноту и согласованность данных.

Vector Database (Векторная база данных) — специализированный тип базы данных, хранящий данные в виде многомерных векторов для эффективного семантического поиска и сравнения объектов.

Векторное представление (эмбеддинг) — метод преобразования данных в многомерные числовые векторы, позволяющий машинным алгоритмам эффективно обрабатывать и анализировать информацию.

VectorOps — методология, ориентированная на обработку, хранение и анализ многомерных векторных данных, особенно актуальная в таких областях как цифровые двойники и семантический поиск.

Visualization (Визуализация) — графическое представление данных для более эффективного восприятия и анализа информации.

Алфавитная разбивка терминов осуществлялась по их названиям на английском языке.

СПИСОК ЛИТЕРАТУРЫ И ОНЛАЙН-МАТЕРИАЛЫ

[1] Gartner, «IT Key Metrics Data 2017: Index of Published Documents and Metrics,» 12 Декабрь 2016. [В Интернете]. Available: https://www.gartner.com/en/documents/3530919. [Дата обращения: 1 Март 2025].

[2] KPMG, «Familiar challenges – new approaches. 2023 Global Construction Survey,» 1 Январь 2023. [В Интернете]. Available: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [Дата обращения: 5 Март 2025].

[3] F. R. Barnard, «A picture is worth a thousand words,» 10 Мапи 1927. [В Интернете]. Available: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Дата обращения: 15 Март 2025].

[4] M. Bastian, «Microsoft CEO Satya Nadella says self-claiming AGI is "nonsensical benchmark hacking",» 21 Февраль 2025. [В Интернете]. Available: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Дата обращения: 15 Март 2025].

[5] W. E. Forum, «Forum Shaping the Future of Construction – A Landscape in Transformation:,» 1 Январь 2016. [В Интернете]. Available: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Дата обращения: 2 Март 2025].

[6] С. Д. Гиллеспи, «Глина: запутанность Земли в эпоху глины,» 2024. [В Интернете]. Available: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] «Папирус III века до н. э. Язык – греческий,» 2024. [В Интернете]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] «Monitoring: making use of the tools which are available,» 1980. [В Интернете]. Available: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Дата обращения: 15 Март 2025].

[9] PWC, «Data driven What students need to succeed in a rapidly changing business world,» 15 Февраль 2015. [В Интернете]. Available: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Дата обращения: 15 Март 2025].

[10] Skanska USA, «Fall Construction Market Trends,» 2 Ноябрь 2023. [В Интернете]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] «Oxford Essential Quotations (4 ed.),» Oxford University Press, 2016. [В Интернете]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236. [Дата обращения: 1 Март 2025].

[12] «Quote: Sondergaard on Data Analytics,» [В Интернете]. Available: https://www.causeweb.org/cause/resources/library/r2493. [Дата обращения: 15 Март 2025].

[13] «How global AI interest is boosting the data management market,» 28 Май 2024. [В Интернете]. Available: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Дата обращения: 15 Март 2025].

[14] И. Маккью, «История ERP,» 2024. [В Интернете]. Available: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, «ERP Price: How much does an ERP system cost?,» [В Интернете]. Available: https://erpscout.de/en/erp-costs/. [Дата обращения: 15 Март 2025].

[16] softwarepath, «What 1,384 ERP projects tell us about selecting ERP (2022 ERP report),» 18 Январь 2022. [В Интернете]. Available: https://softwarepath.com/guides/erp-report. [Дата обращения: 15 Март 2025].

[17] Deloitte, «Data-Driven Management in Digital Capital Projects,» 16 Декабрь 2016. [В Интернете]. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Дата обращения: 1 Март 2025].

[18] Mckinsey, «The data-driven enterprise of 2025,» 28 Январь 2022. [В Интернете]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Дата обращения: 22 Май 2024].

[19] Wikipedia, «Moore's law,» [В Интернете]. Available: https://en.wikipedia.org/wiki/Moore%27s_law. [Дата обращения: 15 Март 2025].

[20] Accenture, «Building More Value With Capital Projects,» 1 Январь 2020. [В Интернете]. Available: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Дата обращения: 3 Март 2024].

[21] Б. Марр, «Сколько данных мы создаем каждый день? The Mind-Blowing Stats Everyone Should Read,» 2018. [В Интернете]. Available: 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] «Сколько данных производится каждый день?,» 2024. [В Интернете]. Available: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] Т. Салливан, «ИИ и глобальная "датасфера": сколько информации будет у человечества к 2025 году?,» 2024. [В Интернете]. Available: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-

2025.html.

[24] Statista, «Total number of printed books produced in various regions of Western Europe in each half century between 1454 and 1800,» [В Интернете]. Available: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Дата обращения: 1 Март 2025].

[25] «Примеры ценообразования,» 2024. [В Интернете]. Available: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, «Enterprises outsource data storage as complexity rises,» 10 Май 2024. [В Интернете]. Available: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Дата обращения: 15 Март 2025].

[27] JETSOFTPRO, «SaaS is Dead? Microsoft CEO’s Shocking Prediction Explained,» 13 Январь 2025. [В Интернете]. Available: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, «Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner,» 12 Декабрь 2024. [В Интернете]. Available: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Дата обращения: 15 Март 2025].

[29] GoodReads, «Tim Berners-Lee,» [В Интернете]. Available: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Дата обращения: 15 Март 2025].

[30] KPMG, «Cue Construction 4.0: Make-or-Break Time,» 1 Январь 2023. [В Интернете]. Available: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Дата обращения: 5 Март 2025].

[31] И. Дайнингер, Б. Кох, Р. Баукнехт и М. Лангханс, «Использование цифровых моделей для декарбонизации производственной площадки: Пример соединения модели здания, производственной модели и энергетической модели,» 2024. [В Интернете]. Available: 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 Февраль 2017. [В Интернете]. Available: 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 Октябрь 2014. [В Интернете]. Available: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

[34] Forbes, «Without An Opinion, You're Just Another Person With Data,» 15 Март 2016. [В Интернете].

Available: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/. [Дата обращения: 15 Март 2025].

[35] Wikiquote, «Charles Babbage,» [В Интернете]. Available: https://en.wikiquote.org/wiki/Charles_Babbage. [Дата обращения: 15 Март 2025].

[36] SAP, «New Research Finds That Nearly Half of Executives Trust AI Over Themselves,» 12 Март 2025. [В Интернете]. Available: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Дата обращения: 15 Март 2025].

[37] The Canadian Construction Association and KPMG in Canada, 2021, «Construction in a digital world,» 1 Май 2021. [В Интернете]. Available: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Дата обращения: 5 Март 2025].

[38] ЗЦС, «Decoding the Fifth Industrial Revolution,» [В Интернете]. Available: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Дата обращения: 15 Март 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, «The Value of Open Source Software,» 24 Январь 2024. [В Интернете]. Available: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Дата обращения: 15 Март 2025].

[41] Naval Center for Cost Analysis Air Force Cost Analysis Agency, «Software Development Cost Estimating Handbook,» 1 Сентябрь 2008. [В Интернете]. Available: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, «Improving construction productivity,» [В Интернете]. Available: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Дата обращения: 15 Март 2025].

[43] A. G. a. C. Syverson, «The Strange and Awful Path of Productivity in the US Construction Sector,» 19 Январь 2023. [В Интернете]. Available: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Дата обращения: 1 Март 2025].

[44] McKinsey, «Delivering on construction productivity is no longer optional,» 9 Август 2024. [В Интернете]. Available: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Дата обращения: 5 Март 2025].

[45] ING Group, «Lagging productivity in construction is driving up building costs,» 12 Декабрь 2022. [В Интернете]. Available: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/. [Дата обращения: 15 Март 2025].

[46] M. Berman, «Microsoft CEO’s Shocking Prediction: “Agents Will Replace ALL Software",» 19 Декабрь 2024. [В Интернете]. Available: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Дата обращения: 15 Март 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 Март 2025. [В Интернете]. Available: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Дата обращения: 30 Март 2025].

[48] Statista, «Popularity comparison of database management systems (DBMSs) worldwide as of June 2024, by category,» Июнь 2024. [В Интернете]. Available: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Дата обращения: 15 Март 2025].

[49] DB-Engines, «DB-Engines Ranking,» [В Интернете]. Available: https://db-engines.com/en/ranking. [Дата обращения: 15 Март 2025].

[50] «Stack Overflow Developer Survey 2023,» 2024. [В Интернете]. Available: https://survey.stackoverflow.co/2023/.

[51] «SQL,» 2024. [В Интернете]. Available: https://en.wikipedia.org/wiki/SQL.

[52] «Структурированные и неструктурированные данные: What's the Difference?,» 2024. [В Интернете]. Available: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, «PDF COMPARISON OF DATA FORMATS FOR CONSTRUCTION PROJECTS,» 23 Апрель 2024. [В Интернете]. Available: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] «Building Information Modeling Whitepaper site,» 2003. [В Интернете]. Available: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] А. Бойко, «Лоббистские войны и развитие BIM. Часть 5: BlackRock – мастер всех технологий. Как корпорации контролируют открытый исходный код,» 2024. [В Интернете]. Available: 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. [В Интернете]. Available: https://isicad.net/articles.php?article_num=14805. [Дата обращения: 02 2025].

[57] C. Eastman и A. Cthers, «Eastman, Charles; And Cthers,» Сентябрь 1974. [В Интернете]. Available: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Дата обращения: 15 Март 2025].

[58] D. Ushakov, «Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies,» 11 Ноябрь 2011. [В Интернете]. Available: https://isicad.net/articles.php?article_num=14805. [Дата обращения: 15 Март 2025].

[59] D. Weisberg, «History of CAD,» 12 Декабрь 2022. [В Интернете]. Available: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Дата обращения: 15 Март 2025].

[60] ADSK, «White Paper Building Information Modeling,» 2002. [В Интернете]. Available: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Дата обращения: 15 Март 2025].

[61] ADSK, «White Paper Building Information Modeling in Practice,» [В Интернете]. Available: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Дата обращения: 15 Март 2025].

[62] А. Бойко, «Лоббистские войны и развитие BIM. Часть 2: открытый BIM VS закрытый BIM. Европа VS остальной мир,» 2024. [В Интернете]. Available: 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. Бойко, «Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs,» 2024. [В Интернете]. Available: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, «Whitepaper BIM,» 2002. [В Интернете]. Available: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Дата обращения: 15 Март 2025].

[65] ADSK, «Integrated Design-Through-Manufacturing: Benefits and Rationale,» [В Интернете]. Available: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Дата обращения: 15 Март 2025].

[66] M. Шеклетт, «Структурированные и неструктурированные данные: Ключевые различия,» 2024. [В Интернете]. Available: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Вулард, «Осмысление роста неструктурированных данных,» 2024. [В Интернете]. Available: 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. [В Интернете]. Available: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf . [Дата обращения: 02 2025].

[69] CrowdFlower, «Data Science Report 2016,» 2016. [В Интернете]. Available: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Дата обращения: 15 Март 2025].

[70] Analyticsindiamag, «6 Most Time Consuming Task For Data Scientists,» 15 Май 2019. [В Интернете]. Available: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, «Report: Data scientists spend bulk of time cleaning up,» 06 Июль 2015. [В Интернете]. Available: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html. [Дата обращения: 5 Март 2025].

[72] S. Hawking, «Science AMA Series: Stephen Hawking AMA Answers!,» 27 Июль 2015. [В Интернете]. Available: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Дата обращения: 15 Март 2025].

[73] Б. Сайферс и К. Доктороу, «Конфиденциальность без монополии: Защита данных и совместимость,» 2024. [В Интернете]. Available: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, «Open data: Unlocking innovation and performance with liquid information,» 1 Октябрь 2013. [В Интернете]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Дата обращения: 15 Март 2025].

[75] А. Бойко, «The struggle for open data in the construction industry. The history of AUTOLISP, intelliCAD, openDWG, ODA and openCASCADE,» 15 05 2024. [В Интернете]. Available: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Дата обращения: 16 02 2025].

[76] Wikipedia, «Microsoft and open source,» [В Интернете]. Available: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Дата обращения: 15 Март 2025].

[77] TIME, «The Gap Between Open and Closed AI Models Might Be Shrinking. Here’s Why That Matters,» 5 Ноябрь 2024. [В Интернете]. Available: https://time.com/7171962/open-closed-ai-models-epoch/. [Дата обращения: 15 Март 2025].

[78] The Verge, «More than a quarter of new code at Google is generated by AI,» 29 Октябрь 2024. [В Интернете]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Дата обращения: 15 Март 2025].

[79] McKinsey Digital, «The business case for using GPUs to accelerate analytics processing,» 15 Декабрь 2020. [В Интернете]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-

processing. [Дата обращения: 15 Март 2025].

[80] PWC, «PwC Open Source Monitor 2019,» 2019. [В Интернете]. Available: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [Дата обращения: 15 Март 2025].

[81] Travers Smith, «The Open Secret: Open Source Software,» 2024. [В Интернете]. Available: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Дата обращения: 15 Март 2025].

[82] Deloitte, «The data transfer process in corporate transformations,» 2021. [В Интернете]. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf. [Дата обращения: 15 Март 2025].

[83] gov.uk, «Data Analytics and AI in Government Project Delivery,» 20 Март 2024. [В Интернете]. Available: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Дата обращения: 5 Март 2025].

[84] «Quote Origin: Everything Should Be Made as Simple as Possible, But Not Simpler,» 13 Май 2011. [В Интернете]. Available: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Дата обращения: 15 Март 2025].

[85] «Transformer (deep learning architecture),» [В Интернете]. Available: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Дата обращения: 15 Март 2025].

[86] «Python Packages Download Stats,» 2024. [В Интернете]. Available: https://www.pepy.tech/projects/pandas.

[87] Interview Bit, «Top 10 Python Libraries,» 2023. [В Интернете]. Available: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Дата обращения: 30 Март 2025].

[88] «NVIDIA and HP Supercharge Data Science and Generative AI on Workstations,» 7 Март 2025. [В Интернете]. Available: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Дата обращения: 15 Март 2025].

[89] P. Орак, «Как обработать DataFrame с миллионами строк за секунды,» 2024. [В Интернете]. Available: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, «What is Kaggle?,» 2024. [В Интернете]. Available: https://www.datacamp.com/blog/what-is-kaggle.

[91] «NVIDIA CEO Jensen Huang Keynote at COMPUTEX 2024,» 2 Июнь 2024. [В Интернете]. Available: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [Дата обращения: 15 Март 2025].

[92] «Members: Основатели и корпоративные члены,» 2024. [В Интернете]. Available: https://www.opendesign.com/member-showcase.

[93] A. Бойко, «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 Ноябрь 2024. [В Интернете]. Available: 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. [Дата обращения: 23 февраль 2025].

[94] A. Бойко, «The post-BIM world. Transition to data and processes and whether the construction industry needs semantics, formats and interoperability,» 20 декабрь 2024. [В Интернете]. Available: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Дата обращения: 23 февраль 2025].

[95] N. I. o. Health, «NIH STRATEGIC PLAN FOR DATA SCIENCE,» 2016. [В Интернете]. Available: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Дата обращения: 23 февраль 2025].

[96] Harvard Business Review, «Bad Data Costs the U.S. $3 Trillion Per Year,» 22 Сентябрь 2016. [В Интернете]. Available: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, «Impacts of Data Quality,» 1 Январь 2025. [В Интернете]. Available: 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,» [В Интернете]. Available: https://www.wbdg.org/resources/design-for-maintainability. [Дата обращения: 15 Март 2025].

[99] O. o. D. C. P. a. Oversight, «Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment,» Апрель 2014. [В Интернете]. Available: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Дата обращения: 15 Март 2025].

[100] Gartner, «Data Quality: Best Practices for Accurate Insights,» 1 Январь 2025. [В Интернете]. Available: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] «For Want of a Nail,» [В Интернете]. Available: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Дата обращения: 15 Март 2025].

[102] McKinsey Global Institute, «Open data: Unlocking innovation and performance with liquid information,» Октябрь 2013. [В Интернете]. Available: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20i

nformation/mgi_open_data_fullreport_oct2013.pdf. [Дата обращения: 15 Март 2025].

[103] EY, «Путь углеродной нейтральности,» 10 Март 2023. [В Интернете]. Available: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Дата обращения: 15 Март 2025].

[104] PWC, «ESG Awareness,» 1 Июль 2024. [В Интернете]. Available: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Дата обращения: 15 Март 2025].

[105] G. Hammond, «Embodied Carbon - The Inventory of Carbon and Energy (ICE),» 2024. [В Интернете]. Available: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] «CO2_calculating the embodied carbon,» 2024. [В Интернете]. Available: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, «Imagining Construction’s Digital Future,» 24 Июнь 2016. [В Интернете]. Available: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Дата обращения: 25 Февраль 2025].

[108] Bund der Steuerzahler Deutschland e.V., «Das Schwarzbuch,» 10 Октябрь 2024. [В Интернете]. Available: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Дата обращения: 15 Март 2025].

[109] SAS, «Data lake and data warehouse – know the difference,» [В Интернете]. Available: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Дата обращения: 15 Март 2025].

[110] ADSK, «Building Information Modeling,» 2002. [В Интернете]. Available: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Дата обращения: 15 Март 2025].

[111] А. Бойко, «Карта истории BIM,» 2024. [В Интернете]. Available: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, «Definitions of BIM by Organizations and Standards,» 27 Декабрь 2023. [В Интернете]. Available: https://encyclopedia.pub/entry/53149. [Дата обращения: 5 Март 2025].

[113] CAD Vendor, «OPEN BIM Program,» 2012. [В Интернете]. Available: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Дата обращения: 30 Март 2025].

[114] Wikipedia, «Industry Foundation Classes,» [В Интернете]. Available: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Дата обращения: 15 Март 2025].

[115] Wikipedia, «IGES,» [В Интернете]. Available: https://en.wikipedia.org/wiki/IGES. [Дата обращения: 30 Март 2025].

[116] А. Бойко, «History of CAD (BIM),» 15 Декабрь 2021. [В Интернете]. Available: https://miro.com/app/board/o9J_laML2cs=/. [Дата обращения: 24 Февраль 2025].

[117] T. K. K. A. O. F. B. C. E. L. H. H. E. L. P. N. S. H. T. J. v. L. 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 Январь 2021. [В Интернете]. Available: https://arxiv.org/pdf/2007.10951. [Дата обращения: 5 Март 2025].

[118] И. Рогачёв, «Поговорим за BIM: Максим Нечипоренко | Renga | IFC | Отечественный BIM,» 13 Апрель 2021. [В Интернете]. Available: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be. [Дата обращения: 5 Март 2025].

[119] D. Ares, «RETS in Real Estate: Why It’s Crucial for Efficiency & Growth,» 17 Декабрь 2024. [В Интернете]. Available: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Дата обращения: 5 Март 2025].

[120] «Flex token cost,» 2024. [В Интернете]. Available: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] А. Бойко, «Забудьте о BIM и демократизируйте доступ к данным (17. Kolloquium Investor – Hochschule – Bauindustrie),» 2024. [В Интернете]. Available: https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagment-und-Bay-Bauindustrie.pdf.

[122] Д. Хилл, Д. Фолдези, С. Феррер, М. Фридман, Э. Лох и Ф. Плашке, «Решение загадки производительности строительной отрасли,» 2015. [В Интернете]. Available: 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 Январь 2018. [В Интернете]. Available: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Дата обращения: 2 Март 2025].

[124] Apple.com, «Pixar, Adobe, Apple and NVIDIA form Alliance for OpenUSD to drive open standards for 3D content,» 1 Август 2023. [В Интернете]. Available: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [Дата обращения: 2 Март 2025].

[125] AECmag, «ADSK’s granular data strategy,» 25 Июль 2024. [В Интернете]. Available: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Дата обращения: 15 Март 2025].

[126] A. Бойко, «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. [В Интернете]. Available: 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. [Дата обращения: 23 февряллb 2025].

[127] A. Бойко, «ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction,» 7 Август 2024. [В Интернете]. Available: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Дата обращения: 15 март 2025].

[128] «Jeffrey Zeldman Presents,» 6 Май 2008. [В Интернете]. Available: https://zeldman.com/2008/05/06/content-precedes-design/. [Дата обращения: 15 Март 2025].

[129] A. Бойко, «DWG Analyse with ChatGPT | DataDrivenConstruction,» 5 Март 2024. [В Интернете]. Available: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Дата обращения: 15 Март 2025].

[130] McKinsey , «The McKinsey guide to outcompeting in the age of digital and AI,» 2023. [В Интернете]. Available: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Дата обращения: 30 Март 2025].

[131] Forbes, «Data Storytelling: The Essential Data Science Skill Everyone Needs,» 31 Март 2016. [В Интернете]. Available: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Дата обращения: 15 Март 2025].

[132] J. Bertin, «Graphics and Graphic Information Processing,» 8 Сентябрь 2011. [В Интернете]. Available: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Дата обращения: 15 Март 2025].

[133] CauseWeb, «Wells/Wilks on Statistical Thinking,» [В Интернете]. Available: https://www.causeweb.org/cause/resources/library/r1266. [Дата обращения: 15 Март 2025].

[134] Ministrymagazine, «How science discovered Creation,» Январь 1986. [В Интернете]. Available: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Дата обращения: 15 Март 2025].

[135] BCG, «Data-Driven Transformation: Accelerate at Scale Now,» 23 Май 2017. [В Интернете]. Available: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [Дата обращения: 15 Май 2024].

[136] «How to build a data architecture to drive innovation—today and tomorrow,» 3 Июнь 2020. [В Интернете]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Дата обращения: 15 Март 2025].

[137] Oxford, «Woodrow Wilson 1856–1924,» [В Интернете]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Дата обращения: 15 Март 2025].

[138] «Convertors,» 2024. [В Интернете]. Available: 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 Январь 2017. [В Интернете]. Available: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Дата обращения: 18 Февраль 2025].

[140] «Трубопровод в строительстве,» 2024. [В Интернете]. Available: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, «Apache NiFi,» 1 Январь 2025. [В Интернете]. Available: https://de.wikipedia.org/wiki/Apache_NiFi. [Дата обращения: 5 Март 2025].

[142] n8n, «Gmail AI Auto-Responder: Create Draft Replies to incoming emails,» 1 Май 2024. [В Интернете]. Available: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/. [Дата обращения: 15 Март 2025].

[143] n8n, «Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail,» 1 Март 2025. [В Интернете]. Available: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Дата обращения: 15 Март 2025].

[144] B. T. O'Neill, «Failure rates for analytics, AI, and big data projects = 85% – yikes!,» 1 Январь 2025. [В Интернете]. Available: 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 Январь 2025. [В Интернете]. Available: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Дата обращения: 1 Май 2024].

[147] Т. Ландсалл-Велфэр, Прогнозирование настроения нации в настоящее время, Significance, 2012.

[148] А. Бойко, «Сан-Франциско. Строительный сектор 1980-2019,» 2024. [В Интернете]. Available: https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019.

[149] А. Бойко, «Kaggle: RVT IFC Files 5000 Projects,» 2024. [В Интернете]. Available: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, «Preparing for the Future with Connected Construction,» [В Интернете]. Available: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Дата обращения: 15 Март 2025].

[151] Cisco, «Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing,» 22 Май 2017. [В Интернете]. Available: 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] «Conditions Required for Plant Fossil Preservation,» 2024. [В Интернете]. Available: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] «Финк из BlackRock об облигациях, слияниях и поглощениях, рецессии в США, выборах: Full Interview,» 2023. [В Интернете]. Available: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, «12 famous AI disasters,» 02 Октябрь 2024. [В Интернете]. Available: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Дата обращения: 15 Март 2025].

[155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.

[156] Wikipedia, «Kaggle,» 1 Январь 2025. [В Интернете]. Available: https://en.wikipedia.org/wiki/Kaggle. [Дата обращения: 15 Март 2025].

[157] Kaggle, «Titanic - Machine Learning from Disaster,» 1 Январь 2025. [В Интернете]. Available: https://www.kaggle.com/competitions/titanic/overview. [Дата обращения: Марта 10 2025].

[158] Ш. Йохри, «Создание ChatGPT: От данных к диалогу,» 2024. [В Интернете]. Available: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Домингос, «Несколько полезных вещей, которые нужно знать о машинном обучении,» 2024. [В Интернете]. Available: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, «Quotable Quote,» [В Интернете]. Available: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Дата обращения: 17 Март 2025].

[161] NVIDIA, «Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models,» 19 Декабрь 2024. [В Интернете]. Available: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [Дата обращения: 25 Март 2025].

[162] «NVIDIA Announces Major Release of Cosmos World Foundation Models and Physical AI Data Tools,» 18 Март 2025. [В Интернете]. Available: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Дата обращения: 25 Март 2025].

[163] NVIDIA, «NVIDIA Isaac Sim,» [В Интернете]. Available: https://developer.nvidia.com/isaac/sim. [Дата обращения: 25 Март 2025].

[164] M. Quarterly, «Why digital strategies fail,» 25 Январь 2018. [В Интернете]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail. [Дата обращения: 15 Марта 2025].

[165] M. J. Perry, «My Favorite Milton Friedman Quotes,» 17 Ноябрь 2006. [В Интернете]. Available: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Дата обращения: 1 Март 2025].

[166] J. A. Wheeler, «Information, physics, quantum: the search for links,» 1990.

[169] А. Бойко, «Лоббистские войны и развитие BIM. Часть 5: BlackRock – мастер всех технологий. Как корпорации контролируют открытый исходный код,» 2024. [В Интернете]. Available: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen и J. Beetz, «A SPARQL query engine for binary-formatted IFC building models,» *Advanced Engineering Informatics,* 2024.

[171] «Количество предприятий в строительном секторе в Великобритании в 2021 году, по размеру предприятия,» 2024. [В Интернете]. Available: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] «5000 проектов IFC&RVT,» 2024. [В Интернете]. Available: 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. [В Интернете]. Available: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Дата обращения: Февраль 2025].

[174] *Lobbying Wars Over Data in Construction | Techno-Feudalism and the History of BIM's Hidden Past.* [Фильм]. Germany: Артём Бойко, 2023.

[175] А. Бойко, «CHATGPT WITH REVIT AND IFC | Automatic retrieval of documents and data from projects,» 16 Ноябрь 2023. [В Интернете]. Available: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Дата обращения: 2 Март 2025].

[176] M. &. Company, «Three new mandates for capturing a digital transformation’s full value,» 22 Январь 2022. [В Интернете]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Дата обращения: 15 Февраль 2025].

[177] KPMG, «Construction in a Digital World,» 1 Май 2021. [В Интернете]. Available: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-

report-en.pdf. [Дата обращения: 5 Апрель 2024].

[178] LLP, KPMG, «Cue Construction 4.0: Make-or-Break Time.,» 17 Март 2023. [В Интернете]. Available: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Дата обращения: 15 Февраль 2025].

[179] O. Business, «Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models,» 10 Январь 2025. [В Интернете]. Available: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Дата обращения: 15 Март 2025].

[180] Forbes, «Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says,» 23 Март 2016. [В Интернете]. Available: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Дата обращения: 15 Март 2025].

[181] Министерство иностранных дел, по делам Содружества и развития Великобритании, «Digital development strategy 2024 to 2030,» 18 Март 2024. [В Интернете]. Available: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Дата обращения: 15 Март 2025].

[182] «Vision and Strategy in the Building Design Industry,» 7 Ноябрь 2003. [В Интернете]. Available: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Дата обращения: 5 Март 2025].

[183] М. Бочаров, «Информационное моделирование,» Март 2025. [В Интернете]. Available: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Дата обращения: 15 Март 2025].

[184] «Integrated Design-Through-Manufacturing: Benefits and Rationale,» 2000. [В Интернете]. Available: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Дата обращения: 25 Март 2025].

[185] CAD Vendor, «Open BIM Program is a marketing campaign,» 12 Март 2012. [В Интернете]. Available: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Дата обращения: 30 Март 2025].

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

УЗНАЙТЕ, КАК ДАННЫЕ ТРАНСФОРМИРУЮТ СТРОИТЕЛЬСТВО

Что находится внутри

- Более 100 ключевых тем, касающихся данных в AEC
- Более 300 уникальных визуализаций и диаграмм
- Более 50 реальных бизнес-кейсов
- Практические применения LLM и ИИ
- Примеры кода и готовые к применению рабочие процессы

Темы

- Строительство, основанное на данных
- Цифровая трансформация в архитектуре, инженерии и строительстве
- Анализ данных и автоматизация
- Качество данных и управление ими
- САПР, BIM и совместимость информации
- Магистратура по искусственному интеллекту и машинному обучению в строительстве
- Прогнозирование стоимости и сроков проекта

Аудитория

- Руководители проектов в строительстве
- Архитекторы и строительные инженеры
- Координаторы BIM и менеджеры по данным
- Лидеры в области цифровой трансформации
- Специалисты в области информационных технологий и искусственного интеллекта в сфере AEC
- Градостроители и эксперты по устойчивому развитию
- Студенты архитектурных и инженерных направлений
- Преподаватели и ученые

- **Веб-сайт**
- **www.datadrivenconstruction.io.**

Комментарии к первому изданию:

«Бойко — Джеймс Карвилл ИТ — в его часто цитируемом «It's the economy, stupid» для этой знаменитой книги достаточно заменить всего одно слово. «It's the data, stupid» (а не программное обеспечение). А чтобы найти свой путь во вселенной данных, поговорка древних римлян, восходящая к греческому языку, актуальна и сегодня: «Navigare necesse est». Автор ведет своих читателей по всем глубинам и отмелям океана данных уверенной рукой и непоколебимым компасом, не говоря уже о всеобъемлющем историческом подходе и, что не менее важно, весьма оригинальной графике...»

— Доктор Буркхард Талебитари

«Книга Артема Бойко представляет собой значимый этап в процесс демократизации цифровизации строительной отрасли и является настоящим переломным моментом для малых и средних предприятий (МСП). Это произведение — призыв к действию! Оно служит ценным руководством для тех, кто стремится не только понять цифровую трансформацию в строительстве, но и активно участвовать в ее формировании — прагматично, эффективно и с дальновидным подходом. Настало время объединить усилия для обмена знаниями и устойчивого повышения производительности строительной отрасли...»

— Доктор Майкл Макс Бюлер

Это практическое руководство помогает как профессионалам, так и новичкам ориентироваться в стремительно меняющемся мире строительства на основе данных. От основ управления данными до продвинутых цифровых рабочих процессов, инструментов ИИ и практических приложений — эта книга станет вашей дорожной картой к более умным, быстрым и эффективным строительным процессам.

ISBN: 978-3-98262-559-1