

12 июня 2025

Отраслевой стандарт и «единая версия правды»: как меняется работа с данными

В цифровой трансформации данных фокус сместился с количества на качество, внимание усилилось к безопасности и внедрению процессов и инструментов для интеграции, извлечения, преобразования и перемещения данных. Сбор и аналитика нереляционных данных позволила снизить простои в производстве на несколько десятков процентов, а также оптимизировать расходы на электроэнергию и логистику. Об этом сообщили эксперты.

ТРЕНДЫ И ПРОБЛЕМЫ



Бизнес стремится к эффективности и повышению качества принимаемых решений: первая достигается за счет автоматизации и оптимизации различных процессов, а второе возможно благодаря data-driven подходу и «единой версии правды».

Руководитель направления Data Warehousing/Machine Learning «ДАР» (входит в ГК «КОРУС Консалтинг»),
Александр Зенькович

Также на тренды в области сбора и обработки данных оказывает влияние растущая конкуренция на многих рынках: бизнес стремится к максимальной персонализации и улучшению клиентского опыта, а для этого необходимо собирать как можно больше данных и правильно работать с ними. Наконец, развитие технологий, которые позволяют дополнительно монетизировать данные, например, интернет вещей, видео- и аудио-аналитика, также является значимым фактором, отметил эксперт.

В последние годы мы наблюдаем рост спроса на сбор нереляционных данных (NoSQL), в том числе в онлайн. Это массовый сбор телеметрии (технология удаленного сбора, передачи и анализа данных от различных устройств или объектов), компьютерное зрение, например, для контроля качества на производстве, анализа трафика в ритейле. Сбор и анализ таких данных позволяет снизить простои в производстве на несколько десятков процентов, а также оптимизировать расходы на электроэнергию и логистику. Для обработки и анализа данных бизнес намного чаще стал использовать облака или гибридные архитектуры. Компании внедряют DataMesh-подход (децентрализованный гибкий подход к работе распределенных команд и распространению информации) с децентрализацией владения данными по доменам. Также растет спрос на потоковую аналитику (Kafka, Flink и прочие инструменты) и полноценные MLOps-системы для индустриализации AI/ML и ускорения получения бизнес-эффекта.

Александр Зенькович,

Руководитель направления Data Warehousing/Machine Learning «ДАР» (входит в ГК «КОРУС Консалтинг»)

В последние два-три года фокус заказчиков был на двух вещах, подчеркнул директор департамента «Облака и данные» Рексофт Руслан Заединов. Первая — сохранение доступа к данным. Поставщики данных, особенно у

промышленных заказчиков, — это контроллеры и другое оборудование уровня АСУ ТП (автоматизированная система управления технологическим процессом, — прим. ред.), которые были и во многом остаются зарубежного производства. В этом контексте ключевая проблема в том, что эти устройства часто защищены паролем, и даже у конечного заказчика нет доступа к их настройкам, что создает трудности в обслуживании и перепрограммировании. Преодоление этого барьера, особенно после ухода производителей, потребовало больших усилий и широко обсуждалось на рынке, подчеркнул эксперт.

«Второй момент касается перехода на российские решения в части сбора данных, включая как аппаратную, так и программную части. Речь идет, в частности, о телематике, когда данные получаются удаленно через публичные сети связи, часто с транспорта или удаленных точек. Эффективность и надежность этих решений на первых этапах оставались под вопросом, что стало своего рода своеобразными «детскими болезнями роста», и отняло у заказчиков много сил и времени. Мы участвовали в этом процессе, помогая заказчикам выявлять особенности российского аппаратного обеспечения, телематических устройств и модулей, на практике сталкиваясь с возникающими сложностями. Сохранение доступа к данным и переход на российские решения по части источников данных — это, на мой взгляд, самый главный технологический вызов, который решался в стране за последние годы», — полагает Руслан Заединов.

Наблюдаются несколько значительных изменений: во-первых, в процессы сбора данных стали внедрять этапы проверки их качества, включая

предварительную (pre-validation) и последующую (post-validation) валидацию, отметил руководитель отдела разработки контентных модулей компании VolgaBlob, разработчика платформы для анализа и обработки данных Тимофей Мельников. Ключевыми критериями качества, на которые обращают внимание, стали актуальность, полнота и достоверность — именно эти факторы наиболее критичны для бизнеса. Во-вторых, процесс сбора данных все чаще включает в себя шаги защиты данных. Это могут быть организационные меры, такие как согласование и предоставление доступа, а также технические меры — маскирование, шифрование, токенизация, пояснил Тимофей Мельников.

«Для этапов сбора и обработки данных появляется все больше процессов и инструментов Data Management. В частности, сейчас внедряются каталоги данных для организованной инвентаризации активов на основе метаданных и инструменты Data Lineage, которые описывают происхождение и изменение данных. Помимо этого, формируются процессы работы с мастер-данными для исключения избыточности и обеспечения высокого качества общих данных, а на каждый этап их сбора назначаются ответственные, что повышает прозрачность и контроль. Эффективность зависит от того, насколько процессы и инструменты настроены в соответствии со зрелостью компании, а также не являются ли они избыточными. При их правильной настройке бизнес может сократить издержки, связанные с принятием решений на основе некачественных данных, избежать репутационных потерь из-за утечек и ускорить развитие хранилища данных, выводя его на новую ступень», — рассказал Тимофей Мельников.

АТАКУ ОБНАРУЖИТ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Цифровая трансформация в сфере данных набирает темп благодаря ряду ключевых факторов: формированию развитых дата-экосистем, устойчивому партнерству бизнеса и государства, росту требований к технологическому суверенитету и спросу на персонализированные сервисы, отметили в пресс-службе Ассоциации больших данных (АБД).

«В последние годы бизнес сосредоточен на замене западного программного обеспечения — компании массово переходят на Open Source и решения российских вендоров. Одновременно, на фоне ускоренной цифровизации, возрос спрос на решения в области кибербезопасности: внедряются современные архитектуры защиты данных, повышается зрелость подходов и ответственность за защиту данных. Так, например, в АБД мы разработали Отраслевой стандарт защиты данных, к которому уже присоединился ряд ведущих российских компаний. Он помогает бизнесу повышать доверие, обеспечивать безопасность и системно совершенствовать практики защиты. Параллельно развиваются технологии повышения конфиденциальности — пока преимущественно на уровне пилотов и экспериментов, но с большим потенциалом. Логичным продолжением этого тренда стоит отметить растущий интерес к синтетическим данным как стратегической альтернативе обезличиванию. Таким образом, мы можем наблюдать, что рынок данных и ИИ движется к зрелой модели, где эффективность цифровых решений напрямую зависит от доверия к инфраструктуре и уровню безопасности», —



пояснили в Ассоциации больших данных.

В технологических сетях, где генерируется основной объем данных, традиционно применяется периметральная защита, этот подход давно и повсеместно внедрен — трудно представить современное предприятие, где бы он отсутствовал, отметил Руслан Заединов. Заказчики продолжают успешно использовать эту модель, и тому есть веские причины. В отличие от финтеха и смежных областей, где утечки данных часто происходят из-за человеческого фактора и социальной инженерии, в чисто технологических процессах и производстве периметральная защита демонстрирует высокую эффективность. По сути, эта технически выверенная система не требует кардинальных усовершенствований, тем не менее есть тенденции развития, уточнил эксперт.

«Наблюдаются определенные тенденции развития, например, внедрение методов искусственного интеллекта для обнаружения атак. Этим направлением активно занимаются российские компании, многие из которых уже вышли из стадии стартапов. Наряду с этим отмечается рост спроса на услуги аудита и консалтинга в сфере информационной безопасности. Это связано с частичным переходом на отечественные решения в области сетевой инфраструктуры и передачи данных. Российские разработки имеют свои особенности и точки интеграции с системами защиты, что обуславливает потребность как в технологическом, так и в организационном сопровождении. Разработка новых методик и внедрение современных средств защиты оставались в приоритете у заказчиков на протяжении последних двух-трех лет», — сообщил Руслан Заединов.

Тимофей Мельников выделил три драйвера цифровой трансформации сбора и обработки данных:

Data Quality — смещение фокуса с количества данных на их качество, Data Security — усиление внимание к безопасности данных, включая меры по предотвращению утечек и Data Management — внедрение процессов и инструментов для интеграции, извлечения, преобразования и перемещения данных.

«Для этапов сбора и обработки данных появляется все больше процессов и инструментов Data Management. В частности, сейчас внедряются каталоги данных для организованной инвентаризации активов на основе метаданных и инструменты Data Lineage, которые описывают происхождение и изменение данных. Помимо этого, формируются процессы работы с мастер-данными для исключения избыточности и обеспечения высокого качества общих данных, а на каждый этап их сбора назначаются ответственные, что повышает прозрачность и контроль», — пояснил Тимофей Мельников.

Эффективность зависит от того, насколько процессы и инструменты настроены в соответствии со зрелостью компании, а также не являются ли они избыточными. При их правильной настройке бизнес может сократить издержки, связанные с принятием решений на основе некачественных данных, избежать репутационных потерь из-за утечек и ускорить развитие хранилища данных, выводя его на новую ступень, резюмировал эксперт.

Роль данных в бизнесе увеличивается, всё больше компаний переходят от data-driven к датацентричному подходу, при котором данные — один из основных и неотъемлемых ресурсов: такой подход стимулирует бизнес внедрять цифровые технологии, менять процессы для повышения своей эффективности и конкурентоспособности, так, особое внимание здесь уделяют инструментам сбора и обработки данных, рассказал менеджер продуктов в Nexign, разработчике высокотехнологичных enterprise-решений для различных отраслей экономики Дмитрий Лемеш.

«Активнее начинает использоваться ИИ (AI/ML) в инструментах сбора и обработки данных. Интеграция этой технологии в системы ETL/ESB помогает не только улучшить качество обработки данных, но и значительно ускорить процесс, повышая его точность и эффективность. ИИ также задействован в настройке интеграций между системами, тестировании, документировании и мониторинге. Уже сейчас это позволяет значительно снизить расходы на запуск новых интеграций и сократить time to market для новых сервисов», — рассказал Дмитрий Лемеш.

Еще недавно на рынке был распространен такой подход: внедряем систему сбора, хранения и обработки данных, а далее думаем, как ее защитить. Сейчас произошел переход к Privacy by Design — встраивание приватности на этапе проектирования решений для минимизации и анонимизации данных. Безусловно, это снижает риски утечек. Сюда же можно отнести правило расширенного доступа — точный контроль того, кто и к каким данным имеет доступ на основе атрибутов (это минимизирует внутренние угрозы утечек).

Александр Зенькович,

Руководитель направления Data Warehousing/Machine

Learning «ДАР» (входит в ГК «КОРУС Консалтинг»)

Бизнес постепенно начинает применять Data Governance-подход, который позволяет классифицировать данные, отслеживать потоки (Data Lineage) и актуализировать использование и востребованность данных, что добавляет прозрачности, позволяет автоматизировать аудит работы с данными и повысить эффективность политики безопасности, резюмировал эксперт.