

Влияние качества данных в бизнесе 2/42/24



Старший специалист по проектам автоматизации, PwC Латвия
Anta Ošiņa



Старший консультант в сфере услуг по налоговой отчетности и бухгалтерскому учёту, PwC Латвия

Artis Vizbelis

В мире бизнеса данным придается огромное значение, однако их подлинная ценность кроется не столько в объеме, сколько в качестве. Данные плохого качества могут препятствовать развитию вашего бизнеса, а также стать причиной принятия неверных решений и утраты возможностей. В данной статье мы углубимся в практические аспекты управления качеством данных, чтобы помочь раскрыть истинный потенциал информации и принимать решения, основанные на надежных и точных данных.

Глобальное видение

Некачественные данные могут вызывать существенные финансовые последствия для предприятий. Например, в 2022 году компания Unity Software понесла убытки, достигшие 110 миллионов долларов США в виде поступлений и 4,2 миллиарда долларов США в виде рыночной стоимости, получив неточные данные от крупного клиента. Исследования IBM свидетельствуют о том, что некачественные данные ежегодно обходятся экономике США приблизительно в 3,1 триллиона долларов. Однако последствия плохого качества данных не ограничиваются финансовым ущербом – они также оказывают влияние на возможности инноваций и репутацию предприятий. Исследования Experian Data Quality показывают, что лишь 3% данных предприятия соответствуют стандартам качества и 90% данных не используются в стратегических целях. Перечисленные факты подчеркивают, насколько критично качество данных в современной бизнес-среде и насколько существенны инвестиции в улучшение качества данных.

Практические стратегии повышения качества данных

Управление качеством данных не является универсальным решением. Оно требует адаптированного подхода, соответствующего уникальным проблемам и потребностям каждой организации. Рассмотрим некоторые проверенные стратегии, позволяющие улучшить качество данных.

Внедрение системы управления основными данными (Master Data Management, или MDM)

MDM позволяет создать единый, надежный источник важнейших бизнес-данных, обеспечивая последовательность и точность в разных системах и отделах. Это помогает устранить раздробленность данных и уменьшить количество ошибок. В число популярных решений MDM входят SAP Master Data Governance, Informatica MDM и IBM InfoSphere MDM.

Стандартизация данных

Используйте средства профилирования (Extract, Transform, Load) данных, например Talend или Informatica Data Quality, чтобы выявлять и исправлять ошибки, расхождения и дубликаты, а также генерировать условия стандартизации данных. Условия стандартизации данных обеспечивают

соответствие данных, определяя форматы, условия именования и критерии валидации конкретных элементов данных. Например, условие может требовать, чтобы все адреса клиентов соответствовали определенному формату или чтобы все коды изделий представляли собой буквенно-цифровую комбинацию.

Валидация и обогащение данных

Используйте средства валидации данных, например Great Expectations или Apache Spark, чтобы обеспечить целостность данных. Apache Spark можно применять для разработки процедур валидации (pipelines), которые выполняют проверку полноты, последовательности и точности данных. При этом они способны обогащать данные, объединяя их с внешними наборами данных или используя модели машинного обучения, чтобы получить новые выводы.

Показатели качества данных

Используя средства надзора за качеством данных, например Collibra или Ataccama, можно отслеживать основные показатели данных и определять области, требующие улучшения. Регулярный мониторинг качества данных позволяет своевременно выявлять проблемы и предпринимать корректирующие действия, тем самым обеспечивая постоянное соответствие данных потребностям и стратегическим целям вашего бизнеса.

Основные показатели качества данных и их влияние

Показатель	Значения в среднем наборе данных	Последствия в случае низкого показателя
Полнота – какая часть записей содержит все необходимые значения.	80–95%	Неполнота информации может привести к принятию ошибочных решений и неэффективной деятельности. Например, если в базе данных клиентов не хватает контактной информации, это может затруднить обслуживание и общение с клиентами.
Точность – какая часть записей соответствует действительности и поддается проверке.	90–99%	Неточные данные могут привести к искажению финансовой отчетности или иметь юридические последствия. Например, если клиент в бухгалтерской системе неверно определен как предприятие ЕС с зарегистрированным номером НДС, к которому должна применяться нулевая ставка НДС, но фактически является физическим лицом, к которому применяется НДС по ставке 21%, это может привести к неверному начислению налогов и неправильно выписанным счетам. Такая ошибка может повлечь не только финансовые последствия, но и юридические проблемы у предприятия, если будут обнаружены несоответствующие налоговые платежи. Расхождения данных могут вызвать неясность и затруднить процесс принятия решений. Например, если цена товара на платформе э-коммерции ниже себестоимости, указанной в бухгалтерской системе предприятия, это может причинить убытки, поскольку предприятие будет продавать товар по цене ниже фактических расходов на него. Если цена на платформе завышена, это может оказать негативное влияние на объемы продаж. Поэтому важно обеспечивать последовательность как цен, так и учетных данных.
Последовательность – каков уровень согласованности значений данных в разных системах.	95–99%	

Актуальность – насколько часто обновляются данные.	Зависит от вида данных, но обычно данные считаются актуальными, если обновлялись в течение предыдущих 24 часов.	Устаревшие данные могут стать причиной принятия неверных решений и утраты возможностей. Например, устаревшая информация об акционных ценах может привести к невыгодным вложениям.
Уникальность – процентная доля уникальных записей.	95-100%	Дубликаты данных могут вызвать неясность, затруднить анализ и привести к ошибочным выводам. Например, если в базе данных клиентов есть несколько записей об одном и том же клиенте, это может вызвать проблемы при обслуживании клиентов и персонализации маркетинговых кампаний.

Указанные показатели качества данных подчеркивают, насколько важно обеспечить, чтобы данные, которыми вы располагаете, были не только объемными, но и точными, полными, последовательными, актуальными и уникальными. Только тогда вы сможете полностью задействовать потенциал своих данных и принимать решения, исходя из надежной информации.

Качество данных – катализатор успеха вашего бизнеса

В современной конкурентной среде качество данных уже не роскошь, а стратегическое требование. Отдавая приоритет качеству данных и внедряя практики проактивного управления, предприятия могут использовать потенциал своих данных, содействовать инновациям и достигать устойчивого успеха. Сделайте качество данных привычкой в своей организации!