

УДК 004.6

DOI 10.51691/2541-8327_2022_12_27

СТРАТЕГИЯ УПРАВЛЕНИЯ КАЧЕСТВОМ ДАННЫХ

Микаева А.С.

к.э.н., доцент,

МИРЭА - Российский технологический университет,

Москва, Россия

Смехнова С.Н.

студент,

Российская академия народного хозяйства и государственной службы при

Президенте Российской Федерации,

Москва, Россия

Аннотация

В статье рассматриваются теоретические аспекты стратегии управления качеством данных и инструменты для обеспечения и повышения качества данных. Приводится пример политики по работе с данными ПАО «Сбербанк России».

Ключевые слова: качество данных, информация, управление данными, стратегия, инструменты.

DATA QUALITY MANAGEMENT STRATEGY

Mikaeva A.S.

Candidate of Economics, Associate Professor,

MIREA - Russian Technological University,

Moscow, Russia

Smekhnova S.N.

student,

The Russian Presidential Academy of National Economy and Public Administration,

Moscow, Russia

Abstract

The article discusses the theoretical aspects of the data quality management strategy and tools for ensuring and improving data quality. An example of the Sberbank of Russia data policy is given.

Keywords: data quality, information, data management, strategy, instruments.

Данные – движущая сила большинства процессов в науке, экономике и бизнесе, на основе которой принимаются управленческие решения. Однако не все данные пригодны для работы, для получения достоверных результатов они должны соответствовать качеству [3]. Согласно ГОСТу Р ИСО 8000-2-2019 качество данных – это степень, с которой набор характеристик, присущих данным, отвечает требованиям, которые представлены на рисунке 1.



Рис. 1 – Требования к качеству данных [2]

Качество данных один из важнейших вопросов для компаний. Низкое качество данных не позволяет получить достоверные результаты от эксплуатации сложных и дорогих систем интеллектуальной автоматизации. Рассмотрим распространенные проблемы, связанные с несоответствием требованиям к качеству данных, на примерах.

Недостоверная информация: производитель считает, что ему известно точное местоположение грузовика, перевозящего его готовую продукцию с производственной площадки в распределительный центр. В компании

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

оптимизируют маршрутизацию, оценивают время доставки. И оказывается, что данные о местоположении неверны, из-за чего грузовик прибывает позже, что нарушает нормальный рабочий процесс в распределительном центре [3].

Неполные данные: в компании работают над оптимизацией управления цепочками поставок. Для того, чтобы оценить поставщиков и понять, какие из них дисциплинированы и заслуживают доверия, отслеживается время доставки. Но в отличие от запланированного времени, поле фактического времени доставки не является обязательным, и сотрудники склада обычно забывают его заполнить. Без этой важной информации (имея неполные данные), компания не сможет понять, как работают ее поставщики.

Устаревшая информация: устаревшие данные могут привести к неправильной сегментации клиентов, плохому знанию рынка и упущенной прибыли. Например, покупатель однажды заполнил анкету на сайте и заявил, что у него нет детей. Однако прошло время – и теперь у него есть ребенок. Покупатель готов потратить свой бюджет на товары для детей, но продавец об этом не знает, так как по результатам опроса клиент не входит в сегмент «Клиенты с детьми».

Поскольку последствия низкого качества данных могут оказаться разрушительными, важно своевременно устранять перечисленные проблемы и последовательно управлять качеством данных.

Управление качеством данных - согласованная деятельность по контролю и управлению организацией, имеющей непосредственное отношение к качеству данных [2]. Другими словами, это внедрение систематической структуры, которая постоянно профилирует источники данных, проверяет качество информации и выполняет ряд процессов для устранения ошибок в качестве данных, делая данные более точными, правильными, достоверными, полными и надежными.

Основной задачей управления качеством данных является удовлетворение требований к измерениям качества данных в соответствии с потребностями бизнеса. Стратегия по управлению качеством данных должна гарантировать, что качество данных соответствует потребностям бизнеса, иначе ресурсы будут потрачены впустую без какой-либо выгоды. Понимание текущих и будущих потребностей бизнеса играет важную роль в постоянном улучшении качества данных [2].

Повышение качества данных - первоочередная задача компании, поэтому все сотрудники должны понимать проблемы, связанные с низким качеством данных. Ответственность за качество несет каждый из сотрудников организации. Высокое качество данных достигается за счет понимания и управления всей деловой и технической деятельностью как взаимосвязанными процессами, функционирующими как целостная экосистема. Комплексная и успешная программа управления качеством данных должна учитывать все бизнес и технические процессы, которые собирают, производят, поддерживают, преобразуют, распространяют и уничтожают данные. Понимание того, как эти процессы взаимодействуют друг с другом и какие результаты они дают, позволит организации оптимизировать свои результаты [4].



Рис. 2 - Место управления качеством данных в общем спектре работ с корпоративной информацией [3]

Можно заметить, что деятельность по управлению качеством в общем объеме работ по управлению корпоративными данными мала (рис. 2). Однако, если в озере данных будут отсутствовать данные хотя бы от одного подразделения, то такое озеро потеряет эффективность. Несмотря на то, что вклад информации именно из этого подразделения может быть меньше 1% от общего объема всех источников.

Помимо этого, нужно понимать, что управление качеством данных - не разовое усилие, а непрерывный процесс. Требуется регулярно пересматривать политики и правила в области качества данных с целью их постоянного улучшения. Это необходимо, поскольку бизнес-среда постоянно меняется.

Поскольку требования и характеристики качества данных индивидуальны, то управление качеством данных также различается на предприятиях. Всё зависит от множества факторов: размера компании, объема данных, задействованных источников. Однако существуют основные принципы внедрения и управления качеством данных [1].

Для решения проблем, связанных с обеспечением качества данных, используются различные комбинации инструментов. Можно выделить следующие категории:

- инструменты синтаксического анализа и стандартизации. Они разбивают данные на компоненты и приводят их к унифицированному формату;
- средства очистки. С их помощью удаляют неправильные или дублирующиеся записи данных или изменяют значения в соответствии с определенными правилами и стандартами;
- инструменты сопоставления. Они объединяют тесно связанные записи данных;
- инструменты профилирования. Они собирают статистику о данных, а затем используют ее для оценки качества данных;
- инструменты мониторинга. Обеспечивают контроль за качеством данных [1].

Рассмотрим пример политики по работе с данными ПАО «Сбербанк России». Она строится на утвержденном правлением банка нормативном документе, своде правил, вокруг которого выстраиваются организационная вертикаль CDO, набор процессов управления данными, роли, их функции и взаимодействие друг с другом. В данном документе прописаны и структурированы основные моменты по работе с данными в банке.

Для удобства исследований данных был создан сервис «Рабочее место D-people», к D-people относят дата-инженеров, дата-сайентистов, дата-аналитиков. Главная цель данного сервиса – повышение эффективности сотрудников. Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

Единое рабочее место позволило сократить время создания и вывода моделей в промышленную эксплуатацию.

При создании «Рабочего места D-people» Сбербанк в первую очередь сосредоточился на том, чтобы сократить время на получение результата, выстроив новую цепочку процессов и совместив их с инструментарием, компоненты которого интегрированы друг с другом. Был сделан супермаркет данных, позволяющий заказать любые доступные в банке данные. Таким образом были выполнены требования к качеству данных: актуальность, своевременность.

На следующем шаге в сервисе были собраны все необходимые инструменты. Рабочее место наполнено всеми современными инструментами обработки данных (spark, hive и пр.), UI (hue, Jupyter Hub) и библиотеками.

Таким образом цикл работы D-people был сжат до одного дня, когда раньше на эту работу требовалось от 3 месяцев.

На данный момент загружено более 270 внешних и внутренних источников данных, а также внешние справочники. Что позволяет качеству данных отвечать целостности, полноте, точности и надежности. Более 600 сотрудников Сбера уже постоянно работают с этой инфраструктурой, свыше 1,3 тыс. подписок на данные поставляются регулярно [5].

Подводя итог вышесказанному, можно сделать вывод, что продуманная стратегия управления данными помогает компаниям получить потенциальные преимущества перед своими конкурентами по бизнесу, как за счет повышения операционной эффективности, так и за счет более эффективного принятия решений. Организации с хорошо управляемыми данными более гибкие, что позволяет выявлять тенденции рынка и быстрее использовать новые возможности. Также эффективное управление данными помогает избежать утечек данных, проблем с конфиденциальностью и соблюдением нормативных

требований, которые могут нанести ущерб репутации организации, увеличить непредвиденные расходы.

Библиографический список:

1. ГОСТ Р 59547-2021 Защита информации. Мониторинг информационной безопасности. Общие положения - Введ. впервые 02.08.2021. – Москва: Российский институт стандартизации, 2021
2. ГОСТ Р ИСО 8000-2-2019 Качество данных. Часть 2. Словарь – Взамен ГОСТ Р ИСО 8000-2-2014; Введ. 29.10. 2019. – Москва: Стандартиформ, 2019
3. Незнанов, А.А. Качество данных: от стратегии к практике / А.А. Незнанов, Д.В. Волков // Конференция «Качество данных 2020». – 2020 [Электронный ресурс] – Режим доступа – URL: <https://www.osp.ru/os/2020/01/13055348>, свободный (Дата обращения 01.12.2022)
4. Смирнов Н. Качество данных — это про людей / Н. Смирнов // Конференция «Качество данных 2021». – 2021 [Электронный ресурс] – Режим доступа – URL: <https://cio.osp.ru/articles/100321-Kachestvo-dannyh--eto-pro-lyudey>, свободный (Дата обращения 10.12.2022)
5. Шеян, И. «Цифровые люди» Сбера / И. Шеян // Конференция «Качество данных 2021»: – 2021 [Электронный ресурс] – Режим доступа – URL: <https://cio.osp.ru/articles/040221-Tsifrovye-lyudi-Sbera>, свободный (Дата обращения 10.12.2022)

Оригинальность 89%