

25 апреля 2024

В океане данных

Национальный проект «Экономика данных» предполагает создание суверенной инфраструктуры для хранения, сбора и управления данными. Залог его успеха – появление современных систем управления данными. В РЖД успешно реализован проект создания и внедрения такой системы – КСУД. О том какие задачи он решил, какие сложности необходимо было преодолеть при создании системы, и как КСУД будет развиваться рассказывает **Ирина Долженко, руководитель проекта, главный эксперт департамента информатизации ОАО «РЖД».**

ЧТО ИЗМЕНИЛА СИСТЕМА КСУД, КАК ВОЗМОЖНОСТИ ДЛЯ УПРАВЛЕНИЯ ДАННЫМИ ОНА ОТКРЫВАЕТ?

КСУД – является той необходимой для компании основой для перехода к управлению на основе данных.

Комплекс инструментов системы позволяет разрабатывать аналитические формы с использованием единого методологического подхода, а также получать информацию о данных смежных подразделений и переиспользовать уже существующие формы в новых системах бизнес-аналитики (BI). Таким образом снижается трудоемкость множества операций,



связанных с работой с данными.

Кроме того, КСУД предоставляет возможность оперативно согласовывать методологию при работе с показателями и аналитическими формами, сотрудники эффективнее взаимодействуют и быстрее решают задачи. Экономия времени и трудозатрат возникает за счет переиспользования промежуточных результатов в ряде этапов внедрения КХД и BI.

Система является базисом для моделирования и построения предиктивных моделей. Для построения предиктивных моделей необходимы качественные данные с единой методологией, над чем мы работали и продолжим работать, масштабируя созданные решения и методологические практики на все бизнес-подразделения РЖД.

КАК ВЛИЯЕТ ВНЕДРЕНИЕ КСУД НА ОПЕРАЦИОННУЮ ДЕЯТЕЛЬНОСТЬ СОТРУДНИКОВ РЖД?

Все специалисты, начиная от руководителей Топ-уровня до рядовых пользователей, теперь работают по единой модели данных. Такой подход помогает сделать сбор информации о методологии формирования показателей единообразным, у сотрудников смежных подразделений снизился риск возникновения бизнес-ошибок, вызванных неверным пониманием бизнес-терминов.

КСУД позволяет сформировать отраслевую аналитику в реальном режиме времени с высокой степенью детализации – всё это возможно благодаря новым стандартам качества и прогрессивным подходам к классификации и хранению данных.

Комплекс инструментов по управлению данными дает возможность разрабатывать аналитические отчеты и формы с использованием единого методологического подхода. Снижается трудоемкость множества операций, связанных с работой с данными.

При этом КСУД исключает использование недостоверных, повторяющихся, невыверенных данных, как с точки зрения методологии, так и с точки зрения систем-источников, в повседневной работе при формировании отчетов, справочно-аналитических материалов для руководства РЖД.

КАК ОРГАНИЗОВАТЬ КОНТРОЛЬ ЗА ХРАНЕНИЕМ КРИТИЧЕСКИ ВАЖНЫХ ДАННЫХ И УПРАВЛЕНИЕМ ИМИ?

Для организации контроля за хранением критически важных данных и управлением ими необходимо разработать строгие политики и процедуры по управлению данными, определить ответственных сотрудников за их выполнение, проводить регулярные аудиты и мониторинг доступа к данным. В РЖД критически важные данные описываются в единых утвержденных форматах. Это модели данных (методологическое бизнес-описание

показателей) и технических метаданных (технологическое и техническое описание всех этапов обработки и формирования данных). Они ведутся в инструментах комплекса систем управления и анализа данных (КС УАД).

Инструменты комплекса систем управления и анализа данных (КС УАД)

- Система для ведения моделей данных - Глоссарий бизнес-терминов.
- Система для ведения моделей метаданных - Банк метаданных.
- Нормативная база при работе с данными - Репозиторий форм и аналитических справок
- Связи моделей данных и моделей метаданных - Диаграмм жизненного цикла и потоков данных

Также стоит предпринять ряд шагов для организации контроля за хранением критически важных данных и управлением ими. В частности, сформировать реестр критичных данных, регламенты по доступности, безопасности, включая ролевые модели, к наличию копий в случае возникновения аварий, организовать процессы мониторинга по качеству, доступности, безопасности критически важных данных.

КСУД СИЛЬНО ИЗМЕНИЛ СОТРУДНИЧЕСТВО С ГОСУДАРСТВЕННЫМИ ОРГАНАМИ? ЗАКЛЮЧЕНИЕ В КОНЦЕ ПРОШЛОГО ГОДА СОГЛАШЕНИЕ С ФЕДЕРАЛЬНОЙ СЛУЖБОЙ СТАТИСТИКИ (РОССТАТОМ)

В ЧАСТИ ОБМЕНА ДАННЫМИ ИЗМЕНИЛО РАБОТУ РЖД?

КСУД позволяет взаимодействовать с государственными органами, а также обмениваться с внешними партнерами качественными данными.

Мы имеем возможность построить данное сотрудничество на нескольких ключевых принципах. В том числе, перейти на автоматический обмен данными между системами, на потоковое представление показателей, ввести единое цифровое окно сбора показателей и данных, автоматизировать согласование каталогов назначений статистической отчетности.

Эти шаги повышают возможности использования административных данных для целей федерального статистического наблюдения и для внутренних целей компании.

Подписанное в прошлом году Соглашение о взаимодействии Росстата и РЖД расширяет возможности для обмена данными. В частности, оно предусматривает организацию обмена данными по принципу «система – система». Сформированный канал позволяет перейти к потоковой передаче данных по показателям.

ЛЕГКО ЛИ БЫЛО СОТРУДНИКАМ КОМПАНИИ ОВЛАДЕТЬ ВОЗМОЖНОСТЯМИ, КОТОРЫЕ ДАЕТ НОВАЯ СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ?

В РЖД количество аналитиков, работающих с отчетностью, достигает 9000 чел. Для эффективной реализации проекта КСУД важно сформировать команду, которая объединит внутреннюю экспертизу и создаст новые компетенции в области управления данными.

Для создания новых компетенций мы проводим обучение для наших специалистов и развиваем внутренний кадровый резерв. Так в 2023 году в Корпоративном университете РЖД прошло обучение для работников компании по новой программе «Бизнес-аналитика». В течение трех месяцев 45 руководителей и специалистов из различных подразделений погружались в теорию и практику работы с данными.

Программа обучения включала разработку проектных идей, основанных на применении аналитического подхода и комплексной работе с корпоративными хранилищами данных. В качестве основного источника данных слушателями активно использовалась система «СКИМ», также разработанная в РЖД. Она предназначена для оперативного мониторинга производственных, операционных и финансово-экономических показателей компании. «СКИМ» обеспечивает единый подход к формированию оперативной управленческой отчетности, что на порядок повышает скорость подготовки данных.

Мы уже получаем много запросов на участие в следующем потоке. Верю в эту программу и ее масштабирование, так как в рамках программы мы не только обучаем, но и объединяем лучших экспертов нашей компании.

Организованная коммуникация и обмен опытом профессионалов в области

аналитики определенно даст огромный эффект в работе с данными.

КАКИЕ СПЕЦИФИЧЕСКИЕ ПРИХОДИЛОСЬ РЕШАТЬ ПРИ РАЗРАБОТКЕ КСУД?

При построении корпоративной системы управления данными (КСУД) мы стремились объединить все лучшие практики. Важно было использовать все актуальные инструменты управления данными — прежде всего программные решения, предназначенные для администрирования, мониторинга, контроля и составления отчетов на протяжении всего жизненного цикла данных.

Но наш подход обладал и рядом уникальных особенностей. Например, мы начали работу именно с разработки такого инструмента управления данными, как глоссарий бизнес-терминов. Такой путь был более сложным. В РЖД большое количество бизнес-направлений и, следовательно, бизнес-терминов, которые требовали единого понимания. Зато в результате такого подхода мы смогли создать структурированные и согласованные с бизнес-процессами правила работы с данными и определить приоритеты и задачи в области ИТ-инфраструктуры.

Кроме того, прорабатывая функциональные требования к Глоссарию бизнес-терминов мы увидели, что при увеличении количества показателей, введенных в систему, растут и трудозатраты на их ведение, что снижало качество и доступность данных. Мы поняли, что Глоссарий должен быть не простой библиотекой карточек, а должен обладать автоматизированной

функцией ведения описаний показателей и за счет среды согласования (workflow) обеспечивать их согласованность. В результате мы не только разрабатываем инструмент, а решаем гораздо более важную для компании задачу - создаем площадку для общения и наращивания экспертизы всех специалистов, причастных к аналитике. Это уникальное решение, не имеющее аналогов на рынке.

Востребованные программные решения для систем управления данными

- 1. Бизнес-гlossарий:** Интерфейс для работы с бизнес-терминами пользователями без специализированной подготовки и возможностью согласования бизнес-терминов.
- 2. Дата каталог:** Количество доступных коннекторов (без доработок) к источникам данных для извлечения технических метаданных. Дополнительно оцениваются: возможность комментирования извлечённых метаданных, как отдельных таблиц, так и витрин данных; версионность метаданных; возможность профилирования; возможность построения моделей данных.
- 3. Дата линедж:** Уровень детализации, возможность построения связей между разными слоями данных, а также разбор скриптов трансформации
- 4. Управление качеством данных:** Возможность самостоятельной настройки проверок с помощью конструктора (минимизации программирования), отображения ошибок в данных в интерфейсе системы и их детализация, нотификация о найденных ошибках, визуальных вид дашбордов показывающих основные показатели по качеству данных по настроенным проверкам.

ЧТО СЕГОДНЯ ДОЛЖНА «УМЕТЬ» СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ, ЧТОБЫ ОТВЕЧАТЬ СОВРЕМЕННЫМ ТРЕБОВАНИЯ И ПОТРЕБНОСТЯМ?

Сегодня существующие инструменты управления данными, такие как например, бизнес-глоссарий, и даже продукты open source, которые изначально делали упор описание только технических метаданных (дата каталог), расширили свою функциональность до возможности описания бизнес-терминов с несколькими параметрами.

Управление именно бизнес-терминами стало наиболее актуальным в последнее время. Возможность построения зависимостей между ними, формирование деревьев показателей для того, чтобы на основании схематического представления можно было отследить, из каких базовых данных состоят расчётные показатели и что влияет на их формирование.

Растут требования к расширенным ролевым моделям. Во многих компаниях появляются дата-офисы с разными функциями, процессами и ролями, такими как дата-партнёры и дата-стюарды. Для всех них требуются отдельные полномочия и возможности системы.

Расширяются и требования к хранилищам метаданных (Дата каталогу). Раньше от них требовалось простое сканирование уже реализованных схем данных. Но уже важно, чтобы они давали возможность проектирования

витрин данных, и других слоёв КХД.

КАК МОЖЕТ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ВНЕДРЯТЬСЯ В СИСТЕМЫ УПРАВЛЕНИЯ ДАННЫМИ? КАКИЕ ЗАДАЧИ ОН ДОЛЖЕН РЕШАТЬ?

На текущем этапе можно ожидать применения ИИ во вспомогательных функциях систем управления данными. Скажем, при работе с семантическими дубликатами в логической модели данных, поиске показателя, например с использованием чат-бота.

В данном случае работа ИИ включает в себя семантический анализ описания и состава модели данных. Под семантическим анализом модели данных подразумевается этап в последовательности алгоритма автоматического понимания текстов. Он может быть использован с целью выявления дубликатов по показателям и справочникам.

Другие перспективные направления применения ИИ – это интеллектуальный анализ текста (проверка содержимого документа на актуальность и наличие противоречий с другими документами), корреляционный анализ значений показателей или анализ связей с другими показателями, предсказания в аналитическом блоке, оптимизация потоков данных. В долгосрочной перспективе часть задач подсистем в цикле управления данными могут взять на себя модули с искусственным интеллектом.