

ДАННЫЕ – НОВАЯ НЕФТЬ

КАК **> самолет** НАУЧИЛСЯ ЕЕ КАЧАТЬ

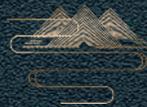
Свитнева Ольга CDO

Свитнева Ольга
CDO Самолет >

18 лет в Data
Моя миссия: превращать
данные в прибыль



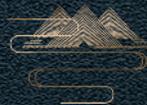
ПЛАН



Тренды



Платформа данных



Управление данными



Результаты



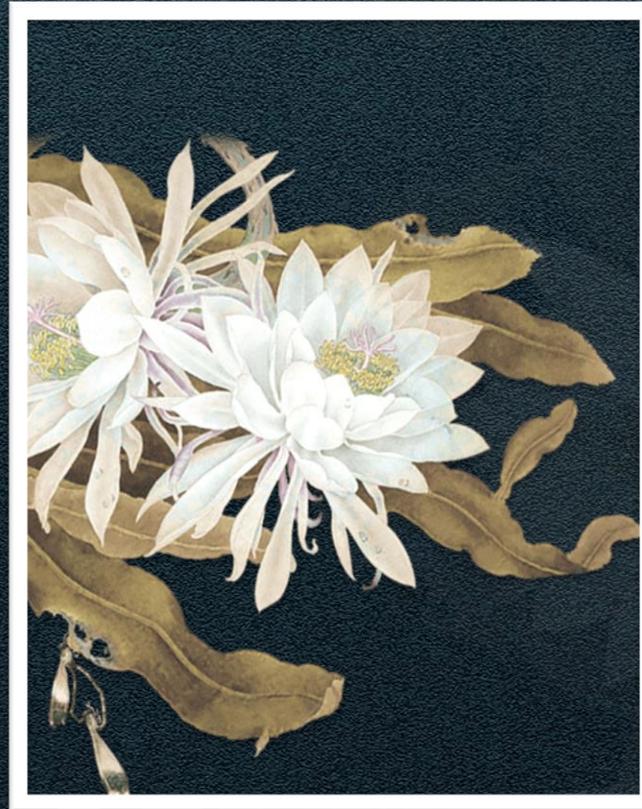
Данные – новая нефть

Основа экономики

Современный цифровой мир
построен на данных

Путь к ИИ и автоматизациям

Данные в основе любого ИИ
решения



Стратегический ресурс

Катализатор эпохи данных

Топливо современных технологий

Данные стали фундаментом для
развития новых технологий

Цифровизация как тренд



Рост рынка больших данных

Согласно исследованию Research and Markets, до 2028 года рынок больших данных будет расти в среднем на 13,3%, эксперты считают, что рост в России составит до 20% в год.

<https://www.researchandmarkets.com/reports/5939585/big-data-analytics-global-market-report>
https://doc.cnews.ru/reviews/analitika_30_2024

Рост применения генеративного ИИ

Темп роста AI-технологий 25% в год.

Не только LLM, но и SLM (требует в разы меньше ресурсов при обучении и анализе данных, чем большая языковая модель, но может обладать лучшей точностью в конкретной задаче)

<https://tass.ru/ekonomika/19406079>

<https://www.vedomosti.ru/technology/news/2023/09/26/997211-mishustin-na-razvitie>

Подробнее на Г.В.К.

Как изменился мир DATA

Data driven management

Принятие решений на основе данных

10V's of Big Data

Широта характеристик данных

- ❁ Объем
- ❁ Скорость обновлений
- ❁ Многообразие форматов
- ❁ Визуализация
- ❁ Ценность
- ❁ Изменчивость = вариативность
- ❁ Валидность = пригодность
- ❁ Достоверность = точность
- ❁ Уязвимость = безопасность
- ❁ Волатильность = нестабильность тенденций
(данные, полученные несколько дней спустя, отличаются от текущих данных)



Смена архитектур

LakeHouse – современная парадигма



Изоляция хранения от вычислений для облегчения масштабирования.



Возможность использовать инструменты бизнес-аналитики для работы непосредственно с исходными данными.



Поддержка разнообразных рабочих нагрузок: от алгоритмов машинного обучения до SQL-запросов и распределённых вычислений

Платформа данных как экосистема

Качество данных

Каталог данных

Платформа качества данных

Устранение ошибок в данных, обеспечение их точности, целостности, согласованности, валидности и полноты

Управление мастер данными и справочниками

Система управления мастер-данными

Сервис НСИ

Единое решение управления бизнес-метаданными и НСИ для стандартизации и поддержания данных в актуальном состоянии с целью интеграции общих корпоративных справочников в процессы компании

DATA GOVERNANCE

Аналитическая платформа LakeHouse

Озеро данных

Кроссистемные витрины

Объединение информации из различных источников в единой аналитической платформе и обогащение данных
Преобразование данных в формат, пригодный для аналитики и моделирования
Единая версия правды по бизнес показателям
Основа для проектов ИИ, BI, Self-Service и Ad-hock

Централизованная отчетность

Сервис BI

Self-service

Корпоративная система отчетности и визуализации данных в удобном для анализа формате, с применением единой согласованной методологии по терминам и расчетам. База для принятия управленческих решений.

Искусственный интеллект

Платформа ИИ

Генерация дополнительных выгод и оптимизация процессов за счёт применения класса решений LLM, нейросетей, прогнозной аналитики, имитационного моделирования, Data Mining, статистического анализа и глубинного анализа данных.

DATA FUSION

Разработка

Data инженерия

Разработка сервисов и автоматизации

ML инженерия

Управление потоком данных, их хранением и архитектурой. Продуктивизация данных, оптимизации вычислений и движков.

Службы эксплуатации

Реализация технологического стека и практик, направленных на повышение качества и скорости анализа данных.
Обеспечение непрерывного процесса интеграции, развертывания и мониторинга сервисом по работе с данными.

SRE (обеспечение стабильности сервисов)

Data Ops (инфраструктура и сервисы данных)

Инфраструктура в собственных ЦОД

Что важно помимо технологий?

Клиентцентричность ➤

Каталог данных

Удобный поиск
необходимой
информации

▶ 📁 sproject	444
▶ 📁 icbit	383
▶ 📁 scenter	328
▶ 📁 crm	272
▶ 📁 stender	269
▶ 📁 scatalog	258
▶ 📁 scontrol	247
▶ 📁 shome	247

Инструкции

Понятные шаги по
заведению задачи и
получению доступа

Завести новую задачу в команду LakeHouse

Если вам необходимо:

- Разработать витрину
- Внести изменение в уже созданную витрину
- Подключить новый источник
- Внести изменения в уже подключенный источник
- Исправить найденную ошибку в уже существующих объектах LakeHouse

Получение доступов к уже существующим витринам LakeHouse

Если вам необходимо:

- Получить доступ к уже существующим объектам LakeHouse
- Создать учетную запись ClickHouse
- Сделать выгрузку данных в Excel

Сроки и результат

Прозрачный процесс

Финансы Закупки 10 Data Преддевелопмент C

READY FOR DQ 1



IN PROGRESS 42



LH-1908
Подключение источника
Mindbox

Обогащение LakeHo...

13

В А

LH-113
Реализация историчности в
витрине по продажам

Обогащение LakeHo...

Нет

В А =

LH-312
Подключение источника
Talent Q

Обогащение LakeHo...

Нет

В А =



Data Governance + Data Management

- Данными нужно управлять
- Данные имеют ценность лишь тогда, когда они качественные



Платформа Качества данных

Обеспечиваем целостность, согласованность, точность, валидность, полноту данных на этапе сбора информации в централизованное хранилище

Система управления мастер-данными

Верификация, дедубликация, формирование золотых записей, создание Глобальных Справочников, которые тиражируются во все системы компании

Платформа качества данных

Автоматизация проверок данных

Запуск тестов по расписанию

УРОВЕНЬ
КАЧЕСТВА
ДАнных **97%**

- Создано **> 72 000** автоматизированных тестов
- Создано **116** очисток данных
- Исправлено **> 2 000** ошибок

Мониторинг результатов тестов

Отслеживание метрик качества данных



Регулярный мониторинг качества данных



Качество данных в КХД | общая информация



Дата

Last 3 months

01.12

09.12

Категория ошибки

(All)

Система-источник

(All)

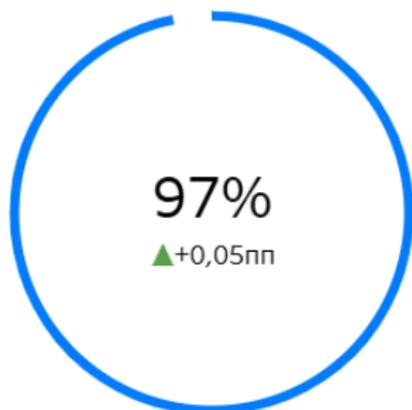
Слой

(All)

Тотал на **09.12.2024**

▲ Исправлено ▼ Новых

Уровень КД

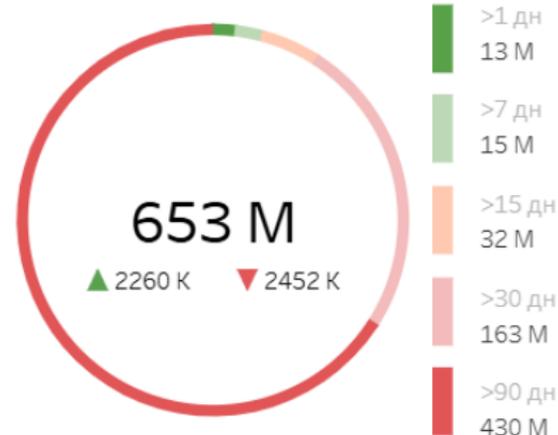


Количество тестов, шт.

Запущено
72 128

Пройдено **66 831** Не пройдено **5 297**

Кол-во ошибок по Валидности



Топ-5 источников по непройденным тестам

Система	Кол-во тестов
1С БИТ ФИНАНС	642
CRM	349
ДОМИЛЕНД	197
UJIN	190
S.PASS (СКУД)	178

Сложности при попытках внедрения MDM

Неочевидность проблемы

Проблемы с данными (дублирование, противоречивость, устаревание) часто воспринимаются как "фоновые" или "незначительные"

Отсутствие измеримых KPI

Сложно измерить, как улучшение качества данных влияет на бизнес-показатели

Высокая стоимость внедрения

Внедрение MDM требует значительных ресурсов: времени, денег и экспертизы. При этом результаты станут ощутимы только через год

Сложность интеграции

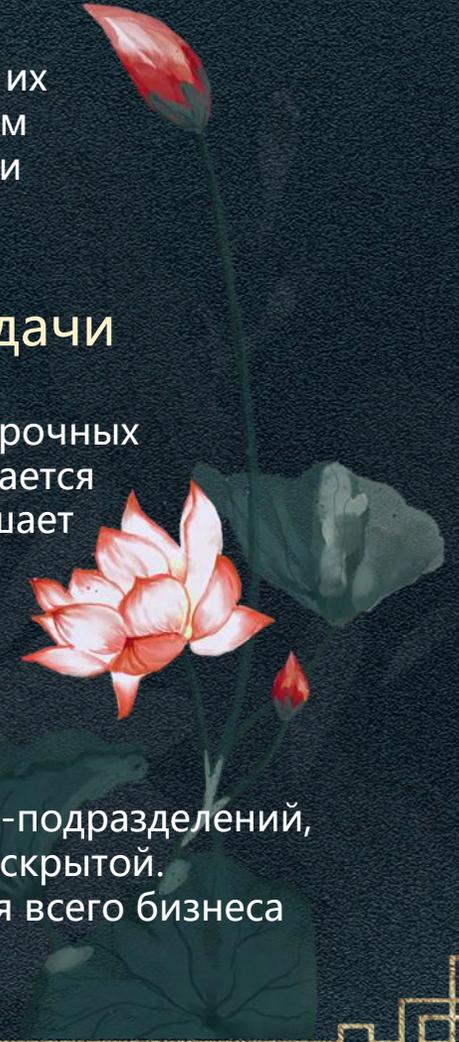
В компаниях уже есть множество систем и их интеграция с MDM может казаться слишком сложной, поскольку потребует перестройки бизнес-процессов

Приоритеты на «горящие» задачи

Компании часто фокусируются на краткосрочных задачах. Управление данными воспринимается как долгосрочный проект, который не решает текущих проблем в данный момент

MDM – не ИТ-проект

Если MDM внедряется без участия бизнес-подразделений, его ценность для компании остается нераскрытой. MDM - это стратегический инструмент для всего бизнеса



Как изменить ситуацию?

Компании не вкладываются в MDM, потому что не видят прямой связи между управлением данными и своими ключевыми задачами

Обучение

Демонстрация ROI

Постепенное внедрение

Фокус на бизнес-ценности

Единый источник
правды

Быстрое внедрение
изменений

Стандартизация
процессов

Эффективная
аналитика



Яркие примеры

Как МДМ помогает бизнесу стать более эффективным и гибким



Проблема	Данные о клиентах хранятся в разных системах (CRM, ERP, маркетинг), и они часто противоречат друг другу	Разрозненные данные затрудняют анализ и прогнозирование	В разных подразделениях используются разные справочники (например, коды товаров или единицы измерения)	Изменения в справочниках (например, новые налоговые ставки или категории товаров) требуют ручного обновления в каждой системе.
Решение	MDM создает единый источник данных, где информация о клиентах всегда актуальна и согласована	MDM объединяет данные из разных источников, обеспечивая полноту и точность аналитики	НСИ обеспечивает единые стандарты для всех систем и процессов	НСИ позволяет централизованно управлять изменениями, которые автоматически применяются во всех связанных системах
Результат	Меньше ошибок, быстрее принятие решений, выше доверие к данным	Лучшее понимание бизнеса, возможность строить точные прогнозы и стратегии	Упрощение взаимодействия между отделами, снижение ошибок и затрат на ручную обработку данных	Снижение времени на внедрение изменений и повышение гибкости бизнеса

Управление мастер-данными в Самолет >

- Центр экспертизы по управлению мастер-данными
- 100% документирование: бизнес-процесс, модель данных, правила качества, интеграционные взаимодействия
- Централизация ведения ключевых справочников всех бизнес-подразделений
- 10 тысяч заявок в месяц на создание/изменение данных
- Более 20 объектов мастер-данных в MDM
- Единый справочник Материалов и Работ для девелоперского блока – проектирование, смета, закупки, активирование
- Более 40 информационных систем потребляют централизованные Мастер данные посредством интеграций
- Автоматизированные DQ по мастер-данным в MDM

Единая база Клиентов,
Поставщиков, Подрядчиков и
прочих контрагентов

Единые справочники блока
Девелопмент

Одновременное обновление
мастер-данных во всех ИС
компании

Обогащение данных из
внешних сервисов

Эффективные
решения

Эффективный
производственный процесс

Единые данные во всех
бизнес-процессах

Регулярная актуализация
данных



Благодарю!
Ваши вопросы

