

Большие данные в промышленности

Михаил Юрьевич Гуненков, старший преподаватель

Какие данные называются большими

- Большие данные (Big Data) — **это** структурированные, частично структурированные или неструктурированные большие массивы данных
- также под Big Data **понимают** обработку, хранение и анализ огромных объемов данных
- большие данные **помогают** анализировать текущее состояние бизнеса, строить прогнозы и автоматизировать рутинные процессы

Классификация данных

- **структурированные данные:** хранятся в таблицах и реляционных базах данных
- **частично структурированные данные:** данные не подходят для таблиц, но могут быть систематизированы; текстовые документы или файлы с записями о событиях.
- **неструктурированные данные:** не имеют организованной структуры; аудио- и видеоматериалы, фото и другие изображения

Основные признаки больших данных (3V)

- **Объем (Volume):** данных действительно много (терабайты, петабайты)
- **Разнообразие (Variety):** они имеют различную структуру, поступают из различных источников
- **Скорость (Velocity):** данные генерируются и обрабатываются с огромной скоростью

Некоторые примеры больших данных

Не Big Data	Big Data
Список пользователей приложения для заказа такси: это просто большая база данных	Логи приложения для заказа такси с геолокацией пользователя, временем заказа, стоимостью поездки и моделью устройства, с которого сделан заказ
Данные одного валидатора в автобусе в течение дня: они могут быть разнообразны, но их слишком мало	Данные всех валидатора в автобусе за последний месяц
Информация библиотеки о статистике посещений и востребованных книгах за день: она генерируется слишком медленно	Информация о просмотрах Википедии за день: ежедневно пользователи просматривают больше 800 млн страниц

Парадокс определения больших данных

Большие данные — это когда **хранить их все дешевле**, чем определять, какая часть из них имеет ценность.

Следует задавать вопросы:

- вы действительно генерируете огромное количество данных?
- если да, то действительно ли все эти данные вам нужны одновременно?
- если да, то уверены ли вы, что просто не храните мусор?

Дополнительные признаки больших данных (6V)

- **Достоверность (Veracity):** важны качество и точность данных, а также корректность способа, с помощью которого их получили
- **Вариативность (Variability):** значение одних и тех же данных может меняться с течением времени, поэтому оценивать их нужно в общем контексте
- **Ценность (Value):** анализ данных способен принести пользу бизнесу

Ещё примеры больших данных

- информация о лайках, комментариях, активности пользователей в социальных сетях за месяц
- массив данных о населении в городской системе распознавания лиц
- медицинские записи и образы, включая результаты анализов за десятилетия в Единой медицинской информационно-аналитической системе
- логи серверов, где каждый день генерируются миллиарды записей
- аналитика поисковых систем, собирающая запросы от миллиардов пользователей
- **показатели датчиков / сенсоров технологического объекта за некоторый период времени**

Какие задачи решаются

ThyssenKrupp AG

Один из ведущих мировых производителей лифтов внедрил систему МАХ, которая собирает данные от датчиков, установленных в лифтах компании, и по ним строит модели на платформе Azure Machine Learning, позволяющие предупредить инцидент до его возникновения и передать технику конкретный код поломки для сокращения времени обслуживания

Nestle

Используя данные по продажам за предыдущие периоды и оптимизационные алгоритмы, компания Nestle с помощью технологий анализа больших данных осуществляет оптимизацию производственной цепочки, это позволяет снизить ошибки при прогнозировании спроса на материалы в два раза, также снижаются убытки от хранения избыточных оборотных средств, от задержек в производстве и т.д.

Магнитогорский металлургический комбинат

На Магнитогорском металлургическом комбинате был внедрен сервис от Yandex Data Factory – «Снайпер», предназначенный для оптимизации технологического процесса производства стали.

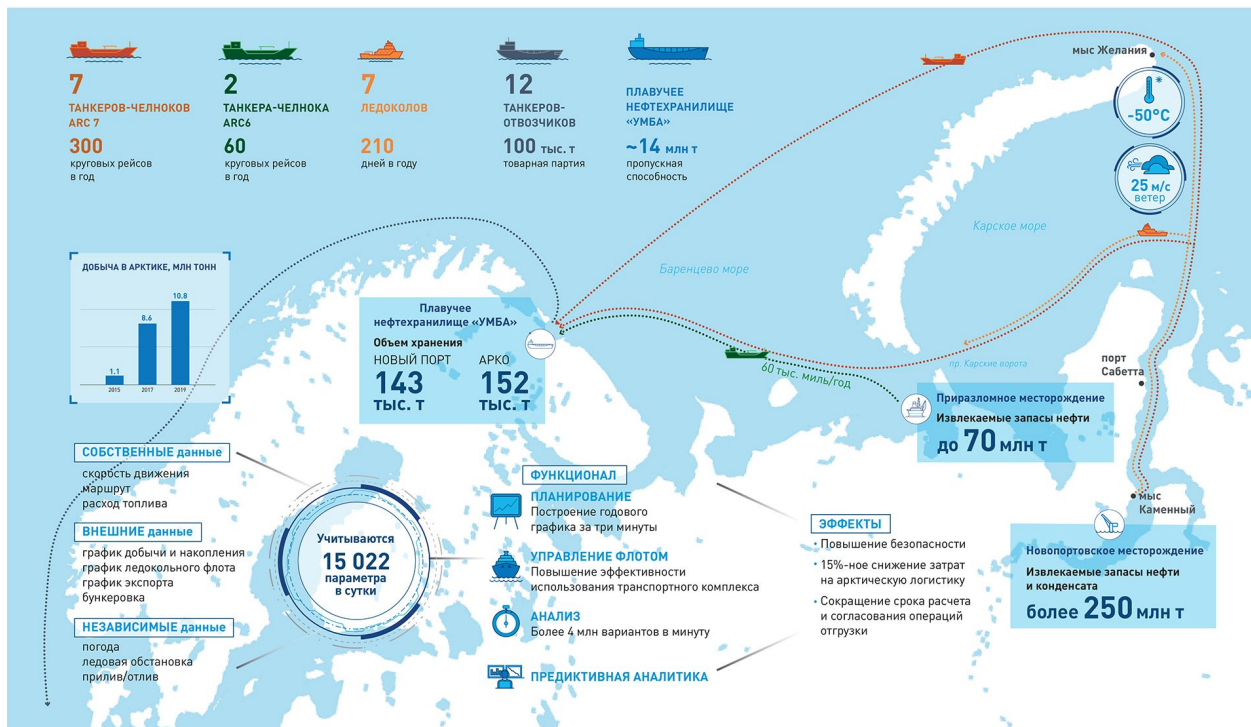
Платформа обрабатывает параметры плавки, а затем выдает соответствующие рекомендации. Специалисты ММК считают, что это приводит к существенной экономии, кроме того, подобная технология может найти применение в целлюлозно-бумажной и химической промышленности

Еще

- На **нефтеперерабатывающих заводах Антипинска и Ильска** были установлены распределенные системы управления для автоматизации производственных процессов, позволяющие безопасно осуществлять технологический процесс, повышать качество продукции и улучшать эффективность производства
- Внедрение платформы SAP HANA для ведения бизнеса в режиме реального времени позволило **«Сургутнефтегаз»** автоматизировать учет продукции, осуществлять расчет скользящих цен онлайн, что привело к значительному увеличению эффективности бизнес-процессов

“Газпром нефть” (“Капитан”)

Цифровая система управления арктической логистикой



Процессы, связанные с большими данными

- сбор
- хранение
- обработка
- анализ

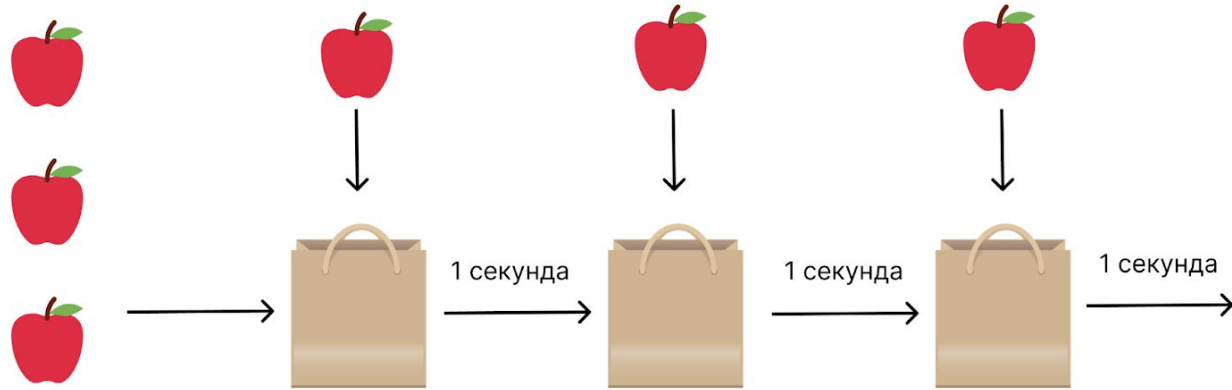
Сбор

Источники больших данных

- приборы (датчики, сенсоры и т. д.)
- видеорекамеры
- интернет
- мобильные устройства
- спутники

Пакетные и потоковые данные

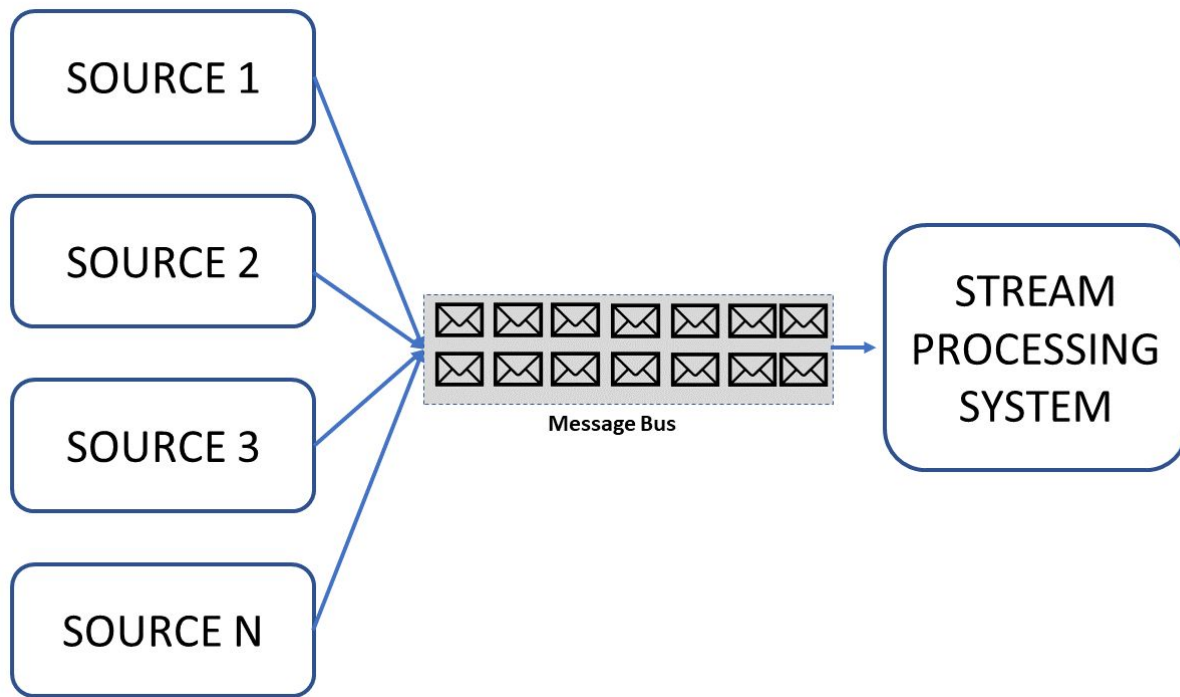
Данные разделяются на микропакеты



Большой набор
данных из
источника

Микропакеты с данными отправляются на обработку

Очереди сообщений



Технологии

- **индустриальные протоколы OPC (OPC DA, OPC UA) для передачи информации с полевого уровня**
- скраперы, краулеры, парсеры для сбора данных из сети интернет
- протокол RTSP для сбора видеоданных
- формат Apache Parquet (для передачи пакетных данных)
- брокеры сообщений Apache Kafka, Apache ActiveMQ, RabbitMQ (для передачи потоковых данных)

Хранение

Варианты хранения больших данных

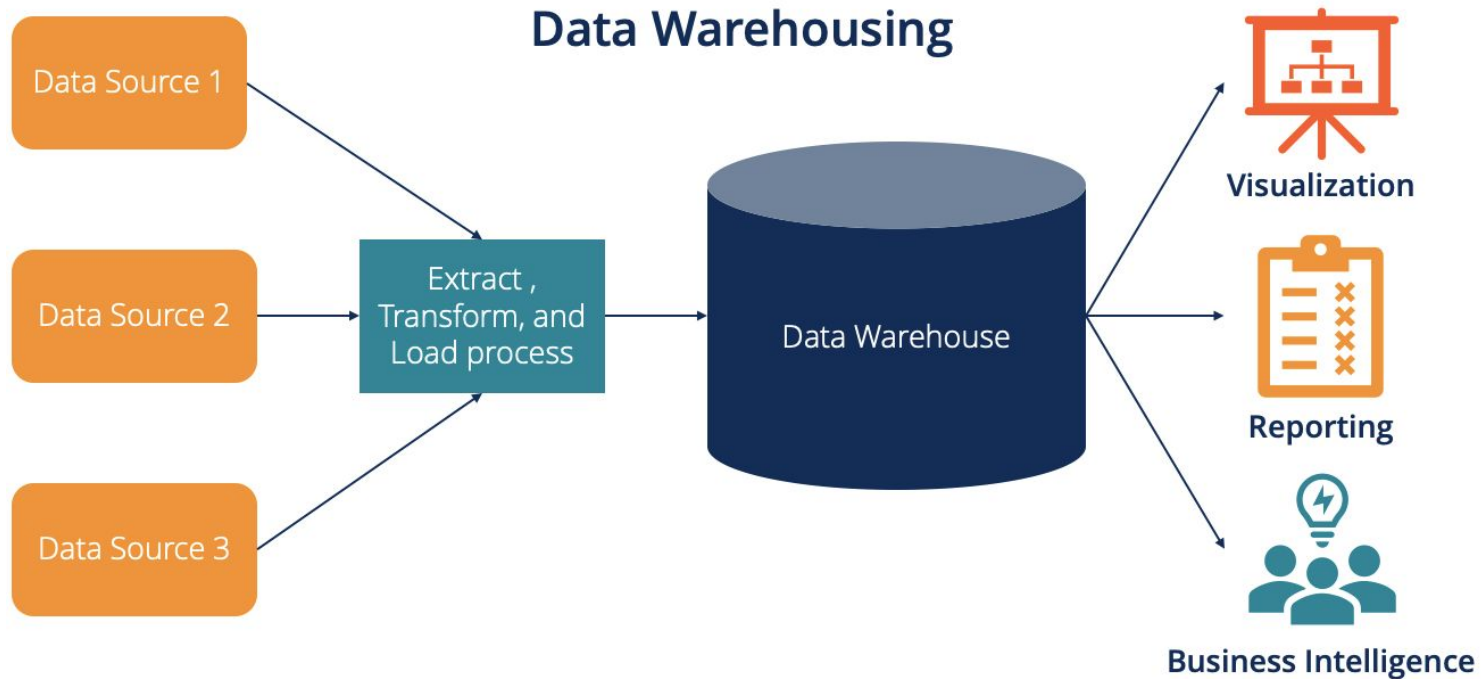
Data Warehouse (DWH): хранение обработанной информации в структурированном виде (готовом для анализа)

Data Lake (DL): хранение данных, которым требуется последующая обработка

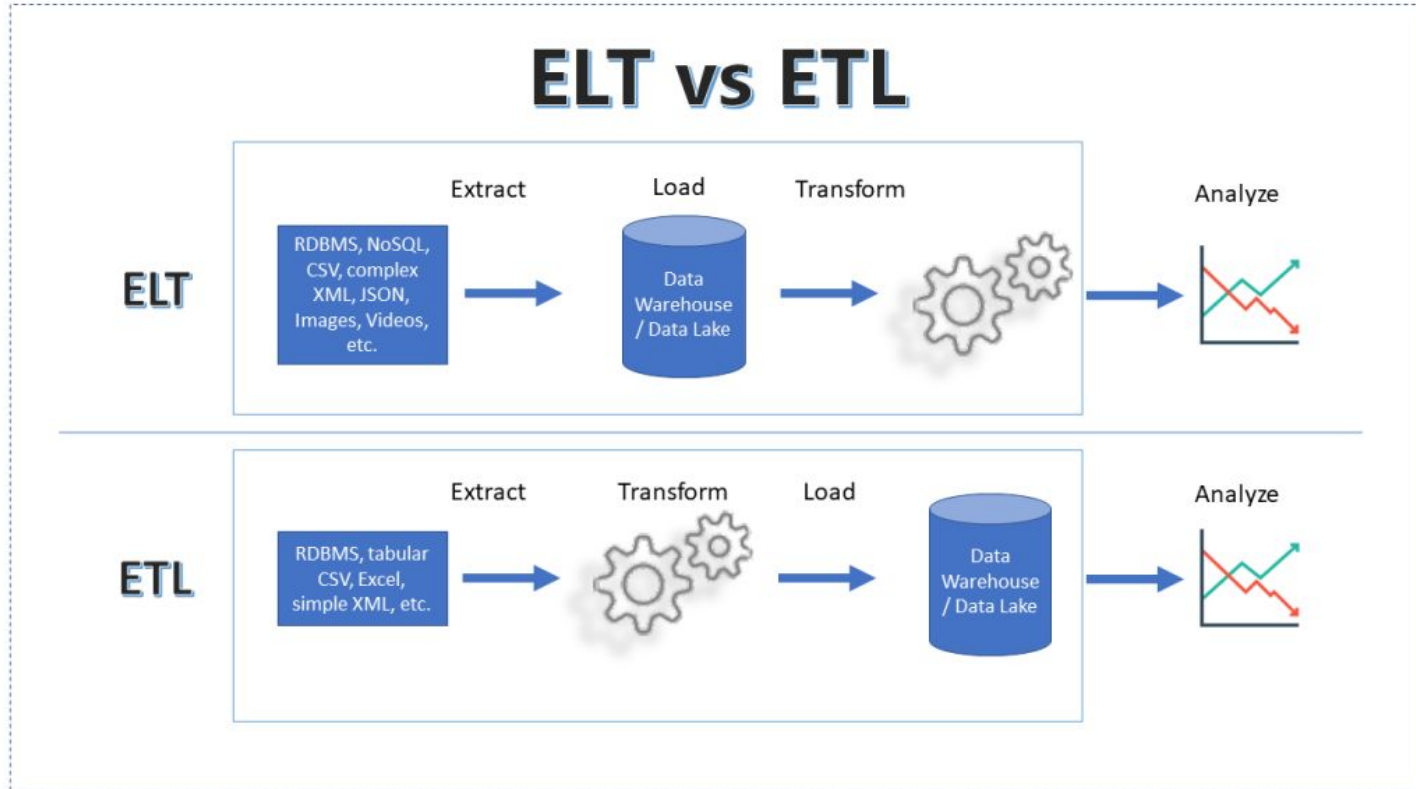
Способы организации DWH: колоночные СУБД (ClickHouse, Greenplum)

Способы организации DL: распределенные ФС (HDFS), NoSQL БД (Apache Cassandra, Redis), объектные хранилища (S3)

Extract Transform Load (ETL)



Extract Load Transform (ELT)

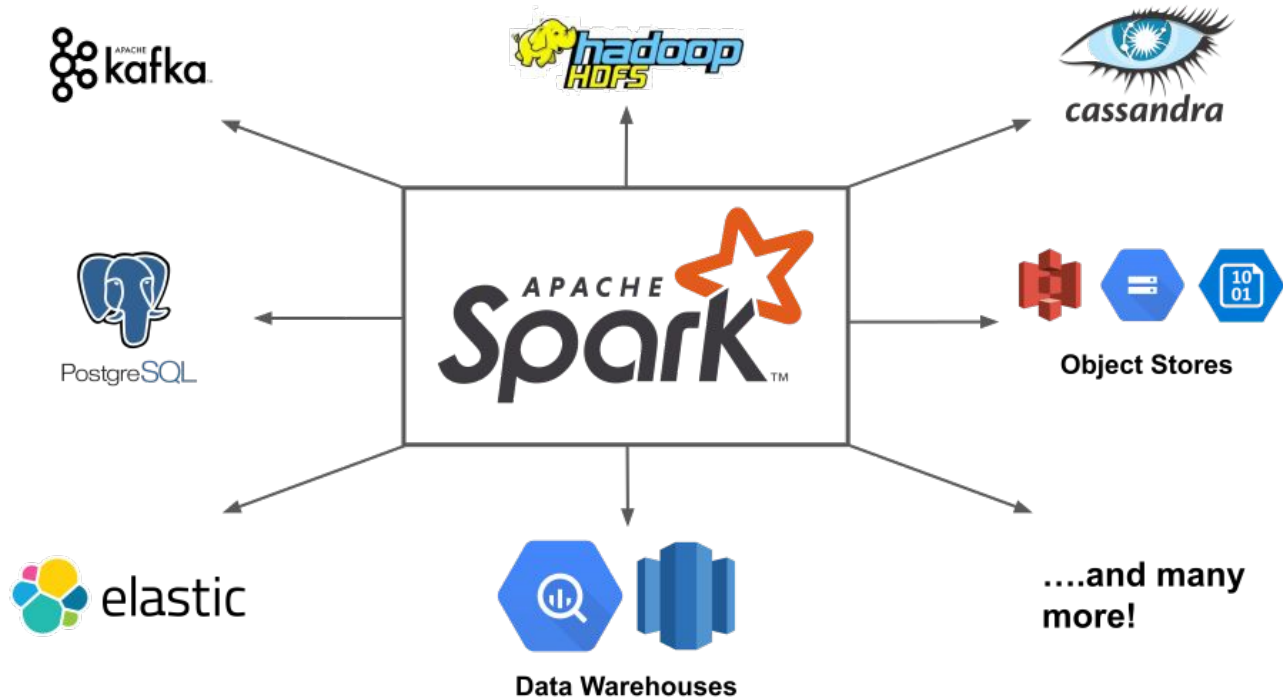


Обработка

Этапы обработки данных

- **очистка:** удаление дубликатов, пропусков, шумов
- **интеграция:** соединение данных (например, таблиц БД)
- **трансформация:** фильтрация, сортировка, группировка, вычисление новых значений
- **редукция:** сжатие избыточной информации
- **агрегация:** вычисление итогов по группам (суммы, средние, максимумы, минимумы)

Основной инструмент

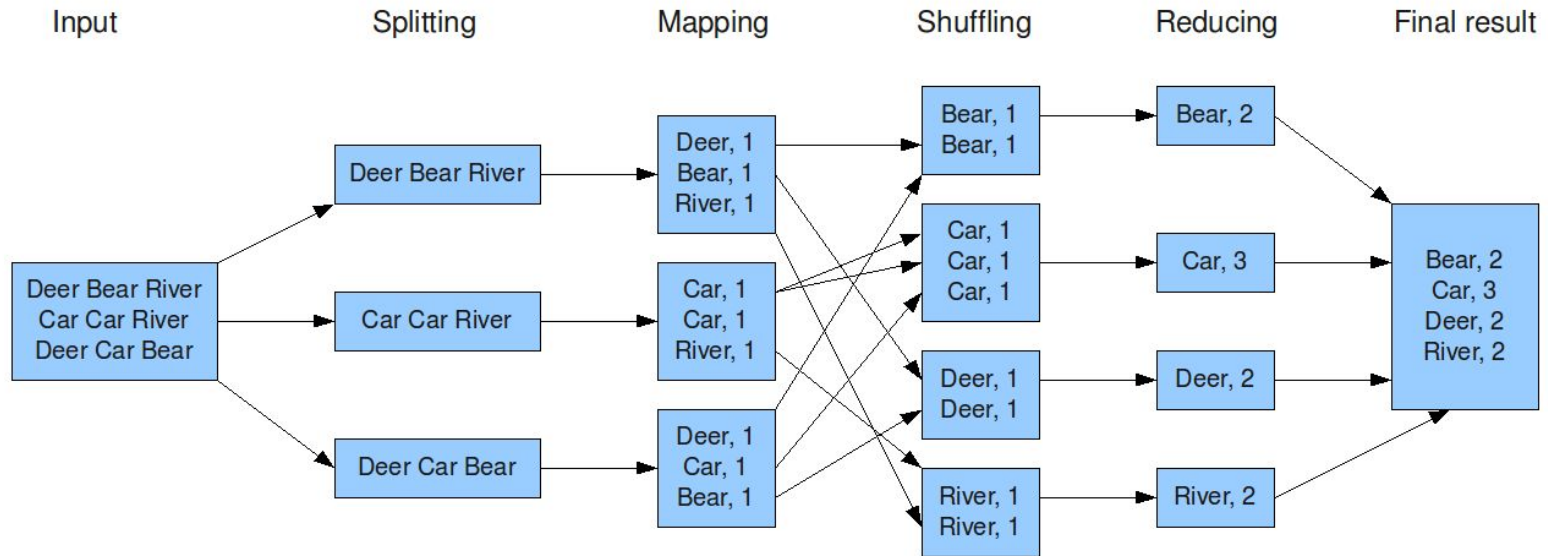


Основные принципы хранения и обработки данных

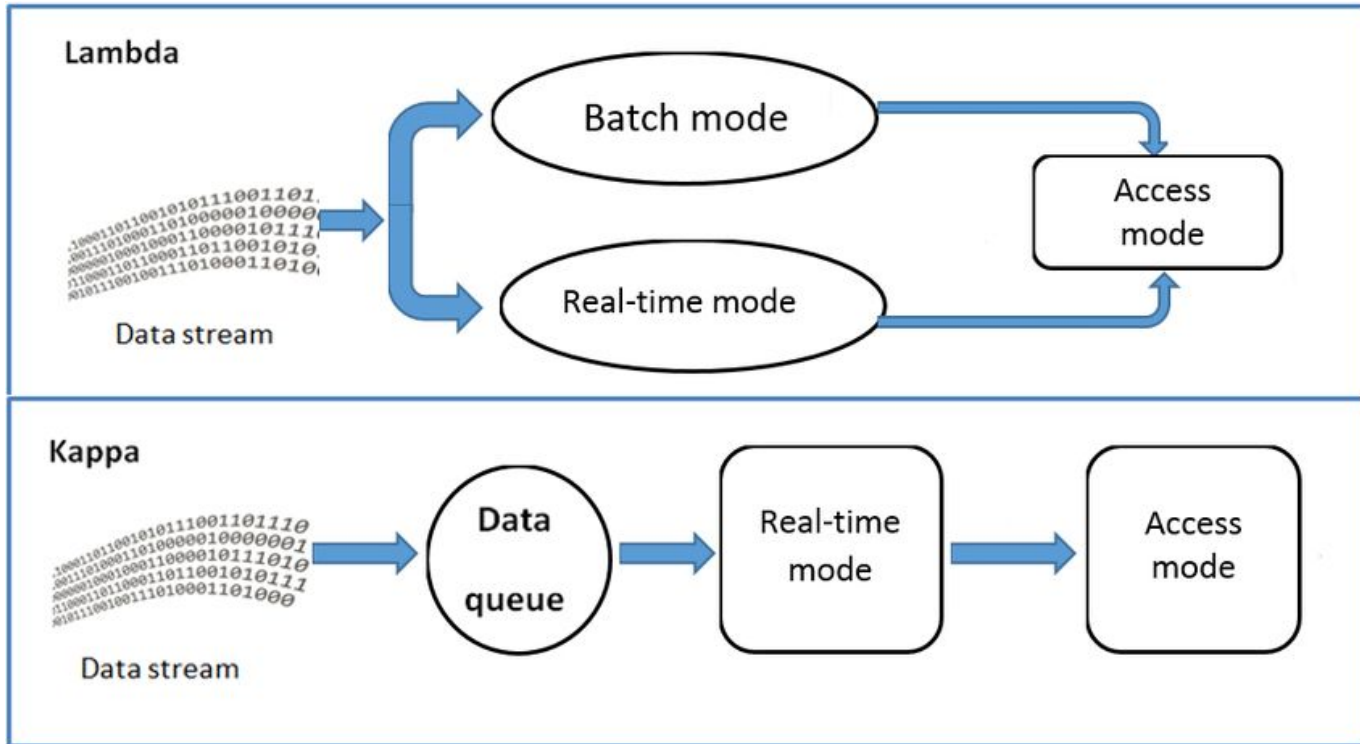
- **масштабируемость:** возможность расширить кластер, добавить новые сервера (горизонтальное масштабирование)
- **отказоустойчивость:** обеспечение непрерывной обработки данных; репликация серверов
- **локальность:** задачи обработки данных выполняются как можно ближе к месту их физического хранения

Подход Map Reduce

The overall MapReduce word count process



Архитектуры обработки больших данных



Анализ

Типы анализа больших данных

- **описательный анализ:** что и когда произошло
- **диагностический анализ:** почему это произошло
- **предиктивный анализ** (на этом уровне ИИ): что будет происходить
- **предписывающий анализ:** как оптимизировать процесс, какой план действий

В рамках модели зрелости анализа организация может характеризоваться уровнем управления большими данными

Формы анализа

- статистический анализ
- отбор признаков
- поиск закономерностей: регрессия, кластеризация
- выявление групп: кластеризация, поиск аномалий
- моделирование: имитационное, языковое, тематическое
- визуализация

Могут применяться инструменты искусственного интеллекта

Промышленный интернет вещей

Промышленный интернет вещей

Интернет вещей (IoT) – система объединенных компьютерных сетей и подключенных физических объектов (вещей) со встроенными датчиками и ПО для сбора и обмена данными, с возможностью удаленного контроля и управления в автоматизированном режиме, без участия человека

Промышленный интернет вещей (IIoT) — интернет вещей для корпоративного / отраслевого применения - система объединенных компьютерных сетей и подключенных промышленных (производственных) объектов со встроенными датчиками и ПО для сбора и обмена данными, с возможностью удаленного контроля и управления в автоматизированном режиме, без участия человека

Принцип работы

- Первоначально **устанавливаются датчики**, исполнительные механизмы, контроллеры и человеко-машинные интерфейсы на ключевые части оборудования
- Осуществляется **сбор информации**, которая впоследствии позволяет компании приобрести объективные и точные данные о состоянии предприятия
- Обработанные данные доставляются во все отделы предприятия, что помогает наладить взаимодействие между сотрудниками разных подразделений и **принимать обоснованные решения**

Zyfra Industrial IoT Platform

Непрерывное производство

Zyfra Industrial IoT Platform

Отечественная цифровая платформа, включающая весь необходимый набор компонентов для создания и внедрения цифровых решений на предприятии.

Решения, созданные и работающие на базе ZIIOT, уже обеспечивают эффективную и бесперебойную работу промышленных предприятий.



ZIIOT – ЛУЧШЕЕ ЦИФРОВОЕ РЕШЕНИЕ ДЛЯ
УПРАВЛЕНИЯ ПРОИЗВОДСТВОМ



IOT AWARD 2020 — ЛУЧШАЯ
ОТЕЧЕСТВЕННАЯ IOT ПЛАТФОРМА



ЛУЧШЕЕ РЕШЕНИЕ ДЛЯ НЕФТЕГАЗОВОЙ
ОТРАСЛИ С ПРИМЕНЕНИЕМ ИНТЕРНЕТА
ВЕЩЕЙ



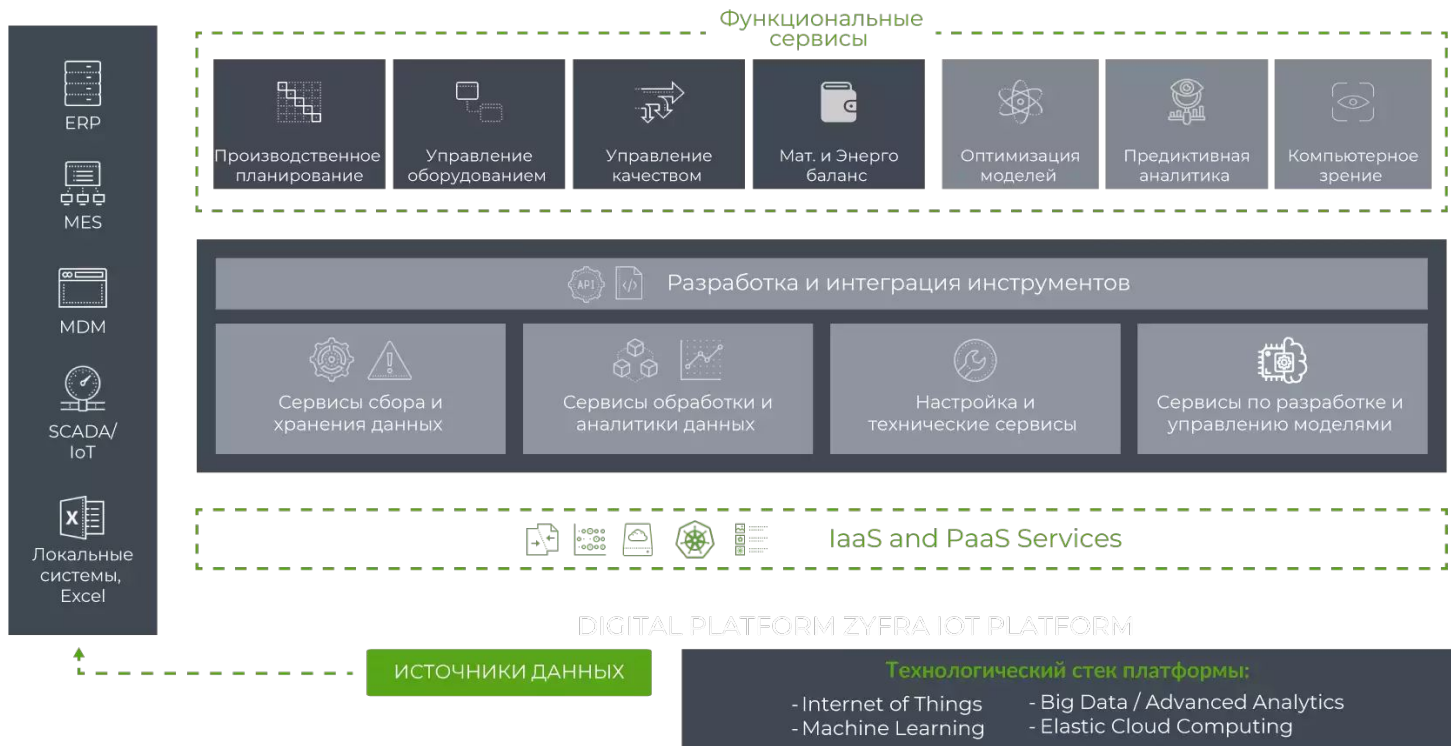
CDO/CDTO
Summit & Award
2022 Russia

DIGITAL РЕШЕНИЕ ГОДА

[Запросить демонстрацию продукта](#)

Zyfra Industrial IoT Platform

Zyfra Industrial IoT Platform



Zyfra Industrial IoT Platform

- Сбор, хранение и управление данными
- Среда для разработки корпоративных бизнес-приложений
- Сервисы администрирования и управления безопасностью
- Готовая платформа для оперативного развертывания, предоставляющая все необходимые инструменты для управления жизненным циклом ИТ-системы

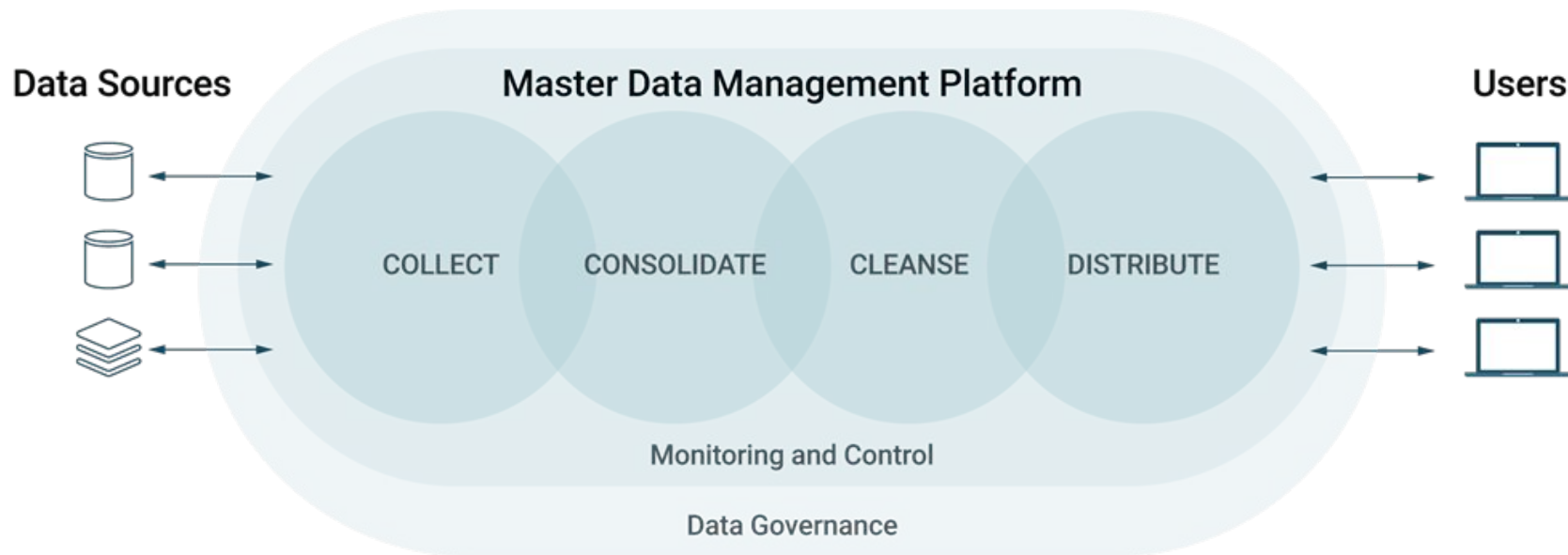
Управление данными

Data Governance — это совокупность процессов определения наиболее важных данных для использования сотрудниками, присвоение им прав доступа и управления для отлаженной работы бизнес-процесса, а также защита данных от внешних воздействий

Data Management — это процесс сбора, хранения и обновления данных

Data Governance формирует стратегию управления данными, а **Data Management** напрямую осуществляет управление данными согласно определенной стратегии

Управление данными



СИБУР запустил собственную платформу промышленного интернета вещей

[Главная](#) > [Пресс-центр](#) > [Новости и пресс-релизы](#) > [СИБУР запустил собственную платформу промышленного интернета вещей](#)

Москва, 18 июня 2020 года. В СИБУРе разработали собственную платформу для промышленного интернета вещей. Она обеспечивает управление устройствами IoT-сети всех уровней, собирает и хранит данные с датчиков и обрабатывает информацию. Теперь оператор установки получает всю необходимую для него информацию – параметры температуры, вибрации, показания работоспособности IIoT-датчиков – а также может управлять удалёнными устройствами в удобном веб-интерфейсе, который он может создать сам.

Спутник-ИОТ

[Главная](#) / [Продукты](#) / [Спутник-ИОТ](#)

Спутник-ИОТ

[Назначение и применение](#)

[Что позволяет](#)

[Преимущества](#)

[Стоимость и сроки](#)

Платформа обеспечивает бесшовную интеграцию различных аппаратных средств с помощью специальных интерфейсов, протоколов связи, сетевых топологий, а также средств хранения, обработки и интеллектуального анализа данных. На первом этапе устройства собирают данные, например, о температуре в помещении, а затем эти данные отправляются в облако. Подключение к облаку может осуществляться различными способами, включая: Wi-Fi, спутниковую, сотовую связь, Bluetooth и другие виды связи. Как только данные переданы в облако, программное обеспечение выполняет их обработку.

Подытожим

Подытожим

- большие данные — данные большого объема, поступающие из различных источников с высокой скоростью, собираемые с целью **решения конкретных задач анализа**
- для работы с большими данными существуют **отдельные подходы, архитектуры и технологии**
- существует тренд интеграции процессов сбора, хранения, обработки и анализа данных в **системы промышленного интернета вещей**

Спасибо за внимание!