



Большие данные в контексте цифровизации образования от понятия к технологиям обработки

Цифровизация» – это средство получения желаемого результата, а именно гибкости образовательного процесса, приносящего обучающимся отличный результат.

Бурное развитие цифровых технологий в сфере образования диктуется актуальностью рассматриваемых при этом проблем:

- цифровые технологии в образовании поддерживаются на государственном уровне и широкой общественностью;
- это инструмент эффективной доставки информации и знаний студентов;
- это инструмент создания учебных материалов;
- это инструмент эффективного способа преподавания;
- это средство построения новой образовательной среды



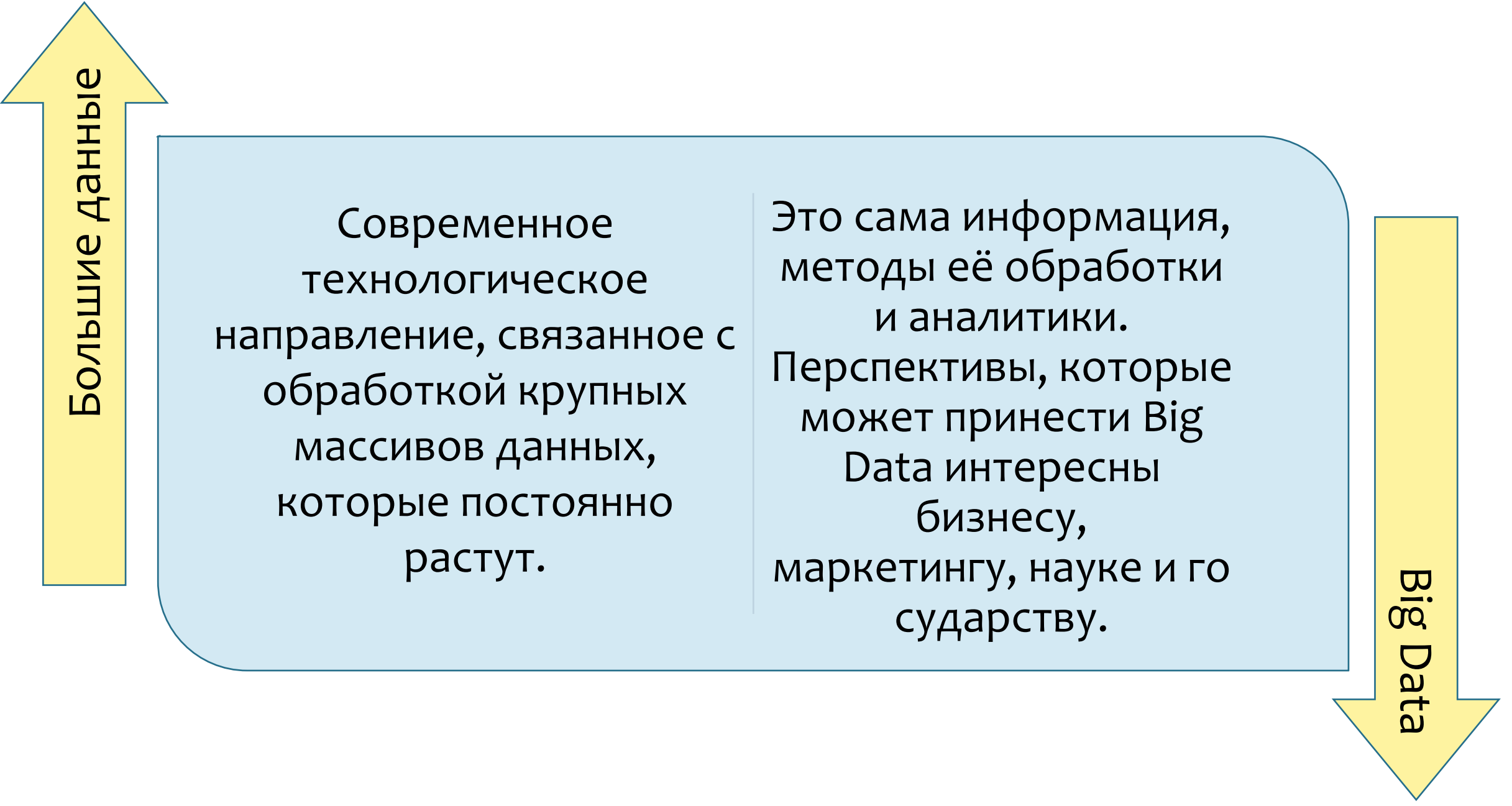
министр просвещения
Ольга Васильева

Цифровизация и активное использование больших данных – это необходимость для любой сферы жизни современного общества, в том числе в школах. Разумеется, никакие цифровые технологии не заменят живого учителя

Тем не менее, сбор и обработка большого массива данных в автоматическом режиме облегчат современным педагогам жизнь: позволят скорректировать образовательные процессы, выстроить актуальные схемы обучения с учетом индивидуальных особенностей каждого ученика и подготовить его ко взрослой жизни.

Что же такое Big Data, какие перспективы они предлагают и где используются?





Большие данные

Современное
технологическое
направление, связанное с
обработкой крупных
массивов данных,
которые постоянно
растут.

Это сама информация,
методы её обработки
и аналитики.
Перспективы, которые
может принести Big
Data интересны
бизнесу,
маркетингу, науке и го
сударству.

Big Data

Два типа информации

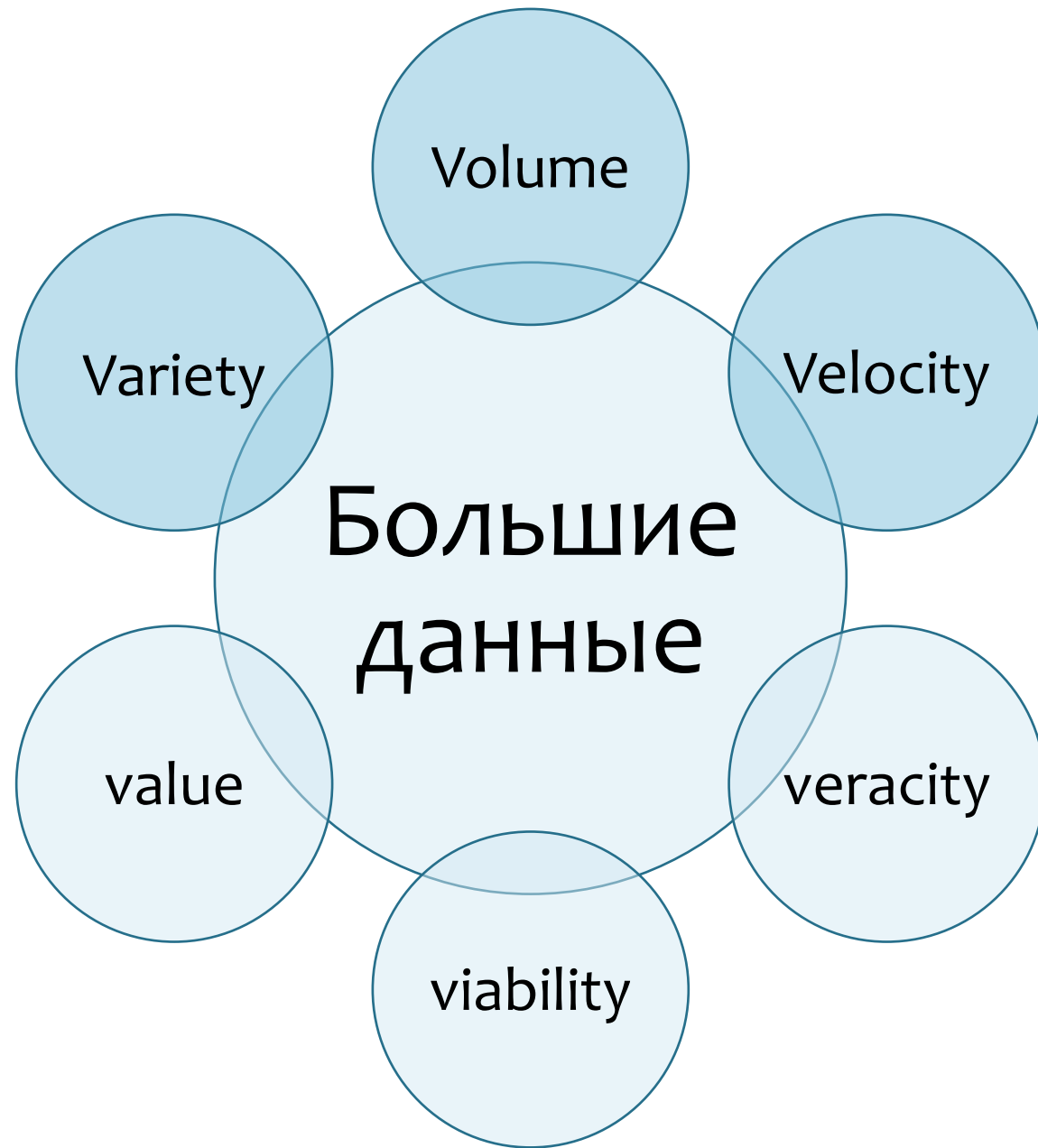


Структурированная

Неструктурированная



- ✓ Данные с сейсмологических станций по всей Земле.
- ✓ База пользовательских аккаунтов Facebook.
- ✓ Геолокационная информация всех фотографий, выложенных за сегодня в Instagram.
- ✓ Базы данных операторов мобильной связи.



Прямо сейчас большие данные помогают в решении таких задач:



повышение производительности труда;

точная реклама и оптимизация продаж;

прогнозирование ситуаций на внутренних и глобальных рынках;

совершенствование товаров и услуг;

улучшение логистики;

качественное таргетирование клиентов в любой сфере бизнеса.

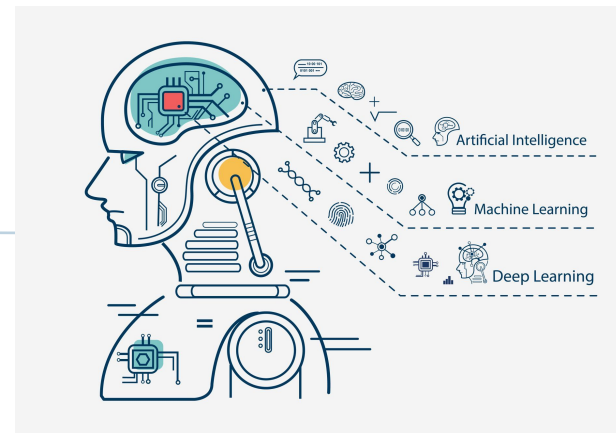
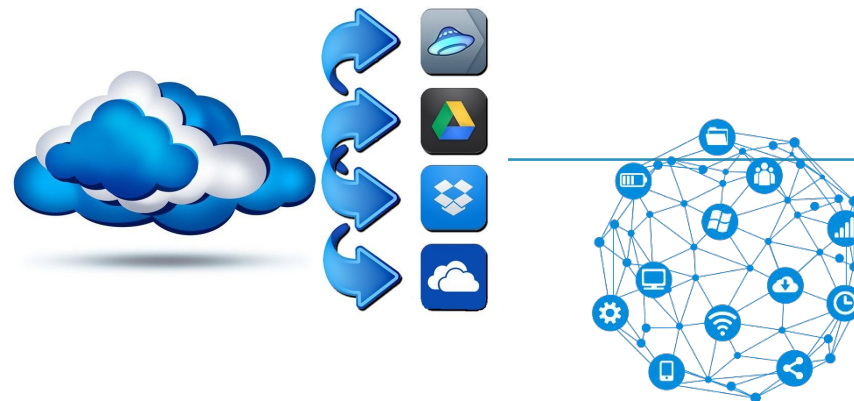
Где используются больше данные?

Облачные хранилища

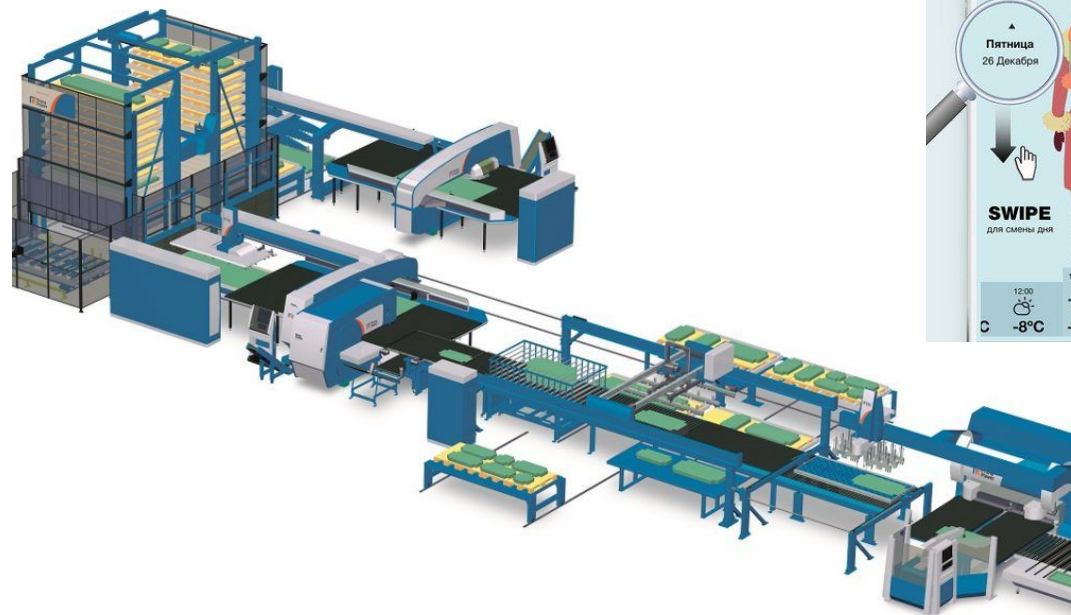
Блокчейн

Самообслуживание

Искусственный интеллект и глубокое обучение



Big Data будет неотъемлемой частью Индустрии 4.0 и интернета вещей, когда сложные системы из огромного числа устройств работают, как единое целое. Вот простые, уже не футуристические, примеры этого:



Кто использует большие данные?



P&G
Procter & Gamble



Среди российских компаний стоит
отметить следующие:

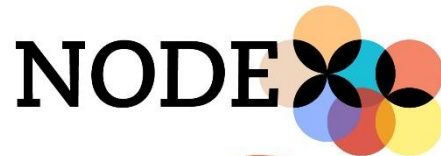


Использование Больших Данных по отраслям



Big Data помогают обработать опыт тысяч преподавателей и студентов, на основе анализа можно получить эффективную методику.







Ключевые
характеристики
Apache
Cassandra:

Возможность функционирования на не очень мощном оборудовании.

Архитектура Cassandra, которая построена на основе технологии Dynamo от Amazon и реализует систему базы данных с использованием ключей.

Архитектура Cassandra, которая построена на основе технологии Dynamo от Amazon и реализует систему базы данных с использованием ключей.

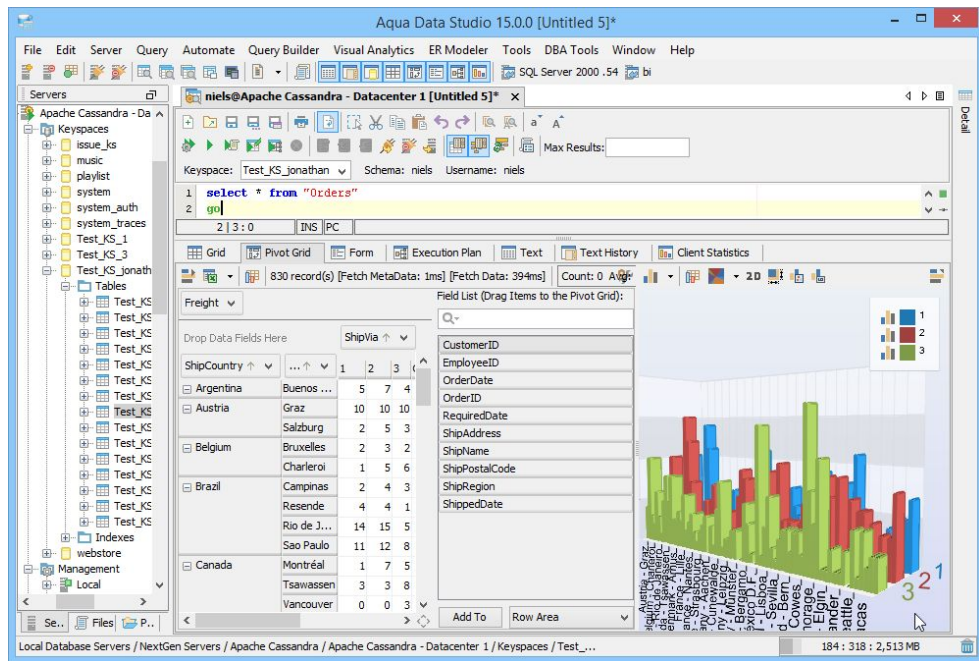
Язык запросов Cassandra.

Развернутое распределение и высокая масштабируемость применения.

Отказоустойчивость и децентрализованная система.

Оперативная запись и считывание данных.

Настраиваемая совместимость и поддержка фреймворка MapReduce.

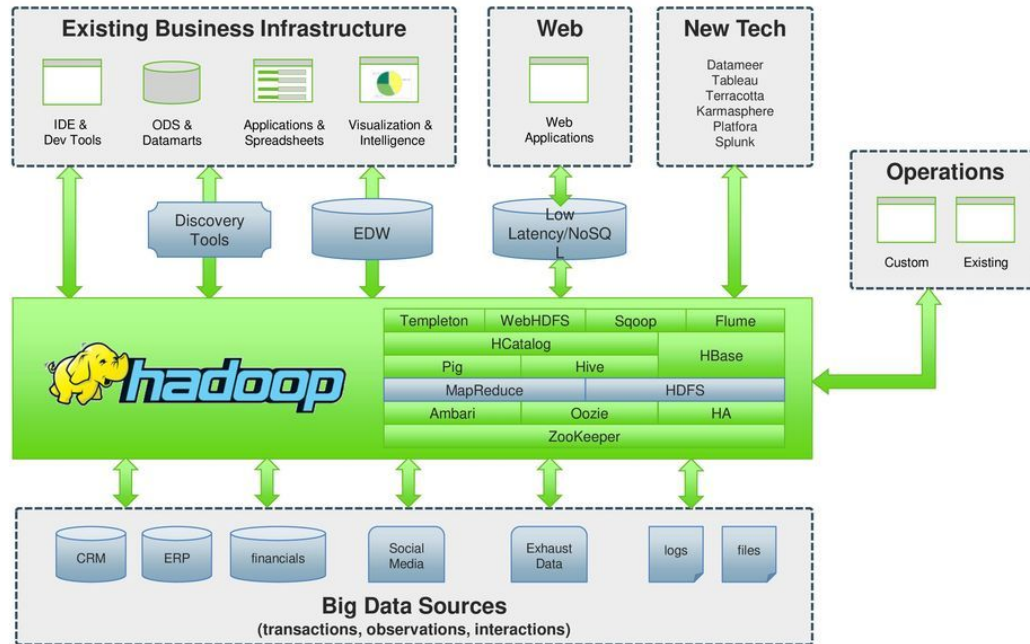




Ключевые
характеристики
Apache Hadoop:

Платформа с высокой масштабируемостью для анализа данных на уровне петабайта.

Hadoop in Enterprise Data Architectures



Редкий отказ узлов в кластере. Но даже если это происходит, система автоматически заново воспроизводит данные и переадресовывает остаточные данные.

Возможность взаимодействовать с другой приоритетной платформой анализа данных. Использование не только NoSQL, но и пакетов, диалогового SQL или доступа с низким значением задержки для бесперебойного процесса обработки данных.



Ключевые характеристики Elasticsearch

Сборка и поддержка программ-клиентов на нескольких языках, таких как Java, Groovy, NET и Python.

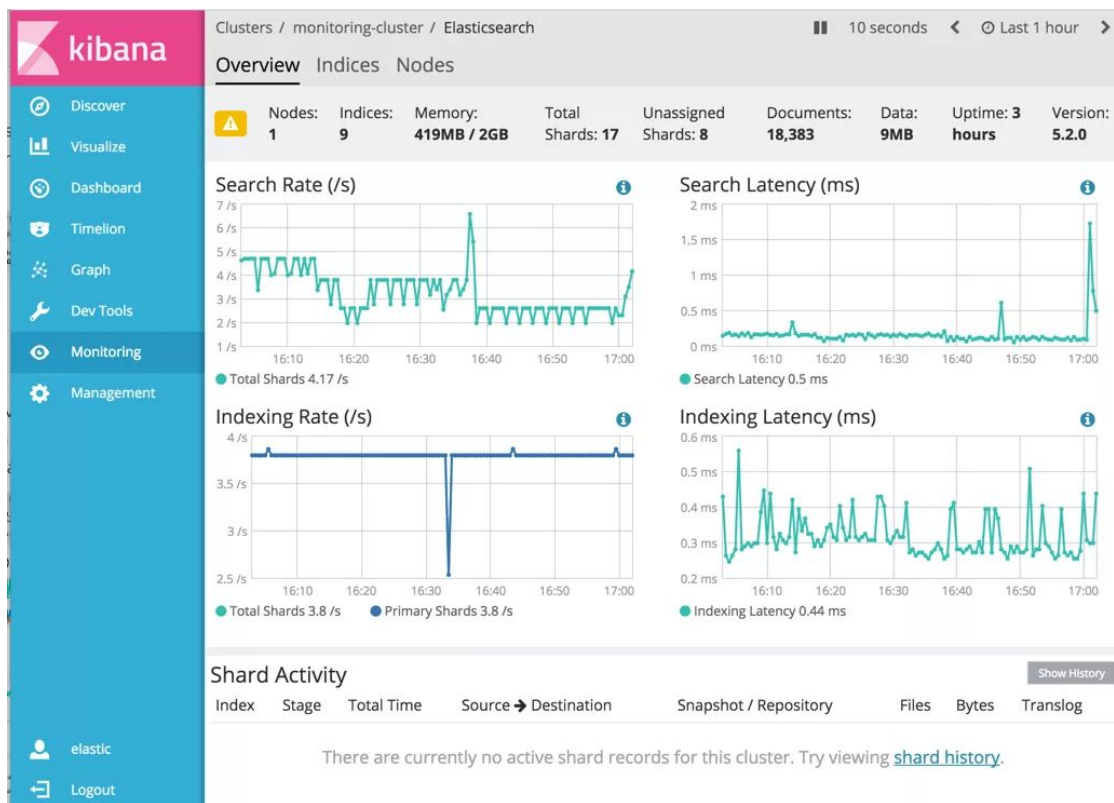
Интуитивно понятный API для управления и мониторинга данных, который обеспечивает полный контроль и наглядность.

Возможность комбинировать несколько видов поиска, включая геопоиск, поиск по метрикам, структурированный и неструктурированный поиск и т. д.

Использование стандартного API и формата JSON на основе архитектуры REST.

Расширенные возможности при анализе данных благодаря машинному обучению, параметрам мониторинга, предоставления отчетов и безопасности.

Актуальная аналитика и параметры поиска для обработки Big Data с помощью Elasticsearch-Hadoop.





Ключевые характеристики presto:

Адаптивная многопользовательская система, поддерживающая одновременное выполнение нескольких операций с памятью машины, операций ввода/вывода (I/O) и запросов с интенсивной вычислительной нагрузкой на CPU.

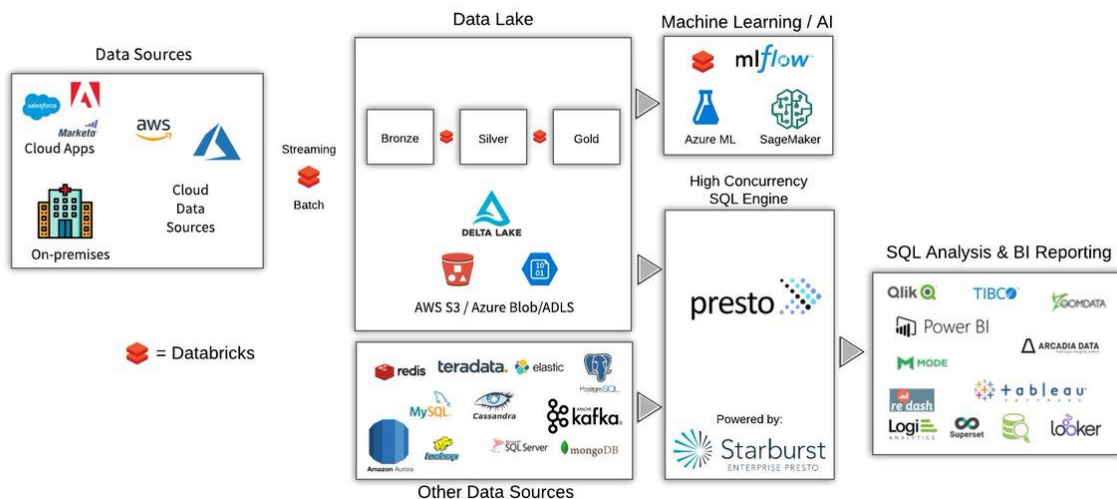
Обеспечение оптимизации для достижения высокой производительности, включая такую важную опцию, как генерация кода.

Возможность расширения и дальнейшей интеграции для создания нескольких кластеров.

Различные настройки и конфигурации для поддержания многочисленных вариантов использования с несколькими ограничениями и параметрами производительности.

Возможность комбинировать в одном запросе данные из множества источников и организовывать анализ Big Data.

Поддержка стандартов ANSI SQL (в дополнение к ARRAY, JSON, MAP и ROW).





Ключевые
характеристики
talend:

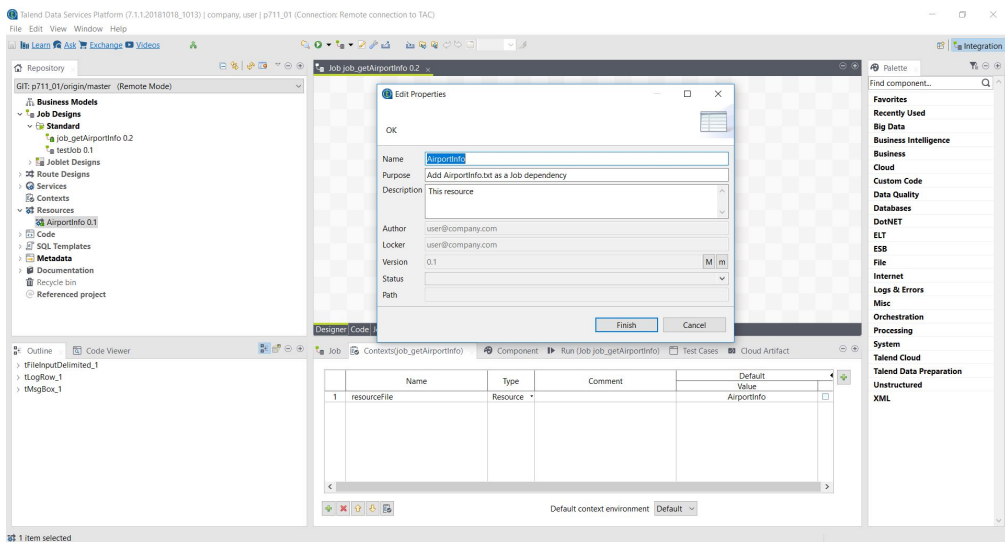
Agile DevOps для ускоренной обработки Big Data.

Упрощение работы Spark и MapReduce за счет генерации
собственных кодов.

Более качественные данные благодаря машинному
обучению и обработке информации на естественном языке.

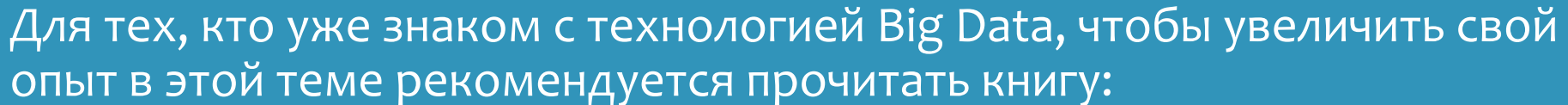
Упрощение процессов ELT (Extract, Load и Transform) и ETL
(Extract, Transform и Load) для Big Data.

Оптимальная настройка всех процессов в DevOps.





- «BIG DATA. Вся технология в одной книге» Андреас Вайгенд;
- «Укрощение больших данных» Билл Фрэнкс;
- «Большие данные» Виктор Майер-Шенбергер, Кеннет Куьер.



«Spark для профессионалов» Сэнди Риза, Ури Лезерсон, Шон Оуэн, Джош Уиллс.



- «Big data простым языком» Алексей Благирев, Наталья Хапаева;
- «Работа с BigData в облаках» Александр Сенько;
- «Теоретический минимум по Big Data» Анналин Ын, Кеннет Су.

Заполните таблицу характеристиками больших данных и традиционных данных

Характеристики	Традиционные базы данных	Большие данные
Область применения		
Характеристика данных		
Способ хранения данных		
Модель хранения и обработки данных		
Количество информации для обработки		

