

BIG DATA



Выполнил: Лахонин Степан, Соловьёв Никита,
Хохлов Михаил

Группа: С2-ИБС-12

2022

Первый компонент определения больших данных

Что же такое Big Data? Возьмём жёсткий диск размером на 500 гигабайт, а это, если задуматься, объём 1 миллиона книг в текстовом формате. Представьте у себя дома такую библиотеку, много? Большие ли это данные? Но те данные, которые могут уместиться на одном компьютере, сложно назвать большими данными. Таким образом, первый компонент для определения больших данных – объём.



Второй компонент определения больших данных

Но нам недостаточно просто хранить эти данные, мы хотим каким-либо образом их обрабатывать. Например, нас интересует вопрос сколько уникальных слов в том или ином языке. В этом случае мы можем пролистать все книги и выписать все слова, которые там встречаются. План есть, берём человека и предоставляем ему доступ к этой библиотеке. Если человек тратит 3 часа на чтении книги размером в 500 страниц, то если не есть, не пить, не спать и не отвлекаться, то спустя каких-то 300 с чем-то лет мы получим ответ. Это наталкивает нас на второй компонент определения больших данных – скорость обработки.



Третий компонент определения больших данных

Третим же компонентом является разнообразие. Когда идёт работа с большими данными, помимо того, что идёт большой поток информации, который нужно оперативно анализировать, этот поток информации разнороден. Там есть информация в текстовом виде, есть в графическом, есть различные аудио- и видеопотоки информации. Например, зайдите на любую поисковую систему и вбейте запрос, на выходе вы увидите сайты, представляющие собой текстовую информацию, картинки - информацию в графическом виде, музыку - аудио, фильмы или ролики - видео.

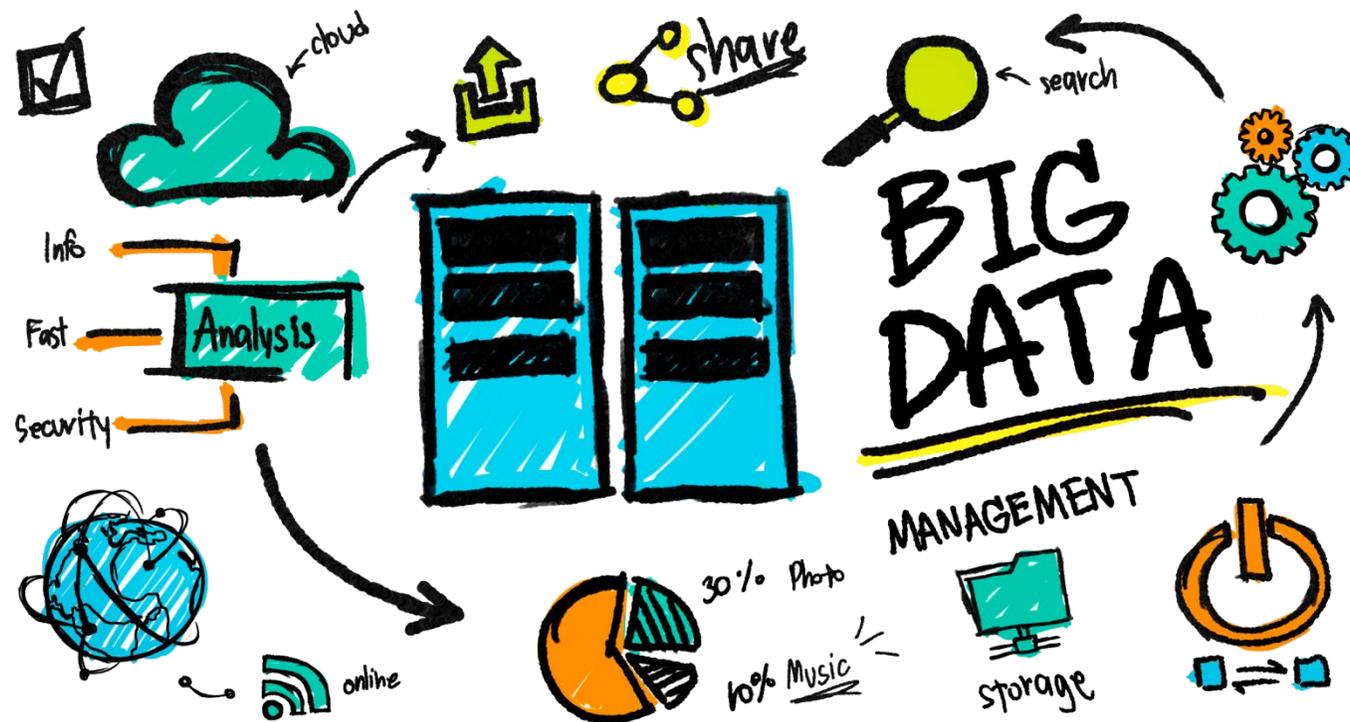
Итоги определения больших данных

Именно пересечением большого количества информации, скорости обработки и её разнообразия в 2001 году компания Gartner выделила в отдельный феномен, который уже прилично позже получил название Big Data. Это определение является неформальным, но в этом есть и свои плюсы, так как, например, то, что в 2000 считалось большими данными, сейчас может считаться нормой, и возможно получится обработать на персональном компьютере. Сравните, если раньше обработка 1 терабайта данных вызывала сложность и относилась к области больших данных, то сейчас в технологических компаниях люди обрабатывают петабайты информации ежедневно и даже не задумываются, а терабайты - уже далеко не большие данные.

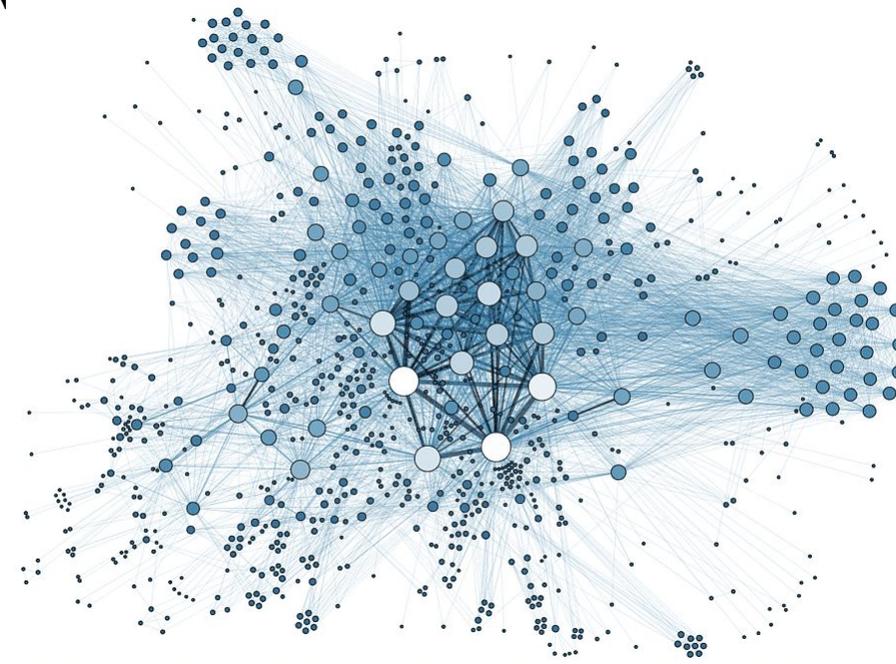


ADVANCED ANALYTICS
BIG DATA
AND VISUALIZATION

Что подразумевается под данными?



Большие данные — это океан информации, в который мы ежедневно погружаемся: зеттабайты данных, поступающих от наших компьютеров, мобильных устройств и аппаратных датчиков. Эти данные используются организациями для принятия решений, оптимизации процессов и политик, а также для создания ориентированных на клиента продуктов, услуг и клиентского опыта. Слово «большие» в этом определении говорит не только об объеме данных, но и о разнообразии и сложности их характера. Как правило, они превышают возможности традиционных баз данных по сбору, управлению и обработке данных. Кроме того, большие данные могут поступать из любой точки земного шара и от любого устройства, которое мы можем отследить в цифровом формате.



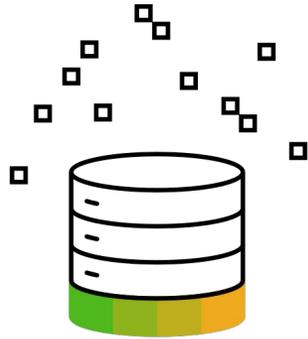
Big Data sources



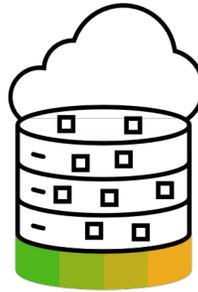
Преимущества больших данных

- **Разработка продуктов и услуг.** Аналитика больших данных позволяет разработчикам продуктов анализировать неструктурированные данные, такие как отзывы клиентов и культурные тенденции, и быстро реагировать на них.
- **Диагностическое техническое обслуживание.** В ходе международного опроса выяснилось, что анализ больших данных с устройств с поддержкой Интернета вещей позволил снизить затраты на техническое обслуживание оборудования на 40%.
- **Клиентский опыт.** Анализ больших данных позволяет компаниям улучшать и персонализировать опыт взаимодействия клиентов с брендом.
- **Устойчивость и управление рисками.** Анализ больших данных позволяет компаниям прогнозировать риски и готовиться к внезапным изменениям.
- **Экономия затрат и повышение эффективности.** Когда компании встраивают расширенную аналитику больших данных во все процессы организации, им удается не только выявлять проблемные аспекты, но и внедрять быстрые и эффективные решения.
- **Повышение конкурентоспособности.** Ценная информация, полученная из больших данных, способна помочь компаниям экономить средства, удовлетворять потребности клиентов, повышать качество продукции и внедрять инновации в свои бизнес-операции.

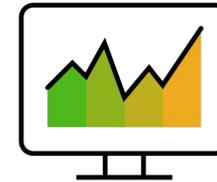
Этапы работы с большими данными



Gather Big Data



Store Big Data



Analyze Big Data

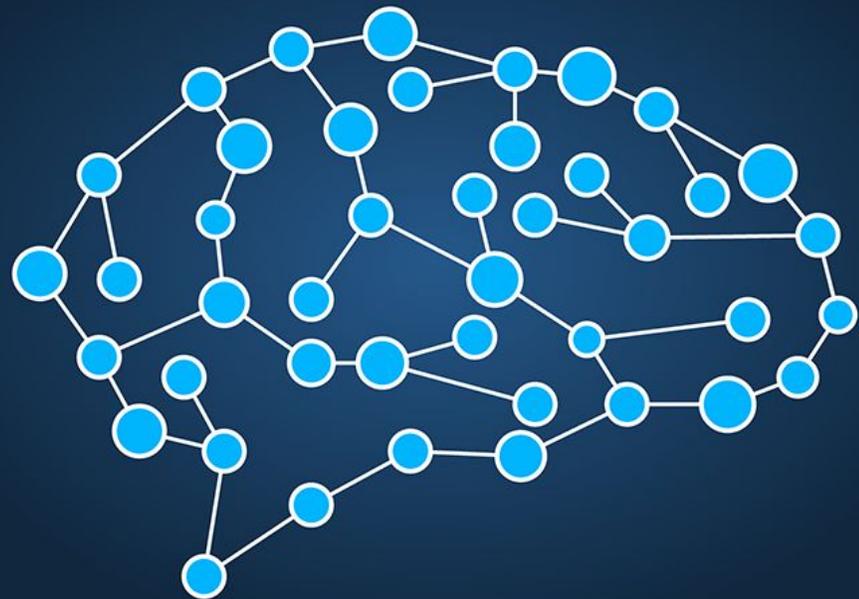
1. Сбор больших данных.
2. Хранение больших данных.
3. Анализ больших данных.

Как работать с Big Data?



Машинное обучение

Машинное обучение — это специализированный способ, позволяющий обучать компьютеры, не прибегая к программированию. Отчасти это похоже на процесс обучения младенца, который учится самостоятельно классифицировать объекты и события, определять взаимосвязи между ними.



Machine Learning

Big Data в бизнесе

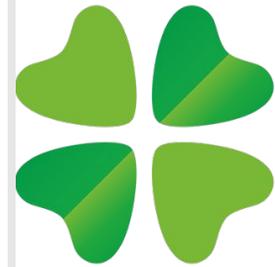
Всех, кто имеет дело с большими данными, можно условно разделить на несколько групп:

- **Поставщики инфраструктуры**
- **Датамайнеры**
- **Системные интеграторы**
- **Потребители**

The SAP logo is displayed in a bold, yellow, sans-serif font against a dark, textured background.The Oracle logo is partially visible, showing the word "ORACLE" in a white, sans-serif font on a dark background.The Microsoft logo consists of a four-colored square (red, green, blue, yellow) above the word "Microsoft" in a white, sans-serif font.The IBM logo is the classic eight horizontal stripes, with the letters "IBM" in a white, sans-serif font below them.

Поставщики инфраструктуры — решают задачи хранения и предобработки данных. Например: IBM, Microsoft, Oracle, Sap и другие.

Датамайнеры — разработчики алгоритмов, которые помогают заказчикам извлекать ценные сведения. Среди них: Yandex Data Factory, «Алгомост», Glowbyte Consulting, CleverData и др.



Clever DATA

glowbyte



Системные интеграторы — компании, которые внедряют системы анализа больших данных на стороне клиента. К примеру: «Форс», «Крок» и др.

КРОК



Потребители — компании, которые покупают программно-аппаратные комплексы и заказывают алгоритмы у консультантов. Это «Сбербанк», «Газпром», «МТС», «Мегафон» и другие компании из отраслей финансов, телекоммуникаций, ритейла.



Технологии Big Data уже обыденность — множество компаний использует их для решения задач своего бизнеса, наряду с автоматизацией и CRM.

