

Основные составляющие Big Data



Любой обмен информацией в социальных сетях неизбежно предполагает её дальнейшую обработку, анализ и активное использование данных компаниями. Пользователи генерируют статистику. Любое подключенное устройство отправляет материал на центральные серверы. Если Big data — технология, то пользователи её ресурсы.

Когда возникла технология Big Data: конец 2000-х годов

Ключевые характеристики:

- децентрализованный способ хранения
- нетрадиционный формат структурирования данных
- полуструктурированные данные
- неструктурированные данные
- горизонтальная модель хранения и обработки данных
- слабая взаимосвязь элементов

Лидеры по объёму по объёму создаваемых данных за последние 12 месяцев: Китай

Прогнозирование: рынок аналитики Big Data будет расти на 35,9% ежегодно

BIG DATA



Преимущества технологии Big Data:

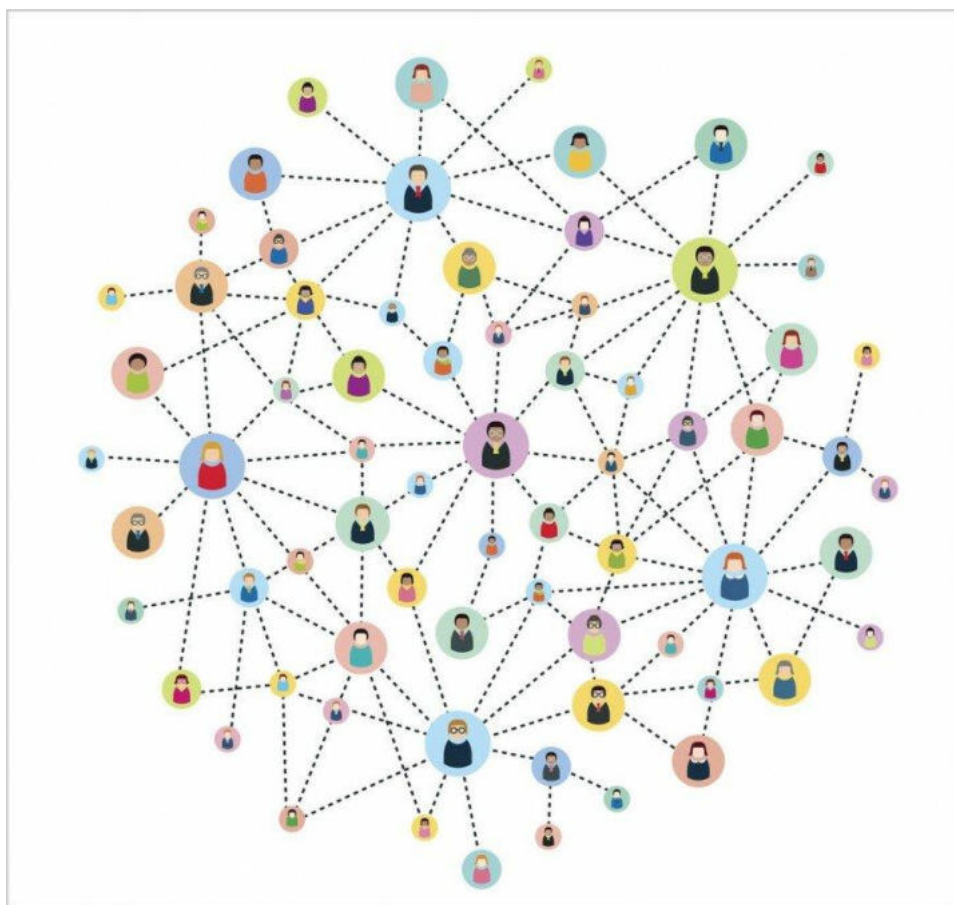
- Наиболее рациональное использование ресурсов
- Фундаментальное изучение портрета клиентской базы
- Предотвращение мошенничества и профилактика хакерских атак

Big Data используют для углублённого взаимодействия с целевой аудиторией. Так, технология оптимизирует операции и предотвращает угрозу взлома программ. За последние несколько лет Google, Uber и Amazon открыли сотни вакансий для специалистов Data science.



Big data — это различные инструменты, подходы и методы обработки структурированных и неструктурированных данных для решения конкретных целей и задач.

Неструктурированные данные — данные, которые находятся в хранилище организации без какого-либо намерения их оркестровки, структурирования или категоризации. Как правило, анализ неструктурированных данных включает анализ каждого объекта. К ним относятся документы, медиа-файлы и изображения. Анализ неструктурированных данных обычно проводится для поиска информации, скрытых тенденций и взаимосвязей внутри элементов информационного потока.

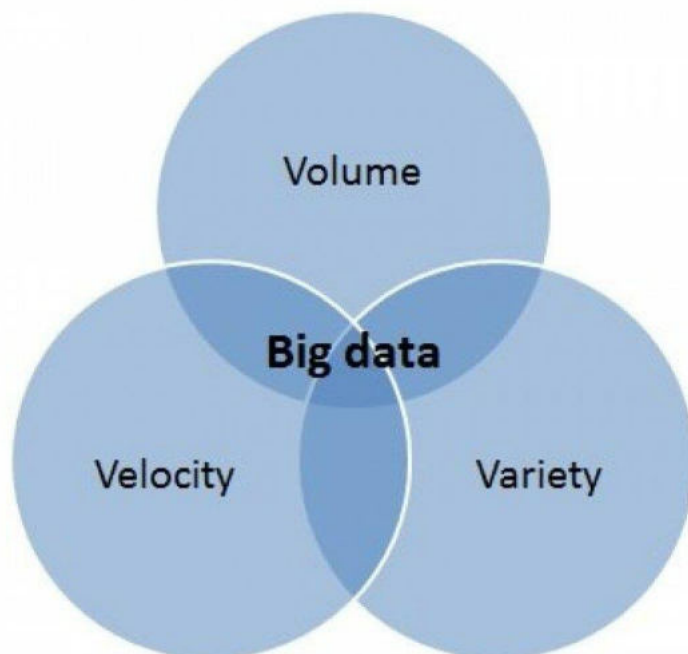


Хотя термин «Big Data» является относительно новым, сам процесс сбора и хранения больших объемов информации для дальнейшего анализа получил развитие в начале 2000-х годов, когда отраслевой аналитик Даг Лейни сформулировал основное определение составляющих Big Data:

1. **Объем.** Организации собирают данные из различных источников, включая бизнес-транзакции и социальные сети. Раньше хранение было бы проблемой, но новые технологии (такие как Hadoop) облегчили эту задачу. **Hadoop** — свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения программ распределенных вычислений.
2. **Скорость.** Данные поступают с беспрецедентной скоростью и должны обрабатываться своевременно. RFID-метки, датчики и

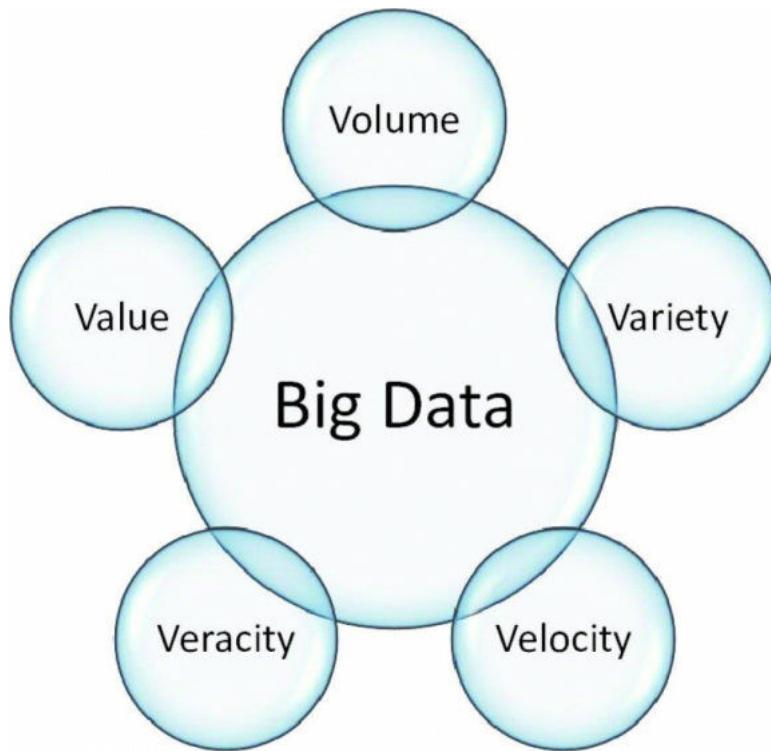
интеллектуальные измерения приводят к необходимости обрабатывать потоки данных в реальном времени.

3. **Разнообразие.** Данные поступают во всех типах форматов: от структурированных числовых данных в традиционных базах до неструктурированных текстовых документов, электронной почты, видео, аудио, биржевых данных и финансовых транзакций.



В SAS (Американская частная компания, разработчик технологического программного обеспечения и приложений углубленной аналитики, исследования данных и управления данными) рассматривается два дополнительных измерения Big data:

4. **Изменчивость.** Помимо увеличения скорости и разнообразия данных, потоки данных могут сильно не соответствовать периодическим пикам. Ежедневные, сезонные и вызванные событиями пиковые нагрузки данных могут быть сложными в управлении. В частности, если речь идёт о неструктурированных данных.
5. **Сложность.** Сегодняшние данные поступают из разных источников. Это затрудняет их связывание и преобразование в разных системах.



Источник: https://zen.yandex.ru/media/ludi_rosta/tehnologiia-big-data-5daf25d23639e600af7a6768