



Цели урока

Узнаете :

Что такое Bigdata (бигдейта).

Положительные и отрицательные стороны использования Bigdata (бигдейта).

Каждые 1,5 года объем информации в интернете удваивался в 2 раза на данный момент общий объем всего контента в интернете составляет 8 Зеттабайт (хотя с учетом развития интернета и увеличения скорости передачи данных это число скорее всего в несколько раз больше)!!!

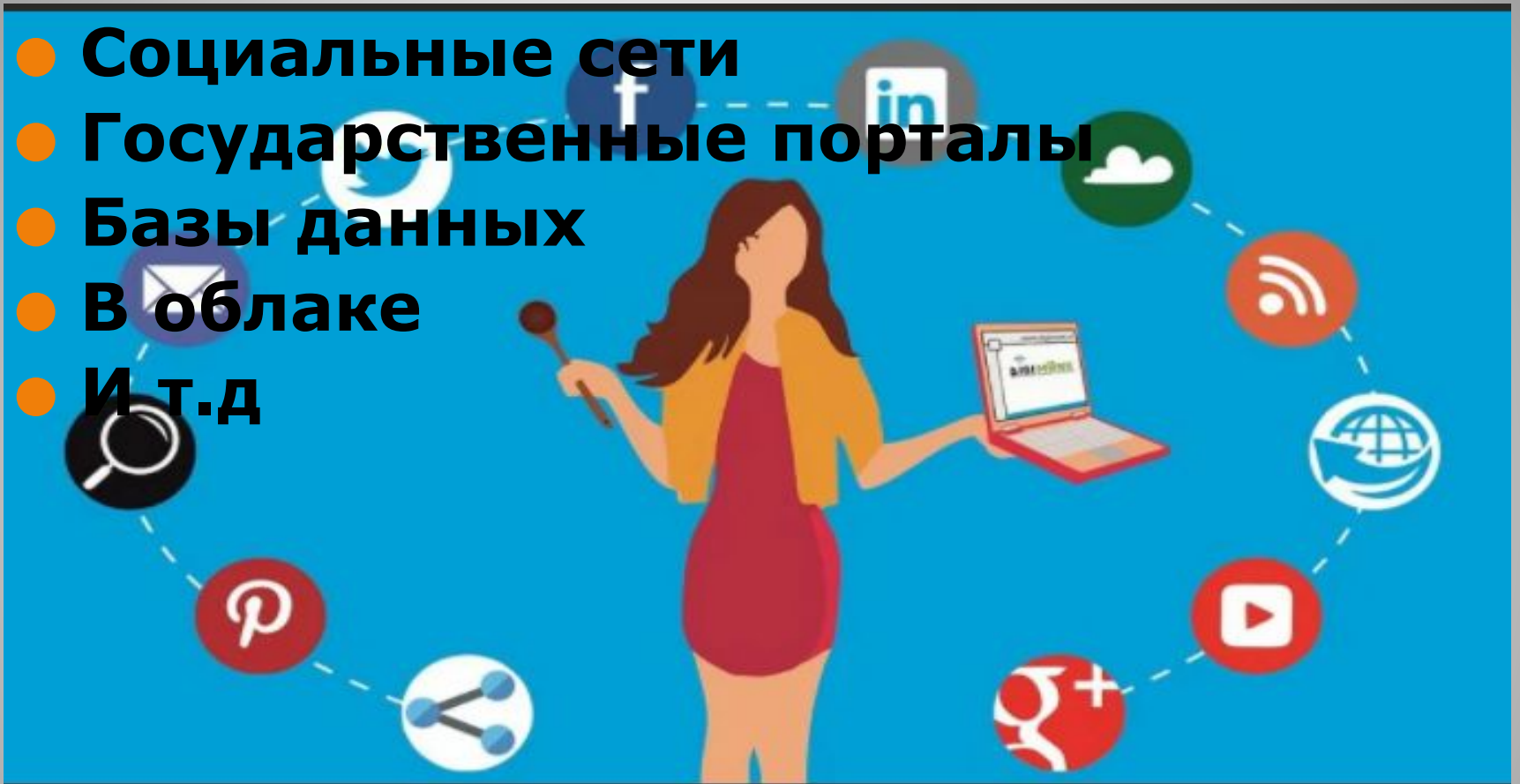
Хранение и анализ больших данных

Big DATA .

Big Data (дословно – «большие данные»), данные – это величины, знаки или символы, которыми оперирует компьютер и которые могут храниться и передаваться в форме электрических сигналов, записываться на магнитные, оптические или механические носители.



- Социальные сети
- Государственные порталы
- Базы данных
- В облаке
- И т.д



Где мы храним данные? И кто их анализирует?

- Knewton — нью-йоркская компания, которая специализируется на индивидуальном обучении. С 2008 года привлекла \$157 млн от 22 инвесторов. Создает цифровые курсы адаптирующиеся к индивидуальным знаниям и особенностям обучающегося. Компания одной из первых стала активно применять технологии анализа данных в сфере образования.
- Платформа анализирует собранные данные и возвращает их приложению в виде рекомендаций преподавателю или указаний о том, какой блок контента нужно показать студенту следующим.

Knewton- индивидуальное обучение

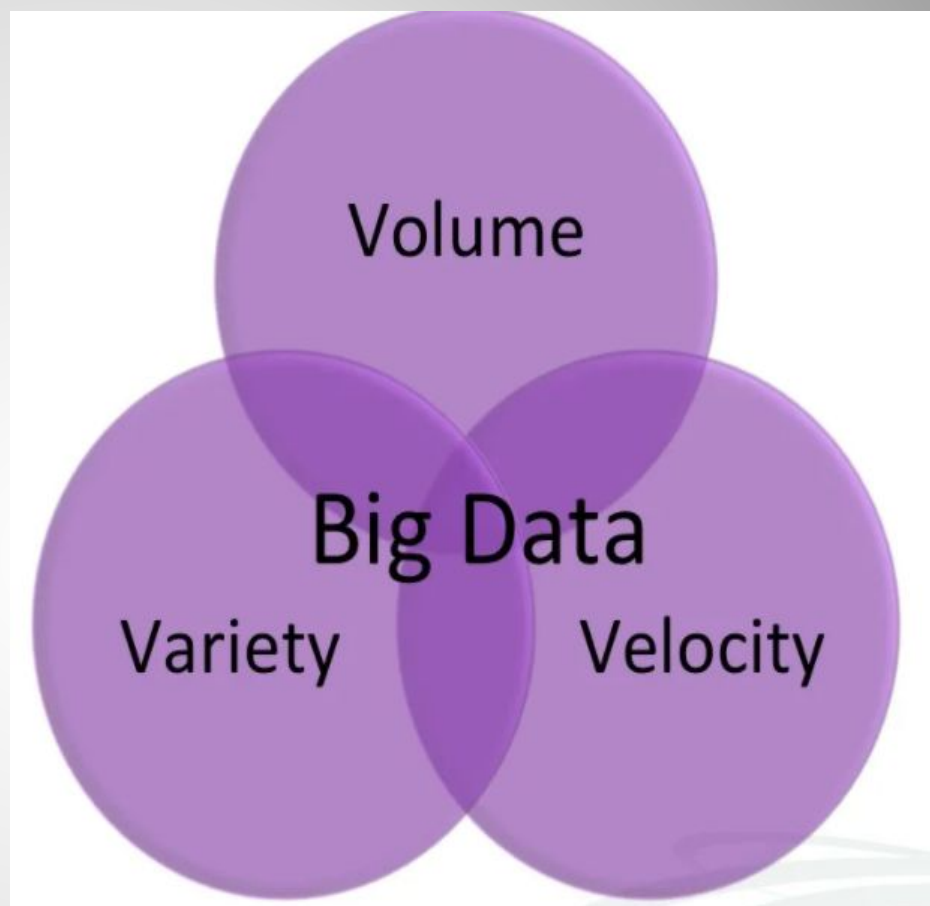
- Калифорнийская компания Cloudera создала собственный инструментарий для работы с большими данными на основе Apache Hadoop (открытое программное обеспечение). Этот сервис предлагает услуги по управлению данными и аналитике на основе технологии, позволяющей в одном месте хранить, обрабатывать и анализировать данные. В компании работает 1400 человек.

**Cloudera — «облачная»
платформа для анализа данных**

- Нью-йоркский стартап Flatiron Health разрабатывает платформу с постоянно обновляющейся информацией о раковых заболеваниях для пациентов и докторов, которые с ее помощью могут заниматься аналитикой и принимать более правильные решения о терапии.
- Flatiron Health был основан бывшими сотрудниками Google Заком Вайнбергом и Нетом Тёрнером в 2010 году. Основатели столкнулись с болезнью лицом к лицу — несколько членов их семей и близкие друзья боролись с раком. Это подтолкнуло их начать исследование индустрии онкологии. В ходе него они обнаружили, что многие медицинские центры и ученые не обладают даже базовыми данными и средствами для аналитики — так и появился этот проект.

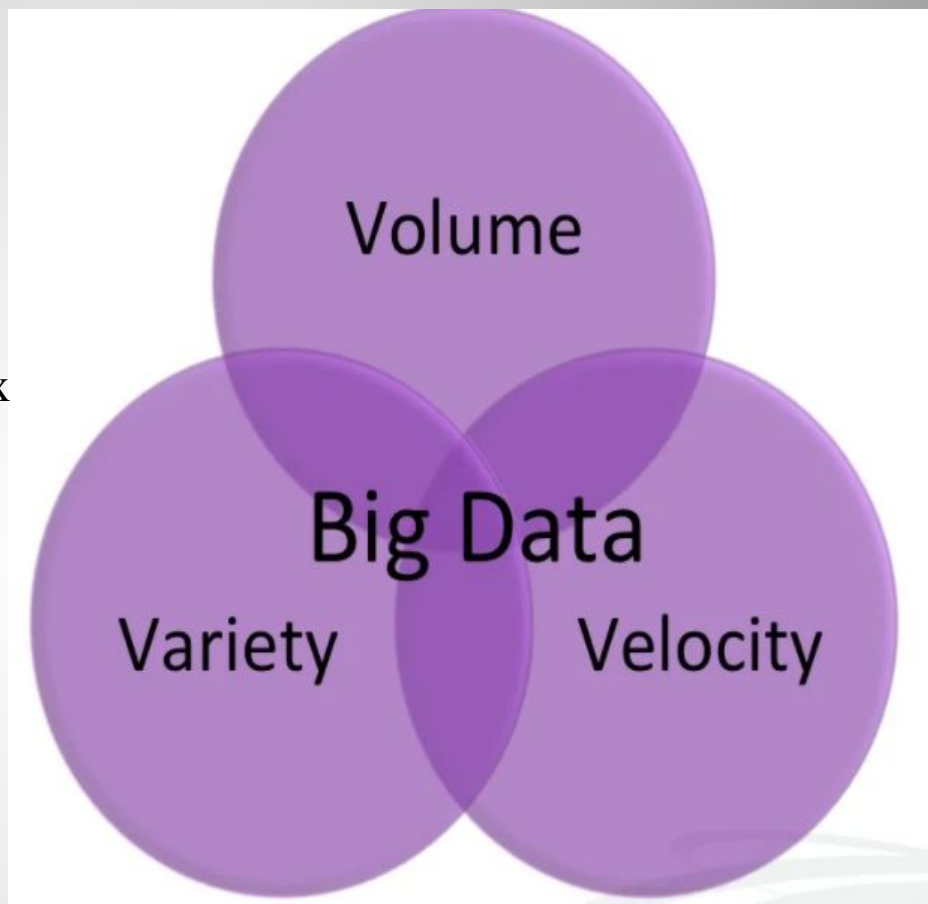
Flatiron Health — медицинская платформа для аналитики раковых заболеваний

1. Объем. Сам по себе термин Big Data связан с большим размером. Размер данных – важнейший показатель при определении возможной извлекаемой ценности. 6 миллионов людей ежедневно используют цифровые медиа, что, по предварительным оценкам, генерирует 2.5 квинтиллиона байт данных. Поэтому объем – первая характеристика для рассмотрения.



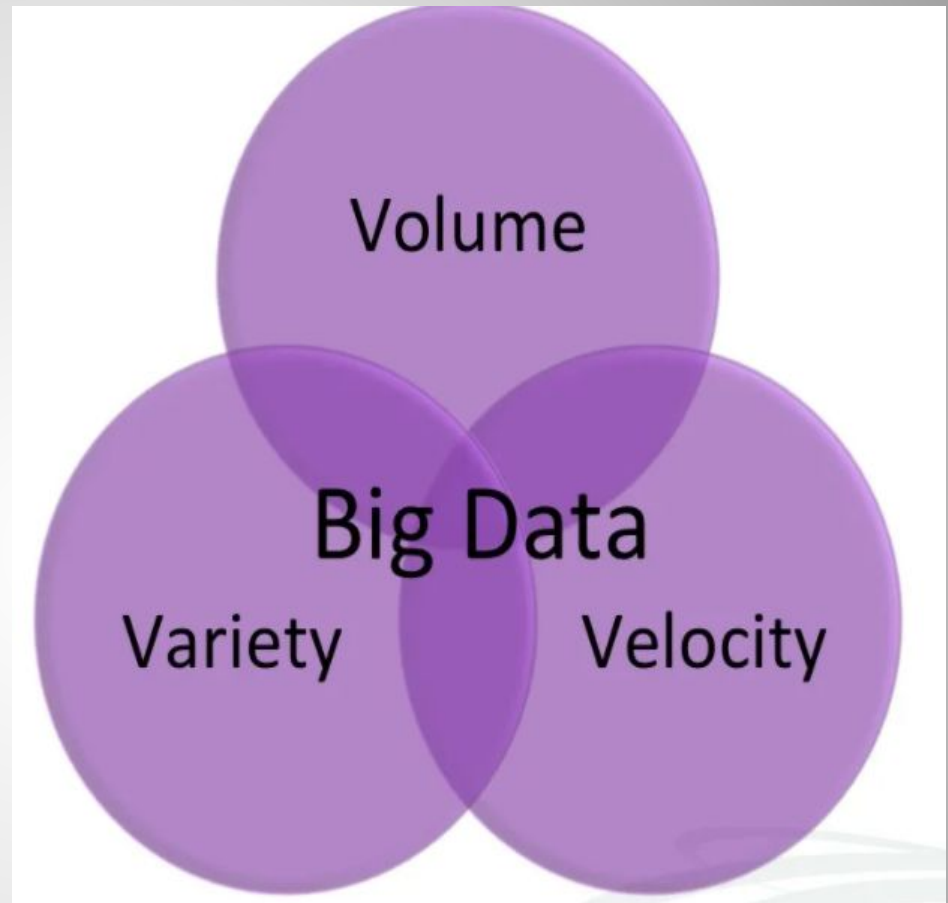
Характеристика больших данных.

2. Разнообразие. Этот аспект характеризуют гетерогенные источники и природу данных, которые могут быть как структурированными, так и неструктурированными. Раньше электронные таблицы и базы данных были единственными источниками информации, рассматриваемыми в большинстве приложений. Сегодня же данные в форме электронных писем, фото, видео, PDF-файлов и аудио также рассматриваются в аналитических приложениях. Такое разнообразие неструктурированных данных приводит к проблемам при сборе, хранении и анализе: 27% компаний не уверен, что работают с подходящими данными.



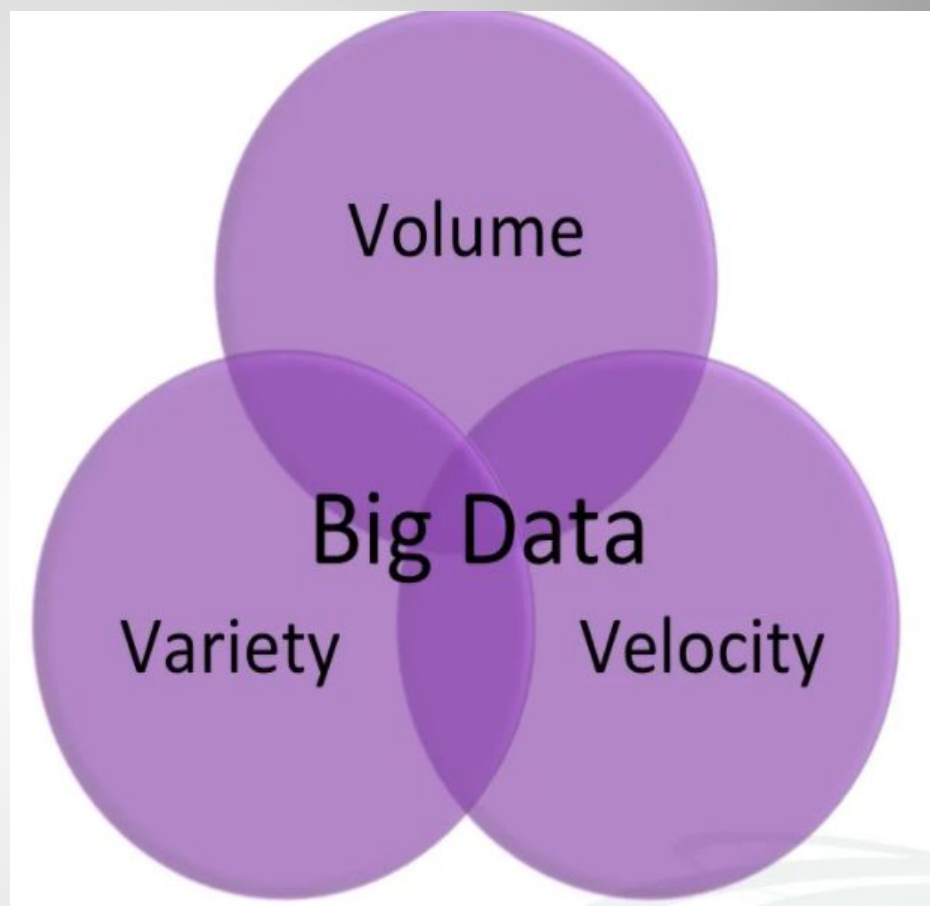
Характеристика больших данных.

3. Скорость генерации. То, насколько быстро данные накапливаются и обрабатываются для удовлетворения требований, определяет их потенциал. Скорость определяет быстроту притока информации из источников – бизнес-процессов, логов приложений, сайтов социальных сетей и медиа, сенсоров, мобильных устройств. Поток данных огромен и непрерывен во времени.



Характеристика больших данных.

4. Изменчивость описывает непостоянство данных в некоторые моменты времени, которые усложняют обработку и управление. Так, например, большая часть данных неструктурирована по своей природе



Характеристика больших данных.

- **Сбор данных** из разных источников.
- Улучшение бизнес-процессов через **аналитику в реальном времени**.
- **Хранение** огромного объема данных.
- **Инсайты**, то есть Big Data более проницательна к скрытой информации при анализе структурированных и полуструктурированных данных.
- Большие данные помогают уменьшать риски и **принимать умные решения** благодаря подходящей риск-аналитике.

Преимущества больших данных.

- **1. Конфиденциальность данных.** Big Data, которую мы сегодня генерируем, содержит много информации о нашей личной жизни, на конфиденциальность которой мы имеем полное право.
- **2. Защита данных.** Даже если мы решаем, что нас устраивает то, что у кого-то есть наши данные для определенной цели, можем ли мы доверить им сохранность и безопасность наших данных?

Недостатки больших данных.

Назовите характеристики больших данных:

Объем

Анализ

Защита

Разнообразие

Конфиденциальность

Скорость генерации

Изменчивость

Задание.

Назовите положительные и отрицательные стороны больших данных

Преимущества	Недостатки

Задание.

Домашнее задание.

Написать краткий конспект

Тема: BigDATA.

Кардикулов Р.А. Информатика 10

класс стр 163-165

Задание.