BIG DATA

МОСКВА 2015

BIG DATA

Термин «большие данные» относится к наборам данных, размер которых превосходит возможности типичных баз данных (БД) по занесению, хранению, управлению и анализу информации.

McKinsey Institute

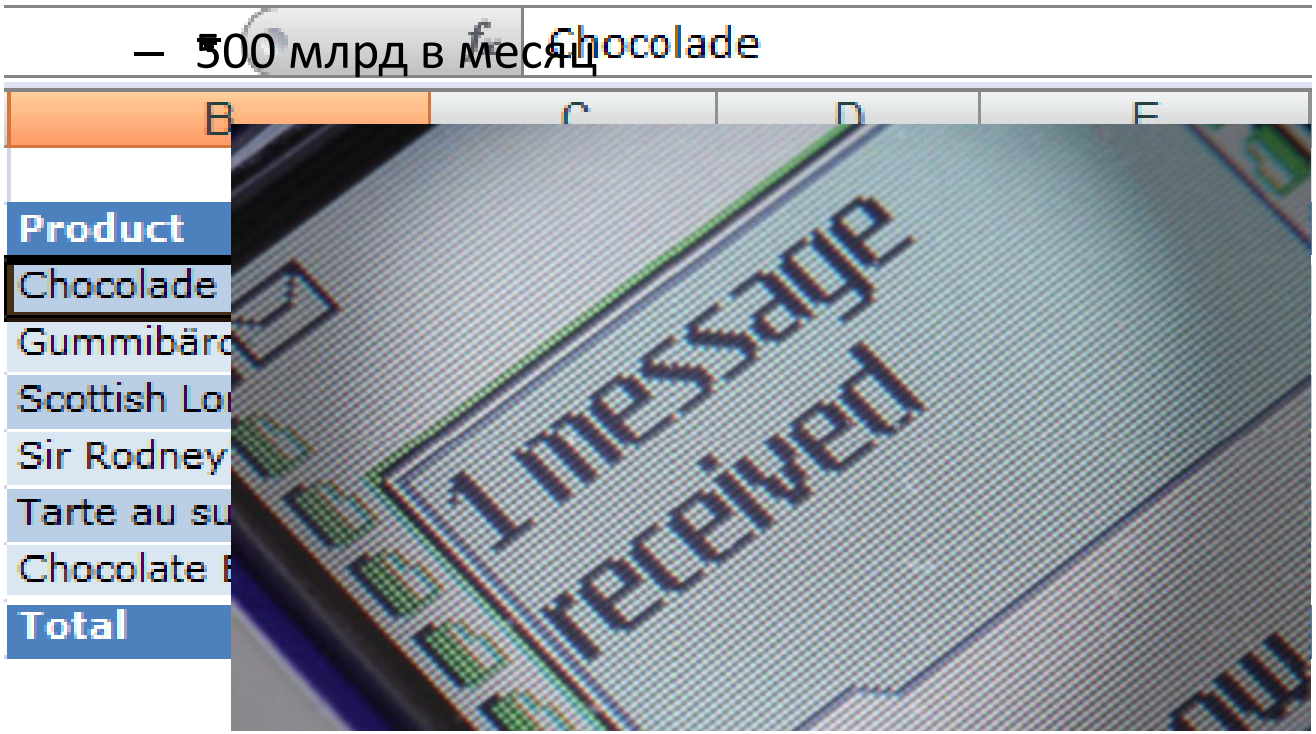
Основные типы BIG DATA

- Структурированные
  - Табличные данные, аналитика
- Медиа
  - 1 GB+ файлы с видео, изображениями
- Объектные
  - Миллионы случайных объектов

БелАЗ
011

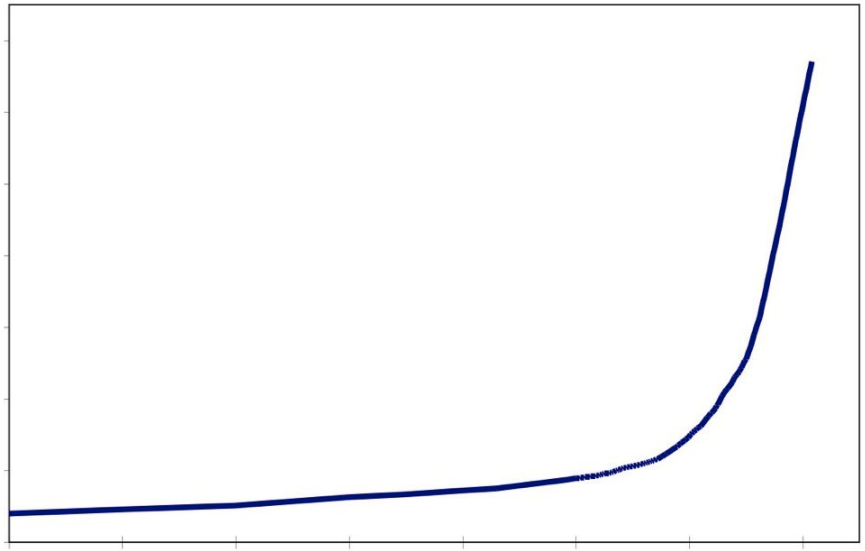
Структурированные

- Структурированные данные
  - Набор однотипных компонентов
- Например
  - SMS
  - 500 млрд в месяц



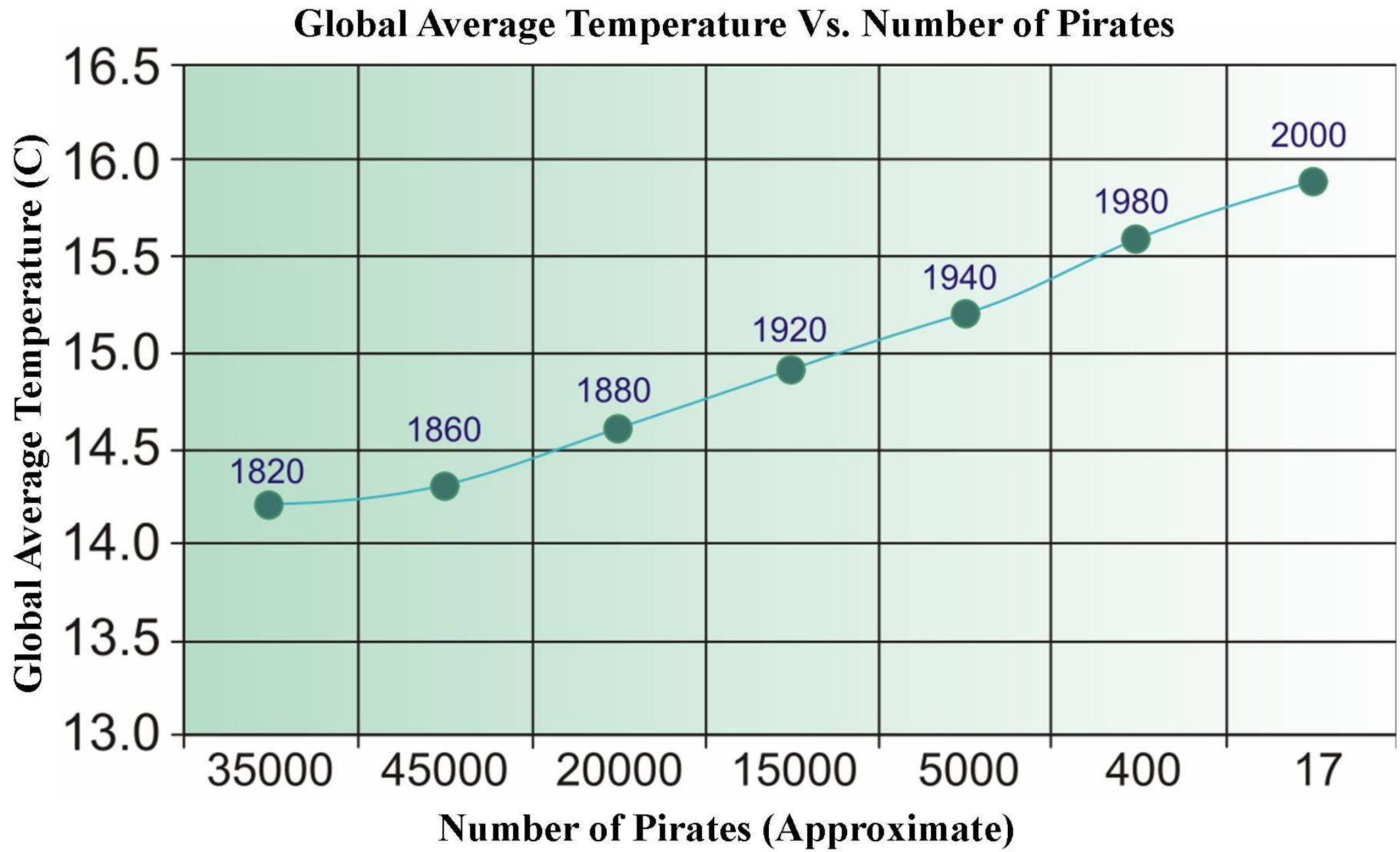
Аналитика

- Есть ложь, есть наглая ложь и статистика
- 1я Мировая Война
  - Обязательное ношение стальных касок
  - Взрывной рост количества ранений в голову



Аналитика

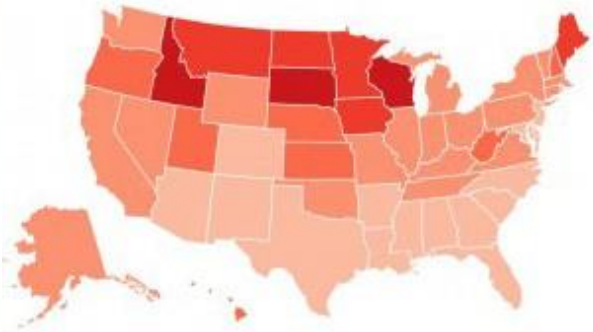
- Есть ложь, есть наглая ложь и статистика
- 1я Мировая Война
  - Обязательное ношение стальных касок
  - Взрывной рост количества ранений в голову
  - Резкое падение смертности
- Ценность аналитики
  - Полнота данных
  - Корректность модели
  - GIGO – мусор на входе, мусор на выходе


Global Average Temperature Vs. Number of Pirates
Global Average Temperature (C)
16.5
16.0
15.5
15.0
14.5
14.0
13.5
13.0
1820
1860
1880
1920
1940
1980
2000
35000
45000
20000
15000
5000
400
17
Number of Pirates (Approximate)

Новые сюрпризы

- 12 летняя девочка
  - Предложения для беременных
  - Родители в ярости
  - Анализ активности в интернет магазине
- Беременность подтвердилась

Новые возможности



- Наблюдение за эпидемией гриппа
  - Twitter 2011-2013
  - 5000 публичных твитов в минуту
- В реальном времени
  - Совпадает с информацией CDC

Медиа данные

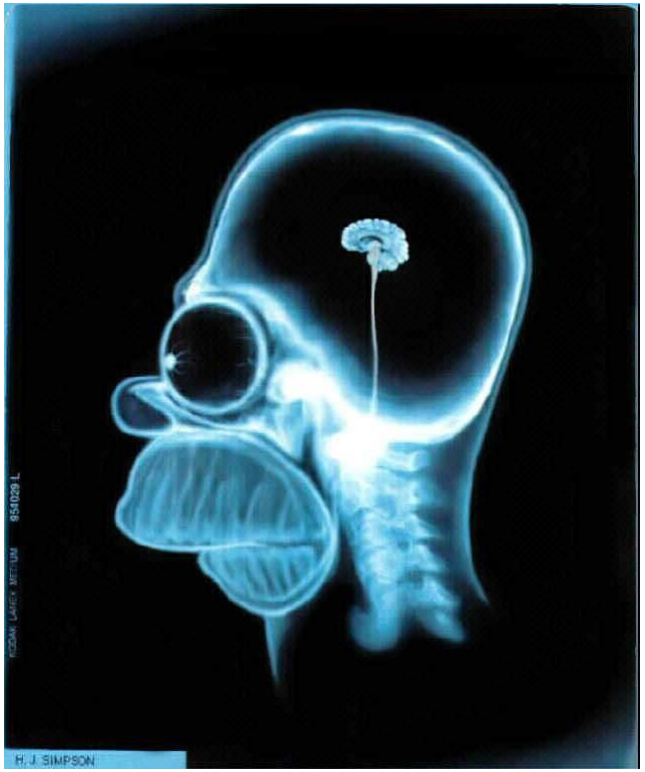
- Много БОЛЬШИХ файлов
  - 1 GB+ файлы с видео, изображениями
  - Видеомонтаж, видеонаблюдение

Медиа данные

- Много БОЛЬШИХ файлов
  - 1 GB+ файлы с видео, изображениями
  - Аэрофотосъемка

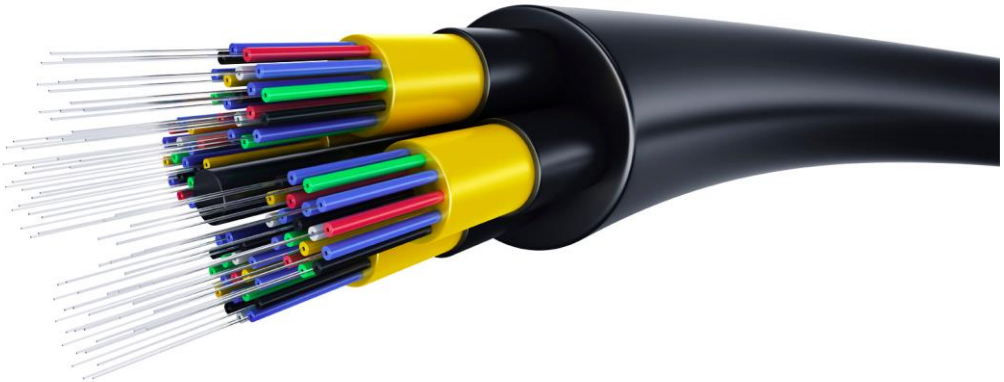
Медиа данные

- Много БОЛЬШИХ файлов
  - 1 GB+ файлы с видео, изображениями
  - Цифровая медицина (рентген, КТ, МРТ)



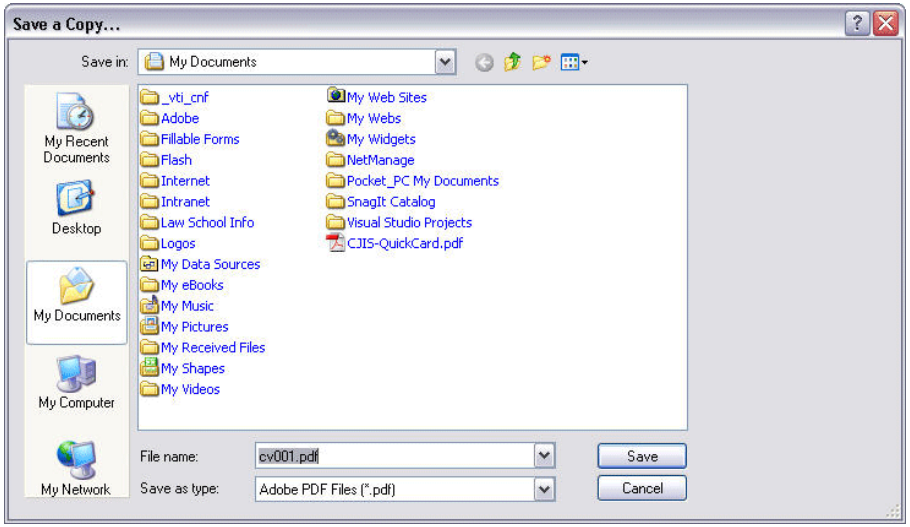
Медиа данные

- Много БОЛЬШИХ файлов
  - 1 GB+ файлы с видео, изображениями
  - Видеомонтаж
  - Аэрофотосъемка
  - Цифровая медицина (рентген, КТ, МРТ)
- Высокая пропускная способность



Объектные

- Миллионы случайных объектов
  - Облачные хранилища



OneDrive

Dropbox

Объектные

- Миллионы случайных объектов
  - Архивные системы

Объектные

- Миллионы случайных объектов
  - Множество совпадающих
- Облачные хранилища
  - Миллионы пользователей
- Архивные системы
  - Годы и десятки лет гарантированного хранения

Трансформация

- Монолитные системы
  - Обработка за разумное время невозможна
  - Максимальный размер ограничен
  - Производительность ограничена
- Массивно-параллельные системы
  - Неограниченный рост
  - Производительность – обратная сторона размера

Что делать?

- BIG DATA – это серьезно
  - Как сэкономить
  - Где заработать
  - Деньги – это всегда серьезно
- BIG DATA – не только аналитика
- BIG DATA – это будущее

Контактная информация:

Россия, 142784, г. Москва,
Киевское ш., 22-й км., вл. 6, стр. 1, БП ComCity
Тел.: + 7(495) 775 31 20, 363 01 33
Факс: + 7 (495) 363 01 34
E-mail: info@step.ru
www.step.ru