



Курс: Аналитическая журналистика  
Преподаватель Шерчалова Е. В.  
Факультет журналистики  
ИМПЭ им. А. С. Грибоедова



# Big data —



это различные инструменты, подходы и методы обработки как структурированных, так и неструктурированных данных для того, чтобы их использовать для конкретных задач и целей.



способность управлять большими объемами разнородных данных со скоростью, достаточной для анализа таких данных в реальном времени и своевременного реагирования.



Представьте такой магазин, где нет системы выкладки: макароны лежат рядом с собачьим кормом и зубными щетками. Как в нем что-то найти?

Это же

невозможно!

Нужна

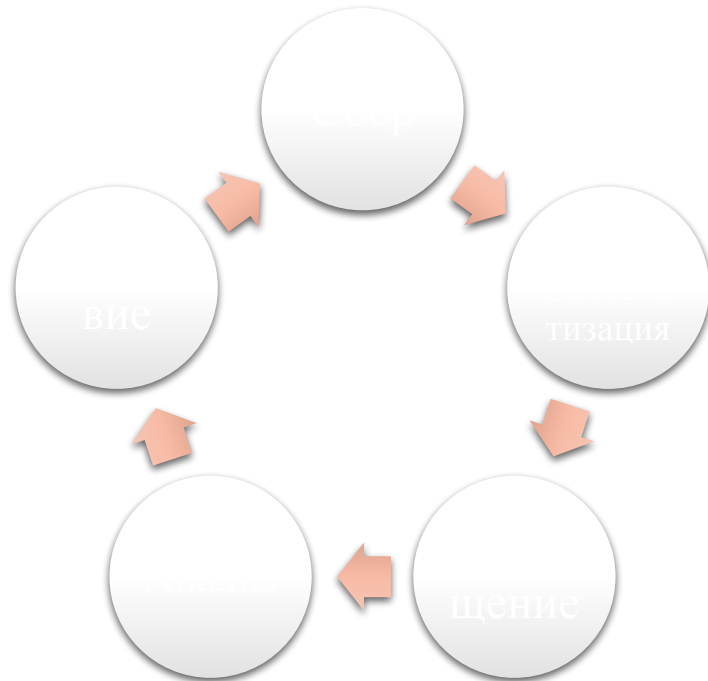
систематизация.

Мир данных часто представляет собой вот такой магазин, и, чтобы разобраться, нужно приложить максимум усилий.

## Признаки:

- Объем: количество данных;
- Скорость: быстрота обработки;
- Вариативность: количество разных типов данных.

Данные в обобщенном виде.



## Влияние больших данных на систему управления данными, факторы:

- Принципиально новые источники данных – смартфоны, планшеты, датчики;
- Эффективные и экономически выгодные технологии сбора, хранения и анализа информации, способные работать со сколь угодно большими объемами данных.



# Структурированные данные имеют определенные длину и формат.

## Источники:

- Данные, генерируемые машинами, создаваемыми компьютерами и другими устройствами без вмешательства человека:
  - Медицинские приборы
  - Метки радиочастотной идентификации (например, отслеживание перемещение контейнеров с продукцией)
  - GPS- датчики в смартфонах (исследование поведения покупателей)
  - Лог-файлы – все виды данных о функционировании серверов, приложений, сетей, собираемые в процессе их работы
  - Данные из точек продаж (штрихкод)
  - Финансовые данные – получаемые из финансовых систем (не все финансовые данные генерируются машиной) – обозначение компании и её стоимость
- Данные, генерируемые человеком: создаются человеком при взаимодействии с компьютером:
  - Входные данные (то, что вводит человек – соц. дем.)
  - Данные о посещении веб-сайтов
  - Игровые данные



**Неструктурированные данные не имеют определенного формата. Примерно 80% всех получаемых данных.**

➔ **Машинные данные:** спутниковые снимки, метеорологические данные, научные данные (графики сейсмической активности), записи с камер наблюдения

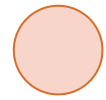
---

➔ **Данные, генерируемые человеком:** тексты документов, соц.сети, информационное наполнение веб-сайтов (YouTube).

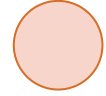
---



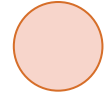
# Техники и методы анализа, применимые к Big data



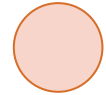
Data Mining



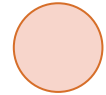
Краудсорсинг



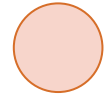
Смешение и интеграция данных



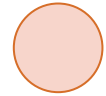
Машинное обучение



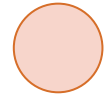
Искусственные нейронные сети



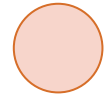
Распознавание образов



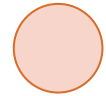
Прогнозная аналитика



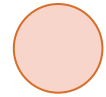
Имитационное моделирование



Пространственный анализ



Статистический анализ



Визуализация аналитических данных

# Накопление данных всем миром



**2003:** 5 эксабайтов данных  
(1 ЭБ = 1 млрд гигабайтов)

**2008:** 0,18 зеттабайта  
(1 ЗБ = 1024 эксабайта)

**2011:** 1,76 зеттабайта

**2013:** 4,4 зеттабайта

**2015:**  
6,5 зеттабайта

**2025:** вырастет в  
10 раз

**2020:** 40-44  
зеттабайтов

\* По данным компании IBS



«Есть пять вариантов использования, которые являются наиболее популярными»,  
генеральный директор Pentaho Квентин Галливэн:

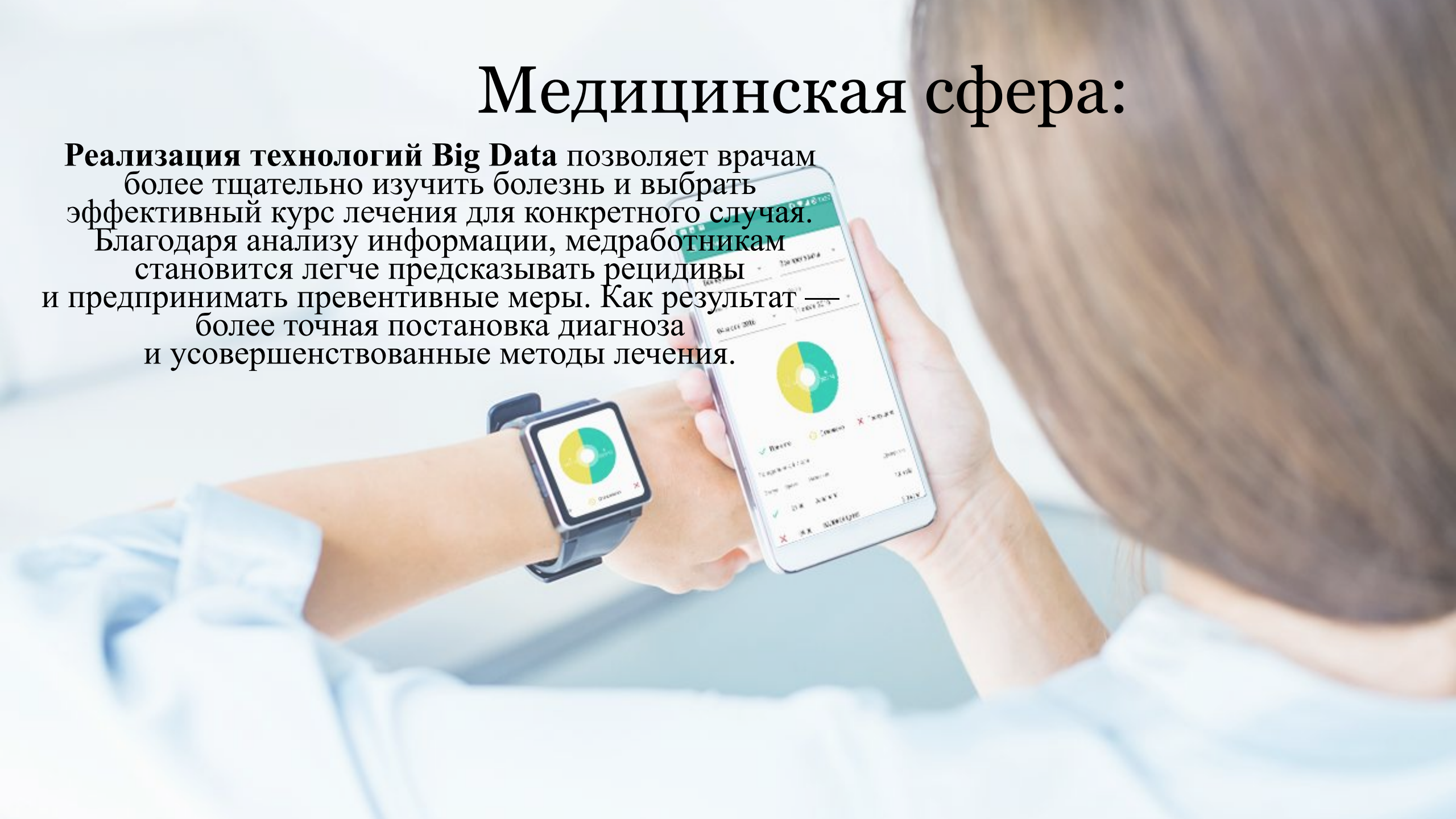
- **Составление портрета клиента.** Такое применение позволяет получить наиболее полную картину о потенциальных и имеющихся клиентах. Заказчики хотят знать, сколько времени посетители проводят на сайте, какие страницы посещают, где дольше задерживаются, когда и, главное, почему уходят.
- **Внедрение в обычные вещи.** Второе популярное направление — управляемые датчики и аппаратные средства, помогающие собирать данные о пользе, здоровье пользователей или безопасности устройств.
- **Оптимизация базы данных.** Мы помогаем клиентам определить, какие данные лучше подходит для более дешевой вычислительной платформы.
- **Повышение корпоративной эффективности.** К примеру, крупный институт захотел перейти к отчетности данных, для этого необходимо было предъявить банку архивы финансового отдела. Поскольку те хранились в нескольких различных базах данных, нам пришлось сначала загрузить их и обработать в Hadoop, что позволило банку получить единую удобную базу данных.
- **Информационная безопасность клиентов.** Использование кейсов Big Data для обеспечения безопасности хранения данных, также помогает обеспечить более эффективный метод хранения большого объема информации.



# Медицинская сфера:

**Реализация технологий Big Data** позволяет врачам более тщательно изучить болезнь и выбрать эффективный курс лечения для конкретного случая.

Благодаря анализу информации, медработникам становится легче предсказывать рецидивы и предпринимать превентивные меры. Как результат — более точная постановка диагноза и усовершенствованные методы лечения.





Сети магазинов Target с помощью  
глубинного анализа данных  
и собственной системы  
прогнозирования удастся с высокой  
точностью  
определить — беременна женщина  
или нет.



На основе технологий Больших Данных в городе Лонг-Бич работают «умные» счетчики воды, которые используются для пресечения незаконного полива. Ранее они применялись с целью сокращения потребления воды частными домовладениями (максимальный результат — сокращение на 80%).



На основании данных, полученных от датчиков дорожных камер, власти производят контроль работы светофоров, что в свою очередь позволяет регулировать трафик. Под управлением компьютеризованной системы находится порядка 4 500 тысяч светофоров по всему городу. Согласно официальным данным, новый алгоритм помог уменьшить заторы на 16%.



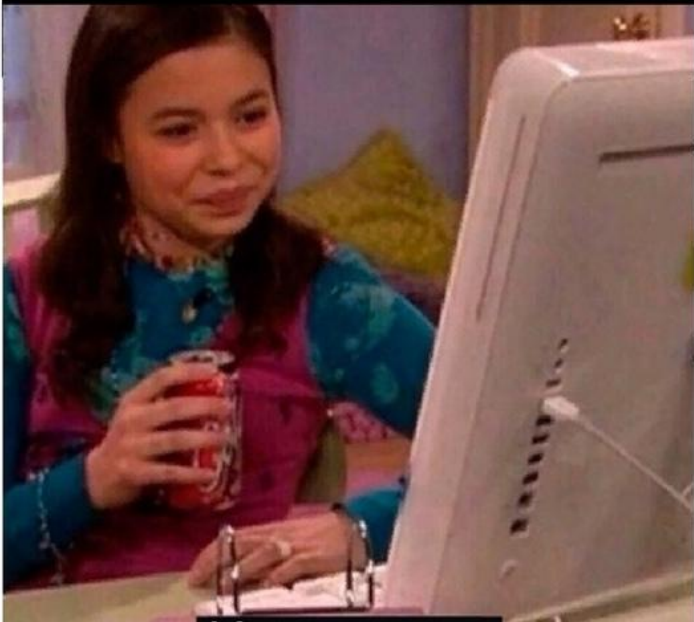
## Продуктивные советы



Yuri Krupenin  
@turbojedi

Советы от гуру биг дата:

1. При разговоре в Zoom надо двигать глазом так, будто вы посматриваете на джиру на соседнем мониторе.
2. Делайте движения курсором, как будто вы УЖЕ листаете рабочую почту.
3. Рычите, порывивайте, урчите: показывайте, что вы **ВОВЛЕЧЕННЫЕ И ПРОДУКТИВНЫЕ**



Интересно

# ПРОБЛЕМА БОЛЬШИХ ДАННЫХ



Даниил Кортес  
@Spanish\_Kortez



Это биг Ксола, большая русская дата

Он не любит просто так увольнять людей, он любит жиры и дашборды



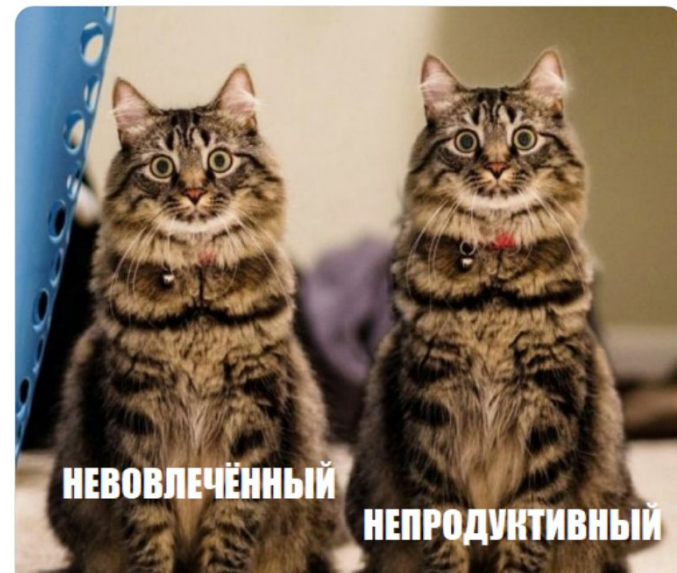
6:05 PM · 4 авг. 2021 г. · Twitter Web App

4 ретвита · 88 отметок «Нравится»



Pavel B. Novikov  
@reinforced\_sc

Кокой ты сегодня?



meduza

ЧТО ВЫ СЛАДКАЯ БУЛОЧКА

«Биг дата показала, что вы сладкая булочка». Только мемы — для наших самых вовлеченных читателей. Про компанию Xsolla, откуда уволили 150 человек

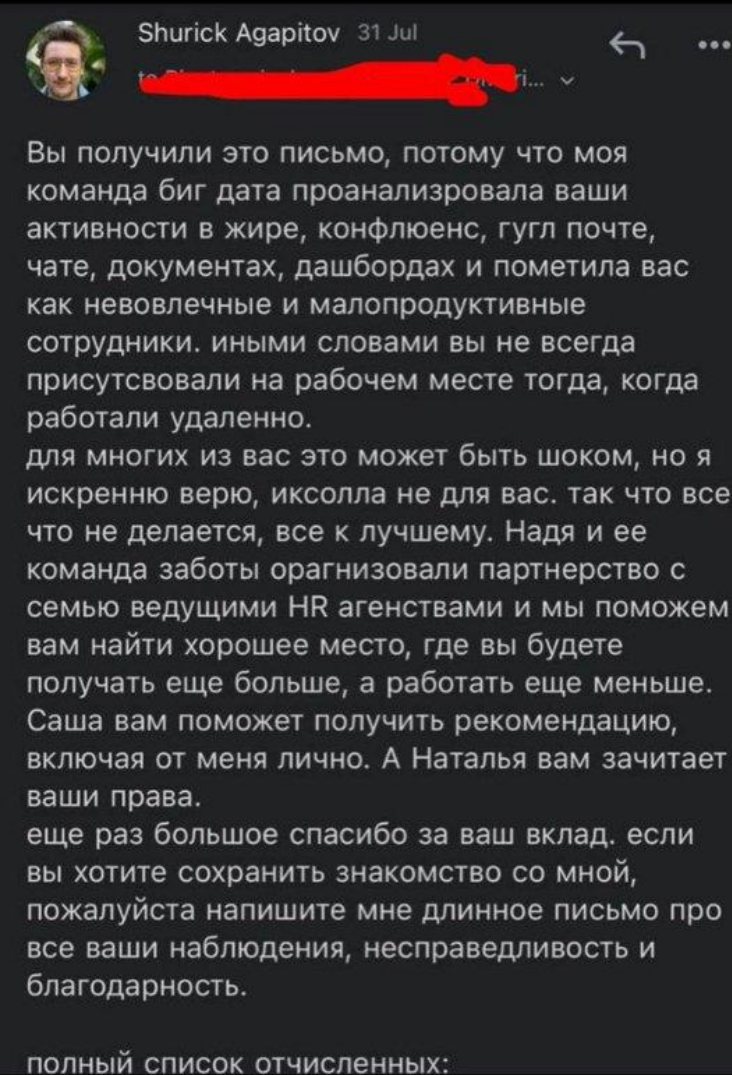
ПАСИБА

## В Перми из компании уволили 150 айтишников: команда Big Data проанализировала их активность на удаленке

Позже Агапитов назвал основной причиной увольнений тот факт, что Xsolla, штаб-квартира которой находится в Лос-Анджелесе, а главный офис разработки — в Перми, перестала показывать 40-процентный рост выручки.

Что говорит сам основатель Xsolla:

<https://meduza.io/feature/2021/08/06/my-formalno-esche-nikogo-ne-uvolili-my-skazali-kompaniya-vas-ne-tsenit>



# Большие данные в ритейле



*«Если посмотреть, какие битвы разыгрываются в интернете вокруг использования различных брендов, — говорит экс-директор по моделированию и анализу данных X5 Retail Group Валерий Бабушкин, — это выглядит как целая патриотическая война. В какие магазины ты ходишь — в эти или в эти? “Да ты что! Как можно в них ходить?!”»*





@ElectricBanned

BIG DATA всё про вас знает. Только понять ничего не может

## Пятерочка

Ваш любимый супермаркет. В этом году вы были там 144 раз и потратили 33 178 ₽. Ваш любимый товар в 2020 — пакет.

Поделиться с друзьями

BIG DATA это...



Это когда ты сливаешь данные 60 млн пользователей...

**НУ ДАВАЙ РАССКАЖИ МНЕ**

**ПРО СЕКРЕТНЫЙ ПРИНЦИП  
АНАЛИТИКИ BIG DATA**

# Большие данные для большого города

<https://postnauka.ru/tv/85785>



# Что еще почитать / посмотреть по теме:

- Big Data в кино: Netflix, Кевин Спейси и генерация сценариев <https://dtf.ru/flood/39198-big-data-v-kino-netflix-kevin-speysi-i-generaciya-scenariev>
- Документальный фильм «Большие данные с человеческим лицом» <https://www.afisha.ru/movie/226995/>
- Документальный фильм «Социальная дилемма» (с большой долей скептицизма)
- Курс «Маленький гид по большим данным» <https://postnauka.ru/courses/83921>
- Курс «Big data смотрит на мир» <https://postnauka.ru/courses/86303>
- Артур Хачуян (генеральный директор SocialDataHub): «Настоящая Big Data в рекламе» [https://www.youtube.com/watch?v=OAzcxDE7X-g&t=680s&ab\\_channel=BBDOGroup](https://www.youtube.com/watch?v=OAzcxDE7X-g&t=680s&ab_channel=BBDOGroup)  
текст статьи: <https://rb.ru/howto/bigdata-advertising/>
- Проблема больших данных в городских исследованиях <https://postnauka.ru/video/83423>

# Литература по теме:

- К. Митник «Искусство быть невидимым. Как сохранить приватность в эпоху Big Data»
- С. Стивен-Давидовиц «Все лгут. Поисковики, Big Data и Интернет знают о вас все»
- У. Девис «Индустрия счастья. Как Big Data и новые технологии помогают добавить эмоцию в товары и услуги»
- А. Благирев «Big Data простым языком»
- В. Мальцев «Карл Маркс и большие данные»
- С. Грингард «Интернет вещей»

# Если тема вас заинтересовала, вы можете почитать материалы по ссылкам:

- Продажи, портреты, маршруты: как компании и города используют Big Data <https://trends.rbc.ru/trends/industry/cmrm/611c26309a794760e2c9927b>
- Big data поглощает мир ретейла [https://m.dp.ru/a/2021/08/09/JA\\_znaju\\_vse\\_tvoi\\_treshhi](https://m.dp.ru/a/2021/08/09/JA_znaju_vse_tvoi_treshhi)
- Как Big Data помогает корпорациям следить за нами <https://trends.rbc.ru/trends/sharing/60e83c249a79476ed8e2593e>
- Big Data. Пять успешных кейсов анализа данных в бизнесе <https://1-a-b-a.com/blog/1110-morozhenoe-bjustgaltery-i-netflix>
- Как и зачем «Ашан» построил платформу для работы с Big Data в публичном облаке <https://habr.com/ru/company/mailru/blog/565664/>
- Мифы и легенды про Big Data <https://habr.com/ru/company/beeline/blog/218669/>
- Что такое Big Data и почему их называют «новой нефтью» [https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c#card\\_5d6c020b9a7947a740fea65c\\_5](https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c#card_5d6c020b9a7947a740fea65c_5)
- Ну, и напоследок – интересная статистика: <https://www.worldometers.info/ru/>