

uplab



Big Data

Большие данные (Big Data, биг дата)

это структурированные и неструктурированные данные огромных объемов и разнообразия, а также методы их обработки, которые позволяют распределенно анализировать информацию.

Анализ больших данных проводят для того, чтобы получить новую, ранее неизвестную информацию. Подобные открытия называют **инсайтом**, что означает озарение, догадку, внезапное понимание.

Признаки, характеристики

1

Volume

величина физического
объёма

2

Velocity

Скорость прироста.
Данные регулярно
обновляются, что требует
их постоянной обработки.

3

Variety

Разнообразие. Данные могут
иметь неоднородные
форматы, быть
неструктурированными или
структурированными
частично.

Источники данных

1

Интернет

соцсети, блоги, СМИ,
форумы, сайты

2

Корпоративные данные

транзакционная деловая
информация, архивы,
базы данных

3

Показания устройств

датчиков, приборов, а также
метеорологические данные,
данные сотовой связи и т. д.

Функции и задачи

Функция	Задача
Big Data — собственно массивы необработанных данных	Хранение и управление большими объемами постоянно обновляющейся информации
Data mining — процесс обработки и структуризации данных, этап аналитики для выявления закономерностей	Структурирование разнообразных сведений, поиск скрытых и неочевидных связей для приведения к единому знаменателю
Machine learning — процесс машинного обучения на основе обнаруженных связей в процессе анализа	Аналитика и прогнозирование на основе обработанной и структурированной информации

Техники и технологии

Техники и методы анализа, применимые к Big data

1. Data Mining;
2. Краудсорсинг;
3. Машинное обучение;
4. Искусственные нейронные сети;
5. Распознавание образов;
6. Прогнозная аналитика;
7. Имитационное моделирование;
8. Пространственный анализ;
9. Статистический анализ;
10. Визуализация аналитических данных.

Технологии

1. NoSQL;
2. MapReduce;
3. Hadoop;
4. R;
5. Аппаратные решения.
6. Столбцовые БД (ClickHouse)

Аппаратные решения

1



2



3

Горизонтальная масштабируемость. любая система, которая обрабатывает большие данные должна быть расширяемой

Отказоустойчивость

Локальность данных — для снижения издержек данные необходимо обрабатывать на том же сервере, где они хранятся.

Разница подходов

Традиционная аналитика

1. Постепенный анализ небольших пакетов данных
2. Редакция и сортировка данных перед обработкой
3. Старт с гипотезы и ее тестирования относительно данных
4. Данные собираются, обрабатываются, хранятся и лишь затем анализируются

Big Data

1. Обработка сразу всего массива доступных данных
2. Данные обрабатываются в их исходном виде
3. Поиск корреляций по всем данным до получения искомой информации
4. Анализ и обработка больших данных в реальном времени, по мере поступления

Big Data в бизнесе

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

1. Поставщики инфраструктуры — решают задачи хранения и предобработки данных. Например: IBM, Microsoft, Oracle, Sap
2. Датамайнеры — разработчики алгоритмов, которые помогают заказчикам извлекать ценные сведения. Среди них: Yandex Data Factory, CleverData;
3. Системные интеграторы — компании, которые внедряют системы анализа больших данных на стороне клиента. К примеру: «Форс», «Крок» и др
4. Потребители — компании, которые покупают программно-аппаратные комплексы и заказывают алгоритмы у консультантов. Это «Сбербанк», «Газпром», «МТС», «Мегафон»
5. Разработчики готовых сервисов — предлагают готовые решения на основе доступа к большим данным. Они открывают возможности Big Data для широкого круга пользователей

Драйверы и ограничители в бизнесе в России

Драйверы	Ограничители
Высокий спрос на Big Data для повышения конкурентоспособности с помощью возможностей технологий	Необходимость обеспечивать безопасность и конфиденциальность данных
Развитие методов обработки медиафайлов на мировом уровне	Нехватка квалифицированных кадров
Реализация отраслевого плана по импортозамещению программного обеспечения	В большинстве российских компаний объем накопленных информационных ресурсов не достигает уровня Big Data
Создание технопарков, которые способствуют развитию информационных технологий	Новые технологии сложно внедрять в устоявшиеся информационные системы компаний
Перенос на территорию России серверов, которые обрабатывают персональную информацию	Высокая стоимость технологий
Государственная программа по внедрению грид-систем — виртуальных суперкомпьютеров, которые распространяются по кластерам	Рост цен на импортную продукцию

Big Data в маркетинге

Зачем нужны большие данные в маркетинге? Анализ массивов информации о компании открывает новые возможности:

1. Понять работу бизнеса в цифрах.
2. Изучить конкурентов.
3. Узнать своих клиентов.

Сервисы Big Data:

1. 1С-Битрикс BigData - сервис персонализации торговых предложений
2. RTB Media - Сервис по управлению закупками цифровой рекламы
3. Alytics - Система сквозной аналитики с автоматизацией контекстной рекламы и интерактивными отчетами

Заключение

Спасибо :))

uplab



+7 499 653 78 83

www.uplab.ru



Москва

127055,
ул.
Новослободская,
д. 61, стр. 2

info@uplab.ru



Чебоксары

428018,
ул. Нижегородская,
д. 6, стр. 2