

Лекция 1. Часть 1.

Базы данных: введение

Базы данных: введение

1. Значение и классификация данных.
2. Базы данных как часть технологий работы с данными.

Что такое данные?

Дословно: Data (англ. *данные*) = «данность», «факт».

ISO, 1996: Формы представления информации, с которыми имеют дело информационные системы и их пользователи.

ISO, 2015: Поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи, или обработки.

Данные – факты, текст, графики, картинки, звуки, аналоговые или цифровые видео-сегменты, представленные в форме, пригодной для хранения, передачи и обработки.

Для чего нужны данные?

Классические бизнес-задачи

- формирование отчетностей
- учетные системы
- описательная аналитика

Современные задачи

- воспроизведение связей между событиями и результатом (machine learning)
- таргетированные коммуникации
- распознавание образов (лиц)
- классификация

Для чего нужны данные?

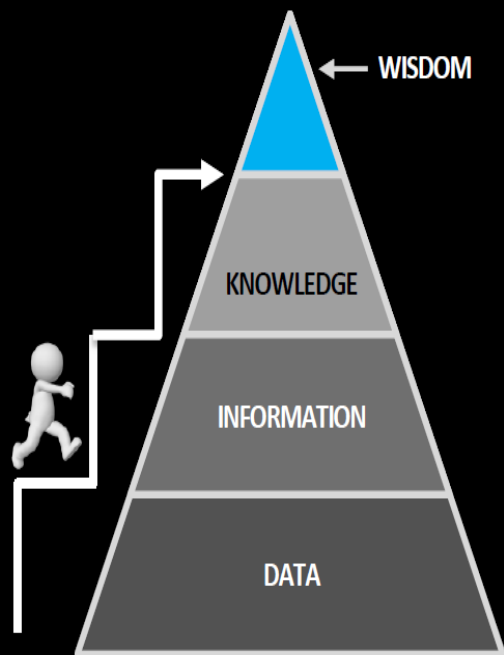
- получение из данных **информации**
 - из информации **знаний**
 - из знаний **мудрости**

Знание – понимание того, ЧТО мы должны делать в определенный момент времени и КАК это нужно делать.

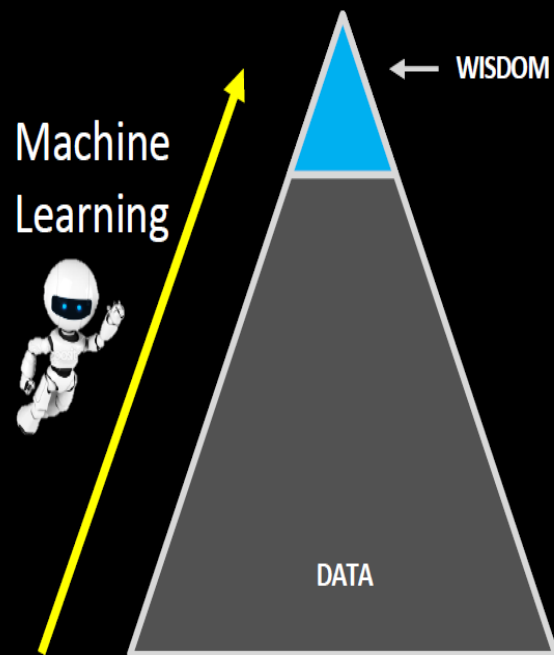
Мудрость – предвидение возможного развития событий.

Данные – это важнейший актив, так как из них можно автоматически получать знание

До Big Data и Deep Learning



После Big Data и Deep Learning



С ростом объёма данных и появлением технологий Big Data, сократился путь от данных к мудрости за счет развития технологий машинного обучения

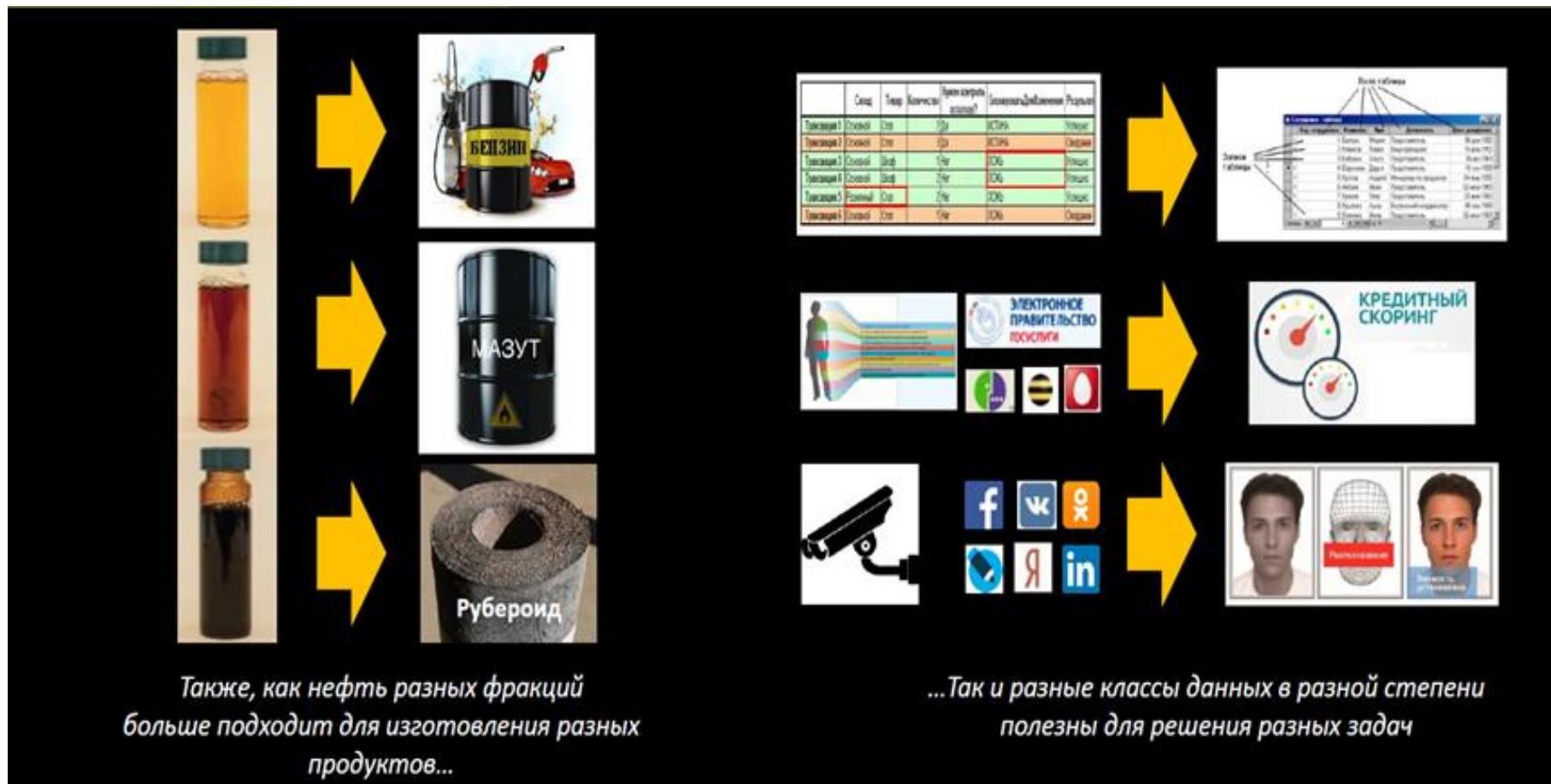
Цель – замена человека машиной и ИИ

- **Механизация** – замена физического труда человека машинным.
- **Автоматизация** – замена управленческого труда человека (машина реализует некоторую программу, алгоритм управления).
- **Кибернетизация** – замена человека в сфере принятия решений.

Данные и базы данных используются в процессах автоматизации и кибернетизации.

Данные – новая нефть

Данные = сырье для создания ценности.
Разные данные – для разных задач.



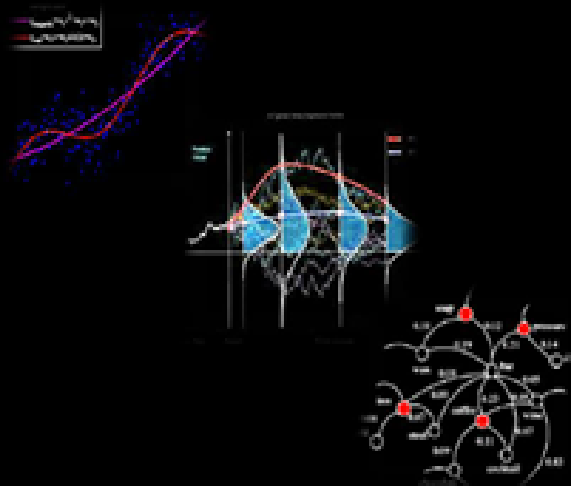
Знания из данных дают модели и аналитика

‘Интеллектуальный анализ данных (Data Mining) включает изучение больших объемов данных и поиска в них закономерностей с применением статистических методов, искусственного интеллекта, а также некоторых технологий управления базами данных’



Машинное обучение

Machine Learning – способы воспроизведения **связей** между событиями и результатом



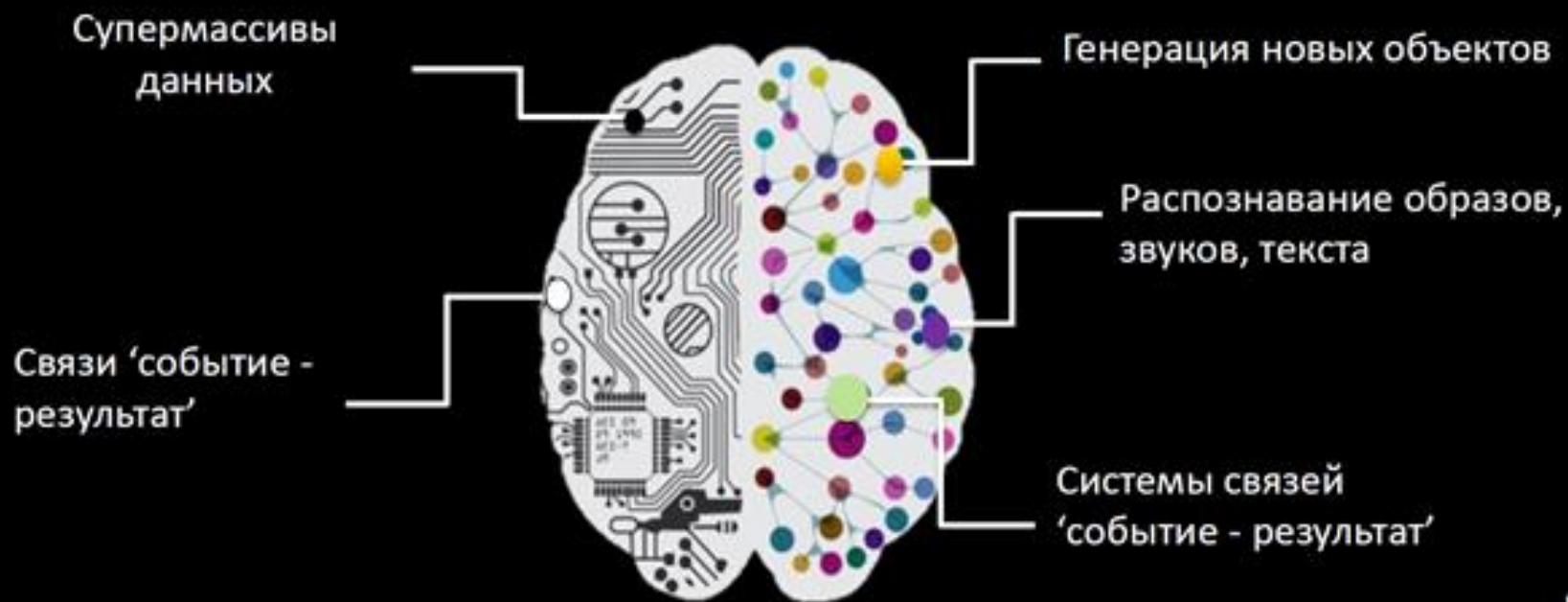
Алгоритмы

+

Технологии

Глубокое обучение

Deep Learning – способы воспроизведения **системы связей** между событиями и результатом, в том числе **скрытых связей**



Задачи технологий работы с данными

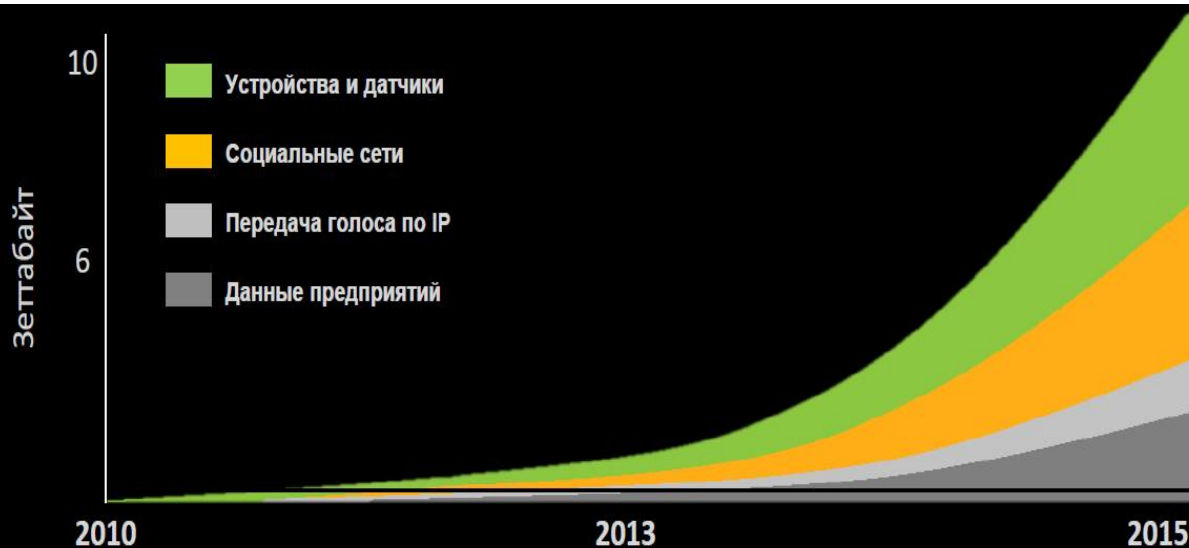


Базы данных и СУБД

Откуда берутся данные?

1

Развитие интернета, социальных сетей, мобильных устройств и M2M стимулировало огромный рост объемов и разнообразия типов данных



2

Каждый человек и каждое его действие создают данные

Финансовые операции



Мобильные устройства



Социальные сети



Государственные учреждения



Аудио и видео



«Умные» вещи



Классификация данных

Тип носителя	Цифровые	Аналоговые	Цифровой или аналоговый носитель данных
Место генерации	Внутренние	Внешние	Данные формируются внутри или вовне организации
Способ генерации	Машина	Человек	Данные создаются машиной или человеком
Объем ¹	Малый	Большой	Гб/Тб. Малые - пригодны к обработке на ПК. Большие требуют решений класса Big Data ¹
Доступность ²	Низкая	Высокая	Показатель доступности
Качество	Низкое	Высокое	Показатель качества
Структурированность	Низкая	Высокая	Простота классификации данных по атрибутам
Однородность	Низкая	Высокая	Число различных форматов данных
Связность	Низкая	Высокая	Возможность сопоставить данные между собой



¹ Малые данные – те, на которых можно анализировать на современном ПК (до 16 Гб) (Том Андерсон, Большие данные – нужны ли они в маркетинговых исследованиях)

² Одни и те же данные могут быть доступными для одного пользователя и не доступными для другого

**Какое значение имеют
базы данных?**

Обработка данных: от сырья к результату



Базы данных – это инструменты, средства достижения конечного результата (месторождения, буровые установки, насосы и нефтеочистительные заводы).

Профессии, требующие знания СУБД

- Разработчик ПО
- Тестировщик (QA инженер)
- Аналитик данных (Data Analyst)
- Инженер машинного обучения (Data Engineer)
- Инженер DevOps
- Project Manager
- Специалист технической поддержки