

## Курс: Аналитика больших данных для руководителей

**Длительность: 24 ак. часа**

### О курсе

Базовый курс по большим данным предоставляет необходимые знания для участия в проектах по анализу больших данных. Включает информацию о фазах жизненного цикла процессов аналитики больших данных при переходе бизнеса к использованию **Big Data**. Материал курса обеспечивает слушателей знанием базовых и расширенных аналитических методов и техник, применяемых для поиска и извлечения знаний из больших массивов разнородных данных. Сравнение различных версий дистрибутивов платформы **Hadoop**, **облачные сервисы для хранения и аналитики данных**, **"open source"** и коммерческие инструменты используемые для хранения, обработки, визуализации и аналитики больших данных.

### Аудитория

Данный курс предназначен для руководителей, топ-менеджеров желающих получить знания по инструментам и методам анализа больших данных.

#### 1. Введение в Big Data(Большие данные)

**Большие данные и цифровая трансформация.** Что такое **Data Lake**. **Обзор технологий основных трендов.** Интернет Вещей **I-o-T**. Методы аналитики больших данных. Отраслевая специфика аналитики больших данных. Жизненный цикл аналитики данных: накопление данных, подготовка данных, планирование модели, построение модели, проверка результатов, внедрение.

#### 2. Data Mining - извлечение знаний из больших данных

Задачи **Data Mining**. Техники **Data Mining**. Классификация и кластеризация. Прогнозирование и визуализация. Ассоциативные правила и обнаружение аномалий. Методология **CRISP-DM**. Инструменты **Data Mining**. Специфика применения **Data Mining** для разных отраслей бизнеса с примерами.

#### 3. Машинное обучение для Data mining

Основные определения. Задачи и область применения машинного обучения. **Supervised/unsupervised** машинное обучение. Инструменты и технологии машинного обучения.

#### 4. Data mining в социальных сетях

Введение в анализ социальных сетей и теорию графов. Базовые алгоритмы на графах и основные возможности графового анализа. Феномен маленького мира. Выделение важных узлов в социальных сетях. Инструменты и методы визуализации графов для применения в бизнесе.

## 5. Инструментарий для работы с Big Data

Специфика работы с **Big Data**. Аналитика для неструктурированных данных с использованием **Hadoop**. Назначение и характеристика компонент экосистемы **Apache Hadoop** для хранения и обработки **Big Data** (**MapReduce, HDFS, YARN, Spark, HBase, Hive**, и т.д.). Сравнительный анализ **Hadoop** дистрибутивов (**Cloudera, Horton Works, MapR**) и инструментария аналитика данных на примерах использования. Сравнительные характеристики программных и аппаратных решений для реализации решений по **Big Data**. Облачные платформы (**AWS, EMR, Azure**) для реализации решений по **Big Data**. Средства визуализации для аналитики данных.

## 6. Интеграция Больших данных

Основные принципы работы с **Big Data**. Импорт и экспорт данных с **Hadoop** (**SQL, NoSQL, HDFS, NFS**, потоковые данные, web content, файлы логов, социальные сети). Пакетная и динамическая загрузка данных.

## 7. Промышленный интернет Вещей (Industrial Internet of Things)

Отличия от Интернета Вещей (**IoT**). Подходы и модели **IIoT**. Применение **IIoT** в различных отраслях. Безопасность **IIoT**. Тенденции и прогнозы. **Industry 4.0**

## 8. Правовые аспекты организации защиты персональных данных

Правовое регулирование в области защиты персональных данных. Международная практика в области защиты персональных данных. Права субъекта и обязанности оператора при обработке персональных данных. Виды нарушений безопасности персональных данных. Стандарт **GDPR**.

## 9. С чего начать?

Формирование команды проекта **Big Data**. Ключевые роли. Отличия подходов **Business Intelligence** и **Data Science**. Специфика рынка данных и аналитики. Тенденции развития технологий **Big Data**.