



## Курс BDAM: Аналитика больших данных для руководителей (Big Data для руководителей и менеджеров)

**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 цифровизация** данных
- ✓ Методы аналитики больших данных
- ✓ Отраслевая специфика аналитики больших данных
- ✓ Сценарии применения технологий больших данных
- ✓ **Жизненный цикл аналитики данных**: получение данных, подготовка данных, планирование модели, построение модели, проверка результатов, внедрение
- ✓ Формирование озера данных **Data Lake**

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

- ✓ Задачи и техники **Data Mining**
- ✓ Классификация и кластеризация
- ✓ Прогнозирование и визуализация
- ✓ Ассоциативные правила и обнаружение аномалий
- ✓ Методология **CRISP-DM, SEMMA**
- ✓ 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 (**Аренадата Hadoop, Cloudera, Horton Works, MapR**) и инструментарий 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 **GDRP**

### 9. С чего начать?

- ✓ Формирование команды проекта **Big Data**. Ключевые роли
- ✓ Специфика рынка данных и аналитики

- ✓ Использование подходов **Agile** и **DevOps**
- ✓ Методологии для стандарты
- ✓ Отличия подходов **Business Intelligence** и **Data Science**
- ✓ Сравнительные характеристики программных и аппаратных решений для реализации решений по **Big Data**.