

**Аннотации модулей программы повышения квалификации  
«Аналитик данных»  
72 часа**

**Содержание модуля 1**

**Тема 1. Задачи систем поддержки принятия решений (СППР).  
Архитектура СППР. Базы данных в СППР.**

Основные задачи, решаемые СППР. Концепция многомерного представления данных. Основные недостатки OLTP-систем при использовании их для анализа данных. Использование концепции многомерного анализа данных для выполнения оперативно-аналитического анализа. Методы классификации моделей хранения данных.

**Тема 2. Неэффективность использования OLTP-систем для анализа данных. Требования к данным.**

Технические сложности работы с большими данными. Распределенная файловая система HDFS. Базовая идея модели MapReduce, пример использования MapReduce. Модель вычислений MapReduce. Реализация MapReduce в Hadoop. Основы Java API. Планирование вычислений. YARN. Примеры реализации алгоритмов на MapReduce (включая умножение матриц, операции реляционной алгебры). Колоночные форматы хранения (на примере ORC).

**Тема 3. Концепция хранилища данных. Свойства хранилищ данных (предметная ориентация, интеграция, поддержка хронологии, неизменяемость).**

Отличия физических и виртуальных хранилищ данных; контент хранилищ данных: метаданные, детальные, агрегированные и архивные данные; отличия хранилищ данных и витрин данных.

**Тема 4. Предварительная обработка данных. Очистка данных. Показатели и документы. Интеграция и преобразование данных. Сокращение данных.**

Разделение данных на два класса: измерения и факты. Технологии переноса данных из оперативных источников в хранилище данных; технологии предварительной очистки данных.

## **Содержание модуля 2**

**Тема 1. Многомерная модель данных. Определение OLAP-систем. Многомерный анализ данных.**

Основные понятия OLAP. OLTP- и OLAP-технологии. Клиентские OLAP-средства. Серверные OLAP-средства. MOLAP, ROLAP, HOLAP. Многомерность в OLAP-приложениях. Тест FAMSИ. Двенадцать законов аналитической обработки в реальном времени.

**Тема 2. Сферы применения Data Mining.**

Введение в Data Mining – понятие, структура, составляющие и сопутствующие науки. Задачи Data Mining и способы их решения. Классификация методов DM. LP. Области применения DM. Классы систем DM. Процесс накопления и анализа данных: Алфавит когнитивного анализа.

**Тема 3. Классификация задач Data Mining.. Предсказательные и описательные модели Data Mining.**

Классификация задач Data Mining. Сферы применения Data Mining. Предсказательные и описательные модели Data Mining. Процесс Data Mining: анализ предметной области, постановка задачи, подготовка данных, построение модели, проверка и оценка моделей, выбор модели, применение модели, коррекция и обновление модели.

## **Содержание модуля 3**

**Тема 1. Классификация. Регрессия. Методы представления результатов анализа. Прогнозирование и визуализация.**

Программное обеспечение для решения задач классификации. Программное обеспечение Data Mining для решения задач оценивания и прогнозирования.

**Тема 2. Методы прогнозирования и классификации.**

Алгоритм построения 1R-алгоритм; метод Naive Bayes; деревья решений, метод опорных векторов, метод «ближайшего соседа», нейронные сети.

### **Тема 3. Поиск ассоциативных правил.**

Характеристики ассоциативных правил: поддержка, достоверность и улучшение. Алгоритм Apriori. Методы кластерного анализа. Меры близости в алгоритмах кластеризации.

И.о. директора филиала  
Смоленского Финуниверситета



А.В. Мосийчук