

Аннотация к рабочей программе дисциплины

1. . .01.01.01 Технологии больших данных

Направление подготовки: 01.03.04 Прикладная математика

Направленности (профили): Математическое и программное обеспечение систем искусственного интеллекта

Квалификация выпускника: бакалавр

Цель освоения дисциплины: изучение студентами технологий и принципов обработки больших данных, практическая работа с современным инструментарием, таким, как фреймворки Hadoop и Hive, модель распределенных вычислений MapReduce.

Объем дисциплины: 4 зачетных единицы, 144 часа

Семестр: 7

Краткое содержание основных разделов дисциплины:

№ п/п раздела	Основные разделы дисциплины	Краткое содержание разделов дисциплины
1	Введение в большие данные. Жизненный цикл аналитики данных	Понятие Big Data. Причины появления, свойства, специфика работы с большими данными. Основные типы данных. Задачи, решаемые с помощью технологий Big Data. Аналитика данных. Сбор, структурирование и анализ на основе Business Intelligence. Инструментарий BI. Оперативная аналитическая обработка (OLAP-технология).
2	Технология Hadoop. СУБД Hive. Вычисления в MapReduce	История и архитектура Hadoop. Файловая система HDFS. Движки MapReduce, Spark, Tez. Реляционные СУБД Hive, Shark, Spark SQL. Нереляционные БД: HBase. Вычисления в MapReduce. Модель распределенных вычислений MapReduce. Реализация для параллельной обработки больших массивов данных.
3	Масштабирование и многоуровневое хранение данных	Концепция СУБД NoSQL. Сравнение с СУБД SQL-типа. Виды, особенности, область применения. Масштабируемость. Репликация. CAP теорема. СУБД MongoDB.
4	Визуализация данных и результатов анализа	Типы, задачи, виды визуализации. Интерактивный сторителлинг, дашборды. Язык R, характеристики и возможности. Дедупликация данных.

Форма промежуточной аттестации: экзамен (7 семестр)