

Б1.В.ДВ.07.01 Технологии организации и управления инфраструктурой больших данных

Цели дисциплины

Целями освоения дисциплины Технологии организации и управления инфраструктурой больших данных является формирование профессиональных компетенций будущих специалистов в области Информационных систем и технологий. Опираясь на знания, полученные при изучении курсов высшей математики, программирования и численных методов рассмотреть основы решения прикладных задач математики методами организации и анализа больших данных.

Задачи дисциплины

Основными задачами изучения дисциплины являются:

- Формирование знаний в области управления и анализа больших данных методов решения экспериментальных и теоретических задач;
- Формирование практических навыков применения математического аппарата и систем управления инфраструктурой больших данных;
- Формирование навыков владения интерпретирования, представления и применения результатов, полученных с использованием анализа больших данных и управления инфраструктурой больших данных для своей предметной области.

Формируемые компетенции и индикаторы их достижения по дисциплине

Код компетенции	Содержание компетенции	Код и наименование индикатора достижения компетенции
ПКС-2.	Способен проводить формализацию предметной области с целью создания информационной системы	ПКС-2.1 - Знает требования к компьютерному программному обеспечению; виды технической спецификации на программные компоненты и их взаимодействие; методы проектирование компьютерного программного обеспечения ПКС-2.2 – Умеет применять требования к компьютерному программному обеспечению; разрабатывать технические спецификации на программные компоненты и их взаимодействие; применять методы проектирования компьютерного программного обеспечения; ПКС-2.3 – Владеет методами разработки требований к компьютерному программному обеспечению, технических спецификаций на программные компоненты, методами проектирования компьютерного программного обеспечения.

Содержание разделов дисциплины

Тема 1 Анализ систем обработки больших данных

Объем генерируемых данных. Многообразие больших данных. Скорость генерирования или обработки данных. Изменчивость, нестабильность данных во времени. Достоверность собранных данных. Трудоемкость процесса корреляции и построения взаимосвязей между данными.

Тема 2 Современные методы визуализации информации Big Data

Сложная столбчатая диаграмма, хордовая диаграмма, солнечная диаграмма, потоковый график матричная диаграмма, диаграмма рассеяния, диаграмма ohlc

Тема 3 Задачи параллельной обработки данных в научных исследованиях

Проблемные точки процесса реализации программы, параллельные структуры алгоритмов, параллелизм на уровне инструкций, параллелизм на уровне задач, параллелизм на уровне алгоритмов

Тема 4 Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы

Технологии кластеризации и классификации. K-means. ЕМ-алгоритм. Другие методы кластеризации. Задачи классификации. Байесовский классификатор. Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия. Нейронные сети: общая архитектура. Многослойные сети. Обратное распространение ошибки. Стохастический градиентный спуск. Генетические алгоритмы.

Тема 5. Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети

Извлечение признаков / выделение особенностей (Feature detection). Преобразования признаков. Нормализация данных. Методы нормализации данных. Нормализация по методу минимакса. Нормализация по Z-показателю. Десятичное масштабирование. Нечеткие множества. Байесовы сети. Задачи байесовского вывода. Методика построения нечеткой байесовой сети