Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 15.08.2024 17:26:39
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8bbe6ede0e1ab78

АННОТАЦИЯ
рабочей программы дисциплины

Технологии анализа больших данных

1. Общая трудоёмкость

Трудоёмкость дисциплины составляет 5 зачётных единиц (180 часов), из них 18 часов лекционных занятий, 36 часов практических занятий.

2. Место дисциплины в структуре образовательной программы

Дисциплина относится к модулю обязательных профессиональных дисциплин обязательной части образовательной программы.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими элементами образовательной программы: Математические методы анализа больших данных; Экспертные системы и базы знаний; Методы машинного обучения.

Результаты обучения, формируемые данной дисциплиной, потребуются при освоении следующих элементов образовательной программы: производственная практика, проектно-технологическая практика; производственная практика, преддипломная практика.

3. Цель изучения дисциплины

формирование у обучающихся способности осуществлять обработку больших объемов данных для решения профессиональных задач, эффективно применять методы, технологии и инструментальные средства анализа больших данных в профессиональной деятельности.

4. Содержание дисциплины

Модуль 1 «Основы построения и использования систем больших данных»

Тема 1.1 «Основы систем больших данных»

Понятие Больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Использование больших данных в науке, бизнесе, государственном управлении.

Тема 1.2 «Методы работы с распределенными информационными системами. Управление качеством данных»

Использование фреймворка MapReduce в распределенной среде. Реализации MapReduce. Состав и возможности программного комплекса Apache Hadoop. Языки поисковых запросов для Hadoop.Управление качеством данных с использованием ETL-конвейеров ApacheAirflow.

Модуль 2 «Разработка и использование приложений на основе распределенных баз данных»

Тема 2.1 «СУБД NoSQL»

Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность. CAP-теорема. Классы NoSQL баз данных. Примеры СУБД NoSQL. Семейства столбцов. Графовые СУБД.Neo4j

Тема 2.2 «Документно-ориентированные распределенные СУБД»

Понятие агрегата. Современные документо-ориентированные СУБД. Запросы к СУБД на языке JSON. MongoDB. Использование фреймворка Map-Reduce в документо-ориентированных СУБД.

5. Дополнительная полезная информация

Дисциплина предназначена для формирования элементов следующих компетенций образовательной программы:

ПК-4. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика.

ПК-6. Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации.

Форма промежуточной аттестации: дифференцированный зачёт.

Наименование оценочного средства: практическая работа №1-3; тест №1, тест №2.