|  |
| --- |
|  Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»  |
|  |  УТВЕРЖДАЮ Директор Таганрогского института имени А.П. Чехова (филиала) РГЭУ (РИНХ) \_\_\_\_\_\_\_\_\_\_\_\_\_ Голобородько А.Ю. «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_\_г. |
|  |
|  |  |
|  **Рабочая программа дисциплины** **Практикум по анализу Big Data** |
|  |  |
|  направление 09.04.03 Прикладная информатика направленность (профиль) 09.04.03.02 Информационные системы и анализ больших данных |
|  |  |
|  Для набора \_\_\_\_\_\_\_\_\_\_ года |
|  |  |
|  Квалификация Магистр |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  УП: 09.04.03.02-22-1-ПИGZ.plx |  |  |  |  стр. 2 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  КАФЕДРА |  |  **информатики** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  **Распределение часов дисциплины по курсам** |  |  |  |  |
|  |  Курс |  **2** |  Итого |  |  |  |  |
|  |  Вид занятий |  УП |  РП |  |  |  |  |
|  |  Лекции |  4 |  4 |  4 |  4 |  |  |  |  |
|  |  Лабораторные |  6 |  6 |  6 |  6 |  |  |  |  |
|  |  Итого ауд. |  10 |  10 |  10 |  10 |  |  |  |  |
|  |  Кoнтактная рабoта |  10 |  10 |  10 |  10 |  |  |  |  |
|  |  Сам. работа |  94 |  94 |  94 |  94 |  |  |  |  |
|  |  Часы на контроль |  4 |  4 |  4 |  4 |  |  |  |  |
|  |  Итого |  108 |  108 |  108 |  108 |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  **ОСНОВАНИЕ** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  Учебный план утвержден учёным советом вуза от 26.04.2022 протокол № 9/1.   Программу составил(и): канд. техн. наук, Зав. каф., Тюшнякова И. А. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Зав. кафедрой: Тюшнякова И. А. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  УП: 09.04.03.02-22-1-ПИGZ.plx |  |  |  |  |  |  стр. 3 |
|  |  |  |  |  |  |  |  |  |
|  **1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ** |
|  1.1 |  сформировать у студентов устойчивый комплекс навыков практической работы с большими данными для решения аналитических задач |
|  |  |  |  |  |  |  |  |  |
|  **2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ** |
|  **ПКР-4.1:Знать современные технологии работы с Big Data, методы решения задач обработки и анализа больших данных, принципы обработки больших данных в распределенных вычислительных системах** |
|  **ПКР-4.2:Уметь использовать и применять углубленные знания в области обработки и анализа больших данных** |
|  **ПКР-4.3:Владеть навыками применения программных систем, предназначенных для анализа больших данных** |
|  **ОПК-5.1:Знать современное программное и аппаратное обеспечение информационных и автоматизированных систем** |
|  **ОПК-5.2:Уметь модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем для решения профессиональных задач;** |
|  |  |  |  |  |  |  |  |  |
|  **В результате освоения дисциплины обучающийся должен:** |
|  **Знать:** |
|  базовые понятия технологии Big Data; характеристики рынка систем анализа неструктурированной информации и перспективы развития сегмента информационно-технологической отрасли «Большие данные» (Big Data), основные методы анализа, применяемые в «Больших данных», а также основные классы и принципы построения информационных систем, применяемых для практической реализации этих методов |
|  **Уметь:** |
|  определять массивы больших данных; анализировать кластеры больших данных; строить различными способами прогнозы развития различных процессов. |
|  **Владеть:** |
|  современными технологиями создания и обслуживания больших данных; методологией и методикой прогнозирования. |
|  |  |  |  |  |  |  |  |  |
|  **3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ** |
|  **Код занятия** |  **Наименование разделов и тем /вид занятия/** |  **Семестр / Курс** |  **Часов** |  **Компетен-** **ции** |  **Литература** |
|  |  **Раздел 1. Практикум по анализу Big Data** |  |  |  |  |
|  1.1 |  Введение в Большие Данные (Big Data). Распределенные файловые системы, Workshop. Hadoop экосистема, MapReduce. /Лек/ |  2 |  2 |  ПКР-4.1 ПКР-4.3 |  Л1.1 Л1.3 Л1.4Л2.1 Л2.4 |
|  1.2 |  Распределенные файловые системы (GFS, HDFS). Их составляющие, достоинства, недостатки и сфера применения.Чтение и запись в HDFS. HDFS APIs: Web, shell. Hadoop Streaming. Элементы Hadoop-задачи (Mapper, reducer, combiner, partitioner, comparator). /Лаб/ |  2 |  2 |  ПКР-4.1 ПКР-4.3 |  Л1.1 Л1.3 Л1.4Л2.1 Л2.4 |
|  1.3 |  Оптимизация MapReduce вычислений. Spark /Лек/ |  2 |  2 |  ПКР-4.1 ПКР-4.3 |  Л1.1 Л1.3 Л1.4Л2.1 Л2.4 |
|  1.4 |  Приложения с несколькими Hadoop-задачами. Тюнинг Hadoop- job (настройка партиционирования, сложные ключи, uber jobs). Задачи с несколькими входами. Joins в Hadoop. /Лаб/ |  2 |  2 |  |  Л1.3 |
|  1.5 |  Spark: основные термины и RDD, Spark DataFrames , оптимизации Spark вычислений /Лаб/ |  2 |  2 |  |  Л1.2Л2.2 Л2.3 |
|  1.6 |  Работа с потоковой обработкой данных /Ср/ |  2 |  28 |  |  Л1.2Л2.2 Л2.3 |
|  1.7 |  Подготовка к лабораторным работам /Ср/ |  2 |  26 |  ПКР-4.1 ПКР-4.2 ПКР-4.3 |  Л1.1 Л1.4Л2.1 Л2.3 Л2.4 |
|  1.8 |  "SQL поверх больших данных (Hive)". Архитектура Hive, виды таблиц, форматы хранения данных.Трансляция Hive-запросов в MapReduce- задачи.Сериализация и десериализация. Тюнинг Join'ов в Hive. Партиционирование, бакетирование, семплирование. User defined functions, Hive Streaming. /Ср/ |  2 |  40 |  ПКР-4.1 ПКР-4.3 |  Л1.1 Л1.4Л2.1 Л2.3 Л2.4 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  УП: 09.04.03.02-22-1-ПИGZ.plx |  |  |  |  |  |  |  |  стр. 4 |
|  1.9 |  /Зачёт/ |  2 |  4 |  ПКР-4.1 ПКР-4.3 |  Л1.1 Л1.2 Л1.3 Л1.4Л2.1 Л2.2 Л2.4 |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  **4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ** |
|  Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины. |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  **5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ** |
|  **5.1. Основная литература** |
|  |  Авторы, составители |  Заглавие |  Издательство, год |  Колич-во |
|  Л1.1 |  Жуковский О. И. |  Информационные технологии и анализ данных: учебное пособие |  Томск: Эль Контент, 2014 |  http://biblioclub.ru/index. php? page=book&id=480500 неограниченный доступ для зарегистрированных пользователей |
|  Л1.2 |  Бутаков Н. А., Петров М. В., Насонов Д. |  Обработка больших данных с Apache Spark: учебно-методическое пособие |  Санкт-Петербург: Университет ИТМО, 2019 |  http://biblioclub.ru/index. php? page=book&id=566771 неограниченный доступ для зарегистрированных пользователей |
|  Л1.3 |  Уайт Т. |  Hadoop: Подробное руководство |  Санкт-Петербург: Питер, 2013 |  https://ibooks.ru/reading. php? short=1&productid=3543 74 неограниченный доступ для зарегистрированных пользователей |
|  Л1.4 |  Шнарева, Г. В., Пономарева, Ж. Г. |  Анализ данных: учебно-методическое пособие |  Симферополь: Университет экономики и управления, 2019 |  http://www.iprbookshop. ru/89482.html неограниченный доступ для зарегистрированных пользователей |
|  **5.2. Дополнительная литература** |
|  |  Авторы, составители |  Заглавие |  Издательство, год |  Колич-во |
|  Л2.1 |  Каган Е. С. |  Прикладной статистический анализ данных: учебное пособие |  Кемерово: Кемеровский государственный университет, 2018 |  http://biblioclub.ru/index. php? page=book&id=573550 неограниченный доступ для зарегистрированных пользователей |
|  Л2.2 |  Риза С., Лезерсон У., Оуэн Ш., Уиллс Д. |  Spark для профессионалов: современные паттерны обработки больших данных |  Санкт-Петербург: Питер, 2017 |  https://ibooks.ru/reading. php? short=1&productid=3543 85 неограниченный доступ для зарегистрированных пользователей |
|  Л2.3 |  Карау Х., Уоррен Р. |  Эффективный Spark. Масштабирование и оптимизация |  Санкт-Петербург: Питер, 2018 |  https://ibooks.ru/reading. php? short=1&productid=3581 42 неограниченный доступ для зарегистрированных пользователей |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  УП: 09.04.03.02-22-1-ПИGZ.plx |  |  |  |  стр. 5 |
|  |  Авторы, составители |  Заглавие |  Издательство, год |  Колич-во |
|  Л2.4 |  Пальмов, С. В. |  Интеллектуальный анализ данных: учебное пособие |  Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017 |  http://www.iprbookshop. ru/75376.html неограниченный доступ для зарегистрированных пользователей |
|  **5.3 Профессиональные базы данных и информационные справочные системы** |
|  |
|  Научная электронная библиотека https://www.elibrary.ru/defaultx.asp |
|  **5.4. Перечень программного обеспечения** |
|  Microsoft Office |
|  **5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья** |
|  При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме. |
|  |  |  |  |  |  |  |
|  **6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)** |
|  Помещения для проведения всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения. Для проведения лекционных занятий используется демонстрационное оборудование. Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными программными средствами и выходом в Интернет. |
|  |  |  |  |  |  |  |
|  **7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)** |
|  Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины. |