ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ

КАФЕДРА КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕХНОЛОГИЙ

ОДОБРЕНО УМС ИИКС  
  
Протокол № УМС-575/08-1   
  
от 28.08.2020 г.

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

НАУКА О ДАННЫХ / DATA SCIENCE

|  |  |
| --- | --- |
| Направление подготовки (специальность) | 09.03.01 Информатика и вычислительная техника |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Семестр** | **Трудоемкость, кред.** | **Общий объем курса, час.** | **Лекции, час.** | **Практич. занятия, час.** | **Лаборат. работы, час.** | **СРС, час.** | **КСР, час.** | **Форма(ы) контроля, экз./зач./КР/КП** |
| 7 | 3 | 108 | 16 | 16 | 16 | 24 | 0 | Э |
| Итого | 3 | 108 | 16 | 16 | 16 | 24 | 0 |  |

АННОТАЦИЯ

В настоящее время процесс генерации новых данных приобрел «лавинный» характер. В результате всеобщей информатизации и активного совершенствования современных вычислительных мощностей темпы роста объема хранимых данных можно охарактеризовать как крайне высокие. Помимо, собственно, хранения данные нуждаются в обработке. При этом анализ данных в общем случае не является разовым. В следствии развития методов DataMining и концепции BigData, старые данные обычно сохраняются в первозданном виде для последующего анализа с учетом новых тенденций и подходов, которые отсутствовали на момент первоначального анализа.

В качестве примера таких данных можно привести: экспериментальные данные, статистику обращений к веб-сервисам, метеоданные, поток информации из социальных сетей, микроблогов и др. В каждом из представленных случаев возможно осуществить параллельную обработку данных на распределенной вычислительной системе. При этом, необходимы системы хранения, способны обрабатывать миллионы довольно простых по форме запросов.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целями освоения учебной дисциплины «Data Science and Big Data Analytic» являются формирование у студентов целостного представления о принципах разработки, анализа и реализации параллельных алгоритмов обработки структур данных; освоение студентами технологий разработки программных продуктов для суперкомпьютерных систем обработки и хранения больших объемов данных данных.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Учебная дисциплина входит в профессиональный цикл, раздел M2. Для освоения дисциплины необходимы компетенции, сформированные у обучающихся в результате освоения дисциплин общенаучного цикла М1 (в частности, дисциплины разделов М1.Б.2 «Методы оптимизации», М1.В.1 «Архитектура информационных систем» и М1.В.2. «Параллельные вычисления»).

Сформированные при изучении данной дисциплины компетенции необходимы для выполнения научно-исследовательской работы и подготовки магистерских диссертаций по проблематике организации высокопроизводительных систем анализа данных.

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Универсальные и(или) общепрофессиональные компетенции:

|  |  |
| --- | --- |
| Код и наименование компетенции | Код и наименование индикатора достижения компетенции |
| ОПК-1 – Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности | З-ОПК-1 – Знать: основы математики, физики, вычислительной техники и программирования У-ОПК-1 – Уметь: решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования В-ОПК-1 – Владеть: навыками теоретического и экспериментального исследования объектов профессиональной деятельности |
| ОПК-2 – Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности | З-ОПК-2 – Знать: современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности У-ОПК-2 – Уметь: выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности В-ОПК-2 – Владеть: навыками применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности |
| ОПК-8 – Способен разрабатывать алгоритмы и программы, пригодные для практического применения | З-ОПК-8 – Знать: алгоритмические языки программирования, операционные системы и оболочки, современные среды разработки программного обеспечения У-ОПК-8 – Уметь: составлять алгоритмы, писать и отлаживать коды на языке программирования, тестировать работоспособность программы, интегрировать программные модули В-ОПК-8 – Владеть: языком программирования; навыками отладки и тестирования работоспособности программы |
| ОПК-9 – Способен осваивать методики использования программных средств для решения практических задач | З-ОПК-9 – Знать: классификацию программных средств и возможности их применения для решения практических задач У-ОПК-9 – Уметь: находить и анализировать техническую документацию по использованию программного средства, выбирать и использовать необходимые функции программных средств для решения конкретной задачи В-ОПК-9 – Владеть: способами описания методики использования программного средства для решения конкретной задачи в виде документа, презентации или видеоролика |
| УКЕ-1 – Способен использовать знания естественнонаучных дисциплин, применять методы математического анализа и моделирования, теоретического и экспериментального исследования в поставленных задачах | З-УКЕ-1 – знать: основные законы естественнонаучных дисциплин, методы математического анализа и моделирования, теоретического и экспериментального исследования У-УКЕ-1 – уметь: использовать математические методы в технических приложениях, рассчитывать основные числовые характеристики случайных величин, решать основные задачи математической статистики; решать типовые расчетные задачи В-УКЕ-1 – владеть: методами математического анализа и моделирования; методами решения задач анализа и расчета характеристик физических систем, основными приемами обработки экспериментальных данных, методами работы с прикладными программными продуктами |
| УКЦ-1 – Способен в цифровой среде использовать различные цифровые средства, позволяющие во взаимодействии с другими людьми достигать поставленных целей | З-УКЦ-1 – Знать: современные информационные технологии и цифровые средства коммуникации, в том числе отечественного производства, а также основные приемы и нормы социального взаимодействия и технологии межличностной и групповой коммуникации с использованием дистанционных технологий У-УКЦ-1 – Уметь: выбирать современные информационные технологии и цифровые средства коммуникации, в том числе отечественного производства, а также устанавливать и поддерживать контакты, обеспечивающие успешную работу в коллективе и применять основные методы и нормы социального взаимодействия для реализации своей роли и взаимодействия внутри команды с использованием дистанционных технологий В-УКЦ-1 – Владеть: навыками применения современных информационных технологий и цифровых средств коммуникации, в том числе отечественного производства, а также методами и приемами социального взаимодействия и работы в команде с использованием дистанционных технологий |

Профессиональные компетенции в соотвествии с задачами и объектами (областями знаний) профессиональной деятельности:

|  |  |  |  |
| --- | --- | --- | --- |
| **Задача профессиональной деятельности (ЗПД)** | **Объект или область знания** | **Код и наименование профессиональной компетенции;** **Основание (профессиональный стандарт-ПС, анализ опыта)** | **Код и наименование индикатора достижения профессиональной компетенции** |
| научно-исследовательский и инновационный |  |  |  |
| Изучение научно-технической информации, отечественного и зарубежного опыта по тематике исследования. Математическое моделирование процессов и объектов на базе стандартных пакетов автоматизированного проектирования и исследований. Проведение экспериментов по заданной методике и анализ результатов. Проведение измерений и наблюдений, составление описания проводимых исследований, подготовка данных для составления обзоров, отчетов и научных публикаций. Составление отчета по выполненному заданию, участие во внедрении результатов исследований и разработок. ? Участие в составе коллектива исполнителей во внедрении результатов научно-технических исследований в высокотехнологичных сферах экономики и коммерциализации разработок. | Вычислительные машины, комплексы, системы и сети; автоматизированные системы обработки информации и управления; системы автоматизированного проектирования и информационной поддержки жизненного цикла промышленных изделий; программное обеспечение средств вычислительной техники и автоматизированных систем (программы, программные комплексы и системы); математическое, информационное, техническое, лингвистическое, программное, эргономическое, организационное и правовое обеспечение перечисленных систем. | ПК-1 - Способен обосновывать принимаемые проектные решения, осуществлять постановку и выполнять эксперименты по проверке их корректности и эффективности  *Основание:* Профессиональный стандарт: 06.001 | З-ПК-1 - Знать: основы верификации и аттестации аппаратного и программного обеспечения, стандарты качества и процессов его обеспечения, способы оптимизации, принципы и виды отладки, методы оценки качества, методики постановки экспериментов; У-ПК-1 - Уметь: разрабатывать и специфицировать требования, осуществлять составление описания проводимых исследований, подготовку данных для составления обзоров и отчетов, обосновывать принимаемые проектные решения, выполнять эксперименты по проверке корректности решений; В-ПК-1 - Владеть: навыками построения моделей объектов профессиональной деятельности с использованием инструментальных средств, навыками тестирования, отладки и верификации |

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Разделы учебной дисциплины, их объем, сроки изучения и формы контроля:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **№ п.п** | **Наименование раздела учебной дисциплины** | **Недели** | **Лекции/ Практ. (семинары )/ Лабораторные работы, час.** | **Обязат. текущий контроль (форма\*, неделя)** | **Максимальный балл за раздел\*\*** | **Аттестация раздела (форма\*, неделя)** | **Индикаторы освоения компетенции** |
|  | *7 Семестр* |  |  |  |  |  |  |
| 1 | Первый раздел | 1-8 | 8/8/8 |  | КИ-8 | 25 |  |
| 2 | Второй раздел | 9-16 | 8/8/8 |  | КИ-16 | 25 |  |
|  | *Итого за 7 Семестр* |  | 16/16/16 |  |  | 50 |  |
|  | **Контрольные мероприятия за 7 Семестр** |  |  |  | Э | 50 |  |

\* – сокращенное наименование формы контроля

\*\* – сумма максимальных баллов должна быть равна 100 за семестр, включая зачет и (или) экзамен

Сокращение наименований форм текущего контроля и аттестации разделов:

|  |  |
| --- | --- |
| **Обозначение** | **Полное наименование** |
| КИ | Контроль по итогам |
| Э | Экзамен |

КАЛЕНДАРНЫЙ ПЛАН

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Недели** | **Темы занятий / Содержание** | **Лек., час.** | **Пр./сем., час.** | **Лаб., час.** |
|  | *7 Семестр* | 16 | 16 | 16 |
| **1-8** | **Первый раздел** | 8 | 8 | 8 |
| 1 | **Вводное занятие** Общие вопросы организации BigData-систем. Проблемы организации BigData-систем. Классификация характера задач для выбора подходящей реализации. Методы декомпозиции BigData-задач. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 1 | 1 | 1 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 2 | **Архитектура BigData систем. Часть 1** Метрики производительности BigData-систем, состав и назначение подсистем внутренних/внешних обменов, подсистем online/offline аналитики. Особенности использования внешних и внутренних каналов связи при передиче больших объемов данных. Основные метриками для оценки производительности BigData-систем. Способы организации внешних сетевых интерфейсов BigData-систем и построения модулей для online/offline обработки. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 1 | 1 | 1 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 3 | **Архитектура BigData систем. Часть 2** Способы интеграции подсистем online/offline обработки, подсистем индексирования и хранения данных. Введение в проблематику построения систем мониторинга вычислительных ресурсов. Состав и назначение (на примере системы мониторинга вычислительных ресурсов) модулей online/offline обработки, подсистем индексирования и хранения данных.  Стандартные средства разработки для организации процессов поточной обработки больших объемов данных.  Принципы построения интеграционного кластерного интерфейса для организации взаимодействия распределенных систем. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 1 | 1 | 1 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 4 - 5 | **Инфраструктура BigData систем** Системы построения виртуализированных сетевых инфраструктур. Системы легковесной контейнеризации. Вопросы виртуализации сетевых функций и построения программно определяемых сетей.  Назначение технологий SDN/NFV. Механизмы работы систем управления виртуализированными контейнерами.  Технология контейнеризации OpenShift для развертывания программных решений в облаке.  Принципы использования технологий OpenStack и OpenShift для организации сетей виртуальных машин и систем управления контейнерами. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 | 2 | 2 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 6 - 7 | **Пакетная распределенная обработка больших объемов данных** Технология организации пакетной обработки больших объемов данных MapReduce.  Фазы и действия, выполняемые программным каркасом Hadoop при исполнении MapReduce-программы.  Разработка Java-приложения, использующие Hadoop для выполнения пакетной обработки данных по таймеру.  Управления жизненным циклом Hadoop-кластера. Способами развертывания MapReduce-программ в Hadoop-окружении. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 | 2 | 2 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 8 | **Оценка производительности BigData-систем** Элементы теории массового обслуживания для оценки производительности распределенных вычислительных BigData-систем  Основы теории массового обслуживания для расчет интенсивности поступления запросов на каждый узел сети.  Оценка необходимой производительности узлов сети исходя из прогнозируемой нагрузки.  Математический аппарат теории вероятностей и теории массового обслуживания для построения моделей потоков данных в BigData-системах. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 1 | 1 | 1 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| **9-16** | **Второй раздел** | 8 | 8 | 8 |
| 9 - 10 | **Асинхронная обработка больших объемов данных** Принципы работы с технологией Apache Spark для выполнения асинхронных вычислительных операций и системы очередей для управления асинхронными процессами в BigData-системах. Состав кластера Apache Spark и Apache Kafka. Организация загрузки/выгрузки информации в системах Apache Spark и Apache Kafka. Администрирование систем Apache Spark и Apache Kafka, способами организации асинхронного взаимодействия нескольких вычислительных задач. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 | 2 | 2 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 11 - 12 | **Хранение больших объемов данных** Высокопроизводительные NoSQL-системы. Предпосылки, типы и характеристика. Состав и характеристики высокопроизводительных файловых систем на примере GFS, HDFS и NFS v4.1. Состав кластера Apache Cassandra. Средства обеспечения согласованности в высокопроизводительных системах хранения данных.  Определение необходимого типа системы хранения и схему упаковки данных в зависимости от задачи.  Принципы трансформации потока входящей информации в поток объектов хранения BigData-системы. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 | 2 | 2 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 13 - 14 | **Построение индексов** Модули индексации данных для BigData-систем. Назначение модулей индексации данных и требования к ним.  Проектирование отказоустойчивых высокопроизводительных модулей индексации данных, предназначенные для поддержки конкретных алгоритмов BigData-аналитики.  Принципы построения и оценки производительности подсистем индексации данных. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 | 2 | 2 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |
| 15 - 16 | **Библиотеки машинного обучения для BigData-систем** Алгоритмы машинного обучения, предназначенные для пакетной и поточной обработки.  Виды алгоритмов машинного обучения, пригодные для использования в offline-модулях, Виды алгоритмов машинного обучения, пригодные для использования в onfline-модулях. Применение гибридных вычислительных технологий в задачах машинного обучения. Способы повышения производительности алгоритмов машинного обучения с применением гибридных вычислительных технологий. | Всего аудиторных часов | Всего аудиторных часов | Всего аудиторных часов |
| 2 | 2 | 2 |
| Онлайн | Онлайн | Онлайн |
|  |  |  |

Сокращенные наименования онлайн опций:

|  |  |
| --- | --- |
| **Обозначение** | **Полное наименование** |
| ЭК | Электронный курс |
| ПМ | Полнотекстовый материал |
| ПЛ | Полнотекстовые лекции |
| ВМ | Видео-материалы |
| АМ | Аудио-материалы |
| Прз | Презентации |
| Т | Тесты |
| ЭСМ | Электронные справочные материалы |
| ИС | Интерактивный сайт |

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Проведение лекций предусматривает использование технических средств обучения (ТСО) для показа презентаций, иллюстрации процесса разработки, отладки и профилирования программ.

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущего, рубежного и промежуточного контроля по дисциплине.

Связь между формируемыми компетенциями и формами контроля их освоения представлена в следующей таблице:

|  |  |
| --- | --- |
| **Компетенция** | **Индикаторы освоения** |
| ОПК-1 | З-ОПК-1 |
| ОПК-1 | У-ОПК-1 |
| ОПК-1 | В-ОПК-1 |
| ОПК-2 | З-ОПК-2 |
| ОПК-2 | У-ОПК-2 |
| ОПК-2 | В-ОПК-2 |
| ОПК-8 | З-ОПК-8 |
| ОПК-8 | У-ОПК-8 |
| ОПК-8 | В-ОПК-8 |
| ОПК-9 | З-ОПК-9 |
| ОПК-9 | У-ОПК-9 |
| ОПК-9 | В-ОПК-9 |
| ПК-1 | З-ПК-1 |
| ПК-1 | У-ПК-1 |
| ПК-1 | В-ПК-1 |
| УКЕ-1 | З-УКЕ-1 |
| УКЕ-1 | У-УКЕ-1 |
| УКЕ-1 | В-УКЕ-1 |
| УКЦ-1 | З-УКЦ-1 |
| УКЦ-1 | У-УКЦ-1 |
| УКЦ-1 | В-УКЦ-1 |

Оценочные средства приведены в Приложении.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ A22 Advances in Data Mining. Applications and Theoretical Aspects : 16th Industrial Conference, ICDM 2016, New York, NY, USA, July 13-17, 2016. Proceedings, Cham: Springer International Publishing, 2016

2. ЭИ A22 Advances in Knowledge Discovery and Data Mining : 20th Pacific-Asia Conference, PAKDD 2016, Auckland, New Zealand, April 19-22, 2016, Proceedings, Part I, Cham: Springer International Publishing, 2016

3. ЭИ Б 82 Основы работы с технологией CUDA : , Москва: ДМК Пресс, 2010

4. 004 П18 Параллельные вычисления на GPU. Архитектура и программная модель CUDA : учебное пособие, Москва: Издательство Московского университета, 2012

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. 510 А45 Алгоритмы : построение и анализ, Москва [и др.]: Вильямс, 2011

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

Специальное программное обеспечение не требуется

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

https://online.mephi.ru/

http://library.mephi.ru/

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

9. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ СТУДЕНТОВ

Приведёны в приложении

Автор(ы):

|  |  |  |
| --- | --- | --- |
|  | Дюмин Александр Александрович |  |

Рецензент(ы):

|  |  |  |
| --- | --- | --- |
|  | Ровнягин М.М. |  |