

ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ
КАФЕДРА КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕХНОЛОГИЙ

ОДОБРЕНО УМС ИИКС

Протокол № УМС-575/01-1

от 30.08.2021 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

НАУКА О ДАННЫХ И АНАЛИЗ БОЛЬШИХ ДАННЫХ (DATA SCIENCE AND BIG DATA ANALYTICS)

Направление подготовки
(специальность)

[1] 09.04.04 Программная инженерия

Семестр	Трудоемкость, кред.	Общий объем курса, час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	В форме практической подготовки/ В СРС, час.	КСР, час.	Форма(ы) контроля, экз./зач./КР/КП	
3	5-7	180-252	32	0	16		96-168	0	Э
Итого	5-7	180-252	32	0	16	0	96-168	0	

АННОТАЦИЯ

В настоящее время процесс генерации новых данных приобрел «лавиный» 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 первоизданном 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 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

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
--------------------------------	--

Профессиональные компетенции в соответствии 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
--	---------------------------	--	---

производственно-технологический			
разработка, тестирование и сопровождение программного обеспечения, применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения, взаимодействие с заказчиком в процессе выполнения	инструменты разработки программного продукта, процессы жизненного цикла программного продукта	ПК-10.2 [1] - Способен использовать технологии, методы и инструментальные средства обработки больших данных <i>Основание:</i> Профессиональный стандарт: 06.042	З-ПК-10.2[1] - Знать технологии, методы и инструментальные средства обработки больших данных; У-ПК-10.2[1] - Уметь пользоваться методами и инструментами получения, хранения, передачи, обработки больших данных; В-ПК-10.2[1] - Владеть методами и инструментами получения, хранения, передачи, обработки больших данных
освоение и применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения	программный продукт, процессы, методы и инструменты разработки программного продукта	ПК-9.2 [1] - Способен применять методы получения, хранения, передачи и обработки больших данных <i>Основание:</i> Профессиональный стандарт: 06.042	З-ПК-9.2[1] - Знать методы получения, хранения, передачи и обработки больших данных; У-ПК-9.2[1] - Уметь использовать методы получения, хранения, передачи и обработки больших данных; В-ПК-9.2[1] - Владеть методами получения, хранения, передачи и обработки больших данных
освоение и применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения	программный продукт, процессы, методы и инструменты разработки программного продукта	ПК-9.3 [1] - Способен использовать методы машинного обучения <i>Основание:</i> Профессиональный стандарт: 06.042	З-ПК-9.3[1] - Знать методы машинного обучения; У-ПК-9.3[1] - Уметь методы машинного обучения; В-ПК-9.3[1] - Владеть методы машинного обучения
научно-исследовательский			
построение моделей объектов профессиональной деятельности с использованием инструментальных средств компьютерного	программный продукт, процессы, методы и инструменты разработки программного продукта	ПК-4 [1] - способен применять существующие методы и алгоритмы решения задач распознавания и обработки данных <i>Основание:</i>	З-ПК-4[1] - Знать: методы и алгоритмы решения задач распознавания и обработки данных ; У-ПК-4[1] - Уметь: применять методы и алгоритмы решения

моделирования		Профессиональный стандарт: 40.011	задач распознавания и обработки данных ; В-ПК-4[1] - Владеть: методами и алгоритмами решения задач распознавания и обработки данных
---------------	--	-----------------------------------	--

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

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

№ п.п	Наименование раздела учебной дисциплины	Недели	Лекции/ Практи. (семинары) / Лабораторные работы, час.	Обязат. текущий контроль (форма*, неделя)	Максимальный балл за раздел**	Аттестация раздела (форма*, неделя)	Индикаторы освоения компетенции
	<i>3 Семестр</i>						
1	Организация Big Data - систем	1-8	16/0/8	ЛР-8 (25)	25	КИ-8	3-ПК-10.2, У-ПК-10.2, В-ПК-10.2, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-9.2, У-ПК-9.2, В-ПК-9.2, 3-ПК-9.3, У-ПК-9.3, В-ПК-9.3
2	Методы и средства организации	9-16	16/0/8	ЛР-16 (25)	25	КИ-16	3-ПК-10.2,

	распределенного анализа данных						У-ПК-10.2, В-ПК-10.2, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-9.2, У-ПК-9.2, В-ПК-9.2, 3-ПК-9.3, У-ПК-9.3, В-ПК-9.3
	<i>Итого за 3 Семестр</i>		32/0/16		50		
	Контрольные мероприятия за 3 Семестр				50	Э	3-ПК-10.2, У-ПК-10.2, В-ПК-10.2, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-9.2, У-ПК-9.2, В-ПК-9.2, 3-ПК-

							9.3, У- ПК- 9.3, В- ПК- 9.3
--	--	--	--	--	--	--	---

* – 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
ЛР	Лабораторная работа
КИ	Контроль по итогам
Э	Экзамен

КАЛЕНДАРНЫЙ ПЛАН

Неделя	Темы занятий / Содержание	Лек., час.	Пр./сем., час.	Лаб., час.
	<i>3 Семестр</i>	32	0	16
1-8	Организация Big Data - систем	16	0	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		
		2		1
		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 передаче 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		
		2		1
		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	Всего аудиторных часов		
		2		1
		Онлайн		

	интерфейса для организации взаимодействия распределенных систем.			
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		
		4		2
		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		
		4		2
		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		
		2		1
		Онлайн		
9-16	Методы и средства организации распределенного анализа данных	16	0	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		
		4		2
		Онлайн		
11 - 12	Хранение больших объемов данных Высокопроизводительные NoSQL-системы. Предпосылки, типы и характеристика. Состав и характеристики высокопроизводительных файловых систем на примере	Всего аудиторных часов		
		4		2
		Онлайн		

	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.			
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		
		4		2
		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 online-модулях. Применение гибридных вычислительных технологий в задачах машинного обучения. Способы повышения производительности алгоритмов машинного обучения с применением гибридных вычислительных технологий.	Всего аудиторных часов		
		4		2
		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

ТЕМЫ ЛАБОРАТОРНЫХ РАБОТ

Недели	Темы занятий / Содержание
	<i>3 Семестр</i>
1 - 8	Организация Big Data - систем. Hadoop Организация Big Data - систем. Hadoop.
9 - 12	Технологии обработки больших данных Технологии обработки больших данных.

13 - 15	Инфраструктура Инфраструктура.
---------	--

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	Аттестационное мероприятие (КП 1)
ПК-10.2	З-ПК-10.2	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	У-ПК-10.2	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	В-ПК-10.2	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
ПК-4	З-ПК-4	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	У-ПК-4	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	В-ПК-4	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
ПК-9.2	З-ПК-9.2	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	У-ПК-9.2	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	В-ПК-9.2	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
ПК-9.3	З-ПК-9.3	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	У-ПК-9.3	Э, КИ-8, КИ-16, ЛР-8, ЛР-16
	В-ПК-9.3	Э, КИ-8, КИ-16, ЛР-8, ЛР-16

Шкалы оценки образовательных достижений

Шкала каждого контрольного мероприятия лежит в пределах от 0 до установленного максимального балла включительно. Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет собой сумму баллов, заработанных студентом при выполнении заданий в рамках текущего и промежуточного контроля.

Итоговая оценка выставляется в соответствии со следующей шкалой:

Сумма баллов	Оценка по 4-ех балльной шкале	Оценка ECTS	Требования к уровню освоению учебной дисциплины
90-100	5 – «отлично»	A	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать

			теорию с практикой, использует в ответе материал монографической литературы.
85-89	4 – «хорошо»	В	Оценка «хорошо» выставляется студенту, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос.
75-84		С	
70-74		Д	
65-69	3 – «удовлетворительно»	Е	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала.
60-64			
Ниже 60	2 – «неудовлетворительно»	Ф	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

Оценочные средства приведены в Приложении.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ А22 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 А22 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. 004 С 36 Основы Data Science и Big Data. Python и наука о данных : , Санкт-Петербург: Питер, 2020
4. ЭИ Б 82 Основы работы с технологией CUDA : , Москва: ДМК Пресс, 2010
5. ЭИ С 18 Технология CUDA в примерах: введение в программирование графических процессоров : , Москва: ДМК Пресс, 2011

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. 510 А45 Алгоритмы : построение и анализ, Москва [и др.]: Вильямс, 2011
2. 004 П18 Параллельные вычисления на GPU. Архитектура и программная модель CUDA : учебное пособие, Москва: Издательство Московского университета, 2012

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

Специальное программное обеспечение не требуется

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

1. Кафедра "Компьютерные системы и технологии" (<http://dozen.mephi.ru>)

<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

1. Указания для прослушивания лекций

Перед началом занятий ознакомиться с учебным планом и списком рекомендованной литературы.

Перед посещением очередной лекции освежить в памяти основные концепции пройденного ранее материала. Подготовить при необходимости вопросы преподавателю. На каждой лекции следует задавать вопросы как по материалу текущей лекции, так и по ранее прочитанным лекциям.

При изучении лекционного материала обязательно следует сопоставлять его с материалом семинарских и лабораторных занятий.

Для более подробного изучения курса следует работать с рекомендованными литературными источниками и материалами из сети Internet.

2. Указания для проведения лабораторного практикума

Соблюдать требования техники безопасности, для чего прослушать необходимые разъяснения о правильности поведения в лаборатории.

Перед выполнением лабораторной работы провести самостоятельно подготовку к работе изучив основные теоретические положения, знание которых необходимо для осмысленного выполнения работы.

В процессе выполнения работы следует постоянно общаться с преподавателем, не допуская по возможности неправильных действий.

При сдаче зачета по работе подготовить отчет о проделанной работе, где должны быть отражены основные результаты и выводы.

3. Указания по выполнению самостоятельной работы

Получить у преподавателя задание и список рекомендованной литературы.

Изучение теоретических вопросов следует проводить по возможности самостоятельно, но при затруднениях обращаться к преподавателю.

При выполнении фронтальных заданий по усмотрению преподавателя работа может быть оценена без письменного отчета на основе ответов на контрольные вопросы, при условии активной самостоятельной работы.

10. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

1. Указания для проведения лекций

На первой вводной лекции сделать общий обзор содержания курса. Дать перечень рекомендованной основной литературы и вновь появившихся литературных источников.

Перед изложением текущего лекционного материала кратко напомнить об основных выводах по материалам предыдущей лекции.

Внимательно относиться к вопросам студентов и при необходимости давать дополнительные более подробные пояснения.

Периодически освещать на лекциях наиболее важные вопросы лабораторного практикума, вызывающие у студентов затруднения.

В середине семестра (ориентировочно после 8-й лекции) обязательно провести контроль знаний студентов по материалам всех прочитанных лекций.

Желательно использовать конспекты лекций, в которых используется принятая преподавателем система обозначений.

Давать рекомендации студентам для подготовки к очередным лабораторным работам.

На последней лекции уделить время для обзора наиболее важных положений, рассмотренных в курсе.

2. Указания для проведения лабораторного практикума

На первом занятии рассказать о лабораторном практикуме в целом (о целях практикума, инструментальных средствах для выполнения лабораторных работ, о порядке отчета по лабораторным работам), провести инструктаж по технике безопасности при работе в лаборатории.

Для выполнения каждой лабораторной работы студентам выдавать индивидуальные задания.

При принятии отчета по каждой лабораторной работе обязательно побеседовать с каждым студентом, задавая контрольные вопросы, направленные на понимание изучаемой в лабораторной работе проблемы.

По каждой работе фиксировать факт выполнения и ответа на контрольные вопросы.

Общий зачет по практикуму должен включать все зачеты по каждой лабораторной работе в отдельности.

Задания на каждую следующую лабораторную работу студенту выдавать по мере выполнения и сдачи предыдущих работ.

Автор(ы):

Ровнягин Михаил Михайлович

Рецензент(ы):
Васильев Н.П.