Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»

ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ

КАФЕДРА КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕХНОЛОГИЙ

ОДОБРЕНО УМС ИИКС

Протокол № 8/1/2025

от 25.08.2025 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Направление подготовки (специальность) [1] 09.03.04 Программная инженерия

Семестр	Трудоемкость, кред.	Общий объем курса, час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	В форме практической подготовки, час.	СРС, час.	КСР, час.	Форма(ы) контроля, экз./зач./КР/КП
5	2	72	16	16	0		40	0	ЗО
Итого	2	72	16	16	0	0	40	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 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**
ОПК-1 [1] – Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	З-ОПК-1 [1] – Знать основные объекты дискретной математики и методы их описания и исследований; проблемы алгоритмической разрешимости задач и эффективной вычислимости чисел. У-ОПК-1 [1] – Уметь решать основные задачи математической логики; однозначно задавать объекты дискретной математики, приводить их к стандартным формам, выполнять эквивалентные преобразования; определять сложности алгоритмов, применение прямых и косвенных доказательств теорем, определение

	принадлежности функций к соответствующим классам В-ОПК-1 [1] – Владеть методами математической логики для решения задач формализации, анализа и синтеза логических схем, для нахождения инвариантов циклических и условных конструкций в информатике, для выполнения эквивалентных преобразований; методами применения логического подхода к решению сложных задач с помощью их декомпозиции.
ОПК-6 [1] – Способен разрабатывать алгоритмы и программы, пригодные для практического использования, применять основы информатики и программирования к проектированию, конструированию и тестированию программных продуктов	З-ОПК-6 [1] – Знать основы информатики и программирования У-ОПК-6 [1] – Уметь разрабатывать алгоритмы и программы; проектировать, конструировать и тестировать программные продукты В-ОПК-6 [1] – Владеть основами информатики и программирования
ОПК-7 [1] – Способен применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой	З-ОПК-7 [1] – Знать основные концепции, принципы, теории и факты, связанные с информатикой У-ОПК-7 [1] – Уметь применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой В-ОПК-7 [1] – Владеть основными концепциями и принципами, связанными с информатикой
УКЕ-1 [1] – Способен использовать знания естественнонаучных дисциплин, применять методы математического анализа и моделирования, теоретического и экспериментального исследования в поставленных задачах	З-УКЕ-1 [1] – знать: основные законы естественнонаучных дисциплин, методы математического анализа и моделирования, теоретического и экспериментального исследования У-УКЕ-1 [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 [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

4. ВОСПИТАТЕЛЬНЫЙ ПОТЕНЦИАЛ ДИСЦИПЛИНЫ

Направления/цели воспитания	**Задачи воспитания (код)**
Профессиональное воспитание	Создание условий, обеспечивающих, формирование ответственности за профессиональный выбор, профессиональное развитие и профессиональные решения (В18)
Профессиональное воспитание	Создание условий, обеспечивающих,

	формирование научного мировоззрения, культуры поиска нестандартных научно-технических/практических решений, критического отношения к исследованиям лженаучного толка (В19)
Профессиональное воспитание	Создание условий, обеспечивающих, формирование профессионально значимых установок: не производить, не копировать и не использовать программные и технические средства, не приобретённые на законных основаниях; не нарушать признанные нормы авторского права; не нарушать тайны передачи сообщений, не практиковать вскрытие информационных систем и сетей передачи данных; соблюдать конфиденциальность доверенной информации (В40)

5. 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
	5 Семестр						
1	Организация Big Data - систем	1-8	8/8/0		25	КИ-8	З-ОПК-1, У-ОПК-1, В-ОПК-1, З-ОПК-6, У-ОПК-6, В-ОПК-6, З-ОПК-7, У-ОПК-7, В-ОПК-7, З-УКЕ-1, У-УКЕ-1, В-УКЕ-1
2	Методы и средства организации распределенного анализа данных	9-16	8/8/0		25	КИ-16	З-ОПК-1, У-ОПК-1, В-ОПК-1, З-ОПК-6, У-ОПК-6, В-ОПК-6, З-ОПК-7,

							У-ОПК-7, В-ОПК-7, З-УКЕ-1, У-УКЕ-1, В-УКЕ-1
	Итого за 5 Семестр		16/16/0		50		
	Контрольные мероприятия за 5 Семестр				50	ЗО	З-ОПК-1, У-ОПК-1, В-ОПК-1, З-ОПК-6, У-ОПК-6, В-ОПК-6, З-ОПК-7, У-ОПК-7, В-ОПК-7, З-УКЕ-1, У-УКЕ-1, В-УКЕ-1

* – 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
ЗО	Зачет с оценкой
З	Зачет

КАЛЕНДАРНЫЙ ПЛАН

Недели	**Темы занятий / Содержание**	**Лек., час.**	**Пр./сем., час.**	**Лаб., час.**
	5 Семестр	16	16	0
1-8	**Организация Big Data - систем**	8	8	0
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		
		1	1	0
		Онлайн		
		0	0	0
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		
		1	1	0
		Онлайн		
		0	0	0
3	**Архитектура BigData систем. Часть 2**	Всего аудиторных часов		

	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.	1	1	0
		Онлайн		
		0	0	0
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		
		2	2	0
		Онлайн		
		0	0	0
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		
		2	2	0
		Онлайн		
		0	0	0
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		
		1	1	0
		Онлайн		
		0	0	0
9-16	**Методы и средства организации распределенного анализа данных**	8	8	0
9 - 10	**Асинхронная обработка больших объемов данных** Принципы работы с технологией Apache Spark для выполнения асинхронных вычислительных операций и системы очередей для управления асинхронными	Всего аудиторных часов		
		2	2	0
		Онлайн		
		0	0	0

	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.			
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		
		2	2	0
		Онлайн		
		0	0	0
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		
		2	2	0
		Онлайн		
		0	0	0
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		
		2	2	0
		Онлайн		
		0	0	0

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 СЕМИНАРОВ

Недели	Темы занятий / Содержание
	5 Семестр
1 - 8	**Технологии обработки больших данных** Технологии обработки больших данных
9 - 16	**Инфраструктура. Организация Big Data - систем. Hadoop.** Инфраструктура. Организация Big Data - систем. Hadoop.

6. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

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

7. 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)
ОПК-1	З-ОПК-1	ЗО, КИ-8, КИ-16
	У-ОПК-1	ЗО, КИ-8, КИ-16
	В-ОПК-1	ЗО, КИ-8, КИ-16
ОПК-6	З-ОПК-6	ЗО, КИ-8, КИ-16
	У-ОПК-6	ЗО, КИ-8, КИ-16
	В-ОПК-6	ЗО, КИ-8, КИ-16
ОПК-7	З-ОПК-7	ЗО, КИ-8, КИ-16
	У-ОПК-7	ЗО, КИ-8, КИ-16
	В-ОПК-7	ЗО, КИ-8, КИ-16
УКЕ-1	З-УКЕ-1	ЗО, КИ-8, КИ-16
	У-УКЕ-1	ЗО, КИ-8, КИ-16
	В-УКЕ-1	ЗО, КИ-8, КИ-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-х балльной шкале	Отметка о зачете	Оценка ECTS

90-100	5 – *«отлично»*	*«зачтено»*	A
85-89	4 – «*хорошо*»		B
75-84			C
70-74			D
65-69	3 – «*удовлетворительно*»		
60-64			E
ниже 60	2 – «*неудовлетворительно*»	*«не зачтено»*	F

Оценка «отлично» соответствует глубокому и прочному освоению материала программы обучающимся, который последовательно, четко и логически стройно излагает свои ответы, умеет тесно увязывать теорию с практикой, использует в ответах материалы монографической литературы.

Оценка «хорошо» соответствует твердым знаниям материала обучающимся, который грамотно и, по существу, излагает свои ответы, не допуская существенных неточностей.

Оценка «удовлетворительно» соответствует базовому уровню освоения материала обучающимся, при котором освоен основной материал, но не усвоены его детали, в ответах присутствуют неточности, недостаточно правильные формулировки, нарушения логической последовательности.

Отметка «зачтено» соответствует, как минимум, базовому уровню освоения материала программы, при котором обучающийся владеет необходимыми знаниями, умениями и навыками, умеет применять теоретические положения для решения типовых практических задач.

Оценку «неудовлетворительно» / отметку «не зачтено» получает обучающийся, который не знает значительной части материала программы, допускает в ответах существенные ошибки, не выполнил все обязательные задания, предусмотренные программой. Как правило, такие обучающиеся не могут продолжить обучение без дополнительных занятий.

8. 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 : , Силен Д., Мейсман А., Али М., Санкт-Петербург: Питер, 2020

4. ЭИ Б 82 Основы работы с технологией CUDA : , Харламов А. А., Боресков А. В. , Москва: ДМК Пресс, 2010

5. ЭИ С 18 Технология CUDA в примерах: введение в программирование графических процессоров : , Кэндрот Э., Сандерс Д., Москва: ДМК Пресс, 2011

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

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

2. 004 П18 Параллельные вычисления на GPU. Архитектура и программная модель CUDA : учебное пособие, Сахарных Н.А. [и др.], 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/

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

10. 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.

11. 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.