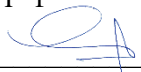


Федеральное государственное бюджетное образовательное
учреждение высшего образования
Московский государственный университет имени М.В.Ломоносова
Филиал Московского государственного университета имени М.В.Ломоносова
в городе Сарове

УТВЕРЖДАЮ

Директор филиала МГУ в городе Сарове

 /В.В. Воеводин/

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Технологии распределённого хранения и обработки данных»

Уровень высшего образования – подготовка магистров (неинтегрированная магистратура)

Направление подготовки – «Прикладная математика и информатика» (01.04.02)

Направленность (профиль) – «Суперкомпьютерные технологии математического моделирования и обработки данных»

Саров, 2021 г.

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 01.04.02 "Прикладная математика и информатика" программы магистратуры в редакции приказа МГУ от 30 декабря 2020 г. №1366.

1. Наименование дисциплины - **Технологии распределенного хранения и обработки данных**
2. Уровень высшего образования – **Подготовка научно-педагогических кадров в магистратуре.**
3. Направление подготовки, направленность: **«Системное программирование и компьютерные науки», «Суперкомпьютерные технологии математического моделирования и обработки данных».**
4. Место дисциплины в структуре ООП: **относится к вариативной части ОПОП.**
5. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников) в соответствии с Картами компетенций выпускников программ аспирантуры МГУ

Формируемые компетенции (код компетенции)	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
<p>Способен разрабатывать и реализовывать проекты, предусматривая и учитывая проблемные ситуации и риски на всех этапах выполнения проекта (УК-3).</p>	<p>Умеет разрабатывать архитектуры распределенных систем с учётом специфики проекта</p> <p>Может разрабатывать программные комплексы, функционирующие в распределённых вычислительных средах</p> <p>Умеет определять оптимальный компонентный состав распределённой системы</p>
<p>Способность разрабатывать и реализовывать масштабируемые параллельные методы и алгоритмы, участвовать в междисциплинарных исследованиях с применением суперкомпьютерных систем (МПК-2)</p>	<p>У1 Уметь реализовывать базовые алгоритмы обработки данных в распределённых средах в том числе с применением суперкомпьютерных технологий</p> <p>В1 Владеть навыками использования базовых распределённых технологий хранения и обработки данных для анализа больших объёмов инфор-</p>

<p>способность разрабатывать эффективное системное и прикладное программное обеспечение для суперкомпьютерных систем и высокопроизводительных кластеров (МПК-3);</p>	<p>мации в том числе на многопроцессорных вычислительных системах и суперкомпьютерах.</p> <p>ЗЗ (МПК-3) Знать: основные методы и подходы для оптимизации как последовательных, так и параллельных программ;</p> <p>УЗ (МПК-3) Уметь: оценивать эффективность распределенных алгоритмов;</p> <p>ВЗ (МПК-3) Владеть: навыками использования современных инструментальных средств для профилирования и анализа производительности параллельных программ.</p>
--	---

6. ОБЪЕМ ДИСЦИПЛИНЫ

Объем дисциплины составляет 4 зачетные единицы, всего 144 часа.

72 часа составляет контактная работа с преподавателем – 36 часов занятий лекционного типа, 36 часов научно-практических занятий.

72 часа составляет самостоятельная работа учащегося.

7. ВХОДНЫЕ ТРЕБОВАНИЯ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ ПРЕДВАРИТЕЛЬНЫЕ УСЛОВИЯ

Учащиеся должны владеть знаниями по линейной алгебре, дискретной математике, алгоритмам и структурам данных, программированию, архитектуре ЭВМ, операционным системам и программному обеспечению вычислительных систем в объеме, соответствующем основным образовательным программам бакалавриата по укрупненным группам направлений и специальностей 01.00.00 «Математика и механика», 02.00.00 «Компьютерные и информационные науки».

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Лекционная часть курса проводится с использованием презентаций лекций.

9. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы из них						Самостоятельная работа обучающегося, часы из них		
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости коллоквиумы, практические контрольные занятия и др)*	Всего	Выполнение практических контрольных заданий	Подготовка рефератов и т.п..	Всего
Тема 1. Теория распределенных систем Определения распределенной системы по Таненбауму и Лэмпорту. Виды распр систем (гриды, инф системы, тотальные (pervasive)). Классификация распределенных СУБД (реляционные, NoSQL, NewSQL, distributed ledgers).Цели и причины построения РС. Требования к распр си-	12	4	4	0	0	0	8	8	0	8

<p>стемам (прозрачность, открытость, масштабируемость, безопасность). Виды архитектур распределенных систем (централизованная, распределенная, гибридная).</p>										
<p>Тема 2. Коммуникация в распределенных системах Разновидности связующего ПО (middleware). Виды коммуникаций - RPC, MPI, брокеры сообщений, потоковая, широковебательная.</p>	12	4	4	0	0	0	8	8	0	8
<p>Тема 3. Модель распределённых вычислений Модель распределённых вычислений. Логические часы Лэмпорта. Векторные часы. Способы эффективной реализации векторных часов.</p>	12	4	4	0	0	0	8	8	0	8
<p>Тема 4. Алгоритмы взаимного исключения в распределенных системах. Алгоритмы на основе разрешений и алгоритмы с маркером. Централизованный, децентрализованный алгоритмы, алгоритм Лэмпорта, алгоритм маркерного</p>	12	4	4	0	0	0	8	8	0	8

окольца.										
Тема 5. Алгоритмы выбора лидера в распределенных системах. Алгоритм Гарсии-Молины, алгоритм Чанга-Робертса в кольцевой топологии, выборы лидера в беспроводных сетях,	12	4	4	0	0	0	8	8	0	8
Тема 6. Репликация и резервирование в распределенных системах Модели согласованности: строгая, последовательная (sequential), причинно-следственная (causal), согласованность в конечном счете (eventual). CAP-теорема Брюера.	12	4	4	0	0	0	8	8	0	8
Тема 7. СУБД Redis Основные характеристики (область применения, конкурирующие продукты, достоинства и недостатки). Репликация master-slave. Сохранение на диск (persistence): снимки (snapshots), AOF. Язык запросов Redis, ос-	16	4	4	0	0	0	8	8	0	8

новные 5 типов данных. Транзакции. Работа с Redis из программ на Python. HyperLogLog и GeoApi(geohash) в Redis.											
Тема 8. СУБД MongoDB Форматы JSON и BSON. Язык запросов MongoDB. Конвейер (aggregation pipeline). Использование MapReduce в MongoDB. Примеры реализации алгоритмов под MapReduce в Mongo.	10	4	2	0	0	0	6	8	0	8	
Тема 9. Технология MapReduce Обзор технологии. Алгоритмы: вычисление среднего, подсчет слов, наивный классификатор Байеса.	18	4	2	0	0	0	6	8	0	8	
Промежуточная аттестация: экзамен	4						4				
Итого	144	36	36	0	0	0	72	72		72	

10. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ МАГИСТРАНТОВ.

Самостоятельная работа учащихся состоит в изучении лекционного материала, учебно-методической литературы, выполнения практических заданий текущего контроля и промежуточной аттестации.

Основная учебно-методическая литература

1. *В.А. Крюков* Операционные системы распределенных вычислительных систем <https://parallel.ru/krukov/>
2. *Э. Таненбаум* Распределенные системы: принципы и парадигмы
3. *М.С. Косяков* Введение в распределенные вычисления
4. *J. Carlson* Redis in action
5. *K. Banker et al* MongoDB in action

Дополнительная учебно-методическая литература

1. *G. Coulouris et al* Distributed Systems: Concepts and Design
2. *S. Ghosh* Distributed Systems: An Algorithmic Approach

Ресурсы информационно-телекоммуникационной сети «Интернет»

1. <https://www.freecodecamp.org/news/a-thorough-introduction-to-distributed-systems-3b91562c9b3c/>
2. <https://martinfowler.com/articles/patterns-of-distributed-systems/>
3. <https://blog.pragmaticengineer.com/distributed-architecture-concepts-i-have-learned-while-building-payments-systems/>

Лицензионное программное обеспечение, в том числе отечественного производства

При реализации дисциплины может быть использовано следующее программное обеспечение:

1. Программный продукт Red Hat Enterprise Linux Server for HPC Compute Node for Power, LE, Self-support 4 шт. №5540331
2. Программный продукт Red Hat Enterprise Linux Server for HPC Head Node for Power, LE, Standard 1 шт. №5540332
3. Операционная система SUSE Linux Enterprise Server 11 SP4 for x86_64 16 шт.
4. Операционная система Red Hat Enterprise Linux Server 5.0 for x86_64 14шт.
5. Операционная система SUSE Linux Enterprise Server 10 SP3 for ppc64 7 шт.
6. Операционная система Ubuntu 18.04.
7. Программное обеспечение для виртуализации Oracle VM VirtualBox

8. Операционная система ALTLinuxMATEStarterkit 9 лицензияGPL
9. Программный продукт JetBrains IntelliJ IDEA Community Edition Free Educational Licenses
10. Программный продукт JetBrains PyCharm Community Edition Free Educational Licenses
11. Программный продукт JetBrains CLion Community Edition Free Educational Licenses
12. Программный продукт UPPAAL (<http://www.uppaal.org/>) академическая лицензия
13. Программный продукт Java 8 (64-bit) Oracle Corporation
14. Программный продукт Java SE Development Kit 8(64-bit) Oracle Corporation
15. Программный продукт NetBeans IDE 8.2 NetBeans.org
16. Программный продукт Dev-C++ Bloodshed Software
17. Программный продукт CodeBlocks The Code::Blocks Team
18. Программный продукт Free Pascal 3.0.0 Free Pascal Team
19. Программный продукт Python 3.5.1 (64-bit) Python Software Foundation
20. Программный продукт R for Windows 3.2.2 R Core Team
21. Программный продукт Haskell Platform 7.10.3 Haskell.org
22. Операционная система Microsoft Windows 7 корпоративная академическая лицензия
23. Операционная система Microsoft Windows 10 Education академическая лицензия
24. Программный продукт Microsoft Project Professional 2013 академическая лицензия
25. Программный продукт Microsoft Visio Professional 2013 академическая лицензия
26. Программный продукт Microsoft Visual Studio Professional 2013 - RUS [Русский (Россия)] академическая лицензия

Профессиональные базы данных и информационные справочные системы

1. <http://www.edu.ru> – портал Министерства образования и науки РФ
2. <http://www.ict.edu.ru> – система федеральных образовательных порталов «ИКТ в образовании»
3. <http://www.openet.ru> - Российский портал открытого образования
4. <http://www.mon.gov.ru> - Министерство образования и науки Российской Федерации
5. <http://www.fasi.gov.ru> - Федеральное агентство по науке и инновациям

Информационные технологии, используемые в процессе обучения

В процессе обучения используются технологии хранения данных Redis и MongoDB

11. РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ

Для преподавания дисциплины требуется лекционная аудитория, оборудованная персональным компьютером и проектором для воспроизведения презентаций лекций, маркерной или меловой доской. У студентов должны быть ноутбуки с возможностью выхода в Интернет.

12. ЯЗЫК ПРЕПОДАВАНИЯ

русский

13. РАЗРАБОТЧИК ПРОГРАММЫ, ПРЕПОДАВАТЕЛЬ

к.ф.-м.н., доцент Никольский Илья Михайлович(inikolsky@cs.msu.ru)

Оценочные средства для промежуточной аттестации по дисциплине «Разработка и реализация образовательных программ на основе ФГОС ВО» на основе Карт компетенций выпускников программ магистратуры МГУ

Промежуточная аттестация проводится в форме устного экзамена.

Каждый магистрант отвечает на 2 вопроса, направленных на оценку полученных им знаний и умений.

Средства для оценивания планируемых результатов обучения, критерии и показатели оценивания приведены в следующей таблице. В случае невыполнения практических заданий оценка за экзамен снижается на балл.

РЕЗУЛЬТАТ ОБУЧЕНИЯ по дисциплине (модюлю)	КРИТЕРИИ и ПОКАЗАТЕЛИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ по дисциплине из соответствующих Карт компетенций					ОЦЕНОЧНЫЕ СРЕДСТВА
	1	2	3	4	5	
	<i>Неудовлетворительно</i>	<i>Неудовлетворительно</i>	<i>Удовлетворительно</i>	<i>Хорошо</i>	<i>Отлично</i>	
31 (СПК-49) <i>Знать</i> основы теории распределенных систем и теории распределенных алгоритмов	отсутствие знаний	фрагментарные представления о теории распределенных систем и теории распределенных алгоритмов	сформированные представления о теории распределенных систем и теории распределенных алгоритмов на приемлемом уровне	сформированные представления о теории распределенных систем и теории распределенных алгоритмов на среднем уровне	систематизированные знания о теории распределенных систем и теории распределенных алгоритмов	Устные вопросы
32 (СПК-49) <i>Знать</i> модели согласованности	отсутствие знаний	фрагментарные представления о моделях согласованности в распределенных си-	сформированные представления о моделях согласованности в распреде-	сформированные представления о моделях согласованности в	системные знания о моделях согласованности в распределенных системах	Устные вопросы

в распределенных системах		стемах	ленных системах на приемлемом уровне	распределенных системах на среднем уровне		
<i>У1 (СПК-49)</i> <i>Уметь</i> моделировать сложные распределенные системы	отсутствие умений	имеет недостаточные умения моделирования сложных распределенных систем	имеет приемлемые умения моделирования сложных распределенных систем	имеет средние умения моделирования сложных распределенных систем	имеет системные умения моделирования сложных распределенных систем	Практическое самостоятельное задание
<i>У2 (СПК-49)</i> <i>Уметь</i> разрабатывать алгоритмы для распределенных алгоритмических моделей (в т.ч. MapReduce)	отсутствие умений	имеет недостаточные умения разработки алгоритмов для распределенных алгоритмических моделей	имеет приемлемые умения разработки алгоритмов для распределенных алгоритмических моделей	имеет средние умения разработки алгоритмов для распределенных алгоритмических моделей	имеет системные умения разработки алгоритмов для распределенных алгоритмических моделей (в т.ч. MapReduce)	Практическое самостоятельное задание
<i>У3 (СПК-49)</i> <i>Уметь</i> оценивать эффективность распределенных алгоритмов	отсутствие умений	имеет недостаточные навыки оценки эффективности распределенных алгоритмов	имеет приемлемые навыки оценки эффективности распределенных алгоритмов	имеет средние навыки оценки эффективности распределенных алгоритмов	имеет системные навыки оценки эффективности распределенных алгоритмов	Практическое самостоятельное задание
<i>В3(СПК-49)</i> Владеть основами работы с NoSQL базами данных Redis и MongoDB	отсутствие навыков	имеет недостаточные навыки работы с NoSQL базами данных Redis и MongoDB	имеет приемлемые навыки работы с NoSQL базами данных Redis и MongoDB	имеет средние навыки работы с NoSQL базами данных Redis и MongoDB	имеет системные навыки работы с NoSQL базами данных Redis и MongoDB	Практическое самостоятельное задание

Фонды оценочных средств

Примерные вопросы

1. Определения распределенной системы по Таненбауму и Лэмпорту.
2. Виды распределенных систем (гриды, информационные системы, тотальные (pervasive))
3. Классификация СУБД (реляционные, NoSQL, NewSQL, distributed ledgers).
4. Цели и причины построения РС (географически распределенная вычислительная среда, увеличение производительности, совместное использование ресурсов, отказоустойчивость)
5. Требования к распределенным системам (прозрачность, открытость, масштабируемость, безопасность)
6. Архитектура распределенных систем (централизованная, одноранговая, гибридная)
7. Разновидности связующего ПО (middleware)
8. Виды коммуникаций - RPC, основанные на сообщениях (transient - MPI, persistent - брокеры сообщений), потоковые, мультикаст (широковещательная)
9. Преимущества децентрализации данных, архитектур вычислительных системы, алгоритмов. Отличие распределенных алгоритмов от параллельных
10. Алгоритмы синхронизации часов (по Таненбауму)
11. Модель распределённых вычислений. Логические часы Лэмпорта.
12. Модель распределённых вычислений. Векторные часы.
13. Эффективная реализация векторных часов - дифференциальная пересылка
14. Алгоритмы взаимного исключения в распределенных системах. Алгоритмы на основе разрешений и алгоритмы с маркером. Централизованный, децентрализованный, алгоритм Лэмпорта, маркерное кольцо.
15. Алгоритмы выбора лидера. Алгоритм громилы, выбор лидера в беспроводных сетях, кольцевой алгоритм.
16. Репликация и резервирование в распределенных системах
17. Модели согласованности: строгая, последовательная (sequential), причинно-следственная (causal), в конечном счете (eventual)
18. CAP-теорема Брюера
19. Redis. Основные характеристики (область применения, конкурирующие продукты; плюсы и минусы). Репликация master-slave. Сохранение на диск (persistence): снимки (snapshots), AOF.
20. Redis. Язык запросов, основные 5 типов данных. Транзакции. Работа с Redis из программ на Питоне. Примеры коротких программ с использованием Redis.
21. Вероятностные алгоритмы (фильтр Блума, MinHash, HyperLogLog). HyperLogLog и GeoApi(geohash) в Redis
22. MongoDB. Формат JSON. Язык запросов. Конвейер (aggregation pipeline).
23. MongoDB. MapReduce, особенности его реализации в Mongo. Примеры реализации алгоритмов под MapReduce в Mongo.
24. Наивный классификатор Байеса на MapReduce

Методические материалы для проведения процедур оценивания результатов обучения *(при наличии)*