

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

	Рабочая программа дисциплины (модуля)
по дисциплине:	Параллельные и распределенные вычисления: дополнительные главы
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Физтех-школа Прикладной Математики и Информатики кафедра анализа данных
курс:	4
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 8 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Количество контрольных работ, заданий: 1

Программу составил: А.В. Бабенко

Программа обсуждена на заседании кафедры анализа данных 06.03.2020

Аннотация

Курс посвящен архитектуре инфраструктурных распределенных систем. В рамках этого курса слушатели узнают, какие принципы стоят в основе современных распределенных систем, познакомятся с такими понятиями, как scalability, fault-tolerance, consensus, distributed storage, map-reduce, zookeeper, etc.

1. Цели и задачи

Цель дисциплины

обучение технологиям параллельных вычислений, работе с распределенными системами обработки и хранения данных, с алгоритмами и архитектурными принципами, применяющимися в программировании распределенных систем.

Задачи дисциплины

- познакомить студентов с параллельными вычислениями и распределёнными системами;
- выработать навыки практического использования соответствующих технологий;
- научить создавать распределенные многопоточные приложения для обработки больших объемов данных.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.3 Знает основные требования информационной безопасности
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.2 Знает основные источники научно-технической и (или) технологической информации в области профессиональной деятельности
	ОПК-4.3 Умеет составлять аннотации, рефераты, библиографические перечни и обзоры информации в области своей профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации

ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценить качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива
	ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- теоретические основы параллельных вычислений. Технологии параллельных вычислений и распределенных систем обработки и хранения больших объемов данных. Области применения, характерные особенности и виды распределенных систем. Облачные вычислительные системы.

уметь:

- использовать средства языков программирования C++ и Java для разработки надежных программных систем. Разрабатывать интерфейсы и протоколы взаимодействия сетевых компонент. Использовать модель программирования MapReduce для параллельной реализации вычислений.

владеть:

- средствами разработки и тестирования программного кода на языках C++ и Java, стандартными библиотеками этих языков. Технологиями многопоточного и распределенного программирования.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение. Основы распределенных систем.	8	8		8
2	Архитектура распределенных stateful сервисов.	8	8		8
3	Иммутабельные данные в распределенных системах.	10	10		10
4	Архитектура систем Map-Reduce.	4	4		4
Итого часов		30	30		30

Подготовка к экзамену	0 час.
Общая трудоёмкость	90 час., 2 зач.ед.

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 8 (Весенний)

1. Введение. Основы распределенных систем.

1. Примеры распределенных систем; типичные проблемы, решаемые с помощью распределенных систем: масштабируемость, отказоустойчивость. Сложности, при проектировании распределенных систем. Понятие распределенного алгоритма, свойства liveness и safety, оценка сложности. Модели отказов в распределенных системах. Постановка задачи mutual exclusion.
2. Понятие времени в распределенной системе. Отношение happens-before, Lamport clock, vector clock. Алгоритм Лэмпарта.
3. Централизованный алгоритм для mutual exclusion, его ограничения. Модели консистентности в распределенных системах. CAP-теорема.

2. Архитектура распределенных stateful сервисов.

1. Отказоустойчивость stateful сервисов. Replicated state machine. Постановка задачи консенсуса, atomic broadcast. FLP impossibility (без строгого доказательства).
2. Алгоритм RAFT.
3. Масштабируемость. Шардирование. DHT.
4. Распределенные транзакции. Двухфазный коммит.
5. Multi-tenant системы. Ограничение нагрузки, квотирование. Gossip-протоколы.

3. Иммутабельные данные в распределенных системах.

1. Репликация данных. Erasure кодирование для репликации.
2. P2P системы, протокол BitTorrent.

4. Архитектура систем Map-Reduce.

1. Map-Reduce с точки зрения пользователя. Поиск узких мест и оптимизация map-reduce программ.
2. Распределенная сортировка в Map-Reduce.
3. Планирование ресурсов в вычислительном кластере. Понятие fair-share.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

6. Перечень рекомендуемой литературы

Основная литература

1. Технологии параллельного программирования MPI и OpenMP [Текст] : учеб. пособие для вузов / А. С. Антонов ; Моск. гос. ун-т им. М. В. Ломоносова .— М. : Изд-во Моск. ун-та, 2012 .— 344 с.

Дополнительная литература

1. Параллельное программирование на C++ в действии. Практика разработки многопоточных программ [Текст]/Э. Уильямс , -М., ДМК Пресс, 2012

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

<http://elibrary.ru/defaultx.asp> Научная электронная библиотека;
<http://www.twirpx.com> Все для студента

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Методические рекомендации позволяют студенту оптимальным образом организовать процесс обучения. В рабочей программе приведено примерное распределение часов аудиторной и внеаудиторной нагрузки по различным темам данной дисциплины.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Информатика и вычислительная техника

профиль подготовки: Физтех-школа Прикладной Математики и Информатики
кафедра анализа данных

курс: 4

квалификация: бакалавр

Семестр, формы промежуточной аттестации: 8 (весенний) - Дифференцированный зачет

Разработчик: А.В. Бабенко

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.3 Знает основные требования информационной безопасности
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.2 Знает основные источники научно-технической и (или) технологической информации в области профессиональной деятельности
	ОПК-4.3 Умеет составлять аннотации, рефераты, библиографические перечни и обзоры информации в области своей профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценить качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива
	ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

2. Показатели оценивания компетенций

В результате изучения дисциплины «Параллельные и распределенные вычисления: дополнительные главы» обучающийся должен:

знать:

- теоретические основы параллельных вычислений. Технологии параллельных вычислений и распределенных систем обработки и хранения больших объемов данных. Области применения, характерные особенности и виды распределенных систем. Облачные вычислительные системы.

уметь:

- использовать средства языков программирования C++ и Java для разработки надежных программных систем. Разрабатывать интерфейсы и протоколы взаимодействия сетевых компонент. Использовать модель программирования MapReduce для параллельной реализации вычислений.

владеть:

- средствами разработки и тестирования программного кода на языках C++ и Java, стандартными библиотеками этих языков. Технологиями многопоточного и распределенного программирования.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Примеры контрольных заданий:

1. Взаимное исключение и условная синхронизация. Модель памяти и низкоуровневые примитивы синхронизации.
2. Альтернативные подходы к реализации одновременных программ.
3. Системы параллельного программирования. Типовые модели программирования и шаблоны.
4. Основы параллельного программирования на системах с общей памятью на примере технологии OpenMP.
5. Основы параллельного программирования на системах с распределенной памятью на примере технологии MPI.
6. Локальная отладка и запуск программ на кластере. Приемы и стратегии реализации MapReduce-программ.
7. Высокоуровневые языки и инструментарии для работы с Hadoop.
8. Практические примеры использования MapReduce. Ограничения модели MapReduce, расширения и альтернативные подходы.
9. Технологии распределенного программирования. Распределенные системы хранения данных, репликация данных, NoSQL-системы.
10. Технологии распределенных вычислений, гриды, добровольные вычисления.
11. Облачные вычислительные системы.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Concurrency (одновременность). Области применения и проблематика. Способы реализации одновременных систем, процессы и потоки, программный инструментарий. Основы многопоточного программирования на примере языков C++ и Java. Типичные ошибки многопоточного программирования.
2. Параллельные вычисления. Области применения и проблематика. Современные параллельные вычислительные системы. Теоретические основы параллельных вычислений. Показатели качества параллельного алгоритма. Принципы разработки и типовые структуры параллельных алгоритмов. Методология PCAM.
3. Параллельная обработка больших массивов данных. Феномен Big Data. Модель программирования MapReduce. Принципы параллельной реализации вычислений. Область применения и примеры задач. Принципы распределенной реализации MapReduce на кластерных системах. Платформа Apache Hadoop. Интерфейсы прикладного программирования и реализация программ для Hadoop.

4. Распределенные системы и вычисления. Области применения, характерные особенности и виды распределенных систем. Проблемы построения распределенных систем. Теоретические основы распределенных вычислений, примеры распределенных алгоритмов. Способы взаимодействия распределенных процессов, сетевые протоколы.

Критерии оценивания

отлично (10) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

отлично (9) - выставляется студенту, показавшему свободное оперирование знаниями учебной программы дисциплины, выполнение заданий творческого характера.

отлично (8) - выставляется студенту, показавшему владение программным учебным материалом с наличием несущественных ошибок в действиях, самостоятельно исправляемых учащимся.

хорошо (7) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускается в ответе или в решении задач некоторые неточности.

хорошо (6) - выставляется студенту если он осознает воспроизведение программного учебного материала, в том числе и различной степени сложности, с несущественными ошибками, затруднения в применении отдельных навыков.

хорошо (5) - выставляется студенту если теоретическое содержание освоено не полностью, некоторые практические навыки сформированы недостаточно, в некоторых случаях были допущены ошибки.

удовлетворительно (4) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

удовлетворительно (3) - выставляется студенту в случае большого количества недочетов и неправильных ответов, а также пассивной работе в ходе занятий, многие учебные задания не выполнены.

неудовлетворительно (2) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

неудовлетворительно (1) - выставляется студенту, который не освоил теоретическое и практическое содержание курса, все выполненные учебные задания содержат грубые ошибки.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Дифференцированный зачет может проводиться по итогам текущей успеваемости и сдачи заданий и других видов работ, предусмотренных программой дисциплины и (или) путем организации специального опроса, проводимого в устной и (или) письменной форме.

При проведении дифференцированного зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на дифференцированном зачете не должен превышать одного астрономического часа.