

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:

Директор
А. В. Замятин

Оценочные материалы по дисциплине

Технологии высокопроизводительной обработки больших данных

по направлению подготовки

01.03.02 Прикладная математика и информатика

Направленность (профиль) подготовки:

Прикладная математика и инженерия цифровых проектов

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2024

СОГЛАСОВАНО:

Руководитель ОП
Д.Д. Даммер

Председатель УМК
С.П. Сущенко

Томск – 2024

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-4 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

ОПК-5 Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения.

ПК-2 Способен собирать, обрабатывать и анализировать данные для проведения научно-исследовательских работ в зависимости от проблемной и предметной области, создавать математическую модель исследуемого объекта.

ПК-3 Способен проектировать и разрабатывать программное обеспечение компьютерных и информационных систем, а также формализовать и алгоритмизировать поставленную задачу в рамках проекта в зависимости от проблемной и предметной области.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-4.1 Обладает необходимыми знаниями в области информационных технологий, в том числе понимает принципы их работы

ИОПК-4.2 Применяет знания, полученные в области информационных технологий, при решении задач профессиональной деятельности

ИОПК-4.3 Использует современные информационные технологии на всех этапах решения задач профессиональной деятельности

ИОПК-4.4 Демонстрирует умение составлять научные обзоры, рефераты и библиографии по тематике научных исследований.

ИОПК-5.1 Обладает необходимыми знаниями алгоритмов, принципов разработки алгоритмов и компьютерных программ

ИОПК-5.2 Разрабатывает алгоритмы и компьютерные программы для решения задач профессиональной деятельности

ИПК-2.1 Осуществляет сбор, обработку, анализ данных в исследуемой предметной области

ИПК-3.1 Способен предложить техническое и алгоритмическое решение для решения поставленной задачи в исследуемой предметной области

ИПК-3.2 Осуществляет оформление программного кода в соответствии с установленными требованиями, разработку процедур проверки работоспособности и измерения характеристик программного обеспечения

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- теоретические вопросы;
- лабораторные работы.

Текущий контроль по дисциплине проводится в форме вопросов по лекционному материалу, лабораторных работ и фиксируется в форме контрольной точки не менее одного раза в семестр.

Вопросы для текущего контроля по лекционному материалу:

Тема 1. Актуальность, базовая терминология и тенденции развития. Параллелизм компьютерных вычислений

1. Дать определение термину «высокопроизводительные вычисления».
2. Назвать основные способы повышения производительности вычислительной системы.
3. С чем связаны сложности дальнейшего повышения тактовой частоты процессора?
4. Дать определение термину «мультипроцессирование».

5. Рассказать о векторной обработке данных.
6. Применяется ли векторная обработка данных в современных персональных компьютерах? Назвать соответствующие технологии / расширения процессора.
7. Рассказать о многофункциональной обработке данных.
8. Рассказать о конвейере команд.
9. В чём разница пиковой и реальной производительности суперкомпьютера?
10. Какой суперкомпьютер на данный момент имеет максимальную пиковую производительность?

Тема 2. Архитектура высокопроизводительных вычислительных систем.

Классификация вычислительных систем.

1. По какому критерию классифицированы вычислительные системы в классификации Флинна?
2. Чем отличаются от прочих SISD системы?
3. Чем отличаются от прочих SIMD системы?
4. Чем отличаются от прочих MISD системы?
5. Чем отличаются от прочих MIMD системы?
6. Чем отличаются мультипроцессоры от мультикомпьютеров?
7. Чем отличаются UMA системы от NUMA систем?
8. Чем отличаются от прочих CC-NUMA системы?
9. Чем отличаются от прочих NCC-NUMA системы?
10. Чем отличаются от прочих SMP системы?
11. Чем отличаются от прочих мультипроцессорных вычислительных систем кластеры?

Тема 3. Облачные технологии, их свойства и типы.

1. Что такое «облачные технологии»?
2. Назовите основные свойства облачных технологий.
3. В чём причина широкого распространения облачных технологий?
4. Назовите основные типы облачных сервисов при проведении классификации по типу предоставляемого ресурса.
5. Приведите примеры облачных сервисов.
6. Почему IaaS (Infrastructure as a Service, Аппаратное обеспечение как сервис) не относится к сфере облачных технологий?
7. В чём отличие DaaS от SaaS?
8. Какие типы облачных сервисов предназначены в первую очередь для конечных пользователей, а какие для разработчиков ПО?
9. Какие модели развёртывания облачных систем вы знаете?
10. С какой целью создаются гибридные облака?

Тема 4. Технология MapReduce.

1. Дайте определение «функции высшего порядка» в программировании.
2. Какие функции высшего порядка используются в технологии облачных вычислений MapReduce?
3. Назовите шаги обобщённого алгоритма MapReduce.
4. Какие типы узлов создаются при использовании технологии MapReduce и каково их назначение?
5. Какими характеристиками должна обладать вычислительная задача, чтобы она могла быть эффективно решена с помощью технологии MapReduce?
6. Приведите пример задачи, которая может быть эффективно решена с помощью технологии MapReduce?
7. Запишите сигнатуру функции map() на псевдоязыке программирования для приведённого примера.
8. Запишите сигнатуру функции reduce() на псевдоязыке программирования для приведённого примера.
9. Назовите наиболее известные реализации MapReduce.

10. Для каких типов вычислительных систем в классификации Флинна имеет смысл использовать технологию MapReduce?

Тема 5. Распределённые файловые системы.

1. Дайте определение «распределённой файловой системы».
2. Дайте определение «репликации данных».
3. С какой целью выполняется репликация данных?
4. Каковы сферы применения распределённых файловых систем?
5. В чём состоит отличие распределённой файловой системы от распределённого хранилища данных?
6. Какие распределённые файловые системы вы знаете?
7. Опишите принцип работы Google File System.
8. Что происходит в Google File System, если узел данных, хранящий основную копию фрагмента файла, выходит из строя?
9. Опишите принцип работы Hadoop distributed file system.
10. Что происходит в Hadoop distributed file system, если узел данных, хранящий основную копию фрагмента файла, выходит из строя?
11. Назовите сходства и различия Google File System и Hadoop distributed file system.

Тема 6. Программирование для высокопроизводительных вычислений.

Методология проектирования параллельных алгоритмов.

1. Назовите основные проблемы параллельного программирования.
2. Расскажите о методологии организации параллельных вычислений для SIMD архитектуры.
3. Расскажите о методологии организации параллельных вычислений для MIMD архитектуры.
4. Назовите основные показатели качества параллельных методов.
5. Как связаны между собой такие показатели как ускорение и эффективность.
6. Может ли значение ускорения превышать количество процессоров, используемых для запуска параллельной программы?
7. Чем отличается сильная масштабируемость от слабой?
8. Для чего используется библиотека MPI, какие основные возможности предоставляет?
9. Что такое OpenMP, какие возможности предоставляет?
10. Имеет ли смысл использовать MPI на вычислителях с общей памятью?

Лабораторные работы:

1. Написать аппаратно ускоренную программу для поэлементного сложения векторов (например, с помощью технологии Nvidia CUDA).
2. Написать аппаратно ускоренную программу, подсчитывающую количество вхождений каждого символа в строке.
3. Создать аппаратно ускоренную реализацию алгоритма сжатия RLE (Run-Length Encoding).
4. Создать аппаратно ускоренную реализацию для алгоритма замены палитры цветов в изображении.

Студенту необходимо ответить правильно и развёрнуто на заданный ему вопрос текущего контроля.

Студенту требуется вовремя сдавать лабораторные работы, ориентироваться в написанном программном коде, правильно и ёмко отвечать на заданные по лабораторной вопросы.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Экзамен по окончании изучения курса проводится в письменно-устной форме. Допуском до экзамена является выполнение двух лабораторных работ. При выполнении менее трёх лабораторных работ максимальная оценка на экзамене – 4.

Примерный перечень теоретических вопросов:

1. Виды параллелизма компьютерных вычислений.
2. Классификацию вычислительных систем Флинна.
3. Классификация MIMD систем.
4. Свойства и типы облачных технологий.
5. Технология распределённых вычислений MapReduce.
6. Реализации MapReduce: сходства и различия.
7. Распределённые файловые системы.
8. Распределённая файловая система Google File System.
9. Распределённая файловая система Hadoop distributed file system.
10. Методология организации параллельных вычислений для SIMD, MIMD архитектур.
11. Показатели качества параллельных методов.
12. OpenMP и MPI: назначение каждой из технологий и их сравнительный анализ.

Оценка «отлично» ставится, если студент полноценно ответил на вопрос билета, а также на дополнительный вопрос, требующий аналитического сопоставления знаний, полученных при изучении различных тем данной дисциплины. Оценка «хорошо» ставится, если студент полноценно ответил на вопрос билета, но не ответил на дополнительный вопрос. Оценка «удовлетворительно» ставится, если студент частично ответил на вопрос билета.

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Тест:

1. Вычислительные системы какого типа чаще всего используются для организации высокопроизводительных вычислений
 - a. SIMD
 - b. MISD
 - c. SISD
 - d. MIMD (правильный ответ)
2. В каком случае можно прогнозировать снижение эффективности обработки данных с помощью технологии Nvidia Cuda?
 - a. Большой объём входных данных
 - b. Большое количество математических операций в алгоритме обработки данных
 - c. Большое количество условных переходов в алгоритме обработки данных
 - d. Входные данные имеют двумерную или трёхмерную природу

Ключи: 1 d), 2 c)

Теоретические вопросы:

1. Для чего нужна операция append в Google File System, почему не пользоваться всегда операцией write?

Ответ должен содержать упоминание того факта, что в GFS одновременно несколько клиентских приложений могут выполнять запись в один файл, поэтому операция write, в общем случае, небезопасна, так как может затереть изменения, вносимые параллельно другим клиентским приложением.

2. В чём принципиальное отличие мультипроцессора и мультикомпьютера (кластера)?

Ответ должен содержать упоминание того факта, что узлы мультипроцессора имеют общую память, а узлы мультикомпьютера не имеют общей памяти.

Информация о разработчиках

Дружинин Денис Вячеславович, канд. техн. наук, доцент кафедры теоретических основ информатики института прикладной математики и компьютерных наук НИ ТГУ.