

МИНОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ

Директор института прикладной
математики и компьютерных наук

А.В. Замятин

« 15 » июня 2023 г.



Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине
(Оценочные средства по дисциплине)

Технологии высокопроизводительной обработки больших данных

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки:

Интеллектуальный анализ больших данных

ОС составил:

канд. техн. наук,

доцент кафедры теоретических основ информатики

Д.В. Дружинин

Рецензент:

д-р техн. наук, профессор,

профессор кафедры теоретических основ информатики

Ю.Л. Костюк

Оценочные средства одобрены на заседании учебно-методической комиссии института прикладной математики и компьютерных наук (УМК ИПМКН).

Протокол от 08.06.2023 г. № 2

Председатель УМК ИПМКН,

д-р техн. наук, профессор

С.П. Сущенко

Оценочные средства (ОС) являются элементом системы оценивания сформированности компетенций у обучающихся в целом или на определенном этапе ее формирования.

ОС разрабатывается в соответствии с рабочей программой (РП) дисциплины.

1. Компетенции и результаты обучения, формируемые в результате освоения дисциплины «Технологии высокопроизводительной обработки больших данных»

Компетенция	Индикатор компетенции	Код и наименование результатов обучения	Критерии оценивания результатов обучения			
			Отлично	Хорошо	Удовлетворительно	Неудовлетворительно

<p>УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий</p>	<p>ИУК-1.1. Выявляет проблемную ситуацию, на основе системного подхода осуществляет её многофакторный анализ и диагностику. ИУК-1.2. Осуществляет поиск, отбор и систематизацию информации для определения альтернативных вариантов стратегических решений в проблемной ситуации. ИУК-1.3. Предлагает и обосновывает стратегию действий с учетом ограничений, рисков и возможных последствий.</p>	<p>ОР-1.1.1: Знать актуальность, базовую терминологию, архитектуру высокопроизводительных систем. ОР-1.2.1: Знать методы, алгоритмы высокопроизводительной обработки данных. ОР-1.3.1: Владеть навыками использования углубленных теоретических и практических знаний в области использования технологий высокопроизводительной обработки данных для решения задач интеллектуального анализа данных и биоинформатики. ОР-1.3.2: Владеть навыками разработки и применения математических методов системного и прикладного программного обеспечения для решения задач научной и проектно-технологической деятельности</p>	<p>Демонстрация высокого уровня знаний; способность самостоятельного анализа и реализации полученных знаний. Знание терминологии, архитектуры высокопроизводительных систем; методов и алгоритмов высокопроизводительной обработки данных. Владение навыками использования углубленных теоретических и практических знаний в области технологий высокопроизводительной обработки данных для решения задач интеллектуального анализа данных и биоинформатики и разработки и применения математических методов системного и прикладного программного обеспечения для решения задач научной и проектно-технологической деятельности.</p>	<p>В целом успешные, но содержащие отдельные пробелы знания терминологии, архитектуры высокопроизводительных систем; методов и алгоритмов высокопроизводительной обработки данных. В целом успешные, но содержащие отдельные пробелы владение навыками использования практических знаний в области технологий высокопроизводительной обработки данных для решения задач интеллектуального анализа данных и разработки и применения математических методов системного и прикладного программного обеспечения для решения задач научной и проектно-технологической деятельности.</p>	<p>Фрагментарное, неполное без грубых ошибок знания терминологии, архитектуры высокопроизводительных систем; методов и алгоритмов высокопроизводительной обработки данных. Фрагментарное, неполное без грубых ошибок владение навыками использования практических знаний в области технологий высокопроизводительной обработки данных для решения задач интеллектуального анализа данных.</p>	<p>Не имеет четкого представления об изучаемом материале, допускает грубые ошибки при использовании практических знаний в области технологий высокопроизводительной обработки данных для решения задач интеллектуального анализа данных.</p>
---	---	---	---	--	---	--

<p>ПК-5. Способен управлять получением, хранением, передачей, обработкой больших данных.</p>	<p>ИПК-5.1. Осуществляет мониторинг и оценку производительности обработки больших данных. ИПК-5.2. Использует методы и инструменты получения, хранения, передачи, обработки больших данных ИУК-5.3. Разрабатывает предложения по повышению производительности обработки больших данных.</p>	<p>ОП-5.1.1: Уметь применять системное и прикладное программное обеспечение для решения научных и практических задач технологической деятельности в области интеллектуального анализа данных и биоинформатики. ОП-5.2.1: Уметь выбирать методы и средства высокопроизводительной обработки данных для решения научно-исследовательских и прикладных задач. ОП-5.3.1: Уметь разрабатывать и применять математические методы для решения научно-исследовательских задач.</p>	<p>Демонстрация высокого уровня знаний; способность применять системное и прикладное программное обеспечение для решения научных и практических задач технологической деятельности в области интеллектуального анализа данных и биоинформатики. Умение выбирать методы, средства и математические методы для решения задач высокопроизводительной обработки данных для решения научно-исследовательских и прикладных задач.</p>	<p>В целом успешные, но содержащие отдельные пробелы умения применять системное и прикладное программное обеспечение для решения научных и практических задач технологической деятельности в области интеллектуального анализа данных и биоинформатики. В целом успешные, но содержащие отдельные пробелы умения выбирать методы, средства и математические методы для решения задач высокопроизводительной обработки данных для решения научно-исследовательских и прикладных задач.</p>	<p>Фрагментарное, неполное без грубых ошибок умение применять системное и прикладное программное обеспечение для решения научных и практических задач технологической деятельности в области интеллектуального анализа данных и биоинформатики.</p>	<p>Не имеет четкого представления об изучаемом материале, допускает грубые ошибки при выборе методов и средств высокопроизводительной обработки данных для решения научно-исследовательских и прикладных задач.</p>
--	---	--	---	---	---	---

2. Этапы формирования компетенций и виды оценочных средств

№	Этапы формирования компетенций (разделы дисциплины)	Код и наименование результатов обучения	Вид оценочного средства (тесты, задания, кейсы, вопросы и др.)
1.	Раздел 1. Введение	ОР-1.1.1, ОР-1.2.1	Вопросы для текущего контроля, лабораторные работы, вопросы для проведения промежуточной аттестации
2.	Раздел 2. Архитектура вычислительных систем	ОР-5.2.1, ОР-5.1.1, ОР-5.3.1	Вопросы для текущего контроля, лабораторные работы, вопросы для проведения промежуточной аттестации
3	Раздел 3. Облачные технологии, их свойства и типы	ОР-1.3.1, ОР-1.3.2	Вопросы для текущего контроля, лабораторные работы, вопросы для проведения промежуточной аттестации
4	Раздел 4. Технология вычислений MapReduce	ОР-5.2.1; ОР-5.1.1; ОР-5.3.1; ОР-1.3.1	Вопросы для текущего контроля, лабораторные работы, вопросы для проведения промежуточной аттестации
5	Раздел 5. Распределённые файловые системы	ОР-5.2.1; ОР-5.1.1; ОР-5.3.1; ОР-1.3.1; ОР-1.3.2	Вопросы для текущего контроля, лабораторные работы, вопросы для проведения промежуточной аттестации
6	Раздел 6. Программирование для высокопроизводительных вычислений.	ОР-5.2.1; ОР-5.1.1; ОР-5.3.1; ОР-1.3.1; ОР-1.3.2; ОР-5.1.1; ОР-5.2.1; ОР-7.3.1	Вопросы для текущего контроля, лабораторные работы, вопросы для проведения промежуточной аттестации

3. Типовые контрольные задания или иные материалы, необходимые для оценки образовательных результатов обучения

3.1. Типовые задания для проведения текущего контроля успеваемости по дисциплине «Технологии высокопроизводительной обработки больших данных»

Вопросы для проведения текущего контроля успеваемости:

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. Облачные технологии, их свойства и типы.

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?

Темы лабораторных работ:

Лабораторная работа 1. Написать программу, реализованную с помощью облачных технологий, для поэлементного сложения векторов (например, с помощью технологии MapReduce).

Цель: научить студентов организовывать обработку данных с помощью технологии MapReduce. Форма защиты работы – презентация преподавателю решения задачи на компьютере, отчет – по требованию преподавателя (при наличии существенных замечаний при устной защите).

Лабораторная работа 2. Написать аппаратно ускоренную программу, подсчитывающую количество вхождений каждого символа в строке.

Цель: научить студентов эффективно решать задачи обработки строковых данных с помощью специализированных технологий и программных комплексов для аппаратного ускорения вычисления. Форма защиты работы – презентация преподавателю решения задачи на компьютере, отчет – по требованию преподавателя (при наличии существенных замечаний при устной защите).

Лабораторная работа 3. Создать реализацию алгоритма сжатия RLE (Run-Length Encoding), ускоренную с помощью облачных технологий.

Цель: научить студентов эффективно решать задачи сжатия данных с помощью облачных технологий. Форма защиты работы – презентация преподавателю решения задачи на компьютере, отчет – по требованию преподавателя (при наличии существенных замечаний при устной защите).

Лабораторная работа 4. Создать облачную реализацию для алгоритма замены палитры цветов в изображении.

Цель: научить студентов эффективно решать задачи обработки изображений с помощью облачных технологий. Форма защиты работы – презентация преподавателю решения задачи на компьютере, отчет – по требованию преподавателя (при наличии существенных замечаний при устной защите).

3.2. Типовые контрольные вопросы для проведения промежуточной аттестации по дисциплине.

Вопросы для промежуточной аттестации (экзамен):

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.

4. Методические материалы, определяющие процедуры оценивания образовательных результатов обучения

4.1. Методические материалы для оценки текущего контроля успеваемости по дисциплине.

Рейтинговая система для оценки текущей успеваемости обучающихся

Таблица – Балльные оценки для элементов контроля

Элементы учебной деятельности	Максимальный балл с начала семестра	Оцениваемая компетенция
Опрос по темам на занятиях	20	УК-1, ПК-5
Лабораторные работы	50	УК-1, ПК-5

4.2. Методические материалы для проведения промежуточной аттестации по дисциплине «Технологии высокопроизводительной обработки больших данных».

Методические материалы включают: порядок формирования оценки при использовании балльно-рейтинговой системы; критерии оценивания результатов за промежуточную аттестацию, учитывающую оценки за компетенции.

Рейтинговая система для оценки промежуточной успеваемости обучающихся

Таблица – Балльные оценки для элементов контроля

Элементы учебной деятельности	Максимальный балл с начала семестра	Оцениваемая компетенция
Опрос по темам на занятиях	20	УК-1, ПК-5
Лабораторные работы	50	УК-1, ПК-5
Экзамен	30	УК-1, ПК-5

Пересчет баллов в оценки промежуточной успеваемости

Баллы на дату контрольной точки	Оценка
$\geq 90\%$ от максимальной суммы баллов	5 (зачтено)
От 70% до 89% от максимальной суммы баллов	4 (зачтено)
От 60% до 69% от максимальной суммы баллов	3 (зачтено)
$< 60\%$ от максимальной суммы баллов	2 (незачтено)