


МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ

Заведующий кафедрой
цифровых технологий

 / Кургалин С.Д.

22.04.2024 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.О.10 ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ И БОЛЬШИЕ ДАННЫЕ

1. Код и наименование направления подготовки/специальности:

02.04.01 Математика и компьютерные науки

2. Профиль подготовки/специализация:

Компьютерное моделирование и искусственный интеллект,

Компьютерные науки и информационные технологии для цифровой экономики

3. Квалификация выпускника: магистр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: цифровых технологий

6. Составители программы:

Борзунов Сергей Викторович, кандидат физико-математических наук, доцент

Рекомендована: НМС ФКН (протокол № 5 от 05.03.2024)

8. Учебный год: 2024-2025 Семестр: 2

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются: формирование знаний, умений и навыков в сфере разработки и эксплуатации аппаратного и программного обеспечения современных высокопроизводительных распределенных систем, формирование у обучающихся теоретических знаний и практических навыков по работе с большими данными.

Задачи учебной дисциплины:

- знакомство с архитектурой многопроцессорных вычислительных систем;
- обзор средств распараллеливания программного кода на системах с общей и распределенной памятью;
- получение начальных знаний и умений по созданию систем обработки Big Data и использованию BigData в системах реального времени.

10. Место учебной дисциплины в структуре ООП:

Обязательная часть, блок Б1. Для успешного освоения дисциплины необходимо предварительное изучение следующих разделов математики и компьютерных наук: математический анализ, программирование.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ОПК-2	Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы	ОПК-2.1	Владеет навыками создания и исследования новых математических моделей в естественных науках	Знать: области применения высокопроизводительных вычислений
		ОПК-2.2	Умеет использовать их в профессиональной деятельности	Уметь: использовать технологии высокопроизводительных вычислений для создания и исследования новых математических моделей в естественных науках
		ОПК-2.3	Имеет практический опыт создания и исследования подобных математических моделей и разработки теорий и методов для их описания	Владеть: навыками создания и исследования математических моделей с использованием технологий высокопроизводительных вычислений
ОПК-3	Способен демонстрировать базовые знания математических и естественных наук, основ программирования и информационных технологий.	ОПК-3.1	Обладает фундаментальными знаниями в области прикладного программирования и информационных технологий	Знать: программные средства и информационные технологии, используемые в высокопроизводительных вычислениях
		ОПК-3.2	Умеет использовать их в профессиональной деятельности	Уметь: использовать полученные знания для решения профессиональных задач

		ОПК-3.3	Имеет практический опыт применения программных средств, используемых при построении математических моделей в естественных науках	Владеть: навыками применения и разработки программных средств в сфере высокопроизводительных вычислений
--	--	---------	--	---

12. Объем дисциплины в зачетных единицах/час. — 3/108.

Форма промежуточной аттестации 2 семестр – экзамен.

13. Трудоемкость по видам учебной работы

Вид учебной работы	Трудоемкость		
	Всего	По семестрам	
		2 сем.	
Аудиторные занятия	48	48	
в том числе:	лекции	16	16
	практические		
	лабораторные	32	32
Самостоятельная работа	24	24	
Экзамен	36	36	
Итого:	108	108	

13.1. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК*
1. Лекции			
1.1	Введение в высокопроизводительные вычисления	Роль и значение высокопроизводительных вычислений в современном мире. Производительность вычислительных систем. Закон Амдала	https://edu.vsu.ru/course/view.php?id=3968
1.2	Технологии параллельного программирования	Параллельное программирование с использованием MPI. Структура MPI- программы. Сообщения, их передача и прием. Синхронное и асинхронное взаимодействие. Коллективный обмен данными. Виды коллективного обмена, барьеры, широковещательная рассылка данных. Коммуникаторы и топологии. Производные типы данных. Компиляция и отладка MPI-программ. Система программирования OpenMP. Распределенные вычисления с использованием GRID технологий	https://edu.vsu.ru/course/view.php?id=3968
1.3	Параллельные алгоритмы	Информационный граф алгоритма. Показатели эффективности параллельного алгоритма. Умножение матрицы на вектор. Матричное умножение. Сортировка	https://edu.vsu.ru/course/view.php?id=3968
1.4	Ведение в анализ больших данных	Основные определения, термины, задачи анализа больших данных. Понятие «Data Mining». Когнитивный анализ данных. Обзор источников информации для Big Data. Открытые источники информации: статистические сборники,	https://edu.vsu.ru/course/view.php?id=3968

		опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных	
1.5	Технологии хранения и обработки больших данных	Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. NoSQL СУБД. Технология Map-Reduce. Ее реализация Hadoop. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных	https://edu.vsu.ru/course/view.php?id=3968
1.6	Статистические методы анализа данных	Содержание темы. Основные понятия теории вероятностей и математической статистики. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы). Обзор современных программных средства анализа данных: Statistica, SPSS, Excel, R- Studio и другие; их преимущества и недостатки. Языки Python и R. Синтаксис языка R, основные типы данных	https://edu.vsu.ru/course/view.php?id=3968
2. Лабораторные занятия			
2.1	Технологии параллельного программирования	Параллельное программирование с использованием MPI. Структура MPI- программы. Сообщения, их передача и прием. Синхронное и асинхронное взаимодействие. Коллективный обмен данными. Виды коллективного обмена, барьеры, широковещательная рассылка данных. Коммуникаторы и топологии. Производные типы данных. Компиляция и отладка MPI-программ. Система программирования OpenMP. Распределенные вычисления с использованием GRID технологий	https://edu.vsu.ru/course/view.php?id=3968
2.2	Параллельные алгоритмы	Информационный граф алгоритма. Показатели эффективности параллельного алгоритма. Умножение матрицы на вектор. Матричное умножение. Сортировка	https://edu.vsu.ru/course/view.php?id=3968
2.3	Ведение в анализ больших данных	Основные определения, термины, задачи анализа больших данных. Понятие «Data Mining». Когнитивный анализ данных. Обзор источников информации для Big Data. Открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных	https://edu.vsu.ru/course/view.php?id=3968
2.4	Технологии хранения и обработки больших данных	Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. NoSQL СУБД. Технология Map-Reduce. Ее реализация Hadoop. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных	https://edu.vsu.ru/course/view.php?id=3968
2.5	Статистические методы анализа данных	Содержание темы. Основные понятия теории вероятностей и математической статистики. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы). Обзор современных программных средства анализа данных: Statistica, SPSS, Excel, R- Studio и другие; их преимущества и недостатки. Языки Python и R. Синтаксис языка R, основные типы данных	https://edu.vsu.ru/course/view.php?id=3968

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Введение в высокопроизводительные вычисления	2		0	2	4
2	Технологии параллельного программирования	2		6	2	10
3	Параллельные алгоритмы	2		6	4	12
4	Ведение в анализ больших данных	2		6	4	12
5	Технологии хранения и обработки больших данных	4		6	6	16
6	Статистические методы анализа данных	4		8	6	18
	Итого:	16		32	24	72

14. Методические указания для обучающихся по освоению дисциплины:

Освоение дисциплины складывается из аудиторной работы (учебной деятельности, выполняемой под руководством преподавателя) и внеаудиторной работы (учебной деятельности, реализуемой обучающимся самостоятельно).

Аудиторная работа состоит из работы на лекциях и выполнения практических (или лабораторных) заданий в объёме, предусмотренном учебным планом. Лекция представляет собой последовательное и систематическое изложение учебного материала, направленное на знакомство обучающихся с основными понятиями и теоретическими положениями изучаемой дисциплины. Лекционные занятия формируют базу для практических (или лабораторных) занятий, на которых полученные теоретические знания применяются для решения конкретных практических задач. Обучающимся для успешного освоения дисциплины рекомендуется вести конспект лекций и практических (лабораторных) занятий.

Самостоятельная работа предполагает углублённое изучение отдельных разделов дисциплины с использованием литературы, рекомендованной преподавателем, а также конспектов лекций, презентационным материалом (при наличии) и конспектов практических (лабораторных) занятий. В качестве плана для самостоятельной работы может быть использован раздел 13.1 настоящей рабочей программы, в котором зафиксированы разделы дисциплины и их содержание. В разделе 13.2 рабочей программы определяется количество часов, отводимое на самостоятельную работу по каждому разделу дисциплины. Больше количество часов на самостоятельную работу отводится на наиболее трудные разделы дисциплины. Для самостоятельного изучения отдельных разделов дисциплины используется перечень литературы и других ресурсов, перечисленных в пунктах 15 и 16 настоящей рабочей программы.

Успешность освоения дисциплины определяется систематичностью и глубиной аудиторной и внеаудиторной работы обучающегося.

При использовании дистанционных образовательных технологий и электронного обучения выполнять все указания преподавателей, вовремя подключаться к online занятиям, ответственно подходить к заданиям для самостоятельной работы.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 4-е изд., стер. — Санкт-Петербург : Лань, 2024. — 188 с. — ISBN 978-5-507-47346-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/362318
2	Борзунов, С.В. Суперкомпьютерные вычисления: практический подход / С.В. Борзунов, С.Д. Кургалин. — СПб. : БХВ, 2019. — 256 с. — (Серия: Учебная литература для вузов). — ISBN 978-5-9909805-2-5. — URL: https://elibrary.ru/item.asp?id=36628828
3	Борзунов, С.В. Языки программирования. Python: решение сложных задач : учебное пособие для ВО / С.В. Борзунов, С.Д. Кургалин. — Санкт-Петербург : Лань, 2023. — 192 с. — ISBN 978-5-507-45923-0. — URL: https://e.lanbook.com/book/319394 (дата обращения: 05.04.2024). — https://www.elibrary.ru/item.asp?id=53735343

б) дополнительная литература:

№ п/п	Источник
1	Методы и модели исследования сложных систем и обработки больших данных : монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ; под редакцией В. А. Смагин, А. Д. Хомоненко. — 2-е изд., стер. — Санкт-Петербург : Лань, 2024. — 236 с. — ISBN 978-5-507-50398-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/425024
2	Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
1	ЗНБ ВГУ: https://lib.vsu.ru/
2	Электронно-библиотечная система "Университетская библиотека online": http://biblioclub.ru/
3	Электронно-библиотечная система "Лань": https://e.lanbook.com/

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 4-е изд., стер. — Санкт-Петербург : Лань, 2024. — 188 с. — ISBN 978-5-507-47346-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/362318

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ, электронное обучение (ЭО), смешанное обучение):

При реализации дисциплины могут использоваться технологии электронного обучения и дистанционные образовательные технологии на базе портала edu.vsu.ru, а также другие доступные ресурсы сети Интернет.

18. Материально-техническое обеспечение дисциплины:

Аудитория для лекционных занятий: мультимедиа-проектор, экран для проектора, компьютер с выходом в сеть «Интернет». Специализированная мебель (столы ученические, стулья, доска). Программное обеспечение: LibreOffice v.5-7, программа для просмотра файлов формата pdf, браузер.

Компьютерный класс: специализированная мебель, персональные компьютеры, мультимедийный проектор, экран.

ПО: ОС Windows v.7, 8, 10, Интерпретатор языка Python, дистрибутив Anaconda, IDE PyCharm, редактор Jupiter.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенции	Индикаторы достижения компетенции	Оценочные средства
1.	Разделы 1-6	ПК-2	ПК-2.1 ПК-2.2 ПК-2.3	Лабораторные работы
2.	Разделы 1-6	ПК-3	ПК-3.1 ПК-3.2 ПК-3.3	Лабораторные работы
Промежуточная аттестация форма контроля – экзамен				Перечень вопросов к экзамену

20. Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1. Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью лабораторных работ.

Перечень лабораторных работ

1. Алгоритмы перемножения матриц.
2. Алгоритмы работы с числами произвольной точности.
3. Алгоритмы на графах.

Типовые задания для лабораторных работ

Лабораторная работа № 1 «Алгоритмы перемножения матриц»

Цель работы: изучить основные алгоритмы перемножения числовых матриц – классический алгоритм и алгоритм Штрассена, сравнить их асимптотическую сложность.

Требования к выполнению работы: обучающиеся должны составить программный код, реализующий основные алгоритмы перемножения матриц.

Критерии оценки: для получения оценки «зачтено» необходимо показать высокий уровень владения теоретическим материалом, уметь объяснить принцип работы написанной программы, верно ответить на дополнительные вопросы.

Задание: алгоритм Штрассена умножения матриц имеет асимптотическую сложность $\Theta(N^{\log 7})$, где N – размер перемножаемых квадратных матриц. При каких N время выполнения реализации алгоритма Штрассена на языках C и Python меньше, чем для алгоритма стандартного перемножения квадратных матриц?

Лабораторная работа № 2 **«Алгоритмы работы с числами произвольной точности»**

Цель работы: изучить основные алгоритмы работы с числами произвольной точности на примере вычисления числа π .

Требования к выполнению работы: обучающиеся должны составить программный код, реализующий алгоритм Чудновских.

Критерии оценки: для получения оценки «зачтено» необходимо показать высокий уровень владения теоретическим материалом, уметь объяснить принцип работы написанной программы, верно ответить на дополнительные вопросы.

Задание: широко используемый способ вычисления числа $\pi = 3,1415\dots$ – алгоритм Чудновских – основан на формуле, полученной как продолжение исследований представления обратной величины π^{-1} в виде гипергеометрических рядов. С помощью алгоритма Чудновских получите десятичное разложение числа $\pi = 3,1415\dots$ с N знаками, где $N = 10^5, 10^6$.

Лабораторная работа № 3 **«Алгоритмы на графах»**

Цель работы: изучить основные алгоритмы на графах, применяемые для анализа больших данных.

Требования к выполнению работы: обучающиеся должны составить программный код, реализующий алгоритм поиска кратчайшего пути в ориентированном графе.

Критерии оценки: для получения оценки «зачтено» необходимо показать высокий уровень владения теоретическим материалом, уметь объяснить принцип работы написанной программы, верно ответить на дополнительные вопросы.

Задание: транспортная сеть города задана в виде взвешенного ориентированного графа. Потенциальный поток автомобилей на каждой из дуг задается случайными величинами, распределенными в соответствии с единым законом распределения. Оцените время, требуемое на прохождение маршрута, начинающегося в узле A и оканчивающегося в узле B .

20.2. Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: перечень вопросов к экзамену.

Перечень вопросов к экзамену

1. Роль и значение высокопроизводительных вычислений в современном мире.
2. Производительность вычислительных систем. Закон Амдала.
3. Параллельное программирование с использованием MPI. Структура MPI-программы.
4. Сообщения, их передача и прием. Синхронное и асинхронное взаимодействие.
5. Коллективный обмен данными. Виды коллективного обмена, барьеры, широковещательная рассылка данных.
6. Коммуникаторы и топологии.
7. Производные типы данных.
8. Компиляция и отладка MPI-программ.
9. Система программирования OpenMP.
10. Распределенные вычисления с использованием GRID технологий.
11. Информационный граф алгоритма.
12. Показатели эффективности параллельного алгоритма.
13. Умножение матрицы на вектор.
14. Матричное умножение. Сортировка.
15. Параллельное программирование с использованием MPI.
16. Система программирования OpenMP.
17. Распределенные вычисления с использованием GRID технологий.
18. Матричное умножение.
19. Сортировка.
20. Базы данных. Системы управления базами данных.
21. NoSQL СУБД.
22. Технология Map-Reduce. Ее реализация Hadoop.
23. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.
24. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).
25. Обзор современных программных средства анализа данных: Statistica, SPSS, Excel, R-Studio.
26. Языки Python и R. Синтаксис языка R, основные типы данных.

Требования к студентам при проведении промежуточной аттестации

Для оценивания результатов обучения на экзамене используется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
---------------------------------	--------------------------------------	--------------

Дан полный, развёрнутый ответ на поставленный вопрос (вопросы), обучающийся свободно оперирует основными понятиями дисциплины, ориентируется в предметной области. Изложение материала не содержит ошибок, отличается последовательностью, грамотностью, логической стройностью.	Повышенный уровень	Отлично
Дан развёрнутый ответ на поставленный вопрос (вопросы), обучающийся свободно оперирует основными понятиями дисциплины, ориентируется в предметной области. Материал изложен в целом последовательно и грамотно, отсутствуют грубые ошибки, однако имеются отдельные неточности в определениях, вычислениях, доказательствах, изложениях положений теории.	Базовый уровень	Хорошо
Ответ на поставленный вопрос (вопросы) содержит изложение только базового теоретического материала, имеются ошибки в определениях, вычислениях, доказательствах, формулировках положений теории. Нарушена логическая последовательность в изложении материала.	Пороговый уровень	Удовлетворительно
Ответ на поставленный вопрос (вопросы) отсутствует, либо содержит грубые ошибки в определениях, вычислениях, доказательствах, формулировках положений теории. Обучающийся не владеет основными понятиями дисциплины. Отсутствует логическая последовательность в изложении материала.	–	Неудовлетворительно

Приведённые ниже задания рекомендуется использовать при проведении диагностических работ для оценки остаточных знаний по дисциплине

Задания с выбором ответа

1. Компиляция программы в рамках стандарта MPI может быть выполнена с помощью команды:

- a) mpirun
- b) mpifort**
- c) mpicxx
- d) mpicc

2. Получить значение размера коммутатора можно с помощью подпрограммы MPI:

- a) MPI_Comm_size**
- b) MPI_Comm_rank
- c) MPI_Init
- d) MPI_Finalize

3. В рамках операции редукции в стандарте MPI можно выполнить

- a) суммирование элементов массива**
- b) умножение двух матриц
- c) определение максимального элемента**
- d) перемножение элементов массива**

4. Операция широковещательной рассылки

- a) пересылает одни и те же данные всем процессам**
- b) выполняет частичную редукцию
- c) пересылает всем остальным процессам разные фрагменты данных

5. Подпрограмма MPI_Gather:

- a) выполняет широковещательную рассылку
- b) выполняет сбор данных**
- c) является операцией коллективного обмена данными**

6. Подпрограмма MPI_Scatter:

- a) выполняет распределение данных между процессами
- b) выполняет широковещательную рассылку
- c) является операцией синхронизации

7. Какие из следующих операций выполняют сбор данных на одном процессе

- a) **MPI_Gather**
- b) **MPI_Gatherv**
- c) MPI_Allgather

8. Какие их перечисленных операций относятся к классу векторных

- a) MPI_Alltoall
- b) **MPI_Alltoallv**
- c) MPI_Allgather
- d) **MPI_Allgatherv**

9. Подпрограмма MPI_Alltoall:

- a) выполняет операцию частичной редукции
- b) выполняет операцию полной редукции
- c) выполняет операцию синхронизации
- d) **выполняет передачу данных по схеме «каждый-всем»**

10. Коммуникатором в MPI называется:

- a) **группа процессов, наделенная общим контекстом обмена данными**
- b) подпрограмма, выполняющая передачу или прием сообщения
- c) подпрограмма для обработки сигналов, поступающих от операционной системы

11. Для создания коммуникатора предусмотрена команда

- a) MPI_Comm_size
- b) **MPI_Comm_create**
- c) MPI_Comm_rank

12. Создать группу коммуникаторов можно с помощью команды

- a) **MPI_Group_excl**
- b) **MPI_Group_incl**
- c) **MPI_Group_difference**
- d) MPI_Group_free

13. Подпрограмма MPI_Comm_size:

- a) является операцией двухточечного обмена
- b) **возвращает количество процессов в указанном коммуникаторе**
- c) задает количество процессов в указанном коммуникаторе
- d) проверяет возможность создания коммуникатора с указанным размером

14. Определить количество процессов в коммуникаторе можно с помощью:

- a) **MPI_Comm_size**
- b) MPI_Comm_rank
- c) MPI_Comm_create

15. В MPI существуют следующие типы двухточечных обменов:

- a) **стандартный обмен**
- b) **обмен по готовности**
- c) **синхронный обмен**
- d) **обмен с буферизацией**

16. Стандартная блокирующая двухточечная передача выполняется

- a) **MPI_Send**
- b) MPI_Ssend
- c) MPI_Bsend

d) MPI_Rsend

17. Двухточечная передача с буферизацией выполняется функцией

- a) MPI_Send
- b) MPI_Ssend
- c) **MPI_Bsend**
- d) MPI_Rsend

18. Неблокирующий вариант операций передачи сообщений существует для:

- a) **стандартной операции обмена данными**
- b) **обмена с буферизацией**
- c) **обмена «по готовности»**

19. Неблокирующий прием в MPI выполняется:

- a) **MPI_irecv**
- b) MPI_Recv
- c) MPI_Reduce_Scatter

20. Подпрограмма MPI_Wait предназначена для:

- a) **блокирующей проверки выполнения обмена**
- b) **неблокирующей проверки выполнения обмена**
- c) **приостановки выполнения программы на заданный период времени**

21. Подпрограмма MPI_Test предназначена для:

- a) **блокирующей проверки выполнения обмена**
- b) **неблокирующей проверки выполнения обмена**
- c) **приостановки выполнения программы на заданный период времени**

22. Подпрограмма MPI_Testall выполняет проверку:

- a) **завершения любой операции обмена из некоторого множества**
- b) **завершения всех обменов**
- c) **завершения одного обмена**

23. Подпрограмма MPI_Testany выполняет проверку:

- a) **завершения любой операции обмена из некоторого множества**
- b) **завершения всех обменов**
- c) **завершения одного обмена**

24. Неблокирующая передача с буферизацией выполняется :

- a) **MPI_Ibsend**
- b) MPI_Bsend
- c) MPI_Test
- d) MPI_Testall

25. Средством разработки многопоточных программ является:

- a) **OpenMP**
- b) PVM
- c) MPI

26. Какие виды зависимости операторов в программе существуют

- a) **зависимость по выходу**
- b) **поточковая зависимость**
- c) **антизависимость**

27. Какие ключевые слова OpenMP можно использовать после директивы parallel

- a) **firstprivate**
- b) **lastprivate**

c) reduction

d) runtime

28. Какие ключевые слова OpenMP можно использовать после директивы single

a) private

b) firstprivate

c) nowait

d) copyprivate

29. Какой термин применяется для обозначения структурированных и неструктурированных данных значительного объема и многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами?

a) **большие данные**

b) суперкомпьютерные вычисления

c) интеллектуальный анализ данных

d) облачные технологии

30. Какие технологии не относятся к технологиям анализа больших данных

a) **конвейерные вычисления**

b) системы управления базами данных категории NoSQL

c) алгоритмы MapReduce

d) методы интеллектуального анализа данных

31. Какой термин не входит в набор признаков VVV, изначально выбранных для характеристики больших данных?

a) **континуальность**

b) объем

c) скорость

d) разнообразие

Задания с кратким ответом

1. Назовите метод исследования больших данных, при котором изучаемая система заменяется моделью, с достаточной точностью описывающей реальную систему, с которой проводятся эксперименты с целью получения информации об этой системе.

Ответ: имитационное моделирование

2. Какой аббревиатурой принято обозначать модель вычислительной системы – параллельную машину с произвольным доступом к памяти?

Ответ: PRAM

3. Как называется величина, равная отношению $T_1(N)/T_p(N)$, где p – количество вычислительных узлов, $T_1(N)$ и $T_p(N)$ – время работы вычислительного алгоритма на системе с одним и p вычислительными узлами соответственно?

Ответ: ускорение **или** ускорение алгоритма

4. Как называется система управления базами данных, существенно отличающаяся от традиционных реляционных СУБД с доступом к данным средствами языка SQL за счет решения проблемы масштабируемости и доступности путем полного или частичного отказа от требований атомарности и согласованности данных.

Ответ: NoSQL

5. Какую аббревиатуру используют для обозначения отдельного устройства вычислительной системы, способное выполнять графический рендеринг, а также высокопроизводительные вычисления наряду с центральным процессором?

Ответ: GPU

Задания с развернутым ответом

1. Время работы последовательной версии некоторого алгоритма B равно $T_1(N) = 2^N t$, где N — размер входных данных, t — время выполнения одной вычислительной операции. В предположении, что алгоритм допускает максимальное распараллеливание, т.е. время работы на вычислительной системе с p процессорами равно $T_p(N) = T_1(N) / p$, вычислите время работы алгоритма B в следующем случае: $N = 32$, $p = 4$.
2. Время работы последовательной версии некоторого алгоритма B равно $T_1(N) = 2^N t$, где N — размер входных данных, t — время выполнения одной вычислительной операции. В предположении, что алгоритм допускает максимальное распараллеливание, т.е. время работы на вычислительной системе с p процессорами равно $T_p(N) = T_1(N) / p$, вычислите время работы алгоритма B в следующем случае: $N = 32$, $p = 16$.
3. Пусть доля последовательных вычислений f в некоторой программе равна $1/10$. Вычислите максимальное ускорение программы $(S_p)_{\max}$ на вычислительной системе с p процессорами с учетом закона Амдала.
4. Пусть доля последовательных вычислений f в некоторой программе равна $1/100$. Вычислите максимальное ускорение программы S_∞ с учетом закона Амдала.
5. Покажите, что условия Бернштейна не являются необходимыми условиями коммутативности двух операторов
6. На экзамене студент утверждает, что в рамках коммуникатора любой процесс может отправить с помощью функции `MPI_Send()` сообщение любому процессу, в том числе и самому себе. Прав ли студент?
7. Пусть организуется попарный обмен сообщениями в рамках коммуникатора `MPI_COMM_WORLD`, и количество процессов в параллельной программе равно N . Запишите множество рангов, доступных для использования в качестве рангов процесса-получателя в `MPI_Send()`;
8. Пусть организуется попарный обмен сообщениями в рамках коммуникатора `MPI_COMM_WORLD`, и количество процессов в параллельной программе равно N . Запишите множество рангов, доступных для использования в качестве рангов процесса-отправителя в `MPI_Recv()`.

Ответы

1. $T_p(N) = 2^{30} t$
2. $T_p(N) = 2^{28} t$
3. $10p/(p+9)$
4. $S_\infty = 100$
5. Рассмотрим следующую пару операторов (o_1, o_2) : $o_1 : a = a + 1$; $o_2 : a = a - 1$;
Ясно, что все три условия Бернштейна в данном случае нарушены. Тем не менее, результат работы программ, содержащих последовательности (o_1, o_2) и (o_2, o_1) , будет одинаков для всех возможных значений переменной a . Таким образом, условия Бернштейна не являются необходимыми условиями коммутативности двух операторов.
6. Да, в среде MPI процесс коммуникатора может отправить сообщение любому процессу, принадлежащему тому же коммуникатору, в том числе и себе. Тем не менее, указанной возможности в программе желательно избегать, т.к. в данном случае возможно появление ситуации «тупика».
7. $\text{rank} \in \{0, 1, \dots, N - 1\} \cup \{\text{MPI_PROC_NULL}\}$
8. $\text{rank} \in \{0, 1, \dots, N - 1\} \cup \{\text{MPI_PROC_NULL}\} \cup \{\text{MPI_ANY_SOURCE}\}$

Критерии оценивания	Баллы
---------------------	-------

Имеется верная последовательность всех этапов решения, обоснованно получен верный ответ.	3
Получен неверный ответ из-за вычислительной ошибки, при этом имеется верная последовательность всех этапов решения.	2
Получен верный ответ, однако имеются пропуски одного или двух этапов решения ИЛИ Решение не завершено, однако верно выполнен хотя бы один из этапов решения.	1
Решение не соответствует ни одному из критериев, перечисленных выше.	0