

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.В.ДВ.01.02
(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Технологии высокопроизводительной обработки данных

(наименование дисциплины)

по направлению подготовки
09.04.03 Прикладная информатика

направленность (профиль)/специализация
Прикладной анализ данных

Форма обучения: очная

Год набора: 2022

Общая трудоемкость: 33Е

Распределение часов дисциплины по семестрам

Семестр	3	Итого
Форма контроля	зачет	
Вид занятий		
Лекции	16	16
Лабораторные		
Практические	16	16
Руководство: курсовые работы (проекты) / РГР		
Промежуточная аттестация	0,25	0,25
Контактная работа	32,25	32,25
Самостоятельная работа	75,75	75,75
Контроль		
Итого	108	108

Рабочую программу составил(и):

профессор, доцент, д.техн.наук, Мкртычев С.В.

(должность, ученое звание, степень, Фамилия И.О.)

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

(должность, ученое звание, степень, Фамилия И.О.)

Рабочая программа дисциплины составлена на основании ФГОС ВО и учебного плана направления подготовки

09.04.03 Прикладная информатика

Срок действия программы дисциплины до «31» августа 2024 г.

УТВЕРЖДЕНО

На заседании кафедры «Прикладная математика и информатика»

(протокол заседания № 2 от «15 » сентября 2021 г.).

1. Цель освоения дисциплины

Цель – формирование у студентов научных представлений о принципах разработки и реализации высокопроизводительных технологий обработки данных и практических навыков применения полученных знаний для решения задач в области прикладного анализа данных.

2. Место дисциплины в структуре ОПОП ВО

Дисциплины и практики, на освоении которых базируется данная дисциплина: Методы и технологии прикладной информатики, Методы и инструментальные средства прикладного анализа данных.

Дисциплины и практики, для которых освоение данной дисциплины необходимо как предшествующее: Методы интерпретации и визуализации анализа больших данных.

3. Планируемые результаты обучения

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
ПК-4. Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации	ПК-4.1. Знает технологию управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации	Знать: технологию управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации
		Уметь: применять знания для управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации
		Владеть: технологией управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации
	ПК-4.2. Умеет управлять этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации	Знать: основы управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации
		Уметь: осуществлять организацию работ для управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации
		Владеть: приемами организации работ для управления этапами

		жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации
	ПК-4.3. Владеет технологическими методами и средствами управления этапами жизненного цикла методологической и технологической инфраструктурой анализа больших данных в организации	Знать: методы и средства управления инфраструктурой анализа больших данных в организации
		Уметь: применять технологические методы и средства для управления инфраструктурой анализа больших данных в организации
		Владеть: навыками аналитики для управления инфраструктурой анализа больших данных в организации

4. Структура и содержание дисциплины

Модуль (раздел)	Вид учебной работы	Наименование тем занятий (учебной работы)	Семестр	Объем, ч.	Баллы	Интерактив, ч.	Формы текущего контроля (наименование оценочного средства)
Модуль 1. Архитектура высокопроизводительных вычислительных систем	Лек	Тема 1. Основы построения высокопроизводительных вычислительных систем	3	4	-	-	
	Лек	Тема 2. Основы многопроцессорных вычислительных систем	3	4			
	Пр	1. Параллельные вычисления с помощью WinAPI 2. Вычисление времени работы алгоритма при распараллеливании	3	8	-	-	Отчеты по практическим работам №№1,2
Модуль 2. Технологии параллельных вычислений	Лек	Тема 3. Оценка производительности высокопроизводительных	3	4	-	-	
	Лек	Тема 4. Программное обеспечение высокопроизводительных вычислительных систем	3	4	-	-	
	Пр	3. Расчет ускорения совместной обработки данных 4. Управление параллелизмом транзакций в СУБД	3	8	-	-	Отчеты по практическим работам №№3,4
	ПА		3	0,25	-	-	
	Ср		3	75,75	-	-	
		Контроль			-	-	
Итого:				108			

5. Образовательные технологии

В рамках изучения дисциплины предусмотрено использование следующих образовательных технологий:

- технология традиционного обучения: лекции и практические работы, самостоятельная работа;
- технология проектного обучения: реализация и защита отчетов по практическим работам.

6. Методические указания по освоению дисциплины

Изучение дисциплины требует систематического и последовательного накопления знаний, следовательно, пропуски отдельных тем не позволяют глубоко освоить предмет.

В ходе лекционных занятий рекомендуется задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Студент может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых и выпускных квалификационных работ.

6.1. Рекомендации по подготовке к практическим занятиям

Студентам следует:

- при подготовке к практическим занятиям следует обязательно использовать не только лекции, учебную литературу, но и другие источники;
- в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения;
- на занятии доводить каждую задачу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по рассмотренному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться студентом на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа

данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

7. Оценочные средства

7.1. Паспорт оценочных средств

Семестр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
3	ПК-4	Отчеты по практическим работам №№1-4

7.2. Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1. Комплект отчетов по практическим работам (примеры). *(наименование оценочного средства)*

Практическая работа №1 «Параллельные вычисления с помощью WinAPI».

Форма отчета по практической работе № 1. В отчет по практической работе должны быть включены следующие пункты:

- титульный лист;
- цель работы;
- краткие теоретические сведения;
- описание хода выполнения работы;
- результаты выполненной работы.

Практическая работа №2 «Вычисление времени работы алгоритма при распараллеливании».

Форма отчета по практической работе № 2. В отчет по практической работе должны быть включены следующие пункты:

- титульный лист;
- цель работы;
- краткие теоретические сведения;
- описание хода выполнения работы;
- результаты выполненной работы.

Практическая работа №3 «Расчет ускорения совместной обработки данных».

Форма отчета по практической работе № 3. В отчет по практической работе должны быть включены следующие пункты:

- титульный лист;
- цель работы;
- краткие теоретические сведения;
- описание хода выполнения работы;
- результаты выполненной работы.

Практическая работа №4 «Управление параллелизмом транзакций в СУБД».

Форма отчета по практической работе № 4. В отчет по практической работе должны быть включены следующие пункты:

- титульный лист;

- цель работы;
- краткие теоретические сведения;
- описание хода выполнения работы;
- результаты выполненной работы.

Требования к оформлению

Отчёт по практической работе выполняется на страницах формата А4 в электронном виде.

При оформлении отчёта используется сквозная нумерация страниц, считая титульный лист первой страницей. Номер страницы на титульном листе не ставится. Номера страницы ставятся по центру сверху.

При оформлении отчёта соблюдать следующие требования:

- Для заголовков: полужирный шрифт, 14 пт, центрированный.
- Для основного текста: нежирный шрифт, 14 пт, выравнивание по ширине.
- Во всех случаях тип шрифта – Times New Roman, отступ абзаца 1.25 см, полуторный междустрочный интервал.
- Поля: левое – 2 см, правое, верхнее и нижнее – 1 см.

Процедура оценивания

Оценка выполненной работы проводится по следующим критериям:

1. Наличие всей существенной информации по работе
2. Точность и полнота предоставляемых сведений
3. Непротиворечивость приводимой информации
4. Правильность интерпретаций и выводов, которые сделаны по результатам работы
5. Степень достижения студентом поставленной цели
6. Обоснованность применяемого решения
7. Грамотность (содержательная) используемых формулировок

Критерии оценки за отчеты по практическим работам:

– оценка «зачтено» ставится студенту, который продемонстрировал результаты выполнения практической работы, соответствующие поставленным задачам, и представил отчет, оформленный должным образом и содержащий краткое описание полученных результатов;

– оценка «не зачтено» ставится студенту, который не продемонстрировал результаты выполнения практической работы или не представил по ней отчет или представленный отчет не соответствует требованиям по оформлению.

7.2.2. _____ Задания для оценки сформированности компетенций (наименование оценочного средства)

ПК-4 - Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации
код и наименование компетенции

ОМ закрытого типа

Задание 1

Выберите один правильный вариант ответ.

Что такое Hive?

- а) SQL-интерфейс доступа к данным для платформы Hadoop;
- б) диалект SQL для BigData;

в) хранилище данных Hadoop;
г) нереляционная СУБД в Hadoop
Правильный ответ: а

Задание 2

Выберите один правильный вариант ответа.

Процесс сбора, обеспечения целостности и хранения больших объемов структурированной и неструктурированной информации – это

- а) управление транзакционными данными;
- б) управление хранилищами данных;
- в) управление большими данными;
- г) управление мастер-данными

Правильный ответ: в.

Задание 3

Выберите правильные ответы.

Укажите этапы обработки больших данных.

- а) агрегирование;
- б) управление;
- в) анализ данных;
- г) интеграция

Правильные ответы: б,в,г

Задание 4

Выберите один правильный вариант ответа.

На каком этапе работы MapReduce каждая мэп-функция получает на вход список пар ключ – значение??

- а) reduce;
- б) map;
- в) input read;
- г) partition

Правильный ответ: б

Задание 5

Выберите один правильный вариант ответа.

На каком этапе работы MapReduce выполняется операция сокращения для каждого уникального ключа в отсортированном списке значений?

- а) reduce;
- б) map;
- в) input read;
- г) partition

Правильный ответ: а.

ОМ открытого типа

Задание 6

Дайте развернутый ответ.

Как процессоры классифицируются по организации управления потоком команд / способу загрузки исполнительных устройств?

Правильный ответ:

- с последовательной обработкой команд;
- с конвейером команд;
- суперскалярные процессоры;
- процессоры с длинным командным словом (VLIW) и т. д..

Задание 7

Дайте развернутый ответ.

Что представляет собой скалярный процессор?

Правильный ответ: это простейший класс микропроцессоров с традиционной фон неймановской архитектурой, оперирующий только со скалярными данными и обрабатывающие один элемент данных за одну инструкцию, типичными элементами данных могут быть целые или числа с плавающей запятой

Задание 8

Дайте развернутый ответ.

Перечислите основные принципы фон Неймана.

Правильный ответ:

- Принцип двоичного кодирования.
- Принцип однородности памяти.
- Принцип адресуемости памяти.
- Принцип программного управления.

Задание 9

Дайте развернутый ответ.

Что представляют собой процессы обращения к ЗУ?

Правильный ответ: это запись - процесс фиксации информации в ЗУ и чтение - процесс выдачи информации.

Задание 10

Дайте развернутый ответ.

Что такое MapReduce?

Правильный ответ: это модель распределённых вычислений от компании Google, используемая в технологиях [Big Data](#) для параллельных вычислений над очень большими (до нескольких петабайт) наборами данных в компьютерных кластерах, и фреймворк для вычисления распределенных задач на узлах (node) кластера

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1. Вопросы к промежуточной аттестации

Семестр 3

№ п/п	Вопросы к зачету
1.	Что такое вычислительная система?
2.	Сколько поколений компьютеров можно выделить?
3.	В чем заключается идея расслоения памяти?
4.	Каковы отличительные особенности архитектуры компьютера CRAY 1?
5.	В чем заключается конвейерный принцип обработки команд?
6.	Что такое центральный процессор?
7.	По каким признакам можно классифицировать процессоры?
8.	В чем заключаются основные принципы фон Неймана?
9.	Что представляют собой архитектура фон Неймана?
10.	В чем заключается принцип однородности памяти?
11.	Что такое устройство управления?
12.	Какие элементы архитектуры фон Неймана включает в себя центральный процессор?
13.	Как характеризуется уровень иерархической памяти?
14.	Какие виды запоминающих устройств может содержать основная память?
15.	Что такое кэш-память?
16.	В чем заключается назначение и логика работы кэш-памяти?
17.	Какие проблемы порождает включение в иерархию ЗУ кэш-памяти?
18.	Чем обусловлено введение дополнительных уровней кэш-памяти?
19.	Какие факторы влияют на выбор емкости кэш-памяти и размера блока?
20.	Как соотносятся характеристики обычной и дисковой кэш-памяти?
21.	На какие классы подразделяется виртуальная память?
22.	Какими средствами обеспечивается виртуализация памяти?
23.	Что представляет собой основная память?
24.	В чем заключаются преимущества и недостатки кольцевой топологии сети?
25.	В чем заключаются преимущества и недостатки линейной топологии сети?
26.	В чем заключается идея CISC архитектуры?
27.н	В чем заключаются особенности RISC архитектуры?
28.	В чем заключается идея конвейерных устройств?
29.	Что понимается под вектором в вычислительной технике?
30.	Что такое векторный процессор?
31.	Какие основные блоки включает в себя суперскалярный процессор?
32.	Что такое суперкомпьютер?
33.	Как классифицируются архитектуры вычислительных систем?
34.	На чем основывается классификация Флинна?
35.	На чем основывается классификация Хокни?
36.	В чем отличие архитектур ВС с общей и разделенной памятью?
37.	Какие уровни параллелизма реализуют симметричные мультипроцессорные системы?
38.	Что представляет собой архитектура NUMA?

39.	Что представляет собой архитектура СОМА?
40.	Что такое SMP?
41.	Какими характеристиками обладают SMP системы?
42.	Какие виды архитектур SMP систем можно выделить?
43.	В чем заключаются достоинства и недостатки различных SMP-архитектур?
44.	В чем заключается кластерная организация вычислительной системы?
45.	Какие задачи возлагаются на специализированное (кластерное) программное обеспечение?
46.	Как организуется взаимодействие между узлами кластерной ВС?
47.	Как измеряется производительность ВС?
48.	В чём заключается суть закона Амдала?
49.	Какие технологии используются для параллельного программирования?
50.	Какие задачи необходимо решить для использования потенциала параллельной архитектуры?

7.3.2. Критерии и нормы оценки

Семестр	Форма проведения промежуточной аттестации	Критерии и нормы оценки	
3	Устная	«зачтено»	ставится студенту, если он на поставленные зачетные вопросы дал грамотные (полные или краткие) ответы или допустил небольшие неточности
		«не зачтено»	ставится студенту, если он не дал ответы на поставленные зачетные вопросы или в ответе содержались фундаментальные ошибки

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Обязательная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Гергель В. П.	Теория и практика параллельных вычислений	Учебное пособие	2020	ЭБС «IPRbooks»

8.2. Дополнительная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Григорьев А.А.	Методы и алгоритмы обработки данных	Учебное пособие	2022	ЭБС "ZNANIUM.COM"
2	Некрасов К.А.	Параллельные вычисления общего назначения на графических процессорах	Учебное пособие	2017	ЭБС "ZNANIUM.COM"

8.3. Перечень профессиональных баз данных и информационных справочных систем

- Springer International Publishing, Part of Springer Science+Business Media
- NVIDIA GPU Educators Program. [Electronic resource] : [Образовательные программы NVIDIA GPU] . – Electronic data. [2018]. – Mode of access : <https://developer.nvidia.com/educators>
- The NVIDIA® CUDA® Toolkit. [Electronic resource] : [Инструментарий NVIDIA® CUDA®] . – Electronic data. [2018]. – Mode of access : <https://developer.nvidia.com/cuda-toolkit>
- Курс лекций по CUDA. [Электронный ресурс] . – Электронные данные. [2018]. – Режим доступа : http://www.nvidia.ru/object/cuda_state_university_courses_new_ru.html

8.4. Перечень программного обеспечения

№ п/п	Наименование ПО	Количество лицензий	Реквизиты договора (дата, номер, срок действия)
1	Visual Studio Community	Не ограничено	Свободное ПО
2	МУБД PostgreSQL	Не ограничено	Свободное ПО

8.5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий	Перечень основного оборудования
1.	Компьютерный класс. Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для проведения лабораторных работ. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации.(УЛК-312)	Столы ученические, переносной проектор, экран, стол преподавательский, стулья, доска аудиторная (маркерная), ПК с выходом в сеть Интернет
2.	Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной	Столы ученические двухместные (моноблок), доска аудиторная 3-х секционная (меловая), стол преподавательский , стулья, проектор Acer

	аттестации. (УЛК-418)	
3.	Компьютерный класс. Помещение для самостоятельной работы. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (Г-401)	Столы ученические, стулья ученические, ПК с выходом в сеть Интернет