МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет математики и компьютерных наук

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Современные методы распределенного хранения
и обработки данных

Кафедра прикладной математики
факультета математики и компьютерных наук

Образовательная программа бакалавриата:

01.03.05 - Статистика

Направленность (профиль) программы:
Анализ больших данных

Форма обучения:
очная

Статус дисциплины:
входит в ***часть ОПОП, формируемую участниками образовательных отношений***

Махачкала, 2023

Рабочая программа дисциплины «**Современные методы распределенного хранения и обработки данных**» составлена в 2023 году в соответствии с требованиями ФГОС ВО - бакалавриат по *направлению подготовки 01.03.05 Статистика* от 14.08.2020 г. №1032

Разработчик: кафедра прикладной математики:

Лугуева А.С, к.ф-м.н., доцент,

Рабочая программа дисциплины одобрена:
на заседании кафедры прикладной математики от «20» января 2023 г., протокол № 5
Зав. кафедрой _____ Кадиев Р.М.

на заседании Методической комиссии факультета математики и компьютерных наук от «25» января 2023 г., протокол № 4 .

Председатель _____ Ризаев М.К.

Рабочая программа дисциплины согласована с учебно-методическим управлением «___» _____2023 г.

Начальник УМУ _____ Гасангаджиева А.Г.
(подпись)

Аннотация рабочей программы дисциплины

Дисциплина «Современные методы распределенного хранения и обработки данных» входит ***в часть*** ОПОП***, формируемую участниками образовательных отношений*** по направлению 01.03.05 - Статистика. Дисциплина реализуется на факультете математики и компьютерных наук ДГУ кафедрой прикладной математики.

Содержание дисциплины охватывает круг вопросов, связанных с подготовкой студентов к эффективному использованию компьютерных систем и информационных технологий в будущей профессиональной деятельности.

Дисциплина нацелена на формирование следующих компетенций выпускника:

общепрофессиональных

- ОПК-2 - Способен формировать упорядоченные сводные массивы статистической информации и осуществлять расчет сводных и производных показателей в соответствии с утвержденными методиками, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ

профессиональных

- ПК-3-. Способен решать задачи профессиональной деятельности.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: *лекции, практические занятия, самостоятельная работа.*

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости в форме *контрольной работы* и промежуточный контроль в форме *зачета.*

Объем дисциплины: 3 зачетные единицы, в том числе в академических часах по видам учебных занятий:

Очная форма обучения

Семестр	Учебные занятия								Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен)
	в том числе:								
	всего	Контактная работа обучающихся с преподавателем						СРС, том чи экзам	
		всег	из них						
			Лекции	Практические занятия	Практически занятия	КСР	конс		
7	108	28	*32*		*18*			*58*	***зачет***

1. Цели освоения дисциплины:

Целями курса «Современные методы распределенного хранения и обработки данных» являются получение необходимых теоретических знаний и практических навыков в распределенном программировании для задач распределенной обработки данных. В курсе рассматриваются теоретические и практические аспекты технологий BIG DATA, методы и технологии параллельных и распределенных вычислений, использование которых постоянно возрастает в современной практической деятельности.

Проблематика параллельных вычислений является чрезвычайно широкой областью теоретических исследований и практически выполняемых работ по анализу и обработке данных. Материал данного курса ориентирован на изучение практических аспектов построения распределенных систем, в основе которой лежит распределённая файловая система HDFS. Предоставляется возможность изучения одной из лидирующих технологий, относящихся к классу BIG DATA которой является открытая платформа Hadoop, позволяющая обрабатывать огромные объемы данных в распределенной среде и обеспечивающая таким образом существенное снижение стоимости такого решения.

2.Место дисциплины в структуре ОПОП бакалавриата

Дисциплина «Современные методы распределенного хранения и обработки данных» входит ***в часть ОПОП, формируемую участниками образовательных отношений*** по направлению 01.03.05 - Статистика.

Для изучения дисциплины необходимы компетенции, сформированныев результате изучения дисциплин " Введение в Data Mining", " Машинное обучение".

Компетенции, сформированные при изучении данной дисциплины, используются при написании выпускной квалификационной работы.

Изучение данной дисциплины логически и содержательно-методически взаимосвязано с другими частями ОПОП.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения).

Код и наименование компетенции из ОПОП	**Код и наименование индикатора достижения компетенций** *(в соответствии с ОПОП*	**Планируемые результаты обучения**	**Процедура освоения**
ОПК-2. Способен формировать упорядоченные сводные массивы статистической информации и осуществлять расчет сводных и производных показателей в соответствии с утвержденными методиками, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ	ОПК- 2.1. Знает основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач.	Знает: основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач. Умеет: применять основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач. Владеет: основными моделями решения функциональных и вычислительных задач, инструментальными средствами для решения прикладных задач.	устный опрос, тестирование, письменный опрос
	ОПК-2.2. Умеет формировать массивы сводной статистической информации, применять математические и статистические методы при решении типовых профессиональных задач.	Знает: как формировать массивы сводной статистической информации, применять математические и статистические методы при решении типовых профессиональных задач. Умеет: формировать массивы сводной статистической информации, применять математические и статистические методы при решении типовых профессиональных задач. Владеет: математическими и статистическими методами для формирования массивов сводной статистической информации для решения типовых профессиональных задач.	устный опрос, тестирование, письменный опрос
	ОПК-2.3. Владеет навыками применения математического и статистического инструментария для решения прикладных задач, методами работы с современной вычислительной техникой . .	Знает: как применить математические и статистические инструментария и современную вычислительную технику для решения прикладных задач. Умеет: применить математические и статистические инструментария и современную вычислительную технику для решения прикладных задач. Владеет: навыками применения математического и	устный опрос, тестирование, письменный опрос

		статистического инструментария для решения прикладных задач, методами работы с современной вычислительной техникой.	
ПК-3. Способен решать задачи профессиональной деятельности.	ПК-3.1. Знает как решать задачи профессиональной деятельности.	Знает: общую методику статистического исследования и способы количественной формализации объекта наблюдений. Умеет: применить общую методику статистического исследования и способы количественной формализации объекта наблюдений при решении профессиональных задач. Владеет: навыками применения общей методики статистического исследования и способы количественной формализации объекта наблюдений при решении прикладных задач.	устный опрос, тестирование, письменный опрос
	ПК-3.2. Умеет решать задачи профессиональной деятельности.	Знает: как применить математический и эконометрический инструментарий для анализа количественных данных, в том числе с применением информационных систем и технологий. Умеет: применять математический и эконометрический инструментарий для анализа количественных данных, в том числе с применением информационных систем и технологий. Владеет: математическим и эконометрическим инструментарием для анализа количественных данных, в том числе с применением информационных систем и технологий, вычислительной техникой.	устный опрос, тестирование, письменный опрос
	ПК-3.3. Владеет навыками решения задач профессиональной деятельности.	Знает: как применить математический и статистический инструментарий и современную вычислительную технику для решения прикладных задач. Умеет: применить математический и статистический инструментарий, и современную вычислительную технику для решения прикладных задач. Владеет: навыками применения математического и статистического инструментария для решения прикладных задач, методами работы с современной вычислительной техникой.	устный опрос, тестирование, письменный опрос

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 3 зачетные единицы, 108 академических часа.

4.2. Структура дисциплины.

4.2.1. Структура дисциплины в очной форме

№ п/п	Разделы и темыдисциплины	Семестр		Виды учебной работы, включая самостоятель-ную работу студентов итрудоемкость (в часах)				Формы теку щего контроля успеваемости *(по неделям семестра)* Форма проме- жуточной аттестации *(по се- местрам)*
				Лекции	Практические занятия	Практические занятия	Самостоятельная	
	Модуль 1. Основы распределенных вычислений							
1	Введение в распределенные вычисления	7		4	2		6	устный опрос, тестирование, письменный опрос
2	Принципы разработки параллельных алгоритмов	7		4	2		6	устный опрос, тестирование, письменный опрос
3	Модель вычислений MapReduce и платформа Apache Hadoop	7		4	2		6	устный опрос, тестирование, письменный опрос
	Итого по модулю 1	7		12	6		18	Контрольная работа
	Модуль 2. Технологии распределенного программирования							
4	Распределенные системы и технологии распределенного программирования	7		4	2		6	устный опрос, тестирование, письменный опрос
5	Модель программирования Hadoop	7		4	2		6	устный опрос, тестирование, письменный опрос
6	Распределенная обработка данных с помощью Hadoop	7		2	2		8	устный опрос, тестирование, письменный опрос
	Итого по модулю 2	7		10	6		20	Контрольная работа
	Модуль 3. Распределенная файловая система							
7	Распределенная файловая система	7		2	2		6	устный опрос, тестирование, письменный опрос
8	Конфигурация Hadoopдля распределенной работы			2	2		6	устный опрос, тестирование, письменный

								опрос
9	Тестирование файловой системы HDFS			2	2		4	устный опрос, тестирование, письменный опрос
10	Текущие задачи в области распределеннойобработки данных			4			4	устный опрос, тестирование, письменный опрос
	Итого по модулю 3	7		10	6		20	Контрольная работа
	Итого	7		32	18		58	

4.3. Содержание дисциплины, структурированное по темам (разделам).

4.3.1. Содержание лекционных занятий по дисциплине

Темы лекций

Модуль 1. Основы распределенных вычислений

Тема 1. Введение в распределенные вычисления

Многопроцессность (concurrency), асинхронность и параллелизм. Процессы и потоки, способы взаимодействия. Основы многопоточного програм мирования с общей памятью на примере языка Java. Недетерминированность,состояние гонки. Синхронизация доступа к общим данным, взаимное исклю- чение, блокировка. Взаимная блокировка. Условная синхронизация потоков. Циклический барьер. Конфигурируемый пул потоков. Модель взаимодейст- вия процессов, основанная на асинхронном обмене сообщениями, на примереязыка Scala.

Тема 2. Принципы разработки параллельных алгоритмов

Цели применения параллельных вычислений. Пути достижения параллелизма. Классификация вычислительных систем. Системы с общей разделяемой памятью. Системы с распределенной памятью. Модель параллельного алгоритма, граф “операции-операнды”, расписание вычислений. Теоретические оценки времени выполнения параллельного алгоритма. Характеристики параллельных алгоритмов: ускорение, эффективность, стоимость. Теоретические оценки максимально достижимого ускорения, законы Амдала и Густавсона-Барсиса. Масштабируемость параллельного алгоритма, функция изоэффективности. Оценка коммуникационной сложности параллельных алгоритмов.

Декомпозиция исходной задачи на подзадачи. Типовые структуры алгоритмов. Функциональный параллелизм. Парадигма “разделяй и властвуй”. Геометрическая деком- позиция. Алгоритмы на рекурсивных структурах данных. Конвейерная обра- ботка. Координация на основе событий. Анализ информационных зависимо- стей между подзадачами. Выбор вычислительной платформы. Масштабиро- вание параллельного алгоритма. Распределение подзадач между исполните- лями. Примеры параллельных алгоритмов: матричные вычисления, сорти- ровка, алгоритмы на графах, задача N тел, рендеринг изображений.

Тема 3. Модель вычислений MapReduce и платформа Apache Hadoop

Основополагающие принципы MapReduce иллюстрируются на примере задачи построения инвертированного индекса. Определяется общая модель программирования и рассматриваются принципы параллельной реализации вычислений. Приводятся другие примеры задач, решение которых укладывается в описанную модель. Рассматриваются принципы распределенной реализации MapReduce на кластерных системах. В заключение обсуждаются истоки модели MapReduce, ее область применения, преимущества и недостатки.

Архитектура платформы ApacheHadoop и ее основные элементы – распределенная файловая система HDFS и реализация модели MapReduce. Интерфейс прикладного программирования и общие принципы реализации при- ложений для Hadoop на языке Java на примере задачи подсчета частоты встречаемости слов в тексте. Процесс установки Hadoop на локальной машине, с последующей компиляцией и запуском приложения. Принципы реализации функций map и reduce на языках, отличных от Java. Работе с данныминиз запуск приложений на учебном кластере, выбору значений параметров запуска заданий и отладке приложений.

Реализации процедуры кластерного анализа. Алгоритм canopyclustering, позволяющий снизить сложность вычислений при анализе данных большого объема и размерности. Реализация алгоритма canopyclustering в рамках модели MapReduce на примере задачи кластериза- ции фильмов из набора данных NetflixPrize. Реализация обхода и эффектив- ного представления графов в MapReduce, поиск кратчайших путей в графе. В качестве основного примера алгоритма на графах рассматривается задачавычисления значений PageRank. Принципы распараллеливания алгоритма и его реализация в рамках MapReduce.

Модуль 2. Технологии распределенного программирования

Тема 4. Распределенные системы и технологии распределенного программирования

Цели построения и виды распределенных систем. Способы взаимодействия распределенных процессов. Обмен сообщениями. Удаленные вызовы процедур. Пространство кортежей. Распределенная общая память. Современные технологии распределенного программирования: Java RMI. Проблемы построения распределенных систем, способы повышения отказоустойчивости.

Тема 5. Модель программирования Hadoop

История создания. MapReduce в Google (закрытая реализация на C++). Основы MapReduce на примере задачи построения инвертированного индек- са. Примеры использования MapReduce. Особенности MapReduce в Hadoop – открытая реализация на Java. Функция Map. Функция Reduce. Списки ключ- значение. Принципы распределенной реализации MapReduce на кластерных системах. Недостатки MapReduce.

Тема 6. Распределенная обработка данных с помощью Hadoop

Понятие и определение BigData. Что такое технология Hadoop. Применение Hadoop. Установка и настройка простого кластера. Настройка. Запуск. Проверка файловой системы HDFS. Тестирование HDFS. Конфигурации Hadoop для распределенной работы. Преимущества решений на базе Hadoop.

Модуль 3. Распределенная файловая система

Тема 7. Распределенная файловая система

Клиент HDFS. Сервер файлов контрольных точек (CheckpointNode),сервер резервных копий (BackupNode). Обновления и снимки файловой сис- темы. Чтение и запись файлов Размещение блоков. Управление репликацией. Балансировщик. Сканер блоков. Копирование данных между кластерами. Практическое использование файловой системы Долговечность хранения данных. Возможности совместного использования ресурсов HDFS. Масшта- бирование и объединение файловой системы.

Тема 8. Конфигурация Hadoop для распределенной работы

Дистрибутивы Hadoop. Аппаратные требования. Установка и настройка. Пример кластера ClouderaHadoop. Потребности приложений Hadoop. Работа с HDFS из Java. Запуск программы. Разделение Hadoop на главный и подчиненные узлы. Конфигурация кластера Hadoop. Обновление конфигурации Hadoop. Записи для узлов Hadoop в файле. Определение хозяина HDFS в файле core-site.xml Запуск демонов MapReduce. Получение информации о запущенных процессах на одном из подчиненных узлов. Тестирование HDFS.

Тема 9 Тестирование файловой системы HDFS

Проверка файловой системы HDFS. Параллельная обработка больших массивов данных. Распределенные хранилища данных. Распределенное про- граммирование. Примеры

Тема 10. Текущие задачи в области распределеннойобработки данных

Текущее состояние в области распределенной обработки данных, примеры технологий, задач и объемов обрабатываемых данных. Обзор актуальных задач в области распределенной обработки данных.

4.3.3. Содержание практическиих занятий по дисциплине.

На практических занятиях студены на практике знакомятся с платформой Apache Hadoop, овладевают навыками использования модели вычисления MapReduce для решения различных задач распределенной обработки данных.

Модуль 1. Основы распределенных вычислений

Тема 1. Введение в распределенные вычисления

Создание программы для иллюстрации ситуации взаимной блокировки.

Тема 2. Принципы разработки параллельных алгоритмов

Создание программы с параллельными матричными вычислениями.

Тема 3. Модель вычислений MapReduce и платформа Apache Hadoop

Создание программы с параллельной сортировкой.

Создание программы с графическим интерфейсом, которая позволяет загружать исходную картинку и изображение «водяного знака», скачивать результирующую картинку с наложенным водяным знаком. Программа должна использовать несколько потоков для указанных операций.

Модуль 2. Технологии распределенного программирования

Тема 4. Распределенные системы и технологии распределенного программирования

Создание программы для построения инвертированного индекса по набору текстовых документов.

Тема 5. Модель программирования Hadoop

Создание программы подсчета частоты встречаемости слов в тексте.

Тема 6. Распределенная обработка данных с помощью Hadoop

Реализация алгоритма canopyclustering в рамках модели MapReduce например задачи кластеризации фильмов из набора данных NetflixPrize.

Создание программы поиска кратчайших путей в графе с ипользованием MapReduce.

Модуль 3. Распределенная файловая система

Тема 7. Распределенная файловая система

Создайте программу с графическим интерфейсом, которая позволяет загружать исходный видеофайл и изображение водяного знака, накладывать «водяной знак» на содержимое видеофайла и скачивать результирующий видеофайл.

Тема 8. Конфигурация Hadoop для распределенной работы

Создание программы обмена сообщениями между распределенными процессами.

Тема 9 Тестирование файловой системы HDFS

Создание программы с графическим интерфейсом, которая позволяет загружать исходный видеофайл и изображение водяного знака, накладывать «водяной знак» на содержимое видеофайла и скачивать результирующий ви- деофайл. В виде распределенной среды выполнения используйте ApacheHadoop.

Тема 10. Текущие задачи в области распределеннойобработки данных

Создание списка новых возможностей, которые будут реализованы в следующих версиях Hadoop и HDFS.

5. Образовательные технологии

Лекции проводятся с использованием меловой доски и мела. Параллельно материал транслируется на экран с помощью мультимедийного проектора. Для проведения лекционных занятий необходима аудитория, оснащенная мультимедиа-проектором, экраном, доской, ноутбуком (с программным обеспечение для демонстрации слайд-презентаций).

Для проведения лабораторных занятий необходима аудитория на 15 человек, оснащена доской, компьютерами.

На лекционном и лабораторном занятиях посредством мультимедийных средств широко используется **демонстрационный материал**, который усиливает ощущения и восприятия обучаемого.

В частности, при изучении дисциплины предусмотрено применение следующих образовательных технологий:

– *Лекция-беседа*, являющаяся наиболее распространенной и сравнительно простой формой активного вовлечения студентов в учебный процесс. Эта лекция предполагает

непосредственный контакт преподавателя с аудиторией. Преимущество лекции-беседы состоит в том, что она позволяет привлекать внимание студентов к наиболее важным.

вопросам темы, определять содержание и темп изложения учебного материала с учетом особенностей студентов.

– *Проблемная лекция*, определяющим признаком которой является постановка и разрешение учебных проблем с различной степенью приобщения к этому слушателей. Такое занятие начинается с вопросов, с постановки проблемы, которую необходимо решить в ходе изложения материала.

– *Лекция-визуализация*, во время которой происходит переработка учебной информации по теме лекционного занятия в визуальную форму для представления студентам через технические средства обучения или вручную (схемы, рисунки, чертежи и т.п.).

Презентация – представление студентом наработанной информации по заданной тематике в виде набора слайдов и спецэффектов, подготовленных в выбранной программе.

– *Творческие задания* – самостоятельная творческая деятельность студента, в которой он реализует свой личностный потенциал, демонстрирует умение грамотно и ясно выражать свои мысли, идеи.

– *Компьютерные технологии* (компьютерный опрос, лекция – презентация, доклады студентов в сопровождении мультимедиа);

– *Диалоговые технологии* (опрос, взаимоопрос, дискуссия между студентами, дискуссия преподавателя и студентов);

– Технологии на основе метода *опережающего обучения* и др.

В ходе изучения дисциплины предусматриваются активные и интерактивные формы проведения занятий, в частности, с использованием разнообразных методов организации и осуществления:

– *учебно-познавательной деятельности* (словесные, наглядные и практические методы передачи информации, проблемные лекции и др.);

– *стимулирования и мотивации учебно-познавательной деятельности* (дискуссии, самостоятельные исследования по обозначенной проблематике, публикация статьи и др.);

– *контроля и самоконтроля* (индивидуального и фронтального, устного и письменного опроса, экзамена).

При ведении практических занятий по данной дисциплине используются такие стандартные методы обучения, как тестирование, фронтальный опрос, индивидуальный опрос, выполнение кейс-заданий,метод малых групп и т.п.

При проведении практических занятий в интерактивной форме используются следующие методы: дебаты, круглый стол, тематическая груп- повая дискуссия, блиц-опрос, научный кружок.

Предусмотрены также встречи с представителями республиканского управления и городских отделов статистики.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

Самостоятельная работа рассматривается как форма организации обучения, которая способна обеспечивать самостоятельный поиск необходимой информации, творческое восприятие и осмысление учебного материала в ходе аудиторных занятий, разнообразные формы познавательной деятельности студентов на занятиях и во внеаудиторное время, развитие аналитических способностей, навыков контроля и планирования учебно-

го времени, выработку умений и навыков рациональной организации учебного труда. Она является формой организации образовательного процесса, стимулирующей активность, самостоятельность и познавательный интерес студентов, а также одним из обязательных видов образовательной деятельности, обеспечивающей реализацию требований Федеральных государственных стандартов высшего профессионального образования (ФГОС).

Самостоятельная работа студента выполняется по заданию и при методическом руководстве преподавателя и реализуется непосредственно в процессе аудиторных занятий – на лекциях и практических занятиях, а также вне аудитории – в библиотеке, на кафедре, дома и т.д.

Аудиторная самостоятельная работа студента осуществляется на лекционных и практических занятиях в форме выполнения различных заданий и научных работ. Внеаудиторная самостоятельная работа студента традиционно включает такие виды деятельности, как проработка ранее прослушанного лекционного материала, конспектирование программного материала по учебникам, подготовка доклада, выполнение реферата, поиск наглядного материала, выполнение предложенных преподавателем заданий в виртуальной обучающей системе в режиме on-line и т.д.

Самостоятельная работа студента должна быть ориентирована на поиск и анализ учебного и научного материалов для подготовки к работе на семинарском занятии и обсуждения заранее заданных и возникающих в ходе занятия вопросов.

Эффективность и конечный результат самостоятельной работы студента зависит от умения работать с научной и учебной литературой, источниками и информацией в сети Интернет по указанным адресам.

При изучении дисциплины используются следующие виды самостоятельной работы студентов:

При оценивании результатов освоения дисциплины (текущей и промежуточной аттестации) применяется балльно-рейтинговая система, внедренная в Дагестанском государственном университете. В качестве оценочных средств на протяжении семестра используется тестирование, контрольные работы студентов, творческая работа, итоговое испытание.

Основными видами самостоятельной работы студентов являются:

1. изучение рекомендованной литературы, поиск дополнительного материала;
2. работа над темами для самостоятельного изучения;
3. подготовка к ***зачету***.

Примерное распределение времени самостоятельной работы студентов

Вид самостоятельной работы	**Примерная трудоёмкость, а.ч.**	
	очная	
Текущая СРС		
работа с лекционным материалом, с учебной литературой	6	
самостоятельное изучение разделов дисциплины	6	
подготовка к лабораторным занятиям	6	
подготовка к контрольным работам	6	
подготовка и сдача зачета	10	

Творческая проблемно-ориентированная СРС		
выполнение научных докладов и рефератов	6	
поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	6	
исследовательская работа, участие в конференциях, семинарах, олимпиадах	6	
анализ информации по теме на основе собранных данных	6	
Итого СРС:	***58***	

Темы, виды и содержание самостоятельной работы по дисциплине

Темы	Виды и содержание самостоятельной работы	Форма контроля
Введение в распределенные вычисления	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос, тестирование, презентация, проверка тетрадей.
Принципы разработки параллельных алгоритмов	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Подготовка к слабораторному занятию по теме, составление конспекта.	Устный опрос, тестирование, презентация, проверка тетрадей.
Модель вычислений MapReduce и платформа Apache Hadoop	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Поиск и анализ дополнительной литературы.	Устный опрос, тестирование, презентация, проверка тетрадей.
Распределенная файловая система	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Подготовить реферат по теме.	Устный опрос, тестирование, презентация, проверка тетрадей..
Конфигурация Hadoopдля распределенной работы	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос, тестирование, презентация, проверка тетрадей.
Тестирование файловой системы HDFS	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов;	Устный опрос, тестирование, презентация, проверка тетрадей..

	2. Разработать электронную презентацию	
Распределенная файловая система	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Подготовить реферат по теме.	Устный опрос, тестирование, презентация, проверка тетрадей.
Конфигурация Hadoopдля распределенной работы	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос, тестирование, презентация, проверка тетрадей.
Тестирование файловой системы HDFS	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос, тестирование, презентация, проверка тетрадей.
Текущие задачи в области распределеннойобработки данных	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос, тестирование, презентация, проверка тетрадей.

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

7.1. **Типовые контрольные задания**

<u>Задания для промежуточного контроля</u>

Задание 1.
Создайте программу для иллюстрации ситуации взаимной блокировки.

Задание 2.
Создайте программу с параллельными матричными вычислениями.

Задание 3.
Создайте программу с параллельной сортировкой.

Задание 4.

Создайте программу с параллельной сортировкой.

Задание 5.

Создайте программу с графическим интерфейсом, которая позволяет загружать исходную картинку и изображение «водяного знака», скачивать результирующую картинку с наложенным водяным знаком. Программа должна использовать несколько потоков для указанных операций.

Задание 6.

Создайте программу для построения инвертированного индекса по на-бору текстовых документов.

Задание 7.

Создайте программу подсчета частоты встречаемости слов в тексте.

Задание 8.

Реализация алгоритма canopyclustering в рамках модели MapReduce например задачи кластеризации фильмов из набора данных NetflixPrize.

Задание 9.

Создайте программу поиска кратчайших путей в графе с использовани-ем MapReduce.

Задание 10.

Реализуйте алгоритм PageRank.

Задание 11.

Создайте программу с графическим интерфейсом, которая позволяетзагружать исходный видеофайл и изображение водяного знака, накладывать

«водяной знак» на содержимое видеофайла и скачивать результирующий видеофайл.

Задание 12.

Создайте программу обмена сообщениями между распределеннымипроцессами.

Задание 13.

Создайте программу с реализацией функций Map и Reduce в платформе Apache Hadoop.

Задание 14.

Создайте программу с графическим интерфейсом, которая позволяет загружать исходный видеофайл и изображение водяного знака, накладывать

«водяной знак» на содержимое видеофайла и скачивать результирующий видеофайл. В виде распределенной среды выполнения используйте ApacheHadoop.

Задание 15.

Сконфигурировать HDFS, установив количество копий в 4 (реплика- ция), и проверить работу файловой системы с этой конфигурацией.

Задание 16.Запустить Hadoop кластер из трех нод – 1 главной и 2 подчиненных.

Задание 17.

Создать программу, которая реализует параллельную версию подсчета слов (WordCount).

Задание 18.

Сделать список новых возможностей, которые будут реализованы вследующих версиях Hadoop и HDFS.

Практические задания

Практическое задание 1

Электронная таблица как стандартного OLAP-клиента для доступа к данныхмхранилища.

Средство «Подбор параметра» Структурирование данных и получение промежуточных итогов. Консолидация данных. Сортировка данных. Фильтрацияданных.

Практическое задание 2.

СУБД. Создание базы данных: создание таблиц, выбор типа полей, свойства, установка ключевых полей, связь между таблицами, ввод информации. Работа с формами.

Практическое задание 3.

СУБД. Создание запросов - режим дизайна, создание запросов режим SQL(использование основных команд SELECT. Оператор Where. Предикаты Top,Distinct).

Практическое задание 4.

СУБД Задание для самостоятельного выполнение (Создание бд «Деканат»,создание таблиц, запросов, форм)

Практическое задание 5.

Основы работы в MySQL Server. Основные типы данных Особенности диалекта SQL в СУБД MySQL. Создание и выбор базы дан-

ных. Создание таблиц. Оператор CREATE TABLE. База данных «Книжный магазин» Вставка, удаление иобновление данных операторы INSERT; DELETE и TRUNCATE, UPDATE.

Практическое задание 6.

Выборка данных из одной таблицы с помощью оператора SELECT Использование в запросах операторов и встроенных функций MySQL.

Использование ключевых слов DISTINCT. Сортировка. Использования функцииcount( )

Практическое задание 7.

Использование объединений в запросах к нескольким таблицам JOIN , CROSS JOIN (перекрестное объединение) или INNER JOIN (внутреннее объединение).

Создание вложенных запросов

Лабораторная работа 8.

Создание хранимых процедур CREATE PROCEDURE Создание триггеров с помощью оператора CREATE trigger

Критерии оценивания:

- (для каждого задания):

10 б. – задание выполнено верно;

9-7 б. – при выполнении задания были допущены неточности, не влияющие нарезультат;

6-3 б. – при выполнении задания были допущены ошибки;

2 - 1 б. – при выполнении задания были допущены существенные ошибки;0 б. – задание не выполнено.

Максимальное количество баллов по лабораторным заданиям – 10

Перечень вопросов для промежуточного опроса

1. К средствам Электронной таблицы, предназначенных для анализа данныхотносятся…
2. Для чего используются фильтры, сортировка данных
3. Для чего предназначено средство «Подбор параметра»

4. Основные компоненты реляционной СУБД
5. Технология работы с реляционной СУБД.
6. Основные модели баз данных.
7. Способы создания таблиц
8. Способы создания запросов
9. Что такое первичный ключ
10. Для чего используются таблицы в БД 11.К
11. Какие типы данных полей есть в MySQL
12. 12.Что такое целостность данных
13. Как устанавливаются связи между таблицами
14. Способы создания отчетов
15. Общая характеристика языка SQL
16. Синтаксис языка SQL
17. Понятия и сущность Olap-технологий
18. Компоненты Olap-систем
19. .Что такое хранилище данных (ХД)
20. Чем отличается Х от БД

Вопросы к зачету

1. Основные понятия распределённой обработки данных
2. Условия и требования к распределенной обработке данных
3. Основные принципы создания и функционирования распределённых базданных
4. Olap- технологии.
5. Понятия и сущность OLAP – технологий
6. OLAP – продукты: требования, компоненты
7. Технологии объектного связывания данных
8. Хранилища данных: свойства, данные, источник данных
9. Витрины данных
10. Компоненты хранилища
11. Выбор метода реализации хранилищ данных
12. Оптимизация хранилищ данных
13. Классификация хранилищ данных
14. Организация защиты конфиденциальной информации
15. Методы защиты данных в распределенных базах
16. Уязвимости и угрозы безопасности данных
17. Средства защиты баз данных
18. Облачные системы обработки данных.
19. Облачные хранилища
20. Системы управления потоками данных.

21. Системы хранения больших данных.
22. Платформы больших данных.
23. Обработка данных в реальном времени.
24. Системы управления большими данными.
25. Аналитические платформы.

Практико-ориентированные задания к зачету

Задание 1

Создать базу данных

БД «Поступление лекарственных средств»

Код лекарства	Код лекарства	Код поставщика
Название лекарства	Код поставщика	Сокращенное название
Показания к применению	Дата поставки	Полное название
Единица измерения	Цена за единицу	Юридический адрес
Количество в упаковке	Количество	Телефон
Название производителя	Код поступления	ФИО руководителя

Разработать следующие запросы:

1. Вывести список лекарств с указанным показанием к применению.
2. Вывести дату поставки, сумму, ФИО руководителя от поставщика иназвание лекарства по коду поступления больше указанного числа.

Задание 2

Создать базу данных

Код клиента	Фамилия клиента	Адрес	Инспектор
101	Алексеев А.А.	пр. Шолохова, д.8	Гаврилова А.К.
102	Иванов И.И.	ул. Металлургическая, 1	Науменко Ю.С.
103	Петрова А.П.	пр. Шолохова, 104/6 кв. 2	Гаврилова А.К.
104	Сидоров С.К.	ул. 14-я линия, д. 14	Антипова К.Р.
105	Яковлева М.Ф.	ул. Сержантова, 12	Науменко Ю.Р.

Выполните запрос в режиме конструктора который выведет информацию: ФИОИнспектора работающего с клиентом Ивановым И.И.

Задание 3

Создать базу данных

Код клиента	Фамилия клиента	Адрес	Инспектор
101	Алексеев А.А.	пр. Шолохова, д.8	Гаврилова А.К.
102	Иванов И.И.	ул. Металлургическая, 1	Науменко Ю.С.
103	Петрова А.П.	пр. Шолохова, 104/6 кв. 2	Гаврилова А.К.
104	Сидоров С.К.	ул. 14-я линия, д. 14	Антипова К.Р.
105	Яковлева М.Ф.	ул. Сержантова, 12	Науменко Ю.Р.

Выполните запрос в режиме SQL который выведет информацию: ФИОИнспектора работающего с клиентом Ивановым И.И.

Задание 4

Создать базу данных

Код клиента	Фамилия клиента	Адрес	Инспектор
101	Алексеев А.А.	пр. Шолохова, д.8	Гаврилова А.К.
102	Иванов И.И.	ул. Металлургическая, 1	Науменко Ю.С.
103	Петрова А.П.	пр. Шолохова, 104/6 кв. 2	Гаврилова А.К.
104	Сидоров С.К.	ул. 14-я линия, д. 14	Антипова К.Р.
105	Яковлева М.Ф.	ул. Сержантова, 12	Науменко Ю.Р.

Создайте запрос, который покажет Фамилии клиентов и Фамилии инспекторов,работающих с этими клиентами.

Задание 5

Создать базу данных

Код клиента	Фамилия клиента	Адрес	Инспектор
101	Алексеев А.А.	пр. Шолохова, д.8	Гаврилова А.К.
102	Иванов И.И.	ул. Металлургическая, 1	Науменко Ю.С.
103	Петрова А.П.	пр. Шолохова, 104/6 кв. 2	Гаврилова А.К.
104	Сидоров С.К.	ул. 14-я линия, д. 14	Антипова К.Р.

105	Яковлева М.Ф.	ул. Сержантова, 12	Науменко Ю.Р.

Используя Мастер форм, создайте форму по столбцам «Фамилия клиента и егоАдрес».

Критерии оценивания:

2 теоретических вопроса (60 баллов), 1 практико-ориентированное задание (40баллов):

1 теоретический вопрос:

30 – 20 баллов - изложенный материал верен, наличие знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;

19 – 10 баллов - наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов;

9 – 0 баллов - ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неточность ответов на дополнительные и наводящие вопросы.

Практико-ориентированное задание:

40 -30 баллов - правильные, уверенные действия по применению полученных знаний на практике, практическое задание выполнено верно и в полном объеме;

29 – 20 баллов – при выполнении задания были допущены неточности, невлияющие на результат;

19-10 баллов – при выполнении задания были допущены ошибки;

9- 0 баллов- при выполнении задания были допущены существенные ошибки; задание не выполнено.

51-100 баллов (зачет)

0-50 баллов (незачет)

7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

1. Общий результат выводится как интегральная оценка, складывающая из текущего контроля – 50% и промежуточного контроля – 50%.

Текущий контроль по дисциплине включает:

- посещение занятий– 20 баллов,

- выполнение лабораторной работы– 10 баллов,
- выполнение самостоятельных, контрольных работ– 20 баллов.

Промежуточный контроль по дисциплине включает:

- письменная контрольная работа - 50 баллов.

2. Критерии оценок при проведении текущего контроля успеваемости

- Выполнение контрольной работы**:**

оценка «отлично» - выставляется студенту, если студент дал подробные ответы на все заданные вопросы. При этом студент должен показать знания не только из основной литературы, но и знания из дополнительной литературы, сети Internet;

оценка «хорошо» - выставляется студенту, если студент дал полные ответы на все вопросы, показав знания из основной литературы. При этом студент допустил несущественнее недочеты в ответах и незначительные нарушения логики изложения материала;

оценка «удовлетворительно»: знание и понимание основного материала, наличие несущественных ошибок (не более 50%) при неспособности их последовательного и логического изложения, вызывает затруднение использование терминологии дисциплины;

оценка «неудовлетворительно»: непонимание сущности вопросов, грубые существенные ошибки в ответе, отсутствие способности к письменному изложению материала.

- Критерии оценки коллоквиума:

оценка «отлично»: ответ полный, правильный, самостоятельный; материал изложен в определенной логической последовательности, демонстрируется многосторонность подходов, многоаспектность обсуждения проблемы, умение находить рациональные пути решения задач, устанавливать причинно- следственные связи, в логическом рассуждении при решении задачи, графических построениях нет ошибок, задача решена рациональным способом с корректным использованием необходимых величин, получен верный ответ. Верные ответы даны на 86-100%

оценка «хорошо»: дан полный, правильный ответ на основе изученных понятий, но допускаются несущественные ошибки. Верные ответы даны на 66-85%.

оценка «удовлетворительно»: дан полный ответ, но при этом есть существенные ошибки указывающие на неумение использовать теоретические знания и умения при решении поставленных задач. Данные пробелы в знаниях не препятствуют дальнейшему обучению. Верные ответы даны на 51-65%

оценка «неудовлетворительно»: ответ обнаруживает незнание основного (порогового) содержания учебного материала Верные ответы даны менее 50%.

Контроль освоения дисциплины и оценка знаний обучающихся на **зачете** производится в соответствии с Положением о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся ДГУ и его филиалов.

оценка «отлично»: ответ полный, правильный, самостоятельный, материал изложен в определенной логической последовательности демонстрируется многосторонность подходов, многоаспектность обсуждения проблемы, умение аргументировать собственную точку зрения, находить пути решения познавательных задач, устанавливать причинно-следственные связи между строением, свойствами и применением веществ, в логическом рассуждении, решении задачи, графических построениях нет

ошибок, задача решена рациональным способом.

оценка «хорошо»: дан полный, правильный, самостоятельный ответ на основе изученных понятий, концепций, закономерностей, но допускаются несущественные ошибки в решении задач.

оценка «удовлетворительно»: дан полный ответ, но при этом есть существенные ошибки указывающие на неумение использовать теоретические знания и умения при решении поставленных задач. Данные пробелы в знаниях не препятствуют дальнейшему обучению.

оценка «неудовлетворительно»: ответ обнаруживает незнание основного (порогвого) содержания учебного материала. менее 50%, уровень не сформирован.

Шкала диапазона для перевода рейтингового балла по дисциплине с учётом итогового контроля в балльную систему.

0 – 50 баллов – «незачтено»;

51 – 100 баллов – «зачтено»

8. Учебно-методическое обеспечение дисциплины.

а) адрес сайта курса:

1. Сайт кафедры прикладной математики ДГУ: http://cathedra.dgu.ru/OfTheDepartment.aspx?id=7

2. Образовательный блог: https://chislen-met.blogspot.com/

б) Основная литература:

1. Машнин Т.С. Современные Java технологии на практике. – Санкт- Петербург: БХВ-Петербург. 2010.

в) Дополнительная литература:

1. Волкова,Т.В. Разработка систем распределенной обработки данных [Электронный ресурс] : учебно-методическое пособие / Волкова Т. В. - Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2012. - 330 с. - Б. ц. Книга находится в базовой версии ЭБС IPRbooks.

2. Головашкин, Д.Л. Методы параллельных вычислений. Ч. 2 : учеб. посо- бие / С.П. Головашкина, Д.Л. Головашкин.— Самара : Издательство СГАУ, 2003. ЭБС Руконт http://rucont.ru/efd/176534

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.

1. Университетская библиотека online : [электронно-библиотечная система] / ООО «ДиректМедиа». — Москва, 2001 — . — URL: http://www.biblioclub.ru — Режим доступа: по подписке. — Текст: электронный

2. eLIBRARY.RU [Электронный ресурс]: электронная библиотека / Науч. электрон. б-ка. — Москва, 1999 – . Режим доступа: http://elibrary.ru/defaultx.– Яз. рус., англ.

3. Электронный каталог НБ ДГУ [Электронный ресурс]: база данных содержит сведения о всех видах лит, поступающих в фонд НБ ДГУ/Дагестанский гос. ун-т. – Махачкала, 2010 – Режим доступа: http://elib.dgu.ru, свободный

4. КонсультантПлюс — студенту и преподавателю : [справочно-правовая система] / ООО Компания «КонсультантПлюс». — Москва, 1997 — . — URL: https://student.consultant.ru/card/ — Режим доступа: для зарегистрир. пользователей. — Текст : электронный

5. Book.ru : электронно-библиотечная система / ООО «КноРус Медиа». — Москва, 2010 — URL: https://www.book.ru/ — Режим доступа: по подписке. — Текст: электронный.

10. Методические указания для обучающихся по освоению дисциплины.

Перечень учебно-методических изданий, рекомендуемых студентам, для подготовки к занятиям представлен в разделе «Учебно-методическое обеспечение. Литература».

Для успешного освоения курса студентам рекомендуется проводить самостоятельный разбор материалов семинарских занятий в течении семестра. В случае затруднений в понимании и освоении каких-либо тем решать дополнительные задания из учебных пособий, рекомендуемых к данному курсу.

Важнейшей задачей учебного процесса в университете является формирование у студента общекультурных и профессиональных компетенций, в том числе способностей к саморазвитию и самообразованию, а также умений творчески мыслить и принимать решения на должном уровне. Выработка этих компетенций возможна только при условии активной учебно-познавательной деятельности самого студента на всём протяжении образовательного процесса с использованием интерактивных технологий.

Такие виды учебно-познавательной деятельности студента как лекции, семинарские занятия и самостоятельная работа составляют систему вузовского образования.

Лекция является главным звеном дидактического цикла обучения в отечественной высшей школе. Несмотря на развитие современных технологий и появление новых методик обучения лекция остаётся основной формой учебного процесса. Она представляет собой последовательное и систематическое изложение учебного материала, разбор какой-либо узловой проблемы. Вузовская лекция ориентирована на формирование у студентов информативной основы для последующего глубокого усвоения материала методом самостоятельной работы, призвана помочь студенту сформировать собственный взгляд на ту или иную проблему.

При изучении дисциплины рекомендуется рейтинговая технология обучения, которая позволяет реализовать комплексную систему оценивания учебных достижений студентов. Текущие оценки усредняются на протяжении семестра при изучении модулей. Комплексность означает учет всех форм учебной и творческой работы студента в течение семестра.

Рейтинг направлен на повышение ритмичности и эффективности самостоятельной работы студентов. Он основывается на широком использовании тестов и заинтересованности каждого студента в получении более высокой оценки знаний по дисциплине.

Рейтинговый балл студента на каждом занятии зависит от его инициативности, качества выполненной работы, аргументированности выступления, характера использованного материала и т.д. Уровень усвоения материала напрямую зависит от внеаудиторной самостоятельной работы, которая традиционно такие формы деятельности, как вы-

полнение письменного домашнего задания, подготовка к разбору ранее прослушанного лекционного материала, подготовка доклада и выполнение реферата.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

Информационные средства обучения: электронные учебники, презентации, технические средства предъявления информации (многофункциональный мультимедийный комплекс) и контроля знаний (тестовые системы). Электронные ресурсы Научной библиотеки ДГУ. Электронно-образовательные ресурсы Дагестанского государственного университета.

Для успешного освоения дисциплины, обучающийся использует следующие программные средства: WINDOWSXP, пакет MSOFFICE.

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

На факультете математики и компьютерных наук Дагестанского государственного университета имеются аудитории, оборудованные интерактивными, мультимедийными досками, проекторами, что позволяет читать лекции в формате презентаций, разработанных с помощью пакета прикладных программ MS Office, использовать наглядные, иллюстрированные материалы, обширную информацию в табличной и графической формах, пакет прикладных обучающих программ, а также электронные ресурсы сети Интернет/