

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук
Кафедра прикладной математики

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
«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

Фонд оценочных средств по дисциплине «Современные методы распределенного хранения и обработки данных» составлена в 2023 году в соответствии с требованиями ФГОС ВО бакалавриата по направлению подготовки 01.03.05 - статистика от 14.08.2020 г. № 1032

Разработчики:

кафедра прикладной математики, Лугуева А.С. к.ф.-м. н., доцент;

Фонд оценочных средств по дисциплине «Современные методы распределенного хранения и обработки данных» одобрен: на заседании кафедры Прикладной математики от «20» января 2023 г., протокол № 5

Зав. кафедрой  Кадиев Р.И.

на заседании Методической комиссии факультета МиКН от «25» января 2023г., протокол №4.

Председатель  Ризаев М.К.

Фонд оценочных средств «Современные методы распределенного хранения и обработки данных» согласован с учебно-методическим управлением

«20» февраля 2023 г. 

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

«Современные методы распределенного хранения и обработки данных»

1.1. Основные сведения о дисциплине

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 академических часов).

Вид работы	Трудоемкость, академических часов		
	7 семестр		всего
Общая трудоёмкость	108		108
Контактная работа:	50		50
Лекции (Л)	32		32
Практические занятия (ПЗ)	18		18
Лабораторные занятия (ЛЗ)			
Консультации			
Промежуточная аттестация (зачет, экзамен)	зачет		
Самостоятельная работа			
1. работа с лекционным материалом, с учебной литературой	8		8
2. опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	8		8
3. выполнение домашних заданий	8		8
4. подготовка к лабораторным работам, к практическим занятиям	8		8
5. подготовка к коллоквиуму	8		8
6. подготовка к контрольным работам	8		8
7. подготовка к зачету	10		10

1.2. Требования к результатам обучения по дисциплине, формы их контроля и виды оценочных средств

ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине «Современные методы распределенного хранения и обработки данных»

№ п/п	Контролируемые модули, разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Оценочные средства		Способ контроля
			наименование	№№ заданий	
1	Модуль 1. Основы распределенных вычислений	ОПК-2 ПК-3	Вопросы для собеседования	1-28	устно
		ОПК-2 ПК-3	Контрольные	1	письменно

			работы		
2	Модуль 2. Технологии распределенного программирования	ОПК-2 ПК-3	Вопросы для собеседования	1-13	устно
		ОПК-2 ПК-3	Контрольные работы	2	письменно
	Модуль 3. Распределенная файловая система	ОПК-2 ПК-3	Вопросы для собеседования	1-14	устно
		ОПК-2 ПК-3	Контрольные работы	3	письменно

1.3. Показатели и критерии определения уровня сформированности компетенций

№ п/п	Код компет енции	Уровни сформированности компетенции			
		Недостаточный	Удовлетворительный (достаточный)	Базовый	Повышенный
		Отсутствие признаков удовлетворительного уровня	Знать: Уметь: Владеть:	Знать: Уметь: Владеть:	Знать: Уметь: Владеть:
1	ОПК-2	Не знает на достаточном уровне основные модели решения функциональных и вычислительных задач, инструментальны е средства для решения прикладных задач. Не умеет на достаточном уровне применять основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач. Не владеет на достаточном уровне основными моделями решения функциональных и	Знает на достаточном уровне основные модели решения функциональны х и вычислительны х задач, инструментальн ые средства для решения прикладных задач. Умеет на достаточном уровне применять основные модели решения функциональных и вычислительных задач, инструментальны е средства для решения прикладных задач. Владеет на достаточном	Знает на хорошем уровне основные модели решения функциональны х и вычислительны х задач, инструментальн ые средства для решения прикладных задач. Умеет на хорошем уровне применять основные модели решения функциональных и вычислительных задач, инструментальны е средства для решения прикладных задач. Владеет на	Знает в совершенстве общую методику статистическог о исследования и способы количественно й формализации объекта наблюдений Умеет в совершенстве применять математический и эконометрическ ий инструментарий для анализа количественных данных, в том числе с применением информационны х систем и технологий Владеет в

		вычислительных задач, инструментальными средствами для решения прикладных задач.	уровне основными моделями решения функциональных и вычислительных задач, инструментальными средствами для решения прикладных задач.	хорошем уровне основными моделями решения функциональных и вычислительных задач, инструментальными средствами для решения прикладных задач.	совершенстве навыками выбора и применения инструментальных средств для обработки количественных данных, навыками интерпретации результатов и формулирования выводов и рекомендаций для подготовки аналитических материалов.
2	ПК-3	Не знает на достаточном уровне общую методику статистического исследования и способы количественной формализации объекта наблюдений. Не умеет на достаточном уровне применить общую методику статистического исследования и способы количественной формализации объекта наблюдений при решении профессиональных задач. Не владеет на достаточном уровне навыками применения общей методики статистического исследования и способы количественной	Знает на достаточном уровне общую методику статистического исследования и способы количественной формализации объекта наблюдений. Умеет на достаточном уровне применить общую методику статистического исследования и способы количественной формализации объекта наблюдений при решении профессиональных задач. Владеет на достаточном уровне навыками применения общей методики статистического исследования и способы	Знает на хорошем уровне общую методику статистического исследования и способы количественной формализации объекта наблюдений. Умеет на хорошем уровне применить общую методику статистического исследования и способы количественной формализации объекта наблюдений при решении профессиональных задач. Владеет на хорошем уровне навыками применения общей методики статистического исследования и способы формализации	Знает на совершенстве общую методику статистического исследования и способы количественной формализации объекта наблюдений. Умеет на достаточном уровне применить общую методику статистического исследования и способы количественной формализации объекта наблюдений при решении профессиональных задач. Владеет на достаточном уровне навыками применения общей методики статистического

		формализации объекта наблюдений при решении прикладных задач.	количественной формализации объекта наблюдений при решении прикладных задач.	объекта наблюдений при решении прикладных задач	исследования и способы количественной формализации объекта наблюдений при решении прикладных задач

2. КОНТРОЛЬНЫЕ ЗАДАНИЯ И ИНЫЕ МАТЕРИАЛЫ ОЦЕНКИ знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения дисциплины «Современные методы распределенного хранения и обработки данных»

**Контрольные работы
Контрольная работа 1**

Практическое задание 1

Электронная таблица как стандартного OLAP-клиента для доступа к данным хранилища.

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

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).

Контрольная работа 2

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 данных. Создание таблиц. Оператор CREATE TABLE. База данных «Книжный магазин» Вставка, удаление и обновление 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. Использование функции count()

Контрольная работа 3

Практическое задание 7.

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

Создание вложенных запросов

Практическое задание 8.

Создание хранимых процедур CREATE PROCEDURE

Создание триггеров с помощью оператора CREATE trigger

Критерии оценки:

- оценка «отлично» выставляется студенту, если верно и правильно выполнено 90%-100% заданий;
- оценка «хорошо» выставляется студенту, если верно и правильно выполнено 70%-80% заданий;
- оценка «удовлетворительно» выставляется студенту, если верно и правильно решено 50%-60% заданий, возможны некоторые исправления при решении;
- оценка «неудовлетворительно» выставляется студенту, если верно выполнено менее 50% заданий;

Вопросы для коллоквиумов, собеседования

Модуль 1. Основы распределенных вычислений

1. Многопроцессность (concurrency), асинхронность и параллелизм.
2. Процессы и потоки, способы взаимодействия.
3. Основы многопоточного программирования с общей памятью на примере языка Java.
4. Синхронизация доступа к общим данным, взаимное исключение, блокировка. Взаимная блокировка.
5. Условная синхронизация потоков. Циклический барьер. Конфигурируемый пул потоков.
6. Модель взаимодействия процессов, основанная на асинхронном обмене сообщениями, на примере языка Scala.
7. Цели применения параллельных вычислений. Пути достижения параллелизма.
8. Классификация вычислительных систем.
9. Системы с общей разделяемой памятью.
10. Системы с распределенной памятью. Модель параллельного алгоритма, граф “операции-операнды”, расписание вычислений.
11. Теоретические оценки времени выполнения параллельного алгоритма.
12. Характеристики параллельных алгоритмов: ускорение, эффективность, стоимость.
13. Теоретические оценки максимально достижимого ускорения, законы Амдала и

Густавсона-Барсиса.

14. Масштабируемость параллельного алгоритма, функция изоэффективности.
15. Оценка коммуникационной сложности параллельных алгоритмов.
16. Декомпозиция исходной задачи на подзадачи. Типовые структуры алгоритмов.
17. Функциональный параллелизм. Парадигма “разделяй и властвуй”.
18. Геометрическая деком- позиция. Алгоритмы на рекурсивных структурах данных. Конвейерная обработка.
19. Координация на основе событий. Анализ информационных зависимостей между подзадачами.
20. Выбор вычислительной платформы
21. Примеры параллельных алгоритмов: матричные вычисления, сортировка, алгоритмы на графах, задача N тел, рендеринг изображений.
22. основополагающие принципы MapReduce иллюстрируются на примерезадачи построения инвертированного индекса.
23. Архитектура платформы ApacheHadoop и ее основные элементы – распределенная файловая система HDFS и реализация модели MapReduce.
24. Интерфейс прикладного программирования и общие принципы реализации при-ложений для Hadoop на языке Java на примере задачи подсчета частотывстречаемости слов в тексте.
25. Процесс установки Hadoop на локальной машине, с последующей компиляцией и запуском приложения.
26. Принципы реализации функций map и reduce на языках, отличных от Java. Работе с данными и запуск приложений на учебном кластере, выбору значений параметров запуска заданий и отладке приложений.
- 27 . Реализации процедуры кластерного анализа.
28. Алгоритм canopyclustering, позволяющий снизить сложность вычислений при анализе данных большого объема и размерности. Реализация алгоритмасanopyclustering в рамках модели MapReduce на примере задачи кластериза- ции фильмов из набора данных NetflixPrize.

Модуль 2. Технологии распределенного программирования

1. Цели построения и виды распределенных систем.
2. Способы взаимодействия распределенных процессов. Обмен сообщениями. Удаленные вызовы процедур.
3. Пространство кортежей. Распределенная общая память.
4. Современные технологии распределенного программирования: Java RMI.
5. Проблемы построения распределенных систем, способы повышения отказоустойчивости.
6. История создания. MapReduce в Google (закрытая реализация на C++).
7. Основы MapReduce на примере задачи построения инвертированного индек- са.
8. Примеры использования MapReduce. Особенности MapReduce в Hadoop – открытая реализация на Java. Функция Map.
9. Функция Reduce. Списки ключ- значение.
10. Принципы распределенной реализации MapReduce на кластерных системах. Недостатки MapReduce.
11. Понятие и определение BigData. Что такое технология Hadoop. Применение Hadoop.
12. Установка и настройка простого кластера. Настройка. Запуск. Проверка файловой системы HDFS. Тестирование HDFS.
13. Конфигурации Hadoop для распределенной работы. Преимущества решений на базе Hadoop.

Модуль 3. Распределенная файловая система

1. Клиент HDFS. Сервер файлов контрольных точек (CheckpointNode),сервер резервных

- копий (BackupNode).
2. Обновления и снимки файловой системы. Чтение и запись файлов Размещение блоков. Управление репликацией.
 3. Балансировщик. Сканер блоков. Копирование данных между кластерами.
 4. Практическое использование файловой системы
 5. Долговечность хранения данных. Возможности совместного использования ресурсов HDFS. Масштабирование и объединение файловой системы.
 6. Дистрибутивы Hadoop. Аппаратные требования. Установка и настройка.
 7. Пример кластера ClouderaHadoop. Потребности приложений Hadoop.
 8. Работа с HDFS из Java. Запуск программы. Разделение Hadoop на главный и подчиненные узлы.
 9. Конфигурация кластера Hadoop. Обновление конфигурации Hadoop. Записи для узлов Hadoop в файле.
 10. Определение хозяина HDFS в файле core-site.xml Запуск демонов MapReduce. Получение информации о запущенных процессах на одном из подчиненных узлов. Тестирование HDFS.
 11. Проверка файловой системы HDFS. Параллельная обработка больших массивов данных.
 12. Распределенные хранилища данных. Распределенное программирование. Примеры
 13. Текущее состояние в области распределенной обработки данных, примеры технологий, задач и объемов обрабатываемых данных.
 14. Обзор актуальных задач в области распределенной обработки данных.

Критерии оценки:

- оценка «отлично» выставляется студенту, если изложение полученных знаний в устной форме полное, в системе, в соответствии с требованиями учебной программы; допускаются единичные несущественные ошибки, самостоятельно исправляемые учащимися;
- оценка «хорошо» выставляется студенту, если изложение полученных знаний в устной форме полное, в системе, в соответствии с требованиями учебной программы; допускаются, отдельные несущественные ошибки, исправляемые учащимися после указания преподавателя на них;
- оценка «удовлетворительно» выставляется студенту, если изложение полученных знаний неполное, однако это не препятствует усвоению последующего программного материала; допускаются отдельные существенные ошибки, исправляемые с помощью преподавателя;
- оценка «неудовлетворительно» выставляется студенту, если изложение учебного материала неполное, бессистемное, что препятствует усвоению последующей учебной информации; существенные ошибки, не исправляемые даже с помощью преподавателя;

Вопросы к зачету

1. Основные понятия распределённой обработки данных
2. Условия и требования к распределенной обработке данных
3. Основные принципы создания и функционирования распределённых баз данных
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 и название 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: ФИО Инспектора 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. Metallургическая, 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: ФИО Инспектора 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. Metallургическая, 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. Metallургическая, 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 его Адрес».

Критерии оценки:

- «зачтено» выставляется студенту, если изложение полученных знаний в устной форме полное, в системе, в соответствии с требованиями учебной программы; допускаются, отдельные несущественные ошибки, исправляемые учащимися после указания преподавателя на них;
- «не зачтено» выставляется студенту, если изложение учебного материала неполное, бессистемное, что препятствует усвоению последующей учебной информации; существенные ошибки, не исправляемые даже с помощью преподавателя.

Рекомендуемые границы оценок:

- «зачтено» - не менее 51% правильных ответов,
- «не зачтено»- менее 51% правильных ответов.