

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«24» июня 2025 г.
Протокол №2

**Рабочая программа дисциплины (модуля)
«Хранилища данных»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Программа двух дипломов НИУ
ВШЭ и ЦУ «Прикладная математика и информатика»

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2025

**Москва
2025**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения	5
3. Тематический план	7
4. Содержание дисциплины (модуля)	7
5. Учебно-методическое обеспечение	8
6. Материально-техническое обеспечение	8
7. Методические и оценочные материалы	10

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Хранилища данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Хранилища данных» важно для эффективного хранения, интеграции и анализа больших объемов информации, что позволяет принимать обоснованные управленческие решения. Это способствует оптимизации бизнес-процессов и повышению конкурентоспособности организаций за счет быстрого доступа к качественным данным.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений.

Дисциплина (модуль) является выборной и доступна для изучения на 3 или 4 курсе в 5, 6, 7 семестрах на выбор.

Цель изучения дисциплины (модуля): освоение методов проектирования, создания и управления системами хранения и анализа больших объемов данных для поддержки принятия решений.

Задачи изучения дисциплины (модуля):

— изучить архитектурные основы хранилищ данных, включая модели данных, компоненты (такие как ETL-процессы, OLAP-кубы и метаданные) и принципы интеграции разнородных источников информации;

— освоить практические навыки проектирования и реализации хранилищ данных с использованием современных технологий, инструментов и стандартов, таких как SQL, BI-платформы и облачные решения;

— понять методы анализа и визуализации данных в хранилищах, включая OLAP-операции, отчетность и применение машинного обучения для извлечения инсайтов из больших данных.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— базовую архитектуру реляционной СУБД (хранение данных, индекс, планировщик, исполнение, транзакции);

— назначение страниц на диске и буферного пула;

— принцип работы B+-дерева как индекса;

— общий процесс обработки SQL-запроса: парсинг → план → исполнение;

— основную идею журнала WAL для сохранности данных;

уметь:

— реализовать простое файловое хранилище страниц и базовый буферный кэш на Java;

— написать минимальный B+-индекс и выполнить поиск/вставку;

— разобрать простой SELECT-запрос в AST с помощью готового парсера (ANTLR);

— построить и запустить элементарный план сканирования таблицы и фильтрации;

— логировать изменения страниц в WAL и восстановить данные после сбоя;

владеть:

- практикой тестирования и отладки компонентов собственной мини-СУБД;
- навыками интеграции модулей (storage, индекс, исполнитель) в рабочий прототип;
- базовыми методами измерения производительности и поиска узких мест.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области разработки, основные принципы критической оценки источников информации и их релевантности
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
ОПК-1.	Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности	ОПК-1.1.	Знает основные концепции и теории в области математического анализа и смежных дисциплин; методы и подходы, используемые в различных областях математики
		ОПК-1.2.	Умеет применять математические методы для решения профессиональных задач
		ОПК-1.3.	Имеет практический опыт разработки и реализации математических моделей в профессиональной деятельности
ОПК-4.	Способен находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем	ОПК-4.1.	Знает базовые основы современного математического аппарата, связанного с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности

		ОПК-4.2.	Умеет использовать этот математический аппарат в профессиональной деятельности
		ОПК-4.3.	Имеет практический опыт применения современного математического аппарата, связанного с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности
ОПК-6.	Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения	ОПК-6.1.	Знает алгоритмы разработки, компьютерные программы, а также алгоритмы вычислительной математики
		ОПК-6.2.	Умеет разрабатывать математические программные продукты и комплексы с использованием современных технологий программирования
		ОПК-6.3.	Имеет практический опыт разработки интеллектуальных информационных систем для визуализации результатов исследований
ПК-3.	Способен использовать методы математического и алгоритмического моделирования при решении теоретических и прикладных задач	ПК-3.1.	Знает основные методы математического и алгоритмического моделирования, а также их применение для решения теоретических и прикладных задач
		ПК-3.2.	Умеет разрабатывать и применять математические модели и алгоритмы для решения различных задач, анализируя полученные результаты
		ПК-3.3.	Имеет практический опыт использования методов математического и алгоритмического моделирования в реальных проектах или исследованиях

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		<i>Очная форма</i>				
		Контактная работа		Контроль	Самостоятельная работа	
Лекции	Семинары (практические занятия)					
1	Хранилище данных	4	4		18	Кейс
2	Каталог и организация данных	4	4		18	Кейс
3	Буферизация и индексы	4	4		18	Кейс
4	Обработка запросов	4	4		18	Кейс
5	Транзакции и восстановление	4	4		20	Кейс
6	Сервер и инструменты	4	4		20	Коллоквиум
7	Проект	4	4		18	Проект
	<i>Зачет с оценкой</i>			4		
	Итого:	28	28	4	130	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	190				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Хранилище данных	Архитектура СУБД, страницы и файловое хранилище. Управление страницами и кэш страниц
2	Каталог и организация данных	Система каталогов метаданных. Неар-таблицы и последовательное сканирование
3	Буферизация и индексы	Буферный менеджер. Индексация: В+-дерево
4	Обработка запросов	SQL-парсинг и построение AST. Планировщик запросов. Движок выполнения запросов. Join-ы и сортировки
5	Транзакции и восстановление	Транзакции и уровни изоляции. WAL и восстановление данных
6	Сервер и инструменты	Серверный слой и сетевой протокол. CLI-интерфейс и утилиты
7	Проект	Интеграция и защита

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Рогов, Е.В. PostgreSQL 16 изнутри : практическое руководство / Е. В. Рогов. – Москва : ДМК Пресс, 2024. - 666 с. – ISBN 978-5-93700-305-8. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2205083>.

2. Карпова, И. П. Базы данных : учебное пособие / И. П. Карпова. - Санкт-Петербург : Питер, 2021. - 240 с. - (Серия «Учебное пособие»). - ISBN 978-5-4461-9681-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1857026>.

3. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2025. — 97 с. — (Высшее образование). — ISBN 978-5-534-21173-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559502>.

4. Советов, Б. Я. Базы данных : учебник для вузов / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. — 4-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 403 с. — (Высшее образование). — ISBN 978-5-534-18479-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559898>.

Дополнительная литература:

1. Советов Б. Я. Базы данных : учебник для вузов / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. — 4-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 403 с. — (Высшее образование). — ISBN 978-5-534-18479-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559898>.

2. Нестеров С. А. Базы данных : учебник и практикум для вузов / С. А. Нестеров. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 258 с. — (Высшее образование). — ISBN 978-5-534-18107-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560753>.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной

мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное

Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Хранилища данных» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, кейсы, коллоквиум, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар — это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Кейс – практическая работа студентов над реальными или смоделированными

задачами, что позволяет студенту применять теоретические знания на практике.

Студент самостоятельно разрабатывает стратегию решения поставленной задачи, что способствует развитию навыков критического мышления и самостоятельного принятия решений. Такой подход помогает подготовить будущих специалистов к реальным вызовам в их профессиональной деятельности.

Коллоквиум – устные ответы на вопросы, список которых известен студенту заранее.

В процессе подготовки к коллоквиуму необходимо проанализировать учебные материалы, ознакомившись с лекциями, учебниками и дополнительными источниками, акцентируя внимание на ключевых темах. Рекомендуется создать структурированные конспекты, выделяя основные идеи, термины и формулы.

Проект – исследовательская работа по курсу и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов, планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Хранилища данных»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета с оценкой*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Хранилища данных» оценивается следующим образом:

Активность	Вес	Количество	Описание
Кейс	70%	11	Практическая работа студентов над реальными или смоделированными задачами, что позволяет студенту применять теоретические знания на практике
Коллоквиум	30%	1	Устные ответы на вопросы, список которых известен студенту заранее

Формула расчёта итоговой оценки по дисциплине (модулю) «Хранилища данных»: $0,7 \times$ среднее за кейсы $+ 0,3 \times$ за коллоквиум».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные задания для кейсов

Кейс 1: Хранилище данных — Архитектура СУБД, страницы и файловое хранилище. Управление страницами и кэш страниц

Сценарий: Вы — разработчик СУБД, и вам нужно оптимизировать хранение данных для приложения с большим объемом транзакций. Система использует страничную организацию данных с кэшированием страниц. Вам предстоит проанализировать архитектуру, управлять страницами и реализовать кэш для улучшения производительности.

- Объясните архитектуру СУБД на уровне хранения данных.** Опишите роль страниц (pages) в файловом хранилище, включая их структуру (заголовки, данные, свободное пространство) и как они организуют файлы на диске. Приведите пример диаграммы простой страницы с 4 КБ данными.
- Опишите управление страницами в СУБД.** Как происходит выделение, чтение и запись страниц? Напишите псевдокод функции, которая загружает страницу из файла в память, проверяет ее валидность по контрольной сумме и обновляет метаданные страницы.
- Что такое кэш страниц и зачем он нужен?** Объясните механизм LRU (Least Recently Used) для замещения страниц в кэше. Напишите код на C (или псевдокод), имитирующий кэш страниц: структуру данных (например, хэш-таблица + список) и функцию для чтения страницы с попаданием в кэш или загрузкой с диска.
- Как взаимодействует кэш страниц с файловым хранилищем?** Опишите политику write-back и ее преимущества для производительности. Приведите практический пример: симулируйте запись данных в страницу кэша и отложенную синхронизацию с диском, измеряя время операции.
- Проектируйте улучшение для управления страницами.** Предложите оптимизацию кэша страниц для системы с высокой нагрузкой на чтение (например, добавление prefetching). Напишите псевдокод алгоритма, который предзагружает соседние страницы при чтении одной, и объясните, как это снижает задержки.

Кейс 2: Каталог и организация данных — Система каталогов метаданных. Неартаблицы и последовательное сканирование

Сценарий: В вашей команде разрабатывается простая СУБД для хранения метаданных о файлах. Вы работаете с системой каталогов, hear-таблицами (неупорядоченными

таблицами) и последовательным сканированием для поиска данных. Вам нужно реализовать и оптимизировать эти компоненты.

1. **Объясните систему каталогов метаданных в СУБД.** Как каталоги хранят информацию о таблицах, индексах и других объектах? Опишите структуру каталога (например, как таблица с записями) и приведите диаграмму иерархии каталогов для базы данных с двумя схемами.
2. **Что такое heap-таблицы и как они организуют данные?** Опишите преимущества и недостатки heap-таблиц по сравнению с индексированными. Напишите псевдокод создания heap-таблицы: выделение страниц, вставка записей без сортировки и обновление метаданных каталога.
3. **Опишите последовательное сканирование (sequential scan) в heap-таблице.** Как оно работает и когда применяется? Напишите код на SQL-подобном языке (или псевдокод), который выполняет полное сканирование таблицы для поиска записей по условию (например, `SELECT * FROM table WHERE column > value`), и объясните его эффективность.
4. **Как интегрируется система каталогов с heap-таблицами?** Опишите процесс создания новой таблицы: обновление каталога, выделение heap-страниц и запись метаданных. Приведите практический пример: код на C (псевдокод), который создает таблицу в каталоге и вставляет данные в heap-структуру.
5. **Оптимизируйте последовательное сканирование.** Предложите улучшение для heap-таблиц с большим объемом данных (например, добавление зонального сканирования). Напишите псевдокод модифицированного алгоритма сканирования, который пропускает нерелевантные страницы, и сравните производительность с базовым sequential scan.

Примерные задания для коллоквиума

1. Опишите архитектуру СУБД на уровне хранения данных. Объясните роль страниц (pages) в файловом хранилище, включая их структуру (заголовки, данные, свободное пространство). Приведите диаграмму простой страницы размером 4 КБ с примерами полей.
2. Как происходит управление страницами в СУБД? Напишите псевдокод функции, которая загружает страницу из файла в память, проверяет её валидность по контрольной сумме и обновляет метаданные страницы.
3. Что такое кэш страниц и зачем он нужен? Объясните механизм LRU (Least Recently Used) для замещения страниц. Напишите псевдокод структуры данных для кэша (хэш-таблица + список) и функции чтения страницы с попаданием в кэш или загрузкой с диска.
4. Как взаимодействует кэш страниц с файловым хранилищем? Опишите политику write-back и её преимущества. Приведите пример симуляции записи данных в страницу кэша и отложенной синхронизации с диском, оценив влияние на производительность.
5. Объясните систему каталогов метаданных в СУБД. Как каталоги хранят информацию о таблицах и индексах? Опишите структуру каталога как таблицы с записями и приведите диаграмму иерархии для базы данных с двумя схемами.
6. Что такое heap-таблицы и как они организуют данные? Сравните их преимущества и недостатки с индексированными таблицами. Напишите псевдокод создания heap-таблицы: выделение страниц, вставка записей и обновление каталога.
7. Опишите последовательное сканирование (sequential scan) в heap-таблице. Когда оно применяется? Напишите SQL-подобный код (или псевдокод) для полного

сканирования таблицы с условием (например, `SELECT * FROM table WHERE column > value`) и объясните его эффективность.

8. Как интегрируется система каталогов с heap-таблицами? Опишите процесс создания новой таблицы. Напишите псевдокод на C, который создаёт таблицу в каталоге и вставляет данные в heap-структуру.

9. Что такое буферный менеджер в СУБД и как он связан с кэшем страниц? Объясните его роль в управлении памятью. Приведите диаграмму взаимодействия буферного менеджера с диском и памятью.

10. Опишите структуру B+-дерева как индекса. Как происходит вставка и поиск ключа? Напишите псевдокод операции вставки в B+-дерево с учётом балансировки.

11. Как буферный менеджер оптимизирует работу с B+-деревом? Объясните, как страницы дерева кэшируются. Приведите пример политики замещения страниц в буфере для индекса с высокой нагрузкой на чтение.

12. Сравните B+-дерево с другими индексами (например, хэш-индексом). Напишите псевдокод поиска по B+-дереву для диапазонного запроса и объясните его преимущества для больших наборов данных.

13. Опишите процесс SQL-парсинга и построения AST. Как парсер преобразует запрос в дерево? Приведите пример AST для простого запроса `SELECT` с `JOIN` и диаграмму дерева.

14. Что делает планировщик запросов и как он выбирает план? Объясните роль оптимизации (например, выбор индексов). Напишите псевдокод алгоритма планировщика для запроса с `JOIN`, оценивающего стоимость планов.

15. Как работает движок выполнения запросов? Опишите этапы выполнения плана. Приведите пример выполнения запроса с сортировкой: псевдокод итератора для сортировки данных в памяти.

16. Объясните алгоритмы `JOIN` (например, `nested loop`, `hash join`). Когда использовать каждый? Напишите псевдокод `hash join` для двух таблиц и объясните его эффективность по сравнению с `nested loop`.

17. Что такое транзакции и уровни изоляции в СУБД? Опишите уровни (`Read Uncommitted`, `Serializable`) с примерами аномалий. Приведите диаграмму, иллюстрирующую проблему "dirty read".

18. Как работает WAL (`Write-Ahead Logging`) для восстановления? Объясните его роль в ACID. Напишите псевдокод записи транзакции в WAL и восстановления после сбоя.

19. Опишите процесс восстановления данных с использованием WAL. Какие типы сбоев он покрывает? Приведите пример сценария восстановления базы данных после внезапного отключения питания.

20. Сравните уровни изоляции по производительности и надёжности. Напишите псевдокод реализации блокировки для уровня `Serializable` и объясните, как это предотвращает аномалии.

Примерное описание и критерии оценивания к итоговому проекту

Описание проекта:

Вам необходимо разработать и реализовать полнофункциональное решение на базе СУБД PostgreSQL, демонстрирующее глубокое понимание архитектуры, возможностей и инструментов работы с данными. Проект должен включать создание и настройку базы данных, реализацию бизнес-логики с использованием процедур и функций PL/pgSQL, применение механизмов безопасности и разграничения доступа, а также предусматривать методы резервного копирования и восстановления данных. Особое внимание уделяется правильному использованию принципов изоляции и многоверсионности, оптимальному управлению ресурсами и интеграции с внешними приложениями.

Критерии оценивания:

1. Установка и настройка PostgreSQL

- Корректность и полнота установки сервера PostgreSQL.
 - Настройка среды для эффективной работы (конфигурационные файлы, параметры).
 - Умение работать с psql и базовыми командами управления.
- 2. Архитектура и управление данными**
- Правильная организация структуры базы данных с учетом физических и логических аспектов.
 - Применение принципов MVCC и уровней изоляции для обеспечения согласованности данных.
 - Использование механизмов буферного кеша и журналирования для повышения производительности и надежности.
- 3. Разработка бизнес-логики**
- Создание и использование функций, процедур и триггеров на PL/pgSQL.
 - Эффективное применение курсоров, массивов и динамических запросов.
 - Корректная обработка ошибок и отладка кода.
 - Использование составных типов данных и продвинутых возможностей SQL.
- 4. Безопасность и разграничение доступа**
- Создание и настройка ролей с соответствующими правами доступа.
 - Реализация механизмов аутентификации и авторизации.
 - Ограничение доступа к данным на уровне таблиц и/или столбцов.
- 5. Интеграция и резервное копирование**
- Организация взаимодействия базы данных с внешними приложениями (например, через клиентские библиотеки).
 - Реализация стратегии резервного копирования и восстановления данных.
 - Демонстрация понимания логического резервирования и восстановления на практике.
- 6. Документация и презентация**
- Наличие подробной документации по проекту, описывающей архитектуру, используемые технологии и инструкции по запуску.
 - Качество и полнота описания бизнес-логики и технических решений.
 - Умение представить и аргументировать выполненную работу.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Укажите базовый элемент хранения данных в СУБД.	Страница / Страница данных (варианты: страница, page)	УК-1
2	Укажите метод управления памятью для страниц в СУБД.	Кэш страниц / Кэш страниц (варианты: кэш страниц, page cache)	УК-1
3	Укажите структуру для хранения метаданных в СУБД.	Система каталогов / Система каталогов (варианты: каталог, catalog system)	УК-1
4	Укажите тип таблицы для неупорядоченного хранения данных.	Heap-таблица / Heap-таблица (варианты: heap table, кучи таблица)	УК-1
5	Укажите математическую модель для оценки производительности кэша.	LRU / LRU (варианты: LRU, least recently used)	ОПК-1
6	Укажите вероятностный метод в буферном менеджере.	FIFO / FIFO (варианты: FIFO, first in first out)	ОПК-1
7	Укажите структуру данных для эффективной индексации.	B+-дерево / B+-дерево (варианты: B+-tree, B+ дерево)	ОПК-1

8	Укажите алгоритм парсинга SQL-запросов.	Построение AST / Построение AST (варианты: AST, abstract syntax tree)	ОПК-4
9	Укажите метод оптимизации плана запроса.	Планировщик запросов / Планировщик запросов (варианты: query planner, оптимизатор запросов)	ОПК-4
10	Укажите операцию для объединения таблиц в движке выполнения.	Join / Join (варианты: join, соединение)	ОПК-4
11	Укажите алгоритм для сортировки в обработке запросов.	Внешняя сортировка / Внешняя сортировка (варианты: external sort, внешняя сортировка)	ОПК-4
12	Укажите механизм логирования для транзакций.	WAL / WAL (варианты: WAL, write-ahead logging)	ОПК-6
13	Укажите уровень изоляции для предотвращения фантомных чтений.	Serializable / Serializable (варианты: serializable, серийный)	ОПК-6
14	Укажите метод восстановления данных после сбоя.	Point-in-time recovery / Point-in-time recovery (варианты: PITR, восстановление в точку времени)	ОПК-6
15	Укажите протокол для сетевого взаимодействия сервера СУБД.	TCP/IP / TCP/IP (варианты: TCP/IP, сетевой протокол)	ОПК-6
16	Укажите инструмент для командной строки в СУБД.	CLI / CLI (варианты: CLI, command-line interface)	ОПК-6
17	Укажите модель для heap-таблиц в организации данных.	Последовательное сканирование / Последовательное сканирование (варианты: sequential scan, seq scan)	ПК-3
18	Укажите алгоритмическая модель для B+-дерева.	Балансированное дерево / Балансированное дерево (варианты: balanced tree, B+ structure)	ПК-3
19	Укажите метод моделирования транзакций с изоляцией.	MVCC / MVCC (варианты: MVCC, multiversion concurrency control)	ПК-3
20	Укажите подход к моделированию серверного слоя.	Утилиты администрирования / Утилиты администрирования (варианты: admin tools, утилиты)	ПК-3