ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол №2

Рабочая программа дисциплины (модуля)
«Хранилища данных»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Дизайн и разработка информационных продуктов*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2026*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2026

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 5
4. **Содержание дисциплины (модуля)** 8
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

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 Дизайн и разработка информационных продуктов, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Хранилища данных» позволяет разрабатывать эффективные и масштабируемые системы хранения данных, глубоко понимая, как устроены СУБД изнутри, что критически важно для решения задач data engineering в условиях роста объемов и требований к производительности. Полученные знания и навыки лежат в основе построения надежных, оптимизированных и легко сопровождаемых аналитических систем.

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 Дизайн и разработка информационных продуктов и входит в обязательную часть Блока 1.

Дисциплина (модуль) изучается на 2 курсе в 4 семестре.

Цель изучения дисциплины (модуля): сформировать у обучающихся системное понимание внутреннего устройства реляционных СУБД, от обработки SQL-запросов до физического хранения данных и управления транзакциями, для эффективного проектирования, анализа и оптимизации хранилищ данных.

Задачи изучения дисциплины (модуля):

- освоить архитектурные компоненты СУБД и проследить путь SQL-запроса от парсинга до выполнения, включая работу анализаторов, оптимизатора и исполнителя;
- изучить принципы физического хранения данных, включая организацию страниц, слотов и layout кортежей, а также различия представления NULL и пустых значений на уровне хранилища;
- понять механизмы управления буферным пулом, включая политики кэширования и вытеснения страниц, а также роль dirty pages и чекпоинтов в обеспечении производительности и надежности;
- научиться интерпретировать планы выполнения запросов, связывая их с физическим доступом к данным, использованием индексов и оценками стоимости операций на основе статистики;
- овладеть навыками диагностики узких мест в работе СУБД и проектирования решений, демонстрирующих ключевые механизмы хранения и обработки данных.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- как Java обеспечивает кроссплатформенность: роль JVM, байткода, спецификаций и верификации;
- как SQL-запрос проходит по модулям СУБД: parser → semantic analyzer → planner/optimizer → executor → buffer/page/IO;
- физическое хранение: страницы, слоты, tuple layout, nullmask; различие NULL и "" на уровне хранения;
- буферный менеджер: buffer pool, cache hit/miss, dirty pages; политика вытеснения (FIFO/LRU/Clock), clock-sweep в PostgreSQL;

— системные каталоги (pg_class/pg_attribute и т.п.) и принцип самоописания БД; взаимодействие catalog manager с page/buffer;
— парсинг SQL: lexer/parser/semantic analysis; AST vs QueryTree; приоритет операторов;
— планирование и исполнение: логический/физический план, Volcano/iterator модель, базовые исполнители;
— базовые алгоритмы: Nested Loop Join без индексов; хеш-индексы и B+-tree; основы cost model, статистика (ANALYZE), EXPLAIN/EXPLAIN ANALYZE;
— транзакции: MVCC, WAL, lock manager, deadlocks; путь запроса внутри транзакции;
— клиент–сервер: сокеты, postmaster/backend, жизненный цикл соединения, сессия vs транзакция;

уметь:

— объяснять по симптомам (план, статистика, нагрузка), какие механизмы СУБД стали бутылочным горлышком;
— интерпретировать планы выполнения и связывать их с физическим доступом к данным и индексами;
— сравнивать варианты исполнения (scan/join/index) и аргументировать выбор через cost model и устройство хранения;
— разбирать поведение буферного пула и причины появления dirty pages/чекпоинтов;

владеть:

— навыками проведения инженерных разборов работы запросов «от SQL до диска»;
— навыками проектирования и реализации (в рамках проекта) прототипов/утилитов, работающих с данными и демонстрирующих ключевые механизмы хранения;
— навыками использования EXPLAIN ANALYZE и базовых диагностических приёмов для поиска проблем производительности.

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 искусственного интеллекта, 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
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач

		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
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 реализация математических моделей в профессиональной деятельности
ПК-2.	Способен решать типовые задачи профессиональной деятельности, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной

			деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области профессиональной деятельности, включая соблюдение норм информационной безопасности
ПК-4.	Способен разрабатывать программное обеспечение для решения прикладных задач в области профессиональной деятельности под руководством более опытного специалиста	ПК-4.1.	Знает основные принципы разработки программного обеспечения и методы решения прикладных задач в сфере профессиональной деятельности
		ПК-4.2.	Умеет разрабатывать программное обеспечение под руководством специалистов более высокой категории, используя современные технологии и инструменты
		ПК-4.3.	Имеет опыт участия в разработке программного обеспечения для решения прикладных задач в команде под руководством более опытных специалистов

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы					ТКУ (текущий контроль успеваемости)
		Очная форма					
		Контактная работа			Контроль	Самостоятельная работа	
		Лекции	Семинары (практические занятия)	Консультации			
1	Внутренняя архитектура СУБД и хранение данных	3	3	2		11	Домашние задания
2	Управление памятью и метаданными	3	3	2		11	Домашние задания
3	Разбор и представление SQL-запросов	4	4	2	6	11	Домашние задания Коллоквиум
4	Исполнение запросов без индексов	4	4	2		11	Домашние задания
5	Индексные структуры доступа	4	4	2		11	Домашние задания
6	Планирование и оптимизация запросов	4	4	2		11	Домашние задания
7	Транзакции и клиент–серверное взаимодействие	4	4	2		12	Домашние задания Проект
	Экзамен				2		
Итого:		***26***	***26***	***14***	***8***	***78***	
Объем дисциплины (модуля) (в ак. ч.)		***152***					
Объем дисциплины (модуля) (в зач. ед.)		***4***					

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Внутренняя архитектура СУБД и хранение данных	Архитектура СУБД и путь SQL-запроса Физическое представление данных
2	Управление памятью и метаданными	Буферизация и доступ к данным Управление памятью и устойчивость данных
3	Разбор и представление SQL-запросов	Системные каталоги и метаданные Разбор и анализ SQL-запросов
4	Исполнение запросов без индексов	Модели исполнения запросов Алгоритмы соединения данных
5	Индексные структуры доступа	Индексные методы доступа Деревья поиска и диапазонные запросы
6	Планирование и оптимизация запросов	Планирование и оптимизация выполнения
7	Транзакции и клиент–серверное взаимодействие	Транзакции, соединения и защита проекта

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. Карау, Х. Изучаем Spark. Молниеносный анализ данных : практическое руководство / Х. Карау, Э. Конвински, П. Венделл, З. Матей ; пер. с англ. - 2-е изд. - Москва : ДМК Пресс, 2023. - 305 с. - ISBN 978-5-89818-320-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102607.

2. Харенслак, Б. Apache Airflow и конвейеры обработки данных : практическое руководство / Б. Харенслак, Д. дэ Руйтер ; пер. с англ. Д. А. Беликова. - Москва : ДМК Пресс, 2022. - 502 с. - ISBN 978-5-97060-970-5. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2155905.

3. Apache Kafka. Потоковая обработка и анализ данных : практическое руководство / Г. Шапира, Т. Палино, Р. Сиварам, К. Петти. - 2-е изд. - Санкт-Петербург : Питер, 2023. - 512 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-2288-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2123357.

Дополнительная литература:

1. 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.

2. Кэмпбелл, Л. Базы данных. Инжиниринг надежности : практическое руководство / Л. Кэмпбелл, Ч. Мейджорс. - Санкт-Петербург : Питер, 2020. - 304 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-1310-1. - Текст : электронный. - URL: https://znanium.com/catalog/product/1756135.

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, коллоквиумы, домашние задания, 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 удается, 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.

Консультации – структурированные встречи, на которых преподаватели предоставляют индивидуальную или групповую помощь в освоении учебного материала, обсуждении вопросов и решении проблем, возникающих в процессе обучения.

Консультации могут включать разъяснение сложных тем, подготовку к экзаменам и помощь в выполнении проектных работ, что способствует более глубокому пониманию предмета и улучшению академической успеваемости.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

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 дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту рекомендуется: четко определить цели и задачи проекта; составить план работы, разбив проект на этапы с указанием сроков выполнения каждого из них; использовать разнообразные источники информации и инструменты для исследования темы; регулярно проверять прогресс и вносить коррективы в план, если это необходимо.

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 ***экзамена***, 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**	**Общая характеристика результата обучения по дисциплине (модулю)**
10	Отлично	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 моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты
9	Отлично	
8	Отлично	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		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	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	
5	Удовлетворительно	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	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

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

Активность	Вес	Количество	Описание
Домашние задания	20%	7	Набор задач по темам недели
Коллоквиум	30%	1	Письменная работа с набором задач, которые нужно решить за ограниченное время
Проект	20%	1	Исследовательская работа по дисциплине (модулю) и презентация результатов
Экзамен	30%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Хранилища данных» «0,2 × среднее за домашние задания + 0,3 × коллоквиум + 0,2 × проект + 0,3 × экзамен».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание по теме «Внутренняя архитектура СУБД и хранение данных».

Напишите краткий технический отчёт (2–3 стр.), в котором опишите путь выполнения простого SELECT * FROM employees WHERE id = 100; в PostgreSQL: от приёма запроса до возврата результата, с указанием этапов (парсер, анализатор, оптимизатор, исполнитель). Дополните отчёт схематичным описанием физического хранения строки в таблице employees: структура страницы, слота, кортежа (tuple), размещение NULL-значений и отличие NULL от пустой строки. Приведите пример pageinspect-запросов для анализа содержимого страницы таблицы.

Домашнее задание по теме «Управление памятью и метаданными».

Настройте PostgreSQL-экземпляр с малым shared_buffers (например, 16 МБ) и выполните серию последовательных и случайных SELECT по большой таблице. С помощью pg_stat_bgwriter и pg_stat_database проанализируйте показатели buffers_hit, buffers_read, checkpoint_write_time. Опишите, как меняется cache hit ratio, объясните появление dirty pages и частоту чекпоинтов. Сравните эффективность политик LRU и Clock-sweep (на основе документации и логики), объясните, почему PostgreSQL использует clock-sweep. Предложите оптимальные параметры checkpoint_timeout и max_wal_size для высоконагруженной OLTP-системы.

Домашнее задание по теме «Разбор и представление SQL-запросов».

Создайте таблицу products с несколькими столбцами (включая NULL-значения). Напишите SQL-запросы к системным каталогам (pg_class, pg_attribute, pg_type), чтобы программно получить: имя таблицы, имена столбцов, их типы, смещения в кортеже, флаги NOT NULL. Нарисуйте AST (Abstract Syntax Tree) для запроса UPDATE products SET price = price * 1.1 WHERE category = 'electronics', выделив узлы: оператор, таблица, условие, выражение. Объясните, как парсер и семантический анализатор проверяют корректность этого запроса (существование таблицы, типы данных, привязка имён).

Домашнее задание по теме «Транзакции и клиент–серверное взаимодействие».

Напишите два скрипта на Python (с использованием psycopg2), имитирующих две параллельные сессии, которые вызывают взаимоблокировку (deadlock) при обновлении строк в таблицах accounts и transactions. Зафиксируйте в логах PostgreSQL сообщение о deadlock. Проанализируйте, как работает lock manager, какие типы блокировок задействованы (row-level, table-level). С помощью pg_locks и pg_stat_activity определите, какие процессы участвуют в конфликте. Объясните, как MVCC и WAL обеспечивают изоляцию и устойчивость изменений. Напишите рекомендации по проектированию приложений, чтобы избегать deadlocks (например, фиксированный порядок обновления строк).

Примерные вопросы для коллоквиума.

1. Опишите, как SQL-запрос проходит через компоненты СУБД: от получения на стороне сервера до возврата результата. В чём разница между логическим и физическим планом выполнения?
2. Что такое страница (page) в контексте физического хранения данных? Как устроены слоты и кортежи (tuple layout) на странице, и как представляются NULL-значения на уровне хранения?
3. Объясните, как работает буферный менеджер в СУБД. Что такое buffer hit/miss, dirty pages и чекпоинт? Почему PostgreSQL использует алгоритм clock-sweep вместо LRU?
4. Как системные каталоги (например, pg_class, pg_attribute) обеспечивают принцип самоописания базы данных? Как catalog manager взаимодействует с буферным пулом при чтении метаданных?
5. В чём разница между AST (Abstract Syntax Tree) и QueryTree? Как семантический анализатор проверяет корректность имён столбцов и типов данных в запросе?
6. Опишите работу алгоритма Nested Loop Join без использования индексов. Как он взаимодействует с Volcano-моделью исполнения? В каких случаях он эффективен, а в каких — нет?
7. Сравните B+-дерево и хеш-индекс по структуре, производительности и поддерживаемым операциям. Почему B+-дерево подходит для диапазонных запросов, а хеш-индекс — нет?
8. Как cost-based оптимизатор выбирает план выполнения запроса? Какую роль играют статистика (через ANALYZE) и функция оценки стоимости (cost model) в этом процессе?
9. Объясните, как механизм MVCC (многоверсионное управление параллелизмом) позволяет обеспечивать изоляцию транзакций без блокировок на чтение. Какие проблемы он решает и какие создаёт?
10. Опишите жизненный цикл клиент-серверного соединения в PostgreSQL: от установки сокета до завершения сессии. В чём разница между сессией и транзакцией? Как работает postmaster и backend-процессы?

Примеры заданий и критерии оценивания проекта.

Цель проекта:

Разработать и представить инженерный разбор выполнения SQL-запроса «от SQL до диска», демонстрирующий глубокое понимание внутренних механизмов СУБД: от парсинга и оптимизации до физического доступа к данным, управления памятью, индексации и обеспечения целостности транзакций. Проект должен показать способность связывать логику запроса с его физическим исполнением и выявлять узкие места производительности.

Задачи и этапы выполнения проекта:

1. **Выбор и формулировка аналитического запроса**
   — Выберите нетривиальный SQL-запрос (например, JOIN нескольких таблиц с фильтрацией, агрегацией и сортировкой).
   — Создайте схему БД и заполните таблицы реалистичными данными (например, с помощью pgbench или скрипта генерации).
2. **Анализ пути выполнения запроса в СУБД**
   — Опишите, как запрос проходит этапы: парсинг, семантический анализ, построение AST/QueryTree, доступ к системным каталогам.
   — Покажите, как СУБД проверяет существование таблиц и столбцов, разрешает

имена, проверяет типы.

3. **Исследование плана выполнения и физического доступа**
   — Используйте EXPLAIN (ANALYZE, BUFFERS, VERBOSE) для получения детального плана.
   — Проанализируйте: какие операции используются (Seq Scan, Index Scan, Nested Loop, Hash Join и т.д.), почему выбран именно этот план.
   — Свяжите операции с физическим хранением: какие страницы читаются, сколько buffer hit/miss, есть ли dirty pages.
4. **Моделирование или имитация ключевых механизмов (на выбор)**
   Выберите **один** из следующих элементов и реализуйте упрощённую модель:
   - Симулятор B+-дерева (вставка, поиск, диапазонный запрос)
   - Имитация Volcano-модели (итераторы: SeqScan, Filter, NestedLoopJoin)
   - Простой буферный менеджер с политикой Clock-sweep
   - Визуализация MVCC: отображение версий строк и видимости для разных транзакций
5. **Диагностика и оптимизация производительности**
   — Найдите узкое место в выполнении запроса (например, отсутствие индекса, плохая статистика, высокий I/O).
   — Внесите изменения (добавьте индекс, выполните ANALYZE, перепишите запрос) и сравните производительность до и после.
   — Объясните, как изменения повлияли на cost model, план и физический доступ.
6. **Подготовка и защита проекта**
   — Подготовьте презентацию (10–12 слайдов), включающую: схему БД, текст запроса, план выполнения, анализ узких мест, результаты оптимизации, демонстрацию модели.
   — На защите ответьте на вопросы, подтверждающие понимание архитектуры СУБД.

Критерии оценивания.

1. Глубина и системность анализа запроса
   — Понимание всех этапов обработки запроса: от SQL до диска.
   — Умение связать логический запрос с физическим исполнением, включая метаданные, буферы, индексы и транзакции.
2. Корректность и содержательность технических решений
   — Адекватность выбранной модели (B+-дерево, итераторы и т.д.).
   — Точность интерпретации EXPLAIN, pg_stat, pageinspect и других диагностических инструментов.
3. Навыки диагностики и оптимизации производительности
   — Умение выявить бутылочное горлышко (I/O, отсутствие индекса, устаревшая статистика).
   — Обоснованность предложенных улучшений и демонстрация их эффекта.
4. Качество презентации и защита
   — Чёткая, логичная подача материала.
   — Способность отвечать на уточняющие и сложные вопросы, демонстрируя понимание внутреннего устройства СУБД.
5. Самостоятельность и творческий подход

— Оригинальность модели или способа визуализации механизмов СУБД.
— Глубина проработки деталей (например, layout кортежа, поведение MVCC, работа clock-sweep).

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Укажите, сколько этапов проходит SQL-запрос в классической архитектуре СУБД, начиная с парсера и заканчивая исполнением. В ответе запишите число и название последнего этапа через пробел.	5 executor	УК-1
2	Сколько байт в PostgreSQL обычно занимает фиксированная часть кортежа (HeapTupleHeader)? В ответе запишите число.	24	УК-1
3	Какое максимальное количество кортежей может храниться на одной странице размером 8 КБ, если каждый кортеж занимает 200 байт, а служебные накладные расходы составляют 100 байт? В ответе запишите число.	39	УК-1
4	Сколько уровней может иметь B+-дерево индекса, если каждый узел содержит до 100 ключей, а индекс включает 1 000 000 записей? В ответе запишите число.	4	УК-1
5	Какой механизм в PostgreSQL отвечает за управление версиями строк и позволяет читать данные без блокировок? В ответе запишите число 1 и название механизма через пробел.	1 MVCC	УК-2
6	Сколько типов блокировок уровня строки выделяют в PostgreSQL (например, Row Share, Row Exclusive и др.)? В ответе запишите число.	8	УК-2
7	Какой параметр в PostgreSQL определяет размер общей памяти, выделенной для кэширования данных? В ответе запишите название параметра.	shared_buffers	УК-2
8	Сколько этапов оптимизации запроса включает cost-based оптимизатор в PostgreSQL (от генерации планов до выбора наилучшего)? В ответе запишите число.	1	УК-2
9	Какой системный каталог в PostgreSQL хранит информацию о таблицах и индексах? В ответе запишите название каталога.	pg_class	ОПК-1
10	Какой алгоритм используется в PostgreSQL для вытеснения страниц из буферного пула? В ответе запишите название алгоритма.	clock-sweep	ОПК-1
11	Сколько бит в null-маске выделяется под хранение информации о NULL для столбца в кортеже? В ответе запишите число.	1	ОПК-1
12	Какой метод доступа используется для выполнения операции равенства в хеш-индексе? В ответе запишите название метода через пробел.	хеширование	ОПК-1
13	Какой оператор в PostgreSQL используется для просмотра физического плана выполнения запроса? В ответе запишите название оператора.	EXPLAIN	ПК-2
14	Сколько соединений одновременно может обрабатывать postmaster в PostgreSQL по умолчанию (стандартное значение max_connections)? В ответе запишите число	100	ПК-2
15	Какой компонент СУБД отвечает за запись всех изменений данных перед их сохранением на диск? В ответе запишите аббревиатуру механизма.	1WAL	ПК-2

16	Какой тип соединения используется по умолчанию при отсутствии индексов и малом размере одной из таблиц? В ответе запишите название алгоритма.	Nested Loop	ПК-2
17	Напишите, сколько строк будет обработано во внешнем цикле Nested Loop Join, если в таблице A 1000 строк, а в таблице B — 50. В ответе запишите число.	1000	ПК-4
18	Какой тип индекса рекомендуется использовать для диапазонных запросов по датам? В ответе запишите название индекса.	B+-дерево	ПК-4
19	Какой системный каталог в PostgreSQL хранит информацию о столбцах таблиц? В ответе запишите название каталога.	pg_attribute	ПК-4
20	Какой итератор в Volcano-модели возвращает строки по одной из таблицы? В ответе запишите название оператора.	SeqScan	ПК-4