

УТВЕРЖДЕНА

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

**Рабочая программа дисциплины (модуля)
«Базы 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) «Базы 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) «Базы данных» позволяет студентам освоить современные методы обработки и анализа больших данных, что является ключевым навыком в условиях стремительного развития технологий. Дисциплина (модуль) формирует фундаментальные навыки работы с данными, которые являются критически важными в современном мире, где информация играет ключевую роль в бизнесе и технологиях. Кроме того, понимание баз данных открывает возможности для карьерного роста в различных областях, включая разработку программного обеспечения, анализ данных и управление информационными системами.

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 обязательную часть Блока 1.

Дисциплина (модуль) изучается на 2 или 3 курсе в 3, 4 или 5 семестре на выбор.

Цель изучения дисциплины (модуля): заключается в формировании у студентов знаний и навыков, необходимых для проектирования, разработки и управления системами управления базами данных.

Задачи изучения дисциплины (модуля):

- освоение языка запросов SQL, устройства реляционных баз данных, версионирования строк и журналирование;
- формирование умения проектировать эффективные структуры баз данных;
- формирование умения выполнять сложные запросы к базе данных;
- формирование умения анализировать бизнес-задачи и строить по ним модели базы данных.

В результате освоения дисциплины (модуля), обучающийся должен:

знать:

- основы реляционной модели данных, включая понятия таблиц, строк, столбцов и связей между таблицами;
- синтаксис и семантику языка запросов SQL, включая DDL, DML и DCL операции;
- принципы версионирования строк и журналирования в реляционных базах данных для обеспечения целостности и восстановления данных;
- методы нормализации и денормализации данных для оптимизации структуры базы данных;

уметь:

- проектировать структуры баз данных, учитывая требования к эффективности хранения и быстройдействию запросов;
- выполнять сложные SQL-запросы с использованием подзапросов, JOIN, агрегатных функций и оконных функций;
- анализировать бизнес-задачи и формализовывать требования для построения адекватных моделей баз данных;
- использовать механизмы версионирования и журналирования для контроля изменений и обеспечения отказоустойчивости;

владеть:

- практическими навыками создания и модификации баз данных в СУБД (например, POSTGRESQL, MYSQL);
- инструментами разработки и отладки SQL-запросов, включая оптимизацию производительности;
- методами проектирования баз данных с применением ER-диаграмм и других средств моделирования;
- технологиями управления транзакциями и обеспечения согласованности данных в многопользовательской среде.

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)
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-5.	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-5.1.	Знаете технологии, необходимые для прикладного программирования, включая современные функциональные языки программирования, а также основные принципы и понятия, применяемыми при использовании компьютерных сетей
		ОПК-5.2.	Умеет пользоваться технологиями прикладного программирования, включая среды высокоуровневого программирования
		ОПК-5.3.	Имеет практический опыт использования технологий прикладного программирования
ПК-3.	Способен применять методы математического и алгоритмического моделирования для решения как теоретических, так и практических задач в рамках профессиональной	ПК-3.1.	Знает основные методы математического и алгоритмического моделирования, а также их применение для решения теоретических и прикладных задач

	деятельности	ПК-3.2.	Умеет применять методы математического и алгоритмического моделирования для анализа и решения различных задач в области математики и компьютерных наук
		ПК-3.3.	Имеет опыт использования методов математического и алгоритмического моделирования при решении теоретических и прикладных задач в профессиональной деятельности

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)					
1	Реляционная алгебра и язык запросов SQL	5	5	2	24	Домашние задания, Контрольная работа Подготовка к семинару
2	Проектирование баз данных и нормальные формы	5	5	2	26	Домашние задания, Контрольная работа Подготовка к семинару
3	Расширенные возможности SQL	5	5	2	26	Домашние задания, Контрольная работа
4	Анализ и улучшение производительности в БД	5	5		24	Домашние задания Подготовка к семинару
5	Хранилище данных	6	6		24	Домашние задания Подготовка к семинару
	<i>Зачет с оценкой</i>			8		Защита проекта
	Итого:	26	26	14	124	
	Объем дисциплины (модуля) (в ак. ч.)	190				
	Объем дисциплины (модуля) (в зач. ед.)	5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Реляционная алгебра и язык запросов SQL	Введение в базы данных и реляционную алгебру. Язык SQL. Транзакции
2	Проектирование баз данных и нормальные формы	Проектирование базы данных. Нормальные формы.
3	Расширенные возможности SQL	Подзапросы. Оконные функции. Версионность таблиц. Дополнительные возможности PostgreSQL
4	Анализ и улучшение производительности в БД	План запроса и анализ производительности. Индексы и оптимизация запросов
5	Хранилище данных	Ведение хранилища данных

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. 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. Файлы, К. SQL. Руководство для использования с любыми SQL СУБД : учебное пособие / К. Файлы ; пер. с англ. А. В. Хаванова. - 2-е изд. - Москва : ДМК Пресс, 2023. - 454 с. - ISBN 978-5-89818-323-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2102610>.

3. Маркин, А. В. Базы данных. PostgreSQL : учебник для вузов / А. В. Маркин. — Москва : Издательство Юрайт, 2025. — 799 с. — (Высшее образование). — ISBN 978-5-534-21779-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/582050>.

4. Молиаро, Э. SQL. Сборник рецептов : практическое руководство / Э. Молиаро. - 2-е изд. - Санкт-Петербург : БХВ-Петербург, 2022. - 592 с. - ISBN 978-5-9775-6759-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123362>.

Дополнительная литература:

1. Основы технологий баз данных: учебное пособие / Б. А. Новиков, Е. А. Горшкова, Н. Г. Графеева; под ред. Е. В. Рогова. — 2-е изд. — М.: ДМК Пресс, 2020. — 582 с. - ISBN 978-5-97060-841-8

2. Агальцов, В. П. Базы данных : в 2 книгах. Книга 2. Распределенные и удаленные базы данных : учебник / В.П. Агальцов. — Москва : ФОРУМ : ИНФРА-М, 2024. — 271 с. — (Высшее образование). - ISBN 978-5-8199-0959-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2138458>.

3. Жао Элис. SQL. Pocket guide. 4-е изд. — Астана: «Спринт Бук», 2024. — 320 с. - ISBN 978-601-08-3728-7

4. Исаченко, О. В. Базы данных : учебное пособие / О.В. Исаченко. — Москва : ИНФРА-М, 2025. — 202 с. — (Высшее образование). - ISBN 978-5-16-020320-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2169541>.

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) «Базы данных» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, проект, домашние задания и контрольные работы, 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.

Участие в семинаре (аудиторная работа) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками,

чтобы обогатить обсуждение и расширить свои знания.

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.

Контрольная работа – письменная работа с набором задач, которые нужно решить за ограниченное время.

Цель контрольной работы - получить специальные знания по одной или нескольким темам дисциплины (модуля) и продемонстрировать навыки их практического применения.

Проект – работа студентов, направленная на достижение общей цели или решение определенной задачи, в ходе которой участники объединяют свои знания и навыки.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, 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) «Базы 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,
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			<p>обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые 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.</p>
7	Хорошо	Зачтено	<p>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.</p>
6	Хорошо	Зачтено	<p>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.</p>
5	Удовлетворительно	Зачтено	
4	Удовлетворительно	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
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) «Базы данных» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	30%	12	Набор задач по темам недели
Контрольные работы	40%	3	Письменная работа с набором задач, которые нужно решить за ограниченное время
Итоговый проект	30%	1	Подготовка проекта и его защита перед аудиторией

Формула расчёта итоговой оценки по дисциплине (модулю) «Базы данных»:
« $0,3 \times$ среднее за домашние задания + $0,4 \times$ среднее за контрольные работы + $0,3 \times$ защита проекта».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные вопросы для подготовки к семинарам

Реляционная алгебра и язык запросов SQL. Введение в базы данных и реляционную алгебру. Язык SQL. Транзакции

- Что такое реляционная модель данных и как она представлена в базе данных?
- Какие основные операции включает реляционная алгебра?
- Чем отличается реляционная алгебра от реляционного исчисления?
- Какие основные типы SQL-запросов существуют?
- Как выполняется операция SELECT в языке SQL?
- Что такое транзакция в базе данных и какие свойства она должна удовлетворять (ACID)?
- Как реализуется управление параллелизмом транзакций?
- Что такое коммит и роллбэк в контексте транзакций?
- Каковы основные типы данных в SQL?
- Как осуществляется соединение таблиц (JOIN) в SQL?
- Что такое первичный ключ и зачем он нужен?
- Какие существуют ограничения целостности данных в реляционных базах?

Проектирование баз данных и нормальные формы. Проектирование базы данных. Нормальные формы.

- Что такое концептуальное, логическое и физическое проектирование базы данных?
- Какие этапы включает процесс проектирования базы данных?
- Что такое ER-диаграмма и какие элементы она содержит?
- Что такое функциональная зависимость в контексте базы данных?
- Какие существуют нормальные формы и в чем их смысл?
- Как определить, что таблица находится в первой нормальной форме (1НФ)?
- Что требует вторая нормальная форма (2НФ)?
- Чем отличается третья нормальная форма (3НФ) от второй?

- Что такое нормальная форма Бойса-Кодда (BCNF)?
- Какие проблемы решает нормализация базы данных?
- В каких случаях применяется денормализация и зачем она нужна?
- Как проектировать эффективные структуры базы данных с учетом нормальных форм?

Расширенные возможности SQL. Подзапросы. Оконные функции. Версионность таблиц. Дополнительные возможности PostgreSQL.

- Что такое подзапрос и как он используется в SQL?
- Какие бывают типы подзапросов (коррелированные и некоррелированные)?
- Как работают оконные функции и в чем их отличие от агрегатных функций?
- Приведите примеры оконных функций и их применения.
- Что такое версионность таблиц и зачем она нужна?
- Какие механизмы версионности поддерживаются в PostgreSQL?
- Как реализовать временное хранение данных в PostgreSQL?
- Какие расширенные типы данных поддерживает PostgreSQL?
- Что такое CTE (Common Table Expressions) и как их использовать?
- Как реализуются триггеры и процедуры в PostgreSQL?
- Какие дополнительные операторы и функции предоставляет PostgreSQL по сравнению со стандартом SQL?
- Как использовать JSON и JSONB типы данных в PostgreSQL?

Анализ и улучшение производительности в БД. План запроса и анализ производительности. Индексы и оптимизация запросов.

- Что такое план запроса и как его получить в PostgreSQL?
- Какие типы индексов существуют и в чем их отличия?
- Как индексы влияют на производительность запросов?
- Какие ситуации могут привести к ухудшению производительности запросов?
- Что такое селективность индекса?
- Как использовать EXPLAIN и EXPLAIN ANALYZE для анализа запросов?
- Какие методы оптимизации запросов вы знаете?
- Что такое полнотекстовый поиск и как его реализовать?
- Как влияет порядок условий в WHERE на производительность?
- Что такое партиционирование таблиц и как оно помогает масштабировать БД?
- Какие существуют методы кэширования запросов?
- Как мониторить и профилировать работу базы данных?

Хранилище данных. Ведение хранилища данных.

- Что такое хранилище данных и чем оно отличается от обычной базы данных?
- Какие основные компоненты входят в архитектуру хранилища данных?
- Что такое ETL-процесс и какова его роль?
- Какие типы хранилищ данных существуют (OLTP, OLAP)?
- Что такое звездчатая и снежинкавая схемы моделирования?
- Как организуется хранение исторических данных в хранилищах?
- Какие инструменты используются для построения хранилищ данных?
- Что такое агрегированные данные и зачем они нужны?
- Как обеспечивается качество данных в хранилище?
- Какие существуют подходы к обновлению данных в хранилище?
- Как осуществляется интеграция данных из разных источников?
- Какие задачи решает хранилище данных в бизнес-аналитике?

Примерные домашние задания

Домашнее задание

ЗАДАНИЕ 1

Классификация баз данных

Определи тип базы данных (реляционная, NoSQL, объектно-ориентированная, графовая или другая) для каждого из следующих сценариев использования. Обоснуй свой выбор, указав преимущества выбранного типа для данного сценария.

1. Банковская система, отслеживающая счета клиентов, транзакции и кредиты.
2. Социальная сеть с миллионами пользователей и сложными связями между ними.
3. Интернет-магазин с каталогом товаров, корзиной покупок и системой заказов.
4. Система рекомендаций фильмов, учитывающая предпочтения пользователей.
5. Медицинская система с историями болезней пациентов и результатами анализов.

ЗАДАНИЕ 2

Найти имена всех хостов и их IP-адреса. Результат должен содержать столбцы **host_name** и **ip_address**.

ЗАДАНИЕ 3

Для каждой компании найти пользователя, который совершил наибольшее количество успешных подключений к хостам этой компании. Результат должен содержать название компании, логин пользователя и количество подключений.

ЗАДАНИЕ 4

Найди все хосты, зарегистрированные в первом квартале 2022 года (с января по март), и отсортируй их по дате регистрации в порядке возрастания.

ЗАДАНИЕ 5

Получи список уникальных названий компаний из таблицы **host**, исключив записи, где название компании не указано (**NULL**). Переименуй столбец в «Название организации» и отсортируй результат по алфавиту.

ЗАДАНИЕ 6

Получи информацию о пользователях, чья роль начинается с буквы **d** и email-адрес заканчивается на **.com**. Выведи логин, имя, email и роль пользователя. Отсортируй результат сначала по роли, затем по имени пользователя. Пропусти первую запись и выведи следующие 2 записи.

ЗАДАНИЕ 7

Таблица IP-адресов содержит полные дубли, выведи их.

Домашнее задание

ЗАДАНИЕ 1

Для компании **DDS** в разбивке по хостам посчитайте все события и выведите дату последнего события для каждого хоста. Ответ должен содержать адрес сервера, количество событий и дату последнего события.

ЗАДАНИЕ 2

Создайте и выведите результат материализованного представления — все **id** подключения, **id** хоста и **id** пользователя, у которых среднее время подключения больше или равно их времени подключения. Вывод ограничить до 10 строк.

ЗАДАНИЕ 3

Вы пишете функцию, которая увеличивает баланс пользователя на **N** единиц. Несколько таких транзакций могут запускаться одновременно, например, при зачислении бонусов или пополнении. Важно, чтобы ни одно увеличение не было потеряно.

Какой уровень изоляции следует установить, чтобы избежать потери обновлений, но при этом минимизировать количество откатов?

Необходимо написать транзакцию обновления баланса пользователя (таблица **users**, поле **balance**) с подходящим уровнем изоляции.

ЗАДАНИЕ 4

Пользователь пополняет счёт. Администратор в это же время может «заблокировать» счёт (**status = 'blocked'**).

Напишите код для пополнения счета (таблица **accounts**, поле **balance**) пользователя 1, обогащенный всеми необходимыми конструкциями поддержания целостности операций.

Пополнение не должно происходить, если счёт уже заблокирован или блокируется во время пополнения.

Примерные задания для контрольных работ

Контрольная работа 1: Проектирование баз данных и нормальные формы

1. **Определение сущностей:** Опишите, что такое сущность в контексте проектирования баз данных, и приведите примеры сущностей для системы управления университетом.

2. **ER-диаграмма:** Нарисуйте ER-диаграмму для системы управления библиотекой, включая сущности "Книги", "Читатели", "Заказы" и их связи.

3. **Нормальные формы:** Объясните, что такое первая нормальная форма (1NF) и приведите пример таблицы, которая не соответствует 1NF, а затем преобразуйте её в 1NF.

4. **Проверка на нормальные формы:** У вас есть таблица Students с полями StudentID, Name, Course, Instructor, InstructorPhone. Определите, соответствует ли она 2NF, и если нет, приведите её к 2NF.

5. **Создание таблиц:** Напишите SQL-скрипт для создания таблицы Courses с полями CourseID, CourseName, Credits, и определите необходимые ограничения (например, первичный ключ).

6. **Денормализация:** Объясните, что такое денормализация, и приведите пример, когда это может быть полезно в проектировании базы данных.

7. **Схема базы данных:** На основе вашей ER-диаграммы из задания 2 создайте SQL-скрипт для создания таблиц с соответствующими полями и ограничениями, включая внешние ключи.

Контрольная работа 2: Расширенные возможности SQL

1. **Подзапросы:** Напишите SQL-запрос, который выбирает все книги из таблицы Books, которые были заказаны более чем 5 раз.

2. **Представления:** Создайте представление TopReaders, которое будет содержать информацию о читателях, которые сделали более 10 заказов, и выведите их имена и количество заказов.

3. **Триггеры:** Опишите, как создать триггер, который будет автоматически обновлять поле LastCheckedOut в таблице Books при каждом новом заказе книги.

4. **Хранимые процедуры:** Напишите хранимую процедуру, которая принимает ReaderID и возвращает список всех книг, заказанных этим читателем.

5. **Использование индексов:** Объясните, что такое индекс в базе данных и как он может улучшить производительность запросов. Приведите пример создания индекса для таблицы Orders.

6. **Оптимизация запросов:** Приведите пример неэффективного SQL-запроса и перепишите его, чтобы улучшить производительность. Объясните, какие изменения вы внесли.

7. **Мониторинг производительности:** Опишите методы мониторинга производительности базы данных и какие метрики следует отслеживать для оценки эффективности работы базы данных.

Примерное задание для проекта

Сдача проекта

Перед демонстрацией решения необходимо загрузить в LMS архив со следующим содержимым:

1. Скриншоты концептуальной и логической схем.
2. csv файл(ы) со сгенерированными данными (файлы должны содержать все необходимые данные для полного восстановления базы).
3. SQL-скрипт, в котором должен находиться следующий код:
 - DDL запросы на создание схемы
 - Запросы с загрузкой данных
 - Запрос с демонстрацией валидации данных (смотри пункт 6 в порядке демонстрации решения)
 - Запрос с демонстрацией загрузки в SCD2 таблицу (смотри пункт 7 в порядке демонстрации решения)
 - 2 аналитических запроса

Защита проекта

Порядок демонстрации решения:

1. Скриншот концептуальной схемы
2. Скриншот логической схемы
3. Демонстрация физической схемы в DBeaver
4. `select * limit 100` из всех таблиц
5. `select count(*)` из всех таблиц
6. Демонстрация загрузки данных в таблицу с триггер функцией на валидацию (загрузить валидные и невалидные данные, в первом случае вывести лог об успешной загрузке, во втором случае должна быть ошибка).
7. Демонстрация загрузки данных в SCD2-таблицу:
 - загрузка новых данных
 - обновлённых данных
 - данных, которых не было ранее
8. Демонстрация созданных индексов / партиций
9. Запуск 2х аналитических запросов с выводом планов выполнения запросов (EXPLAIN ANALYZE)

Пары 14 и 15 недели

Лекции, запланированные на **09.12** и **16.12** в 11:30-12:50, будут заменены форматами консультаций по проекту.

Каждый студент может присутствовать на консультации и задать интересующие вопросы.

Приём проектных работ будет осуществляться **очно** в дни проведения семинарских занятий - по четвергам и пятницам, в две отдельные волны:

- **Первая волна:** 11 декабря в 17:30-19:00 и 12 декабря в 19:00-20:20
- **Вторая волна:** 18 декабря в 17:30-19:00 и 19 декабря в 19:00-20:20

Защита проекта осуществляется **на семинаре вашей учебной группы**. Настоятельно рекомендуется представить проект к защите в рамках **первой волны**.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Что такое реляционная алгебра? А) Язык программирования для создания веб-приложений В) Математическая основа для работы с реляционными базами данных С) Метод хранения данных в виде файлов D) Система управления версиями для баз данных	В	ОПК-4

2.	Какова цель нормализации базы данных? А) Увеличение объема данных В) Минимизация избыточности данных и предотвращение аномалий С) Упрощение структуры базы данных D) Оптимизация производительности запросов	В	ОПК-5
3.	Какой метод используется для устранения избыточности данных в проектировании базы данных?	Нормализация	ОПК-4
4.	Какой тип подзапроса возвращает одно значение?	Скалярный	ПК-3
5.	Какой тип индекса в PostgreSQL обеспечивает уникальность значений?	Уникальный индекс	ОПК-4
6.	Как называется функция, которая позволяет создавать временные таблицы в PostgreSQL?	TEMPORARY	ОПК-4
7.	Какой уровень нормализации устраняет транзитивные зависимости?	Третья нормальная форма	ОПК-5
8.	Какой оператор используется для создания оконной функции в SQL?	OVER	ОПК-5
9.	Как называется процесс, позволяющий отслеживать изменения в таблицах?	Версионность	ОПК-5
10.	Какой метод анализа производительности запросов используется в PostgreSQL для оценки плана выполнения?	EXPLAIN	ОПК-5
11.	Какой алгоритм используется для оптимизации поиска в базе данных через индексы?	В-дерево	ПК-3
12.	Какой тип подзапроса позволяет использовать результаты запроса в основном запросе?	Вложенный подзапрос	ПК-3