Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)

Кафедра информационных технологий
Факультет информационных технологий и анализа больших данных

Документ подписан усиленной неквалифицированной электронной подписью
Организация: Финансовый университет при Правительстве РФ
Утвержено: Проректор по учебной и методической работе Е.А. Каменева
Сертификат: TU5N5WvdGZ4ODkWsFhMDvqqBie3iMSHE
Дата: 13.10.2025 г.

О.П. Тулупова

Реляционные и нереляционные базы данных

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:
01.04.02 - Прикладная математика и информатика,
Образовательная программа «Инженер машинного обучения»
Направленность программы:
«Инженер машинного обучения»

Рекомендовано
Факультет информационных технологий и анализа больших данных
(протокол № 03 от 16.12.2025 г.)
Одобрено
Кафедра информационных технологий
(протокол № 12 от 03.12.2025 г.)

© Москва 2025

СОДЕРЖАНИЕ

1.	Наименование дисциплины	3
2.	Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3.	Место дисциплины в структуре образовательной программы	5
4.	Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	5
5.	Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	6
5.1.	Содержание дисциплины	6
5.2.	Учебно-тематический план	8
5.3.	Содержание семинаров	9
6.	Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	12
6.1.	Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	12
6.2.	Перечень вопросов, заданий, тем для подготовки к текущему контролю	15
7.	Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	20
8.	Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	28
9.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	29
10.	Методические указания для обучающихся по освоению дисциплины	30
11.	Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	30
12.	Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	31

1. Наименование дисциплины

«Реляционные и нереляционные базы данных».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-8	Способность создавать, описывать и качественно контролировать исполнение технологических требований и нормативных документов в профессиональной деятельности	Демонстрирует знания в области создания эффективных стратегий управления: командой, современными информационными системами и процессами разработки программных средств, а также знания нормативных документов в профессиональной деятельности	**знать:** критически важные критерии выбора СУБД для различных доменов;принципы ранжирования требований к безопасности (шифрование, аутентификация) и надежности (репликация, отказоустойчивость) в зависимости от предметной области **уметь:** обосновать выбор между реляционной и документной моделью данных на основе сравнительного анализа их характеристик;аргументировать необходимость использования колоночных хранилищ (например, ClickHouse) для аналитических нагрузок, объясняя архитектурные различия OLTP и OLAP систем
		Разрабатывает эффективные стратегии управления: командой разработчиков, современными информационными системами и процессами разработки программных средств	**знать:** критерии выбора между классическими реляционными и гибридными решениями (например, PostgreSQL с JSONB);принципы оптимизации работы с полуструктурированными данными (индексы GIN/GIST для JSONB в PostgreSQL) **уметь:** разрабатывать архитектуру гибридной модели данных, определяя, какие данные хранить в реляционных таблицах, а какие в документных форматах;проектировать процесс репликации данных для аналитической отчетности, выбирая подходящие ETL/ELT-инструменты

			или встроенные механизмы СУБД
		Управляет командой разработчиков, современными информационными системами и процессами разработки программных средств с учетом технологических требований	**знать:** методы оптимизации производительности через материализованные представления в аналитических СУБД;синтаксис и особенности применения TTL-индексов в MongoDB и ClickHouse **уметь:** управлять производительностью СУБД: создавать и обновлять материализованные представления в ClickHouse для ускорения отчетов;обеспечивать целостность данных в гибридных моделях на PostgreSQL, применяя индексы для JSONB и создавая ограничения для проверки структуры документов

3. Место дисциплины в структуре образовательной программы

Дисциплина «Реляционные и нереляционные базы данных» относится к «Модулю общепрофессиональных дисциплин направления».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Очная форма обучения

Вид учебной работы по дисциплине	**Всего (в з/е и часах)**	**Модуль 3 (в часах)**
Общая трудоёмкость дисциплины	4/144	144
Контактная работа-Аудиторные занятия	32	32
Лекции	8	8
Семинары, практические занятия	24	24
Самостоятельная работа	112	112
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям

Как определять критически важные критерии выбора для разных предметных областей.

Методика создания чек-листов для первичной оценки СУБД.

Ранжирование требований к безопасности и надежности в зависимости от домена.

Тема 2. Архитектурные паттерны. Гибридные модели данных и их практическая реализация

Сравнительный анализ реляционных и документных моделей.

Критерии выбора между классическими и гибридными решениями.

Оптимизация работы с JSONB в PostgreSQL (индексы, запросы).

Обеспечение целостности данных в гибридных моделях.

Тема 3. Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища

Архитектурные различия OLTP и OLAP систем.

Преимущества колоночного хранения для аналитических нагрузок.

Паттерны репликации данных из операционных систем в аналитические.

Оптимизация производительности через материализованные представления.

Тема 4. Управление жизненным циклом данных (Data Lifecycle Management)

Политики хранения данных: классификация данных на "горячие", "теплые" и "холодные".

Инструменты для управления данными: использование партиционирования и табличных пространств в PostgreSQL, TTL (Time-To-Live) индексов в MongoDB и ClickHouse для автоматического удаления устаревших данных.

Архивация и выгрузка исторических данных в объектные хранилища (например, Yandex Cloud (Object Storage)).

5.2. Учебно-тематический план

№ п/п	Наименование тем(разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоятельная работа
			Общая, в т.ч.:	Лекции	Семинары, практические занятия	
1	Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям	36	8	2	6	28
2	Архитектурные паттерны. Гибридные модели данных и их практическая реализация	36	8	2	6	28
3	Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища	36	8	2	6	28
4	Управление жизненным циклом данных (Data Lifecycle Management)	36	8	2	6	28
	Итого	144	32	8	24	112

5.3. Содержание семинаров

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям	Анализ приоритетности критериев выбора СУБД (целостность данных, масштабируемость, доступность, гибкость модели) для различных предметных областей, включая: определение первичного критерия для финтех-проектов и его обоснование; сравнительный анализ изменения приоритетов для социальных сетей; факторы влияния на расстановку приоритетов в разных доменах. Сравнительный анализ совокупной стоимости владения (TCO) при выборе между коммерческими и open-source СУБД, охватывающий: ключевые компоненты TCO (Total Cost of Ownership) и методы их расчета; сравнительные преимущества и недостатки различных моделей лицензирования; критерии оптимизации долгосрочных затрат на эксплуатацию. Разработка структуры чек-листа для первичной оценки применимости документной СУБД, содержащего: 3-5 ключевых критериев отбора для нового проекта; области применения документных СУБД в различных архитектурных решениях; ограничения и риски использования документного подхода. Анализ влияния требований информационной безопасности на выбор СУБД в регулируемых отраслях, включающий: сравнительную характеристику требований в финтехе, здравоохранении и e-commerce; критерии соответствия отраслевым стандартам и сертификациям; архитектурные ограничения и компромиссы при обеспечении безопасности.	Интерактивная форма, практикум по решению задач по тематике занятия, коллективное обсуждение решений или выполнение самостоятельной работы.
Архитектурные паттерны. Гибридные модели данных и их практическая реализация	Критерии и этап принятия архитектурного решения о применении гибридной модели данных (PostgreSQL + JSONB) в противовес классической реляционной или документной СУБД. Оценка эффективности механизмов обеспечения целостности данных (CHECK-ограничения, триггеры) в гибридной модели PostgreSQL и определение критериев целесообразности возврата к реляционной схеме. Сравнительный анализ индексов GIN и B-tree для работы с данными JSONB в PostgreSQL, включая: критерии выбора	Интерактивная форма, практикум по решению задач по тематике занятия, коллективное обсуждение решений или выполнение самостоятельной работы.

	типа индекса для различных паттернов запросов; влияние на производительность операций модификации данных (INSERT/UPDATE); оптимальные сценарии применения каждого типа индексации. Поэтапная стратегия миграции существующей реляционной схемы на гибридную модель данных (PostgreSQL + JSONB), содержащая: ключевые метрики оценки успешности миграции; критерии приемлемости результатов перехода; показатели эффективности внедренной гибридной модели.	
Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища	Определение критических точек роста нагрузки на OLTP-систему, при которых становится целесообразным выделение отдельного OLAP-слоя, включая: анализ пороговых значений производительности операционной системы; критерии экономической эффективности разделения OLTP и OLAP нагрузок; методику оценки операционных затрат при различных архитектурных подходах. Исследование эффективности колоночной организации данных для различных типов запросов, охватывающее: принципы работы колоночного хранения и их влияние на производительность агрегаций; ограничения применимости для операций точечного доступа и частых вставок; сравнительные характеристики производительности для различных рабочих нагрузок. Сравнительный анализ движков семейства MergeTree в ClickHouse, включающий: функциональные особенности MergeTree, ReplacingMergeTree и SummingMergeTree; оптимальные сценарии применения для решения типовых бизнес-задач; критерии выбора движка для задач дедубликации (deduplicatio), агрегации и анализа временных рядов. Формирование системы мониторинга эксплуатационных характеристик кластера ClickHouse, предусматривающей: выбор ключевых метрик для оценки состояния системы; методы анализа утилизации ресурсов (CPU, память, дисковые операции); критерии идентификации проблем производительности и оптимизации нагрузки.	Интерактивная форма, практикум по решению задач по тематике занятия, коллективное обсуждение решений или выполнение самостоятельной работы.
Управление жизненным циклом данных (Data Lifecycle Management)	Политики хранения данных: определение, какие данные являются "горячими", "теплыми" и "холодными". Инструменты для управления данными: использование партиционирования и табличных пространств в PostgreSQL, TTL (Time-To-Live) индексов в MongoDB и ClickHouse для	Интерактивная форма, практикум по решению задач по тематике занятия, коллективное обсуждение решений

	автоматического удаления устаревших данных. Архивация и выгрузка исторических данных в объектные хранилища (например, Yandex Cloud (Object Storage)).	или выполнение самостоятельной работы.

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	**Перечень вопросов, отводимых на самостоятельное освоение**	**Формы внеаудиторной самостоятельной работы**
Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям	Разработка требований к миграционной готовности системы в техническом задании, включая: архитектурные паттерны снижения стоимости изменений (абстракция доступа к данным, ORM); критерии оценки связанности системы с конкретной СУБД; методы минимизации рисков будущей миграции на уровне проектирования. Анализ влияния непроизводственных рисков на стратегию выбора СУБД, охватывающий: оценку юридических и вендорских рисков (лицензионные изменения, зависимость от поставщика); критерии устойчивости технологического стека в долгосрочной перспективе; стратегии минимизации рисков для систем с жизненным циклом 10+ лет. Методика формализации требований к уровню экспертизы команды в техническом задании, включающая: критерии оценки соответствия технологического стека компетенциям команды; анализ целесообразности внедрения перспективных СУБД с высоким порогом вхождения; модели оценки операционных рисков при отсутствии готовых специалистов.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.
Архитектурные паттерны. Гибридные модели данных и их практическая реализация	Анализ границ применимости типа данных JSONB в PostgreSQL, включающий: признаки неэффективного использования JSONB для хранения структурированных данных; критерии целесообразности преобразования JSONB-данных в реляционную схему; методику оценки структурной целостности данных для выбора оптимального формата хранения. Сравнительный анализ эксплуатационных характеристик гибридной модели данных, охватывающий: оценку сложности рефакторинга схемы базы данных с JSONB-полями; особенности обеспечения обратной совместимости API при эволюции JSONB-структур; методы управления	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.

	миграциями в условиях гибридной модели хранения. Исследование производительности операций агрегации и сложных выборок в различных моделях хранения, включающее: сравнительный анализ эффективности обработки данных в JSONB и нормализованных таблицах; критерии оценки нагрузки на вычислительные ресурсы при работе с JSONB-структурами; определение пороговых значений производительности для выбора оптимальной модели хранения.	
Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища	Сравнительный анализ методов репликации данных в ClickHouse, включающий: оценку влияния параметров (задержка, объем данных, требования надежности) на выбор стратегии; сравнительные характеристики пакетной и потоковой передачи данных; оптимизацию процессов буферизации для обеспечения целостности и производительности. Анализ методов обработки изменяемых данных в колоночных системах хранения, охватывающий: сравнительную эффективность механизмов управления данными (ReplacingMergeTree, ALTER TABLE); критерии выбора между append-only моделью и периодическим пересчетом данных; оптимизацию операций обновления и удаления в условиях массовой вставки. Проектирование системы обеспечения отказоустойчивости в ClickHouse, включающее: сравнительный анализ стратегий резервирования на уровне кластера и отдельных узлов; оптимизацию процедур резервного копирования и восстановления в распределенной среде; критерии достаточности репликации и планирование ресурсов для обеспечения надежности.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.
Управление жизненным циклом данных (Data Lifecycle Management)	Стратегия доступа к архивным данным. Механизмы прозрачного доступа к данным, перемещенным в архив (например, через внешние таблицы в PostgreSQL postgres_fdw или табличные функции в ClickHouse для чтения из Yandex Cloud (Object Storage). Мониторинг и аудит эффективности политик DLM. Использование системных представлений и логов для отслеживания перемещения данных между уровнями хранения, оценки экономии средств и соблюдения регламентов. Процедуры восстановления данных из архивов (Data Retrieval): разработка регламентов и оценка временных затрат на извлечение "холодных"	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.

	данных для аудита или выполнения нестандартных аналитических запросов.	

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерный вариант контрольной работы

Цель контрольной работы: проверить умение применять теоретические знания для решения практических задач.

Тема 1. Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям

Задание (Аналитическое)

Вам поступило техническое задание на разработку системы для [конкретный сценарий, напр., «чата поддержки» или «системы сбора показаний с счетчиков»]. На основе приведенных требований (объем данных, RPS (Requests Per Second), требования к согласованности) заполните таблицу, обоснуйте выбор типа СУБД (реляционная, документная, ключ-значение) и предложите конкретную реализацию (например, PostgreSQL, Redis).

Тема 2. Архитектурные паттерны. Гибридные модели данных и их практическая реализация

Задание (Практическое)

В рамках проектирования базы данных интернет-магазина предусмотрена таблица products, содержащая поле attributes типа JSONB для хранения вариативных характеристик товаров. Требуется выполнить следующие задачи:

1) Сформировать SQL-запрос для выборки всех товаров, у которых в атрибутах присутствует характеристика «цвет» со значением «красный».

2) Разработать SQL-запрос для расчета средней цены товаров заданной категории с применением фильтрации по вложенному JSON-полю.

3) Проанализировать и обосновать выбор типа индекса для оптимизации производительности запроса.

Тема 3. Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища

Задание (Сравнительное)

Объясните, почему для создания дашборда в реальном времени по данным из предыдущего задания не подходит операционная PostgreSQL и требуется развернуть аналитический слой на ClickHouse. Сравните два подхода по 2-3 ключевым параметрам (скорость агрегации, нагрузка на источник, актуальность данных).

Тема 4. Управление жизненным циклом данных (Data Lifecycle Management)

Задание (Практическое)

Рассмотрим систему сбора метрик в PostgreSQL. Представьте, что вы разрабатываете систему для хранения метрик производительности серверов (PostgreSQL). В таблицу server_metrics каждую минуту поступают данные: server_id, cpu_usage, ram_usage, timestamp. Объем данных огромен, и вы решили использовать партиционирование.

1) По какому столбцу логичнее всего сделать партиционирование и почему?

2) Напишите SQL-запрос для создания основной таблицы server_metrics с декларативным партиционированием по диапазонам. Включая первичный ключ.

3) Напишите SQL-запрос для создания одной дочерней таблицы (партиции) для хранения данных за май 2024 года.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры информационных технологий Факультета информационных технологий и анализа больших данных.

Дополнительная информация

Примеры тестовых заданий

Тема 1. Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям

1. Для системы кеширования пользовательских сессий лучше подойдет:

а) PostgreSQL

б) Redis

в) ClickHouse

*Обоснуйте свой выбор в 1-2 предложениях.

2. Установите соответствие между типом СУБД и сценарием использования:

Графовая СУБД

Документная СУБД

Колоночная СУБД

Сценарии: А) Система рекомендаций "друзья друзей". Б) Аналитический отчет по продажам за 5 лет. В) Хранение профилей пользователей с произвольными атрибутами.

Тема 2. Архитектурные паттерны. Гибридные модели данных и их практическая реализация

3. Дан фрагмент JSONB в PostgreSQL: {"specs": {"ram": 8, "color": "black"}}. Какой оператор вернет значение ram?

а) attributes->'ram'

б) attributes->'specs'->>'ram'

в) attributes->'specs'->'ram'

4. Для эффективного поиска по любому ключу внутри JSONB-поля attributes необходимо создать:

а) B-tree индекс на все поле attributes

б) GIN-индекс на поле attributes

в) UNIQUE-индекс на поле attributes

Тема 3. Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища

5. Главное архитектурное отличие ClickHouse от PostgreSQL, обеспечивающее высокую скорость агрегаций, это:

а) Поддержка иностранных ключей

б) Колоночное хранение данных

в) Полная поддержка стандарта ACID

6. Выберите верные утверждения для ClickHouse (может быть несколько вариантов):

а) Оптимизирован для частых точечных UPDATE/DELETE

б) Эффективно выполняет запросы с GROUP BY по большим данным

в) Рекомендуется использовать как основную операционную БД для интернет-магазина

Тема 4. Управление жизненным циклом данных (Data Lifecycle Management)

7. Компания разрабатывает систему хранения IoT-данных с датчиков. Ежедневно поступает 5 млн записей. Политика хранения:

• Данные за последние 7 дней: доступ в реальном времени для мониторинга.

• Данные от 8 до 90 дней: доступ для аналитических запросов.

• Данные старше 90 дней: архивное хранение с доступом в течение часов.

Какая архитектура хранения в PostgreSQL наиболее эффективна?

а) Использовать партиционирование по времени.

б) Создать одну таблицу с обычными индексами и регулярно удалять старые данные.

в) Хранить все данные в одной таблице и использовать материализованные представления.

г) Создать отдельные базы данных для каждого временного периода.

8. Для системы логирования веб-приложения требуется автоматическое удаление записей старше 180 дней. Какое решение обеспечит наилучшую производительность?

а) В ClickHouse использовать правило TTL в определении таблицы.

б) В PostgreSQL настроить ежедневный cron-скрипт с DELETE запросом по дате.

б) В MongoDB создать TTL-индекс на поле created_at.

г) В любой СУБД создать триггер, проверяющий дату при вставке новых записей.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. *Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.*

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

ПКН-8 Способность создавать, описывать и качественно контролировать исполнение технологических требований и нормативных документов в профессиональной деятельности

1) Демонстрирует знания в области создания эффективных стратегий управления: командой, современными информационными системами и процессами разработки программных средств, а также знания нормативных документов в профессиональной деятельности

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: критически важные критерии выбора СУБД для различных доменов; принципы ранжирования требований к безопасности (шифрование, аутентификация) и надежности (репликация, отказоустойчивость) в зависимости от предметной области

Уметь: обосновать выбор между реляционной и документной моделью данных на основе сравнительного анализа их характеристик; аргументировать необходимость использования колоночных хранилищ (например, ClickHouse) для аналитических нагрузок, объясняя архитектурные различия OLTP и OLAP систем

Типовые контрольные задания

Разработка архитектурной рекомендации для стартапа

Контекст: К вам обратился руководитель стартапа FinStream, находящегося на стадии активного роста. Компания разрабатывает платформу для личных финансов, которая совмещает два ключевых функционала:

1. Операционная часть (OLTP): Высоконагруженное мобильное приложение. Характеризуется частыми операциями записи и чтения небольших объемов данных: регистрация транзакций, пополнение счетов, переводы. Требуется строгая согласованность данных (ACID).

2. Аналитическая часть (OLAP): Внутренний портал для аналитиков и система персональных отчетов. Предназначена для сложных агрегирующих запросов к большим объемам исторических данных (анализ трат по категориям, выявление паттернов).

Нужно проанализировать требования и предложить стратегию использования СУБД.

Задача 1. Выбор СУБД для операционной части

• Необходимо выбрать тип СУБД для OLTP-части. Рассматриваются кандидаты: классическая реляционная СУБД (например, PostgreSQL) и документная NoSQL СУБД (например, MongoDB).

Задача 2. Выбор решения для аналитической части

• Текущая операционная СУБД не справляется с аналитическими отчетами. Необходимо:

• Аргументировать внедрение отдельной колоночной СУБД (например, ClickHouse).

• Объяснить архитектурные различия между OLTP- и OLAP-системами, сравнив их по принципам хранения данных (строчное и колоночное) и оптимизации под тип нагрузки (запись и чтение).

2) Разрабатывает эффективные стратегии управления: командой разработчиков, современными информационными системами и процессами разработки программных средств

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: критерии выбора между классическими реляционными и гибридными решениями (например, PostgreSQL с JSONB); принципы оптимизации работы с полуструктурированными данными (индексы GIN/GIST для JSONB в PostgreSQL)

Уметь: разрабатывать архитектуру гибридной модели данных, определяя, какие данные хранить в реляционных таблицах, а какие в документных форматах; проектировать процесс репликации данных для аналитической отчетности, выбирая подходящие ETL/ELT-инструменты или встроенные механизмы СУБД

Типовые контрольные задания

Типовое задание: "Проектирование гибридной модели данных для маркетплейса"

Контекст: К вам обратился владелец маркетплейса "QuickBuy". Маркетплейс продает товары из сотен категорий.

Необходимо решить две проблемы:

1. Жесткость схемы: Добавление уникальных атрибутов для новых категорий товаров (например, "диагональ экрана", "размер одежды") требует постоянного изменения структуры БД.

2. Производительность: Операционная БД перегружена из-за аналитических отчетов.

Задача 1. Разработать гибридную модель данных

• Для сущности "Товар" в PostgreSQL необходимо использовать комбинацию реляционных таблиц и JSONB.

• Определите, какие данные о товаре (например, цена, категория) хранить в реляционных колонках, а какие динамические атрибуты — в JSONB. Обоснуйте выбор.

• Предложите, для каких полей внутри JSONB необходимо создать индексы (GIN/GIST) и почему.

Задача 2. Спроектировать процесс репликации для аналитики

• Спроектируйте процесс переноса данных из операционной БД (PostgreSQL) в аналитическое хранилище (ClickHouse).

• Какой механизм репликации или ETL/ELT-инструмент вы предложите (например, логическая репликация, Debezium, Airflow) и почему?

• Как будет обрабатываться JSONB-документ при переносе в колоночную БД?

3) Управляет командой разработчиков, современными информационными системами и процессами разработки программных средств с учетом технологических требований

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: методы оптимизации производительности через материализованные представления в аналитических СУБД; синтаксис и особенности применения TTL-индексов в MongoDB и ClickHouse

Уметь: управлять производительностью СУБД: создавать и обновлять материализованные представления в ClickHouse для ускорения отчетов; обеспечивать целостность данных в гибридных моделях на PostgreSQL, применяя индексы для JSONB и создавая ограничения для проверки структуры документов

Типовые контрольные задания

Оптимизация производительности и управление данными для сервиса аналитики

Контекст: Вы работаете с БД в сервисе WebAnalytics.

Необходимо решить две задачи:

1. ClickHouse: ключевые отчеты выполняются медленно; нужно автоматически удалять данные старше 2 лет.

2. PostgreSQL: в JSONB-поле tracking_settings попадают невалидные данные.

Задача 1. Оптимизация ClickHouse

• Для отчета "Статистика по страницам за месяц" создайте материализованное представление с предварительной агрегацией данных.

• Напишите запрос для его обновления.

• Настройте автоматическое удаление данных старше 2 лет с помощью TTL.

Задача 2. Целостность данных в PostgreSQL

• Для JSONB-поля tracking_settings добавьте CHECK-ограничение для проверки формата данных.

• Создайте GIN-индекс для ускорения запросов.

• Сравните синтаксис TTL в MongoDB и ClickHouse.

Примеры практико-ориентированных заданий

Построение системы оперативной аналитики для онлайн-кинотеатра

1. Контекст и бизнес-проблема: Вы работаете в онлайн-кинотеатре. Основная операционная система работает на PostgreSQL, где в реальном времени фиксируются все просмотры пользователей. Каждый запись в таблице video_plays содержит : user_id, movie_id, start_time, end_time, quality. Менеджеры по контенту жалуются, что получают отчеты о популярности фильмов только на следующий день. Потому они не могут понять в реальном времени, набирает ли новый сериал популярность, или в каком регионе сегодня вечером самый высокий спрос на комедии. Это мешает им оперативно управлять рекламными кампаниями и рекомендациями на главной странице.

2. Задача: Спроектировать и реализовать систему оперативной (near real-time) аналитики, которая позволяет:

В течение 1-2 минут видеть актуальную статистику по просмотрам.

Строить сложные аналитические отчеты (например, «средняя длительность просмотра по жанрам за последние 4 часа») без нагрузки на операционную БД.

3. Архитектура решения и роли СУБД. Ваша задача использовать обе СУБД по назначению, организовав поток данных между ними.

PostgreSQL (OLTP-система, источник правды). Роль PostgreSQL: принимает операционную нагрузку, запись каждого факта просмотра от приложения. Структура: таблица video_plays.

ClickHouse (OLAP-система, аналитическая платформа). Роль ClickHouse: принимает данные из PostgreSQL и позволяет выполнять по ним быстрые агрегирующие запросы. Структура: нужно спроектировать таблицу video_plays_analytics в ClickHouse, оптимизированную под частые запросы на агрегацию.

4. Практические шаги для реализации:

4.1 Проектирование и настройка

В ClickHouse создайте таблицу video_plays_analytics с движком MergeTree. Продумайте ключ сортировки (ORDER BY), чтобы ускорить запросы по дате и жанру. Партиционируйте данные, например, по месяцу.

В PostgreSQL создайте логическую реплику или настройте логирование для таблицы video_plays, чтобы данные можно было вычитывать.

4.2 Организация потока данных

Выберите и настройте инструмент для репликации данных (например, Debezium для захвата изменений из PostgreSQL).

Напишите конфигурацию, которая будет непрерывно отправлять новые записи из video_plays (PostgreSQL) в video_plays_analytics (ClickHouse).

Альтернатива (упрощенная версия): напишите Python-скрипт, который раз в минуту будет выбирать новые записи из PostgreSQL и пачками (INSERT ... VALUES (v1), (v2), ...) вставлять их в ClickHouse.

4.3 Решение бизнес-задач

Напишите SQL-запросы в ClickHouse, которые дадут ответы на вопросы менеджеров:

Топ-5 самых популярных фильмов за последние 3 часа (по количеству уникальных зрителей).

Средний процент просмотра фильма (на основе (end_time - start_time) / movie_duration) по жанрам за сегодня.

Динамика запусков просмотра (количество start_time) за последние 6 часов, сгруппированная по 15-минутным интервалам.

Примерные вопросы для подготовки к экзамену

Тема 1. Стратегия выбора СУБД. От бизнес-требований к техническим спецификациям

1. Назовите не менее пяти ключевых критериев для выбора между реляционными и нереляционными СУБД. Проиллюстрируйте на примере, как критерий «требования к согласованности данных» (ACID vs BASE) по-разному влияет на выбор СУБД для финтех-сервиса и платформы для сбора метрик IoT.

2. Как осуществляется ранжирование требований к безопасности (например, шифрование данных на rest (Data at rest encryption), управление доступом) при выборе СУБД для таких предметных областей, как интернет-банкинг, социальная сеть и сервис для проведения онлайн-опросов?

3. Какова цель создания чек-листа для первичной оценки СУБД и чем его содержание принципиально отличается от раздела «Требования к СУБД» в техническом задании (ТЗ) на разработку системы?

4. Какие стратегические и технические меры можно заложить в корпоративный стандарт выбора СУБД, чтобы минимизировать риски зависимости от вендора и обеспечить долгосрочную поддержку критически

5. При выборе СУБД часто упускают из виду операционные затраты (OPEX). Какие факторы, влияющие на общую стоимость владения (TCO), необходимо оценить помимо первоначальной лицензии, и как их формализовать в требованиях?

Тема 2. Архитектурные паттерны. Гибридные модели данных и их практическая реализация

1. В чем заключаются основные преимущества и недостатки чисто реляционной модели данных по сравнению с гибридной (реляционная + JSONB) при проектировании системы управления каталогом товаров с большим количеством вариативных атрибутов?

2. Назовите три конкретных признака или требования к данным, которые указывают на то, что гибридная модель (использование JSONB в PostgreSQL) будет более предпочтительна, чем классическая реляционная или документная СУБД.

3. Какие механизмы PostgreSQL позволяют обеспечить целостность и валидацию данных, хранящихся в полях типа JSONB? В какой момент сложность этих механизмов нивелирует преимущества гибридного подхода?

4. Для каких сценариев работы с данными в JSONB-поле в PostgreSQL предпочтительнее использовать GIN-индекс, а для каких B-tree индекс на выражение? Приведите примеры запросов для каждого случая.

5. Опишите два типичных антипаттерна использования JSONB в PostgreSQL, когда его применение неоправданно и данные лучше хранить в реляционном виде. К каким операционным проблемам это может привести?

Тема 3. Аналитические СУБД в современной архитектуре. ClickHouse и колоночные хранилища

1. Объясните принципиальные архитектурные различия между OLTP-системами (например, PostgreSQL) и OLAP-системами (например, ClickHouse). Почему прямое использование OLTP-базы для сложной аналитики на больших данных становится неэффективным?

2. В чем заключается основное преимущество колоночного формата хранения данных в ClickHouse для аналитических нагрузок? Объясните на конкретном примере запроса с агрегацией (например, SUM, COUNT).

3. Каковы основные различия между движками MergeTree и SummingMergeTree в ClickHouse? Для решения какой бизнес-задачи оптимален каждый из них?

4. Сравните два подхода к переносу данных из операционной PostgreSQL в аналитический ClickHouse: пакетная ежедневная выгрузка и потоковая репликация. Какие факторы (задержка, объем, надежность) влияют на выбор того или иного подхода?

5. Какую роль играют материализованные представления (Materialized Views) в оптимизации производительности ClickHouse? Опишите типичный сценарий их использования для предварительной агрегации данных.

Тема 4. Управление жизненным циклом данных (Data Lifecycle Management)

1. Опишите критерии классификации данных на "горячие", "теплые" и "холодные" в контексте DLM. Какие характеристики данных (частота доступа, требования к латентности, объем) влияют на отнесение к каждой категории?

2. Объясните архитектурные преимущества использования партиционирования в сочетании с табличными пространствами в PostgreSQL для реализации политик DLM.

3. Сравните механизмы TTL в MongoDB и ClickHouse с точки зрения реализации и производительности. В каких сценариях предпочтительнее использовать TTL в MongoDB, а в каких в ClickHouse?

4. Опишите процесс перехода данных между категориями "горячие" -> "теплые" -> "холодные" в распределенной системе хранения. Какие метрики и триггеры используются для автоматического перемещения данных между категориями?

5. Проанализируйте экономические аспекты классификации данных на "горячие", "теплые" и "холодные".

Пример экзаменационного билета

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

1. Теоретический вопрос **(15 баллов)**

Основные особенности нереляционных баз данных.

2. Теоретический вопрос **(15 баллов)**

В чем заключаются ключевые архитектурные особенности ClickHouse, которые делают его исключительно эффективным для выполнения аналитических запросов (OLAP), и как эти особенности противопоставляются принципам работы классических OLTP-СУБД, таких как PostgreSQL?

3. Практико-ориентированное задание **(30 баллов)**

В базе данных MongoDB «Библиотека» напишите запрос, выводящий перечень всех книг из раздела "Экономика и финансы", поступивших в 2024 году.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Григорьев, Ю. А. Реляционные базы данных и системы NoSQL [Электронный ресурс] / Григорьев Ю. А., Плутенко А. Д., Плужникова О. Ю. Благовещенск : АмГУ, 2018 424 с. Книга из коллекции АмГУ - Информатика https://e.lanbook.com/book/156492 ISBN 978-5-93493-308-2. [БИК ID: RU-LAN-BOOK-156492]

2. Парфенов, Юрий Павлович Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. Электрон. дан. Москва : Юрайт, 2025 97 с (Высшее образование) URL: https://urait.ru/bcode/559502 (дата обращения: 08.04.2025). Режим доступа: Электронно-библиотечная система Юрайт, для авториз. пользователей https://urait.ru/bcode/559502 ISBN 978-5-534-21173-3 : 439.00. [БИК ID: RU2fURAIT2f559502]

3. Советов, Борис Яковлевич Базы данных : учебник для вузов / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. 4-е изд., пер. и доп Электрон. дан. Москва : Юрайт, 2025 403 с (Высшее образование) URL: https://urait.ru/bcode/559898 (дата обращения: 01.09.2025). Режим доступа: Электронно-библиотечная система Юрайт, для авториз. пользователей https://urait.ru/bcode/559898 ISBN 978-5-534-18479-2 : 1959.00. [БИК ID: RU2fURAIT2f559898]

Дополнительная литература:

1. Кондрашов, Ю.Н. Язык SQL. Сборник ситуационных задач по дисциплине «Базы данных» : Учебно-практическое пособие / Ю.Н. Кондрашов Электрон. дан. Москва : Русайнс, 2024 125 с. Режим доступа: book.ru Internet access https://book.ru/book/953691 ISBN 978-5-466-06506-0. [БИК ID: RU\bookru\bibl\953691]

2. Мартишин, Сергей Анатольевич Базы данных.Практическое применение СУБД SQL и NoSOL-типа для применения проектирования информационных систем : Учебное пособие : Учебное пособие / Институт системного программирования Российской академии наук ; Российский государственный социальный университет 1 Москва : Издательский Дом "ФОРУМ", 2024 368 с. (Высшее образование) ВО - Бакалавриат https://znanium.com/catalog/document?id=434322

ISBN 978-5-8199-0946-1 ISBN 978-5-16-104936-5 (электр. издание) ISBN 978-5-16-018662-7 (ISBN соиздателя) . [БИК ID: RU\infra-m\znanium\bibl\2096940]

3. Стасышин, Владимир Михайлович Базы данных: технологии доступа : учебник для вузов / В. М. Стасышин, Т. Л. Стасышина. 2-е изд., испр. и доп Электрон. дан. Москва : Юрайт, 2025 164 с (Высшее образование) URL: https://urait.ru/bcode/562868 (дата обращения: 01.09.2025). Режим доступа: Электронно-библиотечная система Юрайт, для авториз. пользователей https://urait.ru/bcode/562868 ISBN 978-5-534-08687-4 : 749.00. [БИК ID: RU2fURAIT2f562868]

4. Мамедли, Р. Э. Большие данные и NoSQL базы данных [Электронный ресурс] : учебное пособие для вузов / Мамедли Р. Э.,Казиахмедов Т. Б. ; Казиахмедов Т. Б. Санкт-Петербург : Лань, 2024 92 с. Книга из коллекции Лань - Информатика https://e.lanbook.com/book/434051 ISBN 978-5-507-49873-4. [БИК ID: RU-LAN-BOOK-434051]

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

1. Электронная библиотека Финансового университета (ЭБ) http://elib.fa.ru/ (http://library.fa.ru/files/elibfa.pdf)

2. Электронно-библиотечная система BOOK.RU http://www.book.ru

3. Электронно-библиотечная система "Университетская библиотека ОНЛАЙН" http://biblioclub.ru/

4. http://znanium.com - ЭБС издательства «ИНФРА-М»

5. Образовательная платформа "ЮРАЙТ" https://urait.ru/

6. •Электронно-библиотечная система издательства Проспект http://ebs.prospekt.org/books

7. Электронно-библиотечная система издательства "Лань" https://e.lanbook.com/

8. MongoDB Documentation [Электронный ресурс]: сайт. – Режим доступа: https://www.mongodb.com/docs/

10. Методические указания для обучающихся по освоению дисциплины

Организация самостоятельной работы:

1. Самостоятельная работа студентов проходит внеаудиторно.

2. Рекомендуется распределять изучение материала равномерно в течение семестра.

Выполнение домашних заданий:

1. Домашние задания следует выполнять регулярно.

2. Каждое задание направлено на формирование конкретных практических навыков.

3. При возникновении трудностей рекомендуется обращаться к материалам лекций и рекомендованной литературе.

Рекомендуемый порядок работы:

1. Изучение теоретического материала по конспектам лекций.

2. Выполнение практических заданий по теме.

3. Самопроверка по контрольным вопросам.

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

Комплект лицензионного программного обеспечения:

1. Kaspersky
2. Пакет офисных программ

Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Гарант»
2. Информационно-правовая система «Консультант Плюс»
3. Система комплексного раскрытия информации «СКРИН» - http://www.skrin.ru/

Сертифицированные программные и аппаратные средства защиты информации:

1. не предусмотрены

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

1. **Компьютерный класс** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенный оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), персональные компьютеры, набор демонстрационного оборудования (проектор, экран)

2. **Учебная аудитория** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), набор демонстрационного оборудования (проектор, экран)