

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор высшей школы
программной инженерии
А.В. Малеев**

	Рабочая программа дисциплины (модуля)
по дисциплине:	Технологии баз данных. Часть 2
по направлению:	Программная инженерия
профиль подготовки:	Разработка программно-информационных систем высшая школа программной инженерии высшая школа программной инженерии
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 48 час.

Всего часов: 108, всего зач. ед.: 3

Программу составил: А.А. Жмуров, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании высшей школы программной инженерии 19.03.2025

Аннотация

Данный курс посвящён углублённому изучению современных технологий управления базами данных, включая распределённые, графовые, колонночные и in-memory СУБД. Основное внимание уделяется транзакциям в распределённых системах, методам оптимизации SQL-запросов, многопоточной обработке данных, а также анализу и применению промышленных решений.

1. Цели и задачи

Цель дисциплины

- освоение концепций, технологий и инструментов современных и графовых систем управления базами данных, распределённых транзакций, компиляции запросов и оптимизации SQL. Формирование навыков работы с высоконагруженными и распределёнными базами данных, включая изучение популярных промышленных решений.

Задачи дисциплины

- ознакомление с современными и графовыми СУБД, их архитектурой и применением;
- изучение механизмов транзакций в распределённых СУБД, включая протокол Paxos;
- освоение принципов компиляции запросов, включая JIT-компиляцию;
- разбор работы протокола TAPIR и принципов колонночных баз данных;
- формирование навыков оптимизации SQL-запросов, в том числе продвинутых методов;
- изучение многопоточных операторов SQL и их влияния на производительность;
- рассмотрение протокола Raft и аналитических систем MPP;
- анализ особенностей баз данных, работающих в оперативной памяти (Main Memory DB);
- ознакомление с архитектурой и особенностями разработки PostgreSQL;
- изучение промышленных СУБД (Cassandra, ScyllaDB, Tarantool, Picodata) и их применения;
- разбор особенностей распределённой масштабируемой отказоустойчивой СУБД YDB;
- изучение динамических таблиц YTsaurus и их применения в обработке больших данных.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-6 Способен разрабатывать алгоритмы и программы, пригодные для практического использования, применять основы информатики и программирования к проектированию, конструированию и тестированию программных продуктов	ОПК-6.1 Знает алгоритмы решения типовых задач, области и способы их применения
ПК-3 Способен проектировать, разрабатывать, интегрировать, проверять на работоспособность программное обеспечение	ПК-3.1 Различает синтаксис языков программирования, особенности программирования на этих языках, стандартные библиотеки языков программирования
	ПК-3.2 Умеет выбирать языки программирования для написания программного кода с учетом технического задания
	ПК-3.4 Знает, как определять оптимальные методы и средства проектирования программного обеспечения и структур данных
ПК-6 Способен разрабатывать и внедрять стандарты и процессы разработки, производить их мониторинг и обновления	ПК-6.1 Знает, как создавать стандарты и методологии разработки программного обеспечения в организации

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- архитектуру и принципы работы современных реляционных, графовых и распределённых СУБД;
- основные алгоритмы управления транзакциями, включая протоколы Paxos, Raft, TAPIR;
- принципы компиляции SQL-запросов, включая JIT-компиляцию;
- методы оптимизации SQL-запросов и их влияние на производительность;
- особенности колонночных и in-memory баз данных;
- структуру и внутреннюю архитектуру PostgreSQL;
- принципы работы промышленных СУБД (Cassandra, ScyllaDB, Tarantool, Picodata, YDB, YTsaurus).

уметь:

- проектировать и разрабатывать базы данных с учётом их производительности и масштабируемости;
- оптимизировать SQL-запросы, используя продвинутые методы;
- работать с многопоточными операторами SQL;
- применять распределённые транзакции в высоконагруженных системах;
- настраивать и администрировать различные СУБД, включая PostgreSQL и YDB;
- разрабатывать решения на основе графовых и колонночных баз данных;
- работать с динамическими таблицами в YTsaurus.

владеть:

- навыками анализа и выбора оптимальной СУБД для конкретных задач;
- инструментами профилирования и оптимизации запросов;
- методами мониторинга и диагностики производительности баз данных;
- практическими подходами к разработке отказоустойчивых и масштабируемых СУБД;
- технологиями работы с распределёнными базами данных в облачной инфраструктуре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Современные и графовые системы управления базами данных (СУБД)	2	2		3
2	Транзакция в распределённых СУБД. Протокол Paxos	2	2		3
3	Компиляция запросов. JIT-компиляция	2	2		3
4	Протокол TAPIR. Колонночные базы данных	2	2		3
5	Оптимизация SQL-запросов	2	2		4
6	Продвинутые методы оптимизации SQL-запросов	2	2		4
7	Многопоточные операторы SQL	2	2		4
8	Протокол Raft & MPP-аналитика	2	2		4
9	Базы данных Main memory	2	2		4
10	Разработка PostgreSQL	3	3		4
11	Обзор промышленных СУБД: Cassandra, ScyllaDB, Tarantool, Picodata	3	3		4
12	Распределённая масштабируемая отказоустойчивая СУБД с открытым исходным кодом YDB	3	3		4
13	Динамические таблицы YTsaurus	3	3		4
Итого часов		30	30		48

Подготовка к экзамену	0 час.
Общая трудоёмкость	108 час., 3 зач.ед.

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 6 (Весенний)

1. Современные и графовые системы управления базами данных (СУБД)

Изучение современных реляционных и нереляционных СУБД, их архитектуры, особенностей работы с графовыми базами данных и их применения в индустрии.

2. Транзакция в распределённых СУБД. Протокол Raхos

Разбор механизмов управления транзакциями в распределённых системах, консенсусных алгоритмов (Raхos), их преимуществ и ограничений при построении отказоустойчивых баз данных.

3. Компиляция запросов. JIT-компиляция

Принципы компиляции SQL-запросов, оптимизация исполнения с использованием JIT-компиляции, способы ускорения выполнения запросов.

4. Протокол TAPIR. Колонночные базы данных

Рассмотрение протокола TAPIR для управления транзакциями, сравнение с традиционными подходами, а также изучение архитектуры и преимуществ колонночных баз данных.

5. Оптимизация SQL-запросов

Основные методы оптимизации SQL-запросов, работа с индексами, планировщиками запросов и кэшем, практические примеры оптимизации.

6. Продвинутое методы оптимизации SQL-запросов

Углублённое изучение оптимизаций, включая перераспределение нагрузок, материализованные представления, партиционирование данных, адаптивные алгоритмы выполнения запросов.

7. Многопоточные операторы SQL

Принципы многопоточного выполнения SQL-запросов, параллелизм в обработке данных, балансировка нагрузки.

8. Протокол Raft & MPP-аналитика

Изучение алгоритма Raft для управления консенсусом, его применение в СУБД, а также принципы massively parallel processing (MPP) аналитики для работы с большими объёмами данных.

9. Базы данных Main memory

Преимущества и ограничения баз данных, работающих в оперативной памяти, их архитектура, применение в высокопроизводительных системах.

10. Разработка PostgreSQL

Введение в архитектуру PostgreSQL, настройка и расширение функциональности, модификация внутреннего ядра системы.

11. Обзор промышленных СУБД: Cassandra, ScyllaDB, Tarantool, Picodata

Сравнительный анализ популярных промышленных распределённых баз данных, их архитектурные решения, ключевые особенности и области применения.

12. Распределённая масштабируемая отказоустойчивая СУБД с открытым исходным кодом YDB

Изучение YDB как примера высоконагруженной распределённой базы данных, её архитектуры, возможностей и сценариев использования.

13. Динамические таблицы YTsauros

Основные принципы работы с YTsauros, особенности динамических таблиц, их применение для обработки больших данных в распределённых системах.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Аудитория должна быть оснащена:

- проектором или большим экраном для демонстрации материалов;
- компьютерами с предустановленными СУБД (PostgreSQL, ClickHouse) и инструментами для работы с базами данных;
- доступом в интернет для работы с документацией и удалёнными сервисами;
- доской (маркерной или классической) для объяснения концепций.

6. Перечень рекомендуемой литературы

Основная литература

Фонд библиотеки МФТИ:

1. Полтавцева, М. А. Безопасность баз данных : учебное пособие для вузов / М. А. Полтавцева. — Санкт-Петербург : Лань, 2024. — 356 с. — ISBN 978-5-507-49999-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/436274>
2. Разработка приложений на C# с использованием СУБД PostgreSQL / Васюткина И.А., Трошина Г.В., Бычков М.И. - Новосибирск : НГТУ, 2015. - 143 с.: ISBN 978-5-7782-2699-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/556925>
3. Бедердинова, О. И. Создание приложений баз данных в среде Visual Studio : учебное пособие / О.И. Бедердинова, Т.А. Минеева, Ю.А. Водовозова. — Москва : ИНФРА-М, 2021. — 94 с. - ISBN 978-5-16-109411-2. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1243816>
4. Голицына, О. Л. Базы данных : учебное пособие / О.Л. Голицына, Н.В. Максимов, И.И. Попов. — 4-е изд., перераб. и доп. — Москва : ФОРУМ : ИНФРА-М, 2023. — 400 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-00091-516-5. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1937956>
5. Затонский, А. В. Информационные технологии: разработка информационных моделей и систем : учебное пособие / А.В. Затонский. — Москва : РИОР : ИНФРА-М, 2023. — 344 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-369-01183-6. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1931479>
6. Кукарцев, В.В. Теория баз данных : учебник / В.В. Кукарцев, Р.Ю. Царев, О.А. Антамошкин. - Красноярск : Сиб. федер. ун-т, 2017. - 180 с. - ISBN 978-5-7638-3621-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1032103>

Дополнительная литература

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

<https://flutter.dev/>

<https://dart.dev/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Для проведения занятий потребуется оснащение аудитории компьютерами с доступом в интернет и проектором для демонстрации материалов.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Для успешного освоения дисциплины студенту рекомендуется:

- активно участвовать в обсуждениях, задавать вопросы преподавателям;
- практиковаться на реальных примерах, используя кейсы из индустрии;
- анализировать, как изученные технологии используются в современных IT-компаниях;
- развивать навык поиска информации в технической документации и open-source проектах;
- не бояться экспериментировать с настройками и конфигурациями СУБД – практический опыт наиболее ценен в изучении данной дисциплины.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Программная инженерия
профиль подготовки: Разработка программно-информационных систем
высшая школа программной инженерии
высшая школа программной инженерии
курс: 3
квалификация: бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Разработчик: А.А. Жмуров, канд. физ.-мат. наук, доцент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-6 Способен разрабатывать алгоритмы и программы, пригодные для практического использования, применять основы информатики и программирования к проектированию, конструированию и тестированию программных продуктов	ОПК-6.1 Знает алгоритмы решения типовых задач, области и способы их применения
ПК-3 Способен проектировать, разрабатывать, интегрировать, проверять на работоспособность программное обеспечение	ПК-3.1 Различает синтаксис языков программирования, особенности программирования на этих языках, стандартные библиотеки языков программирования
	ПК-3.2 Умеет выбирать языки программирования для написания программного кода с учетом технического задания
	ПК-3.4 Знает, как определять оптимальные методы и средства проектирования программного обеспечения и структур данных
ПК-6 Способен разрабатывать и внедрять стандарты и процессы разработки, производить их мониторинг и обновления	ПК-6.1 Знает, как создавать стандарты и методологии разработки программного обеспечения в организации

2. Показатели оценивания компетенций

В результате изучения дисциплины «Технологии баз данных. Часть 2» обучающийся должен:

знать:

- архитектуру и принципы работы современных реляционных, графовых и распределённых СУБД;
- основные алгоритмы управления транзакциями, включая протоколы Paxos, Raft, TAPIR;
- принципы компиляции SQL-запросов, включая JIT-компиляцию;
- методы оптимизации SQL-запросов и их влияние на производительность;
- особенности колонночных и in-memory баз данных;
- структуру и внутреннюю архитектуру PostgreSQL;
- принципы работы промышленных СУБД (Cassandra, ScyllaDB, Tarantool, Picodata, YDB, YTsaurus).

уметь:

- проектировать и разрабатывать базы данных с учётом их производительности и масштабируемости;
- оптимизировать SQL-запросы, используя продвинутые методы;
- работать с многопоточными операторами SQL;
- применять распределённые транзакции в высоконагруженных системах;
- настраивать и администрировать различные СУБД, включая PostgreSQL и YDB;
- разрабатывать решения на основе графовых и колонночных баз данных;
- работать с динамическими таблицами в YTsaurus.

владеть:

- навыками анализа и выбора оптимальной СУБД для конкретных задач;
- инструментами профилирования и оптимизации запросов;
- методами мониторинга и диагностики производительности баз данных;
- практическими подходами к разработке отказоустойчивых и масштабируемых СУБД;
- технологиями работы с распределёнными базами данных в облачной инфраструктуре.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Текущий контроль – тестирование по теоретическим темам, выполнение практических заданий.
Промежуточная аттестация – защита практических работ, анализ выполненных SQL-запросов, разбор ошибок.

Итоговый контроль включает комплексное задание с теоретическими и практическими аспектами, моделирование реальных ситуаций (настройка СУБД, работа с транзакциями, оптимизация запросов).

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Теоретические вопросы

1. Какие основные типы СУБД существуют и в чём их различия?
2. Каковы ключевые особенности графовых баз данных и их преимущества перед реляционными СУБД?
3. Что такое ACID и BASE? В каких ситуациях предпочтительно использование каждого подхода?
4. Как устроены индексные структуры в современных базах данных?
5. Как работает протокол Paxos и где он применяется в распределённых системах?
6. В чём отличие протокола Raft от Paxos? Какие преимущества он даёт?
7. Как организуется распределённая транзакция в современных СУБД?
8. В чём особенности протокола TAPIR и какие задачи он решает?
9. Как работает компиляция SQL-запросов и что такое JIT-компиляция?
10. Какие методы оптимизации SQL-запросов существуют?
11. Как влияет партиционирование таблиц на производительность запросов?
12. Какие преимущества даёт использование колонночных баз данных?
13. В чём отличия и особенности работы Cassandra, ScyllaDB, Tarantool, Picodata?
14. Как устроена архитектура PostgreSQL и какие её ключевые компоненты?
15. Как обеспечивается отказоустойчивость в YDB?
16. Какие сценарии использования характерны для YTsaurus?
17. Как реализуется многопоточная обработка SQL-запросов?
18. В чём особенности massively parallel processing (MPP) аналитики?
19. Как базы данных, работающие в оперативной памяти (main memory), отличаются от традиционных СУБД?

Практические задания

1. Оптимизация SQL-запроса

Дано несколько SQL-запросов. Определите их узкие места, предложите оптимизации, объясните выбор индексов.

2. Разработка распределённой транзакции

Реализуйте распределённую транзакцию с использованием двухфазного коммита (2PC) или аналогичного механизма.

3. Настройка индексирования

Оптимизируйте базу данных для эффективного выполнения выборки по сложным фильтрам и соединениям таблиц.

4. Работа с многопоточными операторами SQL

Напишите SQL-запрос, который выполняется с использованием многопоточной обработки данных, и измерьте его эффективность.

5. Проектирование отказоустойчивой архитектуры

Разработайте схему масштабируемой распределённой базы данных с высокой доступностью.

6. Работа с PostgreSQL

Настройте PostgreSQL, включите расширенные методы оптимизации запросов, проанализируйте производительность.

7. Развёртывание и тестирование промышленной СУБД

Установите и настройте одну из рассмотренных промышленных СУБД (например, YDB или ScyllaDB), проведите нагрузочное тестирование.

Критерии оценивания

- оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений
- оценка «хорошо (7)» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (6)» выставляется студенту, если он знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (5)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Дифференцированный зачет проводится по итогам текущей успеваемости и сдачи заданий и других видов работ, предусмотренных программой дисциплины и (или) путем организации специального опроса, проводимого в устной и (или) письменной форме.