

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

УТВЕРЖДЕНО
Заместитель директора

Ю.О. Соболев

| | |
|----------------------------|---|
| | Рабочая программа дисциплины (модуля) |
| по дисциплине: | Системы хранения и обработки данных |
| по направлению: | Прикладная математика и информатика |
| профиль подготовки: | Науки о данных |
| | центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" |
| | центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" |
| курс: | 1 |
| квалификация: | магистр |

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Аудиторных часов: 22 всего, в том числе:

лекции: 2 час.

семинары: 20 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 203 час.

Всего часов: 225, всего зач. ед.: 5

Программу составили:

К.А. Лапин, методист

С.Ю. Медведева, преподаватель

С.А. Геворгян, преподаватель

А.М. Куртасов, преподаватель

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и
онлайн-образования "Пуск" 01.03.2025

Аннотация

В этом курсе обучающиеся научатся различать разные типы баз данных (БД) и выбирать оптимальный вариант для своих проектов. Студенты освоят создание и управление БД, научатся извлекать данные из множества таблиц и выполнять сложные запросы с группировкой. Курс также включает знакомство с нереляционными базами данных (NoSQL) и практическое применение MongoDB для управления данными.

1. Цели и задачи

Цель дисциплины

- изучить принципы построения и разработки систем хранения и обработки данных;
- получить навыки настройки систем хранения данных, проектирования и разработки процесса наполнения систем хранения данных, реализации запросов к системам хранения данных.

Задачи дисциплины

- приобретение студентами навыков проектирования архитектур, применения специализированных инструментов и разработки программных систем для работы с данными.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

| Код и наименование компетенции | Индикаторы достижения компетенции |
|--|---|
| УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий | УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними |
| | УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации |
| | УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности |
| УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки | УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности |
| | УК-6.2 Оценивает свою деятельность, соотносит цели, способы и средства выполнения деятельности с её результатами |
| ОПК-1 Способен решать актуальные задачи фундаментальной и прикладной математики | ОПК-1.3 Понимает междисциплинарные связи в области прикладной математики и информатики и способен их применять при решении задач профессиональной деятельности |
| ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач | ОПК-2.1 Имеет представление о современном состоянии математических исследований в рамках тематической области своей профессиональной деятельности |
| | ОПК-2.2 Способен оценить актуальность и практическую значимость прикладных математических исследований в своей профессиональной области |
| | ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации |

| | |
|--|---|
| ОПК-3 Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности | ОПК-3.4 Способен самостоятельно приобретать, развивать и применять математические знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте |
| ОПК-4 Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности | ОПК-4.1 Умеет применять информационно-коммуникационные технологии для поиска и анализа профессиональной информации, выделения в ней главного, структурирования, оформления и представления в виде аналитических обзоров с обоснованными выводами и рекомендациями |
| | ОПК-4.2 Умеет применять знание информационно-коммуникационных технологий для решения поставленной задачи, формулирования выводов и оценки полученных результатов |
| | ОПК-4.3 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий |

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- основные этапы подготовки данных и основы работы с базами данных;
- существующие методы работы с данными, области их применения, достоинства и недостатки;
- основные цифровые технологии и алгоритмы их работы.

уметь:

- применять технологии отбора, обработки, интерпретации и анализа информации;
- отбирать технологии работы с информацией в зависимости от класса задач в области данных.

владеть:

- методами и технологиями обработки и представления информации;
- технологиями извлечения информации из различных источников.

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 | Введение в базы данных (БД) | 2 | 4 | | 29 |
| 2 | PostgreSQL как пример СУБД для работы с реляционными БД | | 2 | | 29 |
| 3 | Группировка данных и оконные функции | | 4 | | 29 |
| 4 | Нереляционные базы данных | | 2 | | 29 |
| 5 | Оркестрация | | 2 | | 29 |
| 6 | Airflow | | 2 | | 29 |
| 7 | Kafka | | 4 | | 29 |
| Итого часов | | 2 | 20 | | 203 |
| Подготовка к экзамену | | 0 час. | | | |
| Общая трудоёмкость | | 225 час., 5 зач.ед. | | | |

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

Семестр: 1 (Осенний)

1. Введение в базы данных (БД)

Понятие БД. Элементы реляционных БД. CAP теорема и нормализация. СУБД: понятие, функции. СУБД для реляционных БД. Основы работы с PostgreSQL. Типы операторов. Понятие Git.

2. PostgreSQL как пример СУБД для работы с реляционными БД

Простые выборки по условию. Выборки по нескольким условиям. Вложенные запросы. Множественное объединение. Типы объединений таблиц.

3. Группировка данных и оконные функции

Группировка данных и агрегатные функции. Оконные функции. Ранжирующие оконные функции. Агрегирующие оконные функции. Функции смещения. Аналитические функции. Временные таблицы. План запроса.

4. Нереляционные базы данных

Введение в концепцию NoSQL. Документориентированные базы данных. Базовые операции с данными в MongoDB. Подключение к MongoDB из Python. Объединение, группировка и агрегация данных в MongoDB.

5. Оркестрация

Extract, Transform, Load. Luigi. Задача. Параметры задач. Зависимости между задачами. Динамические зависимости. Планировщик.

6. Airflow

Docker и контейнеризация. Airflow: введение, архитектура и веб интерфейс. Создание пайплайна. Интеграция с другими системами. Best practices.

7. Kafka

Apache Kafka: введение, архитектура и интерфейс. Работа с Kafka. Интеграция с другими системами. Best practices.

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

ПО для видеосвязи, ноутбук с доступом в интернет

6. Перечень рекомендуемой литературы

Основная литература

1. Хранение информации в кибернетических устройствах [Текст], сб. ст./под ред. Л. П. Крайзера, -М., Советское радио, 1969
2. Обработка данных в управлении [Текст] : в 2-х т. Т.1/Р. Тироф , пер. с англ. с предисл. В. Я. Алтаева, -М., Мир, 1976

3. Введение в реляционные базы данных и язык SQL [Текст] / Т. М. Дадашев [и др.] ; Моск. физико-техн. ин-т (гос. ун-т .— Долгопрудный : МФТИ, 2002 .— 288 с.

Дополнительная литература

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

Не используются

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

1. Google Drive для доступа к материалам курса
2. Zoom
3. Ноутбук для участия в интерактивных занятиях

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

Успешное освоение курса требует большой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- проработку учебного материала (по конспектам лекций, видеоурокам, учебной и научной литературе);
- подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- доказательство отдельных утверждений, свойств;
- подготовку к практическим занятиям.

Промежуточный контроль знаний проводится в виде письменных опросов (мини-тестов) по теории и контрольных работ.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

| | |
|----------------------------|--|
| по направлению: | Прикладная математика и информатика |
| профиль подготовки: | Науки о данных центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" |
| курс: | <u>1</u> |
| квалификация: | магистр |

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Разработчики:

К.А. Лапин, методист
С.Ю. Медведева, преподаватель
С.А. Геворгян, преподаватель
А.М. Куртасов, преподаватель

1. Компетенции, формируемые в процессе изучения дисциплины

| Код и наименование компетенции | Индикаторы достижения компетенции |
|--|---|
| УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий | УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними |
| | УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации |
| | УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности |
| УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки | УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности |
| | УК-6.2 Оценивает свою деятельность, соотносит цели, способы и средства выполнения деятельности с её результатами |
| ОПК-1 Способен решать актуальные задачи фундаментальной и прикладной математики | ОПК-1.3 Понимает междисциплинарные связи в области прикладной математики и информатики и способен их применять при решении задач профессиональной деятельности |
| ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач | ОПК-2.1 Имеет представление о современном состоянии математических исследований в рамках тематической области своей профессиональной деятельности |
| | ОПК-2.2 Способен оценить актуальность и практическую значимость прикладных математических исследований в своей профессиональной области |
| | ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации |
| ОПК-3 Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности | ОПК-3.4 Способен самостоятельно приобретать, развивать и применять математические знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте |
| ОПК-4 Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности | ОПК-4.1 Умеет применять информационно-коммуникационные технологии для поиска и анализа профессиональной информации, выделения в ней главного, структурирования, оформления и представления в виде аналитических обзоров с обоснованными выводами и рекомендациями |
| | ОПК-4.2 Умеет применять знание информационно-коммуникационных технологий для решения поставленной задачи, формулирования выводов и оценки полученных результатов |
| | ОПК-4.3 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий |

2. Показатели оценивания компетенций

В результате изучения дисциплины «Системы хранения и обработки данных» обучающийся должен:

знать:

- основные этапы подготовки данных и основы работы с базами данных;
- существующие методы работы с данными, области их применения, достоинства и недостатки;
- основные цифровые технологии и алгоритмы их работы.

уметь:

- применять технологии отбора, обработки, интерпретации и анализа информации;
- отбирать технологии работы с информацией в зависимости от класса задач в области данных.

владеть:

- методами и технологиями обработки и представления информации;
- технологиями извлечения информации из различных источников.

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

Текущий контроль знаний – тестирование:

Раздел 1

1. Модель логического уровня, представляющая собой отображение логических связей между элементами данных безотносительно к среде хранения, называется:

а) инфологической; б) даталогической ; в) физической; г) семантической.

2. Формализованное описание предметной области называется:

а) инфологической; б) даталогической ; в) физической; г) семантической.

3. Модель, определяющая используемые запоминающие устройства и способы физической организации данных в среде хранения называется:

а) инфологической; б) даталогической ; в) физической; г) семантической.

4. Операция Select предназначена для:

а) извлечения определенных строк из таблицы; б) извлечения определенных столбцов из таблицы; в) соединения двух таблиц на основе значений в общих столбцах; г) поддержания целостности ссылок.

5. Операция Join предназначена для:

а) извлечения определенных строк из таблицы; б) извлечения определенных столбцов из таблицы; в) соединения двух таблиц на основе значений в общих столбцах; г) поддержания целостности ссылок.

6. Использование директивы Group by позволяет:

а) отсортировать записи по любому столбцу; б) задания критерия отбора возвращаемых запросов данных; в) сгруппировать все похожие строки результата выполнения запроса по значениям одного или нескольких столбцов; г) объединения нескольких таблиц вместе.

7. Использование директивы Order by позволяет:

а) отсортировать записи по любому столбцу; б) задания критерия отбора возвращаемых запросов данных; в) сгруппировать все похожие строки результата выполнения запроса по значениям одного или нескольких столбцов; г) объединения нескольких таблиц вместе.

8. Использование директивы Where позволяет:

а) отсортировать записи по любому столбцу; б) задания критерия отбора возвращаемых запросов данных; в) сгруппировать все похожие строки результата выполнения запроса по значениям одного или нескольких столбцов; г) объединения нескольких таблиц вместе.

9. Первичный ключ – это:

а) поле или совокупность полей, используемых для установления связей между таблицами; б) поле или совокупность полей обеспечивающих быстрый доступ к записям таблицы; в) поле или совокупность полей однозначно идентифицирующих запись; г) главное поле записи.

10. Кто не является пользователя базы данных:

а) администраторы; б) конечные пользователи; в) прикладные программисты; г) трансляторы.

11. Концептуальный уровень архитектуры:

а) представление информации базы данных такими, какие они есть на самом деле б) обобщенное представление пользователей в) индивидуальное представление пользователя г) представление в памяти

12. В реляционных системах данные рассматриваются пользователем как:

а) таблицы б) объекты и отношения между объектами в) набор древовидных структур г) компьютеризированная системы хранения записей

13. Репликация данных используется для:

а) изменения структуры БД; б) выполнения операций над записями; в) манипулирования объектами БД; г) создания копии БД на рабочей станции.

14. Что такое SQL-сервер:

а) система управления базами данных; б) язык для разработки СУБД; в) язык описания сетевой модели; г) язык манипулирования данными

Примеры контрольных работ:

1. Написать запрос на создание БД, включающей не менее 3 таблиц
2. Написать SQL запрос на выборку данных из заданной БД
3. Написать запрос, включающий сортировку по заданному полю
4. Написать запрос на удаление записей из таблицы

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

Перечень вопросов к дифференцированному зачету:

1. Архитектура «файл-сервер»
2. Архитектура «клиент-сервер»
3. Архитектура «трехзвенный клиент-сервер»
4. Основные этапы проектирования данных
5. Концептуальная модель данных
6. Представление концептуальной модели моделью БД
7. Сетевая модель данных
8. Иерархическая модель данных
9. Основные понятия реляционной модели данных
10. Многомерная модель данных
11. Элементы реляционной алгебры. Объединение.
12. Элементы реляционной алгебры. Декартово произведение
13. Элементы реляционной алгебры. Пересечение.
14. Элементы реляционной алгебры. Деление.
15. Нормализация БД.
16. Представление данных во внутренней и внешней памяти компьютера
17. Использование индексирования для хранения и выборки данных
18. Использование деревьев индексов
19. Построение связей между таблицами в реляционных БД
20. Типы запросов в реляционных БД.
21. Использование условий и функций в запросах
22. Использование групповых операций в запросах
23. Конструирование многотабличных запросов в БД
24. Конструирование запросов на выборку с помощью языка SQL
25. Фильтрация данных с помощью языка SQL
26. Сортировка данных с помощью языка SQL

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

Итоговый контроль осуществляется в виде письменного дифференцированного зачета на lms.
Время отведенное на дифференцированный зачет: 2 академических часа.