

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»
Обнинский институт атомной энергетики –
филиал федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(ИАТЭ НИЯУ МИФИ)

Утверждено на заседании
УМС ИАТЭ НИЯУ МИФИ
Протокол №2-8/2024 От 30.08.2024

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

SQL для анализа данных

Шифр, название дисциплины

01.04.02 «Прикладная математика и информатика»

Шифр, название специальности/направления подготовки

Математическое моделирование и прикладной анализ данных

Название программы магистратуры

магистр

(Квалификация (степень) выпускника)

Форма обучения: очная

г. Обнинск 2024 г.

Программа составлена в соответствии с требованиями образовательного стандарта высшего образования национального исследовательского ядерного университета «МИФИ» по направлению подготовки 01.04.02 – Прикладная математика и информатика. (квалификация (степень) магистр).

Программу составил:

_____ С.В. Ермаков, доцент, к.ф.-м.н, доцент

Рецензент:

_____ Г.Е. Деев, доцент, к.ф.-м.н, доцент

Программа рассмотрена на заседании ОИКС

(протокол № 5/7 от «30» июля от 2024 г.)

Руководитель направления подготовки 01.04.02
«Прикладная математика и информатика»

_____ Ермаков С.В.

« ____ » _____ 2024 г.

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения ООП магистратуры обучающийся должен овладеть следующими результатами обучения по дисциплине:

Коды компетенций	Результаты освоения ООП <i>Содержание компетенций*</i>	Перечень планируемых результатов обучения по дисциплине**
ОПК-1	Способен решать актуальные задачи фундаментальной и прикладной математики	З-ОПК-1 Знать актуальные задачи фундаментальной и прикладной математики, методы математического моделирования. У-ОПК-1 Уметь использовать методы математического моделирования для решения задач фундаментальной и прикладной математики. В-ОПК-1 Владеть методами математического моделирования и основами их использования.
ОПК-2	Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	З-ОПК-2 Знать основные понятия, математические методы решения прикладных задач, принципы математического моделирования и методы верификации. У-ОПК-2 Уметь применять полученную теоретическую базу для решения практических задач В-ОПК-2 Владеть основными математическими методами решения прикладных задач
ОПК-3	Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	З-ОПК-3 Знать основные методы и принципы математического моделирования, методы построения математических моделей типовых профессиональных задач, способы нахождения решений математических моделей и содержательной интерпретации полученных результатов. У-ОПК-3 Уметь составлять математические модели типовых профессиональных задач, находить способы их решения и профессионально интерпретировать смысл полученного результата. В-ОПК-3 Владеть методами построения математических моделей типовых профессиональных задач, способами нахождения решений математических моделей и содержательной интерпретации полученных результатов

2. Место дисциплины в структуре ООП магистратуры

Дисциплина реализуется в рамках вариативной части.

Для освоения дисциплины необходимы компетенции, сформированные в рамках изучения следующих дисциплин: Математическая статистика для анализа данных, Базовый Python + объектно-ориентированное программирование (ООП) в Python

Дисциплина изучается на 1 курсе в 2 семестре.

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся

Общая трудоемкость (объем) дисциплины составляет 2 зачетных единиц (з.е.), 72 академических часа.

3.1. Объем дисциплины по видам учебных занятий (в часах)

	Семестр		
	№ 1	№ 2	Всего
	Количество часов на вид работы:		
Контактная работа обучающихся с преподавателем			
Аудиторные занятия (всего)		64	64
В том числе:			
<i>лекции</i>		32	32
<i>практические занятия</i>		32	32
<i>лабораторные занятия</i>			
Промежуточная аттестация			
В том числе:			
<i>зачет</i>		Зачет	Зачет
<i>экзамен</i>		-	-
Самостоятельная работа обучающихся (всего)		8	8
В том числе:			
<i>проработка учебного (теоретического) материала</i>			
<i>выполнение индивидуальных заданий</i>			
<i>подготовка ко всем видам контрольных испытаний текущего контроля успеваемости (в течение семестра)</i>			
<i>подготовка ко всем видам контрольных испытаний промежуточной аттестации (по окончании семестра)</i>		8	8
<i>Всего (часы):</i>		72	72
<i>Всего (зачетные единицы):</i>		2	2

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

**4.1. Разделы дисциплины и трудоемкость по видам учебных занятий
(в академических часах)**

№ п/п	Наименование раздела /темы дисциплины	Общая трудоём- кость всего (в часах)	Виды учебных занятий, включая самостоятельную работу обучающихся и трудоёмкость (в часах)				Формы текущего контроля успевае- мости
			Аудиторные учебные занятия			СРО	
			Лек	Сем/Пр	Лаб		
1.			32	32	-	8	
1.1.	Введение в базы данных	4	2	2		-	Практическое задание
1.2.	Основы работы с БД	4	2	2	-	-	Тест
1.3.	Первые шаги в SQL	4	2	2	-	-	ДЗ
1.4.	Как я умею гуглить	4	2	2	-	-	ДЗ
1.5.	JOIN: Объединение таблиц – важнейшая операция в РСУБД	6	3	3		-	ДЗ
1.6.	GROUP BY: расчет агрегированных метрик и описательных статистик	6	3	3	-	-	ДЗ
1.7.	Подзапросы, CTE, correlated subquery и вложенная логика	6	3	3	-	-	Практика
1.8.	Оконные функции: самый мощный инструмент для решения задач	6	3	3	-	-	ДЗ
1.9.	Код стайл: как писать красивый код на SQL	6	3	3		-	практика
1.10.	DDL и DML: не SELECT-ом единым	6	3	3		-	практика
1.11.	Финальный тест по модулю SQL	7	3	3	-	5	тест
1.12.	Финальный проект по модулю SQL	9	3	3		3	Финальный проект

4.2. Содержание дисциплины, структурированное по разделам (темам)

Лекционный курс

№	Наименование раздела /темы дисциплины	Содержание
1.1.	Введение в базы данных	как устроена база данных зачем вообще лезть в базы данных Важность нормализации данных
1.2.	Основы работы с БД	группы операторов (DML, DDL, TCL, DCL) все основные типы данных, с которыми нам предстоит работать (числа, строки, даты и т.д.)
1.3.	Первые шаги в SQL	базовые операторы SQL-запросов: SELECT, WHERE, FROM и другие скалярные функции: обработку даты-времени, функции для работы с текстом и кастинга и т.д операторы COALESCE, NULLIF, CASE и не только
1.4.	Как я умею гуглить	Спец. Урок, который поможет научиться искать информацию
1.5.	JOIN: Объединение таблиц – важнейшая операция в РСУБД	Объединение таблиц Использование джоинов
1.6.	GROUP BY: расчет агрегированных метрик и описательных статистик	Группировка - основная и мощная операция в SQL функции MAX, SUM, AVG и необычные функции/операторы и сценарии их применения
1.7.	Подзапросы, CTE, correlated subquery и вложенная логика	Использование подзапросов в фильтре WHERE Коррелированные подзапросы Подзапросы в блоке FROM Обобщенные табличные выражения CTE Как провести ABC, XYZ анализы и посчитать сочетаемость товаров
1.8.	Оконные функции: самый мощный инструмент для решения задач	Что такое оконные функции и когда они используются Какие бывают оконные функции Какие подводные камни есть при использовании окон

		Как решать типовые задачи с помощью оконных функций Как использовать агрегатные функции с окнами
1.9	Код стайл: как писать красивый код на SQL	«внешний вид» написанного кода
1.10	DDL и DML: не SELECT-ом единым	Как создавать таблицы Как вставлять данные в таблицы Как изменять структуру таблиц
1.11	Финальный тест по модулю SQL	Тест по пройденному материалу по базовым возможностям языка SQL и проектированию баз данных.
1.12	Финальный проект по модулю SQL	Финальный проект на знание SQL

Практические/семинарские занятия

№	Наименование раздела /темы дисциплины	Содержание
1.1.	Введение в базы данных	как устроена база данных зачем вообще лезть в базы данных Важность нормализации данных
1.2.	Основы работы с БД	группы операторов (DML, DDL, TCL, DCL) все основные типы данных, с которыми нам предстоит работать (числа, строки, даты и т.д.)
1.3.	Первые шаги в SQL	базовые операторы SQL-запросов: SELECT, WHERE, FROM и другие скалярные функции: обработку даты-времени, функции для работы с текстом и кастинга и т.д операторы COALESCE, NULLIF, CASE и не только
1.4.	Как я умею гуглить	Спец. Урок, который поможет научиться искать информацию
1.5.	JOIN: Объединение таблиц – важнейшая операция в РСУБД	Объединение таблиц Использование джоинов

1.6.	GROUP BY: расчет агрегированных метрик и описательных статистик	Группировка - основная и мощная операция в SQL функции MAX, SUM, AVG и необычные функции/операторы и сценарии их применения
1.7.	Подзапросы, CTE, correlated subquery и вложенная логика	Использование подзапросов в фильтре WHERE Коррелированные подзапросы Подзапросы в блоке FROM Обобщенные табличные выражения CTE Как провести ABC, XYZ анализы и посчитать сочетаемость товаров
1.8.	Оконные функции: самый мощный инструмент для решения задач	Что такое оконные функции и когда они используются Какие бывают оконные функции Какие подводные камни есть при использовании окон Как решать типовые задачи с помощью оконных функций Как использовать агрегатные функции с окнами
1.9	Код стайл: как писать красивый код на SQL	«внешний вид» написанного кода
1.10	DDL и DML: не SELECT-ом единым	Как создавать таблицы Как вставлять данные в таблицы Как изменять структуру таблиц
1.11	Финальный тест по модулю SQL	Тест по пройденному материалу по базовым возможностям языка SQL и проектированию баз данных.
1.12	Финальный проект по модулю SQL	Финальный проект на знание SQL

Лабораторные занятия

Не предусмотрены.

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

В качестве учебно-методических материалов используется рекомендованная литература.

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

6.1. Паспорт фонда оценочных средств по дисциплине

№ п/п	Контролируемые разделы (темы) дисциплины (результаты по разделам)	Код контролируемой компетенции (или её части) / и ее формулировка	Наименование оценочного средства
1.3	Оконные функции	ОПК-3	Контрольная работа № 1
2.1.	GROUP BY: расчет агрегированных метрик и описательных статистик	ОПК-2	Контрольная работа № 1
2.3.	Основы работы с БД	ОПК-1	Контрольная работа № 2
2.4.	Объединение таблиц	ОПК-1	Контрольная работа № 2

6.2. Типовые контрольные задания или иные материалы

6.2.1. Зачет

В билете два теоретических вопроса и один практический

Теоретические вопросы билета:

1. Что такое база данных и какие основные виды баз данных существуют?
2. Чем отличается реляционная база данных от нереляционной?
3. Каковы основные принципы нормализации базы данных?
4. Что такое первичный ключ и внешние ключи в реляционных базах данных?
5. Какие операторы SQL используются для выборки данных из таблицы?
6. Что такое SQL-запрос и как его выполнить на практике?
7. Как использовать оператор SELECT для выборки данных из одной таблицы?
8. Как с помощью SQL фильтровать данные с помощью оператора WHERE?
9. Как объединить данные из нескольких таблиц с помощью JOIN? Объясните основные виды JOIN.
10. Что такое LEFT JOIN, RIGHT JOIN и INNER JOIN, и как они отличаются друг от друга?
11. Как использовать оператор GROUP BY для группировки данных? Приведите пример с агрегатными функциями.
12. Что такое агрегатные функции в SQL? Приведите примеры.
13. Как использовать подзапросы в SQL? Чем отличается подзапрос в SELECT от подзапроса в WHERE?
14. Что такое CTE (Common Table Expressions) и как их использовать в SQL-запросах?
15. Объясните, что такое correlated subquery и как его использовать в SQL.
16. Что такое оконные функции в SQL и как они могут улучшить анализ данных?
17. Какие задачи решаются с помощью оконных функций? Приведите примеры.
18. Почему важно соблюдать код стайл при написании SQL-запросов? Какие основные принципы хорошего кодирования в SQL?
19. Чем отличаются DDL (Data Definition Language) и DML (Data Manipulation Language)? Приведите примеры команд.
20. Как с помощью SQL создать таблицу, изменить её структуру или удалить?

Критерий оценки – правильность и полнота ответа на вопросы. Оценка выставляется по шкале от 0 до 40 баллов: теоретические вопросы – 30 баллов, 10 баллов – дополнительные вопросы. Зачет считается сданным при оценке не ниже 25 баллов.

6.2.2. Контрольная работа № 1

Нужно посмотреть, как менялось количество уникальных активных пользователей с течением времени. Потому что люди могли регистрироваться и не заходить потом. Или же задачи могли решать постоянно одни и те же люди.

На основании того, как пользователи решают задачи (таблицы coderun и codesubmit), найдите:

- ymd - дата активности
- unique_cnt - количество уникальных пользователей в период с 01.01.2022 по заданный день

Примечание: Здесь учитываем только те дни, в которые активность реально была.

6.2.2. Контрольная работа № 2

Необходимо для каждой тмс группы вывести 3 наиболее дорогих транзакции за 2019 и 2020 год. Если в рамках какой-то тмс группы за год было менее 3 транзакций, то недостающие места должны быть дополнены null.

В результате необходимо вывести столбцы:

- group_name - имя группы
- year - год
- rn - ранг
- transaction_value - сумма транзакции

Результат должен быть отсортирован по имени группы, рангу и году по возрастанию всех полей. Все суммы округлите до 2 знака после запятой.

б) критерии оценивания компетенций (результатов) – правильная работа кода программы, понимание алгоритма метода оптимизации, умение вывести необходимые для алгоритма формулы.

в) описание шкалы оценивания:

Каждая задача оценивается по шкале от 0 до 10 баллов.

Контрольная работа считается выполненной успешно при суммарной оценке не ниже 18 баллов.

6.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Форма аттестации	Наименование оценочного средства	Баллы
Зачет (100 баллов)	Контрольная работа № 1	30
	Контрольная работа № 2	30
	Ответы на билет	40

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

а) основная учебная литература:

1. Алекс Петров: Распределенные данные. Алгоритмы работы современных систем хранения информации
2. Новиков Б. А. Основы технологий баз данных: учеб. пособие / Б. А. Новиков, Е. А. Горшкова, Н. Г. Графеева; под ред. Е. В. Рогова. — 2-е изд. — М.: ДМК Пресс, 2020. — 582 с.
- Рогов Е. В. PostgreSQL 16 изнутри. — М.: ДМК Пресс, 2024. — 664 с.
- Efficient MySQL Performance by Daniel Nichter Released November 2021 Publisher(s): O'Reilly Media, Inc.
- Энтони Молинаро, Роберта де Грааф «SQL. Сборник рецептов» 2021

б) дополнительная учебная литература:

-

8. Перечень ресурсов* информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»), необходимых для освоения дисциплины

-

9. Методические указания для обучающихся по освоению дисциплины

Вид учебного занятия	Организация деятельности студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.
Практические занятия	Проработка рабочей программы, уделяя особое внимание целям и задачам, структуре и содержанию дисциплины. Работа с конспектом лекций, просмотр рекомендуемой литературы. Изучение выбранной предметной области на примерах решения задач семинарских занятий, индивидуальных домашних заданий.
Курсовая работа	Не предусмотрена
Контрольная работа	Ознакомиться с основной и дополнительной литературой, включая справочные издания, зарубежные источники, основополагающие термины. Попрактиковаться в решении аналогичных домашних задач по всем темам контрольных работ.
Лабораторная работа	Не предусмотрена.
Подготовка к зачету	При подготовке к зачету необходимо ориентироваться на конспекты лекций и рекомендуемую литературу.

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

Издательская система LaTeX для подготовки докладов, презентаций и учебного материала.

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

Видеопроектор, компьютер, издательская система LaTeX для подготовки докладов, презентаций и учебного материала.

12. Иные сведения и (или) материалы

12.1. Перечень образовательных технологий, используемых при осуществлении образовательного процесса по дисциплине

Часов в интерактивной форме – 8.

В ходе практических занятий происходит публичное обсуждение каждой решаемой задачи. При этом студенты высказывают свои мнения по выбору наиболее простого способа поиска оптимального решения.

После решения домашних работ на консультациях проводится разбор допущенных студентами ошибок.

12.2. Формы организации самостоятельной работы обучающихся (темы, выносимые для самостоятельного изучения; вопросы для самоконтроля; типовые задания для самопроверки)

Некоторые темы изучаются студентами самостоятельно. Для изучения используется приведённая в списке основная и дополнительная литература. Контроль освоения материала осуществляется при проверке контрольных работ, домашнего задания и на зачете.

№	Тема и часть, изучаемая (осваиваемая) самостоятельно
1.1	Индексы в базах данных
1.2	Транзакции в реляционных базах данных
1.3	Нормализация и денормализация данных
1.4.	Репликация и масштабирование баз данных
1.5	Планирование и оптимизация запросов
1.6	Моделирование данных
1.7	Управление правами доступа в базах данных
1.8	Резервное копирование и восстановление данных
1.9	NoSQL базы данных
1.10	Индексы в базах данных

Вопросы и задания для самоконтроля по всем темам:

1. Что такое база данных и какие типы баз данных существуют?
2. Каковы основные операции, которые можно выполнять с базой данных?
3. Чем отличается реляционная база данных от нереляционной?
4. Какие SQL-операторы используются для выбора данных из базы?
5. Что такое SQL-запрос и как его правильно формировать?
6. Как осуществляется фильтрация данных с помощью оператора WHERE?
7. Что такое JOIN в SQL и какие его типы существуют?
8. Как работает оператор GROUP BY и какие агрегатные функции можно использовать с ним?
9. Как использовать подзапросы в SQL и чем они могут быть полезны?
10. Что такое CTE (Common Table Expression) и в чём его преимущества?
11. В чём отличие между correlated subquery и обычным подзапросом?
12. Какие задачи решаются с помощью оконных функций в SQL?

13. Какие примеры оконных функций ты знаешь и как они применяются на практике?
14. Почему важно соблюдать код-стайл при написании SQL-запросов?
15. Какие команды относятся к DDL, а какие — к DML?

12.3. Краткий терминологический словарь

JOIN	операция в SQL, используемая для объединения данных из двух или более таблиц на основе общего столбца. Существуют различные виды JOIN, такие как INNER JOIN, LEFT JOIN, RIGHT JOIN, и FULL OUTER JOIN, которые определяют, какие строки будут включены в результат объединения.
GROUP BY	оператор SQL, который используется для группировки строк в таблице по одному или нескольким столбцам. Он часто используется в сочетании с агрегатными функциями (например, SUM, AVG, COUNT), чтобы вычислять статистики для каждой группы.