

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
Московский государственный университет имени М.В. Ломоносова  
Факультет вычислительной математики и кибернетики

**УТВЕРЖДАЮ**  
декан факультета вычислительной  
математики и кибернетики

  
/И.А. Соколов /  
«27» сентября 2023г.

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

по дисциплине

**Базы данных для машинного обучения**

---

**Уровень высшего образования:**

**бакалавриат**

**Направление подготовки / специальность:**

**02.03.02 "Фундаментальная информатика и информационные технологии" (3++)**

**Направленность (профиль) ОПОП:**

**Искусственный интеллект и анализ данных**

**Форма обучения:**

**очная**

Рассмотрен и утвержден

*на заседании Ученого совета факультета ВМК*

*(протокол №7, от 27 сентября 2023 года)*

Москва 2023

## 1. ФОРМЫ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В процессе и по завершении изучения дисциплины оценивается формирование у студентов следующих компетенций:

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-7. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	<p>ПК-7.1. Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях</p> <p>ПК-7.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения</p>	<p>ПК-7.1.3-1. Знает виды представления данных, методы поиска и парсинга данных.</p> <p>ПК-7.1. 3-2. Знает уровни представления данных (ODS, DDL, семантический слой, модель данных)</p> <p>ПК-7.1. 3-3. Знает основные инструменты, библиотеки и технологии Data Science</p> <p>ПК-7.1. У-1. Умеет отделять достоверные источники данных от сомнительных, осуществлять критических отбор данных, проверять их на целостность и непротиворечивость</p> <p>ПК-7.1. У-2. Умеет использовать инструменты и библиотеки для Data Science для поиска данных в открытых источниках, специализированных библиотеках и репозиториях</p> <p>ПК-7.2. 3-1. Знает методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных</p> <p>ПК-7.2. 3-2. Знает методы планирования вычислительного эксперимента, формирования обучающей и контрольной выборок</p> <p>ПК-7.2. У-1. Умеет выявлять и исключать из массива данных ошибочные данные и выбросы</p>

		<p>ПК-7.2. У-2. Умеет выделять входные и выходные переменные с целью использования предиктивных моделей</p> <p>ПК-7.2. У-3. Умеет осуществлять разметку структурированных и неструктурированных данных</p> <p>ПК-7.2. У-4. Умеет использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения</p> <p>ПК-7.2. У-5. Умеет использовать методы и технологии массово параллельной обработки и анализа данных</p>
--	--	--

### 1.1. Текущий контроль успеваемости

Текущий контроль успеваемости осуществляется путем оценки результатов выполнения заданий практических (семинарских) занятий, самостоятельной работы, предусмотренных учебным планом и посещения занятий/активность на занятиях.

В качестве оценочных средств текущего контроля успеваемости предусмотрены:

#### Практические занятия

1. Построение нормализованных OLTP схем. Операции SQL DML для OLTP. Работа с транзакциями.
2. Построение OLAP схем с нормализованными и ненормализованными измерениями.
3. Генерация данных. Загрузка больших объемов данных. Загрузка данных в PostgreSQL.
4. Преобразование данных из OLTP схемы в OLAP схему. Полная начальная загрузка. Инкрементная загрузка данных.
5. Оптимальное индексирование атрибутов. Практикум назначения индексов. Исследование ускорения выполнения запросов по индексированным атрибутам.
6. Построение и отладка SQL- запросов с функциями агрегирования данных. Анализ плана выполнения запроса.
7. Визуализация данных в системах Data Warehouse

#### Примеры тестовых заданий

Вопрос №1. Укажите характеристики OLTP-системы

1. Высокая степень нормализации схемы данных
2. Небольшое количество длительных запросов на чтение данных
3. Отсутствие индексов для обеспечения максимально быстрого обновления данных
4. Большое количество одновременно работающих пользователей

Вопрос №2. Дана таблица

test(i number, j number, s varchar2(100))

Выберите индексы, которые могут повысить скорость выполнения SQL запроса:

select \* from test where j between 10 and 20 and s like '%data'

1. Create index I\_TEST1 on test(i, j)
2. Create index I\_TEST2 on test(j)
3. Create index I\_TEST3 on test(s)
4. Create index I\_TEST4 on test(j, s)

Вопрос №3. Выберите из списка атрибуты, которые могут быть представлены в виде иерархических измерений в многомерном OLAP-кубе, хранящим факты продаж:

1. Имя покупателя
2. Дата продажи товара
3. Классификатор товаров
4. Количество проданного товара
5. Адрес торговой точки

Вопрос №4. Укажите количество таблиц в реляционной базе данных, требуемых для хранения трех сущностей (A, B, C) и связей «многие-ко-многим» между A и B, и «один-ко-многим» между B и C:

Вопрос №5. Какие из следующих операторов SQL могут уменьшить количество-строк в результирующей выборке:

1. ORDER BY
2. GROUP BY
3. HAVING
4. INNER JOIN
5. UNION ALL

Вопрос №6. Укажите корректные SQL выражения для таблицы PERSON(ID INTEGER, NAME VARCHAR(100)):

1. SELECT ID, NAME FROM PERSON ORDER BY NAME WHERE NAME LIKE '%a% '
2. SELECT ID, NAME FROM PERSON GROUP BY NAME
3. SELECT ID, NAME FROM PERSON GROUP BY NAME WHERE COUNT(\*) > 1
4. SELECT NAME FROM PERSON WHERE NAME like '%a% ' GROUP BY NAME HAVING COUNT(\*) > 1
5. SELECT ID, NAME FROM PERSON P1 INNER JOIN PERSON P2 ON (P1.NAME = P2.NAME)

Вопрос №7. В таблице лежат объемы продаж по продуктам и по годам SALES(PRODUCT, YEAR, AMOUNT). Укажите выражение оконной аналитической функции, которое позволяет вычисляет разницу объема продаж продукта по сравнению с предыдущим годом:

1. LEAD(AMOUNT) - LAG(AMOUNT) OVER (PARTITION BY PRODUCT ORDER BY YEAR)
2. SALES - LAG(AMOUNT, 1, AMOUNT) OVER (PARTITION BY PRODUCT ORDER BY YEAR)
3. SALES - LAG(AMOUNT) OVER (ORDER BY PRODUCT, YEAR)
4. SALES - LAST(AMOUNT) OVER (PARTITION BY PRODUCT ORDER BY YEAR DESC)

Вопрос №8. Сколько различных группировок будет выполнено при указании условия группировки GROUP BY CUBE(A, B, C)

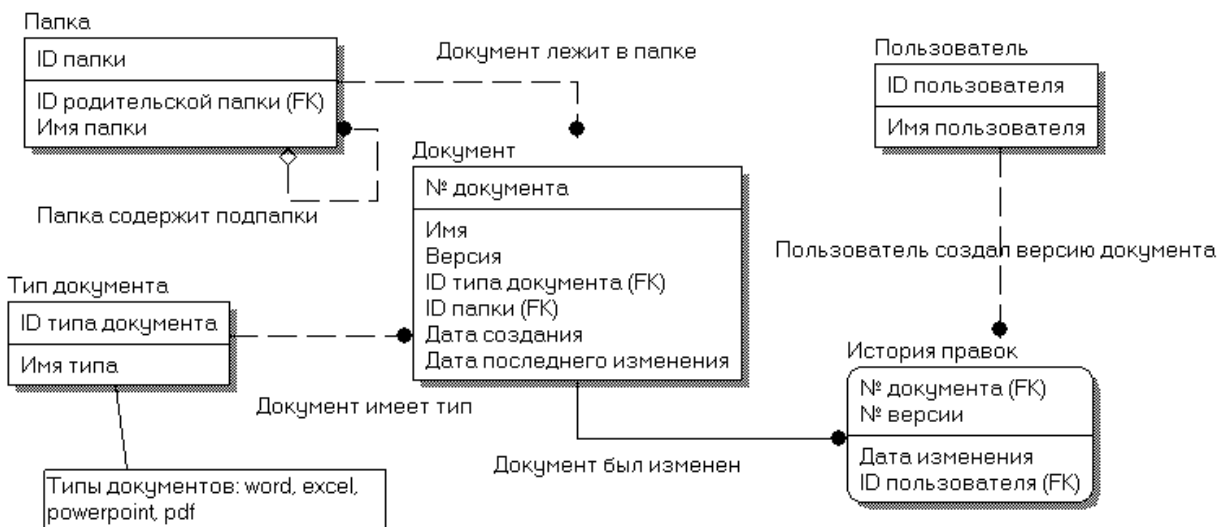
Вопрос №9. Укажите компоненты, необходимые для внедрения Data Warehouse (хранилища данных) в организации на базе OLAP:

1. OLAP-сервер
2. ETL
3. «Озеро данных» (Data Lake)
4. Сервер отчетов (Reporting server)
5. Визуализация OLAP-куба

## Образцы заданий к практическим занятиям:

### Задание № 1: Проектирование ROLAP-схемы

1. Дана ER-схема транзакционной (OLTP) базы данных



Разработайте OLAP-схему с нормализованными (“снежинка”) измерениями: пользователь, тип документа, папка, год, месяц, день изменения документа.

### Задание №2: Трансформация и визуализация данных

Даная исходная схема базы данных сайта <https://ru.stackoverflow.com>

Данные загружены в БД PostgreSQL в схему so\_ru.

Задание:

Спроектировать и заполнить данными ROLAP-схему данными согласно варианту задания.

Визуализировать полученные данные с помощью Tableau Desktop (или аналогичного средства)

Вариант 1: Количества и популярности (ViewCount) вопросов (POSTS) по дате создания вопроса и тегам (TAGS, TAG\_SYNONYMS)

Вариант 2: Количества и репутации пользователей по дате регистрации и классов знаков (Badges.class)

## 1.2. Промежуточная аттестация

Промежуточная аттестация осуществляется в форме экзамена

В качестве средств, используемых на промежуточной аттестации предусматривается:

Билеты

## 1.3. Типовые задания для проведения промежуточной аттестации

Вопросы к экзамену

1. OLAP и хранилища данных. Назначение. Основные характеристики
2. OLAP и OLTP системы. Характеристики и основные отличия
3. Правила Кодда для OLAP систем
4. Многомерное представление данных как основной способ логического представления данных. Назначение, классы задач, преимущества и недостатки в сравнении с реляционной моделью. Срезы многомерных кубов и их визуальное представление
5. Типы OLAP. Преимущества и недостатки
6. 1. Многомерный OLAP – MOLAP. Основные преимущества. Представление данных. Основные операции над данными. Преимущества и недостатки
7. Реляционный OLAP – ROLAP. Основные преимущества. Представление данных. Основные операции над данными. Преимущества и недостатки
8. Гибридный OLAP – HOLAP. Основные преимущества. Представление данных. Основные операции над данными. Преимущества и недостатки
9. Моделирование многомерных кубов на реляционной модели данных. Схема звезда. Преимущества и недостатки
10. Моделирование многомерных кубов на реляционной модели данных. Схема снежинка. Преимущества и недостатки
11. Моделирование многомерных кубов на реляционной модели данных. Подходы к хранению агрегатов
12. Общая схема организации хранилища данных
13. Расширения SQL для организации OLAP запросов. Обзор
14. ROLLUP – расширение к GROUP BY
15. CUBE – расширение к GROUP BY
16. GROUPING SETS – расширение к GROUP BY
17. Составные столбцы в GROUP BY
18. Конкатенация групп в GROUP BY
19. Вспомогательные функции – GROUPING, GROUPING\_ID, GROUP\_ID
20. “Оконные” агрегирующие функции в SQL. Синтаксис, отличие от классических функций агрегирования, основные возможности

## 2. КРИТЕРИИ ОЦЕНКИ ПО ДИСЦИПЛИНЕ

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
виды оценочных средств				
<b>Знания</b> (виды оценочных средств: приведены в п. 1.2.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
<b>Умения</b> (виды оценочных средств: приведены в п. 1.2.)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности не принципиального характера)	Успешное и систематическое умение
<b>Навыки</b> (владения, опыт деятельности) (виды оценочных средств: приведены в п. 1.2..)	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач