

Формализация информации и Big Data

02.03.03 -Математическое обеспечение и администрирование информационных систем,
направленность (профиль) -разработка и администрирование информационных систем

<https://vikchas.ru>

Вопросы 2-ой контрольной точки 28-30.10.2025

Часовских Виктор Петрович
доктор технических наук, профессор кафедры
ШИиКМ, ФГБОУ ВО «Уральский государственный
экономический университет

Екатеринбург 2025

Основы реляционной модели данных

1. Что такое реляционная база данных и в чем ее отличие от других типов баз данных?

- Реляционная БД — система хранения данных, основанная на модели данных в виде таблиц (отношений)
- Отличия: строгая структура, связи между таблицами через внешние ключи, использование SQL, поддержка ACID

2. Объясните концепции отношения, кортежа и атрибута в реляционной модели.

- Отношение — таблица в БД
- Кортеж — строка таблицы (запись)
- Атрибут — столбец таблицы (поле)

3. Что такое первичный ключ и какие требования к нему предъявляются?

- Первичный ключ — атрибут или набор атрибутов, уникально идентифицирующий каждую запись
- Требования: уникальность, не null, неизменяемость, минимальность (желательно)

4. Что такое внешний ключ и как он обеспечивает целостность данных?

- Внешний ключ — поле или набор полей, которые ссылаются на первичный ключ другой таблицы
- Обеспечивает ссылочную целостность, предотвращая недопустимые операции (например, удаление записи, на которую есть ссылки)

SQL и работа с данными

5. Какие категории команд SQL вы знаете и для чего они используются?

- DDL (Data Definition Language): CREATE, ALTER, DROP — для определения структуры
- DML (Data Manipulation Language): INSERT, UPDATE, DELETE — для манипуляции данными

- DQL (Data Query Language): SELECT — для запросов к данным
- DCL (Data Control Language): GRANT, REVOKE — для управления доступом
- TCL (Transaction Control Language): COMMIT, ROLLBACK — для управления транзакциями

6. Напишите пример запроса SELECT с использованием JOIN, GROUP BY, HAVING и ORDER BY.

SQL

```
SELECT c.category_name, COUNT(p.product_id) as product_count,  
AVG(p.price) as avg_price
```

```
FROM products p
```

```
JOIN categories c ON p.category_id = c.category_id
```

```
GROUP BY c.category_name
```

```
HAVING COUNT(p.product_id) > 10
```

```
ORDER BY product_count DESC;
```

7. В чем разница между INNER JOIN и OUTER JOIN? Объясните типы OUTER JOIN.

- INNER JOIN — возвращает строки, когда есть соответствие в обеих таблицах
- OUTER JOIN:
 - LEFT JOIN — все строки из левой таблицы и соответствующие из правой
 - RIGHT JOIN — все строки из правой таблицы и соответствующие из левой
 - FULL JOIN — все строки из обеих таблиц, с соответствиями где возможно

8. Что такое подзапросы в SQL? Приведите пример.

- Подзапрос — запрос SELECT внутри другого SQL-запроса
- Пример:

SQL

```
SELECT employee_name, salary
```

FROM employees

WHERE salary > (SELECT AVG(salary) FROM employees);

Нормализация и проектирование БД

9. Что такое нормализация и какие проблемы она решает?

- Нормализация — процесс организации данных для минимизации избыточности
- Устраняет аномалии вставки, обновления и удаления
- Улучшает целостность данных и эффективность хранения

10. Опишите первые три нормальные формы (1NF, 2NF, 3NF).

- 1NF: атомарные значения, есть первичный ключ, нет повторяющихся групп
- 2NF: в 1NF + не ключевые атрибуты полностью функционально зависят от всего первичного ключа
- 3NF: в 2NF + не ключевые атрибуты не зависят от других неключевых атрибутов

11. Что такое де нормализация и когда к ней прибегают?

- Де нормализация — намеренное введение избыточности для оптимизации операций чтения
- Применяется при:
 - Необходимости повысить производительность часто выполняемых запросов
 - Снижении количества соединений между таблицами
 - Создании отчетов и аналитических систем

12. Какие инструменты используются для проектирования БД? Что такое ERD?

- Инструменты: SQL Workbench, Oracle SQL Developer, Microsoft Visio, dbdiagram.io
- ERD (Entity-Relationship Diagram) — диаграмма сущность-связь, визуализирующая структуру БД, отношения между таблицами и их атрибуты

Транзакции и целостность данных

13. Что такое ACID и какие свойства транзакций она описывает?

- Atomicity (Атомарность): транзакция выполняется полностью или не выполняется вообще
- Consistency (Согласованность): транзакция переводит БД из одного целостного состояния в другое
- Isolation (Изолированность): транзакции независимы друг от друга
- Durability (Долговечность): результаты выполненной транзакции сохраняются независимо от сбоев

14. Какие уровни изоляции транзакций существуют и чем они отличаются?

- Read Uncommitted: видны незафиксированные данные (грязное чтение)
- Read Committed: видны только зафиксированные данные (неповторяющееся чтение)
- Repeatable Read: повторное чтение данных дает тот же результат (фантомное чтение)
- Serializable: полная изоляция, последовательное выполнение транзакций

15. Что такое блокировки в СУБД и как они влияют на производительность?

- Блокировки — механизм контроля конкурентного доступа к данным
- Типы: разделяемые (чтение), исключительные (чтение/запись)
- Влияние: предотвращают конфликты, но могут вызвать задержки или взаимоблокировки

16. Объясните понятие “взаимная блокировка” (deadlock). Как СУБД обнаруживают и разрешают deadlock?

- Deadlock — ситуация, когда две транзакции блокируют ресурсы друг друга
- Обнаружение: граф зависимостей, таймауты

- Разрешение: откат одной из транзакций (жертвы), таймауты, предотвращение через упорядочивание доступа

Индексация и оптимизация запросов

17. Что такое индекс в СУБД и как он влияет на производительность?

- Индекс — структура данных, ускоряющая поиск и сортировку записей
- Ускоряет SELECT, но замедляет INSERT/UPDATE/DELETE
- Занимает дополнительное пространство
- Типы: В-деревья, хеш-индексы, битовые индексы

18. В чем разница между кластерными и некластерными индексами?

- Кластерный: определяет физический порядок хранения данных, только один на таблицу
- Не кластерный: создает отдельную структуру с указателями на данные, можно иметь много на одну таблицу

19. Что такое план выполнения запроса и как его анализировать?

- План выполнения — порядок и способ доступа к таблицам и индексам при выполнении запроса
- Анализируется с помощью EXPLAIN/EXPLAIN PLAN
- Показывает используемые индексы, порядок соединения таблиц, оценку стоимости операций

20. Какие факторы могут негативно влиять на производительность запросов?

- Отсутствие нужных индексов
- Избыточная индексация
- Неэффективные соединения таблиц
- Неоптимальные условия WHERE
- Избыточная выборка данных (SELECT *)
- Фрагментация таблиц и индексов

Практическое применение и особенности СУБД

21. Перечислите несколько известных реляционных СУБД и их особенности.

- SQL: открытый исходный код, высокая производительность, популярность в веб-приложениях
- PostgreSQL: открытый исходный код, расширяемость, поддержка JSON, географические данные
- Oracle: корпоративное решение, высокая надежность, многофункциональность, стоимость
- SQL Server: интеграция с продуктами Microsoft, хорошая поддержка BI-инструментов
- SQLite: встраиваемая БД, не требует сервера, низкое потребление ресурсов

22. Что такое функциональные и материализованные представления?

- Функциональные представления (VIEW) — виртуальные таблицы на основе запроса, вычисляемые при обращении
- Материализованные представления — физически сохраненные результаты запроса, требующие периодического обновления

23. Для чего используются хранимые процедуры и триггеры?

- Хранимые процедуры — набор SQL-команд, хранимых в СУБД, улучшают производительность и безопасность
- Триггеры — программы, автоматически запускаемые при определенных событиях (INSERT, UPDATE, DELETE), обеспечивают целостность данных

24. Как осуществляется резервное копирование и восстановление в СУБД?

- Типы резервного копирования: полное, дифференциальное, инкрементальное
- Методы: логическое (дампы), физическое (копии файлов)
- Восстановление: до определенной точки во времени (Point-in-Time Recovery)
- Инструменты: встроенные утилиты (sqldump, pg_dump), системы резервного копирования

Современные тенденции и расширения

25. Как реализуется поддержка JSON в современных реляционных СУБД?

- Специальные типы данных (JSONB в PostgreSQL)
- Функции для извлечения и модификации JSON (`json_extract`, `json_object`)
- Индексирование JSON-полей
- Проверки валидности JSON

26. Что такое репликация и какие типы репликации существуют?

- Репликация — копирование данных между несколькими серверами
- Типы:
 - Master-Slave (основной-подчиненный)
 - Master-Master (с несколькими ведущими узлами)
 - Синхронная/асинхронная

27. Что такое шардинг и партиционирование? В чем их отличие?

- Партиционирование — разделение таблицы на логические части внутри одной БД по определенному критерию
- Шардинг — распределение данных по нескольким физическим серверам
- Отличие: партиционирование — на одном сервере, шардинг — на разных серверах

28. Как осуществляется миграция схемы БД в современных проектах?

- Инструменты миграции: Flyway, Liquibase, Rails Migrations, Entity Framework Migrations
- Версионирование схемы
- Скрипты для применения и отката изменений
- Интеграция в CI/CD процессы

29. Что такое ORM и какие преимущества и недостатки они имеют?

- ORM (Object-Relational Mapping) — технология связывания БД с объектами в коде
- Преимущества: абстракция от SQL, удобство разработки, меньше ошибок

- Недостатки: возможные проблемы производительности, скрытая сложность, ограниченный контроль над SQL

30. Как обеспечивается безопасность в реляционных СУБД?

- Аутентификация и авторизация пользователей
- Шифрование данных (на диске, при передаче)
- Аудит действий пользователей
- Защита от SQL-инъекций
- Управление правами доступа к объектам БД