

Базы данных

Что такое базы данных и СУБД

Понятия БД и СУБД

База данных (БД) - совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между соответствующими сущностями и поддерживающей одну или более областей применения.

Система управления базами данных (СУБД) - это программное обеспечение для удобного управления базой данных, позволяющее создать базу данных и впоследствии манипулировать данными.

Модель данных - это формальное описание структуры хранения информации.

Что такое базы данных и СУБД

Основные функции СУБД

- Администрирование БД
- Настройка прав доступа
- Управление данными во внешней памяти (на дисках)
- Управление данными в оперативной памяти с использованием дискового кэша
- Журнализация изменений данных
- Резервное копирование и восстановление базы данных
- Поддержка языков БД для манипулирования данными

Что такое базы данных и СУБД

Свойства проектируемой БД

- Целостность
- Восстанавливаемость
- Безопасность
- Эффективность

Что такое базы данных и СУБД

Топ 10 БД

1.

ORACLE

2.



3.



4.



5.



6.



7.



8.



9.



10.



Классификация СУБД

По модели данных

- Иерархические
- Сетевые
- Реляционные
- Нереляционные
- Мультимодельные

По степени распределённости

- Локальные
- Распределенные

По способу доступа

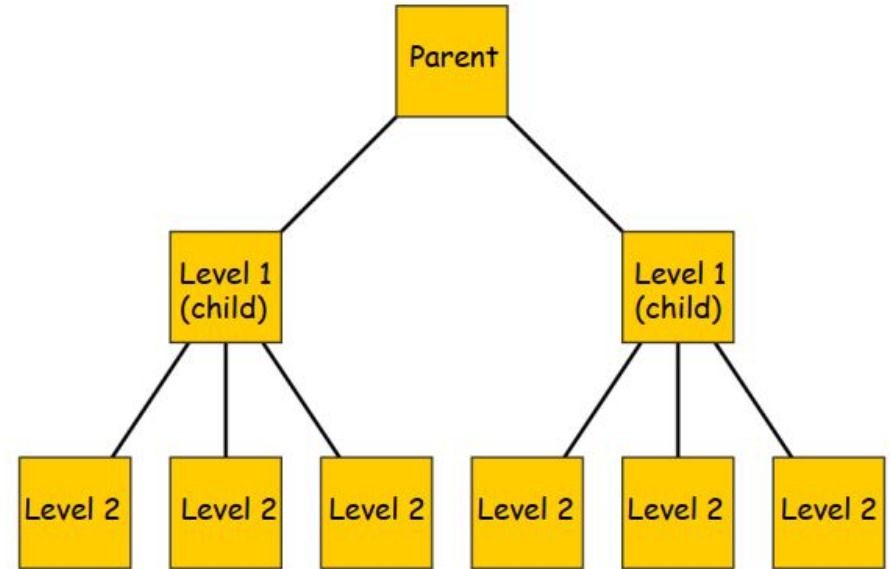
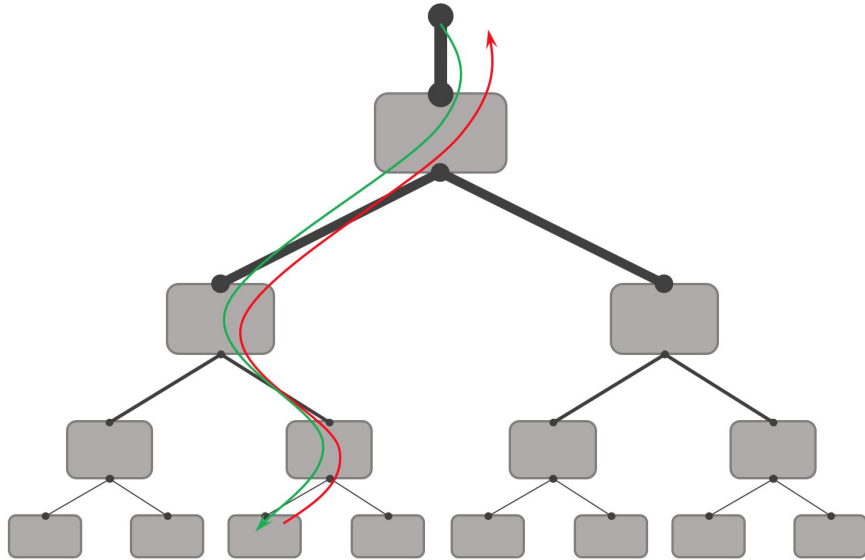
- Файл-серверные
- Клиент-серверные
- Встраиваемые

Виды моделей организации данных

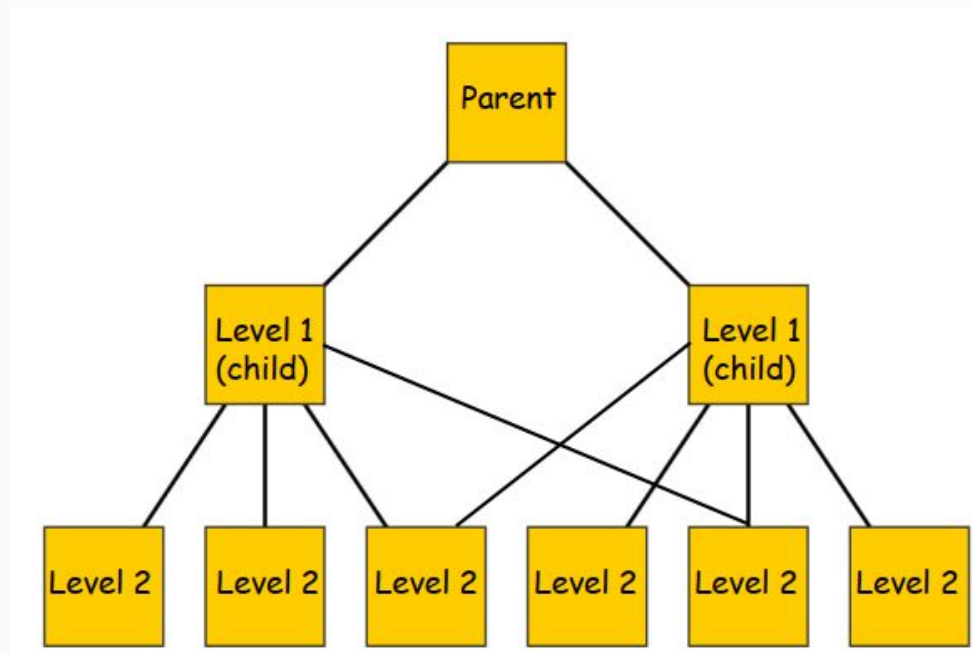
Основные виды моделей организации данных

- Иерархическая модель
- Сетевая модель
- Реляционная модель (SQL)
- Не реляционная модель (NOSQL)

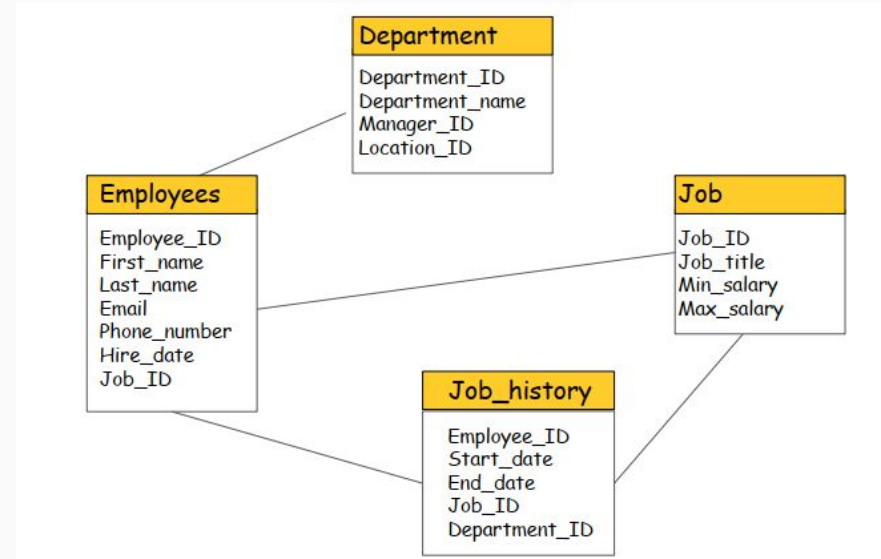
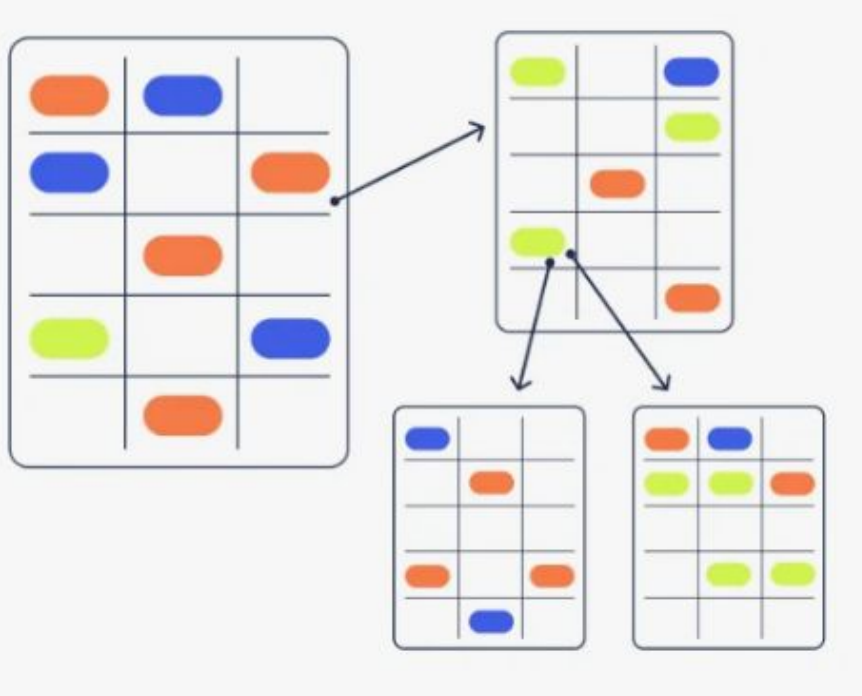
Иерархическая модель



Сетевая модель



Реляционная модель

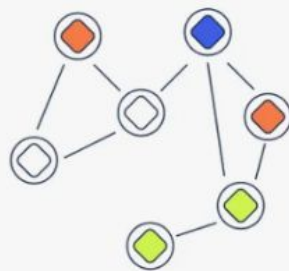


Нереляционная модель

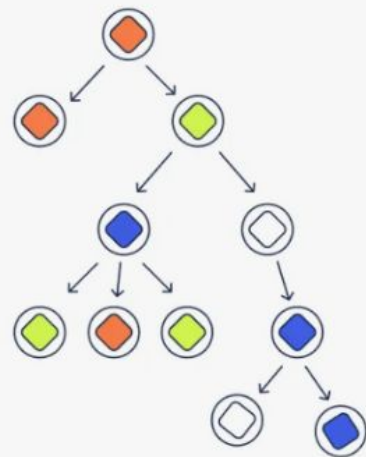
Нереляционная база данных - это база данных, в которой в отличие от большинства традиционных систем баз данных не используется табличная схема строк и столбцов.

В этих базах данных применяется модель хранения, оптимизированная под конкретные требования типа хранимых данных.

Сетевые



Иерархические



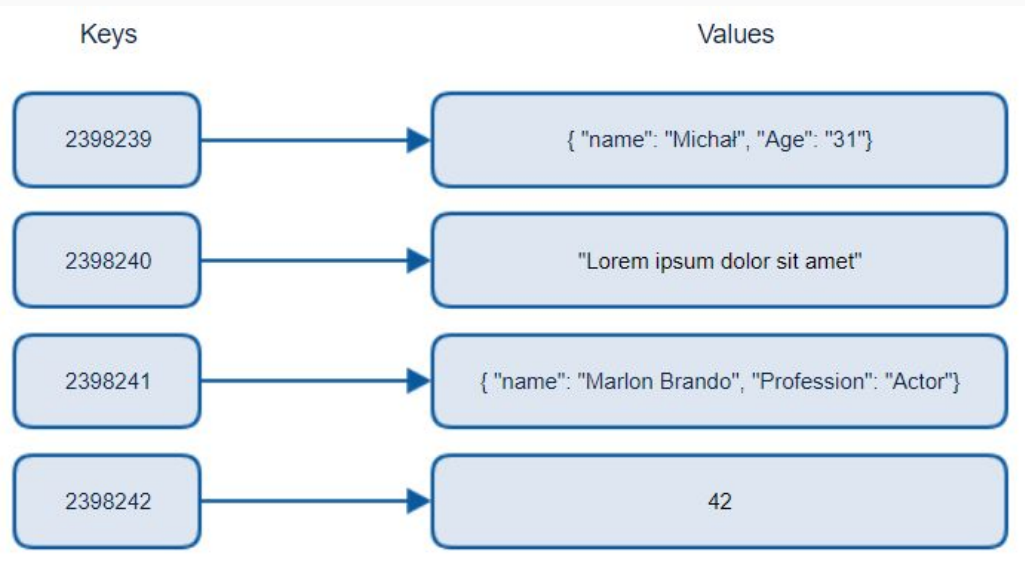
Преимущества нереляционных БД

- Гибкость
- Масштабируемость
- Высокая производительность
- Широкие функциональные возможности

Виды нереляционных моделей

- Key - value storage (хранилище ключ-значение)
- Документоориентированное хранилище
- Колоночное хранение (столбцовое)
- Графовое хранилище

Key-value storage



Для чего используется:

- Кэш
- Сессии
- Счетчик посещений/просмотров
- Хранение изображений
- Брокеры сообщений



Key-value storage

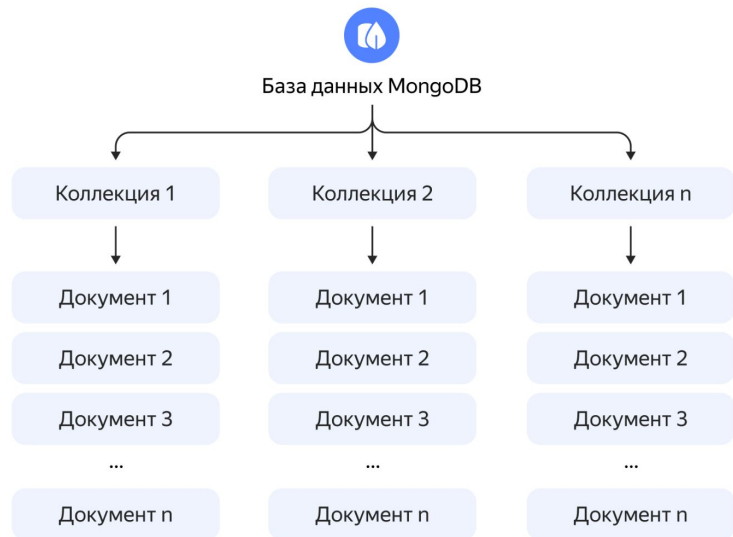
Преимущества:

- Быстрая запись
- Быстрое получение (поиск) данных
- Легкая масштабируемость

Недостатки:

- Невозможность выполнения сложных запросов
- Невозможность выполнения агрегации данных
- Медленный поиск по значениям
- Возможные проблемы с генерацией ключа (при работе с большими данными)

Документоориентированное хранилище



Для чего используется:

- Каталоги
- Профили пользователей
- Управление контентом

Преимущества:

- Отсутствие схемы
- Высокая производительность
- Легкая масштабируемость
- Гибкость
- Отсутствие внешних ключей
- Открытые форматы
- Встроенное управление версиями

Недостатки:

- Ограничения по проверке на согласованность
- Проблемы с атомарностью
- Безопасность

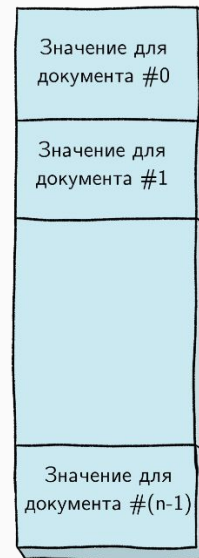
Колоночное хранилище (столбцовое)

Для чего используется:

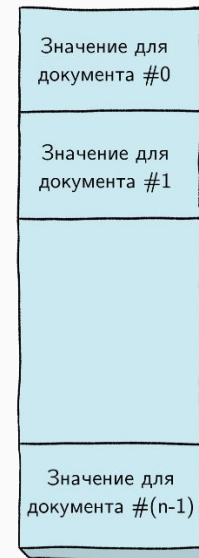
- Большой объем данных
- Аналитика

Колоночное хранение Column store

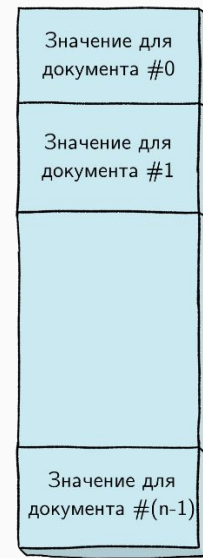
Колонка #0



Колонка #1



Колонка #(m-1)



Колоночное хранилище (столбцовое)

Преимущества:

- Высокая производительность
- Сжатие данных
- Агрегация запросов на больших объемах данных
- Легкая масштабируемость

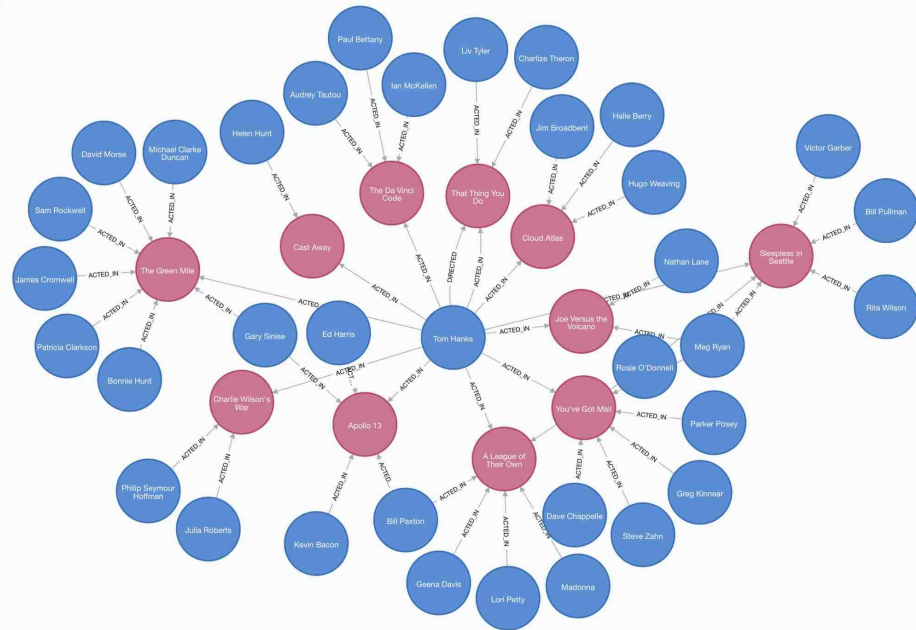
Недостатки:

- Медленно работают на запись
- Ограничения для разработчиков

Графовое хранилище

Для чего используется:

- Большой объем связанных данных
- Соцсети
- Рекомендательные системы



Преимущества:

- Гибкая
- Легкая масштабируемость
- Быстрый поиск

Недостатки:

- Низкая производительность при незначительном количестве связей и больших объемах данных
- Плохо работают в параллельных архитектурах

Реляционная модель - это модель данных, структура которой основана на наборе отношений

Принципы БД:

- Данные хранятся в таблицах (отношениях), состоящих из столбцов (атрибутов, полей) и строк (кортежей)
- У каждого столбца есть свое имя и свой тип
- В таблице может быть минимум 1 столбец. А строк в таблице может и не быть
- Запрос к базе данных возвращает результат в виде таблиц



Столбец_1		Столбец_2					
Строка (кортеж)		Значение					
		Столбец (атрибут)					

Типы данных в реляционной БД

- **INT** или **INTEGER** - целочисленные данные
- **DEC** или **DECIMAL** — десятичные дробные величины
- **FLOAT** — еще один тип для десятичных дробных величин
- **CHAR** или **CHARACTER** — строковый тип данных с фиксированной длиной
- **VARCHAR** — строковые данные переменной длины
- **DATE** — дата
- **DATETIME** — дата и время
- **TEXT** (либо **BLOB**) — большой объем информации

и много много других типов

Ограничения (constraints)

Ограничения - это правила для данных в таблице

- PRIMARY KEY
- FOREIGN KEY
- NOT NULL
- UNIQUE
- CHECK
- DEFAULT
- INDEX

Ограничения (constraints)

Ключ - это идентификатор, являющийся частью набора элементов данных

Клиент				Товар	
Id_кл	Фамилия	Имя	Отчество	Id_тов	Название
15	Иванов	Иван	Иванович	1	Шкаф
16	Петров	Петр	Петрович	2	Стул
17	Николаев	Николай	Николаевич	3	Стол

Заказ				
Id_зак	Клиент	Товар	Дата	Количество
1	15	1	15.09.2003	1
2	17	1	17.09.2003	2
3	15	2	20.09.2003	12

Ограничения (constraints)

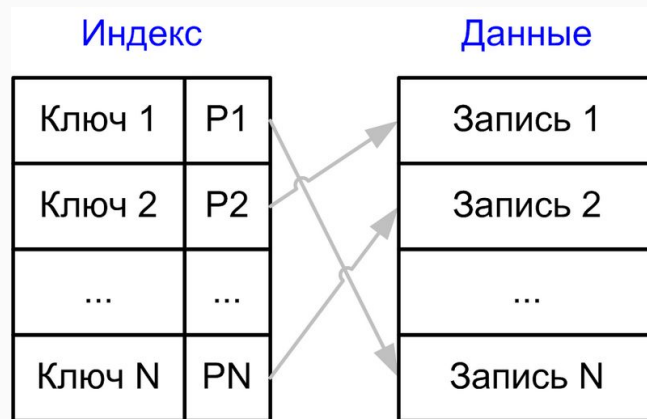
NOT NULL гарантирует, что столбец не может иметь значение NULL.

UNIQUE гарантирует, что все значения в столбце разные. Для каждой таблицы может быть множество ограничений UNIQUE, но только одно ограничение PRIMARY KEY для каждой таблицы.

CHECK ограничивает диапазон значений, которые могут быть записаны в столбец.

DEFAULT определяет значение колонки по умолчанию

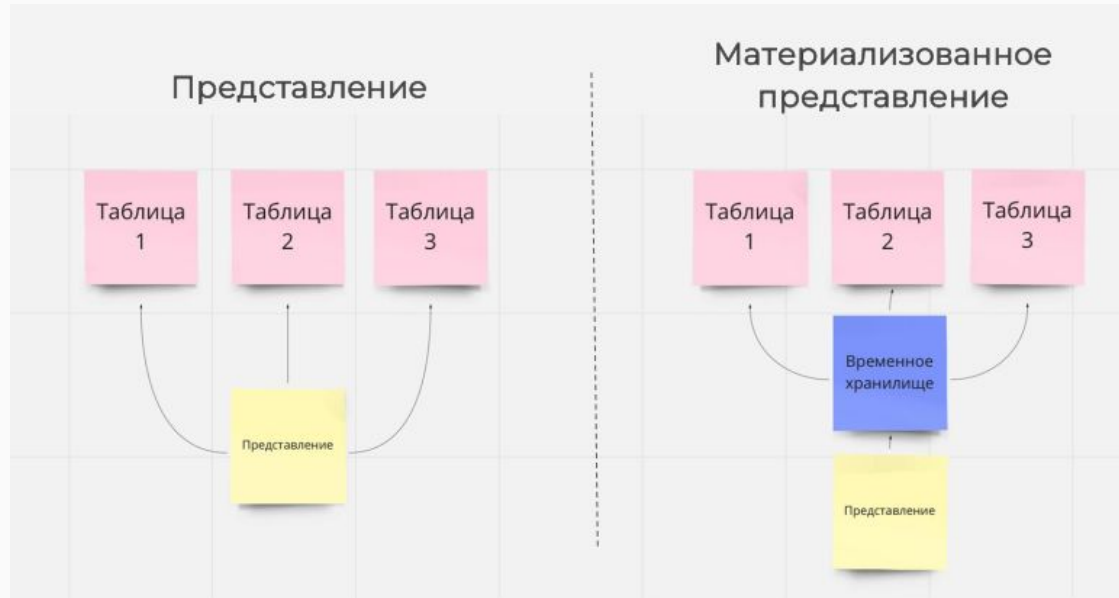
INDEX специальная структура базы данных, которая предназначена для повышения скорости доступа к данным



Представления (view)

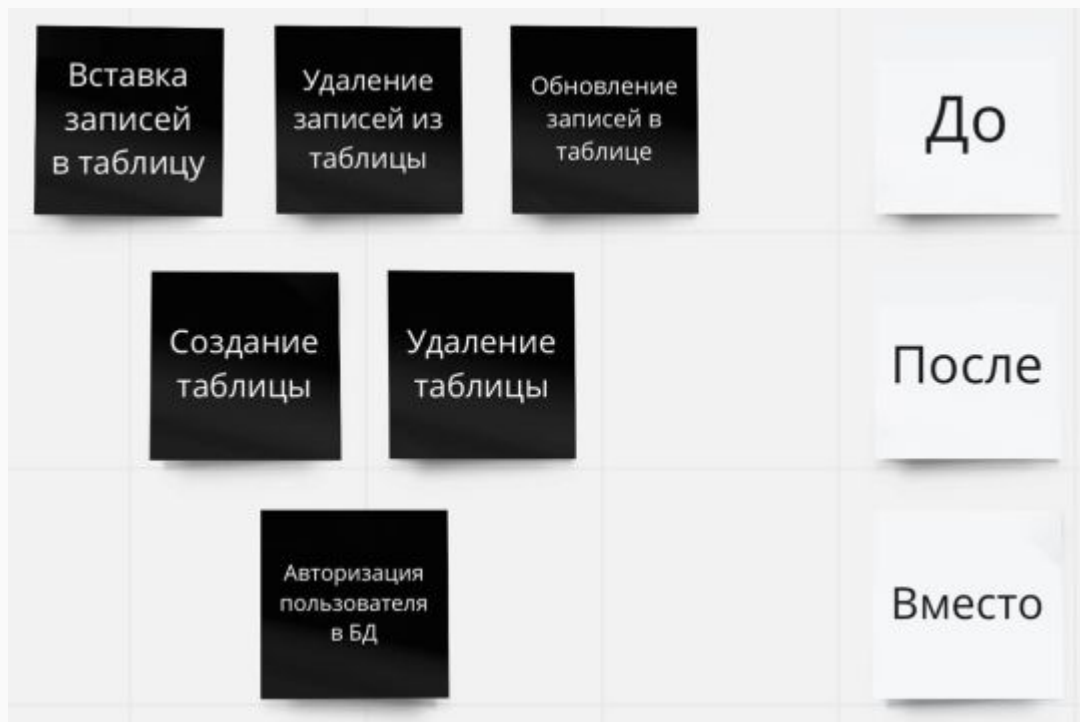
Представление - это SQL-запрос, которому заранее присвоено имя и который можно выполнить, указав имя.

Материализованное представление - это SQL-запрос, которому заранее присвоено имя и который можно выполнить, указав имя. Результат запроса сохраняется для дальнейшего использования.



Триггеры

Триггер - это программа, которая автоматически выполняется или запускается при возникновении некоторых событий



Транзакции

Транзакция - это логически завершенная единица работы, которая содержит один или более sql-запрос. Все действия составляющие транзакцию либо должны выполняться полностью либо полностью не выполняться.

ACID:

ATOMICITY (атомарность) - гарантирует, что транзакция не может выполняться частично

CONSISTENCY (консистентность) - требует, чтобы после завершения транзакции данные оставались консистентными и валидными

ISOLATION (изоляция) - требует, чтобы при параллельном выполнении транзакции они не влияли друг на друга

DURABILITY (надежность) - гарантирует, что если пользователь получил подтверждение, что транзакция выполнена, изменения не будут отменены из-за какого-либо сбоя

Нормализация

Нормализация - это удаление избыточности данных



1-я НФ

Сотрудник	Контакт
Иванов И.И.	123-456-789, 987-654-321
Сергеев С.С.	Рабочий телефон 555-666-777, Домашний телефон 777-888-999
John Smith	123-456-789
John Smith	123-456-789



1NF

Сотрудник	Телефон	Тип телефона
Иванов И.И.	123-456-789	
Иванов И.И.	987-654-321	
Сергеев С.С.	555-666-777	Рабочий телефон
Сергеев С.С.	777-888-999	Домашний телефон
John Smith	123-456-789	

- В таблице не должно быть дублирующих строк
- В каждой ячейке таблицы хранится атомарное значение (одно не составное значение)
- В столбце хранятся данные одного типа
- Отсутствуют массивы и списки в любом виде

2 - я НФ

ФИО	Должность	Подразделение	Описание подразделения
Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
John Smith	Продавец	Отдел реализации	Организация сбыта продукции

1NF



2NF

Табельный номер	ФИО	Должность	Подразделение	Описание подразделения
1	Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	John Smith	Продавец	Отдел реализации	Организация сбыта продукции

- Таблица должна находиться в первой нормальной форме
- Таблица должна иметь первичный ключ
- Все неключевые столбцы таблицы должны зависеть от полного ключа (в случае если он составной)

3 - я НФ

Табельный номер	ФИО	Должность	Подразделение	Описание подразделения
1	Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Sereev С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	John Smith	Продавец	Отдел реализации	Организация сбыта продукции

2NF

Табельный номер	ФИО	Должность	Подразделение
1	Иванов И.И.	Программист	1
2	Sereev С.С.	Бухгалтер	2
3	John Smith	Продавец	3

3NF

Идентификатор подразделения	Подразделение	Описание подразделения
1	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	Отдел реализации	Организация сбыта продукции

- Таблица должна находиться во второй нормальной форме
- В таблицах отсутствует транзитивная зависимость

Транзитивная зависимость - это когда неключевые столбцы зависят от значений других неключевых столбцов

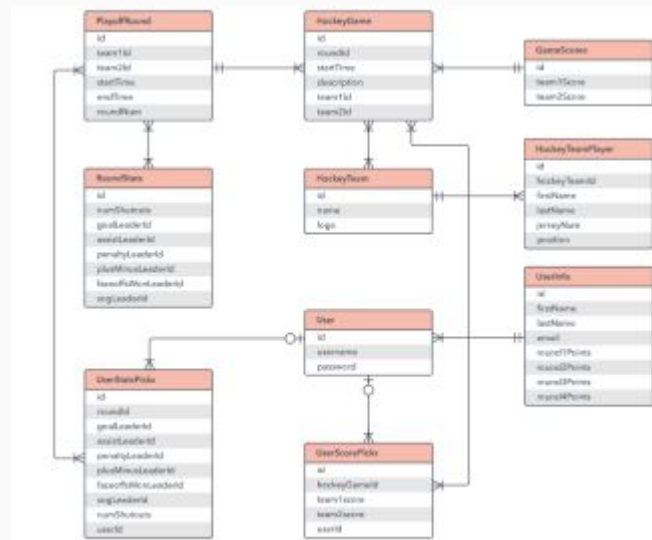
Недостатки реляционных БД

- Нехватка гибкости
- Не всегда предметная область может быть представлена в виде "таблиц"
- Для реляционных БД предусмотрено вертикальное масштабирование

Модель представления базы данных

Модель представления базы данных - это описание базы данных с помощью графического языка, по которому она потом будет создана

ERD или ER-диаграмма - это разновидность блок-схемы, где показано, как разные «сущности» (люди, объекты, концепции и так далее) связаны между собой внутри системы



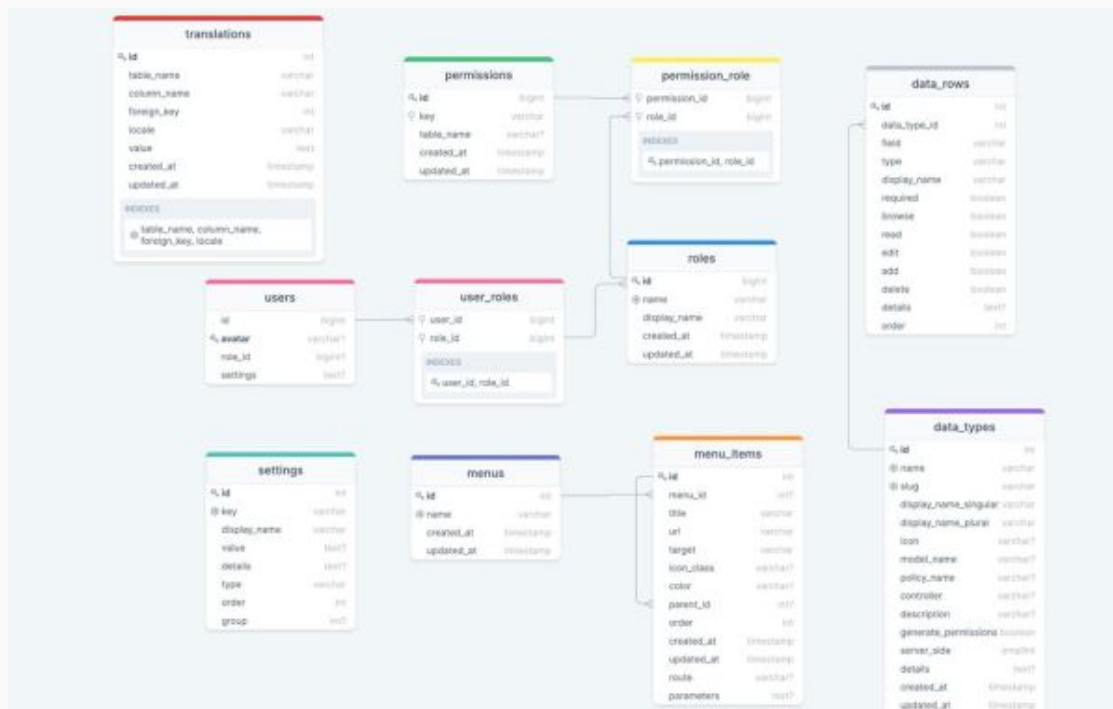
Модель представления базы данных

Инфологический уровень (концептуальный)



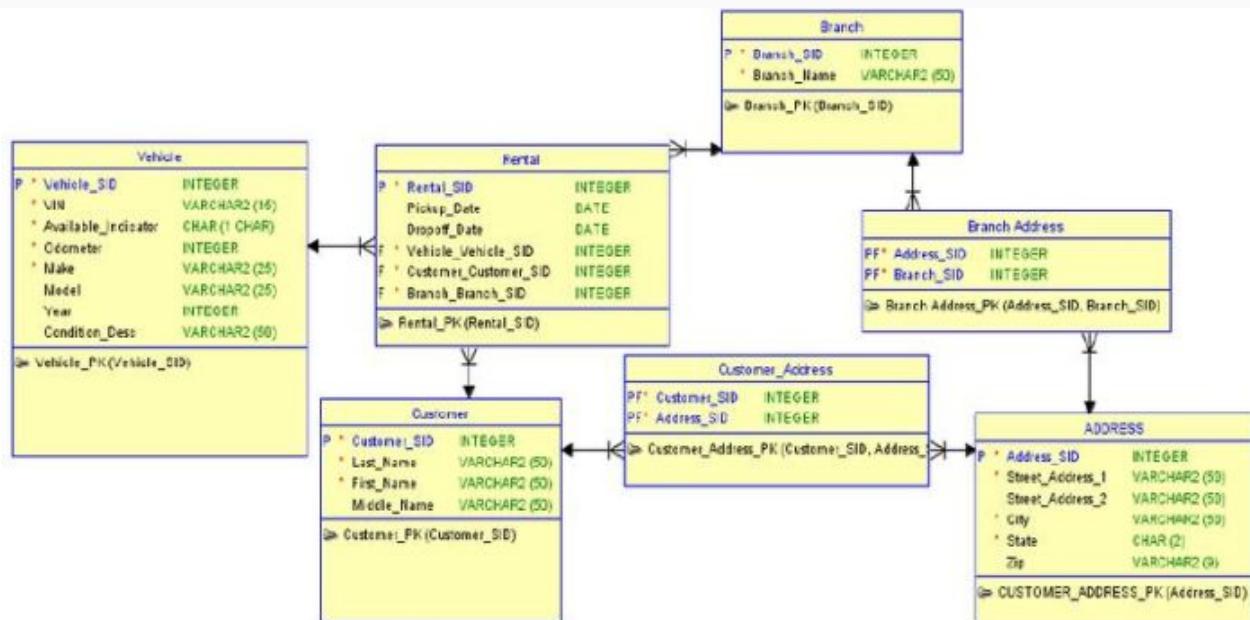
Модель представления базы данных

Даталогический уровень (логический)



Модель представления базы данных

Физический уровень



Виды связей

Один-к-одному

У одного гражданина определенной страны обязательно есть только один паспорт этой страны

Многие-к-одному

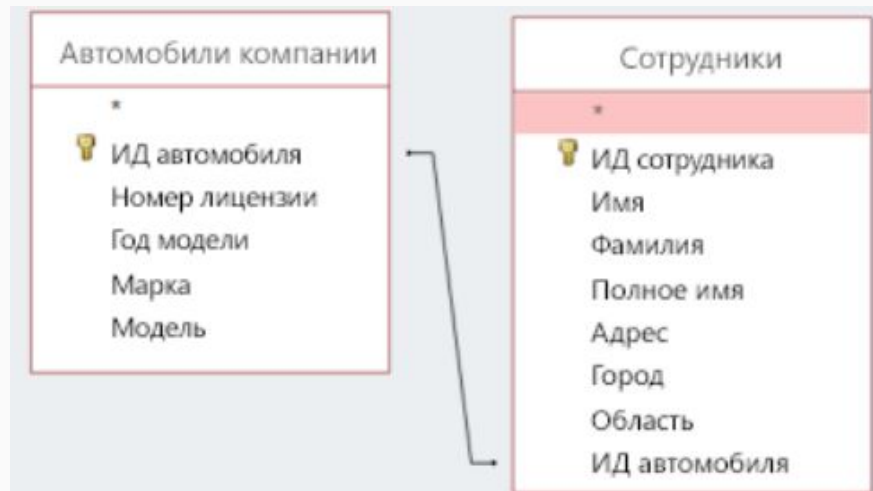
У каждого пользователя может быть один и больше номеров телефонов

Один-к-многим

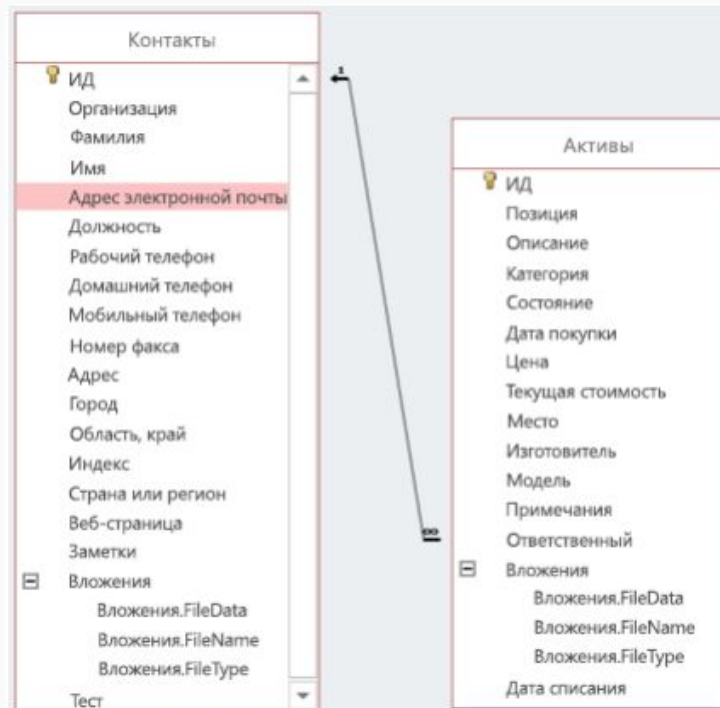
Многие-ко-многим

Каждый учитель обучает многих учащихся, а каждый учащийся может обучаться с несколькими учителями

Отношение один к одному



Отношение один ко многим



Отношение многие ко многим

