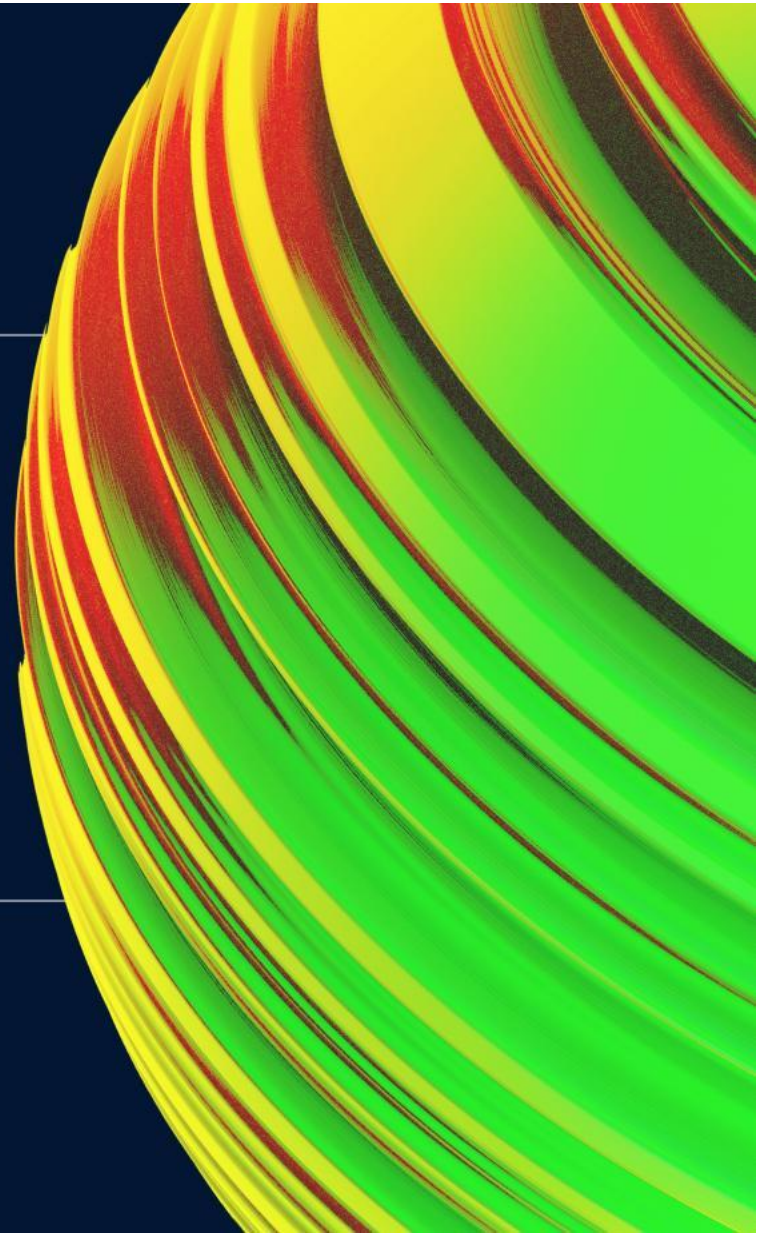
УНИВЕРСИТЕТ ИННОПОЛИС

Реляционные базы данных и SQL

Воробьёва Мария

- maria.vorobyova.ser@gmail.com
- @SparrowMaria



Как можно хранить данные?

Можно так

Как можно хранить данные?

innopolis university

А можно так

Определение базы данных

Базы данных (БД) — это структурная совокупность взаимосвязанных данных определенной предметной области (реальных объектов, процессов, явлений и т.д.).

База данных (БД) – это совокупность массивов и файлов данных, организованная по определённым правилам, предусматривающим стандартные принципы описания, хранения и обработки данных независимо от их вида.

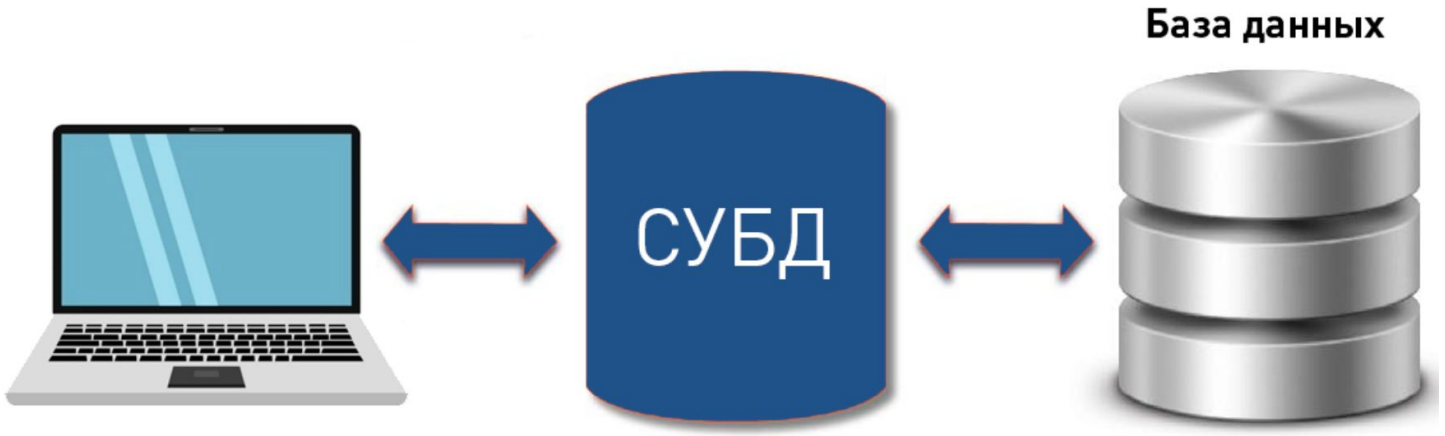
Основная функция базы данных – предоставление единого хранилища для всей информации, относящейся к определенной теме.

https://thecode.media/db/

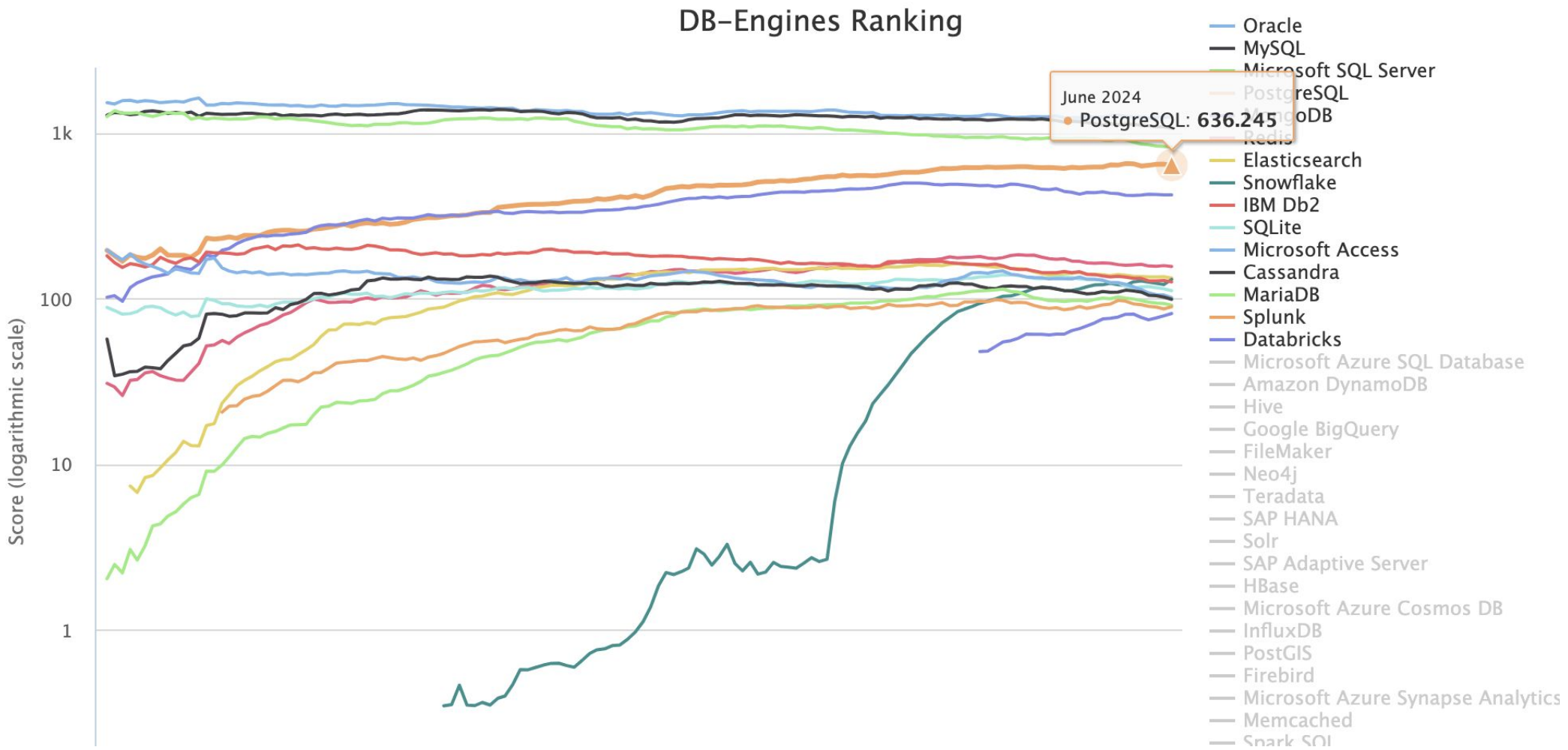
Что такое Системы управления базами данных

Системы управления базами данных — программное обеспечение, используемое для создания, изменения, администрирования баз данных

Функции СУБД - выполнение различных операций с данными: ввод, хранение, манипулирование, обработка запросов, поиск, выборка, сортировка, обновление, защита данных от несанкционированного доступа или потери



DB-Engines Ranking - Trend Popularity



Реляционные СУБД используют 99,5% респондентов

https://db-engines.com/en/ranking_trend

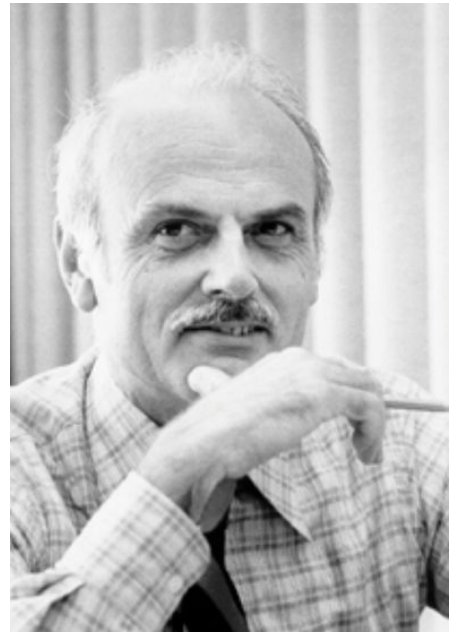
Немного истории

INNOPOLIS UNIVERSITY

Как данные хранить?

Как эффективно манипулировать данными

Понятие «модель данных»

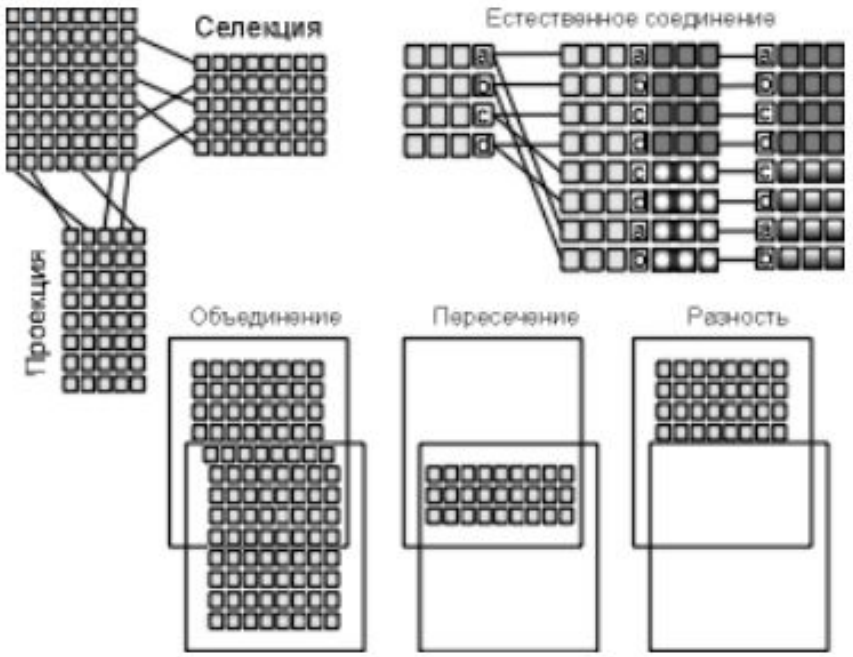


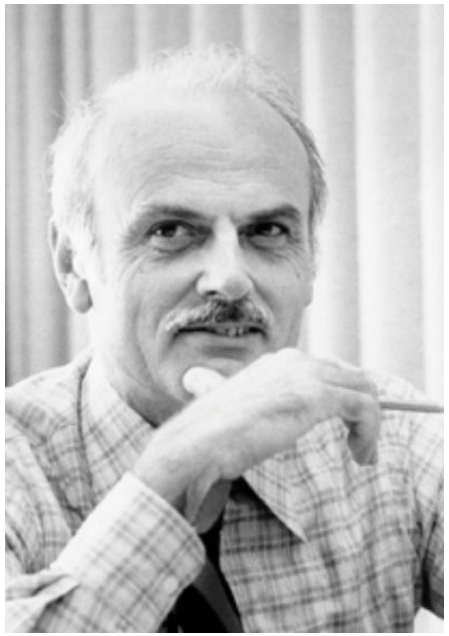
- Понятие модели данных предложено в 1969 г. Эдгаром Коддом для описания реляционного подхода к организации БД. Понятие модели данных оказалось удобным и для реализационно-независимого представления и сопоставления других подходов.

- В классической теории баз данных, ***модель данных*** есть формальная теория представления и обработки данных в системе управления базами данных.

Немного истории

- Предложив реляционную модель данных, Э.Ф.Кодд создал и инструмент для удобной работы с отношениями – **реляционную алгебру**. Каждая операция этой алгебры использует одну или несколько таблиц в качестве ее операндов и продуцирует в результате новую таблицу.
- Созданы языки манипулирования данными, позволяющие реализовать все операции реляционной алгебры.
  - **SQL (Structured Query Language)** – *структуризованный язык запросов*





Основы реляционной модели

INNOPOLIS UNIVERSITY

ACID свойства:

- **Атомарность** - все изменения данных выполняются как единая операция, от начала и до конца.,
- **Согласованность** - после нормального завершения работы (EOT, End of transaction) транзакция фиксирует допустимые результаты.,
- **Изолированность** - параллельные транзакции не должны влиять друг на друга во время выполнения.,
- **Долговечность** - пользователь может быть уверен, что после оповещение об успешной транзакции ничто не отменит результат.

A tomicity ("all or nothing")
C onsistency ("follow the rules")
I solation ("my changes are mine…")
D urability ("death is not a reason")

Немного про BASE

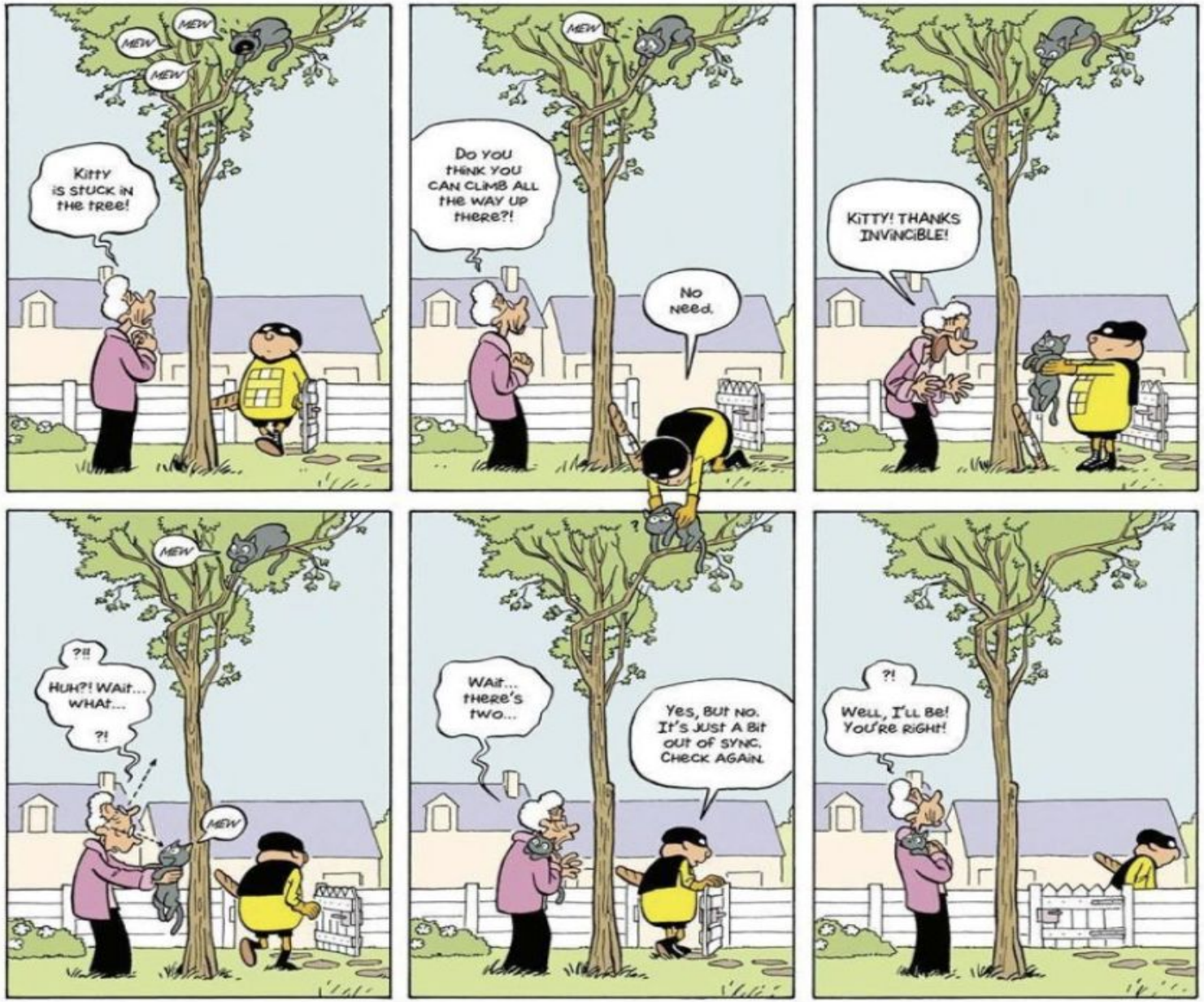
INNOPOLIS UNIVERSITY

BASE(Basically Available, Soft-state, Eventual Consistency) — это своеобразный контраст ACID, который говорит нам, что истинная согласованность не может быть достигнута в реальном мире и не может быть смоделирована в высокомасштабируемых системах.

Основные принципы BASE:

1. Базовая доступность (Basic Availability): Система отвечает на любой запрос, но возможны ошибки или несогласованность данных.
2. Гибкое состояние (Soft-state): Состояние системы может изменяться со временем, что приводит к неконечной согласованности.
3. Конечная согласованность (Eventual Consistency): Система в конечном итоге достигает согласованного состояния, продолжая принимать данные без проверки каждой транзакции на согласованность

Немного про BASE



BASE(Basically Available, Soft-state, Eventual Consistency) говорит, что система может временно находиться в несогласованном состоянии, но в конечном итоге согласуется.

Про CAP теорему

INNOPOLIS UNIVERSITY

Eric Brewer's theorem

Any distributed data store can provide only **two of** the following **three** guarantees

C onsistency ("all nodes have the same data")

A vailability ("answer for every request.
No fresh data... then take old one")

P artition tolerance ("Hydra style")

Про CAP теорему

Теорема Брюера

CAP (Consistency, Availability, Partition tolerance) — теорема о том, что для распределенных вычислений невозможно обеспечить все три свойства: согласованность данных, доступность и устойчивость к разделению.

- **Согласованность.**

Каждый процесс чтения получает последнюю запись или ошибку, соответственно, когда в системе запущено несколько параллельных процессов записи и чтения, то каждое чтение всегда возвращает последнюю запись, сделанную в системе.

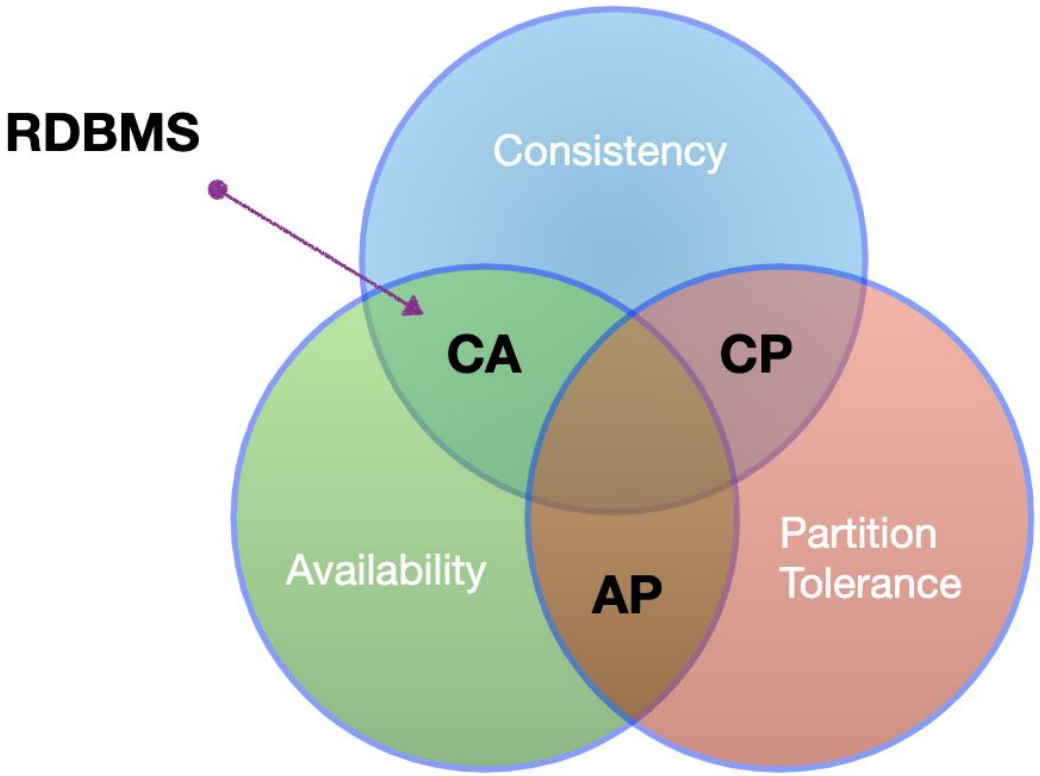
- **Доступность.**

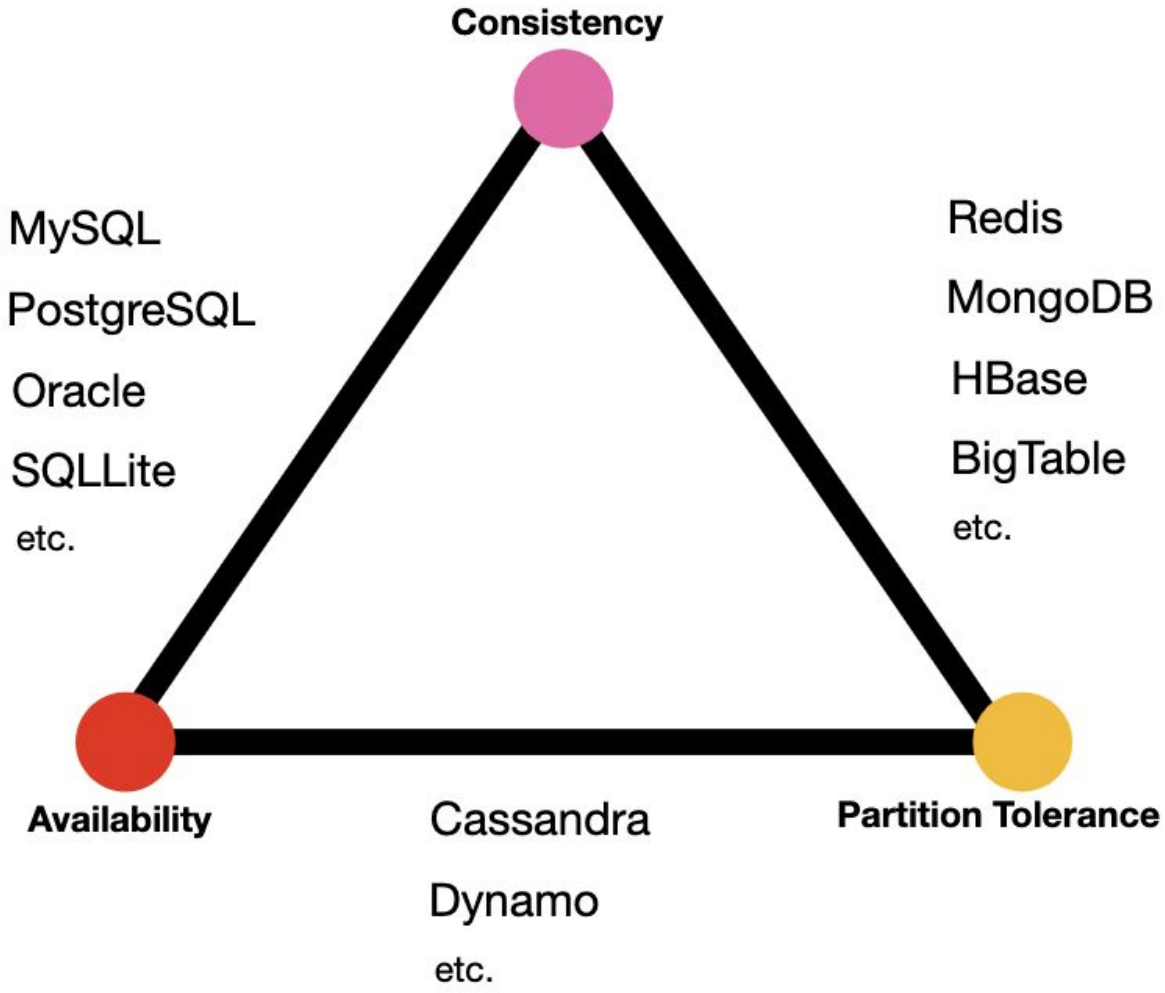
Принцип доступности в распределенной системе гарантирует, что система всегда остается работоспособной. Каждый запрос получает ответ без ошибок, независимо от индивидуального состояния узла. Впрочем, принцип не гарантирует, что ответ содержит самую последнюю запись (смотрите предыдущий пункт “Согласованность”).

- **Устойчивость к разделению.**

Распределенная система продолжает работу, даже когда отдельный узел не отвечает. Вышедший из строя узел подкрепляется вторичным узлом, поэтому вторичный узел заменяет первичный во время сбоев, а система становится отказоустойчивой. Хотя некоторые сообщения все-таки могут выходить из строя.

Про CAP теорему



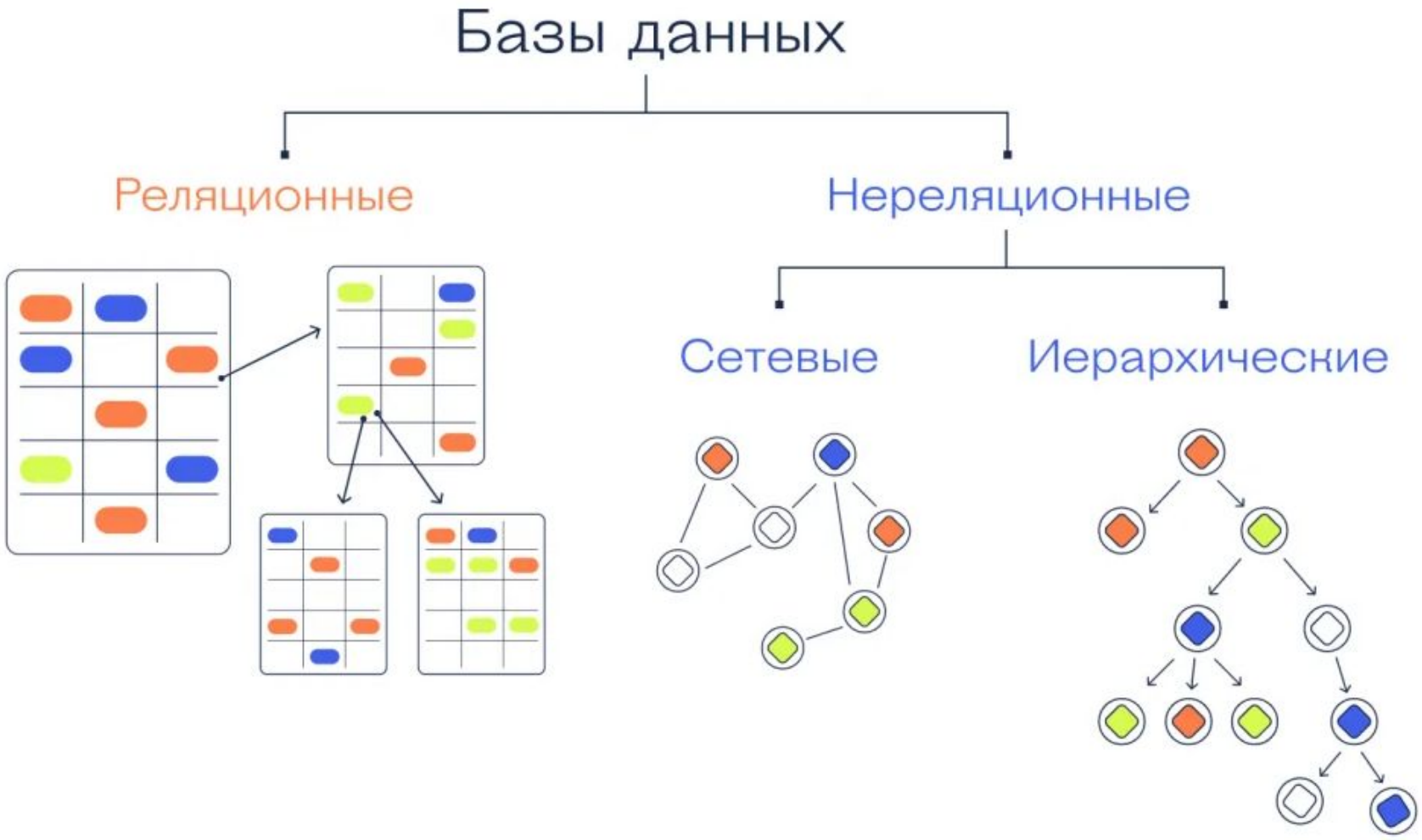


https://habr.com/ru/articles/328792/

Основные типы базы данных

Три основных типа

В зависимости от того, какие данные нужно в ней хранить и как с ними работать, базы делятся на реляционные и нереляционные:



Основные типы базы данных

Реляционные

Реляционные базы данных ещё называют табличными, потому что все данные в них можно представить в виде разных таблиц. Одни таблицы связаны с другими, а другие — с третьими. Например, база данных покупок в магазине может выглядеть так:

Нормализация данных

Нормализация — это процесс организации данных в базе данных, включающий создание таблиц и установление отношений между ними в соответствии с правилами, которые обеспечивают защиту данных и делают базу данных более гибкой, устраняя избыточность и несогласованные зависимости.

Первая нормальная форма

- Устраните повторяющиеся группы в отдельных таблицах.
- Создайте отдельную таблицу для каждого набора связанных данных.
- Идентифицируйте каждый набор связанных данных с помощью первичного ключа.

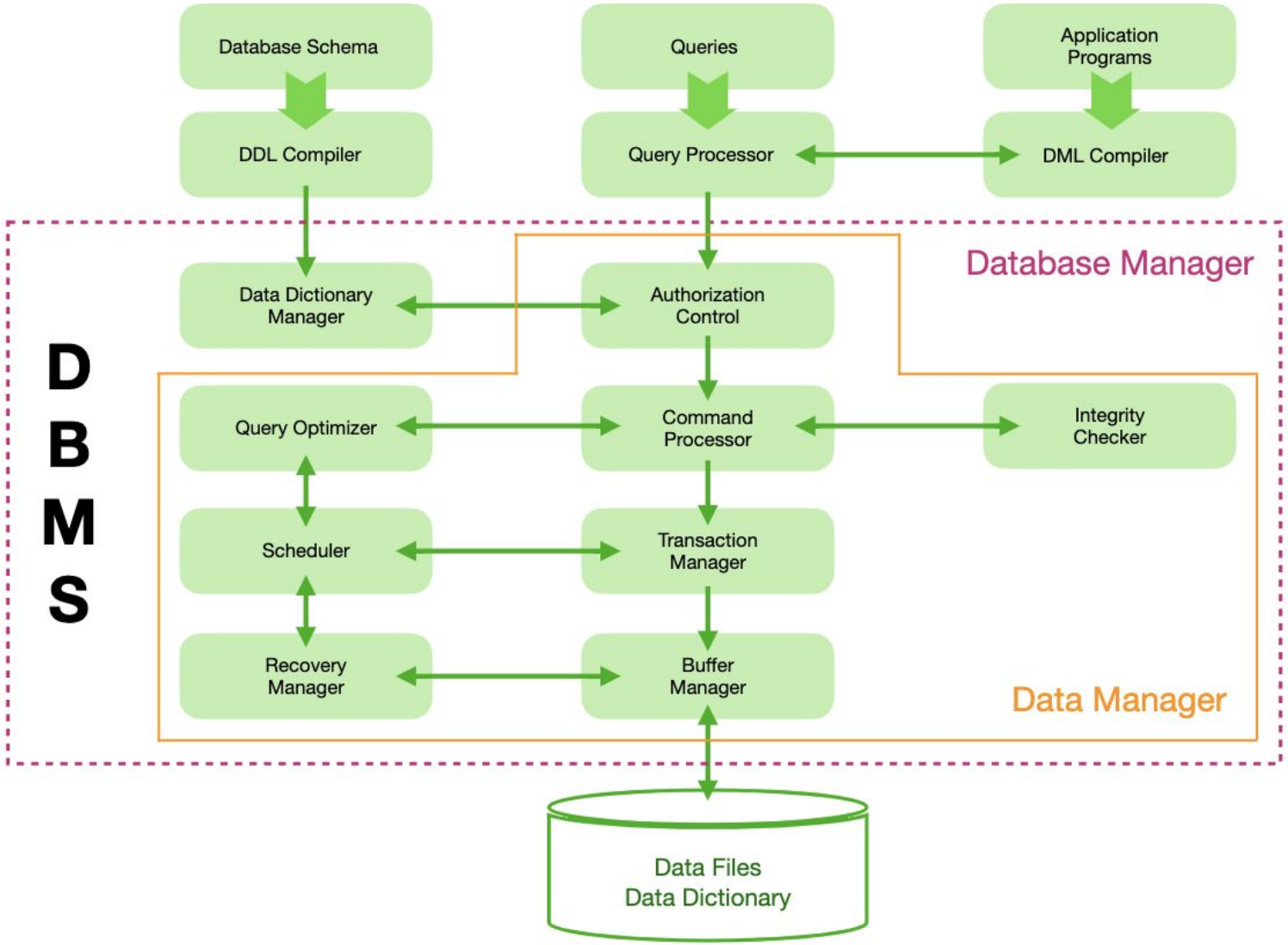
Вторая нормальная форма

- Создайте отдельные таблицы для наборов значений, относящихся к нескольким записям.
- Свяжите эти таблицы с помощью внешнего ключа.

Третья нормальная форма

- Устраните поля, не зависящие от ключа.

Пример РБД


Database Schema
Queries
Application Programs
DDL Compiler
Query Processor
DML Compiler
Database Manager
Data Dictionary Manager
Authorization Control
D
B
M
S
Query Optimizer
Command Processor
Integrity Checker
Scheduler
Transaction Manager
Recovery Manager
Buffer Manager
Data Manager
Data Files
Data Dictionary

Основные понятия реляционной модели. Ключи

INNOPOLIS UNIVERSITY

Первичный ключ (primary key constraint) - это поле или набор полей со значениями, которые являются уникальными для всей таблицы

Внешний ключ (foreign key) - это ограничение, которое поддерживает согласованное состояние данных между двумя таблицами, обеспечивая так называемую ссылочную целостность.

Основные понятия реляционной модели. Ключи

ПЕРВИЧНЫЕ КЛЮЧИ

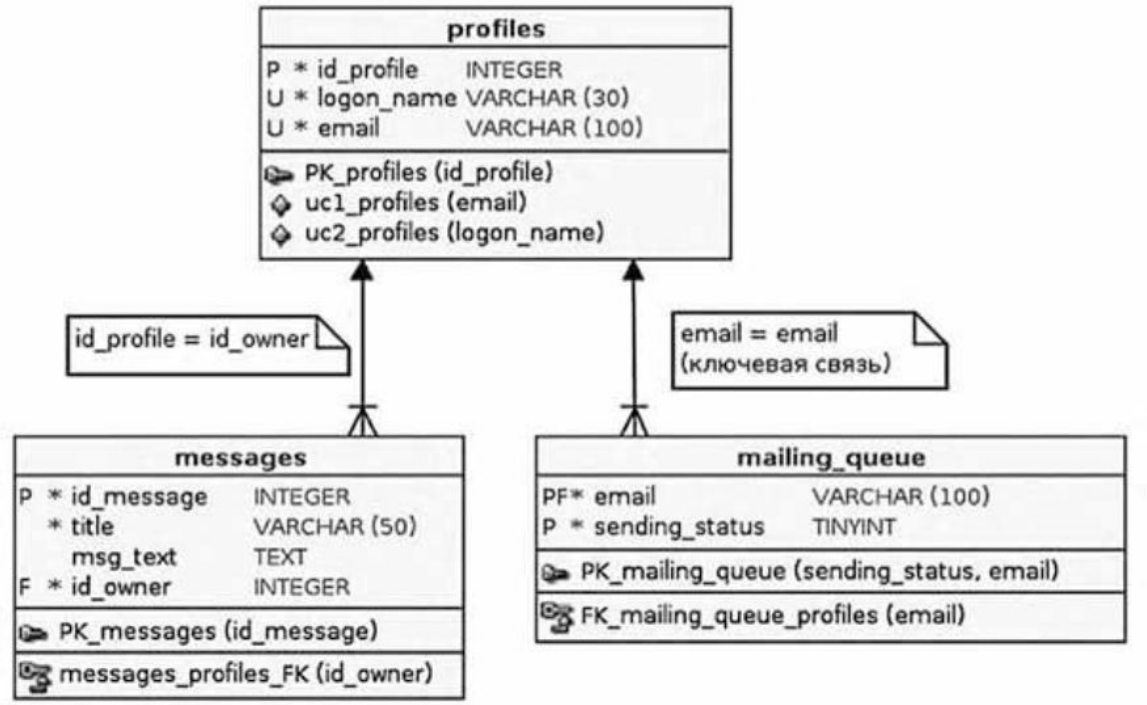
Ключ — минимальный набор атрибутов, совокупность значений которых однозначно определяет кортеж в отношении.

Первичный ключ (primary key constraint) - каждая строка таблицы должна содержать уникальное значение.

Основные понятия реляционной модели. Ключи

ВНЕШНИЕ КЛЮЧИ

Внешний ключ (foreign key) - это ограничение, которое поддерживает согласованное состояние данных между двумя таблицами, обеспечивая так называемую ссылочную целостность.



Основные понятия реляционной модели. Ключи

ТИПЫ ДАННЫХ

Числовые

- integer: хранит числа от -2147483648 до +2147483647. Занимает 4 байта. Имеет псевдонимы int и int4.
- bigint: хранит числа от -9223372036854775808 до +9223372036854775807. Занимает 8 байт. Имеет псевдоним int8.
- real: хранит числа с плавающей точкой из диапазона от 1E-37 до 1E+37. Занимает 4 байта. Имеет псевдоним float4.
- double precision: хранит числа с плавающей точкой из диапазона от 1E-307 до 1E+308. Занимает 8 байт. Имеет псевдоним float8.

Символьные

- character(n): представляет строку из фиксированного количества символов. С помощью параметра задается количество символов в строке. Имеет псевдоним char(n).
- character varying(n): представляет строку из не фиксированного количества символов. С помощью параметра задается максимальное количество символов в строке. Имеет псевдоним varchar(n).
- text: представляет текст произвольной длины.

Основные понятия реляционной модели. Ключи

INNOPOLIS UNIVERSITY

ТИПЫ ДАННЫХ

Дата и время

- timestamp: хранит дату и время. Занимает 8 байт. Для дат самое нижнее значение - 4713 г до н.э., самое верхнее значение - 294276 г н.э.
- timestamp with time zone: то же самое, что и timestamp, только добавляет данные о часовом поясе.
- date: представляет дату от 4713 г. до н.э. до 5874897 г н.э. Занимает 4 байта.
- time: хранит время с точностью до 1 микросекунды без указания часового пояса. Принимает значения от 00:00:00 до 24:00:00. Занимает 8 байт.
- time with time zone: хранит время с точностью до 1 микросекунды с указанием часового пояса. Принимает значения от 00:00:00+1459 до 24:00:00-1459. Занимает 12 байт.
- interval: представляет временной интервал. Занимает 16 байт.

Логические

- Тип boolean может хранить одно из двух значений: true или false.
- Вместо true можно указывать следующие значения: TRUE, 't', 'true', 'y', 'yes', 'on', '1'.
- Вместо false можно указывать следующие значения: FALSE, 'f', 'false', 'n', 'no', 'off', '0'.

Специальные

- json: хранит данные json в текстовом виде.
- jsonb: хранит данные json в бинарном формате.
- xml: хранит данные в формате XML
- массивы

Принципы формирования баз данных

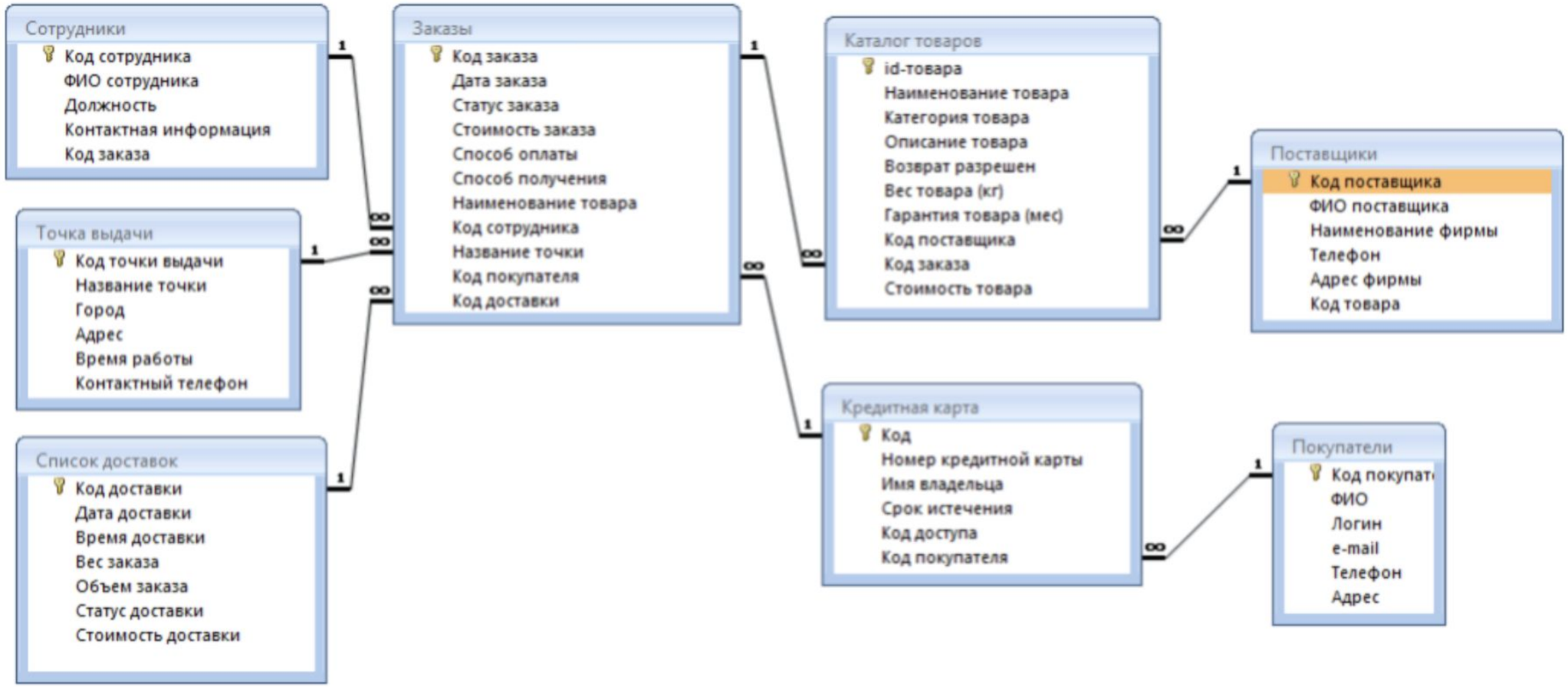
INNOPOLIS UNIVERSITY

- Данные представляются в виде таблиц
- Данные доступны логически. Это означает, что доступ к данным осуществляется не по номерам строк и столбцов
- NULL трактуется как неизвестное значение.
- БД должна включать в себя метаданные. БД хранит два вида таблиц: пользовательские и системные.
- Должен использоваться единый язык для взаимодействия с СУБД. В настоящее время в реляционных базах данных таким языком является SQL.
- Должны поддерживаться операции реляционной алгебры.

Пример реляционной БД

innopolis university

ER ДИАГРАММА



Сущности в БД

- Table

```
create table test as select * from Tablle where
datestart = ‘2020-01-01’
```

- View

```
create view test_v as select * from Tablle
where datestart = ‘2020-01-01’
```

- Materialized view

```
create Materialized view test_mv as select *
from Tablle where datestart = ‘2020-01-01’
```

Сущности в БД

- Routines

```
create function test_function(calculation_start timestamp without time zone,
calculation_end timestamp without time zone)
    returns TABLE(id bigint, start_date date, client_start_date date)
    language plpgsql
as
    $$
    BEGIN
        RETURN QUERY
    select * from Tablle where date_start = calculation_start and date_end =
    calculation_end
    END;
    $$
```

- Temporary table

```
create temporary table test as select * from
Tablle where date_start = '2020-01-01'
```

SQL (Structured Query Language

SQL (Structured Query Language — Структурированный язык запросов) — язык управления базами данных для реляционных баз данных.
SQL состоит из четырех отдельных частей:

язык определения данных (DDL - Data Definition Language) используется для определения структур данных, хранящихся в базе данных:

- Create - создание объекта в БД
- Alter - изменение объекта в БД
- Drop - удаление объекта в БД

язык манипуляции данными (Data Manipulation Language) используется для извлечения и изменения данных в БД:

- SELECT - выбор данных с условием
- INSERT - добавление новых данных
- UPDATE - изменение данных
- DELETE - удаление данных

SQL (Structured Query Language

язык определения доступа к данным (Data Control Language) используется для контроля доступа к данным в БД

- GRANT - предоставление доступа
- REVOKE - отзыв выданного доступа
- DENY запрет доступа

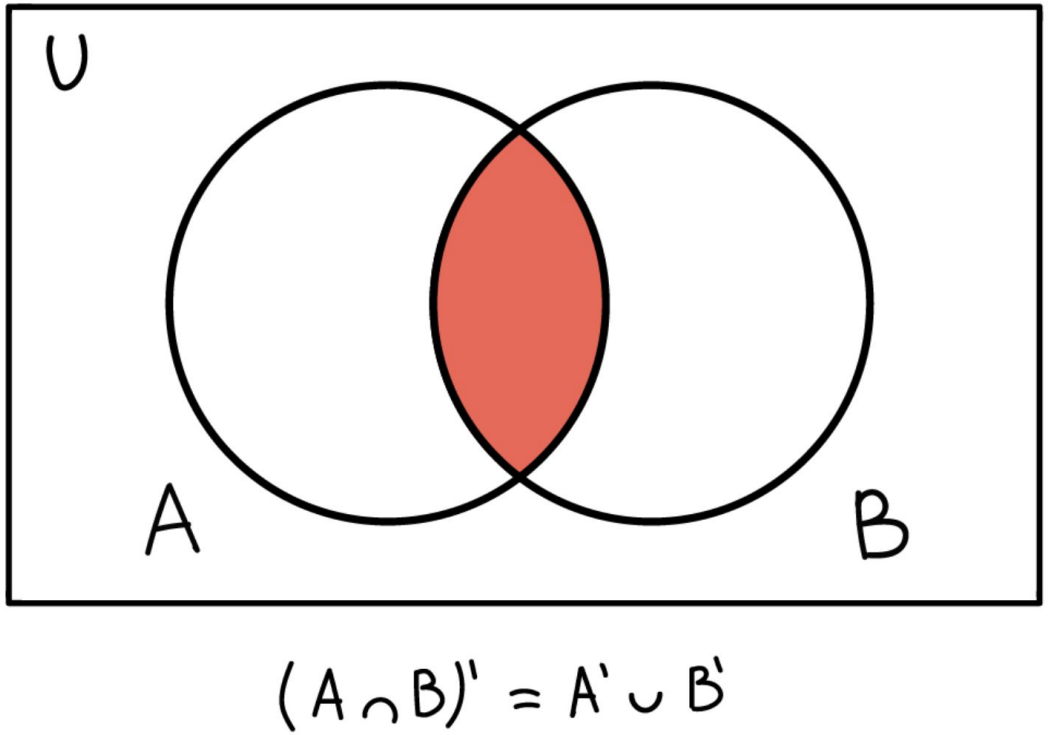
язык управления транзакциями (Transaction Control Language) используется для контроля обработки транзакций в БД

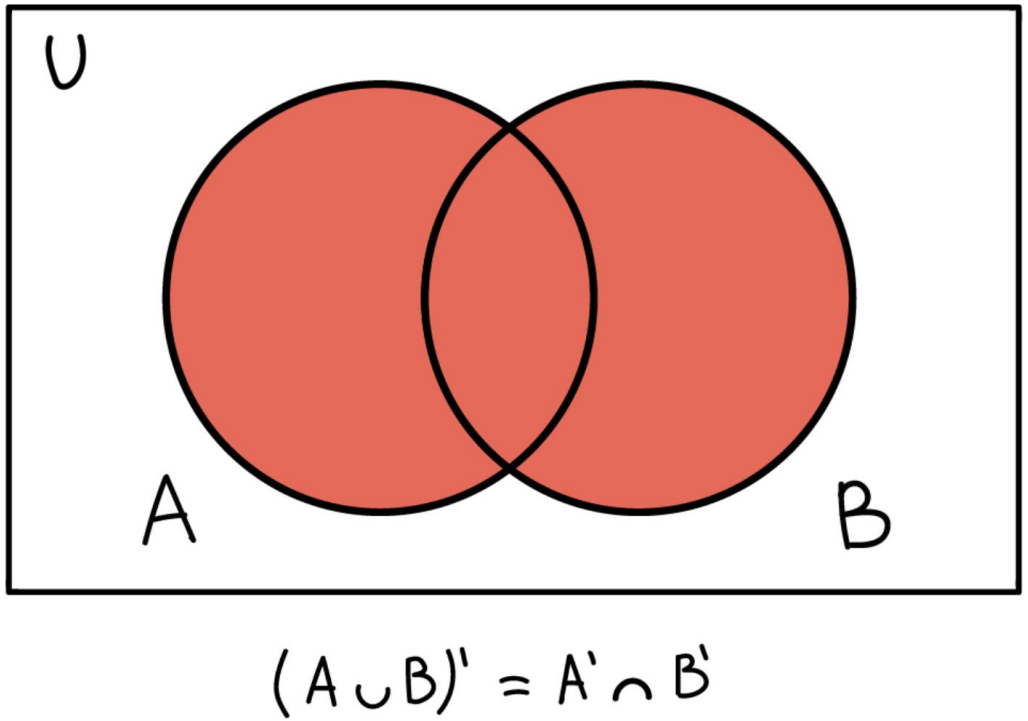
- BEGIN TRANSACTION - начало транзакции
- COMMIT TRANSACTION -
- ROLLBACK TRANSACTION - откат
- SAVE TRANSACTION - промежуточная точка

Основы SQL. Фильтруем данные правильно

Законы де Моргано должен знать каждый, кто работает с данными!!!

не (A и B) - это тоже самое, что не A или не B
не (A или B) - это тоже самое, что не A и не B





источник - https://ru.hexlet.io/courses/set-theory/lessons/de_morgan/theory_unit#:~:text=%D0%97%D0%B0%D0%BA%D0%BE%D0%BD%D1%8B%20%D0%B4%D0%B5%20%D0%9C%D0%BE%D1%80%D0%B3%D0%B0%D0%BD%D0%B0%20%D0%BE%D0%BF%D0%B8%D1%81%D1%8B%D0%B2%D0%B0%D1%8E%D1%82%2C%20%D0%BA%D0%B0%D0%BA,%D0%BE%D0%B1%D1%8A%D0%B5%D0%B4%D0%B8%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5%20%D0%BC%D0%BD%D0%BE%D0%B6%D0%B5%D1%81%D1%82%D0%B2%20%D1%87%D0%B5%D1%80%D0%B5%D0%B7%20%D0%B8%D1%85%20%D0%B4%D0%BE%D0%BF%D0%BE%D0%BB%D0%BD%D0%B5%D0%BD%D0%B8%D1%8F.

Основы SQL. Фильтруем данные правильно

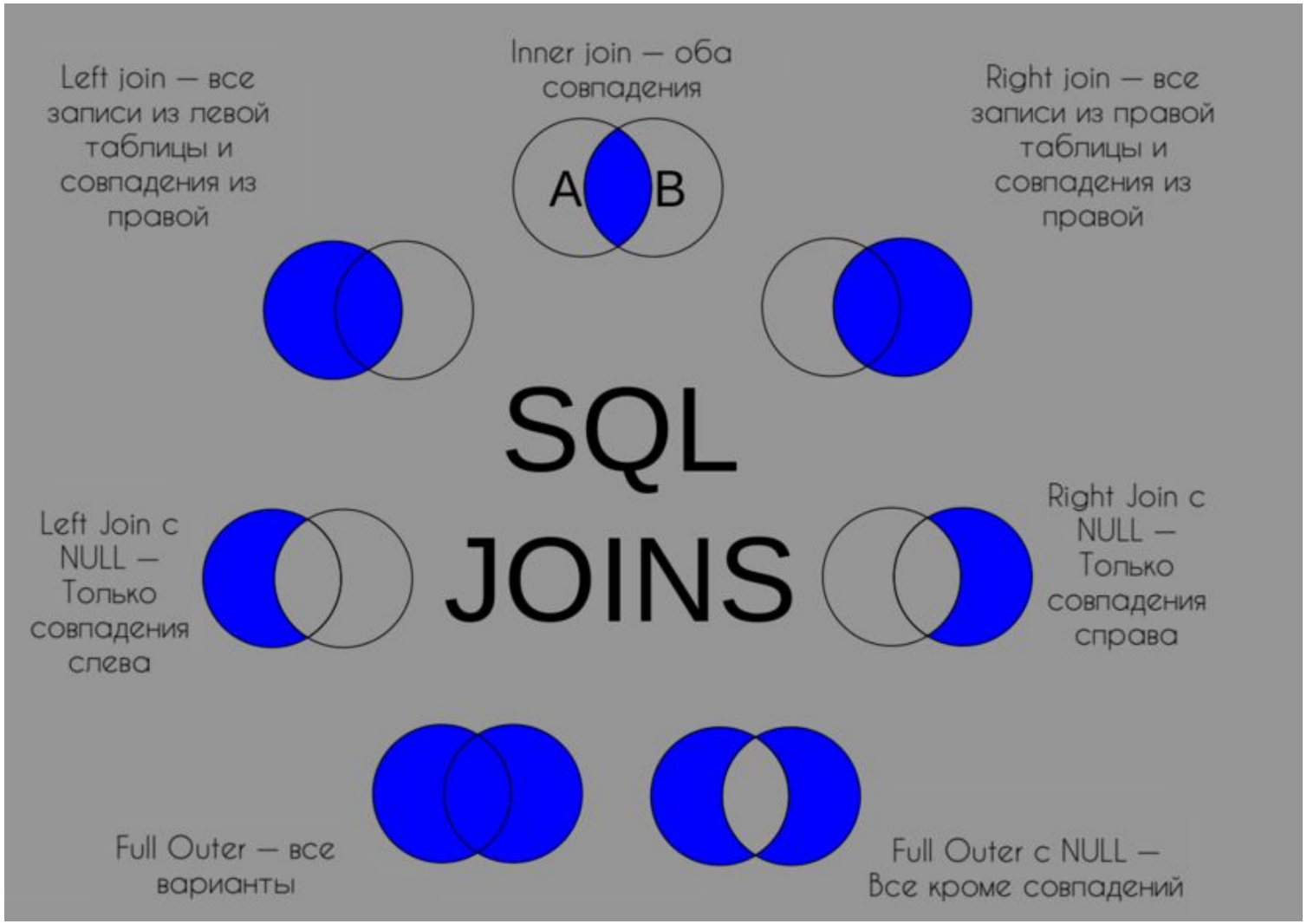
innopolis university

```
select
    distinct income_type, is_empl
from sample
where not(income_type = '2NDFL' and is_empl = 0);
```

ТОЖЕ САМОЕ

```
select
    distinct income_type, is_empl
from sample
where not income_type = '2NDFL' OR not is_empl = 0;
```

Основы SQL. Джойним данные правильно



источник - https://timeweb.com/ru/community/articles/kak-rabotaet-sql-join

Основы SQL. Джойним данные правильно

Как еще можно соединить таблицы?

- Union
- Union all

Бэкап БД

Резервным копированием называется сохранение копии данных где-то вне основного места их хранения.

Главное назначение резервного копирования – восстановление данных после их потери.

Источник фото: https://www.meme-arsenal.com/memes/49744a3f286732121b18344f96748312.jpg

Демо. SQL запросы

Полезные ссылки

тг каналы:

- https://t.me/sqlquestions
- https://t.me/sqlprofi
- https://t.me/prosql

курсы:

- https://www.sql-ex.ru/?Lang=0
- https://learndb.ru/
- https://stepik.org/course/63054/promo
- https://sql-academy.org/ru/guide

онлайн тренажер:

- можно писать sql запросы онлайн - https://sqliteonline.com/

статья

- для нчального погружения https://habr.com/ru/articles/145381/

УНИВЕРСИТЕТ
ИННОПОЛИС

ВОПРОСЫ И ОТВЕТЫ