Database design

Goals of a good data structure design

- Data Integrity: consistency, accuracy, avoiding anomalies
- Query performance: fast data retrieval Allow for future expansion of data types or relationships, evolution business requirements
- Scalability : growth in data volume Trade-offs between read and write performance
- Storage : minimize data redundancy
- Simplicity: Create an understandable structure for developers and analysts

simplicity is IMPORTANT!

How to design the data structure of a database ?

Some background and recap

OLAP vs OLTP : big dashboards or loads of transactions Entity-Relationship Diagrams : ERDs 1-1, 1-many, many-many relations Schemas in postgres

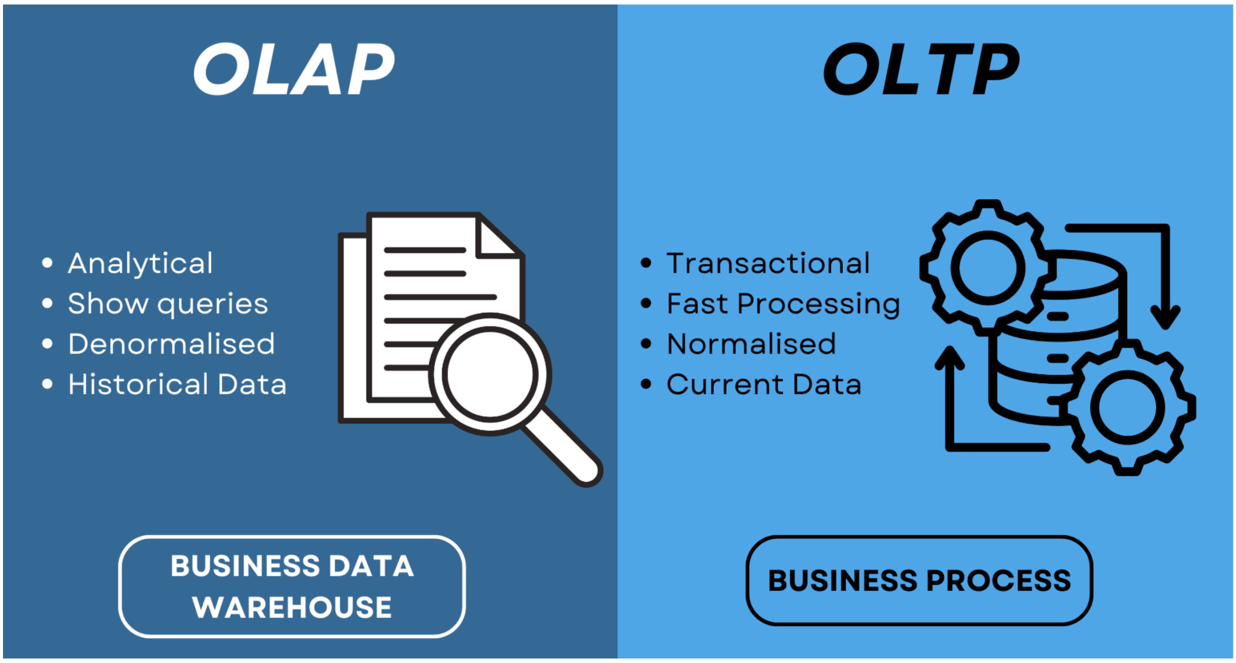
and then Normalization & Denormalization as efficiency tools normal forms 1NF, 2NF, 3NF functional dependency Practice : we normalize a flat dataset on the trees dataset

OLAP vs OLTP

Consider 2 databases. one for dashboards one for high volume of transactions

OLAP Online Analytical Processing analysis, BI, reporting, dashboards, high read volume OLTP: Online Transaction Processing, applications, transactions, high write volume high write volume

reading and Q&A https://www.geeksforgeeks.org/difference-between-olap-and-oltp-in-dbms/



OLAP

- Analytical
- Show queries
- Denormalised
- Historical Data

BUSINESS DATA WAREHOUSE

OLTP

- Transactional
- Fast Processing
- Normalised
- Current Data

BUSINESS PROCESS

quizz

For each scenario, determine whether it's more suited for an OLTP (Online Transaction Processing) or OLAP (Online Analytical Processing) system.

Liking a friend's post on Instagram. Analyzing trending hashtags on Twitter over the past month. Sending a Snapchat message to a friend. Netflix recommending shows based on your viewing history. Ordering food through a delivery app like UberEats or DoorDash. Spotify creating your personal "Wrapped" year-in-review of music. Making an in-app purchase in a mobile game. TikTok analyzing which video types keep users watching longer. Swiping right on a dating app like Tinder or Bumble. A fitness app calculating your average daily steps for the past year.

Liking a friend's post on Instagram. Analyzing trending hashtags on Twitter over the past month. Sending a Snapchat message to a friend. Netflix recommending shows based on your viewing history. Ordering food through a delivery app like UberEats or DoorDash. Spotify creating your personal "Wrapped" year-in-review of music. Making an in-app purchase in a mobile game. TikTok analyzing which video types keep users watching longer. Swiping right on a dating app like Tinder or Bumble. A fitness app calculating your average daily steps for the past year.

OLTP: Liking a post is a simple, immediate transaction. OLAP: Analyzing trends over time involves processing large amounts of historical data. OLTP: Sending a message is a quick, individual transaction. OLAP: Recommendations are based on analyzing viewing patterns across users and time. OLTP: Placing an order is a specific, time-sensitive transaction. OLAP: Creating a year-in-review involves analyzing a user's entire listening history. OLTP: An in-app purchase is a single, immediate financial transaction. OLAP: This involves analyzing user behavior patterns across many videos and users. OLTP: Swiping on a profile is a simple, immediate action. OLAP: Calculating long-term averages involves processing historical data over time.

Entity Relation Diagrams

super easy in pgAdmin

Peter Chen developed the ER diagram in 1976.

The ER model was created to provide a simple and understandable model for representing the structure and logic of databases.

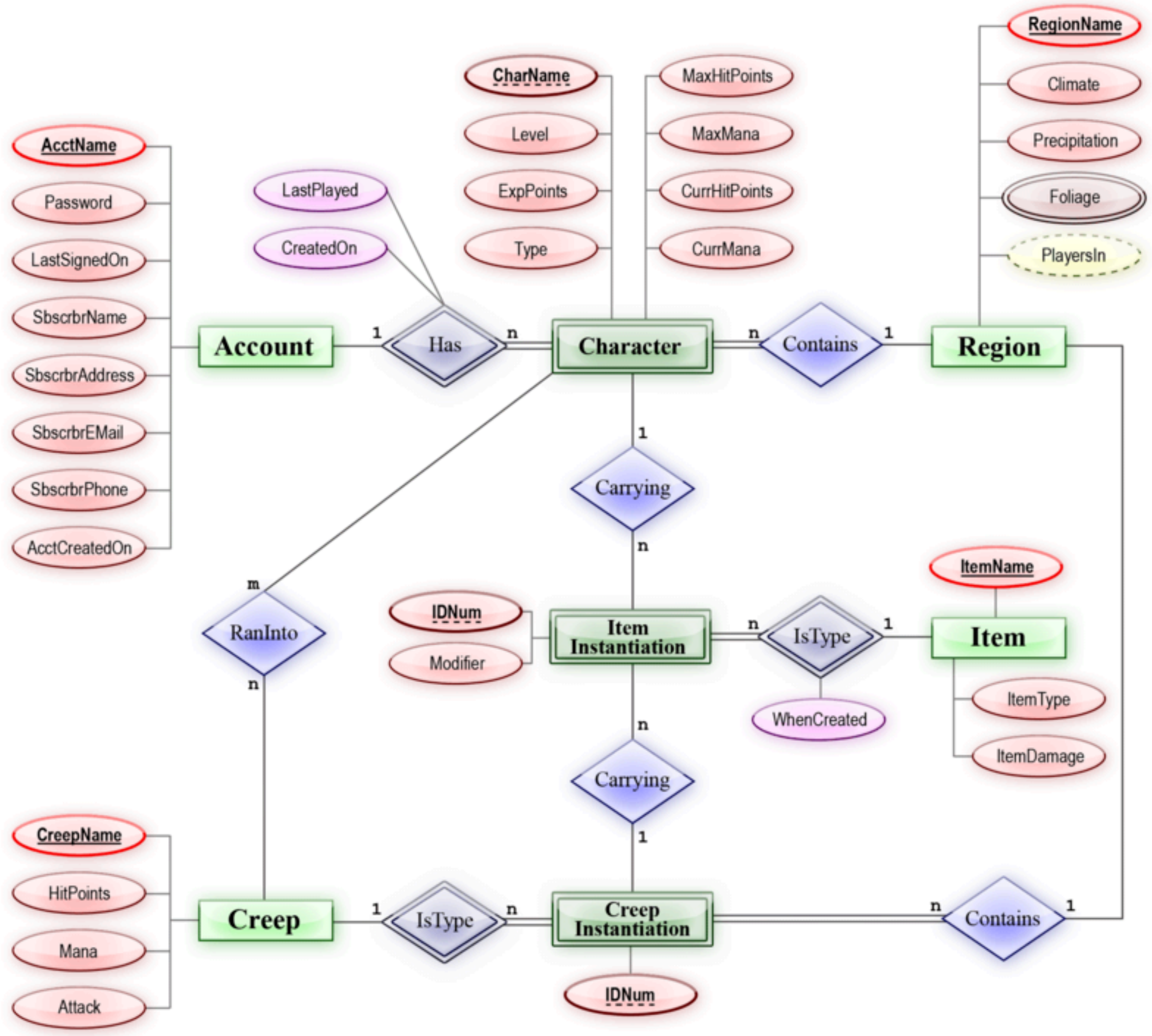
The Entity Relationship Diagram explains the relationship among the entities present in the database.

ER models are used to model real-world objects like a person, a car, or a company and the relation between these real-world objects.

In short, **the ER Diagram is the structural format of the database.**

https://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model

https://www.geeksforgeeks.org/introduction-of-er-model/



The article mentions entities, attributes in different types (string, weak, key, composite, …)

https://www.geeksforgeeks.org/introduction-of-er-model/

We need the equivalence: entity : table attribute : column

and the relations between entities: one to one one to many many to many

next : continue with slides to markdown