Databases, SQL and Querying

Davis Data Science Club

Agenda

1. Slides
    a. Data
    b. Databases
    c. Database management systems
2. SQL Review
    a. Syntax
    b. Aggregation and Filters
3. Mode.com login/overview
4. Interactive Questions

Data, Databases, Data Lakes??

Data

- Value types: datetime, float, int, string
- File types: .csv, .txt, .json, .RData, .xlsx

Data Storage Locations

- Locally
- Server hosted

Databases

- Manages collections of data and relationships between datafiles.
- Warehouse: Structured database with automations.
- Lake: Large collection of unstructured data, missing relation links.

What is a Database
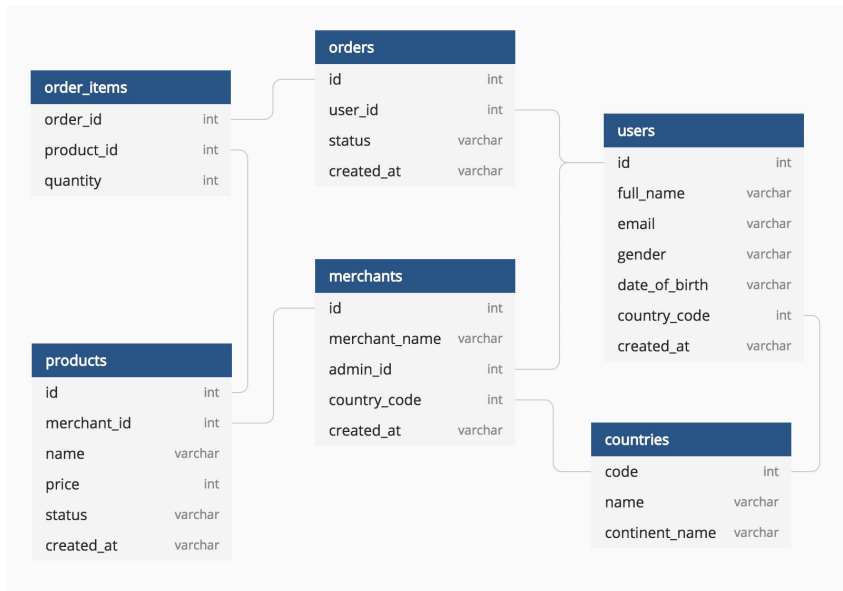
A scalable storage system

SQL vs NoSQL

- Relational
- Non-relational, dynamic schema

Data is added in various ways

- Manually editing records, online forms, website or app data

Schema (Layout/Map of your database)

- Important to plan out ahead of time

Relational vs Document Records

order_items

order_id	int
product_id	int
quantity	int

orders

id	int
user_id	int
status	varchar
created_at	varchar

users

id	int
full_name	varchar
email	varchar
gender	varchar
date_of_birth	varchar
country_code	int
created_at	varchar

merchants

id	int
merchant_name	varchar
admin_id	int
country_code	int
created_at	varchar

products

id	int
merchant_id	int
name	varchar
price	int
status	varchar
created_at	varchar

countries

code	int
name	varchar
continent_name	varchar

Students Document

```
{
    _id : "5373aadadcac133aad5b6660",
    name : "kaushal patel",
    email : "kp@example.com",
    courses : {
        course_name : "java",
        fees : "5000",
        duration : "3",
        professor : "g.r."
    }
}
```

Relational - Tabular
Data split across tables

Non-Relational - Document
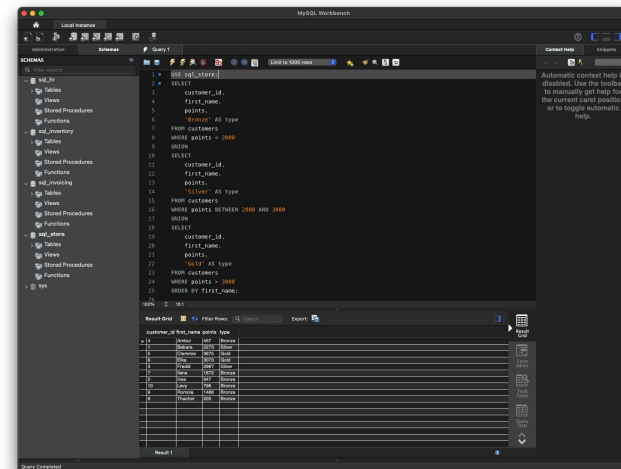Related data contained in a single document

Database Management Systems

Remote Services: MSFT SQL Server, AWS Redshift

- Connect using credentials (IP Address, Port Number, etc.)
- Scripts query and use data in memory.

Apps: MySQL, MongoDB Compass

- Log-In using credentials
- Query within GUI and export data to local as a file.

Data Center Management and Use

On-Site:

- Position: Database Administrator
- Update db, manage hardware load

Remote:

- Position: Data Engineer/Scientist
- Query, build pipelines, and use data

https://www.rgj.com/story/money/business/2018/01/17/sneak-peek-rare-look-inside-apple-reno-data-center/1038984001/

SQL Review

Mode.com SQL Practice

https://mode.com
Questions:
https://github.com/Davis-Data-Science-Club/SQL-workshop-aggretation-and-filtering/blob/main/aggregation-filtering-practice/question-1.md