SQL and Databases

Database

- What is it
  - Collection
  - Storage
  - Management
  - ….. of data
- System designed for storing and retrieving data.

Why you should know

- Data you are interested in or have to work with is stored in a database or multiple databases
- You have to Extract, Transform and Load (ETL) data before you start working with it.

Types

- Relational

- Non-Relational

Relational

- Data is organized and stored in tables - rows and columns.
- Data points are related to one another.
- Each row in the table is a record with a unique ID called the key.
- Columns are the attributes of the data.

- Examples: Sqlite, PostgreSQL, MySQL, Oracle, SQL Server

Non-relational

- Does not use the tabular format/structure - rows and columns
- Storage model is optimized for the specific type of data being stored
- Most likely document oriented
- Example: Key/value pairs, JSON documents, Graph
- NoSQL databases
- Examples: MongoDB, Cassandra, Coachbase

SQL

- Structured Query Language
- Pronounced as “sequel,” or “es-queue-el”
- Standardized language for querying and manipulating data
  - Set of commands for working/interacting/communicating with the database
- Essential part of your toolkit

SQL

- CRUD Operations
  - Create
  - Read
  - Update
  - Delete

- DDL - Data Definition Language
    - CREATE
- DML - Data Manipulation Language
    - INSERT, UPDATE
- DQL - Data Query Language
    - SELECT
- DCL - Data Control Language

Relational DBs

Advantages

- ACID
  - Atomicity
  - Consistency
  - Isolation
  - Durability
- Normalization
  - First Normal
  - Second Normal

Relational DBs

Drawbacks

- Organizing data into a particular structure
- Cannot handle complexity
- Change to one record needs change to all records
- Not suitable for Big Data - Volume, Variety, Velocity
- Scalability

Non-Relational DBs

Advantages

- No schema
- Scalability
- BASE
  - Basically available
  - Soft state
  - Eventually consistent

Non-Relational DBs

Drawbacks

- Lack of consistency
- Lack of analytics
- Lack of standardization

New SQL

- Aims to combine/provide benefits of both relational and non-relational dbs.
- Speed, scalability, consistency
- Ideal for Big Data.