

SQL Fundamentals

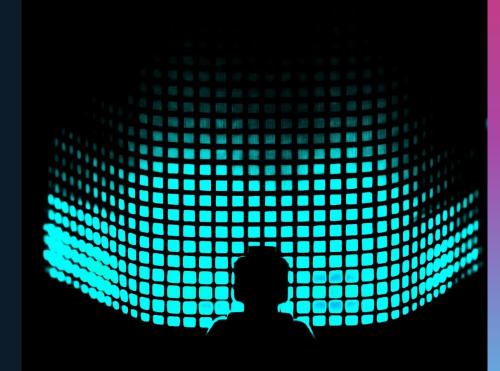
Kick off your career in data science & analytics

Module 1: Introduction to Data and Databases



In this module, we will introduce SQL as a programming language for managing databases. Specifically, we will share with you:

- What is SQL
- Why we need to learn SQL
- What we can achieve with SQL



What is Data?

What is a Database?

Agenda.



Data are a set of values about one or more persons or objects. There are three types:

<u>wikipedia</u>



Reference: <u>Data Types: Structured vs. Unstructured Data</u>





1, 000, 000 rows



1, 000, 000, 000 rows

What can we do if there's more data we want to store than what the program allows for?





Storing Data (Vlookup?)

Introduction to SQL

nfldb Entity-Relationship diagram (condensed) home meta version [smallint, null] season_type [season_phase, null] team ason year [usmallint, null] team id [varchar, not null] ek [usmallint, null] player city [varchar, not null] 0..NO..N prayer id [varchar, not null] **CSV** ne [varchar, not null] game 0..N 0..N full name [varchar, null] drive gsis_id [gameid, not null] team [varchar, not null] CSV start time [utctime, not null] gsis id [gameid, not null] position [player_pos, not null] play_player 0..N week [usmallint, not null] drive id [usmallint, not null] status [player_status, not null] gsis id [gameid, not null] season_year [usmallint, not null] start_field [field_pos, null] drive id [usmallint, not null] season_type [season_phase, not null] start_time [game_time, not null play_id [usmallint, not null] end_field [field_pos, null] finished [boolean, not null] player_id [varchar, not null] play home team [varchar, not null] end time [game time, not null] [varchar, not null] gsis_id [gameid, not null] pos_team [varchar, not null] home_score [usmallint, not null] drive_id [usmallint, not null] team [varchar, not null] pos period, null] CSV play id [usmallint, not null] score [usmallint, not null] time [game_time, not null] CSV CSV pos_team [varchar, not null] yardline [field_pos, null] down [smallint, null]

CSV

to_go [smallint, null]

0.N

What is Data?

What is a Database?

Agenda.



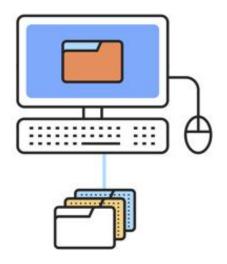


Image: <u>Background vector created by rawpixel.com</u>
<u>- www.freepik.com</u>

A database is an organized <u>collection of data</u>. It is a collection of schemas, tables, queries, reports, views, and other objects.

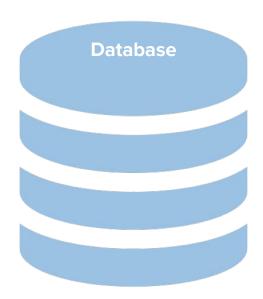
Database designers typically organize the data to model aspects of reality in a way that supports processes requiring information.

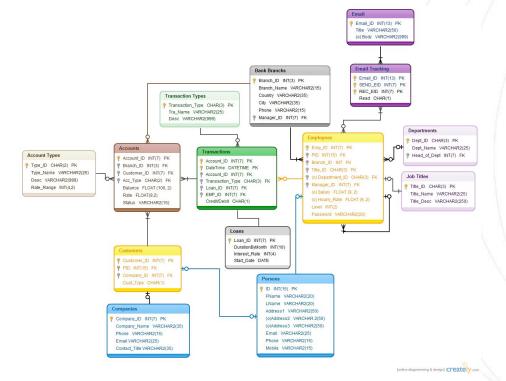
E.g. Modelling the availability of rooms in hotels in a way that supports finding a hotel with vacancies.

wikipedia





















Relational
Tend to be larger, monolithic



















Non-relational
Newer field, lots of players













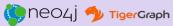
Document Database

Pair each key with a complex data structure known as a document



Graph Database

Storing graphs to model relationships and dependencies





Key-value Store

Collections of key-value pairs



Column-family

Wide-column stores that are optimized for queries over large datasets





Search Engine DB

Full-text search



