School and Pool for Digital Talent

Why learn about Databases?

Why?

- Most of (large) data is stored in databases
- An analyst needs to be able to connect to a database and access its data in order to unlock insights

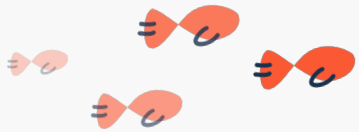
What?

- Understand the basics of databases
- Be able to connect to a database

How?

- Learn about database types and how they are structured
- Connect to a database and explore its content

Introduction to Databases



Databases

are a systematic collection of data

store data on disk (cheaper) or in-memory (faster)

support electronic storage and manipulation of data

Types of Databases

RDBMS: **R**elational **D**ata**b**ase **M**anagement **S**ystems

- **use SQL** to query RDBMS
- have a predefined schema
- data is stored in tabular form of columns and rows
- the connection between data/tables is relational
- Examples: Postgres, MySQL, Oracle, SQLite

NoSQL Databases

- **Use something other than SQL** as the primary language
- have no predefined schema
- Examples: Neo4j, Elasticsearch, MongoDB

RDBMS

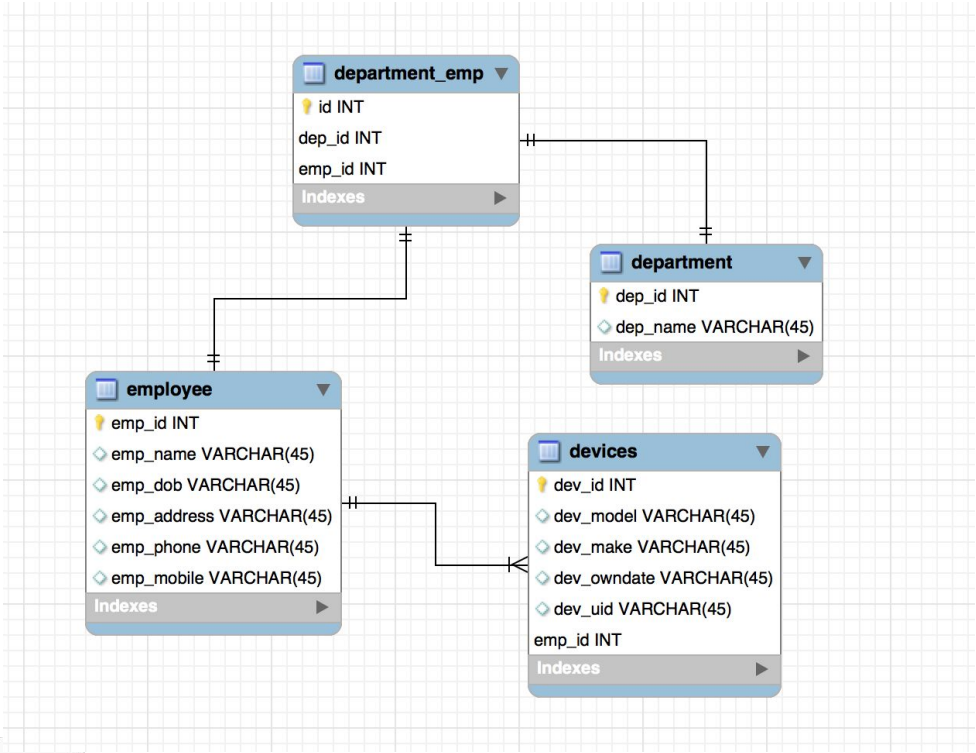
- Many different types of databases exist
  and each uses a different flavour of SQL
- Their syntax can differ, but the core concepts are the same
- Some databases will implement a subset of the functionality
- Some DB will be optimized for speed of read,
  others for speed of write

Database Structure

- A database consists of one/multiple schemas
- Schemas consist of tables
- Tables consist of columns and rows
- A column is a variable and has a unique name
- A row is an observation
- Every cell is a single value

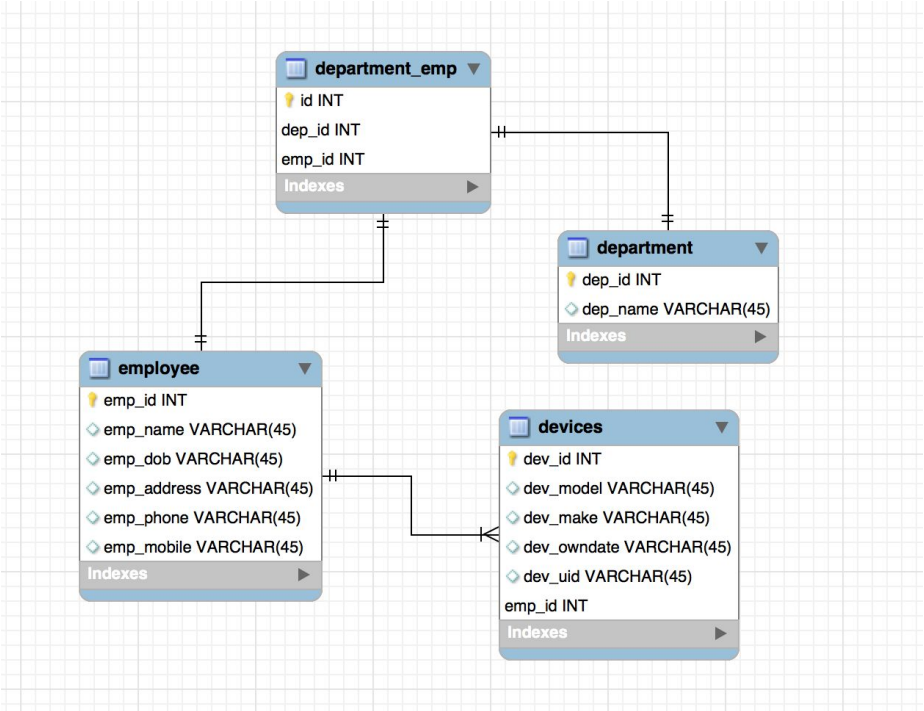
Entity-Relationship model

- Entities can be of type:
  object, class, person or place
- The properties of an entity are described through their attribute(s)
- Relationships describe the relation between entities
- Different types of relationship exist



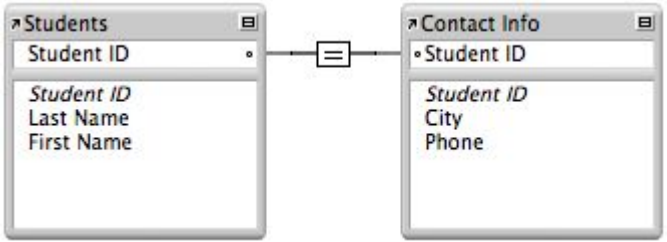
Relational Databases

- Tables are related via primary and foreign keys
- Each table has **one** primary key that is unique for each record
- A foreign key is a field (or collection of fields) in one table, that refers to the primary key in another table



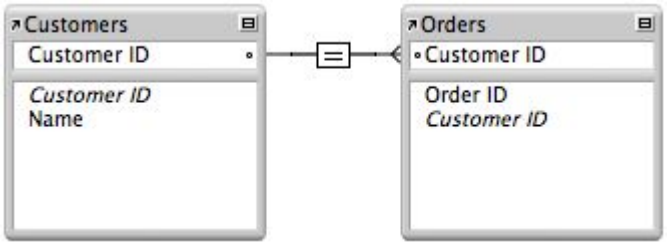
Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)



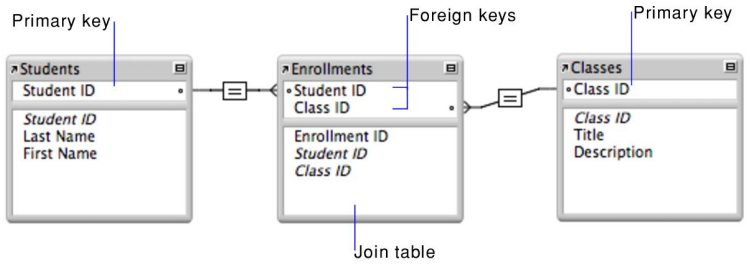
Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)

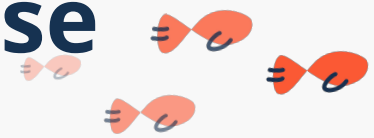


Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)



Connecting to a Database



SQL Client / Database IDE

- IDE = **I**ntegrated **D**evelopment **E**nvironment
- Powerful software that can be used to connect to a database and retrieve and visualise data (and more!)
- Local or in the cloud
- Open-source, free and paid software is available

Local SQL Clients

- Installed and run locally on your machine

Examples:

Cloud SQL Clients

- Deployed in the cloud and accessed via a web-interface

Examples:

DBeaver

- In this course we will use DBeaver

Why?

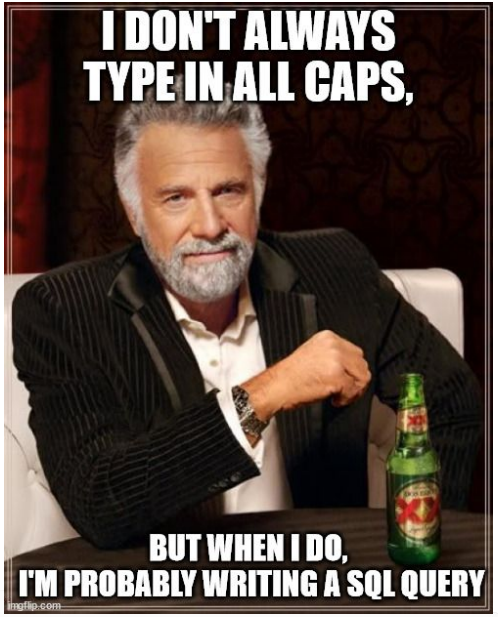
- free and open source universal database tool
- Works for many different types of databases
- Cross platform (Windows, Linux, Mac OS, Solaris)

Setting up DBeaver

Open DBeaver > Window > Preferences > Editors

- Enable upper case:
  SQL Editor > Formatting > Keyword Case > Set to: Upper

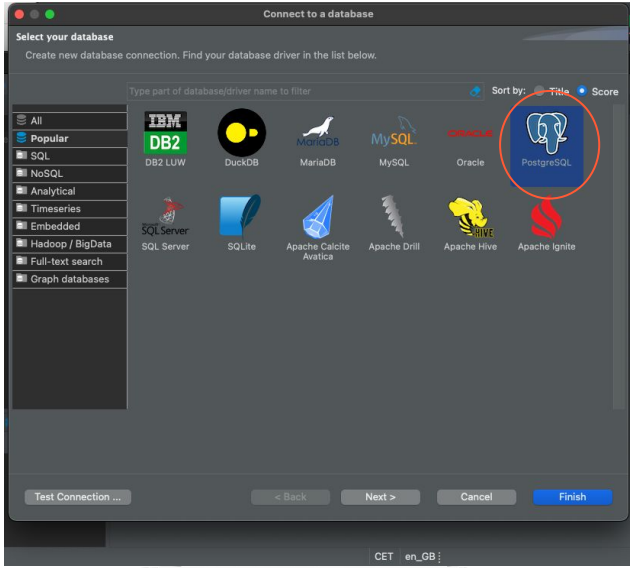
- Add line numbers:
  Text Editors > Show line numbers > Tick box



Connect to a PostgreSQL database

1. Click on “New Database Connection” 
2. Download Driver (if necessary)
3. Search for and select PostgreSQL
4. Enter the connection details below

Host	data-analytics-course-2.c8g8r1deus2v.eu-central-1.rds.amazonaws.com
Port	5432
Database	postgres
Username	Will be sent to you via e-mail / posted in Slack/Zoom Chat
Password	Will be sent to you via e-mail / posted in Slack/Zoom Chat



Exploring the database

Find the database connection in your “**Database Navigator**” pane on the left
Expand it to:

postgres > **Databases** > **postgres** > **Schemas** > **cgn_dp_23_2** > **Tables**

Check out the flights and airports tables and answer the following questions:

1. What happens if you double click on a table?
2. What is the first and last airport listed in the airports table?
3. What data type is the column ‘flightdate’ in the flights table?
4. What type of entity-relationship exists between the flights and airports table
5. Name the primary and foreign keys

Set course schema as default

Set the schema for this class as default to make queries easier
Expand it to:

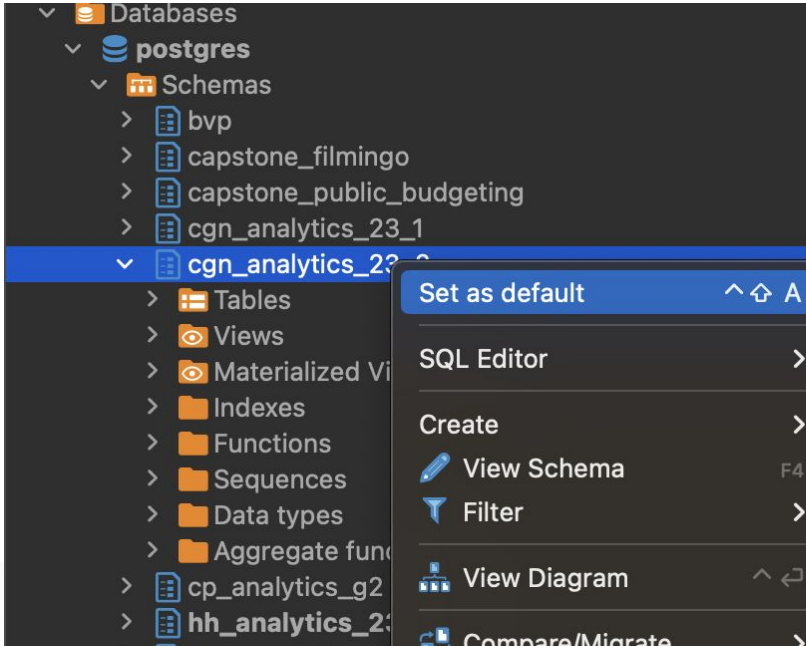
postgres > **Databases** > **postgres** > **Schemas**

click on **cgn_dp_23_2**

1. **ctrl + shift + A**

   or

2. right click and choose set as default



Fork and Clone the repo

https://github.com/neuefische/da-internal_data_sourcing