WARWICK

QSTEP SQL Masterclass

James Tripp, Senior Academic Technologist, CIM (Warwick)

1st December 2020/ Microsoft Teams

James Tripp

- Background in Psychology (BSc, PhD)
- Senior Academic Technologist at CIM (Centre for Interdisciplinary Methodologies)
- SQL?
    - **Large datasets**
    - **Analysis**
    - **System Administration**

WARWICK

SQL?

- Structure Query Language. A domain specific language – designed for a specific purpose
- Relational databases
- Give the database the query. The database then gets the data in an optimal way.

See:

https://en.wikipedia.org/wiki/SQL

https://www.w3schools.com/sql/sql_intro.asp

Our Data

Table name: world_indicators

country	country code	electricity	area
United Kingdom	GBR	100	241930
…	…	…	…

Row

Column

Table name: world_borders

iso3	population	geography
GBR	60244834	01060000 …
…	…	…

WARWICK

E.g.,

```
SELECT column FROM table;

SELECT country FROM world_indicators;

SELECT population FROM world_borders;
```

Why use a database?

- Data storage
- Efficient data querying via SQL
- A little complicated…

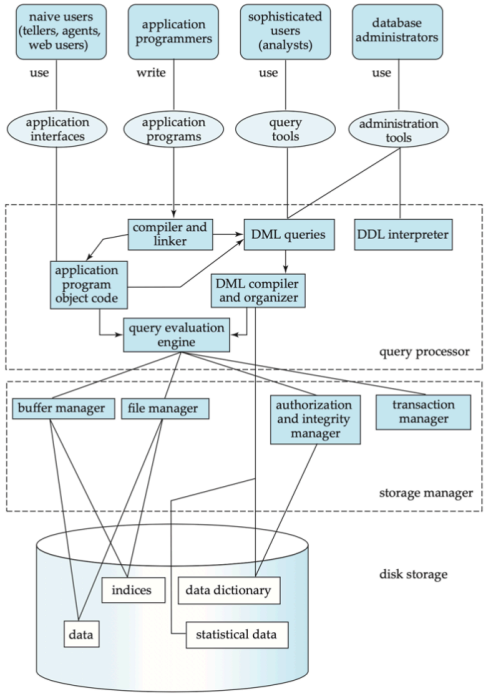
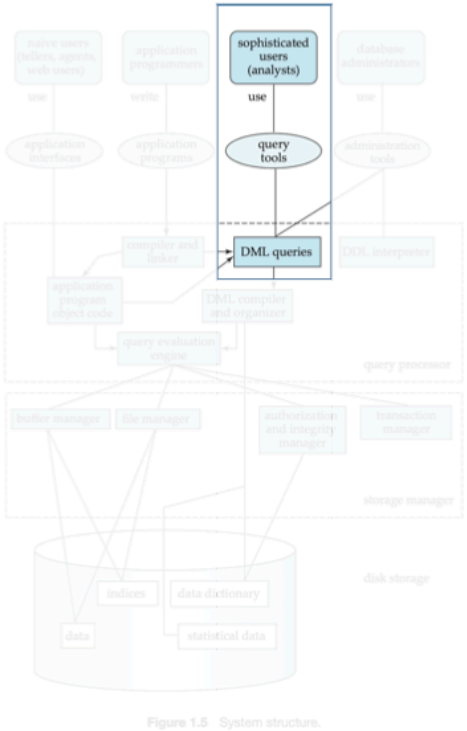


Figure 1.5 System structure.

From Database Systems Concepts (6th Edition)

Why use a database?

- Data storage
- Efficient data querying via SQL
- A little complicated…



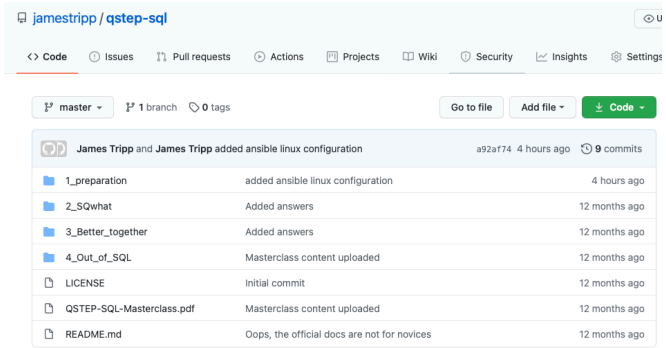
From Database Systems Concepts (6th Edition)

Our brief

- We are data analysts/data scientists
- Question: What is the relationship between two world indicators across countries?
- Our data is in a table called world_indicators in the qstep database
- An additional table called world_borders contains the geospatial borders of countries. This may help with visualization work

Materials

- Located in Github
- https://github.com/jamestripp/qstep-sql



WARWICK

Today

- Introduction (this presentation)
- Local installation of database and data (optional)
- SQWhat?
    - **Basic SQL introduction**
- Better together
    - **Aggregate function for data and joining tables**
- Out of SQL
    - **Taking data from the database**
    - **Using R for data analysis and visualisation**