



# QSTEP SQL Masterclass

James Tripp, Senior Academic Technologist, CIM (Warwick)

---

3<sup>rd</sup> December 2019 / Radcliffe

## James Tripp

- Background in Psychology (BSc, PhD)
- Senior Academic Technologist at CIM (Centre for Interdisciplinary Methodologies)
- SQL?
  - Large data sets
  - Analysis
  - System administration

## SQL?

- Structured Query Language. A domain specific language
- Relational databases.
- You send the query. The database handles getting the data.

See:

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

<https://www.youtube.com/watch?v=FR4QleZaPeM>

## Our Data

Table name:

world\_indicators

Row	country	country code	electricity	area
	United Kingdom	GBR	100	241930
	...	...	...	...

Column

world\_borders

iso3	population	geography
GBR	60244834	01060000 ...
...	...	...

E.g.,

- `SELECT` *column* `FROM` *table*;
- `SELECT` country `FROM` world\_indicators;
- `SELECT` population `FROM` world\_borders;

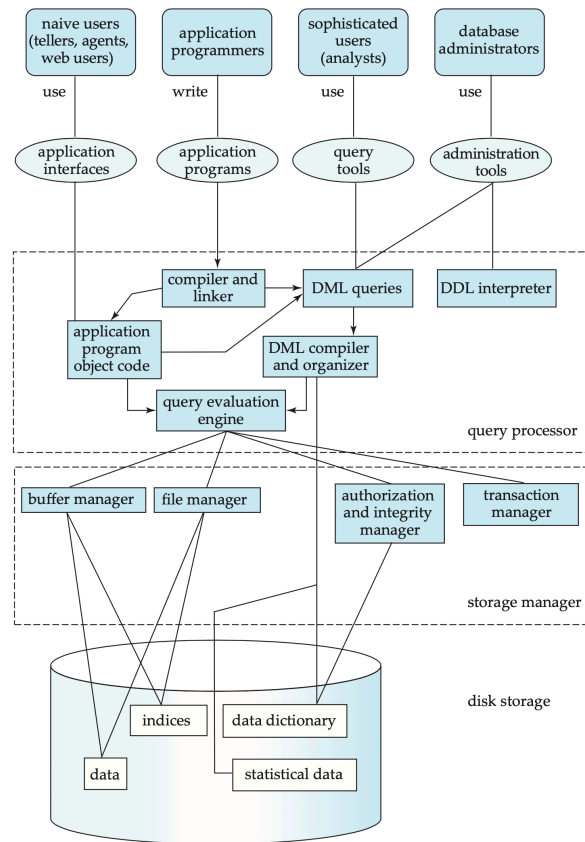


Figure 1.5 System structure.

## Why use a database?

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

## Why use a database?

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

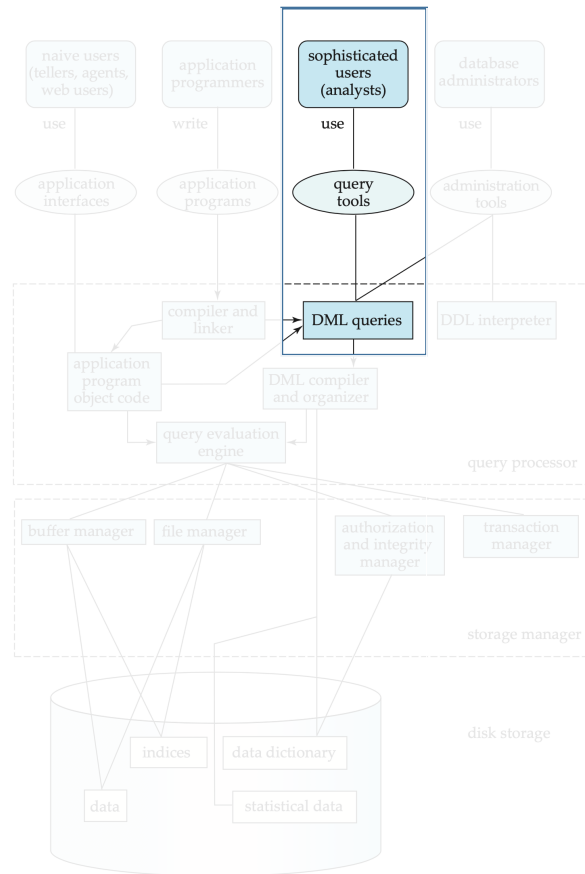


Figure 1.5 System structure.

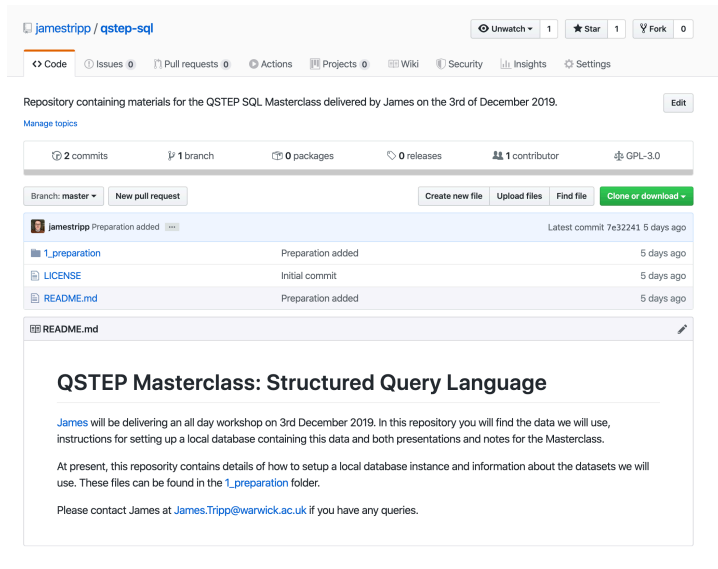
## Our brief

- We are data analysts
- 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 on GitHub
- <https://github.com/jamestripp/qstep-sql>



The screenshot shows the GitHub repository page for `jamestripp/qstep-sql`. The repository description states: "Repository containing materials for the QSTEP SQL Masterclass delivered by James on the 3rd of December 2019." The repository has 2 commits, 1 branch, 0 packages, 0 releases, and 1 contributor. The file list shows:

File	Commit	Time
<code>1_preparation</code>	Preparation added	5 days ago
<code>LICENSE</code>	Initial commit	5 days ago
<code>README.md</code>	Preparation added	5 days ago

The `README.md` file content is as follows:

**QSTEP Masterclass: Structured Query Language**

James will be delivering an all day workshop on 3rd December 2019. In this repository you will find the data we will use, instructions for setting up a local database containing this data and both presentations and notes for the Masterclass.

At present, this repository contains details of how to setup a local database instance and information about the datasets we will use. These files can be found in the `1_preparation` folder.

Please contact James at [James.Tripp@warwick.ac.uk](mailto:James.Tripp@warwick.ac.uk) if you have any queries.

← Download worksheets and datasets

← Each folder is a section

## 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