Querying a medical database

Theodoros Diakonidis

**Email: diakonidis@auth.gr**

ARISTOTLE
UNIVERSITY
OF THESSALONIKI

Purpose of the course

- A medical data analyst-scientist should use all the information in hands to acquire valuable data.

- In order to achieve all the above knowledge of how to handle one of the commonest databases is necessary.

- This lecture is going to introduce you to basic querying of relational databases using SQL language, through an example, a demo, of one of the most useful medical databases so far, the one called MIMIC.

What is a database?

- It is a collection of data residing on a computer.
- The collection is stored in different tables
- Each row in the table is considered as a record
- Every record is broken down into fields that represent single items of data describing a specific thing.
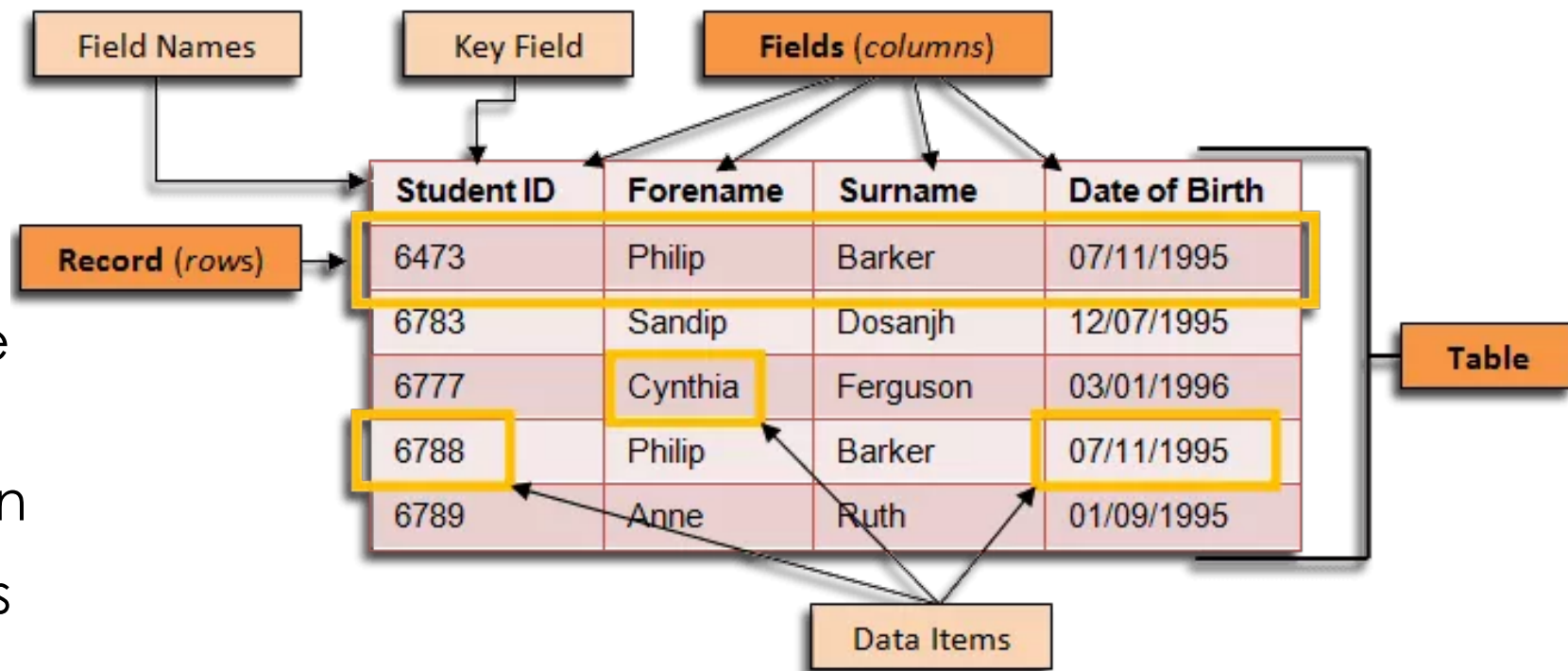
What is SQL?

⊙ Stands for Structured Query Language.

⊙ It is a standard programming language specifically designed for storing, retrieving, managing or manipulating the data inside a Relational Database Management System (RDBMS).

⊙ It is the most widely-implemented database language and it is adopted by a lot of relational database systems you may have already heard:

Oracle,MySQL,Microsoft SQL Server, PostgreSQL ,SQLite ,Teradata, Microsoft Azure SQL Database

What will be our focus in SQL?

Lot of things that one can do using SQL:

- create a database initially
- create tables and populate them
- update, modify them

But we will focus mostly on:

- extracting useful info from database, by querying, calculating by using different functions applying in the fields of the database etc.
- connecting tables together to extract information from common fields.

What is actually a query?

'A query is a statement written in SQL, which may include commands and actions, written to solicit an answer to a question or to perform an action'

- Think of it as a sentence that includes verbs, nouns, prepositions, clauses. E.g.

```
SELECT * FROM prescriptions WHERE drug LIKE '%Magnesium%'
```

It's purpose is to extract the information that is actually needed

What SQL can do? Is the SQL language the same to all RDS's?

- Some features of the SQL standard are implemented differently in different database systems.

- In general though most of the commands that we are going to use in this lecture are the same for all of them with very small variations.

- In our case we are going to use the SQLite RDS.

Why SQLite?

Some of the advantages:

- SQLite is a software library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine

- it is easy to install and maintain

- It is the most widely deployed SQL database engine in the world.

- SQLite comes with zero-configuration, which means no setup or administration needed

What is actually MIMIC?

⊙ MIMIC stands for :

'Medical Information Mart for Intensive Care'

It is described by the people that contributed to it:

"… Is a large, freely-available database comprising deidentified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012"

MIMIC III DEMO
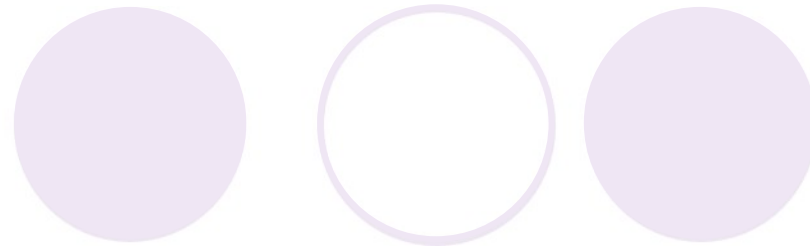
- We are going to deal with the third version of this database MIMIC III.

- What we are actually going to query, is a demo version of the package with fictitious numbers.

Do not by alarmed if you come across with strange numbers!!!

- Our purpose is to understand and learn how to query, extract the info that we are seeking for.

How am I going to follow the lecture?

⊙ What is needed:

Compulsory:

⊙ You need to install Jupyter Notebook according to the instructions that given in the email I sent you.

⊙ There you can run commands of sql, see the results from your queries and in the same time read instructions and info on how to do it.

Advisable:

Use also the sqlitebrowser to take a better look of the tables and run the commands with no python extras.

SQLiteBrowser

- After installing and running the browser you can open the database following the path that you have downloaded it

File->Open Database->

Your_Path_to_the_database/mimic3.db

There you can directly see all your tables from the table dropdown menu(Browse Data tablet) and run sql queries and see the results from the Execute SQL tablet.

SqliteBrowser continued

Execute SQL for running your queries

Browse Data tablet for checking the database

Table Dropdown menu of all tables in database

Using Jupiter Notebook to run SQL

- As Jupiter runs through Python, to execute a statement will require applying minimal amount of python code in our case:

  - the **%sql** before each command and
  - **/** to continue a line of code to a new line.

- But you will do this thing all the time in the future so do not worry it will come handy quite fast.

EXTRAS
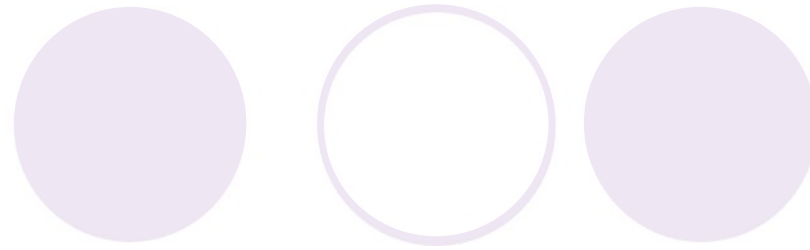
The NULL value

- The NULL value means a missing or unknown value
- Conceptually it is different than the zero "0" value or an empty space string " " value.
- Comparing a NULL with a non-null value will return a NULL
- An arithmetic operator using a NULL value will return a NULL.
- To make it more intriguing, comparing between two nulls will return a NULL as well.
- When ordering / sorting results, the NULL values will appear first (if ascending order).
- Any expression that contains a NULL will return a NULL.