SQL for Data Analysis

Course Contents

- What is Database?
- What is SQL
- How Do Data Analysts Use SQL?
- Jupyter Magics with SQL
- Working with Jupyter Notebook
- Basic SQL Commands

What is Database?

Database is a systematic collection of data. Databases support storage and manipulation of data. Databases make data management easy.

The most innovative structures for storing data today are **NoSQL** and **Relational Databases**.

Relational Database (SQL):

A relational database consists of a collection of **tables**, each having a unique **name**. A **row** in a table represents a **relationship** among a set of values. Thus a table represents a **collection of relationships**.

Relational Database (SQL):

- **Tables** in the database, which are known as relations.
- **Records**, which are the number of rows present in the database. They are also known as tuples.
- **Attributes** are the data categories present in columns.
- **Primary Key** is a unique value that identifies information in the database.
- **Foreign Key** links the primary table to another table.

EMPLOYEE

EMP_ID	EMP_NAME	ADDRESS	DEPT_ID
100	Anup	Nagpur	10
101	Darshan	Hyderabad	20
102	Subhash	Bangalore	10
103	Divija	Nagpur	30
104	Sanket	Pune	30

DEPARTMENT

DEPT_ID	DEPT_NAME
10	Accounting
20	Quality
30	Design

NoSQL Database:

- A NoSQL database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations.

- NoSQL database where each database has collections which in turn has documents. Each document has a different number of fields, size, content, and is stored in a JSON-like format
- NoSQL databases are increasingly used in big data and real-time web applications.

```
[{'name': 'Ashwin Sahane',
  'age': '31',
  'job': 'SAP',
  'email': 'ashwin@example.com'},

 {'name': 'Guru Singh',
  'age': '34',
  'job': 'Crystal Report',
  'email': 'guru@example.com'}]
```

What is SQL?

- SQL (Structured Query Language) is a domain-specific language used in programming and designed for managing data held in a relational database management system.
- It is also the standard for the current big data platforms that use SQL as their key API for their relational databases.
- SQL is used for storing, retrieving, updating and reading data from the database.
- Some common relational database management systems that use SQL are: Oracle, Sybase, Microsoft SQL Server, SQLite, PostgreSQL, MySQL, etc.

How Do Data Analysts Use SQL?

- Data analysis is a process of collecting data and organizing it in a manner where one can draw a conclusion.
- In order to analyze the data, we need to extract it from the database. **Relational Database Management** is an important part of Data Analysis.
- Much of the world's data resides in databases. Many database platforms are modelled after SQL.
- As data collection has increased exponentially, so has the need for people skilled at using and interacting with data; to be able to think critically and provide insights to make better decisions and optimize their businesses.
- The skills necessary to be a good data analysts include being able to retrieve and work with data, and to do that you need to be well versed in SQL, the standard language for communicating with database systems.

Jupyter Magics with SQL

- Jupyter/IPython notebooks can be used for an interactive data analysis with SQL on a relational database.
- This fuses together the advantages of using Jupyter, a well-established platform for data analysis, with the ease of use of SQL and the performance of SQL engines.

- Jupyter magic functions allow us to interact with any other relational databases. Magic functions are pre-defined functions(“magics”) in Jupyter kernel that executes supplied commands.

BeakerX (http://beakerx.com/)

- BeakerX is a collection of kernels and extensions to the Jupyter interactive computing environment.
- BeakerX supports SQL environment with many jupyter magics.
- **Install BeakerX :**
  - Close Jupyter notebook if open**.**
  - Open anaconda prompt, and run as administrator.
  - Run following command on anaconda prompt:

```
conda install -c conda-forge ipywidgets beakerx
```

Working with Jupyter Notebook