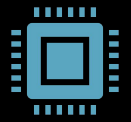
Databases & SQL for Data Science with Python

PROFESSIONAL CERTIFICATE IBM Data Analyst

In this course

Basics of SQL & Relational Databases

Working knowledge of SQL



Connect to a database and run SQL queries

Python and Jupyter notebooks

In this week

- What is SQL?
- Why SQL?
- What is data and what is a database?
- Relational Databases
- DBMS - RDBMS
- Database management systems that use SQL
- Basic SQL Commands
- Explain the advantage of the relational data model

Data Analyst Required Skills

Technical skills: spreadsheets, statistical tools, visualization tools, programming and querying languages, and the ability to work with different types of data repositories and big data platforms.

Functional skills: understanding of Statistics, Analytical techniques, problem-solving, the ability to probe a situation from multiple perspectives, data visualization, and project management skills

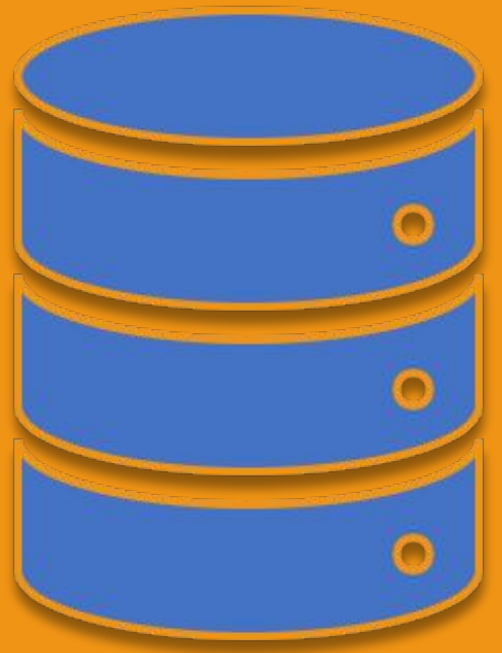
Soft Skills: include the ability to work collaboratively, communicate effectively, tell a compelling story with data, and garner support and buy-in from stakeholders. Curiosity to explore different pathways

Technical skills

- spreadsheets
- statistical tools
- visualization tools
- programming and querying languages
- Ability to work with different types of data repositories and big data platforms.

What is SQL?

- stands for Structured Query language (SQL) pronounced as "S-Q-L" or sometimes as "See-Quel“
- communicating with Relational Databases
- perform tasks such as:
    - Insert data on a database
    - Search & retrieve data from a database
    - Update data on a database



Why SQL!

According to Dataquest blog:

1. SQL is the most universal and commonly used database language.
2. It’s in high demand because so many companies use it.
3. SQL is still the most popular language for data work in 2021.

what is data and what is a database

Data is a collection of facts in the form of words, numbers, or even pictures.

A database is a repository of data, where data can be secured and accessed quickly

A database also provides the functionality for adding, modifying, and querying that data.

Relational Database

- When data is stored in tabular form.
- The data is organized in tables like in a spreadsheet, which is columns and rows.

ID	Name	Phone	Age
1	Alex	587..	21
2	Paki	825..	28

DBMS - RDBMS

A set of software tools for the data in the database is called a database management system or DBMS for short.

RDBMS refers to relational database management system

Database management systems that use SQL

What are examples of relational database management systems?

- My SQL
- Oracle Database
- DB2 Warehouse
- DB2 on Cloud

Basic SQL Commands

Create	Create a table
Insert	Insert data to populate the table
Select	Select data from the table
Update	Update data in the table
Delete	Delete data from the table

- The statement is called a **query.**
- The output we get from executing this query is called **a result set or a result table.**
- Conditions are called **predicates.**

Comparison operators supported by RDBMS

Equal to	=
Greater than	>
Lesser than	<
Greater than or equal to	>=
Less than or equal to	<=
Not equal to	<>

Create Table

```
CREATE TABLE table_name (
    column1 datatype,
    column2 datatype,
    column3 datatype,
   ....
);
```

Exercise

Write a SQL statement to create Student table with 5 columns (StudentID, LastName, FirstName, Email, City)

Exercise

Write a SQL statement to create Student table with 5 columns (StudentID, LastName, FirstName, Email, City)

```
CREATE TABLE Student (
    StudentID int,
    LastName varchar(20),
    FirstName varchar(20),
      Email varchar(50),
      City varchar(255)
);
```

StudentID	LastName	FirstName	Email	City

SELECT Statement

select COLUMN1, COLUMN2, ... from TABLE1 ;

- select * from COUNTRY ;
- select * from COUNTRY **where** ID < 5 ;

INSERT statement

- INSERT INTO *table_name* VALUES (*value1*, *value2*, *value3*, ...);
- INSERT INTO *table_name* (*column1*, *column2*, *column3*, ...)
  VALUES (*value1*, *value2*, *value3*, ...);
- INSERT INTO *table_name* (*column1*, *column2*, *column3*, ...)
  VALUES (*value1*, *value2*, *value3*, ...),
  (*value1*, *value2*, *value3*, ...);

UPDATE & DELETE statements

- UPDATE *table_name*
  SET *column1 = value1, column2 = value2,...*
  WHERE *condition;*

- DELETE FROM *table_name* WHERE *condition;*

COUNT, DISTINCT, LIMIT

- SELECT COUNT(*column_name*) FROM *table_name* WHERE *condition*;
- SELECT DISTINCT *column1, column2, ..*FROM *table_name*;
- SELECT *column_name(s)*FROM *table_name* WHERE *condition* LIMIT *number*;

Practice: True or False

The INSERT statement can be used to insert multiple rows in a single statement.

Practice: **True or False**

Distinct statement is used to retrieve the duplicate values from a table

Practice: True or False

In update statement if you do not specify the WHERE clause, only the first row in the table will be updated.

Practice: **Multiple choice**

You are working on a Film database, with a FilmLocations table. You want to retrieve a list of films that were released in 2019.You run the following query but find that all the films in the FilmLocations table are listed. What is missing?

SELECT Title, ReleaseYear, Locations FROM FilmLocations;

A. A DISTINCT clause to specify a distinct year.

B. Nothing, the query is correct.

C. A WHERE clause to limit the results to films released in 2019.

D. A LIMIT clause to limit the results to films released in 2019.

Practice: **Multiple choice**

Where clause is not used with:

A. Insert
B. Update
C. Delete
D. Limit

Breakout Room Activity

- Take 40 minutes to complete SQL basics activity posted on Schoology
- Discuss the advantage of the relational data model

Any questions