Published in Towards Data Science

Soner Yıldırım Follow

Nov 23, 2020 · 5 min read · ✨ · Listen

11 Must-Know SQL Statements for Data Scientists

Hands-on tutorial with MySQL

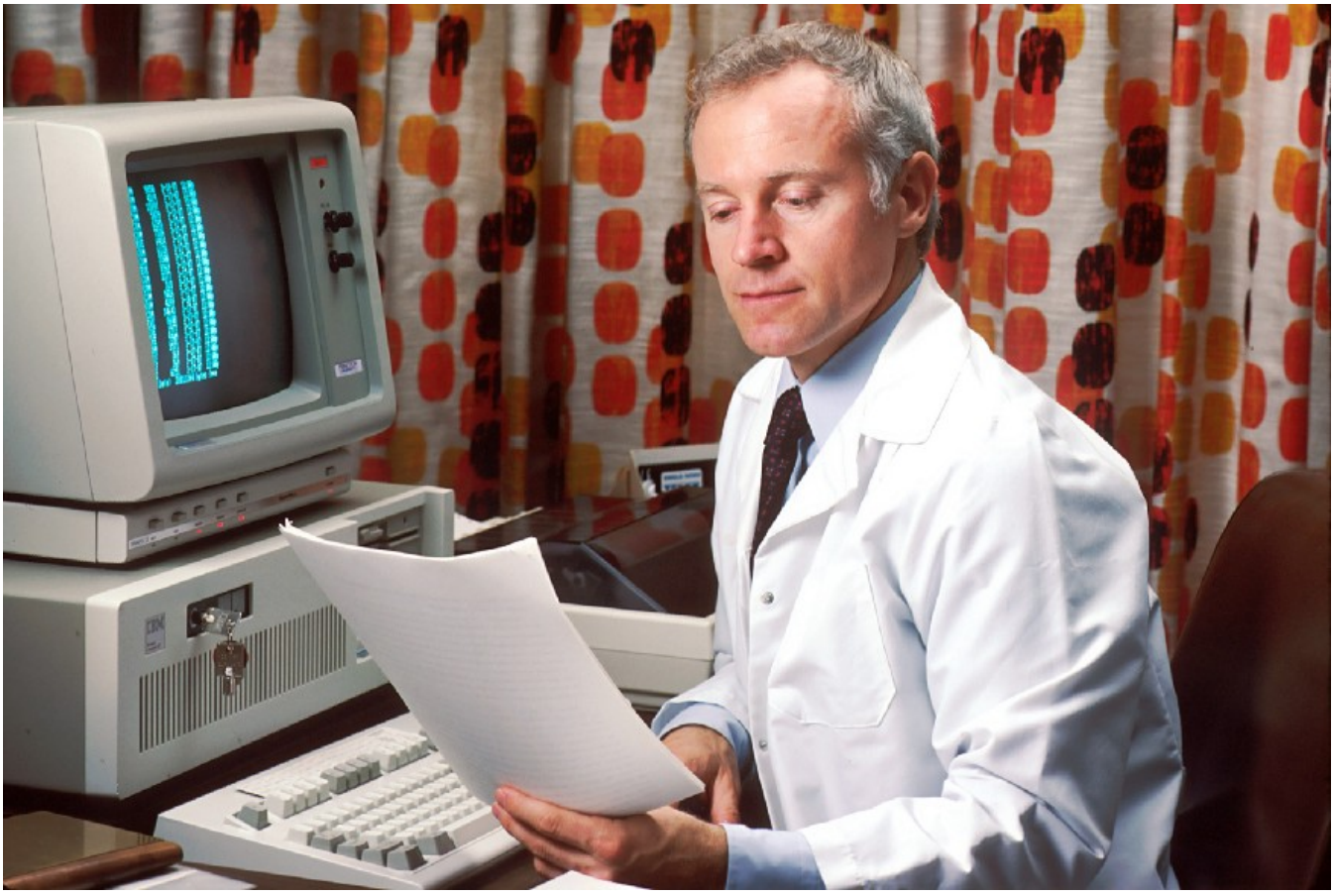

Photo by National Cancer Institute on Unsplash

SQL is a fundamental skill expected from a data scientist. You might argue that it is the job of a data engineer but the data scientist roles are inclined to be full-stack. Besides, as a data scientist, you wouldn't want to depend on a data engineer to retrieve data from a database.

In this post, we will cover the most commonly used SQL statements with many examples.

There are many different RDBMSs that adapted SQL. The SQL syntax is pretty much the same for all with some small differences. I will be using MySQL.

The best way to learn SQL is through practice. You have different options to set up an environment to practice SQL.

If you have a computer with Linux or macOS, you can easily install MySQL server from the terminal:

```
$ sudo apt install mysql-server
```

We can then connect to the server using the following command:

```
$ sudo mysql -u root
```

```
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 9
Server version: 5.7.32-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
```

1. Creating a database

A relational database consists of tables that relate to each other based on columns. Thus, in order to practice or work with SQL, we first need a database.

Let's first check the databases in the server:

```
mysql> SHOW DATABASES;
```

(image by author)

We can create a new one with CREATE DATABASE statement.

```
mysql> CREATE DATABASE medium;
```

We can delete a database with DROP DATABASE statement.

```
mysql> DROP DATABASE medium;
```

```
mysql> USE medium;
```

We are now in the medium database. We will create a student table using the CREATE TABLE statement.

```
mysql> CREATE TABLE student(
    -> Id int primary key,
    -> Name varchar(20),
    -> Major varchar(20),
    -> registered char(1)
    -> );
```

The student table consists of 4 columns. When we define a column, we specify the column name and data type. The data types are:

- Int: integer
- Decimal(M, N): Floating point number. M is total number of digits, N is the number of decimal digits.
- Varchar(N): String (text) of lenght N
- Blob: Binary large object
- Date: 'YYYY-MM-DD'
- Timestamp: 'YYYY-MM-DD HH:MM:SS'

You may have noticed that we used the words "primary key" for the "Id" column. Primary key indicates that this column uniquely identifies each row. It is similar to the index of a pandas dataframe.

Let's check if the student table was created successfully:

```
mysql> SHOW TABLES;
```


(image by author)

3. Deleting a table

It is very simple to delete a table. We use the DROP TABLE statement along with the table name:

```
mysql> DROP TABLE student;
```

4. Adding a new column

We can add a new column with the ALTER TABLE statement with ADD option.

```
mysql> ALTER TABLE student ADD gpa DECIMAL(3,2);
```

We provide the name of the column and data type.

5. Deleting a column

It is also done with the ALTER TABLE statement. We need to specify that we are dropping a column.

6. Describe

DESCRIBE statement provides an overview of the table.

```
mysql> DESCRIBE student;
```

(image by author)

7. Insert into

We have created a table but it is empty. We can populate it by adding rows. The INSERT INTO statement adds new rows to a table.

```
mysql> INSERT INTO student VALUES(1, "John", "Finance", "F");
mysql> INSERT INTO student VALUES(2, "Emily", "Math", "F");
mysql> INSERT INTO student VALUES(3, "Ashley", "Finance", "S");
mysql> INSERT INTO student VALUES(4, "Max", "Chemistry", "S");
```

We specify the table name and the values. The values must be compatible with the pre-defined data types.

The student table now contains 4 rows:

 

(image by author)

8. Delete from

We can also delete rows from a table. The DELETE FROM statement is used to delete rows based on a condition.

If we do not specify a condition, all the rows are deleted.

```
mysql> DELETE FROM student;  #will delete all rows
```

Let's delete a row based on a value in the Id column:

```
mysql> DELETE FROM student WHERE Id = 4;
```

The condition is specified after the WHERE keyword.

9. Updating a row

We can also update the existing rows in a table. The following SQL statement will update the major of the student with Id 4.

We specify the update after the SET keyword. The condition comes after the WHERE keyword.

Here is the current version of our table:

(image by author)

10. Select row or rows

Select is the most frequently used SQL statement. We run queries to retrieve data from a database. The queries are specified with the SELECT statement.

We specify the columns we want to retrieve (or select):

```
mysql> SELECT Name, Major FROM student;
```

(image by author)

A table in a database is likely to contain lots of rows. Thus, it is not a good practice to select all rows in most cases. Besides, it makes the data analysis process easier to apply some filtering and conditions before we get the data.

For instance, we may only need to work on data of the students with finance major.

```
mysql> SELECT * FROM student WHERE Major = "Finance";
```

"*" indicates all columns. The above statement will return the rows in which the major is finance.

I wrote a separate article on the select statement if you'd like to see more advanced queries.

Conclusion

We have covered the basic SQL statements. If you are or plan to be a data scientist, you are likely to use the select statement much more than the other ones.

You will probably need to run more advanced queries but it is essential to cover the basics first. Just like any other subject, practice makes perfect. Thus, I suggest to create your own database and tables. You can then practice as much as you want.

Thank you for reading. Please let me know if you have any feedback.