# [Supplemental] Store and use data with SQL

Programming is a useful skill that can help data analysts manage and interact with data to uncover helpful insights. There are many different programming languages that data analysts use to write code so that they can interact with databases. Structured query language or SQL (often pronounced “sequel”) is a programming language commonly used by data analysts. SQL allows data analysts to store, organize, and access data in relational databases.

In this reading, you’ll learn about the structure of relational databases, and how data analysts use SQL to interact with data in these databases.

## Relational database schema

Relational databases follow a consistent schema that organizes data into one or more tables. In each table, the rows represent single records, such as a customer order, while the columns represent unique attributes for the record, such as the order number, order date, and order revenue. Tables in relational databases contain primary keys, which are identifiers that reference a column in which each value is unique. For example, in the following example tables the Order_ID is the primary key of the Orders table, while the Customer_ID is the primary key for the Customers table.

### Example e-commerce database

**Orders**

| Order_ID | Customer_ID | Order_Date | Order_Revenue | Shipping_Date |
|---|---|---|---|---|
| 1 | 20432 | 12/15/23 | $120.35 | 12/16/23 |
| 2 | 25093 | 12/16/23 | $22.98 | 12/17/23 |
| 3 | 10268 | 12/16/23 | $66.75 | 12/17/23 |

**Customers**

| Customer_ID | Customer_Name | Billing_Address | Shipping_Address |
|---|---|---|---|
| 10268 | Susan | 4467 Mercer Street<br>Elmwood, Wisconsin<br>54740 | 4467 Mercer Street<br>Elmwood, Wisconsin<br>54740 |
| 20432 | Michaela | 4662 Hickory Heights Dr.<br>Belpre, Ohio<br>45714 | 4662 Hickory Heights Dr.<br>Belpre, Ohio<br>45714 |
| 25093 | Juan | 1002 Rocky Road<br>Wayne, Pennsylvania<br>19088 | 1879 Leroy Lane<br>Mitchell, South Dakota<br>57301 |

Tables in relational databases also contain foreign keys, which are columns within a table that are primary keys in another table. When a table has a foreign key, it can be connected to the other table that shares the same column. For example, the Customer_ID column is the foreign key in the order table because it's the primary key of the customer table.

## Interact with relational databases using SQL

In order to interact with relational databases, data analysts write queries that work with how the data is structured. A query is a statement written in a programming language that accesses or manipulates data in a database. Data analysts that understand the schema of relational databases, as well as the basic structure and commands of a SQL query, can efficiently interact with data in relational databases. There are four basic commands that data analysts use when working with data using SQL:

1. Creating a table or database
2. Selecting columns
3. Targeting relevant information
4. Sorting

### Creating a table or database

Data analysts use the creating a table or database command when they need to make space for new data. In order to create a table, a data analyst writes a query that includes the **CREATE TABLE** table_name (column 1 data type, column 2 data type) statement. Create table allows data analysts to identify the title of the table, and then in parentheses specify the titles of each column along with the data type. The following query is an example of how a data analyst would build a table called Products that includes columns for the product's ID number, name, color, and inventory.

Unset

```
CREATE TABLE database.Products (
      Product_ID STRING,
      Product_Name STRING,
      Color STRING,
      Inventory INT64);
```

## Selecting columns

Data analysts can select columns to focus on specific parts of the data to help answer a specific question, or investigate a particular trend when retrieving information from the database. For example, if a data analyst is interested in understanding the relationship between sales and marketing, they could select the columns for sales and marketing data to analyze the relationship between them.

Data analysts select columns by writing queries that include **SELECT** column1, column2, ... **FROM** table_name statements. The select command identifies what columns a data analyst wants to be pulled from the database, and the from command identifies which table the information should be pulled from. To select all the columns in a table, a data analyst would replace the column titles after **SELECT** with an asterisk (*). The following query is an example of how a data analyst could select the order date and order revenue columns from the orders table.

**SELECT** Order_Date, Order_Revenue **FROM** Orders;

## Targeting relevant information

Targeting relevant information involves retrieving specific information that fits designated criteria from one or more tables within a database. A data analyst can target relevant information following the same query structure as selecting a column, but with an added **WHERE** *condition* statement. The **SELECT** statement retrieves the information from specified columns in one or more tables, while the added **WHERE** statement identifies conditions that the selected data must meet in order to be returned as a result. The following query is an example of how a data analyst could retrieve a specific customer's name and shipping address from the example ecommerce database when the Customer_ID is known.

```
SELECT Customer_Name, Shipping_Address FROM Customers WHERE Customer_ID=25093
```

### Sorting

Data analysts use the sorting command to order data within a table based on the information in a specific column. In order to sort data, a data analyst writes a query that selects a column in a specific table, then they add an **ORDER BY** column1, column2, ... ASC|DESC statement. The order by statement identifies whether the selected column should be in ascending (ASC) or descending (DESC) order. The following query is an example of how a data analyst could sort the orders table so that the data was in order from the most recent order.

```
SELECT Order_Number FROM Orders ORDER BY DESC
```

## Key takeaways

Data analysts frequently encounter data in relational databases, which structures the data into one or more tables organized by columns and rows. Relational databases often have large amounts of data, so data analysts use SQL to interact with the data more efficiently by writing queries. When getting started with data analysis, you can use common commands to create a table or database, select columns, retrieve information, and sort data. You can become a more effective data analyst and improve your data management and analysis skills by familiarizing yourself with how to write SQL queries.

## Resource for more information

Review the following resource to explore more SQL commands and statements that you can use to interact with data:

- W3 schools provides definitions and tutorials for writing a variety of SQL queries.