

What is SQL?

SQL (Structured Query Language) is a domain-specific programming language used to manage and manipulate relational databases. SQL is essential for retrieving, modifying, and managing data in a relational database. Data engineers often use SQL to create, modify, and query database tables, views, and indexes. SQL allows data engineers to filter, sort, group, and aggregate data to perform complex calculations and generate reports. SQL is also used to write stored procedures, triggers, and functions to automate repetitive database tasks and ensure data integrity. SQL plays a critical role in data engineering as it enables data engineers to efficiently work with data in relational databases. Data engineers can use SQL to transform raw data into a structured format that is more suitable for analysis and reporting. Additionally, SQL is the standard language used for communication between data engineering teams and other stakeholders, such as data scientists, analysts, and business users.

What is DDL?

DDL (Data Definition Language) commands are SQL statements used to define and modify the structure of database objects such as tables, views, indexes, and sequences. These commands are used to create, alter, or drop database objects, and they do not modify the data contained within the objects.

Some common DDL commands include:

- **CREATE:** This command is used to create a new database object, such as a table, view, or index.
- **ALTER:** This command is used to modify the structure of an existing database object, such as adding a new column to a table.
- **DROP:** This command is used to delete a database object, such as a table, view, or index.
- **TRUNCATE:** This command is used to remove all data from a table, while leaving the table structure intact.

DDL commands are essential for managing the structure of databases and ensuring data integrity. They allow data engineers to create and modify database objects according to business requirements, and they provide a way to enforce data constraints and relationships. However, it's important to use DDL commands carefully, as they can have significant consequences if used incorrectly. For example, dropping a table using the DROP command will permanently delete all data contained within the table.

What is DML?

DML (Data Manipulation Language) commands are SQL statements used to insert, update, delete, and query data in a database. These commands are used to modify the data contained within the database objects created with DDL commands.

Some common DML commands include:

- **SELECT:** This command is used to retrieve data from a database table.
- **INSERT:** This command is used to insert new data into a table.
- **UPDATE:** This command is used to modify existing data in a table.
- **DELETE:** This command is used to delete data from a table.

DML commands are essential for managing the data contained within a database. They allow data engineers to add, modify, and remove data from tables as needed. Additionally, they provide a way to filter, sort, and aggregate data to generate reports and perform complex calculations.

What is DQL?

DQL (Data Query Language) is a subset of SQL (Structured Query Language) used to retrieve data from a database. DQL commands are used to perform read-only operations on database objects, such as tables and views, without modifying their structure or contents. The primary DQL command is **SELECT**, which is used to retrieve data from one or more database tables or views. The **SELECT** command is used in conjunction with other SQL clauses such as **WHERE**, **ORDER BY**, and **GROUP BY** to filter, sort, and aggregate the data retrieved from the database.

Other DQL commands include:

- **UNION:** This command is used to combine the results of two or more **SELECT** statements into a single result set.
- **INTERSECT:** This command is used to return only the rows that are common between two or more **SELECT** statements.
- **EXCEPT:** This command is used to return only the rows that are unique to one **SELECT** statement and not present in another.

DQL commands are essential for retrieving data from a database for reporting and analysis purposes. They allow data engineers to retrieve and manipulate data in a variety of ways, and they provide a way to filter and aggregate data to generate meaningful insights.