

Data Engineering 101

SQL Q&As



Shwetank Singh
GritSetGrow - GSGLearn.com



What is SQL?

SQL (Structured Query Language) is a standard language for storing, retrieving, and manipulating data in relational databases.

SQL is used for managing data in RDBMS (like MySQL, SQL Server, PostgreSQL). It has commands for data manipulation, definition, control, and querying.



Shwetank Singh
GritSetGrow - GSGLearn.com



What are the different types of SQL commands?

SQL commands are divided into:
DDL (Data Definition Language),
DML (Data Manipulation Language),
DCL (Data Control Language), and
TCL (Transaction Control Language).

DDL: Create, Drop, Alter.

DML: Select, Insert, Update, Delete.

DCL: Grant, Revoke.

TCL: Commit, Rollback, Savepoint.



What is a primary key?

A primary key is a column or set of columns that uniquely identifies each row in a table.

It must contain unique values and cannot contain NULL values. There can be only one primary key in a table, which helps maintain data integrity.



What is a foreign key?

A foreign key is a column that creates a relationship between two tables.

It is a field (or collection of fields) in one table that refers to the primary key in another table, establishing a link between them to enforce referential integrity.



What is a JOIN?

A JOIN clause is used to combine rows from two or more tables, based on a related column between them.

There are different types of JOINS: INNER JOIN returns records that have matching values,

LEFT JOIN returns all records from the left table,

RIGHT JOIN, and
FULL JOIN.



What is the difference between INNER JOIN and LEFT JOIN?

INNER JOIN returns records that have matching values in both tables, while LEFT JOIN returns all records from the left table and the matched records from the right table.

If there is no match, INNER JOIN will exclude the record, whereas LEFT JOIN includes unmatched rows from the left table, filling missing values from the right with NULL.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is normalization?

Normalization is the process of organizing data to reduce redundancy and improve data integrity.

Normalization involves dividing a database into two or more tables and defining relationships between them. The different normal forms (1NF, 2NF, 3NF) guide this process.



Shwetank Singh
GritSetGrow - GSGLearn.com



What are indexes?

An index is a database object that speeds up the retrieval of rows by creating pointers to the data in a table.

Indexes improve read performance by reducing the amount of data scanned, but they also add overhead for write operations like INSERT, UPDATE, and DELETE.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is the difference between DELETE and TRUNCATE?

DELETE removes rows one by one, allows WHERE clause filtering, and can be rolled back, whereas TRUNCATE removes all rows without logging individual row deletions.

DELETE is a DML command, while TRUNCATE is a DDL command. TRUNCATE is faster since it does not log each row, and it cannot be rolled back in most databases.



What is a subquery?

A subquery is a query nested inside another query.

Subqueries are used to retrieve data that will be used in the main query. They can be placed in SELECT, WHERE, or FROM clauses, and they return a single value or a set of values.



What is a view?

A view is a virtual table created by a query that joins and simplifies data from one or more tables.

Views can be used to encapsulate complex queries, improve security by exposing only specific columns, and simplify data access for users. They do not store data themselves.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is a stored procedure?

A stored procedure is a set of SQL statements that can be stored in the database and executed later.

Stored procedures improve performance by reducing compilation overhead, ensuring consistency, and encapsulating complex logic that can be reused across applications.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is a UNION operator?

The UNION operator is used to combine the result sets of two or more SELECT statements. Each SELECT statement within UNION must have the same number of columns.

UNION removes duplicate records by default, whereas UNION ALL retains all duplicates. The columns in each SELECT must be compatible in terms of data types and order.



What is a trigger?

A trigger is a special type of stored procedure that automatically runs when certain events occur in the database, like INSERT, UPDATE, or DELETE.

Triggers are used to enforce business rules, validate input, and maintain audit trails. They can run either BEFORE or AFTER the triggering event.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is a transaction in SQL?

A transaction is a sequence of one or more SQL operations that are treated as a single unit of work.

A transaction ensures ACID properties (Atomicity, Consistency, Isolation, Durability).

It can be explicitly committed or rolled back if any step fails to ensure data integrity.



What is the difference between WHERE and HAVING?

WHERE is used to filter rows before any groupings are made, while HAVING is used to filter groups after the GROUP BY clause.

WHERE is used before GROUP BY to filter individual rows, whereas HAVING filters the aggregated data (e.g., using functions like COUNT, SUM).



What is a self join?

A self join is a regular join where a table is joined with itself.

This is useful when dealing with hierarchical data, such as employees and their managers. Aliases are often used to distinguish between instances of the same table.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is a CTE (Common Table Expression)?

A CTE is a temporary result set that you can reference within a SELECT, INSERT, UPDATE, or DELETE statement.

CTEs are helpful for improving readability and reusing queries, and can also be recursive, allowing complex hierarchical queries. It starts with WITH keyword.



Shwetank Singh
GritSetGrow - GSGLearn.com



How can you improve the performance of a SQL query?

Use indexes, avoid using `SELECT *`, use `EXISTS` instead of `IN`, minimize use of `JOINS`, and optimize indexing.

Reducing complexity, retrieving only needed columns, avoiding subqueries, and using efficient data structures can significantly improve query performance.



Shwetank Singh
GritSetGrow - GSGLearn.com



What is the difference between RANK() and DENSE_RANK()?

Both are window functions used for ranking rows. RANK() may have gaps in ranking if there are ties, whereas DENSE_RANK() does not leave gaps.

If two rows have the same rank, RANK() will skip the next rank value, while DENSE_RANK() assigns consecutive ranks, resulting in no gaps in the sequence.



THANK
YOU