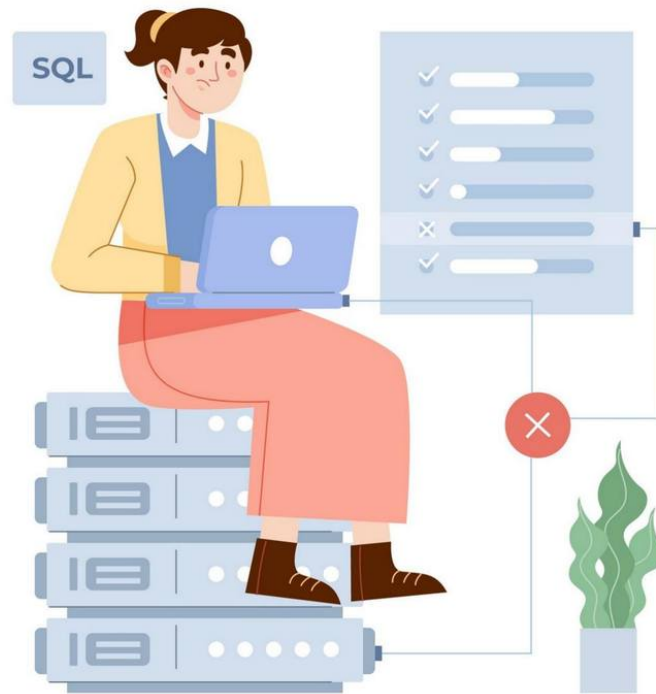


## 25 Advanced SQL Interview Questions



### 1. Explain Normalization?

Normalization is a method of organization of the data in the database to minimize redundancy from a set of relations. The concept of Normal forms is used to perform normalization on a relation.

1NF:- If a relationship has an atomic value it is in 1NF.

2NF:- For a relation to be in 2NF, it should be in 1NF and all the non-key attributes are fully functionally dependent on the primary key.

3NF:- For a relation to be in 3NF, it should be in 2NF and additionally no transition dependency exists.

BCNF:- Boy Codd's Normal form is a strong representation of 3NF.

## **2. Explain the Difference Between DDL and DML?**

DDL stands for Data Definition language whereas, DML is called Data Manipulation Language.

DDL creates a schema with some constraints. DML adds, retrieves or updates the data in that schema.

DDL defines the column attribute of the table. DML provides the row attribute of the table.

Examples of DDL commands are created, alter, drop, rename, etc. and those of DML are merged, update, insert, etc.

The WHERE clause is not used in DDL but is used in DML.

## **3. What are ACID Properties in SQL?**

The ACID properties are as follows:

Atomicity - Changes in the data must be like a single operation.

Consistency - The data must be consistent before and after the transaction.

Isolation - Multiple transactions can be done without any hindrance.

Durability - Transaction gets successful in case of system failures.

## **4. Write a Query to Find all the Duplicates in a Table?**

In terms of a general scenario, the approaches that can be followed to find all duplicates in a table are as follows:

Using the GROUP BY command to group all the rows by target columns. Here target columns are the ones where you can check the duplicates.

Use the HAVING command and integrate the COUNT function to check if any group has more than one entry.

## **5. Difference Between a Clustered and non-clustered Index?**

Below are the differences between a clustered and non-clustered index:

A clustered index is used to define the order to sort the index or table just like a dictionary. A non-clustered index collects all the data in one place and records it at a different place.

The clustered index is faster than the non-clustered index. Less memory is used to perform operations on a clustered index as compared to a non-clustered index.

## **6. Explain Denormalisation?**

The database optimization technique in which we add data(redundant) to the table(s) is called denormalization. The technique is useful as it helps to reduce the costly joins in the database.

## **7. What is Collation?**

In a SQL server, collation provides sorting rules, accent, and case sensitivity properties to the data in the database. It represents each character in the database by defining the bit patterns using collation.

Some of the collation levels are as below:

- Case-sensitive (\_CS)
- Accent-sensitive (\_AS)
- Kana-sensitive (\_KS)
- Width-sensitive (\_WS)
- Variation-selector-sensitive (\_VSS)a
- Binary (\_BIN)
- Binary-code point (\_BIN2)

## **8. What is the Difference Between inner, outer, and full outer join?**

Below are the differences between inner, outer, and full outer join:

In an inner join, the rows that are common in both tables are the outputs.

In a full outer join, all the rows in both tables are returned.

An outer join returns the values from either or both tables.

## **9. Explain the Difference Between a Join and a Union?**

Below are the differences between a join and a union:

Join retrieves the matched records from the tables whereas union is used to combine the set of two different SELECT statements.

Join does not remove duplicate data whereas union removes duplicate data(rows) from selected statements.

## **10. Explain the Difference Between UNION AND UNION ALL?**

Union keeps unique records whereas Union keeps all records including all the duplicates. Also, the union does a deduplication step before returning the data and the union all prints concatenated results

## 11. Write a Query to Print the Highest Salary from a Table?

Consider a table with the below data and table name as EMP-

Name	Salary
------	--------

-----

John	990000
Ellie	100000
Nick	400000
Sam	500000

The query would be:

```
SELECT * FROM EMP WHERE salary=(select Max(salary)
FROM EMP);
```

## 12. What is the Difference Between Shared and Inclusive Locks?

In the shared lock, the lock mode is only in read operation but in the inclusive lock, it is in reading and writes operation.

A shared lock prevents others from updating the data whereas an inclusive lock prevents others from reading or updating the data.

Any number of transactions can hold a shared lock on a particular item whereas, only one transaction can be held in an exclusive lock.

### **13. Explain the Mechanism of Transpose Using SQL?**

Transpose in SQL is defined as changing a row or column into a particular format to visualize the data from a new perspective.

One basic transpose is changing a row to a column or a column to a row. Typical transpose methods are dynamic and joined transposition.

All of the transpose methods are used for data analysis for a fruitful outcome.

### **14. Explain the Working of the B-trees Index?**

A B-tree is defined as a self-balancing search tree in which all leaves are at the same level. To insert data in a B-tree, it is inserted at the leaf node.

In the B-trees index, the tree has a key-value pair which is called a payload. This value is referenced to the actual data record.

When we use a B-tree index, the database searches a given key with correspondence to B-tree and gets the index.

### **15. Explain the Effect of Truncate and Delete on an Identity?**

Below are the differences between DELETE and TRUNCATE commands:

The DELETE command is used to delete a specific row in the table whereas, TRUNCATE deletes all the data from the rows.

DELETE is a DML command and slower than the TRUNCATE command which is a DDL command.

DELETE command active triggers whereas TRUNCATE does not.

## **16. Explain the Cost of Having a Database Index?**

It eats up your space. The larger your table, the larger will be the index.

ADD, DELETE or UPDATE of certain rows in the table must be also done to our index.

As a thumb rule, an index must be created on a table if the table in the indexed column will be required frequently.

## **17. What are the Advantages of Having an Index?**

Below are the advantages of having an index:

- It speeds up the SELECT query.

- Data retrieval becomes quite easy.

- Indexes can be used for sorting.

- It makes a row unique without any duplicates.

- Large string text can be inserted when the index is set to fill-text index.

## **18. Explain Indexing in the Database?**

Indexing is a data structure technique to optimize the database performance by reducing disk access during the time a query is being processed.

Indexing has the following attributes:

Access Types: Based on the search type  
Access Time: Based on search time  
Insertion Time: Based on quick insertion  
Deletion Time: Based on quick deletion  
Space Overhead: Based on additional overhead space.

## **19. What is OLAP and OLTP?**

### **Online Analytical Processing (OLAP)**

Businesses have to comprehend a set of data that helps finance leaders and business teams to get insights into their users and the performance of multiple related factors.

A business can have various data across the business process and strategies that help teams to analyze the performance at a more granular level.

These data are stored in multiple database systems and thus require consistent, real-time, and more flexible processes that help the team analyze the data and process further for more analysis.

Online Analytical Processing or OLAP contains software tools that are used for analyzing business data and help get insights into the database.

For example, OLAP can be used to build the Spotify song recommendation engine that automatically generates a playlist for users based on their song choices and historical listening data.

### **Online Transaction Processing(OLTP)**

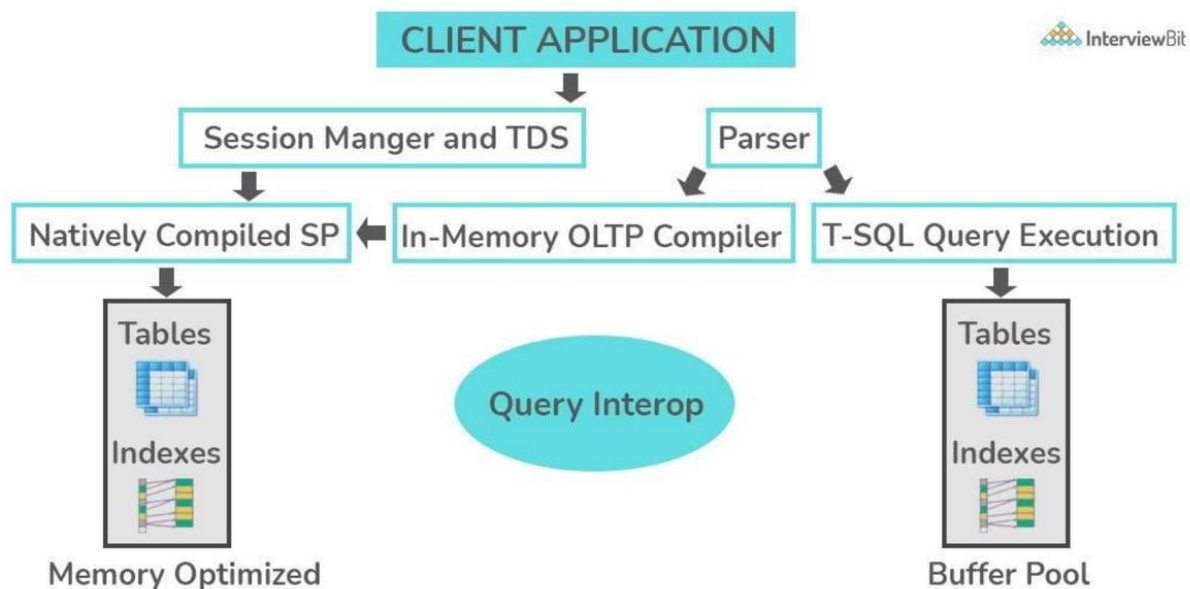
The transaction is an integral part of every business. A business offering services to its customers receive an amount when the customer uses the services.



An invoice gets immediately generated when users pay for the services of their choice.

The transaction database comes into the picture when we have to store the transaction information of all users.

However, Online Transaction Processing (OLTP) offers transaction-oriented applications that are categorized in a 3-tier architecture.



Source: [InterviewBit](#)

OLTP helps businesses to perform certain activities related to the transaction database.

For example, Transaction through online banking serves the same purpose where OLTP can be used to operate the user transaction information.

## 20. In What Conditions Does CASE WHEN Apply?

SQL is used for querying databases and modifying the values in large datasets. It also provides a way to manipulate data when there are multiple cases.

For example, a table contains a list of employees(name) along with experience (EXP) and joining date(ac\_date).

Now, if we want to query the data in such a way that if the experience is greater than(>) “4”, then insert “Is a senior” in a new column named “exp\_level”, otherwise insert NULL in the column.

You can query the database in such cases using the CASE WHEN statement.

The CASE WHEN statement is similar to the If/else condition in programming languages like C, C++, Python, etc.

The CASE statement is followed by one pair of WHEN and THEN statements. However, you can add nested statements based on your requirements.

Once the CASE statement is completed, it is finished by adding the ending statement that starts with ELSE and END AS.

In the example above, we have employee.database that lists out the employees of a company with attributes such as employee name, joining year, experience, and more. Using the CASE WHEN statement in this example,

```
SELECT employee_name,  
       exp,  
       CASE WHEN exp > 4 THEN 'Is a Senior'  
            ELSE NULL END AS exp_level  
FROM xyz.marketing_employees
```

## **21. Explain the Cases to Use HAVING versus WHERE?**

### **HAVING Clause in SQL**

SQL offers HAVING clauses to filter a group of data in a database.

For example, company XYZ has a list of employees who have more than 1 year of experience.

The database contains rows and columns including the employee name, joining data, experience, and experience level.

You can retrieve a list of employees who have more than 5 years of experience.

The HAVING clause provides you with a way to control the information and modify the data with a bunch of data.

The syntax of the HAVING clause is the HAVING condition.

In the example, we can retrieve the list of employees having more than 5 years of experience by querying the database as below:

```
SELECT employee_name, exp_level,  
FROM EMPLOYEE  
GROUP BY experience  
HAVING COUNT(experience) > 5;
```

## **WHERE Clause in SQL**

WHERE clause is used when you want to filter the records based on a particular condition.

For example, in an employee table, we can retrieve the list of all employees who have exactly 5 years of experience in a company. The syntax of the WHERE clause is, WHERE condition.

In the example, we can extract the name of employees who have 5 years of experience by querying the database as below:

```
SELECT employee_name, exp_level,  
FROM EMPLOYEE  
WHERE experience= 5;
```

## **22. What is PL/SQL?**

The PL/SQL is a block-structured language that encourages developers to use the power of SQL using procedural statements.

It is a featured procedural language that has embedded the power of decision-making and many more features of POP.

Using a single block command, PL/SQL can run multiple queries with the support of extensive error checking.

PL/SQL is an extension of SQL that is used to create applications.

## **23. What is ETL in SQL?**

ETL stands for 'Extract, Transform and Load'. It uses the concept of data warehousing to make data visualization and data analysis. In a broader context, it is used for data integration.

Any system where data is taken from one system and stored/copied in another destination system and its data is represented differently is called an ETL process.

## **24. What are Nested Triggers?**

SQL offers Data Manipulation Language (DML) and Data Definition Language(DDL) to perform certain operations while querying the database.

When a particular type of operation is performed on the database, it automatically executes some actions on the database as well.

When these types of actions are triggered in the database, they are known as Nested triggers.

SQL categorizes the nested triggers into two main categories- AFTER triggers and INSTEAD OF triggers.

As the name suggests AFTER trigger is executed after a DML or DDL operation is performed on the database.

However, INSTEAD OF trigger is executed in place of DML and DDL operations.

## **25. What are Commits and Checkpoints?**

The commit makes sure the data is consistent and maintained in the updated state after the current transaction ends. A new record is added to the log memory when a commit is used.

In the case of a checkpoint, it is used to write all the changes that are committed to disk up the system change number in the control files and header files.