

The background of the slide is light gray, decorated with various hand-drawn blue scribbles and shapes. These include loops, swirls, and wavy lines scattered around the central text.

SQL and Power BI Interview Questions

Write a query to fetch customers who made purchases in both January and February.

```
select c.customer_name  
      from customer c  
join orders o ON c.customer_id = o.customer_id  
  where o.purchase IN ('January', 'February')  
 group by c.customer_id, c.customer_name  
  having count(distinct o.purchase) = 2;
```

What's the difference between RANK(), DENSE_RANK() and ROW_NUMBER()?

ROW_NUMBER() → Gives a unique number to each row, even if values are the same. (No ties, no duplicates),

**RANK() → Gives the same rank to rows with the same value, but skips the next rank(s).
Example: if two rows are ranked 1, the next will be 3 (rank 2 is skipped).,**

**DENSE_RANK() → Similar to RANK(), but it does not skip ranks.
Example: if two rows are ranked 1, the next will be 2 (no gap).**

How do you handle NULL values in joins?

To handle NULL values in join, we should check IS NULL/IS NOT NULL, and using coalesce.

- 1. IS NULL/IS NOT NULL– To filter the rows with or without nulls.**
- 2. COALESCE– We can replace NULL with any default value.**

Write a query to find duplicate records in a table.

```
select customer_id, count(*) as duplicate_record from customer group by customer_id  
having count(*)>1;
```



Explain the use of EXISTS vs IN. Which is faster?

EXISTS Checks if a row exists in the subquery, and returns TRUE or FALSE.

IN Checks if a value is inside a list returned by the subquery, and it compares the value with that list.



How do you handle many-to-many relationships in Power BI?

To handle many-to-many relationships in Power BI, we can introduce a bridge table which will contain a column with all the unique list of values of column through which both the tables have many-to-many relation, and we connect 2 fact tables to this bridge table with one-to-many relationships.

What is the difference between calculated column and measure?

Calculated columns works on row context and Measures works on filter context. Calculated columns takes space in the data model and Measures does not takes space in the data model.

Calculated columns are physically available in the data model so we can create relationships based on columns while Measures don't, we can use Measures inside visuals.

How do you optimize a slow Power BI report?

We can follow below steps to optimize a slow Power BI report.

- 1. Keep only those columns which we need in the report and we can remove unnecessary tables.**
- 2. We aggregate the data wherever possible, that means fewer rows, reduced size and hence better performance.**
- 3. Use proper data types.**
- 4. Try to avoid many-to-many relationships when possible, and use star schema instead of snowflake schema.**

When should you use DirectQuery vs Import mode?

Data size: For larger datasets, use DirectQuery to avoid loading all the data into memory.

Real-time updates: If we need real-time updates, then also we use DirectQuery, but it has slower performance because of database speed.

Data source: Complex transformation and DAX calculations may be easier to handle in Import mode.

Write a DAX formula to calculate running total of sales.

**Running Total Sales =
CALCULATE(SUM(Sales[SalesAmount]),FILTER(ALL(Sales[Date]),Sales[Date] <=
MAX(Sales[Date])))**

The background of the slide is decorated with various hand-drawn blue scribbles and shapes. These include loops, swirls, and abstract patterns that frame the central text. The text itself is in a bold, black, sans-serif font with a white drop shadow, making it stand out against the light gray background.

Thank You !