BROADCOM

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 2+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹20-25 LPA
- **Total Rounds:** 3
- **Top Required Skills:**

1. Problem Solving
2. Advanced SQL
3. Apache Spark
4. Data Warehousing
5. Hive & Hadoop
6. Behavioral Skills

Ankita Gulati

Shubh Goyal

Round 1
Preliminary Technical Round

Problem Solving (Python / Coding)

1. Solve an array problem using the sliding window technique.
2. Explain the time and space complexity of your solution.

SQL

1. Explain the difference between RANK() and DENSE_RANK().
2. Write a query using window functions to get the running totals for sales.
3. Complex joins: If two tables each contain one column with duplicate values, how many rows would be returned after:

INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN

Apache Spark:

1. Difference between ORDER BY and SORT BY in Spark SQL.
2. What are partitions in Hive/Spark? How do they impact performance?
3. Compare RDD vs DataFrame vs Dataset in Spark.
4. Explain the Spark architecture — Driver, Executors, and Cluster Manager.
5. Hadoop & Hive:
6. Explain fundamental concepts of HDFS and Hive.
7. How does Hive store data under the hood?

Ankita Gulati

Shubh Goyal

Round 2
Strength Interview

ETL Design:

- Given a scenario where raw clickstream data needs to be cleaned, enriched, and loaded into a data warehouse, design an ETL pipeline.
- How would you handle late-arriving data and schema evolution?

SQL Challenge:

- Write a query using LEAD() and LAG() to compare each employee's salary with the previous and next employee's salary (ordered by hire date).
- Explain practical use cases for LEAD and LAG.

Hive ACID Properties:

- What are ACID properties in Hive?
- Why might a project choose not to enable ACID transactions? (e.g., performance overhead, complexity).

Data Storage Formats:

- Difference between storing data in row-oriented vs columnar format.
- Which format would you use for:
  - Analytical queries (OLAP)?
  - Transactional systems (OLTP)?

Spark Optimization Strategies:

- Explain how to handle data skew using techniques like salting.
- How do partitioning and bucketing improve query performance?
- When would you use caching in Spark?

Ankita Gulati

Shubh Goyal

Round 3
Discover Interview (Hiring Manager)

Project Deep Dive:

- Walk me through one of your recent projects in detail.
- What were the business requirements, and how did your solution help?
- What tech stack and tools were used?

Team Insights:

- Explain your current team's structure.
- How are data pipelines deployed and monitored?
- What tools do you use for CI/CD and orchestration?

Strengths and Weaknesses:

- What do you consider your biggest strengths as a Data Engineer?
- Share an example of an area you're actively working to improve.

Team Fit / Behavioral Scenarios:

- How do you handle conflicts within your team? Provide an example.
- If you strongly disagree with a design decision, how would you approach it with your manager?

Thank You

Best of luck with your upcoming interviews – you've got this!

Ankita Gulati

Shubh Goyal