

McKinsey
& Company

Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



Job Details

- **Position:** Senior Data Engineer
- **Experience:** 5 years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹30+ LPA
- **Total Rounds:** 4
- **Top Required Skills:**
 1. SQL
 2. PySpark / Python
 3. Cloud Data Engineering
 4. ETL / Data Modeling
 5. Big Data & Streaming
 6. System Design

Round 1

Online Assessment

1. **Python** – List Manipulation

- Given a list of integers, find the sum of all even numbers.

2. **Regex** – String Validation

- Write a regex to validate email addresses.

3. **SQL** – Ranking Query

- Write a query to find the second-highest salary in a department.

4. **SQL** – Aggregations & Joins

- Write queries involving GROUP BY, JOIN, and correlated subquerie

Round 2

Technical Telephonic Interview

1. **SQL Window Functions**

- Explain the difference between ROW_NUMBER(), RANK(), and DENSE_RANK() with examples.

2. **Apache Spark Fundamentals**

- How would you handle large-scale data processing using Spark?
- Explain RDDs vs DataFrames.

3. **NoSQL (Cassandra)**

- What are the advantages of Cassandra over relational databases?
- When would you prefer Cassandra?

4. **Project Discussion**

- Walk me through one of your past projects involving data pipelines and Spark.

Round 3

Technical Discussion

1. **SQL** (Ad-hoc Queries):

- Given a user activity table, write a query to find the most active users.

2. **DSA Questions:**

- Implement an algorithm to detect a cycle in a singly linked list.
- Given a sorted array, write a function to construct a balanced binary search tree (BST).

3. **Optimization:**

- How do you optimize slow SQL queries?
- Example techniques: Indexes, partitioning, avoiding cross joins, query rewrites.

Round 4

Behavioral Interview

1. **Leadership & Teamwork:**

- Describe a time when you had to lead a team through a challenging project.

2. **Problem-Solving in Ambiguity:**

- How do you approach problem-solving when faced with incomplete or unclear requirements?

3. **Client Case Study:**

- If a client faces data quality issues in production, how would you identify and resolve them?

Thank You

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

