Snap ANALYTICS

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Senior Data Engineer
- **Experience:** 5-7 years
- **Location:** Bangalore
- **Work mode:** Remote
- **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

Ankita Gulati

Shubh Goyal

Round 1
Online Assessment

Python / DSA

- Given a list of integers, write a function to return the sum of all even numbers.
- Write a regex to validate email addresses.

SQL

- Write a query to find the second-highest salary in a department.
- Write aggregation queries using GROUP BY, JOIN, and subqueries.

Ankita Gulati

Shubh Goyal

Round 2
Technical Telephonic Interview

1. Walk through your past projects and responsibilities.

2. Explain the difference between ROW_NUMBER(), RANK(), and DENSE_RANK() in SQL.

3. How would you handle large-scale data processing in Apache Spark?

4. What are the advantages of using Cassandra (NoSQL) compared to relational databases?

5. Explain Spark concepts: RDDs, DataFrames, and transformations.

Ankita Gulati

Shubh Goyal

Round 3
Machine Coding

Format: Cloud-based environment with datasets.

Task:

- Given F1 race data (CSV format) → perform data cleaning, transformation, and analysis using PySpark.

Skills Tested:

- PySpark DataFrame API (filtering, grouping, aggregations).
- Data wrangling in a real-world dataset.
- Writing clean and efficient PySpark code.

Round 4
Technical Discussion

1. **SQL**: From a user_activity table, find the most active users.

2. **DSA**: Implement an algorithm to detect a cycle in a singly linked list.

3. **DSA**: Given a sorted array
, construct a balanced binary search tree (BST).

4. **SQL Optimization**: How would you optimize queries on large tables (indexes, partitioning, avoiding cross-joins)?

Ankita Gulati Shubh Goyal

Round 4
Behavioral

1. Describe a time when you led a team through a challenging project.

2. How do you approach problem-solving under ambiguity?

3. **Scenario**: A client is facing data quality issues in production. How would you investigate and resolve them?

4. How do you prioritize when multiple stakeholders demand data simultaneously?

Thank You

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

Ankita Gulati

Shubh Goyal