Myntra

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹25-30 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
Technical Screening

1. What is the difference between the HAVING and WHERE clauses in SQL? Provide an example.
2. How would you use a SELF JOIN in SQL? Give a scenario.
3. Explain the use cases of WINDOW functions in SQL with examples.
4. What are the benefits of Indexing in SQL? (True/False + explanation).
5. In a given scenario, how would you decide between using a Fact table and a Dimension table?
6. Explain the implementation of Slowly Changing Dimension (SCD) Type 4 with an example.
7. Write a SQL query to track historical changes in data.
8. What are different loading strategies for incoming data in ETL pipelines?

Ankita Gulati

Shubh Goyal

9. Python Coding: Write a program to calculate the total amount spent, identify the top 5 users based on their spending, and find the most purchased product.

10. Write an SQL query to find the second highest salary from an Employee table.

11. Write a Python function to detect if a string is a valid palindrome after removing spaces and special characters.

Ankita Gulati

Shubh Goyal

Round 2
Projects & SQL Scenarios

1. Walk me through your past projects. What challenges did you face, and how did you resolve them?
2. Can you share an example where you diagnosed and fixed a production data pipeline issue?
3. SQL Challenge: You are given two tables with values and NULLs. Show the results of applying LEFT JOIN, RIGHT JOIN, and INNER JOIN.
4. Write an SQL query to find users who purchased the same product more than 3 times in the last 30 days.
5. Explain how you would optimize a slow SQL query in production.
6. How do you ensure data quality and consistency across multiple pipelines?
7. How do you design a data validation framework in Python/SQL before loading data into production tables?
8. Describe how you would monitor and log ETL jobs in Airflow or other orchestration tools.

Ankita Gulati **Shubh Goyal**

Round 3
System Design & Adv. Technicals

1. Explain the fundamentals of Apache Spark – cores, executors, jobs, stages, transformations vs. actions.
2. What is the difference between REPARTITION vs. COALESCE in Spark? When would you use each?
3. Explain Delta Lake vs. Parquet file formats. When would you use one over the other?
4. What is Z-Ordering in Databricks, and how does it improve query performance?
5. Discuss different JOIN strategies in Spark (Broadcast Join, Shuffle Join, etc.).
6. How would you implement incremental loading using Delta file formats?
7. Write a SQL query using CTEs and conditional joins to get active customers who purchased more than 5 products but never returned an item.

Ankita Gulati Shubh Goyal

8. How would you handle data skewness in Spark? Explain techniques like salting.

9. Explain how you would design a real-time pipeline for processing Myntra's order and inventory updates using Kafka/Spark.

10. How would you build a data lake vs. data warehouse architecture for Myntra? What trade-offs would you consider?

11. How would you implement CI/CD for data pipelines in a cloud-based environment (AWS + Databricks)?

Ankita Gulati

Shubh Goyal

Round 4
HR & Behavioral

1. Walk me through your past experiences, projects, and challenges.
2. How have you leveraged Databricks services in your projects?
3. How do you manage tight deadlines while ensuring data quality?
4. Why are you looking to switch from your current company?
5. What are your long-term career goals in data engineering?

Thank You

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

Ankita Gulati

Shubh Goyal