Myntra

Data Engineering
Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹20-30 LPA
- **Total Rounds:** 5
- **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
Telephonic Screening

1. What is the difference between ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform)?

2. Explain the concept of a data warehouse and its main benefits.

3. How would you approach data cleaning and preprocessing for a large dataset? Provide examples of techniques/tools you would use.

4. What do you know about Myntra's business model and where data plays a role in driving value?

Ankita Gulati   Shubh Goyal

Round 2
Technical Interview 1

Coding

1. Write a function to calculate the most popular product categories based on sales data. Explain the logic and complexity.

SQL

2. Write a SQL query to find the top 10 customers with the highest total purchase value.

3. Explain the concept of joins in SQL and provide examples of:
   • INNER JOIN
   • LEFT JOIN
   • RIGHT JOIN
   • FULL OUTER JOIN

Ankita Gulati   Shubh Goyal

Round 3
Technical Interview II

1. Describe the steps involved in building a data pipeline, from ingestion to final data analysis.

2. What are the key components of Apache Spark, and how do they interact (driver, executors, DAG, jobs, stages, tasks)?

3. How would you use Apache Airflow to orchestrate a complex data pipeline?
• Example: handling dependencies, retries, and monitoring.

4. What are some best practices you follow to ensure scalability, reliability, and monitoring of data pipelines?

Ankita Gulati   Shubh Goyal

Round 4
Technical Interview III

1. What are the key differences between AWS S3 and Google Cloud Storage?

2. Explain the concept of Hadoop and describe its main components (HDFS, YARN, MapReduce).

3. How would you handle large-scale data processing using Apache Hadoop?
 • Include considerations like storage format, partitioning, and job optimization.

4. How would you design a cloud-based pipeline to process real-time vs batch data on a big data platform?

Ankita Gulati   Shubh Goyal

Round 5
Managerial Interview

1. Tell me about a time when you faced a challenging data engineering project. How did you overcome it, and what was the outcome?

2. How do you stay updated with the latest trends and technologies in data engineering (e.g., blogs, open-source contributions, courses)?

3. What are your career goals, and how do you see yourself contributing to Myntra's success in the next few years?

4. How do you collaborate with cross-functional teams (e.g., product managers, analysts, ML engineers) on large-scale projects?

Ankita Gulati   Shubh Goyal

Thank You

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

HIRED

Ankita Gulati   Shubh Goyal