BANK OF AMERICA

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3 years
- **Location:** Mumbai
- **Work mode:** Office
- **Compensation:** ₹20+ 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
HR Screening

1. Walk me through your resume and highlight key data engineering projects.

2. Why are you interested in working at Bank of America?

3. Describe a project where you worked with large datasets.

4. How do you ensure data quality and integrity in your pipelines?

5. How do you handle tight deadlines or conflicts within a team?

6. Discuss your experience with ETL pipelines, data warehouses, and cloud platforms.

Ankita Gulati Shubh Goyal

Round 2
Coding Test

SQL Questions:

1. Find duplicate records in a transactions table.
2. Retrieve the second highest transaction amount per account.
3. Aggregate monthly transactions per account.
4. Use window functions (ROW_NUMBER, RANK, DENSE_RANK) to rank accounts by transaction volume.
5. Join accounts, transactions, and customers tables and filter on specific conditions.

Python / Programming Questions:

6. Read a large CSV/JSON file and compute summaries using Pandas or PySpark.
7. Find the top K most frequent elements in a dataset.
8. Implement an LRU cache or a stack/queue using Python.
9. Handle missing or corrupt data in a dataset.

Ankita Gulati

Shubh Goyal

Round 3
Technical Interview

PySpark Questions:

1. Compute total transaction amount per account per month.
2. Identify the top 3 accounts by transaction amount for a given month.
3. Calculate the average transaction amount by transaction type.

SQL / Data Modeling Questions:

4. Design a star or snowflake schema for banking transactions.
5. Write complex joins and aggregations across multiple tables.
6. Explain and implement Slowly Changing Dimensions (SCD Types 1/2/3).

Ankita Gulati Shubh Goyal

Optimization & Performance:

7. How would you optimize a slow-running SQL query?
8. How do you partition large datasets in Spark or Hive?
9. When would you use caching, broadcasting, or window functions in Spark?

Ankita Gulati

Shubh Goyal

Round 4
System Design

System Design Questions:

1. Design a data pipeline to ingest transactions from multiple sources into a warehouse.

2. How would you handle real-time streaming data using Kafka or Spark Streaming?

3. How do you ensure fault tolerance, scalability, and data integrity?

4. Explain partitioning, bucketing, and storage formats for large financial datasets.

5. Scenario: You need to migrate TBs of transaction data to cloud without downtime. How would you approach it?

Round 5
Behavioral / Managerial

1. Describe a challenging data engineering problem you solved.

2. How do you handle data quality issues in production pipelines?

3. Have you worked with cross-functional teams? How did you collaborate?

4. Give an example of a time you optimized a pipeline for performance or cost.

5. How do you stay updated with emerging data engineering technologies?

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal