Adobe

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 4-6 years
- **Location:** Noida
- **Work mode:** Office
- **Compensation:** ₹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
Online Coding & SQL

1. Consecutive Transactions (SQL):

• Table: Transactions(branch_id, txn_id, amount, txn_date)

• Find customers who made at least 3 consecutive transactions with the same amount.

• Hint: Use ROW_NUMBER() or other windowing logic.

2. 3rd Highest Transaction per Branch (SQL):

• Find the 3rd highest transaction amount per branch, considering ties.

• Hint: Use DENSE_RANK() with PARTITION BY branch_id.

3. Array Update (DSA):

• Given two arrays:
arr1 = {5,2,3,6,7}, arr2 = {9,2,8,4,5}

• Modify arr1 such that if an element exists in arr2, replace it with the corresponding element from arr2.

• Expected Output: {9,8,5,6,7}

Ankita Gulati

Shubh Goyal

4. Path Normalization (DSA – Strings):

- Normalize a given path string.
- Example:
- Input: "\a\b\c\..\..\file.txt"
- Output: "\a\file.txt"

Ankita Gulati

Shubh Goyal

Round 2
Advanced SQL & Data Modeling

1. Word Search in Matrix (SQL/Logic):

• Given a character matrix, write SQL/logic to count how many times the word “HELLO” appears in all 8 possible directions.

2. Schema Design (Adobe Analytics Events):

• Design a schema for Adobe Analytics to store user events such as page visits, clicks, and purchases across platforms (web, mobile).

• Expected points:

• Star vs. Snowflake schema.

• Partitioning & clustering for event logs.

• Indexing frequently queried fields.

3. SQL Query Optimization:

• A query on a 10M+ record table is taking 15 minutes.

• How would you optimize it?

• Expected answers: Indexing, partition pruning, avoiding cross joins, using EXPLAIN PLAN, materialized views.

Ankita Gulati Shubh Goyal

Round 3
Data Engineering & Big Data

1. Real-Time Clickstream Pipeline Design:
• How would you design a real-time clickstream ingestion pipeline for Adobe Experience Cloud?
• Expected: Kafka (ingestion) → Spark Structured Streaming (processing) → Data Lake (S3/ADLS) → Warehouse (Snowflake/BigQuery).

2. Schema Evolution:
• A downstream consumer expects a JSON field, but it is suddenly missing due to schema change.
• How would you handle this?
• Expected: Schema Registry, backward compatibility, default values, alerts.

3. Spark Optimization & Debugging:
• Real-world Spark optimization techniques you've applied.
• Example: Broadcast joins, repartitioning before writes, caching.

Ankita Gulati Shubh Goyal

• How would you debug OutOfMemoryError in Spark?

4. Bloom Filters in Data Pipelines:
• Explain how a Bloom filter works.
• Where would you use it in a data pipeline?

Ankita Gulati

Shubh Goyal

Round 4
Techno-Managerial

1. Project Architecture Deep Dive:

• Draw and explain the architecture of a pipeline you built (e.g., Spark + Snowflake + Airflow + AWS).

• Walkthrough: ingestion → transformation → storage → querying.

• Highlight performance optimizations.

2. Scenario-Based Questions:

• A Spark job that usually takes 10 mins now takes 1 hour. How do you troubleshoot?

• A dashboard shows inconsistent numbers — how do you debug if the issue lies in the pipeline vs. reporting logic?

• Your manager asks you to cut cloud costs by 30%. What strategies do you apply?

3. Behavioral Questions:

• Tell me about a time you handled production failure under tight deadlines.

• How do you prioritize when multiple teams demand data simultaneously?

Ankita Gulati

Shubh Goyal

Round 5
HR Round

Discussion Topics:

- Career aspirations and long-term goals.
- Fit within Adobe’s data platform teams.
- Salary expectations and negotiation.
- Work culture, flexibility, and alignment with Adobe values.

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal