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:** 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 Coding & SQL

SQL Questions

1. Photoshop Revenue Analysis

• Write a query to return all customers who purchased Photoshop.

• For each, calculate the total amount spent on all other products (excluding Photoshop).

• Sort the result by customer_id.

2. Average Subscription Duration

• Table: subscriptions(user_id, product, start_date, end_date).

• Write a query to calculate the average subscription duration (in months) per product.

Concepts Tested: DATEDIFF, AVG(), GROUP BY.

DSA Question

3. Merge Intervals

• Given a list of subscription start and end dates, merge overlapping intervals and return consolidated ranges.

Round 2
Advanced SQL & Data Modeling

SQL Questions

1. Rolling Usage + Reviews

• Given usage_logs(user_id, product, date) and reviews(user_id, product, rating):

• Find users who used any product more than 4 times in a single month AND gave at least one 4-star+ review in the same month.

Concepts Tested: COUNT(), GROUP BY, date truncation, JOINs, filters.

2. Top N Products per Category

• Return the top 3 most used products per country from usage logs.

Concepts Tested: Window functions (ROW_NUMBER(), RANK() with PARTITION BY).

Ankita Gulati

Shubh Goyal

Data Modeling

3. Creative Cloud Subscription System Schema

• Design a schema for:

• Users

• Subscriptions

• Billing

• Payments

• Product features

• Follow-ups:

• Handling Slowly Changing Dimensions (SCDs) for subscription history.

• Partitioning strategy for large logs table.

Ankita Gulati

Shubh Goyal

Round 3
System Design + Data Concepts

System Design Scenario

1. Adobe Analytics Real-Time Usage Tracking Pipeline

• Requirement: Ingest clickstream logs → power real-time dashboards (5-second refresh).

• Must handle retries & failure recovery.

• Tools expected: Kafka / Kinesis (ingestion), Spark Streaming / Flink (processing), Data Lake (S3/ADLS), Warehouse (Snowflake/BigQuery).

• Follow-ups: Partitioning scheme, late-arriving data, deduplication, retries with DLQ (Dead Letter Queue).

Conceptual Questions

2. Explain Clustered vs. Non-Clustered Index and when to use each.

3. Explain Horizontal vs. Vertical Partitioning and how partition pruning works.

4. Compare file formats: CSV vs. Parquet vs. ORC. Which would you choose for query performance and why?

Round 4
Hiring Manager Round

Behavioral & Leadership

1. Tell me about a time you reduced complexity in a data pipeline — what was the impact?
2. How do you balance speed of data delivery vs. accuracy?
3. Describe a data quality issue you faced in production and how you solved it.
4. How do you collaborate with analysts & product teams?
5. Share an example of when you applied Adobe's values (creativity, simplicity, collaboration) in your work.

Technical Discussion

6. Deep dive into past projects: ETL pipelines, optimizations, bottlenecks solved.
7. What trade-offs did you face in cloud cost vs. performance optimization?

Round 5
HR Round

1. Discussion on salary package and components.
2. Explanation of your role in the project and Walmart's expectations.
3. Answering your queries related to policies, work culture, or career growth.
4. Why should Walmart hire you over other candidates?

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal