Meta

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3 years
- **Location:** Bengaluru
- **Work mode:** Hybrid
- **Compensation:** ₹30–40 LPA
- **Total Rounds:** 6
- **Top Required Skills:**
  - Advanced SQL
  - Python Programming
  - Data Modeling
  - ETL & Data Pipelines
  - Big Data Technologies
  - Business Metrics & Product Sense
  - System Design for Data
  - Behavioral & Ownership

Ankita Gulati

Shubh Goyal

Round 1
Referral + Recruiter Screen

SQL Basics (One-liners, Easy)

1. What is the difference between INNER JOIN and LEFT JOIN? Provide an example with Users and Transactions tables.
2. Given a Users table, how do you count how many unique countries users are from?
3. Write the syntax for using the LAG() function to find a user's previous transaction amount.
4. When would you use HAVING instead of WHERE? Provide an example query.
5. What's the difference between COUNT(*), COUNT(1), and COUNT(column_name)?

Python Basics (Easy)

1. What's the difference between a list and a tuple? When would you use each?
2. Strings are immutable in Python – what does this mean? Give an example.
3. Write Python code to remove duplicates from [1,2,2,3,4,4,5].
4. What's the difference between append() and extend()? Show examples with code.
5. How do you iterate over both index and value in a list?

Round 2
Technical Screening Round

SQL Section (Medium)

Tables:

- Users(UserID, UserName, JoinDate)
- Transactions(TransactionID, UserID, Amount, TransactionDate)

1. Find the top 3 spenders in the last 30 days. Return UserName and TotalSpent.
2. Find the average transaction amount for users who joined in 2023.
3. Identify users who made more than one transaction in a month. Return UserID and count.
4. Using ROW_NUMBER(), return the latest transaction per user.

Python Section (Medium)

1. Two Sum: Given an array nums = [2,7,11,15], target = 9, return indices [0,1].
2. First Unique Character: Input = "loveleetcode", Output = 2 (first unique character is v).
3. Flatten Nested List: Input = [1,[2,3],[4,[5]]], Output = [1,2,3,4,5].
4. Word Count: Input = "the day is sunny the the", Output = { "the":3, "day":1, "is":1, "sunny":1 }.

Ankita Gulati

Shubh Goyal

Loop Rounds-Round 1
Data Modeling + SQL

Scenario: Design a schema for a ride-sharing app (Uber/Lyft style).

1. Design fact and dimension tables for Users, Drivers, Trips, Payments.
   - Fact: Trips(TripID, UserID, DriverID, CityID, StartTime, EndTime, Distance, Cost)
   - Dimensions: Users, Drivers, Cities, Payments.
2. Business Metrics:
   - Avg trip distance per city per week.
   - Cancellation % (cancelled trips / total trips).
   - Driver earnings per day.
3. SQL Queries:
   - Find the top 5 drivers by total earnings in the last 7 days.
   - For each city, calculate cancellation %.
   - Find the city with the maximum completed trips last month.

Ankita Gulati Shubh Goyal

Loop Rounds-Round 2
ETL + Python

Scenario: Track Instagram Reels engagement

1. Design a near real-time pipeline to capture user engagement events.
    - Which tools: Kafka for ingestion, Spark/Flink for processing, Hive/Presto for querying.
    - How to partition data (by date, country, content_id).
2. When would you use batch vs streaming?
    - Batch for daily aggregates.
    - Streaming for dashboards (e.g., live likes counter).
3. How do you handle late events (e.g., a like arrives 5 mins late)?
    - Use watermarking & window functions in Spark.

Python Coding:

1. Steps to Reduce a Number to Zero
    - Input: n = 14
    - Output: 6 (14→7→6→3→2→1→0).
2. Check Anagrams
    - Input: "listen", "silent" → Output: True.
    - Input: "hello", "world" → Output: False.
3. Spark Optimization:
    - What causes skew? How do you fix it?
    - Example: One key (e.g., India) has 80% of data → use salting or repartition.

Ankita Gulati Shubh Goyal

Loop Rounds-Round 3
Advanced SQL + Product Metrics

Scenario: Meta launches “Story Reactions” → measure user engagement.

1. What metrics would you track?
    - DAU/WAU/MAU.
    - % of users using reactions vs not.
    - Impact on time spent.
2. SQL Challenges:
    - Daily Active Users (DAU): Count unique users per day.
    - 7-Day Rolling Average of DAU.
    - Find users who reacted but never commented in the same week.
    - Nth highest salary without RANK():
    - Data Quality: How would you detect missing/duplicate data in a dashboard?

Ankita Gulati

Shubh Goyal

Loop Rounds-Round 4
Behavioral / Ownership

1. Why do you want to work at Meta?
2. Tell me about a time you worked with cross-functional teams and had conflicting requirements. How did you resolve it?
3. Describe a project that failed. What did you learn?
4. How do you prioritize when multiple stakeholders demand urgent data?
5. Have you ever disagreed with a manager's decision? How did you handle it?

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal