Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 5+ years
- **Location:** Gurgaon/ Pune/ Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹45+ LPA
- **Total Rounds:** 7
- **Top Required Skills:**

1. Advanced SQL
2. Apache Spark
3. System Design
4. Real-time Pipelines
5. Problem Solving & Coding
6. Behavioral & Leadership (Googliness)

Ankita Gulati

Shubh Goyal

Round 1
Coding (Algorithmic Problem Solving)

1. Minimum Key Presses
    a. You have a machine with 4 buttons:
    b. Button 1 → +60 minutes
    c. Button 2 → +15 minutes
    d. Button 3 → +5 minutes
    e. Button 4 → +1 minute

Input: Current timestamp displayed on the machine in HH:MM format, and a target timestamp.

Task: Write a program to calculate the minimum number of key presses required to reach the target time from the current time.

Follow-ups:

- How would your code handle crossing midnight (e.g., 23:50 → 00:10)?
- Can you optimize it using a greedy approach vs dynamic programming?

Ankita Gulati

Shubh Goyal

Round 2
Coding (String & Combinatorial Logic)

1. Next Closest Time Using Same Digits

- Input: A timestamp in format HH:MM where 0 <= H <= 9, 0 <= M <= 9.
- Task: Find the next closest valid time (in 24-hour format) that can be formed using the same digits from the given timestamp.
- Example:
  - Input → 19:34
  - Output → 19:39
- Follow-ups:
  - What's your approach: brute-force all 1440 possible minutes or use permutations + filtering?
  - How do you ensure your solution is efficient for real-time systems?

Ankita Gulati

Shubh Goyal

Round 3
Technical (Spark + SQL + Coding)

Apache Spark Questions

1. Explain Spark fundamentals: lazy evaluation, RDD vs DataFrame vs Dataset.
2. What happens during a Spark job DAG execution?
3. How do you handle OOM (Out Of Memory) errors in Spark?
4. Techniques: increasing partitions, using broadcast joins, caching strategies, checkpointing.
5. Explain optimized joins in Spark: broadcast join vs sort-merge join vs shuffle join. When would you use each?
6. How do you tune Spark applications for performance (executor memory, partition size, shuffle tuning)?

SQL Dataset Problem

1. Given an Employee table with columns (emp_id, emp_name, manager_id, salary), solve:
2. Find employees who earn more than their manager.
3. Find the hierarchy chain (employee → manager → senior manager).
4. Find the department with the second-highest average salary.

Coding Question

1. Given a string, write code to find the frequency of each character.
2. Output format: character → frequency (sorted by character or frequency based on interviewer's requirement).

Ankita Gulati

Shubh Goyal

Round 4
Application Design & Domain Knowledge

Design Task – Uber-like Application & Data Warehouse

1. Design an Uber-style ride-hailing system.
   - Entities: Drivers, Riders, Trips, Payments, Locations.
   - Key workflows: rider request, driver matching, trip completion, payment.
2. Extend to Data Warehouse design:
   - Fact table: Trip details (rider_id, driver_id, start_time, end_time, fare).
   - Dimensions: Date, Location, Rider, Driver, Payment Method.
3. Follow-up Questions:
   - How would you ensure real-time updates (driver availability)?
   - How would you handle geo-spatial queries like nearest driver search?
   - How would you partition warehouse data for fast analytics?

Ankita Gulati Shubh Goyal

Round 5
System Integration Knowledge

Design Task – IoT Real-time Data Ingestion Pipeline

- Scenario: IoT devices send data every second.
- Task: Design a real-time ingestion pipeline to process this data and store it in Google BigQuery.
- Expected Discussion Points:
  - Source: IoT devices → Pub/Sub.
  - Processing Layer: Dataflow (Apache Beam) or Spark Streaming.
  - Storage: BigQuery tables partitioned by date/hour.
  - Challenges: Data ordering, late-arriving data, cost optimization.

Round 6
Googliness
(Behavioral & Leadership Principles)

Behavioral Questions

1. Tell me about a time you took ownership of a project under tight deadlines.
2. Share an example where you had to deal with conflicting opinions within your team.

Situational Questions

1. Startup Scenario: If you are promoted to CTO after the CTO leaves, what organizational and technical changes would you bring for employees and the business?
2. Peer Review Scenario: Your teammate designs an architecture but is very rigid and rejects your feedback. How would you handle the situation and escalate to the manager if needed?

Google Leadership Principles

- Questions around:
  - Acting with humility.
  - Prioritizing user-first mindset.
  - Thinking 10x (scalability).

Ankita Gulati Shubh Goyal

Round 7
Hiring Manager (Experience & Fitment)

1. Past project experience: Walkthrough of most challenging projects.
2. How I handled scaling issues in pipelines.
3. Experience in designing real-time data systems.
4. My approach to working with cross-functional teams.
5. Why Google and how I see myself contributing as a Data Application Engineer.

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal