Deutsche Telekom

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Gurgaon
- **Work mode:** Hybrid
- **Compensation:** ₹14-20 LPA
- **Total Rounds:** 3
- **Top Required Skills:**

1. Advanced SQL
2. Apache Spark & Hive
3. Python/Scala/Java coding
4. Data Modeling & Pipeline Design
5. Communication, Stakeholder Management, Team Collaboration

Ankita Gulati

Shubh Goyal

Round 1
SQL & Coding

1. Running Sum with Window Function
    a. Given a table sales(id, branch_id, txn_date, amount), write a query to calculate the running total of sales per branch, ordered by txn_date.
    b. Follow-up: How would you handle ties on txn_date?
2. Top-N per Group
    a. From the same sales table, find the top 3 transactions by amount per branch.
    b. Expected: Use ROW_NUMBER() or DENSE_RANK(). Explain the difference and edge cases with ties.
3. De-duplication
    a. Given a user logins table with duplicates, write a query to keep only the latest login per user.
    b. Follow-up: Compare ROW_NUMBER() vs MAX(date) + GROUP BY. Which is more efficient at scale?
4. Complex Join Query
    a. Two tables: employees(id, name, dept_id, salary) and departments(dept_id, dept_name).
    b. Write a query to list employees working in departments where the average salary is higher than the company average.
    c. Expected: Use GROUP BY dept_id with HAVING and a subquery for company average.

Ankita Gulati Shubh Goyal

Coding (Python / Java / Scala)

- Write a function to return indices of all subarrays in an integer array that sum to zero.
- Example:

  Input → [1, 2, -3, 4, -1, 2, -2]

  Output → (0,2), (2,4), (4,6)…
- Expected Discussion:
  - Naïve solution = O(n²) → check all subarrays.
  - Optimized solution = O(n) using prefix sum + hashmap.

Ankita Gulati

Shubh Goyal

Round 2
Big Data / Spark + System Design

1. Broadcast Joins
    a. What is a broadcast join in Spark?
    b. Follow-up: When would you use it? What happens if the broadcast dataset is too large?
2. Data Skew & Shuffle
    a. A Spark join job is running slowly due to data skew. How would you detect and fix it?
    b. Expected: Key salting, repartitioning, map-side joins, skew hint in Spark 3.x.
3. Partitioning vs Bucketing
    a. What's the difference between partitioning and bucketing in Hive?
    b. Follow-up: When would bucketing be more effective than partitioning (e.g., small but high-cardinality columns)?
4. Spark DAG Execution
    a. Explain what happens internally when a Spark job with filter → groupBy → join runs.
    b. Follow-up: Which steps cause shuffles? How do shuffles impact performance?

Ankita Gulati Shubh Goyal

System Design / Data Modeling

Scenario:

Design a data warehouse schema for telecom call records (CDRs). Each record contains:

- caller_id, callee_id, call_duration, call_timestamp.

Expected Discussion:

- Use a fact table for CDRs.
- Dimension tables: Customer, Region, Time.
- Support analytics like “top callers per day”, “avg call duration per region”.
- Partition fact table by date for efficient querying.

Round 3
HR / Managerial

1. Project Experience
    a. Walk me through a recent project where you built a data pipeline.
    b. What challenges did you face? (e.g., scaling, performance tuning, schema evolution)
    c. How did you solve them?
2. Stakeholder Management
    a. Tell me about a time when the business demanded real-time dashboards but infra costs were too high.
    b. How did you manage expectations and balance cost vs. performance?
3. Incident Handling
    a. Describe a time when a production data pipeline failed at night.
    b. What immediate steps did you take?
    c. How did you prevent recurrence?
4. Teamwork & Collaboration
    a. How do you collaborate with remote teams across countries (e.g., Germany vs. India)?
    b. Share an example of overcoming cultural/communication gaps.
5. Why Deutsche Telekom?
    a. Why do you want to join Deutsche Telekom?
    b. What are your expectations from this role and team?

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal