ATLASSIAN

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Bengaluru
- **Work mode:** Remote
- **Compensation:** ₹25–35LPA
- **Total Rounds:** 5
- **Top Required Skills:**

1. SQL (window functions, pivot, advanced queries)
2. Python (DSA, string/list manipulations, problem solving)
3. Apache Spark & SparkSQL (JSON parsing, performance optimization, AQE, caching)
4. Data Modeling (fact/dimension tables, SCD, schema evolution)
5. Delta Lake (merge operations, schema handling)
6. System Design (data pipeline, ingestion, transformations, warehousing)
7. Managerial & Behavioral Skills (project discussion, cultural alignment, teamwork)

Ankita Gulati

Shubh Goyal

Round 1
Screening

Duration: 30 minutes
Mode: Telephonic / Online Screening

Focus Areas: Big Data basics, Python fundamentals, SQL proficiency.

Questions:

1. Can you explain your overall experience with Big Data technologies? Which tools and frameworks have you used in your projects?
2. How comfortable are you with Python for handling data processing tasks? Can you share an example where you automated a data-related task using Python?
3. Write a simple SQL query to fetch the top 10 most recent transactions from a sales table. How would you optimize this query if the table had billions of rows?
4. What are your preferred debugging methods when a SQL query does not return the expected output?

Ankita Gulati

Shubh Goyal

Round 2
Technical Interview

Duration: 60 minutes
Mode: Coding + SQL + SparkSQL
Focus Areas: SQL, DSA, SparkSQL parsing, performance trade-offs.

Questions:

1. You are given a table of customer orders. Write an SQL query using window functions to find the most recent three orders per customer. Explain why you used ROW_NUMBER() or RANK() in your solution.
2. Suppose you need to pivot a dataset of monthly sales to show months as columns. How would you write this query in SQL? What performance considerations must you keep in mind when pivoting large tables?
3. Solve a medium-level LeetCode list problem such as “Merge Two Sorted Lists.” Explain the time and space complexity of your solution.
4. Solve an easy LeetCode string problem, e.g., “Check if a string is a palindrome.” Describe how your approach scales with very long strings.
5. In SparkSQL, you are given a JSON column inside a DataFrame. Write a query to parse this JSON using explode and regex_replace. Also explain the role of concat_ws when flattening arrays.

Ankita Gulati

Shubh Goyal

Round 3

Data Modeling & Advanced Concepts

Duration: 60 minutes
Mode: Whiteboard + SQL + Spark
Focus Areas: Spark performance, pipeline design, Delta Lake, data modeling.

Questions:

1. In Apache Spark, explain the difference between caching and persisting. When would you use each? What happens if you cache a DataFrame and then perform further transformations on it?
2. Suppose a Spark job is suffering from skewed joins. How would you identify skewed keys, and what techniques (like salting) would you apply to fix the issue?
3. Explain Adaptive Query Execution (AQE) in Spark. How does it improve query performance?
4. Design a data pipeline that ingests raw clickstream logs, transforms them for analytics, stores them in a data warehouse, and makes them query-ready for dashboards. Which technologies would you choose for each step, and why?
5. What is a Delta table? Explain how you would implement an upsert operation using the MERGE command in Delta Lake.

Ankita Gulati Shubh Goyal

6. What is schema evolution in Delta Lake? Give an example where schema evolution is useful, and another where it could create risks.
7. In dimensional modeling, how would you design a fact table and its related dimension tables for an e-commerce sales platform? Explain the types of facts (transactional, periodic snapshots, accumulating snapshots) and dimensions (slowly changing vs fixed).
8. You are asked to build a Slowly Changing Dimension (SCD) Type 2 table to maintain employee salary history. Write SQL queries to update the table when a salary changes.

Ankita Gulati

Shubh Goyal

Round 4
Techno-Managerial

Duration: 60 minutes
Mode: Discussion with Manager

Focus Areas: Project depth, decision-making, technical leadership.

Questions:

1. Tell me about one of your most challenging data engineering projects. What was the objective, what technologies did you use, and what was your role?
2. How did you handle situations when a pipeline you built started failing in production? Walk through the steps you took to debug, communicate with stakeholders, and resolve the issue.
3. Describe a time you had to balance performance optimization with cost efficiency. What trade-offs did you make, and what was the outcome?
4. How do you mentor or collaborate with junior engineers in a project setting?

Round 5
Values & Culture

Duration: 45 minutes
Mode: Behavioral Interview

Focus Areas: Team fit, alignment with Atlassian values.

Questions:

1. Tell me about a time when you had to work with a difficult teammate. How did you handle the situation, and what did you learn from it?
2. Atlassian emphasizes openness and collaboration. Can you share an example where your transparent communication helped the team succeed?
3. Describe a time when you failed at a project or task. How did you handle the failure, and what changes did you implement afterward?
4. Why do you want to work at Atlassian, and how do you think you align with our culture of teamwork and learning from failure?

Thank You

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

Ankita Gulati

Shubh Goyal