AUTODESK

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Bengaluru
- **Work mode:** Hybrid
- **Compensation:** ₹15–25LPA
- **Total Rounds:** 4
- **Top Required Skills:**

1. Big Data Technologies (Spark, Hadoop)
2. Data Modeling
3. Cloud Computing (AWS, Snowflake, Lambda)
4. SQL (optimization, advanced queries)
5. DevOps (CI/CD, GitHub, Jenkins)
6. Problem Solving & Algorithms
7. Behavioral Skills (team collaboration, project contributions)

Ankita Gulati

Shubh Goyal

Round 1
Hiring Manager

Duration: 45 minutes

Mode: Video Call

Focus Areas:

- Big Data processing (Spark)
- Data modeling techniques
- Problem-solving with data structures

Questions:

1. Can you walk me through one of your most complex data engineering projects? What tools did you use to handle big data, and how did you leverage Spark for processing?
2. Explain how you would design a data model for a large-scale e-commerce platform to track product sales. What types of facts and dimensions would you include in your schema?
3. How would you optimize a large Spark job to process millions of records? What performance tuning techniques (e.g., partitioning, caching, shuffle operations) would you apply?
4. Can you discuss your approach to problem-solving with data structures? How would you handle a situation where you need to merge overlapping intervals in a list?

Round 2
Deep Dive into Tech

Duration: 60 minutes

Mode: Video Call

Focus Areas:

- Advanced SQL queries and optimization
- Data modeling strategies
- AWS Lambda and Snowflake for cloud data warehousing

Questions:

1. Write an SQL query to find the top 5 customers by total sales for each region in the last quarter. How would you optimize this query for large datasets using window functions or indexing?
2. Explain the data modeling strategies you would apply when building a data warehouse for an analytics platform. How would you ensure that the schema supports efficient querying and scalability?
3. Discuss how AWS Lambda functions could be used to trigger ETL jobs in a serverless data pipeline. How would you manage the scaling of Lambda functions based on incoming data volume?
4. How would you approach designing a cloud-based data warehouse using Snowflake? Discuss the benefits and trade-offs of Snowflake's architecture for handling large-scale analytical workloads.

Round 3
Technical Round

Duration: 60 minutes

Mode: Coding + System Design

Focus Areas:

- Coding assessments (LeetCode problems)
- Data modeling and DevOps practices
- Real-time pipeline data quality maintenance

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

Round 4
Managerial Round

Duration: 60 minutes

Mode: Video Call with Multiple Managers

Focus Areas:

- Behavioral interview
- Past project reflections
- Alignment with Autodesk's values and culture

Questions:

1. Tell us about a time when you failed or encountered a major challenge in a project. What happened, how did you handle it, and what did you learn from the experience?
2. Describe a cross-functional project where you worked with different teams (e.g., product, DevOps, data science). How did you ensure collaboration and successful project delivery?
3. Can you provide an example of how you have demonstrated ownership of a challenging project, particularly when the project was ambiguous or lacked clear direction?
4. Autodesk values collaboration and innovation. Can you share an example where you drove innovation within your team or company? How did your actions align with the company's goals and values?

Thank You

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

Ankita Gulati

Shubh Goyal