accenture

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Lead Data Engineer
- **Experience:** 7+ years
- **Location:** Pan India
- **Work mode:** Hybrid
- **Compensation:** ₹30–35 LPA
- **Total Rounds:** 3
- **Top Required Skills:**

1. SQL
2. PySpark / Python
3. Cloud Data Engineering
4. ETL / Data Modeling
5. Big Data & Streaming
6. System Design

Ankita Gulati

Shubh Goyal

Round 1
Technical Discussion

1. Walk me through your most recent data engineering project. What was the end-to-end architecture and tools used?
2. What types of transformations did you implement in the silver layer of your pipeline?
3. What transformations did you apply overall in your project? Provide examples of complex business logic transformations.
4. Explain the role of Delta Live Tables in Databricks. How is it different from traditional pipelines?
5. What is Unity Catalog in Databricks, and how would you implement it in an ongoing project?
6. How would you retrieve records that are present in one version of a Delta table but not in another version?
7. Which function in Databricks allows you to get statistics about columns and table structure in a Delta table?

Ankita Gulati

Shubh Goyal

8. Difference between workflow and pipeline in Databricks.
9. Difference between RDD and Dataset in Spark.
10. Difference between SparkContext and SparkSession. Why was SparkSession introduced in Spark 2.0?
11. What are User Defined Functions (UDFs) in Spark? When should you use them, and what are the performance implications?
12. How do you handle data skew in Spark? Provide real-world examples.
13. Explain the concept of schema-on-read vs. schema-on-write and how it applies to Databricks.
14. How would you implement slowly changing dimensions (SCD) in Delta Lake?

Ankita Gulati

Shubh Goyal

Round 2
Technical Deep Dive

1. Explain how the Catalyst Optimizer and Tungsten Execution Engine improve Spark performance.
2. What is liquid clustering in Databricks? How does it compare with the Photon execution engine?
3. When would you use Delta caching in Databricks, and what are its benefits?
4. How do you debug and optimize a long-running Spark/Databricks job?
5. Compare AWS Glue vs. Databricks for ETL workloads.
6. How would you implement data lineage and governance using Unity Catalog across multiple workspaces?
7. What are some cost optimization strategies you've applied in Databricks or cloud platforms (AWS/GCP/Azure)?
8. What are the pros and cons of Lambda Architecture vs. Medallion Architecture?

Ankita Gulati

Shubh Goyal

9. Write an efficient program to process a 100GB JSON dataset and extract only required fields into Parquet. How would you optimize memory usage?

10. Given an orders table, write a query to find the 3rd highest order amount per customer using window functions.

11. Write a query to detect gaps in time-series data for each device_id.

12. As a Lead Data Engineer, how do you mentor junior engineers on Spark optimization and coding best practices?

13. How do you ensure observability, monitoring, and alerting for production pipelines?

14. Design a streaming pipeline for ingesting real-time IoT sensor data, ensuring exactly-once delivery semantics.

15. What experience do you have with Databricks Assistant or GenAI features for accelerating development?

Ankita Gulati

Shubh Goyal

Round 3
HR Discussion

1. Tell us about your career journey and key achievements as a Data Engineer.
2. Why do you want to join Accenture at this stage of your career?
3. How do you handle conflicts in a team while leading critical projects?
4. What are your short-term and long-term career goals?
5. Salary expectations and role alignment discussion.

Ankita Gulati

Shubh Goyal

Round 4
HR & Behavioral

1. What are your salary expectations?
2. What is your notice period, and are you open to negotiations?
3. Why do you want to join Accenture as a Data Engineer?

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal