virtusa

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Senior Data Engineer
- **Experience:** 4+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹20-24 LPA
- **Total Rounds:** 4
- **Top Required Skills:**

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

Ankita Gulati

Shubh Goyal

Round 1
Technical Screening

1. Write a query to fetch the second highest salary from an employee table.
2. From a product pricing table, find product names where prices are strictly increasing over months (use LAG() or self-joins).
3. Explain CTEs (WITH clause) and write a query using multiple CTEs for better readability.
4. Difference between Clustered vs Non-Clustered Indexes. When would you use each?
5. Explain partitioning vs bucketing in SQL.
6. Write a Python function to reverse a string without using built-in functions.
7. Implement a program to check if two strings are anagrams in O(n) time (no sorting allowed).
8. Find the first non-repeating character in a string using a dictionary.

Ankita Gulati

Shubh Goyal

9. What is the difference between list, tuple, and set in Python?

10. Explain generators in Python. Why are they useful in data engineering pipelines?

11. What are the differences between AWS S3 and HDFS?

12. How would you implement a basic ETL pipeline in AWS (S3 → Glue → Redshift)?

13. Explain schema-on-read vs schema-on-write. Which one does Hive use?

Ankita Gulati

Shubh Goyal

Round 2
Advanced Data Engineering

1. Difference between RDD, DataFrame, and Dataset in Spark.
2. What is lazy evaluation in Spark? What happens during an action?
3. Difference between repartition vs coalesce. When would you use each?
4. How would you handle data skewness in Spark?
5. Difference between reduceByKey vs groupByKey. Which one is more efficient?
6. Difference between internal and external tables in Hive.
7. What are partitioning and bucketing in Hive? Give use cases.
8. Explain the execution order of an SQL query.
9. How would you design a data ingestion pipeline in AWS for streaming data (Kinesis + Lambda + S3 + Glue)?

Ankita Gulati Shubh Goyal

10. What are Delta Lake tables? How do you handle schema evolution in Delta?

11. Explain partition pruning in AWS Redshift.

12. What are best practices for optimizing Glue jobs?

13. How would you implement data governance in AWS (IAM, Lake Formation, Unity Catalog equivalent)?

Ankita Gulati

Shubh Goyal

Round 3
System Design + Scenario-Based

1. Design a data pipeline to ingest clickstream data from multiple regions into AWS (near real-time).
   → Which AWS services would you use? (Kinesis, Lambda, Glue, Redshift, Athena, EMR)
   → How would you ensure fault tolerance and exactly-once processing?
2. Suppose your pipeline ingests 1TB per day into Redshift, but queries are slow. How do you optimize performance?
3. Design a solution to track changes in a dimension table (Slowly Changing Dimension Type 2) using Spark/Delta.
4. How would you maintain data consistency across multiple AWS regions?
5. If you had to design a log analytics platform in AWS, which services would you choose and why?
6. How do you decide between batch vs streaming processing for a new use case?

Ankita Gulati Shubh Goyal

7. How would you scale the architecture if daily data grew from 1TB → 10TB?

8. How do you ensure data quality checks inside pipelines?

9. Explain eventual consistency in distributed systems.

Ankita Gulati

Shubh Goyal

Round 4
HR Discussion

1. Tell me about a challenging production issue you faced. How did you resolve it?
2. Describe a time when you had conflicting priorities. How did you handle it?
3. Have you worked with cross-functional teams (data scientists, business analysts)? Share an example.
4. Why do you want to join Virtusa?
5. If you receive a better offer elsewhere, why would you still consider Virtusa?
6. Where do you see yourself in the next 3–5 years?

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal