Tiger Analytics

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 4+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹19–22 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
Core Technical & Cloud

1. Tell me about yourself and describe a recent data engineering project you have been part of.
2. What types of transformations have you performed in your projects using AWS Glue or other services?
3. How do you replace spaces in column names with underscores in source files using AWS Glue and S3?
4. How do you read data from S3 using Amazon Redshift Spectrum or Athena?
5. What are the differences between AWS S3 and AWS Redshift in terms of data storage and usage?
6. How do you connect multiple tables from different AWS databases (e.g., RDS, Redshift) using a single connection in AWS Glue?
7. What are the different types of triggers in AWS Glue or AWS Step Functions?
8. How do you deploy code from DEV to QA and PROD environments using AWS services?

9. How do you create a CI/CD pipeline for deployment in AWS using CodePipeline, CodeCommit, and CodeBuild?
10. What is Slowly Changing Dimension (SCD) Type 2, and how can you implement it in AWS using Glue or Redshift?
11. Write a SQL query to fetch the second-highest salary department-wise. Explain multiple approaches.
12. Write a Python function to merge two sorted lists into one sorted list.
13. Explain a scenario where you optimized Glue jobs or Redshift queries for better performance.
14. Suppose you have a very large dataset in S3. How would you decide whether to use Athena, Redshift Spectrum, or loading into Redshift for querying?
15. Write a SQL query to identify duplicate rows in a table and remove them while retaining one record.
16. Write Python code to count vowels in a string and their frequencies.

Round 2
Advanced Tech & System Design

1. How do you create a view in AWS Glue or Amazon Redshift?
2. Write a DDL command in Amazon Redshift to create a table with appropriate data types.
3. What AWS Glue activities have you used in your projects, and in what scenarios?
4. How would you design a pipeline to ingest, transform, and load (ETL) large datasets from S3 into Amazon Redshift using Spark?
5. How would you implement data versioning in a Spark-based pipeline, ensuring that historical data can be tracked?
6. What Spark optimizations have you applied in your projects (e.g., partitioning, caching, broadcast joins)? When should they be used?
7. Are you familiar with AWS S3 and IAM security? How do you secure access to data in S3?

Ankita Gulati Shubh Goyal

8. What are the different authentication methods available in AWS Glue for accessing S3 or RDS?
9. How many members were in your team, and what was your specific role and contribution in the project?
10. What are your key skill sets, roles, and responsibilities in your current data engineering project, particularly with Spark and AWS?
11. Write an SQL query to get employees who earn more than the average salary in their department.
12. Write Python code to implement a LRU Cache class.
13. Explain how partitioning vs bucketing works in Spark, and when would you use one over the other?
14. How would you handle schema evolution in a Spark or Glue pipeline when source schema changes frequently?
15. Explain event-driven ETL pipeline design using AWS Lambda, S3, and Glue.

Round 3
HR & Behavioral

1. Walk me through your experience and highlights from your resume.
2. What do you expect in your next role at Tiger Analytics?
3. Why do you want to join Tiger Analytics specifically?
4. What challenges have you faced in your past projects, and how did you overcome them?
5. Salary/package discussion and career progression expectations.

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal