EPAM AWS Data Engineer Interview Guide – Experienced 4+

Round 1: Technical

Introduction

The interview began with a brief introduction and a discussion about my background in data engineering. The interviewer asked me to elaborate on my recent projects.

- Tell me about yourself and your journey into data engineering.

- Describe a recent project where you used AWS services extensively. What was your role, and what challenges did you face?

The questions here focused on understanding my hands-on experience and problem-solving capabilities. I discussed my project involving data ingestion from multiple sources using AWS Glue and processing using Spark.

Technical Questions

1. How would you fetch data from an external API, and what AWS services would you use to build a scalable data pipeline?

    Follow-up: How do you secure API requests in this setup?

2. Describe the process for migrating data from an on-premises SQL database to AWS. What services and strategies would you use?

    Follow-up: How do you handle large data transfers with minimal downtime?

3. Explain how AWS Glue interacts with on-premises SQL databases to extract data efficiently.

4. In AWS Data Pipeline, how would you design a process to copy only recently modified files from one S3 bucket to another?

    Follow-up: What configurations would you set to ensure the pipeline only copies incremental data?

5. Discuss the key differences between AWS Glue, Lambda, and Data Pipeline for orchestrating data workflows.

Scenario-Based Questions

1. Partitioning and Bucketing in Spark: Explain how you would implement partitioning and bucketing for data stored in S3 to improve query performance.

    Follow-up: When would you choose partitioning over bucketing, or vice versa?

2. Near Real-Time Data Pipeline: Design a data pipeline to ingest and process data from multiple sources (e.g., S3, Kinesis) to Redshift using Spark.

    Follow-up: How would you handle data schema evolution in this pipeline?

3. AWS Glue for Orchestration: Describe how you would use AWS Glue to schedule and manage Spark jobs.

4. Incremental Data Updates: How do you implement incremental updates in a data lake using AWS services and Spark?

> Follow-up: What challenges arise with duplicate records, and how do you address them?

5. Monitoring Spark Jobs: Explain your approach to monitoring and logging Spark jobs in AWS. What tools would you use to identify performance bottlenecks?

> Follow-up: Describe how you would optimize slow-running Spark jobs in a distributed environment.

AWS Service-Specific Questions

1. Data Pipeline: Describe a scenario where AWS Data Pipeline is preferred over Glue. Why?

2. Glue vs. Lambda: Explain when you would use Glue instead of Lambda for a data ingestion use case.

Round 2: Team Fitment and Company Understanding (15 Minutes)

Company Knowledge

1. What do you know about EPAM Systems?

2. Why do you want to join EPAM?

3. What do you think differentiates EPAM from other consulting firms in the data engineering space?

Team Fitment

1. Describe your preferred work environment and collaboration style.

2. How do you handle conflicts within a team? Provide an example.

3. What strategies do you use to stay updated with new technologies?

Reason for Change

1. Why are you looking for a job change?

2. What are you seeking in your next role that your current position does not offer?

Round 3: HR Interview

Experience and Projects

1. Walk me through your resume. What are the key highlights that align with this role?

2. Discuss a project where you significantly impacted performance or cost optimization.

Role Expectations and Fit

1. What are your expectations for this role?

2. What is your preferred location, and how soon can you join?

3. How do you prioritize tasks when working on multiple projects?

Additional HR Questions

1. What are your salary expectations?

2. Where do you see yourself in five years?

Glassdoor EPAM Review –

https://www.glassdoor.co.in/Reviews/EPAM-Systems-Reviews-E15544.htm

EPAM Careers –

https://www.epam.com/careers

Subscribe to my YouTube Channel for Free Data Engineering Content –

https://www.youtube.com/@shubhamwadekar27

Connect with me here –

https://bento.me/shubhamwadekar

Checkout more Interview Preparation Material on –

https://topmate.io/shubham_wadekar