Amazon Data Engineer Interview Guide

Application & Resume Screening

Overview:

- Resume submitted via Amazon Careers portal.
- Recruiters screen profiles based on skills in SQL, Python, ETL, and relevant project experience.
- Strong emphasis on data pipeline experience, data modeling, and exposure to scalable systems.

Tips:

- Highlight ownership in past data engineering projects and any quantifiable impact.
- Use clear bullet points focused on technologies like Redshift, EMR, S3, Glue, etc.
- Leadership principle alignment (bias for action, customer obsession) should reflect in project descriptions.

Round 1: Phone / Video Screening

Focus Areas:

- SQL and Data Modeling basics.
- Python coding fundamentals.
- Behavioral traits aligned with Amazon's Leadership Principles.

Sample Questions:

- Tell me about a data pipeline you've built end-to-end.
- Write a SQL query to find the top 3 selling products per region in the last month.
- How would you identify duplicate records based on a composite key in SQL?
- Given a list of integers, write a Python function to return the number of unique pairs that sum up to a target.

Tips:

- Keep answers concise but structured; walk through the logic clearly.
- Think aloud while solving problems.
- Be prepared to speak to trade-offs and performance considerations.

Round 2: Onsite Interview (Virtual)

1. Technical Interview Rounds (2 Rounds – 45 to 60 mins each)

Focus Areas:

- Advanced SQL, data modeling, and ETL system design.
- Python for transformation logic and edge-case handling.
- Problem-solving using real-world scenarios.

Sample Questions:

- Design a data model for an e-commerce system tracking orders, shipments, and payments.
- Write a SQL query to detect customers who have not placed a second order in 90 days.
- How would you build a pipeline that transforms semi-structured logs into a structured analytics layer?
- In Python, process a large CSV in chunks and remove duplicate records based on email and timestamp.

Tips:

- Break down system design into ingestion, processing, storage, and access layers.
- Ask clarifying questions before coding; they appreciate collaborative problem-solving.
- Explain optimizations such as use of partitioning, sorting, or caching.

2. Bar Raiser Round

Focus Areas:

- Leadership Principles like "Dive Deep," "Invent and Simplify," and "Hire and Develop the Best."
- Behavioral depth and ownership in past roles.
- Problem-solving in ambiguous situations.

Sample Questions:

- Tell me about a time you had to work with incomplete or dirty data. How did you manage it?
- Describe a situation where you made a mistake in a data pipeline. How did you identify and fix it?
- How have you mentored others in your team or improved team-wide engineering practices?
- Explain a project where you had to influence stakeholders without having authority.

Tips:

- Use STAR method: Situation, Task, Action, Result.
- Don’t hide failures—focus on what was learned and how the problem was resolved.
- Demonstrate clear alignment with Amazon's principles through real examples.

3. Final HR / Hiring Manager Round

Focus Areas:

- Role fit and long-term vision alignment.
- Leadership and communication style.
- Cross-functional collaboration and ownership traits.

Sample Questions:

- How do you prioritize tasks when handling multiple stakeholders?
- What motivates you to work on data infrastructure problems?
- How do you keep up with the latest trends or tools in data engineering?
- Describe a scenario where you disagreed with a product or business team. What did you do?

Tips:

- Be authentic and show long-term intent.
- Highlight decision-making strategies and business impact.
- Expect questions around collaboration with product teams and engineering leadership.

More sample Data Engineer Interview questions

1. Describe a complex data engineering project you have worked on. What were the challenges you faced, and how did you overcome them?
2. How would you design a scalable and fault-tolerant data processing pipeline for handling large volumes of streaming data?
3. Explain the process you would follow for optimizing a database query that is running slow.
4. What strategies and technologies would you consider when designing a data warehouse architecture for efficient data storage and retrieval?
5. How would you ensure data quality and integrity in a data pipeline? Discuss the steps you would take to validate and cleanse data.
6. Share your experience in working with big data technologies such as Hadoop, Spark, or AWS EMR. How have you leveraged these tools in your previous projects?
7. Describe a scenario where you had to make trade-offs between data processing speed and accuracy. How did you approach this situation and what was the outcome?
8. How would you handle security and privacy concerns when working with sensitive data in a cloud environment?
9. Discuss your experience with ETL (Extract, Transform, Load) processes. What tools and techniques have you used to ensure efficient data extraction and transformation?
10. Describe a time when you had to collaborate with cross-functional teams to deliver a successful data engineering solution. What was your role, and how did you ensure effective communication and coordination.

Glassdoor Amazon Review –

https://www.glassdoor.co.in/Reviews/Amazon-Reviews-E6036.htm

Amazon Careers –

https://www.amazon.jobs/en/

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