Morgan Stanley Data Engineer Interview Guide – Experienced

Introduction

Morgan Stanley is a globally renowned investment bank with a strong presence across institutional securities, wealth management, and investment management. Securing a Data Engineer role here involves a rigorous five-round interview process, designed to assess both technical depth and your alignment with the firm's values.

This guide is based on real experiences, offering a detailed breakdown of each interview round, actionable tips, and insights to help you navigate the process with confidence.

Interview Process Breakdown

1. **Preliminary Round: Online Test**
   A 2-hour timed test on HackerRank, designed to evaluate your core technical knowledge and problem-solving skills across various domains.
2. **Technical Interview 1**
   A deep dive into SQL, Python, Big Data tools, and cloud technologies with a focus on scenario-based problem-solving.
3. **Technical Interview 2**
   Primarily focuses on data modeling, ETL pipelines, PySpark, and Databricks Lakehouse architecture.
4. **Techno-Managerial Round**
   Assesses your leadership qualities, project experiences, and alignment with Morgan Stanley's core principles.
5. **HR Round**
   A conversational round focusing on your personality, goals, and career aspirations.

Detailed Insights on Each Round

Round 1: Preliminary Round (Online Test)

This online test evaluates your foundational skills and practical knowledge. Conducted on HackerRank, it consists of four individually timed sections:

- **SQL Coding Questions**: Medium-level queries focusing on joins, aggregations, window functions, and subqueries.
- **Python & SQL MCQs**: Questions covering data structures, functions, and database operations.
- **Data Structures Coding Question**: A medium-level problem involving arrays, stacks, queues, or linked lists.
- **Database and Administration Scenarios**: Practical MCQs based on relational database concepts, transactions, and Unix commands.

Example Questions:

1. Write a SQL query to calculate the highest salary in each department using a window function.
2. Identify the Unix command that lists files with specific permissions.

Tips for Success:

- Review SQL concepts like joins, window functions, group by, having clauses, and subqueries.
- Focus on Python basics, including lists, dictionaries, and sets. Deep dive into Pandas for data manipulation.
- Brush up on data structures like arrays, strings, stacks, and queues.
- Practice Unix commands and understand basic shell scripting.

Round 2: Technical Interview 1

This round is a comprehensive evaluation of your expertise across various data engineering tools and concepts.

- **Key Focus Areas:**
    - SQL: Writing queries with joins, aggregations, and window functions.
    - Big Data Tools: Hive (partitioning, bucketing), Sqoop (incremental loads, commands), and HDFS.
    - Cloud Computing: AWS services like S3, EC2, IAM, and Redshift.
    - Spark: Architecture, narrow vs. wide transformations, coalesce vs. repartition, and scheduling jobs.
    - DevOps: CI/CD concepts, Git workflows, and tools like Jenkins.

Example Questions:

1. Given two tables, calculate the row count for different types of joins (inner, left, right, and full outer).
2. Write a Sqoop command to import relational data from MySQL into HDFS.
3. Explain the difference between coalescing and repartitioning in Spark.

Tips for Success:

- Be prepared to explain SQL queries, especially window functions like lead and lag.
- Understand the differences between Hive's managed and external tables.
- Practice writing Sqoop commands and discuss how to handle incremental loads.

Round 3: Technical Interview 2

This round emphasizes your ability to design efficient data models and pipelines.

- **Key Focus Areas:**
    - Data Modeling: Normalize and denormalize schemas based on use cases.
    - ETL Design: Extract, transform, and load pipelines using tools like AWS Glue.
    - PySpark: Handling batch and stream processing, Spark monitoring, and performance optimization.
    - Databricks: Lakehouse architecture and data transformation workflows.

Example Questions:

1. Design a relational data model for a sales database, incorporating normalization techniques.
2. Write pseudo code for an ETL pipeline using Python and Pandas.
3. Explain Spark's narrow vs. wide transformations and when to use each.

Tips for Success:

- Discuss real-world projects to demonstrate your ETL and data modeling expertise.
- Understand the architecture and benefits of Databricks Lakehouse.
- Prepare to write pseudo code and explain its logic clearly.

Round 4: Techno-Managerial Round

This round evaluates your leadership, communication skills, and alignment with Morgan Stanley's values.

- **Key Focus Areas:**
    - Project Experiences: Discuss the challenges and technologies used in your past work.
    - Teamwork: Explain how you handle conflicts and manage teams effectively.
    - Technical Skills: Answer scenario-based questions about Spark, AWS Glue, and Delta Lakes.

Example Questions:

1. How would you manage a disagreement within your team about an ETL pipeline design?
2. What are the advantages of Delta Tables in your project?
3. Explain the configuration of a Spark cluster for optimal performance.

Tips for Success:

- Use examples from past experiences to highlight problem-solving and leadership skills.
- Align your answers with Morgan Stanley's core principles.

Round 5: HR Round

This conversational round assesses your cultural fit and long-term goals.

Questions Asked:

1. What motivates you to join Morgan Stanley?
2. What are your strengths and weaknesses?
3. Why should we hire you?

Tips for Success:

- Be authentic and articulate your career goals clearly.
- Research Morgan Stanley's culture and values to align your answers with their expectations.

Tips for Success Across Rounds

- **Preparation:** Focus on SQL, Python, Big Data, and cloud technologies. Practice coding problems and review key concepts.
- **Communication:** Explain your thought process clearly during coding and scenario-based questions.
- **Examples:** Draw from real-world experiences to demonstrate your skills effectively.

Common Mistakes to Avoid

- Overlooking SQL window functions or Python libraries like Pandas.
- Failing to articulate the reasoning behind your approach to problems.
- Not researching Morgan Stanley's values and aligning your answers accordingly.

Glassdoor Morgan Stanley Review –

https://www.glassdoor.co.in/Reviews/Morgan-Stanley-Reviews-E2282.htm

Morgan Stanley Careers –

https://www.morganstanley.com/careers/career-opportunities-search

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