Myntra Data Engineer Interview Guide – Experienced 3+

Round 1 - Soft Skills & SQL Basics (Recruiter Assessment)

Overview:

The first round focused on foundational technical skills in SQL, Python, and data modeling, coupled with an assessment of communication skills.

Detailed Breakdown:

1. **Soft Skills Assessment:**

   The recruiter began with a quick self-introduction:

   - Focused on presenting your role, tech stack, and recent projects.
   - Highlighted responsibilities like building pipelines, optimizing data systems, and integrating tools like Spark and Databricks.

2. **SQL Basics:**

   Questions were focused on intermediate SQL concepts. Examples include:

   - **HAVING vs WHERE:**
     - Explained how WHERE filters rows before aggregation, whereas HAVING filters groups post-aggregation.
   - **SELF JOIN Applications:**
     - Discussed scenarios like finding manager-employee relationships within the same table.
   - **WINDOW Functions:**
     - Provided use cases for RANK(), DENSE_RANK() for ordering data within partitions.
   - **Indexing:**
     - A True/False question testing the understanding of indexes and their role in query optimization.

3. **Data Modeling (2 out of 5 Questions):**

   Fact vs Dimension Table:

   - Given a real-world problem, explained how to categorize data into fact tables (measurable data) and dimension tables (descriptive data).

   Slowly Changing Dimension (SCD) Type 4:

   - Described the hybrid approach of maintaining a current data table and a historical changes table.

Tracking Historical Data Changes:

- Wrote a SQL query leveraging window functions and timestamps to identify updates over time.

4. **Python Problem:**

   Problem Statement: Write a Python program to calculate total spending, identify top 5 users by spending, and find the most purchased product.

   Solution Approach:

   - Used dictionaries and sorting functions to efficiently calculate total spends.
   - Applied Counter from Python collections to identify the most frequent product.

   Followed up with explanations around list comprehensions, lambda functions, and performance optimizations.

Round 2 - Technical Discussion

Overview:

This round focused on practical experiences, challenges faced in real-world projects, and SQL troubleshooting.

Key Highlights:

1. **Project-Based Discussion:**

   Discussed your previous projects in detail:

   - Explained the architecture of a data pipeline (e.g., ingestion → processing → storage → reporting).
   - Highlighted your role in designing and optimizing Spark jobs for ETL processes.

   Challenges:

   - Talked about issues like data skewness and how you resolved them using techniques like salting and repartitioning.
   - Optimizing Spark jobs with caching, tuning executor memory, and using broadcast joins.

2. **Unsolved Written Test Questions:**

   The interviewer revisited SQL and Python questions from Round 1:

   - Provided alternative approaches for solving JOIN-based SQL queries involving null values.
   - Demonstrated Python solutions emphasizing code readability and efficiency.

3. **SQL Problems:**
    - Solved medium-level SQL questions involving two tables with null values.
    - Required outputs for LEFT JOIN, RIGHT JOIN, and INNER JOIN were discussed with a clear query structure.

Round 3 - System Design

Overview:

This round tested your understanding of Apache Spark, file formats, and advanced SQL design scenarios.

Detailed Topics Discussed:

1. **Apache Spark Fundamentals:**

    Core Concepts:

    - Defined executors, cores, stages, jobs, transformations, and actions in Spark.
    - Highlighted how Spark executes tasks in DAG (Directed Acyclic Graph) format.

    Optimization:

    - Discussed REPARTITION vs COALESCE for managing partitions.
    - Explained data skew resolution using partitioning strategies.

2. **File Format Comparisons:**

    Delta vs Parquet:

    - Explained Delta Lake's ACID transactions, schema enforcement, and time-travel features.
    - Compared it to Parquet, highlighting performance trade-offs.

    Z-Ordering:

    - Provided use cases where Z-Ordering improves query performance for partitioned Delta tables.

3. **Scenario-Based SQL:**

    An advanced SQL problem involving:

    - Common Table Expressions (CTEs): Used for query modularization.
    - Conditional Joins: Applied conditional logic for JOIN operations between two tables.

    Focused on both performance optimization and query clarity.

Round 4 - Hiring Manager Discussion

Overview:

The final round was focused on understanding your personality, motivations, and alignment with the company's goals.

Topics Covered:

1. **Project Experience:**
    - Discussed major projects, highlighting achievements and challenges.
    - Explored the usage of tools like Data bricks, Spark, and Delta Lake in your projects.
2. **Career Motivation:**

    Addressed questions like:

    - Why are you looking to switch roles?
    - What excites you about this opportunity?
    - How does this position align with your long-term career goals?
3. **Relocation Questions:**
    - The interviewer evaluated flexibility regarding relocation (e.g., moving to Bangalore while currently settled in Hyderabad).
    - Emphasized your openness to adapt based on career growth opportunities.

Glassdoor Myntra Review –

https://www.glassdoor.co.in/Reviews/Myntra-Reviews-E508705.htm

Myntra Careers –

https://careers.myntra.com/