S&P Global Senior Data Engineer Interview Guide – Experienced 3+

Round 1 – Technical

Project Explanation

In this round, I was asked to explain my current project in detail, including various aspects such as:

- **Type of Data**: What kind of data am I working with? Structured, semi-structured, or unstructured?
- **Data Flow**: How does data flow through the system? From ingestion to processing and storage?
- **Data Size**: The scale of data being handled—whether it's small, medium, or big data.
- **Project-Specific Details**: Any specific challenges in the project, technologies used, and how we achieved our goals.

Coding Questions

1. **Data Structures & Algorithms (DSA) Question (Easy):**
    - Problem: Given the input string "AAABBBCCCDDDAAA," compress it to output "A3B3C3D3A3."
    - This question tested my understanding of string manipulation and how to efficiently compress a string using a simple iteration and count technique.
2. **SQL Query**:
    - Problem: Write a query to find the 5th highest salary in an employee table and calculate the number of employees whose salary is greater than that of their manager.
    - For this, I used window functions to find the 5th highest salary and a join to calculate the number of employees earning more than their manager.
3. **PySpark Implementation**:
    - Problem: Convert the above SQL logic to PySpark code.
    - This question tested my ability to translate SQL logic to PySpark, and I used DataFrames and window functions in PySpark to achieve the same results as in the SQL query.

Round 2 – Hiring Manager

Scenario-Based Questions

The second round was focused on **scenario-based questions**, where the interviewer dug deeper into my experience with core technologies and practices in the data engineering field.

1. **Spark & Airflow**:

   Discussed how I've worked with Apache Spark for large-scale data processing and Airflow for workflow orchestration. I was asked to explain how I handle performance optimizations, scheduling tasks, and monitoring DAGs in Airflow.

2. **Redshift vs. Snowflake**:

   A deep dive into Redshift and Snowflake, including the differences between these two cloud data warehouses, and how I've used them in previous projects. I was asked to explain the scalability, performance, and cost-efficiency of both tools in different use cases.

3. **PySpark & SQL**:

   Detailed questions around PySpark and SQL, including how I use PySpark for distributed data processing and SQL for querying data efficiently. I was asked to provide specific examples of challenges faced and solutions implemented.

4. **Team Size & Project Sprints**:

   The interviewer asked about the size of the teams I've worked with and how we handled sprints during the project. This was an opportunity for me to explain my experience with Agile methodology and my role in managing project timelines and deliverables.

5. **Reason for Considering a Change**:

   I was asked why I was considering leaving my current company, and I explained that I was looking for a new challenge and the opportunity to work on more complex data systems at scale. I emphasized my desire to grow further in my data engineering career.

Glassdoor S&P Global Review –

https://www.glassdoor.co.in/Reviews/S-and-P-Global-Reviews-E1259396.htm

S&P Global Careers –

https://www.spglobal.com/en/explore-s-p-global/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