Snowflake Data Engineer Interview Guide – Experienced 2+

Screening Round

This round was conducted by the hiring manager for the Data Engineer role. It focused on my previous project experiences and scenario-based questions. Key discussion points included:

- Explanation of how to capture data lineage for Spark code, using a DataHub-based example.
- Questions on **Spark optimization** techniques.
- **Snowflake-specific topics**:
    - Change Data Capture (CDC) in Snowflake.
    - Use cases for internal staging in Snowflake.
    - Integration of Snowflake with external data sources such as S3, GCS, and Blob Storage in various cloud services.

The hiring manager mentioned that the next technical round would cover SQL, Python, Big Data concepts, databases, data warehousing, and Spark optimization.

Technical Round 1

Conducted by two technical interviewers, this round evaluated core technical skills:

- **SQL Coding**: Practical coding challenges.
- **Python Coding**: Data manipulation and problem-solving.
- **Big Data Concepts**: Understanding frameworks and architecture.
- **Spark Architecture**: Execution flow and data processing.
- **APIs**: RESTful API integration with Snowflake.

Technical Round 2

This round focused on detailed discussions around Spark and Snowflake:

- **Spark Optimization Techniques**:
    - Coalesce vs. Repartition.
    - Cache vs. Persistent storage.
    - Adaptive Query Execution.
    - Broadcast Joins and Shuffle Merge Joins.
    - Logical Plan workflow when submitting Spark queries.
- **Airflow Integration**:
    - API calling with Airflow.
    - Airflow operators, hooks, and scheduler functionality

- **ETL Pipeline Design in Snowflake**:
    - Building ETL pipelines to capture changes when new records are inserted into source tables.
    - Using Airflow to trigger and manage ETL jobs.
- **Photon Engine Backend Architecture**:
    - Designing backend architecture for SQL Warehouse, leveraging prior experience.

Techno-Managerial Round

This round assessed both technical knowledge and management skills:

- **Managing Multiple Tasks**: Approaches to handling multiple tasks within a sprint.
- **Communication with Team Leads**: Strategies for working with busy team leads.
- **Motivation for Joining Snowflake**: Personal aspirations and reasons for interest in the company.
- **ETL Pipeline Design**: High-level design using tools like Kafka or Flink for new use cases.
- **Challenges in Requirement to Solution**: Specific challenges faced in translating requirements into technical solutions.
- **Snowflake Tech Stack**: Deployment on Azure, cluster sizing considerations, and overall data warehouse design.
- **Change Data Capture (CDC) in Snowflake**: Importance and implementation.

Key Topics and Questions

- **SQL Proficiency**: Essential for any Data Engineer role, with a focus on:

    Self-joins to compare employee salaries.

    Grouping and aggregation functions.
- **Python Skills**: Important for data manipulation and ETL processes.
- **Big Data and Spark**: Understanding of Spark's architecture, execution features, and data processing capabilities.
- **Data Warehousing Concepts**: Familiarity with data warehousing principles, especially as they relate to Snowflake.
- **ETL Processes**: Questions on how to set up ETL pipelines using tools like Apache Airflow and how to handle change data capture (CDC).

Glassdoor Snowflake Review –

https://www.glassdoor.co.in/Reviews/PayPal-Reviews-E9848.htm

Snowflake Careers –

https://careers.pypl.com/home/

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