Apple Data Engineer Interview Experience

Round 1: Recruiter Screening

The process began with a recruiter screening, which served as an initial alignment discussion. The focus was primarily on past projects, resume walkthrough, and relevance of prior experience to Apple's engineering needs. Discussions often centered around large-scale data pipeline work, infrastructure optimization, and project outcomes. Clear communication of business impact, system design choices, and team collaboration typically helps during this round.

Round 2: Technical Phone Screen

This round focused on SQL proficiency and high-level data pipeline design. The session included problem-solving scenarios that involved writing optimized SQL queries, dealing with aggregations, and explaining the architecture of a data solution for a given use case.

Key areas evaluated:

- Efficient SQL query writing, including use of window functions, CTEs, and joins.
- Understanding of indexing, partitioning, and execution plans.
- Clarity in data pipeline thinking: how data is ingested, transformed, and served.

Demonstrating the ability to balance performance, maintainability, and scalability in query design is essential. Familiarity with SQL optimization patterns and anti-patterns adds value.

Round 3: Onsite (Virtual) Interviews

The onsite loop included a series of technical and behavioral interviews aimed at evaluating core skills in data engineering.

SQL and Python Coding Round

The round tested the ability to write complex SQL queries involving multiple joins, subqueries, and data aggregation logic. Python questions focused on data manipulation, edge case handling, and cleaning semi-structured or inconsistent datasets.

Common expectations:

- Handling nulls, duplicates, and inconsistent timestamp formats.
- Writing maintainable, efficient Pandas or PySpark code.
- Optimizing SQL using indexing and partitioning filters.

System Design: Data Pipeline Architecture

A system design session was conducted to assess the ability to architect data pipelines from end to end. A use case was given, and the expectation was to describe how data would be ingested, processed, stored, and queried.

Important aspects:

- Batch vs real-time processing choices and their trade-offs.
- Technologies like Kafka, Spark, S3, Snowflake, Redshift, etc.
- Strategies for monitoring, retries, idempotency, and validation.
- Emphasis on scalability, data lineage, and fault tolerance.

Big Data and Cloud Infrastructure Round

This round emphasized distributed data processing and familiarity with big data ecosystems.

Topics covered:

- Spark transformations (lazy evaluation, wide vs narrow).
- File formats (Parquet, Avro, ORC) and storage strategies.
- Kafka messaging guarantees and Snowflake schema evolution.
- Performance tuning concepts such as shuffle, skew, and caching.

Glassdoor Apple Review –

https://www.glassdoor.co.in/Reviews/Apple-Reviews-E1138.htm

Apple Careers –

https://www.apple.com/careers/in/

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