Kaseya Azure Data Engineer Interview Guide – Experienced 3+

Overview of the Process

This interview process was structured into three rounds:

1. **Technical Round (90 minutes):** Focused on Azure Data Factory (ADF), Databricks, Snowflake, and Python fundamentals, along with practical SQL concepts.
2. **Techno-Managerial Round (30-45 minutes):** Led by a director-level executive, it involved project-specific discussions and advanced concepts like SCD types in Snowflake.
3. **HR Discussion:** Covered salary negotiation, benefits, and other administrative aspects.

Round 1: Technical (90 minutes)

Key Topics Covered:

- **Azure Data Factory (ADF):**
    - Components of ADF: Pipelines, Activities, Linked Services, Datasets, Triggers, and Integration Runtimes.
    - Integration Runtimes: Explained the differences between Azure IR, Self-hosted IR, and Azure-SSIS IR.
    - Data Flows: Usage in transformations and connecting data sources.
    - Databricks Integration: Steps to link a Databricks notebook to an ADF pipeline using Linked Services and Activities.
    - Automation of Stored Procedures: Configuring Stored Procedure Activities within pipelines.
    - CI/CD in ADF: Explained deployment using Azure DevOps or ARM templates.
- **SQL Concepts:**
    - Temporary Tables vs. Views: Temporary tables are session-specific, while views are stored queries without data.
    - Error Handling in T-SQL: Usage of TRY...CATCH, THROW, and RAISEERROR.
    - Indexing: Differences between clustered (alters table storage) and non-clustered (separate structure) indexes.
    - SQL Query Optimization: Techniques like using proper indexes, avoiding SELECT *, and reducing nested subqueries.
    - CTE (Common Table Expression): Explained as a temporary result set for better query readability.

- APPLY Operator: Explained CROSS APPLY and OUTER APPLY for joining with table-valued functions.
- PIVOT Operator: Used to transform rows into columns in T-SQL.

- **Data Modeling:**

  Normalization: Explained 1NF (removal of duplicate columns), 2NF (removal of partial dependency), and 3NF (removal of transitive dependency) with examples.

- **Big Data and Azure Ecosystem:**
  - Common Services: Discussed Synapse Analytics, Azure Data Lake, and Azure Databricks.
  - Schema-on-Read: Data is interpreted at runtime instead of predefining a schema.
  - Schema Evolution: Discussed dynamic schema adjustments in big data environments.
- **Databricks and Delta Lake:**
  - Medallion Architecture: Explained bronze (raw data), silver (filtered/clean data), and gold (aggregated data) layers.
  - Delta Lake: Discussed ACID compliance, time travel, and streaming support.
  - Role-Based Access Control (RBAC): Configuring permissions for workspaces and data access.
- **Snowflake:**
  - Types of Caching: Metadata, local disk, and cloud services cache.
  - Time Travel vs. Fail-safe: Time Travel allows access to historical data; Fail-safe provides recovery for up to 7 days.
  - Snowpipe: Discussed its capability for continuous data ingestion, including API integration for real-time pipelines.
  - Materialized Views: Precomputed views for faster query execution.
- **Python Concepts:**
  - Common Libraries: pandas, numpy, requests, and pyarrow.
  - break, continue, and pass: Control flow statements.
  - Pandas Operations: Renaming columns (DataFrame.rename()) and exporting to Excel (to_excel()).
  - GIL (Global Interpreter Lock): Discussed its impact on multi-threading in Python.
  - Data Engineering: Passing API responses to a DataFrame and storing DataFrames in Azure Storage.
  - Advanced Python: Explained iterators, decorators, generators, pickling/unpickling, negative indexing, and shallow vs. deep copies.

Round 2: Techno-Managerial (30-45 minutes)

This round focused on real-world project discussions and scenario-based questions.

Key Discussion Points:

- **Azure Projects:**
  - In-depth review of portfolio projects involving ADF, Databricks, and Synapse Analytics.
  - Cross-questions on design decisions, resource optimization, and performance tuning.
- **Snowflake SCD (Slowly Changing Dimensions):**
  - Types of SCDs:
    - **SCD Type 1:** Overwriting data.
    - **SCD Type 2:** Maintaining history with versioning.
    - **SCD Type 3:** Adding new columns for changes.
- **Managerial Scenarios:**
  - Handling tight deadlines for critical pipelines.
  - Communication with cross-functional teams for smooth deployments.

Round 3: HR Discussion

- **Salary Negotiation:**

  Discussed expectations, current market trends, and benefits.
- **Variable Component:**

  Clarity on performance-based bonuses and fixed vs. variable pay.

Glassdoor Kaseya Review –

https://www.glassdoor.co.in/Reviews/Kaseya-Reviews-E262966.htm

Kaseya Careers –

https://www.kaseya.com/careers/

Checkout more Interview Preparation Material on –

https://topmate.io/shubham_wadekar