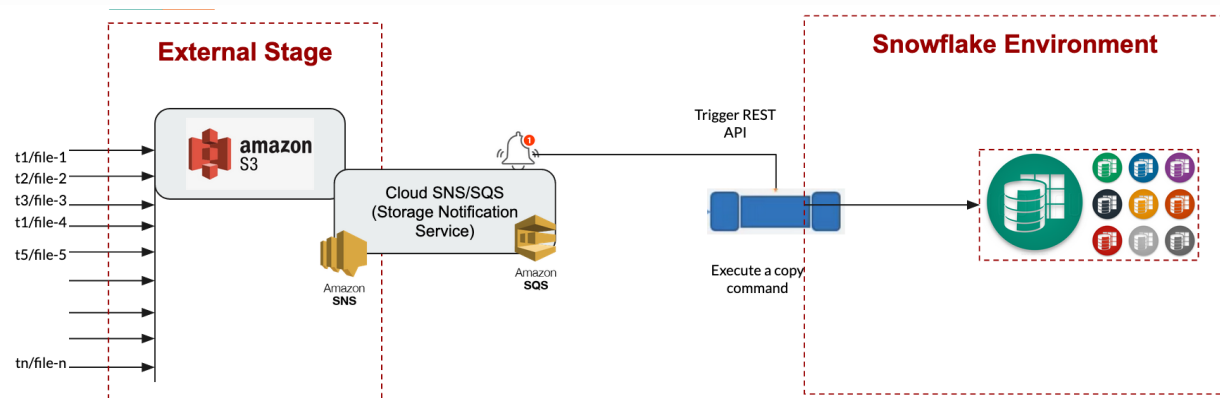


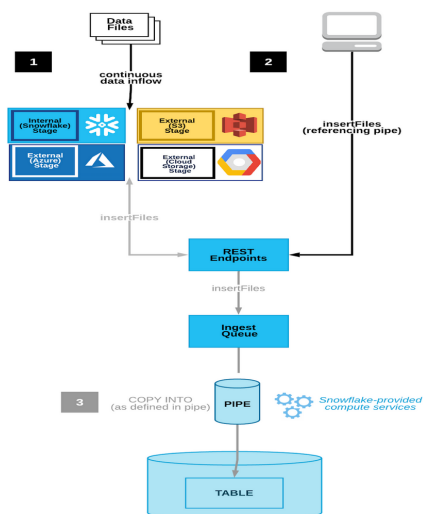
SnowPipe Introduction (Snowflake Certification)

Data Loading & Data Ingestion is one of the key activity in any data warehouse system and that's why Snowflake SnowPro Certification exam ask many questions around this concept. This subject area does not need any hands on knowledge however you will get at least 2 questions around SnowPipe and fundamentals around continuous data loading. With trial edition, we generally don't play a lot with SnowPipe as it needs cloud account to trigger batches and that weakens our knowledge around SnowPipe & continuous data loading.

Overview - Snowflake SnowPipe



Workflow - Snowflake SnowPipe



- Data files are copied to an internal (Snowflake) or external (Amazon S3, Google Cloud Storage, or Microsoft Azure) stage.
- A client calls the *insertFiles* endpoint with a list of files to ingest and a defined pipe.
- The endpoint moves these files to an ingest queue.
- A **Snowflake-provided virtual warehouse** loads data from the queued files into the target table based on parameters defined in the specified pipe.

Note:

- Option 1: Using a Client to Call the REST API
- Option 2: Using AWS Lambda to Call the REST API

Snowflake SnowPipe Cheat Sheet

Snowflake SnowPipe Overview

1. SnowPipe enables loading data from files as soon as they're available in a external stage
2. A pipe is a named, first-class Snowflake object that contains a COPY statement used by Snowpipe
3. Pipe wraps copy commands, so all data type is supported (json,avro etc)
4. File arrival detecting mechanism
 - Using cloud notification
 - Calling REST API endpoint
5. SnowPipe copies the files into a queue, from which they are loaded into the target table in a continuous, serverless fashion based on parameters defined in a specified pipe object.
6. File Size Recommendation
 - The number of load operations that run in parallel cannot exceed the number of data files to be loaded
 - data files roughly 10 MB to 100 MB in size compressed
 - Split large files into a greater number of smaller files to distribute the load among the servers
7. Semi-Structured File size
 - Variant data type imposes 16Mb compressed for individual rows.
 - we recommend enabling the STRIP_OUTER_ARRAY file format option for the COPY INTO <table> command to remove the outer array structure and load the records into separate table rows
8. Other Important Note
 - Snowpipe charges 0.06 credits per 1000 files queued
 - There is no guarantee that files are loaded in the same order they are staged
 - SnowPipe latency is hard to estimate as it depends on the file size, file format and complexity of the copy statement.

SnowPipe/Continuous Loading VS Copy Command/Bulk Loading

- Authentication
 - Bulk Loading - user session
 - Continuous Loading - JSON Web Token using public/private key
- Load History
 - Bulk Loading - 64 days of metadata history

- **Continuous Loading - 14 days of metadata history**
- **Transaction**
 - **Bulk Loading - always single transaction**
 - **Continuous Loading - combine or split based on number of files**
- **Compute Resource**
 - **Bulk Loading - WH is required**
 - **Continuous Loading - uses SF supplied resources**
- **Cost**
 - **Bulk Loading - priced everytime WH is active**
 - **Continuous Loading - bill as per SF supplied resource is used.**