

[Snowflake] 4-1. Data Movement Overview



노션 웹 공유 링크 (댓글 & 상세설명 참고)

References

- Snowflake Learn (SnowPro PREP-CORE Course)_4장 1강
- Snowflake 설명서 (데이터 로딩 개요)

-
- Snowflake 데이터 로딩 (Data Loading) 개요
 - 정형화 데이터 (CSV, TSV ..) 반정형화 데이터 (JSON, Parquet) 등 다양한 데이터 형식을 외부 또는 내부 스테이지 (파일 저장소) 로 이동시키고, Snowflake 내부로 데이터를 가져오는 일련의 과정

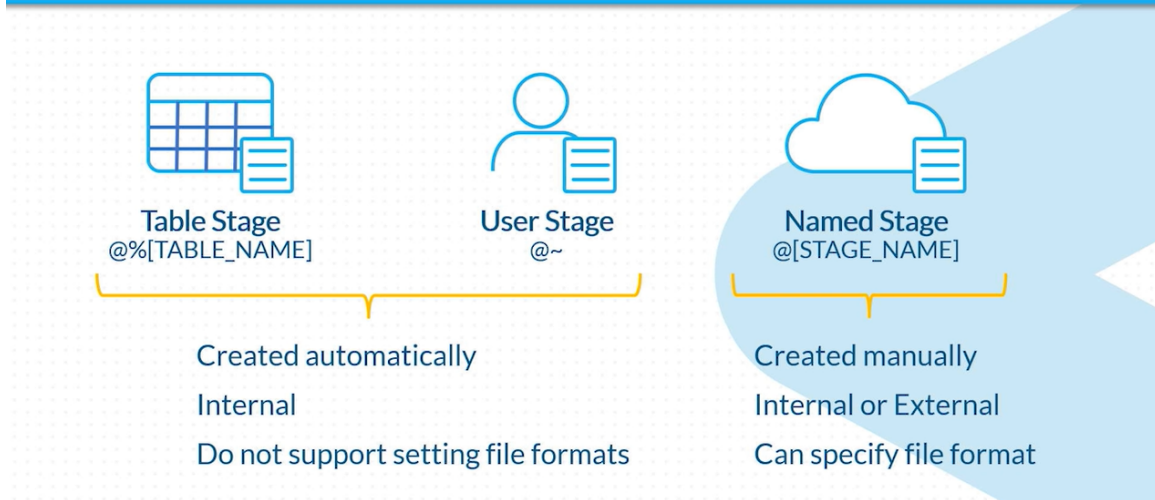
DATA LOADING STEPS



데이터 로딩 단계

- Snowflake의 데이터 로딩은 다음 단계를 거침
 - 데이터 파일을 클라우드 스토리지에 업로드
 - COPY INTO 명령을 사용하여 테이블에 데이터를 로드하기 전 클라우드 영역에 파일을 준비 (스테이징)
 - 데이터를 Snowflake 영역으로 가져옴, 마이크로 파티션을 만들고 메타데이터를 저장
- **Stage (스테이지, 지원되는 파일 위치) 개요 - Snowflake에 로드 되기 전 파일 저장 위치**
 - 외부 스테이지 : Amazon S3, Google Cloud Storage, Microsoft Azure
 - 외부 스테이지는 Snowflake에서 스키마에 생성된 일종의 데이터베이스 오브젝트
 - CREATE STAGE 명령을 사용하여 스테이지 '오브젝트'를 생성
 - STAGE 생성 시 파일 저장위치, URL 또는 URI, Stage에 접근할 수 있는 자격증명 필요

TYPES OF STAGES



- 내부 스테이지 : 사용자, 테이블, 명명된 스테이지 세가지 타입의 저장소 존재
 - 사용자와 테이블 스테이지는 시스템에 의해 자동으로 생성, 항상 존재하는 스테이지, 일부 파일 형식은 지원하지 않음
 - **사용자** : 파일 저장소를 각 사용자에게 할당
 - **테이블** : Snowflake에 생성된 테이블에 대해 테이블 스테이지를 사용, 이 스테이지 타입은 한명 이상의 사용자가 스테이징하고 관리하지만, 단일 테이블에만 로드되는 파일을 저장하도록 설계, 변경 삭제가 불가능
 - **명명된 스테이지** : 스키마에서 생성된 데이터베이스 오브젝트, 한 명 이상의 사용자가 스테이징 및 관리하고 하나 이상의 테이블에 로드되는 파일을 저장, CREATE STAGE 명령을 사용하여 스테이지 생성
- 내부 스테이지로 파일 전송은 PUT 명령을 사용하여 로컬 시스템의 파일을 내부 스테이지로 업로드 할 수 있음.

- 지원되는 파일 형식

FILE FORMAT

A named object within a SCHEMA

Instructs Snowflake how to parse files during load/unload

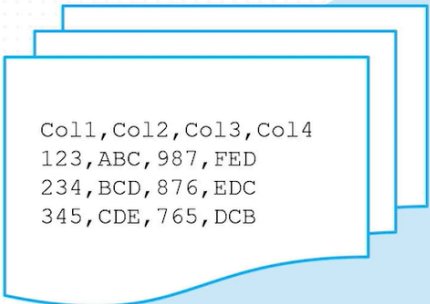
File types: CSV, JSON, Parquet, AVRO, ORC, XML

Specify the FILE FORMAT:

In the COPY INTO command

As part of a named stage or pipe definition

As part of a table definition



```
Col1,Col2,Col3,Col4
123,ABC,987,FED
234,BCD,876,EDC
345,CDE,765,DCB
```

```
CREATE FILE FORMAT MY_CSV_FORMAT
TYPE = 'CSV'
FIELD_DELIMITER = ','
RECORD_DELIMITER = '\n'
```

- Snowflake 스키마 내 명명된 객체
 - 지원 파일 타입은 CSV, JSON, Parquet 등
 - FILE_FORMAT 매개변수를 사용하여 파일 형식 지정 가능
-
- Snowpipe (연속 로드)

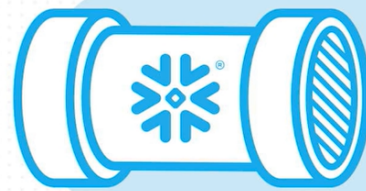
PIPE

Enables streaming/pipeline/micro-batch loading as opposed to bulk loading

A pre-defined object that specifies the COPY INTO statement syntax for a data load

Instructs Snowflake how and where to load/copy/convert data into records in the target table

All data types are supported, including semi-structured data types such as JSON and Avro



- 데이터를 지속적으로 캡처하기 위해 Snowpipe를 사용
- 소량의 데이터 (마이크로 배치) 를 로드하고 증분식으로 분석에 사용할 수 있도록 설계
- Snowpipe는 파일이 스테이지에 추가되면 수집요청으로 제출된 후 몇분 이내에 데이터를 로드, 이를 통해 데이터를 사용할 수 있게 되는 즉시 사용자에게 최신 결과가 제공