



[Snowflake] 4-3. Continuous Data Loading

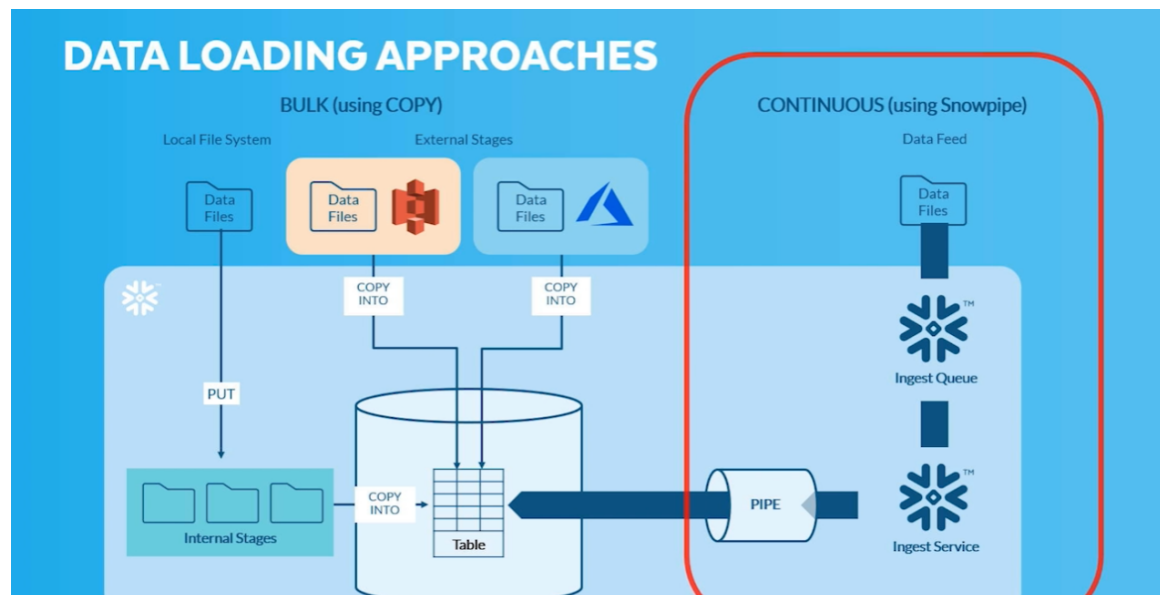


[노션 웹 공유 링크 \(댓글 & 상세설명 참고\)](#)

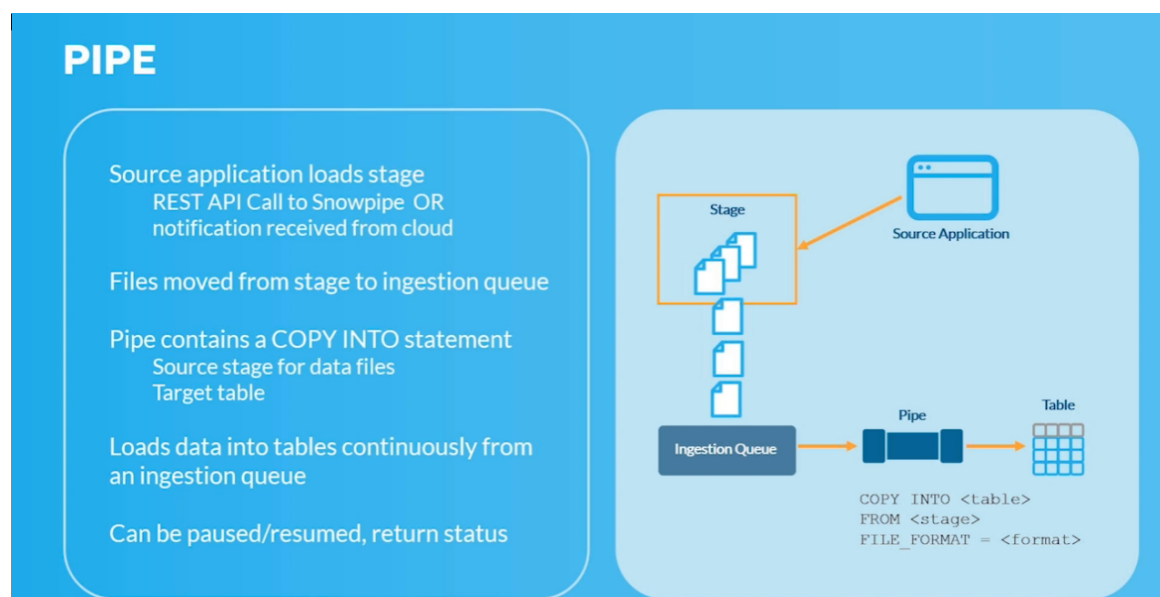
References

- [Snowflake Learn \(SnowPro PREP-CORE Course\)_4장 3강](#)
- [Snowflake 설명서 \(Snowpipe\)](#)

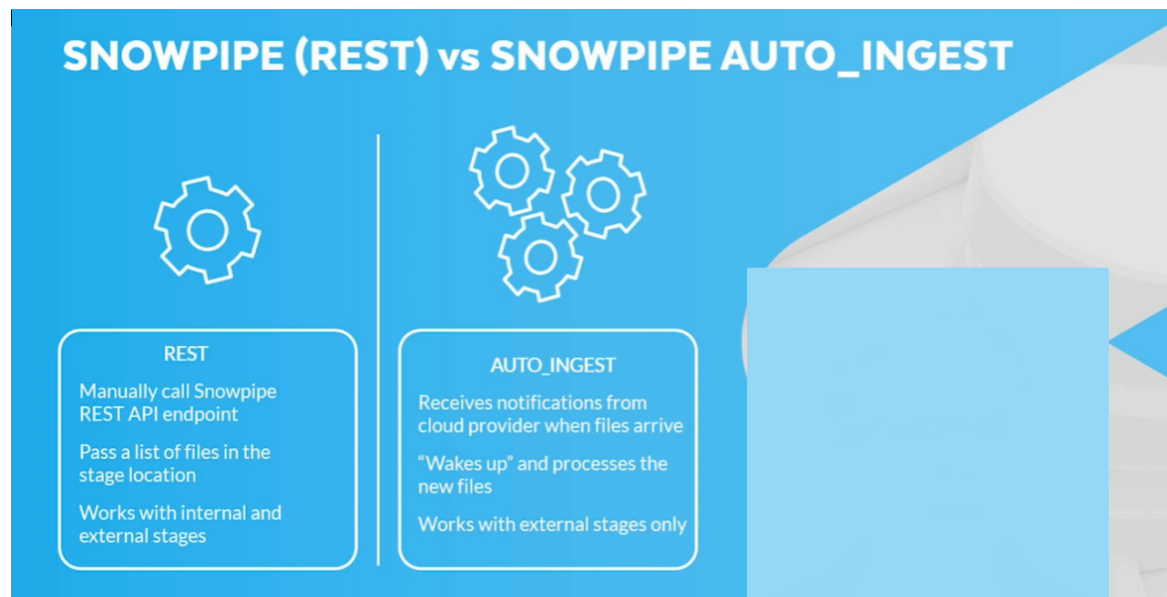
-
- 지속적 데이터 로드 (Continuous Data Loading) - Snowpipe



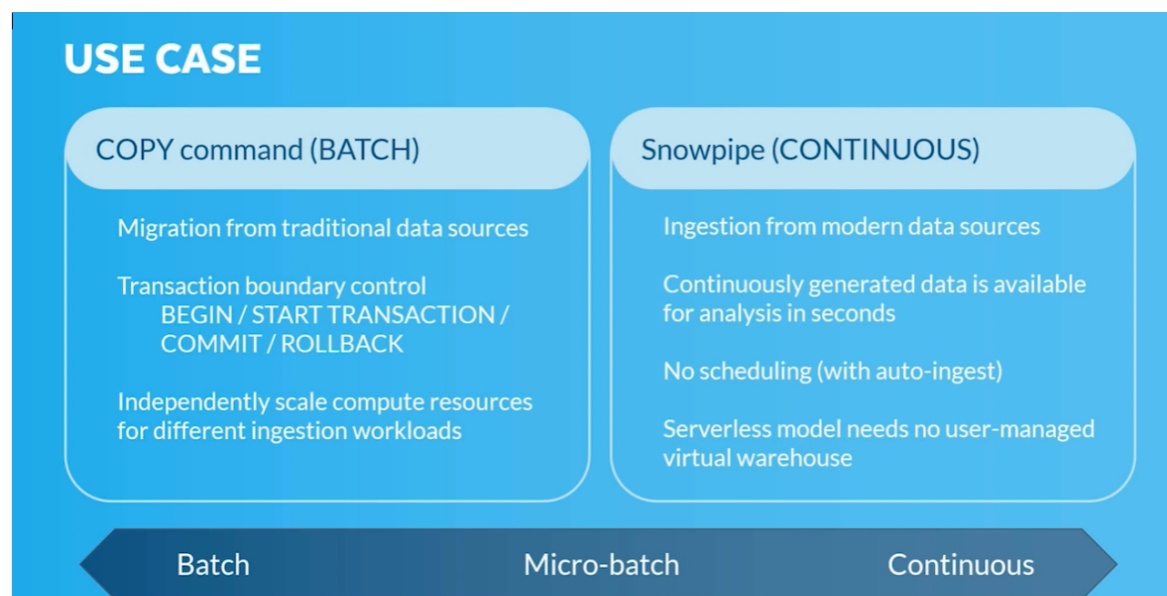
- Snowpipe 는 스테이지에서 파일을 사용할 수 있는 즉시 파일에서 데이터를 로드
- COPY 문을 수동으로 실행하는 방식이 아닌, 사용자가 몇 분 내 사용할 수 있도록 마이크로 배치로 파일에서 데이터를 로드



- Pipe 에 정의 된 COPY 문에 따라 로드
 - Pipe 는 COPY 문이 포함된 Snowflake 오브젝트
- 데이터 파일은 JSON, Avro와 같은 반정형 데이터 타입을 포함한 모든 데이터 타입 지원

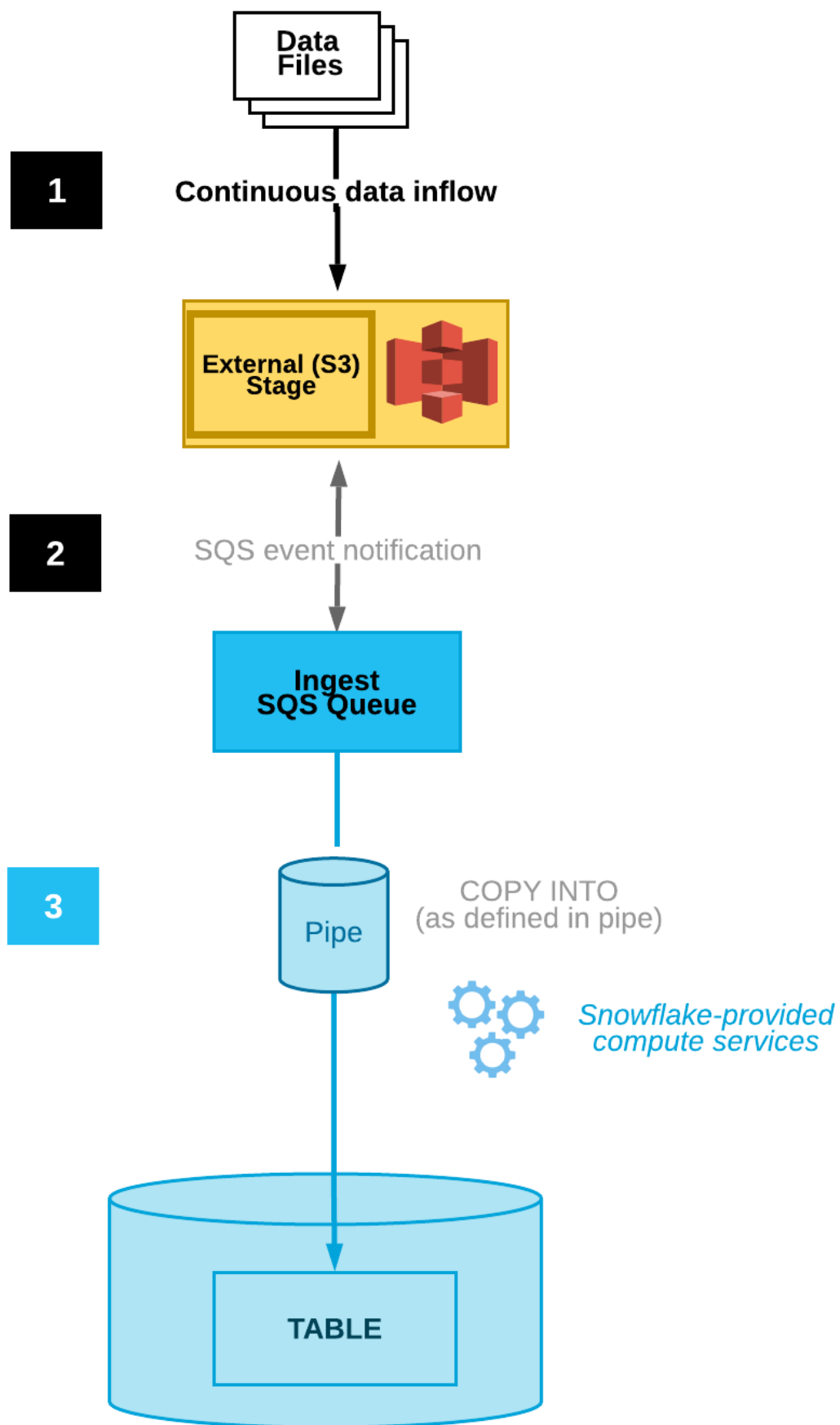


- 스테이징 된 파일을 감지하는 방식
 - 클라우드 메시징을 사용한 자동화 (AUTO_INGEST)
 - REST 엔드포인트 호출 (REST)

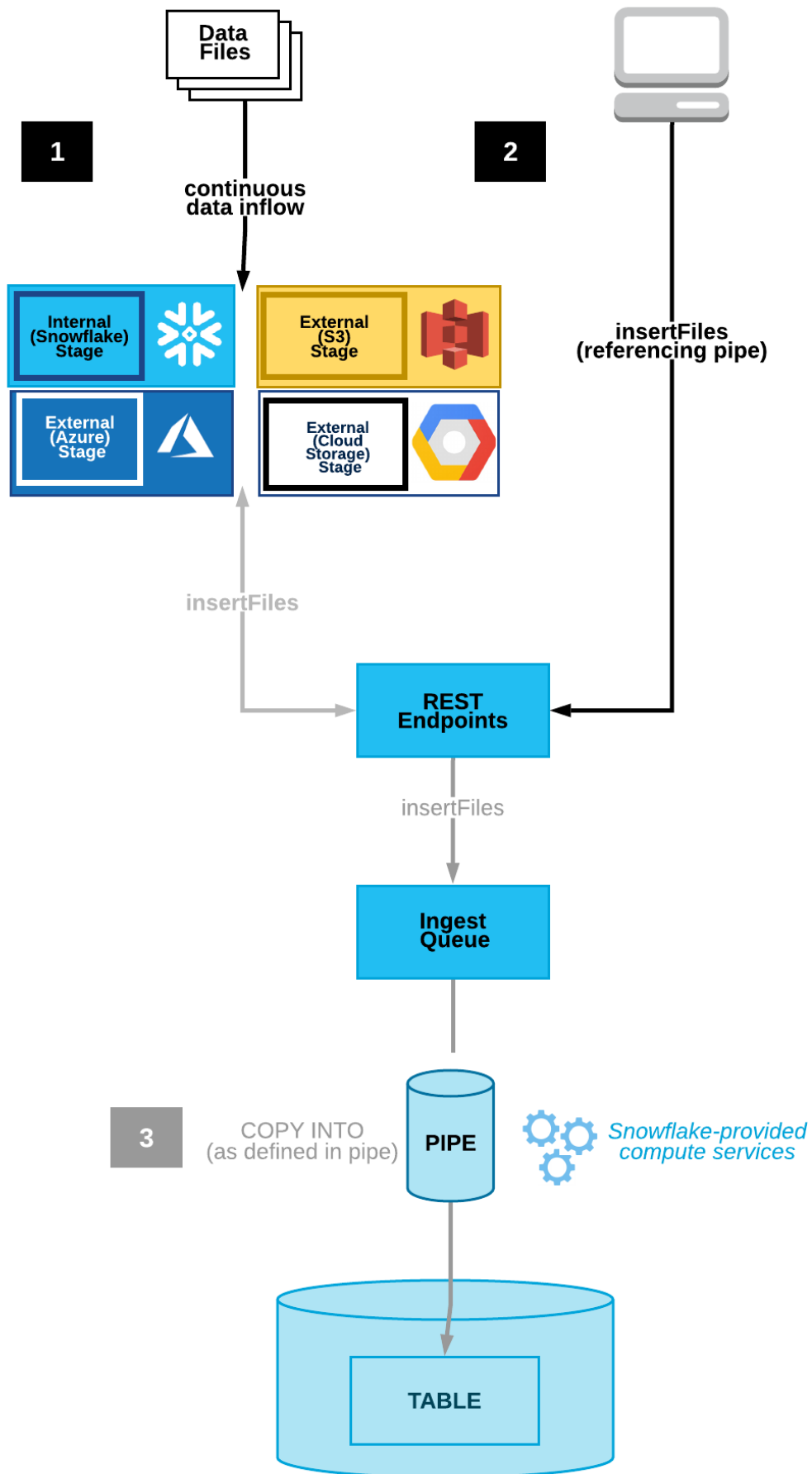


COPY INTO 명령(배치) 와 Snowpipe 의 차이

- Snowpipe 는 서버리스 모델로 가상 웨어하우스가 필요하지 않기에, 컴퓨팅 사용량에 따른 과금정책을 따름 (컴퓨팅 시간 외 INGEST 된 파일 1000개당 0.06 크레딧 사용 비용이 있음)
- 클라우드 메시징을 사용한 자동화



- 클라우드 저장소에 대한 이벤트 알림을 활용하여 Snowpipe에 로드할 새 데이터 파일이 수신 되었음을 알림
 - Snowpipe는 지정된 파이프 오브젝트에 정의된 매개변수를 기반으로 대상 테이블에 로드되는 큐에 쌓인 파일을 복사
 - Pipe가 일시 중지 되는 경우 이벤트 메시지는 14일 동안 보존되며, 14일 초과 시 Snowflake 내부 메타데이터에서 알림을 삭제
 - 비용, 이벤트 노이즈, 대기 시간을 줄이기 위해 Snowpipe에 대해 클라우드 이벤트 필터링을 사용하는 것을 권장
-
- **REST 엔드포인트를 사용한 자동화**



- 공용 Snowpipe REST Endpoints 호출에는 키 기반 인증을 사용
- 워크 플로
 1. 데이터 파일은 내부(Snowflake) 또는 외부(Amazon S3, Google Cloud Storage 또는 Microsoft Azure) 스테이지에 복사
 2. 클라이언트는 수집할 파일 목록과 정의된 파이프를 사용하여 `insertFiles` 엔드포인트를 호출, 엔드포인트는 이러한 파일을 수집 큐로 이동
 3. Snowflake가 제공하는 가상 웨어하우스는 지정된 파이프에 정의된 매개 변수를 기반으로 큐에서 대기 중인 파일의 데이터를 대상 테이블로 로드