

Introduction of Athena

- 서버리스의 S3를 이용하여 쿼리 혹은 분석하는 서비스이다
- ETL 설정은 하지 않고 S3 있는 모든 데이터를 가져온다
- 오직 데이터 스캐닝 하는 것. 즉 실행한 쿼리에 대해서만 비용 지불
- 쿼리 성능 향상으로 비용 절약
- S3에서 직접 스캐닝하므로 S3 비용 말고 추가적인 비용 발생 없음
- Ansi SQL을 지원하는 Presto 사용
- 다양한 포맷 형식과 호환 (CSV, JSON, ORC, Avro, Parquet)
- 어떤 SQL문이든 잘 수행함
- S3를 사용해 데이터 가용성 및 내구성 좋음
- 병렬 방식으로 쿼리를 실행. 쿼리 종류에 따라 더 빠를 수도 있음
- 쿼리 실행은 DML이 아닌 쿼리를 실행한다. 즉(삽입하거나, 업데이트)하지 않는다.
- 쿼리 사용 기록이 History에 남는다.

※Presto는 Facebook 개발자가 대량의 데이터에 대해 대화형 분석을 실행하기 위해 만든 오픈소스 분산 SQL 쿼리 엔진

작동 수행 방식

1. S3 버킷 및 오브젝트 생성
2. 메타데이터 데이터베이스 생성
3. 스키마 생성
4. 시리얼라이저/디시리얼라이저 미세조정 - Serde (선택사항)
5. 쿼리 실행
6. 히스토리에 액세스

Athena는 Apache Web 서버는 S3 cloudwatch를 통해 로그를 생성하는 데, 이를 Athena가 쿼리할 수 있음

또한 엑셀같은 쉽표로 구분된 것도 가능

또한 외부 데이터베이스는 S3에 저장한 다음 쿼리 가능