

AWS Athena

작성일: 2024-08-19

작성자: 이서연

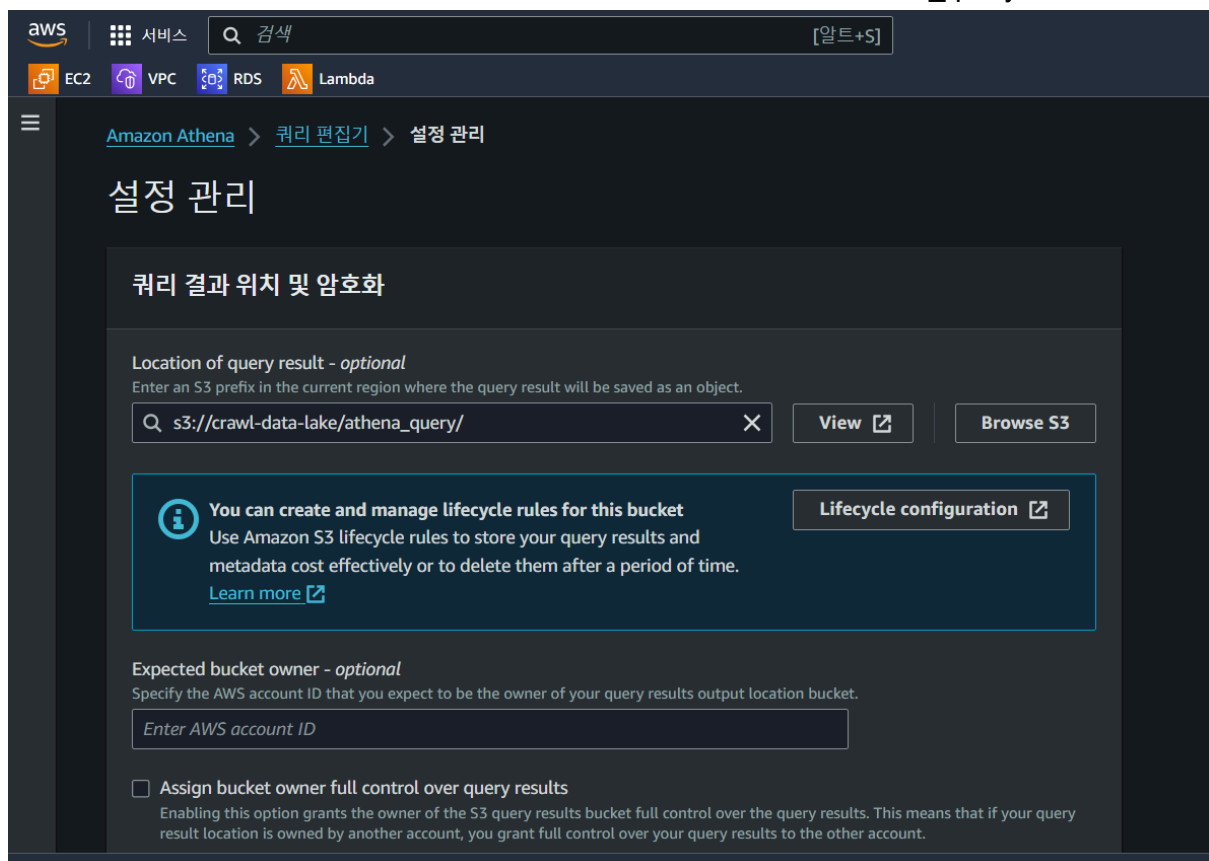
hive와 같은 기능을 제공하며 S3에서 JSON으로 저장한 파일을 SQL의 형태로 쿼리를 할 수 있다. Hive와 유사한 기능을 제공한다.

쿼리 당 청구로 비용이 산정되며, 쿼리로 스캔된 데이터에 대한 요금만 청구된다(실패한 쿼리거나 CREATE, DROP 같은 DDL은 청구되지 않음). TB 당 청구됨

사용법

1. 쿼리 결과 저장 설정

설정에서 쿼리 저장할 S3 버킷의 경로 설정. 예시에서는 athena_query에 저장



2. 편집기에서 데이터베이스 생성

생략가능. 데이터베이스로 default가 생성되어 있음. SQL문과 동일. 쿼리는 하나씩만 가능. 여러 개 중첩해서 쓸 수 없다.

```
CREATE DATABASE rocketpunch;
```

3. 편집기에서 테이블 생성

테이블 생성. 변수값 설정을 자동으로 생성하고 싶다면 AWS의 Glue 사용해야 함.

```
CREATE EXTERNAL TABLE rocketpunch (  
    company_id INT,  
    company_name STRING,  
    description STRING,  
    job_id INT,  
    job_title STRING,  
    job_career STRING,  
    timestamp DATE,  
    crawl_domain STRING,  
    job_url STRING,  
    date_end STRING,  
    date_start STRING,  
    job_task STRING,  
    job_specialties STRING,  
    job_detail STRING,  
    job_industry STRING  
)  
ROW FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe'  
LOCATION 's3://crawl-data-lake/rocketpunch/data/';  
  
# 칼럼에 타입값만 넣어줌  
# LOCATION에는 가져올 파일의 디렉토리를 설정
```

4. 쿼리 조회

쿼리 조회 결과는 하단과 같음. Hive와 유사한 기능 제공

```
SELECT * FROM rocketpunch limit 5;
```

결과 (5)

복사

결과 다운로드

🔍 행 검색

< 1 > ⚙️

# ▾	company_id ▾	company_name ▾	description ▾	job_id ▾	job_title ▾	job_career ▾
1	142538	석세스코리아컨설팅	헤드헌팅 기업	151587	규제대응시스템 (AML, FDS 등) 개발팀 팀장	["경력"]
2	142538	석세스코리아컨설팅	헤드헌팅 기업	151588	App 개발팀 팀장	["경력"]
3	142538	석세스코리아컨설팅	헤드헌팅 기업	151590	블록체인 서비스 개발자	["경력"]
4	142538	석세스코리아컨설팅	헤드헌팅 기업	151586	스테이킹 서비스 개발자 (시니어)	["경력"]
5	142538	석세스코리아컨설팅	헤드헌팅 기업	151589	블록체인 기술 개발자	["경력"]