

```

1 Lab2. Queryingv AWS S3 Data using SQL
2
3 ※AWS Glue 서비스가 아닌 수동으로 Athena SQL 사용하기
4
5 1. S3 버킷 데이터로부터 테이블 생성하기
6 1)[서비스] > [분석] > [Athena] 클릭하여 Amazon Athena 페이지로 이동
7 2)athena-demo.py 실행
8 3){계정}-datalake-bucket/output/sample-json/ 에서 json 파일 생성 확인
9 4)Amazon Athena > [쿼리 편집기] > [편집기] 탭에서 [데이터] > [테이블 및 보기] 섹션의 [생성] 버튼 클릭
10 5)목록에서 [데이터 원본에서 테이블 생성] > [S3 버킷 데이터] 클릭
11
12 6)[S3 버킷 데이터에서 테이블 생성] 페이지에서
13 -[테이블 세부 정보] > [테이블 이름] : {계정}-sample-table
14 -[데이터베이스 구성] > [기존 데이터베이스 선택 또는 새 데이터베이스 생성] > [기존 데이터베이스 선택] > {계정}-glue-db}
15 -[데이터 세트] > [입력 데이터 세트의 위치] > [S3 찾아보기] 클릭
16 --{계정}-datalake-bucket > output > sample-json 폴더 선택
17 -[데이터 형식] > [테이블 유형] : Apache Hive
18 -[데이터 형식] > [파일 형식] : [중첩] > JSON
19 -[열 세부 정보]
20 --[열 이름] : hakbun, [열 유형] : string, [열 추가] 버튼 클릭
21 --[열 이름] : info, [열 유형] : string, [열 추가] 버튼 클릭
22 --[열 이름] : subjects, [열 유형] : map
23 ---[매핑 키 유형] : string, [매핑 값 유형] : smallint
24 -[테이블 쿼리 미리 보기]에서 확인
25 -[테이블 생성] 버튼 클릭
26 -[쿼리 결과] 탭에서 "완료됨" 확인
27
28 7)[테이블 및 보기] > [테이블] > {계정}-sample-table 생성 확인
29 8)테이블 이름 옆 세로점을 클릭 > [쿼리 실행] > [테이블 미리 보기]
30 9)"이 쿼리를 여시겠습니까?" 창에서 [쿼리 열기] 클릭
31 9)스키마대로 값이 [결과]에서 확인할 수 있음.
32
33
34 2. 생성된 테이블에서 데이터 쿼리하기
35 1)[쿼리 편집기]에 다음과 같이 코딩한다.
36 -"Functions in Amazon Athena"로 Googling해서 다음 링크 확인
37 ref to : https://docs.aws.amazon.com/athena/latest/ug/functions.html
38 with tbl AS (
39     SELECT hakbun,
40         # 위의 웹 페이지에서 JSON을 찾고 https://trino.io/docs/current/functions/json.html 페이지에서 'json_extract_scalar()' 검색
41         json_extract_scalar(info, '$.name') as name,
42         json_extract_scalar(info, '$.city') as city,
43         json_extract_scalar(info, '$.gender') as gender,
44         # 위의 웹 페이지에서 Map을 찾고 https://trino.io/docs/current/functions/map.html 페이지에서 'name_to_age_map' 검색
45         subjects['kor'] as Korean,
46         subjects['eng'] as English,
47         subjects['mat'] as Math,
48         subjects['edp'] as Edps
49
50     FROM "{계정}-glue-db"."{계정}-sample-table"
51 )
52
53 SELECT * FROM tbl;
54
55 2)[실행] 버튼 클릭
56
57
58 3. Query 결과를 저장하기
59 1)방금 위에서 실행한 결과를 저장하려면 [결과] 섹션의 [결과 다운로드] 버튼을 클릭한다.
60 2)결과 데이터를 Clipboard에 복사하려면 [결과] 섹션에서 [복사] 버튼을 클릭하면 된다.
61 3)[Amazon Athena] > [쿼리 편집기] 페이지의 상단에 있는 [설정] 탭으로 이동
62 4)[쿼리 결과 및 암호화 설정] 섹션에서 [관리] 버튼 클릭
63 5)먼저, S3 페이지로 이동하여 결과 데이터를 담을 Bucket 생성
64 -[버킷 이름] : {계정}-athena-bucket
65 6)Athena 쿼리 편집기 설정관리 페이지에서
66 -[쿼리 결과 위치 및 암호화] > [쿼리 결과의 위치] > [S3 찾아보기]
67 -방금 생성한 {계정}-athena-bucket 선택
68 -[저장] 버튼 클릭
69 7)다시 [편집기] 탭으로 이동하여 [다시 실행] 클릭
70 8)[Amazon S3] > [버킷] > 위에서 생성한 {계정}-athena-bucket > Unsaved > 날짜별로 폴더 생성 > csv 데이터 확인
71 9)이 csv 파일을 다운로드 받아서 내용 확인
72 10)쿼리를 저장할 수 있는데, 쿼리 번호 옆의 세로점을 클릭하여 [다른 이름으로 저장] 클릭
73 11)[쿼리 저장] 팝업페이지에서
74 -[쿼리 이름] : Select Student
75 -[쿼리 저장] 버튼 클릭
76 -[Amazon Athena] > [쿼리 편집기] > [저장된 쿼리] 탭에서 확인 가능
77
78
79 4. 작업 그룹 관리하기
80 1)[Amazon Athena]의 왼쪽 메뉴 중 [작업 그룹] 페이지로 이동
81 2)[작업 그룹] 페이지에서
82 -primary 작업 그룹 확인
83 3)[작업 그룹 생성] 버튼 클릭
84 -[작업 그룹 생성] 페이지에서

```

85 -[작업 그룹 이름] : develop
86 -[작업 그룹 생성] 버튼 클릭
87 4)[쿼리 편집기] 페이지에서 [작업 그룹]을 "deveop"로 맞춘 후, 쿼리 실행