

```

1 Lab. Using AWS Athena
2
3 [Task1] Log Data처리하기
4 1. 데이터 가져오기
5   log_cord.log
6
7
8 2. S3에 Bucket 생성하기
9   1)버킷 이름 : {계정이름}-s3-{날짜}
10  2)AWS 리전 : 아시아 태평양(서울) ap-northeast-2
11  3)[이 버킷의 퍼블릭 액세스 차단 설정] > [모든 퍼블릭 액세스 차단] 모두 체크해제
12     -[현재 설정으로 인해 이 버킷과 그 안에 포함된 객체가 퍼블릭 상태가 될 수 있음을 알고 있습니다.] 체크
13
14  4)나머지 기본값 그대로 --> [버킷 만들기]
15
16  5)Publicc Access로 수정
17     -[권한] 탭 > [버킷 정책] > [편집] 클릭
18     -[버킷 ARN] 복사
19     -[정책 생성기] 클릭
20     -[AWS Policy Generator] 창에서
21         --Step 1: Select Policy Type > Select Type of Policy : S3 Bucket Policy
22         --Step 2: Add Statement(s)
23             ---Effect : Allow
24             ---Principal : *
25             ---Actions : GetObject 체크
26             ---Amazon Resource Name (ARN) : 버킷 ARN 값 붙여넣기
27             ---[Add Statement] 클릭
28         --Step 3: Generate Policy
29             ---[Generate Policy] 클릭
30         --[Policy JSON Document]에서
31             ---Resource : ARN 뒤에 반드시 /* 추가
32             ex)"Resource": "arn:aws:s3:::henry-mybucket-0219/*",
33             ---코드 전체 복사
34             ---[Close] 클릭
35     -다시 [버킷 정책 편집]창으로 돌아와서 [정책]에 붙여넣기
36     -[변경 사항 저장] 클릭
37     -버킷이 [퍼블릭 액세스 가능] 확인
38
39  6)log_cord.log 버킷에 업로드
40
41
42 3. AWS Glue Crawler 생성
43  1)[서비스] > [분석] > [AWS Glue]
44  2)좌측메뉴에서 [Data Catalog] > [Crawlers] 선택
45  3)[Create crawler] 클릭
46  4)Name : {계정이름}-crawler > Next
47
48  5)Choose data sources and classifiers
49     --Is your data already mapped to Glue tables? : Not yet
50     --[Add a data source] Click
51     --Data source : S3
52     --S3 path : s3://{bucket-name} <--[Browse S3] 클릭해서 선택해도 됨.
53     --[Add an S3 data source] 클릭
54     --[Next] 클릭
55
56  6)Configure security settings
57     -[Create new IAM role]
58     -[Enter new IAM role] : AWSGlueServiceRole-{계정이름}-crawler-role
59     -Next
60
61  7)Set output and scheduling
62     -[Add database] Click
63     --Create a database
64         ---Name : {계정이름}-athena-db
65         ---[Create a database] Click
66     -다시 돌아와서
67     -Target database : {계정이름}-athena-db
68     -Crawler schedule
69         --Frequency : On deman
70     -Next
71
72  8)[Create Crawler] Click
73
74
75 4. Crawler 실행
76  1)생성된 Crawler 클릭
77  2)[Run] 클릭
78  3)실행이 끝나면 [Status]가 "Running" -> "Stopping" -> "Ready"가 되고, [Last run]이 "Succeeded"로 바뀌며, [Table changes from last run]이 "1 created"로 변경됨.
79
80
81 5. Table 확인
82  1)좌측 메뉴에서 [Data Catalog] > [Databases] > [Tables] 클릭
83  2)목록에서 [View data] > "Table data" 클릭

```

84 3)[View data] 팝업창에서 [Proceed] 클릭
85 4)Amazon Athena > 쿼리 편집기 실행
86 5)[데이터 원본] : AwsDataCatalog
87 6)[데이터베이스] : {계정이름}-athena-db
88 7)[테이블] > 버킷이름
89 8)[결과]에 내용 출력됨
90
91
92 6. Athena 실행
93 1)[설정] 탭 > [관리]
94 2)[쿼리 결과의 위치] : [S3 찾아보기]
95 -s3://{버킷이름}
96 3)[저장]
97 4)[편집기] 탭 > [테이블] > 테이블 이름 > ':' >? 테이블 미리 보기
98 5)SELECT * FROM "henry-athena-db"."henry_mybucket_0219" WHERE user_id='jake'
99 6)SELECT * FROM "henry-athena-db"."henry_mybucket_0219" WHERE user_id='jake' ORDER BY time DESC
100
101
102 7. 실습1
103 sungjuk for Athena.json
104
105
106 8. 실습2
107 -Titanic dataset
108 <https://github.com/datasciencedojo/datasets/blob/master/titanic.csv>