

```

1 Lab. Using AWS Athena
2
3 [Task1] Log Data처리하기
4 1. 데이터 가져오기
5   log_cord.log
6
7 2. S3에 Bucket 생성하기
8   1)버킷 이름 : {계정이름}-s3-{날짜}
9   2)[이 버킷의 퍼블릭 액세스 차단 설정] > [모든 퍼블릭 액세스 차단] 모두 체크해제
10      -[현재 설정으로 인해 이 버킷과 그 안에 포함된 객체가 퍼블릭 상태가 될 수 있음을 알고 있습니다.] 체크
11
12   3)나머지 기본값 그대로 --> [버킷 만들기]
13
14   4)Publicc Access로 수정
15      -[권한] 탭 > [버킷 정책] > [편집] 클릭
16      -[버킷 ARN] 복사
17      -[정책 생성기] 클릭
18      -[AWS Policy Generator] 창에서
19          --Step 1: Select Policy Type > Select Type of Policy : S3 Bucket Policy
20          --Steop 2: Add Statement(s)
21              ---Effect : Allow
22              ---Principal : *
23              ---Actions : GetObject 체크
24              ---Amazon Resource Name (ARN) : 버킷 ARN 값 붙여넣기
25              ---[Add Statement] 클릭
26          --Step 3: Generate Policy
27              ---[Generate Policy] 클릭
28          --[Policy JSON Document]에서
29              ---Resource : ARN 뒤에 반드시 /* 추가
30              ex)"Resource": "arn:aws:s3:::henry-mybucket-0219/*",
31              ---코드 전체 복사
32              ---[Close] 클릭
33      -다시 [버킷 정책 편집]창으로 돌아와서 [정책]에 붙여넣기
34      -[변경 사항 저장] 클릭
35      -버킷이 [퍼블릭 액세스 가능] 확인
36
37   4)'시군별유동인구데이터.json' 버킷에 업로드
38
39
40 3. AWS Glue Crawler 생성
41   1)[서비스] > [분석] > [AWS Glue]
42   2)좌측메뉴에서 [Data Catalog] > [Crawlers] 선택
43   3)[Create crawler] 클릭
44   4)Name : {계정이름}-crawler > Next
45   5)Choose data sources and classifiers > [Add a data source]
46      --Data source : S3
47      --S3 path : s3://{bucket-name}
48      --[Add an S3 data source] 클릭
49      --[Next] 클릭
50   6)Configure security settings
51      -[Create new IAM role]
52      -[Enter new IAM role] : AWSGlueServiceRole-{계정이름}-crawler-role
53      -Next
54   7)Set output and scheduling
55      -[Add database]
56          --Create a database
57              ---Name : {계정이름}-athena-db
58      -다시 돌아와서
59      -Target database : {계정이름}-athena-db
60      -Crawler schedule
61          --Frequency : On deman
62      -Next
63   10)[Create Crawler]
64
65
66 4. Crawler 실행
67   1)생성된 Crawler 클릭
68   2)[Run crawler] 클릭
69   3)실행이 끝나면 [Status]가 Complete가 되고 [Table changes]가 1 table change로 변경됨.
70
71
72 5. Table 확인
73   1)좌측 메뉴에서 [Tables] 클릭
74   2)목록에서 [View data] > [Table data] 클릭
75   3)[View data] 팝업창에서 [Proceed] 클릭
76   4)Amazon Athena > 쿼리 편집기 실행
77   5)[데이터 원본] : AwsDataCatalog
78   6)[데이터베이스] : {계정이름}-athena-db
79   7)[테이블] > 버킷이름
80   8)[결과]에 내용 출력됨
81
82
83 6. Athena 실행
84   1)[설정] 탭 > [관리]

```

```

85 2)[쿼리 결과의 위치] : [S3 찾아보기]
86 -s3://{버킷이름}
87 3)[저장]
88 4)[편집기] 탭 > [테이블] > 테이블 이름 > ':' >? 테이블 미리 보기
89 5)SELECT * FROM "henry-athena-db"."henry_mybucket_0219" WHERE user_id='jake'
90 6)SELECT * FROM "henry-athena-db"."henry_mybucket_0219" WHERE user_id='jake' ORDER BY time DESC
91
92
93
94 [Task2] 일반 JSON 데이터 처리하기
95 1. 데이터 가져오기
96 sungjuk for Athena.json
97
98 2. S3에 Bucket 생성하기
99 1)버킷 이름 : {계정이름}-
100 2)[이 버킷의 퍼블릭 액세스 차단 설정] > [모든 퍼블릭 액세스 차단] 모두 체크해제
101 -[현재 설정으로 인해 이 버킷과 그 안에 포함된 객체가 퍼블릭 상태가 될 수 있음을 알고 있습니다.] 체크
102
103 3)나머지 기본값 그대로 --> [버킷 만들기]
104
105 4)Publicc Access로 수정
106 -[권한] 탭 > [버킷 정책] > [편집] 클릭
107 -[버킷 ARN] 복사
108 -[정책 생성기] 클릭
109 -[AWS Policy Generator] 창에서
110 --Step 1: Select Policy Type > Select Type of Policy : S3 Bucket Policy
111 --Steop 2: Add Statement(s)
112 ---Effect : Allow
113 ---Principal : *
114 ---Actions : GetObject 체크
115 ---Amazon Resource Name (ARN) : 버킷 ARN 값 붙여넣기
116 ---[Add Statement] 클릭
117 --Step 3: Generate Policy
118 ---[Generate Policy] 클릭
119 --[Policy JSON Document]에서
120 ---Resource : ARN 뒤에 반드시 /* 추가
121 ex)"Resource": "arn:aws:s3:::henry-mybucket-0219/*",
122 ---코드 전체 복사
123 ---[Close] 클릭
124 -다시 [버킷 정책 편집]창으로 돌아와서 [정책]에 붙여넣기
125 -[변경 사항 저장] 클릭
126 -버킷이 [퍼블릭 액세스 가능] 확인
127
128 4)'sungjuk for Athena.json' 버킷에 업로드
129
130
131 3. AWS Glue Crawler 생성
132 1)[서비스] > [분석] > [AWS Glue]
133 2)좌측메뉴에서 [Data Catalog] > [Crawlers] 선택
134 3)[Create crawler] 클릭
135 4)Name : {계정이름}-crawler > Next
136 5)Choose data sources and classifiers > [Add a data source]
137 --Data source : S3
138 --S3 path : s3://{bucket-name}
139 --[Add an S3 data source] 클릭
140 --[Next] 클릭
141 6)Configure security settings
142 -[Create new IAM role]
143 -[Enter new IAM role] : AWSGlueServiceRole-{계정이름}-crawler-role
144 -Next
145 7)Set output and scheduling
146 -[Add database]
147 --Create a database
148 ---Name : {계정이름}-athena-db1
149 -다시 돌아와서
150 -Target database : {계정이름}-athena-db1
151 -Crawler schedule
152 --Frequency : On deman
153 -Next
154 10)[Create Crawler]
155
156
157 4. Crawler 실행
158 1)생성된 Crawler 클릭
159 2)[Run crawler] 클릭
160 3)실행이 끝나면 [Status]가 Complete가 되고 [Table changes]가 1 table change로 변경됨.
161
162
163 5. Table 확인
164 1)좌측 메뉴에서 [Tables] 클릭
165 2)목록에서 [View data] > [Table data] 클릭
166 3)[View data] 팝업창에서 [Proceed] 클릭
167 4)Amazon Athena > 쿼리 편집기 실행
168 5)[데이터 원본] : AwsDataCatalog

```

```
169 6)[데이터베이스] : {계정이름}-athena-db1
170 7)[테이블] > 버킷이름
171 8)[결과]에 내용 출력됨
172
173
174 6. Athena 실행
175 1)[설정] 탭 > [관리]
176 2)[쿼리 결과의 위치] : [S3 찾아보기]
177 -s3://{버킷이름}
178 3)[저장]
179 4)[편집기] 탭 > [테이블] > 테이블 이름 > ':' >? 테이블 미리 보기
180
181
182 7. 실습
183 1)Titanic dataset
184 https://github.com/datasciencedojo/datasets/blob/master/titanic.csv
```