

AWS 데이터 인프라 구축 실습 평가 (50분)

[실습 평가 개요]

당신은 한 중소기업의 데이터 인프라 엔지니어로 채용되었으며, 사내 데이터 분석 환경을 빠르게 구성하는 임무를 맡았다.

이 과제에서 AWS 서비스를 활용하여, 기본적인 데이터 분석 인프라를 구성하고, 샘플 데이터를 적재하고 확인할 수 있다

[목표]

1. **S3 버킷**을 생성하여 데이터를 저장할 공간을 확보한다.
2. **IAM 역할**을 생성하여 분석 서비스가 S3에 접근할 수 있도록 한다.
3. **AWS Glue Crawler**를 통해 S3에 있는 데이터를 메타데이터로 등록한다.
4. **Athena**에서 Glue Data Catalog를 활용해 쿼리를 실행한다.

[실습 상세]

(문서 뒤에서 제출해야 할 캡처 화면 목록을 먼저 보면서 진행한다)

1. S3 버킷 생성

- AWS S3 콘솔에 접속하여 다음 조건을 만족하는 **새로운 S3 버킷**을 생성하시오.
 - 버킷 이름: student-데이터-이니셜 (예: student-data-bh)
 - 퍼블릭 접근 차단 설정은 **기본값 유지**
 - 버전 관리는 **비활성화 상태 유지**
 - 리전은 **버지니아 북부(us-east-1)**에 생성할 것
 -

2. 샘플 CSV 파일 업로드

- 강사 깃허브(<https://github.com/digicope/aws0414>)에 있는 아래 CSV 파일을 다운로드하여, 생성한 S3 버킷에 /input/ 폴더를 만들고 업로드 하시오.

kaggle_credit_card_lab.csv (42.5MB)

- 업로드 후, S3 콘솔에서 정상 업로드 여부를 확인하고 캡처하시오.

3. IAM 역할 생성 및 정책 연결

- 이름이 GlueS3AccessRole-이니셜인 IAM 역할을 생성하시오.
 - 신뢰할 수 있는 엔터티는 **Glue**로 설정
 - 정책은 다음 두 가지를 연결할 것:
 - AmazonS3ReadOnlyAccess
 - AWSGlueServiceRole
- 생성된 역할 이름과 연결된 정책을 캡처하시오.

4. AWS Glue Crawler 설정 및 실행

- AWS Glue 콘솔에서 다음 조건으로 Crawler를 생성하시오.
 - 이름: student-crawler-이니셜
 - 데이터 소스: 본인이 생성한 S3 버킷의 /input/ 경로
 - IAM 역할: 위에서 생성한 IAM 역할 사용
 - 데이터베이스: student_db (없다면 새로 생성)
 - 크롤러 실행 후, 테이블이 정상 생성되었는지 Glue 콘솔에서 확인하시오.
- Glue 데이터 카탈로그에서 테이블 이름과 스키마 정보를 캡처하시오.

5. Athena 쿼리 실행

- Athena 콘솔로 이동하여 다음 조건으로 쿼리를 수행하시오.
 - 쿼리: 방금 생성된 테이블에서 전체 행을 조회하는 SELECT * FROM 테이블명 LIMIT 10;
 - 결과 저장 위치는 본인이 생성한 S3 버킷의 /query-results/ 디렉터리로 지정
- 쿼리 결과 캡처하시오.

[제출 내용 정리]

아래 항목들을 스크린샷으로 캡처 or 텍스트 파일로 작성하여 압축해서 제출 :

1. 생성된 S3 버킷 이름 및 /input/ 경로에 업로드 된 CSV 파일 스크린샷
2. IAM 역할 생성 완료 화면 및 연결된 정책 목록
3. Glue Crawler 실행 완료 화면 및 테이블 생성 확인(스키마 목록 화면 포함)
4. Athena 쿼리 수행 화면 (설정관리 페이지와 쿼리문 , 결과 출력 포함)
5. 아래 쿼리문을 수행하고 쿼리 결과를 캡처하여 제출하세요

(database와 table 이름은 각자에 맞게 변경해서 사용한다)

```
SELECT COUNT(*) AS count, SUM(amount) AS total_price
```

```
FROM "student_db"."input";
```

6. 아래 쿼리문을 수행하고 결과값에 대하여 설명하세요(텍스트 파일로 제출)

```
SELECT * FROM "student_db"."input"
```

```
WHERE amount >= 1000 LIMIT 10;
```

7. AWS 클라우드 기반 데이터 인프라 구축의 장점과 구축에 사용되는 서비스들에 대하여 자유롭게 설명하세요(서술형)

화면 캡처 파일이름을 차례 대로 숫자를 붙이고 작업 내용을 제목으로 만든 다음 전체를 하나로 압축하여 아래와 같은 파일로 강사 메일로 제출한다

제출 파일명 : 홍길동_과제물.zip , 제출기한 2025-04-18 pm 06:00

강사 메일로 제출 : digicope@aicore.co.kr

[설정 값 예시]

(※ 구성의 기준을 위한 예시이며, 개인별 이름은 다를 수 있음)

항목	설정 값 예시
S3 버킷 이름	student-data-bh
업로드 경로	s3://student-data-bh/input/kaggle_credit_card_lab.csv
IAM 역할 이름	GlueS3AccessRole-bh
연결된 정책	AmazonS3ReadOnlyAccess, AWSGlueServiceRole
Glue 크롤러 이름	student-crawler-bh
생성된 데이터베이스	student_db
테이블 이름	credit_card
Athena 쿼리	SELECT * FROM student_db. credit_cardv LIMIT 10;
쿼리 결과 저장 경로	s3://student-data-bh/query-results/
출력 행 수	10

[수고하셨습니다!!]