

AWS 종합 실습

주제 : AWS를 활용한 데이터 분석 서비스 파이프라인 구축

실습 개요 :

- 상황 : 데이터 분석가/엔지니어로서 현업을 진행하면서 데이터 수집 시 자동으로 분석 코드가 실행되는 파이프라인을 구축하려고 한다.
- 목적 : AWS 서비스를 사용하여 데이터 분석 파이프라인이 어떤 구조로 진행되는지 이해하고, AWS 서비스 사용에 대한 역량 향상을 위함.

제출 안내 :

- 대상 : 빅데이터 9기 교육생
- 일자 : 2024-10-24(목) 09:00 ~ 17:00 (휴게시간 : 12:00 ~ 13:00 / 총 7시간)
 - 13시, 15시에 중간 점검 및 질문 받겠습니다.
- 질문이 있을 시 사유가 명확한 질문 권장드립니다.

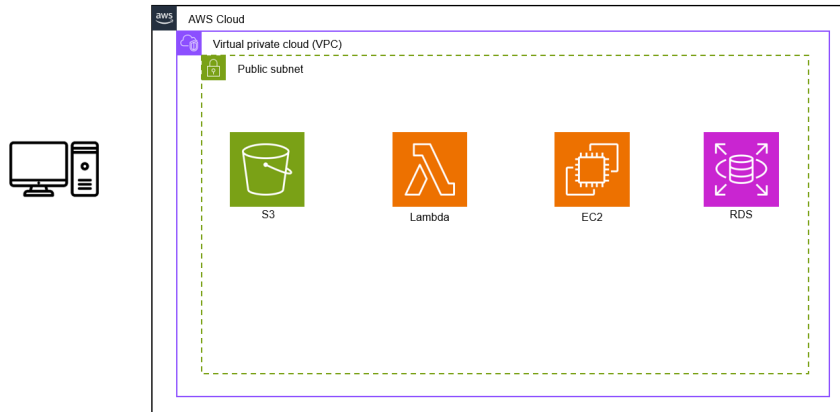
ex. 모범 사례

- EC2 생성을 하려는데(목표점 제시)
 - VPC를 임의로 설정했더니(예상되는 문제의 원인)
 - 인스턴스 유형에서 T2 시리즈가 비활성화 됐어요.(문제 상황 설명)
 - 활성화하는 방법이 궁금합니다(원하는 목표 방향)
-
- 검색, 의논, 공유 모두 자유롭게 진행 가능합니다.

산출물 안내

- 실습에 대한 AWS 아키텍처 작성
- '실습 목표'에 대한 수행 과정 및 결과 보고 (자유 형식)
- AWS 아키텍처란?
 - AWS 서비스 및 인프라 구성요소를 조합하여 솔루션에 대한 이해를 돕기위해서 만드는 구조적 청사진.

- 트래픽의 흐름, 서비스 진행 과정 등을 자유롭게 서술하면 됨.



— 아이콘 + 영역 + 설명 + 화살표 —

실습 목표 :

- 제출 자료는 'AWS_실습_[기수]_[이름].pptx' 형식에 맞춰서 구글 클래스룸에

2024-10-24 23:59 까지 제출 바랍니다.

- 결과물은 pptx 파일 1개 이상으로 제출합니다.
 - 부연 설명을 위한 보고서 파일(docx, hwp, txt, pdf 등), 이미지 파일 모두 함께 제출 가능합니다.
 - 2개 이상의 파일 제출 시 압축하여 제출하시길 바랍니다.
 - 제출 시 'AWS_실습_[기수]_[이름].zip' 의 형식에 맞춰 제출하길 바랍니다.
- 각 항목별 요구사항은 **파란색+ 기울임 글씨체**로 작성됩니다.

가. AWS 환경 구축하기

1. EC2 생성 :

아래 사항대로 생성됐다면 OK

- Ubuntu 24.04 LTS
- 인스턴스 유형 : T2.medium
- 네트워크 설정 :
 - vpc : 기본 값

- ii. 보안 그룹 : ssh(22), http(80), https(443), 5000포트, 8888포트 인바운드 허용 / 아웃바운드 모든 트래픽 허용
 - d. 스토리지 : 32GiB 이상
2. S3 생성 :

아래 사항대로 생성됐다면 OK

- a. 자유 설정. (단, 버킷 퍼블릭 액세스는 차단함.)
 - b. bucket 이름 규칙에 맞게 생성하되, 'big9_test' 라는 문구를 접두 부분에 작성
3. EC2에 Python 개발 환경 구축 (아래 중 자유롭게 선택)

py파일만 실행되면 OK

- a. apt install Python & Python virtual environment
- b. Anaconda 환경 구축
- c. Jupyter Notebook 띄우기 (권장 / 필수x)

나. AWS 데이터 분석 서비스 파이프라인 구축하기

1. Local PC에서 boto3를 사용하여 '가-2' 에서 만든 S3 Bucket로 데이터(csv) 업로드 기능 구현하기

코드를 통해 업로드 가능 시 OK

2. EC2 인스턴스에서 S3 Bucket 데이터(csv) 다운로드 기능 구현하기

EC2 인스턴스 내부에서 코드(or 명령어)를 통해 다운로드 가능 시 OK

3. EC2 인스턴스에서 분석 코드 실행하기

EC2 인스턴스 내부에서 코드를 통해 구현 시 OK

4. EC2 인스턴스에서 S3 Bucket으로 데이터 업로드 기능 구현하기

EC2 인스턴스 내부에서 코드(or 명령어)를 통해 다운로드 가능 시 OK

다. RDS에 데이터 적재하기

1. RDS 인스턴스 생성(자유롭게 생성)
 - a. DB 엔진 선택 자유
 - b. Table 형식 자유
2. 나 부분에서 발생하는 데이터 EC2에서 RDS로 데이터 적재 진행.
 - a. 원본 데이터 or 전처리 데이터 or **분석 결과(권장)** 등 자유롭게 진행
3. 스키마 정의 문서 작성

라. Lambda를 활용한 자동화

1. Lambda 함수 생성
2. Lambda 함수 트리거 생성
3. Lambda 함수 Layer 생성
4. Lambda 함수 정상 작동 테스트 진행
 - a. Test 기능을 사용하지 않아도 무방함.
5. Python 시각화 코드 작성 및 Image File S3 적재 자동화

마. 점검

1. 전체 파이프라인 테스트 및 디버깅+최적화
2. 개선안 제시