

AWS SQS(Simple Queue Service)

작성일: 2024-08-28

작성자: 유정연

1. 목적

본 문서는 **Lambda**와 **Airflow** 데이터 파이프라인 연결을 위한 이벤트 연동 기능을 구현하기 위해 조사한 내용을 기술한다. 문서는 현 데이터 파이프라인의 중점과 데이터 추출 클러스터의 추구하는 방향성을 기술하고, 이를 반영하기 위해 조사한 **AWS SQS**에 대한 내용으로 구성되어 있다.

2. 데이터 파이프라인 중점

채용 공고 데이터 수집 파이프라인은 **Airflow cluster**를 통해 오케스트레이션 되지만, 데이터 추출 단계는 **EventBridge Scheduler**의 배치성 일정으로 작동 된다. 이는 채용 데이터 특성과 **Airflow Cluster** 등 제어쪽에서 장애가 생겼을 때의 위험성 관리를 반영한 것으로, 오케스트레이션 단위에서 장애가 생기더라도 데이터 추출 만큼은 항상성을 보장한다. 따라서 데이터 추출 단계는 전체 데이터파이프라인과 독립성과 폐쇄성을 가지며, 항상성 유지와 외부 접근에 의한 보안 문제를 방지를 위해 이 두 성향을 유지할 필요가 있다. 그러면서도 데이터 파이프라인의 원활한 자동화 및 플로우 이행을 위해서 데이터 추출 단계와 **airflow**간의 이벤트 연동이 필요하다. 데이터 추출은 **EventBridge**를 통해 실행된 **Lambda** 함수가 수행하며, 일회성 이벤트이다. **airflow**측에서 이 이벤트 수신하지 못했을 경우, 이벤트 메시지를 재전송하거나 **TCP**처럼 송수신을 보장하기 어렵다. 그리고 데이터 특성상 해당 데이터를 즉시 반영할 필요는 없으므로, 비동기를 허용하는 손실없는 이벤트 메시지 전달이 필요하다.

3. AWS SQS란?

Amazon Simple Queue Service(Amazon SQS)는 AWS에서 제공하는 메시지 큐 서비스 중 하나로서 내구력 있고 가용성이 뛰어난 보안 호스팅 대기열을 제공하며 이를 통해 분산 소프트웨어 시스템과 구성 요소를 통합 및 분리할 수 있다. **Amazon SQS**는 배달 못한 편지 대기열 및 비용 할당 태그와 같은 공용 구성을 제공하며, **AWS SDK**가 지원하는 모든 프로그래밍 언어를 사용하여 액세스할 수 있는 일반 웹 서비스 **API**를 제공한다.

4. AWS SQS 특이점

AWS SQS는 Producer와 Consumer로 나뉜다는 Message Queue의 특징을 가진다. 하지만 Kafka와 같은 Message Queue와 달리, 메시지 전달 현황 표시를 Check point 방식이 아닌 Message 객체에 대해 ACK/NACK을 보낸다는 차이를 가진다. 즉 Message별로 대한 수신 및 Queue에서 제거(ACK) 액션과 처리 실패시 처리 재시도를 할 수 있도록 SQS에게 취소 요청(NACK)을 보낼 수 있다.

5. 참고

- [AWS SQS DOC]
(https://docs.aws.amazon.com/ko_kr/AWSSimpleQueueService/latest/SQSDeveloperGuide/welcome.html)
- [AWS SQS 사례 및 설명]
(<https://channel.io/ko/blog/tech-backend-aws-sqs-introduction>)