

# 카프카 클러스터 운영

## 카프카 클러스터를 운영하는 여러가지 방법

### 운영 방법에 따른 서비스 형태들



SaaS란 클라우드 서비스 제공업체가 소프트웨어와 관리하면서 플랫폼을 제공하는 것을 뜻한다. 사용자는 웹 대시보드나 CLI(Command Line Interface)를 통해 플랫폼의 세부 설정을 간편하게 할 수 있으며 네트워크, 스토리지, 가상화 등 다양한 종류의 운영상 이슈는 업체에서 관리하기 때문에 매우 편리하게 사용할 수 있다.

### 설치 방법별 서비스 형태 종류

서비스 종류	특징
온프레미스(on-premise)	- 사용자가 자체적으로 보유한 전산실 서버에 직접 설치해 운영 - 기업 상황에 맞게 하드웨어를 커스터마이징 구성 가능 - 초기 도입 비용, 운영 및 관리를 위한 유지보수 비용 발생
IaaS(Infrastructure-as-a-Service)	- 물리/가상 컴퓨팅 리소스를 발급받아서 사용 - 온라인 스토리지, 데이터베이스 등도 포함됨 - 사용자가 운영체제, 애플리케이션 등을 직접 설정, 배포, 운영
PaaS(Platform-as-a-Service)	- 애플리케이션 개발 및 실행 환경을 제공 - 사용자는 컴퓨팅 리소스 관리를 신경 쓰지 않아도 됨

서비스 종류	특징
SaaS(Software-as-a-Service)	- 소프트웨어의 배포, 실행을 업체에서 관리하고 기능을 제공 - 소프트웨어 관리를 업체에 위임하고 기능만 사용할 때 유용

## 서비스 형태별 카프카 운영 방법

서비스 종류	특징
온프레미스(on-premise)	- 물리장비(서버)를 구매, 네트워크 설치 구성 수행 - 물리장비 운영체제 설치 수행 - 오픈소스 카프카를 설치, 운영 - 또는 기업용 카프카(컨플루언트 플랫폼 등) 설치, 운영
IaaS(Infrastructure-as-a-Service)	- AWS, GCP와 같은 클라우드 서비스를 통해 물리/가상 컴퓨팅 리소스를 발급 - 컴퓨팅 리소스에 오픈소스 카프카 설치, 운영 - 또는 기업용 카프카(컨플루언트 플랫폼 등) 설치, 운영
SaaS(Software-as-a-Service)	- 컨플루언트 클라우드 또는 AWS MSK는 대표적인 SaaS - 다양한 주변 생태계(ksqlDB, 모니터링 도구 등)를 옵션으로 제공

## 오픈 소스 카프카를 직접 설치하여 운영하는 경우

IaaS 또는 온프레미스 환경에서 카프카 클러스터를 설치하여 운영하는 것이 가장 흔한 운영 방식. 카프카는 전송된 데이터를 모두 파일 시스템에 저장하고 대규모 데이터 통신이 일어나기 때문에 고성능의 하드웨어를 사용해야 한다. 컨플루언트에서는 상용 환경의 카프카 클러스터 운영시 브로커의 하드웨어를 다음과 같이 설정하는 것을 추천

- 메모리 : 32GB 머신에 힙 메모리 6GB로 설정. 나머지는 OS의 페리지 캐시 영역으로 활용
- CPU : 24core 머신 사용. 만약 SSL과 같은 보안 설정 사용할 경우 추가 더 높은 사용 필요
- 디스크 : RAID 10으로 설정된 디스크 사용. NAS는 사용하면 안됨.
- 네트워크 : 사용하는 데이터 통신량에 따라 다름.
- 파일시스템 : XFS 또는 ext4

## 오픈 소스 카프카를 직접 설치하여 운영하는 경우

항목	개발용 카프카 클러스터	상용 환경 카프카 클러스터
브로커 개수	5개	10개
메모리	16GB(heap memory 6GB)	32GB(heap memory 6GB)

항목	개발용 카프카 클러스터	상용 환경 카프카 클러스터
CPU	16core	24core
디스크	사용량에 따라 달라짐	사용량에 따라 달라짐

## SaaS형 아파치 카프카 소개

### 클라우드 서비스 - 컨플루언트

컨플루언트는 카프카에 대한 개념을 최초로 생각하고 아키텍처를 제안, 개발한 인물인 제이 크랩스와 그의 동료들이 설립한 회사. 컨플루언트는 아파치 카프카의 생태계를 가꾸고 발전시키는 데에 선구적인 역할을 하고 있다.

### 클라우드 서비스 - 컨플루언트 클라우드, 컨플루언트 플랫폼

컨플루언트 클라우드	컨플루언트 플랫폼
-클라우드 기반 카프카 클러스터 -요구사항에 따라 자동으로 늘려주는 클러스터 리소스 제공 - GCP, AWS 등 클러스터 설치 위치 지정(리전 단위) 가능 - 120개가 넘는 커넥터, ksqlDB, 스키마 레지스트리 서비스 제공 - 99.95% SLA - 엔터프라이즈 수준의 보안 수준 제공 - 데이터 적재 제한 없음	- 온프레미스 기반 설치형 카프카 클러스터 - 서버를 내부에서 발급하여 직접 설치 - 필요에 따라 컨플루언트 팀에서 지원, 학습 제공 - 단계별 스토리지 기능(Tiered-storage) 제공 - GUI 기반 모니터링 시스템 제공

### 클라우드 서비스 - AWS MSK

Amazon Managed Streaming for Apache Kafka(MSK)

- 완전관리형 고가용성 Apache Kafka 서비스를 통해 안전하게 데이터를 스트리밍

MSK(Managed Streaming for Apache Kafka)는 AWS에서 제공하는 SaaS형 아파치 카프카 서비스이다. AWS MSK는 AWS 인프라에서 카프카 클러스터를 생성, 업데이트, 삭제 등과 같은 운영 요소를 대시보드를 통해 제공한다. 또한, 안전하게 접속할 수 있도록 클러스터와 연동시 TLS인증 보안을 설정할 수 있다. MSK로 클러스터를 생성하고자 하는 엔지니어는 아파치 카프카 버전을 직접 선택할 수 있다. MSK로 생성한 클러스터는 AWS에서 제공하는 인프라 영역에 구축된다. MSK는 AWS에서 운영하는 애플리케이션과 쉽게 연동할 수 있기 때문에

AWS를 이미 사용 중인 기업에서는 어렵지 않게 아키텍처에 포함시킬 수 있다.

## SaaS형 아파치 카프카 장점과 단점

### SaaS로 카프카 클러스터를 운영할 경우 장점

#### 인프라 관리의 효율화

카프카 클러스터는 사용환경에서 최소 3대 이상의 서버로 운영. 클러스터를 담당하고 운영하는 개발자의 입장에서는 3대 이상의 서버를 모니터링 해야 한다, SaaS를 사용할 경우 인프라 운영 관련 역할에서 자유로울 수 있다. 왜냐하면 카프카 SaaS 서비스를 사용하게 되면 브로커가 올라가는 서버는 자동으로 관리되기 때문. 만약 브로커들 중 일부에 이슈가 발생하여 서버를 재시작하는 상황이 일어나더라도 SaaS 서비스가 이슈를 감지하여 이슈가 생긴 서버를 제외하고 신규 장비에 브로커를 실행하여 복구. 클러스터의 데이터 사용량이 순간적으로 많아지더라도 서비스를 제공하는 업체의 SaaS 대시보드에서 브로커 개수만 설정하여 쉽게 스케일 아웃 할 수 있다. 몇 분 뒤면 더 많은 양의 데이터 소화 가능한 카프카 클러스터 설정

#### 모니터링 대시보드 제공

카프카 클러스터를 직접 설치한다고해서 카프카를 잘 활용할 수 있는 것은 아니다. 브로커들이 제공하는 지표들을 수집하고 적재하고 대시보드화하여 데이터를 시각화해야 카프카 클러스터를 효과적으로 운영하기 위해 필요한 설정들을 수정하고 적용할 수 있기 때문. SaaS형 카프카에서는 자동화되어 만들어진 클러스터로부터 운영에 필요한 지표들을 수집하고 그래프로 보여주는 옵션 제공 만약 직접 카프카 클러스터를 운영했다면 수집한 지표를 저장할 저장소를 구축하고 대시보드를 운영하기 위해 신규로 추가 플랫폼을 설치, 운영해야 한다. 직접 카프카 클러스터를 운영하는 대신에 SaaS형 카프카를 사용함으로써 추가로 플랫폼들을 설치하고 운영할 필요없이, 클릭 몇 번으로 지표를 그래프로 보면서 운영의 짐을 덜수 있을 것이다.

## 보안 설정

보안 설정이 되지 않은 카프카 클러스터는 자물쇠가 달리지 않은 외양 간과 같다. 보안이 설정 되지 않은 클러스터의 경우 호스트와 포트 번호만 알면 모든 토픽의 데이터를 가져갈 수 있다. 카프카 브로커는 SSL, SASL, ACL과 같이 불특정 다수의 침입을 막기 위해 다양한 종류의 보안 설정 방안을 제공하고 있다. SaaS형 카프카에서는 클러스터 접속 시 보안 설정을 기본으로 제공하고 있다. 클러스터 생성 시 보안 설정을 통해 인가된 사용자만 카프카 클러스터에 접근할 수 있도록 할 수 있다.

## SaaS로 카프카 클러스터를 운영할 경우 단점

### 서비스 사용 비용

카프카 클러스터를 직접 설치하고 운영한다면 브로커(그리고 주키퍼)를 실행하는 서버의 사용 비용만 들이면 된다. 그러나 SaaS형 카프카를 사용하려면 SaaS를 제공하는 업체에서 지정한 특정 요금제를 사용해야 한다. 게다가 SaaS형 카프카에 적용된 금액은 서버만 발급해서 사용하는 비용에 비해 현저히 높다. 예를 들어, AWS MSK의 경우 3대 브로커로 인스턴스 (kafka.m5.xlarge: CPU 4, 메모리 16GB)들을 구성하여 사용하면 시간당 1.5달러(서울 리전 기준)로 한 달에 1,080달러(한화 약 120만원)의 비용이 발생한다. 반면 동일 사양의 인스턴스 (m5.xlarge)를 직접 발급하여 설치, 운영하면 시간당 0.7달러로 한 달에 504달러(한화 약 60만원)의 비용이 발생한다. 즉, 직접 서버를 발급하여 설치하고 운영하는 것이 AWS MSK를 사용하는 것에 비해 2배 이상 저렴하다. 인스턴스 발급 및 사용 비용은 클러스터를 단순히 실행만 하더라도 빠져나가는 비용이며 추가적으로 스토리지 요금과 데이터 전송 요금이 발생 한다.

### 커스터마이징의 제한

카프카 클러스터를 직접 운영하다 보면 서버의 최적화 옵션이나 카프카 브로커 옵션 같은 다양한 부분에서 사용자 설정들이 들어간다. 그러나 SaaS 서비스들은 모든 부분이 자동화되어 있고 업체의 아키텍처를 따라가기 때문에 상세 설정 적용과 클러스터 아키텍처 변화의 적용이

매우 어렵다. 특히 멀티 클라우드(2개 이상의 퍼블릭 클라우드를 함께 사용)나 하이브리드 클라우드(사내 서버와 퍼블릭 클라우드를 함께 사용) 형태로 카프카 클러스터를 구성 하는 것은 SaaS형 카프카에서는 불가능하다. 클라우드 종속성을 탈피하는 목적으로 그리고 마이그레이션을 목적으로 멀티 클라우드, 하이브리드 클라우드 형태의 클러스터로 구성해야 할 필요가 있을 때 SaaS형 카프카를 사용하고 있다면 문제가 매우 복잡해진다.

### 클라우드 종속성

SaaS형 카프카를 구축하기 위해서는 클라우드 서비스 업체를 선택하고 클러스터를 운영해야 한다. 클라우드 서비스 업체를 선택하고 클러스터를 운영하는 순간 해당 서비스에 종속된다. 클라우드 종속성을 탈피하기 위해서는 멀티 클라우드 또는 하이브리드 클라우드를 도입하는 방법이 있지만 SaaS형 카프카를 사용할 경우에는 멀티 클라우드, 하이브리드 클라우드를 도입하는 방법이 있지만 SaaS형 카프카를 사용할 경우에는 멀티 클라우드, 하이브리드 클라우드 적용이 불가능하다. 따라서 SaaS형 카프카를 도입할 경우 추후 발생할 수 있는 이슈를 예측해보고 서비스에 미치는 위험 정도를 따져본 뒤 도입 해야한다.

## 그럼에도 불구하고 SaaS형 카프카를 사용하는 것이 나은 경우는?

클라우드로 제공되는 SaaS형 카프카는 직접 브로커들을 설치하고 운영하는 것에 비해 확연한 장단점을 가지고 있다. 그럼에도 불구하고 카프카 클러스터에 대한 운영 노하우가 부족한 상태에서 빠르게 클러스터 인프라를 구축하고 활용할 때는 SaaS형 카프카가 최고의 선택지가 될 수 있다. 다만, SaaS형 카프카도 결국 카프카를 설치하여 구성하는 클러스터이므로 카프카에 대한 노하우가 없다면 세세한 운영 부분에 있어서는 어려움을 겪게 될 수도 있다. 카프카 클러스터에 대한 이해를 충분히 가진 상태에서 SaaS형 카프카를 운영한다면 더욱 효과적으로 활용할 수 있다. 굳이 카프카 클러스터를 직접 설치하고 운영하여 운영상 리소스를 낭비하기보다는 SaaS를 통해 자동화된 클러스터의 인프라를 사용하면서 얻는 이점도 있기 때문이다. 그러므로 SaaS형 카프카를 도입할 때는 현재 인력의 카프카 클러스터 운영에 대한 이해의 깊이, 클러스터 구축 비용, 추후 운영상 이슈 등을 다각도로 검토하는 것이 이후 안정적인 서비스를 운영하는 데에 밑바탕이 될 수 있다.

