

데이터 파이프라인 자동화

Red Hat OpenShift로 실시간 데이터 수집 처리 및 이벤트 생성

주요 내용: 이벤트 기반 데이터 처리

Red Hat OpenShift 및 Red Hat AMQ를 기반으로 완전히 통합된 클라우드 네이티브 데이터 수집 및 자동화된 스트림 프로세싱 솔루션 배포

Red Hat OpenShift Container Storage에서 이벤트를 수집하여 맞춤형 메타데이터를 추가하고, 더욱 풍부한 데이터를 생성하며 실행 가능성 개선

Red Hat OpenShift Serverless를 활용한 데이터 수집률을 바탕으로 애플리케이션 자동 확장 및 축소

인공지능 및 머신 러닝은 데이터 가치를 높이고 다양한 정보를 더욱 유용하고 실행 가능하게 만드는 능력으로 혁신을 주도했습니다. Apache Kafka 같은 분산형 이벤트 스트리밍 기술 및 서버리스(serverless) 모델은 AI/ML 기술을 더욱 강력하고 확장 가능하게 만듭니다. 조직으로 흘러들어가는 데이터의 양이 방대해짐에 따라 소프트웨어 정의 스토리지 기술은 데이터 처리, 파이프라인 애플리케이션, 인프라 자동화 및 스케일링을 위한 핵심 구성 요소로 떠올랐습니다.

Red Hat® OpenShift® Container Storage는 Red Hat OpenShift 및 Red Hat AMQ(Red Hat Integration에 포함)와 결합하여 데이터 파이프라인을 자동화하기 위한 강력한 기반과 아키텍처를 제공합니다. Red Hat Ceph® Storage의 스토리지 기반 이벤트 생성과 더불어 해당 기술은 데이터가 생성될 때부터 데이터를 집계, 수집, 준비, 관리할 수 있으므로 데이터 파이프라인을 자동화하고 필요에 따른 인프라 스케일링을 지원합니다.

확장 가능한 이벤트 기반 데이터 파이프라인 아키텍처

알림 기반 또는 이벤트 기반 아키텍처는 점점 데이터 처리 틀로서 중요성이 커지고 있습니다. 데이터 수집 서비스를 Red Hat OpenShift Serverless 기능에 연결하면 데이터 파이프라인은 자동으로 확장하여 요구 사항을 충족하고, 애플리케이션 인프라를 확장 또는 축소하여 조직의 특정 요구 사항에 따라 수신 데이터를 처리할 수 있습니다. 수집 알림은 수집 시 메타데이터 삽입을 사용자 정의할 수 있는 기능과 함께 새로운 데이터를 자동으로 데이터 파이프라인으로 이전합니다. 이러한 혁신을 통해 조직은 변화하는 정보 또는 고객 행동의 변화에 신속하게 대응할 수 있으며 다양한 활용 사례에 대한 최신 데이터를 기반으로 시의적절한 인사이트를 확보할 수 있습니다.

- 제조업체는 품질 보증에 대한 이상을 감지하고, 소매 제품의 재고 보충을 관리하며, 물류를 파악할 수 있습니다.
- 의료 시설은 AI 추론을 사용하여 감지 및 경고를 통해 이미지를 자동으로 처리할 수 있으며, 데이터를 익명화하여 연구자들이 프로세스를 개선하거나 치료 속도를 개선할 수 있습니다.
- 금융 서비스 기관은 결제를 가속화하거나 사기 방지 기능을 활용하여 고객에 대한 서비스를 향상할 수 있습니다.
- 공공 기관은 장비 유지 관리에서부터 위성 이미지의 지리 공간 변화에 대한 자동 감지에 이르는 광범위한 사례를 활용할 수 있습니다.

컨테이너 및 쿠버네티스 오케스트레이션은 하이브리드 클라우드에서 AI/ML을 배포하는 데 필수적입니다.¹ 그림 1은 통합된 Red Hat 기술을 활용하여 금융 서비스 애플리케이션을 위한 수집 데이터 파이프라인을 구축하는 방법을 보여줍니다. 이 솔루션은 Red Hat Ceph Storage RADOS 게이트웨이(RGW)의 오브젝트 버킷 알림, Red Hat AMQ 스트림에서 제공하는 데이터 스트리밍 서비스, Red Hat OpenShift의 서버리스 기능을 통합한 솔루션입니다.



오브젝트 버킷 알림

Red Hat Ceph Storage의 오브젝트 버킷에서 특정 이벤트가 트리거되면 알림이 HTTP, AMQP(Advanced Message Queueing Protocol) 및 Kafka 엔드포인트로 전송됩니다.

완전히 통합된 아키텍처를 활용하여 조직은 수신 데이터를 실시간으로 자동 처리하는 데이터 파이프라인을 구축할 수 있습니다. 사용자 정의 기능을 통해 데이터 익명화, 태그 지정, 메타데이터 강화 및 필요에 따라 정의하고 적용할 수 있는 기타 수정 작업을 수행할 수 있습니다. 오브젝트 버킷 알림은 파이프라인의 여러 단계에서 처리를 자동화하는 과정 전반에 사용할 수 있습니다.

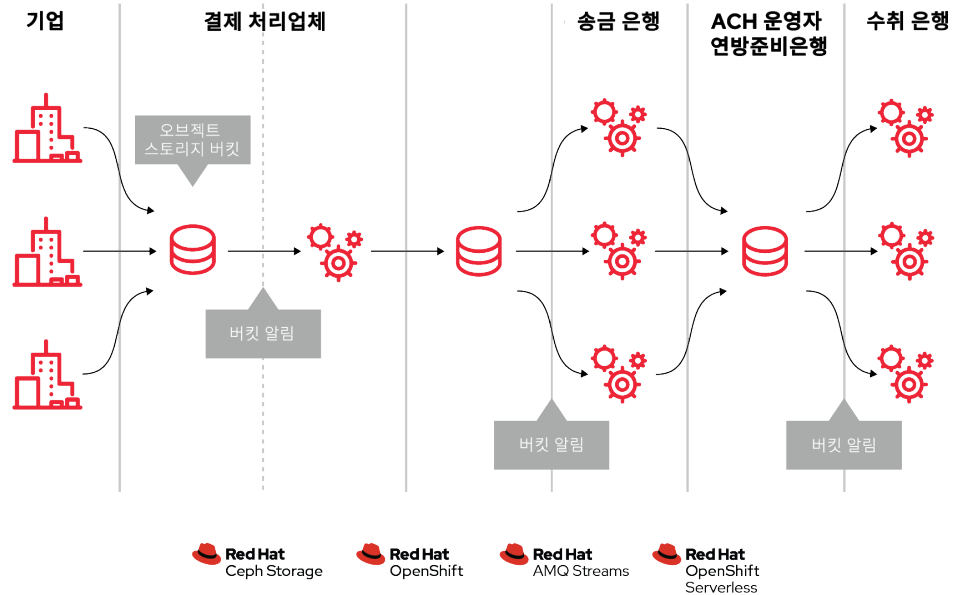


그림 1. 스트림 처리 이벤트를 추진하는 Red Hat Ceph Storage의 오브젝트 버킷 알림

Red Hat Ceph Storage(OpenShift Container Storage의 구성 요소)는 데이터 수집을 Red Hat OpenShift Serverless 기능에 연결하여 확장 가능한 이벤트 기반 아키텍처를 구축하도록 지원합니다.

이 클라우드 네이티브 아키텍처 접근 방식은 다음과 같은 실질적인 이점을 제공합니다.

- **더욱 빠른 실시간 처리.** 데이터 파이프라인을 구성하면 데이터가 수집되는 즉시 실시간으로 데이터를 처리할 수 있습니다. 이를 통해 머신 러닝에 사용할 수 있는 데이터셋 생성 프로세스의 속도를 대폭 높이고, 데이터 과학자들이 모델을 학습시키기 위한 최신 데이터를 확보할 수 있습니다.
- **확장 가능성 및 유연성.** 아키텍처의 각 구성 요소는 관리자 및 사용자의 요구 사항에 따라 사용자 정의하고 개별적으로 스케일링할 수 있습니다. Red Hat OpenShift Serverless 내에서 Knative 기능을 특정 데이터셋 및 조직 프로세스, 요구 사항 및 목표에 맞게 사용자 정의할 수 있습니다.
- **확장성.** 확장 가능한 솔루션 아키텍처는 재설계할 필요 없이 데이터 라이프사이클 프로세스에 더 많은 기능과 자동화를 제공해 줍니다. OpenShift Container Storage, Red Hat OpenShift Serverless 및 Red Hat AMQ 스트림을 결합한 솔루션으로 데이터 카탈로그 작성 및 감사 로깅과 같은 데이터 라이프사이클 관리의 기타 영역에도 활용할 수 있습니다.
- **효율성.** 수집 데이터 파이프라인 자동화를 통해 데이터 엔지니어는 머신 러닝을 위한 데이터 준비에 관련된 많은 운영 작업을 코드화하고 자동으로 수행할 수 있습니다. 설정이 완료되면 파이프라인은 대규모의 다양한 수집 데이터를 관리하도록 확장할 수 있으므로 엔지니어들은 가치가 높은 다른 활동에 집중할 수 있습니다.
- **데이터 라이프사이클 추적을 위한 기반.** 데이터 수집 시부터 실시간으로 처리할 수 있는 기능을 활용하면 데이터 소스에 대한 올바른 태그와 함께 소스 및 시간별 정보도 임베딩할 수 있습니다. 이러한 중요한 정보는 향후에 데이터 분류, 기능 엔지니어링, 카탈로그 작성 등과 같은 데이터 관리 및 머신 러닝 프로세스를 향상시키도록 활용할 수 있습니다.
- **Red Hat AMQ를 위한 복구 가능한 퍼시스턴트 스토리지.** OpenShift Container Storage는 Apache Zookeeper 및 Apache Kafka를 포함하여 Red Hat AMQ 구성 요소와 함께 필요로 하는 데이터 지속성을 고가용성 블록 스토리지 퍼시스턴트 볼륨 클레임(PVC)의 형태로 제공합니다.

OpenShift Container Storage는 Red Hat AMQ를 위한 고가용성 클라우드 네이티브 퍼시스턴트 스토리지를 제공합니다.

Red Hat AMQ를 활용하는 Red Hat OpenShift 기반의 Apache Kafka

Red Hat AMQ는 쿠버네티스(Red Hat OpenShift Container Platform)에서 구동하도록 설계된 Apache Kafka 프로젝트를 기반으로 하는 대규모로 확장 가능한 분산형 고성능 데이터 스트리밍 플랫폼입니다. 광범위한 활용 사례에서 사용되는 Apache Kafka는 웹 활동 추적, 이벤트, 메트릭, 로깅, 데이터 레이크로 데이터 스트리밍 등을 포함하는 애플리케이션에 적합합니다. Kafka는 실시간 데이터 스트림을 처리하여 대량의 데이터 수집 또는 실시간 분석을 수행하거나 둘 모두를 수행할 수 있습니다.

그림 2에서 볼 수 있듯이, Red Hat AMQ는 세 가지의 쿠버네티스 오퍼레이터를 제공합니다.

- **클러스터 오퍼레이터:** Red Hat OpenShift Container Platform 클러스터 내의 Apache Kafka 클러스터를 배포 및 관리합니다.
- **토픽 오퍼레이터:** Red Hat OpenShift Container Platform 클러스터와 함께 실행되는 Kafka 클러스터 내의 Kafka 토픽을 관리합니다.
- **사용자 오퍼레이터:** Red Hat OpenShift Container Platform 클러스터 내에서 실행되는 Kafka 클러스터 내의 Kafka 사용자를 관리합니다.

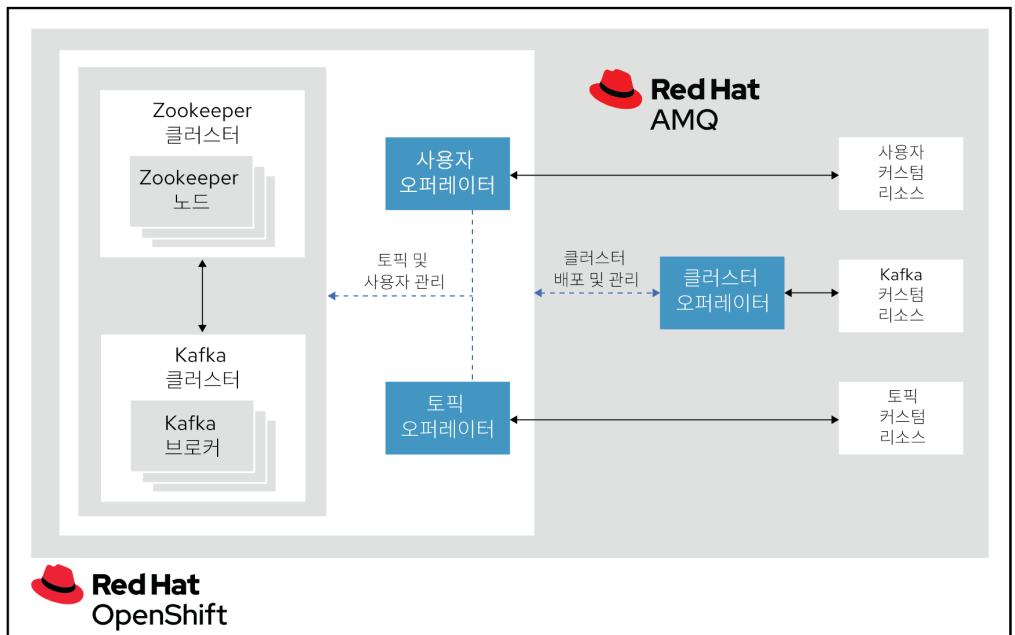


그림 2. Red Hat AMQ로 세 가지 쿠버네티스 오퍼레이터 구현

OpenShift 컨테이너 스토리지로 탄력적인 퍼시스턴트 스토리지 구축

OpenShift Container Storage는 통합 페타바이트급 스토리지 솔루션으로서, OpenShift 애플리케이션에 블록, 오브젝트, 파일 인터페이스를 제공합니다. 스토리지 플랫폼은 Red Hat OpenShift Container Platform과 함께 작동하도록 설계되고 긴밀하게 통합되었습니다. OpenShift Container Storage PVC는 Apache Kafka 토픽을 트리거하는 오브젝트 버킷 알림 생성 이벤트로 Kafka와 Zookeeper 클러스터를 모두 지원합니다(그림 3).

RED HAT 정보

Red Hat은 세계적인 엔터프라이즈 오픈소스 솔루션 공급업체로서 커뮤니티 기반 접근 방식을 통해 신뢰도 높은 고성능 Linux, 하이브리드 클라우드, 컨테이너, 쿠버네티스 기술을 제공합니다. 또한 고객으로 하여금 신규 및 기존 IT 애플리케이션을 통합하고, 클라우드 네이티브 애플리케이션을 개발하며, 업계를 선도하는 Red Hat의 운영 체제를 기반으로 표준화하는 동시에 복잡한 환경의 자동화, 보안 및 관리를 실현할 수 있도록 지원합니다. Red Hat은 전 세계 고객에게 높은 수준의 지원과 교육 및 컨설팅 서비스를 제공하여 권위있는 어워드를 다수 수상한 바 있으며, Fortune 선정 500대 기업의 신뢰를 받는 어드바이저로 인정받고 있습니다. 또한 기업, 파트너, 오픈소스 커뮤니티의 전략적인 파트너로서 고객들이 디지털 미래에 대비할 수 있도록 지원하고 있습니다.

한국레드햇 홈페이지

<https://www.redhat.com/ko>

또한, 전형적인 퍼블릭 클라우드 블록 스토리지 오퍼링과 달리 OpenShift Container Storage는 여러 퍼블릭 클라우드 가용 영역 전반에 최고 수준의 복구 능력을 제공합니다. Apache Kafka는 데이터 복구 기능을 기본 제공하지만 OpenShift Container Storage 퍼시스턴트 볼륨 클레임은 쿠버네티스와 같은 클라우드 네이티브 환경에 배포할 때 필수적인 역할을 합니다. Kafka 메시지는 또한 데이터 분석 및 장기간 보존을 위해 확장 가능한 분산형 Amazon Simple Storage Service(S3)에 호환되는 오브젝트 스토리지로 이전할 수 있습니다.

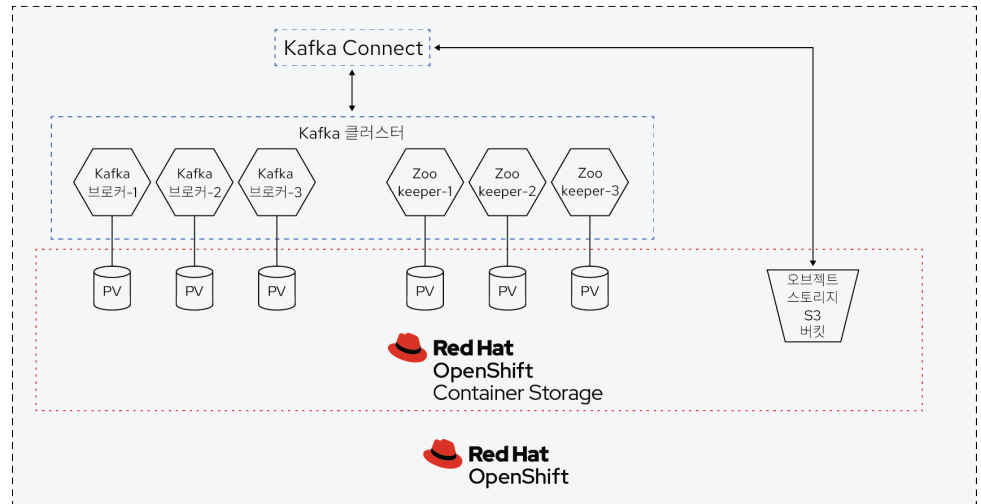


그림 3. Red Hat AMQ 스트림의 스토리지 활용 사례

결론

Red Hat은 자동화된 스트림 프로세싱 솔루션을 배포하기 위해 완전히 통합된 클라우드 네이티브 솔루션을 제공합니다. Red Hat Ceph Storage 및 OpenShift Container Storage를 통해 방대한 양의 데이터를 수집하고 실시간으로 처리할 수 있습니다. 오브젝트 버킷 이벤트 알람이 제공되어 데이터 엔지니어와 데이터 과학자는 스트림 처리를 자동화하고 가장 최신의 데이터로 인사이트를 도출함으로써 사용자 정의 메타데이터를 삽입하여 AI/ML 애플리케이션에 대한 데이터 가치를 높일 수 있습니다. 클라우드 제공업체 가용 영역 전반에 가용성이 지원되므로 가장 까다로운 엔터프라이즈 환경에 준비된 강력한 솔루션입니다.



www.facebook.com/redhatkorea
 구매문의 080 708 0880
buy-kr@redhat.com