



MLflow(최소) NetApp Solutions

NetApp
August 14, 2024

목차

MLflow(최소	1
MLflow 전개	1
NetApp 및 MLflow를 통한 데이터 세트 대 모델 추적성	2

MLflow(최소

MLflow 전개

이 섹션에서는 Kubernetes 클러스터에 MLflow를 배포하기 위해 완료해야 하는 작업에 대해 설명합니다.



Kubernetes 이외의 플랫폼에 MLflow를 구축할 수 있습니다. Kubernetes 이외의 플랫폼에 MLflow를 구축하는 것은 이 솔루션의 범위를 벗어납니다.

필수 구성 요소

이 섹션에 요약된 배포 연습을 수행하기 전에 이미 다음 작업을 수행했다고 가정합니다.

1. Kubernetes 클러스터 작업이 이미 진행 중입니다.
2. Kubernetes 클러스터에 NetApp Astra Trident를 이미 설치하고 구성했습니다. Astra Trident에 대한 자세한 내용은 ["Astra Trident 문서"](#) 참조하십시오.

제어 장치를 설치합니다

MLflow는 Kubernetes의 인기 패키지 매니저인 Helm을 사용하여 구축됩니다. MLflow를 배포하기 전에 Kubernetes 제어 노드에 Helm을 설치해야 합니다. Helm을 설치하려면 공식 Helm 설명서의 ["설치 지침"](#) 따르십시오.

기본 Kubernetes StorageClass를 설정합니다

MLflow를 구축하기 전에 Kubernetes 클러스터 내에 기본 StorageClass를 지정해야 합니다. 클러스터 내에서 기본 StorageClass를 지정하려면 섹션에 설명된 지침을 ["Kubeflow 구축"](#) 따릅니다. 클러스터 내에서 기본 StorageClass를 이미 지정한 경우에는 이 단계를 건너뛸 수 있습니다.

MLflow를 배포합니다

필수 구성 요소가 충족되면 Helm Chart를 사용하여 MLflow 배포를 시작할 수 있습니다.

MLflow Helm Chart Deployment를 구성합니다.

Helm 차트를 사용하여 MLflow를 배포하기 전에 NetApp Trident 저장소 클래스를 사용하도록 배포를 구성하고 * config.yaml * 파일을 사용하여 필요에 맞게 다른 매개 변수를 변경할 수 있습니다. config.yaml* 파일의 예는 다음 웹 사이트에서 찾을 수 있습니다. <https://github.com/bitnami/charts/blob/main/bitnami/mlflow/values.yaml>



config.yaml 파일의 * global.defaultStorageClass* 매개 변수 아래에서 Trident storageClass 를 설정할 수 있습니다(예: storageClass:"ONTAP-FlexVol").

제어 차트 설치

Helm 차트는 다음 명령을 사용하여 MLflow용 custom * config.yaml * 파일과 함께 설치할 수 있습니다.

```
helm install oci://registry-1.docker.io/bitnamicharts/mlflow -f
config.yaml --generate-name --namespace jupyterhub
```



이 명령은 제공된 * config.yaml * 파일을 통해 사용자 지정 구성으로 Kubernetes 클러스터의 MLflow를 배포합니다. MLflow는 지정된 네임스페이스에 배포되며 릴리스의 경우 Kubernetes를 통해 임의 릴리스 이름이 지정됩니다.

배포를 확인합니다

제어 차트 배포가 완료된 후 다음을 사용하여 서비스에 액세스할 수 있는지 확인할 수 있습니다.

```
kubectl get service -n jupyterhub
```



jupyterhub*를 배포 중에 사용한 네임스페이스로 바꿉니다.

다음과 같은 서비스가 표시됩니다.

NAME	TYPE	CLUSTER-IP	EXTERNAL-IP
mlflow-1719843029-minio	ClusterIP	10.233.22.4	<none>
80/TCP, 9001/TCP			
mlflow-1719843029-postgresql	ClusterIP	10.233.5.141	<none>
5432/TCP			
mlflow-1719843029-postgresql-hl	ClusterIP	None	<none>
5432/TCP			
mlflow-1719843029-tracking	NodePort	10.233.2.158	<none>
30002:30002/TCP			



NodePort 서비스를 사용하여 포트 30002의 MLflow에 액세스하기 위해 config.yaml 파일을 편집했습니다.

MLflow에 액세스합니다

MLflow와 관련된 모든 서비스가 가동되어 실행되면 지정된 NodePort 또는 부하 분산 장치 IP 주소(예:)를 사용하여 액세스할 수 있습니다 <http://10.61.181.109:30002>.

NetApp 및 MLflow를 통한 데이터 세트 대 모델 추적성

은 "Kubernetes용 NetApp DataOps 툴킷" MLflow의 실험 추적 기능과 함께 데이터 세트 대 모델 또는 작업 공간 대 모델 추적 기능을 구현할 수 있습니다.

데이터 세트-모델 또는 작업 공간-모델 추적 기능을 구현하려면 다음 코드 조각과 같이 교육 실행의 일부로 DataOps Toolkit을 사용하여 데이터 세트 또는 작업 공간 볼륨의 스냅샷을 생성하기만 하면 됩니다. 이 코드는 데이터 볼륨 이름

및 스냅샷 이름을 MLflow Experiment Tracking Server에 로깅하는 특정 교육 실행과 관련된 태그로 저장합니다.

```
...
from netapp_dataops.k8s import create_volume_snapshot

with mlflow.start_run() :
    ...

    namespace = "my_namespace" # Kubernetes namespace in which dataset
    volume PVC resides
    dataset_volume_name = "project1" # Name of PVC corresponding to
    dataset volume
    snapshot_name = "run1" # Name to assign to your new snapshot

    # Create snapshot
    create_volume_snapshot(
        namespace=namespace,
        pvc_name=dataset_volume_name,
        snapshot_name=snapshot_name,
        printOutput=True
    )

    # Log data volume name and snapshot name as "tags"
    # associated with this training run in mlflow.
    mlflow.set_tag("data_volume_name", dataset_volume_name)
    mlflow.set_tag("snapshot_name", snapshot_name)

    ...
```

저작권 정보

Copyright © 2024 NetApp, Inc. All Rights Reserved. 미국에서 인쇄된 본 문서의 어떠한 부분도 저작권 소유자의 사전 서면 승인 없이는 어떠한 형식이나 수단(복사, 녹음, 녹화 또는 전자 검색 시스템에 저장하는 것을 비롯한 그래픽, 전자적 또는 기계적 방법)으로도 복제될 수 없습니다.

NetApp이 저작권을 가진 자료에 있는 소프트웨어에는 아래의 라이선스와 고지사항이 적용됩니다.

본 소프트웨어는 NetApp에 의해 '있는 그대로' 제공되며 상품성 및 특정 목적에의 적합성에 대한 명시적 또는 묵시적 보증을 포함하여(이에 제한되지 않음) 어떠한 보증도 하지 않습니다. NetApp은 대체품 또는 대체 서비스의 조달, 사용 불능, 데이터 손실, 이익 손실, 영업 중단을 포함하여(이에 국한되지 않음), 이 소프트웨어의 사용으로 인해 발생하는 모든 직접 및 간접 손해, 우발적 손해, 특별 손해, 징벌적 손해, 결과적 손해의 발생에 대하여 그 발생 이유, 책임론, 계약 여부, 엄격한 책임, 불법 행위(과실 또는 그렇지 않은 경우)와 관계없이 어떠한 책임도 지지 않으며, 이와 같은 손실의 발생 가능성이 통지되었다 하더라도 마찬가지입니다.

NetApp은 본 문서에 설명된 제품을 언제든지 예고 없이 변경할 권리를 보유합니다. NetApp은 NetApp의 명시적인 서면 동의를 받은 경우를 제외하고 본 문서에 설명된 제품을 사용하여 발생하는 어떠한 문제에도 책임을 지지 않습니다. 본 제품의 사용 또는 구매의 경우 NetApp에서는 어떠한 특허권, 상표권 또는 기타 지적 재산권이 적용되는 라이선스도 제공하지 않습니다.

본 설명서에 설명된 제품은 하나 이상의 미국 특허, 해외 특허 또는 출원 중인 특허로 보호됩니다.

제한적 권리 표시: 정부에 의한 사용, 복제 또는 공개에는 DFARS 252.227-7013(2014년 2월) 및 FAR 52.227-19(2007년 12월)의 기술 데이터-비상업적 품목에 대한 권리(Rights in Technical Data -Noncommercial Items) 조항의 하위 조항 (b)(3)에 설명된 제한사항이 적용됩니다.

여기에 포함된 데이터는 상업용 제품 및/또는 상업용 서비스(FAR 2.101에 정의)에 해당하며 NetApp, Inc.의 독점 자산입니다. 본 계약에 따라 제공되는 모든 NetApp 기술 데이터 및 컴퓨터 소프트웨어는 본질적으로 상업용이며 개인 비용만으로 개발되었습니다. 미국 정부는 데이터가 제공된 미국 계약과 관련하여 해당 계약을 지원하는 데에만 데이터에 대한 전 세계적으로 비독점적이고 양도할 수 없으며 재사용이 불가능하며 취소 불가능한 라이선스를 제한적으로 가집니다. 여기에 제공된 경우를 제외하고 NetApp, Inc.의 사전 서면 승인 없이는 이 데이터를 사용, 공개, 재생산, 수정, 수행 또는 표시할 수 없습니다. 미국 국방부에 대한 정부 라이선스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

상표 정보

NETAPP, NETAPP 로고 및 <http://www.netapp.com/TM>에 나열된 마크는 NetApp, Inc.의 상표입니다. 기타 회사 및 제품 이름은 해당 소유자의 상표일 수 있습니다.