



Kubeflow

NetApp Solutions

NetApp
May 10, 2024

This PDF was generated from https://docs.netapp.com/zh-tw/netapp-solutions/ai/aicp_kubeflow_deployment_overview.html on May 10, 2024. Always check docs.netapp.com for the latest.

目錄

Kubeflow	1
Kubeflow部署	1
Kubeflow作業與工作範例	2

Kubeflow

Kubeflow部署

本節說明在Kubernetes叢集中部署Kubeflow時、必須完成的工作。

先決條件

在您執行本節所述的部署練習之前、我們假設您已經執行下列工作：

1. 您已經有一個運作中的 Kubernetes 叢集、而且您正在執行的 Kubernetes 版本受到您要部署的 Kubeflow 版本的支援。如需支援的 Kubernetes 版本清單、請參閱中 Kubeflow 版本的相依性 ["官方Kubeflow文件"](#)。
2. 您已在 Kubernetes 叢集中安裝並設定 NetApp Astra Trident 。如需 Astra Trident 的詳細資訊、請參閱 ["Astra Trident文件"](#)。

設定預設Kubernetes StorageClass

在您部署 Kubeflow 之前、建議您在 Kubernetes 叢集中指定預設 StorageClass 。Kubeflow 部署程序可能會嘗試使用預設 StorageClass 來配置新的持續磁碟區。如果未將 StorageClass 指定為預設 StorageClass 、則部署可能會失敗。若要在叢集內指定預設StorageClass、請從部署跨接主機執行下列工作。如果您已在叢集內指定預設StorageClass、則可以跳過此步驟。

1. 將現有的其中一個StorageClass指定為預設StorageClass。後面的命令範例顯示指定的 StorageClass 名稱 `ontap-ai-flexvols-retain` 為預設 StorageClass 。



「ontap-non-flexgroup」Trident後端類型的最小PVC尺寸相當大。根據預設、Kubeflow會嘗試配置大小只有幾GB的PVCS。因此、您不應將使用「ONTAP-NAAS-Flexgroup」後端類型的StorageClass指定為Kubeflow部署的預設StorageClass。

```
$ kubectl get sc
NAME                                PROVISIONER                AGE
ontap-ai-flexgroups-retain         csi.trident.netapp.io     25h
ontap-ai-flexgroups-retain-iface1  csi.trident.netapp.io     25h
ontap-ai-flexgroups-retain-iface2  csi.trident.netapp.io     25h
ontap-ai-flexvols-retain           csi.trident.netapp.io     3s
$ kubectl patch storageclass ontap-ai-flexvols-retain -p '{"metadata": {"annotations":{"storageclass.kubernetes.io/is-default-class":"true"}}}'
storageclass.storage.k8s.io/ontap-ai-flexvols-retain patched
$ kubectl get sc
NAME                                PROVISIONER                AGE
ontap-ai-flexgroups-retain         csi.trident.netapp.io     25h
ontap-ai-flexgroups-retain-iface1  csi.trident.netapp.io     25h
ontap-ai-flexgroups-retain-iface2  csi.trident.netapp.io     25h
ontap-ai-flexvols-retain (default) csi.trident.netapp.io     54s
```

Kubeflow 部署選項

部署 Kubeflow 有許多不同的選項。請參閱 ["官方Kubeflow文件"](#) 如需部署選項清單、請選擇最適合您需求的選項。

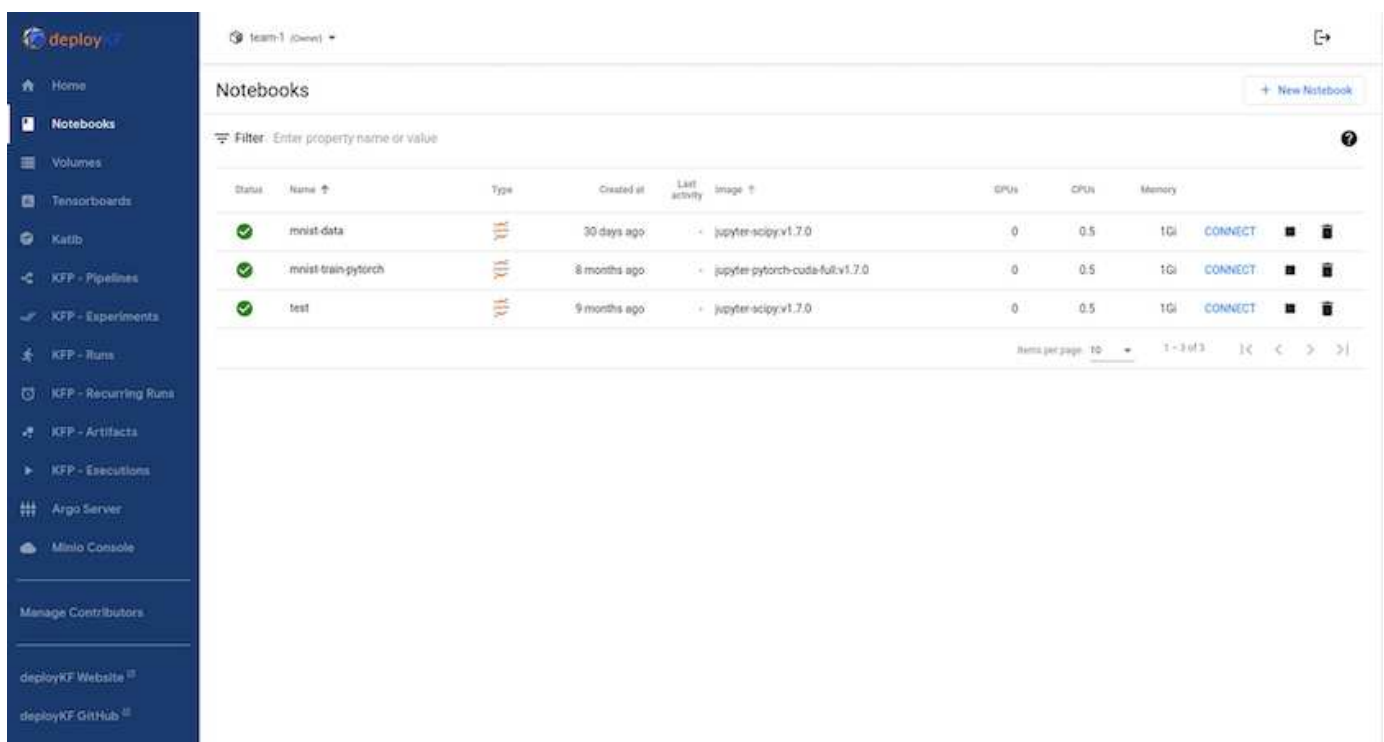


為了驗證目的、我們使用部署 Kubeflow 1.7 ["deployKF"](#) 0.1.1。

Kubeflow作業與工作範例

為資料科學家或開發人員提供Jupyter筆記型電腦工作區

Kubeflow能夠快速配置新的Jupyter筆記型電腦伺服器、做為資料科學家工作空間。如需Kubeflow內容中Jupyter Notebooks的詳細資訊、請參閱 ["官方Kubeflow文件"](#)。



使用 NetApp DataOps Toolkit 搭配 Kubeflow

◦ ["適用於Kubernetes的NetApp Data科學工具套件"](#) 可與Kubeflow搭配使用。搭配Kubeflow使用NetApp Data科學工具套件可提供下列效益：

- 資料科學家可以直接從 Jupyter Notebook 執行進階的 NetApp 資料管理作業、例如建立快照和複本。
- 進階 NetApp 資料管理作業（例如建立快照和複本）可以使用 Kubeflow Pipelines 架構整合到自動化工作流程中。

請參閱 ["Kubeflow範例"](#) NetApp Data科學工具套件GitHub儲存庫中的一節、詳細說明如何搭配Kubeflow使用此工具組。

範例工作流程：使用 Kubeflow 和 NetApp DataOps Toolkit 訓練影像辨識模型

本節說明使用 Kubeflow 和 NetApp DataOps 工具套件來訓練及部署用於影像辨識的神經網路所涉及的步驟。這是一個範例、旨在展示整合 NetApp 儲存設備的訓練工作。

先決條件

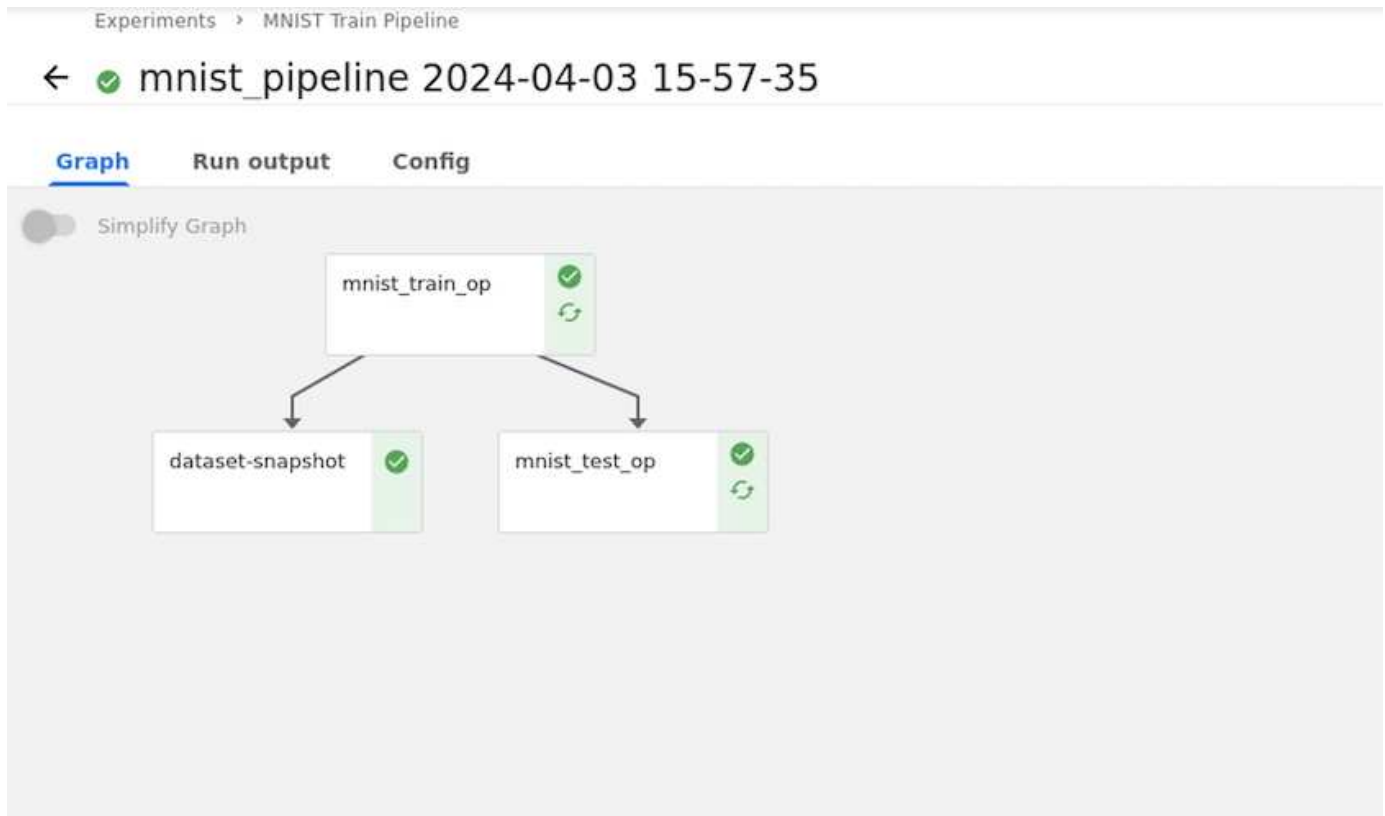
使用必要的組態建立一個 Dockerfile、以用於 Kubeflow 管道中的訓練和測試步驟。以下是一個 Dockerfile 的範例：

```
FROM pytorch/pytorch:latest
RUN pip install torchvision numpy scikit-learn matplotlib tensorboard
WORKDIR /app
COPY . /app
COPY train_mnist.py /app/train_mnist.py
CMD ["python", "train_mnist.py"]
```

視您的需求而定、安裝執行程式所需的所有必要程式庫和套件。在您訓練機器學習模式之前、我們假設您已經有一個有效的 Kubeflow 部署。

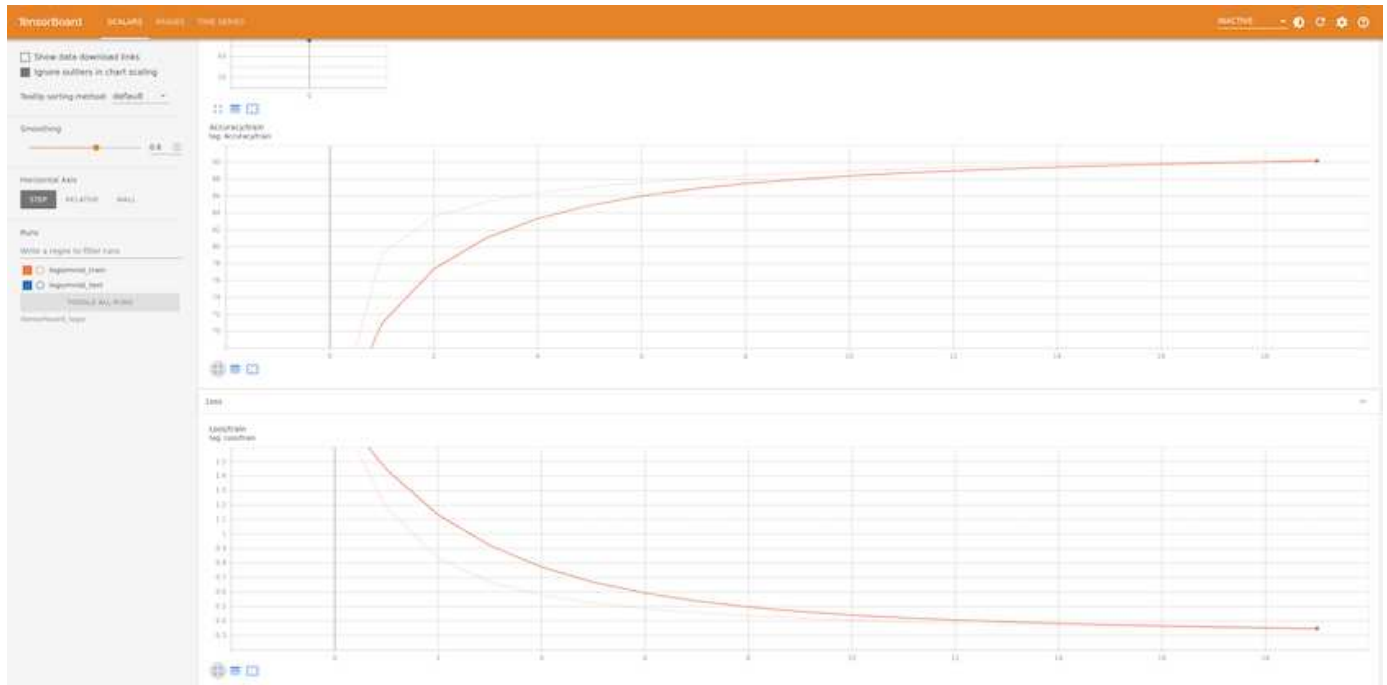
使用 PyTorch 和 Kubeflow Pipelines 訓練一部小型 NN on MIST Data

我們使用小型神經網路的範例來訓練 MIST 資料。MIST 資料集由 0-9 位數的手寫影像組成。影像大小為 28x28 像素。資料集分為 60、000 個訓練影像和 10、000 個驗證影像。此實驗使用的神經網路是 2 層饋送網路。訓練是使用 Kubeflow Pipelines 執行。請參閱文件 ["請按這裡"](#) 以取得更多資訊。我們的 Kubeflow 管道整合了「先決條件」區段中的泊塢視窗影像。



使用 Tensorboard 視覺化結果

一旦模型經過訓練、我們就可以使用 Tensorboard 將結果視覺化。"Tensorboard" 可在 Kubeflow 儀表板上作為功能使用。您可以為工作建立自訂的浮動期管理板。以下範例顯示訓練準確度與的繪圖時期和訓練損失與時期數。



使用 Katib 嘗試使用 Hyperparameters

"Katib" 是 Kubeflow 內的一項工具、可用於實驗模型超參數。若要建立實驗、請先定義所需的指標 / 目標。這通常是測試準確度。一旦定義了度量、請選擇您想要使用的超參數（最佳化器 / 學習率 / 層數）。Katib 會使用使用者定義的值執行超參數掃描、以找出符合所需度量的最佳參數組合。您可以在 UI 的每個區段中定義這些參數。或者、您也可以使用必要的規格來定義 *YAML* 檔案。以下是 Katib 實驗的圖例：

The screenshot shows the 'Experiment details' page for an experiment named 'Validation-accuracy'. The interface includes a sidebar with navigation options like Home, Notebooks, Volumes, Tensorboards, and Katib. The main content area displays the following information:

- Objective:**
 - Name: Validation-accuracy
 - Type: maximize
 - Goal: 0.9
 - Additional metrics: Train-accuracy
- Trials:**
 - Max failed trials: 3
 - Max trials: 12
 - Parallel trials: 3
- Parameters:**
 - lr: Parameter type: double, Min: 0.01, Max: 0.03
 - num-layers: Parameter type: int, Min: 1, Max: 64
 - optimizer: Parameter type: categorical, sgd, adam, trl
- Algorithm:**
 - Name: grid
- Metrics collector:**
 - Collector type: File

The screenshot shows the 'Experiment details' page for an experiment named 'mnist-gytorch'. The interface includes a sidebar with navigation options like Home, Notebooks, Volumes, Tensorboards, and Katib. The main content area displays the following information:

- Message:** Couldn't find any successful Trial.
- OVERVIEW TRIALS DETAILS YAML** (Navigation tabs)
- Name:** mnist-gytorch
- Status:** Experiment is running
- Best trial:** No optimal trial yet
- Best trial's params:** No optimal trial yet
- Best trial performance:**
- User defined goal:** Validation-accuracy > 0.9
- Running trials:** 3
- Failed trials:** 0
- Succeeded trials:** 0
- Experiment Conditions:**
- Filter:** Enter property name or value

使用 NetApp Snapshot 來儲存資料以進行追蹤

在模型訓練期間、我們可能想要儲存訓練資料集的快照、以便追蹤。為達成此目標、我們可以將快照步驟新增至管道、如下所示。若要建立快照、我們可以使用 ["適用於Kubernetes的NetApp DataOps工具套件"](#)。

```

@dsl.pipeline(
    name = 'MNIST Classification Pipeline',
    description = 'Train a simple NN for classification'
)
def mnist_pipeline():
    mnist_train_task = mnist_train_op()
    mnist_train_task.apply(
        kfp.onprem.mount_pvc('mnist-data', 'mnist-data-vol', '/mnt/data/')
    )

    mnist_test_task = mnist_test_op()
    mnist_test_task.apply(
        kfp.onprem.mount_pvc('mnist-data', 'mnist-data-vol', '/mnt/data/')
    )

    volume_snapshot_name = "mnist-pytorch-snapshot"
    dataset_snapshot = dsl.ContainerOp(
        name="dataset-snapshot",
        image="python:3.9",
        command=["/bin/bash", "-c"],
        arguments=["\
            python3 -m pip install netapp-dataops-k8s && \
            echo "" + volume_snapshot_name + "" > /volume_snapshot_name.txt && \
            netapp_dataops_k8s_cli.py create volume-snapshot --pvc-name=" + "mnist-data" + " --snapshot-name=" + str(volume_snapshot_name) + " --namespace={workload_namespace}"],
        file_outputs={"volume_snapshot_name": "/volume_snapshot_name.txt"}
    )
    mnist_test_task.after(mnist_train_task)
    dataset_snapshot.after(mnist_train_task)

```

請參閱 ["NetApp DataOps Toolkit Kubeflow 範例"](#) 以取得更多資訊。

版權資訊

Copyright © 2024 NetApp, Inc. 版權所有。台灣印製。非經版權所有人事先書面同意，不得將本受版權保護文件的任何部分以任何形式或任何方法（圖形、電子或機械）重製，包括影印、錄影、錄音或儲存至電子檢索系統中。

由 NetApp 版權資料衍伸之軟體必須遵守下列授權和免責聲明：

此軟體以 NETAPP「原樣」提供，不含任何明示或暗示的擔保，包括但不限於有關適售性或特定目的適用性之擔保，特此聲明。於任何情況下，就任何已造成或基於任何理論上責任之直接性、間接性、附隨性、特殊性、懲罰性或衍生性損害（包括但不限於替代商品或服務之採購；使用、資料或利潤上的損失；或企業營運中斷），無論是在使用此軟體時以任何方式所產生的契約、嚴格責任或侵權行為（包括疏忽或其他）等方面，NetApp 概不負責，即使已被告知有前述損害存在之可能性亦然。

NetApp 保留隨時變更本文所述之任何產品的權利，恕不另行通知。NetApp 不承擔因使用本文所述之產品而產生的責任或義務，除非明確經過 NetApp 書面同意。使用或購買此產品並不會在依據任何專利權、商標權或任何其他 NetApp 智慧財產權的情況下轉讓授權。

本手冊所述之產品受到一項（含）以上的美國專利、國外專利或申請中專利所保障。

有限權利說明：政府機關的使用、複製或公開揭露須受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中的「技術資料權利 - 非商業項目」條款 (b)(3) 小段所述之限制。

此處所含屬於商業產品和 / 或商業服務（如 FAR 2.101 所定義）的資料均為 NetApp, Inc. 所有。根據本協議提供的所有 NetApp 技術資料和電腦軟體皆屬於商業性質，並且完全由私人出資開發。美國政府對於該資料具有非專屬、非轉讓、非轉授權、全球性、有限且不可撤銷的使用權限，僅限於美國政府為傳輸此資料所訂合約所允許之範圍，並基於履行該合約之目的方可使用。除非本文另有規定，否則未經 NetApp Inc. 事前書面許可，不得逕行使用、揭露、重製、修改、履行或展示該資料。美國政府授予國防部之許可權利，僅適用於 DFARS 條款 252.227-7015(b)（2014 年 2 月）所述權利。

商標資訊

NETAPP、NETAPP 標誌及 <http://www.netapp.com/TM> 所列之標章均為 NetApp, Inc. 的商標。文中所涉及的所有其他公司或產品名稱，均為其各自所有者的商標，不得侵犯。