



クビフロー

NetApp Solutions

NetApp
May 10, 2024

目次

| | |
|---------------------------|---|
| クビフロー | 1 |
| Kubeflow の導入 | 1 |
| 例： Kubeflow の操作とタスク | 2 |

クビフロー

Kubeflow の導入

このセクションでは、Kubernetes クラスタに Kubeflow を導入するために実行する必要のあるタスクについて説明します。

前提条件

このセクションで説明する導入の演習を行う前に、次の作業をすでに実行していることを前提としています。

1. すでに稼働中の Kubernetes クラスタがあり、導入する Kubeflow バージョンでサポートされている Kubernetes のバージョンを実行している。サポートされている Kubernetes バージョンのリストについては、"[Kubeflow の公式ドキュメント](#)"。
2. Kubernetes クラスタに NetApp Astra Trident をインストールして設定しておきます。Astra Trident の詳細については、"[Astra Trident のドキュメント](#)"。

デフォルトの **Kubernetes StorageClass** を設定します

Kubeflow を導入する前に、Kubernetes クラスタ内でデフォルトの StorageClass を指定することを推奨します。Kubeflow の導入プロセスで、デフォルトの StorageClass を使用して新しい永続ボリュームのプロビジョニングが試行されることがあります。StorageClass がデフォルトの StorageClass として指定されていない場合、導入に失敗する可能性があります。クラスタ内のデフォルトの StorageClass を指定するには、導入ジャンプホストから次のタスクを実行します。クラスタ内ですでにデフォルトの StorageClass を指定している場合は、この手順を省略できます。

1. 既存のストレージクラスの 1 つをデフォルトのストレージクラスとして指定します。次のコマンド例は、`ontap-ai-flexvols-retain` という名前の StorageClass の指定を示しています。`ontap-ai-flexvols-retain` をデフォルトの StorageClass として設定します。



「ONTAP-NAS-flexgroup」の Trident バックエンドタイプには、かなり大きな最小 PVC サイズがあります。デフォルトでは、Kubeflow はサイズが数 GB しかない PVC のプロビジョニングを試みます。したがって、Kubeflow の導入の目的で、「ONTAP-NAS-flexgroup」バックエンドタイプをデフォルトの StorageClass として使用する StorageClass を指定しないでください。

```

$ 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の導入オプション

Kubeflowを導入するには、さまざまなオプションがあります。を参照してください ["Kubeflow の公式ドキュメント"](#) を参照し、ニーズに最適な導入オプションを選択してください。

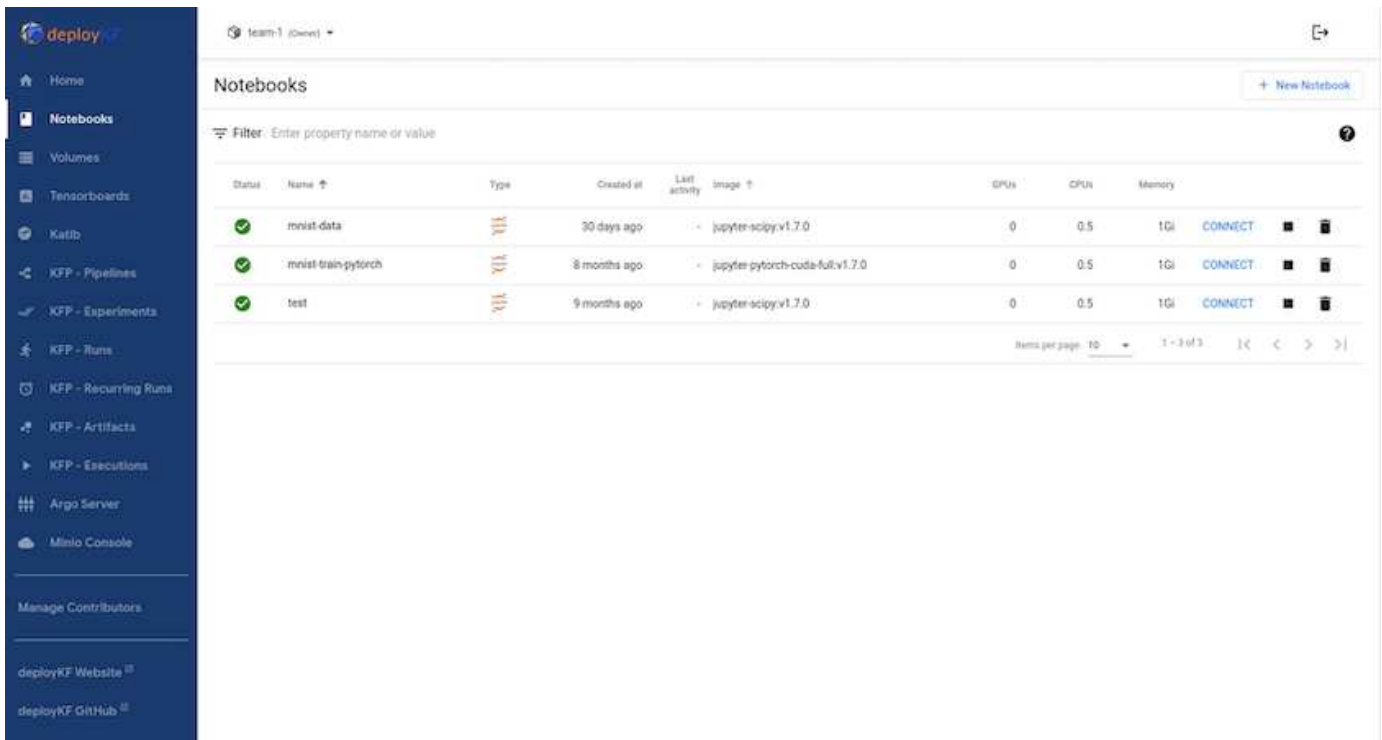


検証のために、次のツールを使用してKubeflow 1.7を導入しました。 ["deployKF" 0.1.1](#)。

例： Kubeflow の操作とタスク

データサイエンティストまたは開発者向けに **Jupyter Notebook Workspace** をプロビジョニングします 使用

Kubeflow は、データサイエンティストのワークスペースとして機能する、新しい Jupyter Notebook サーバの迅速なプロビジョニングを可能にします。Kubeflow コンテキスト内の Jupyter Notebook の詳細については、を参照してください ["Kubeflow の公式ドキュメント"](#)。



KubeflowでのNetApp DataOps Toolkitの使用

。"NetApp Data Science Toolkit for Kubernetes" Kubeflow と組み合わせて使用できます。NetApp Data Science Toolkit と Kubeflow を使用すると、次のようなメリットがあります。

- データサイエンティストは、NetAppの高度なデータ管理操作（スナップショットやクローンの作成など）を、Jupyterノートブックから直接実行できます。
- Kubeflow Pipelinesフレームワークを使用すると、スナップショットやクローンの作成など、NetAppの高度なデータ管理操作を自動化されたワークフローに組み込むことができます。

を参照してください "[Kubeflow の例](#)" ツールキットと Kubeflow を使用する場合の詳細については、NetApp Data Science Toolkit GitHub リポジトリのセクションを参照してください。

ワークフロー例- KubeflowとNetApp DataOps Toolkitを使用した画像認識モデルのトレーニング

このセクションでは、KubeflowとNetApp DataOps Toolkitを使用した画像認識のためのニューラルネットワークのトレーニングと導入に関連する手順について説明します。これは、NetAppストレージを組み込んだトレーニングジョブの例を示すことを目的としています。

前提条件

Kubeflowパイプライン内のトレーニングおよびテストステップに使用するために必要な構成を含むDockerfileを作成します。

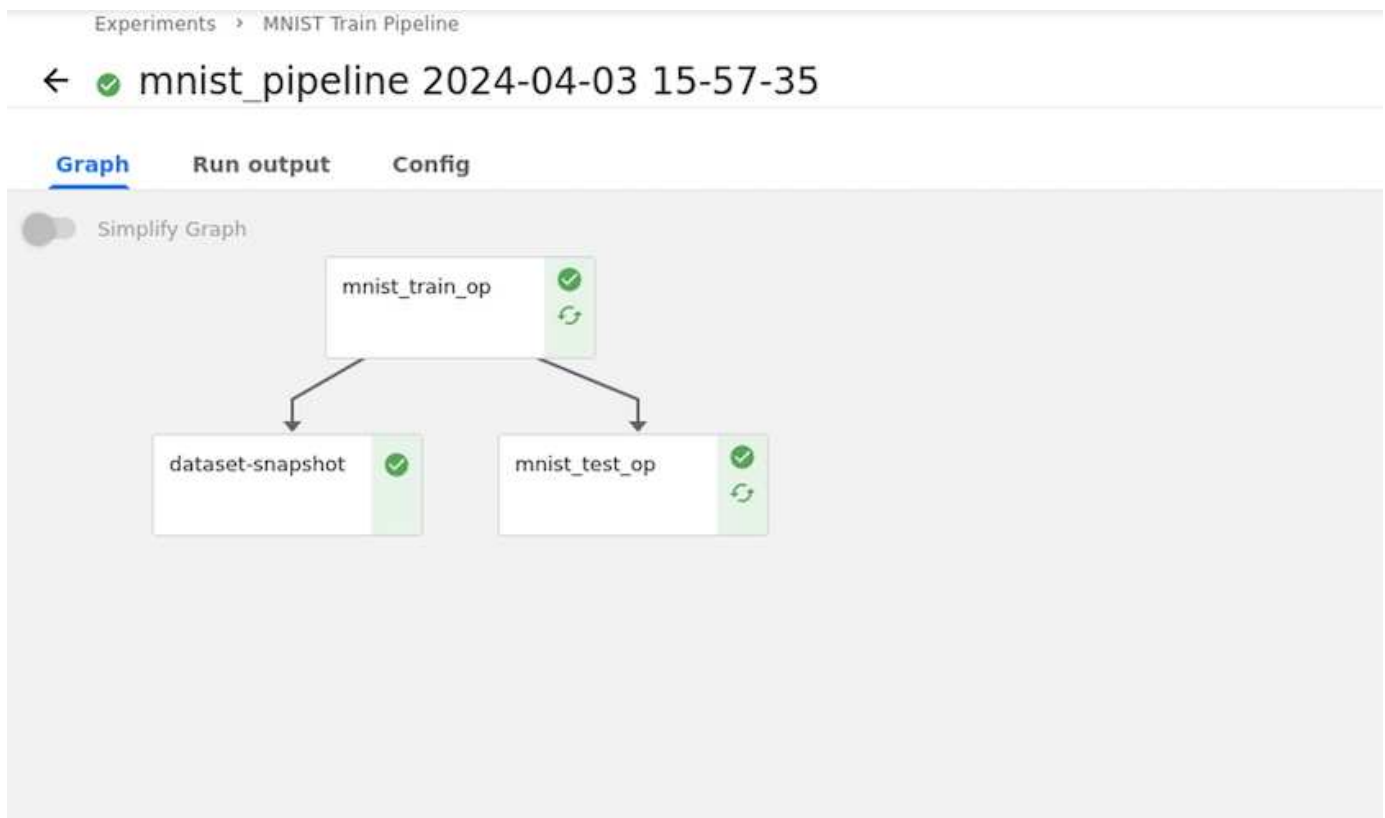
Dockerfileの例を次に示します。

```
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"]
```

必要に応じて、プログラムの実行に必要なすべてのライブラリとパッケージをインストールします。機械学習モデルをトレーニングする前に、すでに稼働しているKubeflowデプロイメントがあることを前提としています。

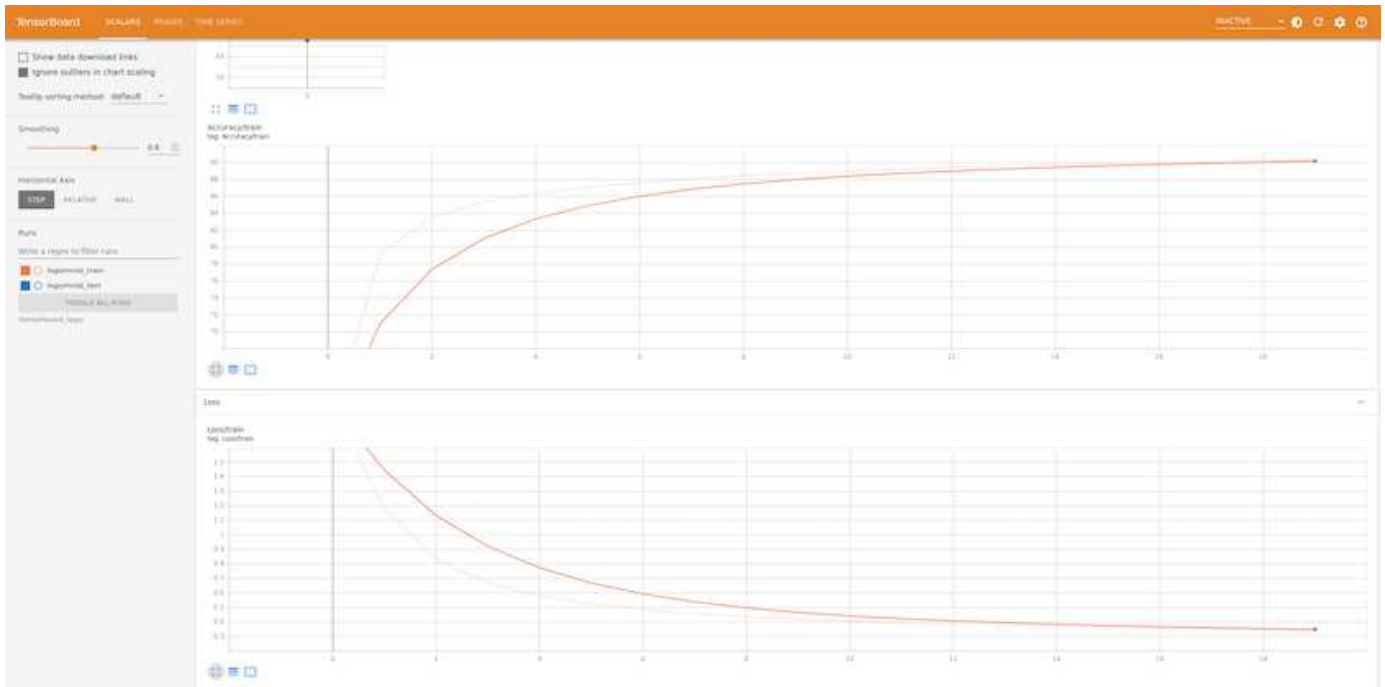
PyTorchとKubeflowパイプラインを使用して、小規模NNをMNISTデータでトレーニング

ここでは、MNISTデータでトレーニングされた小さなニューラルネットワークの例を使用します。MNISTデータセットは、0から9までの数字の手書き画像で構成されています。画像のサイズは28x28ピクセルです。データセットは、60,000枚の列車画像と10,000枚の検証画像に分割されています。この実験で使用されたニューラルネットワークは、2層のフィードフォワードネットワークです。トレーニングはKubeflow Pipelinesを使用して実行されます。のドキュメントを参照してください ["こちらをご覧ください"](#) を参照してください。Kubeflowパイプラインには、PrerequisitesセクションのDockerイメージが組み込まれています。



Tensorboardを使用した結果の表示

モデルのトレーニングが完了すると、Tensorboardを使用して結果を視覚化できます。"Tensorboard"は、Kubeflowダッシュボードの機能として使用できます。ジョブ用のカスタムテンソルボードを作成できます。以下の例は、トレーニングの精度とエポック数、トレーニング損失とエポックの数。



Katibを使用したハイパーパラメータの実験

"カティブ" は、Kubeflow内のツールで、モデルのハイパーパラメータを試すために使用できます。実験を作成するには、まず目的のメトリック/ゴールを定義します。これは通常、テストの精度です。メトリックが定義されたら、再生するハイパーパラメータを選択します (optimizer/learning_rate/number of layers)。Katib は、ユーザー定義の値を使用してハイパーパラメータスイープを実行し、目的のメトリックを満たす最適なパラメータの組み合わせを見つけます。これらのパラメータは、UIの各セクションで定義できます。または、必要な仕様で*YAML*ファイルを定義することもできます。以下はKatib実験の図です。

| 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 KubeFlow Katib interface. On the left is a navigation sidebar with options like Home, Notebooks, Volumes, Tensorboards, and Katib. The main area displays the 'Experiment details' for 'mnist-pytorch'. A message at the top states 'Couldn't find any successful Trial'. Below this is a table with columns for OVERVIEW, TRIALS, DETAILS, and YAML. The table lists various metrics: Name (mnist-pytorch), 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), and Succeeded trials (0). Below the table is the 'Experiment Conditions' section with a filter input.

NetAppスナップショットを使用してデータを保存し、トレーサビリティを確保

モデルのトレーニング中に、トレーサビリティのためにトレーニングデータセットのスナップショットを保存したい場合があります。これを行うには、以下のようにパイプラインにスナップショットステップを追加します。スナップショットを作成するには、["Kubernetes向けNetApp DataOpsツールキット"](#)。

```
@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=["\n",
      "python3 -m pip install netapp-dataops-k8s && \n",
      "echo '' + volume_snapshot_name + '' > /volume_snapshot_name.txt && \n",
      "netapp_dataops_k8s_cli.py create volume-snapshot --pvc-name=" + "mnist-data" + " --snapshot-name=" + str(volume_snapshot_name) + " --namespace={work[0,namespace]}",
      "file_outputs{{volume_snapshot_name}}: /volume_snapshot_name.txt"
    ]
  )
  mnist_test_task.after(mnist_train_task)
  dataset_snapshot.after(mnist_train_task)
```

を参照してください ["KubeflowのNetApp DataOpsツールキットの例"](#) を参照してください。

著作権に関する情報

Copyright © 2024 NetApp, Inc. All Rights Reserved. Printed in the U.S.このドキュメントは著作権によって保護されています。著作権所有者の書面による事前承諾がある場合を除き、画像媒体、電子媒体、および写真複写、記録媒体、テープ媒体、電子検索システムへの組み込みを含む機械媒体など、いかなる形式および方法による複製も禁止します。

ネットアップの著作物から派生したソフトウェアは、次に示す使用許諾条項および免責条項の対象となります。

このソフトウェアは、ネットアップによって「現状のまま」提供されています。ネットアップは明示的な保証、または商品性および特定目的に対する適合性の暗示的保証を含み、かつこれに限定されないいかなる暗示的な保証も行いません。ネットアップは、代替品または代替サービスの調達、使用不能、データ損失、利益損失、業務中断を含み、かつこれに限定されない、このソフトウェアの使用により生じたすべての直接的損害、間接的損害、偶発的損害、特別損害、懲罰的損害、必然的損害の発生に対して、損失の発生の可能性が通知されていたとしても、その発生理由、根拠とする責任論、契約の有無、厳格責任、不法行為（過失またはそうでない場合を含む）にかかわらず、一切の責任を負いません。

ネットアップは、ここに記載されているすべての製品に対する変更を随時、予告なく行う権利を保有します。ネットアップによる明示的な書面による合意がある場合を除き、ここに記載されている製品の使用により生じる責任および義務に対して、ネットアップは責任を負いません。この製品の使用または購入は、ネットアップの特許権、商標権、または他の知的所有権に基づくライセンスの供与とはみなされません。

このマニュアルに記載されている製品は、1つ以上の米国特許、その他の国の特許、および出願中の特許によって保護されている場合があります。

権利の制限について：政府による使用、複製、開示は、DFARS 252.227-7013（2014年2月）およびFAR 5252.227-19（2007年12月）のRights in Technical Data -Noncommercial Items（技術データ - 非商用品目に関する諸権利）条項の(b)(3)項、に規定された制限が適用されます。

本書に含まれるデータは商用製品および/または商用サービス（FAR 2.101の定義に基づく）に関係し、データの所有権はNetApp, Inc.にあります。本契約に基づき提供されるすべてのネットアップの技術データおよびコンピュータソフトウェアは、商用目的であり、私費のみで開発されたものです。米国政府は本データに対し、非独占的かつ移転およびサブライセンス不可で、全世界を対象とする取り消し不能の制限付き使用权を有し、本データの提供の根拠となった米国政府契約に関連し、当該契約の裏付けとする場合にのみ本データを使用できます。前述の場合を除き、NetApp, Inc.の書面による許可を事前に得ることなく、本データを使用、開示、転載、改変するほか、上演または展示することはできません。国防総省にかかる米国政府のデータ使用权については、DFARS 252.227-7015(b)項（2014年2月）で定められた権利のみが認められます。

商標に関する情報

NetApp、NetAppのロゴ、<http://www.netapp.com/TM>に記載されているマークは、NetApp, Inc.の商標です。その他の会社名と製品名は、それを所有する各社の商標である場合があります。