



# **Kubeflow 操作和任务示例**

## **NetApp Solutions**

NetApp  
April 12, 2024

This PDF was generated from [https://docs.netapp.com/zh-cn/netapp-solutions/ai/aicp\\_example\\_kubeflow\\_operations\\_and\\_tasks\\_overview.html](https://docs.netapp.com/zh-cn/netapp-solutions/ai/aicp_example_kubeflow_operations_and_tasks_overview.html) on April 12, 2024. Always check docs.netapp.com for the latest.

# 目录

- Kubeflow 操作和任务示例 ..... 1
  - Kubeflow 操作和任务示例 ..... 1
  - 为数据科学家或开发人员配置 Jupyter 笔记本电脑工作空间 ..... 1
  - 示例笔记本电脑和管道 ..... 9

# Kubeflow 操作和任务示例

本节包括您可能希望使用 Kubeflow 执行的各种操作和任务的示例。

## Kubeflow 操作和任务示例

本节包括您可能希望使用 Kubeflow 执行的各种操作和任务的示例。

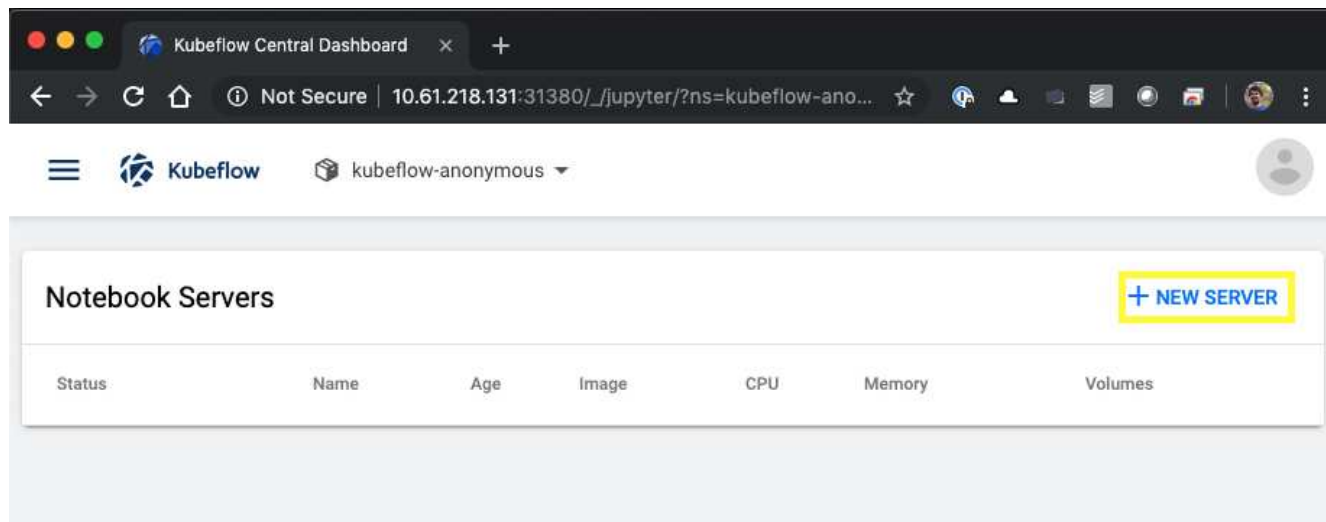
## 为数据科学家或开发人员配置 Jupyter 笔记本电脑工作空间

Kubeflow 能够快速配置新的 Jupyter 笔记本电脑服务器，以充当数据科学家工作空间。要使用 Kubeflow 配置新的 Jupyter 笔记本电脑服务器，请执行以下任务。有关 Kubeflow 上下文中 Jupyter 笔记本电脑的详细信息，请参见 "[Kubeflow 官方文档](#)"。

1. 从 Kubeflow 中央信息板中，单击主菜单中的 Notebook Servers 以导航到 Jupyter 笔记本电脑服务器管理页面。



2. 单击新服务器以配置新的 Jupyter 笔记本电脑服务器。



3. 为新服务器指定一个名称，选择希望服务器基于的 Docker 映像，并指定服务器要预留的 CPU 和 RAM 量。如果命名空间字段为空，请使用页面标题中的选择命名空间菜单选择命名空间。然后，Namespace 字段将自动填充所选命名空间。

在以下示例中，选择了 `kubeflow-anonymous` 命名空间。此外，还接受 Docker 映像，CPU 和 RAM 的默认值。

**Name**

Specify the name of the Notebook Server and the Namespace it will belong to.

Name:  Namespace:

**Image**

A starter Jupyter Docker Image with a baseline deployment and typical ML packages.

☐ Custom Image

Image:

**CPU / RAM**

Specify the total amount of CPU and RAM reserved by your Notebook Server. For CPU-intensive workloads, you can choose more than 1 CPU (e.g. 1.5).

CPU:  Memory:

- 指定工作空间卷详细信息。如果选择创建新卷，则会使用默认 StorageClass 配置该卷或 PVC 。因为在部分中，使用 Trident 的 StorageClass 被指定为默认 StorageClass "[Kubeflow 部署](#)"，卷或 PVC 配置有 Trident 。此卷会自动挂载为 Jupyter 笔记本电脑服务器容器中的默认工作空间。用户在服务器上创建但未保存到单独数据卷的任何笔记本电脑将自动保存到此工作空间卷。因此，这些笔记本电脑在重新启动后会持久存在。

**Workspace Volume**

Configure the Volume to be mounted as your personal Workspace.

☒ Don't use Persistent Storage for User's home

Type:  Name:  Size:  Mode:  Mount Point:

- 添加数据卷。以下示例指定了一个名为 "pt-fg-all" 的现有 PVC 并接受默认挂载点。

**Data Volumes**

Configure the Volumes to be mounted as your Datasets.

[+ ADD VOLUME](#)

Type	Name	Size	Mode	Mount Point
Existing	pb-fg-all	10Gi	ReadWriteOnce	/home/jovyan/data-vol-1

6. \* 可选：\* 请求将所需数量的 GPU 分配给您的笔记本服务器。在以下示例中，请求一个 GPU 。

**Configurations**

Extra layers of configurations that will be applied to the new Notebook. (e.g. Insert credentials as Secrets, set Environment Variables.)

Configurations

**Extra Resources**

Specify extra resources that might be needed in the Notebook Server.

☒ Enable Shared Memory

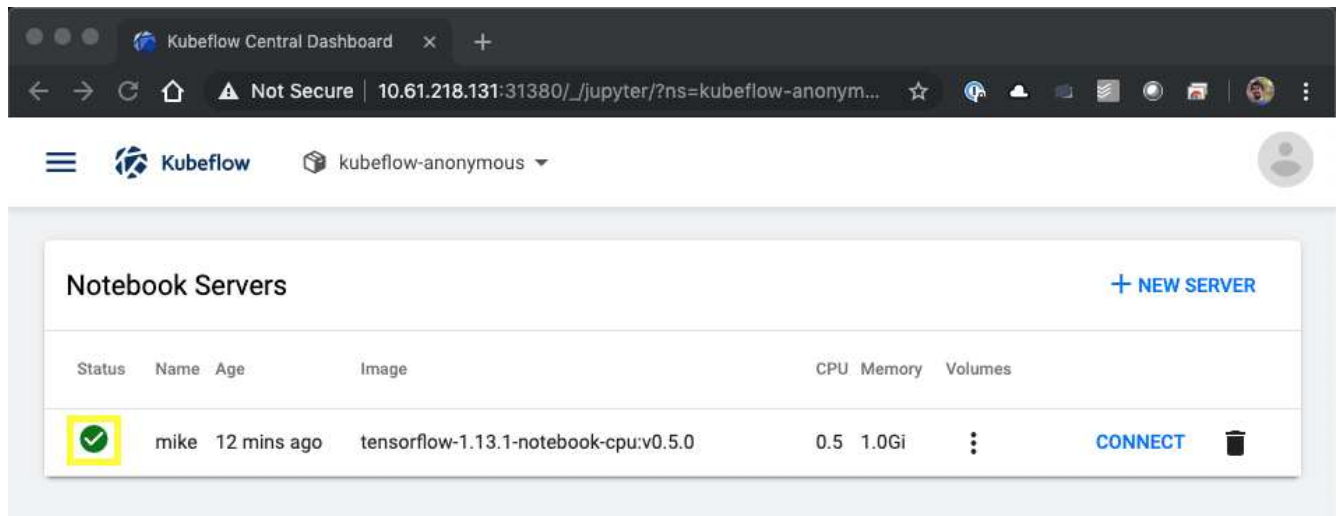
Extra Resources \*

`{"nvidia.com/gpu": 1}`

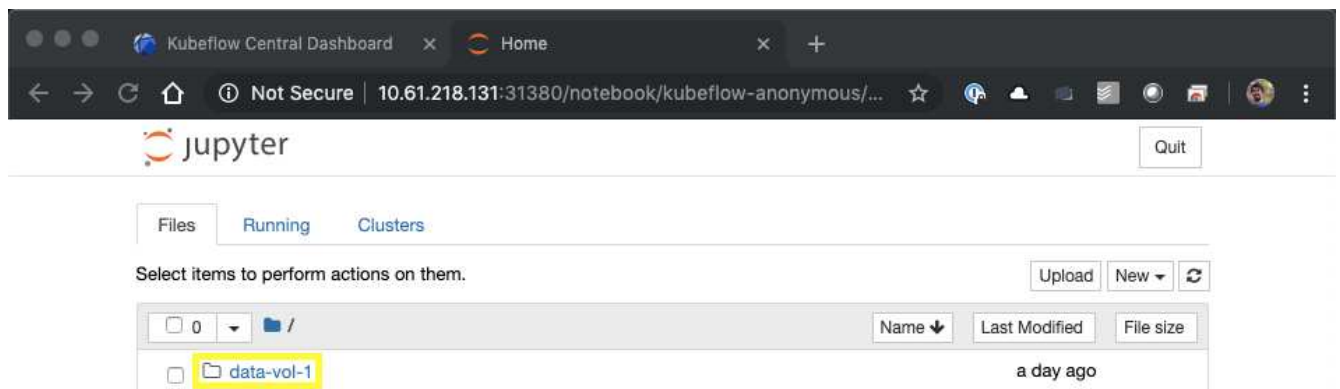
Extra Resources available in the cluster (ex. NVIDIA GPUs)

[LAUNCH](#) [CANCEL](#)

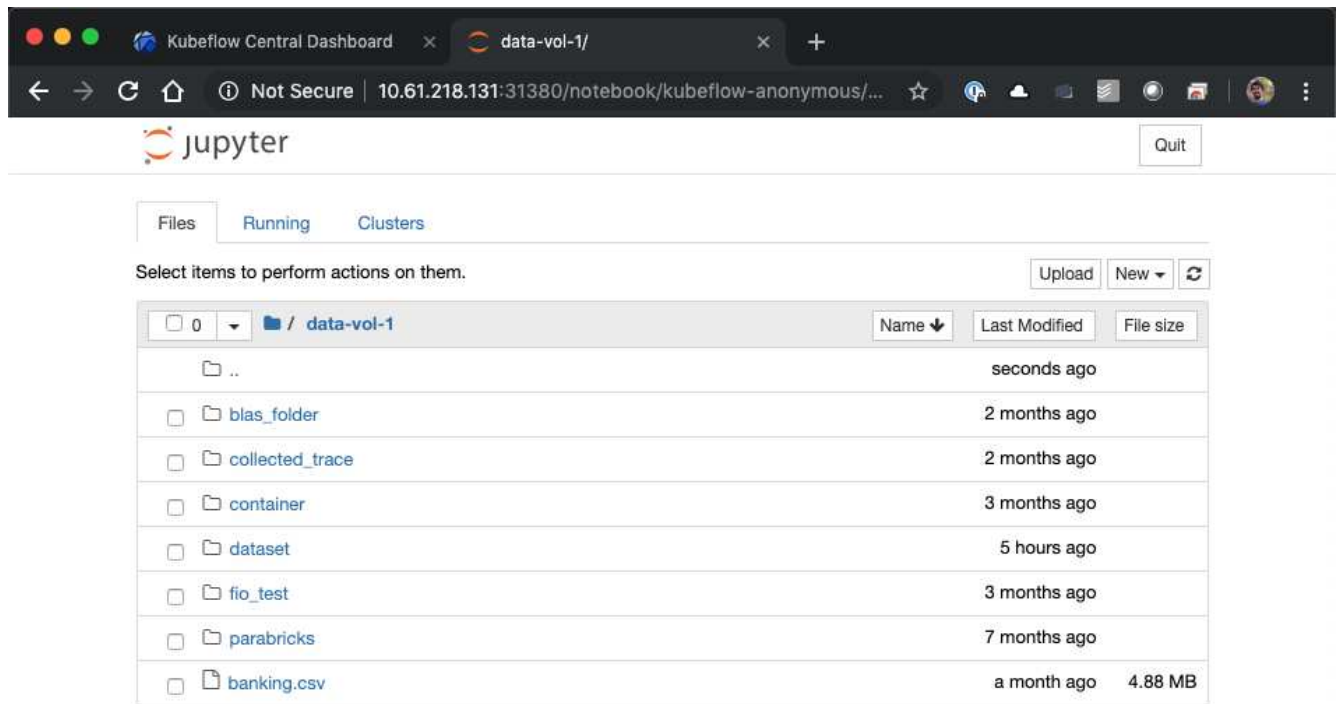
7. 单击启动以配置新的笔记本电脑服务器。
8. 等待笔记本电脑服务器完全配置完毕。如果您从未使用您指定的 Docker 映像配置服务器，则可能需要几分钟的时间，因为需要下载此映像。服务器配置完成后，Jupyter 笔记本电脑服务器管理页面上的状态列会显示一个绿色复选标记。



9. 单击连接以连接到新的服务器 Web 界面。
10. 确认步骤 6 中指定的数据集卷已挂载到服务器上。请注意，默认情况下，此卷会挂载在默认工作空间中。从用户的角度来看，这只是工作空间中的另一个文件夹。用户可能是数据科学家，而不是基础架构专家，因此使用此卷无需具备任何存储专业知识。

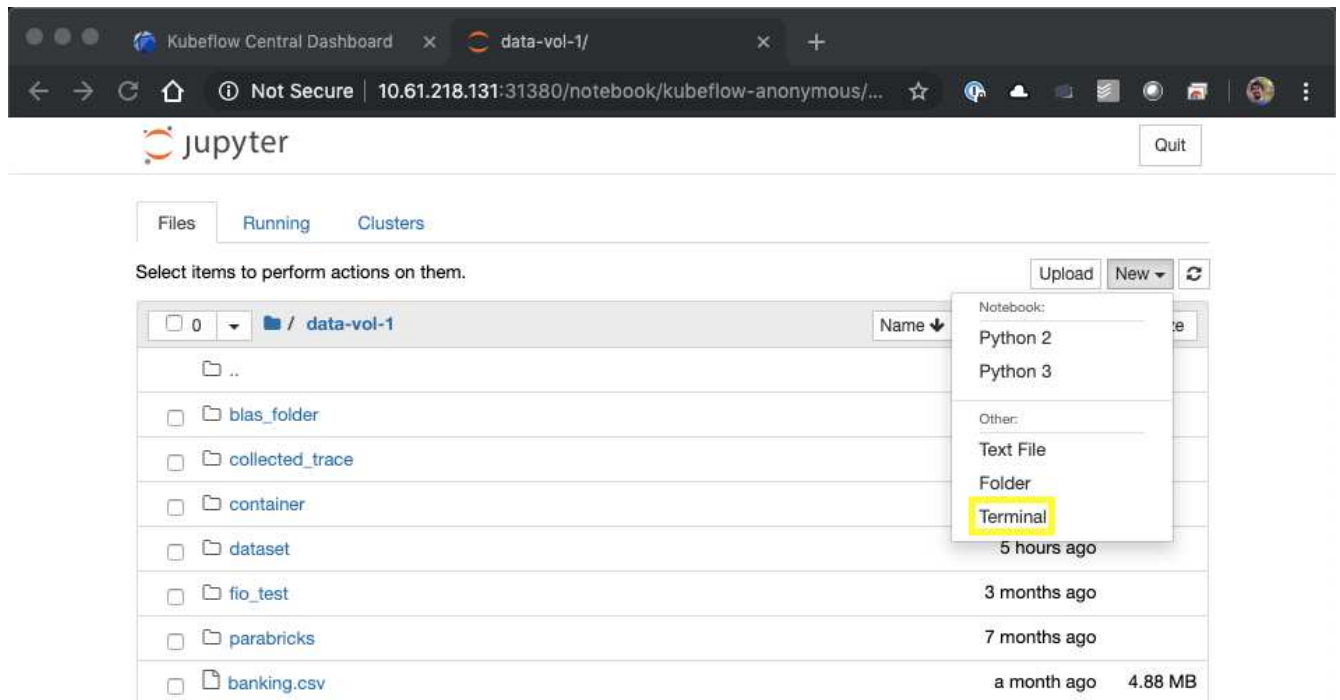






11. 打开一个终端，假设步骤 5 中请求了一个新卷，请执行 `df -h` 以确认已挂载新的 Trident 配置的永久性卷作为默认工作空间。

默认工作空间目录是首次访问服务器的 Web 界面时显示的基目录。因此，使用 Web 界面创建的任何项目都会存储在此 Trident 配置的永久性卷上。



```
Kubeflow Central Dashboard x data-vol-1/ x 10.61.218.131:31380/notebook/ x +
Not Secure | 10.61.218.131:31380/notebook/kubeflow-anonymous/mike/t...
jupyter

$ df -h
Filesystem                                Size  Used Avail
Use% Mounted on
overlay                                  439G   34G  382G
9% /
tmpfs                                     64M    0   64M
0% /dev
tmpfs                                     252G    0  252G
0% /sys/fs/cgroup
/dev/sda2                                439G   34G  382G
9% /etc/hosts
192.168.11.11:/trident_pvc_3dcfe7e5_d5a9_11e9_9b9d_00505681a82d 10G  320K   10G
1% /home/jovyan
tmpfs                                     252G    0  252G
0% /dev/shm
192.168.11.11:/pb_fg_all                 10T   10T   47G
100% /home/jovyan/data-vol-1
tmpfs                                     252G   12K  252G
1% /run/secrets/kubernetes.io/serviceaccount
tmpfs                                     252G   12K  252G
1% /proc/driver/nvidia
tmpfs                                     51G   4.9M   51G
1% /run/nvidia-persistenced/socket
udev                                     252G    0  252G
0% /dev/nvidia5
tmpfs                                     252G    0  252G
0% /proc/acpi
tmpfs                                     252G    0  252G
0% /proc/scsi
tmpfs                                     252G    0  252G
0% /sys/firmware
$
```

12. 使用终端运行 `nvidia-smi` 以确认为笔记本电脑服务器分配了正确数量的 GPU。在以下示例中，已按照步骤 7 中的请求为笔记本电脑服务器分配一个 GPU。

```
Kubeflow Central Dashboard x Home x 10.61.218.131:31380/notebook/ x +
Not Secure | 10.61.218.131:31380/notebook/kubeflow-anonymous/mike/t...
jupyter

$ nvidia-smi
Fri Sep 13 13:52:15 2019
+-----+
| NVIDIA-SMI 410.104                Driver Version: 410.104                CUDA Version: N/A                |
+-----+-----+
| GPU   Name           Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|    0   Tesla V100-SXM2...    On   | 00000000:86:00:0 Off |              0%      |
| N/A   38C    P0     46W / 300W | 0MiB / 32480MiB |              0%      |
+-----+-----+

+-----+
| Processes:                               GPU Memory |
|  GPU       PID    Type    Process name                        Usage |
+-----+-----+
| No running processes found |
+-----+

$
```

## 示例笔记本电脑和管道

。"适用于 Kubernetes 的 NetApp 数据科学工具包" 可与 Kubeflow 结合使用。将 NetApp 数据科学工具包与 Kubeflow 结合使用具有以下优势：

- 数据科学家可以直接在 Jupyter 笔记本电脑中执行高级 NetApp 数据管理操作。
- 可以使用 Kubeflow 管道框架将高级 NetApp 数据管理操作整合到自动化工作流中。

请参见 "[Kubeflow 示例](#)" 有关将工具包与 Kubeflow 结合使用的详细信息，请参见 NetApp Data Science Toolkit GitHub 存储库中的一节。

## 版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

## 商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。