



# Kubeflow

## NetApp Solutions

NetApp  
May 10, 2024

This PDF was generated from [https://docs.netapp.com/es-es/netapp-solutions/ai/aicp\\_kubeflow\\_deployment\\_overview.html](https://docs.netapp.com/es-es/netapp-solutions/ai/aicp_kubeflow_deployment_overview.html) on May 10, 2024. Always check docs.netapp.com for the latest.

# Tabla de contenidos

- Kubeflow ..... 1
- Despliegue de Kubeflow ..... 1
- Ejemplo de operaciones y tareas de Kubeflow ..... 2

# Kubeflow

## Despliegue de Kubeflow

En esta sección se describen las tareas que debe completar para poner en marcha Kubeflow en su clúster de Kubernetes.

### Requisitos previos

Antes de realizar el ejercicio de implementación descrito en esta sección, asumimos que ya ha realizado las siguientes tareas:

1. Ya tiene un clúster de Kubernetes en funcionamiento y está ejecutando una versión de Kubernetes que es compatible con la versión de Kubeflow que desea implementar. Para ver una lista de las versiones de Kubernetes compatibles, consulte las dependencias de su versión de Kubeflow en la "[Documentación oficial de Kubeflow](#)".
2. Ya ha instalado y configurado Astra Trident de NetApp en su clúster de Kubernetes. Si quiere más información sobre Astra Trident, consulte la "[Documentación de Astra Trident](#)".

### Establezca el tipo de almacenamiento de Kubernetes predeterminado

Antes de implementar Kubeflow, recomendamos designar un StorageClass predeterminado en su clúster de Kubernetes. El proceso de puesta en marcha de Kubeflow puede intentar aprovisionar nuevos volúmenes persistentes mediante el StorageClass predeterminado. Si no se designa ningún StorageClass como el StorageClass predeterminado, es posible que falle la implementación. Para designar un StorageClass predeterminado en el clúster, realice la siguiente tarea desde el host de salto de implementación. Si ya ha designado un tipo de almacenamiento predeterminado en el clúster, puede omitir este paso.

1. Designe una de las clases de almacenamiento existentes como clase de almacenamiento predeterminada. Los comandos de ejemplo siguientes muestran la designación de un StorageClass llamado `ontap-ai-flexvols-retain` Como el tipo de almacenamiento predeterminado.



La `ontap-nas-flexgroup` El tipo de backend de Trident tiene un tamaño de RVP mínimo que es bastante grande. De manera predeterminada, Kubeflow intenta suministrar EVs que son sólo unos pocos GBS en tamaño. Por lo tanto, no debe designar un StorageClass que utilice `ontap-nas-flexgroup` Tipo back-end como StorageClass predeterminado para la implementación de Kubeflow.

```

$ 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

```

## Opciones de implementación de Kubeflow

Hay muchas opciones diferentes para implementar Kubeflow. Consulte la ["Documentación oficial de Kubeflow"](#) para acceder a una lista de opciones de puesta en marcha y elija la opción que mejor se ajuste a sus necesidades.

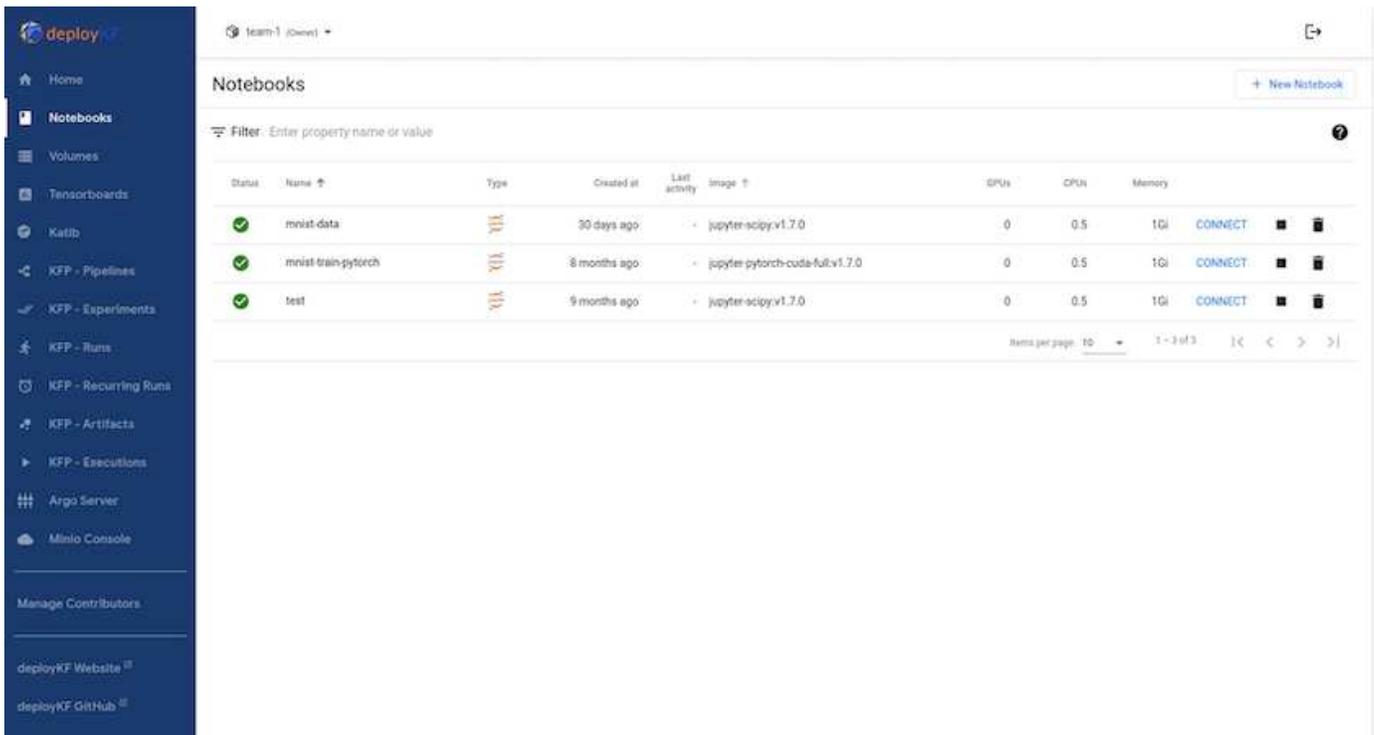


Para fines de validación, implementamos Kubeflow 1,7 utilizando ["Desplegar KF" 0,1.1](#).

## Ejemplo de operaciones y tareas de Kubeflow

### Aprovisione un espacio de trabajo para portátiles Jupyter para uso científico de datos o desarrollador

Kubeflow es capaz de suministrar rápidamente nuevos servidores Jupyter Notebook para actuar como espacios de trabajo de científicos de datos. Para obtener más información acerca de Jupyter Notebooks dentro del contexto de Kubeflow, consulte ["Documentación oficial de Kubeflow"](#).



## Use el kit de herramientas NetApp DataOps con Kubeflow

La "[Kit de herramientas para la ciencia de datos de NetApp para Kubernetes](#)" Se puede utilizar junto con Kubeflow. El uso del kit de herramientas para la ciencia de datos de NetApp con Kubeflow ofrece las siguientes ventajas:

- Los científicos de datos pueden llevar a cabo operaciones avanzadas de gestión de datos de NetApp, como la creación de copias Snapshot y clones, directamente desde un portátil de Jupyter.
- Las operaciones avanzadas de gestión de datos de NetApp, como la creación de snapshots y clones, se pueden incorporar en flujos de trabajo automatizados mediante el marco de Kubeflow Pipelines.

Consulte la "[Ejemplos de Kubeflow](#)" Sección dentro del repositorio de Data Science Toolkit de NetApp, GitHub para obtener información sobre el uso del kit de herramientas con Kubeflow.

## Flujo de trabajo de ejemplo: Formación de un modelo de reconocimiento de imágenes mediante Kubeflow y el kit de herramientas de DataOps de NetApp

En esta sección se describen los pasos que deben seguirse para entrenar y poner en marcha una red neuronal para el reconocimiento de imágenes mediante Kubeflow y el kit de herramientas NetApp DataOps. Esta información pretende servir de ejemplo para mostrar un trabajo de formación que incorpora almacenamiento NetApp.

### Requisitos previos

Cree un Dockerfile con las configuraciones necesarias para usar en los pasos de tren y prueba dentro de la tubería de Kubeflow.

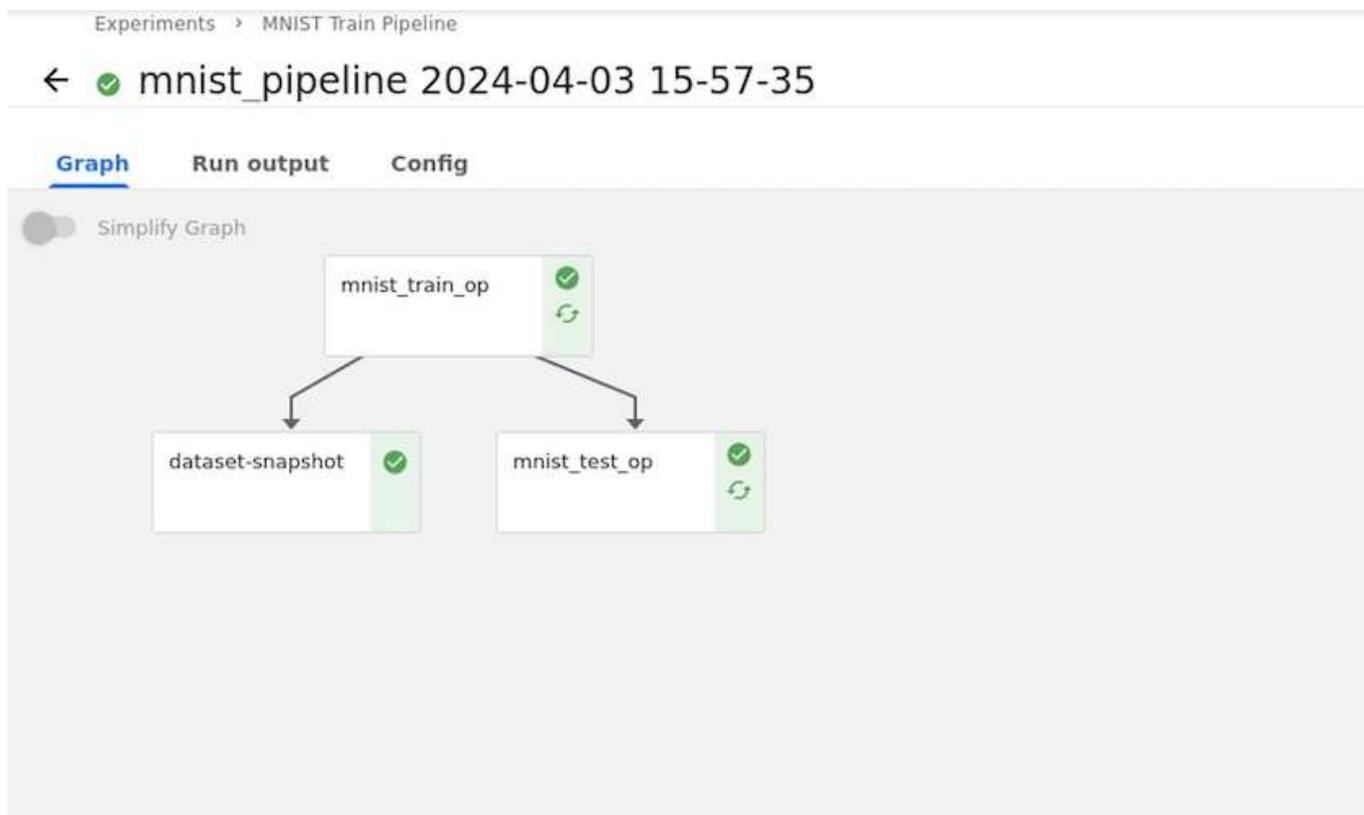
Aquí hay un ejemplo de un Dockerfile -

```
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"]
```

Dependiendo de sus requisitos, instale todas las bibliotecas y paquetes necesarios para ejecutar el programa. Antes de entrenar el modelo de aprendizaje automático, se asume que ya tiene un despliegue de Kubeflow en funcionamiento.

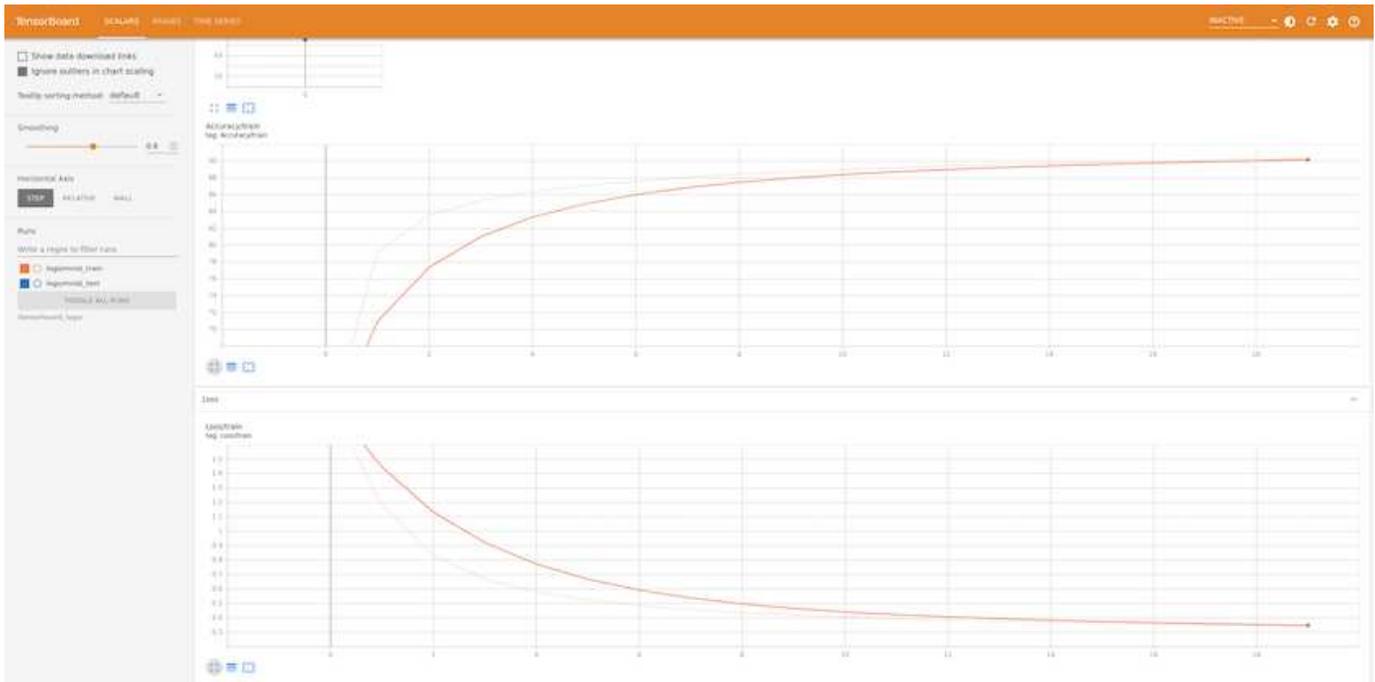
### Entrena un pequeño NN en datos MNIST usando PyTorch y Kubeflow Pipelines

Utilizamos el ejemplo de una pequeña red neuronal entrenada en datos MNIST. El conjunto de datos MNIST consta de imágenes escritas a mano de dígitos del 0 al 9. Las imágenes tienen un tamaño de 28x28 píxeles. El conjunto de datos se divide en 60.000 imágenes de tren y 10.000 imágenes de validación. La red neuronal utilizada para este experimento es una red de avance de 2 capas. El entrenamiento se ejecuta mediante Tuberías Kubeflow. Consulte la documentación "[aquí](#)" si quiere más información. Nuestro Pipeline Kubeflow incorpora la imagen de docker de la sección Requisitos previos.



### Visualiza los resultados usando Tensorboard

Una vez entrenado el modelo, podemos visualizar los resultados usando Tensorboard. "[Tablero de placas](#)" Está disponible como una característica en el Panel de control de Kubeflow. Puede crear un tablero de tenencia personalizado para el trabajo. Un ejemplo a continuación muestra el trazado de la precisión del entrenamiento frente a número de épocas y pérdida de entrenamiento con respecto a número de épocas.



## Experimenta con Hiperparámetros usando Katib

"Katib" Es una herramienta dentro de Kubeflow que se puede utilizar para experimentar con los hiperparámetros del modelo. Para crear un experimento, defina primero una métrica/objetivo deseado. Esta es generalmente la precisión de la prueba. Una vez definida la métrica, elija hiperparámetros con los que desea jugar (optimizador/learning\_rate/número de capas). Katib realiza un barrido de hiperparámetros con los valores definidos por el usuario para encontrar la mejor combinación de parámetros que satisfagan la métrica deseada. Puede definir estos parámetros en cada sección de la interfaz de usuario. Alternativamente, puede definir un archivo **YAML** con las especificaciones necesarias. A continuación se muestra una ilustración de un experimento Katib -

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, ftrl

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, Katib, KFP - Pipelines, KFP - Experiments, KFP - Runs, KFP - Recurring Runs, KFP - Artifacts, KFP - Executions, Argo Server, and Minio Console. The main panel is titled 'Experiment details' and shows a message: 'Couldn't find any successful Trial.' Below this is a table with columns: OVERVIEW, TRIALS, DETAILS, and YAML. The table contains the following information:

OVERVIEW	TRIALS	DETAILS	YAML
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		
Succeeded trials	0		

Below the table is a section for 'Experiment Conditions' and a filter input field with the placeholder text 'Enter property name or value'.

## Utilice Instantáneas de NetApp para guardar datos y rastrearlos

Durante el entrenamiento del modelo, puede que desee guardar una instantánea del conjunto de datos de entrenamiento para la trazabilidad. Para ello, podemos agregar un paso de snapshot a la canalización, tal como se muestra a continuación. Para crear la instantánea, podemos usar la ["Kit de herramientas Data OPS de NetApp para Kubernetes"](#).

```
@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={{workflow.namespace}}",
            file_outputs={'volume_snapshot_name': '/volume_snapshot_name.txt'}
        ]
    )
    mnist_test_task.after(mnist_train_task)
    dataset_snapshot.after(mnist_train_task)
```

Consulte la ["Ejemplo del kit de herramientas NetApp DataOps para Kubeflow"](#) si quiere más información.

## Información de copyright

Copyright © 2024 NetApp, Inc. Todos los derechos reservados. Imprimido en EE. UU. No se puede reproducir este documento protegido por copyright ni parte del mismo de ninguna forma ni por ningún medio (gráfico, electrónico o mecánico, incluidas fotocopias, grabaciones o almacenamiento en un sistema de recuperación electrónico) sin la autorización previa y por escrito del propietario del copyright.

El software derivado del material de NetApp con copyright está sujeto a la siguiente licencia y exención de responsabilidad:

ESTE SOFTWARE LO PROPORCIONA NETAPP «TAL CUAL» Y SIN NINGUNA GARANTÍA EXPRESA O IMPLÍCITA, INCLUYENDO, SIN LIMITAR, LAS GARANTÍAS IMPLÍCITAS DE COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO, CUYA RESPONSABILIDAD QUEDA EXIMIDA POR EL PRESENTE DOCUMENTO. EN NINGÚN CASO NETAPP SERÁ RESPONSABLE DE NINGÚN DAÑO DIRECTO, INDIRECTO, ESPECIAL, EJEMPLAR O RESULTANTE (INCLUYENDO, ENTRE OTROS, LA OBTENCIÓN DE BIENES O SERVICIOS SUSTITUTIVOS, PÉRDIDA DE USO, DE DATOS O DE BENEFICIOS, O INTERRUPTIÓN DE LA ACTIVIDAD EMPRESARIAL) CUALQUIERA SEA EL MODO EN EL QUE SE PRODUJERON Y LA TEORÍA DE RESPONSABILIDAD QUE SE APLIQUE, YA SEA EN CONTRATO, RESPONSABILIDAD OBJETIVA O AGRAVIO (INCLUIDA LA NEGLIGENCIA U OTRO TIPO), QUE SURJAN DE ALGÚN MODO DEL USO DE ESTE SOFTWARE, INCLUSO SI HUBIEREN SIDO ADVERTIDOS DE LA POSIBILIDAD DE TALES DAÑOS.

NetApp se reserva el derecho de modificar cualquiera de los productos aquí descritos en cualquier momento y sin aviso previo. NetApp no asume ningún tipo de responsabilidad que surja del uso de los productos aquí descritos, excepto aquello expresamente acordado por escrito por parte de NetApp. El uso o adquisición de este producto no lleva implícita ninguna licencia con derechos de patente, de marcas comerciales o cualquier otro derecho de propiedad intelectual de NetApp.

Es posible que el producto que se describe en este manual esté protegido por una o más patentes de EE. UU., patentes extranjeras o solicitudes pendientes.

LEYENDA DE DERECHOS LIMITADOS: el uso, la copia o la divulgación por parte del gobierno están sujetos a las restricciones establecidas en el subpárrafo (b)(3) de los derechos de datos técnicos y productos no comerciales de DFARS 252.227-7013 (FEB de 2014) y FAR 52.227-19 (DIC de 2007).

Los datos aquí contenidos pertenecen a un producto comercial o servicio comercial (como se define en FAR 2.101) y son propiedad de NetApp, Inc. Todos los datos técnicos y el software informático de NetApp que se proporcionan en este Acuerdo tienen una naturaleza comercial y se han desarrollado exclusivamente con fondos privados. El Gobierno de EE. UU. tiene una licencia limitada, irrevocable, no exclusiva, no transferible, no sublicenciable y de alcance mundial para utilizar los Datos en relación con el contrato del Gobierno de los Estados Unidos bajo el cual se proporcionaron los Datos. Excepto que aquí se disponga lo contrario, los Datos no se pueden utilizar, desvelar, reproducir, modificar, interpretar o mostrar sin la previa aprobación por escrito de NetApp, Inc. Los derechos de licencia del Gobierno de los Estados Unidos de América y su Departamento de Defensa se limitan a los derechos identificados en la cláusula 252.227-7015(b) de la sección DFARS (FEB de 2014).

## Información de la marca comercial

NETAPP, el logotipo de NETAPP y las marcas que constan en <http://www.netapp.com/TM> son marcas comerciales de NetApp, Inc. El resto de nombres de empresa y de producto pueden ser marcas comerciales de sus respectivos propietarios.