MLRun Pipeline con Iguazio

NetApp Solutions

NetApp
April 19, 2024

This PDF was generated from https://docs.netapp.com/es-es/netapp-solutions/ai/mlrun_introduction.html on April 19, 2024. Always check docs.netapp.com for the latest.

Tabla de contenidos

MLRun Pipeline con Iguazio . . . 1
- TR-4834: NetApp e Iguazio para el suministro de MLRun . . . 1
- Visión general de la tecnología . . . 3
- Requisitos de software y hardware . . . 7
- Resumen de casos de uso de predicción de fallos de dispositivos de red . . . 8
- Descripción general de la configuración . . . 8
- Implementación de la aplicación . . . 13
- Conclusión . . . 28

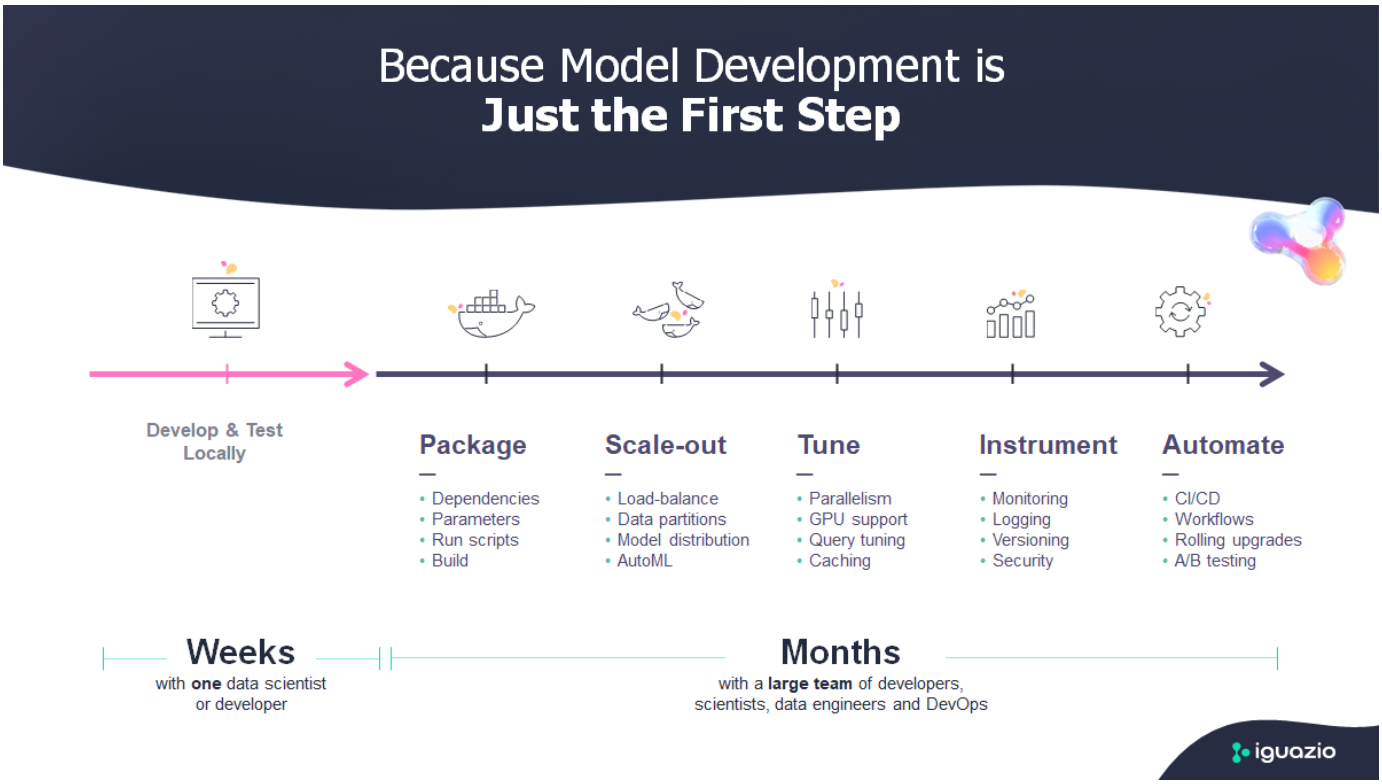
MLRun Pipeline con Iguazio

TR-4834: NetApp e Iguazio para el suministro de MLRun

Rick Huang, David Arnette, NetApp Marcelo Litovsky, Iguazio

Este documento contiene los detalles de la canalización MLRun usando NetApp ONTAP AI, el plano de control de IA de NetApp, el software Cloud Volumes de NetApp y la plataforma de ciencia de datos de Iguazio. Utilizamos la función Nuclio serverless, volúmenes persistentes de Kubernetes, volúmenes Cloud de NetApp, copias Snapshot de NetApp, consola Grafana, Y otros servicios en la plataforma Iguazio para crear una canalización de datos de extremo a extremo para la simulación de la detección de fallos de red. Hemos integrado las tecnologías de Iguazio y NetApp para permitir una rápida puesta en marcha de modelos, replicación de datos y supervisión de la producción tanto en las instalaciones como en el cloud.

El trabajo de un científico de datos debe centrarse en el entrenamiento y el ajuste de los modelos de aprendizaje automático y de inteligencia artificial (IA). Sin embargo, según una investigación realizada por Google, los científicos de datos gastan aproximadamente un 80 % de su tiempo en averiguar cómo hacer que sus modelos funcionen con aplicaciones empresariales y se ejecutan a escala, como se muestra en la siguiente imagen, en la que se ilustra el desarrollo de modelos en el flujo de trabajo de IA/ML.



Para gestionar proyectos de IA y ML integrales, se necesita una comprensión más amplia de los componentes empresariales. Aunque DevOps ha retomado la definición, la integración y la puesta en marcha de estos tipos de componentes, las operaciones de aprendizaje automático siguen un flujo similar que incluye proyectos de IA/ML. Para hacerse una idea de lo que puede tocar una canalización de IA/ML integral en la empresa, consulte la siguiente lista de componentes requeridos:

- Reducida
- Redes
- Oracle
- Sistemas de ficheros
- Contenedores
- Integración continua y canalización de puesta en marcha continua (CI/CD)
- Entorno de desarrollo integrado (IDE)
- Seguridad
- Políticas de acceso a los datos
- Hardware subyacente
- Cloud
- Virtualización
- Conjuntos de herramientas y bibliotecas de ciencia de datos

En este artículo mostramos cómo la asociación entre NetApp y Iguazio simplifica drásticamente el desarrollo de una canalización completa de IA/ML. Esta simplificación acelera el plazo de comercialización de todas sus aplicaciones de IA/ML.

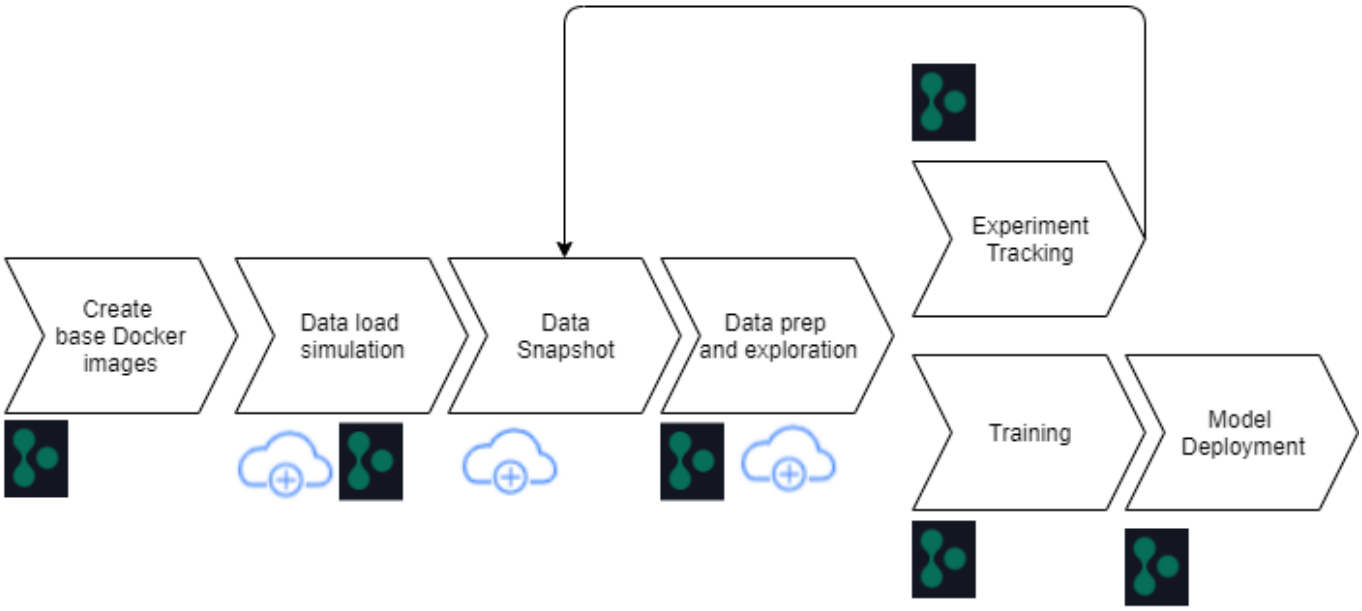
Público objetivo

El mundo de la ciencia de datos tiene múltiples disciplinas de tecnología de la información y negocio.

- El científico de datos necesita la flexibilidad para poder usar las herramientas y las bibliotecas de elección.
- El ingeniero de datos necesita saber cómo fluyen los datos y dónde residen.
- Un ingeniero de DevOps necesita herramientas para integrar nuevas aplicaciones de IA/ML en sus canalizaciones de CI/CD.
- Los usuarios empresariales quieren tener acceso a aplicaciones de IA/ML. Describimos cómo NetApp y Iguazio ayudan a cada una de estas funciones a aportar valor empresarial con nuestras plataformas.

Descripción general de la solución

Esta solución sigue el ciclo de vida de una aplicación de IA/ML. Empezamos con la labor de los científicos de datos para definir los diferentes pasos necesarios para preparar datos y entrenar y poner en marcha modelos. Seguimos con el trabajo necesario para crear una canalización completa con la capacidad de realizar un seguimiento de los artefactos, experimentar con la ejecución y desplegar en Kubeflow. Para completar el ciclo completo, integramos la canalización con NetApp Cloud Volumes para permitir el versionado de datos, tal y como se puede ver en la siguiente imagen.



iguazio

Cloud Volumes ONTAP

Visión general de la tecnología

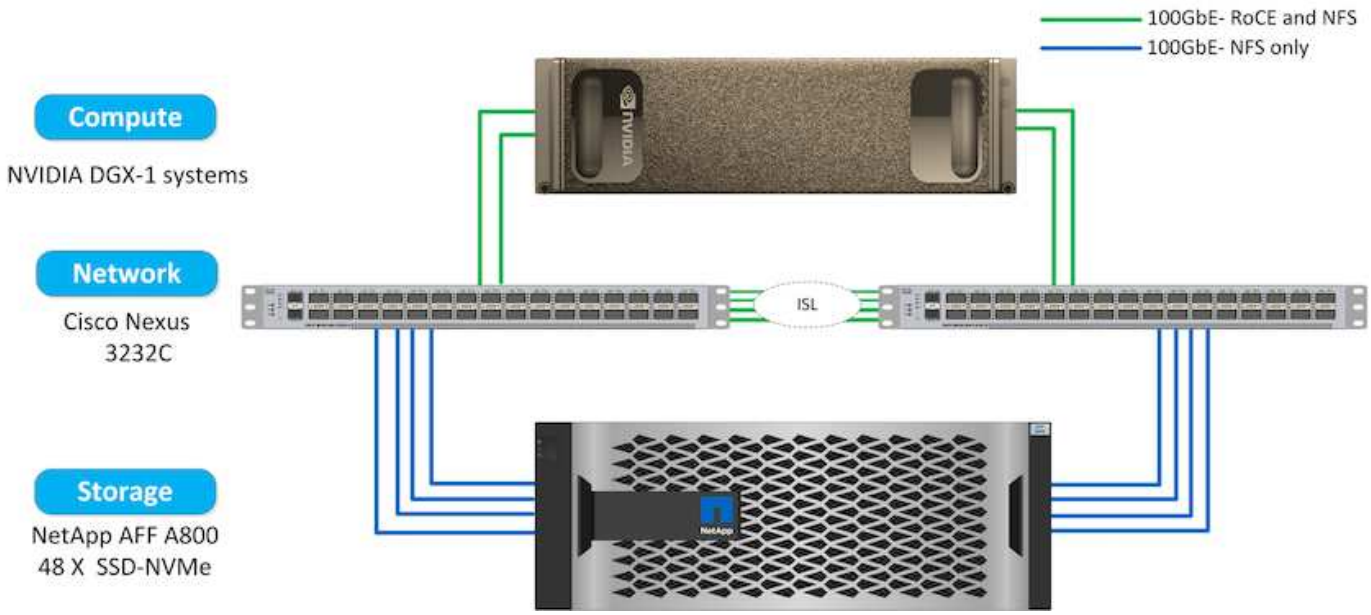
Información general de NetApp

NetApp es un referente en materia de datos para el cloud híbrido. NetApp proporciona una gama completa de servicios de datos en el cloud híbrido que simplifican la gestión de aplicaciones y datos en entornos de cloud y en las instalaciones para acelerar la transformación digital. Junto con nuestros partners, NetApp permite a organizaciones globales aprovechar al máximo todo el potencial de sus datos para ampliar los puntos de contacto con los clientes, fomentar una mayor innovación y optimizar sus operaciones.

ONTAP AI de NetApp

ONTAP AI de NetApp, impulsado por los sistemas NVIDIA DGX y el almacenamiento all-flash conectado al cloud de NetApp, optimiza el flujo de datos con total confianza y acelera el análisis, la formación y la inferencia con su Data Fabric, que abarca desde el perímetro al núcleo y al cloud. Proporciona a las organizaciones DE TI una arquitectura que ofrece las siguientes ventajas:

- Elimina las complejidades de diseño
- Permite un escalado independiente de las capacidades de computación y almacenamiento
- Permite a los clientes empezar con poco y escalar sin problemas
- Ofrece una gama de opciones de almacenamiento para diferentes tipos de rendimiento y coste.ONTAP AI ofrece pilas de infraestructuras convergentes que incorporan NVIDIA DGX-1, un sistema de IA a escala de petaflops y switches Ethernet de alto rendimiento NVIDIA Mellanox para unificar cargas de trabajo de IA, simplificar la puesta en marcha y acelerar el retorno de la inversión. Para este informe técnico, hemos aprovechado la inteligencia artificial de ONTAP con un sistema de almacenamiento DGX-1 y AFF A800 de NetApp. En la siguiente imagen, se muestra la topología de ONTAP AI con el sistema DGX-1 utilizado en esta validación.



Plano de control de IA de NetApp

El plano de control de IA de NetApp permite aprovechar la IA Y EL APRENDIZAJE AUTOMÁTICO con una solución que ofrece una escalabilidad extrema, una puesta en marcha optimizada y una disponibilidad de datos ininterrumpida. La solución de plano de control de IA integra Kubernetes y Kubeflow con una estructura de datos habilitada por NetApp. Kubernetes, la plataforma de orquestación de contenedores estándar del sector para puestas en marcha nativas del cloud, permite la escalabilidad y la portabilidad de las cargas de trabajo. Kubeflow es una plataforma de aprendizaje automático de código abierto que simplifica la gestión y la puesta en marcha, lo que permite a los desarrolladores hacer más ciencia de datos en menos tiempo. Una estructura de datos habilitada por NetApp ofrece una disponibilidad y portabilidad de datos sin concesiones para garantizar que sus datos están accesibles en toda la canalización, desde el perímetro al núcleo y al cloud. En este informe técnico se utiliza el plano de control de IA de NetApp en una canalización MLRun. La siguiente imagen muestra la página de gestión de clústeres de Kubernetes, donde se pueden tener distintos extremos para cada clúster. Conectamos volúmenes persistentes de NFS al clúster de Kubernetes, y las siguientes imágenes muestran un volumen persistente conectado al clúster donde "Trident de NetApp" ofrece compatibilidad con almacenamiento persistente y capacidades de gestión de datos.

Kubernetes Clusters

Discover Cluster

4 Kubernetes Clusters

kubernetes

https://3.20.111.39:6443
Cluster Endpoint

v1.15.5
Cluster Version

19.07.1
Trident Version

0
Working Environments

kubernetes

https://172.31.14.31:6443
Cluster Endpoint

v1.15.5
Cluster Version

19.07.1
Trident Version

1
Working Environments

Persistent Volumes for Kubernetes

Connected with Kubernetes Cluster

Cloud Volumes ONTAP is connected to 1 Kubernetes cluster. View Cluster

You can connect another Kubernetes cluster to this Cloud Volumes ONTAP system. If the Kubernetes cluster is in a different network than Cloud Volumes ONTAP, specify a custom export policy to provide access to clients.

Kubernetes Cluster

Select Kubernetes Cluster

kubernetes

Custom Export Policy *(Optional)*

Custom Export Policy

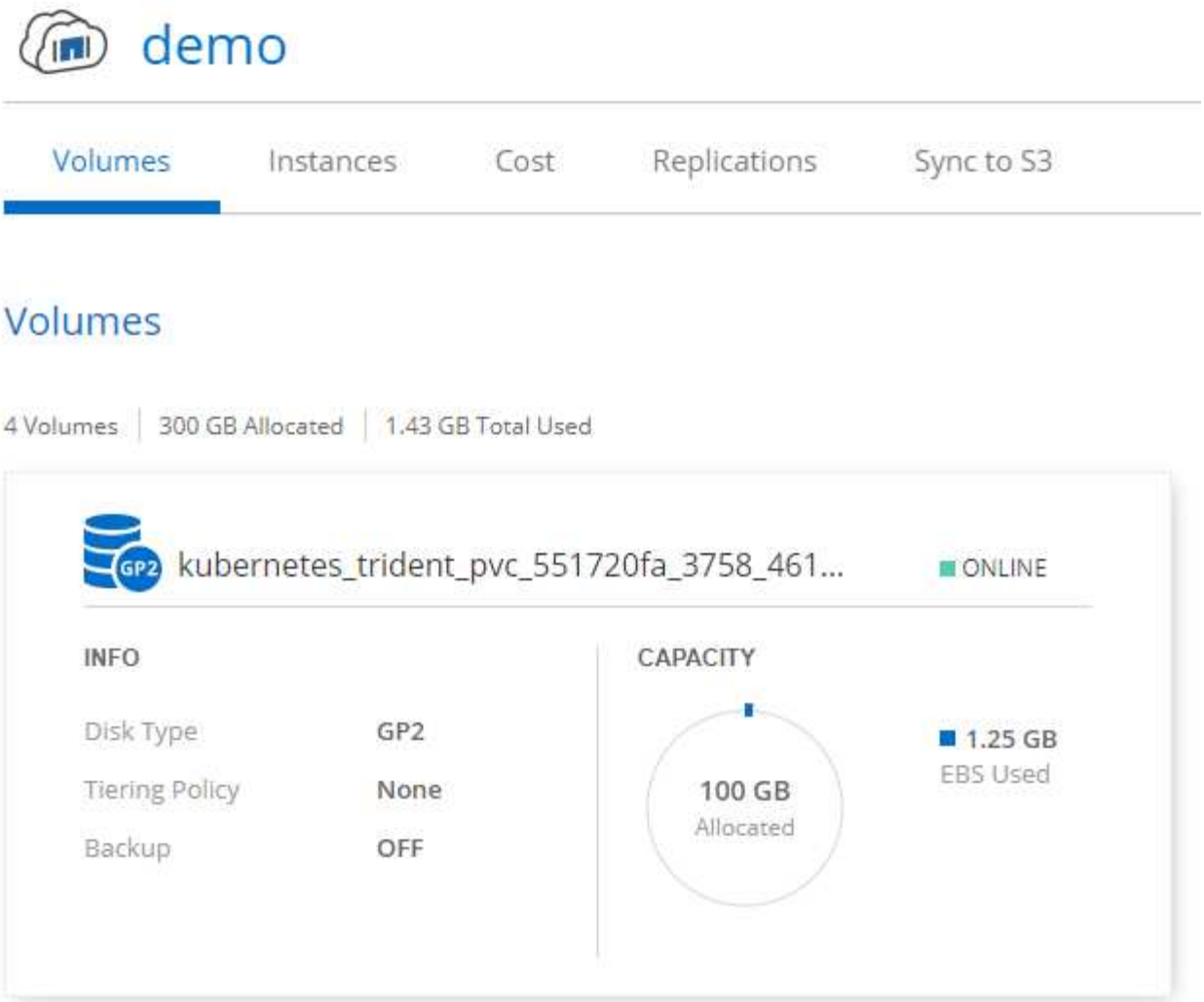
172.31.0.0/16

☑ Set as default storage class

◉ NFS ○ iSCSI

Connect

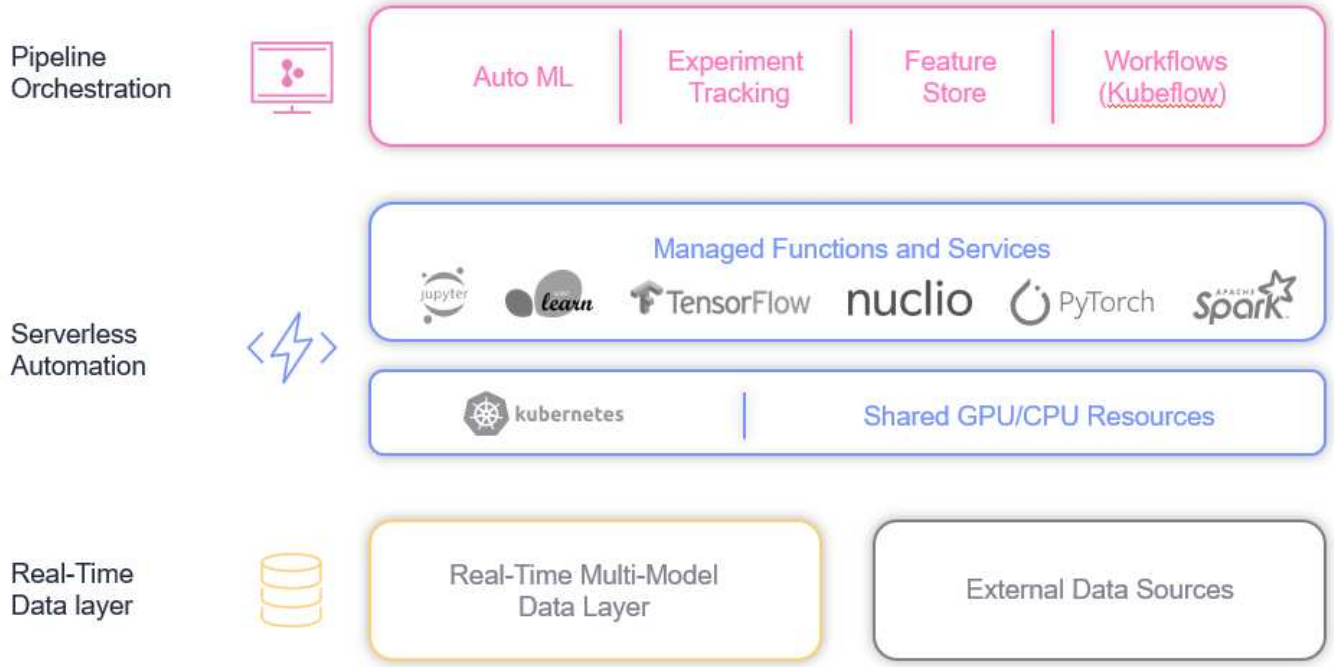
Cancel



Visión General de Iguazio

La Plataforma de Ciencias de datos de Iguazio es una plataforma de ciencia de datos como servicio (PaaS) totalmente integrada y segura que simplifica el desarrollo, acelera el rendimiento, facilita la colaboración y aborda los retos operativos. Esta plataforma incorpora los siguientes componentes, y la Plataforma de Ciencias de datos de Iguazio se presenta en la siguiente imagen:

- Banco de trabajo de ciencia de datos que incluye portátiles Juppyter, motores de análisis integrados y paquetes Python
- Gestión de modelos con seguimiento de experimentos y funcionalidades de canalizaciones automatizadas
- Datos gestionados y servicios ML a través de un clúster de Kubernetes escalable
- Nuclio, un marco de funciones sin servidor en tiempo real
- Una capa de datos extremadamente rápida y segura compatible con SQL, NoSQL, bases de datos de series temporales, archivos (objetos sencillos) y streaming
- Integración con fuentes de datos de terceros como NetApp, Amazon S3, HDFS, bases de datos de SQL y protocolos de transmisión o mensajería
- Paneles en tiempo real basados en Grafana



Requisitos de software y hardware

Configuración de red

A continuación, se muestra el requisito de configuración de red para configurar en el cloud:

- El clúster de Iguazio y los volúmenes cloud de NetApp deben estar en el mismo cloud privado virtual.
- El administrador de la nube debe tener acceso al puerto 6443 en los nodos de la aplicación Iguazio.
- Utilizamos Amazon Web Services en este informe técnico. Sin embargo, los usuarios tienen la opción de poner en marcha la solución en cualquier proveedor de cloud.para las pruebas en las instalaciones en IA de ONTAP con NVIDIA DGX-1, hemos utilizado el servicio DNS alojado en Iguazio para mayor comodidad.

Los clientes deben poder acceder a los dominios DNS creados de forma dinámica. Los clientes pueden utilizar su propio DNS si lo desean.

Requisitos de hardware

Puede instalar Iguazio en sus instalaciones en su propio clúster. Hemos verificado la solución en ONTAP AI de NetApp con un sistema NVIDIA DGX-1. La siguiente tabla enumera el hardware utilizado para probar esta solución.

Hardware subyacente	Cantidad
Sistemas DGX-1	1
Sistema AFF A800 de NetApp	1 pareja de alta disponibilidad (ha), incluye 2 controladoras y 48 SSD NVMe (3,8 TB o superior)
Switches de red Cisco Nexus 3232C	2

En la siguiente tabla se enumeran los componentes de software necesarios para las pruebas in situ:

De NetApp	Versión u otra información
Software de gestión de datos ONTAP de NetApp	9.7
Firmware de switch Cisco NX-OS	7.0(3)I6(1)
SO DGX DE NVIDIA	4.4 - Ubuntu 18.04 LTS
Plataforma contenedora Docker	19.03.5
Versión del contenedor	20.01-tf1-py2
Marco de aprendizaje automático	TensorFlow 1.15.0
Iguazio	La versión es 2.8 o posterior
Servidor ESX	6.5

Esta solución se ha probado totalmente con Iguazio versión 2.5 y NetApp Cloud Volumes ONTAP para AWS. El clúster de Iguazio y el software de NetApp se ejecutan en AWS.

De NetApp	Versión o tipo
Iguazio	La versión es 2.8 o posterior
Nodo de aplicación	M5.4xgrande
Nodo de datos	I3.4xgrande

Resumen de casos de uso de predicción de fallos de dispositivos de red

Este caso de uso se basa en un cliente de Iguazio en el espacio de telecomunicaciones en Asia. Con 100 000 clientes empresariales y 125 000 eventos de interrupción de la red al año, era muy importante predecir y tomar medidas proactivas para evitar que los fallos de red afecten a los clientes. Esta solución ofreció las siguientes ventajas:

- Análisis predictivo para fallos de red
- Integración con un sistema de emisión de boletos
- Tomar medidas proactivas para evitar fallos de la red como resultado de esta implementación de Iguazio, el 60 % de los fallos se impidieron de forma proactiva.

Descripción general de la configuración

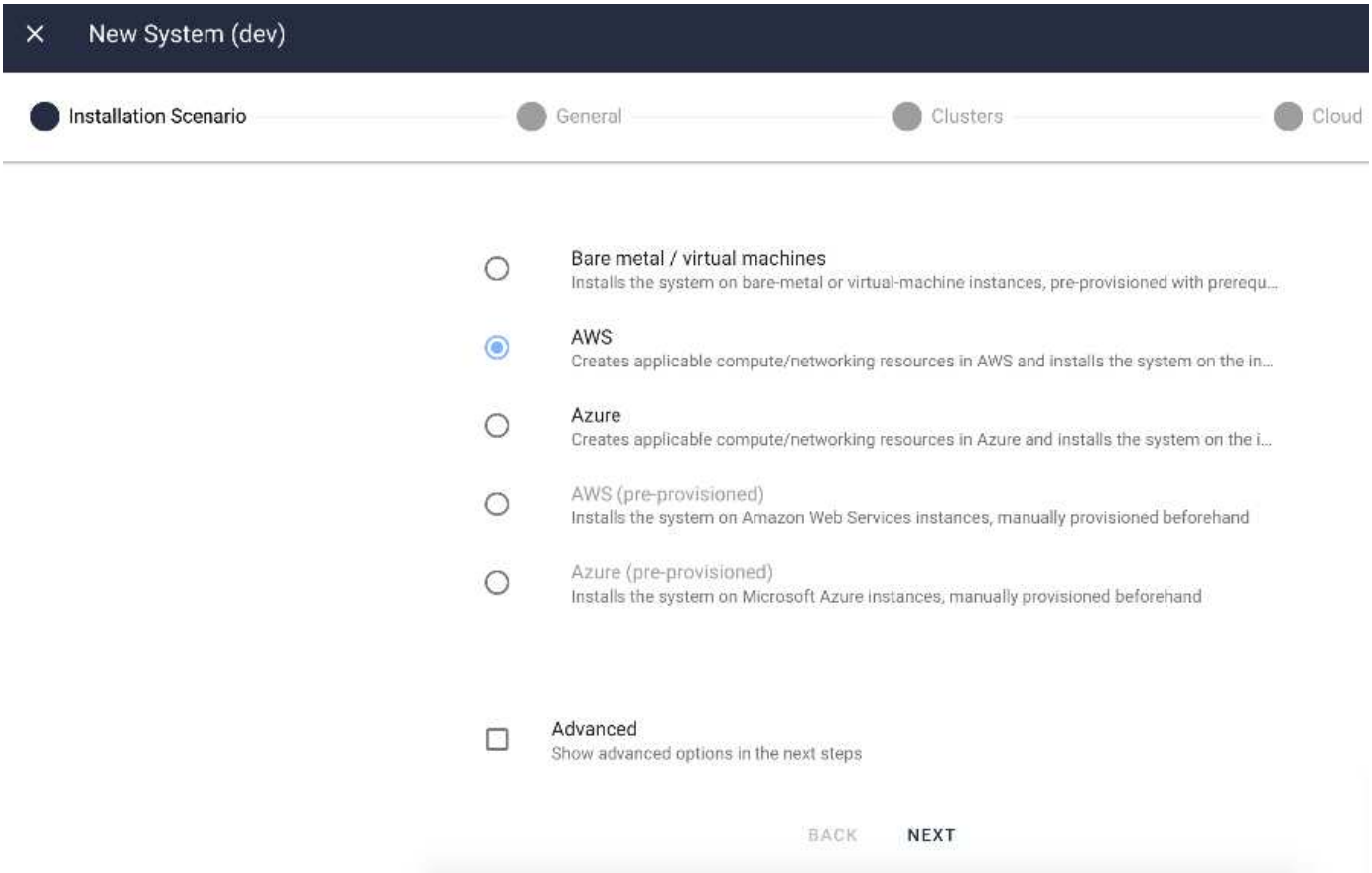
Iguazio se puede instalar de forma local o en un proveedor de cloud.

Instalación de Iguazio

El aprovisionamiento puede realizarse como servicio y gestionado por Iguazio o por el cliente. En ambos casos, Iguazio ofrece una aplicación de despliegue (Proventio) para implementar y gestionar clústeres.

Para la instalación en las instalaciones, consulte "NVA-1121" para la configuración de computación, red y almacenamiento. La implantación en las instalaciones de Iguazio es proporcionada por Iguazio sin costes adicionales para el cliente. Consulte "esta página" Para configuraciones de servidores DNS y SMTP. La

página de instalación de Proventio se muestra a continuación.



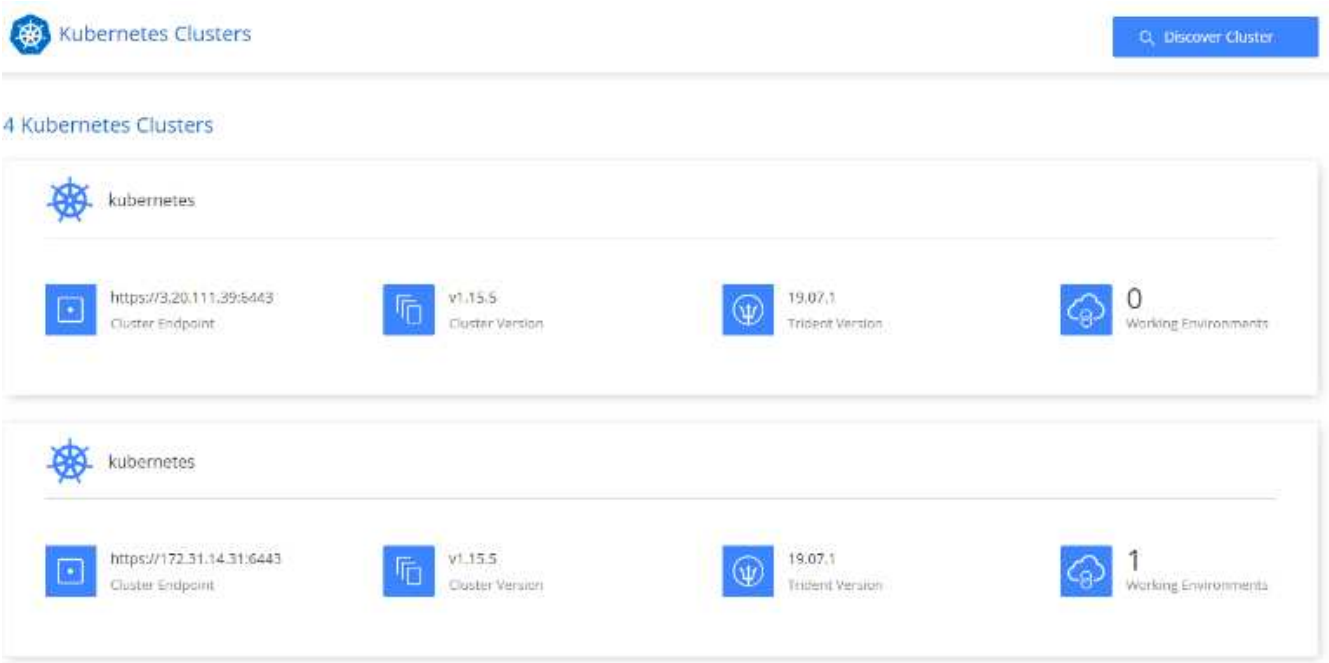
Configurar Kubernetes Cluster

Esta sección se divide en dos partes para la puesta en marcha en el cloud y en las instalaciones, respectivamente.

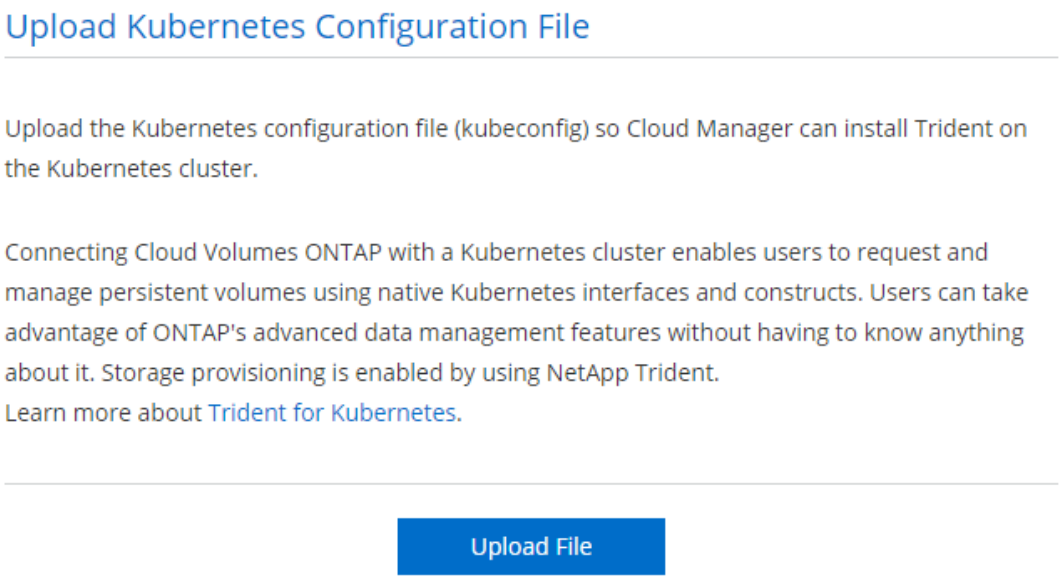
Configuración de Kubernetes de puesta en marcha del cloud

A través de Cloud Manager de NetApp, puede definir la conexión al clúster de Kubernetes del Iguazio. Trident requiere acceso a varios recursos en el clúster para que el volumen esté disponible.

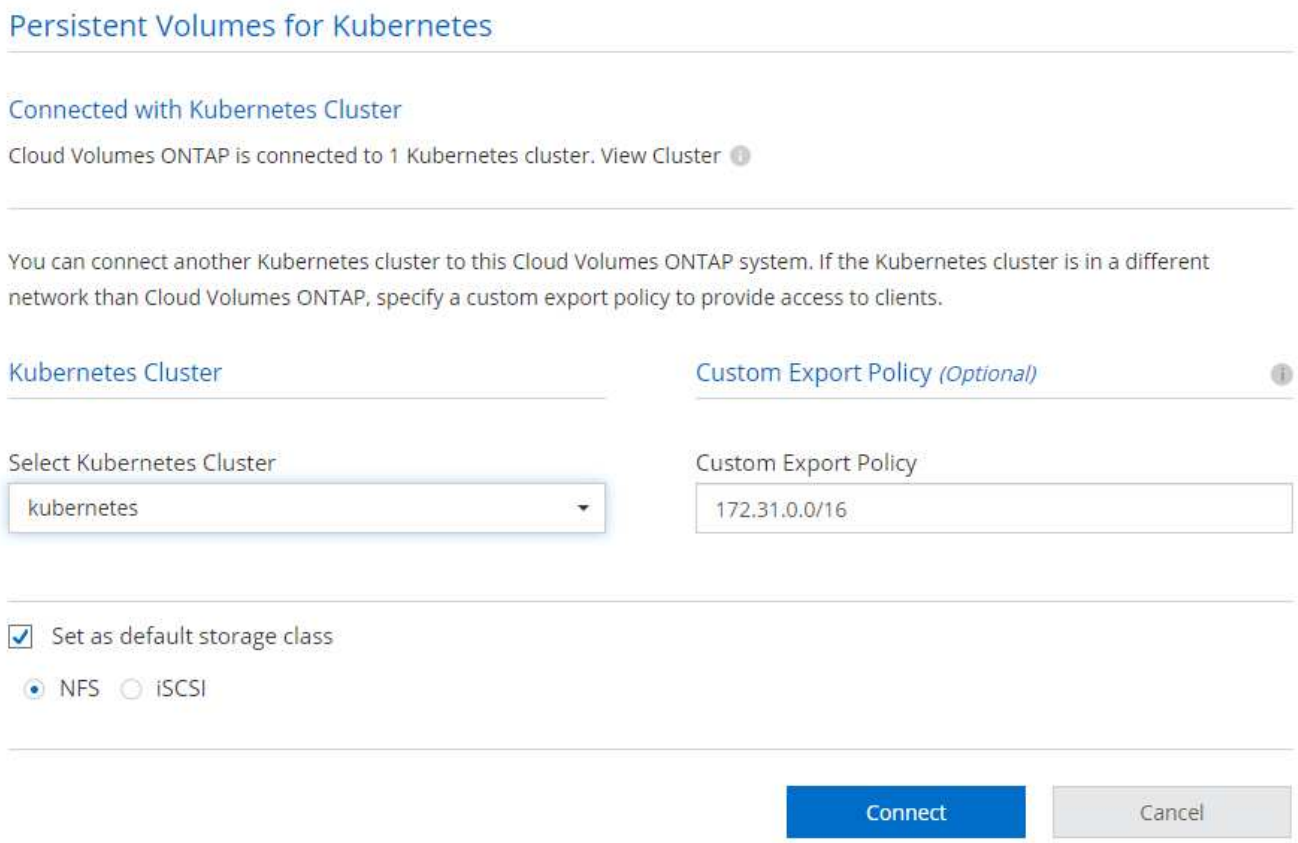
1. Para habilitar el acceso, obtenga el archivo de configuración de Kubernetes desde uno de los nodos de Iguazio. El archivo está ubicado en `/home/Iguazio/.kube/config`. Descargue este archivo en su escritorio.
2. Vaya a detectar clúster para configurar.



3. Cargue el archivo de configuración de Kubernetes. Consulte la siguiente imagen.



4. Ponga en marcha Trident y asocie un volumen con el clúster. Vea la siguiente imagen sobre la definición y asignación de un volumen persistente al clúster de Iguazio. Este proceso crea un volumen persistente (PV) en el clúster Kubernetes de Iguazio. Antes de poder usarlo, debe definir una solicitud de volumen persistente (PVC).



Configuración de Kubernetes de puesta en marcha en las instalaciones

Para ver la instalación local de Trident de NetApp, consulte "TR-4798" para obtener más detalles. Tras configurar su clúster de Kubernetes e instalar Trident de NetApp, puede conectar Trident al clúster Iguazio para habilitar las capacidades de gestión de datos de NetApp, como realizar copias Snapshot de sus datos y su modelo.

Defina una solicitud de volumen persistente

1. Guarde el siguiente YAML en un archivo para crear un PVC del tipo Basic.

```
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: basic
spec:
  accessModes:
    - ReadWriteOnce
  resources:
    requests:
      storage: 100Gi
  storageClassName: netapp-file
```

2. Aplique el archivo YLMA al clúster de Kubernetes de Iguazio.

```
Kubectl -n default-tenant apply -f <your yaml file>
```

Adjunte el volumen de NetApp al ordenador portátil Jupyter

Iguazio ofrece varios servicios gestionados para proporcionar a los científicos de datos una pila completa para el desarrollo y la puesta en marcha de aplicaciones de IA/ML. Puede obtener más información sobre estos componentes en la "Visión General de Iguazio de Servicios y Herramientas de aplicación".

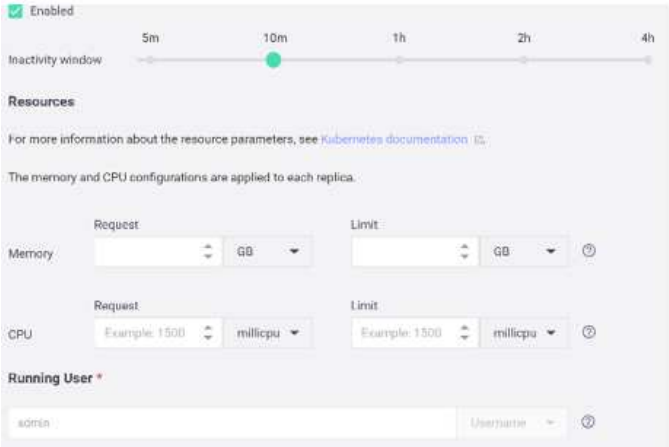
Uno de los servicios gestionados es el portátil Jupyter. Cada desarrollador obtiene su propia implementación de un contenedor de portátiles con los recursos que necesitan para el desarrollo. Para darles acceso a Cloud Volume de NetApp, puede asignar el volumen a su contenedor y la asignación de recursos, la configuración de usuario en ejecución y la variable de entorno para las reclamaciones de volumen persistente se presentan en la siguiente imagen.

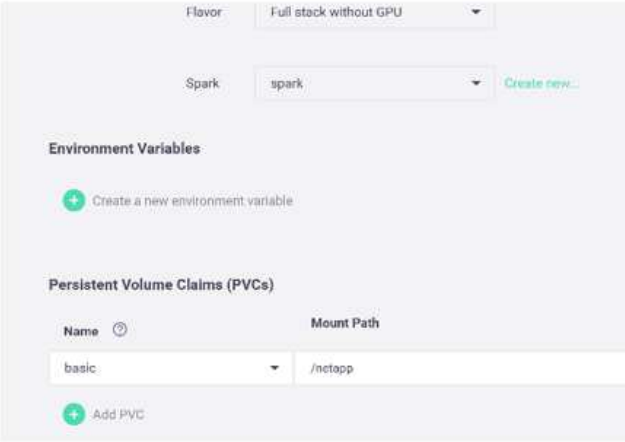
Para obtener información sobre una configuración en las instalaciones, puede consultar "TR-4798" En la configuración de Trident para habilitar las funcionalidades de gestión de datos de ONTAP de NetApp, como realizar copias Snapshot de sus datos o modelos para el control de versiones. Añada la siguiente línea en su archivo de configuración de back-end de Trident para que sean visibles los directorios de Snapshot:

```
{
    …
    "defaults": {
        "snapshotDir": "true"
    }
}
```

Debe crear un archivo de configuración back-end de Trident en formato JSON y, a continuación, ejecutar lo siguiente "Comando Trident" como referencia:

```
tridentctl create backend -f <backend-file>
```





Implementación de la aplicación

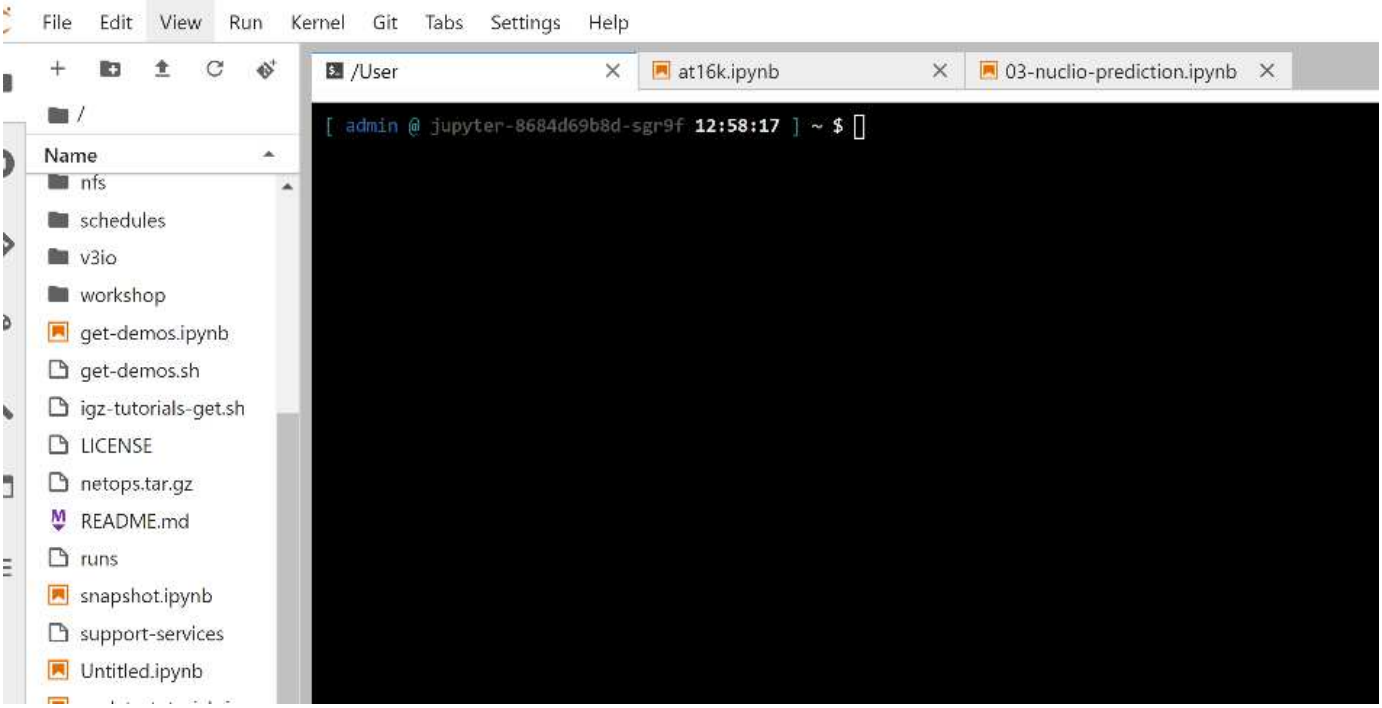
En las siguientes secciones se describe cómo instalar e implementar la aplicación.

Obtenga el código de GitHub

Ahora que el volumen de cloud de NetApp o el volumen de Trident de NetApp están disponibles para el clúster de Iguazio y el entorno de desarrolladores, puede comenzar a revisar la aplicación.

Los usuarios tienen su propio espacio de trabajo (directorio). En cada notebook, la ruta al directorio de usuario es `/User`. La plataforma Iguazio administra el directorio. Si sigue las instrucciones anteriores, el volumen de cloud de NetApp está disponible en `/netapp` directorio.

Obtenga el código de GitHub con una terminal Juppyter.



En el prompt de la terminal de Juppyter, clone el proyecto.

```
cd /User
git clone .
```

Ahora debería ver la `netops- netapp` Carpeta en el árbol de archivos del espacio de trabajo Juppyter.

Configurar entorno de trabajo

Copie el `Notebook set_env-Example.ipynb` como `set_env.ipynb`. Abrir y editar `set_env.ipynb`. Este cuaderno establece variables para las credenciales, ubicaciones de archivos y controladores de ejecución.

Si sigue las instrucciones anteriores, los pasos siguientes son los únicos cambios que se deben realizar:

1. Obtenga este valor desde el panel de servicios de Iguazio: `docker_registry`

   Ejemplo: `docker-registry.default-tenant.app.clusterq.iguaziodev.com:80`

2. Cambiar `admin` A su nombre de usuario de Iguazio:

   ```
   IGZ_CONTAINER_PATH = '/users/admin'
   ```

   A continuación, se muestran los detalles de conexión del sistema ONTAP. Incluya el nombre del volumen que se generó al instalar Trident. La siguiente configuración es para un clúster ONTAP en las instalaciones:

   ```
   ontapClusterMgmtHostname = '0.0.0.0'
   ontapClusterAdminUsername = 'USER'
   ontapClusterAdminPassword = 'PASSWORD'
   sourceVolumeName = 'SOURCE VOLUME'
   ```

   La siguiente configuración es para Cloud Volumes ONTAP:

```
MANAGER=ontapClusterMgmtHostname
svm='svm'
email='email'
password=ontapClusterAdminPassword
weid="weid"
volume=sourceVolumeName
```

Cree imágenes Docker básicas

Todo lo que necesita para construir un ducto ML está incluido en la plataforma Iguazio. El desarrollador puede definir las especificaciones de las imágenes Docker necesarias para ejecutar la canalización y ejecutar la creación de imágenes desde el portátil Jupyter. Abra el portátil `create- images.ipynb` Y ejecute todas las celdas.

Este cuaderno crea dos imágenes que utilizamos en la canalización.

- `iguazio/netapp`. Se utiliza para manejar tareas ML.

  Create image for training pipeline

```
[4]: fn.build_config(image=docker_registry+'/iguazio/netapp', commands=['pip install \
     v3io_frames fsspec>=0.3.3 PyYAML==5.1.2 pyarrow==0.15.1 pandas==0.25.3 matplotlib seaborn yellowb
     fn.deploy()
```

- `netapp/pipeline`. Contiene utilidades para gestionar las copias snapshot de NetApp.

Create image for Ontap utilitites

```
[9]: fn.build_config(image=docker_registry + '/netapp/pipeline:latest',commands=['apt -y update','pip install v3io_frames netapp_ontap'
     fn.deploy()
```

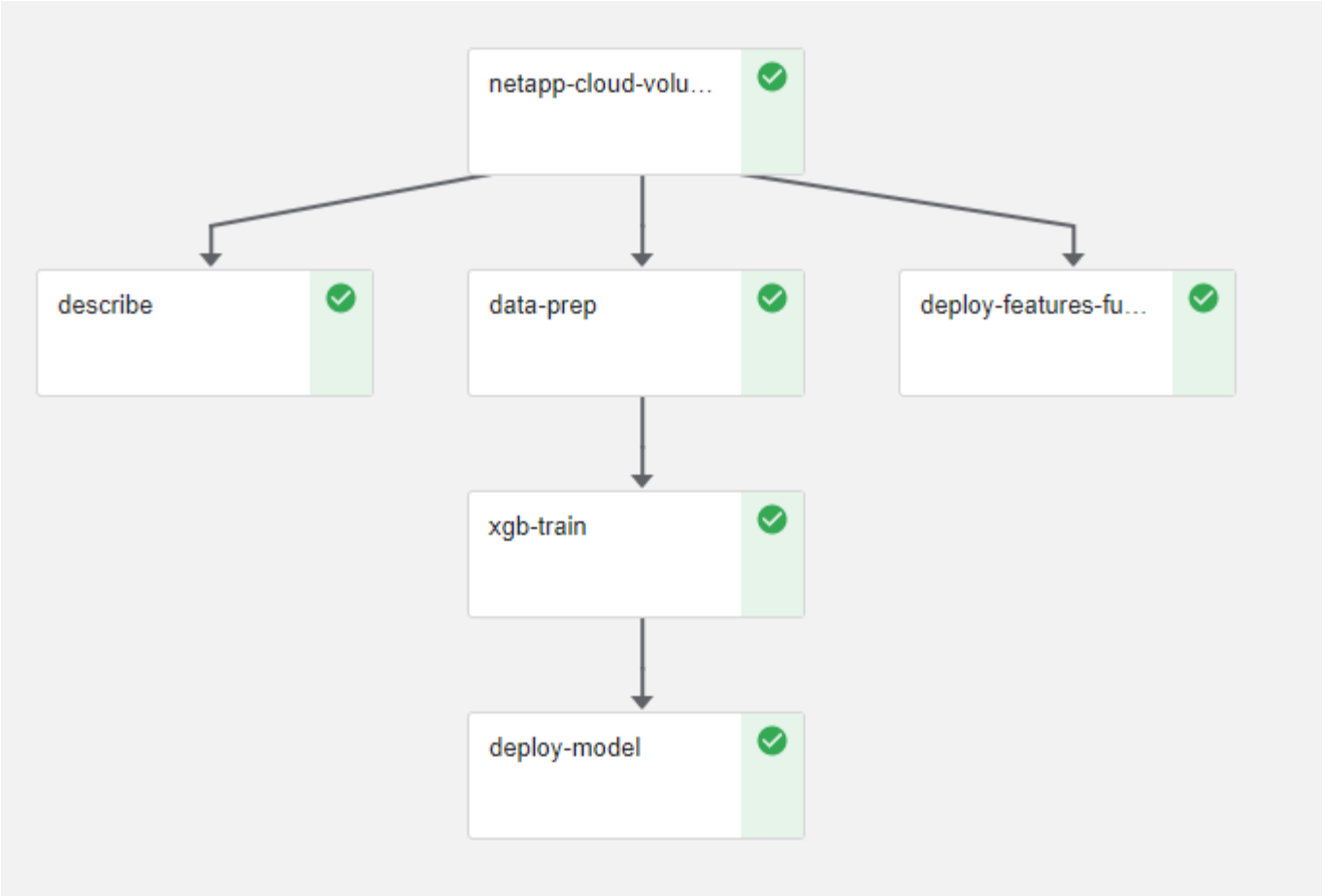
Revise los cuadernos Juppyter individuales

En la siguiente tabla se enumeran las bibliotecas y los marcos que utilizamos para crear esta tarea. Todos estos componentes se han integrado plenamente con los controles de seguridad y acceso basados en funciones de Iguazio.

Bibliotecas/Marco	Descripción
MLRun	Un gestionado por Iguazio para permitir el montaje, ejecución y supervisión de una canalización ML/IA.
Nuclio	Un marco de funciones sin servidor integrado con Iguazio. También disponible como proyecto de código abierto gestionado por Iguazio.
Kubeflow	Un marco basado en Kubernetes para poner en marcha la canalización. Este es también un proyecto de código abierto al que colabora Iguazio. Se integra con Iguazio para una mayor seguridad e integración con el resto de la infraestructura.
Docker	Un registro Docker se ejecuta como servicio en la plataforma Iguazio. También puede cambiar esta opción para conectarse al registro.
Cloud Volumes de NetApp	Cloud Volumes que se ejecuta en AWS nos proporciona acceso a grandes cantidades de datos y la capacidad de realizar copias de Snapshot para versiones de los conjuntos de datos utilizados para el entrenamiento.
Trident	Trident es un proyecto de código abierto gestionado por NetApp. Facilita la integración con los recursos de almacenamiento y computación en Kubernetes.

Hemos utilizado varios portátiles para construir la canalización DE ML. Cada portátil puede probarse individualmente antes de ser reunido en la tubería. Cubrimos cada portátil individualmente tras el flujo de despliegue de esta aplicación de demostración.

El resultado deseado es una canalización que entrena un modelo basado en una copia Snapshot de los datos y pone en marcha el modelo para la inferencia. En la siguiente imagen se muestra un diagrama de bloque de una canalización MLRun completada.



Ponga en marcha la función de generación de datos

En esta sección se describe cómo utilizamos las funciones de Nuclio serverless para generar datos de dispositivos de red. El caso de uso se adapta a partir de un cliente Iguazio que ha implementado la canalización y utiliza servicios Iguazio para supervisar y predecir los fallos del dispositivo de red.

Simulamos los datos procedentes de dispositivos de red. Ejecución del cuaderno Juppyter `data-generator.ipynb` Crea una función sin servidor que se ejecuta cada 10 minutos y genera un archivo Parquet con nuevos datos. Para implementar la función, ejecute todas las celdas de este portátil. Consulte "Sitio Web de Nuclio" para revisar cualquier componente desconocido en este cuaderno.

Al generar la función, se ignora una celda con el siguiente comentario. Se asume que todas las celdas del portátil forman parte de la función. Importe el módulo Nuclio para activarlo `%nuclio magic`.

```
# nuclio: ignore
import nuclio
```

En la especificación para la función, definimos el entorno en el que se ejecuta la función, cómo se activa y los recursos que consume.

```
spec = nuclio.ConfigSpec(config={"spec.triggers.inference.kind":"cron",

"spec.triggers.inference.attributes.interval" :"10m",
                                "spec.readinessTimeoutSeconds" : 60,
                                "spec.minReplicas" : 1},……
```

La `init_context` La función es invocada por el marco Nuclio tras la inicialización de la función.

```
def init_context(context):
    ….
```

Cualquier código que no esté en una función se invoca cuando se inicializa la función. Cuando lo invoca, se ejecuta una función de controlador. Puede cambiar el nombre del controlador y especificarlo en la especificación de función.

```
def handler(context, event):
            …
```

Puede probar la función desde el portátil antes de la implementación.

```
%%time
# nuclio: ignore
init_context(context)
event = nuclio.Event(body='')
output = handler(context, event)
output
```

La función puede desplegarse desde el portátil o puede desplegarse a partir de una canalización CI/CD (adaptando este código).

```
addr = nuclio.deploy_file(name='generator',project='netops',spec=spec,
tag='v1.1')
```

Cuadernos de pipeline

Estos cuadernos no están diseñados para ejecutarse individualmente para esta configuración. Esto es sólo una revisión de cada bloc de notas. Los invocamos como parte del proyecto. Para ejecutarlas individualmente, revise la documentación MLRun para ejecutarlas como trabajos de Kubernetes.

snap_cv.ipynb

Este portátil gestiona las copias snapshot de Cloud Volume al principio de la canalización. Pasa el nombre del volumen al contexto de la canalización. Este cuaderno invoca un script de shell para manejar la copia

Snapshot. Mientras se ejecuta en la canalización, el contexto de ejecución contiene variables que ayudan a localizar todos los archivos necesarios para su ejecución. Mientras escribe este código, el desarrollador no tiene que preocuparse por la ubicación del archivo en el contenedor que lo ejecuta. Como se describe más tarde, esta aplicación se implementa con todas sus dependencias y es la definición de los parámetros de canalización que proporciona el contexto de ejecución.

```
command = os.path.join(context.get_param('APP_DIR'),"snap_cv.sh")
```

La ubicación de la copia Snapshot creada se coloca en el contexto de MLRun que consumirán los pasos de la canalización.

```
context.log_result('snapVolumeDetails',snap_path)
```

Los siguientes tres portátiles se ejecutan en paralelo.

data-prep.ipynb

Las métricas sin formato deben convertirse en funciones para permitir el entrenamiento de modelos. Este cuaderno lee las métricas sin formato del directorio Snapshot y escribe las funciones de entrenamiento de modelos en el volumen de NetApp.

Cuando se ejecuta en el contexto de la canalización, la entrada `DATA_DIR` Contiene la ubicación de la copia Snapshot.

```
metrics_table = os.path.join(str(mlruncontext.get_input('DATA_DIR',
os.getenv('DATA_DIR','/netpp'))),
                             mlruncontext.get_param('metrics_table',
os.getenv('metrics_table','netops_metrics_parquet')))
```

describa.ipynb

Para visualizar las métricas entrantes, implementamos un paso de canalización que proporciona gráficos y gráficos disponibles a través de las interfaces de usuario de Kubeflow y MLRun. Cada ejecución tiene su propia versión de esta herramienta de visualización.

```
ax.set_title("features correlation")
plt.savefig(os.path.join(base_path, "plots/corr.png"))
context.log_artifact(PlotArtifact("correlation",  body=plt.gcf()),
local_path="plots/corr.html")
```

deploy-feature-function.ipynb

Supervisamos continuamente las métricas en busca de anomalías. Este bloc de notas crea una función sin servidor que genera las funciones que necesitan ejecutar la predicción en las métricas entrantes. Este cuaderno invoca la creación de la función. El código de función se encuentra en el portátil `data-prep.ipynb`. Observe que utilizamos el mismo bloc de notas como un paso en la tubería para este propósito.

training.ipynb

Una vez que creamos las funciones, activamos la formación del modelo. El resultado de este paso es el modelo que se va a utilizar para la inferencia. También recopilamos estadísticas para realizar un seguimiento de cada ejecución (experimento).

Por ejemplo, el siguiente comando introduce la puntuación de precisión en el contexto de ese experimento. Este valor es visible en Kubeflow y MLRun.

```
context.log_result('accuracy',score)
```

despliegue-inferencia-function.ipynb

El último paso de la canalización es poner en marcha el modelo como una función sin servidor para la inferencia continua. Este cuaderno invoca la creación de la función sin servidor definida en `nuclio-inference- function.ipynb`.

Revisar y crear el canalización

La combinación de ejecutar todos los portátiles en una tubería permite que la ejecución continua de experimentos reevalúe la precisión del modelo con las nuevas métricas. En primer lugar, abra la `pipeline.ipynb` portátil. Le repasamos por los detalles que muestran cómo NetApp y Iguazio simplifican la puesta en marcha de esta canalización DE ML.

Utilizamos MLRun para proporcionar contexto y manejar la asignación de recursos a cada paso de la canalización. El servicio de API MLRun se ejecuta en la plataforma Iguazio y es el punto de interacción con los recursos de Kubernetes. Cada desarrollador no puede solicitar recursos directamente; la API gestiona las solicitudes y habilita los controles de acceso.

```
# MLRun API connection definition
mlconf.dbpath = 'http://mlrun-api:8080'
```

La canalización puede funcionar con Cloud Volumes de NetApp y los volúmenes en las instalaciones. Hemos creado esta demostración para usar Cloud Volumes, pero puede ver en el código la opción para ejecutarse en las instalaciones.

```
# Initialize the NetApp snap fucntion once for all functions in a notebook
if [ NETAPP_CLOUD_VOLUME ]:
    snapfn =
code_to_function('snap',project='NetApp',kind='job',filename="snap_cv.ipyn
b").apply(mount_v3io())
    snap_params = {
    "metrics_table" : metrics_table,
    "NETAPP_MOUNT_PATH" : NETAPP_MOUNT_PATH,
    'MANAGER' : MANAGER,
    'svm' : svm,
    'email': email,
    'password': password ,
    'weid': weid,
    'volume': volume,
    "APP_DIR" : APP_DIR
       }
else:
    snapfn =
code_to_function('snap',project='NetApp',kind='job',filename="snapshot.ipy
nb").apply(mount_v3io())
….
snapfn.spec.image = docker_registry + '/netapp/pipeline:latest'
snapfn.spec.volume_mounts =
[snapfn.spec.volume_mounts[0],netapp_volume_mounts]
      snapfn.spec.volumes = [ snapfn.spec.volumes[0],netapp_volumes]
```

La primera acción necesaria para convertir un cuaderno Juppyter en un paso de Kubeflow es convertir el código en una función. Una función tiene todas las especificaciones necesarias para ejecutar ese portátil. A medida que se desplaza hacia abajo por el bloc de notas, puede ver que definimos una función para cada paso de la canalización.

Parte del portátil	Descripción
<code_to_function> (parte del módulo MLRun)	Nombre de la función: Nombre del proyecto. se utiliza para organizar todos los artefactos del proyecto. Esto es visible en la interfaz de usuario de MLRun. Amable. En este caso, un trabajo de Kubernetes. Esto podría ser DASK, mpi, sparkk8s, y más. Consulte la documentación de MLRun para obtener más detalles. Archivo. El nombre del portátil. También puede ser una ubicación en Git (HTTP).
imagen	El nombre de la imagen de Docker que estamos utilizando para este paso. Lo hemos creado anteriormente con el bloc de notas create-image.ipynb.
montajes_volúmenes y volúmenes	Detalles para montar el Cloud Volume de NetApp en tiempo de ejecución.

También definimos parámetros para los pasos.

```
params={   "FEATURES_TABLE":FEATURES_TABLE,
           "SAVE_TO" : SAVE_TO,
           "metrics_table" : metrics_table,
           'FROM_TSDB': 0,
           'PREDICTIONS_TABLE': PREDICTIONS_TABLE,
           'TRAIN_ON_LAST': '1d',
           'TRAIN_SIZE':0.7,
           'NUMBER_OF_SHARDS' : 4,
           'MODEL_FILENAME' : 'netops.v3.model.pickle',
           'APP_DIR' : APP_DIR,
           'FUNCTION_NAME' : 'netops-inference',
           'PROJECT_NAME' : 'netops',
           'NETAPP_SIM' : NETAPP_SIM,
           'NETAPP_MOUNT_PATH': NETAPP_MOUNT_PATH,
           'NETAPP_PVC_CLAIM' : NETAPP_PVC_CLAIM,
           'IGZ_CONTAINER_PATH' : IGZ_CONTAINER_PATH,
           'IGZ_MOUNT_PATH' : IGZ_MOUNT_PATH
            }
```

Después de tener la definición de función para todos los pasos, puede construir la canalización. Utilizamos la `kfp` módulo para realizar esta definición. La diferencia entre el uso de MLRun y la construcción por su cuenta es la simplificación y el acortamiento de la codificación.

Las funciones definidas se convierten en componentes de pasos mediante el `as_step` Función de MLRun.

Definición de paso de instantánea

Inicie una función Snapshot, la salida y el montaje v3io como fuente:

```
snap = snapfn.as_step(NewTask(handler='handler',params=snap_params),
name='NetApp_Cloud_Volume_Snapshot',outputs=['snapVolumeDetails','training
_parquet_file']).apply(mount_v3io())
```

Parámetros	Detalles
Nueva tarea	Newtask es la definición de la ejecución de la función.
(Módulo MLRun)	Manipulador. Nombre de la función Python que se va a invocar. Utilizamos el controlador de nombres en el portátil, pero no es necesario. parámetros. Los parámetros que pasamos a la ejecución. Dentro de nuestro código, utilizamos context.get_param («PARAMETER») para obtener los valores.

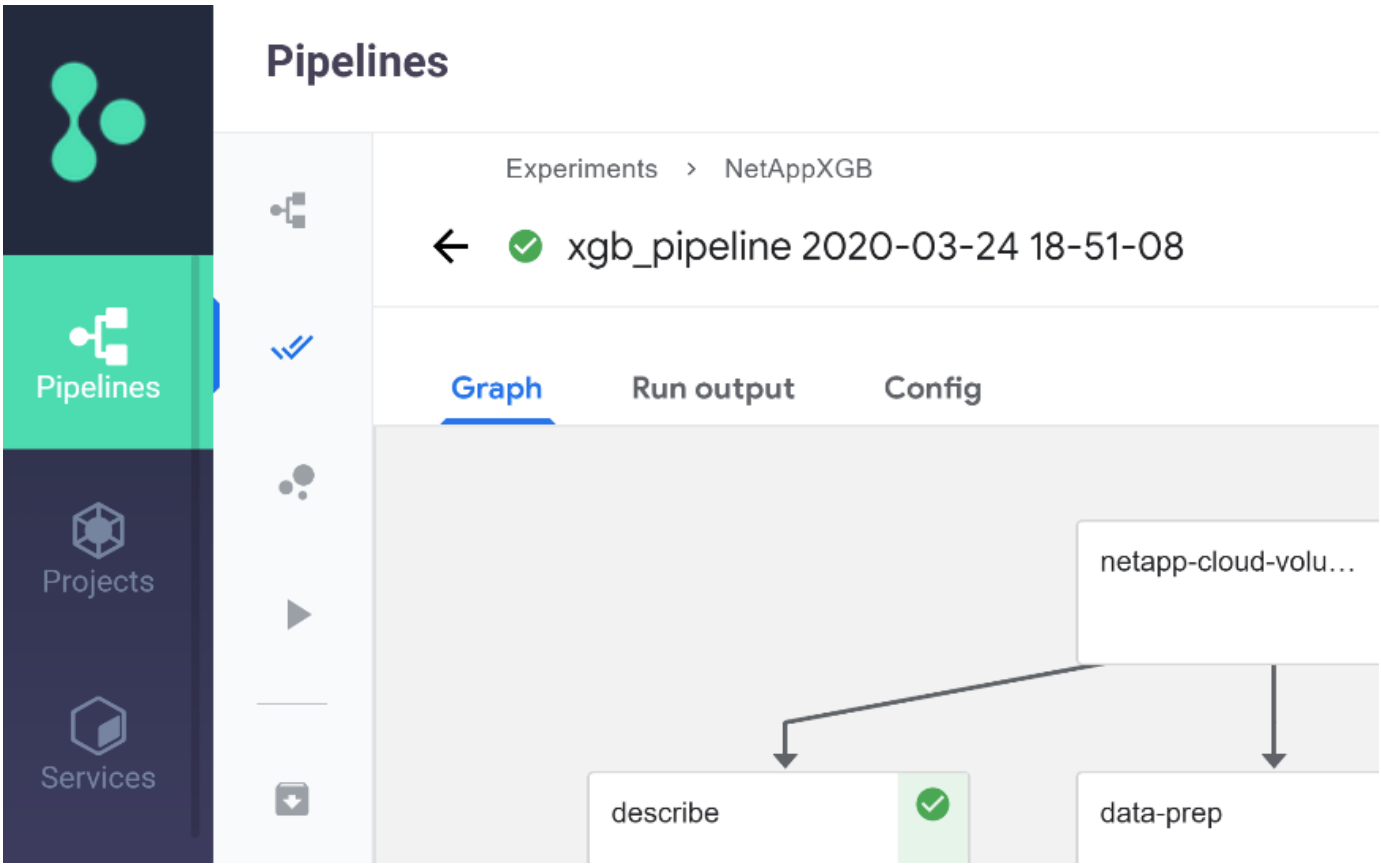
Parámetros	Detalles
paso_as	Nombre. Nombre del paso de la canalización de Kubeflow. salidas. Estos son los valores que el paso agrega al diccionario al terminar. Eche un vistazo al portátil SNAP_cv.ipynb. mount_v3io(). Esto configura el paso para montar /User para el usuario que ejecuta la canalización.

```
prep = data_prep.as_step(name='data-prep',
handler='handler',params=params,
                          inputs = {'DATA_DIR':
snap.outputs['snapVolumeDetails']} ,

out_path=artifacts_path).apply(mount_v3io()).after(snap)
```

Parámetros	Detalles
entradas	Puede pasar a un paso las salidas de un paso anterior. En este caso, snap.outsits[napVolumeDetails] es el nombre de la copia Snapshot que creamos en el paso snap.
ruta de salida	Ubicación para colocar artefactos que generan utilizando el módulo MLRun log_Artifacts.

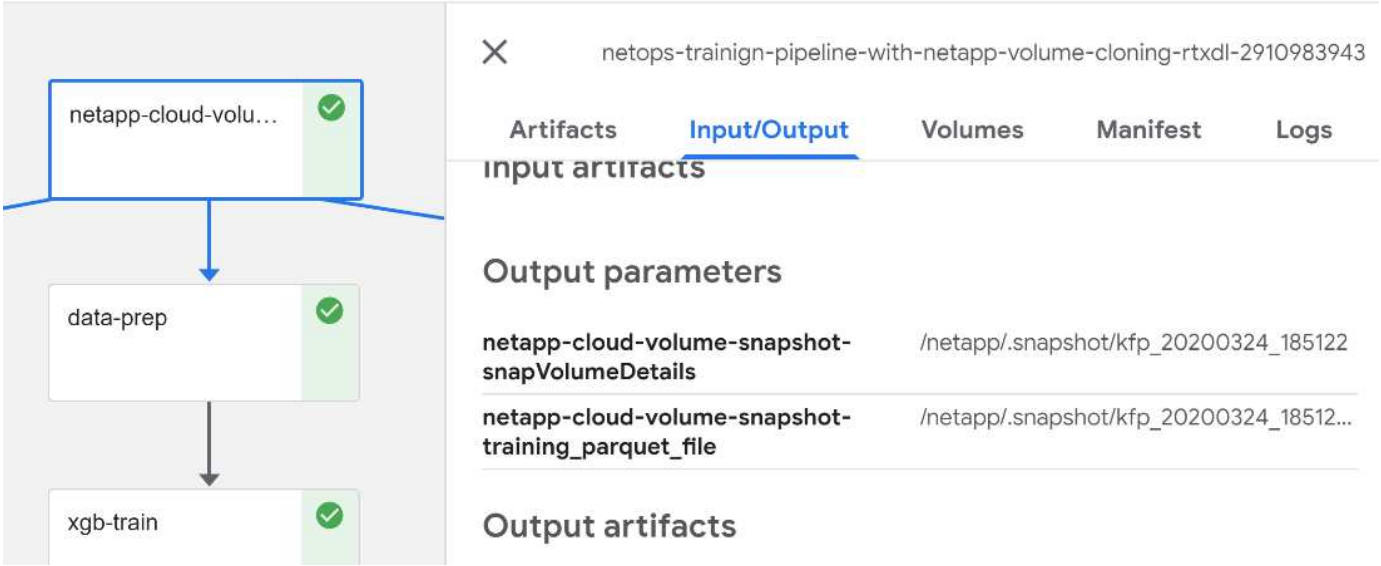
Puede ejecutar `pipeline.ipynb` de arriba a abajo. A continuación, puede ir a la pestaña tuberías desde el panel de control de Iguazio para supervisar el progreso tal y como se ve en la pestaña tuberías del panel de control de Iguazio.



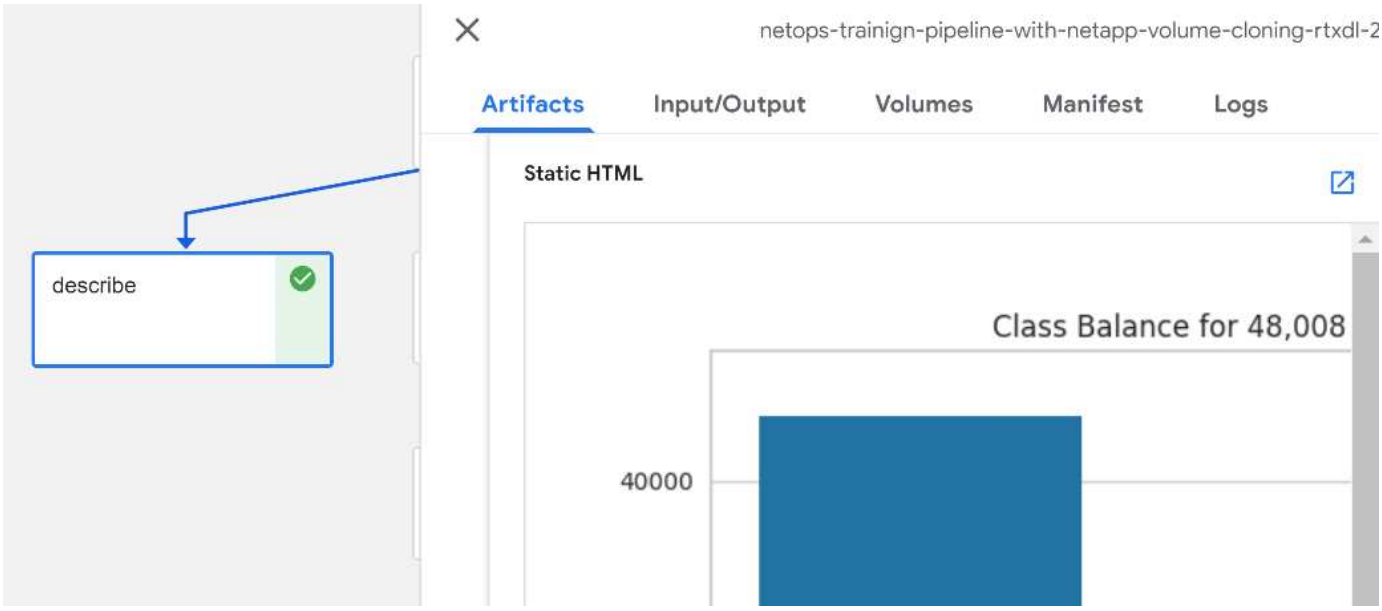
Debido a que hemos registrado la precisión del paso de entrenamiento en cada carrera, tenemos un registro de precisión para cada experimento, como se ve en el registro de precisión de entrenamiento.

Run name	Status	Duration	Pipeline Version	Recurring ...	Start time	accuracy
xgb_pipeline 2020-03-24 18-51-...	✓	0:08:43	[View pipeline]	-	3/24/2020, 2:51:09 PM	0.985
xgb_pipeline 2020-03-19 13-31-...	✓	0:08:14	[View pipeline]	-	3/19/2020, 9:31:19 AM	0.980
xgb_pipeline 2020-03-18 12-56-...	✓	0:08:11	[View pipeline]	-	3/18/2020, 8:56:08 AM	0.990
xgb_pipeline 2020-03-17 19-49-...	✓	0:08:03	[View pipeline]	-	3/17/2020, 3:49:31 PM	0.985
xgb_pipeline 2020-03-17 18-34-...	✓	0:05:54	[View pipeline]	-	3/17/2020, 2:34:56 PM	0.980
xgb_pipeline 2020-03-17 17-34-...	✓	0:04:48	[View pipeline]	-	3/17/2020, 1:34:16 PM	0.982
xgb_pipeline 2020-03-17 17-01-...	✓	0:05:25	[View pipeline]	-	3/17/2020, 1:01:58 PM	0.987
xgb_pipeline 2020-03-16 16-47-...	✓	0:06:08	[View pipeline]	-	3/16/2020, 12:47:19 ...	0.983
xgb_pipeline 2020-03-16 13-57-...	✓	0:05:18	[View pipeline]	-	3/16/2020, 9:57:03 AM	0.980

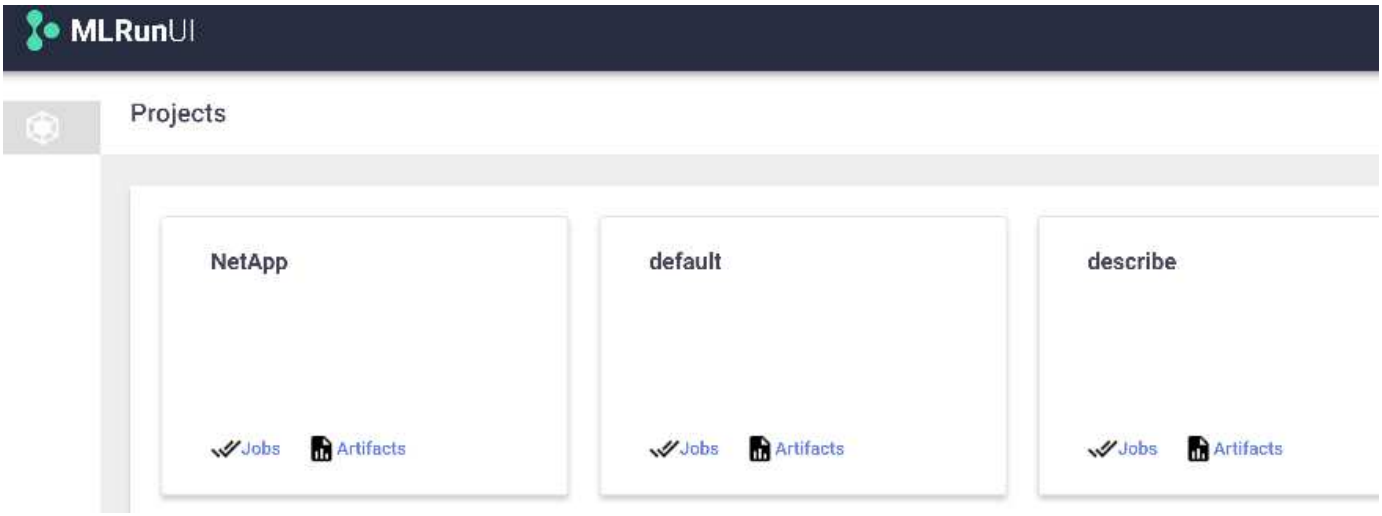
Si selecciona el paso Snapshot, puede ver el nombre de la copia Snapshot que se utilizó para ejecutar este experimento.



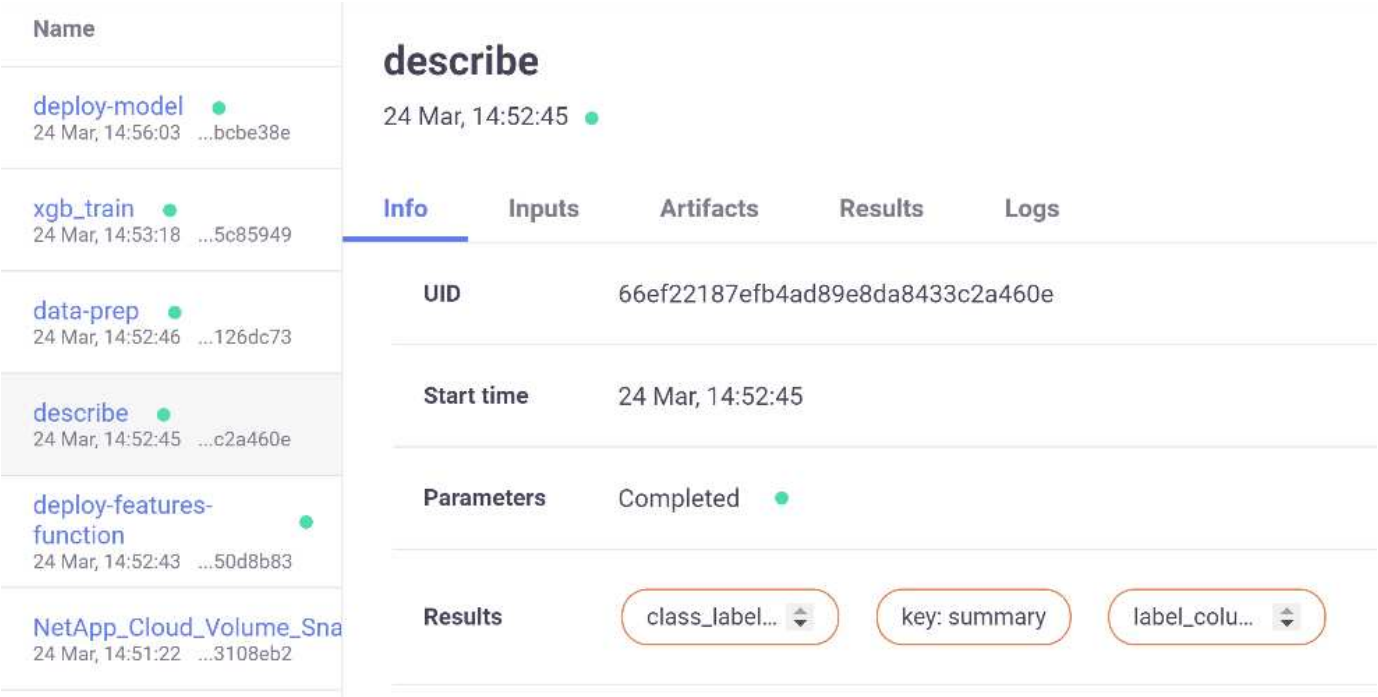
El paso descrito tiene artefactos visuales para explorar las métricas que utilizamos. Puede expandir para ver el trazado completo como se ve en la siguiente imagen.



La base de datos de la API MLRun también realiza un seguimiento de las entradas, salidas y artefactos de cada ejecución organizada por el proyecto. En la siguiente imagen se puede ver un ejemplo de entradas, salidas y artefactos para cada secuencia.



Para cada trabajo, almacenamos detalles adicionales.



Hay más información sobre MLRun que podemos cubrir en este documento. Los artefactos de al, incluida la definición de los pasos y las funciones, se pueden guardar en la base de datos de API, con versiones e invocados individualmente o como un proyecto completo. Los proyectos también se pueden guardar e insertar en Git para su uso posterior. Le animamos a obtener más información en la "Sitio de MLRun GitHub".

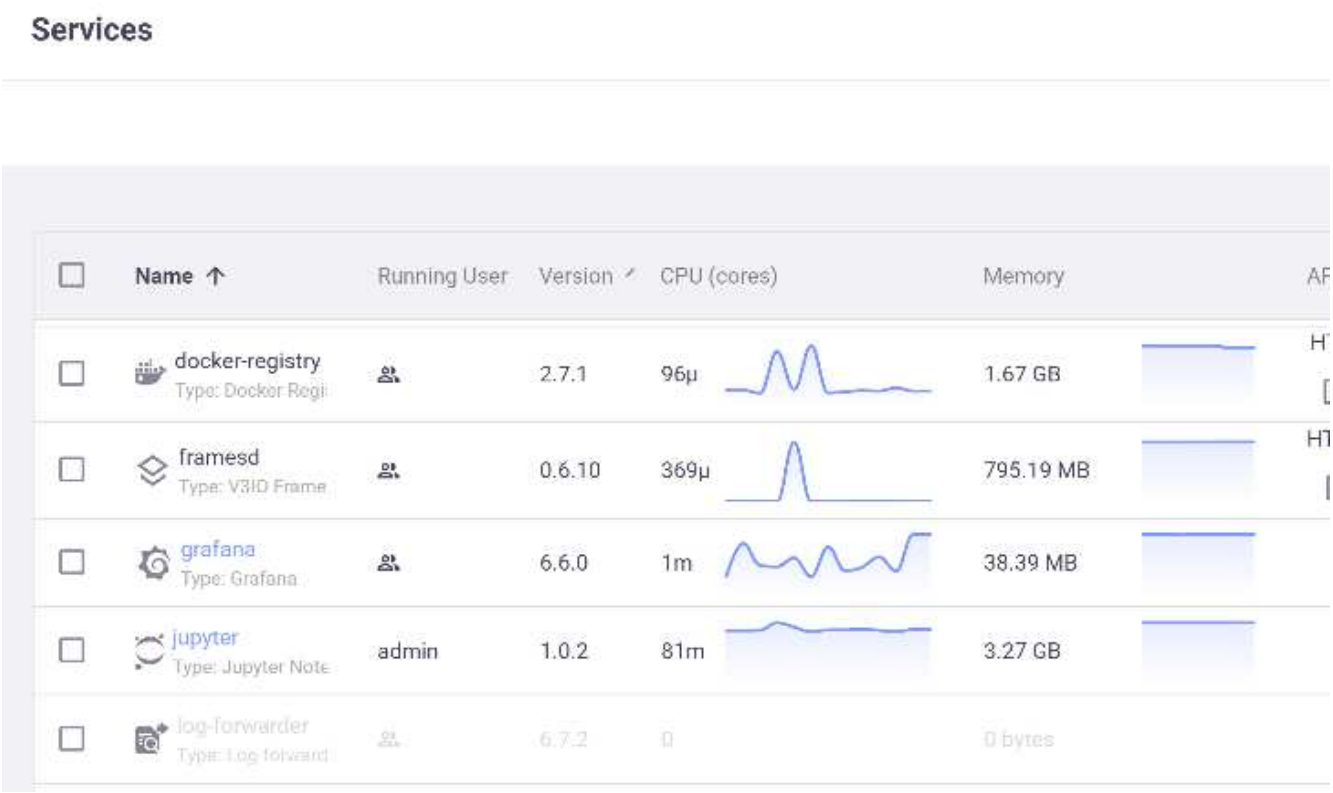
Implemente el panel de Grafana

Una vez que todo se pone en marcha, ejecutamos inferencias sobre nuevos datos. Los modelos predicen fallos en el equipo de dispositivo de red. Los resultados de la predicción se almacenan en una tabla de timbres de Iguazio. Puede visualizar los resultados con Grafana en la plataforma integrada con la política de acceso a datos y seguridad de Iguazio.

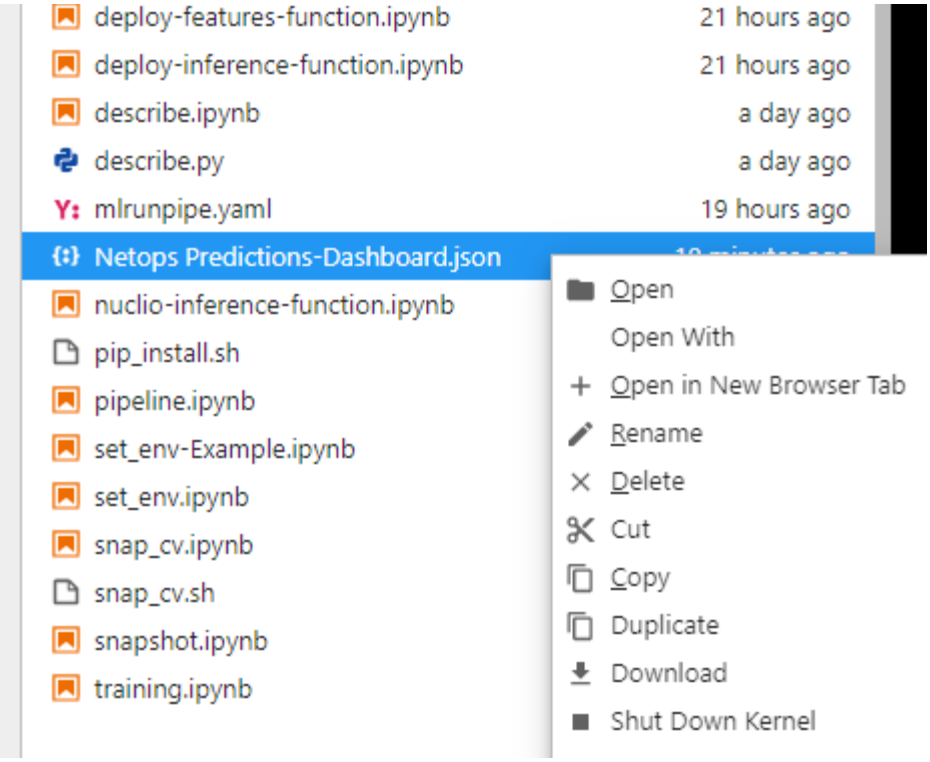
Puede implementar la consola importando el archivo JSON proporcionado en las interfaces de Grafana en el

clúster.

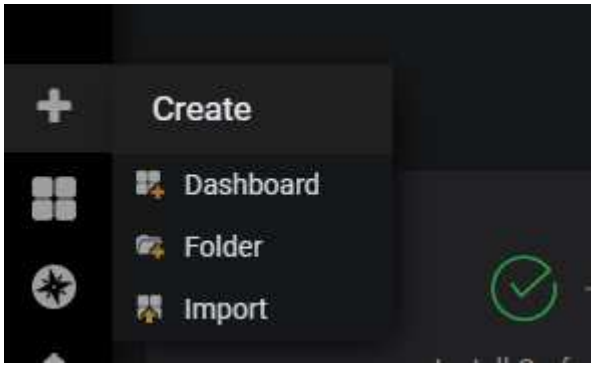
1. Para verificar que el servicio Grafana se está ejecutando, busque Servicios.



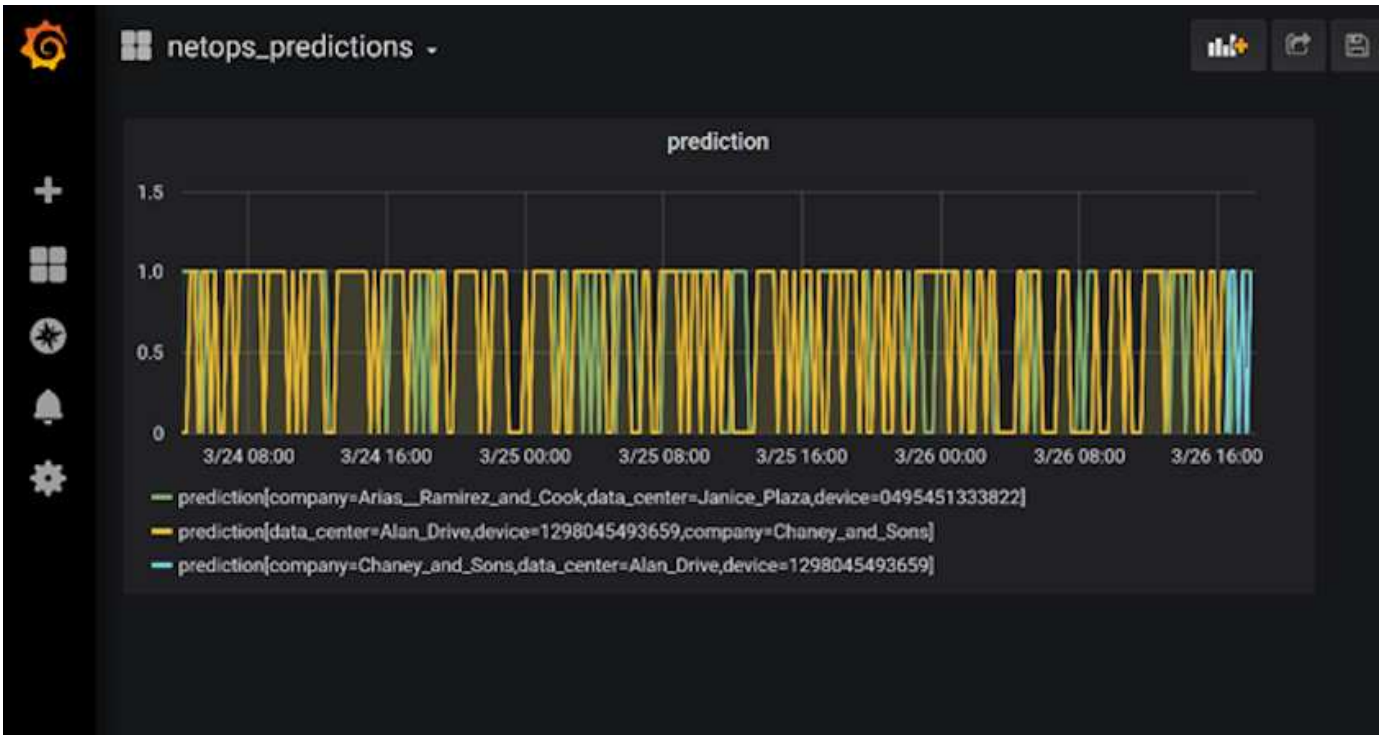
2. Si no está presente, implemente una instancia en la sección Servicios:
   a. Haga clic en Nuevo servicio.
   b. Seleccione Grafana de la lista.
   c. Acepte los valores predeterminados.
   d. Haga clic en Siguiente paso.
   e. Introduzca su ID de usuario.
   f. Haga clic en Guardar servicio.
   g. Haga clic en aplicar cambios en la parte superior.
3. Para implementar el panel de control, descargue el archivo `NetopsPredictions-Dashboard.json` A través de la interfaz Juppyter.



4. Abra Grafana en la sección Servicios e importe el panel de control.



5. Haga clic en Upload `*.json` File (Archivo) y seleccione el archivo que descargó anteriormente (`NetopsPredictions-Dashboard.json`). El panel se muestra una vez finalizada la carga.



Despliegue la función de limpieza

Cuando genera una gran cantidad de datos, es importante mantener las cosas limpias y organizadas. Para ello, implemente la función de limpieza con `cleanup.ipynb` portátil.

Beneficios

NetApp y Iguazio aceleran y simplifican la puesta en marcha de aplicaciones de IA y ML mediante la creación de marcos esenciales como Kubeflow, Apache Spark y TensorFlow, junto con herramientas de orquestación como Docker y Kubernetes. Al unificar la canalización de datos completa, NetApp y Iguazio reducen la latencia y la complejidad inherentes a muchas cargas de trabajo informáticas avanzadas, y esta brecha entre el desarrollo y las operaciones. Los científicos de datos pueden ejecutar consultas en grandes conjuntos de datos y compartir de forma segura datos y modelos algorítmicos con usuarios autorizados durante la fase de entrenamiento. Después de que los modelos en contenedores están listos para la producción, puede moverlos fácilmente desde entornos de desarrollo a entornos operativos.

Conclusión

Cuando cree sus propias canalizaciones de IA/ML, configurar la integración, la gestión, la seguridad y la accesibilidad de los componentes en una arquitectura es una tarea ardua. Dar a los desarrolladores acceso y control de su entorno presenta otro conjunto de retos.

La combinación de NetApp e Iguazio aúna estas tecnologías como servicios gestionados para acelerar la adopción de tecnologías y mejorar los plazos de comercialización de las nuevas aplicaciones DE IA/ML.

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 INTERRUPCIÓN 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.