



Implemente capacitación en IA híbrida con Union.ai y NetApp FlexCache

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Tabla de contenidos

- Implemente capacitación en IA híbrida con Union.ai y NetApp FlexCache 1
 - Descripción general 1
 - Caso de uso del cliente: Capacitación en IA en la nube híbrida 1
 - Valor del cliente 1
 - Habilitación del complemento: requisitos previos 1
 - Arquitectura de referencia 2
 - _Paso 1: Crear un volumen FlexCache 2
 - Paso 2: Configurar Trident* 2
 - Paso 3: Implementar flujos de trabajo de Union.ai* 3
- Cargar y entrenar con datos del PVC 5
 - Paso 4: Validar la integración_ 5
 - Consideraciones de seguridad 5
 - Monitoreo y Optimización 5
 - Enlaces relacionados 6
 - Conclusión 6
 - Union.ai - Guía complementaria** 6

Implemente capacitación en IA híbrida con Union.ai y NetApp FlexCache

Aprenda a implementar un entorno de entrenamiento de IA híbrido utilizando la orquestación de Union.ai con NetApp FlexCache y Trident para el aprovisionamiento de almacenamiento de Kubernetes.

David Espejo, Union.ai Sathish Thyagarajan, NetApp

Descripción general

La plataforma de orquestación híbrida de Union.ai se integra perfectamente con NetApp ONTAP y FlexCache para acelerar los flujos de trabajo de entrenamiento de IA/ML. Esta solución permite que los datos permanezcan seguros en las instalaciones mientras se aprovecha el cómputo de GPU basado en la nube para cargas de trabajo de entrenamiento de IA. NetApp FlexCache garantiza que solo los datos necesarios se almacenen en caché en la nube, lo que permite canales híbridos de IA/ML eficientes, seguros y escalables.

Caso de uso del cliente: Capacitación en IA en la nube híbrida

- Datos locales: almacenados en NetApp ONTAP para garantizar el cumplimiento y la seguridad.
- Computación en la nube: entrenamiento de GPU escalable en EKS/GKE/AKS.
- Orquestación de IA/ML: Union.ai coordina el procesamiento y el entrenamiento de datos en todos los entornos.
- Aprovisionamiento de almacenamiento: NetApp Trident automatiza el aprovisionamiento de PVC/PV.

Valor del cliente

- Ejecute cargas de trabajo de IA en conjuntos de datos masivos utilizando las capacidades de escalamiento horizontal de NetApp ONTAP.
- Mueva y sincronice datos entre las instalaciones locales y la nube utilizando las funciones de nube híbrida de NetApp.
- Almacene rápidamente en caché datos locales en la nube usando FlexCache.
- Union.ai simplifica la orquestación en todos los entornos con control de versiones, seguimiento de linaje y gestión de artefactos.
- Ejecute capacitación en la nube mientras mantiene los datos confidenciales en las instalaciones.

Habilitación del complemento: requisitos previos

Requisito	Detalles
Versión ONTAP	ONTAP 9.7+ (no se requiere licencia FlexCache)
Licencia FlexCache	Requerido en ONTAP 9.6 y anteriores
Kubernetes	Clústeres locales y en la nube (EKS/GKE/AKS)

Trident	Instalado tanto en clústeres locales como en la nube
Unión.ai	Plano de control implementado (Union Cloud o autoalojado)
Redes	Conectividad entre clústeres (si los clústeres ONTAP están separados)
Permisos	Acceso de administrador a los clústeres de ONTAP y Kubernetes. <input type="checkbox"/> Utilice las credenciales ONTAP correctas (por ejemplo, vsadmin)
¿Eres nuevo en Union.ai?	Consulte la guía complementaria al final de este documento.

Arquitectura de referencia

La siguiente figura muestra el plano de control de Union.ai integrado con el almacenamiento de NetApp para el entrenamiento de IA híbrida.

[Arquitectura de entrenamiento de IA híbrida con Union.ai y NetApp] | [ai-hybrid-union-001.png](#)

- Plano de control de Union.ai: orquesta flujos de trabajo, administra el movimiento de datos y se integra con las API de NetApp .
- NetApp ONTAP + FlexCache: proporciona almacenamiento en caché de datos eficiente desde las instalaciones locales hasta la nube.
- Clústeres de entrenamiento híbridos: los trabajos de entrenamiento se ejecutan en clústeres K8s en la nube (por ejemplo, EKS) con datos almacenados en caché desde las instalaciones locales.

_Paso 1: Crear un volumen FlexCache

Uso del Administrador del sistema ONTAP

1. Vaya a Almacenamiento > Volúmenes.
2. Haga clic en Agregar.
3. Seleccione más opciones.
4. Habilitar Agregar como caché para un volumen remoto.
5. Elija sus volúmenes de origen (local) y de destino (nube).
6. Definir QoS o nivel de rendimiento (opcional).
7. Haga clic en Crear.

☐ Si NetApp DataOps Toolkit no funciona debido a problemas de permisos o agregados, cree el volumen FlexCache directamente mediante ONTAP System Manager o CLI.

Paso 2: Configurar Trident

Instalar Trident en ambos clústeres:

☐ [" Guía de instalación de Trident "](#)

Crear el backend de Trident

```
apiVersion: trident.netapp.io/v1
kind: TridentBackendConfig
metadata:
  name: ontap-flexcache
spec:
  version: 1
  storageDriverName: ontap-nas
  managementLIF: <ONTAP-MGMT-IP>
  dataLIF: <ONTAP-DATA-IP>
  svm: <SVM-NAME>
  username: vsadmin
  password: <password>

Apply: kubectl apply -f backend-flexcache.yaml
```

Si recibe un error 401 No autorizado, verifique que el usuario de ONTAP tenga suficientes permisos de API y que utilice el nombre de usuario (vsadmin) y la contraseña correctos.

Definir StorageClass

```
apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
  name: flexcache-sc
provisioner: csi.trident.netapp.io
parameters:
  backendType: "ontap-nas"
Apply:
  kubectl apply -f storageclass-flexcache.yaml
```

Paso 3: Implementar flujos de trabajo de Union.ai

Union utiliza PVC para montar volúmenes FlexCache en trabajos de entrenamiento.

Ejemplo de plantilla de pod

```
apiVersion: v1
kind: PodTemplate
metadata:
  name: netapp-podtemplate
  namespace: flytesnacks-development
template:
  metadata:
    labels:
      default-storage: netapp
  spec:
    containers:
      - name: primary
    volumeMounts:
      - name: flexcache-storage
        mountPath: /data/flexcache
    volumes:
      - name: flexcache-storage
        persistentVolumeClaim:
          claimName: flexcache-pvc
```

Ejemplo de flujo de trabajo

Desde la tarea de importación de la unión, flujo de trabajo

```
@task(pod_template="netapp-podtemplate")

def train_model(pvc_path: str):
```

Cargar y entrenar con datos del PVC

```
@workflow

def training_pipeline():

    train_model(pvc_path="/data/flexcache")
```

El operador sindical deberá:

- Crea el PVC
- Montar el volumen FlexCache
- Programe el trabajo en el clúster K8s en la nube

Paso 4: Validar la integración_

Tarea	Validación
Montaje de PVC	Los pods de entrenamiento deben montar /data/flexcache correctamente
Acceso a datos	Los trabajos de capacitación pueden leer/escribir desde FlexCache
Comportamiento de la caché	Supervisar los aciertos y errores de caché en ONTAP. Asegúrese de que los agregados admitan FlexCache
Actuación	Validar la latencia y el rendimiento para las cargas de trabajo de entrenamiento

Utilice NetApp BlueXP o ONTAP CLI para supervisar el rendimiento.

Consideraciones de seguridad

- Utilice puntos finales de VPC para FSx para NetApp ONTAP
- Habilitar el cifrado en tránsito y en reposo
- Aplicar RBAC/IAM para el acceso a ONTAP
- Union.ai no accede ni almacena datos de los clientes

Monitoreo y Optimización

Herramienta	Objetivo
NetApp BlueXP	Supervisar el uso y el rendimiento de FlexCache
Interfaz de usuario de Union.ai	Seguimiento del estado y las métricas del pipeline
Registros de Trident	Depurar problemas de PVC o backend

Mejoras opcionales

- Automatice la creación de FlexCache mediante las API de BlueXP

- Utilice Union SDK para calentar la caché antes del entrenamiento
- Agregar inferencia por lotes o canalizaciones de servicio de modelos después del entrenamiento
- Si falla DataOps Toolkit, recurra a la creación manual de FlexCache a través del Administrador del sistema

Solución de problemas

Asunto	Resolución
PVC atascado en Pendiente	Verifique los registros de Trident y la configuración del backend
401 No autorizado desde la API de ONTAP	Utilice vsadmin y verifique los permisos
Trabajo fallido: No hay almacenamiento adecuado	Asegúrese de que el agregado de ONTAP admita FlexCache/FabricPool
Rendimiento de entrenamiento lento	Comprobar la tasa de aciertos de caché y la latencia de la red
Los datos no se sincronizan	Validar el estado de la relación de FlexCache en ONTAP

Próximos pasos

1. Validar FlexCache con datos de prueba
2. Implementar los canales de entrenamiento de Union.ai
3. Supervisar y optimizar el rendimiento
4. Documentar la configuración específica del cliente

Enlaces relacionados

- ["Documentos de Union.ai"](#)
- [" Descripción general de NetApp FlexCache "](#)
- [" Controlador Trident CSI"](#)
- ["FSx para NetApp ONTAP"](#)

Conclusión

Ahora tiene un entorno de entrenamiento de IA híbrido validado utilizando Union.ai y NetApp FlexCache. Los trabajos de capacitación se pueden ejecutar en la nube mientras se accede a datos locales de forma segura y eficiente, sin replicar conjuntos de datos completos ni comprometer la gobernanza.

Union.ai - Guía complementaria

Paso 1: Elija el modelo de implementación

Opción A: Nube de la Unión

- Visita: ["consola.union.ai"](#)
- Crear organización → Crear proyecto

Opción B: Autoalojado

- Seguir: <https://docs.union.ai/platform/latest/deployment/self-hosted/>["Guía autoalojada"]
- Implementar mediante Helm:

Repositorio de Helm: añadir Unionai <https://unionai.github.io/helm-charts/>

```
helm install union unionai/union -n union-system -f values.yaml
```

Paso 2: Instalar Union Operator

☐ `kubectl aplicar -f https://raw.githubusercontent.com/unionai/operator/main/deploy/operator.yaml`

`kubectl obtener pods -n sistema de unión`

☐

Paso 3: Instalar Union CLI

☐ `pip instalar unionai`

inicio de sesión del sindicato

☐

Paso 4: Registrar el flujo de trabajo

☐ proyecto sindical para crear inteligencia artificial híbrida

registro sindical `training_pipeline.py --proyecto hybrid-ai`

☐

Paso 5: Ejecutar y supervisar

☐ canalización de entrenamiento de ejecución sindical `--proyecto hybrid-ai`

canal de formación de vigilancia sindical

☐ Ver registros en el ["IU de unión"](#)

Paso 6: Registrar el clúster de cómputo (opcional)

☐ registro de clúster de unión `--name cloud-k8s --kubeconfig ~/.kube/config`

Paso 7: Rastrear artefactos y linaje

Union rastrea automáticamente:

- Parámetros de entrada/salida
- Versiones de datos
- Registros y métricas
- Linaje de ejecución

Información de copyright

Copyright © 2025 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.