



Distribuisci la formazione ibrida sull'intelligenza artificiale con Union.ai e NetApp FlexCache

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Sommario

Distribuisci la formazione ibrida sull'intelligenza artificiale con Union.ai e NetApp FlexCache	1
Panoramica	1
Caso d'uso del cliente: formazione sull'intelligenza artificiale nel cloud ibrido	1
Valore per il cliente	1
Abilitazione del plugin – Prerequisiti	1
Architettura di riferimento	2
_Passaggio 1: creare un volume FlexCache	2
_Passaggio 2: Configura Trident	2
_Passaggio 3: distribuire i flussi di lavoro Union.ai	3
Carica e addestra sui dati dal PVC	5
Passaggio 4: convalidare l'integrazione	5
Considerazioni sulla sicurezza	5
Monitoraggio e ottimizzazione	5
Link correlati	6
Conclusione	6
Union.ai - Guida complementare	6

Distribuisci la formazione ibrida sull'intelligenza artificiale con Union.ai e NetApp FlexCache

Scopri come distribuire un ambiente di formazione AI ibrido utilizzando l'orchestrazione Union.ai con NetApp FlexCache e Trident per il provisioning dello storage Kubernetes.

David Espejo, Union.ai Sathish Thyagarajan, NetApp

Panoramica

La piattaforma di orchestrazione ibrida di Union.ai si integra perfettamente con NetApp ONTAP e FlexCache per accelerare i flussi di lavoro di formazione AI/ML. Questa soluzione consente di mantenere i dati in modo sicuro in sede, sfruttando al contempo l'elaborazione GPU basata su cloud per i carichi di lavoro di formazione dell'intelligenza artificiale. NetApp FlexCache garantisce che solo i dati necessari vengano memorizzati nella cache del cloud, consentendo pipeline ibride AI/ML efficienti, sicure e scalabili.

Caso d'uso del cliente: formazione sull'intelligenza artificiale nel cloud ibrido

- Dati on-premise: archiviati su NetApp ONTAP per conformità e sicurezza.
- Cloud computing: addestramento GPU scalabile su EKS/GKE/AKS.
- Orchestrazione AI/ML: Union.ai coordina l'elaborazione dei dati e la formazione in tutti gli ambienti.
- Provisioning dello storage: NetApp Trident automatizza il provisioning PVC/PV.

Valore per il cliente

- Esegui carichi di lavoro di intelligenza artificiale su enormi set di dati utilizzando le funzionalità di scalabilità orizzontale di NetApp ONTAP.
- Sposta e sincronizza i dati tra locale e cloud utilizzando le funzionalità cloud ibride di NetApp.
- Memorizza rapidamente nella cache i dati locali nel cloud utilizzando FlexCache.
- Union.ai semplifica l'orchestrazione tra ambienti con il controllo delle versioni, il tracciamento della discendenza e la gestione degli artefatti.
- Esegui la formazione nel cloud mantenendo i dati sensibili in sede.

Abilitazione del plugin – Prerequisiti

Requisito	Dettagli
Versione ONTAP	ONTAP 9.7+ (licenza FlexCache non richiesta)
Licenza FlexCache	Richiesto su ONTAP 9.6 e versioni precedenti
Kubernetes	Cluster on-prem e cloud (EKS/GKE/AKS)
Trident	Installato sia su cluster on-premise che cloud
Union.ai	Piano di controllo distribuito (Union Cloud o self-hosted)

Networking	Connettività inter-cluster (se i cluster ONTAP sono separati)
Permessi	Accesso amministrativo ai cluster ONTAP e Kubernetes. <input type="checkbox"/> Utilizzare le credenziali ONTAP corrette (ad esempio, vsadmin)
Nuovo su Union.ai?	Consultare la guida complementare alla fine di questo documento

Architettura di riferimento

La figura seguente mostra il piano di controllo Union.ai integrato con lo storage NetApp per l'addestramento dell'IA ibrida.

[Architettura di formazione AI ibrida con Union.ai e NetApp] | *ai-hybrid-union-001.png*

- Piano di controllo Union.ai: orchestra i flussi di lavoro, gestisce lo spostamento dei dati e si integra con le API NetApp .
- NetApp ONTAP + FlexCache: fornisce un efficiente caching dei dati da locale al cloud.
- Cluster di formazione ibridi: i processi di formazione vengono eseguiti nei cluster cloud K8s (ad esempio, EKS) con dati memorizzati nella cache da locale.

Passaggio 1: creare un volume FlexCache

Utilizzo di ONTAP System Manager

1. Vai su Archiviazione > Volumi.
2. Fare clic su Aggiungi.
3. Seleziona Altre opzioni.
4. Abilita Aggiungi come cache per un volume remoto.
5. Scegli i volumi di origine (on-prem) e di destinazione (cloud).
6. Definire QoS o livello di prestazioni (facoltativo).
7. Fare clic su Crea.

Se NetApp DataOps Toolkit non funziona a causa di problemi di autorizzazione o di aggregazione, creare il volume FlexCache direttamente tramite ONTAP System Manager o CLI.

Passaggio 2: Configura Trident

Installare Trident su entrambi i cluster:

" [Guida all'installazione Trident](#) "

Crea Trident Backend

```

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

```

Se si riceve un errore 401 Non autorizzato, verificare che l'utente ONTAP disponga di autorizzazioni API sufficienti e che vengano utilizzati il nome utente (vsadmin) e la password corretti.

Definisci 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

```

Passaggio 3: distribuire i flussi di lavoro Union.ai

Union utilizza i PVC per montare i volumi FlexCache nei processi di formazione.

Esempio di PodTemplate

```
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
```

Esempio di flusso di lavoro

dall'attività di importazione dell'unione, flusso di lavoro

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

def train_model(pvc_path: str):
```

Carica e addestra sui dati dal PVC

```
@workflow

def training_pipeline():

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

L'operatore dell'Unione:

- Creare il PVC
- Montare il volume FlexCache
- Pianifica il lavoro nel cluster cloud K8s

Passaggio 4: convalidare l'integrazione

Compito	Convalida
Supporto in PVC	I pod di addestramento dovrebbero montare /data/flexcache correttamente
Accesso ai dati	I lavori di formazione possono leggere/scrivere da FlexCache
Comportamento della cache	Monitora i risultati/mancati risultati della cache in ONTAP. Assicurarsi che gli aggregati supportino FlexCache
Prestazione	Convalida la latenza e la produttività per i carichi di lavoro di formazione

Utilizzare NetApp BlueXP o ONTAP CLI per monitorare le prestazioni.

Considerazioni sulla sicurezza

- Utilizzare endpoint VPC per FSx per NetApp ONTAP
- Abilita la crittografia in transito e a riposo
- Applicare RBAC/IAM per l'accesso ONTAP
- Union.ai non accede né memorizza i dati dei clienti

Monitoraggio e ottimizzazione

Attrezzo	Scopo
NetApp BlueXP	Monitorare l'utilizzo e le prestazioni FlexCache
Interfaccia utente Union.ai	Monitorare lo stato e le metriche della pipeline
Tronchi Trident	Debug di problemi PVC o backend

Miglioramenti opzionali

- Automatizza la creazione FlexCache utilizzando le API BlueXP

- Utilizzare Union SDK per riscaldare la cache prima dell'addestramento
- Aggiungere pipeline di inferenza batch o di fornitura di modelli dopo l'addestramento
- Se DataOps Toolkit fallisce, ricorrere alla creazione manuale FlexCache tramite System Manager

Risoluzione dei problemi

Problema	Risoluzione
PVC bloccato in sospeso	Controlla i log Trident e la configurazione del backend
401 Non autorizzato dall'API ONTAP	Utilizzare vsadmin e verificare i permessi
Lavoro fallito: nessun archivio adatto	Assicurarsi che l'aggregato ONTAP supporti FlexCache/ FabricPool
Prestazioni di allenamento lente	Controlla il tasso di hit della cache e la latenza di rete
I dati non si sincronizzano	Convalida dello stato di salute della relazione FlexCache in ONTAP

Passi successivi

1. Convalida FlexCache con dati di prova
2. Distribuisci pipeline di formazione Union.ai
3. Monitorare e ottimizzare le prestazioni
4. Documentare la configurazione specifica del cliente

Link correlati

- ["Documentazione Union.ai"](#)
- [" Panoramica NetApp FlexCache "](#)
- [" Pilota Trident CSI"](#)
- ["FSx per NetApp ONTAP"](#)

Conclusione

Ora disponi di un ambiente di addestramento AI ibrido convalidato utilizzando Union.ai e NetApp FlexCache. I processi di formazione possono essere eseguiti nel cloud, accedendo ai dati locali in modo sicuro ed efficiente, senza replicare interi set di dati o compromettere la governance.

Union.ai - Guida complementare

Passaggio 1: Scegli il modello di distribuzione

Opzione A: Union Cloud

- Visita: ["console.union.ai"](#)
- Crea organizzazione → Crea progetto

Opzione B: Auto-ospitato

- Seguire: <https://docs.union.ai/platform/latest/deployment/self-hosted/> ["Guida auto-ospitata"]
- Distribuisci tramite Helm:

helm repo aggiungi unionai <https://unionai.github.io/helm-charts/>

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

Passaggio 2: Installa Union Operator

□ kubectl applica -f <https://raw.githubusercontent.com/unionai/operator/main/deploy/operator.yaml>

kubectl get pods -n union-system

□

Passaggio 3: Installa Union CLI

□ pip install unionai

accesso sindacale

□

Passaggio 4: Registra il flusso di lavoro

□ progetto sindacale crea intelligenza artificiale ibrida

registro sindacale training_pipeline.py --project hybrid-ai

□

Passaggio 5: Esecuzione e monitoraggio

□ union run training_pipeline --project hybrid-ai

formazione di vigilanza sindacale_pipeline

□ Visualizza i registri in ["Interfaccia utente dell'Unione"](#)

Passaggio 6: Registra Compute Cluster (facoltativo)

□ union cluster register --name cloud-k8s --kubeconfig ~/.kube/config

Passaggio 7: Traccia artefatti e lignaggio

Union tiene traccia automaticamente:

- Parametri di input/output
- Versioni dei dati
- Registri e metriche
- Lignaggio di esecuzione

Informazioni sul copyright

Copyright © 2026 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIALITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEGUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina <http://www.netapp.com/TM> sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.