



Implementieren Sie hybrides KI-Training mit Union.ai und NetApp FlexCache.

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Inhalt

- Implementieren Sie hybrides KI-Training mit Union.ai und NetApp FlexCache. 1
 - Überblick 1
 - Anwendungsfall des Kunden: Hybrid-Cloud-KI-Training 1
 - Kundennutzen 1
 - Aktivierung des Plugins – Voraussetzungen 1
 - Referenzarchitektur 2
 - Schritt 1: Erstellen eines FlexCache -Volumes 2
 - Schritt 2: Trident konfigurieren 2
 - Schritt 3: Union.ai-Workflows bereitstellen. 3
- Laden und Trainieren mit Daten aus dem PVC 5
 - Schritt 4: Integration validieren. 5
 - Sicherheitsüberlegungen 5
 - Überwachung und Optimierung 5
 - Verwandte Links. 6
 - Abschluss 6
 - Union.ai - Begleithandbuch**. 6

Implementieren Sie hybrides KI-Training mit Union.ai und NetApp FlexCache.

Erfahren Sie, wie Sie eine hybride KI-Trainingsumgebung mithilfe der Union.ai-Orchestrierung mit NetApp FlexCache und Trident für die Kubernetes-Speicherbereitstellung implementieren.

David Espejo, Union.ai Sathish Thyagarajan, NetApp

Überblick

Die hybride Orchestrierungsplattform von Union.ai integriert sich nahtlos in NetApp ONTAP und FlexCache, um KI/ML-Trainingsworkflows zu beschleunigen. Diese Lösung ermöglicht es, Daten sicher vor Ort zu speichern und gleichzeitig cloudbasierte GPU-Rechenleistung für KI-Trainingsworkloads zu nutzen. NetApp FlexCache stellt sicher, dass nur die notwendigen Daten in der Cloud zwischengespeichert werden und ermöglicht so effiziente, sichere und skalierbare hybride KI/ML-Pipelines.

Anwendungsfall des Kunden: Hybrid-Cloud-KI-Training

- Lokale Daten: Werden aus Gründen der Compliance und Sicherheit auf NetApp ONTAP gespeichert.
- Cloud-Computing: Skalierbares GPU-Training auf EKS/GKE/AKS.
- KI/ML-Orchestrierung: Union.ai koordiniert die Datenverarbeitung und das Training über verschiedene Umgebungen hinweg.
- Speicherbereitstellung: NetApp Trident automatisiert die PVC/PV-Bereitstellung.

Kundennutzen

- Führen Sie KI-Workloads auf massiven Datensätzen mithilfe der Scale-Out-Funktionen von NetApp ONTAP aus.
- Verschieben und synchronisieren Sie Daten zwischen On-Premise-Systemen und der Cloud mithilfe der Hybrid-Cloud-Funktionen von NetApp.
- Mit FlexCache können Sie lokale Daten schnell in der Cloud zwischenspeichern.
- Union.ai vereinfacht die Orchestrierung über verschiedene Umgebungen hinweg durch Versionierung, Herkunftsverfolgung und Artefaktverwaltung.
- Führen Sie Schulungen in der Cloud durch und bewahren Sie gleichzeitig sensible Daten vor Ort auf.

Aktivierung des Plugins – Voraussetzungen

Erfordernis	Details
ONTAP -Version	ONTAP 9.7+ (FlexCache -Lizenz nicht erforderlich)
FlexCache -Lizenz	Erforderlich für ONTAP 9.6 und frühere Versionen
Kubernetes	On-Premise- und Cloud-Cluster (EKS/GKE/AKS)
Trident	Installiert sowohl auf lokalen als auch auf Cloud-Clustern

Union.ai	Steuerungsebene bereitgestellt (Union Cloud oder selbst gehostet)
Vernetzung	Konnektivität zwischen Clustern (falls die ONTAP Cluster getrennt sind)
Berechtigungen	Administratorzugriff auf ONTAP und Kubernetes-Cluster. <input type="checkbox"/> Verwenden Sie die korrekten ONTAP -Anmeldeinformationen (z. B. vsadmin).
Neu bei Union.ai?	Siehe den zugehörigen Leitfaden am Ende dieses Dokuments.

Referenzarchitektur

Die folgende Abbildung zeigt die Union.ai-Steuerungsebene, die mit NetApp -Speicher für hybrides KI-Training integriert ist.

[Hybride KI-Trainingsarchitektur mit Union.ai und NetApp] | *ai-hybrid-union-001.png*

- Union.ai Control Plane: Orchestriert Workflows, verwaltet Datenbewegungen und integriert sich mit NetApp APIs.
- NetApp ONTAP + FlexCache: Bietet effizientes Daten-Caching von On-Premise bis zur Cloud.
- Hybride Trainingscluster: Trainingsaufträge werden in Cloud-K8s-Clustern (z. B. EKS) mit Daten aus dem On-Premise-Cache ausgeführt.

Schritt 1: Erstellen eines FlexCache -Volumes

Verwendung des ONTAP -Systemmanagers

1. Navigieren Sie zu Speicher > Volumes.
2. Klicken Sie auf Hinzufügen.
3. Wählen Sie „Weitere Optionen“.
4. Aktivieren Sie „Als Cache hinzufügen“ für ein Remote-Volume.
5. Wählen Sie Ihre Quellvolumes (lokal) und Zielvolumes (Cloud).
6. QoS oder Leistungsniveau definieren (optional).
7. Klicken Sie auf Erstellen.

☐ Falls das NetApp DataOps Toolkit aufgrund von Berechtigungs- oder Aggregationsproblemen nicht funktioniert, erstellen Sie das FlexCache Volume direkt mit dem ONTAP System Manager oder der CLI.

Schritt 2: Trident konfigurieren

Installieren Sie Trident auf beiden Clustern:

☐ "[Trident Installationsanleitung](#)"

Trident Backend erstellen

```
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
```

Sollten Sie die Fehlermeldung 401 Unauthorized erhalten, überprüfen Sie, ob der ONTAP Benutzer über ausreichende API-Berechtigungen verfügt und ob der richtige Benutzername (vsadmin) und das richtige Passwort verwendet werden.

Speicherklasse definieren

```
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
```

Schritt 3: Union.ai-Workflows bereitstellen

Union verwendet PVCs, um FlexCache Volumes in Trainingsjobs einzubinden.

Beispiel 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
```

Beispielhafter Arbeitsablauf

aus der Union-Importaufgabe, Workflow

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

def train_model(pvc_path: str):
```

Laden und Trainieren mit Daten aus dem PVC

```
@workflow

def training_pipeline():

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

Union Operator wird:

- PVC erstellen
- FlexCache Volume einbinden
- Planen Sie den Job im Cloud-K8s-Cluster ein.

Schritt 4: Integration validieren

Aufgabe	Validierung
PVC-Halterung	Trainings-Pods sollten /data/flexcache erfolgreich einbinden.
Datenzugriff	Trainingsaufträge können FlexCache lesen und beschreiben.
Cache-Verhalten	Cache-Treffer/Fehler in ONTAP überwachen. Stellen Sie sicher, dass die Aggregate FlexCache unterstützen.
Performance	Latenz und Durchsatz für Trainingsworkloads validieren

Verwenden Sie NetApp BlueXP oder ONTAP CLI, um die Leistung zu überwachen.

Sicherheitsüberlegungen

- VPC-Endpunkte für FSx für NetApp ONTAP verwenden
- Aktivieren Sie die Verschlüsselung während der Übertragung und im Ruhezustand.
- RBAC/IAM für ONTAP -Zugriff anwenden
- Union.ai greift nicht auf Kundendaten zu und speichert diese auch nicht.

Überwachung und Optimierung

Werkzeug	Zweck
NetApp BlueXP	FlexCache Nutzung und -Leistung überwachen
Union.ai-Benutzeroberfläche	Pipeline-Status und Kennzahlen verfolgen
Trident Logs	PVC- oder Backend-Probleme beheben

Optionale Erweiterungen

- Automatisierte FlexCache Erstellung mithilfe der BlueXP -APIs

- Verwenden Sie das Union SDK, um den Cache vor dem Training aufzuwärmen.
- Fügen Sie nach dem Training Batch-Inferenz- oder Modellbereitstellungs-Pipelines hinzu.
- Falls das DataOps Toolkit fehlschlägt, kann die manuelle FlexCache Erstellung über den System Manager durchgeführt werden.

Fehlerbehebung

Ausgabe	Auflösung
PVC steckt in Pending fest	Überprüfen Sie die Trident Protokolle und die Backend-Konfiguration.
401 Nicht autorisiert von der ONTAP -API	Verwenden Sie vsadmin und überprüfen Sie die Berechtigungen.
Auftrag fehlgeschlagen: Kein geeigneter Speicherplatz verfügbar	Stellen Sie sicher, dass ONTAP Aggregate FlexCache/ FabricPool unterstützt.
Langsame Trainingsleistung	Cache-Trefferrate und Netzwerklatenz prüfen
Daten werden nicht synchronisiert	Überprüfen Sie den Zustand der FlexCache Beziehungen in ONTAP

Nächste Schritte

1. FlexCache mit Testdaten validieren
2. Union.ai-Trainingspipelines bereitstellen
3. Leistung überwachen und optimieren
4. Kundenspezifische Einrichtung dokumentieren

Verwandte Links

- ["Union.ai-Dokumentation"](#)
- [" NetApp FlexCache Übersicht"](#)
- [" Trident CSI-Fahrer"](#)
- ["FSx für NetApp ONTAP"](#)

Abschluss

Sie verfügen nun über eine validierte hybride KI-Trainingsumgebung mit Union.ai und NetApp FlexCache. Trainingsprozesse können in der Cloud ausgeführt werden, während gleichzeitig sicher und effizient auf lokale Daten zugegriffen wird – ohne dass ganze Datensätze repliziert oder die Datenverwaltung beeinträchtigt wird.

Union.ai - Begleithandbuch

Schritt 1: Bereitstellungsmodell auswählen

Option A: Union Cloud

- Besuchen: ["console.union.ai"](#)

- Organisation erstellen → Projekt erstellen

Option B: Selbst gehostet

- Folgen: <https://docs.union.ai/platform/latest/deployment/self-hosted/>["Selbstgehosteter Leitfaden"]
- Bereitstellung über Helm:

```
helm repo add unionai https://unionai.github.io/helm-charts/
```

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

Schritt 2: Union Operator installieren

```
☐ kubectl apply -fhttps://raw.githubusercontent.com/unionai/operator/main/deploy/operator.yaml[]
```

```
kubectl get pods -n union-system
```

☐

Schritt 3: Union CLI installieren

```
☐ pip install unionai
```

Union-Login

☐

Schritt 4: Workflow registrieren

****Union-Projekt erstellt Hybrid-KI**

```
union register training_pipeline.py --project hybrid-ai
```

☐

Schritt 5: Ausführen & Überwachen

****Union run training_pipeline --project hybrid-ai**

Ausbildungsprogramm der Gewerkschaftswache

☐ Protokolle anzeigen in der "[Union UI](#)"

Schritt 6: Rechencluster registrieren (optional)

****Union Cluster Register --name cloud-k8s --kubeconfig ~/.kube/config**

Schritt 7: Artefakte und Herkunft verfolgen

Union verfolgt automatisch:

- Eingangs-/Ausgangsparameter
- Datenversionen

- Protokolle und Metriken
- Ausführungslinie

Copyright-Informationen

Copyright © 2026 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtsinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFTE SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGEND EINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter <http://www.netapp.com/TM> aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.