



Implante treinamento híbrido de IA com Union.ai e NetApp FlexCache.

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Índice

Implante treinamento híbrido de IA com Union.ai e NetApp FlexCache.....	1
Visão geral	1
Caso de uso do cliente: Treinamento de IA em nuvem híbrida	1
Valor para o cliente.....	1
Ativar o Plugin – Pré-requisitos	1
Arquitetura de referência	2
_Etapa 1: Criar um volume FlexCache	2
<i>Etapa 2: Configurar o Trident</i>	2
<i>Etapa 3: Implantar fluxos de trabalho do Union.ai</i>	3
Carregar e treinar com dados do PVC.....	5
Etapa 4: Validar a integração.....	5
Considerações de segurança	5
Monitoramento e Otimização	5
Links relacionados	6
Conclusão	6
Guia Complementar do Union.ai	6

Implante treinamento híbrido de IA com Union.ai e NetApp FlexCache.

Aprenda como implantar um ambiente híbrido de treinamento de IA usando a orquestração Union.ai com NetApp FlexCache e Trident para provisionamento de armazenamento Kubernetes.

David Espejo, Union.ai Sathish Thyagarajan, NetApp

Visão geral

A plataforma de orquestração híbrida da Union.ai integra-se perfeitamente com o NetApp ONTAP e o FlexCache para acelerar os fluxos de trabalho de treinamento de IA/ML. Essa solução permite que os dados permaneçam em segurança no local, enquanto se aproveita o poder computacional da GPU na nuvem para cargas de trabalho de treinamento de IA. O NetApp FlexCache garante que apenas os dados necessários sejam armazenados em cache na nuvem, permitindo pipelines híbridos de IA/ML eficientes, seguros e escaláveis.

Caso de uso do cliente: Treinamento de IA em nuvem híbrida

- Dados locais: armazenados no NetApp ONTAP para fins de conformidade e segurança.
- Computação em nuvem: Treinamento escalável em GPU no EKS/GKE/AKS.
- Orquestração de IA/ML: Union.ai coordena o processamento de dados e o treinamento em diversos ambientes.
- Provisionamento de armazenamento: o NetApp Trident automatiza o provisionamento de PVC/PV.

Valor para o cliente

- Execute cargas de trabalho de IA em conjuntos de dados massivos usando os recursos de escalabilidade horizontal do NetApp ONTAP.
- Transfira e sincronize dados entre ambientes locais e na nuvem usando os recursos de nuvem híbrida da NetApp.
- Armazene rapidamente dados locais em cache na nuvem usando o FlexCache.
- Union.ai simplifica a orquestração em diversos ambientes com controle de versão, rastreamento de linhagem e gerenciamento de artefatos.
- Realize treinamentos na nuvem, mantendo os dados confidenciais em servidores locais.

Ativar o Plugin – Pré-requisitos

Exigência	Detalhes
Versão ONTAP	ONTAP 9.7+ (licença FlexCache não necessária)
Licença FlexCache	Requerido no ONTAP 9.6 e versões anteriores.

Kubernetes	Clusters locais e em nuvem (EKS/GKE/AKS)
Trident	Instalado em clusters locais e na nuvem.
Union.ai	Plano de controle implantado (Union Cloud ou auto-hospedado)
Rede	Conectividade entre clusters (se os clusters ONTAP forem separados)
Permissões	Acesso administrativo aos clusters ONTAP e Kubernetes. <input type="checkbox"/> Use as credenciais ONTAP corretas (por exemplo, vsadmin)
É novo(a) na Union.ai?	Consulte o guia complementar no final deste documento.

Arquitetura de referência

A figura a seguir mostra o plano de controle do Union.ai integrado ao armazenamento NetApp para treinamento híbrido de IA.

[Arquitetura híbrida de treinamento de IA com Union.ai e NetApp] | *ai-hybrid-union-001.png*

- Union.ai Control Plane: Orquestra fluxos de trabalho, gerencia a movimentação de dados e integra-se com as APIs da NetApp .
- NetApp ONTAP + FlexCache: Oferece armazenamento em cache de dados eficiente, desde ambientes locais até a nuvem.
- Clusters de treinamento híbridos: as tarefas de treinamento são executadas em clusters K8s na nuvem (por exemplo, EKS) com dados armazenados em cache no ambiente local.

Etapa 1: Criar um volume FlexCache

Utilizando o Gerenciador de Sistemas ONTAP

1. Acesse Armazenamento > Volumes.
2. Clique em Adicionar.
3. Selecione Mais opções.
4. Ative a opção "Adicionar como cache" para um volume remoto.
5. Escolha seus volumes de origem (local) e destino (nuvem).
6. Defina o nível de QoS ou desempenho (opcional).
7. Clique em Criar.

Se o NetApp DataOps Toolkit não estiver funcionando devido a problemas de permissão ou agregação, crie o volume FlexCache diretamente usando o ONTAP System Manager ou a CLI.

Etapa 2: Configurar o Trident

Instale o Trident em ambos os clusters:

"Guia de Instalação do Trident"

Criar Backend 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
```

Se você receber um erro 401 Não autorizado, verifique se o usuário do ONTAP possui permissões de API suficientes e se o nome de usuário (vsadmin) e a senha corretos estão sendo usados.

Defina a classe de armazenamento.

```
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
```

Etapa 3: Implantar fluxos de trabalho do Union.ai

A Union usa PVCs para montar volumes FlexCache em tarefas de treinamento.

Exemplo de modelo 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
```

Exemplo de fluxo de trabalho

da tarefa de importação de união, fluxo de trabalho

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

def train_model(pvc_path: str):
```

Carregar e treinar com dados do PVC

```
@workflow

def training_pipeline():

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

O operador sindicalizado irá:

- Criar o PVC
- Monte o volume FlexCache
- Agende a tarefa no cluster Kubernetes na nuvem.

Etapa 4: Validar a integração

Tarefa	Validação
Suporte de PVC	Os pods de treinamento devem montar /data/flexcache com sucesso.
Acesso a dados	Os trabalhos de treinamento podem ler/gravar no FlexCache.
Comportamento do cache	Monitorar acertos/falhas de cache no ONTAP. Garantir que os agregados suportem FlexCache
Desempenho	Validar a latência e a taxa de transferência para cargas de trabalho de treinamento.

Utilize a CLI do NetApp BlueXP ou do ONTAP para monitorar o desempenho.

Considerações de segurança

- Use endpoints VPC para FSx para NetApp ONTAP
- Habilite a criptografia em trânsito e em repouso.
- Aplicar RBAC/IAM para acesso ao ONTAP
- A Union.ai não acessa nem armazena dados de clientes.

Monitoramento e Otimização

Ferramenta	Propósito
NetApp BlueXP	Monitore o uso e o desempenho do FlexCache.
Interface do usuário Union.ai	Acompanhe o status e as métricas do pipeline.
Registros de Trident	Depurar problemas de PVC ou backend

Aprimoramentos opcionais

- Automatize a criação do FlexCache usando as APIs do BlueXP.

- Use o Union SDK para aquecer o cache antes do treinamento.
- Adicione pipelines de inferência em lote ou de disponibilização de modelos após o treinamento.
- Se o DataOps Toolkit falhar, recorra à criação manual do FlexCache por meio do Gerenciador de Sistemas.

Solução de problemas

Emitir	Resolução
PVC preso em Pendente	Verifique os logs do Trident e a configuração do backend.
401 Não autorizado da API ONTAP	Use o vsadmin e verifique as permissões.
Tarefa falhou: Não há armazenamento adequado.	Garantir que o agregado ONTAP suporte FlexCache/ FabricPool
Desempenho lento no treinamento	Verifique a taxa de acertos de cache e a latência da rede.
Os dados não estão sendo sincronizados.	Validar a integridade do relacionamento do FlexCache no ONTAP

Próximos passos

1. Valide o FlexCache com dados de teste.
2. Implantar pipelines de treinamento da Union.ai
3. Monitorar e otimizar o desempenho
4. Documente a configuração específica do cliente.

Links relacionados

- "[Documentação do Union.ai](#)"
- "[Visão geral do NetApp FlexCache](#)"
- "[Motorista Trident CSI](#)"
- "[FSx para NetApp ONTAP](#)"

Conclusão

Agora você possui um ambiente de treinamento de IA híbrido validado usando Union.ai e NetApp FlexCache. Os trabalhos de treinamento podem ser executados na nuvem, acessando dados locais de forma segura e eficiente, sem replicar conjuntos de dados inteiros ou comprometer a governança.

Guia Complementar do Union.ai

Etapa 1: Escolha o modelo de implantação

Opção A: Union Cloud

- Visita: "[console.union.ai](#)"
- Criar organização → Criar projeto

Opção B: Hospedagem própria

- Seguir:<https://docs.union.ai/platform/latest/deployment/self-hosted/>["Guia de Hospedagem Própria"]
- Implantação via Helm:

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

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

Etapa 2: Instalar o Union Operator

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

kubectl get pods -n union-system

Etapa 3: Instale o Union CLI

pip install unionai

login da união

Etapa 4: Registrar o fluxo de trabalho

projeto de união cria IA híbrida

union register training_pipeline.py --project hybrid-ai

Etapa 5: Executar e Monitorar

union run training_pipeline --project hybrid-ai

Treinamento de vigilância sindical_pipeline

Veja os registros em "[Interface de usuário da União](#)"

Etapa 6: Registrar o cluster de computação (opcional)

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

Etapa 7: Rastrear artefatos e linhagem

O Union rastreia automaticamente:

- Parâmetros de entrada/saída
- Versões de dados
- Registros e métricas
- Linhagem de execução

Informações sobre direitos autorais

Copyright © 2025 NetApp, Inc. Todos os direitos reservados. Impresso nos EUA. Nenhuma parte deste documento protegida por direitos autorais pode ser reproduzida de qualquer forma ou por qualquer meio — gráfico, eletrônico ou mecânico, incluindo fotocópia, gravação, gravação em fita ou storage em um sistema de recuperação eletrônica — sem permissão prévia, por escrito, do proprietário dos direitos autorais.

O software derivado do material da NetApp protegido por direitos autorais está sujeito à seguinte licença e isenção de responsabilidade:

ESTE SOFTWARE É FORNECIDO PELA NETAPP "NO PRESENTE ESTADO" E SEM QUAISQUER GARANTIAS EXPRESSAS OU IMPLÍCITAS, INCLUINDO, SEM LIMITAÇÕES, GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO E ADEQUAÇÃO A UM DETERMINADO PROPÓSITO, CONFORME A ISENÇÃO DE RESPONSABILIDADE DESTE DOCUMENTO. EM HIPÓTESE ALGUMA A NETAPP SERÁ RESPONSÁVEL POR QUALQUER DANO DIRETO, INDIRETO, INCIDENTAL, ESPECIAL, EXEMPLAR OU CONSEQUENCIAL (INCLUINDO, SEM LIMITAÇÕES, AQUISIÇÃO DE PRODUTOS OU SERVIÇOS SOBRESSALENTES; PERDA DE USO, DADOS OU LUCROS; OU INTERRUPÇÃO DOS NEGÓCIOS), INDEPENDENTEMENTE DA CAUSA E DO PRINCÍPIO DE RESPONSABILIDADE, SEJA EM CONTRATO, POR RESPONSABILIDADE OBJETIVA OU PREJUÍZO (INCLUINDO NEGLIGÊNCIA OU DE OUTRO MODO), RESULTANTE DO USO DESTE SOFTWARE, MESMO SE ADVERTIDA DA RESPONSABILIDADE DE TAL DANO.

A NetApp reserva-se o direito de alterar quaisquer produtos descritos neste documento, a qualquer momento e sem aviso. A NetApp não assume nenhuma responsabilidade nem obrigação decorrentes do uso dos produtos descritos neste documento, exceto conforme expressamente acordado por escrito pela NetApp. O uso ou a compra deste produto não representam uma licença sob quaisquer direitos de patente, direitos de marca comercial ou quaisquer outros direitos de propriedade intelectual da NetApp.

O produto descrito neste manual pode estar protegido por uma ou mais patentes dos EUA, patentes estrangeiras ou pedidos pendentes.

LEGENDA DE DIREITOS LIMITADOS: o uso, a duplicação ou a divulgação pelo governo estão sujeitos a restrições conforme estabelecido no subparágrafo (b)(3) dos Direitos em Dados Técnicos - Itens Não Comerciais no DFARS 252.227-7013 (fevereiro de 2014) e no FAR 52.227- 19 (dezembro de 2007).

Os dados aqui contidos pertencem a um produto comercial e/ou serviço comercial (conforme definido no FAR 2.101) e são de propriedade da NetApp, Inc. Todos os dados técnicos e software de computador da NetApp fornecidos sob este Contrato são de natureza comercial e desenvolvidos exclusivamente com despesas privadas. O Governo dos EUA tem uma licença mundial limitada, irrevogável, não exclusiva, intransferível e não sublicenciável para usar os Dados que estão relacionados apenas com o suporte e para cumprir os contratos governamentais desse país que determinam o fornecimento de tais Dados. Salvo disposição em contrário no presente documento, não é permitido usar, divulgar, reproduzir, modificar, executar ou exibir os dados sem a aprovação prévia por escrito da NetApp, Inc. Os direitos de licença pertencentes ao governo dos Estados Unidos para o Departamento de Defesa estão limitados aos direitos identificados na cláusula 252.227-7015(b) (fevereiro de 2014) do DFARS.

Informações sobre marcas comerciais

NETAPP, o logotipo NETAPP e as marcas listadas em <http://www.netapp.com/TM> são marcas comerciais da NetApp, Inc. Outros nomes de produtos e empresas podem ser marcas comerciais de seus respectivos proprietários.