



# **Formazione distribuita in Azure - previsione dei tassi click-through**

NetApp Solutions

NetApp  
April 26, 2024

# Sommario

- Formazione distribuita in Azure - previsione dei tassi click-through. . . . . 1
  - TR-4904: Formazione distribuita in Azure - previsione dei tassi click-through . . . . . 1
  - Panoramica della tecnologia . . . . . 2
  - Requisiti software . . . . . 3
  - Requisiti relativi alle risorse cloud . . . . . 4
  - Riepilogo del caso d'uso con la previsione del tasso di click-through . . . . . 4
  - Setup (Configurazione) . . . . . 5
  - Fare clic per valutare l'elaborazione dei dati di previsione e modellare la formazione . . . . . 16
  - Conclusione . . . . . 22
  - Dove trovare ulteriori informazioni . . . . . 23

# Formazione distribuita in Azure - previsione dei tassi click-through

## TR-4904: Formazione distribuita in Azure - previsione dei tassi click-through

Rick Huang, Verron Martina, Muneer Ahmad, NetApp

Il lavoro di un data scientist dovrebbe essere incentrato sulla formazione e sulla messa a punto di modelli di apprendimento automatico (ML) e intelligenza artificiale (ai). Tuttavia, secondo una ricerca condotta da Google, i data scientist dedicano circa il 80% del loro tempo a capire come far funzionare i propri modelli con le applicazioni aziendali e a eseguirlo su larga scala.

Per gestire i progetti ai/ML end-to-end, è necessaria una maggiore comprensione dei componenti aziendali. Sebbene DevOps abbia assunto il controllo della definizione, dell'integrazione e dell'implementazione, questi tipi di componenti, le operazioni ML hanno come obiettivo un flusso simile che include i progetti ai/ML. Per avere un'idea di ciò che una pipeline ai/ML end-to-end tocca nell'azienda, consulta il seguente elenco di componenti richiesti:

- Storage
- Networking
- Database
- File system
- Container
- Pipeline ci/CD (Continuous Integration and Continuous Deployment)
- Ambiente di sviluppo integrato (IDE)
- Sicurezza
- Policy di accesso ai dati
- Hardware
- Cloud
- Virtualizzazione
- Set di strumenti e librerie per le scienze dei dati

### Pubblico di riferimento

Il mondo della scienza dei dati tocca diverse discipline nell'IT e nel business:

- Il data scientist ha bisogno della flessibilità necessaria per utilizzare i propri strumenti e le librerie preferite.
- Il data engineer deve sapere come i dati scorrono e dove risiedono.
- Un tecnico DevOps ha bisogno dei tool per integrare le nuove applicazioni ai/ML nelle pipeline ci/CD.
- Gli amministratori e gli architetti del cloud devono essere in grado di configurare e gestire le risorse di Azure.

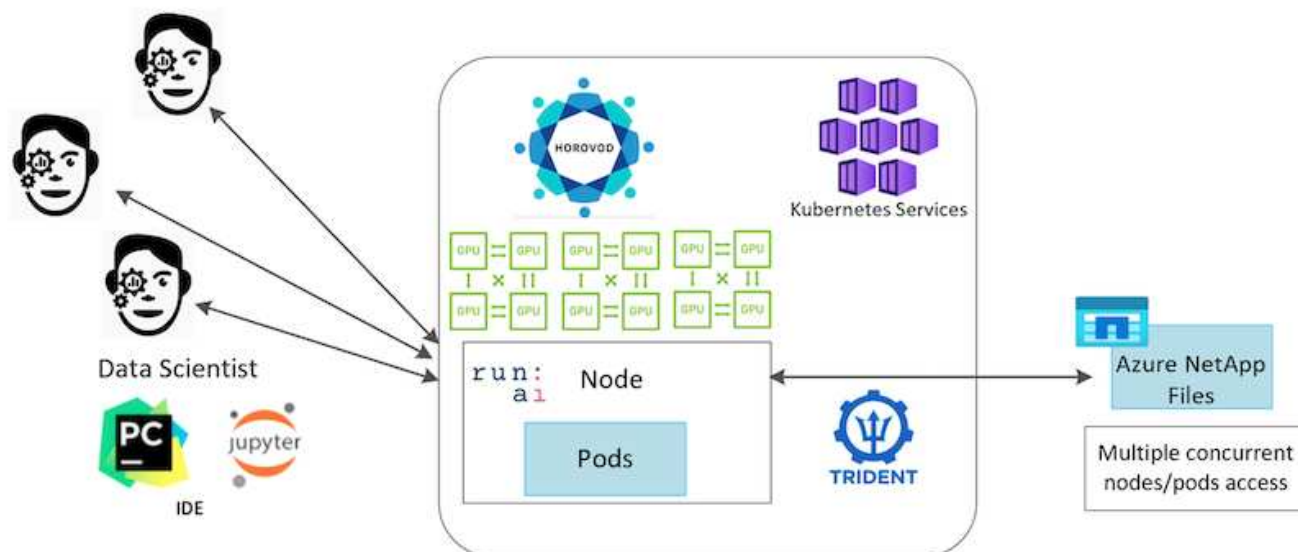
- Gli utenti aziendali desiderano avere accesso alle applicazioni ai/ML.

In questo report tecnico, descriviamo in che modo Azure NetApp Files, RAPIDS ai, DAK e Azure aiutano ciascuno di questi ruoli a portare valore al business.

## Panoramica della soluzione

Questa soluzione segue il ciclo di vita di un'applicazione ai/ML. Iniziamo con il lavoro dei data scientist per definire le diverse fasi necessarie per preparare i dati e formare i modelli. Sfruttando RAPIDS su Dask, eseguiamo training distribuiti nel cluster Azure Kubernetes Service (AKS) per ridurre drasticamente i tempi di training rispetto all'approccio convenzionale di Python scikit-Learn. Per completare il ciclo completo, integriamo la pipeline con Azure NetApp Files.

Azure NetApp Files offre diversi livelli di performance. I clienti possono iniziare con un Tier Standard, scalare e scalare fino a un Tier dalle performance elevate senza interruzioni, senza spostare alcun dato. Questa funzionalità consente agli scienziati dei dati di formare modelli su larga scala senza problemi di performance, evitando i silos di dati nel cluster, come mostrato nella figura seguente.



## Panoramica della tecnologia

Questa pagina fornisce una panoramica della tecnologia utilizzata in questa soluzione.

### Microsoft e NetApp

Da maggio 2019, Microsoft ha fornito un servizio di portale nativo e di prima parte Azure per file service NFS e SMB aziendali basati sulla tecnologia NetApp ONTAP. Questo sviluppo è guidato da una partnership strategica tra Microsoft e NetApp e estende ulteriormente la portata dei servizi dati ONTAP di livello mondiale ad Azure.

### Azure NetApp Files

Il servizio Azure NetApp Files è un servizio di storage di file di livello Enterprise, dalle performance elevate e misurato. Azure NetApp Files supporta qualsiasi tipo di carico di lavoro ed è altamente disponibile per impostazione predefinita. È possibile selezionare i livelli di servizio e di performance e impostare le copie Snapshot tramite il servizio. Azure NetApp Files è un servizio Azure first-party per la migrazione e l'esecuzione

dei carichi di lavoro dei file aziendali più esigenti nel cloud, inclusi database, SAP e applicazioni di calcolo ad alte performance senza modifiche del codice.

Questa architettura di riferimento offre alle organizzazioni IT i seguenti vantaggi:

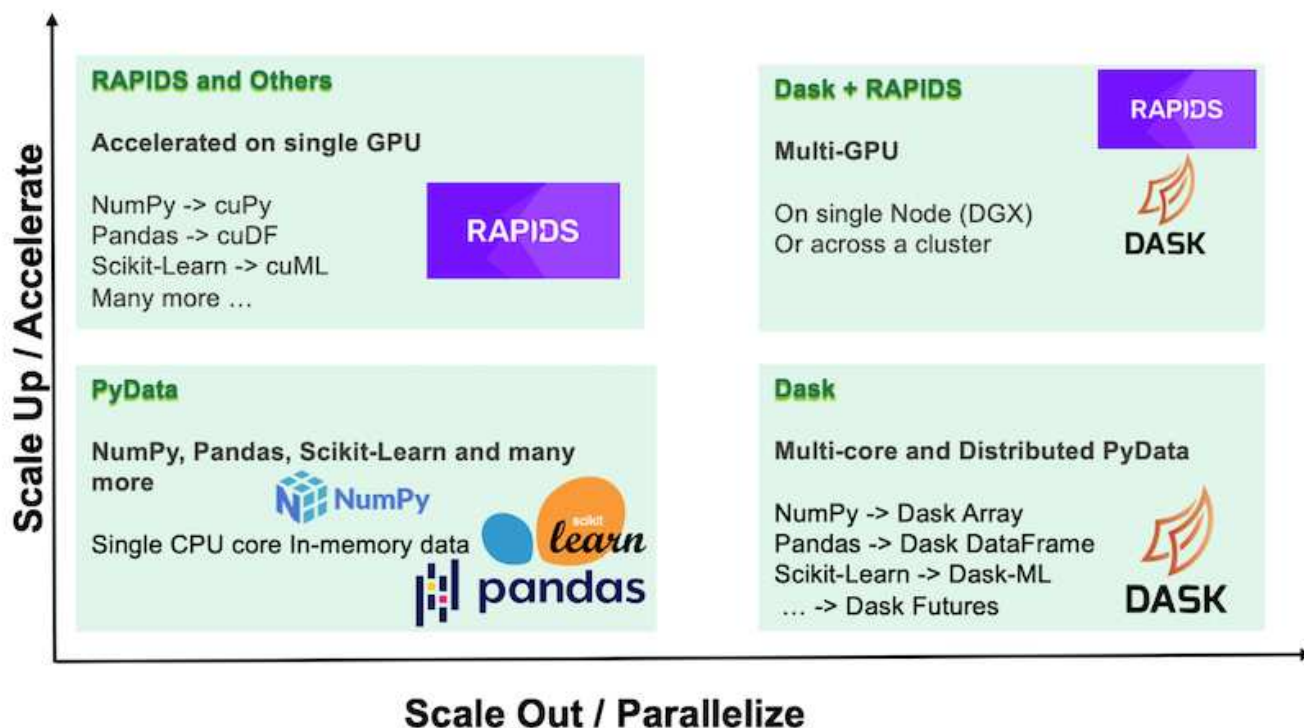
- Elimina le complessità di progettazione
- Consente una scalabilità indipendente di calcolo e storage
- Consente ai clienti di partire da piccoli e scalare perfettamente
- Offre una gamma di Tier di storage per diversi punti di costo e performance

## Panoramica di DAK e NVIDIA RAPIDS

Dask è un tool di calcolo parallelo open-source che scala le librerie Python su più macchine e fornisce un'elaborazione più rapida di grandi quantità di dati. Fornisce un'API simile alle librerie Python convenzionali a thread singolo, come Pandas, Numpy e scikit-Learn. Di conseguenza, gli utenti nativi di Python non sono costretti a modificare molto nel codice esistente per utilizzare le risorse all'interno del cluster.

NVIDIA RAPIDS è una suite di librerie open-source che consente di eseguire FLUSSI DI lavoro END-to-end DI ANALISI DEI dati E ML interamente su GPU. Insieme a Dask, consente di scalare facilmente da workstation GPU (scale-up) a cluster multi-GPU a più nodi (scale-out).

Per l'implementazione di DAK su un cluster, è possibile utilizzare Kubernetes per l'orchestrazione delle risorse. È inoltre possibile scalare in verticale o in orizzontale i nodi di lavoro in base ai requisiti di processo, che a loro volta possono aiutare a ottimizzare il consumo delle risorse del cluster, come illustrato nella figura seguente.



## Requisiti software

La seguente tabella elenca i requisiti software necessari per questa soluzione.

Software	Versione
Servizio Azure Kubernetes	1.18.14
IMMAGINE del container RAPIDS e Dask	Repository: "Rapidsai/rapidsai" Tag: 0.17-cuda11.0-runtime-ubuntu18.04
Trident di NetApp	20.01.1
Timone	3.0.0

## Requisiti relativi alle risorse cloud

Questa pagina descrive la configurazione delle risorse cloud per Azure NetApp Files.

### Configurare Azure NetApp Files

Configurare Azure NetApp Files come descritto in ["QuickStart: Configurazione di Azure NetApp Files e creazione di un volume NFS"](#).

È possibile passare alla sezione ["Crea volume NFS per Azure NetApp Files"](#) perché si stanno creando volumi tramite Trident. Prima di continuare, attenersi alla seguente procedura:

1. Registrare per Azure NetApp Files e per il provider di risorse NetApp (tramite la shell Azure) ( ["collegamento"](#) ).
2. Creare un account in Azure NetApp Files ( ["collegamento"](#) ).
3. Configurare un pool di capacità (un minimo di 4 TB Standard o Premium, a seconda delle esigenze) ( ["collegamento"](#) ). la seguente tabella elenca i requisiti di configurazione di rete per l'impostazione nel cloud. Il cluster e Azure NetApp Files devono trovarsi nella stessa rete virtuale Azure o in una rete VNET peered.

Risorse	Tipo/versione
Servizio Azure Kubernetes	1.18.14
Nodo dell'agente	3x Standard_DS2_v2
Nodo GPU	3x Standard_NC6s_v3
Azure NetApp Files	Pool di capacità standard
Capacità in TB	4

## Riepilogo del caso d'uso con la previsione del tasso di click-through

Questo caso d'utilizzo si basa sui dati pubblicamente disponibili ["Fare clic su Log di terabyte"](#) dataset da ["Criteo ai Lab"](#). Con i recenti progressi nelle piattaforme E nelle applicazioni ML, è ora molto importante concentrarsi sull'apprendimento su larga scala. Il tasso di click-through (CTR) è definito come il numero medio di click-through per cento impressioni di annunci online (espresso in percentuale). È ampiamente adottato come parametro chiave in diversi mercati verticali e casi di utilizzo del settore, tra cui digital marketing, retail, e-commerce e service provider. Di seguito sono riportati alcuni esempi

di utilizzo di CTR come metrica importante per il potenziale traffico dei clienti:

- **Digital marketing:** in "[Google Analytics](#)", CTR può essere usato per misurare come bene un advertiser o mercantile parole chiavi, annunci, ed elenchi liberi stanno eseguendo. Un CTR elevato è una buona indicazione che gli utenti trovano i tuoi annunci e gli elenchi utili e pertinenti. Il CTR contribuisce anche al CTR previsto dalla parola chiave, che è un componente di "[Classifica ad](#)".
- **E-commerce:** oltre a sfruttare "[Google Analytics](#)", ci sono almeno alcune statistiche dei visitatori in un backend di e-commerce. Anche se queste statistiche potrebbero non sembrare utili a prima vista, in genere sono facili da leggere e potrebbero essere più accurate di altre informazioni. I set di dati di prima parte composti da tali statistiche sono proprietari e sono quindi i più rilevanti per i venditori, gli acquirenti e le piattaforme di e-commerce. Questi set di dati possono essere utilizzati per impostare benchmark, confrontando i risultati con l'anno scorso e ieri, creando una serie temporale per ulteriori analisi.
- **Retail:** i retailer Brick-and-mortar possono correlare il numero di visitatori e il numero di clienti al CTR. Il numero di clienti può essere visto dalla loro storia del punto vendita. Il CTR proveniente dai siti web dei rivenditori o dal traffico pubblicitario potrebbe comportare le vendite di cui sopra. I programmi fedeltà sono un altro caso d'utilizzo, perché i clienti reindirizzati dagli annunci online o da altri siti Web potrebbero unirsi per guadagnare premi. I retailer possono acquisire i clienti attraverso programmi fedeltà e registrare i comportamenti dalle cronologie di vendita per creare un sistema di raccomandazione che non solo preveda i comportamenti di acquisto dei consumatori in diverse categorie, ma anche personalizza i coupon e riduce il tasso di abbandono.
- **Fornitori di servizi:** le aziende di telecomunicazioni e i provider di servizi Internet dispongono di numerosi dati telemetrici di prima parte per utenti che utilizzano in maniera approfondita ai, ML e analytics. Ad esempio, una telecomunicazione può sfruttare ogni giorno i log di cronologia dei domini di primo livello per la navigazione sul Web dei propri abbonati mobili per mettere a punto i modelli esistenti e produrre una segmentazione aggiornata del pubblico, prevedere il comportamento dei clienti e collaborare con gli inserzionisti per inserire annunci in tempo reale per una migliore esperienza online. In questo workflow di marketing basato sui dati, il CTR è una metrica importante per riflettere le conversioni.

Nel contesto del digital marketing, "[Registri Click di Criteo terabyte](#)" Sono ora il set di dati di riferimento per la valutazione della scalabilità delle piattaforme E degli algoritmi ML. Prevedendo il tasso di click-through, un inserzionista può selezionare i visitatori che hanno più probabilità di rispondere agli annunci, analizzare la cronologia di navigazione e mostrare gli annunci più rilevanti in base agli interessi dell'utente.

La soluzione fornita in questo report tecnico evidenzia i seguenti vantaggi:

- I vantaggi di Azure NetApp Files nella formazione distribuita o su larga scala
- RAPIDA elaborazione dei dati abilitata per CUDA (cuDF, cuPy e così via) e algoritmi ML (cuML)
- Il framework di calcolo parallelo di Dask per la formazione distribuita

Un workflow end-to-end basato su RAPIDS ai e Azure NetApp Files dimostra il drastico miglioramento dei tempi di training dei modelli di foresta casuali di due ordini di grandezza. Questo miglioramento è significativo rispetto all'approccio Pandas convenzionale quando si tratta di log click reali con 45 GB di dati tabulari strutturati (in media) ogni giorno. Ciò equivale a un DataFrame contenente circa venti miliardi di righe. Illustreremo la configurazione dell'ambiente cluster, l'installazione del framework e della libreria, il caricamento e l'elaborazione dei dati, la formazione convenzionale e distribuita, la visualizzazione e il monitoraggio e il confronto dei risultati critici di runtime end-to-end in questo report tecnico.

## Setup (Configurazione)

## Installare e configurare il cluster AKS

Per installare e configurare il cluster AKS, consultare la pagina Web ["Creare un cluster AKS"](#) quindi completare i seguenti passaggi:

1. Quando si seleziona il tipo di nodo (nodi di sistema [CPU] o di lavoro [GPU]), selezionare quanto segue:
  - a. I nodi di sistema primari devono essere DS2v2 standard (agentpool tre nodi predefiniti).
  - b. Quindi, aggiungere il pool Standard\_NC6s\_v3 del nodo di lavoro (minimo tre nodi) per il gruppo di utenti (per i nodi GPU) denominato gpupool.

+ Add node pool    Delete					
Name	Mode	OS type	Node count	Node size	
<input type="checkbox"/> agentpool	System	Linux	3	Standard_DS2_v2	
<input type="checkbox"/> gpupool	User	Linux	3	Standard_NC6s_v	

2. L'implementazione richiede da 5 a 10 minuti. Al termine, fare clic su Connect to Cluster (Connetti al cluster).
3. Per connettersi al cluster AKS appena creato, installare quanto segue dall'ambiente locale (laptop/pc):
  - a. Lo strumento della riga di comando Kubernetes che utilizza ["Istruzioni fornite per il sistema operativo in uso"](#)
  - b. La CLI di Azure come descritto nel documento, ["Installare Azure CLI"](#)
4. Per accedere al cluster AKS dal terminale, immettere `az login` e immettere le credenziali.
5. Eseguire i due comandi seguenti:

```
az account set --subscription xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxxxx
aks get-credentials --resource-group resourcegroup --name aksclustername
```

6. Invio Azure CLI: `kubectl get nodes`.
7. Se tutti e sei i nodi sono attivi e in esecuzione, come illustrato nell'esempio seguente, il cluster AKS è pronto e connesso all'ambiente locale

```
verronmartina@verron-mac-0 ~ % kubectl get nodes
NAME                                STATUS    ROLES    AGE   VERSION
aks-agentpool-34613062-vmss000000  Ready    agent    22m   v1.18.14
aks-agentpool-34613062-vmss000001  Ready    agent    22m   v1.18.14
aks-agentpool-34613062-vmss000002  Ready    agent    22m   v1.18.14
aks-gpupool-34613062-vmss000000     Ready    agent    20m   v1.18.14
aks-gpupool-34613062-vmss000001     Ready    agent    20m   v1.18.14
aks-gpupool-34613062-vmss000002     Ready    agent    20m   v1.18.14
verronmartina@verron-mac-0 ~ %
```

## Creare una subnet delegata per Azure NetApp Files

Per creare una subnet delegata per Azure NetApp Files, attenersi alla seguente



procedura:

1. Accedere alle reti virtuali all'interno del portale Azure. Trova la tua rete virtuale appena creata. Dovrebbe avere un prefisso come aks-vnet.
2. Fare clic sul nome di VNET.

Microsoft Azure

Search resources, services, and docs (G+)

Dashboard >

## Virtual networks

seanlucelive (Default Directory)

+ Add Manage view Refresh Export to CSV Open query Assign tags Feedback

Filter by name... Subscription == AzureSub01 Resource group == all Location == all Add filter

Showing 1 to 5 of 5 records. No grouping List view

Name	Resource group	Location	Subscription
aks-vnet-22885919	MC_sluce_rg_TridentDemo_eastus2	East US 2	AzureSub01

3. Fare clic su subnet e fare clic su +Subnet nella barra degli strumenti superiore.

Microsoft Azure

Search resources, services, and docs (G+)

Dashboard > Virtual networks > aks-vnet-22885919

## aks-vnet-22885919 | Subnets

Virtual network

Search (Ctrl+/) + Subnet Gateway subnet Refresh Manage users Delete

Search subnets

Name	IPv4	IPv6 (many availab...	Delegated to	Security group
aks-subnet	10.240.0.0/16 (65530 av...	-	-	aks-agentpool-2288591...

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Settings Address space Connected devices Subnets

4. Specificare un nome per la subnet, ad esempio ANF.sn Quindi, sotto delega subnet, selezionare Microsoft.Netapp/volumes. Non cambiare altro. Fare clic su OK.

## Add subnet



Name \*

ANF.sn



Subnet address range \* ⓘ

10.0.0.0/24

10.0.0.0 - 10.0.0.255 (251 + 5 Azure reserved addresses)

☐

Add IPv6 address space ⓘ

NAT gateway ⓘ

None



Network security group

None



Route table

None



### SERVICE ENDPOINTS

Create service endpoint policies to allow traffic to specific azure resources from your virtual network over service endpoints. [Learn more](#)

Services ⓘ

0 selected



### SUBNET DELEGATION

Delegate subnet to a service ⓘ

Microsoft.Netapp/volumes



OK

Cancel

I volumi Azure NetApp Files vengono allocati nel cluster di applicazioni e vengono utilizzati come dichiarazioni di volumi persistenti (PVC) in Kubernetes. A sua volta, questo processo offre la flessibilità di mapparli a diversi servizi, come i notebook Jupyter, le funzioni senza server e così via.

Gli utenti dei servizi possono consumare lo storage dalla piattaforma in molti modi. Mentre questo report tecnico illustra gli NFS, i principali vantaggi di Azure NetApp Files sono:

- Fornire agli utenti la possibilità di utilizzare le copie Snapshot.
- Consente agli utenti di memorizzare grandi quantità di dati su volumi Azure NetApp Files.
- Utilizzo dei vantaggi in termini di performance dei volumi Azure NetApp Files quando si eseguono i modelli su grandi set di file.

## Peer AKS VNET e Azure NetApp Files VNET

Per eseguire il peer di AKS VNET a Azure NetApp Files VNET, attenersi alla seguente procedura:

1. Immettere Virtual Networks nel campo di ricerca.
2. Selezionare `vnet aks-vnet-name`. Fare clic su di esso e immettere Peerings nel campo di ricerca.
3. Fare clic su +Add (Aggiungi).
4. Immettere i seguenti descrittori:
  - a. Il nome del collegamento di peering è `aks-vnet-name_to_anf`.
  - b. SubscriptionID e Azure NetApp Files VNET come partner di peering VNET.
  - c. Lasciare tutte le sezioni senza asterisco con i valori predefiniti.
5. Fare clic su Aggiungi.

Per ulteriori informazioni, vedere ["Creare, modificare o eliminare un peering di rete virtuale"](#).

## Installare Trident

Per installare Trident utilizzando Helm, attenersi alla seguente procedura:

1. Installare Helm (per istruzioni sull'installazione, visitare il ["origine"](#)).
2. Scaricare ed estrarre il programma di installazione di Trident 20.01.1.

```
$wget  
$tar -xf trident-installer-21.01.1.tar.gz
```

3. Modificare la directory in `trident-installer`.

```
$cd trident-installer
```

4. Copia `tridentctl` a una directory del sistema `$PATH`.

```
$sudo cp ./tridentctl /usr/local/bin
```

5. Installare Trident sul cluster Kubernetes (K8s) con Helm ( ["origine"](#)):
  - a. Modificare la directory in `helm` directory.

```
$cd helm
```

- b. Installare Trident.

```
$helm install trident trident-operator-21.01.1.tgz --namespace  
trident --create-namespace
```

c. Controllare lo stato dei pod Trident.

```
$kubectl -n trident get pods
```

Se tutti i pod sono in funzione, Trident viene installato ed è possibile procedere.

6. Impostare il backend Azure NetApp Files e la classe storage per AKS.

a. Creare un principio di servizio Azure.

Il service principal è il modo in cui Trident comunica con Azure per gestire le risorse Azure NetApp Files.

```
$az ad sp create-for-rbac --name ""
```

L'output dovrebbe essere simile al seguente esempio:

```
{  
  "appId": "xxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",  
  "displayName": "netapptrident",  
  "name": "",  
  "password": "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx",  
  "tenant": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"  
}
```

7. Creare un file json backend Trident, nome di esempio `anf-backend.json`.

8. Utilizzando l'editor di testo preferito, completare i seguenti campi all'interno di `anf-backend.json` file:

```
{
  "version": 1,
  "storageDriverName": "azure-netapp-files",
  "subscriptionID": "fakec765-4774-fake-ae98-a721add4fake",
  "tenantID": "fakef836-edc1-fake-bff9-b2d865eefake",
  "clientID": "fake0f63-bf8e-fake-8076-8de91e57fake",
  "clientSecret": "SECRET",
  "location": "westeurope",
  "serviceLevel": "Standard",
  "virtualNetwork": "anf-vnet",
  "subnet": "default",
  "nfsMountOptions": "vers=3,proto=tcp",
  "limitVolumeSize": "500Gi",
  "defaults": {
    "exportRule": "0.0.0.0/0",
    "size": "200Gi"
  }
}
```

9. Sostituire i seguenti campi:

- `subscriptionID`. Il tuo ID di abbonamento Azure.
- `tenantID`. Il tuo ID tenant Azure dall'output di `az ad sp` nella fase precedente.
- `clientID`. Il tuo `appId` dall'output di `az ad sp` nella fase precedente.
- `clientSecret`. La password dall'output di `az ad sp` nella fase precedente.

10. Chiedere a Trident di creare il backend Azure NetApp Files in `trident` namespace con `anf-backend.json` come file di configurazione:

```
$tridentctl create backend -f anf-backend.json -n trident
```

NAME	STORAGE DRIVER	UUID	STATE	VOLUMES
azurenetaappfiles_86181	azure-netapp-files	2ca85462-59ac-4946-be05-c03f5575a2ad	online	0

11. Creare una classe di storage. Kubernetes consente agli utenti di eseguire il provisioning dei volumi utilizzando PVC che specificano una classe di storage in base al nome. Chiedere a K8s di creare una classe di storage `azurenetaappfiles` Che fa riferimento al backend Trident creato nel passaggio precedente.
12. Creare un YAML (`anf-storage-class.yaml`) per la copia e la classe di storage.

```

apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
  name: azurenetappfiles
provisioner: netapp.io/trident
parameters:
  backendType: "azure-netapp-files"
$kubectl create -f anf-storage-class.yaml

```

13. Verificare che la classe di storage sia stata creata.

```
kubectl get sc azurenetappfiles
```

NAME	PROVISIONER	RECLAIMPOLICY	VOLUMEBINDINGMODE	ALLOWVOLUMEEXPANSION	AGE
azurenetappfiles	csi.trident.netapp.io	Delete	Immediate	false	98s

## Impostare Dask con L'implementazione DI RAPIDS su AKS utilizzando Helm

Per configurare la distribuzione di Dask con RAPIDS su AKS utilizzando Helm, attenersi alla seguente procedura:

1. Creare uno spazio dei nomi per l'installazione di Dask con RAPIDS.

```
kubectl create namespace rapids-dask
```

2. Creare un PVC per memorizzare il set di dati del tasso di click-through:

a. Salvare il seguente contenuto YAML in un file per creare un PVC.

```

kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: pvc-criteo-data
spec:
  accessModes:
    - ReadWriteMany
  resources:
    requests:
      storage: 1000Gi
  storageClassName: azurenetappfiles

```

b. Applicare il file YAML al cluster Kubernetes.

```
kubectl -n rapids-dask apply -f <your yaml file>
```

3. Clonare il rapidsai git repository ( "<https://github.com/rapidsai/helm-chart>").

```
git clone https://github.com/rapidsai/helm-chart helm-chart
```

4. Modificare `values.yaml` E includere il PVC creato in precedenza per i lavoratori e l'area di lavoro Jupyter.

- a. Accedere alla rapidsai directory del repository.

```
cd helm-chart/rapidsai
```

- b. Aggiornare `values.yaml` Archiviare e montare il volume utilizzando PVC.

```
dask:
  ...
worker:
  name: worker
  ...
mounts:
  volumes:
    - name: data
      persistentVolumeClaim:
        claimName: pvc-criteo-data
  volumeMounts:
    - name: data
      mountPath: /data
  ...
jupyter:
  name: jupyter
  ...
mounts:
  volumes:
    - name: data
      persistentVolumeClaim:
        claimName: pvc-criteo-data
  volumeMounts:
    - name: data
      mountPath: /data
  ...
```

5. Accedere alla home directory del repository e implementare Dask con tre nodi di lavoro su AKS utilizzando

Helm.

```
cd ..  
helm dep update rapidsai  
helm install rapids-dask --namespace rapids-dask rapidsai
```

## Livelli di performance Azure NetApp Files

È possibile modificare il livello di servizio di un volume esistente spostando il volume in un altro pool di capacità che utilizza il livello di servizio desiderato per il volume. Questa soluzione consente ai clienti di iniziare con un piccolo set di dati e un piccolo numero di GPU nel Tier standard e scalare in orizzontale o in verticale fino al Tier Premium con l'aumentare della quantità di dati e GPU. Il livello Premium offre un throughput per terabyte quattro volte superiore rispetto al livello Standard e la scalabilità verticale viene eseguita senza dover spostare alcun dato per modificare il livello di servizio di un volume.

### Modificare dinamicamente il livello di servizio di un volume

Per modificare dinamicamente il livello di servizio di un volume, attenersi alla seguente procedura:

1. Nella pagina Volumes (volumi), fare clic con il pulsante destro del mouse sul volume di cui si desidera modificare il livello di servizio. Selezionare Cambia pool.

NFSv3	10.28.254.4:/norootfor	Standard	pool0	...
NFSv4.1	NAS-735a.docs.lab:/for	Premium		...
NFSv4.1	NAS-735a.docs.lab:/krt	Premium		...
NFSv3	10.28.254.4:/moveme0	Premium		...
NFSv3	10.28.254.4:/placeholder	Premium		...

Resize

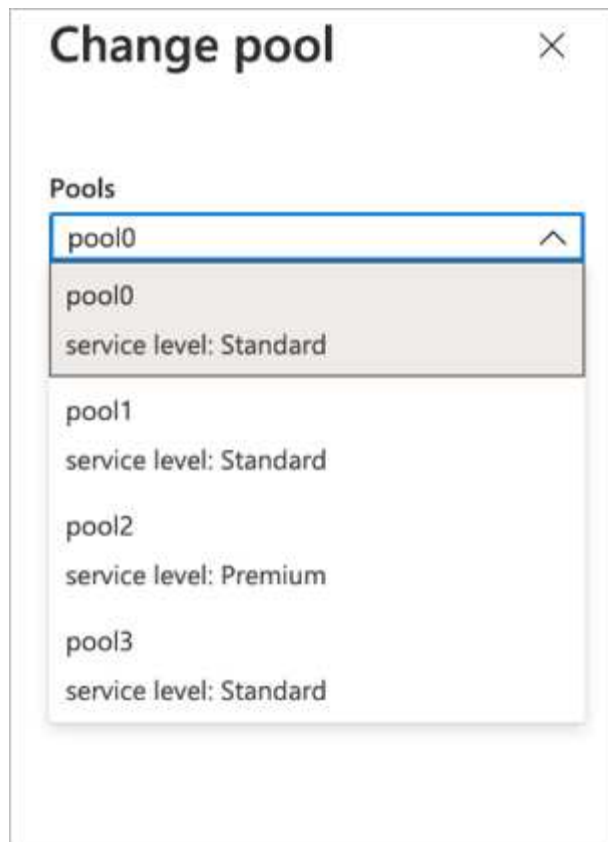
Edit

Change pool

Delete

2. Nella finestra Change Pool, selezionare il pool di capacità in cui si desidera spostare il volume.





3. Fare clic su OK.

### Automatizza la modifica del Tier delle performance

Sono disponibili le seguenti opzioni per automatizzare le modifiche del Tier di performance:

- La modifica dinamica del livello di servizio è ancora in Public Preview (Anteprima pubblica) e non è attivata per impostazione predefinita. Per attivare questa funzione nell'abbonamento Azure, consultare la presente documentazione ["Modificare dinamicamente il livello di servizio di un volume"](#).
- I comandi di modifica del pool di volumi Azure CLI sono forniti in ["documentazione per la modifica del pool di volumi"](#) e nel seguente esempio:

```
az netappfiles volume pool-change -g mygroup --account-name myaccname  
--pool-name mypoolname --name myvolname --new-pool-resource-id  
mynewresourceid
```

- PowerShell ["Set-AzNetAppFilesVolumePool cmdlet"](#) Modifica il pool di un volume Azure NetApp Files e viene mostrato nell'esempio seguente:

```
Set-AzNetAppFilesVolumePool
-ResourceGroupName "MyRG"
-AccountName "MyAnfAccount"
-PoolName "MyAnfPool"
-Name "MyAnfVolume"
-NewPoolResourceId 7d6e4069-6c78-6c61-7bf6-c60968e45fbf
```

## Fare clic per valutare l'elaborazione dei dati di previsione e modellare la formazione

### Librerie per l'elaborazione dei dati e la formazione sui modelli

La tabella seguente elenca le librerie e i framework utilizzati per creare questa attività. Tutti questi componenti sono stati completamente integrati con i controlli di sicurezza e accesso basati sui ruoli di Azure.

Librerie/framework	Descrizione
CuML di Dask	Per CONSENTIRE A ML di lavorare su GPU, il <a href="#">"Libreria cuML"</a> Fornisce l'accesso al pacchetto RAPIDS cuML con DAK. RAPIDS cuML implementa i più diffusi algoritmi ML, tra cui clustering, riduzione delle dimensioni e approcci di regressione, con implementazioni basate su GPU ad alte performance, che offrono velocità fino a 100 volte superiori rispetto agli approcci basati su CPU.
Dask cuDF	CuDF include diverse altre funzioni che supportano l'estrazione, la trasformazione, il carico (ETL) con accelerazione GPU, come il sottosetting dei dati, le trasformazioni, la codifica one-hot e molto altro ancora. Il team RAPIDS gestisce un <a href="#">"libreria dask-cudf"</a> Sono inclusi i metodi di supporto per l'utilizzo di Dask e cuDF.
Scikit Impara	Scikit-Learn offre decine di algoritmi e modelli di apprendimento automatico integrati, chiamati stimatori. Ciascuno <a href="#">"stimatore"</a> può essere adattato ad alcuni dati utilizzando its <a href="#">"adatta"</a> metodo.

Abbiamo utilizzato due notebook per costruire LE pipeline ML per il confronto; uno è l'approccio convenzionale Pandas scikit-Learn e l'altro è la formazione distribuita con RAPIDS e Dask. Ciascun notebook può essere testato singolarmente per verificarne le prestazioni in termini di tempo e scalabilità. Copriamo ogni notebook singolarmente per dimostrare i vantaggi della formazione distribuita utilizzando RAPIDS e Dask.

### Load Criteo fare clic su Logs giorno 15 in Pandas e formare un modello di foresta casuale scikit-learn

In questa sezione viene descritto come abbiamo utilizzato Pandas e Dask DataFrame per

caricare i dati dei registri Click dall'insieme di dati Criteo Terabyte. Il caso d'utilizzo è importante nella pubblicità digitale per gli scambi di annunci per creare i profili degli utenti prevedendo se gli annunci verranno cliccati o se lo scambio non utilizza un modello accurato in una pipeline automatica.

Abbiamo caricato i dati del giorno 15 dal set di dati Click Logs, per un totale di 45 GB. Eseguire la seguente cella nel notebook Jupyter CTR-PandasRF-collated.ipynb Crea un Pandas DataFrame che contiene i primi 50 milioni di righe e genera un modello di foresta casuale scikit-Learn.

```
%%time
import pandas as pd
import numpy as np
header = ['col'+str(i) for i in range (1,41)] #note that according to
criteo, the first column in the dataset is Click Through (CT). Consist of
40 columns
first_row_taken = 50_000_000 # use this in pd.read_csv() if your compute
resource is limited.
# total number of rows in day15 is 20B
# take 50M rows
"""
Read data & display the following metrics:
1. Total number of rows per day
2. df loading time in the cluster
3. Train a random forest model
"""
df = pd.read_csv(file, nrows=first_row_taken, delimiter='\t',
names=header)
# take numerical columns
df_sliced = df.iloc[:, 0:14]
# split data into training and Y
Y = df_sliced.pop('col1') # first column is binary (click or not)
# change df_sliced data types & fillna
df_sliced = df_sliced.astype(np.float32).fillna(0)
from sklearn.ensemble import RandomForestClassifier
# Random Forest building parameters
# n_streams = 8 # optimization
max_depth = 10
n_bins = 16
n_trees = 10
rf_model = RandomForestClassifier(max_depth=max_depth,
n_estimators=n_trees)
rf_model.fit(df_sliced, Y)
```

Per eseguire la previsione utilizzando un modello di foresta casuale con formazione, eseguire il paragrafo seguente in questo notebook. Abbiamo preso gli ultimi un milione di righe dal giorno 15 come set di test per evitare qualsiasi duplicazione. La cella calcola anche la precisione della previsione, definita come la

percentuale di occorrenze che il modello prevede accuratamente se un utente fa clic su un annuncio o meno. Per esaminare eventuali componenti non familiari presenti in questo notebook, consultare la sezione ["documentazione ufficiale scikit-learn"](#).

```
# testing data, last 1M rows in day15
test_file = '/data/day_15_test'
with open(test_file) as g:
    print(g.readline())

# dataframe processing for test data
test_df = pd.read_csv(test_file, delimiter='\t', names=header)
test_df_sliced = test_df.iloc[:, 0:14]
test_Y = test_df_sliced.pop('coll')
test_df_sliced = test_df_sliced.astype(np.float32).fillna(0)
# prediction & calculating error
pred_df = rf_model.predict(test_df_sliced)
from sklearn import metrics
# Model Accuracy
print("Accuracy:", metrics.accuracy_score(test_Y, pred_df))
```

## Caricare il giorno 15 in Dask e formare un modello di foresta casuale di Dask cuML

In modo simile alla sezione precedente, caricare Criteo Click Logs Day 15 in Pandas e formare un modello di foresta casuale scikit-learn. In questo esempio, è stato eseguito il caricamento di DataFrame con Dask cuDF e il training di un modello di foresta casuale in Dask cuML. Nella sezione abbiamo confrontato le differenze di tempo e di scala per la formazione ["Confronto tra i tempi di formazione"](#).

### criteo\_dask\_RF.ipynb

Questo notebook importa numpy, cuml e il necessario `dask` librerie, come mostrato nell'esempio seguente:

```
import cuml
from dask.distributed import Client, progress, wait
import dask_cudf
import numpy as np
import cudf
from cuml.dask.ensemble import RandomForestClassifier as cumlDaskRF
from cuml.dask.common import utils as dask_utils
```

Avviare Dask Client().

```
client = Client()
```

Se il cluster è configurato correttamente, è possibile visualizzare lo stato dei nodi di lavoro.

```
client
workers = client.has_what().keys()
n_workers = len(workers)
n_streams = 8 # Performance optimization
```

Nel nostro cluster AKS viene visualizzato il seguente stato:

Client	Cluster
<b>Scheduler:</b> tcp://rapidsai-scheduler:8786	<b>Workers:</b> 3
<b>Dashboard:</b> <a href="http://proxy/rapidsai-scheduler:8787/status">/proxy/rapidsai-scheduler:8787/status</a>	<b>Cores:</b> 3
	<b>Memory:</b> 354.55 GB

Si noti che Dask utilizza il paradigma di esecuzione pigro: Invece di eseguire il codice di elaborazione istantaneamente, Dask crea invece un DAG (Directed Acyclic Graph) di esecuzione. IL DAG contiene una serie di attività e le relative interazioni che ciascun lavoratore deve eseguire. Questo layout significa che i task non vengono eseguiti finché l'utente non dice a Task di eseguirli in un modo o nell'altro. Con Dask hai tre opzioni principali:

- **Call compute() su un DataFrame.** questa chiamata elabora tutte le partizioni e restituisce i risultati allo scheduler per l'aggregazione finale e la conversione in cuDF DataFrame. Questa opzione deve essere utilizzata con parsimonia e solo in caso di risultati fortemente ridotti, a meno che il nodo dello scheduler non esaurisca la memoria.
- **Call Persistent() su un DataFrame.** questa chiamata esegue il grafico, ma, invece di restituire i risultati al nodo scheduler, li mantiene in memoria nel cluster in modo che l'utente possa riutilizzare questi risultati intermedi lungo la pipeline senza dover eseguire nuovamente la stessa elaborazione.
- **Call head() su un DataFrame.** proprio come con cuDF, questa chiamata restituisce 10 record al nodo Scheduler. Questa opzione consente di verificare rapidamente se il DataFrame contiene il formato di output desiderato o se i record stessi hanno senso, a seconda dell'elaborazione e del calcolo.

Pertanto, a meno che l'utente non chiami una di queste azioni, i lavoratori sono inattivi in attesa che lo scheduler avvii l'elaborazione. Questo paradigma di esecuzione pigro è comune nei moderni framework di calcolo distribuiti e paralleli come Apache Spark.

Il paragrafo seguente forma un modello di foresta casuale utilizzando Dask cuML per il calcolo distribuito con accelerazione GPU e calcola la precisione di previsione del modello.

```

Adsf
# Random Forest building parameters
n_streams = 8 # optimization
max_depth = 10
n_bins = 16
n_trees = 10
cuml_model = cumlDaskRF(max_depth=max_depth, n_estimators=n_trees,
n_bins=n_bins, n_streams=n_streams, verbose=True, client=client)
cuml_model.fit(gdf_sliced_small, Y)
# Model prediction
pred_df = cuml_model.predict(gdf_test)
# calculate accuracy
cu_score = cuml.metrics.accuracy_score( test_y, pred_df )

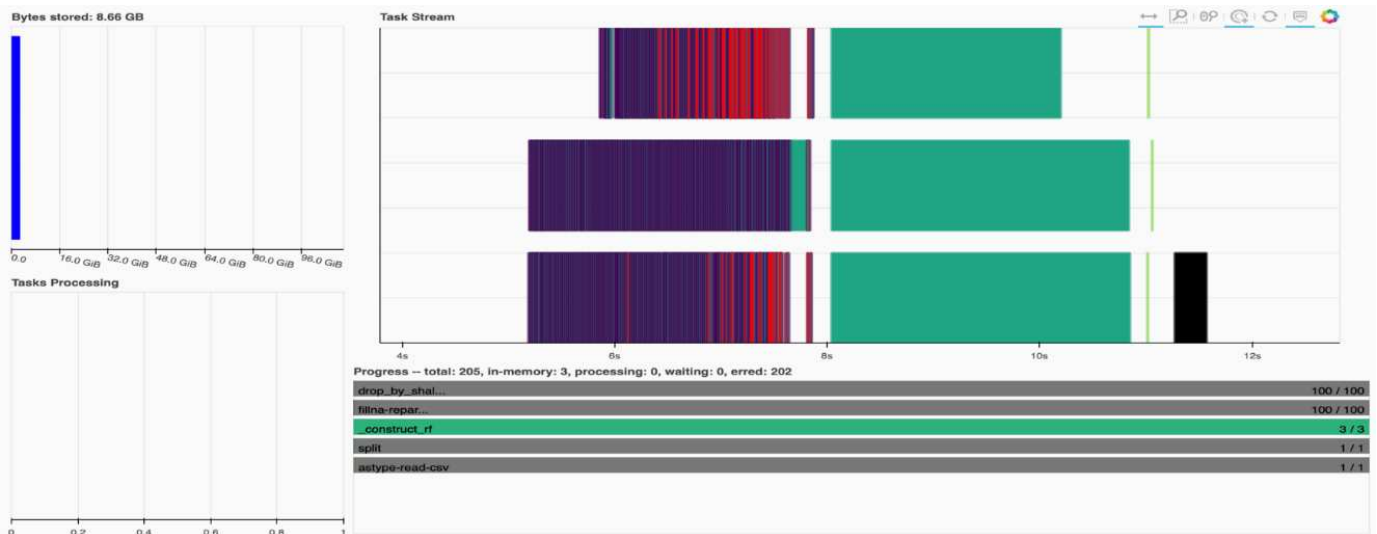
```

## Monitorate la Task utilizzando la dashboard nativa dei Task Streams

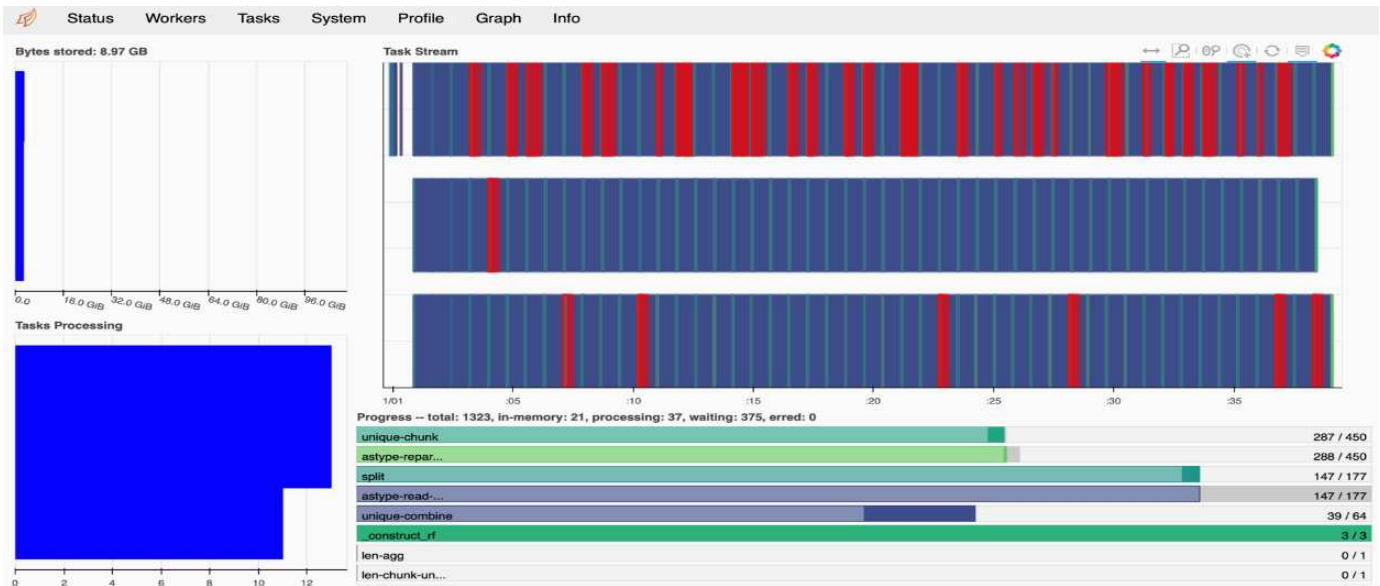
Il "[Scheduler distribuito di Dask](#)" fornisce feedback live in due forme:

- Una dashboard interattiva contenente numerosi grafici e tabelle con informazioni in tempo reale
- Barra di avanzamento adatta per l'utilizzo interattivo in console o notebook

Nel nostro caso, la figura seguente mostra come è possibile monitorare l'avanzamento del task, inclusi i byte memorizzati, il Task Stream con una dettagliata suddivisione del numero di flussi e l'avanzamento in base ai nomi delle attività con le funzioni associate eseguite. Nel nostro caso, poiché abbiamo tre nodi di lavoro, ci sono tre blocchi principali di flusso e i codici colore indicano attività diverse all'interno di ogni flusso.



È possibile analizzare le singole attività ed esaminare il tempo di esecuzione in millisecondi o identificare eventuali ostacoli o ostacoli. Ad esempio, la figura seguente mostra i flussi di attività per la fase di adattamento del modello di foresta casuale. Le funzioni eseguite sono notevolmente più numerose, tra cui il chunk unico per l'elaborazione di DataFrame, `_Construct_rf` per l'adattamento della foresta casuale e così via. La maggior parte del tempo è stato dedicato alle operazioni DataFrame a causa delle grandi dimensioni (45 GB) di un giorno di dati provenienti dai Click Logs di Criteo.



### Confronto dei tempi di training

In questa sezione viene confrontato il tempo di training del modello utilizzando i Panda convenzionali rispetto a quello di Dask. Per Pandas, abbiamo caricato una quantità inferiore di dati a causa della natura del tempo di elaborazione più lento per evitare l'overflow della memoria. Pertanto, abbiamo interpolato i risultati per offrire un confronto equo.

La tabella seguente mostra il confronto dei tempi di training raw quando i dati utilizzati per il modello di foresta casuale Pandas sono significativamente inferiori (50 milioni di righe su 20 miliardi al giorno 15 del set di dati). Questo esempio utilizza solo meno del 0.25% di tutti i dati disponibili. Mentre per Dask-cuML abbiamo addestrato il modello di foresta casuale su tutti i 20 miliardi di righe disponibili. I due approcci hanno consentito di ottenere tempi di formazione comparabili.

Approccio	Tempo di training
Scikit-Learn: Utilizzando solo 50M righe nel giorno 15 come dati di training	47 minuti e 21 secondi
RAPIDS-Dask: Utilizzo di tutte le 20B righe del giorno 15 come dati di training	1 ora, 12 minuti e 11 secondi

Se si interpolano i risultati dei tempi di training in modo lineare, come mostrato nella tabella seguente, si ha un vantaggio significativo nell'utilizzo della formazione distribuita con Dask. L'approccio convenzionale Pandas scikit-Learn richiede 13 giorni per elaborare e formare 45 GB di dati per un singolo giorno di log click, mentre L'approccio RAPIDS-Dask elabora la stessa quantità di dati 262.39 volte più velocemente.

Approccio	Tempo di training
Scikit-Learn: Utilizzando tutte le 20B righe del giorno 15 come dati di training	13 giorni, 3 ore, 40 minuti e 11 secondi
RAPIDS-Dask: Utilizzo di tutte le 20B righe del giorno 15 come dati di training	1 ora, 12 minuti e 11 secondi

Nella tabella precedente, è possibile osservare che, utilizzando RAPIDS con Dask per distribuire

l'elaborazione dei dati e modellare la formazione su più istanze GPU, il tempo di esecuzione è significativamente più breve rispetto all'elaborazione convenzionale di Pandas DataFrame con il training del modello scikit-Learn. Questo framework consente la scalabilità verticale e orizzontale nel cloud e on-premise in un cluster multi-GPU a più nodi.

## Monitoraggio di Dask e RAPIDE con Prometheus e Grafana

Una volta implementato tutto, esegui le inferenze sui nuovi dati. I modelli prevedono se un utente fa clic su un annuncio in base alle attività di navigazione. I risultati della previsione sono memorizzati in un cuDF di Dask. Puoi monitorare i risultati con Prometheus e visualizzarli nelle dashboard Grafana.

Per ulteriori informazioni, consulta questa sezione ["RAPIDS ai Medium post"](#).

## Versione di set di dati e modelli con NetApp DataOps Toolkit

Il NetApp DataOps Toolkit per Kubernetes astratta le risorse di storage e i carichi di lavoro Kubernetes fino al livello di spazio di lavoro per la scienza dei dati. Queste funzionalità sono integrate in un'interfaccia semplice e facile da usare, progettata per data scientist e data engineer. Utilizzando la forma familiare di un programma Python, il Toolkit consente a data scientist e ingegneri di eseguire il provisioning e la distruzione delle aree di lavoro di JupyterLab in pochi secondi. Queste aree di lavoro possono contenere terabyte, o persino petabyte, di capacità di storage, consentendo agli scienziati dei dati di memorizzare tutti i set di dati di training direttamente nelle aree di lavoro dei progetti. Sono finiti i tempi della gestione separata degli spazi di lavoro e dei volumi di dati.

Per ulteriori informazioni, visitare il Toolkit ["Repository di GitHub"](#).

## Notebook Jupyter come riferimento

Al report tecnico sono associati due notebook Jupyter:

- **"CTR-PandasRF-collated.ipynb."** Questo notebook carica il giorno 15 dal set di dati Click Logs di Criteo Terabyte, elabora e formatta i dati in un Pandas DataFrame, forma un modello di foresta casuale Scikit-learn, esegue la previsione e calcola la precisione.
- **"criteo\_dask\_RF.ipynb."** Questo notebook carica il giorno 15 dal set di dati Click Logs di Criteo Terabyte, elabora e formatta i dati in un cuDF Dask, forma un modello di foresta casuale cuML Dask, esegue la previsione e calcola la precisione. Sfruttando nodi di lavoro multipli con GPU, questo approccio di elaborazione e formazione dei dati distribuiti e dei modelli è altamente efficiente. Maggiore è il numero di dati elaborati, maggiore è il risparmio di tempo rispetto a un approccio ML convenzionale. È possibile implementare questo notebook nel cloud, on-premise o in un ambiente ibrido in cui il cluster Kubernetes contiene calcolo e storage in posizioni diverse, purché la configurazione di rete consenta il libero spostamento dei dati e la distribuzione dei modelli.

## Conclusione

Azure NetApp Files, RAPIDS e Dak accelerano e semplificano l'implementazione dell'elaborazione E della formazione ML su larga scala integrandosi con strumenti di



orchestrazione come Docker e Kubernetes. Unificando la pipeline di dati end-to-end, questa soluzione riduce la latenza e la complessità inerenti a molti carichi di lavoro di calcolo avanzati, colmando efficacemente il divario tra sviluppo e operazioni. I data scientist possono eseguire query su set di dati di grandi dimensioni e condividere in modo sicuro dati e modelli algoritmici con altri utenti durante la fase di training.

Quando si creano pipeline ai/ML personalizzate, configurare l'integrazione, la gestione, la sicurezza e l'accessibilità dei componenti in un'architettura è un'attività complessa. Fornire agli sviluppatori l'accesso e il controllo del proprio ambiente presenta un'altra serie di sfide.

Creando un modello di training distribuito end-to-end e una pipeline di dati nel cloud, abbiamo dimostrato un miglioramento di due ordini di grandezza nel tempo totale di completamento del workflow rispetto a un approccio open-source convenzionale che non ha sfruttato i framework di elaborazione e di elaborazione dei dati accelerati dalla GPU.

La combinazione di NetApp, Microsoft, framework di orchestrazione open-source e NVIDIA riunisce le più recenti tecnologie come servizi gestiti con grande flessibilità per accelerare l'adozione della tecnologia e migliorare il time-to-market per le nuove applicazioni ai/ML. Questi servizi avanzati vengono forniti in un ambiente cloud nativo che può essere facilmente trasferito per architetture di implementazione on-premise e ibride.

## Dove trovare ulteriori informazioni

Per ulteriori informazioni sulle informazioni descritte in questo documento, consultare le seguenti risorse:

- Azure NetApp Files:
  - Pagina dell'architettura delle soluzioni per Azure NetApp Files  
["https://docs.microsoft.com/azure/azure-netapp-files/azure-netapp-files-solution-architectures"](https://docs.microsoft.com/azure/azure-netapp-files/azure-netapp-files-solution-architectures)
- Storage persistente Trident per container:
  - Azure NetApp Files e Trident  
["https://netapptrident.readthedocs.io/en/stablev20.07/kubernetes/operations/tasks/backends/anf.html"](https://netapptrident.readthedocs.io/en/stablev20.07/kubernetes/operations/tasks/backends/anf.html)
- Dask e RAPIDE:
  - Dek  
["https://docs.dask.org/en/latest/"](https://docs.dask.org/en/latest/)
  - Installare Dask  
["https://docs.dask.org/en/latest/install.html"](https://docs.dask.org/en/latest/install.html)
  - API di Dask  
["https://docs.dask.org/en/latest/api.html"](https://docs.dask.org/en/latest/api.html)
  - Apprendimento automatico di Dask  
["https://examples.dask.org/machine-learning.html"](https://examples.dask.org/machine-learning.html)

- DAK Distributed Diagnostics

["https://docs.dask.org/en/latest/diagnostics-distributed.html"](https://docs.dask.org/en/latest/diagnostics-distributed.html)

- Framework E tool ML:

- TensorFlow: Un framework di apprendimento automatico open-source per tutti

["https://www.tensorflow.org/"](https://www.tensorflow.org/)

- Docker

["https://docs.docker.com"](https://docs.docker.com)

- Kubernetes

["https://kubernetes.io/docs/home/"](https://kubernetes.io/docs/home/)

- Kubeflow

["http://www.kubeflow.org/"](http://www.kubeflow.org/)

- Jupyter notebook Server

["http://www.jupyter.org/"](http://www.jupyter.org/)

## Informazioni sul copyright

Copyright © 2024 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 COMMERCIALIZZABILITÀ 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.