



Formation distribuée dans Azure - prévision de taux par clic

NetApp Solutions

NetApp
April 25, 2024

Sommaire

- Formation distribuée dans Azure - prévision de taux par clic. 1
 - Tr-4904 : formation distribuée dans Azure - prévision de taux par clic 1
 - Présentation de la technologie 2
 - Configuration logicielle requise 4
 - Ressources cloud nécessaires 4
 - Récapitulatif de l'utilisation des prédictions de taux par clic 4
 - Configuration 6
 - Cliquez par le biais du traitement des données de prédiction de taux et de l'entraînement des modèles . . 16
 - Conclusion 23
 - Où trouver des informations complémentaires 24

Formation distribuée dans Azure - prévision de taux par clic

Tr-4904 : formation distribuée dans Azure - prévision de taux par clic

Rick Huang, Verron Martina, Muneer Ahmad, NetApp

Le travail d'un data Scientist doit se concentrer sur l'entraînement et l'ajustement des modèles de machine learning (ML) et d'intelligence artificielle (IA). Toutefois, selon une étude menée par Google, les data Scientists consacrent environ 80 % de leur temps à comprendre comment exploiter leurs modèles pour les applications d'entreprise et leur fonctionnement à grande échelle.

Pour gérer des projets d'IA et DE ML de bout en bout, il faut avoir une vision plus large des composants de l'entreprise. Bien que le DevOps ait pris en charge la définition, l'intégration et le déploiement, ces types de composants ciblent LES opérations DE ML un flux similaire qui inclut les projets d'IA/ML. Pour découvrir ce qu'un pipeline IA/ML de bout en bout touche dans l'entreprise, consultez la liste suivante de composants :

- Stockage
- Mise en réseau
- Les bases de données
- Systèmes de fichiers
- Conteneurs
- Intégration continue et pipeline de déploiement continu (ci/CD)
- Environnement de développement intégré (IDE)
- Sécurité
- Règles d'accès aux données
- Sous-jacent
- Le cloud
- Virtualisation
- Bibliothèques et jeux d'outils de science des données

Public visé

L'univers de la science des données touche plusieurs disciplines au NIVEAU DE L'INFORMATIQUE et des activités :

- Les data Scientists doivent donc pouvoir utiliser leurs outils et leurs bibliothèques de choix.
- L'ingénieur doit savoir comment les données circulent et où elles résident.
- Un ingénieur DevOps doit disposer des outils nécessaires pour intégrer les nouvelles applications d'IA et DE ML dans son pipeline ci/CD.
- Les administrateurs et architectes du cloud doivent pouvoir configurer et gérer les ressources Azure.

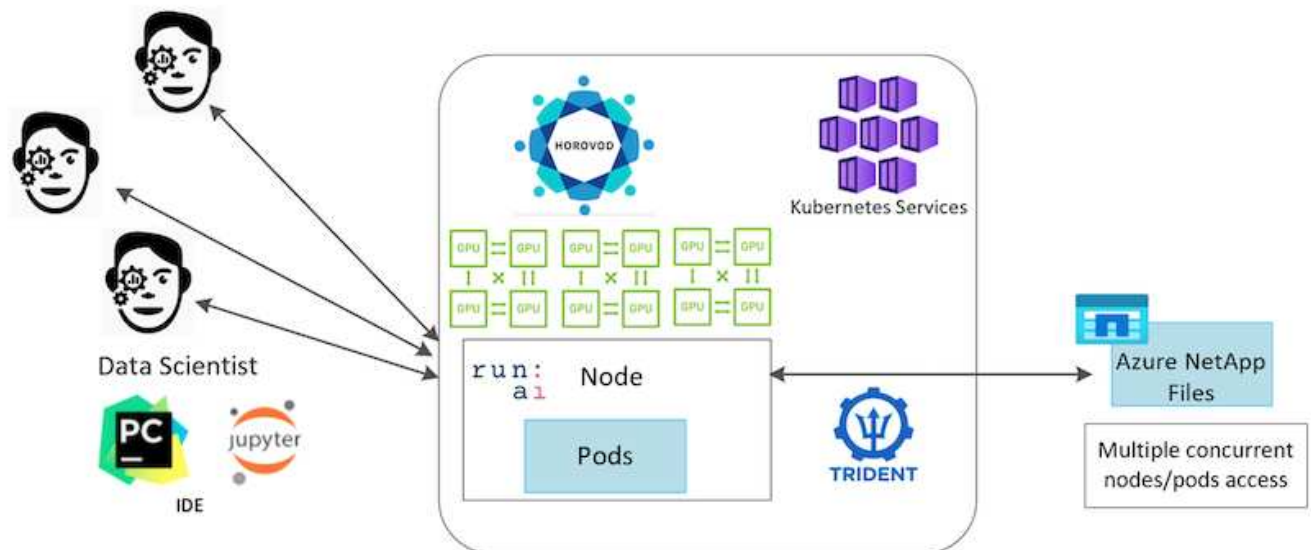
- Les utilisateurs professionnels veulent avoir accès aux applications d'IA et DE ML.

Dans ce rapport technique, nous décrivons comment Azure NetApp Files, RAPIDS IA, DASK et Azure aident chacun de ces rôles à générer de la valeur pour le business.

Présentation de la solution

Cette solution suit le cycle de vie d'une application d'IA/DE ML. Nous commençons par le travail des data Scientists pour définir les différentes étapes requises pour préparer les données et entraîner les modèles. Grâce À RAPIDS on DASK, nous procédons à une formation distribuée dans le cluster Azure Kubernetes Service (AKS) afin de réduire considérablement le temps d'entraînement par rapport à l'approche conventionnelle Python scikit d'apprentissage. Pour terminer le cycle complet, nous intégrons le pipeline à Azure NetApp Files.

Azure NetApp Files offre plusieurs tiers de performance. Les clients peuvent commencer avec un niveau standard, puis évoluer horizontalement et verticalement jusqu'à un niveau hautes performances sans interruption, sans déplacer aucune donnée. Ainsi, les data Scientists peuvent entraîner des modèles à grande échelle sans problèmes de performances, en évitant les silos de données sur le cluster, comme l'illustre la figure ci-dessous.



Présentation de la technologie

Cette page présente la technologie utilisée dans cette solution.

Microsoft et NetApp

Depuis mai 2019, Microsoft propose un service de portail Azure propriétaire natif pour les services de fichiers NFS et SMB d'entreprise basés sur la technologie NetApp ONTAP. Ce développement est régi par un partenariat stratégique entre Microsoft et NetApp, et étend encore davantage la portée des services de données ONTAP à Azure.

Azure NetApp Files

Le service Azure NetApp Files est un service de stockage de fichiers hautes performances et mesuré. Azure

NetApp Files prend en charge tous les types de charges de travail et est extrêmement disponible par défaut. Vous pouvez sélectionner des niveaux de service et de performances et configurer des copies Snapshot via le service. Azure NetApp Files est un service Azure propriétaire pour la migration et l'exécution des workloads de fichiers d'entreprise les plus exigeants dans le cloud, y compris les bases de données, SAP et les applications de calcul haute performance sans modifier le code.

Et présente plusieurs avantages pour les SERVICES IT :

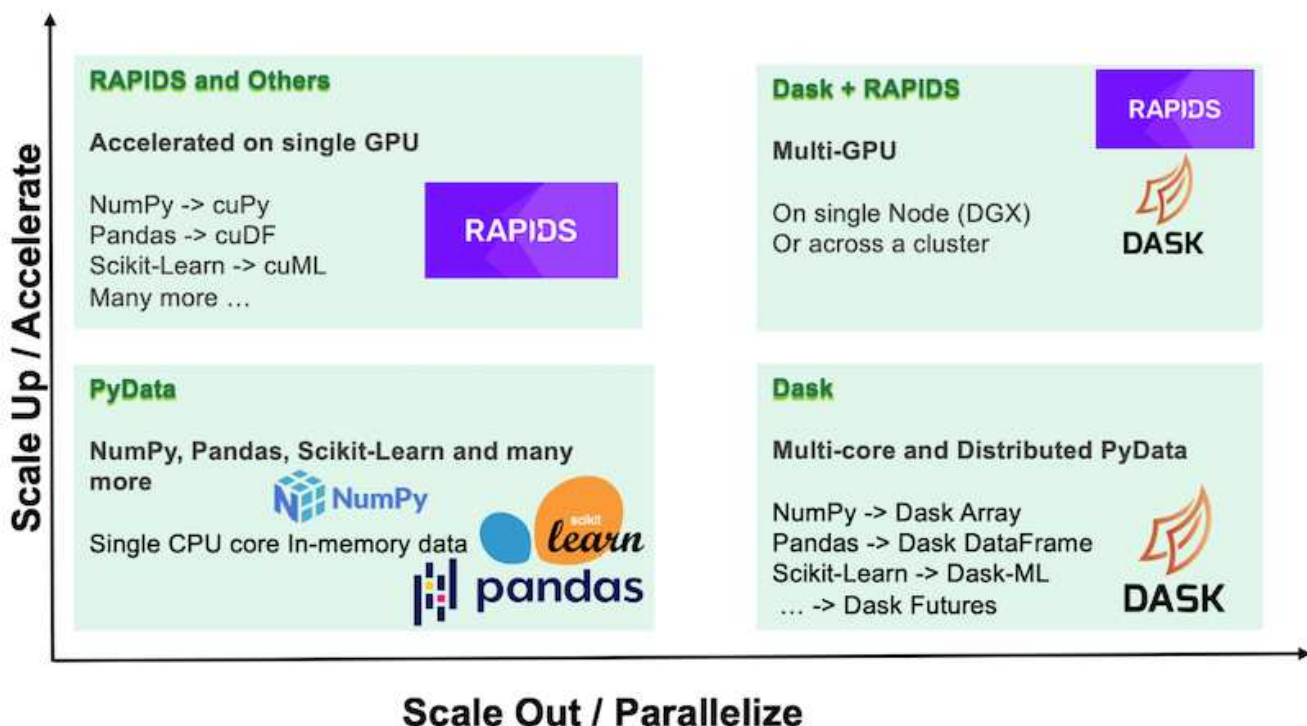
- Simplifie la conception
- Évolutivité indépendante des ressources de calcul et de stockage
- Possibilité de faire évoluer de manière fluide une infrastructure initiale de petite taille
- Propose plusieurs tiers de stockage pour répondre à des exigences variées de coûts et de performance

Présentation de DASK et NVIDIA RAPIDS

DASK est un outil informatique parallèle open source qui permet de faire évoluer les bibliothèques Python sur plusieurs machines et d'accélérer le traitement de grandes quantités de données. Il fournit une API similaire aux bibliothèques Python conventionnelles à thread unique, telles que Pandas, NumPy et scikit-Learn. Par conséquent, les utilisateurs Python natifs n'ont pas à modifier un grand nombre de leur code existant en vue d'utiliser les ressources sur l'ensemble du cluster.

NVIDIA RAPIDS est une suite de bibliothèques open source qui permet d'exécuter facilement les workflows D'AM et d'analytique de bout en bout sur les GPU. Avec DASK, il vous permet de passer facilement d'une station de travail GPU (évolutivité verticale) à des clusters multi-nœuds et multi-processeurs graphiques (évolutivité horizontale).

Pour le déploiement de DASK sur un cluster, vous pouvez utiliser Kubernetes pour l'orchestration des ressources. Vous pouvez également augmenter ou réduire les nœuds workers selon les besoins process, ce qui peut vous aider à optimiser la consommation des ressources du cluster, comme le montre la figure suivante.



Configuration logicielle requise

Le tableau suivant répertorie la configuration logicielle requise pour cette solution.

Logiciel	Version
Service Azure Kubernetes	1.18.14
Image conteneur RAPIDS et Dapose	Référentiel: "Rapidsai/rapidsai" tag: 0.17-cuda11.0-runtime-ubuntu18.04
NetApp Trident	20.01.1
Gouvernail	3.0.0

Ressources cloud nécessaires

Cette page décrit la configuration des ressources cloud pour Azure NetApp Files.

Configurez Azure NetApp Files

Configurez Azure NetApp Files comme décrit à la section ["Démarrage rapide : configurez Azure NetApp Files et créez un volume NFS"](#).

Vous pouvez passer outre la section « Créer un volume NFS pour Azure NetApp Files », car vous allez créer des volumes via Trident. Avant de continuer, procédez comme suit :

1. Inscrivez-vous à Azure NetApp Files et au fournisseur de ressources NetApp (via le shell Azure) (["lien"](#)).
2. Créer un compte dans Azure NetApp Files (["lien"](#)).
3. Configurez un pool de capacité (un minimum de 4 To Standard ou Premium, selon vos besoins) (["lien"](#)).le tableau suivant répertorie les configurations réseau requises pour la configuration dans le nuage. Le cluster DASK et Azure NetApp Files doivent se trouver sur le même réseau virtuel Azure (vnet) ou un réseau vnet pêche.

Ressources	Type/version
Service Azure Kubernetes	1.18.14
Nœud agent	3 x Standard_DS2_v2
Nœud de processeur graphique	3 x Standard_NC6s_v3
Azure NetApp Files	Pool de capacité standard
Capacité en To	4

Récapitulatif de l'utilisation des prédictions de taux par clic

Il s'agit d'un cas d'utilisation basé sur le cloud public ["Fichiers de clic sur un téraoctet"](#) jeu de données de ["Criteo ai Lab"](#). Avec les récentes avancées des plates-formes ET des applications DE ML, une grande attention se porte maintenant sur l'apprentissage à grande échelle. Le taux de clics (CTR) est défini comme le nombre moyen de clics par cent impressions de publicités en ligne (exprimé en pourcentage). Elle est largement

adoptée comme indicateur clé dans différents secteurs d'activité et champs d'application, notamment le marketing digital, la vente au détail, l'e-commerce et les fournisseurs de services. Voici quelques exemples d'utilisation de CTR comme mesure importante pour le trafic potentiel des clients :

- **Marketing numérique:** in "[Google Analytics](#)", CTR peut être utilisé pour évaluer la manière dont bien un annonceur ou les mots-clés, annonces, et les listes libres sont en cours de réalisation. Un CTR élevé est une bonne indication que les utilisateurs trouvent vos annonces et listes utiles et pertinentes. CTR contribue également à votre mot clé CTR attendu, qui est un composant de "[Rang d'annonce](#)".
- *** E-commerce:** en plus de tirer parti "[Google Analytics](#)", il y a au moins quelques statistiques de visiteurs dans un back-end de commerce électronique. Bien que ces statistiques ne semblent pas utiles à première vue, elles sont généralement faciles à lire et peuvent être plus précises que d'autres informations. Les ensembles de données de première partie composés de ces statistiques sont propriétaires et sont donc les plus pertinents pour les vendeurs, les acheteurs et les plates-formes de commerce électronique. Ces ensembles de données peuvent être utilisés pour établir des bancs d'essai, en comparant les résultats à l'année dernière et à la veille, en élaborant une série chronologique pour une analyse plus approfondie.
- **Retail:** les détaillants Brick-et-mortier peuvent mettre en corrélation le nombre de visiteurs et le nombre de clients avec le CTR. Le nombre de clients est visible depuis leur historique des points de vente. Le CTR sur les sites Web des détaillants ou le trafic publicitaire peut entraîner les ventes susmentionnées. Les programmes de fidélité sont un autre cas d'utilisation, car les clients redirigés depuis des publicités en ligne ou d'autres sites Web peuvent se joindre pour gagner des récompenses. Les détaillants peuvent acquérir des clients par le biais de programmes de fidélité et enregistrer des comportements liés aux histoires de vente pour élaborer un système de recommandations qui non seulement prédit les comportements d'achat des consommateurs dans différentes catégories, mais également personnalise les bons de réduction et diminue le nombre de départs.
- **Fournisseurs de services:** les entreprises de télécommunications et les fournisseurs de services Internet disposent d'une abondance de données de télémétrie utilisateur de première partie pour des cas d'utilisation instructifs de l'IA, DU ML et de l'analytique. Par exemple, un opérateur télécom peut utiliser quotidiennement ses journaux d'historique de domaine de premier niveau pour affiner les modèles existants afin de générer une segmentation de l'audience à jour, d'anticiper le comportement des clients et de collaborer avec les annonceurs pour placer des publicités en temps réel et améliorer l'expérience en ligne. Dans ce flux de travail de marketing axé sur les données, CTR est une mesure importante pour refléter les conversions.

Dans le cadre du marketing numérique, "[Criteo Terabyte cliquez sur journaux](#)" Sont désormais le dataset de référence pour évaluer l'évolutivité des plateformes ET des algorithmes DE ML. En prédisant le taux de clics, un annonceur peut sélectionner les visiteurs qui sont les plus susceptibles de répondre aux annonces, analyser leur historique de navigation, et montrer les annonces les plus pertinentes en fonction des intérêts de l'utilisateur.

La solution proposée dans ce rapport technique présente les avantages suivants :

- Avantages de Azure NetApp Files dans le cadre d'une formation distribuée ou à grande échelle
- RAPIDES, traitement de données CUDA (cuDF, cuPy, etc.) et algorithmes DE ML (cuML)
- Le cadre informatique parallèle DASK pour la formation distribuée

Un workflow complet basé SUR RAPIDS dans l'IA et Azure NetApp Files démontre l'amélioration drastique du temps d'entraînement des modèles forestiers aléatoires de deux ordres de grandeur. Cette amélioration est très significative par rapport à l'approche Pandas classique lors du traitement des journaux de clics du monde réel avec 45 Go de données tabulaires structurées (en moyenne) chaque jour. Cela équivaut à un DataFrame contenant environ vingt milliards de lignes. Dans ce rapport technique, nous allons présenter la configuration de l'environnement de cluster, l'installation de la structure et de la bibliothèque, le chargement et le traitement

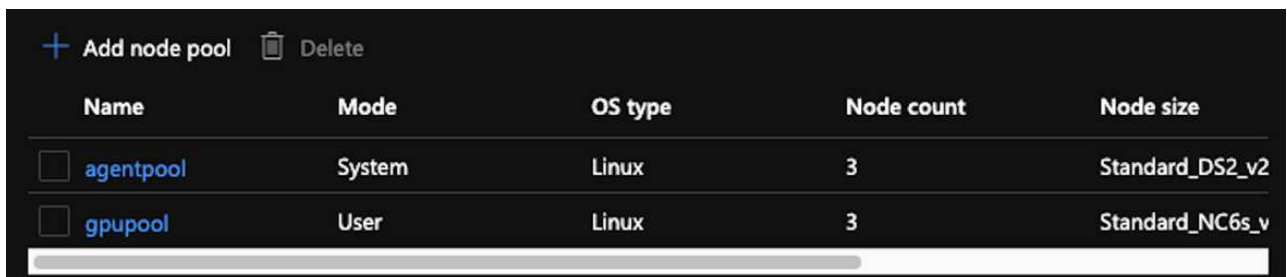
des données, les méthodes conventionnelles par rapport à la formation distribuée, la visualisation et la surveillance, et comparer les résultats critiques de l'exécution de bout en bout.

Configuration

Installez et configurez le cluster AKS

Pour installer et configurer le cluster AKS, reportez-vous à la page Web ["Créer un cluster AKS"](#) puis procédez comme suit :

1. Lors de la sélection du type de nœud (nœuds du système [CPU] ou du worker [GPU]), sélectionnez ce qui suit :
 - a. Les nœuds du système primaire doivent être de type Standard DS2v2 (`agentpool` trois nœuds par défaut).
 - b. Ajoutez ensuite le nœud de travail Standard_NC6s_v3 pool (trois nœuds minimum) pour le groupe d'utilisateurs (pour les nœuds GPU) nommé `gpupool`.



Name	Mode	OS type	Node count	Node size
<input type="checkbox"/> <code>agentpool</code>	System	Linux	3	Standard_DS2_v2
<input type="checkbox"/> <code>gpupool</code>	User	Linux	3	Standard_NC6s_v

2. Le déploiement prend entre 5 et 10 minutes. Lorsque l'opération est terminée, cliquez sur Connect to Cluster.
3. Pour vous connecter au cluster AKS nouvellement créé, installez les éléments suivants à partir de votre environnement local (ordinateur portable/pc) :
 - a. L'outil de ligne de commande Kubernetes à l'aide de ["Instructions fournies pour votre système d'exploitation"](#)
 - b. L'interface de ligne de commandes Azure, comme décrit dans le document, ["Installez l'interface de ligne de commande Azure"](#)
4. Pour accéder au cluster AKS à partir du terminal, entrez `az login` et entrez les informations d'identification.
5. Exécutez les deux commandes suivantes :

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

6. Entrez Azure CLI: `kubectl get nodes`.
7. Si les six nœuds sont actifs, comme illustré dans l'exemple suivant, votre cluster AKS est prêt et connecté à votre environnement local


```
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 ~ %
```

Créez un sous-réseau délégué pour Azure NetApp Files

Pour créer un sous-réseau délégué pour Azure NetApp Files, effectuez les opérations suivantes :

1. Accédez aux réseaux virtuels depuis le portail Azure. Trouvez votre nouveau réseau virtuel. Il doit avoir un préfixe tel que aks-vnet.
2. Cliquez sur le nom du 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. Cliquez sur sous-réseaux et sur +sous-réseau dans la barre d'outils supérieure.

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

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

Search subnets

Name ↑↓	IPv4 ↑↓	IPv6 (many availab... ↑↓	Delegated to ↑↓	Security group ↑↓
aks-subnet	10.240.0.0/16 (65530 av...	-	-	aks-agentpool-22885919...

4. Indiquez au sous-réseau un nom tel que `ANF.sn` Et, sous l'en-tête délégation de sous-réseau, sélectionnez `Microsoft.Netapp/volumes`. Ne rien changer. Cliquez sur 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

Les volumes Azure NetApp Files sont alloués au cluster d'applications et utilisés en tant que demandes de volume persistant dans Kubernetes. Par conséquent, ce processus vous offre la possibilité de les associer à différents services, tels que les ordinateurs portables Jupyter, les fonctions sans serveur, etc.

Les utilisateurs des services peuvent consommer le stockage depuis la plateforme de différentes manières. Dans ce rapport technique, NFSS, les principaux avantages de Azure NetApp Files sont les suivants :

- Possibilité d'utiliser des copies Snapshot pour les utilisateurs.
- Possibilité pour les utilisateurs de stocker d'importants volumes de données sur des volumes Azure NetApp Files.
- Avantages des performances des volumes Azure NetApp Files lors de l'exécution de modèles sur de

vastes ensembles de fichiers.

Peer AKS vnet et Azure NetApp Files vnet

Pour passer le réseau VNet AKS à Azure NetApp Files VNet, procédez comme suit :

1. Entrez les réseaux virtuels dans le champ de recherche.
2. Sélectionnez `vnet aks-vnet-name`. Cliquez dessus et saisissez Peerings dans le champ de recherche.
3. Cliquez sur +Ajouter.
4. Saisissez les descripteurs suivants :
 - a. Le nom de la liaison de peering est `aks-vnet-name_to_anf`.
 - b. Indice de connexion et Azure NetApp Files vnet en tant que partenaire de peering vnet.
 - c. Laissez toutes les sections non astérisque avec les valeurs par défaut.
5. Cliquez sur Ajouter.

Pour plus d'informations, voir ["Créez, modifiez ou supprimez un peering de réseau virtuel"](#).

Installation de Trident

Pour installer Trident à l'aide de Helm, effectuez les opérations suivantes :

1. Installez Helm (pour obtenir des instructions d'installation, consultez le ["source"](#)).
2. Téléchargez et extrayez le programme d'installation de Trident 20.01.1.

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

3. Définissez le répertoire sur `trident-installer`.

```
$cd trident-installer
```

4. Copier `tridentctl` dans un répertoire de votre système `$PATH`.

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

5. Installez Trident sur le cluster Kubernetes (K8s) avec Helm (["source"](#)) :
 - a. Changez le répertoire en `helm` répertoire.

```
$cd helm
```

- b. Installation de Trident.

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

c. Vérifier l'état des pods Trident.

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

Si tous les modules sont opérationnels, Trident est installé et vous pouvez passer à l'étape supérieure.

6. Configurer le système Azure NetApp Files backend et la classe de stockage pour AKS.

a. Créer un principe de service Azure

Le principal service est de la façon dont Trident communique avec Azure pour manipuler vos ressources Azure NetApp Files.

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

Le résultat de la commande doit ressembler à l'exemple suivant :

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

7. Créez un fichier json backend Trident, par exemple un nom anf-backend.json.

8. À l'aide de votre éditeur de texte préféré, renseignez les champs suivants à l'intérieur du anf-backend.json fichier :

```
{
  "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. Remplacez les champs suivants :

- `subscriptionID`. Votre ID d'abonnement Azure.
- `tenantID`. Votre ID de locataire Azure à partir des résultats de `az ad sp` à l'étape précédente.
- `clientID`. Votre AppID à partir de la sortie de `az ad sp` à l'étape précédente.
- `clientSecret`. Votre mot de passe à partir de la sortie de `az ad sp` à l'étape précédente.

10. Demandez à Trident de créer le back-end Azure NetApp Files dans le système `trident` espace de noms avec `anf-backend.json` comme fichier de configuration :

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

NAME	STORAGE DRIVER	UUID	STATE	VOLUMES
azurenetafiles_86181	azure-netapp-files	2ca85462-59ac-4946-be05-c03f5575a2ad	online	0

11. Créer une classe de stockage. Les utilisateurs Kubernetes provisionnent les volumes à l'aide des demandes de volume qui spécifient une classe de stockage par nom. Demandez à K8s de créer une classe de stockage `azurenetafiles` Qui fait référence au système back-end Trident créé dans l'étape précédente.
12. Créez un YAML (`anf-storage-class.yaml`) fichier pour la classe de stockage et la copie.

```

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. Vérifiez que la classe de stockage a été créée.

```
kubectl get sc azurenetappfiles
```

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

Mise en place de DASK avec RAPIDS sur AKS à l'aide de Helm

Pour configurer le déploiement de DASK avec RAPIDS sur AKS à l'aide de Helm, procédez comme suit :

1. Créer un espace de noms pour l'installation de DASK avec RAPIDS.

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

2. Création d'une demande de volume persistant pour stocker le dataset de taux de clics :

a. Enregistrez le contenu YAML suivant dans un fichier pour créer une demande de volume persistant.

```

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

```

b. Appliquez le fichier YAML sur votre cluster Kubernetes.

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

3. Cloner le rapidsai git référentiel ("<https://github.com/rapidsai/helm-chart>").

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

4. Modifier `values.yaml` Et inclure le PVC créé précédemment pour les travailleurs et l'espace de travail Jupyter.

- a. Accédez au rapidsai répertoire du référentiel.

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

- b. Mettez à jour le `values.yaml` Fichier et montage du volume à l'aide de la demande de volume persistant.

```
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. Accédez au répertoire personnel du référentiel et déployez DASK avec trois noeuds de travail sur AKS en utilisant Helm.

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

Tiers de performance Azure NetApp Files

Vous pouvez modifier le niveau de service d'un volume existant en déplaçant ce volume vers un autre pool de capacité qui utilise le niveau de service souhaité pour le volume. Elle permet de commencer avec un petit dataset et un petit nombre de GPU dans le Tier standard, puis de monter en charge horizontalement ou verticalement jusqu'à Premium Tier en fonction du volume de données et du nombre de GPU. Le niveau Premium offre un débit par téraoctet avec un niveau standard, et il est nécessaire de faire évoluer l'infrastructure sans déplacer de données pour modifier le niveau de service d'un volume.

Modification dynamique du niveau de service d'un volume

Pour modifier de manière dynamique le niveau de service d'un volume, procédez comme suit :

1. Sur la page volumes, cliquez avec le bouton droit de la souris sur le volume dont vous souhaitez modifier le niveau de service. Sélectionnez Modifier le pool.

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

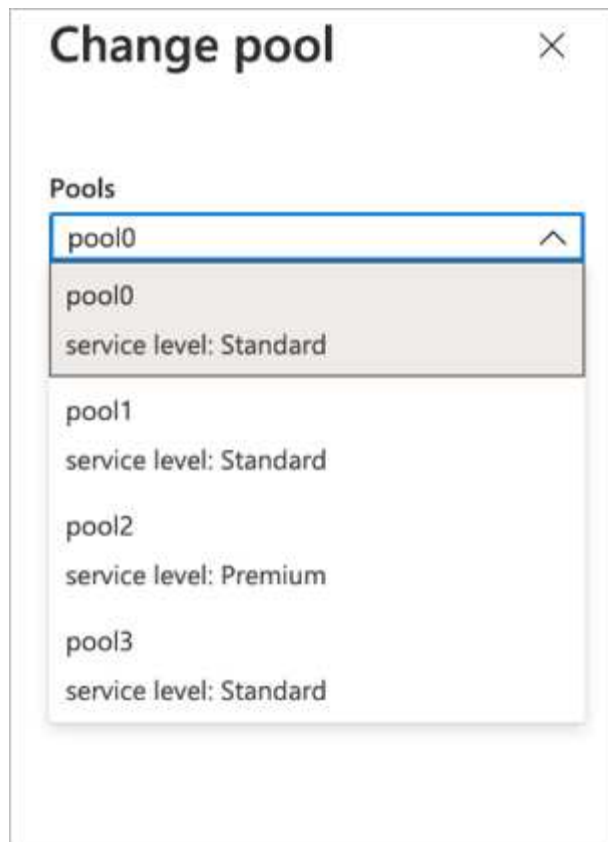
Resize

Edit

Change pool

Delete

2. Dans la fenêtre change Pool, sélectionnez le pool de capacité vers lequel vous souhaitez déplacer le volume.



3. Cliquez sur OK.

Automatisez les changements au niveau des tiers de performance

Plusieurs options sont disponibles pour automatiser les modifications des tiers de performance :

- Le changement de niveau de service dynamique est toujours dans l'aperçu public à ce moment et n'est pas activé par défaut. Pour activer cette fonctionnalité sur l'abonnement Azure, consultez cette [documentation](#) sur la procédure à suivre "[Modification dynamique du niveau de service d'un volume](#)".
- Les commandes de modification du pool de volumes de l'interface de ligne de commande Azure sont fournies dans le "[documentation sur les modifications apportées au pool de volumes](#)" et dans l'exemple suivant :

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

- PowerShell : le "[Cmdlet Set-AzNetAppFilesVolumePool](#)" Modifie le pool d'un volume Azure NetApp Files et est présenté dans l'exemple suivant :

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

Cliquez par le biais du traitement des données de prédiction de taux et de l'entraînement des modèles

Bibliothèques de traitement de données et d'entraînement des modèles

Le tableau suivant répertorie les bibliothèques et les structures utilisées pour créer cette tâche. Tous ces composants ont été entièrement intégrés aux contrôles d'accès et de sécurité basés sur les rôles d'Azure.

Bibliothèques/structures	Description
DMASK cuML	Pour QUE LE ML fonctionne sur le GPU, le "Bibliothèque cuML" Donne accès au forfait RAPIDS cuML avec Dask. RAPIDS cuML implémente des algorithmes DE ML courants, notamment la mise en cluster, la réduction de la dimensionnalité et les approches de régression, avec des implémentations basées sur GPU haute performance, offrant des vitesses jusqu'à 100 fois supérieures aux approches basées sur les processeurs.
DASK cuDF	CuDF inclut diverses autres fonctions prenant en charge l'extraction accélérée par GPU, la transformation, la charge (ETL), telles que la sous-définition de données, les transformations, l'encodage à chaud unique, etc. L'équipe DE RAPIDS a "bibliothèque dask-cudf" Cela inclut des méthodes d'aide pour utiliser DASK et cuDF.
Apprendre Scikit	Scikit-Learn fournit des dizaines d'algorithmes et de modèles d'apprentissage machine intégrés, appelés estimateurs. Chacun "estimateur" peut être installé sur certaines données à l'aide de son "ajustement" méthode.

Nous avons utilisé deux ordinateurs portables pour construire les pipelines DE ML à des fins de comparaison. L'une est l'approche classique Pandas scikit d'apprentissage et l'autre, une formation distribuée à RAPIDS et DASK. Chaque ordinateur portable peut être testé individuellement pour connaître les performances en termes de temps et d'échelle. Nous recouvrons chaque ordinateur portable individuellement pour démontrer les avantages de la formation distribuée à L'aide DE RAPIDS et DASK.

Charger Criteo Click Logs Day 15 dans Pandas et former un modèle de forêt aléatoire de scikit

Cette section décrit comment nous avons utilisé Pandas et DASK DataFrames pour charger les données de journaux de clic du dataset Criteo Terabyte. Le cas d'utilisation est pertinent dans la publicité numérique pour les échanges publicitaires afin de créer les profils des utilisateurs en prédisant si les annonces seront cliqué ou si l'échange n'utilise pas un modèle précis dans un pipeline automatisé.

Nous avons chargé les données du jour 15 à partir du jeu de données Click Logs, soit un total de 45 Go. Exécution de la cellule suivante dans le bloc-notes Jupyter `CTR-PandasRF-collated.ipynb` Crée un Pandas DataFrame contenant les 50 premiers millions de lignes et génère un modèle aléatoire d'apprentissage de forêt de scikit.

```

%%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)

```

Pour effectuer une prédiction à l'aide d'un modèle de forêt aléatoire entraîné, exécutez le paragraphe suivant dans ce bloc-notes. Nous avons effectué le dernier million de lignes à partir du jour 15 comme jeu de tests pour éviter toute duplication. La cellule calcule également la précision de la prévision, définie comme le pourcentage d'occurrences que le modèle prédit avec précision si un utilisateur clique ou non sur une annonce. Pour passer en revue tous les composants inconnus de cet ordinateur portable, reportez-vous au ["documentation officielle d'apprentissage du kit de science"](#).

```
# 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('coll1')
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))
```

Charger le jour 15 à DASK et former un modèle forestier aléatoire de DASK cuML

D'une manière similaire à la section précédente, chargez Criteo Click Logs Day 15 dans Pandas et entraînez un modèle de forêt aléatoire d'apprentissage de scikit. Dans cet exemple, nous avons effectué le chargement de DataFrame avec DASK cuDF et formé un modèle de forêt aléatoire dans DASK cuML. Nous avons comparé les différences en termes de temps et d'échelle d'entraînement dans la section ["Comparaison du temps de formation."](#)

criteo_dAsk_RF.ipynb

Cet ordinateur portable importe numpy, cuml, et le nécessaire dask bibliothèques, comme illustré dans l'exemple suivant :

```
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
```

Lancez DASK client().

```
client = Client()
```

Si le cluster est configuré correctement, vous pouvez afficher l'état des nœuds workers.

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

Dans notre cluster AKS, l'état suivant s'affiche :

Client	Cluster
Scheduler: tcp://rapidsai-scheduler:8786	Workers: 3
Dashboard: /proxy/rapidsai-scheduler:8787/status	Cores: 3
	Memory: 354.55 GB

Notez que DASK utilise le paradigme d'exécution paresseux : plutôt que d'exécuter le code de traitement instantanément, DASK construit plutôt un graphe acyclique dirigé (DAG) d'exécution. DAG contient un ensemble de tâches et leurs interactions dont chaque employé a besoin pour s'exécuter. Cette disposition signifie que les tâches ne sont pas exécutées tant que l'utilisateur n'a pas dit à DASK de les exécuter d'une manière ou d'une autre. Avec DASK, vous disposez de trois options principales :

- **Call Compute() sur un DataFrame.** cet appel traite toutes les partitions, puis renvoie les résultats au planificateur pour l'agrégation finale et la conversion en cuDF DataFrame. Cette option doit être utilisée avec parcimonie et avec des résultats fortement réduits, à moins que le nœud du planificateur ne manque de mémoire.
- **Call persistent() sur un DataFrame.** cet appel exécute le graphique, mais, au lieu de renvoyer les résultats au nœud du planificateur, il les maintient dans le cluster en mémoire afin que l'utilisateur puisse réutiliser ces résultats intermédiaires dans le pipeline sans avoir besoin de réexécuter le même traitement.
- **Call head() sur un DataFrame.** tout comme avec cuDF, cet appel renvoie 10 enregistrements au nœud du planificateur. Cette option permet de vérifier rapidement si votre DataFrame contient le format de sortie souhaité ou si les enregistrements eux-mêmes ont un sens, selon votre traitement et votre calcul.

Par conséquent, à moins que l'utilisateur n'appelle l'une ou l'autre de ces actions, les travailleurs restent inactifs en attendant que le planificateur lance le traitement. Ce modèle d'exécution paresseux est courant dans les infrastructures informatiques modernes parallèles et distribuées telles qu'Apache Spark.

Le paragraphe suivant forme un modèle de forêt aléatoire en utilisant DASK cuML pour le calcul accéléré par GPU distribué et calcule la précision de prévision des modèles.

```

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 )

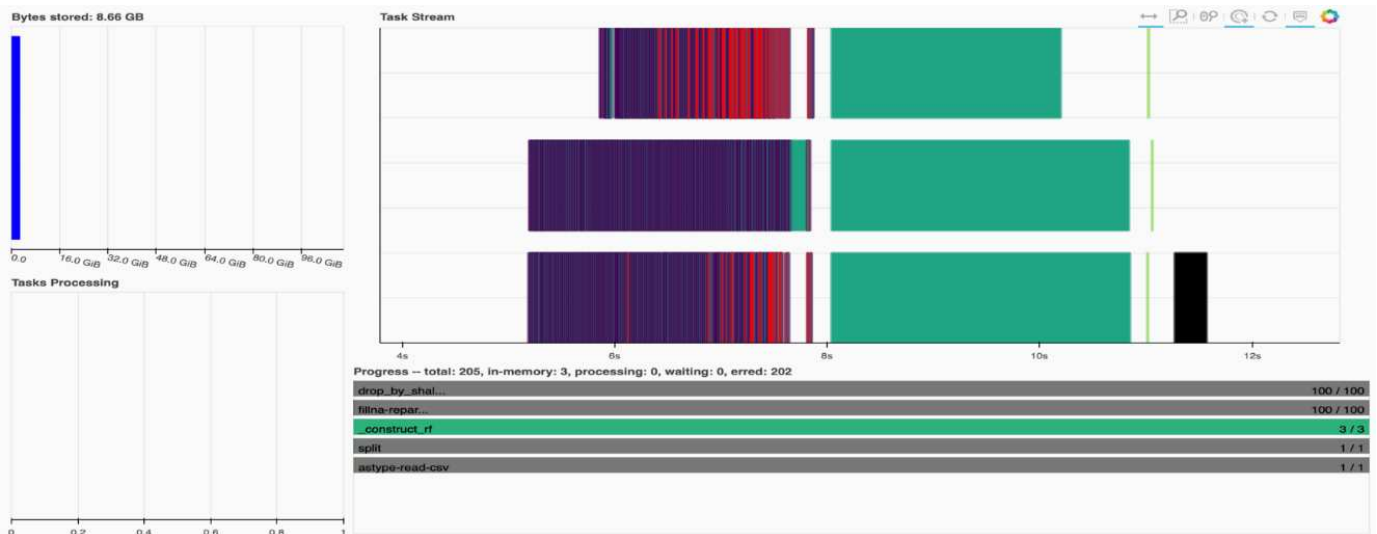
```

Surveiller le DASK à l'aide du tableau de bord des flux de tâches natifs

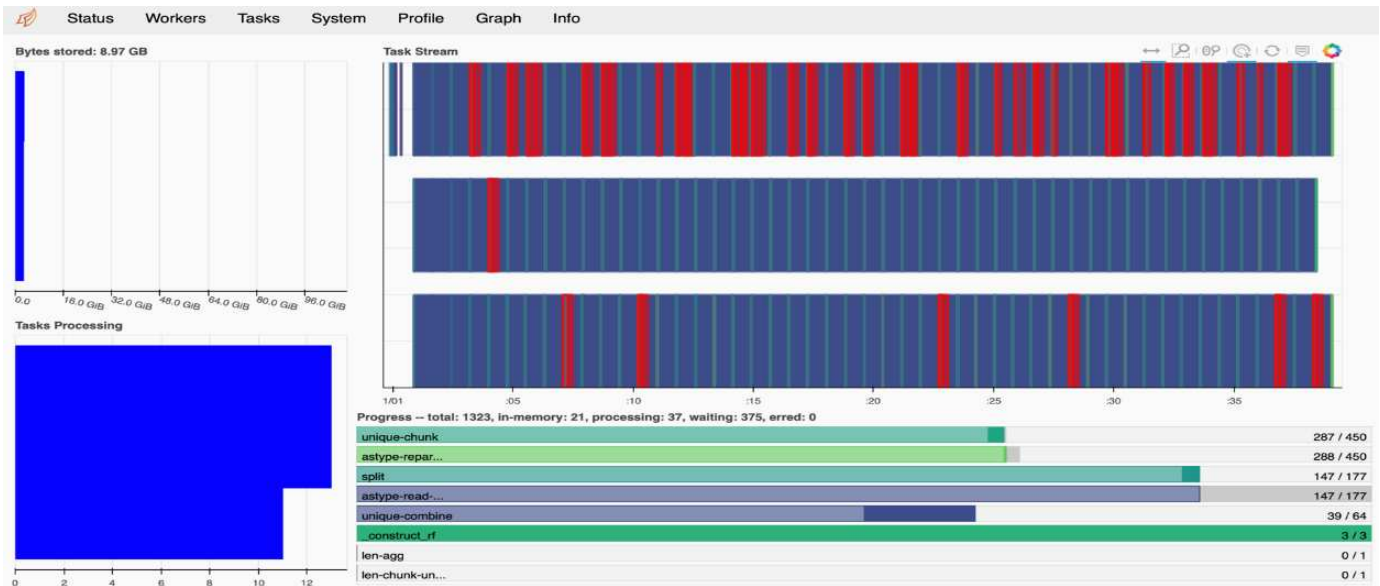
Le "[Planificateur distribué DASK](#)" fournit des commentaires en direct sous deux formes :

- Un tableau de bord interactif contenant de nombreux tracés et tableaux avec des informations en direct
- Barre de progression adaptée à une utilisation interactive dans les consoles ou les ordinateurs portables

Dans notre cas, la figure suivante montre comment surveiller la progression de la tâche, y compris les octets stockés, le flux de tâches avec une répartition détaillée du nombre de flux et la progression par nom de tâche avec les fonctions associées exécutées. Dans notre cas, étant donné que nous avons trois nœuds workers, il y a trois principaux segments de flux et les codes de couleurs indiquent des tâches différentes dans chaque flux.



Vous avez la possibilité d'analyser des tâches individuelles et d'examiner le temps d'exécution en millisecondes ou d'identifier tout obstacle ou obstacle. Par exemple, la figure suivante montre les flux de tâches pour le stade d'ajustement du modèle forestier aléatoire. Il y a beaucoup plus de fonctions exécutées, y compris le bloc unique pour le traitement de DataFrame, `_construct_rf` pour l'adaptation de la forêt aléatoire, et ainsi de suite. La plupart du temps a été consacré à des opérations DataFrame en raison de la grande taille (45 Go) des données d'une journée provenant des journaux Criteo Click.



Comparaison de la durée de formation

Cette section compare le temps d'entraînement du modèle à l'aide de Pandas conventionnels par rapport à DASK. Pour Pandas, nous avons chargé une plus petite quantité de données en raison de la nature du temps de traitement plus lent pour éviter le débordement de mémoire. Par conséquent, nous avons interpolé les résultats pour offrir une comparaison équitable.

Le tableau suivant montre la comparaison du temps d'entraînement brut lorsque le modèle Pandas utilise beaucoup moins de données pour sa forêt aléatoire (50 millions de lignes sur une base de 20 milliards par jour et 15 du dataset). Cet échantillon utilise uniquement moins de 0.25 % de toutes les données disponibles. Alors que pour DASK-cuML nous avons formé le modèle forestier aléatoire sur les 20 milliards de lignes disponibles. Les deux approches ont donné un temps de formation comparable.

Approche	Temps de formation
Scikit-Learn: L'utilisation de seulement 50M rangées en jour 15 comme données d'entraînement	47 minutes et 21 secondes
RAPIDS-DASK : utiliser les 20B rangées en jour 15 comme données d'entraînement	1 heure, 12 minutes et 11 secondes

Si nous interpoler les résultats de temps d'entraînement de façon linéaire, comme le montre le tableau suivant, il y a un avantage significatif à utiliser la formation distribuée avec DASK. L'approche classique Pandas scikit-Learn prendra 13 jours pour traiter et entraîner 45 Go de données pour une seule journée de journaux, tandis QUE L'approche RAPIDS-DASK traite la même quantité de données 262.39 fois plus vite.

Approche	Temps de formation
Scikit-Learn: Utiliser les données d'entraînement de toutes les lignes 20B en jour 15	13 jours, 3 heures, 40 minutes et 11 secondes
RAPIDS-DASK : utiliser les 20B rangées en jour 15 comme données d'entraînement	1 heure, 12 minutes et 11 secondes

Le tableau précédent montre que, GRÂCE À RAPIDS et DASK, le traitement des données et l'entraînement

des modèles sur plusieurs instances GPU, la durée d'exécution est considérablement plus courte que le traitement classique Pandas DataFrame avec entraînement des modèles scikit. Cette structure permet une évolutivité verticale et horizontale dans le cloud et sur site au sein d'un cluster multinœud et multiprocesseur graphique.

Surveillez DAsk et RAPIDS avec Prometheus et Grafana

Une fois que tout est déployé, exécutez des inférences sur les nouvelles données. Les modèles prédisent si un utilisateur clique sur une publicité en fonction des activités de navigation. Les résultats de la prédiction sont stockés dans un DASK cuDF. Vous pouvez surveiller les résultats avec Prometheus et visualiser dans les tableaux de bord Grafana.

Pour plus d'informations, reportez-vous à ce document ["RAPIDS ai moyen post"](#).

Gestion des versions de datasets et de modèles avec le kit NetApp DataOps

Le kit NetApp DataOps pour Kubernetes extrait les ressources de stockage et les workloads Kubernetes jusqu'à l'espace de travail data-science. Ces fonctionnalités sont packagées dans une interface simple et conviviale conçue pour les data Scientists et les ingénieurs de données. Ce kit, qui se présente sous la forme familière d'un programme Python, permet aux scientifiques et aux ingénieurs des données de provisionner et de déployer des espaces de travail JupyterLab en quelques secondes. Ces espaces de travail peuvent contenir des téraoctets, voire des pétaoctets, de capacité de stockage, ce qui permet aux data Scientists de stocker tous leurs datasets d'entraînement directement dans leur espace de travail de projet. Il ne fait plus que gérer séparément les espaces de travail et les volumes de données.

Pour en savoir plus, visitez la boîte à outils ["Référentiel GitHub"](#).

Ordinateurs portables Jupyter à titre de référence

Deux ordinateurs portables Jupyter sont associés à ce rapport technique :

- **"CTR-PandasRF-Assemblé.ipynb."** Cet ordinateur portable charge le jour 15 à partir du jeu de données de journaux Criteo Terabyte Click, traite et formate les données dans un Pandas DataFrame, forme un modèle de forêt aléatoire d'apprentissage Scikit, effectue des prédictions et calcule la précision.
- **"criteo_dAsk_RF.ipynb."** Cet ordinateur portable charge le jour 15 à partir du jeu de données de journal Criteo Terabun Click Logs, traite et formate les données dans un DASK cuDF, forme un modèle forestier aléatoire DASK cuML, effectue des prévisions et calcule la précision. L'exploitation de plusieurs nœuds workers avec des processeurs graphiques permet de tirer parti de cette approche de traitement et d'entraînement des modèles distribués pour les données. Plus vous traitez de données, plus vous gagnez de temps par rapport à une approche ML classique. Vous pouvez déployer cet ordinateur portable dans le cloud, sur site ou dans un environnement hybride dans lequel le cluster Kubernetes contient des ressources de calcul et de stockage à différents emplacements, à condition que la configuration réseau permet la libre circulation des données et de la distribution des modèles.

Conclusion

Azure NetApp Files, RAPIDS et DASK accélèrent et simplifient le déploiement du

traitement DE ML et de la formation à grande échelle en intégrant les outils d'orchestration tels que Docker et Kubernetes. En unifiant le pipeline de données de bout en bout, cette solution réduit la latence et la complexité inhérentes à de nombreuses charges de travail informatiques avancées, afin de combler efficacement l'écart entre le développement et les opérations. Les data Scientists peuvent exécuter des requêtes sur d'importants jeux de données et partager en toute sécurité les données et les modèles algorithmiques avec d'autres utilisateurs au cours de la phase d'entraînement.

Lors de la création de vos propres pipelines d'IA et DE ML, la configuration de l'intégration, de la gestion, de la sécurité et de l'accessibilité des composants d'une architecture représente un véritable défi. Le fait de donner aux développeurs accès à leur environnement et de le contrôler constitue un autre ensemble de défis.

En créant un modèle d'entraînement distribué complet et un pipeline de données dans le cloud, nous avons mis en avant deux ordres d'amélioration considérable au niveau du temps d'exécution total des workflows, par rapport à une approche open source classique qui n'exploite pas les frameworks de calcul et de traitement des données accélérés par GPU.

L'association de NetApp, Microsoft, des frameworks d'orchestration open source et NVIDIA rassemble les dernières technologies en tant que services gérés, tout en offrant une grande flexibilité pour accélérer l'adoption de technologies et améliorer le délai de mise sur le marché de nouvelles applications d'IA et DE ML. Ces services avancés sont fournis dans un environnement cloud natif facilement porté pour les architectures de déploiement sur site ainsi que pour les architectures de déploiement hybrides.

Où trouver des informations complémentaires

Pour en savoir plus sur les informations données dans ce document, consultez les ressources suivantes :

- Azure NetApp Files :

- Architecture des solutions pour 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)

- Trident pour le stockage persistant pour les conteneurs :

- Azure NetApp Files et 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 and RAPIDS :

- DASK

["https://docs.dask.org/en/latest/"](https://docs.dask.org/en/latest/)

- Installer DASK

["https://docs.dask.org/en/latest/install.html"](https://docs.dask.org/en/latest/install.html)

- API DASK

["https://docs.dask.org/en/latest/api.html"](https://docs.dask.org/en/latest/api.html)

- DASK machine Learning

["https://examples.dask.org/machine-learning.html"](https://examples.dask.org/machine-learning.html)

- Diagnostic distribué DASK

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

- Structure ET outils ML :

- TensorFlow : un framework d'apprentissage machine open source pour tous

["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/)

Informations sur le copyright

Copyright © 2024 NetApp, Inc. Tous droits réservés. Imprimé aux États-Unis. Aucune partie de ce document protégé par copyright ne peut être reproduite sous quelque forme que ce soit ou selon quelque méthode que ce soit (graphique, électronique ou mécanique, notamment par photocopie, enregistrement ou stockage dans un système de récupération électronique) sans l'autorisation écrite préalable du détenteur du droit de copyright.

Les logiciels dérivés des éléments NetApp protégés par copyright sont soumis à la licence et à l'avis de non-responsabilité suivants :

CE LOGICIEL EST FOURNI PAR NETAPP « EN L'ÉTAT » ET SANS GARANTIES EXPRESSES OU TACITES, Y COMPRIS LES GARANTIES TACITES DE QUALITÉ MARCHANDE ET D'ADÉQUATION À UN USAGE PARTICULIER, QUI SONT EXCLUES PAR LES PRÉSENTES. EN AUCUN CAS NETAPP NE SERA TENU POUR RESPONSABLE DE DOMMAGES DIRECTS, INDIRECTS, ACCESSOIRES, PARTICULIERS OU EXEMPLAIRES (Y COMPRIS L'ACHAT DE BIENS ET DE SERVICES DE SUBSTITUTION, LA PERTE DE JOUISSANCE, DE DONNÉES OU DE PROFITS, OU L'INTERRUPTION D'ACTIVITÉ), QUELLES QU'EN SOIENT LA CAUSE ET LA DOCTRINE DE RESPONSABILITÉ, QU'IL S'AGISSE DE RESPONSABILITÉ CONTRACTUELLE, STRICTE OU DÉLICTELLE (Y COMPRIS LA NÉGLIGENCE OU AUTRE) DÉCOULANT DE L'UTILISATION DE CE LOGICIEL, MÊME SI LA SOCIÉTÉ A ÉTÉ INFORMÉE DE LA POSSIBILITÉ DE TELS DOMMAGES.

NetApp se réserve le droit de modifier les produits décrits dans le présent document à tout moment et sans préavis. NetApp décline toute responsabilité découlant de l'utilisation des produits décrits dans le présent document, sauf accord explicite écrit de NetApp. L'utilisation ou l'achat de ce produit ne concède pas de licence dans le cadre de droits de brevet, de droits de marque commerciale ou de tout autre droit de propriété intellectuelle de NetApp.

Le produit décrit dans ce manuel peut être protégé par un ou plusieurs brevets américains, étrangers ou par une demande en attente.

LÉGENDE DE RESTRICTION DES DROITS : L'utilisation, la duplication ou la divulgation par le gouvernement sont sujettes aux restrictions énoncées dans le sous-paragraphe (b)(3) de la clause Rights in Technical Data-Noncommercial Items du DFARS 252.227-7013 (février 2014) et du FAR 52.227-19 (décembre 2007).

Les données contenues dans les présentes se rapportent à un produit et/ou service commercial (tel que défini par la clause FAR 2.101). Il s'agit de données propriétaires de NetApp, Inc. Toutes les données techniques et tous les logiciels fournis par NetApp en vertu du présent Accord sont à caractère commercial et ont été exclusivement développés à l'aide de fonds privés. Le gouvernement des États-Unis dispose d'une licence limitée irrévocable, non exclusive, non cessible, non transférable et mondiale. Cette licence lui permet d'utiliser uniquement les données relatives au contrat du gouvernement des États-Unis d'après lequel les données lui ont été fournies ou celles qui sont nécessaires à son exécution. Sauf dispositions contraires énoncées dans les présentes, l'utilisation, la divulgation, la reproduction, la modification, l'exécution, l'affichage des données sont interdits sans avoir obtenu le consentement écrit préalable de NetApp, Inc. Les droits de licences du Département de la Défense du gouvernement des États-Unis se limitent aux droits identifiés par la clause 252.227-7015(b) du DFARS (février 2014).

Informations sur les marques commerciales

NETAPP, le logo NETAPP et les marques citées sur le site <http://www.netapp.com/TM> sont des marques déposées ou des marques commerciales de NetApp, Inc. Les autres noms de marques et de produits sont des marques commerciales de leurs propriétaires respectifs.