Distributed Azure training - Lane detection

NetApp Solutions

NetApp
July 24, 2024
# Table of Contents

TR-4896: Distributed training in Azure: Lane detection - Solution design ........................................ 1
   Solution overview ......................................................................................................................... 1
   Lane detection – Distributed training with RUN:AI ................................................................. 3
   Conclusion ....................................................................................................................................... 29
   Additional information ............................................................................................................... 29
Since May 2019, Microsoft delivers an Azure native, first-party portal service for enterprise NFS and SMB file services based on NetApp ONTAP technology. This development is driven by a strategic partnership between Microsoft and NetApp and further extends the reach of world-class ONTAP data services to Azure.

NetApp, a leading cloud data services provider, has teamed up with RUN: AI, a company virtualizing AI infrastructure, to allow faster AI experimentation with full GPU utilization. The partnership enables teams to speed up AI by running many experiments in parallel, with fast access to data, and leveraging limitless compute resources. RUN: AI enables full GPU utilization by automating resource allocation, and the proven architecture of Azure NetApp Files enables every experiment to run at maximum speed by eliminating data pipeline obstructions.

NetApp and RUN: AI have joined forces to offer customers a future-proof platform for their AI journey in Azure. From analytics and high-performance computing (HPC) to autonomous decisions (where customers can optimize their IT investments by only paying for what they need, when they need it), the alliance between NetApp and RUN: AI offers a single unified experience in the Azure Cloud.

Solution overview

In this architecture, the focus is on the most computationally intensive part of the AI or machine learning (ML) distributed training process of lane detection. Lane detection is one of the most important tasks in autonomous driving, which helps to guide vehicles by localization of the lane markings. Static components like lane markings guide the vehicle to drive on the highway interactively and safely.

Convolutional Neural Network (CNN)-based approaches have pushed scene understanding and segmentation to a new level. Although it doesn’t perform well for objects with long structures and regions that could be occluded (for example, poles, shade on the lane, and so on). Spatial Convolutional Neural Network (SCNN) generalizes the CNN to a rich spatial level. It allows information propagation between neurons in the same layer, which makes it best suited for structured objects such as lanes, poles, or truck with occlusions. This compatibility is because the spatial information can be reinforced, and it preserves smoothness and continuity.

Thousands of scene images need to be injected in the system to allow the model learn and distinguish the various components in the dataset. These images include weather, daytime or nighttime, multilane highway roads, and other traffic conditions.

For training, there is a need for good quality and quantity of data. Single GPU or multiple GPUs can take days to weeks to complete the training. Data-distributed training can speed up the process by using multiple and multinode GPUs. Horovod is one such framework that grants distributed training but reading data across clusters of GPUs could act as a hindrance. Azure NetApp Files provides ultrafast, high throughput and sustained low latency to provide scale-out/scale-up capabilities so that GPUs are leveraged to the best of their computational capacity. Our experiments verified that all the GPUs across the cluster are used more than 96% on average for training the lane detection using SCNN.
Target audience

Data science incorporates multiple disciplines in IT and business, therefore multiple personas are part of our targeted audience:

• Data scientists need the flexibility to use the tools and libraries of their choice.
• Data engineers need to know how the data flows and where it resides.
• Autonomous driving use-case experts.
• Cloud administrators and architects to set up and manage cloud (Azure) resources.
• A DevOps engineer needs the tools to integrate new AI/ML applications into their continuous integration and continuous deployment (CI/CD) pipelines.
• Business users want to have access to AI/ML applications.

In this document, we describe how Azure NetApp Files, RUN: AI, and Microsoft Azure help each of these roles bring value to business.

Solution technology

This section covers the technology requirements for the lane detection use case by implementing a distributed training solution at scale that fully runs in the Azure cloud. The figure below provides an overview of the solution architecture.

The elements used in this solution are:

• Azure Kubernetes Service (AKS)
• Azure Compute SKUs with NVIDIA GPUs
• Azure NetApp Files
• RUN: AI
• NetApp Trident

Links to all the elements mentioned here are listed in the Additional information section.
Cloud resources and services requirements

The following table lists the hardware components that are required to implement the solution. The cloud components that are used in any implementation of the solution might vary based on customer requirements.

<table>
<thead>
<tr>
<th>Cloud</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKS</td>
<td>Minimum of three system nodes and three GPU worker nodes</td>
</tr>
<tr>
<td>Virtual machine (VM) SKU system nodes</td>
<td>Three Standard_DS2_v2</td>
</tr>
<tr>
<td>VM SKU GPU worker nodes</td>
<td>Three Standard_NC6s_v3</td>
</tr>
<tr>
<td>Azure NetApp Files</td>
<td>4TB standard tier</td>
</tr>
</tbody>
</table>

Software requirements

The following table lists the software components that are required to implement the solution. The software components that are used in any implementation of the solution might vary based on customer requirements.

<table>
<thead>
<tr>
<th>Software</th>
<th>Version or other information</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKS - Kubernetes version</td>
<td>1.18.14</td>
</tr>
<tr>
<td>RUN:AI CLI</td>
<td>v2.2.25</td>
</tr>
<tr>
<td>RUN:AI Orchestration Kubernetes Operator version</td>
<td>1.0.109</td>
</tr>
<tr>
<td>Horovod</td>
<td>0.21.2</td>
</tr>
<tr>
<td>NetApp Trident</td>
<td>20.01.1</td>
</tr>
<tr>
<td>Helm</td>
<td>3.0.0</td>
</tr>
</tbody>
</table>

Lane detection – Distributed training with RUN:AI

This section provides details on setting up the platform for performing lane detection distributed training at scale using the RUN: AI orchestrator. We discuss installation of all the solution elements and running the distributed training job on the said platform. ML versioning is completed by using NetApp SnapshotTM linked with RUN: AI experiments for achieving data and model reproducibility. ML versioning plays a crucial role in tracking models, sharing work between team members, reproducibility of results, rolling new model versions to production, and data provenance. NetApp ML version control (Snapshot) can capture point-in-time versions of the data, trained models, and logs associated with each experiment. It has rich API support making it easy to integrate with the RUN: AI platform; you just have to trigger an event based on the training state. You also have to capture the state of the whole experiment without changing anything in the code or the containers running on top of Kubernetes (K8s).

Finally, this technical report wraps up with performance evaluation on multiple GPU-enabled nodes across AKS.
Distributed training for lane detection use case using the TuSimple dataset

In this technical report, distributed training is performed on the TuSimple dataset for lane detection. Horovod is used in the training code for conducting data distributed training on multiple GPU nodes simultaneously in the Kubernetes cluster through AKS. Code is packaged as container images for TuSimple data download and processing. Processed data is stored on persistent volumes allocated by NetApp Trident plug-in. For the training, one more container image is created, and it uses the data stored on persistent volumes created during downloading the data.

To submit the data and training job, use RUN: AI for orchestrating the resource allocation and management. RUN: AI allows you to perform Message Passing Interface (MPI) operations which are needed for Horovod. This layout allows multiple GPU nodes to communicate with each other for updating the training weights after every training mini batch. It also enables monitoring of training through the UI and CLI, making it easy to monitor the progress of experiments.

NetApp Snapshot is integrated within the training code and captures the state of data and the trained model for every experiment. This capability enables you to track the version of data and code used, and the associated trained model generated.

AKS setup and installation

For setup and installation of the AKS cluster go to Create an AKS Cluster. Then, follow these series of steps:

1. When selecting the type of nodes (whether it be system (CPU) or worker (GPU) nodes), select the following:
   a. Add primary system node named agentpool at the Standard_DS2_v2 size. Use the default three nodes.
   b. Add worker node gpupool with the Standard_NC6s_v3 pool size. Use three nodes minimum for GPU nodes.

2. After deployment is complete, click Connect to Cluster. To connect to the newly created AKS cluster, install the Kubernetes command-line tool from your local environment (laptop/PC). Visit Install Tools to install it as per your OS.

3. Install Azure CLI on your local environment.

4. To access the AKS cluster from the terminal, first enter az login and put in the credentials.

5. Run the following two commands:
az account set --subscription xxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx
aks get-credentials --resource-group resourcegroup --name aksclustername

6. Enter this command in the Azure CLI:

kubectl get nodes

If all six nodes are up and running as seen here, your AKS cluster is ready and connected to your local environment.

Create a delegated subnet for Azure NetApp Files

To create a delegated subnet for Azure NetApp Files, follow this series of steps:

1. Navigate to Virtual networks within the Azure portal. Find your newly created virtual network. It should have a prefix such as aks-vnet, as seen here. Click the name of the virtual network.

2. Click Subnets and select +Subnet from the top toolbar.
3. Provide the subnet with a name such as `ANF.sn` and under the Subnet Delegation heading, select `Microsoft.NetApp/volumes`. Do not change anything else. Click OK.
Azure NetApp Files volumes are allocated to the application cluster and are consumed as persistent volume claims (PVCs) in Kubernetes. In turn, this allocation provides us the flexibility to map volumes to different services, be it Jupyter notebooks, serverless functions, and so on.

Users of services can consume storage from the platform in many ways. The main benefits of Azure NetApp Files are:

- Provides users with the ability to use snapshots.
- Enables users to store large quantities of data on Azure NetApp Files volumes.
- Procure the performance benefits of Azure NetApp Files volumes when running their models on large sets of files.
Azure NetApp Files setup

To complete the setup of Azure NetApp Files, you must first configure it as described in Quickstart: Set up Azure NetApp Files and create an NFS volume.

However, you may omit the steps to create an NFS volume for Azure NetApp Files as you will create volumes through Trident. Before continuing, be sure that you have:

2. Created an account in Azure NetApp Files.
3. Set up a capacity pool (minimum 4TiB Standard or Premium depending on your needs).

Peering of AKS virtual network and Azure NetApp Files virtual network

Next, peer the AKS virtual network (VNet) with the Azure NetApp Files VNet by following these steps:

1. In the search box at the top of the Azure portal, type virtual networks.
2. Click VNet aks- vnet-name, then enter Peerings in the search field.
3. Click +Add and enter the information provided in the table below:

<table>
<thead>
<tr>
<th>Field</th>
<th>Value or description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peering link name</td>
<td>aks-vnet-name_to_anf</td>
</tr>
<tr>
<td>SubscriptionID</td>
<td>Subscription of the Azure NetApp Files VNet to which you’re peering</td>
</tr>
<tr>
<td>VNet peering partner</td>
<td>Azure NetApp Files VNet</td>
</tr>
</tbody>
</table>

Leave all the nonasterisk sections on default

4. Click ADD or OK to add the peering to the virtual network.

For more information, visit Create, change, or delete a virtual network peering.

Trident

Trident is an open-source project that NetApp maintains for application container persistent storage. Trident has been implemented as an external provisioner controller that runs as a pod itself, monitoring volumes and completely automating the provisioning process.

NetApp Trident enables smooth integration with K8s by creating and attaching persistent volumes for storing training datasets and trained models. This capability makes it easier for data scientists and data engineers to use K8s without the hassle of manually storing and managing datasets. Trident also eliminates the need for data scientists to learn managing new data platforms as it integrates the data management-related tasks through the logical API integration.

Install Trident

To install Trident software, complete the following steps:
1. **First install helm.**

2. Download and extract the Trident 21.01.1 installer.

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

3. Change the directory to `trident-installer`.

   ```
   cd trident-installer
   ```

4. Copy `tridentctl` to a directory in your system `$PATH`.

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

5. Install Trident on K8s cluster with Helm:
   a. Change directory to helm directory.
      ```
      cd helm
      ```
   b. Install Trident.
      ```
      helm install trident trident-operator-21.01.1.tgz --namespace trident --create-namespace
      ```
   c. Check the status of Trident pods the usual K8s way:
      ```
      kubectl -n trident get pods
      ```
   d. If all the pods are up and running, Trident is installed and you are good to move forward.

**Set up Azure NetApp Files back-end and storage class**

To set up Azure NetApp Files back-end and storage class, complete the following steps:

1. Switch back to the home directory.

   ```
   cd ~
   ```
2. Clone the project repository `lane-detection-SCNN-horovod`.

3. Go to the `trident-config` directory.

   ```bash
   cd ./lane-detection-SCNN-horovod/trident-config
   ```

4. Create an Azure Service Principle (the service principle is how Trident communicates with Azure to access your Azure NetApp Files resources).

   ```bash
   az ad sp create-for-rbac --name
   ```

   The output should look like the following example:

   ```json
   {
     "appId": "xxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
     "displayName": "netapptrident",
     "name": "http://netapptrident",
     "password": "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx",
     "tenant": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxx"
   }
   ```

5. Create the Trident backend json file.

6. Using your preferred text editor, complete the following fields from the table below inside the `anf-backend.json` file.

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>subscriptionID</td>
<td>Your Azure Subscription ID</td>
</tr>
<tr>
<td>tenantID</td>
<td>Your Azure Tenant ID (from the output of <code>az ad sp</code> in the previous step)</td>
</tr>
<tr>
<td>clientId</td>
<td>Your appID (from the output of <code>az ad sp</code> in the previous step)</td>
</tr>
<tr>
<td>clientSecret</td>
<td>Your password (from the output of <code>az ad sp</code> in the previous step)</td>
</tr>
</tbody>
</table>

   The file should look like the following example:
7. Instruct Trident to create the Azure NetApp Files back-end in the trident namespace, using anf-backend.json as the configuration file as follows:

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

8. Create the storage class:

   a. K8 users provision volumes by using PVCs that specify a storage class by name. Instruct K8s to create a storage class azurenetappfiles that will reference the Azure NetApp Files back end created in the previous step using the following:

```bash
kubectl create -f anf-storage-class.yaml
```

   b. Check that storage class is created by using the following command:

```bash
kubectl get sc azurenetappfiles
```

The output should look like the following example:

<table>
<thead>
<tr>
<th>NAME</th>
<th>PROVISIONER</th>
<th>RECLAIMPOLICY</th>
<th>VOLUMEBINDINGMODE</th>
<th>ALLOWVOLUMEEXPANSION</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>azurenetappfiles</td>
<td>csi.trident.netapp.io</td>
<td>Delete</td>
<td>Immediate</td>
<td>false</td>
<td>98s</td>
</tr>
</tbody>
</table>

**Deploy and set up volume snapshot components on AKS**

If your cluster does not come pre-installed with the correct volume snapshot components, you may manually install these components by running the following steps:
AKS 1.18.14 does not have pre-installed Snapshot Controller.

1. Install Snapshot Beta CRDs by using the following commands:

```bash
ekubectl create -f https://raw.githubusercontent.com/kubernetes-csi/external-snapshotter/release-3.0/client/config/crd/snapshot.storage.k8s.io_volumesnapshotclasses.yaml
kubectl create -f https://raw.githubusercontent.com/kubernetes-csi/external-snapshotter/release-3.0/client/config/crd/snapshot.storage.k8s.io_volumesnapshotcontents.yaml
kubectl create -f https://raw.githubusercontent.com/kubernetes-csi/external-snapshotter/release-3.0/client/config/crd/snapshot.storage.k8s.io_volumesnapshots.yaml
```

2. Install Snapshot Controller by using the following documents from GitHub:

```bash
```

3. Set up K8s `volumesnapshotclass`: Before creating a volume snapshot, a `volume snapshot class` must be set up. Create a volume snapshot class for Azure NetApp Files, and use it to achieve ML versioning by using NetApp Snapshot technology. Create `volumesnapshotclass netapp-csi-snapclass` and set it to default `volumesnapshotclass` as such:

```bash
ekubectl create -f netapp-volume-snapshot-class.yaml
```

The output should look like the following example:

```
volumesnapshotclass.snapshot.storage.k8s.io/netapp-csi-snapclass created
```

4. Check that the volume Snapshot copy class was created by using the following command:

```bash
ekubectl get volumesnapshotclass
```

The output should look like the following example:

```
NAME       DRIVER                      DELETIONPOLICY AGE
netapp-csi-snapclass  csi.trident.netapp.io Delete 63s
```
**RUN:AI installation**

To install RUN:AI, complete the following steps:

1. **Install RUN:AI cluster on AKS.**

2. Go to app.runai.ai, click create New Project, and name it lane-detection. It will create a namespace on a K8s cluster starting with runai- followed by the project name. In this case, the namespace created would be runai-lane-detection.

3. **Install RUN:AI CLI.**

4. On your terminal, set lane-detection as a default RUN: AI project by using the following command:

   ```bash
   runai config project lane-detection
   ```

   The output should look like the following example:

   ```bash
   Project lane-detection has been set as default project
   ```
5. Create ClusterRole and ClusterRoleBinding for the project namespace (for example, lane-detection) so the default service account belonging to runai-lane-detection namespace has permission to perform volumesnapshot operations during job execution:

a. List namespaces to check that runai-lane-detection exists by using this command:

```
kubectl get namespaces
```

The output should appear like the following example:

```
NAME             STATUS AGE
default          Active 130m
kube-node-lease  Active 130m
kube-public      Active 130m
kube-system      Active 130m
runai            Active 4m44s
runai-lane-detection Active 13s
trident          Active 102m
```

6. Create ClusterRole netappsnapshot and ClusterRoleBinding netappsnapshot using the following commands:

```
kubectl create -f runai-project-snap-role.yaml
kkubectl create -f runai-project-snap-role-binding.yaml
```

Download and process the TuSimple dataset as RUN:AI job

The process to download and process the TuSimple dataset as a RUN: AI job is optional. It involves the following steps:

1. Build and push the docker image, or omit this step if you want to use an existing docker image (for example, muneer7589/download-tusimple:1.0)
   a. Switch to the home directory:

   ```
   cd ~
   ```

   b. Go to the data directory of the project lane-detection-SCNN-horovod:

   ```
   cd ./lane-detection-SCNN-horovod/data
   ```

   c. Modify build_image.sh shell script and change docker repository to yours. For example, replace muneer7589 with your docker repository name. You could also change the docker image name and
d. Run the script to build the docker image and push it to the docker repository using these commands:

```
chmod +x build_image.sh
./build_image.sh
```

2. Submit the RUN: AI job to download, extract, pre-process, and store the TuSimple lane detection dataset in a `pvc`, which is dynamically created by NetApp Trident:

a. Use the following commands to submit the RUN: AI job:

```
runai submit
--name download-tusimple-data
--pvc azurenetappfiles:100Gi:/mnt
--image muneer7589/download-tusimple:1.0
```

b. Enter the information from the table below to submit the RUN:AI job:

<table>
<thead>
<tr>
<th>Field</th>
<th>Value or description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-name</td>
<td>Name of the job</td>
</tr>
<tr>
<td>Field</td>
<td>Value or description</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------</td>
</tr>
<tr>
<td>PVC</td>
<td>PVC of the format [StorageClassName]:Size:ContainerMountPath. In the above job submission, you are creating an PVC based on-demand using Trident with storage class azurenetappfiles. Persistent volume capacity here is 100Gi and it's mounted at path /mnt.</td>
</tr>
<tr>
<td>Image</td>
<td>Docker image to use when creating the container for this job.</td>
</tr>
</tbody>
</table>

The output should look like the following example:

```
The job 'download-tusimple-data' has been submitted successfully.
You can run `runai describe job download-tusimple-data -p lane-detection` to check the job status.
```

c. List the submitted RUN:AI jobs.

```
runai list jobs
```

d. Check the submitted job logs.

```
runai logs download-tusimple-data -t 10
```

e. List the PVC created. Use this PVC command for training in the next step.

```
kubectl get pvc | grep download-tusimple-data
```

The output should look like the following example:

```
pvc-downloadd-tusimple-data-8 Bound pvc-b863b7ac-2c17-4eb4-a44d-79f3de4f5d65 1000i RWO azurenetappfiles 4M47s
```

f. Check the job in RUN: AI UI (or app.run.ai).
Perform distributed lane detection training using Horovod

Performing distributed lane detection training using Horovod is an optional process. However, here are the steps involved:

1. Build and push the docker image, or skip this step if you want to use the existing docker image (for example, muneer7589/dist-lane-detection:3.1):
   a. Switch to home directory.

   ```bash
   cd ~
   ```
   b. Go to the project directory lane-detection-SCNN-horovod.

   ```bash
   cd ./lane-detection-SCNN-horovod
   ```
   c. Modify the build_image.sh shell script and change docker repository to yours (for example, replace muneer7589 with your docker repository name). You could also change the docker image name and TAG (dist-lane-detection and 3.1, for example).
d. Run the script to build the docker image and push to the docker repository.

```
chmod +x build_image.sh
./build_image.sh
```

2. Submit the RUN: AI job for carrying out distributed training (MPI):
   
a. Using submit of RUN: AI for automatically creating PVC in the previous step (for downloading data) only allows you to have RWO access, which does not allow multiple pods or nodes to access the same PVC for distributed training. Update the access mode to ReadWriteMany and use the Kubernetes patch to do so.
   
b. First, get the volume name of the PVC by running the following command:

```
kubectl get pvc | grep download-tusimple-data
```

   c. Patch the volume and update access mode to ReadWriteMany (replace volume name with yours in the following command):

```
kubectl patch pv pvc-bb03b74d-2c17-40c4-a445-79f3de8d16d5 -p '{"spec":{"accessModes":["ReadWriteMany"]}}'
```

   d. Submit the RUN: AI MPI job for executing the distributed training job using information from the table below:
runai submit-mpi
   --name dist-lane-detection-training
   --large-shm
   --processes=3
   --gpu 1
   --pvc pvc-download-tusimple-data-0:/mnt
   --image muneer7589/dist-lane-detection:3.1
   -e USE_WORKERS="true"
   -e NUM_WORKERS=4
   -e BATCH_SIZE=33
   -e USE_VAL="false"
   -e VAL_BATCH_SIZE=99
   -e ENABLE_SNAPSHOT="true"
   -e PVC_NAME="pvc-download-tusimple-data-0"

<table>
<thead>
<tr>
<th>Field</th>
<th>Value or description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Name of the distributed training job</td>
</tr>
<tr>
<td>large shm</td>
<td>Mount a large /dev/shm device</td>
</tr>
<tr>
<td></td>
<td>It is a shared file system mounted on RAM and provides large enough shared memory</td>
</tr>
<tr>
<td></td>
<td>for multiple CPU workers to process and load batches into CPU RAM.</td>
</tr>
<tr>
<td>processes</td>
<td>Number of distributed training processes</td>
</tr>
<tr>
<td>gpu</td>
<td>Number of GPUs/processes to allocate for the job</td>
</tr>
<tr>
<td></td>
<td>In this job, there are three GPU worker processes (--processes=3), each allocated</td>
</tr>
<tr>
<td></td>
<td>with a single GPU (--gpu 1)</td>
</tr>
<tr>
<td>pvc</td>
<td>Use existing persistent volume (pvc-download-tusimple-data-0) created by previous</td>
</tr>
<tr>
<td></td>
<td>job (download-tusimple-data) and it is mounted at path /mnt</td>
</tr>
<tr>
<td>image</td>
<td>Docker image to use when creating the container for this job</td>
</tr>
<tr>
<td>Define environment variables to be set in the container</td>
<td></td>
</tr>
<tr>
<td>USE_WORKERS</td>
<td>Setting the argument to true turns on multi-process data loading</td>
</tr>
<tr>
<td>NUM_WORKERS</td>
<td>Number of data loader worker processes</td>
</tr>
<tr>
<td>BATCH_SIZE</td>
<td>Training batch size</td>
</tr>
<tr>
<td>USE_VAL</td>
<td>Setting the argument to true allows validation</td>
</tr>
<tr>
<td>VAL_BATCH_SIZE</td>
<td>Validation batch size</td>
</tr>
<tr>
<td>Field</td>
<td>Value or description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ENABLE_SNAPSHOT</td>
<td>Setting the argument to true enables taking data and trained model snapshots for ML versioning purposes</td>
</tr>
<tr>
<td>PVC_NAME</td>
<td>Name of the pvc to take a snapshot of. In the above job submission, you are taking a snapshot of pvc-download-tusimple-data-0, consisting of dataset and trained models</td>
</tr>
</tbody>
</table>

The output should look like the following example:

```
The job 'dist-lane-detection-training' has been submitted successfully
You can run runai describe job dist-lane-detection-training -p lane-detection' to check the job status
```

e. List the submitted job.

```
runai list jobs
```

<table>
<thead>
<tr>
<th>SERVICE URL(S)</th>
<th>STATUS</th>
<th>AGE</th>
<th>NODE</th>
<th>IMAGE</th>
<th>TYPE</th>
<th>PROJECT</th>
<th>USER</th>
<th>GPUs Allocated (Requested)</th>
<th>PODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>download-tusimple-data</td>
<td>Succeeded</td>
<td>1d</td>
<td>muenner7589/download-tusimple:1.0</td>
<td>Train lane-detection</td>
<td>verrornarting</td>
<td>- (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dist-lane-detection-training</td>
<td>Init:0/1</td>
<td>2m</td>
<td><em>multiple</em></td>
<td>muenner7589/dist-lane-detection:3.1</td>
<td>Train lane-detection</td>
<td>root</td>
<td>3 (3)</td>
<td>4 (4)</td>
<td></td>
</tr>
</tbody>
</table>

f. Submitted job logs:

```
runai logs dist-lane-detection-training
```

```
root@ai - wgpu-2:~/runai# runai logs dist-lane-detection-training
Running with 3 workers
2021-03-04 17:29:23.158449: I tensorflow/stream_executor/platform/default/dso_loader.cc:148] Successfully opened dynamic library libcudart.so.10.3
+ [ 0 --]
+ shift
+ /opt/kube/kubectl cp /opt/kube/hosts dist-lane-detection-training-worker-0:/etc/hosts_of_nodes
+ POD_NAME=dist-lane-detection-training-worker-0
+ [ 0 --]
+ shift
+ /opt/kube/kubectl cp /opt/kube/hosts dist-lane-detection-training-worker-2:/etc/hosts_of_nodes
+ POD_NAME=dist-lane-detection-training-worker-2
+ [ 0 --]
+ shift
+ /opt/kube/kubectl cp /opt/kube/hosts dist-lane-detection-training-worker-3:/etc/hosts_of_nodes
+ POD_NAME=dist-lane-detection-training-worker-3
```

g. Check training job in RUN: AI GUI (or app.runai.ai): RUN: AI Dashboard, as seen in the figures below. The first figure details three GPUs allocated for the distributed training job spread across three nodes on AKS, and the second RUN:AI jobs:
h. After the training is finished, check the NetApp Snapshot copy that was created and linked with RUN: AI job.

```
runai logs dist-lane-detection-training --tail 1
```

```
kubectl get volumesnapshots | grep download-tusimple-data-0
```
Restore data from the NetApp Snapshot copy

To restore data from the NetApp Snapshot copy, complete the following steps:

1. Switch to home directory.
   
   ```
   cd ~
   ```

2. Go to the project directory `lane-detection-SCNN-horovod`.
   
   ```
   cd ./lane-detection-SCNN-horovod
   ```

3. Modify `restore-snapshot-pvc.yaml` and update `dataSource` name field to the Snapshot copy from which you want to restore data. You could also change PVC name where the data will be restored to, in this example its `restored-tusimple`.

   ```yaml
   apiVersion: v1
   kind: PersistentVolumeClaim
   metadata:
     name: restored-tusimple
   spec:
     storageClassName: azurenetappfiles
     dataSource:
       kind: VolumeSnapshot
       apiGroup: snapshot.storage.k8s.io
     accessModes:
     - ReadWriteMany
     resources:
     requests:
     storage: 100Gi
   ```

4. Create a new PVC by using `restore-snapshot-pvc.yaml`.
   
   ```
   kubectl create -f restore-snapshot-pvc.yaml
   ```

   The output should look like the following example:

   ```
   persistentvolumeclaim/restored-tusimple created
   ```

5. If you want to use the just restored data for training, job submission remains the same as before; only replace the `PVC_NAME` with the restored `PVC_NAME` when submitting the training job, as seen in the following commands:
Performance evaluation

To show the linear scalability of the solution, performance tests have been done for two scenarios: one GPU and three GPUs. GPU allocation, GPU and memory utilization, different single- and three-node metrics have been captured during the training on the TuSimple lane detection dataset. Data is increased five-fold just for the sake of analyzing resource utilization during the training processes.

The solution enables customers to start with a small dataset and a few GPUs. When the amount of data and the demand of GPUs increase, customers can dynamically scale out the terabytes in the Standard Tier and quickly scale up to the Premium Tier to get four times the throughput per terabyte without moving any data. This process is further explained in the section, Azure NetApp Files service levels.

Processing time on one GPU was 12 hours and 45 minutes. Processing time on three GPUs across three nodes was approximately 4 hours and 30 minutes.

The figures shown throughout the remainder of this document illustrate examples of performance and scalability based on individual business needs.

The figure below illustrates 1 GPU allocation and memory utilization.
The figure below illustrates single node GPU utilization.

The figure below illustrates single node memory size (16GB).
The figure below illustrates single node GPU count (1).

The figure below illustrates single node GPU allocation (%).

The figure below illustrates three GPUs across three nodes – GPUs allocation and memory.
The figure below illustrates three GPUs across three nodes utilization (%).

The figure below illustrates three GPUs across three nodes memory utilization (%).
Azure NetApp Files service levels

You can change the service level of an existing volume by moving the volume to another capacity pool that uses the service level you want for the volume. This existing service-level change for the volume does not require that you migrate data. It also does not affect access to the volume.

Dynamically change the service level of a volume

To change the service level of a volume, use the following steps:

1. On the Volumes page, right-click the volume whose service level you want to change. Select Change Pool.

2. In the Change Pool window, select the capacity pool you want to move the volume to. Then, click OK.
Automate service level change

Dynamic Service Level change is currently still in Public Preview, but it is not enabled by default. To enable this feature on the Azure subscription, follow these steps provided in the document “Dynamically change the service level of a volume.”

• You can also use the following commands for Azure: CLI. For more information about changing the pool size of Azure NetApp Files, visit az netappfiles volume: Manage Azure NetApp Files (ANF) volume resources.

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

• The set-aznetappfilesvolume pool cmdlet shown here can change the pool of an Azure NetApp Files volume. More information about changing volume pool size and Azure PowerShell can be found by visiting Change pool for an Azure NetApp Files volume.
Conclusion

NetApp and RUN: AI have partnered in the creation of this technical report to demonstrate the unique capabilities of the Azure NetApp Files together with the RUN: AI platform for simplifying orchestration of AI workloads. This technical report provides a reference architecture for streamlining the process of both data pipelines and workload orchestration for distributed lane detection training.

In conclusion, with regard to distributed training at scale (especially in a public cloud environment), the resource orchestration and storage component is a critical part of the solution. Making sure that data managing never hinders multiple GPU processing, therefore results in the optimal utilization of GPU cycles. Thus, making the system as cost effective as possible for large-scale distributed training purposes.

Data fabric delivered by NetApp overcomes the challenge by enabling data scientists and data engineers to connect together on-premises and in the cloud to have synchronous data, without performing any manual intervention. In other words, data fabric smooths the process of managing AI workflow spread across multiple locations. It also facilitates on demand-based data availability by bringing data close to compute and performing analysis, training, and validation wherever and whenever needed. This capability not only enables data integration but also protection and security of the entire data pipeline.

Additional information

To learn more about the information that is described in this document, review the following documents and/or websites:

- Dataset: TuSimple
  https://github.com/TuSimple/tusimple-benchmark/tree/master/doc/lane_detection
- Deep Learning Network Architecture: Spatial Convolutional Neural Network
  https://arxiv.org/abs/1712.06080
- Distributed deep learning training framework: Horovod
  https://horovod.ai/
- RUN: AI container orchestration solution: RUN: AI product introduction
  https://docs.run.ai/home/components/
- RUN: AI installation documentation
• Submitting jobs in RUN: AI CLI
  https://docs.run.ai/Researcher/cli-reference/runai-submit/
  https://docs.run.ai/Researcher/cli-reference/runai-submit-mpi/

• Azure Cloud resources: Azure NetApp Files
  https://docs.microsoft.com/azure/azure-netapp-files/

• Azure Kubernetes Service
  https://azure.microsoft.com/services/kubernetes-service/-features

• Azure VM SKUs
  https://azure.microsoft.com/services/virtual-machines/

• Azure VM with GPU SKUs
  https://docs.microsoft.com/azure/virtual-machines/sizes-gpu

• NetApp Trident
  https://github.com/NetApp/trident/releases

• Data Fabric powered by NetApp

• NetApp Product Documentation
  https://www.netapp.com/support-and-training/documentation/