

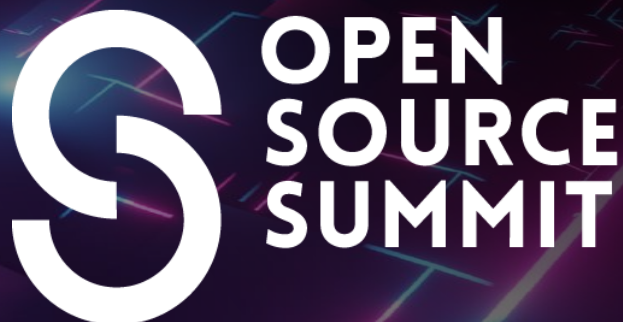


KubeCon



CloudNativeCon

THE LINUX FOUNDATION



AI_dev
Open Source GenAI & ML Summit

China 2024



KubeCon



CloudNativeCon



China 2024

Deep Dive Into Windows CSI Driver HostProcess Containers

Andy Zhang, Principal Software Engineer @Microsoft

Intro

Andy Zhang (张夏)

- Storage Dev lead @AKS(Azure Kubernetes Service) Team
- Maintainer of multiple k8s projects
 - [Azure Disk/File/Blob](#) CSI drivers
 - [SMB, NFS](#) under github.com/kubernetes-csi
 - [Local volume provisioner](#)
 - [csi-proxy](#)



Andy Zhang
andyzhangx

- Github: [andyzhangx](#)
- Email: xiazhang@microsoft.com
- Slack: andyzhang [@kubernetes](#)

Agenda



China 2024

- Introduction
- Background of Windows csi-proxy project
- Introduction of HostProcess container
- Migrate CSI driver on Windows from csi-proxy dependent to HostProcess container deployment
- Key learnings and gotchas
- Migration progress of cloud providers
- Conclusion and plans

- Most CSI drivers rely on csi-proxy process on Windows node
 - csi-proxy performs privileged storage operations on behalf of CSI driver containers
- Windows HostProcess Container
 - Introduced in k8s 1.23 and stable on k8s 1.26
 - run directly on the Windows host as a regular process with privileges
 - Switching to HostProcess container deployment will make Windows CSI driver development and deployment easier

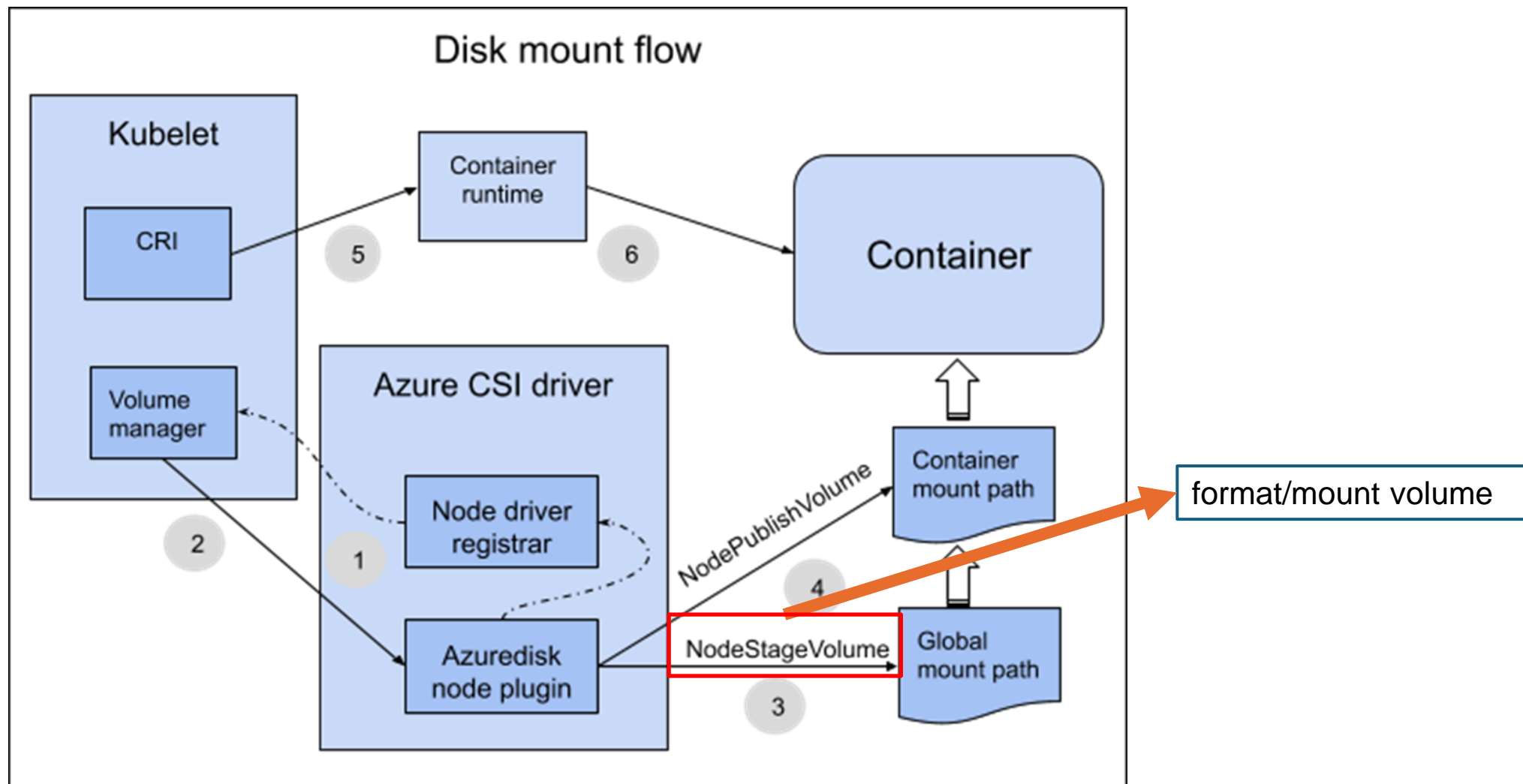
Background of windows csi-proxy project



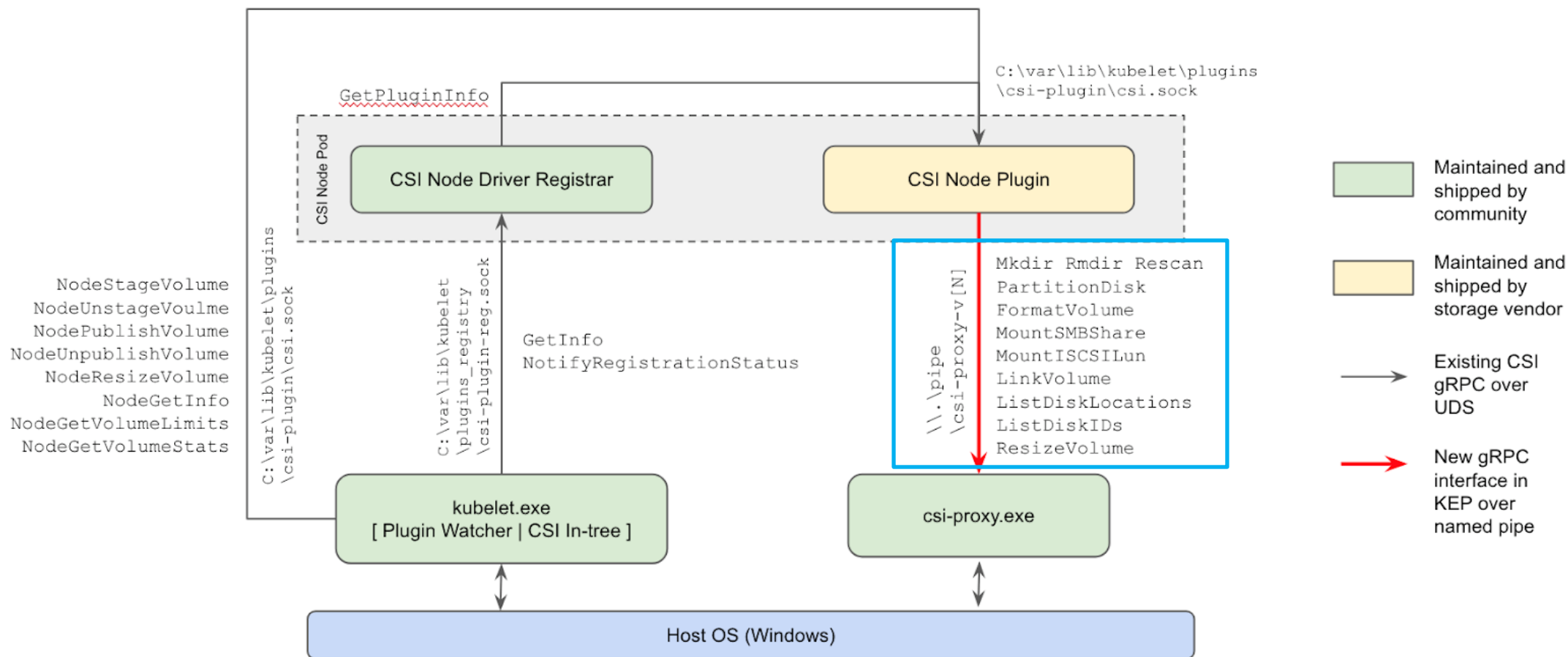
China 2024

- in-tree(built-in) storage driver
 - Kubelet handles privileged operations for in-tree storage drivers on both Linux and Windows nodes
 - e.g. disk format, mount/unmount
- CSI driver
 - introduced in k8s 1.13
 - in-tree driver removal and migrate to CSI driver
 - Node plugin containers require elevated privileges to perform storage operations
 - Prior to Kubernetes v1.23, privileged pods are only supported on Linux nodes and not on Windows nodes

How CSI driver interacts with k8s on Linux node



How CSI driver interacts with k8s on Windows node using csi-proxy



Reference: <https://raw.githubusercontent.com/kubernetes/enhancements/master/keps/sig-windows/csi-proxy3.png>

Introduction of HostProcess container

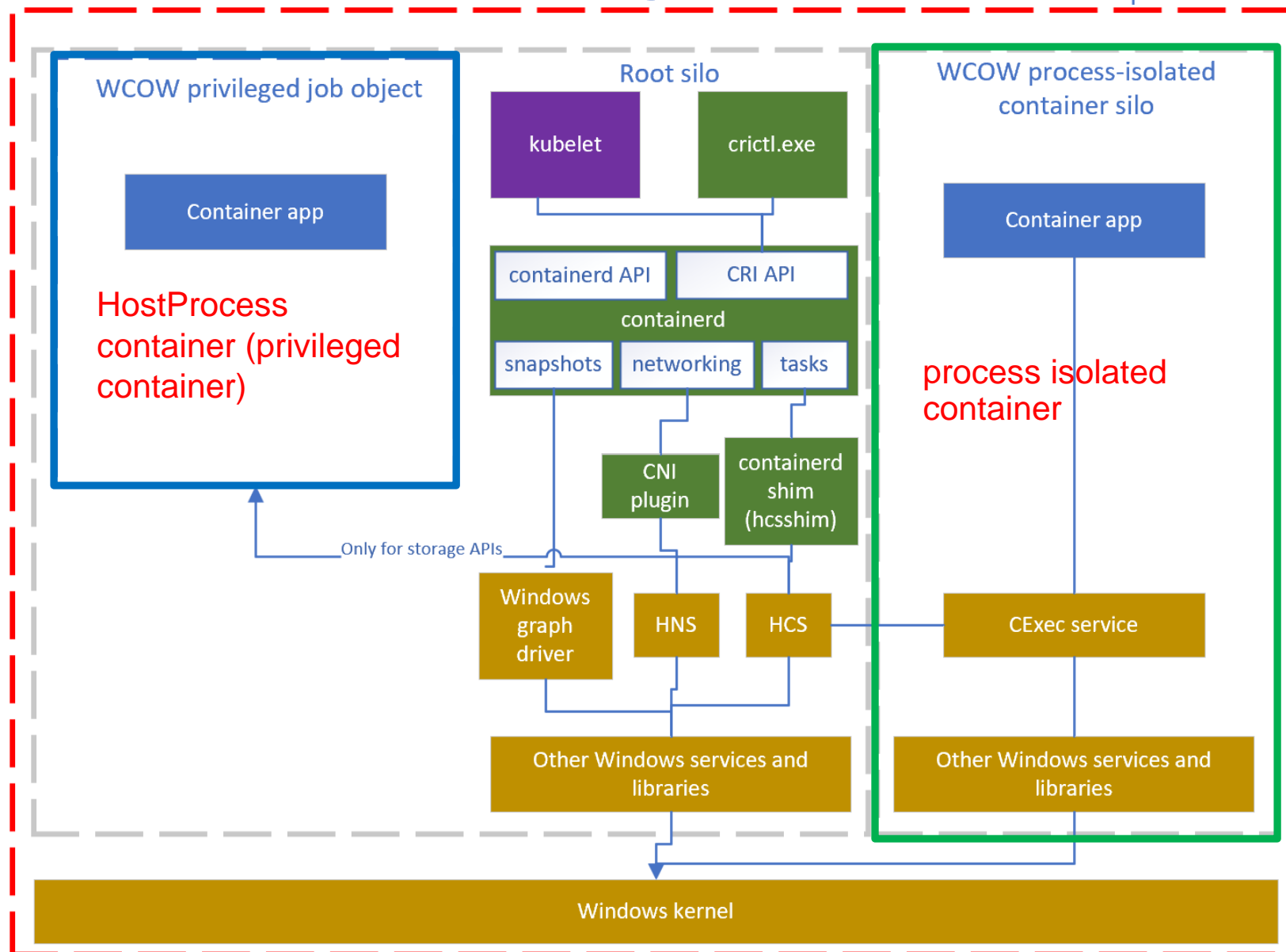


China 2024

- HostProcess container
 - Introduced in k8s v1.23, GA in v1.26
 - Run directly on the host as a regular process with privileges
- Benefits against csi-proxy for CSI driver on Windows node
 - Simplify development and deployment process
 - Simplify troubleshooting
 - Smaller image size

Window Process isolated Containers vs Windows Privileged Containers

Root partition



Benefits of HostProcess against csi-proxy



China 2024

- Development and deployment workflow with csi-proxy
 1. add new API in both csi-proxy server and client
 2. publish new csi-proxy version
 3. adopt the new csi-proxy version in CSI driver
 4. deploy new csi-proxy version to the Windows nodes
 5. upgrade new CSI driver version on all Windows nodes

Pain points:

- in certain Kubernetes clusters, in-place replacement is not supported and requires upgrading the whole node image on all nodes to get csi-proxy version upgrade
- new CSI driver version must be compatible with all recent csi-proxy versions

Benefits of HostProcess against csi-proxy



China 2024

- pain point in troubleshooting with csi-proxy
 - user needs to log on to the Windows node and ensure that csi-proxy service is healthy and then obtain csi-proxy logs as an extra step in troubleshooting

[csi-proxy](#) troubleshooting

- ssh to the windows node(e.g. 512k8s010), and then run Powershell command

```
azureuser@2512k8s010 C:\Users\azureuser>powershell

cd c:\k
ls csi-proxy.exe

PS C:\k> Get-Service csi-proxy

Status  Name      DisplayName
-----
Running csi-proxy  csi-proxy

PS C:\k> cat C:\k\csi-proxy.err.log
I0905 13:22:22.341048 3880 main.go:55] Starting CSI-Proxy Server ...
I0905 13:22:22.398066 3880 main.go:56] Version: v0.2.2-0-gffb169f
```

Benefits of HostProcess against csi-proxy



China 2024

- Development and deployment workflow with HostProcess CSI driver
 1. Add new feature in CSI driver (no new API required)
 2. publish new CSI driver version
 3. deploy new CSI driver version on all Windows nodes

status of csi-proxy project



China 2024

- <https://github.com/kubernetes-csi/csi-proxy>

CS Proxy

There's a [KEP](#) about moving CSI Drivers to HostProcess containers which involves making CSI Proxy a Go library, development for this work is happening in the `library-development` branch

CS Proxy (which might be more aptly named "csi-node-proxy") is a binary that exposes a set of gRPC APIs (over named pipes) around local storage operations for nodes in Windows. A container, such as CSI node plugins, can mount the named pipes depending on operations it wants to exercise on the host and invoke the APIs. This allows a storage plugin to run as if were a CSI plugin on linux which have the ability to perform "privileged" actions on the windows host OS.

Limitations of HostProcess container



KubeCon



CloudNativeCon



China 2024



- require Kubernetes 1.23+
- require containerd 1.6+
- a limited set of host user accounts are available for Host Process containers by default
 - Local System account for CSI driver
- Filesystem and Hyper-V isolation aren't supported for HostProcess containers
- HostProcess pods can only contain HostProcess containers

Steps for migrating to HostProcess container deployment



China 2024

- HostProcess container interface change
- HostProcess container deployment change
- Windows Image build for HostProcess container image

Migration – interface change



China 2024

- Code changes in CSI driver interfaces
 - NodeStageVolume
 - NodeUnstageVolume
 - NodePublishVolume
 - NodeUnpublishVolume
 - NodeExpandVolume
 - **NodeGetVolumeStats**

Migration – interface change



China 2024

- Replace csi-proxy calls with direct calls inside driver container

AzureDisk change:

FormatAndMount
ExistsPath
Rmdir
Rescan
FindDiskByLun
GetDeviceNameFromMount
GetVolumeSizeInBytes
ResizeVolume
GetVolumeStats

AzureFile(SMB mount) change:

SMBMount
ExistsPath
MakeDir
Rmdir
IsMountPointMatch
EvalHostSymlinks
GetVolumeStats

Migration – HostProcess container deployment



China 2024

<https://kubernetes.io/docs/tasks/configure-pod-container/create-hostprocess-pod/>

```
kind: DaemonSet
apiVersion: apps/v1
metadata:
  name: csi-azurefile-node-win
spec:
  ...
  securityContext:
    windowsOptions:
      hostProcess: true
      runAsUserName: "NT AUTHORITY\\SYSTEM"
  hostNetwork: true
```

Migration – Windows Image build



China 2024

- <https://github.com/microsoft/windows-host-process-containers-base-image>
- HostProcess container base image
mcr.microsoft.com/oss/kubernetes/windows-host-process-containers-base-image:v1.0.0
- Benefits against nanoserver image
 - Size
 - a few KB vs a few hundred MB
 - OS compatible
 - a single hostprocess container image tag is compatible with Windows Server 2019 and later os version

- Dockerfile example of Windows HostProcss image

```
FROM mcr.microsoft.com/oss/kubernetes/windows-host-process-containers-base-image:v1.0.0
LABEL description="CSI Azure disk plugin"

ARG ARCH=amd64
ARG binary=./_output/${ARCH}/azurediskplugin.exe
COPY ${binary} /azurediskplugin.exe
ENV PATH="C:\Windows\system32;C:\Windows;C:\WINDOWS\System32\WindowsPowerShell\v1.0\;"
USER ContainerAdministrator
ENTRYPOINT ["/azurediskplugin.exe"]
```

Key learnings and gotchas



China 2024

```
PS C:\Windows\System32> kubectl top po --containers -n kube-system | Select-String -Pattern "disk|file"
```

csi-azure←[7mdisk←[0m-node-win-m2vrh	azuredisk	841m	707Mi
csi-azure←[7mdisk←[0m-node-win-m2vrh	node-driver-registrar	0m	5Mi
csi-azure←[7mfile←[0m-node-win-xwp4r	azurefile	0m	10Mi
csi-azure←[7mfile←[0m-node-win-xwp4r	node-driver-registrar	0m	5Mi

for details, refer to <https://github.com/kubernetes-sigs/azuredisk-csi-driver/issues/2235>

Key learnings and gotchas



China 2024

- Tips for troubleshooting performance issues with the HostProcess container
 - Use wpr(Windows Performance Recorder) to capture resource usage on Windows node
- for details, refer to <https://github.com/kubernetes-sigs/azuredisk-csi-driver/issues/2235>

```
kubectl exec -it csi-azurefile-node-win-xxx -n kube-system -c azurefile -- cmd  
wpr -start cpu  
# sleep 30s  
wpr -stop 'c:\cpu.etl'  
kubectl cp kube-system/csi-azurefile-node-win-xxx:/cpu.etl -c azurefile /tmp/cpu.etl
```


Key learnings and gotchas

- Use Windows performance analyzer to analyze the cpu consumption on a per-process basis
 - <https://learn.microsoft.com/en-us/windows-hardware/test/wpt/windows-performance-analyzer>

Line #	Process Name	Process	% Weight	Sum S
1	▼ powershell.exe		60.08	
2		▷ powershell.exe (15096)	44.15	
3		▷ powershell.exe (6424)	11.66	
4		▷ powershell.exe (9488)	2.55	
5		▷ powershell.exe (13660)	1.57	
6		▷ powershell.exe (1828)	0.13	
7	▷ svchost.exe		19.01	

NodeGetVolumeStats

(Get-Item -Path \$Env:mount).Target

Line #	Process	Sessio...	Duration (s)	Sum	Command Line	Start Time (s)	Min	End Time (s)	Max
1	powershell.exe (16260)	11	137.082586900		powershell.exe -File C:\entrypoint\entrypoint.ps1	0.000000000		137.082586900	
2	powershell.exe (8516)	15	137.082586900		powershell.exe -File C:\entrypoint\entrypoint.ps1	0.000000000		137.082586900	
3	powershell.exe (6424)	10	137.082586900		powershell.exe -File C:\entrypoint\entrypoint.ps1	0.000000000		137.082586900	
4	powershell.exe (6228)	8	137.082586900		powershell C:\opt\scripts\main.ps1	0.000000000		137.082586900	
5	powershell.exe (4296)	16	137.082586900		C:\Windows\System32\WindowsPowerShell\v1.0\powershell.exe -noexit -Command "Invoke-SCo	0.000000000		137.082586900	
6	powershell.exe (4192)	0	137.082586900		"C:\Windows\System32\WindowsPowerShell\v1.0\powershell.exe" c:\kubeproxy\start.ps1	0.000000000		137.082586900	
7	powershell.exe (1828)	0	137.082586900		"C:\Windows\System32\WindowsPowerShell\v1.0\powershell.exe" c:\kubenet\start.ps1	0.000000000		137.082586900	
8	powershell.exe (9488)	0	0.683055300		powershell -Mta -NoProfile -Command "(Get-Item -Path \$Env:mount).Target"	97.892886300		98.575941600	
9	powershell.exe (15096)	0	0.600230200		powershell -Mta -NoProfile -Command "(Get-Item -Path \$Env:mount).Target"	98.602900200		99.203130400	
10	powershell.exe (13660)	0	0.700041500		powershell -Mta -NoProfile -Command "(Get-Item -Path \$Env:mount).Target"	99.240264500		99.940306000	

Key learnings and gotchas



China 2024

- how to reduce cpu and memory usage of **GetVolumeStats** on Windows node
 - **Tips: PowerShell command calls are expensive, minimize their usage as much as possible**
 - add a 10min expiration cache in **GetVolumeStats**
<https://github.com/kubernetes-sigs/azuredisk-csi-driver/pull/2273>
 - `(Get-Item -Path $Env:mount).Target`

Key learnings and gotchas



China 2024

- Replace PowerShell command calls with Golang API calls as more as possible

Powershell command	Golang API
Get-Volume -Uniqueid \$volumeID Select SizeRemaining,Size	windows.GetDiskFreeSpaceEx
New-Item -ItemType SymbolicLink	os.Symlink

Migration progress of cloud providers



China 2024

- Azure disk and file CSI drivers
 - Migrated and GA from Kubernetes v1.27 on AKS
- AWS ebs driver
 - Alpha stage
- GCP disk driver
 - Working in progress

Conclusion and future work



China 2024

- Benefits of HostProcess container deployment
 - Simplify development and deployment process
 - Simplify troubleshooting
 - Quite smaller image size

- Replace more Powershell cmdlets with Golang APIs in CSI Drivers
 - (Get-Item -Path \$Env:mount).Target
 - Test-Path \$Env:remotepath
 - <https://github.com/kubernetes-sigs/azuredisk-csi-driver/issues/2436>
- Make csi-proxy as a golang library for CSI driver using HPC
 - Upstream performance optimization changes to csi-proxy project
- Support host process image build in upstream csi-release-tools repo
 - <https://github.com/kubernetes-csi/csi-release-tools>
- Migrate more CSI driver on Windows to HostProcess container deployment
 - <https://github.com/kubernetes-csi/csi-driver-smb>

- <https://github.com/kubernetes-csi/csi-proxy>
- <https://kubernetes.io/docs/tasks/configure-pod-container/create-hostprocess-pod/>
- <https://github.com/microsoft/windows-host-process-containers-base-image>
- Example of CSI driver using HostProcess container
 - https://github.com/kubernetes-sigs/azuredisk-csi-driver/blob/master/pkg/mounter/safe_mounter_host_process_windows.go
 - <https://github.com/kubernetes-sigs/azuredisk-csi-driver/blob/master/deploy/csi-azuredisk-node-windows-hostprocess.yaml>

A high-altitude mountain landscape featuring a vibrant turquoise lake nestled in a valley. In the foreground, a hiker wearing an orange jacket stands on a rocky outcrop, holding a trekking pole. The surrounding slopes are covered in sparse green vegetation and rocks. In the background, majestic snow-capped mountain peaks rise against a clear blue sky with some light clouds. The overall scene conveys a sense of adventure and natural beauty.

Thanks!
Q & A