SUSEcondigital '20

Using Ceph for persistent storage on a Kubernetes platform

BP-1133

Cameron Seader

Technology Strategist

SUSE

Agenda

1. What are we trying to solve?
2. Types of Volumes
3. What is the CSI
4. SUSE Enterprise Storage Prereqs and Setup
5. Rook Prereqs and Setup
6. Demo Using RBD & CephFS in Kubernetes

What are we trying to solve?

On-disk files in a Container are ephemeral, which presents some problems for critical applications when running in Containers.

First, when a Container crashes, kubelet will restart it, but the files will be lost and the Container starts with a clean state.

Second, when running Containers together in a Pod it is often necessary to share files between those Containers.

The Kubernetes Volume abstraction solves both of these problems.

Types of Volumes

Standard / Open:

- Hostpath, iscsi, local, nfs, fibre channel, rbd, cephfs …

3rd Party:

- Linstor, Glusterfs, PortworxVolume, Quobyte, scaleIO, storageos, vspherevolume, NetApp, Nutanix, EMC, …

CSP Specific:

- Azure, GCP, AWS …

What is the CSI?

Container Storage Interface

In a nutshell

“Standard interface to expose storage to your container workload.”

https://github.com/container-storage-interface/spec/blob/master/spec.md

Introduced in K8s ver. 1.9, but most stable in ver. 1.18

New drivers available for CSI for just about everything

https://kubernetes-csi.github.io/docs/drivers.html

https://github.com/ceph/ceph-csi

What is the CSI?

Typical Driver Features

- Topology, Secrets & Creds, Raw Block Volume, Skip Attach, Pod Info, Expansion, Data Sources {Snapshot, Cloning}, Volume Limit, Ephemeral

Sidecar Containers

- Provisioner, attacher, node-driver-registrar, snapshotter

Ceph CSI Support Matrix

Plugin	Features	Feature Status	CSI Driver Version	CSI Spec Version	Ceph Cluster Version	Kubernetes Version
RBD	Dynamically provision, de-provision Block mode RWO volume	GA	>= v1.0.0	>= v1.0.0	Mimic (>=v13.0.0)	>= v1.14.0
	Dynamically provision, de-provision Block mode RWX volume	GA	>= v1.0.0	>= v1.0.0	Mimic (>=v13.0.0)	>= v1.14.0
	Dynamically provision, de-provision File mode RWO volume	GA	>= v1.0.0	>= v1.0.0	Mimic (>=v13.0.0)	>= v1.14.0
	Creating and deleting snapshot	Alpha	>= v1.0.0	>= v1.0.0	Mimic (>=v13.0.0)	>= v1.14.0
	Provision volume from snapshot	Alpha	>= v1.0.0	>= v1.0.0	Mimic (>=v13.0.0)	>= v1.14.0
	Provision volume from another volume	-	-	-	-	-
	Resize volume	Beta	>= v2.0.0	>= v1.1.0	Mimic (>=v13.0.0)	>= v1.15.0
	Metrics Support	Beta	>= v1.2.0	>= v1.1.0	Mimic (>=v13.0.0)	>= v1.15.0
CephFS	Dynamically provision, de-provision File mode RWO volume	Beta	>= v1.1.0	>= v1.0.0	Nautilus (>=14.2.2)	>= v1.14.0
	Dynamically provision, de-provision File mode RWX volume	Beta	>= v1.1.0	>= v1.0.0	Nautilus (>=v14.2.2)	>= v1.14.0
	Creating and deleting snapshot	-	-	-	-	-
	Provision volume from snapshot	-	-	-	-	-
	Provision volume from another volume	-	-	-	-	-
	Resize volume	Beta	>= v2.0.0	>= v1.1.0	Nautilus (>=v14.2.2)	>= v1.15.0
	Metrics	Beta	>= v1.2.0	>= v1.1.0	Nautilus (>=v14.2.2)	>= v1.15.0

SUSE Enterprise Storage
Prereqs

Software on each Kubernetes worker

- ceph-common, xfsprogs

Kubernetes Cluster must communicate with Ceph Monitors, OSD nodes, and metadata nodes.

Provisioned RBD Pool

Provisioned Cephfs Volume

Ceph RBD Setup

Create Ceph CSI ConfigMap

Create ceph RBD secret in Kubernetes

Create Ceph CSI Plugin (Provisioner & Node)

Create Ceph RBD StorageClass

Create PersistentVolumeClaim

Create a Pod that uses your PVC

Sample RBD StorageClass yaml

```
apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
   name: csi-rbd-sc
provisioner: rbd.csi.ceph.com
parameters:
   clusterID: b9127830-b0cc-4e34-aa47-9d1a2e9949a8
   pool: kubernetes
   csi.storage.k8s.io/provisioner-secret-name: csi-rbd-secret
   csi.storage.k8s.io/provisioner-secret-namespace: default
   csi.storage.k8s.io/node-stage-secret-name: csi-rbd-secret
   csi.storage.k8s.io/node-stage-secret-namespace: default
reclaimPolicy: Delete
mountOptions:
   - discard
```

Ceph RBD Quick Setup

Use the Helm Chart

https://github.com/ceph/ceph-csi/tree/master/charts/ceph-csi-rbd

Fill in your values in the values.yaml file about your ceph cluster

```
# helm install --namespace "ceph-csi-rbd" --name "ceph-csi-rbd" ceph-csi/ceph-csi-rbd
```

Provides kubernetes sidecar

- Provisioner, attacher, node-driver-registrar, snapshotter

CephFS Setup

Create CephFS CSI Plugin

Create Cephfs Storage Class

Create PersistentVolumeClaim

Create a Pod that uses your PVC

CephFS Quick Setup

Use the Helm Chart

https://github.com/ceph/ceph-csi/tree/master/charts/ceph-csi-cephfs

Fill in your values in the values.yaml file about your ceph cluster

```
# helm install --namespace "ceph-csi-cephfs" --name "ceph-csi-cephfs" ceph-csi/ceph-csi-cephfs
```

Provides kubernetes sidecar

- Provisioner, attacher, node-driver-registrar, snapshotter

SUSE Enterprise Storage

Rook Prereqs

At least 3 worker nodes

1 extra disk per worker node

SUSE Rook manifests on K8s management node

Rook Ceph Cluster setup on K8s

Rook RBD and CephFS Setup

Create RBD Pool and Filesystem Pools

Create RBD and CephFS StorageClass

Create PersistentVolumeClaim

Create Pod using PVC

Demo of RBD and CephFS use in Kubernetes

References

CaaS Platform Documentation for SES Integration

https://documentation.suse.com/suse-caasp/4.1/html/caasp-admin/_integration.html#ses-integration

Rook Ceph on CaaS Platform (Technical Preview)

https://documentation.suse.com/ses/6/html/ses-all/cha-container-kubernetes.html

Helpful Upstream Docs

https://rook.io/docs/rook/v1.0/ceph-csi-drivers.html

https://github.com/ceph/ceph-csi

General Disclaimer

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of SUSE, LLC, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.


SUSEcondigital
'20