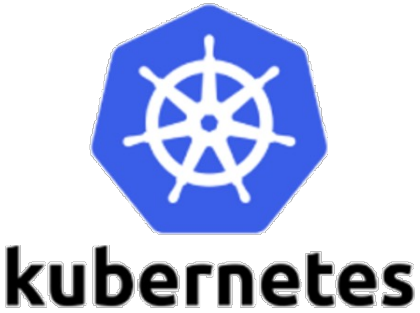

AI and Scientific Research Computing with Kubernetes
Storage

A tutorial at PEARC24

July 22, Providence, Rhode Island

Presented by Mahidhar Tatineni and Dmitry Mishin
University of California San Diego – San Diego Supercomputer Center

Ref: Tutorials at PEARC, SC, 5NRP by Igor Sfiligoi, Dmitry Mishin, and Mahidhar Tatineni

Ephemeral storage

(work area
while the pod is running)

Storage inside the container image

- All areas inside the container are writable (typically)
- You can write data straight into the directories provided by the image

Ephemeral partition

- Sometimes you need a larger and faster partition
- Kubernetes allows for an explicit ephemeral mount
- Known as an emptyDir volume

RAM disk

- As with all Linux systems,
  RAM disk is mounted in all containers
- But (typically) by default limited to 64M
- Must explicitly request a larger one
  (memory-based emptyDir)

https://kubernetes.io/docs/concepts/storage/volumes/#emptydir

Using external storage

External storage essential for persistency

- Remember, ephemeral storage is gone once the pod is gone
- Most applications will need some persistency

Kubernetes provides several hooks at Pod launch time

- Remote filesystem (e.g. NFS, CEPH)
- Block storage (seen as a block device in the pod)
- Local storage, typically ephemeral but can be persistent
- Direct access to external services (e.g. S3, HTTP/WebDAV, Globus, scp)

Not really k8s-native but still useful

https://kubernetes.io/docs/concepts/storage/volumes/
https://kubernetes-csi.github.io/docs/

Mounting storage

Pick the volume to mount

- You may be able to create it at Pod creation type
- But most persistent storage pre-created as Persistent Volume Claims (PVC)

Mount it inside the container

- Any directory path will work
- Whatever works for you

Example PVC creation yaml

```
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
  name: vol-mahidhar
spec:
  storageClassName: rook-cephfs
  accessModes:
  - ReadWriteMany
  resources:
    requests:
      storage: 1Gi
```

Example PVC mount yaml

```
apiVersion: batch/v1
kind: Job
metadata:
  name: s3-mahidhar
spec:
  completionMode: Indexed
  completions: 10
  parallelism: 10
  ttlSecondsAfterFinished: 1800
  template:
    spec:
      restartPolicy: OnFailure
      containers:
      - name: mypod
        image: rockylinux:8
        resources:
           limits:
             memory: 100Mi
             cpu: 0.1
           requests:
             memory: 100Mi
             cpu: 0.1
        command: ["sh", "-c", "let s=2*$JOB_COMPLETION_INDEX; d=`date +%s`; date; sleep $s; (echo Job
$JOB_COMPLETION_INDEX; ls -l /mnt/mylogs/)  > /mnt/mylogs/log.$d.$JOB_COMPLETION_INDEX; sleep 1000"]
        volumeMounts:
        - name: mydata
          mountPath: /mnt/mylogs
      volumes:
      - name: mydata
        persistentVolumeClaim:
          claimName: vol-mahidhar
```

Example CVMFS mount yaml

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: s4-mahidhar
spec:
  containers:
  - name: mypod
    image: rockylinux:8
    resources:
      limits:
        memory: 1Gi
        cpu: 1
      requests:
        memory: 100Mi
        cpu: 100m
    command: ["sleep", "1000"]
    volumeMounts:
    - name: cvmfs
      mountPath: /cvmfs
      readOnly: true
      mountPropagation: HostToContainer
  volumes:
  - name: cvmfs
    persistentVolumeClaim:
      claimName: cvmfs
```

Fetching the output

Stdout and stderr can be accessed at any time

- `kubectl logs <pod name>`

If you used persistent storage, output can be stored and remain there

Or you can explicitly copy it out

- Pick you favorite (non-K8S) tool (e.g. S3, Globus, scp)

Acknowledgements

This work was partially funded by
US National Science Foundation (NSF) awards
OAC-2112167, OAC-1826967, OAC-1541349,
OAC-2030508 and CNS-1730158.

And now the hands-on session

AI and Scientific Research Computing with Kubernetes - Storage