Introduction to Data Protection WG

Xing Yang, VMware & Xiangqian Yu, Google

Speakers

KubeCon CloudNativeCon Europe 2020 Virtual

Xing Yang, VMware
Data Protection WG Co-Chair
K8S SIG-Storage Co-Chair
Slack: xyang
Github: xing-yang
Twitter: @2000Xyang

Xiangqian Yu, Google
Data Protection WG Co-Chair
Slack: xiangqian
Github: yuxiangqian

Agenda

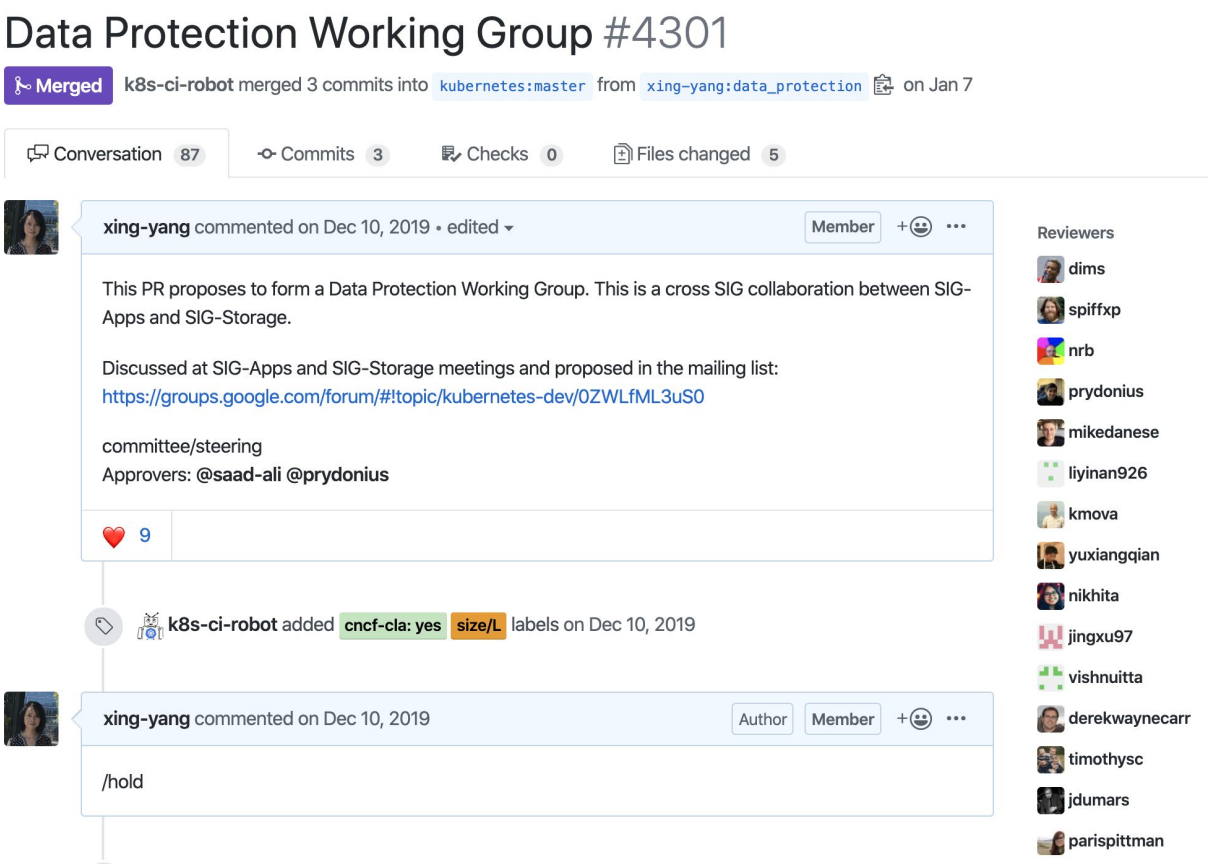
- Motivation and history
- Who are involved
- Charter
- Data Protection definition
- Existing building blocks
- Missing building blocks
- What are we working on
- How to get involved

Motivation

- Day One Operation involving workloads and volume provisioning is well known
- How to support Day Two Operations to protect your workloads is still a challenge.

History

- VolumeSnapshot was promoted to Beta in Kubernetes 1.17.
- Started discussions at KubeCon in San Diego, 2019.
- Established in January 2020



Who are involved?

- The following companies are supporting this initiative:

Arrikto, Catalogic Software, Cohesity, Commvault, Dell EMC, Druva, Google, IBM, Kasten, LINBIT, Linkedin, MayaData, Microsoft, Mongo, NetApp, PortWorx, Pure Storage, Red Hat, Rubrik, SUSE, Trilio, Veeam, Veritas, VMware

Charter

- The Data Protection Working Group is organized with the goal of providing a cross SIG forum to discuss how to support data protection in Kubernetes, identify missing functionalities, and work together to design features that are needed to achieve the goal.
- Stakeholder SIGs: SIG-Apps and SIG-Storage

Data Protection Definition

- The main purpose of Data Protection is to ensure that an application and its data can be restored to a previously preserved state after any corruption or loss.
- In Kubernetes context, it mainly involves backup and recovery of two types of entities:
  - API resources
  - Persistent volume data
- We consider it a complicated and layered problem, including backup and recovery at persistent volume level, application level, and cluster level.

Definition Cont'd

Part of the WG's charter is to define a list of Kubernetes native constructs to enable backup/recovery at different levels:

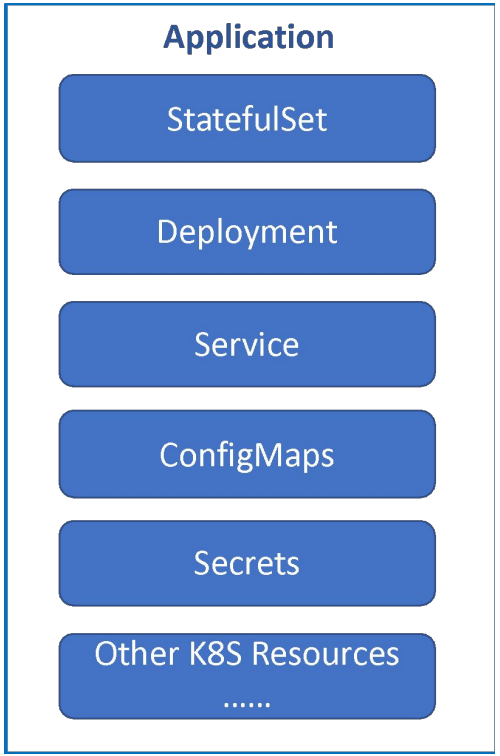
- Persistent Volume Level
  - Volume Snapshot/Backup
  - Volume Restore
- Application Level
  - Application definition
  - Quiesce/Unquiesce hooks
- Cluster Level

Existing Building Blocks

- Building blocks in Application
  - Workload APIs: StatefulSet, Deployment, DaemonSet, etc.
  - Application CRD
- Building blocks in Storage
  - Volume Snapshots

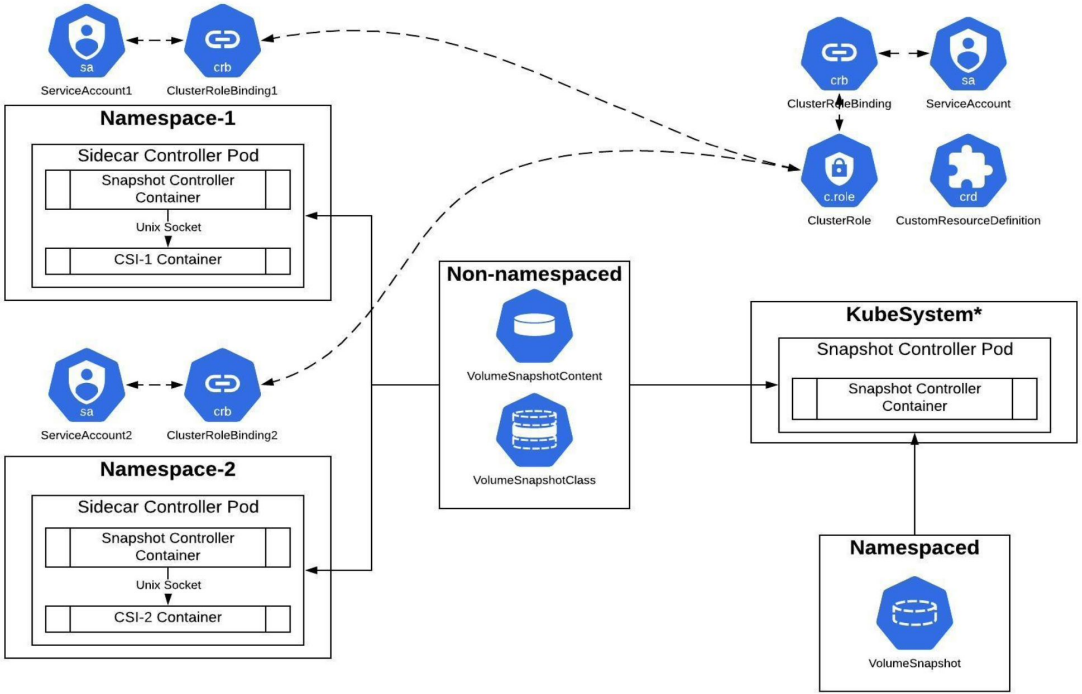
Application CRD

- Provides an API for creating, viewing, and managing applications in Kubernetes.
- Aggregates individual Kubernetes components and manages them together.




```yaml
apiVersion: app.k8s.io/v1beta1
kind: Application
metadata:
  name: mongo
  namespace: default
labels:
  app.kubernetes.io/name: "mongo"
spec:
  selector:
    matchLabels:
      app.kubernetes.io/name: "mongo"
  componentKinds:
  - group: ""
    kind: Service
  - group: apps
    kind: StatefulSet
```

Volume Snapshot



- Volume Snapshot moves to beta in Kubernetes 1.17 release
- Planning to bring it to GA in 1.20:
  - Observability
  - Webhooks
  - E2E tests

Missing Building Blocks

- Volume Backup
- Backup repository
- Quiesce and unquiesce hooks
- Application Snapshot and Backup
- ……

What are we working on?

- Generic Data Populator (design)
  - https://github.com/kubernetes/enhancements/blob/master/keps/sig-storage/20200120-generic-data-populators.md
- Quiesce/unquiesce Hooks (design)
  - ExecutionHook CRD
    https://github.com/kubernetes/enhancements/blob/master/keps/sig-storage/20190120-execution-hook-design.md
  - ContainerNotifier
    https://docs.google.com/document/d/1SWSlZoxY5zFjBKFKaATP07s3q02UenSp8R9-yRkCcwg/edit#
- Volume Group and Group Snapshot (design)
  - https://github.com/kubernetes/enhancements/pull/1551
- Object Bucket Provisioning (design)
  - https://github.com/kubernetes/enhancements/pull/1383
  - New SIG-Storage subproject kubernetes-cosi:
    - https://github.com/kubernetes-sigs/container-object-storage-interface-spec
    - https://github.com/kubernetes-sigs/container-object-storage-interface-csi-adapter

What are we working on?

- Volume Backups (design)
  - https://docs.google.com/document/d/1esDeaQgk5j9w_5j3_IU9NM7QBkMOpawIG4xVyqAd_ZE/edit#
- Application Snapshot, Backup, and Restore (design)
  - https://github.com/kubernetes/enhancements/pull/1051
- Data Protection Workflows (document)

Other Potential Topics

- Diffs between two Snapshots (not started)
  - Change blocks or changed files
- Data Protection Policy (not started)

How to get involved?

- Start at the Data Protection WG community page: https://github.com/kubernetes/community/tree/master/wg-data-protection
- Join the meeting: Bi-weekly meeting on Wednesdays at 9am Pacific Time. Meeting recordings available on YouTube.
  - https://docs.google.com/document/d/15tLCV3csvjHbKb16DVk-mfUmFry_Rlwo-2uG6KNGsfw/edit
- Join the mailing list: https://groups.google.com/forum/#!forum/kubernetes-data-protection
- Join the slack channel: #wg-data-protection


KubeCon
CloudNativeCon
Europe 2020
Virtual
HELM
KEEP CLOUD NATIVE
CONNECTED