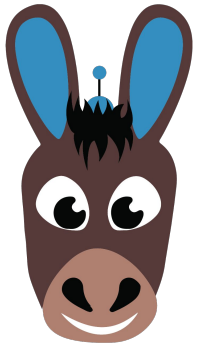


OpenEBS Architecture and Design

v0.6

February 28, 2018

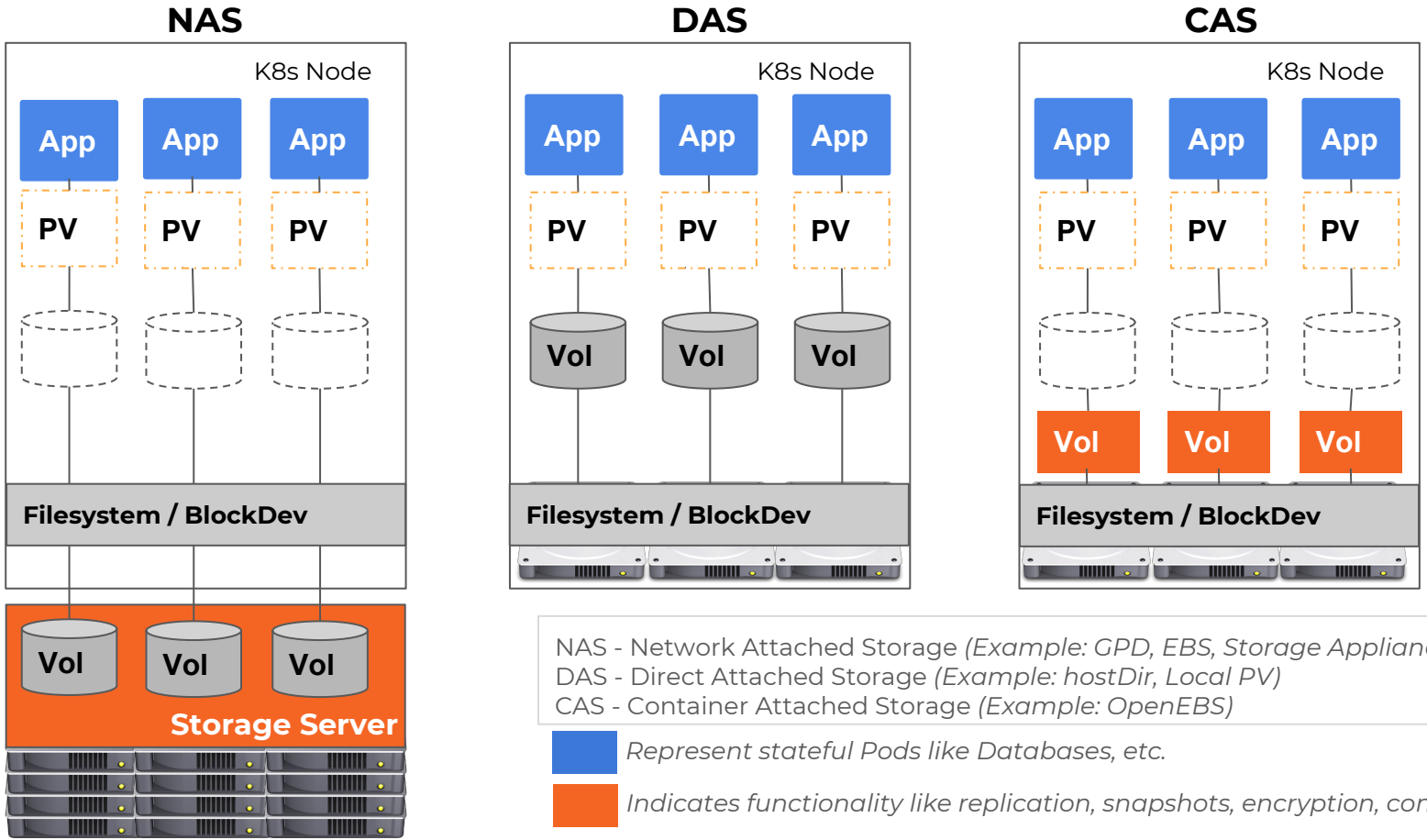
Prerequisites

- **Kubernetes Concepts :** Namespaces, RBAC,CRD, Taints & Tolerations, Pod Anti Affinity, etc.,
- **Kubernetes Storage Concepts:** Running Stateful Workloads via PV, PVC, Storage Classes and Dynamic Provisioner
- ***Kubernetes and CNCF Management Tools:*** *Kube* Dashboard, *Prometheus, Grafana, Opentracing & Jaeger*
- ***Kubernetes Incubator Projects:*** *Node Exporter, Node Problem Detector*

Design Goals and Constraints

- Storage optimized for Containerized Applications
- Stable, Secure and Scalable - Horizontally scalable to millions of Containers, Fault tolerant and Secure by default
- Seamless integration into any private and public cloud environments. Vendor independent.
- Non-disruptive software upgrades
- Easy to setup. Low entry barrier. Developer and Operators Friendly.

Persistent Volume Categories

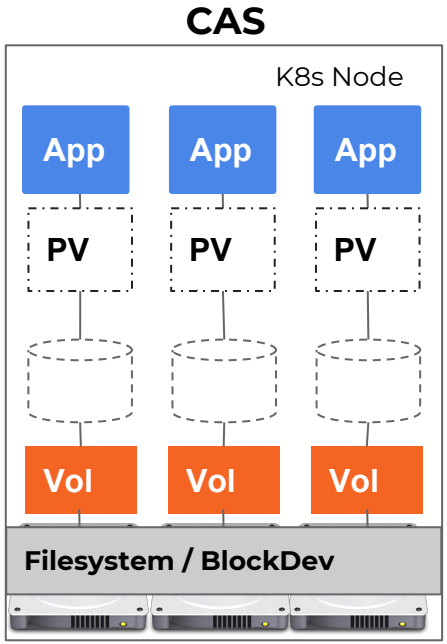


NAS - Network Attached Storage *(Example: GPD, EBS, Storage Appliances)*
DAS - Direct Attached Storage *(Example: hostDir, Local PV)*
CAS - Container Attached Storage *(Example: OpenEBS)*

Represent stateful Pods like Databases, etc.

Indicates functionality like replication, snapshots, encryption, compression, etc.

Container Attached Storage



- Storage Controllers (Targets) are running in containers.
- These Storage Containers are orchestrated by Kubernetes, just like any other workloads.
  - `Installation and Upgrades`
  - `Monitoring`
  - `Debuggability`
- Storage Containers are mainly dealing with:
  - `Disk/Storage Management`
  - `Data - High Availability and`
  - `Data - Protection`

*"OpenEBS is a **CAS** solution, that provides storage as a service to stateful workloads. OpenEBS hooks-into and extends the capabilities of Kubernetes to orchestrate storage services (workloads)"*

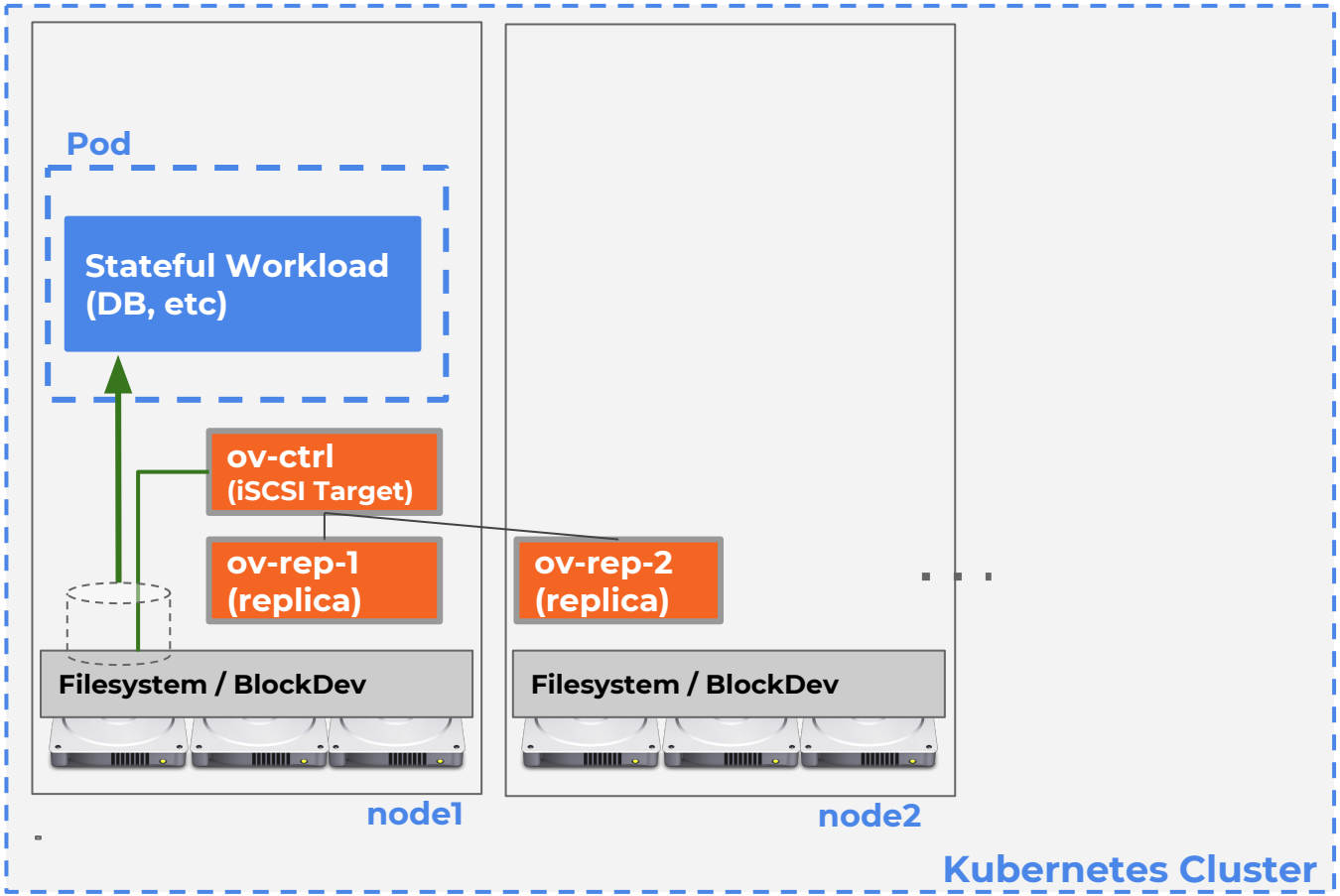
Using OpenEBS Volumes

Application
(Deployment, PVC)

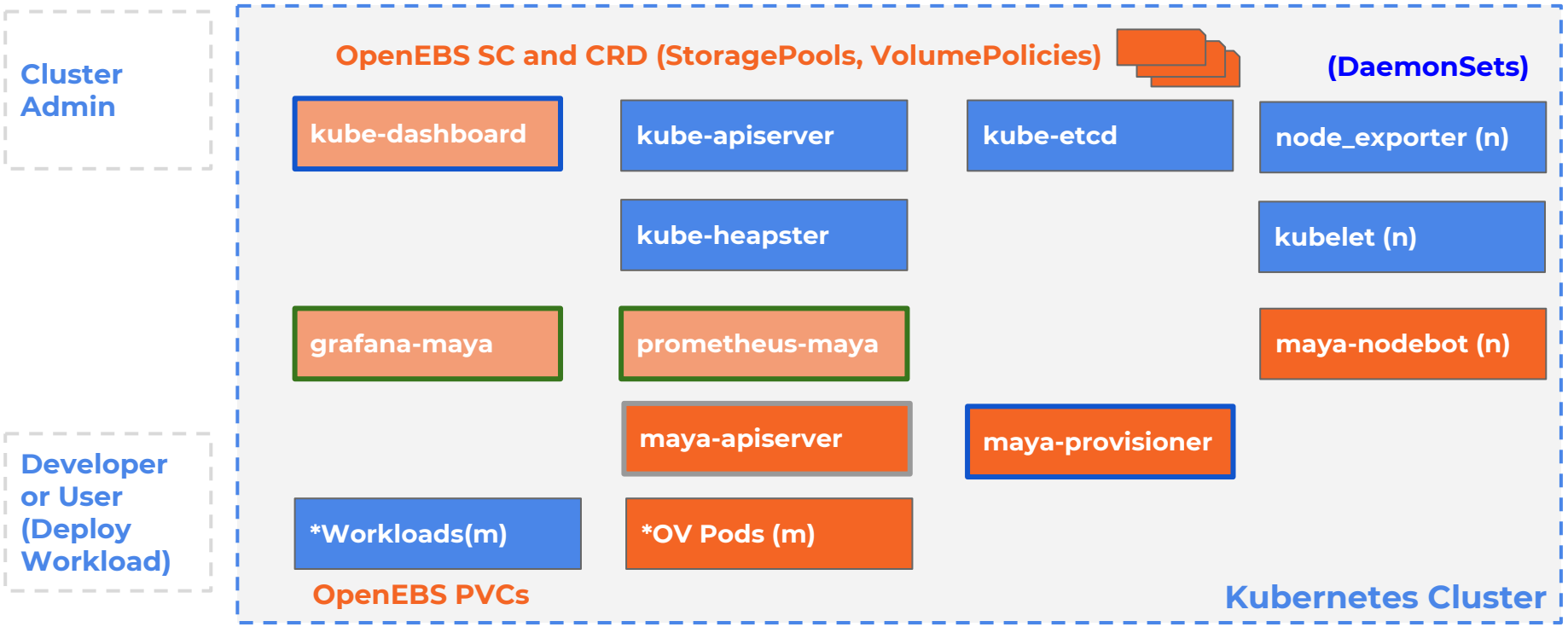
OpenEBS Volume
(iSCSI PV)

Setup OpenEBS
(StorageClasses
StoragePools, Disks)

(iSCSI Initiator)

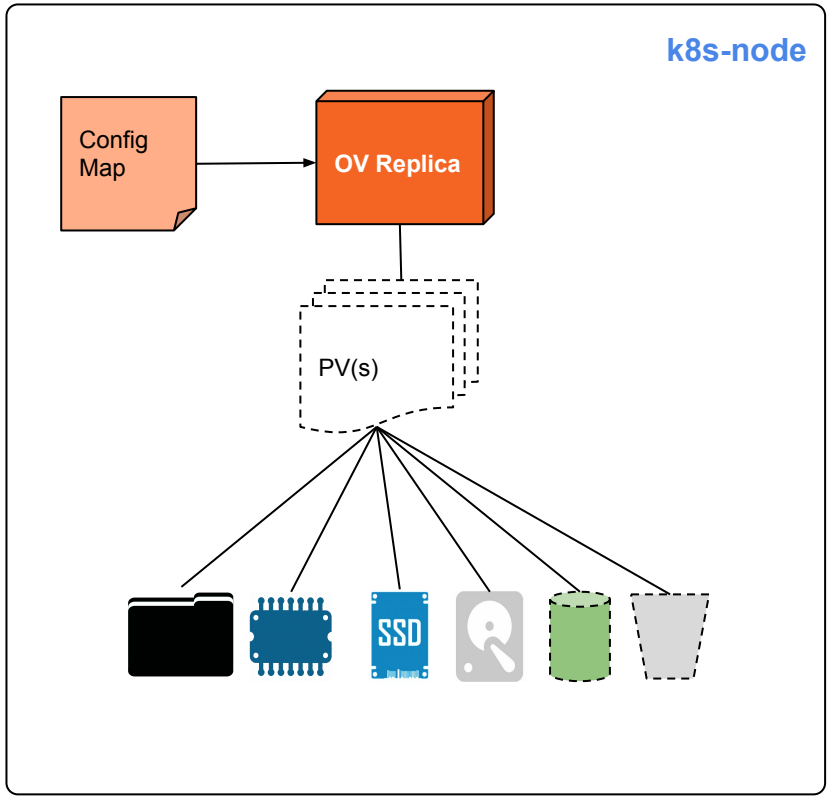


OpenEBS Components



The source code for the control plane components is located in mainly **openebs/maya** repository.

Node Storage Management



K8s-node can be configured with ***Persistent Storage*** as:

- Additional space on OS directory
- NVMe Disks
- SSD Disks
- SAS Disks
- SAN/NAS volumes
- Object Store mounts

OpenEBS Volume Replica (***OV Replica***) is a container running within a K8s Pod. OV Replica will be granted access to local storage on the K8s-node using Local PV Options like:

- hostDir
- Block Disks by mounting (/dev) *

OpenEBS Control Plane manages the discovery, allocation to the OV Replica's and monitoring of the storage attached to the node.

**https://github.com/kubernetes/kubernetes/issues/58569*
**https://docs.google.com/document/d/1fG-KwUQNsuPYY40ByoBFqKJKpxzgyk7cQ5gqsGRXxfk/edit*

OpenEBS - CRDs (extended Schema)

OpenEBS Control Plane uses Kubernetes etcd cluster to store the configuration information about OpenEBS related objects.

While K8s primitives like PV, PVC, SC are used, they can in-turn refer to OpenEBS specific objects for further details.

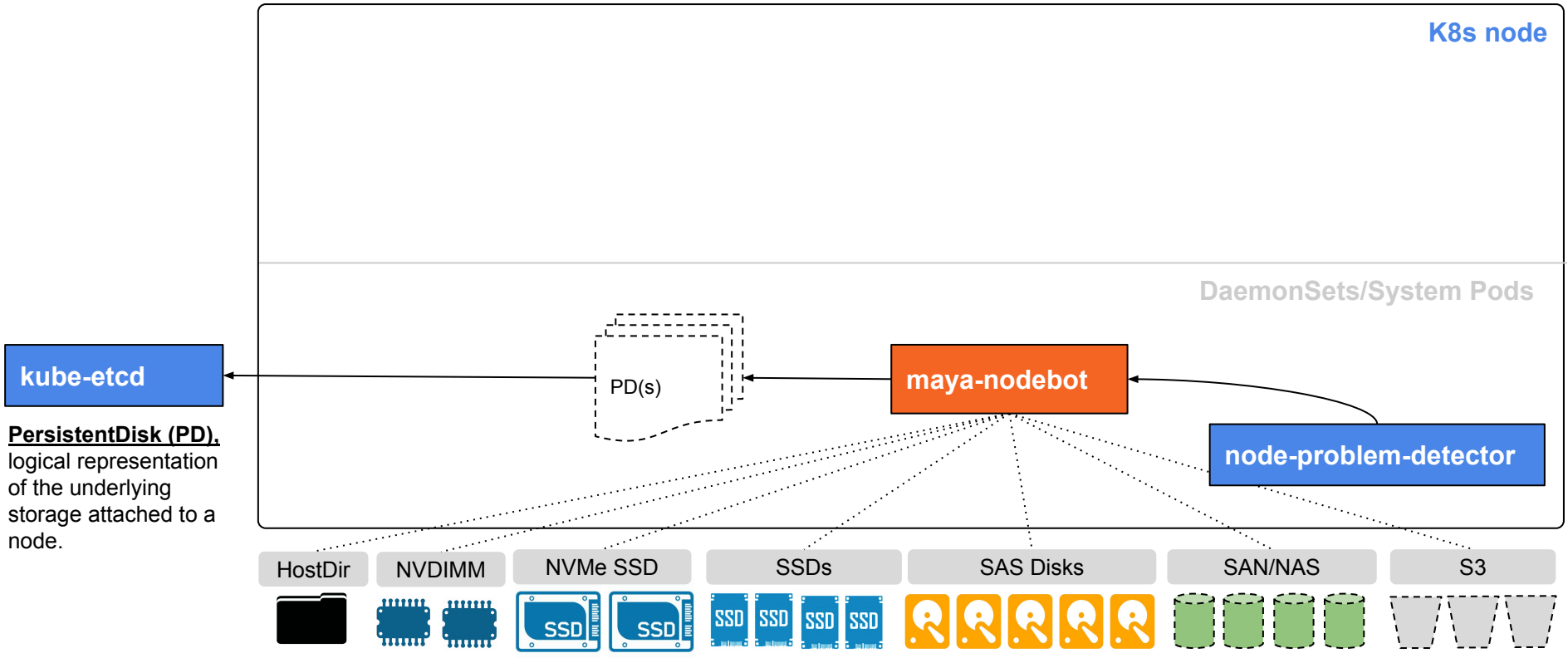
Some of the OpenEBS objects (CRDs) created by the OpenEBS Operator are:

- Persistent Disk
- Storage Pool
- Storage Pool Claim
- Persistent Volume Policy

In addition to the above generic CRDs used by the openebs control plane, there can be custom CRDs specific to the Storage Engines like cStor:

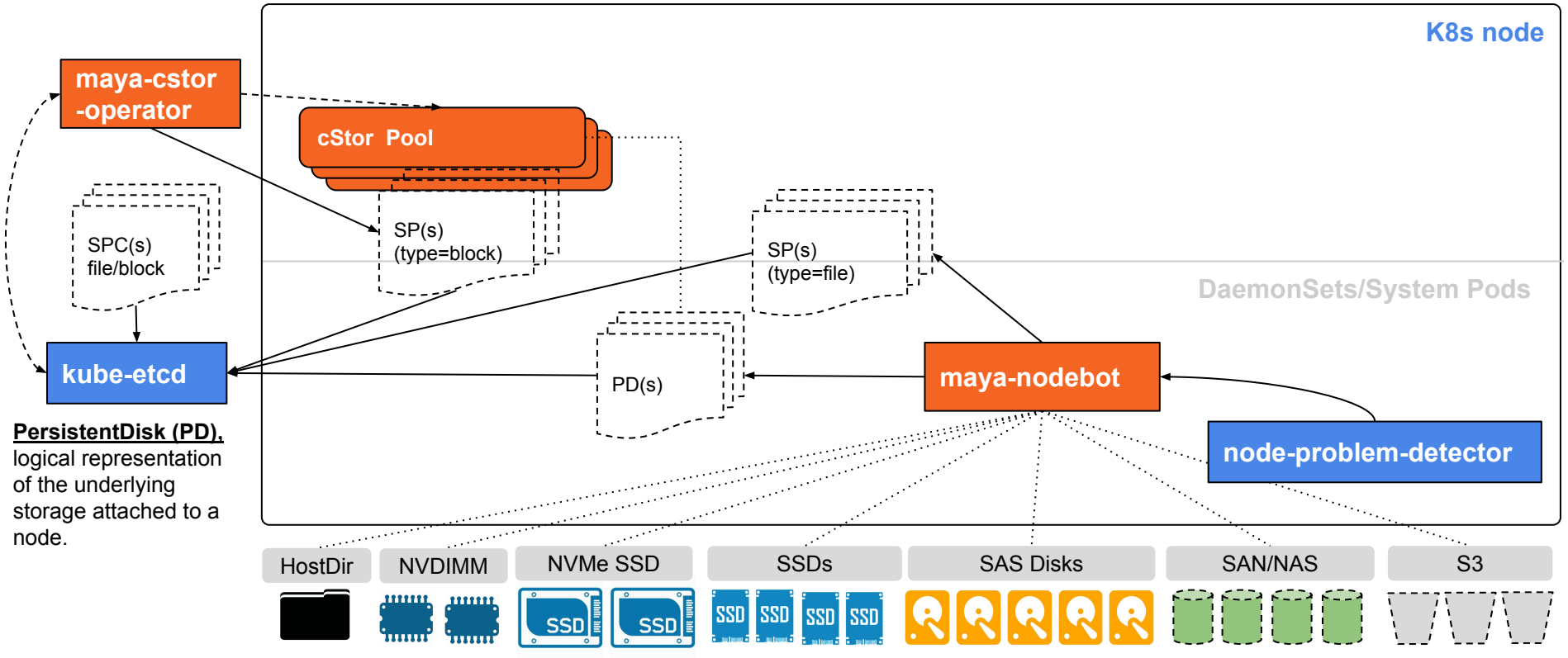
- cStorPool
- cStorVolume

Storage Schema - PersistentDisk



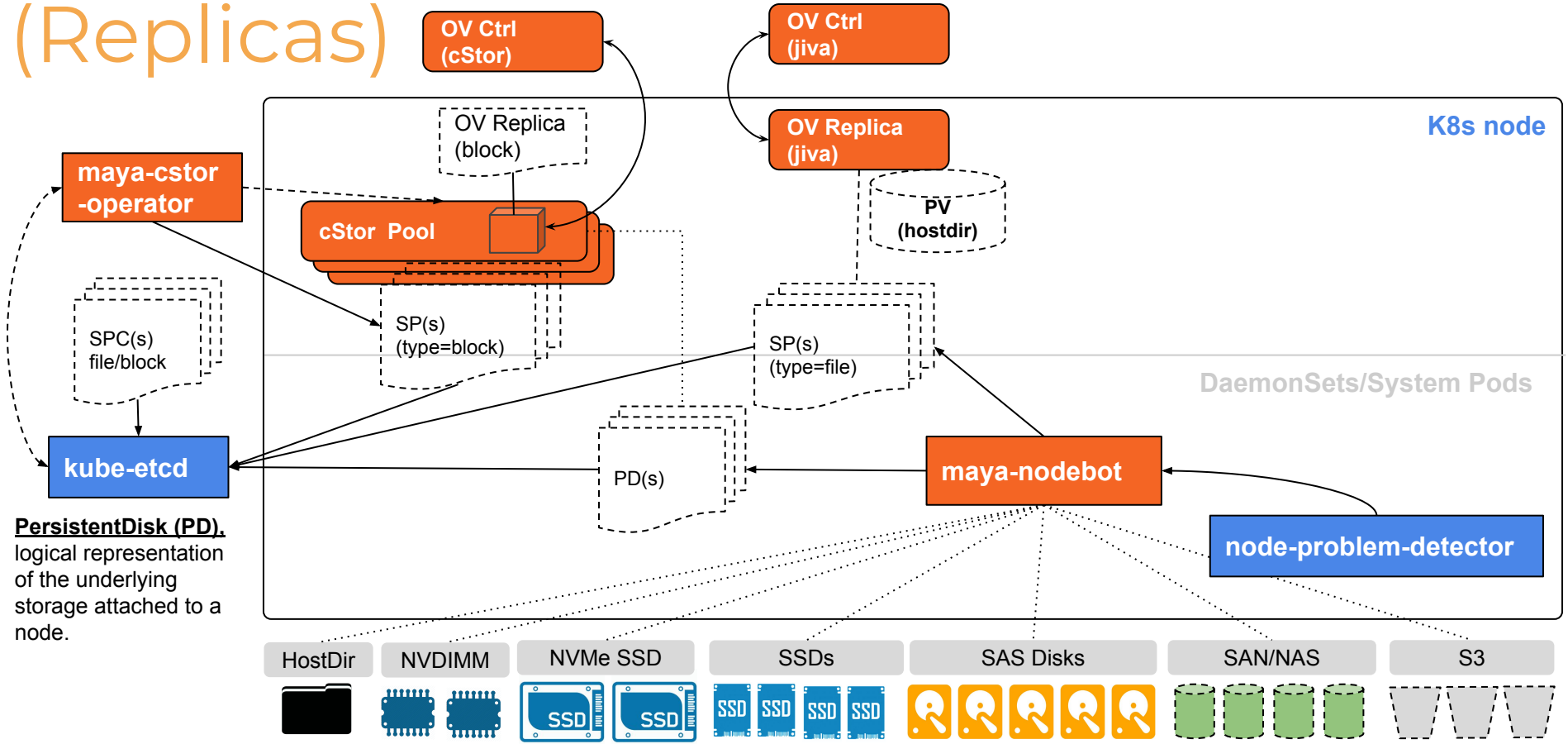
PersistentDisk (PD), logical representation of the underlying storage attached to a node.

Storage Schema - StoragePools



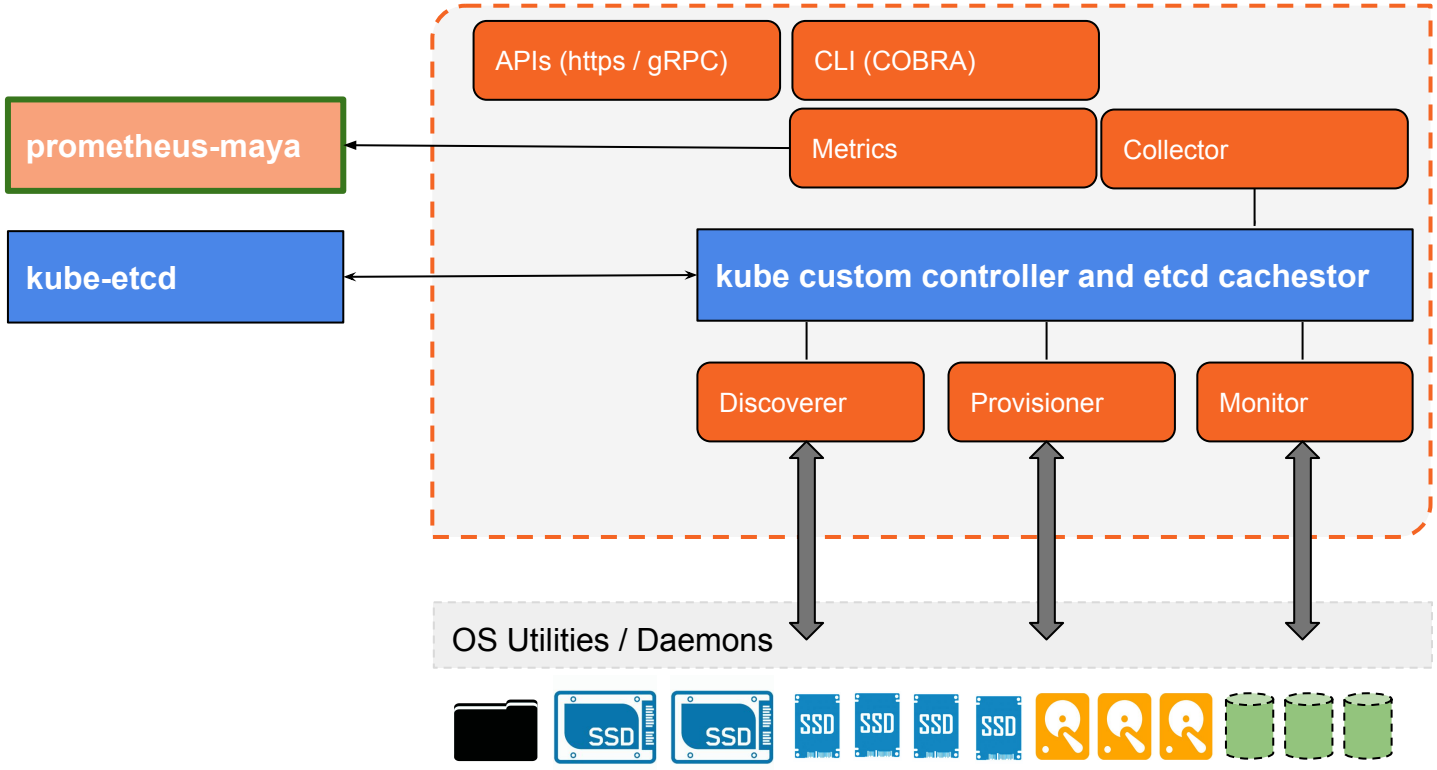
PersistentDisk (PD), logical representation of the underlying storage attached to a node.

Storage Schema - OpenEBS Volumes (Replicas)



PersistentDisk (PD), logical representation of the underlying storage attached to a node.

maya-nodebot



Maya-nodebot runs on each node as a daemon-set.

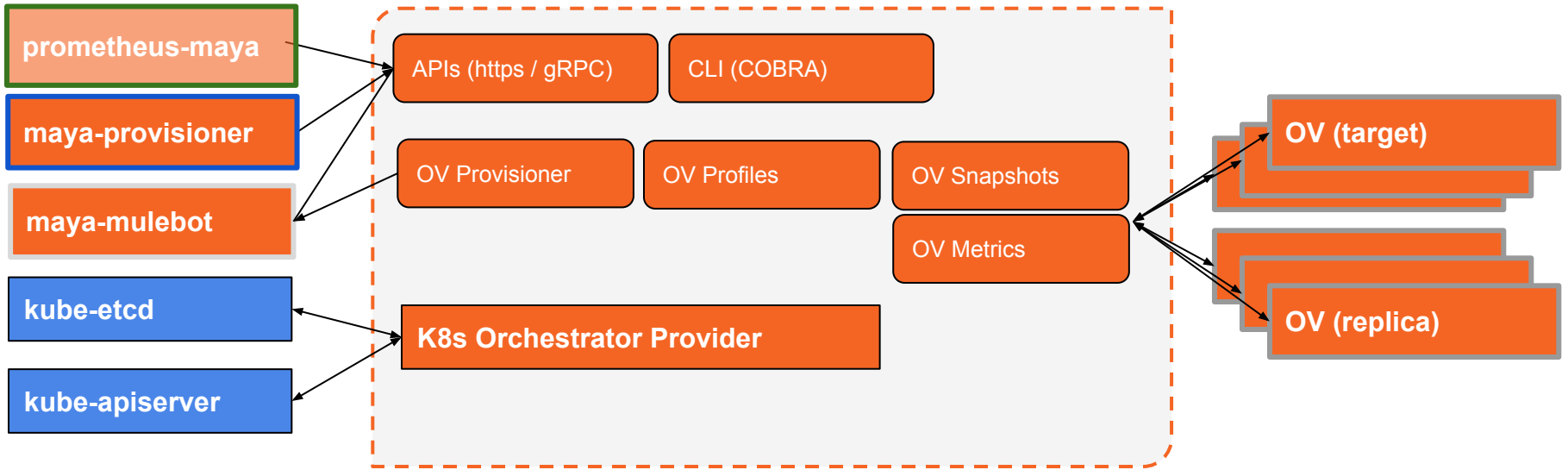
Has access to the underlying disk subsystem and can handle the disk add/lost events and can monitor for disk errors.

Disk details and status (lost/deleted/healthy) are updated as CR into the etcd.

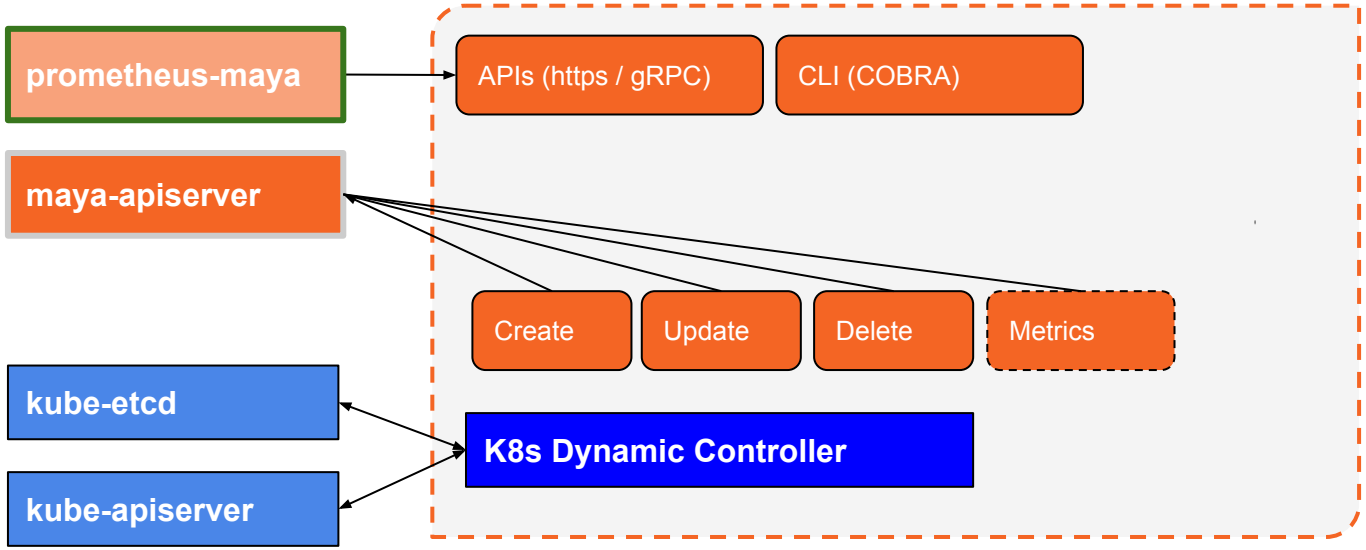
Disks are monitored for IO stats and smart and metrics can be collected via prometheus.

When critical events are observed (say disk lost from an active pool) calls maya-apiserver for handling or can call the Replica impacted.

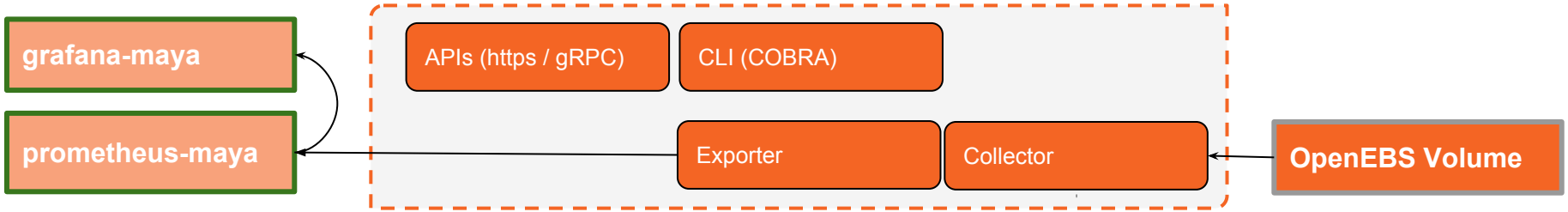
maya-apiserver



maya-provisioner



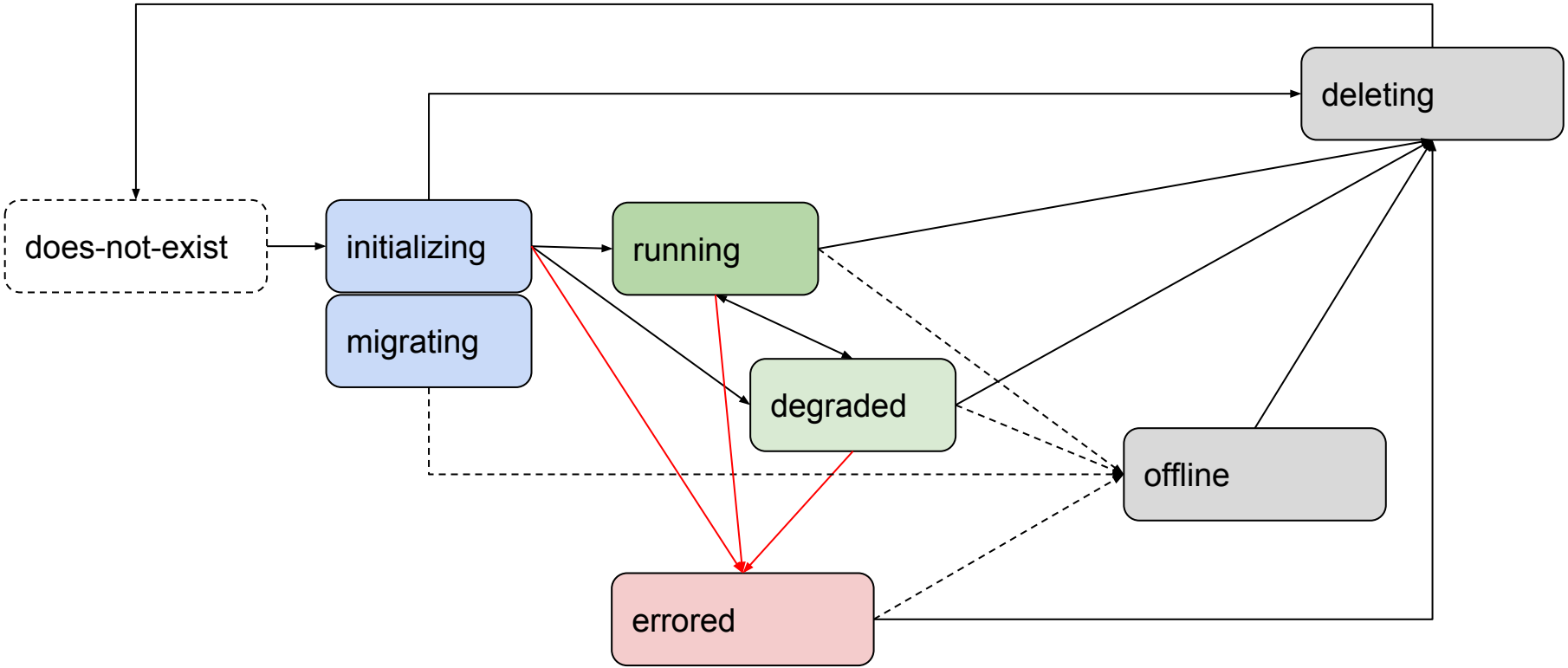
maya-volume-exporter



maya-mulebot (TBD)

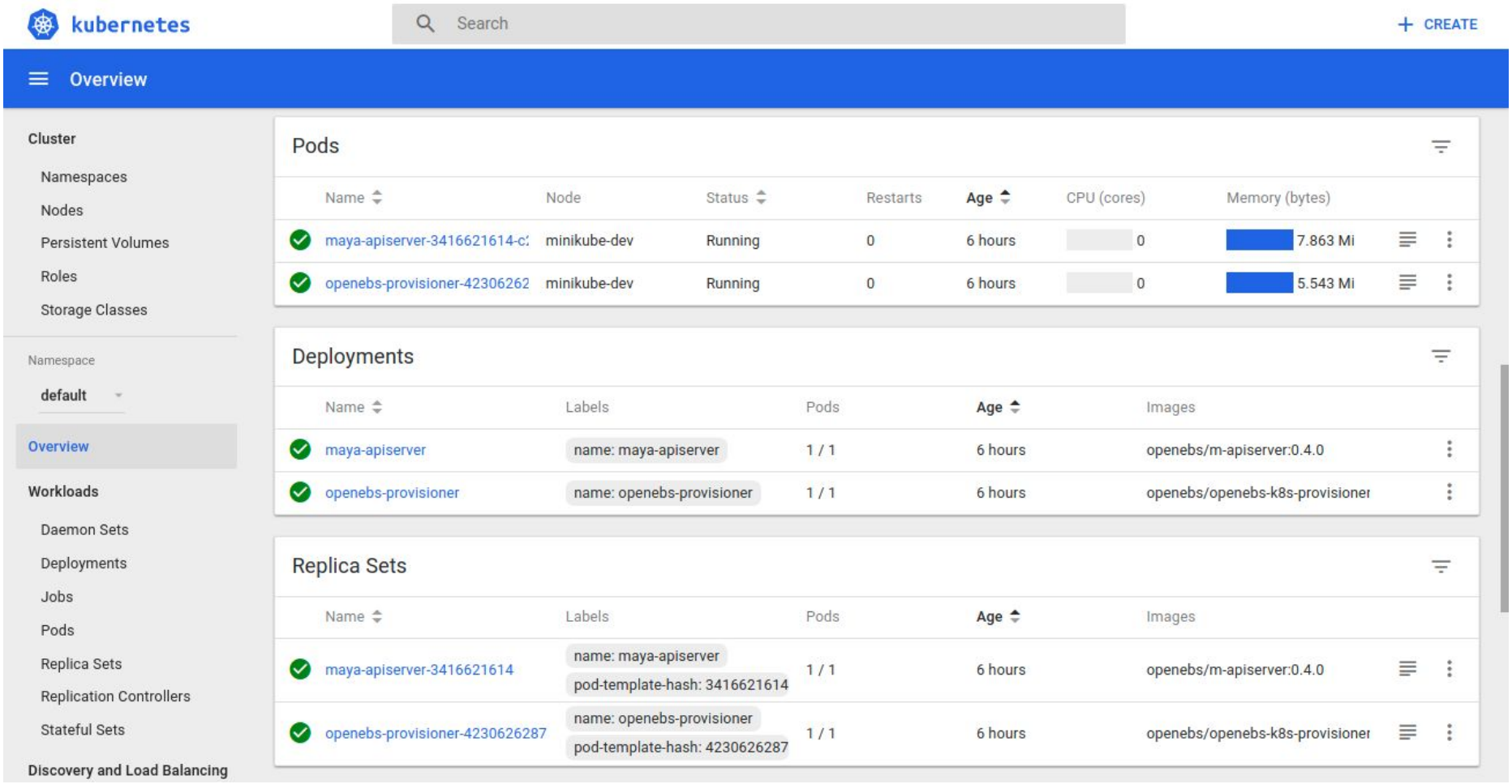
State Diagrams

OpenEBS Volume - State Diagram



Sequence Diagrams (TBD)

UI (Kubernetes Dashboard)


kubernetes
Search
+ CREATE
Overview
Cluster
Namespaces
Nodes
Persistent Volumes
Roles
Storage Classes
Namespace
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Discovery and Load Balancing
Pods
Name
Node
Status
Restarts
Age
CPU (cores)
Memory (bytes)
maya-apiserver-3416621614-c
minikube-dev
Running
0
6 hours
0
7.863 Mi
openebs-provisioner-42306262
minikube-dev
Running
0
6 hours
0
5.543 Mi
Deployments
Name
Labels
Pods
Age
Images
maya-apiserver
name: maya-apiserver
1 / 1
6 hours
openebs/m-apiserver:0.4.0
openebs-provisioner
name: openebs-provisioner
1 / 1
6 hours
openebs/openebs-k8s-provisioner
Replica Sets
Name
Labels
Pods
Age
Images
maya-apiserver-3416621614
name: maya-apiserver
pod-template-hash: 3416621614
1 / 1
6 hours
openebs/m-apiserver:0.4.0
openebs-provisioner-4230626287
name: openebs-provisioner
pod-template-hash: 4230626287
1 / 1
6 hours
openebs/openebs-k8s-provisioner

kubernetes

Search

+ CREATE

Workloads > Pods > maya-apiserver-3416621614-c2116

EXEC LOGS EDIT DELETE

Cluster

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

Namespace

default

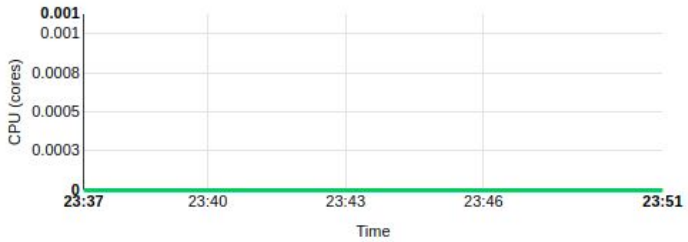
Overview

Workloads

- Daemon Sets
- Deployments
- Jobs
- Pods
- Replica Sets
- Replication Controllers
- Stateful Sets

Discovery and Load Balancing

CPU usage



Memory usage

Details

Name: maya-apiserver-3416621614-c2116

Namespace: default

Labels: name: maya-apiserver pod-template-hash: 3416621614

Annotations: Created by: ReplicaSet maya-apiserver-3416621614

Creation time: 2017-10-12T11:46

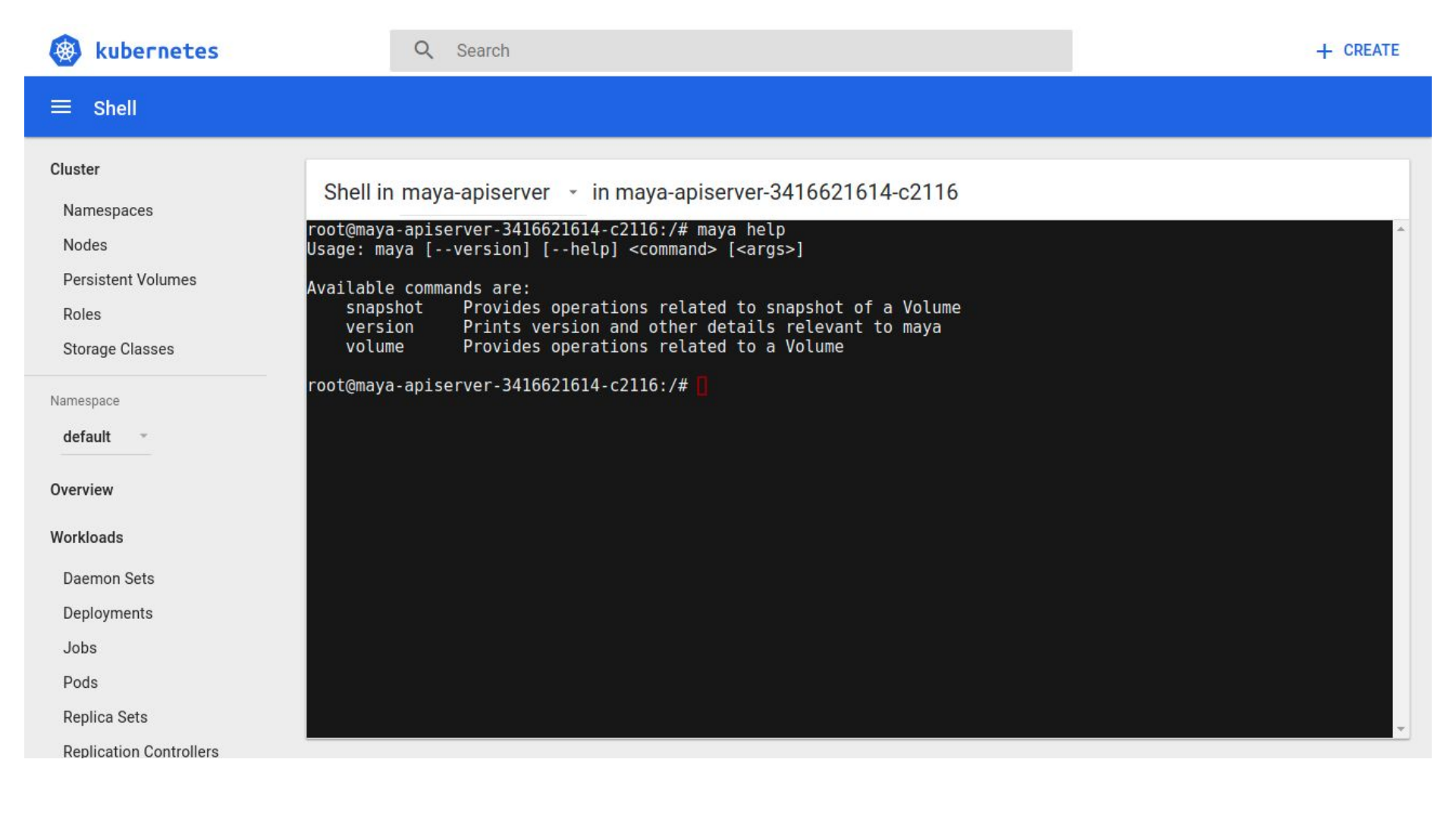
Status: Running

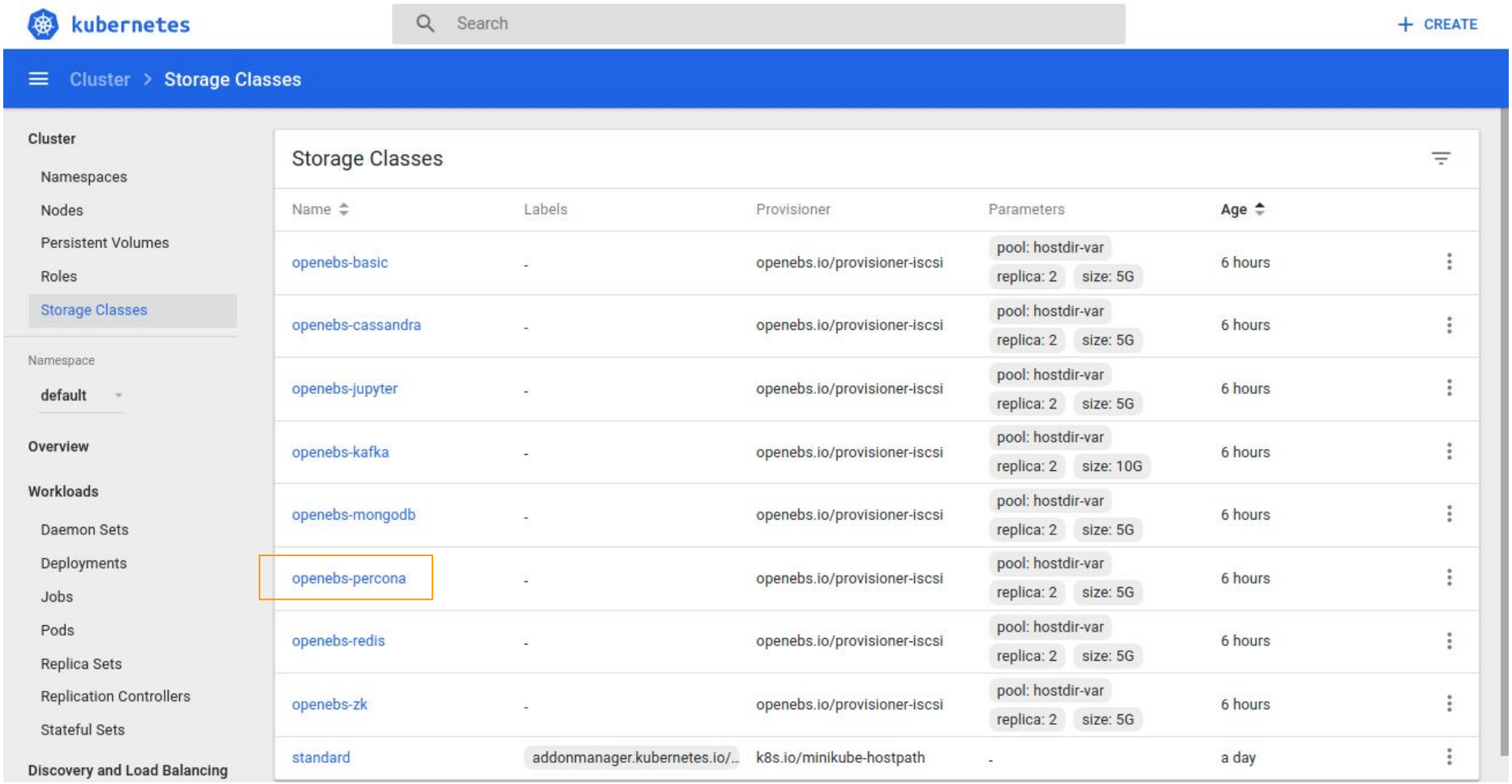
Network

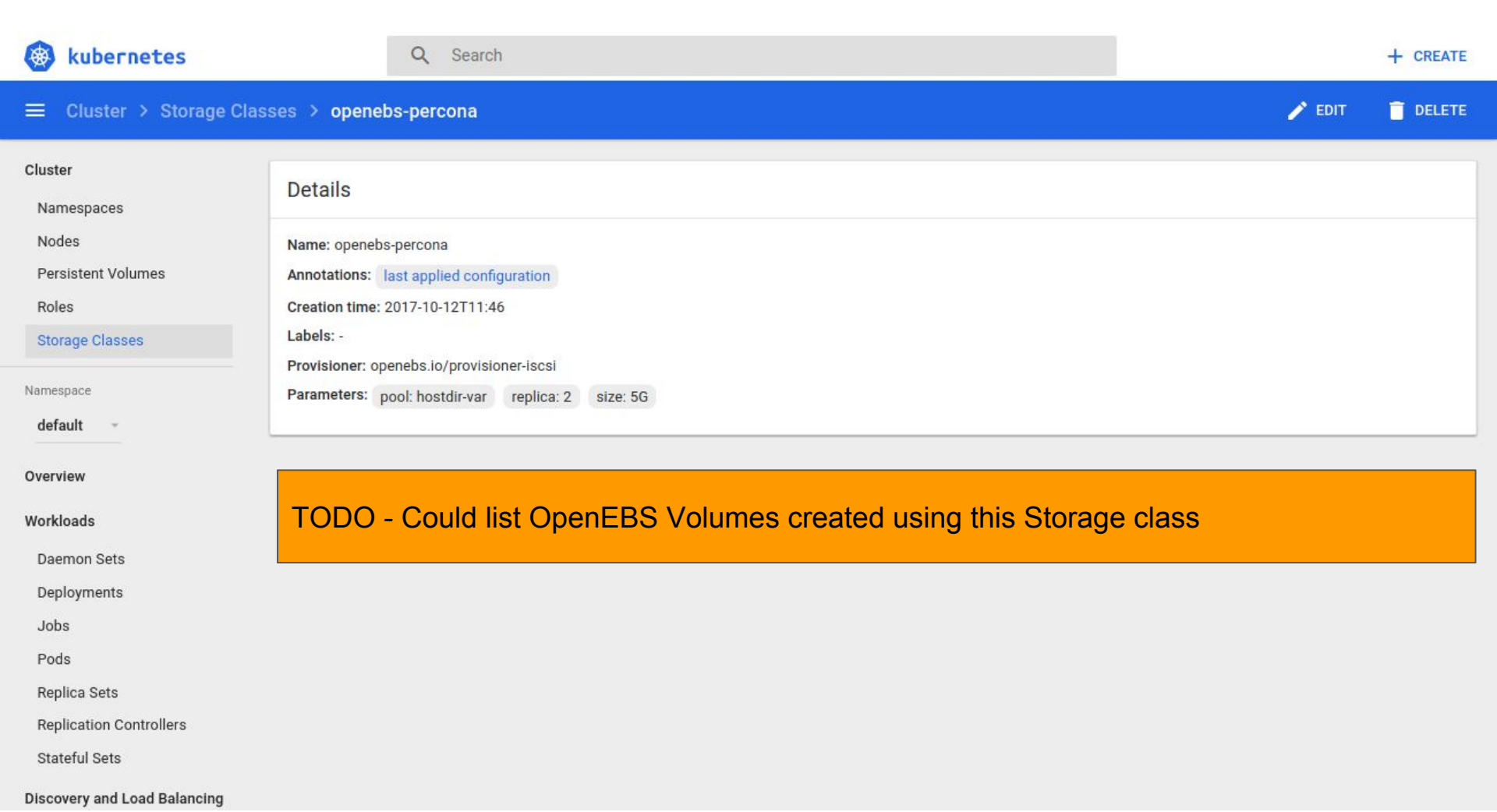
Node: minikube-dev

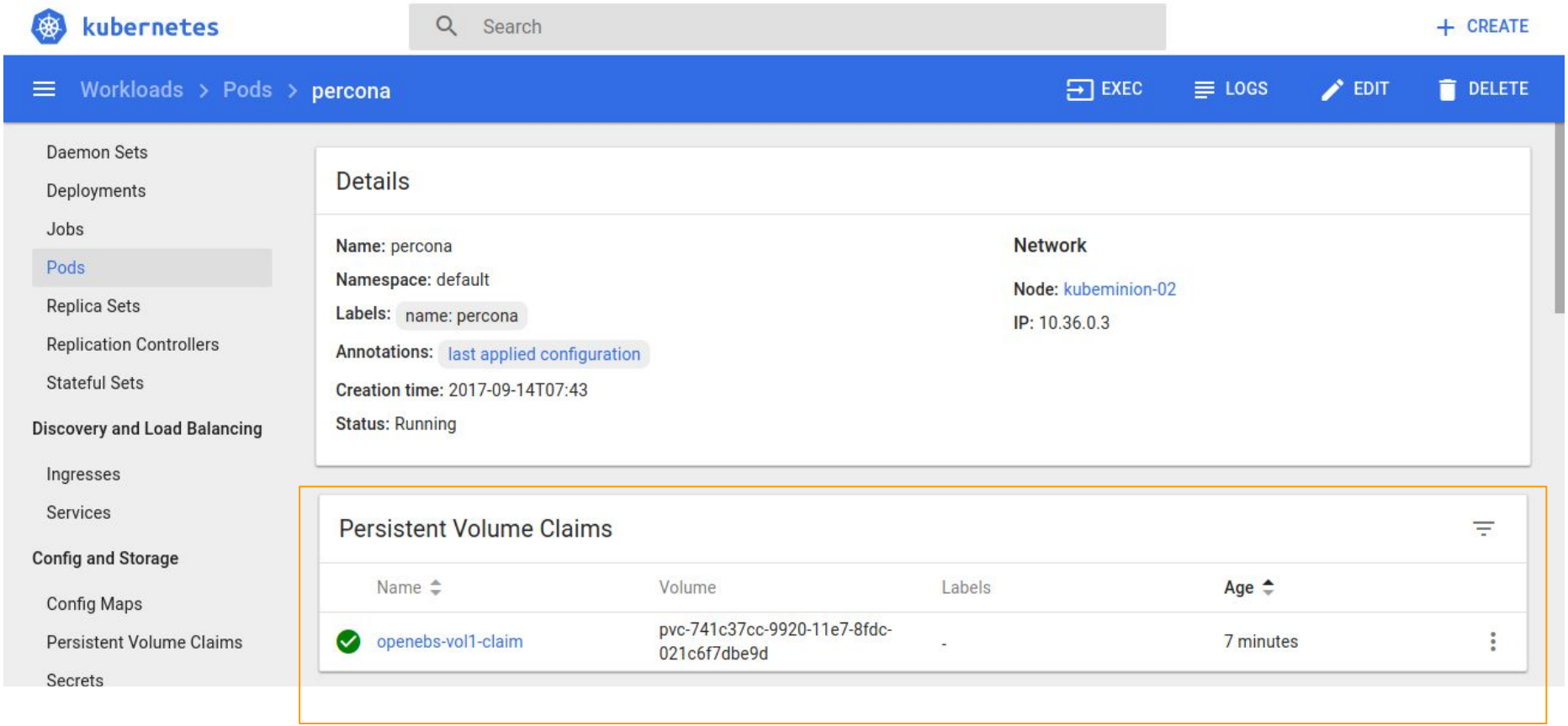
IP: 172.17.0.6

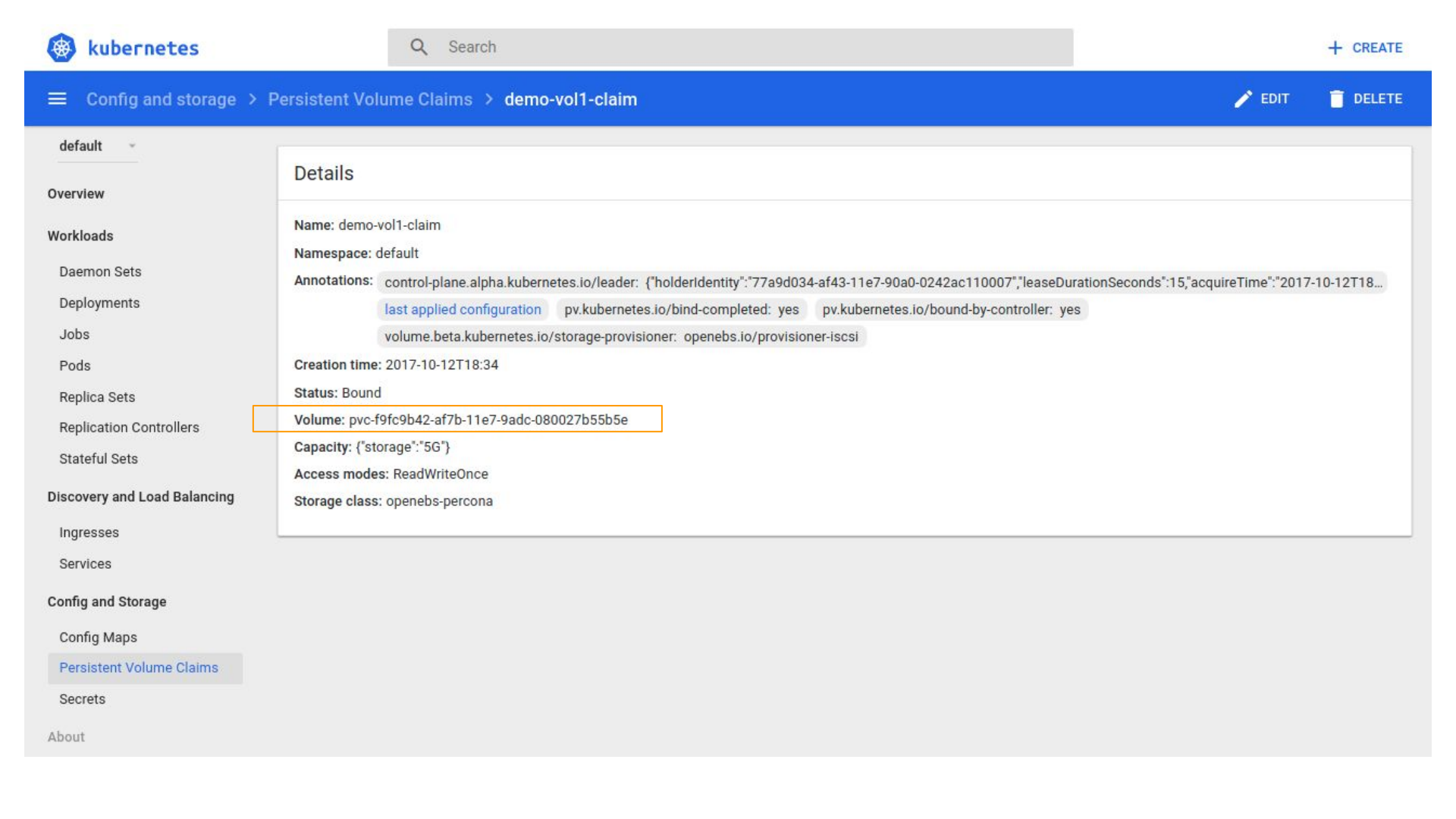
Containers


kubernetes
Search
CREATE
Shell
Cluster
Namespaces
Nodes
Persistent Volumes
Roles
Storage Classes
Namespace
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Shell in maya-apiserver in maya-apiserver-3416621614-c2116
root@maya-apiserver-3416621614-c2116:/# maya help
Usage: maya [--version] [--help] <command> [<args>]
Available commands are:
snapshot Provides operations related to snapshot of a Volume
version Prints version and other details relevant to maya
volume Provides operations related to a Volume
root@maya-apiserver-3416621614-c2116:/#


kubernetes
Search
+ CREATE
Cluster > Storage Classes
Cluster
Namespaces
Nodes
Persistent Volumes
Roles
Storage Classes
Namespace
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Discovery and Load Balancing
Storage Classes
Name
Labels
Provisioner
Parameters
Age
openebs-basic - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
openebs-cassandra - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
openebs-jupyter - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
openebs-kafka - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 10G 6 hours
openebs-mongodb - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
openebs-percona - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
openebs-redis - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
openebs-zk - openebs.io/provisioner-iscsi pool: hostdir-var replica: 2 size: 5G 6 hours
standard addonmanager.kubernetes.io/... k8s.io/minikube-hostpath - a day


kubernetes
Search
CREATE
Cluster > Storage Classes > openebs-percona
EDIT
DELETE
Cluster
Namespaces
Nodes
Persistent Volumes
Roles
Storage Classes
Namespace
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Discovery and Load Balancing
Details
Name: openebs-percona
Annotations: last applied configuration
Creation time: 2017-10-12T11:46
Labels: -
Provisioner: openebs.io/provisioner-iscsi
Parameters: pool: hostdir-var replica: 2 size: 5G
TODO - Could list OpenEBS Volumes created using this Storage class


kubernetes
Search
CREATE
Workloads > Pods > percona
EXEC
LOGS
EDIT
DELETE
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Discovery and Load Balancing
Ingresses
Services
Config and Storage
Config Maps
Persistent Volume Claims
Secrets
Details
Name: percona
Namespace: default
Labels: name: percona
Annotations: last applied configuration
Creation time: 2017-09-14T07:43
Status: Running
Network
Node: kubeminion-02
IP: 10.36.0.3
Persistent Volume Claims
Name
Volume
Labels
Age
openebs-vol1-claim
pvc-741c37cc-9920-11e7-8fdc-021c6f7dbe9d
-
7 minutes


kubernetes
Search
CREATE
Config and storage > Persistent Volume Claims > demo-vol1-claim
EDIT
DELETE
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Discovery and Load Balancing
Ingresses
Services
Config and Storage
Config Maps
Persistent Volume Claims
Secrets
About
Details
Name: demo-vol1-claim
Namespace: default
Annotations: control-plane.alpha.kubernetes.io/leader: {"holderIdentity":"77a9d034-af43-11e7-90a0-0242ac110007","leaseDurationSeconds":15,"acquireTime":"2017-10-12T18...
last applied configuration
pv.kubernetes.io/bind-completed: yes
pv.kubernetes.io/bound-by-controller: yes
volume.beta.kubernetes.io/storage-provisioner: openebs.io/provisioner-iscsi
Creation time: 2017-10-12T18:34
Status: Bound
Volume: pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e
Capacity: {"storage":"5G"}
Access modes: ReadWriteOnce
Storage class: openebs-percona

kubernetes

Search

+ CREATE

Cluster > Persistent Volumes > pvc-741c37cc-9920-11e7-8fdc-021c6f7dbe9d

EDIT DELETE

Cluster

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

Namespace

default

Overview

Workloads

- Daemon Sets
- Deployments
- Jobs
- Pods
- Replica Sets
- Replication Controllers
- Stateful Sets

Allocated resources

CPU

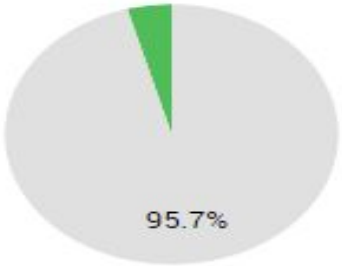
Memory

Up Time

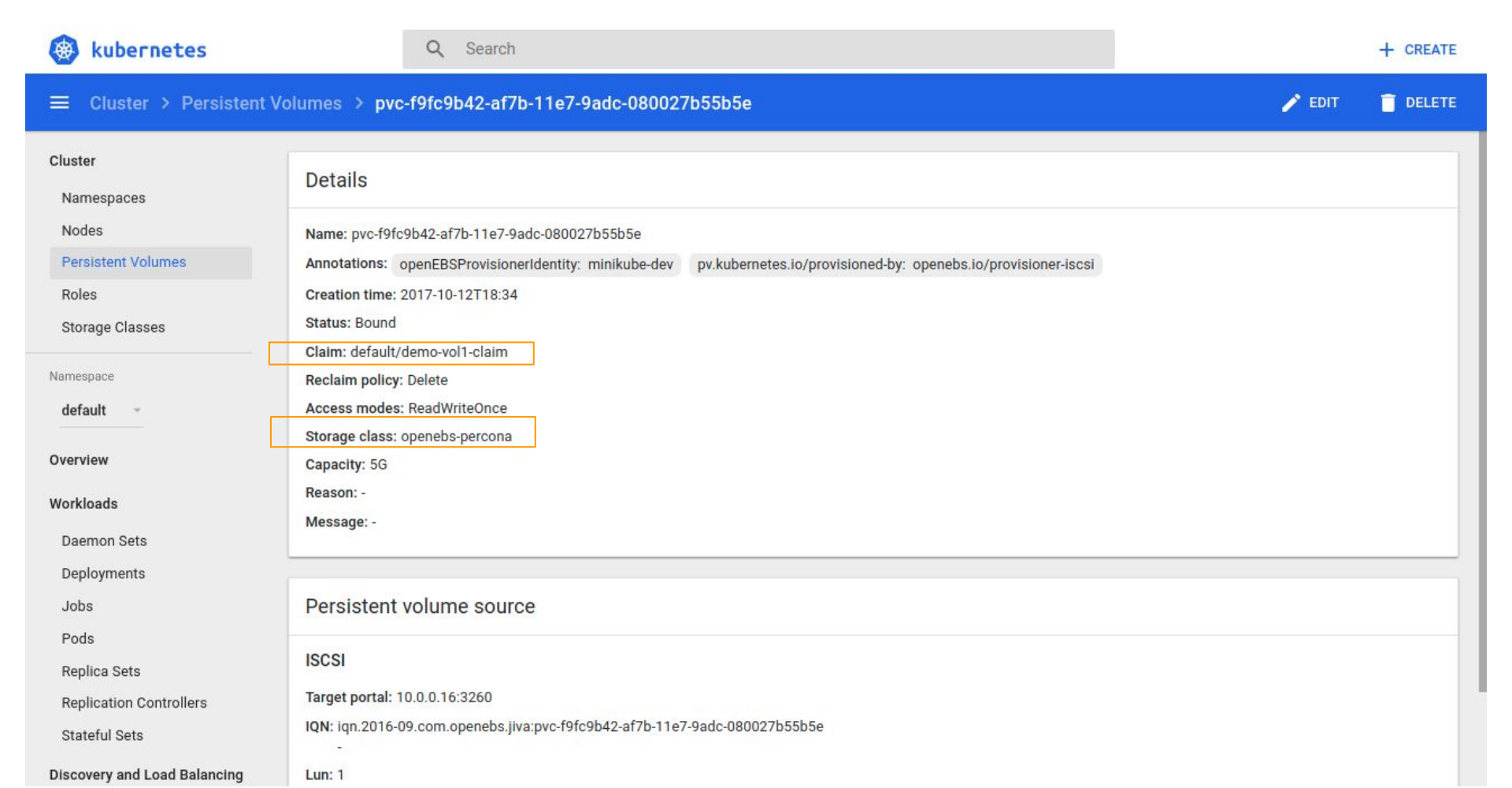
Since:
2017 Sep 14 14:56 IST
(1 hour 21 minutes)

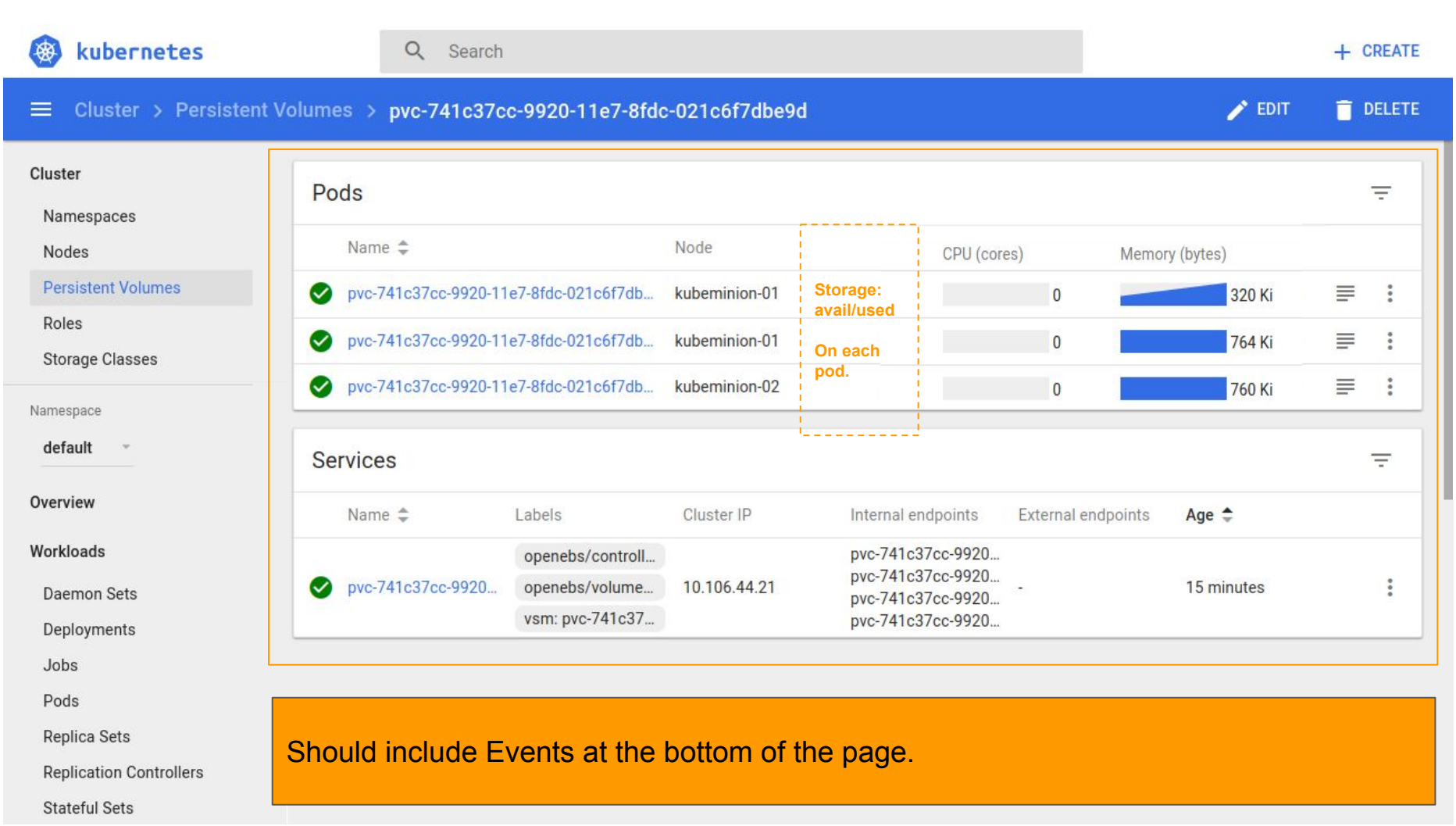
Current Time: 4:17 PM

Storage Capacity



Should include outstanding alerts at the top of the page.


kubernetes
Search
CREATE
Cluster > Persistent Volumes > pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e
EDIT
DELETE
Cluster
Namespaces
Nodes
Persistent Volumes
Roles
Storage Classes
Namespace
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Discovery and Load Balancing
Details
Name: pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e
Annotations: openEBSProvisionerIdentity: minikube-dev
pv.kubernetes.io/provisioned-by: openebs.io/provisioner-iscsi
Creation time: 2017-10-12T18:34
Status: Bound
Claim: default/demo-vol1-claim
Reclaim policy: Delete
Access modes: ReadWriteOnce
Storage class: openebs-percona
Capacity: 5G
Reason: -
Message: -
Persistent volume source
ISCSI
Target portal: 10.0.0.16:3260
IQN: iqn.2016-09.com.openebs.jiva:pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e
Lun: 1


kubernetes
Search
CREATE
Cluster > Persistent Volumes > pvc-741c37cc-9920-11e7-8fdc-021c6f7dbe9d
EDIT
DELETE
Cluster
Namespaces
Nodes
Persistent Volumes
Roles
Storage Classes
Namespace
default
Overview
Workloads
Daemon Sets
Deployments
Jobs
Pods
Replica Sets
Replication Controllers
Stateful Sets
Pods
Name
Node
CPU (cores)
Memory (bytes)
pvc-741c37cc-9920-11e7-8fdc-021c6f7db...
kubeminion-01
0
320 Ki
pvc-741c37cc-9920-11e7-8fdc-021c6f7db...
kubeminion-01
0
764 Ki
pvc-741c37cc-9920-11e7-8fdc-021c6f7db...
kubeminion-02
0
760 Ki
Storage: avail/used
On each pod.
Services
Name
Labels
Cluster IP
Internal endpoints
External endpoints
Age
pvc-741c37cc-9920...
openebs/controll...
openebs/volume...
vsm: pvc-741c37...
10.106.44.21
pvc-741c37cc-9920...
pvc-741c37cc-9920...
pvc-741c37cc-9920...
pvc-741c37cc-9920...
-
15 minutes
Should include Events at the bottom of the page.

kubernetes

Search

+ CREATE

Workloads > Pods > pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e-rep-1715657804-r0bhq

EXEC LOGS EDIT DELETE

Cluster

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

Namespace

default

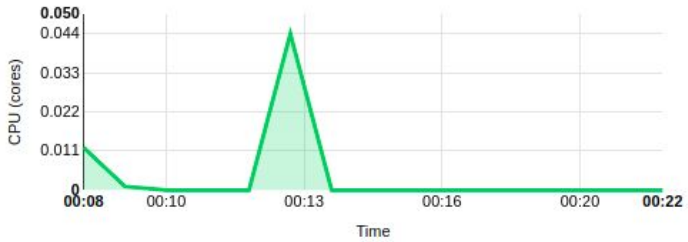
Overview

Workloads

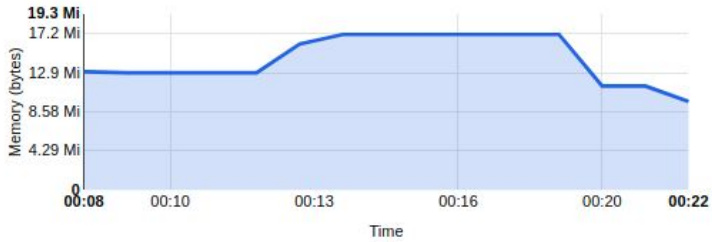
- Daemon Sets
- Deployments
- Jobs
- Pods
- Replica Sets
- Replication Controllers
- Stateful Sets

Discovery and Load Balancing

CPU usage



Memory usage



Details

Name: pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e-rep-1715657804-r0bhq

Namespace: default

Labels: openebs/replica: jiva-replica | pod-template-hash: 1715657804 | vsm: pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e

Annotations: Created by: ReplicaSet pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e-rep-1715657804

Creation time: 2017-10-12T18:34

Status: Running

Network

Node: minikube-dev

IP: 172.17.0.8

Containers

Cluster

Namespaces

Nodes

Persistent Volumes

Roles

Storage Classes

Namespace

default

Overview

Workloads

Daemon Sets

Deployments

Jobs

Pods

Replica Sets

Replication Controllers

Stateful Sets

Discovery and Load Balancing

Containers

pvc-f9fc9b42-af7b-11e7-9adc-080027b55b5e-rep-con

Image: openebs/jiva:0.4.0

Environment variables: -

Commands: launch

Args: replica
--frontendIP
10.0.0.16
--size
5G
/openebs

Conditions

Type	Status	Last heartbeat time	Last transition time	Reason	Message
Initialized	True	-	21 minutes	-	-
Ready	True	-	19 minutes	-	-
PodScheduled	True	-	21 minutes	-	-