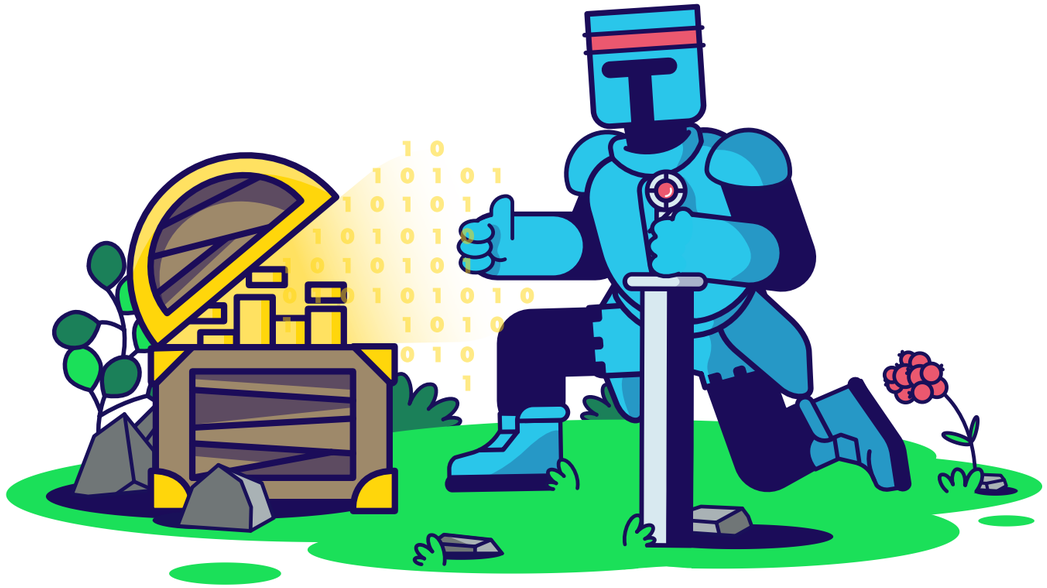
Introducing Rook

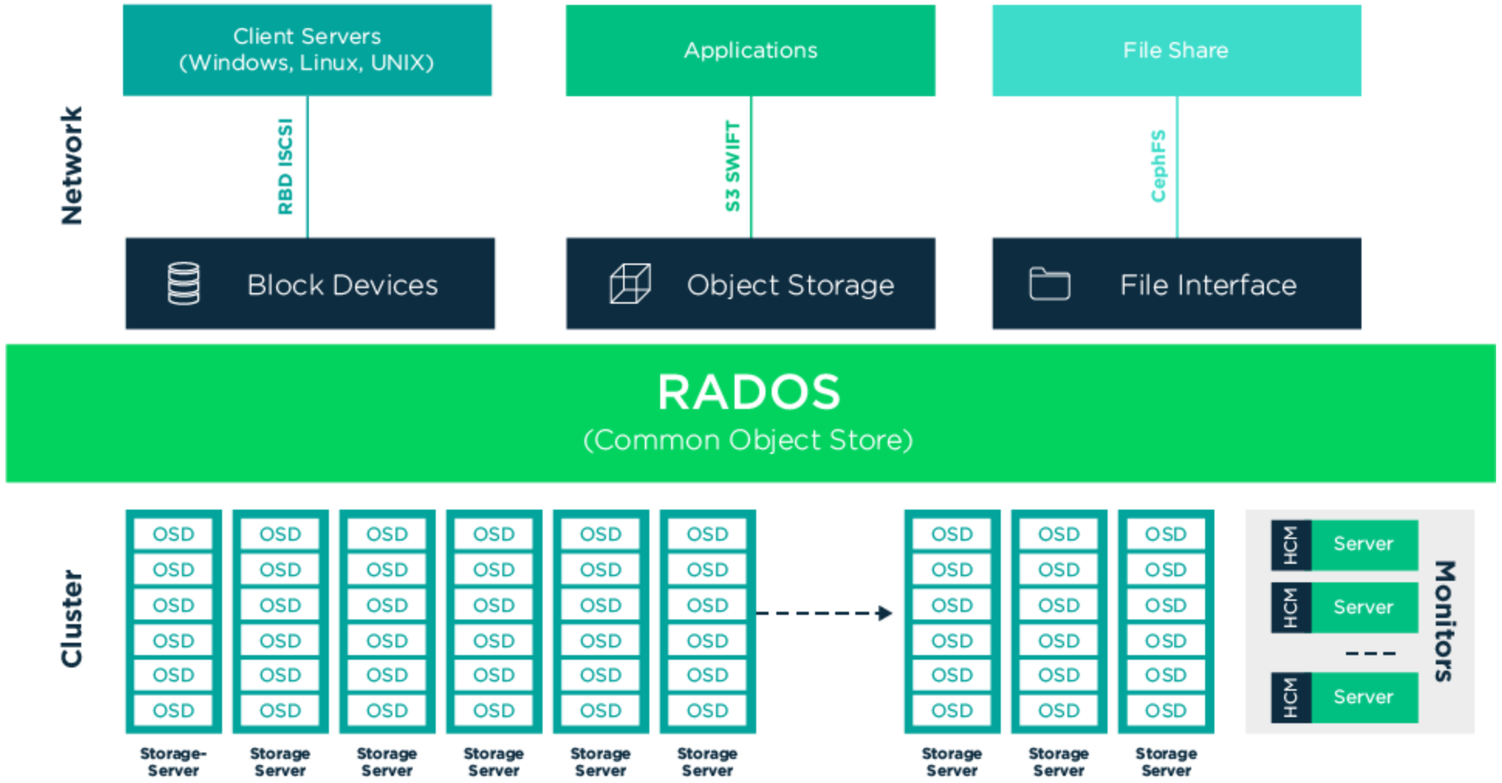


Kristoffer Grönlund <kgronlund@suse.com>

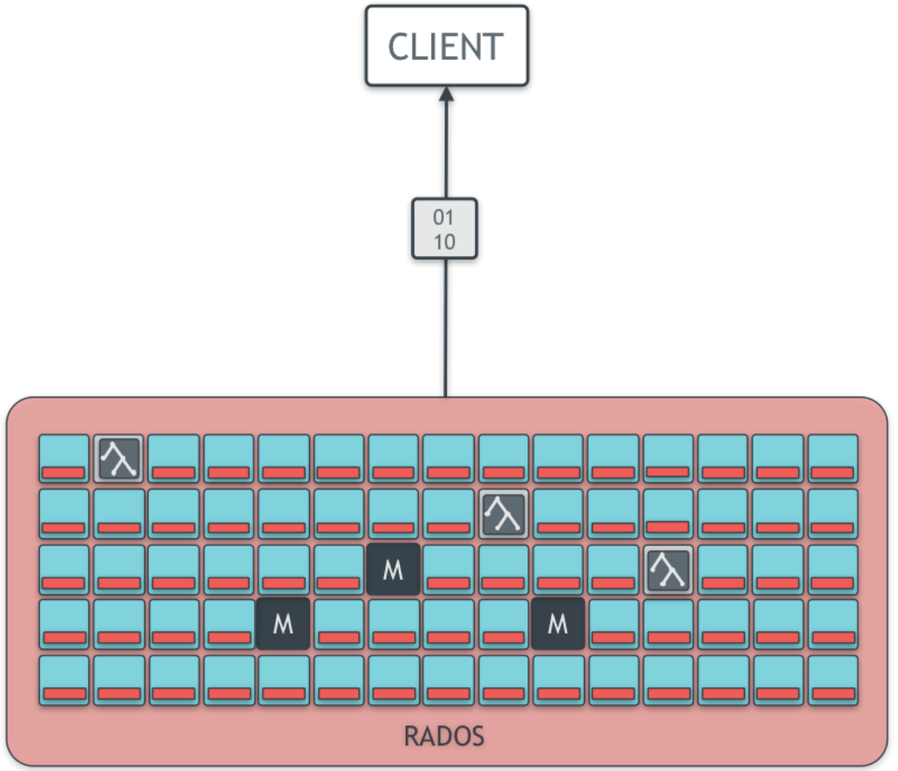
Ceph • **Rook** • Kubernetes

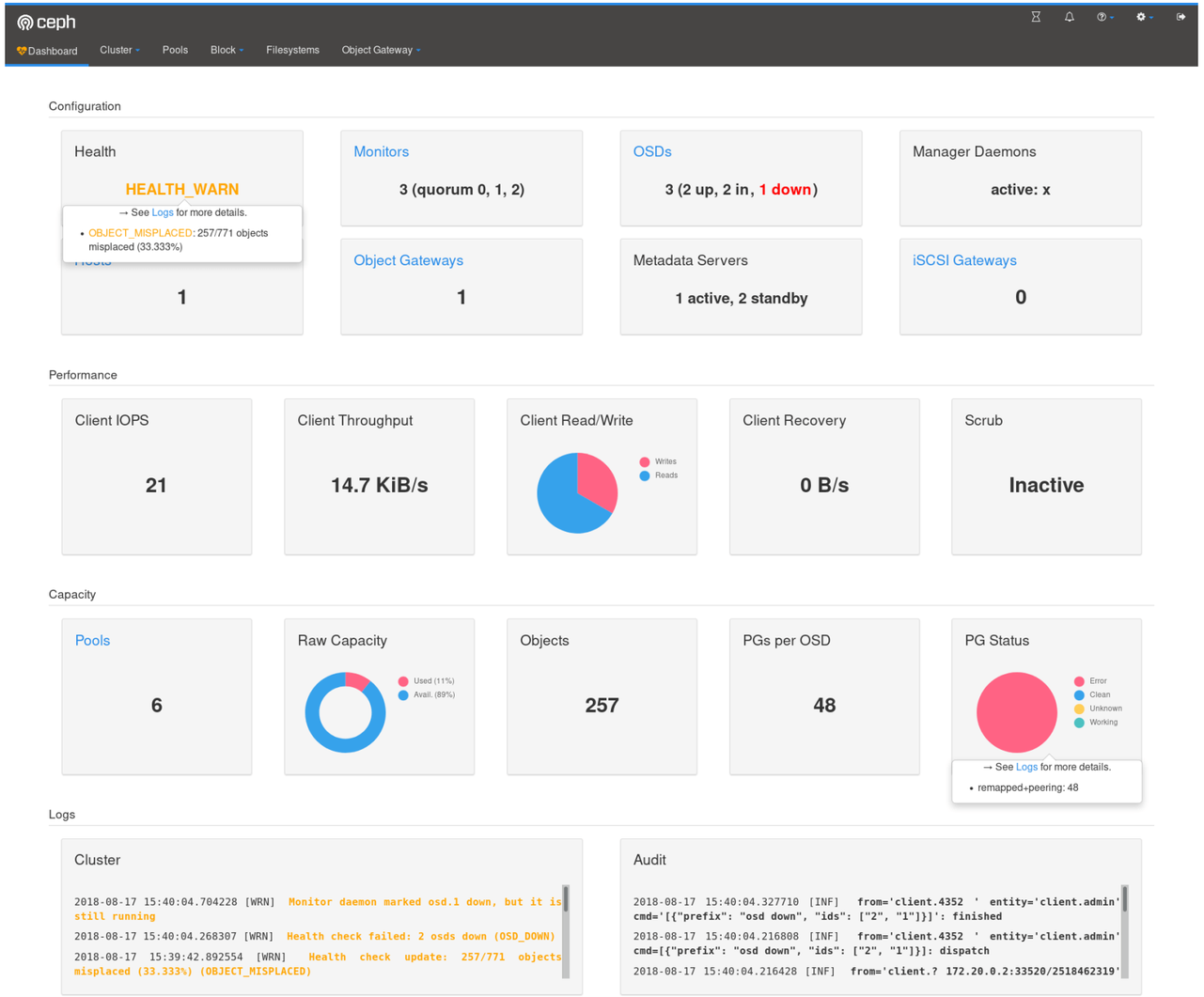
Software Defined Storage

- Open source
- Distributed
- Massively scalable
- Self healing
- Runs on commodity hardware / public cloud


Network
Client Servers
(Windows, Linux, UNIX)
Applications
File Share
RBD ISCSI
S3 SWIFT
CephFS
Block Devices
Object Storage
File Interface
RADOS
(Common Object Store)
Cluster
OSD
Storage-
Server
Storage
Server
HCM
Server
Monitors

No Single Point Of Failure




ceph
Dashboard
Cluster
Pools
Block
Filesystems
Object Gateway
Configuration
Health
HEALTH_WARN
→ See Logs for more details.
• OBJECT_MISPLACED: 257/771 objects misplaced (33.333%)
1
Monitors
3 (quorum 0, 1, 2)
OSDs
3 (2 up, 2 in, 1 down)
Manager Daemons
active: x
Object Gateways
1
Metadata Servers
1 active, 2 standby
iSCSI Gateways
0
Performance
Client IOPS
21
Client Throughput
14.7 KiB/s
Client Read/Write
Writes
Reads
Client Recovery
0 B/s
Scrub
Inactive
Capacity
Pools
6
Raw Capacity
Used (11%)
Avail. (89%)
Objects
257
PGs per OSD
48
PG Status
Error
Clean
Unknown
Working
→ See Logs for more details.
• remapped+peering: 48
Logs
Cluster
2018-08-17 15:40:04.704228 [WRN] Monitor daemon marked osd.1 down, but it is still running
2018-08-17 15:40:04.268307 [WRN] Health check failed: 2 osds down (OSD_DOWN)
2018-08-17 15:39:42.892554 [WRN] Health check update: 257/771 objects misplaced (33.333%) (OBJECT_MISPLACED)
Audit
2018-08-17 15:40:04.327710 [INF] from='client.4352 ' entity='client.admin' cmd='[{"prefix": "osd down", "ids": ["2", "1"]}]': finished
2018-08-17 15:40:04.216808 [INF] from='client.4352 ' entity='client.admin' cmd=[{"prefix": "osd down", "ids": ["2", "1"]}]: dispatch
2018-08-17 15:40:04.216428 [INF] from='client.? 172.20.0.2:33520/2518462319'

Kubernetes

Concepts

- Pod
- Service
- Namespace
- Controller
- Volume
- Custom Resource Definition (CRD)

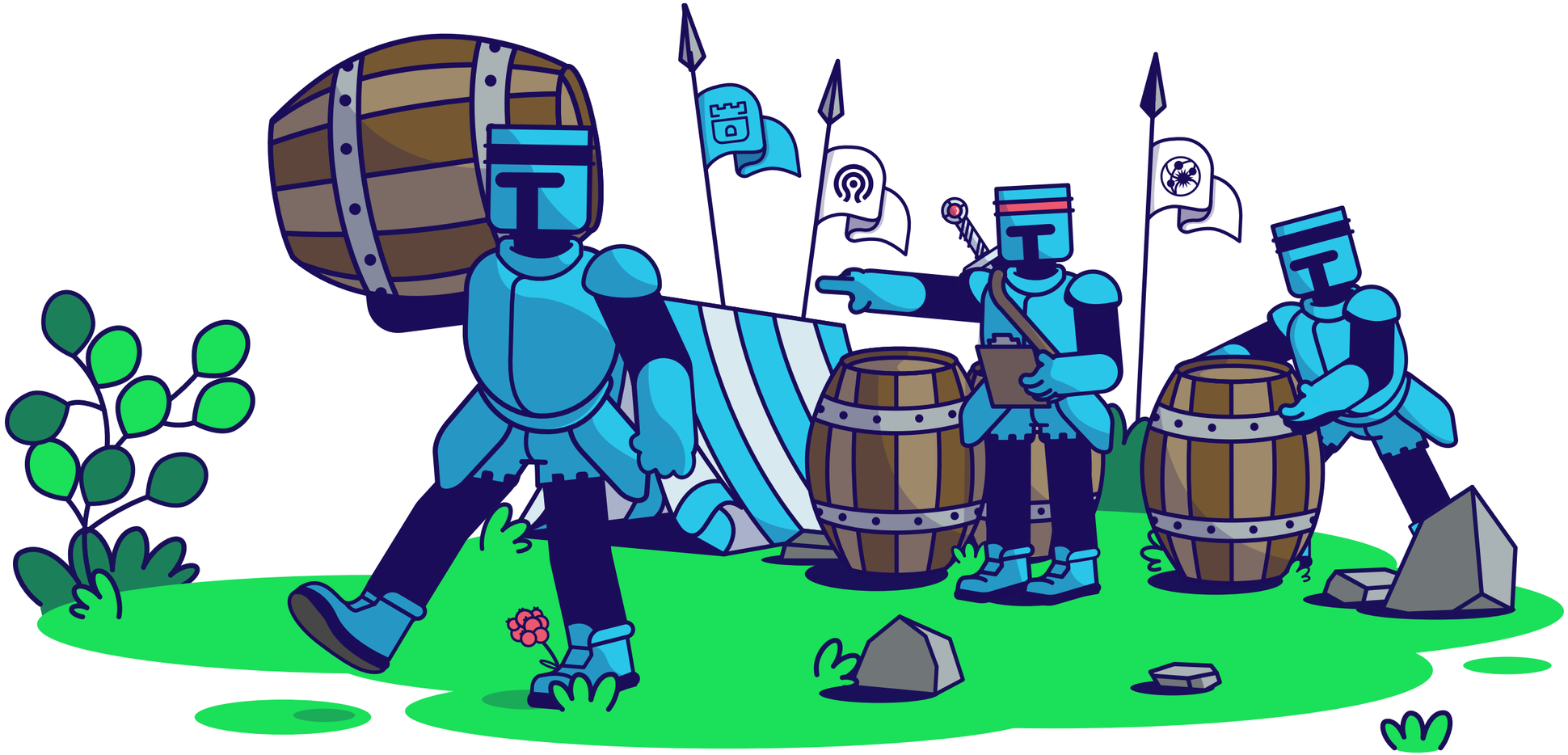
Kubernetes Storage Story

1. Volume Plugins
2. FlexVolume *
3. CSI *

Challenges

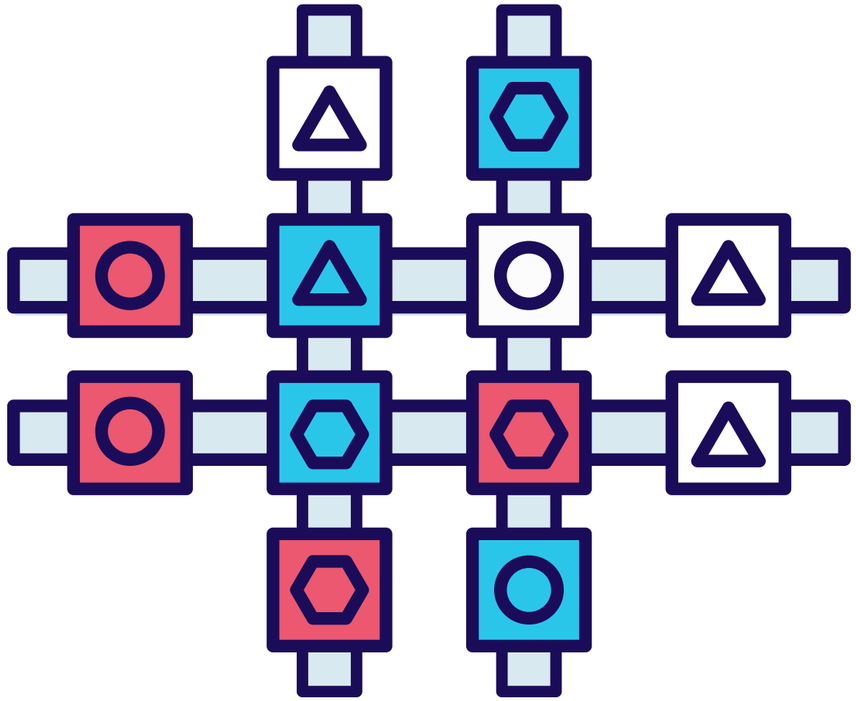
External Storage / Cloud Vendor

- Vendor lock-in
- Portability
- Connectivity
- Deployment
- Administration



Multiple storage providers

- **Ceph**
- **EdgeFS**
- CockroachDB
- Cassandra
- NFS
- Yugabyte DB
- Apache Ozone



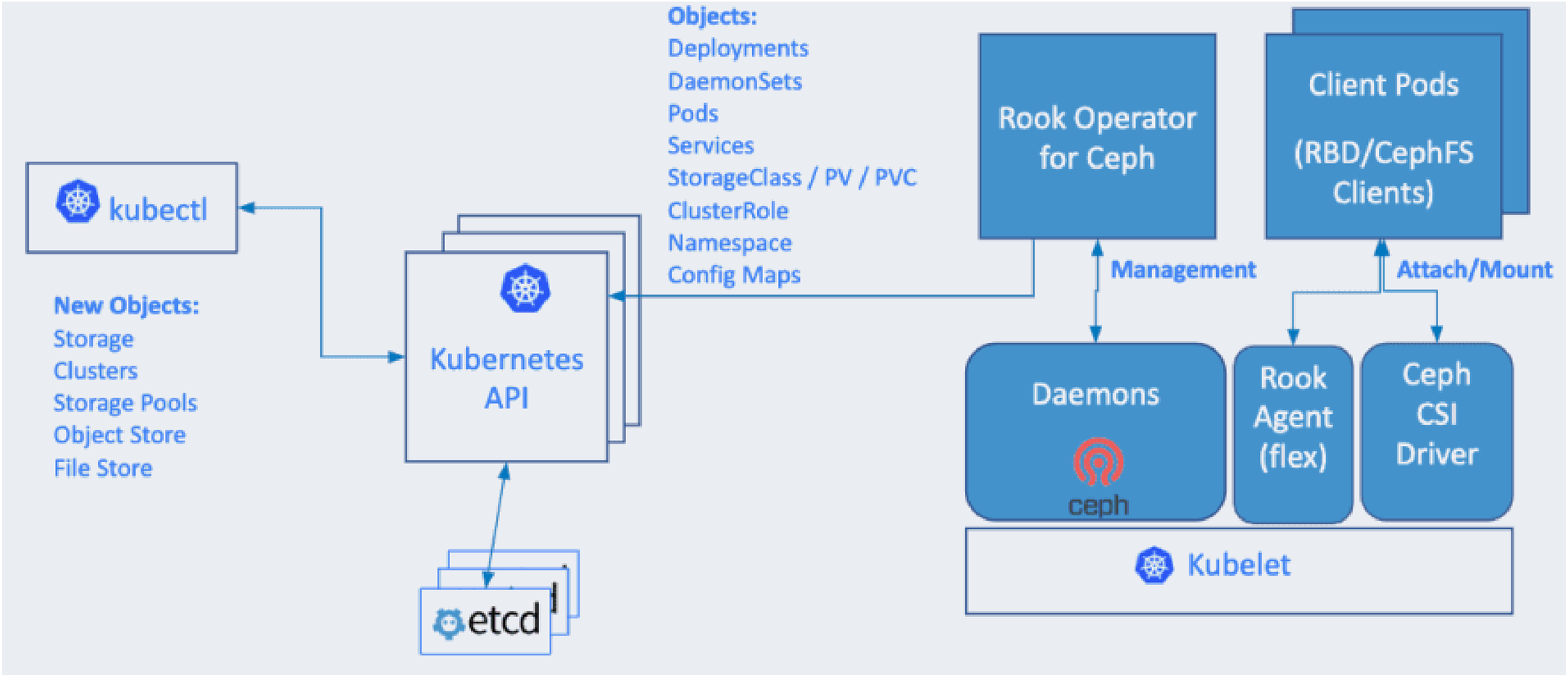
Rook Operator

- Bootstrap and Monitor
- Manages Ceph
- Creates Agents

Rook Agents

- Runs on all nodes
- Handles storage operations

- Persistent Volume (PV)
- Persistent Volume Claim (PVC)
- Storage Classes


Objects:
Deployments
DaemonSets
Pods
Services
StorageClass / PV / PVC
ClusterRole
Namespace
Config Maps
kubectl
New Objects:
Storage
Clusters
Storage Pools
Object Store
File Store
Kubernetes API
etcd
Rook Operator for Ceph
Client Pods
(RBD/CephFS Clients)
Management
Attach/Mount
Daemons
ceph
Rook Agent (flex)
Ceph CSI Driver
Kubelet

Use Case: Pure Ceph Cluster

- Dedicated K8S cluster for running Ceph
- Can serve multiple K8S application clusters
- Not common

Use Case: Shared Cluster

- One partitioned K8S cluster
- Dedicated storage nodes for Ceph
- Dedicated compute nodes for Apps

Use Case: Unified Cluster

- All nodes run both Ceph and workloads
- Most common

Use Case: External Cluster

- Rook as interface to external Ceph cluster
- Added in Rook v1.1

Try it out

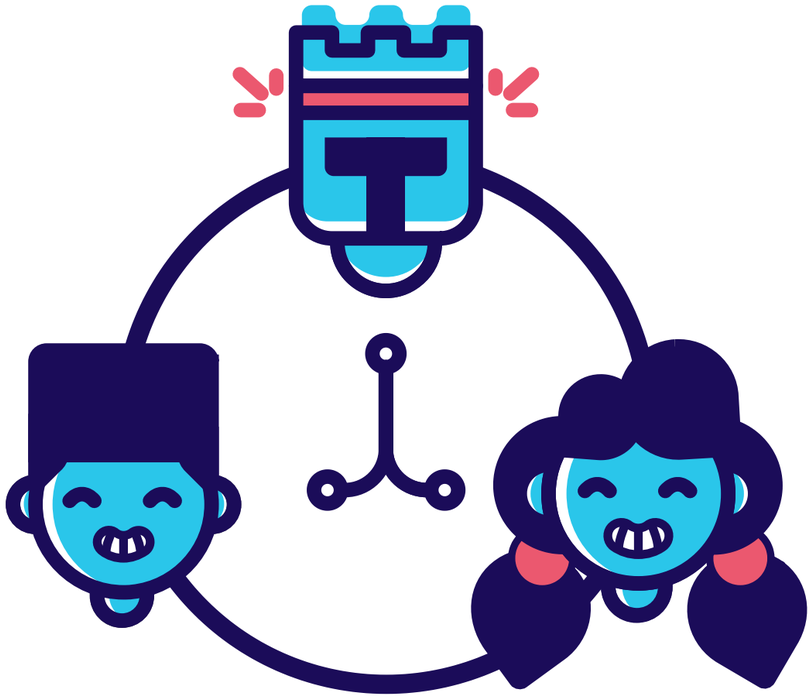
```
# cat object.yaml
apiVersion: ceph.rook.io/v1
kind: CephObjectStore
metadata:
  name: my-store
  namespace: rook-ceph
spec:
  metadataPool:
    failureDomain: host
    replicated:
      size: 3
  dataPool:
    failureDomain: host
    erasureCoded:
      dataChunks: 2
      codingChunks: 1
  preservePoolsOnDelete: true
  gateway:
    type: s3
    sslCertificateRef:
    port: 80
    securePort:
    instances: 1
```

```
# Create the object store
kubectl create -f object.yaml

# To confirm the object store is configured, wait for the rgw pod to start
kubectl -n rook-ceph get pod -l app=rook-ceph-rgw
```

rook.io

Conclusion



suse.com/storage

- SUSE Storage 6
  - Technology Preview (feedback welcome)
- SUSE Storage 7
  - Rook on CaaSP as supported stack