

Rook: Storage for Kubernetes

Blaine Gardner, IBM
Parth Arora, IBM



Raise your hand if…

- Are you here to learn about Rook for the first time?
- Have you experimented with Rook?
- Have you deployed Rook in production?



Agenda

- Introduction to Rook and Ceph
- Maximizing Storage Efficiency with EC
- Recent features
    - Ceph-CSI and its operator
    - OSD Replacement
    - Ceph Umbrella
    - NVMe-oF
- CephX authorization CVE

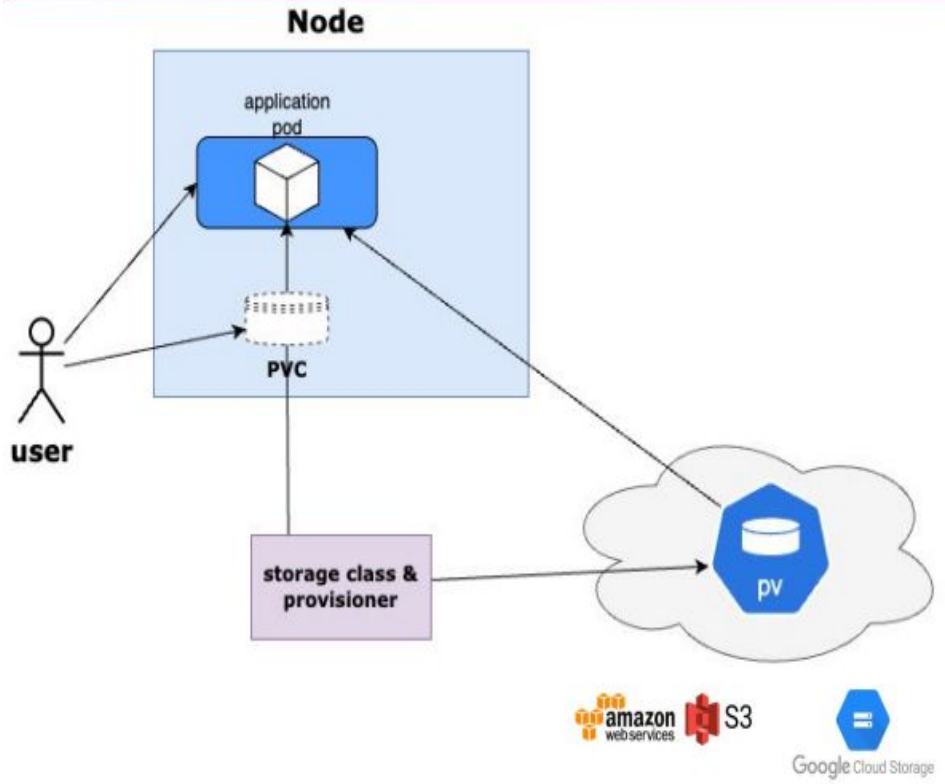
KubeCon CloudNativeCon
North America 2026

Introduction to Rook



Where did Rook begin?

- Storage is commonly provided by cloud providers.
- What about storage in your datacenter?
- Why not manage storage as any other K8s application?
- What about storage in your datacenter?

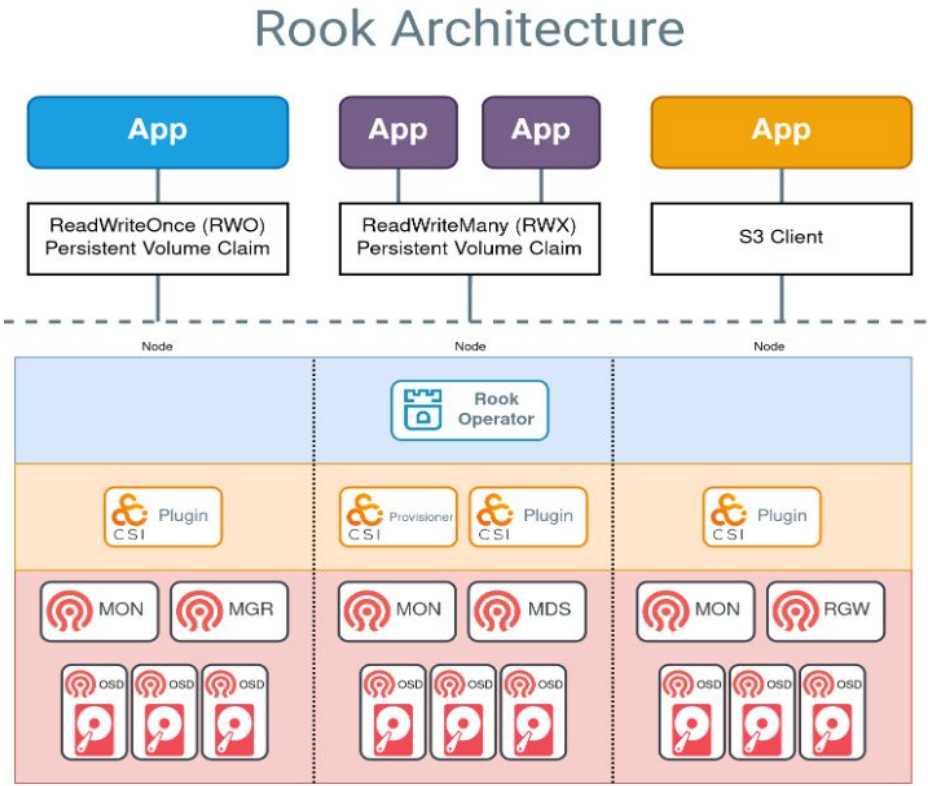


What is Rook?

- Brings Ceph storage into your Kubernetes cluster
- Manages Ceph storage with an operator and CRDs
- Automated deployment, configuration, upgrades
- Allows apps to consume storage like any other K8s storage
  - Storage Classes, PVCs
- Open Source (Apache 2.0)
- Happy 10th Birthday!
  - Created November 2016

Architectural Layers

- Rook
  - **Deploys** and **manages** Ceph
- CSI
  - Ceph CSI dynamically **provisions** and **mounts** storage for user Pods
- Ceph
  - **Data** layer



What is Ceph?

- Distributed Software-Defined Storage solution: https://ceph.io/
- All-in-one open-source storage platform

RBD	Block storage	RWO
CephFS	Shared File System	RWX
RGW	Object storage	S3 Protocol



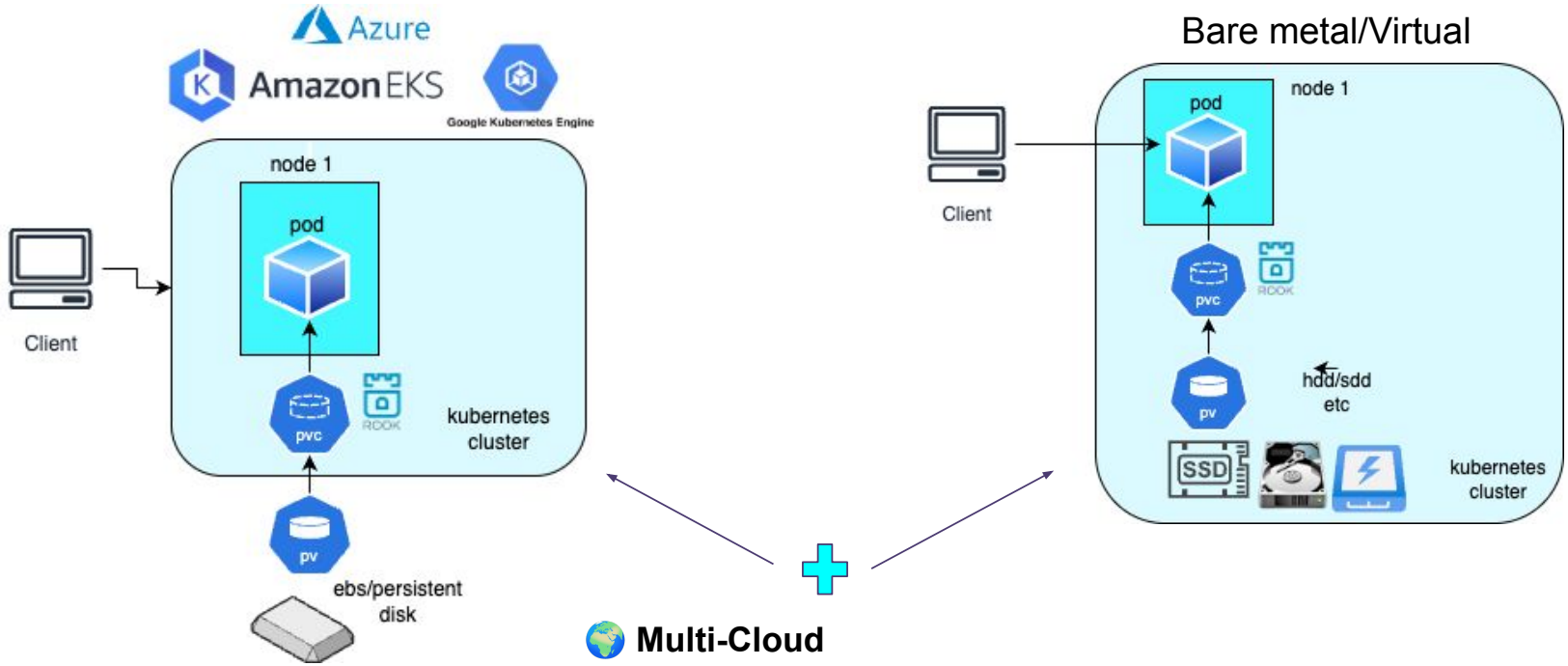
Why Ceph

- Scale up to many Petabytes in a single cluster!
- Upstream users with
    - 2K+ OSDs, 10s PiBs
- Add capacity without downtime
    - Scale up and out
- Proven history of enterprise adoption and support
- Thin provisioning → Efficient storage use
- Data is sharded across AZs, racks, nodes, disks
- Ceph designed to be consistent, not eventually consistent

Installation Environments

- Anywhere Kubernetes runs!
- Bare Metal (Use local SSDs/HDDs)
- Cloud Environments (Use cloud block storage (EBS, Persistent Disks, etc.))
- Install Rook via:
    - Helm charts
    - Example manifests for many configurations

Installation Environments

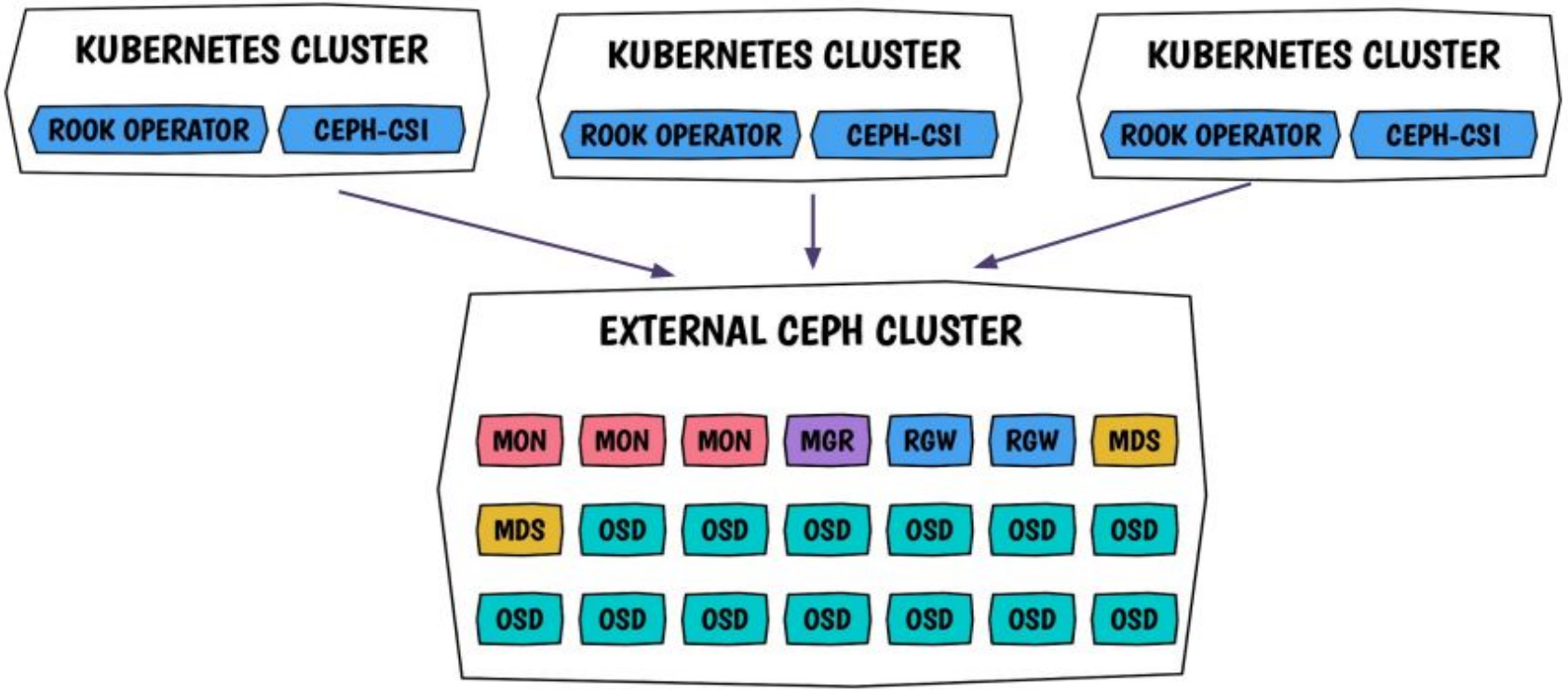


Why Deploy Rook in the Cloud?

- Overcomes limitations of cloud providers
  - Storage not available across AZs
  - Number of PVs per node
  - Poor performance of small PVs
- Allows consistency across clouds/on-prem



External Cluster: Multi Tenant Support



KubeCon CloudNativeCon North America 2026

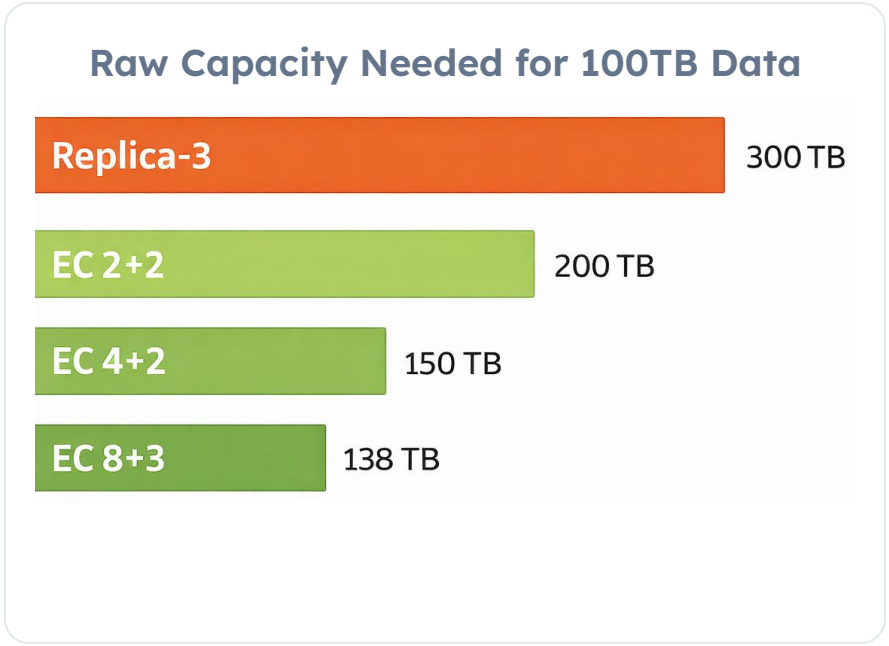
Maximise Storage Efficiency with: Erasure Coding



Why Erasure Coding

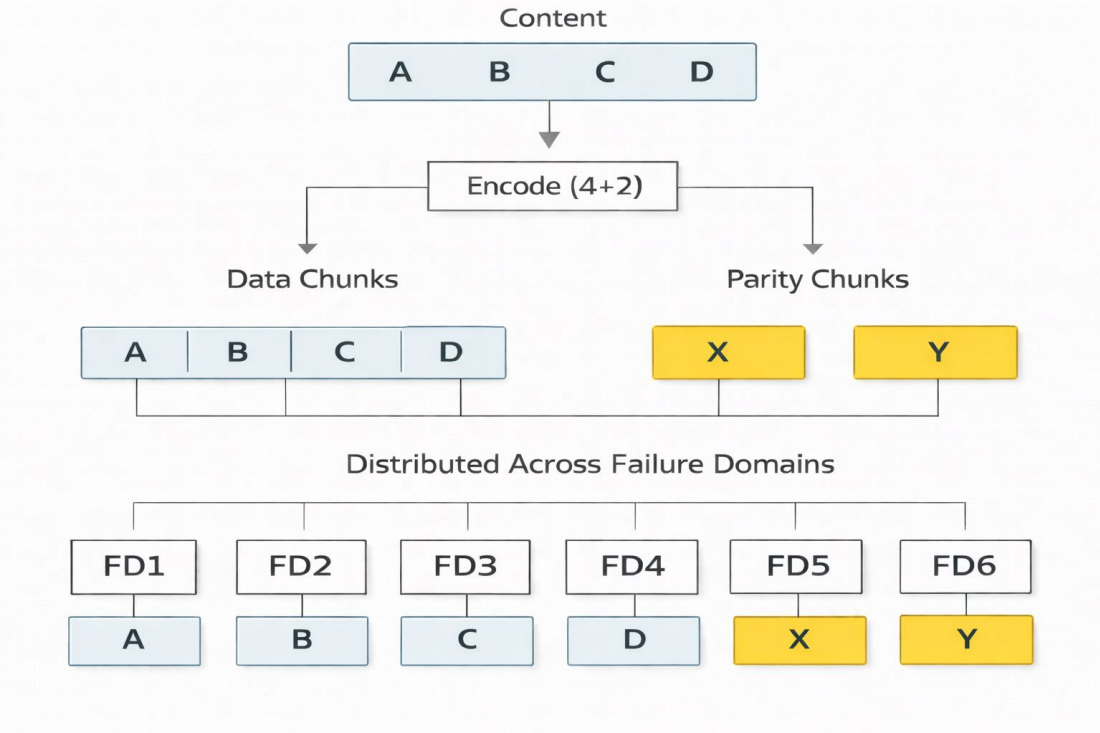
Why make 3 copies when maths can do the job?

- Most Ceph clusters use **replica-3** for durability meaning **3x storage requirement**
- Erasure Coding provides **similar durability** with significantly lower storage overhead
- **Same durability, ~50% less storage (EC 4+2 vs Replica-3)**



EC Overview

Data is split into chunks and extra parity chunks are generated for recovery.



EC Profile: 4+2

k = 4 data chunks
m = 2 coding chunks

Minimum failure domains:

k + m = 6

Failure tolerance:

m = 2

Perf Comparison: EC vs Replica-3

With Large Read & Write, EC performance is almost identical to Replica-3

KubeCon CloudNativeCon North America 2026

Recent Features



Ceph-CSI

- Rook v1.20 fully transitioned to Ceph-CSI's new(ish) operator
  - Allows more flexibility and first-class CRD configuration for drivers
    - RBD, CephFS, NFS, NVMe-oF
  - Required transition for many users

- Rate limiting client I/O per Pod added in Ceph-CSI v3.18.0
  - cgroup v2 limiting via **`io.max`** enforcement
  - Configured via VolumeAttributesClass

Ceph Messenger v2 Default

- “Messenger” is Ceph’s authentication protocol
- Messenger v2 (msgr2) is the latest, most performant, and most secure
- Allows encryption and compression on the wire (disabled by default)
- Requires Linux kernel 5.11+ (over 5 years old)
- Enabled by default since Rook v1.21

OSD Replacement

- OSDs can be replaced with more efficient backfill
  - Avoids unnecessary data movement prior to replacing the disk
- OSDs sharing a metadata device
  - A single failed data disk can be replaced
  - No need to replace all OSDs on the same metadata device

Ceph Umbrella

- Rook v1.21 supports Ceph Umbrella (released Oct 28)
- Erasure-coded pool improvements
  - Read performance improved, comparable to replicated
  - Write performance improved, better than replication for large block
  - No more need for an additional replicated pool for metadata

- Coming in Ceph Vampire next year
  - Automated pool conversion from replicated to erasure-coded

Biology Class

Flapjack Umbrella Octopus



Actually orange

Vampire Squid



Bioluminescent skin

NVMe-oF Gateway (experimental)

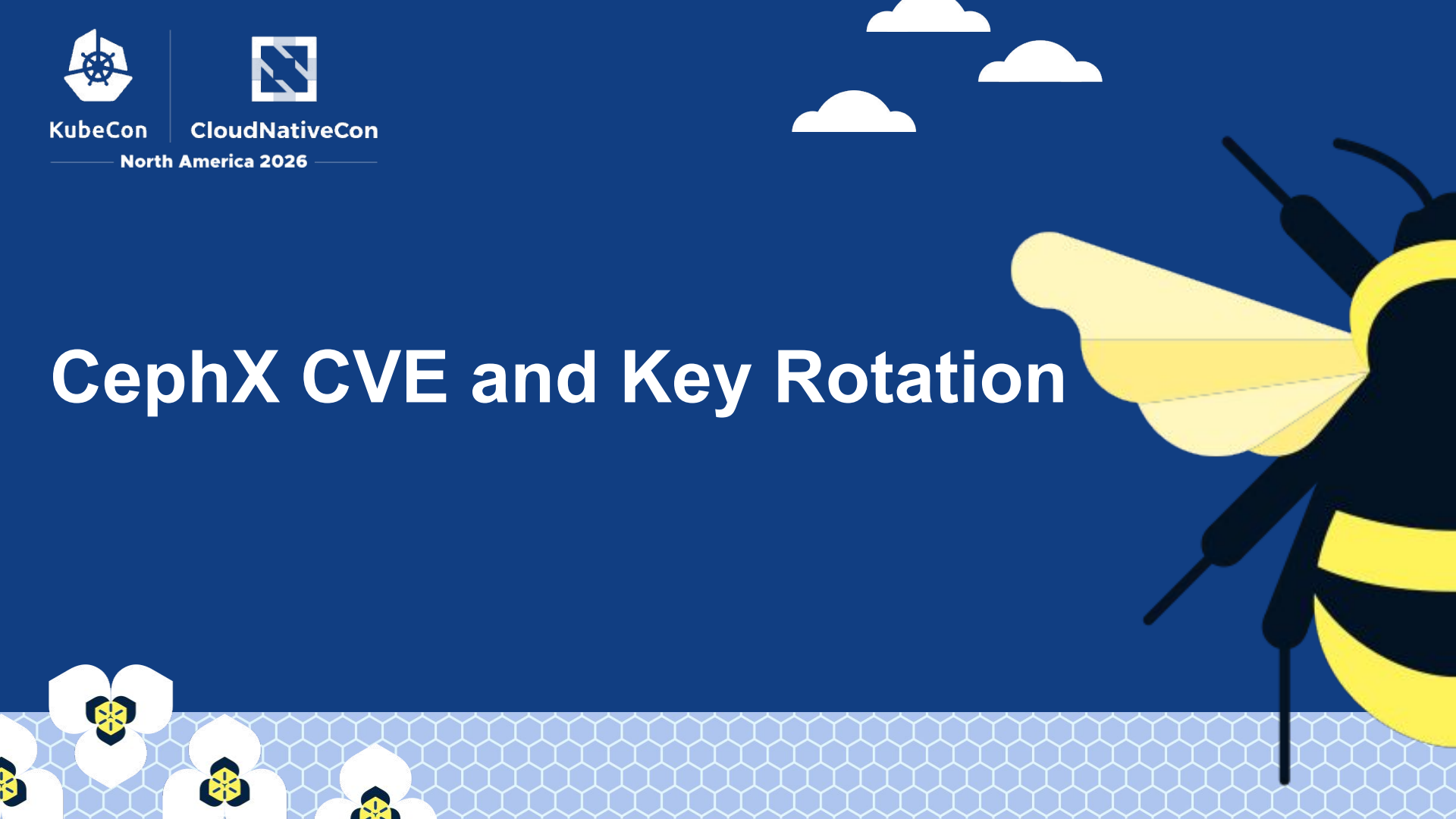
- NVMe over Fabrics
- Delivers low-latency, high-throughput block storage access over existing network infrastructure
- Ceph RBD (block) volume access via the NVMe/TCP protocol
- External clients outside of Kubernetes can connect to Ceph block storage using standard NVMe-oF initiators



Two Node Clusters (experimental)

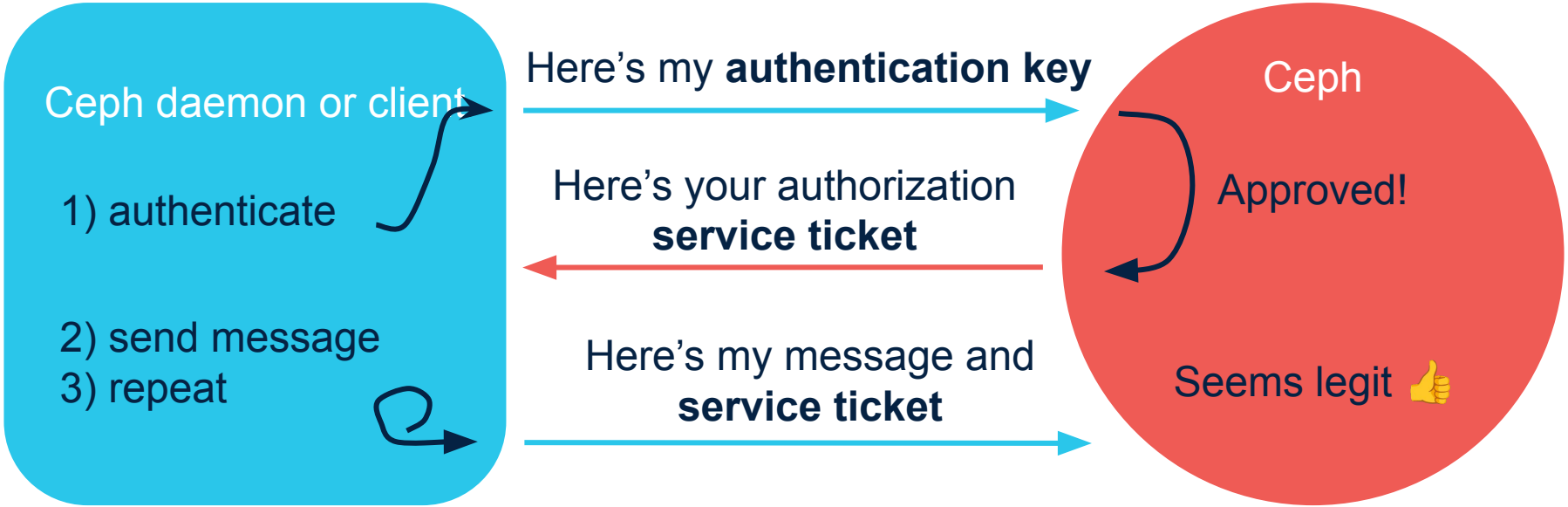
- Edge clusters with only two nodes
- Requires Pacemaker for node failure handling
- A floating mon can move between nodes (using DRBD in experiments)
- Parth presented at Kubecon India 2026 on this topic (QR/link below)

https://www.youtube.com/watch?v=Cw8sFDdNu9Y



CephX CVE and Key Rotation

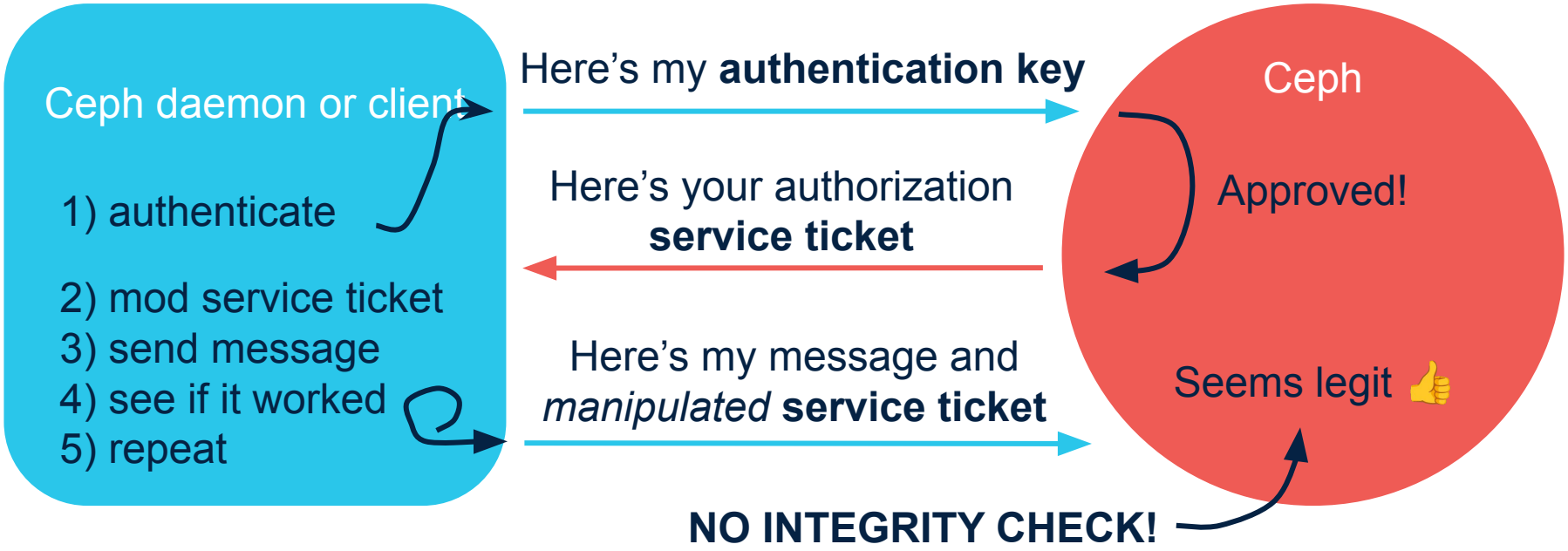
CephX Overview



To the client, service ticket is opaque blob

Ticket validated by Mon, Mgr, OSD, MDS daemons

CVE-2025-30156 Overview



CVE Resolution Overview

- Ceph introduced new AES256K CephX key type for both authentication keys and service tickets
- Mon, Mgr, OSD, and MDS daemons must upgrade to support service tickets with new key type
- After, service tickets now use integrity checks, CVE is resolved
- Clients that continue to use AES authentication keys will still be secure due to upgraded service tickets (FAQ)
- Ceph recommends upgrading all keys to latest AES256K type when possible, without specific urgency

Resolution for Rook Users

- Rook can automatically rotate CephX keys when initiated by user
- Every daemon whose keys can be rotated safely without PVC or client downtime is done so automatically afterwards
- Four required daemons and Ceph admin key are included
- Not rotated (also not critical):
  - Ceph-CSI key rotation requires Linux Kernel 7.0+ support
  - CephClient key and RBD mirror peer key rotations require updated client-side Ceph library

Steps for Rook Users

- Update Rook to latest v1.19, v1.20, or v1.21
- Apply this patch to CephCluster

```
spec:
  cephVersion:
    image: quay.io/ceph/ceph:v20.2.4-20260818 # or >
  security:
    cephx:
      daemon:
        keyRotationPolicy: KeyGeneration
        keyGeneration: 2 # must be >than current keyGeneration
```

blog.rook.io

- Wait for update to finish, and done
- Health warnings (not errors) are informational and can be muted

Survey Form

KubeCon CloudNativeCon North America 2026



KubeCon CloudNativeCon
North America 2026

Questions?

Website + Docs	https://rook.io
Slack	https://slack.rook.io
X	@rook_io



KubeCon CloudNativeCon
North America 2026

Appendix

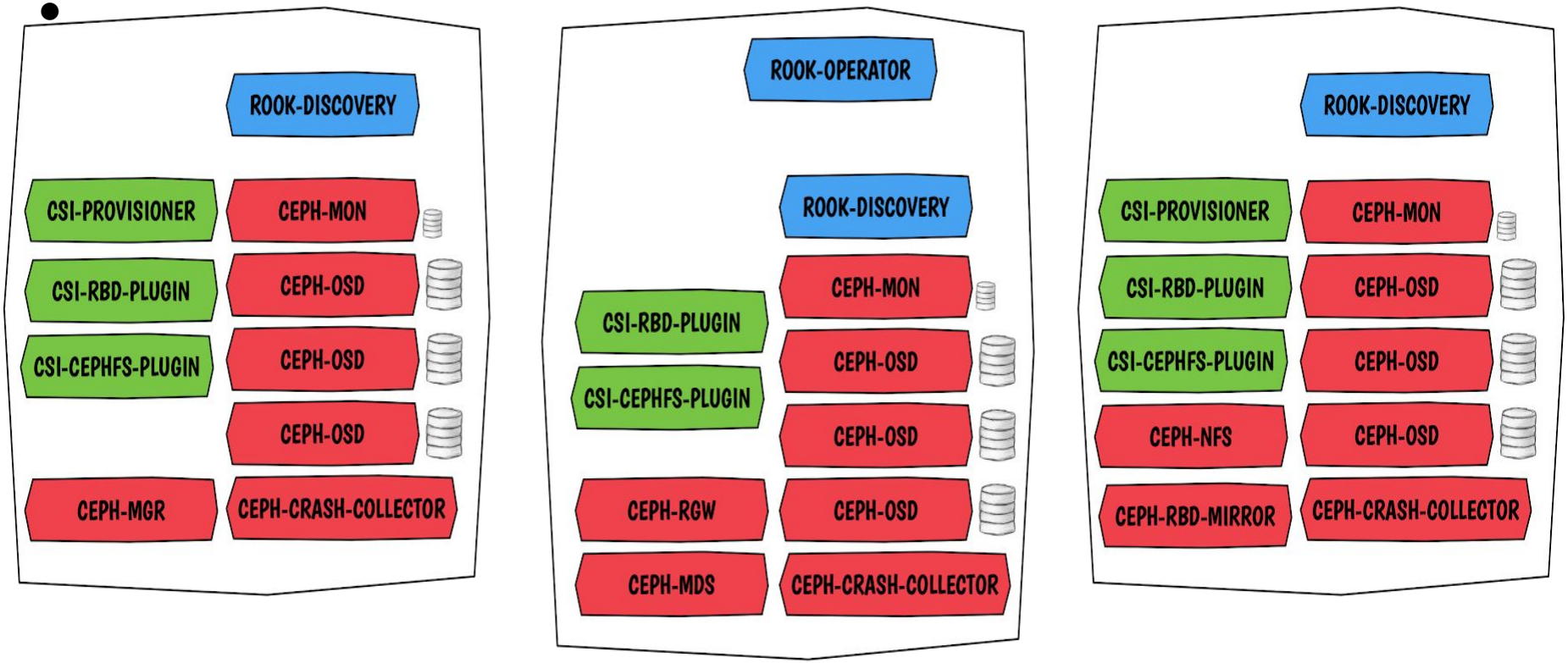


KubeCon CloudNativeCon
North America 2026

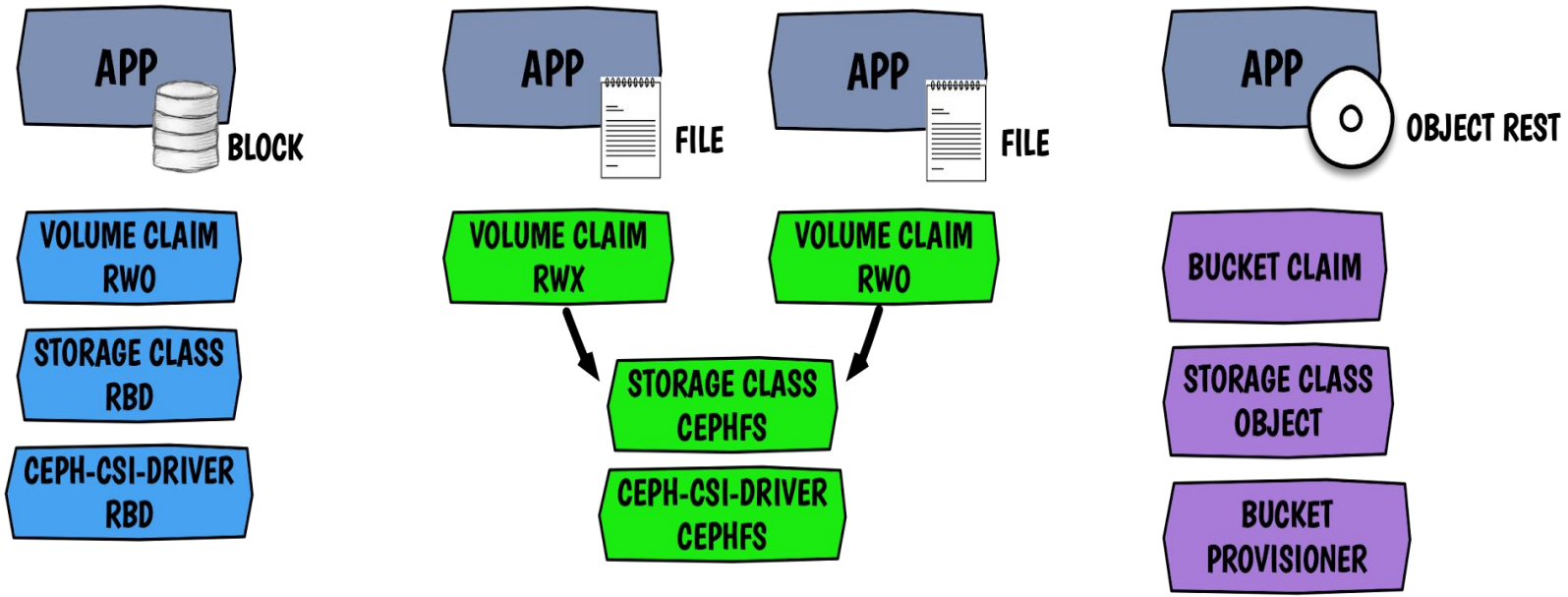
Survey From



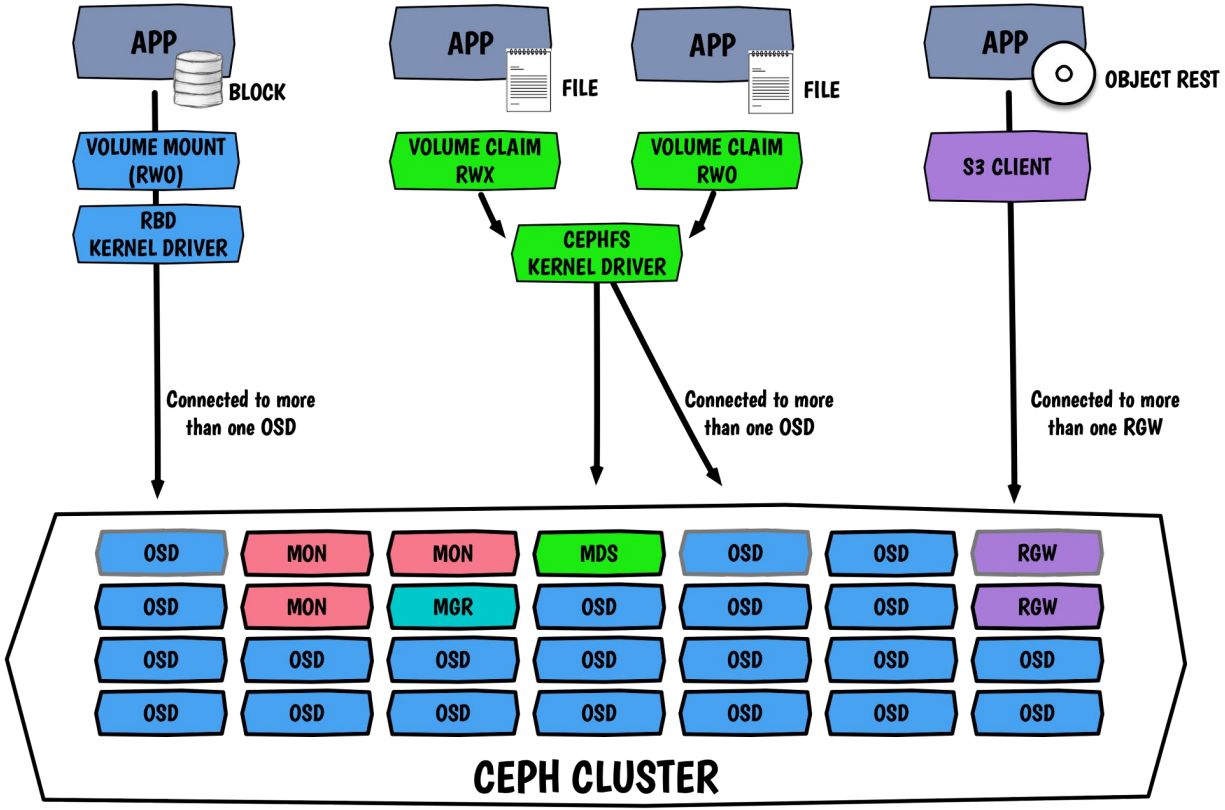
Rook Pods


ROOK-DISCOVERY
CSI-PROVISIONER
CEPH-MON
CSI-RBD-PLUGIN
CEPH-OSD
CSI-CEPHFS-PLUGIN
CEPH-OSD
CEPH-OSD
CEPH-MGR
CEPH-CRASH-COLLECTOR
ROOK-OPERATOR
ROOK-DISCOVERY
CSI-RBD-PLUGIN
CEPH-MON
CSI-CEPHFS-PLUGIN
CEPH-OSD
CEPH-OSD
CEPH-RGW
CEPH-OSD
CEPH-MDS
CEPH-CRASH-COLLECTOR
ROOK-DISCOVERY
CSI-PROVISIONER
CEPH-MON
CSI-RBD-PLUGIN
CEPH-OSD
CSI-CEPHFS-PLUGIN
CEPH-OSD
CEPH-NFS
CEPH-OSD
CEPH-RBD-MIRROR
CEPH-CRASH-COLLECTOR

CSI Provisioning

- 

Ceph Data Path

- 

Survey Form

KubeCon CloudNativeCon North America 2026

