

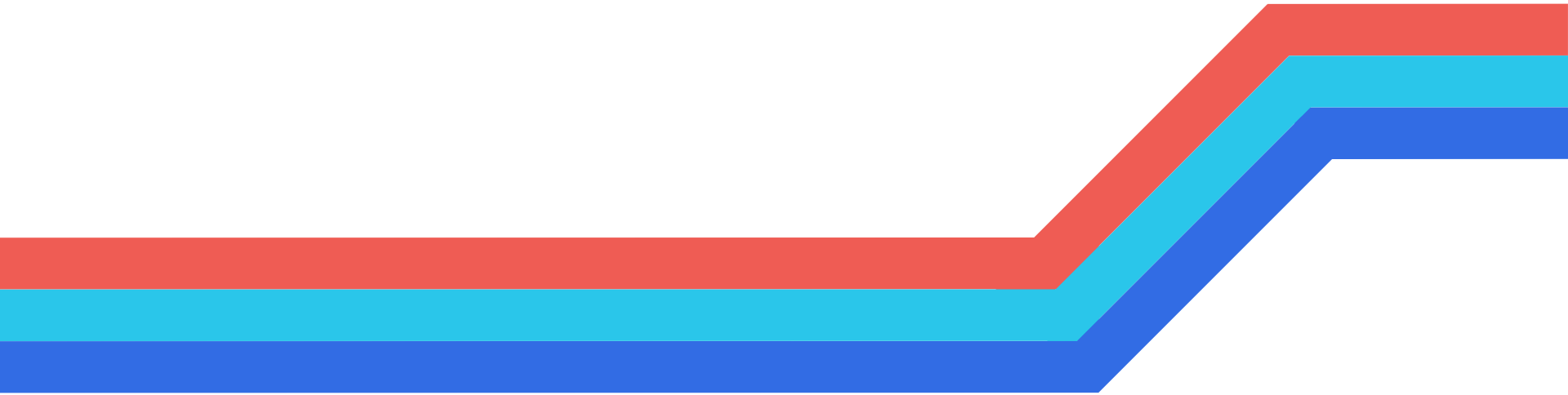
Rook: Storage for Kubernetes

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
- Erasure Coding
- Recent features
    - Ceph-CSI and its operator
    - OSD Replacement
    - Ceph Umbrella
    - NVMe-oF
    - Security Processes
- CephX authorization CVE

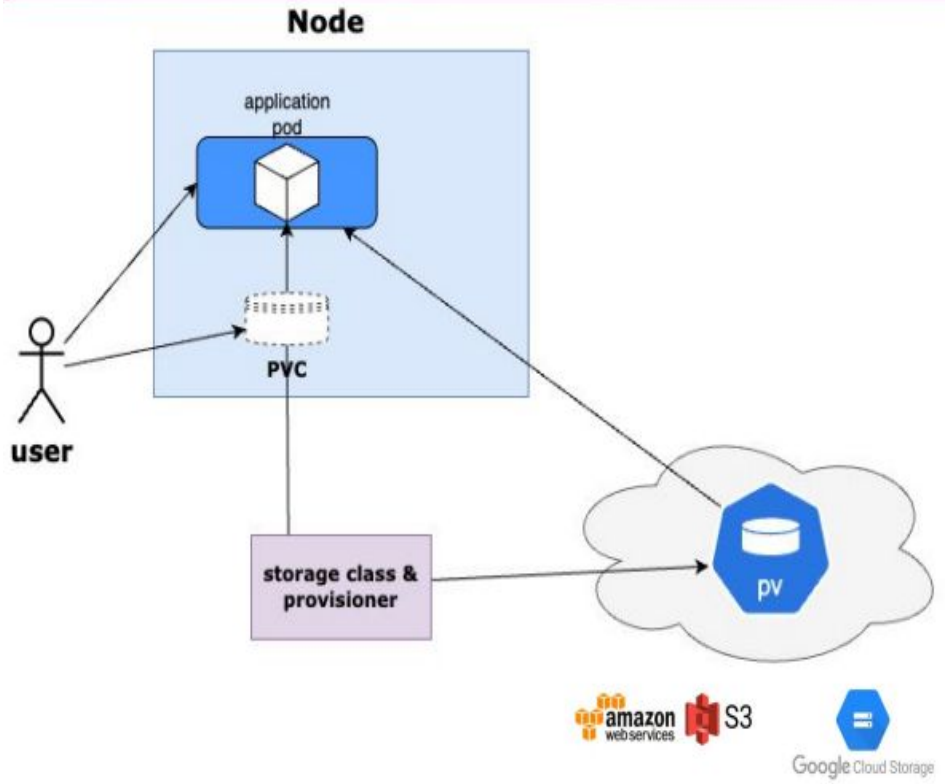
KubeCon CloudNativeCon
North America 2026

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?

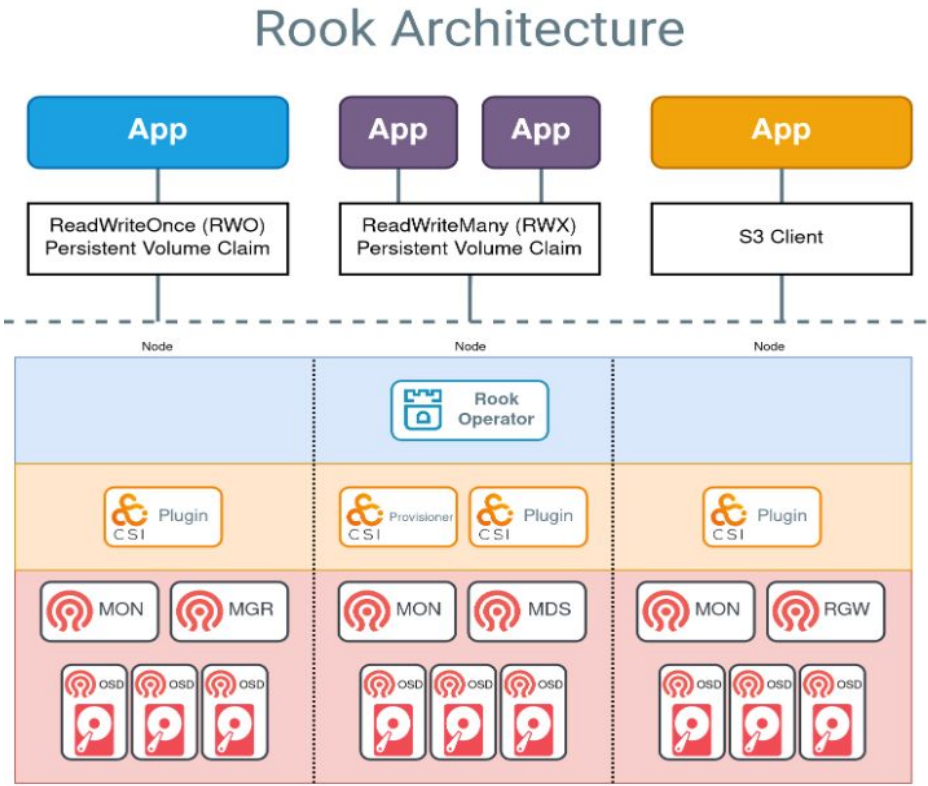


What is Rook?

- 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



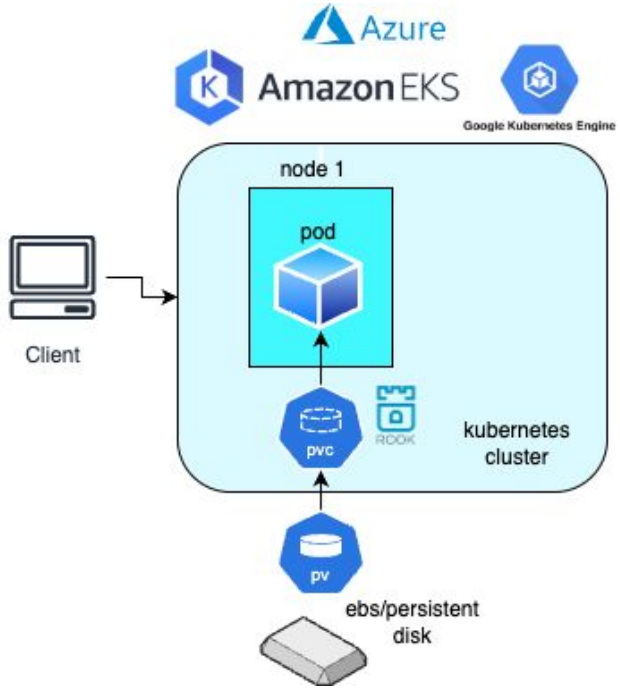
Why Ceph

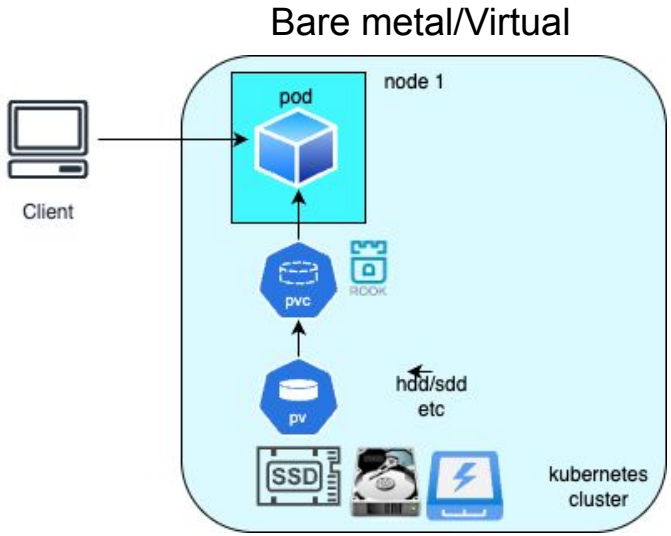
- Scale up to many Petabytes in a single cluster!
- Upstream users with
    - 2K+ OSDs, 10s PiBs
- Add capacity without downtime
    - Scale up and out
- Proven history of enterprise adoption and support
- Thin provisioning → Efficient storage use
- Data is sharded across AZs, racks, nodes, disks
- Ceph 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



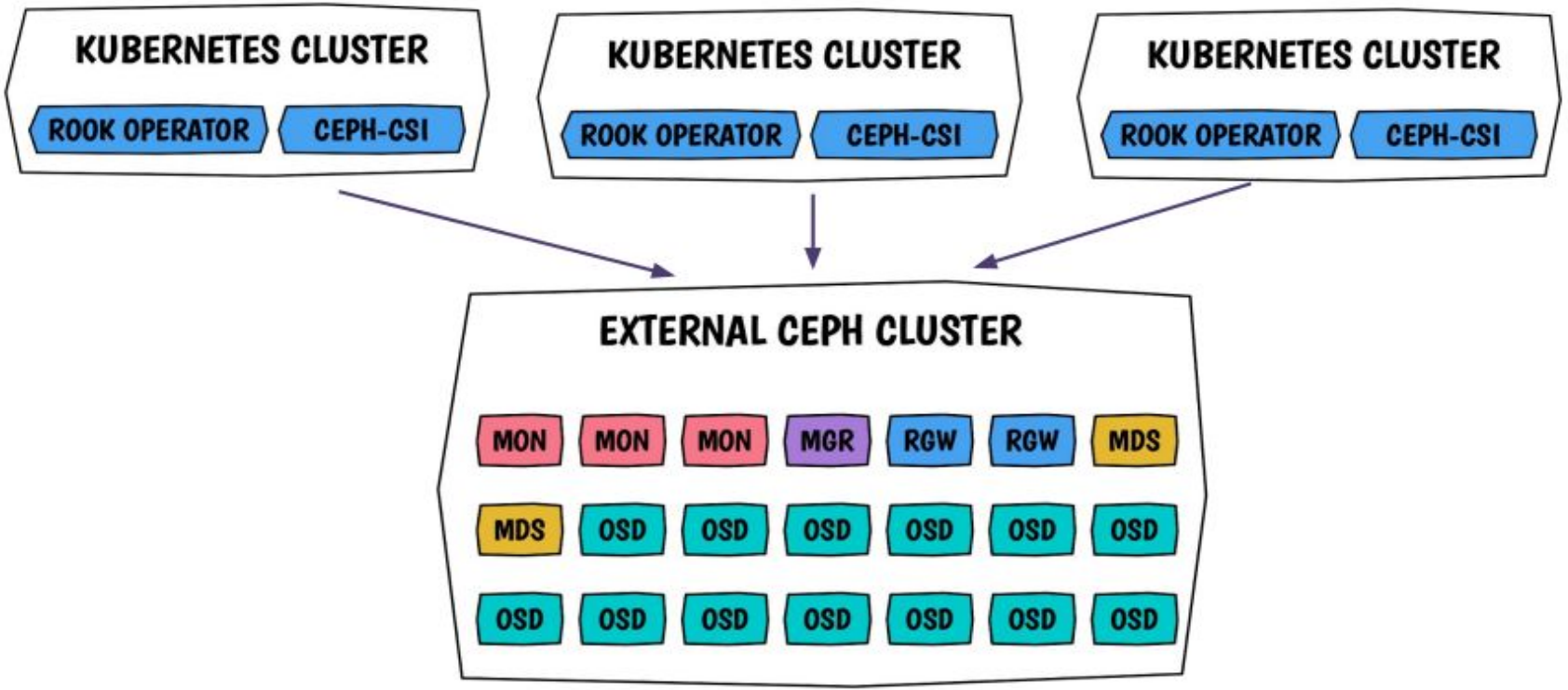


Why Deploy Rook in the Cloud?

- 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

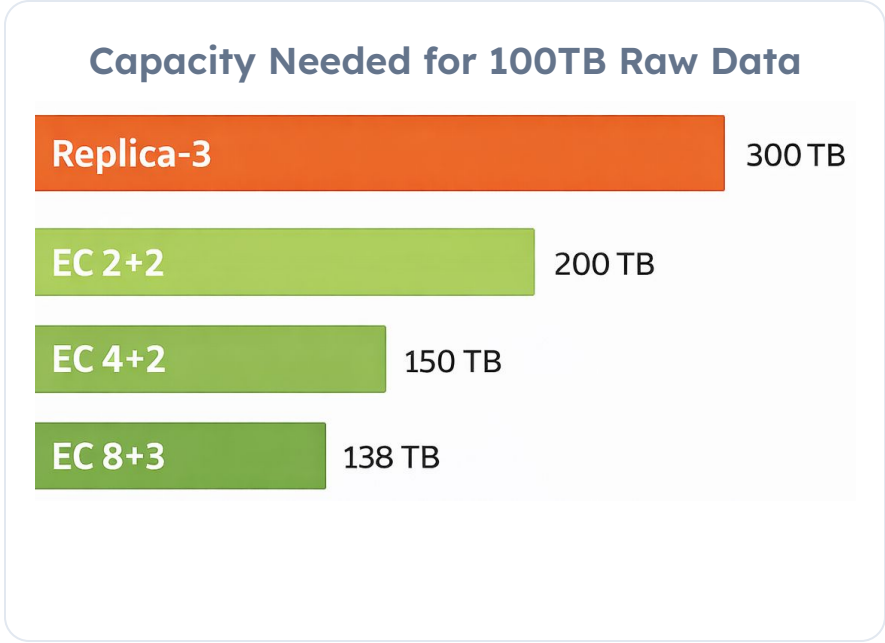
Erasure Coding



Why Erasure Coding

Why make 3 copies when math can reduce the space?

- Most Ceph clusters use **replica-3** for durability, requiring **3x storage requirement**
- Erasure Coding provides **the same or greater durability** with significantly lower storage overhead
- **Same durability, half the storage consumption (Replica-3 vs EC 4+2)**

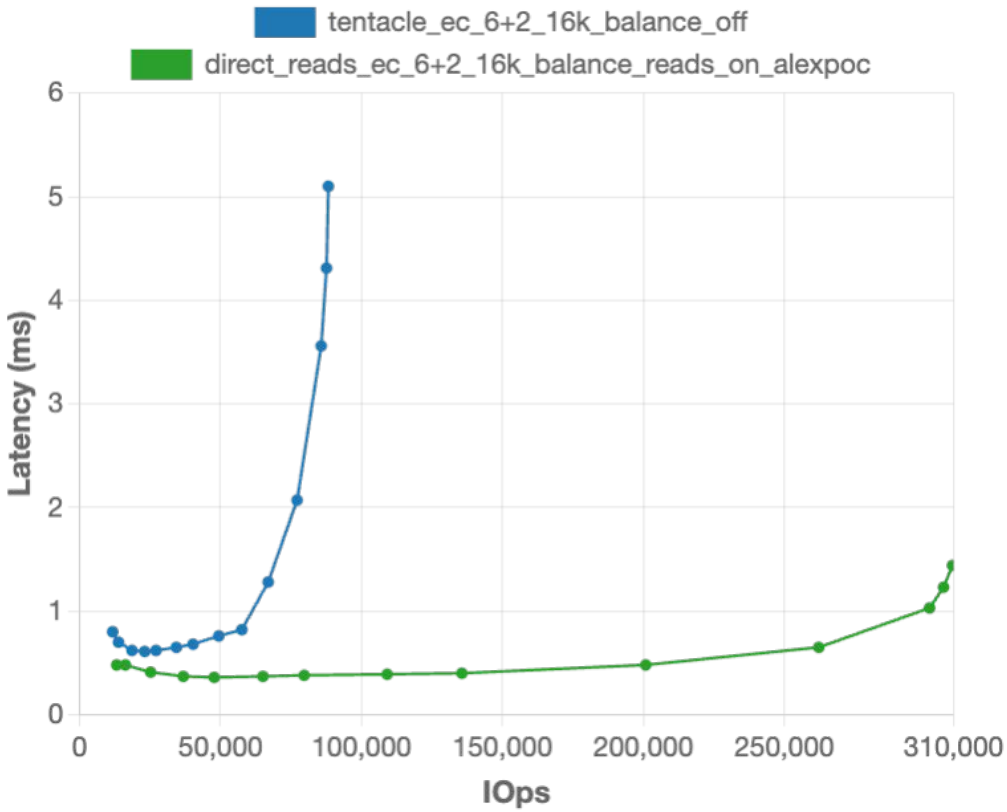


EC Performance Improvements

- Tentacle v20: “FastEC”
  - Read/write performance acceptable for many block/file workloads
- Umbrella v21
  - Read performance: Similar to replicated with “Direct Reads”
  - Write performance:
    - Even **better than replication** for large block sizes.
    - Replication is still better for smaller block sizes
- Vampire v22 (2027)
  - Performance improvements continuing…

With large block sizes, EC performance is better than Replica-3

EC: Performance - 8K Reads Tentacle vs Umbrella+POC



Ceph Days London Talk: All about EC Performance improvements

EC: Test it out!

- Significantly increase your cluster capacity
- Review the performance improvements in the Ceph blog
  - “FastEC” improvements
  - Umbrella blog → coming soon



Trust your workloads with EC

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 with Rook v1.21
  - 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



NVMe-oF Gateway (experimental)

- NVMe over Fabrics
- High-throughput block storage 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



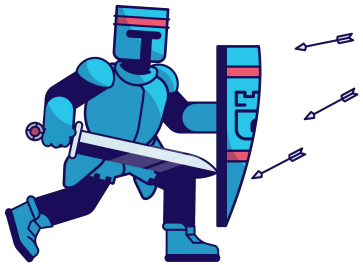
Two Node Clusters (experimental)

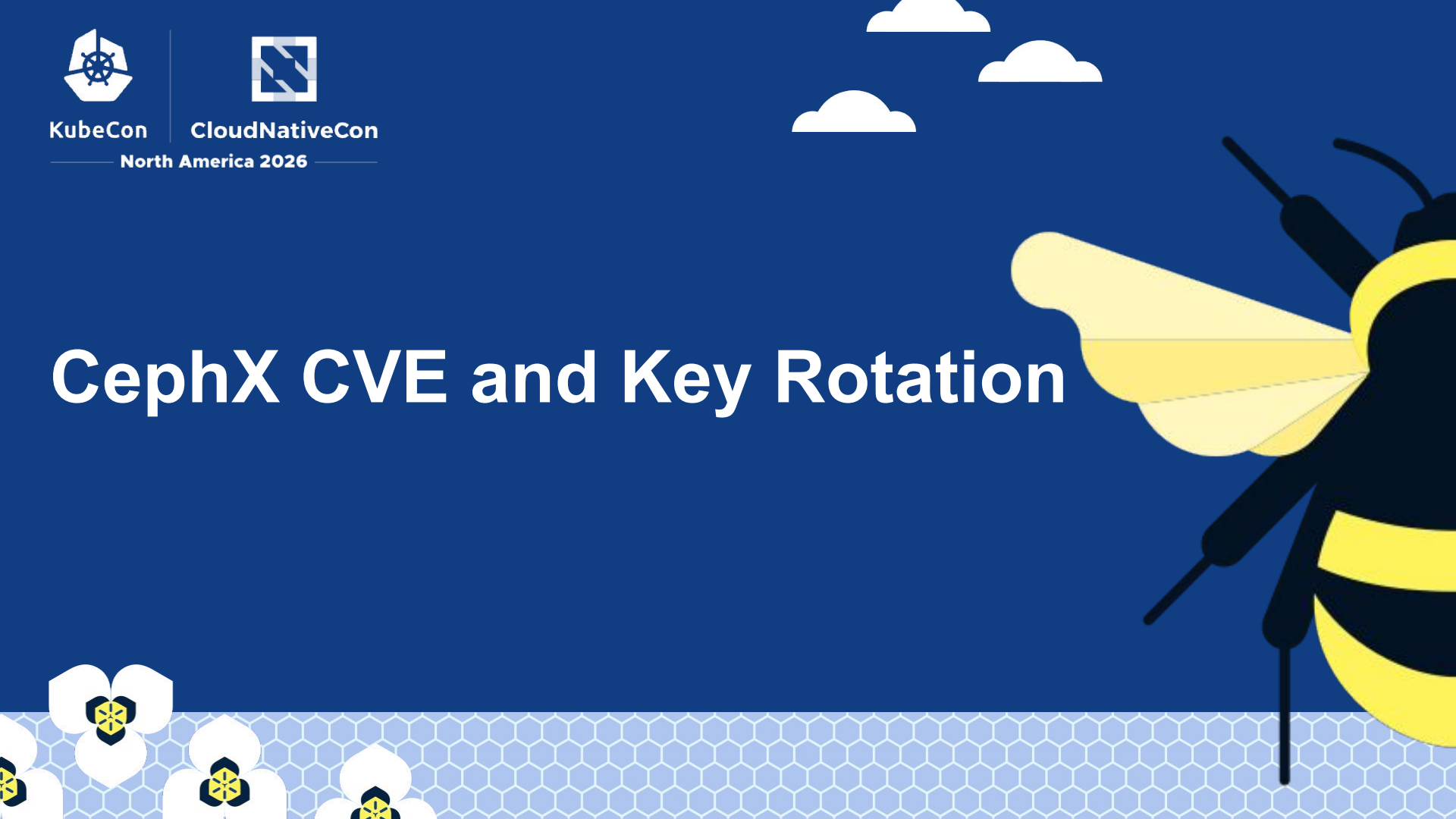
- Edge clusters with only two nodes
- Requires Pacemaker for node failure handling
- A floating mon can move between nodes (using DRBD in experiments)
- Parth presented at Kubecon India 2026 on this topic (QR/link below)

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

Improving Security Processes

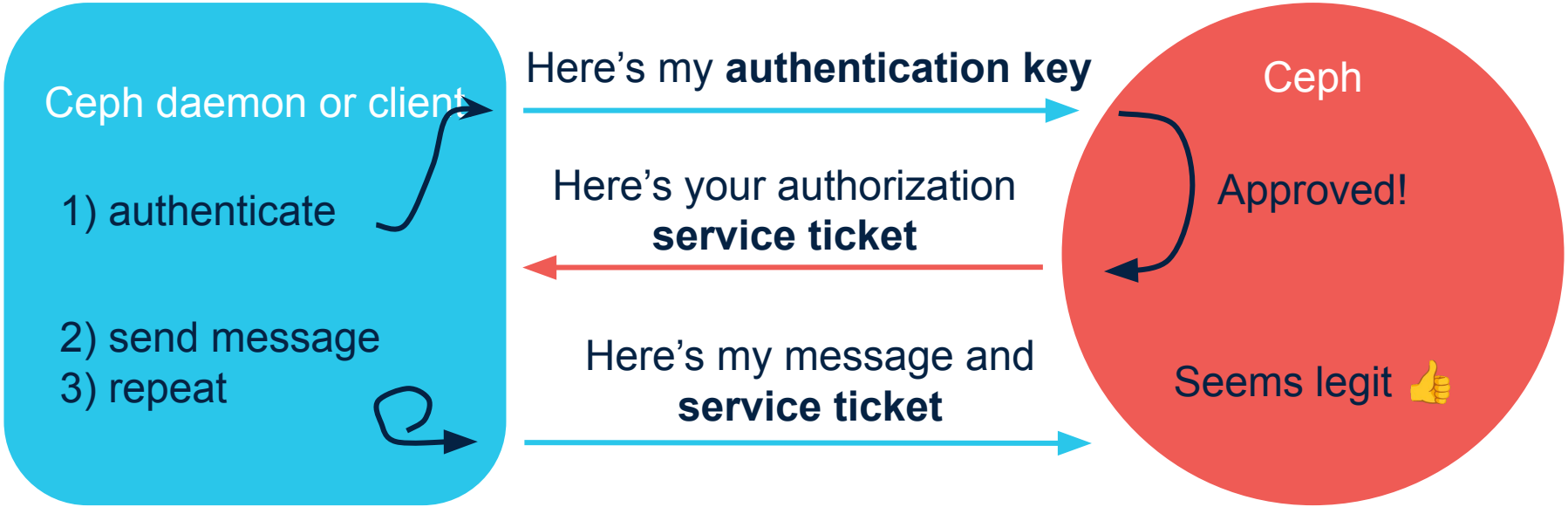
- Regression tests, static analysis, and reviewer focus
- Using latest LLM models to check for security issues every release
- Few major issues compared to other large projects
  - Attribute this to few end-user resources, most admin-owned
- If you discover a security issue, check security doc: https://github.com/rook/rook/security/policy
- Thank you to users submitting issues to our security mailing list





CephX CVE and Key Rotation

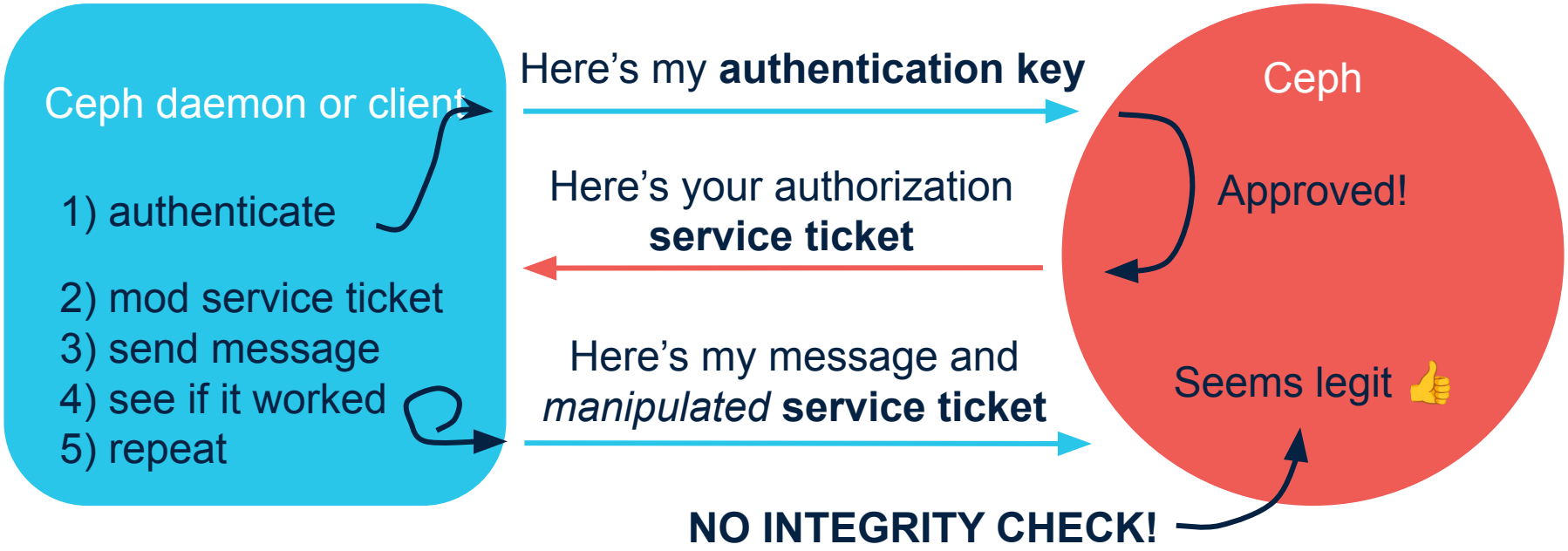
CephX Overview



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** daemon upgrade **required** to support service tickets with new key type (others optional)
- Afterwards, **service tickets use integrity checks**, CVE 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 automatically **rotates 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 (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 (no kernel requirement)

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
```

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

blog.rook.io

KubeCon | CloudNativeCon
North America 2026

User Survey

Questions?

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



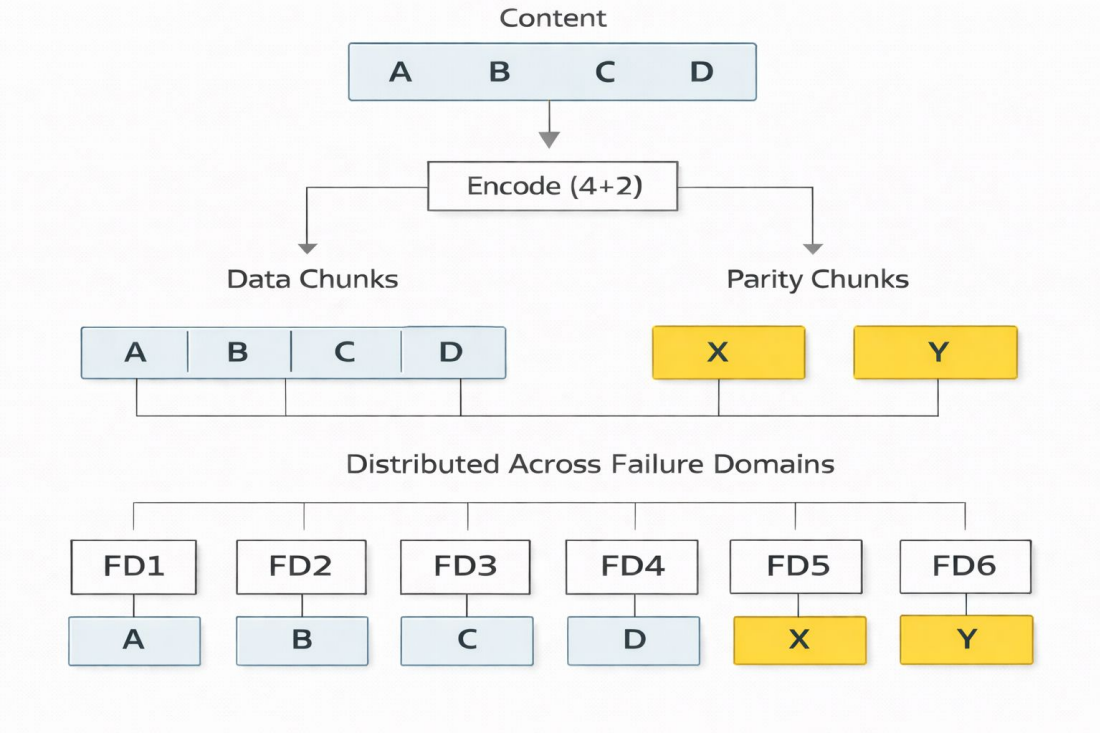
KubeCon CloudNativeCon
North America 2026

Appendix



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

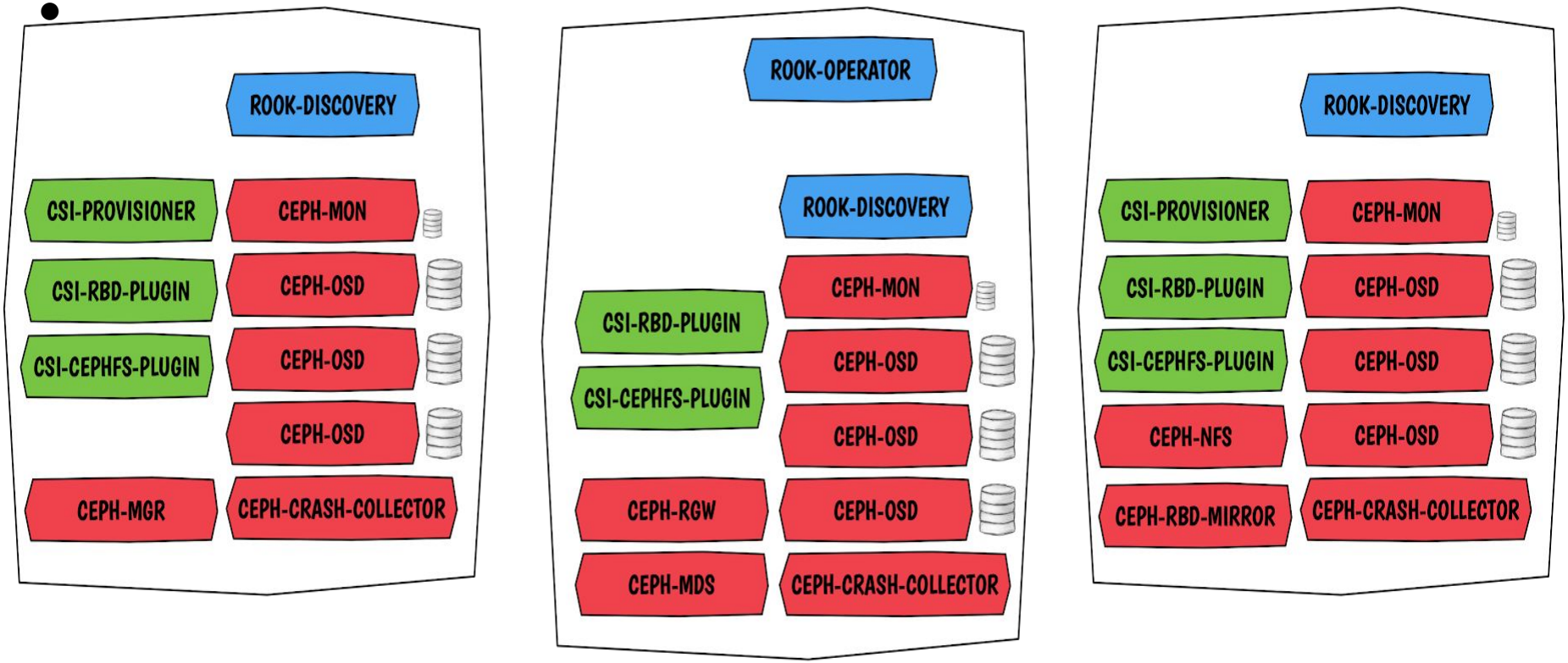
Minimum failure domains:

k + m = 6

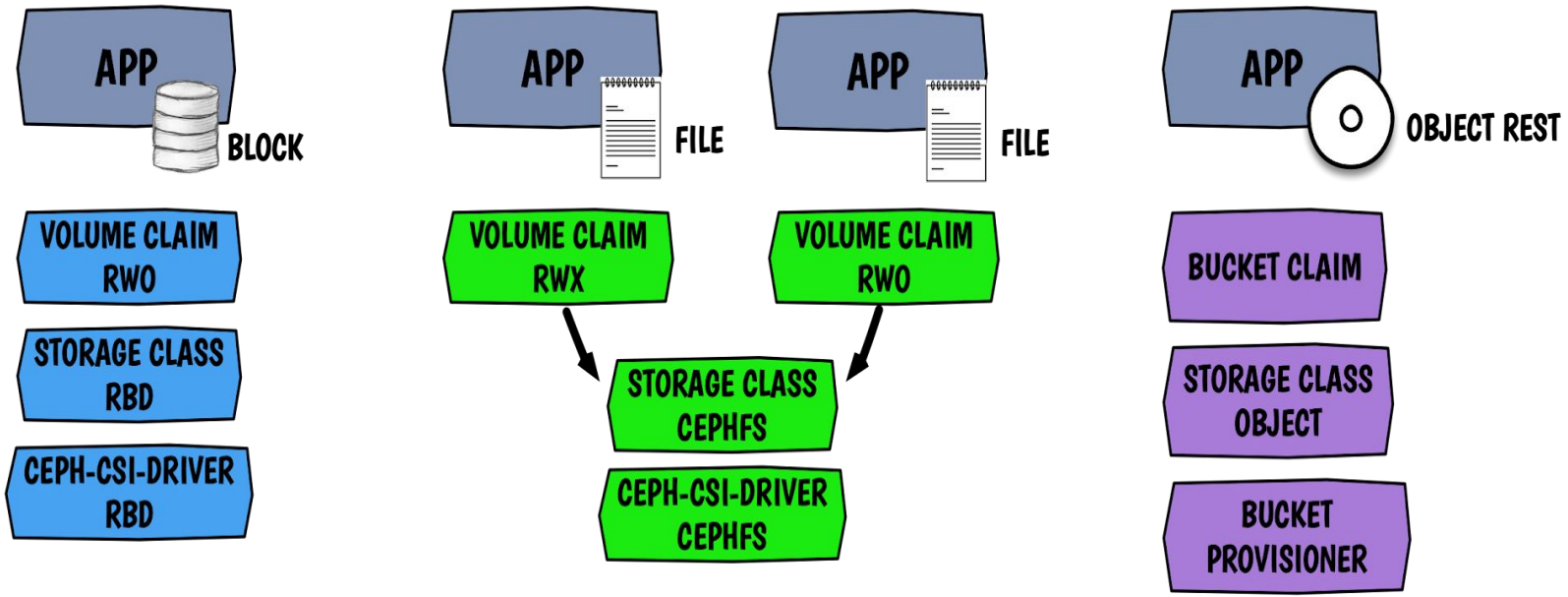
Failure tolerance:

m = 2

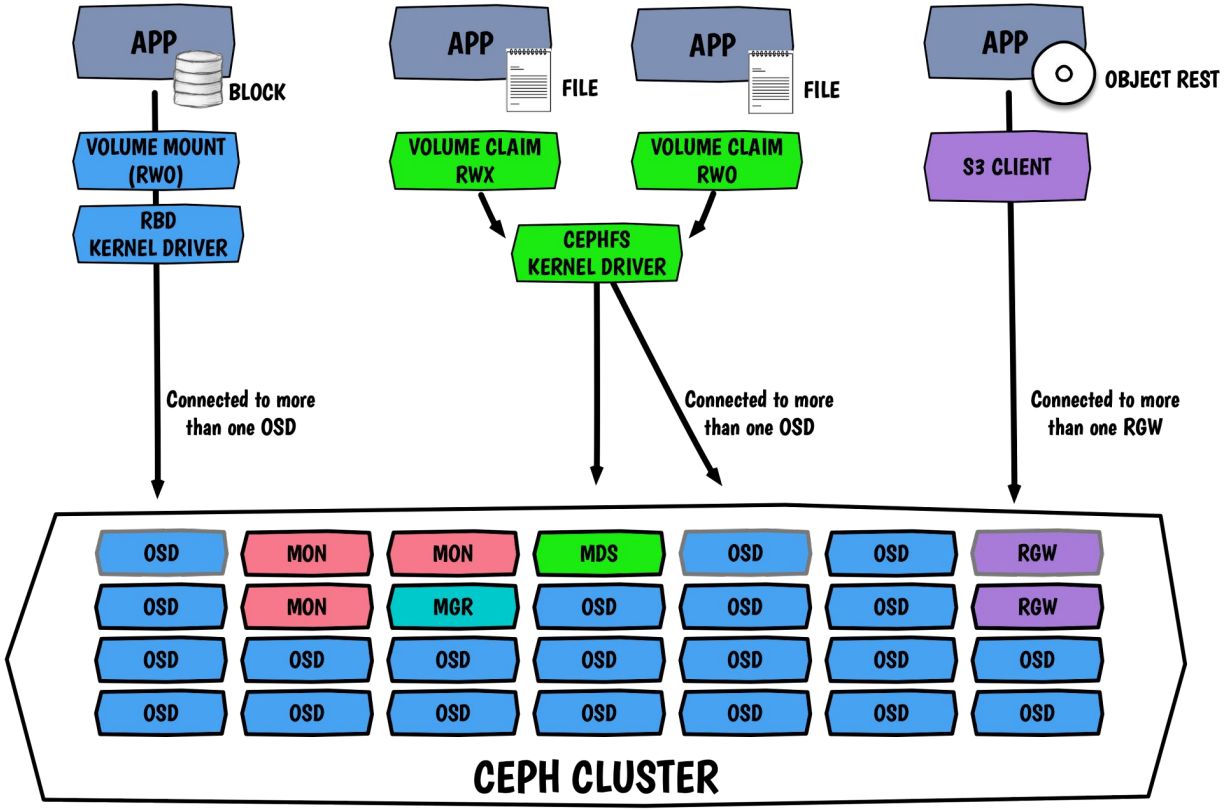
Rook Pods


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

- 

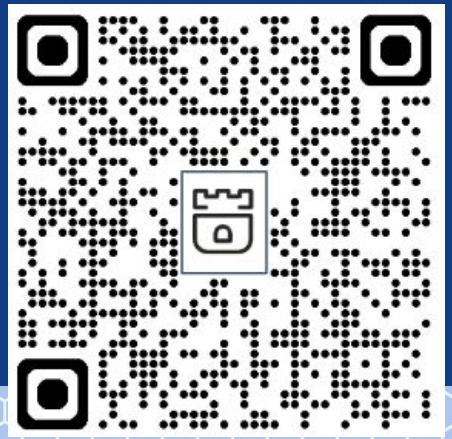
Ceph Data Path

- 

KubeCon CloudNativeCon North America 2026

Survey Form

https://docs.google.com/forms/d/e/1FAIpQLScNEXNfuIOLMbPXzA7DUnvnGkDcFXwm2gv2eJeQ_bQob8xKvw/viewform?usp=publish-editor



Ceph Umbrella

- Rook v1.21 supports Ceph Umbrella (released Oct 28)
- Erasure-coded pool improvements
  - Read and write performance improvements
  - No more need for an additional replicated pool for metadata

- Coming in Ceph Vampire next year
  - Automated pool conversion from replicated to erasure-coded
  - Erasure Coding for stretch clusters
    - Fully erasure-coded replica in each zone

Biology Class

Flapjack Umbrella Octopus



Actually orange

Vampire Squid



Bioluminescent skin