KubeCon CloudNativeCon India 2026

# When the Edge Can’t Afford a Third Node: A Storage Solution for Two-Node Kubernetes Cluster

KubeCon CloudNativeCon India 2026


# Parth Arora

Software Engineer @IBM Storage

**Specialization**

Distributed storage systems

**Connect**

linkedin.com/in/parth-arora-1449ab14a

KubeCon India 2026

# Agenda



- Edge Clusters(Bare Metal) Challenge
- Rook Ceph
- Ceph Quorum
- The Two-Node & Split Brain Problem
- Introducing Two-Node Fencing(TNF)
- How TNF Prevents Split-Brain
- TNF + Rook Architecture
- Demo


# Edge Clusters Challenge

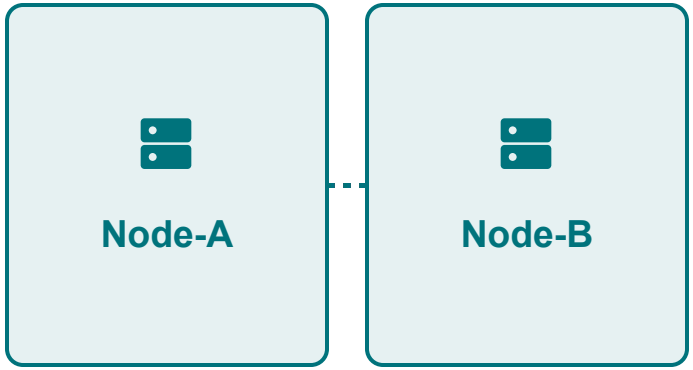
# Edge Clusters



## Edge Constraints

- Space and power constraints
- Limited hardware availability
- Often only two servers available
- Storage still needs HA

## Edge Site Layout




Typical 2-Node Configuration


# Introduction to Rook

# What is Rook?

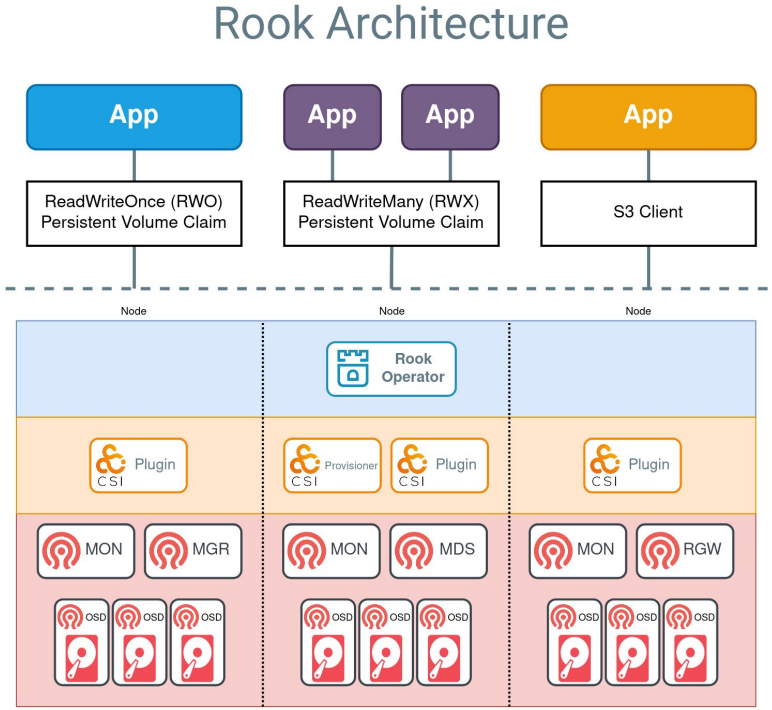


- Brings Ceph storage into your Kubernetes cluster
- Manages Ceph storage with an operator and CRDs
- Automated deployment, configuration, upgrades
- Allows apps to consume storage like any other K8s storage
  - Storage Classes, PVCs
- Open Source (Apache 2.0)
- CNCF Graduated Project since 2020
- Learn more: https://rook.io
- Happy 10th Birthday!
  - Created November 2016

# Architectural Layers



- Rook
  - Operator **deploys** and **manages** Ceph
- CSI
  - Ceph CSI driver dynamically **provisions** and **mounts** storage to user application pods
- Ceph
  - **Data** Layer



# Ceph: Open-source, distributed Enterprise storage platform



- **Block** (RWO)
  Ceph RBD
- **File** (RWX)
  CephFS
- **Object** (S3)
  Ceph RGW


# Ceph Architecture Requirements

# Ceph Node Requirements



### Standard Production

Three nodes are expected for production clusters.

### Data Safety

Three data copies are recommended for highest data safety.

### Edge Constraint

Only two nodes are available with TNF clusters in the edge.

***Question: Is it possible to fit a three-node architecture into two nodes?***

# Ceph Mons: Brain of the Cluster

KubeCon CloudNativeCon India 2026

## Cluster State

Ceph MONs maintain cluster state. They are the source of truth for the cluster map.

## Quorum Requirement

Ceph requires a majority of MONs (quorum) to communicate and establish consensus.

## Split-Brain Prevention

This prevents split-brain scenarios and ensures a consistent cluster state across all nodes.

## Persistence

Mons critical metadata is persisted directly to the local disk for durability.

# Why Ceph Require 3 Node?



## 3-Node Configuration

**Standard Production**

- 3 unique MONs (one on each node)
- Quorum: 2/3 MONs active
- **Tolerance: 1 Node Failure**

## 2-Node Constraint

**Edge Scenario**

- 2 unique MONs only
- Quorum requires 2/2 MONs
- **Tolerance: 0 Node Failures**

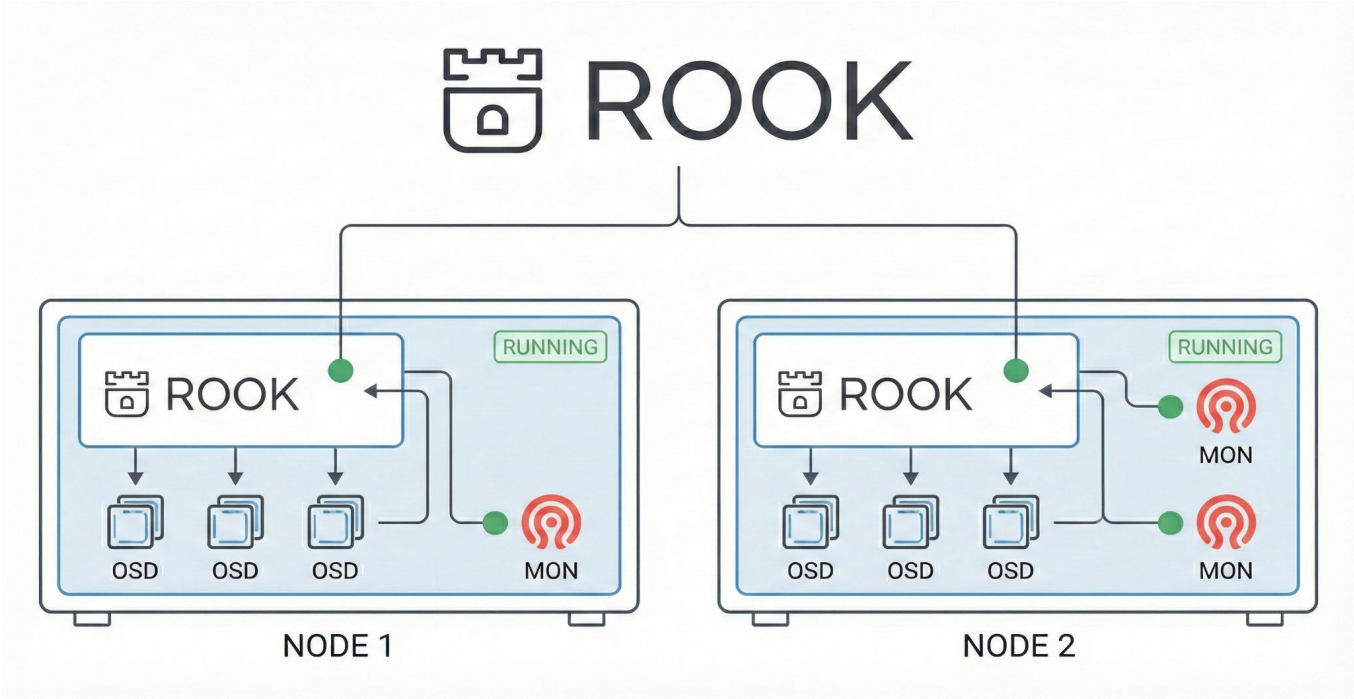
### *The "Floating MON" Concept*

What if we could run a third monitor that "floats" between nodes to provide a dynamic quorum?


# Lets talk about a Two-Node Failure Scenario

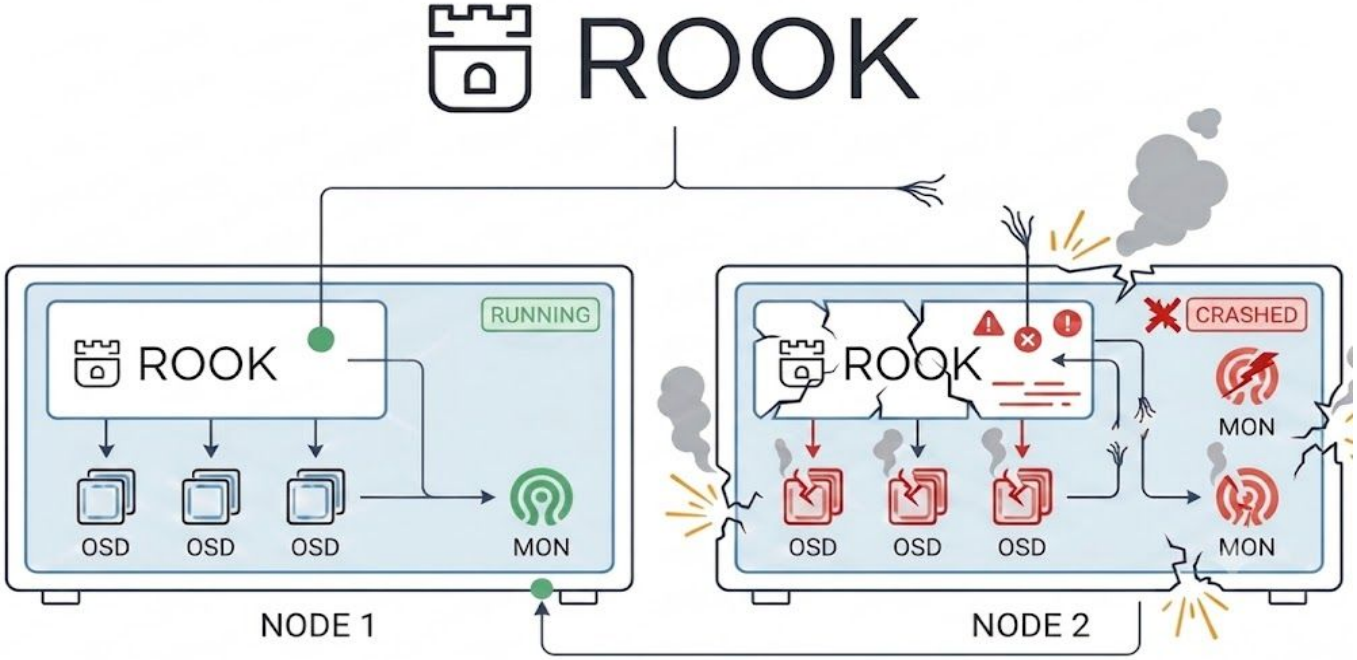
# Two Node Cluster: Online





# Non Graceful Node Shutdown

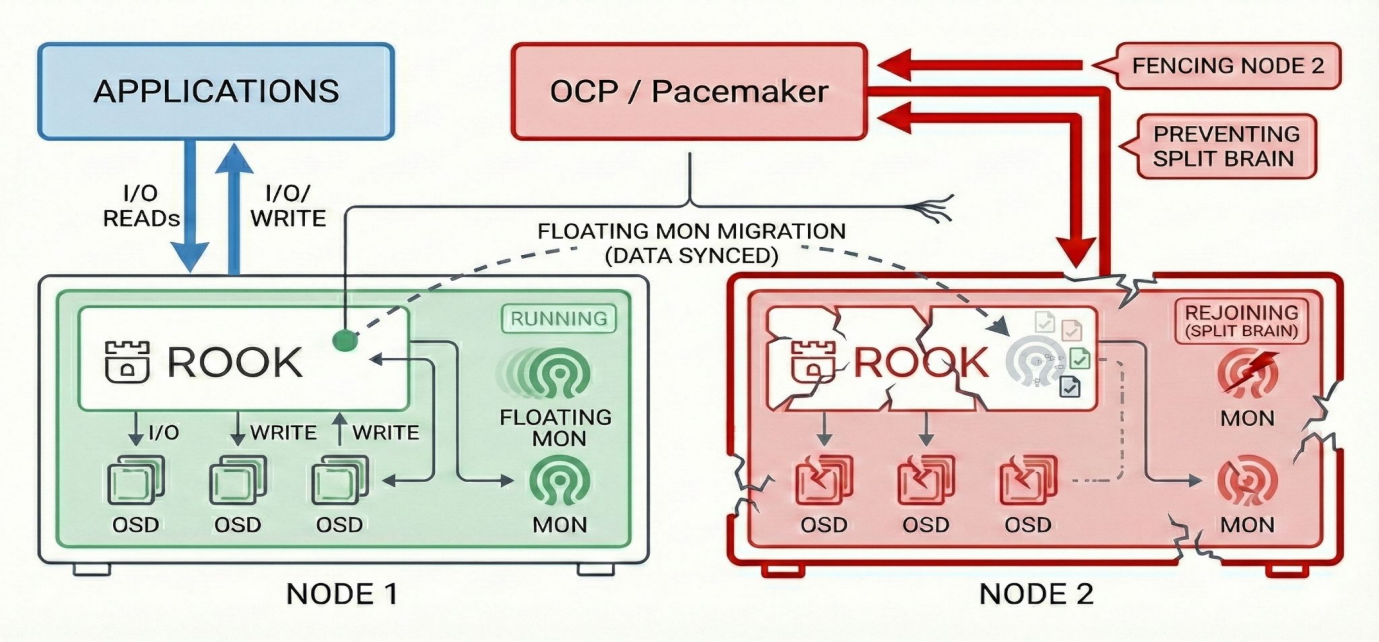




# Cluster i/o Recovery



Floating mon will move from one node to another, to maintain the Quorum.



# Fencing node using Pacemaker(STONITH)

KubeCon CloudNativeCon India 2026

### Active Continuity

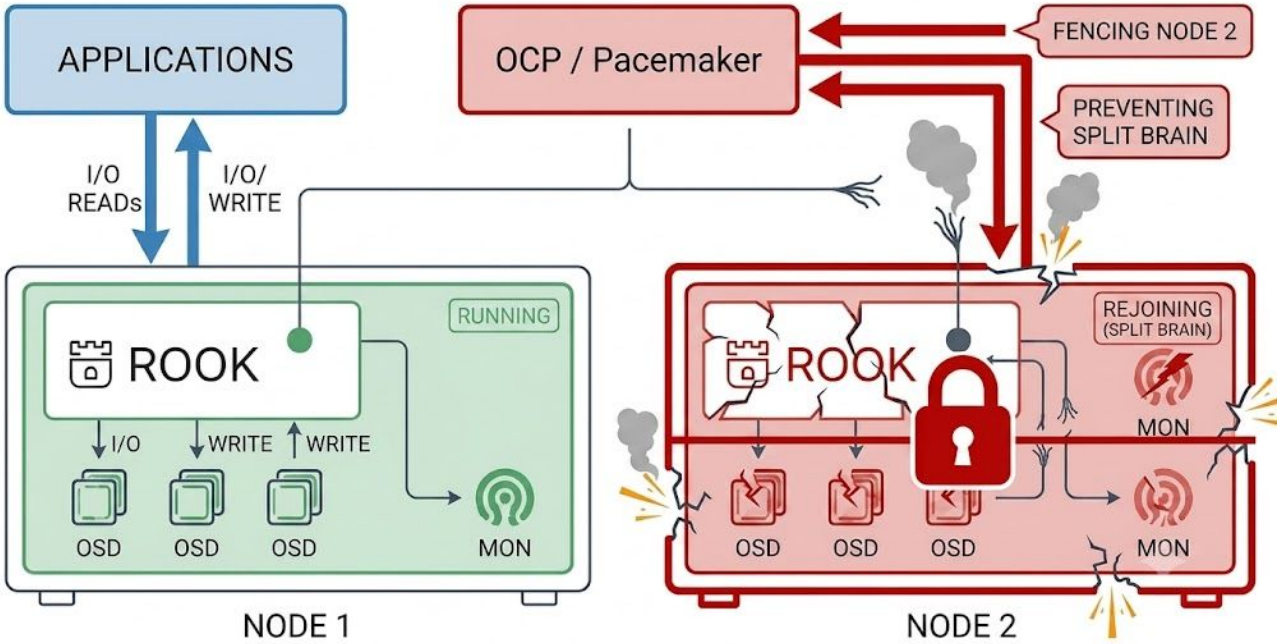
The remaining active node continues to serve I/O to applications during the outage.

### Split-Brain Risk

A split-brain scenario can occur if the failed node re-establishes connectivity and attempts to resume operations without coordination.

### Data Consistency

OCP utilizes Pacemaker to fence the failed node via its Base Management Controller (BMC) using the RedFish API.




# Solution For Floating Mon

# Floating Mon Design



## Configure a floating mon, ensuring data is replicated in real time across both nodes.

### Node Flexibility

A floating mon is allowed to run on either of the two nodes.

### Storage Replication

The mon metadata store on disk must be replicated between the two nodes.

### Networking Constraint

Must work with **host networking: false**.

**Key Challenge:**
How to reliably replicate the data?

# DRBD: Distributed Replicated Storage

KubeCon CloudNativeCon India 2026

**DRBD** is open source distributed replicated block storage software for the Linux platform.

Typically used for high performance and high availability in demanding environments.

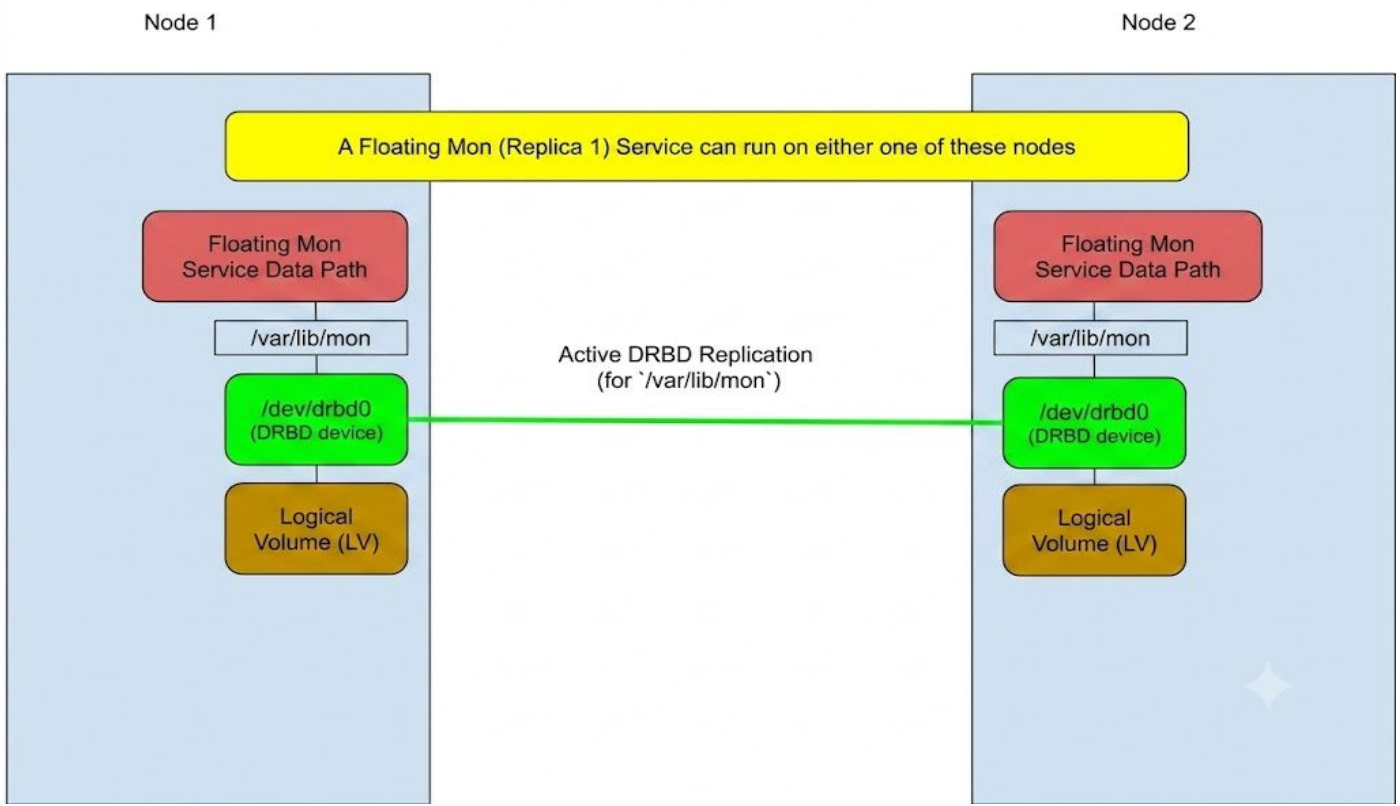
## High Performance

## High Availability

READ DOCS

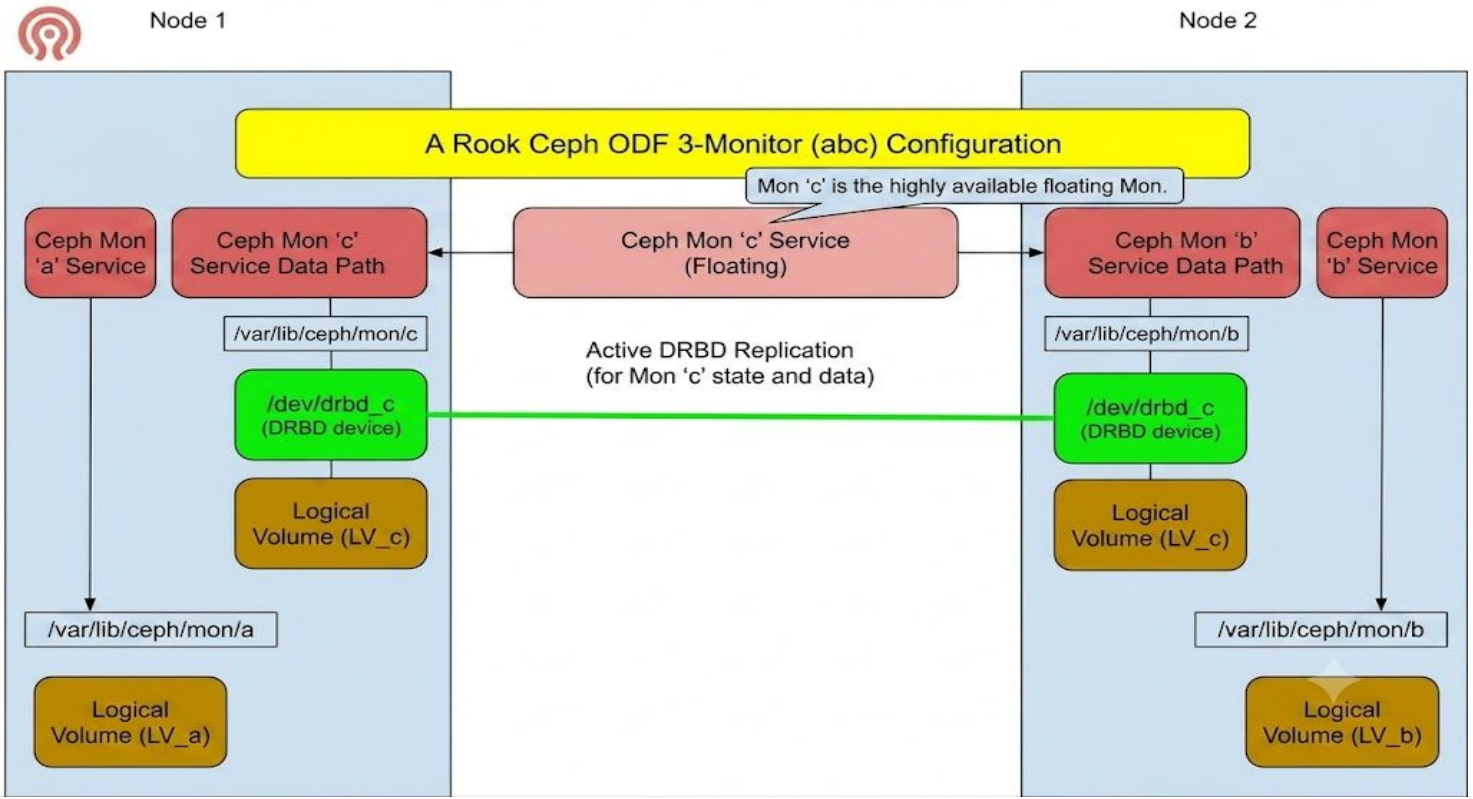
# Floating Mon design





# Floating Mon design





# Zooming In: Container Workflow

KubeCon CloudNativeCon India 2026

## Init Container

**Unmount:**

- Handles leftovers from failed pre-stop containers.

**DRBD Primary:**

- Promotes the resource to primary state.

## Main Container

**Mount DRBD Disk**

Mounts the disk to **hostpath** and executes normal operations.

## Floating-mon-shutdown

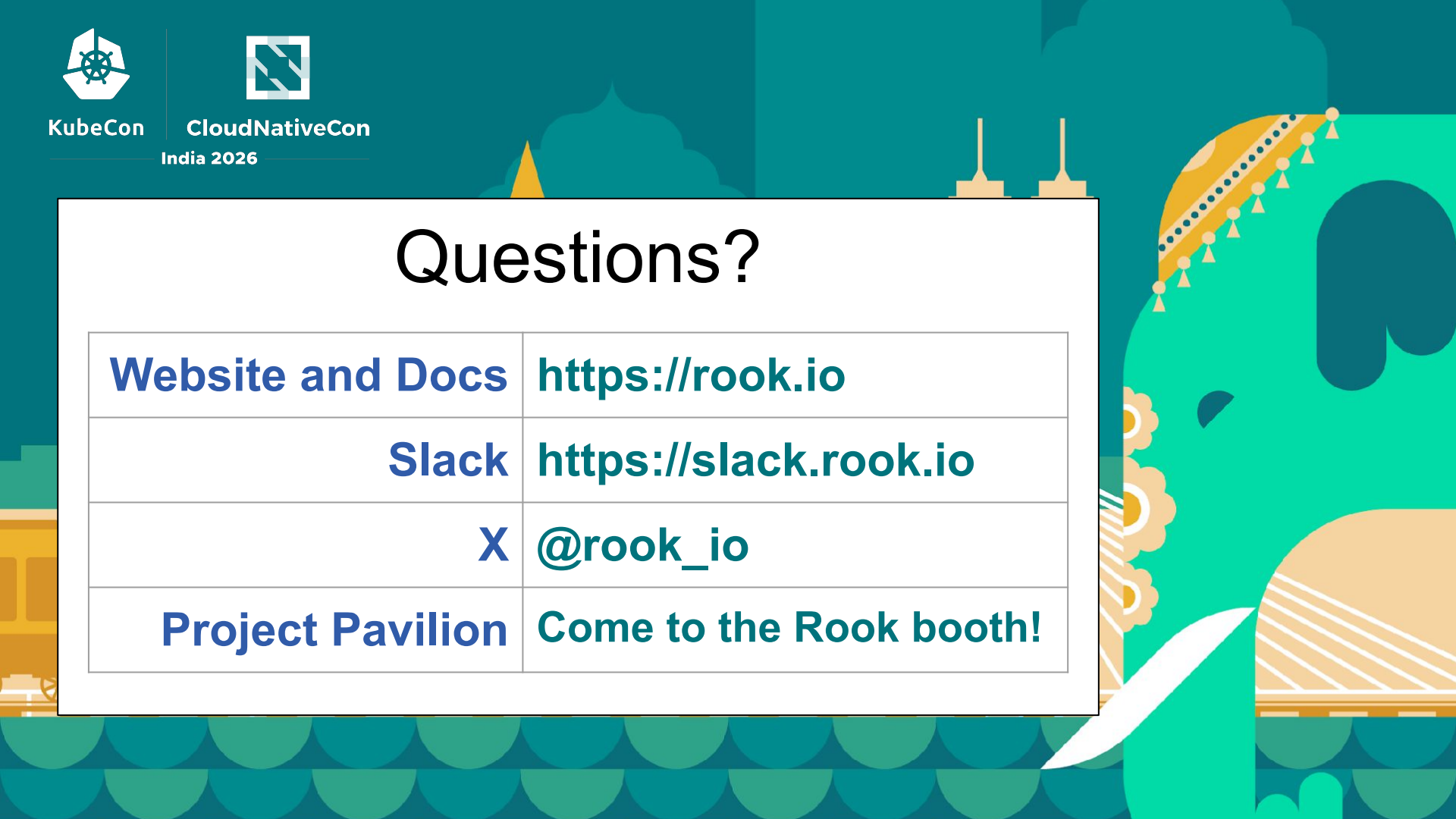
**preStop Hook:**

- Demotes DRBD to secondary
- Unmounts hostpath

*Note: All other containers run without modifications.*

KubeCon CloudNativeCon India 2026


# Demo Link




# Questions?

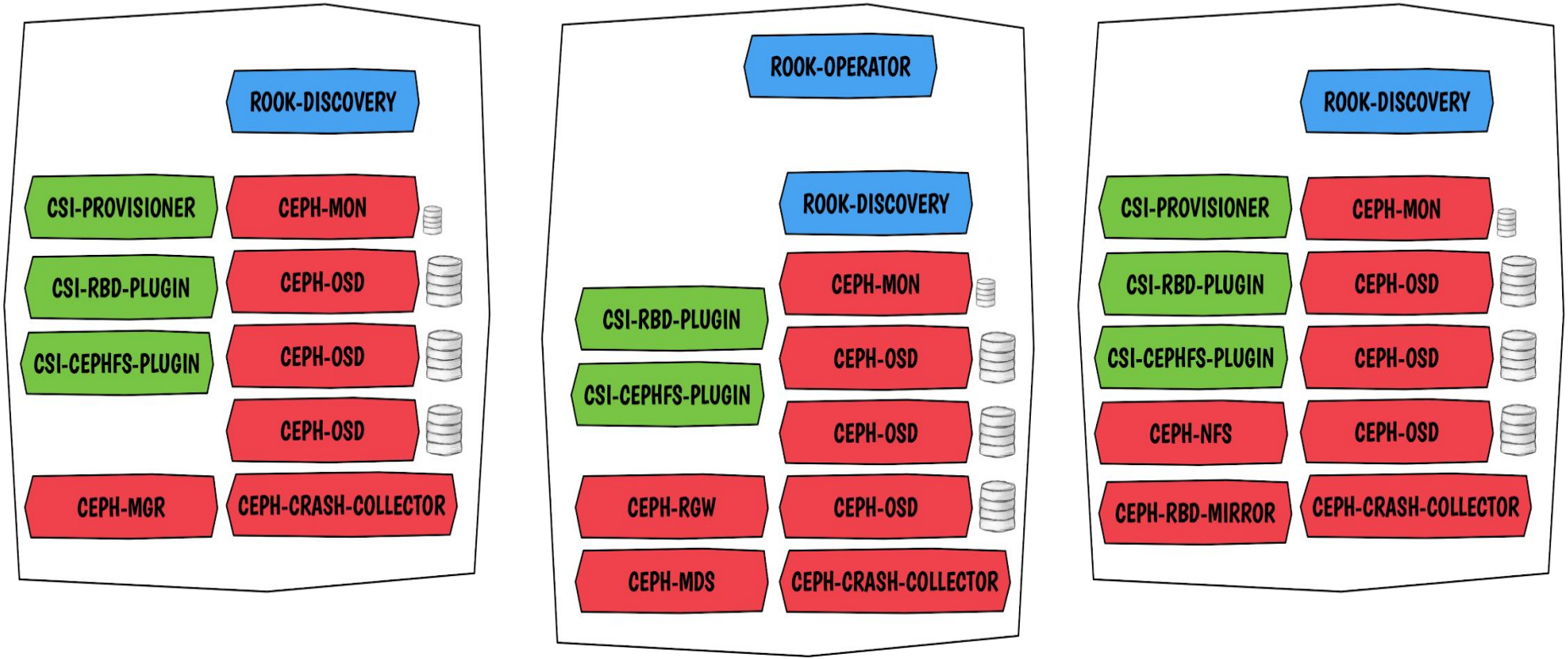
| | |
|---|---|
| Website and Docs | https://rook.io |
| Slack | https://slack.rook.io |
| X | @rook_io |
| Project Pavilion | Come to the Rook booth! |

KubeCon CloudNativeCon India 2026

# Appendix

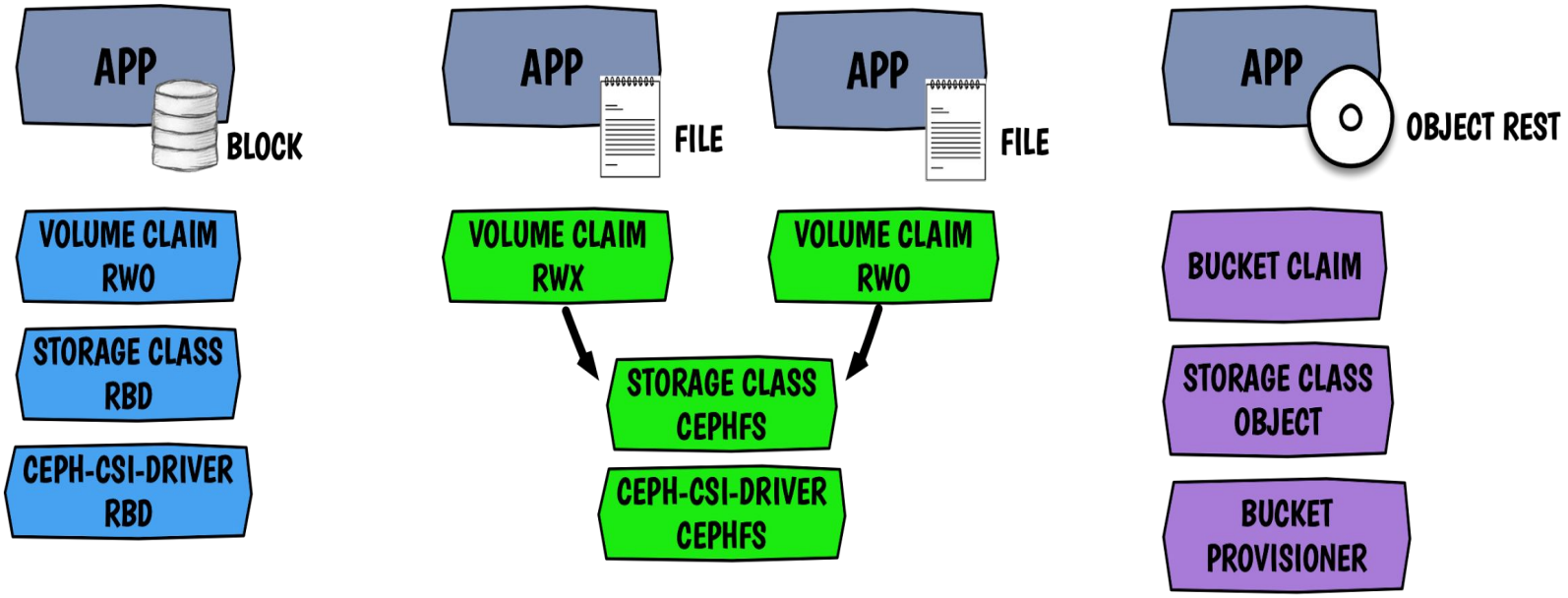
# Rook Pods




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