


KubeCon | CloudNativeCon

Europe 2023

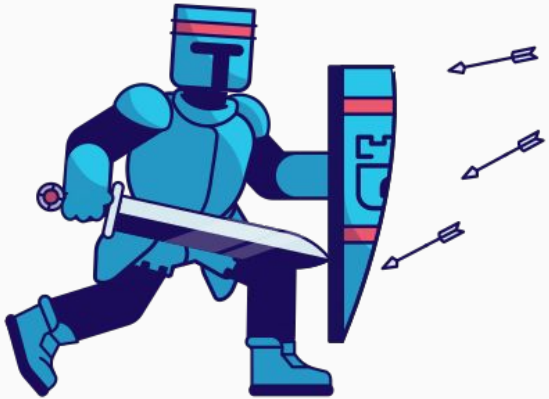
Rook: Storage for Kubernetes

Travis Nielsen, IBM Storage
Blaine Gardner, IBM Storage
Alexander Trost, Koor Technologies
Deepika Upadhyay, Koor Technologies

April 2023

Agenda

- Welcome to our Rook panel
- Our goal is to familiarize you with storage for Kubernetes
- Question for the audience:
  - Is Rook a potential storage solution for me?
- Time at the end for your questions
- Come talk to us at the Rook booth in the project pavilion



How should someone new to Kubernetes think about storage?

Storage in Kubernetes

- Kubernetes easily manages distributed apps
- For a user application, storage should be easy
  - Application failover/scale-out, storage “follows” pod
  - Across nodes and partitions (availability zones)
- Kubernetes has no distributed storage options built in for admins
  - External storage - Who manages that?
  - Relying on cloud provider - Vendor lock-in?
  - Run distributed storage on Kubernetes - Can I handle that?

How did Rook Get Started?

Questions that led to Rook

- Storage is commonly provided by cloud providers
- What about storage in your datacenter?
- Storage is traditionally not part of the cluster
  - Why should storage be external to K8s?
- Why not manage storage as any other K8s application?

Storage Platform

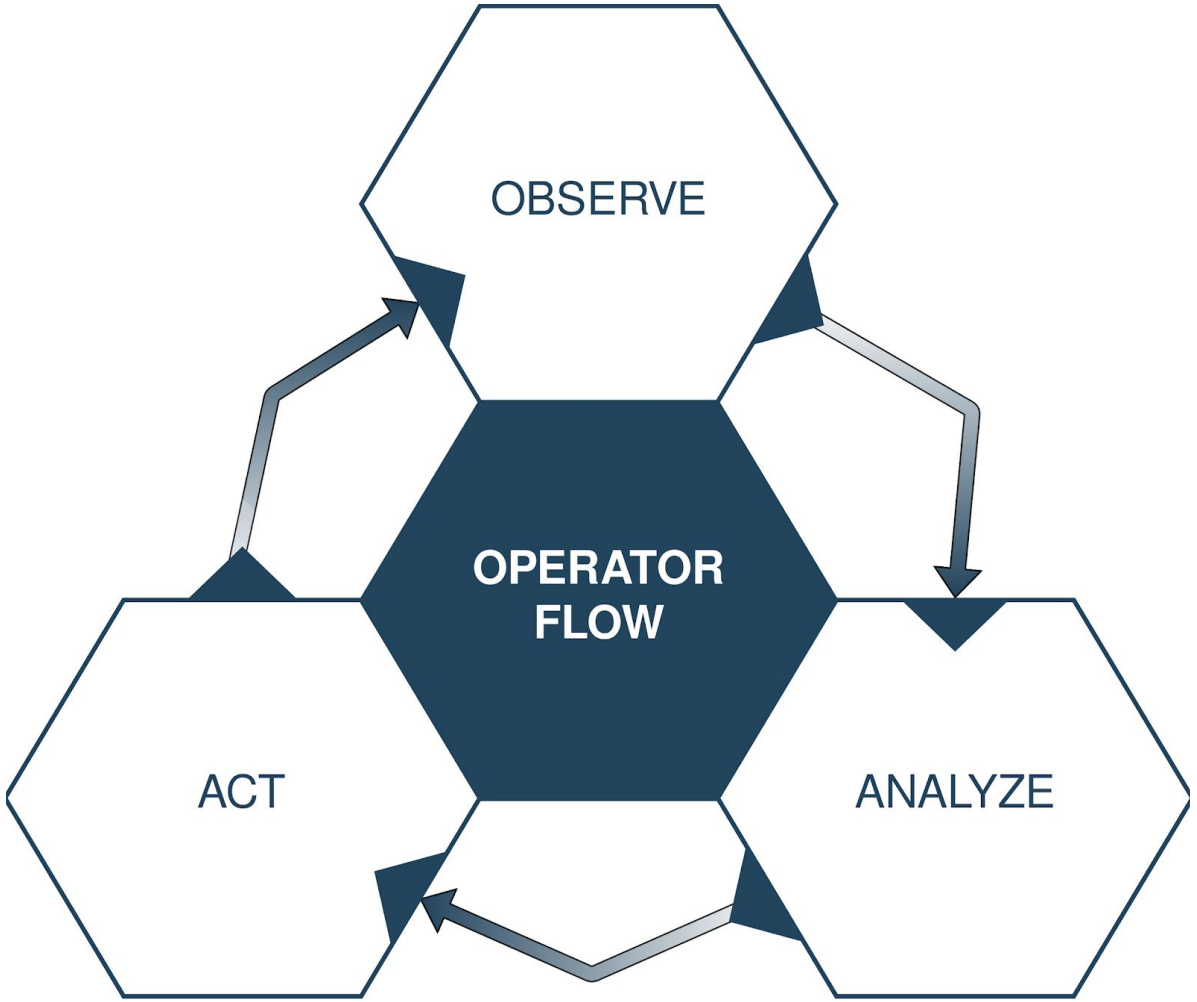
- Which storage platform to trust?
- Enterprises don’t trust a new data platform
- We didn’t want to build a new storage platform
- We made the decision to build on **Ceph**

What is Rook?

What is Rook?

- Makes storage available inside your Kubernetes cluster
- Manages Ceph storage with an operator and Custom Resource Definitions (CRDs)
- Automates deployment, configuration, upgrades
- Allows apps to consume storage like any other K8s storage
  - Storage Classes, Persistent Volume Claims
- Open Source (Apache 2.0)

Photo by Gabriel Santos Fotografia from Pexels


OBSERVE
OPERATOR
FLOW
ACT
ANALYZE

What is Rook’s CNCF status?

CNCF Graduated

- Rook is a CNCF graduated project!
  - Sandbox: January 2018
  - Incubation: September 2018
  - Graduation: October 2020

Who is involved in the Rook community?

Rook Community

- Community first
- Open Source (Apache 2.0)
- Maintainers from four companies
  - Cybozu, IBM/Red Hat, Koor, Upbound
- 400+ contributors to the Github project
- 280M container downloads

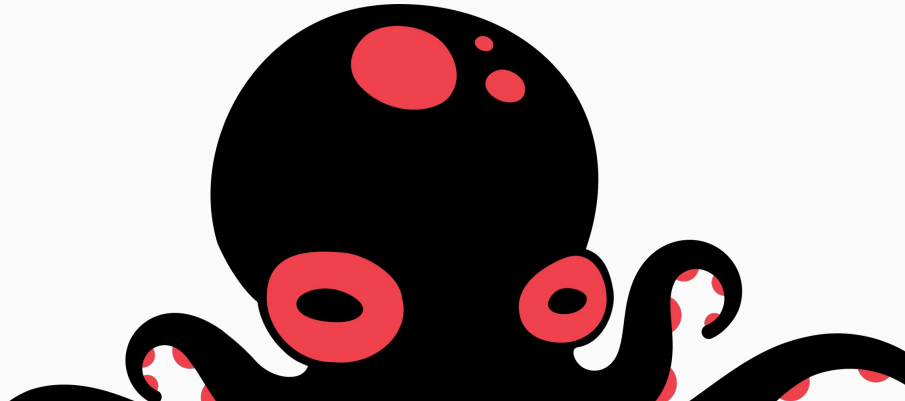
Raise your hand if…

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

What is Ceph?

What is Ceph?

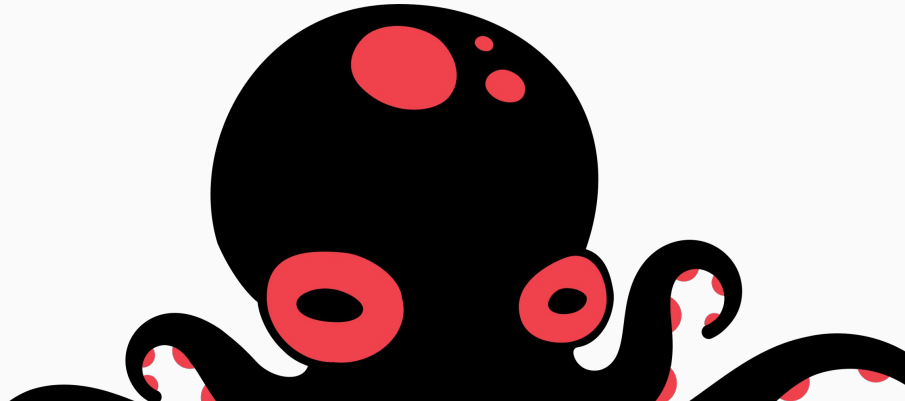
- Open source
- Distributed Software-Defined Storage solution
  - Block
  - Shared File System
  - Object (S3 compliant)
- https://ceph.io/



Why Ceph?

Why Ceph?

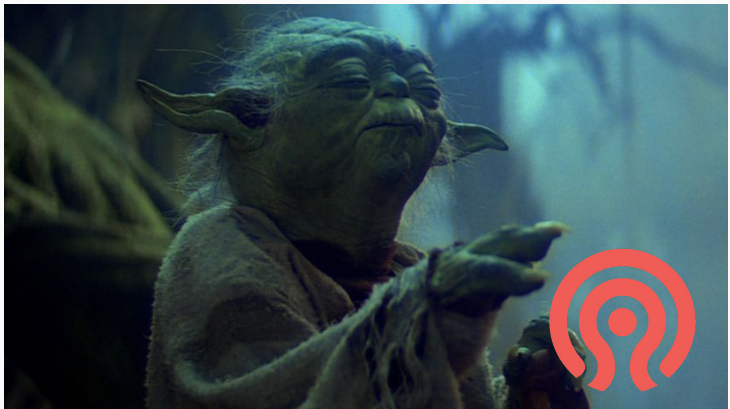
- Block, Object, and File storage in a single platform
- Proven history of enterprise adoption and support
  - First release in July 2012
  - CERN’s Large Hadron Collider!



Does Rook support storage providers other than Ceph?

Storage providers

- Ceph is the only supported storage provider
  - The force is strong with the Ceph community
- Rook maintainers explored supporting other storage platforms in the past
  - Lacked community support



How stable is Rook?

Rook Stability

- In our 3rd year since CNCF graduation
- In our 5th year since declared stable for production
- Many upstream users running in production
- Many downstream deployments running in production

Case Study

- National Research Platform deployed Rook starting 5+ years ago
- **6PB** storage across clusters throughout the world
- Largest cluster:
  - **3PB**
  - **237 OSDs**

“Rook significantly simplifies our persistent storage needs in Kubernetes by automating all Ceph setup, monitoring, and config management.”

“With Rook, adding and using a new Ceph cluster requires almost no efforts and becomes a trivial task”

– Dmitry Mishin, National Research Platform

How do you install Rook?

How do you install Rook?

- Helm charts
- Example manifests for many configurations
- Intro and quickstart guide
  - https://rook.io and click Get Started
- Multi-node cluster installs in about 12 minutes
  - Install demo: https://youtu.be/TwGJsTa3F2g

In what environments can Rook be installed?

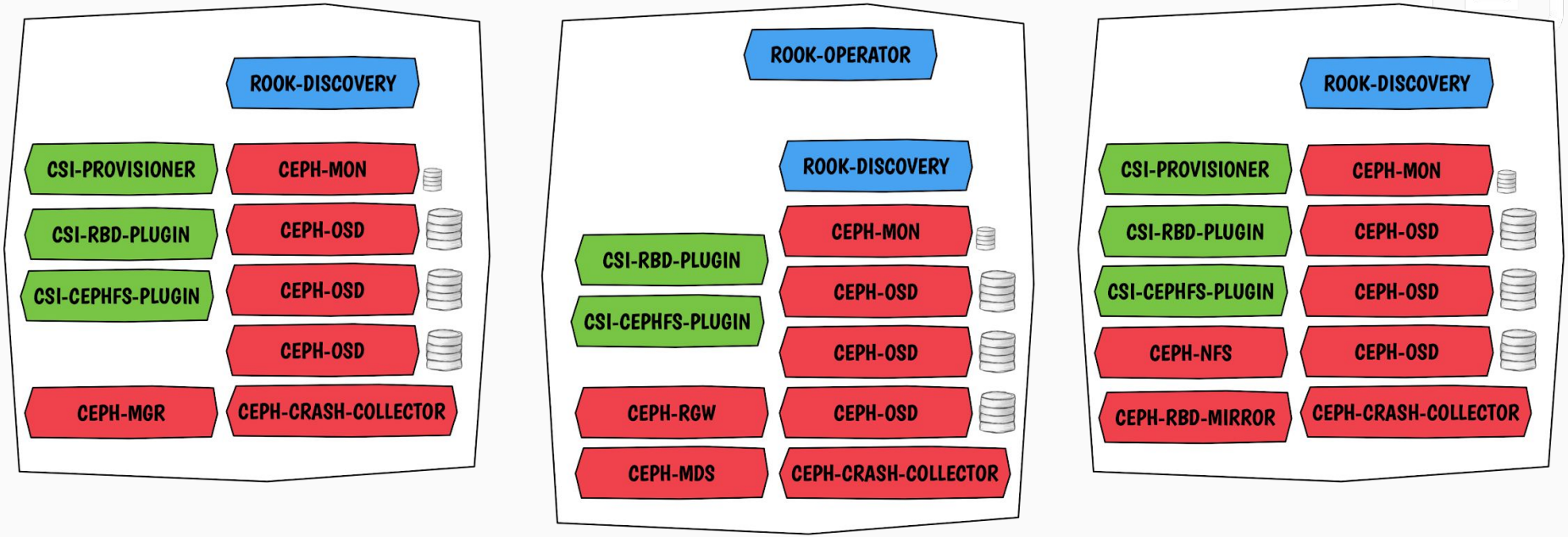
Rook installation environments

- Anywhere Kubernetes runs
  - Cloud or on-premises
  - Virtual or bare metal hardware
  - Underlying storage can be node-attached devices, cloud volumes, or loopback devices
- Rook helps enable cross-cloud support

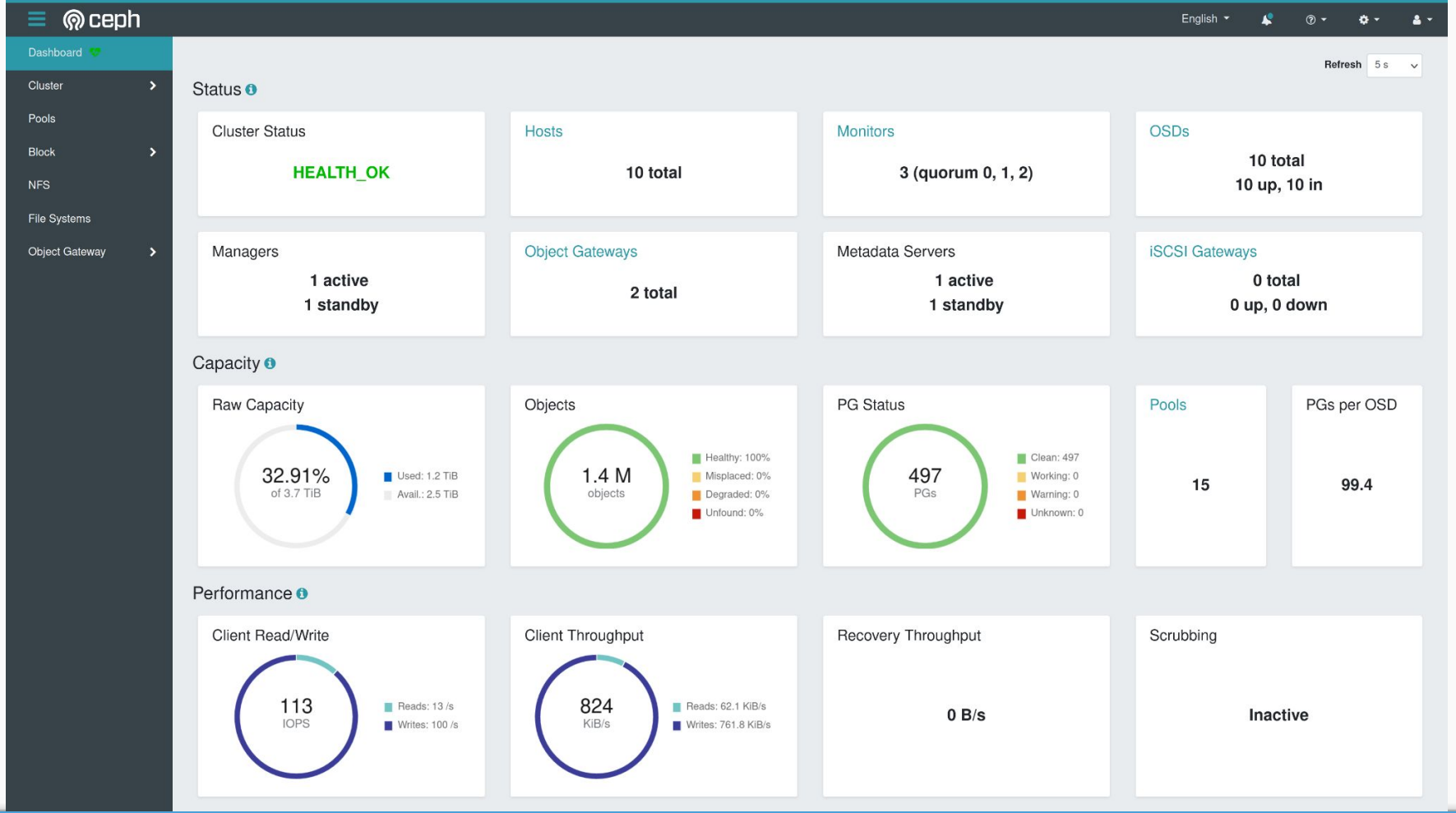
What should I know about the architecture?

Architectural Layers

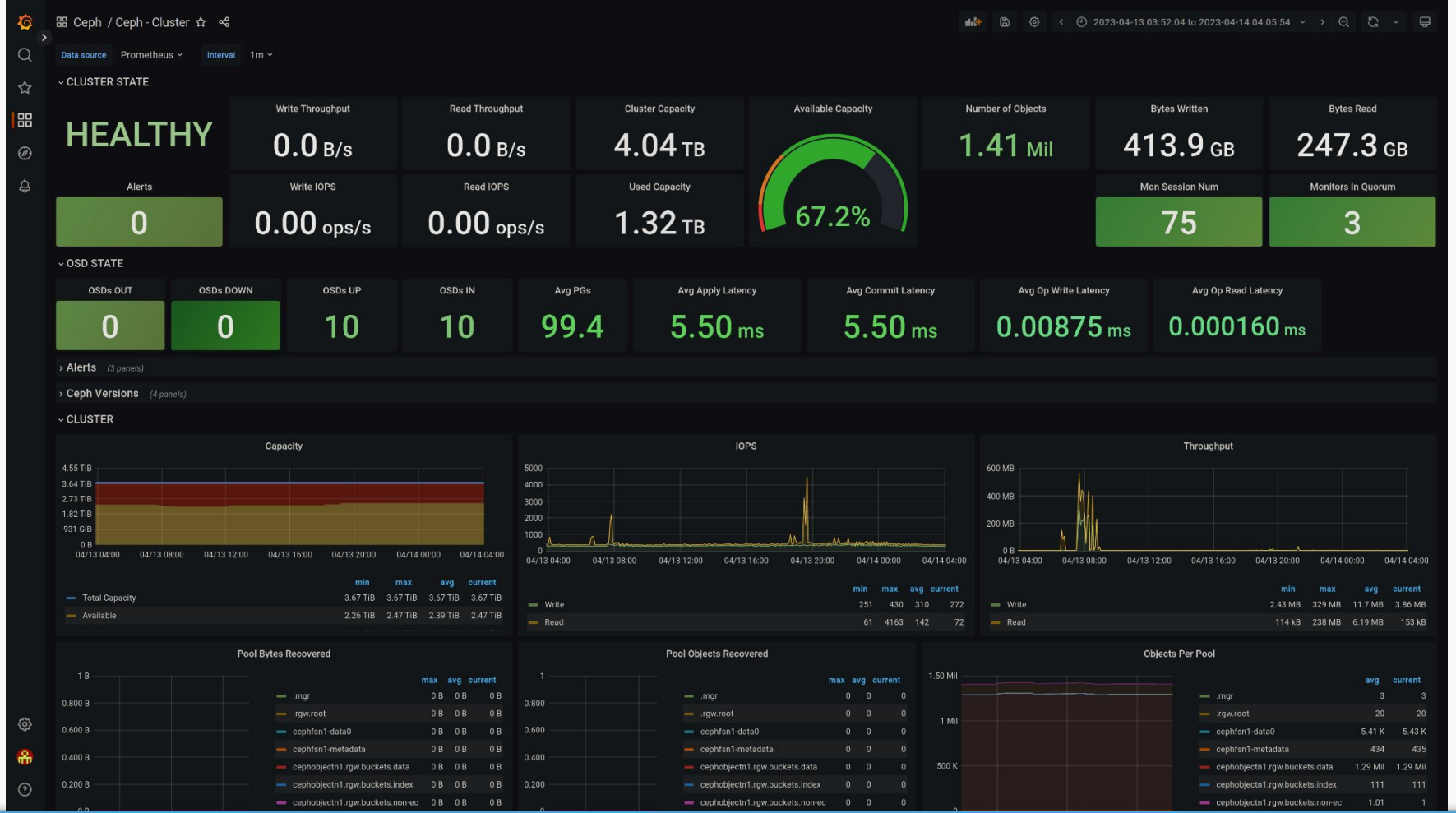
- Rook
  - Operator owns the **management** of Ceph
- CSI
  - Ceph CSI driver dynamically **provisions** and **mounts** storage to user application Pods
- Ceph
  - **Data** layer


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

How can I monitor the Ceph storage?



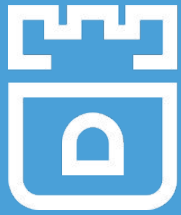
Ceph MGR Dashboard



Prometheus Metrics + Grafana

How can I troubleshoot Rook clusters?

How do we troubleshoot Rook clusters?



- Troubleshooting guide in Rook documentation
  - https://rook.io
- Krew plugin is designed for troubleshooting Rook clusters
  - From simple status to advanced operations
    - Rook & Ceph cluster status
    - Maintenance operations
  - https://github.com/rook/kubectl-rook-ceph

How safe is my data in Ceph?

Ceph data durability

- Ceph designed to be consistent, not eventually consistent
- Data sharded across partitions (AZs), racks, nodes, disks
- Shard replication is configurable
- Proven highly durable
- Even in extreme disasters, data can be recovered manually

What Rook features are we most excited about?

COSI

- Container Object Storage Interface (COSI)
- Like CSI drivers but for object storage
- Pods can request and access buckets (or blob storage)
- Make object storage as cloud-agnostic as block/file
- Alpha in Kubernetes v1.25
- Beta design in progress
- On track to add to Rook for upcoming v1.12 release

Krew Improvements

- Get all cluster status at one place
- Disaster Recovery
  - Reduce Dependency on Ceph Expertise for failure Scenarios
- Install rook-ceph Krew Plugin:
  - `kubectl krew install rook-ceph`

OSD Encryption

- Kinda like “Encryption at REST” for data in Ceph
- Ceph utilizes `dmcrypt` under the hood

Recovering from Disaster

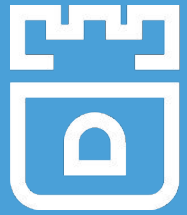
- Disaster is always possible
- Ceph is prepared for disaster
  - No single point of failure
  - Data consistency
- Survive node loss & disk loss
  - Data stays **available** in most scenarios
  - Ceph restores the data to full redundancy automatically
- Even with multiple node or full cluster loss?
  - Ceph can be recovered from disk

Disaster Recovery (DR) features

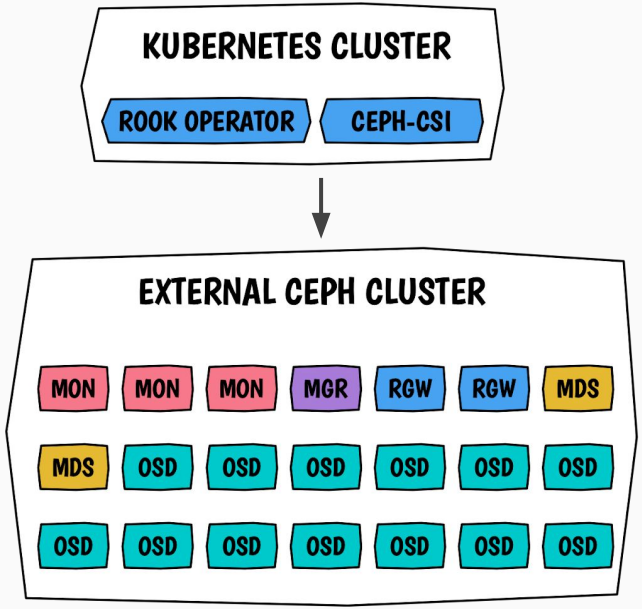
- Ceph supports mirroring data across clusters
  - RBD Mirroring
  - CephFS Mirroring
  - RGW Multisite

What if I have a Ceph cluster deployed outside K8s? Can I connect to it from K8s?

External Cluster Connection



- Connect to a Ceph cluster outside of the current K8s cluster
- Dynamically create Block/File/Object storage consumable by K8s applications



Can I provision a bucket with an S3 endpoint?

Object Storage Provisioning

- Object Bucket Claim (OBC)
  - Similar pattern to a Persistent Volume Claim (PVC)
  - The operator creates a bucket when requested
  - Give access via K8s Secret
- Container Object Storage Interface (COSI)
  - Implementation coming soon…

Can Rook be configured for clients to access storage outside the cluster?

Yes, but …

Requires either `HostNetwork` or `Multus`

How does Rook keep data available during K8s upgrades?

Data Availability during K8s upgrades

- Rook manages `PodDisruptionBudgets` (PDBs) to keep the cluster online even during node or zone outages
- In a cloud environment, all daemons can be easily drained to new nodes

How often does Rook have releases?

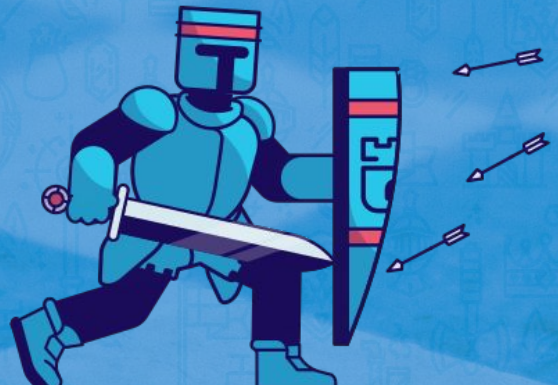
Release Cycle

- Minor releases are about **every 4 months**
  - v1.11 was in March
  - v1.12 potentially in July
- Regular patch releases
  - Usually biweekly unless there is a critical need

Where did the Rook name originate?

“Castle”
Original project name

A secure place to store data



“Is Rook your next move?”

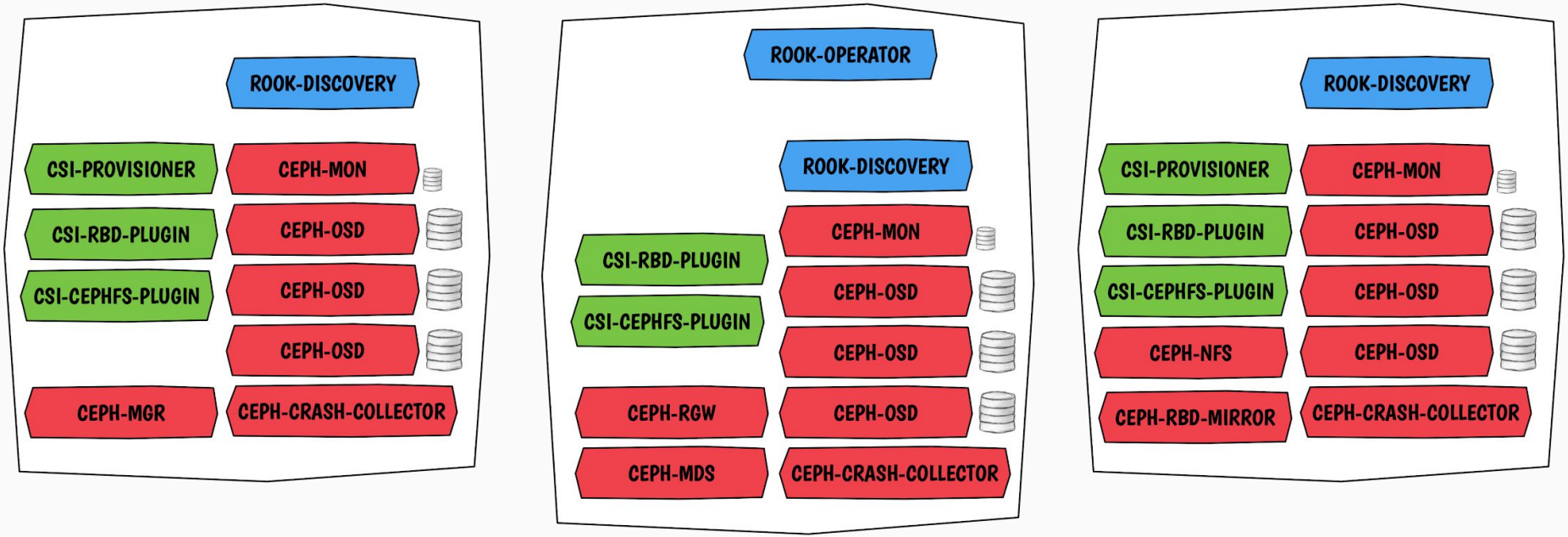
Thank you!

Website	https://rook.io/
Documentation	https://rook.io/docs/rook/v1.11
Slack	https://slack.rook.io/
Contributions	https://github.com/rook/rook
Twitter	@rook_io
Community Meeting	https://github.com/rook/rook#community-meeting

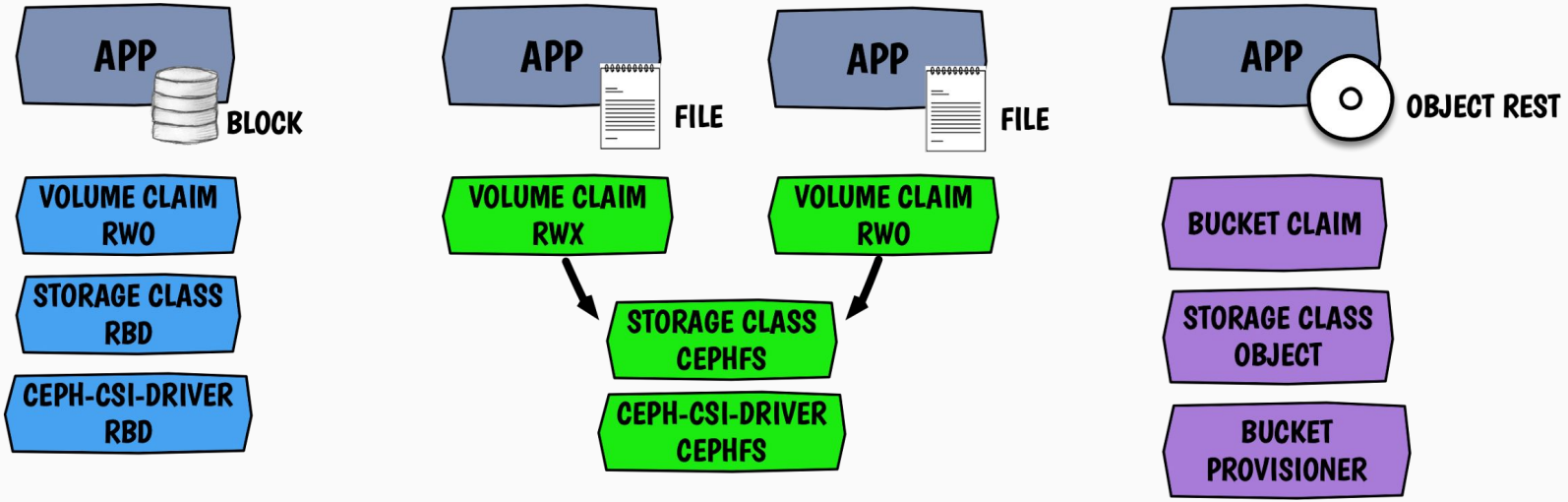

KubeCon
CloudNativeCon
Europe 2023
TiKV

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

Layer 2: CSI Provisioning



Layer 3: Ceph Data Path