Rook: Storage for Kubernetes

KubeCon | CloudNativeCon
North America 2022
BUILDING FOR THE ROAD AHEAD
DETROIT 2022

KubeCon | CloudNativeCon
North America 2022

BUILDING FOR THE ROAD AHEAD

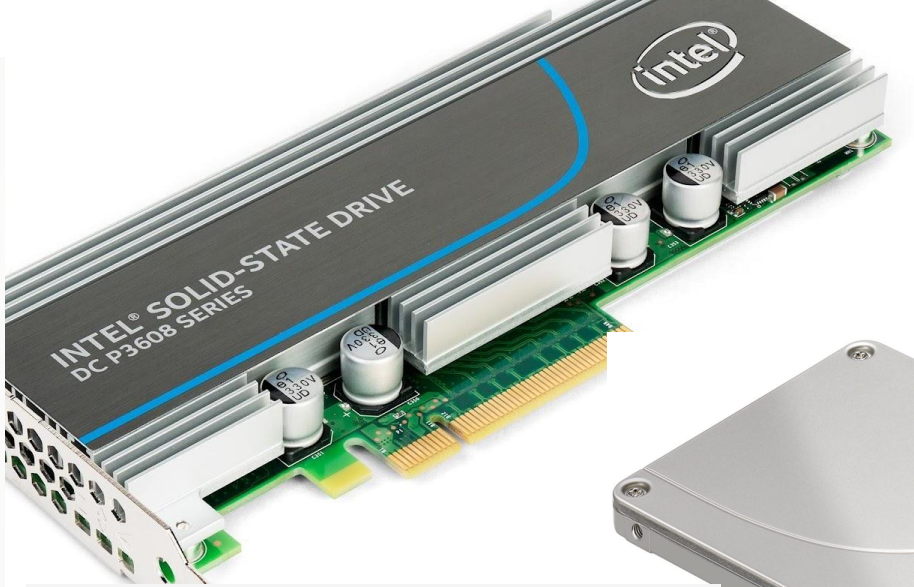
DETROIT 2022

October 24-28, 2021

Alexander Trost
Founding Engineer
Koor

Travis Nielsen
Rook Maintainer
Red Hat

Photo by PublicDomainPictures from Pixabay




TERMINAL BURCHARDKAI
Automation, CI/CD,
etc.

Applications / Containers
(e.g., Kubernetes)

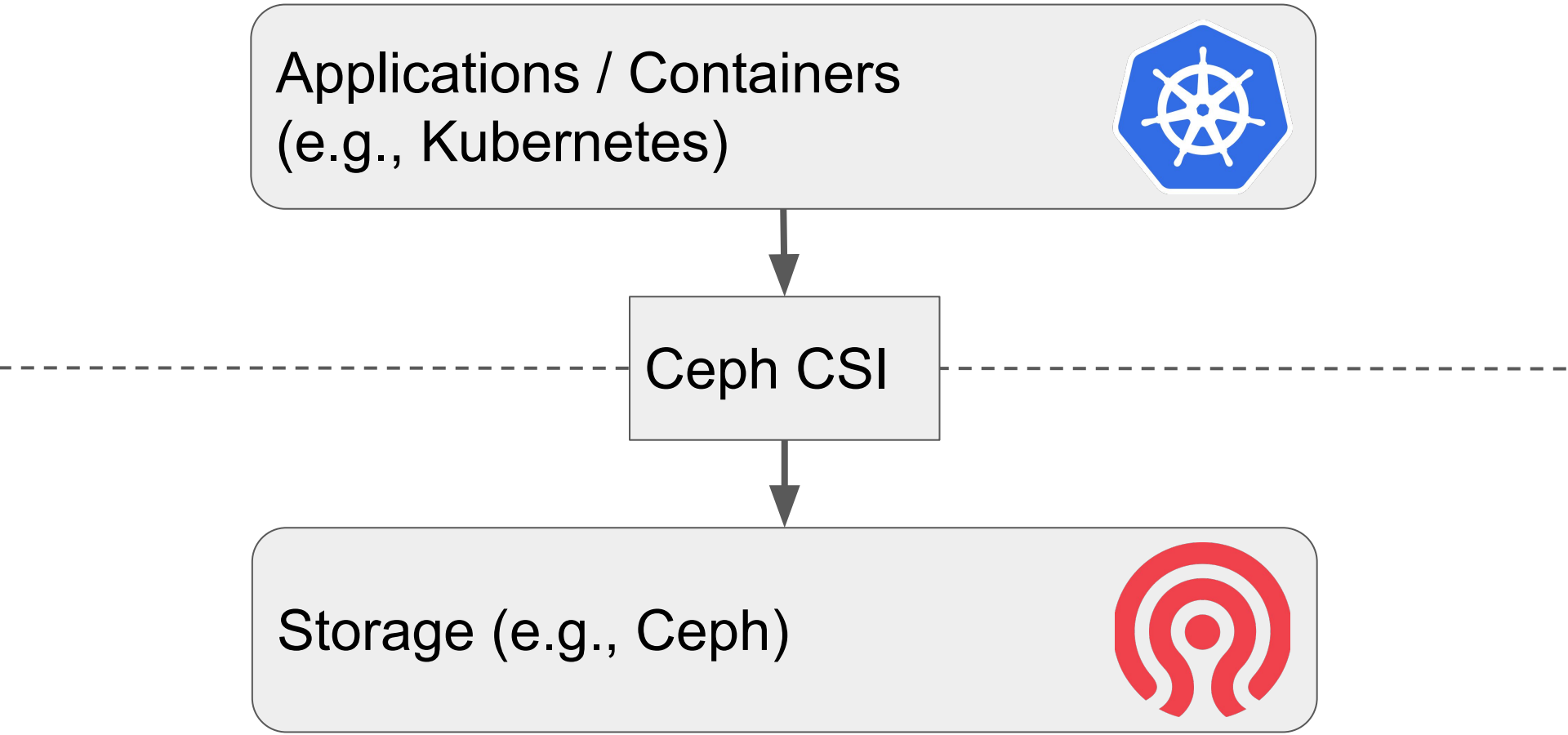
?

Storage (e.g., Ceph)

Container Storage Interface (CSI)


In a Nutshell...

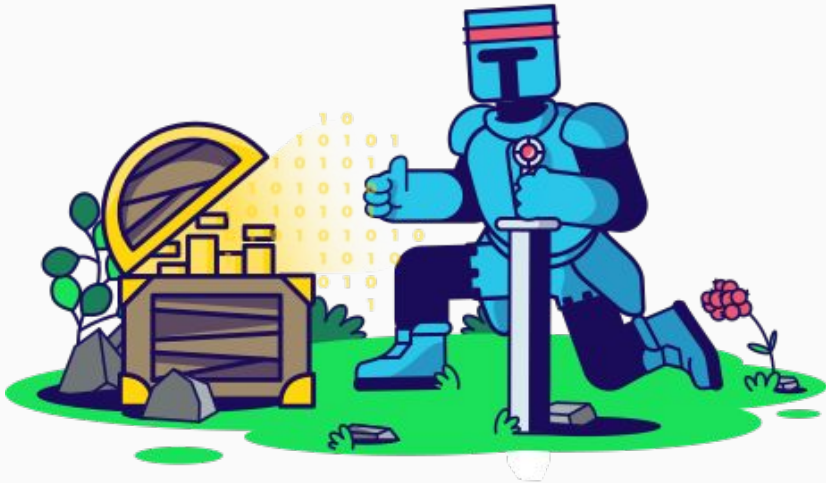
One interface for storage backend and container orchestrator to implement. For users to be able to **consume storage**.


Applications / Containers
(e.g., Kubernetes)
Ceph CSI
Storage (e.g., Ceph)

What is Rook?

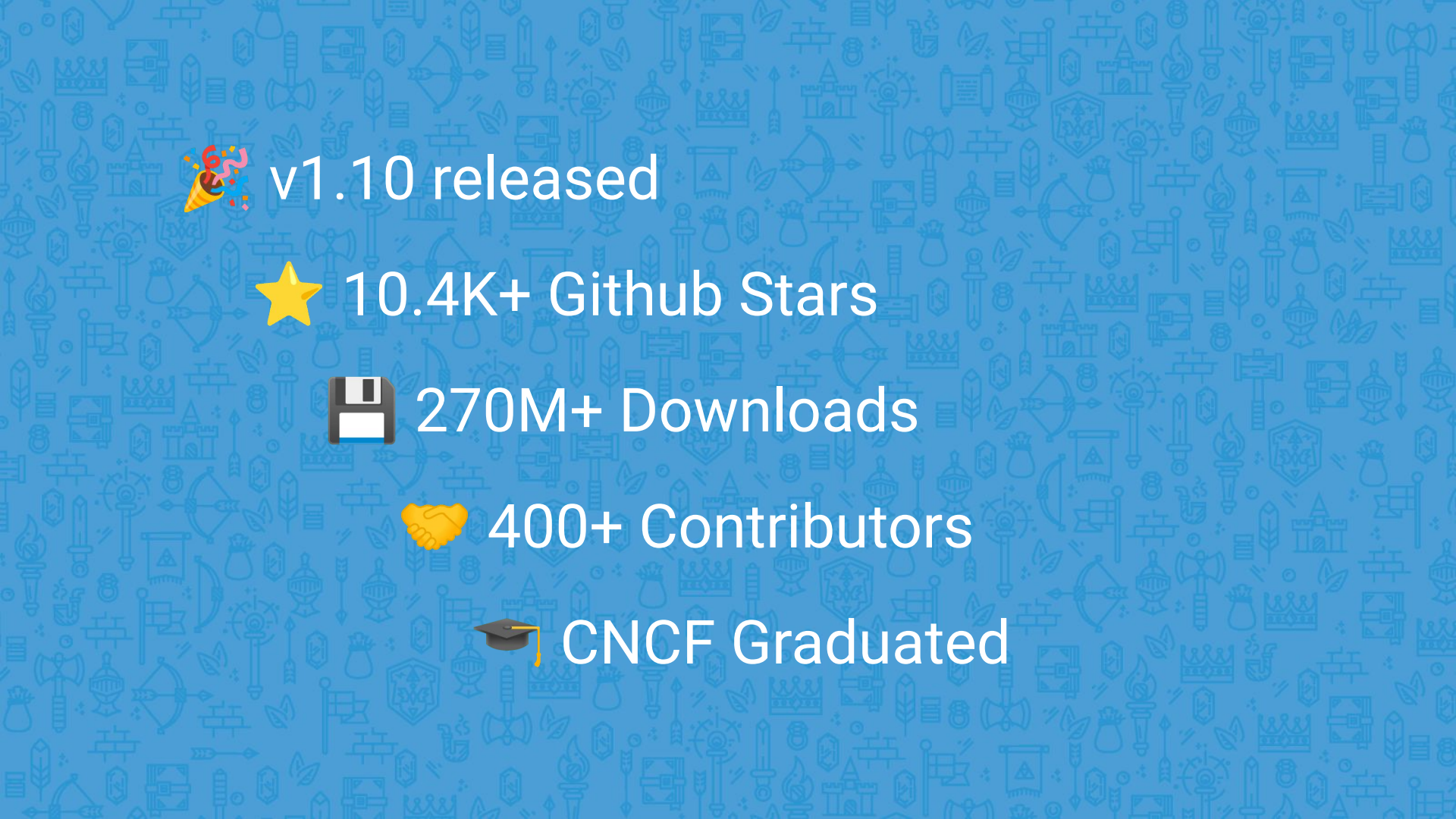
What is Rook?

- Ceph **Operator** for Kubernetes
- Automate
  - Deployment
  - Bootstrapping
  - Configuration
  - Upgrading
- Provision
  - Consume Storage by PVCs



Ceph Storage Platform

- Block Storage (RWO)
  - Ceph RBD
- Shared Filesystem (RWX)
  - Ceph FS
- Object Store (S3)
  - Ceph RGW


🎉 v1.10 released
⭐ 10.4K+ Github Stars
💾 270M+ Downloads
🤝 400+ Contributors
🎓 CNCF Graduated

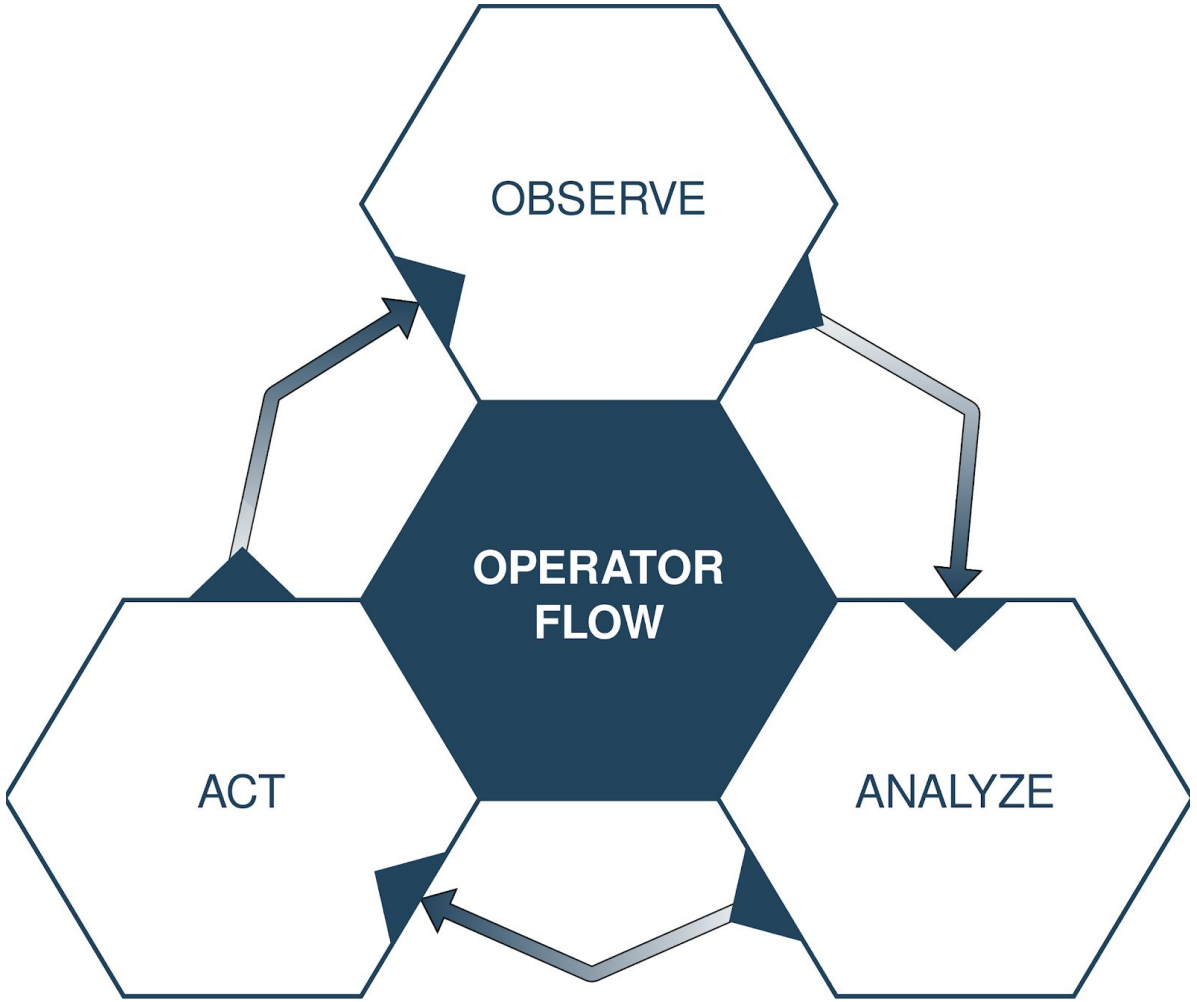
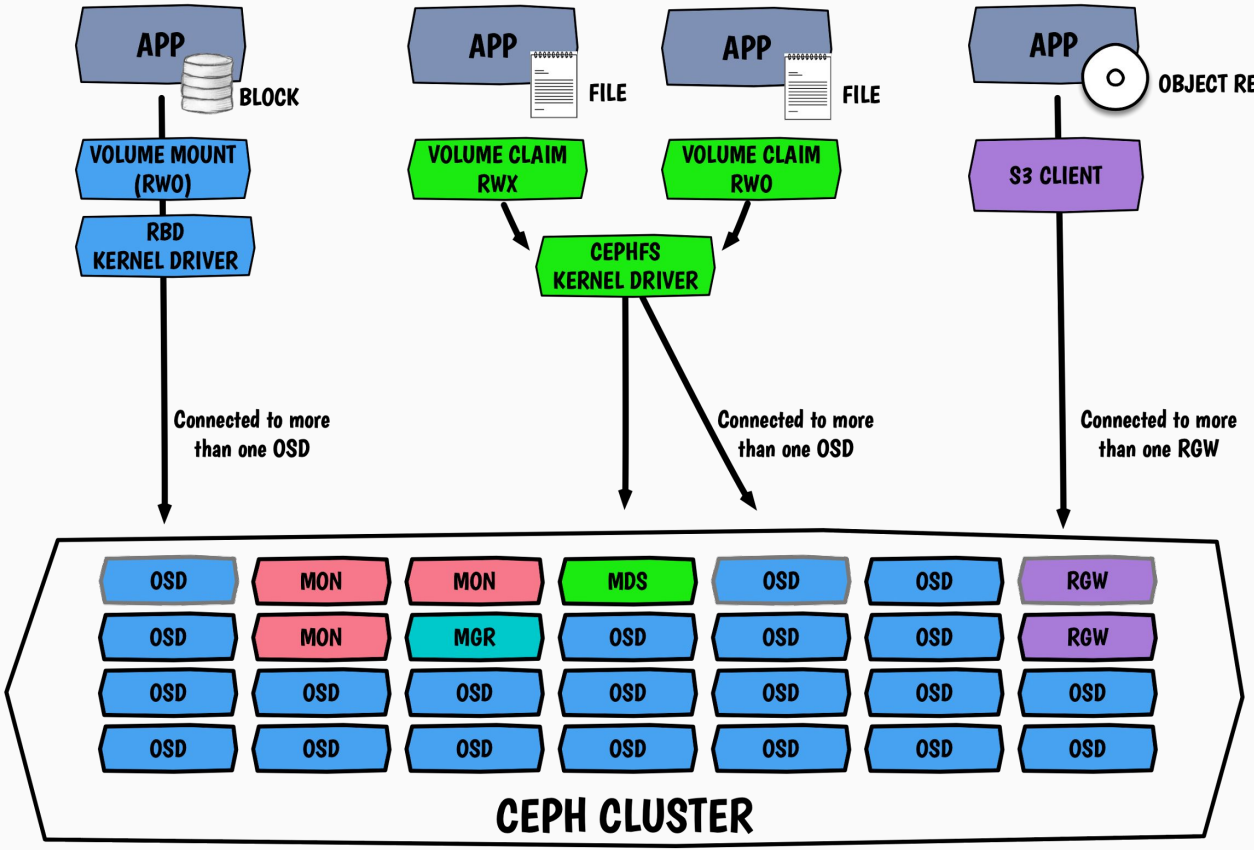
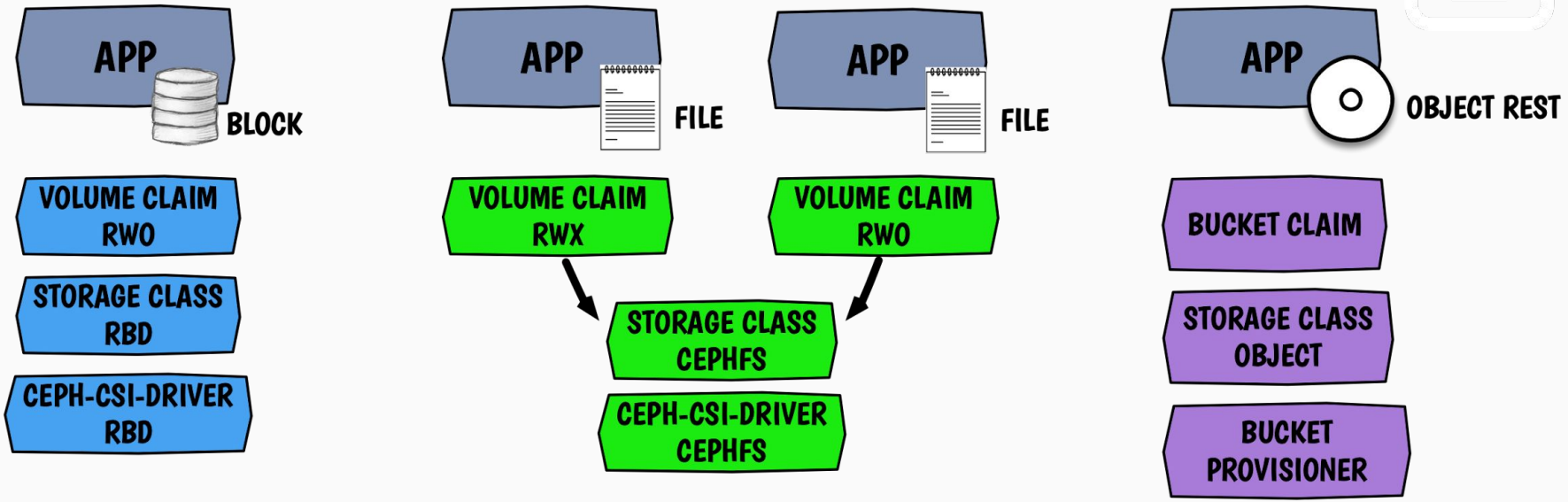
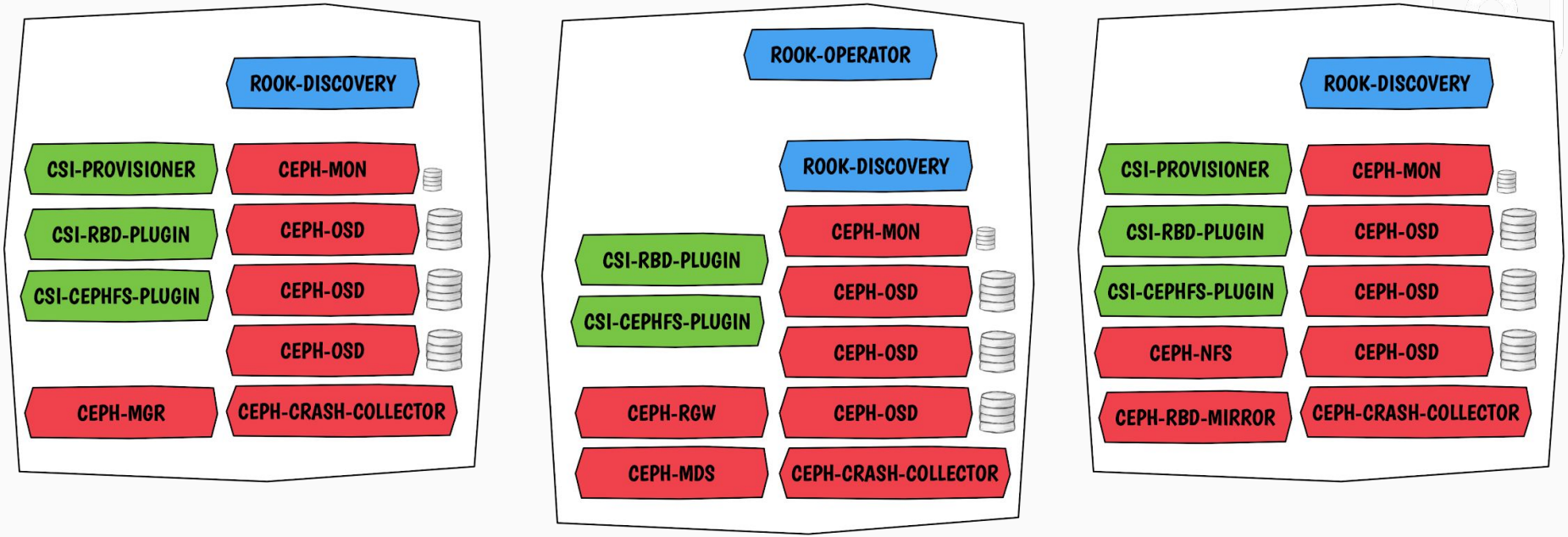

OBSERVE
OPERATOR
FLOW
ACT
ANALYZE

Photo by Gabriel Santos Fotografia from Pexels

Rook Architecture


APP
BLOCK
VOLUME MOUNT (RWO)
RBD KERNEL DRIVER
Connected to more than one OSD
APP
FILE
APP
FILE
VOLUME CLAIM RWX
VOLUME CLAIM RWO
CEPHFS KERNEL DRIVER
Connected to more than one OSD
APP
OBJECT REST
S3 CLIENT
Connected to more than one RGW
OSD
MON
MON
MDS
OSD
OSD
RGW
OSD
MON
MGR
OSD
OSD
OSD
RGW
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
OSD
CEPH CLUSTER


APP
BLOCK
VOLUME CLAIM
RWO
STORAGE CLASS
RBD
CEPH-CSI-DRIVER
RBD
APP
FILE
APP
FILE
VOLUME CLAIM
RWX
VOLUME CLAIM
RWO
STORAGE CLASS
CEPHFS
CEPH-CSI-DRIVER
CEPHFS
APP
OBJECT REST
BUCKET CLAIM
STORAGE CLASS
OBJECT
BUCKET
PROVISIONER


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

Desired State Examples

```yaml
apiVersion: ceph.rook.io/v1
kind: CephCluster
metadata:
  name: rook-ceph
  namespace: rook-ceph
spec:
  cephVersion:
    image: quay.io/ceph/ceph:v17.2.3
    allowUnsupported: false
  dataDirHostPath: /var/lib/rook
```

```
storage:
  useAllNodes: true
  useAllDevices: true
  #deviceFilter:
  config:
    # osdsPerDevice: "1"
  nodes:
    - name: "172.17.4.201"
      devices:
        - name: "sdb"
        - name: "nvme01"
          config:
            osdsPerDevice: "5"
        - name: "/dev/disk/by-id/ata-ST4000DM004-XXXX"
```

```
rook-ceph-mds-myfs-a-5c6f6dbf8c-kqq95          1/1     Running     0               18d
rook-ceph-mds-myfs-b-6d7cd9469c-n5hl2          1/1     Running     0               18d
rook-ceph-mgr-a-c5ff977d-wqbxk                 1/1     Running     0               18d
rook-ceph-mon-b-7f6bf57db6-nlprb               1/1     Running     0               18d
rook-ceph-mon-f-85b77d9f8f-nd724               1/1     Running     0               18d
rook-ceph-mon-g-fc6888db9-mswkq                1/1     Running     0               18d
rook-ceph-operator-6f4fdc4448-xzdd4            1/1     Running     0               16d
rook-ceph-osd-0-6d8db96856-qw6ld               1/1     Running     0               18d
rook-ceph-osd-1-5c76b5466-bncdt                1/1     Running     0               18d
rook-ceph-osd-2-7fcf9745c5-t7vwf               1/1     Running     0               18d
rook-ceph-rgw-my-store-a-6875885f49-c6lcn      1/1     Running     0               18d
rook-ceph-rgw-my-store-a-6875885f49-x4sk5      1/1     Running     0               18d
```

TL;DR

Kubernetes + Ceph = 

Ceph Storage Features

Deployment

- Two Helm Charts
  - Operator chart: RBAC, CRDs, operator
  - Cluster chart: Cluster, Pools, Filesystems, Object Stores, …
- Manifests
  - YAML files for raw access to config

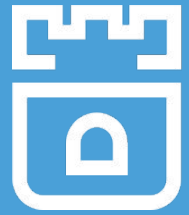
Configurable for Cluster Topologies

- Rook can be easily customized across cluster topologies
- Failure domains: High availability and durability
  - Zones, Data Centers, Racks, ...

Ceph in a Cloud Environment

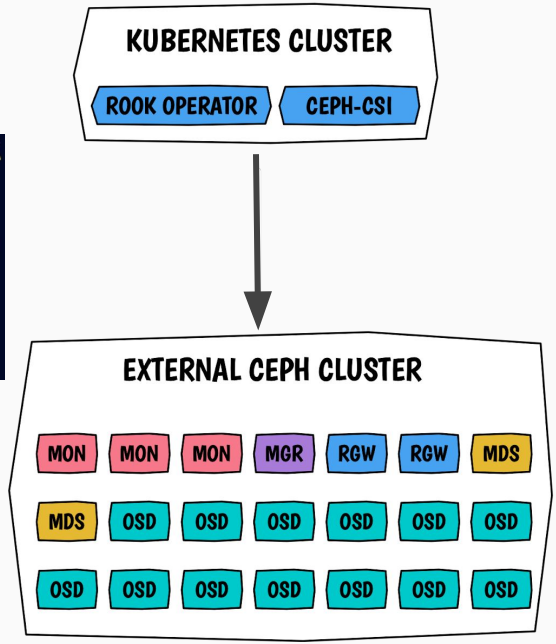
- Consistent Storage Platform wherever K8s is deployed
- Overcome shortcomings of the cloud provider's storage
  - Storage across AZs
  - Slow failover times (seconds instead of minutes)
  - Limitations of number of PVs per node (many more than ~30)
  - Perf characteristics of large volumes
- Ceph Monitors and OSDs run on PVCs
  - No need for direct access to local devices

External Cluster Connection



- Connect to a Ceph cluster that you've configured separately from Kubernetes
- Inject into Kubernetes:
  - Ceph Monitors list
  - Ceph Keyring
  - Ceph Cluster FSID
- Create the cluster-external CR

```yaml
apiVersion: ceph.rook.io/v1
kind: CephCluster
metadata:
  name: rook-ceph-external
spec:
  external:
    enable: true
```



Ceph CSI

- RWO (Ceph RBD)
- RWX (CephFS)
- Snapshots and Clones
- Volume Encryption
- Volume Expansion
- Ephemeral Volumes

Object Bucket Provisioning

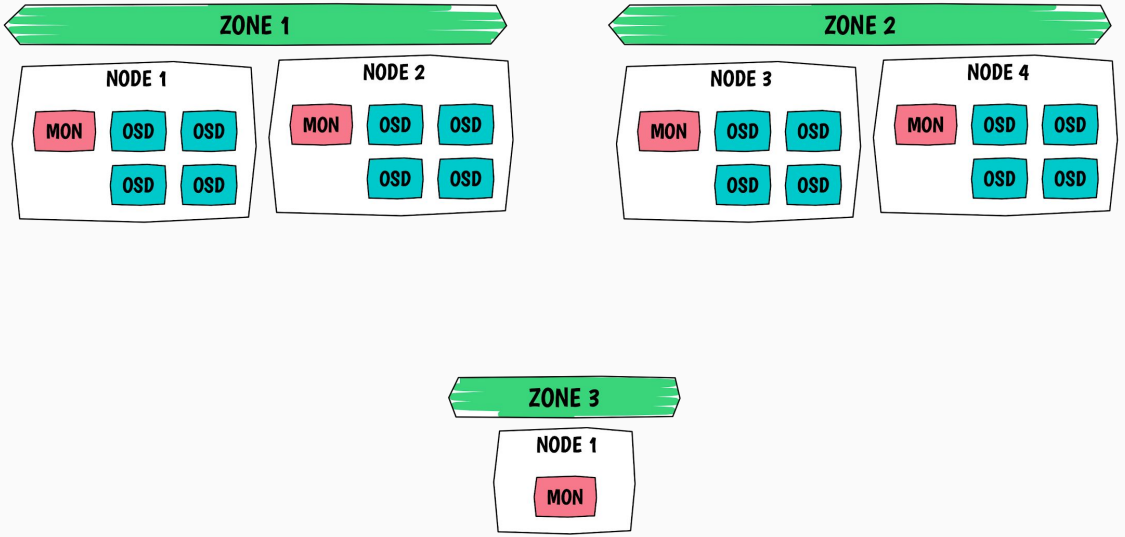
- Define a Storage Class for object storage
- Create an “object bucket claim”
  - The operator creates a bucket when requested
  - Similar pattern to a Persistent Volume Claim (PVC)

KMS Encryption

- OSD encryption keys stored by:
  - KMS HashiCorp Vault
  - IBM KeyProtect
  - Key Management Interoperability Protocol (KMIP)

Stretch Cluster

- Two zones available for storage
- Ceph has greater resiliency to network partitions



Data Mirroring

- Mirror your data to another Ceph cluster
  - Block
  - Shared Filesystem
  - Object (RGW Multisite)
- Asynchronous replication
- Enables DR scenarios

Recent Features 2022
Version: 1.9 & 1.10

Krew Plugin

- Troubleshooting features are very helpful to admins, whether novice or experienced
  - Ideas are flowing for useful commands
- Show health of the cluster
- Enable “debug” pods for advanced operations on Ceph mons and OSDs
- Purge an OSD
- Restart the operator

Documentation Facelift

- MkDocs framework
- Updated look and feel
- Searchable
- Reorganization of content: Tabs, subtopic links, etc

Ceph Updates

- Support constant updates to Ceph data layer
- Added support for Ceph Quincy
- Dropped support for Ceph Octopus

CSI Improvements

- KMIP integration for RBD PVC encryption
- NFS support for snapshot, restore and clone, and volume expansion
- PV and snapshot metadata
- Shallow read-only support without cloning the underlying snapshot

NFS

Support for more enterprise NFS features

- Client ID management via SSSD (tested LDAP)
- Kerberos client authentication support
- Supported only for external clients *for now*

Roadmap

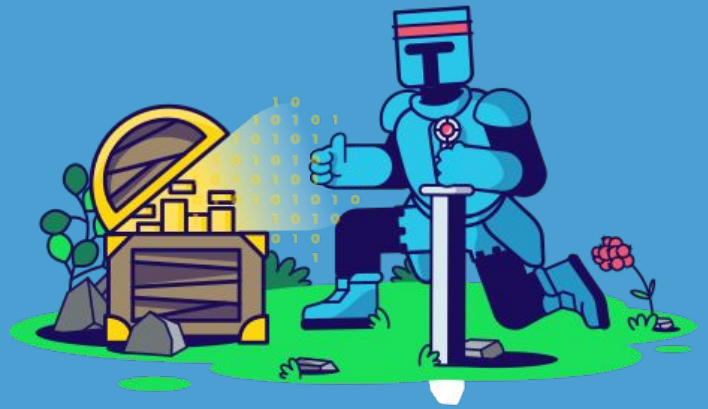
Goals

- Support K8s features
- Support Ceph features
- Respond to community input

Focus Areas

- Implementation of COSI
  - OBCs (Object Bucket Claims) will eventually be deprecated
- Encryption
- Disaster Recovery
- and more…

How to get involved?



Visit the Rook booth at KubeCon!
Contribute to Rook - https://github.com/rook/rook
Slack - https://rook-io.slack.com/ #conferences
Twitter - @rook_io
Community Meetings

Thank you!

https://github.com/rook/rook

https://rook.io/