Open Infrastructure FOUNDATION

INDONESIA OpenInfra Days

Block and Object Storage Solution with Ceph in Production

Talk #4 IOID2021

Aji Arya

Cloud Engineer at **IMTEK**

Jakarta, August 21, 2021

About Me

Aji Arya

Cloud Engineer at **IMTEK**

imtek

@ajiaryar

https://www.linkedin.com/in/ajiarya/

arya@imtek.id

INDONESIA OpenInfra Days

Agenda

- Community
- Ceph
- Deployment Planning
- Block Storage - RBD
- Object Storage - RGW
- Dashboard & Monitoring

INDONESIA OpenInfra Days

Community

/kəˈmjuːnəti/ a group of people who have the same interests, religion, etc.

Open Infrastructure FOUNDATION

BRI EasyStack Biznet GIOCloud boer technology NVIDIA

Join us

INDONESIA OpenInfra Days

Telegram Group: https://t.me/cephid
Current Members: 40

INDONESIA
OpenInfra Days

Ceph

/ˈsɛf/

Open Infrastructure
FOUNDATION

BRI EasyStack Biznet GioCloud boer technology NVIDIA

What is Ceph?

Ceph is open source software-defined storage which provides 3-in-1 access types: **block, object, & file**. You can make storage cluster with **Ceph** on many nodes.

Ceph Logo
https://ceph.io/

Why use Ceph?

- Scalable storage
- Business continuity
- Long term solution
- **Object**, **Block**, and **File** Storage in one unified system

Ceph Logo
https://ceph.io/

Common Ceph Term

- **Ceph Monitor (ceph-mon)**
  The Ceph monitor software, which maintain the cluster state
- **Ceph Manager (ceph-mgr)**
  The Ceph manager software, which provide additional monitoring and interfaces to external monitoring and management systems
- **Object Storage Device (OSD)**
  A physical or logical storage unit used by **ceph-osd** daemon
- **CRUSH**
  Controlled Replication Under Scalable Hashing
- **CRUSH rule**
  The CRUSH data placement rule that applies to a particular pool(s)
- **Pool**
  Pools are logical partitions for storing objects
- **RADOS (Reliable Autonomic Distributed Object Store)**
  The core set of storage software which stores the user's data (MON+OSD).

INDONESIA OpenInfra Days

Deployment Planning

Open Infrastructure FOUNDATION

BRI EasyStack Biznet GioCloud boer technology NVIDIA

Production Checklist (1)

INDONESIA OpenInfra Days

Ceph Monitor + Ceph Manager Hardware

- **6** cores for Ceph Monitor + Ceph Manager Daemon
- **32GB** RAM Per Ceph Monitor + Ceph Manager Daemon (Below 300 OSD)
- **10GbE+** NIC (larger is better)

Ceph OSD Hardware

- **1** core per OSD Daemon
- **4GB+** RAM Per OSD Daemon (more is better)
- **10GbE+** NIC (larger is better)

Production Checklist (2)

Ceph Object Gateway / RADOS Gateway

- **6** cores for RadosGW Daemon
- **32GB** RAM per RadosGW Daemon

Ceph Cluster Component

- **5** Ceph Monitor + Ceph Manager (**2** Nodes fail tolerance)
- **2 - 3** HAProxy + Keepalived Node (For **HA Object Storage**)
- **3 - 5** RADOS Gateway Node (For **HA Object Storage**)
- Enable Monitoring Feature

System Tuning

- Operating System Kernel Tuning

Operating System

This is based on **Ceph Pacific** Release (Latest)

Recommendation for LTS:

- **Ubuntu 20.04** (Focal)
  - Ubuntu Cloud Archive
- **RHEL 8** (Ootpa)
- **Debian Buster**
- **Debian Stretch**

Deployment Methods

Recommended Methods:

- cephadm
- rook (Operator for Kubernetes)

Others Methods:

- ceph-ansible
- ceph-deploy (**deprecated**)
- Ceph Juju
- Manual

Data Distribution

Create CRUSH rule for your data storing method:

Replica

- Store the data with N number of replica
- Faster than Erasure Code

Erasure Code

- Similar to RAID-6
- More Capacity but More Compute Cost
- Split data to some chunks, **data chunk** & **coding chunk**

INDONESIA OpenInfra Days

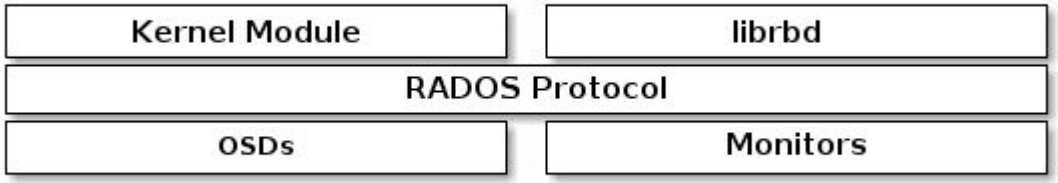
Block Storage - RBD

Open Infrastructure FOUNDATION

BRI EasyStack Biznet GioCloud boer technology NVIDIA

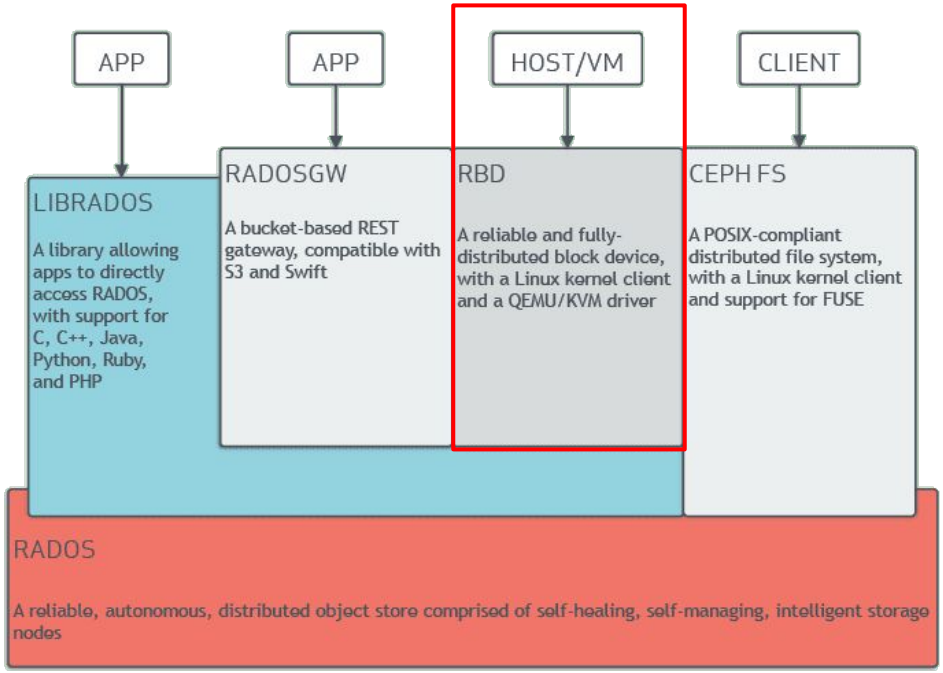
RBD

Ceph block devices are thin-provisioned, resizable, and store data striped over multiple OSDs. Ceph block devices leverage RADOS capabilities including snapshotting, replication and strong consistency.



RBD Architecture
https://docs.ceph.com/en/latest/rbd/

RBD



Ceph Stack
https://docs.ceph.com/en/latest/architecture/

RBD Uses

INDONESIA OpenInfra Days

- Block Storage Backend for **Cloud Platform** (OpenStack, CloudStack, & Proxmox)
- Attach to Linux Machine using RBD Kernel Module (rbd.ko)
- Attach to Windows Server 2019 Machine using **rbd-wnbd**

INDONESIA OpenInfra Days

Object Storage - RGW

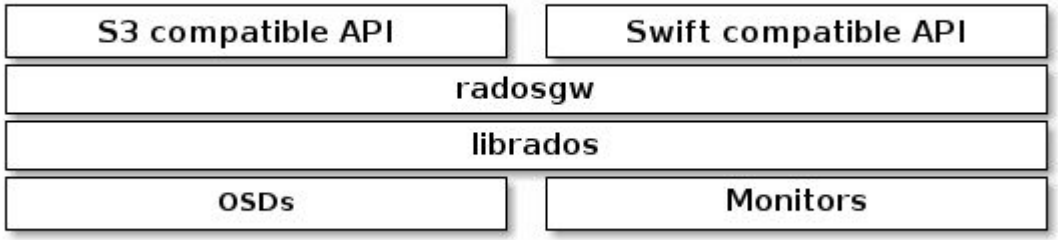
Open Infrastructure FOUNDATION

BRI EasyStack Biznet GioCloud boer technology NVIDIA

RADOS Gateway

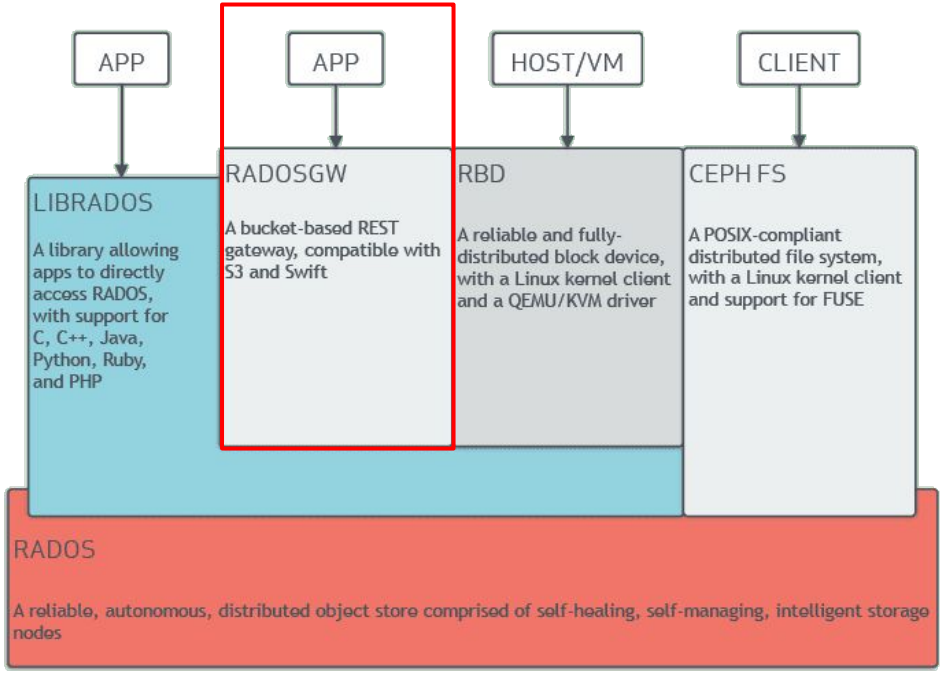
INDONESIA OpenInfra Days

RADOS Gateway is HTTP server for interacting with a Ceph Storage Cluster. Since it provides interfaces compatible with OpenStack Swift and Amazon S3, the Ceph Object Gateway has its own user management.



RADOSGW Architecture
https://docs.ceph.com/en/latest/radosgw/

RADOS Gateway

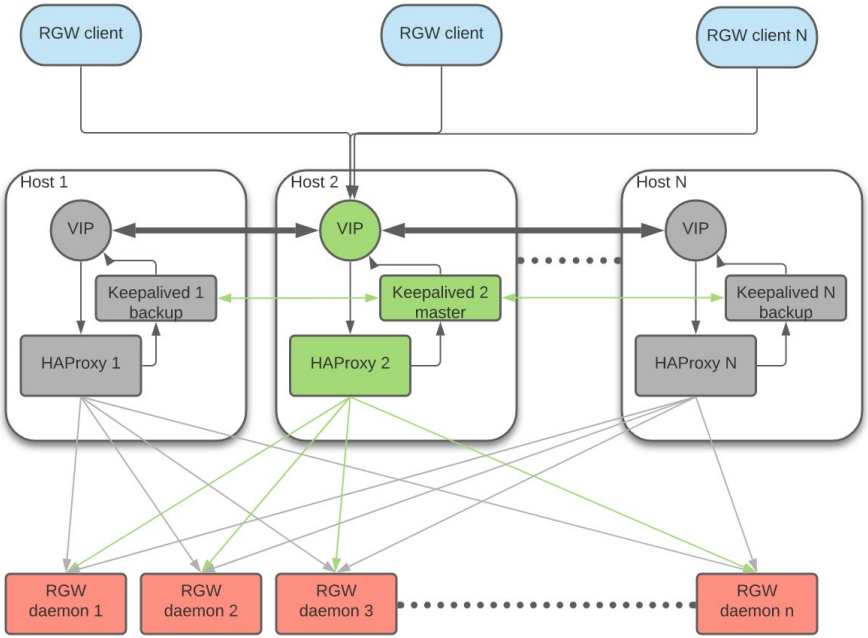


Ceph Stack
https://docs.ceph.com/en/latest/architecture/

High Availability (1)

- HAProxy
- Keepalived
- Multiple RGW Daemon

High Availability (2)



HA RADOS Gateway
https://docs.ceph.com/en/latest/_images/HAProxy_for_RGW.svg

RGW Uses

- S3 Object Storage for App (Thanos, NextCloud, OwnCloud, etc)
- Integrate with OpenStack Keystone for Enable Swift

OpenStack + Ceph

INDONESIA OpenInfra Days

Category	OpenStack Service	Ceph Solution
Object Storage	Swift	RGW Swift Interface
Ephemeral Storage	Nova	RBD
Persistent Storage	Cinder	RBD
Image Storage	Glance	RBD, RGW S3 Interface
Shared Filesystems	Manila	CephFS

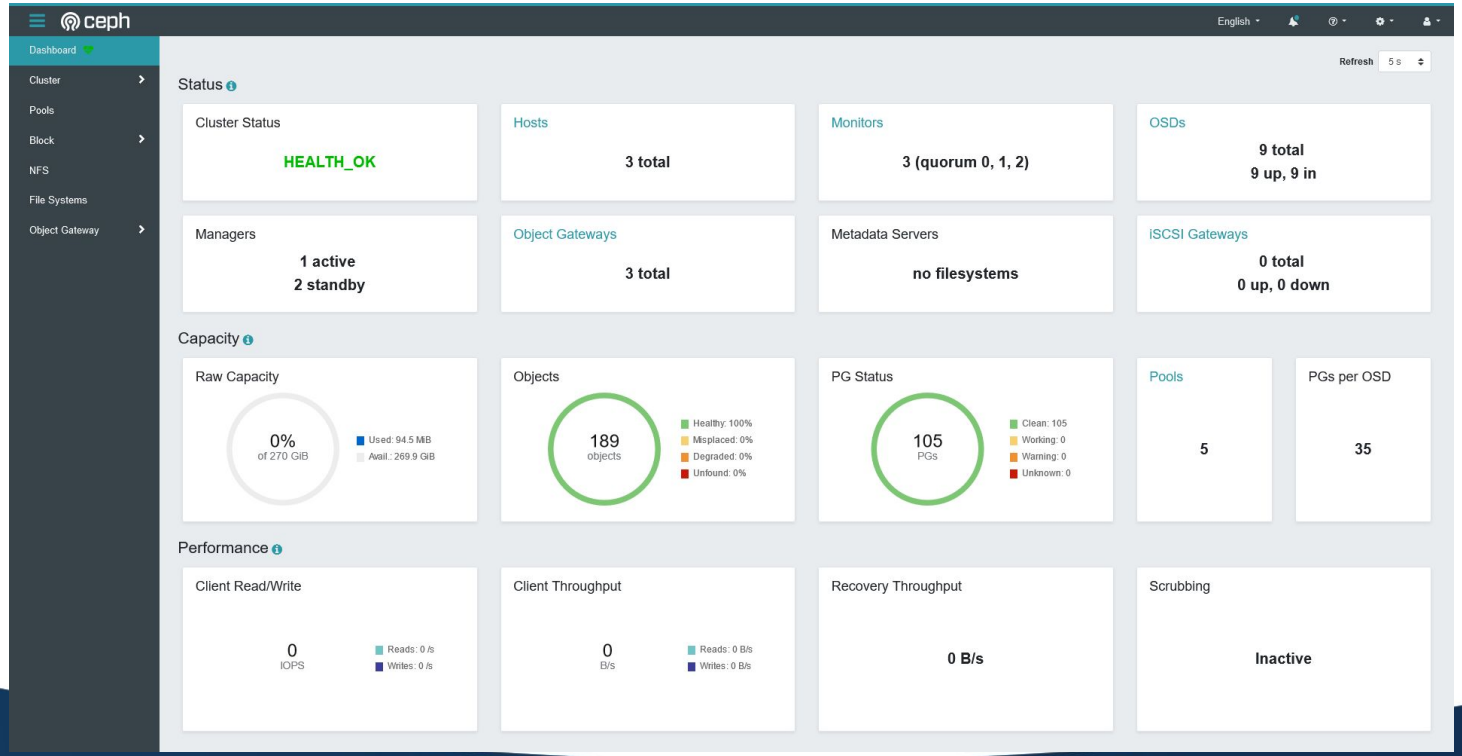
INDONESIA OpenInfra Days

Dashboard & Monitoring

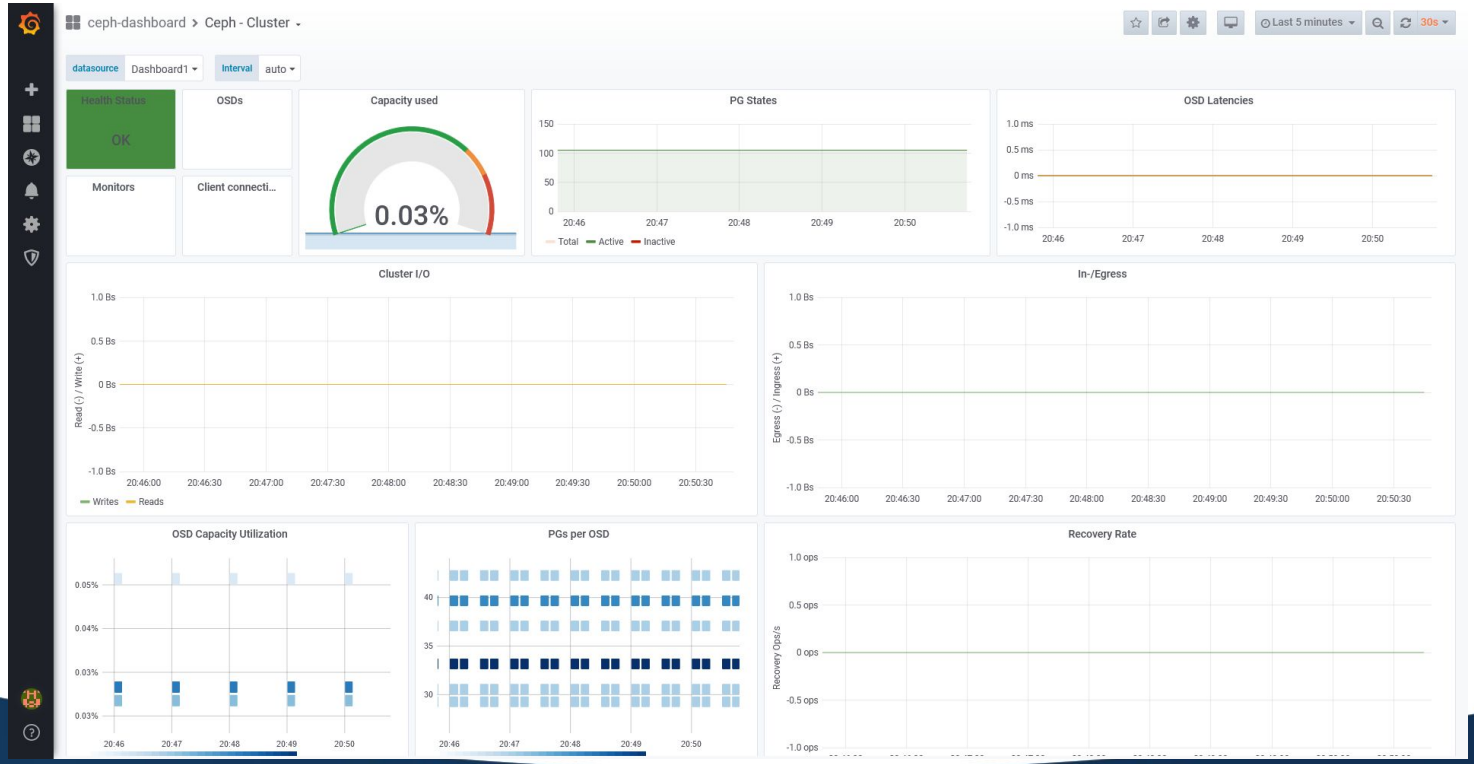
Open Infrastructure FOUNDATION

BRI EasyStack Biznet GioCloud boer technology NVIDIA

Ceph Dashboard



Grafana Dashboard



Sponsored by:

Open Infrastructure
FOUNDATION

Open Networking
Indonesia

Hosted by:

OpenStack Indonesia
Indonesia OpenStack Foundation Community
www.openstack.id

Community Partners:

sdX Indonesia

openSUSE ID

KLAS

gimpscape ID

CLOUD NATIVE INDONESIA

Adaptive Network Laboratory

NETWORK EVOLUTION

Open Infrastructure FOUNDATION

INDONESIA OpenInfra Days

Thanks!

Do you have any questions?
Need experts?

arya@imtek.id (Personal)
info@imtek.id (Business)
+62 816-232-262
imtek.id