

Ceph Mon: a technical overview

Paolo VIOTTI

25th March 2015

Abstract

Ceph is a free software storage platform designed to provide object, block and file storage using computer clusters running on commodity hardware. Ceph's main design goals include high scalability, fault tolerance and low maintenance requirements. This document provides an in-depth technical overview of the design of Ceph Monitor, i.e. the Ceph component in charge of maintaining a map of the cluster along with authorization information.

1 Ceph: an introduction

Ceph [1] is a free software storage platform that provides object, block and file storage using computer clusters running on commodity hardware. Since its origin as a research project around 2006, Ceph has undergone constant and substantial development, thus gaining popularity which is reflected by an ever increasing adoption as storage backend in state-of-the-art high-end computing systems [2].

As illustrated in Fig. 1, Ceph exposes to application clients multiple APIs:

- a POSIX-compatible distributed file system (**Ceph FS**), built as Linux kernel module or user-space FUSE client;
- a REST-based storage gateway (**RADOSGW**) compatible with OpenStack Swift and Amazon S3;
- a block storage device (**RDB**) suitable for virtualization platforms making use of kernel virtualization technologies such as QEMU or KVM.

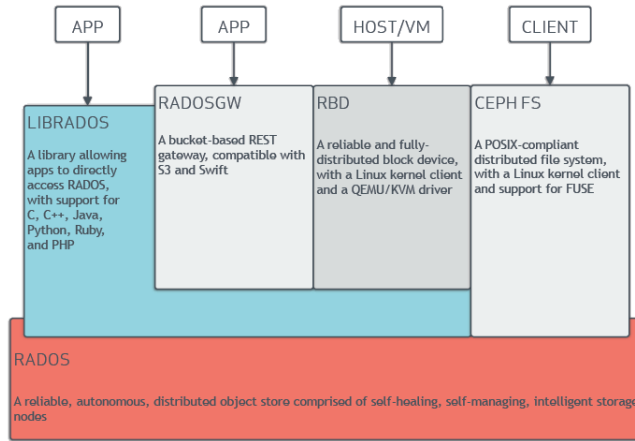


Figure 1: Ceph application stack

Beside, an application developer may even hook into the low level API exposed through **librados** and offered in several programming languages in order to directly connect with **RADOS** (*Reliable Autonomic Distributed Object Store*), i.e. the inner object storage layer that acts as foundation of the interfaces mentioned above.

The Ceph cluster consists of different components, running as distributed daemons:

- Cluster monitors (*ceph-mon*, or Mon) that keep track of active and failed cluster nodes;
- Metadata servers (*ceph-mds*) that store the metadata of inodes and directories;
- Object storage devices (*ceph-osd*) that actually store data on local filesystems;
- RESTful gateways (*ceph-rgw*) that expose the object storage layer as an interface compatible with Amazon S3 or OpenStack Swift APIs.

2 Ceph Mon

During runtime operations, Ceph OSD Daemons check up on other Ceph OSD Daemons and report their findings to the Ceph Monitor. You do not have to provide any settings.

2.1 Architecture

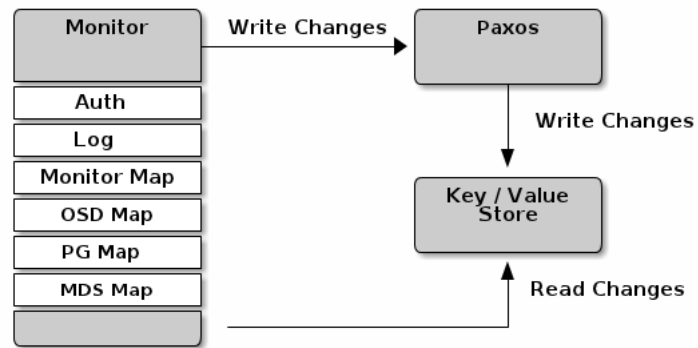


Figure 2: Ceph Monitor high-level architecture

2.2 Code

2.3 Additional notes

References

- [1] Ceph storage platform. <http://ceph.com/>
- [2] OpenStack User Survey Insights: November 2014.
<http://perma.cc/367D-G5YN>