

Ceph: An Open Source Object Store

Evan Harvey, Gustavo Rayos, Nicholas Schuchhardt
Mentors: David Bonnie, Christopher Hoffman, Dominic Manno

Purpose: To test the performance and viability of Ceph as an object storage solution.

What is Ceph?

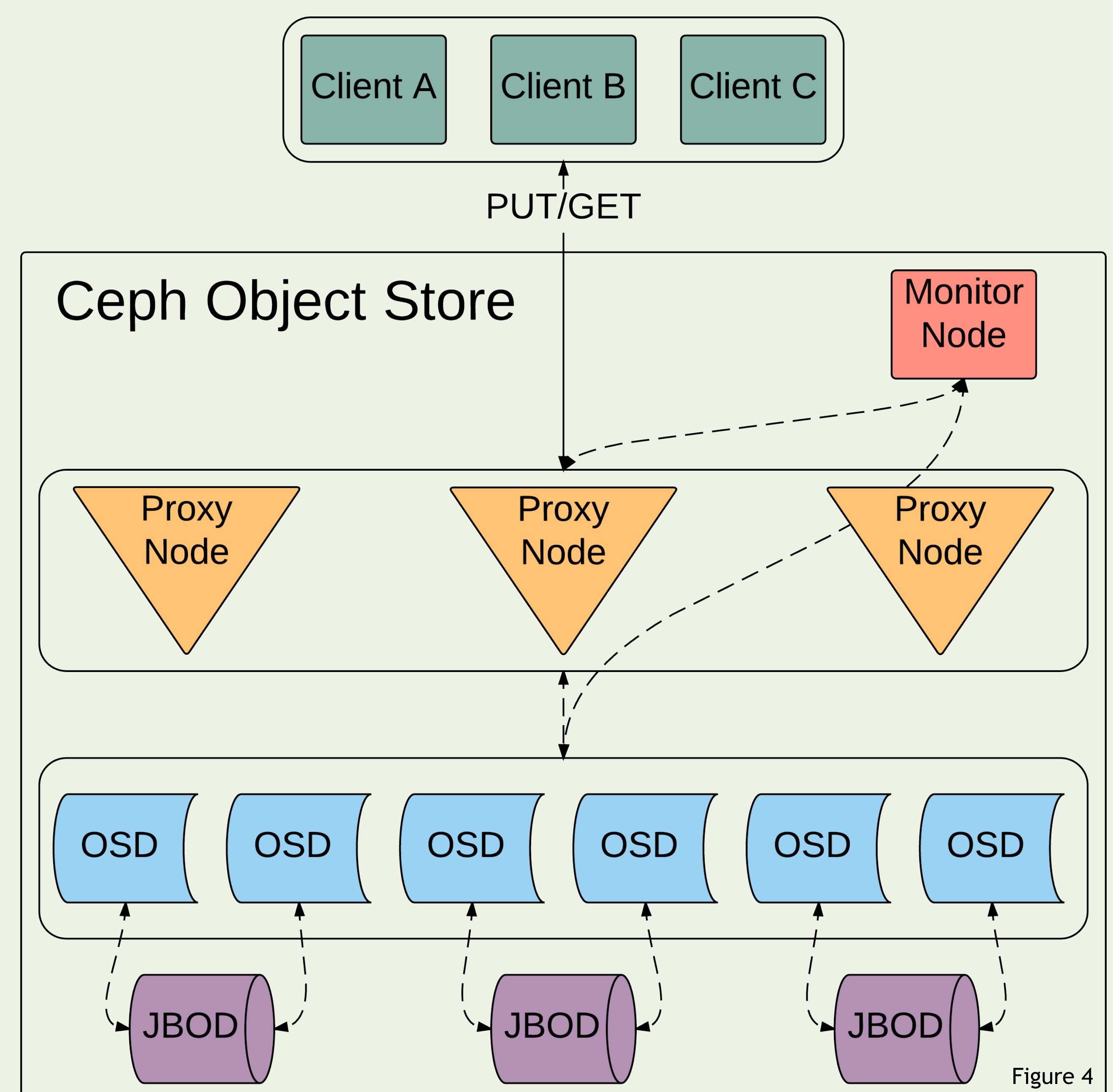
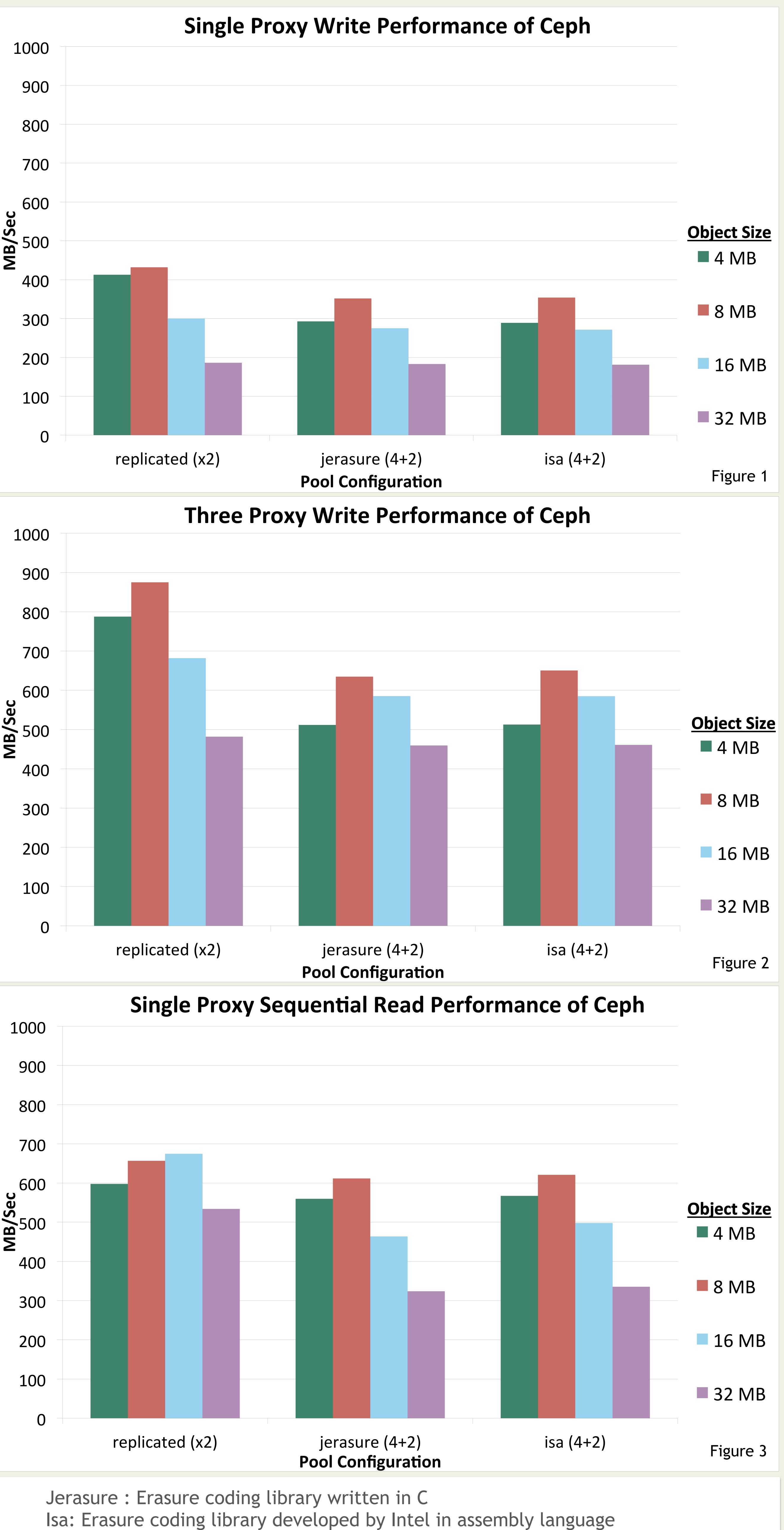
Ceph is an open source object store and filesystem that is designed to be scalable to the exabyte level. Ceph also provides higher resilience and scalability via erasure coding techniques. It uses pools and placement groups in order to specify where and how data is stored. A pool is a logical volume for grouping placement groups. Placement groups are used to collect objects since it is resource intensive to track individual object metadata and placement.

Test Environment

Ten HP ProLiant nodes compose the Ceph object store in Figure 4. The proxy nodes are running a RADOS gateway and apache server providing clients with a HTTP interface to the object store. Object storage devices (OSDs) are connected to 135 hard drives formatted as EXT4. Each of the drives are capable of read and write speeds around 100 megabytes per second. The object store has a capacity of 540 terabytes.

Results

During benchmarking hard drive failure occurred, and the object store was unable to recover on its own or with manual intervention. As a result, the data was lost during reconfiguration. Shutting the object store down improperly also resulted in data loss. Once the failed drives had been removed from the cluster the object store performed well.



Conclusions

At the moment, Ceph is difficult to install, configure and administer. Additionally, Ceph did not provide the stability required for a production environment. The support of, and performance with, erasure coded pools in Ceph is on the right track toward a performant, scalable, and resilient exabyte level filesystem. As of now, the use of Ceph as an object storage solution is not advised without further research.

Future Work

- Determine bottleneck of performance testing
- Further investigate pool configurations
- Explore Ceph monitoring solutions
- Compare ZFS with BTRFS and XFS with EXT4
- Investigate the performance and implementation of the Ceph block device