



Contribution ID: 402

Type: oral presentation

POSIX and Object Distributed Storage system – performance Comparison studies and real-life usage in an experimental data taking context leveraging OpenStack/Ceph.

Thursday, 16 April 2015 11:15 (15 minutes)

The STAR online computing environment is an intensive ever-growing system used for first-hand data collection and analysis. As systems become more sophisticated, they result in a more detailed dense collection of data output and inefficient limited storage systems have become an impediment to fast feedback to the online shift crews relying on data processing at near real-time speed. Motivation for a centrally accessible, scalable and redundant storage solution was led from an expansion of data growth and user data processing necessity. However, standard solutions (NAS, SAN ··· GPFS) are “expensive” solutions and it became clear a balance of affordability and cost effectiveness was needed. Furthermore, the vast amount of sparse and distributed storage (disk attached to individual nodes) made the aggregation of the storage an attractive path. Yet, as shift crews are often composed of novice members sustaining the daily operations, providing a POSIX compliant interface within a standard namespace was, for this environment, a strong requirement for any retained solution.

The acquisition of reused hardware has offered STAR an opportunity to deploy a storage strategy at minimal to no cost. We have analyzed multiple open source object oriented cloud inspired solutions and a POSIX compliant storage system. Openstack Swift Object Storage and Ceph Object Storage were put to the test with single and parallel I/O tests emulating real world scenario for data processing and workflows. The Ceph file system storage, offering a POSIX compliant file system mounted similar to an NFS share which maintain owner/group permissions and historical lookup, was of particular interest as aligned with our requirements and was retained as our solution.

The Ceph storage system will allow scalability and redundancy without user or system interruption. Initial configuration for user setup is minimal while maintaining security and data integrity across the entire cluster. A distributed storage system becomes an interconnected web of machines utilizing load balancing and redundancy, however if a subset of machines or all machines go down data loss is unlikely to occur. The expansion of a cloud storage system is trivial and can allow system administrators to add capacity and load balancing in a complete transparent manner.

In this report, we will review the necessary steps and requirements for our system, tools leveraged for performance testing and present comparative IO performance results between Swift and Ceph Object storage approach in similar context as well as a cross comparison of performance between the Object and POSIX compliant Ceph approaches. We will also discuss the benefit of a backbone private network for enhanced performance and fast communication between the distributed storage system components and scalability considerations for Meta-Data access and report on actual user’s experience in a data taking environment. Finally, we will also present in great details the essential steps necessary to setup a Ceph storage cluster, including tweaks and hardware changes made along the way –the recipes we will present are not always easy to find in the manual and we hope our presentation will serve well the community’s interest for distributed storage solutions.

Primary authors: Dr LAURET, Jerome (BROOKHAVEN NATIONAL LABORATORY); Mr POAT, Michael

(Brookhaven National Laboratory); BETTS, Wayne (Brookhaven National Laboratory)

Presenter: Mr POAT, Michael (Brookhaven National Laboratory)

Session Classification: Track 3 Session

Track Classification: Track3: Data store and access