Contribution ID: **25** Type: **Oral Presentation**

Data I/O at edge sites for traditional experiments in a distributed system

Friday, March 25, 2022 2:10 PM (20 minutes)

"One platform, multi Centers" is a distributed computing platform in China managed by manpower of computing center IHEP. It consists of 10 distributed computing centers which belongs to HEP related institutes and departments. The computing center of IHEP at Beijing and the big data center of IHEP-CSNS-branch at Guangdong Province contribute to 90% of its computing and storage resources, while the other small and medium scale sites contribute the rest 10%. The platform is also capable of adding opportunistic computing resources rented from public cloud for the peak of data processing requirement.

In such a system, the small and medium sites and public cloud are designed as edge sites which have only computing resources and volatile disk storage. Therefore, remote data I/O with the two data centers is a necessity. For modern HEP experiments which access data via XRootD protocol and have mature data management systems such as RUCIO, technology stack adapted from WLCG can be reused directly. However, for traditional experiment which relies on POSIX I/O protocol and organizes dataset with namespaces of distributed file systems, data I/O at edge sites is a new and tricky task. Many HEP experiments at China belongs to this category, and have the intention to join the distributed computing platform.

We propose a solution for data I/O at edge sites of traditional experiments. This solution re-use members of XROOTD family as communication protocol over WAN, data proxy, read only cache manager and FUSE file system. Therefore, all the sites of the platform can see a same set of file systems of the two central data centers. The cache resources managed by XCache can be shared by clients of multiple file systems and experiments at the same time. A user-mapping plugin adopted from OSG is used on proxy nodes. With this plugin user identities embedded in XRootD requests can be read out and translated to local POSIX usernames. Currently, small size output is transferred back by HT-Condor while big size output has to be uploaded manually by using xrdcp. To overcome the inconvenience, a writable data cache and a data synchronization mechanism is imperative and under development.

Results of small scale tests with BESIII simulation and reconstruction at CSNS sites will be presented at the end of the presentation. It validates the feasibility and performance acceleration effect of our design.

Primary authors: LI, Haibo (Chinese); WANG, Lu (Institute of High Energy Physics, Chinese Academy of Scicece); HU, Qingbao (IHEP); CHENG, Yaodong (IHEP, CAS); BI, Yujiang (IHEP)

Presenter: WANG, Lu (Institute of High Energy Physics, Chinese Academy of Scicece)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data