Contribution ID: **102** Type: **Poster**

Distributed Metadata Management of Mass Storage System in High Energy of Physics

Tuesday 11 October 2016 16:30 (15 minutes)

The new generation of high energy physics(HEP) experiments have been producing gigantic data. How to store and access those data with high performance have been challenging the availability, scalability, and I/O performance of the underlying massive storage system. At the same time, a series of researches focusing on big data have been more and more active, and the research about metadata management is one of them. Metadata management is quite important to overall system performance in large-scale distributed storage systems, especially in the big data era. Metadata performance would produce a big effect on the scalability, availability and high performance of the massive storage system. In order to manage metadata effectively, so that data can be allocated and accessed efficiently, we design and implement a dynamic and scalable distributed metadata management system to HEP mass storage system.

In this contribution, the open source file system Gluster is reviewed and the architecture of the distributed metadata management system are introduced. Particularly, we discuss the key technologies of the distributed metadata management system and the way to optimize the metadata performance of Gluster file system by modifying the DHT(Distributed Hash Table) layer. We propose a new algorithm named Adaptive Directory Sub-tree Partition(ADSP) for metadata distribution. ADSP divides the filesystem namespace into sub-trees with directory granularity. Sub-trees will be stored on storage devices in flat structure, whose locality information and file attributes are recorded as extended attributes. The placement of sub-tree is adjusted adaptively according to the load of metadata cluster so that the load balance could be improved and metadata cluster could be extended dynamically. ADSP is an improved sub-tree partition algorithm with low computational complexity, also easy to be implemented. Experiments show that ADSP achieves higher metadata performance and scalability compared to Gluster and Lustre. The performance evaluation demonstrates the performance of metadata of Gluster file system is greatly improved. We also propose a new algorithm called Distributed Unified Layout(DULA) to improve dynamic scalability and efficiency of data positioning. A system with DULA could provide uniform data distribution and efficient data positioning. DULA is an improved consistent hashing algorithm which is able to locate data in O(1) without the help of routing information. Experiments prove that the better uniform data distribution and efficient data access can be achieved by DULA. This work is validated in YBJ experiment. In addition, three evaluation criteria of hash algorithm in massive storage system are presented. And a comparative analysis of legacy hash algorithms has been carried out in both theory and software simulation according to those criteria. The results of that provide the theoretical basis for the choice of hash algorithm of DULA.

Primary Keyword (Mandatory)

Storage systems

Secondary Keyword (Optional)

Distributed data handling

Tertiary Keyword (Optional)

Primary author: HUANG, Qiulan (Chinese Academy of Sciences (CN))

Co-authors: SHI, Jingyan (Chinese Academy of Sciences (CN)); CHENG, Yaodong (IHEP)

Presenter: HUANG, Qiulan (Chinese Academy of Sciences (CN))

Session Classification: Posters A / Break

Track Classification: Track 4: Data Handling