Contribution ID: **20** Type: **Poster Presenation**

Visual analysis system for large-scale data storage

With the development of new generation high energy physics experiment facilities, a large amount of data has been produced, which brings great challenge to the storage management of large-scale data. Traditional storage system can't find what data need to be cleaned up in time. Idle data takes up a lot of storage space, which leads to poor utilization of storage system. Storage system often adopts hierarchical storage architecture, the higher the hierarchy, the faster the access speed, the smaller the capacity. Traditional system can't know which level the file to be stored should be placed on. When people suddenly find out that the data is placed in the wrong storage location , they often need to spend a large amount of computing resources, manpower and time to transfer large amounts of data. To solve these problems, this paper put forward to establish a visual analysis system for large-scale data storage, which could manage files dynamically. The machine learning is used to accurately predict the storage location of the files to be stored by using the file information. Fast index mechanism is established, which supports fuzzy check for file information. Collecting file information is critical for these tasks. There are several ways to collect information: accessing log, EOS(EOS open storage) metadata dump and multi-thread scan. This paper discusses the advantages and disadvantages of these methods. File information is written in time series database in bulk, and InfluxDB is used here. By writing in database, data can be quickly extracted for analysis and processing compared to getting file information from disk. Using the file information collected, the monitoring system can display the usage of the storage system in real time according to various attributes, so as to manage the storage data timely. Finally, this paper introduces how machine learning is used to improve the efficiency of the storage system.

Summary

This paper put forward to establish a visual analysis system for large-scale data storage, which could manage files dynamically. The machine learning is used to accurately predict the storage location of the files to be stored by using the file information. Fast index mechanism is established, which supports fuzzy check for file information.

Primary authors: Mr HU, Qingbao (IHEP); Mr ZHANG, Wentao (IHEP); Mr CHENG, Yaodong (IHEP, CAS)

Presenter: Mr ZHANG, Wentao (IHEP)

Track Classification: Big Data & Data Management