Contribution ID: **358**   Type: **Oral**

Automated and Intelligent Data Migration Strategy in High Energy Physical Storage Systems

Monday 11 March 2019 18:00 (20 minutes)

As a data-intensive computing application, high-energy physics requires storage and computing for large amounts of data at the PB level. Performance demands and data access imbalances in mass storage systems are increasing. Specifically, on one hand, traditional cheap disk storage systems have been unable to handle high IOPS demand services. On the other hand, a survey found that only a very small number of files have been active in storage for a period of time. Most files have never been accessed. Some enterprises and research organizations are beginning to use tiered storage architectures, such as tape, disk or solid state drives to reduce hardware purchase costs and power consumption.

As the amount of stored data grows, tiered storage requires data management software to migrate less active data to lower cost storage devices. Thus an automated data migration strategy is needed. At present, automatic data migration strategies such as LRU, CLOCK, 2Q, GDSF, LFUDA, FIFO, etc., are usually based on files'recent access mode(such as file access frequency, etc.), are mainly used to resolve data migration between memory and disk. They need to run in the operating system kernel, so the rules are relatively simple. For file access mode does not take file life cycle trend into account, some regularly accessed files are often not predicted accurately. In addition, file history access records are not considered.

Data access requests are not completely random. They are driven by the behavior of users or programs. There must be association between different files that are accessed consecutively. This paper proposes a method of file access heat prediction. Data heat trend is used as the basis for migration to a relatively low-cost storage device. Due to the limitations of traditional models, it is difficult to achieve good results in predicting at such nonlinear scenes. This paper attempts to use the deep learning algorithm model to predict the evolution trend of data access heat. This paper discussed the implementation of some initial parts of the system, in particular the trace collector and the LSTM model. Then some preliminary experiments are conducted with these parts.

Authors: CHENG, Zhenjing (INSTITUE OF HIGH ENERGY PHYSICS); CHENG, Yaodong (IHEP); WANG, Lu (Computing Center,Institute of High Energy Physics, CAS); Mr HU, Qingbao (IHEP); LI, Haibo (Institute of High Energy Physics Chinese Academy of Science); XU, Qi (Institute of High Energy Physics,Chinese Academy of Sciences); CHEN, Gang (INSTITUTE OF HIGH ENERGY PHYSICS)

Presenter: CHENG, Zhenjing (INSTITUE OF HIGH ENERGY PHYSICS)

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research