

Simulation of the cache hit rate for data readout at the Tokyo Tier-2 center

Friday, 5 April 2019 09:40 (20 minutes)

The Tokyo Tier-2 center, which is located in the International Center for Elementary Particle Physics at the University of Tokyo, is providing computer resources for the ATLAS experiment in the Worldwide LHC Computing Grid (WLCG). The official site operation in the WLCG was launched in 2007 after several years of development. The site has been achieving a stable and reliable operation since then.

We replaced almost all hardware devices in every three years in order to satisfy the requirement of the ATLAS experiment. The next hardware replacement will be performed in December 2018. In the current system, 6144 CPU cores (256 worker nodes) and 7392 TB disk storages are reserved for the ATLAS experiment. The disk storage consists of 48 sets of a disk array and a file server, where each disk array consists of 24 SATA HDDs. The worker nodes and the file servers are connected to a central network switch. The internal network bandwidth between the worker nodes and the central switch is 1040 Gbps in total (10 Gbps \times 104 links). The file servers are also connected to the central switch by 480 Gbps in total (10 Gbps \times 48 links).

We recently observe that the total throughput of data readout from the disk storage to the worker nodes is limited to about 100 Gbps in the current system even though the enough internal network bandwidth is available. The I/O performance of the disk arrays is one of the reasons of this throughput limitation because the disk array I/O utilization, which is measured by iostat command in Linux, is saturated during the heavy readout. The load of the data readout should increase in the next system because the number of CPU cores will increase, while the number of disk arrays will be the same with the current system due to the limited space of server racks. Therefore, we are discussing about a possibility to introduce a cache hierarchy for the storage system using SSDs to improve the I/O performance in the future system. However, we can not build an efficient caching system without the knowledge of the data readout patterns. The cache hit rate, which is defined as the number of data readout from cache area divided by the total number of data readout, is a good guideline to evaluate the data readout patterns for the caching system. In this presentation, simulation results of the cache hit rate will be reported. The simulation is performed using the real data access logs in the storage element at the Tokyo Tier2 center. We will discuss whether we can build an efficient cache system based on the simulation results.

Primary author: Mr KISHIMOTO, Tomoe (University of Tokyo)

Co-authors: Prof. TANAKA, Junichi (University of Tokyo); Dr KANEDA, Michiru (ICEPP, the University of Tokyo); Mr MATSUI, Nagataka (University of Tokyo); Prof. MASHIMO, Tetsuro (University of Tokyo)

Presenter: Mr KISHIMOTO, Tomoe (University of Tokyo)

Session Classification: Physics & Engineering Application

Track Classification: Physics (including HEP) and Engineering Applications