Contribution ID: **0** Type: **Oral presentation**

Solving Small Files Problem in Enstore

Thursday, 31 October 2013 14:00 (30 minutes)

Enstore is a tape based Mass Storage System originally designed for Run II Tevatron experiments at FNAL (CDF, D0). Over the years it has proven to be reliable and scalable data archival and delivery solution, which meets diverse requirements of variety of applications including US CMS Tier 1, High Performance Computing, Intensity Frontier experiments as well as data backups. Data intensive experiments like CDF, D0 and US CMS Tier 1 generally produce huge amount of data stored in files with the average size of few Gigabytes, which is optimal for writing and reading data to/from tape. In contrast, much of the data produced by Intensity Frontier experiments, Lattice QCD and Cosmology is sparse, resulting in accumulation of large amounts of small files.

Reliably storing small files on tape is inefficient due to file marks writing which takes significant amount of the overall file writing time (few seconds). There are several ways of improving data write rates, but some of them are unreliable, some are specific to the type of tape drive and still do not provide transfer rates adequate to rates offered by tape drives (20% of the drives potential rate). In order to provide good rates for small files in a transparent and consistent manner, the Small File Aggregation (SFA) feature has been developed to provide aggregation of files into containers which are subsequently written to tape. The file aggregation uses reliable internal Enstore disk buffer. File grouping is based on policies using file metadata and other user defined steering parameters.

If a small file, which is a part of a container, is requested for read, the whole container is staged into internal Enstore read cache thus providing a read ahead mechanism in anticipation of future read requests for files from the same container. SFA is provided as service implementing file aggregation and staging transparently to user.

The SFA is has been successfully used since April 2012 by several experiments. Currently we are preparing to scale up write/read SFA cache.

This paper describes Enstore Small Files Aggregation feature and discusses how it can be scaled in size and transfer rates.

Primary author: Dr MOIBENKO, Alexander (Fermi NAtiona Accelerator Laboratoy)

Co-authors: Mr KULYAVTSEV, Alexander (FNAL); LITVINTSEV, Dmitry (FNAL); Dr OLEYNIK, Gene (Fermilab); HENDRY, John (Fermi National Accelerator Laboratory); NAYMOLA, Stan (Fermi Naitional Accelerator Laboratory)

Presenter: Dr MOIBENKO, Alexander (Fermi NAtiona Accelerator Laboratoy)

Session Classification: Storage and file systems

Track Classification: Storage & Filesystems