

Resilient dCache: Replicating Files for Integrity and Availability.

Monday, 13 February 2006 11:00 (20 minutes)

dCache is a distributed storage system currently used to store and deliver data on a petabyte scale in several large HEP experiments. Initially dCache was designed as a disk front-end for robotic tape storage file systems. Lately, dCache systems have been increased in scale by several orders of magnitude and considered for deployment in US-CMS T2 centers lacking expensive tape robots. This necessitated storing data for extended periods of time on disk-only storage systems, in many cases using very inexpensive commodity (non-RAID) disk devices purchased specifically for storage or opportunistically exploiting spare disk space in computing farms. Hundreds of Terabytes of storage may be added for little additional cost. The large number of nodes in computing clusters and the lesser reliability of commodity disks and computers leads to a higher likelihood for individual files to become lost or unavailable in normal operations.

Resilient dCache is a new top level dCache service created to address these reliability and file availability issues by keeping several replicas of each logical file on elements of different dCache disk hardware. The Resilience Manager automatically keeps the number of copies in the system within a specified range when files are stored in or removed from dCache, or disk pool nodes are found to have crashed, been removed from, or added to the system. The Resilience Manager maintains a local file replica catalog and disk pool configuration in Postgres DB.

The paper describes the design of dCache Resilience Manager and experience in the production deployment and operations in US-CMS T1 and T2 centers. We use the configuration “all pools are resilient” in US-CMS T2 centers to store generated data before they are stored in T1 center. The US-CMS T1 center has some pools in the single dCache system configured as resilient, while the other pools are tape-backed or volatile. Such a configuration simplifies the administration of the system and data exchange. We attribute the increase in amount of data delivered to compute nodes from dCache US-CMS T1 center (0.2 PB/day in October 2005) to the data stored in resilient pools.

Primary author: Mr KULYAVTSEV, Alexander (FNAL)

Co-authors: Mr LITVINSEV, Dmitry (FNAL); Mr PETRAVICK, Don (FNAL); Mrs BERMAN, Eileen (FNAL); Mr MANDRICHENKO, Igor (FNAL); Mr BAKKEN, Jon (FNAL); Mr DE RIESE, Mathias (DESY); Mr ERNST, Michael (DESY); Mr FUHRMANN, Patrick (DESY); Mr KENNEDY, Robert (FNAL); Mr MKRCHAN, Tigran (DESY); Mr PERELMUTOV, Timur (FNAL); Mr PODSTAVKOV, Vladimir (FNAL)

Presenter: Mr PERELMUTOV, Timur (FNAL)

Session Classification: Poster

Track Classification: Grid middleware and e-Infrastructure operation