_UC2.png _UC2.bb _UC2.png”

Contribution ID: **186** Type: **Poster presentation**

Analysis and improvement of data-set level file distribution in Disk Pool Manager

Monday, 14 October 2013 15:00 (45 minutes)

Of the three most widely used implementations of the WLCG Storage Element specification, Disk Pool Manager (DPM) has the simplest implementation of file placement balancing (StoRM doesn't attempt this, leaving it up to the underlying filesystem, which can be very sophisticated in itself). DPM uses a round-robin algorithm (with optional filesystem weighting), for placing files across filesystems and servers. This does a reasonable job of evenly distributing files across the storage array provided to it. However, it does not offer any guarantees of the evenness of distribution of that subset of files associated with a given “dataset” (which often maps onto a “directory” in the DPM namespace (DPNS)). It is useful to consider a concept of 'balance', where an optimally balanced set of files indicates that the files are distributed evenly across all of the pool nodes. The best case performance of the round robin algorithm is to maintain balance, it has no mechanism to improve balance.

In the past year or more, larger DPM sites have noticed load spikes on individual disk servers, and suspected that these were exacerbated by excesses of files from popular datasets on those servers.
We present here a software tool which analyses file distribution for all datasets in a DPM SE, providing a measure of the poorness of file location in this context. Further, the tool provides a list of file movement actions which will improve dataset-level file distribution, and can action those file movements itself.
We present results of such an analysis and movement on the UKI-SCOTGRID-GLASGOW DPM.

Summary

Primary author: Dr SKIPSEY, Samuel Cadellin

Co-authors: Prof. BRITTON, David (University of Glasgow (GB)); SMITH, David (CERN); MITCHELL, Mark (University of Glasgow); PURDIE, Stuart (University of Glasgow); BHIMJI, Wahid (University of Edinburgh (GB))

Presenter: Dr SKIPSEY, Samuel Cadellin

Session Classification: Poster presentations

Track Classification: Data Stores, Data Bases, and Storage Systems