Contribution ID: **171** Type: **Parallel Talk**

XCFS - an analysis disk pool & filesystem based on FUSE and xroot protocol

Tuesday, 4 November 2008 17:25 (25 minutes)

One of the biggest challenges in LHC experiments at CERN is data management for data analysis. Event tags and iterative looping over datasets for physics analysis require many file opens per second and (mainly forward) seeking access. Analyses will typically access large datasets reading terabytes in a single iteration.
A large user community requires policies for space management and a highly performant, scalable, fault-tolerant and higly available system to store user data. While batch job access for analysis can be done using remote protocols experiment users expressed a need for a direct filesystem integration of their analysis (output) data to support file handling via standard unix tools, browsers,scripts etc.

XCFS - the xroot based catalog file system is an attempt to implement the above ideas based on xroot protocol. The implementation is done via a filesystem plugin for FUSE using the xroot posix client library which has been tested on LINUX and MAC OSX platform.

Filesystem meta data is stored on the head node in a XFS filesystem using sparse files and extended attributes. XCFS provides synchronous replica creation during write operations, a distributed unix quota system, krb5/gsi and voms authentication with support for secondary groups (via xroot remote protocol and through the mounted filesystem).

High availability of the headnode is achieved using a heartbeat setup and filesystem mirroring using DRBD. The first 80 TB test setup allowing to store a maximum number of 800 million files has shown promising results with thousands of file open and meta data operations per second and saturation of gigabit ethernet executing single 'cp' commands on the mounted file system. The average latency for meta data commands is in the order of ˜1ms, for file open operations it is <4ms.
The talk will discuss results of typical LHC analysis applications using remote or mounted filesystem access. A comparison will be made between XCFS and other filesystem implementations like AFS or Lustre. Strength and weaknesses of the approach and its possible usage in CASTOR - the CERN mass storage system - will be discussed.

Primary author: Mr PETERS, Andreas Joachim (CERN)

Presenter: Mr PETERS, Andreas Joachim (CERN)

Session Classification: Computing Technology for Physics Research - Session 2

Track Classification: 1. Computing Technology