Storage at the National Analysis Facility

Uwe Ensslin
4. 12. 2007

Overview

- **Wide area filesystem: AFS**
- **Cluster File System: Lustre**
- **GRID Storage: dCache**

AFS - The wide area filesystem

- With the NAF, we will provide a new AFS cell: naf.desy.de
- Use cases
  - Wide area access to NAF and other AFS cells
  - User home directories
  - Working Group directories
  - Provide space for private versions of VO software
    - Managed by the VO
  - Safe storage of and (remote) access to 20GB-200GB per user for private data (Initially 6 TB total)
    - Initial: Desaster recovery only

NAF Cluster Filesystem - Concept

- With the NAF, we will introduce a cluster filesystem
- The cluster filesystem will provide high bandwidth (O(GB/s)) access to a large (O(10TB)) 'working space' for temporary data
- Usage Model: Copy/write your data to the working area provided by the CFS, process, then save valuable results to wide area or GRID storage
- Data life time in the working area: Larger than job run, limited by available space, policy TBD

NAF Cluster Filesystem - Concept

- **For HEP, it is quite new technology**
- **Our objective is to evaluate such a filesystem in the NAF environment in cooperation with users**

NAF Cluster Filesystem - Implementation

- We chose to evaluate lustre®
  - Most promising candidate providing transport via native Infiniband
  - It is on it's way to the HEP community (Sites are evaluating)
  - Until recently maintained by Cluster Files Sytems, Inc. (CFS)
- CFS was bought by SUN in October (Good !)
- The roadmap was considerably delayed
  - Version 1.8 will come in 2nd Quarter of 2008
- Some features are not yet available or usable (ACLs, kerberization, ZFS)

NAF Cluster Filesystem - Implementation

- We cannot wait for 1.8..
- So we will base our CFS user evaluation on Lustre 1.6
  - Which is the current production version, used at several sites
- In parallel we will actively follow the development of 1.8
- Restrictions in the 1.6 evaluation instance
  - The available total disk space will be 50TB
  - There will be no ACLs, Quotas, kerberized user authentification, backup
- Access to the CFS will be over Infiniband only
- There will be one Filesystem per VO (Atlas, CMS, LHCb, ILC)

NAF Cluster Filesystem - Development

- **We will set up an internal Lustre 1.8 development instance, if possible in cooperation with SUN, to ..**
  - **evaluate 1.8 stability, functionality, performance, ..**
  - **test features like ZFS, ACLs, Quotas, Kerberization, Solaris based storage servers, ..**
  - **do some more hardware evaluation**
  - **develop a roadmap for further use of Lustre in the NAF**

dCache

- NAF users can access data from the Tier-2 dCache instances
- With the usual protocols and tools
  - gsidcap, gridftp/srm, dccp, srmcp ..
- dCache is the central data entry point for physics data..
- .. and also the common point for data exchange
  - Within VOs
  - Between working groups
  - Between GRID and NAF

- **Thank you..**