



Proposal for dCache based Analysis Disk Pool for CDF

presented by
Krzysztof Genser
Fermilab/CD/REX

on behalf of the CDF Offline Group



Outline

- The Problem & Proposed Solution
- Deployment Process
 - Collaborative and phased approach with feedback from users and experts at each stage



The Problem

- Current solution used to support physics analysis use cases (see Ray's talk), i.e. ~45 separate file servers with static disk areas is
 - Fragmented and non transparent and therefore
 - hard to oversee, use, support and manage
- Specialized version of rootd used to serve data; the person who was supporting it has left CDF; standard rootd will need investigation



Proposed Solution

- Replace the majority of the static disk space with dCache based pool (the analysis disk pool)
 - Use it for large files for which dCache is known to work well
 - Store small files, e.g. log files on other disk based system e.g. on nfs mounted disks visible from Interactive Login Pool nodes



Advantages

- Solution adopted/supported at Fermilab and within HEP
 - Allows for unified support and expertise sharing and accumulation
- Global name space and space management
 - more efficient use of disk resources
 - more transparent maintainability
- Decoupling of the name and disk space
- Scalability
- Client software already used by CDF



Risks

- In a centralized, pnfs based system a small group of users may inadvertently affect all users of that system
- Limited experience serving ntuples using dCache on a large scale (many clients)
- User expectations may not match system performance/capabilities
- Lack of personnel – risk present for the existing model as well as for the proposed one



Risk Management

- Staged deployment with reviews between stages
- Study impact of the use cases on the system
- Find system limits and communicate them to the users
- Monitor the system to make sure it stays within the stable limits
- Establish usage guidelines and limit exposures when possible (e.g. limit pnfs mounts)
- Using more widely supported solution
 - Experiment modified rootd -> dCache



Proposed Support Model

- Three tier approach:
 - Day-to-day operations and trouble shooting by CDF power users & CDF offline operations managers
 - Diagnosing difficult problems, evolving and reconfiguring system when needed by a group from a CDF institution (per a to be established MOU) – also serving as a point of contact between the experiment and CD
 - Expert level consultations within to be agreed upon limits by CD dCache development team



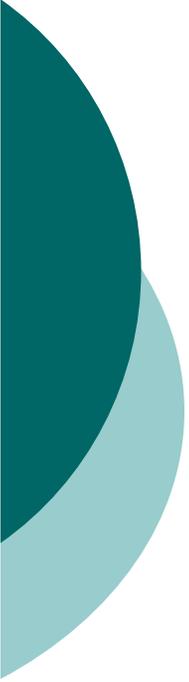
Proposed Support Model (cont'd)

- Hardware, OS support by REX/SA
 - No change
- Given similarities of the system to the other dCache systems one may want to revisit the support model once the system reaches stable level of operations and the support effort is known



Deployment Plan

- Staged approach; Three initial phases:
 - I. Prototyping
 - II. Pre-production
 - III. Production



Phase I: Prototyping

- Time scale: till ~end of January 2006
- Goals:
 - Develop resource loaded schedule
 - Understand use cases and system requirements
 - Understand technical characteristics and limits of the system (using ~50TB of disks)
 - Recruit and train power users (with needs after winter conferences)
 - Train CDF Offline Operations Managers
 - Establish sufficient system monitoring
 - Develop usage rules and guidelines
 - Investigate a possibility of building a common knowledge base (repository?) to be shared among CD, CDF, CMS, MINOS
 - Develop specification for the hardware to be used in phase II
 - Develop support agreements for phase II
 - Hold/pass pre-production readiness review



Phase II: Pre-Production

- Time scale: ~4 months
- Goals
 - Deploy production hardware (system size ~100TB)
 - Establish automatic system monitoring
 - Define/Perform load tests at the anticipated production level
 - Gradually expand user base to the full collaboration with the understanding that each expansion may need to be reverted should system become unstable; use authentication mechanisms to limit access
 - Revise usage rules and guidelines as needed
 - Develop support agreements for phase III
 - Hold/pass production readiness review



Phase III: Production

- Desirable Time Scale: end of Tevatron Shutdown ~July 2006
- Goals:
 - Maintain stable operations (system size ~100TB)
 - Allow for user base expansion within stable limits
 - Evolve the system to support approved use cases
 - Perform needed tests during scheduled downtimes to validate configuration and policy changes



Summary

- The process of arriving at a stable Analysis Pool solution was described
- The staged deployment with the reviews between the phases should help risk mitigation and with managing user expectations