



Light weight Disk Pool Manager experience and future plans

Jean-Philippe Baud, IT-GD, CERN

September 2005





Agenda



- Goals for the DPM
- Current functionality offered
- Current sites using DPM
- Experience with DPM at Tier2s
- Service Challenge 3
- Missing functionality
- Plans for future versions
- Discussion on priorities for new features



Goals for the DPM



- Provide a solution for the small Tier-2s in LCG-2
 - This implies a few tens of Terabytes in 2005
- Focus on manageability
 - Easy to install
 - Easy to configure
 - Low effort for ongoing maintenance
 - Easy to add/remove resources
- Integrated security (authentication/authorization)
- Replacement of 'Classic SE'



Current functionality offered



- Management of disk space
- Management of name space (including ACLs)
- Control interfaces: socket, SRM v1.0 (no srmCopy method yet), SRM v2.1 without global space reservation
- Data access protocols: secure RFIO, GsiFTP (Globus 2.4)



Current sites using DPM



- - NDGF (Denmark-Sweden)
- - Glasgow (UK)
- - Edinburgh (UK)
- - QMUL (UK)
- - RAL (UK)
- - INFN-Catania (Italy)
- - INFN-Legnaro (Italy)
- - INFN-Bari (Italy)
- - INFN-Padova (Italy)
- - CERN (Switzerland)
- - CSCS (Switzerland)
- - LAL-DAPNIA-LPNHE (France)
- - NIKHEF (Netherlands)
- - IFCA (Spain)
- - BNL (USA)
- - HIP (Finland)
- - TW-NCUHEP (Taiwan)
- - ASCC (Taiwan)



Experience with DPM at Tier2s



- 3 ways to install the DPM: manual, Yaim, Quattor (LAL)
- Sites using Yaim had almost no problem for the installation
- Sites doing a manual installation had quite a few problems in the security area
- Very stable, but a few problems
 - free space reported
 - TCP ports used by RFIO v3 (firewall)
 - Filesystems larger than 2TB not supported
 - Insufficient documentation about ACLs



Service Challenge 3



- NDGF (Denmark/Norway/Sweden) used a distributed Disk Pool with a single “head” node
- very stable
- Constant data rate (150 MB/s)
- Good resiliency observed after a network fiber accidental cut
- Some restarts needed because of incorrect free space reported
- Bad disk nodes not blacklisted for all error conditions (needs integration with monitoring), but FTS could also disable filesystems in case of errors



Missing functionality



- Administrator point of view
 - Support for large filesystems (UK)
 - Support for disk servers spread over multiple subnets
 - Support for RFIO port range
 - VOMS integration
- "User" point of view
 - Incompatibility with srmcp (Phedex)



VOMS integration



- DNs will be mapped to virtual UIDs: the virtual uid is created on the fly the first time the system receives a request for this DN (no pool account)
- VOMS roles will be mapped to virtual GIDs
- A given user may have one UID and several GIDs
- Integration with CSEC and CGSI
- Administrative tools will be offered to update the DB mapping table
- A paper has been submitted: "Issues in implementing Storage Security for the LHC Computing Grid"



Plans for future versions



- Short term (LCG 2.7.0)
- Medium term (end of 2005)
- Longer term (first half of 2006)
- Twiki page:
<https://uimon.cern.ch/twiki/bin/view/LCG/DataManagementWorkPlan>



Disk Pool Manager short term plan



- Bugfix release for incorrect free space reported
- Support for large filesystems (> 2 TB)
- Procedure to convert a Classic SE into a DPM

- Integration with VOMS/LCAS/LCMAPS
- RFIO_PORT_RANGE



Disk Pool Manager medium term plan



- MySQL backups
- DPM with disk servers on several network domains (distributed Tier2)
- srmCopy
- DPM DSI plugin for Globus 4 gridFTP server
- "du" command
- Integration with Information System (DESY)
- Drain of a pool, a server or a filesystem
- Integration with fabric monitoring
- Limit number of streams per disk server (may be useful for some applications like bulk replication, so pool dependent)
- Support for ROOTD/XROOTD transfer protocol



Disk Pool Manager longer term plan



- RFIO client library compatible with CASTOR
- Quotas (INFN/DESY)
- (automatic) replication inside a pool
- Global space reservation with max lifetime
- Streaming mode (SRM v3)
- implementation of a migrator/recaller to either recall/migrate files automatically between Tier1 and Tier2 or interface to a tape/DVD backend
- Problem of ACLs propagation during replication between SEs (RRS/RMS ?)



Discussion on priorities for new features

