



FIO

Fabric Infrastructure
and Operations

CERN IT
Department

Experiment top 5 issues: CASTOR Status & Plans

GDB – May 2nd 2007

Tony Cass



Experiment	Issue 1	Issue 2	Issue 3	Issue 4
<i>CMS</i>	Single request queue bottleneck	Priorities must work; need for capacity	Performance at CERN & Tier1s	Disk mover limitations
<i>ATLAS</i>	Tier0 shielded from users	SRM v2.2 availability	quotas	
<i>LHCb</i>		"Resource busy" from CASTOR	SRM v2.2 availability	
<i>Alice</i>	xrootd interface	Predictable latency for tape recall		

- Tier-0
 - Stability & Performance, isolation from general users
- Tier-1
 - Stability & Performance
- SRM v 2.2 interface availability
 - including VOMS integration
- xrootd interface availability
- Quotas
- Predictable latency

- Stability & Performance

- Major problems occur if CASTOR cannot schedule new requests
 - queue builds up and stager breaks down.

- New LSF plugin improved ability to handle large queues.

- BUT: Why do the queues build up?

- No free slots on disk server
 - Main cause of problems with CUS in 2006 due to various bugs & misunderstandings. All now fixed.

- All filesystems full & our requests cannot be satisfied

- Generally due to use of *disk1* pools; see later

- Slow or unresponsive database server

- main reason for ATLAS problems in March; improvement seen when memory added. New servers being introduced.

Address the underlying problems.
Improved resilience with new LSF plugin.

- Isolation from Analysis users / Single queue bottleneck
 - LSF can schedule requests according to priorities...
 - it does so for the batch system
 - ... but only if it sees the requests
 - new LSF plugin works and moves queue limit from 1000 to at least 30,000 requests (>50,000 shown in tests) and many requests (e.g. *put_data*) now bypass LSF.
 - // there are still problems we could have separate stagers
 - but with increased operational overhead and reduced resource usage efficiency (disk pools & tape drives).

Requests for scheduling now maintained in shared memory to avoid Oracle query limitations.
New plugin being deployed now.
Fallback available.

- **Stability & Performance**
 - As for Tier0, but with better version control for outside sites.
 - An improved build, test and deployment procedure is in place. However, CVS changes and increased modularity are not a priority compared to some other issues. Multiple version support not before end-2007.
- **Improved Disk1 support**
 - Highest priority after Tier-0 isolation issues
 - Many issues already addressed (e.g. server file system load balancing, issues with file system space limit)
 - Restricting write access to authorised users also needed
 - Reduces risk of cache filling up with unwanted files
 - » and therefore problems of searching through filesystems to identify unwanted and uncatalogued files to delete.

Improving disk1 support is the highest priority after delivering an improved scheduling interface

- SRM v2.2 Interface Availability
 - CASTOR interface passes all basic and use case tests.
 - Stress testing can start, but test stager instance at CERN is currently devoted to tests of the new LSF plugin in the context of the ATLAS taskforce.
 - Strong Authentication and VOMS integration next priority after disk1 issues
 - but no deployment in production before Q2 2008 (and probably later).
- Quotas
 - Introducing support for quotas is not a priority
 - so nothing before end-2008 or 2009
 - and, even if done, quota checking performed only at moment of initial write request.
- Predictable Latency
 - Predictable latency for access to disk resources (a CMS requirement) believed achievable with scheduler interface.
 - Predictable latency for access to files on tape is not achievable given likely resource levels
 - would require dedicated read-only drives per stager (experiment)
 - Production reload of data from tape should be managed
 - No guarantees for reload of user data from tape. Indeed, user requests (especially for small files) may have to be held until we have enough requests for a given volume.

- The xrootd interface to Castor is developed by SLAC, not CERN.
 - CERN made some slight changes to the underlying CASTOR software last year, but the key interface is developed by Andy Hanushovsky.
 - We believe the interface is now working
 - but extensive testing is needed to be sure that the overall Castor/xrootd combination functions correctly in all use cases and can support (more than) the likely load.
 - Long term support for this interface has yet to be agreed.
 - SLAC seem willing to provide support but are not funded for this, and
 - support currently depends on one person.
- Please*
- Test the new interface*
- Discuss support with SLAC contacts*

