Mass Storage Progress: Overview

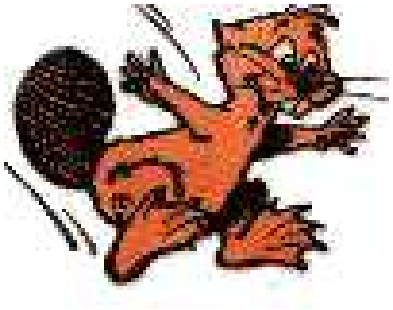
Greig A. Cowan

University of Edinburgh

Outline

1. Brief introduction to the storage technologies.
2. Status of storage deployment in WLCG.
3. Introduction to SRM2.2.
   - Outcomes from SRM2.2 deployment workshop.
4. Current issues.
5. Storage accounting.
6. Summary.

Storage middleware



CASTOR

CERN, RAL

Tier-0 and Tier-1s

dCache

DESY, FNAL, NDGF

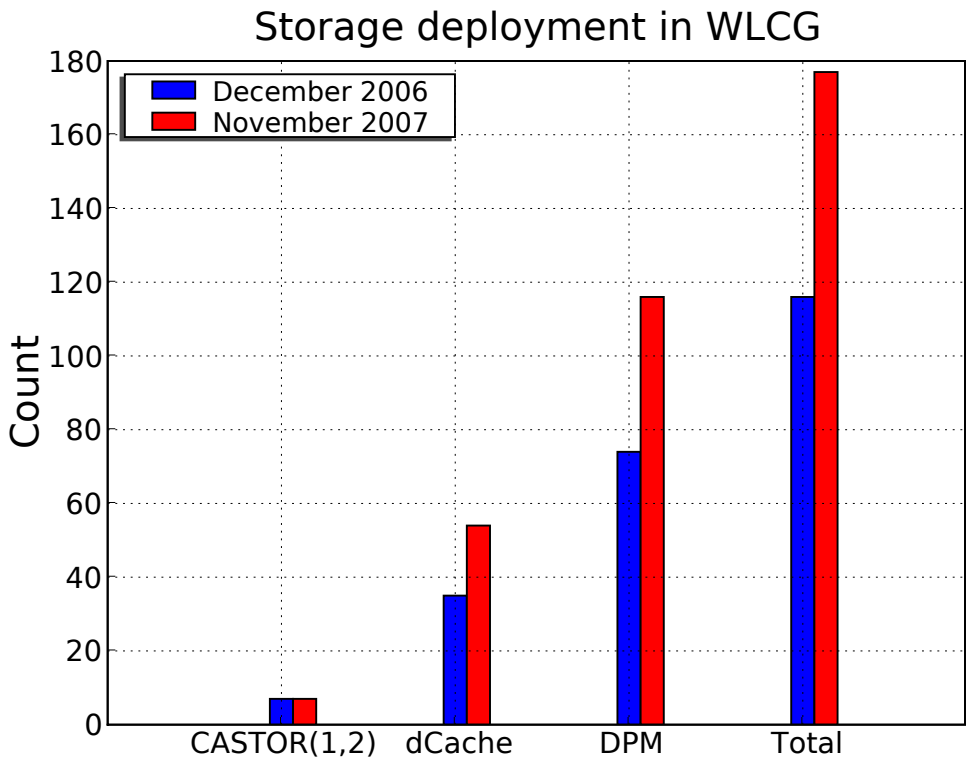
Tier-1s and large Tier-2s

Disk Pool Manager

CERN

Tier-2s, no tape

Storage deployment in WLCG

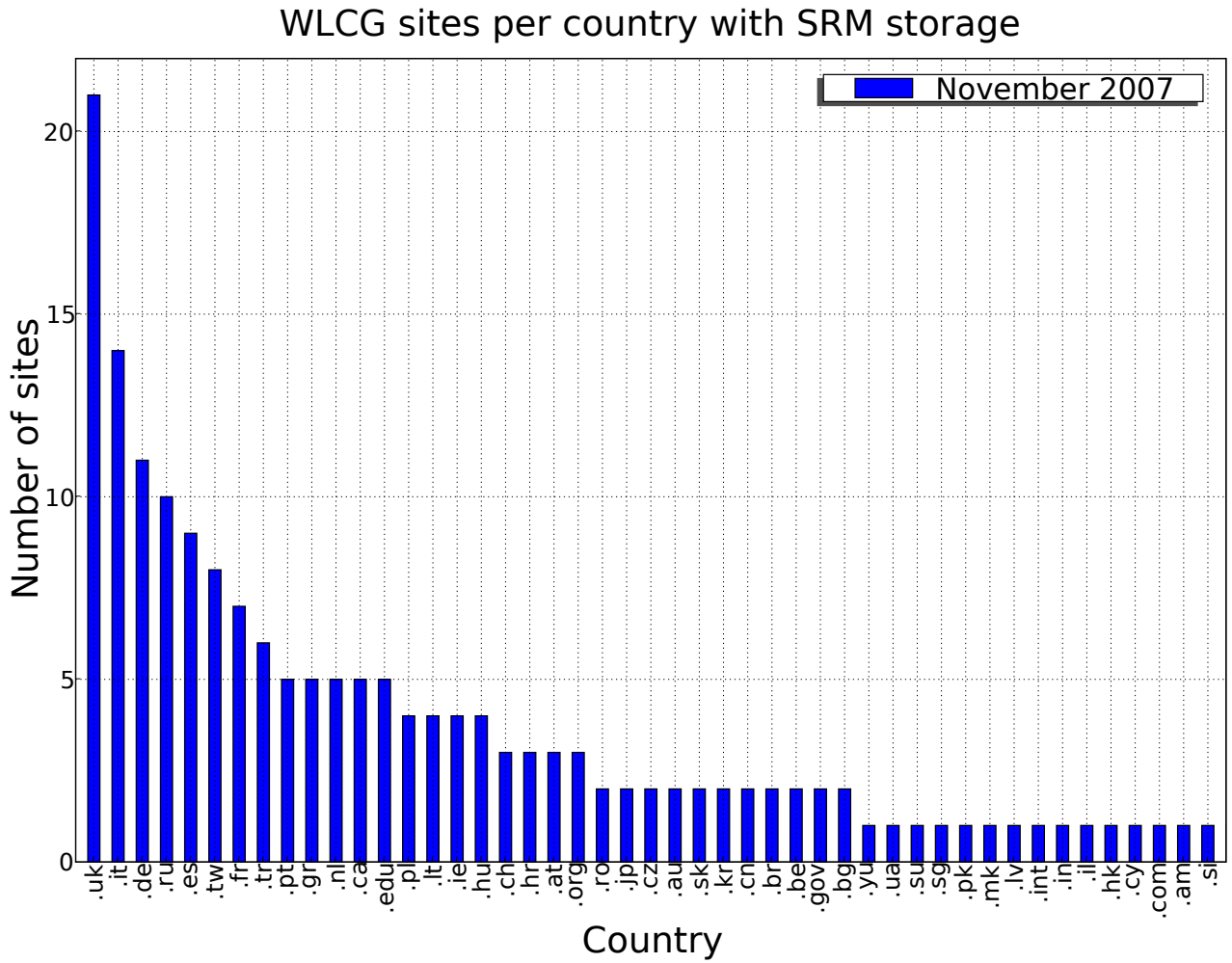


- **WLCG grid information system** tells us the number of SRMs deployed in production.
- Query for particular string in each site record.
- Clear **increase** in the use of storage middleware.
- Middleware also used at non-WLCG institutes.

http://www.gridpp.ac.uk/wiki/WLCG_SRM_usage

Storage deployment in WLCG

- **177** sites in **47** countries relying on WLCG storage middleware.
- Must ensure that we have the correct **support** infrastructure in place prior to LHC start-up.
- As we continue to scale up we need new structures in addition to just the developers.

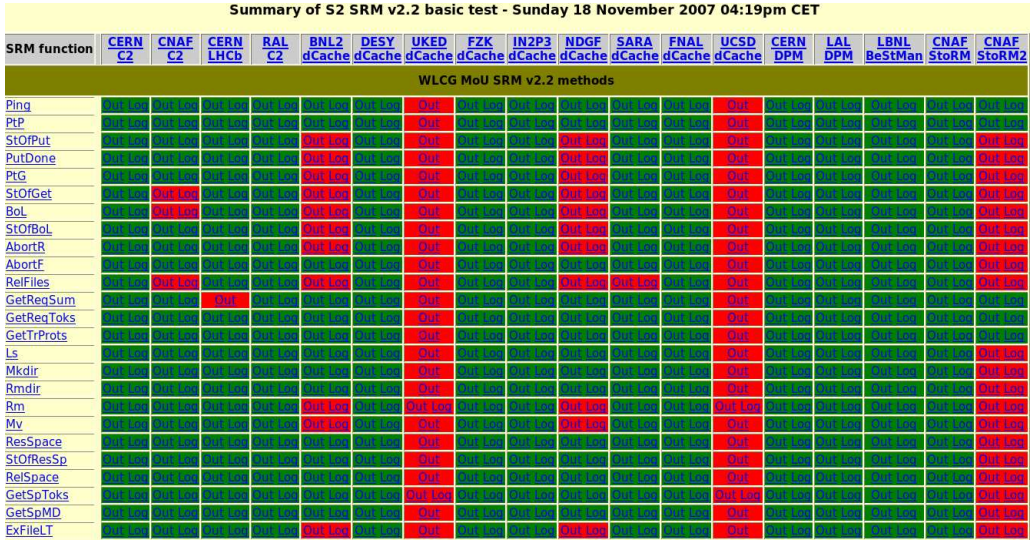


Storage Resource Manager

- SRM is the standard interface to Grid storage.
- CASTOR, dCache and DPM all provide production SRM v1.1 endpoints.
- SRM v2.2 introduces **new concepts** for managing storage resources.
  - Retention policy, access latency, space token descriptions.
  - Will provide new functionality that has been requested by the experiments.
- All software providers have been deeply involved in defining the SRM v2.2 spec.

SRM2.2 deployment in WLCG

- CASTOR, dCache and DPM instances were installed at CERN and various Tier-1 and Tier-2s.
- CERN have been running an **inter-operability** test suite for over a year.
  - Crucial in ensuring the different server and client implementations follow same spec.
  - Very successful at identifying problems.



Summary of S2 SRM v2.2 basic test - Sunday 18 November 2007 04:19pm CET

SRM function	CERN C2	CNAF C2	CERN LHCb	RAL C2	BNL2 dCache	DESY dCache	UKED dCache	FZK dCache	IN2P3 dCache	NDGF dCache	SARA dCache	FNAL dCache	UCSD dCache	CERN DPM	LAL DPM	LBNL BeStMan	CNAF StoRM	CNAF StoRM2
WLCG MoU SRM v2.2 methods																		
Ping	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
PtP	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
StOfPut	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
PutDone	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
PtG	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
StOfGet	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
BoL	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
StOfBoL	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
AbortR	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
AbortF	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
RelFiles	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
GetReqSum	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
GetReqToks	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
GetTrProts	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
Ls	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
Mkdir	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
Rmdir	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
Rm	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log
Mv	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
ResSpace	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
StOfResSp	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
RelSpace	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
GetSpToks	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log
GetSpMD	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log
ExFileLT	Out Log	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log	Out	Out Log	Out Log	Out Log	Out Log	Out Log

- **Configuration of the test systems has been difficult.**
  - It is inevitable that there will be similar problems at the Tier-2s and other Tier-1s.
- WLCG timescale: **all sites deployed SRM2.2 by February 2008**.
  - SRM2.2 endpoints in production at FZK, NDGF (dCache).

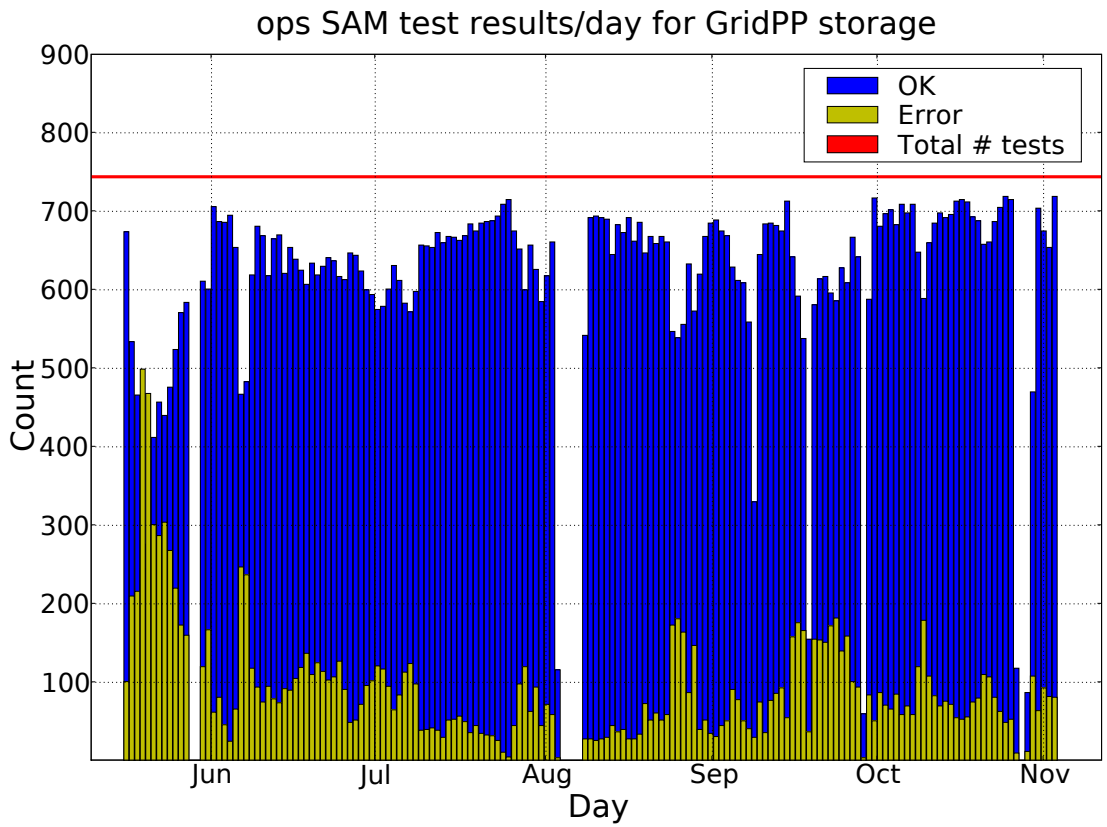
SRM2.2 deployment workshop

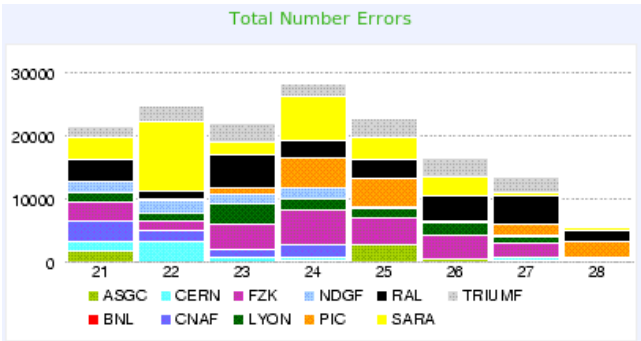
- National eScience Centre, Edinburgh, November 13-14th.
- $>$60 participants from across WLCG, primarily Europe.
- Purpose was to bring together sites, developers, experiments and storage “experts” in order to share information about the new SRM2.2 service.
- Good place to **collect information** for future reference and to get sites thinking about SRM2.2 deployment.
- ATLAS and CMS are only now starting to plan how they will use SRM2.2, in particular at the Tier-2s.

http://indico.cern.ch/conferenceDisplay.py?confId=21405

Experiment issues

- Strong concerns about the **reliability** and **stability** of storage.
- One measure of this comes from Service Availability Monitoring ⟶
- VO experience is often different to that reported by SAM.
- e.g., the software may function correctly, but the configuration for a particular VO gives errors.
- Possibly need more experiment specific tests to really highlight their problems.





Site issues (1)

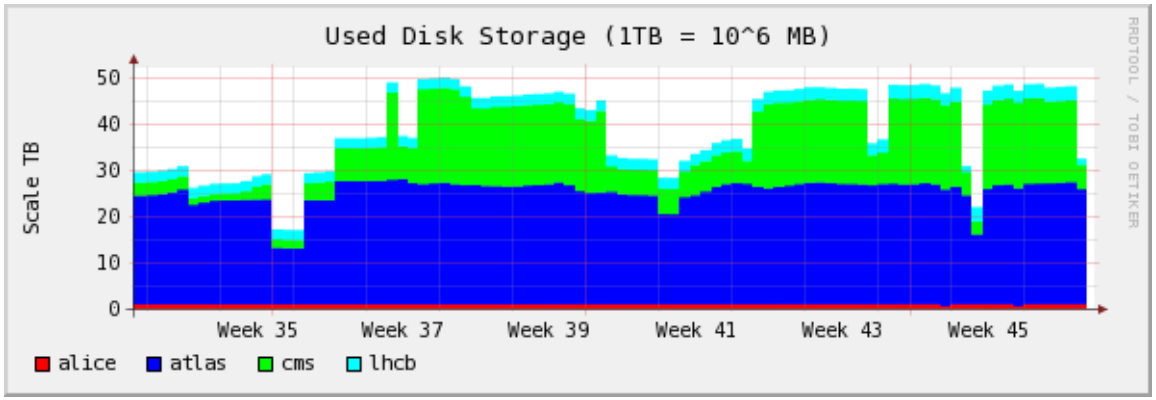
- **Management** of storage systems.
  - For a basic system, YAIM performs well.
  - Tools for management of the disk pools, reconfiguration of the service, permissions management, quotaing still not quite there.
  - Documentation and training are essential for understanding new features (SRM2.2).
- **Scalability** of the software.
  - Work is ongoing to investigate the effects of stressing the storage with LAN data access simultaneously with WAN transfers.

Site issues (2)

- Utilisation of **cluster file systems** (GPFS, Lustre).
  - Not all Grid sites are dedicated to HEP, e.g., shared University research cluster.
  - Often use parallel filesystems which sites want to take advantage of to give good local data access performance.
  - StoRM from INFN gives an SRM interface on top of GPFS.
- Support for **SL4, 64bit** (almost ready to go now).
  - For months, sites have forced the SL3 middleware onto SL4 and installed 64bit versions before production release.
  - Sites have already started to ask about SL5.

Storage accounting

Used disk storage by LHC VOs in France over 3 months ($\sim$ 46TB)



- Collecting grid storage usage/availability figures from the **information system**.
- Provides **high-level** overview of status of deployed resources on the grid.
- Discussion ongoing as to how this should be extended to account for SRM2.2 spaces and new information that will be available in v1.3 of GLUE.

http://goc02.grid-support.ac.uk/storage-accounting/view.php

Summary

- Data storage is central to the success of the **data grid** that we are building.
- 177 sites are now using these SRM middleware.
  - CASTOR, dCache and DPM are performing **useful work** - they store TB of “real” data.
  - Community must **support** these complex systems before and during LHC operation.
- **Reliability/stability** of storage software is a concern for experiments and sites.
  - This is partly the software itself, partly the management of the software at sites.
  - **Communication** between sites and experiments is essential for identifying and resolving problems.
  - Training and documentation essential to provide the required service.
- **SRM2.2** deployment has started.
  - We need to manage the migration from SRM1.1 to SRM2.2 with space management.