



WLCG Jamboree – Summary of Day 2

Simone Campana, Maria Girone,
Maarten Litmaath, Gavin McCance,
Andrea Sciaba, Dan van der Ster

IEEE MSST Summary – Jean Philippe

- + Use standard protocols (see also next talks...)
- + Use standard building blocks
- Multi-tiered storage must be flexible (future-proof) but is difficult to implement
 - Tapes will be actively developed - mainly used for archiving
 - SSDs not ready for production yet, are expensive and deliver relatively small performance advantages
- Fast development cycle is a key point
- Cost of ownership discussion...
- **Potential demonstrators?**

File System Tests - Andrei

- The majority of the HEP data is stored in dCache and CASTOR but there is a zoo of solutions adopted by the sites
- Benchmarked a number of technologies:
 - GPFS, Xrootd, AFS, Lustre, dCache, Hadoop, AFS/VICGPFS, AFS/VICELU
 - Two analyses tested (events processed/30 minutes):
 - A CMS analysis doing rather linear reading (Datascan)
 - An ATLAS analysis doing rather random I/O (Hammercloud)
- Three messages:
 - The underlying storage is relevant: there is a very large difference between the best and least best
 - “Silver bullet”: Not all solutions good for one pattern are good for the other.
 - But, GPFS and the AFS++ solutions seem to work well for both. Lustre well performing for CMS but not for ATLAS
 - Storage tuning can bring big improvements, but is application dependent
 - Performance sensitive to caching on client-side
 - (Posix) filesystems in general look more efficient and require less tuning
- **Waiting on input from community. Potential demonstrators?**

NFS 4.1 – Gerd

- A Standard supported by many distributions
 - No lock-in to vendors (unlike GPFS/Lustre)
 - Open standard, backed by industry
- Key features
 - Striping and replication supported
 - Caching comes for “free”
- Clients widely provided
- Now standard settled and clients available industry is adding NFS4.1 to their storage solutions
- Funding available (EMI, dCache, ...)
- **Potential demonstrators?**

Xrootd - Fabrizio

- Hierarchical storage aggregation and access
- Site inter-operability, seamless WAN data access & site aggregations
- Plug-in based architecture, allowing overriding of default behaviour + new features
- P2P-like internal file discovery, no internal catalogue, subset of POSIX semantics, authentication and authorization
- Support: part of VDT + OSG support for US-ATLAS primarily; ALICE
- Discussion on protocol decoupling from implementation
- **Potential demonstrator(s)?**

FTS / LFC - Gavin

- FTS and LFC function as intended at WLCG scale
- The conceptual model of WLCG's transfer system is too simple
 - Need to consider storage bandwidth
 - But global $n \times n$ optimisation is hard
 - A global system would allow more magic (strawman)
 - FTS | SRM interface is too loosely coupled
 - Is the “file” the correct primitive for unordered bulk operations?
 - Ability to do partial transfers and restart missing
 - [Future developments..]
- Consistency with storage and among catalogues is the key challenge for cataloguing
 - Add GUIDs to storage catalogue
 - Use industry standard messaging a backbone of reliability and storage / catalogue integration [Proposed demonstrator]

AliEn File Catalogue – Pablo

- Provides a Global catalogue
 - Translation from LFN/GUID to (closest) PFN
 - Automatic SE selection
 - Unix-like file namespace
- Catalogue contains everything
 - Metadata, file collections, quotas, job output
 - Multiple storage protocols
 - Used by several experiments
 - ALICE, CBM, PANDA (HI), MAGIC V (LS)
 - Authorization, authentication
 - Most advanced use: Alice conditions accessed by every job (100 concurrent clients..)
 - 150M entries for ALICE: 100GB DB
- Scalability? Performance? (comparison with LFC?), Portability?

P2P and DAaM - JT

- WLCG tier model first step toward Content Distribution Network (CDN)
 - Producing site is origin server for its data
 - Each site has a proxy / cache
 - All files in global name space
- Features: collaboration between caches, distributed hash tables for cache resolution (transient catalogues?)
- CoralCDN: widely deployed, collaborative, paper attached to agenda [possible demonstrator?]
 - Security? Protocols? Storage / bandwidth quotas?
- Need to size caches, consider cache turn-over and hit missing, integrate with job management



Java/J2EE
Developer vacature

GRAYDON
CREDIT MANAGEMENT SERVICES

[Bekijk vacature](#)

Ads by Google

"You make it fun; we'll make it run"



The Coral Content Distribution Network

- [Home](#) ** Project home page
- [Overview](#) ** Brief overview and news
- [Usage](#) ** Coral Wiki and FAQ
- [Lists](#) ** Mailing Lists
- [Pubs](#) ** Publications and people
- [Download](#) ** Plugins and source code
- [Illuminati](#) ** Network measurement project
- [OASIS](#) ** OASIS anycast service

Browser Plugins

Click [here](#) for information regarding browser plugins.

Source Code

Our CVS server went down a few months ago, with significant disk corruption. We have yet to bring back a new source repository online.

No official release for Coral is yet available, but you are welcome to check out the latest version from the CVS repository. This "alpha" version is experimental and under active development.

SRM2.2 Lessons – Markus

- Problems:
 - Conflicting goals; rushed specification and implementation; tests arrived late; large effort to re-align implementations
 - Severe stability and performance problems of initial implementations; ACLs & permissions not well covered; no standard library
 - Space management – overly complex; quotas missing
 - Some methods not meaningful for all implementations
 - etc.
- SRM now ~same quality as storage systems
- Development of a protocol takes time, resources and commitment – treat different use cases separately
 - e.g. data management and data access interfaces...
- “Will we ever learn?” - Next steps ?
- Common data access protocol would be extremely useful...

ROOT – René

- 3 new developments:
 1. Caches;
 2. Speed-ups (10k reduction in network round-trips!)
 3. Parallel merge
- File accesses in WAN: local caches & proxies is way to go
- Could Squid caches play a role? Partial file caching?

NDGF/ARC - Josva

- Input (output) files stage in (out) taken care by the ARC CE
 - Used by ATLAS – happy!
- Data stored at T1, analyzed & cached at T2s
- Caching
 - A global index stores locations of cached files
 - Index kept up to date through Bloom Filter
 - Cache-aware brokering essential
- To be understood
 - Interaction with pilot framework not straightforward

KIT: xrootd+tape b/e – Artem

- Used by ALICE – happy!
 - Second largest ALICE SE after CERN In both allocated and used space
 - 25% of GridKa storage (~2.5PB)
- Stateless, scalable
- Low maintenance
- But good deal of integration efforts
- SRM frontend and tape backend
- No single point of failure

CNAF: GEMSS – Vincenzo

- StoRM+GPFS used in production since a few years now
 - Interface with TSM implemented by INFN (GEMSS)
 - GPFS/TSM/StoRM for ATLAS, CMS, LHCb
 - xrootd interface for ALICE
- GPFS Information Lifecycle Management is used to identify files for migration and recall
 - Basically a Rule Engine
- CASTOR full dismissed for LHC experiments



hadoop@Tier2

- [HDFS](#) mounted into WNs via FUSE
- BestMan used as SRM stripped down implementation
- To integrate GridFTP, in-memory data reordering has been implemented
 - HDFS supports “append” and not random writes
- Reliability
 - Hot blocks and lost blocks replication, block integrity
- Strong authentication missing (?)

Summary – Day 2

- Many interesting ideas exposed: several common themes and some differences of opinion (or at least interpretation...)
 - **Speakers: please upload slides if not done already**
 - Modify key: jam (plus upload rights by name)
- A few “early demonstrators” explicitly mentioned (catalogues, file transfer system, data access, storage management building blocks...)
- A small # of demonstrators – even “competing” – could help pin-down which ideas to pursue further...
- For discussion tonight and tomorrow...

Jamboree & WLCG Workshop

- A 30' [summary](#) of this jamboree is scheduled for day 1 of the WLCG Collaboration workshop 7 – 9 July 2010 in Imperial College, London
- Plus a [panel discussion](#) on Tier1/Tier2 storage options (also first day)
- A 90' [slot](#) is also foreseen for updates on “Early Demonstrators” (last day)
- Please [register](#) asap!
- [27th IEEE](#) – May 2011 – volunteers for PC?