Storage Interfaces and Access: Interim report

Wahid Bhimji
University of Edinburgh

On behalf of WG:
Brian Bockelman, Philippe Charpentier , Simone Campana, Dirk Duellmann, Michel Jouvin, Oliver Keeble,
Markus Schulz
And other participants (see meeting agenda/ minutes)

Storage Interfaces WG: Background and Mandate

- Little of current management interface (SRM) is used. Leads to performance overheads for experiments; work on developers to maintain; restricts sites technology choices.
- Building on Storage/Data TEG, clarify for disk-only systems the minimal functionality for WLCG Storage Management Interface.
- Evaluate alternative interfaces as they emerge, call for the need of tests whenever interesting, and recommend those shown to be interoperable, scalable and supportable. Help ensure that these alternatives can be supported by FTS and lcg_utils to allow interoperability.
- Meetings coincide with GDBs with extras on demand:
  - Contributions from developers, sites and experiments
  - Not designing a replacement interface to SRM but there are already activities so bringing together and coordinating these.

Functionality required

- Experiment's Data Management plans:
  - CMS: no "blockers" for non-SRM usage; Nebraska SRM-free site; Example ways of doing things that can be used by others.
  - ATLAS: some issues some of which will be resolved in next gen of data management: rucio. Open to trying controlled non-SRM sites using common solutions / interfaces.
  - LHCb have some concerns but also have not very different requirements than Atlas.
- Sites Perspective: CERN: Bestman SRM doesn't scale for them. Also RAL future disk only technology, choice ideally wouldn't be hampered by SRM options. Other sites see advantages…
- Middleware and tool development : e.g. FTS3; gFal2 support non-SRM interfaces (some testing by VOs to be done).

Areas still requiring development

See also tables in backup slides

Needed by?	Issue	Solution proposed in Annecy pre-GDB
ATLAS/ LHCb/ Sites	Reporting of space used in space tokens. Protection of space.	JSON publishing currently used in some places on ATLAS – temporary measure. WebDav quotas?
ATLAS/ LHCb	Targeting upload to space token.	Could just use namespace but certain SEs would need to change the way they report space to reflect. (Or use e.g. http)
ATLAS/L HCb	Deletion	gFal2 will help provide an abstract layer.
LHCb (ATLAS)	Surl->Turl	Require a redirecting protocol and SURL = Turl for sites that want no SRM.
ATLAS/L HCb	Service query: checksum check etc.	Some service check needed as is some “srm-ls”. gFal2 will help
All?	Redirecting protocol on different storage	Performant gridftp redirection already on dCache – soon on DPM (also have xrootd/http and FTS3 will /does support these).

Developments and remaining actions.

Developments since start of WG

- FTS3 now in production and its xrootd and http interfaces are almost at VO testing stage. gFal2 well developed and will be supported to satisfy the ongoing need described in last table.
- CMS now require xrootd at all sites; ATLAS soon need both xrootd (for federation) and webdav (for rucio – initially renaming but more envisaged)
  - These interfaces now prevalent (after EMI upgrades etc.)
  - Note: currently means **more** complexity with new and legacy interfaces used

Actions:

- ATLAS will test their required functionality in gFAL2 for deletion and service discovery and use of a few (more) non-SRM demonstrator sites.
- For using namespace instead of space-token (space usage , uploading, quota ..) – iterate on more concrete description of needs (both for VOs (LHCb/ATLAS) but also sites) alongside proposals / demonstrators.

Conclusions

- Activity on storage interfaces is progressing well. Transition from SRM is happening. Remaining issues are tractable via a combination of rules (SURL=TURL); redirecting protocols and movement to client side rather than server side abstraction (gFal2).
- Dooable – but not done: as it is not peoples primary focus it will not be an overnight transition so continuing need for group to stay engaged, monitor and ensure interoperability.
- There are also a number of wider issues, such as data access interface (see next slides), and cloud storage which did have TEG recommendations and which this group is discussing (though not in original mandate).

Data access protocols: WLCG direction

Reminder: TEG Recommendation(s)

[LAN] Protocol support and evolution:

- Both remote I/O (direct reading from local storage) and streaming of the file (copy to WN) should be supported in the short/medium term. However the trend to move towards remote IO should be encouraged by both experiments and storage solution providers, and should be accompanied by an increase in resilience of protocols.

- LHC experiments are able to support all protocols supported by ROOT and expect to be able to continue to do so in the future. This support should be maintained but the current direction of travel towards fewer protocols (in particular the focus on file://, xrootd and http://) is encouraged. Specifically both the work on current implementations of file:// access through Nfs-4-1 and that on testing ROOT performance with direct access via http, should be continued

Where are we now.

- Server wise http and xrootd options are more mature
- Experiment side still in the same zoo with even more animals:
  - Rfio, file, dcap now joined by http, xrootd, S3 …
  - With copy-to-scratch, direct and federation flavours
- Given xrootd is now established (on e.g. DPM) rfio is the most obvious candidate for retirement but still used very widely in production.
- For WAN gridFTP is already standard – here things are also diverging (as forseen) with FTS3 supporting xrootd; http.

Escaping the Zoo?

- Do we want to escape: Yes … there are support and performance issues, also makes for random user failures on different sites.
- Some solutions
  - Forced (gradual) retirement of rfio. Supported by WG
    - CMS is now requiring this. ATLAS phased transition (now used in prod at Taiwan; Edinburgh and Oxford) (see also backup slides).
    - Initial retirement monitored by this WG then WLCG ops or GDB.
  - Concept of a "core" protocol that is required for all sites? (even if another is used for performance.)
    - This is de-facto currently xrootd. However systems should be flexible enough to allow transitions.

The End

The following are extra slides….

Table of used functions from TEG

	Is this feature used by ...				Tier	
	Atlas	***CMS***	***LHCb***	***FTS only***		**SRM function[2]**
Transfer Management						
Upload / download a complete file	Yes	Yes	Yes	No	All	srmPrepareToPut/Get//Put/GetDone
Manage transfers.	Yes	Yes	Yes	Yes	T1/2	srmAbort/Suspend/ResumeRequest
Balance over multiple transfer servers.	Yes	Yes	Yes	Yes	T1/2	srmPrepareToGet [3]
Manage third-party copy	Yes	Yes	Yes	Yes [5]	T1/2	
Negotiating a transport protocol	No	No	No			srmGetTransferProtocols
Namespace Interaction						
Querying information about a file (stat)	No	No	Yes[1]	Yes[6]	T1/2	srmLs
Upload data integrity information (chksums)	No	No	No	No	T1/2	
Check integrity information	Yes	Yes	Yes	Yes		srmLs
Creating/Deleting data and directories	Yes	Yes	Yes[1]	Yes[7]	All	srmMkdir srmRmdir srmRm srmMv
Changing ownership, perms and ACLs	No	No	No	No	-	srmSet/Check/GetPermission
Storage Capacity Management						
Query used capacity (like df)	Yes	No	Yes	No	T1/2	srmGetSpaceMetaData/Tokens
Create/remove reservations; assign characteristics	No	No	No	No	-	srmReserve/Update/ReleaseSpace
Targeting uploads to specific reservation	Yes	Yes	Yes	No	T1/2	srmPrepareToPut
Moving files between reservations	No	No	Yes	No	T1/2	srmChangeSpaceForFiles
Server Identification						
Test service availability and information	Yes	Yes	No	No		srmPing

- Somewhat simplified and removed those only relevant for Archive/T1
- Still probably can't read it (!) but a couple of observations:
  - Not that much is needed – e.g. space management is only querying and not even that for CMS

Brief functionality table – focussing on areas where there are issues

Function	Used by ATLAS	CMS	LHCb	Is there an existing Alternative or Issue (to SRM)
Transfer: 3rd Party (FTS)	YES	YES	YES	Using just gridFTP in EOS (ATLAS) and Nebraska (CMS) What about on other SEs?
Transfer: Job in/out (LAN)	YES	YES	YES	ATLAS and CMS using LAN protocols directly
Negotiate a transport protocol	NO	NO	YES	LHCb use lcg-getturls;
Transfer: Direct Download	YES	NO	NO	ATLAS use SRM via lcg-cp, Alternative plugins in rucio
Namespace: Manipulation / Deletion	YES	YES	YES	ATLAS: Deletion would need plugin for an alternative
Space Query	YES	NO	YES?	Development Required
Space Upload	YES	NO	YES?	Minor Development Required

Rfio -> Xrootd on DPM

- All DPM sites use(d) rfio for local file access
  - Perceived performance issues
  - Therefore mainly used in copy mode (for atlas)
  - Sometimes there are issues with client libs (e.g. EMI 32bit ones; link to libshift.so; CMS many issues)
  - DPM developers want to move away
- WLCG “Storage interfaces” group supported decision to “retire” rfio.
- CMS requested / required it in WLCG ops mtg

Testing…

- Testing xrootd in UK and ASGC for almost a year
- Initially some issues but everything stable on the DPM server side for a long while now.
- Some issues seen on ATLAS FAX tests were N2N / Fax specific and shouldn't affect local xrootd ops.
- HammerCloud test so far didn't show any problems – not clear if performance gains
- Experiment configuration – production experience will sort these quicker
- Changed for ATLAS at Oxford, Edinburgh last week (for both production and analysis) to add to ANALY_TAIWAN

A few debateables

- Direct or copy-to-scratch: using copy at first but would be good for direct access to work well
- Should we push http copy at some sites?

As further steps can also use xrdcp for

- Federation access
- Stage-out (lfc registering taken care of elsewhere. What about SpaceTokens?).
- FTS transfers (soon in FTS3 (again possible ST issue))