

2015

DPM

Workshop

Highlights

A. Sartirana

6th DPM Workshop.

- 7–8 December 2015 at CERN;
- 22 participants;
 - ❖ developers + national communities
 - Italy, France, UK, Czech R., Australia, Russia, Switzerland;
 - ❖ for France: C. Biscarat, M. Jouvin, A.S
 - who was remotely connected?
- we had the (usual) general impression of an active and dynamical community;
- here is a partial summary, for more details:
 - ❖ <https://indico.cern.ch/event/432642/> .



The collaboration status.

- 60PBs in total, 176 instances
 - ❖ 6 are >2PB, 18 are >1PB. The largest:3.3PB
- run as collaborative effort
 - ❖ committed: CERN, France, GridPP, Italy, Japan
 - ❖ call for new contributors/contributions
 - sites with more than 1PB
 - possibility of master stages (e.g. DPMBox)
- reorganization: join FTS, EOS, Castor....
 - ❖ review of DPM in March
 - invited to show our support
- new website: <http://lcgdm.web.cern.ch/>



Answers to our questions.

- Yes, we “forgot” to announce 1.8.10
 - ❖ but we are on twitter now ... ☺ ;
- UMD nightmare: planning to drop MP from 1.9.0;
- the request about docs on pkgs/vers compat
 - ❖ ... echoed by other admins;
- Atlas green light for 1.8.10;
- requested draining tools may be already
 - ❖ dpm-replica-move (see later);
- plan to get rid of ST along with SRM
 - ❖ directories quotas instead;
- your contribution is very useful
 - ❖ good idea to take part to the MW readiness
 - ❖ always much to test (CentOS7/HTTP-only);
 - ❖ LAL --> WebDav.

- 3 sites (Atlas), 4.5PB
- Puppet/Foreman config
 - ❖ general reusable modules;
 - ❖ use foreman to pass parameters to modules;
- issues with atlas monthly dump;
- tested functionality HTTP/DynFed.



- 1 Atlas site (~1PB) + Belle II (Melbourne) + Belle I (Adelaide)
- good year
- not confident with upgrades (now on 1.8.9) and with migration from yaim to puppet
 - ❖ ask for more release docs

- 3 sites (Atlas), 4.5PB
- Puppet/Foreman config
 - ❖ general reusable modules;
 - ❖ use foreman to pass parameters to modules;



- issues with atlas monthly dump;

- test

- this seems to be an hot topic in atlas;
- there are currently, at least, 3 version of the script
 - ❖ one (very) old from atlas;
 - ❖ one from CMS;
 - ❖ one from Fabrizio;
- we should probably have a look and unify the efforts
 - ❖ fix slowness problems.



- many sites. Few very big (~2PB)
- uneasy with upgrades and with moving to puppet
 - ❖ many 1.8.8 (and/or SL5) and xrootd-3;
 - ❖ afraid to harm UK metrics;
- requests:
 - ❖ monitoring;
 - ❖ intelligent file placement (CEPH/Lustre backend);
 - ❖ throttle io per DS/proto.



- DMLite + HDFS at Bristol_T2
 - ❖ decommissioned GPFS+StoRM
 - ❑ already have an (underused) HDFS cluster;
 - ❖ Setup DMLite+HDFS SE + 2 gFTP
 - ❑ gFTP cache in tmpfs (need RAM;)
 - ❑ pbs with staling gftp procs;
 - ❑ BDII/Apel adapted to understand replication;
 - ❖ first CMS SRM-less site
 - ❑ first site with gridftp redirection in prod;
 - ❑ adapted the CMS configuration;
 - ❑ soon enable HTTP writes.



➤ Belle II Computing

- ❖ 1 VO (belle);
- ❖ 30 sites, 1.9PB (30% DPM)
 - ❑ ST and ACL's required;
- ❖ DIRAC;
- ❖ LFC/AMGA catalogs;
- ❖ sw (BASF2) via cvmfs;
- ❖ remote analysis
 - ❑ using "index files";
- ❖ plans to use HTTP.



1.8.10 (*dmlite* 0.7.3).

➤ Released mid-October

- ❖ MP sent to EMI end of October, available mid Nov;
- ❖ dmlite bugfix release beginning of November;
- ❖ DPM-dsi (gridFTP) recompiled for the new Globus, beginning of Nov;

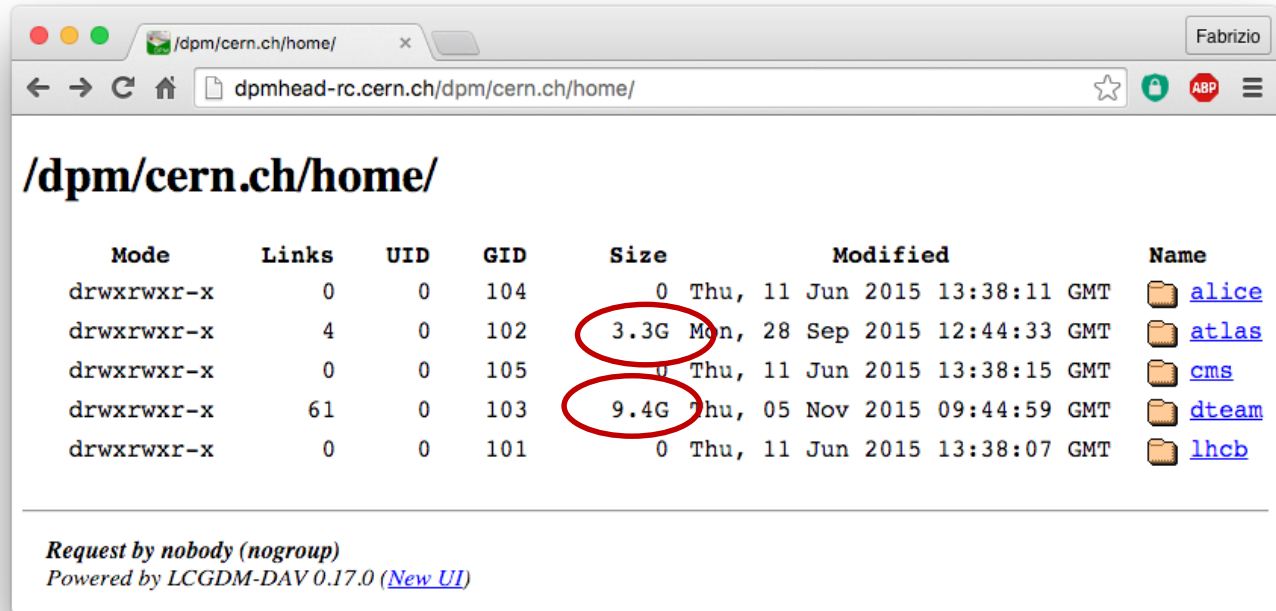
➤ highlights

- ❖ many improvements on dmlite logging (readability);
- ❖ multiple checksum from multiple frontends;
- ❖ important ACL fixes;
- ❖ sendfile() support in disk nodes;
- ❖ DPMBox, new fancy WebDAV interface;
- ❖ puppet standalone setup is consolidated;
- ❖ space reporting on top directories;
- ❖ drain and replication through HTTP;
- ❖ many improvements to dmlite-shell;
- ❖ gridFTP redirection improved.

Next slides

- Feature-rich toolbox for DPM administration
 - ❖ made for admins and devs not for users;
 - ❖ python-based, expandable, scriptable, command history, autocompletion, ...
 - ❖ you can manage directories, files, replicas, pools, users, groups, ... (48 commands)
 - ❑ `command lines are being "moved" to dmlite-shell;`
- drain (*drainpool/drainserver/drainfs* cmds)
 - ❖ multithreaded and http based;
- *replicamove* allows moving specific replica folders to a specific server/FS
 - ❖ similar to drain, but with the possibility to specify the destination (and to change spacetoken when moving as well);
 - ❖ *dpm-replica-move* in admin tools.

Space reporting.



Mode	Links	UID	GID	Size	Modified	Name
drwxrwxr-x	0	0	104	0	Thu, 11 Jun 2015 13:38:11 GMT	alice
drwxrwxr-x	4	0	102	3.3G	Mon, 28 Sep 2015 12:44:33 GMT	atlas
drwxrwxr-x	0	0	105	0	Thu, 11 Jun 2015 13:38:15 GMT	cms
drwxrwxr-x	61	0	103	9.4G	Thu, 05 Nov 2015 09:44:59 GMT	dteam
drwxrwxr-x	0	0	101	0	Thu, 11 Jun 2015 13:38:07 GMT	lhcb

Request by nobody (nogroup)
 Powered by LCGDM-DAV 0.17.0 ([New UI](#))

- populated by `dmlite-mysql-dirspaces.py`
 - ❖ designed depth 4/5 and 1 run/month;
- coincides with the SRM numbers only if `dirs~ST`;
- writing via SRM aren't accounted;
- disabled by default.

gridFTP redirection.

➤ What is it?

- ❖ allows for **SRM-less** gftp transfers;
- ❖ **not all clients** support the DP mode needed
 - ❑ **old clients** will trigger **rfio** background **copies**;
 - ❑ **FTS3** and **gfal-2** are ok;
- ❖ no yaim conf, only puppet;

➤ current status

- ❖ ok on testbeds since ~1y;
- ❖ non trivial integration with exps;
- ❖ need to recompile dpm-dsi each new version of globus (which is quite often) as it relies on internal API.

- Final stages of a 4y long smooth transition
 - ❖ make DPM the lowest cost grid storage;
- cut DMLite and LCGDM deps, LCGDM non optional
 - ❖ challenge: the DMLite relies on LCGDM, through the adapter plugin. Remove this dep;
 - ❖ Eric Cheung did this as a proof-of-concept prototype using fastCGI, codename DPMRest;
 - ❖ expected for Q4/2016;
- address some historically difficult features
 - ❖ checksum calculations, recalculations, (re)checks;
 - ❖ file pull/caching callouts
 - explore lightweight DPMs only working as file caches;
 - ❖ freespace reporting/quotas on directories. Can work as spacetokens.

HTTP access.

- All the pieces are now in place for HEP exploitation of the HTTP protocol
 - ❖ HTTP infra for Atlas and LHCb;
 - ❖ storage systems provide HTTP access;
 - ❖ ROOT accessible via davix;
- Need a deployment effort (HTTP Depl. TF)
 - ❖ define minimal requirements
 - ❑ <https://twiki.cern.ch/twiki/bin/view/LCG/HTTPTFStorageRecommendations> ;
 - ❖ test, monitor and validate
 - ❑ via xrootd f-stream for DPM;
 - ❑ https://etf-atlas-preprod.cern.ch/etf/check_mk/index.py ;
 - ❖ lead sites in deployment
 - ❑ start ticketing in 2016.

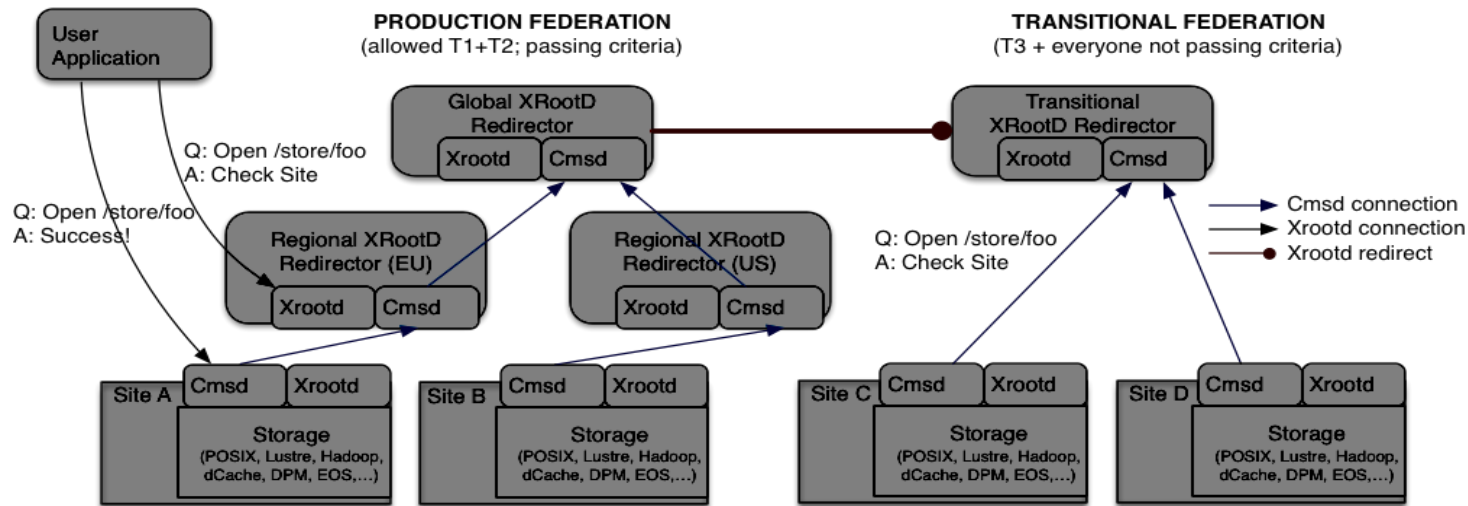
➤ Next release 3.6.0

- ❖ epel-testing before Christmas/stable in Jan;
- ❖ RPM dependency fixed;
- ❖ add option to the configuration to make the cluster ID unique;
- ❖ introduce statinfo library and support a new Name2Name interface
 - ❑ avoid the use of the xroot proxy at the cmsd: use a new plugin called statinfo;
 - ❑ the other source of calls is internal to some N2N libs: will be removed by use of a new interface, which dpm-xrootd 3.6.0 will support.

AAA federation.

➤ Transitional/Production Federations for CMS

- ❖ isolate sites which might affect AAA
 - scaling tests and other criteria;
- ❖ keep production activities intact
 - allow one-way access to the data which are not anywhere else in production.



- Major reorganization in CERN IT
- a very interesting feedback from the community
 - ❖ not so confident with upgrade and puppet;
 - ❖ some interesting setup: e.g. HDFS;
- 1.8.10 is out
 - ❖ highlights: dmlite shell, gftp redir, space report, ...;
- coming soon ...
 - ❖ LCGDM-free core based on fastCGI;
 - ❖ HTTP access TF;
 - ❖ new dpm-xrootd setup for federations
 - (aside) CMS transitional federation.

