Disk Pool Manager (DPM): From DOME to LHC Run-3

Fabrizio Furano
Oliver Keeble
Andrea Manzi
Gianfranco Sciacca (speaker)

Status - November 2019

- Disk space grows: 97PB
- Number of instances decreases: 94

- The old components (dpnsdaemon, srm, dpm-daemon, rfio…) are deprecated since September
  - No urgency, they will stay in the EPEL7 repos, simply not being ported to EPEL8

- The DPM upgrade TF is tracking/promoting the upgrade progress and the enabling of the new components (DOME) and of the WLCG Storage Reporting Record
  - https://twiki.cern.ch/twiki/bin/viewauth/LCG/DPMupgrade

- Roadmap of stability: the tech goals have been reached, hence no revolutions are foreseen in the DPM development. We expect the effort to be incrementally more into support, and discussions about the WLCG directions (e.g. sustainability and future TPC/authz)

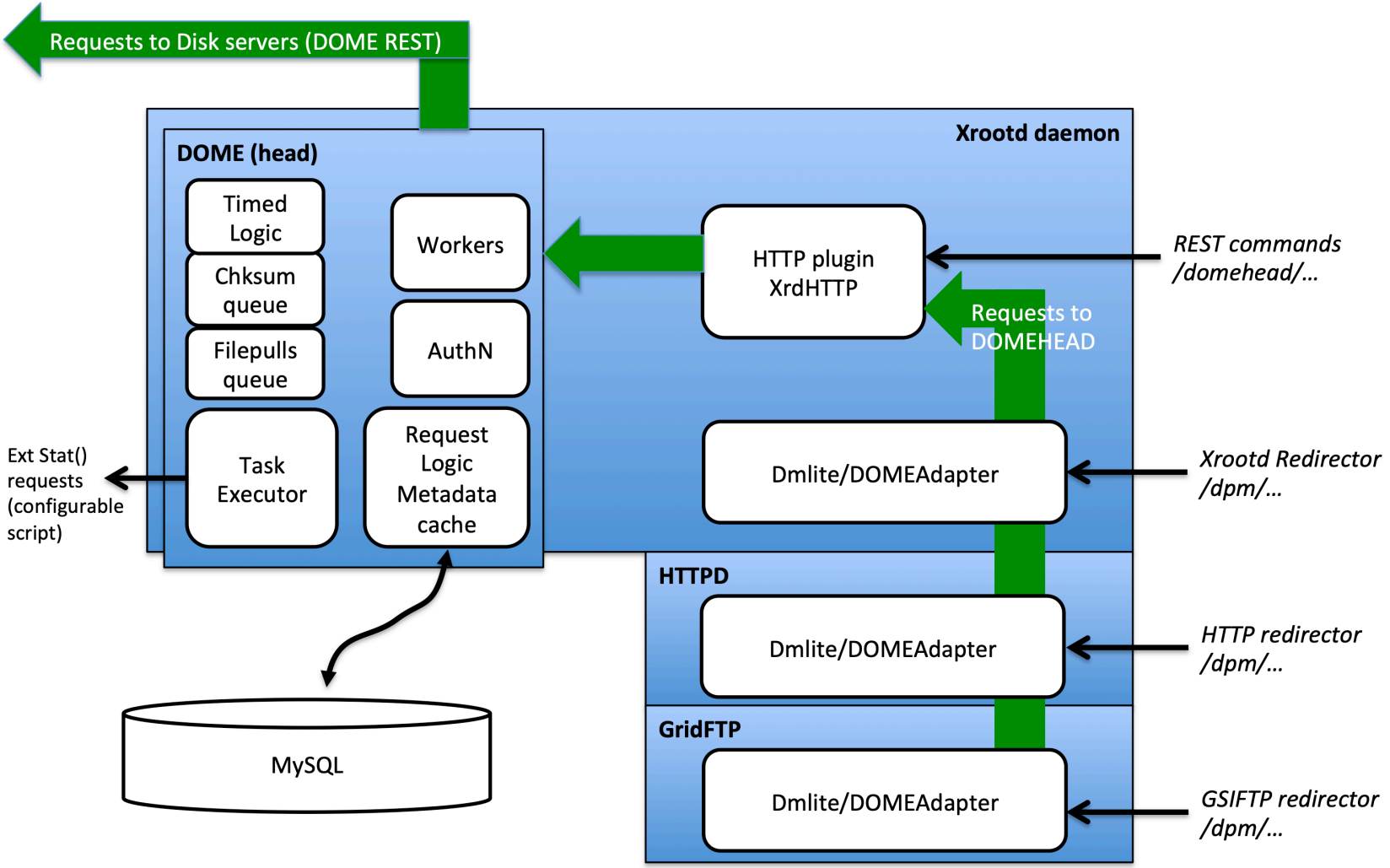
- Given the stability of the platform, the dev deltas that we see will mostly be little fixes, polishments and additions, e.g. bearer tokens, CLI commands, etc.

Direction: DOME stability and LTS

- A modern DPM is based on a version >= v1.13, with DOME enabled (and hence also the gridftp redirection)
  - DOME: ***Disk Operations Management Engine*** is the DPM core
    - manages all the metadata activity and the actions to be performed on it (e.g. checksums)
    - caches the namespace, relieving the MySQL backend
    - manages the realtime disk space and quotas

- When DOME is enabled, the site will be able to contribute to the WLCG plans for LHC Run-3 and beyond
  - Feature support: e.g. cross-protocol checksums, TPC, easier multi-site, pools as caches
  - Core performance and scalability: more than one order of magnitude than the historical components, plenty of headroom for letting sites grow
  - Future things, e.g. bearer token-based authorization

DOME in a HEAD node



DOMA

- DOMA (Data Organization, Management and Access) is the ongoing WLCG evolution activity
- DOMA produces requirements

- DPM is very much up to date with the current ones, e.g.
  - Third Party COPY (xrootd, HTTP(s), macaroons)
  - Multi-site deployments
  - DPM pools as caches of remote SEs

- The platform is very healthy and ready to accommodate the future ones and contribute to their refinement

Xrootd & HTTP(s) third party copy

- DPM supports all the current flavours of TPC
- Xrootd, HTTP(s), (plus gridftp, SRM)
- X509 delegation, VOMS
- Bearer tokens: macaroons
- SRM is easy to switch off, with big benefits

- DPM is technically well placed to adopt the “WLCG tokens” spec for TPC with Xrootd and HTTP(s)
  - A test infrastructure (and docs to use it) for that will be an useful trigger

Remote pools - disk-only sites

- Theoretically it has always been possible, yet quite tricky with libshift (among the oldest components from CERN IT!) and rfio (not much younger)
  - Without forgetting firewall rules and reliability

- UniBE has been a pioneer of this, more details here (DPM workshop)
  - https://indico.cern.ch/event/776832/contributions/3378586/

- In pure DOME mode the setup of a disk-only site becomes simpler and more robust, because the intercluster communication is more solid and simpler on the firewalls

Cache mode - Volatile pools

- There since Q1/2018, works interchangeably with all the protocols (modulo SRM/rfio)
- INFN-NA has an advanced testbed that was described at the workshop
  - https://indico.cern.ch/event/776832/contributions/3378598/

- Functionally it's quite solid and well integrated in the idea of the DPM pools
- It's a full-file buffer, that gives the 'cache experience', i.e. files get pulled from elsewhere and purged when there's the need to make space
- Supports pre-populating by construction, can also be written into normally, like a regular DPM pool

- If/when there is any content that is worth caching we will be able to understand if its cache purging algorithm needs more sophistication
  - Only prod feedback will tell the final word

DOMA and the WLCG security evolution

- DPM has pioneered macaroons together with dCache, a few years ago. They are being used in the DOMA-TPC exercise, and they are a first example of “bearer tokens”

- The WLCG working group on security has published its specifications for the WLCG profile
  - In practice, standard security tokens, plus WLCG fields

- DPMs are well positioned towards this kind of evolution in direction “WLCG tokens”. Some things will need to be better understood and agreed, e.g.
  - how will the tokens be transmitted, through the HTTP and Xrootd protocols. OpenID-Connect is simple with DPM/HTTP, but for Xrootd by now it’s unclear
  - the roadmap of the experiments for adopting this technology, and which subset of the spec will be used as a first step
  - how to match the new authz rules to the DPM model (ACLs, ownership, etc.)

- The first beta service distributing preliminary WLCG tokens for the LHC experiments is welcome. Support in DPM is in development.

Post-RAID redundancy

- Many sites use RAIDs as their backend storage
- Discussions can be heard about implementing erasure encoding in the storage system
- There are no known technical showstoppers for DPM to go to this direction, “only” its real usefulness, plus development and maintenance cost

- Also the 10x increase in data foreseen for the far-future has to be considered, together with the data loss statistics. Will this push sites to reduce the redundancy?
- It’s a balance between
  - reducing the probabilities of data loss (through various sw/hw techniques)
  - reducing the cost of dealing with data loss, at the site and in the experiments’ central services. If this cost is low, there would be little need for sophistications

- So far we see no clear answer, and clear requirements have not yet emerged for DPM sites

Conclusions

- Modern DPMs profit from a healthy technical platform
- The system accommodates all the current requirements and is technically well-placed to be extended to accommodate the future ones

- Discussions are ongoing about the directions for the DPM support, both from the point of view of the sites and of the WLCG Operations team