DPM community White Paper

F. Chollet (LAPP, CNRS/IN2P3), A. Sartirana (GRIF-LLR, CNRS/IN2P3)
on behalf of the DPM community
(a.k.a sites using the Disk Pool Manager as storage technology)

GDB, 11/03/2020

Contents

- Motivations and status of the DPM White Paper
- DPM current status *
- Perspectives and plans of DPM sites *
    - Sites evolution
    - R&D activities
    - Medium-term requirements (& long-term view)
- Statements on DPM sustainability *

* *from the sites' point of view*

Motivations

- The idea of a community White Paper was put forward during the [2019 DPM workshop](#) in Bern (13-14 june 2019)
  - Site round table and discussion « Looking into the future »
- An initiative from the DPM sites in order to
  - share views and plans
  - get and describe a global vision of the DPM infrastructure, technology and collaboration
  - identify some area of concerns and expectations for the future
  - Try a collective forward-looking reflection and put together some statements on DPM sustainability

Status of the DWP

Final version document available as [Google doc](#)

- Kick-off meeting in 13/09/2019

- Final meeting 24/01/2020

- [35 signers](#) from INFN (IT), Wigner, TW, CEA (FR), IN2P3 (FR), EGI, Cyfronet, TIFR (India) and Universities (Prague, Tokyo, Glasgow)

- Sent on the 10/02/2020 to DPM team (CERN), CERN IT, WLCG Chair, GDB &MB, LHC and BELLE II Computing coord., WLCG Ops coord. and EGI

    "Thanks to all of you who have already answered in a way or another"

- Signers were invited to send the DWP to the concerned persons in their countries or communities (funding agencies…)

- Special thanks go to Fabrizio Furano, Oliver Keeble, Marteen Litmaath and Matthew Nguyen for their useful discussions, remarks and corrections.

DPM, as a storage technology

- Today, about 85 DPM sites providing roughly 100 PB of disk storage
    - 35 DPM instances above 1PB, represent 90 % of the total capacity.
    - The largest single instance deployed by the Univ. of Tokyo provides 10 PB.
    - DPM sites represent significant resources : 40% of WLCG Tier-2 storage
- The DPM project has been able to attract and integrate into WLCG small or medium size sites with relatively limited manpower.
- The new DOME (Disk operations Management Engine), a successful reengineering of the DPM software :
    - Running a single 10PB instance, Tokyo Tier 2 explored the limits of the legacy DPM.
    - Looking into the future, it is fully admit that DOME is the way to go. Just forget about legacy stack and SRM. Greats assets of the plugin-based architecture and the puppet configuration modules
- Higher stability and much better performances for all SRM-less operations :
    - Intensive stress tests performed at Prague Tier-2 showed that metadata operations are at least one order of magnitude faster with DOME.
    - With DOME, sites expect to scale up to 10-20PB without major problems and to ~100PB with some more effort.

DPM, as a collaboration

- The DPM Collaboration started in 2013
  - More or less an agreement to ensure the software maintenance, features implementation and the support of the DPM software.
  - CERN, Czech Republic, France, Italy and the UK engaged as key partners.
- The DPM team at CERN has undertaken all the critical tasks (development, certification, integration, release and maintenance).
- The majority of contributions outside CERN came from site admins doing mainly testing on a "best effort" basis.
- The DPM collaboration was expected to attract new contributions, especially in case of new development projects but this goal was not met.
- EGI has not been explicitly part of the DPM collaboration but has been testing and distributing DPM as part of its middleware distribution.
- There is no certainty that the DPM collaboration, as it is, is capable to address the long-term challenges.
- Beside the DPM collaboration, clear benefits in setting-up WLCG Task Forces (DPM Ugrade TF), many exchanges and reactive support via the DPM users forum.

Sites evolution

DPM remains an appropriate storage technology for Tier 2 sites at least for the next 5 years.

- Manpower is unlikely to increase ; consolidated solutions that reduced operational costs are to be preferred. DPM has been one of such solutions.

- Today, it is worthwhile taking advantage of DOME.
  - Performances are there;
  - Sites have a longstanding experience in managing DPM and they have just invested effort in the DOME migration

- In 5 years from now, the community will count fewer DPM instances, each providing from 10 to 20 PB.
  - Most of the medium size sites are planning to roughly double their capacity in 5 years.
  - Other groups of sites are planning to consolidate multiple instances into a single distributed storage endpoint. This is quite easy with DPM.
  - The smaller sites providing less than 1 PB will move to a federation or to cache-only storage

Sites evolution

- Sites have in mind moving from RAID-6 ; preference would be given to industry-standard storage solutions which support distributed resilience
  - But deployment scenarios and cost savings are not straightforward
- On a longer time scale (up to the HL-LHC horizon), much depends on the outcomes of the DOMA project
  - Will the medium size Tier-2s still participate in the WLCG storage?
  - Some sites will still integrate resources and funding for non-LHC experiments
  - Local communities, end-users will still need dedicated disk space
  - Will DPM be able to evolve and provide the required functionalities?
- Sites need to think about the coherence between their current purchases and their future evolution
  - Lifetime of today's hardware puts some constraints on possible technology changes
  - Disruptive change of technology is not really compatible with yearly flat budget

R&D activities

- Italy, Czechia, France and the UK have ongoing R&D activities involving DPM technology or planned contributions to DOMA and/or ESCAPE, which is a very positive sign.
    - CERN DPM and Prague teams involved in DOMA TPC and QoS activities,
    - INFN and IN2P3 sites have deployed DPM testbed within the ESCAPE Data Lake prototype, evaluating DPM volatile pools and distributed configurations,
    - UK is working on the central banning (Argus integration with DOME flavour),
    - Please refer to the DWP for more details.
- Obviously, the goal is to identify as soon as possible the potential limitations or the further needed developments of the DPM technology
- Contribution to DPM R&D could be of interest for contractual positions as part of the ongoing or future EU projects. But this would require to suitably define priorities and long-term road-map for DPM .

Medium-term requirements

- The support for token-based authentication will probably become a critical requirement for storage:
  - Experimental support of macaroons has been enabled in DPM
  - Full support for the WLCG future token-based authorization will require some work in adaptation and integration and testing
- Still various topics for further discussion :
  - Sustainability of the standard raid-6 based HW configuration,
  - Status of the existing S3 backend plugin for DMLite currently used with Dynafed
  - Some suitable features may be missing even with the new stack (head-node and/or DB HA)
- Long-term view of DPM evolution
  - DPM volatile pool providing an integrated cache-like behavior will probably have to be improved
  - DPM over different backend solutions with different resilience profiles : a topic of interest to be discussed…
  - QOS : long-term requirements as possible outcomes of DOMA
- More discussions will be needed to suitably define priorities and possible road-map for DPM.

Statements on DPM sustainability

- DOME is a major achievement which may be converted into Petabytes of performing storage in the medium term.

- The great majority of DPM sites are not in favour of a migration to a different storage technology in the short-medium term.

- The presence of an active team with sufficient manpower to assure high level support and urgent developments is a mandatory condition.

- The DPM long-term future depends on decisions and actions taken now and in the next 2 or 3 years.

- Providing storage services requires planning over 5-6 years and thus a clear visibility, on the same time scale, of the future of the technology deployed. Having such visibility on DPM future is thus a necessary condition.

- Alternatives to DPM are available but the same above condition applies and the overall cost of a migration should be considered with great attention (not just of matter of sites).

Statements on DPM sustainability

- DPM sites and the CERN DPM team have a common interest in participating as much as possible in ESCAPE, DOMA R&D, WLCG Task Forces and to share information as much as possible.

- New requirements from experiments or DOMA, or further decrease of CERN support are identified as high risks for the future of DPM.

- If DPM is to remain as one of the technologies providing storage at HL-LHC time scale, it requires an engagement from CERN to keep the leading role in DPM development, or at least a clarification if it is not the case.

- Being able to share views among the DPM sites represents the first outcome of the DWP.

- There is a need for sites to identify any show stoppers and reasonably anticipate the necessity to change their plans or not.

Everything is up for discussion.

Sites expect some feedback before thinking about the next steps.