GRIF Storage: Perspectives after DPM

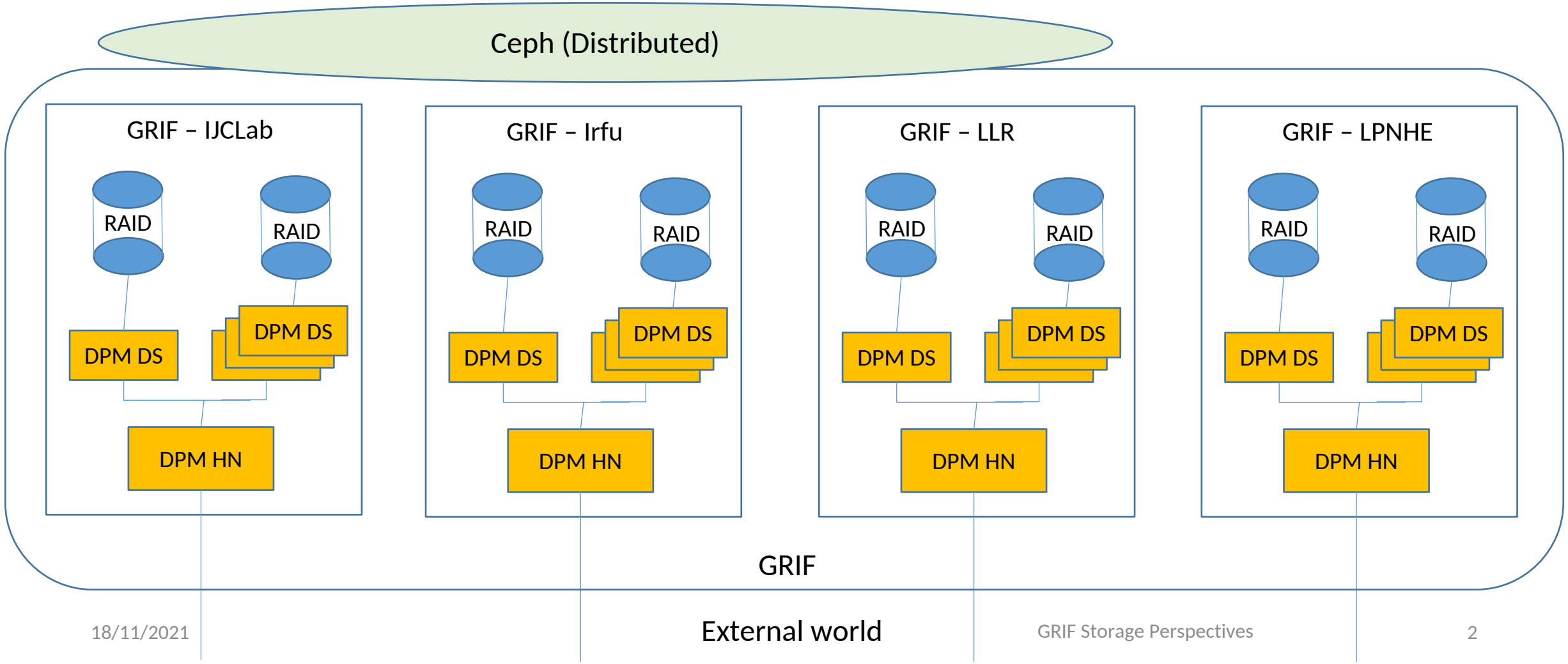
18/11/2021

Journées LCG France

Andrea Sartirana, Emmanouil Vamvakopoulos, Michel Jouvin, Guillaume Philippon

GRIF : Current Storage Infrastructure

Janvier 2021

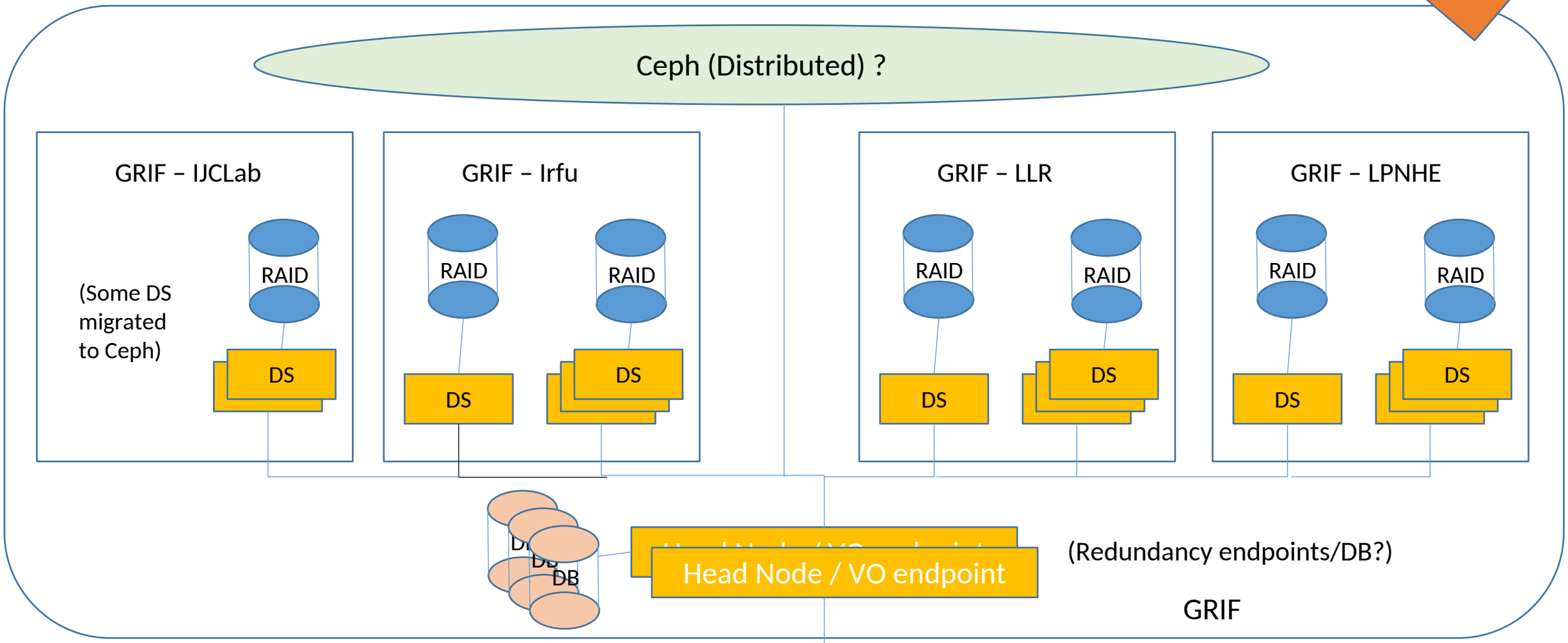


Motivations for changing

- GRIF represents a total ~10 PB but is seen as 4 medium-size sites
  - Some data duplication between GRIF subsite
  - Impact on the possible contribution to some experiments (e.g. ATLAS)
- Datalake perspective makes GRIF configuration inappropriate
  - Has the potential to be a major player of a French datalake if it can expose one GRIF endpoint for each VO
- Management not optimal: we can share experience/tools but each subsite has to be managed independently
  - Manpower/expertise not increasing is the best option… tends to decrease
- Work started on a distributed Ceph may open the way for more things in common but grid integration remains unclear

GRIF future storage configuration?

Janvier 2021



Constraints

- Moving from 5 DPM to 1 storage service prevent doing an inplace migration
  - 2 serious options identified: dCache and EOS
- Resilience target: outage at a subsite should not impact more than the data hosted at the site
  - Partial unavailability of a VO
- Possibility to run 2 instances of a VO endpoint at different subsites
- Main difficulty identified: DB redundancy
  - Multi-site relational DB clusters seems to be a possible source of tricky problems…
  - Are in-memory/non-SQL DB clusters easier?

Work done so far

- Winter 2021 : discussions with dCache and EOS experts
  - dCache : Nordugrid (M. Wadenstein, V. Garonne), AGLT2 (S. McKee)
  - EOS : L. Mascetti, A.J. Peters, E.A. Sindrilaru (developers)
  - Both products can fulfill our needs, similar configurations already existing for dCache, less for EOS (CERN Meyrin-Budapest) despite an easier database replication/failover
- Decided to start 2 PoC to evaluate the features of each product in a distributed context
  - Evaluation focused on building a resilient distributed infrastructure : multiple instances of services, failover, impact of a site outage
  - No performance evaluation : both products already demonstrated their ability to run at scales larger than GRIF
  - No functional tests: both products already demonstrated their compliance in terms of protocols

PoC : general remarks

- Work started slowly in Spring 2021 : overloaded by many other tasks, confinement didn’t help
- For each PoC, 3 sites : IJCLab, LLR, Irfu
  - A production service will also have LPNHE, may have some impact on quorum strategies
  - dCache PoC : Andrea
  - EOS PoC : Emmanouil
- Most services implemented in VMs : performance is not the issue for the PoC
- Fully functional setup for each PoC is recent: not everything tested yet
- Some initial work done to integrate configuration of these services in Quattor
  - Puppet (Irfu) still needs to be done
  - dCache more advanced than EOS

PoC : dCache

Configuration (dcache-6.2.23 + zookeeper-3.6.3 + postgres-13.1)

- 1 or 2 pools per site
- 1 door per site supporting all protocols and seeing all pools
- Central services replicated on 2 instances, both at LLR (should be strightforward to have 1 per site)
- Multipath cell message passing topology with 2 cores at LLR (will be 1 per site)
- 3 zookeeper instances at LLR
  - which allows to lose 1 instance without affecting services (see below)
- Database : 1 master and 1 standby both at LLR
  - Using WAL records and a warm-standby with streaming replication
  - Needs a HA shared FS on primary and the standby. For the moment just using an NFS mount. (see below)
- HA : set of watchdog services which detect a primary DB failure
  - Shut down the primary and promote the standby
  - Redefine hostname resolution on the central services machines to point to the new DB server
  - Homemade scripts (see below)
  - Made some tests and the services are back in few minutes

PoC : dCache

Open questions

- Zookeeper has 2n + 1 topology (allows for n failures) : how to map it to GRIF topology ?
- Database failover requires a common and HA file system between sites: use Ceph?
- So far we haven’t found an « official » HA system that fits our needs
  - E.g. Pacemaker-like solutions migrate a service IP between HA nodes. Not that easy in a WAN setup. Also fenching (STONITH) may not be trivial to implement in a our WAN setup.
  - Using an homemade solution does not sound like a good idea. The implemented set of scripts in the PoC is mostly a way to define « how it should work» and guide our quest for a suitable tool.
  - HA « policy » decisions may also impact our choice of the proper tool
    - Desired reaction time : e.g. should a DB node reboot trigger a migration ?
    - Level of automation : are we ok with manual operations e.g. to recover from a migration ?
- Local caching for reads at each site: already done at AGLT2, no plan to reinvent the wheel
- Andrea departure: need to find a replacement for taking over the work done or to decide before he leaves!

PoC : EOS

- Configuration : 2 VMs at IJCLab, 1 VM/machine at LLR and Irfu
  - EOS v5.0.2 based on xrootd version 5.x.x
  - 1 MGM per site (+ 1 MQ + 1 FST +1 Quarkdb instance)
  - one (1) «space» over one group over three fs (1 fs per node)
- The database based on QuarkDB : an efficient key/value datastore above persistent storage (rocksdb)
  - Memory cache is supported
  - A clustered database is easy to set up, with a short failover time (~1m) : a dynamically-elected master responsible for writes (base on Raft algorithm via Redis implementation)
  - Cluster quorum requires an odd number of nodes : with 3 nodes, allow the failure of 1 node (compatible with GRIF needs)
- EOS MGM access endpoint failover currently managed by DNS
  - When the master MGM node goes offline a standby MGM will become a master, this trigger also the DNS alias exchange
  - A script runs, detects if the current MGM is down, and update the DNS alias if necessary
  - 5 mn latency that could be reduced by using HAProxy in production (to be checked, limitation due to WAN topology like dCache use-case)
- One access endpoint for the three (3) sites : **eostest.grif.fr**

PoC : EOS

- Focus only on HTTPS and XROOT protocols
  - Check the voms attributes extraction and grid map file mapping for legacy compatibility
  - Maccaroons (via x509 auth) works fine for http(s),
  - Tests with TPC smoke tests and various HTTPS clients (e.g. Curl and/or gfal tools)
- The failover mechanism will be improved with native HTTP(S) redirection from slave to master on the next releases of EOS v5.
  - With this feature, we could use the simplest form of DNS access endpoint alias (e.g. round-robin)
  - For the moment it works for the XROOT protocol
- Tests with geographical access and data placement are coming soon
- OAuth v.2 is supported on EOS v5 via mapping (for the moment)
- We do not have major issues up to now but few open questions :
  - How we reorganize the disks « spaces » amongst the 3/4 sites for WLCG and other EGI VO (e.g. Space - Group - FS) ?
  - How and when we will incorporate Erasure Code Technique (EC) and get rid off of local raid configurations ?
- We are following the EOS v5 releases to test and adapt as soon as possible the new features and bug fixes in our tests.

Next steps

- Need to reach a decision at the end of the year, January at the latest
  - No inplace migration possible: means a long VO by VO migration of the current DPM (1 year planned)
  - If we want to shutdown down DPM before end of 2023, not much spare time left…
- Next GRIF meeting dedicated to storage future is next week: not sure we'll be ready to take a final decision yet…
- Dec. 8 GDB: French presentation scheduled, will see if we can meet this target!