

Impact of Data Lake Model on total cost of ownership : GRIF



***Frédéric Derue, LPNHE Paris
(on behalf of GRIF members)***

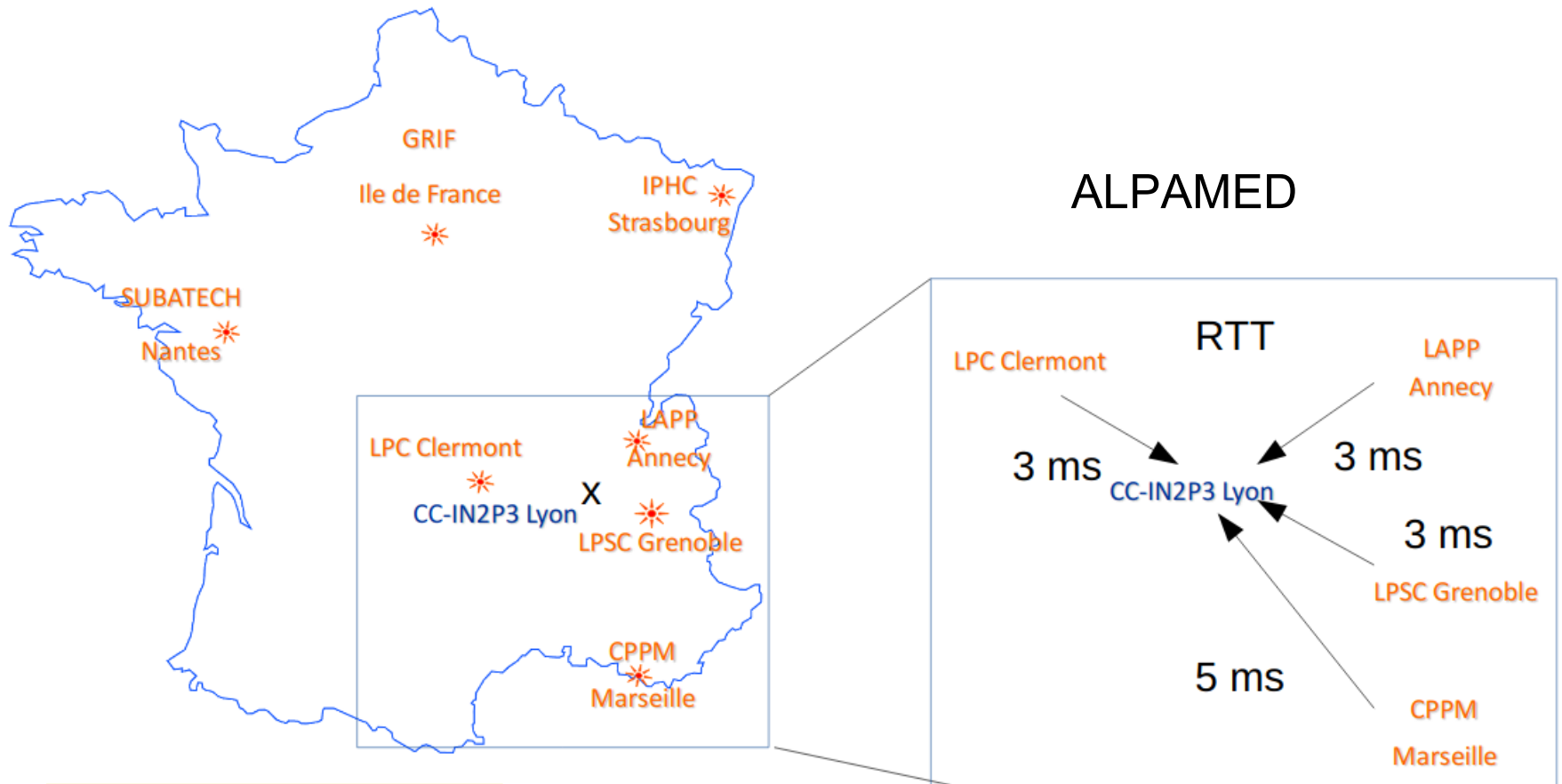
DOMA Access meeting,
20th October 2020

Existing DPM storage federation in France

- **ALPAMED**

- part of DOMA-FR project
- testbed built in spring 2019 integrated in ATLAS Grid

- **What about GRIF in Ile de France ?**



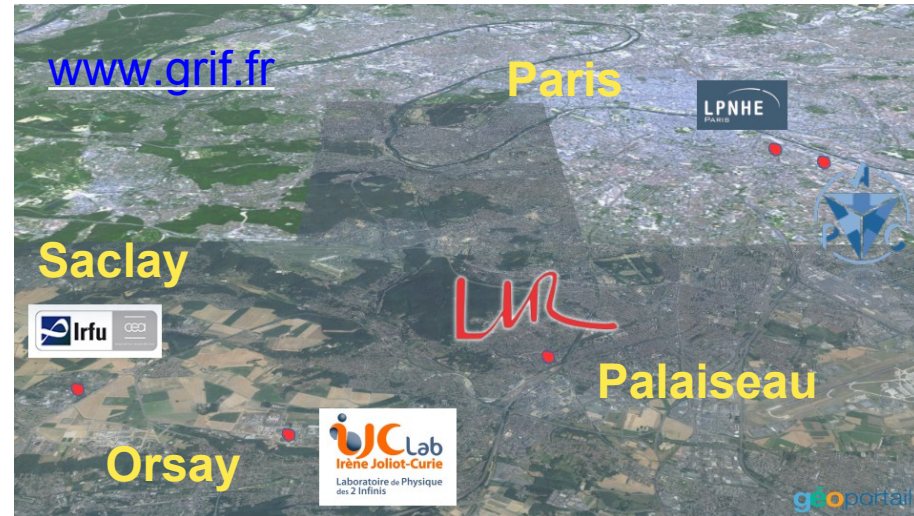
S. Jézéquel, CHEP 2019

All sites connected through CC-IN2P3 routers

GRIF a distributed Tier 2 in Ile de France

- **GRIF federates (since 2005) the effort of HEP laboratories in Ile de France**

- **Tier-2** of the 4 LHC experiments as part of WLCG/LCG-FR and some other major physics experiments (Belle II, CTA)
- provides grid resources for other experiments (VOs) of the laboratories and for interdisciplinary VOs (biomed, complex-system *ISC-PIF*)
- provides also **cloud** resources
- each sub-site responsible of its scientific project (VO, fair-share etc.)
- technical work done in common



- **Human resources**

- 1 scientific deputy by laboratory
- engineers : 10 syst admins representing ~6 FTEs (e.g ~2 FTEs for ATLAS)

- **Funding Agencies (person-power, hardware, infrastructure, ...)**

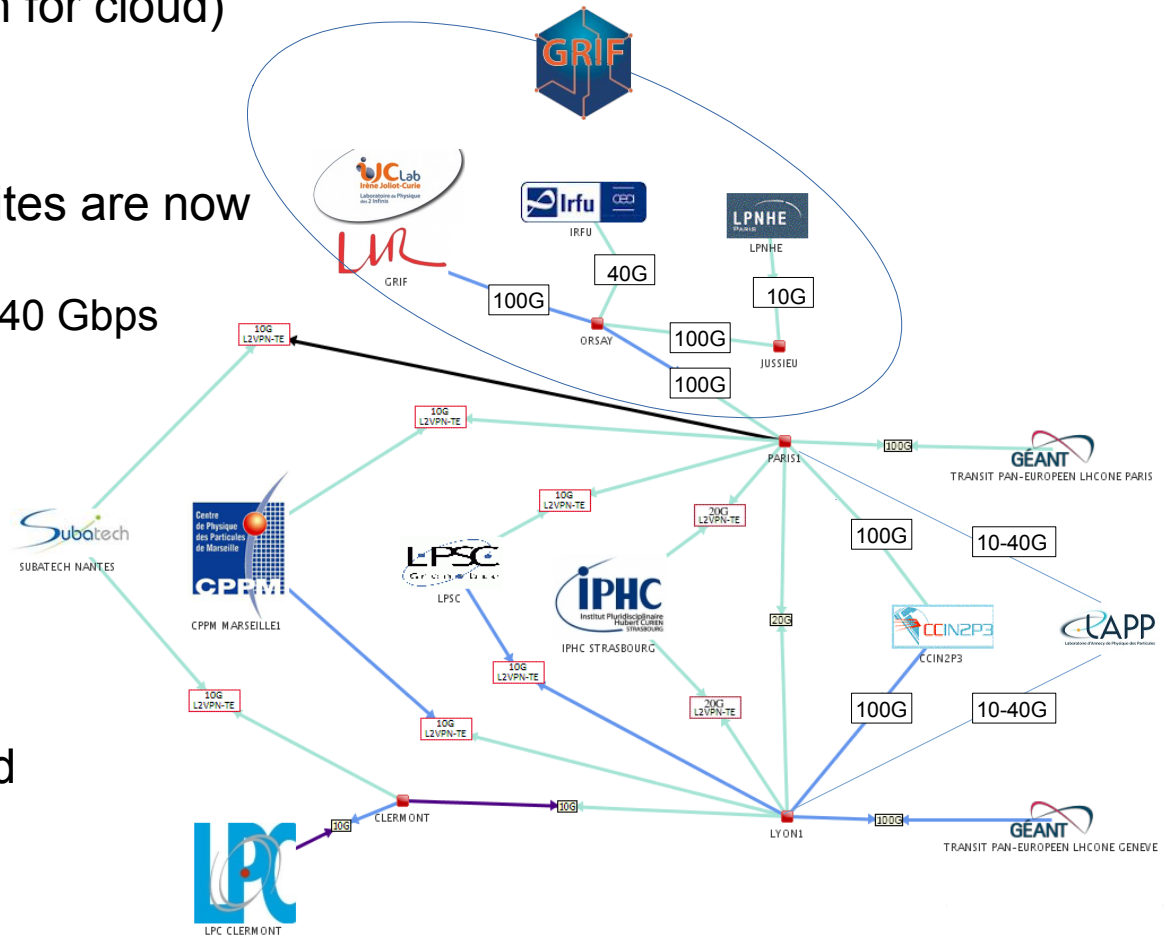
- CEA/IRFU and CNRS/IN2P3 (through LCG-FR)
- but also Ecole Polytechnique, Université Paris Saclay, Sorbonne Université, ...

● Infrastructure

- **4 computing rooms** : IJCLab, IRFU, LLR and LPNHE → importance of local fundings
- different type of CEs (ARC, CREAM, HT-Condor) and SEs (DPM for grid, Ceph for cloud)

● Network / LHCONE

- links in between GRIF sub-sites are now at 10-40-100 Gbps
 - foreseen to upgrade to at least 40 Gbps by ~early 2021
- links to other Tier 2s : mostly 10 Gbps
- links to Tier 1 : 100 Gbps (Orsay↔Paris1) and 100 Gbps (Paris1↔CCIN2P3↔Lyon1)
- future to 100 Gbps discussed between LCG-FR, CC-IN2P3, RENATER



Last update: Mon Oct 12 10:29:34 CEST 2020

● Storage end-points per supported VO

- **Tier-2** for the 4 LHC experiments : ALICE (IJCLab, IRFU), ATLAS (IJCLab, IRFU, LPNHE), CMS (IRFU, LLR), LHCb (IJCLab, LPNHE)
- **Tier-2** for CTA (IJCLab (no storage), IRFU, LPNHE)
 - soon LLR will support also CTA as other sites
 - eCheops with only one SE at LLR (+IJCLab)
- other VOs (not on all sub-sites) :
 - HEP : Auger, Belle, Calice, Clas12, Comet, HESS, ILC, T2K
 - non HEP : biomed, compchem, complex-system, france-grilles, vo.ipno

● Available resources

- Global pledge (2021)
 - for the 4 LHC experiments : 130,000 HS06, 10,000 TB
 - for CTA (6,000 HS06, 500 TB), for ISC-PIF (~20,000 HS06)
- Pledge for **ATLAS** : 52,600 HS06 and **4,900 TB**
 - 3 CE and SE (IRFU, IJCLab, LPNHE) – each SE is 1.5-2 PB
- Pledge for **CMS** : 42,700 HS06 and **3,140 TB**
 - 2 CE and SE (IRFU, LLR) – each SE is ~1.5 PB
- Non pledged resources
 - e.g 3 ATLAS LOCALGROUPDISK >1 PB (mostly beyond warranty servers)
 - also existing for non-LHC experiments
 - (also non pledged cpu)

● Human resources

- existing organisation in GRIF, and LCG-FR, aimed to optimize HR
- most (not to say all) members (engineers and physicists) are staff members
- can expect same level of HR in future ... but for sure no increase
 - details on DataLake model has little impact
 - but of course can help to redirect priorities

● Support to VOs

- increase of pledges for the 4 LHC experiments (LCG-FR and other FAs)
- continue to support non-LHC VOs, including with (increasing) storage

● Computing (CEs)

- **short term** : 2 pools for computing
 - 1 for CNRS/IN2P3 (IJCLab, LLR, LPNHE) for the 4 LHC VOs,
 - 1 for CEA/IRFU for ALICE, ATLAS, CMS
- local resources already included in grid/cloud but also batch cluster at IRFU

● Storage (SEs)

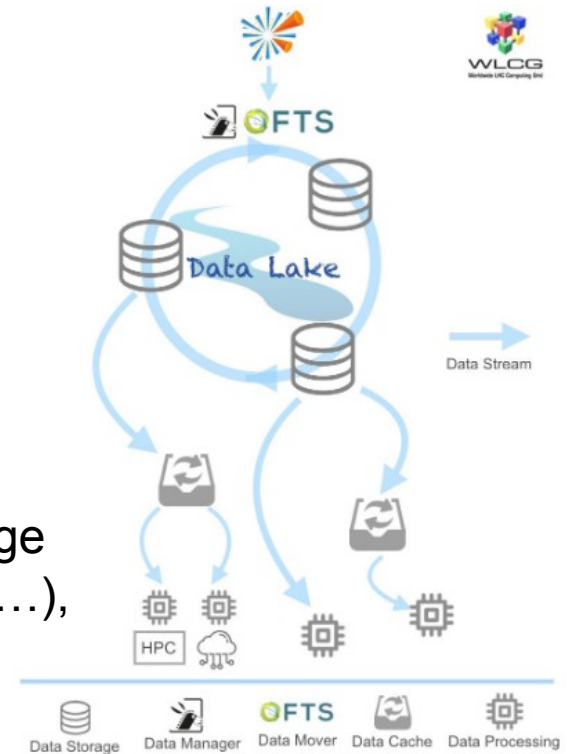
- will to switch from end-points for each sub-site to global end-points to allow VOs to access all/most of storage through a single end-point
- **middle term** : target summer 2021 for a first prototype of unified storage at GRIF, before complete deployment

● existing DataLake components

- many components already included in ATLAS Grid (Rucio, FTS)
 - allow adiabatic changes from existing infrastructure
- recent evolution for http-tpc protocols, macaroons etc
 - done for 2 sub-sites (other 3)

● Feedback

- with the **probable phasing out of DPM**, another storage technology will have to be chosen (dcache, Eos, Echo ...), which should be compatible with DOMA plans
 - also discussed inside LCG-FR sites
- remote access, protection with xcache in front of WN or in network
 - we already have 1 pool HTCondor on several sub-sites and a pattern for jobs to access data via WAN inside GRIF
 - what CMS is doing with [xrootd/AAA](#)
- hardware problem / lost files
 - ~2 power cut or pb with cooling per year
 - how many disks are lost (per month) ?
 - could we rely more on disks without warranty ?
 - automatic recovery of lost files ?



• GRIF

- gives resources to many different projects (through grid, cloud) for many different collaborations (4 LHC experiments, Belle II, CTA and other HEP, non-HEP), even incorporates computing servers of non HEP projects
- several players/FAs involved : CEA/IRFU+CNRS/IN2P3 (LCG-FR) for LHC, but many others (universities, Ecole Polytechnique), labs/groups, Ile de France Region etc...
- middle/long term evolution is driven by LHC experiments – but not only
 - syst admins have to follow needs of many different projects
- from ATLAS (CMS) point of view it is still seen as 3 (2) CE and SE
 - on short time scale reduce the number of pools for CE
 - for summer 2021 expect a first proototype to unify SEs

• DataLake model

- Diskless site model is not interesting for GRIF
- handling of storage will be modified in 2021 (less end-points) and will rely on more powerful network
- GRIF installs the different tools **needed/required by ATLAS and CMS**
- then can rely on the existing know-how from our colleagues from DOMA-FR and ALPAMED for an ATLAS DataLake

