LAL and GRIF Site Report

Michel Jouvin
LAL, Orsay
jouvin@lal.in2p3.fr

HEPiX, Budapest, April 2017
Except when noticed explicitly, the information in these slides applies to LAL only.
Infrastructure

- No major change since 18 months...

- Infrastructure built in 2013 is running smoothly
  - Hosting 250 kW IT in 30 racks currently:
    - 400 kW max, racks almost full (42U free)
  - Shared by several labs P2IO + Univ. Paris Sud: ~15
    - P2IO is the main user: 25 racks over 30, LAL: 11 racks
  - No issue so far
  - Cooling system based exclusively on rear door heat exchangers working very well: racks up to 25 kW without problems

- Extension finally starting... delayed 18 months because of administrative problems
  - Goal: 900 kW IT in 50 racks, 3 chillers + 3 power circuits in N+1 configuration
  - Presentation on Friday (9:50 am)
Infrastructure

remote monitoring
Manpower

- Quote from last site report 18 months ago: “a better period ahead of us?”... It happened!
  - Would still be better to if we were more... Main servers, grid and cloud managed by a team of 3-4 people
  - Cloud apprentice very successful and ending it June but we’ll hire him with a 3-year temporary contract
  - One other apprentice joined this year to support the Spark activity: not clear if we’ll manage to fund it longer
  - No retirements affecting the operation team in the coming years
  - Several senior person left the development team (~20 people) but we managed to get replacement for the key positions
Compute Hardware

- Servers: standardized on dual-twin servers to optimize the footprint
  - Dell C6xxx + Dell FX2 mainly
  - Dell FX2: a blade-like chassis with a shared network switch and ability to chain the switch (with redundancy) with 10 GbE uplinks
    - Replacement of the top-of-the-rack switch with a more flexible oversubscription policy

- Still 2-3 racks of IBM 3550 (8 years)… they don’t want to die!
  - Mostly replaced for critical/pledged resources
  - Keeping them as long as we have the room for them for non-critical needs

- Increasing service virtualization (cloud) to use efficiently “big boxes” rather than buying small ones
Storage:

- Storage: mainly Dell MD3xxx (iSCSI or SAS) and Dell Compellent
  - Grid: 100 TB per front-end server
  - Proved to be performing well and very resilient/dependable
  - 15-20% more expensive than a storage-in-a-box solution: can we continue to afford it?

Still have to replace a HP-based storage (200 TB): not very dependable but operational issues than in the past
- A challenge as we also need to increase significantly pledged grid storage resources

NetApp filer for critical file systems (~100 TB)
- Very reliable but not performing as well as Dell iSCSI systems
- Future not yet clear…
Ceph: about to deploy a 1 PB distributed infrastructure

- Backend for cloud storage
- Experiment (non-grid) data storage: cloud storage
- Evaluate as an alternative (or backend) for grid storage (DPM)
- Evaluate ability to host critical distributed file systems
- More with presentation on Data-NG on Wednesday (11:55) by Guillaume Philippon
OS Changes

- **Linux: business as usual... mainly SL6 systems currently, CentOS7 growing**
  - ~5 SL5 machines left... the usual long tail...
  - CentOS7: mainly service machines, e.g. web servers
    - Dependency of experiment SW for user-accessible machines
  - No specific issue in managing CentOS 7 machines (Quattor)
  - A few Ubuntu machines for accelerator controls (Tango): currently unmanaged, Quattor in the future?

- **Windows: mix of W7, W8 and W10**
  - Still a few WXP machines in electronics division: constraint of specific drivers
  - Upgrade mainly at HW renewal
  - An unidentified problem with certificate-based Eduroam authentication: only French certificates?
About to do a major change: replace the LAL-managed mail service by the Zimbra-based service operated by CCIN2P3

- Adresses will remain the same (@lal.in2p3.fr)
- Already several labs did this migration at IN2P3 but none with a size comparable with LAL
  - ~600 accounts
  - Some users still using it as an archival system: 1 user has 1 million messages in his inbox!
- A lot of testing done in the past months
- No possibility of migrating everybody at one at a low risk: migration will be done progressively, typical group by group
  - No change required on user side, except adjusting mail server if they don’t use the web interface (recommended for standard users)
- A migration report at a next HEPiX?
(Cloud) Projects

- H2020 CYCLONE: enabling technologies for cloud federations
  - Complementary (with some overlap!) to INDIGO Datacloud
  - LAL operating a testbed and doing some developments
  - Ending at the end of the year: end of cloud R&D at LAL?

- Irfu involved in H2020 INDIGO Datacloud with similar responsibilities
  - INDIGO: develop an integrated data/computing platform for scientific communities
    - SW components allowing execution of applications on Cloud and Grid as well as on HPC clusters.
  - 1 person hired recently

- EOSC-Hub (post EGI-Engage, not yet funded): LAL as a cloud provider for a Marine Competence Center
One grid site in Paris region involving LAL, Irfu and 4 other labs
- 10000 cores, 7 PB of disks
- Distributed on 4 geographical sites

Like others, have to live with tighter budgets and manpower but remains a successful site supporting all the LHC VOs and tens of others...

Main focus of current work: provide a more unified view of the site to experiments
- Currently 1 CE and 1 SE per site
- Recent advancement: a unified HTCondor pool used by different CEs
  - Currently in test between 3 sites
  - Not very different from what CMS is doing worldwide...

Next on the list: unified storage: DPM? Ceph? A combination?
Major decision taken 18 months ago: StratusLab stopped and existing cloud was migrated to OpenStack
- LAL was the only developer left in the StratusLab project: no sense
- Migration completed in 4 months beginning of 2016
  - Still ~10 service VMs in StratusLab: time missing for their migration...
- OpenStack cloud managed by Quattor

Significant increase of cloud resources funded mainly by University Paris Sud
- 2000+ cores, 400 TB
- A key component of the scientific computing infrastructure at the University
- Also used at LAL and CSNSM (P2IO) for service virtualization
Very good feedback from users since migration to OS
  > Much easier user interfaces thanks to dashboards

New use case explored: Spark on demand
  > Driven by a LSST R&D done at LAL
  > Still at the beginning: the hardest part is coupling with storage
    • Critical for Spark performances
    • Would like to use Ceph rather than HDFS: S3 interface?
  > Growing interest across scientific fields at Univ. Paris Sud
    • A new person will be hired at LAL at the end of the year (proactive replacement of a retirement in 2018)

LAL is still considering virtualization of all its grid WNs
  > With HTCondor, much easier to do
  > A good use case for containers?
“Refondation” is the name of a major project discussed presently: possible merge of 5 (over 8) P2IO labs from Univ. Paris Sud
- Potentially a 800 person lab covering nuclear physics, HEP, astrophysics and cosmology, accelerators
- 2 labs out (Irfu and LLR) are not part of Univ. Paris Sud
- Target date 2020

Goal: be in a better position to face the future challenges and have the ability to play a greater role in experiments
- E.g: potential “computing division” is 60 people

Still uncertain: a process in progress involving all personels interested through topical WGs
- Half of the concerned people participating
- Time consuming! But healthy: attempt to define collectively our future
- Computing: a lot of collaboration exists through P2IO VirtualData WG