



Data Processing and Simulation (R&D3)

Michel Jouvin (LAL)

P2IO Scientific Council , Sept. 12, 2013

Challenges (from 2012)

- Huge volume of data in all our scientific fields
 - LHC : 15 PB of raw data + 15 PB of simulated data per year
 - Upcoming experiments in other fields with similar volumes
- Both storage and computing are challenging
 - Distributed computing (DCI) emerged as the main approach
 - Some applications require HPC resources
- Emergence of clouds as a key technology for DCIs
 - Paradigm change from the original grid approach
- Infrastructures are a major issue for hosting required resources in P2IO laboratories
 - Many small, inefficient computing rooms: costly to operate, preventing synergies at P2IO level around a single large resource
- Give an increased visibility to computing expertise in P2IO

P2IO Assets

- **130 people** covering the main computing expertises...
 - All P2IO laboratories involved in VirtualData
 - Resource operations (55): strong expertise in DCI (grids and clouds)
 - SW Development (75): visualization, data management, databases, numeric computations, parallelism, new HW architectures (GPU), online and real-time environments...
- ... but expertise dispersed in 8 laboratories
 - Some partnerships already exists but more could be developed
 - Example: P2IO R&D project GridCL about GPUs and their integration into DCIs
- The existing, large computing resources
- P2IO could be a **major** contributor to experiment computing if the synergies become more active
 - A common computing platform should help to foster these synergies

P2IO Datacenters

- P2IO decided in 2010 to replace its existing computing rooms (~10) by 2 datacenters shared by all P2IO labs
 - New datacenters designed around energy efficiency to lower the operating cost of computing: at least 30% improvement
 - 1 located in Orsay (Univ. Paris Sud) and 1 located in Palaiseau (Ecole Polytechnique) for higher availability of critical services
- Funding has been an issue
 - 2 years were necessary to « solve » it
 - At the end funded by laboratories with the help of the LABEX
 - ~22% of the total infrastructure cost for the Orsay datacenter
- 2013 : major step forward for both datacenters
 - Orsay datacenter will be ready in 2 weeks!
 - Palaiseau datacenter will be ready in April 2014

Palaiseau Datacenter

- No medium or long term solution identified yet: more time needed
 - No appropriate existing building
 - Moving context due to new partners of Univ. Paris Saclay coming in the next years
- Short term: ~1/3 of the refurbished computing room in Ecole Polytechnique (Avril 2014)
 - Refurbishment in progress and driven by Polytechnique DSI
 - ~70 m² : 12 racks
 - Not extensible
- Discussions started last Spring at Univ. Paris Saclay (FCS) around datacenter needs
 - Neglected in the original plans
 - Polytechnique proposing to build a building up to 1000 m²: discussion ongoing with potential partners (may include P2IO?)

Orsay Datacenter

- Project started early 2012: reuse of an existing free 2-storey building
- Initial phase: 100 m² with a capacity of 400 kW IT (30 racks)
 - Extensible up to 220 m² and a capacity of 1,5 MW IT (84 racks)
 - Modular design allowing increment of 300 KW IT
 - Redundancy in the production of chilled water
 - Future: redundancy in the electrical power
- Delay and cost inline with expectations
 - ~950 k€ HT for the initial phase
 - Total cost of final target estimated around 2,5 M€ HT
- Move of existing resources to start in November
 - 15 racks in production from 6 labs expected at the end of 2013 (~200 kW)
 - LAL will move its entire existing computing rooms

Orsay : March 2013



Orsay : 30/6/2013



Orsay : 4/9/2013



12/09/2013

P2IO VirtualData

9

Ongoing Actions...

- Build initial governance model for this shared computing platform
 - Collective work started last Spring: a first draft discussed in each lab
 - A second draft being finalized: will be discussed with laboratory directors during autumn
 - Goal: have it ready by the end of the year
- Discussions with non-P2IO labs at Univ. Paris Sud
 - A lot of interest for this modern, efficient facility
 - University management wants to take advantage of the P2IO VirtualData process to foster synergies all around the university
 - Interest for using some P2IO computing platforms, like StratusLab cloud
 - May act as a demonstrator for further discussions at the Univ. Paris Saclay level

... Ongoing Actions

- Discussions on the future common computing platforms
 - Cloud technology should play a major role
 - Experience and expertise gained with StratusLab project should help
- Create a network of computing expertise around P2IO and its 130 people involved in computing
 - **Bottom-up approach**: helps to get people involved
 - Build upon success of people sharing the proposed vision
 - Shared R&D actions, e.g. **GridCL (P2IO R&D 2011, 60 k€)**
 - General meetings of computing people in P2IO held in 2012 and 2013 with a good participation (~50)
 - Need even more time than building a datacenter...!!!

Conclusion

- P2IO members have a long tradition of involvement in the computing of their scientific communities
 - Operating several significant resources
 - Playing an important role in SW development
 - Involved in several R&D projects/activities (SW and HW)
- P2IO has the potential to build a world-class network of computing expertise around a large, common flexible platform
 - Governance to be defined: **must keep people in their labs**, close to the users
 - Take new responsibilities in the computing of future experiments
- **Optimizing the operational cost** of infrastructures is critical
 - 2 new shared computing rooms for replacing the existing ones (10+)
 - Strong focus on energy efficiency for new rooms
 - Major step forward accomplished in 2013