



# Data Processing and Simulation (R&D3)

## Michel Jouvin (LAL)

P2IO Scientific Council , June 27-28 2012

# Challenges

- 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
  - 130 people covering all major computing areas

# Computing Projects in P2IO

- GRIF, a large grid site
  - 4 P2IO laboratories involved in GRIF
  - 80% of GRIF resources located in P2IO
  - **A successful experience of 6 years in the common technical and scientific management of the platform**
- StratusLab : an IaaS cloud testbed
  - 250 cores, 50 TB of disks
  - Possible doubling this year
- 1 HPC machine for astrophysics simulations
  - 600 cores, 50 TB disks, Infiniband interconnect
  - Doubling planned this year
- IDOC : Integrated Data and Operation Centre
  - Astrophysics

# 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): visualisation, data management, databases, numeric computations, parallelism, new HW architectures (GPU), online and DAQ...
- ... but expertise dispersed in 8 laboratories
  - Some partnerships already exists but could be promoted
  - 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 VirtualData Actions

- Create a network of computing expertise around P2IO
  - **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, 70 k€)**
  - First meeting planned beginning of July
- Build a shared datacenter to host P2IO computing resources
  - In fact one datacenter on 2 sites to allow redundancy for critical services
  - 2 chosen locations: Orsay (U..Paris Sud) and Palaiseau (Polytechnique)
    - Building in progress for the Orsay site: readiness Oct. 2013 ?
    - Several possible rooms under evaluation for the Palaiseau site
    - **Supported by LABEX P2IO (R&D 2012, 250 k€), 500 k€ by the labs**
- Use the platform and the network of expertise around it to foster relationships with University Paris Sud and Ecole Polytechnique
  - In the future, IDEX partners

# Expected Results

- Short term (1-2 years)
  - Network of expertise: shared evaluation of new technologies like GPUs, clouds... rather than duplication
  - Infrastructure: replacement solution for existing ones with too many problems or too small/inappropriate, better level of resilience, capacity for the needs of the next 3 years
  - Services: redundancy for a small set of critical services
- Medium term (3-5 years)
  - ROI for first phase of the two new computing rooms
  - Flexible common computing platform for P2IO managed by a “common team” of people
- Long term (6+ years): new responsibilities in experiment computing
  - T0/T1 for some small/medium size experiments
  - Remote control of experiments (satellites, detectors...)

# 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
    - Leverage on GRIF experience success
  - 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