

P2IO VirtualData goal, supported by R&D3 program of LABEX P2IO, is to build a computing platform common to all P2IO laboratories to address their needs in simulation, processing and storage of very large data volumes. Since 2010, this work has been driven by a working group involving members of the eight laboratories of the LABEX (~15 people).

To reach this goal, VirtualData has defined three objectives:

- *Foster synergies between all the computing expertizes present in the LABEX*, around 130 people, representing an important asset for our scientific communities. The present fragmentation of these competencies in our different laboratories often prevent them to have the visibility to take major computing responsibilities in experiments, despite the large number of P2IO scientists and engineers involved in computing activities. In particular during these last years, P2IO people were involved in several R&D activities around distributed computing infrastructures (grid & cloud), new processor architectures and software challenges resulting from these new infrastructures and architectures.
- *Optimize the operational cost of P2IO computing resources*. The key factor for this optimization is the energy efficiency of computing rooms used to host our resources. Until last year, every laboratory operated one or several small rooms, most of them quite inefficient (PUE >2) and having limited or no built-in redundancy, with in several cases infrastructure problems and instabilities. VirtualData proposed to build 2 modern, complementary machine rooms designed for hosting efficiently scientific computing resources, in particular high density for dissipation (25 kW/rack) or weight (1.5T/rack), one located at Université Paris Sud (Orsay valley) and the other at Ecole Polytechnique (Saclay plateau) to allow implementation of high-availability services.
- *Maximize effective usage of resources* by developing common platforms. P2IO is operating a large quantity of computing resources and operation cost can be optimized by maximizing their effective usage. The idea is to enable a dynamic sharing of resources, taking advantage of the improvements of virtualization technologies in the last years, in particular for its use in the scientific computing context. The cloud technology for example may allow to deliver a flexible common platform appropriate for many P2IO computing needs. Delivering and operating such a common platform would also help to foster the synergies between the different computing expertizes in our laboratories, both for operations and development activities.

VirtualData Achievements in 2014

The main achievements of this last year are related to the new machine rooms with both the first year of operation for the new Salle Vallée (Orsay) machine room and the end of the building phase for the Salle Plateau (Ecole Polytechnique) machine room:

- Salle Vallée : what has been done is the first phase of a high energy efficiency infrastructure (PUE < 1,3), in an existing free technical building. Completed in time on inline with the planned budget in October 2013, it can host 30 racks and 400 kW of IT hardware. Since October 2013, we ramped up progressively the installation of resources in the new computing room (move of existing resources) and P2IO is now occupying 24 racks (over a

capacity of 30). All six P2IO laboratories from Université Paris Sud (CSNSM, IAS, IPNO, IMNC, LAL, LPT) have moved part of their resources (all the resources for LAL) to the new computing room. The existing computing room can be extended up to a capacity of 84 racks and 1.5 MW of IT hardware if necessary, by increment of 300 kW. The first year of operation allowed to confirm pertinence of the design choices and already improved a lot the infrastructure availability, compared to previous machine rooms, thanks to the redundancy of the most critical elements (in particular the cooling infrastructure). We are not starting to work on the possible funding scheme for the first extension (~800 k€).

- Salle Plateau : for the short term, P2IO is involved in the refurbishment of the Ecole Polytechnique computing room (200 m²) that has been initiated by Ecole Polytechnique DSI in which P2IO will be able to use ~70 m². P2IO brought his experience gained with Salle Vallée in the technical discussions to design an efficient computing room in the constraints of the existing environment and participated to the follow-up of the building work. The existing room being occupied, the work had to be planned in two phases: first one has been completed at the end of the August 2014 and the second phase is about to be completed (end of November). This computing room will host hardware from other P2IO laboratories, in addition to LLR resources, for the services requiring high availability. This possibility will be explored during the next year.

These two rooms, in addition to the Irfu recent computing rooms currently dedicated to Irfu resources, should allow P2IO to host in good conditions its resources during the next 2 to 3 years. Current Salle Plateau will make possible for P2IO to wait for a more long-term and extensible solution, as currently discussed between partners of Université Paris Saclay, including P2IO. In Salle Vallée a significant part of the moved resources are already pretty old and their planned renewal in the coming years by denser hardware should allow to absorb the planned increase of resources.

The first year of operation of the Salle Vallée in common by the six laboratories from Université Paris Sud fostered the links between people from all these laboratories, already established during the construction phase and permitted to involve new people in charge of computing operations (computing infrastructures, network, server management). In particular a rota ("tour de garde") has been organized for the every day monitoring and management of the computing room.

The VirtualData working group, beyond the computing room work, has been meeting regularly, every 3 weeks, to follow-up all the other related actions and in particular to discuss how to foster synergies in areas other than computing operations, for example online activities or software development for physics.

During this last year, we also tried to formalize the new platform governance, based on the experience gained with the first year of Salle Vallée. Unfortunately, it was a too busy period and we lacked of time to complete the governance definition, based on the principles agreed between VirtualData and the laboratory directors. This remains an important and urgent topic for the coming year: even though it doesn't impact shared computing room operations, in particular because a MoU was signed between involved directors on the funding of the Salle Vallée operations, it is necessary to go further in discussions with our partners, in particular at University Paris Sud.

VirtualData is a project supported by computing persons from our eight laboratories. Because of this, the working group is very sensitive to the needs and challenges of a computing resource closed to

physicists and other users from our laboratories. We continue to believe that our initiative is very complementary, rather than in competition, to the French national computing centers. For this reason, we decided on a model based on people remaining attached to their laboratories, working as a network of expertise around the common infrastructure and platform. We strongly believe that this must remain the model for the future.

VirtualData in its Local Environment

Since the very beginning of our project of shared computing rooms, VirtualData paid a great attention to inform our local partners, beyond P2IO and discuss our initiative with them to ensure that it was consistent with other related initiatives locally and to allow a wider collaboration with the partners who could be interested by our infrastructure and the dynamics around it. These main contacts have been with the following partners:

- Université Paris Sud: Beginning of 2013, university management setup a working group on Scientific Computing to prepare the university policy on this topic for the next 5 years (quinquennal). Historically, there was no real scientific computing policy at Université Paris Sud. All the departments/communities from the university participates to this group. One of the explicit objective of this group is to leverage on P2IO initiative to initiate a similar dynamics around a shared hosting facility and a common platform university-wide. Since the group inception, there were a lot of fruitful discussions on the needs and requirements and on the respective role of P2IO, Direction Informatique of the university and other partners (P2IO made clear that it had no vocation to substitute the Direction Informatique). In the last year, there has been a growing interest by several laboratories and institutes to use the Salle Vallée to host their resources or to use the StratusLab cloud currently operated by LAL. To support this interest, University Paris Sud funded a project presented in common by all departments of the university to double the resources in the StratusLab cloud: this will be deployed in 2015.
- Université Paris Saclay: the coordination of the different initiatives around datacenters has resurrected this year to prepare a project (HPC-Data@UPSa) submitted to the CPER (Contrat de Plan Etat-Région). Despite VirtualData was not directly involved in this project (VirtualData funding is part of P2IO project in the CPER), it was associated to preliminary discussions that were convened by CNRS. Shared hosting as done by VirtualData has been used as a reference for the shared datacenter proposed by Ecole Polytechnique in the context of this project. We are still waiting the final decisions for the CPER projects and hope that it will allow the emergence of a shared, extensible facility similar to the Salle Vallée, suitable for P2IO needs, on “plateau area” to replace the current non-extensible machine room at Ecole Polytechnique. From the discussions around this project, it is now clear that the Salle Vallée will play an important role in the future datacenter infrastructures of Université Paris Saclay, building on what has started at Université Paris Sud.

Another asset of P2IO VirtualData in its scientific environment is the strong links established in the last ten years between some P2IO laboratories and the computer science community, structured around DIGITEO. This collaboration is now expanding with the Center for Data Science (CDS), funded

by Université Paris Saclay in the framework of its AAP Recherche 2014. CDS is gathering together a large number of data scientists from different communities and disciplines of the university and has emerged as one of the visible center of excellence for data science in the world. One of the two co-leaders is Balasz Kegl from LAL who defines himself as an “embedded data scientist”. This strong link with the computer science community on one side, the coexistence on the same campus of an advanced expertise in (big) data analysis and in storage and data management technologies on the other side, give P2IO a unique opportunity to play a major role in these key challenges for our scientific communities.