

VirtualData (R3) : Perspectives

Journée du LABEX P2IO

20 novembre 2015

S. Anvar, V. Givaudan, P. Hennebelle, M.
Jouvin, G. Poulleau

Sommaire

- Infra et ressources P2IO
- Notre vision toujours d'actualité !
- Passer de la gestion des ressources au développement
- Cycle de vie des données
- La simulation
- Développement On-line
- Contexte Paris-Saclay
- 1^{er} bilan du travail collectif
- Actions à entreprendre

Infra et ressources P2IO...

Mise en production d'une plateforme d'hébergement mutualisée sur 3 sites:

- Création d'une salle dite « vallée »
- Jouvence de la salle École Polytechnique
- Prise en compte d'une 3^e salle Irfu

Au départ un **besoin commun**

Important travail en réseau de tous les laboratoires du Labex, autour de la gestion des ressources

En lien avec la DI de l'université Paris Sud

Exemple de mutualisation de ressources réussie sur Paris-Saclay...qui commence à se savoir!

... Infra et ressources P2IO

Ne pas laisser retomber les efforts...

- Travailler à la mise en place de *services mutualisés*
- Volonté d'exploiter les 3 salles pour faire du fail-over de services
- Reste encore du travail sur le monitoring des ressources

Meilleure intégration des ressources Irfu:

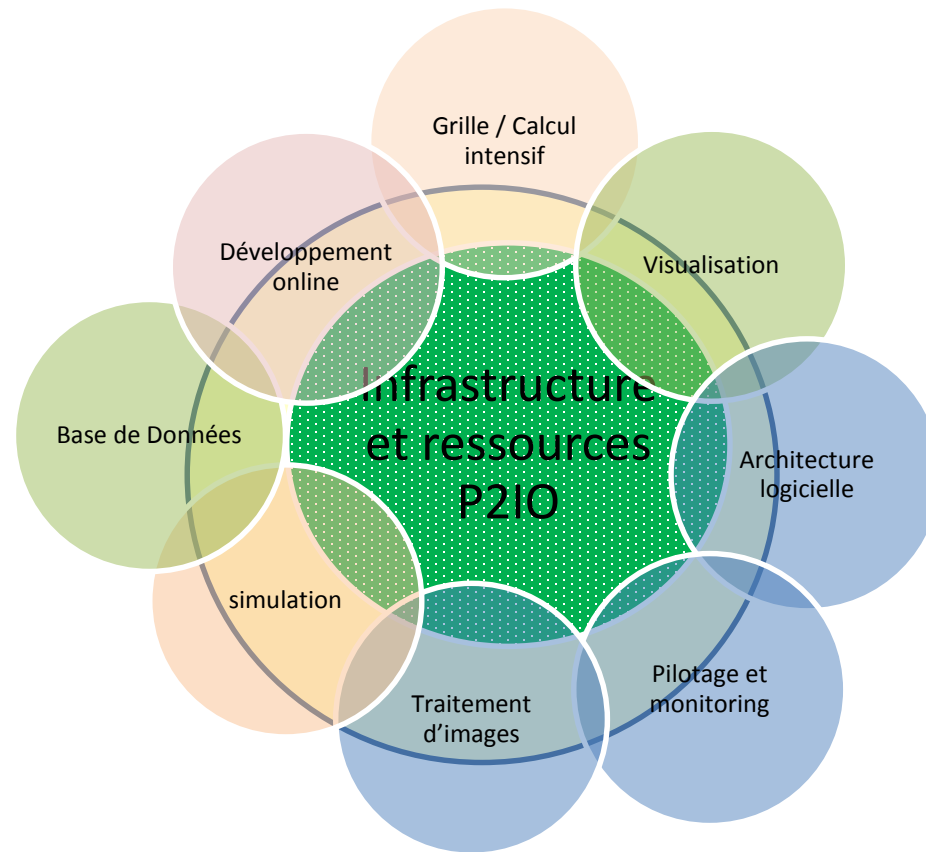
- Pb de sécurité qui devraient se résoudre avec la segmentation réseau
- Machines pour la mise au point des codes

Avenir de la dynamique enclenché par P2IO, particulièrement Vallée

- Extension financée par CPER, en lien avec UPSud (UPSay?)

La gestion des ressources et des services restera probablement le côté le plus visible de notre travail

Notre vision toujours d'actualité !



Challenge : passer de la gestion commune des ressources à la synergie des compétences

- Objectif : augmenter notre expertise et notre capacité d'intervention dans les expériences
 - Tout en restant totalement ancré dans les laboratoires, au contact des utilisateurs
- Mobiliser nos compétences en développement
- Production, accès, mise à disposition, qualification et valorisation des données, architecture logicielle,...
- S'insérer dans le contexte de Paris-Saclay : modélisation, simulation, HPC, Center for Data Science, Maison de la Simulation...
- Créer les conditions pour que les gens se rencontrent, échangent et mettent en commun expertises, méthodes et outils.
- Plus compliqué que la gestion de ressources...

Cycle de vie des données

Enjeu: données à haute valeur ajoutée qu'il faut rendre disponibles via des interfaces ergonomiques et dont les volumes augmentent considérablement !

IAS (IDOC) , Irfu(COAST) , IMNC (données patients), IPNO/LAL/LLR (Grille/LHC), ...

Traiter les notions de gestion, de confidentialité, de sécurité et d'archivage des données (Data Management Plan)

- Mise en place d'une architecture distribuée, évolutive , résiliente et performante de ces données (projet emblématique Data-NG)

Besoin de développer des interfaces d'accès, des bases de données (ex: mise en ligne de résultats de simulation numériques), d'outils communs de présentation de jeux de données → **recensement des besoins à faire.**

Contexte Paris-Saclay:

- SPU : projet d'un centre de données spatiales: Observatoire Virtuel
- P2I: données LHC, nucléaires, astro, médicales, etc.

Simulation/modélisation

Il existe des compétences réelles et diverses: Utilisateurs et contributeurs des codes

- Simulation essentiellement sur des machines nationales ou internationale mais besoin d'élaborer des modèles et de tester les simulations sur des ressources locales à VirtualData
 - HPC : GENCI (France), PRACE (Europe), certains supercalculateurs US
 - HTC: grille (ressource mondiale), dont une composante locale (GRIF : 10K cœurs) + « petits » clusters de calcul avec GPU
- Contribution aux logiciels des grandes expériences: des développements généralement très globalisés (GEANT4, GATE, MOKKA, NPTOOL, MUR, SLAM, COAST, Solarsoft,...) mais des responsabilités/contributions locales importantes
 - Nombreuses synergies possibles, recenser les expertises et les besoins

Développement On-line

Le on-line historiquement proche de l'électronique mais se rapproche de plus en plus de l'informatique offline

Déjà des actions en cours:

- **Collaborations autour de projets, à élargir dans P2IO Irfu/CSNSM :**
gateway pour DAQ
 - CSNSM/IPNO : ANDROMEDE
 - CSNSM/LAL : CORTO
- Echanges de connaissance, en particulier via des réunions de type « réseau » P2IO, mais aussi Paris-Saclay (loops, ...)
- Architectures, matériels, langages,.. spécialisés,
- Discussions électroniciens/informaticiens sur les protocoles de sortie

Contexte Paris-Saclay (1)

Paris-Saclay propose un environnement unique pour interagir avec d'autres communautés, en particulier la recherche en informatique

- Faire émerger de nouvelles approches pour nos défis
- Fournir à la RI un « terrain d'expérimentation »

3 partenaires principaux avec qui des liens existent déjà et gagneraient à être développés

- Maison de la Simulation : problématique HPC et parallélisme, lien avec la communauté GENCI. Contact principalement via l'Irfu/AIM.
- LRI : forte expertise en parallélisme, data science et visualisation. Lien historique avec le LAL.
- Center for Data Science : « knowledge extraction ».
 - Réseau très diversifié de visibilité mondiale pour explorer de nouvelles approches pour l'analyse de données.
 - Plusieurs points de contact avec P2IO/SPU/P2i.

Contexte Paris-Saclay (2)

Paris Saclay, c'est aussi les départements dans lequel P2IO est impliqué...

- P2I : réflexion dans le cadre du document de stratégie autour des enjeux informatiques couvrant la perspective VD
 - Personnes communes impliquées : forte cohérence avec la vision VD, à la fois développement et infrastructure
- SPU : Structuration différente, plus répartie (GTs), mais identification effectuée des axes stratégiques et des projets majeurs à conduire au cours des prochaines années
 - Ouverture vers d'autres communautés...P2IO ?

Bilan du travail collectif

Après 2 ans d'expérience, très positif :

- Projets communs (financement puis réalisation)
 - Des actions parfois un peu limitées aux même personnes.
- Elargissement de la dynamique au développement a commencé avec le projet emblématique AccelSW
 - LAL, LLR, Irfu/AIM, Irfu/SPHN
- Contacts avec les informaticiens de l'Irfu moins fluides que dans les autres labos, probablement à cause de la répartition dans les différents services
 - Evolution récente grâce au projet AccelSW
- Projet d'enseignement (License pro) en stand-by qu'on souhaite reprendre
 - Manque de perspective et manque de temps...

Actions à entreprendre

- Construire une plate-forme de services au-dessus de la plateforme infrastructure : 1 an
 - Le plus urgent : stockage distribué (ex projet emblématique)
- Réflexion développements communs logiciels/outils autour des Données : 1 an
- Mise en œuvre de serveurs de données de simulation sur infrastructures P2IO (e.g. ERC MagMist Irfu) : 2 ans
- Production d'un document recensant les besoins autour de services pour segment sol de manips spatiales : 1 an
- Elargissement et approfondissement du travail en réseau: 1 an
 - Renforcer les réseaux existants sur Paris-Saclay en incitant les actions émanant de P2IO à aller vers ces réseaux (Loops, Argos,..)
 - Mise en œuvre de services et d'outils permettant à la communauté d'organiser facilement des rencontres d'experts de taille modeste (support conférence, salle dédiée, environnements reconfigurables,..)

Le Role du LABEX

- Historiquement très important comme incitateur
 - Même si le financement essentiellement le fait des labos (~80%)
- Un financement par les seuls laboratoires : risque de fragmentation
 - Difficile d'attendre le même soutien de tous les laboratoires
 - Particulièrement vrai pour l'infrastructure de stockage distribué
- Echec des 2 projets emblématiques pose la question du soutien à la dynamique d'une plateforme emblématique de l'ambition P2IO
 - Retour malgré tout positif sur le fond des projets emblématiques
 - Volonté forte des acteurs de continuer la dynamique
 - Quel financement du LABEX après les projets emblématiques ?
- Renforcer la dynamique au niveau développement logiciel
 - Un potentiel considérable dans P2IO mais peu visible (fragmenté)
 - Des enjeux majeurs (ex: parallélisation) pour nos disciplines