

# Traitement de Données et Simulation

Michel Jouvin  
LAL, Orsay  
[jouvin@lal.in2p3.fr](mailto:jouvin@lal.in2p3.fr)

Journée Inaugurale LABEX P2IO



# Sommaire

- Les besoins
- Le capital existant
  - › Compétences et ressources humaines
  - › Les ressources
- Les nouvelles ambitions
- Les projets
- Conclusions

# Les Besoins

- Les thématiques P2IO reposent sur la production et l'exploitation de grandes masses de données
  - › Volume des données simulées ~ données expérimentales
  - › LHC : **15 PB** de données expérimentales par an
  - › Futures expériences des autres domaines : volume similaire
- Stocker et traiter (efficacement) ces données est un défi technologique
  - › **Centres de calcul isolés** : Impossible, besoin d'infrastructures de calcul distribuées
  - › **HPC et/ou GPU** : indispensables pour certains besoins mais leur utilisation efficace est complexe et leur coût élevé
  - › **Grille** : solution LHC n'est pas applicable à tous les besoins
  - › **Cloud** : élargit les promesses de la grille en apportant une flexibilité accrue tout en optimisant les coûts

# Des Compétences Multiples

- Concentration relativement unique de compétences couvrant l'essentiel des domaines informatiques (~130 personnes)
  - **Développement** (~75) et **mise en oeuvre** de ressources (~55)
  - Visualisation, gestion de données, parallélisme, base de données, online et acquisition de données, calcul numérique, ...
- Rôle important dans les développements logiciels des expériences et projets de nos laboratoires
  - Inclut veille technologique et R&D logicielle
- Participation à des actions de R&D sur les architectures du futur
  - GPU : PetaQCD (ANR), GridCL (LABEX P2IO)
  - Grilles : forte participation aux projets européens EGEE (2001-2010)
  - Cloud : StratusLab (EU), EDGI (EU)

# Les Ressources P2IO

- GRIF, un site majeur de la grille EGI/WLCG
  - › 4 laboratoires P2IO impliqués dans sa mise en œuvre
  - › 8000 cœurs, 2,5 PB de disque dont 80% localisés dans P2IO
  - › 6 ans d'expérience de travail en commun quotidien pour la gestion scientifique et technique du site
- StratusLab : embryon d'infrastructure cloud « IaaS »
  - › 250 cœurs, 50 TB de disque
  - › Extension prévue dans le cadre de GRIF
- 1 machine HPC (calcul parallèle)
  - › 624 cœurs, 52 TB de stockage rapide, interconnexion Infiniband
- Des clusters de calcul significatifs dans plusieurs laboratoires
- IDOC : centre d'opérations et de données spatiales (IDOC)

# Un Peu d'Histoire

- Groupe de travail réunissant tous les laboratoires du LABEX depuis 2009 dans le cadre du Plan Campus
  - › VirtualData : CSNSM, IAS, IMNC, IPNO, Irfu, LAL, LLR, LPT
- Des réunions régulières pour préparer chacune des échéances importantes de P2IO depuis sa "préhistoire"
  - › Rapport sur les plateformes technologiques P2IO (2010)
  - › EQUIPEX COGIS (2010)
  - › LABEX P2IO (2010)
  - › Réflexion sur une infrastructure informatique mutualisée (2011)
- Depuis l'origine, volonté de mettre en synergie les compétences présentes autour d'une infrastructure flexible et performante
  - › Mission "Transform" du LABEX !

# Les Ambitions

- Constituer en France un pôle informatique capable de relever de nouveaux défis
  - › Centre T0/T1 de données : exigence de disponibilité inatteignable pour un laboratoire seul
  - › Chef de projet dans l'informatique des expériences
  - › Responsabilité dans le contrôle à distance des expériences
- Augmenter la disponibilité des services tout en optimisant le coût et l'impact environnemental des infrastructures
  - › Mise en place d'infrastructures d'hébergement mutualisées
  - › Permettre une allocation flexible ("élastique") des ressources unifiées
- Participer à la R&D sur les architectures du futur
  - › Calcul distribué, processeur et parallélisme, logiciel

# Les Projets...

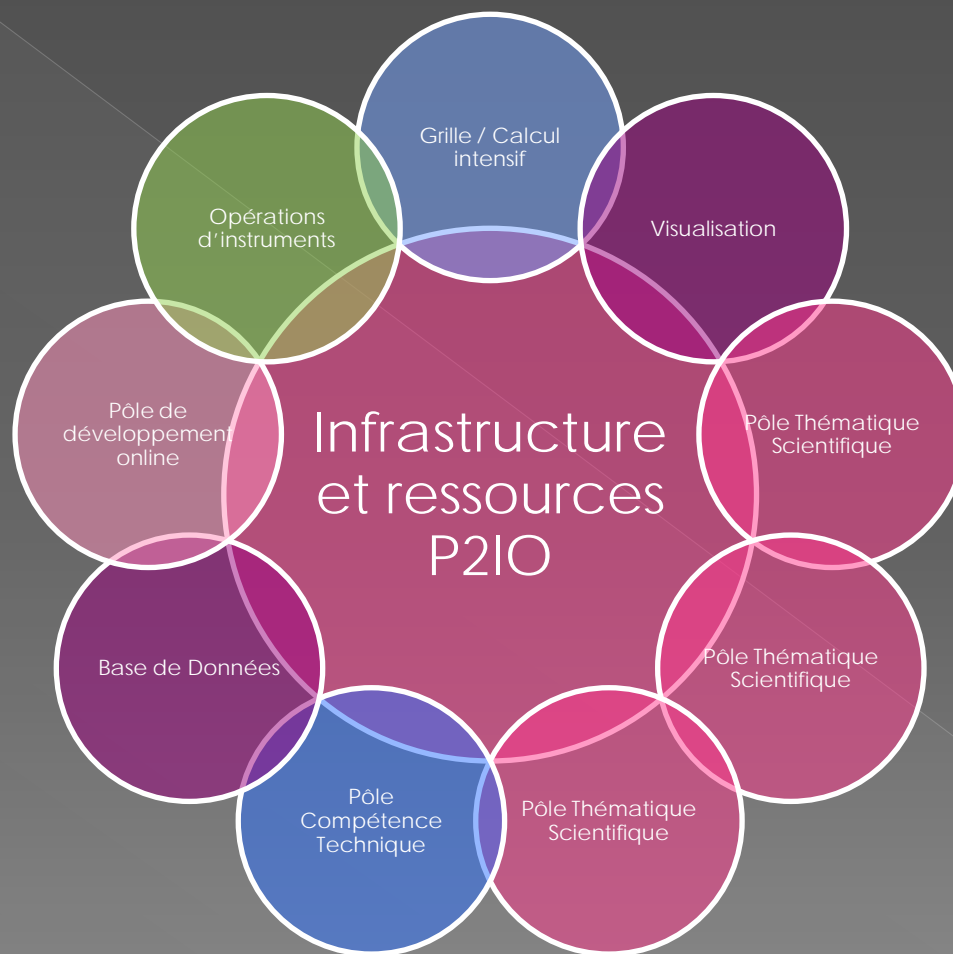
- Mise en place d'une structure de collaboration entre les "informaticiens" du LABEX P2IO
  - Mise en place d'un réseau d'experts dans les différents domaines informatiques (développement et exploitation)
  - Rencontres thématiques régulières
  - Accroître les échanges de compétences inter laboratoires, avec une ouverture à d'autres laboratoires de l'université intéressés
- Salle machine mutualisée à "2 jambes"
  - Une ressource unique sur 2 sites
    - Projet actuel : Ecole Polytechnique et vallée d'Orsay
  - Redondance des services les plus critiques entre les 2 salles
  - Optimisation des coûts de fonctionnement et de l'efficacité énergétique



# ... Les Projets

- Poursuite des projets R&D sur les architectures matérielles et les infrastructures de calcul distribuées
  - PetaQCD : besoins spécifiques à LQCD (calculs massivement parallèles à fort taux d'échanges, 9 partenaires Physique + Informatique)
  - GridCL : impact de l'intégration de ressources many-core/GPU dans les infrastructures de calcul distribuées sur des applications phares de nos disciplines
  - Gestion et accès performant à de grands volumes de données dans les infrastructures distribuées, en particulier les clouds

# La Vision



# Conclusions

- La capacité de produire et traiter des grandes masses de données est critique pour nos thématiques scientifiques
  - › Le volume de données rend impossible l'externalisation
- Le LABEX P2IO fédère une grande diversité de compétences
  - › R&D, développement et exploitation de ressources
  - › Augmenter les synergies de ce **"capital humain" rare**
  - › Intéraction locale fertile entre experts techniques et chercheurs
- Le LABEX P2IO offre un cadre pour la mutualisation des infrastructures
  - › Mettre en place de nouveaux services et prendre de plus grandes responsabilités dans les projets
  - › Optimiser le coût global des infrastructures pour nos laboratoires
  - › Prendre en compte les contraintes environnementales par la mise en oeuvre de ressources énergétiquement efficaces