

# Le projet HPCDA@UGA

Pierre Neyron (LIG, CNRS)



# Convergence HPC - BigData

Beaucoup de disciplines scientifiques sont confrontées à des *flots de plus en plus massifs de données* : en volume, variété et vitesse  
→ *méthodes* et *outils* d'analyse à *réinventer*

## Géants du WEB :

- paradigmes (MapReduce...)
  - outils (Hadoop, Spark, Flink, ...)
- = *BigData*

« déplacer au maximum les traitements vers les données »

Evolution du Big Data relativement *indépendante* du HPC :

- propres infrastructures (clouds vs. supercomputers),
- applications (data analytics vs. simulation scientifique)
- logiciels (MapReduce vs. MPI/OpenMP)

Big Data de plus en plus *gourmand en calcul* (deep learning)  
Gestion des données massives → calcul scientifique

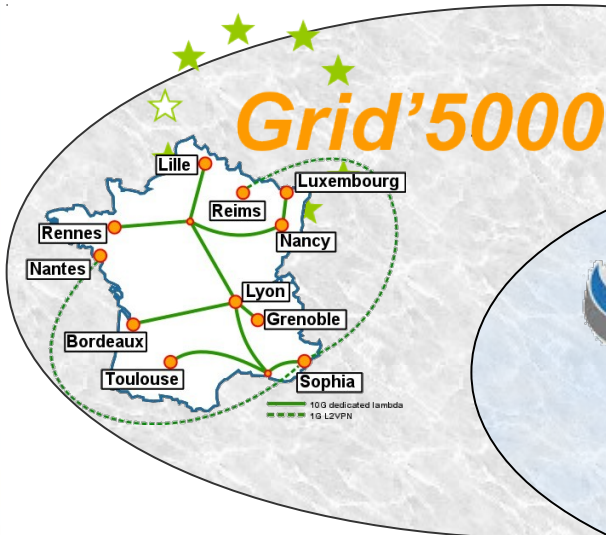
Une nécessité (cf. **BDEC**) :

- convergence des paradigmes et des outils
- expérimenter des plates-formes technologiques convergées



# HPCDA : une double convergence

EXPÉRIMENTATION



*proximité* des plates-formes



*mutualisation* des efforts et compétences



PRODUCTION

Convergence **HPC – Data Analytics (BigData)**



→ *le calcul intensif rencontre les technologies des hyperscalers*

*Mais aussi*

Convergence **Expérimentation – Production**

→ *machine commune pour les 2 communautés **Grid'5000** et **CIMENT***

# Challenge Expérimentation VS Production

		
Objectif général	Contribution à la recherche informatique «L'objectif est la méthode »	Calcul Scientifique «L'objectif est le résultat »
Communauté utilisateurs	Recherche informatique Plate-forme nationale	Toutes les disciplines scientifiques Plate-forme régionale
Domaines d'utilisation	Expérimentation HPC, Cloud, Big Data, ... Informatique distribuée au sens large	Traitement Intensif de Calculs et de Données
Particularité de l'Infrastructure	« Expérimentation » Interactivité, contrôle, reconfigurabilité	« Production » Optimisée pour la puissance de calcul, traitement par lot

***Machine commune = un beau challenge technique !***

- *Une plate-forme de production orientée vers l'expérimentation*
- *Une plate-forme expérimentale production-proof*

# HPCDA : win-win

---

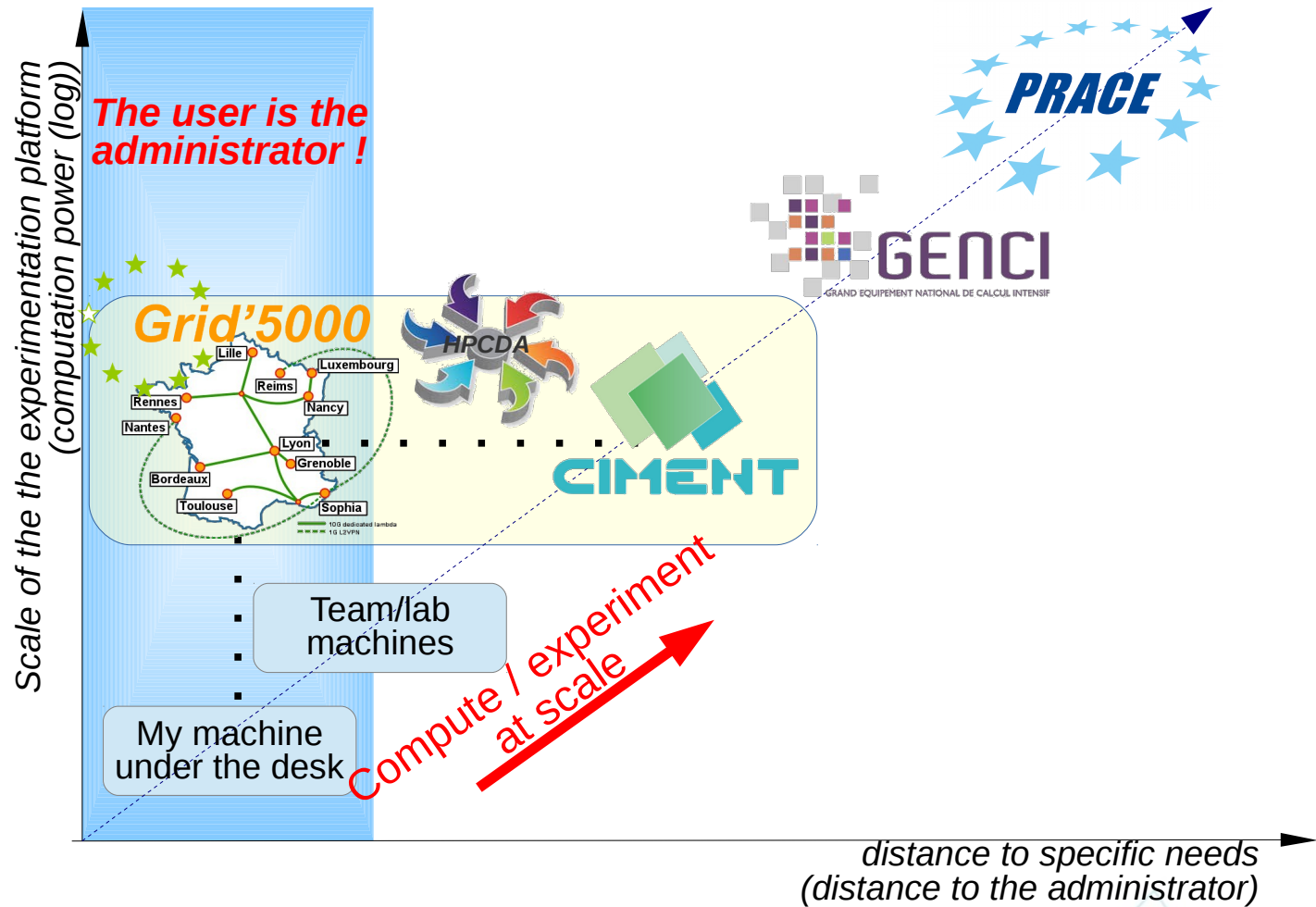
## CIMENT : Un cluster « expérimental »

- Mécanisme Grid'5000 de **déploiement système** et applicatif  
→ *s'affranchir de certaines contraintes des environnement de « production »*
- Ouverture sur les **nouvelles technologies** matérielles et **logicielles** : NVRAM, burst buffers, piles BigData  
→ *lever certains verrous des technos HPC classiques*
- Maîtrise de la plate-forme → *instrumentation/reproductibilité, monitoring énergétique, ...*

## Grid'5000 : Un compromis gagnant-gagnant

- Plate-forme d'envergure (2000 coeurs) dimensionnante → validation expérimentale
- Développement des collaborations transversales → cas d'usage réels
- Échange bi-directionnel : transfert de solutions ↔ traces de productions

# HPCDA dans l'écosystème HPC

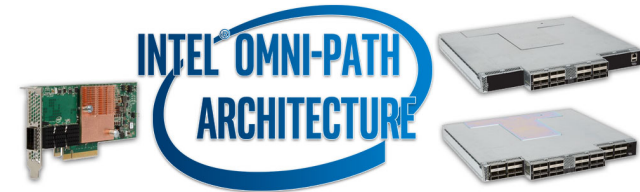




# Nouvelles technologies matérielles

Impact des *nouvelles technologies matérielles*

- *mémoire non-volatile* de type NVM Express (NVMe)
- *réseaux rapides nouvelle génération* 100 Gbps (OmniPath)
- + augmentation
  - nombre de coeurs
  - mémoire RAM
  - tailles des disques SSD et HDD



*Modification profonde de l'équilibre des grandes plateformes de traitements intensifs*

En particulier → *Burst Buffers*

- mémoire/stockage temporaire → amortir les pics de transfert vers les DFS
- sauvegarde de points de reprise (checkpointing)
- analyse in transit
- stage-in de données
- ...



SSD Burst Buffer



Permanent Storage

# La machine HPCDA

Financement CPER/Inria, Idex, GrenobleINP, Isterre : ~550K€  
Achat sur Matinfo4 → Dell, soit a priori :

## Noeuds Calcul et Données :

- 64 Noeuds Dell Poweredge C6420 : 2048 coeurs
  - Bi-cpu Xeon Gold 6130 (2.2 GHz, 2x 16 cores)
  - 192 GB RAM DDR4-2666
  - **2x SSD 446 GB + HDD 1TB**
  - **Omnipath 100Gbps** + Ethernet 10Gbps

## Noeuds Burst buffers :

- 4 Noeuds Dell poweredge R740
  - Bi-cpu Xeon Silver 4114 (2.1 GHz, 2x 10 cores)
  - 192 GB RAM DDR4-2400
  - **2 (4?) x NVMe 1.6TB HHHL PCIe 8x + HDD 4TB**
  - **Omnipath 100Gbps** + Ethernet 10Gbps

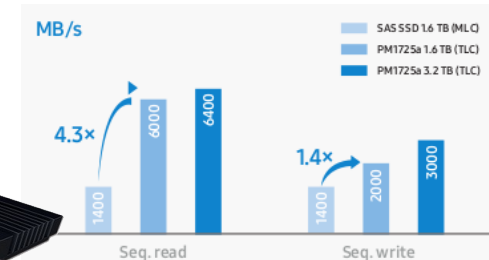
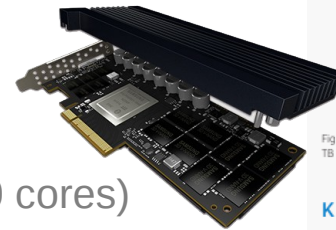


Figure 1. Sequential R/W performance comparison between an MLC SAS SSD and the 1.6 and 3.2 TB TLC PM1725a

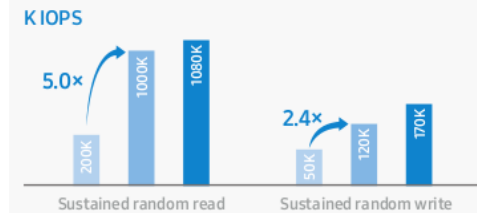


Figure 2. Sustained random R/W performance comparison between an MLC SAS SSD and the 1.6 and 3.2 TB PM1725a

Installation dans le nouveau **datacentre UGA**

Proximité des **autres équipements CIMENT** (interco ACI UGA) :

- BeeGFS, Irods, Froggy, Luke, Summer



# Support logiciel pour l'exploration

## Utilisation de la pile logicielle Grid'5000

Taillée pour le **développement** et **validation** expérimentale du passage à l'échelle des algorithmes, logiciels et systèmes  
→ **HPC, Cloud Computing, Big Data, Networking**

## Expérimentation sur toutes les couches de la pile logicielle

→ capacité assez unique pour cette taille de plateforme de **changer le système d'exploitation** (ou l'hyperviseur !)  
L'utilisateur gagne les **privileges d'administrateur** !

Configuration par l'utilisateur de **sa propre topologie du réseau** et isolation (création de VLANs, réservation de pool d'IP, routage)

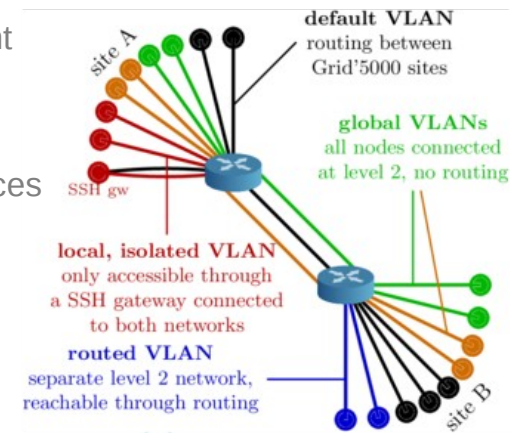
## Maitrise de l'environnement d'expérimentation ( ≠ AWS !):

- Description, vérification et traçabilité de la plateforme et de l'environnement d'expérimentation
- Monitoring: sondes réseau, énergie

**Interrogeabilité programmatique** (Rest API), instrumentation des expériences

→ **Reproductibilité**

+ couche applicative CIMENT (modules, Nix, ...)



# Plan d'action

---

## *Groupe de Travail G5K-CIMENT*

- *rédaction d'un document technique (hal-01511285)*
- *dossier Action de Développement Technologique Inria*

⇒ *Des compromis, du pragmatisme...*

## **Scénario n°1 : intégration dans l'infrastructure Grid'5000 + couche HPC CIMENT**

### **Calendrier :**

2017 :

- **Financements** : OK
- **Achat** : En cours
- **Recrutement** : En cours

2018 :

- ***Installation***

# Plan d'action

Aperçu des tâches :

- *Hébergement matériel → nouveau datacentre UGA*
- *Déploiement Grid'5000 (déménagement du site Grenoblois)*
- *Interconnexion avec les autres équipements CIMENT*
- *Installation et instrumentation des technos Omnipath et NVMe*
- *Interface de compatibilité des comptes utilisateurs CIMENT → Grid'5000*
- *Gestion des ressources et des taches : arbitrage expérimentation vs. production*
- *Portage des outils HPC CIMENT dans Grid'5000*
- *Sécurisation des services pour la production*
- *Convergence des systèmes de monitoring et de suivi des incident*
- *Instrumentation de l'infrastructure réseau SDN du datacentre*
- *Instrumentation monitoring énergétique*

*Mise en place de la collaboration entre les équipes techniques !*

*→ **Projet « pilote » pour le rapprochement de Grid'5000 et des mésocentres***

# Questions

---

Questions ?