Inauguration HPCDA

Équipement scientifique

B. Bzeznik (GRICAD)
P. Neyron (LIG/Grid'5000)

28 Novembre 2018

GRICAD
GRENOBLE ALPES RECHERCHE
INFRASTRUCTURE DE CALCUL INTENSIF ET DE DONNÉES

Le projet
HPCDA@UGA

Un projet communautaire grenoblois pour une nouvelle plateforme de calcul pour la Recherche

- Convergence **HPC - BigData**
- Convergence **"Production" / "Experimentation"**

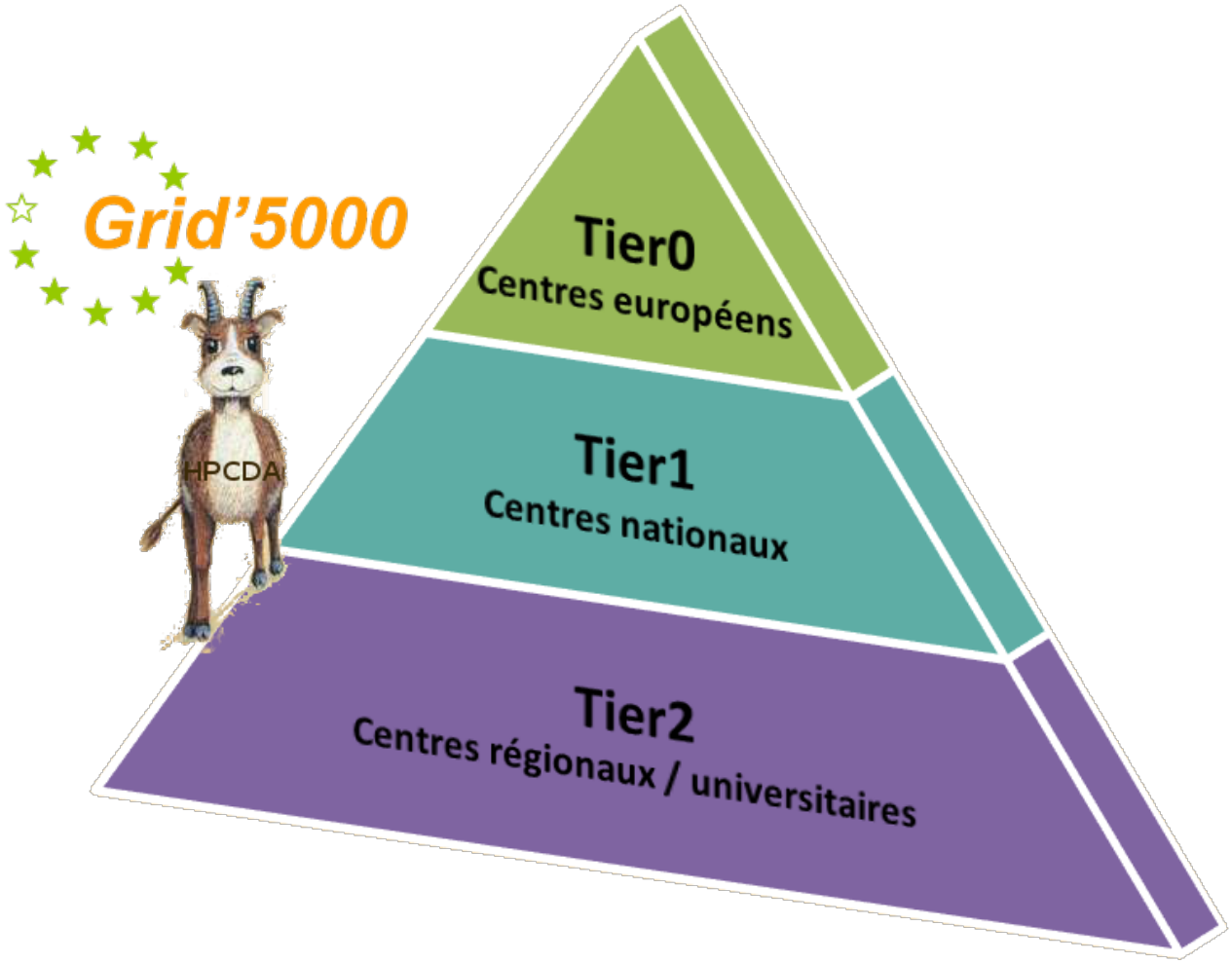
Rappel des financements obtenus

- Inria/CPER, projet LECO++ → 270 k€
- LIG, équipe ERODS → 10 k€
- AAP IDEX UGA → 100 k€
- AAP G-INP Équipement → 70 k€
- ERC OSUG/IPAG → 150 k€
- ISTERRE, projet Seiscope → 100 k€

→ **Total 700 k€**

Positionnement de l'équipement

Grid'5000 / la pyramide du calcul intensif GENCI



- Montage d'une plateforme en *équilibre* entre GENCI et Grid'5000.

Caractéristique des machines
DAHU/YETI

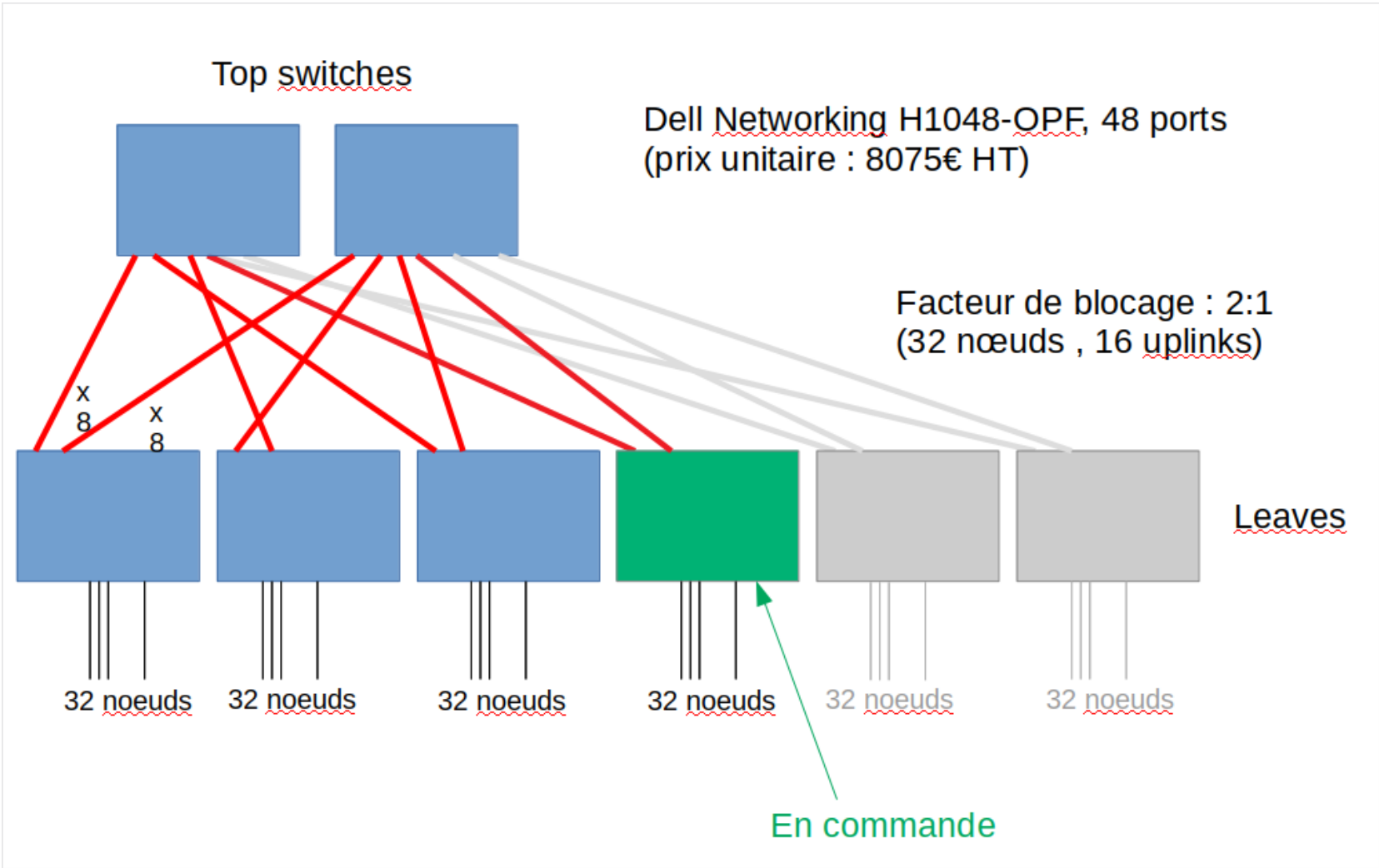
- 2560 coeurs de calcul
  - DAHU: 72 noeuds bi-Xeon Skylake Gold 6130 (16 cores, 2.1Ghz)
    - 192Go de RAM dans tous les noeuds
    - 1 x SSD de 240Go (système)
    - 1 x SSD de 480Go (scratch rapide)
    - 1 x HDD de 4To (scratch)
    - connectivités Ethernet 10Gbps et Omnipath 100Gbps
  - YETI: 4 noeuds *fat burst buffers* quad-Xeon Skylake Gold 6130
    - 768Go de RAM
    - 1 x SSD 240Go (système)
    - 2 x NVMe 1,6To (scratch ultra-rapide)
    - 3 x HDD 2To (stockage local réservable)
    - connectivités Ethernet 10Gbps et Omnipath 100Gbps
- 5 switchs Omnipath (OPA), réseau fat-tree semi-bloquant 2:1 à 100Gb/s
- 1 noeuds beegfs-data de 80 To pour extension de Bettik (scratch partagé avec la machine "Luke")
- Dispositifs wattmètres haute fréquence (50Hz)

Les machines
Points forts

- 2560 coeurs de calcul
  - DAHU: **72** noeuds bi-Xeon **Skylake** Gold 6130 (16 cores, 2.1Ghz)
    - **192Go de RAM dans tous les noeuds**
    - 1 x SSD de 240Go (système)
    - **1 x SSD de 480Go (scratch rapide)**
    - **1 x HDD de 4To** (scratch)
    - connectivités Ethernet 10Gbps et **Omnipath 100Gbps**
  - YETI: 4 noeuds *fat burst buffers* **quad**-xeon **Skylake** Gold 6130
    - **768Go de RAM**
    - 1 x SSD 240Go (système)
    - **2 x NVMe 1,6To** (scratch ultra-rapide)
    - 3 x HDD 2To (stockage local réservable)
    - connectivités Ethernet 10Gbps et **Omnipath 100Gbps**
- 5 switchs Omnipath (OPA), réseau fat-tree semi-bloquant 2:1 à 100Gbps
- 1 noeuds beegfs-data de 80 To pour extension de **Bettik (scratch partagé avec la machine "Luke")**
- Dispositif **wattmètres haute fréquence** (50Hz)

Une infrastructure réseau Omnipath évolutive

Fat-tree 2:1



- Possibilité d'évolution du nombre de ports Omnipath

	Ports utilisés	Ports libres	coût
Aujourd'hui	79[1]	96	
Fin 2018	94[2]	128	
Debut 2019	100	128	
?	?	160[3]	+8k€
?	?	192[3]	+8k€

[1] : 72 dahu + 4 yeti + 3 bigfoot
[2] : 76 dahu + 4 yeti + 3 bigfoot + 11 bettik
[3] : ajout d'1, puis 2 switchs OPA

- Une infrastructure évolutive jusqu'à 192 nœuds

Mode opératoire
CIMENT + Grid'5000

Le projet **HPCDA** :

- Convergence technologique : **HPC - BigData**
- Convergence de plateformes : **"Production" / "Experimentation"**

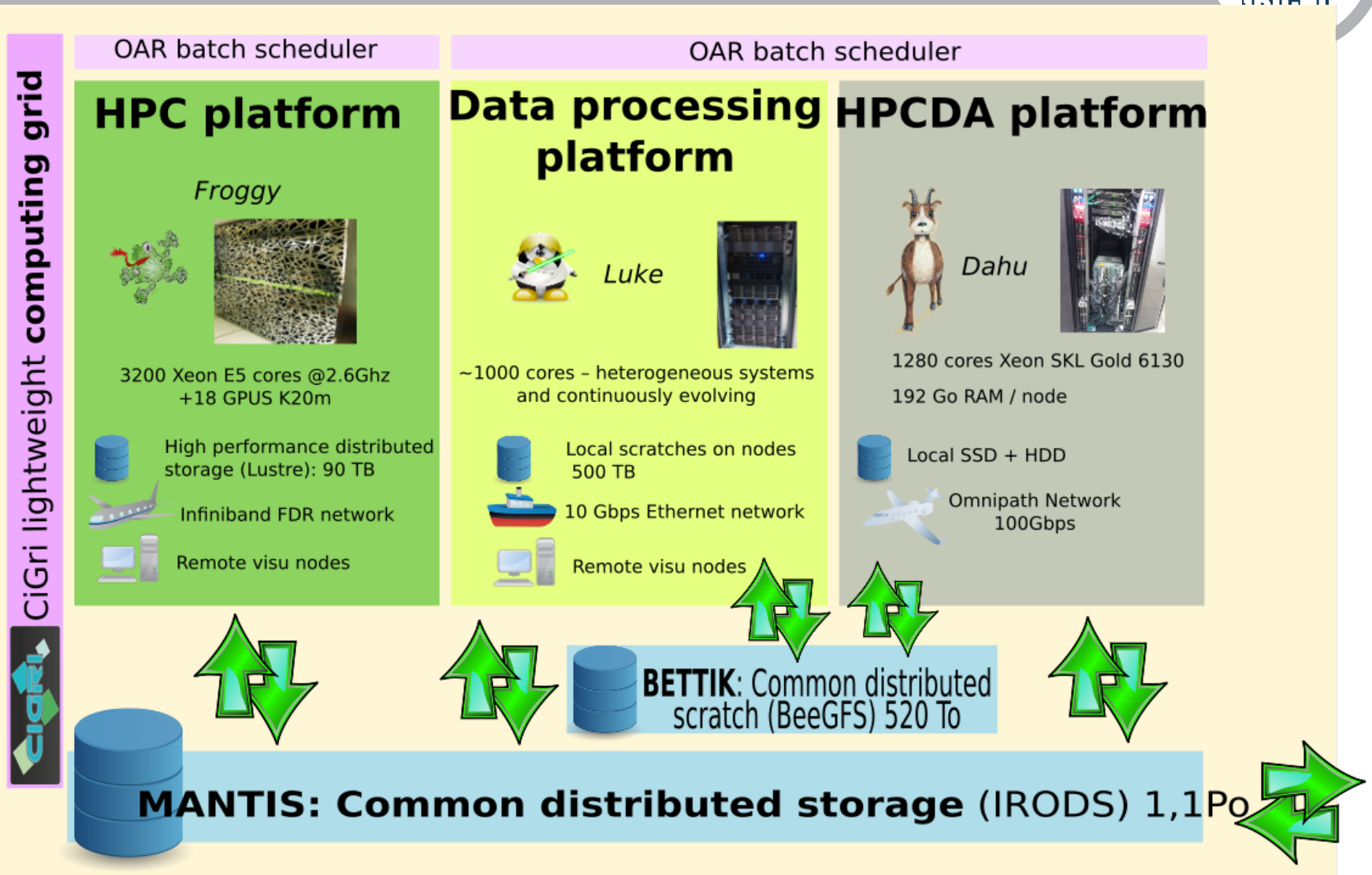
Un chantier commun alliant les équipes techniques :

- de **GRICAD** - Mésocentre CIMENT
- de l'Infrastructure de Recherche **Grid'5000**
- avec l'aide des CT Spring et datacentre

Plusieurs scénarios opératoires envisagés pour permettre ces convergences... $\rightarrow$ Un fonctionnement retenu, résultat de compromis techniques et humains :

- une partition orientée **production** opérée par **CIMENT/GRICAD**
- une partition orientée **expérimentation** opérée par **Grid'5000**
- **une passerelle entre les 2 !**

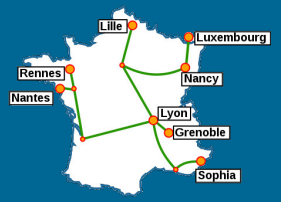
Intégration dans CIMENT
40 dahus



- 40 nœuds de Dahu intégrés dans l'infrastructure CIMENT

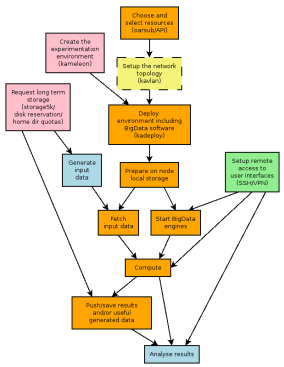
Intégration dans Grid'5000

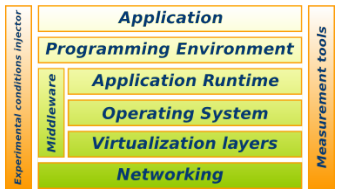
32 dahus, 4 yetis

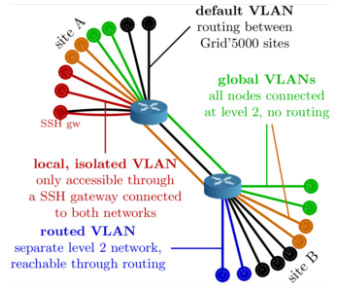


Intégration dans le **site grenoblois de Grid'5000** → Déploiment de la panoplie complète d'**instrumentation Grid'5000**

- Offre aux utilisateurs un **controle complet pour l'expérimentation**
- Une plateforme **conçue pour la Recherche en Informatique Distribuée** : *HPC, BigData, Cloud, Réseau*
  - Expérimentation possible sur toutes les couches de la pile logicielle
  - Capacité unique pour cette taille de plateforme de changer le système d'exploitation ou l'hyperviseur
  - Configuration par l'utilisateur de sa propre topologie réseau
  - Description, vérification et traçabilité de la plateforme et de l'environnement d'expérimentation
  - Accès aux sondes réseau et de consommation énergétique
  - Moteur d'instrumentation pour scripter les expériences → *reproductibilité*







- Par ses specificités matérielles *(Skylake, NVMe, Omnipath, Wattmètres)*, HPCDA apporte un **nouvel équipement scientifique majeur** à la communauté scientifique.

Passerelles et collaborations interdisciplinaire

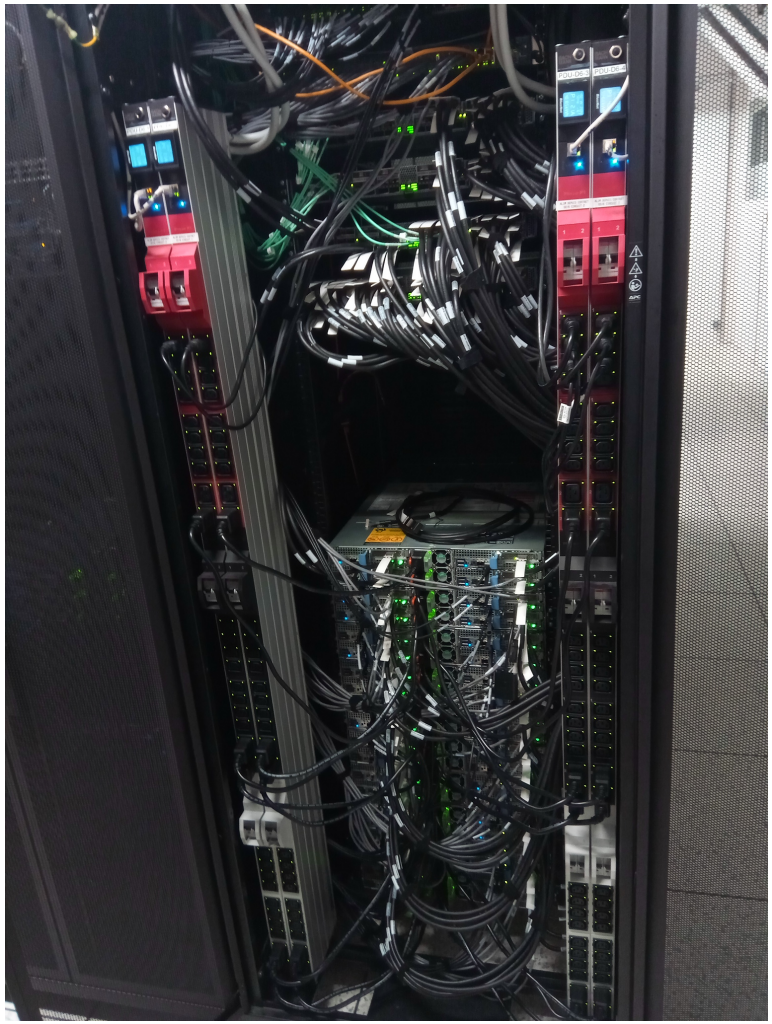
CIMENT → o ← Grid'5000

- Pour la communauté Grid'5000 : Accès à une plateforme de plus grande envergure
  - Essentiel pour la validation expérimentale à une plus grande échelle → dépasser le millier de cœurs
  - En particulier : exploiter l'intégralité du réseau Omnipath
  - Une proximité réaffirmée de la communauté du mésocentre

  COMMENT ? mécanique pour déployer ponctuellement des environements d'expérimentation utilisateur Grid'5000 sur CIMENT
- Pour la communauté CIMENT : Accès à une plateforme « expérimentale »
  - développer des nouveaux procédés en s'affranchissant des contraintes des environnements de « production »
  - lever des verrous grâce aux technos nouvelles : frameworks BigData, burst-buffers, NVRAM
  - maîtriser l'environnement de calcul : monitoring énergétique, ...

  COMMENT ? en se rapprochant des chercheurs en informatique distribuée et en devenant un utilisateur Grid'5000

En photos
Dahu@CIMENT



En photos
Dahu@G5K+Yeti

Merci !

questions? ...

IPAG
OSUG
ISTerre
LIG
LECO++
SEISCOPE
SOLIDI
Communauté UNIVERSITÉ Grenoble Alpes
La Région Auvergne-Rhône-Alpes
Investissements d'Avenir
Grenoble INP
erc
Inria
UNIVERSITÉ Grenoble Alpes
cnrs