



**HAL**  
open science

## Grid'5000 @ IPL HPC BigData

Pierre Neyron

► **To cite this version:**

| Pierre Neyron. Grid'5000 @ IPL HPC BigData. 2018. hal-01805122

**HAL Id: hal-01805122**

**<https://inria.hal.science/hal-01805122>**

Preprint submitted on 1 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Kickoff IPL HPC-BigData

Pierre Neyron (LIG, CNRS)

Grid'5000 technical team



# Grid'5000 en bref

Un instrument de recherche national pour la recherche en informatique distribuée : HPC, Cloud et BigData

- 8 sites, 30 clusters, 850 serveurs, 9000 coeurs
- réseau d'interconnexion dédié 10Gbps
- 600 utilisateurs et 100 publications par an

Expérimentation sur toutes les couches de la pile logicielle

- en particulier : capacité assez unique pour cette taille de plateforme de **changer le système d'exploitation** (ou l'hyperviseur!)

L'utilisateur gagne les **privileges d'administrateur**

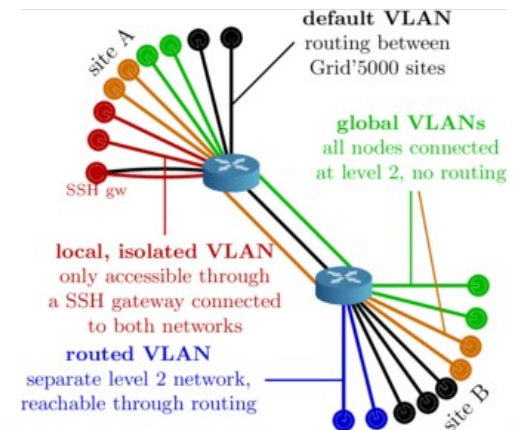
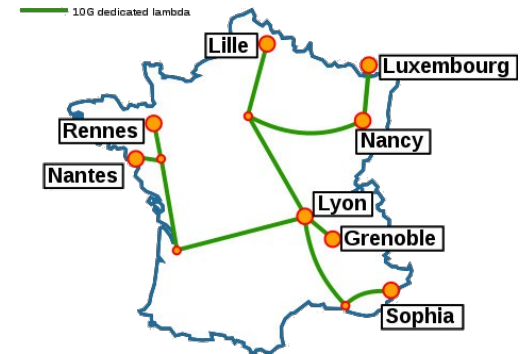
Configuration par l'utilisateur de **sa propre topologie réseau** et isolation (création de VLANs, réservation de pool d'IP, routage)

**Maîtrise 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**

**Une plateforme conçue pour la Recherche en Informatique Distribuée**



# Grid'5000 dans l'écosystème HPC

Un datacentre expérimental originaire du monde HPC/Grid

Variété de processeurs, réseau rapides, accélérateurs

→ <http://www.grid5000.fr/w/Hardware>

Reconfiguration des middleware et du système

→ runtimes, gestion de tâche et de ressources, filesystems parallèles, ...

Monitoring énergétique → **Green computing**

Outil de gestion des ressources de l'instrument de type **Batch Scheduler** (OAR)

→ compatible avec les usages de type « **production** »



# Expériences BigData

Un datacentre expérimental qui évolue pour répondre aux problématiques : Cloud, **BigData**

Matériel et services adaptés au BigData

→ variété de **stockage local sur les nœuds** (x HDD, y SSD, « in memory »)

→ **persistance des données** :

- Services de stockage longue durée (NFS, Ceph, ...)
- Réserveation des disques locaux (HDFS...)

Un support pour une **orchestration reproductible des expériences**

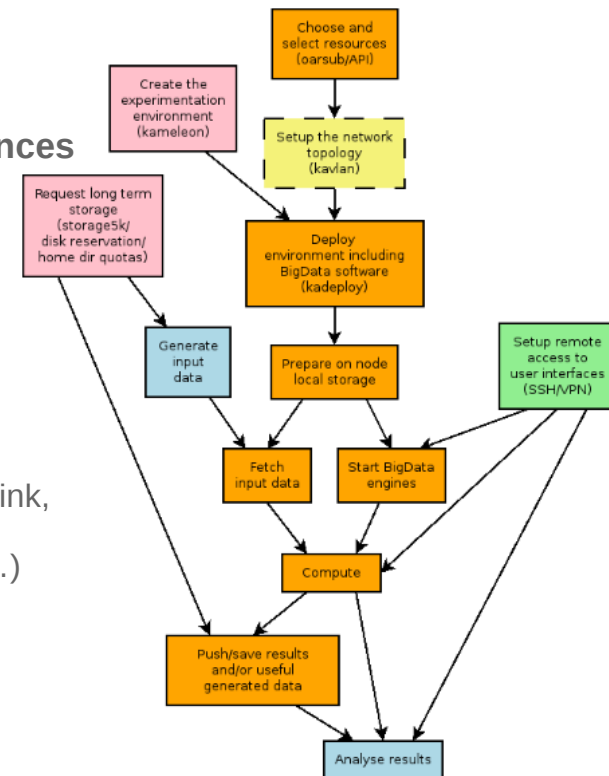
- Choix des ressources : compute / storage / network
- Déploiement de la pile logicielle (pile de services)
- Gestion des données (génération, stage-in, ...)
- Accès aux interfaces (web) des services BigData ou notebooks
- Support pour l'automatisation de l'expérience

Exemples :

- Déploiement de Hadoop (HDFS/Yarn/MR), Spark, Cassandra, Flink, ...
- Déploimenet de suites de benchmarks (YCSB, BigDataBench, ...)
- BigData in Edge computing
- Prototypage, évaluation de performance, ...

Tutoriel lors de l'école Grid5000 →

[http://www.grid5000.fr/w/BigData\\_hands-on\\_tutorial](http://www.grid5000.fr/w/BigData_hands-on_tutorial)



# Le projet HPCDA@Grenoble

Projet commun Grid'5000 / Mésocentre CIMENT(Gricad) [CPER/Inria/Idex/labos]

→ **Double convergence :**

**HPC – Data Analytics (BigData)** → *le calcul intensif rencontre les technologies des hyperscalers*

*Mais aussi*

**Expérimentation – Production** → *infrastructure commune pour les 2 communautés*

**Pour la communauté mésocentre : Un cluster « expérimental »**

- *s'affranchir de certaines contraintes des environnements de « production »*
- *lever certains verrous grâce aux nouvelles technos : frameworks BigData, burst-buffers, NVRAM*
- *maîtriser l'environnement : monitoring énergétique, ...*
- *favoriser les collaborations avec la communauté informatique*

**Pour la communauté Grid'5000 : Un compromis gagnant-gagnant**

- *plate-forme de plus grande envergure → validation expérimentale*
- *support pour l'expérimentation HPC-BigData*
- *développement des collaborations transversales → cas d'usage réels*
- *transfert bi-directionnel : technologies d'avant garde → o ← traces de production*

→ **Nouveaux clusters en cours d'installation sur le site Grid'5000 de Grenoble :**

**Dahu : Noeuds convergés HPC-BigData :**

- 72 Dell C6420 (2304 coeurs) - 2x Intel Skylake Gold 6130 - 192 GB RAM
- SSD 240GB + **SSD 446GB** + **HDD 4TB**

**Yeti : Noeuds « Fat Burst buffers » :**

- 4 Dell R940 - 4x Intel Skylake Gold 6130 (64 coeurs/noeud) - 768 GB RAM
- SSD 446GB + 2 x **NVMe 1.6TB** + 3x **HDD 2TB**

Réseau **Intel Omnipath 100Gbps** + Ethernet 10Gbps

Convergence des solutions logicielles : couche « système » Grid'5000 + couche HPC CIMENT (BeeGFS, Nix, ...)



SSD Burst Buffer



Permanent Storage



# Grid'5000 vs. Deep Learning

Comme pour les autres plateformes, l'usage « Deep Learning » est nouveau pour Grid'5000

- Comment le caractériser ?
- Quel support apporter ?

## Deep Learning → GPU ?

Un usage historique des GPUs qui se tourne effectivement vers le Deep Learning

- Traitement de la parole
- Indexing multimedia
- ...

## Des efforts en cours pour améliorer le support pour le Deep Learning sur Grid'5000

→ Un tutoriel est disponible :  
[https://www.grid5000.fr/w/User:lbada/Tuto\\_Deep\\_Learning](https://www.grid5000.fr/w/User:lbada/Tuto_Deep_Learning)

→ L'équipe technique est à l'écoute pour comprendre les besoins

**Quelle place pour Grid'5000 parmi les infrastructures pour le Deep Learning ?**  
**Quelle valeur ajoutée de Grid'5000 face aux autres offres ?**

## Coté matériel :

Quelques clusters GPUs actuellement sur Grid'5000, fortement sollicités notamment pour du Deep Learning

- **Nancy** : 28x GTX 1080Ti, 10x GTX 980, 2x Titan black, 12x K40M (**queue « production »**, financés par les équipes locales → règles d'accès spécifiques)
- **Lille** : 16x GTX 1080TI + **achat en cours** (CPER) → choix technologique difficile
- (**Lyon** : 4x M2075)

# Grid'5000 → SILECS

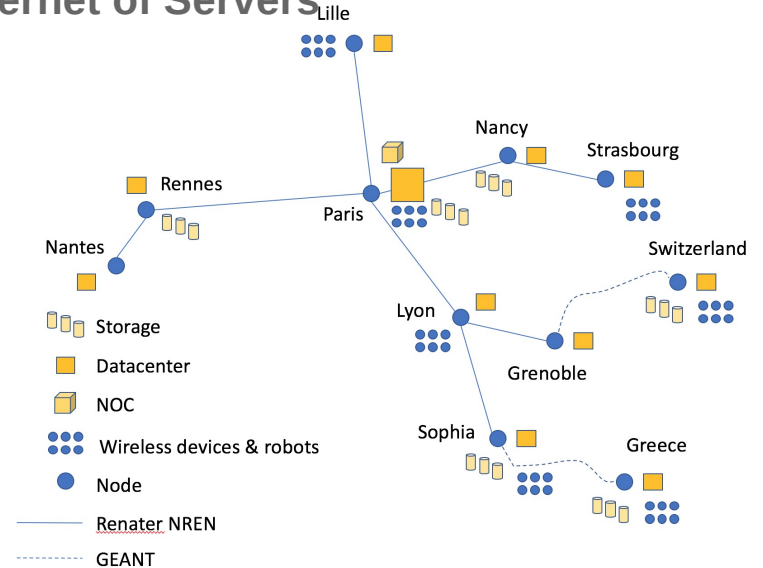
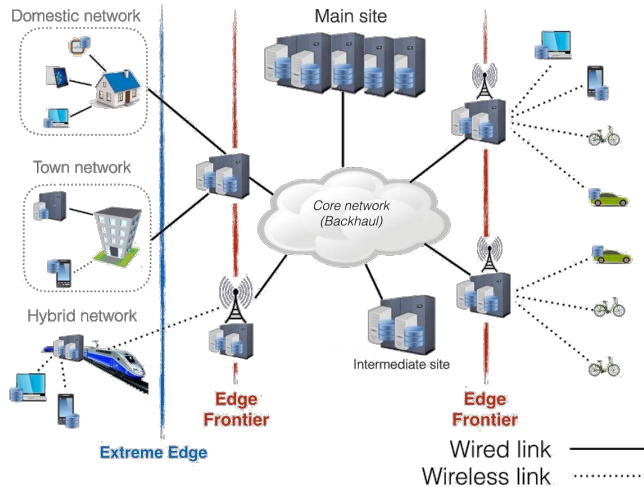
Montage en cours :

Nouvel Instrument de Recherche, produit de la réunion de **Grid'5000 + FIT** (equipex, IoT/network)



**SILECS : Super Infrastructure for Large-scale Experimental Computer Science**

→ Convergence **Internet of Things + Internet of Servers**



**SILECS** support pour un périmètre d'expérimentation élargi :

→ **Expérimenter la convergence HPC-BigData jusqu'à l'IoT (Fog Computing) ?**





# Questions

---

Questions ?



# 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*