

Retour du sondage sur les moyens de calcul

Jean Feydy
Équipe HeKA

2 juin 2022
CEP INRIA Paris

Le modèle actuel de gestion par centre (à Paris, CLEPS) a des limites :

- Difficile à défendre **politiquement** auprès de l'Etat.
- Empêche la **mutualisation** du travail de gestion.
- Freine la mise en conformité aux exigences de **sécurité** des données :
 - Partenariats industriels (NDA).
 - Données sensibles (RGPD).
 - Données de santé (HDS).

Comité dédié (Albin Petit, Lucas Nussbaum, Laurent Mirtain, Brice Goglin) avec:

- 2018 : Mise en route du projet.
- 2020 : **Choix technique** engageant l'avenir (plateforme Grid5000/SILECS)
- 2022 : Création d'un comité des **utilisateurs** des moyens de calcul.
- 2025 : Fin du déploiement ?

Sélection de 2-3 représentants des utilisateurs par centre.

À Paris : Michel Kern (HPC), Simon Legrand (SED), Jean Feydy (IA santé).

Pourquoi lancer un sondage ?

Première réunion du comité des utilisateurs le 31 mars 2022. Plusieurs points positifs :

- Volonté de faire **mieux**, pas moins cher.
- Le volet **sécurité des données** est déjà bien en route.

Mais plusieurs inquiétudes de notre part :

- Premier contact utilisateur **4 ans** après le début du projet.
- Choix techniques **contraignants** déjà entérinés : OAR et pas SLURM, etc.
- **Documentation** : nette surestimation des compétences “systèmes” des utilisateurs.
- **Stockage** : avec une bande passante de 10 Go/s par centre Inria, la gestion de jeux de données conséquents (1 To - 10 To) semble être un angle mort du projet.

Il nous est donc paru essentiel de faire remonter les besoins **du terrain** avec un sondage.

- Participation : 87 réponses complètes à Paris, **25 équipes / 31**.
- La problématique du **stockage** est indissociable de celle du calcul :
 - Volumétrie : >1Go = 70%, >100Go = 34%.
 - Données sensibles à **sécuriser** : 34%.
- Besoins très variables sur la **priorité des queues** :
les jobs **interactifs** (31%) sont aussi importants que les **calculs longs** (37%).
- Les utilisateurs sont très majoritairement **néophytes (>75%)** en gestion système.

On identifie 5 grands profils d'utilisation, qui représentent chacun **au moins 10%** des utilisateurs (en pondérant l'importance des chefs d'équipes).

D'abord, deux profils "classiques" et **bien connus** par les admin Inria, qui consomment aujourd'hui la très grande majorité des ressources sur CLEPS :

1. **HPC/HTC** : simulations au quotidien.

- Priorités : perf. **CPU, stockage, légèreté** de l'interface, support **MPI**.

2. **IA open data** : entraînements de réseaux de neurones "toutes les 48h".

- Priorités : perf. **GPU, stockage, bande passante**, support **docker/conda/python**.

Le sondage fait aussi ressortir l'importance démographique de 3 profils “moins bien dotés”, qui considèrent que le calcul est **essentiel à leurs activités** mais qui font face à d'importantes **barrières à l'utilisation** des ressources Inria :

3. **Maths** : du laptop au cluster “avant la soumission”.

- Priorités : **simplicité** pour les néophytes, **documentation** “premiers pas”.

4. **IA en santé** : entraînements et visualisation interactive de données sensibles.

- Priorités : **sécurité** des données, sessions **GPU interactives**, interface “notebook”.

5. **Système** : tests sur de multiples plate-formes.

- Priorités : contrôle **bas-niveau** de l'attribution des jobs, **documentation avancée**.

Le sondage a permis de faire ressortir :

- Le rôle capital du **stockage** et de la bande passante.
Est-il pertinent de lancer à Nancy un calcul sur des données hébergées à Paris ?
- Le besoin de **faciliter l'accès aux ressources** : de nombreux doctorants, ingénieurs ou chercheurs souhaitent utiliser les clusters Inria, mais n'y arrivent pas.
Il faut donc **étoffer la documentation** et **simplifier les démarches** administratives.
- La grande **diversité** des utilisations des moyens de calcul à Inria, au-delà du “leaderboard” des heures consommées sur CLEPS.

Le projet de mutualisation des moyens de calculs est **porteur de vraies avancées** :

- Maintenance optimisée.
- Facilitation des collaborations entre centres.
- Travail sur des données sécurisées.

Néanmoins, la **diversité** de nos recherches est difficile à accomoder.

Le calcul à Inria ne se résume pas à l'énumération des décimales de pi !

Il s'agit donc de **re-centrer le projet sur l'expérience des utilisateurs**.

Dépasser l'accumulation de "Flops" (#coeurs CPU, GPU, RAM) et mettre l'accent sur :

- Les temps d'attente et de transfert, les quotas de stockage.
- Le support logiciel.
- La **simplicité** d'utilisation.

