

GLiCID

Groupement Ligérien pour le Caclul Intensif Distribué

Yann Capdeville, A. Bihouée, A. Blondel, P. Bondia-Luttiau, J Coatanea, L. Cournede,
H. Dignonnet, Y. Dupont, G. Dupuy, J.C. Feufeu, P.E. Guérin, J.F. Guillaume, S.
Meigner, G. Moebis, M. Molière, F. Saubion, L. Silva

U. d'Angers, Centrale Nantes, Le Mans U., Nantes U.

10-12 octobre 2022



Plan

Contexte de la création de GLiCID

Les origines de GLiCID

DaCaS

Le projet GLiCID

Contexte – les établissements MESR de La région Pays de Loire (PdL)

- ▶ Nantes Université (NU)
43000 étudiants, 2100 EC
- ▶ Université d'Angers (UA)
25000 étudiants, 1100 EC
- ▶ Le Mans Université (LMU)
11000 étudiants, 630 EC
- ▶ École Centrale de Nantes (ECN)
2400 étudiants, 550 EC
- ▶ Université de Laval, La Roche-Sur-Yon,
Université Gustave Eiffel, ENSAM,
IFREMER, INRAe, ESEO ...



Contexte – les ressources HPC en PdL avant GLiCID

mesocentres

- ▶ CCIPL (NU)
- ▶ BiRD (INSERM, NU, Biologie)
- ▶ ICI (ECN)
- ▶ MathStic (UA)
- ▶ INFRALAB (LMU)

Stockage

- ▶ CEPH recherche (NU):
3Po de stockage
CEPH mutualisé
CCIPL/BiRD/ 3 labos



	# utilisateurs actifs	cœurs	GPU	loc
CCIPL	155	6500	27	DC NU (130kW)
BiRD	142	450	9	DC NU
ICI	130	6400	24	salle ECN (80kW)
UA	30	1000	10	salle labo (40kW)
LMU	50	1900	90	salle labo (30kW)

Contexte – les ressources HPC en PdL avant GLiCID

Les modèles économiques

- ▶ CCIPL (NU): accès sur projet. Fonds CPER et mutualisation d'équipements (ANR, Région, Europe ...)
- ▶ BiRD (NU): accès sur projet (public) / payant (privé); stockage supp payant
- ▶ ICI (ECN): facturé à l'heure (public et privé), prise en charge par l'établissement des projets non financés; fonds CPER, ECN ...
- ▶ MathStic (UA): mutualisation d'équipements (Région, ANR ...)
- ▶ INFRALAB (LMU): mutualisation d'équipements (Région, ANR ...)

Les publics

de l'ordre de 450 utilisateurs actifs cumulés (Univ, écoles, CNRS, INSERM, CNAP, ...); tous les domaines, les pratiques et le niveau en HPC très hétérogènes

Les équipes

effectifs cumulés (Univ, écoles, CNRS), (statutaires, CDI, CDD), ingénieurs + direction: 7.5 ETP sur 16 personnes (8 IE, 4 IR, 4 PR/DR; dont 4 CNRS; 10 statutaires)

Contexte – les ressources HPC en PdL avant GLiCID

Les matériel et logiciel actuellement

Très varié:

- ▶ Réseau : IB, OPA ... et **RoCE** (Fabric EVPN/VxLAN, switches «**whitebox**» +**SONIC**)
- ▶ Stockage : GPFS, BeeGFS et CEPH (HDD répartis sur 3 salles, NVMe)
- ▶ CPU + GPU : Intel et AMD, P100,T4,A40,A100...
- ▶ Logiciels : Modules et GUIX (**GUIX system** sur une partie de l'infra), SLURM ...



Origine de GLiCID

Différents points nous ont collectivement amener à agir:

- ▶ 2016: présentation du ministère aux Journées des mésocentres: projets et objectifs de labellisation des datacenters uniques / région
- ▶ volonté des financeurs de rationaliser les financements pour le CPER 2021-2027
- ▶ nécessité répondre à toujours plus de demandes (heures.cœurs, GPU, stockage, services) à moyen constant;
- ▶ notre volonté d'aller vers plus de mutualisation

Nous avons proposé le projet GLiCID, partie du projet plus global DaCaS (**D**atacenter et **C**alcul Scientifique + réseau) avec pour principaux objectifs:

- ▶ n'avoir qu'un seul acteur calcul Tier2 en Région PdL
- ▶ avoir plus de ressources et de services pour un coût équivalent
- ▶ Mutualiser les ressources et les RH

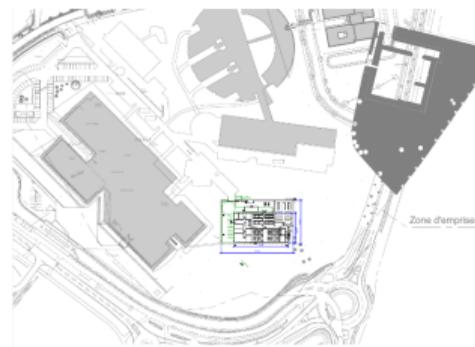
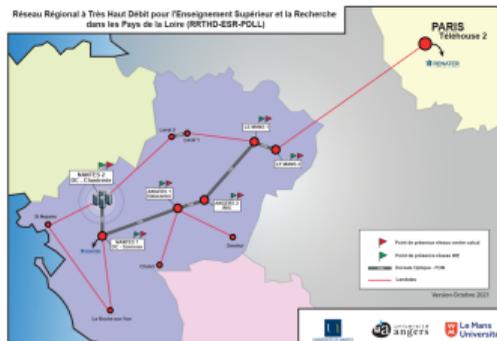
Le projet DaCaS et les services qui en découlent

Projet DaCaS + RRTHD

Projet CPER 2021-2027 avec 3 axes:

- ▶ construire un nouveau datacenter régional (10M€)
- ▶ équiper GLiCID en CPU, GPU et stockage (6M€) (et héberger tout ça dans DaCaS)
- ▶ connecter les différents sites à DaCaS avec un bon réseau (4M€)

Le projet a été accepté et financé. La partie réseau est bien avancée et le DC DaCaS est prévu pour 2025.



Le projet DaCaS et les services qui en découlent

Les services DaCaS et GLiCID

Il a été choisi d'avoir deux entités distinctes pour gérer DaCaS et GLiCID:

- ▶ la partie DaCaS + réseau est gérée par un “service inter-établissement” le SIEN DaCaS, avec pour tutelles les 3U des PdL. Direction: Stéphane Amiard (UA);
- ▶ pour la partie calcul, GLiCID est une unité de Service régionale avec pour tutelle les 3U + l'ECN. Une tutelle de l'INSERM est en discussion et le CNRS ne s'est actuellement pas positionné

Le datacenter actuel NU et le futur DaCaS ont obtenu le label “datacenter régional” le

11/12/2020

Le projet GLiCID

Missions de GLiCID

- ▶ Promouvoir le calcul scientifique ;
- ▶ Assurer l'exploitation, la maintenance, l'évolution des moyens et des équipements
- ▶ veiller à la mutualisation des équipements
- ▶ proposer un support **de proximité** au calcul et aux données, assurer la formation des utilisateurs
- ▶ accompagner les utilisateurs dans leur développement
- ▶ assurer une veille technologique
- ▶ participer aux actions proposées par les infrastructures nationales et internationales autour du calcul.

Le projet GLiCID

Modèle économique de GLiCID

- ▶ Accès aux heures de calcul sur projet (simple) pour les ayants droits
- ▶ Accès payant pour les acteurs privés et les extérieurs
- ▶ possibilité d'acheter des heures de calcul prioritaires (max 40% des heures dispo)
- ▶ Stockage gratuit pour le quota de base (3To), payant ensuite (pour les gros volumes)
- ▶ Partage des coûts de fonctionnement entre partenaires
- ▶ Mutualisation des ingénieurs des partenaires

L'objectif de la facturation est de financer les postes CDI de GLiCID

Pour les heures prioritaires et le stockage, il y a 4 tarifs:

- ▶ ayants droits et leurs collaborateurs mutualisant des RH
- ▶ ayants droits et leurs collaborateurs
- ▶ extérieurs académiques
- ▶ extérieurs privés

GLiCID: équipes

Équipe technique

- ▶ Aymeric Blondel (IGE, CNRS)
- ▶ Pablo Bondia-Luttiau (IGR, ECN)
- ▶ Jerome Coatanea (IGE, NU)
- ▶ Laurent Cournede (IGE, CNRS)
- ▶ Yann Dupont (IGR, NU)
- ▶ Grégor Dupuy (IGE, LMU)
- ▶ Jean-Christian Feufeu (IGE, NU)
- ▶ Pierre-Emmanuel Guérin (IGR, ECN)
- ▶ Jean-Francois Guillaume (IGE, BiRD, NU)
- ▶ Guy Moebs (IGR, CNRS)
- ▶ Mélanie Molière (Gestionnaire, ECN)
- ▶ Poste MesoNET vacant (IGR, ECN)
- ▶ Poste UA vacant (IGR, UA)

Équipe de direction

Direction adjointe:

- ▶ Audrey Bihouée (NU)
- ▶ Sylvain Meignier (LMU)
- ▶ Luisa Rocha-Da-Silva (ECN)
- ▶ Frédérique Saubion (UA)

Direction :

Yann Capdeville (CNRS-NU)

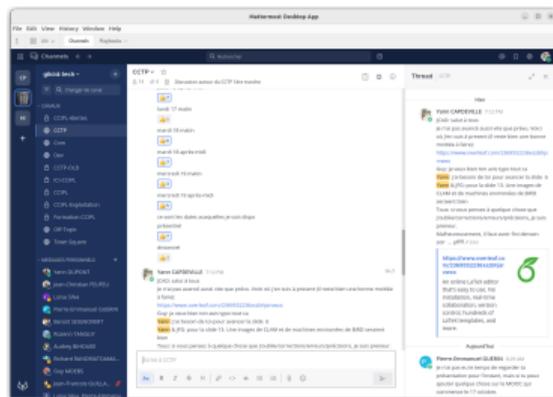
Difficultés/avantages par construction

- ▶ équipe multi-sites
- ▶ des cultures différentes
- ▶ beaucoup d'expérience et de compétences disponibles

GLiCID: équipes

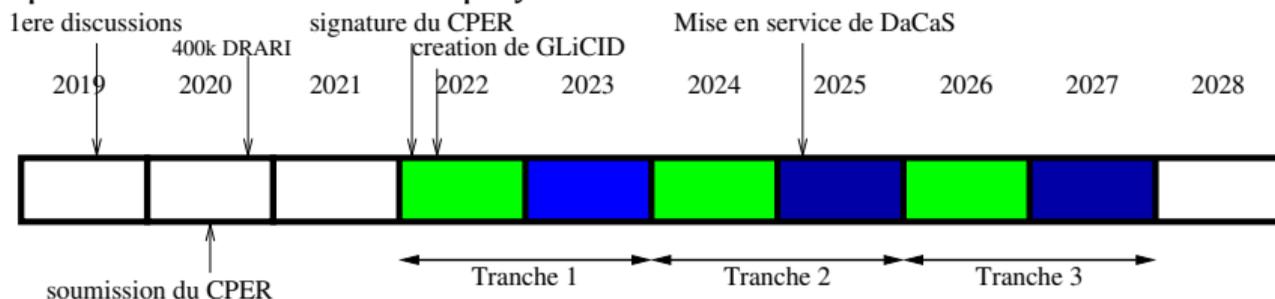
Une équipe distribuée sur 4 sites. Méthode de travail:

- ▶ Une réunion zoom par semaine
- ▶ Une réunion physique quand on peut (idéalement une fois tous les deux mois)
- ▶ des échanges permanents sur mattermost



Le projet GLiCID: les budgets

- ▶ équipements (doivent être installés dans le DC DaCaS):
 - ▶ CPER21-27: 6M€ répartis en 3 tranches
 - ▶ les équipements des PIA3 MesoNET (porté par l'ECN) et MuDi4LS (porté par l'NU) seront mutualisés dans GLiCID.
- ▶ fonctionnement:
 - ▶ récurrent: 45k€/an venant des 4 tutelles
 - ▶ fluide et postes statutaires pris en charge par les tutelles
 - ▶ postes CDI: facturation et projets



L'AO de la 1ere tranche (2M€) est terminé (lauréat: BULL-ATOS). Difficultés: tranche avancée d'un an par l'État, délais courts et stricts (50% du budget vient de REACT-EU), crise des composants ... installation temporaire salle ECN avant DaCaS

Le projet GLiCID

Les solutions en cours de mise en place

- ▶ fédération d'identité et création de compte en un clic (CLAM)
- ▶ infrastructure HA adaptée à l'implantation multi-site, administrable collégialement
- ▶ machines virtuelles à la demande (OpenStack, à partir de la solution BiRD)
- ▶ jupyter-hub...
- ▶ augmentation volumétrie CEPH, volumes SSD exclusifs à la volée, S3...

```

--- RAW STORAGE ---
CLASS  SIZE      AVAIL      USED      RAW USED  %RAW USED
hdd    3.9 PiB    2.1 PiB    1.8 PiB    1.8 PiB    46.67
ssd    43 TiB    24 TiB    19 TiB    19 TiB    43.51
TOTAL  3.9 PiB    2.1 PiB    1.8 PiB    1.8 PiB    46.64
  
```

La formation des utilisateurs

Des formations utilisateurs sont proposées par les différents établissements (en cours de fusion), mais aussi plus largement avec le MOOC "Dopez vos calculs" proposé par l'ECN sur la plateforme FUN MOOC :

https://cutt.ly/cn-mooc-dopez_vos_calculs-v2



Nous avons actuellement deux postes à pourvoir:

- ▶ un poste IGR "Responsable de la cellule calcul scientifique" à Angers (CDD 1 an puis CDI)
- ▶ un poste IGR "Ingénieur.e système et support HPC / IA / technologies innovantes" MesoNET à l'ECN (CDD puis CDI)

FIN