

XIII^E FORUM CENTRE DE CALCUL DU CNES

2023
Toulouse



Sommaire

Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

Les nouveaux services HPC6G :

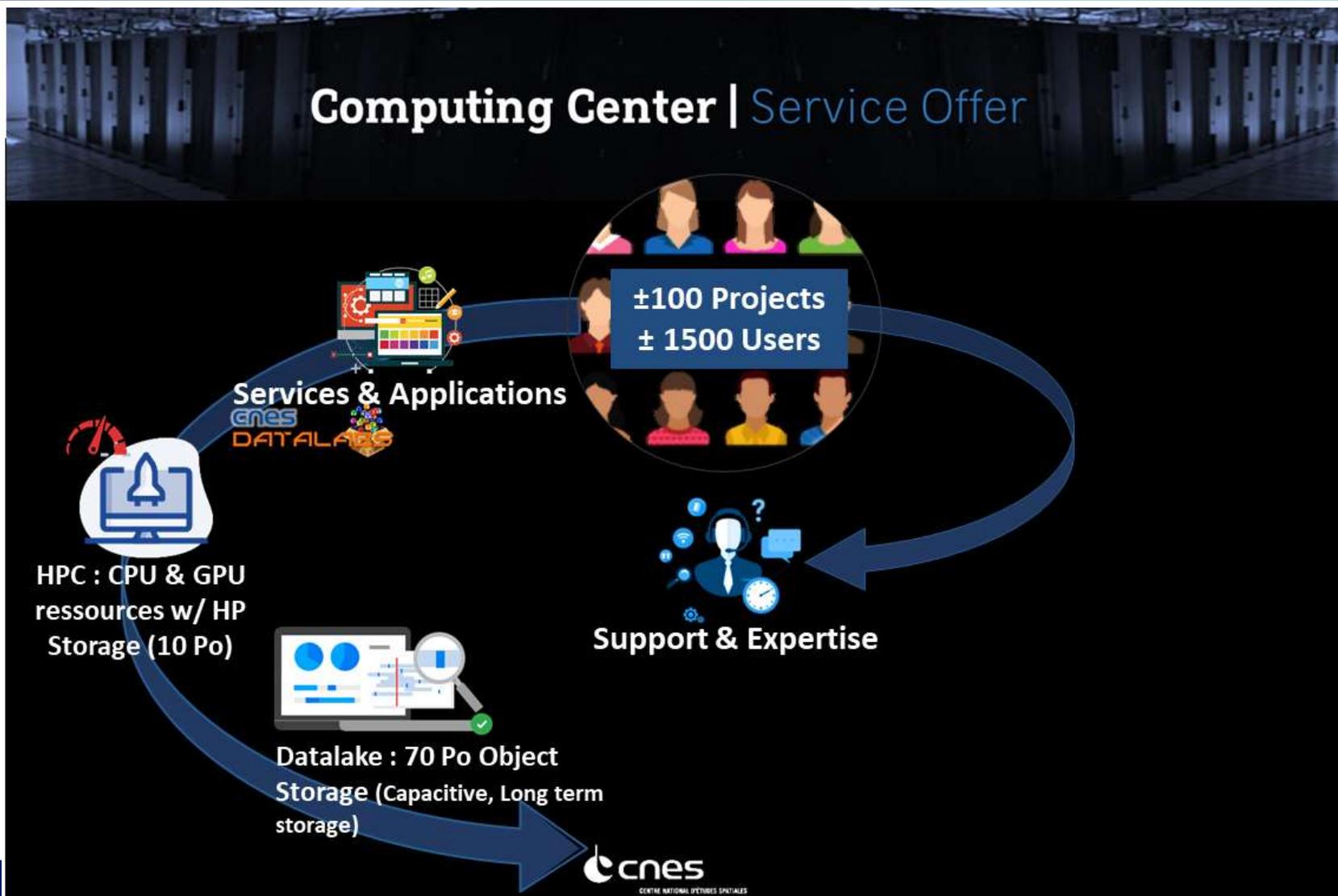
- Kubernetes
- Monitoring
- Energie



CENTRE DE CALCUL
DU **CNES**

Offre de service de l'équipe fondation Calcul

Service cohérent orienté utilisateurs, intégré au sein du SIS et SIDR



+ Info sur :



Offre de service de l'équipe fondation Calcul

Service cohérent orienté utilisateurs, intégré au sein du SIS et SIDR

- ❑ **Mise à disposition d'un environnement de calcul distribué de type High Performance Computing basé sur Linux (HPC) :**
 - ✓ Grande capacité de calcul via une ferme de serveurs (12000 cœurs disponibles sur HAL).
 - ✓ Espace de stockage haute performance POSIX type Spectrum Scale (ex GPFS) : 12 Po partagés, avec une bande passante de 200 Go/s).
 - ✓ Des nœuds mutualisés pour les développements et l'interfaçage avec le cluster.
 - ✓ Ouverture de la plateforme depuis l'extérieur du CNES.

- ❑ **Mise à disposition d'un stockage capacitif objet permettant la diffusion, visualisation et valorisation des données (DATALAKE) :**
 - ✓ Ce stockage capacitif Objet (standard AWS) 70 Po possède un lien privilégié avec la plateforme de calcul.

- ❑ **Mise à disposition d'un écosystème logiciel adapté au calcul scientifique propre à la plateforme (SIS) :**
 - ✓ plus de 200 produits (Compilateurs Intel/gcc, bibliothèques scientifiques, éditeurs, logiciels comme QGis, ENVI ...)

- ❑ **Mise à dispositions d'accès simplifiés et d'environnements de travail virtuels au travers de son offre Datalabs, Sessions interactives, etc. :**
 - ✓ Cet espace de travail Web offre un accès transparent à tous les composants/services dont un utilisateur a besoin pour accéder à ses données, les traiter et collaborer/échanger avec sa communauté.



Accompagnement des projets spatiaux



Support et expertise HPC



WikiHPC



Offres de service

ARCHITECTES SEGMENT SOL



Ma Vie Numérique

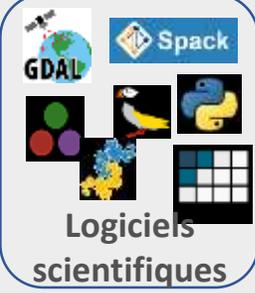
Conteneurs



Métrologie



Amalthée



Logiciels scientifiques

Cloud



Services et applications

SERVICES ET OUTILS MISSIONS

- Catalogue de données (Regards)
- Orchestrateur (Spirit)
- Archivage pérenne (STAF)

CELLULE IA

- Support projets
- Usine IA

OUTILS TRANSVERSES

- Usine logiciel (intégration continue)
- Annuaire centralisé
- Monitoring (Centreon)
- Base de données mutualisées

Infrastructures

SYSTEMES ET SERVICES D'INFRASTRUCTURE

- Virtualisation (VM projets)
- Stockage et sauvegarde NAS

RESEAU ET CONNECTIVITE

- communication entre les composants du SI Scientifique
- accès externe au SI-S



Calcul



Stockage

Rappel des objectifs du Centre de Calcul

- ✓ **Augmentation en puissance et en performance au plus juste**
 - Traitements plus complexes et plus nombreux,
 - Nombre d'utilisateurs en augmentation continue,
 - Augmentation du volume de données et des débits réseaux
- ✓ **Monitoring** pour le suivi des ressources **accessible pour les projets et les utilisateurs**
- ✓ **Agilité et scalabilité** des plateformes d'un point de vue matériel et dimensionnement
- ✓ **Inter-opérabilité : Compatibilité avec les plateformes partenaires (centre nationaux) et les plateformes de type Cloud**
- ✓ **SLA clairement affichés** et attention portée sur des **architectures évitant les SPOF**
- ✓ **Maîtrise de la consommation énergétique**
 - Consommation des plateformes
 - Performances des codes
 - Bonnes pratiques pour le stockage

Sommaire

Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie

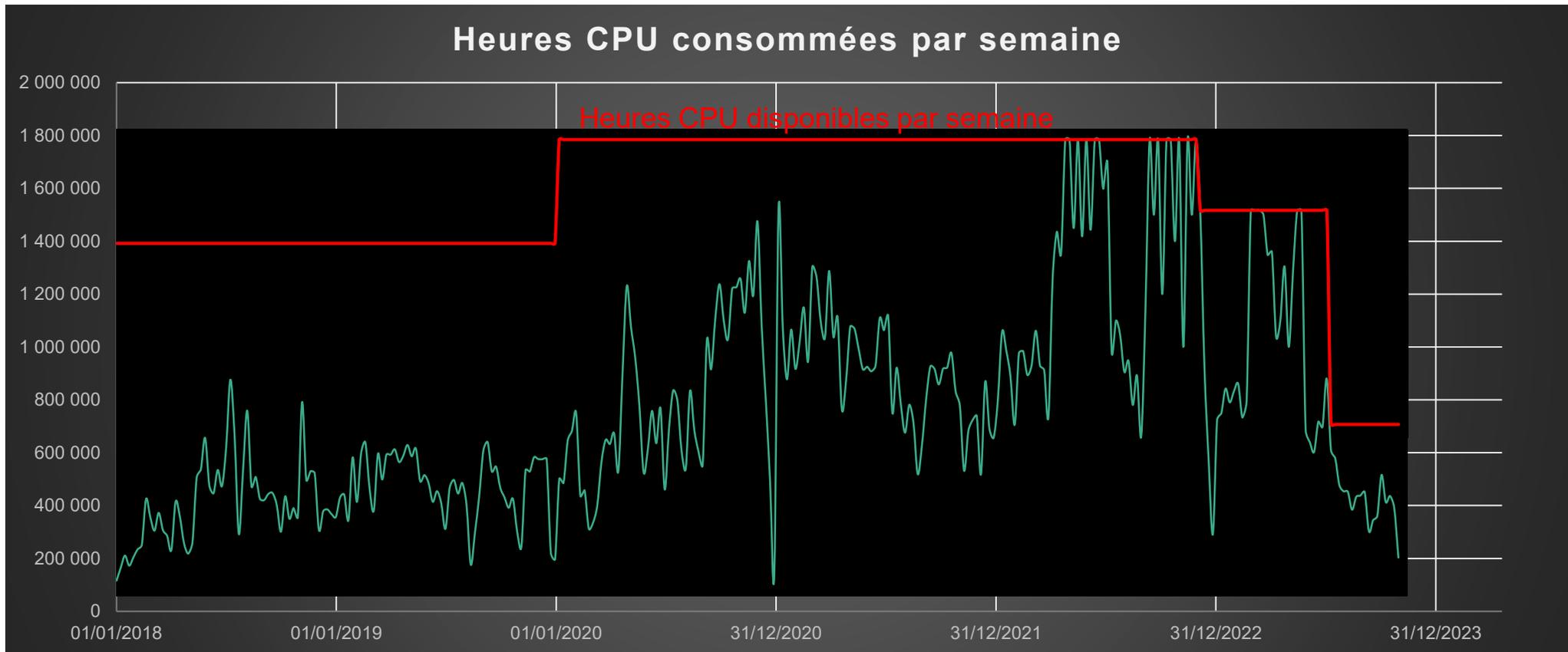


CENTRE DE CALCUL
DU **CNES**

HAL : Utilisation de la plateforme | Vue heures de calculs sur l'année

57 557 301

Heures CPU consommées sur 12 mois 1^{er} nov 2022 – 31 oct 2023



REX : Incidents HAL et arrêts de prod. sur l'année 2023

Janvier		Février		Mars		Avril		Mai		Juin		Juillet		Août		Septembre		Octobre		Novembre		Décembre	
1 D		1 M		1 M		1 S		1 L		1 J		1 S		1 M		1 V		1 D		1 M		1 V	
2 L		2 J		2 J		2 D		2 M		2 V		2 D		2 M		2 S		2 L		2 J		2 S	
3 M		3 V		3 V		3 L		3 M		3 S		3 L		3 J	1 J	3 D		3 M		3 V		3 D	
4 M		4 S		4 S		4 M		4 J		4 D		4 M		4 V		4 L		4 M		4 S		4 L	
5 J		5 D		5 D		5 M		5 V		5 L		5 M		5 S		5 M		5 J		5 D		5 M	
6 V		6 L		6 L		6 J		6 S		6 M		6 J		6 D		6 M		6 V		6 L		6 M	
7 S		7 M		7 M		7 V		7 D		7 M		7 V		7 L		7 J		7 S		7 M		7 J	
8 D		8 M		8 M		8 S		8 L		8 J		8 S		8 M		8 V		8 D		8 M		8 V	
9 L		9 J		9 J		9 D		9 M		9 V	3 H	9 D		9 M		9 S		9 L		9 J		9 S	
10 M		10 V		10 V		10 L		10 M		10 S		10 L		10 J		10 D		10 M		10 V		10 D	
11 M		11 S		11 S		11 M		11 J															
12 J		12 D		12 D		12 M		12 V															
13 V		13 L		13 L		13 J		13 S															
14 S		14 M		14 M		14 V		14 D															
15 D		15 M		15 M		15 S		15 L															
16 L		16 J		16 J		16 D		16 M															
17 M		17 V		17 V		17 L		17 M															
18 M		18 S		18 S		18 M		18 J															
19 J		19 D		19 D		19 M		19 V															
20 V		20 L		20 L		20 J		20 S															
21 S		21 M		21 M		21 V		21 D															
22 D		22 M		22 M		22 S	3 J	22 L															
23 L		23 J		23 J		23 D		23 M		23 V		23 D		23 M		23 S		23 L		23 J		23 S	
24 M		24 V		24 V		24 L		24 M		24 S		24 L		24 J		24 D		24 M		24 V		24 D	
25 M		25 S		25 S		25 M		25 J		25 D		25 M		25 V		25 L		25 M		25 S		25 L	
26 J		26 D		26 D		26 M		26 V		26 L		26 M		26 S		26 M		26 J		26 D		26 M	
27 V		27 L		27 L		27 J		27 S		27 M		27 J		27 D		27 M		27 V		27 L		27 M	
28 S		28 M		28 M		28 V		28 D		28 M		28 V		28 L		28 J		28 S		28 M		28 V	
29 D				29 M		29 S		29 L		29 J		29 S		29 M		29 V		29 D		29 M		29 V	
30 L				30 J		30 D		30 M		30 V		30 D		30 M		30 S		30 L		30 J		30 S	
31 M				31 V				31 M	3 H			31 L		31 J				31 M					

- 0 arrêt de production pour maintenance (plateforme en fin de vie)
- 4 Incidents



Plateforme opérationnelle **99%*** sur 2023

* Sur une base de 365 jours comptant 4 jours d'indisponibilité

- Incidents (sans AP)
- Incidents (avec AP)
- Arrêts de production prévus (avec AP)

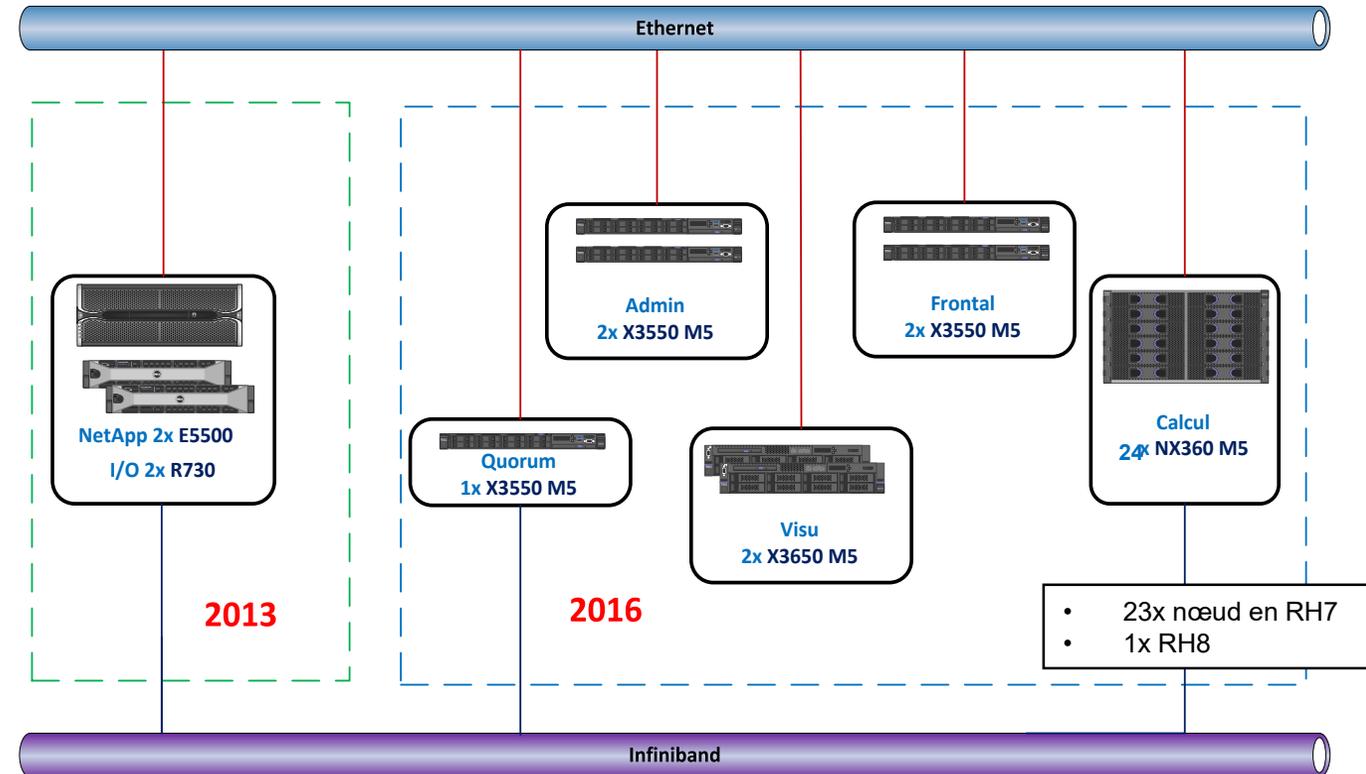
Focus sur la plateforme KTULU (DRSF)

822 875

Heures CPU consommées sur 12 mois
1^{er} nov 2022 – 31 oct 2023

Espaces de travail work/ :

- **Utilisation actuelle** :
85 TB/137.5 TB **(62%)**
- **Total quotas attribués** : 150 To /
137.5To **(110%)**



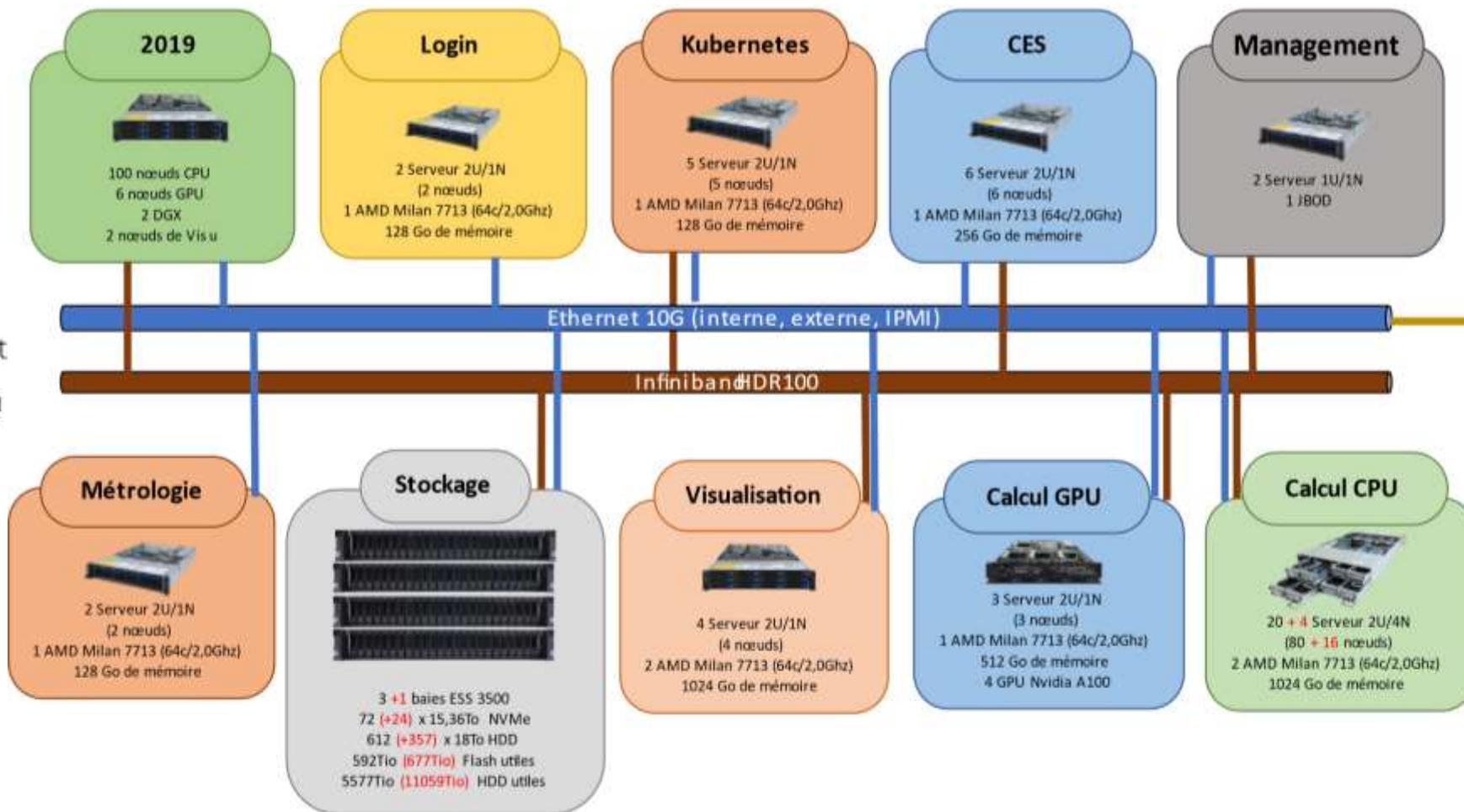
Aucun incident ni arrêt de production pour maintenance sur l'année 2023

Planning HPC6G

- ❑ Mise en production plateforme diffusion limitée (TREX) le 15/09/2023
- ❑ Migration de cœur de réseau de la plateforme SWOT entre le 16/10/2023 et le 19/10/2023
- ❑ Réintégration du matériel g2019 de HAL dans TREX, finalisation prévue fin novembre 2023
 - 100% des nœuds CPU ont été réintégrés dans TREX
 - Migration des cartes GPU en cours
- ❑ Plateforme DR-SF (ARYX) : début de VSR prévue le 15/12/2023
- ❑ Plateforme de métrologie ouverte aux utilisateurs
 - La plateforme est déjà accessible, des améliorations sont en cours
- ❑ Plateforme Kubernetes orientée traitement des données : début de VSR prévue le 15/12/2023
 - POC en cours avec le Campus de la Donnée sur OpenEO
- ❑ Arrêt de HAL prévu le 22/12/2023



P/F Diffusion limitée 6G : TREX (vue fonctionnelle)



Cœur de réseau CNES 100G

12 288 CPU : Processeurs AMD
(128 Cœurs par serveurs, 1TiB de RAM, 3GHz par core)

4000 CPU : Processeurs Intel (40 Cœurs par serveur et 187 GB de RAM)

12 GPU : NVIDIA A100 80Go Pci-e

12 PiB de stockage POSIX : Système de fichier // et distribué **IBM Spectrum Scale** basées sur baies intelligentes **ESS-3500**

Réseau Faible latence **Mellanox Infiniband HDR-100 : 100Gb**

Réseau **Ethernet 10G**

Orchestration Conteneurs : **Kubernetes**

Système d'exploitation : **RedHat 8.X et RedHat 7,9**

Orchestrateur Jobs HPC : **SLURM**

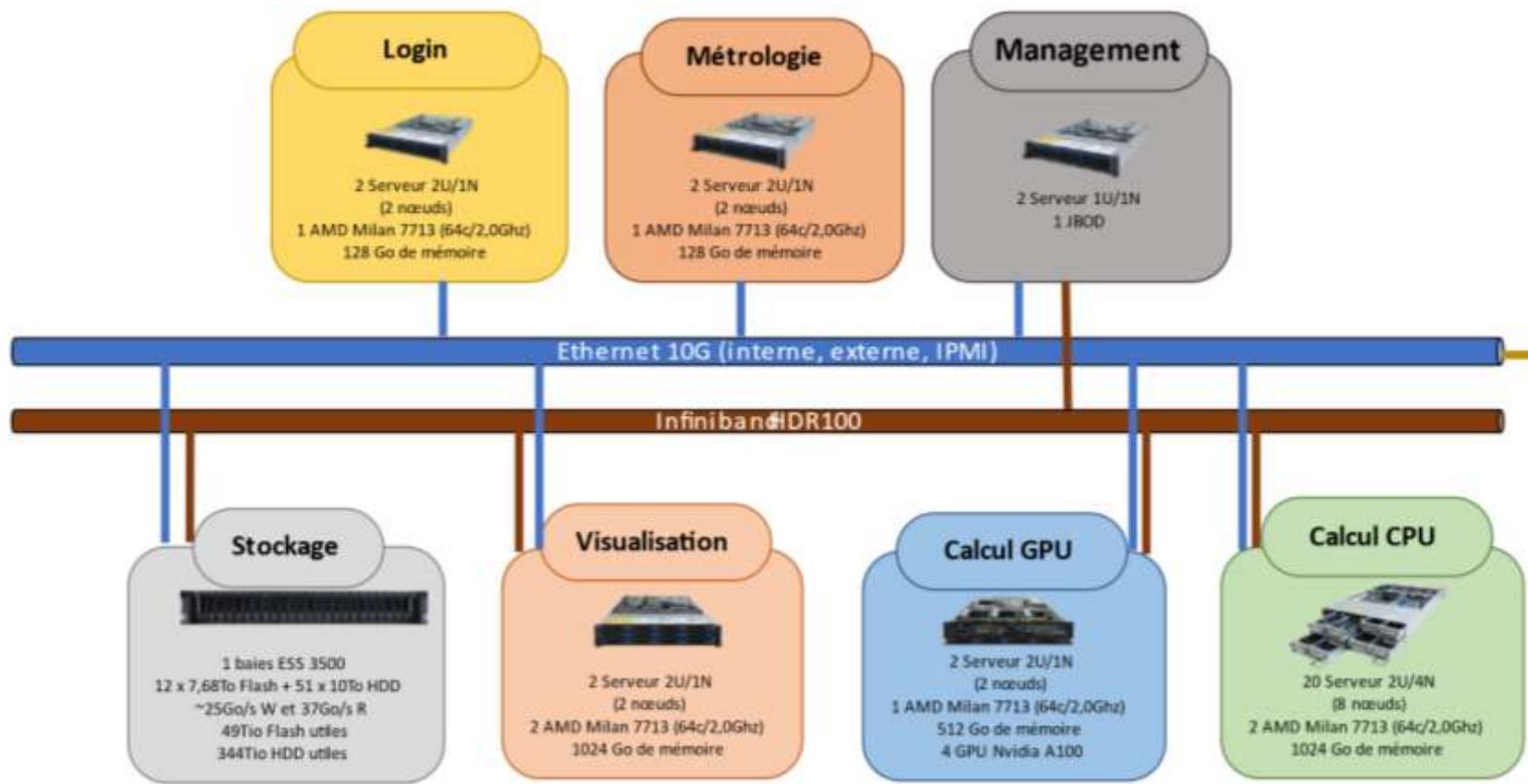
REX : TREX incidents et arrêts de prod (AP) depuis la mise en production 15/09/2023

Janvier		Février		Mars		Avril		Mai		Juin		Juillet		Août		Septembre		Octobre		Novembre		Décembre	
1 D		1 M		1 M		1 S		1 L		1 J		1 S		1 M		1 V		1 D		1 M		1 V	
2 L												2 D		2 M		2 S		2 L		2 J		2 S	
3 M												3 L		3 J		3 D		3 M		3 V		3 D	
4 M												4 M		4 V		4 L		4 M		4 S		4 L	
5 J												5 M		5 S		5 M		5 J		5 D		5 M	1 J
6 V												6 J		6 D		6 M		6 V		6 L		6 M	
7 S												7 V		7 L		7 J		7 S		7 M		7 J	
8 D												8 S		8 M		8 V		8 D		8 M		8 V	
9 L												9 D		9 M		9 S		9 L		9 J		9 S	
10 M		10 V		10 V		10 L		10 M		10 S		10 L		10 J		10 D		10 M		10 V		10 D	
11 M		11 S		11 S		11 M		11 J		11 D		11 M		11 V		11 L		11 M	2 H	11 S		11 L	
12 J		12 D		12 D		12 M		12 V		12 L		12 M		12 S		12 M		12 J		12 D		12 M	
13 V		13 L		13 L		13 J		13 S		13 M		13 J		13 D		13 M		13 V		13 L		13 M	
14 S		14 M		14 M		14 V		14 D		14 M		14 V		14 L		14 J		14 S		14 M		14 J	
15 D		15 M		15 M		15 S		15 L		15 J		15 S		15 M		15 S	MEP	15 D		15 M		15 V	
16 L		16 J		16 J		16 D		16 M		16 V		16 D		16 S	VSR	16 S		16 L		16 J		16 S	
17 M		17 V		17 V		17 L		17 M		17 S		17 L		17 J		17 D		17 M		17 V		17 D	
18 M		18 S		18 S		18 M		18 J		18 D		18 M		18 V		18 L		18 M		18 S		18 L	
19 J		19 D		19 D		19 M		19 V		19 L		19 M		19 S		19 M		19 J		19 D		19 M	
20 V		20 L		20 L		20 J		20 S		20 M		20 J		20 D		20 M		20 V		20 L		20 M	
21 S		21 M		21 M		21 V		21 D		21 M		21 V		21 L		21 J		21 S		21 M		21 J	
22 D		22 M		22 M		22 S		22 L		22 J		22 S		22 M		22 V		22 D		22 M		22 V	
23 L		23 J		23 J		23 D		23 M		23 V		23 D		23 M		23 S		23 L		23 J		23 S	
24 M		24 V		24 V		24 L		24 M		24 S		24 L		24 J		24 D		24 M		24 V		24 D	
25 M		25 S		25 S		25 M		25 J		25 D		25 M		25 V		25 L		25 M		25 S		25 L	
26 J		26 D		26 D		26 M		26 V		26 L		26 M		26 S		26 M		26 J		26 D		26 M	
27 V		27 L		27 L		27 J		27 S		27 M		27 J		27 D		27 M		27 V		27 L		27 M	
28 S		28 M		28 M		28 V		28 D		28 M		28 V		28 L		28 J		28 S		28 M		28 S	
29 D				29 M		29 S		29 L		29 J		29 S		29 M		29 V		29 D		29 M		29 D	
30 L				30 J		30 D		30 M		30 V		30 D		30 M		30 S		30 L		30 M		30 L	
31 M				31 V				31 M				31 L		31 J				31 M				31 M	

- 1 arrêt de production pour maintenance programmé le 5 décembre
- 1 incident bloquant (2h le 11/10) depuis la mise en production

Incidents (sans AP)
 Incidents (avec AP)
 Arrêts de production prévus (avec AP)

P/F Diffusion limitée 6G : ARYX (vue fonctionnelle)



1 024 CPU : Processeurs AMD (128 Cœurs par serveurs, 1TiB de RAM, 3GHz par core)

8 GPU: NVIDIA A30 24Go Pci-e

400 TiB de stockage Posix : Système de fichier parallèle et distribué **IBM Spectrum Scale** basées sur baies intelligentes **ESS-3500**

Réseau HP Infiniband HDR-100

Réseau **Ethernet 10G**

Système d'exploitation : **RedHat 8.X**

Orchestrateur Jobs : **SLURM**

Sommaire



Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie

Le Datalake : Focus sur la P/F de stockage synchrone

Le stockage synchrone (disque / Tier2) :

- ❖ 3 salles datacenter
- ❖ 9 baies (3 par salle)
- ❖ 69 serveurs de stockages (23 par salle)
- ❖ 30 Po (pétaoctet) utile de données

Performance du stockage synchrone (disque / Tier2) :

- ❖ Débits globaux :

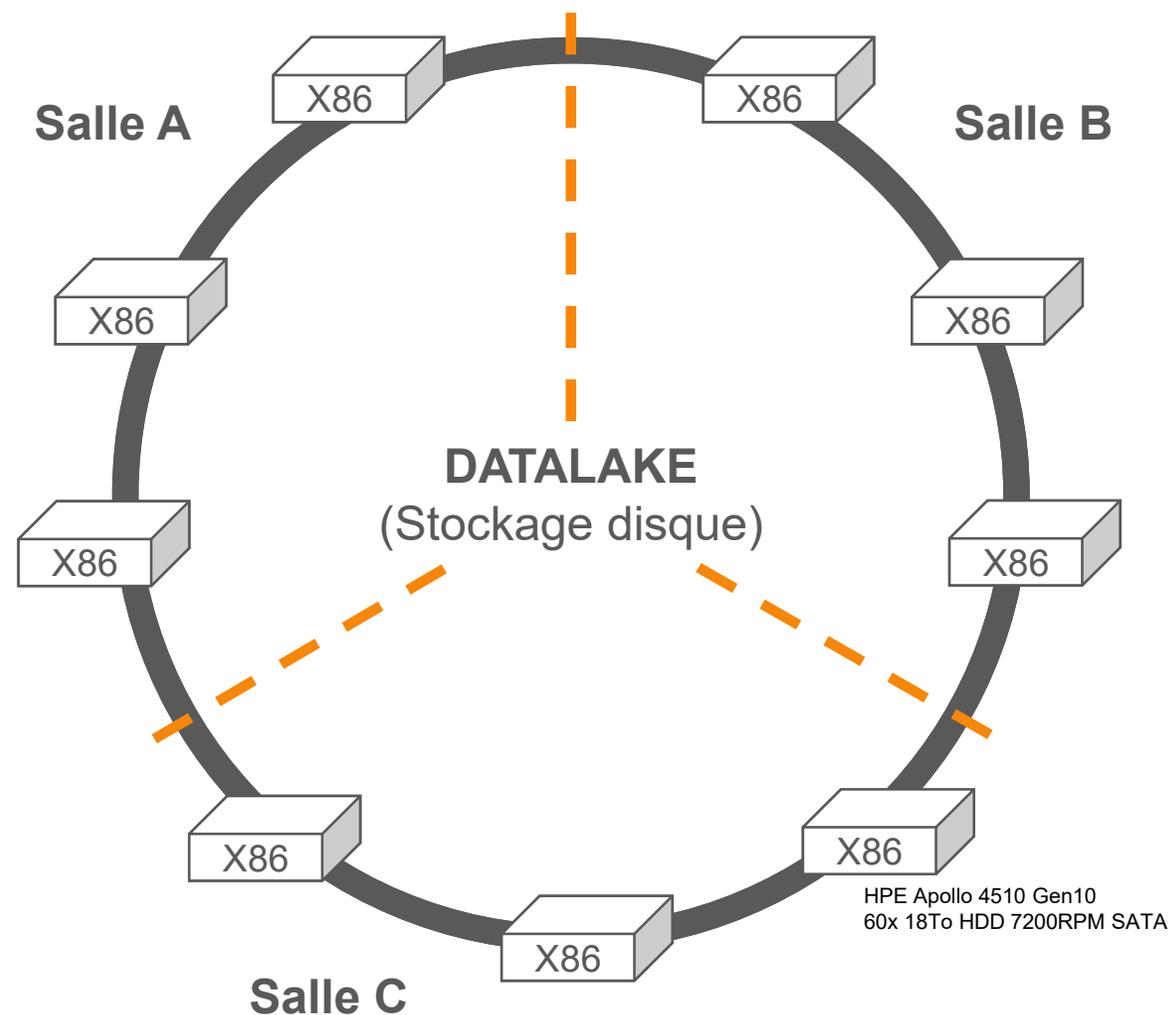
Lecture	Ecriture
8 Go/s	5 Go/s

- ❖ Exemple de débits projet :

PEPS	Ecriture
<ul style="list-style-type: none"> • Nombre de VM : 1 • Interface réseau : 10 Gbit/s • Client S3 : SDK Java 1.12 • Thread MPU : 20 • MPU chunk size : 20 Mo 	490 Mo/s (790 Mo/s en pic)

- ❖ Impact environnemental :

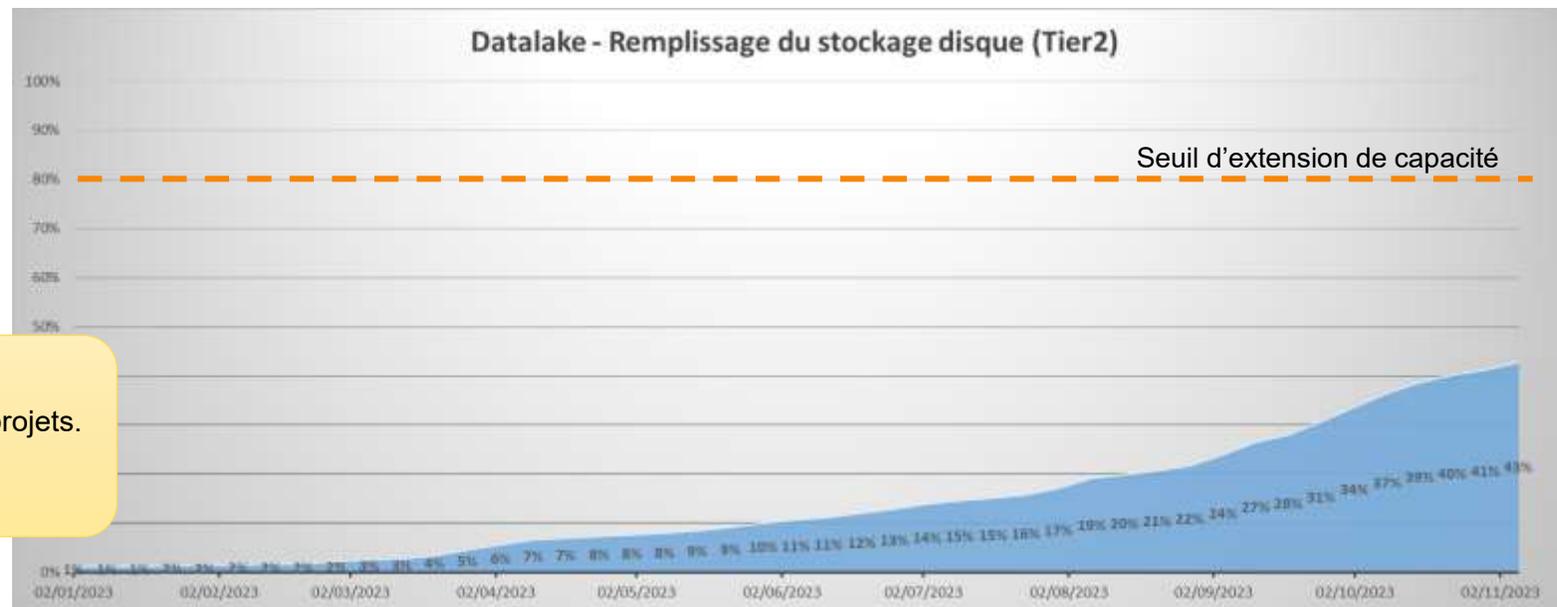
Tier2	Tier3
4 kWh/Po (24/24h)*	0,4 kWh/Po **



Datalake : Utilisation de la plateforme | Vue stockage sur l'année

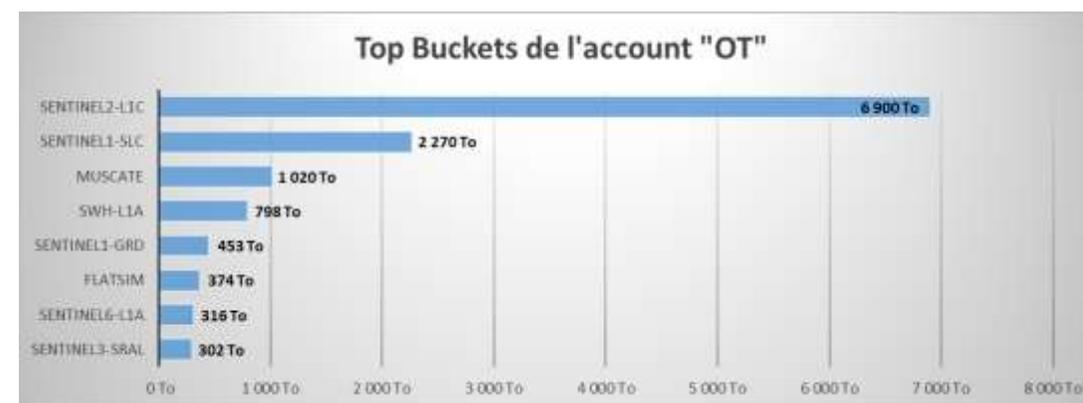
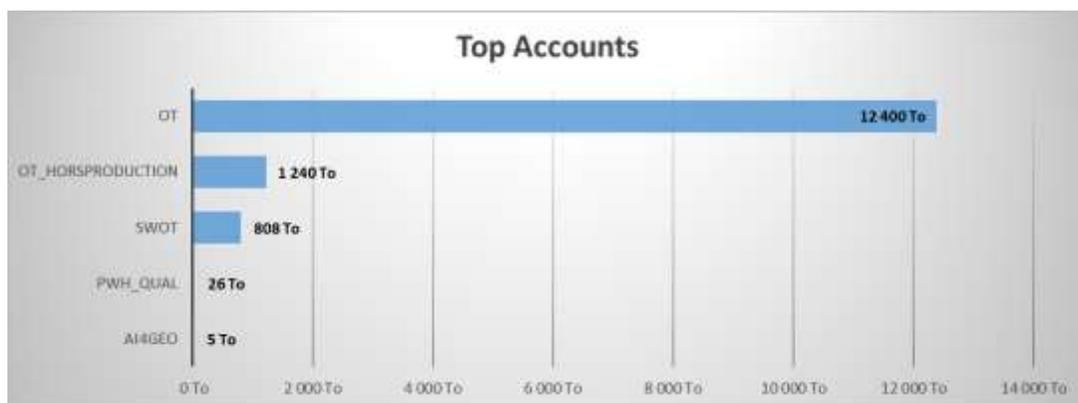
Datalake en quelques chiffres * :

- Accounts projet : 21
- Buckets projet : 117



Extension de capacité :

- Prérequis : Actualiser les besoins de volumétrie projets.
- Délais organisationnel : plusieurs mois.
- Extension groupée : 1 par année.



(* Chiffres de début novembre 2023

REX : Incidents DATALAKE et arrêts de prod. sur l'année

- 0 arrêt de production pour maintenance
- 3 Incidents avec dégradation des accès / performances
- 1 crise sur les performances : root cause hors Datalake



Plateforme opérationnelle **100%** sur 2023

Performance nominale 92,32% sur 2023

Les SLA sont applicables depuis Avril 2023

Janvier		Février		Mars		Avril		Mai		Juin		Juillet		Août		Septembre		Octobre		Novembre		Décembre	
1 D		1 M		1 M		1 S		1 L		1 J		1 S		1 M		1 V		1 D		1 M		1 V	
2 D												2 D		2 M		2 S		2 L		2 J		2 S	
3 L												3 L		3 J		3 D		3 M		3 V		3 D	
4 M												4 M		4 V		4 L		4 M		4 S		4 L	
5 M												5 M		5 S		5 M		5 J		5 D		5 M	
6 J												6 J		6 D		6 M		6 V		6 L		6 M	
7 V												7 V		7 L		7 J		7 S		7 M		7 J	
8 S												8 S		8 M		8 V		8 D		8 M		8 V	
9 D												9 D		9 M		9 S		9 L		9 J		9 S	
10 L												10 L		10 J		10 D		10 M		10 V		10 D	
11 M												11 M		11 V		11 L		11 M	4,5 H	11 S		11 L	
12 M												12 M		12 S		12 M		12 J		12 D		12 M	
13 J												13 J		13 D		13 M		13 V		13 L		13 M	
14 V												14 V		14 L		14 J		14 S		14 M		14 J	
15 D		15 M		15 M		15 S		15 L		15 J		15 S		15 M		15 V		15 D		15 M		15 V	
16 L		16 J		16 J		16 D		16 M		16 V		16 D		16 M		16 S		16 L		16 J		16 S	
17 M		17 V		17 V		17 L		17 M		17 S	26 H	17 L		17 J		17 D		17 M		17 V		17 D	
18 M		18 S		18 S		18 M		18 J		18 D		18 M		18 V		18 L		18 M		18 S		18 L	
19 J		19 D		19 D		19 M		19 V		19 L		19 M		19 S		19 M		19 J		19 D		19 M	
20 V		20 L		20 L		20 J		20 S		20 M		20 J		20 D		20 M		20 V		20 L		20 M	
21 S		21 M		21 M		21 V		21 D		21 M		21 V		21 L		21 J		21 S		21 M		21 J	
22 D		22 M		22 M						22 J		22 S		22 M		22 V		22 D		22 M		22 V	
23 L		23 J		23 J						23 V		23 D		23 M		23 S		23 L		23 J		23 S	
24 M		24 V		24 V						24 S		24 L		24 J		24 D		24 M		24 V		24 D	
25 M		25 S		25 S						25 D		25 M		25 V		25 L		25 M		25 S		25 L	
26 J		26 D		26 D		26 M		26 V		26 L		26 M		26 S		26 M		26 J		26 D		26 M	
27 V		27 L		27 L		27 J		27 S		27 M		27 J		27 D		27 M		27 V		27 L		27 M	
28 S		28 M		28 M		28 V		28 D		28 M		28 V		28 L		28 J		28 S		28 J		28 S	
29 D				29 M		29 S		29 L		29 J		29 S		29 M		29 V		29 D		29 D		29 M	
30 L				30 J		30 D		30 M		30 V		30 D		30 M		30 S		30 L		30 L		30 M	
31 M				31 V				31 M				31 L		31 J				31 M				31 M	

 Incidents (sans AP)

 Incidents (avec AP)

 Arrêts de production prévus (avec AP)

Le Datalake : Planning et Evolution

- ❖ **T4 2022 : Production Tier2 ouverte**
- ❖ **T2 2023 : Qual Tier3 (Premiers tests et interfaces avec les projets)**
- ❖ **T4 2023 : MEP Tier3 PROD = KO (Pb passage à l'échelle sur transition glacier)**
- ❖ **T1 2024 : Extensions de capacité du stockage sur Disque (Tier2) = limiter l'impact projet des retards Tier3**
- ❖ **T1 2024 : Ouverture de service du stockage sur Bande (Tier3) = MEP Partielle (initialement prévue 09/2023)**

➤ Performances cibles :

Lecture	Ecriture
2,5 Go/s	1,5 Go/s (100 To / 24h)

➤ Projets prioritaires : Migrations sur bandes des collections PEPS

- ❖ **Début T3 2024 : MEP définitive Tier3:**
 - Corrections de l'ensemble des anomalies
 - Tests PRA (avec Arrêt de Production) : *Dates à confirmer*

- ❖ **2024 : P/F Métrologie – Tableau de bord orientée métriques d'utilisation pour projets spatiaux**
 - L'objectif est de fournir aux responsables de projets spatiaux et utilisateurs du Datalake, une interface unique pour visualiser les métriques propres à leurs utilisations de la plateforme et de leurs espaces.
 - Ex: Par Account et Bucket :
 - Quota, espace consommé / disponible,
 - Statistiques : Nb objets, Nb accès, Nb transition / restauration, ...
 - Etc.

Sommaire



Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie

Bascule des utilisateurs et des projets de HAL => TREX

Le changement de plateforme de calcul a beaucoup mobilisé les équipes EVIDEN, THALES et CNES pour accompagner les utilisateurs.

- Organisation de la migration des filesets de HAL vers TREX en concertation avec chaque projet
- Portage des chaînes de traitements (PBS => Slurm, RH7 => RH8, architecture Intel vers AMD)
- Adaptation aux changements liés à l'architecture des nœuds AMD et à Slurm (accounts, partitions, QoS) pour les réservation de ressources :
 - Nombre de cœurs par nœuds : 128
 - Mémoire par cœur : 4,5/5 Go => 8 Go par cœur
- Accompagnement renforcé des projets producteurs de données en continu (SWOT, SSALTO, THEIA)

TREX : difficultés rencontrées

➤ Export des filesets du /work sur les machines clientes (VMs projets)

- Fonctionnement différent de l'existant sur HAL, avec une nouvelle règle de non-imbrication des exports [Lien Confluence pour plus d'informations sur le sujet](#)

➤ Conversion ACL NFS V3 vers ACL NFS V4

- Nouveau mécanisme d'export (CES) impose un passage aux ACL NFS V4
- Conversion automatique récursive impossible pour certains filesets contenant beaucoup d'inodes => choix de repartir sur une organisation plus simple des permissions => action qui nécessite des échanges avec le support

➤ Sessions de visualisation

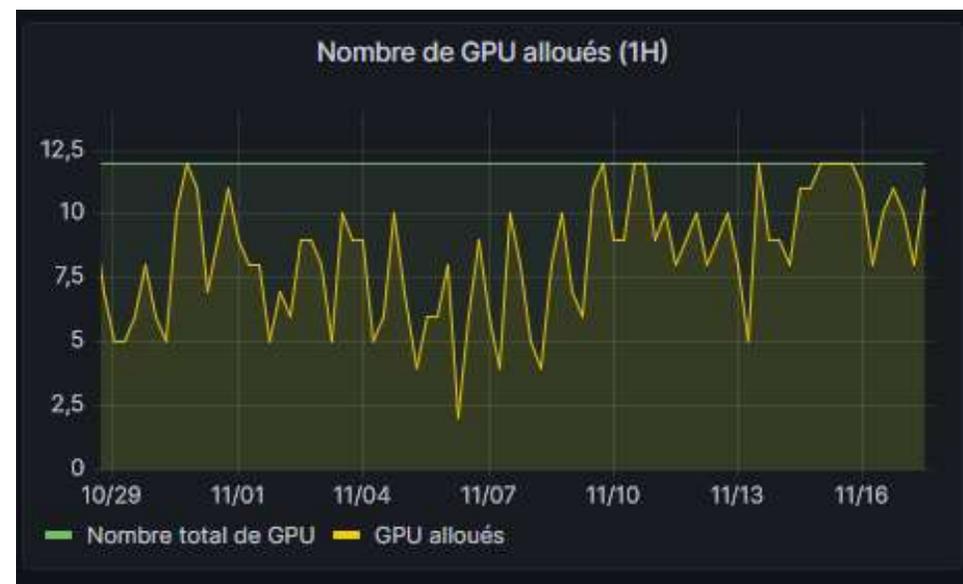
- Incompatibilité avec la version de MobaXterm disponible sur les postes CNES => nécessité d'installer un client léger

➤ Saturation des nœuds GPU

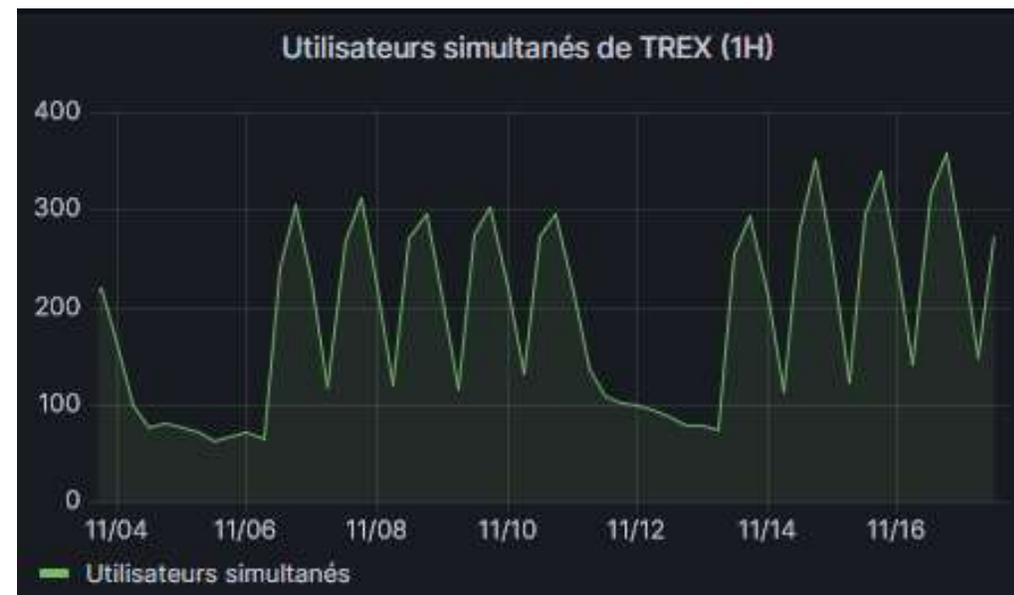
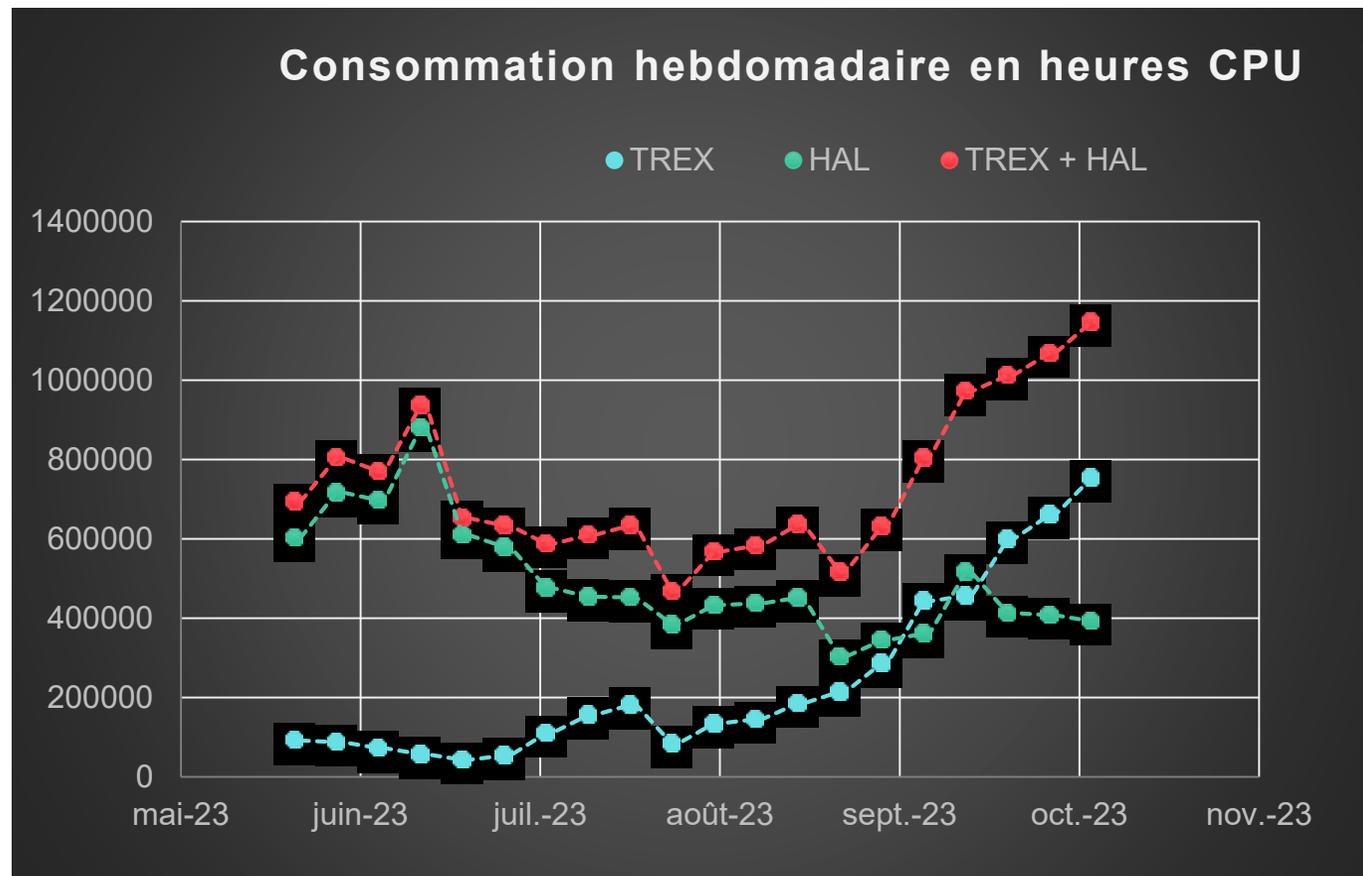
- Difficultés à réintégrer les nœuds DGX de HAL vers TREX

➤ Sujet en cours d'investigation

- Calculs parallèles MPI multi-nœuds qui gèlent de façon aléatoire

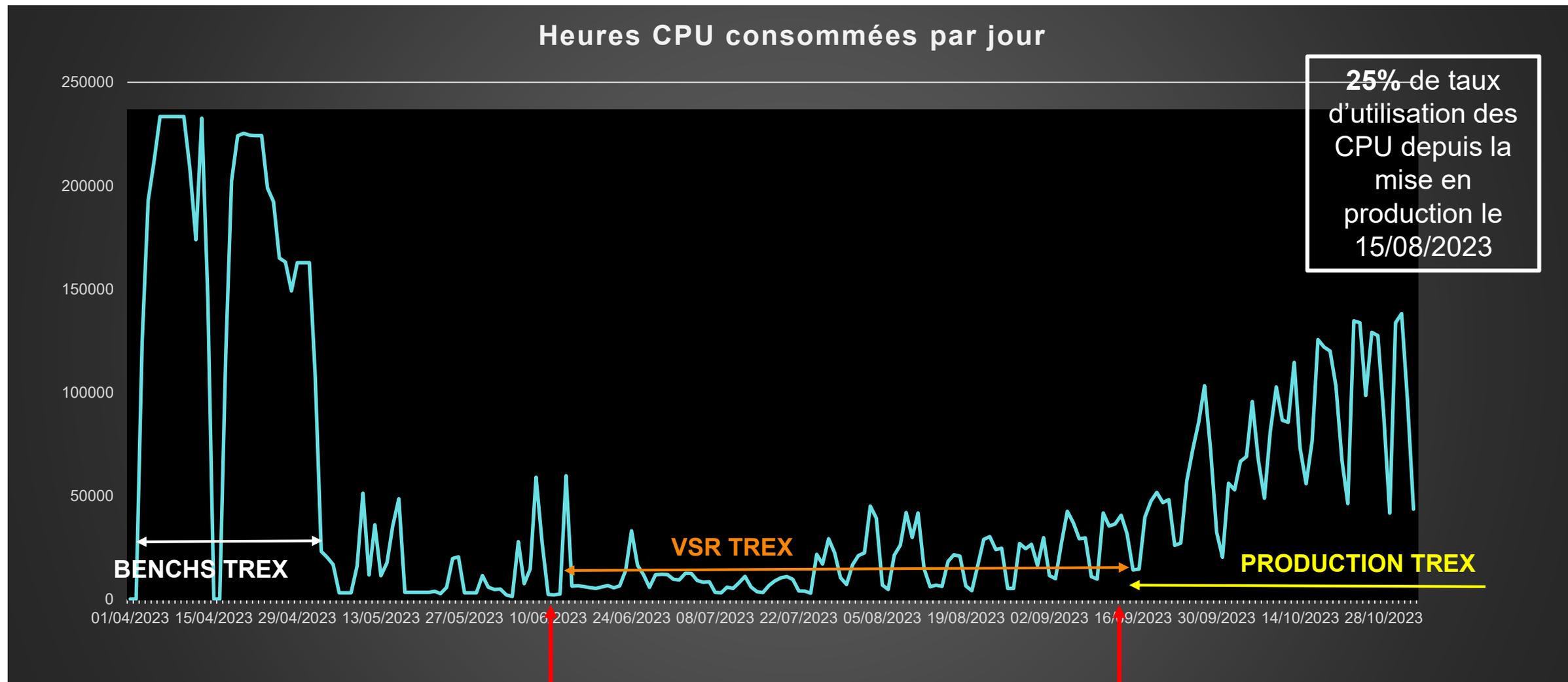


TREX : Utilisation depuis le début de la VSR



Au 3 novembre 2023, 880 logins IPA différents se sont déjà connectés à la plateforme TREX

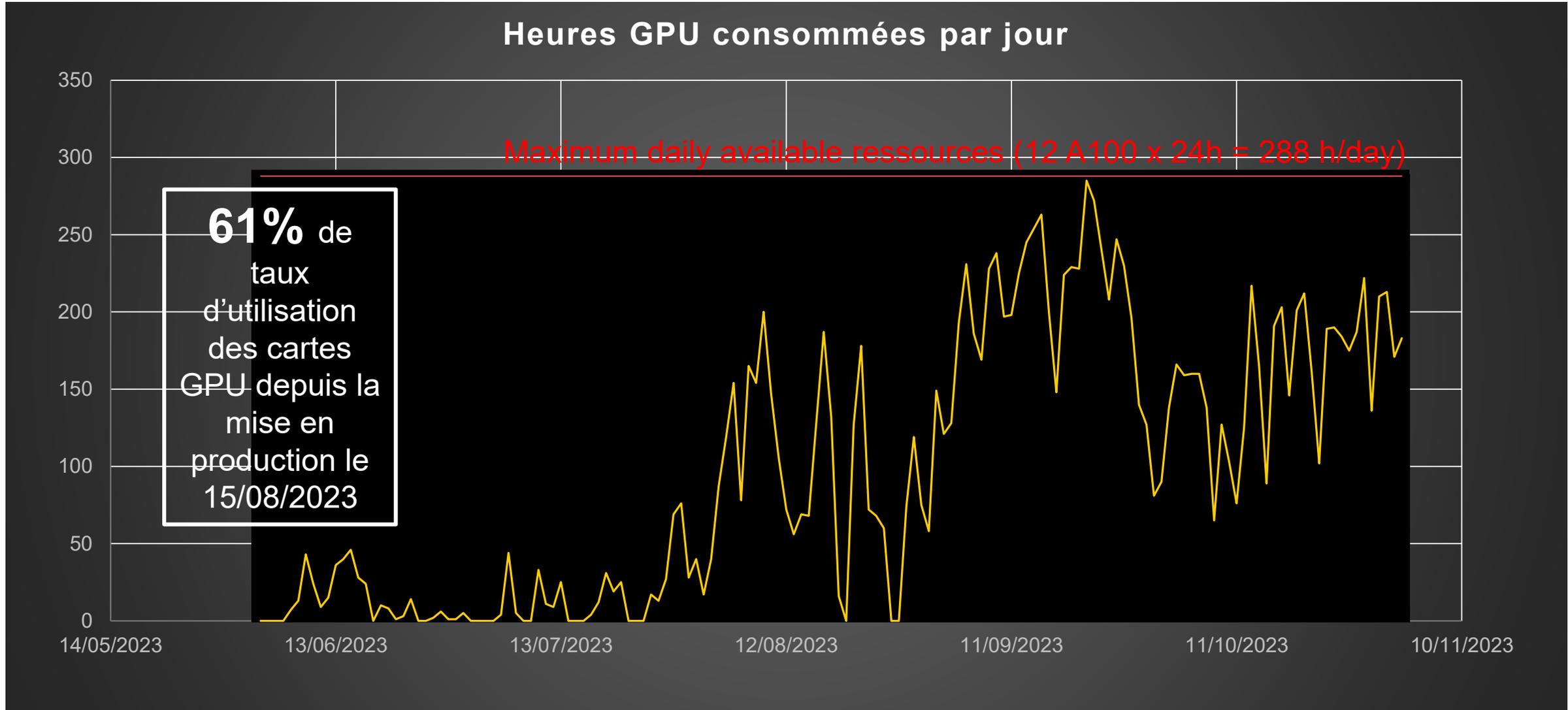
TREX : Consommation journalière (variabilité jours ouvrés / week-end !)



Début VSR
12/06/2023

Mise en production
15/09/2023

TREX : Consommation journalière des ressources GPU



TREX : Utilisation de la plateforme | Vue stockage

Presque la moitié des 12 PiB disponible est déjà utilisée (correspond à la volumétrie des filesets migrés depuis HAL)

Une commande d'une cinquième tête ESS a été effectuée fin 2023 pour permettre d'ajouter de la volumétrie si besoin (sur financement projet).



HPC6G : Merci à l'implication de nombreux acteurs

La mise en production de TREN ainsi que le travail toujours en cours sur les plateformes ARYX et Kubernetes nécessitent une collaboration étroite de nombreuses équipes!

L'équipe CNES du Centre de Calcul remercie :

- ✓ **Les équipes « Delivery » et « MCO » d'EVIDEN**
- ✓ **L'équipe Thales SI du support HPC**
- ✓ **L'équipe Réseau CNES et l'infogérant SPIE**
- ✓ **L'équipe projet et le groupe de revue HPC6G**
- ✓ **Le service DTN/ISA/SSM**
- ✓ **Les équipes segments sols de projets spatiaux avec qui nous avons itéré et réitéré... 😊**
- ✓ **Vous tous et toutes, utilisateurs du HPC pour votre patience et votre compréhension !**



Sommaire



Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

Les nouveaux services HPC6G :

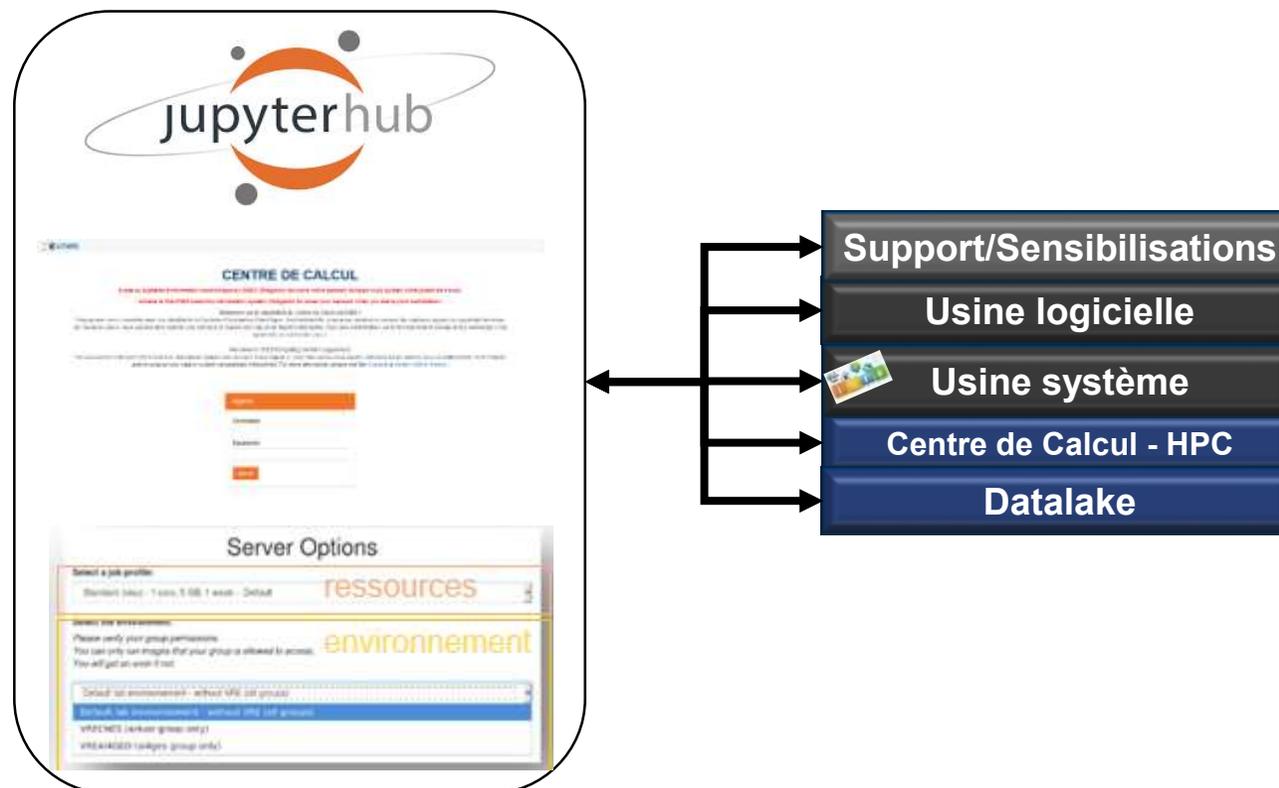
- Kubernetes
- Monitoring
- Energie

Datalabs | Présentation

Un Datalab (ou VRE) est un **espace de travail Web** offrant un **accès transparent à tous les composants/services** dont un utilisateur a besoin pour accéder à ses données, les traiter et collaborer/échanger avec sa communauté du POC jusqu'à la Production*.

Objectif : Fournir un écosystème user-friendly pour accéder et utiliser les principales briques du SIS.

<https://jupyterhub.cnes.fr>



+ Info sur :



* Plateforme pas assez mature pour accueillir des projets opérationnels pour le moment. Si le besoin est là, faite le nous savoir.

ⓘ Le Datalabs **ne se substitue pas aux structures ou projets**. Il fourni un cadre fonctionnel et organisationnel. **Chaque projet ou structure reste décisionnaire sur ses choix stratégiques !**

Datalabs | Environnements disponibles

Environnements disponibles :

- **Non containerisés :**
 - Environnement Natif (non containerisé) :

Check to use latest version of VRE (might be unstable)

- [HPC6G] Default lab environnement - without VRE (all groups)
- [HPC6G] Default lab environnement - without VRE (all groups)
- [HPC6G] VRECNES (vrecnes group only)
- [HPC6G] VREOT (vreet group only)
- [HPC5G] Default lab environnement - without VRE (all groups)
- [HPC5G] VRECNES (vrecnes group only)
- [HPC5G] VREOT (vreet group only)

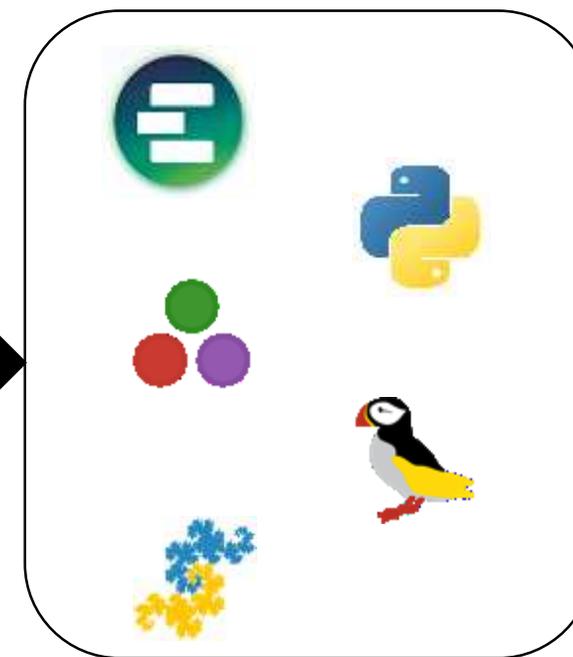
1. Jupyterhub



2. Jupyterlab



3. Kernels



Slurm



Datalabs | Environnements disponibles

Environnements disponibles :

- **Containerisés** : Virtual Runtime Environment

Check to use latest version of VRE (might be unstable)

[HPC6G] Default lab environnement - without VRE (all groups)

[HPC6G] Default lab environnement - without VRE (all groups)

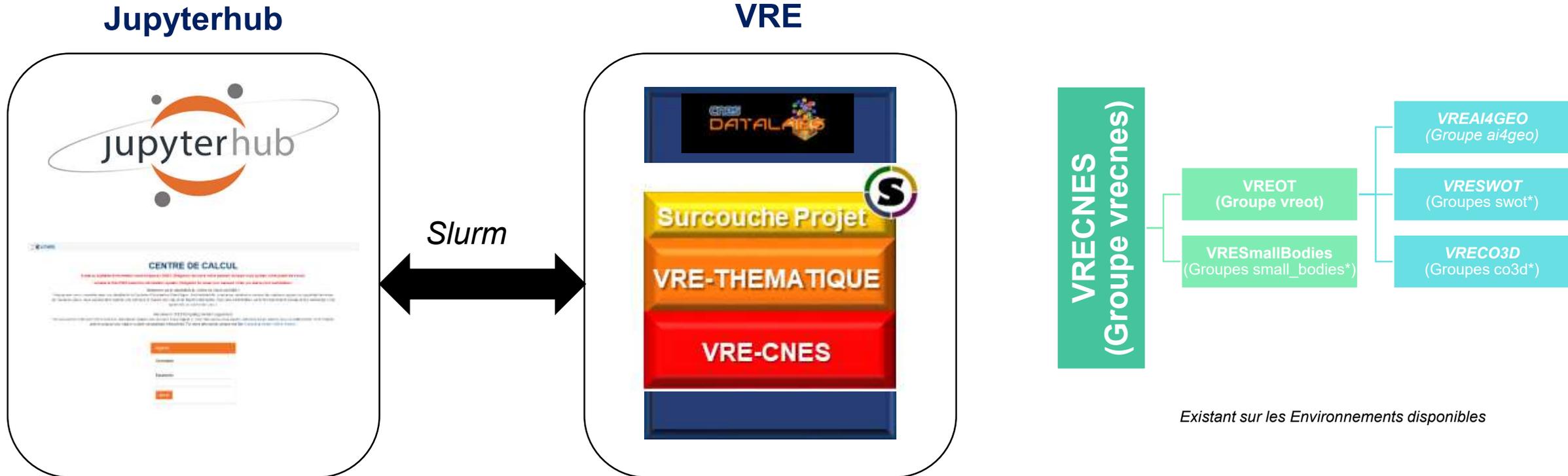
[HPC6G] VRECNES (vrecnes group only)

[HPC6G] VREOT (vreet group only)

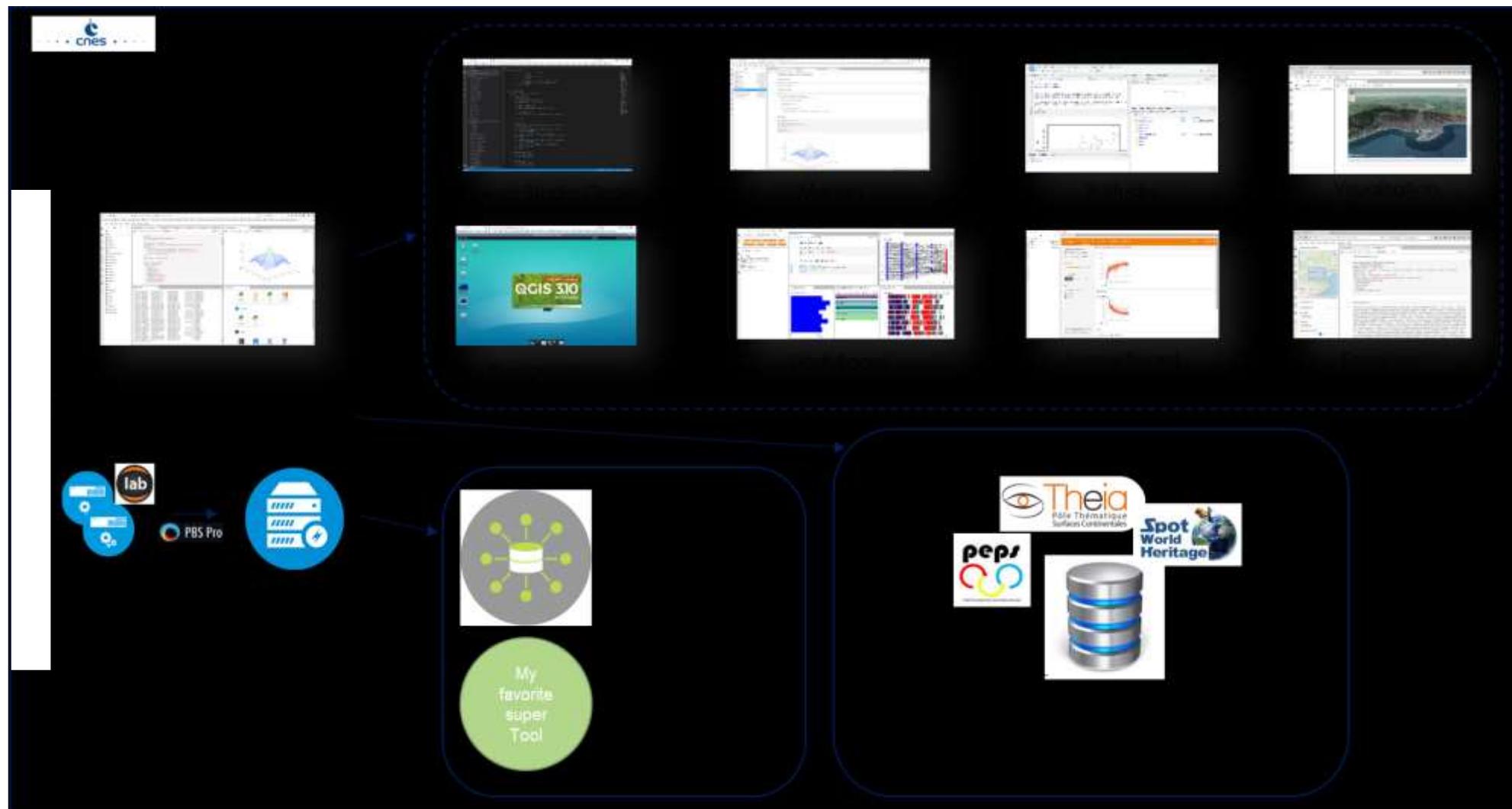
[HPC5G] Default lab environnement - without VRE (all groups)

[HPC5G] VRECNES (vrecnes group only)

[HPC5G] VREOT (vreet group only)



Datalabs | Fonctionnalités disponibles depuis une VRE



Datalabs | Rappel important

Tout projet, utilisateur (en lien avec le CNES) a accès au Datalabs et aux moyens du centre de calcul sur simple demande.

Natif

Seule condition : Avoir un compte nominatif IPA

⚠ Prérequis

Afin de pouvoir utiliser les différentes VRE à votre disposition il vous faut :

- Avoir un compte sur le cluster de calcul du cnes (HAL)
- Appartenir au groupe : **vrecnes** (pour les briques de bases)
- Additionnellement :
 - Appartenir au groupe : **vreot** (pour accès aux briques thématiques de l'observation de la Terre)
 - Appartenir au groupe : **vresc** (pour accès aux briques thématiques des sciences)
 - Appartenir au projet cible (ai4geo, hysopell, etc) pour avoir accès aux surcouches projets

Pour ce faire il faut passer par une demande catalogue : [Liens vers la demande d'ajout de groupe.](#)

Note : Il faut effacer toutes mentions à des VRE customisés dans le .bashrc pour que le chargement des VRE se fassent sans encombre

Demande
de groupe :
[MVN](#)

+ Info sur :



① Pour toutes questions ou demandes de support, n'hésitez pas à contacter : SIS-supportHPC@cnes.fr

Datalabs | Ressources disponibles

Service JupyterHub : VM

Nom	vCPU	Memory (GB)	Disk (GB)	OS	Status
tu-juphub-p01	1	4 Go	52 Go	CentOS 4/5 or later (64-bit)	Powered On
tu-juphub-q01	1	4 Go	52 Go	CentOS 4/5 or later (64-bit)	Powered On

Ressource de calculs de TREX :

- **CPU : Nœuds de calcul (AMD/RH8)**
- **GPU**

[HPC6G] Batch g2022 (RH8/AMD) - 1 core, 8 GB, 12 Hours

[HPC6G] Batch g2022 (RH8/AMD) - 4 core, 32 GB, 12 Hours

[HPC6G] Batch g2022 (RH8/AMD) - 32 core, 256 GB, 12 Hours

[HPC6G] GPGPU - 1 gpgpu A100, 16 cores, 128 GB, 4 hours

[HPC6G] Batch g2022 (RH8/AMD) - 1 core, 8 GB, 12 Hours

ⓘ Point d'attention sur l'utilisation des ressources : ressources mutualisées et à utiliser selon la réservation faite

Stockage : Espaces projets, espaces scratch/ (**temporaire**) et Datalake V2 (T1 2023)



Haute Performance/GPFS POSIX (HPC – Tiers 1)



Capacitif Objet (Datalake – Tiers 2 et 3)

Datalabs | Ressources disponibles

Nouveauté TREX : Slurm et la notion d'account

Le Jupyter Hub (hors VRE) n'est désormais accessible aux comptes nominatifs uniquement (pas aux comptes techniques), pour 2 raisons :

- La nécessité de soumettre sur un account `cnes_level*` que les comptes techniques ne possèdent pas,
- Pour une question d'usages cibles du service JupyterHub

Sujet ouvert sur la mise à disposition de profils restreints pour les comptes techniques.

Des profils de ressources dynamiques et différents selon les `cnes_level*` (1, 2 ou 3). Ex, pour `cnes_level2` :

[HPC6G] Batch g2022 (RH8/AMD) - 1 core, 8 GB, 12 Hours

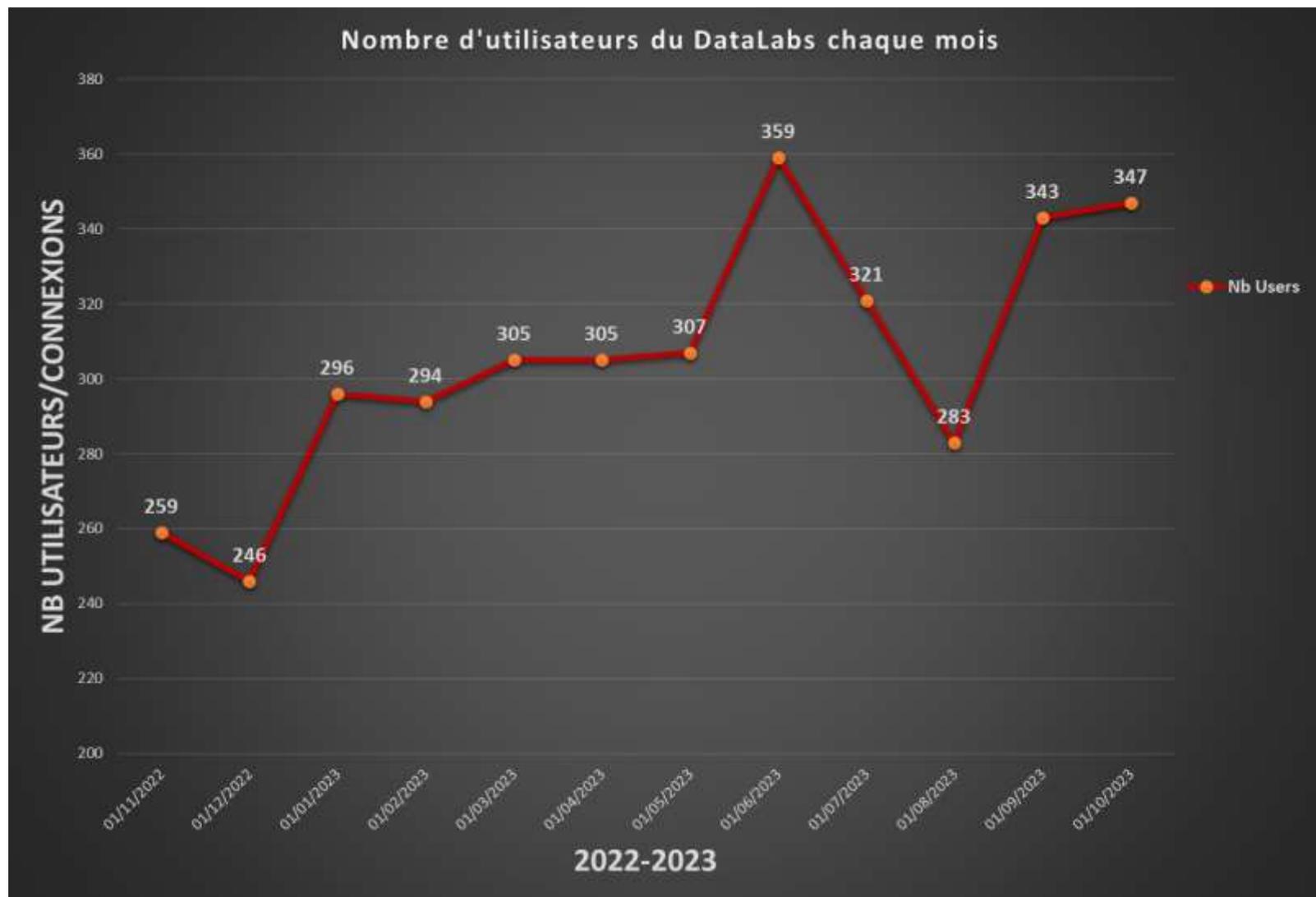
[HPC6G] Batch g2022 (RH8/AMD) - 4 core, 32 GB, 12 Hours

[HPC6G] Batch g2022 (RH8/AMD) - 32 core, 256 GB, 12 Hours

[HPC6G] GPGPU - 1 gpgpu A100, 16 cores, 128 GB, 4 hours

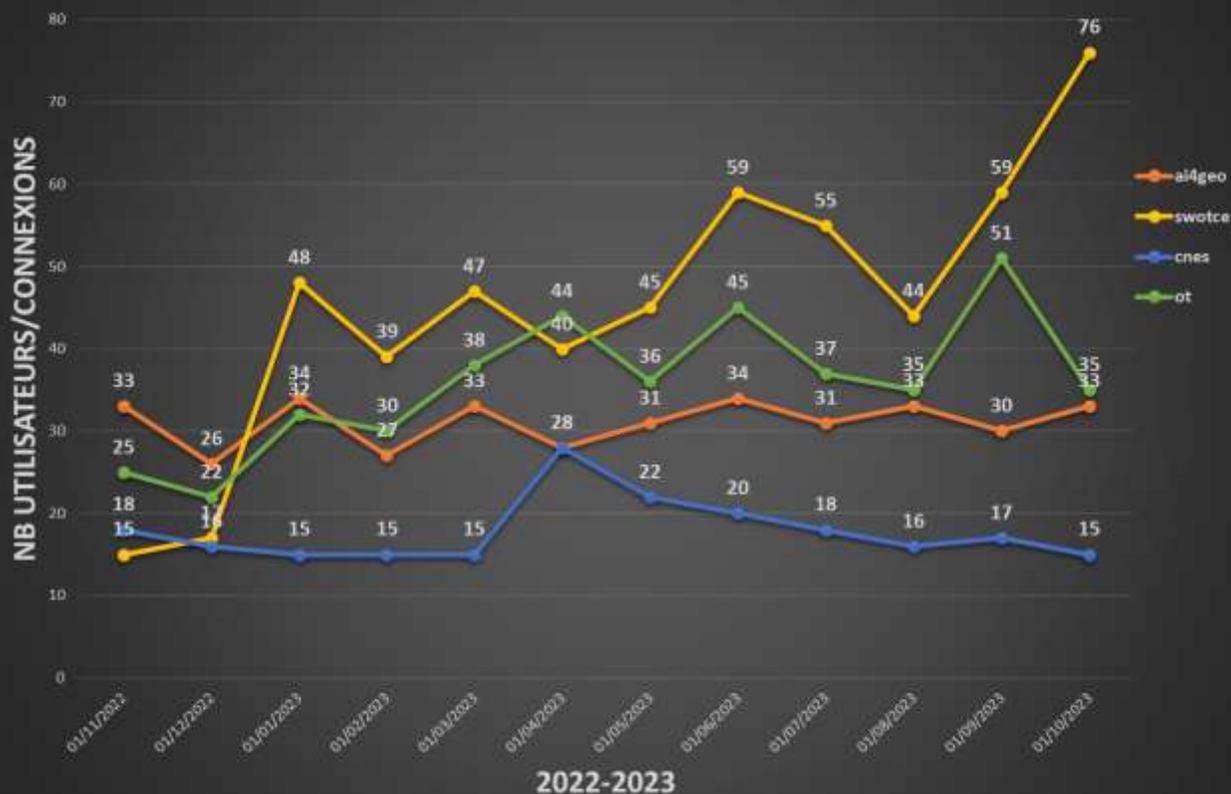
[HPC6G] Batch g2022 (RH8/AMD) - 1 core, 8 GB, 12 Hours

Datalabs | Utilisation du Datalabs 2022-2023



Datalabs | Utilisation du Datalabs 2022-2023

Utilisation des VREs chaque mois

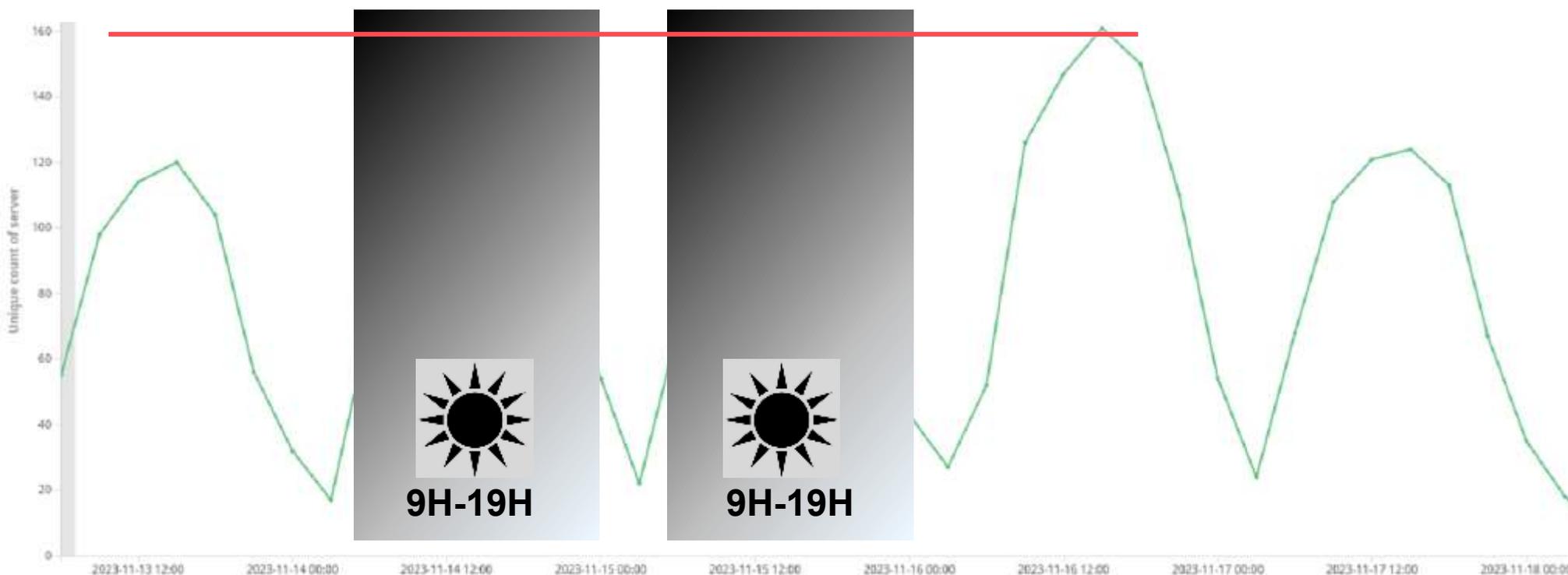


Utilisation de l'environnement Natif chaque mois



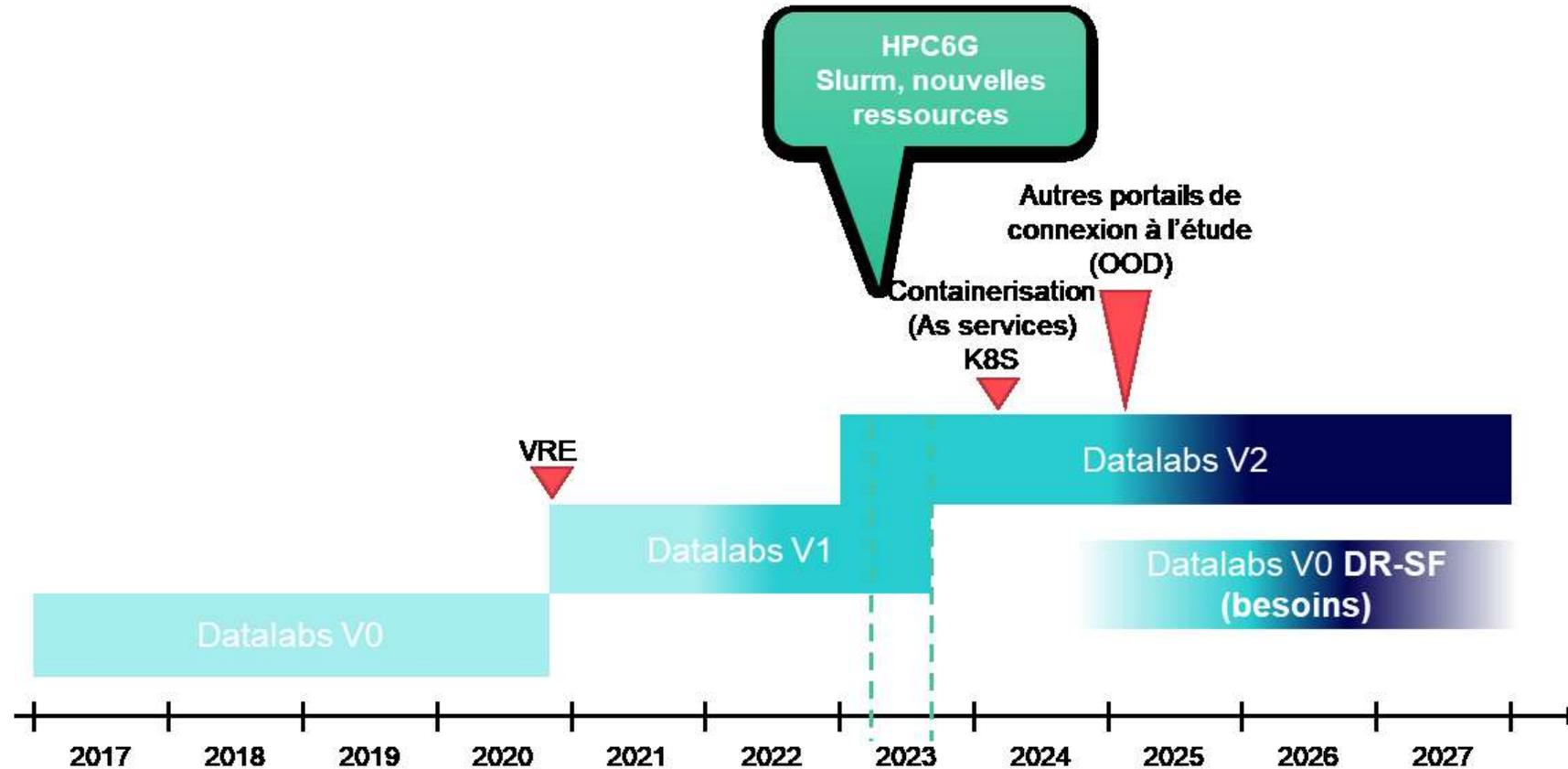
Datalabs | Utilisation journalière du Datalabs 2022

Plus de 150 sessions (utilisateurs) actives par jour ouvré et en heures ouvrées !



ⓘ Pic d'activité ciblé sur les horaires travaillés et sur les ressources mutualisées du cluster TREX => Il peut avoir des latences voir des timeouts au démarrage de votre serveur JupyterHub (notamment pour les ressources GPU).

Datalabs | Roadmap



① Des communications sont et seront faites à travers le blog Confluence de [l'espace DataLabs](#).

Conteneurisation

Multitude de technologies : Docker, Singularity, Apptainer, PODMAN, Sarus.

Une grande question : Laquelle choisir pour un environnement HPC mutualisé ?



Plusieurs conditions pour avoir la solution/technologie idéale :

- ✓ La sécurité :
 - ✓ Mode rootless
 - ✓ Avoir une bonne encapsulation : user namespace, permission sur les espaces de stockage ...
- ✓ Facilité d'utilisation, interface avec les services (forge logicielle), respect des spécifications OCI
- ✓ Utilisation des ressources de calcul (performance) : Infiniband, GPU



Pas de technologie qui remplit toutes les cases.

Etude métier fin 2022 pour être en veille sur les technologies et trouver le meilleur compromis

Conteneurisation

Les deux technologies utilisées et mises à disposition pour le moment : Docker et **Singularity** (recommandé).



Pour l'utiliser, il faut :

- ✓ Utiliser le process de BUILD spécifique forge logiciel : cf. [Usine Logiciel](#)
- ✓ Pas d'utilisation en RUN sur les nœuds mutualisés



Pour utiliser, il faut :

- ✓ Pour BUILDER ses images : utilisation du mode *fakeroot* sur les nœuds de visualisation exclusivement (visu01-visu06)
- ✓ Lancer ses conteneurs depuis **toutes** les ressources HPC 😊.

La sécurité !

Les solutions build et run proposées sont des contournements/wrappers.

Non OCI !

Interface non fonctionnelle avec certains outils : Artifactory

Besoin de conteneurisation sur HPC avec maîtrise du build en mode **rootless** :

- **Singularity** : <https://gitlab.cnes.fr/hpc/wikiHPC/-/wikis/page-containers#singularity>

Conteneurisation : le compromis

Utiliser un pipeline complet (BUILD + RUN + REGISTRY) en utilisant les moyens de la forge logicielle et du Centre de Calcul

Il faut :

- ✓ Utiliser le process de **BUILD** spécifique forge logiciel à partir de DockerFile : cf. [Usine Logiciel](#)
- ✓ Placer l'image Docker dans un dépôt Artifactory
- ✓ Pour la phase de **RUN** sur HAL :
 - ✓ Convertir l'image Docker récupéré de l'Artifactory en image Singularity
 - ✓ Lancer ses conteneurs Singularity



D'autres technologies disponibles en version Beta sur TREX :

- Sarus : Pour le RUN uniquement. Possibilité d'exécuter une image Docker directement tout en gardant l'uid/gid de l'utilisateur au sein du conteneur
- Apptainer : Fork de Singularity

+ d'infos sur le wiki HPC : <https://gitlab.cnes.fr/hpc/wikiHPC/-/wikis/page-containers#outils-de-conteneurisation>



Sommaire

Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie



CENTRE DE CALCUL
DU **cnès**

Support utilisateurs : les logiciels

- ❖ /softs est un espace mutualisé, installation utilisateurs/projets possible dans /softs/projets
 - /softs : > 100 applications différentes (163 modules disponibles)
 - /softs/projets : ~20 applications différentes (48 modules disponibles)
- ❖ Des environnements mis à disposition
 - Conda pour des environnements personnalisés et pas seulement pour les environnements Python (ex : qgis, gdal ...).
 - Pour éviter les problèmes de conflits et avoir une construction plus rapide, possibilité d'utiliser Mamba (déjà installé sur HAL).
- ❖ Pensez aussi à Singularity si besoin de personnalisation et environnement complexe.

[Plus d'informations sur l'espace Confluence de l'équipe Fondation Calcul](#)



Installation automatique reproductible avec Spack

Spack est un gestionnaire de paquets adapté pour le déploiement sur HPC avec compilation des sources compatible avec l'environnement module (Lmod).

Cet outil permet une meilleure reproductibilité, une automatisation et a permis d'anticiper la migration vers RH8.

- Alternative intéressante à Anaconda :
 - **Spack permet de créer ses environnements logiciels en s'appuyant la pile mutualisée (possibilité de « stacker » contrairement à Conda)**
 - **Spack est libre contrairement à Anaconda qui a changé ses Terms of Service (depuis janvier 2023 le CNES paye une licence Anaconda avec un prix proportionnel au nombre d'utilisateurs)**



```
# Install a specific version by appending @
$ spack install hdf5@1.10.1

# Specify a compiler (and optional version), with %
$ spack install hdf5@1.10.1 %gcc@4.7.3

# Add special boolean compile-time options with +
$ spack install hdf5@1.10.1 %gcc@4.7.3 +szip

# Add custom compiler flags
$ spack install hdf5@1.10.1 %gcc@4.7.3 cppflags="-O3 -floop-block"
```

```
# This is a Spack Environment file.
#
# It describes a set of packages to be installed, along with
# configuration settings.
spack:
  packages:
    all:
      providers:
        mpi: [mpich]

# add package specs to the `specs` list
specs: [tcl, trilinos]
```

Installation mutualisée des logiciels RedHat 8 pour HPC6G

Sur RH8, le support HPC a déjà installé sur TREX :

- la dernière version des logiciels présents (et toujours utilisés) sur HAL (/softs)
- uniquement des versions maintenues (exemple pour les compilateurs Intel : année N, N-1 et N-2)
- pour Python : une ou deux versions "standard" avec les paquets les plus demandés/utilisés comme sur HAL (Python 3.8 avec Spack) puis utilisation des environnements virtuels pour le reste

Puis au fur et à mesure de la vie de la plateforme :

- une nouvelle version de chaque logiciel une fois par an par défaut (LTS)
- une version plus récente tous les 6 mois à la demande des projets
- de nouveaux logiciels si utilisation commune à plusieurs projets sinon => /softs/rh8/projets

Support utilisateurs : logiciels RH7 et RH8

Le catalogue d'applications disponibles sur les nœuds g2019 de TREX **est identique à celui de HAL**

Arborescence (point de montage)	Source	Monté sur
/softs/rh7	NAS	sur g2019
/softs/projets	NAS	sur g2019
/softs/rh8	/work	sur g2022
/softs/rh8/projets	/work	sur g2022
/softs/rhall	/work et NAS	sur g2019 et sur g2022

Pour soulager les connexions au NAS mutualisé et améliorer les performances, sur RH8 les logiciels sont à présent installés sur le /work. (le support HPC ne réalise plus d'installations RH7)

Support utilisateurs

Depuis janvier 2021 : Société Thales SI
Polyvalence de l'équipe ~10 ETP sur les deux lots (support du quotidien + prestation projets).

- ❑ **Contacter le Support HPC : [MaVieNumérique](https://www.cnes.fr/ma-vie-numerique) ou SIS-supportHPC@cnes.fr**
 - **Support quotidien pour l'utilisation du Centre de Calcul**
 - **Support optimisation et parallélisation de codes (< 3 jours)**
 - **Monitoring et supervision**
 - **Suivi des incidents**
 - **Support spécifique pour les projets (monitoring, queue dédiée, ...)**
 - **Installation de logiciels**

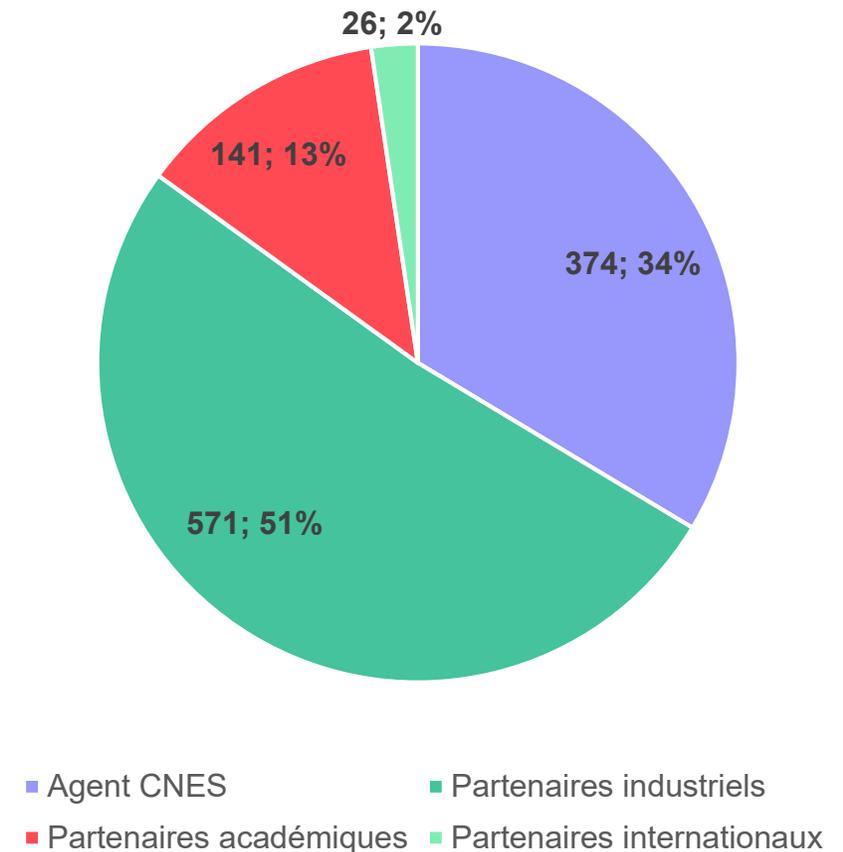
- ❑ **Support Portage et Optimisation de Codes**
 - **Prestations entre 1 et 12 mois sur bons de commande**



Qui sont les utilisateurs du Centre de Calcul?

- ~ 1300 utilisateurs en 2023
- provenance diversifiée (CNES, industriels, académiques)
- compétences HPC « variées » : du débutant à l'expert
- des usages très divers : HTC, HPC, IA, ... pour de la production, du développement, des Proof of Concept, ...

1122 utilisateurs différents se sont connectés à TREX les 3 derniers mois



Authentification

- ❖ **Authentification via l'annuaire IPA**
 - Plus de 2500 comptes actifs, ~2100 utilisateurs nominatifs
 - ~stable sur un an

- ❖ **Campagnes de désactivation automatique des comptes obsolètes**
 - Basée sur le statut GUID (date de fin de mission...)
 - Attention aux notifications pour les responsables CNES

- ❖ **Cycle de vie des comptes dans l'annuaire IPA-DR également**

- ❖ **Responsabilité de l'annuaire déléguée à l'infogérant CNES**
 - Contacter le Guichet-IT en cas de problème



Support utilisateur : les bonnes pratiques

❖ Les principaux types d'incidents

- Connexion / transfert / mot de passe
- Job PBS/Slurm : performances, temps d'attente, jobs en erreur
- Logiciels : environnement utilisateur, proxy, licences



❖ Les solutions

- Participation aux sessions de sensibilisation
- Lecture du Wiki HPC
- Monitoring des jobs
- Bonne gestion des ressources (CPU, mémoire, I/O)
- Bonne gestion de son environnement (quotas disques, .bashrc)
- Contacter le support : MaVieNumérique, GuichetIT (40) , SIS-supportHPC@cnes.fr



Support utilisateur : bilan des sensibilisations

❖ Sensibilisations pour les nouveaux utilisateurs du Centre de Calcul et pour faciliter la transition vers TREX

- deux modules proposés (niveau 1 et niveau 2)
- 12 demi-journées dispensées en 2023
- ~300 inscriptions (200 personnes différentes sensibilisées)
- inscriptions en ligne sur le [wiki HPC](#)



❖ Tutoriels Pangeo/Jupyter disponibles : <https://gitlab.cnes.fr/inno/pangeo-jupyter-demo>



❖ Une sensibilisation OpenMP prête à être dispensée



Visios à thèmes et codes sprint

Format : visio 1h ouvert à tout le monde + 2h de TP en salle sur inscription

- **Concepts de base du Datalake** : mardi 20 juin 2023
Une centaine de personnes connectées, 11 personnes ont assisté au TP
- **Mesurer l'efficacité d'un code Python** : mardi 28 mars 2023
82 personnes connectées, 8 personnes ont assisté au TP
- **Concepts de base du Datalake** : mardi 14 février 2023
132 personnes connectées à la visio, 9 personnes ont assisté au TP
- **Sensibilisation pour les nouveaux utilisateurs du Datalabs** : mardi 31 janvier 2023
52 personnes connectées à la visio, 8 personnes ont assisté au TP
- **Introduction à Slurm** : mardi 6 décembre 2022
50 personnes connectées à la visio, 8 personnes ont assisté au TP

Supports disponibles ici :

<https://gitlab.cnes.fr/hpc/wikiHPC/-/wikis/Visios-a-themes-et-retours-d'experiences>

Support utilisateur : la supervision

Objectif : Anticiper pour éviter des incidents bloquants.

❖ Les alertes remontées par la supervision

- Charge des espaces disques NAS : /home et /softs
- Quota des espaces \$HOME et des espaces partagés (fileset, /work/scratch/data, /work/scratch/env)
- Jobs Slurm en erreurs
- Charge des nœuds de visualisation
- Espace \$TMPDIR des nœuds de calcul
- Expirations de mot de passe et désactivations de comptes

❖ Les solutions

- Mails automatiques envoyés aux utilisateurs ou responsables d'espace projet
- Contact des utilisateurs

Nouveauté TREX : Service de métrologie via Grafana

- Consommation de job
- Logs



Bonjour,

Nous vous contactons car le compte `fa` **FileSetScratch** : 819.245GB. Par volume occ votre compte sur l'ensemble du **FileSet Scratch** scratch mais possiblement d'autres répertoires des écritures.

Voici un petit rappel de la page du Wiki HPC (<https://gitlab.cnes.fr/hpc/wikiHPC/-/wikis/espaces-stockages>) :

/work/scratch/\$USER : espace utilisateur de travail partagé

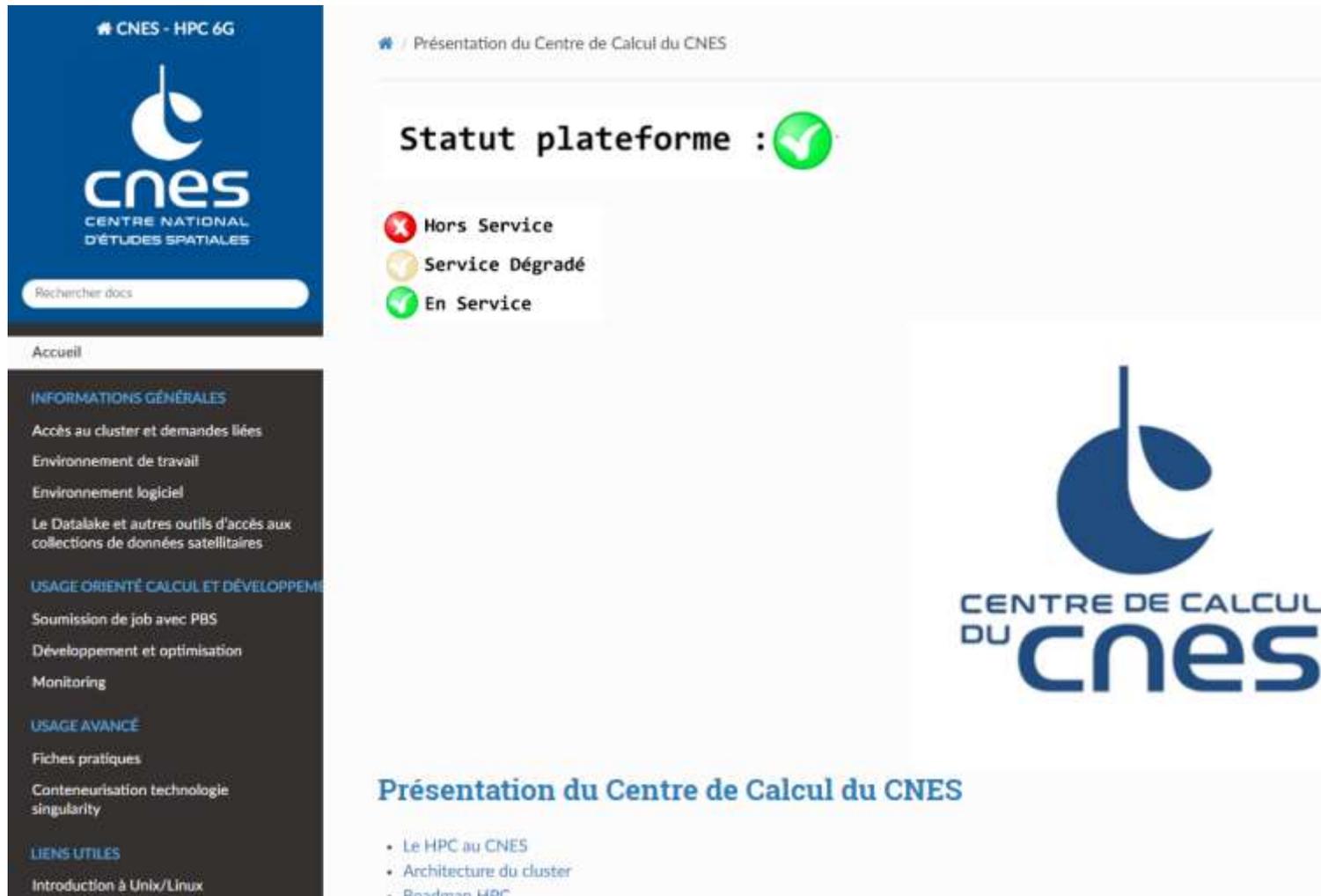
- Espace de stockage non sauvegardé
- Volume capacitif utilisateur (limité à 2To par défaut)
- Visible depuis tous les noeuds du cluster
- Utilisable par les jobs de calcul SSI les accès aux données sont conformes aux règles de bonnes pratiques

Cibles d'utilisation :

- hébergement de fichiers de données (produits spatiaux, modèle, maillage, etc.)
- évitez d'héberger des sources logicielles

Cet espace est temporaire. A tout moment un administrateur peut vous demander de supprimer vos fichiers, voir **supprimer lui-même les données si l'espace de vient critique suite à plusieurs relances**.

Refonte du WikiHPC et ouverture vers l'extérieur

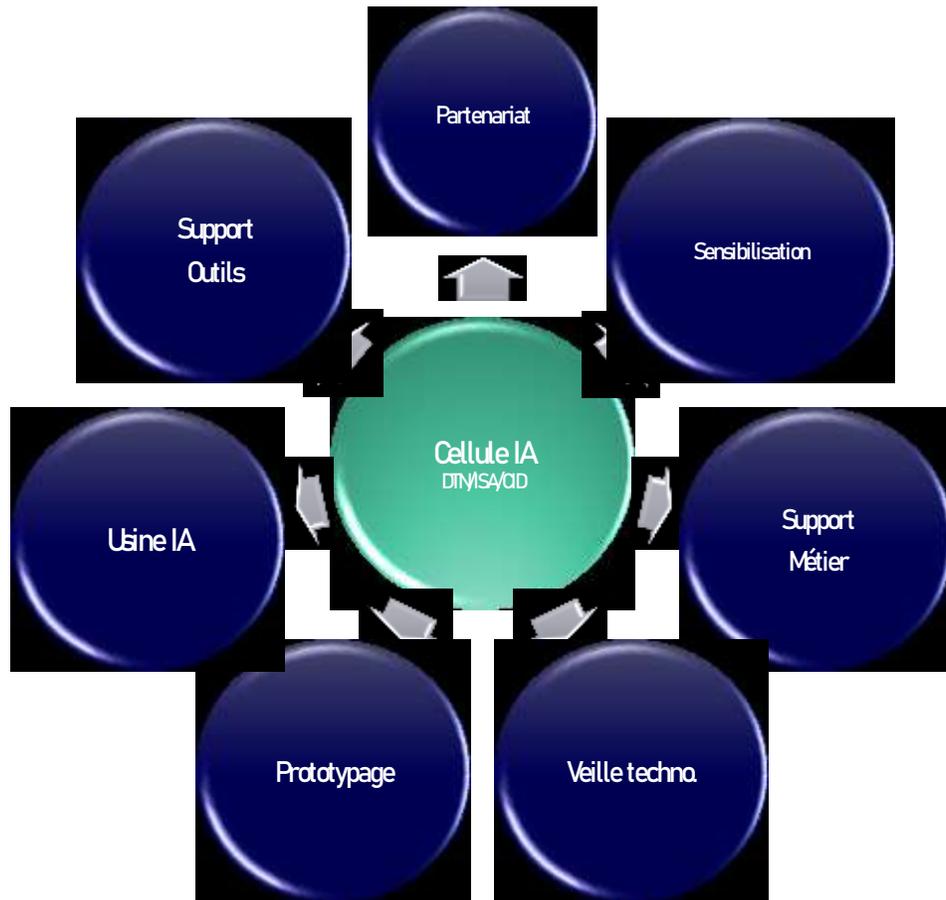


The screenshot displays the CNES HPC 6G website interface. On the left is a dark blue sidebar with the CNES logo and a search bar. The main content area shows the platform status as 'En Service' with a green checkmark icon. Below this, there are three status options: 'Hors Service' (red X), 'Service Dégradé' (yellow warning), and 'En Service' (green checkmark). The central part of the page features a large blue logo for the 'CENTRE DE CALCUL DU CNES'. At the bottom, there is a section titled 'Présentation du Centre de Calcul du CNES' with a list of links: 'Le HPC au CNES', 'Architecture du cluster', and 'Roadmap HPC'.

Une documentation plus accessible et ouverte sur l'extérieur

- Refonte du site WEB avec template Read The Doc
- Accessibilité via Gitlab Pages début 2024
- Ouverture vers l'extérieur à la fin du premier semestre 2024

La Cellule IA



- ❖ La cellule Intelligence Artificielle (IA) de DTN/ISA/CID a été créée en 2017 afin de coordonner les activités CNES autour de l'intelligence artificielle :
 - Sensibiliser les projets et structures du CNES au sujet de l'intelligence artificielle (potentiel et limites)
 - Proposer des services basés sur le support/accompagnement (R&T, stages,...)
 - Prototyper des solutions orientées IA pour la résolution de problèmes métier
- ❖ Grâce à sa vision d'ensemble des activités IA menés au CNES, la cellule peut s'appuyer sur un réseau d'experts répartis dans les différentes sous-directions et structures du CNES afin d'aborder des thématiques transverses (traitement d'images, détection d'anomalies, IA embarquée, ...)
- ❖ La cellule IA s'appuie sur:
 - L'infrastructure de calcul HPC du CNES
 - L'usine logicielle du CNES et l'usine IA
- ❖ La cellule est composée de 4 collaborateurs (3,5 ETP)



<https://confluence.cnes.fr/x/syT6B>

Usine Logicielle du CNES

❖ L'offre de service

- Outils : Gitlab, Gitlab-CI, Jenkins, Sonarqube, Artifactory, Xray



- Documentation : <https://confluence.cnes.fr/display/USINELOG> (dont le [blog](#) associé)

❖ Organisation et activités

- L'UL n'est pas rattachée à l'équipe C3 mais les équipes travaillent en collaboration.
- DTN/ISA/IDA (Infrastructure de l'UL) + DTN/QE/NEO (Processus d'Ingénierie Logicielle).

❖ Support utilisateur sur l'UL

- UsineLogicielle@cnes.fr
- Communication via le blog confluence :
pensez à vous abonner
- Nouvelles demandes « MaVieNumérique »
- Pas de support sur les développements.

Usine Logicielle - Demande de création d'espace	Usine Logicielle - Demande de support	Usine Logicielle - Demandes standards
		
Cette demande permet la création d'espace dans les outils de l'usine logicielle (Gitlab, Jenkins, Artifactory)	Demande de support pour l'utilisation des services de l'Usine Logicielle : Gitlab, Jenkins, Artifactory, Xray, Sonarqube	Cette demande regroupe les actions suivantes : * ajout d'un dépôt distant à Artifactory
Durée de traitement : 3 jours ouvrés métropole	Durée de traitement : 3 jours ouvrés métropole	Durée de traitement : 3 jours ouvrés métropole
Faire ma demande En savoir plus	Faire ma demande En savoir plus	Faire ma demande En savoir plus

Usine Logicielle du CNES

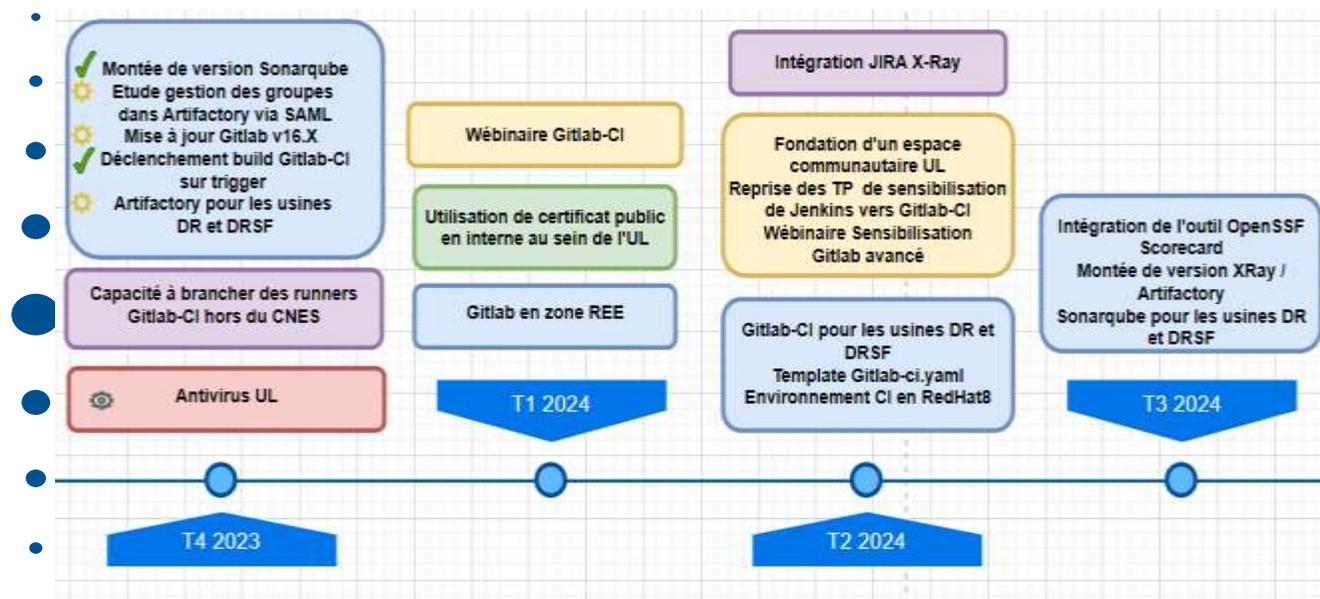
❖ Nouveautés 2023 / Roadmap 2024

- Usines DR/DRSF
- Ouvertures à l'externe
- Montées de version
- Nouvelles sensibilisations / webinaires
- Antivirus UL

❖ Quelques chiffres (01/2021 à 11/2023)

- + de 2200 utilisateurs
- + de 2900 tickets de supports traités
- + de 450 personnes formées

<https://confluence.cnes.fr/display/USINELOG/Roadmap>



Sommaire

Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (documentation et bonnes pratiques)

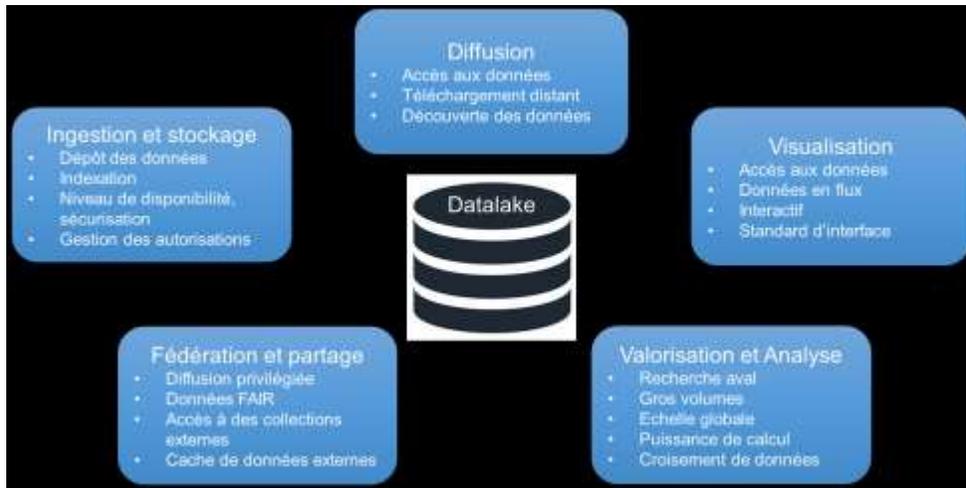
Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie



Le Datalake : Système de stockage de masse pour les données scientifiques

❖ Usages multiples

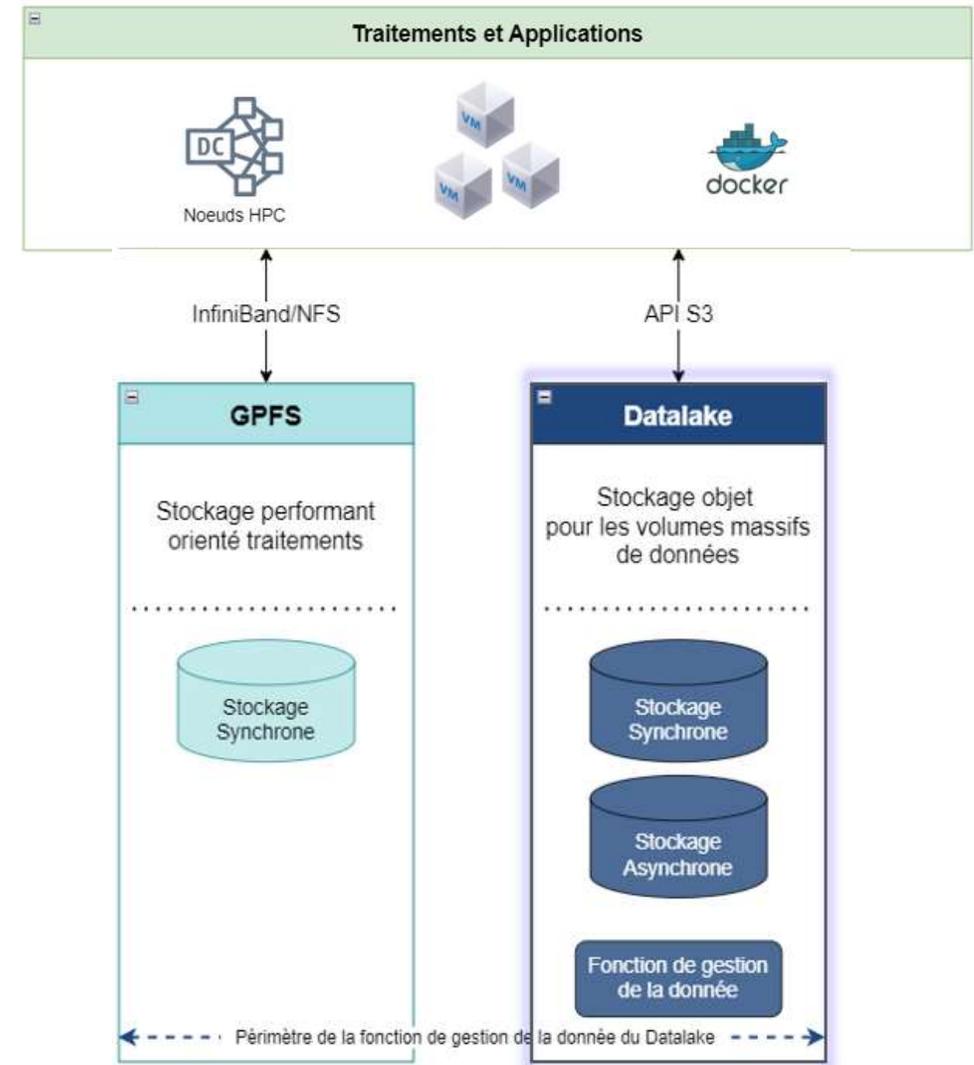


❖ Offre de service Datalake :

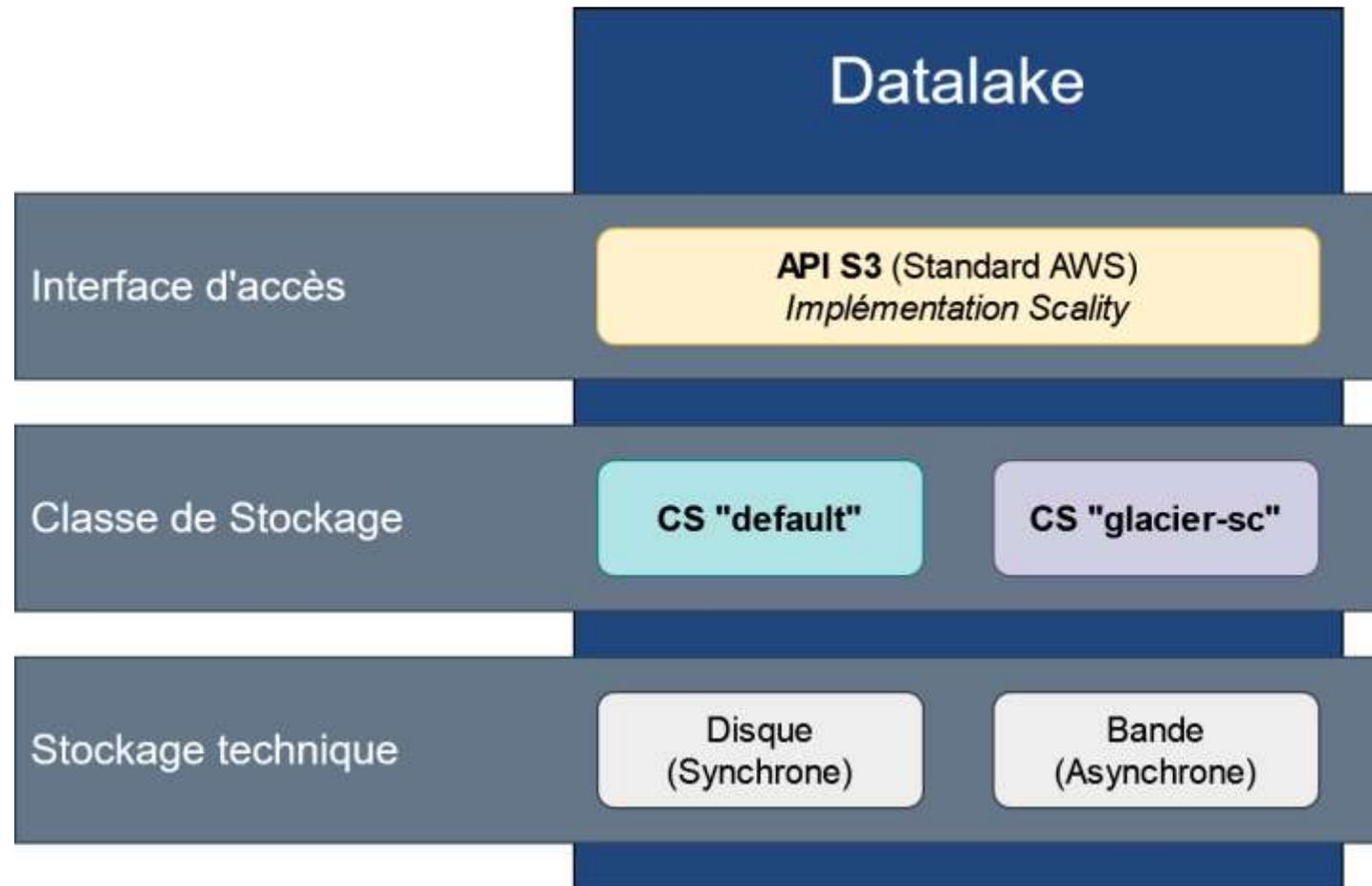
- Système intégré dans l'écosystème du SIS CNES
- Equipe support et accompagnement aux projets
- Données de sensibilité Diffusion Limitée Projet

❖ Tier2 : Opérationnel depuis le **10/11/2022**

❖ Tier3 : Toujours en cours d'intégration. **Cible : T2 2024**



Le Datalake : Différents services



❖ Interface « Simple Storage Service » (S3) :

- API native : S3 – Standard AWS.
- Pas d'interface NFS « native ».

❖ Plusieurs classes de stockage en fonction du besoin :

- « default » (par défaut) : Pour un accès temps réel.
- « glacier-sc » et « glacier-dc » : Plus économique pour des accès moins fréquents.
- Un seul espace de nommage global au Datalake : Accès aux données avec un même "chemin" d'accès quelle que soit la classe de stockage.

❖ Fonctions « Tiering » :

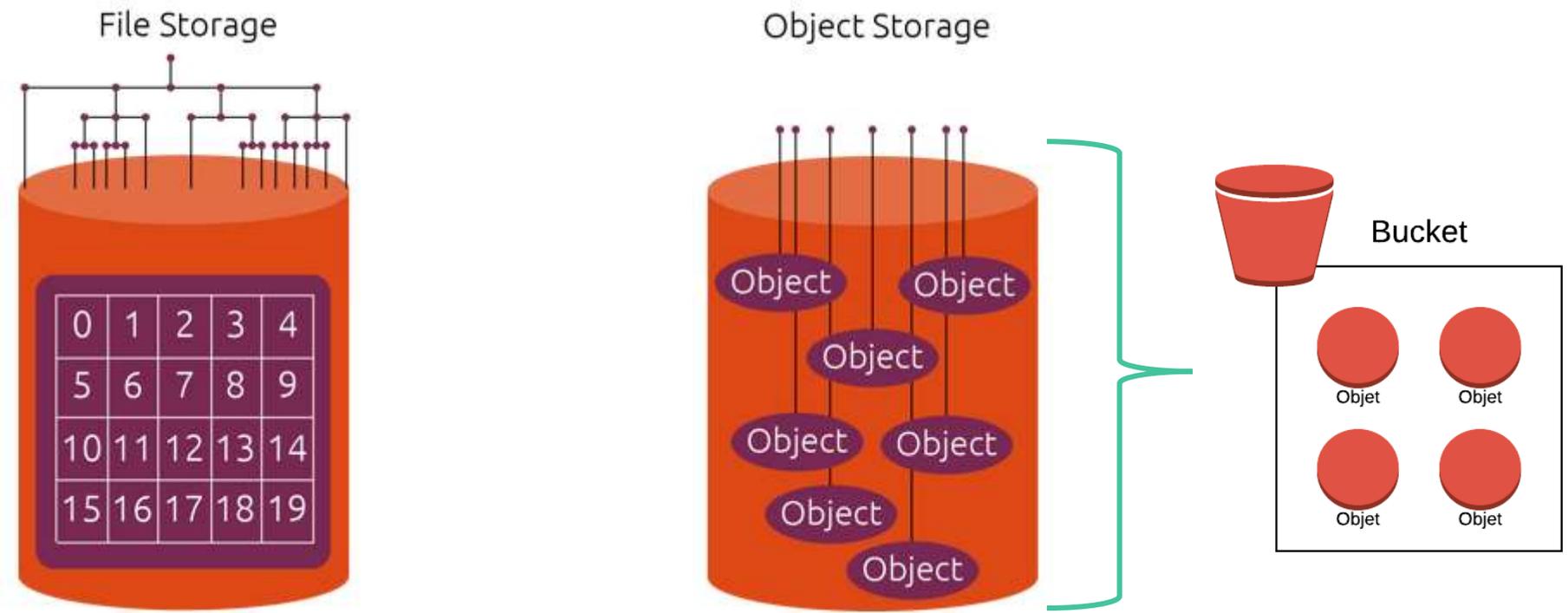
- Gestion du cycle de vie (ILM : Information Lifecycle Management).
- Possibilité de définir des règles de purge / déplacement de données, en fonction de différents critères.

❖ Description détaillée (offre de service):

- [Datalake V2 - Description détaillée - Equipe Fondation Centre de Calcul - Confluence \(cnes.fr\)](#)

Le Datalake : Protocole S3

système de fichiers : POSIX
VS
AWS S3 (Simple Storage Service) : stockage objet



Le Datalake : Protocole S3

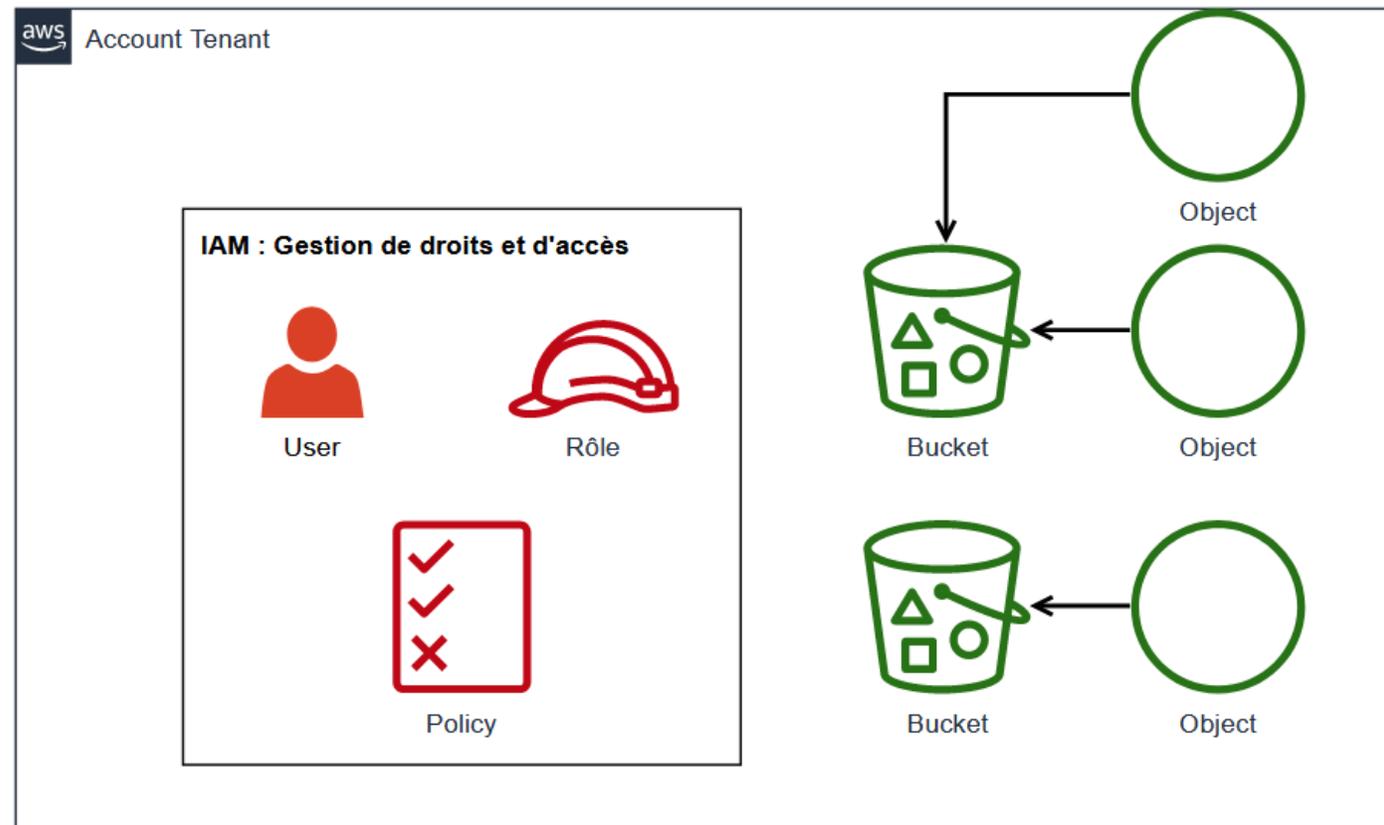
Des nouveaux concepts :

❖ Organisation des données

- Tenant / Projet (Account au sens AWS)
- Buckets à l'intérieur des tenants
- Des objets placés à plat au sein des buckets

❖ Les accès via entités IAM :

- Policy : les politiques d'accès
- User/Group : utilisateurs/groupes IAM authentifiés avec une paire de clé
- Rôle : Entités détenant des droits d'accès à des ressources (bucket/objet) et pouvant être endossé temporairement



Le Datalake : Au sein du SIS

Différents usages et modes d'authentification :

❖ Producteurs des données : projet

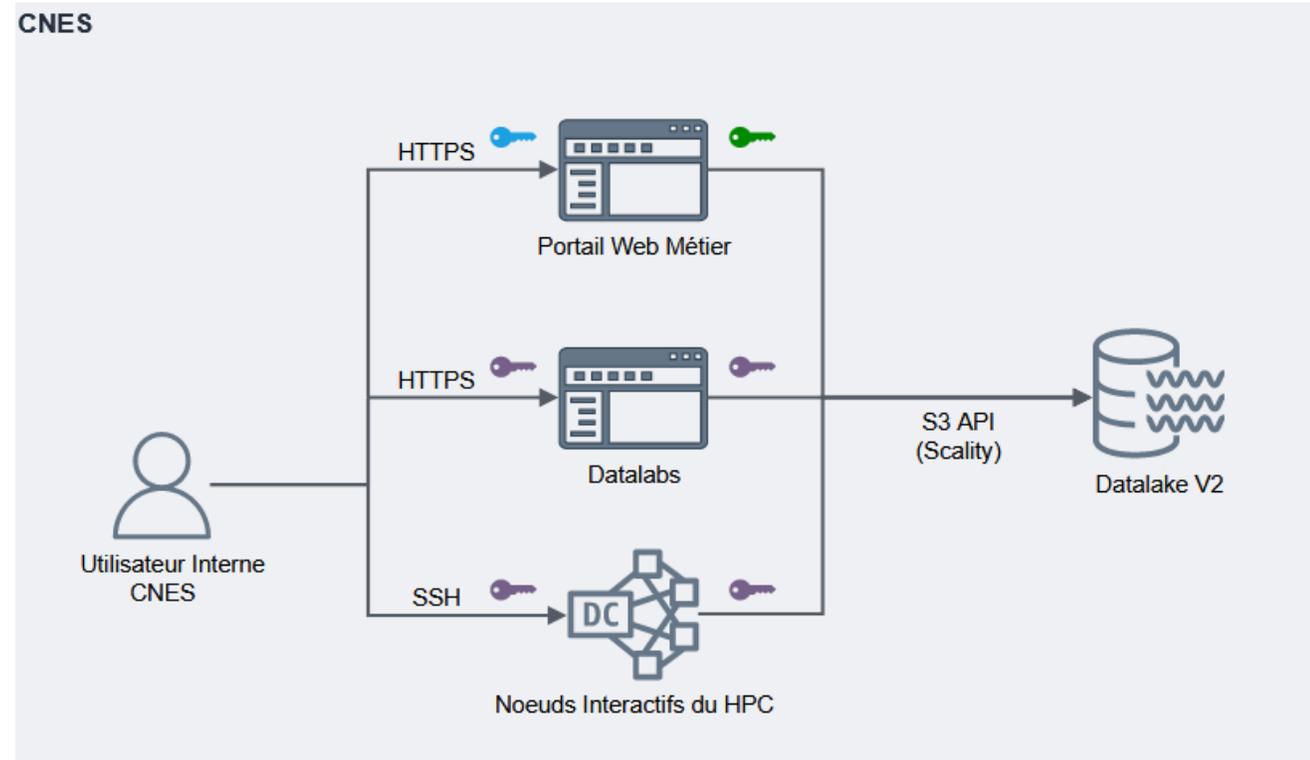
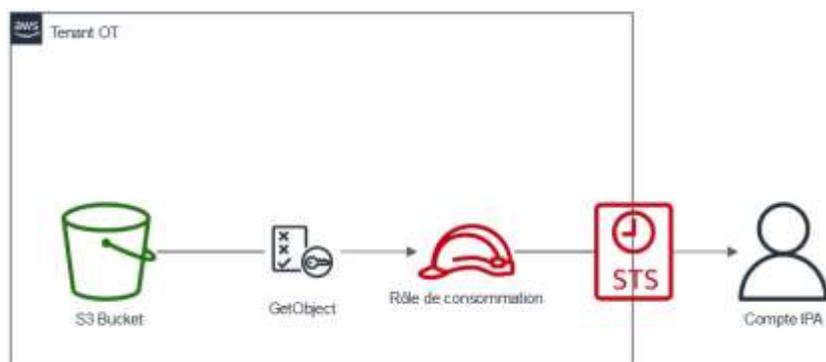
➤ IAM Users :

- Access key (ex : ERTL1OCG46A721LYAGH6)
- Secret key (ex : H42G6Q5yVsT2VerlYw0NHPx8x6BClr0VvUIMNkpO)

❖ Consommateurs de données :

➤ Compte IPA

- En endossant un rôle
- Des credentials (paire de clé + token) tmp



Légende :

-  Utilisateurs gérés par l'application métier. Inconnu du Datalake V2.
-  Utilisateurs gérés par l'annuaire IPA CNES. A accès aux services du SIS. Utilisateurs disposants des groupes IPA permettant un accès au Datalake V2.
-  Utilisateurs techniques disposants des groupes permettant un accès au Datalake V2.

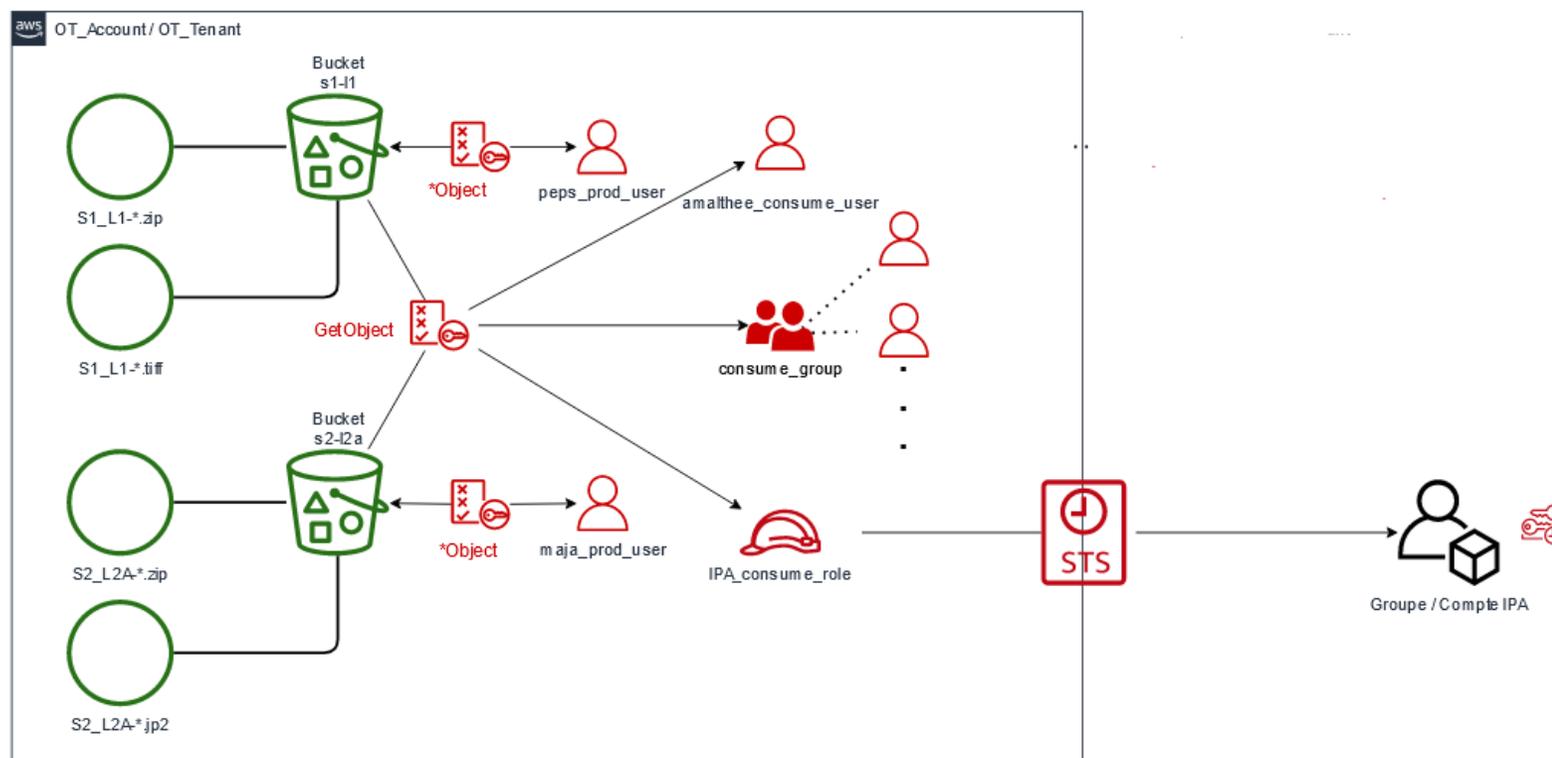
Le Datalake : demande d'accès

❖ Producteurs/Projets des données : User IAM

- Demande de mise en place de Account/bucket, par mail (L-SIS-Datalake@cnes.fr)
 - Account/tenant cible
 - La liste de buckets avec la gestion des droits :
 - Read-only, Read-Write
 - Savoir comment organiser ses données : prefix, tags, metadata

❖ Consommateurs de données

- Compte IPA
 - Gestion de permissions pour endosser un rôle
 - Groupe IPA



Le Datalake : Manipuler les données

Le Datalake respecte pour les APIs les plus répandues, le standard AWS.
La [documentation AWS S3](#) est une excellente base pour commencer !

❖ Utiliser des clients S3 qui permettent de manipuler les données

- Des APIs haut-niveaux :
 - Simuler les commandes POSIX (ls, cp ...)
 - Uploads/downloads partitionnés (MPU et range request)
- Des APIs bas niveaux : get-object, put-object, head-object, list-object

❖ Prendre en compte les caractéristiques du stockage objet

- très performant en lecture mais comparativement peu performant en écriture.
- organisation des données : metadata personnalisable, tags ...
- formats de données plus ou moins adaptés (en terme de performance) pour le S3

SDKs



Android



iOS



Java



JavaScript



.NET



Node.js



PHP



Python (boto)



Ruby



Xamarin



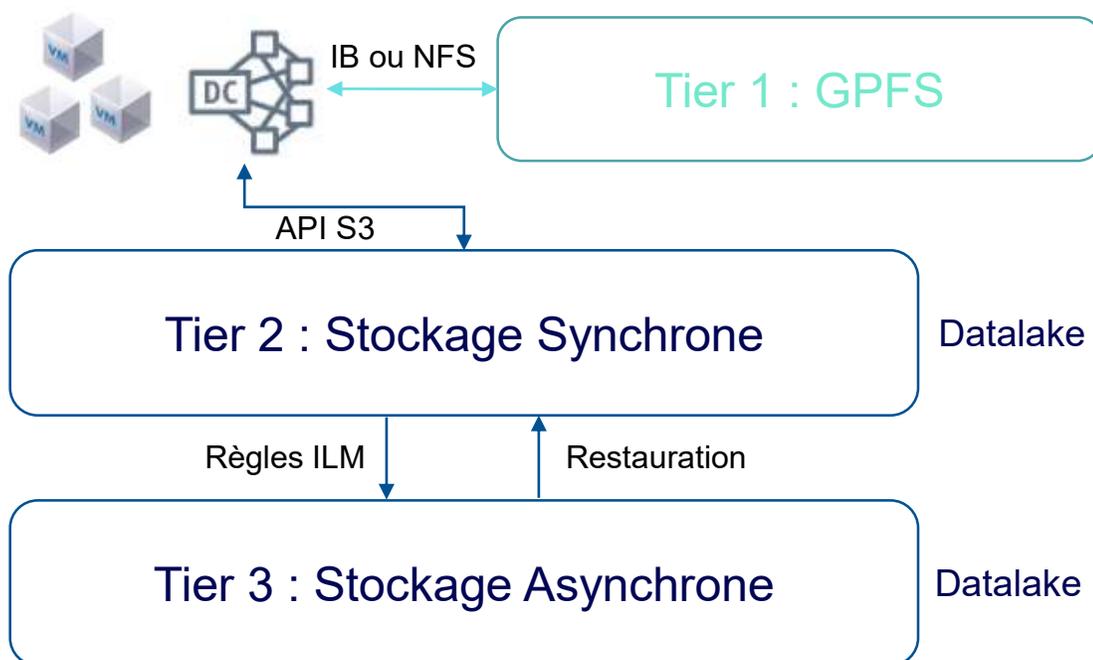
AWS CLI

AWS Toolkit
for EclipseAWS Toolkit
for Visual StudioAWS Tools
for Windows
PowerShell

Le Datalake : Manipuler les données avec l'arrivée du Tier 3

Le Tier 3 :
D'autres concepts et une accès à la donnée de manière asynchrone !

Ressource du SIS



- Aucune interaction directe avec le Tier 3 (passage par le Tier 2 nécessaire)
- Deux types d'interactions :
 - Tier 2 vers Tier 3 : **Transitions** actionnées par des ILMs sans actions utilisateurs
 - Tier 3 vers le Tier 2 : **Restaurations** à la demande utilisateurs.

Le Datalake : Rappel important

Tout utilisateur (avec un compte IPA) a accès au Datalake et peut endosser des rôles selon les autorisations attribuées à son compte

Projet

Faire la demande d'Account/Bucket, réfléchir à l'organisation des données, formats ... : L-SIS-Datalake@cnes.fr

+ Info sur Confluence :

<https://confluence.cnes.fr/pages/viewpage.action?spaceKey=EFCALC&title=Datalake+V2+-+Principe+de+fonctionnement>



Sensibilisations dédiées :

<https://gitlab.cnes.fr/hpc/wikiHPC/-/wikis/Sensibilisations-et-formationen#introduction-datalake>

① Pour toutes questions ou demandes de support, n'hésitez pas à contacter : SIS-supportHPC@cnes.fr

Le Datalake : Rappel important

Tout utilisateur (avec un compte IPA) a accès au Datalake et peut endosser des rôles selon les autorisations attribuées à son compte

❖ Producteurs/Projets des données : Tests en amont des phases de développement

- Demande par mail (L-SIS-Datalake@cnes.fr)
 - Account de tests avec des buckets déjà créés

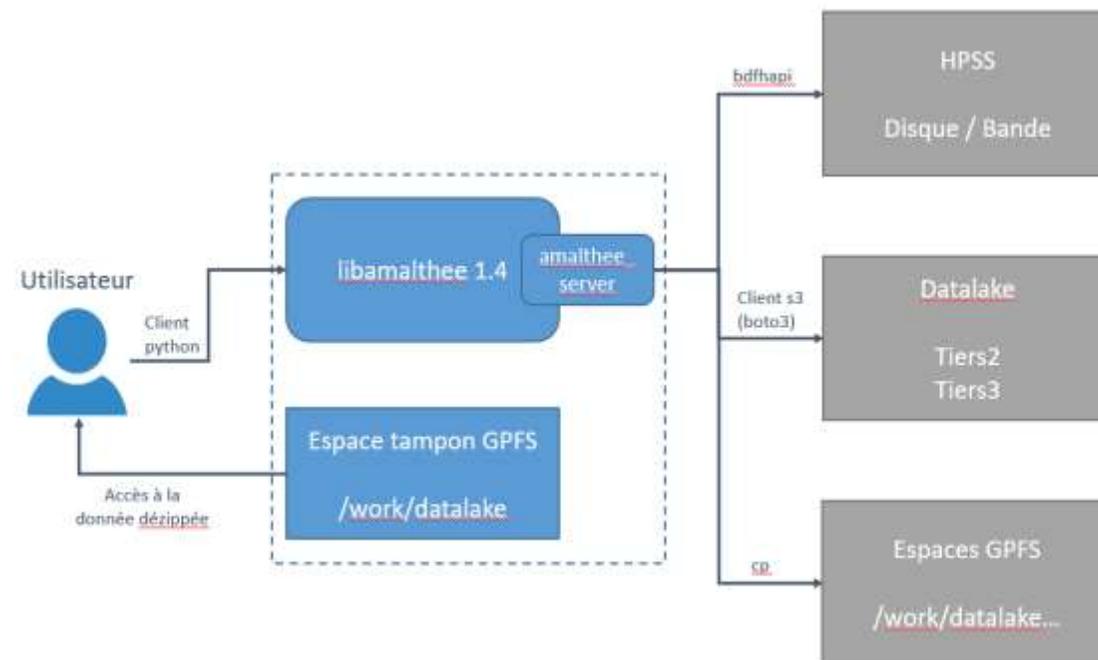
❖ Consommateurs de données

- Rôles publiques disponibles à tous les comptes IPA (pas de demandes particulières) :
 - Données OT (sentinel, HII ...)
 - Données SWOTCE (zcollection)
- Des outils pour la manipulation des rôles disponible sur TREX :
 - `/softs/tools/datalake/roles_list.sh`
 - `/softs/tools/datalake/use_role.sh`

① Pour toutes questions ou demandes de support, n'hésitez pas à contacter : SIS-supportHPC@cnes.fr

Amalthee : interface python pour le Datalake

- ❖ Bibliothèque Python qui facilite la mise à disposition des produits PEPS et THEIA. Outil étroitement lié à la plateforme HPC du CNES et son espace d'écriture /datalake
- ❖ Prestation Thales SI pour le développement et le maintien de l'outil
- ❖ Permet la transition entre les anciens et nouveaux moyens de stockage (HPSS/GPFS → Datalake) – transparent côté utilisateur
- ❖ Permet de lier des requêtes par critères à des chemins d'accès à la donnée dans le datalake le temps que les futures plateformes OT (GDH) ingèrent l'intégralité des collections PEPS & THEIA.
- ❖ V 1.4 compatible TRES RH7 RH8 disponible en qualification depuis 09/23 et en production depuis 11/23



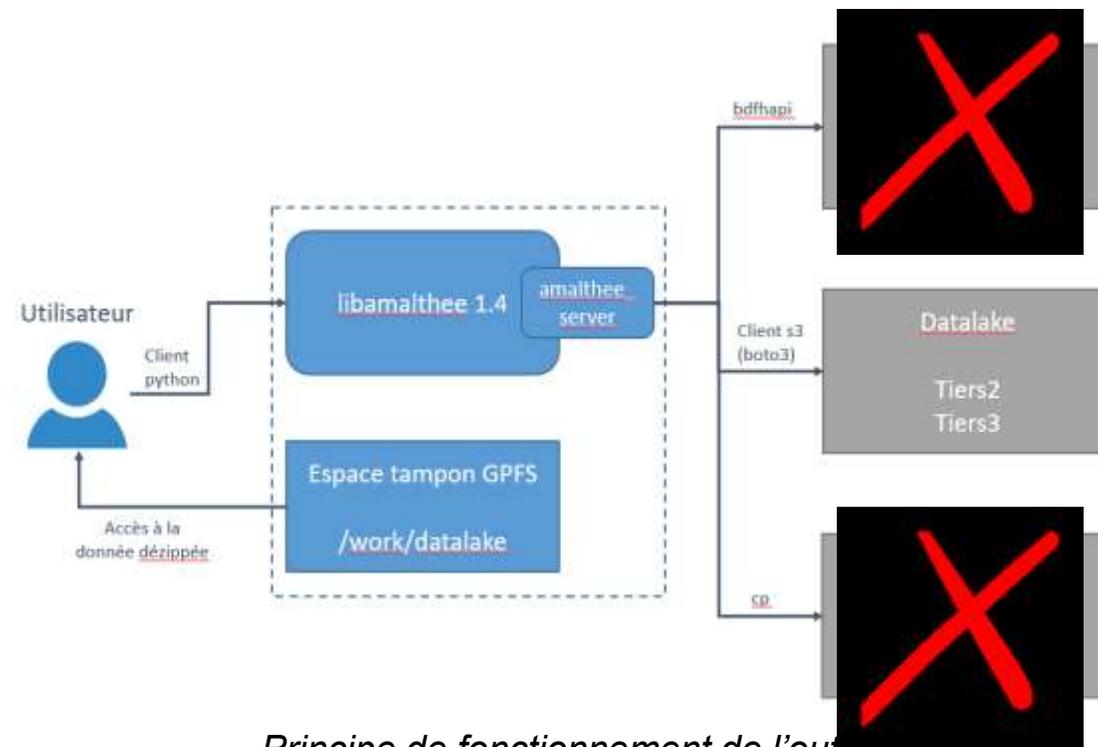
Principe de fonctionnement de l'outil Amalthee

Une documentation est disponible ici :

<https://gitlab.cnes.fr/datalake/libamalthee/-/wikis/home>

Amalthee : interface python pour le Datalake

- ❖ Bibliothèque Python qui facilite la mise à disposition des produits PEPS et THEIA. Outil étroitement lié à la plateforme HPC du CNES et son espace d'écriture /datalake
- ❖ Prestation Thales SI pour le développement et le maintien de l'outil
- ❖ Permet la transition entre les anciens et nouveaux moyens de stockage (HPSS/GPFS → Datalake) – transparent côté utilisateur
- ❖ Permet de lier des requêtes par critères à des chemins d'accès à la donnée dans le datalake le temps que les futures plateformes OT (GDH) ingèrent l'intégralité des collections PEPS & THEIA.
- ❖ V 1.4 compatible TRES RH7 RH8 disponible en qualification depuis 09/23 et en production depuis 11/23



Principe de fonctionnement de l'outil Amalthee

Une documentation est disponible ici :

<https://gitlab.cnes.fr/datalake/libamalthee/-/wikis/home>

Amalthee : interface python pour le Datalake

❖ Quelques statistiques :

HAL (Depuis 03/2023 : passage datalake)

- PEPS :
 - 143114 produit différents (186748 demandés)
 - **Derniers 30 jours :**
 - **6312 produits**
 - **11 utilisateurs**
- THEIA :
 - 108436 produits différents (142538 demandés)
 - 5 principaux utilisateurs
 - **Derniers 30 jours :**
 - **7388 produits**
 - **4 utilisateurs**

TREX - production (11/2023)

- PEPS :
 - 4019 produits différents
 - 9 utilisateurs différents
- THEIA :
 - 3907 produits différents
 - 3 utilisateurs différents

❖ Evolution moyen/long terme

- Outil utile tant qu'il n'existe pas de catalogue permettant de récupérer le chemin d'accès à la donnée sur le Datalake
- Peu utile à moyen terme avec l'arrivée de GDH, la possibilité de requêtes partielles, stockage en mode dézippé COG/ZARR/autre sur le datalake
- Intégration du Datalake en tant que plugin S3 dans **eodag** pour les utilisateurs du SIS ?
- Autres ?

Sommaire

Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

Le Datalake et initiatives liées (planning, documentation, bonnes pratiques)

Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie



K8S

DES APIS ET TECHNOLOGIES STANDARDS



Kubernetes



UI

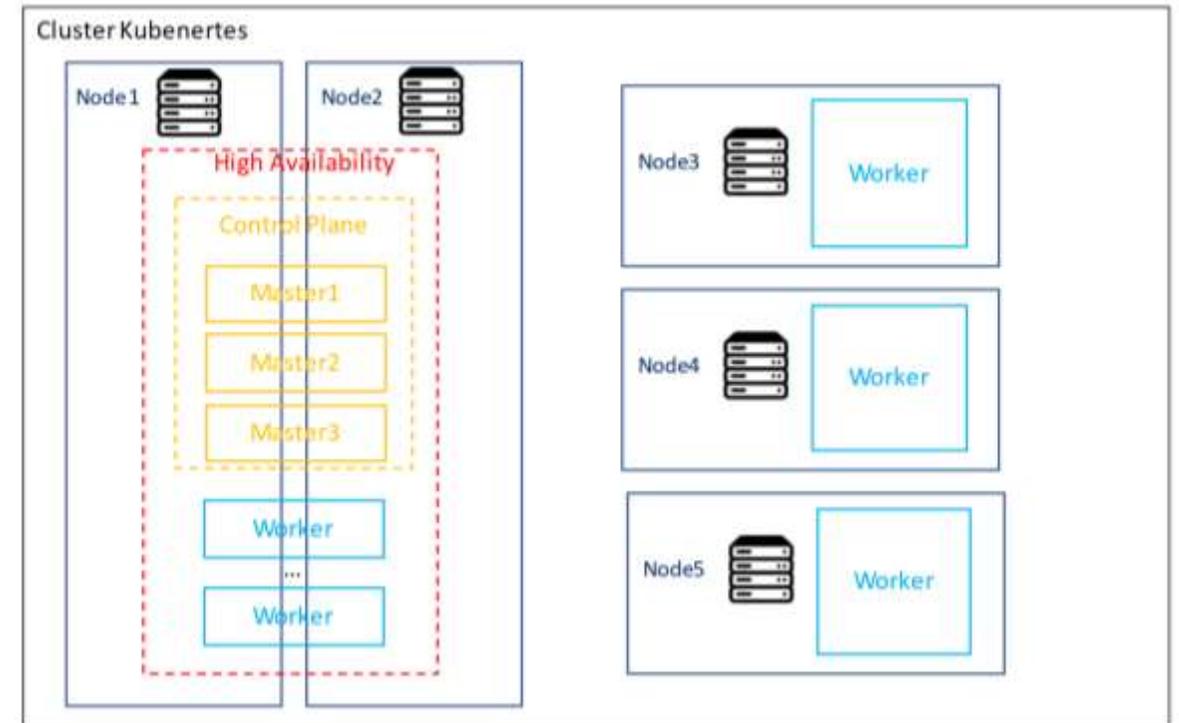
DES NOUVEAUX CONCEPTS

Une encapsulation projet par namespace avec des politiques de RBAC

Des quotas de ressources attribués par namespace

Un accès à la donnée plus restreint

UN CLUSTER K8S MUTUALISÉ



5 noeuds :
AMD EPYC Milan 7713, 64C (2.0GHz-225W) avec
64 cœurs et 128 Go de RAM

Métrologie et monitoring orienté utilisateurs

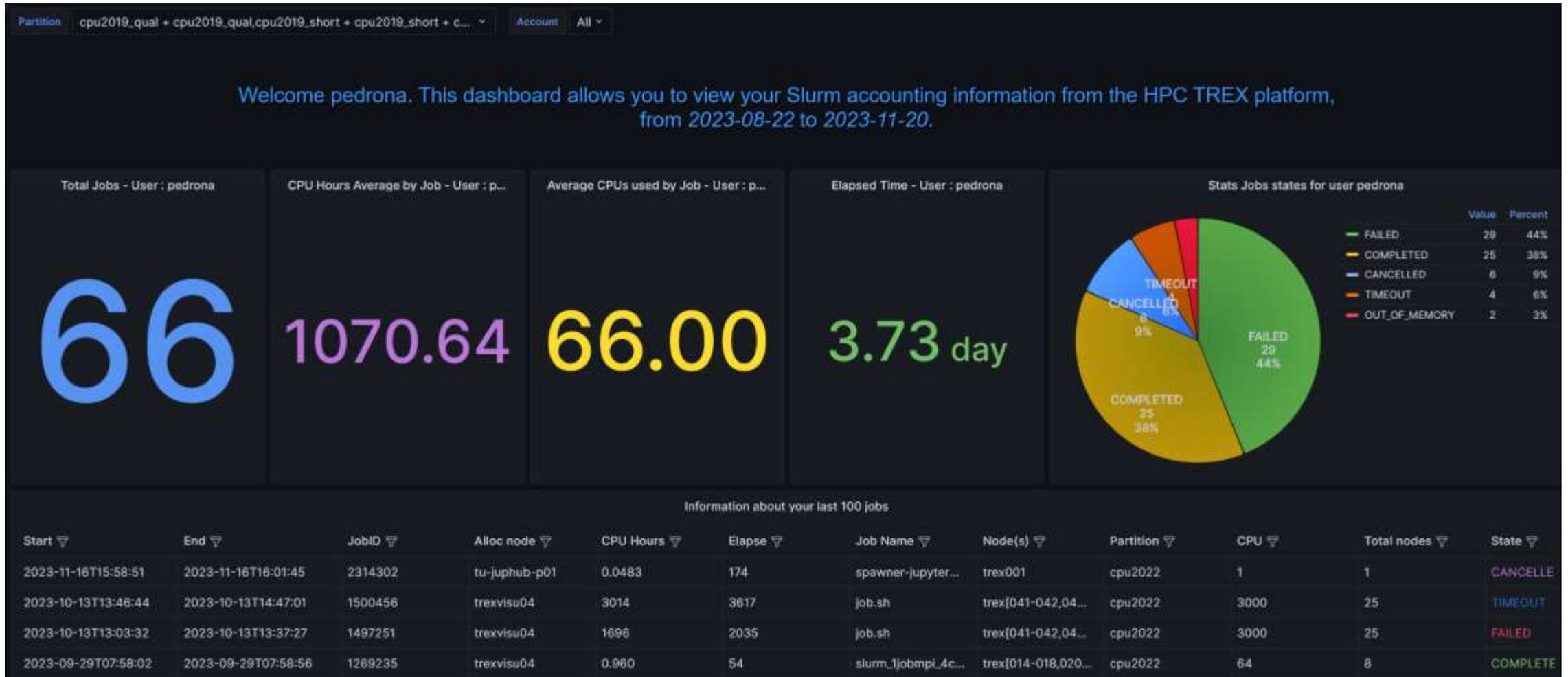
<https://trexmetro.sis.cnes.fr> (accessible depuis un PC CNES ou une session de visualisation)

Une vision d'ensemble de la plateforme :

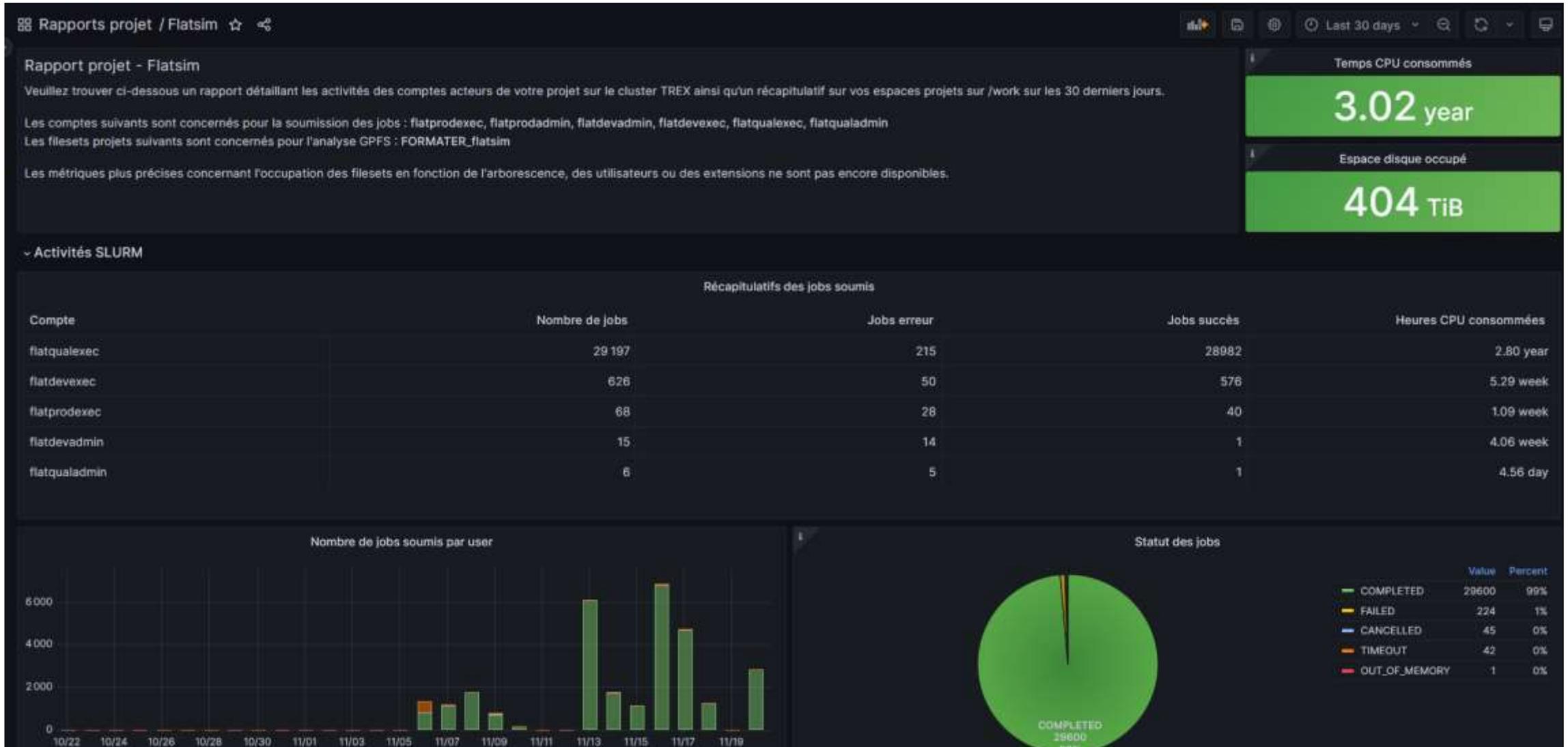


Métrologie et monitoring orienté utilisateurs

Une vue spécifique de vos jobs Slurm :



Métrologie et monitoring orienté projets (accounting Slurm)



Métrologie et monitoring orienté projets (utilisation /work)



Sommaire

Centre de calcul et ses Équipes

Bilan annuel d'activité et d'incidents sur les plateformes

- HPC5G (HAL & KTULU)
- HPC6G (TREX & ARYX à venir)
- Datalake (tier 2)
- Retours d'expériences HPC5G vers HPC6G

Les services avancés

- Le Datalabs
- La conteneurisation pour le calcul
- Les logiciels
- Le support utilisateur et expertise de code,
- Les sensibilisations
- La forge et l'Usine Logicielle

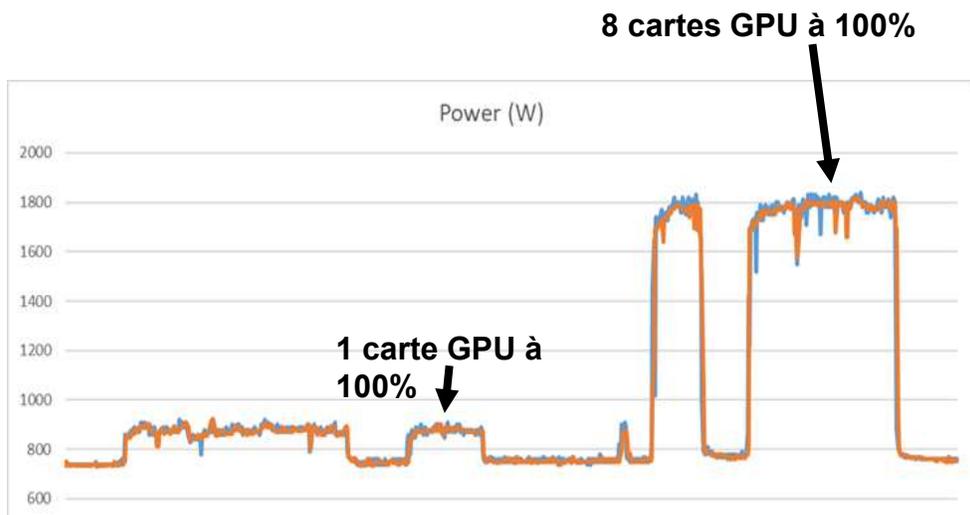
Le Datalake et initiatives liées (planning, documentation, bonnes pratiques)

Les nouveaux services HPC6G :

- Kubernetes
- Monitoring
- Energie



Consommation électrique : mesurer



Exemple de consommation instantanée sur le nœud gpgpu07 avec 8 cartes GPU A30

Idle (au repos) : 750 W

1 GPU à 100% : max 900W

8 GPU à 100% : max 1840 W

Energy efficiency and Power management

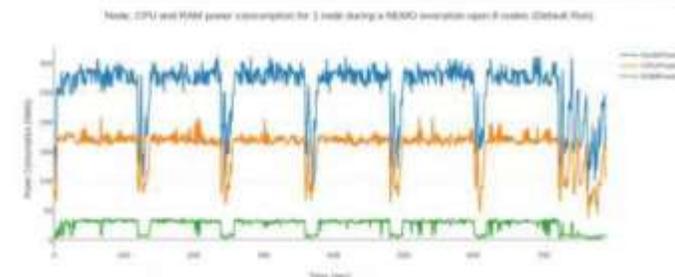
BEO (Bull Energy Optimizer) and Slurm Plugin

► Infrastructure monitoring based on BEO

- BEO allows to follow power or energy of each:
 - Compute node (IPMI)
 - Switch (Ethernet + IB)

► Energy job profiling achieved through SLURM

- Energy consumed by a job collected for:
 - Compute node
 - CPUs
 - RAM



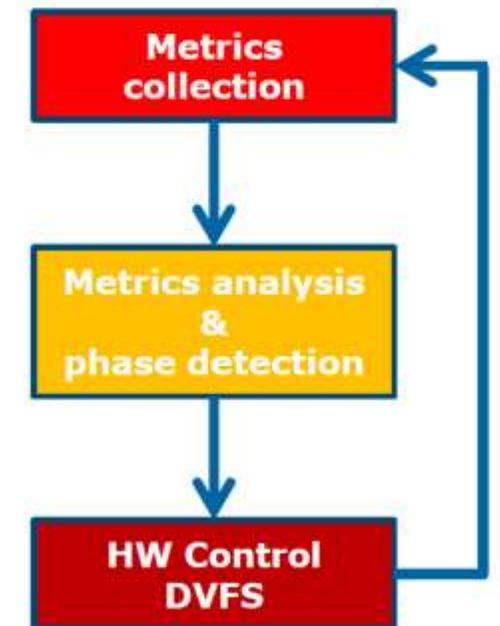
Outillage pour HPC6G BEO/BDPO

Mesurer la consommation électrique du Centre de Calcul du CNES

=> Offre de stage Master 2024 publiée sur le [site du CNES](https://www.cnes.fr)

Consommation électrique : optimiser

- **Optimiser les codes de calcul : c'est de loin le point où il y a le plus de marge de manœuvre**
 - Coder de manière efficace, utiliser des bibliothèques optimisées, se poser la question du langage et des algos, ...
- Mesurer l'énergie consommée et adapter la configuration des ressources au besoin
 - Etude CERFACS/CNES 2022 sur l'outil EAR pour optimiser l'énergie en adaptant selon l'application la fréquence des CPU, de la mémoire et des GPU
 - ⇒ gain conso espéré dans les cas favorables : ~10% pour des temps de restitution qui augmentent de 2%
 - HPCG 6G : mise en place de la suite d'outils BULL : BEO (mesure) / BDPO (optimisation)
- Porter les codes sur des solutions moins énergivores
 - HPC6G : choix des processeurs AMD
 - efficacité énergétique X2 par rapport aux processeurs Intel
 - Veille pour portage d'applications sur les nouveaux matériels
 - GPU Intel Ponte-Vecchio, nécessite de transformer les directives OpenACC en OpenMP qui n'est pas encore très mature
 - GPU AMD (AMD Mi250) – supercalculateur du CINES
 - ARM + GPU Nvidia (Tests du projet LISA sur la machine Turpan à CALMIP)
 - FPGA



Retour JCAD 2023

- Présentation CNES du Centre de Calcul et description des nouvelles plateformes TREX et Datalake aux JCAD 2023
Programme et planches : <https://jcad2023.sciencesconf.org/resource/page/id/9>

Points d'intérêt pour le CNES :

- **Réseau MésoNet** pour accéder à des machines de tests + des machines « codes formation » :
 - Quantique : machine à Reims (ROMEO) + organisation de formation Quantique en régions
 - Vectorielle Boréale : Rouen au CRIHAN
 - ARM/GPU Nvidia : Turpan à Toulouse
 - OpenStack Grenoble CINORA
 - Strasbourg GPU AMD (MI*)Connexion avec l'identité EduGAIN (des membres du projet LISA ont réussi à se connecter à la machine Turpan à CALMIP depuis le Centre de Calcul du CNES)
Des axes de travaux commun (Datalake, Fédérations de données, identité, Hébergement Datacenter, Backbone Renater, CPU Européen,...)
- **Ateliers de la Donnée de Recherche Data Gouv**
 - Des ateliers de la donnée sont labellisés dans toute la France. A Toulouse, il y a l'atelier ADOO : trajectoire de labellisation, fairisation, valorisation des données : porté pour le moment par CALMIP + Sciences de l'Homme
 - Plan Science Ouverte v2 (PNSO) : Surement des perspectives pour les usages du Centre de Calcul (Politique de logiciels libre, Archivage universitaire,...)

Retour JCAD 2023 (suite)

➤ GENCI

- Projet CLUSSTER pour l'IA avec pour objectif d'adresser à la fois la recherche ouverte, confidentielle et à but lucratif
- QLM au TGCC : Pasqal disponible fin 2023 (possibilité de demander des heures sur eDARI pour le simulateur Quantique)
- La France a obtenu un futur ordinateur européen qui sera installé au TGCC, l'objectif est d'attendre la disponibilité du processeur européen Rhea (Sipearl) pour l'installer.

➤ Production de logiciels

- Rapport MESR sur les forges logicielles à paraître le 29 novembre 2023
- A part la Forge in2p3 ouverte à toute la fédération d'identité EduGAIN, les autres sont réservées à leur institution
Difficulté à partager des codes, du coup, tout le monde utilise github 😞!



Des questions ?

L-sis-poleHPC@cnes.fr

Présentation disponible ici : Confluence