Plateforme VirtualData : cloud et hébergement

Michel Jouvin, michel.Jouvin@ijclab.in2p3.fr

Inauguration du mésocentre Paris Saclay

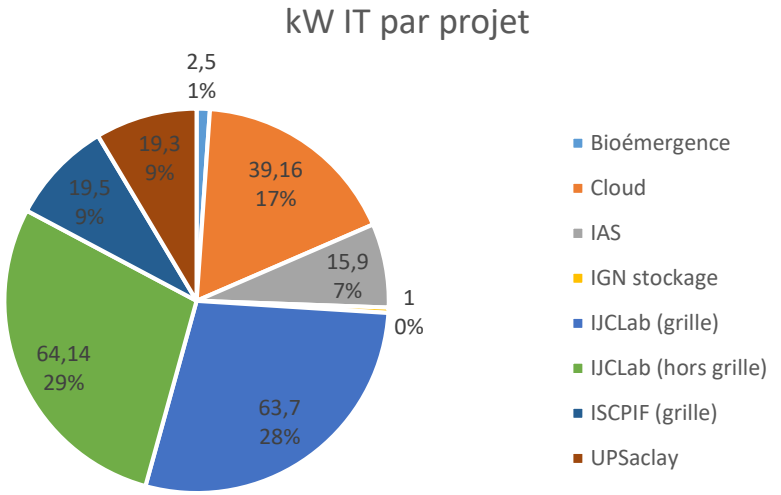
6 décembre 2022

Au commencement, le datacenter

- Initiative des laboratoires du Labex P2IO en 2011
  - P2IO : Physique des 2 Infinis et des Origines
  - 5 laboratoires devenus IJCLab en 2020 + IAS
  - Objectif : construire un datacentre centré sur l'efficacité énergétique (PUE = 1,25) pour héberger l'informatique de nos laboratoires
  - Localisé au bat. 206 (vallée) : opérationnel depuis octobre 2013
- Infrastructure modulaire pouvant être étendue jusqu'à 90 racks et 1,5 MW IT
  - Actuellement 51 racks [~20 libres) déployés, 600 kW IT, double alimentation 300 kW
  - Possibilité d'héberger des services critiques : secours groupe électrogène pour 80 kW (10 racks)
- Héberge des ressources mutualisées et des ressources spécifiques à un établissement/laboratoire
  - Toute l'informatique d'IJCLab, une partie de l'informatique de l'IAS
  - DSI Paris Saclay, AgroParisTech, CentraleSupelec
  - Noeud de la grille EGI/WLCG : principalement experience LHC + Institut des Systèmes Complexes (ISCPIF)
  - Cloud pour l'informatique scientifique

Hébergement : utilisation

- Electricité totale = 225 kW IT (24h/24)
  - ~280 kW avec la climatisation (2,5 MWh/an) soit 345 k€ (0,145€/kWh, + 16%)
- Haute disponibilité
  - Double alimentation : 120 kW sur 300 kW
  - Secours groupe électrogène : 40 kW sur 80 kW
- Croissance des ressources UPSaclay hébergées hors cloud (5 racks)
- Beaucoup de (très) vieilles machines : arrêt ou remplacement devient un enjeu majeur
  - IJCLab en cours : ~40 kW IT (17%, 60 k€) représentant une fraction marginale en termes de puissance CPU
  - Favoriser les ressources mutualisées : meilleur taux d'utilisation



- 1 kWh : 100 gCO2 (ADEME)
- PUE ~1,25 →1 kW IT = 1,25 kW réel

Cloud VirtualData…

- Infrastructure de calcul virtualisée : permet de répondre à des besoins très variés…
  - L'utilisateur acquiert dynamiquement les ressources dont il a besoin au sein des ressources partagés : régulé par des quotas
  - L'utilisateur contrôle complètement l'environnement d'exécution : OS (Linux), librairies…
  - Possibilité de données persistant aux arrêts/redémarrages de VMs
  - Possibilité d'orchestrer le démarrage de plusieurs VMs coopérant entre elles
- … mais pas de parallélisme impliquant plusieurs machines
  - Pas de connexion faible latence (Infiniband,Omnipath) entre les machines
- Une ressource mutualisée conséquente au service de l'informatique scientifique de Paris Saclay opéré par IJCLab
  - Basé sur OpenStack, démarré en 2016 dans le cadre de l'Université Paris Sud
  - Actuellement 12000 cœurs associés à un stockage permanent de 1 PB utile : 2/3 ont moins de 2 ans (serveur de 256 cœurs/512 GB mémoire, processeurs AMD 7702 2 Ghz)
  - 4 serveurs avec 40 GB de mémoire par cœur (soit 1,7 TB/serveur au lieu de 2 GB/cœurs pour les autres) pour l'astrophysique et la cosmologie

… Cloud VirtualData

- Mode d'utilisation basique : création/exécution d'une machine virtuelle pour exécuter une application
  - D'autres modes d'utilisations possibles à travers des services avancés
  - Possibilité de construire des clusters de containers (Kubernetes, Swarm)
  - Durée des VMs peut être de plusieurs mois : l'utilisateur doit confirmer périodiquement qu'il souhaite prolonger l'utilisation de sa VM pour permettre le recyclage des ressources
- Point fort : traitement de grandes masses de données
  - Possibilité de traiter en parallèle les données avec un grand nombre de ressources
  - (potentiellement) pas de restriction dans la connectivité externe
- Prise en charge de programmes parallèles nécessitant moins de 256 cœurs
  - Insuffisant pour les grandes simulations
- Pas (encore) de GPUs : 1 projet prévoit d'en acquérir prochainement
- Financement depuis 2016 : laboratoires/utilisateurs, région IdF (DIM), Paris Sud/Saclay…
  - Capacité d'accroitre les ressources au fil de l'eau en fonction des besoins et des opportunités

Stockage

- Plateforme de stockage VirtualData basée sur la technologie Ceph
  - Cluster de stockage dont la haute disponibilité et la performance reposent sur une forte distribution et la réplication des données
  - Fournit différents types de stockage : file system (~NFS), disques virtual (~iSCSI) ou de l'object storage (S3)
  - Actuellement : 1 PB utile
- Principalement utilisable à travers le cloud
  - Pas destiné à faire de l'archivage de long terme
  - Pas de backup (par défaut)
  - Pas de mutualisation possible des octets…
- Actuellement, utilisation principale = disque non volatile des VMs du cloud
  - Quelques utilisations de S3 pour l'import/export de gros volumes de données
- Technologie retenue pour la future plateforme de stockage globale du mésocentre

Services avancés : Spark

- Service pour le traitement de grand volume de données avec le paradigme map-reduce
  - Très grande scalabilité par ajout de ressources tant que l'essentiel du traitement peut être fait en // sur des sous-ensembles de données (map) et que la consolidation du résultat reste courte (reduce)
  - Basé sur le framework Apache Spark, très utilisé dans le monde du big data : ecosystème très actif
- Un service potentiellement ouvert à des groupes d'utilisateurs distincts
  - Actuellement, utilisateur principal est le projet Fink, un « broker » d'évènements astrophysiques pour le Vera Rubin Observatory (voir l'exposé de Julien Peloton)
  - Démarrage dynamique des worker nodes en fonction de la demande
- Projet Astrolab : développements basés sur Spark pour l'astrophysique
  - https://astrolabsoftware.github.io

Services avancées : JupyterHub

- Un service ouvert à tous pour l'exécution et le partage de notebooks Jupyter
  - Notebook : « document computationnel » qui permet de mixer du code, du text (ex: documentation) et des images (ex: graphiques produit par le code du notebook)
  - S'est fortement développé dans le monde de l'analyse de données et de l'enseignement
  - Plusieurs langages supportés : Python, R, C++ (de C++ 11 à C++ 20)
  - Authentification intégrée avec eduGAIN mais demander l'autorisation à info-scientifique.di@unviersite-paris-saclay.fr
- Configuration actuelle : 1 VM de 256 cœurs / 512 GB de mémoire
  - Travail en cours pour un provisioning dynamique des workers dans un cluster de containers, en fonction de la demande
  - Les environnements JupyterHub et JupyterLab supportés
- Fortement utilisé par l'enseignement Paris Saclay
  - 20 UE, ~50 enseignants, O(10000) étudiants en 2021-22 (https://jupyterhub.ijclab.in2p3.fr/dashboard.html)

Batch System

- Pour les besoins scientifiques, souvent le souhait de pouvoir lancer une/des applications en mode non interactif mais sans avoir à configurer/maintenir le système d'une VM
  - Mode d'utilisation classique d'un cluster de calcul
- Une solution basée sur HTCondor déployée dans le cloud VirtualData
  - 1 instance privée IJCLab connectée aux espaces locaux des utilisateurs
  - 1 instance pour les utilisateurs extérieurs nécessitant l'import/export des données
- Encore à un stade préliminaire : intéressé par connaitre les besoins utilisateurs pour mieux définir les évolutions futures du service
- Axes d'évolution planifiés
  - Provisioning dynamique des ressources en fonction des jobs en attente pour une meilleure optimisation de l'utilisation du cloud
  - Déploiement du service CVMFS pour permettre le déploiement des logiciels par les utilisateurs indépendamment de la soumission des jobs

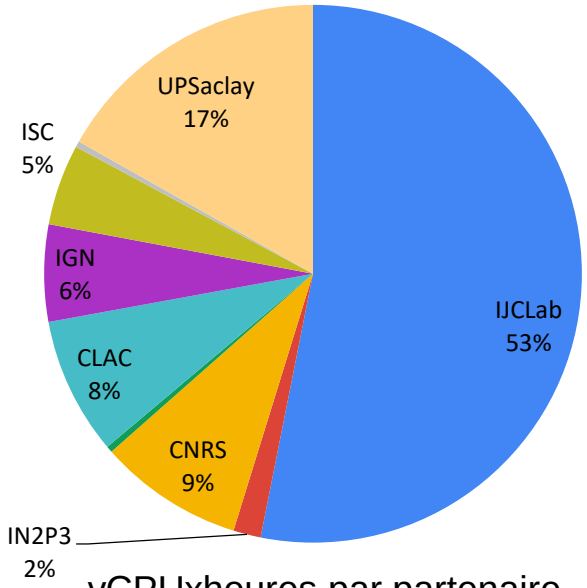
Utilisation : cloud

Ressource mutualisée en croissance

- +10%, 12000 cœurs
- 43 MheuresxCPU (4900 joursxCPU) délivrés : 2021 + 25%
- Utilisation mensuelle moyenne : ~50% cet automne
- Principaux financeurs : IJCLab, ISCPIF et UPSaclay
- 80 « projets » actifs

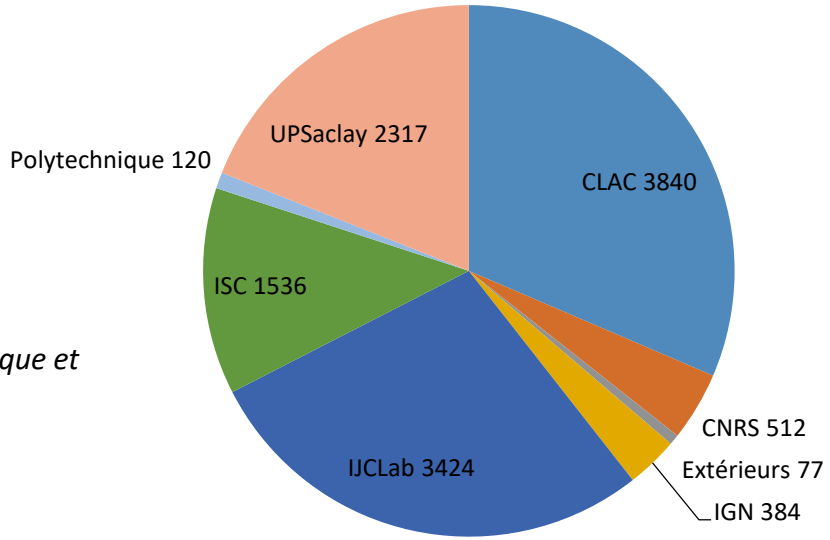
Quelques services critiques

- CNRS BBB, Mathrice BBB, GRANDMA, SVOM…
- Tous les hyperviseurs double alimentés mais pas secourus



vCPUxheures par partenaire

CLAC : activité astrophysique et cosmologie (DIM ACAV+)



Nombres de coeurs par partenaire

Support

- IJCLab se charge de l'administration du cloud, du stockage associé et des services avancés
  - ~4 FTE
  - Helpdesk dédié pour ces services : https://cloud-support.ijclab.in2p3.fr. Accessible à toute personne ayant un compte cloud.
- Pas de support aux applications/services spécifiques d'un utilisateur, d'un laboratoire ou d'une communauté
  - Une expertise de ces applications doit exister dans le laboratoire ou la communauté
  - IJCLab/mésocentre favorise la mise en réseau des expertises similaires communes à plusieurs groupes
- Obtenir un compte pour accéder au cloud : https://registration.lal.in2p3.fr
  - Auto-enregistrement : pas de validation
  - Contacter le helpdesk cloud pour créer un projet ou être associé à un projet existant (plutôt passer par le responsable du projet)

Modèle économique

- Cloud ouvert à tous les membres de Paris Saclay sans garantie de ressources
  - Contribution des utilisateurs pour se voir garantir un niveau de ressources défini : permet de démarrer une activité et valider que le cloud répond aux besoins
- Modèle de contribution centré sur les machines et non pas les heures de calcul
  - Reflète le fait qu'on alloue des machines (virtuelles) au lieu de soumettre des jobs
  - Achat du nombre de cœurs qu'on souhaite se voir garantir en moyenne : possibilité d'en utiliser plus si/quand il y en a de disponible, utilisation par les autres de ce qui n'est pas utilisé
  - Financement de l'hébergement des machines achetées pendant la période d'utilisation (typiquement 5 ou 7 ans)
- Cas particulier du stockage : les octets ne sont pas mutualisables
  - Nécessite d'acheter les TBs qu'on souhaite pouvoir utiliser
- Matériel acheté est du matériel standard, disponible au marché MATINFO5
  - Possibilité d'ajouts de ressources au fil de l'eau
  - IJCLab prescripteur du type de matériel

VirtualData dans le mésocentre

- Une ressource ouverte à tous les membres de Paris Saclay, complémentaire de RUCHE
- Financement CPER va permettre une évolution des ressources financées hors projet pour accueillir les « petits » utilisateurs et permettre les expérimentations
  - Le besoin de disposer de ressources significatives régulièrement devra toujours faire l'objet de contributions spécifiques
  - Les ressources disponibles permettront d'allouer des ressources significatives pour une courte durée à un projet avec des besoins exceptionnels
  - Les ressources CPER incluront des ressources de calcul et du stockage
- Plateforme de stockage globale du mésocentre : un service important pour l'utilisation combinée de VirtualData et RUCHE
  - Va s'appuyer sur l'expertise IJCLab autour de la technologie Ceph dans un environnement distribué
  - Ouverture planifiée pour l'automne 2023 : première action financée par le CPER
- Hébergement hors cloud : possibilité d'héberger du matériel spécifique dans un environnement résilient et avec une consommation énergétique/électrique optimisée