CINES – Département Archivage et Diffusion

3ème journée des utilisateurs de l'archivage

9 juin 2015


JOURNEE
INTERNATIONALE
des ARCHIVES
Mardi
9 juin
2015
Et on la fête absolument
partout...
...même au CINES !
CINES
ICA
International Council
on Archives

L'approche datacentrique du CINES,

une politique de gestion intelligente des données

3ème journée des utilisateurs de l'archivage – 9 juin 2015

Environnement HPC

Renater

Environnements de pré/post-traitement

YODA : IBM Power7 (arrêt en juin)

8 nodes P755 6.6 Tflops

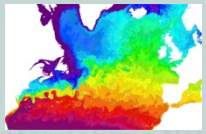
- 32 cores: IBM Power7 @ 3.3 GHz
- 128 GB / node
- Infiniband DDR
- GPFS , 10 TB @ 3 GB/s (/scratch, /home)

CRISTAL : Bull s6030 (Visualisation)

2 nodes s6030 3 Tflops

- 32 cores : Intel x7560 @ 2.27 GHz
- 128 GB node1
- 256 GB node2
- 2 GPUs Nvidia

Quadro Plex D2 (FX5800)



10 Gb Backbone

Ressources HPC

OCCIGEN : Bull DLC

2106 nodes bullx B720 2.1 Pflops

- 50544 cores Intel E5-2690 @ 2.6 GHz – 64 and 128 GB/node
- Infiniband FDR
- Lustre 4.6 PB @ 100 GB/s (/scratch), Panasas 240 TB @ 5 GB/s (/home)

FDR Infiniband gateways

Ressources de stockage

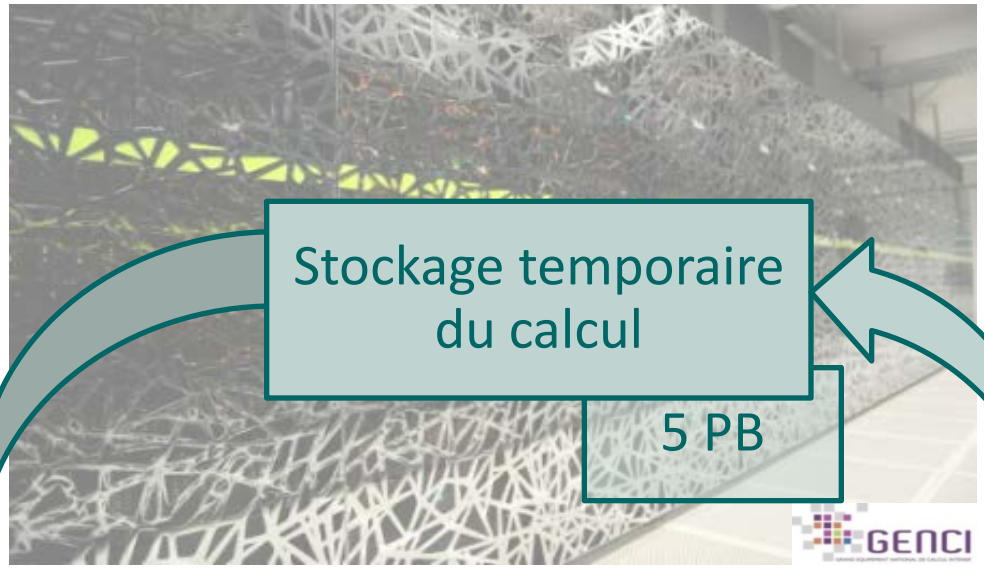
Lustre HSM 2 PB @ 50 GB/s, DMF, 2 x IBM TS3500 lib. ~2 .3 PB, 2 x 1000 cartridges, 9 x Jaguar3 & 10 x LTO4 drives (/store)

En 2015 : passage en 9 Jaguar4 (gain 60% vol) et 8 LTO6 + 2 LTO4 (gain = x 3)

Cycle de vie des données au CINES

Occigen : supercalculateur parallèle Bull DLC , Rmax = 2.1 Pflops/s

étape 1 : chargement code source + données en entrée



étape 2 : Transfert données sur espace de travail

étape 3 : Transfert des données pour sécurisation à court-terme

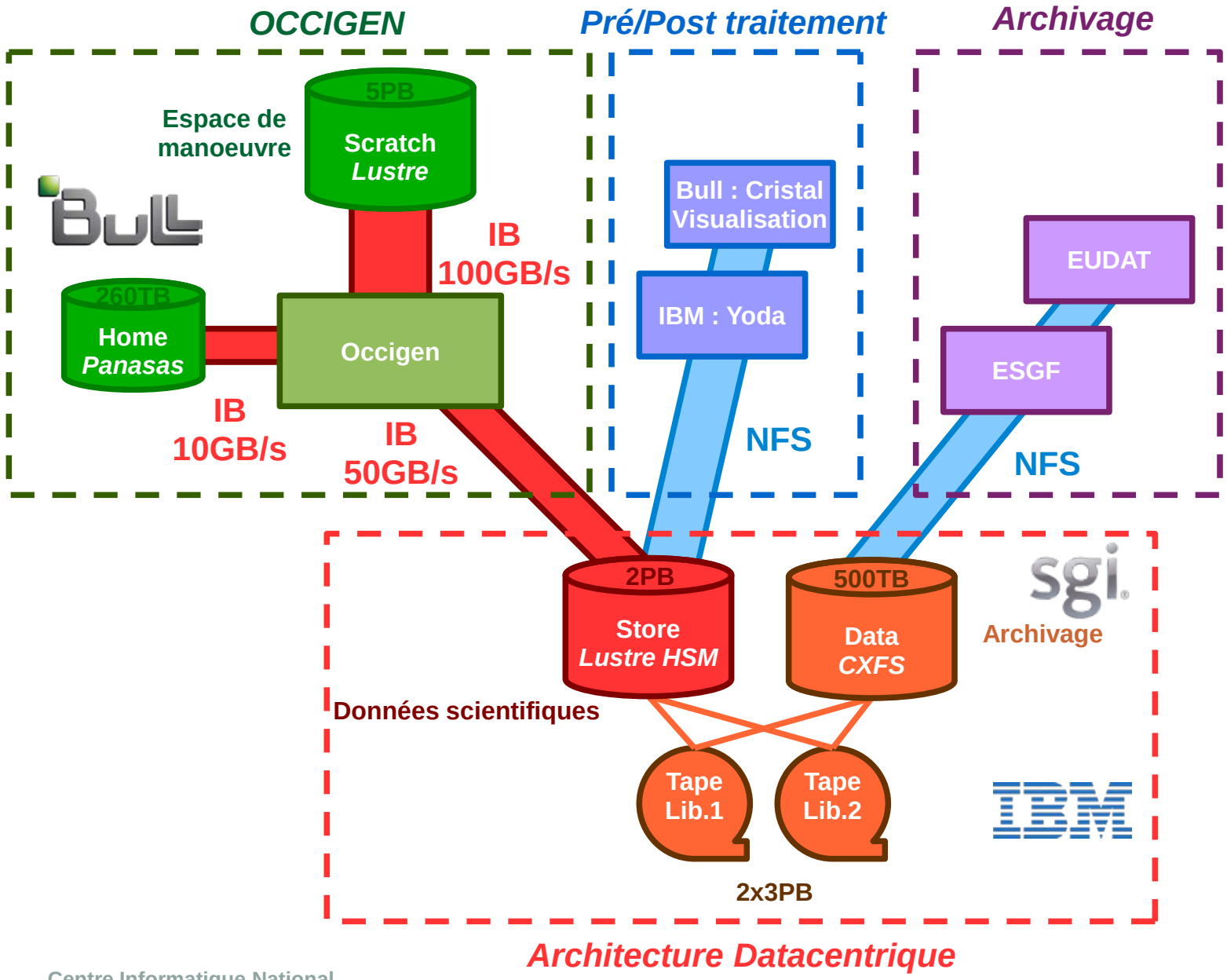
Espace de stockage pour la durée du projet (home & store)

2 PB +

Espace pour préservation à moyen/long termes

étape 4 : Transfert des données pour préservation à moyen/long termes

Organisation des données



Classification par durée de conservation des données

Durée de conservation	<18 mois	18 mois à 5 ans	> 5ans
Cadre	DARI	ISAAC/Eudat/ESGF/…	PAC
Accès	Fréquent	Régulier	Occasionnel
Disponibilité	Immédiate	Retardée	Différée
Support	Disques (2 Po) + cartouches (HSM)	Disques ou cartouches	Cartouches
Metadata	Non exigées	Légères	Complètes
Modèle économique	Gratuit pour projet DARI	Facturé	Facturé

- **Projet DARI = 1 an + 6mois de délais pour récupérer les données**
- **ISAAC/EUDAT** = service de conservation des données à moyen terme (archivage intermédiaire)
- **PAC** = plateforme d'archivage pérenne du CINES (conservation à long terme)

Données scientifiques, du stockage à l'archivage pérenne :

présentation des services EUDAT, ESGF, et PAC

3ème journée des utilisateurs de l'archivage – 9 juin 2015

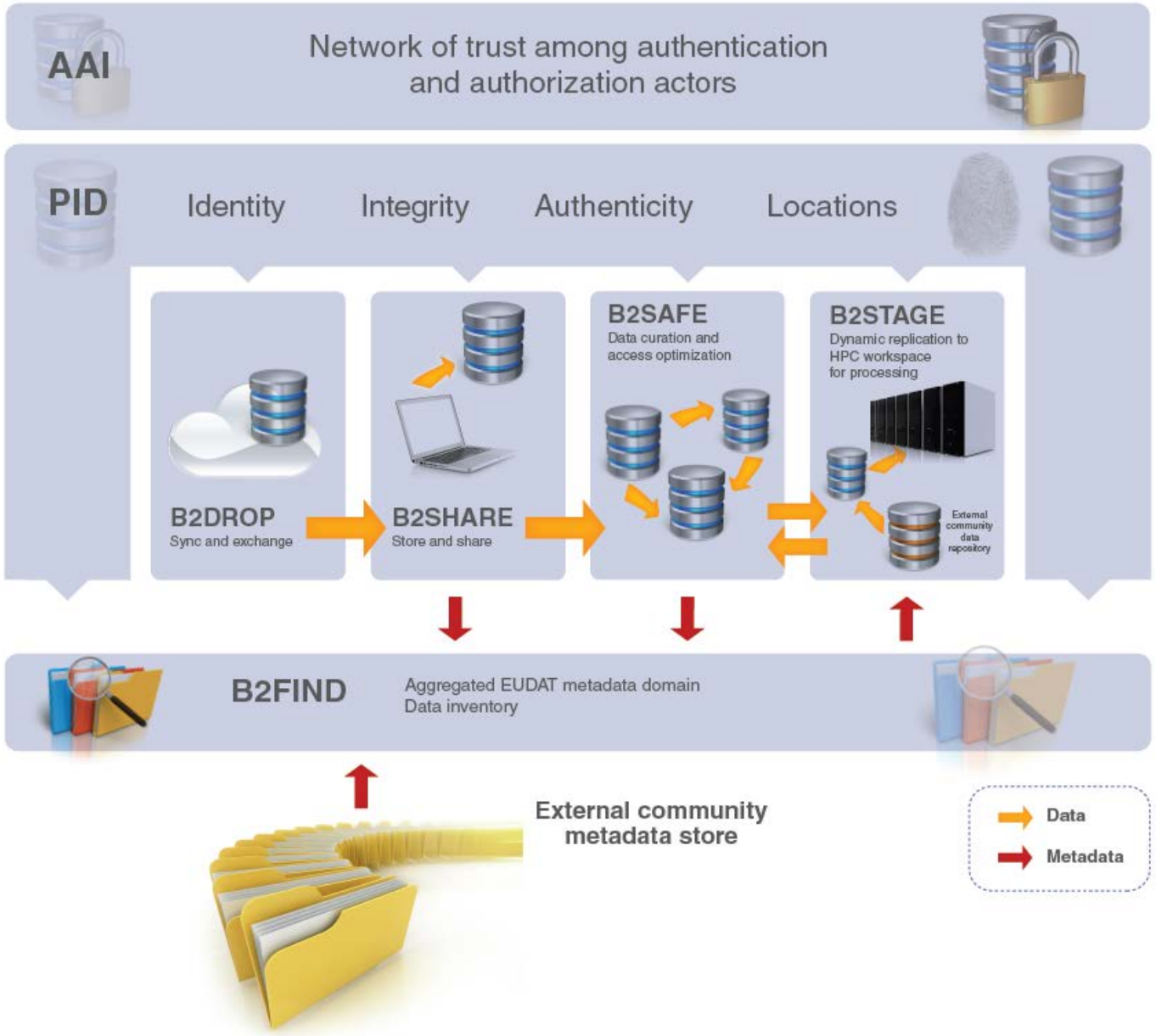
Eudat « European Data for e-science »

- **Démarrage du projet le 1 octobre 2011 pour 3 ans**
- **Renouvelé pour 3 ans en 2015**
- **Objectif : fournir une Infrastructure Collaborative de Données (CDI) européenne qui adresse le cycle de vie de la donnée (Stockage, Traitement, Accès, Echange, Conservation à moyen et long termes)**
- **Aux communautés de la recherche dans toutes les disciplines**
- **A travers un réseau de 35 partenaires (centres de calcul , centres de données, communautés scientifiques) à travers 15 pays européens**
- **CINES particulièrement impliqué dans :**
  - WP2 : Qualité des services (Catalogue de services ITIL, portfolio, SLA/OLA, DSA), coûts et business model.
  - WP 5 : Conservation, gestion et analyse des données
  - WP 6 : Validation des services sur des communautés / intégration de nouvelles communautés
  - WP 7 : Interface Eudat / HPC (PRACE - « Joint calls »)

Les partenaires EUDAT




Les services EUDAT



L'appel à collaboration EUDAT

- **POURQUOI FAIRE ?**
  - *Data synchronisations and exchange*
  - *Data repository and data sharing*
  - *Data replication and preservation*
  - *Data staging for analysis and processing*
  - *Data discovery and search*
  - *Data typing and visualization*
  - *New services or tools in the area of Big Data Analytics, Semantic web, etc.*

→ **Possibilités de financement par la CE**

- **POUR QUI ? Toutes les initiatives, infrastructures et communautés de recherche européennes.**

- **CRITERES d'EVALUATION :**
  - Faisabilité technique du pilote (considérant le calendrier et les ressources dédiés) [40%]
  - Participation et bénéfices attendus pour la communauté de recherche ciblée [20%]
  - Valeur ajoutée pour EUDAT (développement de services, de communautés) [20%]
  - Contribution à l'open access [10%]
  - Développement de solutions selon des approches générique telles que RDA [10%]

- **CALENDRIER :**
  - Date limite des soumissions : 30/09 – 17h CET
  - Implémentation du pilote : 01/01/2016 – 30/06/2017


Home
Plan
Satellite
Legenda
Compute
Idp
Index
Data

- **Nouveau service CINES : pour la publication de données (ESGF public)**

- **Master des données CMIP5 décennales du groupe de modélisation CNRM-CERFACS : mis à disposition au travers de esgf.cines.fr**

- **Participation à ESGF-France**

- **Déploiement d'un « ESGF privé » (à l'étude)**

Et l'archivage pérenne / PAC dans tout ça ?

- Quèsaco ?
- Triple engagement :

Préservation de l'intégrité

SUPPORT : Checksum + RAID + contrôles CRC + VEILLE + MIGRATION PHYSIQUE + copie sur SITE DISTANT

Préservation de la lisibilité

ENVIRONNEMENT MATERIEL : VEILLE TECHNO et ANTICIPATION

ENVIRONNEMENT LOGICIEL : privilégier les FORMATS DURABLES + MIGRATION LOGIQUE (→ validation préalable de format)

Préservation de la capacité à comprendre le contenu du fichier

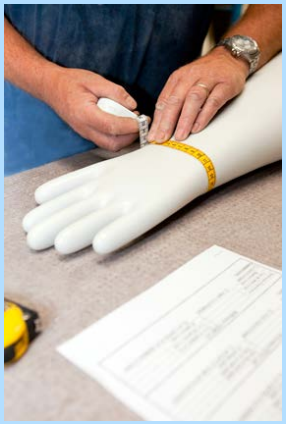
METADONNEES DESCRIPTIVES + IDENTIFICATION UNIQUE et PERENNE

Et l'archivage pérenne / PAC dans tout ça ?

Plateforme d'archivage (agréée SIAF + Santé + DSA + ISO 16363)

1. Réception

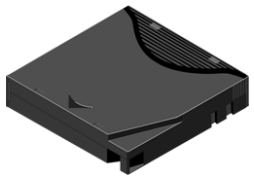
2. Vérification de la qualité des données reçues



3. Ajout d'informations (PID, empreintes, date d'archivage…)

4. Traitements complémentaires (récupération d'informations associées, etc.)

5. Stockage de l'archive en plusieurs exemplaires

Salle machine 1	Salle machine 2	Salle machine 3	Site distant (> 300km)
			

6. Vérification périodique de tous les exemplaires archivés

+ **migration** quand il y a lieu

Quand la climatologie utilise les services ESGF et EUDAT pour la gestion de leurs données

Marie-Pierre LEMOINE

CERFACS

3ème journée des utilisateurs de l'archivage – 9 juin 2015

INES