

Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

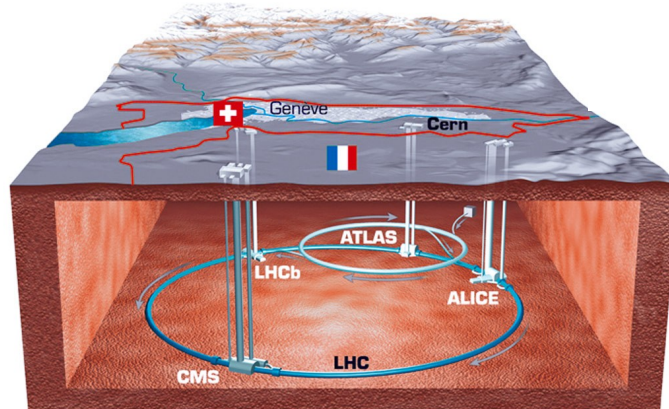
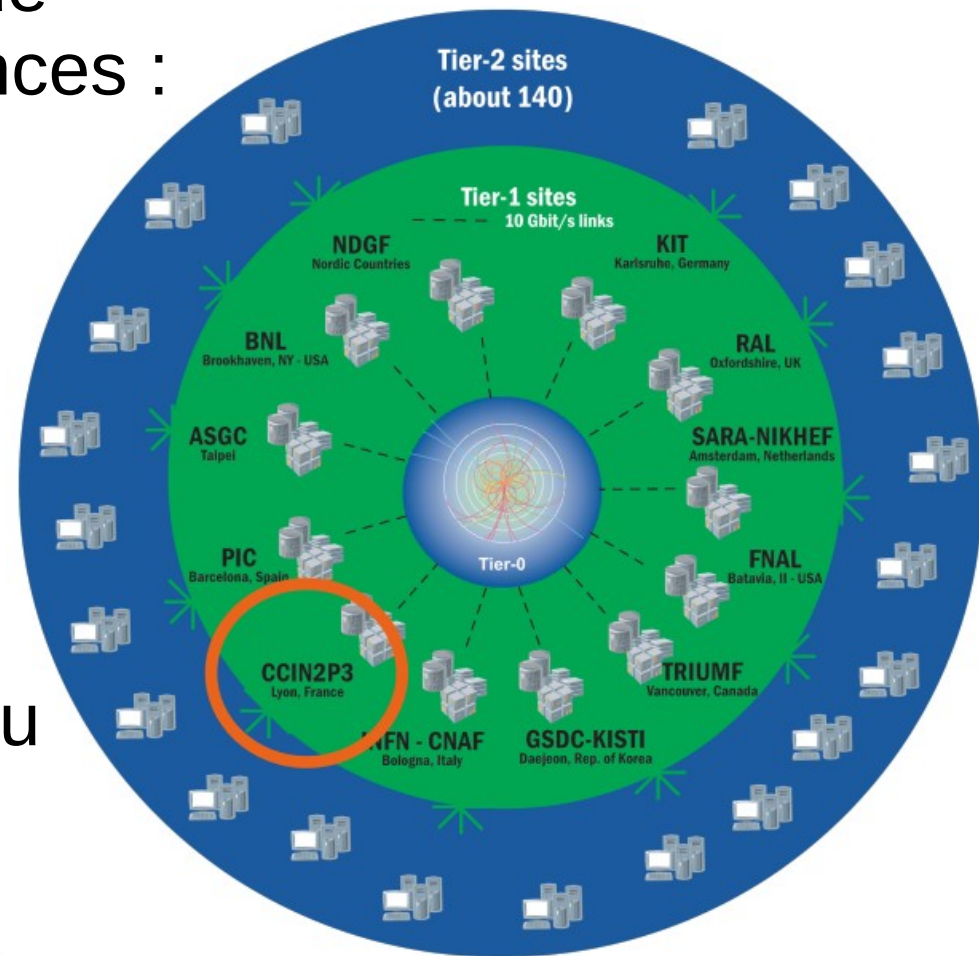
Présentation du CC-IN2P3

David Bouvet – Loïc Tortay
DU Data Science
13/05/2019

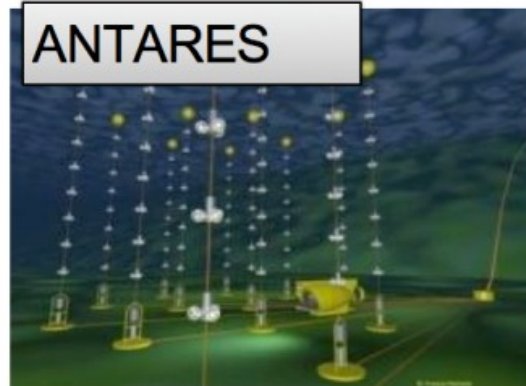
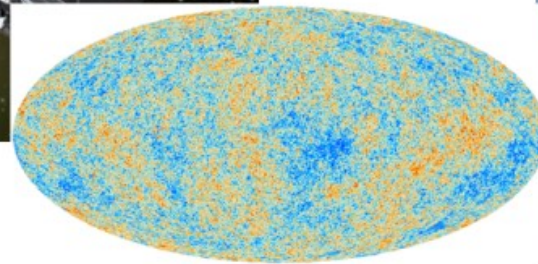
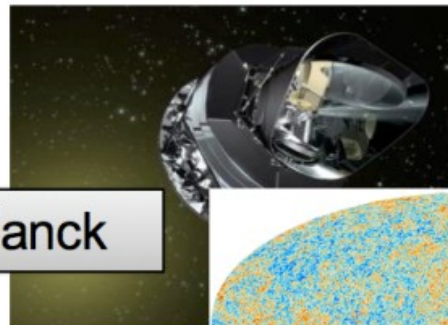
- Institut National de Physique Nucléaire et Physique des Particules
- L'un des 10 instituts du CNRS
- Composé de 20 laboratoires
 - 1000 chercheurs, 1500 ingénieurs et techniciens
- Domaines scientifiques
 - Physique des particules et hadronique
 - Physique nucléaire
 - Physique des astroparticules

- Centre de Calcul de l'IN2P3 / CNRS
- Installations informatiques et de stockage de données dédiées pour l'IN2P3
 - Missions : fournir des ressources IT pour la communauté française de physique des hautes énergies
 - Déploie aussi une infrastructure commune pour les services institutionnels (GED, outils collaboratifs...)
- 84 personnes dont 68 ingénieurs IT
 - Fonctionnement 24/24, 7/7, 365/365
- Participe à plus de 70 expériences
 - Dont la plupart d'envergure internationale

- Réseau de grands centres de données pour les 4 expériences : ALICE, ATLAS, CMS, LHCb
- Maillon de premier niveau
- Le CC-IN2P3 fournit 10% des ressources mondiales du calcul pour le LHC



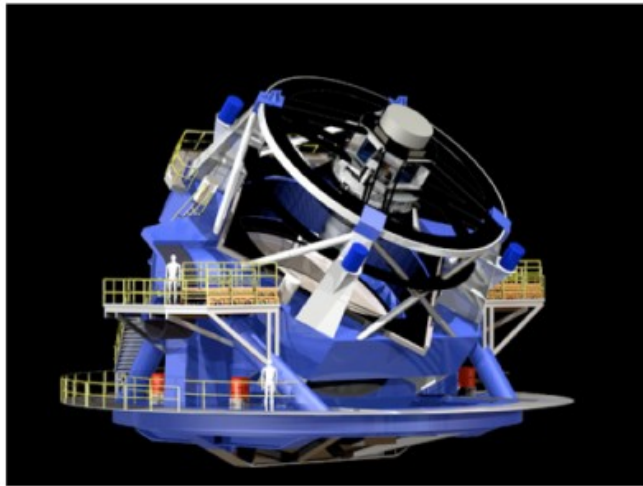
Principales expériences actuelles...



LSST

Totalité des données disponibles au CC-IN2P3

50 % du traitement par le CC-IN2P3 et 50 % par NCSA (USA)



EUCLID

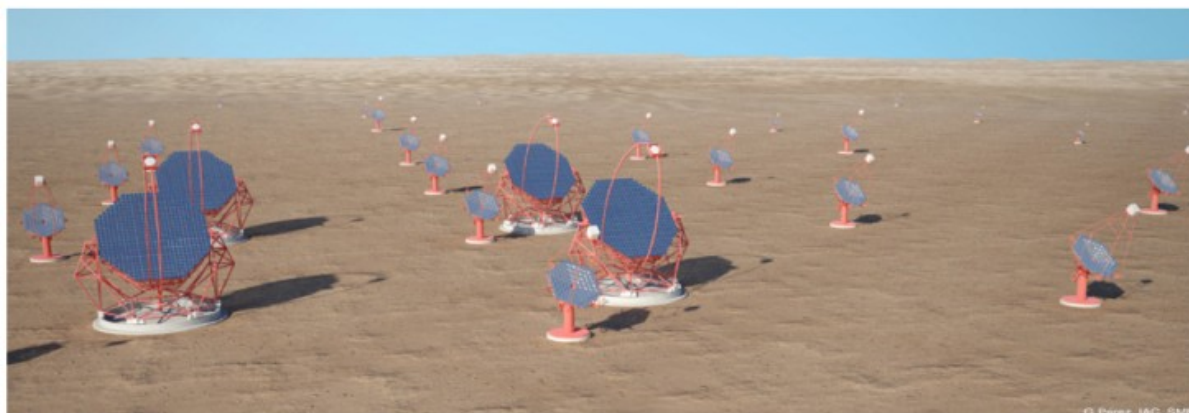
Le CC-IN2P3 est le centre de données français pour le traitement et la gestion des données



Énergie et matière noires

CTA

Le CC-IN2P3 jouera un rôle clé dans le traitement des données de CTA



Rayons gamma

- 2 salles machines de 850 m² chacune
 - VIL1 : bâtiment principal, depuis 1986
 - héberge les services et les systèmes de stockage
 - VIL2 : depuis 2011
 - héberge la ferme de calcul et le stockage depuis 2015
- Les 2 salles machines sont reliées par un lien à 400 Gbps



400 Gbps



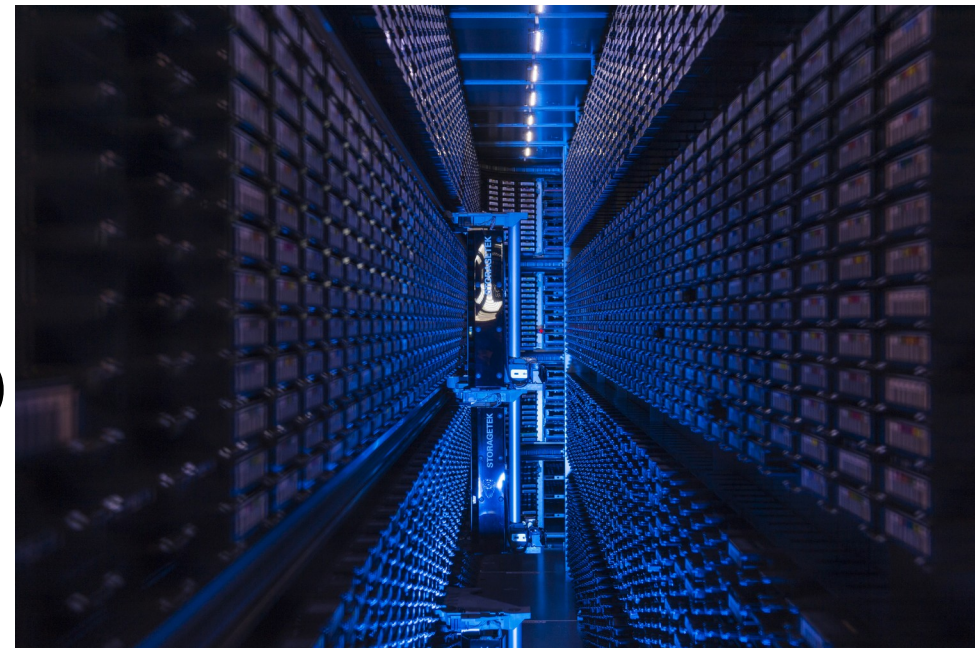
- 2 lignes EDF HTA
- 5 transfos HTA
- 6 groupes froids
- Consommation électrique : 1,2 MW

- 3 fermes :
 - HTC (High Throughput Computing)
 - jobs séquentiels, multicoeurs et tableau
 - jobs interactifs
 - HPC (High Performance Computing)
 - jobs parallèles
 - GPGPU (General Purpose GPU)
 - jobs GPU
 - jobs GPU interactifs

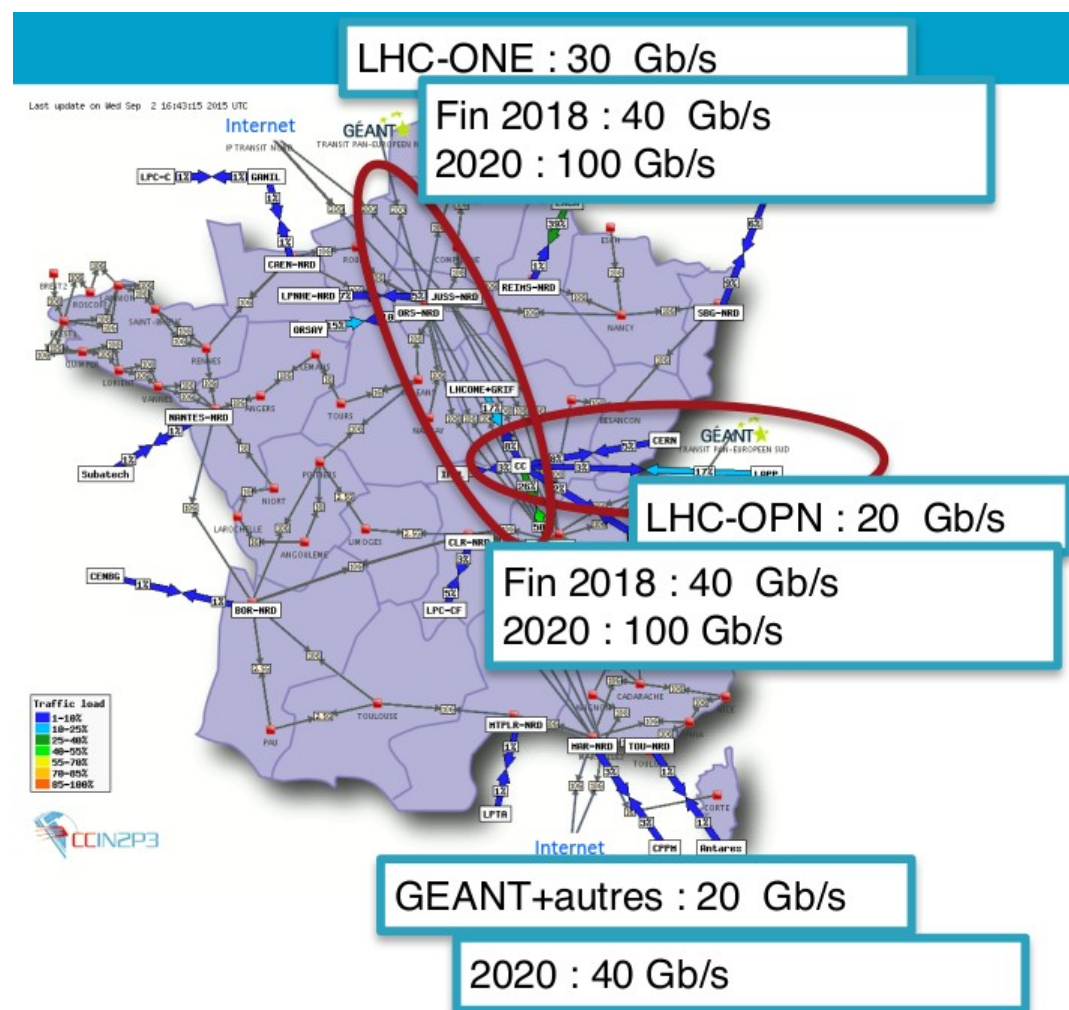
- Univa Grid Engine : ordonnanceur de tâche
 - Point d'entrée unique
 - Gère les 3 fermes pour optimiser l'utilisation des ressources en fonction des besoins et des priorités des expériences



- Stockage de masse
 - Bandes magnétiques et disques pour le cache
 - utilisées par HPSS dans des bibliothèques StorageTek
- Stockage disque
 - DAS (Direct Attached Storage)
 - utilisé par dCache, XRootD, iRODS
 - frontal pour l'accès au stockage de masse
 - NAS (Network Attached Storage)
 - utilisé par Isilon
 - accès POSIX via montage NFS
 - SDS (Shared Disk Storage)
 - utilisé par GPFS
 - accès POSIX



- Hébergement de points de présence opérateurs réseau
 - Nationaux : RENATER
 - Régionaux : Lyres, Amplivia...
- Liaison dédiée avec le CERN à 40 Gbps



- Bases de données

- SGDB

- Oracle
 - MariaDB : fork de MySQL
 - PostgreSQL

- NoSQL

- MongoDB
 - Cassandra

- Plateforme SPARK

- Outils collaboratifs

- Mail

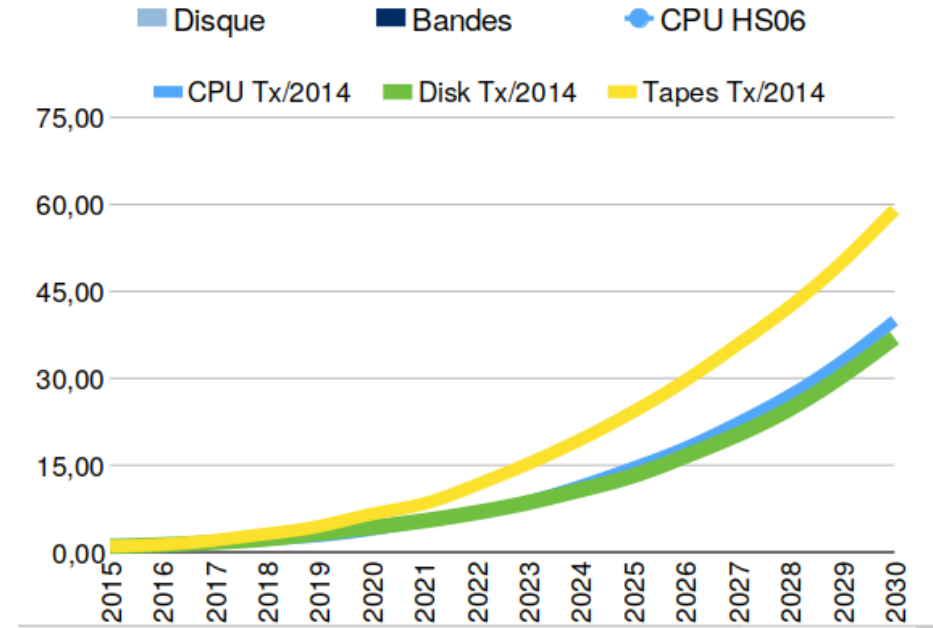
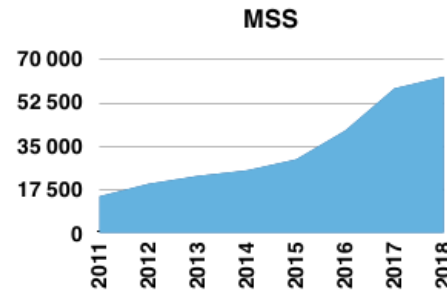
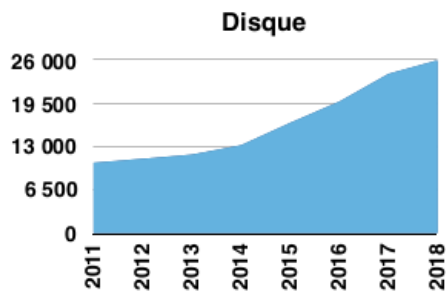
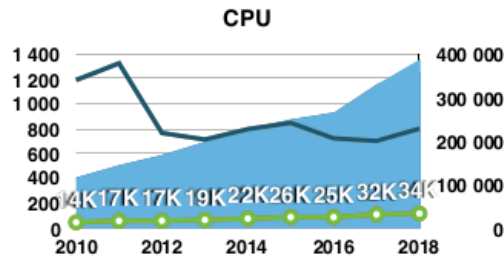
- GED (Gestion Electronique de Documents)

- Gitlab

- Forge

- Outil d'intégration continue

- Besoins en constante augmentation



- Conserver la mutualisation des ressources
- Proposer de nouveaux services pour répondre à la demande