



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

# L'impact des données sur l'infrastructure

Journées Plateformes  
6 Octobre 2016

Rachid Lemrani





- ▶ **Introduction : Le CC-IN2P3**
- ▶ Impact du volume des données
- ▶ Impact de l'accès données
- ▶ Impact de la variété des données
- ▶ Impact de la variété des traitements
- ▶ Futur au CC-IN2P3

- ▶ **CC-IN2P3** : Unité de Service et de Recherche du CNRS fonctionnant en partenariat avec le CEA / DRF/ IRFU
- ▶ **Mission principale** : déployer et opérer les moyens informatiques nécessaires à la mise en œuvre de la politique scientifique de l'IN2P3



## Le CC-IN2P3



- ▶ Fonctionnement 24/24, 7/7, 365/365
- ▶ 2 000 m<sup>2</sup> de bureaux ~100 personnes
- ▶ 4 000 m<sup>2</sup> de locaux techniques
- ▶ 2 salles informatiques de 850 m<sup>2</sup> :
  - CPU : ~26 000 vcores – 253 kHS06, 239 TFlops)
  - Disque : ~22 Po (~ 2 Po haute performance)
  - Bandes magnétiques : 31 Po (capacité 340 Po)
  - Sauvegarde (TSM) : 1.1 Po (capacité 5 Po)



## Les expériences (~75 groupes actifs)

### ▶ Physique des particules

- Modèle standard et au delà : **ATLAS**, **CMS**, D0, H1, ...
- Violation de symétries : **LHCb**, Babar, ...

### ▶ Physique des astroparticules

- Neutrinos : **DOUBLE-CHOOZ**, **NEMO**, **OPERA**, ...
- Cosmologie : **PLANCK**, **EDELWEISS**, **SNLS**, **SNF**, ...
- Rayons cosmiques : **AMS**, **ANTARES**, **AUGER**, **FERMI**, **HESS**, ...
- Ondes gravitationnelles : **VIRGO**

### ▶ Physique Hadronique / Nucléaire

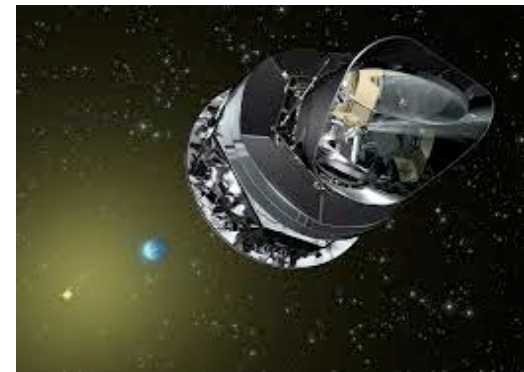
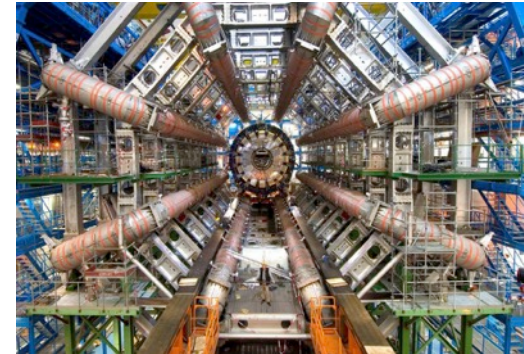
- Plasma quark-gluon : **ALICE**, Phenix, ...
- Structure nucléaire : **AGATA**, **INDRA**, ...
- Radiobiologie, Imagerie, ... : **HADRONTHERAPIE**, ...

### ▶ Physique théorique

- **QCD** (interaction forte), **NANTHEO** (HEP), ...

### ▶ Ouvertures aux autres disciplines

- Science de la vie, Santé, SHS ...





- ▶ Introduction : Le CC-IN2P3
- ▶ **Impact du volume des données**
- ▶ Impact de l'accès données
- ▶ Impact de la variété des données
- ▶ Impact de la variété des traitements
- ▶ Futur au CC-IN2P3

# Stockage de masse sur bandes magnétiques moins onéreuses que les disques



### Mass Storage System: HPSS

- **37 PB**
- Max traffic (from HPSS): **100 TB / day**
- Interfaced with our disk services

Backup service: TSM (**1 PB**)

### 4 Oracle/STK SL8500 librairies:

- **40,000** slots (T10K and LTO4)
- Max capacity: **320 PB** (with T10KD tapes)
- **111** tape drives

### 1 IBM TS3500 library:

- **3500** slots (LTO6)



- Les utilisateurs accèdent de manière transparente aux fichiers stockés sur bandes magnétiques
- Les fichiers non présents sur disque sont récupérés automatiquement depuis les bandes magnétiques

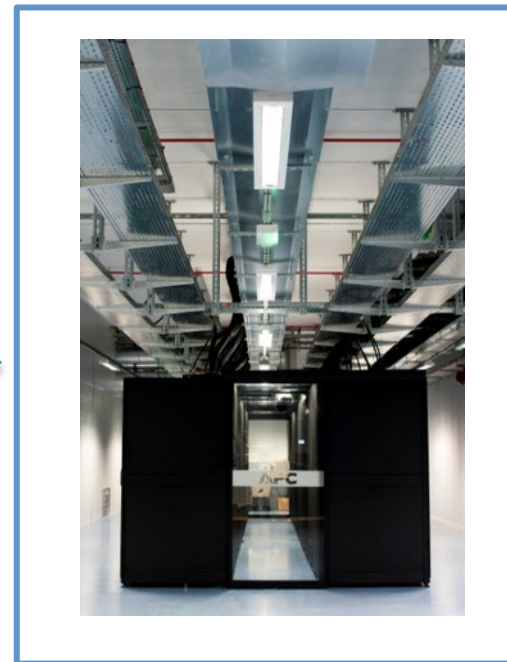
Bandes magnétiques



Disque

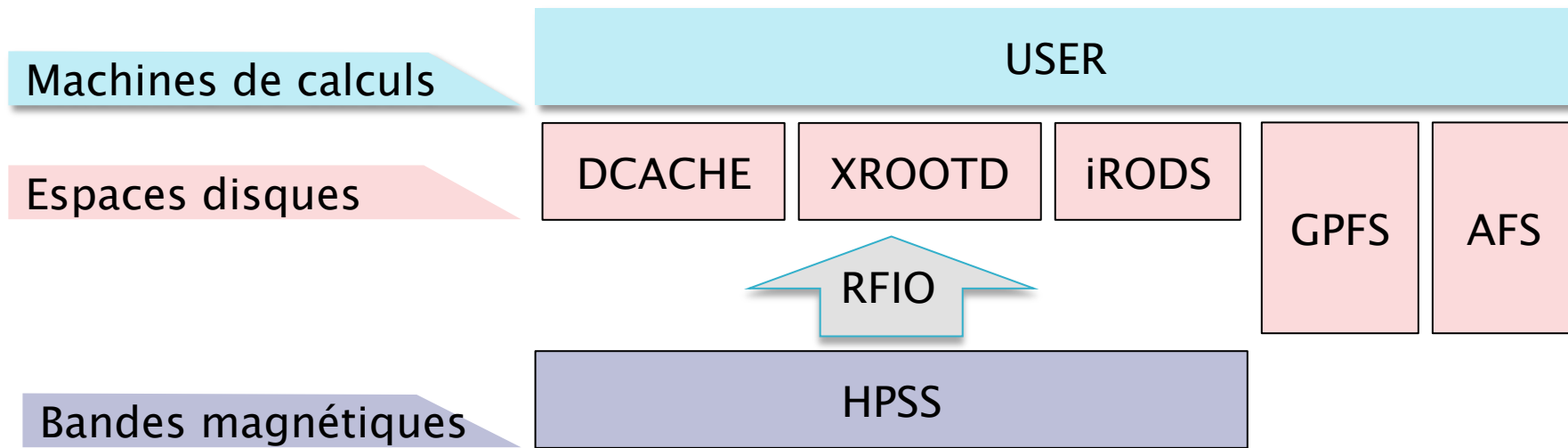


Calcul



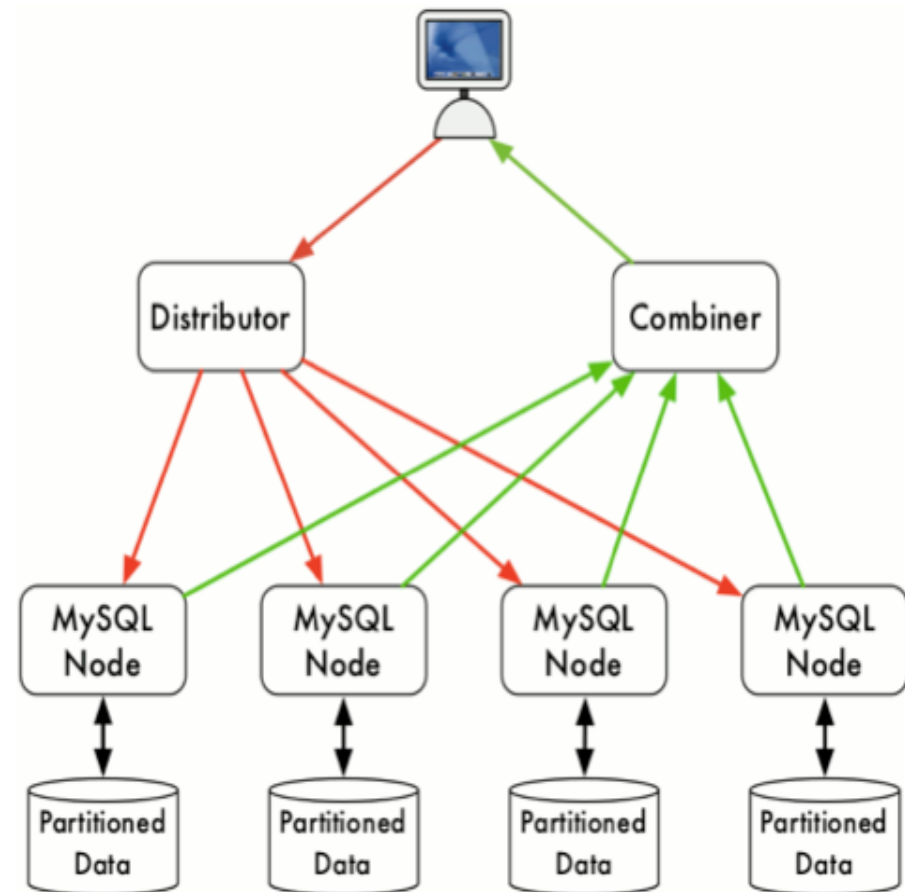
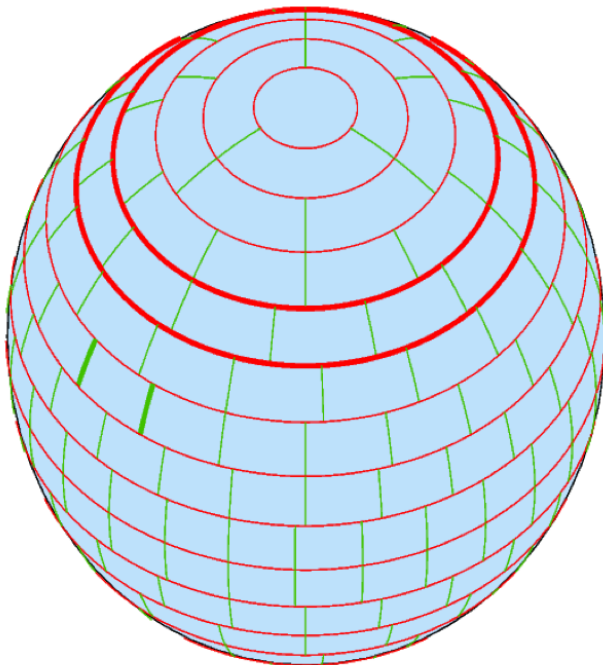


Les utilisateurs accèdent de manière transparente aux fichiers stockés sur bandes magnétiques



## Expérience LSST

**QSERV** : base de données distribuée sur plusieurs nœuds

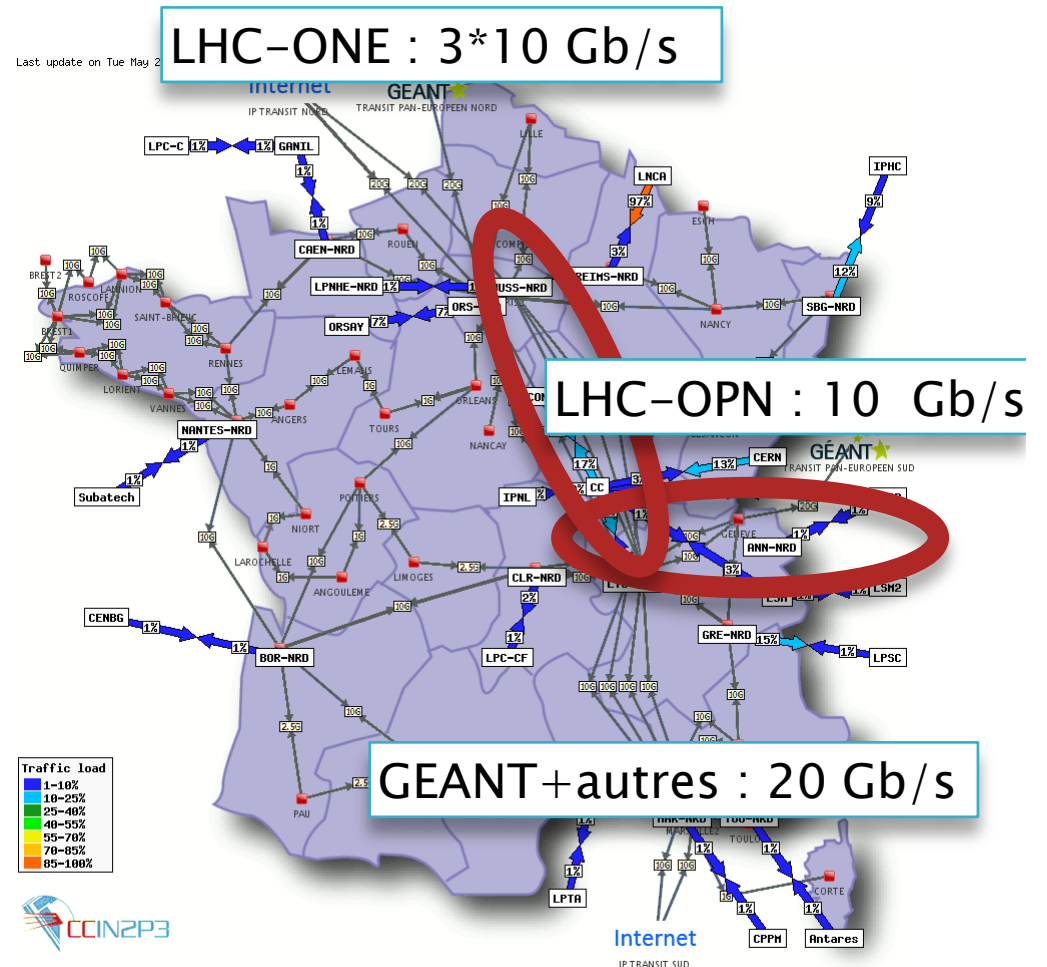
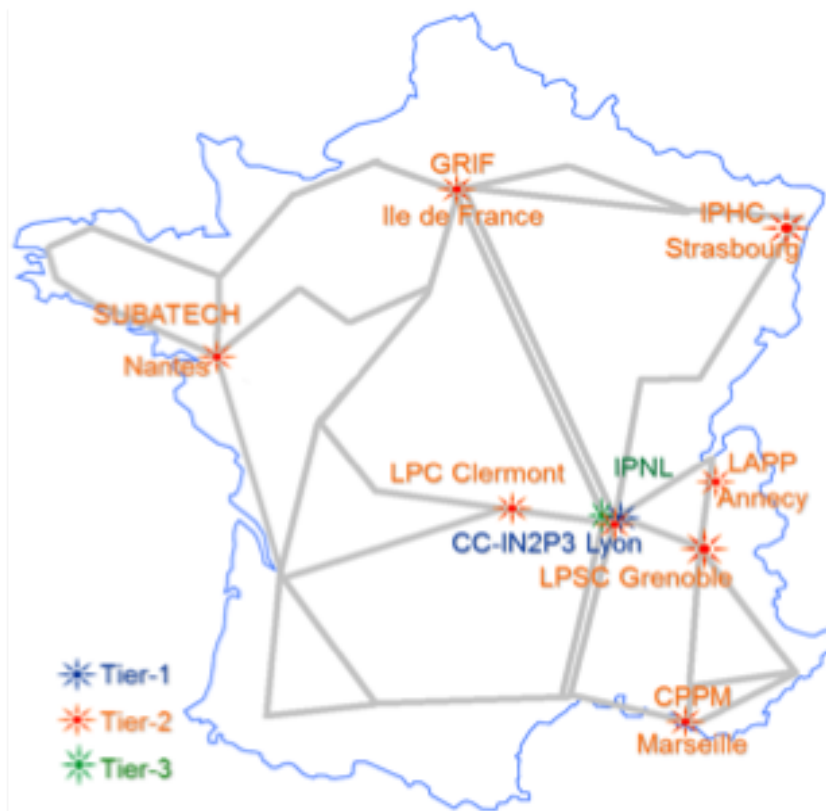




- ▶ Introduction : Le CC-IN2P3
- ▶ Impact du volume des données
- ▶ **Impact de l'accès données**
- ▶ Impact de la variété des traitements
- ▶ Impact de la variété des données
- ▶ Futur au CC-IN2P3

# Transferts des données entre sites

- ▶ Transferts : Site → CC-IN2P3 ↔ Autres centres de données
- ▶ Distribution : CC-IN2P3 → Laboratoires
- ▶ Nécessite de bons réseaux



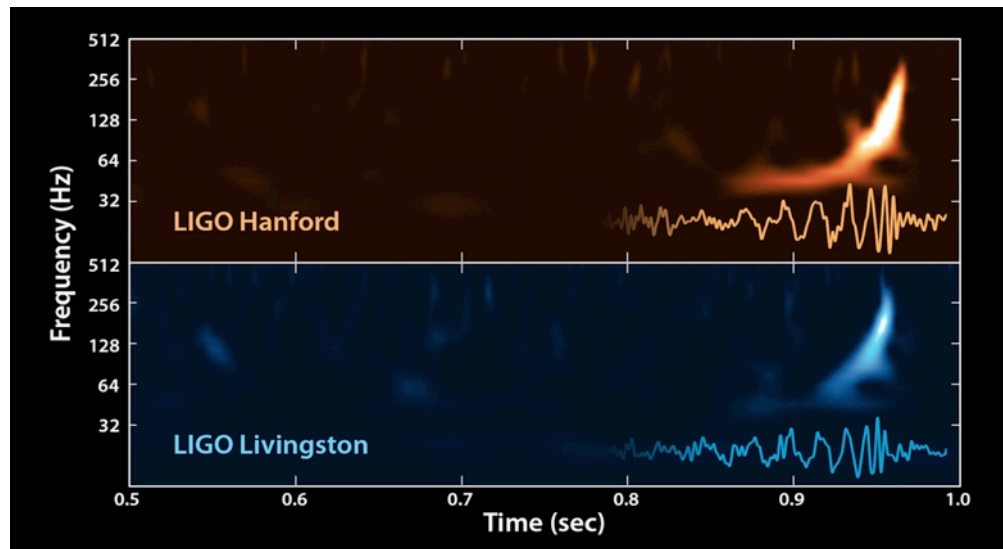
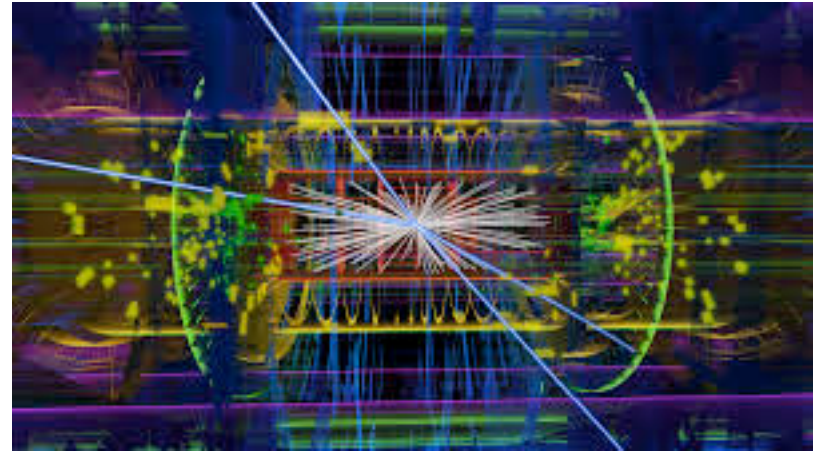




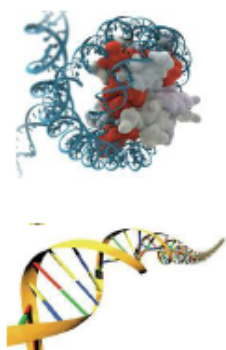
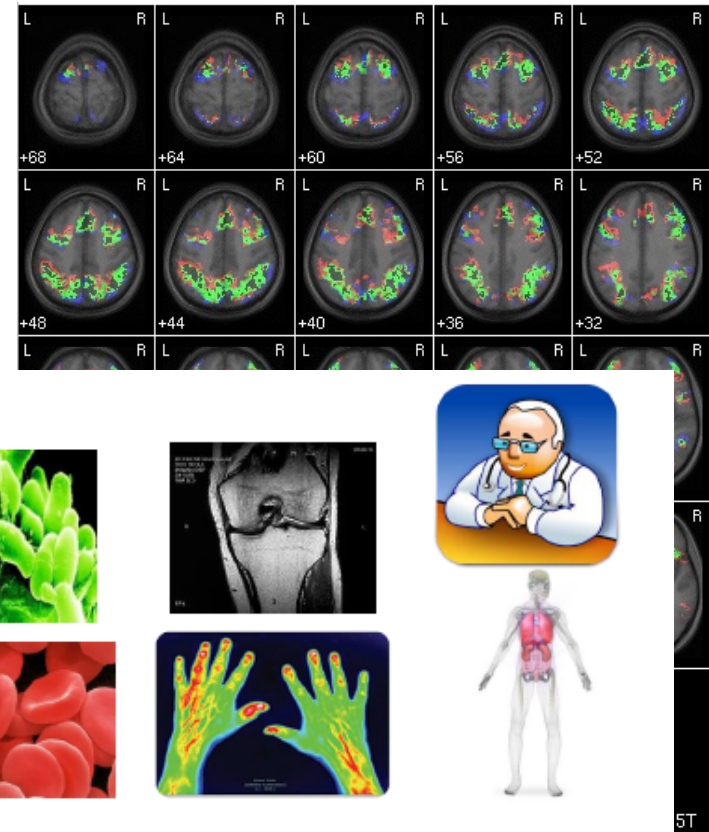
- ▶ Introduction : Le CC-IN2P3
- ▶ Impact du volume des données
- ▶ Impact de l'accès données
- ▶ **Impact de la variété des données**
- ▶ Impact de la variété des traitements
- ▶ Futur au CC-IN2P3

## Type de données

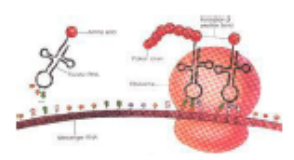
- ▶ **Événements** : (format ROOT)  
particules, gamma, neutrinos,  
rayons cosmiques,
- ▶ **Images 2D** : (format FITS)  
Télescopes CFHT, ...
- ▶ **Signaux 1D** : virgo, planck, ...



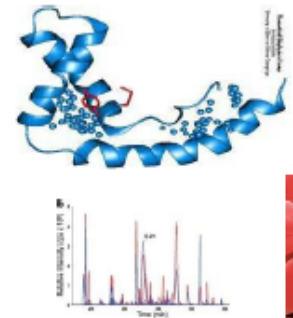
- ▶ **Biologie, Médicale :**
  - ▶ Etriks, Bioaster, ...
- ▶ **Sciences humaines et Sociales**



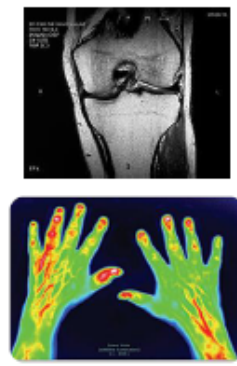
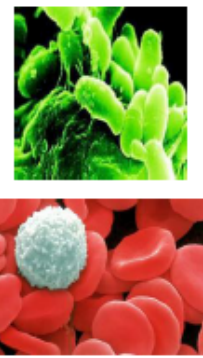
**Epi/Genetics**  
DNA variants & modification



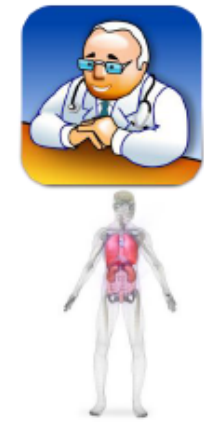
**Transcriptome**  
mRNA, ncRNA  
miRNA



**Peripheral markers**  
Proteins, metabolites,  
cells, microbes



**Organs Systems**



**Clinical Observations**

➔ **Hétérogénéité des données : solutions spécifiques que permet la flexibilité de la virtualisation**



**XROOTD** (“accès aléatoire”):  
fichiers root, productions, ...  
**GPFS** (“accès séquentiel”):  
images, analyses,...  
etc



## Hardware

Direct Attached Storage servers (DAS):

- Dell servers (R720xd + MD1200)
- ~ **240** servers
- Capacity: **15 PBs**

Disk attached via SAS:

- Dell servers (R620 + MD3260)
- Capacity: **1.9 PBs**

Storage Area Network disk arrays (SAN):

- IBM V7000 and DCS3700, Hitachi HUS 130.
- Capacity: **240 TBs**

## Software

Parallel File System: GPFS (**1.9 PBs**)

File servers: xrootd, dCache (**14 PBs**)

- Used for High Energy Physics (LHC etc...)

Mass Storage System: HPSS (**600 TBs**)

- Used as a disk cache in front of the tapes.

Middlewares: SRM, iRODS (**840 TBs**)

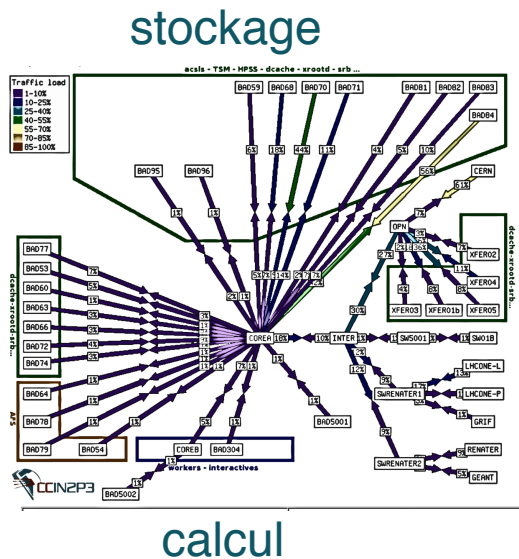
Databases: MySQL, PostGres, Oracle (**57 TBs**)



- ▶ Introduction : Le CC-IN2P3
- ▶ Impact du volume des données
- ▶ Impact de l'accès données
- ▶ Impact de la variété des données
- ▶ **Impact de la variété des traitements**
- ▶ Futur au CC-IN2P3

# Accès aux données depuis les machines de calcul

- ▶ Réseaux performants entre les machines de calcul et les espaces de stockage

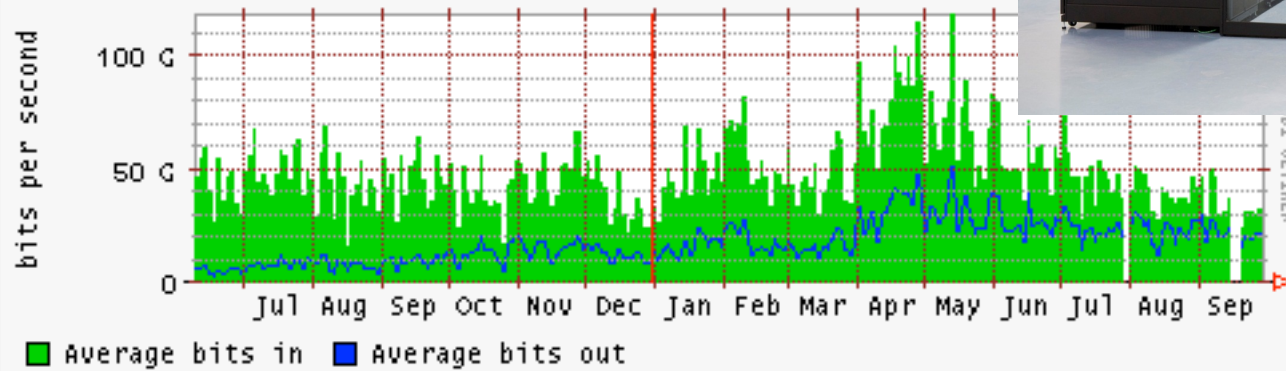


extérieur

ferme de calcul :  
~26 000 vœurs  
3 Go de RAM / cœur  
jusqu'à 48 vœurs par machine



jusqu'à 180 Gb/s utilisés sur plusieurs jours





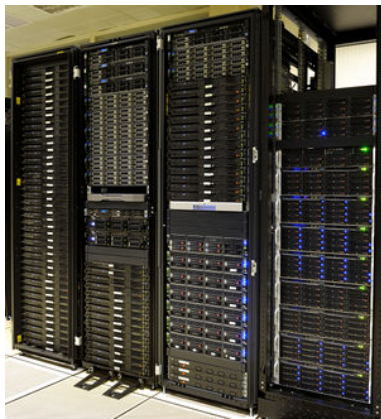
- ▶ Introduction : Le CC-IN2P3
- ▶ Impact du volume des données
- ▶ Impact de l'accès données
- ▶ Impact de la variété des données
- ▶ **Impact de la variété des traitements**
- ▶ Futur au CC-IN2P3

- ▶ **Productions** (collaborations) de données reconstruites, réduites, de simulations Monte Carlo , ...
- ▶ **Analyses** (physiciens) **des** données pour publications
- ▶ **Exemples** :
  - Traitement des événements : reconstruction, calibration, ...
  - Traitement du signal : ondes gravitationnelles, ...
  - CMB : 1D → cartes 2D (planck)
  - Traitement d'images : Supernovae
  - Simulations monte-carlo : geant4 (lhc, ...), gerbes atmosphériques (auger, cta, ...), bruits de fond événements rares (edelweiss, supernemo, ...), ...

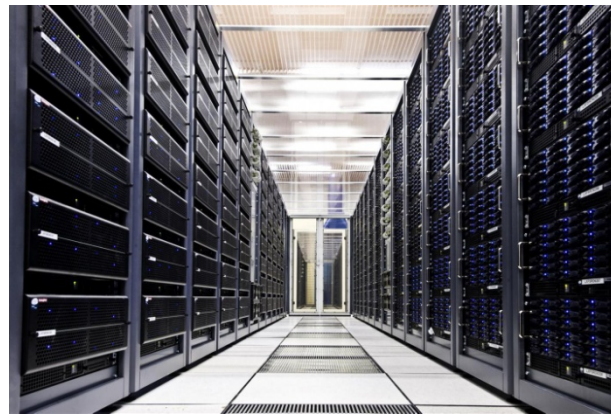
## ► High-Throughput Computing (HTC)

- Ressources à disposition continuellement pour des projets durant plusieurs années
- Beaucoup de « petits jobs » indépendants
- HTC en mode monocoeurs ou multicoeurs (8 cœurs par job pour le LHC actuellement)
- Calcul distribué : les jobs peuvent être soumis via la grille

Laboratoire



Centre de calcul



Grille de calcul  
(plusieurs centre de calculs)



### Atypique pour répondre à des besoins spécifiques non massifs

- ▶ **MPI** : Ferme de 16 machines x 32 cœurs avec Infiniband :

Utilisations de plusieurs machines en parallèle pour les gros jobs

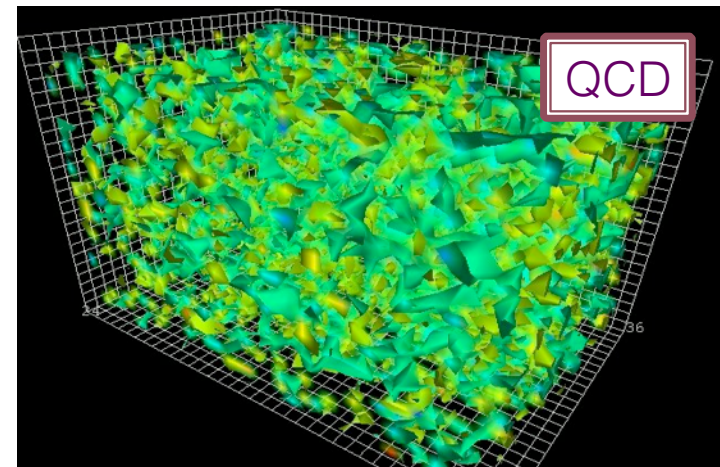
Besoin de beaucoup RAM (PLANCK),

Besoin HPC / Infiniband (QCD),

Petites simulations N-Corps (CRAL),

Simulations neutroniques (MCNPX),

Jobs de simulations en Biologie



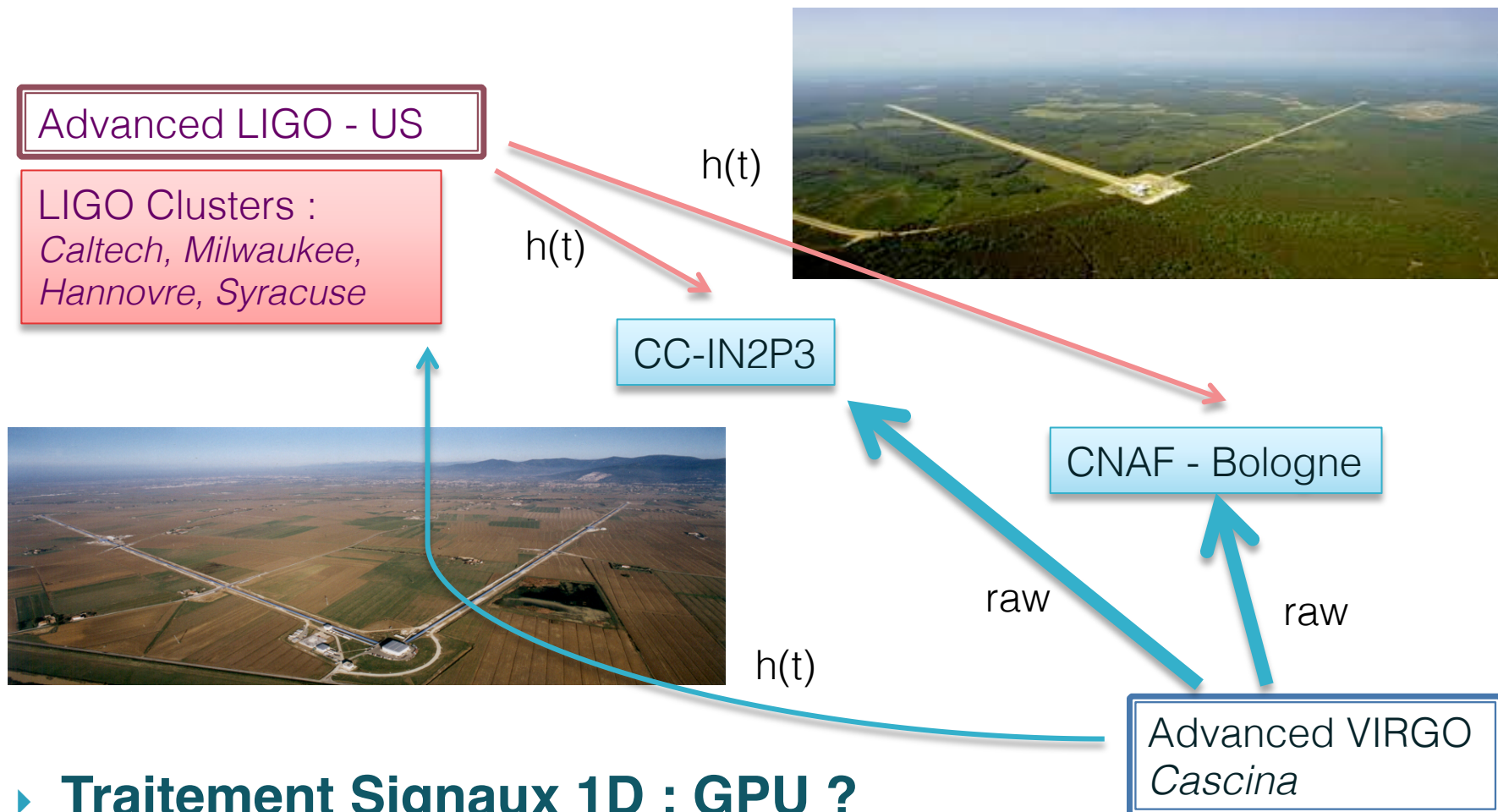
- ▶ **GPU** : 10 machines x 4 GPU - 16 cœurs :

Prochainement : Traitement du signal (VIRGO),

Propagation de particules (AUGER, CTA, KM3NET, ...)

Traitement d'images, biologie, ...

## Transferts et partage des données avec LIGO pour analyse combinée : iRODS



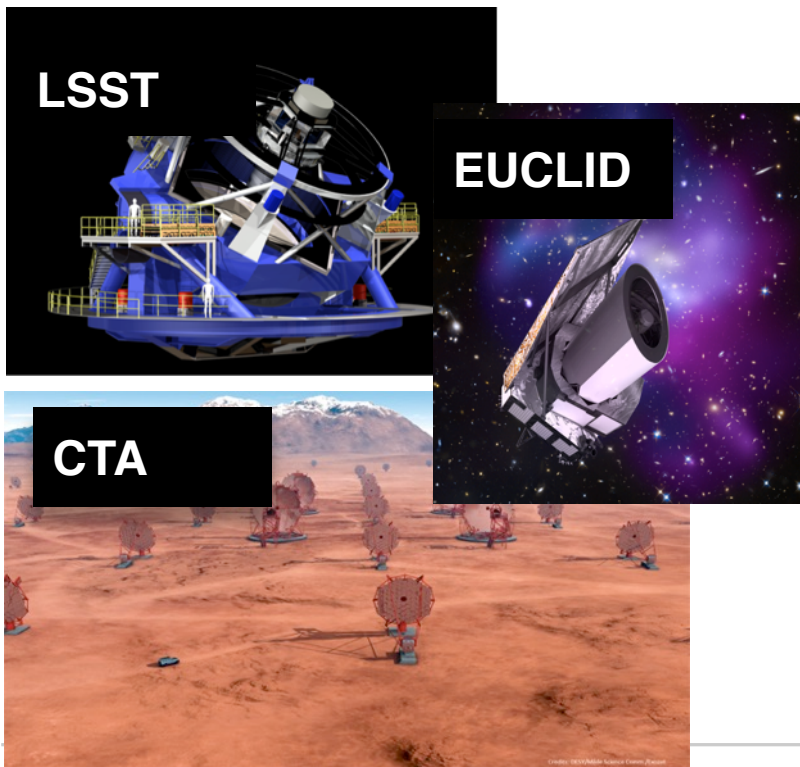
- ▶ **Traitement Signaux 1D : GPU ?**
- ▶ **Stockage de masse : HPSS accédés via XROOTD**



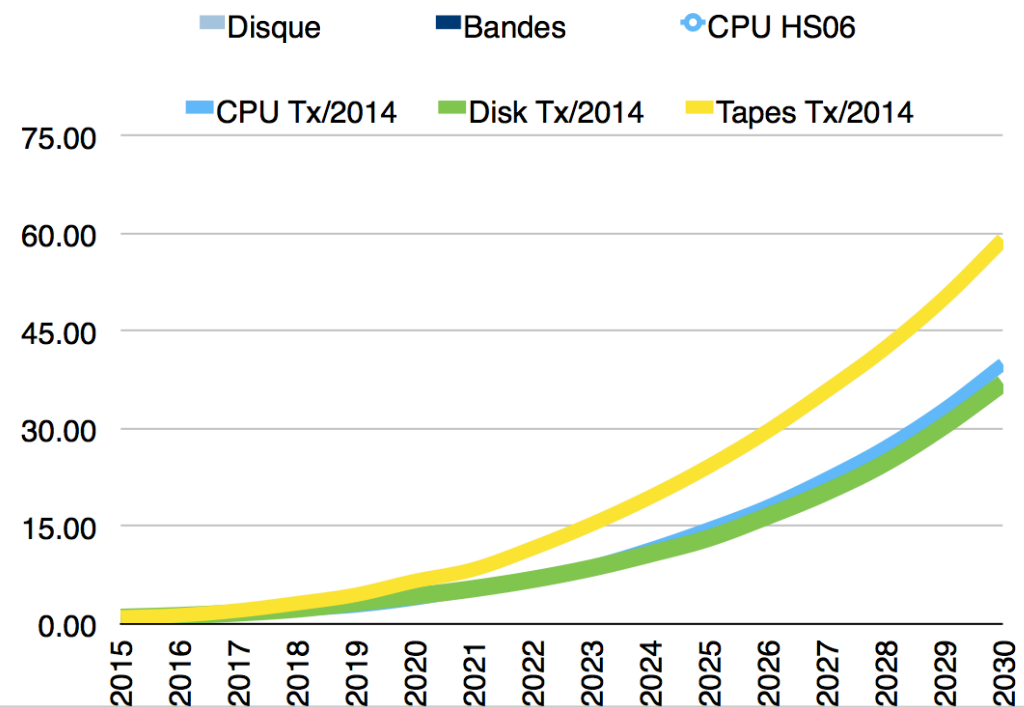
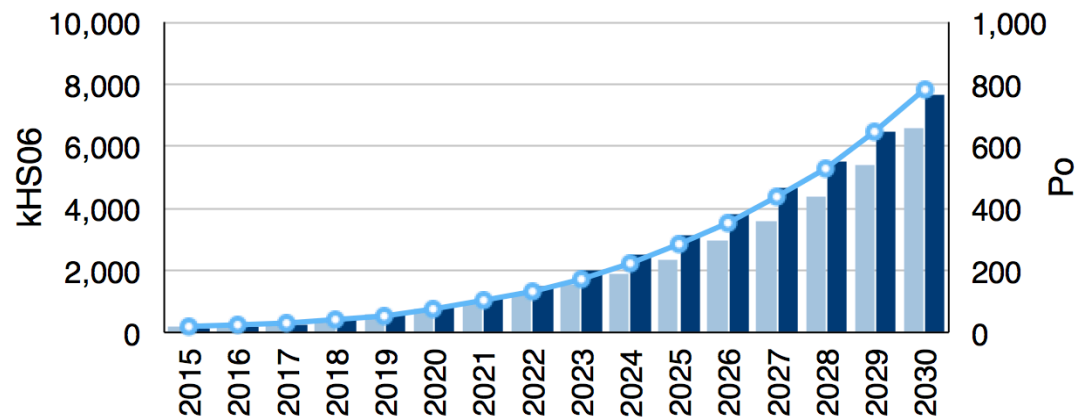


- ▶ Introduction : Le CC-IN2P3
- ▶ Impact du volume des données
- ▶ Impact de l'accès données
- ▶ Impact de la variété des données
- ▶ Impact de la variété des traitements
- ▶ **Futur au CC-IN2P3**

# Futur : Quelques « gros » besoins en ressources



Somme des besoins annuels de capacité pour ces 7 expériences

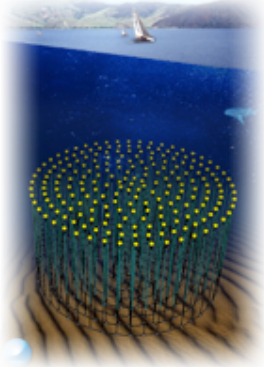


# Expériences Astroparticules futures

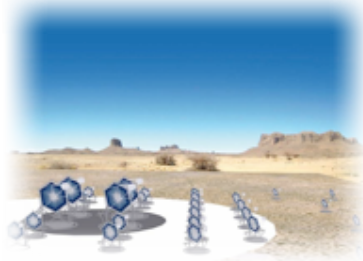
		En cours	Futur ?
<b>Energies extrêmes</b> <b>Origine</b> <b>Composition</b> <b>Anti-matière</b> <b>Matière Noire</b> <b>Gravité</b>	<b>Gamma</b>	<b>HESS-II / FERMI</b>	<b>CTA</b>
	<b>Neutrino</b>	<b>ANTARES</b>	<b>KM3NET</b>
	<b>Rayons cosmiques</b>	<b>AUGER / AMS02</b>	<b>JEM-EUSO</b>
	<b>Ondes gravitationnelles</b>	<b>VIRGO</b>	<b>Adv VIRGO / LISA / ET</b>
<b>Cosmologie : Matière Noire</b> <b>Energie noire</b> <b>Paramètres</b>	<b>Matière Noire</b> <b>SuperNovae, WL,...</b> <b>Fond diffus cosmologique</b>	<b>EDELWEISS</b> <b>SNLS / SNF</b> <b>PLANCK</b>	<b>EURECA</b> <b>LSST / EUCLID</b>
<b>Neutrinos : Majorana/Masse</b> <b>Mixing</b> <b>Désintégration du proton</b>	<b>v-less double <math>\beta</math> Decay</b> <b>Oscillation Neutrino</b>	<b>NEMO</b> <b>DCHOOZ/OPERA/T2K</b>	<b>SuperNEMO</b> <b>LAGUNA</b>

# Expériences Astroparticules futures

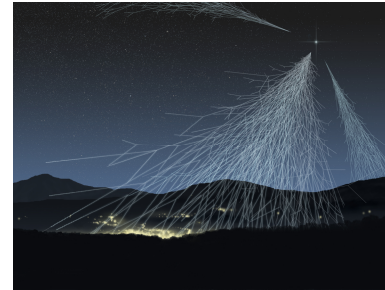
KM3NET  
high energy neutrinos



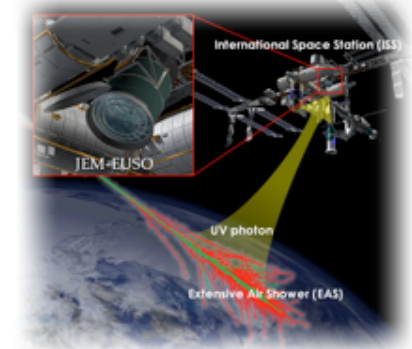
CTA : gamma rays



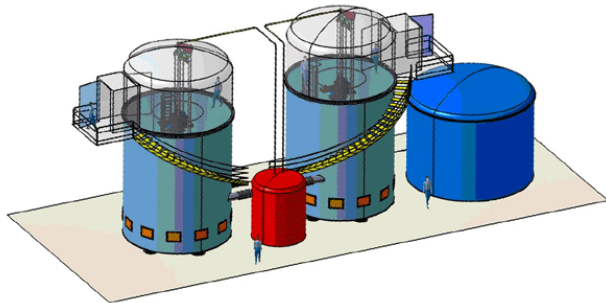
AUGER upgrade/extension



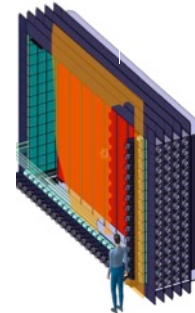
JEM-EUSO : extreme energies



EURECA : Dark Matter



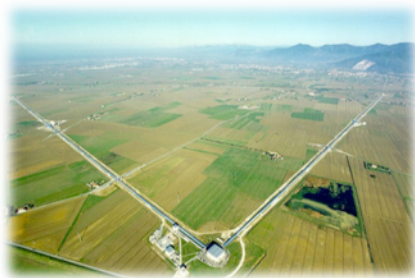
SUPERNEMO  
 $\nu$ -less double beta decay



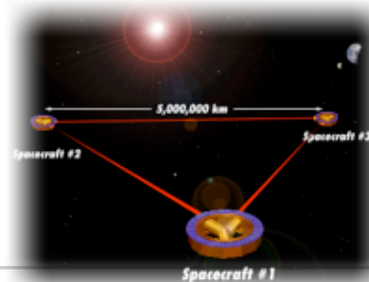
LAGUNA proton decay, neutrino



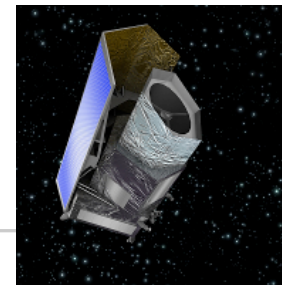
advVIRGO /ET Gravitational waves



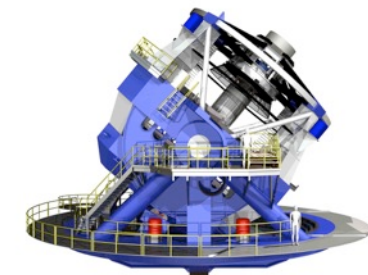
LISA gravitational waves



Euclid,  
dark energy



LSST, Chile  
dark energy



### ▶ Le calcul au CC-IN2P3

- Domaines de physiques → données hétérogènes
- Traitement et analyse de données expérimentales

### ▶ Impact des données

- Volume : utilisations de **bandes magnétiques**, ...
- Transferts : bons **réseaux** inter-sites, ...
- Traitements massifs : bon **réseau** interne, ...
- Variétés des données : Software **disques** adaptés, ...
- Variétés des traitements : Infrastructure de **calcul** adapté, ...

### ▶ Futur

- Avalanches de données pour le LHC et les astroparticules
- Nécessité de veille et d'évolution technologique