



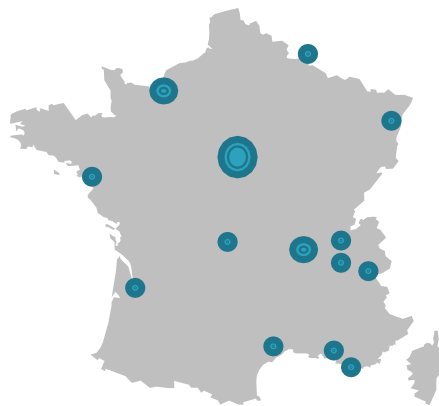
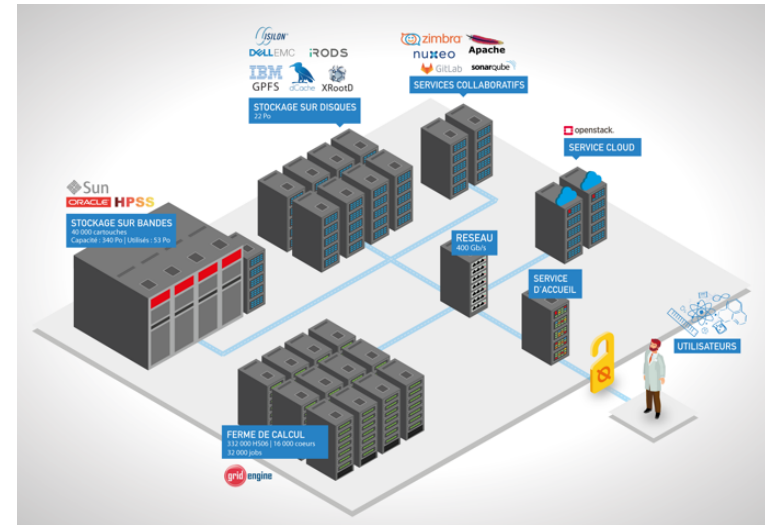
Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

FJPPL 2018

Centre de Calcul IN2P3 site report



- ▶ CC-IN2P3 missions
 - Massive data storage resources
 - Computing facilities
 - Common collaborative services



- ▶ CC is main contact hub for RENATER to organize and deploy network accesses.
- ▶ Responsible of the network entry points of 24 sites of IN2P3.

- ▶ News on Organisation
- ▶ Infrastructure
 - Computing and CPU
 - Disk Storage
 - Mass Storage
 - Network
- ▶ Summary

Changes in the global organization



Chargés de missions

Assistant de Prévention :
X. Canehan

Correspondantes formation :
V. Delebarre Dutruel

Réseaux métiers :
J-R. Rouet

Sécurité Informatique :
B. Delaunay

Urbanisation Salles machines :
X. Canehan

Comités

CHS **CU**

Direction

P-E. Macchi
Directeur

B. Delaunay
Directeur adjoint

C. Aulas
Responsable administrative

Deputy
Director



New Team

Domaine informatique

Applications	CCLab	Opération	Réseau	Stockage	Support	Système
J-R. Rouet Adj. C. L'Orphelin A. Bourges D. Cathala-Martinez E. Cervera B. Chambon P. Correia O. Drevon C. Evesque D. Mège J. Moutarde A. Perrier S. Reynaud T. Salanon L. Schwarz M. Walter	F. Suter A. Chai M. Khannouz B. Rigaud	F. Azevedo X. Canehan H. Cordier E. Fede N. Lajili P. Larrieu J-C. Soret	L. Caillat-Vallet T. Balde J. Bernier C. Rondelet	J-Y. Nief O. Aïdel P-E. Brinette Y. Calas Y. Cardenas P. Cheynet A. Dubois A. Georget Z. Goutali L. Tortay	D. Bouvet P. Calvat S. Gadrat T. Kachelhoffer A. Khalfa Q. Le Boulc'h R. Lemrani G. Marchetti E. Vamvakopoulos R. Vernet	M. Puel J. Babel C. Eloto R. Ferrand P-Y. Fontanière N. Fournials V. Hamar A. Khouder X. Niu F. Wernli

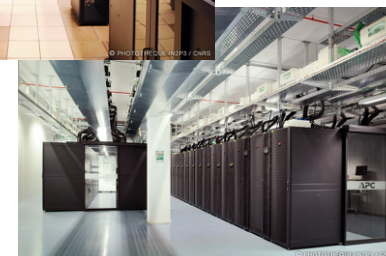
- ▶ Factual observation and incentive:
 - Increasing number of projects coming from European funding
 - They are short or mid-term projects with(or without) resources
 - segmentation in the resources and lack of communication between projects
 - Overlaps in the topics
- ▶ Try to overcome and benefit from by:
 - Global Coordination of our participation together with CNRS/IN2P3
 - Better coordination with other international Computing and Data centres
 - Better and more focused answer to the calls in line with our own strategy
 - Better optimization and integration of the HR and the R&D
 - Better investment return

► Portfolio:

- EGI-Engage:
 - Developments for Ops Portal (Nagios/GGUS), Monitoring resources (VAPOR)
- ARGO:
 - Monitoring services (e.g. Nagios)
- HNSciCloud Pre-Commercial Procurement to drive an innovative market place :
 - participating into assessment criteria.
- EOSCPilot: pilot for an open science cloud
 - Participating in gap analysis and setup of a pilot for specific use case.
- eTRIKS:
 - setup of a platform for biomedical translational research
- XDC: Computing e-infrastructure for extreme large datasets
 - Tests of ONEDATA for CTA use case
- EOSC-Hub:
 - Continuation of the development for Ops Portal and ARGO
- HPC-Europa3: Containerization for HPC
- ...

INFRASTRUCTURE

- ▶ 2 computing rooms for a total of 1 700 m².
- ▶ VIL1 (oldest) hosts mass storage and critical services.
- ▶ VIL2 (newest) hosts computing resources and disk storage systems.
- ▶ Network point of view :
 - VIL1 hosts the regional point of presence of RENATER.
 - Both rooms are connected with a 400Gbps link.
 - The production network is distributed across the two rooms.



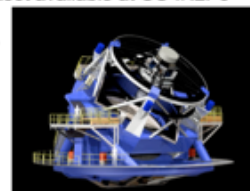
- ▶ Working for 90 experiments
- ▶ Tier1 for WLCG: Alice, Atlas, CMS, LHCb



LSST

Whole dataset available at CC-IN2P3

50% of the processing by CC-IN2P3
other 50% by NCSA



EUCLID

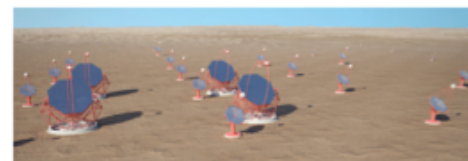
CC-IN2P3 is the French Data Center for processing and data management



dark energy and dark matter

CTA

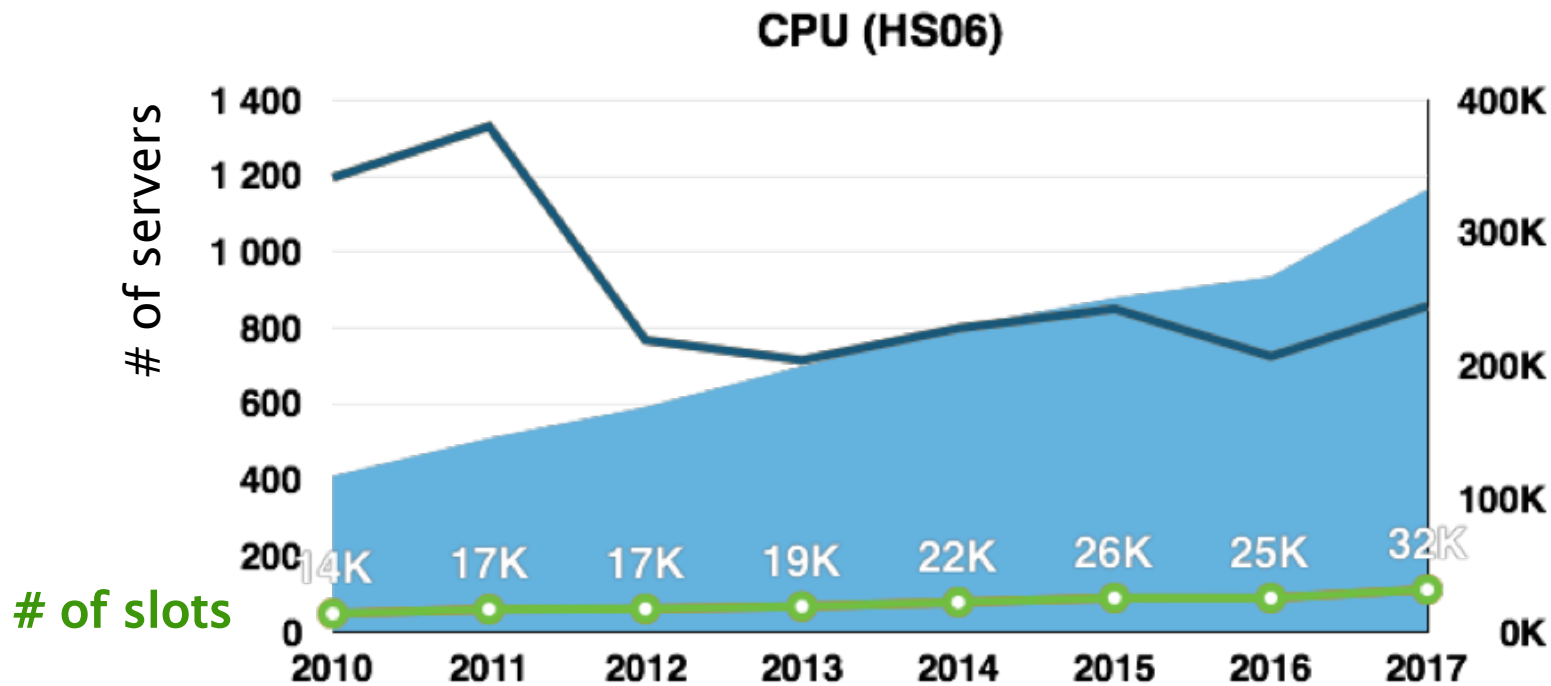
CC-IN2P3 should play a key role in the CTA data processing



Gamma rays

COMPUTING and CPU

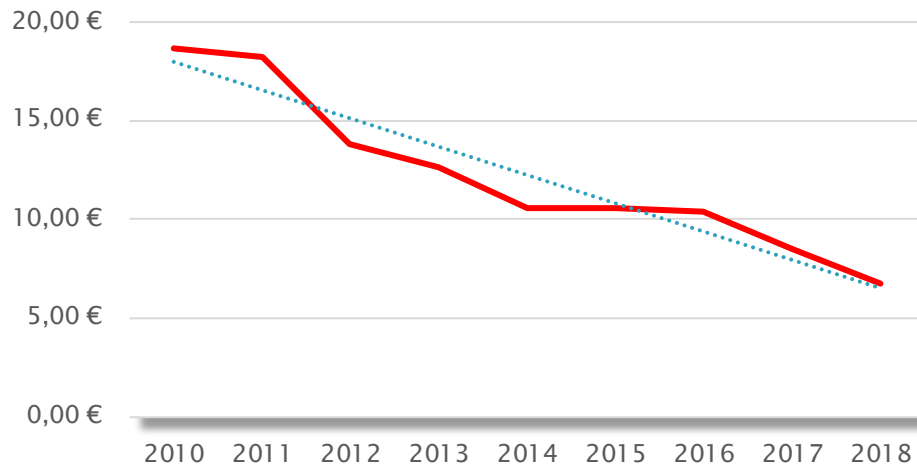
- ▶ CC-IN2P3 is providing 3 computing facilities in production.
 - Main: The High Throughput Computing (HTC) farm.
 - From 234 589 HS06 in 2015 to 319 045 in 2017 (+36%).



- ▶ The 2 other computing facilities:
 - Simba, a High Performance Computing (HPC) cluster.
 - Previous aging cluster replaced in June 2016.
 - +16% HS06, remains with 512 physical cores.
 - Relies on 16 DELL C6320 servers and an QDR IB interconnect.
 - Nala, a GPGPU cluster.
 - In production since September 2016.
 - Relies on new DELL servers C4130 with K80 GPU.
 - High speed interconnect using IB network.

- ▶ Activity on HPC and GPGPU:
 - After one year, still a very low use comparing to the HTC farm.
 - Some HPC users are looking at multi-cores HTC jobs.

CPU - COST EVOLUTION- €/HS06



2011	2012	2013	2014
18,24 €	13,86	12,68	10,62

2015	2016	2017	2018
10,62 €	10,37 €	8,49 €	6,76 €

- ▶ Even if the decrease is up to 20,4 % between March and December 2017, the trend should be more like 12,5 %/year.
- ▶ 6,76€/HS06 under EL6 where 6,12 €/HS06 under EL7 (-27,9 %)

- ▶ The DELL partnership started in 2012 has to be renewed.
 - 50 servers QSERV
 - 2 servers for data storage technologies assessments
- ▶ This year, the DELL EMC partnership is going to be renewed.
 - LSST always in the scope
 - Radiotherapy type-of-computing also in the target

Univa Grid Engine 8.4.4:

- ▶ Resource management and scheduler for the 3 platforms.
 - License acquired for 16 000 cores (15 776 cores in use).
 - Maintenance and license contract renewed in April 2017 for 2 years.
- ▶ Continuous improvement on batch scheduling.
 - Give back freedom to the batch system by reducing the number of manual adjustments of the scheduling.
 - Filling the batch farm is similar and acceptable for less efforts.
- ▶ Change in the HR organization:
 - Split between the maintenance of UGE platform (system/infra team) and its operation (operation team)

- ▶ HTCondor as an alternative to Univa Grid Engine ?
 - Keep the control on financial and human costs.
 - A first study conducted in 2016.
 - Proof of concept targeted at the end of 2017, first for the HTC grid part.
- ▶ Transition from EL6 (Scientific Linux) to EL7 (CentOS).
 - HPC and GPGPU clusters are running EL7.
 - EL7 will be the default OS for HTC in April 2018, targeted to have 100% in June.
- ▶ Singularity as a container solution.
 - Can be an answer to the problematic migration from EL6 to EL7.
 - Software still young, evolving every day.
 - Currently deployed, waiting for the experiments to validate the solution in production.

▶ Computing cloud platform based on Openstack.

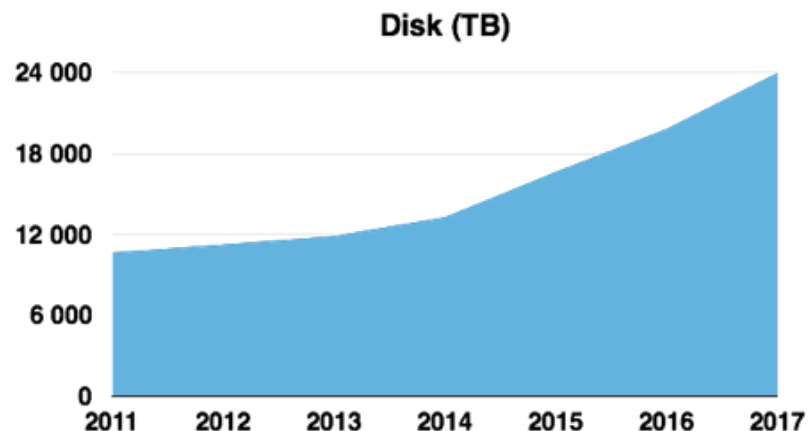
- Used only for prospective and assessment purposes.
- Test bed of the Synergy INDIGO-Datacloud project.
- Used by Atlas as opportunistic computing resources.
- No planned take off.



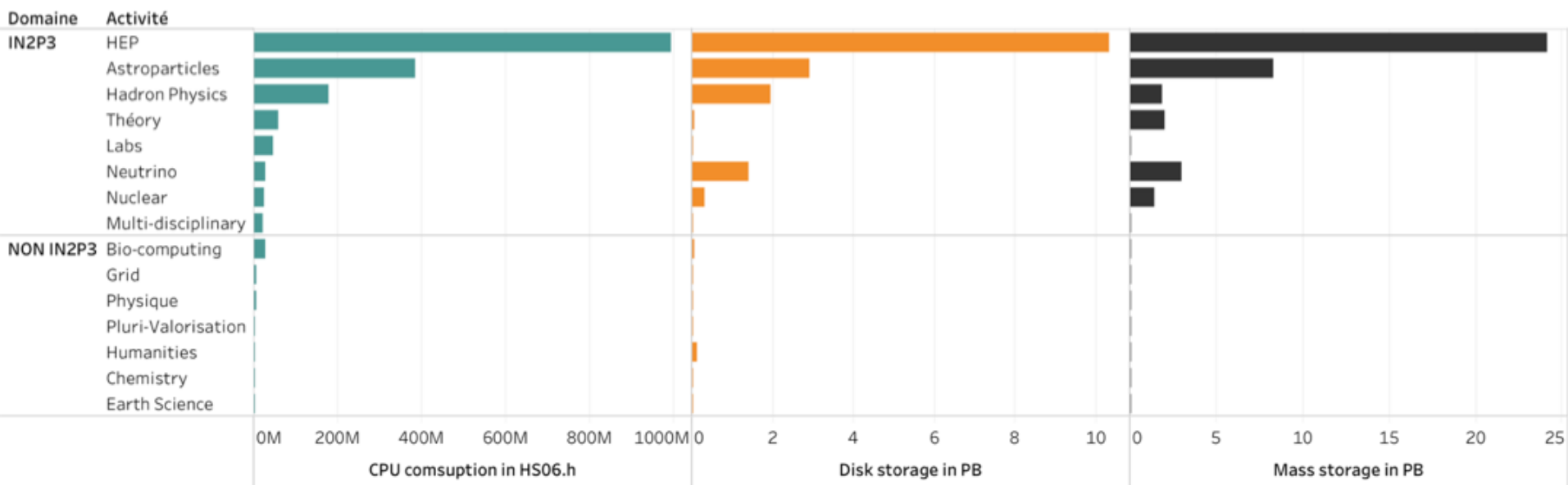
DISK STORAGE

Disk storage, what's new ?

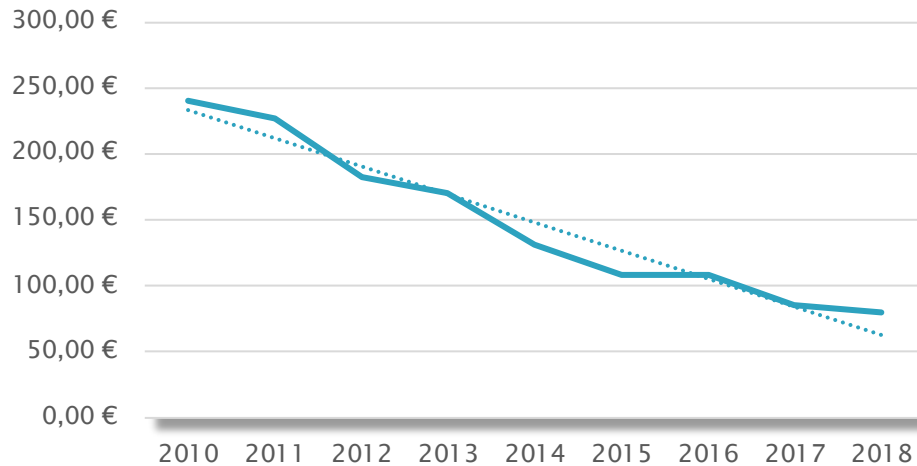
- ▶ Current systems grew up.
 - DCACHE (13,8PB), XROOTD (6,4PB), IRODS (1,8PB), SPS/GPFS (2,6PB).
 - Server bandwidth increased from 2 to 10Gbps.



2016 Ressource Usage



DISK STORAGE – COST EVOLUTION – €/TB



2011	2012	2013	2014
227,50 €	183,33 €	170,00 €	131,25 €

2015	2016	2017	2018
107,92 €	108,28 €	85,71 €	80,07 €

- ▶ No definitive prices for 2018, still has to be discussed with DELL by the end of this year.
- ▶ Decrease 13,72 %/year.

- ▶ Database services.
 - NO SQL database provided (mongodb).
 - MySQL service increased while Oracle RDB decreases.
- ▶ Framework with Hadoop model, spark cluster.
- ▶ Progressing towards end of AFS.
 - Since 2016, AFS group space and SPS/GPFS have been merged.
 - 2017, throng space and user software repository moved on ISILON cluster.
 - 2018, home directories are planed to be moved on ISILON cluster.
 - End of 2018, AFS will only used to access other cells.

- ▶ CEPH
 - A proof of concept has been started in 2017.
 - Storage backend for the cloud platform Openstack.
 - CEPH-FS, S3 API, will follow in the study.

- ▶ IBM/GPFS
 - Study alternative solutions In order to control financial and human costs.
 - Lighten the administration of the service.

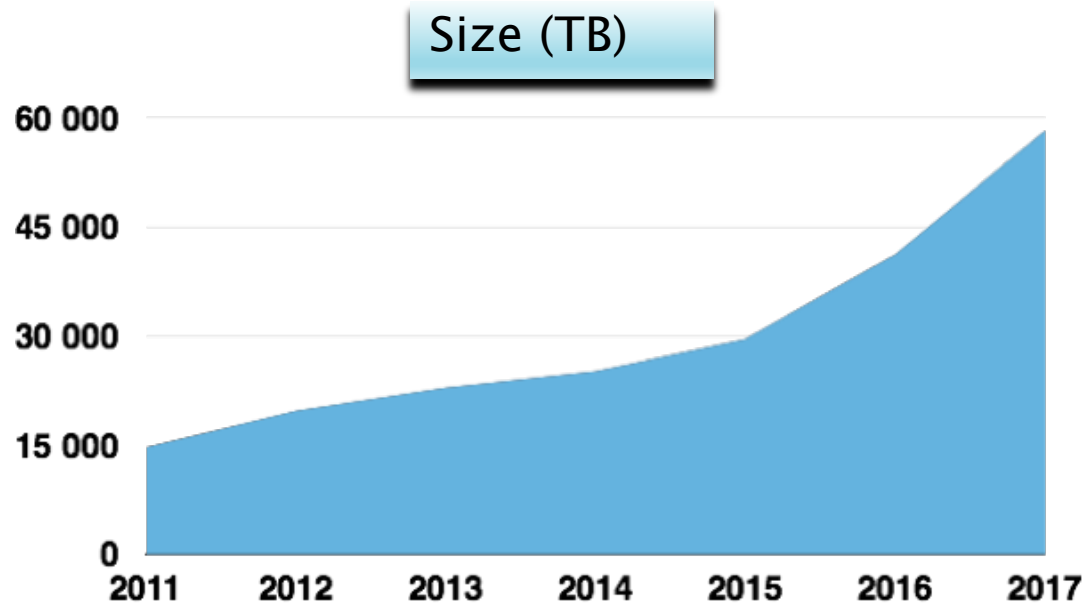
- ▶ IBM/TSM, transferring backup system management to the storage team, studying alternatives. Problems to renew support.

- ▶ XDC project, OneData a storage middleware for CTA use case.

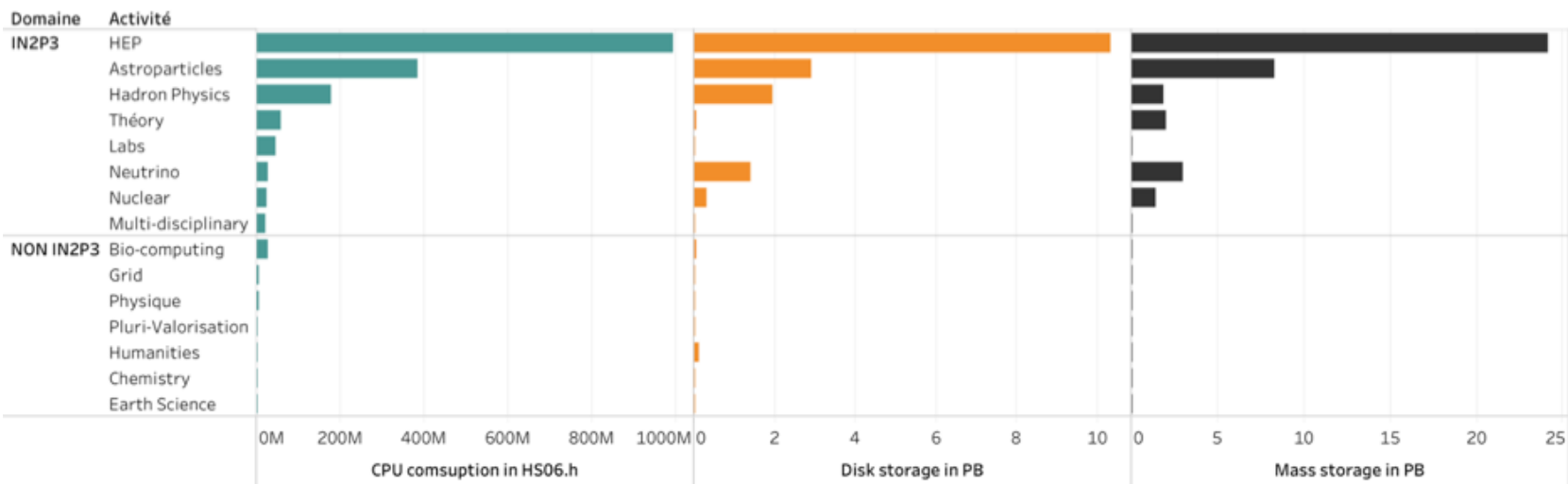
- ▶ Data Management Plan
 - To be applied to all experiments: began in 2017 and will be applied to all experiments by the end of 2018.

MASS STORAGE

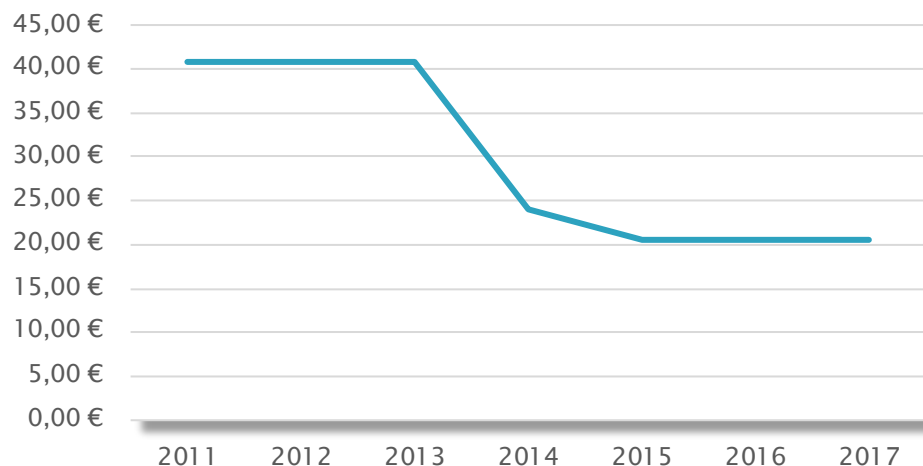
More details in Pierre-Emmanuel's presentation



2016 Ressource Usage



TAPE STORAGE – COST EVOLUTION – €/TB



2011	2012	2013	2014
40,80 €	40,80 €	40,80 €	24,00 €

2015	2016	2017	2018
20,54 €	20,54 €	20,54 €	20,54 €

- ▶ The tape cartridge price went from 204,00€ to 174,60€ in 7 years.
- ▶ The price decrease is mainly due to the replacement of the tape drives.
 - T10KC (5TB) to T10KD (8,5).
 - Reuse the same cartridges.

NETWORK

- ▶ Increase the throughput for internal and external links in 2018.
 - No more modification required on the current infrastructure to cope with future requirements.
 - RENATER, internet access upgrade to 2 x 100 Gbps (active/backup).
 - LHCOPN upgrade to 40 Gbps.
 - 100 Gbps dedicated link for future projects with IDRIS (CNRS institute of supercomputers infrastructure)
- ▶ Work in progress on firewalling.
- ▶ Installation of VRF routing for CC

SUMMARY

▶ Network

- Infrastructure ready to cope with experiment requirements, waiting for RENATER to upgrade external links.

▶ Computing

- Ongoing work on HTCondor and Singularity.
- Migration plan to EL7 defined and launched.

▶ Storage

- Replacement of AFS on going.
- Concerns with IBM contracts renews (GPFS & TSM).
- Mass storage system, technology assessments and plans.

▶ Applications and services

- We have to work on user experience, providing tools to access and monitor resources. Identity management and user portal are the foundations of these tools, and must be completed by others.

- ▶ We have good tools to purchase hardware and software, having these public procurements is a real gain.
- ▶ The cost trend is encouraging even if we of course expect better prices. A question is pending concerning the disk storage cost in 2018.
- ▶ The cost evolution of the tape technology is flat for years. Still competitive with the disk storage technology ? We will try to answer to this question in 2018.
- ▶ One part of the strategy is to reduce the recurring costs (mainly maintenance contracts) to provide more resources.

