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

Panorama des infrastructures des projets de l'institut

Conseil Scientifique IN2P3

Disclaimer

I am computing coordinator at CC IN2P3

- In charge of coordinating the experiments requirements and the resources at CC-IN2P3.

I don't have a complete view of the resources provided by other IN2P3 laboratories and their constraints.

Not all the detailed relationship and specificities between computing infrastructures and applications can be highlighted on this talk.

Overview

Computing, storage and service at IN2P3 institute.

- CC-IN2P3.
- Platforms at our laboratories.
- External platforms.
- Infrastructures of our infrastructures.

Computing and storage resources for applications.

- IN2P3 computing infrastructure on experiments computing models.
- Our infrastructures on international landscape.

Evolution and opportunities.

- Growing of infrastructures.
- Investigate news approachs.

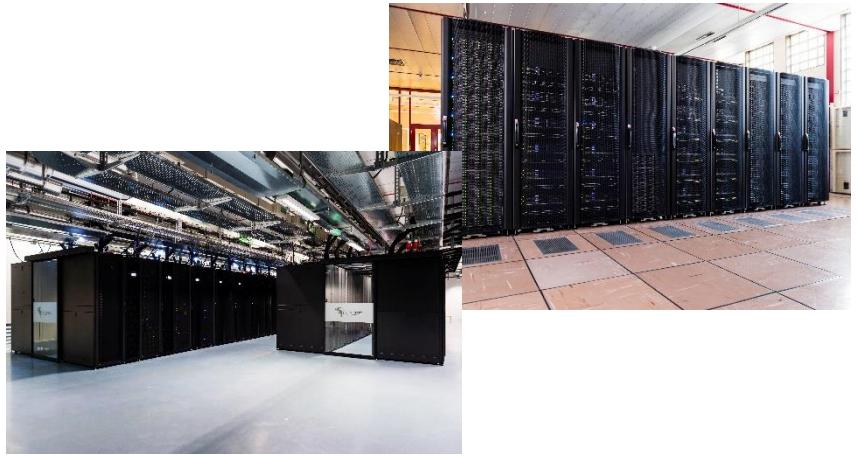
Summary

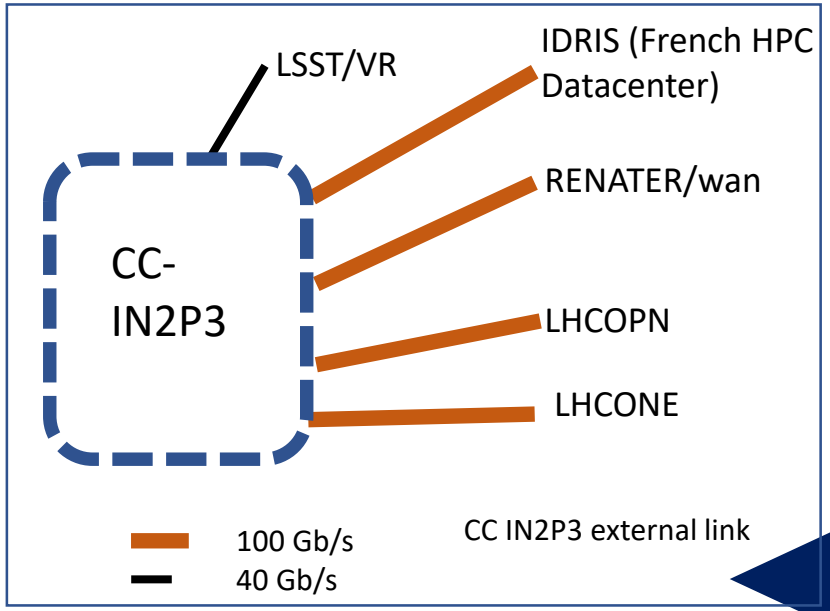
CC-IN2P3

Mission : Providing data processing and services for experiments supported by IN2P3.

- ~85 collaborations supported at CC-IN2P3
  - HEP, Astroparticle physics, Nuclear physics, Cosmology, …
  - A small fraction < 5% of computing time open to other sciences.
- 84 agents mainly engineers.

- Computing centre with a focus on data aspect.
- 1500 m2 (2*750) of computing room.
- Dedicated network links
- 1.2 MW of electric power
- High availability





Site	Availability		Reliability	
IN2P3-CC		99.92%		99.98%

ATLAS Avail/relia last 6 months

CC-IN2P3 : Computing

Shared facilities

- HTC (High Throughput Computing)
  - Main computing resource
  - ~60 kslot : ~800 000 HS06
- HPC (High Performance Computing)
  - Limited usage
  - 512 cores
- GPU (Graphics Process Unit)
  - 68 GPU V100

Specifics/dedicated facilities

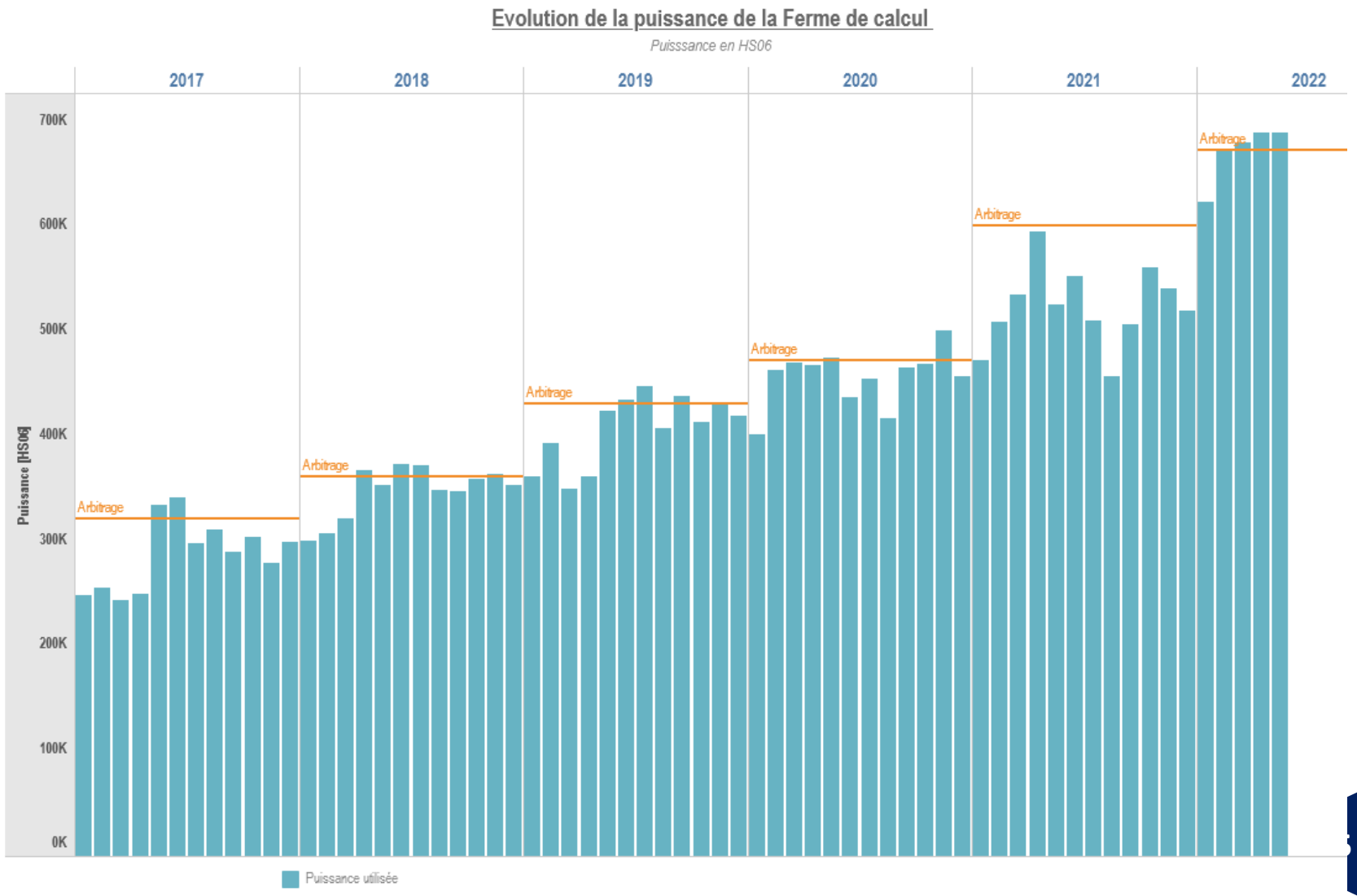
- Virtual hosts, dedicated GPU,…

Challenges to address :

- Providing resources requested during the collaboration lifetime.
- Adapting computing farm to the new requests (new devices).

22/06/2022

Computing evolution at CC IN2P3



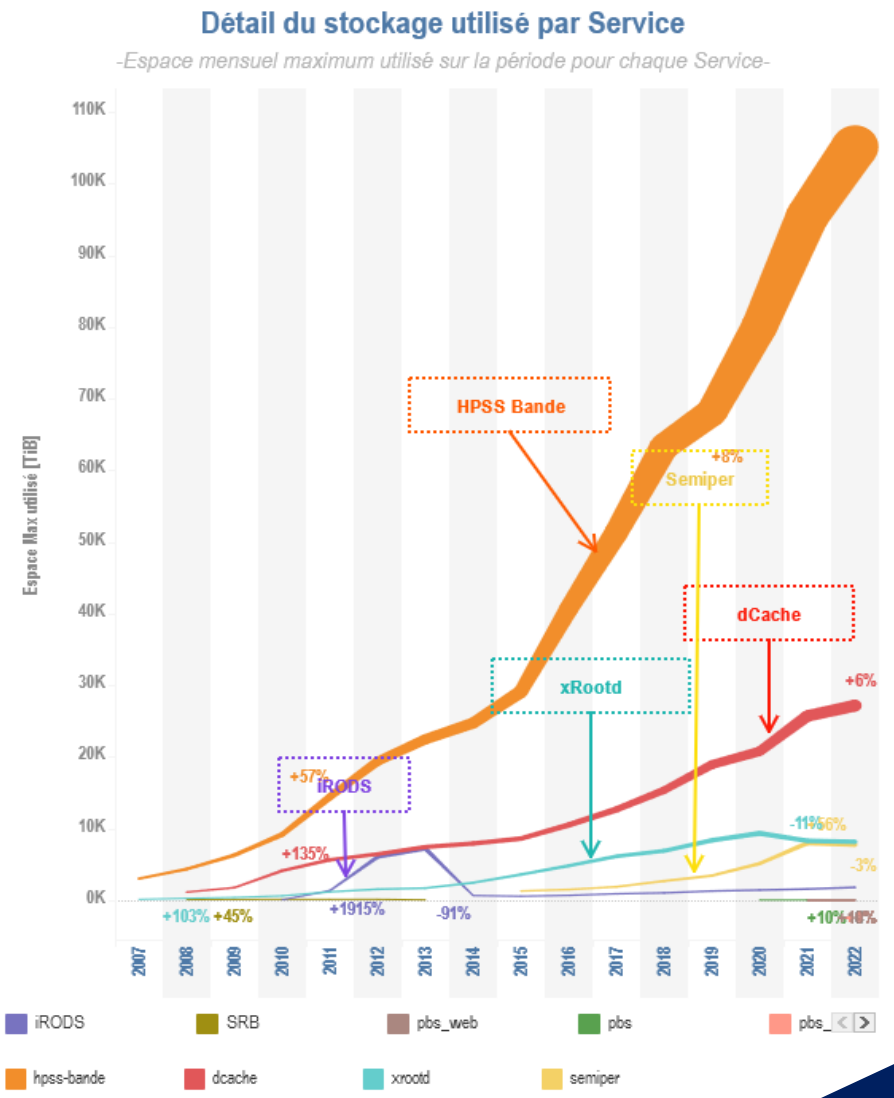
CC-IN2P3 : Storage

- Disk infrastructure
  - High capacities ~ 45 PB (dCache, Irods, Xrootd)
  - High performance ~ 15 PB (Ceph-FS, Isilon, GPFS)
  - Oriented service : Ceph
- Tape infrastructure
  - ~130 PB of data on tape
  - A large set of experiments
- Backup infrastructure
  - Double copy (~ 2*3PB)
  - Mainly to store IN2P3 laboratories backup and some experiments data.
- Other
  - Ceph for cloud
  - Database : Oracle, MariaDB, postgres, nosql,…

Challenges to address:

- Providing resources requested during the collaboration lifetime.
- Be able to ingest the data volume deluge of the next years and be able to serving them.

Data storage usage evolution at CC-IN2P3



CC-IN2P3 : Services

CC-IN2P3 provides also a large set of services to the users/collaborations.

Some of them are shared but some of them are dedicated to a collaboration and the main part of the time are a critical services for the collaboration.

Two ways to provide theses services are available :

- CC-IN2P3 provides infrastructure only, services are managed by user/collaboration
  - Mainly via internal cloud infrastructure/ kubernetes cluster.
    - Web, VObox,…
- CC-IN2P3 provide infrastructure and management of services
  - Via dedicated infrastructure or virtual machine on internal cloud
    - Job submission (Dirac), data management tool, Databases, repository (cvmfs stratum 0),….

Challenges to address:

- Having/keeping/developing the skill to provide this service.
- Keeping a close relationship with experiments computing team.
- Watching out about new services : ex Analysis facilities

Platform on IN2P3 laboratories

Some IN2P3 computing resources can be found outside the Computing Center, on labs

4 Platforms of computing certified by IN2P3

- Provide computing resource.
- Provide storage resource.
- Human resource for support and skill sharing.

- FACe (APC)
- Virtual Data (ORSAY)
- MUST (LAPP)
- SCIGNE (IPHC)

New certification on going (CPPM)

IN2P3 laboratories resources

- Some of them are Tier2 for WLCG (LPC Clermont, CPPM, Subatech) and provide an important part of computing/storage resources.

Other platforms

Specifics facilities can be found outside of the IN2P3.

This is mainly High Performance Computing facilities requested by some IN2P3 activities

- QCD computing
- Machine learning (on large system)
- …

The natural infrastructure to use is the French CNRS HPC facilities “Jean Zay” hosted at IDRIS.

Some facilities on universities can also be used rarely.

Challenges to address :

- Simplify access to this type of infrastructure to our users.

Core Infrastructures

To provide compute, storage and services you need to have some core infrastructures which are key points to define the computing infrastructures capabilities.

- Computing room/datacenter.
  - Define the usage and next decade capability of evolution.
  - Define the efficiency of the computing/storage.
  - Define a large part of the cost (electric power).
  - Define the level of redundancy which can be satisfied.
  - Challenges to address : Retaining control on our datacenter
- Connection to the partners across the world (aka Network):
  - Connection to the data provider (experiment site).
  - Connection to the user/collaboration.
  - Implication on computing model today and future (mainly true on distributed computing).
  - Network for French research is provided by our partner RENATER (NREN). We have to deal with that.
  - Challenges to address : Helping/encouraging/allowing RENATER to satisfy our network requests.
- And human skills to do that.

IN2P3 computing infrastructure for collaborations

Three kinds of usage of our computing infrastructure can be highlighting.

First one : When the computing model of the collaboration is based on the usage of a large scale distributed infrastructure(Grid).

- LHC, JUNO, KM3net, Belle2,… ~all HEP collaborations

Second one : When the computing model is build around a small set of sites.

- LSST/VR, Virgo, Euclid…..

The last : When the usage of the computing is focused on only one site.

- All collaborations on which the computing resource request can be satisfied by one site.
  - Nuclear, theory,…

IN2P3 computing infrastructure for collaborations

On distributed computing models (worldwide scale) : GRID

- Commonly a MoU (IN2P3/Collaboration contract) define the computing commitments.
- Collaboration contact person is a team = “computing team”
- IN2P3 (and CC) are often asked to participate to these infrastructures.
- CC-IN2P3 is considered as a stable and performant site.
  - Tier 1 for WLCG : stored a part of the raw data and provide ~13 % of the Tiers 1 resources requested by applications (Alice, ATLAS, CMS, LHCb)
  - 15 % of the Belle2 Raw data at CC IN2P3.
  - Solicited to store a part of the JUNO, KM3NET,… raw data.
  - Involved on simulations and analyses of these collaborations.
  - Providing collaboration main services (Dirac for CTA, DB for HyperK,…).
- Others IN2P3 labs can also be involved on this distributed infrastructure.
- IN2P3 visibility depends of
  - Respect of our commitment with a significant level of participation on the resources.
  - Keeping a quality/performance that allow collaboration to use us.
  - Keeping the IN2P3 infrastructure in phase with the computing models.

IN2P3 computing infrastructure for collaborations

IN2P3 computing as main partner of a collaboration.

- CC-IN2P3 is defined as a keystone of the computing model of the collaboration
  - Collaboration looking for a specificity at the site
  - LSST and Virgo : 50 % of the data at CC-IN2P3.
  - Provide some specifics services for the collaboration.
- More a relation per to per with the collaboration and one or two partner sites.
  - Sometimes require to deploy a specific infrastructure.
  - This kind of model could evolve to a distributed model depending of the collaboration growth.
- IN2P3 visibility depend of
  - Respect of our commitment (critical for the collaboration).
  - Ability to provide a specific/new infrastructure.
    - Ex: Infrastructure for AI/ML, multi-messenger service, non standard workflow,…

IN2P3 as main infrastructure for the collaboration.

- A close and strong relation between collaboration and infrastructure.
- To the opposite to the previous cases, the main contact person is often a human and not a « computing team ».

- IN2P3 visibility depends of
  - Ability to propose and provide the best approach to satisfy the request.
    - A real issue because the landscape of the requests can be wide.
  - Ability to answer to the request with a solution enough open to be able to move to more distributed models.
    - Sharing skill and knowledge with the collaboration.

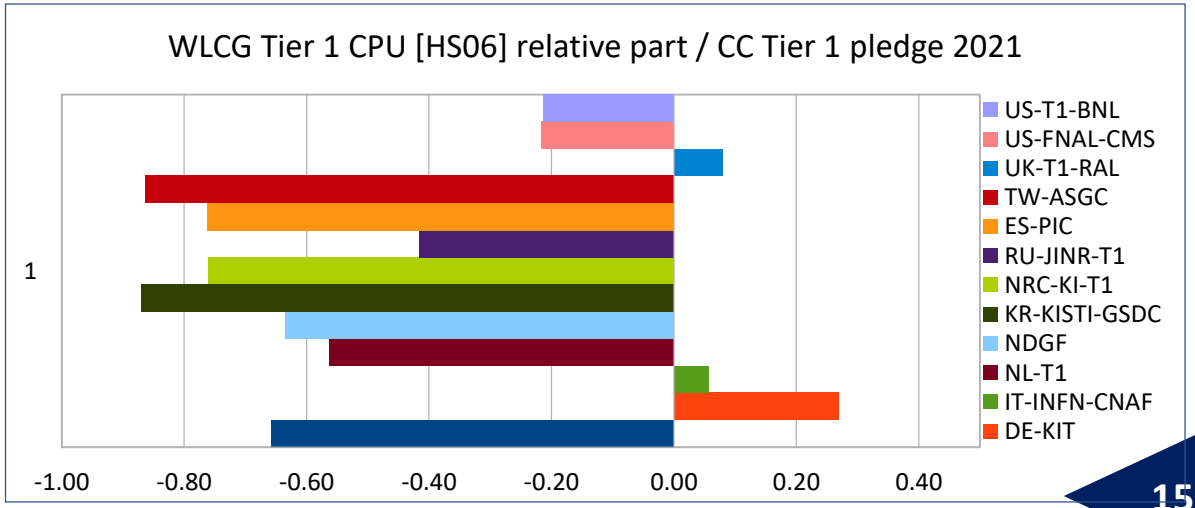
National/International landscape

Landscape of nation datacentre for sciences (Cf : stratégie nationale des infrastructures de recherche)

- TGCC
- CINES
- IDRIS
- CC-IN2P3

IN2P3 on international landscape.

- Tier 1 for WLCG
- Tiers 2 for WLCG
- Regional center for Belle2/JUNO.
- Main data center for Virgo/LSST.
- Datacenter for a whole set of nuclear experiments.
- End user facilities.



Infrastructure growth

We have to consider whole lifetime of the collaboration (sometimes many decade).

- A multi year engagement between IN2P3 and collaborations concerning the computing (MoU) has to be established.

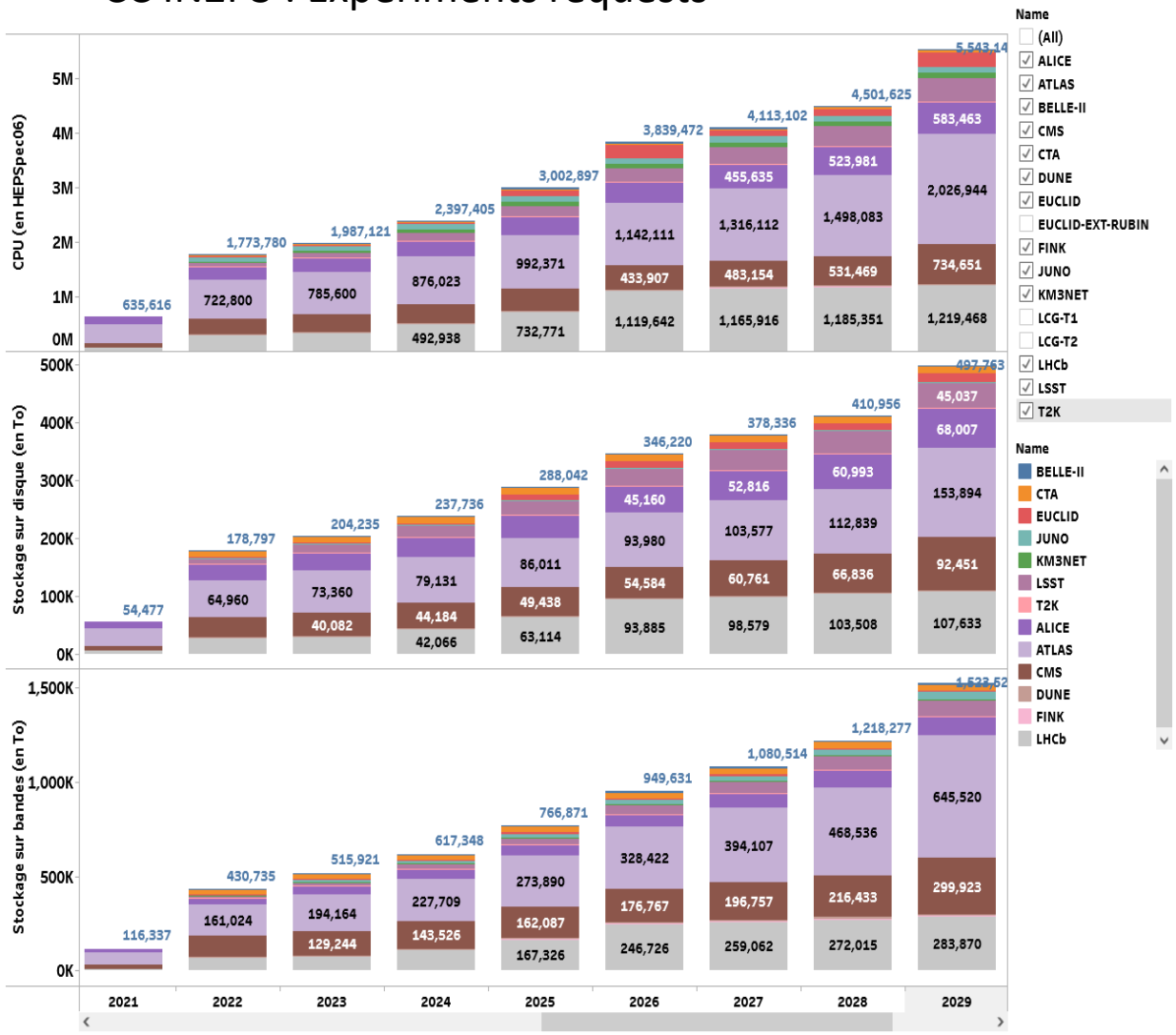
Concerning the budget many aspects have to be considered.

- Investments costs
- Running costs
- Operational costs

New technologies investigation/deployment

- More FTE

CC IN2P3 : Experiments requests



Infrastructure growth

One opportunity to cover some item about infrastructure growing is the FITS (CNRS Federated IT services for Research Infrastructures) project :

Target : Federate CC IN2P3 (HTC) and IDRIS (HPC) , respecting their specific mission , via a distributed infrastructure for data storage.

Project approved and funded (period mid-2021-mid 2029)

In practice

- A new computing room at CC (an upgrade at IDRIS)
  - To replace the old one.
  - Increase the efficiency of the computing room -> better PUE -> better running cost
- A sharing storage between two sides that can be used by the HTC and HPC facilities.
- 4 use cases identified : LSST, HL-LHC, Soleil (synchrotron), French Biology Institute

In the spirit

- A context to increase the collaboration between CC-IN2P3 and IDRIS in order to allow the usage of the IDRIS HPC computing resource easily by the IN2P3 collaborations.
  - A new type of infrastructure for our collaborations.

Investigate new approach

Solving next years for achieve computing request R&D activities on software : see next talks

But a part of the challenges is also on the data access performance and how the data is accessed by applications.

These R&D are covered by many projects ESCAPE or DOMA (Data Organization, Management and Access) and are based on an evolution of the storage infrastructure.

Aim is to built a storage infrastructure for scientific data as a federation of storage (aka datalake).

From application point of view this storage will be an infrastructure with different qualities of service (availabilities, performances, price,…).

This federation of storage (built on many datacenter) could

- Decrease the operational aspect on data management
- Simplify the data access for users and applications workflow
- …..

Summary

IN2P3 computing infrastructure are involving on every computing model of our collaborations.

Our computing centre (CC-IN2P3, platforms and sites) are engaged with a major role on the computing an data management of collaborations.

- Productions
- R&D and infrastructure evolution

Skills and technical capability of our teams are recognized.

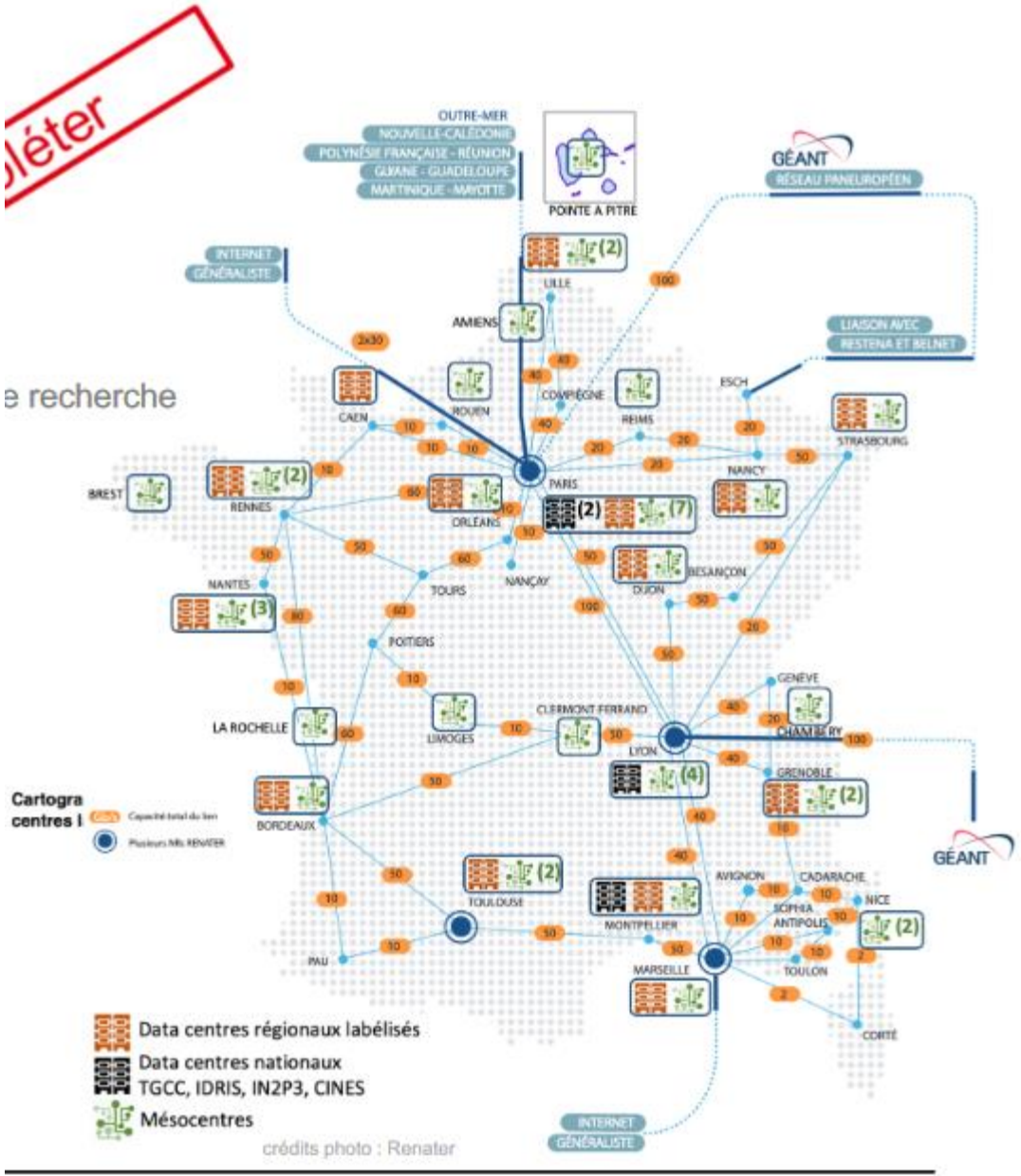
Two main challenges

- Providing today and tomorrow infrastructures requested by collaboration.
- Adapting infrastructure to the requests.
- This both challenges require time and strong relationship with collaboration

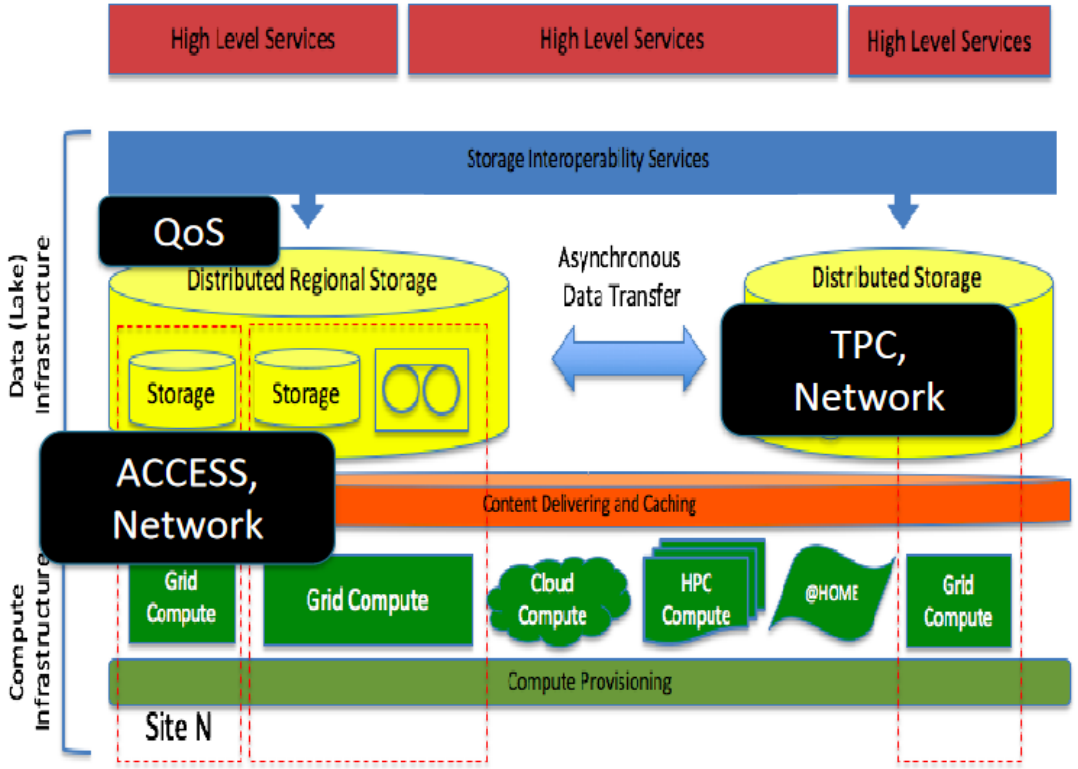
Question

Backup

National datacenter



Backup



DATALAKE