

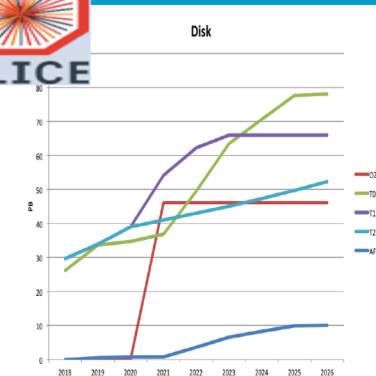
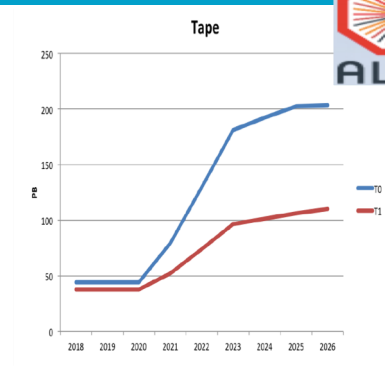
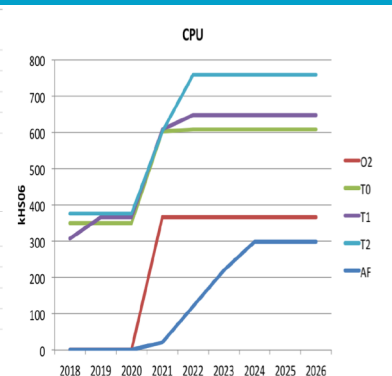
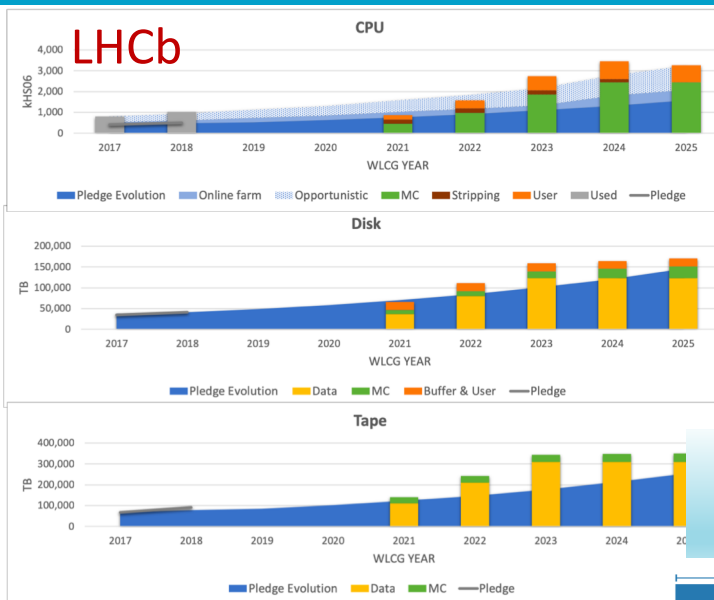
Sustainable computing models for IN2P3

Ghita Rahal, CC-IN2P3

The game players

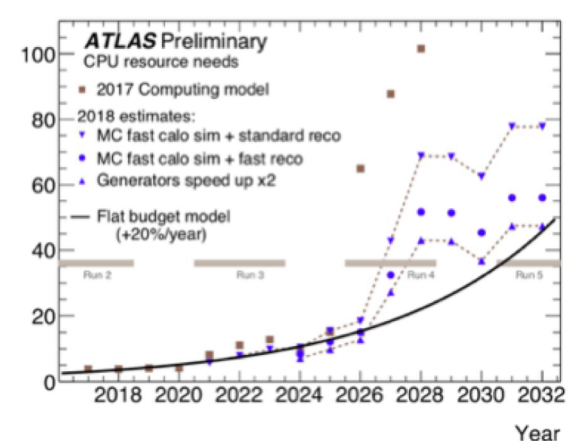
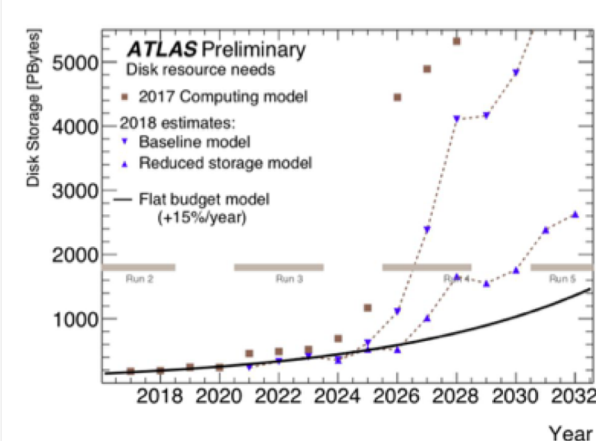
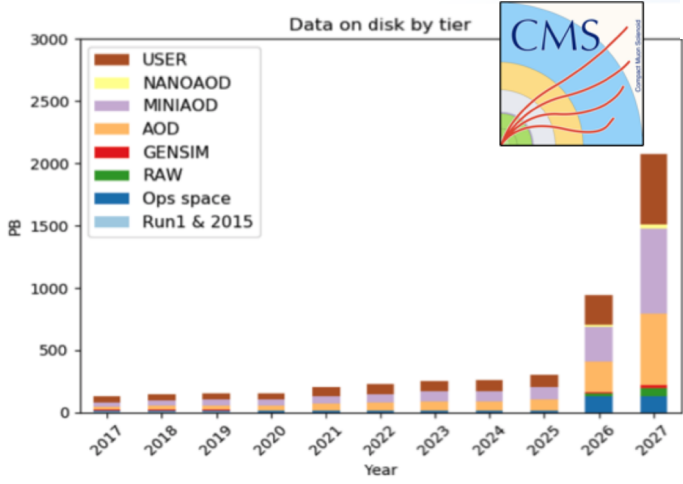
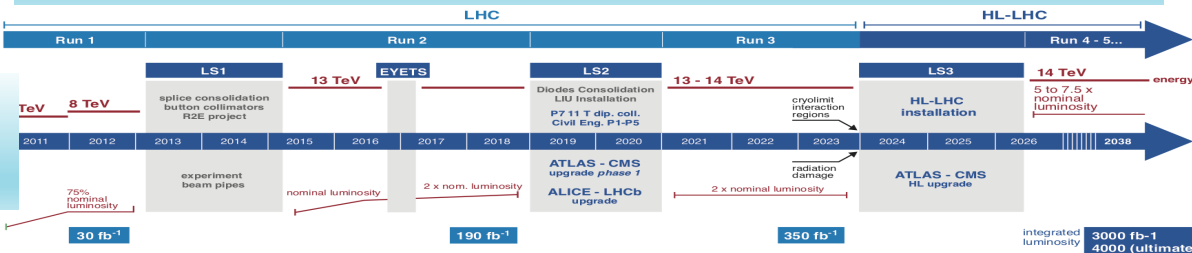
Les communautés scientifiques

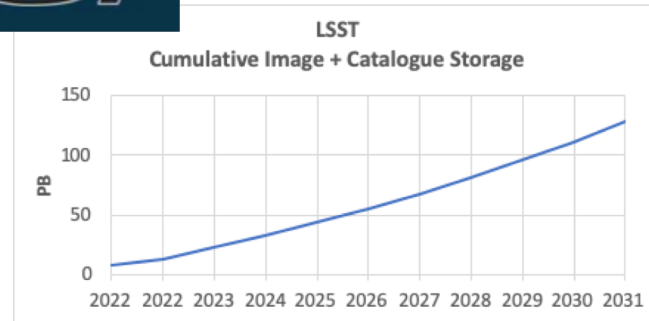
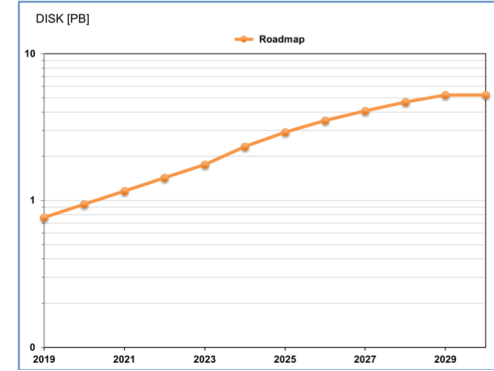
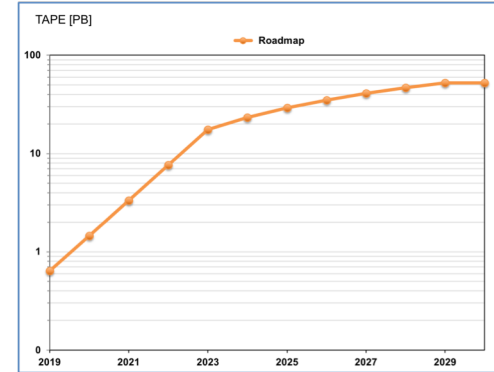
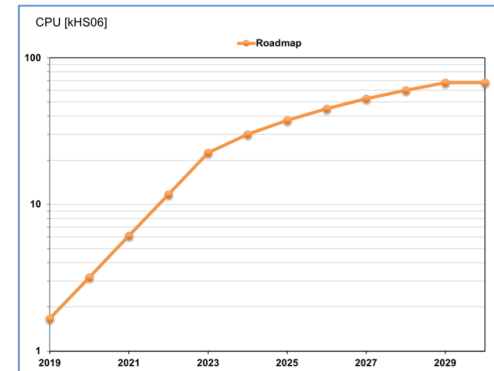
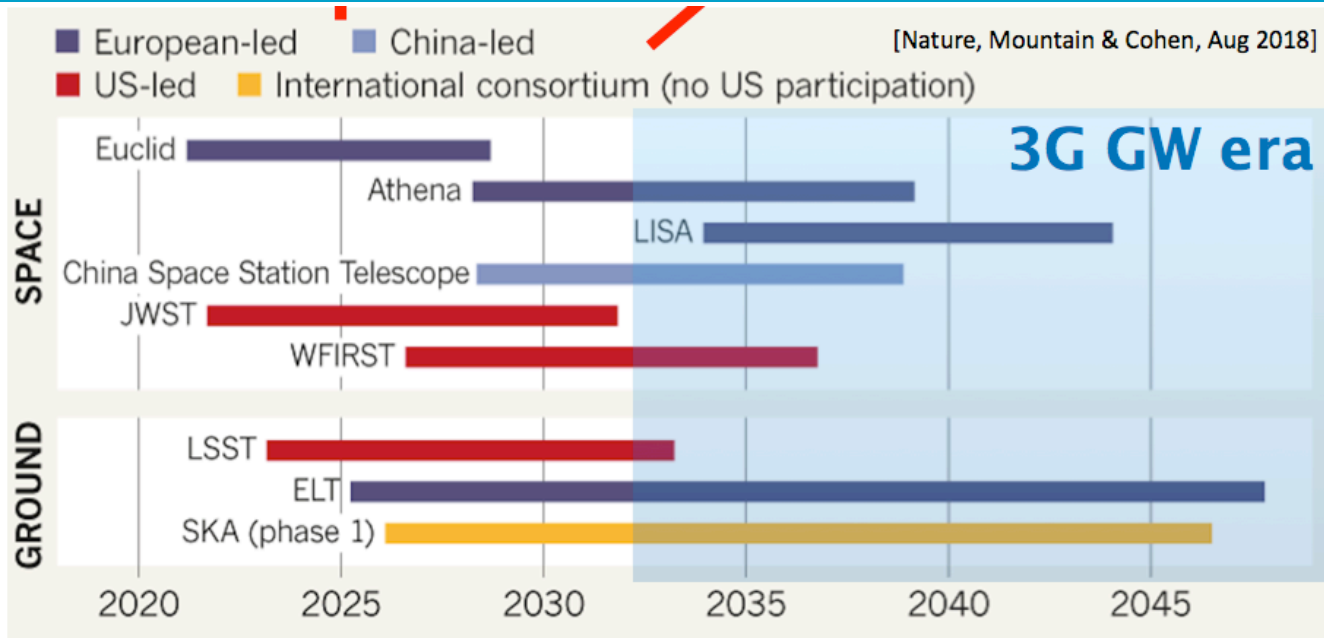
Le LHC dans le futur: projections



Run III: augmentation réelle mais maîtrisée puis explosion à l'après-2025

Run III: déjà quelques EB de données





Data and compute sizes:
 Final volume of raw image data = 60 PB
 Final image collection (DR11) = 0.5 Exabytes
 Final catalog size (DR11) = 15 PB
 Final disk storage = 0.4 Exabytes
 Peak number of nodes = 1750 nodes
 Peak compute power in LSST data centers = about 2 PFLOPS



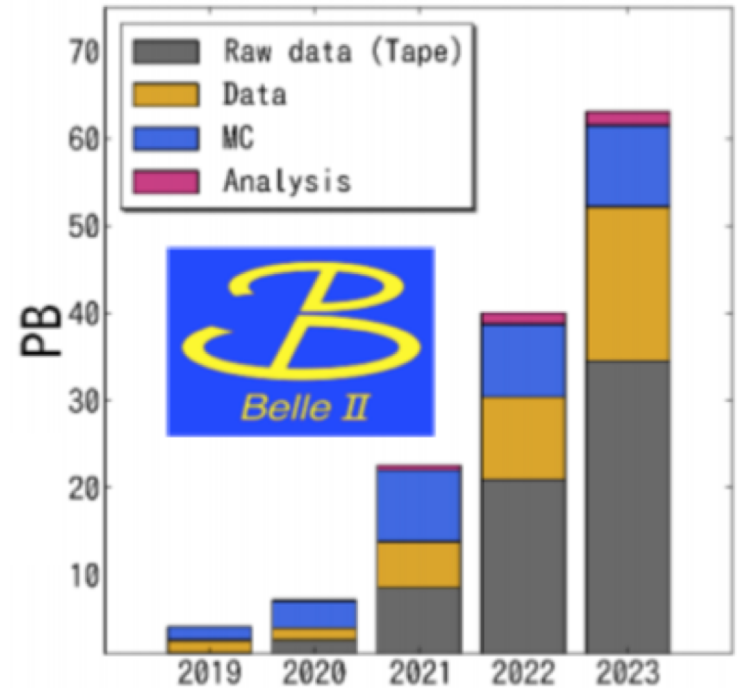
DUNE foresees to produce
~70PB/year in the mid 2020s



~20-aine de PB/an, et >1000-iers de cœurs/an En cours d'étude.

Ressources importantes dans le futur pour le calcul et le stockage.
On rentre dans l'ère de l'Exabyte.

Storage



Y. Kato @ HOW 2019

Le paysage numérique

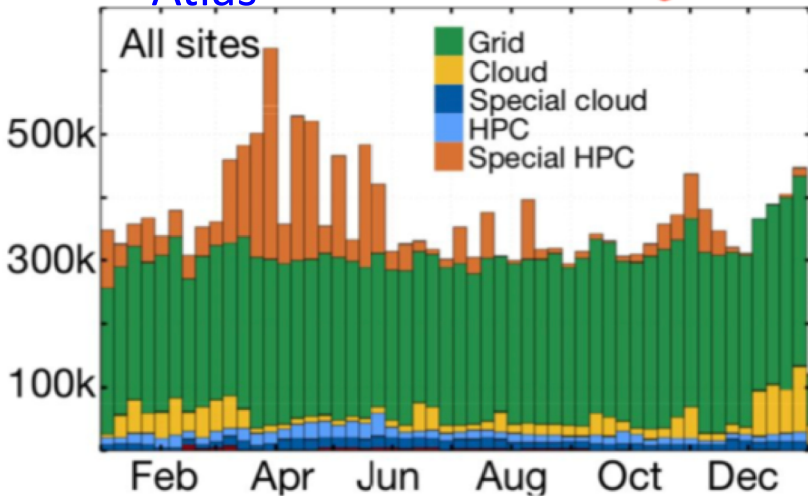
Calcul et Stockage

Plateformes hétérogènes de Calcul: Grille, Cloud, local

Provenance des ressources multiples: public, commercial, opportuniste

Atlas

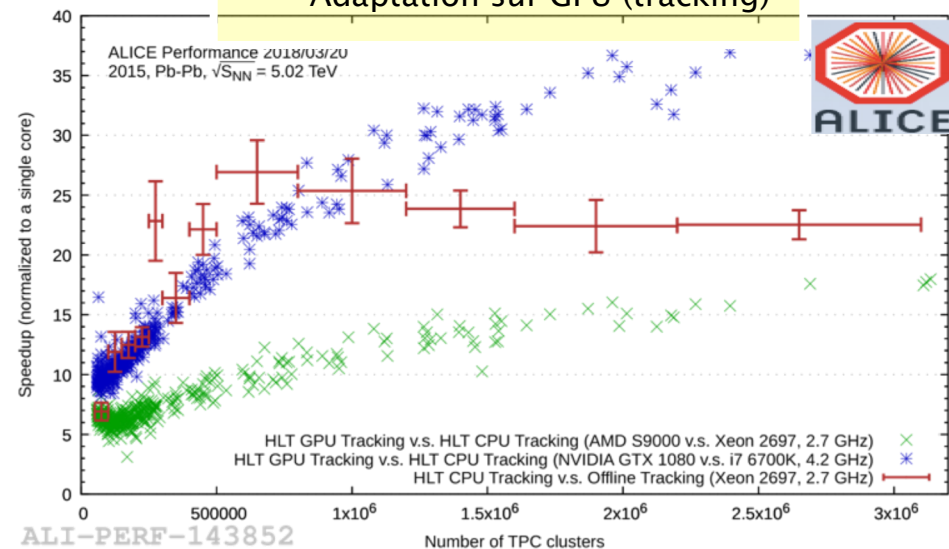
CPU usage 2018



HPC: De nouvelles utilisations du ML/AI amplifient son utilisation.

HPC: ressource importante pour non-HEP et HEP. Adaptation parfois couteuse en termes humains pour son exploitation.

Adaptation sur GPU (tracking)



Dans tous les cas, nécessité d'exploiter tous les types de ressources disponibles pour faire face → Uniformisation des accès.

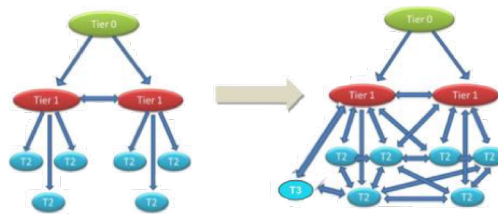
C'est la ressource la plus incompressible!

Les données –uniques– doivent être conservées.

Plusieurs types de données: Données exclusives, publiques, Open Data, croisement/échanges de données ...

Les services de stockage requièrent une expertise et une taille critique pour garantir leur fiabilité.

Rôle dévolu aux Tier0s, Tier1s.



Le réseau a montré sa grande capacité et disponibilité → Duplication et multiplication de la réplication des données

Mais le modèle n'est plus soutenable avec l'explosion en volume des données primaires.

La tendance: data lake pour stocker les données et y accéder.

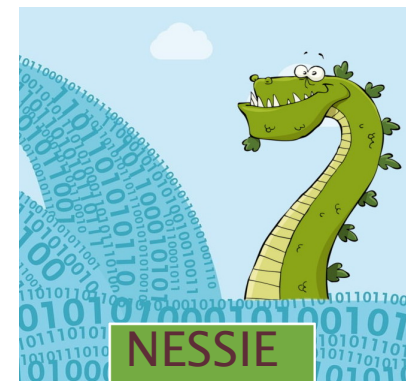
Wikipedia: Un **lac de données** (en anglais **data lake**) est une méthode de stockage des données utilisée par le big data¹ (mégadonnées en français). Ces données sont gardées dans leurs formats originaux ou sont très peu transformées.

Comment y faire face?

Faire évoluer le modèle de calcul

Par le développement d'outils et de plateformes communs: Projets européens et internationaux

Par le partage des plateformes (et des ressources) entre les communautés scientifiques.



Spécialisation des types d'infrastructures et de sites:
Données et calculs n'ont plus (et ne peuvent plus) être co-hébergés sauf cas de très grands centres.

Sites spécialisés dans le calcul: HPC, HTC, commerciaux.
Mais il faut intégrer leur accès dans le paysage en développement.

Regroupements des compétences régionales et thématiques.

Haute qualité des réseaux et de l'interconnectivité: Indispensable!

Interopérabilité, rationalisation, optimisation des datacenters

Partage, interdisciplinarité, mutualisation

Banalisation de l'accès aux ressources: pas d'adaptation spécifique à chaque site.

Prise en compte de l'importance du développement de la pile logicielle

Coupler étroitement le développement au niveau français à celui au niveau européen.

Peu de projets purement franco-français

Nécessité d'optimiser l'utilisation des ressources

Tendre vers une utilisation normalisée multi-sciences

Participer activement à la mise en place du paysage européen,
France ↔ Europe ↔ Monde

Les 5 années qui viennent sont cruciales pour préparer l'ère d'après 2025. Des pistes d'adaptation et d'évolution:

- Une réorganisation et restructuration des services que chaque site peut proposer.
- Prise en compte de l'exigence d'interopérabilité et de standardisation des accès aux ressources
- Participation aux définitions des standards et aux évolutions à l'échelle européenne.

- ▶ HOW 2019: <https://indico.cern.ch/event/759388>
- ▶ European Strategy for particle physics: <https://indico.cern.ch/event/808335>
- ▶ JCAD 2019: <https://jcad2019.sciencesconf.org>