



IN2P3

Institut national de physique nucléaire
et de physique des particules

Institut National de Physique Nucléaire et de Physique des Particules



Utilisation du CC-IN2P3: Point vue de l'IN2P3

Volker.Beckmann@IN2P3.FR



IN2P3

Institut national de physique nucléaire
et de physique des particules

DAS calcul & données

Objectifs principaux

- Infrastructure calcul pour nos projets scientifique
- Recherche dans la domaine informatique appliqué
- Coordination recherche dans IN2P3 et avec partenaires (nationaux, internationaux)
- Formation, master, doctorat, HDR
- Coordination utilisation/participation FGI, EGI, CC-IN2P3
- Concentration optimale (meso-)centres de calcul





IN2P3

Institut national de physique nucléaire
et de physique des particules

Hors périmètre DAS calcul

Service informatique comme un service technique labo

→ DAT* Christian Olivetto



Soutiens infrastructure calcul labo → SBNA*

* DAT: Directeur Adjoint Technique

* SBNA: Soutien de base non affecté



IN2P3
Institut national de physique nucléaire
et de physique des particules

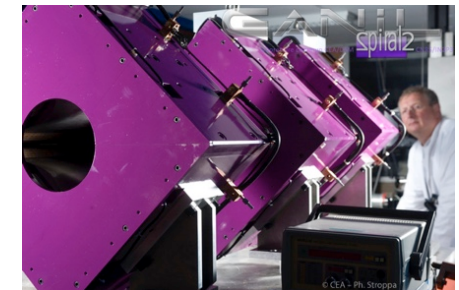
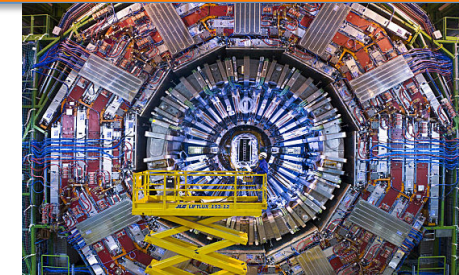
Calcul & Données

Les défis pour le calcul scientifique

- LHC, CTA, LSST, Euclid, ...
 - Big Data (volume, variété, vitesse et véracité*)
 - Infrastructure (CC-IN2P3, meso centre, Grille, HPC)
 - Logiciels (Geant4, DIRAC, ...)
-
- Il n'est pas suffisant d'agrandir les solutions existantes
 - De nouvelles approches sont nécessaire sur tous les niveaux

1. CAIF: Computing Architectures and Infrastructures for Nuclear, Particle and Astroparticle Physics
2. DCAP: Data and Computational Science, Algorithms and Programming for Nuclear, Particle and Astroparticle Physics

* La véracité du Big Data fait référence à la validité des données, l'existence des biais, la problématique du bruit et la présence de toute anomalie dans les données





IN2P3

Institut national de physique nucléaire
et de physique des particules

Calcul & Données Projets existants

LHC Computing Grid (LCG); Laurent Dufлот
Tier-1 (CC-IN2P3), Tier-2s (CPPM, IPHC, LAPP, LPC, LPSC, Subatech, GRIF*)



FGI (France Grid Infrastructure; France Grilles); Vincent Breton
participation EGI



MUST (plateforme calcul LAPP); Nadine Neyroud

FACe (plateforme calcul APC); Cécile Cavet



- * GRIF: Grille de production pour la recherche en Ile de France (APC, CEA/Irfu, LAL, LLR, IPNO, LPNHE)
- * EGI: European Grid Infrastructure



IN2P3
Institut national de physique nucléaire
et de physique des particules

Calcul & Données Projets nouveaux

DIRAC* (CPPM, LUPM, [CC-IN2P3, Univ. Bordeaux, CREATIS])



Andrei Tsaregorodtsev

- Interware: lancer processus sur la Grille
- Evolution BigData: gestion workflow, gestion données en masse, gestion des ressources (calcul / stockage)

Geant4 (LLR, LAPP, LAL, IPNO, CENBG); Marc Verderi

Geant 4

- Simulation du transport des particules dans la matière
- Méthodes de réduction de variance
- Prospective quant à intérêt à introduire micro-parallélisme
- Rapprochement avec la communauté de la physique nucléaire

* DIRAC: Distributed Infrastructure with Remote Agent Control; diracgrid.org

* LRI: Laboratoire de Recherche en Informatique, Université Paris Sud



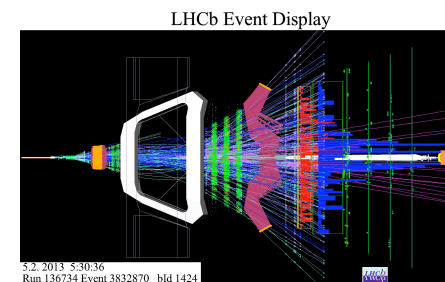
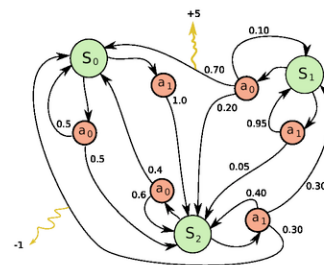
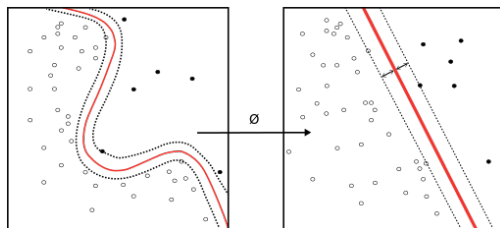
IN2P3

Institut national de physique nucléaire
et de physique des particules

Calcul & Données Projets nouveaux

CompStat: Computational Statistics and Machine Learning (LAL, LRI);
Balazs Kegbl

- Reconnaissance de formes et *machine learning* pour calorimètres imageurs
- *Deep learning*
- Apprendre pour découvrir: entraîner modèles de ML à la découverte, en tenant compte des incertitudes systématiques
- *Lean classifiers* pour détecteurs et triggers
- Remplacer la cascade par le processus de décision markovien
- Adapter au *trigger design*, LHCb





IN2P3

Institut national de physique nucléaire
et de physique des particules

Calcul & Données Projets nouveaux

BETA* (CC-IN2P3, LAPP); Frédéric Suter

- Comment transporter BigData?
- Mettre en place un prototype pour un système de transport de données de masse fortement automatisé et performant
- HTTP2; *user-centred design*

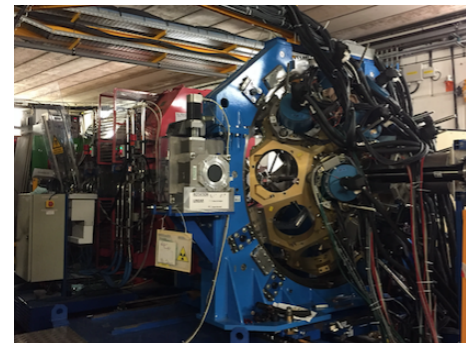


Big Data Workflow 4 Nuclear Physics (CSNSM, GANIL, CC-IN2P3, IPNL, IPNO, LAL); Eric Legay

- Gestion des données et analyse pour la physique nucléaire (AGATA*, S3, Ganil)
- Plateforme d'analyse
- Gestion des données

* BETA: Bulk Efficient Data Transport for Astroparticle Physics

* AGATA: Advanced GAMMA Tracking Array





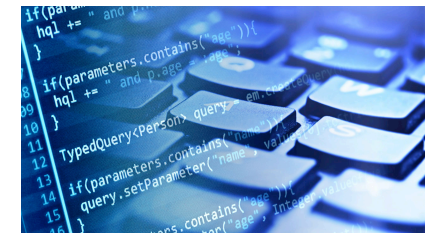
IN2P3

Institut national de physique nucléaire
et de physique des particules

Calcul & Données Projets nouveaux

PRESOFT: Pérennisation du logiciel (IdGC, LPC, CC-IN2P3, LIGM);
Geneviève Romier

- Plan de gestion des données → plan de gestion du logiciel
- Développer des procédures pour les plans de gestion du logiciel
- Étudier faisabilité, bénéfices, acceptabilité et conditions
- Évaluer impact





IN2P3

Institut national de physique nucléaire
et de physique des particules

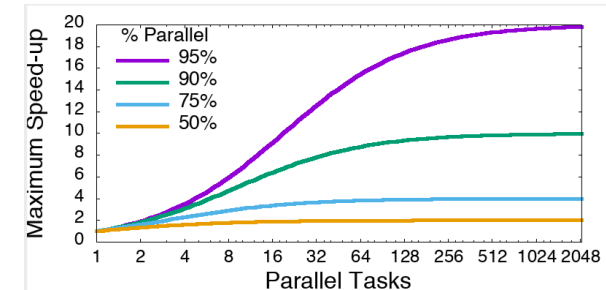
Calcul & Données

Groups de travail nouveaux

HTC / HPC

Groupe Travail HTC/HPC (LAL, LAPP, IPNO, IPHC, APC, LLR, CC-IN2P3, CPPM, ...)

- *Embarrassingly parallel computing*
- Transfer de systèmes HTC* à systèmes HPC (GPU)
- Evoluer l'application interaction laser/plasma d'électrons
- Possibilités, impact, vision
- Groupe travail → focus → possibilité projets 2018
- 22 février 22, CNRS/IN2P3 Paris
- <https://indico.in2p3.fr/event/14008/>



* HTC: High-Throughput Computing; HPC: High-Performance Computing



IN2P3

Institut national de physique nucléaire
et de physique des particules

Calcul & Données

Groups de travail nouveaux

Machine Learning

Groupe Travail Machine Learning (LPC Clermont, LAL ,LLR, CPPM, IPHC, LAPP, IMNC, LUPM)

- Deep learning
- Potentiel de découverte du Machine Learning
- Lean classifiers
- Synergie astroparticule (LSST / Euclid), physique particules,
- ...

Première réunion: mars? Paris ou Lyon ou ...?



IN2P3

Institut national de physique nucléaire
et de physique des particules

IN2P3 et utilisateurs calcul

Quel est le meilleur modèle informatique pour mon projet?
Laptop, cluster locale, CC-IN2P3, grille, cloud ... ?

Data Management Plan (DMP):

- Info administratives
- Description données
- Métadonnées et documentation
- Responsabilité
- Ethique, cadre légal (droits de propriété)
- Stockage sécurisé et sauvegarde
- Partage des données
- Préservation a long terme (archivage)
- Budget

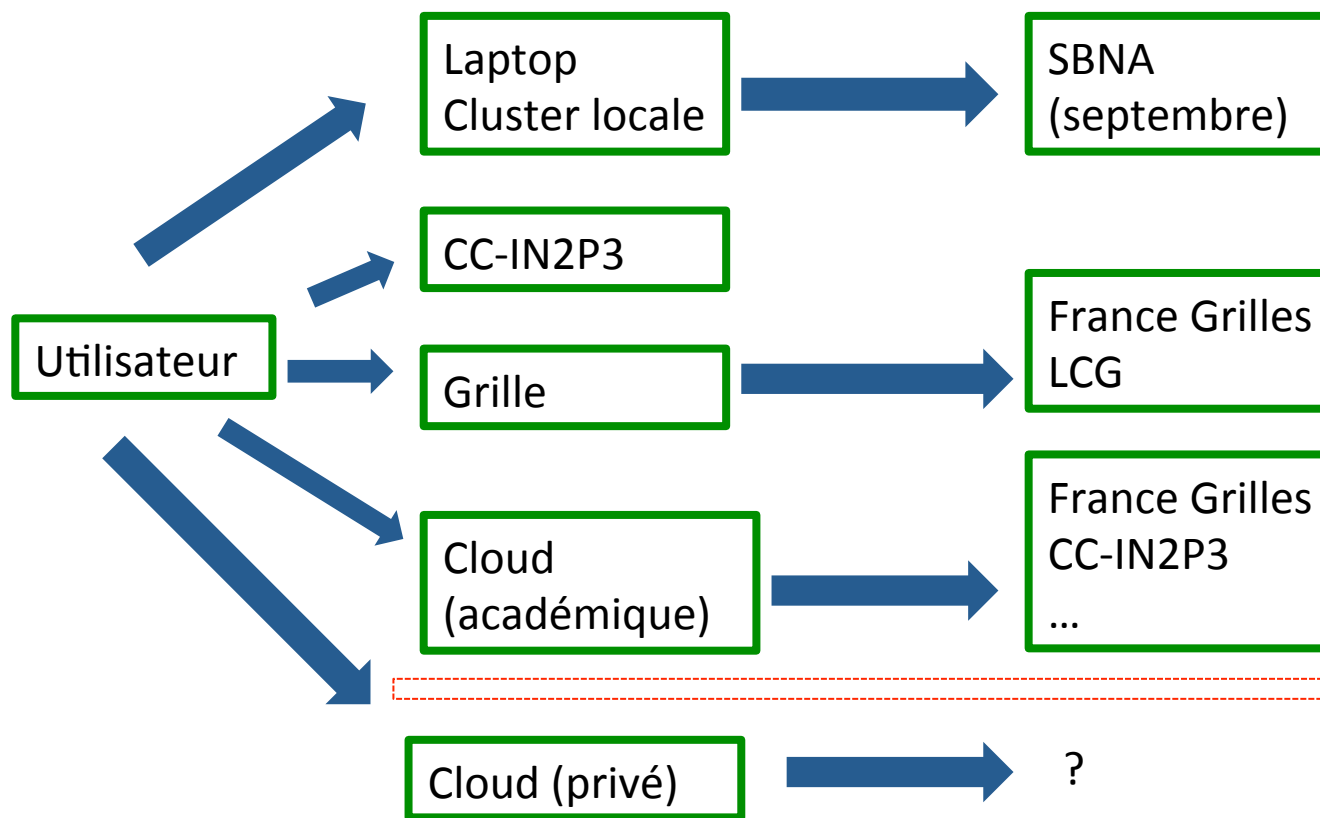




IN2P3

Institut national de physique nucléaire
et de physique des particules

Financement calcul Projets IN2P3



Comment l'IN2P3 vous soutient-il dans vos besoins calcul?

- Financement RH dans les labos
- Infra labo: soutien direct à travers SBNA
- Soutien à travers projet
- LCG: soutien Tier-1 (CC-IN2P3) et Tier-2, 2 Meuros, ~10 ETP
- France Grilles: 200 kEuros

- Coordination demandes
- Proposition H2020: information, coordination, avis
- Infrastructure calcul hors IN2P3



IN2P3

Institut national de physique nucléaire
et de physique des particules

Concentration des budgets

- Budget IN2P3 constant
- Demande croissante de ressources calcul: LHC, LSST, Euclid, Spiral-2, CTA, ...
- Virtualisation des infrastructures
- Centralisation des services (mail, Atrium, Redmine, ...)
- Infrastructure centralisée, expertise locale
- Coordination des infrastructures au niveau régional
- Notre infrastructure informatique dans le contexte européen

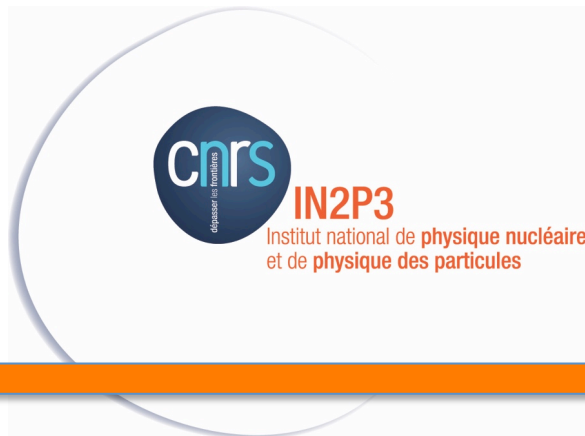


IN2P3

Institut national de physique nucléaire
et de physique des particules

Summary

- Projets existants : FGI, LCG, FACe, MUST
- Projets nouveaux: Geant4, DIRAC, PRESOFT, CompStat, Workflows 4NP, BETA
- Data Management Plan: Un outil pour structurer vos besoins calcul
- Infrastructure et services centralisée, expertise locale
- CC-IN2P3 comme navire phare de notre infrastructure



European Open Science Cloud (EOSC)

- Integration and consolidation of e-infrastructures
- Federation of existing research infrastructures and scientific clouds
- Development of cloud-based services for Open Science
- Connection of [ESFRI](#)s* to the EOSC

[ESFRI roadmap](#): “The research process is becoming more and more dependent on [...] large amounts of data, often acquired in a very short time, and on the availability of effective **on-line analysis** and **high throughput, high performance** computing and the strongly emerging shift towards cloud computing.”



* ESFRI: European Strategy Forum on Research Infrastructures. E.g. [CTA](#), [KM3NeT](#), [Euro-Biolmaging](#), [HL-LHC](#), FAIR, SPIRAL2, SKA, ...