



**A LA UNE**

## Le Higgs passe au vert avec le calcul haute performance



Une méthode d'analyse pour CMS

L'application CMS-MEM est un outil d'analyse basé sur la Méthode des Éléments de Matrice (MEM) [1]. Elle permet de combiner de manière optimale l'information théorique décrivant des processus physiques avec l'information expérimentale décrivant la résolution d'un détecteur. Elle a été développée dans sa version « Htautau » au LLR, et elle est utilisée dans l'expérience CMS pour caractériser les propriétés du boson de Higgs, produit en association avec des quarks top, et se désintégrant en paires de leptons tau.

A l'opposé des méthodes supervisées (réseau de neurones, arbre de décision, machines à vecteur de support), la MEM permet de définir des discriminants pour la classification d'événements, sans apprentissage, en s'appuyant sur la physique pour calculer les probabilités d'observer les états finaux attendus. Cette nouvelle méthode plus précise est, néanmoins, extrêmement consommatrice de temps CPU.

Dans ce contexte, l'équipe CMS du LLR, par le biais du travail de recherche de T. Streblor sur le canal ttH, a relevé le défi de développer un application dite "HPC" [2]. Dans une première étape, elle a été exploitée sur des plates-formes multi-nœuds / multi-cœurs, puis dans un second temps sur des architectures multi-nœuds / multi-GPUs [3]. L'implémentation multi-nœuds / multi-cœurs a été utilisée avec (...)

**Cycle de séminaires sur la science des données au CERN**  
Retrouvez ici une nouvelle série intéressante de séminaires sur la science des données au CERN. (...)

[en savoir plus](#)

**De l'intérêt des forums ORAP**  
Le forum ORAP a lieu tous les six mois en France et il est à mes yeux, presque l'équivalent (...)

[en savoir plus](#)

**558 inscrits aux JDEV 2017 à Marseille**

L'édition 2017 des Journées du Développement Logiciel de l'Enseignement Supérieur (ESR) et de la (...)

[en savoir plus](#)

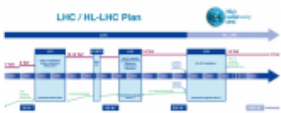
[lire la suite](#)

**Appel à projets CC-IN2P3**

Le CC-IN2P3 fournit des ressources d'analyse et de traitement de données aux laboratoires de (...)

[en savoir plus](#)

**LHC**



### Scientific Computing Forum : La planification du futur de notre infrastructure

Après deux premières sessions en février et mai dernier, le prochain Scientific Computing Forum, qui réunit des experts dans le domaine des applications de calcul intensif du secteur public, se réunira le 27 octobre prochain au CERN. L'une des principales motivations du CERN dans l'organisation du Scientific Computing Forum est le défi lancé par le futur accélérateur LHC haute luminosité (HL-LHC). En effet, le traitement des données des expériences organisées au CERN a toujours poussé les limites des infrastructures informatiques associées. Cela a conduit par exemple au développement de la grille informatique, qui sert la communauté depuis 15 ans maintenant [1]. Grâce à cette infrastructure internationale de grille de calcul (WLCG), nous pourrions traiter les 50 Po de données attendues en 2017. Mais les expériences au LHC doivent évoluer si la physique des particules au CERN veut continuer à produire des résultats scientifiques de première classe. Dans un an, le LHC cessera de prendre des données (...)

[lire la suite](#)

**Logiciel**



### A la découverte de SPARK

Un nouvel outil s'installe dans le paysage de la gestion des ressources du calcul distribué : Spark. Plusieurs actions ont démarré depuis un an, à l'IN2P3 et à l'université Paris-Sud (U-PSud), ayant comme objectif de comprendre et évaluer ce nouvel écosystème de programmation et de calcul distribué. En effet, Spark semble proposer des solutions tout à fait complémentaires aux outils de calcul intensif et haut-débit (HPC, HTC) déjà implantés dans nos environnements. Spark a démarré comme projet de recherche à l'université UC Berkeley en 2009, puis a été pris en charge par la fondation Apache. Il est maintenant une plateforme logicielle Open Source de traitement des données. Développé en Scala, il s'interface naturellement avec les langages Java, Python et R (et Scala). Spark se propose d'optimiser et d'élargir le concept déjà bien connu de « Map-Reduce » (déployé dans le framework Hadoop), tout en optimisant de façon cohérente l'utilisation de la mémoire et les communications entre processus, grâce à une analyse du graphe (DAG) des

**Colloque**



### Journées Sciences de Données du GdR MaDICS

Les journées MaDICS ont eu lieu les 22 et 23 juin à l'École de Management de Marseille (EDM), co-organisées localement par des laboratoires CNRS / Aix-Marseille Université (CPPM, IMBE, LAM). MaDICS (Masses de Données, Informations et Connaissances en Sciences) est un groupement de recherche (GdR) qui permet d'encourager les activités d'animation interdisciplinaires sur les masses de données scientifiques. Les journées se sont déroulées en deux temps : d'abord une session plénière le 22 juin a mis en avant des thématiques phares, parmi lesquelles des avancées obtenues au sein des actions du GdR. La séance a inclus des présentations sur de sujets de recherche innovants : nouveau concept et plateforme RAMP Rapid Analytics and Model Prototyping (Balasz Kegl, Université Paris-Saclay), Recherches en masses de Données Bioacoustiques (Herve Glotin, Univ. Toulon), Analyse statistique (machine learning) en partenariat industriel (Charlotte Laclau, Laboratoire LIG Grenoble) et Flux de travail scientifiques et outils complémentaires pour la

**Simulation**



### PSPA : Design your Accelerator !

PSPA est un projet né il y a quelques années sur une idée d'Alessandro Variola, alors responsable du département Accélérateurs, au Laboratoire de l'Accélérateur Linéaire d'Orsay. Il vise à fournir aux concepteurs d'accélérateurs de particules un outil universel et convivial permettant d'exécuter différents codes de simulation d'accélérateurs existants de la discipline en les combinant facilement entre eux. Aujourd'hui, pour concevoir ou étudier un accélérateur de particules, un « physicien machine » dispose d'un nombre important de codes de modélisation et de simulation, essentiellement open source, produits par la communauté des accélérateurs. L'usage de la plupart d'entre eux est libre, mais leur facilité de mise en œuvre (disponibilité, systèmes d'exploitation supportés, documentation, convivialité d'utilisation...) est très variable, avec un grand nombre de paramètres de configuration. Ce physicien machine doit dès lors faire face à une multitude de problèmes : Il doit récupérer le code qui l'intéresse et l'installer sur les ressources informatiques dont il dispose Une étude de (...)

[lire la suite](#)

**Agenda**

**Conférence XLDB-2017, du 10 au 12 octobre 2017 - Clermont-Ferrand**

Pour la première fois de son histoire, la conférence XLDB (Extremely Large Databases) se tiendra (...)

[en savoir plus](#)

**Journées SUCCES 2017, 16 et 17 octobre - Grenoble**

Dédiées aux utilisateurs des mésocentres et des infrastructures de grilles ou de cloud, les (...)

[en savoir plus](#)

**10èmes journées méso-centres, 26 et 27 septembre 2017 - Paris**

Les 10èmes journées méso-centres auront lieu les 26 et 27 septembre à l'Institut Henri Poincaré à (...)

[en savoir plus](#)

**HEPIX Fall/Autumn 2017, du 16 au 20 octobre - KEK, Japon**

La conférence d'automne HEPiX est organisée du 16 au 20 octobre à KEK, au Japon. Programme, appel (...)

[en savoir plus](#)

**Archives**

**Abonnement**

Pour vous abonner/désabonner, suivez ce [lien](#).

**Proposer un article**

Vous souhaitez proposer un article ? Envoyez un mail à [LettreInformatique@in2p3.fr](mailto:LettreInformatique@in2p3.fr).

tâches et des transferts de reproductibilité en bio-  
jeux de données, en informatique (...)  
s'appuyant largement sur la (...)

[lire la suite](#)

[lire la suite](#)



**Equipe**

**Directeur de la publication** : Alain FUCHS.

**Responsable éditorial** : Volker BECKMANN, Pierre-Etienne MACCHI.

**Comité de rédaction** : David CHAMONT, Frédérique CHOLLET LE FLOUR, Virginie DELEBARRE DUTRUEL, Dirk HOFFMANN et Gaëlle SHIFRIN.

---

**A LA UNE**

## Le Higgs passe au vert avec le calcul haute performance

Infrastructure



Détails au niveau de l'événement des échanges entre l'hôte et un GPU. Pour chaque événement, plusieurs scénarii sous forme d'intégrales sont évalués. Chaque intégrale est calculée plusieurs fois (5 fois) pour effectuer un test de khi-deux.

### Une méthode d'analyse pour CMS

L'application CMS-MEM est un outil d'analyse basé sur la Méthode des Éléments de Matrice (MEM) [1]. Elle permet de combiner de manière optimale l'information théorique décrivant des processus physiques avec l'information expérimentale décrivant la résolution d'un détecteur. Elle a été développée dans sa version « Htautau » au LLR, et elle est utilisée dans l'expérience CMS pour caractériser les propriétés du boson de Higgs, produit en association avec des quarks top, et se désintégrant en paires de leptons tau.

A l'opposé des méthodes supervisées (réseau de neurones, arbre de décision, machines à vecteur de support), la MEM permet de définir des discriminants pour la classification d'événements, sans apprentissage, en s'appuyant sur la physique pour calculer les probabilités d'observer les états finaux attendus. Cette nouvelle méthode plus précise est, néanmoins, extrêmement consommatrice de temps CPU.

Dans ce contexte, l'équipe CMS du LLR, par le biais du travail de recherche de T. Strebler sur le canal ttH, a relevé le défi de développer un application dite "HPC" [2]. Dans une première étape, elle a été exploitée sur des plates-formes multi-nœuds / multi-cœurs, puis dans un second temps sur des architectures multi-nœuds / multi-GPUs [3]. L'implémentation multi-nœuds / multi-cœurs a été utilisée avec succès dans les comptes-rendus publics de la collaboration CMS [4].

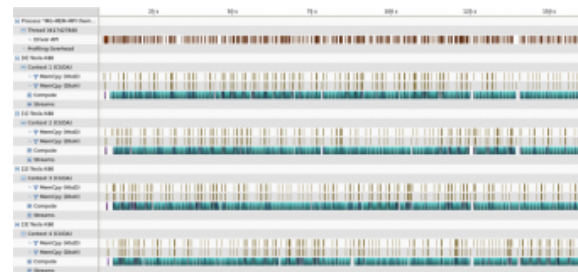
### Contraintes et défis technologiques

Nous avons souhaité dès le début des développements, initialement réalisés sur la plate-forme GridCL [5], relever un certain nombre de défis techniques. Pour traiter de l'analyse du canal ttH par MEM dans des délais raisonnables, nous avons conçu l'application autour des standards portables MPI et OpenCL pour agréger, respectivement, la puissance de calcul des nœuds et celle des GPUs ou autres accélérateurs de calcul rattachés à chaque nœud. L'autre choix de conception consiste à déployer tous les calculs sur les cartes accélératrices ou GPUs pour minimiser les communications « host/devices ». Cette stratégie nous a amené à développer une extension OpenCL/CUDA dans le générateur de code MadGraph [6], qui engendre l'élément de matrice au cœur de la méthode sous la forme d'un code source. Ceci nous a conduit à manipuler des "kernels" [7] allant de  $10^4$  à  $2.10^4$  lignes de code, caractéristique peu courante pour ces technologies, permettant ainsi d'éprouver la robustesse des compilateurs pour accélérateurs de calcul.

Plus récemment encore, notre participation à la Cellule de Veille Technologique (CVT) GENCI [8], nous a mis face à une difficulté supplémentaire. Leur plate-forme "pré-exascale" [9] combine des processeurs IBM OpenPOWER à instructions non x86, et des cartes de génération Pascal, pour lesquelles NVIDIA ne fournit pas encore d'implémentation d'OpenCL. Soucieux de profiter des évolutions de CUDA et des outils NVIDIA (non accessibles depuis OpenCL), nous avons construit une passerelle OpenCL 1.1/CUDA, pérennisant ainsi nos développements OpenCL face aux aléas. Cette passerelle a été largement validée et exploitée sur GridCL, sur les GPUs Pascal P100 de l'IDRIS et sur la plate-forme GPU du CC-IN2P3.

CC-IN2P3 a été une aubaine pour effectuer les analyses requises. Cette machine dispose d'une puissance de calcul importante, répartie sur 10 nœuds équipés chacun de 4 accélérateurs NVIDIA K80. Le portage de CMS-MEM sur cette plate-forme a été effectué sans difficulté et le mode de soumission a été adapté pour répondre aux contraintes du système de queues.

Comme illustré dans la figure ci-dessous, le taux d'occupation des GPUs est excellent : les *kernels* calculent la majorité du temps d'exécution et les communications *host/devices* sont négligeables.



chronologie des communications (traits noirs) et kernels actifs (en bleu) sur un nœud (4 GPUs)

Pour évaluer le gain global en performance de l'application, nous avons pris comme référence l'analyse de 3000 événements. Le temps CPU pour traiter de cette configuration, s'évalue approximativement à 15 heures sur une plate-forme type MPI du CC-IN2P3 comportant 100 cœurs (100 processus MPI). Après quelques difficultés de mise au point des classes (le point délicat étant de coupler des classes « parallèle » avec la réservation des ressources GPUs), nous avons pu lancer nos premières exécutions de production sur 2, 4, puis 6 nœuds. Pour cette dernière configuration, 6 nœuds totalisant 24 GPUs, l'application a requis 1767 secondes. Le facteur d'accélération substantiel obtenu, approximativement 30, n'est pas seulement imputable aux GPUs. La réécriture du calcul des éléments de matrice C++, en kernels C/OpenCL entre dans une part importante de ce gain. Ainsi la puissance de calcul de la configuration avec 6 nœuds est équivalente à une plate-forme MPI comportant 3000 cœurs pour notre application CMS-MEM. Outre le gain en puissance de calcul, l'exploitation des accélérateurs de calcul ou GPU permet de réduire significativement la facture de la consommation électrique des centres de calcul, dans notre cas, cette consommation est réduite d'un facteur 5 [10].

Cette version, première du genre, est en cours d'exploitation sur les nœuds GPUs du CC (prochainement 8 nœuds), et alimente les résultats d'analyse du canal ttH pour la thèse de T.Strebler, libérant ainsi la grille d'une charge significative de travail.

Dans les années à venir, son évolution profitera des axes de réflexion du groupe de travail RI3 « CodeursIntensifs » [11] et des projets IN2P3 en cours de constitution autour des thèmes des conteneurs, de la génération de code, de la précision et de la reproductibilité numérique dans un contexte de calcul parallèle.

### Remerciements

Nous adressons nos sincères remerciements au LabEx P2IO pour avoir financé (avec une contribution de l'équipe CMS et du LLR) la plate-forme GridCL, brique indispensable aux développements de CMS-MEM. Nous remercions le Centre de Calcul de l'IN2P3 pour l'élaboration du contexte de production de CMS-MEM, ainsi que la CVT GENCI pour les échanges lors des groupes de travail autour d'experts HPC.

Parmi les possibilités de déploiement de CMS-MEM en mode production, l'acquisition d'une plate-forme GPUs par le

[1] A titre d'exemple <https://arxiv.org/abs/hep-ex/0406031>

[2] High Performance Computing

[3] Graphical Processing Unit

[4] <https://cds.cern.ch/record/2257067>

[5] Plate-forme de développement, financé principalement par le LabEx P2IO

[6] <https://launchpad.net/mg5amcnlo>

[7] Terme utilisé pour désigner les codes téléchargés sur l'accélérateur de calcul, une fois compilés

[8] Cellule de Veille Technologique (CVT) GENCI est en charge de préparer les communautés scientifiques françaises à l'émergence des prochaines génération de supercalculateurs

[9] Les architectures dites exascales désignent les futurs supercalculateurs qui disposeront d'une puissance de calcul supérieure à  $10^{18}$  opérations en arithmétique flottante par seconde (ExaFlops)

[10] Comparaison effectuée en prenant uniquement en compte la consommation électrique des processeurs (E5-2698 v3 16 cœurs) de 135 W et la consommation par GPU de 150 W

[11] <https://gitlab.in2p3.fr/CodeursInte...>, accessible sur demande.



## Scientific Computing Forum : La planification du futur de notre infrastructure



Le calendrier pour le High-Luminosity LHC. © CERN

Après deux premières sessions en février et mai dernier, le prochain Scientific Computing Forum, qui réunit des experts dans le domaine des applications de calcul intensif du secteur public, se réunira le 27 octobre prochain au CERN.

L'une des principales motivations du CERN dans l'organisation du Scientific Computing Forum est le défi lancé par le futur accélérateur LHC haute luminosité (HL-LHC).

En effet, le traitement des données des expériences organisées au CERN a toujours poussé les limites des infrastructures informatiques associées. Cela a conduit par exemple au développement de la grille informatique, qui sert la communauté depuis 15 ans maintenant [1].

Grâce à cette infrastructure internationale de grille de calcul (WLCG), nous pourrions traiter les 50 Po de données attendues en 2017. Mais les expériences au LHC doivent évoluer si la physique des particules au CERN veut continuer à produire des résultats scientifiques de première classe. Dans un an, le LHC cessera de prendre des données afin d'effectuer une série de mises à niveau sur une période de 18 mois. Les principales mises à jour seront effectuées sur ALICE et LHCb, bien que tous les détecteurs et l'accélérateur soient concernés par des changements. Puis, en 2020, le « run 3 » débutera. Bien que cela représente déjà une augmentation du volume de données et des besoins de calcul, la communauté est assez confiante quant à sa capacité à traiter ces données dans les délais. La prise de données de 3 ans du « run 3 » sera suivie d'un arrêt plus long, d'un ordre de grandeur différent des précédents. ATLAS et CMS subiront de vastes changements et le LHC lui-même produira un faisceau beaucoup plus fort, c'est-à-dire qu'il augmentera sa luminosité d'un facteur 20 par rapport à aujourd'hui. Lors de la mise en production du LHC haute luminosité (HL-LHC), cette combinaison de détecteurs de résolution spatiale et temporelle plus fine avec un faisceau plus fort produira plus d'événements et entraînera une importante augmentation du débit de données. Des estimations du volume attendu sont difficiles, car les détails de toutes les mises à jour doivent encore être définis. Mais une augmentation d'un facteur d'environ 100 par rapport à aujourd'hui est d'ores et déjà considérée, tant pour le volume de données à traiter que la puissance de calcul nécessaire.

### Le statut du Scientific Computing Forum

Le Scientific Computing Forum réunit des experts dans le domaine des applications de calcul intensif du secteur public pour définir les solutions répondant aux besoins de la prochaine décennie. En développant le succès de la grille mondiale LHC (WLCG) et englobant les derniers développements des *Scientific Clouds*, la science continuera à conduire et à diriger le développement d'une installation de calcul de pointe. Le forum prend en compte les développements de l'informatique et du calcul dans le but d'identifier les options et les synergies pour des solutions adéquates et abordables.

Puisque la tâche est plutôt difficile, la discussion est ouverte à de nouvelles approches. Par exemple, certaines propositions indiquent un modèle dans lequel les données seront stockées

invitation, les présentations et les comptes rendus des réunions sont accessibles au public.

### En quoi cela vous intéresse ?

Si la façon dont WLCG organise le calcul change, il aura un impact certain sur l'activité non seulement des 30 collègues environ qui travaillent directement pour LCG-France mais également pour tous les membres du réseau. À ce jour, notre infrastructure de calcul est dominée par la nécessité de servir LCG-France. Si nous devions changer le modèle de calcul WLCG, cela pourrait avoir un impact significatif sur la mise en place de l'infrastructure informatique et des ressources humaines.

Un autre aspect important est que nous ne devons pas seulement considérer les expériences LHC dans le contexte de notre infrastructure de calcul. Au cours des prochaines années, d'autres expériences avec calcul et données intensives deviendront opérationnelles, telles que LSST, Euclid, CTA et KM3NeT. En outre, la physique nucléaire, surtout en vue de Spiral-2, se prépare à centraliser plus fortement sa gestion des données et éventuellement ses plates-formes d'analyse. Pour relever ce défi, un groupe de travail au niveau de l'IN2P3 discute des perspectives d'une infrastructure informatique commune pour la physique nucléaire. De plus, en janvier 2017, un projet pilote dans ce contexte a été lancé afin de fournir une plate-forme d'analyse pour AGATA, une autre expérience de physique nucléaire.

Ainsi, nous devons partager l'infrastructure de calcul afin de servir tous les projets scientifiques soutenus par l'IN2P3.

Le Scientific Computing Forum du CERN jouera un rôle majeur dans la définition de ce qui sera calculé et stocké, où et comment. Suivre cette discussion est vital pour nous afin de préparer notre propre infrastructure. Participer ici offre la possibilité d'influencer l'avenir de l'informatique en France. Vous êtes invités à donner votre opinion - soit directement à moi, soit dans le cadre du travail pour la perspective de l'IN2P3, qui commencera cette année et qui fournira dans l'année à venir une feuille de route pour la recherche en informatique et l'évolution de notre infrastructure de calcul.

Forum février 2017 : <https://indico.cern.ch/event/581096/>  
Forum mai 2017 : <https://indico.cern.ch/event/619438/>

**Volker BECKMANN (Direction IN2P3)**



dans un petit nombre (5-10) de grands centres du monde entier. De plus, l'informatique devrait alors être concentré dans les centres Tier-1 (par exemple le CC-IN2P3) et dans quelques-uns des grands centres Tier-2. Ce **point de vue a par exemple été présenté par Ian Bird** (CERN).

L'IN2P3 est représenté dans le forum et participe activement aux discussions. A titre d'exemple, voir la **présentation sur IN2P3 Computing Outlook** lors du dernier Computing Forum. Cette participation est signe que l'Institut souhaite maintenir sa structure avec une forte expertise dans les centres Tier-1 et Tier-2.

Bien que la participation à la réunion soit faite uniquement par



Spiral-2 au GANIL. © V. Beckmann

[1] Par exemple Foster & Kesselman 2003, "*The Grid 2*", Elsevier



## A la découverte de SPARK



Un nouvel outil s'installe dans le paysage de la gestion des ressources du calcul distribué : **Spark**. Plusieurs actions ont démarré depuis un an, à l'IN2P3 et à l'université Paris-Sud (U-PSud), ayant comme objectif de comprendre et évaluer ce nouvel éco-système de programmation et de calcul distribué. En effet, **Spark** semble proposer des solutions tout à fait complémentaires aux outils de calcul intensif et haut-débit (HPC, HTC) déjà implantés dans nos environnements.

**Spark** a démarré comme projet de recherche à l'université UC Berkeley en 2009, puis a été pris en charge par la fondation Apache. Il est maintenant une plateforme logicielle Open Source de traitement des données. Développé en Scala, il s'interface naturellement avec les langages Java, Python et R (et Scala). **Spark** se propose d'optimiser et d'élargir le concept déjà bien connu de « Map-Reduce » (*déployé dans le framework Hadoop*), tout en optimisant de façon cohérente l'utilisation de la mémoire et les communications entre processus, grâce à une analyse du graphe (DAG) des tâches et des transferts de jeux de données, en s'appuyant largement sur la programmation fonctionnelle. L'importance de cette technologie nous est apparue en mars 2016 lors d'une journée Loops organisée par Julien Nauroy (à l'époque à l'U-PSud), où avait été présenté un travail effectué au CDS (Centre de Données astronomiques) de Strasbourg sous la conduite de André Schaaff et François-Xavier Pineau. Puis, Cécile Germain (*du LRI*) a lancé un projet ERM-MRM de l'U-PSud centré sur Spark, s'appuyant sur la plateforme mutualisée VirtualData des laboratoires de physique d'Orsay (*dont le LAL*) et réunissant plusieurs équipes de recherche (Astrophysique - LSST, Génomique - [Médecine, Bio-Informatique], Ecologie, Machine Learning). Ce projet (*ERM-MRM Spark*) a été officiellement lancé en janvier 2017 avec la mise en place d'une structure collaborative autour d'une plateforme OpenStack (*pilotée par l'équipe d'exploitation du LAL autour de Guillaume Philippon*), et moi-même (Christian Arnault) pour l'animation (*Liste de diffusion, Wiki, réunions techniques récurrentes, formation Spark*) et la liaison avec LSST.

### Architecture SPARK

Un cluster SPARK se compose d'un ensemble de machines (*workers*) homogènes ou hétérogènes dont les ressources sont pilotées par un gestionnaire de ressource (*master*).

Pour utiliser SPARK, c'est aussi simple qu'une connexion à une base de données, il vous suffit de charger dans votre code JAVA, PYTHON ou SCALA les bibliothèques SPARK pour accéder à l'ensemble des ressources de votre plateforme SPARK. Lorsque vous établissez une connexion à une plateforme SPARK, votre programme va tout d'abord discuter avec le gestionnaire de ressources pour réclamer ce dont il a besoin en mémoire et en CPU. Une fois les ressources obtenues, votre programme est compilé par le driver SPARK afin de construire un graphe de tâches (**DAG**). Ce graphe indique l'ordre et les relations entre toutes les tâches à effectuées par SPARK. C'est aussi le **SPARK Driver** qui est en charge de la coordination et l'exécution des tâches sur les **workers**.

Nous pourrions de plus, étudier la collaboration entre Spark et MongoDB à travers la production et l'utilisation de catalogues d'objets. En effet, un connecteur pour cette base de données NoSQL est disponible. Ainsi une vision ensembliste des données structurées par MongoDB est accessible, avec les opérateurs traditionnels des bases SQL (Jointures, groupes, tris, ...).

Evidemment, cette étude se structure autour de la communauté astrophysique, incluant en particulier la collaboration LSST, ou d'autres projets associés comme le projet Petasky à Clermont, et en collaboration naturelle avec les experts du CC-IN2P3.

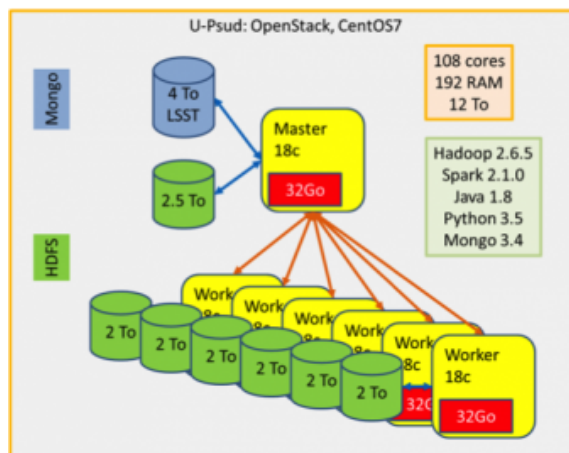
### Organisation d'une école Spark

Pour initier le projet ERM-MRM Spark, une école Spark a été organisée au LAL les 14 et 15 mars 2017, avec l'aide d'un intervenant extérieur d'une société partenaire de Databricks (*un des contributeurs d'origine du logiciel Spark*). Trente étudiants étaient présents, essentiellement les membres des équipes de recherche du projet ERM-Spark, mais aussi quelques personnes de la communauté IN2P3. Cette école s'est structurée autour d'une plateforme pédagogique basée sur les notebooks Jupyter, opérée sur le cluster OpenStack de VirtualData.

### Statut de la plateforme

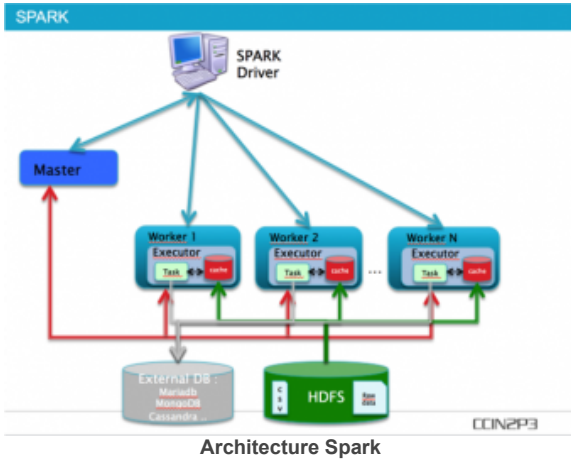
Le cluster de développement Spark est mis à disposition sous forme de machines virtuelles sur la plateforme OpenStack de VirtualData avec un certain nombre d'outils logiciels :

- API standards : Scala, Python, Java
- Built-In Components :
- SparkR (programmation R)
  - Spark SQL
  - MLib (bibliothèque pour le Machine Learning)
  - Spark GraphX (traitement des graphes)
- Mongo (→ Mongo sharded),
- Genomics/Adam (plateforme pour la génomique)
- Système de fichier distribué HDFS
- ssh, Jupyter notebook
- HT Condor (pour la gestion des ressources de plusieurs clusters Spark)
- Ganglia, pour le monitoring



Le cluster Spark sur la plateforme Virtual Data

A ce stade, nous avançons dans la compréhension du paramétrage de Spark (équilibre de la mémoire, des I/O, des cœurs, ...). La collaboration avec les experts du génome, qui démarre cet été, va nous permettre d'explorer une plateforme quasiment opérationnelle, en mesurant les performances



réalistes sur des données de taille significative (1To). D'un autre côté, l'exploration de plusieurs cas d'utilisation dans le contexte LSST va permettre de comprendre comment l'architecture d'une application de calcul et d'analyse mais aussi les structures de données, vont devoir s'adapter à la nouvelle façon de penser une application distribuée, où la programmation fonctionnelle joue un rôle central.

#### Références

- Apache Spark, <http://spark.apache.org/>
- Apache Hadoop, <http://hadoop.apache.org/>
- Spark : Cluster Computing with Working Sets, Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scop Shenker, Ion Stoica, University of California, Berkeley, <http://static.usenix.org/legacy/eve...>
- Application en astrophysique, « crossmatch » de catalogues de sources, <https://reseau-loops.github.io/pres...>
- LSST, <http://lsst.in2p3.fr/>
- MongoDB, <https://www.mongodb.com/fr>
- Plateforme VirtualData du Mésocentre U-PSud, [www.mesocentre.u-psud.fr](http://www.mesocentre.u-psud.fr)
- France génomique, <https://www.france-genomique.org/>

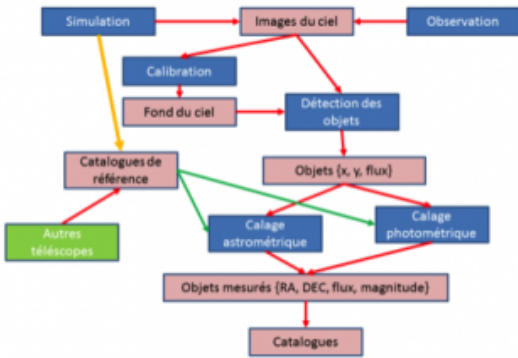
Christian ARNAULT (LAL) et Osman AIDEL (CC-IN2P3)

#### Les développements pour LSST

Dans la perspective des grandes quantités d'images en provenance du futur télescope LSST, un axe de R&D consiste à confronter les flots de traitements de données avec les logiques de distribution du calcul proposées par Spark.

Notre objectif ici sera d'évaluer les performances de Spark, en fonction de ses multiples paramètres d'optimisation. Pour y parvenir, une application émulant l'architecture proposée par la plateforme standard de traitement des images (*Stack*) a été développée, permettant de personnaliser à volonté le dimensionnement des données, les étapes de calcul, les transferts entre nœuds du graphe de calcul.

Des outils de monitoring peuvent être associés avec l'environnement Spark (ex : *Ganglia*) et offrent de puissants outils de visualisation des performances (Mémoire, IO, CPU, Network, Disque).



Un exemple de flots de traitement des données pour LSST





n°37  
Juillet  
2017

# La lettre IN2P3 Informatique

Réseau des Informaticiens de l'IN2P3 et de l'IRFU



## Journées Sciences de Données du GdR MaDICS



Les journées MaDICS ont eu lieu les 22 et 23 juin à l'École de Management de Marseille (EDM), co-organisées localement par des laboratoires CNRS / Aix-Marseille Université (CPPM, IMBE, LAM). MaDICS (Masses de Données, Informations et Connaissances en Sciences) est un groupement de recherche (GdR) qui permet d'encourager les activités d'animation interdisciplinaires sur les masses de données scientifiques.

Les journées se sont déroulées en deux temps : d'abord une session plénière le 22 juin a mis en avant des thématiques phares, parmi lesquelles des avancées obtenues au sein des actions du GdR. La séance a inclus des présentations sur des sujets de recherche innovants : nouveau concept et plateforme RAMP Rapid Analytics and Model Prototyping (Balasz Kegl, Université Paris-Saclay), Recherches en masses de Données Bioacoustiques (Herve Glotin, Univ. Toulon), Analyse statistique (machine learning) en partenariat industriel (Charlotte Laclau, Laboratoire LIG Grenoble) et Flux de travail scientifiques et outils complémentaires pour la reproductibilité en bio-informatique (Sarah Cohen Boulakia, Université Paris-Saclay). Une discussion sur les programmes de financement, ainsi que sur la structure et l'organisation du GdR MaDICS ont animé la mi-journée. Une conférence sur la « Gouvernance des masses de données, Questions éthiques et juridiques », présentée par Danièle Bourcier (CNRS, Commission d'éthique du numérique, Alliance ALLISTENE) a clôturé la première journée de conférences. En fin de journée, la session « poster » a donné à une quinzaine d'étudiants d'horizons très différents la possibilité de présenter leurs sujets de recherche.

La deuxième journée a démarré par une conférence d'introduction à la technologie « Blockchain » par Jean-Luc Parouty (Institut de Biologie Structurale). Ensuite, plusieurs groupes de travail MaDICS (actions) se sont réunis en session parallèle : GRAMINEES (GRaph data Mining in Natural, Ecological and Environnemental Sciences) organisé par le consortium IndexMEED (Indexing for Mining Ecological and

Environmental Data), PREDON (Préservation des données scientifiques), MAESTRO (MAsses de données En aSTRonomie et astrophysique), RoD (Raisonnement sur les données - Reasoning on Data). Plusieurs ateliers se sont également déroulés autour des thématiques comme : Masses de Données et santé, Qualité des masses de données scientifiques, Indexation de grandes masses de données (en lien avec le GdR ISIS). Une réunion de travail, incluant des démonstrations de la technologie Blockchain (Blockfest 1.3) a également eu lieu.



© C.Diaconu/ CNRS

Avec une participation de plus de cent quarante personnes, les journées ont constitué une excellente occasion de passer en revue les sujets phares en science de données, d'échanger avec des experts invités ainsi que de continuer le dialogue entre les communautés très différentes représentées au sein du GdR MaDICS.

Pour plus de détails : <https://www.madics.fr>.

**Cristinel DIACONU (CPPM), Romain DAVID (IMBE), Christian SURACE (LAM)**



## PSPA : Design your Accelerator !



PSPA est un projet né il y a quelques années sur une idée d'Alessandro Variola, alors responsable du département Accélérateurs, au Laboratoire de l'Accélérateur Linéaire d'Orsay. Il vise à fournir aux concepteurs d'accélérateurs de particules un outil universel et convivial permettant d'exécuter différents codes de simulation d'accélérateurs existants de la discipline en les combinant facilement entre eux.

Aujourd'hui, pour concevoir ou étudier un accélérateur de particules, un « physicien machine » dispose d'un nombre important de codes de modélisation et de simulation, essentiellement open source, produits par la communauté des accélérateurs. L'usage de la plupart d'entre eux est libre, mais leur facilité de mise en œuvre (disponibilité, systèmes d'exploitation supportés, documentation, convivialité d'utilisation...) est très variable, avec un grand nombre de paramètres de configuration.

Ce physicien machine doit dès lors faire face à une multitude de problèmes :

- Il doit récupérer le code qui l'intéresse et l'installer sur les ressources informatiques dont il dispose
- Une étude de machine étant un problème complexe, elle fait généralement appel à l'enchaînement de plusieurs programmes différents, incompatibles entre eux : au problème précédent, multiplié par le nombre de programmes en jeu, s'ajoute celui de la conversion des données du format d'un programme à l'autre, un processus fastidieux, sujet à erreurs et consommateur de temps
- Il doit gérer lui-même les données générées par les programmes, aucune automatisation de cette gestion n'étant fournie, du fait même de la nature « stand-alone » des programmes
- Souvent, il souhaite comparer, en général sur une partie précise de la machine, les simulations fournies par deux programmes différents, soit que ces programmes représentent deux implémentations indépendantes de la même modélisation du processus physique étudié, soit qu'ils fassent chacun appel à des modélisations différentes ; cette opération combine toutes les difficultés évoquées ci-dessus, l'amenant généralement à renoncer.

PSPA s'est donné pour ambition de répondre à ces problématiques, en fournissant une plateforme accessible sur le WEB qui permette, de façon interactive :

- de définir graphiquement tout ou partie d'un accélérateur à l'aide d'un ensemble de composants élémentaires traditionnels de la discipline, que l'utilisateur peut combiner à sa guise, un peu comme dans un jeu de construction ; alternativement, PSPA peut aussi importer une machine décrite dans certains formats existants ;
- d'indiquer à quels composants, individuellement ou regroupés en secteurs, quel programme de simulation de fonctionnement, choisi dans un catalogue aussi large que possible, doit être

appliqué ;

- après en avoir fixé les paramètres, de lancer l'exécution de la simulation du système ainsi défini, et récupérer les sorties des programmes mis en jeu, soit sous forme de données numériques, soit sous forme de courbes.

L'état d'avancement actuel de PSPA peut être considéré au stade « prototype », puisqu'il a été utilisé avec succès pour simuler le fonctionnement d'un accélérateur complexe, ThomX, actuellement en cours de construction au LAL. Cependant, les défis à relever pour en faire une plateforme pleinement opérationnelle au sein de la communauté des accélérateurs restent nombreux. Parmi eux, on peut citer :

- La représentation réaliste de l'accélérateur étudié ; pour l'instant, on ne dispose que d'un alignement de boîtes symbolisant chaque composant ou secteur
- L'amélioration de l'interface utilisateur, à la fois sous son aspect « design » et sous son aspect ergonomique ; les compétences dans ce domaine étant rares dans notre environnement, nous envisageons sérieusement de sous-traiter ce problème à un prestataire extérieur, avec transfert de compétence
- La refonte du moteur d'orchestration des programmes gérés par le système (les codes de la discipline), afin d'améliorer (faciliter, rendre robuste) le processus d'ajout de nouveaux codes, et surtout de permettre la prise en compte de codes développés par l'utilisateur lui-même
- La mise en place du système de gestion des données évoqué ci-dessus, permettant à l'utilisateur d'automatiser ses campagnes de simulations, et d'archiver les résultats pour une consultation ultérieure efficace
- La possibilité de lancer le système en mode déconnecté, pour permettre l'exécution de simulations longues (plusieurs heures ou jours)
- La possibilité d'exécution en mode « élastique » sur le cloud
- Une version stand-alone
- Une version didactique pour l'enseignement et la dissémination

Note : la « *dockerisation* » de PSPA, en cours de réalisation actuellement, prépare ces dernières étapes.

Face à ces perspectives, nous sommes à la recherche de forces vives (et enthousiastes) d'une part chez les informaticiens pour participer aux développements, et d'autre part chez les physiciens machine pour tester la plateforme et surtout partager leurs pratiques d'étude et conception d'accélérateurs pour adapter le plus possible l'interface de PSPA aux manières de faire de la discipline. Si vous êtes intéressé, n'hésitez pas à **nous contacter**.

PSPA est un projet ambitieux. Durant ces derniers mois, des contacts positifs ont été établis tant avec la communauté accélérateurs nationale qu'internationale. Sans réel équivalent, même si certaines applications traitent une partie du problème, il suscite un vif intérêt. Son adoption par les concepteurs d'accélérateur en physique des hautes énergies contribuerait de façon significative à la renommée de l'IN2P3 au sein de cette communauté. Rejoignez-nous pour participer à cette aventure !

**Christian HELFT et l'équipe PSPA**



n°37  
Juillet  
2017

# La lettre IN2P3 Informatique

Réseau des Informaticiens de l'IN2P3 et de l'IRFU



## Conférence XLDB-2017, du 10 au 12 octobre 2017 - Clermont-Ferrand



Pour la première fois de son histoire, la conférence XLDB (Extremely Large Databases) se tiendra en dehors des Etats-Unis, à Clermont-Ferrand (ou plus exactement Royat), en France. L'organisation de cet événement est co-pilotée par le **LIMOS (Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes)**, rattaché à l'INS2I, et le Laboratoire de Physique de Clermont, rattaché à l'IN2P3. Fortement impliqués dans les Big Data, du fait de leurs contributions en recherche et développement pour la gestion des grandes masses de données cosmologique du télescope LSST et habitués à collaborer, ils se sont associés pour accueillir l'édition 2017 en France.

Grâce à l'évolution rapide et simultanée de la puissance de calcul, des capacités de stockage de données, des réseaux de capteurs et des algorithmes de simulation, les volumes de données atteignent aujourd'hui des échelles extrêmes de l'ordre du PétaOctet, voire de l'ExaOctet. Les solutions de gestion et d'analyse de ces grandes masses de données constituent la pointe des technologies associées à l'explosion actuelle du e-commerce et des activités scientifiques liées au Big Data.

La communauté XLDB (Extremely Large Databases) relève les défis inhérents à la gestion de ces grandes masses de données.

Les principales activités de cette communauté sont :

- L'identification des tendances, des éléments communs et des verrous liées à la gestion et à l'analyse de ces volumes de données.
- La définition de solutions et de stratégies permettant le rapprochement des différentes communautés qui utilisent des systèmes de gestion de données à grande échelle et cela grâce aux experts académiques et industriels du domaine.
- La facilitation du développement et de la croissance des technologies Big-Data, y compris, mais sans s'y limiter, les bases de données.

Initié en 2007 à Stanford, dans la Silicon Valley, XLDB a rapidement suscité un intérêt croissant de la part des acteurs majeurs du Big-Data. C'est pour cela que ses organisateurs ont créé des événements tels qu'une conférence internationale annuelle, des conférences satellites et des événements ciblés sur des sujets à la pointe du domaine.

La conférence ouverte XLDB attire généralement 200 participants. De plus, les événements satellites permettent à la communauté XLDB, centrée sur les États-Unis, de mieux se connecter avec la communauté internationale. Ainsi, des événements ont été organisés au CERN à Genève, à Édimbourg, à Pékin et à Rio de Janeiro. Un workshop axé sur la santé a également été mis en place.

La communauté XLDB est composée :

- à 50% d'utilisateurs industriels, incluant tous les acteurs majeurs du Web, des réseaux sociaux, du pétrole et du gaz, du commerce de détail, de la santé, de la recherche pharmaceutique, des télécommunications, des finances et d'autres secteurs commerciaux,
- à 15% de scientifiques relevant des défis impliquant la gestion de grande masses de données, incluant les laboratoires de recherche majeurs des États-Unis et de nombreux instituts à l'international,
- à 15% d'industriels fournisseurs de solutions Big-Data - incluant tous les principaux éditeurs logiciels (base de données et autres) et les fournisseurs d'infrastructure de stockage de pointe,
- à 15% de chercheurs du monde académique,
- à 5% d'autres acteurs, incluant les organismes de financement et les capital-risqueurs.

L'impact des activités XLDB comprend à la fois des résultats tangibles :

- le démarrage de SciDB, un logiciel libre de gestion de base de donnée, basé sur les tableaux multi-dimensionnels, dans lequel s'implique notamment Mike Stonebraker, titulaire du prix Turing,
- une communauté de plus de 1000 personnes,
- l'initialisation d'une liste de cas d'utilisation, en évolution rapide. et d'autres avancées moins tangibles mais importantes tel que la mise en place d'un réseau pluridisciplinaire et une influence certaine sur l'orientation des technologies de type Big-Data.

Les possibles activités à venir incluent :

- la définition de solutions génériques dédiées à des cas d'utilisation communs,
- la création d'un groupement d'experts XLDB sur lesquels la communauté entière pourrait s'appuyer,
- la communication auprès de la nouvelle génération, notamment auprès des étudiants.

Nous invitons tous les acteurs du domaine du Big-Data à nous rejoindre pour participer à cet événement d'exception.

Pour plus d'informations, consulter [notre site web](#), ou [contactez-nous](#).



## Journées SUCCES 2017, 16 et 17 octobre - Grenoble

Dédiées aux utilisateurs des méso-centres et des infrastructures de grilles ou de cloud, les journées SUCCES 2017 sont organisées par les GIS FRANCE GRILLES et GRID5000, le Groupe Calcul et le GDR RSD.

Les objectifs de ces rencontres sont de présenter des travaux scientifiques, dans toutes les disciplines, réalisés grâce au soutien des infrastructures de grilles de calcul, de méso-centres ou de cloud.

Elles sont aussi l'occasion de présenter aux utilisateurs l'évolution des outils et infrastructures, via la recherche en informatique associée.

Ces journées associeront exposés pléniers et tables rondes sur des sujets d'actualité.

Ces journées se dérouleront les lundi 16 et mardi 17 octobre 2017 dans les locaux de GRICAD à Grenoble.

Les inscriptions et l'appel à contribution sont ouverts sur [le site de l'événement](#).



## 10èmes journées méso-centres, 26 et 27 septembre 2017 - Paris

Les 10èmes journées méso-centres auront lieu les 26 et 27 septembre à l'Institut Henri Poincaré à Paris.

Au programme cette année, vous trouverez entre autres des présentations techniques, des ateliers, des débats, et des événements centrés sur les 10 ans de ces journées.

Les inscriptions sont ouvertes, gratuites et obligatoires pour des questions d'organisation.

Programme prévisionnel et inscriptions sur [le site de l'événement](#).



## HEPiX Fall/Autumn 2017, du 16 au 20 octobre - KEK, Japon

La **conférence d'automne HEPiX** est organisée du 16 au 20 octobre à KEK, au Japon.

Programme, appel à contributions et inscriptions sur [le site indicatif de l'événement](#).



n°37  
Juillet  
2017

# La lettre IN2P3 Informatique

Réseau des Informaticiens de l'IN2P3 et de l'IRFU



## 558 inscrits aux JDEV 2017 à Marseille

L'édition 2017 des Journées du Développement Logiciel de l'Enseignement Supérieur (ESR) et de la Recherche s'est terminée le 7 juillet dans les locaux d'Aix-Marseille Université sur la Canebière. Presque 400 participants inscrits et plus de 100 exposants, organisateurs, intervenants et sponsors ont fait de cet événement bisannuel encore une fois une des manifestations les plus rayonnantes de la communauté.

En attendant l'article plus détaillé sur le programme et les impressions de quelques participants sélectionnés de l'IN2P3, dans la prochaine édition de votre Lettre Informatique, rappelons les grandes lignes du programme, les huit thématiques : **Systemes embarqués, les réseaux de capteurs et l'internet des objets, Ingénierie et web des données, Programmation de la matière, fabriques personnelles, Usines logicielles et outils de production de code, Infrastructures logicielles et science ouverte, Méthodes et techniques de production de logiciel, Science des données et apprentissage automatique, Parallélisme itinérant, virtualisation et reproductibilité.**

Les JDEV sont organisées depuis 2011 comme Action Nationale de Formation (ANF) du CNRS, avec le concours de l'INRA, d'Inria, de l'IRD et de l'Irstea. Elles sont de plus en plus ancrées dans le paysage techno-scientifique au-delà même de l'ESR, ce qui s'est exprimé par la présence de sponsors de renom à cette édition, ainsi qu'une allocution de bienvenue donnée par le Délégué au Numérique, Daniel Sperling, de la Mairie de Marseille, représentant le Sénateur-Maire Jean-Claude Gaudin.



## De l'intérêt des forums ORAP

Le forum **ORAP** a lieu tous les six mois en France et il est à mes yeux, presque l'équivalent européen de la conférence Super-Computing qui se tient aux États-Unis ! Des chercheurs en informatique viennent y présenter leurs travaux (nouveaux algorithmes hautes performances, nouvelles architectures ou micro-architectures machine, nouveaux cas d'utilisation, nouvelles méthodes de génération de code, nouvelles méthodes d'adaptation d'un code à une architecture données, etc.) dans des domaines variés, tels que la physique des particules (simulation de trajectoire de particules chargées dans différents milieux), l'astrophysique (problème à N corps) mais aussi, la biologie (calcul de dose pour des patients, simulation du cerveau humain, simulation de forêts), la physique des plasmas, la mécanique des fluides, la simulation de population, etc.

Personnellement, je travaille sur le projet ESFRI H2020 ASTERICS et je suis allé à tous les forums ORAP depuis la 35<sup>ème</sup> édition. Le contenu est, à chaque fois, très instructif. Le **39<sup>ème</sup> forum** s'est tenu en mars dernier au siège du CNRS ; le 40<sup>ème</sup> annoncé pour la mi-octobre est « à ne pas manquer ».

Aujourd'hui, le constat est sans appel : la moyenne de l'utilisation du processeur par toutes les applications dans le monde est inférieure à 4%, et les applications dédiées à l'analyse des données physique n'échappent pas à la règle. Mon travail consiste à fournir des fonctions ultra-optimisées et des méthodes de création de format de données qui vont permettre des calculs rapides et qui utilisent le processeur à 100% dans tous les cas !

En règle générale, les applications présentées dans le cadre des forums ORAP font chacune une utilisation différente des architectures informatiques existantes (CPU, GPU, GPGPU, FPGA, multicore, manycore et architectures hybrides) et c'est là tout l'intérêt !

Tous ces retours d'expériences me semblent utiles pour mettre au point des outils d'analyse qui seront utilisés, notamment dans les expériences du **programme ASTERICS**, telles que CTA, SKA ou LSST. Ces expériences, de par leur conception, posent le problème de l'analyse de données à une échelle 1000 fois supérieure que celle des expériences précédentes. Ils sont utiles à relativement court terme car ils rendent compte d'une utilisation particulière d'architectures relativement bien connues et utilisables, ce qui permet d'orienter à quelques années les développements pour leur permettre de ne pas tomber dans une totale obsolescence comme ce fut le cas de plusieurs programmes d'analyse par le passé.

Mais lorsque l'on parle d'expériences maintenues pendant 30 ans, comme CTA ou LSST, il est indispensable d'avoir aussi une vision à long terme. Lors des forums ORAP, les présentations les plus attendues sont celles des grands industriels du domaine de l'informatique hautes performances, tel que Intel, Nvidia ou Bull qui présentent leurs futures architectures CPU, GPU et centre de calcul respectivement. D'autres industriels tel que Google, Facebook ou Total, présentent les méthodes qu'ils utilisent pour traiter massivement les données qu'ils reçoivent chaque seconde (par exemple 100 heures de vidéo toutes les secondes sur Youtube). Ces retours sont importants car ils nous permettent d'envisager ce que seront les flux de données des expériences CTA ou LSST, et d'avoir une première estimation des architectures à utiliser pour résoudre ce genre de problème.

D'autres informations cruciales proviennent des plus grands centres de calcul au monde. Par exemple, lors du forum ORAP du 28 octobre 2016 a été présenté le retour d'expérience sur le plus grand centre de calcul au monde, en Chine. Ce super ordinateur a une puissance de 93Tflops et a été testé avec des algorithmes de calcul LAPACK (multiplication matrice-matrice, matrice-vecteur, vecteur-vecteur, décomposition SVD, QR, etc). Ce test a révélé que les algorithmes actuels rencontrent des problèmes de passage à l'échelle lorsque l'on se rapproche de la barrière des 100 Tflops. Ce résultat n'est absolument pas intuitif et ne pouvait pas être mis en évidence qu'avec un test réel sur une machine existante. L'accès à ce genre de résultat peut se révéler très utile pour qui souhaite définir les architectures des futurs centres de calcul utilisés pour des expériences du programme ASTERICS comme du monde HEP.

Pierre AUBERT (LAPP)





## Cycle de séminaires sur la science des données au CERN

Retrouvez *ici* une nouvelle série intéressante de séminaires sur la science des données au CERN. Toutes les présentations sont transmises via webcast.



## Appel à projets CC-IN2P3

Le CC-IN2P3 fournit des ressources d'analyse et de traitement de données aux laboratoires de l'Institut au travers de sa ferme de calcul. Le CC-IN2P3 a par ailleurs un partenariat avec la société DELL EMC qui lui permet d'évaluer de nouveaux matériels, en particulier des architectures spécifiques (GPU, processeurs many core, nouveaux processeurs, stockage rapide etc.).

Le CC-IN2P3 souhaite aujourd'hui élargir l'utilisation de la plateforme DELL EMC et recenser les besoins en architecture matérielle en avance de phase qui ne seraient pas couverts dans le cadre habituel de la fourniture de calcul et stockage.

Pour ce faire, le CC-IN2P3 lance à titre exploratoire un appel à projets et étudiera toute demande informatique qui pourrait faire l'objet d'un prototype sur la plateforme DELL EMC.

Les réponses à cet appel à projets doivent être envoyées au plus tard le 15 septembre 2017 à [aap@cc.in2p3.fr](mailto:aap@cc.in2p3.fr), sous un format d'une page expliquant le contexte et les objectifs de la demande, les caractéristiques techniques souhaitées et les résultats attendus.

Le CC-IN2P3 étudiera chaque proposition et se réserve le droit de sélectionner les demandes qui lui sembleront pouvoir rentrer dans le partenariat qui le lie à DELL EMC. Le travail d'évaluation qui sera réalisé devra faire l'objet d'un rapport qui pourra être utilisé dans le cadre d'une communication conjointe avec DELL EMC.