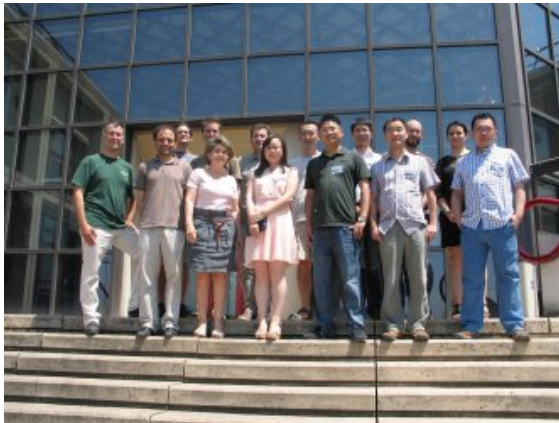




A LA UNE

Une délégation chinoise en visite à l'IN2P3



Les collaborations entre l'IN2P3 et l'IHEP (Institute of High Energy Physics) sont placées depuis 2007 dans un cadre institutionnel, sous la forme d'un laboratoire international associé. Le FCPL (pour France China Particle Physics Laboratory) réunit l'IN2P3, l'Institut de

recherche sur les lois fondamentales de l'Univers du CEA et des universités françaises avec leurs homologues chinois et a pour objectif de développer la coopération entre les deux pays dans les domaines de la physique des particules, de l'astroparticule, des accélérateurs, et les technologies associées.

Depuis plusieurs années déjà, des échanges ont lieu entre le CC-IN2P3 et le Centre de Calcul de l'IHEP de Pékin. Cette collaboration s'est encore renforcée lorsque Fabio Hernandez, ingénieur et ancien directeur adjoint du CC-IN2P3, a été missionné par le Ministère des Affaires Étrangères et Européennes pour développer davantage les liens entre les organismes de recherche français et chinois dans le domaine du traitement de données informatiques. C'est par le biais de son contact que six valeureux globe-trotters, nommés YaoDong CHENG, Lu WANG, BeiJiang LIU, ZhiJia SUN, ZhiMin WANG et Yi ZHANG, tous membres de l'IHEP, ont été amenés à parcourir la France du 9 au 29 juin, et ce malgré la grève des transports, pour découvrir le savoir-faire des (...)

Interview
"La communauté de l'apprentissage automatique est affamée de données provenant de projets scientifiques"



Balázs Kégl, chercheur en informatique au Laboratoire de l'Accélérateur Linéaire.

[lire l'interview](#)

[lire la suite](#)

Ecole GEANT4 - Bilan
La seconde formation Geant4 organisée par l'école doctorale MIPEGE (Modélisation et Instrumentation (...))

[en savoir plus](#)

CERN openlab publie un livre blanc sur les futurs défis informatiques pour la recherche scientifique

Genève, le 22 mai 2014. CERN openlab [1], le partenariat public-privé entre le CERN [2], des (...)

[en savoir plus](#)

Agenda

WLCG collaboration workshop, 7-9 juillet - Barcelone

Le prochain workshop de collaboration WLCG se déroulera à Barcelone du 7 au 9 juillet 2014. Plus (...)

[en savoir plus](#)

Journées Informatiques, 13-16 octobre - Grau du Roi

Réunions du réseau des informaticiens de l'IN2P3 et de l'IRFU qui se tiennent tous les deux ans, (...)

[en savoir plus](#)

Ecole Informatique pour le calcul scientifique : bases et outils, 22-26 septembre - Saint-Dié

Le Groupe Calcul organise une école intitulée "Informatique pour le calcul scientifique : bases (...)

[en savoir plus](#)

ENVOL 2014, 18-21 novembre - Provence

ENVOL - Formation pour le développement et la Valorisation des Logiciels en environnement de (...)

[en savoir plus](#)

Infrastructure



Configuration automatisée avec Puppet

Le gestionnaire de configuration Puppet a bien évolué depuis son tout premier déploiement au Centre de Calcul de l'IN2P3 en 2009. Afin d'exploiter au mieux ses nouvelles fonctionnalités, nous repensons actuellement notre installation. C'est l'occasion parfaite pour une revue de cet outil à la mode. Puppet est un logiciel open source développé par Luke Kanies en 2005. Il se présente comme une alternative à l'écriture de multiples scripts pour automatiser les tâches répétitives liées à la configuration des systèmes d'exploitation (Unix-like et Windows). Produit aujourd'hui par la compagnie PuppetLabs, en version gratuite (3.6.2) et en version commerciale (Puppet Enterprise 3.2), il est utilisé par de nombreuses sociétés renommées telles que Google, Dell, Mozilla, Oracle, Twitter et Disney. Il est également pressenti comme le successeur de YAİM pour la configuration des nœuds de grille de calcul LCG/EGI.

Collaboration



HEP Software Collaboration

Les expériences HEP, dont celles du LHC après les upgrades prévus, vont faire face à une masse de données sans précédent (à la fin de la décennie, facteur 10 par rapport au volume produit actuellement) dont l'analyse représente un challenge pour l'informatique de ces expériences. Un challenge technique d'abord car il est devenu de plus en plus difficile pour les logiciels existants d'utiliser efficacement les machines modernes dont la puissance repose sur un fort parallélisme interne (nombre de cœurs, instructions vectorielles). Mais aussi un challenge financier car, dans la situation budgétaire contrainte partout dans le monde, les expériences LHC peuvent espérer un budget pour l'achat de ressources informatiques au mieux stable, ce qui permet d'espérer une augmentation de la puissance installée correspondant au mieux à la loi de Moore. Sur la base des prévisions actuelles,

Stockage



Bandes magnétiques : le stockage du futur !

Bien que l'utilisation de la bande magnétique soit régulièrement remise en cause, un avenir prometteur semble néanmoins assuré pour ce type de média de stockage, même de nos jours. Alors que la capacité des disques durs traditionnels évolue lentement depuis quelques années, et malgré l'apparition des premiers disques de 6To en 2014 dans nos boutiques en lignes préférées, les capacités stockées sur bandes magnétiques ne cessent de croître avec une progression nettement plus importante que celle constatée pour les disques classiques. Alors que le Centre de Calcul de l'IN2P3 s'apprête à mettre en service la technologie StorageTek T10KD, avec une capacité de stockage non compressée native sur la bande de 8,5To et 252Mo/sec de débit nominal, la société Sony annonce avoir réussi à écrire 185To [1] sur une bande

Challenge



Le Higgs boson machine learning challenge

Le Higgs boson machine learning challenge (HiggsML) est en cours, et s'achèvera le 15 septembre 2014. Pour la première fois, l'expérience ATLAS au CERN a rendu publique une partie des données de simulation utilisées par les physiciens pour optimiser une analyse, en l'occurrence l'analyse qui a mis en évidence la désintégration du boson de Higgs en paires tau+tau- annoncée en novembre 2013. Le challenge est organisé par une équipe interdisciplinaire de physiciens d'ATLAS et d'informaticiens du LAL (Université Paris Sud et CNRS/IN2P3), du LRI (Université Paris Sud et CNRS), d'Inria, de Royal University University College et de ChaLearn. Il est hébergé par la société Kaggle. Le but est d'explorer le potentiel des méthodes avancées d'apprentissage automatique pour améliorer la signification statistique de l'expérience. Au delà, il vise à promouvoir la collaboration entre physiciens

Puppet est une application Ruby pouvant fonctionner de manière autonome (mode standalone), mais plus souvent déployée en environnement client/serveur (mode pull). Le client vérifie régulièrement sa (...)

[lire la suite](#)

les expériences LHC estiment qu'il faudra améliorer la performance des logiciels d'un facteur 10 en 10 ans pour espérer analyser les données avec les ressources qui seront disponibles. Pour les autres expériences de la discipline qui (...)

[lire la suite](#)

magnétique et jusqu'à 370To avec des données compressées. Les perspectives de progression en terme d'évolution de densité au « cm² » sur une bande magnétique restent particulièrement encourageantes (33% pour la bande contre 15% pour le disque [2]). Certes, l'utilisation (...)

[lire la suite](#)

des hautes énergies et informaticiens spécialistes de l'apprentissage automatique. Le challenge se déroule de mi-mai à septembre 2014. Aucune connaissance de physique des particules n'est (...)

[lire la suite](#)

 [Archives](#)

 [Abonnement](#)

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

 [Proposer un article](#)

Vous souhaitez proposer un article ? Envoyez un mail à LettreInformatique@in2p3.fr.



Equipe

Directeur de la publication : Alain FUCHS.

Responsables éditoriaux : Giovanni LAMANNA et Pierre-Etienne MACCHI.

Comité de rédaction : Virginie DELEBARRE DUTRUEL, Christian HELFT, Dirk HOFFMANN et Gaëlle SHIFRIN.



© 2014 CCIN2P3



n°28
Juillet
2014

La lettre IN2P3 Informatique

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

À LA
UNE

Une délégation chinoise en visite à l'IN2P3



Les membres de l'IHEP ont démarré leur séjour par une visite du CC-IN2P3. © CC-IN2P3 / CNRS

Les collaborations entre l'IN2P3 et l'IHEP (*Institute of High Energy Physics*) sont placées depuis 2007 dans un cadre institutionnel, sous la forme d'un laboratoire international associé. Le FCPPL (pour *France China Particle Physics Laboratory*) réunit l'IN2P3, l'Institut de recherche sur les lois fondamentales de l'Univers du CEA et des universités françaises avec leurs homologues chinois et a pour objectif de développer la coopération entre les deux pays dans les domaines de la physique des particules, de l'astroparticule, des accélérateurs, et les technologies associées.

Depuis plusieurs années déjà, des échanges ont lieu entre le CC-IN2P3 et le Centre de Calcul de l'IHEP de Pékin. Cette collaboration s'est encore renforcée lorsque Fabio Hernandez, ingénieur et ancien directeur adjoint du CC-IN2P3, a été missionné par le Ministère des Affaires Etrangères et Européennes pour développer davantage les liens entre les organismes de recherche français et chinois dans le domaine du traitement de données informatiques. C'est par le biais de son contact que six valeureux globe-trotters, nommés YaoDong CHENG, Lu WANG, BeiJiang LIU, ZhiJia SUN, ZhiMin WANG et Yi ZHANG, tous membres de l'IHEP, ont été amenés à parcourir la France du 9 au 29 juin, et ce malgré la grève des transports, pour découvrir le savoir-faire des laboratoires de l'IN2P3.

Financé par l'Académie des Sciences de Chine, ce séjour a été proposé dans le cadre d'un programme d'incitation aux jeunes chercheurs innovants à participer à des actions de formation à l'étranger.

Contraint par la réglementation chinoise, ce séjour a dû se limiter à trois villes : Lyon, Paris et Marseille. Néanmoins un programme riche les attendait. Ils ont ainsi visité les laboratoires

suivants : le CC-IN2P3, le LMA, l'IPNL, le LAL, l'IPNO, l'APC, le LPNHE et le CPPM. Ils ont également assisté à une présentation des activités du LLR. Un parcours dense mais qui leur a permis d'avoir un aperçu de nombreux projets dans lesquels les équipes IN2P3 sont impliquées et de dessiner d'éventuelles collaborations futures.

Les membres de cette délégation ont également pu profiter de leur présence en France pour visiter la plateforme OMEGA, Organisation for micro-electronics design and applications, basée à Massy Palaiseau, ainsi que le Synchrotron SOLEIL. Ils ont également pris une journée pour aller au CERN rencontrer Jamie Shiers du centre de calcul et descendre dans la caverne du détecteur CMS, en compagnie de Denis Linglin, directeur du CC-IN2P3 de 1997 à 2005 et membre de cette collaboration.



Dans la caverne de CMS, en compagnie de Denis Linglin, directeur du CC-IN2P3 de 1997 à 2005. © CC-IN2P3 / CNRS

Dans la caverne de CMS, en compagnie de Denis Linglin, directeur du CC-IN2P3 de 1997 à 2005. © CC-IN2P3 / CNRS

L'objectif de ces visites était de faire connaître le savoir-faire des laboratoires IN2P3 dans plusieurs domaines de la recherche en physique et en computing, et éventuellement donner lieu à des collaborations plus spécifiques à l'avenir. Au vu de la richesse des échanges, gageons que le but a été atteint.

Retrouvez le programme complet de leur visite et leur cv [ici](#). Ainsi qu'un précédent article avec des informations complémentaires sur les relations IN2P3 et IHEP publié dans le **n°23 d'IN2P3 Informatique**.

GUELLE SHIFRIN



n°28
Juillet
2014

La lettre IN2P3 Informatique

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



"La communauté de l'apprentissage automatique est affamée de données provenant de projets scientifiques"

Balázs Kégl, chercheur en informatique au Laboratoire de l'Accélérateur Linéaire.



- Balázs, vous êtes l'un des deux chercheurs en informatique qui exercent leur activité au sein de l'IN2P3. Pourriez-vous nous dire quel a été votre parcours pour rejoindre notre communauté ?

J'ai obtenu mon diplôme d'ingénieur en génie électrique, spécialité informatique, en Hongrie en 1994. Dès le début de mes études, j'ai été attiré par l'apprentissage automatique (machine learning), qui est, pour citer Andrew Ng, la science de permettre aux ordinateurs d'agir sans être explicitement programmés. C'était au temps des premiers succès des réseaux de neurones sur des problèmes pratiques, par exemple la reconnaissance d'écriture manuscrite. Les connexions avec l'intelligence humaine et le cerveau m'intriguaient. Il y avait très peu de théorie derrière ces méthodes, et du fait que j'étais entouré de brillants mathématiciens travaillant sur les statistiques et le traitement du signal, j'ai commencé à faire de la recherche principalement en théorie de l'apprentissage. Au cours de mon doctorat et post-doctorat au Canada, ma fibre technologique a peu à peu pris le dessus et j'ai commencé à travailler plus sur le côté algorithmique-méthodologique de l'apprentissage automatique, mais mon bagage mathématique m'a beaucoup aidé dans la formalisation des problèmes et des solutions tout au long de ma carrière.

J'ai été recruté comme professeur adjoint en 2001 par l'Université de Montréal, où j'étais un chercheur « classique » de l'apprentissage automatique, publiant des papiers dans nos deux grandes conférences annuelles (ICML et NIPS), et travaillant soit sur des problèmes propres à la discipline (amélioration et analyse des méthodes) soit sur des problèmes motivés par des applications pratiques en génie logiciel, en traitement de signal musical et en traitement de l'image. Pendant cette période, j'étais régulièrement (mais pas de façon très intensive) à la recherche d'opportunités pour revenir en Europe, principalement pour des raisons familiales. En 2006, un poste de CR1 sur le thème de la fouille de données (data mining) s'est ouvert au LAL, et j'ai postulé et été recruté. J'ai plongé directement dans le grand bain parce que je n'avais pas suivi de formation académique en physique depuis le lycée, mais j'ai apprécié le défi. Je dois admettre que c'était un tournant risqué dans ma carrière, mais je pensais que la flexibilité et la liberté d'un poste CNRS me permettraient de revenir dans un laboratoire d'apprentissage automatique traditionnel au cas où l'expérience n'aurait pas fonctionné.

Je pense qu'à cette époque seul Guy Wormser, alors directeur du LAL, avait une vision claire de la place que pouvait prendre en physique des hautes énergies la recherche en apprentissage automatique et en fouille de données. Il est frappant de constater que cette vision, un chercheur en science des données immergé dans un laboratoire scientifique où les

choses que je peux maintenant voir clairement en raison de mon expérience en physique.

- Comment voyez-vous l'avenir de la fertilisation croisée entre les sciences statistiques et la physique des hautes énergies ?

Ces 2-3 dernières années, j'ai commencé peu à peu à travailler avec d'autres groupes de physique des particules. Avec l'expérience LHCb, nous travaillons sur l'« apprentissage avec budget » (budgeted learning) pour concevoir des triggers. J'ai deux projets sur un thème que j'appelle « apprendre pour découvrir ». Avec le groupe ATLAS, nous travaillons sur des méthodes multivariées pour optimiser la signification statistique d'une découverte. Nous avons récemment lancé le défi de données publiques HiggsML (<https://www.kaggle.com/c/higgs-boson>), qui a attiré 700 équipes de chercheurs en science des données en un mois ! Il est intéressant de constater combien la communauté de l'apprentissage automatique est affamée de données provenant de projets scientifiques. Le projet scientifique le plus futuriste est celui avec l'équipe Calice (ILC) au LAL et au LLR : nous travaillons sur l'adaptation des techniques d'apprentissage de la représentation profonde (deep representation learning) à la représentation graphique des données du calorimètre, essentiellement en apprenant la physique des particules à des ordinateurs en leur montrant des événements. Ces méthodes ont révolutionné la reconnaissance vocale et la vision par ordinateur au cours des cinq dernières années, et ont le potentiel de réaliser le rêve de 50 ans d'intelligence artificielle.

- Vous avez proposé, puis on vous a demandé de mettre en place un "Center for Data Science" dans le cadre de la Fondation de Coopération Scientifique (FCS) du Campus Paris Saclay, sans doute la plus grosse concentration scientifique jamais réalisée en France. Qu'est ce qu'un "Center for Data Science" ? Comment envisagez-vous de relever ce défi ?

Comme je l'ai déjà dit, ce qui était visionnaire il y a huit ans est devenu un courant majeur aujourd'hui. Les Centres de Données Scientifiques (ou initiatives portant un nom similaire) fleurissent un peu partout dans le monde (New-York, Berkeley, UWashington, Amsterdam, Edimbourg, pour ne citer que les principaux). L'idée est très exactement une généralisation d'AppStat. Aujourd'hui, la communauté de la science des données est dispersée dans différentes disciplines. Nous faisons essentiellement la même recherche en statistique (mathématiques), en apprentissage automatique, en fouille de données, en visualisation de données (informatique), et en traitement de signal (génie électrique). La preuve : nous nous rencontrons régulièrement lors de conférences scientifiques. Le premier objectif du CDS est de former une communauté de chercheurs en science des données à la nouvelle Université Paris-Saclay. Le deuxième grand défi est que les données sont aujourd'hui omniprésentes, et les disciplines qui étaient plutôt pauvres en données dans le passé sont submergées par les données aujourd'hui. Notre objectif est de créer une "agora" où les scientifiques qui détiennent des données (et ont des problèmes d'analyse) peuvent rencontrer des scientifiques qui connaissent (et font de la recherche sur) la méthodologie d'analyse des données. Un troisième et important objectif est d'organiser la construction et la maintenance des outils logiciels qui pourront être utilisés pour l'analyse des données dans plusieurs disciplines. La communauté de la physique a une longue expérience dans ce domaine, paradoxalement beaucoup plus d'expérience que la communauté de l'informatique, et nous sommes impatients d'apprendre et de généraliser cette expérience en forgeant des outils pour une communauté scientifique plus vaste.

données sont produites et analysées, est de plus en plus courante aujourd'hui.

- Pourriez-vous nous décrire brièvement vos domaines de recherche et votre équipe ? Depuis votre arrivée à l'IN2P3, quelle a été votre interaction avec les physiciens de votre discipline ? Comment décrire l'impact de vos propres recherches sur les méthodes informatiques des expériences de physique des particules ? Inversement, avez-vous tiré un enrichissement scientifique de cette interaction ?

Quand je suis arrivé à LAL, j'ai rejoint l'expérience Pierre Auger et le groupe Auger du LAL. Il n'y avait rien de conscient dans ce choix, mais rétrospectivement j'étais probablement impressionné d'approcher une des plus grandes expériences de physique des particules. Cela s'est avéré une sage décision : Auger était un petit groupe avec des physiciens astucieux et ouverts d'esprit, en particulier Marcel Urban, qui a été assez patient pour m'initier aux bases de la physique des particules et de la physique expérimentale en général, principalement au cours de déjeuners à la cantine. J'ai beaucoup apprécié le défi intellectuel d'apprendre une nouvelle discipline, et du coup la phase d'apprentissage a été assez courte. Je peux dire avec une certaine fierté que dans Auger je peux passer pour un physicien sans problème. Il s'avère également qu'une expérience en phase finale a surtout besoin de chercheurs qui savent comment extraire de la connaissance à partir de données, de sorte que mon expertise dans le domaine s'est révélée assez efficace dans beaucoup de cas où l'analyse des données n'était pas simple.

La mission de mon groupe au LAL (AppStat) s'est décantée assez rapidement : apporter à la physique des techniques d'analyse à l'état de l'art, et aiguillonner la recherche fondamentale en apprentissage automatique et en statistique par de vrais problèmes de physique. J'ai construit rapidement une équipe, grâce à deux projets ANR, MetaModel et Siminole. Pendant ces huit années, j'ai dirigé la thèse de trois étudiants en informatique et (partiellement) de deux étudiants en physique, et encadré quatre postdocs, un ingénieur et un chercheur invité. Je collabore étroitement avec des groupes d'apprentissage automatique et de statistique à Saclay, principalement avec Michèle Sebag et Cécile Germain du LRI et Olivier Cappé et Gersende Fort de Telecom ParisTech. Environ la moitié de mes publications sont en apprentissage automatique « pur » (sans motivation par de la physique) ; il était important de rester en contact avec ma communauté. Fait intéressant, la physique m'a aussi motivé à élargir mon horizon sur le plan méthodologique, puisque certains problèmes n'auraient pas pu être résolus par des techniques que je connaissais auparavant. Un autre bénéfice de travailler avec des physiciens, c'est que j'ai appris à leur contact l'humilité et la rigueur dans la conduite d'expériences sur ordinateur. En apprentissage automatique, c'est la créativité qui prime : nous inventons des méthodes qui doivent améliorer de façon significative les techniques existantes pour être publiées. Habituellement une amélioration est mesurée sur des jeux de données de référence, mais souvent ces expériences sont loin d'être rigoureuses, quelque

Le CDS, pour le moment, est un projet de deux ans. Nos deux outils principaux sont le financement de 10 à 15 projets interdisciplinaires et de 3 à 6 thèses, et l'organisation de journées thématiques et de séances de brainstorming informelles. Nous sommes actuellement dans le processus d'évaluation des propositions à notre premier appel. Nous allons lancer un deuxième appel à projets au niveau doctorat à l'automne prochain, et, s'il nous reste encore de l'argent, nous aurons un deuxième appel à projets au début de 2015. Nous prévoyons d'organiser 8 à 10 journées thématiques au cours des deux prochaines années, autour de thèmes tant méthodologiques que scientifiques. Outre la physique des particules, nous voyons se former de fortes collectivités en sciences de données autour des neurosciences, des sciences de l'environnement et de la terre, de l'économie, de la biologie et de la chimie, et de l'astrophysique et de la cosmologie.

Outre ces objectifs à court terme, nous allons également rapidement commencer à concevoir une stratégie à long terme. Nous avons une forte motivation venant de la base, et, en même temps, la pression de la FCS pour mener à bien ce travail. C'est un moment passionnant : en un sens, nous avons à inventer l'avenir de la science des données pour les données de la science. La tâche est difficile : nous devons trouver une façon d'initier puis d'organiser des projets interdisciplinaires temporaires et à inciter les gens à investir dans la construction d'outils. Saclay est un endroit idéal pour cela en France : nous avons la masse critique (plus de 250 chercheurs sont associés au CDS), le soutien formel de la FCS et d'autres acteurs (laboratoires, écoles, universités, et Labexes), et un paysage institutionnel particulièrement fluide en ce moment, par ailleurs en évolution du fait de la mise en place d'UPSay.

Un défi majeur qui semble passer largement inaperçu de nos tutelles est la "fuite des cerveaux" sans précédent des chercheurs en science des données vers la recherche privée. La plupart des grandes entreprises dites « IT » (technologie de l'information) comme Google, Microsoft, Facebook, Amazon, Baidu, Criteo, etc. mettent sur pied rapidement des laboratoires de recherche à part entière, offrant des salaires et des infrastructures technologiques avec lesquels le secteur public ne peut pas rivaliser, et, en même temps, nous apportent des problèmes scientifiques redoutables et excitants dans les sciences sociales, l'ingénierie et la recherche en intelligence artificielle. Et ceci au moment où les étudiants commencent également à découvrir la science des données et s'inscrivent en masse dans nos classes. Sans une stratégie concertée des instituts nationaux et de l'enseignement supérieur, nous allons être confrontés à une grave pénurie dans la recherche et l'enseignement des sciences de données publiques (*open data*) dans un proche avenir. Les CDS ne peuvent pas résoudre tous ces problèmes, mais nous pouvons être une part importante de la solution.

Propos recueillis et traduits de l'anglais par le comité de rédaction



Configuration automatisée avec Puppet



Le gestionnaire de configuration Puppet a bien évolué depuis son tout premier déploiement au Centre de Calcul de l'IN2P3 en 2009. Afin d'exploiter au mieux ses nouvelles fonctionnalités, nous repensons actuellement notre installation. C'est l'occasion parfaite pour une revue de cet outil à la mode.

Puppet est un logiciel *open source* développé par Luke Kanies en 2005. Il se présente comme une alternative à l'écriture de multiples scripts pour automatiser les tâches répétitives liées à la configuration des systèmes d'exploitation (Unix-like et Windows).

Produit aujourd'hui par **la compagnie PuppetLabs**, en version gratuite (3.6.2) et en version commerciale (Puppet Enterprise 3.2), il est utilisé par de nombreuses sociétés renommées telles que Google, Dell, Mozilla, Oracle, Twitter et Disney. Il est également pressenti comme le successeur de YAIM pour la configuration des nœuds de grille de calcul LCG/EGI.

Puppet est **une application Ruby** pouvant fonctionner de manière autonome (mode *standalone*), mais plus souvent déployée en environnement client/serveur (mode *pull*). Le client vérifie régulièrement sa configuration en s'adressant au serveur Puppet, via le protocole HTTP en mode sécurisé SSL, pour récupérer le descriptif de sa configuration, appelé catalogue de ressources (fichiers, services, paquets...).

Le catalogue est généré par le serveur Puppet à partir des manifestes contenant le code de configuration et d'informations issues du client, appelées *facts* (nom d'hôte, OS, noyau, IP, adresse mac...). Des informations additionnelles peuvent aussi être fournies par un tiers, l'*External Node Classifier* (ENC). Le service sur le client compare l'état des ressources locales à celui du catalogue reçu et procède aux modifications nécessaires, décrites dans les manifestes.

Au Centre de Calcul de l'IN2P3, quatre serveurs Puppet gèrent la configuration de plus de 1300 clients (*worker nodes*, nœuds de grille, serveurs de sauvegarde/données/cloud...), à partir de 138 manifestes. L'ENC est la base de données d'une application interne de gestion de parc (**SMURF**). Elle fournit l'information indispensable pour créer une configuration spécifique au rôle de chaque machine.

Parmi les principaux atouts de Puppet, il faut citer :

- Sa couche d'abstraction de ressource ou *Resource Abstraction Layer* (RAL), rendant le langage de configuration indépendant de l'implémentation système. Le code est donc

intuitif. Il suffit de décrire dans les manifestes l'état désiré des ressources, sans se préoccuper de la façon de parvenir à cet état.

- Les environnements qui offrent un moyen simple d'appliquer, à un ou plusieurs clients, différentes sortes de configuration : test, production, développement...

- Une gestion des données séparée du code avec **Hiera**, une base de données hiérarchique, supportée nativement depuis la version 3 de Puppet. Cet outil permet de factoriser les données de configuration des différents clients et de retrouver facilement les informations relatives à une machine.

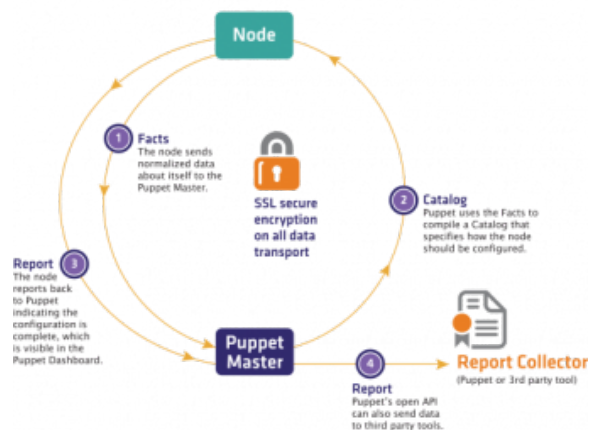
- Le partage de codes grâce aux modules, qui contiennent tout ce qui est nécessaire à la configuration d'un composant (NTP, Apache, MySQL, firewall...). D'autres apportent des fonctionnalités supplémentaires aidant à l'écriture des manifestes (p. ex. création de fichiers, manipulation de chaînes de caractères, de nombres). Beaucoup sont disponibles gratuitement dans **Puppet Forge** ou **GitHub**.

- La PuppetDB qui stocke les ressources externes de chaque client. Ce sont les seules accessibles par les autres clients. Elles sont utilisables, par exemple, pour mettre à jour la configuration d'un serveur Nagios à l'installation d'un nouveau nœud à *monitorer*.

- Une communauté active fournissant une documentation complète et accessible.

Code et données peuvent être intégrés à un gestionnaire de version historisant les changements et autorisant les administrateurs d'un site à déléguer une partie de la configuration aux utilisateurs.

En résumé, Puppet est un outil pragmatique offrant, pour la gestion automatique de la configuration système, des solutions élégantes et applicables à bien des situations.



Christelle ELOTO (CC-IN2P3)



n°28
Juillet
2014

La lettre IN2P3 Informatique

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



HEP Software Collaboration



Les expériences HEP, dont celles du LHC après les upgrades prévus, vont faire face à une masse de données sans précédent (à la fin de la décennie, facteur 10 par rapport au volume produit actuellement) dont l'analyse représente un challenge pour l'informatique de ces expériences. Un challenge technique d'abord car il est devenu de plus en plus difficile pour les logiciels existants d'utiliser efficacement les machines modernes dont la puissance repose sur un fort parallélisme interne (nombre de cœurs, instructions vectorielles). Mais aussi un challenge financier car, dans la situation budgétaire contrainte partout dans le monde, les expériences LHC peuvent espérer un budget pour l'achat de ressources informatiques au mieux stable, ce qui permet d'espérer une augmentation de la puissance installée correspondant au mieux à la loi de Moore. Sur la base des prévisions actuelles, les expériences LHC estiment qu'il faudra améliorer la performance des logiciels d'un facteur 10 en 10 ans pour espérer analyser les données avec les ressources qui seront disponibles. Pour les autres expériences de la discipline qui verront le jour, la contrainte sera vraisemblablement similaire.

Pour faire face à ce défi, le CERN a pris l'initiative de créer un "Concurrency Forum" il y a deux ans avec des experts des différentes expériences pour mener une R&D sur les voies technologiques possibles. Plusieurs ont été explorées et aucune ne suffira à elle seule. Mais une conclusion s'est imposée : HEP n'échappera pas un reengineering majeur de son « patrimoine logiciel ».

D'un autre côté, HEP n'est pas seul à faire face au défi du déluge de données : c'est sans doute le grand changement de la dernière décennie. Cela ne signifie pas que des solutions toutes faites, prêtes à l'emploi, existeraient et qu'il suffirait de les importer. La physique HEP conserve des caractéristiques particulières qui rendent une grande partie des problèmes fondamentalement séquentiels, ce qui constitue un handicap pour exploiter des architectures de processeurs de plus en plus parallèles ! Cependant, l'expertise autour des approches BigData et de la parallélisation est présente dans d'autres communautés scientifiques et au-delà dans des sociétés privées comme les « géants » de l'Internet (Google, Facebook...). HEP a donc tout à gagner à établir des collaborations avec ces autres communautés ayant développé un savoir-faire sur ces technologies.

C'est à partir de ce constat qu'a germé l'idée de créer une « HEP Software Collaboration » qui aiderait à organiser les synergies entre les différents projets logiciels de la communauté HEP pour éviter les duplications coûteuses et faire émerger les composants communs spécifiques à la communauté. Il s'agit bien d'une initiative à l'échelle de HEP et non pas du CERN, puisqu'une bonne partie de ce patrimoine logiciel est commun à

l'ensemble de la communauté (GEANT4 et ROOT par exemple, pour ne citer qu'eux). Cette collaboration fournirait aussi un cadre pour établir des relations plus formelles avec d'autres communautés et pour constituer les « consortiums » ad hoc à même de postuler à des demandes de financement dans les programmes de type H2020 ou équivalent dans d'autres pays/continents, programmes où l'aspect large communauté est souvent important.

Le premier acte du processus de création de cette communauté s'est déroulé début avril au CERN, sous la forme d'un « **kick-off meeting** » qui a réuni environ 150 personnes, principalement mais pas uniquement de la communauté. Ce fut l'occasion de nombreuses présentations (voir agenda) et de discussions assez animées ! Il a été en particulier rappelé par beaucoup d'intervenants que les logiciels phares de la communauté (ROOT en est l'exemple le plus significatif sans doute) ne sont pas nés d'une décision du management mais au contraire de la volonté de répondre à un besoin non pris en compte par les « décideurs » de la communauté, souvent contre la volonté des décideurs. De ce constat et du challenge que nous avons devant nous, a émergé le principal élément de consensus : une éventuelle collaboration doit favoriser l'émergence de solutions novatrices, même si leur intérêt ne semble pas évident au prime abord. Elle doit donc reposer sur une approche très bottom-up et agile, où tout projet a le droit d'exister et bénéficie à la fois d'une infrastructure de développement et d'une certaine publicité au sein de la communauté (sur le mode des incubateurs de certaines fondations comme Apache ou Eclipse). Elle doit aussi permettre que s'exerce un « darwinisme » qui fera que certains projets se développeront quand d'autres finiront par disparaître. Et les discussions sur la gouvernance de cette collaboration ont fait ressortir le souhait d'une structure légère, dont les projets, qui conserveraient leur autonomie, seraient les acteurs principaux.

Depuis le kick-off meeting, nous sommes entrés dans une phase d'appel à contributions (« white papers ») exprimant les contours souhaités de la collaboration, ses buts, ses modalités qui se terminera à la fin du printemps. **Ces contributions** peuvent être soit individuelles, soit plus institutionnelles. Nous sommes quelques français de la région parisienne à avoir pris l'initiative d'une contribution disponible sur ce site.

La suite n'est pas encore très claire mais a priori, il va s'agir à partir de ces contributions initiales de définir le consensus autour duquel peut se construire cette fondation et d'en jeter les bases pratiques. Cela passera sans doute par d'autres réunions, peut-être plus restreintes, et prendra probablement encore quelques mois. Je pense qu'il s'agit néanmoins d'une dynamique importante qui permettrait de sortir de deux écueils rencontrés par le développement logiciel dans HEP : le poids du CERN qui a en plusieurs occasions été peu réceptif aux idées novatrices dans la communauté et la forte fragmentation du développement logiciel qui conduit à une assez forte duplication de code qui pourrait bénéficier de packages communs utilisés par des logiciels plus modulaires.

Cette initiative est ouverte à tous ceux qui sont intéressés. Le vecteur de communication principal est une liste de discussion, **HEP-Software-Interest@cern.ch**. N'hésitez pas à vous inscrire **ici**, onglet "Members", lien "Add Me", en obtenant auparavant un "light weight account" **ici**, pour les personnes n'ayant pas de compte CERN.

Michel JOUVIN (LAL)



Bandes magnétiques : le stockage du futur !



Certes, l'utilisation de la bande magnétique nécessite plus d'attention que le disque. En particulier, une gestion étroite des recopies des données de bandes en bandes s'avère nécessaire, à la fois pour récupérer l'espace libéré par l'effacement des données mais également pour revitaliser le média et assurer la longévité du stockage des données. La bande magnétique s'impose cependant comme la solution de stockage de masse de données et comme le champion du rapport consommation électrique sur capacité de stockage. L'avenir s'annonce radieux et prometteur pour le stockage de grande capacité.

Quelques chiffres :

Le stockage sur bandes magnétiques au Centre de Calcul est assuré par :

- 4 bibliothèques StorageTek SL8500 interconnectées entre elles, d'une capacité totale de 40000 bandes magnétiques et de 256 lecteurs de bandes.
- 1 bibliothèque IBM TS3500 d'une capacité maximale de 2500 bandes magnétiques et de 12 lecteurs de bandes.
- Capacité de stockage maximale de 340Po (85 millions de DVD !).

C'est également :

- 20 lecteurs LTO4 => 800Go/bande
- 6 lecteurs LTO6 => 2,5Go/bande
- 16 lecteurs T10KA => 500Go/bande
- 64 lecteurs T10KB => 1To/bande
- 22 lecteurs T10KC => 5To/bande
- 12 lecteurs T10KD => 8,5To/bande

Plus de 30000 bandes de types T10K-T1, T10K-T2, LTO4 et LTO6 pour une capacité utilisée de 30Po.

Benoit DELAUNAY (CC-IN2P3)

Bien que l'utilisation de la bande magnétique soit régulièrement remise en cause, un avenir prometteur semble néanmoins assuré pour ce type de média de stockage, même de nos jours. Alors que la capacité des disques durs traditionnels évolue lentement depuis quelques années, et malgré l'apparition des premiers disques de 6To en 2014 dans nos boutiques en lignes préférées, les capacités stockées sur bandes magnétiques ne cessent de croître avec une progression nettement plus importante que celle constatée pour les disques durs classiques.

Alors que le Centre de Calcul de l'IN2P3 s'apprête à mettre en service la technologie **StorageTek T10KD**, avec une capacité de stockage non compressée native sur la bande de 8,5To et 252Mo/sec de débit nominal, la société Sony annonce avoir réussi à écrire 185To [1] sur une bande magnétique et jusqu'à 370To avec des données compressées. Les perspectives de progression en terme d'évolution de densité au « cm² » sur une bande magnétique restent particulièrement encourageantes (33% pour la bande contre 15% pour le disque [2]).

[1] <http://www.escapistmagazine.com/new...>

[2] <https://tapepower.fujifilmrmd.com/p...> et <http://www.lto.org/technology/gener...>



n°28
Juillet
2014

La lettre IN2P3 Informatique

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



Le Higgs boson machine learning challenge



leaderboard. Un forum très actif permet l'échange entre participants et avec les organisateurs.

Au 20 juin, plus de 700 équipes ont commencé à soumettre des solutions, plus de 10 soumissions par équipe en moyenne. Dès la première semaine, les benchmarks créés avec des algorithmes traditionnellement utilisés dans ATLAS étaient largement battus. Le profil des participants est très divers et tout indique que l'objectif de toucher une grande communauté en dehors de la physique des particules est atteint.

Pour en savoir plus : <http://higgsml.lal.in2p3.fr>

Le Higgs boson machine learning challenge (HiggsML) est en cours, et s'achèvera le 15 septembre 2014.

Pour la première fois, l'expérience ATLAS au CERN a rendu publique une partie des données de simulation utilisées par les physiciens pour optimiser une analyse, en l'occurrence l'analyse qui a mis en évidence la **désintégration du boson de Higgs en paires tau+tau-** annoncée en novembre 2013.

Le challenge est organisé par une équipe interdisciplinaire de physiciens d'ATLAS et d'informaticiens du LAL (Université Paris Sud et CNRS/IN2P3), du LRI (Université Paris Sud et CNRS), d'Inria, de Royal University University College et de ChaLearn. Il est hébergé par la société Kaggle.

Le but est d'explorer le potentiel des méthodes avancées d'apprentissage automatique pour améliorer la signification statistique de l'expérience. Au delà, il vise à promouvoir la collaboration entre physiciens des hautes énergies et informaticiens spécialistes de l'apprentissage automatique.

Le challenge se déroule de mi-mai à septembre 2014. Aucune connaissance de physique des particules n'est nécessaire pour participer. Il est doté de trois prix de plusieurs milliers de dollars ; en outre, les auteurs des méthodes les plus intéressantes pourront être invités au CERN pour discuter de leurs résultats avec des physiciens des hautes énergies.

Les participants sont invités à télécharger sur le site du challenge un échantillon d'entraînement de 250.000 événements (30 variables par événement), où chaque événement est étiqueté « signal » ou « bruit de fond ». Ils mettent au point un algorithme de séparation, basé sur des techniques réseau de neurone, arbres de décision boosté ou autres, dont la qualité est qualifiée par un score, qui est une estimation de la signification statistique de l'analyse. Puis ils téléchargent un échantillon de test de 550.000 événements où les étiquettes ont été masquées, appliquent leur algorithme, et soumettent les solutions sur le site, qui calcule en ligne le score et met à jour un

The poster features the text: "Higgs challenge the HiggsML challenge May to September 2014 When High Energy Physics meets Machine Learning". It includes logos for ATLAS EXPERIMENT, LAL, INRIA, kaggle, and Google. At the bottom, it lists the organization and advisory committees.

Organization committee:
 Balázs Egri - ATLAS/LAL, David Rousseau - ATLAS/LAL, Isabelle Esque - CERN, Thomas Wengler - ATLAS/CERN, Jung Seok Lee - ATLAS/CERN, Cedric Geronzi - I2D/UP, Gino Casse - ATLAS/IRFU, Claire Adam-Bourdoux - ATLAS/LAL, Andrea Bocchi - ATLAS/CERN, Marc Schwaninger - IRFU

David ROUSSEAU (LAL)



n°28
Juillet
2014

La lettre IN2P3 Informatique

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



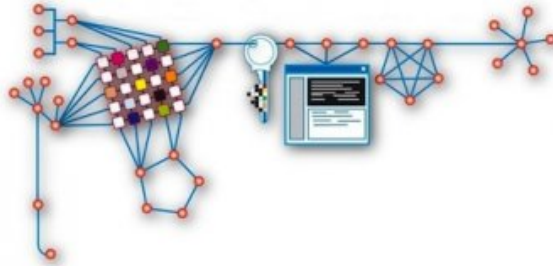
WLCG collaboration workshop, 7-9 juillet - Barcelone

Le prochain workshop de collaboration WLCG se déroulera à Barcelone du 7 au 9 juillet 2014.

Plus d'informations et agenda.



Journées Informatiques, 13-16 octobre - Grau du Roi



Réunions du réseau des informaticiens de l'IN2P3 et de l'IRFU qui se tiennent tous les deux ans, les JI sont l'occasion de mieux connaître ce qui se fait en informatique dans nos instituts (et au-delà), souvent de prendre conscience de technologies émergentes importantes pour l'avenir de nos disciplines, et de réfléchir ensemble au futur de nos activités et aux grandes orientations à envisager.

Cette édition 2014 se déroulera au Grau du Roi, près de Montpellier, **du 13 au 16 octobre** prochain.

Les inscriptions sont désormais ouvertes et seront closes le 3 octobre 2014.

Comme d'habitude, les séjours sont pris en charge par l'organisation, et les voyages sont à la charge des participants.

Programme et inscriptions.



Ecole Informatique pour le calcul scientifique : bases et outils, 22-26 septembre - Saint-Dié

Le Groupe Calcul organise une école intitulée "Informatique pour le calcul scientifique : bases et outils", du 22 au 26 Septembre, à proximité de Saint-Dié (Vosges).

Cette école s'adresse en particulier aux ingénieurs, chercheurs, doctorants, ayant besoin de comprendre les outils de base nécessaires au développement de projets s'appuyant sur du calcul scientifique : bibliothèques mathématiques, compilation, débogage, visualisation scientifique, calcul parallèle. Les exemples traités seront tous issus du domaine du calcul scientifique.

Programme et modalités d'inscriptions.

Les pré-inscriptions sont ouvertes jusqu'au 4 Juillet 2014.



ENVOL 2014, 18-21 novembre - Provence

ENVOL - Formation pour le dEveloppement et la ValOrisation des Logiciels en environnement de recherche : Méthodes de test et validation des logiciels

L'édition 2014 de la formation ENVOL est prévue **du mardi 18 au vendredi 21 novembre** en Provence. Co-organisée par le CNRS (via PLUME, la DSI et des laboratoires du CNRS) avec la participation de l'INRA et d'Inria, l'appel aux candidatures pour une cinquantaine de places sera lancé fin juin.

Depuis 2008, PLUME propose tous les 2 ans environ, une formation pour le dEveloppement et la ValOrisation des Logiciels en environnement de recherche. Les trois premières éditions ont permis de faire un point sur les méthodes et outils de développement de logiciels, sur les moyens de valorisation de ces logiciels (ENVOL2008), avec un aspect plus concret de mise en oeuvre et d'utilisation (ENVOL2010) et enfin la collaboration entre utilisateurs et/ou entre développeurs de codes (libres dans la majorité des cas), en se focalisant sur le partage et la réutilisation de modèles et de codes logiciels dans la communauté ESR (ENVOL2012, **couvert dans la LI24**).

La formation fin 2014 se focalisera sur une introduction aux méthodologies et outils modernes de tests automatiques et systématiques des logiciels, qui permettent d'augmenter et la qualité et la productivité du développement.

Cette édition d'ENVOL, financée dans le cadre des ANF CNRS, a pour ambition de mieux faire connaître la méthodologie de tests systématiques des logiciels, en se limitant dans un premier temps à un survol des techniques et une découverte des méthodes et outils utiles pour la communauté ESR, dans des sessions de travaux pratiques (50% du temps alloué). L'objectif est de montrer que la communauté peut mieux répondre aux besoins de développement logiciel exprimés par les utilisateurs et les développeurs en termes de fonctionnalités, en termes de technologies et de méthodes utilisées, en ayant une meilleure connaissance des méthodes et outils de tests automatiques et systématiques.

Cible

Cette formation s'adresse à des personnels de laboratoires de recherche, chercheurs, ingénieurs ou doctorants, ayant une activité de développement logiciel. Les personnels CNRS seront pris en charge dans le cadre du financement de cette action nationale de formation (ANF) ; les conditions d'inscription pour les personnels d'unités mixtes et hors CNRS seront annoncées sous peu. Chaque participant est invité à venir avec un portable permettant de faire tourner une machine virtuelle utilisée pour les TP proposés.

Pour en savoir plus ...

... allez sur le site PLUME : <https://www.projet-plume.org/envol-2014> qui contiendra des mises à jour régulières.



n°28
Juillet
2014

La lettre IN2P3 Informatique

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



Ecole GEANT4 - Bilan

La seconde formation Geant4 organisée par l'école doctorale MIPEGE (Modélisation et Instrumentation en physiques, Énergies, Géosciences et Environnement) s'est tenue du 19 au 23 mai 2014 dans les locaux du LAL à Orsay. À destination des doctorants et plus largement à tous les utilisateurs de Geant4 (outil logiciel de simulation d'interactions de particules à travers la matière), cette formation a réuni 21 participants (dont 13 doctorants) provenant de toute la France.

Les objectifs pédagogiques étaient de couvrir l'utilisation du framework Geant4 pour les utilisateurs afin de développer leur application. L'accent était mis sur les travaux pratiques. Les supports de cours sont disponibles ici : <http://groups.lal.in2p3.fr/ED-geant4/>

Animée par trois développeurs de Geant4 au sein de l'IN2P3 - Ivana Hrivnacova (IPNO), Laurent Garnier (LAL) et Marc Verderi (LLR) - cette formation répond à une forte demande et s'est remplie en un mois sans publicité particulière. Elle sera probablement reconduite l'année prochaine.



CERN openlab publie un livre blanc sur les futurs défis informatiques pour la recherche scientifique

Genève, le 22 mai 2014. CERN openlab [1], le partenariat public-privé entre le CERN [2], des entreprises leaders du secteur informatique et des instituts de recherche, publie aujourd'hui un livre blanc sur les futurs défis informatiques pour la recherche scientifique, afin de définir des orientations pour sa nouvelle phase, d'une durée de trois ans, qui s'ouvrira en 2015.

Quatre-vingt-seize pour cent de notre Univers est encore inconnu, et les défis qui attendent la communauté scientifique sont gigantesques. Plus que jamais, l'informatique joue un rôle essentiel dans les efforts entrepris pour élucider les mystères de l'Univers. Au cours des dernières années, le volume et le taux de production des données recueillies par les instruments, les détecteurs et les capteurs se sont accrus de façon considérable. Les détecteurs du LHC au CERN produisent une quantité de données impressionnante, de l'ordre d'un pétaoctet par seconde, chiffre appelé à augmenter encore au cours de la prochaine campagne d'exploitation du LHC, qui commencera en 2015. De nouvelles infrastructures de recherche internationales dans différentes disciplines scientifiques telles que la neurologie, la radioastronomie ou la génétique, se mettent en place. S'appuyant sur des instruments aussi divers que les satellites d'observation de la Terre, les séquenceurs de génome haute performance, les diffractomètres de neutrons ou les télescopes à rayons X, elles devraient produire des volumes de données comparables, voire supérieurs à ceux du LHC. Plus que jamais, la collaboration jouera un rôle primordial pour faciliter les futures découvertes.

Dans ce contexte, CERN openlab, ainsi qu'un certain nombre de laboratoires européens, tels que l'EMBL-EBI, l'ESA, l'ESRF, l'ILL et des chercheurs du projet Human Brain Project, aidés de contributions d'entreprises leaders du secteur informatique, ont publié un livre blanc. Celui-ci a pour objet de définir les défis ambitieux que suppose la couverture des besoins les plus cruciaux en matière d'infrastructures informatiques dans des domaines tels que l'acquisition de données, les plateformes de calcul, les architectures de stockage de données, la mise en service et la gestion des capacités de calcul, les réseaux et la communication, et l'analyse des données. Un certain nombre de scénarii d'utilisation dans différents domaines scientifiques et technologiques sont décrits pour chacun des six grands champs de recherche.

Une collaboration continue entre les infrastructures de recherche et les entreprises informatiques est plus que jamais nécessaire, pour assurer l'adéquation entre les objectifs scientifiques et les feuilles de route technologiques. Dans la phase actuelle du projet CERN openlab, Huawei, Intel, Oracle et Siemens sont des partenaires du projet, tandis que Rackspace a le statut de contributeur et Yandex celui d'associé. Le livre blanc, qui est le résultat de six mois de réflexion entre experts du domaine de l'informatique et de la recherche scientifique, constitue un cadre très stimulant pour le partenariat public-privé CERN openlab dans les années à venir. Le document fixe des objectifs, définit le savoir-faire technique et les programmes éducatifs requis, ouvrant ainsi des opportunités pour une future collaboration entre le CERN, d'autres laboratoires européens, des projets scientifiques internationaux et de grandes entreprises informatiques, afin de repousser plus avant les limites et favoriser les découvertes scientifiques majeures de demain.

Pour en savoir plus :

Contact :

Responsable communication CERN openlab
melissa.gaillard@cern.ch
+41 (0)22 767 50 49

[1] CERN openlab est un partenariat public-privé unique entre le CERN et des entreprises leaders du secteur informatique. Sa mission est d'accélérer le développement de solutions de pointe utiles à la communauté mondiale du LHC ainsi qu'aux instituts de recherche partenaires. Dans ce cadre, les partenaires ont accès à l'expérience du Laboratoire en matière d'ingénierie et à son infrastructure informatique complexe, étendue dans certains cas à des instituts de recherche dans le monde entier. Les tests effectués dans des environnements exigeants et des conditions extrêmes poussent la technologie à ses limites, donnant ainsi aux partenaires de l'industrie informatique un précieux retour d'information sur leurs produits, tout en permettant au CERN et aux laboratoires de recherche associés d'évaluer les mérites de nouvelles technologies dans leurs premiers stades de développement, en vue d'une éventuelle utilisation future.

[2] Le CERN, Organisation européenne pour la Recherche nucléaire, est le plus éminent laboratoire de recherche en physique des particules du monde. Il a son siège à Genève. Ses États membres actuels sont les suivants : Allemagne, Autriche, Belgique, Bulgarie, Danemark, Espagne, Finlande, France, Grèce, Hongrie, Israël, Italie, Norvège, Pays-Bas, Pologne, Portugal, République tchèque, Royaume-Uni, Slovaquie, Suède et Suisse. La Roumanie a le statut de candidat à l'adhésion. La Serbie est État membre associé en phase préalable à l'adhésion. Les États-Unis d'Amérique, la Fédération de Russie, l'Inde, le Japon, la Turquie, la Commission européenne et l'UNESCO ont le statut d'observateur.