



A LA
UNE

Journée des expériences 2017



La Journée des Expériences a eu lieu le 30 janvier au Centre de Calcul de l'IN2P3. Lors de cette réunion est invité l'ensemble des représentants des expériences utilisatrices des ressources du CC-IN2P3 pour échanger sur les services mis à disposition et leur utilisation.

Pierre-Etienne Macchi, directeur du CC-IN2P3, a présenté un bilan de l'année 2016, évoquant notamment la capacité du centre et l'utilisation des ressources. Les demandes des expériences pour 2017 seront couvertes. Volker Beckmann, Directeur Adjoint Scientifique Calcul et Données de l'IN2P3, a ensuite donné le point de vue de l'Institut sur l'utilisation du Centre de Calcul à travers les projets et groupes de travail « Calcul et données ». Il a également souligné l'importance du Data Management Plan (DMP) pour structurer les données et leur gestion.

S'en est suivi un bilan des activités et des évolutions des services du Centre en 2016 qui a rappelé notamment la refonte des plate-formes interactive et parallèle, la création d'une ferme GPU, la mise à disposition du NoSQL via MongoDB et à partir de cette année l'envoi aux groupes d'utilisateurs d'un DMP à remplir. Les perspectives pour 2017 prévoient, entre autres, une optimisation de la configuration du système de batch, du travail sur l'exécution et l'orchestration des conteneurs, la poursuite des travaux en vue de l'abandon (...)

Interview

"Il va falloir opérer des choix et nous souhaitons le faire avec le maximum de paramètres et d'informations."



Christian Olivetto, Directeur Adjoint Technique à l'IN2P3

[lire l'interview](#)

[lire la suite](#)

Rencontres IN2P3 autour du HTC/HPC et du Machine Learning

Le 22 février dernier s'est tenue une rencontre IN2P3 pour discuter de nos activités autour du (...)

[en savoir plus](#)

ENVOL 2016 - Nouvelles technologies et méthodes pour gérer le cycle de vie d'un logiciel

La cinquième édition des ANF ENVOL a eu lieu du 28 novembre au 2 décembre 2016 à La Rochelle. Cette (...)

[en savoir plus](#)

Webinaires du RI3

Le retour des webinaires dans le paysage de l'informatique de l'IN2P3 semble avoir été un succès (...)

[en savoir plus](#)

Agenda

Journée ARAMIS « Dev & Ops : les outils du dialogue », 13 avril 2017 - Saint Etienne

Le réseau ARAMIS organise sa 9ème journée annuelle le 13 avril 2017 à Saint-Etienne sur le thème : (...)

[en savoir plus](#)

CIVEMSA 2017, du 26 au 28 juin 2017 - Nancy

La conférence IEEE internationale "Computational Intelligence and Virtual Environments for (...)

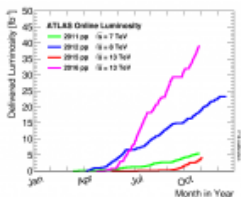
[en savoir plus](#)

ACAT, du 21 au 25 août 2017 - Seattle

Le 18ème workshop international ACAT 2017 (Advanced Computing and Analysis Techniques in Physics (...)

[en savoir plus](#)

LHC



Quand la performance du LHC dépasse les attentes...

...inévitablement les ressources informatiques nécessaires dépassent un peu les prévisions. L'environnement Le Run-2 du LHC, qui voit l'énergie disponible passer de 8 TeV au Run-1 à 13 TeV, s'étend sur trois ans, de 2015 à 2018. Après un démarrage difficile en 2015, dû à l'optimisation délicate de l'accélérateur à une telle énergie et luminosité, l'année 2016 a vu un fonctionnement en tous points exceptionnel, comme le montre la figure ci-dessus. Par rapports aux précédentes estimations, le LHC a fourni fin 2016, et donc pour 2017 et 2018, environ 50% de plus de données que prévues initialement (7.8M secondes de collision pp en 2016, contre 5.5M attendues). Cette situation, excellente pour les découvertes potentielles au Run-2, a un impact important sur les ressources : plus d'événements à traiter et conserver (données et

Tendance



"Blockchain" : Décentraliser l'Internet

A sa conception, Internet avait été imaginé comme un grand magma décentralisé de services informatiques communicants dont les protocoles de routage promettaient d'éviter les situations d'engorgement et de rupture de flux. Aujourd'hui, certains acteurs majeurs ont mis en place des services centralisés qui s'accaparent une grande partie de deux ressources essentielles de l'Internet : la bande passante et les utilisateurs. La « Blockchain » est la mise en œuvre d'un ensemble de technologies éprouvées (les certificats électroniques, les communications « pair à pair » ou « P2P »...) promettant de revenir aux bases de l'Internet : décentralisation des services informatiques et réappropriation des droits des utilisateurs. Une Blockchain est une structure de données qui circule sur un réseau pair à pair dans lequel les participants, évoluant hors de toute hiérarchie, assurent l'exécution de transactions et les enregistrent dans un « ledger » (registre en anglais) entièrement distribué : un « livre des comptes » inaltérable

Projet européen



Quand l'Astronomie et l'astrophysique européennes s'organisent pour relever les défis BigData de demain

ASTERICS (Astronomy ESFRI and Research Infrastructure Cluster) est une infrastructure de recherche financée par la Commission Européenne Horizon 2020 à hauteur de 15 M€. Le projet rassemble pour la première fois 22 instituts d'astronomie, d'astrophysique et de physique des particules afin de créer une synergie autour des défis auxquels les grandes expériences font face. Il est organisé en 4 work packages scientifiques chargés de la dissémination auprès du grand public, de la gestion du temps et des alertes temps réel, de l'interopérabilité des données et de leur accès, et enfin de la gestion des volumes croissants de données générés par les différents projets européens de recherche inscrits dans la feuille de route de l'ESFRI (le

Communauté



ANF ProDev : Les Problématiques du Développement Logiciel de la création à la diffusion

De nos jours, la plupart des projets scientifiques inclut des développements logiciels. La gestion de ceux-ci, tout au long de leur cycle de vie et dans un contexte de recherche, révèle des difficultés spécifiques qu'il faut anticiper. Alors que la démarche « qualité » a été adoptée dans de nombreuses disciplines techniques, elle l'est moins souvent dans le domaine informatique, peut-être trop récent et usant de technologies et méthodes évoluant souvent trop rapidement pour pouvoir s'assimiler par exemple aux méthodes de conception et production mécanique, rodées et consolidées depuis des siècles. De ce constat est né l'idée de co-organiser une formation « qualité logiciel » entre les réseaux DevLOG (CNRS/MI) et Qualité en Recherche (CNRS/MI), avec le concours du Pepi (INRA) et la Délégation Aquitaine (CNRS). Elle s'est déroulée du

simulation), donc plus de CPU et de stockage nécessaires. Cependant cet accroissement de ressources doit s'effectuer dans le cadre financier propre au calcul depuis le démarrage du LHC, le 'flat budget', qui permet grâce aux progrès de la technologie et (...)

[lire la suite](#)

apportant la preuve d'existence, d'identité et d'horodatage. (...)

[lire la suite](#)

Forum stratégique européen pour les infrastructures de recherches) comme CTA, SKA, KM3NeT et d'autres projets majeurs, notamment Euclid, EGO-VIRGO, E-ELT et LSST. En particulier, toutes les problématiques de (...)

[lire la suite](#)

7 au 10 novembre 2016 au Domaine de Françon à Biarritz et a attiré 44 participants dont 37 stagiaires et 7 intervenants avec une répartition (...)

[lire la suite](#)

CTD/WIT17, du 6 au 9 mars 2017 - LAL

Les événements Connecting The Dots et Intelligent Trackers 2017 sont colocalisés et organisés du 6 (...)

[en savoir plus](#)

HEPIX Spring 2017, du 24 au 28 avril 2017 - Budapest

Les inscriptions sont ouvertes jusqu'au 17 avril sur le site de l'événement.

[en savoir plus](#)

Journées LCG France, du 27 au 29 mars 2017 - CC-IN2P3 (Villeurbanne)

Plus d'informations.

[en savoir plus](#)

JDEV2017, du 4 au 7 juillet 2017 - Marseille

La 4ème édition des JDEVs se tiendra du 4 au 7 juillet au centre de Marseille. Ces journées (...)

[en savoir plus](#)

 **Offres d'emploi**

CDD : Ingénieur en informatique (H/F), CC-IN2P3 - Villeurbanne

[en savoir plus](#)

CDD : Ingénieur Systèmes / DevOps (H/F), CC-IN2P3 - Villeurbanne

[en savoir plus](#)

Post-Doc Data Science (H/F), CC-IN2P3 - Villeurbanne

[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.



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.



© 2017 CCIN2P3



n°36
Mars
2017

La lettre IN2P3 Informatique

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

A LA
UNE

Journée des expériences 2017

Communauté



La Journée des Expériences a eu lieu le 30 janvier au Centre de Calcul de l'IN2P3. Lors de cette réunion est invité l'ensemble des représentants des expériences utilisatrices des ressources du CC-IN2P3 pour échanger sur les services mis à disposition et leur utilisation.

Pierre-Etienne Macchi, directeur du CC-IN2P3, a présenté un bilan de l'année 2016, évoquant notamment la capacité du centre et l'utilisation des ressources. Les demandes des expériences pour 2017 seront couvertes. Volker Beckmann, Directeur Adjoint Scientifique Calcul et Données de l'IN2P3, a ensuite donné le point de vue de l'Institut sur l'utilisation du Centre de Calcul à travers les projets et groupes de travail « Calcul et données ». Il a également souligné l'importance du Data Management Plan (DMP) pour structurer les données et leur gestion. S'en est suivi un bilan des activités et des évolutions des services du Centre en 2016 qui a rappelé notamment la refonte des plate-formes interactive et parallèle, la création d'une ferme GPU, la mise à disposition du NoSQL via MongoDB et à partir de cette année l'envoi aux groupes d'utilisateurs d'un DMP à remplir. Les perspectives pour 2017 prévoient, entre autres, une optimisation de la configuration du système de *batch*, du travail sur l'exécution et l'orchestration des

conteneurs, la poursuite des travaux en vue de l'abandon d'AFS. Jean-Yves Nief, responsable de l'équipe Stockage, a d'ailleurs présenté, plus tard dans la matinée, les choix techniques pour le remplacement des différents espaces AFS d'ici à 2 ans. Autre sujet débattu lors d'une session interactive, l'évolution de la plate-forme en charge des rapports de consommations envoyés aux responsables des groupes d'utilisateurs. Cécile Evesque et Dominique Cathala-Martinez, de l'équipe Applications, ont sollicité un retour des gens présents sur les besoins qui sera complété par un sondage en ligne. Enfin, Osman Aïdel et Philippe Cheynet, de l'équipe Stockage, ont présenté les nouveautés du service Bases de données avec un focus sur MongoDB.

L'après-midi était réservé aux expériences. Ken Ganga, de l'APC, et Maurice Poncet, du CNES, ont présenté le projet Euclid, satellite qui cartographiera les galaxies, à la recherche des structures de l'univers et de la matière noire. Le modèle de calcul du consortium et la place du CC-IN2P3 ont été décrits. Maurice Poncet a aussi parlé de l'utilisation du Centre par Euclid et les moyens mis en place dans le cadre du développement du centre de traitement scientifique français, le SDC-FR (Scientific Data Centre). Chercheur au Laboratoire d'Innovation Thérapeutique à l'Université de Strasbourg, Florian Koensgen a donné un retour d'expérience sur l'utilisation de la nouvelle ferme de GPU du CC-IN2P3. Les perspectives de gain de temps sont intéressantes pour pouvoir étudier des systèmes plus complexes. Enfin, Klaus Werner, chercheur en physique théorique du Subatech à Nantes, a parlé du programme EPOS de génération d'événements pour les collisions proton-proton et d'ions lourds. Ce générateur lancé sur la ferme du CC est très consommateur en CPU et est utilisé par les expériences du LHC mais aussi pour les simulations de gerbes cosmiques par Auger.

Vous pouvez retrouver **les présentations de cette journée.**

David BOUVET (CC-IN2P3)



n°36
Mars
2017

La lettre IN2P3 Informatique

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



"Il va falloir opérer des choix et nous souhaitons le faire avec le maximum de paramètres et d'informations."

Christian Olivetto, Directeur Adjoint Technique à l'IN2P3



- Vous avez été nommé Directeur Adjoint Technique le 1er mars 2016. Pourriez-vous retracer brièvement votre parcours ?

Electronicien de formation, je suis rentré en 1983 au CNRS, au Laboratoire de Physique des Solides (INP). Trois ans plus tard, j'ai rejoint l'IPNO où j'ai travaillé durant 5 ans sur le projet de Cyclotron AGOR. J'ai ensuite réussi un concours interne affecté au CPPM où j'ai cette fois travaillé sur les liaisons fibres optiques analogiques pour ATLAS puis les développements de la distribution d'horloges du projet ANTARES, dont j'assurais aussi la responsabilité de la coordination technique. Puis j'ai travaillé à l'IRES (aujourd'hui IPHC) à Strasbourg en tant que Directeur Technique tout en maintenant une activité technique.

J'encourage d'ailleurs fortement tous les Directeurs Techniques et Chefs de Service à poursuivre une activité de développement, même si évidemment je suis conscient que ce n'est pas simple.

Durant six années à l'IPHC, je me suis notamment occupé de l'organisation de la production de CMS. J'ai également démarré les activités sur SPIRAL et développé les diagnostics faisceaux qui ont notamment fait l'objet de vente de licences. J'ai enfin rejoint l'APC pour les projets spatiaux en tant que Directeur Technique et ai travaillé sur les projets ATHENA et TARANIS, dont j'ai développé l'électronique de vol de l'instrument XGRE.

J'ai donc travaillé à la fois sur des projets en physique nucléaire, physique des particules, astrophysique et spatial, et ce dans plusieurs laboratoires. Ce parcours diversifié m'est aujourd'hui très utile pour mon rôle de Directeur Adjoint Technique de l'IN2P3.

- Quel est le périmètre de votre mission ?

Le rôle d'un Directeur Technique, que ce soit dans un laboratoire ou à l'IN2P3, est d'une part d'assurer la coordination des projets techniques et d'autre part, d'avoir une vision à moyen et long termes des choix stratégiques concernant ces activités.

Dans la nouvelle organisation de l'IN2P3, il était déjà important qu'il y ait un Directeur Adjoint Technique (DAT), cette fonction a aujourd'hui toute sa place. Nous formons une équipe collégiale avec les Directeurs Adjointes Scientifiques (DAS). Dès qu'il y a une notion technique dans un domaine (électronique, informatique, mécanique, etc.), je suis impliqué dans la réflexion et dans la prise de décision et le suivi des projets scientifiques incorporant des développements techniques.

Je précise que l'objectif n'est pas de créer un nouveau répertoire des métiers mais d'identifier les expertises spécifiques en matière d'informatique, comme pour les autres métiers du CNRS. Pour chaque expertise identifiée, on précisera combien de personnes sont concernées et on établira une pyramide des âges. Ensuite, l'objectif sera de définir les expertises qu'on doit maintenir, celles qu'on doit acquérir ou celles qui, suite à l'évolution des métiers, n'ont plus lieu d'être. C'est un sujet difficile mais aussi un challenge important.

- Qui sont vos interlocuteurs au-delà des Directeurs d'Unité pour la mise en œuvre de ces actions ?

C'est vrai, il y a d'abord les Directeurs d'Unité mais je complète cette diffusion d'information au réseau des RT. J'associe également les canaux des réseaux IN2P3, via les coordinateurs de ces structures avec lesquelles je fais des points réguliers. Pour le RI3, mon interlocutrice est par exemple Frédérique Chollet.

Enfin, il y a aussi tous les chefs de projet techniques au travers des revues et je me rends aussi régulièrement dans les laboratoires pour échanger de façon ouverte et franche avec les membres des services techniques.

D'une façon institutionnelle, je me sers d'ATRIUM : je rédige les documents et à chaque fois qu'un document est rédigé et validé, il est diffusé dans la zone publique / direction technique. Cela nous permet de diffuser l'information le plus largement possible. C'est le cas par exemple des lettres de mission des Directeurs Techniques/Responsables Techniques (**ATRIUM-136477**). On va aussi proposer des lettres de mission pour tous les chefs de service, notamment informatiques. Ces lettres de mission seront mises à disposition des Directeurs d'Unités qui pourront les utiliser en tant que telles ou les adapter à leur laboratoire. Notre objectif ici est de diffuser un message de coordination afin que tous les chefs de service aient une mission claire, identifiée et écrite.

Un autre document structurant porte sur le processus d'affectation des ressources humaines dans les projets, et est disponible là encore dans ATRIUM en zone publique (**ATRIUM-101283**). On y décrit ce que l'on pense être important lors de la mise en place de l'affectation des ressources humaines à un projet.

De façon générale, on dit les choses, on les écrit et on les publie de manière officielle pour que toutes les personnes aient accès à l'information (**ATRIUM Zone Publique DAT**). Il est important que ce soit fait ainsi pour que les informations soient bien transmises et bien interprétées.

Nous sommes dans l'optique de piloter et donc faire des choix. Il est important d'accompagner ces choix et les changements. Ceci se fait par l'échange, la diffusion des actions, dans le souci d'améliorer le processus si besoin. Mieux vaut avoir un retour négatif qu'aucun retour et qu'au bout deux ans, on s'aperçoive que les choses ont été mal interprétées.

- Qu'appelle-t-on un projet de R&D transverse ?

J'ai proposé que le budget Projet de l'institut permette de financer quelques projets de R&D technique, disons 4 ou 5 par an. On s'engage alors à ce qu'ils soient clairement considérés comme des projets IN2P3 et que les gens qui travaillent dessus soient officiellement affectés à ces projets. On garantit également que le budget soit pluriannuel (3 ans maximum).

La différence entre un projet de R&D « transverse » et un projet de R&D dit scientifique, c'est d'abord que le projet R&D transverse peut être porté par un ingénieur. Les critères

J'ai également un rôle important dans l'organisation projets mise en place : nous voulons piloter, c'est-à-dire coordonner et faire des choix. Coordonner sous-entend savoir ce qui se fait dans les laboratoires donc je me déplace, comme tous les DAS, afin de discuter avec les équipes et services techniques. Je m'appuie aussi beaucoup sur le réseau des Responsables Techniques (RT) afin d'obtenir le maximum d'informations, d'échanges avec eux et de les impliquer dans les études préalables. Tous les ans, j'essaie aussi de rencontrer les services techniques et de mieux connaître les expertises dans les laboratoires et leurs retours sur l'organisation et les actions mises en place.

- Pourriez-vous nous en dire plus sur la gestion de projets mise en place à l'IN2P3 ?

Ce qu'il faut comprendre c'est que le système qu'on a mis en place est le même pour tous les DAS et donc pour toutes les thématiques scientifiques.

La direction de l'IN2P3 a demandé à identifier les projets selon des critères précis, à savoir : un projet qui dure au moins 2 ans, qui est financé à hauteur de minimum 10 000 euros et qui implique 1 FTE au minimum.

En 2016, l'objectif était de mettre en place une cartographie des quelque 500 « lignes projets » de l'IN2P3 (**ATRIUM-152359 accès Dus**) et de faire en sorte qu'on ait tous la même vision, les mêmes critères dans toutes les thématiques scientifiques. Aujourd'hui, nous sommes arrivés à une organisation cohérente où l'on sait qui est responsable d'un projet, d'un master projet, etc. ce qui va nous permettre de faire des choix sur ces projets.

Nous réalisons par ailleurs des revues simplifiées, qui durent 3 à 4 heures, lorsqu'un DAS demande à avoir une vision synthétique d'un projet ou souhaite faire un point pour une étape clef. Concrètement, nous mettons en place une revue suivie d'une séance fermée durant laquelle on analyse le projet. Nous nous engageons ensuite à faire un retour officiel au groupe projets dans un maximum de deux semaines avec des recommandations. Clairement, on incite les gens à suivre les bonnes pratiques de la gestion de projets mais on essaye d'accompagner sans pour autant imposer un carcan quel qu'il soit. Du moment que le projet fonctionne, que les gens s'entendent bien, il n'y a pas de problème.

En ce qui concerne l'informatique, Volker Beckmann, le DAS Calcul et Données, essaye d'impulser tout ce qui concerne la recherche en informatique. Au niveau métier informatique cette fois, il est prévu d'ici le mois de juillet de réaliser une cartographie des expertises techniques. Un membre du RI3 sera associé à cette démarche et participera à un groupe de travail que je vais bientôt lancer.

d'acceptation de ces projets, hormis les critères énoncés plus haut et qui sont propres à tous les projets, c'est qu'ils doivent porter sur une thématique technique et impliquer plusieurs laboratoires. L'objectif est d'avoir des projets transverses, qui touchent le maximum de laboratoires et qui soient structurants pour une activité.

Le projet DAQGEN, par exemple, est porté par le LAL, le CPPM, le LPSC et le LPC Caen dont les ingénieurs travaillent sur le développement de cartes et de logiciels d'acquisition générique autour du standard micro TCA. L'idée est, qu'au bout de trois ans, on puisse fournir à l'ensemble des laboratoires IN2P3 la possibilité de faire ces cartes et ces logiciels associés en vue d'un système d'acquisition autour du standard Micro TCA. Les livrables sont clairs.

- Comment les projets sont-ils sélectionnés ?

A tout moment dans l'année, il est possible de remplir la demande de soutien de projet disponible sur **ATRIUM-110259** et flot de création nouveau projet : **ATRIUM-104488** précisant quels laboratoires et ingénieurs sont concernés et quel financement est demandé. J'en discute ensuite avec les DAS concernés afin de vérifier que le projet ne fasse pas doublon avec d'autres existants. Dans les deux mois, je m'engage à donner une réponse de principe puis l'acceptation est officiellement annoncée lors de la journée projets par le DAS incorporant ce projet dans son budget.

N'ayez pas peur de nous soumettre des projets, surtout s'ils sont structurants. Nous faisons tout pour piloter les choses et accompagner les changements. Mais lorsqu'on sent au niveau de la direction qu'on est cohérent, qu'on sait qu'un projet va concerner plusieurs thématiques alors il y a beaucoup plus de forces pour le soutenir.

- Un message à la communauté des informaticiens ?

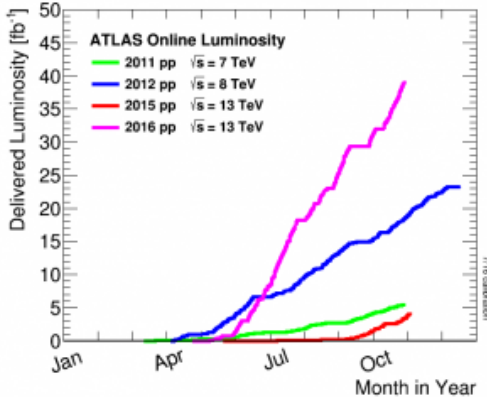
Que les années qui viennent continueront probablement à être difficiles en ce qui concerne les budgets et les ressources humaines. Il va falloir opérer des choix et nous souhaitons le faire avec le maximum de paramètres et d'informations.

Par ailleurs, je voudrais rappeler que les demandes de financements anciennement appelés « AP informatiques » passent dorénavant par DIALOG (équipements informatiques) et qu'il faut nous donner un maximum d'informations afin que nous puissions faire les bons arbitrages.

Propos recueillis par Gaëlle SHIFRIN



Quand la performance du LHC dépasse les attentes...



...inévitavelmente les ressources informatiques nécessaires dépassent un peu les prévisions.

L'environnement

Le Run-2 du LHC, qui voit l'énergie disponible passer de 8 TeV au Run-1 à 13 TeV, s'étend sur trois ans, de 2015 à 2018.

Après un démarrage difficile en 2015, dû à l'optimisation délicate de l'accélérateur à une telle énergie et luminosité, l'année 2016 a vu un fonctionnement en tous points exceptionnel, comme le montre la figure ci-dessus. Par rapports aux précédentes estimations, le LHC a fourni fin 2016, et donc pour 2017 et 2018, environ 50% de plus de données que prévues initialement (7.8M secondes de collision pp en 2016, contre 5.5M attendues).

Cette situation, excellente pour les découvertes potentielles au Run-2, a un impact important sur les ressources : plus d'événements à traiter et conserver (données et simulation), donc plus de CPU et de stockage nécessaires.

Cependant cet accroissement de ressources doit s'effectuer dans le cadre financier propre au calcul depuis le démarrage du LHC, le 'flat budget', qui permet grâce aux progrès de la technologie et la baisse des coûts associée, de bénéficier de plus d'équipements à budget constant.

Prenons maintenant l'exemple de l'expérience ATLAS pour montrer comment faire face au problème.

Les demandes

Par rapport aux demandes initiales pour 2017, qui étaient déjà très exigeantes en termes de ressources, ce surcroît de 50% en données du LHC se traduit par une augmentation de 20%, tant en disque qu'en CPU, et de seulement 10% en bande (grâce notamment au nettoyage permanent des fichiers inutilisés). Les 20% s'expliquent par le fait qu'une grande partie des ressources est dédiée à la simulation qui ne croît pas linéairement avec la quantité de vraies données.

Globalement, la croissance annuelle des demandes des 4 expériences LHC en stockage et CPU pour l'ensemble des sites Tier-1 et Tier-2 (voir les diagrammes ci-dessus), montre clairement un fort accroissement en 2017, moindre en 2018 (demandes qui ne sont cependant pas encore finalisées), et se situe au-delà de ce que permet le modèle à budget constant.

Les solutions envisagées

Les expériences LHC comme ATLAS ont déjà effectué un énorme effort d'optimisation tant pour ce qui est du temps de calcul que du stockage (simulation et reconstruction optimisées, placement de données dynamique, destruction systématique des vieux fichiers).

Deux solutions peuvent être envisagées qui consisteraient soit à réduire le nombre d'événements sauves sur disque, i.e. réduire le dernier niveau de déclenchement de 1kHz à 750 Hz, soit à stocker les données prises et les traiter et analyser avec retard, pendant le prochain long arrêt du LHC. Mais ces deux solutions auraient un impact très négatif sur le programme de physique, et les sujets de thèse accessibles aux étudiants notamment.

La seule solution viable envisagée serait de réduire la quantité d'événements simulés d'environ 20%, avec un impact très négatif sur la précision des résultats de physique produits.

Dans les faits

Fin 2016, grâce à l'effort énorme fourni par les agences de financement, en particulier l'IN2P3 (qui a augmenté le budget 2017 de LCG de 11%) et l'IRFU, le déficit en ressource pour ATLAS en 2017 est de 15% pour le CPU, 5% pour le disque et 5% pour les bandes.

Plus précisément, le choix a été de privilégier le stockage au détriment du CPU, afin de pouvoir stocker de façon optimale les données produites par toutes les sources possibles de CPU, y compris provenant des sites très modestes en taille, ainsi que des ressources de calcul opportunistes (hpc, nuage, calcul 'volontaire').

ATLAS poursuit en parallèle les efforts pour optimiser l'utilisation de ses ressources, comme utiliser encore plus le stockage sur bande, effectuer les campagnes de retraitement des données depuis les bandes, et réduire la quantité d'informations stockée pour chaque événement.

Le futur proche

Une fois encore, grâce aux efforts des instituts, et en particulier de l'IN2P3 et de l'IRFU, les expériences LHC comme ATLAS devraient pouvoir, malgré des ressources inférieures aux besoins, réaliser leur programme de physique en 2017.

L'année 2018 devrait voir un fonctionnement du LHC stable, ce qui peut laisser présager des demandes de ressources abordables, mais à évaluer avec soin.

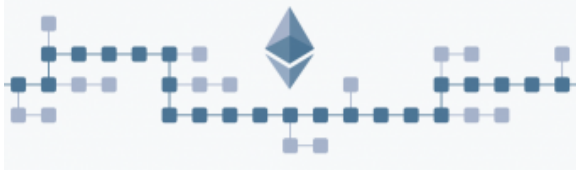
Le problème va devenir beaucoup plus critique pour le Run-3 du LHC en 2021, et encore plus pour la phase à haute luminosité HL-LHC. L'année 2017 a montré les limites du modèle à budget constant, qui doit certainement être revisité pour le futur du LHC.

Luc POGGIOLI (LAL)





"Blockchain" : Décentraliser l'Internet



A sa conception, Internet avait été imaginé comme un grand magma décentralisé de services informatiques communicants dont les protocoles de routage promettaient d'éviter les situations d'engorgement et de rupture de flux. Aujourd'hui, certains acteurs majeurs ont mis en place des services centralisés qui s'accaparent une grande partie de deux ressources essentielles de l'Internet : la bande passante et les utilisateurs. La « *Blockchain* » est la mise en œuvre d'un ensemble de technologies éprouvées (les certificats électroniques, les communications « pair à pair » ou « *P2P* »...) promettant de revenir aux bases de l'Internet : décentralisation des services informatiques et réappropriation des droits des utilisateurs.

Une Blockchain est une structure de données qui circule sur un réseau pair à pair dans lequel les participants, évoluant hors de toute hiérarchie, assurent l'exécution de transactions et les enregistrent dans un « *ledger* » (registre en anglais) entièrement distribué : un « *livre des comptes* » inaltérable apportant la preuve d'existence, d'identité et d'horodatage. Les participants valident ensemble les transactions par un protocole de consensus et les enregistrent de manière inaltérable dans le *ledger* dont ils possèdent chacun une copie. Le protocole est robuste : il s'accommode de pertes de connexion ou de participant, et il a su résister, jusqu'à présent, aux attaques malveillantes.

La Blockchain a été introduite par la crypto monnaie « *BitCoin* », mais son utilisation peut aller bien au-delà de la seule crypto monnaie : elle est un ensemble de données certifiées et pérennes qui circulent sur un réseau entièrement distribué et tolérant aux pannes. Il ne faut toutefois pas se contenter de voir

la Blockchain comme une base de données distribuée et sécurisée, car son concept révolutionnaire réside dans l'algorithme de consensus pour la validation des transactions.

Les middleware actuels de Blockchain (comme « *Ethereum1* ») permettent d'écrire des « *Smart Contract* » (littéralement contrat intelligent) avec le langage « *Solidity2* ». Ces contrats autonomes, auto exécutant et immuables ouvrent la voie à une nouvelle forme de relation de confiance, vers des horizons dont il est difficile de prendre la mesure à la date de cet article.

Il existe aujourd'hui un écosystème très actif autour de cette technologie. Afin de vous faire une idée, je vous invite à jeter un œil sur la page [Ethereum-dApps](#), celle de [Venture-Radar](#), ou encore celle du [projet « Blockchain Infrastructure for e-Science »](#).

De notre côté, nous pensons que cette technologie peut nous permettre de proposer un cloud entièrement distribué. Dans le cadre de l'ANR « Emergence », l'IN2P3 a participé de janvier 2013 à septembre 2015 au projet « CloudPower » d'Inria. L'objectif était de valoriser la technologie « *XtremWeb-HEP* », une plate-forme de calcul global développée en coopération par les deux instituts³. Entre autres résultats, nous avons pu prouver qu'il est possible d'utiliser la Blockchain pour mettre en œuvre un paradigme de cloud entièrement distribué capable de collecter les ressources informatiques sous-exploitées. Ainsi, avec les blockchains, nous sommes à même de proposer une solution alternative à l'utilisation des ressources centralisées par les acteurs dominant le marché du cloud, et tout un chacun est invité à entrer dans un grand « *market place* » pour acheter, vendre, louer des ressources IT, des applications et de l'espace de stockage. Notre plate-forme de test a eu un grand succès lors de la conférence SuperComputing 2016. Nous vous invitons à visiter [notre page](#) et restons à votre disposition pour toute question.

Oleg LODYGENSKY (LAL)



n°36
Mars
2017

La lettre IN2P3 Informatique

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



Quand l'Astronomie et l'astrophysique européennes s'organisent pour relever les défis BigData de demain



ASTERICS (Astronomy ESFRI and Research Infrastructure Cluster) est une infrastructure de recherche financée par la Commission Européenne Horizon 2020 à hauteur de 15 M€. Le projet rassemble pour la première fois 22 instituts d'astronomie, d'astrophysique et de physique des particules afin de créer une synergie autour des défis auxquels les grandes expériences font face. Il est organisé en 4 *work packages* scientifiques chargés de la dissémination auprès du grand public, de la gestion du temps et des alertes temps réel, de l'interopérabilité des données et de leur accès, et enfin de la gestion des volumes croissants de données générés par les différents projets européens de recherche inscrits dans la feuille de route de l'**ESFRI** (le Forum stratégique européen pour les infrastructures de recherches) comme CTA, SKA, KM3NeT et d'autres projets majeurs, notamment Euclid, EGO-VIRGO, E-ELT et LSST.

En particulier, toutes les problématiques de gestion des grands volumes de données sont gérées par le *work package* OBELICS (Observatory E-environments LINKed by common ChallengeS) dirigé par Giovanni Lamanna (LAPP). OBELICS cherche à encourager les développements conjoints et l'adoption de standards et de solutions communes pour le traitement, l'archivage, l'analyse et l'accès aux données des projets internationaux de l'ESFRI.

Le plus important sera sans doute de constituer un réseau de partenaires et un environnement de travail collaboratif. C'est dans cette optique que s'est tenu le **1er workshop d'OBELICS** à Rome en décembre 2016 autour du « *Science Data Cloud* » et des « *Computing models in Astronomy and Astroparticle Physics* ». Son organisation scientifique a été réalisée par le LAPP et son organisation locale par l'**INAF**, partenaire du projet. Le site de l'événement est disponible [ici](#).

Les sujets traités ont permis de rassembler les principaux représentants des projets scientifiques internationaux fédérés dans le consortium ASTERICS, des experts d'autres projets européens majeurs dans le domaine des infrastructures numériques (**Indigo-DataCloud**, **HNSCiCloud**, **AARC2**), de l'industrie (ATOS, NVIDIA, ORACLE Europe) et des consortia des institutions partenaires comme **EU-T0**, **APPEC** et **EGI.eu**.

Le workshop a initié la réflexion et le dialogue entre tous ces partenaires sur des problématiques d'intérêt commun. En particulier, une session dédiée à l'Authentification et l'Autorisation (A&A) des utilisateurs accédant aux données s'est organisée autour de **CTA** et **LOFAR** servant de cas d'études.

Des solutions faisant appel à différentes technologies pour l'Authentification (Shibboleth, certificats X.509, enID connect...), l'Autorisation (Grouper, GMS...) et l'intégration des organisations, communautés ou VO (VOSpace, ObsCore) ainsi que des démonstrations et tutoriels ont été présentés. L'enjeu est clairement de déterminer comment harmoniser les mécanismes A&A même si de grandes lignes comme l'authentification fédérée (type EduGain) font déjà consensus. Le problème du stockage et du transfert de données a été étudié à la lumière de l'expérience **SKA** qui produira 1 Po de données par jour. Des solutions de très grandes bases de données performantes, comme **Qserv** développée pour le **LSST**, ont été présentées, ici encore sous forme de démonstrations. Des systèmes de gestion des processus distribués efficaces et fiables seront également incontournables pour les grandes expériences comme Euclid qui devra traiter ses données sur 8 *data centers* (dont le CC-IN2P3) ou CTA dont le modèle de calcul se base sur un nombre restreint de *data centers* associés en un seul « centre virtuel » ; entre d'autres solutions, une démonstration du service **DIRAC** appliqué d'ailleurs avec succès dans le cadre de la production des simulations de Monte Carlo de CTA, a été présentée. Enfin, la dernière session fut dédiée à l'interopérabilité entre CPU et GPU afin d'utiliser ces ressources de façon optimale.

Un nouveau workshop aura lieu à l'automne 2017 et devra fortifier les liens et projets développés à la suite de cette première rencontre.

Autre événement d'OBELICS à venir : l'organisation d'une école est en cours. L'école « *Advanced software programming for astrophysics and astroparticle physics* » aura lieu au LAPP (Annecy-le-Vieux) du 6 au 9 juin 2017 et s'adressera aux doctorants et chercheurs désirant acquérir des techniques de programmation avancées et efficaces afin d'optimiser leurs codes en Python. Les sessions mêleront cours théoriques et travaux pratiques autour de trois grandes thématiques : « *Efficient code design, Parallel and GPU programming and Scientific libraries* ». Pour en savoir plus sur le programme et les inscriptions, rendez-vous sur [le site](#).



Thomas VUILLAUME (LAPP)



n°36
Mars
2017

La lettre IN2P3 Informatique

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



ANF ProDev : Les Problématiques du Développement Logiciel de la création à la diffusion



Groupe-ProDev2016-Biarritz

De nos jours, la plupart des projets scientifiques inclut des développements logiciels. La gestion de ceux-ci, tout au long de leur cycle de vie et dans un contexte de recherche, révèle des difficultés spécifiques qu'il faut anticiper. Alors que la démarche « qualité » a été adoptée dans de nombreuses disciplines techniques, elle l'est moins souvent dans le domaine informatique, peut-être trop récent et usant de technologies et méthodes évoluant souvent trop rapidement pour pouvoir s'assimiler par exemple aux méthodes de conception et production mécanique, rodées et consolidées depuis des siècles. De ce constat est né l'idée de co-organiser une formation « qualité logiciel » entre les réseaux DevLOG (CNRS/MI) et Qualité en Recherche (CNRS/MI), avec le concours du Pepi (INRA) et la Délégation Aquitaine (CNRS). Elle s'est déroulée du 7 au 10 novembre 2016 au Domaine de Françon à Biarritz et a attiré 44 participants dont 37 stagiaires et 7 intervenants avec une répartition variée par organisme (31 CNRS, 6 INRA, 3 Universités et 2 Renater, 1 Inria, 1 Ecole Polytechnique) et par compétences, voire métiers, également très diverses : développeurs, chargés de valorisation, chefs de projets et qualitiens.

Cette formation a présenté les difficultés rencontrées en faisant appel à un panel d'outils et de méthodes pour gérer un projet de développement logiciel selon une approche systématique intégrant les aspects de valorisation, capitalisation et génie logiciel du projet. Le fil conducteur de mise en application de la méthodologie proposée était la construction d'un site Web, retraçant le déroulement du projet en suivant les préceptes présentés.

Divisés en trois groupes pour trois livrables distincts du site Web, les stagiaires jouaient chacun un rôle particulier (développeur, acteur de la valorisation, ...) pour mener à bien le projet. Le rôle de chaque participant étant un rôle à contre-emploi, les développeurs ont joué les chargés de valorisation et vice-versa. Dans chaque groupe respectivement, trois domaines

ont été abordés : « Génie Logiciel », « Protection des logiciels » et « Diffusion des logiciels » sous l'angle de la qualité. Les participants se sont parfaitement prêtés au jeu et ont pu être confrontés ainsi aux difficultés spécifiques de leurs collègues. Ainsi on peut espérer qu'au retour dans les laboratoires, les interactions de chaque participant avec ses interlocuteurs locaux (Développeurs, Chefs de projets, Chargés de valorisation, Juristes, Acteurs de médiation) du développement logiciel seront facilitées suite à cette sensibilisation.

Selon l'avis des participants, cette formation originale a su marier tous les aspects d'un projet de développement logiciel : aspects techniques et organisationnels (amélioration continue, qualité, gestion de projet et d'équipe), mais aussi des aspects liés à la communication et la propriété intellectuelle (valorisation et diffusion de nos produits logiciels). L'agenda équilibré en cours et TP a également permis aux développeurs loin du développement Web de mettre en œuvre des techniques et pratiques nouvelles pour eux, comme les moteurs et cadriciels Web ou le CSS. Le projet collaboratif avait des objectifs jugés plutôt ambitieux par les participants, au vu des attentes des organisateurs et du temps imparti pour la réalisation. L'organisation et le partage des tâches au sein des équipes constituées pour l'occasion sont certainement en partie à l'origine de ce constat. D'autre part, les stagiaires ont certainement surévalué les objectifs à atteindre dans un temps forcément limité. Avec le recul, certains de ces problèmes auraient été affrontés plus efficacement par des livraisons plus régulières (même à fonctionnalités réduites) des produits des groupes de travail. C'est bien là une approche à promouvoir au sein de sa propre unité/équipe, en veillant à échanger avec ses clients ou utilisateurs plus régulièrement (en forçant les rencontres si besoin), afin de leur garantir la bonne prise en compte de leurs besoins et leurs demandes. C'est du moins la leçon positive qu'on peut tirer de cette formation. La « maquette de projet » vécue sans véritable enjeu dans la formation, avec toutes ses imperfections et irréalismes, servira à tout un chacun d'éviter ce qui s'est révélé comme problème dans la vie professionnelle, moyennant les outils et astuces découverts pendant cette petite semaine au pays basque.

Encore bravo et merci aux organisateurs, aux intervenants et non moins collègues stagiaires.

Thierry CHAVENTRE (LPC Caen, participant), Dirk HOFFMANN (CPPM Marseille, participant), Jean-Christophe SOUPLLET (CRMBM Marseille, co-organisateur), Henri VALEINS (CRMSB Bordeaux, co-organisateur et animateur du réseau « Qualité en Recherche »)



n°36
Mars
2017

La lettre IN2P3 Informatique

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



Journée ARAMIS « Dev & Ops : les outils du dialogue », 13 avril 2017 - Saint Etienne

Le **réseau ARAMIS** organise sa 9ème journée annuelle le 13 avril 2017 à Saint-Etienne sur le thème : « Dev & Ops : les outils du dialogue ».

Le mouvement Devops, apparu il y a maintenant quelques années, plébiscite l'application de méthodes de développement agiles et itératives à l'ensemble d'une organisation, mêlant ainsi développeurs (les Devs) et les administrateurs systèmes et réseau (les Ops).

Les Dev souhaitent des cycles de développement et déploiements courts afin de répondre rapidement aux besoins de l'organisation et des utilisateurs là où les Ops tiennent à la fiabilité et la stabilité de leurs infrastructures.

Comment répondre aux besoins des premiers sans mettre en péril les prérogatives des seconds ? DevOps plaide alors pour une culture, des process et des outils communs.

L'événement sera webcasté par la cellule webcast du CC-IN2P3.



CIVEMSA 2017, du 26 au 28 juin 2017 - Annecy

La conférence IEEE internationale "Computational Intelligence and Virtual Environments for Measurement Systems and Applications" (CIVEMSA 2017) a lieu du du 26 au 28 juin 2017 à Annecy.

Plus d'informations.



ACAT, du 21 au 25 août 2017 - Seattle

Le 18ème workshop international ACAT 2017 (Advanced Computing and Analysis Techniques in Physics Research) est organisé du 21 au 25 août 2017, à Seattle, University of Washington.

Plus d'informations.



CTD/WIT17, du 6 au 9 mars 2017 - LAL

Les événements Connecting The Dots et Intelligent Trackers 2017 sont colocalisés et organisés du 6 au 9 mars au LAL, Orsay.

Plus d'informations.



HEPiX Spring 2017, du 24 au 28 avril 2017 - Budapest

Les inscriptions sont ouvertes jusqu'au 17 avril sur *le site de l'événement*.



Journées LCG France, du 27 au 29 mars 2017 - CC-IN2P3 (Villeurbanne)

Plus d'informations.



JDEV2017, du 4 au 7 juillet 2017 - Marseille

La 4ème édition des JDEVs se tiendra du 4 au 7 juillet au centre de Marseille.

Ces journées s'intéressent à tous les aspects du développement logiciel au travers des thématiques suivantes :

- T1 - **Systèmes embarqués, les réseaux de capteurs et l'internet des objets**
- T2 - **Ingénierie et web des données**

- T3 - *Programmation de la matière, fabriques personnelles*
- T4 - *Usines logicielles et outils de production de code*
- T5 - *Infrastructures logicielles et science ouverte*
- T6 - *Méthodes et techniques de production de logiciel*
- T7 - *Science des données et apprentissage automatique*
- T8 - *Parallélisme itinérant, virtualisation et reproductibilité* et de manière transverse, le big data et la sécurité.

Les inscriptions avant le 15 mars bénéficient d'un tarif remisé.

Les agents des organismes partenaires (CNRS, INRA, INRIA, IRD, IRSTEA) peuvent bénéficier d'une place attribuée dans la limite des places disponibles.

Plus d'informations

Site d'inscription



n°36
Mars
2017

La lettre IN2P3 Informatique

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



CDD : Ingénieur en informatique (H/F), CC-IN2P3 - Villeurbanne

Contexte

Le Centre de Calcul de l'Institut national de physique nucléaire et de physique des particules (CC-IN2P3) est une unité de service et de recherche du CNRS. Classé parmi les grandes infrastructures françaises de recherche, il a pour vocation de fournir des moyens de calcul et de stockage de données aux chercheurs impliqués dans les expériences de physique corpusculaire. Plus de 4000 utilisateurs organisés en plus de 70 collaborations internationales utilisent ses services 24h/24, 7j/7.

Plus d'informations.

Missions

Le CC-IN2P3 renforce son équipe Opération, composée de 7 ingénieurs, par un ingénieur chargé de l'administration, de l'exploitation et de l'amélioration continue du service de calcul. L'Opération ayant un rôle central dans la supervision et l'optimisation de l'utilisation des ressources informatiques, le candidat participera aux diverses tâches communes de l'équipe. Le travail se déroulera dans un environnement multiculturel intellectuellement stimulant.

Activités

- Participer à l'administration et à l'exploitation du système de batch (Univa Grid Engine) dans un environnement Linux.
- Assurer l'amélioration continue du service de calcul.
- Participer aux activités communes de l'équipe :
 - Superviser l'utilisation des ressources informatiques.
 - Planifier et coordonner les maintenances trimestrielles.
 - Participer à la gestion d'événements et d'incidents.

Qualifications requises

- Bonne connaissance d'un OS Linux (plus particulièrement RHEL ou centos7),
- Maîtrise d'un langage de Scripting/programmation : Shell, python, perl,
- Qualités humaines telles que l'autonomie, la rigueur, le dynamisme, l'esprit d'équipe et l'initiative,
- Maîtriser l'anglais technique (écrit et oral),

Qualifications appréciées

- Connaissance d'un système de gestion de configurations (Puppet, ...),
- Connaissance d'un système de batch (Htcondor, Grid Engine, ...),
- Connaissance de langages de programmation orientés objet,
- Connaissance d'outils de monitoring et de supervision (Nagios, Elastic Search, ...)

Formation

Diplôme BAC +3 informatique universitaire ou école d'ingénieur en informatique

Conditions Contractuelles

Le poste proposé est à durée limitée d'un an renouvelable, à pourvoir dès que possible. La rémunération sera fonction du diplôme et de l'expérience conformément aux règles en vigueur au CNRS.

Dépôt des Candidatures

Envoyez votre candidature (CV et lettre de motivation) par courrier électronique à l'adresse Job77@cc.in2p3.fr.

Les candidatures sont acceptées jusqu'à l'attribution du poste. Les candidats retenus seront convoqués à un premier entretien dans les locaux du Centre de Calcul de l'IN2P3 à Villeurbanne.



CDD : Ingénieur Systèmes / DevOps (H/F), CC-IN2P3 - Villeurbanne

Formation : Diplôme en informatique, ou dans une discipline scientifique connexe

Type de contrat : Contrat à durée déterminée de 8 mois (fin du contrat le 30 septembre 2017)

Salaire : En fonction de l'expérience et de la qualification

Disponibilité : Poste à pourvoir immédiatement

Localisation : CC-IN2P3 – Campus de la DOUA – Villeurbanne

Des missions à l'étranger de durée variable sont à prévoir.

Fonctions

Le candidat participera au déploiement et à l'opération de la plateforme eTRIKS de calcul pour la recherche biomédicale et à l'intégration et déploiement d'applications dans le domaine biomédical. eTRIKS est un projet européen de 5 ans s'achevant le 30 septembre 2017.

Vous participerez avec d'autres ingénieurs à l'installation, configuration et opération de la plateforme eTRIKS. Vous prendrez part au cycle de vie de la plateforme, avec un accent particulier sur la préparation d'une migration possible de la plateforme vers un nouveau centre de calcul.

Le candidat retenu travaillera avec les membres du groupe eTRIKS et en collaboration avec les équipes locales ainsi qu'en coordination avec des développeurs d'application et des chercheurs dans le domaine biomédical.

Qualifications requises

Le poste à pourvoir requiert de fortes compétences techniques en informatique et technologie de l'information. Le candidat devra maîtriser tout ou partie des compétences dans les domaines listés ci-dessous :

- Vous avez de l'expérience dans l'administration et la maintenance de systèmes Linux.
 - Idéalement, vous avez déjà opéré sur des systèmes de type RedHat ou Debian (CentOS & Ubuntu)
- Vous avez l'expérience des technologies de type Cloud et Virtualisation
 - Une expérience avec KVM et Openstack serait un plus.
- Vous avez des connaissances en langages de script autre que bash et en développement d'applications.
 - Vous avez utilisé des systèmes de contrôle de versions (Git de préférence)
 - Vous êtes familier avec le déploiement et la maintenance de piles applicatives (LAMP, Java/Tomcat, NodeJS ou autres)
- Vous avez de l'expérience dans le déploiement de logiciels et les outils d'intégration de code, test et déploiement d'applications et de logiciels
- Vous avez des connaissances dans les outils de gestion de configurations
 - Avoir écrit quelques playbooks Ansible serait un plus.

Outre les compétences techniques, nous attendons du candidat :

- Une forte implication dans la résolution de problèmes et fourniture de solutions.
- Une bonne maîtrise du français et de l'anglais parlé et écrit,
- Un excellent relationnel, le sens de l'organisation et la faculté de travailler en collaboration avec des partenaires d'autres domaines.
- Des capacités d'adaptation, de disponibilité et un haut degré d'autonomie.

Candidature

Les candidats sont invités à soumettre leur CV ainsi que les lettres de motivation et de recommandation (ou des adresses mail de référents) à job78@cc.in2p3.fr.

Les candidatures seront considérées jusqu'à ce que le poste soit pourvu. Les sélections se feront au fur et à mesure de la réception des candidatures. Les candidats retenus seront convoqués à un premier entretien dans les locaux du site. L'entretien se tiendra en français et partiellement en anglais.

Lieu de travail

Le Centre de Calcul de l'Institut national de physique nucléaire et de physique des particules (CC-IN2P3) est une unité de service et de recherche du CNRS. Classé parmi les grandes infrastructures françaises de recherche, il a pour vocation de fournir des moyens de calcul et de stockage de données aux chercheurs impliqués dans les expériences de physique corpusculaire. Plus de 4000 utilisateurs organisés en plus de 70 collaborations internationales utilisent ses services 24h/24, 7j/7. **Plus d'informations.**

Le CC-IN2P3 emploie environ 80 personnes qui sont responsables du bon fonctionnement des infrastructures permettant la gestion et le traitement des données des expériences scientifiques. La ferme de calcul est composée de plus de 16500 cœurs. La capacité de stockage est de plus de 30 000 téraoctets et sera démultipliée dans les prochaines années.

Dans le cadre de sa ouverture à de nouvelles disciplines, ici les sciences biomédicales, le Centre participe au projet eTRIKS (<http://www.etriks.org>). Son but est de fournir une plateforme de gestion de connaissance intégrée qui propose un service en support à l'innovation à l'usage de la recherche translationnelle européenne. La plateforme eTRIKS est basée sur le logiciel « open source » transSMART qui propose les services requis de « curation », de standardisation et d'analyse des données.



Post-Doc Data Science (H/F), CC-IN2P3 - Villeurbanne

Le Centre de Calcul de l'Institut national de physique nucléaire et de physique des particules (CC-IN2P3) est une unité de service et de recherche du CNRS. Classé parmi les grandes infrastructures françaises de recherche, il a pour vocation de fournir des moyens de calcul et de stockage de données aux chercheurs impliqués dans les expériences de physique corpusculaire. Plus de 4000 utilisateurs organisés en plus de 70 collaborations internationales utilisent ses services 24h/24, 7j/7. **Plus d'informations.**

Fonctions

Depuis 2008, le CC-IN2P3 dispose d'une équipe de recherche en informatique dont les thématiques portent sur la simulation d'applications et de systèmes distribués, l'ordonnancement de workflows scientifiques et dans les systèmes de batch et la préparation à la mise en production de futures expériences par des activités liées à des aspects stockage, réseau et de pipelines applicatifs. Cette équipe a également pour mission de tisser des liens entre la communauté de recherche en informatique et les scientifiques utilisateurs du centre de calcul en développant de nouvelles façons de collaborer.

L'objectif pour le CC-IN2P3 est d'enrichir son équipe de recherche par l'apport de nouvelles compétences dans le domaine de la science des données. Pour le candidat, ce poste est l'occasion de développer une activité de recherche propre au contact proche d'expériences scientifiques produisant des grandes masses de données et à l'origine de découvertes scientifiques majeures issues de l'analyse de ces données (Boson de Higgs, ondes gravitationnelles, ...).

La principale mission confiée au candidat consistera à mettre à profit ses compétences pour aider les scientifiques à optimiser le traitement informatique de leurs données. L'interaction avec les scientifiques, utilisateurs des ressources informatiques du Centre de Calcul de l'IN2P3, sera donc un point clé de ce post-doc. En complément de l'activité de recherche propre menée par le candidat, les aspects liés à ces interactions permettront de contribuer à l'établissement d'une méthodologie, appelée *scientific-centered design*, adaptée au contexte particulier du CC-IN2P3 et de ses utilisateurs.

Qualifications requises

Formation

Titulaire d'un doctorat en informatique.

Compétences

La science des données est à la frontière entre la recherche scientifique, la recherche en informatique et les mathématiques et les statistiques. Les compétences recherchées couvrent donc un large spectre.

Le candidat aura des compétences dans les domaines ci-dessous :

- Expérience en bases de données SQL et no-SQL
- Maîtrise de langages de programmation tels que C/C++, Python, Java et R
- Expérience en analyse de données et en extraction de connaissance

Les compétences suivantes seront appréciées sans être indispensables pour la candidature :

- Connaissance de base sur les systèmes de stockage
- Maîtrise d'environnements de travail collaboratif (git, forge, ...)
- Bonnes compétences en rédaction d'article scientifique

Outre les compétences techniques, le candidat doit :

- Avoir de bonnes capacités d'écoute et de synthèse
- Faire preuve d'autonomie et de prise d'initiative
- Maîtriser l'Anglais comme langue de travail

Conditions contractuelles

Ce poste d'une durée de 24 mois est à pourvoir au **1er novembre 2016** et la rémunération sera de l'ordre de 2540€ brut/mois (2045€ net/mois).

Dépôt des candidatures

Envoyez votre candidature (CV et lettre de motivation) par courrier électronique (format PDF de préférence, MS Word ou texte également acceptés) à l'adresse **Job74@cc.in2p3.fr**.

Les candidatures sont acceptées jusqu'à l'attribution du poste. Les candidats retenus seront convoqués à un premier entretien dans les locaux du Centre de Calcul de l'IN2P3 à Villeurbanne.



n°36
Mars
2017

La lettre IN2P3 Informatique

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



Webinaires du RI3

Le retour des webinaires dans le paysage de l'informatique de l'IN2P3 semble avoir été un succès : six webinaires ont été présentés jusqu'à aujourd'hui (aussi bien des retours d'expériences que des retours de conférences), et ont à la fois suscité de l'intérêt (6 laboratoires connectés en moyenne, et jusqu'à 45 personnes au total en prenant en compte les connexions individuelles) et des questions (certaines sessions de discussion ont duré près d'une heure).

Le prochain webinaire sera donné le vendredi 17 mars par Fabrice Jammes du LPC Clermont-Ferrand sur le sujet "Analyse de petabytes dans LSST", sujet qu'il avait **déjà pu aborder lors des JI**, et qui avait suscité beaucoup d'intérêt parmi l'audience (par manque de temps, la discussion avait dû être interrompue).

Le webinaire suivant (qui devrait se tenir le 18 mai, suivez les **annonces de webinaires RI3 en ligne**) portera sur Blockchain et sera donné par Oleg Lodygensky du LAL. Nous cherchons encore un deuxième intervenant qui serait prêt à compléter, illustrer voire contraster les points de vue du premier orateur. Pour vous mettre l'eau à la bouche, lisez aussi l'article d'Oleg dans cette édition de la Lettre.

Si vous avez sujets de webinaires à proposer ou à soumettre (en tant qu'orateur) ou voulez simplement faire un vœu, n'hésitez pas à **utiliser le forum prévu à cette fin**.

Votre cellule webinaire du RI3.



ENVOL 2016 - Nouvelles technologies et méthodes pour gérer le cycle de vie d'un logiciel

La cinquième édition des ANF ENVOL a eu lieu du 28 novembre au 2 décembre 2016 à La Rochelle. Cette action, biennale depuis 2008, a fait interagir 35 participants pendant 8 demi-journées (et quelques soirées animées autour du développement et la découverte) sur quelques sujets des plus « chauds » dans le monde du développement logiciel, selon le thème **Nouvelles technologies et méthodes pour gérer le cycle de vie d'un logiciel**. L'objectif d'ENVOL, qui est non seulement d'informer sur des technologies mais également de faciliter les échanges entre les participants, a été largement rempli.

Voici quelques sujets abordés :

- Notre manière de diffuser des logiciels sera impactée par la **loi du 28-12-15**, traduisant une directive de l'EU ("Open Data") en droit français et la **loi du 07-10-16 « République numérique »**. Les décrets d'application sont attendus pour le premier trimestre 2017. Va-t-on vers l'obligation de mettre à disposition libre codes et données dès lors que nous sommes dans la fonction publique ?
- Le « langage » **UML** fait partie si ce n'est du quotidien de chaque développeur, au moins du monde informatique depuis bientôt 25 ans. Le **Consortium Papyrus** a développé des outils permettant implémenter et d'exécuter des diagrammes comportementaux (*behavioural UML*).
- Remis au goût du jour par les méthodes agiles, le développement piloté par les tests (*Test Driven Development*) a pu révéler toute sa pertinence.
- Lors de la maturation d'un logiciel le développeur rencontre souvent l'aspect manufacturier de la diffusion et du déploiement. L'outil inter-plateforme CMake, depuis 17 ans, est cependant toujours aussi pertinent en diffusion et déploiement grâce à une évolution constante, même si c'est au prix d'une compatibilité rétroactive non assurée.
- Les intervenants en Go et Git ont su familiariser leur public avec ces « nouvelles technologies » qui bénéficient chacune d'une communauté croissante d'utilisateurs. Go est devenu **"Language of the Year" selon l'index Tiobe**.
- Quoi de plus naturel que de parler de Docker quand on parle de Go ? Surtout que le cycle de vie se conclut, dans le cas d'un succès, avec une automatisation de la production et de la maintenance et l'assurance qualité à long terme. Un panel d'experts et responsables dans le domaine de l'intégration continue a permis de conclure cette formation par un panorama de retours d'expériences et de vues sur les différentes options qui s'offrent au développeur.
- Le programme était complété avec les petits extras favorisant l'échange entre participants et dont certains sont devenus gourmands : des soirées de coding en Go, découverte de l'Arduino avec une interface en python, voire en Go, TP prolongés de Papyrus/UML ainsi qu'un séminaire sur le **projet européen CityMobil** visant à automatiser certains transports urbains dans une dizaine de villes (dont La Rochelle).

Tous les supports des cours et TD sont disponibles sur le **site Indico des formations ENVOL**.



Rencontres IN2P3 autour du HTC/HPC et du Machine Learning

Le 22 février dernier s'est tenue une rencontre IN2P3 pour discuter de nos activités autour du calcul à haut débit (HTC), du calcul intensif (HPC), et du rapprochement entre ces deux mondes. Ce fut l'occasion de réaffirmer l'importance de ces thématiques dans le contexte technologique actuel, et de constater la diversité des activités menées dans les laboratoires de l'institut. **Les présentations et le compte-rendu de l'atelier sont accessibles.**

Au-delà des collaborations évidentes autour de la formation et de la veille, il s'avère encore difficile de dégager des projets communs avec des besoins et des objectifs précis. Quelques sujets ont cependant émergé du débat, tels que les langages dédiés (DSLs), le calcul intensif en conteneurs, ou la soumission de travaux sur des infrastructures Grille vs Supercalculateurs.

Cette rencontre aura notamment permis de réactiver le groupe de discussion RI3 "Programmation Parallèle et Performance du Calcul", avec la promesse de se revoir au moins une fois par semestre, et l'objectif de définir quelques projets communs d'ici l'automne. Si vous êtes intéressés, demandez à rejoindre le **projet Ri3Pppc sur le Gitlab de l'IN2P3**.

De la même façon que pour le HTC/HPC, nous allons discuter de Machine Learning dans un atelier le 17 mars. **Pour plus d'informations et pour s'inscrire.**