



n°6
Mai
2009

La lettre IN2P3 Informatique

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

**A LA
UNE**

Rhône-Alpes se dote d'une grille régionale



l'intergiciel EGEE mutualise plusieurs unités situées sur la région Rhône-Alpes, conformément aux objectifs définis par l'Institut des Grilles consistant à développer plusieurs grilles régionales sur lesquelles s'appuierait la grille nationale française.

Du nouveau du côté des grilles de calcul ! Alors que la grille européenne EGEE (Enabling Grids for E-Science) entame sa phase finale et que les pays européens se dotent de plus en plus de grilles nationales (National Grid Initiative), une petite dernière est née : la grille Rhône-Alpes.

La grille rhônalpine basée sur les ressources informatiques de plusieurs unités situées sur la région Rhône-Alpes, conformément aux objectifs définis par l'Institut des Grilles consistant à développer plusieurs grilles régionales sur lesquelles s'appuierait la grille nationale française.

La grille rhônalpine met potentiellement à disposition des chercheurs plus de 7000 cœurs (CPUs). Elle est actuellement constituée des nœuds suivants : le LAPP (Annecy), le LPSC (Grenoble), l'IPNL (Lyon), l'IBCP (Lyon) et le CC-IN2P3 (Lyon) qui est à l'initiative de ce projet. A terme, tous les organismes de recherche de la région Rhône-Alpes qui ont une infrastructure informatique à partager sont susceptibles de participer à cette grille, comme l'explique Yonny Cardenas, en charge du déploiement de la grille régionale au CC-IN2P3 : « La grille s'est naturellement constituée à partir des nœuds de la grille EGEE présents sur la région Rhône-Alpes. Mais l'objectif à terme est de rassembler tous les moyens de calcul des universités et des (...)

[lire la suite](#)

Interview

Wojciech Wojcik

« Le CC-IN2P3 est à l'origine du premier serveur web français ! »



Ancien ingénieur dédié au support aux utilisateurs au CC-IN2P3

[lire l'interview](#)



La préservation des données dans la physique des hautes énergies.

Au cours des dernières décennies, les expériences des hautes énergies, expériences sondant la structure de la matière à de très courtes distances, sont devenues de plus en plus complexes. La durée de vie d'une expérience récente, en incluant la conception, le développement, la construction, la prise de données et l'analyse, s'étend sur plus de 20 ans. Répéter de telles expériences pour, par exemple, améliorer la précision est devenu quasiment impossible. Une partie très importante des expériences actuelles, notamment celles associées aux collisionneurs, seront en fait les dernières de ce type. Les lots de données accumulées par ces expériences sont par conséquent uniques. Ces données sont utilisées pour obtenir des résultats de physique, qui seront par la suite publiés dans des revues scientifiques. Mais que deviennent les données brutes à la fin de la collaboration ? Est-ce que l'analyse de ces données reste toujours possible, ou souhaitable, après quelques années ? En effet, il est déjà arrivé que des améliorations dans la compréhension des (...)

[lire la suite](#)



Catherine Biscarat, physico-informaticienne au CC-IN2P3

Les informaticiens font partie d'une espèce à part entière, diverse et variée, c'est bien connu. Car derrière leur code, leur langage propre se cachent en réalité ceux qui développent, ceux qui administrent, ceux qui font de la production et même ceux qui tirent les câbles, etc. Et puis à l'IN2P3, il y a aussi ceux qui essaient de réconcilier deux communautés : les informaticiens d'un côté, les physiciens de l'autre. Cette nouvelle 'espèce' est apparue depuis quelques temps au Centre de Calcul de l'IN2P3 : nous l'appellerons le physico-informaticien. Au CC-IN2P3, il n'est pas rare de rencontrer un informaticien ancien physicien. Mais depuis plus d'un an, on peut maintenant y rencontrer des informaticiens qui sont toujours physiciens. Et là est la nouveauté. C'est le cas de Catherine Biscarat, physico-informaticienne donc, et membre de l'équipe user-support. Elle a bien voulu se prêter au jeu du portrait pour IN2P3 Informatique et nous éclairer sur ce nouveau métier. Née à Vichy en 1973, elle a suivi ses études en Auvergne, à Vichy d'abord puis à Clermont-Ferrand. Après un (...)

[lire la suite](#)



CHEP : Retour

Les conférences Computing in High Energy and Nuclear Physics (CHEP) sont organisées tous les 18 mois. La 17ème a eu lieu cette année à Prague, du 21 au 27 mars, dans le Centre des Congrès de Prague. Ce dernier a été construit par l'ancien régime communiste pour accueillir les congrès du Parti communiste, il se dresse sur la colline de Vysehrad et offre une belle vue sur Prague. La conférence a réuni 615 participants de 41 pays. Il y a eu 30 discours en plénière, près de 300 posters et plus de 200 présentations orales dans 7 sessions parallèles : Collaborative Tools Distributed Processing and Analysis Event Processing Grid Middleware and Networking Technologies Hardware and Computing Fabrics Online Computing Software Components, Tools and Databases Réunion WLCG La conférence a été précédée par la réunion WLCG avec 240 participants. Sans surprise, le thème principal de cette conférence a été la préparation pour les expériences LHC. Une plus petite partie a concerné les expériences qui ont déjà recueilli des données. La plupart des présentations peuvent être caractérisées par plusieurs mots-clés : Distribution, Web Service, (...)

[lire la suite](#)



Un service de visioconférence CNRS-IN2P3-INSERM opéré par le CC-IN2P3

Le Centre de Calcul de l'IN2P3 est opérateur d'un service de visio-conférence offrant la possibilité d'organiser simultanément plusieurs conférences, chacune pouvant réunir plusieurs participants ou sites. La capacité en service de l'ensemble des ponts est de 120 connexions en vidéo, et 90 en audio, ce service permettant l'accès sur une même conférence aussi bien avec un équipement de visioconférence dédié qu'avec un simple téléphone. Historique Ce service est né au sein du CC IN2P3. En 1999, la DSI a alors souhaité la création d'un service ouvert à tout le CNRS ; compte tenu des coûts, nous avons recherché des partenaires pour mutualiser le projet. L'opération initiale concernant une infrastructure de niveau national, a donc été financée par le CNRS (DSI), le CC IN2P3 et l'INSERM (convention avec le CC IN2P3), puis le CERN. Le MCU actuel a été intégralement financé par le CNRS. Usages La configuration des conférences, fréquemment inter-établissements, a largement contribué à la diffusion du service et à la reconnaissance de sa qualité. Nous (...)

[lire la suite](#)

Le CC-IN2P3 inaugure son nouveau calculateur haute performance basé sur une technologie 'verte'

14 mai 2009 - Lyon Depuis mars 2009, le Centre de Calcul de l'IN2P3/CNRS (CC-IN2P3), dispose (...)

[en savoir plus](#)

Agenda

"Les processeurs multicœurs - Le dessous des cartes"

19 mai 2009 - INRIA Rennes - Bretagne Atlantique Le centre Inria Rennes - Bretagne Atlantique (...)

[en savoir plus](#)

Emerging grid middleware standards

8 au 19 juin 2009 - Saint-Lambert-des-Bois L'édition 2009 des écoles d'été d'informatique (...)

[en savoir plus](#)

Josy "Réunion à distance (visioconf, web, chat)"

26 mai 2009 - Paris Cette journée est destinée à donner quelques clés aux administrateurs système (...)

[en savoir plus](#)

Archives

Abonnement

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

Proposer un article

Vous souhaitez proposer un article ? Envoyez un mail à IN2P3informatique@in2p3.fr.



© 2009 CCIN2P3



Equipe

Responsables éditoriaux : Dominique Boutigny et Cristinel Diaconu

Comité de rédaction : Virginie Dutruel, Sébastien Grégoire, Eric Legay et Gaëlle Shifrin



La lettre IN2P3 Informatique

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



Rhône-Alpes se dote d'une grille régionale



Du nouveau du côté des grilles de calcul ! Alors que la grille européenne EGEE (Enabling Grids for E-Science) entame sa phase finale et que les pays européens se dotent de plus en plus de grilles nationales (National Grid Initiative), une petite dernière est née : la grille Rhône-Alpes.

La grille rhônalpine basée sur l'intergiciel EGEE mutualise les ressources informatiques de plusieurs unités situées sur la région Rhône-Alpes, conformément aux objectifs définis par l'Institut des Grilles consistant à développer plusieurs grilles régionales sur lesquelles s'appuierait la grille nationale française.

La grille rhônalpine met potentiellement à disposition des chercheurs plus de 7000 cœurs (CPUs). Elle est actuellement constituée des nœuds suivants : le LAPP (Annecy), le LPSC (Grenoble), l'IPNL (Lyon), l'IBCP (Lyon) et le CC-IN2P3 (Lyon) qui est à l'initiative de ce projet. A terme, tous les organismes de recherche de la région Rhône-Alpes qui ont une infrastructure informatique à partager sont susceptibles de participer à cette grille, comme l'explique Yonny Cardenas, en charge du déploiement de la grille régionale au CC-IN2P3 : « La grille s'est naturellement constituée à partir des nœuds de la grille EGEE présents sur la région Rhône-Alpes. Mais l'objectif à terme est de rassembler tous les moyens de calcul des universités et des laboratoires de la région ayant des ressources à mettre à disposition. On a commencé à faire tourner quelques applications pilotes sur la grille, mais l'idée est d'ensuite attirer de nouvelles communautés, parmi celles qui ont un important besoin de calcul ou de stockage de données. »

Avec cette nouvelle grille, le Centre de Calcul de l'IN2P3 confirme donc l'ouverture de ses ressources à d'autres disciplines scientifiques puisque les chercheurs de tous les domaines scientifiques sont susceptibles d'y faire tourner leurs calculs. Les premiers utilisateurs de la grille rhônalpine ne sont d'ailleurs pas physiciens mais chercheurs au Laboratoire de Biométrie et Biologie Evolutive (

LBBE).

Les recherches du LBBE conduisent les scientifiques à maintenir de nombreuses bases de données dédiées à la phylogénie [1] et à l'évolution moléculaire. La construction et la mise à jour de ces bases reposent sur la comparaison entre elles de centaines de milliers de séquences biologiques (voire des millions pour certaines bases) ainsi que sur le calcul de centaines de milliers d'alignements de séquences et d'arbres phylogénétiques.

Ces comparaisons des séquences nécessitent un temps de calcul très important. Le LBBE a utilisé la grille rhônalpine pour la construction de l'une de ces bases (HOVERGEN, base de familles de gènes de vertébrés) : ce calcul a été réalisé en une semaine, sans la grille il l'aurait été en deux ans environ. Il représente aussi un excellent cas d'utilisation de la grille régionale car le calcul et les données ont été distribués sur différents sites, grâce à la technologie glite. « Cette méthode de calcul est exceptionnelle pour ce type d'applications dans le domaine des sciences du vivant, explique Pascal Calvat, responsable de l'activité Ouverture au CC-IN2P3, exceptionnelle car au-delà de l'aspect technique, les mentalités freinent considérablement l'accès à ce genre d'outils, qui permettent pourtant d'accélérer considérablement les recherches. »

Autre point fort de cette grille : le CC-IN2P3 s'efforce de proposer un accompagnement personnalisé. Les nouveaux utilisateurs sont en effet accompagnés dans toutes les étapes de leurs calculs, de l'expression de leurs besoins jusqu'à la mise en production et aux résultats scientifiques. « Cet accompagnement est un véritable atout, selon Yonny Cardenas. Et en cela, la grille rhônalpine se démarque clairement des infrastructures existantes en promouvant l'accès réel aux grilles auprès de chercheurs de tous les domaines qui n'utilisaient pas jusque là la grille. »

Enfin, le CC-IN2P3 propose dans le cadre de cette grille un certain nombre de services liés à son expertise dans différents domaines : du stockage massif de données avec iRODS, jusqu'à l'interopérabilité des grilles avec JSAGA, en passant par un service de support aux utilisateurs confirmé. Services qui pourront encore s'étoffer lorsque le rapprochement avec les grilles de recherche (ex : CIMENT ou Grid'5000) se concrétisera.

Gaëlle Shiffrin

[1] La phylogénie est l'étude de la formation et de l'évolution des organismes vivants en vue d'établir leur parenté.



n°6
Mai
2009

La lettre IN2P3 Informatique

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



« Le CC-IN2P3 est à l'origine du premier serveur web français ! »

Wojciech Wojcik

Ancien ingénieur dédié au support aux utilisateurs au CC-IN2P3



Vous avez été le premier à introduire le web en France, il y a près de vingt ans. Pourriez-vous nous rappeler comment il a été créé ?

Le web est né au CERN par l'intermédiaire de Tim Berners-Lee et de son compère Robert Cailliau. A l'origine, l'idée était de faciliter l'échange d'informations entre physiciens, éparpillés déjà à l'époque sur plusieurs sites. En physique, les informations sont vite périmées. Il fallait donc avoir accès à la dernière version, le plus facilement et plus rapidement possible. En 1989, Tim a donc eu l'idée de créer un endroit où se trouverait un exemplaire-maître auquel on aurait accès, qu'on pourrait modifier et qui serait donc toujours à jour. Cette idée est fondamentale dans l'histoire du web, le fameux www.

Pourquoi l'avoir appelé www ?

L'anecdote veut que Tim se trouvait à la cafétéria du CERN avec Bernard Cailliau, et qu'il cherchait un nom pour son invention.

D'après lui, les noms grecs étaient passés de mode... donc 'en attendant', il s'est dit qu'il allait l'appeler le WWW (World Wide Web). Vingt ans plus tard, il s'appelle toujours comme ça !

A quoi servait le web à son origine ?

A échanger de façon rapide et conviviale des informations entre physiciens, comme des dessins, des graphes, des résultats d'analyses scientifiques, etc.

Et comment est-il arrivé en France ?

En tant qu'informaticien au CC-IN2P3, je travaillais évidemment avec le CERN et connaissais déjà Tim... Puis en septembre 1992, Tim Berners-Lee et Robert Cailliau ont participé à CHEP, organisé cette année-là à Annecy, conjointement par le CERN et l'IN2P3. J'étais d'ailleurs l'un des organisateurs. A cette occasion, ils nous ont fait une démonstration de leur invention sur une machine Next. On avait justement plusieurs Next stations et deux Next serveurs au CC-IN2P3... Quelques semaines plus tard, on installait le premier serveur web français, le fameux info.in2p3.fr, dont l'adresse est inspirée de celui du CERN, info.cern.ch. Le CC-IN2P3 est donc à l'origine du premier site web français, le 4e ou 5e mondial... Avec mon collègue Daniel Charnay, on a commencé par y déposer quelques informations sommaires comme l'annuaire de l'IN2P3. On avait environ une centaine de connexions par jour.

Comment expliquer l'ampleur qu'a pris le web par la suite ?

Je pense que le génie de Tim Berners-Lee tient d'abord au fait qu'il a su rassembler des technologies existantes (comme l'internet) pour en faire un outil qui répondait parfaitement aux besoins de la communauté des physiciens. D'autre part, il a eu l'intelligence de créer un standard suffisamment large pour qu'on puisse rajouter de nouveaux protocoles. C'était donc un outil ouvert au développement futur. Et comme tous les développements réalisés au sein de la communauté scientifique, il n'était soumis à aucune licence.

© Sophie Blais pour les photos

PROPOS RECUEILLIS PAR GAËLLE SHIFRIN



n°6
Mai
2009

La lettre IN2P3 Informatique

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



La préservation des données dans la physique des hautes énergies.



Au cours des dernières décennies, les expériences des hautes énergies, expériences sondant la structure de la matière à de très courtes distances, sont devenues de plus en plus complexes. La durée de vie d'une expérience récente, en incluant la conception, le développement, la construction, la prise de données et l'analyse, s'étend sur plus de 20 ans. Répéter de telles expériences pour, par exemple, améliorer la précision est devenu quasiment impossible. Une partie très importante des expériences actuelles, notamment celles associées aux collisionneurs, seront en fait les dernières de ce type. Les lots de données accumulées par ces expériences sont par conséquent uniques. Ces données sont utilisées pour obtenir des résultats de physique, qui seront par la suite publiés dans des revues scientifiques. Mais que deviennent les données brutes à la fin de la collaboration ? Est-ce que l'analyse de ces données reste toujours possible, ou souhaitable, après quelques années ?

En effet, il est déjà arrivé que des améliorations dans la compréhension des fondements théoriques, la simulation ou l'avènement de nouvelles techniques d'analyse permettent de mieux analyser les "anciennes" données brutes. Récemment, un exemple très convaincant a été produit avec une nouvelle analyse des données de l'expérience JADE. La ré-analyse des données de production de jets dans des collisions électron-positron dans un domaine qui est considéré maintenant comme "basses énergies" (10-20 GeV) a conduit à la mise en évidence de la variation du couplage de la force forte et à l'amélioration de la précision de la mesure de cette constante fondamentale. Mise à part la démonstration cinglante de ce que l'amélioration de l'analyse et des fondements théoriques peuvent apporter, un autre aspect apparaît clairement : les données à plus basse énergie fournissent une information précieuse, à ne pas sous-estimer. "Bigger" n'est pas "better" en absolu, mais plutôt "complementary". (Figure 1)

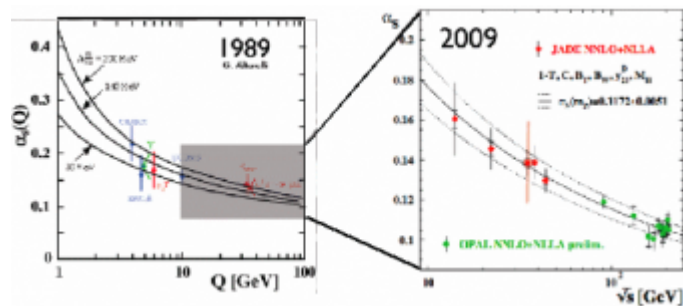
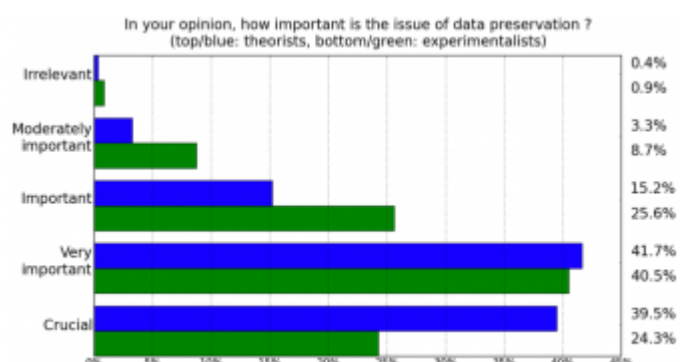


Figure 1

La ré-analyse des données de JADE de 1989, 20 ans après : La mesure originale indiquée dans le carré gris à gauche, est identifiée par une seule barre verticale rouge dans la figure de droite. Les mêmes données, ré-analysées après la mise à jour du software et l'implémentation des nouveaux développements théoriques, sont montrées par les points rouges dans la figure de droite. L'amélioration de la précision est évidente et conduit à l'observation de la variation de la constante de couplage fort (α_s), alors que la mesure originale ne permet pas une telle conclusion. Les données provenant de la ré-analyse sont combinées avec les données plus récentes du LEP pour obtenir une mesure très précise de la constante fondamentale α_s . (S.Bethke, MPI Munich)

Etant donnée la myriade de données accumulées dans la physique des particules, a-t-on la capacité de réanalyser ces données, en supposant que l'exemple de JADE pourrait s'appliquer à d'autres cas ? La réponse est effrayante : la plupart des données anciennes sont purement et simplement perdues. Effacées, anéanties ou, dans les meilleurs des cas, enfouies dans des caves sur des cassettes préhistoriques qui ne sont plus identifiables et/ou pour lesquelles la technique de lecture a été perdue. Cela devient encore plus impressionnant quand on réalise que des expériences relativement récentes (comme par exemple celles du LEP au CERN) n'ont pas de stratégie clairement identifiée quand à la sauvegarde des données. Et en effet, certains lots de données sont perdus, ou en train de devenir inaccessibles ou impossible à interpréter pour cause de perte d'expertise dans le software.

Pourtant, l'attente des physiciens est contraire à cette situation. Dans un sondage récent mené par le groupe PARSE (<http://www.parse-insight.eu>), plus de 90% des physiciens répondent que la préservation des données est importante, voir cruciale (Figure 2). Cette différence entre l'attente et la situation sur le terrain est le reflet de la non-inclusion, dans le programme scientifique des expériences, du problème de la préservation des données. La prise de conscience du rôle de la préservation n'apparaît qu'en urgence, à la fin de l'expérience, lorsque les ressources pour mettre en place un programme cohérent de préservation ne sont plus disponibles.



Si on veut avoir une chance de réussir dans cette entreprise, il convient en effet de définir précisément ce qu'on entend par préservation des données dans le domaine de la physique des hautes énergies. Il apparaît comme évident que la notion de "donnée" comprend plus que la simple suite de fichiers collectés par un détecteur. Au moins trois composantes essentielles doivent être prises en compte : la stratégie de stockage, le logiciel de lecture et d'analyse et l'ensemble des connaissances nécessaires à une interprétation correcte de ces données. Si pour le stockage et le logiciel le problème relève de l'organisation et de la cohérence du projet - problème toujours non résolu d'ailleurs - en ce qui concerne le "know how" le problème apparaît comme infiniment plus compliqué. Il y a en effet à conserver la documentation, mais aussi les méta-connaissances, c'est à dire cette partie de connaissance qui existe dans un ensemble de chercheurs sans en être pour autant formalisées ou documentées.

Il est donc grand temps de réfléchir au problème. Non seulement parce que le domaine de la physique des hautes énergies est en retard par rapport à d'autres domaines, comme la biologie ou l'astrophysique, où des centres de données existent depuis longtemps. Mais aussi parce que, pour plusieurs expériences cruciales, les analyses des données collectées auprès d'accélérateurs de particules uniques (Tevatron, PEP, HERA etc.) arrivent à terme.

Pour affiner la réflexion sur la préservation des données dans la physique des hautes énergies, un groupe d'études international a été mis en place fin 2008. Il est formé par des représentants des expériences de physique de particules : BaBar, Belle, BES-III, Cleo, CDF, DØ, H1 and ZEUS. Les centres de calcul des laboratoires associés SLAC (USA), KEK (Japan), IHEP (China), Fermilab (USA), DESY (Germany) et CERN sont également représentés. Un premier Workshop du groupe a eu lieu en janvier 2009 à DESY (Hambourg, Allemagne).

Les axes de réflexions ont été raffinés. Il s'agit tout d'abord d'identifier des cas de physique qui pourrait justifier une telle entreprise. Des exemples existent, mais il faut aller au delà et documenter une palette complète d'utilisations possibles des données préservées. Il faut ensuite proposer des modèles de préservation qui prennent en compte la collecte des informations nécessaires et leur structuration adéquate en fonction du but d'utilisation. La technologie est sans doute un facteur clé dans la préservation et des solutions techniques peuvent interférer très fortement et éventuellement définir la durée de vies des lots de données. Le modèle de migration des données doit être viable dans le contexte de l'évolution permanente du calcul : alors que le calcul sur grille semble à présent la tendance dictée par le démarrage du LHC, d'autres solutions peuvent se montrer également intéressantes, comme par exemple l'utilisation massive de la virtualisation dans le modèle de "cloud" qui offre la possibilité de s'affranchir ou au moins de réduire les difficultés liées au changement de hardware.

Il y a surtout l'organisation des expériences à très long terme, car préserver les données sans avoir une structure de supervision est vouée à l'échec, comme nous le montre les exemples du passé. C'est donc le problème de la gouvernance d'une collaboration à long terme qui doit être aussi réglé, en même temps que les aspects techniques. Au delà de la fin de vie d'une expérience, l'utilisation libre des données ("open access") n'a jamais été mise en pratique dans notre domaine, contrairement à d'autres domaines comme l'astrophysique, où les données rejoignent très vite le domaine public. Une telle révolution dans la physique des particules ne semble pas facile à mettre en pratique, c'est pourquoi une forme de supervision adéquate, incluant un transfert d'expertise et une capacité de vérification doivent être mises en œuvre.

Le travail de réflexion ne fait que commencer. En effet un deuxième Workshop est programmé à la fin mai au SLAC. Il semble plausible que cette réflexion doit se poursuivre et que l'activité de préservation de données doit prendre une place officielle et pérenne dans les activités des centres de calcul des grands laboratoires. Cela implique aussi, en l'occurrence, des ressources supplémentaires, et plusieurs laboratoires ont déjà inclus ce projet dans leur plan de R&D. Ceci conduira à un nouveau métier "HEP Data Custodian", à la frontière entre la physique et les technologies de l'information, qui participera à l'effort de préservation digitale des connaissances.

Pour plus d'information : <http://www.dphep.org>



Figure 3

Une cassette contenant des données de HERA, prises en 1992. Ces données continuent à être analysées à présent grâce à la migration sur d'autres supports de stockage.

Figure 2
Sondage effectué par le groupe PARSE sur un échantillon de
1500 chercheurs dans le domaine de la physique des hautes
énergies.

Cristinel Diaconu, CPPM



n°6
Mai
2009

La lettre IN2P3 Informatique

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



Catherine Biscarat, physico-informaticienne au CC-IN2P3



Les informaticiens font partie d'une espèce à part entière, diverse et variée, c'est bien connu. Car derrière leur code, leur langage propre se cachent en réalité ceux qui développent, ceux qui administrent, ceux qui font de la production et même ceux qui tirent les câbles, etc. Et puis à l'IN2P3, il y a aussi ceux qui essaient de réconcilier deux communautés : les informaticiens d'un côté, les physiciens de l'autre. Cette nouvelle 'espèce' est apparue depuis quelques temps au Centre de Calcul de l'IN2P3 : nous l'appellerons le physico-informaticien.

Au CC-IN2P3, il n'est pas rare de rencontrer un informaticien ancien physicien. Mais depuis plus d'un an, on peut maintenant y rencontrer des informaticiens qui sont toujours physiciens. Et là est la nouveauté. C'est le cas de Catherine Biscarat, physico-informaticienne donc, et membre de l'équipe user-support. Elle a bien voulu se prêter au jeu du portrait pour IN2P3 Informatique et nous éclairer sur ce nouveau métier.

Née à Vichy en 1973, elle a suivi ses études en Auvergne, à Vichy d'abord puis à Clermont-Ferrand. Après un DEUG de mathématiques et physique, elle a poursuivi son cursus jusqu'à obtenir un DEA option Physique des particules et nucléaire. Des études universitaires adaptées, selon elle, 'à l'acquisition d'une certaine ouverture d'esprit', peut-être déjà signe de son métier actuel. C'est au cours d'un module d'initiation aux techniques nucléaires qu'elle commence à s'immerger dans le monde du boson de Higgs, du CERN, etc., notions qui lui sont aujourd'hui tout à fait familières. S'ensuit alors l'obtention d'une bourse de thèse pour travailler sur ATLAS, l'un des détecteurs du LHC (« calibration du détecteur avec des données simulées [1] »), thèse qu'elle soutient en 2001 ; elle travaille ensuite au CERN pour l'Argonne National Laboratory, puis à Barcelone toujours sur ATLAS (« sur la recherche d'un boson de Higgs »).

En janvier 2003, ses recherches s'orientent vers le détecteur D0, installé auprès du Tévatron (situé à Fermilab, près de Chicago) où elle effectue un post-doc pour Lancaster, une université anglaise. Elle travaille à la préparation des échantillons de données simulées en mettant en œuvre son expertise sur l'implémentation des générateurs dans l'infrastructure software des expériences. Elle s'assure aussi de la distribution des données simulées aux utilisateurs après avoir contrôlé leur qualité. En parallèle, Catherine travaille sur les données de collision où elle contribue notamment à améliorer l'identification

des leptons taus dans le détecteur. En 2006, elle effectue un nouveau post-doc d'un an à l'Institut de Physique Nucléaire de Lyon, où elle met à profit ses connaissances sur les taus pour effectuer une recherche des quarks [2], dont elle publiera d'ailleurs très prochainement les résultats.

Son arrivée au CC-IN2P3 date d'octobre 2007 : la décision avait alors été prise de constituer un groupe de personnes travaillant à 50% sur l'informatique, à 50% sur la physique. Aux côtés de deux autres physiciens spécialisés dans les expériences CMS et LHCb, Catherine a aujourd'hui pour mission d'assurer le support du Tier-1 d'ATLAS et des utilisateurs d'ATLAS, mais aussi de poursuivre ses recherches sur ce détecteur. Selon elle, cette double mission lui permet de rester connectée à la communauté des physiciens, de comprendre spécifiquement leurs besoins, tout en essayant de trouver la solution la plus adéquate auprès des experts informaticiens. Une mission délicate s'il en est car, n'ayant pas suivi un cursus classique en informatique, Catherine avoue certaines lacunes qu'elle met un point d'honneur à combler. En pratique, et alors qu'elle est censée passer la moitié de son temps à effectuer des recherches en physique, Catherine a donc passé sa première année au CC-IN2P3 à s'immerger dans le monde du software d'ATLAS et de la grille : « C'est un monde nouveau ; il m'a fallu ce temps-là pour l'appréhender, explique-t-elle. Il me faut comprendre les choses une par une. On reçoit beaucoup de demandes de la part d'ATLAS et chaque problème rencontré me permet d'avancer. Mais pour cela, il faut faire preuve de curiosité et de débrouillardise. »

Une mission qui l'amène à travailler en étroite collaboration avec les autres équipes du CC-IN2P3, « les experts » selon elle, et une attitude qui la pousse à garder une certaine humilité devant eux. Malgré cela, la collaboration avec les informaticiens purs et durs ne se fait pas toujours sans heurts : « notre intégration a pris du temps car nos approches sont parfois différentes, mais nous arrivons de mieux en mieux à trouver une juste mesure et une bonne coopération se met en place pour travailler plus efficacement. »

Car même si son cœur penche plutôt vers la physique, pour Catherine, « le computing, c'est le nerf de la guerre ; sans lui, il n'y a pas d'analyse et pas de résultats de physique... ». Catherine se verrait bien entourée d'une véritable équipe de physiciens au CC-IN2P3 avec qui elle pourrait non seulement avancer dans la résolution de problèmes types mais aussi travailler sur des thématiques de physique. Cette vision est d'ailleurs partagée par la direction du CC-IN2P3 qui souhaite une imbrication beaucoup plus étroite entre la physique et le computing.

Gaëlle Shiffrin

[1] Les données simulées sont établies à partir des modèles théoriques et de la description des détecteurs. Elles permettent aux physiciens expérimentaux de confronter les données issues des détecteurs aux théories.

[2] Le squark est une particule supersymétrique, partenaire du quark.



n°6
Mai
2009

La lettre IN2P3 Informatique

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



CHEP : Retour



Les conférences Computing in High Energy and Nuclear Physics (CHEP) sont organisées tous les 18 mois. La 17ème a eu lieu cette année à Prague, du 21 au 27 mars, dans le Centre des Congrès de Prague. Ce dernier a été construit par l'ancien régime communiste pour accueillir les congrès du Parti communiste, il se dresse sur la colline de Vysehrad et offre une belle vue sur Prague.

La conférence a réuni 615 participants de 41 pays. Il y a eu 30 discours en plénière, près de 300 posters et plus de 200 présentations orales dans 7 sessions parallèles :

1. Collaborative Tools
2. Distributed Processing and Analysis
3. Event Processing
4. Grid Middleware and Networking Technologies
5. Hardware and Computing Fabrics
6. Online Computing
7. Software Components, Tools and Databases

Réunion WLCG

La conférence a été précédée par la réunion WLCG avec 240 participants.

Sans surprise, le thème principal de cette conférence a été la préparation pour les expériences LHC. Une plus petite partie a concerné les expériences qui ont déjà recueilli des données. La plupart des présentations peuvent être caractérisées par plusieurs mots-clés : Distribution, Web Service, Données et Performance.

► **Distribution** : Tout semble être maintenant à certains points liés à la Grille. Toutefois, le nouveau paradigme de l'informatique distribuée est arrivé - le Cloud. Il n'est pas encore clair, si le Cloud sera utile dans l'environnement HEP, mais l'intérêt est certainement présent.

► **Web Service** : Presque tous les services sont aujourd'hui présentés à un utilisateur sous la forme de Web Service. Les développeurs ont choisi un large spectre de technologies standard disponibles, le plus souvent avec l'aide de XML et Python. Certains projets inventent leur propre technologie pour mettre en œuvre des fonctionnalités spécifiques.

► **Données** : Toutes les données sont distribuées. Alors que

l'idée originale de la base de données fédérées orientées objet, est abandonnée, une architecture générale consiste en un réseau faiblement couplé de bases de données relationnelles et des réseaux de serveurs de fichiers. Les données sont répliquées. L'accès se fait par les proxies et les caches afin d'accélérer l'accès et de distribuer la charge. De nombreuses expériences parlent de données de Conditions (données décrivant l'expérience) et de Métadonnées (données qui décrivent d'autres données). L'idée de l'indépendance sur la technologie de la base de données relationnelle semble être abandonnée. Tout le monde converge vers une base de données Oracle au centre et SQLite pour la distribution des données. XRootd est généralement utilisé pour fournir un accès aux fichiers.

► **Performance** : En plus de l'optimisation classique du code et des algorithmes, de nouvelles méthodes sont évaluées afin d'augmenter la performance disponible. Il existe de nombreux projets sur la parallélisation et l'utilisation des architectures multi-cœurs. Un autre sujet chaud est la virtualisation, qui permet de lancer une expérience framework dans une machine virtuelle hébergée dans n'importe quel système d'exploitation.

Retour sur le CHEP'09

La plupart des projets sont fondés sur Root, LCG ou le framework Gaudi, le plus souvent sur l'ensemble d'entre eux. Le choix du langage semble être C++ avec un intérêt croissant pour Python et une utilisation constante de XML. Les lacunes du C++ ont toutefois été mises en avant dans plusieurs présentations, où les auteurs tentent en effet de créer une variante utilisable de C++.

Une évolution positive est le fait qu'il y ait moins de Status Reports. Plus de temps a été consacré à des problèmes réels et de vraies solutions.

Il n'y avait pas de programme scientifique le mercredi après-midi et les participants ont pu s'essayer à l'un des voyages thématiques proposés. Le banquet de clôture a eu lieu dans le très beau palais Zofin.

Tous les présentations sont disponibles à l'adresse <http://www.particle.cz/conferences/...>, les présentations des séances plénières sont également disponibles avec l'enregistrement vidéo. Les proceedings seront publiés en libre accès dans le Journal of Physics : Conference Series (CPM), publié par l'IOP Publishing. Les proceedings vont probablement contenir plus de 4000 pages (ils n'existent que sous forme numérique).

Julius Hrivnac



La lettre IN2P3 Informatique

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



Un service de visioconférence CNRS-IN2P3-INSERM opéré par le CC-IN2P3



Le Centre de Calcul de l'IN2P3 est opérateur d'un service de visio-conférence offrant la possibilité d'organiser simultanément plusieurs conférences, chacune pouvant réunir plusieurs participants ou sites. La capacité en service de l'ensemble des ponts est de 120 connexions en vidéo, et 90 en audio, ce service permettant l'accès sur une même conférence aussi bien avec un équipement de visioconférence dédié qu'avec un simple téléphone.

Historique

Ce service est né au sein du CC IN2P3. En 1999, la DSI a alors souhaité la création d'un service ouvert à tout le CNRS ; compte tenu des coûts, nous avons recherché des partenaires pour mutualiser le projet. L'opération initiale concernant une infrastructure de niveau national, a donc été financée par le CNRS (DSI), le CC IN2P3 et l'INSERM (convention avec le CC IN2P3), puis le CERN. Le MCU actuel a été intégralement financé par le CNRS.

Usages

La configuration des conférences, fréquemment inter-établissements, a largement contribué à la diffusion du service et à la reconnaissance de sa qualité. Nous avons fait le choix de laisser ce service ouvert à la communauté enseignement/recherche dans la mesure des disponibilités et compte tenu qu'aucun coût de connexion n'incombe à l'opérateur (réseau Internet ou appels à la charge de l'appelant pour la téléphonie). Le système est donc largement utilisé par les EPST et institutions françaises et étrangères qui se le sont approprié.

Technologie

L'originalité du système est de s'appuyer sur une technologie éprouvée et normalisée (h323) auquel est adjoint un système d'accès conçu pour les utilisateurs. Présentation administrative et technique du service

Le Centre de Calcul a développé une plate-forme Web de réservation (instantanée ou différée) qui filtre les autorisations d'accès, pilote les ponts de visioconférence en temps réel, émet des mails vers les utilisateurs et offre un certain nombre de services à valeur ajoutée (réservations de salles, recherches LDAP, espaces de partage de documents/stockage, etc.). » L'équipe projet en charge de ce service au Centre de Calcul IN2P3 est composée de 3 personnes (1 IR, 1 IE 1 T), qui ne travaillent pas à temps plein sur ce service.

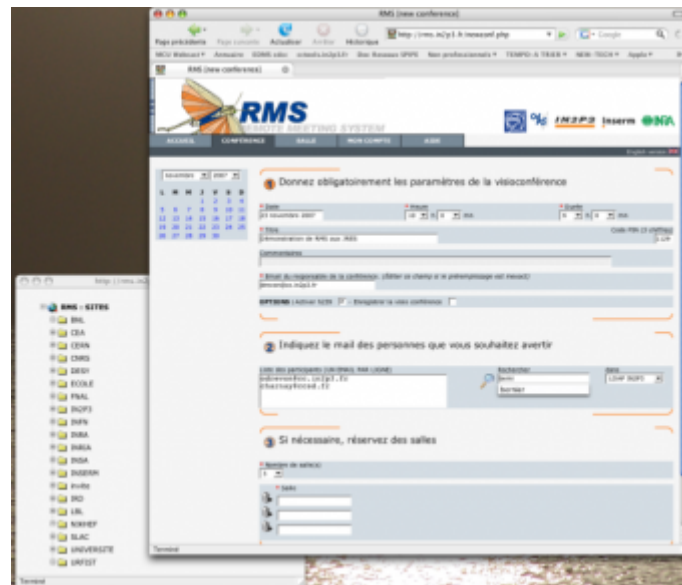
Le logiciel de réservation RMS (prononcer Hermès)

Accessible à l'adresse <http://rms.cnrs.fr>, également en version anglaise, il permet d'enregistrer une réservation de vidéoconférence sur le MCU, en gérant jusqu'à l'envoi d'invitation par mail, à chacun des participants souhaités et connus. Celle-ci contient toutes les informations nécessaires à la connexion du participant le jour dit.

Réservation ? En fait, pas vraiment !

Une des particularités importantes de RMS : le système de MCU (Codian) sur lequel il est basé est volontairement considéré comme un système dit « ad hoc ». Ce qui signifie notamment qu'il n'y a pas vraiment réservation de ressources sur le MCU, tant qu'elles ne sont pas réellement consommées. Le principe qui prévaut donc est « premier connecté, premier servi » et non pas « premier ayant réservé, premier servi » ! Nous garantissons par contre la disponibilité permanente d'un nombre suffisant de voies pour éviter les conflits et congestion, en anticipant raisonnablement la demande par une surveillance constante de la « consommation ». En échange, une souplesse extrême est possible, vous pourrez réserver ainsi dans la minute, une conférence multi-points inopinée via RMS. De même pour la durée, qui peut être rallongée à la demande par vous-même. Autre souplesse permise : le nombre de participants n'est pas figé, toute personne ayant connaissance des coordonnées de la conférence reçue par mail peut s'inviter,

d'où la présence d'un PIN Code d'accès, et aussi d'une indication sonore et visuelle à l'arrivée d'un nouveau participant. C'est surtout intéressant dans le cas de réunions périodiques à assistance variable.



Un exemple de réservation par l'interface web de RMS

Sur le mail figure également une URL de page web dite 'Compagnon', qui permet de suivre de façon passive en streaming une conférence, voire de 'chatter' avec les participants, et/ou partager des documents communs avec eux. Quelques chiffres de fréquentation et références pour terminer ce chapitre : Il est important de dire que la croissance est continue, à un rythme soutenu, depuis la mise en service du système actuel en Novembre 2005. En Avril 2009, nous servons environ 40 conférences quotidiennes, pour tous les organismes, et tous les jours ouvrés de la semaine, pour une moyenne de douze en Septembre 2007, de six en janvier 2007 et de quatre, fin 2006. Dans cet ensemble, la part des conférences purement téléphoniques est en forte augmentation également. Le Centre de Calcul IN2P3 héberge également le second NR de Renater, ce qui permet d'assurer une bande passante quasi « illimitée » au service de visioconférence, et d'envisager le passage à la HD sans souci de ce côté-là.

Pré-requis techniques

Les postes de visioconférence :

Outre un simple téléphone, les matériels possibles comme poste de visioconférence sont nombreux et doivent répondre à la norme de visioconférence sur IP, H323 (ou H320 pour accès par ISDN ce qui devient plus rare). On les divise en deux catégories principales :

* Les stations dédiées : les marques les plus répandues sont Polycom, Tandberg, Aethra, LifeSize. D'autres existent... Ce genre de matériel, en général très performant, est destiné à équiper des salles de réunion ou dédiées, pour les marques plus coûteuses. * Les logiciels fonctionnant sur poste de travail (Windows, Linux ou Mac OS X) ;

A cela il faut ajouter le coût d'achat d'une caméra vidéo (une webcam peut suffire !) et d'un micro(-casque c'est mieux !).

En résumé, nous avons souhaité vous présenter ici un service pleinement opérationnel, en exploitation de longue date, qui s'inscrit dans une logique très actuelle de réduction des coûts de fonctionnement, (et de réduction du temps et des nuisances dues aux transports), qui permet une plus grande efficacité pour les collaborations nationales et internationales, en apportant plus de souplesse et d'ouverture aux relations de votre laboratoire avec l'extérieur. Ceci au coût d'un investissement quasi-inexistant, ou minimal, qui peut dans ce cas, être amorti dans le semestre.

Pour tout contact, et renseignements complémentaires : visioconf@cc.in2p3.fr

Equipe Visioconf



La lettre IN2P3 Informatique

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



"Les processeurs multicoeurs - Le dessous des cartes"

19 mai 2009 - INRIA Rennes - Bretagne Atlantique

Le centre Inria Rennes - Bretagne Atlantique organise en partenariat avec le pôle de compétitivité Images & Réseaux et la technopôle Rennes Atalante, une journée thématique consacrée au portage des applications logicielles sur les "multicoeurs".

Plus d'infos à <http://www.irisa.fr/irisatech>.



Emerging grid middleware standards

8 au 19 juin 2009- Saint-Lambert-des-Bois

L'édition 2009 des écoles d'été d'informatique CEA-EDF-INRIA se déroulera au Centre Port-Royal à Saint Lambert des Bois, du 8 au 19 juin 2009, et portera sur les grilles informatiques.

Elles s'adressent à des chercheurs, des ingénieurs et des doctorants. Elles leur permettent de faire le point sur l'état d'avancement des sujets proposés et de confronter leur expérience. L'enseignement est dispensé en français ou en anglais. Il est complété par des travaux dirigés, en petits groupes, animés par des assistants.

À travers des cours magistraux, des travaux pratiques s'appuyant sur les infrastructures grid5000, EGEE et DEISA, et une sélection de conférences, cette école sera l'occasion pour les participants de

- ▶ comprendre les fondements des grilles informatiques,
- ▶ se familiariser avec les différents fonctionnements des grilles informatiques,
- ▶ être conscients des défis et perspectives scientifiques des grilles informatiques,
- ▶ être en mesure d'utiliser un environnement grille et de porter ses applications sur une infrastructure de grille.

Vous trouverez les informations préliminaires ainsi que l'inscription en ligne sur le site : <http://www.inria.fr/actualites/coll...> avec un prix attractif pour les doctorants (nombre de places limité)



Josy "Réunion a distance (visioconf, web, chat)"

26 mai 2009 - Paris

Cette journée est destinée à donner quelques clés aux administrateurs système et réseau sur les infrastructures et le support utilisateur impliqués par le déploiement grandissant de cette pratique.

Date : le mardi 26 mai 2009 de 9h30 à 17h15

Lieu : Paris, amphi de l'institut du monde anglophone, 5 rue de l'Ecole de Médecine, 75006 Paris

Pour s'inscrire à cette journées JoSY : envoyer un mail à josy.inscription@grenoble.cnrs.fr.

Vous pourrez également suivre cette journée en direct à l'adresse <http://webcast.in2p3.fr/JoSy/RAD>.

Pour plus d'informations sur l'événement, consultez le site à l'adresse <http://www.resinfo.cnrs.fr/spip.php...>



La lettre IN2P3 Informatique

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



Le CC-IN2P3 inaugure son nouveau calculateur haute performance basé sur une technologie 'verte'

14 mai 2009 - Lyon

Depuis mars 2009, le Centre de Calcul de l'IN2P3/CNRS (CC-IN2P3), dispose dans sa salle informatique de 350 serveurs supplémentaires, basés sur une technologie novatrice permettant une réduction significative de la consommation électrique.

Ces machines IBM iDataPlex type dx360, octocoeurs, viennent compléter un parc de plus de 10 000 coeurs de processeurs dédiés aux expériences de recherche dans les domaines de la physique corpusculaire, et dans une moindre mesure celles des sciences humaines et sociales et des sciences du vivant. Situé sur le domaine scientifique de la Doua (à Lyon - Villeurbanne), le CC-IN2P3 accroît depuis plusieurs années considérablement ses ressources informatiques afin de satisfaire aux besoins d'expériences d'envergure internationale, telles que celles du LHC, le plus grand accélérateur de particules du monde.

Le système nouvellement installé est constitué de serveurs IBM iDataPlex intégrés par Serviware. La solution proposée par Serviware et IBM, parfaitement en phase avec les priorités actuelles du CC-IN2P3, donne le ton d'une nouvelle approche dans la conception des Data Centers actuels : un maximum de puissance sur un minimum de surface au sol et une consommation énergétique limitée.

Le choix des serveurs iDataPlex permet d'obtenir une consommation totale de 205W par serveur. Les châssis iDataPlex favorisent la bonne circulation de l'air et réduisent la consommation des ventilateurs. Le système de refroidissement basé sur les nouvelles portes à eau IBM RDHx installées à l'arrière des châssis et connectées par échangeurs Eaton-Williams à un circuit d'eau réfrigérée, absorbe la totalité de la chaleur dégagée par les serveurs. Ainsi, tout en atteignant une très haute densité et de très bonnes performances, la ferme de calcul du CC-IN2P3 reste dans une enveloppe thermique et électrique très raisonnable en conformité avec les contraintes d'alimentation et de place disponible.

C'est la deuxième installation de ce type en Europe, toutes les deux situées en France et réalisées par Serviware. Elle sera officiellement inaugurée le jeudi 14 mai, à partir de 16h30.

Vous pourrez suivre l'inauguration en direct en vous connectant à l'adresse <http://webcast.in2p3.fr/iDataPlex>.