



Vue artistique d'une bibliothèque de bandes du Centre de calcul utilisée pour le stockage à long terme. (Photo Jeff Frost)

# Informatique

## Amélioration et migration

En vue de sa deuxième période d'exploitation, les équipes des expériences du LHC et de la Grille de calcul mondiale pour le LHC (WLCG) ont modernisé leurs infrastructures et services informatiques. Pendant le LS1, le département IT a doublé la capacité pour les expériences du LHC grâce à l'installation de quelque 100 pétaoctets (Po) supplémentaires de mémoire disque et près de 60 000 nouveaux cœurs de processeurs. La capacité de calcul de l'infrastructure en nuage privée du CERN a ainsi presque doublé l'an dernier, avec plus de 150 000 cœurs. Le LS1 était l'occasion idéale de migrer les données archivées sur bandes magnétiques vers des supports de plus haute densité. Environ 85 Po de données ont ainsi été copiés en 2014 et 2015. Au total, cela a permis de libérer 30 000 emplacements pour cartouches afin d'y stocker plus de données.

## Nouvelle exploitation, nouveaux records

Avec le début de la deuxième période d'exploitation du LHC, la collecte des données a atteint des records. Environ 40 Po de

données ont été enregistrés au CERN en 2015. Sur les 30 Po produits par les expériences du LHC, 7,3 Po ont été collectés en octobre, un record, et jusqu'à 0,5 Po ont été enregistrés sur bandes chaque jour durant l'exploitation avec ions lourds. À titre de comparaison, le système d'archivage sur bandes du CERN s'était enrichi d'environ 70 Po de données durant la totalité de première période d'exploitation du LHC. La Grille WLCG a également battu un nouveau record avec 51,1 millions de calculs effectués en octobre.

## Préservation des données

Le CERN gère le plus important volume d'archives de données scientifiques de la physique des hautes énergies. Fin 2015, le système d'archivage sur bandes contenait 132 Po de données, soit 410 millions de fichiers répartis sur environ 20 000 cartouches de bandes, entreposées dans sept bibliothèques équipées de 86 lecteurs. La plupart de ces données, qui proviennent des expériences passées et présentes de physique des hautes énergies, doivent être préservées *ad vitam aeternam*.



Les participants à la 39<sup>e</sup> École d'informatique du CERN en septembre 2015. (OPEN-PHO-CCC-2016-001-1)

Les bits sur une bande magnétique sont plus petits que des bactéries ; la moindre éraflure sur une bande peut donc détruire un nombre important de données. Par conséquent, il est important de protéger l'environnement matériel contre tout risque de contamination pour assurer la préservation des données. En 2015, le CERN a mis en place un système de surveillance des flux d'air basé sur des capteurs d'ambiance. Ces capteurs ont été conçus au CERN et placés dans les emplacements destinés à accueillir les bandes, dans les bibliothèques de sauvegarde. La précision et la rapidité de ces capteurs sont comparables aux systèmes vendus sur le marché, mais sont cinquante fois moins chers et ne nécessitent aucune maintenance. La conception matérielle, les logiciels et les micrologiciels sont accessibles gratuitement sous licence *open source*.

En juin a eu lieu au CERN le premier atelier de la collaboration DPHEP (*Data-Preservation and Long-Term Analysis in High-Energy Physics*), dont le but est de répondre aux problèmes de la préservation des données dans la physique des hautes énergies. Une nouvelle étape a été franchie avec la publication du rapport de situation sur les progrès réalisés depuis la publication en mai 2012 du plan d'action de la collaboration DPHEP.

## Logiciels libres pour une science ouverte

Les efforts du CERN pour développer des logiciels *open source* ont été fructueux. Zenodo, un système d'archivage de données ouvertes ayant bénéficié d'un cofinancement de la Commission européenne, permettant aux chercheurs de partager des articles, des données et des logiciels, a pris son essor en 2015. Il est devenu la plateforme de référence pour publier des logiciels scientifiques en lien avec Github, site de partage de logiciels, assurant ainsi la distribution de plus de 3 000 contributions scientifiques.

Le logiciel *open source* Invenio, initialement développé au CERN, a été modifié pour améliorer les possibilités de réemploi de chaque module, permettant ainsi à de nouveaux projets de logiciels de voir le jour dans différents pays. L'intérêt pour TIND, une *spin-off* du CERN, a également été croissant. Sa clientèle,

qui inclut CALTECH et les Nations Unies, fait appel à ses services pour effectuer l'installation d'Invenio.

En 2015, plus de 200 sites utilisaient le logiciel de gestion de conférences et d'événements Indico, ce qui témoigne de l'engagement du CERN pour favoriser la collaboration. Le CERN a aussi publié un logiciel *open source* qui enrichit d'un tableau de bord son infrastructure de visioconférence basée sur Vidy. Ce dernier a été adopté par des universités, hôpitaux et entreprises. Via de multiples collaborations, EOS, le système de stockage sur disques pour l'informatique du LHC, a également été mis à la disposition d'entreprises et de communautés d'utilisateurs pour qu'ils l'intègrent à leurs systèmes traitant de grands volumes de données ou l'utilisent pour construire des systèmes de stockage distribués.

## Participation à des projets *open source*

Le CERN a aussi contribué au développement de logiciels *open source* extérieurs pour les adapter aux besoins de l'Organisation. La contribution à ownCloud compte parmi les plus importantes. Ce dernier est utilisé pour fournir le service CERNBox, un équivalent de Dropbox sécurisé et évolutif. Le CERN contribue également à OpenStack et Ceph. OpenStack est essentiel au développement des services en nuage du Centre de calcul du site de Meyrin et de celui de Wigner à Budapest, géré à distance. Ceph est par ailleurs le système de stockage en réseau le plus utilisé pour OpenStack.

Le CERN a apporté plus de 90 contributions à la dernière version d'Openstack et a travaillé sur de nouvelles fonctionnalités importantes de Ceph. L'objectif initial avec Ceph était de fournir un service de stockage pour le nuage OpenStack du CERN, mais il a été étendu afin de s'attaquer aux défis à venir liés au stockage des données du LHC. En 2015, l'équipe de développement a réalisé des essais à grande échelle en collaboration étroite avec les concepteurs de Ceph afin d'explorer les limites de ce système. Les résultats pourront potentiellement servir à d'autres partenaires du CERN.

---

## Données ouvertes

En novembre 2014, le CERN avait inauguré son portail de données ouvertes, permettant aux expériences du LHC de partager leurs données avec deux cibles principales : d'une part, la communauté scientifique, y compris les chercheurs ne faisant pas partie des expériences du CERN et les citoyens volontaires ; d'autre part, les élèves, étudiants et personnes en formation par le biais de ressources sélectionnées. Toutes les données sont mises dans le domaine public sous licence Creative Commons zéro, une initiative inédite au CERN. Les scientifiques peuvent les citer grâce à un identificateur d'objet numérique unique.

En 2015, ce portail de données ouvertes a été enrichi de nouveaux codes et données. Son utilisation a augmenté et, dans un esprit de libre accès à la science, un certain nombre d'applications inattendues sont nées, telles que l'entraînement de systèmes d'analyses de données à grande échelle. L'équipe chargée des données ouvertes a transmis son expérience et inspiré de nombreuses équipes du monde entier en participant à des conférences et à des forums en ligne.

---

## La science dans les nuages

Le CERN contribue à l'établissement d'un « Nuage européen pour la science ouverte » en s'appuyant sur l'expérience tirée de différents projets, dont Helix Nebula, un partenariat public-privé. Les travaux de Helix Nebula ont conduit le CERN à diriger un programme Horizon 2020 nommé PICSE. Son objectif est de constituer un réseau de passation de marchés mutualisant les organisations de recherche publiques souhaitant utiliser des services en nuage commerciaux pour leurs programmes de recherche. En 2015, PICSE a étudié la faisabilité d'un système international de pré-commercialisation mutualisant différentes organisations publiques. La Commission européenne a décidé de cofinancer le projet d'achat pré-commercialisation HNSciCloud dirigé par le CERN, qui a débuté en janvier 2016.

## Éducation et partage

Depuis le début des années 1970, l'École d'informatique du CERN vise à offrir un enseignement de haut niveau et promeut l'échange de connaissances dans le domaine de l'informatique scientifique auprès de jeunes chercheurs et ingénieurs impliqués dans la physique des particules ou autres sciences. Elle se compose de trois écoles distinctes, dont chacune a sa spécialité. En 2015, l'École principale, qui dure deux semaines et tient lieu d'université d'été dispensant des cours magistraux et pratiques, a eu lieu à Kavala, en Grèce. Parmi les 76 participants venus de 47 instituts, 68 ont réussi l'examen et reçu 6 crédits ECTS.

Les Centres de calcul de Meyrin et de Wigner hébergent environ 15 000 serveurs, renouvelés tous les 4 à 5 ans, quand ils ne permettent plus d'atteindre les objectifs des programmes de recherche du CERN. Cependant, ils sont toujours utiles pour d'autres applications moins exigeantes. En mars, 224 serveurs et 30 commutateurs réseau ont ainsi été donnés au COMSATS *Institute of Information Technology* d'Islamabad, Pakistan, afin de servir aux scientifiques travaillant sur l'expérience ALICE du LHC. En août, le CERN a donné 384 serveurs, 24 commutateurs et 26 racks à des instituts mexicains, qui les

utilisent pour divers projets scientifiques et pédagogiques dans le domaine de la physique, des mathématiques de l'énergie et de l'environnement. À ce jour, neuf pays ont bénéficié de tels dons : la Bulgarie, l'Égypte, le Ghana, le Maroc, le Mexique, le Pakistan, les Philippines, le Sénégal et la Serbie.

## Nouvelle phase pour CERN openlab

En janvier, CERN openlab a entamé sa cinquième phase triennale. Par le biais de ce partenariat public-privé unique, le CERN travaille avec des entreprises informatiques et des instituts de recherche afin d'accélérer le développement de technologies de pointe pour la communauté mondiale du LHC. Huawei, Intel, Oracle et Siemens sont des entreprises partenaires ; Brocade, Cisco, IDT, Rackspace et Seagate sont des contributeurs, et Comtrade et Yandex sont des membres associés. Pour la première fois, d'autres organisations de recherche publiques ont rejoint le CERN openlab. Il s'agit du *European Bioinformatics Institute* (Royaume-Uni), du *GSI Helmholtz Centre for Heavy Ion Research* (Allemagne) et de l'Université de Newcastle (Royaume-Uni). Les thèmes retenus pour cette nouvelle phase sont, entre autres, les systèmes d'acquisition des données nouvelle génération, les plateformes de calcul matérielles et logicielles optimisées pour la simulation et l'analyse, la gestion et le stockage de données à grande échelle, les opérations et la fourniture de l'informatique en nuage, et les plateformes et applications d'analyse de données.

CERN openlab a tenu deux nouveaux événements : une journée portes ouvertes en juin, première du genre, et un événement intitulé « Innovation et entrepreneuriat » en octobre, en collaboration avec le groupe Transfert de connaissances du CERN et Ideasquare. Le programme des étudiants d'été, organisé par CERN openlab, a également continué à s'affirmer, avec 40 étudiants de 27 nationalités.