

# CERN: le grid digérera les données produites par le LHC

Développé par le CERN et inauguré début octobre, le grid traitera l'énorme quantité d'informations produites par le nouveau collisionneur LHC. L'ensemble composé de centres de calcul du monde entier sera également exploité pour d'autres recherches. *Rodolphe Koller*

Petit récapitulatif: le Large Hadron Collider (LHC) est installé dans un tunnel de 27 km de circonférence enterré à une profondeur de 50 à 150 mètres. L'accélérateur produira des collisions frontales entre des faisceaux de particules de la même espèce (protons ou ions) - chaque faisceau contenant 300 000 milliards de particules et pouvant parcourir pendant 10 heures un trajet équivalent à un aller-retour de la Terre à Neptune. Bien que les chances de collisions soient minces (1 pour 10 milliards), le LHC générera près de 600 millions de collisions par seconde. Quatre détecteurs seront chargés d'observer ces collisions à haute-énergie dans des conditions proches de celles qui prévalaient après le Big Bang pour permettre aux physiciens de répondre à des questions telles que l'origine de la masse ou la nature de l'énergie noire.

## Des volumes de données gigantesques

Dès le départ, un des défis majeurs consécutifs des expériences du LHC a été le traitement des énormes volumes de données générés par les détecteurs. Ceux-ci produiront en effet pas moins de 15 pétaoctets de données par année (15 millions de gigaoctets). Avec quelques 7000 physiciens analysant ces données aux quatre coins du globe, le CERN s'est trouvé face à un besoin d'environ 100 000 processeurs et d'un espace mémoire de 45 pétaoctets pour la première année d'opération. Vu les contraintes budgétaires et le fait que la plupart des laboratoires ayant à traiter les données du LHC disposaient de leurs propres ressources informatiques, les responsables IT du CERN ont décidé de met-

tre en place une structure de type grid (voir glossaire).

## 150 sites sur la planète

Le grid mis en place par le CERN est composé de plus de 150 sites dans plus de 60 pays. Les divers centres de calcul sont organisés en 3 couches: une couche zéro (tier 0) qui est le site du CERN, une première couche composée de 11 sites (tier 1) reliés au CERN par des liens dédiés de 10 Gbits, et une seconde couche faite de plus de 140 sites (tier 2). Le site du CERN accomplit un premier traitement des données et les conserve dans leur intégralité. Les données traitées sont envoyées sous forme de jobs (un ensemble de données et de tâches à effectuer) aux sites tier 1, qui sont de grands centres de calcul à même de sauvegarder les données reçues et de procéder à des analyses intensives. Ils servent également de centres de données aux sites tier 2, lesquels doivent répondre aux besoins d'analyse et de simulation des physiciens utilisateurs finaux. Les défis majeurs posés par la mise en place du grid sont l'orchestration complexe d'un ensemble de centres de calcul dispersés et hétérogènes, les exigences en termes de fiabilité et de sécurité, et le fait que le grid doit être en mesure de traiter les demandes non-planifiables des utilisateurs, de sorte que sa structure doit être facilement extensible.

## Pas uniquement pour les physiciens

Il n'y a pas que les physiciens qui s'intéressent à la capacité offerte par le grid du LHC. Il est par conséquent mis à disposition de projets d'autres disciplines gourmandes en termes de puissance de calcul et de capacité de stoc-



Le docteur Ian Bird, responsable du projet de grid pour le LHC, suppose que les technologies de grid computing et de cloud computing vont converger.

kage. Deux initiatives, l'Enabling Grids for E-science (EGEE) et l'Open Science Grid (OSG), permettent ainsi aux chercheurs de multiples sciences de profiter du grid: agronomie (sources et diffusion des pesticides), lutte contre les épidémies (découverte de nouveaux médicaments), astronomie (données fournies par les observatoires du monde entier) ou encore énergie (projet ITER pour la production d'énergie basée sur la fusion).

## HP, Intel et Oracle sont de la partie

Le développement du grid a également bénéficié de la collaboration intensive d'entreprises privées du secteur IT. L'openlab a ainsi pour fonction de faciliter l'accès pour le CERN aux technologies d'avant-garde des fournisseurs et pour les entreprises de tester ces innovations dans un environnement extrême et en interaction avec les éléments d'autres fabricants. Durant la journée d'inauguration du grid, des prix ont été remis aux principaux partenaires d'openlab: HP, Intel et Oracle.

Interrogé sur les liens entre grid computing et cloud computing, le docteur Ian Bird, responsable du projet, nous a déclaré: «Le grid a surtout une visée scientifique, alors que le cloud a des applications commerciales, mais il est probable que ces technologies convergent.»

 Grids: Le grid computing est un service de partage de capacité d'ordinateurs. Tout comme le web et l'e-mail, les grids (grilles) se basent sur internet. Ils sont toutefois plus complexes puisqu'il ne s'agit pas de partager des informations ou des messages, mais bien de partager la capacité de mémoire et de processeurs des ordinateurs qui en font partie. Les grids comprenant souvent des ordinateurs appartenant à des personnes de diverses institutions et de divers pays, leur fonctionnement soulève des défis importants pour garantir la sécurité, gérer les accès et autoriser les utilisateurs.