

L'Action Concertée Incitative GRID (Globalisation des Ressources Informatiques et des Données)

Michel Cosnard

Directeur de l'ACI GRID
Michel.Cosnard@inria.fr

Les premières technologies de grilles de calcul sont apparues il y a quelques années sous le nom de Calcul Global ou de Peer to Peer (Pair à Pair). Les deux plus célèbres sont [SETI@home](#) pour la recherche de signaux extra terrestres et Napster pour l'échange de musiques sur Internet. Entrent dans la même classe, les méthodes utilisant des calculateurs répartis pour casser des codes de cryptographie. Le procédé est le suivant : un serveur distribue des données à des ordinateurs clients ; chaque ordinateur récupère une donnée et exécute un programme de traitement et envoie le résultat au serveur ; le serveur rassemble les résultats, relance de nouvelles données vers les ordinateurs et en tire les informations pertinentes. Au même moment, le système Globus était développé pour relier ensemble des supercalculateurs américains et construire ainsi un calculateur parallèle encore plus puissant. Dans ce cas, il ne s'agit plus de technologies client-serveur puisque tous les ordinateurs peuvent lancer des calculs et utiliser la puissance de tous les autres. Les applications traitées sont d'une plus grande difficulté de mise en œuvre. Les projets européens DataGrid et EuroGrid s'appuient sur ce type de technologie.

Ces modèles et logiciels sont les plus célèbres et les plus utilisés. Cependant, ils ne constituent que les premières versions (encore très rudimentaires) des modèles de calcul sur la grille. L'ACI Globalisation des ressources informatiques et des données (GRID), mise en place en février 2001 par le Ministre de la Recherche Roger-Gérard Schwartzenberg, a pour objectif le développement des activités de recherche dans ce domaine pour inventer et mettre au point de nouveaux modèles. Il s'agit de mettre en réseau sur internet la puissance de calcul et de stockage des ordinateurs installés. Ce "super web" vise à répondre aux besoins des scientifiques en capacité de stockage et de calcul. L'action GRID constitue un véritable moteur d'expérimentation qui pousse la recherche dans les environnements de programmation distribués, dans les réseaux à très haut débit comme dans les nouvelles interfaces d'accès en langage naturel.

Financée par le Fonds National de la Science, l'action GRID est destinée à accroître l'effort de recherche national dans le domaine de la Globalisation des Ressources Informatiques et des Données. L'objectif de l'ACI est de dynamiser et de rendre opérationnelle, à un moment critique, la contribution des équipes de recherche françaises dans le domaine, en soutenant les équipes déjà actives, en attirant de nouveaux acteurs, en favorisant les rencontres entre concepteurs de nouvelles solutions et utilisateurs. En s'appuyant sur des plates-formes de réseaux à haut débit fournies par RENATER et le VTHD (réseau à très haut débit) mis en œuvre par le RNRT (Réseau national de recherche en télécommunications), et sur des centres de calcul nationaux et régionaux, sur des infrastructures régionales ou à l'échelle d'un campus, plusieurs axes de recherche sont développés pour contribuer à la mise en œuvre de "grilles expérimentales" : déploiement d'outils logiciels (accès distribué, authentification, sécurité et gestion des échanges et des ressources, interfaces pour le "metacomputing" ou l'exploration de gisements d'informations, aide à la décision, etc.), systèmes et environnements pour le calcul distribué ou pour l'exploitation d'ensembles de données de très grande taille, modélisation, algorithmique, couplage de codes, visualisation, pré et post-traitements. Les objectifs sont de contribuer à la création des outils logiciels permettant la mise en œuvre d'un "réseau ambiant", rendant l'accès totalement transparent à des ressources informatiques « potentiellement illimitées » et à la création d'entreprises, en particulier dans le domaine des ASP (application services providers), et de permettre aux équipes de recherche française de se positionner au plan international, en particulier européen en préparant les réponses au 6ème programme européen de recherche dans les technologies de l'information.

La programmation des applications sur des grilles de calcul est difficile puisque les ressources informatiques sont distribuées (géographiquement) et sont aussi parallèles (super-calculateurs). Globus

permet de masquer la nature distribuée de ces ressources en considérant la grille comme un super-calculateur parallèle virtuel dont la programmation est assurée à l'aide de bibliothèques de communication par échange de messages. Mais il ne permet pas de traiter facilement toutes les applications qui pourraient utiliser la grille. Une approche, qui sera développée dans le cadre de l'ACI GRID, est de combiner les modèles du calcul parallèle (à base de tâches et de messages) et les modèles du calcul distribué (à base d'objets distribués Java et de composants logiciels Corba). Dans cette approche une application pour la grille est un ensemble d'objets ou de composants logiciels en interaction, ces objets pouvant être eux-mêmes des programmes parallèles.

Les projets de Calcul Global (SETI@home, Entropia, ..) offrent essentiellement les capacités de calcul des ordinateurs volontaires à des applications à parallélisme trivial. D'un autre côté, les systèmes Pair à Pair comme (Napster, Gnutella, ...) exploitent les ressources de stockage de PC volontaires. Cependant aucun système Pair à Pair ne propose une infrastructure permettant de lancer des calculs sur les PC participants. Cette approche, qui sera développée dans l'ACI GRID, vise à augmenter les fonctionnalités des systèmes de Calcul Global en intégrant la possibilité d'exploiter les ressources de stockage des PC participants et en offrant la possibilité à n'importe quel PC participant de lancer lui-même des calculs sur le système.

Une approche de plus haut niveau se base sur le concept d'ASP (Application Service Provider) dans lequel des clients accèdent à des ressources logicielles à distance. Il s'agit de déployer un ensemble de serveurs de calcul sur les ordinateurs d'une grille de calcul et d'en offrir l'accès à des clients répartis sur le réseau. L'objectif est de permettre l'accès à des machines à fortes puissances de calcul et à grande capacité mémoire à distance sans forcément être un expert du parallélisme, sans avoir à installer sur sa machine des logiciels compliqués et aux interfaces hermétiques, et en pouvant utiliser des codes (et des données) confidentiels sur les serveurs de ceux qui les développent et ceci de manière sécurisée.

Les systèmes d'information contiennent maintenant de gros volumes d'information sous des formes très variées. Les grilles de données permettent de répartir ces données sur un ensemble d'équipements évoluant dynamiquement et accessibles au plus grand nombre. Pour gérer ces grilles, il faut des systèmes permettant d'accéder à de multiples sources de données préexistantes, autonomes et potentiellement hétérogènes. Offrir un accès transparent à un ensemble de ressources passe par la définition d'une infrastructure qui rend la complexité de l'architecture sous-jacente transparente à l'utilisateur en offrant des facilités de conception, d'intégration, d'interrogation et d'administration d'une manière fiable et efficace entre les applications.

Les applications des grilles de calcul et de données se retrouvent dans tous les domaines scientifiques. C'est aussi un des buts de l'ACI GRID que de constituer des consortiums pluridisciplinaires permettant de résoudre des problèmes nouveaux ou d'une taille considérable.