

Grilles de calcul et Réseaux pair-à-pair

Giovanna Di Marzo Serugendo
Université de Genève

Motivation

- ▶ Les ordinateurs sont toujours plus rapides, mais ...
- ▶ ... ils sont toujours trop lents pour résoudre certains types de calculs scientifiques
- ▶ Idées:
 - grouper des ordinateurs ensembles
 - construit à un endroit pour résoudre un problème donné
 - utiliser le temps de calcul inutilisé des ordinateurs dispersés à travers le monde
 - mise en commun de ressources par des communautés scientifiques en vue de résoudre des problèmes communs

Grilles de Calcul

« Une grille de calcul est un environnement persistant qui permet à des applications logicielles d'intégrer des instruments, des affichages, des données, des ressources de calcul et de stockages, gérées par différentes organisations dispersées à travers le monde »

--The Globus Project

3

Grilles de Calcul (Grid)

- ▶ Calculs distribués à l'échelle mondiale impliquant la collaboration de plusieurs sites
- ▶ Mise en commun de ressources par des organisations qui désirent collaborer (instruments, logiciels, données, CPU, disques)
- ▶ Faire bénéficier les utilisateurs de la mise en commun de la puissance de calcul et de stockage
- ▶ Pourquoi cela devient-il possible?
 - réseaux à haut débit sur de longues distances
 - ordinateurs individuels sont de plus en plus puissants

4

Grilles de Calcul

► Analogie: réseau électrique

- fournit à chaque utilisateur les ressources nécessaires au moyen d'une interface standard (prise de courant)
- la complexité du réseau électrique allant de la centrale électrique à l'utilisateur est cachée
- les composants qui forment le réseau électrique sont hétérogènes
 - gérés par différentes organisations
 - composants physiques différents
 - les consommateurs ont des besoins différents

5

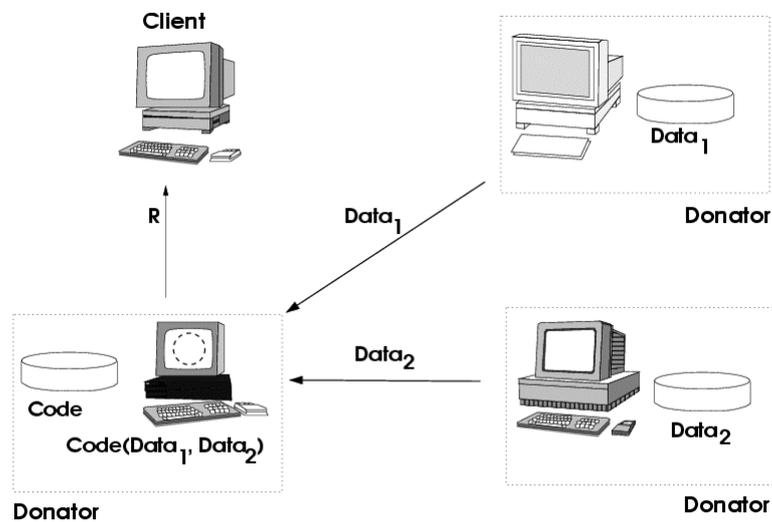
Grilles de Calcul

► Grille de calcul:

- infrastructure logicielle et matérielle qui procure à un utilisateur final un accès à des capacités de calcul et de stockage de masse hautement distribué
- utilisateur accède la grille à travers une interface standard
- l'accès aux ressources est fiable, efficace et bon marché
- hétérogénéité: matériel, logiciel, politiques de sécurité, contraintes d'exploitation, profil des utilisateurs
- la complexité liée aux différences doit être invisible pour l'utilisateur final

6

Grilles de Calcul



7

Domaines d'applications

- ▶ **Calcul intensif distribué**
 - additionner les capacités de plusieurs machines afin de résoudre un problème donné (e.g., simulations)
- ▶ **Calcul à haut débit**
 - ordonnancer un grand nombre de tâches peu couplées en utilisant les cycles de calcul inutilisés d'ordinateurs au repos (e.g., résolution de problèmes cryptographiques)
- ▶ **Calcul à la demande**
 - utilisation temporaire de ressources dont l'utilisation permanente ne serait pas rentable

8

Domaines d'applications

▮ Traitement intensif de données

- production de nouvelles données à partir de données géographiquement dispersées
- implique calcul et communication intensive
- Cern Grid, prévisions météorologiques

▮ Calcul collaboratif

- interactions entre humains en vue de partager des bases de données ou des simulations
- e.g., exploration conjointe de bases de données géographiques

9

Applications

▮ Entropia (www.entropia.com)

- Exploiter le temps de calcul non utilisé par les ordinateurs afin de résoudre des problèmes d'intérêt scientifique
- 30 '000 ordinateurs, 1Tflop/sec, identification du plus grand nombre premier
- Entreprise qui vend l'accès à des ordinateurs non utilisés

▮ Parabon (www.parabon.com)

- Fait appel à des propriétaires d'ordinateurs afin de mettre leur temps CPU inutilisé à disposition d'autres utilisateurs

10

Applications

► SETI@home (www.setiathome.ssl.berkeley.edu)

- utilise des ordinateurs connectés à Internet, pour analyser des données en provenance d'un radio télescope.
- Recherche d'intelligence extraterrestre
- 500 000 PCs, 1000 ans CPU/ jour
- le plus grand « ordinateur » distribué du monde

11

Applications

► Data Grid Computing (www.eu-datagrid.org)

- Grille de calcul complète intégrant plusieurs organisations
- Calcul intensif sur des quantités énormes de données
- Physique des hautes énergies
 - le nouvel accélérateur de particules du Cern produira des petabytes de données par an (10^{15})
 - nécessite 20Teraflops/sec (10^{12}) de puissance de calcul
- Biologie (données relatives au génomes)
- Observation de la terre (100GBytes de données par jour)

12

Infrastructure

► Globus Toolkit (www.globus.org)

- localisation des ressources
- réservation des ressources
- séquence des calculs
- gestion du code
- configuration
- tolérance aux fautes
- sécurité, contrôle d'accès personnalisé par site
- mécanismes de paiement

13

Réseaux Pair-à-pair

► Mise en commun des ressources et des services par des particuliers (anonymes) avec échange direct entre les systèmes

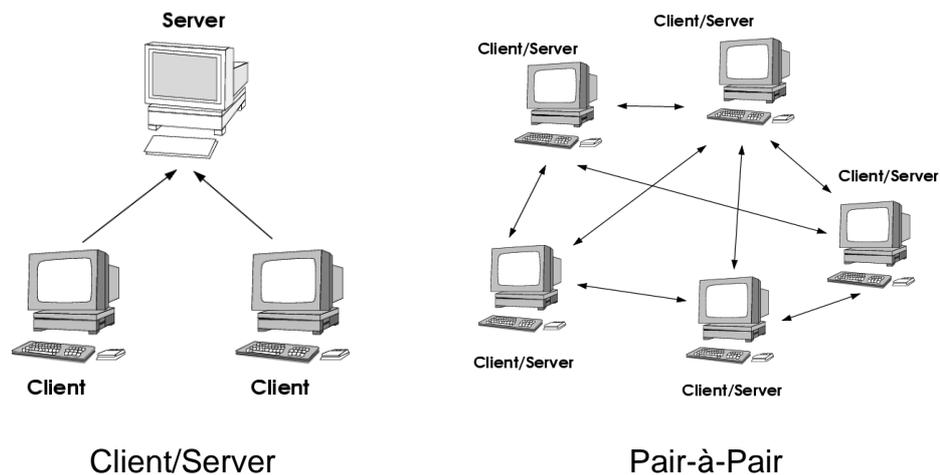
- CPU, disque, cache, données (fichiers)
- les particuliers acceptent que leur PC soit utilisé par un utilisateur distant (CPU, disque), et prennent une copie des informations désirées

► Pair-à-pair (peer-to-peer, P2P)

- relation d'égal à égal entre deux ordinateurs
- Chaque machine peut se comporter à la fois comme un client et un serveur

14

Réseaux Pair-à-Pair



15

Domaines d'applications

► Pour les utilisateurs:

- Partage de fichiers
- Utilisation du temps CPU inutilisé

► Pour des entreprises

- collaboration entre équipes (administration et gestion en-dehors et à l'intérieur de l'entreprise)
- délocalisation des services fournis par l'entreprise
- calculs distribués (pour les entreprises ayant de grands besoins en calculs)

16

Applications

- ▶ Napster (www.napster.com)
 - partage de fichiers entre utilisateurs
 - un serveur maintient un index des fichiers stockés sur chacun des PCs, permettant aux utilisateurs de retrouver leur chanson préférée
 - le fichier est ensuite transféré directement entre les deux machines
 - pas un vrai P2P

17

Applications

- ▶ Gnutella (www.gnutella.com)
 - similaire à Napster, mais ...
 - n'utilise pas un serveur central
- ▶ Groove Networks (www.groove.net)
 - plate-forme décentralisée permettant le développement d'applications inter-entreprises
 - environnement de groupe (calendriers, forums de discussion, ...)
 - services financiers, commerce électronique

18

Infrastructure

- ▶ Protéger l'hôte qui met son CPU à disposition pour des calculs
 - éviter des programmes malicieux qui endommagerait le système hôte
- ▶ Protéger les calculs et les résultats de l'utilisateur
 - éviter que le système hôte intercepte ou modifie les calculs qu'il accepte d'effectuer et les résultats qu'il produit

19

Infrastructure

- ▶ Distribuer un calcul sur plusieurs machines
 - un même calcul sur des données différentes (SIMD)
 - calculs parallèles sur des données différentes (MIMD)
- ▶ Système de Paiement
 - L'hôte qui met ses ressources à disposition reçoit une indemnité
 - L'hôte qui fait effectuer des calculs à distance doit payer pour la puissance et le temps de calcul utilisé

20

Conclusion

- ▶ Ce n'est que le début de la future utilisation d'Internet ...
- ▶ Accès à Internet à partir de téléphones portables ou PDAs (iPaq, PalmPilot, etc.)
 - piconets
- ▶ Semantic Grid
 - Ajouter plus de connaissance et de sémantique au Grid

21

Pour plus d'informations

- ▶ Descriptions de Grid
 - www.globus.org
- ▶ Descriptions P2P
 - www.peer-to-peerwg.org

22