



HAL
open science

L'essor et l'accessibilité du calcul distribué

Bel Dumé, Frédéric Desprez, Yves Robert

► **To cite this version:**

Bel Dumé, Frédéric Desprez, Yves Robert. L'essor et l'accessibilité du calcul distribué. Collection "20 ans d'avancées et de perspectives en sciences du numérique", 2012, 2 p. hal-00813227

HAL Id: hal-00813227

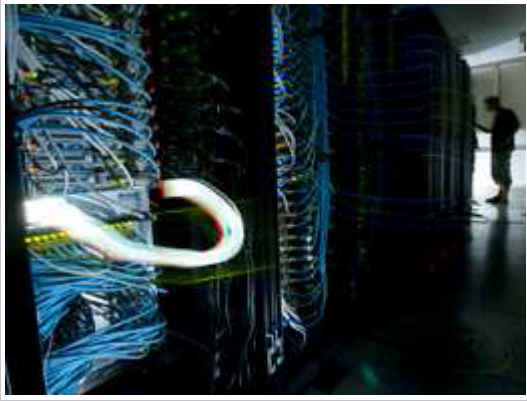
<https://inria.hal.science/hal-00813227>

Submitted on 15 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'essor et l'accessibilité du calcul distribué



En 10 ans, le calcul distribué sur plusieurs processeurs, c'est à dire le calcul parallèle en général, a pu se développer avec les avancées de l'algorithmique pour ce domaine. Le calcul distribué est en effet aujourd'hui utilisé dans les applications touchant la physique, la bio-informatique ou la recherche de données par des moteurs de recherche de type Google pour améliorer leurs performances. Il se loge également dans les Ipad qui, avec leurs quatre cœurs, constituent déjà une machine parallèle.

Témoignage de Frédéric Desprez et Yves Robert, chercheurs en algorithmique et intergiciels de calcul réparti.

Le but de nos recherches est de trouver des algorithmiques suffisamment génériques pour effectuer des calculs sur des modèles d'architectures variés, non spécifiques à un type de réseau ou à un nombre de processeurs donné.

Aujourd'hui, nous atteignons la limite des performances des ordinateurs car il devient trop difficile d'améliorer les composants sans que cela soit très coûteux. Une solution consiste donc à faire travailler les processeurs en parallèle pour les applications qui nécessitent des calculs complexes et performants.

Dans cette démarche, le défi principal de la fin des années 1980, a d'abord consisté à montrer qu'un assemblage de plusieurs ordinateurs pouvait fonctionner ensemble. La progression vers cette démonstration de l'algorithmique pour le calcul distribué n'a pas été rapide ni régulière ! Toutes les sociétés impliquées dans ce domaine dans les années 1990 ont fait faillite, car le marché de l'époque n'était qu'une niche. Les réseaux n'étaient pas aussi performants pour permettre des calculs à distance et, surtout, le parallélisme était réservé à la simulation numérique dans des grands centres de calculs nationaux. C'était vraiment une démarche de précurseur car, à l'époque, les composants continuaient de s'améliorer.

“ Le cloud computing et les machines Exascale prennent le relais. ”

Une des voies de développement du calcul distribué réside dans le cloud computing. Certes très commercial aujourd'hui, il soulève toutefois de nombreux axes fondamentaux de recherche. Nous y travaillons depuis quelques années au sein d'Inria, particulièrement sur la gestion dynamique des ressources, la virtualisation, le stockage, et la gestion de l'énergie car des centaines de milliers de processeurs multi-cœurs doivent travailler efficacement et simultanément ensemble. Ce modèle va plus loin que la simulation numérique « classique » avec des applications dans tous les domaines. Il faut cependant être capable de gérer l'échelle de ces plateformes, leur dynamique et enfin les quantités de données à traiter.



La grappe de PC conçue par HPLaboratories Grenoble et Inria en 2001 - © Inria / Photo R. Lamoureux

Un autre axe de recherche réside dans les machines Exascale (10^{18} opérations/seconde) qui devraient voir le jour d'ici 2020. Nous disposons déjà de systèmes dans lesquels plusieurs milliers de processeurs réalisent chacun un milliard d'opérations/seconde. Mais nous voyons encore plus grand et souhaiterions atteindre un milliard de milliards d'opérations/seconde, voire plus. Cependant de nombreuses problématiques seront à résoudre car le développement de nouveaux algorithmes pouvant fonctionner à si grande échelle, requiert de tolérer les pannes inexorables de certains composants, et accroît mécaniquement la consommation d'énergie.

ET DANS 20 ANS ?



Frédéric Desprez, directeur de Recherche Inria, équipe Avalon

« Mon rêve reste de collaborer avec des chercheurs d'autres disciplines, de participer à des transferts vers l'industrie et comme toujours d'explorer des nouvelles voies et algorithmes qui auront peut être des applications dans un futur lointain »



Yves Robert, professeur à l'Ecole Normale Supérieure de Lyon, équipe Roma

« C'est une chance de pouvoir travailler dans un domaine dont les avancées ont un réel impact. Ceci dit, j'aimais autant l'algorithmique parallèle quand elle ne servait à rien ! Mon plus grand espoir est que la recherche fondamentale continue à être soutenue et financée par Inria et les autres institutions.»

Dates clés



- 1990 : Arrivée de l'iPSC860 et du T-node d'Archipel à l'IMAG
- 1994 : Beowulf clusters (« grappes » de PC)
- 2006 : Premiers processeurs multi-cœurs
- 2006 : Premiers « clouds » déployés
- 2020 : Future machine Exascale

Numérique & société

- 1999 : Seti@home, un projet de calcul distribué utilisant la puissance de calcul des ordinateurs d'internautes volontaires. Deux objectifs : prouver la fonctionnalité et la viabilité du calcul distribué, détecter de la vie intelligente non terrestre. Ce projet de calcul distribué avec 5,2 millions de participants a accumulé plus de 2 millions d'années de temps d'analyse d'ordinateur. Source - <http://fr.wikipedia.org/>
- 2001 : Decryphon Lors du Téléthon 2001, l'AFM et IBM lancent un appel à la mobilisation des internautes. Objectif : réaliser la première cartographie du protéome (l'ensemble des protéines/molécules produite par les cellules). En seulement 2 mois, le défi est relevé avec 75 000 internautes alors qu'il aurait fallu plus de 1170 années pour le réaliser à l'aide d'un seul ordinateur. Source <http://www.decryphon.fr>
- 2011 : Google possède un parc de plus de 900 000 serveurs ce qui en fait le parc de serveurs le plus important au monde (2 % du nombre total de machines réparties sur 32 sites) pour collecter, traiter, indexer les données puis répondre aux requêtes des internautes. Source - <http://fr.wikipedia.org/>

1992 - 2012



- Collection "20 ans d'avancées et de perspectives en sciences du numérique" par les chercheurs d'équipes Inria de Grenoble et Lyon.
- www.inria.fr/20ansgrenoble

© Inria - Editions
Victoria