

Sommaire

- Le 16^{ème} Forum : ORAP aura 10 ans !
- Supercomputing 2003
- Evolution des moyens de calcul de l'IFP
- TOP500 : édition de novembre 2003
- Actualités BI-ORAP
- Agenda

16^{ème} Forum : l'ORAP aura 10 ans !

Le premier Forum de l'ORAP a eu lieu dans la salle de conférence du CNRS, rue Michel-Ange, le 24 juin 1994. Il y aura donc bientôt 10 ans !

Beaucoup de choses se sont passées dans le domaine qui nous intéresse et le "parallélisme", qui était au centre de la création de l'ORAP, a beaucoup évolué (clusters, grilles de calcul, P2P, etc). La petite communauté intéressée par les activités de l'ORAP pense davantage au calcul à grande échelle qu'à telle ou telle technologie parallèle. La déclinaison du nom ORAP n'est donc plus adaptée et nous avons décidé de modifier le logo et la déclinaison du nom; le site Web va être revu ainsi que le style de BI-ORAP. Ces travaux sont en cours .. et vous en verrez les résultats dès le mois de juin !

Le Conseil scientifique et le bureau de l'ORAP ont également souhaité marquer cet anniversaire en organisant un Forum un peu exceptionnel par sa durée, son lieu et, nous l'espérons, la qualité de son programme.

Ce forum aura donc lieu les 3 et 4 juin dans la grande salle de conférences de l'UNESCO à Paris.

Le programme sera consultable sur le site Web de l'ORAP dans les semaines qui viennent.

Compte-tenu de l'importance de l'organisation de cette manifestation, nous souhaitons que les personnes intéressées prennent contact dès maintenant avec Chantal Le Tonquèze : chantal.letonqueze@irisa.fr, 02 99 84 75 33, fax : 02 99 84 74 99

Supercomputing 2003

L'édition 2003 de la conférence et de l'exposition Supercomputing a eu lieu à Phoenix (Arizona) du 15 au 21 novembre. Les records d'affluence ont été battus avec plus de 7000 participants (dont près de 3000 inscrits à la conférence technique) et 10000 mètres carré d'exposition (certaines demandes d'exposants n'ont pu être satisfaites par manque d'espace). Pour la première fois, l'exposition "recherche" et l'exposition "industrie" étaient regroupées, ce qui a donné une meilleure visibilité aux 98 stands "recherche" (avec 128 stands "industrie").

Exposition recherche

Les deux pays les plus représentés dans cette exposition étaient les Etats-Unis et le Japon. Plus de 50 stands pour les Etats-Unis avec les laboratoires nationaux (LANL, LLNL, ORNL, Sandia, etc), les grands centres de calcul (NCSA, San Diego, ...) et de nombreuses universités (Boston, Cornell, Indiana, Utah, ...). Pour le Japon : 18 stands avec RIKEN, JAXA, Tohoku University, les université de Tokyo, de Tsukuba, Notons aussi la présence de Taiwan qui montrait comment une application en grille (Access Grid) leur avait permis de gérer l'épisode du SRAS (Syndrome Respiratoire Aigu Sévère), en interconnectant avec divers médias hôpitaux, généralistes, ambulances, patients sous surveillance (capteurs, téléphones mobiles).

La participation européenne était significative, avec une présence importante et structurée (stands grands et bien organisés) du Royaume-Uni (sous la bannière "e-Science"), de l'Allemagne et des Pays-Bas. Le stand des Pays-Bas regroupait plusieurs universités, avec beaucoup de résultats originaux en bioinformatique, en particulier ceux du groupe d'Henri Bal de l'université Vrije d'Amsterdam. Présence aussi de l'Espagne, de la Hongrie et de la France.

La France était représentée par l'INRIA dont le stand de 50 m² accueillait quatre projets de recherche et leurs démonstrations :

- Le projet grenoblois APACHE présentait une démonstration de transferts optimisés sur réseaux à haut débit. APACHE a développé un système de fichiers distribué utilisant des noeuds d'une grappe avec le protocole NFS. Cette solution de stockage peut ainsi être

déployée en quelques minutes sur une configuration existante sans en interrompre le fonctionnement. Au dessus de ce système, a été développé un outil de transfert efficace qui permet d'utiliser toute la puissance du réseau avec des vitesses de plus de 100MB/s (un CD en 7 secondes !), et ce sans avoir recours à de coûteux systèmes dédiés de stockage pour assurer l'accès aux données¹.

- Les démonstrations de l'équipe lyonnaise RESO concernaient les travaux menés dans le projet VTHD++ sur les réseaux actifs programmables hautes performances autour d'une épine dorsale Gigabit. La suite logicielle haute performance TAMANOIR² présentée permet le déploiement dynamique de services sur des équipements actifs programmables.
- Le déploiement et l'utilisation des grilles de calcul posent des problèmes de modèle de calcul, de déploiement, de sécurité. La bibliothèque ProActive³ (Open source, consortium ObjectWeb), et ses composants parallèles et répartis apportent des solutions novatrices à ces problèmes. En particulier la version présentée à SC'03 comprend :
 - des optimisations importantes pour les communications de groupe,
 - des outils de déploiement XML permettant d'exécuter facilement les applications sur des clusters ou la Grille (interfaces avec Globus, LSF, etc.),
 - des schémas pour la validation des Descriptors XML.
- Le projet XtremWeb⁴ de l'équipe Grand-Large vise à développer une plate-forme pour l'expérimentation des Desktop Grid dans ce contexte de la grande échelle. Cette plate-forme est utilisée comme base dans plusieurs projets en France comme l'ACI Grid CGP2P, le RNTL Casper, le projet Pluri-Formations Augernome XtremWeb et à l'étranger par différentes universités (UCSD, Tsukuba, Genève). XtremWeb faisait aussi l'objet d'une présentation invitée lors du Workshop DGRID 2003 (associé à SC'03), d'un poster, et d'une démonstration sur le stand INRIA.

Exposition industrielle

Les principaux acteurs industriels du domaine étaient présents : constructeurs d'ordinateurs, fournisseurs de technologies réseaux, intégrateurs, éditeurs de logiciels, ... SC'03 était donc l'occasion de "faire le point" sur leurs activités et projets. Voici quelques "échos des stands" qui complètent les articles parus dans le numéro 35 (avril 2003) de BI-ORAP.

1. Présentation détaillée : <http://www.inria.fr/actualites/sc03/apache.pdf>

2. <http://www.ens-lyon.fr/LIP/RESO/Tamanoir/>

3. <http://www-sop.inria.fr/oasis/ProActive/>

4. <http://www.lri.fr/%7Efedak/XtremWeb/introduction.php3>

Bull

Bull disposait d'un stand sur lequel était présentée la nouvelle gamme de serveurs NovaScale⁵ conçue autour des processeurs Intel Itanium2 et de l'architecture FAME (Flexible Architecture for Multiple Environments). Ils permettent l'exploitation simultanée de plusieurs environnements sur un même serveur, en particulier Windows, Linux et GCOS.

Cray

Cray affichait une certaine satisfaction, les ventes de systèmes X1 dépassant les prévisions (plus de 20 armoires installées). Deux installations X1 en Europe : la météo espagnole et l'université de Varsovie. Le modèle X1e devrait être disponible fin 2004, l'étape suivante étant le système *Black Widow* en 2006 (300 Tflops).

Par ailleurs, Cray a annoncé qu'un nouveau produit, développé à partir de la plate-forme *Red Storm* réalisée pour Sandia, serait commercialisé fin 2004. Il utilisera des processeurs Opteron d'AMD, un micro-noyau Linux sur les noeuds de calcul (4 ou 8 processeurs) pour améliorer la performance. Le réseau d'interconnexion sera un tore 3D développé par Cray.

Cray aura alors deux lignes de produits :

- X1, *Black Widow* ..., orientée vecteur/parallèle, avec des processeurs "propriétaires"
- RS, orientée scalaire, avec des processeurs du marché

Pour le plus long terme, Cray travaille sur le projet *CASCADE* de la DARPA qui devrait déboucher en 2010 sur un système ayant une performance crête d'un Petaflops.

HP

HP est n°1 (pour le nombre de systèmes) dans le TOP500 et va poursuivre sa stratégie basée sur :

- les processeurs Intel (Itanium 2 ...),
- les systèmes d'interconnexion Myrinet et Quadrics,
- Linux et les logiciels standard du marché,

avec deux lignes de produits :

- les gros serveurs à architecture SMP (maximum 64 processeurs),
- les clusters (2 à 4 processeurs par noeud).

HP présentait aussi LUSTRE, un "file system" développé par HP dans le cadre du programme *ASCI Path-Forward*. Ce file system Linux Open Source est destiné aux environnements de calcul de haute performance et peut gérer des milliers de noeuds de calcul, des petaoctets de stockage, une bande passante de plusieurs centaines de gigaoctets par seconde.

IBM

IBM partage avec HP la première place dans le TOP500, IBM étant n°1 pour la performance installée. La compagnie a regroupé dans une structure appelée "*Deep*

5. <http://www.bull.com/novascale/>

Computing” ses activités “calcul haute performance”, visualisation, fouille de données ...

IBM va continuer de proposer une grande diversité de solutions : grands serveurs SMP et systèmes distribués utilisant diverses technologies d’interconnexion (internes ou externes telles que Quadrics) et une gamme diversifiée de processeurs (Intel Xeon ou Itanium, AMD Opteron et bien entendu sa propre architecture Power).

L’architecture Power se décline en trois lignes de produits :

- pour les serveurs : Power 2, 3, 4, 4+, 5, ... ;
- pour les postes de travail et consoles de jeu : PPC 603e, 750, 970 (qui est le processeur G5 utilisé par Apple) ;
- pour les systèmes enfouis : PPC 401, 405GP, 440GP, 440GX.

Les “*Blade centers*” sont la base de l’offre IBM dans le calcul distribué. Ils sont développés à travers une collaboration (“*Blade Center Alliance*”) avec Intel, Microsoft, Myricom, Suse, RedHat, ... L’évolution des processeurs Power destinés aux serveurs serait la suivante : *

- Power4+ en 2003 : technologie 130 nm, 1,5 GHz
- Power5 en 2004-2005 : 130 nm, fréquence améliorée
- Power5+ : 2005-2006 : 90 nm, amélioration de la fréquence
- Power6 : 2006-2007 : 65 nm, nouvelle amélioration de la fréquence

Le module Power5 comprendra 4 chips (et sera vu comme un SMP à 4 voies), un cache de niveau 2 plus important (1,9 Mo), un cache L3 extérieur au chip, une amélioration de la bande passante. Les serveurs basés sur le Power5 pourront avoir jusqu’à 64 processeurs en architecture SMP.

Comme Cray, IBM est impliqué dans le projet de la DARPA qui doit déboucher sur le Petaflops en 2010.

Enfin, IBM présentait Blue Gene. Un rapide rappel de ses caractéristiques :

- Chip ASIC à deux processeurs (2,8 Gflops par processeur)
 - caches L2 et L3 sur le chip
 - 2 unités flottantes SIMD (load, store, compare, arith,.. sur les 2 éléments d’un vecteur)
 - interconnect (tore 3D et arbre de diffusion) sur le chip
- Carte à 2 chips (interconnect 1x2x1)
- Noeud à 16 cartes (32 chips, 4x4x2)
- Armoire à 32 noeuds (8x8x16)
- Système à 64 armoires : 65536 noeuds 64x32x32, performance crête de 360 Teraflops

Le réseau d’interconnexion est un tore 3D, *virtual cut through*, avec un arbre global pour les diffusions. Il y a des noeuds de service, des noeuds d’entrées-sorties et des noeuds de calcul.

Sur les noeuds de calcul, pas de pagination (mémoire physique directement “mappée” avec l’espace adresse

de l’application) pour améliorer l’efficacité. Pas de temps partagé, un seul thread ayant le contrôle d’un processeur de calcul.

Un premier système à 1024 processeurs, de la taille d’un grand poste de télévision, fonctionne (avec une démonstration de protein folding assez impressionnante) et est à la 73ème place dans le Top500 (Performance Rmax : 1435 Gflops, size(N) : 110592). Un système complet devrait être fourni à Livermore avant la fin 2005.

SGI

Un peu de morosité sur le stand SGI, le constructeur ayant vu ses parts de marché baisser dans le domaine du HPC (place dans le Top500 en baisse sensible : 41 systèmes).

Le “dernier né”, dans les serveurs HPC, est l’Altix 3000 basé sur le processeur Intel Itanium 2. Il peut avoir jusqu’à 128 processeurs et un seul noyau Linux en mémoire partagée. Au delà, la “clusterisation” des serveurs s’impose (8 noeuds de 128 processeurs actuellement, 16 noeuds en 2004, 523 noeuds un peu plus tard). Total dispose, à Pau, d’un cluster de 4 serveurs de 64 processeurs chacun (classé 192 dans le TOP500).

Enfin, SGI présentait Visual Area Networking, qui permet d’apporter sur le bureau de chacun la puissance graphique d’une Onyx qui intervient en serveur. Cette solution, qui permet de traiter les données sur un serveur central et de visualiser les images à distance, semble particulièrement attractive pour le travail coopératif. Elle est basée sur OpenGL Vizserver et a déjà été expérimentée avec succès par le CINES et le CERFACS.

Apple, Dell

Ces deux constructeurs disposaient chacun, pour la première fois, d’un stand important. Ils entendaient confirmer ainsi leur présence dans le domaine du HPC et plus précisément dans le monde des clusters.

Le cluster construit par l’Université de Virginie, avec des processeurs G5 d’Apple, et classé en 3^{ème} position dans le TOP500, était bien entendu une référence de premier plan. Notons aussi qu’Apple nous a indiqué travailler sur un environnement de calcul en grille XGrid pour ses serveurs.

Le nombre de clusters utilisant des PC fournis par Dell est aussi intéressant : 18 de ces clusters font partie du TOP500 !

Fujitsu

Fujitsu a abandonné le vectoriel et propose maintenant deux lignes :

- le serveur *PrimePower* HPC2500 qui utilise le processeur Sparc64 à 1,3 GHz (performance théorique : 666 Gflops pour un noeud de 128 processeurs, 85 Tflops en configuration maximale),

- les *High Performance Linux Clusters* à base de processeurs Intel Xeon et diverses technologies d'interconnexion (InfiniBand, Myrinet, ...)

Fujitsu présentait également un serveur original pour le calcul intensif : BioServer. Cette machine exécute en parallèle sur 1920 processeurs (qui tiennent dans un rack 2U) un logiciel de dynamique moléculaire (Gromacs de l'université de Groningen, disponible sous licence GNU).

Les processeurs sont des VLIW dérivés de ceux utilisés par les architectures parallèles de Fujitsu. Le système est un clone de Linux, développé par une société japonaise (Axe) pour les PDA et les mobiles. Un frontal contrôle la machine.

Interconnexion

Les vendeurs de solutions réseau "classiques" étaient présents : Myricom (myrinet), Quadrics, Cisco Systems, Foundry, Extreme Networks, ... mais le fait marquant est que la technologie InfiniBand est présente sur presque tous les stands des fournisseurs de clusters.

Mellanox Technologies a d'ailleurs annoncé lors du salon la disponibilité de la 3^{ème} version de son chip ASIC Infiniscale. Mellanox présentait aussi un switch 24 ports 10 Gb/s ou 8 ports 30 Gb/s. Les latences annoncées sont inférieures à 200 nano secondes, et le prix annoncé est de 300 \$ par port.

Une machine TOTS (Teraflops Off-The-Shelf) comprenant 3 baies : 192 processeurs Xeon Intel (128 pour le cluster Ciara et 64 pour celui de Rackable) et une interconnexion par InfiniBand (switches 4x24 ports Mellanox/ Infinicon pour le niveau 1, et un switch ISR9600 de Voltaire pour le niveau 2) était démontrée sur une application de l'université de l'état d'Ohio.

Stockage

Nous remarquons qu'en plus des solutions "classiques" de stockage réseau (NAS), de nombreux vendeurs proposent des solutions originales (smart disks) de disques utilisant des chips de calcul, soit des FPGA (Xilinx) ou des processeurs Celeron (Intel). L'objectif étant d'augmenter la bande passante en faisant éventuellement du pré-traitement local sur les données.

BlueArc propose ainsi son SiliconServer (FPGA) avec du fiberchannel entre les noeuds de stockage et du Gb ethernet en entrée sortie du serveur.

Panasas propose son StorageBlade dont les noeuds sont des lames comprenant 2 disques, un celeron et de la mémoire. Le système de fichiers est un système à objets s'appuyant sur la puissance de calcul disponible dans chaque lame pour détecter d'éventuels goulets d'étranglement et capable de faire migrer (automatic load balancing) des objets (fichiers, répertoires, ...) d'une lame vers une autre. Les performances sont annoncées pouvant aller jusqu'à 7 fois celles des meilleurs systèmes NAS classiques ...

Jean-Loïc Delhaye, Hugues Leroy

Evolution des moyens centraux de calcul de l'IFP

L'Institut Français du Pétrole (IFP) exerce une triple action de recherche et développement (R&D), de formation et d'information au service de l'industrie des hydrocarbures - pétrole et gaz - et de l'industrie automobile. Unique en Europe, ses activités couvrent l'ensemble de la chaîne pétrolière, de l'exploration-production jusqu'à l'utilisation des produits pétroliers dans les moteurs en passant par le raffinage et la pétrochimie.

Avec un budget de 300 Meuros, l'IFP emploie 1800 personnes réparties sur 3 sites : Rueil -Malmaison, Solaize (Lyon) et Pau.

Depuis octobre 2003, l'IFP a mis en service une nouvelle configuration de supercalculateurs en remplacement des systèmes SGI O2000 et NEC SX5 qui étaient en place depuis plus de 5 ans. L'IFP se dote ainsi d'un outil de calcul de premier plan qui répond aux besoins croissants en modélisation de ses programmes de recherche.

La modélisation à l'IFP

L'Institut Français du Pétrole est un important consommateur de ressources de calcul ; parmi les utilisations majeures, on peut citer :

- l'imagerie sismique 3D en exploration pétrolière qui permet d'obtenir une image des structures du sous-sol à partir de mesures sismiques réalisées en surface (chaque campagne sismique représente un volume de données de plusieurs centaines de Go) ;
- la simulation de la formation de bassins sédimentaires sur plusieurs dizaines de millions d'années afin de mieux situer les zones potentiellement riches en hydrocarbures ;
- la simulation des écoulements dans les réservoirs pétroliers (écoulements de fluides en milieux poreux) en physique complexe et avec des grilles de plusieurs millions de cellules afin d'optimiser l'exploitation des gisements et d'améliorer le taux de récupération des hydrocarbures ;
- la modélisation 3D dans les moteurs à combustion interne (mécanique des fluides multi-phasiques réactives) permettant la prédiction et la réduction des consommations et des polluants ;
- la conception, grâce aux techniques de modélisation moléculaire, de nouveaux catalyseurs chimiques utilisés dans les unités de raffinage-pétrochimie.

Ces applications présentent des caractéristiques très différentes en termes de besoins CPU, de besoins mémoire, de volumes d'entrées-sorties, de type de code (code développé en interne ou logiciel commercial, code parallèle ou vectoriel). Il convient donc d'offrir à chacune les ressources de calcul les mieux adaptées.

Un peu d'histoire ...

Depuis les années 60, et afin de répondre aux besoins croissants en modélisation de l'exploration et la production pétrolière, l'IFP a toujours disposé de puissants moyens de calcul centraux.

Jusqu'en 1987 l'IFP était équipé d'un CDC 7600 doté d'un processeur d'une puissance de 36 MFlops et d'une mémoire centrale de 0.5 Mo. Un Cray X-MP/14 (240 Mflops et 36 Mo de mémoire) lui a succédé en 1987 puis cette machine a été remplacée par un Fujitsu VP2400 (machine vectorielle monoprocesseur à 2.5 Gflops et 1 Go de mémoire) en mars 1992.

En février 1995, un Fujitsu VPP500 à 4 processeurs (1.6 Gflops – et 512 Mo de mémoire par processeur) est venu compléter la configuration.

Enfin en 1998 les machines Fujitsu ont été remplacées par une plateforme vectorielle NEC SX5 à 5 processeurs (8 Gflops par processeur et 32 Go de mémoire) et une machine scalaire parallèle SGI Origin 2000 à 20 processeurs.

La nouvelle configuration

Octobre 2003 marque le remplacement complet de la configuration des supercalculateurs de l'IFP avec un changement important en termes d'architecture. L'IFP a en effet décidé de ne plus poursuivre sa politique d'équipement en moyens centraux de calcul vectoriel et a pour objectif de migrer à moyen terme vers une solution tout cluster Linux.

Cette décision a été justifiée notamment par les raisons suivantes :

- les principaux codes industriels utilisés en production sur le NEC SX5 ont tous été parallélisés ;
- les performances actuelles des meilleurs processeurs scalaires donnent sur les applications IFP une équivalence de 1 processeur vectoriel pour 2 à 4 processeurs scalaires ;
- les clusters Linux équipant de plus en plus les clients et partenaires de l'IFP dans l'ensemble de la chaîne pétrolière, il est important pour l'IFP de supporter le même type de plates-formes que ces derniers.

Cette stratégie moyen terme se traduit dans un premier temps par un remplacement des moyens centraux de calcul actuels (NEC SX5 et SGI O2000) par :

- un cluster Linux IBM xseries composé de
 - 32 noeuds x335, chaque noeud étant composé de 2 processeurs Intel Xeon 3.06 Ghz, 2 Go de mémoire et 36 Go de disque SCSI ;
 - un réseau d'interconnexion haut débit faible latence de type Myrinet 2000 ;
 - une baie IBM FastT 700 en SAN avec 1.5 To de disque FC sous GPFS (General Parallel File System, le système de fichiers partagé parallèle IBM) ;
 - un réseau de service Ethernet 100 ainsi que 4 noeuds de services sécurisés et utilisés pour des

tâches de connexion, compilation, soumission et administration du cluster Linux.

- un cluster SMP IBM pseries composé de
 - 4 noeuds p655+, chaque noeud étant composé de 4 processeurs IBM Power4+ 1.7 Ghz, 16 Go de mémoire (32 Go sur un des noeuds) et 280 Go de disque SCSI ;
 - un réseau d'interconnexion haut débit faible latence de type IBM Federation ;
 - un accès à la baie SAN FastT700 IBM sous GPFS ;
 - un noeud de service IBM p630.

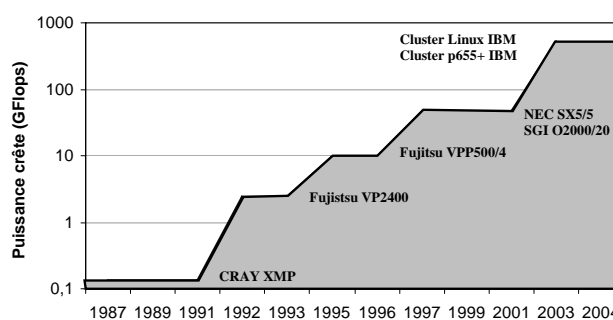
Ces 2 machines sont adressées par les utilisateurs de l'IFP à travers l'environnement de soumission de travaux Platform LSF 5.1.

Sur le cluster Linux s'exécute une grande variété d'applications issues de toute la chaîne pétrolière avec des profils très hétérogènes. Contrairement aux utilisations classiques des clusters Linux, celui de l'IFP exécute en production un ensemble d'applications multi disciplinaires adaptées à ce type d'architecture.

Sur le cluster SMP s'exécutent des applications qui nécessitent de grandes bandes passantes (mémoire, entrées sorties), des applications parallèles OpenMP ou des codes séquentiels 64 bits qui profitent des 16 ou 32 Go de mémoire disponibles sur les noeuds p655+.

Cette approche complémentaire se justifie par le fait que pour le moment les clusters Linux par certaines limitations matérielles et logicielles ne permettent pas de faire tourner efficacement sur une même machine tous les codes de calcul utilisés à l'IFP.

Evolution des moyens centraux de calcul IFP



Perspectives

La mise en place de la stratégie en calcul hautes performances de l'IFP se déclinera sur 2004-2006 en plusieurs actions :

- par une évolution annuelle de la puissance de calcul du cluster Linux ;
- par la mise en place d'un projet exploratoire visant à définir et à qualifier les différents éléments matériels (noeuds, réseaux d'interconnexion, systèmes d'entrées sorties parallèles, ...) et logiciels (système Linux, outils de développement, outils de soumission, checkpoint/restart, outils d'administration centralisés, ...) qui composeront la future architecture de calcul en

2006. La ligne directrice étant l'unification des deux types de machines actuelles dans une même plateforme qui, sous Linux sera composée de ressources hétérogènes et adressera l'intégralité des besoins de l'IFP.

Conclusion

Les nouveaux moyens de calcul de l'IFP ont été définis et dimensionnés pour traiter des problèmes R&D de plus en plus lourds et complexes, dans de meilleurs délais et à des coûts moindres.

Cette évolution marquée par l'arrêt du vectoriel et par une migration à moyen terme vers le tout cluster Linux permet ainsi de tenir compte de l'évolution rapide des technologies dans ce domaine et des choix de nos partenaires pétroliers et automobiles en matière de ressources de calcul.

Pour en savoir plus :
Stephane.Requena@ifp.fr, 01 47 52 56 05

TOP500 : 22^{ème} édition

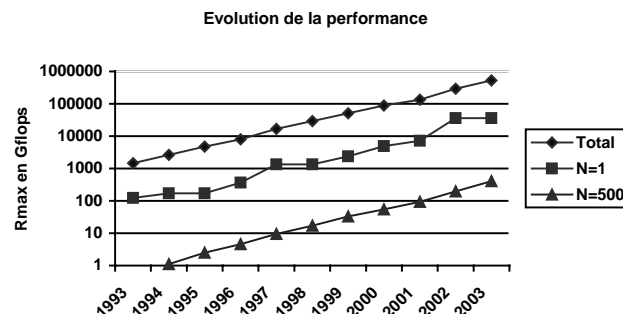
Comme chaque année, l'édition de novembre du TOP500, la liste des 500 ordinateurs les plus puissants installés dans le monde, a été publiée dans le cadre de la Conférence Supercomputng.

Eléments importants de cette 22^{ème} édition

- On retrouve en première position reste la machine japonaise, construite par NEC, "Earth Simulator" (35,8 Tflops Linpack)
- Les 9 machines suivantes sont installées aux USA
- La position 2 est conservée par la machine ASCI de Los Alamos, construite par HP (10,3 Tflops)
- Le système en position 3 dépasse aussi les 10 Tflops et c'est un cluster construit ("home made") avec des Apple G5 par l'université de Virginie avec un réseau InfiniBand
- Le quatrième système est encore un cluster, au NCSA, construit cette fois à partir de systèmes Dell avec des processeurs Intel Xeon et une interconnexion Myrinet
- 131 systèmes ont une performance Linpack supérieure à 1 Tflops. La performance cumulée des 500 systèmes de la liste est supérieure à 500 Tflops (0,5 Pflops) alors qu'elle n'était "que" de 375 Tflops il y a six mois. La machine la moins performante (classée en dernière position du TOP500) fait plus de 400 Gflops, contre 245 il y a six mois !
- IBM domine le TOP500 en terme de performance (35,4 %) devant HP (22,7%) et NEC. Par contre, HP domine le TOP500 en terme de systèmes installés (165 systèmes) devant IBM (159) et SGI (41).
- Le Cray X1 apparaît dans le Top500 avec une machine en 19^{ème} position (le X1 est la seule machine, avec le Earth Simulator, à avoir une architecture vectorielle).

- 50% des 500 systèmes et 57% de la performance totale sont installés sur le territoire américain. Par ailleurs, 90% des 500 systèmes ont été construits aux Etats-Unis, ce qui montre, si besoin était, que ce pays domine très largement ce secteur.

Evolution de la performance



L'échelle de l'ordonnée (puissance Rmax sur Linpack) est logarithmique.

En Europe

L'Europe perd du terrain Elle ne dispose plus que de 27% des systèmes et 13% de la puissance installée. En Europe, l'Allemagne domine très largement (25% des systèmes et de la puissance installés) devant le Royaume-Uni (23% du nombre de systèmes mais 33% de la puissance installée).

En France

La position de la France s'affaiblit. 16 systèmes (3% du parc mondial, 11% du parc européen) et 15 Tflops (3% du monde, 15% de l'Europe). Elle est devancée par l'Italie pour le nombre de systèmes.

Les sites premiers sites français :

- Le CEA avec 3 systèmes dont la machine à 4 Tflops (classée en position 15)
- TotalFinaElf avec 2 systèmes (dont un rang 55)
- La Société Générale (classée au rang 61)
- L'IN2P3 (au rang 167)
- L'IDRIS (au rang 264)
- Météo-France (au rang 280)
- ID-IMAG/INRIA (au rang 283)
- Le CINES avec deux systèmes dont l'un au rang 349

Actualités Bi-Orap

➔ Allemagne

Les trois centres nationaux allemands (LRZ à Munich, NIC à Jülich et HLRS à Stuttgart) ont lancé un magazine, *inSiDE*, destiné à faire connaître les nouveaux concepts du calcul de haute performance et les initiatives de la communauté scientifique allemande.

<http://www.hlrs.de/news-events/inside/>

➔ Belgique

La Belgique a lancé un projet de grille nationale, BEGrid, s'appuyant sur le réseau national de la recherche Belnet (backbone à 2,5 Gbit/s). Cette grille utilisera le logiciel EDG.

➔ Royaume-Uni

Le conseil britannique pour la recherche en physique des particules et en astronomie (PPARC) a annoncé qu'il allait financer à hauteur de 22,74 millions d'euros une grille de calcul qui sera mise en place grâce à une collaboration entre des chercheurs en physique des particules et des chercheurs en informatique. Cette grille devrait être intégrée à un projet de grille européen qui traitera les grands volumes de données résultant des expériences du LHC Large Hadron Collider) du CERN à partir de 2007.

<http://www.pparc.ac.uk>

➔ Etats-Unis

Nous signalions, dans le précédent numéro de BI-ORAP, deux rapports intéressants publiés sur les investissements fédéraux dans le calcul de haute performance. David Keyes, qui avait fait une présentation dans le Forum de mars 2003, nous signale un autre rapport intéressant : SCA-LeS (A Science-base Case for Large-scale Simulation). Merci à David ! Le volume 1 est disponible à l'URL suivante :

<http://www.pnl.gov/scales>

➔ Edition 2003 du rapport "Overview of recent supercomputers"

La 13ème édition du rapport "Overview of recent supercomputers" a été publiée par Aad van der Steen et Jack Dongarra.

<http://www.nwo.nl/ncf/overview-src>

➔ TRIPS : un futur microprocesseur ?

Des chercheurs de l'Université du Texas à Austin, ont conçu un microprocesseur de haute performance. En collaboration avec IBM, ils construisent un prototype basé sur cette architecture appelée TRIPS (Tera-Op Reliable Intelligently Adaptive Processing System). TRIPS utilise une nouvelle approche appelée "polymorphisme" qui permet une très grande souplesse d'adaptation à différents types de logiciels. Ces travaux sont financés par la DARPA avec un budget de 11 millions de dollars.

➔ Apple

Virginia Tech vient de construire un cluster Infini-Band de 1100 ordinateurs Apple Power Mac G5 à 64 bits comprenant chacun deux processeurs IBM PowerPC à 2 GHz. Le système d'exploitation est Mac OS X. Sa puissance théorique : plus de 10 Tflops. Le coût ? un peu plus de 5 millions de dollars !

➔ Bull

- Bull et Quadrics ont annoncé des grappes de serveurs associant la puissance des nouveaux serveurs Bull NovaScale à la performance des solutions d'interconnexion de Quadrics. L'objectif de Bull est de fournir aux centres de calcul demandant des puissances de plusieurs Tflops des solutions associant, grâce au réseau d'interconnexion QsNet, jusqu'à plusieurs centaines de grands serveurs.
- De nouvelles solutions de stockage en réseau dédiées au calcul de haute performance. S'appuyant sur les technologies de DataDirect Networks et destinées aux grands centres de calcul, elles ont vocation à être connectées aux grandes configurations de type grappes incluant de plusieurs dizaines à plusieurs centaines de serveurs NovaScale
- Le centre océanographique de l'Université de Southampton a commandé un système comprenant un serveur NovaScale 5160 à 16 voies et 6 serveurs NovaScale 4040 de 4 voies.

<http://www.soc.soton.ac.uk>

- L'Université de Stuttgart a également installé un serveur NovaScale, la possibilité de fonctionner simultanément avec plusieurs environnements système ayant été particulièrement appréciée.

➔ Cray

Boeing a passé commande d'un Cray X1 (800 Gflops) destiné à des applications de calcul de structure et de mécanique des fluides.

➔ Dell

Le NCSA (National Center for Supercomputing Applications), à l'Université de l'Illinois, acquiert un cluster de 1450 serveurs Dell PowerEdge (processeurs Intel Xeon). Sa puissance théorique serait proche de 10 Tflops.

➔ IBM

Le Japan National Institute of Advanced Industrial Science and Technology a commandé un système IBM comprenant 2636 processeurs dont 1058 eServers 325 (processeurs Opteron). Fonctionnant sous Linux, sa performance crête sera supérieure à 11 Tflops.

➔ NEC

- NEC a commencé la commercialisation de ses nouvelles séries TX7 et NX7700
- Un système SX-6 a été installé à l'université de Dresden (Allemagne). Il est destiné à des applications de bioinformatique
- La météo australienne a commandé un SX/6 d'une performance crête supérieure à 1 Tflops.

➔ SGI

- Le laboratoire national d'Oak Ridge a reçu un système SGI Altix 3000 de 256 processeurs. La puissance crête est de 1,5 Tflops.
- Le cluster SGI Altix 3700 de 416 processeurs Itanium-2 à 1,3 GHz du centre de calcul d'Amsterdam (SARA) est maintenant complet et ouvert aux utilisateurs. Sa puissance crête est de 2,2 Tflops.

<http://www.sara.nl>

➔ OSCAR : la version 3.0 est disponible

OSCAR (Open Source Cluster Application Resources) est une bibliothèque logicielle "Open Source" destinée à simplifier la configuration, l'installation, la gestion des clusters. La version 3.0 fonctionne avec les processeurs IA32 sous Linux (Red Hat et Mandrake). Le support sur les processeurs Itanium devrait être disponible rapidement.

<http://oscar.openclustergroup.org/>

Agenda

- 19 janvier : **Globus World2** (San Francisco, Ca, Etats-Unis)
- 28 au 30 janvier : **AxGrids 2004** : The 2nd European across Grids Conference (Nicosie, Chypre)
- 29 au 30 janvier : **HealthGrid** : Second European HealthGrid Conference (Clermont-Ferrand)
- 11 au 13 février : 17 au 19 février : **PDP 2004** : 12th Euromicro Conference on Parallel, Distributed and Network-based Processing (A Coruna, Espagne)
- 17 au 19 février : **PDCN 2004** : the IASTED International Conference on Parallel and Distributed Computing and Networks (Innsbrück, Autriche)
- 10 au 12 mars : **ISPASS 2004** : International Symposium on Performance Analysis of Systems and Software (Austin, Tx, Etats-Unis)
- 14 au 17 mars : **PerCom 2004** : Second IEEE International Conference on Pervasive Computing and Communications (Orlando, Fl, Etats-Unis)
- 12 au 15 avril : **GP2PC-04** : Fourth International Scientific Workshop on Global and Peer-to-Peer Computing (Chicago, Il, Etats-Unis)
- 14 au 16 avril : **CF'04** : 2004 ACM International Conference on Computing Frontiers (Ischia, Italie)
- 19 au 22 avril : **CCGGrid 2004** : 4th International Symposium on Cluster Computing and the Grid (Chicago, Il, Etats-Unis)
- 19 au 24 avril : **CHT-04** : International Symposium on Advances in Computational Heat Transfer (Norvège)
- 26 au 30 avril : **IPDPS 2004** : International Parallel and Distributed Processing Symposium (Santa Fe, NM, Etats-Unis)
- 26 au 30 avril : **PMEO-PDS 2004** : 3rd International Workshop on Performance Modeling, Evaluation, and Optimization of Parallel and Distributed Systems (Santa Fe, NM, Etats-Unis)
- 26 avril : **HiCOMB 2004** : 3rd International Workshop on High Performance Computational Biology (Santa Fe, NM, Etats-Unis)
- 26 avril : **HIPS 2004** : 9th International Workshop on High-Level Parallel Programming Models and Supportive Environments (Santa Fe, NM, Etats-Unis)
- 26 au 27 avril : **RAW 2004** : The 11th Reconfigurable Architectures Workshop (Santa Fe, NM, Etats-Unis)
- 26 au 27 avril : **PACGrid-04** : The First Workshop on Partitioning Applications for Computational Grids (Santa Fe, NM, Etats-Unis)
- 26 au 30 avril : **WMPP** : Fourth Workshop on Massively Parallel Processing (Santa Fe, NM, Etats-Unis)
- 10 au 12 mai : **I-SPAN 2004** : 7th International Symposium on Parallel Architectures, Algorithms and Networks (Hong Kong)
- 16 au 19 mai : **PADS 2004** : 18th Workshop on Parallel and Distributed Simulation Issues (Kufstein, Autriche)
- 17 au 20 mai : **LCI-04** : The 5th International Conference on Linux Clusters : the HPC Revolution 2004 (Austin, Tx, Etats-Unis)
- 24 au 26 mai : **GT04** : Grid Today 2004 (Philadelphie, Pa, Etats-Unis)
- 24 au 28 mai : Ecole **Druide 2004** : Distribution de données à grande échelle (Le Croisic)
- 4 au 6 juin : **HPDC 13** : 13th International Symposium on High Performance Distributed Computing (Honolulu, Hi, Etats-Unis)

Des informations complémentaires, en particulier les adresses http de ces manifestations, sont disponibles sur le serveur Web d'ORAP.



HOISE - *Europe On-line Information Service*

PRIMEUR ! - *Advancing European Technology Frontiers*

<http://www.hoise.com/primeur/>

**Organisation Associative du Parallélisme
Structure de collaboration créée par
le CEA, le CNRS et l'INRIA.**

Secrétariat : chantal.letonqueze@irisa.fr

IRISA, campus de Beaulieu, 35042 Rennes cedex

Tél : 02.99.84.75.33, Fax : 02.99.84.74.99

<http://www.irisa.fr/orap>