

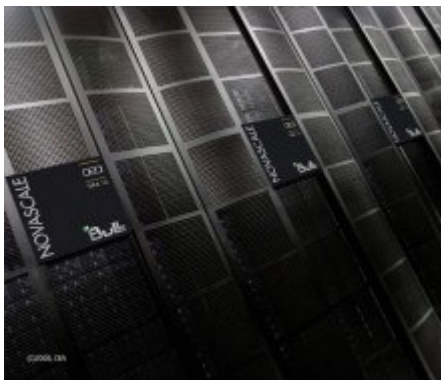
Bull va construire le premier supercalculateur PetaFlops français pour le CEA

Par Christophe Bardy Le 30 juillet 2008 (19:30)

Rubriques : HPC Tags : supercalculateurs - bull - cea

A la suite d'un appel à candidature lancé le 29 janvier dernier, la direction des applications militaires du CEA a finalement retenu Bull pour élaborer son premier supercalculateur Petaflops. Attendu en 2010, la machine fera l'objet d'un programme de R&D commun entre le CEA et Bull avec à la clé un accord de partage de la propriété intellectuelle. Une condition qui aurait effrayé plus d'un constructeur américain.

Bull va construire le premier supercalculateur PetaFlops français pour le CEA



Le CEA et Bull ont annoncé aujourd'hui avoir signé un contrat de collaboration pour la conception et l'acquisition d'un supercalculateur pétaflopique. Le contrat, noué entre la Direction des applications militaires du CEA (CEA-DAM) et le constructeur français porte sur la réalisation de Tera 100, la troisième génération de supercalculateur Tera destiné au Programme de simulation français (après Tera1 et Tera 10).

Tera 100 devrait être assemblé d'ici à l'été 2010 et entrera officiellement en production à la fin 2010. D'ici là, Bull et les ingénieurs du CEA vont collaborer au sein d'un laboratoire commun pour développer les technologies nécessaires à la construction du supercalculateur. Cette phase de R&D, une première du genre en Europe, verra le CEA et Bull partager la propriété intellectuelle produite lors du programme. Cette clause est d'ailleurs vraisemblablement l'une des raisons qui a poussé certains constructeurs américains à ne pas participer à l'appel à candidatures. Plusieurs centaines d'ingénieurs et de chercheurs de très haut niveau devraient être mobilisés pour ce projet.

Dans leur communiqué les deux partenaires indiquent que les technologies pétaflopiques sont un enjeu majeur aussi bien pour la recherche universitaire que pour l'industrie et pour l'emploi. "La simulation numérique Haute Performance est devenu incontournable pour la modélisation et la simulation, notamment dans l'aéronautique, l'énergie, la climatologie, les sciences de la vie, la finance, le traitement de l'information, et également pour le développement durable et les économies d'énergie. Le Calcul Haute Performance est devenu un moyen d'investigation et de simulation indispensable, un atout majeur pour la compétitivité de la recherche et de l'industrie, enfin un élément fondamental de la souveraineté des Etats" explique ainsi le communiqué.

C'est aussi l'avis du MagIT mais avec une petite pincée de sel. Si un tel investissement est aussi précieux pour toutes ces industries, on aurait aimé que le premier supercalculateur de ce type ne soit pas cantonné aux seuls usages militaires. Les chercheurs "civils" pourraient toutefois avoir rapidement de quoi se consoler : Tera 100 pourrait être rapidement suivi d'un autre supercalculateur de classe Petaflops dans le cadre du programme européen Prace (Partnership for Advanced Computing in Europe).

Plus de 100 000 cœurs x64, 300 To de mémoire et 20 petaoctets de stockage



En principe la machine Tera 100 se présentera sous la

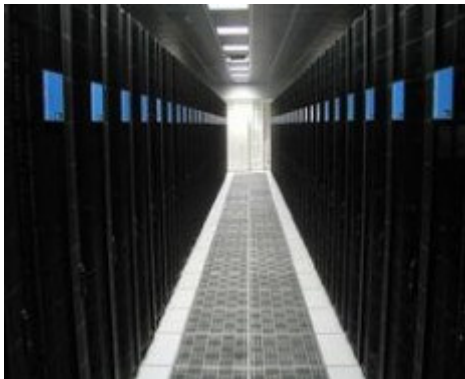
forme d'un cluster composé de près de 100 000 cœurs processeurs, sans doute des puces Xeon de génération "WestMere", une évolution des futures puces Nehalem gravées en technologie 32 nm. Chacune de ces puces devrait intégrer 6 cœurs processeurs. Tirant les leçons de son expérience en matière de calculateur, le CEA s'est fixé pour contrainte de limiter le nombre de nœuds dans le cluster. Dans le cadre de Tera 100, le nombre de nœuds serveurs devrait être d'environ 3000, ce qui suppose que nombre de nœuds intégreront 8 processeurs. Au total, la machine disposera de 300 To de mémoire et s'appuiera sur la technologie Infiniband pour les interconnexions (dans une architecture fat tree amaigrie).

Les entrées/sorties ne devraient pas être en reste au vu des volumes de données à traiter par le supercalculateur. 60 des 3000 serveurs du cluster seront ainsi dédiés à la gestion des seules entrées/sorties. Au total le supercalculateur disposera de 20 Petaoctets de stockage (géré par Lustre). 5 Po seront réservés au seul usage de la machine (stockage privé) tandis que 15 Po seront aussi accessibles par les utilisateurs. Le débit spécifié par le CEA entre le supercalculateur et sa composante stockage est de 300 Go/s, un débit qui devrait nécessiter la mise en service en parallèle de près de 10 000 disques durs.

La composante stockage du cluster devrait ainsi mobiliser pas moins de 162 nœuds de services, dont 6 nœud d'administration, 6 nœud de métadonnées, 100 nœud pour les objets data server et 50 pour les routeurs. Ce qui veut dire que 5% de la puissance du cluster sera mobilisé pour le seul stockage.

Le supercalculateur devrait occuper 600 à 700 m² dans un nouveau bâtiment dont la construction devrait débuter en janvier 2009. Il devrait consommer environ 5 MW (infrastructure de refroidissement comprise). Les différents nœuds seront enfermés dans des racks refroidis par eau.

Bull va construire le premier supercalculateur PetaFlops français pour le CEA



Des défis technique à la mesure du projet

Comme l'explique Jean Gonnort, le chef de projet simulation numérique au CEA DAM, la conception de Tera 100 pose des problèmes uniques. "Il ne s'agit pas avec cette machine de battre un record. Tera 100 n'est pas une machine de recherche, c'est une machine de production qui par contrat devra au minimum tourner à 90% de sa capacité pendant 95% du temps, sous peine de pénalités." Pour atteindre un tel objectif, le CEA et Bull doivent donc résoudre des problèmes nouveaux.

" Tera 100 est la 3e machine du programme Tera" explique ainsi Jean Gonnord. "Pour la première, on ne savait pas si on arriverait au Tflops. Ça a été fait en 2001 en utilisant le parallélisme massif. Nous avons alors buté sur la question des débits et des quantités de données. Aucun OS ne savait gérer avec la fiabilité nécessaire les volumes de données générés. On a donc du travailler en R&D avec les laboratoires américains Lawrence Livermore pour spécifier Lustre. Lustre est aujourd'hui dans les mains de Sun mais reste ouvert et les spécifications et évolutions sont faites par un board dans lequel on a un poids non négligeable. Pour Tera 10, il nous a fallu développer des architectures d'entrée/sorties optimales pour Lustre. Bull a fait des progrès non négligeables en la matière avec Tera10. Ce supercalculateur a établi un certain nombre de records d'entrées/sorties. Aujourd'hui c'est un problème que nous considérons comme maîtrisé."

Tera 100 pose toutefois une autre classe de problèmes à commencer par celui de la consommation énergétique. " Certes les futures puces multi-cœurs offrent une perspective de réduction de la consommation, mais les travaux des fabricants de processeurs ne sont pas suffisant" explique Jean Gonnort. Il s'agit donc de mener une chasse au gaspi sur l'ensemble des équipements composant la

machine. *"Un autre problème pour nous est lié à l'augmentation de la complexité au niveau hardware : Tera 100 aura trois fois plus de composants que tera10, ce qui pose des problèmes de complexité et de fiabilité non négligeables pour une machine de production qui doit travailler presque tous les jours sans interruption"*. Pour résoudre ce problème CEA et Bull vont travailler en R&D sur des architectures logicielles tolérantes aux pannes. Comme les précédents cluster Tera, ces architectures s'appuieront sur des systèmes d'exploitation libre (jusqu'alors, les cluster Tera s'appuient sur Red hat Linux).

Une première pour la R&D européenne

La collaboration R&D à cette échelle est d'ailleurs une première en Europe, la France reprenant là les principes qui ont fait le succès de l'industrie de l'informatique à haute performance aux Etats-Unis qui profite à plein des subventions et autres "grants" de la NSF et du département à l'énergie. Selon Jean Gonnord, " C'est une première de lancer un programme de R&D en parallèle d'un cluster. Nous n'achetons pas cette machine sur étagère. Cette fois le challenge est suffisamment fort pour nécessiter une R&D importante impliquant nos équipes et celles d'un constructeur".

Pour ce programme, le CEA avait fixé pour contrainte que la R&D soit menée en Europe et que la propriété intellectuelle générée par le programme soit partagée. Ces conditions ont, semble-t-il, refroidi l'enthousiasme des constructeurs américains, dont la plupart mènent leurs recherches sur le calcul à haute performance aux Etats-Unis, souvent dans le cadre de programmes de recherche fédéraux américains. Ce principe ne devrait d'ailleurs pas rester exclusif au CEA, puisque le programme européen Prace prévoit lui aussi de tels partenariats de recherche.

Reste que côté CEA, on se félicite à être les premiers à avoir noué un tel accord et qu'on se réjouit surtout d'avoir réussi à mettre en place un programme de R&D, qui permettra à des chercheurs européens de haut niveau de mener des recherches sur le calcul scientifique sur le territoire européen. La collaboration avec Bull devrait ainsi assurer que l'Europe disposera d'au moins une entreprise capable de répondre à la demande HPC très haut niveau.