



Biogemma accélère la recherche génomique avec Lenovo

Lenovo™



Fondée en 1997 par des semenciers et des producteurs de grandes cultures, Biogemma est une société de biotechnologie végétale. L'entreprise s'est donné pour mission de découvrir ou d'acquérir des gènes et des technologies qui contrôlent l'expression de caractéristiques de cultures présentant une valeur commerciale, telles que la tolérance à la sécheresse, l'efficacité d'utilisation d'azote, la résistance aux maladies et le rendement élevé.

Olivier Dugas, directeur général de Biogemma, explique pourquoi la société doit disposer de ressources de calcul haute performance (HPC) : « Nos recherches en bioinformatique génèrent et traitent d'énormes volumes de données génétiques, ce qui se traduit par une augmentation constante de nos besoins en calcul. Les génomes que nous analysons sont complexes et souvent polyploïdes, ce qui peut les rendre beaucoup plus vastes que le génome humain. Par ailleurs, nous axons davantage nos efforts autour de la caractérisation du matériel génétique, qui consiste à comparer un génome de référence avec un grand nombre de points de données de séquençage, ce qui implique des calculs parallèles entre plusieurs cœurs simultanément. »

Synthèse

L'entreprise de biotechnologie Biogemma avait besoin d'augmenter sa puissance de calcul afin de traiter plus rapidement l'analyse de jeux de données génomiques plus importants et plus complexes. En déployant un cluster HPC basé sur Lenovo NeXtScale nx360 M5 et un système à grande capacité de mémoire basé sur un serveur Lenovo System x3850 X6, l'entreprise a considérablement gagné en performances : les temps d'exécution sont généralement réduits à quelques minutes contre plusieurs heures auparavant, soit jusqu'à 99 % d'amélioration, et l'accès aux ressources informatiques s'effectue désormais sans aucune file d'attente.

« Nous avons constaté une très nette augmentation de la vitesse de calcul avec notre cluster Lenovo NeXtScale System. »

— Nathalie Rivière,
Coordonnatrice de génomique en amont /
Manager bioinformatique,
Biogemma



Un défi croissant

Comprendre la manière dont les génomes de récolte sont exprimés en caractéristiques physiques, telles que la résistance aux maladies, se révèle extrêmement précieux, autant sur le plan sociétal que financier. Les travaux de Biogemma contribueront à assurer des récoltes plus résistantes, à haut rendement, pour nourrir une population mondiale de plus en plus nombreuse. Pour déterminer quelles sont les séquences génétiques qui produisent les caractéristiques souhaitées, l'entreprise doit analyser et croiser d'énormes volumes de données de séquençage provenant de génomes complexes.

« Dans l'assemblage du génome de novo, nous pouvons être amenés à comparer un million de séquences avec une autre million de séquences, ce qui exige un système possédant une énorme capacité de mémoire, » commente Nathalie Rivière, coordonnatrice de génomique en amont / Manager bioinformatique chez Biogemma. « La caractérisation du matériel génétique relève d'une autre catégorie de problèmes. En l'occurrence, nous devons mettre en parallèle des calculs statistiques pour obtenir des résultats plus rapides, ce qui exige un cluster de calcul. »

Depuis le déploiement de son précédent cluster de 200 cœurs, Biogemma avait constaté une augmentation considérable des charges de travail bioinformatiques : le nombre d'échantillons dans une analyse type était alors passé de 10 à 400. Sur le plan statistique, le cluster précédent ne pouvait faire face à la croissance en termes à la fois du nombre d'analyses et de la complexité des modèles.

Cultiver le changement

Après avoir réalisé un inventaire des besoins internes, Biogemma s'est adressé à plusieurs grands fournisseurs HPC afin d'obtenir des recommandations en matière d'architecture. « Nous n'avons pas donné de cahier des charges précis, » se souvient Nathalie Rivière. « À la place, nous avons expliqué nos exigences en termes de calculs à grande mémoire et parallélisés, et avons fourni un ensemble de cas d'utilisation représentatifs à des fins d'analyse comparative. »

Elle ajoute : « Nous avons choisi la solution proposée par Lenovo et son partenaire, Serveware, d'une part pour son rapport qualité-prix avantageux et d'autre part pour l'expertise que les deux prestataires ont démontrée lors du processus d'analyse comparative, ce qui nous a aidés à optimiser la solution et à adapter nos applications. Leur expertise nous a également permis d'améliorer considérablement la planification et le niveau de priorité des tâches de calcul. »

Pour traiter les charges de travail bioinformatiques extrêmement gourmandes en capacité de mémoire par cœur, Biogemma a déployé un serveur Lenovo System x3850 X6 équipé de processeurs Intel® Xeon® E7 intégrant 48 cœurs et 1,5 To de mémoire. Pour les charges de travail statistiques hautement parallélisées, Biogemma a déployé un cluster Lenovo NeXtScale nx360 M5 à 36 nœuds, chacun doté de deux processeurs Intel® Xeon® E5 à 12 cœurs, soit un total de 864 cœurs et 34,5 Tflops de performances. Les interconnexions InfiniBand à haut débit relient les

Les composants de la solution

Matériel

Lenovo NeXtScale nx360 M5

Lenovo System x3850 X6

IBM Storwize V3700

Logiciel

IBM Spectrum Scale (IBM GPFS)



nœuds et les connectent à un serveur de stockage IBM Storwize V3700 d'une capacité de 120 To. Dans cet environnement, l'accès aux fichiers parallèle à haute performance est assuré par IBM Spectrum Scale (IBM GPFS).

Une transformation qui porte ses fruits

Le cluster de Lenovo et l'environnement à grande capacité de mémoire ont permis à Biogemma de gagner considérablement en efficacité, et d'aider l'entreprise à répondre à la forte hausse des besoins internes en termes de quantité et de complexité des analyses informatiques. « Nous avons constaté une très nette augmentation de la vitesse de calcul avec notre cluster Lenovo NeXtScale System, qui fournit beaucoup plus de cœurs s'exécutant à des vitesses d'horloge plus rapides, » précise Nathalie Rivière. « Un travail qui s'exécutait généralement en une journée s'effectue désormais en quelques heures, voire quelques minutes, soit jusqu'à 99 % plus vite ! »

Ces performances accrues contribuent également à améliorer la productivité globale de Biogemma en matière de recherche, comme l'explique Nathalie Rivière : « Auparavant, lorsque les ressources étaient limitées, nous étions obligés de planifier nos sessions de calcul, en mettant les projets en file d'attente. Aujourd'hui, nous pouvons lancer les analyses dès que les données sont disponibles, travailler plus efficacement et obtenir les résultats plus rapidement. Nous sommes également en mesure de traiter beaucoup plus d'analyses simultanées, et d'entreprendre des études bien plus ambitieuses. C'est pourquoi, grâce à notre solution HPC Lenovo, Biogemma dispose des outils nécessaires pour conserver sa place au premier rang de l'industrie biotechnologique mondiale. »

Pour plus d'information

Pour en savoir plus sur Lenovo Enterprise Systems, contactez votre représentant commercial Lenovo ou votre partenaire commercial Lenovo ou rendez-vous sur : lenovo.com/systems

Pour en savoir plus sur Biogemma, rendez-vous sur : www.biogemma.com



© 2016 Lenovo. Tous droits réservés.

Disponibilité : les offres, les prix, les spécifications et la disponibilité des produits peuvent être modifiés sans préavis. Lenovo n'est pas responsable des erreurs photographiques ou typographiques. **Garantie** : Pour obtenir un exemplaire des garanties applicables, écrivez à : Warranty Information, 500 Park Offices Drive, RTP, NC, 27709, Attn: Dept. ZPYA/B600. Lenovo ne fait aucune déclaration et n'offre aucune garantie quant aux produits ou services de tiers. **Marques** : Lenovo, le logo Lenovo, NeXtScale, et System x sont des marques commerciales ou déposées de Lenovo. Intel, le logo Intel, Xeon et Xeon Inside sont des marques déposées d'Intel Corporation aux Etats-Unis et dans d'autres pays. D'autres noms de sociétés, de produits et de services mentionnés peuvent être des marques commerciales ou des désignations de services peuvent être des marques de services ou appartenir à des tiers. Visitez régulièrement <http://www.lenovo.com/lenovo/us/en/safecom.html> pour vous tenir informé des bonnes pratiques de sécurité informatique.



« Aujourd'hui, nous pouvons lancer les analyses dès que les données sont disponibles, travailler plus efficacement et obtenir les résultats plus rapidement. »

— Nathalie Rivière,
Coordonnatrice de génomique en amont /
Manager bioinformatique,
Biogemma