

MINES ParisTech analyse des données complexes pour comprendre et améliorer les matériaux

Une grande école d'ingénieurs française mène des recherches de pointe dans l'industrie grâce à l'ActiveStor® de Panasas

RÉSUMÉ

CLIENT

MINES ParisTech

SIZE

2 000 personnes en tout, dont 200 au Centre de recherche des matériaux

ADRESSE

Evry, France

SECTEUR

Enseignement supérieur et recherche académique

CHALLENGE

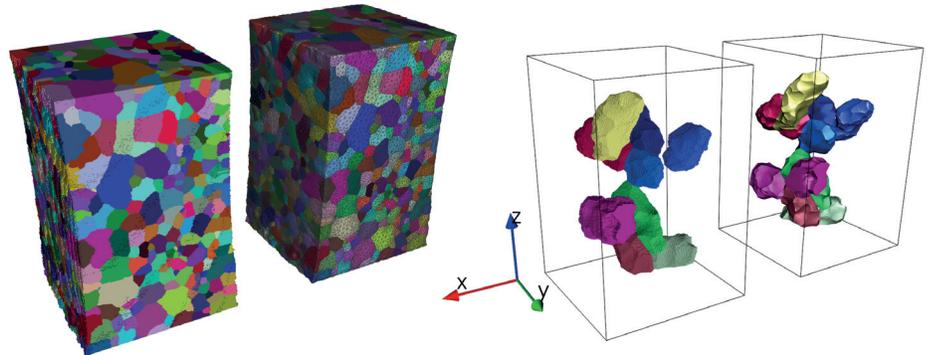
- Collecter, analyser et stocker d'énormes volumes de données issues d'observations expérimentales
- Faciliter l'analyse et la simulation de données à grande échelle, grâce à une solution de stockage unifiée
- Améliorer la productivité des chercheurs en éliminant les goulets d'étranglement dans l'accès aux données

SOLUTION

- Assure des performances constantes et un accès rapide aux données, même en cas de charges de travail très lourdes, simultanées et mixtes
- Favorise les techniques de recherche nécessitant beaucoup de données, comme la tomographie, la corrélation d'images en 3D, et les études statistiques en 4D
- Echelonnable pour répondre aux besoins d'une communauté croissante de chercheurs, avec des volumes de données en hausse
- Evite d'avoir à déplacer ou copier les fichiers, ce qui simplifie l'accès aux données

RÉSULTATS

- Vitesse d'accès aux données augmentée d'un facteur 4 à 5, améliorant ainsi la productivité des chercheurs
- Gains de performances du système grâce à une distribution plus équilibrée des données
- Élimination des arrêts système non-planifiés et amélioration de la disponibilité et de la résilience
- Allègement des efforts d'administration et de maintenance du système
- Besoins de stockage immédiats satisfaits et possibilités d'évolution en prime



Accès 4 à 5 fois plus rapide aux données pour le calcul haute performance

« Grâce à l'ActiveStor de Panasas, les chercheurs gagnent du temps puisqu'ils n'ont plus à déplacer ni à copier les fichiers de données avant de pouvoir les traiter. Ils gagnent en productivité puisqu'il n'y a plus de goulets d'étranglement au niveau HPC pour accéder au système de stockage. La bande passante est quatre à cinq fois plus élevée qu'auparavant. Et le partage de fichiers offre plus de confort, grâce à des temps de réponse plus courts et à un débit plus élevé. »

Grégory Sainte-Luce

Responsable Informatique du Centre de recherche sur les matériaux et membre de la Direction Informatique de MINES ParisTech

MINES ParisTech, l'une des plus prestigieuses écoles d'ingénieurs françaises, travaille avec des entreprises et des universités du monde entier. L'institution publique détient plus de 30 millions d'euros de contrats de recherche annuels.

L'augmentation du nombre de partenariats est essentielle pour MINES ParisTech, car ces contrats permettent non seulement d'accroître les efforts de recherche de l'école, mais aussi de faire des découvertes créatives qui améliorent la vie des gens dans le monde entier.

Les chercheurs s'appuient sur des techniques d'observation expérimentales à base d'imagerie 2D et d'imagerie 3D non-destructive à haut débit. Ces technologies génèrent d'énormes volumes de données, que les chercheurs collectent et analysent quotidiennement à l'aide d'outils sophistiqués, comme l'analyse d'images et la simulation mécanique.

Pour traiter les données, les chercheurs devaient les déplacer à partir de plusieurs systèmes de stockage, ce qui nuisait à la productivité et entraînait souvent des duplications inutiles. La gestion et la maintenance de multiples systèmes de stockage s'avéraient également fastidieuses pour le service informatique.

Pour accueillir de nouveaux outils tels que la tomographie et la corrélation d'images 3D, le Centre avait besoin d'une solution de stockage unifiée, capable de fournir d'excellentes performances, tout en permettant une croissance continue du volume de données. L'autre priorité était d'offrir une expérience utilisateur positive à tous – y compris aux 70 chercheurs utilisateurs de calcul haute performance (HPC) et aux 130 utilisateurs d'applications de partage de fichiers du système.

MINES ParisTech a évalué plusieurs solutions de stockage de fichiers parallèles et de stockage d'objets avant de choisir l'ActiveStor de Panasas, un équipement HPC plug-and-play sur lequel tourne le système de fichiers parallèle PanFS.

« Nous avons choisi ActiveStor pour sa combinaison unique de haute performance, d'échelonnabilité et de simplicité d'administration, » explique Gregory Sainte-Luce, Responsable Informatique du Centre de recherche sur les matériaux et membre du conseil d'administration informatique de MINES ParisTech.

Une technologie de stockage réactive et échelonnée

Installé au Centre de recherche des matériaux, le système ActiveStor fonctionne 24 heures sur 24 et 7 jours sur 7, et offre aux utilisateurs un accès parallèle réactif au stockage, avec des performances élevées et constantes, même en cas de charges de travail lourdes, simultanées et mixtes.

Les tâches HPC du Centre varient de quelques heures à plusieurs jours, 24 heures sur 24. La tâche HPC la plus gourmande en ressources est l'analyse de données de tomographie, qui peut durer plusieurs jours. L'exécution d'une seule tâche de tomographie peut mobiliser de 10 à 100 Go de stockage, et le Centre en exécute des centaines dans le cadre de ses expériences. Les simulations par éléments finis à grande échelle du comportement des matériaux nécessitent jusqu'à 1 téraoctet (To) de stockage par simulation. Le Centre doit s'assurer que les chercheurs ont un accès rapide aux données.

La solution Panasas a satisfait et dépassé les attentes des utilisateurs et du service informatique. « Les chercheurs gagnent du temps puisqu'ils n'ont plus à déplacer ni à copier les fichiers de données avant de pouvoir les traiter, » selon M. Sainte-Luce. « Ils gagnent en productivité puisqu'il n'y a plus de goulets d'étranglement au niveau PC pour accéder au système de stockage. La bande passante est quatre à cinq fois plus élevée qu'auparavant. Et le partage de fichiers offre plus de confort, grâce à des temps de réponse plus courts et à un débit plus élevé. »

Performances élevées et constantes

Avec ActiveStor, les utilisateurs ont désormais un accès rapide et simultané au stockage, même en cas de charges de travail sont mixtes ou lorsque le taux d'utilisation est très élevé.

Le Centre mène des études statistiques 4D, qui évaluent les contraintes mécaniques qui se produisent au cours du

temps. Le processus d'essai 4D génère automatiquement un jumeau numérique de l'échantillon de matériau, ce qui permet aux analystes de simuler les performances et de valider les modèles de comportement utilisés par leurs partenaires industriels.

« L'analyse d'essais 4D faisant appel à de très gros volumes de données peut prendre plusieurs mois, voire un an, » explique Henry Proudhon, chercheur au Centre. « Nous traitons en général 1 To de données pour chaque jour de l'expérience, avec 15 à 20 jours d'expériences par an. Une étude moyenne peut nécessiter une centaine de tâches HPC qui durent en moyenne six heures par jour. »

En tant qu'utilisateur, M. Proudhon apprécie la façon dont ActiveStor améliore la disponibilité des données HPC, et accélère ainsi l'obtention des résultats. « Panasas permet de traiter de gros volumes que l'on peut augmenter ou réduire rapidement, en fonction des besoins, » explique M. Sainte-Luce. « Nous fonctionnons actuellement à 80% de notre capacité de stockage et notre niveau de charge est de 17 à 20%, selon l'outil de suivi de Panasas. L'accès aux données est cinq fois plus rapide pour le stockage HPC, et nous avons même constaté une amélioration des performances et du confort pour les autres utilisateurs qui n'utilisent pas de HPC. Nous n'avons pratiquement plus de données dupliquées. ActiveStor gère toute la charge de travail du laboratoire de recherche sans rencontrer le moindre problème de performance. »

De plus, la centralisation des serveurs et des matrices de stockage au sein d'une même solution convergente a permis au Centre de réduire les besoins des utilisateurs en matière de transfert de données entre des dispositifs de stockage distincts.

Gestion simplifiée

ActiveStor a considérablement réduit les besoins du Centre en matière de gestion du stockage. La simplicité de gestion de la solution a permis d'éliminer de gros efforts d'ingénierie, tout en réduisant le temps d'administration global. L'agrégation automatique des ressources de stockage libère les administrateurs de la nécessité de configurer des règles hiérarchiques complexes de gestion du stockage.

L'interface graphique intuitive à base web permet à un administrateur unique de gérer et de maintenir les quotas et les volumes en consacrant à cette tâche seulement une quarantaine d'heures par an. ActiveStor envoie automatiquement

des avertissements de quota aux administrateurs par e-mail, ce qui leur permet d'ajuster rapidement et dynamiquement l'espace disque en fonction des besoins, sans aucune interruption de service.

« L'ajout d'un nouveau rack est un processus sans risque et complètement transparent, qui prend moins d'une heure, » explique M. Sainte-Luce. « ActiveStor rééquilibre automatiquement les données stockées entre tous les serveurs de stockage du nouveau système, sans aucune dégradation des performances. La distribution plus équilibrée des données a considérablement amélioré les performances de notre système. »

Stockage plus fiable et moins cher

Depuis le déploiement d'ActiveStor, les utilisateurs n'ont connu aucun temps d'arrêt du système. Plusieurs serveurs de gestion de métadonnées et de systèmes de fichiers assurent la redondance, ce qui augmente la disponibilité et la résilience.

« Les administrateurs passent moins de temps à gérer le stockage, ce qui leur permet de se concentrer sur d'autres activités, » explique M. Sainte-Luce. « Dans l'ensemble, le fonctionnement sans faille d'ActiveStor nous a permis de réduire nos coûts de gestion du stockage. »

Flexibilité pour soutenir la croissance

ActiveStor offre à MINES ParisTech l'échelonnabilité nécessaire pour répondre à ses besoins de stockage présents et à venir. Panasas gère la charge de travail actuelle de l'ensemble du groupe de recherche, et peut facilement s'adapter à l'évolution des besoins, en augmentant ou en réduisant les quotas des volumes physiques ou logiques « L'échelonnabilité est l'un des principaux avantages de la solution Panasas, » déclare M. Sainte-Luce. « C'est une technologie que l'on peut échelonner pratiquement sans effort pour répondre à la demande. »

ActiveStor permet également aux chercheurs d'accomplir davantage. « Lorsque les chercheurs peuvent accéder plus rapidement à leurs données scientifiques, atteindre un débit de travail plus élevé et effectuer davantage de calculs, ils peuvent accomplir davantage de tâches, » explique M. Sainte-Luce. « Leur productivité nous permet de vendre plus de contrats à nos partenaires et de développer notre activité. Nous sommes particulièrement ravis d'avoir Panasas comme partenaire. »