

UNIVERSITÉ DE VIENNE – UN CLUSTER HIGH PERFORMANCE DE TRANSTEC SERT DE PILIER POUR LA RECHERCHE BIOLOGIQUE

Traitement et dépouillement d'énormes volumes de données au service de la science

« transtec nous a livré un équipement IT de base pour notre nouveau département à l'université de Vienne. Pour un prix modéré, cet équipement nous fournit une performance optimale ainsi que la capacité dont nous avons besoin pour les travaux de recherche et l'enseignement dans la bioinformatique. Nous pouvons ainsi réaliser également de très vastes projets de coopération avec nos partenaires nationaux et internationaux, sans atteindre les limites techniques. En raison du prix raisonnable, une partie du budget consacré aux investissements reste disponible pour les extensions futures du système.

Rétrospectivement, nous confirmons d'une part la fiabilité absolue de la solution et attestons d'autre part que toutes les déclarations faites par transtec sont pertinentes aujourd'hui encore. Une preuve manifeste donc de l'extraordinaire compétence de transtec en matière de conseil compte tenu des conditions générales vraiment difficiles ».

PROF. DR. THOMAS RATTEI, DIRECTEUR DU DÉPARTEMENT POUR COMPUTATIONAL SYSTEMS BIOLOGY DE LA FACULTÉ DES SCIENCES DE LA VIE DE L'UNIVERSITÉ DE VIENNE

L'ENTREPRISE

Le Department pour Computational Systems Biology est l'un des vingt-cinq départements que compte la faculté des sciences de la vie de l'université de Vienne. Les sciences de la vie se caractérisent par une grande diversité de disciplines. Elles se consacrent à tous les aspects de la vie, son développement, son évolution, sa diversité et son interaction avec l'environnement. L'approche systématique élémentaire pour les sciences de la vie revêt une importance centrale. Les ouvertures sont très diversifiées. Elles s'étendent d'une approche génétique et gnomique (fonctionnelle) à une approche analytique, physiologique et écologique marquée et englobent également la théorisation, l'informatique et les biomathématiques. Les biomathématiques sont une discipline scientifique évoluant rapidement qui combine les méthodes de la biologie, des mathématiques et de l'informatique. Elles sont essentielles pour le traitement et le dépouillement d'énormes volumes de données qui sont générés par la recherche biologique moderne. A lui seul, le lot de séquences ADN publiées double presque chaque année. Entre le fait de recueillir les données et de pouvoir les comprendre et les utiliser pour tous les domaines des sciences de la vie, il se passe une longue période que seuls des outils logiciels spéciaux et des méthodes mathématiques peuvent dominer.

Le Department pour Computational Systems Biology créé récemment est l'un des vingt-cinq départements que compte la faculté des sciences de la vie de l'université de Vienne. Le département à pour

tâche principale notamment de recueillir et de dépouiller des données biologiques provenant de projets de recherche les plus divers.

EQUIPEMENT INITIAL À HAUTE DISPONIBILITÉ POUR LE DÉPARTEMENT CRÉÉ RÉCEMMENT

« Les exigences auxquelles était soumise la nouvelle structure du Department for Computational Systems Biology étaient tellement complexes, que notre priorité en début de projet était de développer en concertation avec les fabricants des idées de solutions pour l'équipement initial du département. Ces idées, complétées par les résultats obtenus à l'aide d'une Benchmark-Suite pour applications bioinformatiques typiques, pouvaient alors nous servir de référence afin de lancer l'appel public d'offres concernant une solution HPC. Outre le fait élémentaire de sélectionner des composants appropriés pour l'environnement universitaire, nous prenions également en considération les facteurs prix, performance/capacité, disponibilité, consommation de ressources ainsi que frais de maintenance » explique le Prof. Dr. Thomas Rattei, directeur du Department pour Computational Systems Biology de la faculté des sciences de la vie de l'université de Vienne.

Dans ces deux aspects de l'appel d'offres, transtec s'est rapidement dégagé comme favori. Le fabricant implanté à Tübingen s'est nettement imposé sur la concurrence sur le plan du rapport prix/performance. L'offre répondait également aux conditions générales difficiles, telles par exemple la prise en compte de la charge de plancher la plus grande possible, la puissance absorbée ainsi que les dimensions disponibles pour la salle des serveurs.



« Tous nos projets de recherche se caractérisent par des volumes de données très importants. Notre profil de travail scientifique requiert une solution capable de garantir le stockage absolument sûr et efficace de toutes les données recueillies pour la recherche. La disponibilité permanente de ces données stratégiques ainsi que de toute l'infrastructure IT constituait un élément tout aussi important. La solution de transtec répond avec brio à toutes les exigences ».

PROF. DR. THOMAS RATTEI, DIRECTEUR DU DÉPARTEMENT POUR COMPUTATIONAL SYSTEMS BIOLOGY DE LA FACULTÉ DES SCIENCES DE LA VIE DE L'UNIVERSITÉ DE VIENNE

LE PROFIL DU TRAVAIL SCIENTIFIQUE EXIGE UN MAXIMUM DE PERFORMANCE ET DE DISPONIBILITÉ

La solution High Performance Computing conçue et implémentée par transtec se distingue par un haut degré de performance, de capacité de mémoire ainsi que de disponibilité. Le cluster est basé sur six serveurs transtec CALLEO 341 et huit serveurs transtec CALLEO 642 Twin², ces derniers comprenant respectivement 4 nœuds de calcul. En outre, la capacité de stockage est de 192 téraoctets que trois systèmes RAID PROVIGO 550 mettent à disposition par liaison Fibre Channel.

UN CLUSTER ACTIF-ACTIF GARANTIT UN RENDEMENT ÉLEVÉ ET QUASI-MENT 100% DE SÉCURITÉ FACE À LA PANNE

Une des particularités de la solution réside dans le fait qu'elle a été conçue et construite en tant que cluster actif-actif. Une unité sans Single-Point-of-Failure est constituée respectivement de 2 serveurs NFS et d'une unité de stockage. « Malgré la structure complexe d'un cluster actif-actif, nous nous sommes décidés à bon escient en faveur de cette forme de cluster, afin d'assurer un maximum de rendement et de disponibilité de notre infrastructure. Par une répartition ciblée de la charge (load balancing), nous pouvons influencer au cas par cas la performance du cluster. Ceci est possible en activant simultanément plusieurs unités de cluster et en regroupant des capacités » explique le Prof. Rattei. Avec cette solution, l'université de Vienne est parée au mieux pour faire face à d'éventuelles pannes de serveurs. Les tâches du système touché par une éventuelle défaillance sont prises immédiatement en charge par un autre nœud actif. L'équipement initial du département constitue la base de travail pour toutes les activités du département : recherche, enseignement et administra-

tion. L'implémentation et la mise en service du hardware dont l'investissement se chiffre à environ 200.000 euros sont complètement achevées.

LA SOLUTION

II 32 nœuds cluster, composés de huit serveurs transtec CALLEO 642 Twin² :

cluster 4-in-1, respectivement 2 processeurs Intel® Xeon™ E5620 Quadcore, mémoire principale de 24 gigaoctets, 3 disques durs de 1 téraoctet

II Six serveurs transtec CALLEO 341, utilisés en tant que serveur NFS :

respectivement 2 processeurs Intel® Xeon™ E5620 Quadcore, mémoire principale 24 gigaoctets, 2 disques durs de 0,5 téraoctet, contrôleur 4 Gbit Fibre Channel à deux ports

II Deux PROVIGO 550 RAID plus Extension de transtec :

respectivement 24 disques durs de 2 téraoctets, 4 connexions hôte x 4Gbit FC

II PROVIGO 550 SUMO RAID de transtec :

48 disques durs de 2 téraoctets, 4 connexions hôte x 4Gbit FC

Le reste du réseau compte également onze Thin-Clients ainsi que quelques PC portables (réseau Office). Deux nœuds High Memory (RAM 128 Go), deux serveurs de login, deux serveurs de virtualisation ainsi que quatre serveurs de base de données fonctionnent en outre au sein du Département pour Computational Systems Biology. Tous les serveurs peuvent être gérés au moyen de l'interface IPMI.