



Présentation d'un transfert de mégadonnées de 100 Gb/s – Documentation

Il existe une demande de connexions stables, dédiées et à large bande passante pour l'échange de données au sein de la collectivité canadienne de recherche et d'innovation. Pourtant, l'utilisation efficace d'un réseau à grande vitesse pour les transferts de données sur des réseaux étendus (WAN) représente un défi. C'est pourquoi Calcul Canada a établi un partenariat avec RISQ, CANARIE et ORION ainsi qu'avec des fabricants d'importance pour l'optimisation du premier et du dernier réseau de liaison afin de relier le travail scientifique des chercheurs à la cyberinfrastructure nationale de données.

Chaque installation de Calcul Canada offrira une architecture de réseau scientifique avec zone démilitarisée permettant une connectivité de 100 Gbit/s aux réseaux NREN (Réseau national de recherche et d'éducation) et aux appareils de transferts de données spécialisés aussi appelés « nœuds de transfert des données » (DTN). Un DTN fournit une plateforme optimisée pour les flux de données de grandes capacités et de haute performance qui sont nécessaires au traitement de grands volumes de données scientifiques.

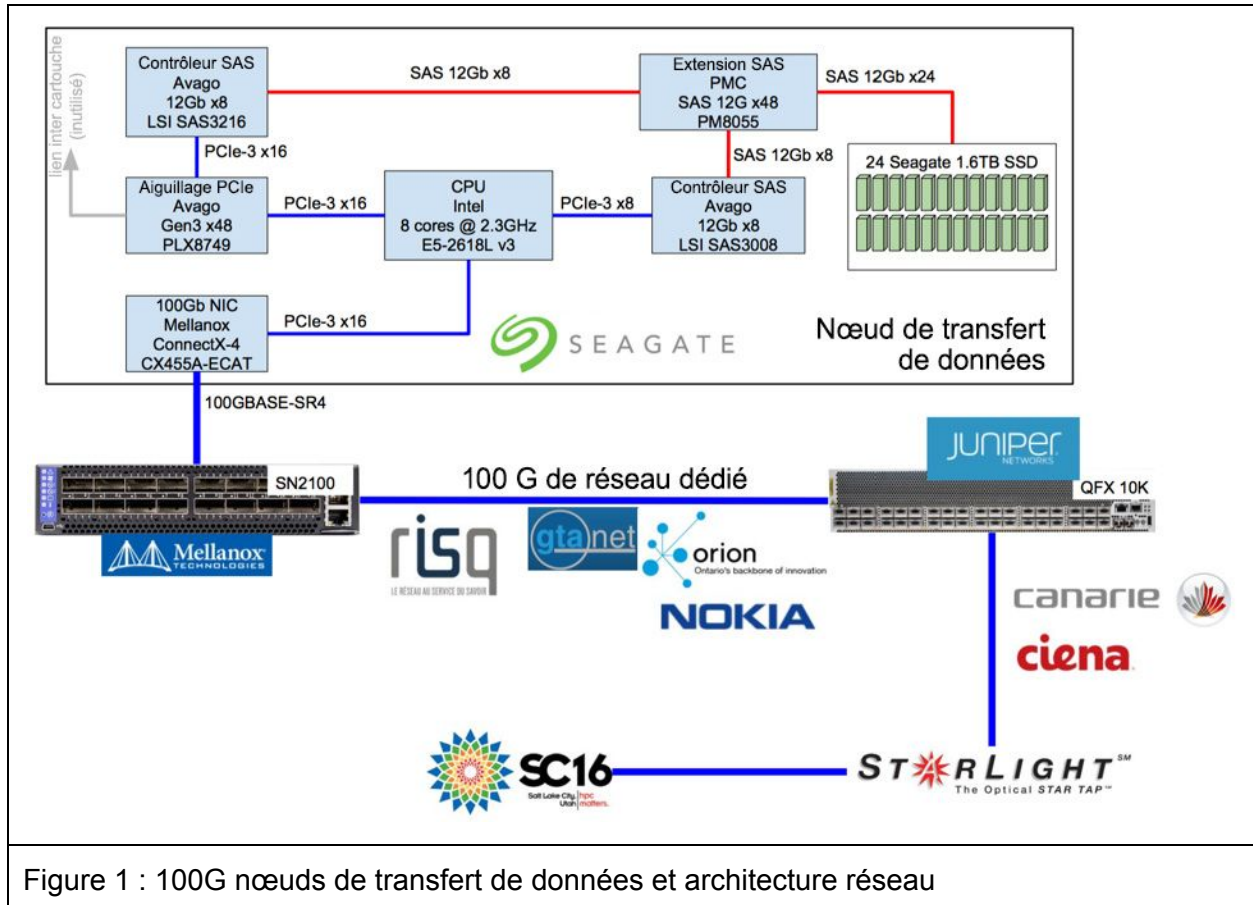
Cette présentation, effectuée à la CS16, utilise des nœuds de transferts de données (DTN) optimisés pour le transfert de données de disque à disque à 100 Gbit/s et démontre comment une telle plateforme favorise les travaux scientifiques exigeants en données dans des secteurs comme la biologie computationnelle (recherche sur le cancer, médecine personnalisée, génomique), la physique des hautes énergies, l'astronomie, la science de l'environnement et la fabrication de pointe.

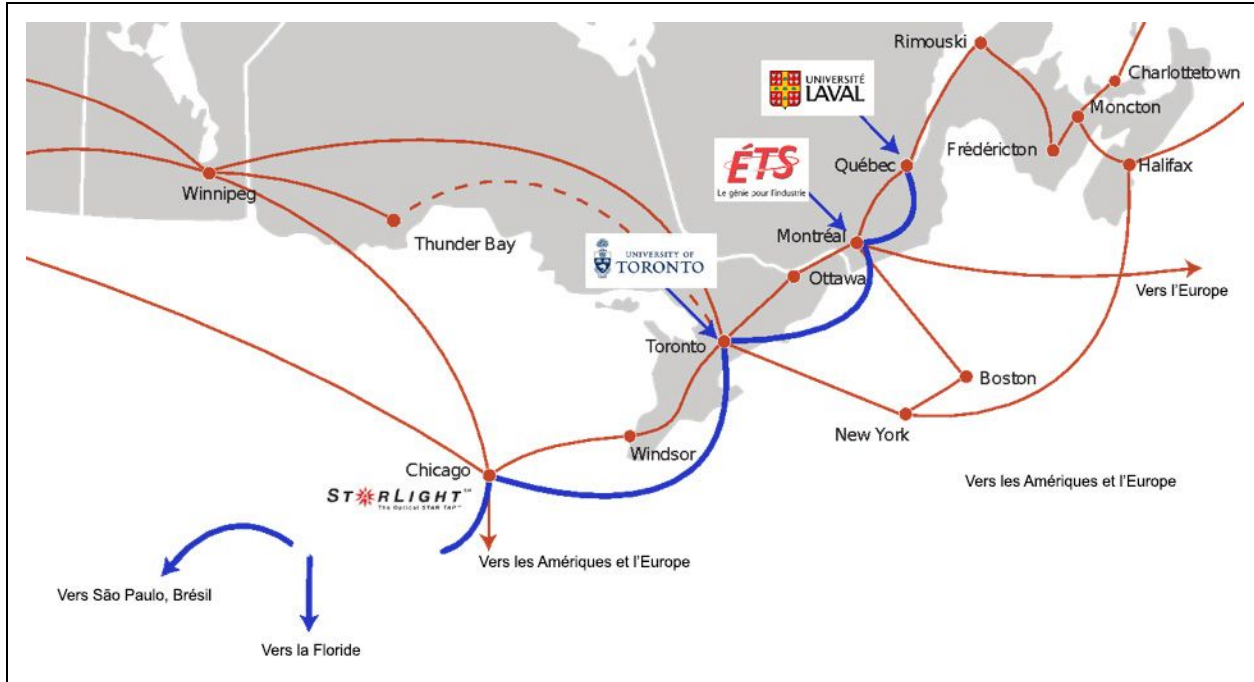
La science aux prises avec des activités à forte concentration en données s'appuie sur des outils de pointe et sur l'accès à une infrastructure numérique robuste pour assurer sa réussite.

Des réseaux rapides et fiables créent des possibilités

- la mise en place d'ensembles de données pour le calcul de haute performance (CHP);
- des copies de sauvegarde, y compris les sauvegardes redondantes sur des sites multiples;
- la réplique des données pour le stockage en mode objet distribué;
- des systèmes de fichiers croisés pour faciliter l'accès aux données persistantes comme celles d'un projet;

- la portabilité de la charge de travail, y compris pour la migration de machines virtuelles entre les hôtes et pour le placement du méta-ordonnanceur des tâches de CHP;
- l'amélioration de la qualité du service et de la disponibilité des services indispensables.





Partenaires

Voici les organisations qui collaborent avec Calcul Canada à cette présentation.



Citations

« La superinformatique est l'infrastructure de l'innovation. À l'ère numérique, l'accès à des outils capables de gérer des ensembles de données massifs est essentiel pour notre collectivité de chercheurs et d'innovateurs et pour leurs partenaires industriels et internationaux. Nous devons continuer à travailler avec nos partenaires pour rester à l'avant-garde de cette technologie et pour demeurer compétitifs sur la scène internationale, » précise Mark Dietrich, président et chef de la direction de Calcul Canada.

« Avec l'explosion des données en génomique, les capacités de transfert rapide de données sont absolument essentielles. Par exemple, nous avons un projet avec des collaborateurs japonais qui nous envoient 200 To de données sur le génome entier. Pour de tels projets, le taux de transfert de données est devenu l'un des principaux facteurs limitatifs de la collaboration, » mentionne Guillaume Bourque, un des chercheurs qui a participé à cette démonstration, professeur agrégé au département de génétique humaine à l'Université McGill et directeur du Centre canadien de génomique computationnelle.

« La possibilité de transférer des données à grande vitesse et de façon fiable donne une toute nouvelle perspective pour les chercheurs en général et spécialement pour nous, en métagénomique, qui avons généralement affaire à de grandes bases de données. Les chercheurs doivent être en mesure de transférer et de traiter facilement cette information de grande envergure afin de préserver l'avantage concurrentiel du Canada et de démocratiser l'accès aux grands ensembles de données. À titre d'exemple de grands ensembles de données, le séquençage métagénomique donne un aperçu de la biologie des communautés microbiennes qui peut être appliquée pour comprendre les maladies humaines ou les questions environnementales. Les jeux de données relatifs à ces projets génèrent des centaines de téraoctets de données, essentiellement l'équivalent de 10 millions de téléchargements de films. Ce type de recherche nous fournit de nouvelles aptitudes afin de chercher à mieux appréhender notre environnement biologique et la façon dont nous interagissons avec lui, » indique Jacques Corbeil, titulaire d'une chaire de recherche du Canada de niveau 1 en génomique médicale, professeur de médecine au Département de médecine moléculaire du Centre de recherche en données massives de l'Université Laval.