

PRÉSENTATION DE LA SOLUTION DE RÉSEAU DE DATACENTER DÉDIÉ À L'IA OUVERTE ET BASÉE SUR ETHERNET

Rapidité, Efficacité Et Économies : Notre Réseau IA Basé Sur L'intention Est Multifournisseur De Bout En Bout

Problématique

Déployer des clusters d'IA est un processus complexe, long et coûteux. Les entreprises qui investissent dans l'IA manquent souvent de ressources pour déployer leurs propres datacenters dédiés à l'IA et optimiser les ressources GPU sans enfermement propriétaire.

Solution

[La solution Juniper de datacenter dédié à l'IA](#) offre le moyen le plus rapide et le plus flexible de déployer des clusters performants pour l'apprentissage, l'inférence et le stockage de l'IA, avec un personnel IT restreint.

Avantages

- Les opérations simplifiées de conception, de déploiement et de gestion vous font économiser du temps et de l'argent avec une équipe réduite.
- Notre conception réseau flexible et ouverte accepte tous les GPU à l'aide de technologies et de produits éprouvés qui évitent l'enfermement propriétaire, avec les mêmes performances qu'InfiniBand.
- Choisissez des produits éprouvés ultrarapides à déployer avec nos solutions validées complètes et clé en main.

Les charges de travail IA exigent des performances, une capacité et une latence d'exception ; hélas, l'infrastructure des datacenters traditionnels n'est pas à la hauteur. Les fabricants d'apprentissage back-end et d'inférence front-end des datacenters IA nécessitent des solutions de calcul, de stockage et de mise en réseau de pointe. Alors que la demande en GPU puissants a explosé, l'attention s'est principalement portée sur les réseaux InfiniBand propriétaires, provenant d'une seule marque. Conséquence malheureuse : les prix ont augmenté et la chaîne logistique n'a pas pu suivre. Pour les entreprises, il est donc nécessaire de réduire les coûts en favorisant l'innovation et la flexibilité. Et pour optimiser les performances GPU à moindre coût, elles s'appuient de plus en plus sur Ethernet dans les datacenters IA.

La problématique

Dans de nombreux secteurs, l'IA passe des bancs d'essai à la production, amenant les architectes de l'IA et leurs homologues IT à se trouver soudainement face à l'inconnu. Pour tirer pleinement parti de l'IA, ils doivent d'abord gérer les complexités de son déploiement et faire face aux réalités de ses coûts et délais.

Les options GPU étant limitées et InfiniBand étant propriétaire, les retards de la chaîne logistique et les faibles stocks ont augmenté les coûts de l'infrastructure IA. Par exemple, un seul GPU peut coûter plus de 30 000 dollars, et les serveurs IA utilisant ces puces peuvent atteindre 400 000 dollars. Avec de tels prix, les petits clusters GPU coûtent des millions de dollars, contre des centaines de millions pour les plus grands.

Véritables moteurs de l'IA, les réseaux de datacenters sont essentiels pour interconnecter et maximiser les GPU. La réduction du délai de traitement des jobs (le temps nécessaire pour terminer l'apprentissage de l'IA) est essentielle pour gagner en rapidité et réduire les coûts.

La démocratisation de l'IA nécessitait une réponse rapide du secteur. Pour réduire la dépendance aux solutions propriétaires et favoriser l'ouverture et la compétition, les acteurs du secteur ont donc diversifié les GPU et se sont appuyés sur la technologie de niveau 2 la plus déployée au monde : Ethernet.

Solution Juniper Networks de réseau de datacenter dédié à l'IA

La solution Juniper de datacenter dédié à l'IA offre le moyen le plus rapide et le plus flexible de déployer des clusters performants pour l'apprentissage, l'inférence et le stockage de l'IA, avec un personnel IT restreint. Les opérations uniques basées sur l'intention,

l'assistant de réseau virtuel IA natif et les conceptions validées Juniper simplifient la conception, le déploiement et le dépannage des réseaux de datacenters dédiés à l'IA. Pucés électroniques diverses et variées, gestion multifournisseur des commutateurs, fabrics Ethernet ouvertes et standardisées... notre solution évite l'enfermement propriétaire et garantit une flexibilité sans égal.

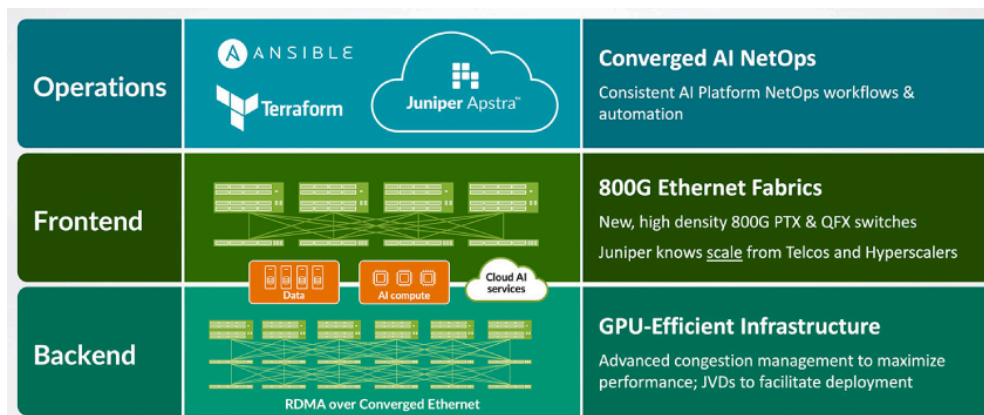


Figure 1 : Solutions Juniper Networks de réseau de datacenter dédié à l'IA

Simplification des opérations

Les opérations simplifiées de conception, de déploiement et de gestion vous font économiser du temps et de l'argent avec une équipe réduite. Seule plateforme d'automatisation multifournisseur de datacenter basée sur l'intention du secteur, le logiciel Juniper accélère le déploiement et le dépannage en évitant les problématiques liées à la configuration manuelle. Oubliez les CLI : Apstra applique des fonctions avancées d'équilibrage de charge, de contrôle de priorité des flux (PFC) et de notification explicite de congestion (ECN) à toute la fabric et garantit ainsi un réseau ultraperformant et sans perte pour l'IA. Apstra accélère la résolution des problèmes en vous permettant de visualiser les flux réseau en temps réel. Vous pouvez ainsi détecter les points de congestion dans votre datacenter IA à l'aide de heat maps et d'analyses basées sur l'intention.

En tant que source unique de vérité disposant d'une assurance en boucle fermée, Apstra offre des capacités d'analyse puissantes qui permettent d'anticiper et de résoudre les problèmes en amont afin d'éviter les interruptions de service. Basée sur des modèles appelés « blueprints », Apstra assure la fiabilité, la cohérence et la reproductibilité des réseaux, avec à la clé une diminution des OpEx pouvant atteindre 90 % et un déploiement accéléré de 85 %.

En intégrant [Mist AI™](#) à Apstra, Juniper optimise l'expérience des datacenters en renforçant l'automatisation et les analyses par [l'AI Ops](#). L'[assistant de réseau virtuel Marvis™](#) pour datacenter étend la télémétrie en streaming et la surveillance en temps réel d'Apstra au tableau de bord Marvis pour offrir un dépannage et des analyses proactifs dans tous les domaines opérationnels. Il pose ainsi les

premiers jalons qui permettront à l'IA de délivrer des informations exploitables dans les datacenters multifournisseurs. Les opérateurs de campus et de sites distants qui utilisent Marvis peuvent désormais consulter les problèmes de leurs datacenters sur le même tableau de bord.

Ouverture et flexibilité

Les solutions Ethernet ouvertes et flexibles vous permettent d'utiliser des technologies et des produits éprouvés qui évitent l'enfermement propriétaire. Les nouvelles fabrics « leaf-and-spine » 800GbE à haute densité de Juniper sont optimisées par l'IA et proposent des fonctionnalités avancées de gestion du trafic pour garantir des performances Ethernet haut débit, sans perte, évolutives et à faible latence. Avec un débit de 1,6 téraoctet et une compatibilité multifournisseur pour prendre en charge tous les GPU, Ethernet réduit les coûts, accélère l'innovation, maximise la flexibilité de conception et évite les problèmes d'approvisionnement. Grâce à RDMA over Converged Ethernet v2 (RoCEv2), la fabric Ethernet de Juniper est la seule solution multifournisseur de gestion et d'automatisation de fabric de datacenter optimisée pour l'apprentissage back-end et l'inférence front-end de l'IA.

Solutions clé en main

Nos solutions clé en main sont entièrement validées et accélèrent les déploiements de datacenters en toute fiabilité. Avec des solutions réseau de bout en bout, les entreprises peuvent

facilement construire des datacenters très performants pour l'IA. Le nouveau [commutateur fixe PTX10002-36QDD](#) 800GbE haute densité et les nouvelles cartes d'interface 800GbE pour le châssis [PTX10000](#) viennent compléter la gamme de routeurs [Juniper Networks® PTX Series pour les datacenters. Conçu avec la puce personnalisée Express 5, le plus grand châssis PTX10000 prend désormais en charge jusqu'à 576 ports 800GbE pour les architectures spine à haute densité et super spine. En outre, le QFX5240, notre nouvelle plateforme fixe](#) 800GbE haute densité basée sur le dernier ASIC Broadcom Tomahawk 5, diversifie plus encore les puces utilisées et assure des performances, une évolutivité et une efficacité énergétique de pointe dans les datacenters dédiés à l'IA.

Les plateformes QFX et PTX Series de Juniper garantissent un fonctionnement optimal des clusters d'IA en offrant des performances, une évolutivité et un contrôle de la congestion exceptionnels. Dans les tests de référence MLPerf, les solutions Ethernet optimisées par l'IA de Juniper ont obtenu des performances de traitement JCT similaires à InfiniBand pour les charges de travail d'entraînement, offrant ainsi une fabric alternative ouverte et performante évitant l'enfermement propriétaire. Pour simplifier les déploiements de l'IA, Juniper a introduit de nouvelles conceptions validées Juniper (JVD) : des fabrics Clos multicouches « rail-optimized » qui forment une solution de datacenter complète incluant la commutation, les opérations et la sécurité. Les JVD sont rigoureusement testées, validées et documentées par les professionnels de Juniper Labs et peuvent être utilisées en tant que conceptions prêtes à l'emploi ou en tant que lignes directrices pour réduire les risques et dimensionner et budgétiser chaque cluster d'IA.

Fonctionnalités et avantages

Simplification des opérations : simplifiez les opérations de conception, d'implémentation et de gestion pour gagner du temps et de l'argent avec une équipe réduite. Technologie de pointe, notre mise en réseau basée sur l'intention accélère le temps de déploiement et de dépannage en évitant les problèmes de configuration manuelle (quels que soient les fournisseurs de matériel).

- La source unique de vérité avec assurance en boucle fermée permet de prédire et de résoudre les problèmes en amont afin d'éviter les interruptions de service.
- L'AIOPS pour les datacenters offre une automatisation et des informations supplémentaires pour optimiser les expériences des datacenters.
- Notre solution rentable réduit les charges d'exploitation de 90 % et accélère le déploiement de 85 %.

Ouverture et flexibilité : concevez votre réseau à l'aide de technologies et de produits éprouvés qui évitent l'enfermement propriétaire.

- Les fonctionnalités avancées de gestion du trafic garantissent des performances Ethernet évolutives et sans perte, à haut débit et faible latence.
- L'utilisation d'Ethernet permet de réduire les coûts, d'accélérer l'innovation, de maximiser la flexibilité de conception et d'éviter les problèmes liés à la chaîne logistiqu.
- Notre solution multifournisseur unique pour la gestion et l'automatisation des fabrics de datacenter maximise la flexibilité de conception.
- La compatibilité est garantie avec les principaux GPU, fabrics réseau et commutateurs.

Solutions clés en main : les solutions validées de bout en bout garantissent la fiabilité des produits pour accélérer le déploiement.

- La grande diversité des circuits intégrés garantit une évolutivité, une performance et une flexibilité maximales. L'interconnexion qu'offrent les puces Broadcom et les puces personnalisées permettent de prendre en charge des milliers de GPU à des vitesses maximales (800 Gbit/s), et toutes les configurations des fabrics de datacenter.
- Notre portefeuille complet de sécurité pour les datacenters assure une sécurité de bout en bout.
- Les conceptions validées Juniper forment une solution de datacenter complète couvrant la commutation, les opérations et la sécurité.
- La gamme complète de commutateurs Juniper (intégrant des circuits divers) favorise l'évolutivité et les performances.
- Le portefeuille de sécurité complet pour datacenter unifie la prévention des menaces pour une sécurité Zero Trust.
- La fiabilité des réseaux de datacenters est décuplée.

Composants de la solution

Apstra : la solution d'automatisation de datacenter de Juniper permet aux opérateurs d'automatiser le cycle de vie complet du réseau de datacenters dans les environnements multifournisseurs, de la conception à la gestion opérationnelle. Apstra fournit une validation continue, une source unique de vérité, des analyses puissantes, ainsi qu'une identification et une correction rapides des causes racines.

Commutateurs QFX Series : cette gamme exceptionnelle de commutateurs réseau QFX offre un débit et une évolutivité supérieurs, une pile de routage complète, la programmabilité ouverte du système d'exploitation [Junos® OS](#) et le plus large éventail de fonctionnalités de fabric IP et [EVPN-VXLAN](#).

Routeurs PTX Series : ils constituent la base des plus grandes architectures centrales et WAN au monde. Propulsés par la dernière génération d'ASIC de la famille Express, ces routeurs sont optimisés pour le cloud afin de permettre des migrations 800GbE fluides. Ils offrent simultanément l'évolutivité, la flexibilité et le retour sur investissement nécessaires pour répondre aux besoins actuels du marché.

Assistant de réseau virtuel Marvis pour datacenter : Marvis est le premier et le seul assistant de réseau virtuel IA natif pour datacenter. Il fournit des informations fiables tout au long du cycle de vie, quel que soit le fournisseur du matériel. De plus, il garantit une visibilité et une assurance de bout en bout dans tous les domaines de l'entreprise, des campus et des sites distants jusqu'aux datacenters.

Junos OS Evolved : ce système d'exploitation réseau unifié de bout en bout offre fiabilité, agilité et programmabilité ouverte pour des déploiements réussis à l'échelle du cloud.

En bref : des opérations simplifiées, une flexibilité ouverte et des réseaux clé en main pour les datacenters IA

Comment exploiter au mieux l'infrastructure du datacenter pour se mettre à l'IA ? Aujourd'hui, peu d'entreprises ne se posent pas la question. La solution AI Data Center de Juniper est le moyen le plus simple et le plus rapide de déployer un réseau évolutif et performant pour l'entraînement et l'inférence de l'IA. Avec les opérations multifournisseurs basées sur l'intention, l'assistant de réseau virtuel IA natif et les conceptions validées Juniper, les réseaux de datacenter IA ont tout d'un jeu d'enfant.

Corporate and Sales Headquarters

Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089 USA **Téléphone :**
888.JUNIPER (888.586.4737)
ou +1.408.745.2000
www.juniper.net

APAC and EMEA Headquarters

Juniper Networks International B.V.
Boeing Avenue 240
1119 PZ Schiphol-Rijk
Amsterdam, Pays-Bas **Téléphone :**
+31 0 207 125 700