

# Solutions Ethernet ouvertes de Dell Technologies pour l'IA générative

## Repousser les limites de l'infrastructure IT

« Les fabricants d'IA (commutation back-end pour la connectivité GPU-à-GPU) devraient connaître une croissance de 1,2 milliard de dollars (2022) à 15,2 milliards de dollars (2027), avec un TCAC de 65 % sur cinq ans.

Ethernet devrait atteindre 32 % de part de chiffre d'affaires et représenter 37 % des expéditions de ports pour les fabricants d'IA (2027)

### Étude Dell'Oro<sup>1</sup>

Les performances des processeurs graphiques dépendent fortement des performances du réseau. De nombreuses charges applicatives d'IA s'exécutent sur de grands clusters de serveurs qui nécessitent une communication constante entre chaque nœud de calcul et système de stockage, et ont donc besoin d'une gestion de réseau robuste pour éviter les goulots d'étranglement. Si les performances de votre réseau ne sont pas suffisantes pour la charge applicative, les processeurs graphiques deviennent inactifs, et l'entraînement et l'inférence prennent plus de temps, ce qui ralentit le traitement des données et le délai d'obtention des informations.

### Introduction : exigences réseau de l'IA générative

Alors que les solutions d'IA générative (GenAI) continuent d'évoluer, repoussant les limites du traitement de données et des besoins de calcul, les infrastructures IT sont confrontées à un nouveau défi : trouver des moyens de prendre en charge les immenses exigences de ces environnements. Ces modèles, en particulier les grands modèles de langage (LLM), exigent non seulement plus d'infrastructure, mais également des systèmes soigneusement architecturés pour gérer les besoins de connectivité massifs entre les clusters de processeurs graphiques. Les solutions réseau traditionnelles forment rapidement des goulots d'étranglement qui menacent la viabilité et la réussite des initiatives d'IA générative. Les fabricants d'IA nécessitent une faible latence, des performances sans perte et une bande passante maximale. Les exigences massives en matière de traitement des données et des applications entraînent une augmentation des exigences pour les fabricants front-end et back-end.

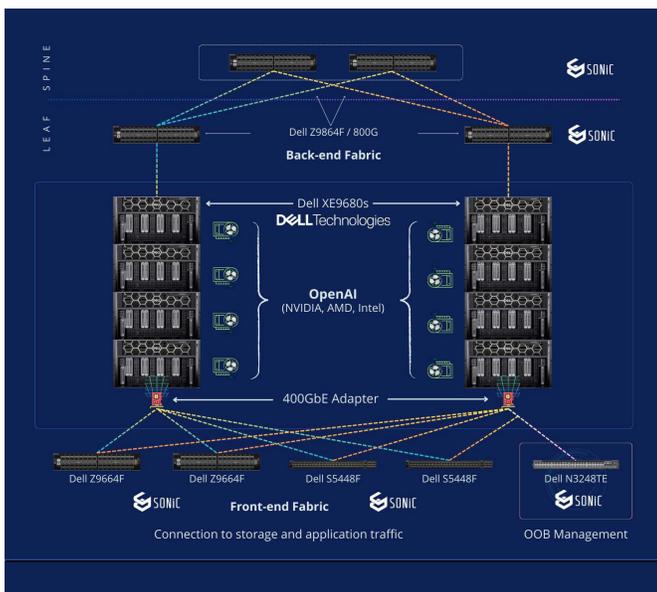
### Défis et besoins liés au déploiement d'une infrastructure d'IA générative

Le déploiement des technologies d'IA générative présente de nombreux défis, allant des complexités techniques associées aux nouvelles architectures au manque de professionnels qualifiés capables de gérer de tels déploiements. Les solutions qui s'appuient sur des technologies propriétaires, telles qu'InfiniBand, ajoutent une couche de complexité supplémentaire, limitant la disponibilité des ressources et compliquant l'intégration avec les plateformes de surveillance ou d'orchestration existantes. En outre, les coûts élevés, la longueur des délais d'évaluation et la dépendance vis-à-vis des fournisseurs associés aux solutions propriétaires constituent des obstacles importants, en particulier à une époque pleine d'incertitudes concernant la chaîne logistique. Ces défis soulignent le besoin urgent de solutions d'infrastructure d'IA générative ouvertes, flexibles et robustes, capables de répondre aux exigences uniques des charges applicatives d'IA générative.

### Approche de Dell Technologies en matière de gestion réseau pour l'IA générative

Dell Technologies a été le pionnier des solutions complètes et ouvertes basée sur Ethernet, conçues pour relever ces défis et répondre aux exigences complexes de l'infrastructure de l'IA générative. Tirant parti de sa vaste expérience dans les environnements d'IA, de modélisation et de calcul haute performance (HPC), Dell Technologies propose une suite de solutions qui répondent aux exigences front-end et back-end. Des systèmes de calcul modulaires optimisés pour l'accélération comme les serveurs Dell PowerEdge XE, aux solutions de stockage axées sur l'IA telles que PowerScale, Dell Technologies fournit les composants essentiels pour un déploiement réussi de l'IA générative. Au cœur de cette approche se trouve le déploiement de structures Ethernet de nouvelle génération alimentées par une puce silicium réseau avancée. Le **Dell PowerSwitch Z9864-ON** offre **800 GbE** de performances réseau non bloquantes, essentielles pour les applications d'IA générative. Les clients peuvent ainsi déployer des clusters d'IA avec une faible latence et un débit élevé à l'aide d'une commutation à bande passante élevée et les nouvelles fonctionnalités de **la solution Enterprise SONiC Distribution de Dell Technologies**, telles que le routage avancé, RoCEv2, le hachage amélioré et le contrôle de flux prioritaire, pour de meilleures performances de fabric et une meilleure surveillance de la congestion.

<sup>1</sup> Présentation de Dell'Oro : Advanced Research Report on AI Networks for AI Workloads



Exemple d'architecture de fabric l'IA générative

La série Dell PowerSwitch Z, qui utilise une puce silicium de pointe, constitue le pilier d'un réseau évolutif hautes performances capable de prendre en charge des milliers de nœuds, répondant ainsi aux défis de connectivité inhérents aux applications d'IA générative.

### Accélérer le déploiement de l'IA générative avec Dell Technologies

L'essor de l'IA générative a entraîné une multitude de défis pour les infrastructures IT, exigeant une nouvelle approche de la gestion réseau à la fois innovante et flexible. Dell Technologies relève le défi en proposant des solutions ouvertes basées sur Ethernet qui répondent aux besoins immédiats des déploiements d'IA générative et posent les bases des futures avancées.

Afin d'éviter les approximations des solutions matérielles d'IA, Dell propose des architectures de référence validées en laboratoire et optimisées pour les charges applicatives d'IA. Ces conceptions validées comprennent des concepts architecturaux, des présentations de solutions complètes, des validations de performances et d'autres validations en laboratoire prouvant les capacités de la solution sur la charge applicative pour laquelle elle a été conçue. Faites de l'IA une réalité éprouvée avec des solutions validées qui permettent de fournir des informations plus approfondies, plus rapidement.

En choisissant Dell Technologies, les organisations bénéficient d'un partenaire disposant de l'expertise, d'une pile de solutions de bout en bout et de l'engagement nécessaires pour garantir la réussite de leurs initiatives d'IA générative. Avec Dell Technologies, les entreprises sont équipées pour affronter les complexités des architectures d'IA générative, ce qui garantit la viabilité et la réussite de leurs projets.

Accélérez le déploiement et le délai de rentabilisation de vos environnements l'IA générative en réduisant les risques et la complexité opérationnelle avec Dell Technologies. Découvrez comment transformer vos initiatives d'IA générative et faire entrer votre entreprise dans une nouvelle ère d'innovation et d'efficacité grâce à une solution réseau ouverte, flexible et durable.

### Innovations Dell Technologies en matière de gestion réseau pour l'IA générative

Dell Technologies est à la pointe de l'innovation en matière de gestion réseau pour l'IA générative, en proposant des solutions qui répondent aux exigences des environnements d'IA générative d'aujourd'hui et de demain, du datacenter au Cloud, en passant par la périphérie. En se concentrant sur des solutions ouvertes et extensibles tirant parti des puces silicium grand public et des logiciels Open Source, Dell Technologies garantit des performances et une flexibilité maximales.

L'utilisation de systèmes d'exploitation réseau Open Source disponibles dans le commerce tels que SONiC, associée à la participation active et à la contribution de Dell Technologies au **consortium Ultra Ethernet (UEC)**, souligne son engagement en faveur des normes ouvertes et du développement collaboratif dans l'espace Ethernet. Ces efforts visent à faire en sorte qu'Ethernet continue de jouer un rôle central dans la prise en charge de la prochaine génération d'environnements d'IA.



[En savoir plus](#) sur les solutions Dell Networking



[Contacter un expert](#) Dell Technologies



[Lire le rapport d'analyste](#) d'ESG



[Lire le rapport d'analyste](#) |d'IDC