NetApp AIPod avec NVIDIA

NetApp Solutions

NetApp
August 14, 2024

Sommaire

NetApp ai avec NVIDIA

Présentation des solutions d'infrastructure convergée ONTAP ai de NetApp et NVIDIA.

NetApp AIPod avec les systèmes NVIDIA DGX

- "FlexPod pour IA NetApp avec les systèmes NVIDIA DGX"

NetApp ONTAP ai avec les systèmes NVIDIA DGX A100

- "Guide de conception"
- "Guide de déploiement"

NetApp ONTAP ai avec les systèmes NVIDIA DGX A100 et les switchs Ethernet Mellanox Spectrum

- "Guide de conception"
- "Guide de déploiement"

NetApp AIPod avec les systèmes NVIDIA DGX - Présentation

Cette section présente NetApp AIPod avec des systèmes NVIDIA DGX.

Ingénierie de solutions NetApp

NetApp&#8482 ; AIPod avec NVIDIA DGX&#8482 ; les systèmes et les systèmes de stockage NetApp connectés au cloud simplifient les déploiements d'infrastructure pour les workloads de machine learning (ML) et d'intelligence artificielle (IA) en éliminant la complexité de la conception et les approximations. Reposant sur la conception NVIDIA DGX BasePOD pour offrir des performances de calcul exceptionnelles pour les charges de travail nouvelle génération, AIPod avec les systèmes NVIDIA DGX ajoute des systèmes de stockage NetApp AFF qui permettent aux clients de commencer avec un déploiement de petite taille, puis d'évoluer de manière non disruptive tout en gérant intelligemment les données de la périphérie au cœur, et jusqu'au cloud, et inversement. NetApp AIPod fait partie du portefeuille plus vaste de solutions d'IA de NetApp, illustré dans la figure ci-dessous :

NetApp ai Solutions Portfolio image::aipod_nv_Portfolio.png[]

Ce document décrit les principaux composants de l'architecture de référence AIPod, les informations sur la connectivité du système et les conseils sur le dimensionnement de la solution. Ce document est destiné aux ingénieurs de solutions partenaires et NetApp, ainsi qu'aux décideurs stratégiques des clients intéressés par le déploiement d'une infrastructure haute performance pour les workloads de ML/DL et d'analytique.

NetApp AIPod avec les systèmes NVIDIA DGX - Présentation

Cette section présente NetApp AIPod avec des systèmes NVIDIA DGX.

Ingénierie de solutions NetApp

NetApp&#8482 ; AIPod avec NVIDIA DGX&#8482 ; les systèmes et les systèmes de stockage NetApp connectés au cloud simplifient les déploiements d'infrastructure pour les workloads de machine learning (ML) et d'intelligence artificielle (IA) en éliminant la complexité de la conception et les approximations. Reposant sur la conception NVIDIA DGX BasePOD pour offrir des performances de calcul exceptionnelles pour les charges de travail nouvelle génération, AIPod avec les systèmes NVIDIA DGX ajoute des systèmes de stockage NetApp AFF qui permettent aux clients de commencer avec un déploiement de petite taille, puis d'évoluer de manière non disruptive tout en gérant intelligemment les données de la périphérie au cœur, et jusqu'au cloud, et inversement. NetApp AIPod fait partie du portefeuille plus vaste de solutions d'IA de NetApp, illustré dans la figure ci-dessous :

NetApp ai Solutions Portfolio image::aipod_nv_Portfolio.png[]

Ce document décrit les principaux composants de l'architecture de référence AIPod, les informations sur la connectivité du système et les conseils sur le dimensionnement de la solution. Ce document est destiné aux ingénieurs de solutions partenaires et NetApp, ainsi qu'aux décideurs stratégiques des clients intéressés par le déploiement d'une infrastructure haute performance pour les workloads de ML/DL et d'analytique.

NetApp AIPod avec les systèmes NVIDIA DGX - composants matériels

Cette section s'intéresse aux composants matériels de NetApp AIPod avec les systèmes NVIDIA DGX.

Systèmes de stockage NetApp AFF

Les systèmes de stockage de pointe NetApp AFF permettent aux services IT DE répondre aux besoins de stockage des entreprises grâce aux performances de pointe, à la flexibilité supérieure, à l'intégration au cloud et à la gestion des données optimale. Conçues spécifiquement pour les systèmes Flash, les baies AFF contribuent à accélérer, gérer et protéger les données stratégiques.

Systèmes de stockage AFF A900

Le système NetApp AFF A900 optimisé par le logiciel de gestion des données NetApp ONTAP offre une protection intégrée des données, des fonctionnalités anti-ransomware en option, ainsi que les performances et la résilience élevées requises pour prendre en charge les workloads les plus stratégiques. Il élimine les interruptions des opérations stratégiques, limite les ajustements des performances et protège vos données contre les attaques par ransomware. Il offre :
• Rendement de pointe de l'industrie
• Sécurité des données sans compromis
• Mises à niveau simplifiées sans interruption

NetApp AFF A900 Storage System image::aipod_nv_A900.png[]

Performances exceptionnelles

Le système AFF A900 gère facilement des workloads nouvelle génération tels que le deep learning, l'IA et l'analytique ultra-rapide, ainsi que des bases de données d'entreprise classiques comme Oracle, SAP HANA, Microsoft SQL Server et des applications virtualisées. Elle assure le fonctionnement optimal des applications stratégiques avec jusqu'à 2,4 millions d'IOPS par paire HA et une latence aussi faible que 100 µs, et augmente les performances jusqu'à 50 % par rapport aux modèles NetApp précédents. Grâce à NFS over RDMA, pNFS et à session Trunking, les entreprises peuvent atteindre le haut niveau de performances réseau requis pour les applications nouvelle génération en utilisant l'infrastructure réseau de data Center existante. Les clients peuvent également évoluer et évoluer grâce à la prise en charge multiprotocole unifiée des environnements SAN, NAS et de stockage objet. Ils bénéficient d'une flexibilité maximale grâce au logiciel unique de gestion des données ONTAP unifiée, pour les données sur site ou dans le cloud. De plus, l'état du système peut être

optimisé grâce aux analyses prédictives basées sur l'IA fournies par Active IQ Digital Advisor (également appelé Digital Advisor) et Cloud Insights.

Sécurité des données sans compromis

Les systèmes AFF A900 comprennent une suite complète de logiciels de protection des données NetApp intégrés et cohérents au niveau des applications. Il offre une protection des données intégrée et des solutions anti-ransomware de pointe pour l'anticipation et la reprise en cas d'attaque. Les fichiers malveillants peuvent être bloqués afin d'être écrits sur le disque. Les anomalies du stockage sont facilement contrôlées pour obtenir des informations exploitables.

Mises à niveau simplifiées sans interruption

Le système AFF A900 est disponible sous forme de mise à niveau du châssis sans interruption pour les clients A700. NetApp simplifie les mises à jour et élimine les interruptions des opérations stratégiques grâce à ses fonctionnalités avancées de fiabilité, de disponibilité, de facilité de maintenance et de gestion. En outre, NetApp renforce l'efficacité opérationnelle et simplifie les activités quotidiennes des équipes IT, car le logiciel ONTAP applique automatiquement les mises à jour de firmware à tous les composants système.

Pour les déploiements les plus vastes, les systèmes AFF A900 offrent les meilleures performances et capacités, tandis que d'autres systèmes de stockage NetApp, tels que AFF A800, AFF C800, AFF A400, AFF C400 et AFF A250, proposent des options pour les déploiements de plus petite taille à moindre coût.

NVIDIA DGX BasePOD

NVIDIA DGX BasePOD est une solution intégrée qui comprend des composants matériels et logiciels NVIDIA, des solutions MLOps et du stockage tiers. En tirant parti des bonnes pratiques de conception de systèmes scale-out avec les produits NVIDIA et les solutions de partenaires validées, les clients peuvent implémenter une plateforme efficace et gérable pour le développement de l'IA. La Figure 1 présente les différents composants de NVIDIA DGX BasePOD.

NVIDIA DGX BasePOD solution image::aipod_nv_basepod_layods.png[]

SYSTÈMES NVIDIA DGX H100

Le système NVIDIA DGX H100 et#8482 ; est la puissance de l'IA accélérée par les performances révolutionnaires du processeur graphique NVIDIA H100 Tensor Core.

NVIDIA DGX H100 system image::aipod_nv_H100_3D.png[]

Principales spécifications du système DGX H100 :
• Huit processeurs graphiques NVIDIA H100.
• 80 Go de mémoire GPU par GPU, pour un total de 640 Go.
• Quatre puces NVIDIA NVSwitch™.
• Processeurs Intel® Xeon® Platinum 8480 double cœur à 56 cœurs avec prise en charge de PCIe 5.0.
• 2 To de mémoire système DDR5.
• Quatre ports OSFP desservant huit adaptateurs NVIDIA ConnectX-7 (InfiniBand/Ethernet) à un port et deux adaptateurs NVIDIA ConnectX-7 (InfiniBand/Ethernet) à deux ports.
• Deux disques NVMe M.2 de 1.92 To pour le système d'exploitation DGX, huit disques NVMe U.2 de 3.84 To pour le stockage/cache.
• Puissance maximale de 10.2 kW.
Les ports arrière du plateau du processeur DGX H100 sont illustrés ci-dessous. Quatre ports OSFP servent huit adaptateurs ConnectX-7 pour la structure de calcul InfiniBand. Chaque paire d'adaptateurs ConnectX-7 à deux ports fournit des voies parallèles aux structures de stockage et de gestion. Le port hors bande est utilisé pour l'accès au contrôleur BMC.

NVIDIA DGX H100 panneau arrière image::aipod_nv_H100_REAR.png[]

Mise en réseau NVIDIA

Commutateur NVIDIA Quantum-2 QM9700

NVIDIA Quantum-2 QM9700 InfiniBand switch image::aipod_nv_QM9700.png[]

Les switchs NVIDIA Quantum-2 QM9700 dotés d'une connectivité InfiniBand 400 Go/s alimentent la structure de calcul des configurations NVIDIA Quantum-2 InfiniBand BasePOD. Les adaptateurs à port unique ConnectX-7 sont utilisés pour la structure de calcul InfiniBand. Chaque système NVIDIA DGX possède des connexions doubles à chaque switch QM9700, offrant ainsi plusieurs chemins à large bande passante et à faible latence entre les systèmes.

Commutateur NVIDIA Spectrum-3 SN4600

NVIDIA Spectrum-3 SN4600 Switch image::aipod_nv_SN4600_HiRes_LLEGER.png[]

Les switchs NVIDIA Spectrum-3 SN4600 offrent 128 ports au total (64 par switch) pour assurer une connectivité redondante pour la gestion intrabande du serveur DGX BasePOD. Le commutateur NVIDIA SN4600 peut fournir des vitesses comprises entre 1 GbE et 200 GbE. Les switchs NVIDIA SN4600 sont également utilisés pour les appliances de stockage connectées via Ethernet. Les ports des adaptateurs NVIDIA DGX à deux ports ConnectX-7 sont utilisés à la fois pour la gestion intrabande et pour la connectivité du stockage.

Commutateur NVIDIA Spectrum SN2201

NVIDIA Spectrum SN2201 switch image::aipod_nv_SN2201.png[]

Les commutateurs NVIDIA Spectrum SN2201 disposent de 48 ports pour assurer la connectivité pour la gestion hors bande. La gestion hors bande assure une connectivité de gestion consolidée pour tous les composants du DGX BasePOD.

Adaptateur NVIDIA ConnectX-7

Adaptateur NVIDIA ConnectX-7 image::aipod_nv_CX7.png[]

L'adaptateur NVIDIA ConnectX-7 offre un débit de 25/50/100/200/400G. Les systèmes NVIDIA DGX utilisent à la fois les adaptateurs ConnectX-7 à un et deux ports pour assurer la flexibilité des déploiements DGX BasePOD avec InfiniBand 400 Go/s et Ethernet 100 Gb.

NetApp AIPod avec les systèmes NVIDIA DGX - composants logiciels

Cette section s'intéresse aux composants logiciels de NetApp AIPod avec des systèmes NVIDIA DGX.

Logiciel NVIDIA

Commande de base NVIDIA

NVIDIA base Command&#8482 ; optimise chaque DGX BasePOD en permettant aux entreprises d'exploiter le meilleur des innovations logicielles NVIDIA. Les entreprises peuvent exploiter tout le potentiel de leur investissement grâce à une plateforme à l'efficacité prouvée qui inclut l'orchestration haute performance et la gestion des clusters, des bibliothèques qui accélèrent l'infrastructure de calcul, de stockage et de réseau, ainsi

qu'un système d'exploitation optimisé pour les workloads d'IA.

NVIDIA BaseCommand solution image::aipod_nv_BaseCommand_New.png[]

NVIDIA GPU CLOUD (NGC)

Le logiciel NVIDIA NGC™ permet de répondre aux besoins des data Scientists, des développeurs et des chercheurs qui possèdent divers niveaux d'expertise en IA. Les logiciels hébergés sur NGC sont soumis à des analyses en fonction d'un ensemble agrégé de vulnérabilités et d'expositions courantes, de clés cryptographiques et privées. Elle a été testée et conçue pour évoluer vers plusieurs GPU et, dans de nombreux cas, vers un système multi-nœuds, afin d'assurer aux utilisateurs qu'ils investissent pleinement dans les systèmes DGX.

NVIDIA GPU Cloud image::aipod_nv_ngc.png[]

NVIDIA ai Enterprise

NVIDIA ai Enterprise est la plateforme logicielle de bout en bout qui met l'IA générative à la portée de toutes les entreprises. Elle assure ainsi le temps d'exécution le plus rapide et le plus efficace pour les modèles de base d'IA générative optimisés pour s'exécuter sur la plateforme NVIDIA DGX. Grâce à sa sécurité, sa stabilité et sa facilité de gestion de qualité production, elle rationalise le développement de solutions d'IA générative. NVIDIA ai Enterprise est inclus avec DGX BasePOD pour que les développeurs accèdent à des modèles pré-entraînés, à des frameworks optimisés, à des microservices, à des bibliothèques accélérées et à un support d'entreprise.

Logiciel NetApp

NetApp ONTAP

ONTAP 9, la dernière génération de logiciel de gestion du stockage de NetApp, permet aux entreprises de moderniser l'infrastructure et de passer à un data Center prêt pour le cloud. Avec des capacités de gestion des données à la pointe du secteur, ONTAP permet de gérer et de protéger les données avec un seul ensemble d'outils, quel que soit leur emplacement. Vous pouvez aussi déplacer vos données librement partout où elles sont nécessaires : la périphérie, le cœur ou le cloud. ONTAP 9 comprend de nombreuses fonctionnalités qui simplifient la gestion des données, accélèrent et protègent les données stratégiques, et permettent d'utiliser des fonctionnalités d'infrastructure nouvelle génération dans toutes les architectures de cloud hybride.

Accélération et protection des données

ONTAP offre des niveaux supérieurs de performances et de protection des données et étend ces fonctionnalités aux méthodes suivantes :

- Des performances élevées et une faible latence. ONTAP offre le débit le plus élevé possible à la latence la plus faible possible, y compris la prise en charge de NVIDIA GPUDirect Storage (GDS) utilisant NFS over RDMA, Parallel NFS (pNFS) et l'agrégation de sessions NFS.
- Protection des données. ONTAP offre des fonctionnalités de protection des données intégrées et la garantie anti-ransomware la plus forte du secteur avec une gestion commune sur toutes les plateformes.
- NetApp Volume Encryption (NVE). ONTAP offre un chiffrement natif au niveau du volume avec un support de gestion des clés interne et externe.
- La colocation du stockage et l'authentification multifacteur ONTAP permet le partage des ressources d'infrastructure avec les plus hauts niveaux de sécurité.

Gestion simplifiée

La gestion des données est cruciale pour les opérations IT et les data Scientists, de sorte que les ressources appropriées sont utilisées pour les applications d'IA et pour l'entraînement des datasets d'IA/DE ML. Les informations supplémentaires suivantes sur les technologies NetApp ne sont pas incluses dans cette validation, mais elles peuvent être pertinentes en fonction de votre déploiement.

Le logiciel de gestion des données ONTAP comprend les fonctionnalités suivantes pour rationaliser et simplifier les opérations et réduire le coût total d'exploitation :

- Les copies Snapshot et les clones permettent la collaboration, l'expérimentation en parallèle et une gouvernance améliorée des données pour les workflows de MACHINE LEARNING et de deep learning.
- SnapMirror permet de déplacer les données de manière transparente dans les environnements de cloud hybride et multi-sites, et de les transférer à tout moment et en tout lieu.
- Compaction des données à la volée et déduplication étendue La compaction des données réduit le gaspillage d'espace à l'intérieur des blocs de stockage, et la déduplication augmente considérablement la capacité effective. Cela s'applique aux données stockées localement et à leur placement dans le cloud.
- Qualité de service (AQoS) minimale, maximale et adaptative. Les contrôles granulaires de la qualité de service (QoS) permettent de maintenir les niveaux de performance des applications stratégiques dans des environnements hautement partagés.
- Les FlexGroups NetApp permettent de répartir les données entre les différents nœuds du cluster de stockage, offrant ainsi une capacité massive et des performances supérieures pour les jeux de données extrêmement volumineux.
- NetApp FabricPool Tiering automatique des données inactives vers des options de stockage de cloud public et privé, notamment Amazon Web Services (AWS), Azure et la solution de stockage NetApp StorageGRID. Pour plus d'informations sur FabricPool, voir "Tr-4598 : meilleures pratiques de FabricPool".
- NetApp FlexCache : Offre des fonctionnalités de mise en cache de volume à distance qui simplifient la distribution de fichiers, réduisent la latence des réseaux WAN et diminuent les coûts de bande passante WAN. FlexCache permet le développement de produits distribués sur plusieurs sites et l'accès accéléré aux jeux de données de l'entreprise à partir de sites distants.

Une infrastructure pérenne

ONTAP permet de répondre aux besoins métier en constante évolution grâce aux fonctionnalités suivantes :

- Évolutivité transparente et opérations non disruptives. ONTAP prend en charge l'ajout en ligne de capacité aux contrôleurs et l'évolution scale-out des clusters. Les clients peuvent effectuer la mise à niveau vers les technologies les plus récentes, telles que NVMe et FC 32 Gb, sans migration des données ni panne coûteuse.
- Connexion cloud. ONTAP est le logiciel de gestion de stockage le plus connecté au cloud, avec des options de stockage SDS (ONTAP Select) et des instances natives de cloud (NetApp Cloud Volumes Service) dans tous les clouds publics.
- Intégration avec les applications émergentes ONTAP propose des services de données d'entreprise pour les plateformes et applications nouvelle génération, telles que les véhicules autonomes, les Smart cities et Industry 4.0, en utilisant la même infrastructure prenant en charge les applications d'entreprise existantes.

Kit NetApp DataOps

Le kit NetApp DataOps est un outil Python qui simplifie la gestion des espaces de travail de développement/formation et des serveurs d'inférence, lesquels sont basés sur un stockage NetApp haute performance et scale-out. Le kit DataOps Toolkit peut fonctionner comme un utilitaire autonome et est encore

plus efficace dans les environnements Kubernetes en tirant parti d'NetApp Astra Trident pour automatiser les opérations de stockage. Les fonctionnalités principales comprennent :

- Provisionnez rapidement de nouveaux espaces de travail JupyterLab haute capacité, soutenus par un stockage NetApp haute performance et scale-out.
- Provisionnez rapidement les nouvelles instances NVIDIA Triton Inférence Server, qui sont sauvegardées par un système de stockage NetApp de grande qualité.
- Clonage quasi instantané des espaces de travail JupyterLab haute capacité afin de permettre l'expérimentation ou l'itération rapide.
- Snapshots quasi instantanés des espaces de travail JupyterLab haute capacité pour la sauvegarde et/ou la traçabilité/référence.
- Provisionnement quasi instantané, clonage et copies Snapshot de volumes de données hautes performances de grande capacité.

NetApp Astra Trident

ASTRA Trident est un orchestrateur de stockage open source entièrement pris en charge pour les conteneurs et les distributions Kubernetes, notamment Anthos. Trident fonctionne avec l'ensemble de la gamme de stockage NetApp, y compris NetApp ONTAP, et prend également en charge les connexions NFS, NVMe/TCP et iSCSI. Trident accélère le workflow DevOps en permettant aux utilisateurs d'approvisionner et de gérer le stockage à partir de leurs systèmes de stockage NetApp, sans intervention de l'administrateur de stockage.

NetApp AIPod avec les systèmes NVIDIA DGX - Architecture de la solution

Cette section est consacrée à l'architecture de NetApp AIPod avec les systèmes NVIDIA DGX.

NetApp ai Pod avec les systèmes DGX H100

Cette architecture de référence utilise des structures distinctes pour l'interconnexion des clusters de calcul et l'accès au stockage, avec une connectivité InfiniBand (IB) de 400 Go/s entre les nœuds de calcul. L'illustration ci-dessous présente la topologie globale de la solution NetApp AIPod avec les systèmes DGX H100.

NetApp topologie de la solution AIpod image::aipod_nv_a900topo.png[]

Configuration du réseau

Dans cette configuration, la structure du cluster de calcul utilise une paire de commutateurs IB 400 Go/s QM9700, qui sont connectés ensemble pour une haute disponibilité. Chaque système DGX H100 est connecté aux switchs par huit connexions, avec des ports paires connectés à un switch et des ports impaires connectés à l'autre switch.

Pour l'accès au système de stockage, la gestion intrabande et l'accès client, une paire de commutateurs Ethernet SN4600 est utilisée. Les commutateurs sont connectés avec des liaisons inter-commutateurs et configurés avec plusieurs VLAN pour isoler les différents types de trafic. Pour les déploiements de plus grande envergure, le réseau Ethernet peut être étendu à une configuration Leaf-Spine en ajoutant des paires de switchs supplémentaires pour les commutateurs Spine et des lames supplémentaires si nécessaire.

Outre l'interconnexion de calcul et les réseaux Ethernet haut débit, tous les périphériques physiques sont également connectés à un ou plusieurs commutateurs Ethernet SN2201 pour la gestion hors bande. Pour plus d'informations sur la connectivité du système DGX H100, reportez-vous au "Documentation NVIDIA BasePOD".

Configuration client pour l'accès au stockage

Chaque système DGX H100 est provisionné avec deux adaptateurs ConnectX-7 à deux ports pour la gestion et le trafic de stockage. Pour cette solution, les deux ports de chaque carte sont connectés au même switch. Un port de chaque carte est ensuite configuré en une liaison LACP MLAG avec un port connecté à chaque switch, et les VLAN pour la gestion intrabande, l'accès client et l'accès au stockage au niveau utilisateur sont hébergés sur cette liaison.

L'autre port de chaque carte est utilisé pour la connectivité aux systèmes de stockage AFF A900 et peut être utilisé dans plusieurs configurations en fonction des exigences des workloads. Pour les configurations utilisant NFS sur RDMA pour prendre en charge le stockage GPUDirect d'E/S NVIDIA Magnum, les ports sont configurés sur une liaison active/passive, car RDMA n'est pris en charge sur aucun autre type de liaison. Pour les déploiements qui ne nécessitent pas de RDMA, les interfaces de stockage peuvent également être configurées avec des liaisons LACP afin d'offrir une haute disponibilité et une bande passante supplémentaire. Avec ou sans RDMA, les clients peuvent monter le système de stockage à l'aide de NFS v4.1 pNFS et de l'agrégation de sessions afin de permettre un accès parallèle à tous les nœuds de stockage du cluster.

Configuration du système de stockage

Chaque système de stockage AFF A900 est connecté à l'aide de quatre ports 100 GbE depuis chaque contrôleur. Deux ports de chaque contrôleur sont utilisés pour l'accès aux données de workload à partir des systèmes DGX, et deux ports de chaque contrôleur sont configurés en tant que groupe d'interface LACP pour la prise en charge de l'accès depuis les serveurs du plan de gestion pour les artéfacts de gestion du cluster et les répertoires locaux des utilisateurs. Tout accès aux données à partir du système de stockage s'effectue via NFS, avec une machine virtuelle de stockage (SVM) dédiée à l'accès aux workloads d'IA et un SVM distinct dédié aux utilisations du cluster management.

La SVM de workload est configurée avec un total de huit interfaces logiques (LIF), avec deux LIF sur chaque port physique. Cette configuration offre une bande passante maximale ainsi que les moyens pour chaque LIF de basculer vers un autre port du même contrôleur, de sorte que les deux contrôleurs restent actifs en cas de défaillance du réseau. Cette configuration prend également en charge NFS sur RDMA pour activer l'accès au stockage GPUDirect. La capacité de stockage est provisionnée sous la forme d'un grand volume FlexGroup unique qui s'étend à tous les contrôleurs de stockage du cluster, avec 16 volumes constitutifs sur chaque contrôleur. Ce FlexGroup est accessible depuis n'importe quelle LIF du SVM. De plus, grâce à NFSv4.1 avec pNFS et mise en circuit de session, les clients établissent des connexions à chaque LIF du SVM, ce qui permet d'accéder en parallèle aux données locales sur chaque nœud de stockage afin d'améliorer considérablement les performances. Le SVM de charge de travail et chaque LIF de données sont également configurés pour l'accès au protocole RDMA. Pour plus de détails sur la configuration RDMA pour ONTAP, reportez-vous au "Documentation ONTAP".

Le SVM de gestion ne nécessite qu'une seule LIF, hébergée sur les groupes d'interface à 2 ports configurés sur chaque contrôleur. D'autres volumes FlexGroup sont provisionnés sur le SVM de gestion pour héberger les artéfacts de gestion de cluster tels que les images de nœud de cluster, les données historiques de surveillance du système et les répertoires locaux des utilisateurs. Le schéma ci-dessous présente la configuration logique du système de stockage.

NetApp A900 Storage cluster Logical configuration image::aipod_nv_A900Logical.png[]

Serveurs de plan de gestion

Cette architecture de référence comprend également cinq serveurs basés sur processeurs pour l'utilisation d'un plan de gestion. Deux de ces systèmes sont utilisés comme nœuds principaux pour NVIDIA base Command Manager pour le déploiement et la gestion du cluster. Les trois autres systèmes sont utilisés pour fournir des services de cluster supplémentaires tels que les nœuds maîtres Kubernetes ou les nœuds de connexion pour les déploiements utilisant Slurm pour la planification des tâches. Les déploiements qui utilisent

Kubernetes peuvent exploiter le pilote NetApp Astra Trident CSI pour fournir un provisionnement automatisé et des services de données avec un stockage persistant pour la gestion et les workloads d'IA sur le système de stockage AFF A900.

Chaque serveur est connecté physiquement aux switchs IB et Ethernet pour permettre le déploiement et la gestion du cluster, et configuré avec des montages NFS sur le système de stockage via la SVM de gestion pour le stockage des artéfacts de gestion de cluster, comme décrit précédemment.

NetApp AIPod avec les systèmes NVIDIA DGX - conseils de validation et de dimensionnement de la solution

Cette section est consacrée aux conseils sur la validation et le dimensionnement de la solution NetApp AIPod avec les systèmes NVIDIA DGX.

Validation des solutions

La configuration du stockage dans cette solution a été validée à l'aide d'une série de charges de travail synthétiques à l'aide de l'outil open source FIO. Ces tests incluent des modèles d'E/S de lecture et d'écriture destinés à simuler le workload de stockage généré par les systèmes DGX exécutant des tâches d'entraînement de deep learning. La configuration du stockage a été validée à l'aide d'un cluster de serveurs CPU à 2 sockets qui exécutent simultanément les workloads d'E/S pour simuler un cluster de systèmes DGX. Chaque client a été configuré avec la même configuration réseau décrite précédemment, avec l'ajout des détails suivants.

Les options de montage suivantes ont été utilisées pour cette validation :

vers=4.1	Active pNFS pour l'accès parallèle à plusieurs nœuds de stockage
proto=rdma	Définit le protocole de transfert sur RDMA au lieu du TCP par défaut
port=20049	Spécifier le port correct pour le service NFS RDMA
max_connect=16	Permet l'agrégation de la bande passante des ports de stockage via l'agrégation de sessions NFS
write=eager	améliore les performances d'écriture des écritures mises en tampon
rsize=262144,wsize=262144	Définit la taille du transfert d'E/S sur 256 Ko

En outre, les clients ont été configurés avec une valeur NFS max_session_slots de 1024. Comme la solution a été testée à l'aide de NFS sur RDMA, les ports des réseaux de stockage ont été configurés avec une liaison actif-passif. Les paramètres de liaison suivants ont été utilisés pour cette validation :

mode=sauvegarde-active	définit le lien en mode actif/passif
primaire=<interface name>	les interfaces principales de tous les clients ont été distribuées sur les commutateurs
intervalle-moniteur-mii=100	spécifie un intervalle de surveillance de 100 ms.
stratégie-de-basculement-mac=active	Indique que l'adresse MAC de la liaison active est l'adresse MAC de la liaison. Ceci est nécessaire pour le bon fonctionnement de RDMA sur l'interface liée.

Le système de stockage a été configuré comme décrit avec deux paires HA A900 (4 contrôleurs) et deux tiroirs disque NS224 de 24 disques NVMe de 1,9 To reliés à chaque paire haute disponibilité. Comme indiqué dans

la section Architecture, la capacité de stockage de tous les contrôleurs a été combinée à l'aide d'un volume FlexGroup, et les données de tous les clients ont été distribuées sur l'ensemble des contrôleurs du cluster.

Conseils de dimensionnement des systèmes de stockage

NetApp a obtenu la certification DGX BasePOD et les deux paires HA A900 testées peuvent facilement prendre en charge un cluster de huit systèmes DGX H100. Pour les déploiements de plus grande envergure nécessitant des performances de stockage plus élevées, des systèmes AFF supplémentaires peuvent être ajoutés au cluster NetApp ONTAP jusqu'à 12 paires haute disponibilité (24 nœuds) dans un seul cluster. Grâce à la technologie FlexGroup décrite dans cette solution, un cluster à 24 nœuds peut fournir plus de 40 po et un débit pouvant atteindre 300 Gbit/s dans un seul namespace. D'autres systèmes de stockage NetApp, tels que les systèmes AFF A400, A250 et C800, offrent des performances inférieures et/ou des capacités supérieures pour les déploiements de moindre envergure et à moindre coût. Comme ONTAP 9 prend en charge les clusters à modèles mixtes, les clients peuvent commencer avec une empreinte réduite et ajouter au cluster des systèmes de stockage plus nombreux ou plus grands selon l'évolution des besoins en capacité et en performance. Le tableau ci-dessous présente une estimation approximative du nombre de GPU A100 et H100 pris en charge sur chaque modèle AFF.

Conseils de dimensionnement du système de stockage NetApp

		Throughput[2]	Raw capacity *(typical / max)*	Connectivity	# NVIDIA A100 GPUs supported[3]	# NVIDIA H100 GPUs supported[4]
NetApp® AFF A900	**1 HA pair[1]**	28GB/s	182TB / 14.7PB	**100 GbE**	1 - 64	1-32
	12 HA pairs	336GB/s	2.1PB / **176.4PB**		**768**	**384**
AFF A800	**1 HA pair**	25GB/s	368TB / 3.6PB	**100 GbE**	1 - 64	1-32
	12 HA pairs	300GB/s	4.4PB / **43.2PB**		**768**	**384**
AFF C800	**1 HA pair**	21GB/s	368TB / 3.6PB	**100 GbE**	1-48	1-24
	12 HA pairs	252GB/s	4.4PB / **43.2PB**		**576**	**288**
AFF A400	**1 HA pair**	11GB/s	182TB / 14.7PB	**40/100 GbE**	1 - 32	1-16
	12 HA pairs	132GB/s	2.1PB / **176.4PB**		**384**	**192**
AFF C400	**1 HA pair**	8GB/s	182TB / 14.7PB	**40/100 GbE**	1 - 16	1-8
	12 HA pairs	128GB/s	2.1PB / **176.4PB**		**192**	**96**
AFF A250	**1 HA pair**	7.4GB/s	91.2TB / 4.4PB	**25 GbE 40/100GbE**	1 - 16	1-8
	4 HA pairs	29.6GB/s	364.8TB / 17.6PB		64	32
AFF C250	**1 HA pair**	5 GB/s	91.2TB / 4.4PB	**25 GbE 40/100GbE**	1-8	1-4
	4 HA pairs	20 GB/s	364.8TB / 17.6PB		32	8

1 – 1 AFF = 1 HA pair = 2 Nodes. 12 HA pairs = 24 nodes
2 – 100% sequential read
3 – Based on workload testing in NVA-1153
4 – Based on BasePOD validation test results

NetApp AIPod avec les systèmes NVIDIA DGX - conclusion et informations supplémentaires

Cette section contient des références d'informations supplémentaires sur NetApp AIPod avec les systèmes NVIDIA DGX.

Conclusion

L'architecture DGX BasePOD est une plateforme de deep learning nouvelle génération qui requiert des

fonctionnalités avancées de stockage et de gestion des données. En combinant DGX BasePOD avec les systèmes NetApp AFF, l'architecture NetApp AIPod avec les systèmes DGX peut être implémentée à presque toutes les échelles jusqu'à 48 systèmes DGX H100 sur un cluster AFF A900 de 24 nœuds. Associé aux fonctionnalités SDS et d'intégration au cloud d'NetApp ONTAP, AFF couvre l'ensemble des pipelines de traitement de données de la périphérie au cœur jusqu'au cloud pour l'apprentissage profond.

Informations supplémentaires

Pour en savoir plus sur les informations fournies dans ce document, veuillez consulter les documents et/ou sites web suivants :

- Le logiciel de gestion des données NetApp ONTAP : bibliothèque d'informations ONTAP

  "https://docs.netapp.com/us-en/ontap-family/"

- Systèmes de stockage NetApp AFF A900,

  "https://www.netapp.com/data-storage/aff-a-series/aff-a900/"

- Informations NetApp ONTAP RDMA -

  "https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html"

- Kit NetApp DataOps

  "https://github.com/NetApp/netapp-dataops-toolkit"

- NetApp Astra Trident

  "Présentation"

- Blog NetApp GPUDirect Storage :

  "https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/"

- NVIDIA DGX BasePOD

  "https://www.nvidia.com/en-us/data-center/dgx-basepod/"

- SYSTÈMES NVIDIA DGX H100

  "https://www.nvidia.com/en-us/data-center/dgx-h100/"

- Mise en réseau NVIDIA

  "https://www.nvidia.com/en-us/networking/"

- NVIDIA Magnum IO GPUDirect Storage

  "https://docs.nvidia.com/gpudirect-storage"

- Commande de base NVIDIA

  "https://www.nvidia.com/en-us/data-center/base-command/"

- Gestionnaire de commande de base NVIDIA

"https://www.nvidia.com/en-us/data-center/base-command/manager"

- NVIDIA ai Enterprise

  "https://www.nvidia.com/en-us/data-center/products/ai-enterprise/"

Remerciements

Ce document est le travail des équipes de solutions NetApp et d'ingénieurs ONTAP, David Arnette, Olga Kornievskaia, Dustin Fischer, Srikanth Kaligotla, Mohit Kumar et Rajeev Badrinath. Les auteurs tiennent également à remercier NVIDIA et l'équipe d'ingénierie NVIDIA DGX BasePOD pour leur support continu.

NVA-1151-DESIGN : guide de conception de NetApp ONTAP ai avec les systèmes NVIDIA DGX A100

David Arnette et Sung-Han Lin, NetApp

NVA-1151-DESIGN décrit une architecture vérifiée NetApp pour les workloads de machine learning et d'intelligence artificielle à l'aide des systèmes de stockage NetApp AFF A800, des systèmes NVIDIA DGX A100 et des switchs réseau NVIDIA Mellanox. Il inclut également les résultats des tests d'évaluation pour l'architecture mise en œuvre.

"NVA-1151-DESIGN : guide de conception de NetApp ONTAP ai avec les systèmes NVIDIA DGX A100"

NVA-1151-DEPLOY : NetApp ONTAP ai avec systèmes NVIDIA DGX A100

David Arnette, NetApp

NVA-1151-DEPLOY inclut des instructions de déploiement de systèmes de stockage pour un workload d'architecture vérifiée NetApp (NVA) pour le machine learning (ML) et l'intelligence artificielle (IA) à l'aide de systèmes de stockage NetApp AFF A800, de systèmes NVIDIA DGX A100 et de switchs réseau NVIDIA Mellanox. Celui-ci contient également des instructions pour l'exécution des tests de validation une fois le déploiement terminé.

"NVA-1151-DEPLOY : NetApp ONTAP ai avec systèmes NVIDIA DGX A100"

NVA-1153-DESIGN : NetApp ONTAP ai avec des systèmes NVIDIA DGX A100 et des switchs Ethernet Mellanox Spectrum

David Arnette et Sung-Han Lin, NetApp

La conception NVA-1153 décrit une architecture vérifiée NetApp pour les workloads de machine learning (ML) et d'intelligence artificielle (IA) utilisant les systèmes de stockage NetApp AFF A800, les systèmes NVIDIA DGX A100 et les switchs Ethernet de 200 Go

NVIDIA Mellanox Spectrum SN3700V. Cette conception comprend le protocole RDMA over Converged Ethernet (RoCE) pour la structure d'interconnexion des clusters de calcul. Elle offre aux clients une architecture entièrement basée sur ethernet pour les charges de travail haute performance. Ce document inclut également les résultats des tests de performance pour l'architecture mise en œuvre.

"NVA-1153-DESIGN : NetApp ONTAP ai avec des systèmes NVIDIA DGX A100 et des switchs Ethernet Mellanox Spectrum"

NVA-1153-DEPLOY : NetApp ONTAP ai avec systèmes NVIDIA DGX A100 et switchs Ethernet Mellanox Spectrum

David Arnette, NetApp

NVA-1153-DEPLOY inclut des instructions de déploiement de système de stockage pour un workload d'architecture vérifiée NetApp pour le machine learning (ML) et l'intelligence artificielle (IA) utilisant les systèmes de stockage NetApp AFF A800, les systèmes NVIDIA DGX A100 et les switchs Ethernet NVIDIA Mellanox Spectrum SN3700V 200 Gb. Celui-ci contient également des instructions pour l'exécution des tests de validation

"NVA-1153-DEPLOY : NetApp ONTAP ai avec systèmes NVIDIA DGX A100 et switchs Ethernet Mellanox Spectrum"

Informations sur le copyright

Informations sur les marques commerciales