

Prise en charge des serveurs Cisco AI UCS contenant des processeurs graphiques Nvidia

Table des matières

[Introduction](#)

[Informations générales](#)

[Présentation du logiciel NVIDIA AI Enterprise](#)

[Qu'est-ce que NVIDIA AI Enterprise ?](#)

[Composants clés](#)

[Modèles de déploiement](#)

[Avantages de l'assistance avec licence NVIDIA AI Enterprise](#)

[Informations de commande](#)

[Comment faire appel au support](#)

[Quand contacter l'assistance Cisco](#)

[Quand contacter l'assistance NVIDIA](#)

[Prise en charge d'autres logiciels liés à AI](#)

Introduction

Ce document décrit la suite logicielle d'entreprise NVIDIA AI (NVIE) ainsi que les avantages d'assistance complets disponibles lorsqu'elle est déployée sur des plates-formes Cisco Unified Computing System (UCS). Ce document décrit également comment faire appel à l'assistance Cisco et NVIDIA pour un déploiement optimal et une assistance opérationnelle.

Informations générales

Les charges de travail liées à l'intelligence artificielle et à l'apprentissage automatique devenant critiques pour les opérations de l'entreprise, les entreprises ont besoin non seulement d'une infrastructure puissante, mais également de logiciels et d'une assistance de qualité professionnelle. NVIDIA AI Enterprise Software offre une plate-forme d'IA complète, sécurisée et prise en charge, optimisée pour le déploiement sur l'infrastructure Cisco UCS.

Ce document décrit :

- Ce que NVIDIA AI Enterprise Software implique
- Avantages de la licence de la suite logicielle NVAIE
- Comment utiliser correctement les canaux d'assistance pour NVAIE sur Cisco UCS

Présentation du logiciel NVIDIA AI Enterprise

Qu'est-ce que NVIDIA AI Enterprise ?

NVIDIA AI Enterprise (NVIE) est une suite logicielle complète qui offre une sécurité, une stabilité, une facilité de gestion et une prise en charge de niveau entreprise pour les charges de travail IA exécutées dans des environnements virtualisés et natifs du cloud. Il est optimisé et certifié pour fonctionner sur les plates-formes d'entreprise de pointe, notamment Cisco UCS.

Composants clés



Remarque : Pour obtenir une liste complète et actualisée, visitez [Nvidia AI Enterprise](#)

Cadres et outils d'IA :

- Flux de travail NVIDIA AI - Flux de travail prédéfinis de bout en bout pour les applications d'IA courantes
- NVIDIA TAO Toolkit - Boîte à outils d'apprentissage par transfert pour la vision par ordinateur, l'IA conversationnelle, et plus encore
- NVIDIA Riva - SDK pour la création d'applications d'IA vocale
- NVIDIA Morpheus - Cadre d'IA de cybersécurité
- NVIDIA Metropolis - Plate-forme Vision AI pour l'analytique vidéo intelligente
- NVIDIA Clara - Cadre d'application IA pour la santé

Logiciels d'infrastructure :

- Logiciel vGPU NVIDIA - Technologie de virtualisation GPU (sous licence séparée ou incluse avec NVIE)
- NVIDIA GPU Operator - Opérateur Kubernetes pour la gestion des GPU
- NVIDIA Network Operator - Gère les ressources réseau NVIDIA dans Kubernetes
- NVIDIA Container Toolkit - Outils de création et d'exécution de conteneurs accélérés par GPU

Pile de développement AI :

- Structures d'IA optimisées (TensorFlow, PyTorch, JAX, etc.)
- NVIDIA RAPIDS pour la science des données
- Serveur d'inférence NVIDIA Triton pour le déploiement d'IA de production
- NVIDIA TensorRT pour une inférence optimisée
- Bibliothèques NVIDIA cuDNN, cuBLAS et CUDA

Gestion et sécurité :

- NVIDIA Base Command Manager Essentials - Gestion de l'infrastructure pour l'IA
- NVIDIA License System - Gestion centralisée des licences
- Correctifs de sécurité et versions LTS (Long Term Support)

Modèles de déploiement

1. Environnements virtualisés - VMware vSphere, virtualisation Red Hat, processeur graphique virtuel NVIDIA
2. Sans système d'exploitation : déploiement direct sur des serveurs physiques
3. Cloud-Native - Kubernetes et déploiements en conteneur
4. Hybride et multi-cloud : plate-forme d'IA cohérente dans tous les environnements

Avantages de l'assistance avec licence NVIDIA AI Enterprise

Assistance d'entreprise complète

En concédant une licence à NVIDIA AI Enterprise, vous bénéficiez d'avantages d'assistance significatifs qui vont au-delà de l'accès aux pilotes GPU de base.

1. Assistance logicielle de niveau production
 - Mises à jour logicielles et correctifs - Mises à jour de sécurité régulières, correctifs de bogues et améliorations des fonctionnalités
 - Versions de support à long terme - Versions LTS stables avec fenêtres de support étendues (jusqu'à 2 ans)



Remarque : Les niveaux d'assistance Nvidia (Basic vs 24x7) sont basés sur la licence spécifique achetée comme indiqué [ici](#)

Composant de service	Norme	24x7
Accès au support technique	Heures ouvrables locales	24 heures sur 24, 7 jours sur 7
Temps de réponse initial de gravité 1	4 heures ouvrables	1 heure
Temps de réponse initial de gravité 2	4 heures ouvrables	2 heures
Temps de réponse initial de gravité 3	1 jour ouvrable	4 heures ouvrables
Gravité 4 Délais de réponse initiaux	2 jours ouvrables	1 jour ouvrable
Accès au portail client	24x7	24x7
Web	24x7	24x7
Téléphone	Heures ouvrables locales	24x7

2. Expertise et conseils techniques

- Assistance au déploiement AI - Aide à la conception d'architecture et aux meilleures pratiques

- Optimisation des performances : conseils sur le réglage des charges de travail IA pour des performances maximales
- Prise en charge du cadre - Assistance avec TensorFlow, PyTorch et d'autres cadres d'IA inclus dans la suite
- Assistance pour les conteneurs et Kubernetes - Aide pour les déploiements d'IA natifs du cloud
- Accès à la base de connaissances - Documentation, guides et ressources de dépannage complets

3. Licences et avantages d'accès

- Logiciels vérifiés et validés - Versions testées et certifiées par l'entreprise
- Portabilité des licences : possibilité de déplacer les licences dans l'infrastructure
- Gestion centralisée des licences - Système de licence NVIDIA pour la gestion des déploiements
- Modèle d'abonnement - Options de licence annuelle ou pluriannuelle
- Software Download Access - Portail de téléchargement officiel pour tous les composants NVAIE

4. Sécurité et conformité

- Correctifs de sécurité - Correction rapide des vulnérabilités
- Suivi CVE - Notification proactive des problèmes de sécurité
- Documentation de conformité - Prise en charge des exigences réglementaires
- Chaîne logistique logicielle sécurisée - Composants logiciels vérifiés et signés

Couverture de soutien

Fonctionnalité	NVIDIA AI Enterprise
Accès au pilote GPU	✓
Documentation de base	✓
Forums communautaires	✓
Suite logicielle AI	✓
Assistance aux entreprises (24h/24, 7j/7)	✓
Temps de réponse SLA	✓

Fonctionnalité	NVIDIA AI Enterprise
Correctifs de sécurité	✓
Assistance à long terme	✓
Certification de production	✓
Gestion de compte technique	Module complémentaire facultatif
Accès au workflow AI	✓
Outils de gestion	✓

Informations de commande

Cisco UCS avec NVIDIA AI Enterprise

Étape 1: Sélectionner le matériel Cisco UCS

Choisissez la plate-forme Cisco UCS et la configuration de processeur graphique NVIDIA appropriées :

Exemples de configuration :

- UCS C885A M8 avec processeurs graphiques NVIDIA HGX H200
- UCS C845A M8 avec processeurs graphiques NVIDIA RTX6000
- UCS C880A M8 avec processeurs graphiques NVIDIA HGX B300
- UCS C240 M7 avec processeurs graphiques NVIDIA L40S
- UCS série X avec processeurs graphiques NVIDIA H100 pour des déploiements modulaires



Remarque : Chaque processeur graphique NVIDIA H200 NVL, H100 PCIe et H100 NVL est fourni avec un abonnement de cinq ans à NVIDIA Enterprise, qui inclut NVIDIA AI Enterprise. L'activation logicielle est requise. Vous devez activer la licence et ouvrir un dossier d'assistance logicielle avec NVIDIA si nécessaire.

Cliquez [ici](#) pour plus de détails.

Par exemple, UCS C845A M8, UCSX-210C-M8, UCS C240 M7 avec processeurs

graphiques NVIDIA H200 incluent un abonnement de cinq ans à NVIDIA Enterprise.

Étape 2: Licence d'entreprise NVIDIA AI

Les licences NVIDIA AI Enterprise sont vendues séparément et peuvent être achetées auprès de :

- Cisco Commerce - Offre groupée avec les achats Cisco UCS
-



Remarque : La licence NVIDIA peut également être achetée en dehors de Cisco.

Types de licence :

- NVAIE- Standard - Abonnement annuel standard (par socket GPU ou par utilisateur)
- NVAIE - Enseignement - Abonnement annuel à l'enseignement (par interface de connexion GPU ou par utilisateur)
- NVAIE-vGPU - Licence combinée NVAIE + vGPU

Pour plus d'informations, consultez [le guide de commande NVIDIA](#)



Remarque : Pour connaître les références et les tarifs spécifiques, contactez votre équipe de compte Cisco ou votre représentant commercial NVIDIA.

Étape 3: Options de support

Assurez-vous d'avoir :

- Contrat d'assistance Cisco UCS - Assistance Cisco (standard, amélioré, signature)
- Abonnement NVIDIA AI Enterprise - Prise en charge NVIDIA incluse

Configuration de référence

Exemple de commande d'infrastructure d'IA :

1. Matériel : UCS C885A M8 avec GPU NVIDIA H200 (8x H200)
1. le logiciel Cisco IOS: NVIDIA AI Enterprise (8 licences GPU)
2. Soutien:
 - Assistance Cisco - Améliorée pour le matériel UCS
 - Abonnement NVIDIA AI Enterprise (inclut une prise en charge 24 h/24, 7 j/7)
3. Services : Services professionnels en option pour le déploiement

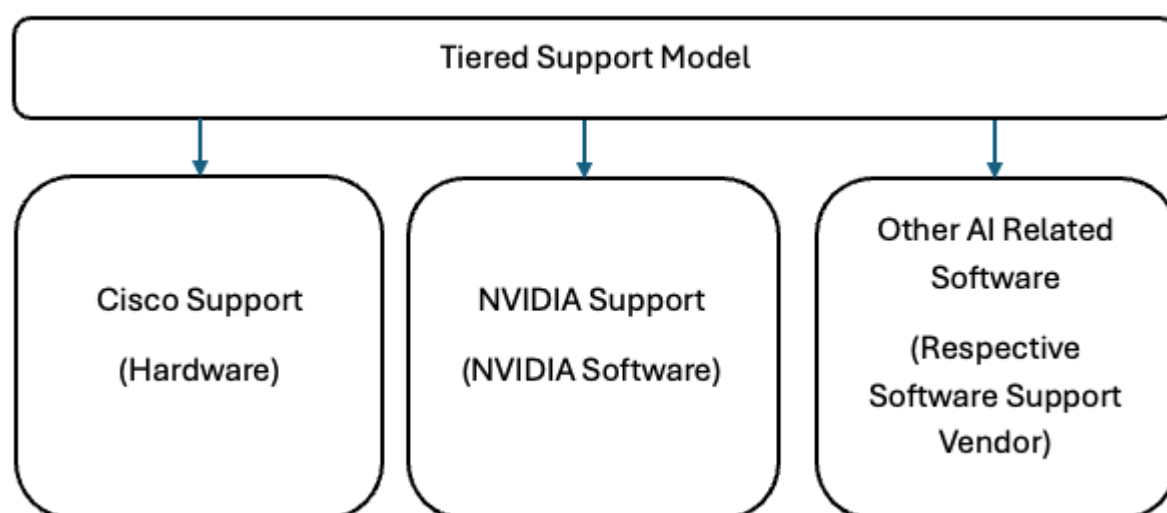
Pour obtenir des guides de commande détaillés, reportez-vous à :

- [Guide de commande NVIDIA AI Enterprise pour Cisco UCS](#)

Comment faire appel au support

Présentation du modèle de support

NVIDIA AI Enterprise sur Cisco UCS adhère à un modèle d'assistance hiérarchisée qui inclut une distinction claire entre l'assistance matérielle et l'assistance logicielle :



Remarque : Pour les problèmes de logiciel NVIDIA, vous devez créer un dossier d'assistance directement auprès de NVIDIA. Si nécessaire, veuillez également inclure l'assistance Cisco pour les efforts de dépannage conjoints.

Quand contacter l'assistance Cisco

- Problèmes matériels :
 - Défaillances ou erreurs du serveur Cisco UCS
 - Diagnostics et remplacement du matériel
 - Mises à jour du micrologiciel pour les composants UCS
 - Problèmes de configuration d'UCS Manager ou Intersight
 - Problèmes d'installation ou de détection de GPU physique
 - Problèmes de connectivité réseau avec l'infrastructure UCS
- Logiciels Cisco :
 - Problèmes liés à UCS Manager ou Cisco Intersight
 - Configuration réseau Cisco pour l'infrastructure AI

Comment ouvrir un dossier d'assistance Cisco :

1. Visite : [Coordonnées du service d'assistance Cisco à l'échelle mondiale](#)

2. Appelez directement l'assistance Cisco (hotline 24/7)
3. Utiliser Cisco Support Case Manager
4. Contact via Cisco Intersight (pour les périphériques connectés au cloud)

Informations requises avant la création du dossier d'assistance :

- Modèle et numéro de série Cisco UCS
- Numéro de contrat de service
- Description détaillée du problème
- Fichiers d'assistance technique (show tech-support)
- Inventaire et configuration des processeurs graphiques

Quand contacter l'assistance NVIDIA

Logiciel d'entreprise NVIDIA AI :

- Problèmes d'installation ou de configuration du cadre AI
- Questions de workflow NVIDIA AI
- Déploiement de Triton Inference Server
- Problèmes de logiciel vGPU NVIDIA
- Problèmes d'opérateur GPU ou d'opérateur réseau
- Problèmes d'exécution du conteneur avec NVIDIA Container Toolkit
- Activation ou gestion de licence
- Réglage des performances pour les charges de travail IA
- Bogues logiciels ou demandes de fonctionnalités
- Questions de sécurité et de vulnérabilité

Problèmes de pilote et de logiciel :

- Installation ou compatibilité du pilote NVIDIA
- Problèmes de boîte à outils CUDA
- Problèmes avec cuDNN ou TensorRT
- Problèmes d'utilisation ou de performances GPU

Comment ouvrir un dossier :

1. Visitez le portail d'[assistance](#) NVIDIA Enterprise
2. Connectez-vous avec votre compte NVIDIA Enterprise
3. Sélectionnez « Créer un dossier » et indiquez :
 - Informations sur la licence NVIDIA AI Enterprise
 - Détails du serveur et du processeur graphique
 - Versions logicielles
 - Description détaillée du problème
 - Journaux et résultats de diagnostic

Accès au portail de support :

- NVIDIA AI Enterprise requiert un abonnement actif
- Configuration du compte lors du provisionnement des licences
- Accès à la base de connaissances, aux téléchargements et à la gestion des dossiers

Prise en charge d'autres logiciels liés à AI

Respectez les procédures d'assistance fournies par le Fournisseur de logiciels/d'applications d'IA concerné

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. Si votre réseau est en ligne, assurez-vous de bien comprendre l'incidence possible des commandes.

À propos de cette traduction

Cisco a traduit ce document en traduction automatisée vérifiée par une personne dans le cadre d'un service mondial permettant à nos utilisateurs d'obtenir le contenu d'assistance dans leur propre langue.

Il convient cependant de noter que même la meilleure traduction automatisée ne sera pas aussi précise que celle fournie par un traducteur professionnel.