

Ondersteuning inschakelen voor Cisco AI UCS-servers met Nvidia-GPU's

Inhoud

[Inleiding](#)

[Achtergrondinformatie](#)

[Overzicht van NVIDIA AI Enterprise-software](#)

[Wat is NVIDIA AI Enterprise?](#)

[Belangrijkste onderdelen](#)

[Implementatiemodellen](#)

[Ondersteuningsvoordelen met NVIDIA AI Enterprise-licentie](#)

[Bestelgegevens](#)

[Hoe om ondersteuning te betrekken](#)

[Wanneer contact opnemen met CiscoSupport](#)

[Wanneer contact opnemen met NVIDIA Support](#)

[Ondersteuning voor andere AI-gerelateerde software](#)

Inleiding

In dit document wordt de NVIDIA AI Enterprise Software (NVAIE) Suite beschreven, evenals de uitgebreide ondersteuningsvoordelen die beschikbaar zijn wanneer deze op Cisco Unified Computing System (UCS)-platforms worden geïmplementeerd. In dit document wordt ook beschreven hoe u zowel Cisco- als NVIDIA-ondersteuning kunt inschakelen voor een optimale implementatie en operationele ondersteuning.

Achtergrondinformatie

Aangezien kunstmatige intelligentie en machine learning-workloads bedrijfskritiek worden voor bedrijfsactiviteiten, hebben organisaties niet alleen een krachtige infrastructuur nodig, maar ook software en ondersteuning op bedrijfsniveau. NVIDIA AI Enterprise Software biedt een uitgebreid, veilig en ondersteund AI-platform dat is geoptimaliseerd voor implementatie op de Cisco UCS-infrastructuur.

Dit document beschrijft:

- Wat houdt NVIDIA AI Enterprise Software in?
- De voordelen van het licentiëren van de NVAIE Software Suite
- Hoe u ondersteuningskanalen voor NVAIE op Cisco UCS correct kunt inschakelen

Overzicht van NVIDIA AI Enterprise-software

Wat is NVIDIA AI Enterprise?

NVIDIA AI Enterprise (NVAIE) is een uitgebreide softwaresuite die beveiliging, stabiliteit, beheerbaarheid en ondersteuning op bedrijfsniveau biedt voor AI-workloads die worden uitgevoerd in gevirtualiseerde en cloud-native omgevingen. Het is geoptimaliseerd en gecertificeerd voor gebruik op toonaangevende bedrijfsplatforms, waaronder Cisco UCS.

Belangrijkste onderdelen



Opmerking: Ga voor een actuele en uitgebreide lijst naar [Nvidia AI Enterprise](#)

AI-frameworks en -tools:

- NVIDIA AI-workflows - Vooraf gebouwde end-to-end workflows voor algemene AI-toepassingen
- NVIDIA TAO Toolkit - Transfer learning toolkit voor computervisie, conversationele AI en meer
- NVIDIA Riva - SDK voor het bouwen van spraak-AI-toepassingen
- NVIDIA Morpheus - AI-framework voor cyberbeveiliging
- NVIDIA Metropolis - Vision AI-platform voor intelligente videoanalyse
- NVIDIA Clara - Toepassingskader voor AI in de gezondheidszorg

Infrastructuursoftware:

- NVIDIA vGPU-software - GPU-virtualisatietechnologie (afzonderlijk gelicentieerd of meegeleverd met NVAIE)
- NVIDIA GPU Operator - Kubernetes operator voor GPU-beheer
- NVIDIA Network Operator - beheert NVIDIA-netwerkbronnen in Kubernetes
- NVIDIA Container Toolkit - Tools voor het bouwen en uitvoeren van GPU-versnelde containers

AI Development Stack:

- Geoptimaliseerde AI-frameworks (TensorFlow, PyTorch, JAX, enzovoort)
- NVIDIA RAPIDS voor data science
- NVIDIA Triton Inference Server voor productie-AI-implementatie
- NVIDIA TensorRT voor geoptimaliseerde inferentie
- NVIDIA cuDNN-, cuBLAS- en CUDA-bibliotheken

Beheer en beveiliging:

- NVIDIA Base Command Manager Essentials - Infrastructuurbeheer voor AI
- NVIDIA-licentiesysteem - Gecentraliseerd licentiebeheer
- Beveiligingspatches en LTS-versies (Long-Term Support)

Implementatiemodellen

1. Gevirtualiseerde omgevingen - VMware vSphere, Red Hat-virtualisatie, NVIDIA Virtual GPU
2. Bare Metal - Directe implementatie op fysieke servers
3. Cloud-native - Kubernetes en gecontaineriseerde implementaties
4. Hybride en multi-cloud - consistent AI-platform in verschillende omgevingen

Ondersteuningsvoordelen met NVIDIA AI Enterprise-licentie

Uitgebreide Enterprise Support

Door NVIDIA AI Enterprise te licentiëren, ontvangt u aanzienlijke supportvoordelen die verder gaan dan eenvoudige GPU-drivertoegang.

1. Softwaresupport van productiekwaliteit
 - Software-updates en -patches - Regelmatige beveiligingsupdates, bugfixes en functieverbeteringen
 - Langetermijnversies van ondersteuning - Stabiele LTS-versies met uitgebreide ondersteuningsvensters (tot 2 jaar)



Opmerking: Nvidia-ondersteuningsniveaus (Basic versus 24x7) zijn gebaseerd op de specifieke licentie die is gekocht zoals [hier](#) wordt beschreven

servicecomponent	standaard	24x7
Toegang tot technische support	Lokale kantooruren	24 uur, 7 dagen per week
Ernst 1 Initiële responstijden	4 kantooruren	1 uur
Ernst 2 Initiële responstijden	4 kantooruren	2 uur
Ernst 3 Initiële responstijden	1 werkdag	4 kantooruren
Ernst 4 Initiële responstijden	2 werkdagen	1 werkdag
Toegang tot het klantenportaal	24x7	24x7
Web	24x7	24x7
Telefoon	Lokale kantooruren	24x7

2. Technische expertise en begeleiding

- AI Deployment Assistance - Hulp bij het ontwerpen van architectuur en best practices
- Optimalisatie van prestaties - Richtlijnen voor het afstemmen van AI-werklasten voor maximale prestaties

- Framework Support - Hulp bij TensorFlow, PyTorch en andere AI-frameworks die in de suite zijn opgenomen
- Ondersteuning voor containers en Kubernetes - Hulp bij AI-implementaties in de cloud
- Toegang tot Knowledge Base - Uitgebreide documentatie, handleidingen en bronnen voor probleemoplossing

3. Voordelen van licenties en toegang

- Geverifieerde en gevalideerde software - op ondernemingen geteste en gecertificeerde releases
- Licentieportabiliteit - Flexibiliteit om licenties over de infrastructuur te verplaatsen
- Gecentraliseerd licentiebeheer - NVIDIA-licentiesysteem voor het beheer van implementaties
- Abonnementsmodel - Jaarlijkse of meerjarige licentieopties
- Software Download Access - Officiële downloadportal voor alle NVAIE-componenten

4. Beveiliging en naleving

- Beveiligingspatches - Tijdige oplossing voor kwetsbaarheden
- CVE Tracking - Proactieve melding van beveiligingsproblemen
- Nalevingsdocumentatie - Ondersteuning voor wettelijke vereisten
- Secure Software Supply Chain - Geverifieerde en ondertekende softwarecomponenten

Ondersteuningsdekking

Feature	NVIDIA AI Enterprise
Toegang tot GPU-stuurprogramma	✓
basisdocumentatie	✓
Communityforums	✓
AI-software suite	✓
Enterprise Support (24/7)	✓
SLA-responstijden	✓
Beveiligingspatches	✓

Feature	NVIDIA AI Enterprise
ondersteuning op lange termijn	✓
productiecertificering	✓
Technisch accountbeheer	Optionele invoegtoepassing
Toegang tot AI-workflow	✓
Beheertools	✓

Bestelgegevens

Cisco UCS met NVIDIA AI Enterprise

Stap 1: Selecteer Cisco UCS-hardware

Kies het juiste Cisco UCS-platform en de NVIDIA GPU-configuratie:

Voorbeeldconfiguraties:

- UCS C885A M8 met NVIDIA HGX H200-GPU's
- UCS C845A M8 met NVIDIA RTX6000-GPU's
- UCS C880A M8 met NVIDIA HGX B300-GPU's
- UCS C240 M7 met NVIDIA L40S-GPU's
- UCS X-serie met NVIDIA H100 GPU's voor modulaire implementaties



Opmerking: elke NVIDIA H200 NVL-, H100 PCIe- en H100 NVL-GPU wordt geleverd met een vijfjarig NVIDIA Enterprise-abonnement, inclusief NVIDIA AI Enterprise. Softwareactivering is vereist. U moet de licentie activeren en indien nodig een softwareondersteuningscase openen met NVIDIA.

Klik [hier](#) voor meer details.

Bijvoorbeeld, UCS C845A M8, UCSX-210C-M8, UCS C240 M7 met NVIDIA H200 GPU's bevatten een vijfjarig NVIDIA Enterprise-abonnement.

Stap 2: NVIDIA AI Enterprise-licentie

NVIDIA AI Enterprise-licenties worden afzonderlijk verkocht en kunnen worden gekocht via:

- Cisco Commerce - Gebundeld met Cisco UCS-aankopen



Opmerking: de NVIDIA-licentie kan ook buiten Cisco worden gekocht.

Typen licenties:

- NVAIE-standaard - standaard jaarabonnement (per GPU-socket of per gebruiker)
- NVAIE-Educational - Educatief jaarabonnement (per GPU-socket of per gebruiker)
- NVAIE-vGPU - Gecombineerde NVAIE + vGPU-licentie

Ga voor meer informatie naar [de NVIDIA-bestelgids](#)



Opmerking: neem voor specifieke onderdeelnummers en prijzen contact op met uw Cisco-accountteam of NVIDIA-verkoper.

Stap 3: Ondersteuningsopties

Zorg ervoor dat u:

- Cisco UCS-supportcontract – Cisco Support (standaard, uitgebreid, handtekening)
- NVIDIA AI Enterprise-abonnement - Inclusief NVIDIA-ondersteuning

Referentieconfiguratie

Voorbeeldorder voor AI-infrastructuur:

1. Hardware: UCS C885A M8 met NVIDIA H200-GPU's (8x H200)
1. Software: NVIDIA AI Enterprise (8 GPU-licenties)
2. Ondersteuning:
 - Cisco Support — Verbeterde ondersteuning voor UCS-hardware
 - NVIDIA AI Enterprise-abonnement (inclusief 24/7 ondersteuning)
3. Services: optionele professionele services voor implementatie

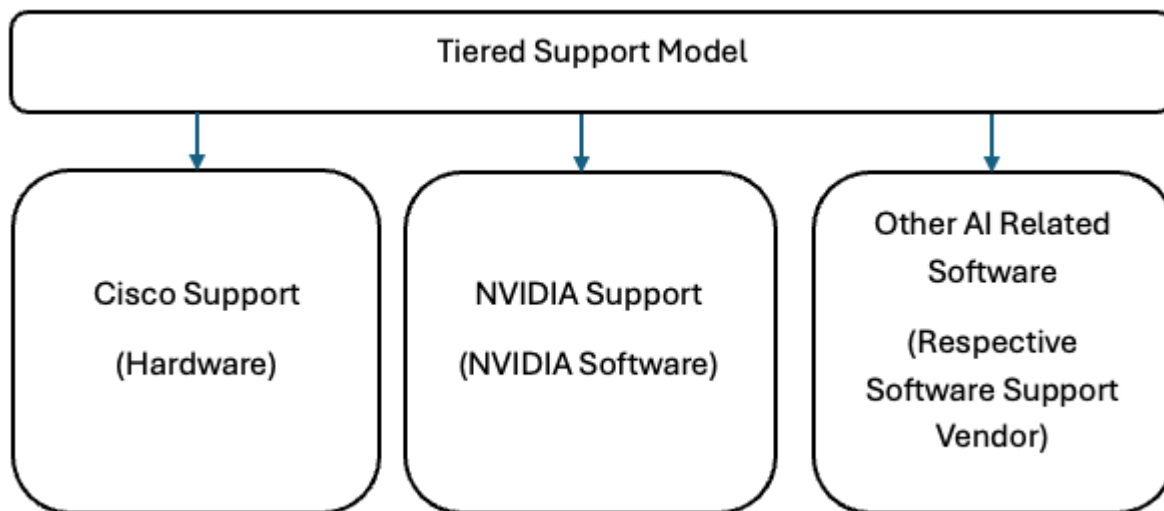
Voor gedetailleerde bestelgidsen raadpleegt u:

- [NVIDIA AI Enterprise-bestelgids voor Cisco UCS](#)

Hoe om ondersteuning te betrekken

Overzicht van supportmodel

NVIDIA AI Enterprise op Cisco UCS hanteert een gelaagd supportmodel dat een duidelijke afbakening tussen hardware- en softwaresupport omvat:



Opmerking: voor NVIDIA-softwareproblemen moet u rechtstreeks met NVIDIA een supportcase maken. Voeg, indien nodig, ook Cisco Support toe voor gezamenlijke inspanningen voor probleemoplossing.

Wanneer contact opnemen met Cisco Support

- Hardwareproblemen:
 - Cisco UCS-serverfouten of -fouten
 - Hardwarediagnostiek en vervanging
 - Firmware-updates voor UCS-onderdelen
 - Configuratieproblemen met UCS Manager of Intersight
 - Fysieke GPU-installatie of detectieproblemen
 - Problemen met netwerkconnectiviteit met UCS-infrastructuur
- Cisco-software:
 - Problemen met UCS Manager of Cisco Intersight
 - Cisco-netwerkconfiguratie voor AI-infrastructuur

Hoe open ik een Cisco Support Case:

1. Bezoek: [Contactpersonen voor wereldwijde ondersteuning van Cisco](#)
2. Bel Cisco Support rechtstreeks (24/7 hotline)
3. Cisco Support Case Manager gebruiken
4. Contact via Cisco Intersight (voor apparaten die in de cloud zijn aangesloten)

Vereiste informatie voordat de supportcase wordt gemaakt:

- Cisco UCS-model en serienummer
- Servicecontractnummer
- Gedetailleerde problembeschrijving
- Bestanden voor technische ondersteuning (toon technische ondersteuning)
- GPU-inventaris en configuratiegegevens

Wanneer contact opnemen met NVIDIA Support

NVIDIA AI Enterprise-software:

- Problemen met de installatie of configuratie van het AI-framework
- Vragen over de NVIDIA AI-workflow
- Triton Inference Server-implementatie
- NVIDIA vGPU-softwareproblemen
- Problemen met GPU-operator of netwerkoperator
- Problemen met de runtime van containers met de NVIDIA Container Toolkit
- Licentie activeren of beheren
- Prestaties afstemmen voor AI-werkbelastingen
- Softwarefouten of functieverzoeken
- Veiligheidskwetsbaarheidsvragen

Problemen met stuurprogramma en software:

- Installatie of compatibiliteit van NVIDIA-stuurprogramma
- CUDA-toolkit problemen
- cuDNN- of TensorRT-problemen
- GPU-gebruik of prestatieproblemen

Hoe een zaak te openen:

1. Bezoek het NVIDIA Enterprise [Support Portal](#)
2. Log in met uw NVIDIA Enterprise-account
3. Selecteer "Case maken" en geef het volgende op:
 - NVIDIA AI Enterprise-licentiegegevens
 - Server- en GPU-gegevens
 - Softwareversies
 - Gedetailleerde beschrijving van probleem
 - Logboeken en diagnostische uitvoer

Toegang tot ondersteuningsportal:

- Vereist actief NVIDIA AI Enterprise-abonnement
- Account instellen tijdens licentietoewijzing
- Toegang tot kennisbank, downloads en casemanagement

Ondersteuning voor andere AI-gerelateerde software

Houd u aan de ondersteuningsprocedures van de toepasselijke leverancier van AI-software/apps

De informatie in dit document is gebaseerd op de apparaten in een specifieke laboratoriumomgeving. Alle apparaten die in dit document worden beschreven, hadden een opgeschoonde (standaard)configuratie. Als uw netwerk live is, moet u zorgen dat u de potentiële impact van elke opdracht begrijpt.

Over deze vertaling

Cisco heeft dit document vertaald via een combinatie van machine- en menselijke technologie om onze gebruikers wereldwijd ondersteuningscontent te bieden in hun eigen taal. Houd er rekening mee dat zelfs de beste machinevertaling niet net zo nauwkeurig is als die van een professionele vertaler. Cisco Systems, Inc. is niet aansprakelijk voor de nauwkeurigheid van deze vertalingen en raadt aan altijd het oorspronkelijke Engelstalige document (link) te raadplegen.