

Unterstützung für Cisco AI UCS-Server mit NVIDIA GPUs

Inhalt

[Einleitung](#)

[Hintergrundinformationen](#)

[NVIDIA AI Enterprise-Software - Überblick](#)

[Was ist NVIDIA AI Enterprise?](#)

[Wichtige Komponenten](#)

[Bereitstellungsmodelle](#)

[Support-Vorteile mit NVIDIA AI Enterprise-Lizenz](#)

[Bestellinformationen](#)

[So erreichen Sie den Support](#)

[Kontaktaufnahme mit dem Cisco Support](#)

[Wann Sie NVIDIA Support kontaktieren](#)

[Unterstützung für andere KI-bezogene Software](#)

Einleitung

In diesem Dokument werden die NVIDIA AI Enterprise Software (NVAIE) Suite sowie die umfassenden Supportleistungen bei Bereitstellung auf Cisco Unified Computing System (UCS)-Plattformen beschrieben. In diesem Dokument wird auch beschrieben, wie Sie den Support von Cisco und NVIDIA in Anspruch nehmen, um optimale Unterstützung bei Bereitstellung und Betrieb zu erhalten.

Hintergrundinformationen

Da künstliche Intelligenz und Workloads für maschinelles Lernen zu einem geschäftskritischen Faktor werden, benötigen Unternehmen nicht nur eine leistungsstarke Infrastruktur, sondern auch Software und Support der Enterprise-Klasse. Die NVIDIA AI Enterprise-Software bietet eine umfassende, sichere und unterstützte KI-Plattform, die für die Bereitstellung in der Cisco UCS-Infrastruktur optimiert ist.

In diesem Dokument wird Folgendes beschrieben:

- Vorteile der NVIDIA AI Enterprise Software
- Die Vorteile der Lizenzierung der NVAIE Software Suite
- Wie können die Support-Kanäle für NVAIE auf dem Cisco UCS richtig genutzt werden?

NVIDIA AI Enterprise-Software - Überblick

Was ist NVIDIA AI Enterprise?

NVIDIA AI Enterprise (NVAIE) ist eine umfassende Software-Suite, die Sicherheit, Stabilität, Verwaltbarkeit und Unterstützung für AI-Workloads der Enterprise-Klasse bietet, die in virtualisierten und Cloud-nativen Umgebungen ausgeführt werden. Die Lösung ist für die Nutzung auf führenden Unternehmensplattformen wie dem Cisco UCS optimiert und zertifiziert.

Wichtige Komponenten



Anmerkung: Eine aktuelle und vollständige Liste finden Sie unter [Nvidia AI Enterprise](#)

KI-Frameworks und -Tools:

- NVIDIA AI Workflows - Vorgefertigte End-to-End-Workflows für gängige KI-Anwendungen
- NVIDIA TAO Toolkit - Transfer Learning Toolkit für Computervision, konversationale KI und mehr
- NVIDIA Riva - SDK für die Erstellung von Sprach-KI-Anwendungen
- NVIDIA Morpheus - KI-Framework für Cybersicherheit
- NVIDIA Metropolis - Vision AI-Plattform für intelligente Videoanalysen
- NVIDIA Clara - KI-Anwendungs-Framework für das Gesundheitswesen

Infrastruktursoftware:

- NVIDIA vGPU Software - GPU-Virtualisierungstechnologie (separat lizenziert oder im Lieferumfang von NVIDIA enthalten)
- NVIDIA GPU Operator - Kubernetes Operator für GPU Management
- NVIDIA Network Operator - verwaltet NVIDIA-Netzwerkressourcen in Kubernetes
- NVIDIA Container Toolkit - Werkzeuge zum Bau und Betrieb von GPU-beschleunigten Containern

KI-Entwicklungspaket:

- Optimierte KI-Frameworks (TensorFlow, PyTorch, JAX usw.)
- NVIDIA RAPIDS für Datenwissenschaften
- NVIDIA Triton Inference Server für die Bereitstellung von Produktions-KI
- NVIDIA TensorRT für optimierte Inferenz
- NVIDIA cuDNN, cuBLAS und CUDA-Bibliotheken

Management und Sicherheit:

- NVIDIA Base Command Manager Essentials - Infrastrukturmanagement für KI
- NVIDIA Lizenzsystem - Zentrales Lizenzmanagement
- Sicherheits-Patches und LTS-Versionen (Long Term Support)

Bereitstellungsmodelle

1. Virtualisierte Umgebungen - VMware vSphere, Red Hat Virtualization, NVIDIA Virtual GPU
2. Bare Metal - Direkte Bereitstellung auf physischen Servern
3. Cloud-nativ - Kubernetes und containerisierte Bereitstellungen
4. Hybrid- und Multi-Cloud - konsistente KI-Plattform für alle Umgebungen

Support-Vorteile mit NVIDIA AI Enterprise-Lizenz

Umfassender Enterprise Support

Durch die Lizenzierung von NVIDIA AI Enterprise profitieren Sie von umfangreichen Supportleistungen, die über den grundlegenden GPU-Treiberzugriff hinausgehen.

1. Software-Support der Produktionsklasse
 - Software-Updates und Patches - Regelmäßige Sicherheits-Updates, Bugfixes und Funktionsverbesserungen
 - Langzeit-Support-Versionen - Stabile LTS-Versionen mit erweiterten Support-Fenstern (bis zu 2 Jahre)



Anmerkung: Die NVIDIA-Supportstufen (Basic und Rund-um-die-Uhr) basieren auf der jeweiligen, wie [hier](#) dargestellt erworbenen Lizenz.

Service-Komponente	Standard	Rund um die Uhr
Zugriff auf technischen Support	Geschäftszeiten vor Ort	24 Stunden, 7 Tage die Woche
Anfängliche Reaktionszeiten mit Schweregrad 1	4 Geschäftsstunden	1 Stunde
Anfängliche Reaktionszeiten mit Schweregrad 2	4 Geschäftsstunden	2 Stunden
Anfängliche Reaktionszeiten mit Schweregrad 3	1 Geschäftstag	4 Geschäftsstunden
Anfängliche Reaktionszeiten mit Schweregrad 4	2 Werktage	1 Geschäftstag
Zugriff auf das Kundenportal	Rund um die Uhr	Rund um die Uhr
Web	Rund um die Uhr	Rund um die Uhr
Telefon	Geschäftszeiten vor Ort	Rund um die Uhr

2. Technische Kompetenz und Beratung

- Unterstützung bei der Bereitstellung von KI - Unterstützung bei Architekturdesign und Best Practices
- Leistungsoptimierung - Anleitung zur Optimierung von KI-Workloads für maximale Leistung

- Framework-Unterstützung - Unterstützung für TensorFlow, PyTorch und andere in der Suite enthaltene KI-Frameworks
- Container- und Kubernetes-Support - Hilfe bei Cloud-nativen KI-Bereitstellungen
- Zugriff auf die Knowledge Base - Umfassende Dokumentation, Leitfäden und Ressourcen zur Fehlerbehebung

3. Lizenzierung und Zugriffsvorteile

- Verifizierte und validierte Software - Enterprise-getestete und zertifizierte Versionen
- Übertragbarkeit von Lizenzen - flexible Verschiebung von Lizenzen über die gesamte Infrastruktur hinweg
- Zentrales Lizenzmanagement - NVIDIA-Lizenzsystem zur Verwaltung von Bereitstellungen
- Abonnementmodell - jährliche oder mehrjährige Lizenzierungsoptionen
- Software Download Access - Offizielles Download-Portal für alle NVAIE-Komponenten

4. Sicherheit und Compliance

- Sicherheits-Patches - Rechtzeitige Behebung von Schwachstellen
- CVE-Tracking - Proaktive Benachrichtigung bei Sicherheitsproblemen
- Compliance-Dokumentation - Unterstützung bei der Erfüllung gesetzlicher Auflagen
- Sichere Software-Lieferkette - geprüfte und signierte Softwarekomponenten

Support-Abdeckung

Funktion	NVIDIA AI Enterprise
GPU-Treiberzugriff	✓
Basisdokumentation	✓
Community-Foren	✓
AI-Software-Suite	✓
Enterprise Support (7x24h)	✓
SLA-Reaktionszeiten	✓
Sicherheitspatches	✓
Langfristiger Support	✓

Funktion	NVIDIA AI Enterprise
Produktions-Zertifizierung	✓
Technisches Kundenmanagement	Optionales Add-on
Zugriff auf AI-Workflow	✓
Verwaltungstools	✓

Bestellinformationen

Cisco UCS mit NVIDIA AI Enterprise

Schritt 1: Cisco UCS-Hardware auswählen

Wählen Sie die geeignete Cisco UCS-Plattform und NVIDIA GPU-Konfiguration aus:

Beispielkonfigurationen:

- UCS C885A M8 mit NVIDIA HGX H200 GPU
- UCS C845A M8 mit NVIDIA RTX P6000 GPUs
- UCS C880A M8 mit NVIDIA HGX B300 GPUs
- UCS C240 M7 mit NVIDIA L40S GPUs
- UCS X-Serie mit NVIDIA H100 GPUs für modulare Bereitstellungen



Anmerkung: Jede NVIDIA H200 NVL, H100 PCIe und H100 NVL GPU wird mit einem fünfjährigen NVIDIA Enterprise-Abonnement ausgeliefert, das NVIDIA AI Enterprise umfasst. Eine Softwareaktivierung ist erforderlich. Sie müssen die Lizenz aktivieren und bei Bedarf ein Support-Ticket für NVIDIA erstellen.

Klicken Sie [hier](#), um weitere Details anzuzeigen.

Beispielsweise beinhalten UCS C845A M8, UCSX-210C-M8, UCS C240 M7 mit NVIDIA H200 GPUs ein fünfjähriges NVIDIA Enterprise-Abonnement.

Phase 2: NVIDIA AI Enterprise-Lizenz

NVIDIA AI Enterprise-Lizenzen sind separat erhältlich und können erworben werden über:

- Cisco Commerce - Im Paket mit dem Cisco UCS



Anmerkung: Die NVIDIA-Lizenz kann auch außerhalb von Cisco erworben werden.

Lizenztypen:

- NVAIE-Standard - Standard-Jahresabonnement (pro GPU-Socket oder pro Benutzer)
- NVAIE- Educational - Jahresabonnement für Schulungen (pro GPU-Socket oder pro Benutzer)
- NVAIE-vGPU - Kombinierte NVAIE- und vGPU-Lizenz

Weitere Informationen finden Sie in [der Bestellanleitung für NVIDIA](#).



Anmerkung: Informationen zu Teilenummern und Preisen erhalten Sie von Ihrem Cisco Account Team oder einem Vertriebsmitarbeiter von NVIDIA.

Schritt 3: Support-Optionen

Vergewissern Sie sich,

- Cisco UCS-Supportvertrag - Cisco Support (Standard, Erweitert, Signatur)
- NVIDIA AI Enterprise-Abonnement - mit NVIDIA-Unterstützung

Referenzkonfiguration

Beispielreihenfolge für KI-Infrastruktur:

1. Hardware: UCS C885A M8 mit NVIDIA H200 GPUs (8 x H200)
1. Software: NVIDIA AI Enterprise (8 GPU-Lizenzen)
2. Support:
 - Cisco Support - Erweitert für UCS-Hardware
 - NVIDIA AI Enterprise-Abonnement (mit Support rund um die Uhr)
3. Services: Optionale Professional Services für die Bereitstellung

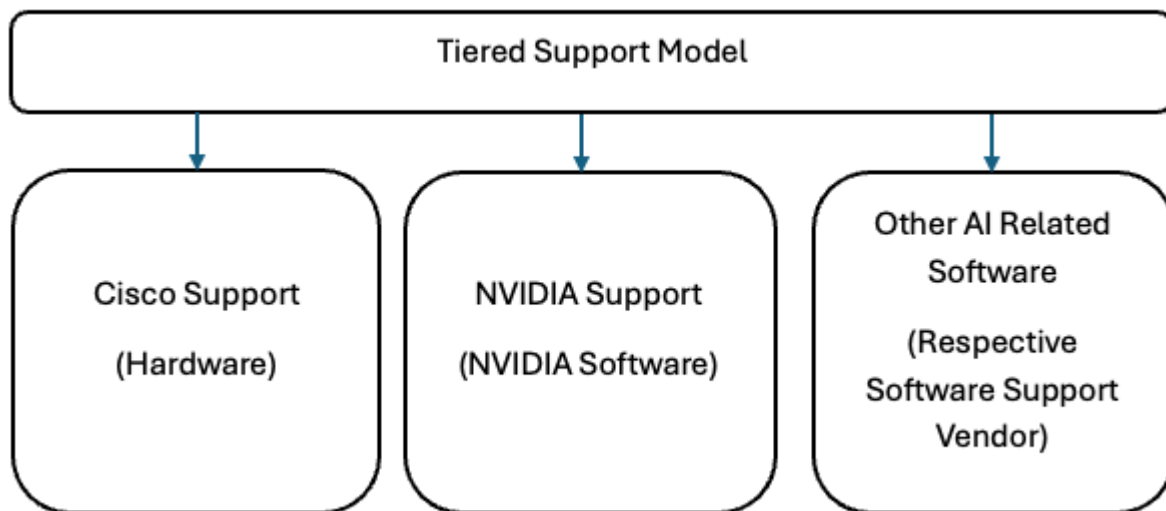
Ausführliche Bestellanleitungen finden Sie hier:

- [NVIDIA AI Enterprise Bestellanleitung für Cisco UCS](#)

So erreichen Sie den Support

Support-Modell - Übersicht

NVIDIA AI Enterprise auf dem Cisco UCS basiert auf einem mehrstufigen Support-Modell, das eine klare Abgrenzung zwischen Hardware- und Software-Support beinhaltet:



Anmerkung: Bei Problemen mit der NVIDIA-Software müssen Sie ein Support-Ticket direkt bei NVIDIA erstellen. Fügen Sie bei Bedarf auch den Cisco Support für gemeinsame Fehlerbehebungsmaßnahmen hinzu.

Kontaktieren des Cisco Supports

- Hardwareprobleme:
 - Cisco UCS-Serverausfälle oder -fehler
 - Hardwarediagnose und -austausch
 - Firmware-Updates für UCS-Komponenten
 - UCS Manager- oder Intersight-Konfigurationsprobleme
 - Physische GPU-Installations- oder Erkennungsprobleme
 - Probleme mit der Netzwerkverbindung in der UCS-Infrastruktur
- Cisco Software:
 - Probleme mit UCS Manager oder Cisco Intersight
 - Cisco Netzwerkkonfiguration für die KI-Infrastruktur

So öffnen Sie ein Cisco Support-Ticket:

1. Website: [Weltweiter Kontakt zum Cisco Support](#)
2. Direkter Cisco Support (24/7 Hotline)
3. Cisco Support-Ticketmanager verwenden
4. Kontaktaufnahme über Cisco Intersight (bei Geräten mit Cloud-Verbindung)

Vor Erstellung des Support-Tickets erforderliche Informationen:

- Cisco UCS-Modell und Seriennummer
- Servicevertragsnummer

- Detaillierte Problembeschreibung
- Dateien für den technischen Support (technischen Support anzeigen)
- GPU-Bestands- und Konfigurationsdetails

Wann Sie NVIDIA Support kontaktieren

NVIDIA AI Enterprise-Software:

- Installations- oder Konfigurationsprobleme mit dem AI-Framework
- Fragen zum NVIDIA AI-Workflow
- Triton Inference Server-Bereitstellung
- NVIDIA vGPU-Softwareprobleme
- Probleme mit GPU-Operator oder Netzwerkbetreiber
- Probleme mit der Container-Laufzeit mit dem NVIDIA Container Toolkit
- Lizenzaktivierung oder -verwaltung
- Performance-Tuning für KI-Workloads
- Software-Bugs oder Funktionsanforderungen
- Fragen zu Sicherheitslücken

Probleme mit Treibern und Software:

- NVIDIA-Treiberinstallation oder -kompatibilität
- CUDA-Toolkit-Probleme
- cuDNN- oder TensorRT-Probleme
- GPU-Auslastungs- oder Leistungsprobleme

So öffnen Sie ein Ticket:

1. Besuchen Sie das NVIDIA Enterprise [Support Portal](#)
2. Melden Sie sich mit Ihrem NVIDIA Enterprise-Konto an.
3. Wählen Sie "Create Case" aus, und geben Sie Folgendes an:
 - NVIDIA AI Enterprise Lizenzinformationen
 - Server- und GPU-Details
 - Softwareversionen
 - Detaillierte Problembeschreibung
 - Protokolle und Diagnoseausgaben

Zugriff auf das Support-Portal:

- Aktives NVIDIA AI Enterprise-Abonnement erforderlich
- Kontoeinrichtung während der Lizenzbereitstellung
- Zugriff auf die Wissensdatenbank, Downloads und Ticket-Management

Unterstützung für andere KI-bezogene Software

Befolgen Sie die Support-Verfahren des entsprechenden Anbieters der AI-Software/Anwendung.

Die Informationen in diesem Dokument beziehen sich auf Geräte in einer speziell eingerichteten Testumgebung. Alle Geräte, die in diesem Dokument benutzt wurden, begannen mit einer gelöschten (Nichterfüllungs) Konfiguration. Wenn Ihr Netzwerk in Betrieb ist, stellen Sie sicher, dass Sie die möglichen Auswirkungen aller Befehle kennen.

Informationen zu dieser Übersetzung

Cisco hat dieses Dokument maschinell übersetzen und von einem menschlichen Übersetzer editieren und korrigieren lassen, um unseren Benutzern auf der ganzen Welt Support-Inhalte in ihrer eigenen Sprache zu bieten. Bitte beachten Sie, dass selbst die beste maschinelle Übersetzung nicht so genau ist wie eine von einem professionellen Übersetzer angefertigte. Cisco Systems, Inc. übernimmt keine Haftung für die Richtigkeit dieser Übersetzungen und empfiehlt, immer das englische Originaldokument (siehe bereitgestellter Link) heranzuziehen.