

Supporto per i server Cisco AI UCS contenenti GPU Nvidia

Sommario

[Introduzione](#)

[Premesse](#)

[Panoramica del software aziendale NVIDIA AI](#)

[Che cos'è NVIDIA AI Enterprise?](#)

[Componenti principali](#)

[Modelli di distribuzione](#)

[Vantaggi del supporto con licenza NVIDIA AI Enterprise](#)

[Informazioni per l'ordinazione](#)

[Come ottenere assistenza](#)

[Quando contattare il supporto Cisco](#)

[Quando contattare il supporto NVIDIA](#)

[Supporto per altri software correlati all'intelligenza artificiale](#)

Introduzione

Questo documento descrive la suite NVIDIA AI Enterprise Software (NVIDIA) e i vantaggi completi del supporto disponibili quando implementata su piattaforme Cisco Unified Computing System (UCS). Questo documento descrive anche come coinvolgere il supporto di Cisco e NVIDIA per un'installazione ottimale e assistenza operativa.

Premesse

Poiché i carichi di lavoro di intelligenza artificiale e apprendimento delle macchine diventano mission-critical per le operazioni aziendali, le organizzazioni necessitano non solo di infrastrutture potenti, ma anche di software e supporto di livello enterprise. Il software aziendale NVIDIA AI offre una piattaforma AI completa, sicura e supportata ottimizzata per l'installazione su infrastrutture Cisco UCS.

In questo documento vengono spiegate:

- Cosa implica il software aziendale NVIDIA AI
- Vantaggi della licenza di NAVISPHERE Software Suite
- Come attivare correttamente i canali di supporto per NVIALE su Cisco UCS

Panoramica del software aziendale NVIDIA AI

Che cos'è NVIDIA AI Enterprise?

NVIDIA AI Enterprise (NVAIE) è una suite software completa che fornisce sicurezza, stabilità, gestibilità e supporto di livello enterprise per i carichi di lavoro dell'IA in esecuzione in ambienti virtualizzati e nativi del cloud. È ottimizzato e certificato per l'esecuzione sulle principali piattaforme aziendali, tra cui Cisco UCS.

Componenti principali



Nota: Per un elenco completo, visitare [Nvidia AI Enterprise](#)

Framework e strumenti IA:

- Flussi di lavoro NVIDIA AI - Flussi di lavoro preconfigurati e completi per le comuni applicazioni AI
- NVIDIA TAO Toolkit - Transfer learning toolkit per la visione dei computer, l'intelligenza artificiale conversazionale e altro ancora
- NVIDIA Riva - SDK per la creazione di applicazioni di riconoscimento vocale
- NVIDIA Morpheus - Framework IA per la cibersecurity
- NVIDIA Metropolis - Piattaforma Vision AI per l'analisi intelligente delle immagini
- NVIDIA Clara - Framework dell'applicazione AI per il settore sanitario

Software di infrastruttura:

- Software NVIDIA vGPU - Tecnologia di virtualizzazione GPU (concessa in licenza separatamente o inclusa con NVIDIA)
- Operatore GPU NVIDIA - Operatore Kubernetes per la gestione GPU
- NVIDIA Network Operator - Gestisce le risorse di rete NVIDIA a Kubernetes
- NVIDIA Container Toolkit - Strumenti per la creazione e l'esecuzione di contenitori con accelerazione GPU

Stack di sviluppo dell'IA:

- Framework di IA ottimizzati (TensorFlow, PyTorch, JAX e così via)
- NVIDIA RAPIDS per la scienza dei dati
- NVIDIA Triton Inference Server per l'installazione di intelligenza artificiale di produzione
- NVIDIA TensorRT per un'inferenza ottimizzata
- Librerie NVIDIA cuDNN, cuBLAS e CUDA

Gestione e sicurezza:

- NVIDIA Base Command Manager Essentials - Gestione dell'infrastruttura per AI
- Sistema di licenze NVIDIA - Gestione centralizzata delle licenze
- Patch per la sicurezza e versioni per il supporto a lungo termine (LTS)

Modelli di distribuzione

1. Ambienti virtualizzati - VMware vSphere, virtualizzazione Red Hat, GPU virtuale NVIDIA
2. Bare Metal - Installazione diretta su server fisici
3. Nativo del cloud - Kubernetes e installazioni containerizzate
4. Hybrid e Multi-Cloud - Piattaforma AI coerente in tutti gli ambienti

Vantaggi del supporto con licenza NVIDIA AI Enterprise

Supporto Enterprise Completo

Con la licenza di NVIDIA AI Enterprise, si ottengono vantaggi significativi che vanno oltre l'accesso di base ai driver GPU.

1. Supporto software di livello produzione
 - Patch e aggiornamenti software - Aggiornamenti periodici della sicurezza, correzioni di bug e miglioramenti delle funzionalità
 - Versioni di supporto a lungo termine - Versioni LTS stabili con finestre di supporto estese (fino a 2 anni)



Nota: I livelli di supporto Nvidia (Basic vs 24x7) si basano sulla licenza specifica acquistata come descritto [qui](#)

Componente del servizio	Standard	24x7
Accesso al supporto tecnico	Orario lavorativo locale	24 ore, 7 giorni alla settimana
Tempi di risposta iniziali di gravità 1	4 ore lavorative	1 ora
Tempi di risposta iniziali di gravità 2	4 ore lavorative	2 ore
Gravità 3 Tempi di risposta iniziali	1 giorno lavorativo	4 ore lavorative
Gravità 4 - Tempi di risposta iniziali	2 giorni lavorativi	1 giorno lavorativo
Accesso al portale clienti	24x7	24x7
Web	24x7	24x7
Telefono	Orario lavorativo locale	24x7

2. Competenze tecniche e guida

- Assistenza all'installazione di AI - Assistenza nella progettazione dell'architettura e nelle best practice
- Ottimizzazione delle prestazioni - Linee guida per ottimizzare i carichi di lavoro dell'IA per

ottenere le massime prestazioni

- Supporto framework - Assistenza con TensorFlow, PyTorch e altri framework di intelligenza artificiale inclusi nella suite
- Supporto per contenitori e Kubernetes - Assistenza per le installazioni di intelligenza artificiale native del cloud
- Accesso alla Knowledge Base - Documentazione completa, guide e risorse per la risoluzione dei problemi

3. Vantaggi di licenze e accesso

- Software verificato e convalidato - Versioni certificate e testate a livello aziendale
- Portabilità delle licenze - Flessibilità per lo spostamento delle licenze nell'infrastruttura
- Gestione centralizzata delle licenze - Sistema di licenze NVIDIA per la gestione delle installazioni
- Modello di abbonamento - Opzioni di licenza annuali o pluriennali
- Accesso al download del software - Portale di download ufficiale per tutti i componenti NVAIE

4. Sicurezza e conformità

- Patch di sicurezza - Correzione tempestiva della vulnerabilità
- Monitoraggio CVE - Notifica proattiva dei problemi di sicurezza
- Documentazione sulla conformità - Supporto per i requisiti normativi
- Catena di fornitura software sicura - Componenti software verificati e firmati

Copertura supporto

Funzionalità	NVIDIA AI Enterprise
Accesso ai driver GPU	✓
Documentazione di base	✓
Forum della community	✓
Suite software AI	✓
Supporto Enterprise (24 ore su 24, 7 giorni su 7)	✓
Tempi di risposta SLA	✓

Funzionalità	NVIDIA AI Enterprise
Patch di sicurezza	Źaddio
Supporto a lungo termine	Źaddio
Certificazione di produzione	Źaddio
Gestione tecnica degli account	Componente aggiuntivo opzionale
Accesso flusso di lavoro AI	Źaddio
Strumenti di gestione	Źaddio

Informazioni per l'ordinazione

Cisco UCS con NVIDIA AI Enterprise

Passaggio 1: Seleziona hardware Cisco UCS

Scegliere la piattaforma Cisco UCS e la configurazione NVIDIA GPU appropriate:

Configurazioni di esempio:

- UCS C885A M8 con GPU NVIDIA HGX H200
- UCS C845A M8 con GPU NVIDIA RTX6000
- UCS C880A M8 con GPU NVIDIA HGX B300
- UCS C240 M7 con GPU NVIDIA L40S
- UCS serie X con GPU NVIDIA H100 per installazioni modulari



Nota: Ogni NVIDIA H200 NVL, H100 PCIe e H100 NVL GPU viene fornita con un abbonamento quinquennale a NVIDIA Enterprise, che include NVIDIA AI Enterprise. Attivazione software necessaria. Se necessario, è necessario attivare la licenza e aprire una richiesta di assistenza per il software con NVIDIA.

Fare clic [qui](#) per ulteriori informazioni.

Ad esempio, UCS C845A M8, UCSX-210C-M8, UCS C240 M7 con GPU NVIDIA H200 includono un abbonamento quinquennale a NVIDIA Enterprise.

Passaggio 2: Licenza NVIDIA AI Enterprise

Le licenze NVIDIA AI Enterprise sono vendute separatamente e possono essere acquistate tramite:

- Cisco Commerce - Acquisti Cisco UCS in bundle



Nota: La licenza NVIDIA può essere acquistata anche al di fuori di Cisco.

Tipi di licenza:

- NVAIE- Standard - Abbonamento annuale standard (per socket GPU o per utente)
- NVAIE- Educational - Abbonamento annuale didattico (per socket GPU o per utente)
- NVAIE-vGPU - Licenza combinata NVAIE + vGPU

Per maggiori informazioni, consultare [la Guida agli ordini di NVIDIA](#)



Nota: Per informazioni specifiche sui numeri di parte e sui prezzi, contattare l'Account Team Cisco o il rappresentante commerciale NVIDIA.

Passaggio 3: Opzioni di supporto

Assicurati di avere:

- Contratto di supporto Cisco UCS - Supporto Cisco (standard, migliorato, firma)
- Abbonamento NVIDIA AI Enterprise - Include il supporto NVIDIA

Configurazione di riferimento

Ordine di esempio per l'infrastruttura AI:

1. Hardware: UCS C885A M8 con GPU NVIDIA H200 (8x H200)
1. Software: NVIDIA AI Enterprise (8 licenze GPU)
2. Supporto:
 - Supporto Cisco - Migliorato per hardware UCS
 - Abbonamento NVIDIA AI Enterprise (supporto 24 ore su 24, 7 giorni su 7 incluso)
3. Servizi Servizi professionali opzionali per l'installazione

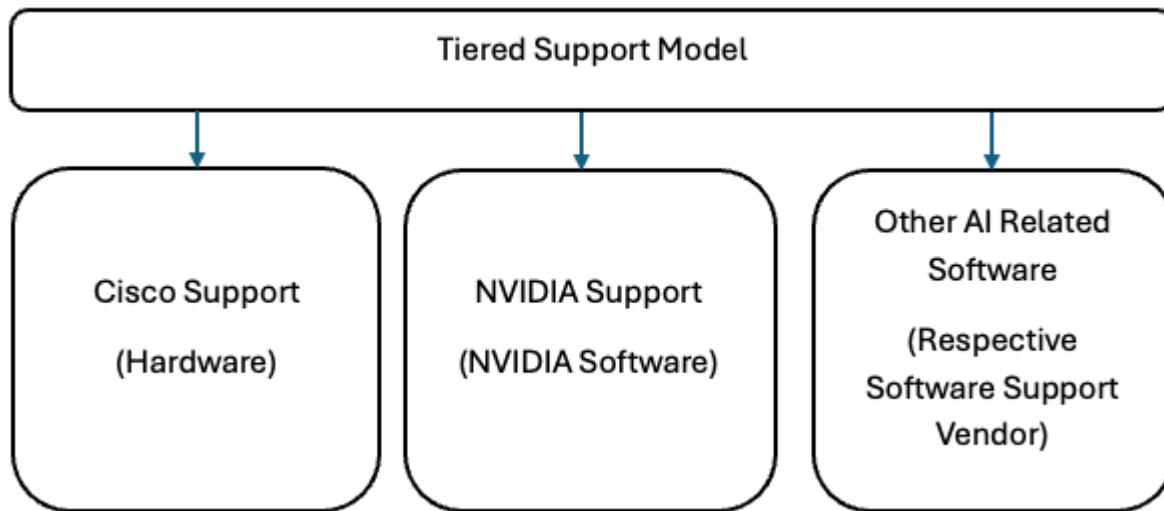
Per le guide dettagliate agli ordini, consultare:

- [Guida agli ordini di NVIDIA AI Enterprise per Cisco UCS](#)

Come ottenere assistenza

Panoramica del modello di supporto

NVIDIA AI Enterprise su Cisco UCS aderisce a un modello di supporto su più livelli che include una chiara distinzione tra supporto hardware e software:



Nota: Per i problemi relativi al software NVIDIA, è necessario creare una richiesta di assistenza direttamente con NVIDIA. Se necessario, includere anche il supporto Cisco per attività congiunte di risoluzione dei problemi.

Quando contattare il supporto Cisco

- Problemi hardware:
 - Errori o errori del server Cisco UCS
 - Diagnostica hardware e sostituzione
 - Aggiornamenti del firmware per i componenti UCS
 - UCS Manager o problemi di configurazione di Intersight
 - Problemi fisici di installazione o rilevamento della GPU
 - Problemi di connettività di rete con l'infrastruttura UCS
- Software Cisco:
 - Problemi di UCS Manager o Cisco Intersight
 - Configurazione di rete Cisco per l'infrastruttura AI

Come aprire una richiesta di assistenza Cisco:

1. Visita: [Contatti del supporto Cisco internazionali](#)
2. Chiama direttamente il supporto Cisco (24 ore su 24, 7 giorni su 7)
3. Uso di Cisco Support Case Manager

4. Contatta tramite Cisco Intersight (per dispositivi connessi al cloud)

Informazioni richieste prima della creazione della richiesta di assistenza:

- Modello e numero di serie Cisco UCS
- Numero di contratto di assistenza
- Descrizione dettagliata del problema
- File di supporto tecnico (show tech-support)
- Dettagli inventario e configurazione GPU

Quando contattare il supporto NVIDIA

Software aziendale NVIDIA AI:

- Problemi di installazione o configurazione del framework AI
- Domande sul flusso di lavoro NVIDIA AI
- Distribuzione di Triton Inference Server
- Problemi relativi al software NVIDIA vGPU
- Problemi dell'operatore GPU o dell'operatore di rete
- Problemi di runtime del contenitore con NVIDIA Container Toolkit
- Attivazione o gestione delle licenze
- Ottimizzazione delle prestazioni per i carichi di lavoro AI
- Bug software o richieste di funzionalità
- Domande sulla vulnerabilità della sicurezza

Problemi relativi a driver e software:

- Installazione o compatibilità del driver NVIDIA
- Problemi relativi al toolkit CUDA
- problemi cuDNN o TensorRT
- Problemi di prestazioni o utilizzo della GPU

Come aprire una richiesta:

1. Visita il [portale di supporto](#) aziendale NVIDIA
2. Accedere con l'account NVIDIA Enterprise
3. Selezionare "Create Case" (Crea richiesta) e fornire:
 - Informazioni sulle licenze NVIDIA AI Enterprise
 - Dettagli su server e GPU
 - Versioni software
 - Descrizione dettagliata del problema
 - Registri e output di diagnostica

Accesso al portale di supporto:

- Richiede l'abbonamento attivo a NVIDIA AI Enterprise
- Impostazione dell'account durante il provisioning della licenza
- Accesso alla knowledge base, download e gestione dei casi

Supporto per altri software correlati all'intelligenza artificiale

Rispettare le procedure di supporto fornite dal fornitore di software/app AI applicabile

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Informazioni su questa traduzione

Cisco ha tradotto questo documento utilizzando una combinazione di tecnologie automatiche e umane per offrire ai nostri utenti in tutto il mondo contenuti di supporto nella propria lingua. Si noti che anche la migliore traduzione automatica non sarà mai accurata come quella fornita da un traduttore professionista. Cisco Systems, Inc. non si assume alcuna responsabilità per l'accuratezza di queste traduzioni e consiglia di consultare sempre il documento originale in inglese (disponibile al link fornito).