

支援包含Nvidia GPU的Cisco AI UCS伺服器

目錄

[簡介](#)

[背景資訊](#)

[NVIDIA AI企業版軟體概述](#)

[什麼是NVIDIA AI Enterprise?](#)

[關鍵元件](#)

[部署模式](#)

[NVIDIA AI企業許可證的支援優勢](#)

[訂購資訊](#)

[如何吸引支援人員](#)

[何時聯絡思科支援](#)

[何時聯絡NVIDIA支援](#)

[支援其他AI相關軟體](#)

簡介

本檔案介紹NVIDIA AI企業軟體(NVAIE)套件以及在思科統一計算系統(UCS)平台上部署時可用的全面支援優勢。本文檔還介紹了如何利用Cisco和NVIDIA支援獲得最佳部署和操作幫助。

背景資訊

隨著人工智慧和機器學習工作負載成為企業運營的關鍵任務，組織不僅需要強大的基礎設施，還需要企業級軟體和支援。NVIDIA AI企業軟體提供全面、安全和受支援的AI平台，該平台針對在Cisco UCS基礎設施上的部署進行了最佳化。

本檔案將說明：

- NVIDIA AI企業版軟體的需求是什麼
- 許可NVAIE軟體套件的支援優勢
- 如何正確接洽Cisco UCS上的NVAIE支援管道

NVIDIA AI企業版軟體概述

什麼是NVIDIA AI Enterprise?

NVIDIA AI Enterprise(NVAIE)是一個全面的軟體套件，為運行在虛擬和雲原生環境中的AI工作負載提供企業級安全性、穩定性、可管理性和支援。它經過最佳化和認證，可在包括Cisco UCS在內的領先企業平台上運行。

關鍵元件



附註：有關最新和詳盡的清單，請訪問[Nvidia AI Enterprise](#)

AI框架和工具：

- NVIDIA AI工作流程 — 針對常見AI應用的預構建、端到端工作流程
- NVIDIA TAO工具包 — 用於電腦視覺、對話AI等的遷移學習工具包
- NVIDIA Riva — 用於構建語音AI應用程式的SDK
- NVIDIA Morpheus — 網絡安全AI框架
- NVIDIA Metropolis — 用於智慧影片分析的Vision AI平台
- NVIDIA Clara — 醫療保健AI應用框架

基礎架構軟體：

- NVIDIA vGPU軟體- GPU虛擬化技術（單獨許可或隨NVAIE提供）
- NVIDIA GPU操作員 — 適用於GPU管理的Kubernetes操作員
- NVIDIA網路運營商 — 管理Kubernetes中的NVIDIA網路資源
- NVIDIA容器工具包 — 用於構建和運行GPU加速容器的工具

AI開發堆疊：

- 最佳化的AI框架（TensorFlow、PyTorch、JAX等）
- 適用於資料科學的NVIDIA RAPIDS
- 適用於生產AI部署的NVIDIA Triton Inference Server
- NVIDIA TensorRT用於最佳化推理
- NVIDIA cuDNN、cuBLAS和CUDA庫。

管理和安全：

- NVIDIA Base命令管理器基本版 — 適用於AI的基礎架構管理
- NVIDIA許可證系統 — 集中式許可證管理
- 安全補丁和長期支援(LTS)版本

部署模式

1. 虛擬化環境 — VMware vSphere、Red Hat虛擬化、NVIDIA虛擬GPU
2. 裸機 — 在物理伺服器上直接部署
3. 雲原生 — Kubernetes和容器化部署
4. 混合和多雲 — 跨環境的一致AI平台

NVIDIA AI企業許可證的支援優勢

全面的企業支援

通過許可NVIDIA AI Enterprise，您將獲得超越基本GPU驅動程式訪問的重大支援優勢。

1. 生產級軟體支援

- 軟體更新和補丁 — 定期安全更新、錯誤修正和功能增強
- 長期支援版本 — 具有擴展支援視窗（長達2年）的穩定LTS版本



附註：Nvidia支援級別（基礎版與24x7版）基於購買的特定許可證，如下所述

服務元件	標準	24x7
技術支援訪問	本地工作時間	一週7天，24小時
嚴重級別1初始響應時間	4個工作小時	1小時
嚴重級別2初始響應時間	4個工作小時	2小時
嚴重級別3初始響應時間	1個工作日	4個工作小時
嚴重程度為4的初始響應時間	2個工作日	1個工作日
訪問客戶門戶	24x7	24x7
Web	24x7	24x7
電話	本地工作時間	24x7

2. 技術專長和指導

- AI部署幫助 — 幫助進行架構設計和最佳實踐
- 效能最佳化 — 有關調整AI工作負載以實現最高效能的指導
- 框架支持 — 對TensorFlow、PyTorch和套件中包含的其他AI框架的協助
- 容器和Kubernetes支援 — 幫助進行雲原生AI部署
- 知識庫訪問 — 全面的文檔、指南和故障排除資源

3. 許可和訪問優勢

- 經驗證和驗證的軟件 — 經企業測試和認證的版本
- 許可證可移植 — 靈活地在基礎設施之間移動許可證
- 集中許可證管理 — 用於管理部署的NVIDIA許可證系統
- 訂購模式 — 年度或多年許可選項
- 軟體下載訪問 — 所有NVAIE元件的官方下載門戶

4. 安全性和合規性

- 安全修補程序 — 及時修補漏洞
- CVE跟蹤 — 主動通知安全問題
- 合規性文件 — 對法規要求的支援
- 安全軟體供應鏈 — 經過驗證並簽名的軟體元件

支援範圍

功能	NVIDIA AI企業版
GPU驅動程式訪問	✓
基本文檔	✓
社群論壇	✓
AI軟體套件	✓
企業支援(24/7)	✓
SLA響應時間	✓
安全補丁程式	✓
長期支援	✓
生產認證	✓
技術帳戶管理	可選載入項
AI工作流訪問	✓
管理工具	✓

訂購資訊

採用NVIDIA AI Enterprise的Cisco UCS

步驟 1:選擇Cisco UCS硬體

選擇適當的Cisco UCS平台和NVIDIA GPU配置：

配置示例：

- 採用NVIDIA HGX H200 GPU的UCS C885A M8
- 採用NVIDIA RTX P6000 GPU的UCS C845A M8

- 採用NVIDIA HGX B300 GPU的UCS C880A M8
- 採用NVIDIA L40S GPU的UCS C240 M7
- 採用NVIDIA H100 GPU的UCS X系列，用於模組化部署



附註：每個NVIDIA H200 NVL、H100 PCIe和H100 NVL GPU均提供為期五年的NVIDIA Enterprise訂閱，其中包括NVIDIA AI Enterprise。軟體啟用是必需的。如果需要，您需要啟用許可證並使用NVIDIA開啟軟體支援案例。

按一下[Here](#)以瞭解其他詳細資訊。

例如，採用NVIDIA H200 GPU的UCS C845A M8、UCSX-210C-M8、UCS C240 M7包括五年NVIDIA企業版訂閱。

步驟 2:NVIDIA AI企業版許可證

NVIDIA AI Enterprise許可證單獨銷售，可通過以下方式購買：

- 思科商務 — 與思科UCS採購捆綁在一起



附註：NVIDIA許可證也可以在思科之外購買。

許可證型別：

- NVAIE-Standard — 標準年度訂閱（每個GPU插槽或每個使用者）
- NVAIE教育 — 教育年度訂閱（按GPU插槽或按使用者）
- NVAIE-vGPU -組合式NVAIE + vGPU許可證

有關更多詳情，請參閱[NVIDIA訂購指南](#)



附註：有關具體部件號和價格，請聯絡您的思科客戶團隊或NVIDIA銷售代表。

步驟 3:支援選項

確保您具有：

- Cisco UCS支援合約- Cisco支援（標準、增強、簽名）
- NVIDIA AI企業版訂閱 — 包括NVIDIA支援

參考配置

AI基礎設施訂單示例：

1. 硬體：採用NVIDIA H200 GPU的UCS C885A M8(8x H200)
1. 軟體：NVIDIA AI企業版（8個GPU許可證）
2. 支援：
 - 思科支援 — 增強型UCS硬體
 - NVIDIA AI Enterprise訂閱（包括全天候支援）
3. 服務：適用於部署的可選專業服務

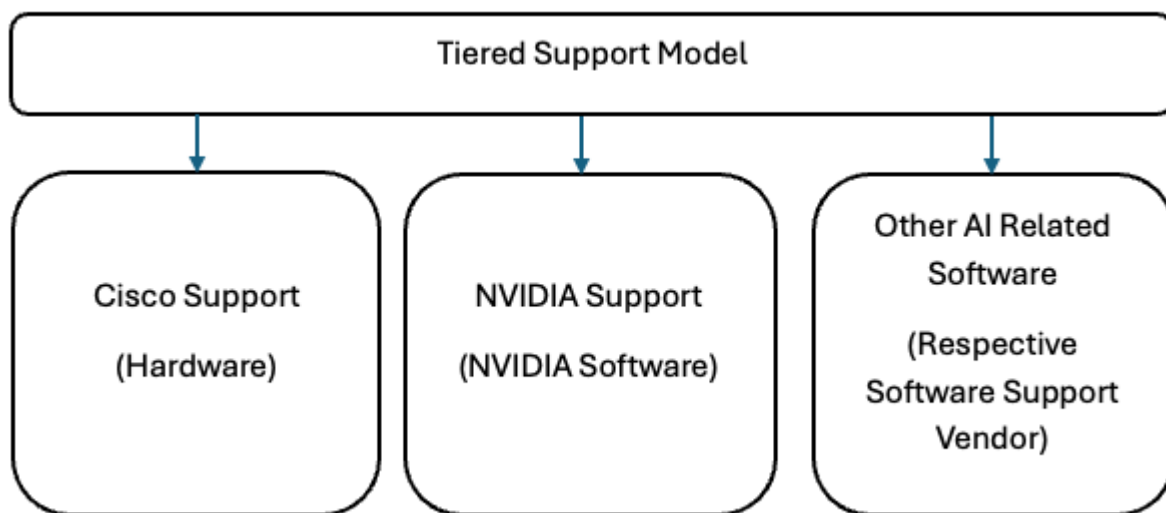
有關詳細的訂購指南，請參閱：

- [適用於Cisco UCS的NVIDIA AI企業訂購指南](#)

如何吸引支援人員

支援模式概述

Cisco UCS上的NVIDIA AI Enterprise遵循分層支援模式，該模式明確區分了硬體和軟體支援：



附註：對於NVIDIA軟體問題，您必須直接使用NVIDIA建立支援案例。如果需要，還應包括思科支援以開展聯合故障排除工作。

何時聯絡思科支援

- 硬體問題：
 - Cisco UCS伺服器故障或錯誤
 - 硬體診斷和更換
 - UCS元件的韌體更新
 - UCS Manager或Intersight配置問題
 - 物理GPU安裝或檢測問題

- UCS基礎設施的網路連線問題

- 思科軟體：
 - UCS Manager或Cisco Intersight問題
 - Cisco AI基礎設施網路配置

如何開啟思科支援問題單：

1. 請訪問：[思科全球支援聯絡人](#)
2. 直接致電思科支援 (24/7熱線)
3. 使用思科支援個案管理器
4. 通過Cisco Intersight聯絡 (適用於雲連線裝置)

建立支援案例之前需要的資訊：

- Cisco UCS型號和序列號
- 服務合約編號
- 詳細問題描述
- 技術支援檔案(show tech-support)
- GPU資產和配置詳細資訊

何時聯絡NVIDIA支援

NVIDIA AI企業版軟體：

- AI框架安裝或配置問題
- NVIDIA AI工作流程問題
- Triton推理伺服器部署
- NVIDIA vGPU軟體問題
- GPU操作員或網路操作員問題
- NVIDIA Container Toolkit的容器運行時間問題
- 許可證啟用或管理
- 針對AI工作負載的效能調整
- 軟體錯誤或功能要求
- 安全漏洞問題

驅動程式和軟體問題：

- NVIDIA驅動程式安裝或相容性
- CUDA工具包問題
- cuDNN或TensorRT問題
- GPU利用率或效能問題

如何建立案例：

1. 訪問NVIDIA企業支援[門戶](#)
2. 使用您的NVIDIA企業版帳戶登入
3. 選擇「建立案例」並提供：
 - NVIDIA AI企業版許可證資訊
 - 伺服器 and GPU詳細資訊
 - 軟體版本
 - 詳細問題描述
 - 日誌和診斷輸出

支援門戶訪問：

- 需要有效的NVIDIA AI企業版訂用
- 許可證調配期間的帳戶設定
- 訪問知識庫、下載和案例管理

支援其他AI相關軟體

遵守適用的AI軟體/應用供應商提供的支援程式

本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除（預設）的組態來啟動。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。