

# 支持包含Nvidia GPU的Cisco AI UCS服务器

## 目录

---

[简介](#)

[背景信息](#)

[NVIDIA AI企业版软件概述](#)

[什么是NVIDIA AI Enterprise?](#)

[关键组件](#)

[部署模式](#)

[NVIDIA AI企业许可证的支持优势](#)

[订购信息](#)

[如何吸引支持人员](#)

[何时联系思科支持](#)

[何时联系NVIDIA支持](#)

[支持其他AI相关软件](#)

---

## 简介

本文档介绍NVIDIA AI企业软件(NVAIE)套件以及在思科统一计算系统(UCS)平台上部署时可用的全面支持优势。本文档还介绍了如何利用Cisco和NVIDIA支持获得最佳部署和运营帮助。

## 背景信息

随着人工智能和机器学习工作负载成为企业运营的关键任务，组织不仅需要强大的基础设施，还需要企业级软件和支持。NVIDIA AI Enterprise Software提供全面的、安全且受支持的AI平台，该平台经过优化，可在Cisco UCS基础设施上部署。

本文档介绍了以下内容：

- NVIDIA AI Enterprise软件的作用
- 许可NVAIE软件套件的支持优势
- 如何正确接洽Cisco UCS上的NVAIE支持渠道

## NVIDIA AI企业版软件概述

### 什么是NVIDIA AI Enterprise?

NVIDIA AI Enterprise(NVAIE)是一个全面的软件套件，为虚拟化和云本机环境中运行的AI工作负载提供企业级安全性、稳定性、可管理性和支持。它经过优化和认证，可在包括Cisco UCS在内的领先企业平台上运行。

### 关键组件



注意：有关最新和详尽的列表，请访问[Nvidia AI Enterprise](#)

#### AI框架和工具：

- NVIDIA AI工作流程 — 适用于常见AI应用的预构建端到端工作流程
- NVIDIA TAO工具包 — 用于计算机视觉、会话AI等的迁移学习工具包
- NVIDIA Riva — 用于构建语音AI应用的SDK
- NVIDIA Morpheus — 网络安全AI框架
- NVIDIA Metropolis -用于智能视频分析的Vision AI平台
- NVIDIA Clara — 医疗保健AI应用框架

#### 基础设施软件：

- NVIDIA vGPU软件- GPU虚拟化技术（单独许可或随NVAIE提供）
- NVIDIA GPU操作员- Kubernetes操作员，用于GPU管理
- NVIDIA网络运营商 — 管理Kubernetes的NVIDIA网络资源
- NVIDIA容器工具包 — 用于构建和运行GPU加速容器的工具

#### AI开发栈：

- 优化的AI框架（TensorFlow、PyTorch、JAX等）
- NVIDIA RAPIDS，用于数据科学
- 用于生产AI部署的NVIDIA Triton Inference服务器
- NVIDIA TensorRT用于优化推理
- NVIDIA cuDNN、cuBLAS和CUDA库。

#### 管理和安全：

- NVIDIA Base命令管理器基本版 — 适用于AI的基础设施管理
- NVIDIA许可证系统 — 集中式许可证管理
- 安全补丁和长期支持(LTS)版本

#### 部署模式

1. 虚拟化环境 — VMware vSphere、Red Hat虚拟化、NVIDIA虚拟GPU
2. 裸机 — 直接部署在物理服务器上
3. 云本地 — Kubernetes和容器化部署
4. 混合云和多云 — 跨环境的一致AI平台

## NVIDIA AI企业许可证的支持优势

#### 全面的企业支持

通过许可NVIDIA AI Enterprise，您将获得超越基本GPU驱动程序访问的重大支持优势。

1. 生产级软件支持

- 软件更新和补丁 — 定期安全更新、漏洞修复和功能增强
- 长期支持版本 — 具有扩展支持窗口（最长2年）的稳定LTS版本



注意：Nvidia支持级别（基础版与24x7版）基于此处列出的购买特定许可证[链接](#)

服务组件	标准	24x7
技术支持访问	当地办公时间	每周7天，每天24小时
严重级别1初始响应时间	4个工作小时	1 小时
严重级别2初始响应时间	4个工作小时	2 小时
严重程度为3的初始响应时间	1个工作日	4个工作小时
严重程度为4的初始响应时间	2个工作日	1个工作日
访问客户门户	24x7	24x7
网络	24x7	24x7
电话	当地办公时间	24x7

2. 技术专业知识和指南

- AI部署协助 — 帮助进行架构设计和最佳实践
- 性能优化 — 有关调整AI工作负载以实现最高性能的指导
- 框架支持 — 对TensorFlow、PyTorch和套件中包含的其他AI框架的帮助
- 容器和Kubernetes支持 — 帮助进行云原生AI部署
- 知识库访问 — 全面的文档、指南和故障排除资源

3. 许可和访问优势

- 经过验证和验证的软件 — 经过企业测试和认证的版本
- 许可证可移植 — 灵活地在基础设施之间移动许可证
- 集中式许可证管理 — 用于管理部署的NVIDIA许可证系统
- 订用模式 — 年度或多年期许可选项
- 软件下载访问 — 所有NVAIE组件的官方下载门户

4. 安全性和合规性

- 安全补丁 — 及时修复漏洞
- CVE跟踪 — 主动通知安全问题
- 合规性文档 — 对法规要求的支持
- 安全软件供应链 — 经过验证并签名的软件组件

支持范围

功能	NVIDIA AI企业版
GPU驱动程序访问	✓
基本文档	✓
社区论坛	✓
AI软件套件	✓
企业支持(24/7)	✓
SLA响应时间	✓
安全补丁	✓
长期支持	✓
生产认证	✓
技术客户管理	可选附加设备
AI工作流程访问	✓
管理工具	✓

## 订购信息

采用NVIDIA AI Enterprise的Cisco UCS

步骤 1：选择Cisco UCS硬件

选择适当的Cisco UCS平台和NVIDIA GPU配置：

配置示例：

- UCS C885A M8,搭配NVIDIA HGX H200 GPU
- UCS C845A M8,搭配NVIDIA RTX P6000 GPU

- UCS C880A M8，带NVIDIA HGX B300 GPU
  - UCS C240 M7,搭配NVIDIA L40S GPU
  - 采用NVIDIA H100 GPU的UCS X系列，用于模块化部署
- 



注意：每个NVIDIA H200 NVL、H100 PCIe和H100 NVL GPU均随附为期五年的NVIDIA Enterprise订用，其中包括NVIDIA AI Enterprise。软件激活是必需的。如果需要，您需要激活许可证，并使用NVIDIA打开软件支持案例。

单击[此处](#)了解更多详细信息。

例如，配备NVIDIA H200 GPU的UCS C845A M8、UCSX-210C-M8和UCS C240 M7包括五年的NVIDIA企业版订用。

---

## 步骤 2：NVIDIA AI企业版许可证

NVIDIA AI企业版许可证单独销售，可通过以下方式购买：

- 思科商务 — 与Cisco UCS采购捆绑
- 



注意：NVIDIA许可证也可以在思科之外购买。

---

许可证类型：

- NVAIE标准 — 标准年度订用（每个GPU插槽或每个用户）
- NVAIE教育 — 教育年度订用（每个GPU插槽或每个用户）
- NVAIE-vGPU -组合NVAIE + vGPU许可证

有关更多详情，请访问[NVIDIA订购指南](#)

---



注意：有关特定部件编号和定价，请联系您的思科客户团队或NVIDIA销售代表。

---

## 步骤 3：支持选项

确保您拥有：

- Cisco UCS支持合同 — 思科支持（标准、增强、签名）
- NVIDIA AI企业版订用 — 包含NVIDIA支持

## 参考配置

AI基础设施订单示例：

1. Hardware:UCS C885A M8 , 搭配NVIDIA H200 GPU(8x H200)

1. 软件：NVIDIA AI企业版（8个GPU许可证）

2. 支持:

- 思科支持 — 增强版UCS硬件
- NVIDIA AI Enterprise订用（包括全天候支持）

3. 服务：用于部署的可选专业服务

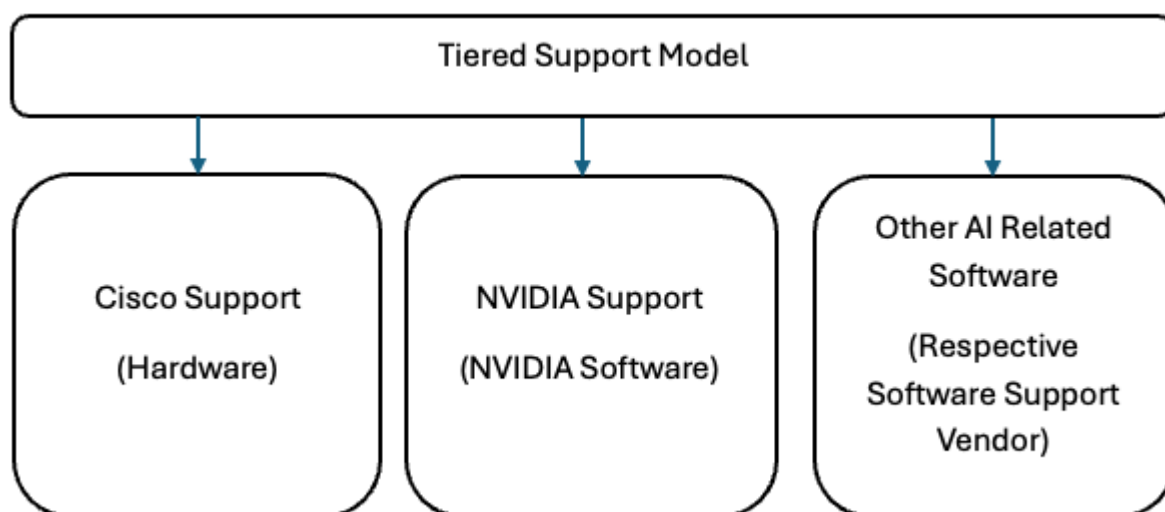
有关详细的订购指南，请参阅：

- [适用于Cisco UCS的NVIDIA AI企业订购指南](#)

## 如何吸引支持人员

### 支持模式概述

Cisco UCS上的NVIDIA AI Enterprise遵循分层支持模式，该模式明确区分了硬件和软件支持：



注意：对于NVIDIA软件问题，您必须直接使用NVIDIA创建支持案例。如果需要，请同时加入思科支持团队进行联合故障排除。

### 何时联系思科支持

- 硬件问题:
  - Cisco UCS服务器故障或错误
  - 硬件诊断和更换
  - UCS组件的固件更新
  - UCS Manager或Intersight配置问题
  - 物理GPU安装或检测问题

- UCS基础设施的网络连接问题
- Cisco 软件:
  - UCS Manager或Cisco Intersight问题
  - 适用于AI基础设施的思科网络配置

如何创建思科支持案例：

1. 请访问：[思科全球支持联系方式](#)
2. 直接致电思科支持（24/7热线）
3. 使用思科支持案例管理器
4. 通过Cisco Intersight联系（适用于云连接设备）

创建支持案例之前需要的信息：

- Cisco UCS型号和序列号
- 服务合同编号
- 详细的问题描述
- 技术支持文件(show tech-support)
- GPU资产和配置详细信息

何时联系NVIDIA支持

NVIDIA AI企业版软件：

- AI框架安装或配置问题
- NVIDIA AI工作流程问题
- Triton推理服务器部署
- NVIDIA vGPU软件问题
- GPU操作员或网络操作员问题
- NVIDIA Container Toolkit的容器运行时问题
- 许可证激活或管理
- 针对AI工作负载的性能调整
- 软件错误或功能请求
- 安全漏洞问题

驱动程序和软件问题：

- NVIDIA驱动程序安装或兼容性
- CUDA工具包问题
- cuDNN或TensorRT问题
- GPU利用率或性能问题

如何创建案例：

1. 访问NVIDIA企业支持[门户](#)
2. 使用您的NVIDIA企业帐户登录
3. 选择“创建案例”，并提供：
  - NVIDIA AI企业版许可证信息
  - 服务器和GPU详细信息
  - 软件版本
  - 详细问题描述
  - 日志和诊断输出

支持门户访问：

- 需要有效的NVIDIA AI企业版订用
- 许可证调配期间的帐户设置
- 访问知识库、下载和案例管理

## 支持其他AI相关软件

遵守适用的AI软件/应用供应商提供的支持程序

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。



## 关于此翻译

思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。