

# Nvidia GPU가 포함된 Cisco AI UCS Server 지원 참여

## 목차

---

[소개](#)

[배경 정보](#)

[NVIDIA AI 엔터프라이즈 소프트웨어 개요](#)

[NVIDIA AI Enterprise란 무엇입니까?](#)

[주요 구성 요소](#)

[구축 모델](#)

[NVIDIA AI Enterprise 라이센스를 통한 지원 혜택](#)

[주문 정보](#)

[지원 참여 방법](#)

[Cisco 지원에 문의하는 시기](#)

[NVIDIA 지원 문의 시기](#)

[기타 AI 관련 소프트웨어 지원](#)

---

## 소개

이 문서에서는 NVAIE(NVIDIA AI Enterprise Software) Suite와 Cisco UCS(Unified Computing System) 플랫폼에 구축할 경우 제공되는 종합적인 지원 혜택에 대해 설명합니다. 이 문서에서는 최적의 구축 및 운영 지원을 위해 Cisco와 NVIDIA의 지원을 모두 이용하는 방법에 대해서도 설명합니다.

## 배경 정보

인공 지능 및 기계 학습 워크로드가 기업 운영에 있어 미션 크리티컬한 요소가 되면서 조직에는 강력한 인프라뿐만 아니라 엔터프라이즈급 소프트웨어 및 지원도 필요합니다. NVIDIA AI Enterprise Software는 Cisco UCS 인프라에 구축하도록 최적화된 포괄적이고 안전하고 지원되는 AI 플랫폼을 제공합니다.

이 문서에서는 다음 내용에 대해 설명합니다.

- NVIDIA AI 엔터프라이즈 소프트웨어에 수반되는 사항
- NVAIE Software Suite 라이센스의 지원 혜택
- Cisco UCS에서 NVAIE에 대한 지원 채널을 적절하게 활용하는 방법

## NVIDIA AI 엔터프라이즈 소프트웨어 개요

NVIDIA AI Enterprise란 무엇입니까?

NVAIE(NVIDIA AI Enterprise)는 엔터프라이즈급 보안, 안정성, 관리 편의성 및 가상화된 클라우드 네이티브 환경에서 실행되는 AI 워크로드를 지원하는 포괄적인 소프트웨어 제품군입니다. Cisco UCS를 비롯한 주요 엔터프라이즈 플랫폼에서 실행되도록 최적화되고 인증되었습니다.

## 주요 구성 요소



참고: 최신 전체 목록을 보려면 [Nvidia AI Enterprise](#)를 방문하십시오.

### AI 프레임워크 및 툴:

- NVIDIA AI 워크플로 - 일반적인 AI 애플리케이션을 위한 사전 구축된 엔드 투 엔드 워크플로
- NVIDIA TAO 툴킷 - 컴퓨터 비전, 대화형 AI 등을 위한 이전 학습 툴킷
- NVIDIA Riva - 음성 AI 애플리케이션 구축을 위한 SDK
- NVIDIA Morpheus - 사이버 보안 AI 프레임워크
- NVIDIA Metropolis - 지능형 비디오 분석을 위한 Vision AI 플랫폼
- NVIDIA Clara - 의료 AI 애플리케이션 프레임워크

### 인프라 소프트웨어:

- NVIDIA vGPU 소프트웨어 - GPU 가상화 기술(별도 라이센스 또는 NVAIE에 포함)
- NVIDIA GPU Operator - GPU 관리용 Kubernetes 연산자
- NVIDIA Network Operator - Kubernetes에서 NVIDIA 네트워킹 리소스 관리
- NVIDIA Container Toolkit - GPU 가속 컨테이너를 구축하고 실행하기 위한 도구

### AI 개발 스택:

- 최적화된 AI 프레임워크(TensorFlow, PyTorch, JAX 등)
- 데이터 과학을 위한 NVIDIA RAPIDS
- 프로덕션 AI 구축을 위한 NVIDIA Triton Inference 서버
- 최적화된 추론을 위한 NVIDIA TensorRT
- NVIDIA cuDNN, cuBLAS 및 CUDA 라이브러리

### 관리 및 보안:

- NVIDIA Base Command Manager Essentials - AI용 인프라 관리
- NVIDIA License System - 중앙 집중식 라이센스 관리
- 보안 패치 및 LTS(Long-Term Support) 버전

## 구축 모델

1. 가상화된 환경 - VMware vSphere, Red Hat Virtualization, NVIDIA Virtual GPU
2. 베어 메탈 - 물리적 서버에 직접 구축
3. 클라우드 네이티브 - Kubernetes 및 컨테이너화된 구축
4. 하이브리드 및 멀티 클라우드 - 환경 전반에 걸쳐 일관된 AI 플랫폼

# NVIDIA AI Enterprise 라이센스를 통한 지원 혜택

## 포괄적인 엔터프라이즈 지원

NVIDIA AI Enterprise 라이센스를 취득하면 기본 GPU 드라이버 액세스 이상의 중요한 지원 혜택을 누릴 수 있습니다.

### 1. 프로덕션 등급 소프트웨어 지원

- 소프트웨어 업데이트 및 패치 - 정기적인 보안 업데이트, 버그 수정, 기능 향상
- 장기 지원 릴리스 - 확장 지원 기간이 있는 안정적인 LTS 버전(최대 2년)



참고: Nvidia 지원 레벨(기본 및 24x7)은 [여기](#)에 설명된 대로 구매한 특정 라이센스를 기반으로 합니다.

| 서비스 구성 요소      | 표준        | 24x7       |
|----------------|-----------|------------|
| 기술 지원 액세스      | 현지 업무 시간  | 주 7일, 24시간 |
| 심각도 1 초기 응답 시간 | 4시간(영업시간) | 1시간        |
| 심각도 2 초기 응답 시간 | 4시간(영업시간) | 2시간        |
| 심각도 3 초기 응답 시간 | 영업일 1일    | 4시간(영업시간)  |
| 심각도 4 초기 응답 시간 | 영업일 2일    | 영업일 1일     |
| 고객 포털 액세스      | 24x7      | 24x7       |
| 웹              | 24x7      | 24x7       |
| 전화             | 현지 업무 시간  | 24x7       |

### 2. 기술 전문 지식 및 지침

- AI 구축 지원 - 아키텍처 설계 및 모범 사례 지원
- 성능 최적화 - AI 워크로드를 조정하여 성능을 극대화하는 방법
- 프레임워크 지원 - TensorFlow, PyTorch 및 기타 AI 프레임워크 지원
- 컨테이너 및 Kubernetes 지원 - 클라우드 네이티브 AI 구축 지원
- 기술 자료 액세스 - 포괄적인 문서, 가이드 및 문제 해결 리소스

### 3. 라이센스 및 액세스 혜택

- 검증 및 검증된 소프트웨어 - 엔터프라이즈 테스트 및 인증 릴리스
- 라이센스 이동성 - 인프라 간에 라이센스를 이동할 수 있는 유연성
- 중앙 집중식 라이센스 관리 - 구축 관리를 위한 NVIDIA 라이센스 시스템
- 서브스크립션 모델 - 연간 또는 다년 라이센스 옵션
- 소프트웨어 다운로드 액세스 - 모든 NVAIE 구성 요소에 대한 공식 다운로드 포털

#### 4. 보안 및 규정 준수

- 보안 패치 - 적시의 취약성 개선
- CVE 추적 - 보안 문제에 대한 사전 대응적 알림
- 규정 준수 문서 - 규정 요구 사항 지원
- 보안 소프트웨어 공급망 - 검증 및 서명된 소프트웨어 구성 요소

#### 지원 커버리지

| 기능              | 엔비디아 AI 엔터프라이즈 |
|-----------------|----------------|
| GPU 드라이버 액세스    | ✓              |
| 기본 문서           | ✓              |
| 커뮤니티 포럼         | ✓              |
| AI 소프트웨어 제품군    | ✓              |
| 엔터프라이즈 지원(24/7) | ✓              |
| SLA 응답 시간       | ✓              |
| 보안 패치           | ✓              |
| 장기 지원           | ✓              |
| 생산 인증           | ✓              |
| 기술 어카운트 관리      | 선택적 애드온        |
| AI 워크플로 액세스     | ✓              |
| 관리 툴            | ✓              |

#### 주문 정보

## Cisco UCS 및 NVIDIA AI Enterprise

### 1단계: Cisco UCS 하드웨어 선택

적절한 Cisco UCS 플랫폼 및 NVIDIA GPU 구성을 선택합니다.

컨피그레이션 예:

- UCS C885A M8(NVIDIA HGX H200 GPU 포함)
- UCS C845A M8(NVIDIA RTXP6000 GPU 포함)
- UCS C880A M8(NVIDIA HGX B300 GPU 포함)
- UCS C240 M7(NVIDIA L40S GPU 포함)
- 모듈형 구축을 위한 NVIDIA H100 GPU가 포함된 UCS X-시리즈



참고: 각 NVIDIA H200 NVL, H100 PCIe 및 H100 NVL GPU는 NVIDIA AI Enterprise가 포함된 5년 NVIDIA Enterprise 서브스크립션과 함께 제공됩니다. 소프트웨어 활성화가 필요합니다. 필요한 경우 라이센스를 활성화하고 NVIDIA에서 소프트웨어 지원 케이스를 열어야 합니다.

자세한 내용을 보려면 [여기](#)를 클릭하십시오.

예를 들어, UCS C845A M8, UCSX-210C-M8, UCS C240 M7(NVIDIA H200 GPU 포함)에는 5년 NVIDIA Enterprise 서브스크립션이 포함됩니다.

### 2단계: NVIDIA AI 엔터프라이즈 라이센스

NVIDIA AI Enterprise 라이센스는 별도로 판매되며 다음을 통해 구매할 수 있습니다.

- Cisco Commerce - Cisco UCS 구매와 함께 번들로 제공



참고: NVIDIA 라이센스는 Cisco 외부에서 구매할 수도 있습니다.

라이선스 유형:

- NVAIE-Standard - 표준 연간 서브스크립션(GPU 소켓당 또는 사용자당)
- NVAIE-Educational - 교육용 연간 서브스크립션(GPU 소켓당 또는 사용자당)
- NVAIE-vGPU - NVAIE + vGPU 라이센스 결합

자세한 내용은 [NVIDIA 주문 가이드를 참조하십시오](#)



참고: 구체적인 부품 번호 및 가격은 Cisco 어카운트 팀 또는 NVIDIA 영업 담당자에게 문의하십시오.

### 3단계: 지원 옵션

다음을 갖추어야 합니다.

- Cisco UCS 지원 계약 - Cisco 지원(표준, 고급, 서명)
- NVIDIA AI Enterprise 서비스옵션 - NVIDIA 지원 포함

### 참조 구성

AI 인프라의 주문 예:

1. 하드웨어: UCS C885A M8(NVIDIA H200 GPU 포함)(H200 8개)
1. 소프트웨어: NVIDIA AI Enterprise(8개의 GPU 라이센스)
2. 지원:
  - Cisco 지원 - UCS 하드웨어에 대해 개선됨
  - NVIDIA AI Enterprise 서비스옵션(24/7 지원 포함)
3. Services: 구축을 위한 선택적 전문 서비스

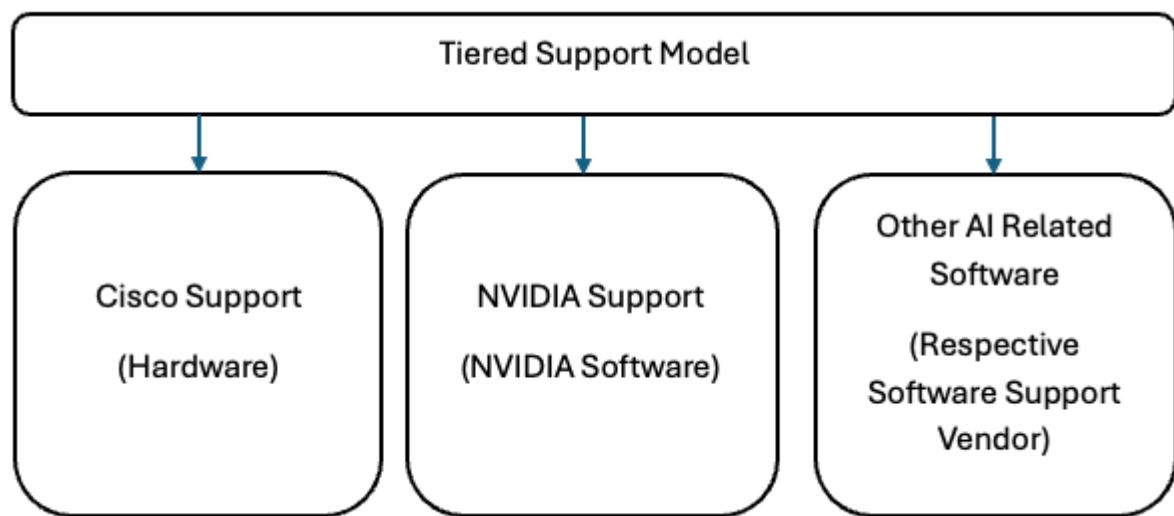
자세한 주문 가이드는 다음을 참조하십시오.

- [Cisco UCS용 NVIDIA AI 엔터프라이즈 주문 가이드](#)

## 지원 참여 방법

지원 모델 개요

Cisco UCS의 NVIDIA AI Enterprise는 하드웨어와 소프트웨어 지원의 명확한 구분이 포함된 계층형 지원 모델을 준수합니다.



참고: NVIDIA 소프트웨어 문제의 경우 NVIDIA에서 직접 지원 케이스를 생성해야 합니다.

---

필요한 경우 공동 트러블슈팅 작업에 Cisco Support도 함께 제공해 주십시오.

---

## Cisco 지원에 문의하는 시기

- 하드웨어 문제:
  - Cisco UCS 서버 장애 또는 오류
  - 하드웨어 진단 및 교체
  - UCS 구성 요소에 대한 펌웨어 업데이트
  - UCS Manager 또는 Intersight 컨피그레이션 문제
  - 물리적 GPU 설치 또는 탐지 문제
  - UCS 인프라의 네트워크 연결 문제
- Cisco 소프트웨어:
  - UCS Manager 또는 Cisco Intersight 문제
  - AI 인프라를 위한 Cisco 네트워크 구성

## Cisco 지원 케이스를 여는 방법:

1. 방문: [Cisco 전 세계 지원 문의처](#)
2. Cisco 지원에 직접 전화(24/7 핫라인)
3. Cisco Support Case Manager 사용
4. Cisco Intersight(클라우드 연결 디바이스용)를 통해 연결

## 지원 케이스를 생성하기 전에 필요한 정보:

- Cisco UCS 모델 및 일련 번호
- 서비스 계약 번호
- 자세한 문제 설명
- 기술 지원 파일(show tech-support)
- GPU 인벤토리 및 컨피그레이션 세부사항

## NVIDIA 지원 문의 시기

### NVIDIA AI 엔터프라이즈 소프트웨어:

- AI 프레임워크 설치 또는 구성 문제
- NVIDIA AI 워크플로 질문
- Triton Inference Server 배포
- NVIDIA vGPU 소프트웨어 문제
- GPU 운영자 또는 네트워크 운영자 문제
- NVIDIA Container Toolkit의 컨테이너 런타임 문제
- 라이센스 활성화 또는 관리
- AI 워크로드의 성능 튜닝
- 소프트웨어 버그 또는 기능 요청
- 보안 취약성 질문

## 드라이버 및 소프트웨어 문제:

- NVIDIA 드라이버 설치 또는 호환성
- CUDA 룰킷 문제
- cuDNN 또는 TensorRT 문제
- GPU 사용률 또는 성능 문제

## 케이스 접수 방법:

1. NVIDIA 엔터프라이즈 [지원 포털 방문](#)
2. NVIDIA Enterprise 계정으로 로그인
3. "Create Case(케이스 생성)"를 선택하고 다음을 제공합니다.
  - NVIDIA AI Enterprise 라이센스 정보
  - 서버 및 GPU 세부 정보
  - 소프트웨어 버전
  - 자세한 문제 설명
  - 로그 및 진단 출력

## 포털 액세스 지원:

- 활성 NVIDIA AI Enterprise 서브스크립션 필요
- 라이선스 프로비저닝 중 계정 설정
- 기술 자료, 다운로드 및 사례 관리 액세스

## 기타 AI 관련 소프트웨어 지원

해당 AI 소프트웨어/앱 벤더에서 제공하는 지원 절차를 준수합니다.

이 문서의 정보는 특정 랩 환경의 디바이스를 토대로 작성되었습니다. 이 문서에 사용된 모든 디바이스는 초기화된(기본) 컨피그레이션으로 시작되었습니다. 현재 네트워크가 작동 중인 경우 모든 명령의 잠재적인 영향을 미리 숙지하시기 바랍니다.

## 이 번역에 관하여

Cisco는 전 세계 사용자에게 다양한 언어로 지원 콘텐츠를 제공하기 위해 기계 번역 기술과 수작업 번역을 병행하여 이 문서를 번역했습니다. 아무리 품질이 높은 기계 번역이라도 전문 번역가의 번역 결과물만큼 정확하지는 않습니다. Cisco Systems, Inc.는 이 같은 번역에 대해 어떠한 책임도 지지 않으며 항상 원본 영문 문서([링크 제공됨](#))를 참조할 것을 권장합니다.