EdgeXpert-Serie

Kleiner Server

MS-C931

Benutzerhandbuch

Inhalt

Erste Schritte 4
- Lieferumfang 4
- Tipps zur sicheren und komfortablen Bedienung 4
- Systemdimension 5
- Systemübersicht 6
- Hardware-Einstellungen 8
- Platzierung des Systems 9
- Stacking des Systems 10

Ersteinrichtung 11
- Was ist NVIDIA DGX™ OS 11
- Merkmale 11

Einrichtung beim ersten Start 12
- Was Sie tun werden 12
- Wählen Sie Ihren Einrichtungsmodus 12
- Vorbereitung 13
- Installationsassistenten starten 13
- Erste Schritte 13
- Was Sie während der Einrichtung erwarten können 14

System-Clusterbildung 16
- Systemanforderungen 16
- Netzwerkkonfiguration zwischen den Systemen 16
- System-Discovery-Skript ausführen 17
- Erforderliche Software installieren und Konfiguration überprüfen 18
- NCCL für zwei Systeme 18
- Fehlerbehebung 21

Aktualisierung des NVIDIA DGX™ OS 22

Neuinstallation (Reimaging) des NVIDIA DGX™ OS 22
- Erstellen eines bootfähigen USB-Flash-Laufwerks 22
- Starten des NVIDIA DGX™ OS ISO-Images 22

Revision

V1.1, 2026/10

NVIDIA Sync 23
Installation 23
Unterstützte Anwendungen 23
Zusätzliche Verbindungsmethoden 23
DGX™ Dashboard 24
Integriertes JupyterLab 24
Zugriff auf das Dashboard 25
NVIDIA Container Runtime für Docker 25
Optional: Benutzer zur Docker-Gruppe hinzufügen 26
Verwendung 26
Validierung 27
Fehlerbehebung 27
NGC 28
Erste Schritte 29
Grundlegende Nutzung 30
Entwicklungsumgebung 30
Bewährte Verfahren 30
Fehlerbehebung 31
Hilfe erhalten 31
Beziehen und Aktivieren eines KI-Modells von der offiziellen NVIDIA-Website . 32
Firmware-Update 32
Empfohlene Methode 32
Manuelle Methode 33
Fehlerbehebung 33
Zusätzliche Ressourcen 33
Safety Instructions 34
Regulatory Notices 37

Erste Schritte

Dieser Teil bietet Ihnen Informationen zur ersten Inbetriebnahme. Bitte achten Sie beim Anschließen des Bildschirms darauf, ihn vorsichtig zu greifen und ein Antistatik-Armband zu tragen, um statische Aufladung zu vermeiden.

Lieferumfang

Kleiner Server	**MS-C931**
Dokumentation	Kurzanleitung
Zubehör	USB-PD-Adapter
	Netzkabel

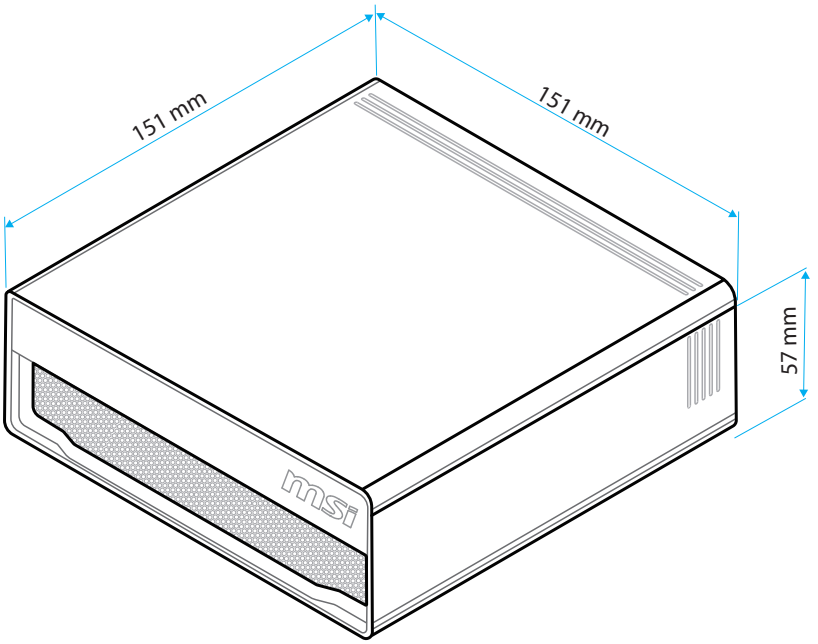
Wichtig

- *Bitte kontaktieren Sie Ihre Verkaufsstelle oder lokalen Händler, falls Teile fehlen oder beschädigt sind.*
- *Der Lieferumfang kann je nach Land variieren.*
- *Das mitgelieferte Netzkabel ist ausschließlich für dieses Gerät bestimmt und sollte nicht mit anderen Produkten verwendet werden.*

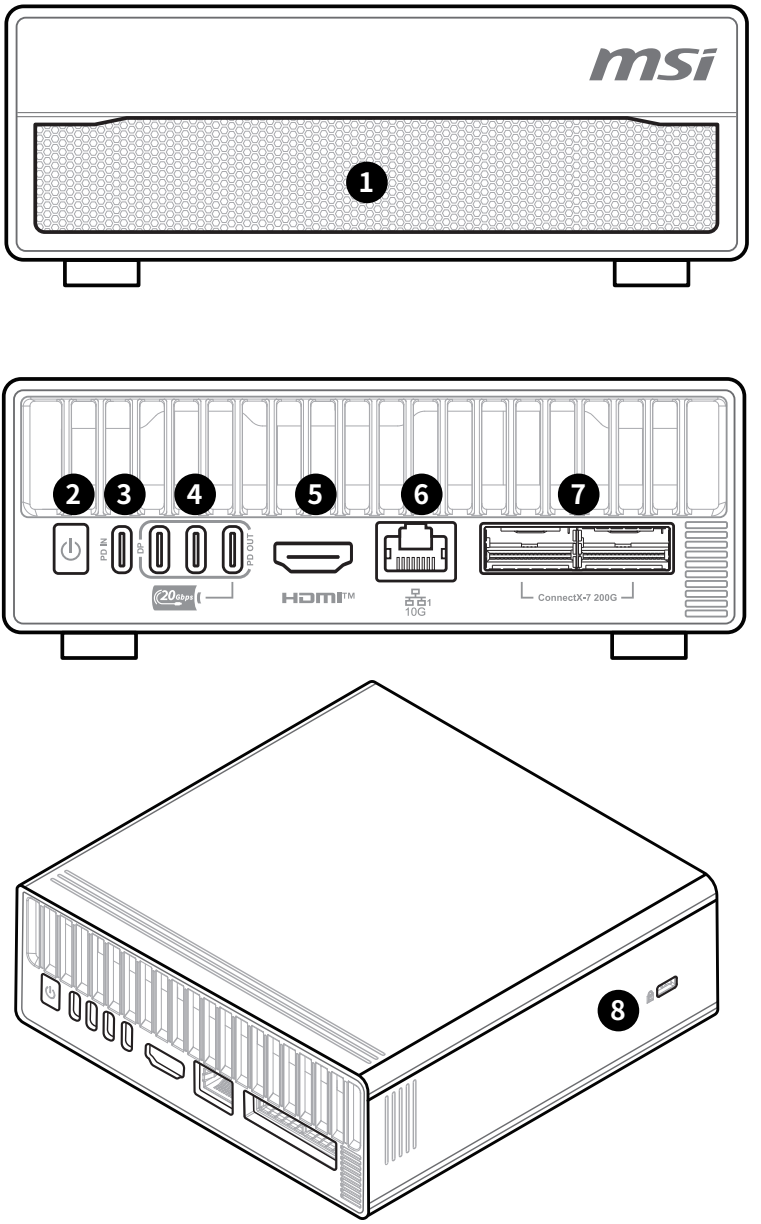
Tipps zur sicheren und komfortablen Bedienung

- Bei längeren Nutzungszeiträumen dieses Geräts ist es besonders wichtig, auf eine ergonomische und angenehme Arbeitsumgebung zu achten, um gesundheitliche Belastungen zu vermeiden.
- Der Arbeitsplatz sollte ausreichend beleuchtet sein.
- Wählen Sie einen geeigneten Schreibtisch und einen guten Stuhl, passen Sie die Höhe an Ihren individuellen Körperbau an.
- Wenn Sie einen Stuhl benutzen, stellen Sie die Rückenlehne so ein, dass diese Ihren Rücken bequem stützt.
- Stellen Sie Ihre Füße flach und in natürlicher Haltung auf den Boden - so, dass Knie und Ellbogen bei der Arbeit um etwa 90° abgewinkelt sind.
- Legen Sie die Hände so auf den Schreibtisch auf, dass Ihre Handgelenke bequem gestützt werden.
- Bitte konfigurieren Sie dieses Gerät so, dass es die Gesundheit nicht beeinträchtigt und Ihre Sitzposition optimal unterstützt.
- Dieses Gerät ist ein Elektrogerät. Bitte beachten Sie die Hinweise und gehen Sie sorgsam mit dem Gerät um.

Systemdimension



Systemübersicht

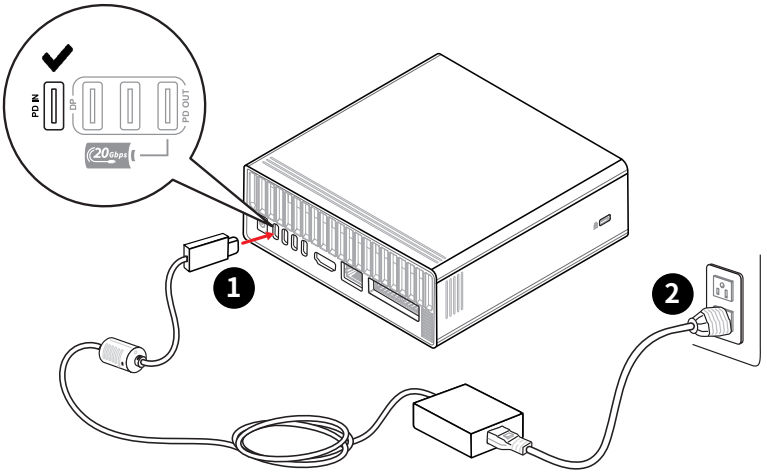


1	**Lüfter** Der Lüfter im Geräteinneren dient der Luftzirkulation und bewahrt das Gerät vor Überhitzung. Blockieren Sie den Lüfter nicht.
2	**Power Button (Ein-/Austaste)** Mit der Ein-/Aus-Taste schalten Sie das System an und aus.
3	**Stromanschluss** Diese Buchse versorgt Ihr Gerät mit Strom.
4	**USB 20 Gbit/s Typ-C Anschluss** Jeder Anschluss kann bis zu 5 V / 3 A Leistung liefern, mit einer maximalen kombinierten Ausgangsleistung von 30 W für drei angeschlossene Geräte.
5	**HDMI™ Anschluss** Unterstützt HDMI™ 2.1.
6	**10 Gbit/s LAN-Anschluss** Der Standard-RJ-45-LAN-Anschluss dient der Verbindung mit einem lokalen Netzwerk (LAN). Hier können Sie ein Netzwerkkabel anschliessen.
7	**200 Gbit/s QSFP-LAN-Anschluss** Verwenden Sie DAC-/AOC-Kabel, um eine Verbindung zu kompatiblen Systemen herzustellen.
8	**Kensington-Schloss** Ihr Gerät ist mit einem Schlitz für ein Kensington-Schloss ausgestattet, damit können Sie Ihr Gerät mit einem festen Gegenstand verbinden und vor Diebstahl schützen. Am Ende des Kabels befindet sich eine kleine Schleife, mit deren Hilfe Sie das Gerät an einem unverrückbaren Gegenstand - zum Beispiel einem schweren Tisch befestigen können, damit es nicht gestohlen wird.

Hardware-Einstellungen

Netzteil anschliessen

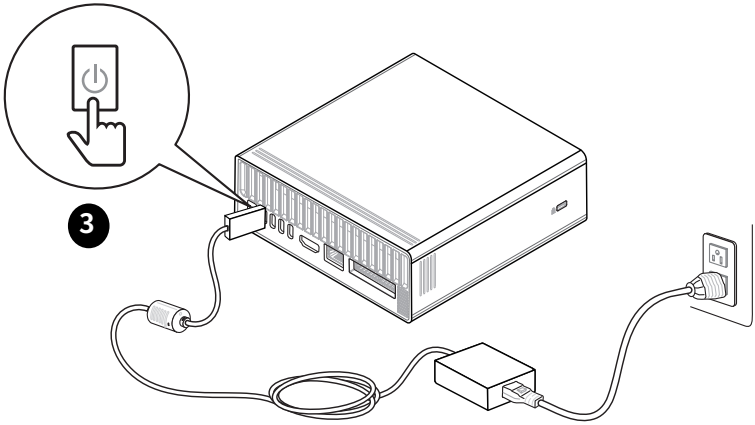
- Externes Netzteil: 240 W, 48,0 V
  - Eingang: 110–120 V AC, 50/60 Hz, 3,5 A / 200–240 V AC, 50/60 Hz, 2,5 A
  - Ausgang: 48,0 V ⎓ 5,0 A



Wichtig

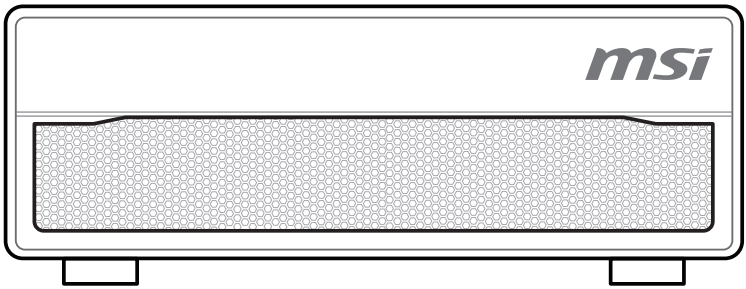
- *Bitte den Adapter verwenden, welcher mit Ihrem Gerät ausgeliefert wird. Die Verwendung eines anderen oder geringer ausgelegten Netzteils kann zu einer verringerten Systemleistung, einem Fehlstart oder zu unerwarteten Abschaltungen führen.*
- *Bitte beachten Sie stets, dass der in Betrieb befindliche Adapter Wärme abgeben kann.*
- *Achten Sie darauf, grundsätzlich den Stecker des Stromkabels zu fassen, wenn Sie den Stecker aus der Steckdose ziehen. Ziehen Sie niemals am Kabel.*

Schalten Sie das System ein.



Platzierung des Systems

Benutzer können das System horizontal aufstellen.

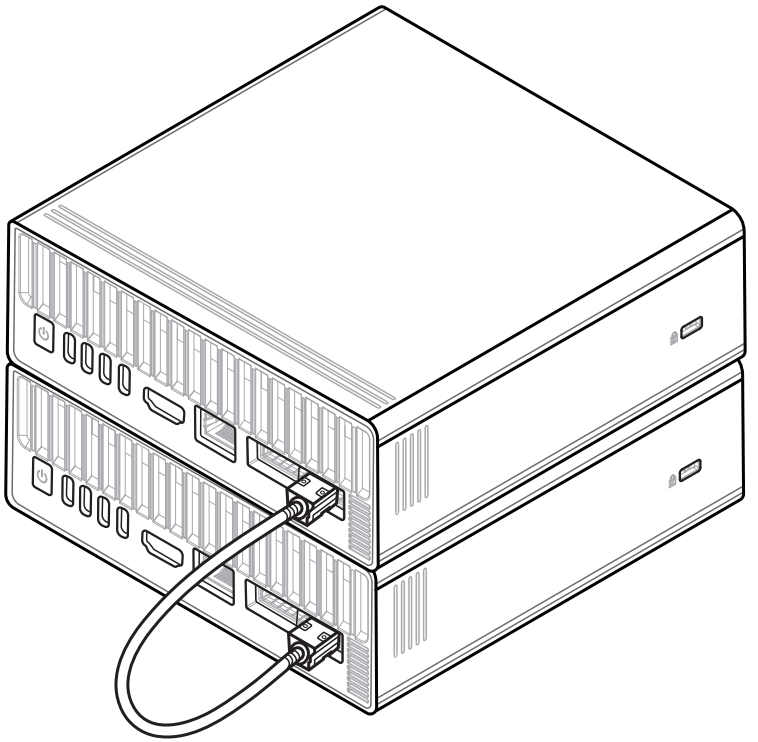


Stacking des Systems

Mit dem optionalen QSFP-Kabel können bis zu zwei Systeme gestackt werden.

Wichtig

- *Das zweite System und das dargestellte QSFP-Kabel dienen ausschließlich zu Illustrationszwecken und sind nicht im Lieferumfang enthalten.*
- *Bitte beachten Sie auch den Abschnitt „System Clustering“ für detailliertere Informationen zum Verbindungsprozess.*



Ersteinrichtung

Wichtig

Alle Informationen und Screenshots können ohne vorherige Ankündigung geändert werden.

Was ist NVIDIA DGX™ OS

Dieses Gerät ist mit NVIDIA DGX™ OS vorinstalliert, um eine schlüsselfertige Lösung für den Betrieb von KI- und Analyse-Workloads bereitzustellen. Die anfängliche Systemkonfiguration erfolgt über einen Einrichtungsassistenten, der nach dem ersten Start ausgeführt wird. Der Einrichtungsassistent bietet den Benutzern ein schnelles Onboarding-Erlebnis für die Nutzung von DGX™-Systemen.

NVIDIA DGX™ OS stellt eine angepasste Installation von Ubuntu Linux mit systemspezifischen Optimierungen und Konfigurationen, zusätzlichen Treibern sowie Diagnose- und Überwachungstools bereit. Es bietet ein stabiles, vollständig getestetes und unterstütztes Betriebssystem, um KI-, Machine-Learning- und Analyseanwendungen auf diesem Gerät auszuführen.

Merkmale

- Vorinstallierte NVIDIA-Treiber und CUDA-Toolkit
- Unterstützt Deep-Learning-Frameworks (z. B. TensorFlow, PyTorch).
- Containerisierte Unterstützung (NVIDIA GPU Cloud Containers + Docker)
- Systemüberwachungs- und Diagnosetools (z. B. Data Center GPU Manager, NVIDIA System Management)
- Unterstützung für die Integration von NGC-Cloud-Ressourcen Ermöglicht Entwicklern, KI/ML-Workloads (Machine Learning) effizient und nahtlos in der Cloud auszuführen.
- Optimierter Kernel, Netzwerk-Stack und I/O zur Steigerung der Gesamtleistung.

Einrichtung beim ersten Start

Diese Anleitung führt Sie durch die Einrichtung Ihres Systems beim ersten Start. Sie wählen aus, wie Sie Ihr System verwenden möchten, und führen den Installationsassistenten aus, um alle Einstellungen zu konfigurieren.

Was Sie tun werden

Dieser Einrichtungsprozess umfasst:

- Auswahl zwischen Desktop-Modus oder Netzwerk-Appliance-Modus.
- Vorbereitung Ihres Systems und der Verbindungen.
- Ausführung des Installationsassistenten zur Konfiguration Ihres Systems.

Wählen Sie Ihren Einrichtungsmodus

Ihr System kann auf zwei Arten konfiguriert werden:

Desktop-Modus

- Schließen Sie Tastatur und Maus über USB oder Bluetooth an.
- Arbeiten Sie direkt über den Ubuntu-Desktop.

Wichtig

Ein USB-C-zu-USB-Adapter ist erforderlich, um eine standardmäßige USB-Tastatur oder -Maus anzuschließen.

Network-Appliance-Modus

- Greifen Sie remote über das Netzwerk auf das System zu.
- Verwenden Sie es als Server oder Rechenknoten.
- Verwalten Sie es ohne lokales Display.

Wichtig

Der Modus, den Sie hier auswählen, wird während des gesamten anfänglichen Einrichtungsprozesses verwendet. Nach Abschluss der Einrichtung können Sie jederzeit frei zwischen Desktop- und Network-Appliance-Modus wechseln. Sie sind nicht dauerhaft an Ihre ursprüngliche Auswahl gebunden.

Vorbereitung

Bevor Sie beginnen, stellen Sie sicher, dass Folgendes vorhanden ist:

- Stromversorgung ist mit dem System verbunden.
- Entweder eine Ethernet-Verbindung mit gültiger Internetverbindung oder ein verfügbares WLAN-Netzwerk mit gültiger Internetverbindung ohne Captive Portal (z. B. in einem Hotel/am Flughafen).
- Für den Desktop-Modus: Display, Tastatur und Maus sind angeschlossen (oder über Bluetooth verfügbar).
- Für den Netzwerk-Appliance-Modus: Ein Computer im selben Netzwerk für den Remote-Zugriff.

Wichtig

Display-Fehlerbehebung: Einige Displays können anfänglich Probleme mit dem System haben. Wenn Sie über USB-C/DisplayPort anschließen und keine Anzeige erhalten, versuchen Sie stattdessen HDMI zu verwenden.

Wichtig

Wenn Sie eine kabelgebundene Netzwerkverbindung verwenden möchten, schließen Sie das Netzwerkkabel vor Beginn der Installation an. Dies hilft, Verbindungsprobleme zu vermeiden, die später im Prozess auftreten könnten.

Installationsassistenten starten

Der Installationsassistent führt Sie durch folgende Schritte:

- Einschalten und Initialisieren des Systems.
- Auswahl Ihres bevorzugten Einrichtungsmodus.
- Herunterladen und Installieren kritischer Updates.
- Abschließen Ihrer Erstkonfiguration.

Wichtig

Kritisch: Schalten Sie das System während des Updatevorgangs nicht aus und starten Sie es nicht neu. Die Installation kann nach Beginn des Downloads nicht unterbrochen werden, und ein Ausschalten während der Updates kann das System beschädigen.

Erste Schritte

Die Art, wie Sie die Installation starten, hängt von Ihrem gewählten Modus ab:

Desktop-Modus

1. Schalten Sie den Computer ein.
2. Der Installationsassistent wird automatisch auf dem angeschlossenen Display gestartet.
3. Verwenden Sie Ihre kabelgebundene Tastatur und Maus (bereits angeschlossen), um zu navigieren.
4. Wenn keine Tastatur oder Maus erkannt wird, werden Sie aufgefordert, Ihre Bluetooth-Geräte in den Kopplungsmodus zu versetzen.

USB-Geräte können jederzeit angeschlossen werden und sollten funktionieren, selbst wenn sie zunächst nicht korrekt erkannt werden. Bluetooth-Geräte können in den Kopplungsmodus versetzt werden und lassen sich in der Regel auch noch auf dem Bildschirm „Erste Schritte“ koppeln (Ausnahme: Tastaturen, die zur Kopplung einen Passcode eingeben müssen; diese funktionieren auf diesem Bildschirm nicht). Sobald Sie auf „Erste Schritte“ klicken, wird die Bluetooth-Kopplung beendet. Um es erneut zu versuchen, müssen Sie das System neu starten.

Network-Appliance-Modus

1. Schalten Sie den Computer ein.
2. Verbinden Sie sich mit dem System über eine der folgenden Methoden:

- Ein Captive-Portal-Bildschirm wird automatisch angezeigt und zeigt die HTTP-Adresse für die Einrichtung an. Diese Einrichtungsseite ist auch auf der Quick-Start-Karte angegeben und sollte etwa so formatiert sein: http://spark-abcd.local
- Öffnen Sie einen Webbrowser und navigieren Sie zu der auf dem Captive-Portal-Bildschirm angezeigten Adresse.
- Schließen Sie bei Bedarf ein Ethernet-Kabel an (optional).

Was Sie während der Einrichtung erwarten können

Der Installationsassistent führt Sie durch mehrere Konfigurationsschritte. Folgen Sie einfach den Anweisungen auf dem Bildschirm, um jeden Schritt abzuschließen.

Schritte des Einrichtungsprozesses:

1. Sprach- und Zeitzonenauswahl

   Wählen Sie Ihre bevorzugte Sprache und Zeitzone für das System aus.

2. Tastaturlayout-Auswahl (nur Desktop-Modus)

   Wählen Sie das gewünschte Tastaturlayout (z. B. US-Tastatur oder russische Tastatur). Dieser Bildschirm erscheint nur im Desktop-Modus.

3. Bedingungen und Konditionen

   Lesen und akzeptieren Sie die Nutzungsbedingungen, um mit der Installation fortzufahren.

4. Benutzerkonto erstellen

   Legen Sie Ihren Benutzernamen und Ihr Passwort für den Systemzugang fest. Hinweis: Die Eingabefelder filtern automatisch beim Tippen, da sie recht lang sind.

5. Einstellungen zur Informationsfreigabe (optional)

   Konfigurieren Sie Analysen und Crash-Reporting nach Ihren Wünschen. Dieser Schritt kann übersprungen werden, falls gewünscht.

6. WLAN-Netzwerkauswahl

   Wählen Sie Ihr WLAN-Netzwerk aus. Dieser Schritt wird automatisch übersprungen, wenn ein Ethernet-Kabel mit Internetverbindung angeschlossen ist.

7. WLAN-Passwort

   Geben Sie das Passwort für Ihr ausgewähltes WLAN-Netzwerk ein.

8. Mit dem WLAN verbinden

   Das System verbindet sich mit Ihrem WLAN und schließt den Access Point. Ihr Computer stellt automatisch die Verbindung zu Ihrem Standardnetzwerk wieder her.

Wichtig

- *Probleme mit der Netzwerkverbindung.*
- *Wenn Ihr Computer automatisch wieder mit demselben Netzwerk wie das System verbunden wird, sollte die Installation nahtlos fortgesetzt werden.*
- *Wenn dies nicht der Fall ist, müssen Sie Ihren Computer mit demselben Netzwerk wie das System verbinden, während die Setup-App auf den Abschluss des Netzwerk-Setups wartet.*
- *Wenn die Einrichtung fehlschlägt, müssen Sie Display/Tastatur/Maus anschließen, um fortzufahren.*
- *Das angezeigte Modal weist Sie an, zu versuchen, sich erneut mit dem Hotspot des Systems zu verbinden und es erneut zu versuchen. Dies funktioniert, wenn das System tatsächlich das Netzwerk nicht beitreten konnte (z. B. falsches Passwort) – im Gegensatz zu Problemen, bei denen Ihr Laptop nicht mit dem System kommunizieren kann.*
- *Wenn der Hotspot nicht verfügbar ist, wenn dieses Fehlermodal erscheint, bedeutet dies, dass das System dem Netzwerk beigetreten ist, Ihr Laptop jedoch keine Kommunikation herstellen kann. Dies kann folgende Ursachen haben:*

  ▶Gerätesegmentierung (Device Isolation)

  ▶Sie sind nicht demselben Netzwerk wie Ihr System beigetreten, oder mDNS funktioniert in Ihrem Netzwerk aufgrund seiner Konfiguration nicht (z. B. in komplexen Unternehmensnetzwerken)

9. Software-Download und Installation

   Sobald das System mit dem Netzwerk verbunden ist, lädt es automatisch das vollständige Software-Image herunter und installiert es.

Wichtig

Schalten Sie das System während dieses Vorgangs nicht aus und starten Sie es nicht neu. Die Installation kann nach Beginn des Downloads nicht unterbrochen werden.

10. Installation abgeschlossen

    Das Gerät wird nach Abschluss der Installation automatisch neu gestartet und kann dann normal verwendet werden.

System-Clusterbildung

Dieser Leitfaden erklärt, wie zwei Systeme zu einem virtuellen Compute-Cluster verbunden werden können, unter Verwendung einer vereinfachten Netzwerkkonfiguration und eines QSFP/CX7-Kabels für die Hochleistungsverbindung.

Ziel ist es, verteilte Workloads über die Grace Blackwell GPUs auszuführen, unter Nutzung von MPI (für die CPU-zu-CPU-Kommunikation zwischen Prozessen) und NCCL v2.28.3 (für GPU-beschleunigte kollektive Operationen).

Weitere Informationen finden Sie im Playbook Connect Two Sparks.

Systemanforderungen

Bevor Sie beginnen, stellen Sie bitte Folgendes sicher:

- Beide Systeme verfügen über Grace Blackwell GPUs, sind über ein QSFP/CX7-Kabel miteinander verbunden und laufen unter Ubuntu 24.04 (oder neuer) mit installierten NVIDIA-Treibern.

Wichtig

- *Diese Anschlüsse unterstützen nur Ethernet-Konfiguration. Zugelassene Kabel für diese Anschlüsse sind:*
- *Amphenol: NJAAKK-N911 (QSFP zu QSFP112, 32AWG, 400 mm, LSZH), NJAAKK0006 ist die 0,5-m-Version dieses Kabels.*
- *Luxshare: LMTQF022-SD-R (QSFP112 400G DAC-Kabel, 400 mm, 30AWG).*

- Die Systeme haben Internetzugang für die anfängliche Softwareeinrichtung.
- Sie verfügen auf beiden Systemen über sudo-/Root-Zugriff.

Netzwerkkonfiguration zwischen den Systemen

Option 1: Führen Sie diese Schritte auf beiden Systemknoten aus, um die Netzwerkschnittstellen mit „netplan“ zu konfigurieren. Die folgenden Befehle müssen in einer Terminal-Sitzung ausgeführt werden (lokal oder remote).

1. Laden Sie die Netplan-Konfigurationsdatei herunter:

   sudo wget -O /etc/netplan/40-cx7.yaml https://github.com/NVIDIA/dgx-spark-playbooks/raw/main/nvidia/connect-two-sparks/assets/cx7-netplan.yaml

2. Setzen Sie die entsprechenden Berechtigungen für die Konfigurationsdatei:

   sudo chmod 600 ／etc／netplan／40-cx7.yaml

3. Wenden Sie die Netplan-Konfiguration an:

   sudo netplan apply

Option 2: Manuelle IP-Zuweisung (Erweitert). Befolgen Sie diese Schritte, um manuell IP-Adressen für das dedizierte Cluster-Netzwerk zu vergeben.

1. Auf Knoten 1 (Node 1): Statische IP-Adresse zuweisen und Schnittstelle aktivieren:
   sudo ip addr add 192.168.100.10／24 dev enP2p1s0f1np1
   sudo ip link set enP2p1s0f1np1 up
2. Auf Knoten 2: Statische IP-Adresse zuweisen und Schnittstelle aktivieren:
   sudo ip addr add 192.168.100.11／24 dev enP2p1s0f1np1
   sudo ip link set enP2p1s0f1np1 up
3. Überprüfen Sie von Knoten 1 aus die Verbindung zu Knoten 2:
   ping -c 3 192.168.100.11
4. Überprüfen Sie von Knoten 2 aus die Verbindung zu Knoten 1:
   ping -c 3 192.168.100.10

System-Discovery-Skript ausführen

Dieser Schritt erkennt automatisch miteinander verbundene Systeme und richtet die SSH-Authentifizierung ein, ohne dass ein Passwort erforderlich ist.

Die folgenden Befehle müssen in einer Terminal-Sitzung (lokal oder remote) auf beiden Knoten ausgeführt werden:

1. Laden Sie das Discovery-Skript herunter:
   wget https://github.com/NVIDIA/dgx-spark-playbooks/raw/refs/heads/main/nvidia/connect-two-sparks/assets/discover-sparks
2. Machen Sie das Skript ausführbar:
   chmod +x discover-sparks
3. Führen Sie das Discovery-Skript aus:
   ./discover-sparks

Beispielausgabe:

```
Found: 192.168.100.10 (spark-1b3b.local)
Found: 192.168.100.11 (spark-1d84.local)
Copying your SSH public key to all discovered nodes using ssh-copy-id.
Auf jedem Knoten werden Sie ggf. nach Ihrem Passwort gefragt.
Copying SSH key to 192.168.100.10 ...
Copying SSH key to 192.168.100.11 ...
nvidia@192.168.100.11' s password:
```

Der SSH-Schlüsselkopierprozess ist abgeschlossen. Die beiden Systeme können nun miteinander kommunizieren.

Erforderliche Software installieren und Konfiguration überprüfen

Nachdem die Netzwerkkonfiguration abgeschlossen ist und die Systeme miteinander kommunizieren können, besteht der nächste Schritt darin, die erforderliche Software für verteilte Workloads zu installieren und Test-Workloads auszuführen. Dadurch wird überprüft, ob die GPU-zu-GPU-Kommunikation korrekt funktioniert, und die Leistung über die gestapelten Systeme hinweg gemessen.

Für vollständige Anweisungen zum Erstellen von NCCL, zum Ausführen der NCCL-Testsuite und zur Interpretation der Ergebnisse siehe das Playbook „NCCL für zwei Sparks“: NCCL for two Sparks playbook.

NCCL für zwei Systeme

Installieren und testen Sie NCCL auf zwei Systemen.

1. Netzwerkkonnektivität konfigurieren

Folgen Sie den Anweisungen zur Netzwerkeinrichtung, um die Konnektivität zwischen Ihren Systemknoten herzustellen. Dies beinhaltet:

- Physische QSFP-Kabelverbindung.
- Konfiguration der Netzwerkschnittstelle (automatische oder manuelle IP-Zuweisung).
- Passwortlose SSH-Einrichtung.
- Überprüfung der Netzwerkkonnektivität.

2. NCCL mit Blackwell-Unterstützung erstellen

Führen Sie diese Befehle auf beiden Knoten aus, um NCCL aus dem Quellcode mit Unterstützung für die Blackwell-Architektur zu erstellen.

```
# Install dependencies and build NCCL
sudo apt-get update && sudo apt-get install -y libopenmpi-dev
git clone -b v2.28.3-1 https://github.com/NVIDIA/nccl.git ~/nccl/
cd ~/nccl/
make -j src.build NVCC_GENCODE="-gencode=arch=compute_121,code=sm_121"

# Set environment variables
export CUDA_HOME="/usr/local/cuda"
export MPI_HOME="/usr/lib/aarch64-linux-gnu/openmpi"
export NCCL_HOME="$HOME/nccl/build/"
export LD_LIBRARY_PATH="$NCCL_HOME/lib:$CUDA_HOME/lib64/:$MPI_HOME/lib:$LD_LIBRARY_PATH"
```

3. NCCL-Testsuite erstellen

Kompilieren Sie die NCCL-Testsuite, um die Kommunikationsleistung zu validieren.

```
# Clone and build NCCL tests
git clone https://github.com/NVIDIA/nccl-tests.git ~/nccl-tests/
cd ~/nccl-tests/
make MPI=1
```

4. Aktive Netzwerkschnittstelle und IP-Adressen finden

Führen Sie den Multi-Node-NCCL-Leistungstest mithilfe der aktiven Netzwerkschnittstelle aus. Identifizieren Sie zunächst, welche Netzwerkports verfügbar und aktiv sind.

```
# Check network port status
ibdev2netdev
```

Beispielausgabe:

```
roceP2p1s0f0 port 1 ==> enP2p1s0f0np0 (Down)
roceP2p1s0f1 port 1 ==> enP2p1s0f1np1 (Up)
rocep1s0f0 port 1 ==> enp1s0f0np0 (Down)
rocep1s0f1 port 1 ==> enp1s0f1np1 (Up)
```

Verwenden Sie eine Schnittstelle, die in Ihrer Ausgabe als „(Up)“ angezeigt wird. In diesem Beispiel verwenden wir enp1s0f1np1. Hinweis: Sie können Schnittstellen mit dem Präfix enP2p<...> ignorieren und sollten nur Schnittstellen berücksichtigen, die mit enp1<...> beginnen.

Sie müssen die IP-Adressen für die aktiven Schnittstellen finden. Führen Sie auf beiden Knoten den folgenden Befehl aus, um die IP-Adressen zu finden, und notieren Sie diese für den nächsten Schritt.

```
ip addr show enp1s0f0np0
ip addr show enp1s0f1np1
```

Beispielausgabe:

```
# In this example, we are using interface enp1s0f1np1.
nvidia@dgx-spark-1:~$ ip addr show enp1s0f1np1
  4: enp1s0f1np1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq
state UP group default qlen 1000
    link/ether 3c:6d:66:cc:b3:b7 brd ff:ff:ff:ff:ff:ff
    inet **169.254.35.62**/16 brd 169.254.255.255 scope link noprefixroute
enp1s0f1np1
     valid_lft forever preferred_lft forever
    inet6 fe80::3e6d:66ff:fecc:b3b7/64 scope link
     valid_lft forever preferred_lft forever
```

In diesem Beispiel lautet die IP-Adresse für Knoten 1 169.254.35.62. Wiederholen Sie den Vorgang für Knoten 2.

5. NCCL-Kommunikationstest ausführen.

Führen Sie die folgenden Befehle auf beiden Knoten aus, um den NCCL-Kommunikationstest zu starten. Ersetzen Sie die IP-Adressen und Schnittstellennamen durch die im vorherigen Schritt gefundenen.

```
# Set network interface environment variables (use your Up interface from the
previous step)
export UCX_NET_DEVICES=enp1s0f1np1
export NCCL_SOCKET_IFNAME=enp1s0f1np1
export OMPI_MCA_btl_tcp_if_include=enp1s0f1np1

# Run the all_gather performance test across both nodes (replace the IP addresses
with the ones you found in the previous step)
mpirun -np 2 -H <IP for Node 1>:1,<IP for Node 2>:1 \
 --mca plm_rsh_agent “ssh -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no” \
 -x LD_LIBRARY_PATH=$LD_LIBRARY_PATH \
 $HOME/nccl-tests/build/all_gather_perf
```

Sie können Ihr NCCL-Setup auch mit einer größeren Puffergröße testen, um die 200 Gbit/s Bandbreite besser auszunutzen.

```
# Set network interface environment variables (use your active interface)
export UCX_NET_DEVICES=enp1s0f1np1
export NCCL_SOCKET_IFNAME=enp1s0f1np1
export OMPI_MCA_btl_tcp_if_include=enp1s0f1np1

# Run the all_gather performance test across both nodes
mpirun -np 2 -H <IP for Node 1>:1,<IP for Node 2>:1 \
 --mca plm_rsh_agent “ssh -o UserKnownHostsFile=/dev/null -o StrictHostKeyChecking=no” \
 -x LD_LIBRARY_PATH=$LD_LIBRARY_PATH \
 $HOME/nccl-tests/build/all_gather_perf -b 16G -e 16G -f 2
```

Hinweis: Den IP-Adressen im mpirun-Befehl folgt :1. Zum Beispiel: mpirun -np 2 -H 169.254.35.62:1,169.254.35.63:1

6. Aufräumen und Zurücksetzen

```
# Rollback network configuration (if needed)
rm -rf ~/nccl/
rm -rf ~/nccl-tests/
```

7. Nächste Schritte

Ihre NCCL-Umgebung ist nun bereit für Multi-Node Distributed Training Workloads auf Ihrem System. Sie können nun größere verteilte Workloads ausführen, z. B. TRT-LLM oder vLLM-Inferenz.

Fehlerbehebung

- Stellen Sie sicher, dass die QSFP/CX7-Schnittstelle aktiv ist und für die IP-Zuweisung verwendet wird.
- Überprüfen Sie die Konnektivität zwischen den Knoten via „ping“.
- Überprüfen Sie Ihre Schnittstellen-Bindungen mit „ip a“ und „ethtool“.
- Wenn das Discovery-Skript fehlschlägt, überprüfen Sie die SSH-Konnektivität zwischen den Knoten manuell.
- Für weitere Anleitungen zur Fehlerbehebung und Unterstützungsmöglichkeiten siehe: Maintenance and Troubleshooting.

Aktualisierung des NVIDIA DGX™ OS

Wenn Sie auf die neueste OS- oder Software-Version aktualisieren möchten, besuchen Sie bitte:

https://ipc.msi.com/product_download/Industrial-Computer-Box-PC/AI-Supercomputer/EdgeXpert-MS-C931

Neuinstallation (Reimaging) des NVIDIA DGX™ OS

Wichtig

Eine Neuinstallation löscht alle auf den OS-Laufwerken gespeicherten Daten. Dies schließt die /home-Partition ein, in der sich alle Benutzerdokumente, Softwareeinstellungen und andere persönliche Dateien befinden.

NVIDIA DGX™ OS ist bereits auf Ihrem Gerät vorinstalliert und erfordert nur in Ausnahmefällen eine Neuinstallation, z. B.:

- Austausch von Speicherlaufwerken
- Wiederaufbau von Cluster-Knoten
- Wiederherstellung nach Systemfehlern

Erstellen eines bootfähigen USB-Flash-Laufwerks

Unter Windows-Systemen finden Sie die Anleitung hier:

https://ipc.msi.com/product_download/Industrial-Computer-Box-PC/AI-Supercomputer/EdgeXpert-MS-C931

Starten des NVIDIA DGX™ OS ISO-Images

1. Stecken Sie das USB-Flash-Laufwerk mit dem OS-Image in das System ein.
2. Schließen Sie Monitor und Tastatur direkt am System an.
3. Starten Sie das System neu und drücken Sie **F2**, sobald das NVIDIA-Logo erscheint, um das Boot-Menü zu öffnen.
4. Wählen Sie den USB-Volumennamen, der dem eingesteckten USB-Flash-Laufwerk entspricht, und booten Sie das System davon.

NVIDIA Sync

NVIDIA Sync ist ein System-Tray-Dienstprogramm, das einen einfachen Zugriff auf Ihr System von einem anderen Rechner ermöglicht, wenn es als Headless-Appliance betrieben wird (ohne Monitor oder Tastatur).

Installation

1. Laden Sie die neueste Version von NVIDIA Sync von der Seite https://build.nvidia.com/spark herunter. Installationsprogramme sind für Windows, macOS und Linux verfügbar.
2. Führen Sie das Installationsprogramm aus.
3. NVIDIA Sync sucht nach kompatiblen Anwendungen, die sich remote mit dem System verbinden können. Wählen Sie die Anwendungen aus, die Sie verwenden möchten, und klicken Sie auf „Weiter“.
4. Geben Sie den Namen des Systems und Ihre Login-Daten ein.

Unterstützte Anwendungen

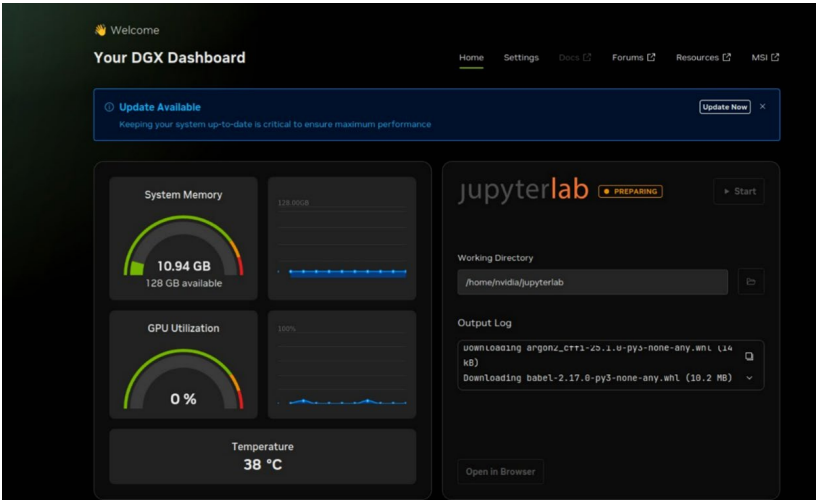
- AI Workbench
- Cursor IDE
- VSCode
- Windsurf

Zusätzliche Verbindungsmethoden

- DGX™ Dashboard (über Webbrowser)
- SSH-Terminal (RSA-Schlüssel werden automatisch von NVIDIA Sync verwaltet)

DGX™ Dashboard

Das System verfügt über ein integriertes Dashboard, das einen Überblick über die aktuellen Betriebsmetriken des Systems bietet, die Möglichkeit, Updates anzuwenden, einige Systemeinstellungen zu ändern und auf lokale Jupyter Notebooks zuzugreifen.



Das DGX™ Dashboard ermöglicht Realtime-Systemüberwachung und bietet integrierten Zugriff auf JupyterLab.

Wichtig

Um Updates auszuführen und den Gerätenamen zu ändern, benötigen Sie „sudo“-Zugriff. Das bei der Ersteinrichtung erstellte Konto verfügt über diesen Zugriff.

Integriertes JupyterLab

Das Dashboard enthält eine integrierte JupyterLab-Instanz, die eine bequeme Entwicklungsumgebung bietet:

- Beim Start erstellt JupyterLab eine virtuelle Umgebung im angegebenen Arbeitsverzeichnis und installiert automatisch eine Reihe empfohlener Pakete.
- Wenn Sie ein neues Arbeitsverzeichnis eingeben und JupyterLab starten, wird eine neue Umgebung erstellt.
- Jedem Benutzerkonto auf dem Gerät wird ein Port zugewiesen, der in der Datei /opt/nvidia/dgx-dashboardservice/jupyterlab_ports.yaml gespeichert ist.
- Um remote auf JupyterLab zuzugreifen, muss ein Tunnel eingerichtet werden – genauso wie beim Dashboard selbst. Der zu tunnelnde Port ist in der Ports-Datei angegeben. Mit NVIDIA Sync wird dieser Tunnel automatisch verwaltet und funktioniert direkt ohne weitere Konfiguration.

Zugriff auf das Dashboard

Das Dashboard kann lokal über die Schaltfläche „Show Apps“ in der unteren linken Ecke des Ubuntu-Desktops geöffnet werden. Wählen Sie anschließend im App-Raster die Verknüpfung „DGX Dashboard“, um das Dashboard im Standard-Webbrowser zu öffnen.

Remote kann auf das Dashboard mit NVIDIA Sync oder über einen manuell erstellten SSH-Tunnel zugegriffen werden.

Bei Verwendung von NVIDIA Sync: Klicken Sie nach dem Verbinden einfach auf die Schaltfläche „DGX Dashboard“, und das Dashboard öffnet sich in Ihrem Standard-Webbrowser unter http://localhost:11000.

Für den manuellen Zugriff über SSH: Öffnen Sie zuerst einen Tunnel, z. B. ssh -L 11000:localhost:11000 <username>@<IP or spark-abcd.local>. Öffnen Sie anschließend das Dashboard in Ihrem Webbrowser unter http://<spark-host-ip>:11000.

NVIDIA Container Runtime für Docker

Die NVIDIA Container Runtime ermöglicht Docker-Containern den Zugriff auf GPU-Ressourcen auf den Systemen. Diese Runtime fungiert als Brücke zwischen Docker und den NVIDIA-Treibern und erlaubt es Containern, GPU-Beschleunigung für AI/ML-Workloads, CUDA-Anwendungen und andere GPU-beschleunigte Software zu nutzen.

Hauptvorteile:

- Nahtloser GPU-Zugriff innerhalb von Containern
- Automatische Treiber- und Bibliotheksverwaltung
- Unterstützung für Multi-GPU-Konfigurationen
- Kompatibilität mit gängigen Container-Orchestrierungsplattformen

Die Runtime arbeitet zusammen mit dem NVIDIA Container Toolkit, das die notwendigen Komponenten bereitstellt, um GPU-Geräte und CUDA-Bibliotheken für containerisierte Anwendungen verfügbar zu machen.

Installation

Das NVIDIA Container Toolkit ist auf dem System bereits vorinstalliert und konfiguriert. Dies umfasst:

- NVIDIA Container Runtime.
- Docker-Integration
- Konfiguration des GPU-Gerätezugriffs
- Verwaltung der CUDA-Bibliotheken.

Die Runtime ist sofort einsatzbereit, um GPU-beschleunigte Container auszuführen.

Optional: Benutzer zur Docker-Gruppe hinzufügen

Standardmäßig erfordert Docker sudo-Rechte, um Befehle auszuführen. Das Hinzufügen Ihres Benutzers zur docker-Gruppe ermöglicht es Ihnen, Docker-Befehle ohne sudo auszuführen, was folgende Vorteile bietet:

- Komfort: Kein Eingeben von sudo vor jedem Docker-Befehl erforderlich.
- Besserer Workflow: Nahtlose Integration mit Entwicklungswerkzeugen und Skripten.
- Geringere Reibung: Schnellere Iterationen beim Arbeiten mit Containern.

So fügen Sie Ihren Benutzer zur Docker-Gruppe hinzu:

```
sudo usermod -aG docker $USER
```

Wichtig

- *Sie müssen sich ab- und wieder anmelden (oder die Sitzung neu starten), damit die Gruppenmitgliedschaft wirksam wird.*
- *Dieser Schritt ist optional. Wenn Sie möchten, können Sie Docker weiterhin mit sudo verwenden, ohne die Gruppenmitgliedschaft zu ändern.*

Verwendung

Grundlegender GPU-Zugriff. Führen Sie einen Container mit GPU-Zugriff über das Flag --gpus aus:

```
docker run -it --gpus=all nvcr.io/nvidia/cuda:13.0.1-devel-ubuntu24.04 nvidia-smi
```

Dieser Befehl: Startet einen interaktiven Container (-it) – ermöglicht Zugriff auf alle GPUs (--gpus=all) – verwendet das NVIDIA CUDA Development Image – führt nvidia-smi aus, um GPU-Informationen anzuzeigen.

GPU-Funktionen festlegen. Steuern Sie, welche GPU-Funktionen dem Container zur Verfügung stehen.

```
docker run -it --gpus '"capabilities=compute,utility"' nvcr.io/nvidia/cuda:13.0.1-devel-ubuntu24.04 nvidia-smi
```

CUDA-Bibliotheken einbinden Für Anwendungen, die bestimmte CUDA-Bibliotheken benötigen, binden Sie diese vom Host ein.

```
docker run -it --gpus=all \
  -v /usr/local/cuda:/usr/local/cuda:ro \
  nvcr.io/nvidia/cuda:13.0.1-devel-ubuntu24.04 bash
```

Validierung

GPU-Zugriff testen.

1. Führen Sie den Testbefehl aus, um den GPU-Zugriff zu überprüfen.
   docker run -it --gpus=all nvcr.io／nvidia／cuda:13.0.1-devel-ubuntu24.04 nvidia-smi

Die erwartete Ausgabe sollte anzeigen: - GPU-Geräteinformationen - Treiberversion - CUDA-Version - Speichernutzung und Temperatur.

2. Runtime-Konfiguration überprüfen.
   docker info | grep -A 10 “Runtimes”
3. Stellen Sie sicher, dass die NVIDIA-Runtime verfügbar ist.
   docker run --rm --runtime=nvidia nvcr.io／nvidia／cuda:13.0.1-devel-ubuntu24.04 nvidia-smi

GPU-Zugriff im Container prüfen. Überprüfen Sie, welche GPU-Ressourcen in einem laufenden Container verfügbar sind.

```
docker run -it --gpus=all nvcr.io／nvidia／cuda:13.0.1-devel-ubuntu24.04 bash

# Inside the container:

nvidia-smi

ls ／dev／nvidia*
```

Fehlerbehebung

Wenn „Runtime not found“-Fehler auftreten.

1. Überprüfen Sie, ob das NVIDIA Container Toolkit installiert ist.
   nvidia-ctk --version
2. Überprüfen Sie die Konfiguration des Docker-Daemons.
   cat ／etc／docker／daemon.json
3. Starten Sie den Docker-Dienst neu.
   sudo systemctl restart docker

Wenn CUDA-Versionskonflikte auftreten.

1. Überprüfen Sie die CUDA-Treiberversion des Hosts.
   nvidia-smi
2. Verwenden Sie ein Container-Image mit kompatibler CUDA-Version.
   docker run -it --gpus=all nvcr.io／nvidia／cuda:12.0.1-devel-ubuntu24.04 nvidia-smi

Wenn Berechtigungsfehler auftreten.

1. Stellen Sie sicher, dass Ihr Benutzer in der docker-Gruppe ist (wenn Sie sudo nicht verwenden).

   groups $USER

2. Überprüfen Sie die Geräteberechtigungen.

   ls -la ／dev／nvidia*

3. Überprüfen Sie, ob der Docker-Daemon Zugriff auf die GPU-Geräte hat.

   sudo docker run -it --gpus=all nvcr.io／nvidia／cuda:13.0.1-devel-ubuntu24.04 nvidia-smi

Wenn Container nicht starten.

1. Überprüfen Sie die Docker-Logs.

   docker logs <container_id>

2. Überprüfen Sie, ob GPU-Geräte auf dem Host verfügbar sind.

   ls ／dev／nvidia*

3. Testen Sie mit einem minimalen Container.

   docker run --rm --gpus=all nvcr.io／nvidia／cuda:13.0.1-devel-ubuntu24.04 echo “GPU test successful”

NGC

NVIDIA GPU Cloud (NGC) ist ein umfassendes Registry-System für GPU-optimierte Container, vortrainierte Modelle und AI/ML-Software, das die schnelle Entwicklung und Bereitstellung von KI-Anwendungen ermöglicht. Für Benutzer bietet NGC Zugriff auf die neuesten Frameworks, Tools und optimierten Umgebungen, die speziell für die Grace Blackwell-Architektur entwickelt wurden.

Wichtige Vorteile für Benutzer:

- Optimierte Container: Vorgefertigte Umgebungen mit den neuesten AI/ML-Frameworks, CUDA und Bibliotheken, optimiert für Grace Blackwell GPUs.
- Vortrainierte Modelle: Zugriff auf State-of-the-Art-Modelle und Modellsammlungen für verschiedene KI-Aufgaben.
- Schnelle Entwicklung: Komplexe Umgebungseinrichtung entfällt – Sie können sich direkt auf Ihre AI/ML-Projekte konzentrieren.
- Spitzen-Software: Zugriff auf den neuesten NVIDIA-Software-Stack und experimentelle Funktionen.

NGC ist besonders wertvoll für Benutzer, da es den aktuellsten und optimierten Softwarestack für diese neue Plattform bereitstellt und somit die neuesten Leistungsoptimierungen und Funktionen verfügbar macht.

Erste Schritte

Erstellen eines NGC-Kontos.

1. Besuchen Sie die NGC-Website.
2. Klicken Sie auf „Sign Up“ und erstellen Sie ein kostenloses Konto.
3. Bestätigen Sie Ihre E-Mail-Adresse.
4. Vervollständigen Sie Ihre Profilinformationen.

API-Schlüssel generieren

1. Melden Sie sich bei Ihrem NGC-Konto an.
2. Navigieren Sie zu „Setup API Key“.
3. Klicken Sie auf „Generate API Key“.
4. Kopieren Sie den API-Schlüssel und speichern Sie ihn sicher.

Wichtig

Ihr API-Schlüssel wird benötigt, um Container herunterzuladen und auf NGC-Ressourcen zuzugreifen. Sichern Sie Ihren API-Schlüssel und geben Sie ihn niemals öffentlich weiter.

NGC CLI installieren (optional). Die NGC CLI bietet einen bequemen Zugriff auf NGC-Ressourcen über die Kommandozeile.

```
# Download and install NGC CLI
wget https://ngc.nvidia.com/downloads/ngccli_linux.zip
unzip ngccli_linux.zip
sudo mv ngc-cli/ngc /usr/local/bin/
ngc config set
```

Mit Docker authentifizieren. Konfigurieren Sie Docker, um auf die NGC-Registries zuzugreifen.

```
# Login to NGC with Docker
docker login nvcr.io
# Username: $oauthtoken
# Password: <your-api-key>
```

Grundlegende Nutzung

Container herunterladen und ausführen. Starten Sie mit einem gängigen AI/ML-Framework-Container.

```
# Pull a PyTorch container optimized for Grace Blackwell
docker pull nvcr.io/nvidia/pytorch:24.08-py3
# Run the container with GPU access
docker run -it --gpus=all nvcr.io/nvidia/pytorch:24.08-py3
```

Verfügbare Ressourcen erkunden Durchsuchen Sie die NGC-Ressourcen über die Weboberfläche.

- Containers: AI/ML-Frameworks, Entwicklungsumgebungen und spezialisierte Tools.
- Modelle: Vortrainierte Modelle für Computer Vision, Natural Language Processing und weitere Aufgaben.
- Helm Charts: Kubernetes-Bereitstellungskonfigurationen.
- Jupyter Notebooks: Interaktive Tutorials und Beispiele.

Entwicklungsumgebung

Entwicklungsumgebung. Verwenden Sie NGC-Container als Ihre Entwicklungsumgebung.

```
# Run a development container with persistent storage
docker run -it --gpus=all \
  -v /path/to/your/project:/workspace \
  nvcr.io/nvidia/pytorch:24.08-py3
```

Modell-Inferenz und Training. Zugriff auf vortrainierte Modelle und Training-Skripte.

```
# Pull a model from NGC
ngc registry model download-version nvidia/bert-base-uncased:1
# Or use models directly in containers
docker run -it --gpus=all \
  nvcr.io/nvidia/tensorflow:24.08-tf2-py3
```

Bewährte Verfahren

Container-Verwaltung.

- Versionen festlegen: Verwenden Sie spezifische Container-Tags für reproduzierbare Umgebungen.
- Regelmäßige Updates: Aktualisieren Sie regelmäßig auf neuere Container-Versionen für die neuesten Optimierungen.
- Ressourcenbegrenzungen: Set appropriate memory and CPU limits for your workloads.

Datenpersistenz.

- Volume-Mounts: Binden Sie Ihre Datenverzeichnisse in Container ein (mount) für die Persistenz.
- Modellspeicher: Speichern Sie trainierte Modelle und Checkpoints außerhalb der Container.
- Konfiguration: Halten Sie Konfigurationsdateien in der Versionskontrolle.

Sicherheit.

- API-Schlüsselsicherheit: Speichern Sie Ihren NGC-API-Schlüssel sicher und ändern Sie ihn regelmäßig.
- Container-Scanning: Scannen Sie Container vor der Verwendung auf Schwachstellen.
- Netzwerksicherheit: Verwenden Sie angemessene Netzwerkkonfigurationen für Ihre Umgebung.

Fehlerbehebung

Authentifizierungsfehler.

```
# Verify your API key is correct
docker login nvcr.io
# Check if your account has access to the requested resource
```

Probleme beim Herunterladen von Containern.

```
# Check network connectivity
ping nvcr.io
# Verify container name and tag
docker search nvcr.io／nvidia／
```

GPU-Zugriffsprobleme.

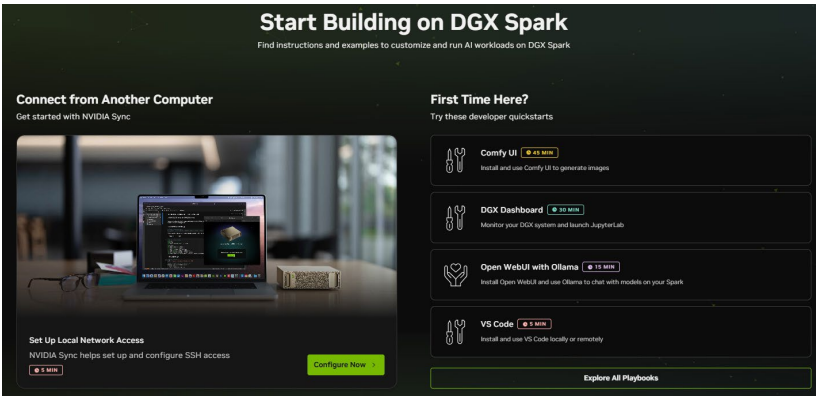
```
# Verify NVIDIA Container Runtime is installed
docker run --rm --gpus=all nvidia／cuda:12.0-base-ubuntu20.04 nvidia-smi
```

Hilfe erhalten

- NGC-Dokumentation: Besuchen Sie die NGC-Dokumentation.
- Community-Foren: Treten Sie den NVIDIA Developer Foren bei.

Beziehen und Aktivieren eines KI-Modells von der offiziellen NVIDIA-Website

Um Anleitungen und Beispiele zum Anpassen und Ausführen von AI-Workloads zu finden, besuchen Sie bitte die NVIDIA Developer Website: https://build.nvidia.com/spark



Firmware-Update

Dieser Abschnitt enthält Anleitungen zum Aktualisieren der Firmware-Komponenten Ihres Systems.

Wichtig

Diese Update-Informationen gelten nur für die Founders Edition. Geräte anderer Hersteller können andere Firmware-Update-Verfahren haben.

Empfohlene Methode

NVIDIA empfiehlt, das DGX™ Dashboard zu verwenden, um Firmware-Updates auf Ihrem System durchzuführen. Das DGX™ Dashboard bietet eine benutzerfreundliche Oberfläche für die Verwaltung von Firmware-Updates und Systemwartungsaufgaben.

Für detaillierte Informationen zum Zugriff auf und zur Nutzung des DGX™ Dashboards siehe Abschnitt DGX™ Dashboard.

Wichtig

- *Stellen Sie sicher, dass Ihr System an eine stabile Stromquelle angeschlossen ist.*
- *Schließen Sie alle laufenden Anwendungen und speichern Sie Ihre Arbeit.*
- *Halten Sie einen Wiederherstellungsplan bereit.*
- *Planen Sie Updates nach Möglichkeit während Wartungsfenstern.*

Manuelle Methode

Wenn Sie das DGX™ Dashboard nicht verwenden können, können Sie die Firmware manuell mithilfe der folgenden Schritte aktualisieren:

1. Öffnen Sie ein Remote- oder lokales Terminal auf dem System.
2. Führen Sie die folgenden Befehle aus:

   sudo apt update

   sudo apt upgrade

   sudo fwupdmgr refresh

   sudo fwupdmgr upgrade

   sudo reboot

Fehlerbehebung

Wenn während der Firmware-Updates Probleme auftreten:

- Stellen Sie eine stabile Stromversorgung während des Update-Vorgangs sicher.
- Für weitere Anleitungen zur Fehlerbehebung und Support-Optionen siehe: spark-maintenance-troubleshooting.

Zusätzliche Ressourcen

- Besuchen Sie das NVIDIA Spark Developer Portal unter https://build.nvidia.com/spark
  für die neuesten Anleitungen, Tutorials und Updates.
- Siehe spark-release-notes für die neuesten Software-Updates und Funktionen.
- Siehe spark-known-issues, um häufige Probleme zu beheben.

Ihr System ist nun bereit, Ihre AI-Entwicklungs- und Bereitstellungs-Workflows zu unterstützen!

Safety Instructions

- Read the safety instructions carefully and thoroughly.
- All cautions and warnings on the device or User Guide should be noted.
- Refer servicing to qualified personnel only.
- IEC 60825-1 :2014 transfer to FDA/CDRH Complies with FDA performance standards for laser products except for conformance with IEC 60825-1 Ed.3., as described in Laser Notice No. 56, dated May 8, 2019.
- The SFP ports should use UL Listed Optional Transceiver product, Rated 3.3Vdc, Laser Class 1.

Power

- Make sure that the power voltage is within its safety range and has been adjusted properly to the value of 100~240V before connecting the device to the power outlet.
- If the power cord comes with a 3-pin plug, do not disable the protective earth pin from the plug. The device must be connected to an earthed mains socket-outlet.
- Please confirm the power distribution system in the installation site shall provide the circuit breaker rated 120/240V, 20A (maximum).
- Always unplug the power cord before installing any add-on card or module to the device.
- Always disconnect the power cord or switch the wall socket off if the device would be left unused for a certain time to achieve zero energy consumption.
- Place the power cord in a way that people are unlikely to step on it. Do not place anything on the power cord.
- If this device comes with an adapter, use only the MSI provided AC adapter approved for use with this device.

Battery

Please take special precautions if this device comes with a battery.

- Danger of explosion if battery is incorrectly replaced. Replace only with the same or equivalent type recommended by the manufacturer.
- Avoid disposal of a battery into fire or a hot oven, or mechanically crushing or cutting of a battery, which can result in an explosion.
- Avoid leaving a battery in an extremely high temperature or extremely low air pressure environment that can result in an explosion or the leakage of flammable liquid or gas.
- Do not ingest battery. If the coin/button cell battery is swallowed, it can cause severe internal burns and can lead to death. Keep new and used batteries away from children.

European Union:



Batteries, battery packs, and accumulators should not be disposed of as unsorted household waste. Please use the public collection system to return, recycle, or treat them in compliance with the local regulations.

Battery Recycle:



廢電池請回收

For better environmental protection, waste batteries should be collected separately for recycling or special disposal.

California, USA:



The button cell battery may contain perchlorate material and requires special handling when recycled or disposed of in California.
For further information please visit: https://dtsc.ca.gov/perchlorate/

⚠WARNING

- **INGESTION HAZARD**: This product contains a button cell or coin battery.
- **DEATH** or serious injury can occur if ingested.
- A swallowed button cell or coin battery can cause **Internal Chemical Burns** in as little as **2 hours**.
- **KEEP** new and used batteries **OUT OF REACH OF CHILDREN**
- **Seek immediate medical attention** if a battery is suspected to be swallowed or inserted inside any part of the body.



- Remove and immediately recycle or dispose of used batteries according to local regulations and keep away from children. Do NOT dispose of batteries in household trash or incinerate.
- Even used batteries may cause severe injury or death. Call a local poison control center for treatment information.
- Battery type: CR2032
- Battery voltage: 3V
- Non-rechargeable batteries are not to be recharged.
- Do not force discharge, recharge, disassemble, heat above (manufacturer' s specified temperature rating) or incinerate. Doing so may result in injury due to venting, leakage or explosion resulting in chemical burns.
- This product contains an irreplaceable battery.
- This icon indicates that a swallowed button battery can cause serious injury or death. Please keep batteries out of sight or reach of children.

Environment Information

- To reduce the possibility of heat-related injuries or of overheating the device, do not place the device on a soft, unsteady surface or obstruct its air ventilators.
- Use this device only on a hard, flat and steady surface.
- To prevent fire or shock hazard, keep this device away from humidity and high temperature.
- Do not leave the device in an unconditioned environment with a storage temperature above 60°C or below -20°C, which may damage the device.
- The operating temperature range is approximately 0°C to 35°C.
- When cleaning the device, be sure to remove the power plug. Use a piece of soft cloth rather than industrial chemical to clean the device. Never pour any liquid into the opening; that could damage the device or cause electric shock.
- Always keep strong magnetic or electrical objects away from the device.
- If any of the following situations arises, get the device checked by service personnel:
  - The power cord or plug is damaged.
  - Liquid has penetrated into the device.
  - The device has been exposed to moisture.
  - The device does not work well or you can not get it working according to the User Guide.
  - The device has dropped and damaged.
  - The device has obvious sign of breakage.

Regulatory Notices

CE Conformity

Products bearing the CE marking comply with one or more of the following EU Directives as may be applicable:

- RED 2014/53/EU
- Low Voltage Directive 2014/35/EU
- EMC Directive 2014/30/EU
- RoHS Directive 2011/65/EU
- Implementing measure Directive 2009/125/EC of ESPR Regulation (EU) 2024/1781

Compliance with these directives is assessed using applicable European Harmonized Standards.

For any support regarding the EU General Product Safety Regulation (GPSR), please contact MSI Computer Europe B.V. via gpsr@msi.com Churchilllaan 202, 5705 BK Helmond, the Netherlands.

Products with Radio Functionality (EMF)

This product incorporates a radio transmitting and receiving device. For computers in normal use, a separation distance of 20 cm ensures that radio frequency exposure levels comply with EU requirements. Products designed to be operated at closer proximities, such as tablet computers, comply with applicable EU requirements in typical operating positions. Products can be operated without maintaining a separation distance unless otherwise indicated in instructions specific to the product.

Restrictions for Products with Radio Functionality (select products only)



CAUTION: IEEE 802.11x wireless LAN with 5.15~5.35 GHz frequency band is restricted for indoor use only in all European Union member states, EFTA (Iceland, Norway, Liechtenstein), and most other European countries (e.g., Switzerland, Turkey, Republic of Serbia). Using this WLAN application outdoors might lead to interference issues with existing radio services.



Radio frequency bands and maximum power levels

- Features: Wi-Fi 7, BT
- Frequency Range:

  2.4 GHz: 2.412~2.484GHz

  5 GHz: 5.180~5.895GHz

  6 GHz: 5.925~7.125GHz
- Max Power Level:

  2.4 GHz: 20dBm

  5 GHz: 23dBm

  6 GHz: 23dBm

FCC-B Radio Frequency Interference Statement



This equipment has been tested and found to comply with the limits for a Class B digital device, pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference in a residential installation. This equipment generates, uses and can radiate radio frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. However, there is no guarantee that interference will not occur in a particular installation. If this equipment does cause harmful interference to radio or television reception, which can be determined by turning the equipment off and on, the user is encouraged to try to correct the interference by one or more of the measures listed below:

- Reorient or relocate the receiving antenna.
- Increase the separation between the equipment and receiver.
- Connect the equipment into an outlet on a circuit different from that to which the receiver is connected.
- Consult the dealer or an experienced radio/television technician for help.

Notice 1

The changes or modifications not expressly approved by the party responsible for compliance could void the user' s authority to operate the equipment.

Notice 2

Shielded interface cables and AC power cord, if any, must be used in order to comply with the emission limits.

This device complies with Part 15 of the FCC Rules. Operation is subject to the following two conditions:

- this device may not cause harmful interference, and
- this device must accept any interference received, including interference that may cause undesired operation.

MSI Computer Corp.

901 Canada Court, City of Industry, CA 91748, USA

(626) 913-0828

www.msi.com

- Any changes or modifications not expressly approved by the party responsible for compliance could void the authority to operate equipment.
- This device and its antenna must not be co-located or operating in conjunction with any other antenna or transmitter.
- End-users and installers must be provided with antenna installation instructions and transmitter operating conditions for satisfying RF exposure compliance.

- In the users manual of the end product, the end user has to be informed to keep at least 20cm separation with the antenna while this end product is installed and operated. The end user has to be informed that the FCC radio frequency exposure guidelines for an uncontrolled environment can be satisfied. The end user has to also be informed that any changes or modifications not expressly approved by the manufacturer could void the user' s authority to operate this equipment.
- If the size of the end product is smaller than 8x10cm, then additional FCC part 15.19 statement is required to be available in the users manual: This device complies with Part 15 of FCC rules. Operation is subject to the following two conditions: (1) this device may not cause harmful interference and (2) this device must accept any interference received, including interference that may cause undesired operation.
- FCC regulations restrict the operation of this device to indoor use only. Operation prohibited on oil platforms, cars, trains, boats, and aircraft, except that operation of this device is permitted in large aircraft while flying above 10,000 feet.

WEEE Statement



European Union: This symbol on the product indicates that this product cannot be discarded as municipal waste. Instead, it is your responsibility to dispose of your waste electrical and electronic equipment by handing it over to a designated collection point for recycling. For more information about where you can drop off your waste equipment for recycling, please contact your local city office, your household waste disposal service or the shop where you purchased the product.

Thailand Compliance Statement

"เครื่องวิทยุคมนาคมนี้มีระดับการแผ่คลื่นแม่เหล็กไฟฟ้าสอดคล้องตามมาตรฐานความปลอดภัยต่อสุขภาพของมนุษย์จากการใช้เครื่องวิทยุคมนาคมที่คณะกรรมการกิจการโทรคมนาคมแห่งชาติประกาศกำหนด"

NCC無線設備警告聲明

取得審驗證明之低功率射頻器材，非經核准，公司、商號或使用者均不得擅自變更頻率、加大功率或變更原設計之特性及功能。低功率射頻器材之使用不得影響飛航安全及干擾合法通信；經發現有干擾現象時，應立即停用，並改善至無干擾時方得繼續使用。前述合法通信，指依電信管理法規定作業之無線電通信。低功率射頻器材須忍受合法通信或工業、科學及醫療用電波輻射性電機設備之干擾。

Canadian Compliance Statement

This device complies with Industry Canada license-exempt RSSs. Operation is subject to the following two conditions:

1) This device may not cause interference, and

2) This device must accept any interference, including interference that may cause undesired operation of the device.

Le présent appareil est conforme aux CNR d' Industrie Canada applicables aux appareils radio exempts de licence . L' exploitation est autorisée aux deux conditions suivantes :

1) l' appareil ne doit pas produire de brouillage ;

2) l' utilisateur de l' appareil doit accepter tout brouillage radioélectrique subi, meme si le brouillage est susceptible d' en compromettre le fonctionnement .

Caution

1) Any devices capable of operating in the band 5150–5250 MHz shall only be used indoors to reduce the potential for harmful interference to co-channel mobile satellite systems (this requirement does not apply to OEM devices installed in vehicles by vehicle manufacturers);

2) For devices with detachable antenna(s), the maximum antenna gain permitted for devices in the bands 5250-5350 MHz and 5470-5725 MHz shall be such that the equipment still complies with the e.i.r.p. limit;

3) For devices with detachable antenna(s), the maximum antenna gain permitted for devices in the band 5725-5850 MHz shall be such that the equipment still complies with the e.i.r.p. limits as appropriate; and

4) Where applicable, antenna type(s), antenna models(s), and worst-case tilt angle(s) necessary to remain compliant with the e.i.r.p. elevation mask requirement set forth in section 7.3.2.4 or 7.3.5.3 shall be clearly indicated.

Avertissement

1) tout dispositif capable de fonctionner dans la bande de 5150 à 5250 MHz ne doit être utilisé qu' à l' intérieur des bâtiments afin de réduire les risques d' interférences nuisibles avec les systèmes mobiles par satellite à canaux multiples (cette exigence ne s' applique pas aux dispositifs FEO installés dans les véhicules par les constructeurs automobiles);

2) pour les dispositifs munis d' antennes amovibles, le gain maximal d' antenne permis pour les dispositifs utilisant les bandes de 5 250 à 5 350 MHz et de 5 470 à 5 725 MHz doit être conforme à la limite de la p.i.r.e;

3) pour les dispositifs munis d' antennes amovibles, le gain maximal d' antenne permis (pour les dispositifs utilisant la bande de 5 725 à 5 850 MHz) doit être conforme à la limite de la p.i.r.e. spécifiée, selon le cas;

4) lorsqu' il y a lieu, les types d' antennes (s' il y en a plusieurs), les numéros de modèle de l' antenne et les pires angles d' inclinaison nécessaires pour rester conforme à l' exigence de la p.i.r.e. applicable au masque d' élévation, énoncée à la section 7.3.2.4 ou 7.3.5.3, doivent être clairement indiqués.

Radiation Exposure Statement

This equipment complies with IC RSS-102 radiation exposure limits set forth for an uncontrolled environment. This equipment should be installed and operated with minimum distance 20cm between the radiator & your body.

Déclaration d' exposition aux radiations

Cet équipement est conforme aux limites d' exposition aux rayonnements IC établies pour un environnement non contrôlé. Cet équipement doit être installé et utilisé avec un minimum de 20 cm de distance entre la source de rayonnement et votre corps.

This product meets the applicable Innovation, Science and Economic Development Canada technical specifications.

Devices shall not be used for control of or communications with unmanned aircraft systems.

Les dispositifs ne doivent pas être utilisés pour commander des systèmes d' aéronef sans pilote ni pour communiquer avec de tels systèmes.

Operation shall be limited to indoor use only.

Operation on oil platforms, automobiles, trains, maritime vessels and aircraft shall be prohibited except for on large aircraft flying above 3,048 m (10,000 ft).

leur utilisation doit être limitée à l' intérieur seulement;

leur utilisation à bord de plateformes de forage pétrolier, d' automobiles, de trains, de navires maritimes et d' aéronefs doit être interdite, sauf à bord d' un gros aéronef volant à plus de 3 048 m (10 000 pi) d' altitude.

Chemical Substances Information

In compliance with chemical substances regulations, such as the EU REACH Regulation (Regulation EC No. 1907/2006 of the European Parliament and the Council), MSI provides the information of chemical substances in products at:
https://csr.msi.com/global/index

RoHS Statement

日本JIS C 0950材質宣言

日本工業規格JIS C 0950により、2006年7月1日以降に販売される特定分野の電気および電子機器について、製造者による含有物質の表示が義務付けられます。

https://csr.msi.com/tw/Japan-JIS-C-0950-Material-Declarations

India RoHS

This product complies with the “India E-waste (Management and Handling) Rule 2016” and prohibits use of lead, mercury, hexavalent chromium, polybrominated biphenyls or polybrominated diphenyl ethers in concentrations exceeding 0.1 weight % and 0.01 weight % for cadmium, except for the exemptions set in Schedule 2 of the Rule.

Türkiye EEE yönetmeliği

Türkiye Cumhuriyeti: EEE Yönetmeliğine Uygundur.

Україна обмеження на наявність небезпечних речовин

Обладнання відповідає вимогам Технічного регламенту щодо обмеження vvi використання деяких небезпечних речовин в електричному та електронному обладнані, затвердженого постановою Кабінету Міністрів України від 3 грудня 2008 № 1057.

Việt Nam RoHS

Kể từ ngày 01/12/2012, tất cả các sản phẩm do công ty MSI sản xuất tuân thủ Thông tư số 30/2011/TT-BCT quy định tạm thời về giới hạn hàm lượng cho phép của một số hóa chất độc hại có trong các sản phẩm điện, điện tử”

产品中有害物质的名称及含有信息表

部件名称	有害物质									
	Pb	Hg	Cd	Cr(VI)	PBBs	PBDEs	DBP	DIBP	BBP	DEHP
电路板组件*	×	○	○	○	○	○	○	○	○	○
处理器和散热器	×	○	○	○	○	○	○	○	○	○
内存条/硬盘	×	○	○	○	○	○	○	○	○	○
电缆/连接器	×	○	○	○	○	○	○	○	○	○
输出输入设备	×	○	○	○	○	○	○	○	○	○
电源供应器/适配器	×	○	○	○	○	○	○	○	○	○
金属机构件	×	○	○	○	○	○	○	○	○	○

注1：○：表示该有害物质在该部件所有均质材料中的含量均不超出电器电子产品有害物质限制使用国家标准要求。

×：表示该有害物质至少在该部件的某一均质材料中的含量超出电器电子产品有害物质限制使用国家标准要求。

注2：以上未列出的部件，表明其有害物質含量均不超出电器电子产品有害物质限制使用国家标准要求。

注3：上述表格标注"×"之部件，皆符合达标管理目录限用物质应用例外清单之限值要求。

* 电路板组件：包括印刷电路板及其构成的零部件。

限用物質含有情況標示聲明書

單元	限用物質及其化學符號					
	鉛 (Pb)	汞 (Hg)	鎘 (Cd)	六價鉻 (Cr^{+6})	多溴聯苯 (PBB)	多溴二苯醚 (PBDE)
電路板總成	—	○	○	○	○	○
儲存裝置	—	○	○	○	○	○
輸出/入裝置	—	○	○	○	○	○
電源供應器	—	○	○	○	○	○
金屬機構件	—	○	○	○	○	○
塑膠機構件	○	○	○	○	○	○
風扇	—	○	○	○	○	○
配件(例:電源線等)	—	○	○	○	○	○

備考1. “超出0.1 wt %” 及 “超出0.01 wt %” 係指限用物質之百分比含量超出百分比含量基準值。

備考2. “○” 係指該項限用物質之百分比含量未超出百分比含量基準值。

備考3. “—” 係指該項限用物質為排除項目。

Environmental Policy

- The product has been designed to enable proper reuse of parts and recycling and should not be thrown away at its end of life.
- Users should contact the local authorized point of collection for recycling and disposing of their end-of-life products.
- Visit the MSI website <https://csr.msi.com/global/pevn_ewaste> and locate a nearby distributor for further recycling information.
- Please visit <https://us.msi.com/page/recycling> for information regarding the recycling of your product in the US.

Warranty

For any further information about the product users purchased, please contact the local dealer. Do not attempt to upgrade or replace any component of the product.

Copyright and Trademarks Notice

Copyright © Micro-Star Int' l Co., Ltd. All rights reserved. The MSI logo used is a registered trademark of Micro-Star Int' l Co., Ltd. All other marks and names mentioned may be trademarks of their respective owners. No warranty as to accuracy or completeness is expressed or implied. MSI reserves the right to make changes to this document without prior notice.

HIGH-DEFINITION MULTIMEDIA INTERFACE

The terms HDMI™, HDMI™ High-Definition Multimedia Interface, HDMI™ Trade dress and the HDMI™ Logos are trademarks or registered trademarks of HDMI™ Licensing Administrator, Inc.

Technical Support

If a problem arises with your system and no solution can be obtained from the user' s manual, please contact your place of purchase or local distributor. Alternatively, please try the following help resources for further guidance. Visit the MSI website for technical guide, BIOS updates, driver updates and other information via https://www.msi.com/support/