

Leistungsbeschreibung „Nvidia GPU-Server 2025-183“

Inhaltsverzeichnis

Inhaltsverzeichnis	1
1. Ziel der Ausschreibung.....	1
2. Allgemein	2
3. Garantiebedingungen und Ersatzteillieferung	2
4. Gehäuse	2
5. Netzteile.....	2
6. Deep Learning Server	3
6.1. Grundlegendes	3
6.2. CPU	3
6.3. GPU	3
6.4. Arbeitsspeicher	3
6.5. Massenspeicher.....	3
6.6. Board Management Controller (BMC).....	3
6.7. Anschlusstechnik für Daten- und Energieversorgung	4
6.8. Software und Softwareserviceleistungen.....	4
7. Option	4
8. Zuschlagskriterium	4
9. Lieferung	4
10. Rechtsfolgen bei Terminüberschreitung.....	5
11. Elektrische Sicherheit	5
12. Abnahmeverfahren	5
13. Dokumentation	5

1. Ziel der Ausschreibung

Die TU Dortmund beabsichtigt den Kauf eines GPU Servers zur Erweiterung und Erneuerung der bestehenden IT-Infrastruktur.

Mitarbeiter*innen soll erlaubt werden im Bereich der künstliche Intelligenz Modelle hoher Komplexität in akzeptablen Zeiträumen zu erforschen. Durch die Beschränkung der Ausschreibung auf Server mit Rechenbeschleunigern mit hoher Speicher-, insbesondere GPU-Speicherkapazität, soll die effiziente Ausführung speicherintensiver Anwendungen, insbesondere Deep Neural Networks, gewährleistet werden. Der Server stellt eine Ergänzung der bestehenden IT-Infrastruktur der Technischen Universität Dortmund dar und unterstützt die

Erforschung komplexer KI-Modelle innerhalb akzeptabler Zeiträume.

Um den Administrationsaufwand zu reduzieren und für die Forschenden die volle Kompatibilität mit vorhandener Hardware zu gewährleisten wird der Kern der Ausschreibung auf Beschleuniger des Herstellers Nvidia eingeschränkt. Dies erlaubt durch die weite Verbreitung einen besseren Wissenstransfer, aber auch die Nutzung exklusiver Technologien wie z.B. Nvidia CUDA, cuDNN, Omniverse, NeMo.

Um zukünftigen forschungsrelevanten Problem- und Fragestellungen gewachsen zu sein, ist eine Server interne Vernetzung der Beschleuniger (z.B. Nvidia NVSwitch) unabdinglich um den nötigen Gesamtspeicher von mehr als 2TB bereitzustellen. Derzeit existiert bereits ein System aus DGX A100-Knoten, in das sich der neue Server einfügen soll. Zurzeit ist das Nvidia DGX B300 System die einzig verfügbare Lösung der die obigen Anforderungen erfüllt und damit Kerngegenstand dieser Ausschreibung.

2. Allgemein

- Die Ausschreibung erfolgt über ein einzelnes Los.
- Der Server muss die Fernwartbarkeit in Bezug auf Temperatur-, Systemfehlerüberwachung und Powermanagement unterstützen.
- Jede Komponente muss für den 24/7-Betrieb ausgelegt sein.

3. Garantiebedingungen und Ersatzteillieferung

Der Auftragnehmer gewährt eine Garantie von mindestens drei Jahren auf die volle Funktionsfähigkeit der gelieferten Systeme einschließlich aller Komponenten und Verschleißteile, gerechnet ab dem Datum der erfolgreichen Abnahme gemäß Leistungsbeschreibung.

Im Garantiefall erfolgt auf Anforderung des Auftraggebers und in Absprache mit diesem eine Reparatur innerhalb von 6 Werktagen (Montag bis Freitag, gesetzliche Feiertage ausgenommen) nach Eingang der Störungsmeldung. Transport von allen benötigten Teilen, inklusive Rücksendung der ausgetauschten Komponenten liegt in der Verantwortung des Auftragnehmers.

Im Falle eines „einfachen“ Austausch (Einschübe, PCIe, M.2, U.2, U.3, Netzteile, USB, SATA, SAS, DIM) hat der Auftraggeber (TU Dortmund) die Möglichkeit einen eigenständigen Tausch, ohne Auswirkungen auf die Garantie, anzubieten um den Prozess zu vereinfachen. In diesem Fall hat der Auftragnehmer. In diese Fall hat der Auftragnehmer alle für den Austausch Kleinteil, Hilfsstoffe und Spezialwerksein ohne Mehrkosten zu Verfügung zu stellen.

Der Auftraggeber ist berechtigt, im Rahmen des Auftrags enthaltene Herstellersupportverträge unmittelbar und ohne Mitwirkung des Auftragnehmers in Anspruch zu nehmen.

4. Gehäuse

- Die Gehäuse aller Server müssen so gebaut sein, dass diese für den Einbau in symmetrische 19“ Racks geeignet sind.
- Die Kühlung der Server muss ausschließlich durch Kaltluft erfolgen und so ausgelegt sein, dass die Kaltseite vorne und die Warmseite (Anschlussseite) hinten liegt und dadurch eine saubere Trennung von Warm- und Kaltseite gewährleistet ist.
- Alle Einbau-, Befestigungsmaterialien müssen im Lieferumfang enthalten sein, dazu zählen auch geeignete Teleskopschienen für ein symmetrisches Rack.

5. Netzteile

Netzteile / Stromanschluss für die Server:

- Die Netzteile der Server müssen mindestens 80-PLUS-Gold-zertifiziert sein.

- Jeder Server muss über Netzteile mit N+N Redundanz verfügen, sodass beim Ausfall einzelner Netzteile der Volllastbetrieb jederzeit und ohne Umschaltzeit gewährleistet ist.
- Die Netzteile müssen für den Anschluss an zwei unabhängige Strompfade (A und B) ausgelegt sein.
- Jeder Strompfad muss allein den Volllastbetrieb sicherstellen, sodass das Abschalten oder der Ausfall eines Pfades jederzeit ohne Beeinträchtigung möglich ist.

6. LOS 1: Deep Learning Server

6.1. Grundlegendes

Benötigt wird ein Server, alle definierten Werte sind als Mindestanforderung zu betrachten.

- Alle Server müssen auf x86-64 Bit-Systemarchitektur basieren.
- Alle Server müssen PCI Express ab 5.0 nativ unterstützen.
- Vollständige Garantie auf alle Komponenten und Austausch

6.2. CPU

Für das System können physikalisch ein einzelner oder zwei vernetzte CPUs verbaut werden, die zusammen mindestens folgende Anforderungen erfüllen:

- Kernanzahl: mindestens 128 physische Kerne
- Cache: mindestens 256 MB L3-Cache
- Basistaktfrequenz: mindestens 2,3 GHz
- Turbo-Taktfrequenz: mindestens 3,50 GHz
- Speicherunterstützung: DDR5-Speichermodule mit einer Datenrate von mindestens 6400 MT/s und ECC-Unterstützung (Error-Correcting Code)
- Speicherkanäle: mindestens 8 unabhängige Speicherkanäle je CPU

6.3. GPU

Das System soll je 8 über NVLink vernetzte GPUs beinhalten mit folgenden (oder besseren) Spezifikationen

- Speicher: 288GB je GPU
- PCI Gen6
- NVLink: 1,8 TB/s

6.4. Arbeitsspeicher

Der Arbeitsspeicher des Servers muss mindestens die folgenden Anforderungen erfüllen:

- Gesamtkapazität: mindestens 2TB
- Fehlerkorrektur: ECC-kompatibel (Error Correcting Code)

6.5. Massenspeicher

Alle nachfolgenden Werte sind Mindestanforderungen pro Server:

- OS: mindestens 2 Stück je 1,92TB (NVMe)
- Datenspeicher: mindestens 8 Stück je 3,85TB (NVMe E1.S)

6.6. Board Management Controller (BMC)

Das System verfügt über einen dedizierten Board Management Controller

- Der BMC bietet ein dediziertes min. 1 Gbe Ethernet Interface (RJ45) für out-of-band Systemmanagement.
- Der BMC unterstützt IPMI und SNMP für Hardware Steuerung(remote Power ON, -Off, Systemreboot und BMC reset/reboot)
- Der BMC bietet ein BMC Benutzermanagement

6.7. Anschluss Technik für Daten- und Energieversorgung

Pro Server sind folgende Anschlüsse und Kabel mitzuliefern:

- 2x CAT6a-Patchkabel, Länge 10 Meter, für Management-Port.
- 4x passende 200 Gbit/s-Transceiver mit nötigen Adaptern um das angebotene System mit voller Kompatibilität an vorhandene 200GBe Nvidia/Mellanox SN3700 Switches anzuschließen.
Je 2 Transceiver für Node und Switch
- 2x OM4-Glasfaserkabel, Länge 20 Meter, zur Datenanbindung, müssen mit den mitgelieferten 200GBe Transceiver kompatibel sein.
- Je Netzteil 1x Kaltgerätekabel (IEC C14), Länge 2 Meter.

6.8. Software und Softwareserviceleistungen

Neben Zugriff auf ein fertig eingerichtet Betriebssystem inklusive Updates (z.B. DGX OS) wird der Zugriff auf die Nvidia Container Registry (NGC) für 3, optional 5 Jahre benötigt, hierbei soll dem Auftraggeber durch die Nutzung keine weiteren Kosten entstehen. Nach Ablauf der Zeit besteht die Möglichkeit einer Verlängerung der Serviceleistungen.

7. Option

Eine Erweiterung der angebotenen Garantie (nach Punkt 3.) und Softwareserviceleistung (nach 6.8) von 3 auf 5 Jahre.

Hinweis zur Option:

Die Option ist eine Zusatzkomponente, die separat angeboten und bepreist werden muss. Es besteht kein Anspruch auf Beauftragung.

Die Entscheidung über eine Beauftragung der Option erfolgt mit der Auftragserteilung.

Im Falle einer Beauftragung gelten die gleichen Liefer-, Abnahme- und Garantiebedingungen wie für das Hauptlos.

8. Zuschlagskriterium

Die Bewertung der eingegangenen Angebote erfolgt anhand des Angebotspreises.

Das bedeutet, dass unter den Angeboten, die die formalen und inhaltlichen Voraussetzungen (Ausschlusskriterien gemäß diesem Dokument und den übrigen Vergabeunterlagen) erfüllen, das Angebot mit dem niedrigsten Angebotspreis den Zuschlag erhält.

9. Lieferung

Die betriebsfertige Anlieferung der Komponenten aller Lose muss frühestens zum 01.02.2026 und zwingend bis zum 01.12.2026 in das Data Center (Untergeschoss Otto Hahn-Str 12; 44227 Dortmund) der TU Dortmund erfolgen, dies bedeutet frei Verwendungsstelle / DDP.

Ein verbindlicher Liefertermin ist im Angebot anzugeben.

Für die Anlieferung steht keine Ladebühne zur Verfügung, allerdings sind Hilfsmittel wie ein Palettenhubwagen und Lastenaufzug(vom Erdgeschoss in das Untergeschoss) verfügbar. Engstellen müssen bei Verwendung üblicher Europaletten (max. Höhe 2m) von der Anlieferstelle (Einfahrt 36) aus bis in das Datacenter mit einem manuellen

Palettenhubwagen nicht berücksichtigt werden. Eine Gewichtslimitierung von maximal 600kg/m² ist durch die Traglast des Bodens gegeben und darf nicht überschritten werden.

Die Lieferung muss in den üblichen Geschäftszeiten (Werktags von 8:00 bis 16:00) unter vorherige Rücksprache mit Mitarbeitern des Institut zur Terminfindung erfolgen (Details werden nach Zuschlag bekanntgegeben). Zugang zu den Räumlichkeiten ist nur unter Begleitung möglich.

Teillieferungen sind – nach vorheriger Absprache – vor dem vereinbarten Stichtag möglich. Eine Teilabnahme erfolgt nicht; die Abnahme und Vergütung erfolgen ausschließlich nach vollständiger Lieferung aller betriebsbereiten Komponenten.

Etwaige Verpackungen werden durch die Technische Universität dem Recycling zugeführt.

10. Rechtsfolgen bei Terminüberschreitung

Der vereinbarte Liefertermin ist ein Fixtermin.

Eine frühere Lieferung ist jederzeit möglich, bedarf jedoch der vorherigen Abstimmung mit dem Auftraggeber.

Wird der Liefertermin aus Gründen, die der Auftragnehmer zu vertreten hat, nicht eingehalten, ist der Auftraggeber berechtigt, vom Vertrag zurückzutreten und Schadensersatz zu verlangen. Hierzu zählen insbesondere auch Schäden, die durch den Verlust eingeplanter Haushaltsmittel oder Fördergelder entstehen.

11. Elektrische Sicherheit

Mit Abgabe des Angebots erklärt der Auftragnehmer, dass die gelieferten Waren den Verordnungen (speziell 1. ProdSV¹) zum Produktsicherheitsgesetz ProdSG² entsprechen und die einschlägigen elektrotechnischen Regeln eingehalten sind (DGUV³ Vorschrift 3).

12. Abnahmeverfahren

Die Abnahmeüberprüfung der Server dauert 10 Werktage (Montag bis Freitag ohne gesetzliche Feiertage in NRW), wobei der Tag der Anlieferung nicht mitgerechnet wird.

13. Dokumentation

Handbücher für Server müssen vor der Abnahme im PDF Format ausgehändigt oder zugänglich gemacht werden.

¹ Verordnung über elektrische Betriebsmittel vom 17. März 2016 (BGBI. I S. 502), die durch Artikel 20 des Gesetzes vom 27. Juli 2021 (BGBI. I S. 3146) geändert worden ist"

² Produktsicherheitsgesetz vom 27. Juli 2021 (BGBI. I S. 3146, 3147), das durch Artikel 2 des Gesetzes vom 27. Juli 2021 (BGBI. I S. 3146) geändert worden ist.

³ DGUV Vorschrift 3 Elektrische Anlagen und Betriebsmittel Unfallverhütungsvorschrift vom 1. April 1979 in der Fassung vom 1. Januar 1997 mit Durchführungsanweisungen vom April 1997 Stand: Januar 2012