HPE Private Cloud AI

Neuerungen

- HPE Private Cloud AI ist jetzt mit einer aktualisierten kleinen Konfiguration mit HPE ProLiant Compute DL380a Gen12 basierte KI Worker Nodes mit NVIDIA RTX Pro 6000 Blackwell Server Edition GPUs verfügbar.
- Wenn Datenisolierung unverzichtbar ist, ist eine Air-Gap-Version von HPE Private Cloud AI mit einer Konfiguration der Größe Medium die beste Lösung, um eine Exposition gegenüber externen Netzwerken zu verhindern.
- HPE Private Cloud AI ist jetzt mit aktualisierten Konfigurationen der Größen Medium und Large mit HPE ProLiant Compute DL380a Gen12 basierte KI Worker Nodes mit NVIDIA® H200 GPUs verfügbar.
- Durch Multi-Tenancy im Unternehmen können mehrere Teams sicher auf einer einzigen,

Übersicht

Das Potenzial der KI wird oft durch Komplexität und Risiken blockiert. HPE Private Cloud AI ist die strategische Antwort – eine komplette AI Factory, die die technischen Risiken und Leistungsengpässe herkömmlicher Lösungen beseitigt. Dies bedeutet, dass Sie eine schnellere Time-to-Value und einen vorhersehbaren ROI erzielen können.
Unsere gemeinsam mit NVIDIA® entwickelte Lösung bietet die Grundlage für zuverlässige KI-Innovationen:
• Kostentransparenz: Unser On-Premises-Modell beseitigt betriebliche Engpässe und verschafft Ihnen die finanzielle Klarheit, die Sie brauchen, um über Pilotprojekte hinauszugehen.
• Optimierte Innovation: Vorab validierte Tools und Notebooks standardisieren Workflows und Modellentwicklung und bieten gleichzeitig vom ersten Tag an konsistente Governance und Zero-Touch-Sicherheit.
• Zukunftssichere Skalierbarkeit: Erweitern Sie Ihre KI-Infrastruktur nahtlos über verschiedene Computing- und GPU-Architekturen hinweg, damit Sie ohne Risiko Innovationen vorantreiben und sich an die KI-Technologien der Zukunft anpassen können.

gemeinsam genutzten KI-Plattform zusammenarbeiten. In isolierten Umgebungen können Administratoren Ressourcen effizient einzelnen Teams zuweisen.

Funktionen

Sofort nutzbare AI Factory

HPE Private Cloud AI ist eine von Grund auf neu entwickelte AI Factory, die Komplexität reduziert und Ihren Weg zur KI beschleunigt. Wir helfen Ihnen, KI-Modelle schneller zu erstellen und bereitzustellen, sodass Sie schnell und sicher vom Pilotprojekt zur Produktion übergehen können.

Ersetzt ein mehrmonatiges Infrastrukturprojekt durch unsere vorentwickelte AI Factory und geht in Tagen – nicht Monaten – vom Konzept zur Produktion über.

Entwickelt zur Reduktion von Komplexität und Time-to-Value sowie zur Gewährleistung der für unternehmensweite Skalierung von KI-Innovationen erforderlichen Kostentransparenz.

Einer der schnellsten Wege, KI-Ambitionen in die Realität umzusetzen.

Umfassende Suite sofort einsatzbereiter KI-Tools

Ein umfassendes, direkt nutzbares KI-Ökosystem reduziert die Notwendigkeit, KI-Stacks von Grund auf neu zu erstellen.

Beschleunigt die KI-Einführung, unterstützt Kunden beim Übergang vom Pilotprojekt zur Produktion und reduziert Komplexität, Zeitaufwand und Risiken in Verbindung zu KI-Initiativen.

Eine konsistente, vorintegrierte Softwareplattform beseitigt die Reibungsverluste und Komplexität, die durch eine fragmentierte Toolchain entstehen.

Data Scientists und Entwickler können mit Single Sign-On nahtlos zwischen einer Reihe beliebter KI- und Open-Source-Tools – von JupyterLab bis Apache Spark – wechseln und dabei vom ersten Tag an eine konsistente Governance und Sicherheit gewährleisten.

Unterstützt sowohl Anfänger als auch erfahrene Data Scientists mit Funktionen wie Low-Code-Funktionen, Endpunkt-APIs, JupyterLab und vorgefertigten Jupyter-Notebooks.

Alle Daten für KI nutzen

Erhalten Sie mit einem föderierten Data Lakehouse nahtlosen Zugriff auf Daten in Ihren heterogenen Speichersystemen. Mit diesem Ansatz können Sie Ihre KI-Modelle anhand einer einzigen, einheitlichen Ansicht Ihrer Daten trainieren und optimieren, ohne dass Zeitaufwand, Kosten und Risiken durch Datenmigration entstehen.

Der globale Namespace ist der zentrale Hub, der die Probleme zeitaufwendiger Datenintegrationsprojekte löst. Ein einziger, nahtloser Access Point für Ihre Daten reduziert die Zeit erheblich, die für die Vorbereitung und Verwendung von Daten für KI benötigt wird, und beschleunigt Ihren gesamten KI-Lebenszyklus.

Beschleunigen Sie Einblicke und KI-Modell-Entwicklung und bieten Sie Ihren Teams die Möglichkeit, ihre bevorzugten Analyse-Engines – von SQL bis spezialisierten KI/ML-Tools – auf die Daten in Ihren Data Lakes anzuwenden.

Bietet konsistente, zuverlässige KI-Modelle, indem Inkonsistenzen und Versionskontrollprobleme beseitigt werden, die Projekte verlangsamen. Zentralisierte Governance und Sicherheit bedeuten, dass Sie vom ersten Tag an über eine einzige zuverlässige Quelle für alle Ihre Daten verfügen.

Flexibles Cloud-Erlebnis für kontinuierliche KI-Innovation

Wir helfen Ihnen, Ihre KI-Infrastruktur vom ersten Tag an richtig zu dimensionieren. Wählen Sie aus einer Auswahl validierter Ausgangspunkte, die für Ihre spezifischen Anwendungsfälle optimiert sind.

Beginnen Sie klein und skalieren Sie nahtlos auf bewährten Konfigurationen. So stellen Sie sicher, dass Ihre Plattform mit Ihren Anforderungen an mehr Benutzer, höhere Durchsätze und neue KI-Initiativen wächst.

Machen Sie Ihre KI mit einer dauerhaften Plattform zukunftsfähig. HPE Private Cloud AI wird regelmäßig mit den neuesten CPUs, GPUs und Hardware-Innovationen aktualisiert, während mit einem Klick verfügbare Upgrades dafür sorgen, dass Ihre Software auf dem neuesten Stand bleibt.

Technische Daten	HPE Private Cloud AI
Konfiguration 1	Small (4 x L40s GPUs) – KI-Inferenz 1 x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (4 x NVIDIA L40s GPUs pro Node) 3 x HPE ProLiant Compute DL325 Gen11 Kontroll-Nodes 109 TB HPE GreenLake for File Storage mit aktiviertem Object NVIDIA SN4600cM Switches (100 GbE Networking) 42U-Rack mit PDUs HPE AI Essentials mit NVIDIA KI Unternehmenssoftware 3- oder 5-Jahres-Abonnement
Konfiguration 2	Expanded Small (8 x L40s GPUs) – KI-Inferenz 2 x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (4 x NVIDIA L40s GPUs pro Node) 3 x HPE ProLiant Compute DL325 Gen11 Kontroll-Nodes 109 TB HPE GreenLake for File Storage mit aktiviertem Object NVIDIA SN4600cM Switches (100 GbE Networking) 42U-Rack mit PDUs HPE AI Essentials mit NVIDIA KI Unternehmenssoftware 3- oder 5-Jahres-Abonnement
Konfiguration 3	Medium (8 x L40s GPUs) – KI-Inferenz und Retrieval Augmented Generation (RAG) 2x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (4 x NVIDIA L40s GPUs pro Node) 3 x HPE ProLiant DL325 Gen11 Kontroll-Nodes 217 TB HPE GreenLake for File Storage mit aktiviertem Object NVIDIA SN4700M Switches (200 GbE Networking) 42U-Racks mit PDUs HPE AI Essentials mit NVIDIA AI Enterprise Software 3- oder 5-Jahres-Abonnement
Konfiguration 4	Expanded Medium (16 x L40S GPUs) – KI-Inferenz und Retrieval Augmented Generation (RAG) 4x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (4 x NVIDIA L40s GPUs pro Node) 3 x HPE ProLiant DL325 Gen11 Kontroll-Nodes 217 TB HPE GreenLake for File Storage mit aktiviertem Object NVIDIA SN4700M Switches (200 GbE Networking) 42U-Racks mit PDUs HPE AI Essentials mit NVIDIA AI Enterprise Software 3- oder 5-Jahres-Abonnement

Technische Daten	HPE Private Cloud AI
Konfiguration 5	Large (16 x H100NVL GPUs) – KI-Inferenz, Retrieval Augmented Generation (RAG) und Modell-Feinabstimmung 4 x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (4 x NVIDIA H100NVL GPUs pro Node) 3 x HPE ProLiant Compute DL325 Gen11 Kontroll-Nodes 670 TB HPE GreenLake for File Storage mit aktiviertem Object NVIDIA SN4700M Switches (400 GbE Networking) 2 x 42U-Racks mit PDUs HPE AI Essentials mit NVIDIA AI Enterprise Software 3- oder 5-Jahres-Abonnement
Konfiguration 6	Expanded Large (32 x H100 NVL GPUsl) – KI-Inferenz, Retrieval Augmented Generation (RAG) und Modell-Feinabstimmung 8 x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (4 x NVIDIA H100NVL GPUs pro Node) 3 x HPE ProLiant Compute DL325 Gen11 Kontroll-Nodes 670 TB HPE GreenLake for File Storage mit aktiviertem Object NVIDIA SN4700M Switches (400 GbE Networking) 2 x 42U-Racks mit PDUs HPE AI Essentials mit NVIDIA AI Enterprise Software 3- oder 5-Jahres-Abonnement
Konfiguration 7	Developer System – KI-Inferenz, Retrieval Augmented Generation (RAG), KI und Sandbox-Entwicklung 1 x HPE ProLiant Compute DL380a Gen11 KI-optimierter Node (2 x NVIDIA H100NVL GPUs insgesamt) 1 x HPE ProLiant Compute DL325 Gen11 Kontroll-Node 32 TB integrierter Datei-/Object-Storage 2x 200GbE Netzwerkanschlüsse HPE AI Essentials mit NVIDIA AI Unternehmenssoftware 3- oder 5-Jahres-Abonnement Switches, Racks und PDUs sind nicht enthalten

[1] Interne HPE Berichte. Vergleich zwischen GPT-4 über OpenAI API und selbst gehosteten Llama3, ausgehend von einem Unternehmensaccount mit 5.000 Benutzern, 5 Chat-Sitzungen pro Tag, 8.000 Tokens pro Chat

HPE Services

Egal, wo Sie sich auf Ihrem Weg zur Transformation befinden, können Sie darauf zählen, dass Ihnen HPE Services die erforderliche Fachkompetenz bietet, wann, wo und wie auch immer Sie sie benötigen. Von der Strategie und Planung über die Bereitstellung bis hin zum laufenden Betrieb und darüber hinaus können Ihnen unsere Experten dabei helfen, Ihre digitalen Ambitionen zu verwirklichen.

Advisory & Professional Services

Lassen Sie sich von Experten dabei unterstützen, Ihren Weg zur Hybrid Cloud zu planen und Ihren Betrieb zu optimieren.

Managed Services

HPE verwaltet Ihren IT-Betrieb und bietet Ihnen eine einheitliche Steuerung, damit Sie sich auf Innovationen konzentrieren können.

Support-Services

Optimieren Sie Ihre gesamte IT-Umgebung und treiben Sie Innovationen voran. Bewältigen Sie die täglichen IT-Betriebsaufgaben und setzen Sie wertvolle Zeit und Ressourcen frei.

— **HPE Complete Care Service:** ein modularer Service, mit dem wir Sie dabei unterstützen Ihre gesamte IT-Umgebung zu optimieren und vereinbarte IT-Ergebnisse und Unternehmensziele zu erreichen. Der gesamte Service wird durch speziell geschulte und zugewiesene HPE Experten bereitgestellt.

— **HPE Tech Care Service:** das operative Serviceerlebnis für HPE Produkte. Der Service bietet Ihnen Zugang zu produktspezifischen Experten, ein KI-gestütztes digitales Erlebnis und allgemeine technische Anleitungen als Hilfestellung, Risiken zu reduzieren und ständig nach Wegen zu suchen, die Dinge besser zu machen.

— **HPE Multivendor Services:** Zentrale Verantwortung für die Verwaltung des Hardware- und Software-Supports vor Ort für Produkte verschiedener Anbieter. Die HPE Experten helfen Ihnen, Ihre IT über Technologien und Plattformen von HPE und anderen Herstellern hinweg zu verwalten, und fungieren als zentrale Ansprechpartner für Ihre Anforderungen hinsichtlich des IT-Betriebs.

Lifecycle Services

Erfüllen Sie die spezifischen Projektanforderungen Ihrer IT-Bereitstellung mit maßgeschneiderten Projektmanagement- und Bereitstellungsservices.

HPE Education Services

Schulungen und Zertifizierungen für Fachkräfte im IT- und Geschäftsbereich in allen Branchen. Erstellen Sie Lernpfade, um Ihre Kenntnisse in einem bestimmten Fachgebiet zu erweitern. Planen Sie die Schulungen so, wie es für Ihr Unternehmen am besten ist, mit flexiblen Optionen für kontinuierliches Lernen.

Die **Einbehaltung defekter Datenträger** ist ein optionaler Service: Sie können Festplatten oder entsprechende SSD/Flash-Laufwerke behalten, die von HPE aufgrund einer Fehlfunktion ausgetauscht wurden.

HPE GreenLake

HPE GreenLake Edge-to-Cloud-Plattform ist das marktführende As-a-Service-Angebot von HPE, das ortsunabhängig (in Rechenzentren, Multi-Clouds und am Edge) das Beste der Cloud für Anwendungen und Daten bietet, mit einem einheitlichen Betriebsmodell, On-Premises, mit vollständiger Verwaltung und einer nutzungsabhängigen Bezahlung.

Informationen zu weiteren Services wie IT-Finanzierungslösungen finden Sie hier.

Weitere technische Informationen, verfügbare Modelle und Optionen finden Sie in den QuickSpecs

HPE.com besuchen

Jetzt chatten

Teile und Materialien: HPE stellt von HPE unterstützte Ersatzteile und Materialien bereit, die für die vertraglich abgedeckte Hardware erforderlich sind.

Teile und Komponenten, die ihre maximal unterstützte Lebensdauer und/oder die maximale Nutzungsbeschränkung gemäß der Beschreibung im Betriebshandbuch des Herstellers, in den QuickSpecs für das Produkt oder im technischen Produktdatenblatt erreicht haben, werden im Rahmen dieser Service nicht bereitgestellt, repariert oder ausgetauscht.

Bild kann vom tatsächlichen Produkt abweichen.

PSN1014847366DEDE, Januar, 2026.

HEWLETT PACKARD ENTERPRISE

hpe.com