

OFFENE, ETHERNET-BASIERTE KI-DATENCENTER-NETZWERKE – LÖSUNGSBESCHREIBUNG

Anbieterübergreifende, Absichtsbasierte End-To-End-KI-Netzwerke Für Geschwindigkeit, Effizienz Und Wirtschaftlichkeit

Herausforderung

Die Bereitstellung von KI-Clustern ist komplex, zeitaufwändig und kostspielig. Unternehmen, die in KI-Anwendungen investieren, haben begrenzte Ressourcen, um KI-Datencenter bereitzustellen und die teuren GPUs ohne Anbieterbindung zu optimieren.

Lösung

[Die Lösung von Juniper für KI-Datencenter](#) ist die schnellste und flexibelste Möglichkeit, leistungsstarke Cluster für das KI-Training, -Inferenz oder als Speicher bereitzustellen, die einfach und mit limitierten IT-Ressourcen betrieben werden können.

Vorteile

- Simplifizierter „Day 0/1/2+“-Betrieb verbraucht weniger Ressourcen und spart Zeit und Geld.
- Völlige Flexibilität beim Design von von GPU unabhängigen Netzwerken mithilfe von bewährten Technologien und Produkten ohne Anbieterbindung, die eine Leistung bieten, welche InfiniBand in nichts nachsteht.
- Einsatzbereite Lösungen mit End-to-End-Validierung, damit Sie die Produkte voller Zuversicht auswählen können und eine schnelle Bereitstellung sichergestellt ist.

Traditionelle Datencenter-Technologien und -Designs können nur schwer mit den anspruchsvollen Performance-, Kapazitäts- und Latenzanforderungen umgehen, die KI-Workloads an die Infrastruktur stellen. KI-Datencenter, die sich durch spezialisierte Fabric-Designs für Back-End-Training und Front-End-Inferenz auszeichnen, erfordern hochmoderne Lösungen für Rechenleistung, Speicher und Netzwerke. Im Zentrum dieser Entwicklungen steht ein sprunghafter Anstieg der Nachfrage nach leistungsfähigen Grafikprozessoren (GPUs) und eine Abhängigkeit von proprietärem InfiniBand für Netzwerke aus einer Hand. Dies hat leider zu einem eskalierten Preisgefälle und Engpässen in der Lieferkette geführt. Für Unternehmen kommt es nun darauf an, die Kosten zu senken und zugleich Innovationen sowie Flexibilität voranzutreiben. Da die Unternehmen die Notwendigkeit zur Optimierung der GPU-Performance erkannt haben, sehen sie im Ethernet verstärkt eine gute offene Netzwerkalternative für KI-Datencenter.

Die Herausforderung

In verschiedenen Branchen ist KI nicht mehr nur ein Thema im Techniklabor sondern auch auf Führungsebene. KI-Architekten und die jeweils entsprechenden IT-Verantwortlichen befinden sich plötzlich auf ganz neuem Gebiet. Um den vollen Wert von KI nutzbar zu machen, müssen sie zuerst die komplexen Sachverhalte der KI-Bereitstellung sowie die Fakten von KI-Kosten und Durchlaufzeiten in den Griff bekommen.

Die Kosten von ohnehin bereits teurer und seltener KI-Infrastruktur haben sich nicht zuletzt aufgrund von Verzögerungen in der Lieferkette und niedrigem Lagerbestand sowie nur wenigen praktikablen GPU-Optionen und proprietärem InfiniBand-Netzwerk weiter erhöht. Eine einzige GPU kostet beispielsweise über 30.000 US-Dollar. KI-Server, die diese Chips verwenden, können über 400.000 US-Dollar kosten. Vor diesem preislichen Hintergrund können selbst kleine GPU-Cluster bereits Millionen kosten, große GPU-Cluster können Hunderte Millionen von Dollar kosten.

Als Motoren der KI spielen Datencenter-Netzwerke eine kritische Rolle bei der Verbindung und Optimierung von GPUs. Die Reduzierung der Job-Abschlusszeit (JCT), also der Zeit, die für ein KI-Training benötigt wird, ist der Schlüssel zu mehr Geschwindigkeit und Kosteneinsparungen.

Eine erfolgreiche KI-Einführung hängt wiederum von einer schnellen Marktantwort auf die Nachfrage ab. Um die Abhängigkeit von der Lösung eines einzelnen Anbieters zu reduzieren, hat die Branche bereits die Vorkehrungen getroffen, die zu einem offenen, wettbewerbsfähigen Markt mit GPU-Vielfalt und zur am umfangreichsten bereitgestellten L2-Technologie der Welt führen sollen: dem Ethernet.

Die Lösung von Juniper Networks für KI-Datencenter-Netzwerke

Die Lösung von Juniper für KI-Datencenter ist die schnellste und flexibelste Möglichkeit, leistungsstarke Cluster für das KI-Training, -Inferenz oder als Speicher bereitzustellen, die einfach und mit limitierten IT-Ressourcen betrieben werden können. Anhand von einzigartigem absichtsbasiertem Betrieb, einem KI-nativen virtuellen

Netzwerkassistenten und validierten Designs von Juniper simplifiziert Juniper das Design, die Bereitstellung und die Fehlerbehebung von KI-Datencenter-Netzwerken. Außerdem bietet die Lösung den Kunden unerreichte Flexibilität und vermeidet durch Siliziumvielfalt, anbieterübergreifende Verwaltung sowie offene, standardbasierte Ethernet-Fabrics eine Anbieterbindung.

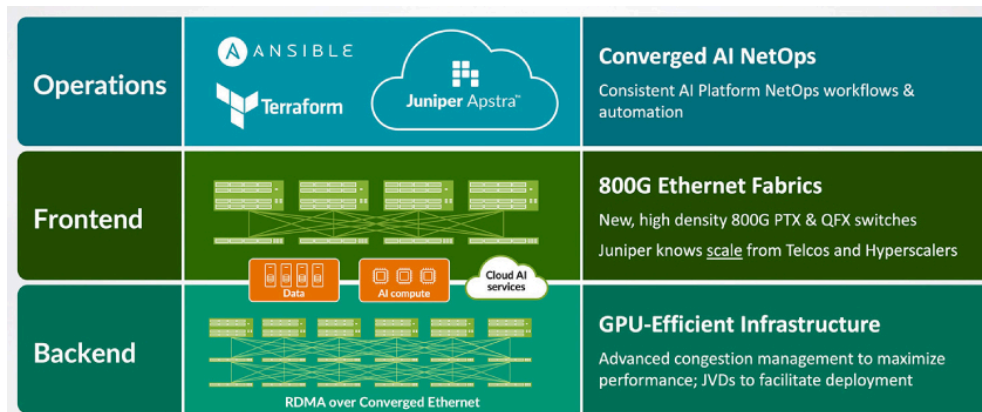


Abbildung 1: Die Lösungen von Juniper Networks für KI-Datencenter-Netzwerke

Simplifizierte Betriebsprozesse

Simplifizierter „Day 0/1/2+“-Betrieb verbraucht weniger Ressourcen und spart Zeit und Geld. Juniper Software ist die einzige anbieterübergreifende Plattform für die Automatisierung von Datencentern mit branchenführendem Intent-based Networking, das dank weniger manueller Konfigurationen für eine schnellere Bereitstellung und Fehlerbehebung sorgt. Aufgrund von Funktionen für fortschrittliches Load Balancing, explizite Überlastungsbenachrichtigung (ECN) und Priority Flow Control (PFC), die für hohe Leistung essenziell sind, können verlustfreie KI-Netzwerke mithilfe von Apstra und ohne einen einzigen CLI-Befehl über die gesamte Fabric angewendet werden. Die Echtzeitvisualisierung der Datenströme im Netzwerk hebt Überlastungs-Hotspots im KI-Datencenter mit Heatmaps hervor und absichtsbasierte Analysen ermöglichen eine schnelle Fehlerbehebung.

Mit einer Single Source of Truth (SSoT) und Closed-Loop-Assurance bietet Apstra leistungsstarke Analysen, um Probleme vorherzusagen und proaktiv zu beheben und so Unterbrechungen zu vermeiden. Apstra sorgt mit Entwurfsvorlagen für Zuverlässigkeit, Konsistenz und Reproduzierbarkeit und damit für bis zu 90 % niedrigere Betriebskosten und eine 85 % schnellere Bereitstellung.

Juniper bringt [AIOps](#) mithilfe von [Juniper Mist AI™](#) und Apstra-Integration ins Datencenter. Dies sorgt für verbesserte Automatisierung und Einblicke sowie optimierte Datencenter-Erfahrungen. [Marvis™](#), der virtuelle Netzwerkassistent für Datencenter, weitet die umfangreiche Streaming-Telemetrie und

Echtzeitüberwachung in Apstra auf das Marvis-Dashboard aus, um proaktive Fehlerbehebung und Analysen über betriebliche Domänen zu gewährleisten. Für die Bereitstellung von zukünftigen KI-gestützten, verwertbaren Einblicken für anbieterübergreifende Datencenter ist diese Integration ein grundlegender erster Schritt. In Campus- und Zweigstellenprozessen, die Marvis einsetzen, können Probleme mit dem Datencenter auf demselben Dashboard angezeigt werden.

Völlige Flexibilität

Offene, flexible Ethernet-Lösungen ermöglichen den Einsatz von bewährten Technologien und Produkten ohne Anbieterbindung. Die neuesten KI-optimierten High-Radix-800-GbE-Leaf-and-Spine-Fabrics für Datencenter von Juniper bieten fortschrittliche Funktionen zur Verwaltung von Datenverkehr, die eine große Bandbreite, geringe Latenz sowie verlustfreie und skalierbare Leistung über das Ethernet garantieren. Ethernet verfügt über eine Runway von 1,6 Terabit und anbieterübergreifenden Support für Systeme, die von GPUs unabhängig sind, und senkt so die Kosten, ermöglicht schnellere Innovationen, maximiert die Flexibilität beim Design und vermeidet Probleme innerhalb der Lieferkette. Die Unterstützung für Remote Direct Memory Access over Converged Ethernet v2 (RoCEv2) bedeutet, dass Ethernet-Fabrics von Juniper die einzige anbieterunabhängige Lösung für die Verwaltung von DC-Fabrics sind, die zudem für die Automatisierung von Back-End-KI-Trainings- und Front-End-Inferenzmodellen optimiert sind.

Einsatzbereite Lösungen

Einsatzbereite Lösungen mit End-to-End-Validierung, damit Sie die Produkte voller Zuversicht auswählen können und eine schnelle Bereitstellung sichergestellt ist. Wenn Unternehmen End-to-End-Netzwerklösungen nutzen können, sind sie einfach und flexibel imstande, leistungsfähige KI-Datencenter zu erstellen. [Die Router der PTX-Serie für Datencenter von Juniper Networks®](#) wurden mit dem neuen 800GbE [PTX10002-36QDD Switch](#) mit fester Konfiguration und hoher Dichte sowie neuen 800GbE Linecards für das [PTX10000 Gehäuse](#) erweitert. Das größte PTX10000 Gehäuse wird von speziellem Express-5-Silizium angetrieben und unterstützt nun bis zu 576 x 800-GbE-Ports für eine High-Radix-Spine- und Super-Spine-Architektur. Außerdem bietet eine neue, auf modernstem Broadcom Tomahawk 5 ASIC basierende 800GbE [QFX5240 Plattform](#) mit fester Konfiguration und hoher Dichte die nötige Siliziumvielfalt für leistungsstarke, energieeffiziente und skalierbare KI-Datencenter.

Die Plattformen der Juniper Serien QFX und PTX bieten außergewöhnliche Leistung, Skalierbarkeit und Überlastungskontrolle, wodurch sie auch für KI-Cluster Spitzenleistung erzielen. In Benchmark-Tests der ML-Performance zeigte die für KI optimierte Ethernet-Lösung von Juniper eine Job-Abschlusszeit (JCT) bei Trainings-Workloads, die der von InfiniBand ebenbürtig war. Kunden haben so eine offene, leistungsfähige Fabric-Alternative ohne Anbieterbindung. Um die KI-Bereitstellung zu simplifizieren, wird Juniper neue validierte Designs von Juniper (JVDs) veröffentlichen. KI-JVDs ermöglichen vorgeschriebene, optimierte oder vielschichtige Clos-Fabrics, um eine vollständige Datencenter-Lösung zu gewährleisten, einschließlich Switching, Prozessen und Sicherheit. JVDs werden von Spezialisten der Juniper Labs streng getestet, validiert und dokumentiert und können als gebrauchsfertige Designs oder Richtlinien verwendet werden, um Risiken zu verringern und KI-Cluster richtig zu dimensionieren und zu budgetieren.

Funktionen und Vorteile

Simplifizierte Betriebsprozesse: Der simplifizierte „Day 0/1/2+“-Betrieb verbraucht weniger Ressourcen und spart Zeit und Geld. Das branchenführende Intent-based Networking beschleunigt die Bereitstellungs- und Fehlerbehebungszeit, indem Probleme der manuellen Konfiguration umgangen werden (und das hardwareunabhängig).

- Mit SsOT und Closed-Loop-Assurance lassen sich Probleme vorhersagen und proaktiv beheben, wodurch Unterbrechungen vermieden werden.

- Nur AIOps für Datencenter sorgt für die zusätzliche Automatisierung und Informationen, mit denen sich Datencenter-Erfahrungen optimieren lassen.
- Kosteneffiziente Lösung mit 90 % niedrigeren Betriebskosten und 85 % schnellerer Bereitstellung

Völlige Flexibilität: Entwerfen Sie Ihr Netzwerk mithilfe von bewährten Technologien und Produkten ohne Anbieterbindung.

- Funktionen für erweitertes Datenverkehrsmanagement sorgen für eine verlustfreie und skalierbare Performance über das Ethernet, die zudem über eine hohe Bandbreite und eine geringe Latenz verfügt.
- Die Nutzung des Ethernets bringt niedrigere Kosten und maximierte Design-Flexibilität mit sich. Zudem lassen sich Innovationen schneller vorantreiben und Probleme innerhalb der Lieferkette werden vermieden.
- Die einzige anbieterunabhängige Lösung für DC-Fabric-Management und Automatisierung, die völlige Design-Flexibilität bietet.
- Kompatibilität mit allen GPUs, Fabrics und Switches

Einsatzbereite Lösungen: Lösungen mit End-to-End-Validierung, damit Sie die Produkte voller Zuversicht auswählen können und eine schnelle Bereitstellung sichergestellt ist.

- Siliziumvielfalt verbessert die Skalierung, Performance und Flexibilität. Unterstützung für Broadcom und benutzerdefinierte Verbindungen für Tausende von GPUs zu höchsten Geschwindigkeiten (800 Gbps) sowie alle Konfigurationen von Datencenter-Fabrics
- Umfassendes Angebot an Datencenter-Sicherheitslösungen für flächendeckende Sicherheit
- Validierte Designs von Juniper gewährleisten eine vollständige Lösung für Datencenter, vom Switching bis hin zu Betrieb und Sicherheit.
- Junipers umfangreiches Switches-Portfolio (mit Siliziumvielfalt) verbessert die Skalierung und Performance
- Umfangreiches Portfolio zur Datencenter-Sicherheit mit einheitlicher Bedrohungsabwehr für Zero-Trust-Datencenter
- Verbessert die Zuverlässigkeit des Datencenter-Netzwerks um das Zehnfache

Lösungskomponenten

Apstra: Die Datencenter-Automatisierungslösung von Juniper ermöglicht es Betreibern, den gesamten Lebenszyklus des Datencenter-Netzwerks in anbieterübergreifenden Umgebungen vom Design bis zum Betrieb zu automatisieren. Apstra bietet eine kontinuierliche Validierung, eine Single Source of Truth (SSOT),

leistungsstarke Analysen und schnelle Ursachenermittlung sowie Problembeseitigung.

Switches der QFX-Serie: Branchenführende Netzwerk-Switches bieten überragenden Durchsatz und Skalierbarkeit, einen umfangreichen Routing-Stack, die offene Programmierbarkeit von Junos® OS und die umfassendste Palette an [EVPN-VXLAN-](#) und IP-Fabric-Funktionen.

Die Router der PTX-Serie: Diese Router bilden die Grundlage für die weltweit größten Core- und WAN-Architekturen. Die neueste Generation der benutzerdefinierten ASICs der Express-Familie ist für die Cloud optimiert und ermöglicht reibungslose 800-GbE-Migration. Sie bieten gleichzeitig die Skalierbarkeit, Flexibilität und den Investitionsschutz, um den heutigen Marktanforderungen gerecht zu werden.

Virtueller Netzwerkassistent Marvis für Datencenter: Marvis ist der erste und einzige KI-native virtuelle Netzwerkassistent für Datencenter. Er bietet die besten Einblicke in den gesamten Lebenszyklus des Datencenters – und das völlig anbieterübergreifend. Zusätzlich liefert Marvis die umfassende Visibilität und Zuverlässigkeit in allen Unternehmensdomänen von Campus und Zweigstelle bis zum Datencenter.

Junos OS weiterentwickelt: Dieses einheitliche End-to-End-Netzwerkbetriebssystem bietet Zuverlässigkeit, Agilität und offene Programmierbarkeit und ermöglicht erfolgreiche Implementierungen im Cloud-Maßstab.

Zusammenfassung: Simplifizierter Betrieb, völlige Flexibilität und einsatzbereite Netzwerke für das KI-Datencenter

Jede Firma hat eine eigene Strategie, wie sie die Datencenter-Infrastruktur am besten nutzen kann, um unternehmensweite KI-Initiativen auszuführen. Mit der KI-Lösung für Datencenter von Juniper sorgen Sie schnell und einfach für leistungsstarke, skalierbare Netzwerke, die fit für KI-Training und -Inferenz sind. Anhand von anbieterübergreifendem, absichtsbasiertem Betrieb, einem KI-nativen virtuellen Netzwerkassistenten und validierten Designs von Juniper gehört die Komplexität von KI-Datencenter-Netzwerken dank Juniper der Vergangenheit an.

Corporate and Sales Headquarters

Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089 USA **Telefon: +1-888-**

JUNIPER (+1-888-586-4737)

oder +1-408-745-2000

www.juniper.net

APAC and EMEA Headquarters

Juniper Networks International B.V.
Boeing Avenue 240
1119 PZ Schiphol-Rijk

Amsterdam, Niederlande **Telefon:**

+31-0-207-125-700