

Kooperation von Arista mit NVIDIA für ganzheitliche KI-Lösungen

Optimale Performance durch koordinierte GPUs und Netzwerke

Santa Clara (USA), 29.05.2024 – Arista Networks (NYSE: ANET) präsentiert in Zusammenarbeit mit [NVIDIA](#) eine technische Demonstration, wie KI-Rechenzentren die Compute- und Netzwerkdomeänen zu einer einzigen gemanagten KI-Einheit zusammenführen können. Um optimierte Netzwerke für generative KI mit geringen Bearbeitungszeiten aufzubauen, können Unternehmen KI-Cluster einheitlich über wichtige Komponenten wie Netzwerke, NICs und Server konfigurieren, verwalten und überwachen. Dies ist der erste Schritt auf dem Weg zu einem herstellerübergreifenden, interoperablen Ökosystem, das die Steuerung und Koordination zwischen KI-Netzwerken und KI-Recheninfrastrukturen ermöglicht.

Bedarf an einheitlichen Kontrollinstrumenten

Mit der Vergrößerung von KI-Clustern und Large-Language-Modellen (LLMs) nehmen auch die Komplexität und die schiere Menge der verschiedenen Komponenten des Systems rapide zu. GPUs, NICs, Switches, optische Komponenten und Kabel müssen alle zusammenarbeiten, um ein ganzheitliches Netzwerk zu bilden. Kunden benötigen aber eine einheitliche Steuerung zwischen ihren KI-Servern, die NICs und GPUs hosten und den KI-Netzwerk-Switches auf den verschiedenen Ebenen.

All diese Komponenten sind für die einwandfreie Bearbeitung von KI-Jobs voneinander abhängig, arbeiten jedoch unabhängig voneinander. Dies kann zu Fehlkonfigurationen oder einer falschen Abstimmung zwischen den einzelnen Modulen des gesamten Ökosystems führen, z. B. zwischen den NICs und dem Switch-Netzwerk, was sich erheblich auf die Dauer der Bearbeitung von KI-Jobs auswirken kann, da Netzwerkfehler nur sehr schwer zu diagnostizieren sind. Große KI-Cluster erfordern außerdem ein koordiniertes Lastmanagement, um Datenpaketverluste oder eine zu geringe Auslastung der GPUs zu vermeiden. Zudem benötigen sie ein koordiniertes Management und Monitoring, um Rechen- und Netzwerkressourcen im Einklang zu optimieren.

Der neue Arista AI Agent

Das Herzstück dieser Lösung ist ein Agent auf Basis von [Arista EOS](#), der es dem Netzwerk und dem Host ermöglicht, miteinander zu kommunizieren und Konfigurationen zu koordinieren, um so die KI-Cluster zu optimieren. Mithilfe eines Remote-AI-Agenten kann EOS, das auf Arista-Switches läuft, auch auf die direkt angeschlossenen NICs und Server ausgeweitet werden, um als ganzheitliche Lösung einen einzigen Kontroll- und Übersichtspunkt in einem KI-Rechenzentrum zu ermöglichen.

Dieser Remote-KI-Agent, der direkt auf einer [NVIDIA BlueField-3 SuperNIC](#) gehostet wird oder auf dem Server läuft und Telemetriedaten von der SuperNIC sammelt, ermöglicht es EOS auf dem Netzwerk-Switch selbst, Netzwerkprobleme auf dem Server zu konfigurieren, zu überwachen und zu debuggen – und so eine durchgängige Netzwerkkonfiguration und QoS-Konsistenz sicherzustellen. AI-Cluster lassen sich damit als eine einzige homogene Lösung steuern und optimieren.

„Arista verfolgt das Ziel, die Effizienz der Kommunikation zwischen dem eingesetzten Netzwerk und der GPU-Topologie zu verbessern, um die Job- Bearbeitungszeiten durch koordinierte Orchestrierung, Konfiguration, Validierung und Überwachung von NVIDIA Accelerated Compute, NVIDIA SuperNICs und der Arista-Netzwerkinfrastruktur zu verbessern“, sagt John McCool, Chief Platform Officer bei Arista Networks.

End-to-End-KI-Kommunikation und -Optimierung

Die Demonstration der neuen Technologie zeigt, wie ein Arista EOS-basierter Remote-KI-Agent es ermöglicht, einen kombinierten, interdependenten KI-Cluster als eine einzige Lösung zu verwalten. EOS, das im Netzwerk läuft, kann mittels Remote-KI-Agenten auf Server oder SuperNICs ausgeweitet werden, um ein sofortiges Tracking und Reporting von Performance-Verschlechterungen oder Ausfällen zwischen Hosts und Netzwerken zu ermöglichen, so dass diese schnell isoliert und die Auswirkungen minimiert werden können. Da EOS-basierte Netzwerk-Switches ständig die genaue Netzwerktopologie kennen, ermöglicht die Ausweitung von EOS auf SuperNICs und Server mit dem Remote-KI-Agenten eine koordinierte Optimierung der Ende-zu-Ende-QoS zwischen allen Elementen im KI-Rechenzentrum, um so die Zeit für die Job-Bearbeitung zu reduzieren.

„Die branchenführenden Netzwerkplattformen von Arista mit den Compute-Plattformen und SuperNICs von NVIDIA ermöglichen koordinierte KI-Rechenzentren. Die neue Möglichkeit, das EOS-Betriebssystem von Arista mit Remote-KI-Agenten auf Hosts zu erweitern, verspricht, ein kritisches Problem von großen KI-Clustern zu lösen, indem ein einziger Punkt zur Steuerung und Kontrolle bereitgestellt wird, um KI-Verfügbarkeit und -Leistung als ganzheitliche Lösung zu managen“, sagt Zeus Kerravala, Principal Analyst bei ZK Research.

Arista wird die KI-Agententechnologie auf der Feier zum 10-jährigen Bestehen des Arista-Börsengangs am 5. Juni in NYSE demonstrieren, Kundentests werden für das zweite Halbjahr 2024 erwartet.

Analysten und andere Mitglieder der Finanzbranche, die an einer Teilnahme an der NYSE-Veranstaltung interessiert sind, können sich hier anmelden:

<https://events.arista.com/aristas-10-year-ipo-anniversary-analyst-and-investor-luncheon>

Weitere Informationen über die neuen KI-Zentren finden Sie im Blog des CEO und Chairman Jayshree Ullal hier:

<https://blogs.arista.com/blog/ai-center>

Über Arista

Arista Networks ist ein branchenführender Anbieter von datengesteuerten Client-to-Cloud-Netzwerken für große Rechenzentrums-, Campus- und Routing-Umgebungen. Die mehrfach ausgezeichneten Plattformen von Arista bieten Verfügbarkeit, Agilität, Automatisierung, Analyse und Sicherheit durch CloudVision® und das fortschrittliche Netzwerkbetriebssystem Arista EOS®.

Weitere Informationen finden Sie unter: www.arista.com

ARISTA, EOS, CloudVision, NetDL und AVA gehören zu den eingetragenen und nicht eingetragenen Marken von Arista Networks, Inc. in allen Ländern der Welt. Andere Firmennamen oder Produktnamen können Marken ihrer jeweiligen Eigentümer sein. Weitere Informationen und Materialien finden Sie unter www.arista.com.

Diese Pressemitteilung enthält zukunftsgerichtete Aussagen, einschließlich, aber nicht beschränkt auf, Aussagen zu Kosteneinsparungen, Leistung, Funktionen und Sicherheit. Alle Aussagen, die sich nicht auf historische Fakten beziehen, sind Aussagen, die als zukunftsgerichtete Aussagen betrachtet werden können. Zukunftsgerichtete Aussagen unterliegen Risiken und Ungewissheiten, die dazu führen können, dass die tatsächliche Leistung oder die Ergebnisse wesentlich von den in den zukunftsgerichteten Aussagen zum Ausdruck gebrachten abweichen, einschließlich des raschen Technologie- und Marktwandels, der Kundenanforderungen und der Industriestandards sowie anderer Risiken, die in unseren bei der SEC eingereichten Unterlagen aufgeführt sind, die auf der Website von Arista unter www.arista.com und auf der Website der SEC unter www.sec.gov verfügbar sind. Arista lehnt jede Verpflichtung ab, zukunftsgerichtete Aussagen öffentlich zu aktualisieren oder zu revidieren, um Ereignisse oder Umstände widerzuspiegeln, die nach dem Datum, an dem sie gemacht wurden, eintreten.

Pressekontakt

Amanda Jaramillo
Corporate Communications
Tel: (408) 547-5798
amanda@arista.com

Investorenkontakt

Liz Stine
Investor Relations
Tel: (408) 547-5885
liz@arista.com