



Flüssigkeitskühlung: Wie Know-how energieeffizientes Computing für KI und mehr ermöglicht

AUTOR

Steven Dickens

Chief Technology Advisor | The Futurum Group

Ron Westfall

Research Director | The Futurum Group

AUGUST 2024

IN PARTNERSCHAFT MIT





Zusammenfassung

Künstliche Intelligenz (KI) und ganz besonders Anwendungen mit generativer KI (GenAI) erobern die Welt. KI ermöglicht Durchbrüche in allen Bereichen, vom Einzelhandel bis zur Fertigung, von Finanzdienstleistungen bis zum Gesundheitswesen, von Grundlagenforschung bis zur Cybersicherheit.

Diese Entwicklungen werden durch Computertechnologie ermöglicht, die noch vor wenigen Jahren unsere Vorstellungskraft übertraf. Jedoch ist mit dem Fortschritt dieser Technologie der Stromverbrauch von praktisch allen Komponenten eines Servers – CPUs, Arbeitsspeicher, Beschleuniger, Networking usw. – in den letzten zehn Jahren durchschnittlich um 200% gestiegen. Diese energieintensiven Systeme, die für KI-Workloads erforderlich sind, füllen Rechenzentren auf der ganzen Welt und tragen wesentlich zum Anstieg des Stromverbrauchs im DataCenter-Bereich bei. Allein in den USA wird die Energienachfrage bis zum Jahr 2030 vor allem durch KI-Rechenzentren von [5.1% auf 9.1% des gesamten Energiebedarfs der USA ansteigen](#). Heute beträgt der Energieverbrauch der USA [nur 4 % laut dem Electric Power Research Institute](#).

Der Joule-Effekt besagt im Wesentlichen, dass elektrische Energie in Wärme umgewandelt wird, wenn elektrischer Strom durch einen Leiter mit Widerstand fließt. Dies bedeutet also, dass bei einem höherem Stromverbrauch mehr Wärme erzeugt wird. Der Umgang mit der Wärme, die durch den erhöhten Strombedarf von KI und High-Performance Computing (HPC) entsteht, ist für viele Unternehmen zur Herausforderung geworden. Große Leistungssteigerungen haben nachgelagerte Effekte. Komponenten mit höherer Leistung wie CPUs, GPUs, Speicher usw. erzeugen mehr lokalisierte Wärme, und Wärme ist der Feind von Silizium. Nicht abgeführte Wärme begrenzt die Leistung von CPUs und GPUs und kann das Silizium selbst beschädigen und/oder schwere Probleme wie Überhitzung und Feuer verursachen.

Vor diesem Hintergrund arbeiten Siliziumanbieter nicht nur an der Effizienzverbesserung einzelner Komponenten, sondern auch die Flüssigkeitskühlung rückt immer mehr in den Fokus, da Flüssigkeiten Wärme schneller und mit höheren Kapazitäten aufnehmen können als herkömmliche Luftkühlungsmethoden. Flüssigkeitskühlung kann auch dazu beitragen, die steigende Nachfrage nach einer verbesserten Energieeffizienz im Rechenzentrum für wachsende KI-Arbeitsumgebungen zu erfüllen. Bspw. durch nachgelagerte Effekte, wie den Wegfall des Stromverbrauchs für Klimaanlage, oder den Wegfall des Stromverbrauchs von Ventilatoren zur Luftbewegung. Dies beinhaltet die Verwaltung und Optimierung des Energieverbrauchs durch den Einsatz von Flüssigkeitskühlungstechnologien mit höheren Eingangstemperaturen (kein Kühlen von Wasser notwendig). Gleichzeitig muss sichergestellt werden, dass die Systemleistung und die Kerndichte pro Rechenzentrum nicht beeinträchtigt wird.

Der Wechsel zur Flüssigkeitskühlung kann auch dazu beitragen, dass Unternehmen auf ESG-Ziele (Umwelt-, Sozial- und Governance) hinarbeiten können, die sich auf Kohlenstoffemissionen und nachhaltige Praktiken beziehen. Wenn ein Unternehmen oder ein Anbieter von Rechenzentrums-Colocation-Lösungen nicht modernisiert, sondern den Status quo der Luftkühlung beibehält, wird der Stromanstieg von KI-Siliziumstrom durch den Bedarf an kälterer Luft, die sich schneller bewegen muss, noch verschärft. Für die Luftkühlung wird mehr Strom benötigt als für den IT-Verbrauch, und das wird ein

Unternehmen indirekt daran hindern, seinen CO₂-Fußabdruck zu reduzieren. Alles in allem stellen die Umweltauswirkungen von KI-Infrastruktur, insbesondere aufgrund des hohen Energiebedarfs, Unternehmen vor Herausforderungen, die technologischen Fortschritte mit ihren ESG-Verpflichtungen in Einklang bringen möchten. Flüssigkeitskühlung kann zur Behebung dieser Probleme beitragen.

Wir bewerten, warum Lenovo Neptune® Flüssigkeitskühlungslösungen der 6. Generation zur KI-Workload-Optimierung überlegen sind, ohne die unternehmensweiten Nachhaltigkeitsziele zu beeinträchtigen. Dieser Bericht konzentriert sich auf die langjährigen Investitionen, die Lenovo in Technik und Innovationen in der Flüssigkeitskühlung tätigt. Wir untersuchen, warum die System- und Lösungsentwicklung für Lenovo Neptune® Flüssigkeitskühlung eine höhere Qualität und Metriken aufweist, die sich im Vergleich zu Industriestandards bewähren.

Lenovo kann auf eine lange Geschichte mit Innovationen für Flüssigkeitskühlung für x86-basierte Systeme zurückblicken, die bereits vor der Übernahme von Systems X Business durch IBM im Jahr 2014 begannen. Lenovos 12-jährige Erfahrung bei der Bereitstellung von Flüssigkeitskühlung in x86-basierten Systemen (sowohl mit als auch ohne GPUs) spiegelt sich im Ansatz des Unternehmens wider - von Designdetails, Herstellung und Lieferungsprozess bis hin zu der Gesamtqualität und den Supportprozessen.

Die technischen Fortschritte des Unternehmens bieten letztendlich einen Wettbewerbsvorteil mit höherer Qualität und Durchbrüchen bei der Wärmeableitung. Lenovos umfassende Erfahrung und das Verständnis von leitfähigen Materialien und dem Verhalten von Flüssigkeiten in diesem Bereich sowie sein einzigartiger Ansatz bei der Entwicklung flüssigkeitsgekühlter Lösungen bieten eine bahnbrechende Systemqualität verbunden mit einem Wettbewerbsvorteil bei der Bereitstellung von energieeffizientem Computing.



Abschnitt 1: Zukünftige Entwicklung der Flüssigkeitskühlung

Es ist nicht übertrieben zu sagen, dass KI eine Vielzahl von Branchen revolutioniert. KI hilft Einzelhändlern, das Kundenerlebnis zu verbessern, Bestandsverluste zu reduzieren und den Gewinn zu steigern. Sie hilft Forschern, neue Arzneimittel zu entwickeln, die Ursprünge des Universums zu erkunden und den Klimawandel einzudämmen. KI hilft Herstellern, die Qualität zu verbessern, die Prozessüberwachung zu optimieren und vorausschauende Wartung bereitzustellen. So können Mediziner die Gemeinkosten reduzieren, Diagnosen verbessern und mehr Zeit mit ihren Patienten verbringen. Und im Bereich der Finanzdienstleistungen verbessert KI die Betrugserkennung und ermöglicht schnellere, genauere Investitionsentscheidungen.

Diese Durchbrüche werden durch Computertechnologie ermöglicht, die vor wenigen Jahren unsere Vorstellungskraft übertraf. Mit dem Fortschritt dieser Technologie ist der Stromverbrauch jedoch von praktisch allen Komponenten eines Servers – CPUs, Arbeitsspeicher, Beschleuniger, Networking usw. – in den letzten zehn Jahren um ca. 200 % gestiegen. Die Rechenleistung wurde auf bis zu 500 W für CPUs und über 1000 W für GPUs erhöht, mit entsprechenden Leistungssteigerungen für andere Serverkomponenten. Der Gesamtstrombedarf für ein einzelnes Rack ist von etwa 15 kW auf über 100 kW gestiegen.

Der Umgang mit der Wärme, die durch all diese elektrische Energie entsteht, ist für viele Kunden zur Herausforderung geworden. Es sind definitiv alternative Kühlmethoden erforderlich, da Kunden versuchen, ihren Wunsch nach neuen KI-Anwendungen, höherer IT-Leistung und den wirtschaftlichen und ökologischen Gegebenheiten der Server-, Rack- und Rechenzentrums Kühlung auszugleichen. Deshalb hat sich Flüssigkeitskühlung als ein wichtiger Faktor für die neue Generation von KI- und GenAI-Anwendungen herausgestellt.

Werfen wir einen genaueren Blick auf die Beweggründe für den Einsatz von Flüssigkeitskühlung. Nicht nur die einzelnen Serverkomponenten, die wir erwähnt haben, verbrauchen zunehmend Strom, sondern auch KI- und HPC-Workloads ändern den Strombedarf auf Rechenzentrumsebene. KI-Training erfolgt synchron, das heißt alle Elemente im Cluster arbeiten mit immenser Intensität gemeinsam an einem einzigen Dienst. Die Größe eines KI-Trainingsclusters kann enorm sein; er kann mindestens 5.000 Server erfordern, wobei jeder Server mehrere CPUs und GPUs verwendet. Und mit einem einzigen Server, der ~10 kW Leistung benötigt, ist es leicht zu erkennen, wie umfangreich das Problem ist. KI-Workloads können bis zu dreimal so viel Leistung wie typische Cloud-Workloads verbrauchen und entsprechend mehr Wärme erzeugen.

Wir haben bereits aufgezeigt, wie eine Erhöhung der Geräteleistung das gesamte Rechenzentrum beeinflusst. Schauen wir uns nun noch weitere Zusammenhänge an, die für diese Systeme der nächsten Generation bestehen, und die diese fortschrittlichen Computing-Workloads unterstützen.

Advanced Computing der nächsten Generation stellt das Systemdesign vor Herausforderungen in drei Schlüsselbereichen: thermisch, mechanisch und elektrisch:

- **Thermisch:** Heißere Komponenten (z. B. Spannungsregler, CPUs, GPUs, Speicher, Netzwerke usw.), niedrigere Silizium-Schwellenwerte, die von Generation zu Generation notwendig sind, geografische Herausforderungen (heiße und trockene Rechenzentrumsstandorte)
- **Mechanisch:** Dramatische Zunahme der Sockelgrößen, um höhere Transistoranzahlen, mehr duale In-Line-Speichermodule (DIMM) pro CPU oder GPU, mehr Spannungsregler (VR), größere Netzwerkkarten sowie mehr Board-Layer und eine größere Siliziumhöhe unterzubringen
- **Elektrisch:** Hohe Nachfrage nach $\geq 600\text{-W-CPU}$ s, $\geq 1200\text{-W-GPU}$ s und Übergang zu 48-V-Eingang. Ein KI-System kann über 10 kW hinausgehen.

Als Nächstes wollen wir uns ein wenig genauer damit beschäftigen, wie sich die Flüssigkeitskühlung von der Luftkühlung unterscheidet und warum es für KI wichtig ist. Flüssigkeit hat eine höhere Wärmekapazität als Luft. Thermische Kapazität beschreibt, wie viel Wärme ein Medium aufnehmen kann, bevor sich seine Temperatur ändert. Die Wärmekapazität von Flüssigkeit ist viel größer als die von Luft, dies bedeutet, dass Flüssigkeit eine deutlich höhere Wärmemenge aufnehmen kann, bevor die Flüssigkeit ihre Temperatur erhöht, so dass der Wärmeableitungsprozess deutlich effizienter ist.

Außerdem gibt es ein Problem mit dem Fußabdruck im Rechenzentrum. Bei der Luftkühlung ist die Kühlkörperleistung an die Gesamtoberfläche gebunden. Je größer die Oberfläche, desto besser ist die Kühlkörperleistung. Eine Vergrößerung der Oberfläche des Kühlkörpers kann durch eine Erweiterung dieser horizontal oder vertikal oder beides erreicht werden. Letztlich führt diese Vergrößerung der Oberfläche zu größeren Systemen in der Höhe und verringert so die Rechenleistung, die in ein Rechenzentrum gelangen kann. Das bedeutet weniger Computing für eine höhere Leistung!

Flüssigkeitskühlung kann jedoch mehr Rechenleistung in traditionellen Serverformaten beibehalten, weil sie Wärme so effizient absorbiert und der Effekt der Oberfläche im Verhältnis zur Leistung drastisch verringert wird, was eine höhere Rechendichte auf herkömmlicher Fläche ermöglicht. Flüssigkeitsgekühlte Server benötigen weniger Platz, erzeugen weniger Lärm (geringere Lüfterleistung durch geringeren Bedarf an einem hohen Luftstrom) und ermöglichen eine insgesamt kühlere Rechenzentrums Umgebung.

All diese Faktoren haben zu intensiven Investitionen in die Flüssigkeitskühlung in der Rechenzentrumsindustrie geführt. Die allgemeine Auffassung ist, dass alle Flüssigkeitskühlungslösungen gleich sind und es nur wenige Unterschiede zwischen den Anbietern gibt. Wir haben jedoch festgestellt, dass dies überhaupt nicht der Fall ist. Lenovos 12-jährige Erfahrung in der Entwicklung von Flüssigkeitskühlungstechnologien hat es dem Unternehmen ermöglicht, Lösungen zu entwickeln, die allgemeine Standards nicht nur erfüllen, sondern deutlich übertreffen und nach unserer Analyse deutlich besser sind in puncto Design, Fertigung, Lieferung und Services als die der Konkurrenten.



Abschnitt 2. Warum Lenovo Neptune® die bewährte Lösung für KI-Workloads ist: Überlegene Flüssigkeitskühlungstechnologie

Wir untersuchen, warum Lenovos Neptune® Flüssigkeitskühlungslösung der 6. Generation eine robuste Lösung bietet, die heute verfügbar ist, um unterschiedliche KI-Workload-Anforderungen zu erfüllen. Die Lenovo Lösung bietet eine deutlich höhere Qualität und ist in wichtigen Bereichen, wie der Wärmeabfuhr, effizienter als eine Luftkühlung - und das bei gewährleisteter Spitzenleistung und erheblich gesteigeter Energieeffizienz. Im Wesentlichen bietet die Lenovo Neptune®-Lösung einen ganzheitlichen Kühlungsansatz, den wir als marktführend im gesamten Rechenzentrumsbereich identifizieren.

Erstens sticht das Lenovo Neptune® Technologieportfolio heraus. Lenovo Neptune® basiert auf mehr als einem Jahrzehnt Erfahrung mit Flüssigkeitskühlung mit mehr als 40 Patenten, die eine wesentliche Rolle bei der Implementierung von Supercomputing und KI-Clustern spielen, sodass Unternehmen leistungsstarke KI in jeder Größenordnung bereitstellen können. Lenovo war der erste Anbieter, der im Juni 2012 im Leibniz Supercomputing Centre (LRZ) einen flüssigkeitsgekühlten Petascale-Supercomputer mit hoher Eingangstemperatur lieferte, der auf Platz 4 der Liste der Top500 Supercomputer (www.top500.org) debütierte, die die schnellsten 500 Systeme der Welt misst.

Diese langjährige Erfahrung im Bereich Flüssigkeitskühlung hat es Lenovo ermöglicht, seine Technologie und seine Fertigungskompetenz zu optimieren. Einige der herausragenden Merkmale der Lenovo Technologie, die wir beobachtet haben, sind:

- **Schläuche (intern und extern):** Lenovo verwendet hochwertige Ethylen-Propylen-Dien-Monomer-Schläuche (EPDM). EPDM ist für seine hohe PSI-Festigkeit und seinen niedrigen Durometerbereich bekannt, was auf eine größere Flexibilität hinweist als die Kunststoffschläuche, die von einigen anderen auf dem Markt verwendet werden. Diese Schläuche werden zudem intern mit Peroxid behandelt, wodurch Korrosion und Schäden am Schlauch verhindert werden.
- **Hauptkühlungskreislauf:** Die internen Hauptschleifen und Kühlplatten von Lenovo verwenden Kupfer anstelle von fluoriertem Ethylenpropylen (FEP), da Kupfer seit langem als ausgezeichneter elektrischer Leiter wegen seiner Festigkeit, Haltbarkeit und effizienterer Wärmeableitung bekannt ist.
- **Optimierte, patentierte Kühlplatten:** Lenovos Kühlplattendesign mit geringem Druckabfall maximiert die Wärmeextraktion für Beschleuniger mit einem Verbrauch von ~700 W jetzt und über 1000W in Zukunft.
- **Wasser:** Lenovo Neptune® verwendet Wasser anstelle von anderen Flüssigkeiten wie Polyethylenglykol, da Wasser eine der höchsten Wärmekapazitäten hat, leicht beschafft werden kann, für Pflanzen und Tiere sicher ist und keine mühsame Reinigung von Gefahrstoffen erfordert.
- **Warmwasser vs. Kühler:** Lenovo Neptune® Direktwasserkühlungstechnologie kann Eintrittstemperaturen von bis zu 45 °C aufnehmen, anstatt vorgekühltes Wasser zu benötigen, wodurch der Energiebedarf für das gesamte Rechenzentrum reduziert und spezielle Rechenzentrum-Klimaanlagen überflüssig werden. Darüber hinaus kann das heiße Ausgangswasser innerhalb der Anlage wiederverwendet werden, wodurch Abwärme aufgrund des hohen Energieinhalts, der in Form von Wärme gespeichert wird, die im Bereich von 65 °C liegen kann, nutzbar gemacht werden kann.
- **Gelötete Verbindungen:** Lenovo verwendet gelötete Verbindungen anstelle von ORFS-Fittings (O-Ring-Gleitdichtung), da sie für ihre leckfreie Leistung, Unterstützung und Stabilität (wodurch ein Verdrehen des Rohrs verhindert wird) bekannt sind.

Zu den Branchenvorteilen zählt zudem Lenovos etablierter Supercomputing-Stammbaum gemäß der TOP500-Supercomputing-Liste. Lenovo hat mehr TOP500-Systeme als jeder andere Anbieter (Stand der neuesten Liste im Juni 2024). Neun von zehn der schnellsten TOP500-Systeme von Lenovo verwenden eine Neptune®-basierte Flüssigkeitskühlung.

Zweitens entsprechen der Fertigungsprozess und der Lieferansatz von Lenovo den entstehenden Anforderungen an nachhaltige Rechenzentren. Lenovo hat Supply Chain Intelligence (SCI) entwickelt, eine KI-gestützte Lösung, die Lieferkettendaten kontinuierlich analysiert, um potenzielle Probleme zu identifizieren und schnell zu beheben.

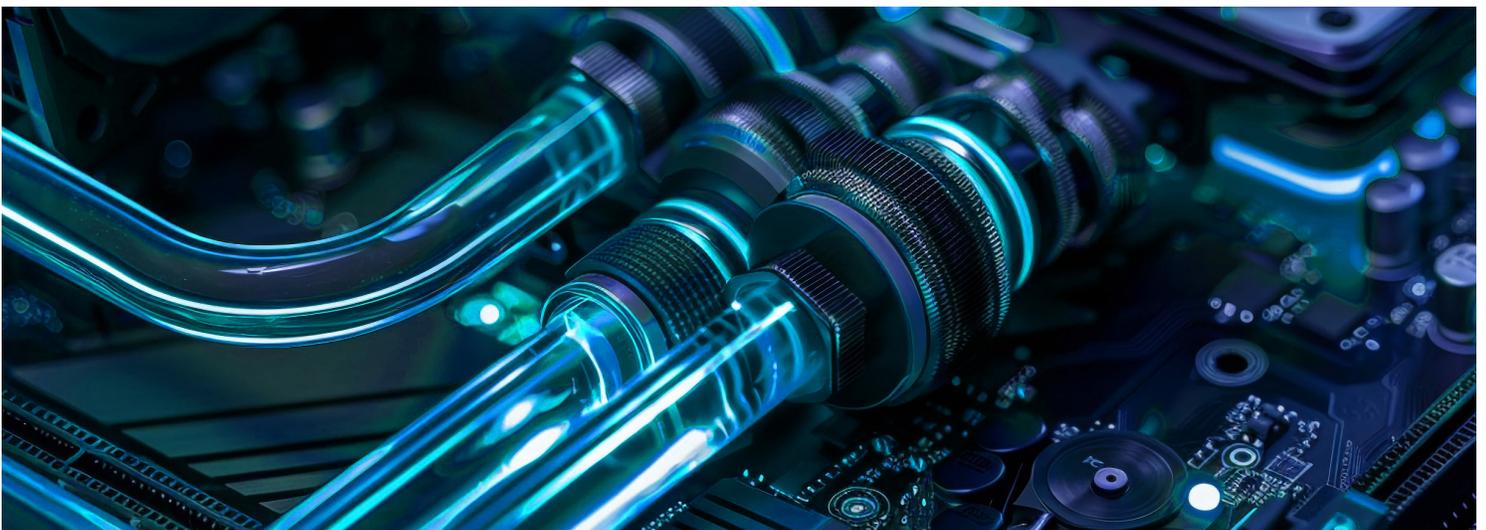
Von zentraler Bedeutung ist, dass Lenovo Neptune® Server zusammen als Lösung getestet werden und nicht nur als einzelne Systeme. Systeme werden mit deionisiertem Wasser gefüllt, online gebracht und stressgetestet, um sicherzustellen, dass der gesamte Systembetrieb und der Wasserfluss hohen Standards entsprechen. Und das bevor sie verpackt und für den Versand an einen Kunden vorbereitet werden, indem die Flüssigkeit abgelassen und der Kreislauf mit Stickstoff unter Druck gesetzt wird. Der Versand der Geräte ohne Flüssigkeit im Kreislauf führt zu einem Wettbewerbsvorteil gegenüber anderen Anbietern, die ihre Lösungen mit Flüssigkeit ausliefern und schon Komponentenschäden und Flüssigkeitslecks während des Transports erlebt haben.

Drittens stellen wir fest, dass Lenovos langjährige Erfahrung mit Flüssigkeitskühlung auch in Bezug auf Support und Wartungsfreundlichkeit der Lösung zum Ausdruck kommt. Lenovo bietet First-Line-Support für alle Geräte und Komponenten der Neptune®-Lösung, von Servern bis zu Cooling Distribution Units (CDUs), im Gegensatz zu anderen Anbietern, bei denen im Allgemeinen jeder Anbieter für den Support seiner eigenen Lösungskomponenten sorgen muss. Und Neptune® „Chassis-enclosed“-Systeme verbinden sich bei der Installation automatisch mit Wasser, sodass einzelne Neptune® Server gewartet werden können, während andere in Produktion bleiben.

Die Lenovo Neptune® Flüssigkeitskühlungslösung wird durch die Fähigkeit des Unternehmens gestärkt, voroptimierte, getestete und zertifizierte IT-Lösungen zu liefern. Dies hilft sicherzustellen, dass Lenovo Neptune® durch die Integration von werksseitig getesteten Komponenten leistungsstarke, skalierbare und energieeffiziente Lösungen bietet, um Zuverlässigkeit und Kosteneffizienz zu gewährleisten.

Schließlich bietet Lenovo Neptune® Flüssigkeitskühlung eine flüsterleise, hohe Leistung bei geringerem Stromverbrauch und kompakter Stellfläche, sodass Kunden eine höhere Dichte und mehr Output mit ihrem Rechenzentrum erreichen können. Während des Betriebs recycelt die direkte Wasserkühlungslösung Warmwasserkreisläufe, um Rechenzentrumssysteme zu kühlen und hält alle Serverkomponenten kühl, so dass der Bedarf an lauten und stromhungrigen Systemlüftern im Rechenzentrumsbetrieb verringert wird.

Es ist bemerkenswert, dass Digital Realty sich kürzlich für Lenovo Neptune® Kühltechnologie beim Aufbau seines Colocation-Angebots mit hoher Dichte in mehr als der Hälfte seiner Rechenzentren weltweit entschieden hat. Aus unserer Sicht ist dies eine wichtige Bestätigung bei der Validierung der Produktionsbereitschaft des Lenovo Neptune® Portfolios der 6. Generation zur Bewältigung der besonderen Herausforderungen von Workloads mit hoher Dichte, die durch KI entstehen.





Abschnitt 3: Vorteile der Lenovo Neptune® Flüssigkeitskühlungstechnologie und Wettbewerbsvorteile – Eine nachgewiesene Erfolgsbilanz bei Lösung, Marktführerschaft und technischer Innovation

Lenovo Neptune® ist keine einzelne Technologie, die einheitlich im gesamten ThinkSystem-Portfolio angewendet wird. Neptune® bietet einen ganzheitlichen Kühlungsansatz, der nicht nur CPUs kühlt; er umfasst Optionen für die direkte Warmwasserkühlung ganzer Systeme, einzelner Komponenten und eine Wasserkühlung auf Rack-Ebene für die Hintertür. Ein solch umfassender Ansatz untermauert die Fähigkeit, Leistung und Effizienz zu optimieren und Lösungen für einzelne Kunden und Installationen anzupassen:

- **Direkte Wasserkühlung (DWC):** Die Lösung bietet DWC einzelner Systemkomponenten, die im gesamten ThinkSystem-Portfolio von Lenovo verfügbar sind, sowie vollständig systemgekühlte Server, bei denen DWC einen parallelen Flüssigkeitsfluss über alle wärmeerzeugenden Komponenten (CPUs, Beschleuniger, DIMMs, Netzwerke usw.) mit einem Wasserkreislauf nutzt, der bis zu 98 % der Systemwärme aufnehmen und den Systempegel sowie den Stromverbrauch im Rechenzentrum um bis zu 40 % reduzieren kann.
- **Flüssigkeitsunterstützte Kühlung:** Lenovo Neptune® flüssigkeitsunterstützte Kühlung bietet die Vorteile von Flüssigkeit in einem luftgekühlten System. Mit entweder einem Thermotransfermodul (TTM) oder einem Flüssigkeit-Luft-Wärmetauscher (L2A) profitieren traditionelle luftgekühlte Systeme von der Flüssigkeitskühlung mit speziell entwickelter Wärmeableitung, alles ohne zusätzliche Installationseinheiten. Dies unterstützt Kunden, die sich dafür entscheiden, den Status quo für luftgekühlte Rechenzentren beizubehalten, ohne Kompromisse bei der Einführung von Flüssigkeitskühlung einzugehen.
- **Rack-Wasserkühlung:** Der Wärmetauscher an der Hintertür (RDHX) bietet 100 % Wärmeabfuhereffizienz, ohne bewegliche Teile oder Strom zu benötigen. RDHX funktioniert mit standardmäßigen luftgekühlten Servern, Speicher und Netzwerken ohne Änderung, wodurch die Assimilation in die bestehende Rechenzentrumsinfrastruktur erleichtert wird. So können die Vorteile von In-Rack-CDUs genutzt werden, die Effizienzgewinne gegenüber herkömmlichen Computerraum-Klimaanlagen (CRAC) bieten.

Wir stellen fest, dass sich die Lenovo Neptune®-Technologie als entscheidend für die Unterstützung von Anwendungsfällen mit höchster Priorität erwiesen hat, einschließlich Engineering, Modellierung, Simulationsworkloads und Animationsfilmproduktion, die alle Rechenleistung bei Spitzenleistung erfordern. Lenovo Neptune® ist beispielsweise ein entscheidender Faktor in Schlüsselsektoren wie Fintech, computergestütztem Engineering und computergestützter Fluidodynamik (CAE/CFD), elektronischer Designautomatisierung (EDA), Wetter- und Klimamodellierung und -vorhersage, Bioinformatik, Georaum/ Energie, Geowissenschaften, Forschung und Animations-Render-Farmen. Lenovos lange Geschichte im HPC-Bereich hat das Unternehmen gut positioniert, um die Zukunft der KI und aller nachgelagerten Herausforderungen für Stromversorgung, Kühlung und Nachhaltigkeit im Zusammenhang mit Rechenzentrumserweiterungen zur Unterstützung von KI anzugehen.

Darüber hinaus hat die Lenovo Neptune® Flüssigkeitskühlungslösung den Preis für Nachhaltigkeit der Business Intelligence Group 2024 gewonnen und ist 2024 CRNs bestes Produkt für grüne Energie des Jahres, Gewinner des SEAL Sustainable Product Award 2024 und HPCwires bestes HPC-Serverprodukt oder -technologie für 2023, wodurch seine ökosystemweiten ESG-Zertifizierungen weiter gestärkt werden.

Neptune® Flüssigkeitskühlungslösungen profitieren immens von Lenovos Bestreben für Partnerschaften, wie das Beispiel der NVIDIA-Partnerschaft zeigt. Dies umfasst neue umfassende Services, die von NVIDIA über das Lenovo AI Center of Excellence für GenAI bereitgestellt werden, das als die weltweit wichtigste Investitionspriorität für Geschäfts- und IT-Entscheidungsträger in Technologieinvestitionen identifiziert wurde. Zum Beispiel verwendet der Lenovo ThinkSystem SR780a Server Neptune® Flüssigkeitskühlung, um eine ultra-effiziente Stromnutzungseffektivität (wie viel Strom ein Rechenzentrum insgesamt im Vergleich zum Stromverbrauch der einzelnen IT-Geräte verbraucht) von 1,1 zu erreichen. Durch die Implementierung einer direkten Wasserkühlung von CPUs, GPUs und NVIDIA NVSwitch-Technologie kann dieses System maximale Leistung erzeugen, ohne an thermische Grenzen zu stoßen.

Der Lenovo AI Fast Start Service bietet Live-Lösungen, um die Geschäfts-, Betriebs- und Technologieergebnisse von GenAI zu präsentieren. So können Unternehmen KI mithilfe von Full-Stack-NVIDIA-basierten Technologien durch Lenovo AI Fast Start for NVIDIA AI Enterprise schnell skalieren und optimieren. Die neuen Lenovo AI Fast Start for NVIDIA NIM Inferenz-Microservices bieten Entwicklern einfach zu verwaltende containerisierte und optimierte Inferenz-Engines für NVIDIA AI Foundation-Modelle, die von NVIDIA angeboten werden.





Abschnitt 5: Schlussfolgerungen und Empfehlungen

KI-Anwendungen, die auf speziell entwickelten Beschleunigertechnologien basieren, verbessern die Ergebnisse in vielen Branchen erheblich. Die Flüssigkeitskühlungstechnologie entwickelt sich zu einem wesentlichen Faktor für KI mit höchster Leistung. Basierend auf der Analyse von Futurum ist Lenovo Neptune® Flüssigkeitskühlung der Branchenführer im Bereich der Flüssigkeitskühlung. Lenovo hat mehr als zehn Jahre Erfahrung mit Flüssigkeitskühlung und bietet daher erstklassiges Design, Fertigung, Lieferung und Support.

Aus unserer Sicht wird das Lenovo Neptune® Flüssigkeitskühlungs-Wertversprechen durch Kunden in Produktionsnetzwerken gründlich validiert. Geely Auto F&E stellte eine HPC-Plattform auf der Basis von Lenovo ThinkSystem SD650 V3-Servern mit Neptune® Wasserkühlung bereit, die die Leistung um 35 % steigerte und eine Grundlage für mehr F&E-Innovationen schaffte. Geely wählte Lenovo für die Bereitstellung eines On-premises-HPC-Clusters, um Cloud-Kosten und Komplexität zu vermeiden. Der Wechsel brachte sofortige Vorteile, da der neue Cluster die HPC-Leistung um 35 % steigerte und den Stromverbrauch um 1 Million kWh/Jahr reduzierte.

Daher können wir bei der Bewertung von Lenovos Neptune®-Lösung zur Erfüllung der Flüssigkeitskühlungsanforderungen Folgendes feststellen:

Ganzheitlicher Ansatz. Lenovo Neptune® umfasst eine direkte Warmwasserkühlung ganzer Systeme, Wasserkühlung einzelner Komponenten und flüssigkeitsunterstützte Kühlung für luftgekühlte Systeme, was einen umfassenden Ansatz zur Optimierung der Leistung darstellt.

Wettbewerbsvorteile. Lenovos 12-jährige Erfahrung in der Bereitstellung von Flüssigkeitskühlungslösungen auf Produktionsebene, einschließlich EPDM-Schläuche, wasserorientierte, kupferbasierte Kühlkreisläufe und Verwendung von gelöteten Verbindungen, bietet Wettbewerbsvorteile bei Design, Herstellung und Lieferungsprozess, der Gesamtqualität und dem Supportprozess.

Nachhaltigkeit/ESG-Erfüllung. Lenovo Neptune® wurde für wichtige Nachhaltigkeitsziele wie der Kohlenstoffreduzierung entwickelt, ohne die Rechenleistung zu beeinträchtigen, die Unternehmen bei der Bewertung von Flüssigkeitskühlungssystemen priorisieren müssen.

Wichtige Informationen zu diesem Bericht

MITWIRKENDE

Steven Dickens

Chief Technology Advisor | The Futurum Group

Ron Westfall

Research Director | The Futurum Group

HERAUSGEBER

Daniel Newman

CEO | The Futurum Group

ANFRAGEN

Kontaktieren Sie uns gerne, wenn Sie über diesen Bericht sprechen möchten. The Futurum Group wird Ihnen umgehend antworten.

ZITIERUNGEN

Dieser Artikel kann von akkreditierten Medien und Analysten zitiert werden, muss jedoch im Kontext zitiert werden, den Namen des Autors, den Titel des Autors und „The Futurum Group“ enthalten. Alle anderen Personen müssen sich vorher von The Futurum Group eine schriftliche Genehmigung einholen.

LIZENZIERUNG

Dieses Dokument, einschließlich aller unterstützenden Materialien, gehört The Futurum Group. Diese Publikation darf ohne vorherige schriftliche Genehmigung von The Futurum Group in keiner Form reproduziert, verbreitet oder geteilt werden.

OFFENLEGUNG

The Futurum Group bietet vielen Hightech-Unternehmen ihre Forschungs-, Analysen und Beratungsdienste an, einschließlich der in diesem Artikel genannten Unternehmen. Mitarbeitende des Unternehmens halten keine Aktienpositionen bei den in diesem Dokument genannten Unternehmen.



ÜBER LENOVO UND NVIDIA

Lenovo bringt die neue Ära KI-gestützter Innovationen für alle. Unser Full-Stack-Portfolio bietet leistungsstarke, flexible und verantwortungsvolle KI-Lösungen, um Branchen zu transformieren und Einzelpersonen zu unterstützen. Wir schaffen eine Zukunft mit intelligenterer KI für alle. Bei Lenovo sind wir davon überzeugt, dass KI für die Öffentlichkeit und KI für Unternehmen in Zukunft koexistieren wird. Lenovo bringt KI zu Ihnen und Ihren Daten.

In Partnerschaft mit NVIDIA werden hybride KI-Lösungen durch technische Zusammenarbeit speziell entwickelt, um KI effizient für Kundendaten zu nutzen, wo und wann Benutzer sie am meisten benötigen. Dies wird die Vision von Lenovo weiterentwickeln, KI für alle zu ermöglichen und Unterstützung für die Einführung bahnbrechender Technologien und Architektur für die nächste Generation generativer KI zu bieten. Lenovo Hybridlösungen, die bereits für die Ausführung von NVIDIA KI Enterprise-Software für sichere, unterstützte und stabile KI-Produktion optimiert sind, bieten Entwicklern Zugang zu NVIDIA Microservices, einschließlich NVIDIA NIMs und NeMo Retriever.



ÜBER THE FUTURUM GROUP

[TheFuturum Group](#) ist ein unabhängiges Forschungs-, Analyse- und Beratungsunternehmen, das sich auf digitale Innovationen und marktverändernde Technologien und Trends spezialisiert hat. Unsere Analysten, Forscher und Berater helfen Führungskräften aus der ganzen Welt dabei, tektonische Veränderungen in ihren Branchen zu antizipieren und disruptive Innovationen zu nutzen, um sich einen Wettbewerbsvorteil in ihren Märkten zu verschaffen oder zu erhalten.



KONTAKTINFORMATIONEN

The Futurum Group LLC | futurumgroup.com | (833) 722-5337 |