



Vloeistofkoeling: hoe diepgaande expertise energie-efficiënte compute mogelijk maakt voor AI en andere

AUTEUR

Steven Dickens

Technologisch hoofdadviseur | The Futurum Group

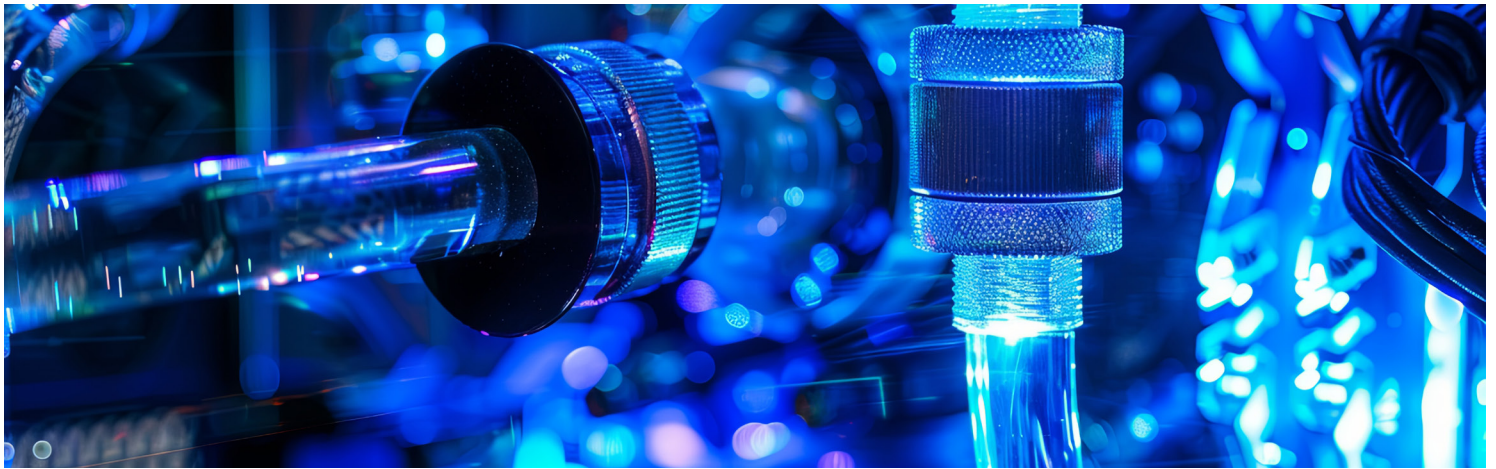
Ron Westfall

Directeur onderzoek | The Futurum Group

AUGUSTUS 2024

IN PARTNERSCHAP MET





Samenvatting

Kunstmatige intelligentie (AI) en in het bijzonder generatieve AI (GenAI)-toepassingen nemen de wereld stormenderhand over. In een breed scala van sectoren maakt AI doorbraken mogelijk in alles, van kleinhandel tot productie, van financiële diensten tot gezondheidszorg, van fundamenteel onderzoek tot cyberbeveiliging.

Deze doorbraken worden mogelijk gemaakt door computertechnologie die we ons een paar jaar geleden nog niet konden voorstellen. Naarmate deze technologie vordert, is het energieverbruik voor vrijwel elk onderdeel van een server - CPU's, geheugen, versnellers, netwerken, enz. - het afgelopen decennium met 200% gestegen. Deze energieverslindende systemen die nodig zijn voor AI-workloads vullen datacenters over de hele wereld en leveren een belangrijke bijdrage aan de totale stroomtoename in de datacenterruimte. Alleen al in de VS is de energievraag vooral te wijten aan de [zal toenemen met 5.1% tot 9.1% van de totale energie van de VS](#) vraag naar AI-datacenters tegen 2030. Vandaag de dag is het energieverbruik van de VS [slechts 4% volgens het Electric Power Research Institute](#).

Het Joule-effect stelt in wezen dat wanneer een elektrische stroom door een geleider met weerstand loopt, die elektrische energie wordt omgezet in warmte. Dit betekent dat naarmate er meer stroom wordt verbruikt, er meer warmte wordt gegenereerd. Omgaan met de warmte die wordt gegenereerd door de toegenomen energievraag van AI en High-Performance Computing (HPC) is een uitdaging geworden voor veel organisaties. Grote toenames in energieverbruik hebben stroomafwaartse effecten. Componenten met hoger vermogen, zoals CPU's, GPU's, geheugen, enz. genereren meer gelokaliseerde warmte en warmte is de vijand van silicium. Warmte die niet wordt afgevoerd, beperkt de prestaties van CPU's en GPU's en kan het silicium zelf beschadigen en/of ernstige problemen veroorzaken zoals oververhitting en brand.

Tegen deze achtergrond werken siliciumleveranciers niet alleen aan het verbeteren van de efficiëntie van individuele componenten, maar richten ontwerpers van silicium en systemen zich ook steeds meer op vloeistofkoelingsmethoden. Vloeistof kan namelijk sneller en met een hogere capaciteit warmte absorberen dan traditionele luchtkoelingsmethoden. Vloeistofkoeling kan ook helpen voldoen aan de toenemende vraag naar het verbeteren van de energie-efficiëntie van datacenters voor uitbreidende AI-workloadomgevingen. Dit komt door de downstream-effecten, zoals het wegvallen van het stroomverbruik van airconditioning, van de energie die nodig is om water te koelen en van de energie die nodig is voor ventilatoren om lucht te verplaatsen. Dit omvat het beheren en optimaliseren van het energieverbruik door het gebruik van vloeistofkoelingstechnologieën met hogere inlaattemperaturen (geen noodzaak om water te koelen), terwijl er tegelijkertijd wordt gegarandeerd dat de systeemprestaties niet in het gedrang komen en de kerndichtheid per datacenter niet wordt beïnvloed.

De overstap naar vloeistofkoeling kan ook bijdragen aan het vermogen van een bedrijf om te werken aan ESG-doelen (Environmental, Social and Governance) die betrekking hebben op koolstofemissies en duurzame praktijken. Als een bedrijf of aanbieder van co-locatie datacenters niet moderniseert, maar de status quo van luchtkoeling handhaaft, dan wordt de energie-explosie van AI-siliciumkracht verergerd door de behoefte aan koudere lucht die sneller moet bewegen. Er zal meer stroom nodig zijn voor luchtbehandeling dan voor IT-verbruik en dat zal een bedrijf indirect belemmeren in het verminderen van zijn ecologische voetafdruk. Alles bij elkaar vormt de milieu-impact van AI-infrastructuur,

met name als gevolg van de hoge energievraag, een uitdaging voor bedrijven die ernaar streven om technologische vooruitgang in evenwicht te brengen met hun ESG-verplichtingen. Vloeistofkoeling kan helpen deze problemen aan te pakken.

We onderzoeken waarom Lenovo Neptune® 6^e generatie vloeistofkoelingsoplossingen superieur zijn in het mogelijk maken van AI-workloadoptimalisatie zonder de duurzaamheidsdoelstellingen van de organisatie in gevaar te brengen. Dit rapport richt zich op de diepgaande en langdurige investeringen die Lenovo heeft gedaan in techniek en innovatie als het gaat om vloeistofkoeling. We onderzoeken hoe de engineering op systeem- en oplossingsniveau voor Lenovo Neptune® vloeistofkoeling hogere kwaliteit en betere statistieken laat zien dan de industrienormen.

Lenovo heeft een lange geschiedenis van innovatie in vloeistofkoeling voor x86-gebaseerde systemen die begon in de IBM-dagen vóór de overname van IBM's System X-activiteiten in 2014. Lenovo's 12 jaar ervaring in het leveren van vloeistofkoeling op productieniveau op x86-gebaseerde systemen (zowel met als zonder GPU's) komt tot uiting in de aanpak van het bedrijf, van de details in het ontwerp-, productie- en leveringsproces tot de algemene kwaliteits- en ondersteuningsprocessen.

De technische verschillen van het bedrijf zorgen uiteindelijk voor een concurrentievoordeel met hogere kwaliteit en doorbraken op het gebied van warmteopvang. Lenovo's diepgaande ervaring en begrip van geleidende materialen en vloeistofgedrag in het hele veld en hun unieke aanpak bij het ontwerpen van vloeistofgekoelde oplossingen leveren een baanbrekende systeemkwaliteit plus een concurrentievoordeel bij het leveren van energie-efficiënte compute.



Deel 1: Het evoluerende landschap van vloeistofkoeling

Het is niet moeilijk om te zeggen dat AI een revolutie teweegbrengt in een groot aantal sectoren. AI helpt kleinhandelaars de klantervaring te verbeteren, voorraadverliezen te verminderen en de winst te verhogen. Het helpt onderzoekers bij het ontwikkelen van nieuwe farmaceutische producten, het verkennen van de oorsprong van het universum en het beperken van de klimaatverandering. AI helpt fabrikanten de kwaliteit te verhogen, de procesbewaking te verbeteren en voorspellend onderhoud te bieden; het stelt zorgverleners in staat overhead te verminderen, diagnoses te verbeteren en meer tijd met hun patiënten door te brengen. En op het gebied van financiële diensten verbetert AI de fraudedetectie en maakt het snellere, nauwkeurigere investeringsbeslissingen mogelijk.

Deze doorbraken worden mogelijk gemaakt door computertechnologie die we ons een paar jaar geleden nog niet konden voorstellen. Naarmate deze technologie vordert, is het energieverbruik voor vrijwel elk onderdeel van een server – CPU's, geheugen, versnellers, netwerken, enz. – het afgelopen decennium met ruwweg 200% gegroeid. Het verwerkingsvermogen is toegenomen tot 500 W voor CPU's en meer dan 1000 W voor GPU's, met een overeenkomstige toename van het vermogen voor andere servercomponenten. De totale vermogensvereisten voor een enkel rack zijn toegenomen van ongeveer 15 kW tot 100 kW en meer.

Omgaan met de warmte die al deze elektrische energie produceert, is voor veel klanten een uitdaging geworden. Uiteraard zijn alternatieve methoden voor koeling nodig omdat klanten proberen een evenwicht te vinden tussen hun verlangen naar nieuwe AI-toepassingen, betere IT-prestaties en de economische en ecologische realiteit van server-, rack- en datacenterkoeling. Vloeistofkoeling is naar voren gekomen als een belangrijke factor voor de nieuwe generatie AI- en GenAI-toepassingen.

Laten we eens dieper ingaan op de drijfveren voor vloeistofkoeling. Niet alleen verbruiken de individuele servercomponenten die we hebben genoemd steeds meer energie, maar AI- en HPC-workloads veranderen ook de energievraag op datacenterniveau. AI-training verloopt synchroon, waarbij alle elementen in de cluster met enorme intensiteit samenwerken aan één enkele service. De omvang van een AI-trainingscluster kan enorm zijn; er zijn minimaal 5.000 servers nodig, waarbij elke server meerdere CPU's en GPU's gebruikt. En met een enkele server die ongeveer 10 kW vermogen vereist, is het eenvoudig om te zien hoe het probleem op schaal nog groter wordt. AI-workloads kunnen tot drie keer de energie van typische cloud-workloads verbruiken, wat een overeenkomstige toename van warmte produceert.

We hebben besproken hoe een toename van het vermogen van apparaten het totale datacenter beïnvloedt. Laten we eens kijken naar andere relaties die bestaan voor deze systemen van de volgende generatie die zijn ontworpen om deze geavanceerde computeworkloads te ondersteunen.

Geavanceerde compute van de volgende generatie stelt uitdagingen op het gebied van systeemontwerp op hoog niveau op drie belangrijke gebieden: thermisch, mechanisch en elektrisch:

- **Thermisch:** warmere componenten in het algemeen (bijv. spanningsregelaars, CPU's, GPU's, geheugen, netwerken, enz.), lagere drempeltemperaturen van silicium die van generatie tot generatie nodig zijn, geografische uitdagingen (hete, droge datacenterlocaties)
- **Mechanisch:** dramatische toename van socketgroottes om hogere aantallen transistors te kunnen verwerken, toename van het aantal dual in-line memory modules (DIMM) per CPU of GPU, toename van het aantal spanningsregelaars (Voltage Regulator of VR) en toename van netwerkkaartgroottes, samen met een toename van het aantal boardlagen en een grotere siliciumhoogte
- **Elektrisch:** grote vraag naar CPU's van ≥ 600 W, GPU's van ≥ 1200 W en verschuiving naar 48 V-ingang. Eén AI-systeem kan meer dan 10 kW verbruiken.

Laten we vervolgens wat dieper ingaan op hoe de twee methoden, vloeistofkoeling en luchtkoeling, verschillen en waarom het belangrijk is voor AI. Vloeistof heeft een hogere thermische capaciteit dan lucht. Onder thermische capaciteit wordt verstaan de hoeveel warmte die een medium kan absorberen voordat de temperatuur ervan verandert. De thermische capaciteit van vloeistof is veel groter dan die van lucht, wat simpelweg betekent dat vloeistof een aanzienlijk hogere hoeveelheid warmte kan absorberen voordat de temperatuur van de vloeistof stijgt. Hierdoor verloopt de warmteafvoer aanzienlijk efficiënter.

Er is ook een probleem met de voetafdruk van het datacenter veroorzaakt door koeling. Bij luchtkoeling zijn de prestaties van het koellichaam afhankelijk van het totale oppervlak. Hoe groter het oppervlak, hoe beter de prestaties van het koellichaam. Het vergroten van het oppervlak van het koellichaam kan worden bereikt door het horizontaal of verticaal of beide uit te breiden. Uiteindelijk resulteert deze toename in oppervlakte in grotere systemen qua hoogte, waardoor de hoeveelheid rekenkracht die in een datacenter kan worden gestopt, afneemt. Vertaald betekent dit minder rekenkracht voor een toename van het vermogen.

Vloeistofkoeling kan echter meer rekenkracht behouden in traditionele servermodellen omdat het zo efficiënt is in het absorberen van warmte en het effect van oppervlakte in relatie tot prestaties drastisch wordt verminderd, waardoor dichtere compute mogelijk wordt in een traditionele computerruimte. Vloeistofgekoelde servers vereisen minder ruimte, genereren minder geluid (minder ventilatorvermogen vanwege een verminderde behoefte aan hoge luchtstroom) en zorgen voor een algeheel koelere datacenteromgeving.

Al deze factoren hebben geleid tot een enorme hoeveelheid recente investeringen in vloeistofkoeling in de datacenterindustrie. We merken dat er een algemene perceptie is dat alle oplossingen voor vloeistofkoeling hetzelfde zijn en dat er nauwelijks verschillen zijn tussen leveranciers. We hebben echter vastgesteld dat dit helemaal niet het geval is. Dankzij de 12 jaar ervaring van Lenovo in de ontwikkeling van vloeistofkoelingstechnologie heeft het bedrijf oplossingen kunnen bouwen die niet alleen voldoen aan de algemene normen, maar deze zelfs ruimschoots overtreffen. Volgens onze vaststelling zijn ze aanzienlijk beter dan de concurrentie op het gebied van ontwerp, productie, levering en service.



Deel 2: Waarom Lenovo Neptune® de bewezen oplossing is voor AI-workloads: superieure vloeistofkoelingstechnologie

We onderzoeken waarom Lenovo's 6^e generatie Neptune® Liquid Cooling-oplossing een robuuste oplossing biedt die vandaag beschikbaar is om te voldoen aan verschillende eisen van AI-workloads. De oplossing van Lenovo biedt aanzienlijk hogere kwaliteit en levert meer efficiëntie dan luchtkoeling op belangrijke gebieden zoals warmteafvoer, terwijl dit alles zorgt voor topprestaties en een sterk verbeterde energie-efficiëntie. In wezen biedt de Lenovo Neptune®-oplossing een holistische koelaanpak die we als marktleidend beschouwen voor alle datacenters.

Ten eerste valt de Lenovo Neptune®-technologieportfolio op. Lenovo Neptune® is gebaseerd op meer dan tien jaar expertise op het gebied van vloeistofkoeling en heeft meer dan 40 patenten op zijn naam staan. Het systeem speelt een essentiële rol bij grootschalige implementaties van supercomputers en AI-clusters, waardoor organisaties hoogwaardige AI op elke schaal kunnen implementeren. Lenovo was de eerste die in juni 2012 een vloeistofgekoelde petaschaal-supercomputer met hoge inlaattemperatuur leverde aan het Leibniz Supercomputing Centre (LRZ). Deze computer debuteerde op nummer 4 van de Top 500 Supercomputers (www.top500.org), een lijst met de 500 snelste systemen ter wereld.

Deze lange ervaring met vloeistofkoeling heeft Lenovo in staat gesteld zijn technologie en productieproces te verfijnen. Enkele van de superieure onderscheidende kenmerken van Lenovo-technologie die we hebben waargenomen, zijn onder meer:

- **Slangen (intern en extern):** Lenovo gebruikt hoogwaardige EPDM-slangen (ethyleen propyleen dieen monomeer). EPDM staat bekend om zijn hoge PSI-sterkte en laag durometerbereik, wat duidt op een grotere flexibiliteit dan de plastic buizen die door sommige anderen op de markt worden gebruikt. Deze slangen zijn ook intern bewerkt met peroxide, wat roestvorming en slijtage van de slang voorkomt.
- **Hoofdkoellus:** de belangrijkste interne lussen en koude platen van Lenovo gebruiken koper in plaats van gefluoreerd ethyleen propyleen (FEP) vanwege de historische reputatie van koper als een uitstekende elektrische geleider, sterkte plus duurzaamheid en efficiëntere warmteafvoer.
- **Geoptimaliseerde, gepatenteerde koude platen:** Lenovo's ontwerp van koude platen met lage drukval maximaliseert de warmteafvoer voor versnellers die nu ongeveer 700 W en in de toekomst meer dan 1000 W verbruiken.
- **Water eerst:** Lenovo Neptune® gebruikt water in plaats van andere vloeistoffen zoals polyethyleenglycol, omdat water een van de hoogste thermische capaciteiten heeft, gemakkelijk te verkrijgen is, veilig is voor planten en dieren en geen rommelige reiniging met gevaarlijke stoffen vereist.
- **Warm water versus koelers:** Lenovo Neptune® directe waterkoelingstechnologie kan inlaattemperaturen gebruiken tot 45 °C in plaats van voorgekoeld, waardoor de energiebehoefte voor het totale datacenter wordt verminderd en de noodzaak voor gespecialiseerde airconditioningapparatuur voor datacenters wordt geëlimineerd. Bovendien kan het warme uitgangswater herbruikt worden binnen de faciliteit, waarbij afvalwarmte omgezet wordt in waarde dankzij de hoge energie-inhoud die opgeslagen is in de vorm van warmte die kan variëren tot 65 °C.
- **Gesoldeerde verbindingen:** Lenovo gebruikt gesoldeerde verbindingen in plaats van ORFS-fittingen (O-ring face seal), omdat deze bekend staan om hun lekvrije prestaties, ondersteuning en stabiliteit (waardoor verdraaien van de buis wordt voorkomen).

Tot de referenties in de sector behoort de gevestigde supercomputer-stamboom van Lenovo volgens de TOP500-lijst van supercomputers. Lenovo heeft meer TOP500-systemen dan elke andere leverancier (volgens de laatste lijst die in juni 2024 werd uitgebracht). Negen van de tien snelste TOP500-systemen van Lenovo gebruiken vloeistofkoeling op basis van Neptune®.

Ten tweede sluiten het productieproces en de leveringsaanpak van Lenovo aan bij de opkomende eisen van duurzame datacenters. Lenovo heeft Supply Chain Intelligence (SCI) ontwikkeld, een AI-oplossing die continu gegevens over de toeleveringsketen analyseert om mogelijke problemen te identificeren en snel op te lossen.

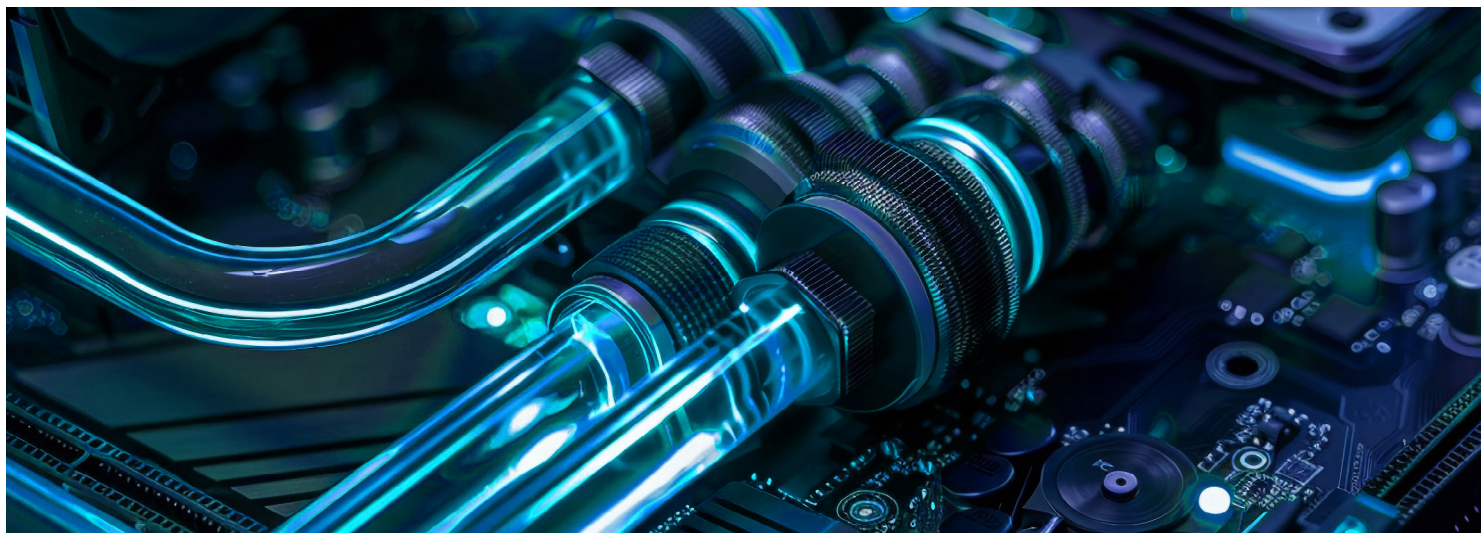
Van groot belang is dat Lenovo Neptune®-servers samen worden getest als een oplossing, en niet als afzonderlijke systemen. Systemen worden gevuld met gedemineraliseerd water, in bedrijf gesteld en onderworpen aan stresstests om te garanderen dat de algehele werking van het systeem en de waterstroom voldoen aan hoge normen. Vervolgens worden ze verpakt en gereedgemaakt voor verzending naar een klant. Dit gebeurt door de vloeistof af te tappen en de lus onder druk te zetten met stikstof. Verzenden zonder vloeistof in de lus levert een concurrentievoordeel op ten opzichte van andere leveranciers die wel met vloeistof verzenden en schade aan onderdelen en vloeistoflekage tijdens het transport hebben ervaren.

Ten derde vinden we dat Lenovo's lange ervaring met vloeistofkoeling ook doorschemert in de ondersteuning en onderhoudbaarheid van de oplossing. Lenovo biedt one-stop, eerstelijns ondersteuning voor alle apparatuur en componenten van Neptune®-oplossingen, van servers tot koeldistributie-eenheden (Cooling Distribution Units of CDU's), in tegenstelling tot andere providers die over het algemeen elke leverancier vragen om zijn eigen oplossingscomponenten te ondersteunen. En Neptune® chassis-omhulde systemen worden na installatie automatisch aangesloten op water, zodat afzonderlijke Neptune®-servers onderhouden kunnen worden terwijl andere in productie blijven.

De Lenovo Neptune® Liquid Cooling-oplossing wordt versterkt door het vermogen van het bedrijf om vooraf geoptimaliseerde, geteste en gecertificeerde IT-oplossingen te leveren. Dit helpt ervoor te zorgen dat Lenovo Neptune® hoogpresterende, schaalbare en energie-efficiënte oplossingen biedt door de integratie van in de fabriek geteste componenten om betrouwbaarheid en kosteneffectiviteit te waarborgen.

Ten slotte levert Lenovo Neptune® Liquid Cooling fluisterstille hoge prestaties met een lager energieverbruik in een compacte behuizing, zodat klanten een hogere dichtheid en meer output uit hun datacenter kunnen halen. In werking recycleert de oplossing voor directe waterkoeling warmwaterlussen om datacentersystemen te koelen en houdt alle serveronderdelen koel. Hierdoor is er minder behoefte aan lawaaiërende en energievervlindende systeemventilatoren in datacenters.

Het is opmerkelijk dat Digital Realty onlangs heeft gekozen voor Lenovo Neptune®-koeltechnologie bij het bouwen van zijn aanbod van co-locatie met hoge dichtheid in meer dan de helft van zijn datacenters wereldwijd. Naar onze mening is dit een belangrijke bevestiging van de productieklaarheid van het Lenovo 6^e generatie Neptune®-portfolio om te voldoen aan de specifieke uitdagingen van workloads met hoge dichtheid die AI met zich meebrengt.





Deel 3: Voordelen en concurrentievoordelen van Lenovo Neptune® Liquid Cooling-technologie - een bewezen staat van dienst op het gebied van oplossingen, leiderschap en technologische innovatie

Lenovo Neptune® is geen unieke technologie die uniform wordt toegepast in de hele ThinkSystem-portfolio. Neptune® biedt een holistische koelaanpak die niet alleen CPU's koelt, maar ook opties bevat voor directe warmwaterkoeling van complete systemen, afzonderlijke componenten en waterkoeling via de achterdeur op rackniveau. Een dergelijke alomvattende aanpak ondersteunt het vermogen om prestaties en efficiëntie te optimaliseren en oplossingen op maat te maken voor individuele klanten en installaties:

- **Directe waterkoeling (DWC):** De oplossing biedt DWC van afzonderlijke systeemcomponenten, beschikbaar in de ThinkSystem-portfolio van Lenovo, evenals volledig gekoelde servers waarin DWC parallelle vloeistofstroom gebruikt over alle warmteproducerende componenten (CPU's, versnellers, DIMM's, netwerken, enz.) met een waterlus die tot 98% van de systeemwarmte kan opvangen en het stroomverbruik op systeemniveau en in het datacenter tot 40 procent kan verminderen.
- **Vloeistofondersteunde koeling:** Lenovo Neptune® vloeistofondersteunde koeling biedt de voordelen van vloeistof in een luchtgekoeld systeem. Traditionele luchtgekoelde systemen profiteren van vloeistofkoeling met speciaal ontworpen warmtebehandeling, dankzij een thermische overdrachtmodule (TTM) of een vloeistof-lucht-warmtewisselaar (L2A), en dat alles zonder extra leidingen. Dit ondersteunt klanten die ervoor kiezen om de status quo voor luchtgekoelde datacenters te behouden zonder concessies te doen voor de introductie van vloeistofkoeling.
- **Waterkoeling van de rack:** de achterdeurwarmtewisselaar (Rear-Door Heat Exchanger of RDHX) levert 100% warmteafvoerefficiëntie zonder dat er bewegende componenten of stroom nodig zijn. RDHX werkt met standaard luchtgekoelde servers, opslag en netwerken zonder aanpassingen, wat de integratie in bestaande datacenterinfrastructuur vergemakkelijkt en voordeel haalt uit CDU's in racks die meer efficiëntie bieden ten opzichte van traditionele CRAC-eenheden (Computer Room Air Conditioning).

We merken dat de Lenovo Neptune®-technologie nuttig is gebleken bij het ondersteunen van use cases met topprioriteit, zoals engineering, modellering, simulatieworkloads en de productie van animatiefilms, die allemaal rekenkracht met piekprestaties vereisen. Lenovo Neptune® is bijvoorbeeld doorslaggevend in belangrijke sectoren zoals fintech, computerondersteund ontwerpen en computational fluid dynamics (CAE/CFD), elektronische ontwerpautomatisering (EDA), weer- en klimaatmodellering en -voorspelling, bio-informatica, geospaatial/energie, aardwetenschappelijk onderzoek en animatie renderfarms. De lange geschiedenis van Lenovo in HPC heeft hen goed gepositioneerd voor het aanpakken van de toekomst van AI en alle stroomafwaartse uitdagingen op het gebied van stroom, koeling en duurzaamheid die gepaard gaan met uitbreidingen van datacenters om AI te ondersteunen.

Bovendien heeft de Lenovo Neptune® vloeistofkoelingsoplossing de 2024 Business Intelligence Group Sustainability Product of the Year award gewonnen en is het CRN's 2024 Best Green Energy Product of the Year, winnaar van de 2024 SEAL Sustainable Product Award en HPCwire's Best HPC Server Product of Technology voor 2023, wat zijn ESG-certificaten voor het hele ecosysteem verder versterkt.

De Neptune® vloeistofkoelingsoplossingen profiteren enorm van het inzicht van Lenovo in partnerschappen, zoals wordt geïllustreerd door het partnerschap met NVIDIA. Dit omvat nieuwe uitgebreide diensten die worden aangedreven door NVIDIA via het Lenovo AI Center of Excellence, gericht op GenAI, dat is geïdentificeerd als de wereldwijde topprioriteit van zakelijke en IT-besluitvormers op het gebied van technologie-investeringen. De Lenovo ThinkSystem SR780a-server gebruikt bijvoorbeeld Neptune® vloeistofkoeling om een ultra-efficiënte Power Usage Effectiveness (PUE, wat de verhouding is tussen het totale stroomverbruik van een datacenter en het stroomverbruik van alleen de IT-apparatuur) van 1,1 te bereiken. Door directe waterkoeling van CPU's, GPU's en NVIDIA NVSwitch-technologie te implementeren, kan dit systeem maximale prestaties leveren zonder thermische grenzen te bereiken.

De Lenovo AI Fast Start-service levert live oplossingen om de zakelijke, operationele en technologische resultaten van GenAI te presenteren. Bedrijven kunnen AI snel opschalen en een boost geven met behulp van full-stack NVIDIA-gebaseerde technologieën via Lenovo AI Fast Start voor NVIDIA AI Enterprise, terwijl de nieuwe Lenovo AI Fast Start voor NVIDIA NIM-inferentie microservices ontwikkelaars voorzien van eenvoudig te beheren, gecontaineriseerde en geoptimaliseerde inferentie-engines voor NVIDIA AI Foundation-modellen die beschikbaar zijn via NVIDIA.





Deel 5: Conclusies en aanbevelingen

AI-toepassingen, gebaseerd op speciaal ontwikkelde versnellingstechnologieën, verbeteren de resultaten in veel sectoren aanzienlijk. Vloeistofkoelingstechnologie is in opkomst als een essentiële factor voor AI met de hoogste prestaties, en op basis van de analyse van Futurum is Lenovo Neptune® vloeistofkoeling de marktleider op het gebied van vloeistofkoeling. De meer dan tien jaar ervaring van Lenovo met vloeistofkoeling heeft geleid tot de beste ontwerpen, productie, levering en ondersteuning in hun klasse.

Vanuit ons perspectief is de waardepropositie van Lenovo Neptune® vloeistofkoeling grondig gevalideerd door de goedkeuring van klanten in productienetwerken. Geely Auto R&D heeft een HPC-platform geïmplementeerd op basis van Lenovo ThinkSystem SD650 V3-servers met Neptune®-waterkoeling, waardoor de prestaties met 35% zijn verbeterd en er een startbaan is gecreëerd voor meer O&O-innovatie. Geely heeft Lenovo gekozen om een on-premises HPC-cluster te implementeren om cloudkosten en complexiteit te vermijden. Deze verandering leverde onmiddellijke resultaten op, aangezien het nieuwe cluster de HPC-prestaties met 35% verhoogde en tegelijkertijd het energieverbruik met 1 miljoen kWh/jaar verminderde.

Met deze overwegingen maken we de volgende opmerkingen aan organisaties bij het evalueren van de Neptune®-oplossing van Lenovo om te voldoen aan hun vereisten voor vloeistofkoeling.

Holistische benadering. Lenovo Neptune® omvat directe warmwaterkoeling van volledige systemen, waterkoeling van afzonderlijke componenten en vloeistofondersteunde koeling voor luchtgekoelde systemen. Dit resulteert in een allesomvattende aanpak die prestaties en vermogen optimaliseert.

Concurrentievoordelen. De 12 jaar ervaring van Lenovo in het leveren van vloeistofkoelingsoplossingen op productieniveau, waaronder EPDM-slangen, water-eerst, op koper gebaseerde koellussen en het gebruik van gesoldeerde verbindingen, levert concurrentievoordelen op het gebied van ontwerp, productie en leveringsproces, algemene kwaliteit en ondersteuningsproces.

Duurzaamheid/ESG-ervulling. Lenovo Neptune® is gebouwd om aan te sluiten bij belangrijke duurzaamheidsdoelstellingen, zoals CO2-reductie, zonder dat dit ten koste gaat van de rekenkracht die organisaties prioriteit moeten geven bij de evaluatie van vloeistofkoelsystemen.

Belangrijke informatie over dit rapport

MEDEWERKERS

Steven Dickens

Technologisch hoofdadviseur | The Futurum Group

Ron Westfall

Directeur onderzoek | The Futurum Group

UITGEVER

Daniel Newman

CEO | The Futurum Group

VRAGEN

Neem contact met ons op als je dit rapport wilt bespreken. The Futurum Group zal snel antwoorden.

CITATEN

Dit artikel kan worden geciteerd door de geaccrediteerde pers en analisten, maar moet in de context worden geciteerd, met vermelding van de naam van de auteur, de titel van de auteur en 'The Futurum Group'. Niet-pers- en niet-analysten moeten vooraf schriftelijke toestemming krijgen van The Futurum Group voor eventuele citaten

LICENTIES

Dit document, inclusief eventueel ondersteunend materiaal, is eigendom van The Futurum Group. Deze publicatie mag niet worden gereproduceerd, verspreid of gedeeld in welke vorm dan ook zonder voorafgaande schriftelijke toestemming van The Futurum Group.

OPENBAARMAKING

De Futurum Group levert onderzoek, analyse, advies en consultancy aan veel hightechbedrijven, waaronder die welke in dit artikel worden genoemd. Er zijn geen werknemers van het bedrijf die aandelen bezitten in de bedrijven die in dit document worden genoemd.



OVER LENOVO EN NVIDIA

Lenovo brengt het nieuwe tijdperk van AI-aangedreven innovatie voor iedereen. Onze full-stack portfolio levert krachtige, flexibele en verantwoorde AI-oplossingen om sectoren te transformeren en individuen te versterken. We creëren een toekomst van slimmere AI voor iedereen. Bij Lenovo geloven we dat de toekomst van AI bestaat uit het naast elkaar bestaan van publieke en zakelijke AI. Lenovo brengt AI naar jou en je gegevens.

In samenwerking met NVIDIA worden hybride AI-oplossingen speciaal ontwikkeld door technische samenwerking om AI efficiënt naar klantgegevens te brengen, waar en wanneer gebruikers deze het meest nodig hebben. Dit bevordert de visie van Lenovo om AI voor iedereen mogelijk te maken en time-to-market ondersteuning te bieden van baanbrekende technologieën en architectuur voor de volgende generatie generatieve AI. Hybride oplossingen van Lenovo, die al zijn geoptimaliseerd voor het uitvoeren van NVIDIA AI Enterprise-software voor veilige, ondersteunde en stabiele productie-AI, bieden ontwikkelaars ook toegang tot NVIDIA-microservices, waaronder NVIDIA NIM's en NeMo Retriever.



OVER THE FUTURUM GROUP

[The Futurum Group](#) is een onafhankelijk onderzoeks-, analyse- en adviesbureau dat zich richt op digitale innovatie en marktverstorende technologieën en trends. Elke dag helpen onze analisten, onderzoekers en adviseurs bedrijfsleiders uit de hele wereld bij het anticiperen op tektonische verschuivingen in hun sectoren en het benutten van ingrijpende innovatie om een concurrentievoordeel in hun markten te behalen of te behouden.



CONTACTGEGEVENS

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

© 2024 The Futurum Group. Alle rechten voorbehouden.

