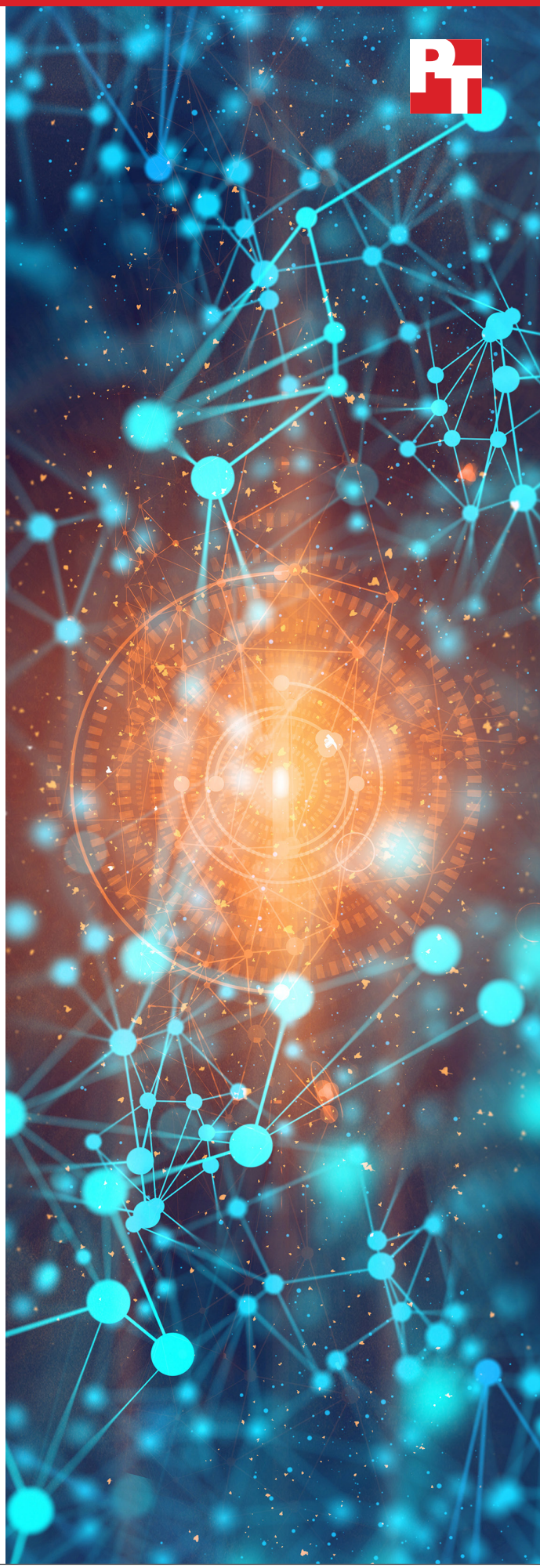
De weg vinden naar AI-succes met het Dell Ai-portfolio

Een vergelijking van het Dell AI-portfolio met het vergelijkbare aanbod van Supermicro

Kunstmatige intelligentie (AI) is de nieuwe heilige graal en is klaar om bedrijfsactiviteiten in verschillende sectoren te transformeren. Hoewel allerlei organisaties verkennen hoe ze AI kunnen gebruiken om hun bedrijfsactiviteiten te verbeteren, is het belangrijk om te onthouden dat de implementatie van AI en de voordelen ervan niet van de ene op de andere dag zijn gerealiseerd. Omdat elk bedrijf uniek is, moet elk bedrijf haar data en bedrijfsdoelen onder de loep nemen om te bekijken hoe het gebruik van AI met haar data de gewenste resultaten op kan leveren. Samenwerken met een bedrijf zoals Dell dat een uitgebreid AI-portfolio biedt, inclusief planning, datavoorbereiding, geschikte hardwareselectie, AI-modelontwerp, concepttests, referentiearchitecturen en end-to-end support, kan zich vertalen in succesvolle AI-projecten.

Met de vele opties beschikbaar op de markt, kan het vinden van een partner om te helpen bij deze beslissingen het verschil betekenen tussen een succesvolle AI-implementatie en een dure fout. In dit artikel bekijken we de AI-portfolio's van Dell en Supermicro nader om lezers te informeren over de voordelen die Dell kan bieden tijdens het AI-traject van een klant. We richten ons eerst op server- en computingopties, waarvan beide bedrijven een ruim en gevarieerd aanbod hebben. Vervolgens bekijken we hoe Dell verder gaat dan alleen hardwareoverwegingen voor bedrijven die op zoek zijn naar training, planningsservices, partnerecosystemen en meer.



Servers en prestatieresultaten voor AI-workloads

Servers, de basisinfrastructuur voor computing die AI-workloads aandrijven, kunnen CPU's, GPU's of beide gebruiken als computingresources, afhankelijk van de grootte of het type workload. Voor grotere of veeleisende workloads zoals HPC of AI bieden GPU's topprestaties. GPU's zijn verkrijgbaar in verschillende modellen, waaronder universele PCIe, Open Compute Project Accelerator module (OAM) en bedrijfseigen NVIDIA SXM architectuur, die momenteel topprestaties biedt.[1] Grote geheugencapaciteit en serverontwerpfuncties, zoals koelingsarchitectuur en energie-efficiëntie, hebben ook invloed op de prestaties. De meeste datacenters gebruiken nog steeds luchtkoeling, wat betekent dat AI-workloads servers nodig hebben die zo effectief mogelijk met lucht kunnen koelen. Hieronder belichten we het aanbod van Dell PowerEdge-servers op het gebied van componenten, koelingsopties en meer, samen met hun gepubliceerde MLCommons® MLPerf®-scores.

Testresultaten

MLPerf® is een benchmarksuite die AI-prestaties test voor zowel training als inferentie. Als een organisatie officiële MLPerf®-resultaten wil publiceren, moeten de resultaten voldoen aan de specifieke voorwaarden die zijn vastgesteld door de benchmarkontwikkelaar, MLCommons®.[2] Deze nalevingsrichtlijnen bieden standaarden die het makkelijker maken om prestaties te vergelijken. Voor inferentietests gebruikt MLPerf® de datasets Datacenter, Edge, Mobile en Tiny, en worden AI-scores en watt aan energie gepubliceerd die tijdens het testen zijn verbruikt. De benchmarksuite voor inferentie bevat tests voor veel gangbare AI-, ML- en DL-modellen (zie tabel 1).

Tabel 1: AI-, ML- en DL-modellen inbegrepen in MLPerf®, met tests en typische gebruiksscenario's voor elk model. Bron: Principled Technologies.

Gangbare AI-modellen	Typische gebruiksscenario's
ResNet	Een beeldclassificatiemodel waarmee computers verschillende beelden kunnen leren, onthouden en identificeren voor gebruiksscenario's zoals medische beeldvorming, contentmoderatie op sociale media en gezichtsherkenning
RetinaNet	Een type objectdetectie dat meer complexiteit kan verwerken dan ResNet. Het helpt computers om objecten in afbeeldingen of videoframes te identificeren en te lokaliseren en kan ze classificeren op belangrijkheid. Gebruikt voor zaken zoals autonoom rijden, auto-assist-technologie in voertuigen, bewaking, gezichtsherkenning
3D-UNet	Specifiek voor segmentatie van medische beelden
RNN-T	Spraakherkenning voor gebruiksscenario's zoals geautomatiseerde vertalingen
BERT	Natuurlijke taalverwerking voor gebruiksscenario's zoals tekstsamenvattingen, vertalingen en automatische voltooiing van taken
DLRM-v2-99.9	Aanbevelingsmodel voor gebruiksscenario's zoals gerichte advertenties en gepersonaliseerde productaanbevelingen
GPTJ-99 en 99.9	LLM voor natuurlijke taalverwerking dat uitblinkt bij het genereren van tekst voor gebruiksscenario's zoals chatbots en chatgebaseerde AI-tools

Over MLPerf

MLPerf®-resultaten bevatten verschillende parameters naast de AI-modellen zelf, wat ertoe kan leiden dat er veel data in één grafiek of tabel worden geparseerd. Hier volgt een beknopt overzicht van deze parameters:

- 99.0 en 99.9: Deze cijfers verwijzen naar de nauwkeurigheid waarmee het model is getraind. Hoe nauwkeuriger de uitvoer moet zijn, hoe complexer het model en hoe langer het kan duren om data te verwerken.
- Offline samples/sec: Modus waarbij de benchmark alle query's aan het begin van de test verzendt waarmee data worden simuleert die al in het systeem aanwezig zijn.
- Server queries/sec: Modus waarbij de benchmark query's verzendt tijdens de hele duur van de test, waarbij de analyse van een live datastroom wordt gesimuleerd.

Zie https://mlcommons.org/benchmarks/inference-datacenter/ voor meer informatie over de resultaten van MLCommons® en MLPerf®.

De resultaten in dit rapport zijn afkomstig van MLPerf® v3.1 Inference Datacenter-resultaten die in november 2023 op de MLCommons®-website zijn gepubliceerd.[3] Deze resultaten bevatten inzendingen van technologiefabrikanten en cloudserviceproviders en gaan over verschillende configuraties. Vergeleken met openbaar beschikbare inzendingen van Supermicro, leverden Dell PowerEdge-servers vergelijkbare resultaten op. Tabel 2 bevat servergegevens.

Tabel 2: Dell- en Supermicro-servers verwerkt in de vanaf november 2023 gepubliceerde resultaten van MLCommons® MLPerf® 3.1. Bron: Principled Technologies.

Indiener	Servermodel	aantal en model van GPU's	Omschrijving
Dell[4]	PowerEdge XE9680	8 x NVIDIA H100 SXM	Voor AI-training en -inferentie met grote workloads, zoals grote taalmodellen
	PowerEdge XE9640	4 x NVIDIA H100 SXM	Voor het trainen van grote AI-modellen in datacenters met een hoge dichtheid en vloeistofkoeling
	PowerEdge XE8640	4 x NVIDIA H100 SXM	Voor traditionele AI-training, HPC en data analyse-apps in een 4U-model voor luchtgekoelde datacenters
Supermicro[5]	AS-8125GS-TNHR	8 x NVIDIA H100 SXM	Voor grootschalige AI-training en HPC-workloads met AMD-processors
	SYS-821GE-TNHR	8 x NVIDIA H100 SXM	Voor grootschalige AI-training en HPC-workloads met Intel-processors
	SYS-421GU-TNXR	4 x NVIDIA H100 SXM	Modulair ontwerp voor flexibiliteit ter ondersteuning van HPC- en AI-workloads

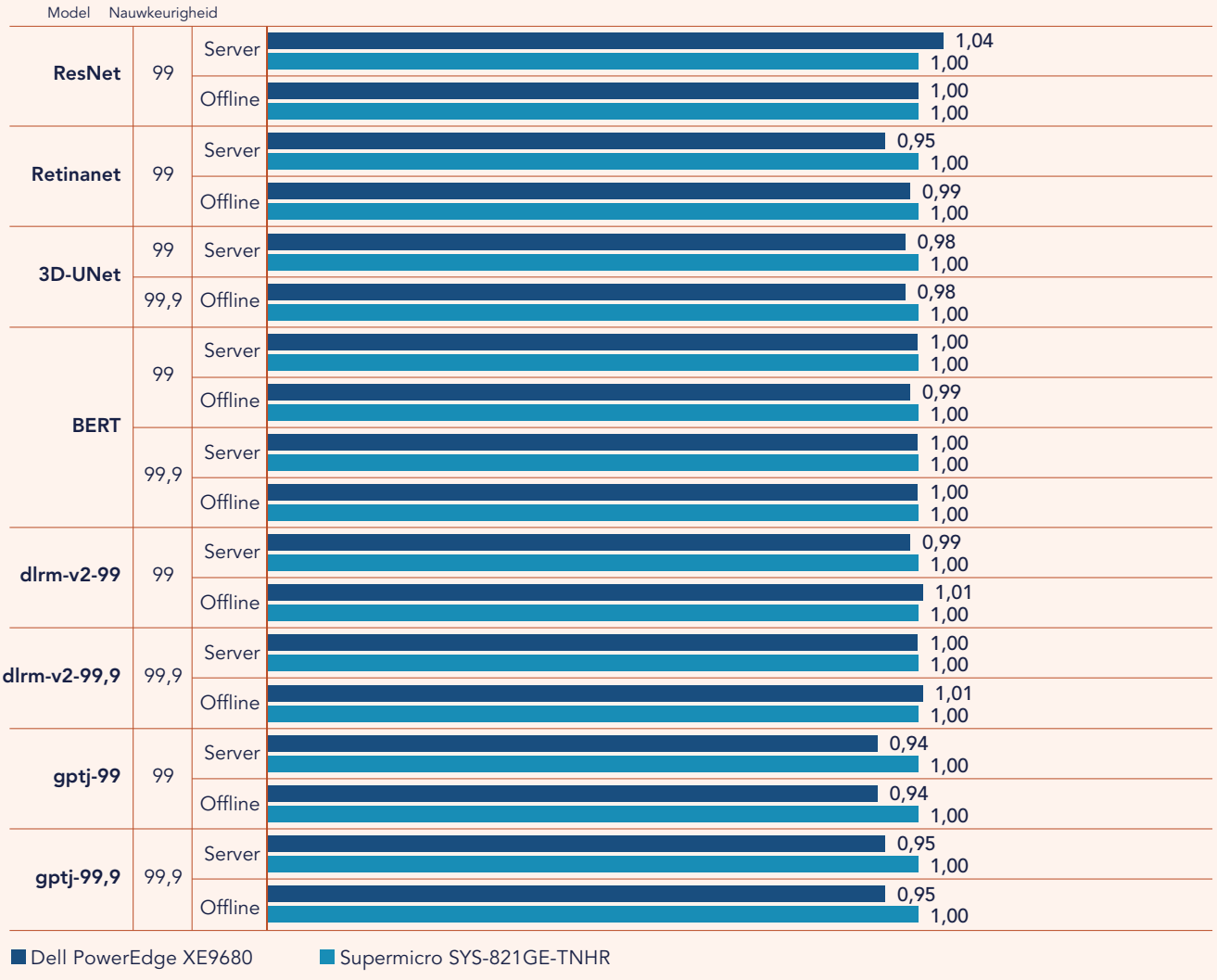
Omdat Dell en Supermicro resultaten hebben ingediend met apples-to-apples GPU-configuraties, zijn prestaties eenvoudig te vergelijken. Zoals uit afbeelding 1 en 2 blijkt, leiden de vrijwel gelijke configuraties van beide leveranciers tot resultaten die grotendeels met elkaar overeenkomen. Voor andere configuraties, zoals die in afbeelding 3 en 4, presteerde Dell beter dan Supermicro met het gptj-99.9-model tijdens de 4-GPU-tests. Dell heeft resultaten ingediend voor alle beschikbare modellen met alle drie de servers, terwijl Supermicro dit niet heeft gedaan. We vergelijken alleen de modellen waarvoor voor beide servers resultaten zijn. Ga naar de MLCommons® MLPerf®-resultaten voor alle Dell-resultaten.

Resultaten met acht GPU's

De Dell PowerEdge XE9680 biedt ondersteuning voor maximaal acht NVIDIA H100 SXM5-GPU's voor AI-versnelling en maximaal twee 4e generatie Intel® Xeon® schaalbare processors. De PowerEdge XE productlijn heeft een modulaire architectuur die SXM4 of SXM5 NVIDIA-GPU's of Open Compute Project Accelerator module (OAM) GPU-constructies ondersteunt, waarmee de prestaties kunnen verbeteren vergeleken met een standaard PCIe-GPU. De Dell PowerEdge XE9680 biedt ook de AMD Instinct™ MI300X versneller.[6] De PowerEdge XE9680 neemt slechts 6U aan rackruimte in en is een compacte acht-richtings NVIDIA H100 SXM5-server.

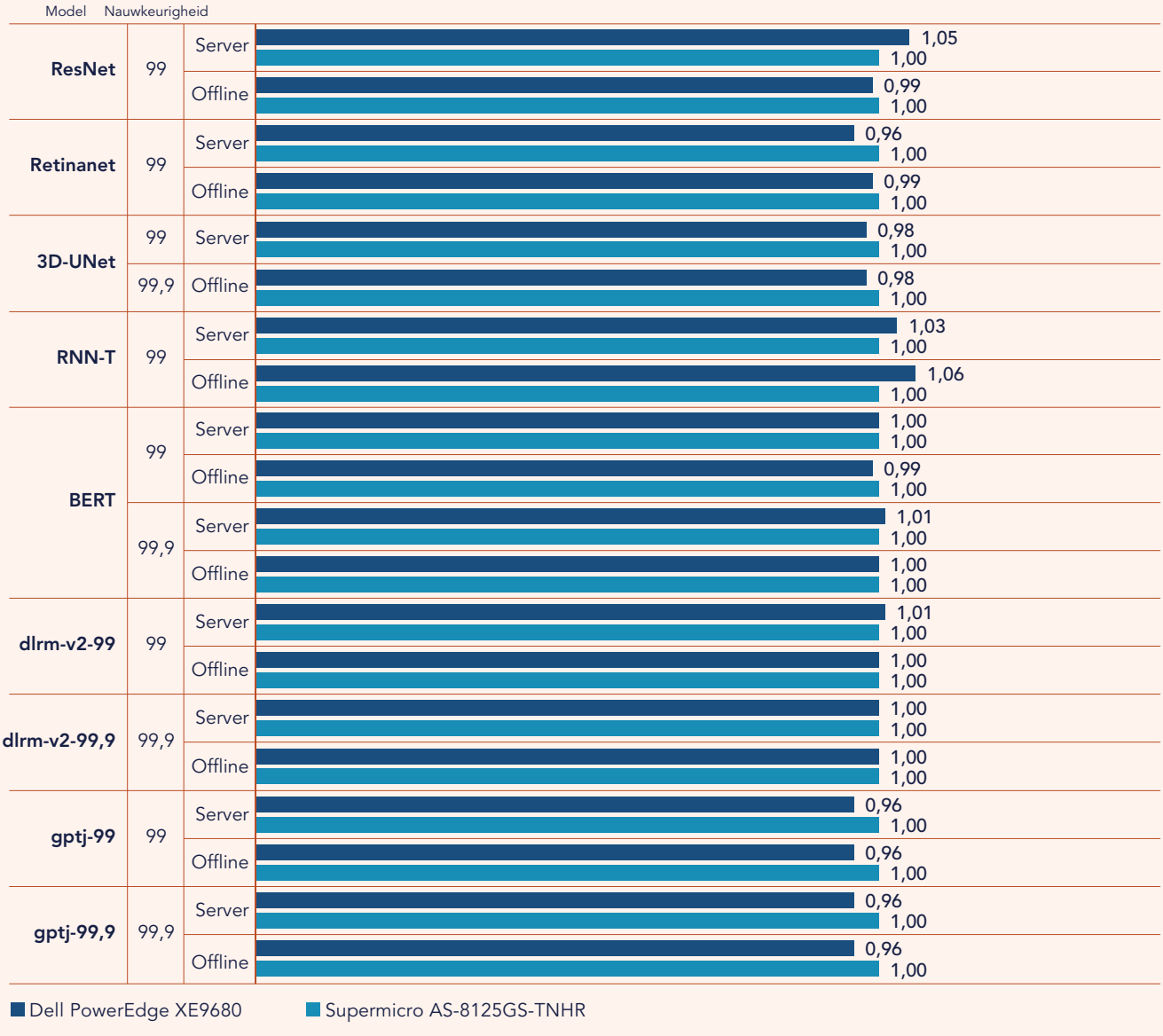
De 8-GPU Supermicro-server vereist daarentegen 33 procent meer rackruimte bij 8U en biedt ook het SXM-model voor NVIDIA GPU's. Het verschil in grootte betekent dat u zeven Dell PowerEdge XE9680-servers in een rack kunt plaatsen en slechts vijf Supermicro-servers. In afbeelding 1 en 2 vergelijken we de resultaten van de Dell PowerEdge XE9680 met die van twee configuraties van de Supermicro 8-GPU server: de SYS-821GE-TNHR met Intel-processors en de AS-8125GS-TNHR met AMD-processors. Supermicro heeft echter geen resultaten ingediend voor RNN-T op de SYS-821GE-TNHR, dus sluiten we dit model uit van de grafiek in afbeelding 1.

Genormaliseerde MLPerf®-resultaten: Dell PowerEdge XE9680 met H100 SXM5 versus Supermicro SYS-821GE-TNHR met H100 *(groter is beter)*



Afbeelding 1: Gepubliceerde MLPerf®-resultaten voor de Dell PowerEdge XE9680 en Supermicro SYS-821GE-TNHR per 29-11-2023. Beide systemen gebruiken het SXM-model van de NVIDIA H100 GPU. Bron: Principled Technologies met gegevens van MLCommons®.[7,8]

Genormaliseerde MLPerf®-resultaten: Dell PowerEdge XE9680 met H100 SXM5 versus Supermicro AS-8125GS-TNHR met H100 SXM5 *(Larger is better)*



Afbeelding 2: Gepubliceerde MLPerf®-resultaten voor de Dell PowerEdge XE9680 en Supermicro AS-8125GS-TNHR per 29-11-2023. Beide systemen gebruiken het SXM-model van de NVIDIA H100 GPU. Bron: Principled Technologies met gegevens van MLCommons®.[9,10]

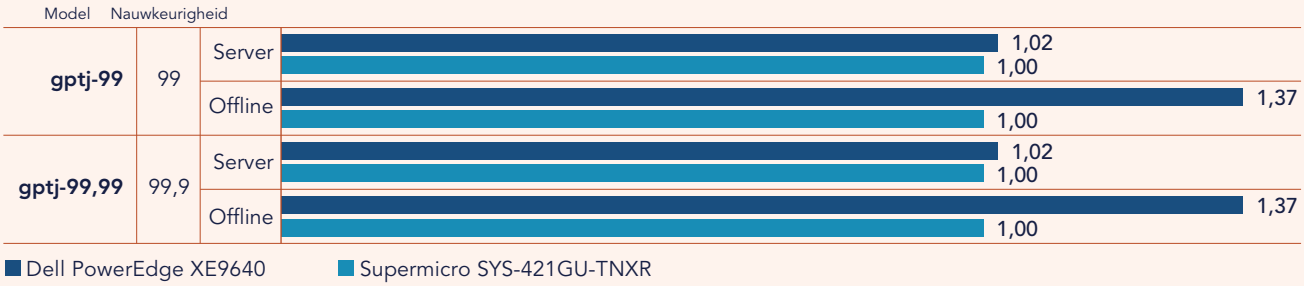
Zoals uit deze resultaten blijkt, biedt de Dell PowerEdge XE9680 vergelijkbare prestaties voor een groot aantal workloads met AI-inferentie en neemt deze minder datacenterruimte in beslag.

Resultaten met vier GPU's

Wanneer het belangrijk is energie of ruimte te besparen in het datacenter, kan de 2U Dell PowerEdge XE9640 het antwoord zijn. Met tot wel vier NVIDIA H100 SXM-GPU's biedt de PowerEdgeXE9640 de helft van de GPU-rekenkracht van de PowerEdge XE9680 en neemt één derde van de ruimte in beslag.[11] De compacte Dell PowerEdge XE9640 bevat Dell Smart Cooling-technologie, die verschillende thermische technologieën biedt, waaronder directe vloeistofkoeling voor CPU's en GPU's.[12] Het 2U-chassis van de PowerEdge XE9640 is geschikt voor verbeterde luchtstroommechanismen, waaronder grotere ventilatoren en koelplaten, om de andere essentiële componenten, zoals PCIe-kaarten en geheugen, te koelen.[13]

Supermicro biedt een oudere SYS-220GQ-TNAR+ server die vier NVIDIA A100 HGX GPU's in een 2U-model biedt, maar we konden geen 2U Supermicro-servers vinden met vier van de recentere H100 HGX GPU's die overeenkomen met de PowerEdge XE9640.[14] De 4-GPU-server met HGX H100 NVIDIA GPU's van Supermicro in de MLPerf®-inzendingen is de SYS-421GU-TNXR, een 4U-server. Zoals eerder gezegd, heeft Supermicro MLPerf® 3.1 resultaten ingediend voor alleen de SYS-421GU-TNXR op het gptj-99.9 AI-model, dus kunnen we deze niet vergelijken met de PowerEdge XE9640 op de andere modellen. In de gepubliceerde resultaten presteerde de PowerEdge XE9640 echter beter dan de Supermicro-server in de offline tests en behaalde tot 1,37 keer de score (zie afbeelding 3).

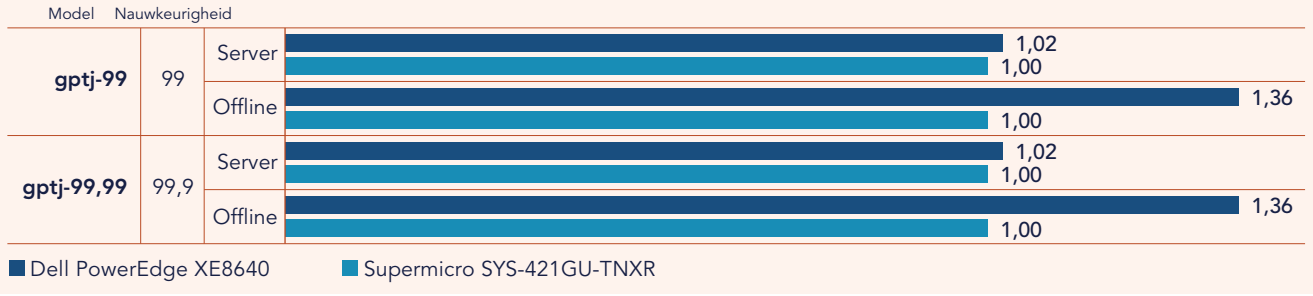
Genormaliseerde MLPerf® Results: Dell PowerEdge XE9640 met H100 SXM5 versus Supermicro SYS-421GU-TNXR met H100 SXM5 *(groter is beter)*



Afbeelding 3: Gepubliceerde MLPerf®-resultaten voor de Dell PowerEdge XE9640 en Supermicro SYS-421GU-TNXR per 29-11-2023. Beide systemen gebruiken het SXM-model van de NVIDIA H100 GPU. Bron: Principled Technologies met gegevens van MLCommons®.[15,16]

We kunnen de resultaten van de Supermicro SYS-421GU-TNXR-server ook vergelijken met de Dell PowerEdge XE8640, een 4U-server met 4 GPU's die ook NVIDIA H100 HGX GPU's ondersteunt. Hoewel de PowerEdge XE9640 groter is dan de PowerEdge XE8640, heeft deze geen directe vloeistofkoeling nodig, waardoor het een compromis is tussen dichtheid- en koelingstechnologieën voor datacenters die geen toegang hebben tot waterkoeling. De PowerEdge XE8640 biedt een luchtkoeling voor processors en een Liquid Assisted Air Cooling Radiator voor de GPU's, waardoor er geen water-naar-rack-beschikbaarheid nodig is.[17] De Dell PowerEdge XE8640 is uitgerust met de nieuwste 4e generatie Intel Xeon schaalbare processors en tot 4 TB geheugen om de grote datasets en complexe berekeningen aan te kunnen die veel voorkomen bij AI en data-analyse.[18] Met zijn 4U-model is de PowerEdge XE8640 vergelijkbaar met de Supermicro SYS-421GU-TNXR wat betreft dichtheid en GPU-mogelijkheden. Zoals we echter ook zagen met de PowerEdge XE9640, behaalde de Dell PowerEdge XE8640 betere gptj-99 scores in de offline tests dan de Supermicro-server (afbeelding 4).

Genormaliseerde MLPerf®-resultaten: Dell PowerEdge XE8640 met H100 SXM5 versus Supermicro SYS-421GU-TNXR met H100 SXM5 *(groter is beter)*



Afbeelding 4: Gepubliceerde MLPerf®-resultaten voor de Dell PowerEdge XE8640 en Supermicro SYS-421GU-TNXR per 29-11-2023. Beide systemen gebruiken het SXM-model van de NVIDIA H100 GPU. Bron: Principled Technologies met gegevens van MLCommons®.[19,20]

Zoals we hebben gezien, zijn de MLPerf®-prestatieresultaten van het Supermicro- en Dell GPU-serveraanbod over het algemeen vergelijkbaar, met een opvallend voordeel voor de Dell PowerEdge XE8640- en PowerEdge XE9640-servers op één AI-model.

Omdat prestaties slechts één aspect van AI-implementatie zijn, hebben we ook gekeken naar de rest van het AI-aanbod van Dell en Supermicro, van workstations, storage en netwerken tot services, support, training en meer. We hebben ontdekt dat het Dell AI-portfolio ruimer is dan het Supermicro-aanbod en oplossingen biedt voor veel van de obstakels waarmee bedrijven worden geconfronteerd, naast computingprestaties.

Workstations en storage-aanbod voor klanten

Extra computingopties met workstations

Sommige AI-gebruiksscenario's vereisen een andere computingmethode en niet iedereen die toegang vereist tot AI-compatibele hardware kan gebonden blijven aan het datacenter. Wetenschappers die in laboratoria werken, hebben mogelijk geen ruimte voor een rack met servers en het is onpraktisch voor mensen die in de buitendienst werken om met een groot, zwaar desktopsysteem te moeten slepen. Dat is waar workstations om de hoek komen kijken. Het Dell AI-portfolio bevat verschillende AI-ready Precision workstations, waaronder tower-, mobiele en rackconfiguraties, om aan verschillende behoeften te voldoen.[21] Supermicro biedt daarentegen geen mobiele workstations voor eenvoudige toegang onderweg. Het aanbod GPU-workstations bestaat uit verschillende towerconfiguraties, waarvan Supermicro beweert dat gebruikers deze in een rack kunnen monteren.[22] Uit ons onderzoek blijkt dat er twee workstations met rackmontage zijn, maar ze lijken ouder te zijn en alleen beschikbaar met NVIDIA A100 GPU's en mogelijk niet meer te koop.[23] Als uw bedrijf flexibiliteit nodig heeft wat betreft het type en de mobiliteit van te implementeren GPU-workstations, is het Dell AI-portfolio veel beter voor u.

Overwegingen bij storage

Storage kan net zo belangrijk zijn als computing voor het uitvoeren van AI-workloads. Meer data verbetert de nauwkeurigheid van AI-modellen, maar het opslaan en beheren van enorme datasets is niet altijd mogelijk bij veel datacenters. Omdat modellen doorgaans worden getraind met behulp van niet-gestructureerde data moeten AI-ready storagesystemen bovendien veel verschillende datatypen met gemak kunnen verwerken.[24] Om capaciteit en schaalbaarheid te bieden voor AI-, ML- en DL-datasets, biedt Dell de PowerScale™ serie voor bestandstorage en Elastic Cloud Storage (ECS) of softwaregedefinieerde ObjectScale voor objectstorage.

Het Dell PowerScale all-flash NAS portfolio biedt capaciteitsopties variërend van 3,84 TB tot 720 TB onbewerkte capaciteit per knooppunt, met geclusterde all-flash-capaciteiten met een onbewerkte capaciteit van wel 186 PB. De flexibiliteit en schaal van PowerScale werken voor een breed scala aan klanten en gebruiksscenario's voor AI.[25] Alle drie de all-flash PowerScale modellen (F200, F600 en F900) bevatten inline datacompressie en deduplicatie om de storage-efficiëntie te verbeteren.[26] Elk PowerScale storagemodel gebruikt het Dell OneFS™-bestandssysteem, dat geavanceerd tieringbeleid gebruikt om te zorgen dat de meest gebruikte data zich op de best presterende storagelagen bevinden.[27] Dell biedt ook OneFS-software in de Amazon Web Services (AWS) marketplace met APEX File Storage voor AWS. Klanten kunnen OneFS gebruiken met hun AWS-computinginstanties voor een consistente gebruikerservaring met dezelfde functies die beschikbaar zijn in on-premise OneFS-arrays.[28]

Het storage-aanbod van Supermicro bestaat uit storageservers: Rackservers met hoge storagedichtheid in verschillende groottes en dichtheid.[29] Voor bestandsstorage van Supermicro kunnen klanten kiezen uit verschillende softwaregedefinieerde storageproducten van derden, zoals WekaIO, Scality RING of OSNEXUS.[30] Hoewel Scality RING en OSNEXUS opties voor bestandsstorage bieden als onderdeel van hun platformbeschrijvingen, lijkt WekaIO de primaire optie voor klanten op zoek naar basisstorage van bestanden. Supermicro biedt verschillende referentiearchitecturen voor veel verschillende gebruiksscenario's, maar klanten moeten een WekaIO-softwareabonnement of -licentie hebben, wat de totale kosten van de oplossing kan verhogen.[31]

Opties voor objectstorage van Dell bevatten Dell ECS Enterprise Object Storage, dat "speciaal is gemaakt om niet-gestructureerde data op public cloud-schaal op te slaan".[32] Samen met ingebouwde compatibiliteit met Amazon S3-objectstorage voor hybrid cloud-functionaliteit leveren ECS storageknooppunten capaciteiten tot 14 PB per rack.[33] Zoals bij het opslaan van bestanden, vereisen de producten voor objectstorage van Supermicro configuraties van derden. Het OSNEXUS-platform is een gecombineerd storageplatform voor bestands-, blok- en objectstorage, terwijl de Scality RING-oplossing bestands- en objectstorage combineert. Beide vereisen licenties van de externe leveranciers.[34][35] Voor alleen objectstorage kunnen klanten de Supermicro-oplossing voor Quantum ActiveScale voor private cloud-objectstorage aanschaffen met een Quantum-softwareabonnement.[36] Op het moment van publiceren konden we geen flexibele consumptie-/operationele opties voor uitgaven vinden die Supermicro aanbiedt.

Omdat voor het Supermicro-storageaanbod klanten in contact moeten komen met externe softwareleveranciers, moeten ze waarschijnlijk extra licentie- of abonnementskosten betalen en kunnen ze problemen ondervinden bij support, probleemoplossing en meer. Het Dell AI-portfolio biedt storageklanten van Dell één betrouwbare service- en supportoplossing voor ieder aspect van hun storageoplossing.

Over Dell APEX

Voor klanten die hun bestandsstorage als service willen gebruiken, biedt Dell APEX Data Storage Services, waaronder bestands-, blok- en back-upopslag. Met de Dell APEX Console kunnen klanten nieuwe abonnementen bestellen, storagecapaciteit aanpassen en monitoren, en meer. Volgens Dell kunt u met deze oplossing "het gemak en de flexibiliteit van de cloudervaring krijgen met meer controle over uw applicaties en data".[37]

Ga voor meer informatie over Dell APEX naar https://www.dell.com/en-us/dt/apex/storage/data-storage-services/index.htm

Netwerkopties

Netwerken vormen ook een essentieel onderdeel van de AI-infrastructuur. Met veel AI-workloads die draaien op grote clusters van servers, waarbij constante communicatie tussen servers en storage nodig is, hebben AI-workloads robuuste netwerken nodig om knelpunten te vermijden. Als uw netwerk onvoldoende capaciteit biedt voor de AI-workload, zullen training- en inferentietijden toenemen, wat de verwerking van data en de tijd tot inzichten vertraagt. Dell biedt PowerSwitch Data Center top-of-rack (Tor) switches en PowerEdge MX I/O-modules voor Ethernet- en fabric-netwerken.[38] Het PowerSwitch-aanbod varieert van 1 GbE tot 400 GbE om aan verschillende behoeften te voldoen. Bovendien bieden de switches uit de Dell PowerSwitch Z-serie 100 GbE- en 400 GbE-verbindingen die zijn geoptimaliseerd voor leaf/spine fabrics.[39]

Supermicro biedt ook switches met Ethernet-poorten tot 400 GbE voor Tor en andere applicaties zoals Data Center spine en leaf.[40] Dell-netwerkservices bieden echter verschillende gebruiksvriendelijke en flexibele voordelen die Supermicro niet biedt. Services zoals het Dell Fabric Design Center kunnen helpen netwerkfouten, gaten of inefficiënties te voorkomen, zodat klanten netwerkfabrics kunnen plannen en implementeren met automatisering.[41] Voor specifieke omgevingen zoals VMware VxRail, VMware ESXi en Dell PowerStore-configuraties biedt Dell SmartFabric Services, die softwaregedefinieerde infrastructuurimplementatie en levenscyclusbeheer mogelijk maken. Met PowerStore kunnen SmartFabric services "tot 99% van de LAN-connectiviteitstaken automatiseren met een plug-and-play fabric."[42] Services zoals deze, die automatisering, begeleiding en meer bieden voor netwerkontwerp en -implementatie, ondersteunen klanten bij het navigeren door het AI-acceptatieproces.

Services, training en meer

De primaire niet-hardware-uitdaging bij de implementatie van AI is de behoefte aan interne expertise voor strategie, planning, datavoorbereiding en beheer. AI-workloads beheren en onderhouden vereist specifieke kennis, waaronder meer traditionele kennis over hardware en machine learning-activiteiten en datawetenschap. Mensen die de AI-strategie ontwerpen en implementeren, moeten ook inzicht hebben in de unieke operationele doelen van het bedrijf om te zorgen dat de nieuwe AI-workloads deze doelen bereiken.[43]

Een andere aanzienlijke hindernis kan de naadloze integratie van AI in bestaande operationele systemen zijn. Deze integratie vereist een strategische afstemming van nieuwe AI-technologieën op de huidige bedrijfsprocessen, zodat de introductie van AI bestaande workflows intact houdt. Samenwerken met een bedrijf zoals Dell dat verschillende geoptimaliseerde, gevalideerde referentiearchitecturen voor oplossingen, trainingen, beheeropties en een groot partnerecosysteem biedt, kan uw AI-acceptatietraject vereenvoudigen.

Professionele services voor AI

Voor training en planning biedt Dell verschillende services specifiek voor AI.[44] Dell-services die de implementatie van AI ondersteunen bestaan uit consulting, datavoorbereiding, implementatie, support en training, elk gericht op specifieke aspecten van de overstap naar AI. Dell Advisory services voor generatieve AI helpen klanten een plan op te stellen waarin gebruiksscenario's worden geïdentificeerd en bedrijven worden geholpen hun processen te stroomlijnen.[45] Op dezelfde manier bieden Adoption services voor generatieve AI-workshops met Dell professionals om uw behoeften en unieke uitdagingen te beoordelen om een vooraf getraind model voor uw bedrijf te bepalen en sessies voor kennisoverdracht te organiseren om uw IT-personeel te trainen.[46] Dell biedt ook Implementation, Scale en Managed services voor generatieve AI die verschillende support- en trainingsniveaus bieden, tot een volledig beheerde AI-infrastructuur, zodat uw IT-personeel zich kan richten op de modellen en data, terwijl Dell de hardware beheert.[47] ProSupport services zorgen voor optimale systeemprestaties en bieden essentiële hardware- en softwareondersteuning voor lopende AI-activiteiten, waarbij technische problemen worden opgelost.[48]

Educational services zijn essentieel voor het bevorderen van de benodigde vaardigheden en kennis voor AI-gebruik. Het Dell-trainingsaanbod bestaat uit uitgebreide trainingsprogramma's op het gebied van datawetenschappen, certificeringen voor geavanceerde analyse en workshops over specifieke AI-technologieën, zoals machine learning.[49]

De services van Supermicro zijn daarentegen meestal beperkt tot probleemoplossing, handleidingen, RMA's (Return Merchandise Authorizations) en garanties.[50] We hebben geen ontwerp-, implementatie-, beheer- of educatieve services aangetroffen in het Supermicro AI-portfolio. Voor bedrijven op zoek naar een trainingspartner voor hulp bij de complexiteit van AI-acceptatie, is Dell de beste keuze.

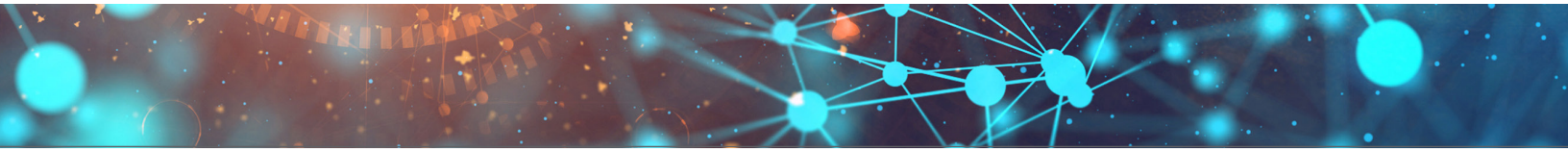
Samenwerkingen met derden voor AI-workloads

Dell Technologies en NVIDIA werken samen om Dell gevalideerde ontwerpen te bieden, die een uitgebreide oplossing voor generatieve AI in bedrijfsomgevingen willen bieden. Dit project creëert een schaalbare, krachtige infrastructuur op basis van technologieën en software van Dell en NVIDIA, in combinatie met een AI-modelframework waarmee bedrijven aangepaste AI-modellen kunnen bouwen en uitvoeren. Met de oplossing kunnen klanten snel en gemakkelijk GenAI-workloads opzetten.[51] Lees voor meer informatie het gedeelte Dell gevalideerde ontwerpen hieronder.

Dell werkt samen met verschillende bedrijven om applicaties voor AI-technologie te verbeteren. Samen met Hugging Face maakt Dell het eenvoudiger om grote taalmodellen (LLM's) op locatie in te stellen. Deze samenwerking combineert expertise op het gebied van AI met Dell-servers en storagesystemen. Een Dell-specifieke Hugging Face-portal biedt tools voor eenvoudige, veilige implementatie van open source AI-modellen voor Hugging Face. Het continue doel is om deze modellen voor Dell-systemen te blijven verbeteren, de prestaties te verbeteren en nieuwe AI-applicaties te ondersteunen.[52]

Dell en Starburst werken aan een krachtig, schaalbaar data lake dat Starburst-analyses integreert met Dell computing- en storagetechnologie, op zoek naar één toegangspunt tot alle gegevensbronnen voor AI- en ML-tools. Klanten kunnen profiteren van deze samenwerking om datasilo's te helpen elimineren.[53]

Volgens ons onderzoek heeft Supermicro veel beperktere samenwerkingen voor AI. SiMa.ai en Supermicro hebben samengewerkt aan de ontwikkeling van de Supermicro SYS-E300-13AD, een compacte Edge ML-server ontworpen voor de verwerking van multi-stream video-analyse. Deze server, uitgerust met SiMa.ai ML pipeline op een chip, verwerkt efficiënt meerdere videokanalen, verlaagt de totale eigendomskosten en verbetert de betrouwbaarheid en beveiliging. De server biedt computing ontworpen voor de verwerking en analyse van tal van videostreams, en levert edge-intelligentie die geschikt is voor verschillende bedrijfsapplicaties.[54]



Door Dell gevalideerde ontwerpen

Om het giswerk uit AI-hardwareoplossingen te halen, biedt Dell in het lab gevalideerde referentiearchitecturen die zijn geoptimaliseerd voor verschillende AI- en andere workloads. Deze gevalideerde ontwerpen bestaan uit architectuurconcepten, volledige oplossingsoverzichten, prestaties en andere labvalidaties die de mogelijkheden van de oplossing bewijzen voor de workload waarvoor deze is ontworpen. Deze workloads bestaan uit gevirtualiseerde omgevingen, MLOps, machine learning, conversationele AI, GenAI Inference, GenAI-modelafstemming, NVIDIA Fleet Command en OpenShift AI.[55]

Hierbij een voorbeeld: AI voor gevirtualiseerde omgevingen gevalideerd ontwerp combineert VMware-enabled AI met NVIDIA AI Enterprise op Dell infrastructuur, waardoor AI in virtuele instellingen wordt geoptimaliseerd.[56] De handleiding voor gevalideerd ontwerp bevat prestatieresultaten die ResNet-modeltraining tonen die aan klanten bewijst dat het ontwerp werkt en laat zien welke prestaties ze kunnen verwachten.[57] Deze validaties bieden klanten meer waarde dan alleen samenwerkende hardware te vermelden, door concepten uit te leggen, configuraties aan te bevelen en klanten te helpen met overwegingen en prestatieverwachtingen.[58]

Supermicro biedt oplossingen op basis van gebruiksscenario's, maar kan niet tippen aan het referentiearchitectuurniveau van de Dell gevalideerde ontwerpen. In plaats van een speciaal ontwikkelde oplossing aan te bevelen voor specifieke workloads, deelt Supermicro haar servers en GPU's in categorieën in, zoals AI-inferentie en -training, HPC/AI, visualisatie en ontwerp, en meer.[59] In hun brochures en datasheets zijn deze categorieën een paar servers en GPU's die ze aanbevelen als het meest geschikt voor de taak, verschillende gebruiksscenario's en lijsten met belangrijke gebruikte technologieën en softwaresuggesties.[60] In tegenstelling tot de Dell gevalideerde ontwerpen, lijken ze geen netwerkarchitecturen en prestatie- of validatiedata te bevatten. Supermicro biedt ook verschillende referentieontwerpen die een meer gedetailleerde referentiearchitectuur bieden voor sommige AI-oplossingen, zoals een grootschalige AI-training met een briefing voor een oplossing voor vloeistofkoeling die Supermicro samen met NVIDIA heeft gepubliceerd.[61] Klanten kunnen mogelijk een uitgebreidere referentiearchitectuur vinden voor een specifiek scenario, maar op het moment van ons onderzoek vonden we er slechts drie: De vloeistofkoelingsarchitectuur die we eerder hebben genoemd, een AI-workstationarchitectuur en een RedHat OpenShift-architectuur.[62]

Over het algemeen hebben we vastgesteld dat Dell gevalideerde ontwerpen meer AI-workloads dekten en meer uitgebreide begeleiding boden dan het Supermicro-aanbod.

Beheerservices en iDRAC

Volgens een rapport van Principled Technologies uit april 2023 biedt de Integrated Dell Remote Access Controller (iDRAC) verschillende geavanceerde functies via de Supermicro Intelligent Platform Management Interface (IPMI), met name op het gebied van automatisering, beveiliging en configuratie.[63] Tabel 3 toont een vergelijking van de beheerfuncties van Dell en Supermicro uit dat rapport, waaruit blijkt hoe iDRAC eenvoudigere implementatie, eenvoudigere firmware-updates en meer beveiligingsfuncties kan bieden dan Supermicro IPMI. Sommige resultaten kunnen zijn gewijzigd sinds de oorspronkelijke publicatie.

Tabel 3: Samenvatting van de Principled Technologies vergelijking uit april 2023 van de beveiligingsfuncties in de beheertools van Dell en Supermicro. Sommige bevindingen kunnen zijn gewijzigd sinds de publicatie. Bron: Principled Technologies https://facts.pt/V5fDf06.

	Wat is het verschil met Dell beheertools	Hoeveel beter
Eenvoudigere firmware-updates *iDRAC9 vergeleken met Supermicro IPMI* *OME vergeleken met Supermicro SSM*	• Geautomatiseerde online updates met iDRAC9, met planningsopties • Met OME kunnen aangepaste firmware-repositories worden gemaakt en het kan de firmware van BIOS, BMC en andere servercomponenten bijwerken zonder extra tools of agents	• We hebben automatische updates in iDRAC in **slechts 74 seconden ingesteld** • Supermicro IPMI heeft **geen automatische updatefunctie**, dus moeten beheerders handmatig bijwerken • SSM ondersteunt alleen BIOS- en BMC-firmware-updates en vereist SUM om andere componenten bij te werken
Meer beveiligingsfuncties *iDRAC9 vergeleken met Supermicro IPMI* *OME vergeleken met Supermicro SSM*	• iDRAC9 biedt MFA en dynamische uitschakeling van USB-poorten zonder systeemdowntime • OME biedt op rollen gebaseerde toegangscontrole (RBAC) en scope-gebaseerde toegangscontrole (SBAC) om apparaatbeheer te beperken tot een subset van apparaatgroepen	• Supermicro IPMI heeft **geen MFA-functies** • Supermicro IPMI vereist dat het systeem opnieuw wordt opgestart en de BIOS-configuratie wordt geopend om USB-poorten uit te schakelen • Supermicro SSM biedt RBAC, maar niet meer restrictieve SBAC
Eenvoudiger beheer van de levenscyclus *OME vergeleken met Supermicro SSM*	• Volledig beheer van de levenscyclus zonder agent via OME om beheer en bewaking te vereenvoudigen	• SSM vereist de SuperDoctor5-agent voor gedetailleerde statistieken van de lokale systeemstatus en Supermicro Update Manager (SUM) om extra componenten bij te werken
Eenvoudigere serverimplementatie *iDRAC9 vergeleken met Supermicro IPMI*	• Importeer in slechts 12 stappen een volledig Dell-serverprofiel met behulp van iDRAC9 • Robuuste opties voor BIOS-configuratie met iDRAC9 met 52 BIOS-functies en ondersteuning voor componentconfiguratie zoals RAID NIC en iDRAC	• Met Supermicro IPMI konden we alleen de IPMI-configuratie opslaan en herstellen in plaats van het volledige serverprofiel • iDRAC9 heeft 52 BIOS-functies, terwijl IPMI **geen opties voor BIOS-configuratie biedt**
Meer opties voor rapportage en analyse *iDRAC9 vergeleken met Supermicro IPMI* *OME vergeleken met Supermicro SSM*	• iDRAC9 biedt telemetriestreaming, waarmee gebruikers eenvoudig serverdata kunnen verzenden naar analysetools zoals Splunk • OME stuurt telemetriedata rechtstreeks naar CloudIQ voor eenvoudigere monitoring	• IPMI biedt alleen een SYSLOG-functie die beheerders kunnen gebruiken om berichten te verzenden voor aggregatie en uiteindelijke analyse • SSM heeft geen cloudgebaseerde beheeroplossing die vergelijkbaar is met Dell CloudIQ
Meer duurzaamheidsfuncties *OME vergeleken met Supermicro SSM*	• Meer statistieken voor monitoring in OME Power Manager, inclusief gegevens over de ecologische voetafdruk	• SSM heeft minder robuuste gebruiksstatistieken en **geen manier om de ecologische voetafdruk te volgen**
Meer manieren om te monitoren *OME vergeleken met Supermicro SSM*	• Beheer Dell-servers vanaf elke locatie via de OpenManage Mobile-app • Monitor apparaten van derden met OME met behulp van server-IP's en referenties met ondersteuning voor het importeren van SNMP MIB'S van derden	• SSM heeft **geen mobiele app** • SSM **staat** monitoring van apparaten van derden met server-IP's niet toe

Conclusie

Als het gaat om het ontwerpen, implementeren, beheren en onderhouden van AI-oplossingen in uw bedrijf, moet er met veel factoren rekening worden gehouden. Om u te helpen verstandig te investeren en alles uit uw AI-oplossing te halen, bent u wellicht op zoek naar een leverancier die meer is dan alleen maar een hardwareleverancier. Uit ons onderzoek blijkt dat Dell services biedt die u als partner tijdens het hele traject kunnen helpen. Overweeg dus om te investeren met Dell terwijl u zich bezighoudt met AI.

---

1. Vipera, "NVIDIA's H100 and A100 GPU Cards: Exploring the Intricacies of SXM and PCI-E Connections", geraadpleegd op 5 januari 2024, https://www.viperatech.com/unraveling-the-mysteries-sxm-vs-pci-e-connections-in-nvidias-high-end-h100-and-a100-gpus/.
2. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results", geraadpleegd op 5 januari 2024, https://mlcommons.org/en/inference-datacenter-31/.
3. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results".
4. Dell, "PowerEdge XE Servers", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/servers/specialty-servers/poweredge-xe-servers.htm.
5. Supermicro, "Next Leap of AI Infrastructure is Here", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/accelerators/nvidia.
6. Dell, "PowerEdge XE9680", geraadpleegd op 5 januari 2024, https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9680-spec-sheet.pdf.
7. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0069. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
8. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0135. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
9. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0069. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.

10. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0132. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
11. Dell, "PowerEdge XE9640 Rack Server", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640.
12. Accelsius, "Enabling the AI Revolution with Liquid Cooling", geraadpleegd op 5 januari 2024, https://www.accelsius.com/blog/enabling-the-ai-revolution-with-liquid-cooling.
13. Dell, "PowerEdge XE9640 Technical Guide", geraadpleegd op 5 januari, 2024, https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9640-technical-guide.pdf.
14. Supermicro, "GPU Server Systems", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/products/gpu?pro=pl_grp_type%3D1.
15. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0067. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
16. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0133. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
17. Dell, "PowerEdge XE8640", geraadpleegd op 5 januari 2024, https://www.delltechnologies.com/asset/nl-nl/products/servers/technical-support/poweredge-xe8640-spec-sheet.pdf.
18. Dell, "PowerEdge XE8640 Rack Server", geraadpleegd op 5 januari 2024, https://www.dell.com/nl-nl/shop/ipovw/poweredge-xe8640.
19. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0066. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
20. Geverifieerde MLPerf®-score van v3.1 Inference Closed. Opgehaald van https://mlcommons.org/benchmarks/inference-datacenter/ op 5 december 2023, vermelding 3.1-0133. De naam en het logo van MLPerf zijn geregistreerde en ongedeponeerde handelsmerken van MLCommons Association in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.
21. Dell, "Artificial Intelligence (AI) technologies powered by Dell Precision workstations," geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/ai-technologies/index.htm?hve=explore+dell+precision+for+ai#tab0=0.
22. Supermicro, "Super Workstations", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/products/superworkstation.
23. Supermicro, "Rackmount Workstations", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/products/rackmount-workstations.
24. Stephen Pritchard, "Storage requirements for AI, ML and analytics in 2022", geraadpleegd op 5 januari 2024, https://www.computerweekly.com/feature/Storage-requirements-for-AI-ML-and-analytics-in-2022.
25. Dell, "PowerScale AI-Ready Data Platform", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/shop/powerscale-family/sf/powerscale.
26. Dell, "Dell PowerScale all-flash", geraadpleegd op 5 januari 2024, https://www.delltechnologies.com/asset/nl-nl/products/storage/technical-support/h15963-ss-powerscale-all-flash-nodes.pdf.
27. Dell, "Dell PowerScale OneFS Software Features", geraadpleegd op 5 januari 2024, https://www.delltechnologies.com/asset/nl-nl/products/storage/technical-support/h18275-onefs-software-features-data-sheet.pdf.
28. Dell, "Dell ECS Enterprise Object Storage", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/storage/ecs/.

29. Supermicro, "Accelerating AI Data Pipelines", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/products/storage.
30. Supermicro, "Supermicro Software-Defined Storage and Memory Solutions", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/solutions/software-defined-storage.
31. Supermicro, "Supermicro WEKA Distributed Storage Solution", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/solutions/wekaio.
32. Dell, "Dell ECS Enterprise Object Storage", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/storage/ecs/.
33. Dell, "Dell ECS Enterprise Object Storage".
34. Supermicro, "Supermicro OSNEXUS Software-Defined Storage Solution", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/solutions/osnexus.
35. ASBIS, "Supermicro Solution for Scality RING", geraadpleegd op 5 januari 2024, https://news.asbis.com/news/suppliers/supermicro-renewed-the-line-of-scality-ring-solution/.
36. Supermicro, "Supermicro solution for Quantum ActiveScale", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/solutions/activescale.
37. Dell, "Scalable and elastic Storage as-a-Service", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/apex/storage/data-storage-services/.
38. Dell, "Flip the Switch to Open Networking with PowerSwitch", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/networking/.
39. Dell, "Dell PowerSwitch Data Center Switches", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/networking/data-center-switches/.
40. Supermicro, "SSE-T7132S - 400Gb Ethernet Switch", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/products/accessories/Networking/SSE-T7132SR.php.
41. Dell, "Dell EMC Networking SmartFabric Services Deployment with VxRail 4.7—Fabric Design Center", geraadpleegd op 5 januari 2024, https://infohub.delltechnologies.com/l/dell-emc-networking-smartfabric-services-deployment-with-vxrail-4-7-1/fabric-design-center-26/.
42. Dell, "Dell SmartFabric Services", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/networking/smartfabric/.
43. Penny Madsen, "Scaling Skills for AI: Lessons from Early Adopters", geraadpleegd op 5 januari 2024, https://www.delltechnologies.com/asset/en-us/products/servers/industry-market/idc-brief-importance-of-skills-for-ai-dell.pdf.
44. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Overview", geraadpleegd op 5 januari 2024, https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/overview-5381/.
45. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Advisory Services for Generative AI", geraadpleegd op 5 januari 2024, https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/advisory-services-for-generative-ai-1/.
46. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Adoption Services for Generative AI", geraadpleegd op 5 januari 2024, https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/adoption-services-for-generative-ai-1/.
47. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Adoption Services for Generative AI".
48. Dell, "Artificial Intelligence (AI) Ready Solution Services", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/services/solutions/artificial-intelligence-services.htm.
49. Dell, "Comprehensive AI Training Modules Tailored for You", geraadpleegd op 5 januari 2024, https://education.dell.com/content/emc/en-us/home/training/aiml.html.
50. Supermicro, "Services and Support", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/support.
51. Travis Vigil, "Dell and NVIDIA: Bringing Generative AI to the Enterprise", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/blog/dell-and-nvidia-bringing-generative-ai-to-the-enterprise/.
52. Dell, "Dell Technologies and Hugging Face to Simplify Generative AI with On-Premises IT", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/corporate/newsroom/announcements/detailpage.press-releases~usa~2023~11~20231114-dell-technologies-and-hugging-face-to-simplify-generative-ai-with-on-premises-it.htm.

53. Richard DeMare, "Starburst and Dell expand partnership to accelerate AI efforts with more intelligent data collection", geraadpleegd op 5 januari 2024, https://www.starburst.io/blog/starburst-and-dell-expand-partnership-to-accelerate-ai-efforts-with-more-intelligent-data-collection/.

54. Business Wire, "SiMa.ai and Supermicro Announce Partnership to Accelerate Power Efficient ML at the Edge", geraadpleegd op 5 januari 2024, https://www.businesswire.com/news/home/20231129609794/en/SiMa.ai-and-Supermicro-Announce-Partnership-to-Accelerate-Power-Efficient-ML-at-the-Edge.

55. Dell, "Dell AI Solutions", geraadpleegd op 5 januari 2024, https://www.dell.com/en-us/dt/solutions/artificial-intelligence/index.htm#accordion0&tab0=0

56. Dell, "Unlock the power of AI in virtualized environments", geraadpleegd op 5 januari 2024, https://www.delltechnologies.com/asset/en-us/products/ready-solutions/briefs-summaries/ai-vxrail-powerscale-brief.pdf.

57. Dell, "Design Guide—Virtualizing GPUs for AI with VMware and NVIDIA Based on Dell Infrastructure—Performance Results", geraadpleegd op 5 januari 2024, https://infohub.delltechnologies.com/l/design-guide-virtualizing-gpus-for-ai-with-vmware-and-nvidia-based-on-dell-infrastructure-1/performance-results-15/.

58. Dell, "Design Guide—Virtualizing GPUs for AI with VMware and NVIDIA Based on Dell Infrastructure—Design Considerations", geraadpleegd op 5 januari 2024, https://infohub.delltechnologies.com/l/design-guide-virtualizing-gpus-for-ai-with-vmware-and-nvidia-based-on-dell-infrastructure-1/design-considerations-105/.

59. Supermicro, "Accelerate Every Workload", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/solutions/ai-deep-learning.

60. Supermicro, "Supermicro Enterprise AI Inference & Training", geraadpleegd op 5 januari 2024, https://www.supermicro.com/datasheet/Datasheet_AI-Workloads_Enterprise_AI_Inferencing_and_Training.pdf.

61. Supermicro, "SUPERMICRO RACK SCALE SOLUTIONS: LARGE SCALE AI TRAINING WITH LIQUID COOLING", geraadpleegd op 5 januari 2024, https://www.supermicro.com/solutions/Solution-Brief_Rack_Scale_AI.pdf.

62. Supermicro, "Accelerate Every Workload", geraadpleegd op 5 januari 2024, https://www.supermicro.com/en/solutions/ai-deep-learning.

63. Principled Technologies, "Dell management tools made server deployment and updates easier, offered more comprehensive security, and provided more robust infrastructure analytics", geraadpleegd op 5 januari 2024, https://www.principledtechnologies.com/Dell/Management-tools-vs-Supermicro-0423.pdf.

▶ Bekijk de oorspronkelijke, Engelstalige versie van dit rapport via https://facts.pt/q9p46K9

Dit project is uitgevoerd in opdracht van Dell Technologies.