

Executive summary



Dell biedt meer oplossingen, services en verbruiksopties om AI-implementaties van klanten te ondersteunen



Tot 37% betere prestaties met de helft van de rackruimte

voor natuurlijke taalverwerking (GPT-J) volgens MLperf-data



Dell gevalideerde ontwerpen bieden meer AI-gerichte referentiearchitecturen dan gevalideerde oplossingen van Supermicro



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

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

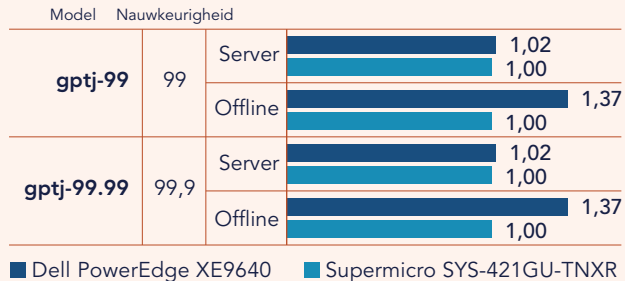
Om kunstmatige intelligentie (AI) te implementeren en de vruchten ervan te plukken, moet u rekening houden met alle aspecten van een oplossing. Het is cruciaal om een partner te vinden die end-to-end ondersteuning kan bieden, van planning, dataverwerking, hardwareselectie, AI-modelontwerp, concepttests, referentiearchitecturen en end-to-end support. In deze samenvatting vergelijken we de prestatie- en ruimtevoordelen van servers in de Dell™ AI-portfolio's met die van Supermicro op basis van openbaar beschikbare MLPerf® Benchmark-testresultaten. MLPerf® test de prestaties voor training en inferentie op verschillende AI-modellen. De data in deze samenvatting zijn afkomstig van resultaten van het MLPerf® v3.1 Inference Datacenter die vanaf november 2023 zijn gepubliceerd op de MLCommons® website.¹ We vergelijken ook openbaar beschikbare informatie over het productaanbod en de oplossingen die elke leverancier aanbiedt om AI-implementaties te ondersteunen.

Servers met vier GPU's: Dell PowerEdge-servers boden betere MLPerf-prestaties dan Supermicro-servers

Met maximaal vier NVIDIA H100 SXM GPU's biedt de 2U Dell PowerEdge XE9640 de helft van de GPU-rekenkracht van de PowerEdge XE9680 in een derde van de ruimte.² De server met vier NVIDIA HGX H100 GPU's van Supermicro in de inzendingen van MLPerf® is de SYS-421GU-TNXR, een 4U-server. In gepubliceerde gptj-99.9 AI-modelresultaten presteerde de PowerEdge XE9640 beter dan de Supermicro-server in de offline tests en behaalde tot 1,37 keer de score (zie afbeelding 1).

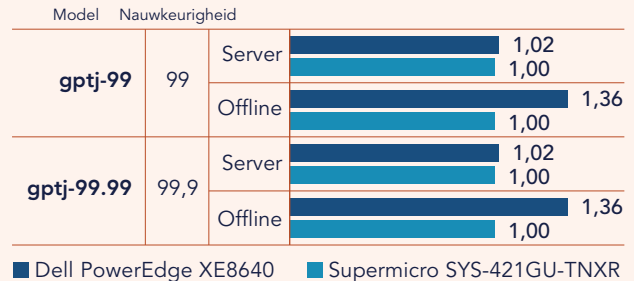
We hebben ook de resultaten van de Supermicro SYS-421GU-TNXR-server vergeleken met de Dell PowerEdge XE8640, een 4U-server met vier GPU's die ook NVIDIA H100 HGX GPU's ondersteunt. De Dell PowerEdge XE8640 is uitgerust met de nieuwste schaalbare 4e generatie Intel Xeon processors en tot 4 TB geheugen om de grote datasets en complexe berekeningen in AI en data-analyse aan te kunnen.³ 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 (zie afbeelding 2).

Genormaliseerde MLPerf®-resultaten: Dell PowerEdge XE9640 met H100 SXM5 versus Supermicro SYS-421GU-TNXR met H100 SXM5 (Larger is better)



Afbeelding 1: 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 data van MLCommons®.*

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



Afbeelding 2: 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 data van MLCommons®.*

Meer dan alleen hardware: Het grotere geheel in ogenschouw nemen

Organisaties op zoek naar flexibiliteit in het aanbod van GPU-workstations, kunnen meer opties vinden in het Dell AI-portfolio. Het grotere Dell AI-portfolio bestaat uit meer dan alleen hardware; het biedt klanten één service- en supportoplossing voor elke fase van hun implementatie. Het Supermicro AI-portfolio biedt daarentegen services die beperkt zijn tot handleidingen, probleemoplossing en het retourneren van producten. Tijdens ons onderzoek konden we geen ontwerp-, implementatie-, beheer- of educatieve services vinden in het Supermicro AI-portfolio. Omdat de implementatie van AI zo complex is, maken de services van Dell het tot een veelbelovende trainingspartner voor het implementeren van deze workloads. Bovendien hebben we vastgesteld dat Dell gevalideerde ontwerpen meer AI-gerichte referentiearchitecturen bevatten dan gevalideerde oplossingen van Supermicro en meer uitgebreide begeleiding bieden.

Lees ons volledige rapport, waarin uitgebreid wordt ingegaan op de voordelen van het Dell AI-portfolio op het gebied van computing-, storage- en netwerkovertuigen, professionele services voor AI, samenwerkingen met derden voor AI-workloads, Dell gevalideerde ontwerpen, die helpen om giswerk uit AI-hardwareoplossingen te halen en beheerservices en iDRAC.

*Geverifieerde MLPerf. score van v3.1 Inference Closed. Uit <https://mlcommons.org/benchmarks/inference-datacenter/> op 5 december 2023, vermelding 3.1-0133, 3.1-0066 en 3.1-0067. De naam en het logo van MLPerf. zijn geregistreerde en niet-geregistreerde handelsmerken van MLCommons. Associatie in de Verenigde Staten en andere landen. Alle rechten voorbehouden. Ongeautoriseerd gebruik is strikt verboden. Zie www.mlcommons.org voor meer informatie.

1. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results," geraadpleegd op 7 februari 2024, <https://mlcommons.org/en/inference-datacenter-31/>.
2. Dell, "PowerEdge XE9640 Rack Server," geraadpleegd op 5 januari 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640>.
3. Dell, "PowerEdge XE8640 Rack Server," geraadpleegd op 5 januari 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe8640>.

► Bekijk de oorspronkelijke Engelstalige versie van dit rapport op <https://facts.pt/kRTwky3>

Klik hier om het rapport te lezen ►



Facts matter.®

Principled Technologies is een geregistreerd handelsmerk van Principled Technologies, Inc. Alle andere productnamen zijn de handelsmerken van hun respectieve eigenaren. Voor aanvullende informatie, bekijk het rapport.