



Sammanfattning



Dell erbjuder en bredare portfölj av lösningar, tjänster och förbrukningsalternativ för att stödja kundernas AI-driftsättningar



Upp till 37 % bättre prestanda på halva rackutrymmet

För naturlig språkbearbetning (GPT-J) per MLPerf-data



Dells validerade utformningar omfattar mer AI-centrerade referensarkitekturer än Supermicro-validerade lösningar

Hitta rätt väg till AI-framgångar med Dells AI-portfölj

En jämförelse av Dells AI-portfölj jämfört med liknande erbjudanden från Supermicro

För att implementera artificiell intelligens (AI) och dra nytta av dess fördelar måste du överväga alla aspekter av en lösning. Hitta en partner som kan stödja hela kedjan: från planering, dataförberedelse och val av hårdvara till AI-modelldesign, koncepttestning, referensarkitekturer och heltäckande support. Det är avgörande. I den här sammanfattningen lyfter vi fram prestanda- och utrymmesfördelar för servrar i Dells™ AI-portföljer jämfört med dem från Supermicro baserat på allmänt tillgängliga testresultat från MLPerf®. MLPerf® testar prestanda för både träning och slutledning på flera AI-modeller. Data i denna sammanfattning kommer från MLPerf® v3.1 Inference Datacenter, resultat som publicerades på MLCommons®-webbplatsen från november 2023.¹ Vi jämför också offentligt tillgänglig information om den mängd produkter och lösningar som varje leverantör erbjuder för att stödja AI-implementeringar.

Fyra GPU-servrar: Dell PowerEdge-servrar erbjöd bättre MLPerf-prestanda än Supermicro-servrar

Med upp till fyra NVIDIA H100 SXM GPU:er erbjuder 2U Dell PowerEdge XE9640 hälften så mycket GPU-beräkningskraft som PowerEdge XE9680 har på en tredjedel av utrymmet.² Fyra GPU-servrar med NVIDIA HGX H100-GPU:er från Supermicro i MLPerf®-inlämningarna är SYS-421GU-TN XR, som är en 4U-server. I publicerade resultat av gptj-99.9 AI-modellen överträffade PowerEdge XE9640 Supermicro-servern i offlinetesterna och uppnådde upp till 1,37 gånger poängen (se bild 1).

Vi jämförde även Supermicro SYS-421GU-TN XR-serverresultaten med Dell PowerEdge XE8640, en 4U-server med fyra GPU:er som även stöder NVIDIA H100 HGX-GPU:er. Dell PowerEdge XE8640 har den senaste 4:e generationens Intel Xeon skalbara processorer och upp till 4 TB minne för att hantera stora datauppsättningar och komplexa beräkningar som är vanliga inom AI och dataanalys.³ Med sin 4U-formfaktor liknar PowerEdge XE8640 Supermicro SYS-421GU-TN XR i både densitet- och GPU-funktioner. Men precis som vi såg med PowerEdge XE9640 fick Dell PowerEdge XE8640 bättre gptj-99-poäng i offline-testerna än Supermicro-servern (se bild 2).

Normaliserade MLPerf®-resultat: Dell PowerEdge XE9640 med H100 SXM5 jämfört med Supermicro SYS-421GU-TNXR med H100 SXM5 (Högre är bättre)

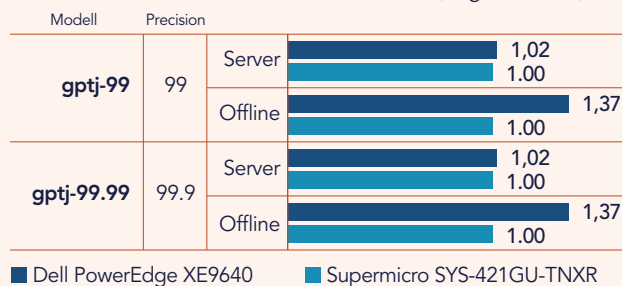


Bild 1: Publicerade MLPerf®-resultat för Dell PowerEdge XE9640 och Supermicro SYS-421GU-TNXR från och med den 29 november 2023. Båda systemen använder SXM-formfaktorn i NVIDIA H100-grafikprocessorerna. Källa: Principled Technologies med data från MLCommons®.*

Normaliserade MLPerf®-resultat: Dell PowerEdge XE8640 med H100 SXM5 jämfört med Supermicro SYS-421GU-TNXR med H100 SXM5 (Högre är bättre)

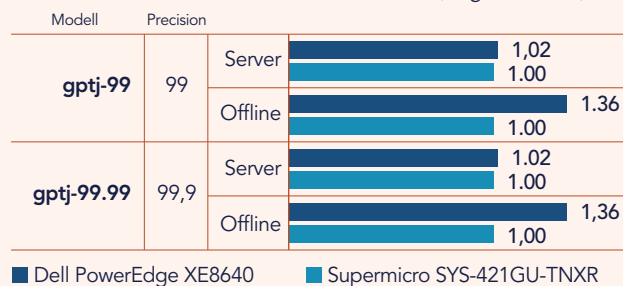


Bild 2: Publicerade MLPerf®-resultat för Dell PowerEdge XE8640 och Supermicro SYS-421GU-TNXR från och med den 29 november 2023. Båda systemen använder SXM-formfaktorn i NVIDIA H100-grafikprocessorerna. Källa: Principled Technologies med data från MLCommons®.*

Utöver hårdvara: se helhetsbilden

Företag som vill ha flexibilitet i GPU-workstation-erbjudanden kan hitta ett bredare utbud av alternativ i Dells AI-portfölj. Den större Dell AI-portföljen omfattar mer än bara hårdvara. Det ger kunderna en enda tjänst- och supportlösning för varje steg av driftsättningen. Supermicro AI-portföljen erbjuder däremot tjänster som är begränsade till handböcker, felsökning och produktreturer. I vår forskning kunde vi inte hitta design-, implementerings-, hanterings- eller utbildningstjänster i Supermicro AI-portföljen. Eftersom AI-implementeringen är så komplex gör de tjänster som Dell erbjuder det till en mer lovande utbildningspartner för att driftsätta dessa arbetsbelastningar. Dessutom upptäckte vi att Dells validerade utformningar omfattar mer AI-centrerade referensarkitekturer än Supermicro-validerade lösningar och erbjöd mer djupgående vägledning.

Läs vår fullständiga rapport, som utforskar de djupgående fördelar som Dells AI-portfölj erbjuder inom beräknings-, lagrings- och nätverksalternativ; professionella tjänster för AI; tredjepartspartnerskap för AI-arbetsbelastningar; Dell-validerade utformningar som hjälper till att eliminera gissningar från AI-maskinvarulösningar; och hanteringstjänster och iDRAC.

*Verifierad MLPerf-poäng för v3.1 Inference Closed. Hämtad från [den https://mlcommons.org/benchmarks/inference-datacenter/](https://mlcommons.org/benchmarks/inference-datacenter/) 5 december 2023, posterna 3.1-0133, 3.1-0066 och 3.1-0067. MLPerf. namn och logotyp är registrerade och oregistrerade varumärken som tillhör MLCommons. Association i USA och andra länder. Med ensamrätt. Obehörig användning är strängt förbjuden. Besök www.mlcommons.org för mer information.

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

► Se den ursprungliga, engelska versionen av den här rapporten på <https://facts.pt/kRTwky3>

Klicka här om du vill läsa rapporten ►



Facts matter.®

Principled Technologies är ett registrerat varumärke som tillhör Principled Technologies, Inc. Alla andra produktnamn är varumärken som tillhör respektive ägare. Mer information finns i rapporten.