



Firma Dell oferuje szerszą ofertę rozwiązań, usług i opcji wykorzystania, aby wspierać wdrożenia sztucznej inteligencji u klientów



Nawet o 37% lepsza wydajność przy połowie miejsca zajmowanego w szafie serwerowej w przypadku przetwarzania języka naturalnego (GPT-J) na podstawie danych MLPerf



Projekty Dell Validated Designs zawierają więcej architektur referencyjnych skoncentrowanych na sztucznej inteligencji niż zatwierdzone rozwiązania Supermicro

Droga do sukcesu w zakresie sztucznej inteligencji dzięki ofercie rozwiązań AI firmy Dell

Porównanie oferty rozwiązań Dell w zakresie sztucznej inteligencji z podobnymi ofertami Supermicro

Aby wdrożyć sztuczną inteligencję (AI) i czerpać z niej korzyści, należy wziąć pod uwagę wszystkie aspekty rozwiązania. Kluczowe znaczenie ma znalezienie partnera, który zapewni kompleksowe wsparcie w zakresie planowania, przygotowania danych, wyboru sprzętu, projektowania modeli AI, testowania weryfikacji modeli koncepcyjnych, architektur referencyjnych i pomocy technicznej. W tym podsumowaniu podkreślamy korzyści wynikające z wydajności i przestrzeni serwerów z™ portfolio Dell AI w porównaniu z serwerami Supermicro na podstawie oficjalnych® wyników testów MLPerf Benchmark. Testy MLPerf® sprawdzają wydajność zarówno w zakresie treningu, jak i inferencji na kilku modelach sztucznej inteligencji. Dane zawarte w tym podsumowaniu pochodzą z® wyników MLPerf v3.1 Inference Datacenter opublikowanych na® stronie MLCommons w listopadzie 2023 r.¹ Porównamy również publicznie dostępne informacje o gamie produktów i rozwiązań oferowanych przez każdego dostawcę w celu obsługi wdrożeń sztucznej inteligencji.

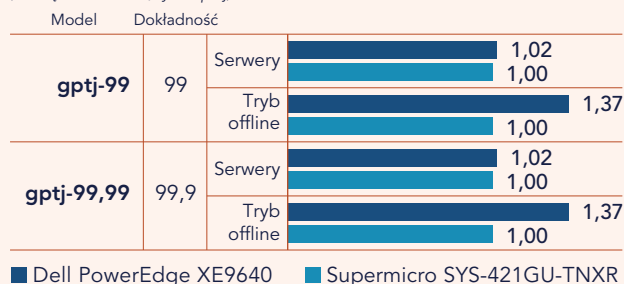
Serwery z czterema procesorami graficznymi: Serwery Dell PowerEdge zapewniają lepszą wydajność MLPerf niż serwery Supermicro

Dzięki maksymalnie czterem procesorom graficznym NVIDIA H100 SXM serwer Dell PowerEdge XE9640 o wysokości 2U zapewnia połowę mocy obliczeniowej procesora graficznego PowerEdge XE9680 na przestrzeni o dwie trzecie mniejszej.² Serwer z czterema procesorami graficznymi NVIDIA HGX H100 zgłoszony w MLPerf® przez firmę Supermicro to SYS-421GU-TNXR, który jest serwerem 4U. W opublikowanych wynikach modelu sztucznej inteligencji gptj-99.9 serwer PowerEdge XE9640 przewyższył serwer Supermicro w testach offline, osiągając nawet 1,37-krotny wynik (patrz rysunek 1).

Porównaliśmy również wyniki serwera Supermicro SYS-421GU-TNXR z Dell PowerEdge XE8640, serwerem 4U z czterema procesorami graficznymi, który obsługuje również procesory graficzne NVIDIA H100 HGX. Serwer Dell PowerEdge XE8640 jest wyposażony w najnowsze skalowalne procesory Intel Xeon 4. generacji i do 4 TB pamięci do obsługi dużych zbiorów danych oraz złożonych obliczeń typowych dla sztucznej inteligencji i analizy danych.³ Dzięki obudowie 4U serwer PowerEdge XE8640 jest podobny do Supermicro SYS-421GU-TNXR zarówno pod względem gęstości, jak i możliwości GPU. Jednak podobnie jak w przypadku serwerów PowerEdge XE9640 serwer Dell PowerEdge XE8640 uzyskał lepsze wyniki gptj-99 w testach offline niż serwer Supermicro (patrz rysunek 2).

Znormalizowane wyniki MLPerf®: Dell PowerEdge XE9640 z procesorem H100 SXM5 w porównaniu z

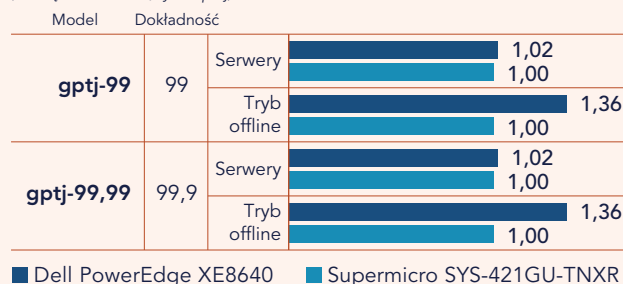
Supermicro SYS-421GU-TNXR z procesorem H100 SXM5
(im większa wartość, tym lepiej)



Rysunek 1. Opublikowano wyniki MLPerf® dla serwerów Dell PowerEdge XE9640 i Supermicro SYS-421GU-TNXR według stanu na dzień 29/11/2023. Oba układy wykorzystują format SXM procesorów graficznych NVIDIA H100. Źródło: Principled Technologies na podstawie danych z MLCommons®.*

Znormalizowane wyniki MLPerf®: Dell PowerEdge XE8640 z procesorem H100 SXM5 w porównaniu z

Supermicro SYS-421GU-TNXR z procesorem H100 SXM5
(im większa wartość, tym lepiej)



Rysunek 2. Opublikowano wyniki MLPerf® dla serwerów Dell PowerEdge XE8640 i Supermicro SYS-421GU-TNXR według stanu na dzień 29/11/2023. Oba układy wykorzystują format SXM procesorów graficznych NVIDIA H100. Źródło: Principled Technologies na podstawie danych z MLCommons®.*

Więcej niż sprzęt: Spojrzenie na szerszy obraz

Organizacje, które potrzebują elastyczności w ofercie stacji roboczych z procesorami graficznymi, mogą znaleźć szerszy zakres opcji w ofercie rozwiązań Dell AI. Większe portfolio rozwiązań Dell AI obejmuje nie tylko sprzęt — zapewnia klientom pojedyncze usługi i rozwiązanie pomocy technicznej na każdym etapie wdrażania. Z kolei oferta Supermicro AI obejmuje usługi ograniczone do instrukcji obsługi, rozwiązywania problemów i zwrotów produktów. W naszych badaniach nie udało nam się znaleźć usług w domenie projektowania, wdrażania, zarządzania ani edukacji w portfolio sztucznej inteligencji Supermicro. Ponieważ wdrażanie sztucznej inteligencji jest bardzo złożone, usługi oferowane przez firmę Dell sprawiają, że jest ona bardziej obiecującym partnerem szkoleniowym w zakresie wdrażania tych obciążeń roboczych. Ponadto zauważyliśmy, że projekty Dell Validated Design zawierają więcej architektur referencyjnych zorientowanych na sztuczną inteligencję niż zatwierdzone rozwiązania Supermicro i zapewniają bardziej szczegółowe wskazówki.

Zapoznaj się z pełnym raportem, w którym szczegółowo omówiono korzyści płynące z oferty rozwiązań Dell AI w zakresie rozwiązań obliczeniowych, pamięci masowej i sieci, profesjonalnych usługi w zakresie sztucznej inteligencji, partnerstwa innych firm w zakresie obciążeń roboczych związanych ze sztuczną inteligencją, projektami Dell Validated Design, które pomagają wyeliminować niepewność w przypadku rozwiązań sprzętowych z zakresu sztucznej inteligencji oraz usług zarządzania i kontrolerów iDRAC.

*Zweryfikowany wynik MLPerf. w wersji 3.1 — wnioskowanie zamknięte. Pobrano ze strony <https://mlcommons.org/benchmarks/inference-datacenter/> 5 grudnia 2023 r.; wpisy 3.1-0133, 3.1-0066 i 3.1-0067. Nazwa i logo MLPerf. są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons. Stowarzyszenia w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.

1. MLCommons, „MLPerf Inference: Datacenter Benchmark Suite Results”, dostęp 7 lutego 2024 r., <https://mlcommons.org/en/inference-datacenter-31/>.
2. Dell, „PowerEdge XE9640 Rack Server”, dostęp 5 stycznia 2024 r., <https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640>.
3. Dell, „PowerEdge XE8640 Rack Server”, dostęp 5 stycznia 2024 r., <https://www.dell.com/en-us/shop/ipovw/poweredge-xe8640>.

► Zobacz oryginalną, angielską wersję tego podsumowania na stronie <https://facts.pt/kRTWky3>

Kliknij tutaj, aby przeczytać raport ►



Facts matter.®

Principled Technologies jest zarejestrowanym znakiem towarowym firmy Principled Technologies, Inc. Pozostałe nazwy produktów są znakami towarowymi odpowiednich właścicieli. Aby uzyskać dodatkowe informacje, zapoznaj się z raportem.