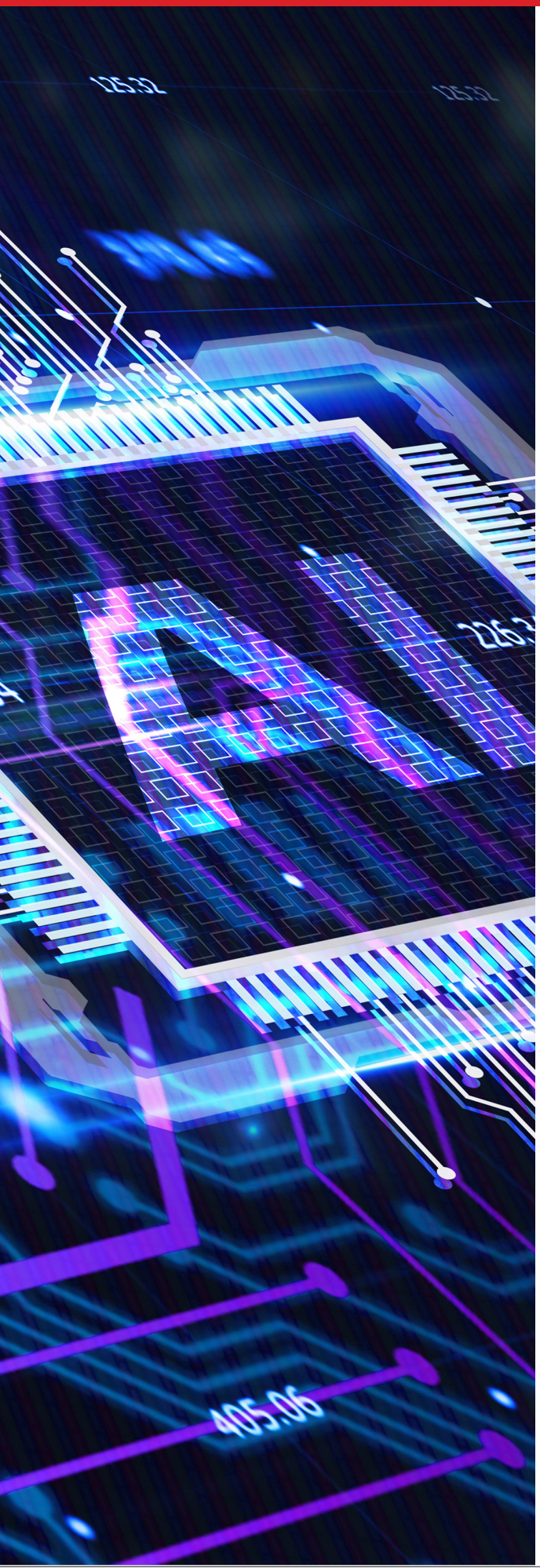


Sprostanie wyzwaniom związanym z obciążeniami wynikającymi z SI dzięki ofercie rozwiązań Dell AI

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

Nie ma wątpliwości, że sztuczna inteligencja (SI) zmieniła krajobraz biznesowy i umożliwiła branżom oraz organizacjom każdej wielkości uzyskanie głębszego wglądu w dane, automatyzację procesów biznesowych, dostarczanie wrażeń dostosowanych do potrzeb klientów i użytkowników oraz większą konkurencyjność w danej branży. Aby skutecznie wykorzystać możliwości sztucznej inteligencji, organizacje potrzebują dostawcy infrastruktury, który może zaoferować kompleksowe i zintegrowane portfolio rozwiązań obejmujące cały cykl życia sztucznej inteligencji.

Dostawcy infrastruktury, tacy jak Dell Technologies i Hewlett Packard Enterprise (HPE), pomagają klientom sprostać rosnącym wymaganiom związanym ze sztuczną inteligencją i radzić sobie z jej nieodłączną złożonością. Prezentując portfolio gotowe do pracy nad sztuczną inteligencją, dostawcy ci oferują różne poziomy rozwiązań w zakresie sztucznej inteligencji, które łączą wysokowydajne rozwiązania infrastruktury lokalnej i chmurowej z partnerstwami strategicznymi oraz szeregiem usług wsparcia i konsultacji.

W niniejszym raporcie przeanalizowano publicznie dostępne informacje na temat oferty firm Dell i HPE w dziedzinie sztucznej inteligencji* w celu zwrócenia uwagi na konkretne zalety w zakresie architektury, wydajności i pomocy technicznej, z których mogą skorzystać klienci wybierający technologie firmy Dell dla swoich potrzeb związanych z SI. Porównujemy szczegółowe informacje o serwerach zaprojektowanych przez firmę Dell do obsługi wdrożeń sztucznej inteligencji i odwołujemy się do wyników testów porównawczych w branży z ML Commons®. Badamy również dodatkowe oferty oprogramowania i usług, które wspierają klientów na każdym etapie eksploatacji sztucznej inteligencji.

*Uwaga: PT przeprowadziło wszystkie badania w dniu 5 grudnia 2023 r. lub wcześniej, więc ten dokument nie będzie odzwierciedlał ofert ani zmian wprowadzonych przez firmy Dell lub HPE po tej dacie.

Wyzwania związane z wdrażaniem sztucznej inteligencji

Przyjęcie strategii sztucznej inteligencji wiąże się z wieloma nowymi wyzwaniami dla centrów danych i personelu IT, który je obsadza, w tym:

- uzupełnieniem istniejących luk w umiejętnościach obecnego personelu poprzez wewnętrzne szkolenia w zakresie sztucznej inteligencji lub zatrudnianie pracowników z zewnątrz;
- zrozumieniem potrzeb sztucznej inteligencji w zakresie przygotowywania danych, w tym jakości, ilości, lokalizacji i bieżącego stanu danych firmy;
- oceną konkretnych celów biznesowych w zakresie sztucznej inteligencji w celu lepszego określenia, które modele i implementacje sztucznej inteligencji przyniosą korzyści;
- oceną potrzeb obliczeniowych, sieciowych i pamięci masowej planowanych systemów sztucznej inteligencji oraz określeniem planu ich nabycia.

To tylko kilka przykładów wielu często znacznych przeszkód, z którymi musi się zmierzyć firma, chcąc czerpać korzyści z wdrożenia sztucznej inteligencji w swoich centrach danych.

Oferta sztucznej inteligencji firmy Dell ma na celu pomóc klientom w sprostaniu tym wyzwaniom poprzez usługi profesjonalne i konsultingowe, które pomagają im w tworzeniu planów wdrożenia i przygotowywaniu danych dla modeli sztucznej inteligencji.[1] W ofercie znajdują się również szkolenia, które obejmują koncepcje uczenia maszynowego (ML) i inne tematy edukacyjne. Ponadto dostępne są sprawdzone projekty dla sztucznej inteligencji, aby pomóc zapewnić sukces wdrożenia.[2] Co więcej Dell współpracuje z innymi firmami, aby zapewnić klientom dodatkowe narzędzia sztucznej inteligencji, takie jak niestandardowy portal Dell w ramach społeczności Hugging Face z dedykowanymi kontenerami i skryptami do wdrażania[3] oraz łatwe wdrażanie dużego modelu językowego Meta Llama 2 (LLM).[4] Poza szerokim wyborem rozwiązań obliczeniowych i komputerowych, od mobilnych stacji roboczych po serwery obsługujące do 8 wysokiej klasy procesorów graficznych NVIDIA, firma Dell zapewnia również niestrukturalne przechowywanie danych obejmujących szereg wysokowydajnych macierzy do przechowywania plików i obiektów, których wymaga sztuczna inteligencja. Te rozwiązania pamięci masowej, w tym Dell PowerScale, ObjectScale, ECS i wbudowana pamięć masowa, mogą obsługiwać nieustrukturyzowane dane, które często są wykorzystywane w obciążeniach związanych ze sztuczną inteligencją.[5] Firma Dell nawiązała również współpracę ze Snowflake, aby dostarczyć swoim klientom rozwiązanie pamięci masowej w chmurze hybrydowej.[6] Według analizy z sierpnia 2023 r. Dell oferuje „najszerszy wybór generatywnej sztucznej inteligencji", który wykracza poza same serwery i pamięć masową, zapewniając zasoby w całym procesie wdrażania sztucznej inteligencji.[7]

Wydajność sztucznej inteligencji i opcje przyspieszenia mocy obliczeniowej: Dell w porównaniu z HPE

Obciążenia związane ze sztuczną inteligencją mogą używać procesorów CPU, GPU lub obu tych podzespołów jako zasobów obliczeniowych — w zależności od rozmiaru lub typu obciążenia. Niektóre procesory zawierają akceleratory przeznaczone specjalnie do sztucznej inteligencji, takie jak Intel Advanced Matrix Extensions (Intel AMX) w najnowszych skalowalnych procesorach Intel Xeon.[8] Procesory graficzne są często lepsze w przypadku większych i/lub bardziej wymagających obciążeń, ale format procesora graficznego może wpływać na poziom wydajności. Na przykład niektóre procesory graficzne NVIDIA A100 i H100 są dostępne w uniwersalnych obudowach PCIe lub zastrzeżonych formatach SXM, przy czym te ostatnie wykorzystują wydajniejszą architekturę NVIDIA SXM.[9] Duże pojemności pamięci i cechy konstrukcyjne serwera, takie jak architektura chłodzenia i wydajność energetyczna, również wpływają na wydajność. Większość centrów danych nadal korzysta z chłodzenia powietrzem, co oznacza, że obciążenia systemów obliczeniowych o wysokiej wydajności (HPC) wymagają serwerów zbudowanych z myślą o jak najbardziej skutecznym chłodzeniu powietrzem. Poniżej przedstawiamy ofertę serwerów PowerEdge z podziałem na podzespoły, opcje chłodzenia i inne elementy, a także opublikowane wyniki MLCommons® MLPerf®.

Test porównawczy wydajności modelu SI: porównanie wyników MLPerf

MLPerf® to pakiet testów porównawczych, który testuje wydajność sztucznej inteligencji zarówno na potrzeby trenowania, jak i wnioskowania. Aby organizacja mogła opublikować oficjalne wyniki MLPerf®, muszą być one zgodne z warunkami określonymi przez firmę odpowiedzialną za programowanie testu porównawczego — MLCommons®.[10] Te wytyczne dotyczące zgodności zawierają standardy, które ułatwiają porównywanie wyników. Do testowania wnioskowania MLPerf® używa zestawów danych Datacenter, Edge, Mobile i Tiny oraz raportuje wyniki sztucznej inteligencji i waty energii zużytej podczas testowania. Pakiet testów porównawczych wnioskowania obejmuje testowanie wielu popularnych modeli sztucznej inteligencji, uczenia maszynowego i uczenia głębokiego; patrz tabela 1.

Tabela 1. Modele sztucznej inteligencji, uczenia maszynowego i uczenia głębokiego zawarte w testach MLPerf® i typowe przypadki użycia dla każdego z nich. Źródło: Principled Technologies.

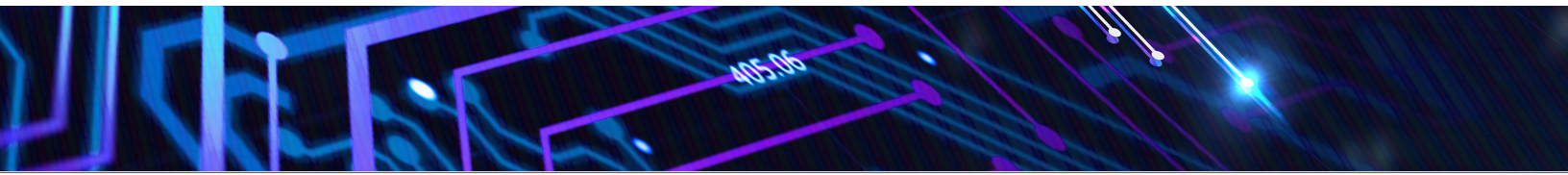
Typowe modele sztucznej inteligencji	Typowe zastosowania
ResNet	Model klasyfikacji obrazów, który pomaga komputerom się uczyć, zapamiętywać i identyfikować różne obrazy na potrzeby przypadków użycia, takich jak obrazowanie medyczne, moderowanie treści w mediach społecznościowych i rozpoznawanie twarzy.
RetinaNet	Typ wykrywania obiektów, który może obsłużyć dodatkową złożoność w porównaniu z ResNet. Pomaga komputerom identyfikować i lokalizować obiekty na obrazach lub w klatkach wideo i klasyfikować je według ważności. Używany w takich dziedzinach jak autonomiczna jazda, technologia automatycznego wspomagania pojazdów, monitoring, rozpoznawanie twarzy.
3D-UNet	Właściwe dla segmentacji obrazów medycznych
RNN-T	Rozpoznawanie mowy w przypadkach użycia takich jak tłumaczenie automatyczne
BERT	Przetwarzanie języka naturalnego w przypadkach użycia takich jak streszczanie tekstu, tłumaczenie i automatyczne wykonywanie zadań
DLRM-v2-99.9	Model rekomendacji dla przypadków użycia takich jak ukierunkowane reklamy i spersonalizowane rekomendacje produktów
GPTJ-99 oraz 99.9	Duży model językowy (LLM) do przetwarzania języka naturalnego, który doskonale sprawdza się w generowaniu tekstu w przypadkach użycia takich jak chatboty i narzędzia SI oparte na czacie

MLPerf

Wyniki MLPerf® obejmują kilka parametrów oprócz samych modeli sztucznej inteligencji, co może sprawić, że wiele danych będzie trzeba przeanalizować na jednym wykresie lub w jednej tabeli. Oto krótki opis tych parametrów:

- 99.0 i 99.9: te liczby odnoszą się do dokładności, do której model został wytrenowany. Im dokładniejsze mają być dane wyjściowe, tym bardziej złożony jest model i tym dłużej może trwać przetwarzanie danych.
- Próbki offline/s: tryb, w którym test porównawczy wysyła wszystkie zapytania na początku testu, symulując dane już obecne w systemie.
- Zapytania serwera/s: tryb, w którym test porównawczy wysyła zapytania przez cały czas trwania testu, symulując analizowanie strumienia danych na żywo.

Aby uzyskać więcej informacji na temat wyników MLCommons® i MLPerf®, zobacz https://mlcommons.org/benchmarks/inference-datacenter/.



Wyniki zawarte w tym raporcie pochodzą z wyników MLPerf® v3.1 Inference Datacenter opublikowanych na stronie MLCommons® w listopadzie 2023 r.[11] Wyniki te obejmują zgłoszenia od producentów technologii i dostawców usług w chmurze i obejmują szereg konfiguracji. W porównaniu z publicznie dostępnymi zgłoszeniami HPE serwery firmy Dell uzyskały lepsze wyniki w niektórych modelach sztucznej inteligencji. (Uwaga: różne konfiguracje procesorów graficznych na różnych serwerach mogą utrudniać bezpośrednie porównania). Szczegółowe informacje znajdują się w tabeli 2.

Tabela 2. Serwery firm Dell i HPE uwzględnione w wynikach MLCommons® MLPerf® 3.1 opublikowanych w dniu 29/11/2023. Źródło: Principled Technologies.

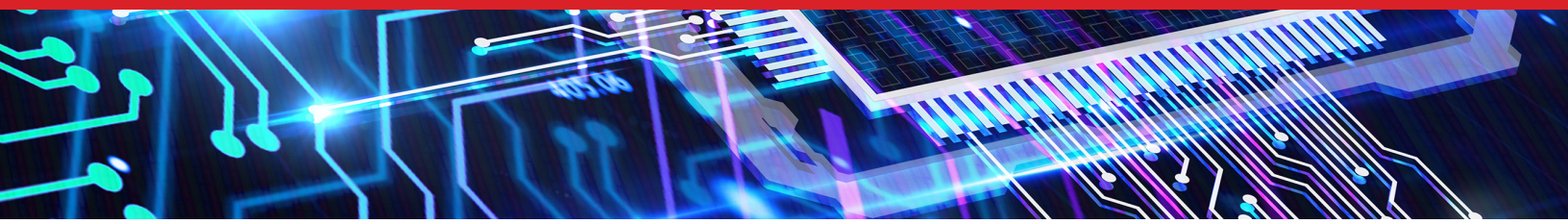
Zgłaszający	Model serwera	Liczba i model procesorów graficznych	Opis
Dell[12]	PowerEdge XE9680	8 × NVIDIA H100 SXM	Do trenowania sztucznej inteligencji i wnioskowania przy dużych obciążeniach, takich jak duże modele językowe
	PowerEdge XE9640	4 × NVIDIA H100 SXM	Do trenowania dużych modeli sztucznej inteligencji w centrach danych o dużej gęstości chłodzonych cieczą
	PowerEdge XE8640	4 × NVIDIA H100 SXM	Do obsługi tradycyjnych szkoleń SI, HPC i aplikacji do analizy danych w obudowie 4U dla centrów danych chłodzonych powietrzem
	PowerEdge R760xa	4 × NVIDIA H100 PCIe	Do szerokiej gamy obciążeń o dużej mocy obliczeniowej, w tym szkolenia i wnioskowania AI-ML/DL, które nie wymagają procesorów GPU o najwyższej wydajności
HPE[13,14]	ProLiant XL675d Plus 10. generacji	8 × NVIDIA A100 SXM	Do systemów obliczeniowych o wysokiej wydajności i sztucznej inteligencji
	ProLiant DL380a 11. generacji	4 × NVIDIA H100 PCIe	Serwer 2U do umiarkowanych obciążeń związanych ze sztuczną inteligencją

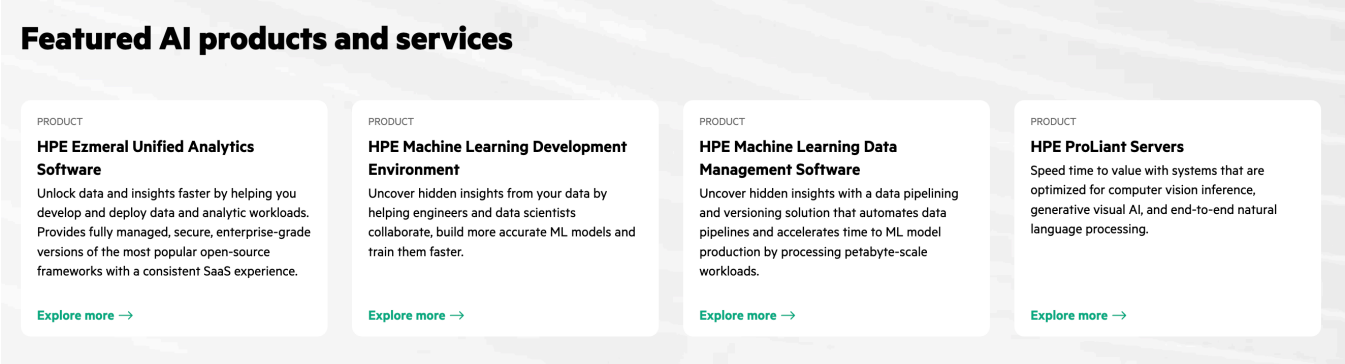
Bezpośrednie porównanie serwerów Dell i HPE

Choć kompleksowa strategia sztucznej inteligencji to coś więcej niż sprzęt, zapewnienie najwyższej wydajności jest jednym z najważniejszych czynników sukcesu obciążeń roboczych związanych ze sztuczną inteligencją. Wraz z pojawieniem się nowszych procesorów graficznych i innych technologii ewoluują również możliwości obciążeń roboczych związanych ze sztuczną inteligencją. W czasie, gdy po raz pierwszy opublikowano wyniki badania MLPerf® v3.1, najlepszym dostępnym procesorem graficznym NVIDIA był H100 Tensor Core. Firma Dell opublikowała wyniki wydajności MLPerf® kilku serwerów zarówno w obudowach PCIe, jak i SXM5, w których zastosowano ten procesor.[15] Opublikowane wyniki HPE obejmowały tylko jedno zgłoszenie z procesorem H100 i jedynie w obudowie PCIe. Nasze badania wykazały, że niewiele z dostępnych serwerów HPE z procesorami graficznymi obsługuje format SXM5 H100 zapewniający najwyższą wydajność procesorów graficznych NVIDIA. Takiej możliwości nie zapewnia żaden z serwerów HPE ProLiant.[16] Jak pokazujemy poniżej, zastosowanie lepszych procesorów graficznych zazwyczaj poprawia wydajność obciążeń związanych ze sztuczną inteligencją.

Wyniki MLPerf z ośmioma procesorami graficznymi

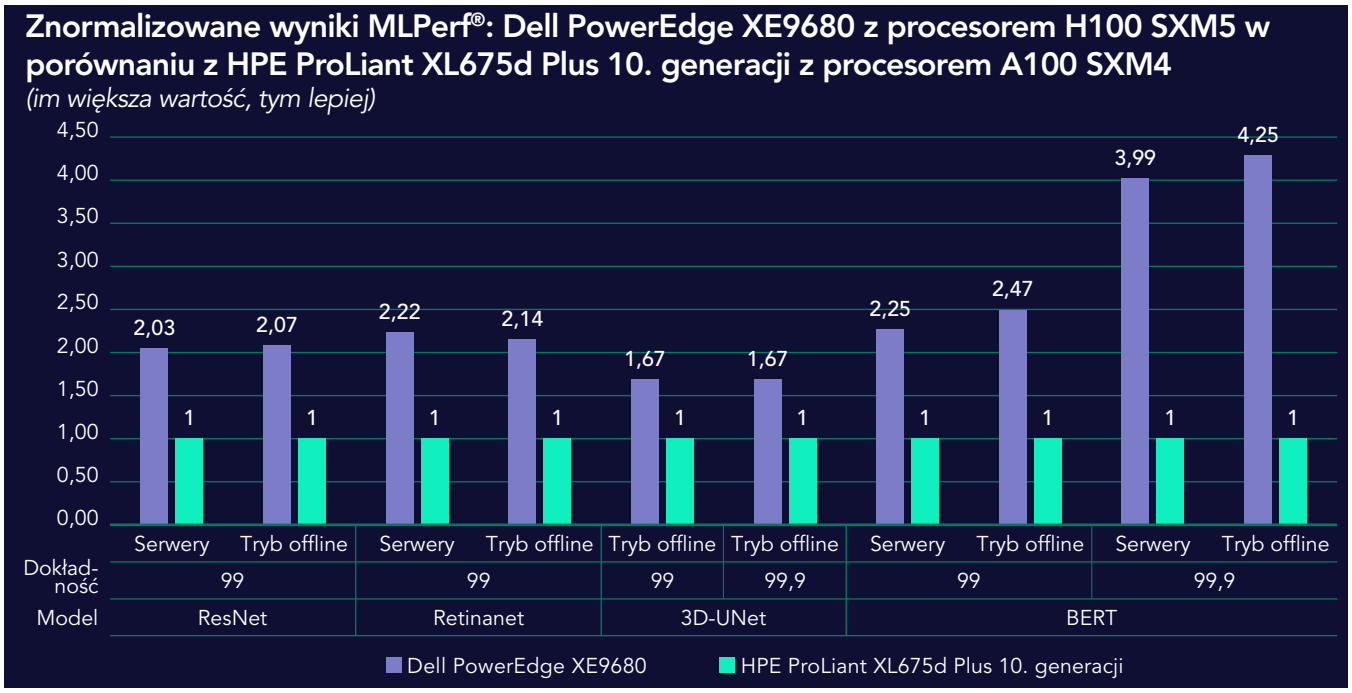
Serwer Dell PowerEdge XE9680 obsługuje do ośmiu procesorów graficznych NVIDIA H100 SXM5 w celu przyspieszenia SI oraz maksymalnie dwa skalowalne procesory Intel® Xeon® czwartej generacji. Rodzina produktów PowerEdge XE ma modułową architekturę obsługującą procesory graficzne NVIDIA SXM4 lub SXM5 bądź zestawy procesorów graficznych Open Compute Project Accelerator Module (OAM) firmy AMD, co może zwiększyć wydajność w porównaniu ze standardowymi procesorami graficznymi PCIe.[17] Model PowerEdge XE9680 to kompaktowy serwer z ośmioma procesorami NVIDIA H100 SXM5 zajmujący zaledwie 6U miejsca w szafie serwerowej. Najnowsze serwery HPE ProLiant 11. generacji nie obsługują obecnie formatu H100 SXM,[18] chociaż niektóre serwery HPE Cray Supercomputing dają taką możliwość.[19] Ponieważ firma HPE nie przesłała żadnych wyników MLPerf® dla serwerów Cray, a na stronie z ofertą sztucznej inteligencji wyróżnia tylko serwery ProLiant, w tym artykule skupimy się na serwerach ProLiant. (Patrz Rysunek 1).





Rysunek 1. Zrzut ekranu przedstawiający polecane produkty i usługi SI w https://www.hpe.com/us/en/solutions/ai-artificial-intelligence.html z wyróżnieniem serwerów HPE ProLiant według stanu na dzień 05/12/2023.

W wynikach MLPerf® v3.1 opublikowanych po raz pierwszy w listopadzie 2023 r. dla serwerów z ośmioma procesorami graficznymi model Dell PowerEdge XE9680 z procesorami graficznymi NVIDIA SXM5 H100 przewyższył nawet 4,25-krotnie model HPE ProLiant XL675d Plus 10. generacji z procesorami graficznymi NVIDIA SXM4 A100 (patrz rysunek 2).



Rysunek 2. Opublikowano wyniki MLPerf® dla serwerów Dell PowerEdge XE9680 i HPE ProLiant XL675d Plus 10. generacji według stanu na dzień 29/11/2023. System firmy Dell korzysta z procesora graficznego NVIDIA H100, podczas gdy procesory graficzne w systemie HPE są o jedną generację starsze. Źródło: Principled Technologies na podstawie danych z MLCommons®.[20,21]

Dla ułatwienia porównania znormalizowaliśmy wyniki testów na rysunkach od 2 do 5. Oznacza to, że przypisujemy wartość 1 do każdego wyniku testu HPE ProLiant DL380a 11. generacji i pokazujemy odpowiadający mu wynik Dell PowerEdge R760xa. Jak pokazują te wyniki, nawet jedna generacja różnicy między modelami procesorów graficznych może mieć znaczny wpływ na wydajność, której można oczekiwać w wielu obciążeniach związanych ze sztuczną inteligencją.

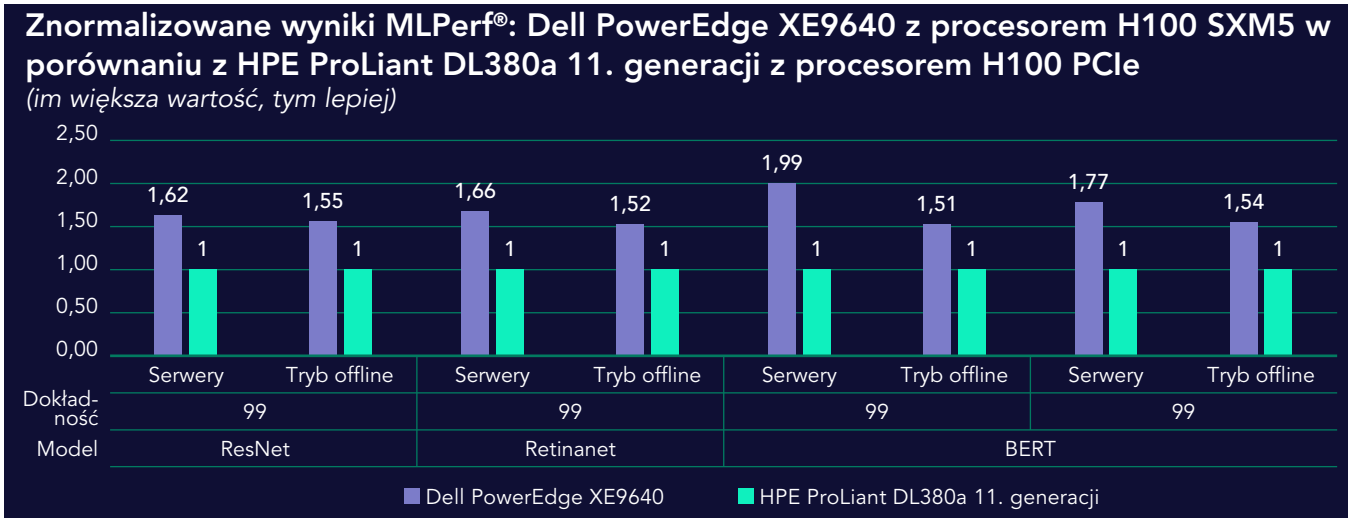
Wyniki MLPerf z czterema procesorami graficznymi

Model Dell PowerEdge XE9640 2U może być właściwym wyborem, gdy oszczędność energii lub miejsca w centrum danych ma kluczowe znaczenie. Dzięki maksymalnie czterem procesorom graficznym NVIDIA H100 SXM serwer PowerEdgeXE9640 zapewnia o połowę mniejszą moc obliczeniową niż model XE9680 na przestrzeni o dwie trzecie mniejszej.[22] Gęsto upakowany serwer Dell PowerEdge XE9640 jest wyposażony w technologię Dell Smart Cooling, która zapewnia szereg rozwiązań termicznych, w tym bezpośrednie chłodzenie cieczą dla procesorów i kart graficznych.[23]

Obudowa 2U serwera PowerEdge XE9640 jest wyposażona w ulepszone mechanizmy przepływu powietrza, w tym większe wentylatory i radiatory, które pomagają chłodzić inne ważne podzespoły, takie jak karty PCIe i pamięć.[24] PowerEdge XE9640 jest obecnie jedynym serwerem 2U oferowanym przez firmę Dell lub HPE, który jest wyposażony w cztery procesory graficzne HGX H100. Oferta HPE w zakresie SI obejmuje serwery ProLiant 11. generacji w obudowach 1U i 2U, ale są one ograniczone do procesorów graficznych w formacie PCIe.[25]

Serwer Dell PowerEdge XE9640 obsługuje również procesory graficzne Intel Max OAM z serii 1550 o niskim poborze mocy i dużej gęstości, które obejmują kartę PCIe i moduł OpenCompute Accelerator Module (OAM).[26] Chociaż nie udało nam się ustalić, czy na dzień 05/12/2023 firma HPE oferowała serwer z tymi procesorami graficznymi, w jej portfolio znajdują się serwery ProLiant DL380 11. generacji i DL380a 11. generacji z procesorami graficznymi Intel Data Center GPU Max 1100.[27] Oznacza to, że Dell PowerEdge XE9640 może być obecnie jedynym produktem z czterema procesorami graficznymi Intel Max 1550 OAM w obudowie 2U. Serwer 2U z czterema procesorami graficznymi Intel Max 1550 łączy wysoką wydajność obliczeniową i energooszczędność bez poświęcania miejsca w centrum danych — to idealne rozwiązanie dla firm zainteresowanych oszczędnością miejsca i wydajnością energetyczną.

W opublikowanym badaniu MLPerf® 3.1 serwer Dell PowerEdge XE9640 z czterema procesorami graficznymi HGX H100 przewyższył niemal dwukrotnie model HPE ProLiant DL380a z czterema procesorami graficznymi PCIe H100 pod względem wydajności (patrz rysunek 3).

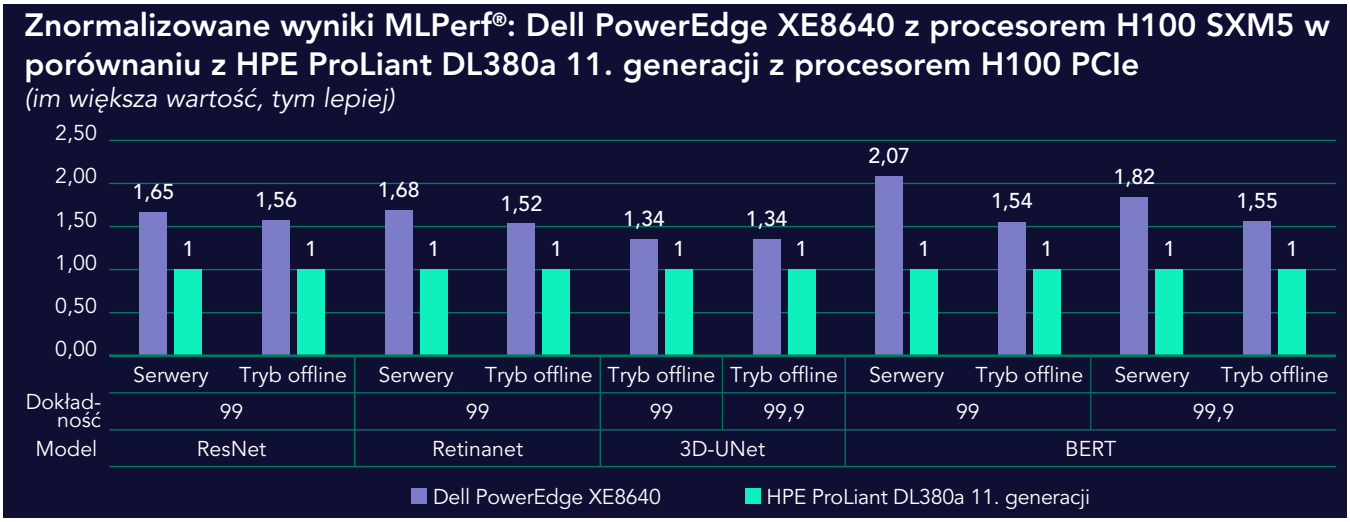


Rysunek 3. Opublikowano wyniki MLPerf® dla serwerów Dell PowerEdge XE9680 i HPE ProLiant XL675d Plus 10. generacji według stanu na dzień 29/11/2023. System firmy Dell korzysta z procesora graficznego NVIDIA H100, podczas gdy procesory graficzne w systemie HPE są o jedną generację starsze. Źródło: Principled Technologies na podstawie danych z MLCommons®.[28,29]

Serwer PowerEdge XE8640 jest dostępny w konfiguracji z czterema procesorami graficznymi z układem chłodzenia procesorów powietrzem i układem chłodzenia procesorów graficznych powietrzem wspomaganym cieczą, który nie wymaga dopływu wody do szafy serwerowej. W przypadkach bez użycia lub możliwości użycia zewnętrznego chłodziwa[30] sprawdzi się serwer Dell PowerEdge XE8640 w obudowie 4U obsługujący cztery procesory graficzne NVIDIA H100 SXM5, który zapewnia taką samą moc obliczeniową jak model PowerEdge XE9640 bez konieczności bezpośredniego chłodzenia cieczą.[31]

Serwer Dell PowerEdge XE8640 jest wyposażony w najnowsze skalowalne procesory Intel Xeon 4. generacji i do 4 TB pamięci[32] do obsługi dużych zbiorów danych oraz złożonych obliczeń typowych dla sztucznej inteligencji i analizy danych. Również w tym przypadku firma HPE oferuje procesory graficzne NVIDIA H100 SXM5 w systemach HPE Cray, ale serwery HPE ProLiant ich nie obsługują.

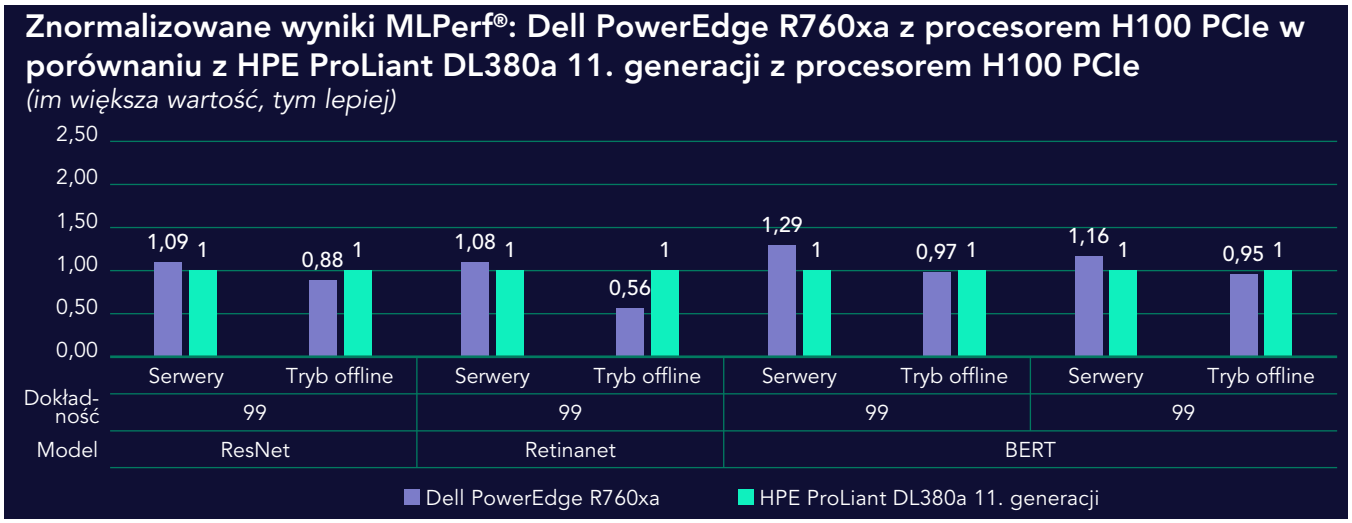
W porównaniu z danymi MLPerf® opublikowanymi w listopadzie 2023 r. serwer PowerEdge XE8640 z czterema procesorami graficznymi NVIDIA H100 SXM5 osiągnął najwyższą przepustowość w zakresie sztucznej inteligencji spośród wszystkich zgłoszonych modeli z czterema procesorami graficznymi w dziewięciu różnych kategoriach. Jak pokazano na rysunku 4, w porównaniu z serwerem HPE ProLiant DL380a uzyskał on nawet 2,07 raza lepszy wynik.



Rysunek 4. Opublikowano wyniki MLPerf® dla serwerów Dell PowerEdge XE8640 i HPE ProLiant DL380a 11. generacji według stanu na dzień 29/11/2023. System firmy Dell korzysta z formatu NVIDIA H100 SXM, podczas gdy system HPE korzysta ze słabszego formatu PCIe. Źródło: Principled Technologies na podstawie danych z MLCommons®.[33,34]

Serwer Dell PowerEdge R760xa o wysokości 2U obsługuje szereg procesorów graficznych firm NVIDIA, AMD i Intel, w tym do czterech procesorów graficznych PCIe 5. generacji o podwójnej szerokości lub 12 procesorów graficznych PCIe o pojedynczej szerokości — idealne rozwiązanie dla organizacji, które poszukują systemu z możliwością skalowania w przyszłości.[35] Model ten jest wyposażony w 32 gniazda DIMM, wnękę na osiem dysków 2,5-calowych i 12 gniazd PCIe, zapewniając skalowalną pamięć masową, która może rosnąć wraz ze wzrostem wymagań dotyczących danych związanych ze sztuczną inteligencją. Ponadto obsługuje do 12 procesorów graficznych PCIe o pojedynczej szerokości lub czterech procesorów graficznych PCIe o podwójnej szerokości, takich jak NVIDIA H100 lub L40S.[36] Ta skalowalność oznacza, że serwer można dostosowywać do zmieniających się zadań sztucznej inteligencji — od trenowania modelu uczenia maszynowego po zaawansowane przetwarzanie danych.

Układ chłodzenia powietrzem serwera PowerEdge R760xa obsługuje środowiska obliczeniowe o dużej gęstości i może obsługiwać akceleratory o wyższej mocy termicznej (TDP) do 350 W.[37] Jest to cecha, która może pomóc w utrzymaniu wydajności przy intensywnych obciążeniach obliczeniowych. W wynikach testów MLPerf® ResNet, RetinaNet i Server opublikowanych w listopadzie 2023 r. w trybie „serwera" model PowerEdge R760xa z czterema procesorami graficznymi NVIDIA H100 PCIe okazał się lepszy od serwera HPE ProLiant DL380a 11. generacji wyposażonego również w cztery procesory graficzne PCIe H100 (patrz rysunek 5).



Rysunek 5. Opublikowano wyniki MLPerf® dla serwerów Dell PowerEdge R760xa i HPE ProLiant DL380a 11. generacji według stanu na dzień 29/11/2023. Oba układy wykorzystują format PCIe procesorów graficznych NVIDIA H100. Źródło: Principled Technologies na podstawie danych z MLCommons®.[38,39]

Podsumowując, wyniki MLPerf® pokazują, że wydajność różni się znacznie w zależności od serwerów i podzespołów, dlatego wybór odpowiednich opcji do obsługi obciążeń i ich wymagań dotyczących wydajności ma kluczowe znaczenie. Serwery Dell PowerEdge do obsługi obciążeń związanych ze sztuczną inteligencją zapewniają wiele opcji chłodzenia i gęstości, które można dopasować do różnorodnych potrzeb centrum danych firmy, zapewniając jednocześnie wysoką wydajność MLPerf®.

Bardziej szczegółowe informacje na temat oferty sztucznej inteligencji firmy Dell

Chociaż wydajność obliczeniowa ma kluczowe znaczenie, jest to tylko jeden z czynników branych pod uwagę podczas planowania obciążeń związanych ze sztuczną inteligencją. Rozpoczynając wdrażanie sztucznej inteligencji, należy również wziąć pod uwagę resztę portfolio rozwiązań SI oferowanych przez dostawcę. Poniżej omówiono dodatkowe kategorie krytyczne dla tych rozwiązań SI, w tym klienckie stacje robocze, produkty natywne dla chmury, pamięć masową i inne. Zwracamy również uwagę na obszary, w których oferta firmy Dell może mieć przewagę nad HPE.

Stacje robocze

Stacje robocze Dell Precision Data Science oferują procesory graficzne NVIDIA RTX™ i procesory Intel Xeon®, a także zestaw narzędzi do analizy danych przeznaczone dla programistów SI i analityków danych.[40] Systemy te wykorzystują profesjonalne opcje obliczeniowe dzięki procesorom graficznym NVIDIA certyfikowanym dla ponad 100 specjalistycznych zastosowań[41] oraz skalowalne akceleratory procesorów Intel Xeon, takie jak Intel DL Boost.[42] Stacje robocze Precision są dostępne w formacie mobilnym, Tower i do montażu w szafie, aby zaspokoić potrzeby od większej, stacjonarnej analizy danych po modelowanie w terenie.

Oferta stacji roboczych HPE jest węższa i obejmuje przede wszystkim pojedyncze stacje robocze w obudowie typu tower wyposażone w NVIDIA L4s; firma HPE nie oferuje opcji mobilnej stacji roboczej.[43] Chociaż produkty te są odpowiednie do wielu zadań, oferta stacji roboczych w obudowie typu tower nie daje takiej samej elastyczności i pokrycia obciążeń jak szerszy zakres rozwiązań oferowany przez firmę Dell. Różnorodność rozmiarów i opcji przenośności stacji roboczych Dell Precision pozwala na tworzenie bardziej dostosowanych rozwiązań, uwzględniających różne potrzeby w takich warunkach, jak laboratoria, biura i działalność w terenie.

Pamięć masowa

Pamięć masowa może być równie ważna jak moc obliczeniowa w kontekście obsługi obciążeń związanych ze sztuczną inteligencją. Więcej danych zwiększa dokładność modelu sztucznej inteligencji, ale przechowywanie ogromnych zbiorów danych i zarządzanie nimi może stanowić wyzwanie dla wielu centrów danych. Ponadto systemy pamięci masowej gotowe do pracy ze sztuczną inteligencją muszą z łatwością obsługiwać wiele różnych typów danych, ponieważ modele są zwykle trenowane przy użyciu danych nieustrukturyzowanych.[44] Aby zapewnić pojemność i skalowalność zestawów danych sztucznej inteligencji, uczenia maszynowego i uczenia głębokiego, firma Dell oferuje serię PowerScale™ do przechowywania plików oraz rozwiązanie Elastic Cloud Storage (ECS) lub definiowane programowo narzędzie ObjectScale do obiektowej pamięci masowej.

Oferta serwerów NAS PowerScale typu all-flash obejmuje opcje pojemności od 3,84 TB do 720 TB pojemności fizycznej na węzeł, przy czym klastrowane pojemności all-flash osiągają 186 PB pojemności fizycznej. Dzięki elastyczności i skalowalności PowerScale może obsługiwać szeroką gamę klientów i przypadków użycia sztucznej inteligencji.[45] W klastrze model PowerScale F900 może osiągnąć do 186 PB całkowitej pojemności.[46] Wszystkie trzy modele PowerScale typu all-flash — F200, F600 i F900 — obsługują wbudowaną kompresję oraz deduplikację danych w celu zwiększenia wydajności pamięci masowej.[47] Każdy model pamięci masowej PowerScale używa systemu plików Dell OneFS™, który wykorzystuje zasady warstwowania pamięci masowej w celu nadania priorytetu najważniejszym danym w najszybszych warstwach w celu optymalizacji obciążeń.[48] Firma Dell oferuje również oprogramowanie OneFS w AWS Marketplace z pamięcią masową Dell APEX File Storage dla platformy AWS. Klienci mogą wykorzystać OneFS w swoich instancjach obliczeniowych AWS, aby uzyskać spójne środowisko użytkownika z tymi samymi funkcjami, które są dostępne w lokalnych macierzach OneFS.[49] Choć firma HPE oferuje integrację z chmurą publiczną dla hybrydowych rozwiązań pamięci masowej, nie znaleźliśmy wśród jej ofert opcji natywnej dla chmury, takiej jak Dell APEX File Storage dla platformy AWS.

Opcje obiektowej pamięci masowej firmy Dell obejmują rozwiązanie Dell Enterprise Object Storage (ECS), które „stworzono specjalnie do przechowywania nieustrukturyzowanych danych w skali chmury publicznej”.[50] Wraz z wbudowaną kompatybilnością z obiektową pamięcią masową Amazon S3 umożliwiającą korzystanie z funkcjonalności chmury hybrydowej węzły pamięci masowej ECS zapewniają pojemność do 14 PB na szafę.[51] Firma HPE oferuje również niestrukturalną pamięć masową z opcjami przechowywania plików i obiektów, ale jej oferta jest realizowana poprzez partnerstwo ze Scality. Klienci mogą zakupić rozwiązania HPE dla Scality od HPE.[52]

Usługi profesjonalne

Firma Dell oferuje szeroki zakres usług profesjonalnych, w tym doradztwo, przygotowywanie danych, wdrażanie, pomoc techniczną i usługi zarządzane wspierające wdrażanie sztucznej inteligencji. Dla organizacji poszukujących sprawdzonych architektur i rozwiązań firma Dell oferuje zweryfikowane projekty sztucznej inteligencji, które są ukierunkowane na konkretne przypadki użycia, aby wyeliminować niepewność podczas projektowania i wdrażania zasobów związanych ze sztuczną inteligencją. Te zatwierdzone przez firmę Dell rozwiązania w zakresie sztucznej inteligencji obejmują pakiety sprzętu i oprogramowania, konwersacyjne modele sztucznej inteligencji, operacje uczenia maszynowego i wiele innych. Łącząc wstępnie skonfigurowane, specjalnie zaprojektowane rozwiązania z usługami związanymi ze sztuczną inteligencją, Dell oferuje kompleksowe rozwiązania w zakresie sztucznej inteligencji w szerokim spektrum potrzeb. Rozwiązania te mogą zapewnić szybszą i łatwiejszą drogę do pomyślnego wdrożenia SI w porównaniu z budowaniem rozwiązań ad-hoc.

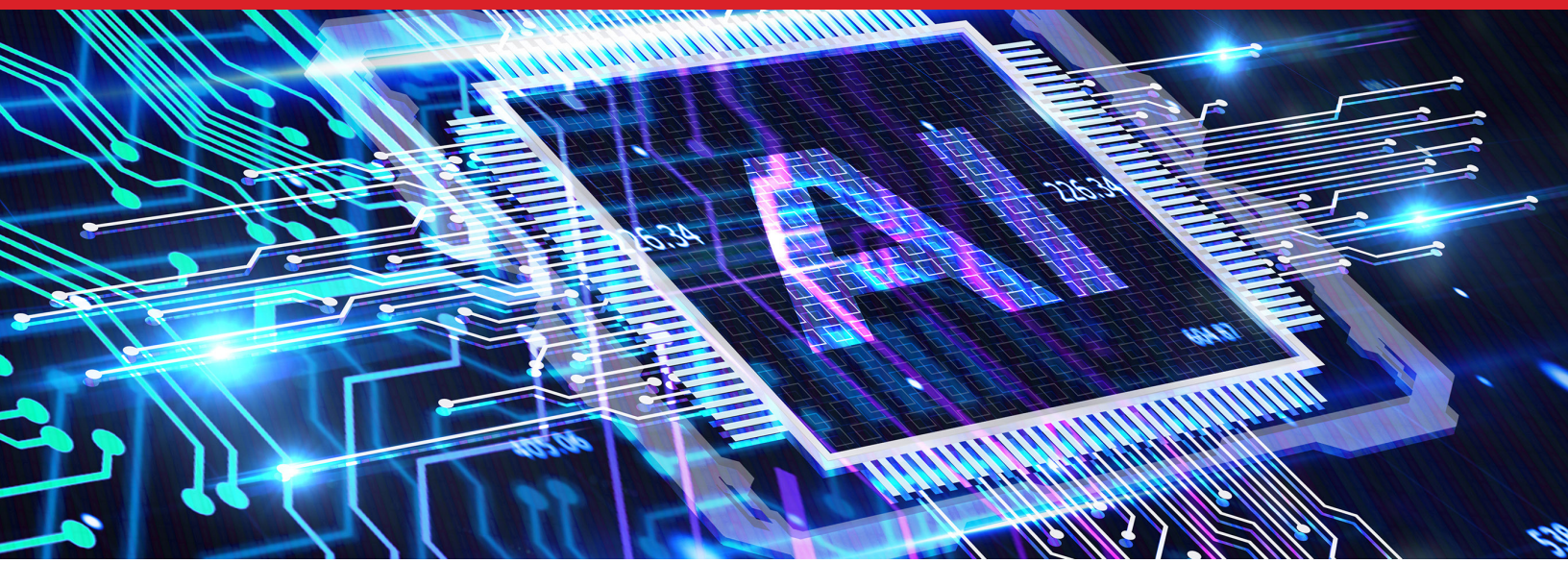
Usługi Dell mogą również pomóc klientowi w całym procesie związanym ze sztuczną inteligencją — od doradztwa do wdrożenia. Usługi doradcze Dell ProConsult pomagają klientom określić, gdzie użytkownicy mogą odnieść korzyści z przyjęcia procesów GenAI i stworzyć mapę drogową obejmującą wymagane rozwiązania oraz umiejętności w zakresie IT. Usługi firmy Dell mogą obejmować przygotowanie danych do integracji z dużym modelem językowym i szkolenie zespołów IT w zakresie wiedzy na temat sztucznej inteligencji. Aby w pełni wdrożyć GenAI, zespoły firmy Dell analizują konkretne przypadki użycia i określają, wdrażają oraz konfigurują najlepszy model sztucznej inteligencji zgodnie z potrzebami. HPE oferuje również profesjonalne usługi wspierające firmy w ich wysiłkach związanych ze sztuczną inteligencją.[53,54]

Zagadnienia dotyczące zarządzania

Serwery wymagają ciągłego zarządzania, które zajmuje czas administratora. Oprogramowanie sprzętowe, oprogramowanie i sterowniki wymagają okresowych aktualizacji; personel IT musi optymalizować i utrzymywać wydajność, poziomy temperatur i wiele więcej. W poprzednich testach firmy Principled Technologies (PT) oceniliśmy możliwości zarządzania serwerami firmy Dell ze zintegrowanym kontrolerem Dell Remote Access Controller 9 (iDRAC9).[55] Administratorzy mogą polegać na automatycznych aktualizacjach online dzięki oprogramowaniu iDRAC9 OpenManage™ Enterprise (OME) z konfigurowalnym planowaniem, aby zapewnić aktualność serwerów i używać profili do szybkiego i łatwego dołączania nowych serwerów w miarę wzrostu obciążenia. Dzięki kontrolerom iDRAC i OME klienci firmy Dell mogą uzyskać dostęp do większej liczby funkcji zdalnego zarządzania, łatwiej wdrażać serwery i aktualizować oprogramowanie wewnętrzne łatwiej niż w przypadku korzystania z rozwiązań HPE OneView oraz HPE iLO. Serwery Dell PowerEdge są dostarczane z usługami i funkcjami zarządzania, które mogą pomóc organizacjom poprzez „ograniczenie czasu i wysiłku związanego z zadaniami takimi jak monitorowanie stanu systemu lub aktualizowanie oprogramowania sprzętowego", zwalniając cykle pracy zespołów IT na innowacje i inne zadania.[56]

Tabela 3. Podsumowanie porównania narzędzi do zarządzania firm Dell i HPE z raportu PT z listopada 2022 r.[57]
Źródło: Principled Technologies.

	Różnice w dostępie do narzędzi do zarządzania firmy Dell	Porównanie
Więcej funkcji zarządzania zdalnego Porównanie kontrolera iDRAC z iLO	Więcej funkcji konfiguracji konsoli HTML5 i systemu BIOS dla większej liczby funkcji zdalnych w kontrolerze iDRAC	2,5 razy więcej funkcji konsoli HTML5 i 13 razy więcej funkcji BIOS
Łatwiejsze wdrażanie serwerów Porównanie OME i OneView	Wdrażanie profilu „jeden do wielu" za pomocą narzędzia OME	O 52% krótszy czas wdrożenia serwera niż w przypadku OneView
Łatwiejsze aktualizacje oprogramowania sprzętowego Porównanie OME i OneView	Automatyczne aktualizacje online za pomocą OME	Aktualizuj wiele serwerów, łącząc się z witryną Dell.com, oszczędzając czas potrzebny na aktualizację serwerów poprzez ręczne przesyłanie pakietów za pomocą OneView
Łatwiejsze powiadamianie Porównanie OME i OneView	Konfigurowanie zasad alertów w OME i wykonywanie działań automatycznych na podstawie alertów	Automatyzacja tego procesu oszczędza czas i zmniejsza ryzyko błędów w porównaniu z ręcznym wykonywaniem działań za każdym razem, gdy otrzymasz alert w OneView
Łatwiejsze w użyciu funkcje bezpieczeństwa (blokada systemu i dynamiczne USB) Porównanie kontrolera iDRAC z iLO	Mniej kroków, mniej czasu, brak konieczności ponownego uruchamiania przy użyciu kontrolera iDRAC	1/4 kroków, o 91% mniej czasu potrzebnego na blokowanie systemu
Bardziej niezawodne narzędzia analityczne CloudIQ dla serwerów PowerEdge w porównaniu z InfoSight	Konfigurowalne raporty i więcej wskaźników stanu zapewniające administratorowi większą kontrolę dzięki CloudIQ dla PowerEdge	Ponad 15 razy więcej wskaźników do wyboru w porównaniu z InfoSight



Wnioski

Wykorzystanie mocy sztucznej inteligencji do usprawnienia działalności gospodarczej może być trudnym zadaniem, które ma poważne implikacje biznesowe. Ponieważ technologia rozwija się szybciej niż kiedykolwiek, kluczowe znaczenie ma współpraca z odpowiednim dostawcą sztucznej inteligencji. Wybierając firmę taką jak Dell, która nie tylko oferuje kompleksowe portfolio sztucznej inteligencji, ale może również świadczyć usługi planowania, przygotowania, wdrażania i zarządzania, klienci mogą stawić czoła tym wyzwaniom. Testy porównawcze MLPerf® pokazują, że oferty z gamy rozwiązań sztucznej inteligencji firmy Dell zapewniają stałą, wysoką wydajność obciążeń związanych ze sztuczną inteligencją. Dzięki wydajnym i elastycznym opcjom serwerów, a także wielu opcjom pamięci masowej, sprawdzonym rozwiązaniom i profesjonalnym usługom dostosowanym specjalnie do sztucznej inteligencji firma Dell może pomóc firmom w wykorzystaniu sztucznej inteligencji oraz jej zalet.

---

1. Dell, „Increasing Your Data Value with Dell Generative AI Solutions”, dostęp 19 grudnia 2023 r., https://www.dell.com/en-us/blog/increasing-your-data-value-with-dell-generative-ai-solutions/.
2. Dell, „Dell AI solutions”, dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/dt/solutions/artificial-intelligence/index.htm#accordion0&tab0=0.
3. Dell, „Dell Technologies and Hugging Face to Simplify Generative AI with On-Premises IT”, dostęp 12 grudnia 2023 r., 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#/filter-on/Country:en-us.
4. Dell, „Dell and Meta Collaborate to Drive Generative AI Innovation”, dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/blog/dell-and-meta-collaborate-to-drive-generative-ai-innovation/.
5. Dell, „Dell AI-Ready Data Platform”, dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/dt/solutions/artificial-intelligence/storage-for-ai.htm?have=explore+unstructured+storage#tab0=0.
6. Dell, „Snowflake and Dell Partnership Gains Tempom”, dostęp 19 grudnia 2023 r., https://www.dell.com/en-us/blog/snowflake-and-dell-partnership-gains-momentum/.
7. Robert McNeal, „Dell, VMware and NVIDIA Bring AI to Your Data”, dostęp 17 stycznia 2024 r., https://www.dell.com/en-us/blog/dell-vmware-and-nvidia-bring-ai-to-your-data/. Pod powyższym linkiem: „Na podstawie analizy firmy Dell, sierpień 2023 r. Firma Dell Technologies oferuje rozwiązania przeznaczone do obsługi zadań sztucznej inteligencji — od stacji roboczych (stacjonarnych i mobilnych) po serwery w zakresie HPC, przechowywania danych, infrastruktury chmurowej definiowanej programowo, przełączników sieciowych, ochrony danych, HCI oraz usług”.
8. Intel, „Accelerate Artificial Intelligence (AI) Workloads with Intel Advanced Matrix Extensions (Intel AMX)”, dostęp 12 grudnia 2023 r., https://www.intel.com/content/www/us/en/content-details/785250/accelerate-artificial-intelligence-ai-workloads-with-intel-advanced-matrix-extensions-intel-amx.html.

9. Vipera, „NVIDIA's H100 and A100 GPU Cards: Exploring the Intricacies of SXM and PCI-E Connections", dostęp 12 grudnia 2023 r., https://www.viperatech.com/unraveling-the-mysteries-sxm-vs-pci-e-connections-in-nvidias-high-end-h100-and-a100-gpus/.
10. GitHub, „MLPerf® Results Messaging Guidelines", dostęp 16 stycznia 2024 r., https://github.com/mlcommons/policies/blob/master/MLPerf_Results_Messaging_Guidelines.adoc.
11. MLCommons®, „MLPerf® Inference: Datacenter Benchmark Suite Results", dostęp 12 grudnia 2023 r., https://mlcommons.org/en/inference-datacenter-31/.
12. Dell, „PowerEdge XE Servers", dostęp 12 grudnia 2023 r . https://www.dell.com/en-us/dt/servers/specialty-servers/poweredge-xe-servers.htm?have=explore+poweredge+xe#tab0=0.
13. HPE, „HPE ProLiant XL675d Gen10 Plus Configure-to-order Server", dostęp 12 grudnia 2023 r., https://www.hpe.com/us/en/product-catalog/compute/proliant-servers/.1013142988.html.
14. HPE, „HPE ProLiant DL380a Gen11", dostęp 12 grudnia 2023 r., https://www.hpe.com/us/en/product-catalog/compute/proliant-servers/.proliant-dl380-server.1014696168.html.
15. MLCommons®, „MLPerf® Inference: Datacenter Benchmark Suite Results v 3.1", dostęp 12 grudnia 2023 r., https://mlcommons.org/benchmarks/inference-datacenter/.
16. HPE, „NVIDIA Acceleratory for HPE ProLiant Servers", dostęp 12 grudnia 2023 r., https://www.hpe.com/psnow/doc/c04123180.html?jumpid=in_pdp-psnow-qs.
17. Dell, „PowerEdge XE9680 Specification Sheet", dostęp 19 stycznia 2024 r., https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9680-spec-sheet.pdf.
18. HPE, „HPE & NVIDIA financial services solution sets new records in performance", dostęp 12 grudnia 2023 r., https://community.hpe.com/t5/alliances/hpe-amp-nvidia-financial-services-solution-sets-new-records-in/ba-p/7197388.
19. HPE, „QuickSpecs: HPE Cray Supercomputing XD670", dostęp 12 grudnia 2023 r., https://www.hpe.com/psnow/doc/a50004292enw.
20. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0069. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
21. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0085. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
22. Dell, „PowerEdge XE9640 Rack Server", dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640.
23. Accelsius, „Enabling the AI Revolution with Liquid Cooling", dostęp 12 grudnia 2023 r., https://www.accelsius.com/blog/enabling-the-ai-revolution-with-liquid-cooling.
24. Dell, „Dell PowerEdge XE9640 Technical Guide", dostęp 12 grudnia 2023 r., https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9640-technical-guide.pdf.
25. HPE, „HPE ProLiant DL380a Gen11", dostęp 12 grudnia 2023 r., https://www.hpe.com/psnow/doc/PSN1014696168WWEN.pdf?jumpid=in_pdp-psnow-dds.
26. Intel, „Intel® Data Center GPU Max Series Technical Overview", dostęp 12 grudnia 2023 r., https://www.intel.com/content/www/us/en/developer/articles/technical/intel-data-center-gpu-max-series-overview.html#gs.08874l.
27. HPE, „Intel Data Center GPU Max 1100 48GB Accelerator for HPE Data sheet", dostęp 12 grudnia 2023 r., https://www.hpe.com/psnow/doc/PSN1014779728WWEN.
28. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0066. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
29. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0084. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.

30. Dell, „AI and HPC —With Air or Liquid Cooling", dostęp 12 grudnia 2023 r., https://www.delltechnologies.com/asset/en-us/products/servers/briefs-summaries/poweredge-xe9640-and-xe8640-infographic.pdf.
31. Dell, „PowerEdge XE8640 : Drive AI, HPC modeling and simulation workloads with superior performance", dostęp 12 grudnia 2023 r., https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe8640-spec-sheet.pdf.
32. Dell, „PowerEdge XE8640 Rack Server", dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/shop/ipovw/poweredge-xe8640.
33. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0067. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
34. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0084. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
35. Dell, „PowerEdge R760xa Rack Server", dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/shop/dell-poweredge-servers/poweredge-r760xa-rack-server/spd/poweredge-r760xa/pe_r760xa_16902_vi_vp#features_section.
36. SANStorageWorks, „Dell EMC PowerEdge R760xa: Powerful and scalable for GPU workloads", dostęp 12 grudnia 2023 r., https://www.sanstorageworks.com/PowerEdge-R760xa.asp.
37. Dell, „Dell PowerEdge Servers and NVIDIA GPUs", dostęp 12 grudnia 2023 r., https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-inferencing/dell-poweredge-servers-and-nvidia-gpus-1/.
38. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0064. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
39. Zweryfikowany wynik MLPerf® w wersji 3.1 — wnioskowanie zamknięte. Pobrane z https://mlcommons.org/benchmarks/inference-datacenter/ 5 grudnia 2023 r., wpis 3.1-0084. Nazwa i logo MLPerf® są zarejestrowanymi i niezarejestrowanymi znakami towarowymi MLCommons® Association w Stanach Zjednoczonych i innych krajach. Wszelkie prawa zastrzeżone. Użycie bez upoważnienia jest surowo wzbronione. Więcej informacji można znaleźć na stronie www.mlcommons.org.
40. Dell, „Workstations for AI", dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/dt/ai-technologies/index.htm?have=explore+dell+precision+for+ai#pdf-overlay=//www.delltechnologies.com/asset/en-us/products/workstations/briefs-summaries/ai-industry-brochure.pdf.
41. NVIDIA, „NVIDIA RTX in Professional Workstations", dostęp 12 grudnia 2023 r., https://www.nvidia.com/en-us/design-visualization/desktop-graphics/.
42. Intel, „Intel® Deep Learning Boost (Intel® DL Boost)", dostęp 12 grudnia 2023 r., https://www.intel.com/content/www/us/en/artificial-intelligence/deep-learning-boost.html.
43. HPE, „HPE ProLiant ML350 Gen11", dostęp 12 grudnia 2023 r., https://buy.hpe.com/us/en/compute/tower-servers/proliant-ml300-servers/proliant-ml350-server/hpe-proliant-ml350-gen11/p/1014696172.
44. ComputerWeekly.com, „Storage requirements for AI, ML and analytics in 2022", dostęp 12 grudnia 2023 r., https://www.computerweekly.com/feature/Storage-requirements-for-AI-ML-and-analytics-in-2022.
45. Dell, „PowerScale AI-Ready Data Platform", dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/shop/powerscale-family/sf/powerscale.
46. Dell, „Compare PowerScale", dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/shop/powerscale-family/sf/powerscale#compare-module.
47. Dell, „Dell PowerScale All-Flash", dostęp 12 grudnia 2023 r., https://www.delltechnologies.com/asset/en-us/products/storage/technical-support/h15963-ss-powerscale-all-flash-nodes.pdf.
48. Dell, „Dell PowerScale OneFS Software Features", dostęp 12 grudnia 2023 r., https://www.delltechnologies.com/asset/en-us/products/storage/technical-support/h18275-onefs-software-features-data-sheet.pdf.
49. Dell, „Dell APEX File Storage for AWS", dostęp 12 grudnia 2023 r., https://www.delltechnologies.com/asset/en-us/products/storage/briefs-summaries/h19575-so-apex-file-storage-for-aws.pdf.

50. Dell, „Dell ECS Enterprise Object Storage”, dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/dt/storage/ecs/index.htm?have=explore+ecs#tab0=0&tab1=0.
51. Dell, „Dell ECS Enterprise Object Storage”, dostęp 12 grudnia 2023 r., https://www.dell.com/en-us/dt/storage/ecs/index.htm#tab0=0&tab1=0&accordion0.
52. HPE, „Storage Solutions for Scality”, dostęp 12 grudnia 2023 r., https://www.hpe.com/us/en/storage/file-object/scality.html.
53. HPE, „Make AI work for you”, dostęp 16 stycznia 2024 r., https://www.hpe.com/us/en/solutions/ai-artificial-intelligence.html.
54. HPE, „HPE AI Services – Generative AI Implementation”, dostęp 16 stycznia 2024 r., https://www.hpe.com/us/en/services/generative-ai-implementation-service.html.
55. Principled Technologies, „Simplify administrator tasks and improve security and health monitoring with tools from the Dell management portfolio vs. comparable tools from HPE”, dostęp 12 grudnia 2023 r., https://www.principledtechnologies.com/Dell/Management-tools-vs-HPE-1122.pdf.
56. Principled Technologies, „Simplify administrator tasks and improve security and health monitoring with tools from the Dell management portfolio vs. comparable tools from HPE”, dostęp 12 grudnia 2023 r., https://www.principledtechnologies.com/Dell/Management-tools-vs-HPE-1122.pdf.
57. Principled Technologies, „Simplify administrator tasks and improve security and health monitoring with tools from the Dell management portfolio vs. comparable tools from HPE”.

Zobacz oryginalną, angielską wersję tego raportu na stronie https://facts.pt/zPmSx4c

Projekt ten został zlecony przez Dell Technologies.

Facts matter.®