

# Rozwiązania Open Ethernet firmy Dell Technologies do generatywnej sztucznej inteligencji

## Poruszanie się w nowych granicach w infrastrukturze IT

„Według prognoz obroty na rynku infrastruktury sztucznej inteligencji (przełączanie zaplecza pod kątem komunikacji GPU–GPU) wzrosną z 1,2 mld USD (2022 r.) do 15,2 mld USD (2027 r.), co oznacza pięcioletni skumulowany roczny wskaźnik wzrostu (CAGR) na poziomie 65%.

Przewiduje się, że do roku 2027 produkty i usługi związane z Ethernetem będą stanowić 32% całkowitych przychodów i 37% całkowitej liczby wysyłek portów na rynku infrastruktury sztucznej inteligencji”.

### Badanie Dell’Oro<sup>1</sup>

Wydajność GPU w dużym stopniu zależy od wydajności sieci. Ponieważ wiele obciążeń roboczych związanych ze sztuczną inteligencją działa na dużych klastrach serwerów, które wymagają stałej komunikacji między węzłami obliczeniowymi i pamięcią masową, potrzebują one solidnej sieci, aby uniknąć wąskich gardeł. Jeśli wydajność sieci jest niewystarczająca dla obciążenia, procesory GPU przestaną działać, a czas trenowania i wnioskowania wydłuży się, spowalniając przetwarzanie danych i czas uzyskiwania szczegółowych informacji.

### Wprowadzenie: wymagania sieciowe generatywnej sztucznej inteligencji

Ponieważ rozwiązania generatywnej sztucznej inteligencji (GenAI) wciąż ewoluują, przesuując granice przetwarzania danych i potrzeb obliczeniowych, infrastruktury IT stają przed wyzwaniem znalezienia sposobów na zaspokojenie ogromnych wymagań tych środowisk. Modele te, zwłaszcza duże modele językowe (LLM), wymagają nie tylko większej infrastruktury, ale także starannie zaprojektowanych systemów, aby sprostać ogromnym potrzebom w zakresie łączności między klastrami GPU. Tradycyjne rozwiązania sieciowe szybko stają się wąskimi gardłami, zagrażając rentowności i powodzeniu inicjatywy generatywnej sztucznej inteligencji. Infrastruktura sztucznej inteligencji wymaga niskich opóźnień, wydajności bez strat i maksymalnej przepustowości.

Ogromne wymagania dotyczące przetwarzania danych i aplikacji powodują wzrost wymagań dotyczących zarówno infrastruktury front-end, jak i zaplecza.

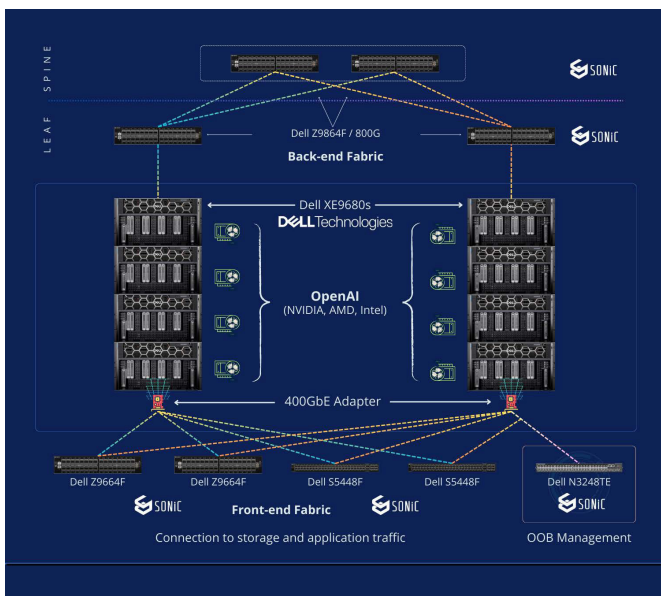
### Wyzwania i potrzeby związane z wdrażaniem infrastruktury generatywnej sztucznej inteligencji

Wdrażanie technologii generatywnej sztucznej inteligencji wiąże się z szeregiem wyzwań, od złożoności technicznej związanej z nowymi architekturami po niedobór wykwalifikowanych specjalistów zdolnych do zarządzania takimi wdrożeniami. Rozwiązania oparte na zastrzeżonych technologiach, takich jak InfiniBand, zwiększają złożoność, ograniczając dostępność zasobów i komplikując integrację z istniejącymi platformami monitorowania lub planowania. Ponadto wysokie koszty, długi czas oceny i uzależnienie od dostawcy związane z zastrzeżonymi rozwiązaniami stanowią istotne bariery, zwłaszcza w dobie niepewności łańcucha dostaw. Wyzwania te wskazują na pilną potrzebę otwartych, elastycznych i solidnych rozwiązań infrastrukturalnych generatywnej sztucznej inteligencji, które mogą sprostać wyjątkowym wymaganiom jej obciążeń roboczych.

### Podejście firmy Dell Technologies do sieci generatywnej sztucznej inteligencji

W odpowiedzi na te wyzwania firma Dell Technologies wprowadziła pionierskie kompleksowe i otwarte rozwiązania oparte na sieci Ethernet, które spełniają zaawansowane wymagania infrastruktury generatywnej sztucznej inteligencji. Korzystając ze swojego bogatego doświadczenia w zakresie sztucznej inteligencji, modelowania i środowisk obliczeń o wysokiej wydajności (HPC), firma Dell Technologies oferuje pakiet rozwiązań, które spełniają zarówno wymagania front-end, jak i zaplecza. Firma Dell Technologies zapewnia podstawowe elementy pomyślnego wdrożenia generatywnej sztucznej inteligencji – od modułowych systemów obliczeniowych zoptymalizowanych pod kątem akceleracji, takich jak serwery Dell PowerEdge XE, po rozwiązania pamięci masowej skoncentrowane na sztucznej inteligencji, takie jak PowerScale. Centralnym elementem tego podejścia jest wdrożenie infrastruktury Ethernet nowej generacji z zaawansowanymi sieciami krzemowymi. Dzięki **przełącznikowi Dell PowerSwitch Z9864-ON** zapewniamy wydajność sieci bez blokad o przepustowości **800 GbE**, która ma kluczowe znaczenie dla zastosowań generatywnej sztucznej inteligencji, klienci mogą wdrażać klastry sztucznej inteligencji z małymi opóźnieniami i wysoką przepływnością przy użyciu przełączania o dużej przepustowości i nowych funkcji dostępnych w **Enterprise SONiC Distribution by Dell Technologies firmy Dell**, takich jak Advanced Routing, RoCEv2, Enhanced Hashing i Priority Flow Control, w celu zwiększenia wydajności infrastruktury i lepszego monitorowania zagęszczenia.

<sup>1</sup> Opis Dell’Oro: Raport z zaawansowanych badań nad sieciami sztucznej inteligencji dla obciążeń roboczych związanych ze sztuczną inteligencją



Przykład architektury infrastruktury generatywnej sztucznej inteligencji

Przełączniki Dell PowerSwitch serii Z wykorzystują najnowocześniejszy krzem, tworząc szkielet wydajnej, skalowalnej sieci, która może obsługiwać tysiące węzłów, rozwiązując w ten sposób problemy związane z połączeniami typowe dla zastosowań generatywnej sztucznej inteligencji.

### Przyspieszone wdrażanie generatywnej sztucznej inteligencji dzięki rozwiązaniom firmy Dell Technologies

Gwałtowny wzrost popularności generatywnej sztucznej inteligencji przyniósł ze sobą szereg wyzwań dla infrastruktury IT, które wymagają nowego, zarówno innowacyjnego, jak i elastycznego podejścia do sieci. Firma Dell Technologies odpowiada na tę sytuację, oferując otwarte rozwiązania oparte na sieci Ethernet, które nie tylko spełniają najpilniejsze potrzeby związane z wdrażaniem generatywnej sztucznej inteligencji, ale także kładą podwaliny pod przyszłe postępy.

Aby wyeliminować niepewność w przypadku rozwiązań sprzętowych z zakresu sztucznej inteligencji, firma Dell oferuje architektury referencyjne zatwierdzone w laboratoriach, zoptymalizowane pod kątem obciążeń roboczych związanych ze sztuczną inteligencją. Te projekty Validated Design obejmują koncepcje architektoniczne, pełne przeglądy rozwiązań, weryfikacje wydajności i inne weryfikacje laboratoryjne potwierdzające możliwości rozwiązania w obciążeniu roboczym, dla którego zostało ono zaprojektowane. Sprawdzone rozwiązania, które ułatwiają uzyskiwanie informacji szybciej i dokładniej, umożliwiają przejście od potencjalnej do działającej sztucznej inteligencji.

Wybierając firmę Dell Technologies, organizacje zyskują partnera z wiedzą ekspercką, kompleksowym zestawem rozwiązań i zaangażowaniem w realizację inicjatyw związanych z generatywną sztuczną inteligencją. Dzięki rozwiązaniom Dell Technologies firmy są przygotowane do poruszania się po złożonych architekturach generatywnej sztucznej inteligencji, dzięki czemu ich projekty są nie tylko rentowne, ale i ukierunkowane na powodzenie.

Firma Dell Technologies przyspiesza wdrażanie i zapewnia szybki zwrot z inwestycji w środowiskach generatywnej sztucznej inteligencji, zmniejszając ryzyko i złożoność operacyjną. Zapraszamy do zapoznania się z tym, w jaki sposób otwarte, elastyczne i zrównoważone rozwiązanie sieciowe może przekształcić inicjatywy związane z generatywną sztuczną inteligencją, wprowadzając firmę w nową erę innowacji i wydajności.

### Innowacje firmy Dell Technologies w zakresie sieci generatywnej sztucznej inteligencji

Firma Dell Technologies stoi na czele innowacji sieciowych generatywnej sztucznej inteligencji, oferując rozwiązania, które spełniają wymagania dzisiejszych i przyszłych środowisk generatywnej sztucznej inteligencji – od brzegu przez rdzeń do chmury. Koncentrując się na otwartych i rozszerzalnych rozwiązaniach, wykorzystując technologię Merchant Silicon i oprogramowanie na bazie open source, firma Dell Technologies zapewnia maksymalną wydajność i elastyczność.

Wykorzystanie dostępnych na rynku sieciowych systemów operacyjnych typu open source, takich jak SONiC, w połączeniu z aktywnym udziałem i wkładem firmy Dell Technologies w **Ultra Ethernet Consortium (UEC)**, wskazuje na jej zaangażowanie na rzecz otwartych standardów i wspólnego rozwoju w obszarze Ethernetu. Wysiłki te służą temu, aby Ethernet nadal odgrywał kluczową rolę we wspieraniu środowisk sztucznej inteligencji nowej generacji.



Dowiedz się więcej o Dell Networking



Skontaktuj się z ekspertem firmy Dell Technologies



Przeczytaj artykuł analityczny ESG



Przeczytaj artykuł analityczny IDC