

Dell Technologies öppna Ethernet-lösningar för generativ AI

Navigera över nya gränser inom IT-infrastruktur

”AI-strukturer (backend-switch för GPU-till-GPU-anslutningsmöjligheter) förväntas växa från \$ 1,2 miljarder (2022) till \$ 15,2 miljarder (2027), vilket ger en genomsnittlig avkastning på 65 % över fem år.

Ethernet förväntas uppnå en intäktsandel på 32 % och en porttrafik på 37 % för AI-strukturer (2027)”

Dell'Oro Research¹

GPU-prestanda är i hög grad beroende av nätverksprestanda. Eftersom många AI-arbetsbelastningar körs på stora kluster av servrar som kräver konstant kommunikation mellan beräkningsnoder och lagring behöver de en robust nätverkshandling för att undvika flaskhalsar. Om nätverksprestandan inte räcker till för arbetsbelastningen blir GPU:erna inaktiva och tiderna för utbildning och inferens ökar, vilket gör databehandlingen långsammare och tiden till insikt längre.

Inledning: Nätverkskrav för generativ AI

I takt med att lösningarna för generativ AI (GenAI) fortsätter att utvecklas och tänjer på gränserna för databehandlings- och beräkningsbehov, utmanas IT-infrastrukturer att hitta sätt att stödja de enorma krav som ställs i dessa miljöer. Dessa modeller, särskilt stora språkmodeller (LLM), behöver inte bara bättre infrastruktur utan även noggrant utformade system för att hantera de enorma anslutningsbehoven som GPU-kluster har. Traditionella nätverkslösningar utvecklar flaskhalsar i hög takt, vilket hotar lönsamheten och framgången för GenAI-initiativ. För AI-strukturer krävs låg latens, förlustfri prestanda och maximal bandbredd.

Dessa höga krav på data- och programbearbetning driver fram ökade krav på både frontend- och backend-strukturer.

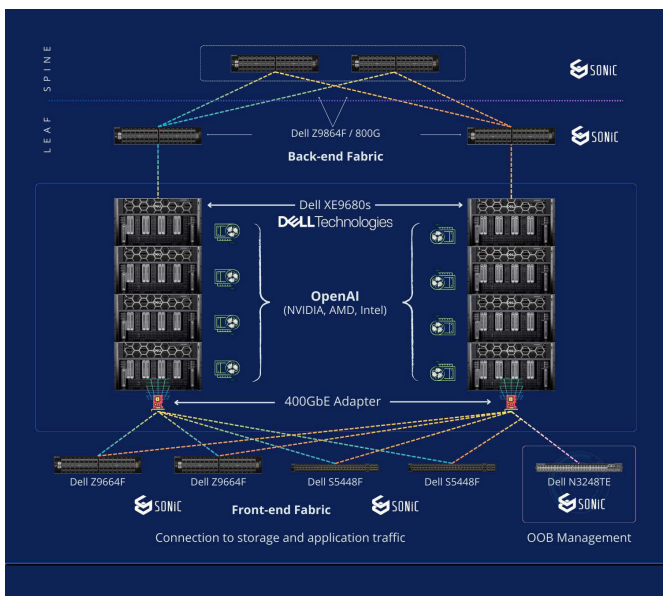
Utmaningar och behov vid distribution av GenAI-infrastruktur

Att installera GenAI-tekniken medför en rad utmaningar, från den tekniska komplexitet som är förknippad med nya arkitekturer till bristen på kvalificerade yrkespersoner som kan hantera sådana distributioner. Lösningar som förlitar sig på varumärkesskyddade tekniker, som t.ex. InfiniBand, tillför ytterligare ett lager komplexitet, begränsar resurstillgängligheten och gör integreringen mer med befintliga övervaknings- eller orkestreringsplattformar mer komplicerad. Dessutom utgör de höga kostnaderna, de långa utvärderingstiderna och leverantörslåsningen i samband med varumärkesskyddade lösningar betydande hinder, särskilt i en tid med så mycket osäkerhet i leveranskedjan. Dessa utmaningar understryker det brådskande behovet av öppna, flexibla och robusta GenAI-infrastrukturlösningar som kan tillgodose de unika kraven som GenAI-arbetsbelastningarna innebär.

Dell Technologies strategi för GenAI Networking

Som svar på dessa utmaningar har Dell Technologies banat väg för omfattande och öppna Ethernet-drivna lösningar som utformats för att uppfylla de intrikata kraven på GenAI-infrastruktur. Dell Technologies utnyttjar sin omfattande erfarenhet av AI, modellering och HPC-beräkningsmiljöer (HighPerformance Computing) och erbjuder en serie lösningar som uppfyller både frontend- och backend-kraven. Dell Technologies tillhandahåller de grundläggande komponenterna för en framgångsrik GenAI-distribution – från modulära beräkningssystem som är optimerade för acceleration, som Dell PowerEdge XE-servrar, till AI-fokuserade lagringslösningar som PowerScale. Centralt för detta tillvägagångssätt är distributionen av nästa generations Ethernet-strukturer med avancerade nätverkschips. Eftersom **Dell PowerSwitch Z9864-ON** levererar icke-blockerande nätverksprestanda på **800 GbE**, vilket är avgörande för GenAI-program, kan kunder distribuera AI-kluster med låg latens och hög genomströmning med hjälp av hög bandbreddsväxling och de nya funktioner som finns i **Dells Enterprise SONiC Distribution by Dell Technologies**, till exempel avancerad routning, RoCEv2, förbättrad hashning och prioritetsflödeskontroll, för förbättrad strukturprestanda och bättre övervakning av överbelastning.

¹Oversikt från Dell'Oro: Advanced Research Report on AI Networks for AI Workloads



Exempel på arkitektur med GenAI-struktur

Innovationer inom GenAI-nätverkshantering från Dell Technologies

Dell Technologies ligger i framkant när det gäller GenAI-nätverksinnovation och erbjuder lösningar som uppfyller kraven för GenAI-miljöer idag och imorgon, från kant till kärna till molnet. Genom att fokusera på öppna och utökningsbara lösningar och användande av mjukvara från detaljhandeln och öppen källkod kan Dell Technologies garantera maximal prestanda och flexibilitet.

Användningen av kommersiellt tillgängliga nätverksoperativsystem baserade på öppen källkod som SONiC, tillsammans med Dell Technologies aktiva deltagande i och bidrag till **Ultra Ethernet Consortium (UEC)**, understryker företagets engagemang för öppna standarder och gemensam utveckling inom Ethernet-området. Dessa insatser syftar till att säkerställa att Ethernet fortsätter att spela en central roll för att stödja nästa generations AI-miljöer.

Dell PowerSwitch Z-serien, som använder toppmoderna chip, utgör stamnätet för ett högpresterande, skalbart nätverk med stöd för tusentals noder och möter på så sätt anslutningsutmaningarna som är inneboende i GenAI-program.

Förstärka GenAI-distributionen med Dell Technologies

Den ökade användningen av GenAI har medfört en mängd utmaningar för IT-infrastrukturer, vilket leder till att nya innovativa såväl som flexibla nätverkshanteringsmetoder nu behövs. Dell Technologies svarar på detta genom att tillhandahålla öppna Ethernet-drivna lösningar som inte bara tillgodoser de omedelbara behoven hos GenAI-distributioner utan även lägger grunden för framtida framsteg.

För att eliminera behovet av att gissa vilka AI-hårdvarulösningar som behövs erbjuder Dell labbvaliderade referensarkitekturer som är optimerade för AI-arbetsbelastningar. Dessa validerade utformningar omfattar arkitekturkoncept, fullständiga lösningsöversikter, prestandavalideringar och andra labbvalideringar som visar att lösningen har kapacitet för den arbetsbelastning som den har utformats för. Gå från AI-möjlig till AI-bevisad med validerade lösningar som gör det enklare att leverera snabbare och djupare insikter.

Genom att välja Dell Technologies får organisationer en partner med den kompetens som behövs, en heltäckande lösningsstack och det nödvändiga engagemanget för att säkerställa framgången för sina GenAI-initiativ. Med Dell Technologies är företag utrustade för att hantera komplexiteten i GenAI-arkitekturer och säkerställa att deras projekt inte bara är genomförbara utan även positionerade för framgång.

Accelerera distribution och tid till värde för dina GenAI-miljöer, minska riskerna och driftskomplexiteten med Dell Technologies. Ta chansen och utforska hur en öppen, flexibel och hållbar nätverkslösning kan omvandla dina GenAI-initiativ och driva ditt företag in i en ny era av innovation och effektivitet.



Mer information
om Dell
Networking



Kontakta en Dell
Technologies-expert



Läs
analysöversikten
från ESG



Läs
analysöversikten
från IDC