DELL Technologies

Resumo da solução Dell Networking

Soluções abertas de Ethernet da Dell Technologies para IA generativa

Navegando por novas fronteiras na infraestrutura de TI

"A previsão é de que o fabric de IA (switches de back-end para conectividade de GPU para GPU) cresça de US$ 1,2 bilhão (em 2022) para US$ 15,2 bilhões (em 2027), com CAGR de cinco anos de 65%.

Espera-se que a Ethernet atinja 32% de participação na receita e 37% de remessa de portas para fabrics de IA (em 2027)"

Pesquisa da Dell'Oro[1]

O desempenho da GPU está intimamente ligado ao desempenho da rede. Com muitas cargas de trabalho de IA em execução em grandes clusters de servidores que exigem comunicação constante entre nós de computação e armazenamento, é necessário ter um sistema de rede eficiente para evitar gargalos. Se o desempenho do sistema de rede for insuficiente para a carga de trabalho, as GPUs ficarão ociosas e os tempos de treinamento e inferência aumentarão, retardando o processamento de dados e o tempo para obter insights.

Introdução: Demandas de rede da IA generativa

À medida que as soluções de IA generativa (GenAI) continuam a evoluir, expandindo os limites do processamento de dados e as necessidades computacionais, as infraestruturas de TI precisam encontrar maneiras de assumir os imensos requisitos desses ambientes. Esses modelos, especialmente grandes modelos de linguagem (LLMs), exigem mais infraestrutura e também sistemas cuidadosamente arquitetados para gerenciar as enormes necessidades de conectividade em clusters da GPU. As soluções de rede tradicionais estão rapidamente se tornando gargalos, ameaçando a viabilidade e o sucesso das iniciativas de GenAI. Os fabrics de IA exigem baixa latência, desempenho sem perdas e o máximo de largura de banda.
Os enormes requisitos de processamento de dados e aplicativos aumentam a necessidade de fabrics de front-end e back-end.

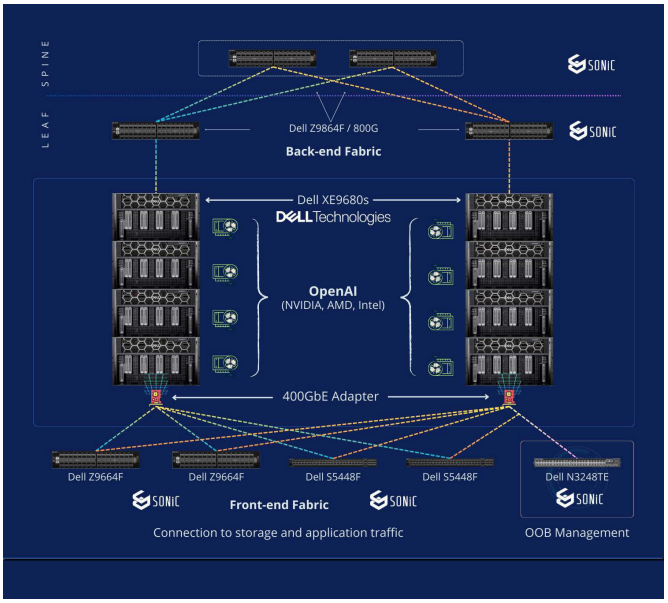
Desafios e necessidades na implementação de infraestrutura de GenAI

A implementação de tecnologias de GenAI apresenta uma série de desafios, desde as complexidades técnicas associadas às novas arquiteturas até a escassez de profissionais qualificados capazes de gerenciar tais implementações. Soluções que dependem de tecnologias próprias, como InfiniBand, adicionam mais uma camada de complexidade, limitando a disponibilidade dos recursos e complicando a integração com plataformas existentes de monitoramento ou orquestração. Além disso, os altos custos, os longos tempos de avaliação e a restrição do fornecedor associados às soluções próprias representam barreiras significativas, especialmente em uma era de incertezas na cadeia de suprimentos. Esses desafios enfatizam a necessidade urgente de soluções de infraestrutura de GenAI abertas, flexíveis e eficientes que possam acomodar as demandas exclusivas das cargas de trabalho de GenAI.

Abordagem da Dell Technologies ao sistema de rede de GenAI

Em resposta a esses desafios, a Dell Technologies foi pioneira em soluções abrangentes e abertas com tecnologia Ethernet criadas para atender às complexas demandas da infraestrutura de GenAI. Aproveitando a ampla experiência em ambientes de IA, modelagem e computação com alto desempenho (HPC), a Dell Technologies oferece uma suíte de soluções que atendem aos requisitos tanto de front-end quanto de back-end. De sistemas computacionais modulares otimizados para aceleração, como os servidores Dell PowerEdge XE, até soluções de armazenamento com foco em IA, como o PowerScale, a Dell Technologies fornece os componentes essenciais para uma implementação bem-sucedida de GenAI.
O fundamento dessa abordagem é a implementação de fabrics Ethernet de última geração com tecnologia de silício de rede avançado. Graças aos **800 GbE** de desempenho de rede sem bloqueios essenciais para aplicativos de GenAI fornecidos pelo **Dell PowerSwitch Z9864-ON**, os clientes podem implementar clusters de IA com baixa latência e alto throughput usando switches de grande largura de banda e novos recursos encontrados na **Enterprise SONiC Distribution by Dell Technologies**, como roteamento avançado, RoCEv2, hashing aprimorado e controle de fluxo prioritário. Isso melhora o desempenho do fabric e o monitoramento de congestionamentos.

[1] Artigo da Dell'Oro: Advanced Research Report on AI Networks for AI Workloads.



Exemplo de arquitetura de fabric de GenAI

Inovações em sistema de rede de GenAI da Dell Technologies

A Dell Technologies está na vanguarda de inovações em sistema de rede de GenAI, oferecendo soluções que atendem aos requisitos dos ambientes de GenAI atuais e futuros, da borda ao núcleo e à nuvem. Ao focar em soluções abertas e extensíveis, aproveitando o silício do comerciante e software baseado em código aberto, a Dell Technologies garante flexibilidade e desempenho máximos.

O uso de sistemas operacionais de rede de código aberto disponíveis comercialmente, como o SONiC, em conjunto com a contribuição e a participação ativa da Dell Technologies no **Ultra Ethernet Consortium (UEC)**, ressaltam o compromisso com padrões abertos e desenvolvimento colaborativo no espaço Ethernet. O objetivo desses esforços é garantir que a Ethernet continue desempenhando um papel fundamental no suporte à última geração de ambientes de IA.

O Dell PowerSwitch Série Z usa silício de última geração e fornece a estrutura para uma rede escalável e de alto desempenho compatível com milhares de nós, resolvendo, assim, os desafios de conectividade inerentes aos aplicativos de GenAI.

Acelerando a implementação de GenAI com a Dell Technologies

O aumento da GenAI trouxe uma série de desafios para as infraestruturas de TI, exigindo uma nova abordagem de sistemas de rede que seja inovadora e flexível. A Dell Technologies atende a esse chamado com soluções abertas e que usam Ethernet que, além de atenderem às necessidades imediatas das implementações de GenAI, também estabelecem a base para avanços futuros.

Para eliminar as suposições das soluções de hardware de IA, a Dell oferece arquiteturas de referência validadas em laboratório otimizadas para cargas de trabalho de IA. Esses Validated Designs incluem conceitos de arquitetura, visões gerais completas da solução, desempenho e outras validações no laboratório que comprovam os recursos da solução na carga de trabalho para a qual ela foi projetada. Migre de "possível" para "comprovado" com IA graças às soluções validadas que facilitam a entrega de insights mais rápidos e profundos.

Ao escolher a Dell Technologies, as organizações ganham um parceiro com o conhecimento especializado, um pacote de soluções completo e o compromisso de garantir o sucesso das iniciativas de GenAI. Com a Dell Technologies, as empresas conseguem navegar pelas complexidades das arquiteturas de GenAI, garantindo projetos viáveis e posicionados para o sucesso.

Acelere a implementação e o time-to-value para os ambientes de GenAI, reduzindo o risco e a complexidade operacional com a Dell Technologies. Convidamos você a explorar como uma solução de rede aberta, flexível e sustentável pode transformar suas iniciativas de GenAI, impulsionando sua empresa para uma nova era de inovação e eficiência.

Saiba mais sobre Dell Networking

Entre em contato com um especialista da Dell Technologies

Confira o resumo do analista da ESG

Confira o resumo do analista da IDC

DELL Technologies