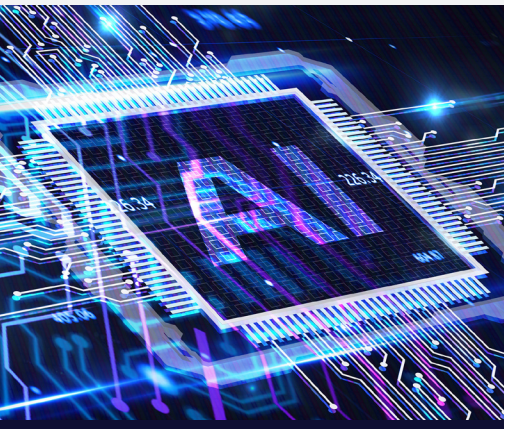
Executive Summary



In base alla nostra ricerca, Dell offre:

Il più ampio portafoglio di prodotti per l'AI generativa

In base ai risultati di MLPerf®, il server Dell PowerEdge XE8640 con quattro GPU NVIDIA H100 SXM5 ha raggiunto il throughput di AI più elevato tra tutti gli invii a quattro GPU in nove diverse categorie

Un portafoglio più ampio di servizi professionali su misura per l'AI

Affrontare le sfide dei carichi di lavoro di AI con il portafoglio AI Dell

Confronto tra il portafoglio AI Dell e le offerte analoghe di HPE

L'adozione dell'AI introduce nuove sfide per le aziende e il personale IT del loro data center, tra cui:

- Colmare le lacune di competenze esistenti del personale attuale tramite formazione interna o assunzioni esterne
- Comprendere le esigenze in termini di preparazione dei dati dell'AI, tra cui la qualità, la quantità, la posizione e lo stato attuale dei dati aziendali
- Valutare gli obiettivi aziendali per stabilire quali modelli e implementazioni di AI offrono vantaggi
- Valutare i requisiti computazionali, di rete e di storage dei sistemi AI pianificati e acquisire tali sistemi

Attraverso portafogli predisposti per l'AI, i vendor di infrastrutture come Dell Technologies e HPE offrono soluzioni integrate che coprono l'intero ciclo di vita dell'AI. Dopo aver ricercato informazioni pubblicamente disponibili sui portafogli AI di Dell e HPE ed esaminato i test di benchmark di MLPerf®, abbiamo rilevato che Dell è pronta ad aiutare le aziende nell'adozione dell'AI con un portafoglio che comprende opzioni di server e storage a prestazioni elevate, soluzioni convalidate e servizi professionali che guidano il processo dalla pianificazione alla produzione.

Prestazioni di benchmark dei modelli di AI: confronto dei risultati di MLPerf

I test di benchmark di MLPerf® pubblicamente disponibili dimostrano che le offerte del portafoglio AI Dell offrono prestazioni coerenti e solide per i carichi di lavoro di AI. MLPerf® testa le prestazioni relative sia all'addestramento che all'inferenza su diversi modelli di AI. I dati a cui si fa riferimento in questo report si basano sui risultati di MLPerf® v3.1 Inference Datacenter pubblicati sul sito web MLCommons® a partire da novembre 2023[1]. Abbiamo confrontato i server a 8 e 4 GPU e incluso in questo riepilogo un risultato per ciascuna categoria. Per visualizzare tutti i risultati, è possibile leggere il report completo qui.

* Dati basati su analisi Dell, agosto 2023. Dell Technologies offre soluzioni progettate per supportare i carichi di lavoro AI dai PC delle workstation (fisse e portatili) ai server per High Performance Computing, storage dei dati, infrastruttura software-defined nativa per il cloud, switch di rete, protezione dei dati, HCI e servizi. Dati tratti da: Robert McNeal, "Dell, VMware and NVIDIA Bring AI to Your Data", consultato il 17 gennaio 2024, https://www.dell.com/en-us/blog/dell-vmware-and-nvidia-bring-ai-to-your-data/.

1. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results", consultato il 12 dicembre 2023, https://mlcommons.org/en/inference-datacenter-31/.

Confronto delle prestazioni MLPerf per i server a otto GPU

Nei risultati di MLPerf® v3.1 per i server a otto GPU, le prestazioni di Dell PowerEdge XE9680 con GPU NVIDIA SXM5 H100 sono fino a 4,25 volte superiori rispetto a quelle di HPE ProLiant XL675d Gen10 Plus con GPU NVIDIA SXM4 A100 (Figura 1).

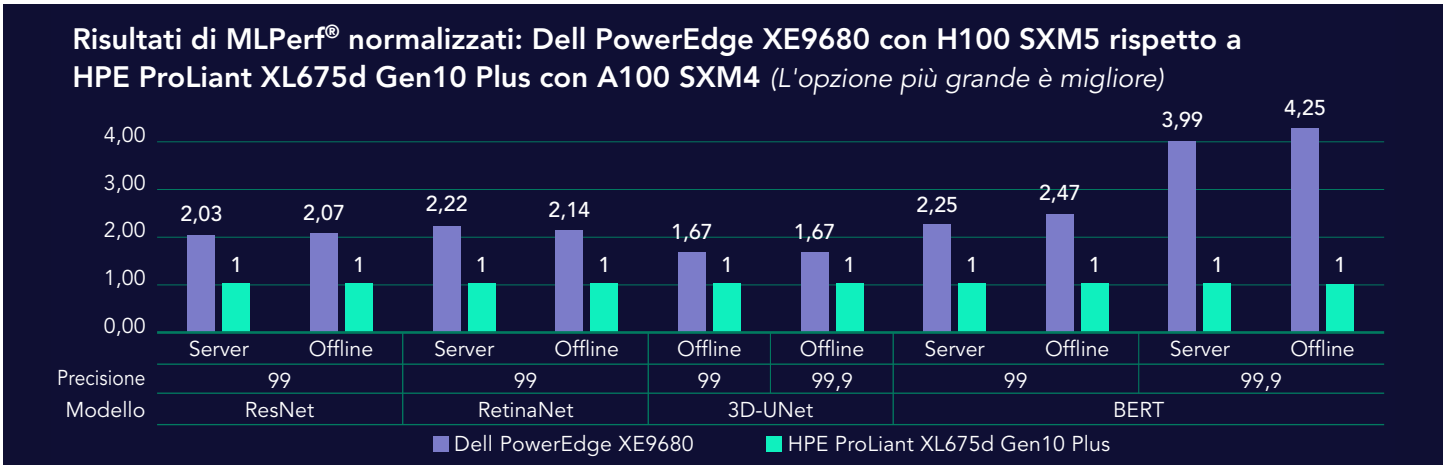


Figura 1. Risultati di MLPerf® pubblicati per Dell PowerEdge XE9680 e HPE ProLiant XL675d Gen10 Plus al 29/11/2023. Il sistema Dell utilizza GPU NVIDIA H100, mentre le GPU del sistema HPE sono di una generazione precedente. Fonte: Principled Technologies con i dati di MLCommons®.*

Confronto delle prestazioni MLPerf per i server a quattro GPU

Il server Dell PowerEdge XE8640 con quattro GPU NVIDIA H100 SXM5 ha raggiunto il throughput di AI più elevato tra tutti gli invii a quattro GPU in nove diverse categorie. Come mostra la Figura 2, ha un punteggio fino a 2,07 volte superiore rispetto al server HPE ProLiant Dl380a nel benchmark MLPerf®.

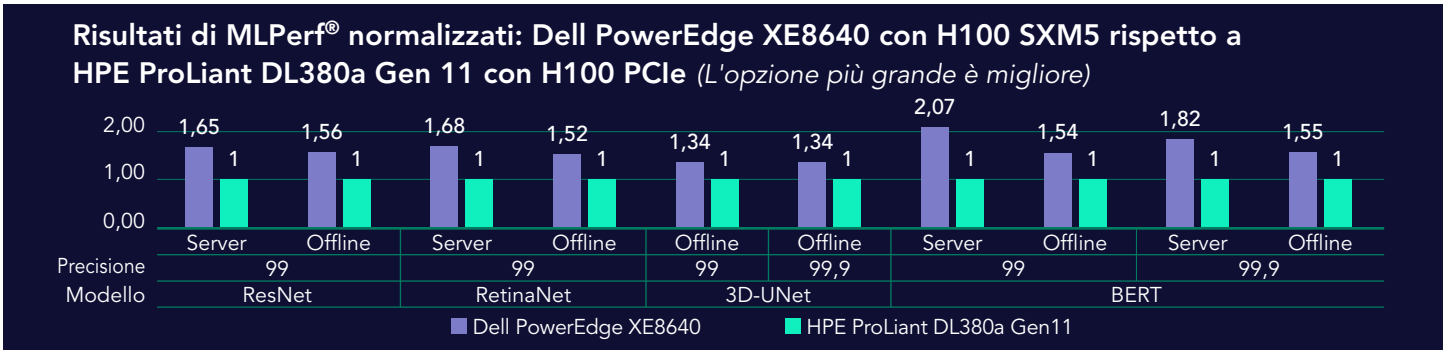


Figura 2. Risultati di MLPerf® pubblicati per Dell PowerEdge XE8640 e HPE ProLiant DL380a Gen11 al 29/11/2023. Il sistema Dell utilizza il fattore di forma NVIDIA H100 SXM, mentre il sistema HPE utilizza il fattore di forma PCIe meno potente. Fonte: Principled Technologies con i dati di MLCommons®.*

In breve: altri elementi che contribuiscono al successo dell'AI

Per avere successo, i modelli di AI hanno bisogno non solo di server a prestazioni elevate. È necessario considerare anche lo storage per i dati non strutturati, nonché servizi professionali per pianificare, preparare, implementare e gestire la soluzione di AI. Il portafoglio Dell offre soluzioni di storage per i dataset di AI con la serie PowerScale per lo storage dei file e lo storage Elastic Cloud Storage o ObjectScale per lo storage a oggetti.

Le organizzazioni possono avvalersi dei servizi professionali e di consulenza Dell per l'AI che offrono alcune attività non fornite da HPE, ad esempio la preparazione dei dati. Il portafoglio Dell include inoltre Validated Design per l'AI, che eliminano le incertezze dalla progettazione e dal deployment dell'AI.

Leggi il report ▶

*Punteggio MLPerf verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 su https://mlcommons.org/benchmarks/inference-datacenter/, voci 3.1-0069, 3.1-0085, 3.1-0067 e 3.1-0084. Il nome e il logo MLPerf sono marchi registrati e non registrati di MLCommons. Association negli Stati Uniti e in altri Paesi. Tutti i diritti riservati. L'uso non autorizzato è severamente vietato. Consultare www.mlcommons.org per maggiori informazioni.

▶ La versione originale in inglese di questo report è disponibile all'indirizzo https://facts.pt/PiSgyY7

Principled Technologies®

Facts matter.®