

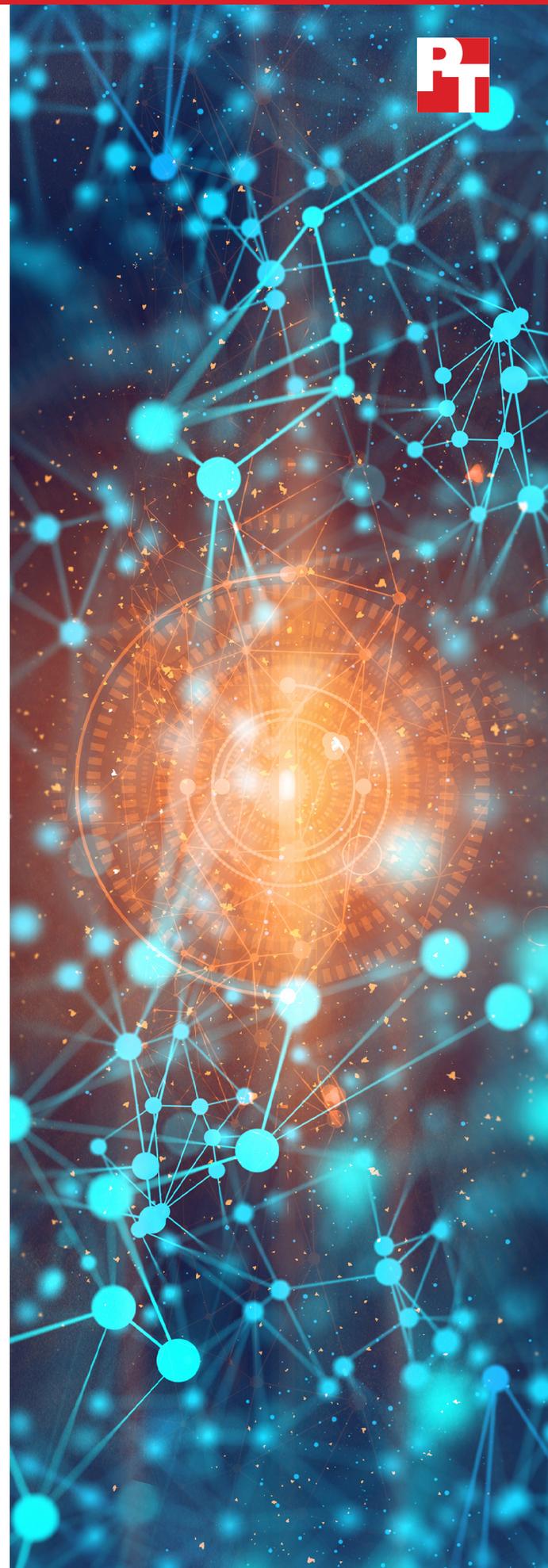


Trovare il percorso per il successo dell'AI con il portafoglio AI Dell

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

L'intelligenza artificiale (AI) è la nuova frontiera, pronta a rivoluzionare le operazioni aziendali in diversi settori. Organizzazioni di ogni tipo stanno riflettendo su come sfruttarla per migliorare le operazioni aziendali, ma è importante ricordare che implementare l'AI e beneficiare dei suoi vantaggi sono risultati che si raggiungono nel tempo. Poiché ogni azienda è unica, ognuna deve valutare i propri dati e obiettivi aziendali per vedere in che modo l'utilizzo dell'AI con i propri dati può portare ai risultati desiderati. Collaborando con un'azienda come Dell che offre un portafoglio completo di AI, tra cui pianificazione, preparazione dei dati, selezione hardware appropriata, progettazione dei modelli di AI, test di proof-of-concept, architetture di riferimento e supporto end-to-end si possono creare progetti di AI di successo.

Con la miriade di opzioni disponibili sul mercato, trovare un partner in grado di fornire assistenza in tutte queste decisioni potrebbe fare la differenza tra un'implementazione dell'AI di successo e un errore che comporta un costo elevato. In questo documento, esaminiamo più da vicino i portafogli AI di Dell e Supermicro con l'obiettivo di informare i lettori sui vantaggi che Dell può offrire durante il percorso di adozione dell'AI di un cliente. Ci concentreremo innanzitutto sulle opzioni di server ed elaborazione, per cui ogni azienda dispone di offerte ampie e diversificate per i clienti. In seguito, scopriremo come Dell vada oltre l'aspetto hardware per le aziende che cercano formazione, servizi di pianificazione, ecosistemi dei partner e altro ancora.



Server e risultati delle prestazioni per i carichi di lavoro di AI

I server, l'infrastruttura informatica di base che alimenta i carichi di lavoro di AI possono utilizzare CPU, GPU o entrambe le opzioni come risorse computazionali a seconda delle dimensioni o del tipo di carico di lavoro. Per carichi di lavoro più grandi o più complessi come HPC o AI, le GPU offrono prestazioni ottimali. Le GPU sono disponibili in una varietà di fattori di forma, tra cui PCIe universale, Open Compute Project Accelerator Module (OAM) e architettura NVIDIA SXM proprietaria, che attualmente offre prestazioni eccellenti.¹ Anche le grandi capacità di memoria e le funzionalità di progettazione dei server, come l'architettura di raffreddamento e l'efficienza energetica, influiscono sulle prestazioni. La maggior parte dei data center utilizza ancora il raffreddamento ad aria e ciò significa che i carichi di lavoro di AI necessitano di server progettati per raffreddare con aria nel modo più efficace possibile. Di seguito, mettiamo in evidenza le offerte di server Dell PowerEdge in termini di componenti, opzioni di raffreddamento e altro ancora, insieme ai punteggi MLPerf® di MLCommons® pubblicati.

Risultati del test

MLPerf® è una suite di benchmark che testa le prestazioni dell'AI sia per l'addestramento che per l'inferenza. Affinché un'organizzazione pubblichi i risultati di MLPerf® ufficiali, questi devono essere conformi alle condizioni specifiche stabilite dallo sviluppatore di benchmark, MLCommons®.² Queste linee guida di conformità forniscono standard che semplificano il confronto delle prestazioni. Per i test di inferenza, MLPerf® utilizza dataset di Datacenter, Edge, Mobile e Tiny e indica i punteggi dell'AI e i watt di energia consumati durante i test. La suite di benchmark per l'inferenza include il test di molti modelli diffusi di AI, ML e DL (vedere la Tabella 1).

Tabella 1: Modelli di AI, ML e DL inclusi in MLPerf®, con test e casi d'uso tipici per ciascuno di essi. Fonte: Principled Technologies.

Modelli di AI comunemente utilizzati	Use case tipici
ResNet	Un modello di classificazione delle immagini che aiuta i computer a imparare, ricordare e identificare immagini diverse per casi d'uso come l'imaging medicale, la moderazione dei contenuti sui social media e il riconoscimento facciale
RetinaNet	Un tipo di rilevamento degli oggetti in grado di gestire più complessità rispetto a ResNet. I computer sono così in grado di identificare e localizzare gli oggetti all'interno di immagini o frame video e di classificarli in base all'importanza. Utilizzato per applicazioni come la guida autonoma, la tecnologia di assistenza automatica dei veicoli, la sorveglianza e il riconoscimento facciale
3D-UNet	Specifico per la segmentazione delle immagini per uso medicale
RNN-T	Riconoscimento vocale per casi d'uso come la traduzione linguistica automatizzata
BERT	Elaborazione del linguaggio naturale per casi d'uso come sintesi dei testi, traduzione linguistica e completamento automatico delle attività
DLRM-v2-99.9	Modello di raccomandazione per casi d'uso come annunci mirati e consigli personalizzati sui prodotti
GPTJ-99 e 99.9	LLM per l'elaborazione del linguaggio naturale che eccelle nella generazione di testo per casi d'uso come chatbot e strumenti di AI basati su chat



Informazioni su MLPerf

I risultati di MLPerf® includono diversi parametri oltre ai modelli di AI stessi e ciò può consentire l'analisi di molti dati in un singolo grafico o tabella. Di seguito è riportato un riferimento rapido a questi parametri:

- 99.0 e 99.9: questi numeri si riferiscono alla precisione per la quale il modello è stato addestrato. Maggiore è la precisione necessaria per l'output, maggiori sono la complessità del modello e il tempo necessario per elaborare i dati.
- Esempi offline/sec: modalità in cui il benchmark invia tutte le query all'inizio del test simulando i dati già presenti nel sistema.
- Query del server/sec: modalità in cui il benchmark invia query per tutta la durata del test simulando l'analisi di un flusso di dati in tempo reale.

Per ulteriori informazioni sui risultati MLCommons® e MLPerf®, visitare <https://mlcommons.org/benchmarks/inference-datacenter/>.

I risultati di questo report derivano dai risultati di MLPerf® v3.1 Inference Datacenter pubblicati sul sito MLCommons® nel novembre 2023.³ Questi risultati includono gli invii di produttori di tecnologia e Cloud Service Provider e coprono una vasta gamma di configurazioni. Rispetto agli invii disponibili pubblicamente di Supermicro, i server Dell PowerEdge hanno ottenuto risultati comparabili. La Tabella 2 fornisce i dettagli del server.

Tabella 2: Server Dell e Supermicro inclusi nei risultati MLCommons® MLPerf® 3.1 pubblicati a novembre 2023. Fonte: Principled Technologies.

Mittente	Modello server	Numero e modello delle GPU	Descrizione
Dell ⁴	PowerEdge XE9680	8 NVIDIA H100 SXM	Per l'addestramento e l'inferenza dell'AI con carichi di lavoro elevati, ad esempio modelli linguistici di grandi dimensioni
	PowerEdge XE9640	4 NVIDIA H100 SXM	Per l'addestramento di modelli di AI di grandi dimensioni in data center ad alta densità e con raffreddamento a liquido
	PowerEdge XE8640	4 NVIDIA H100 SXM	Per promuovere applicazioni tradizionali di addestramento dell'AI, HPC e analisi dei dati in un fattore di forma 4U per data center con raffreddamento ad aria
Supermicro ⁵	AS-8125GS-TNHR	8 NVIDIA H100 SXM	Per l'addestramento dell'AI su larga scala e i carichi di lavoro HPC con processori AMD
	SYS-821GE-TNHR	8 NVIDIA H100 SXM	Per l'addestramento dell'AI su larga scala e i carichi di lavoro HPC con processori Intel
	SYS-421GU-TNHR	4 NVIDIA H100 SXM	Il design modulare è stato progettato per garantire flessibilità e supporto per i carichi di lavoro HPC e AI

Poiché sia Dell che Supermicro hanno inviato risultati con configurazioni GPU con confronto tra elementi omogenei, confrontare le prestazioni è semplice. Come mostrano le Figure 1 e 2, la comunanza tra le configurazioni di entrambi i vendor porta a risultati in gran parte uguali tra loro. Per altre configurazioni, come quelle illustrate nelle Figure 3 e 4, Dell ha superato Supermicro sul modello gptj-99.9 nel test a 4 GPU. Si noti che, diversamente da Supermicro, Dell ha inviato i risultati per tutti i modelli disponibili con tutti e tre i server. Confrontiamo solo i modelli per i quali entrambi i server presentano risultati. Per visualizzare l'intero set di risultati Dell, visita i risultati di MLCommons® MLPerf®.

Risultati dei server con otto GPU

Dell PowerEdge XE9680 offre supporto per un massimo di otto GPU NVIDIA H100 SXM5 per l'accelerazione dell'AI e fino a due processori scalabili Intel® Xeon® di quarta generazione. La famiglia di prodotti PowerEdge XE dispone di un'architettura modulare che supporta le GPU NVIDIA SXM4 o SXM5 o gruppi GPU Open Compute Project Accelerator Module (OAM), che possono migliorare le prestazioni rispetto a una GPU PCIe standard. Dell PowerEdge XE9680 offre anche l'acceleratore AMD Instinct™ MI300X.⁶ Occupando solo 6U di spazio su rack, PowerEdge XE9680 è un server compatto NVIDIA H100 SXM5 a otto vie.

Al contrario, il server Supermicro a 8 GPU richiede il 33% di spazio su rack in più a 8U, offrendo al contempo il fattore di forma SXM per le GPU NVIDIA. La differenza di dimensioni significa che è possibile inserire sette server Dell PowerEdge XE9680 in un rack rispetto a soli cinque server Supermicro. Nelle Figure 1 e 2, confrontiamo i risultati di Dell PowerEdge XE9680 con quelli di due configurazioni del server Supermicro a 8 GPU: SYS-821GE-TNHR con processori Intel e AS-8125GS-TNHR con processori AMD. Si noti che Supermicro non ha presentato risultati per RNN-T sul modello SYS-821GE-TNHR, quindi escludiamo questo modello dalla tabella in Figura 1.

Risultati di MLPerf® normalizzati: Dell PowerEdge XE9680 con H100 SXM5 rispetto a Supermicro SYS-821GE-TNHR con H100 (L'opzione più grande è migliore)

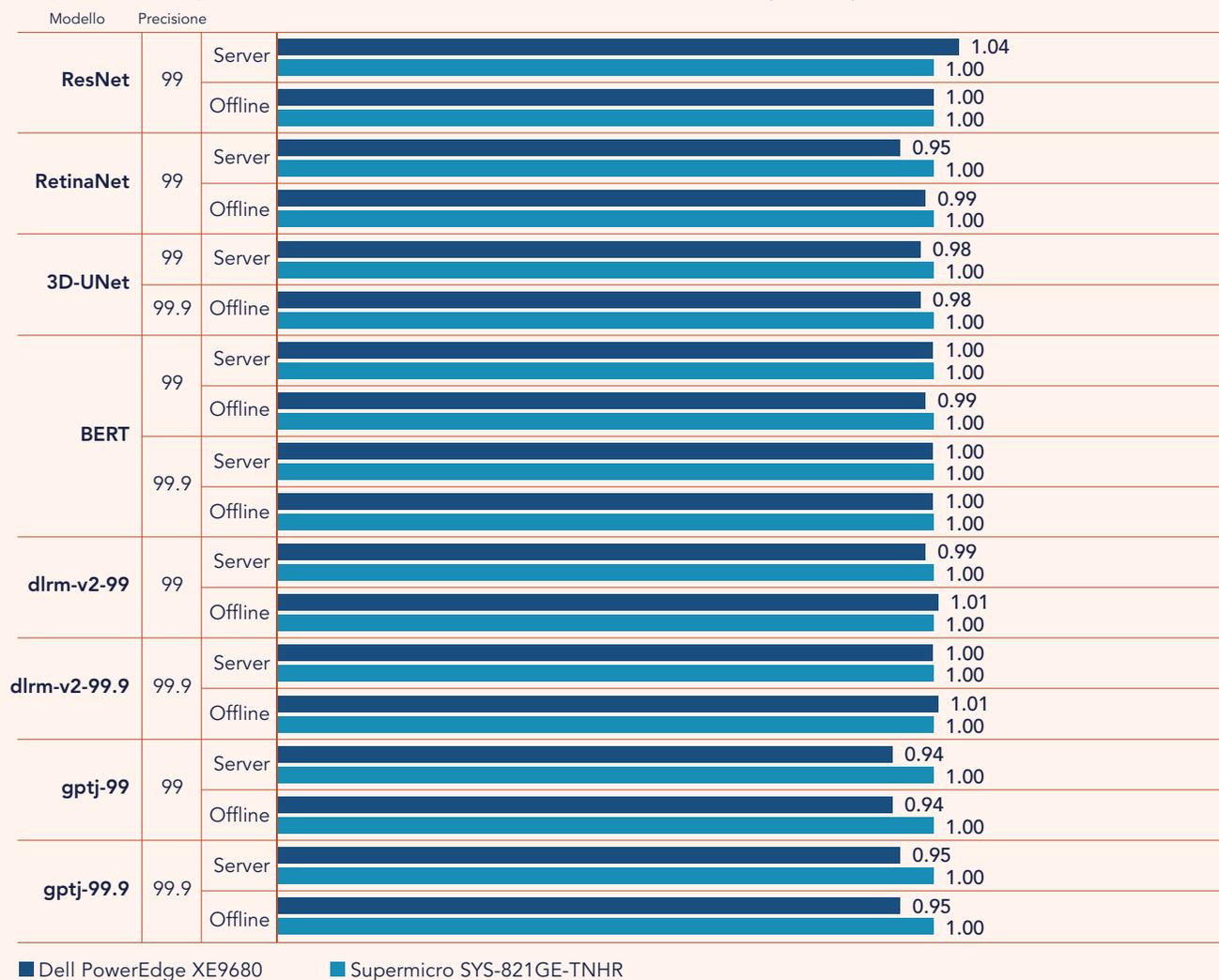


Figura 1. Risultati pubblicati di MLPerf® relativi a Dell PowerEdge XE9680 e Supermicro SYS-821GE-TNHR al 29/11/2023. Entrambi i sistemi utilizzano il fattore di forma SXM della GPU NVIDIA H100. Fonte: Principled Technologies con i dati di MLCommons®.^{7,8}



Risultati di [®] normalizzati: Dell PowerEdge XE9680 con H100 SXM5 rispetto a Supermicro AS-8125GS-TNHR con H100 SXM5 (L'opzione più grande è migliore)

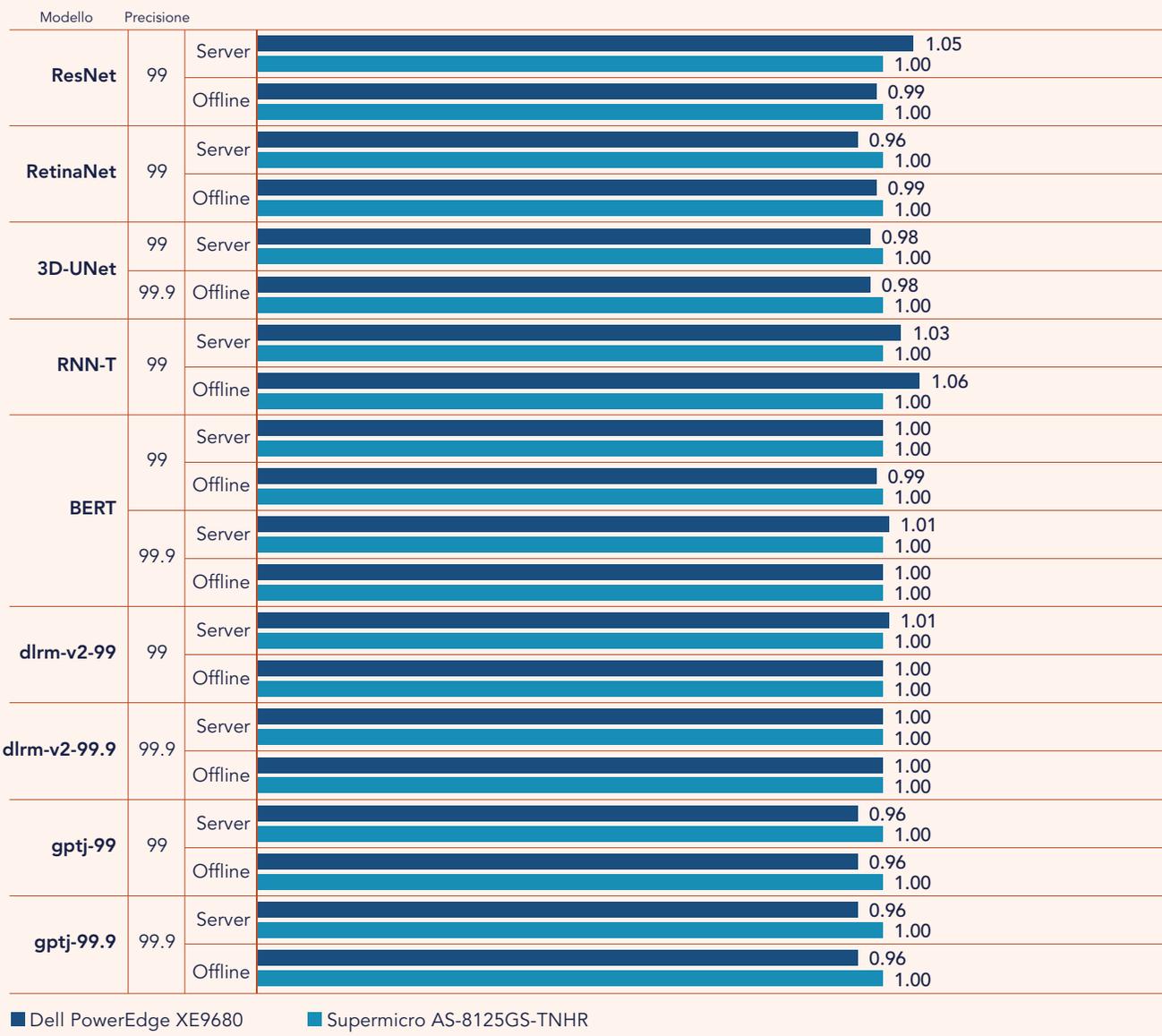


Figura 2. Risultati pubblicati di MLPerf[®] relativi a Dell PowerEdge XE9680 e Supermicro AS-8125GS-TNHR al 29/11/2023. Entrambi i sistemi utilizzano il fattore di forma SXM della GPU NVIDIA H100. Fonte: Principled Technologies con i dati di MLCommons^{® 9,10}

Come mostrato da questi risultati, la scelta di Dell PowerEdge XE9680 offre prestazioni simili su una miriade di carichi di lavoro di inferenza dell'AI, occupando al contempo meno spazio nel data center.



Risultati dei server con quattro GPU

Quando ridurre al minimo il consumo di energia o spazio nel data center rappresenta un fattore chiave, Dell PowerEdge XE9640 2U potrebbe essere la soluzione giusta. Con un massimo di quattro GPU NVIDIA H100 SXM, PowerEdge XE9640 offre la metà della potenza di elaborazione della GPU di PowerEdge XE9680, in un terzo dello spazio.¹¹ Il server Dell PowerEdge XE9640, ad alta densità, integra la tecnologia Dell Smart Cooling, che fornisce una serie di tecnologie termiche, tra cui il raffreddamento a liquido diretto per CPU e GPU.¹² Lo chassis 2U di PowerEdge XE9640 supporta meccanismi di circolazione dell'aria migliorati, tra cui ventole e dissipatori di calore più grandi, per raffreddare gli altri componenti essenziali, come schede PCIe e memoria.¹³

Supermicro offre un server SYS-220GQ-TNAR+ meno recente che fornisce quattro GPU NVIDIA A100 HGX in un fattore di forma 2U, ma non siamo riusciti a trovare alcun server Supermicro 2U con quattro delle GPU H100 HGX più recenti che corrisponderebbero a PowerEdge XE9640.¹⁴ Il server a 4 GPU con GPU NVIDIA HGX H100 di Supermicro negli invii MLPerf® è SYS-421GU-TNXR, che è un server 4U. Come accennato in precedenza, Supermicro ha inviato a MLPerf® 3.1 solo i risultati relativi a SYS-421GU-TNXR sul modello di AI gptj-99.9, quindi non possiamo confrontarli con PowerEdge XE9640 sugli altri modelli. Tuttavia, nei risultati pubblicati, PowerEdge XE9640 ha superato le prestazioni del server Supermicro nei test offline, ottenendo un punteggio fino a 1,37 volte maggiore (vedi Figura 3).

Risultati di MLPerf® normalizzati: Dell PowerEdge XE9640 con H100 SXM5 rispetto a Supermicro SYS-421GU-TNXR con H100 SXM5 (L'opzione più grande è migliore)

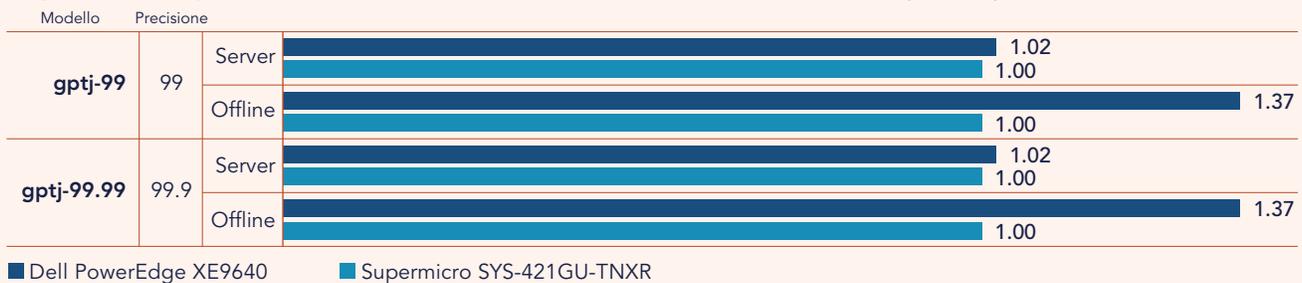


Figura 3. Risultati pubblicati di MLPerf® relativi a Dell PowerEdge XE9640 e Supermicro SYS-421GU-TNXR al 29/11/2023. Entrambi i sistemi utilizzano il fattore di forma SXM della GPU NVIDIA H100. Fonte: Principled Technologies con i dati di MLCommons®.^{15,16}

Possiamo anche confrontare i risultati del server Supermicro SYS-421GU-TNXR con Dell PowerEdge XE8640, un server 4U a quattro GPU che supporta anche GPU NVIDIA H100 HGX. Sebbene più grande di PowerEdge XE9640, PowerEdge XE8640 non richiede il raffreddamento a liquido diretto, il che lo rende un compromesso tra le tecnologie di densità e raffreddamento per i data center che non hanno accesso al raffreddamento ad acqua. PowerEdge XE8640 offre raffreddamento ad aria per i processori e un radiatore con raffreddamento ad aria assistita a liquido per le GPU che non richiede la disponibilità di alcuna struttura di approvvigionamento di acqua per la rack.¹⁷ Dell PowerEdge XE8640 include i più recenti processori scalabili Intel Xeon di quarta generazione e fino a 4 TB di memoria per gestire dataset di grandi dimensioni e calcoli complessi diffusi nell'AI e nell'analisi dei dati.¹⁸ Con il suo fattore di forma 4U, PowerEdge XE8640 è simile a Supermicro SYS-421GU-TNXR sia in termini di densità che di funzionalità GPU. Tuttavia, come abbiamo visto con PowerEdge XE9640, Dell PowerEdge XE8640 ha ottenuto punteggi gptj-99 migliori nei test offline rispetto al server Supermicro (vedi Figura 4).

Risultati di MLPerf® normalizzati: Dell PowerEdge XE8640 con H100 SXM5 rispetto a Supermicro SYS-421GU-TNXR con H100 SXM5 (L'opzione più grande è migliore)

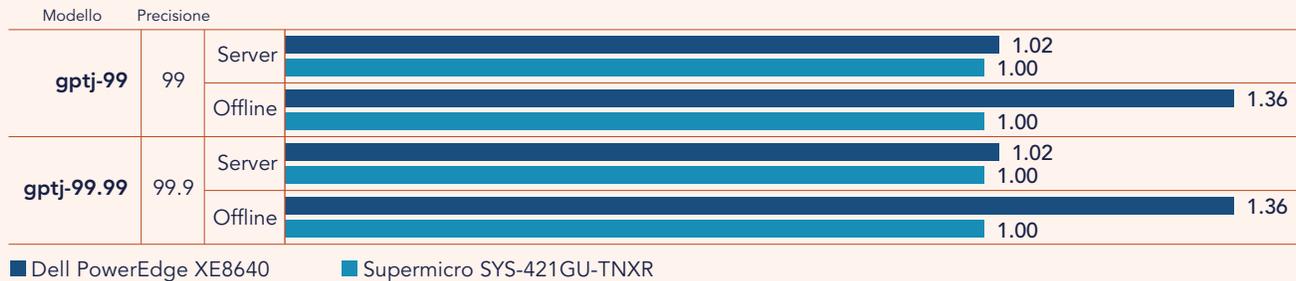


Figura 4. Risultati pubblicati di MLPerf® relativi a Dell PowerEdge XE8640 e Supermicro SYS-421GU-TNXR al 29/11/2023. Entrambi i sistemi utilizzano il fattore di forma SXM della GPU NVIDIA H100. Fonte: Principled Technologies con i dati di MLCommons®.^{19,20}

Come abbiamo visto, i risultati MLPerf® relativi alle prestazioni delle offerte di server basati su GPU di Supermicro e Dell sono complessivamente simili, con un notevole vantaggio per i server Dell PowerEdge XE8640 e PowerEdge XE9640 su un unico modello di AI.

Poiché le prestazioni sono solo un aspetto del percorso di implementazione dell'AI, abbiamo esaminato anche il resto delle offerte di AI di Dell e Supermicro: workstation, storage, rete, servizi, supporto, formazione e altro ancora. Abbiamo rilevato che il portafoglio Dell AI è più ampio rispetto all'offerta Supermicro, oltre a offrire soluzioni a molti degli ostacoli che non riguardano le prestazioni di elaborazione.

Workstation client e offerte di storage

Opzioni di elaborazione aggiuntive con le workstation

Alcuni casi d'uso dell'AI richiedono un approccio di elaborazione diverso e non tutti coloro che hanno bisogno di accedere a hardware compatibile con l'AI possono rimanere legati al data center. Gli scienziati che lavorano nei laboratori potrebbero non avere spazio per un rack di server e per coloro che lavorano nell'edge è poco pratico trasportare un sistema desktop grande e pesante. È qui che entrano in gioco le workstation. Il portafoglio Dell AI include diverse workstation Precision predisposte per l'AI, tra cui configurazioni tower, mobili e rack, per soddisfare una varietà di esigenze.²¹ Al contrario, Supermicro non offre workstation portatili per un facile accesso durante gli spostamenti. Le sue offerte di workstation GPU sono costituite da diverse configurazioni tower, e Supermicro afferma che alcune di esse possono essere montate su rack.²² La nostra ricerca ha rilevato due workstation montabili su rack, ma sembrano essere meno recenti, disponibili solo con GPU NVIDIA A100 e forse non più in vendita.²³ Se l'azienda ha bisogno di flessibilità nel tipo e nella mobilità delle workstation GPU implementate, il portafoglio AI Dell è più adatto a queste esigenze.

Considerazioni sullo storage

Lo storage può essere fondamentale tanto quanto l'elaborazione per l'esecuzione dei carichi di lavoro di AI. Un numero maggiore di dati migliora la precisione dei modelli AI, ma lo storage e la gestione di enormi dataset possono rappresentare una sfida per le capacità di molti data center. Inoltre, poiché i modelli sono generalmente addestrati utilizzando dati non strutturati, i sistemi di storage predisposti per l'AI devono gestire con facilità molti tipi di dati diversi.²⁴ Per offrire capacità e dimensionamento per i dataset di AI, ML e DL, Dell offre la serie PowerScale™ per lo storage su file e l'Elastic Cloud Storage (ECS) o ObjectScale software-defined per lo storage a oggetti.

Il portafoglio Dell All-Flash NAS PowerScale offre capacità che vanno da 3,84 TB fino a 720 TB di capacità raw per nodo, con capacità All-Flash in cluster che raggiungono i 186 PB di capacità raw. La flessibilità e la scalabilità di PowerScale possono supportare un'ampia varietà di clienti e casi d'uso di AI.²⁵ Tutti e tre i modelli All-Flash PowerScale, F200, F600 e F900, includono la compressione e la deduplica dei dati in linea per migliorare l'efficienza dello storage.²⁶ Ogni modello di storage PowerScale utilizza il file system Dell OneFS™, che sfrutta le policy di tiering avanzate per garantire che i dati a cui si accede più di frequente risiedano sugli storage Tier con le prestazioni più elevate.²⁷ Dell offre inoltre il software OneFS nel Marketplace AWS (Amazon Web Services), insieme ad APEX File Storage for AWS. I clienti possono utilizzare OneFS con le istanze di elaborazione AWS per un'esperienza utente coerente con le stesse funzioni disponibili negli array OneFS on-premise.²⁸

Le offerte di storage Supermicro sono costituite da server di storage, montati su rack e ad alta densità, di varie dimensioni e densità.²⁹ Per ottenere lo storage su file da Supermicro, i clienti devono scegliere tra varie offerte di software-defined storage di terze parti, come WekaIO, Scality RING o OSNEXUS.³⁰ Sebbene Scality RING e OSNEXUS, in base alle descrizioni della piattaforma, includano opzioni di storage, WekaIO sembra essere l'opzione principale per i clienti che desiderano storage su file di base. Supermicro offre diverse architetture di riferimento che coprono un'ampia gamma di casi d'uso, ma i clienti devono disporre di un abbonamento o di una licenza software WekaIO, il che può aumentare il costo complessivo della soluzione.³¹

Le opzioni di storage a oggetti di Dell includono Dell ECS (Enterprise Object Storage), che è "progettato in maniera specifica per archiviare i dati non strutturati sul public cloud".³² Insieme alla compatibilità integrata con lo storage a oggetti Amazon S3 per la funzionalità di hybrid cloud, gli storage node ECS offrono capacità fino a 14 PB per rack.³³ Come per lo storage su file, le offerte di storage a oggetti Supermicro richiedono configurazioni di terze parti. La piattaforma OSNEXUS è una piattaforma di storage combinata per file, blocchi e storage a oggetti, mentre la soluzione Scality RING unisce file e storage a oggetti; entrambi richiedono licenze con vendor di terze parti.^{34,35} Per lo storage a oggetti, i clienti possono acquistare la soluzione Supermicro per Quantum ActiveScale per lo storage a oggetti su private cloud con un abbonamento software Quantum.³⁶ Nel momento in cui scriviamo, non siamo riusciti a trovare opzioni di consumo/spesa operativa flessibili offerte da Supermicro.

Poiché le offerte di storage Supermicro richiedono ai clienti di interagire con fornitori di software di terze parti, è probabile che debbano sostenere ulteriori costi di licenza o abbonamento e potrebbero riscontrare difficoltà in termini di supporto, risoluzione dei problemi e altro ancora. Il portafoglio Dell AI offre ai clienti di storage Dell un'unica soluzione di servizi e supporto affidabile in ogni aspetto della soluzione di storage selezionata.

Informazioni su Dell APEX

Per i clienti che desiderano utilizzare lo storage su file as-a-Service, Dell offre APEX Data Storage Services, che include storage su file, blocchi e di backup. Utilizzando Dell APEX Console, i clienti possono ordinare nuovi abbonamenti, regolare e monitorare la capacità di storage e altro ancora. Per Dell, questa soluzione consente di "ottenere la semplicità e l'agilità dell'esperienza del cloud con maggiore controllo su applicazioni e dati".³⁷

Ulteriori informazioni su Dell APEX sono disponibili all'indirizzo <https://www.dell.com/en-us/dt/apex/storage/data-storage-services/index.htm>



Opzioni di rete

La rete è un altro componente vitale dell'infrastruttura AI. Con molti carichi di lavoro AI in esecuzione su cluster di server di grandi dimensioni e che comunicano costantemente tra di loro e con lo storage, occorrono reti robuste per evitare colli di bottiglia. Se le prestazioni di rete sono insufficienti per il carico di lavoro AI, i tempi di addestramento e inferenza aumentano, rallentando l'elaborazione dei dati e il tempo di analisi degli stessi. Dell offre switch top-of-rack (ToR) per data center PowerSwitch e moduli I/O PowerEdge MX per reti Ethernet e fabric.³⁸ Le offerte PowerSwitch vanno da 1 GbE a 400 GbE per soddisfare una vasta gamma di esigenze. Inoltre, gli switch Dell PowerSwitch serie Z offrono connessioni da 100 GbE e 400 GbE ottimizzate per fabric leaf/spine.³⁹

Anche Supermicro offre switch con porte Ethernet fino a 400 GbE per ToR e altre applicazioni come spine e leaf nei data center.⁴⁰ Tuttavia, i servizi di rete Dell offrono diversi vantaggi in termini di facilità d'uso e flessibilità che Supermicro non offre. Servizi come Dell Fabric Design Center possono aiutare a evitare mancate corrispondenze, lacune o inefficienze di rete, aiutando i clienti a pianificare e implementare fabric di rete con automazione.⁴¹ Per ambienti specifici come VMware VxRail, VMware ESXi e configurazioni Dell PowerStore, Dell offre SmartFabric Services, che consente il deployment dell'infrastruttura software-defined e la gestione del ciclo di vita. Con PowerStore, SmartFabric Services può automatizzare "fino al 99% delle attività di connettività LAN con una fabric plug-and-play".⁴² Servizi come questi, che forniscono automazione, indicazioni e altro ancora per la progettazione e l'implementazione della rete, supportano i clienti durante il processo di adozione dell'AI.

Servizi, formazione e altro ancora

La principale sfida non hardware nell'implementazione dell'AI è la necessità di competenze interne per la strategia, la pianificazione, la preparazione dei dati e la gestione. La gestione e la manutenzione dei carichi di lavoro di AI richiede un set di conoscenze unico, che includa competenze hardware più tradizionali, operazioni di apprendimento automatico e data science. Anche le persone che progettano e implementano la strategia di intelligenza artificiale hanno bisogno di una profonda comprensione degli obiettivi operativi specifici dell'azienda per garantire che i nuovi carichi di lavoro di AI soddisfino tali obiettivi.⁴³

Un'altra sfida significativa può essere l'integrazione dell'AI in modo fluido nei sistemi operativi esistenti. Questa integrazione richiede un allineamento strategico delle nuove tecnologie di AI con gli attuali processi aziendali, garantendo che l'introduzione dell'intelligenza artificiale mantenga intatti i flussi di lavoro consolidati. Collaborare con un'azienda come Dell che fornisce diverse architetture di riferimento per soluzioni ottimizzate e convalidate, corsi di formazione, opzioni di gestione e un ampio ecosistema di partner può semplificare il percorso di adozione dell'AI.

Professional Services for AI

Per coprire l'istruzione e la pianificazione, Dell offre una varietà di servizi specifici per l'AI.⁴⁴ I servizi Dell che supportano l'implementazione dell'AI comprendono consulenza, preparazione dei dati, deployment, supporto e istruzione, e ognuno di essi si rivolge ad aspetti specifici dell'adozione dell'AI. Dell Advisory Services for Generative AI aiuta i clienti a creare una roadmap che identifica i casi d'uso e aiuta le aziende a semplificare i processi.⁴⁵ Analogamente, Adoption Services for Generative AI offre workshop con professionisti Dell per esaminare le esigenze e le sfide specifiche per determinare un modello preaddestrato per il business e condurre sessioni di trasferimento delle conoscenze per formare il personale IT.⁴⁶ Dell offre inoltre servizi di implementazione, scalabilità e gestione per l'AI generativa che forniscono diversi livelli di supporto e formazione, fino a un'infrastruttura di AI completamente gestita, che consente al personale IT di concentrarsi sui modelli e sui dati mentre Dell gestisce l'hardware.⁴⁷ I servizi ProSupport garantiscono prestazioni ottimali del sistema e forniscono assistenza hardware e software essenziale per le operazioni di AI in corso, risolvendo i problemi tecnici.⁴⁸

I servizi di formazione sono fondamentali per promuovere le competenze e le conoscenze necessarie per l'utilizzo dell'intelligenza artificiale. Le offerte di formazione Dell includono programmi di formazione completi in data science, certificazioni di analisi avanzata e workshop su tecnologie di AI specifiche, come l'apprendimento automatico.⁴⁹

I servizi Supermicro, d'altra parte, sono principalmente limitati alla risoluzione dei problemi, ai manuali, alle autorizzazioni RMA (Return Merchandise Authorization) e alle garanzie.⁵⁰ Non siamo riusciti a trovare servizi di progettazione, implementazione, gestione o formazione nel portafoglio Supermicro AI. Per le aziende che cercano un partner di formazione per affrontare le complessità dell'adozione dell'AI, Dell è la scelta giusta tra queste due soluzioni.

Collaborazioni di terze parti per carichi di lavoro AI

Dell Technologies e NVIDIA collaborano per offrire i Dell Validated Design, che puntano a fornire una soluzione completa per l'AI generativa nelle impostazioni aziendali. Questo progetto mira a creare un'infrastruttura scalabile e ad alte prestazioni basata su tecnologie e software Dell e NVIDIA insieme a un framework di modelli di AI che consente alle aziende di creare ed eseguire modelli di AI personalizzati. La soluzione consente ai clienti di rendere rapidamente operativi i carichi di lavoro di AI generativa.⁵¹ Per ulteriori informazioni, consulta la sezione Dell Validated Design di seguito.

Dell ha stretto partnership con diverse aziende per sviluppare e implementare nuove tecnologie e applicazioni di intelligenza artificiale. Con Hugging Face, Dell semplifica la configurazione di modelli linguistici di grandi dimensioni (LLMs) on-site. Questa partnership combina l'esperienza di Hugging Face AI con i server e i sistemi di archiviazione di Dell. Un portale specifico per Dell di Hugging Face fornirà strumenti per una distribuzione semplice e sicura di modelli AI open-source di Hugging Face. L'obiettivo costante è quello di continuare a migliorare questi modelli per i sistemi Dell, aumentando le prestazioni e supportando nuove applicazioni AI.⁵²

Dell e Starburst stanno lavorando a un data lake scalabile e a prestazioni elevate che integra l'analisi Starburst con la tecnologia di elaborazione e storage Dell, cercando di offrire un unico punto di accesso a tutte le origini dati per gli strumenti di AI e ML. I clienti potranno sfruttare questa partnership per eliminare i silos di dati.⁵³

Secondo la nostra ricerca, Supermicro ha partnership molto più limitate per l'intelligenza artificiale. SiMa.ai e Supermicro hanno collaborato allo sviluppo di Supermicro SYS-E300-13AD, un server Edge ML compatto progettato per l'elaborazione di analisi video multi-stream. Questo server, dotato di pipeline ML SiMa.ai su un chip, gestisce in modo efficiente più canali video, riduce i costi complessivi di gestione e migliora l'affidabilità e la sicurezza. Il server offre un'impostazione computazionale progettata per l'elaborazione e l'analisi di numerosi flussi video, fornendo intelligenza edge adatta a varie applicazioni aziendali.⁵⁴

Dell Validated Design

Per eliminare le incertezze dalle soluzioni hardware AI, Dell offre architetture di riferimento convalidate in laboratorio e ottimizzate per diverse AI e altri carichi di lavoro. Queste progettazioni convalidate includono concept di architettura, panoramiche complete della soluzione, convalide delle prestazioni e altre convalide di laboratorio, che dimostrano le capacità della soluzione sul carico di lavoro per il quale è stata progettata. Questi carichi di lavoro includono ambienti virtualizzati, MLOps, apprendimento automatico, AI conversazionale, inferenza dell'AI generativa, ottimizzazione dei modelli di AI generativa, NVIDIA Fleet Command e OpenShift AI.⁵⁵

Ad esempio, l'AI per ambienti virtualizzati Validated Design unisce l'AI abilitata per VMware con NVIDIA AI Enterprise sull'infrastruttura Dell, ottimizzando l'AI nelle impostazioni virtuali.⁵⁶ La Guida a Validated Design include i risultati delle prestazioni che mostrano la formazione del modello ResNet, dimostrando ai clienti che la progettazione funziona e il tipo di prestazioni che possono aspettarsi.⁵⁷ Queste convalide offrono ai clienti un valore che va oltre la semplice elencazione dell'hardware che funziona in modo collaborativo, spiegando concetti, consigliando configurazioni e guidando i clienti attraverso considerazioni e aspettative in termini di prestazioni.⁵⁸

Supermicro offre soluzioni in base ai casi d'uso, ma non raggiunge il livello di architettura di riferimento che abbiamo riscontrato con Dell Validated Design. Anziché consigliare una soluzione ad hoc per carichi di lavoro specifici, Supermicro organizza i server e le GPU in categorie, come inferenza e formazione dell'AI, HPC/AI, visualizzazione e progettazione e altro ancora.⁵⁹ Nelle brochure e nei data sheet, queste categorie sono costituite da alcuni server e GPU presentati come più adatti per l'attività, con diversi casi d'uso ed elenchi di tecnologie chiave in gioco e suggerimenti software.⁶⁰ A differenza di Dell Validated Design, non sembrano includere architetture di rete e dati sulle prestazioni o sulla convalida. Supermicro fornisce inoltre diversi progetti di riferimento che offrono un'architettura di riferimento più dettagliata per alcune soluzioni di AI, ad esempio un training sull'AI su larga scala con solution brief di raffreddamento a liquido Supermicro pubblicato con NVIDIA.⁶¹ I clienti potrebbero essere in grado di trovare un'architettura di riferimento più approfondita per uno scenario specifico, ma al momento della nostra ricerca ne abbiamo trovati solo tre: l'architettura di raffreddamento a liquido menzionata in precedenza, un'architettura di workstation AI e un'architettura RedHat OpenShift.⁶²

Nel complesso, abbiamo riscontrato che Dell Validated Design ha preso in considerazione più carichi di lavoro AI e offerto indicazioni più approfondite rispetto alle offerte Supermicro.

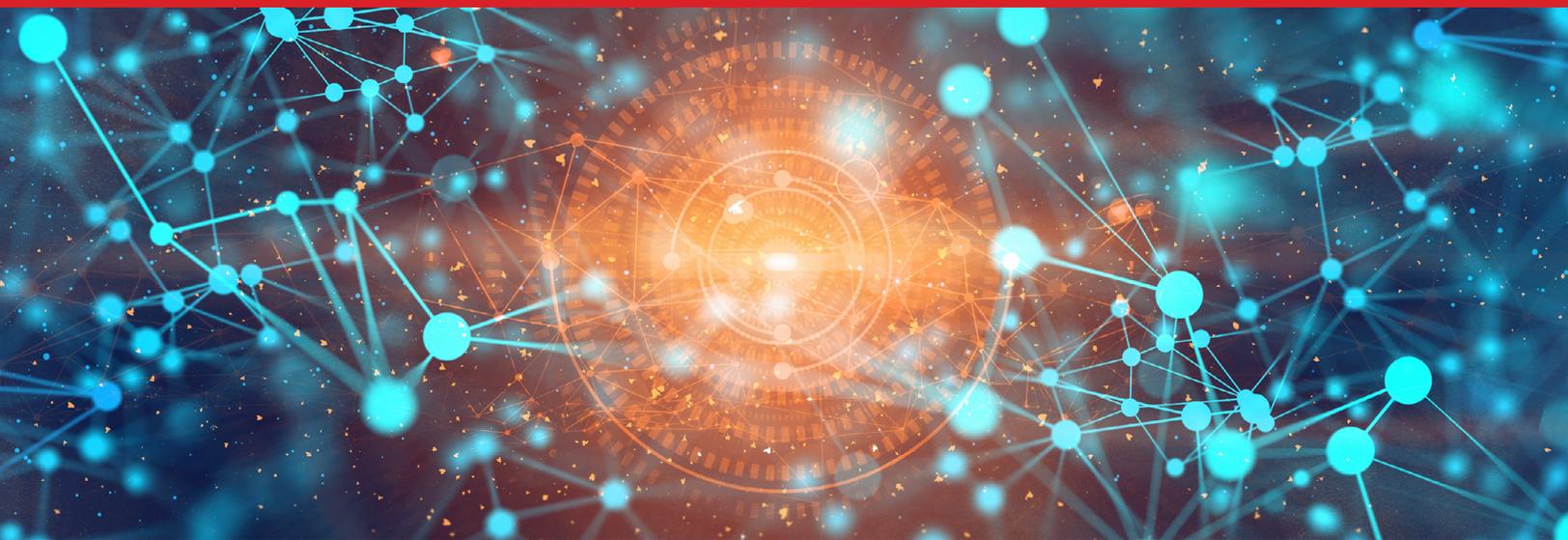


Servizi di gestione e iDRAC

Secondo un [report di aprile 2023](#) a cura di Principled Technologies, Integrated Dell Remote Access Controller (iDRAC) offre diverse funzionalità avanzate rispetto all'interfaccia IPMI (Intelligent Platform Management Interface) Supermicro, in particolare per l'automazione, la sicurezza e la configurazione.⁶³ La Tabella 3 presenta un confronto tra le funzionalità di gestione Dell e Supermicro proveniente dal report, che mostra come iDRAC possa fornire un deployment più semplice, un aggiornamento del firmware più facile e un maggior numero di funzionalità di protezione rispetto a Supermicro IPMI. Alcuni risultati potrebbero variare rispetto alla pubblicazione originale.

Tabella 3: Riepilogo del confronto tra gli strumenti di gestione Dell e Supermicro di aprile 2023 a cura di Principled Technologies. Alcuni risultati potrebbero essere cambiati dalla pubblicazione. Fonte: Principled Technologies <https://facts.pt/V5fDf06>.

	Quali sono le differenze con gli strumenti di gestione Dell	Quanto sono migliori
Aggiornamenti più semplici del firmware <i>iDRAC9 rispetto a Supermicro IPMI</i> <i>OME rispetto a Supermicro SSM</i>	<ul style="list-style-type: none"> • Aggiornamenti online automatizzati con iDRAC9, con opzioni di pianificazione • OME consente la creazione di repository firmware personalizzati e può aggiornare il firmware del BIOS, del BMC e di altri componenti server senza strumenti o agent aggiuntivi 	<ul style="list-style-type: none"> • Abbiamo configurato gli aggiornamenti automatici in iDRAC in soli 74 secondi • Supermicro IPMI non dispone di una funzione di aggiornamento automatico, pertanto gli amministratori devono eseguire l'aggiornamento manualmente • SSM supporta solo gli aggiornamenti del firmware del BIOS e del BMC e richiede SUM per aggiornare altri componenti
Più funzioni di sicurezza <i>iDRAC9 rispetto a Supermicro IPMI</i> <i>OME rispetto a Supermicro SSM</i>	<ul style="list-style-type: none"> • iDRAC9 offre MFA e disabilitazione dinamica delle porte USB senza downtime del sistema • OME offre sia il controllo dell'accesso basato su ruoli (RBAC) che il controllo dell'accesso basato sulla portata (SBAC) per limitare la gestione dei dispositivi a un sottoinsieme di gruppi di dispositivi 	<ul style="list-style-type: none"> • Supermicro IPMI non dispone di funzioni MFA • Supermicro IPMI richiede un riavvio del sistema e l'accesso alla configurazione del BIOS per disabilitare le porte USB • Supermicro SSM offre RBAC ma non il più restrittivo SBAC
Gestione semplificata del ciclo di vita <i>OME rispetto a Supermicro SSM</i>	<ul style="list-style-type: none"> • Gestione completa e senza agent del ciclo di vita tramite OME per semplificare la gestione e il monitoraggio 	<ul style="list-style-type: none"> • SSM richiede l'agent SuperDoctor5 per metriche dettagliate locali sulla salute del sistema e il Supermicro Update Manager (SUM) per aggiornare ulteriori componenti
Deployment dei server più semplice <i>iDRAC9 rispetto a Supermicro IPMI</i>	<ul style="list-style-type: none"> • Importa un profilo server Dell completo in soli 12 passaggi utilizzando iDRAC9 • Robuste opzioni di configurazione del BIOS con iDRAC9 con 52 funzioni del BIOS e supporto per la configurazione dei componenti come RAID NIC e iDRAC 	<ul style="list-style-type: none"> • Supermicro IPMI ci ha consentito di salvare e ripristinare solo la configurazione IPMI anziché l'intero profilo del server • iDRAC9 dispone di 52 funzioni del BIOS, mentre IPMI non offre opzioni di configurazione del BIOS
Più opzioni per il reporting e l'analisi <i>iDRAC9 rispetto a Supermicro IPMI</i> <i>OME rispetto a Supermicro SSM</i>	<ul style="list-style-type: none"> • iDRAC9 offre lo streaming di telemetria, che consente agli utenti di inviare facilmente i dati del server a strumenti di analisi come Splunk • OME invia i dati di telemetria direttamente a CloudIQ per un monitoraggio più semplice 	<ul style="list-style-type: none"> • IPMI offre solo una funzione SYSLOG che gli amministratori possono utilizzare per inviare messaggi per l'aggregazione e l'analisi finale • SSM non dispone di una soluzione di gestione basata su cloud equivalente a Dell CloudIQ
Più funzioni di sostenibilità <i>OME rispetto a Supermicro SSM</i>	<ul style="list-style-type: none"> • Più metriche per il monitoraggio in OME Power Manager, inclusi i dati relativi all'impatto ambientale 	<ul style="list-style-type: none"> • SSM dispone di metriche di utilizzo meno solide e non consente di monitorare l'impatto ambientale
Più modi di monitorare <i>OME rispetto a Supermicro SSM</i>	<ul style="list-style-type: none"> • Gestisci i server Dell da qualsiasi luogo tramite l'app mobile OpenManage • Controlla i dispositivi di terze parti con OME utilizzando gli indirizzi IP e le credenziali del server con supporto per l'importazione di SNMP MIBS di terze parti. 	<ul style="list-style-type: none"> • SSM non dispone di un'app mobile • SSM non consente il monitoraggio di dispositivi di terze parti con IP del server



Conclusioni

Quando si tratta di progettare, implementare, gestire, mantenere e migliorare le soluzioni di intelligenza artificiale nel tuo business, ci sono molti fattori da considerare. Per aiutarti a investire in modo oculato e ottenere il massimo dalla tua soluzione di AI, potresti voler cercare un vendor che possa offrirti di più di una fornitura di hardware. La nostra ricerca indica che Dell offre servizi che possono aiutarti in qualità di partner durante l'intero percorso, quindi prendi in considerazione l'investimento con Dell mentre esplori il mondo dell'AI.

1. Vipera, "NVIDIA's H100 and A100 GPU Cards: Exploring the Intricacies of SXM and PCI-E Connections", consultato il 5 gennaio 2024, <https://www.viperatech.com/unraveling-the-mysteries-sxm-vs-pci-e-connections-in-nvidias-high-end-h100-and-a100-gpus/>.
2. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results", consultato il 5 gennaio 2024, <https://mlcommons.org/en/inference-datacenter-31/>.
3. MLCommons, "MLPerf Inference: Datacenter benchmark Suite results".
4. Dell, "PowerEdge XE Servers", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/servers/specialty-servers/poweredge-xe-servers.htm>.
5. Supermicro, "Next Leap of AI Infrastructure is Here", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/accelerators/nvidia>.
6. Dell, "PowerEdge XE9680", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9680-spec-sheet.pdf>.
7. Punteggio MLPerf[®] verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0069. 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.
8. Punteggio MLPerf[®] verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0135. 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.
9. Punteggio MLPerf[®] verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0069. 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.
10. Punteggio MLPerf[®] verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0132. 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.

-
11. Dell, "PowerEdge XE9640 Rack Server", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640>.
 12. Accelsius, "Enabling the AI Revolution with Liquid Cooling", consultato il 5 gennaio 2024, <https://www.accelsius.com/blog/enabling-the-ai-revolution-with-liquid-cooling>.
 13. Dell, "PowerEdge XE9640 Technical Guide", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9640-technical-guide.pdf>.
 14. Supermicro, "GPU Server Systems", consultato il 5 gennaio 2024, https://www.supermicro.com/en/products/gpu?pro=pl_grp_type%3D1.
 15. Punteggio MLPerf® verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0067. 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.
 16. Punteggio MLPerf® verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0133. 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.
 17. Dell, "PowerEdge XE8640", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe8640-spec-sheet.pdf>.
 18. Dell, "PowerEdge XE8640 Rack Server", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe8640>.
 19. Punteggio MLPerf® verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0066. 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.
 20. Punteggio MLPerf® verificato di v3.1 Inference Closed. Consultato il 5 dicembre 2023 da <https://mlcommons.org/benchmarks/inference-datacenter/>, voce 3.1-0133. 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.
 21. Dell, "Artificial Intelligence (AI) technologies powered by Dell Precision Workstations", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/ai-technologies/index.htm?hve=explore+dell+precision+for+ai#tab0=0>.
 22. Supermicro, "Super Workstations", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/products/superworkstation>.
 23. Supermicro, "Rackmount Workstations", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/products/rackmount-workstations>.
 24. Stephen Pritchard, "Storage requirements for AI, ML and analytics in 2022", consultato il 5 gennaio 2024, <https://www.computerweekly.com/feature/Storage-requirements-for-AI-ML-and-analytics-in-2022>.
 25. Dell, "PowerScale AI-Ready Data Platform", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/shop/powerscale-family/sf/powerscale>.
 26. Dell, "Dell PowerScale All-Flash", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/storage/technical-support/h15963-ss-powerscale-all-flash-nodes.pdf>.
 27. Dell, "Dell PowerScale OneFS Software Features", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/storage/technical-support/h18275-onefs-software-features-data-sheet.pdf>.
 28. Dell, "Dell ECS Enterprise Object Storage", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/storage/ecs/>.
 29. Supermicro, "Accelerating AI Data Pipelines," consultato il 5 gennaio 2024, <https://www.supermicro.com/en/products/storage>.
 30. Supermicro, "Supermicro Software-Defined Storage and Memory Solutions", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/solutions/software-defined-storage>.
 31. Supermicro, "Supermicro WEKA Distributed Storage Solution", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/solutions/wekaio>.
 32. Dell, "Dell ECS Enterprise Object Storage", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/storage/ecs/>.
 33. Dell, "Dell ECS Enterprise Object Storage".

-
34. Supermicro, "Supermicro OSNEXUS Software-Defined Storage Solution", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/solutions/osnexus>.
 35. ASBIS, "Supermicro Solution for Scalify RING", consultato il 5 gennaio 2024, <https://news.asbis.com/news/suppliers/supermicro-renewed-the-line-of-scalify-ring-solution/>.
 36. Supermicro, "Supermicro Solution for Quantum ActiveScale", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/solutions/activescale>.
 37. Dell, "Scalable and elastic Storage as-a-Service", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/apex/storage/data-storage-services/>.
 38. Dell, "Flip the Switch to Open Networking with PowerSwitch", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/networking/>.
 39. Dell, "Dell PowerSwitch Data Center Switches", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/networking/data-center-switches/>.
 40. Supermicro, "SSE-T7132S - 400Gb Ethernet Switch", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/products/accessories/Networking/SSE-T7132SR.php>.
 41. Dell, "Dell EMC Networking SmartFabric Services Deployment with VxRail 4.7—Fabric Design Center", consultato il 5 gennaio 2024, <https://infohub.delltechnologies.com/l/dell-emc-networking-smartfabric-services-deployment-with-vxrail-4-7-1/fabric-design-center-26/>.
 42. Dell, "Dell SmartFabric Services," consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/networking/smartfabric/>.
 43. Penny Madsen, "Scaling Skills for AI: Lessons from Early Adopters", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/industry-market/idc-brief-importance-of-skills-for-ai-dell.pdf>.
 44. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Overview", consultato il 5 gennaio 2024, <https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/overview-5381/>.
 45. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Advisory Services for Generative AI", consultato il 5 gennaio 2024, <https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/advisory-services-for-generative-ai-1/>.
 46. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Adoption Services for Generative AI", consultato il 5 gennaio 2024, <https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/adoption-services-for-generative-ai-1/>.
 47. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Adoption Services for Generative AI".
 48. Dell, "Artificial Intelligence (AI) Ready Solution Services", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/services/solutions/artificial-intelligence-services.htm>.
 49. Dell, "Comprehensive AI Training Modules Tailored for You", consultato il 5 gennaio 2024, <https://education.dell.com/content/emc/en-us/home/training/aiml.html>.
 50. Supermicro, "Servizi e Supporto", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/support>.
 51. Travis Vigil, "Dell and NVIDIA: Bringing Generative AI to the Enterprise", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/blog/dell-and-nvidia-bringing-generative-ai-to-the-enterprise/>.
 52. Dell, "Dell Technologies and Hugging Face to Simplify Generative AI with On-Premises IT", consultato il 5 gennaio 2024, <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>.
 53. Richard DeMare, "Starburst and Dell expand partnership to accelerate AI efforts with more intelligent data collection", consultato il 5 gennaio 2024, <https://www.starburst.io/blog/starburst-and-dell-expand-partnership-to-accelerate-ai-efforts-with-more-intelligent-data-collection/>.
 54. Business Wire, "SiMa.ai and Supermicro Announce Partnership to Accelerate Power Efficient ML at the Edge", consultato il 5 gennaio 2024, <https://www.businesswire.com/news/home/20231129609794/en/SiMa.ai-and-Supermicro-Announce-Partnership-to-Accelerate-Power-Efficient-ML-at-the-Edge>.
 55. Dell, "Dell AI Solutions", consultato il 5 gennaio 2024, <https://www.dell.com/en-us/dt/solutions/artificial-intelligence/index.htm#accordion0&tab0=0>.
 56. Dell, "Unlock the power of AI in virtualized environments", consultato il 5 gennaio 2024, <https://www.delltechnologies.com/asset/en-us/products/ready-solutions/briefs-summaries/ai-vxrail-powerscale-brief.pdf>.

-
57. Dell, "Design Guide—Virtualizing GPUs for AI with VMware and NVIDIA Based on Dell Infrastructure—Performance Results", consultato il 5 gennaio 2024, <https://infohub.delltechnologies.com/l/design-guide-virtualizing-gpus-for-ai-with-vmware-and-nvidia-based-on-dell-infrastructure-1/performance-results-15/>.
 58. Dell, "Design Guide—Virtualizing GPUs for AI with VMware and NVIDIA Based on Dell Infrastructure—Design Considerations", consultato il 5 gennaio 2024, <https://infohub.delltechnologies.com/l/design-guide-virtualizing-gpus-for-ai-with-vmware-and-nvidia-based-on-dell-infrastructure-1/design-considerations-105/>.
 59. Supermicro, "Accelerate Every Workload", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/solutions/ai-deep-learning>.
 60. Supermicro, "Supermicro Enterprise AI Inference & Training", consultato il 5 gennaio 2024, https://www.supermicro.com/datasheet/Datasheet_AI-Workloads_Enterprise_AI_Inferencing_and_Training.pdf.
 61. Supermicro, "SUPERMICRO RACK SCALE SOLUTIONS: LARGE SCALE AI TRAINING WITH LIQUID COOLING", consultato il 5 gennaio 2024, https://www.supermicro.com/solutions/Solution-Brief_Rack_Scale_AI.pdf.
 62. Supermicro, "Accelerate Every Workload", consultato il 5 gennaio 2024, <https://www.supermicro.com/en/solutions/ai-deep-learning>.
 63. Principled Technologies, "Dell management tools made server deployment and updates easier, offered more comprehensive security, and provided more robust infrastructure analytics", consultato il 5 gennaio 2024, <https://www.principledtechnologies.com/Dell/Management-tools-vs-Supermicro-0423.pdf>.

► La versione originale in inglese di questo report è disponibile all'indirizzo <https://facts.pt/q9p46K9>

Questo progetto è stato commissionato da Dell Technologies.



Facts matter.®

Principled Technologies è un marchio registrato di Principled Technologies, Inc. Tutti gli altri nomi di prodotto sono marchi dei rispettivi proprietari.

ESCLUSIONE DI GARANZIE; LIMITAZIONE DI RESPONSABILITÀ:

Principled Technologies, Inc. si è ragionevolmente impegnata per assicurare la precisione e la validità dei test di cui nel presente documento, tuttavia Principled Technologies, Inc. declina specificamente qualsiasi garanzia, espressa o implicita, in merito ai risultati di test e analisi e alla relativa precisione, completezza o qualità, inclusa qualsiasi garanzia implicita di adeguatezza a un determinato scopo. Tutte le persone e le entità che si basano sui risultati di un test lo fanno a proprio rischio e riconoscono che Principled Technologies, Inc., i suoi dipendenti e i suoi subappaltatori non hanno alcun tipo di responsabilità inerente a rivendicazioni per perdite o danni sulla base di presunti errori o difetti nella procedura o nei risultati dei test.

Principled Technologies, Inc. non sarà in alcun caso responsabile per danni indiretti, speciali, incidentali o consequenziali in relazione ai test eseguiti, anche se a conoscenza della possibilità del verificarsi di tali danni. La responsabilità di Principled Technologies, Inc. non supererà in alcun caso, incluso per danni diretti, gli importi versati in relazione ai test di Principled Technologies, Inc. Gli unici ed esclusivi rimedi dei clienti sono definiti nel presente documento.