



Studio di ricerca tecnica



Prestazioni da record mondiale per AI e ML

La ricerca Prowess stabilisce dei collegamenti tra il solido benchmark dei server Dell™ e le prestazioni reali nell'ambito dell'addestramento e dell'inferenza su modelli di intelligenza artificiale (AI) e apprendimento automatico (ML).

Executive Summary

Le organizzazioni di un'ampia varietà di settori dipendono dall'intelligenza artificiale (AI) e dall'apprendimento automatico (ML) per prevedere le vendite, segmentare le loro basi di clienti, identificare i rischi, gestire reti di fornitura complesse, ottimizzare i costi e migliorare l'efficienza. Indipendentemente dall'applicazione, tutti i casi d'uso di AI e ML dipendono dalle prestazioni di elaborazione, oltre che dalla velocità e dalla quantità di memoria e dalla larghezza di banda delle interconnessioni e della rete. Inoltre, a causa della natura dei dati utilizzati dalle aziende per attività come l'analisi predittiva dei clienti o il rilevamento delle frodi, la sicurezza è spesso altrettanto critica.

Le esigenze legate alle prestazioni di AI e ML, rendono l'infrastruttura nel suo insieme di capacità di elaborazione, memoria locale, larghezza di banda di rete e storage dei dati per supportare questi carichi di lavoro, un investimento significativo, che determina la necessità di una valutazione rigorosa prima dell'acquisto. I benchmark degli standard di settore sono utili per questo aspetto e i record a livello mondiale possono rappresentare una soluzione addirittura migliore se valutati nel modo corretto.

Le aziende devono acquisire, archiviare ed elaborare rapidamente i propri dati. I benchmark possono fornire informazioni dettagliate sulla velocità con cui i dati possono essere raccolti, elaborati e a cui è possibile accedere dopo averli archiviati. Per analizzare la relazione tra le prestazioni elevate di benchmark e il valore potenziale per il business nel mondo reale, Prowess Consulting esamina in modo più approfondito il significato dei solidi risultati a livello di benchmark di settore per le aziende che implementano server da record mondiale. Grazie all'enorme quota di mercato di Dell Technologies e al numero di record a livello mondiale che ha raggiunto negli scenari di AI e ML, rappresentati da benchmark riconosciuti a livello di settore, Prowess ha esaminato in modo specifico i server Dell™ PowerEdge™.

Tra i server Dell PowerEdge di ultima generazione, ne abbiamo identificati tre che hanno recentemente stabilito record a livello mondiale in un benchmark progettato appositamente per garantire dati sulle prestazioni reali sui carichi di lavoro di AI e ML:

- **Server Dell PowerEdge R6625 (record mondiale, TPC Express Benchmark™ AI [TPCx-AI])**
- **Server Dell PowerEdge R7615 (record mondiale, TPCx-AI)**
- **Server Dell PowerEdge R7515 (record mondiale, TPC Express Benchmark™ IoT [TPCx-IoT])**

Grazie all'aumento delle ottimizzazioni delle prestazioni di Dell Technologies per i carichi di lavoro di AI e ML, abbiamo anche scoperto che i processori AMD EPYC™ e le schede di rete Broadcom® supportano le prestazioni di Big Data e analisi sia per i benchmark sia per le applicazioni reali. Ad esempio, il supporto del formato numerico bfloat16 nei processori AMD EPYC di quarta generazione consente l'esecuzione di modelli di AI di dimensioni maggiori con dataset più grandi, mentre il supporto dei numeri INT8 può accelerare l'inferenza sui modelli di AI. Inoltre, le doppie schede NVM Express® (NVMe®) del controller RAID Dell™ PowerEdge™ (PERC Dell™) e il supporto per PCIe® 4.0 con schede di rete Broadcom abilitano i controller dell'interfaccia di rete (NIC) 100 Gigabit Ethernet (GbE) a due porte, contribuiscono a eliminare i vincoli della larghezza di banda e accelerano ulteriormente le prestazioni dei carichi di lavoro di AI e ML.

Nel presente studio vengono trattati i seguenti argomenti:

- [Panorama del settore](#)
- [Metodologia di ricerca di Prowess](#)
- [Benchmark di AI e ML](#)
- [Alla base dei risultati delle prestazioni](#)



Panorama del settore: AI e ML

AI e ML hanno assunto un ruolo centrale nelle operazioni aziendali per le organizzazioni di una varietà di settori. Che si tratti di un monitoraggio bancario per le frodi, di un rivenditore che prevede le vendite, di un ospedale alla ricerca di diagnosi più accurate o di un produttore di medie dimensioni che implementa la manutenzione predittiva sulle linee di produzione, le organizzazioni di tutti i tipi e dimensioni si affidano all'AI per ottenere modelli che potrebbero essere invisibili alle persone. I modelli di AI in generale e i modelli di apprendimento approfondito (DL) in particolare, sono generalmente più accurati quando è disponibile un maggior numero di dati per la fase di addestramento. Questa propensione per i dati determina la necessità di uno storage di dimensioni maggiori e più funzionale, una rete più veloce e server dalle prestazioni più solide per individuare un maggior valore nei dati, che, a loro volta, devono anche essere protetti.

Le aziende spesso si affidano a server on-premise anziché a implementazioni su cloud per una serie di motivi. Per AI e ML, tali motivi spesso sono incentrati su data gravity e minore latenza per l'utilizzo dei modelli di AI. Di frequente risulta più facile e veloce avvicinare le funzioni di addestramento dell'AI ai dati, anziché sostenere i costi e le tempistiche necessarie per lo spostamento di grandi quantità di dati verso un sistema di elaborazione centralizzata. L'utilizzo dei dati in prossimità della posizione in cui risiedono può anche ridurre la latenza per l'addestramento dell'AI, che, a sua volta, permette di accelerare il processo. Inoltre, i requisiti richiesti dalle normative vigenti e le leggi in materia di sovranità dei dati possono anche rappresentare dei motivi validi per mantenere i dati on-premise, a seconda del settore e della posizione di un'organizzazione. In tutti i casi, le prestazioni rappresentano un requisito fondamentale per le aziende quando si occupano dei dati e toccano il valore analitico contenuto dai dati tramite AI e ML.

Le esigenze relative alle prestazioni dei carichi di lavoro, come l'analisi, indicano che l'infrastruttura dei dati deve essere ottimizzata per soddisfare gli accordi sui livelli di servizio (SLA). È fondamentale l'interazione di processore, dimensioni della memoria, larghezza di banda di rete e sottosistemi di storage. Uno strumento importante per il confronto delle prestazioni del server per questa interazione è rappresentato dai risultati di benchmark. Poiché i benchmark generano risultati numerici, i confronti tra i sistemi concorrenti possono sembrare semplici.

Proprio perché i benchmark generano risultati chiari e apparentemente obiettivi, tuttavia, risulta cruciale la comprensione di quali elementi misurano e, di conseguenza, delle informazioni effettive sulle piattaforme server. Le organizzazioni che ignorano le sfumature di tali benchmark e si affidano ciecamente alla sicurezza di chi secondo questi offre prestazioni migliori possono rimanere deluse nel caso in cui il ritorno sul capitale investito (ROI) non soddisfi le aspettative.



Metodologia di ricerca di Prowess

Per analizzare la relazione tra le prestazioni elevate di benchmark e il valore potenziale per il business nel mondo reale, Prowess Consulting esamina in modo più approfondito il significato dei solidi risultati a livello di benchmark di settore per le aziende che implementano server da record mondiale. Per semplificare la nostra indagine e focalizzarci sul modo in cui i singoli benchmark possono fornire informazioni dettagliate sulle prestazioni, in particolare per quanto riguarda le sfaccettature di AI e ML, abbiamo esaminato in modo specifico i server Dell PowerEdge. Abbiamo scelto questo metodo sia per l'ampia quota di mercato che hanno i server Dell Technologies sia per il numero di record a livello mondiale detenuti da Dell Technologies in una varietà di benchmark di AI e ML.

È straordinario un record mondiale su un unico benchmark, ma il tratto distintivo per i carichi di lavoro di AI e ML consiste nel fatto che le piattaforme Dell raggiungono record a livello mondiale su più benchmark. Ogni benchmark può essere considerato un pezzo del puzzle del carico di lavoro e il raggiungimento di più record a livello mondiale fornisce ottime informazioni dettagliate sulla modalità di funzionamento delle piattaforme Dell negli ambienti reali.

Abbiamo esaminato i risultati di benchmark per contribuire a stabilire quali piattaforme offrano le prestazioni migliori per diversi aspetti dei carichi di lavoro di AI e ML. La nostra ricerca si è concentrata sui server Dell™ con montaggio su rack. Dell Technologies possiede la maggiore quota di mercato di server a livello globale (17,2%)¹ e i server Dell PowerEdge sono dei modelli dall'operatività consolidata molto diffusi, progettati per le esigenze dei carichi di lavoro da standard a medio/intensi. In particolare, per il presente studio abbiamo esaminato le piattaforme con montaggio su rack 1U (server Dell PowerEdge R6625) e 2U (server Dell PowerEdge R7615 e PowerEdge R7515).

Durante l'esame dei risultati di benchmark, è fondamentale visualizzarli tenendo conto dei fattori più importanti relativi alle prestazioni. Nell'ambito dei carichi di lavoro di AI e ML, sono inclusi:

- Prestazioni
- Rapporto prezzo/prestazioni

Dell Technologies ha ottimizzato la piattaforma PowerEdge, basata su processori AMD®, per le soluzioni di AI e ML. I carichi di lavoro di AI e ML possono presentare esigenze specifiche e i server PowerEdge offrono ai clienti un elevato livello di flessibilità per l'esecuzione dei carichi di lavoro univoci in base ai requisiti specifici. Inoltre, Dell Technologies ha ottimizzato le prestazioni per le piattaforme PowerEdge tramite l'integrazione delle schede Dell PERC più recenti e delle schede di rete Broadcom NVMe che apportano notevoli miglioramenti in termini di larghezza di banda in questo quadro generale.

Benchmark di AI e ML

I benchmark riconosciuti a livello di settore possono fornire informazioni dettagliate sugli utilizzi comuni di una piattaforma server e possono contribuire a informare i clienti sulla capacità della piattaforma di soddisfare le esigenze dei carichi di lavoro del cliente in fase di esecuzione. Per il presente studio, abbiamo esaminato in particolare i benchmark applicabili più direttamente ai carichi di lavoro di AI e ML.

Nota: MLPerf™ Inference Benchmark Suite è comunemente utilizzata per la misurazione delle prestazioni di ML. Prowess ha scelto di non utilizzare MLPerf per l'analisi perché tale benchmark si concentra sulle prestazioni dell'unità di elaborazione grafica (GPU), mentre il nostro studio si è concentrato sull'analisi delle prestazioni della CPU a costi contenuti per l'inferenza dell'AI.

TPC Express Benchmark™ AI (TPCx-AI)	Misura le prestazioni end-to-end dei carichi di lavoro di AI e ML rappresentativi del settore
TPC Express Benchmark™ IoT (TPCx-IoT)	Misura le prestazioni, il rapporto costo/prestazioni e la disponibilità per i sistemi che acquisiscono enormi quantità di dati provenienti da un numero elevato di dispositivi

TPC Express Benchmark™ (TPCx-AI)

TPCx-AI misura le prestazioni di una piattaforma data science o di ML end-to-end. Il benchmark è stato progettato per emulare il comportamento delle soluzioni di AI rappresentative del settore presenti nei data center di produzione e negli ambienti cloud.

TPCx-AI valuta le prestazioni di AI tramite una serie di casi d'uso che si riferiscono a singoli problemi risolti dalla pipeline di data science di DL e ML nell'ambito di TPCx-AI. La pipeline è indipendente da qualsiasi framework o sintassi specifica e può essere implementata secondo varie modalità. In genere, tra i casi d'uso nell'ambito di TPCx-AI sono inclusi la generazione dei dati, la gestione dei dati, l'addestramento, la valutazione e le fasi di esecuzione del servizio.



Figura 1. Pipeline di data science TPC Express Benchmark™ AI (TPCx-AI) DL e ML

Tra i casi d'uso nell'ambito di TPCx-AI sono inclusi la segmentazione dei clienti, la trascrizione di conversazioni con i clienti, le previsioni di vendita, il rilevamento di spam, la previsione dei prezzi, la classificazione e il rilevamento delle frodi.

Dell Technologies detiene quattro record mondiali per **TPCx-AI** con processori AMD EPYC di quarta generazione²:

- Server Dell PowerEdge R7615 (con fattore di scala 3)
- Server Dell PowerEdge R7615 (con fattore di scala 10)
- Server Dell PowerEdge R6625 (con fattore di scala 30)
- Server Dell PowerEdge R6625 (con fattore di scala 100)

I server PowerEdge R7615 hanno raggiunto un throughput della pipeline di AI di 408,36 GB/secondo con fattore di scala 3 (3 GB di dati) e 425,31 GB/secondo con fattore di scala 10 (10 GB di dati), con un rapporto prezzo/prestazioni di \$ 118,56/GB/secondo con fattore di scala 3².

Il server PowerEdge R6625 ha raggiunto 365,59 GB/secondo con fattore di scala 30, mentre un cluster di quattro server PowerEdge R6625 ha raggiunto 868,49 GB/secondo con fattore di scala 100, con rapporti prezzo/prestazioni rispettivamente di \$ 196,38/GB/secondo e \$ 356,56/GB/secondo².

Questi server sono basati su processori AMD EPYC 9374F, EPYC 9174F ed EPYC 9354 di quarta generazione. Oltre a disporre di un numero di core e thread più elevato rispetto alla generazione precedente, i processori AMD EPYC di quarta generazione garantiscono anche il supporto della memoria DDR5 per una memoria di capacità maggiore, il supporto PCIe di quinta generazione per un throughput dei dati più elevato all'interno di server e cluster di server, il supporto per le più recenti AVX-512 (Advanced Vector Extensions 512) per un throughput dei dati più rapido nel registro del processore e il supporto per tipi numerici efficienti (ad esempio bfloat16 e INT8) per accelerare l'addestramento e l'inferenza dei modelli. Tutte queste funzioni sono state utili per i server Dell in ogni fase della pipeline di data science nell'ambito di TPCx-AI.

Internet of Things (IoT): TPC Express Benchmark™ IoT (TPCx-IoT)

L'Internet of Things (IoT) rappresenta un'ampia origine di dati per AI e ML, in particolare per l'inferenza in tempo reale da parte dei modelli di AI, come quelli che eseguono i controlli della qualità sulle linee di produzione. Inoltre, l'IoT rappresenta un'origine essenziale di dati sull'addestramento per i modelli di AI. Tale aspetto si applica all'addestramento iniziale o ai nuovi modelli, ma anche al nuovo addestramento dei modelli esistenti che mostrano una deriva.

Sebbene non faccia parte della pipeline di data science incapsulata nel benchmark di TPCx-AI, l'addestramento continuo dei modelli di AI in fase di produzione è fondamentale. Nel corso del tempo, la composizione dei dati in ambienti di produzione come vendita al dettaglio, finanza o settore sanitario può mutare in modi impercettibili. La conseguenza può essere una riduzione costante dell'accuratezza dei modelli di AI. L'acquisizione efficiente dei dati aggiornati sull'addestramento al fine di correggere la deviazione dei modelli di AI rappresenta una parte cruciale di qualsiasi flusso di lavoro di AI o ML.

Ciò che contraddistingue l'IIoT dagli altri tipi di dati non strutturati consiste nel fatto che i dispositivi IIoT spesso generano i dati dal proprio ambiente, come gli altoparlanti smart che ascoltano i comandi o i droni industriali che raccolgono dati sui terreni agricoli. Pertanto, i dati di IIoT devono essere acquisiti e utilizzati in modo efficiente per i carichi di lavoro di AI e ML a causa dell'enorme volume, che in genere è maggiore di quello di altre origini di dati, come per le transazioni delle vendite al dettaglio o per i record del settore sanitario.

Il benchmark TPCx-IIoT consente di confrontare direttamente diverse soluzioni software e hardware per i gateway IIoT. Proprio perché posizionati tra l'architettura edge e il data center back-end, i sistemi gateway eseguono funzioni come l'aggregazione dei dati per AI e ML in tempo reale. Il benchmark TPCx-IIoT è stato progettato appositamente per garantire metriche verificabili di prestazioni, rapporto prezzo/prestazioni e disponibilità per i sistemi disponibili sul mercato che in genere acquisiscono enormi quantità di dati provenienti da un numero elevato di dispositivi.

Dell Technologies detiene un record mondiale per **TPCx-IIoT** per i server Dell PowerEdge 7515 con processori AMD EPYC di terza generazione. Il cluster ha raggiunto un throughput di 1.617.545.000 record/secondo con un rapporto prezzo/prestazioni di \$ 329,75/milioni di record/sec³.

Oltre alla potenza di elaborazione fornita dal processore AMD EPYC 75F3 di terza generazione, la nostra analisi dei risultati di benchmark indica anche che le prestazioni dei server Dell sono state potenziate dalle schede di rete Broadcom da 25 GbE utilizzate nel cluster. Tutte queste funzioni sono state utili per i server Dell per il raggiungimento delle prestazioni nel benchmark TPCx-IIoT, con i risultati di Dell Technologies particolarmente rilevanti per le organizzazioni che utilizzano Cloudera®, come ha fatto il detentore di record Dell Technologies, per archiviare grandi quantità di dati nell'edge per l'analisi e l'AI.

Alla base dei risultati delle prestazioni

Grazie all'ottimizzazione della piattaforma PowerEdge basata su processori AMD, Dell Technologies detiene numerosi record mondiali nella misurazione delle prestazioni di AI e ML. Nell'ambito di un'implementazione on-premise, i processori AMD EPYC di terza generazione offrono già prestazioni elevate, prestazioni per watt e prestazioni per CPU in dollari. Nel cloud, i sistemi AMD EPYC on chip (SoC) alimentano istanze IaaS (Infrastructure-as-a-Service) ottimizzate per HPC (High Performance Computing) per molti provider di servizi cloud (CSP), tra cui Amazon Web Services® (AWS®), Microsoft® Azure®, Google Cloud Platform™ e altri.

I processori AMD EPYC di quarta generazione utilizzati nelle configurazioni di server da record mondiale di Dell per TPCx-AI garantiscono prestazioni ottimizzate che si traducono in diversi miglioramenti della piattaforma rispetto alla piattaforma della generazione precedente, tra cui:

- Aumento del 50% del numero di core⁴, maggior numero di thread e frequenze più elevate, che possono aumentare direttamente le prestazioni di elaborazione.
- 12 DIMM/socket (rispetto a 8), che consentono alle organizzazioni di aumentare in modo significativo la memoria disponibile. Tale aspetto si traduce nell'elaborazione più rapida di dataset di dimensioni maggiori, in particolare per l'analisi in-memory come quelle eseguite da Apache Spark™.
- Supporto della memoria DDR5 per un accesso più rapido ai dati.
- Supporto AVX-512, che consente ai processori AMD EPYC di quarta generazione il completamento simultaneo di più calcoli nei propri registri.
- Maggiore capacità della cache L2, raddoppiata da 512 KiB a 1 MiB per core, che accelera anche le operazioni all'interno della memoria.
- Supporto PCIe di quinta generazione, che consente interconnessioni più rapide per spostare un maggior numero di dati con minore latenza.
- Miglioramenti progettati appositamente per i carichi di lavoro di AI e ML, compreso il supporto per il tipo numerico bfloat16 per accelerare l'addestramento dei modelli di AI e il supporto per l'inferenza INT8 per aumentare le prestazioni dei modelli già addestrati nella fase di produzione.

In generale, i processori AMD EPYC di quarta generazione funzionano in modo più efficiente rispetto ai loro predecessori. I risultati della frequenza in virgola mobile per SPEC CPU® 2017 di Standard Performance Evaluation Corporation mostrano un aumento in termini di prestazioni del 121% nei test eseguiti in un sistema basato su processori AMD EPYC di quarta generazione, rispetto a un sistema basato su processori AMD EPYC di terza generazione⁵. I risultati della frequenza con numero intero per SPEC CPU 2017 hanno mostrato aumenti del 102%⁶. Tali risultati delle prestazioni dei processori si riflettono nei risultati di benchmark da record mondiale ottenuti da diverse piattaforme PowerEdge che abbiamo esaminato.

Il numero di core in questi processori è aumentato del 50%, rispetto alla generazione precedente, determinando di conseguenza un aumento delle prestazioni. Allo stesso tempo, le specifiche pubblicate da AMD mostrano un aumento in termini di consumo energetico predefinito massimo di solo il 42%, da un TDP (Thermal Design Power) di 280 watt a un TDP massimo di 400 watt⁷. Rispetto ai risultati delle prestazioni SPEC indicate in precedenza, questi numeri relativi all'alimentazione mostrano la capacità dei server basati su processori AMD EPYC di quarta generazione di fornire un vantaggio in termini di prestazioni di alimentazione fino al 55% per le aziende che eseguono carichi di lavoro di AI e ML⁸.

Il benchmark TPCx-AI, che misura le prestazioni di AI e ML, riflette anche le prestazioni presentate nella tabella dalle schede di rete Broadcom e dalle schede Dell PERC. Le organizzazioni che implementano carichi di lavoro di AI e ML necessitano di controller RAID per le ridondanze al fine di soddisfare i requisiti interni o richiesti dalle normative vigenti. I server PowerEdge R6625 e PowerEdge R7615 che hanno stabilito i record a livello mondiale di TPCx-AI sono stati dotati di schede Dell PERC con supporto NVMe RAID rapido.

Sicurezza basata sull'hardware AMD®

Per tutti i carichi di lavoro valutati nel presente studio di ricerca, sono fondamentali le considerazioni sulla sicurezza. I processori AMD EPYC™ di terza generazione e AMD EPYC di quarta generazione sono in grado di garantire la sicurezza basata sull'hardware per carichi di lavoro di AI e ML. AMD® Secure Memory Encryption (AMD® SME) crittografa la memoria di sistema per proteggere i dati in uso. AMD® Secure Encrypted Virtualization (AMD® SEV) protegge le macchine virtuali (VM) in esecuzione in modo che vengano crittografate e isolate l'una dall'altra e dall'hypervisor del sistema host. AMD® Secure Encrypted Virtualization-Encrypted State (AMD® SEV-ES) crittografa il contenuto del registro della CPU delle VM che sono state arrestate per contribuire a proteggere i dati archiviati al loro interno. Inoltre, AMD® Secure Boot contribuisce a proteggere i server durante il processo di avvio, garantendo le difese da rootkit, bootkit e firmware nei momenti in cui i server sono più vulnerabili.

Le schede di rete di Broadcom accelerano il flusso di dati per i carichi di lavoro di AI/ML. Il supporto per PCIe di quinta generazione in processori AMD EPYC di quarta generazione e NIC Broadcom consente l'utilizzo di schede di rete Broadcom da 100 GbE basate sul fattore di forma Open Compute Project (OCP) NIC 3.0. Queste progettazioni moderne riflettono un rapido spostamento nel settore verso le schede da 100 GbE basate su un fattore di forma più efficiente e abilitate PCIe 4.0 e PCIe 5.0. Inoltre, il supporto per PCIe 4.0 e PCIe 5.0 può garantire i numeri relativi alle prestazioni da una singola NIC allo stesso livello di due NIC da 100 Gbps. Le specifiche tecniche di OCP NIC 3.0 consentono ai produttori di server come Dell Technologies di utilizzare progettazioni più compatte in grado di supportare schede ad alte prestazioni con funzionalità avanzate di accelerazione hardware per velocizzare ulteriormente i carichi di lavoro di AI e ML⁹.

Il controller RAID Dell™ PowerEdge™ (Dell™ PERC) protegge i dati e aumenta le prestazioni di storage

Le moderne interfacce RAID PCIe® di quarta generazione funzionano con unità SSD NVM Express® (NVMe®) ad alta larghezza di banda per aumentare in modo significativo le prestazioni di storage. Le due schede 11 e 12 del controller RAID Dell PowerEdge (PERC 11 e PERC 12) e le schede NVMe con host PCIe di quarta generazione e interfacce di storage PCIe di quarta generazione possono contribuire a rimuovere i vincoli di larghezza di banda e latenza.



Conclusioni

Per i produttori di server, i risultati dei benchmark in generale (e i risultati da record mondiale in particolare) vanno oltre il semplice vantarsi. Quando vengono interpretati correttamente, i migliori risultati del settore a livello di benchmark sono in grado di offrire informazioni dettagliate sulla modalità di esecuzione dei server nei casi d'uso reali. Grazie alla quota di mercato di Dell Technologies e al numero di record mondiali detenuto dall'azienda, i server PowerEdge offrono una naturale opportunità di esaminare in che modo i risultati di benchmark possono mapparsi ai vantaggi in termini di prestazioni per le organizzazioni in fase di produzione. Sebbene nessun mapping delle prestazioni di benchmark (record mondiale o altro) sia univoco, la nostra indagine mostra che le prestazioni dei server Dell PowerEdge in un benchmark riconosciuto a livello di settore indicano elevate prestazioni di AI e ML per diversi casi d'uso in una varietà di settori. Inoltre, il numero di record a livello mondiale detenuto da Dell Technologies su diversi benchmark dimostra come l'azienda abbia sviluppato piattaforme che sfruttano i punti di forza dei singoli componenti per offrire un valore reale ai propri clienti per una varietà di carichi di lavoro.

Appendice A: link alle prestazioni di benchmark

- Risultati delle migliori prestazioni TPCx-AI: www.tpc.org/tpcx-ai/results/tpcxai_perf_results5.asp
- Risultati delle migliori prestazioni TPCx-IoT V2: www.tpc.org/tpcx-iot/results/tpcxiot_perf_results5.asp?version=2

Appendice B: link alle specifiche di sistema di Dell Technologies

- Specifiche tecniche dei server Dell PowerEdge: www.dell.com/en-us/dt/servers/poweredge-rack-servers.htm

¹ History-Computer. "The 10 Largest Server Companies In The World, And What They Do". Settembre 2022. <https://history-computer.com/largest-server-companies-in-the-world-and-what-they-do/>.

² TPC. "TPCx-AI Top Performance Results". Consultato il 1° novembre 2022. www.tpc.org/tpcx-ai/results/tpcxai_perf_results5.asp

³ TPC. "TPCx-IoT V2 Top Performance Results". Consultato il 1° novembre 2022. www.tpc.org/tpcx-iot/results/tpcxiot_perf_results5.asp?version=2.

⁴ Tom's Hardware. "Zen 4 Madness: AMD EPYC Genoa With 96 Cores, 12-Channel DDR5 Memory, and AVX-512." Agosto 2021. www.tomshardware.com/news/zen4-madness-amd-epyc-genoa-with-96-cores-12-channel-ddr5-memory-and-avx-512.

⁵ Fino al 121% di prestazioni in virgola mobile SPEC® superiori tramite il confronto dei migliori processori AMD EPYC™ di quarta generazione con i migliori processori AMD EPYC di terza generazione in base al punteggio della frequenza in virgola mobile SPEC pari a 1.410 ottenuto da un server Dell™ PowerEdge™ R7625 basato su processori AMD EPYC 9654, rispetto a un punteggio di 636 ottenuto da un server Dell PowerEdge R7525 basato su processori AMD EPYC 7763. Punteggi consultati il 10 novembre 2022. Consultare i risultati di benchmark di Standard Performance Evaluation Corporation. <http://spec.org/benchmarks.html>.

⁶ Fino al 102% di prestazioni della frequenza con numero intero SPEC® superiori tramite il confronto tra i migliori processori AMD EPYC™ di quarta generazione e i migliori processori AMD EPYC di terza generazione in base al punteggio della frequenza con numero intero SPEC pari a 1.660 ottenuto da un server Dell™ PowerEdge™ R7625 basato su processori AMD EPYC 9654, rispetto a un punteggio di 821 ottenuto da un server Dell PowerEdge R7525 basato su processori AMD EPYC 7763. Punteggi consultati il 10 novembre 2022. Consultare i risultati di benchmark di Standard Performance Evaluation Corporation. <http://spec.org/benchmarks.html>.

⁷ AMD. Pagina web relativa alle specifiche dei processori AMD EPYC serie 7003. www.amd.com/en/processors/epyc-7003-series.

⁸ Miglioramento del 55% delle prestazioni della CPU per watt calcolato utilizzando il punteggio in virgola mobile SPEC® pari a 1.410 ottenuto da un server Dell™ PowerEdge™ R7625 basato su processori AMD EPYC 9654 con cTDP del processore di 400 watt, rispetto a un punteggio di 636 ottenuto da un server Dell PowerEdge R7525 basato su processori AMD EPYC 7763 con cTDP del processore di 280 watt.

⁹ Broadcom. Compendio sulle schede OCP NIC 3.0 Ethernet serie E NetXtreme. 2021. <https://docs.broadcom.com/doc/12395120>.

