

Implementazione della piattaforma AI

Supera la complessità delle soluzioni AI per accelerare la creazione di valore



La potenza trasformativa dell'AI e dell'AI generativa richiede un'azione rapida. Molte organizzazioni hanno difficoltà a gestire infrastrutture IT obsolete, modelli operativi inefficienti e carenze di competenze in aumento, mentre le crescenti richieste di carichi di lavoro AI creano una notevole complessità. Le competenze tecniche sono fondamentali per implementare e integrare la tecnologia in modo da non compromettere velocità, efficienza dei costi e scalabilità.

Ottenere soluzioni personalizzate, sicure e affidabili che soddisfino i casi d'uso e gli obiettivi aziendali è possibile tramite fornitori di servizi di fiducia. Superando rapidamente queste sfide, la tua organizzazione sbloccherà il pieno potenziale dell'AI, promuovendo valore misurabile e successo a lungo termine.

92%

delle organizzazioni prevede che la percentuale di dipendenti con accesso alla piattaforma AI aumenterà nel prossimo anno¹

50%

manca di fiducia nella propria capacità di cercare, selezionare, implementare e gestire soluzioni AI per soddisfare i requisiti aziendali²

84%

delle organizzazioni preferisce esternalizzare/collaborare con un fornitore di servizi per implementare soluzioni AI/GenAI³



I servizi Dell affrontano le sfide legate alla progettazione, all'implementazione e all'utilizzo di soluzioni AI per prestazioni, scalabilità ed efficienza migliorate. Affidati agli esperti per valutare le tue esigenze e fornire le soluzioni giuste.

- ✓ Operazioni semplificate
- ✓ Adozione migliorata
- ✓ Massimizzazione del ROI
- ✓ Scalabilità

Complessità dell'implementazione

Mancanza di chiarezza sulla soluzione AI giusta

Mancanza di competenze tecniche per l'implementazione

Integrazione nei flussi di lavoro esistenti

Possibilità di dimensionamento e miglioramento continuo

Incertezza sugli investimenti



Effetti sul ROI

Risultati di business non allineati

Creazione di valore ritardata

Inefficienza operativa

Adozione lenta e battute d'arresto della concorrenza

Opportunità di crescita perse

Accelerata il time-to-value delle soluzioni on-premise a costi contenuti

- ✓ Su misura in base alle esigenze aziendali
- ✓ Ottimizzato per le prestazioni
- ✓ Massimizzazione per la scalabilità
- ✓ Adozione e utilizzo migliorati
- ✓ Operazioni semplificate tramite orchestration e automazione
- ✓ Output migliorati dai dati integrati
- ✓ Ottimizzazione dei risultati grazie alle competenze dell'ecosistema aperto

Gli esperti dei servizi Dell ti aiutano a superare la complessità dell'implementazione e a raggiungere l'eccellenza delle soluzioni AI. Dalla configurazione dell'infrastruttura e dalla selezione della piattaforma, all'integrazione e all'ottimizzazione fino all'efficienza operativa con possibilità di dimensionamento.

	Focus Areas	Use Cases
Application Infrastructure	Training Frameworks	Inference
	Foundational Models / Model Training	
	MLOps / LLMOps	
Software Infrastructure	Workload Manager	
	Driver / Compiler / Toolchain	
	Operating Systems	
	Deployment Tools	
Hardware Infrastructure	Infrastructure from device to data center Dell compute, storage and networking	Partner GPUs

86%

time-to-value più rapido
grazie a soluzioni full-stack
pronte all'uso con servizi⁴



Progettazione e pianificazione

- Identifica i modelli di utilizzo dell'AI, la strategia di ottimizzazione GPU/rete, i componenti dello stack AI per la scelta del vendor, le considerazioni sulle prestazioni e così via
- Esamina esigenze di conformità e predisposizione dei dati
- Valutazione completa della predisposizione del data center alle funzionalità di AI
- Assicurati che le soluzioni siano allineate agli obiettivi e che possano essere facilmente scalate per la produzione e per più casi d'uso
- Workshop per definire esigenze di integrazione specifiche



Implementazione dell'infrastruttura

- Creazione e integrazione di rack in fabbrica
- Cablaggio e connettività tra rack GPU e switch
- Gestione dell'alimentazione e connettività del sistema di raffreddamento (raffreddamento a liquido o ad aria)
- Creazione del cluster, test e convalida del rack completo
- Configurazione e test della fabric di rete
- Test di accettazione dell'intero cluster (HPL, NCCL, ecc.)
- Trasferimento delle conoscenze sull'hardware implementato



Implementazione della piattaforma

- Implementa e configura lo stack software selezionato
- Implementa e dimostra il server di inferenza con LLM pre-addestrato
 - Crea un esempio di carico di lavoro di addestramento del modello (ottimizzazione) per dimostrare il funzionamento dell'ambiente
 - Fornisci strumenti/ configurazioni aggiuntivi in base ai risultati desiderati
 - Script di test per dimostrare la funzionalità
- Piano di adozione della piattaforma AI



Adozione e scalabilità

- Integra con piattaforme di storage, Observability e registrazione. Miglioramento delle prestazioni del sistema con monitoraggio per consentire un'allocazione ottimale della GPU
- Sviluppa procedure operative e flussi di lavoro a supporto di operazioni di produzione migliorate
- Sviluppo avanzato con strumenti containerizzati, ottimizzazione dei modelli e risoluzione dei problemi
- Porta le competenze a un livello superiore con esperienza in materia per dimensionare i carichi di lavoro AI
- Migliora l'utilizzo con esperti residenti dedicati
- Semplifica le operazioni con soluzioni completamente gestite

Giorno 0

Giorno 1

Dopo
il 2° giorno

Non sai da dove iniziare? Prendi in considerazione un [Accelerator Workshop](#) o un progetto pilota per casi d'uso per iniziare rapidamente, parla oggi stesso con un rappresentante Dell per saperne di più

1. ESG, "There's no time to waste: Confidently moving ahead with Enterprise AI", Mike Leone, novembre 2024

2. ESG, "The AI Maturity Playbook: Driving Business Impact and Innovation", Mike Leone, commissionato da Dell Technologies, i risultati effettivi possono variare

3. Studio su commissione condotto da Forrester Consulting per conto di Dell Services, settembre 2024

4. Stima basata su un'analisi condotta da Dell a maggio 2024, in cui è stato confrontato il tempo necessario per configurare un cluster Kubernetes® a due nodi per un LLM per scopi generici utilizzando script automatizzati e l'implementazione manuale di una progettazione comune. Il tempo di configurazione include solo l'installazione di base. Il tempo di configurazione effettivo varia a seconda della configurazione della soluzione



Scopri i [Servizi AI Dell](#)



[Contatta](#) un esperto Dell Technologies



Partecipa alla conversazione con #DellTechnologies