

Implementierung von KI-Plattformen

Schneller zum Mehrwert durch weniger Komplexität bei KI-Lösungen



Angesichts der transformativen Möglichkeiten von KI und GenAI ist schnelles Handeln entscheidend. Veraltete IT-Infrastrukturen, ineffiziente Betriebsabläufe und wachsende Kompetenzlücken stellen viele Unternehmen vor Herausforderungen – hinzu kommt eine deutlich höhere Komplexität, die mit den zunehmenden Anforderungen an KI-Anwendungen einhergeht. Technisches Fachwissen ist entscheidend, um Technologien so zu implementieren und zu integrieren, dass Geschwindigkeit, Kosteneffizienz und Skalierbarkeit nicht beeinträchtigt werden.

Vertrauenswürdige Serviceanbieter ermöglichen dabei maßgeschneiderte, sichere und zuverlässige Lösungen für Ihre Anwendungsfälle und Geschäftsziele. Wenn Ihr Unternehmen diese Herausforderungen schnell bewältigt, können Sie das volle Potenzial von KI ausschöpfen und dadurch einen messbaren Mehrwert und langfristigen Erfolg erzielen.

92 %

der Unternehmen rechnen damit, dass der Anteil der MitarbeiterInnen mit Zugriff auf KI-Plattformen im Laufe des nächsten Jahres steigt¹

50 %

sehen sich nicht ausreichend befähigt, geeignete KI-Lösungen zu recherchieren, auszuwählen, bereitzustellen und zu managen, um geschäftliche Anforderungen zu erfüllen²

84 %

der Unternehmen setzen bei der Implementierung von KI- und GenAI-Lösungen bevorzugt auf externe Dienstleister oder eine Umsetzung im Rahmen von Partnerschaften³



Dell Services unterstützt Unternehmen dabei, die Herausforderungen bei Entwicklung, Implementierung und Betrieb von KI-Lösungen zu meistern – für mehr Performance, Skalierbarkeit und Effizienz. Verlassen Sie sich bei der Bewertung Ihrer Anforderungen und der Bereitstellung der richtigen Lösungen auf ExpertInnen.

- ✓ Optimierte Betriebsabläufe
- ✓ Verbesserte Einführung
- ✓ Maximierter ROI
- ✓ Flexible Skalierung

Implementierungskomplexität

Mangelnde Klarheit über die richtige KI-Lösung

Lücken bei technischen Kompetenzen für Implementierungen

Integration in bereits vorhandene Workflows

Skalierbarkeit und kontinuierliche Weiterentwicklung

Unsicherheit in Hinblick auf Investitionen



Auswirkungen auf den ROI

Falsch ausgerichtete Geschäftsergebnisse

Verzögerte Wertschöpfung

Betriebliche Ineffizienz

Langsame Einführung und negative Folgen für die Wettbewerbsfähigkeit

Verpasste Wachstumschancen

Schnellere Time-to-Value bei kosteneffizienten On-Premise-Lösungen

- ✓ Zugeschnitten auf Geschäftsanforderungen
- ✓ Optimiert für Performance
- ✓ Auf Skalierung ausgerichtet
- ✓ Verbesserte Einführung und Nutzung
- ✓ Optimaler Betrieb dank Orchestrierung und Automatisierung
- ✓ Verbesserte Ergebnisse durch integrierte Daten
- ✓ Maximierte Ergebnisse dank Fachwissen zu offenen Ökosystemen

Die ExpertInnen von Dell Services unterstützen Sie dabei, komplexe Implementierungsprozesse zu bewältigen und leistungsstarke KI-Lösungen erfolgreich umzusetzen – von der Einrichtung der Infrastruktur und der Auswahl, Integration und Optimierung der Plattform bis hin zur Betriebseffizienz mit der Möglichkeit zur Skalierung.

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

86 %

schnellere Time-to-Value durch gebrauchsfertige Full-Stack-Lösungen mit Services⁴



Entwicklung und Planung

- Identifizierung von KI-Nutzungsmustern, GPU-/Netzwerkoptimierungsstrategien, KI-Stack-Komponenten für die Anbieterwahl, Performanceüberlegungen usw.
- Überprüfung der Datenqualität und der Compliance-Anforderungen
- Umfassende Bewertung der Rechenzentrumsbereitschaft für KI
- Sicherstellung von Zielausrichtung der Lösungen sowie einfacher Skalierbarkeit für Produktion und Anwendungsvielfalt
- Workshop zur Bestimmung spezifischer Anforderungen für Integrationen



Bereitstellung der Infrastruktur

- Rack-Aufbau und Integration im Werk
- Konnektivität zwischen Racks und Verkabelung zwischen GPU-Racks und Switches
- Energiemanagement und Konnektivität des Kühlungssystems (Flüssigkeits- oder Luftkühlung)
- Clusteraufbau sowie vollständige Rack-Tests und -Validierung
- Netzwerk-Fabric-Konfiguration und -Tests
- Annahmetests des gesamten Clusters (HPL, NCCL usw.)
- Wissenstransfer im Zusammenhang mit der bereitgestellten Hardware



Implementierung der Plattform

- Implementierung und Konfiguration ausgewählter Software-Stacks
 - Implementierung und Demonstration des Inferencing-Servers mit vortrainiertem LLM
 - Erstellung einer Beispielmodell-Workload für Feinabstimmung und Training zur Veranschaulichung der Umgebung
 - Bereitstellung zusätzlicher Tools/Konfigurationen je nach gewünschten Ergebnissen
 - Testen von Skripten zur Veranschaulichung der Funktionalität
- Plan für die Einführung der KI-Plattform



Einführung und Skalierung

- Integration in Plattformen für Storage, Observability und Protokollierung; verbesserte Systemperformance mit Monitoring für eine optimale GPU-Zuweisung
- Entwicklung von Betriebsabläufen und Workflows zur Unterstützung verbesserter Produktionsprozesse
- Vorantreiben der Entwicklung mit containerbasierten Tools, Modell-Tuning und Troubleshooting
- Weiterentwicklung von Kompetenzen durch Fachkenntnisse zur Skalierung von KI-Workloads
- Optimierte Auslastung durch die Einbindung von fest eingeteilten ExpertInnen vor Ort
- Optimierung des Betriebs mit vollständig gemanagten Lösungen

Tag 0

Tag 1

Ab Tag 2

Sie sind sich nicht sicher, wann der günstigste Zeitpunkt für den Einstieg ist?
Ein [Accelerator Workshop](#) oder ein Pilotprojekt zu konkreten Anwendungsfällen hilft Ihnen beim unkomplizierten Einstieg. Sprechen Sie noch heute mit einem/r Dell VertriebsmitarbeiterIn, um mehr zu erfahren.

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

2. ESG „The AI Maturity Playbook: Driving Business Impact and Innovation“, Mike Leone im Auftrag von Dell Technologies, tatsächliche Ergebnisse können abweichen

3. Studie von Forrester Consulting im Auftrag von Dell Services, September 2024

4. Schätzung basierend auf einer Analyse von Dell vom Mai 2024, in der die Zeit für die Einrichtung eines Kubernetes®-Clusters mit zwei Nodes für ein allgemeines LLM mithilfe automatisierter Skripte mit der manuellen Bereitstellung eines gemeinsamen Designs verglichen wurde. Die Einrichtungszeit umfasst nur die Basisinstallation. Die tatsächliche Einrichtungszeit variiert je nach Lösungskonfiguration.



Dell [AI Services](#) erkunden



[Kontakt](#) zu einem/r
Dell Technologies ExpertIn



Reden Sie mit:
#DellTechnologies