

AI 플랫폼 구현

AI 솔루션의 복잡성을 극복하여 가치 창출 가속화



AI와 Generative AI의 혁신적인 성능은 신속한 조치를 요구합니다. 많은 조직이 노후화된 IT 인프라스트럭처, 비효율적인 운영 모델, 그리고 심화되는 기술 격차로 어려움을 겪고 있는 가운데, AI 워크로드에 대한 수요 증가는 상당한 복잡성을 야기합니다. 속도, 비용 효율성, 확장성을 저해하지 않으면서 기술을 구현하고 통합하려면 기술적 전문 지식이 필수적입니다.

신뢰할 수 있는 서비스 공급업체를 통해 활용 사례와 비즈니스 목표를 충족하는 안전하고 신뢰할 수 있는 맞춤형 솔루션을 구현할 수 있습니다. 이러한 과제를 신속하게 극복함으로써 귀사는 AI의 잠재력을 최대한 활용하여 측정 가능한 가치와 장기적인 성공을 달성할 수 있습니다.

92%

내년에 AI 플랫폼에 액세스할 수 있는 직원 비율이 증가할 것으로 예상하는 조직의 비율¹

50%

비즈니스 요구 사항을 충족하기 위해 AI 솔루션을 연구, 선택, 구축 및 관리할 수 있는 능력에 대한 확신이 부족함²

84%

AI/GenAI 솔루션을 구현하기 위해 서비스 공급업체와 아웃소싱/파트너십을 맺는 것을 선호하는 조직의 비율³



Dell 서비스는 향상된 성능, 확장성 및 효율성을 위한 AI 솔루션의 설계, 구현 및 운영 과제를 해결합니다. 요구 사항을 평가하고 최적의 솔루션을 제공하는 전문가의 도움을 받을 수 있습니다.

- ✓ 운영 간소화
- ✓ 도입 개선
- ✓ ROI 극대화
- ✓ 유연한 확장

구현 복잡성

적합한 AI 솔루션에 대한 명확성 부족

구현을 위한 기술 역량 격차

기존 워크플로에 통합

확장 및 지속적인 개선 능력

투자 관련 불확실성



ROI에 미치는 영향

요구 사항에 맞지 않는 비즈니스 성과

가치 창출 지연

운영 비효율성

느린 도입 및 경쟁력 저하

성장 기회 상실

비용 효율적인
온프레미스 솔루션의
가치 실현 시간 단축

- ✓ 비즈니스 요구 사항에 맞춤화
- ✓ 성능을 위한 최적화
- ✓ 규모에 맞게 극대화
- ✓ 도입 및 활용도 개선
- ✓ 오케스트레이션 및 자동화를 통한 운영 간소화
- ✓ 통합 데이터의 향상된 출력
- ✓ 개방형 생태계 전문 지식으로 성과 극대화

Dell 서비스 전문가가 구현 복잡성을 극복하고 탁월한 AI 솔루션을 구현할 수 있도록 지원합니다. 인프라스트럭처 설정 및 플랫폼 선택, 통합 및 최적화부터 확장 가능한 운영 효율성까지 제공합니다.

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%

서비스가 포함된 턴키 방식의 풀스택 솔루션을 통해 가치 실현 시간 단축⁴



설계 및 계획

- AI 사용 패턴, GPU/네트워크 최적화 전략, 공급업체 선택을 위한 AI 스택 구성 요소, 성능 고려 사항 등을 파악
- 데이터 준비 상태 및 규정 준수 요구 사항 검토
- AI 역량을 갖추기 위한 데이터 센터 준비 상태를 종합적으로 평가
- 솔루션이 목표에 부합하고 운영 및 여러 활용 사례에 맞게 쉽게 확장될 수 있도록 보장
- 특정 통합 요구 사항을 정의하는 워크숍



인프라스트럭처 배포

- 공장 랙 구축 및 통합
- 랙 간 연결성 및 GPU 랙과 스위치 간 케이블 연결
- 전원 관리 및 냉각 시스템 연결(액체 또는 공기 냉각식)
- 클러스터 구축과 전체 랙 테스트 및 검증
- 네트워크 패브릭 구성 및 테스트
- 전체 클러스터(HPL, NCCL 등)의 인수 테스트
- 구축된 하드웨어에 대한 지식 전달



플랫폼 구축

- 선택한 소프트웨어 스택 구현 및 구성
- 사전 학습된 LLM으로 추론 서버 구현 및 시연
- 워크로드를 교육(미세 조정)하는 견본 모델을 생성하여 환경 운영 시연
- 원하는 결과를 기반으로 추가 튜닝/구성 제공
- 기능 시연을 위한 테스트 스크립트
- AI 플랫폼 도입 계획



도입 및 확장

- 스토리지, Observability 및 로깅 플랫폼과 통합하고, 최적의 GPU 할당을 가능하게 하는 모니터링을 통해 시스템 성능 개선
- 향상된 생산 운영을 지원하기 위한 운영 절차 및 워크플로 개발
- 컨테이너화된 툴링, 모델 튜닝 및 문제 해결을 통한 고급 개발
- AI 워크로드 확장을 위한 분야별 전문 지식으로 역량 강화
- 전담 상주 전문가를 통해 활용도 향상
- 완전 관리형 솔루션으로 운영 간소화

0일차

1일차

2일 차 이후

어디서부터 시작해야 할지 잘 모르시나요? [Accelerator Workshop](#) 또는 활용 사례 파일럿으로 빠르게 시작해 보십시오. 지금 Dell 담당자에게 문의하여 자세한 내용을 알아보십시오.

1. ESG, "There's no time to waste: Confidently moving ahead with Enterprise AI", Mike Leone, 2024년 11월

2. Dell Technologies의 의뢰로 ESG의 Mike Leone가 수행한 "The AI Maturity Playbook: Driving Business Impact and Innovation" 연구 기준, 실제 결과는 다를 수 있음.

3. Dell 서비스의 의뢰로 Forrester Consulting에서 실시한 연구, 2024년 9월

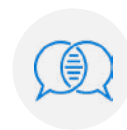
4. 범용 LLM을 위한 2노드 Kubernetes® 클러스터를 설치하는 데 필요한 시간을 자동화된 스크립트를 사용하는 경우와 공용 설계를 수동으로 배포하는 경우로 나눠서 비교한 2024년 5월 Dell 분석 결과를 기준으로 추정했습니다. 설치 시간은 기본 설치만 포함합니다. 실제 설치 시간은 솔루션 구성에 따라 다릅니다.



Dell [AI Services](#) 둘러보기



Dell Technologies
전문가에게 [문의](#)



대화에 참여:
[#DellTechnologies](#)