

AI 平台实施

解决 AI 解决方案的复杂性难题，加快价值创造



AI 和生成式 AI 带来的变革冲击需要组织快速采取行动。过时的 IT 基础架构、低效的运营模式以及不断扩大的技能差距让许多组织疲于应对，同时 AI 工作负载需求的日益增长更是带来了显著的复杂性。采用不影响速度、成本效益和可扩展性的方式来实施和集成技术至关重要。

选择值得信赖的服务提供商，打造量身定制、安全且可靠的解决方案，可以满足应用场景要求并实现业务目标。通过迅速克服这些挑战，贵组织将释放 AI 的全部潜力，推动可衡量的价值和长期成功。

92%

的组织预计未来一年内可访问
AI 平台的员工比例将增加¹

50%

的组织对研究、选择、部署和
管理 AI 解决方案以满足业务需
求的能力缺乏信心²

84%

的组织更倾向于外包或与服务
提供商合作以实施 AI/生成式
AI 解决方案³



戴尔服务致力于解决设计、实施和运营 AI 解决方案的相关难题，从而提升性能、可扩展性和效率。
贵组织可以依靠专家来评估需求并交付合适的解决方案。

- ✓ 简化运营
- ✓ 提高采用率
- ✓ 显著提升 ROI
- ✓ 可灵活调整规模

实施的复杂性

缺乏对合适 AI 解决方案的清晰认识

实施方面的技术技能差距

与现有工作流的整合问题

扩展和持续改进的能力不足

投资的不确定性



对 ROI 的影响

业务成果不一致

价值创造延迟

运营效率低下

采用速度缓慢，竞争优势不足

错失增长机会

采用经济高效的本地解决方案，加速实现价值

- ✓ 根据业务需求量身定制
- ✓ 优化性能
- ✓ 规模最大化
- ✓ 提高采用率和利用率
- ✓ 通过编排和自动化来简化运营
- ✓ 从集成数据中增强输出
- ✓ 借助开放生态系统专业知识实现最大化成果

戴尔服务专家可帮助您克服实施相关的复杂性难题，打造卓越的 AI 解决方案。从基础架构设置和平台选择到集成和优化，再到运营效率的提升及扩展空间的支持，我们可以提供全方位服务。

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%

的价值实现速度提升幅度 — 通过一站式全堆栈解决方案实现⁴



设计和规划

- 确定 AI 使用模式、GPU/网络优化策略、AI 堆栈组件（用于选择供应商）、性能注意事项等
- 审查数据的就绪性和合规性需求
- 全面评估数据中心是否为 AI 功能做好准备
- 确保解决方案与目标保持一致，并且可以轻松扩展以支持生产及多种应用场景
- 举办研讨会，明确具体集成需求



部署基础架构

- 工厂机架构建和集成
- 机架间连接及 GPU 机架与交换机之间的布线
- 电源管理和冷却系统连接（液冷或风冷）
- 群集构建与全机架测试及验证
- 网络构造配置和测试
- 整个群集的验收测试（HPL、NCCL 等）
- 部署硬件相关的知识传授



实施平台

- 实施和配置选定的软件堆栈
- 使用预先训练的 LLM 实施和演示推理服务器
 - 创建示例模型训练（微调）工作负载，以演示环境的运行
 - 根据期望的成果提供额外的工具/配置
 - 测试脚本以演示功能
- AI 平台采用计划



采用和扩展

- 与存储、可观测性和日志平台集成；通过监控提升系统性能，实现最佳 GPU 分配
- 制定运营程序和工作流，为改进生产运营提供支持
- 通过容器化工具、模型调优和故障处理，推动开发进程
- 通过扩展 AI 工作负载相关的主题专业知识，提升技能
- 通过专属的派驻专家，提高利用率
- 通过完全托管的解决方案，简化运营

第 0 天

第 1 天

第 2 天以后

不知道从何处着手？考虑参加 [Accelerator Workshop](#) 或应用场景试点以快速着手开始，请立即联系戴尔代表了解详情

1.ESG, 《There's no time to waste: Confidently moving ahead with Enterprise AI》, Mike Leone, 2024 年 11 月

2.ESG, 《The AI Maturity Playbook: Driving Business Impact and Innovation》, Mike Leone, 受 Dell Technologies 委托撰写, 实际结果可能有所不同

3.Forrester Consulting 受委托代表 Dell Services 进行的研究, 2024 年 9 月

4.基于戴尔在 2024 年 5 月进行的分析得出的估计值，这项分析比较了使用自动化脚本为通用 LLM 设置双节点 Kubernetes® 群集与手动部署通用设计群集所需的时间。设置时间仅包括基础安装时间。

实际设置时间将因解决方案配置而异



探索戴尔 [AI 服务](#)



联系 Dell
Technologies 专家



参与
#DellTechnologies 交流