

Dell AI Factory

与 AMD 联手推出的戴尔生成式 AI 解决方案

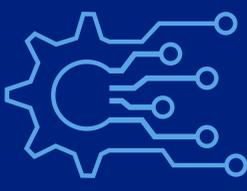
利用可扩展的模块化体系结构支持复杂的生成式 AI，从而加速创新、降低成本并保护数据。



关键应用场景需要强大的性能、灵活性和规模



智能助手、聊天机器人和内容创建



加速器即服务



多模态检索增强生成 (RAG)



化繁为简

借助经验证的成熟解决方案，精简生成式 AI 部署，这些解决方案历经超过 34 万小时的工程开发，拥有坚实保障。

优化性能

高性能加速器、开放式体系结构和 AI 优化型结构

无处不在的 AI

借助多云存储的灵活性，随处利用数据

全包式多节点

经验证的全堆栈 AI 基础，可更快地交付成果



量身定制

AMD ROCm™ 开源软件和开放式生态系统可推动 AI 开发和运营。

创新更迅捷

使用开源软件和生态系统开发独特的应用程序。

加快开发速度

利用行业标准框架和灵活的技术堆栈。

激活您的数据

同时高效运行多个 AI 应用场景。



备受信赖

82% 的 ITDM 更倾向于采用本地或混合模式。³ 数据决定结果。应加以保护！

快速部署

具有信任根安全性和完全控制力的本地基础

简化连接

安全、功能丰富的结构，具有可扩展性和优化的传输流量

自动配置

为部署和管理高性能群集奠定开放源代码基础

采用 AMD 技术的戴尔生成式 AI 解决方案



推理

在单个 AMD Instinct™ MI300X 加速器上运行 70B 参数模型。⁴

定制

在单个 Dell PowerEdge XE9680 上部署和微调八个并发 70B 模型。⁴

扩增

将您的数据整合到生成过程中。

采用经验证的开放式解决方案规模化提供安全的本地 AI 应用程序，从而获得竞争优势。

[了解详情](#)

¹Enterprise Strategy Group, 充分提高 AI ROI: 与公有云相比, 使用 Dell Technologies 解决方案进行本地推理的经济效益可提高 75%, 2024 年 4 月。
²基于戴尔于 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设置双节点 Kubernetes 群集以部署通用 LLM 与手动部署通用设计所需的时间。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。

³Dell Technologies, Generative AI Pulse Survey, 2023 年 8 月和 9 月。
⁴Dell Technologies 博客, Silicon Diversity: Deploy GenAI on the PowerEdge XE9680 with AMD Instinct MI300X Accelerators, 2024 年 5 月。