

## 借助生成式 AI，更快获享高价值洞察

快速部署全堆栈解决方案，以进行生成式人工智能 (GenAI) 大型语言模型推理

### 提高生产效率和推动获得洞察

该联合体系结构采用灵活的模块化设计，可为各种应用场景提供支持并满足计算需求。您可以根据具体应用需求，选配组件并对组件进行独立扩展。

以下是受支持推理应用场景的一些典型示例：

**自然语言生成：**生成模型可用以完成各种文本生成任务，如文档编写、对话生成、摘要撰写或内容创建等

**聊天机器人和虚拟助理：**生成式 AI 根据用户查询或指令生成自然语言回复，为会话虚拟角色、聊天机器人和虚拟助理提供支持。

**代码开发：**在软件开发过程中使用补全代码、生成单元测试代码或聊天（用以说明代码含义）等功能，获取帮助。

Dell Technologies 携手 NVIDIA 推出功能强大的生成式 AI 解决方案，可生成质量更高、更快实现价值的预测和输出，帮助决策者更快做出决策。这款联合开发的解决方案可以解决延迟、响应度和计算需求等推理方面的难题，帮助企业将数据转化为更为智能的高价值成果。

借助创新的技术、全面的专业服务和广泛的合作伙伴生态系统，贵组织可以在企业规模加快生成式 AI 的利用速度。现在，IT 组织、数据科学家和 AI DevOps 可以针对生成式 AI 和 LLM 推理交付可扩展的模块化平台。



### 借助经验证的解决方案缩短收效时间

借助旨在简化采用且经验证的设计和参考体系结构，快速构建本地基础架构，以满足您应用程序的需求。通过降低采用过程中各个步骤的复杂度，您可以获取更多洞察，更快地做出决策，同时提高工作效率。

## 了解详情

- [请参阅《设计指南》](#)
- [AI InfoHub](#)
- [delltechnologies.com/ai](#)
- [Dell Technologies](#) 和 [NVIDIA](#)

## 什么是推理？

AI 推理是指使用经过训练的模型，根据输入数据生成预测、做出决策或产生输出的过程。在此期间，模型会将在训练阶段习得的知识模式和模式应用到初次接触的新数据上。

在推理过程中，经训练的模型会读取输入数据并利用其计算算法或神经网络体系结构对数据加以处理，从而生成输出或进行预测。该模型会利用习得的参数、权重或规则将输入数据转化为有意义的信息或行动。

在 AI 系统的生命周期中，推理阶段至关重要。在使用标记或未标记数据来训练模型，以此习得数据之间的模式和相关性之后，模型将在推理阶段归纳所学知识，并对训练阶段未见过的真实数据做出预测或响应。

## 我们帮助您更快取得成果

借助一系列服务，戴尔服务专家可以在您采用生成式 AI 的各个阶段为您提供帮助，让您更快地针对数据实现生成式 AI 的价值。

- **制定战略** — 制定路线图，实现 IT 和业务利益相关者的创新目标
- **实施** — 建立平台，利用戴尔经验验证的设计来实施生成式 AI 推理硬件和软件
- **采用** — 实施预先经过训练的推理模型，加速实现生成式 AI 应用场景的价值
- **扩展** — 戴尔将通过向贵方派遣常驻技术专家以及培训贵方团队，助您更好地掌握相关技能，以管理您的生成式 AI 创新产品组合

## 技术规格

经验验证的设计配置基于经 AI 加速优化的全新戴尔 [PowerEdge XE](#) 和机架式服务器，搭载有全新 NVIDIA GPU 和 NVIDIA AI Enterprise，配备 Triton 推理服务器和 NeMo 框架。由 [戴尔 PowerScale](#) 全闪存或混合存储阵列提供快速、充足的数据湖存储，可满足生成式 AI 和大型语言模型的需求。

计算	加速器	网络产品	软件	存储
戴尔 PowerEdge R760xa 服务器	NVIDIA A100 或 H100 GPU	NVIDIA 网络产品、戴尔 PowerSwitch S5232F-ON 或 S5248F-ON	Dell OpenManage Enterprise、Power Manager、CloudIQ。NVIDIA AI Enterprise，配备针对 LLM 的 Nemo 框架和 Triton 推理服务器；NVIDIA Base Command Manager Essentials	由戴尔 PowerScale、ECS 和 ObjectScale 提供支持

## Dell Technologies 和 NVIDIA

Dell Technologies 与 NVIDIA 携手合作，共同支持并加速生成式 AI 工作负载，提供经过工程验证的硬件和软件以加速 AI、ML 和 DL 工作负载，从而满足各个行业和垂直领域客户的需求。借助这款经过验证的面向 LLM 推理的设计方案，您可以运用实时数据大规模改善重大决策，从而加速数字化转型过程；您还可以通过制定经过优化的解决方案，更快速实现各项 AI 计划的价值。



详细了解  
戴尔解决方案



联系 Dell  
Technologies 专家



查看更多资源



加入 #HashTag 对话

© 2023 Dell Inc. 或其子公司。保留所有权利。Dell 和其他商标为 Dell Inc. 或其子公司的商标。SAP、SAP HANA、SAP S/4HANA 和 SAP Business One 是 SAP SE 在德国和其他国家/地区的注册商标。其他商标可能是其各自所有者的商标。