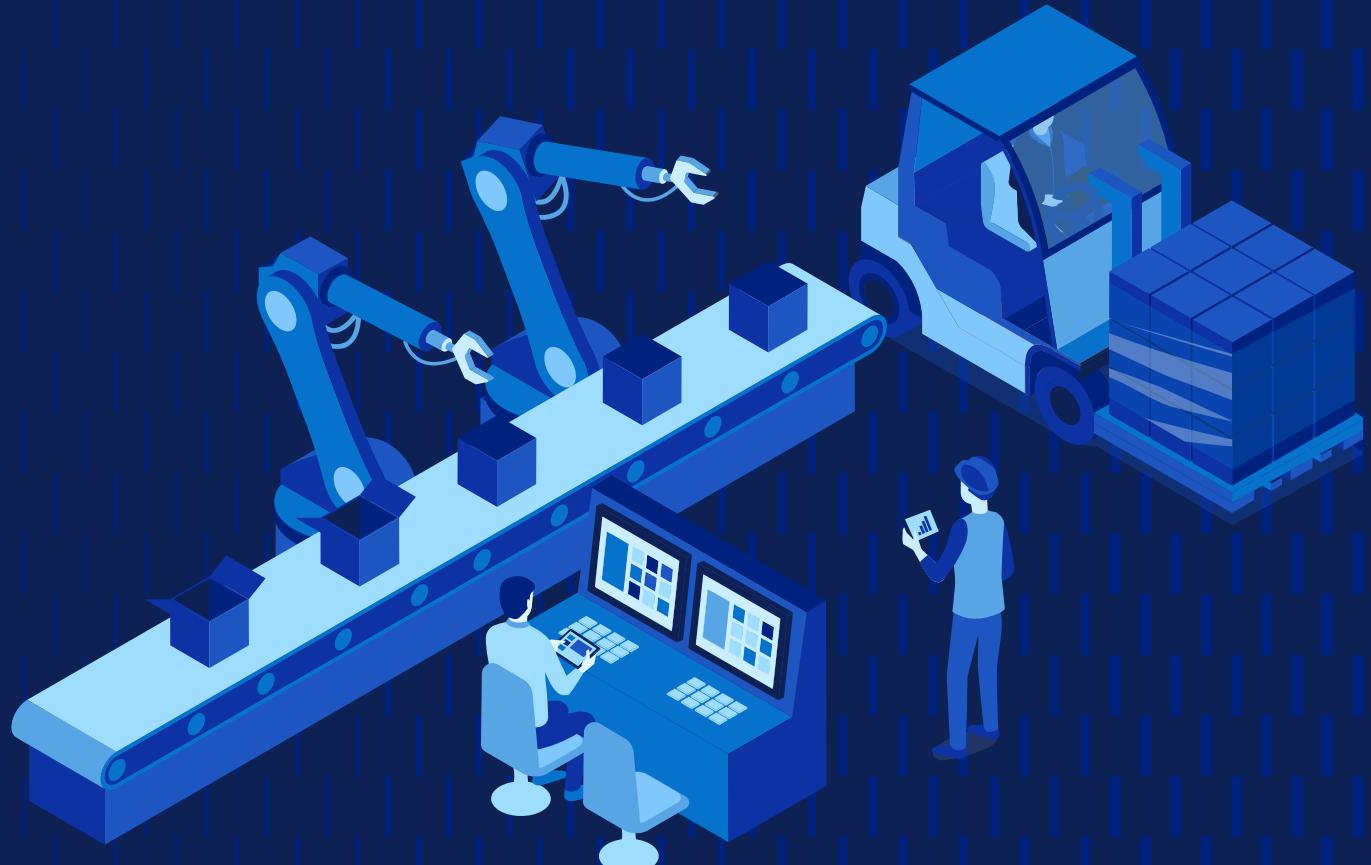


# 通向 AI 之路

Dell AI Factory with NVIDIA：利用尖端 AI 创新技术并以可扩展的方式实现 AI 工作流自动化，大幅提升业务增长速度



# Dell AI Factory with NVIDIA

## 加速采用 AI，使您可以充分挖掘数据的全部潜能

Dell AI Factory with NVIDIA 可加快 AI 采用速度，它集成了戴尔和 NVIDIA® 的诸多功能，可加快 AI 赋能应用场景的运行速度，集成数据和工作流，并让您能够自主设计这一旅程，实现可重复、可扩展的成果。



**更快实现价值**



**转化见解**



**提高生产力**



**交付值得信赖的成果**

一站式提供全面涵盖桌面、数据中心和云端的生成式 AI 解决方案组合。<sup>2</sup>

**75%**

推理的成本效益比公有云 IaaS 更高。<sup>3</sup>

**86%**

通过全包式全堆栈解决方案和服务，更快实现价值。<sup>4</sup>



的 IT 和企业领导者认为，生成式 AI 将为其组织带来巨大的转型价值。<sup>1</sup>

**Dell Technologies 和 NVIDIA 拥有超过 25 年的合作伙伴关系，双方致力于推动创新，提供前沿的平台、解决方案和软件，助力我们的共同客户实现突破性成果。**

<sup>1</sup> Dell Technologies Innovation Catalyst 研究，2024 年 2 月。

<sup>2</sup> 基于戴尔在 2023 年 8 月进行的分析。Dell Technologies 提供精心设计的解决方案，可在（移动和固定）工作站 PC 和服务器上运行 AI 工作负载提供支持，适用于高性能计算、数据存储、云原生软件定义的基础架构、网络交换机、数据保护、HCI 和服务等应用领域。

<sup>3</sup> 基于戴尔委托 Enterprise Strategy Group 进行的研究，比较了本地戴尔基础架构与原生公有云基础架构即服务，2024 年 4 月。分析的模型显示，对于拥有 5,000 名用户的组织，利用 RAG 的 70 亿参数 LLM 的成本效益提高了 38%；而对于拥有 50,000 名用户的组织，利用 RAG 的 700 亿参数 LLM 的成本效益提高了 75%。实际结果可能有所不同。

<sup>4</sup> 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。

# Dell AI Factory with NVIDIA

## 业界较早推出的端到端 NVMe/AI 解决方案

Dell AI Factory with NVIDIA 提供全面的 AI 技术组合、经过验证的全包式解决方案以及相关专家服务，可帮助您更快地实现 AI 成果。

借助戴尔种类丰富的 AI 基础架构和服务组合，结合 NVIDIA 业界卓越的加速计算能力以及包含 GPU、网络、NVIDIA AI Enterprise 软件、NVIDIA 推理微服务 (NIM)、模型和代理蓝图的完整堆栈，在您的企业中大规模应用 AI 和生成式 AI。

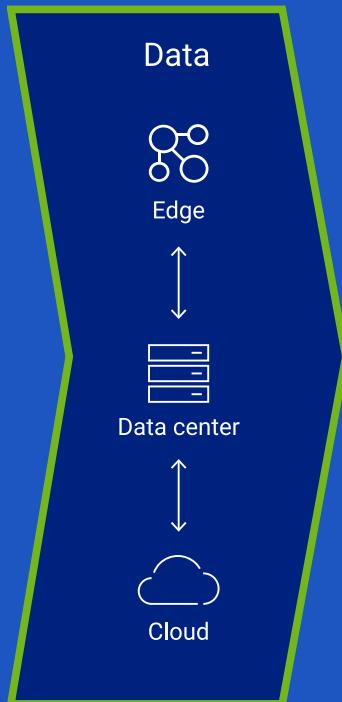


# 您的数据就是竞争优势

运用 AI，从更好的数据管理和准备开始。如果能利用强大、准确且全面的数据集加速业务决策，业界领先企业将从一众公司中脱颖而出。

## 数据为 AI 工厂注入动力

您最为重要的数据往往驻留在本地和边缘，即您拥有专属所有权的位置。这能更大限度地降低延迟，并确保敏感信息安全无忧。



## 在存储、保护和管理数据的先行者帮助下，将 AI 应用于您宝贵的数据。

您的数据管理之旅不是一蹴而就的任务，而是一个持续的过程，需要利用一系列实践和工具来帮助您组织持续地从数据中解锁价值。有效的数据管理对于任何成功的生成式 AI 计划都至关重要。

Dell Technologies 携手 NVIDIA，帮助像您这样的组织轻松完成数据管理之旅。在 AI 驱动的数据工作负载领域，这一旅程与目的地同样重要。

Dell AI Factory with NVIDIA 与 Dell Data Lakehouse 战略协同工作，通过戴尔服务和合作伙伴帮助管理和可视化数据，并为采用生成式 AI 做好准备。AI 就绪型 Dell PowerScale 进一步增强了数据存储能力，为 AI 工厂以及 NVIDIA DGX Pod、DGX Superpod 和 OVX 系统提供高性能的认证平台。

# 应用场景带来丰硕成果

通过经验证的解决方案和量身定制的服务，简化重要 AI 应用场景的部署。



## 内容和代码生成

- 编制品牌内容。
  - 增强内容个性化。
  - 更新市场报告。
  - 汇总行业新闻。
  - 监控竞争活动。
- 集成企业系统。
  - 自动执行测试流程。
  - 协助代码重构。
  - 增强安全协议。
  - 精简数据库查询。



## 数字助手

- 提供面向客户的支持。
- 利用培训和教育。
- 使用助手优化日程安排。
- 跟踪项目截止日期。
- 自动执行日常任务。



## 数据创建

- 增强商业智能。
- 精简项目研究。
- 协助法律咨询。
- 优化采购搜索。
- 改善财务分析。



## 计算机视觉

- 检测盗窃。
- 目视检查设备和产品。
- 自动化机械运行。
- 分析员工安全性。
- 监控路况。



## 数字孪生

- 客户体验。
- 产品开发和原型设计。
- 维护和运营。
- 物流和供应链。
- 培训和指导。



# 针对 AI 优化的基础架构

基础架构是 AI 工厂的基础。借助戴尔种类丰富的 AI 产品组合（涵盖从桌面到数据中心），灵活调整 AI 投资规模，并在任意位置运行工作负载。

## 确定应用场景

您期望的业务成果对基础架构有何要求？

## 确定您需要的 AI 模型/类型

了解了业务目标和应用场景后，您就可以确定如何适当平衡性能和成本。

## 合理调整投资规模

有些模型可以部署在笔记本电脑上，有些模型则需要大量 GPU，如 Dell PowerEdge XE9680 上的 8 个 GPU。这其实取决于您要完成的任务。这正是 AI 工厂框架的用武之地。

## 产品特色

**Dell NativeEdge** 是戴尔率先推出的边缘编排平台，可在边缘自动交付 NVIDIA AI Enterprise。<sup>5</sup>

**Dell PowerEdge XE9680L 服务器** 在紧凑的 4U 外形规格中配备了冷板式液冷和 8 个新一代 NVIDIA Blackwell Tensor Core GPU。

**Dell PowerScale** 是戴尔首款获得 NVIDIA DGX SuperPOD 认证的基于以太网的存储解决方案。

## 基础架构



笔记本电脑



工作站



计算



网络和数据保护



存储

### 基础架构

数据笔记本电脑  
和工作站

戴尔服务器

Dell storage

NVIDIA GPU

NVIDIA 网络产品

<sup>5</sup> 基于戴尔在 2024 年 5 月进行的分析。

# NVIDIA AI Enterprise — 适用于 AI 的操作系统

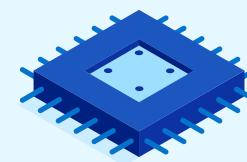
## 灵活稳健且适用于企业 AI 的端到端平台



云原生软件平台旨在加速数据科学管道，简化生产级 AI 应用程序的开发和部署。



易于使用的微服务可提供优化的模型性能和行业标准 API，以确保从原型到生产的平稳过渡。



安全的软件生命周期、端到端可管理性和企业级支持使组织能够自信地构建和运行 AI 解决方案。



# 加速推进 AI 采用 — 从桌面到数据中心

快速投入工作，并为您的团队提供本地原型环境的实践体验。在清晰了解数据管道和业务流程的基础上，开始培养组织技能，使您能够随着需求的增长无缝扩展。



## 易于使用的开发平台

在安全可靠的环境中执行本地原型设计



## 开发并准备扩展

使用基于容器的微服务快速构建 AI 和数据工作流。



## 在企业中部署

利用可扩展的基础架构框架进行协调和优化。



### AI 工工作站

Precision



### 计算

PowerEdge 和 NVIDIA GPU



### 存储

PowerScale/PowerStore



### 网络

NVIDIA Spectrum-X /  
NVIDIA Quantum InfiniBand

利用专为各种工作负载设计的广泛产品组合，随业务增长而扩展。

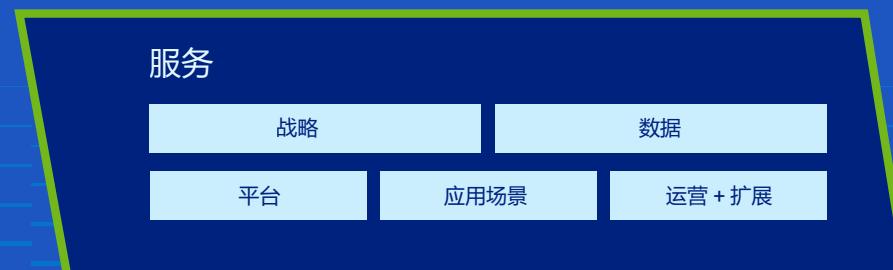
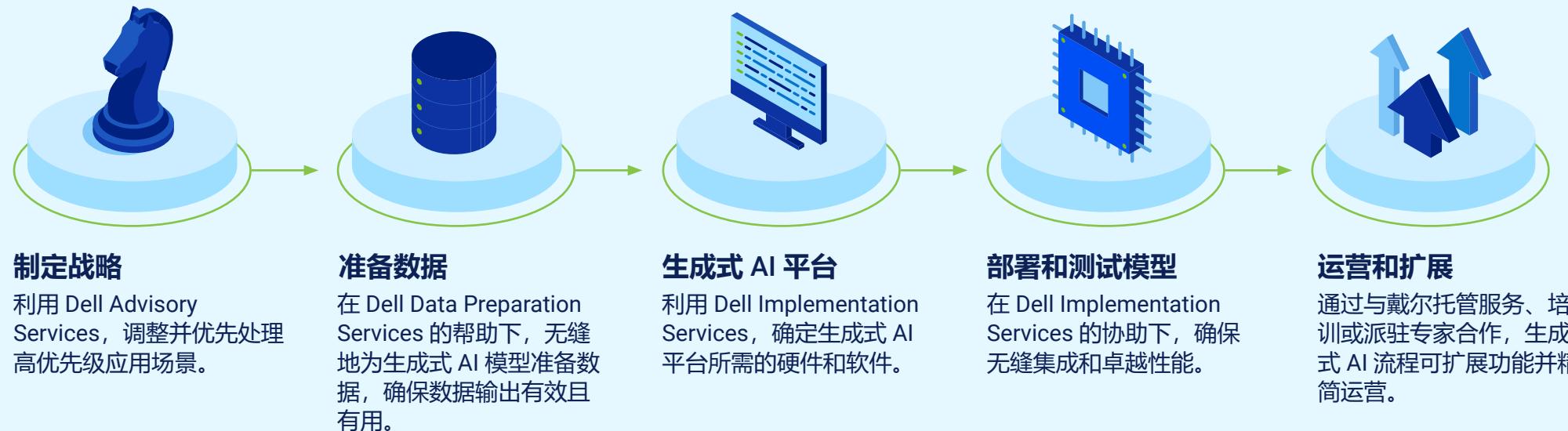
# 组合式、可扩展的基础架构和微服务 服务体系结构



在使用开放源代码构建您自己的生成式 AI 工具时，无需再凭空猜测。相反，利用这些不同方面的集成和连接，通过一个全面的解决方案大规模部署生成式 AI 应用程序，从而更轻松地获得成功。

# 有效采用生成式 AI 任重道远

Dell Technologies 在指导客户推进 AI 之旅方面拥有丰富的经验，能够帮助您将业务目标与适当的技术解决方案大规模结合，快速实现成果。



# 立即开始行动的两种方式

**利用 Dell AI Factory with NVIDIA，快速开始实现组织价值。**

## 战略规划

### 面向生成式 AI 的 Accelerator Workshop

- 参加为期半天的免费研讨会，开始制定制胜战略。
- 解决挑战、弥补差距，确定目标的优先顺序，发现机会。
- 此外，我们还提供了更全面的就绪性评估，以便更深入地了解基础架构要求、AI 模型、运营集成等。



## 技术储备

**利用即用型移动实验室助您快速开启成功之旅 — 包括戴尔 Precision 移动工作站和为期两天的咨询服务。**

- 便携地测试和演示生成式 AI
- 在预先验证的环境中进行快速原型设计。
- 经济高效、低风险地探索生成式 AI 应用场景。
- 弥合生成式 AI 技能差距。



### 搭载 NVIDIA GPU 的戴尔 Precision 移动工作站 5690/7780



**用于生成式 AI 实验的沙盒环境**  
已部署和配置 RAG 体系结构

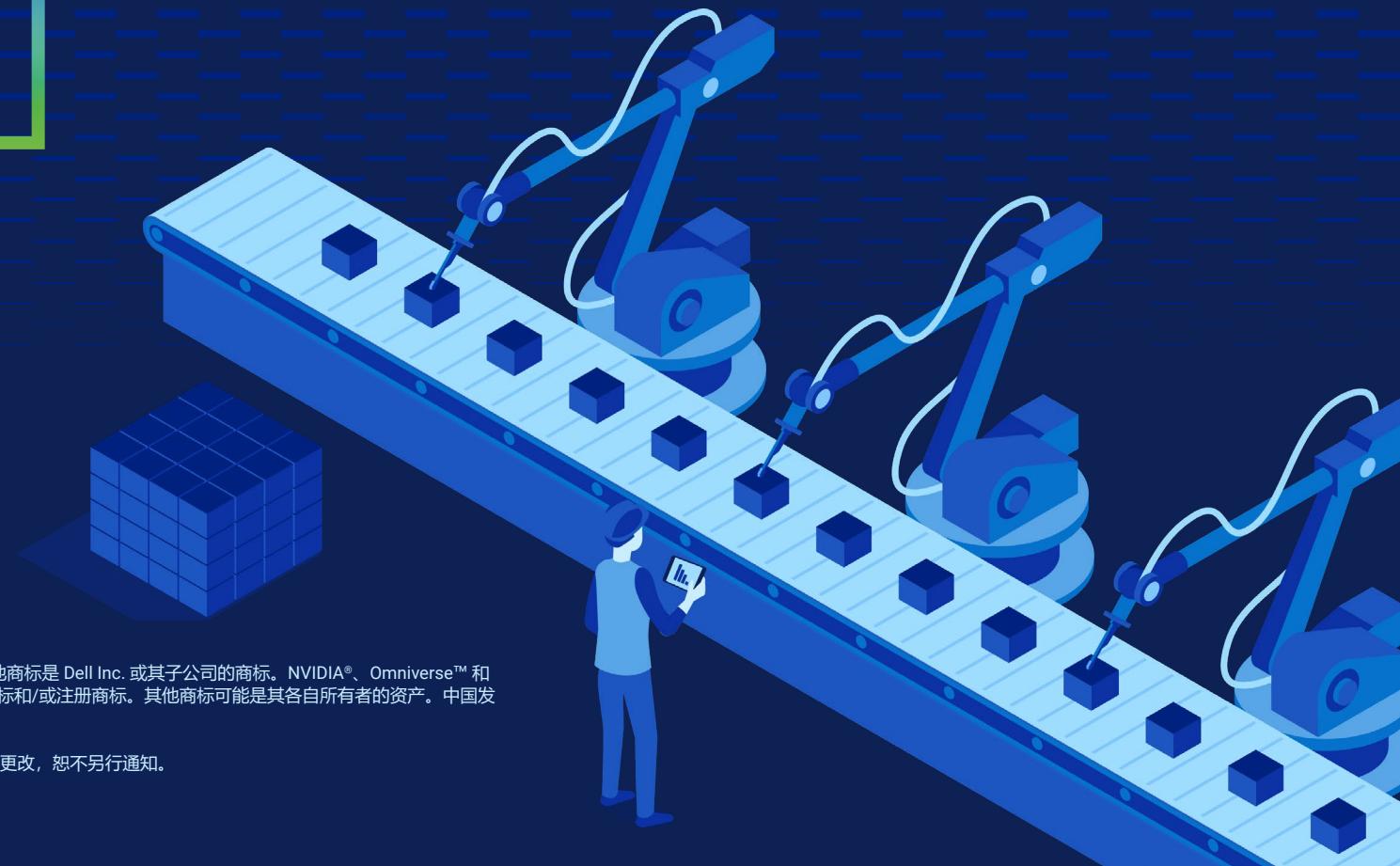
**面向开发人员的 NVIDIA AI Workbench 平台**  
已为运行软件而预先配置并经过性能验证

**支持进一步探索的基础应用场景**  
利用您的数据实施的初始聊天机器人应用场景

# 与 Dell Technologies 和 NVIDIA 合作， 充分释放生成式 AI 的潜力

访问我们的网站 [Dell.com/en-us/dt/nvidia-ai.htm](https://Dell.com/en-us/dt/nvidia-ai.htm)

了解更多信息并立即开始您的 AI 之旅。



版权所有 © 2024 Dell Inc. 或其子公司。保留所有权利。Dell 和其他商标是 Dell Inc. 或其子公司的商标。NVIDIA®、Omniverse™ 和 Spectrum™-X 是 NVIDIA Corporation 在美国和其他国家/地区的商标和/或注册商标。其他商标可能是其各自所有者的资产。中国发布，2024 年 9 月，电子书

Dell Technologies 确信本文档中的信息在发布之日是准确的。如有更改，恕不另行通知。