



借开放生态系统东风，利用 AI 的强大力量 更快获得洞察

采用 AMD Instinct™ MI300X 加速器和 ROCm™ 6
开放软件平台的戴尔 PowerEdge XE9680 服务器

ChatGPT 和其他生成式 AI (GenAI) 应用程序日益普及，促使人工智能 (AI) 技术快速走向普罗大众。越来越多的组织希望投身其中。最近的 IDC 研究表明：



77% 的组织

要么在探索 AI 应用场景，要么已经对 AI 计划进行了重大投资。¹

组织希望利用 AI 更快地获得洞察，并享受 AI 带来的其他好处，如果您尚未采取行动，务必尽快行动起来。时机已经成熟，尤其是考虑到开放生态系统高度致力于推动 AI 发展，新的解决方案已经准备好推动 AI 创新和您的业务向前发展。

¹IDC, 《未来企业弹性和支出调查第 6 轮》，2023 年 7 月。



“开放”就是未来发展之道

随着生成式 AI 和其他高级应用程序的兴起，芯片技术变得越来越强大，也越来越复杂。但在基础层面上，专有系统给一线开发人员制造了障碍。

这种局面即将改变。Dell Technologies 与 AMD 携手合作，利用基于开源技术构建的基础架构解决方案，帮您更快地利用这项技术。在戴尔 PowerEdge XE9680 服务器上运行 AMD Instinct™ MI300X 加速器和 AMD Radeon Open Compute (ROCm™) 开放软件平台，您可以获得强大的生态系统选项，乘“开放”理念的东风迈向未来，激发机遇，释放潜能。

Dell Technologies 和 AMD 助您获得性能和灵活性

Dell Technologies 和 AMD 打造能让开发人员为成功做好准备的系统，帮助全球企业更快取得成果。



戴尔 PowerEdge XE9680 服务器是专为 AI 任务而打造的数据处理主力机型。它支持八个加速器，非常适合机器学习 (ML)/深度学习 (DL) 训练和推理工作负载，特别适合需要训练大语言模型 (LLM) 的用户。

AMD Instinct MI300X 加速器采用第 3 代 AMD CDNA™(Compute DNA) 体系结构，提供 GPU 市场中出类拔萃的高带宽内存容量，每台服务器可提供 1.5 TB 的一致性 GPU 加速器显存，可训练十分庞大的 AI 模型。它还使您能使用更少的服务器来增加推理容量。

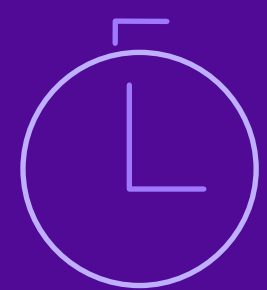
利用 ROCm 6 解锁加速力

ROCm 则进一步释放 AMD Instinct 的计算能力。ROCm 是一种开放、灵活、可扩展的加速计算软件平台，它配备驱动程序、API 和其他 GPU 工具，支持 AMD Instinct 加速器的代码优化，并简化了现有 GPU 加速代码的可移植性。

它也从根本上简化了您的体验。ROCm 在设备级别进行了优化，具有自由共享的代码库，该代码库可随着系统及其部件的发展而演变。这种开放的灵活性还支持无忧移植，所需编码接近于零。

无界生态系统

AMD Instinct MI300X 加速器和 ROCm 基于开放标准和互操作性而构建，多种 AI 框架和库为其提供原生支持。对于既想受益于 AMD 的强大实力，又想保持与现有软件系统的兼容性的用户，这提供了更丰富的选择和更高的灵活性。

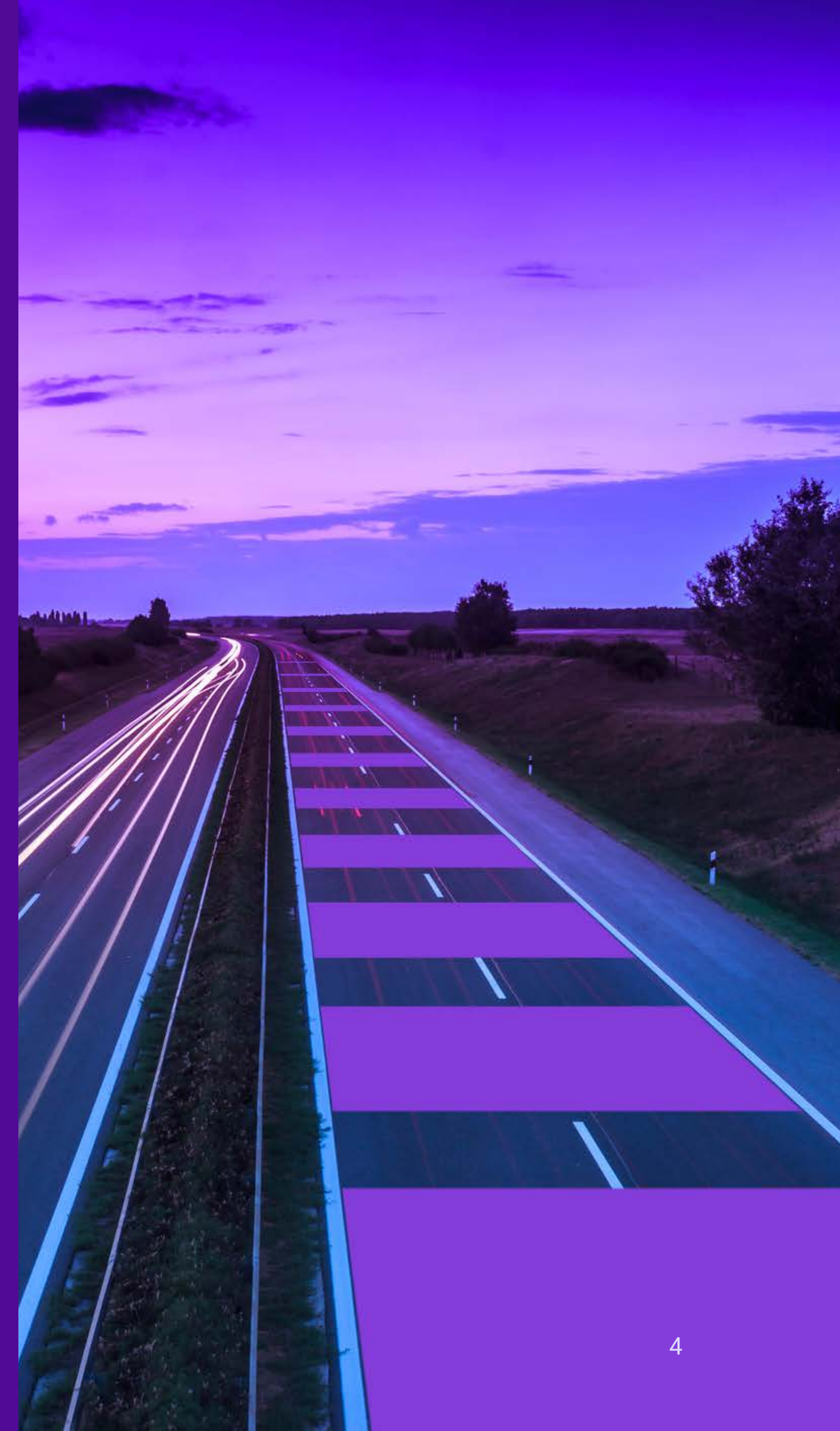


AMD 为您扫清前路，赋予您快速推进项目的自由度。

ROCm 的优势

- 解锁 GPU 加速计算任务的强大力量
- 针对大规模 ML/DL/生成式 AI 工作负载进行了优化
- 为主流型号和框架（包括 PyTorch®、TensorFlow™ 等）提供开箱即用的支持
- 支持快速采用创新（包括 JAX、Open Neural Network Xchange (ONNX)、OpenAI Triton 等框架中的原生支持）
- 与 OSS 开发社区建立联系，从而推动创新、树立差异化优势并开展协作
- 优越的灵活性为开放移植能力筑基

详细了解 ROCm 中的新增功能。



利用 ROCm 为 AMD Instinct 轻松移植和优化应用程序

- 在摩擦阻力较小的情况下将软件移植到 AMD 平台。HIP 允许开发人员通过单一源代码为 AMD Instinct MI300X 创建可移植的应用程序。
- 提供常用库和运算符的所有功能
- 快速移植并优化大多数应用程序，提升其性能

不断增加的上游支持

ROCm 为当今备受欢迎的框架提供上游支持，包括但不限于：

- PyTorch
- TensorFlow
- Open Neural Network Xchange (ONNX)
- JAX



利用开放移植能力节省宝贵的时间

有些开发人员可能对新生态系统及其对现有代码的影响心存顾虑，ROCm 可让其安心无忧。ROCm 平台的一个关键优势是，您可在尽可能减少现有代码返工量的前提下移植应用程序。移植轻而易举，可为您节省大量时间。

以四种不同的方式将现有应用程序移植到 AMD Instinct MI300X 加速器：

1 利用 ROCm 为其提供上游支持的现有 AI 框架。

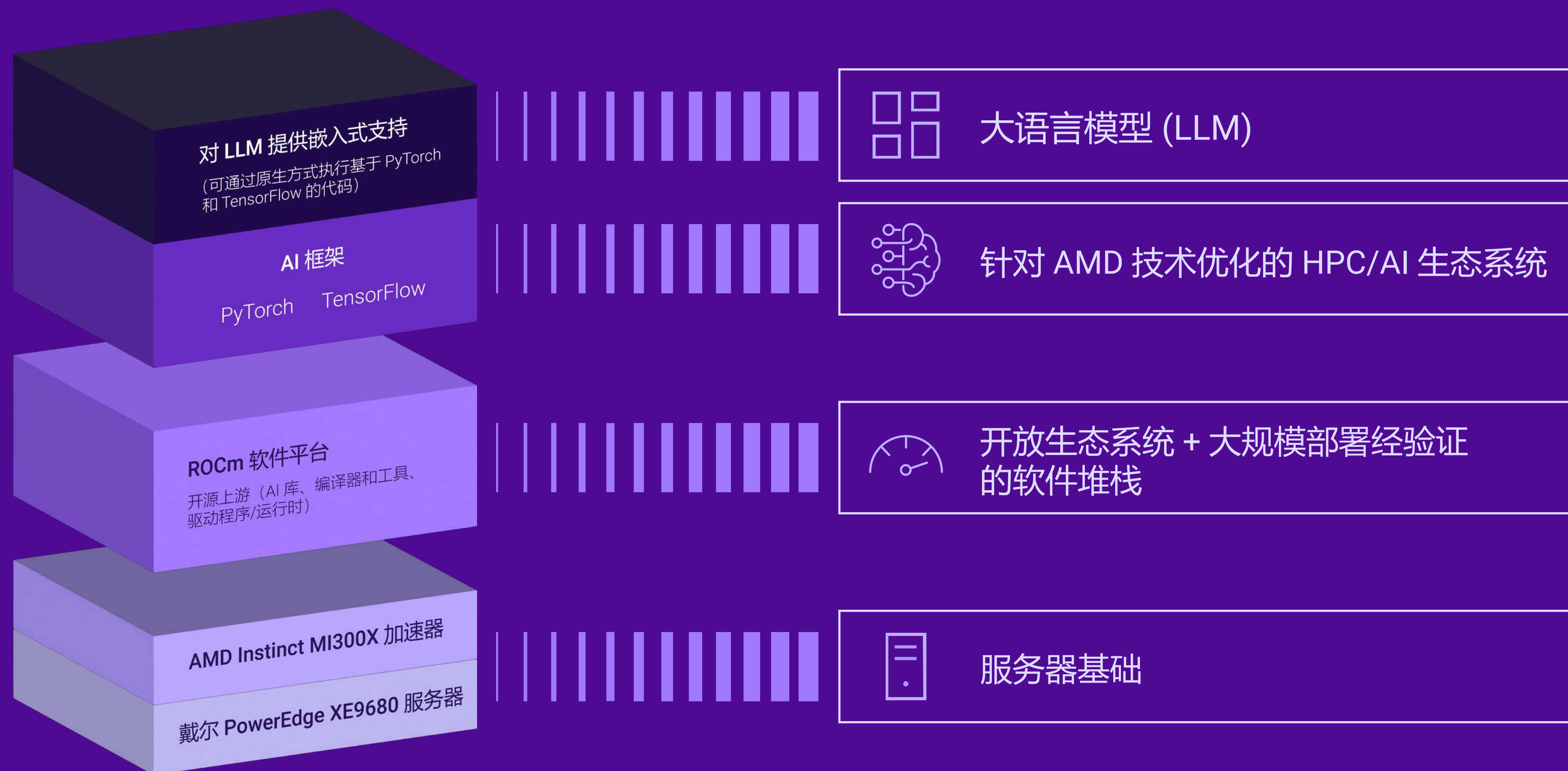
2 使用针对 AMD Instinct MI300X 优化的插入式 ML 内核。或是利用 AMD HIP 转换工具，在基于 ROCm 的 GPU 上优化 CUDA 应用程序。

3 使用为 ROCm 优化和维护的现有 CUDA 库软件存储库的镜像版本。

4 访问 AMD 预先打包的 AI 应用程序和容器。

立即行动

利用 Dell Technologies 和 AMD 提供的开放框架，为现有 AI 项目和新的生成式 AI 项目提速。戴尔 PowerEdge 支持新 AMD 加速器（包括 AMD Instinct MI300X 加速器），您可依托这一经过验证的基础为您的工作负载提供助力。敬请访问 [ROCm Developer Hub](#)。



AMD Instinct MI300X 加速器可直接部署

AMD Instinct MI300X 加速器能实现突破性能，可提供当今所需的速度、可扩展性和灵活的开放生态系统，尤其可以帮助企业克服资源、复杂性、延迟和体系结构方面的难题。AMD Instinct MI300X 加速器在性能方面具有显著优势：



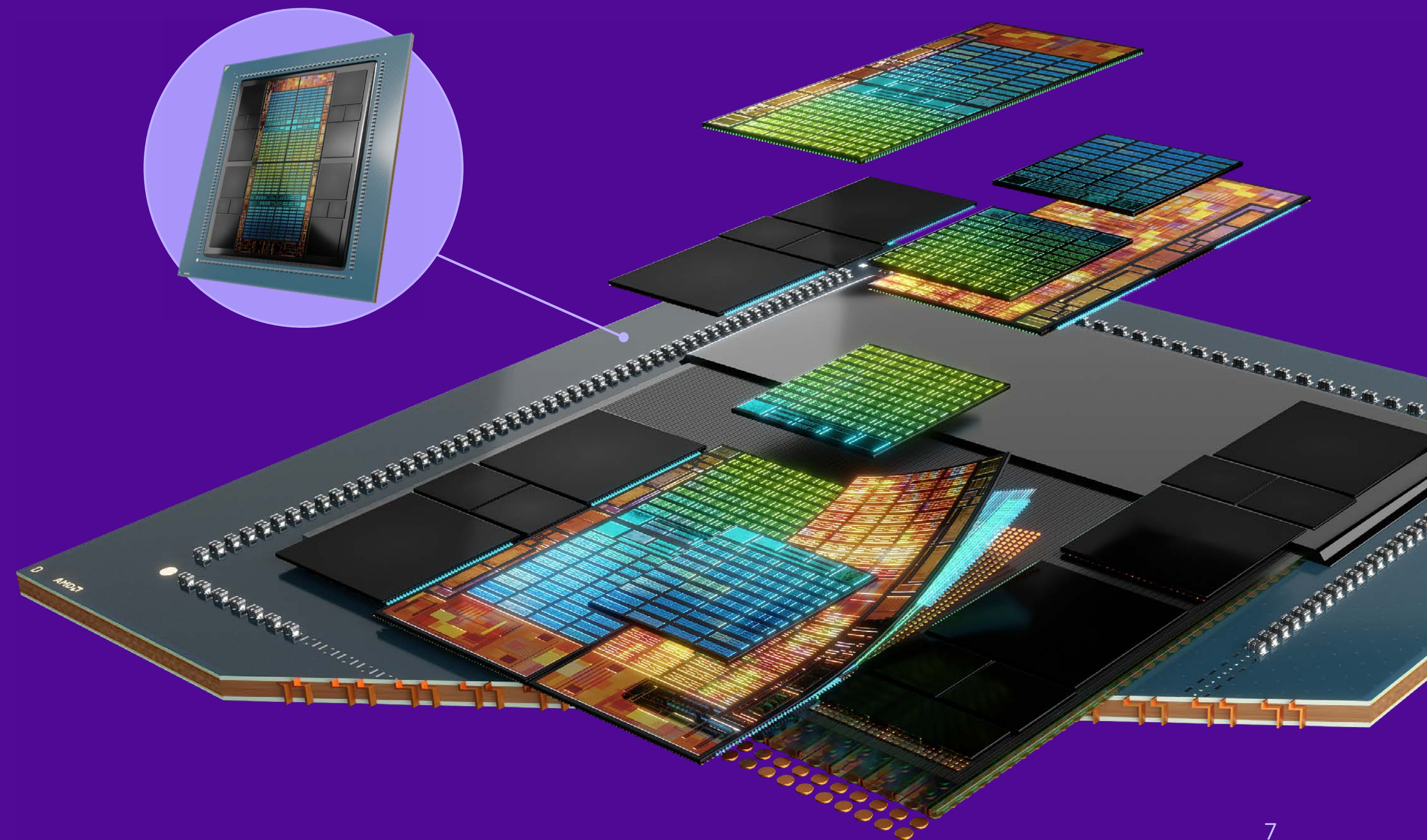
性能：利用每个 GPU 高达 192 GB 的堆叠式 HBM3 显存（显存带宽可达到每秒 5.3 TB），成就新型 AI 工作流。



密度：利用先进的芯片堆叠和芯片组技术减少移动数据的需要，为每台服务器提供更多的算力和内存容量。



开放：利用注重便捷访问和灵活适应的生态系统，轻松取得成果。



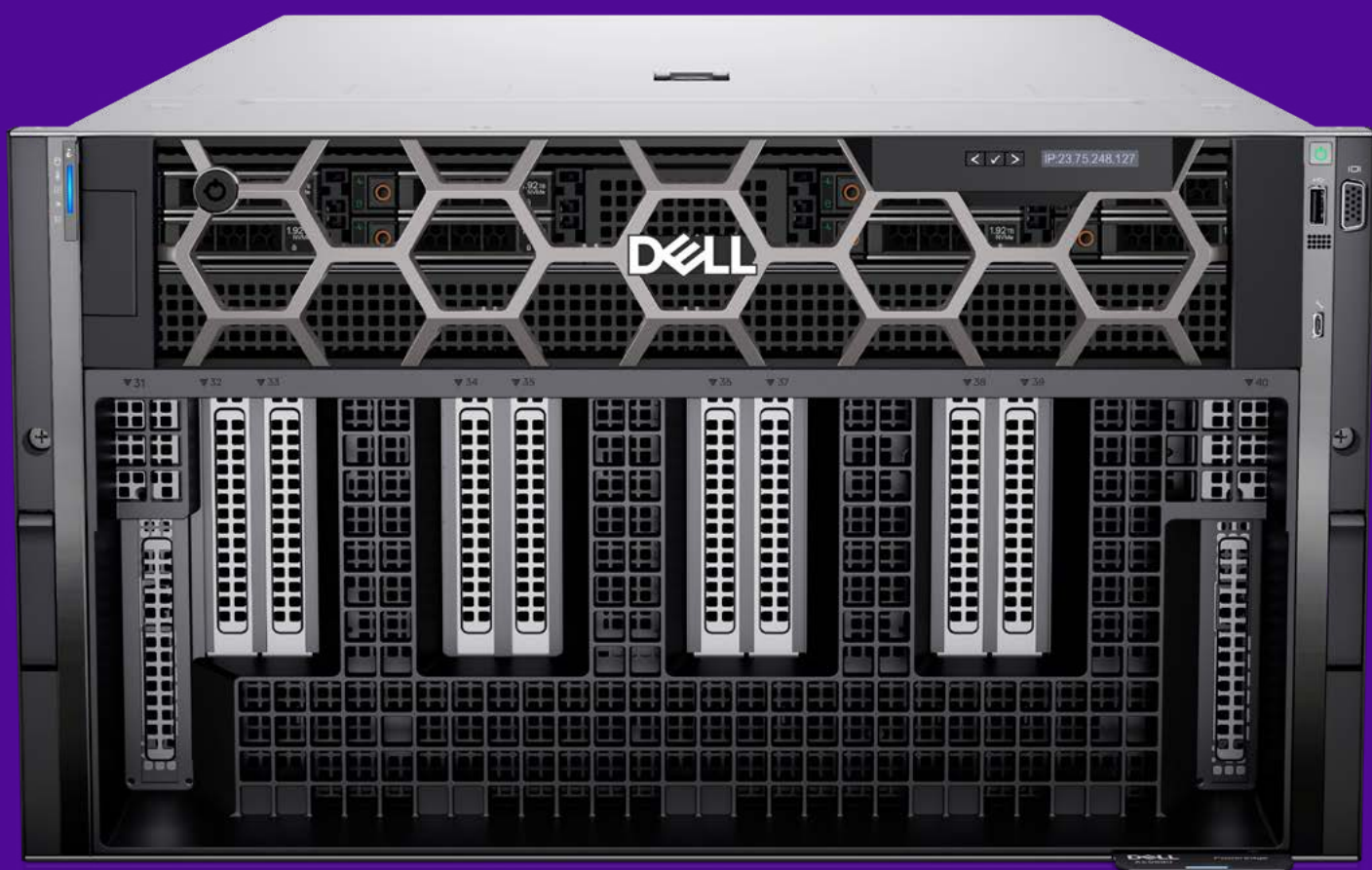
利用戴尔 PowerEdge XE9680 服务器充分发掘 AI 优势

借助专为 AI 设计的服务器平台，加速发展之路，更快速地取得更智能的成果。
借助可加速获得洞察、提供可信赖的 AI 并简化运营的解决方案，增强数据驱动型业务、交付可视化成果并加快整个企业的 AI 运营。



值得信赖的 AI

利用值得信赖、质量出众的解决方案基础架构，加速 AI 生命周期，从而降低风险并节省时间。PowerEdge XE9680 内置硅片信任根，可在产品的整个生命周期内为其保驾护航，直至资产报废。



简化运营

利用 AI 基础架构自动化，全方位有效控制和管理您的 AI 基础架构和性能工作负载。Dell Technologies 提供智能系统，这些系统可协同工作和独立工作，从而为您加速任务并简化体验。



加速形成洞察

帮助在整个 AI 生命周期内提高性能，并确保 AI 运营的速度跟上业务的步伐。Dell Technologies 提供创新计算技术，可助驾驭 AI 之力，实现出色性能。

助推创新

Dell Technologies 和 AMD 正在携手为我们宝贵的客户推动创新。我们共同通过一种切实可行的解决方案为市场带来芯片多样性，加强开发人员的能力。我们的联合解决方案让您能够：



- 受益于更多的选择和更少的阻碍，从而**加快进入市场的速度**。

- 利用不会阻碍您前进脚步的开放生态系统，**更快地获得洞察**。

- **选择适合您的产品**。依靠戴尔供应链的优势，不断推出创新技术，满足市场的迫切需要。

Dell Technologies 和 AMD 共同助您全速前行

AI 已走进现实，它必将加速发展，影响力也必定会不断扩大。对于希望充分利用其功能的企业而言，可能性无穷无尽。

正因如此，您不会希望自己的潜力受到技术障碍的牵制，不会希望您的发展因此受阻。您可以使用专门面向 AI 工作负载的强大解决方案堆栈，获得开放的选择机会。

基于戴尔 PowerEdge XE9680 服务器的 AMD Instinct MI300X 加速器和 ROCm 软件就是答案。加快新应用程序的开发过程、更快地进入市场，并开始推动对于您的业务和全世界意义非凡的成果。

如需了解详情，请访问 Dell.com/GPU

版权所有 © 2023 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。AMD、Instinct™、ROCm™ 及其组合是 Advanced Micro Devices, Inc. 的商标。PyTorch® 是 PyTorch 或 PyTorch 许可方的商标或注册商标。TensorFlow™ 和任何相关标识均为 Google Inc. 的商标。CUDA® 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财产。在中国发布，2023 年 12 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。