



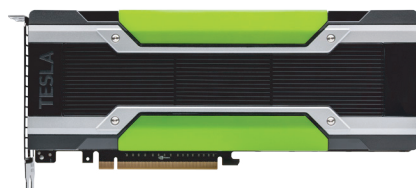
NVIDIA TESLA GPU 为数据中心量身打造

NVIDIA TESLA GPU 为数据中心量身打造

人们在深度学习领域中取得的突破成为了引爆人工智能革命的导火索。在科学界,计算性能正在向百万兆级别的方向迈进。但无论是人工智能革命还是百万兆级别计算,都需要通过 GPU 来提供强劲的计算性能。

NVIDIA GPU 是目前应用最广泛的高性能计算 (HPC) 处理器。无论用户是在 PC 平台还是云平台上进行计算,都可以找到一款 GPU 用来加速处理重要的工作。NVIDIA TITAN X 消费级 GPU 专为 PC 平台而设计,而 NVIDIA® Tesla® 数据中心 GPU 则为数据中心而打造,在可预测的行为以及正常运行时间等方面更能满足数据中心的要求。

TESLA GPU: 专为服务器平台而打造

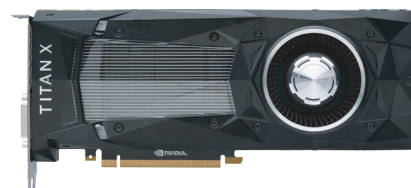


数据中心的特点是针对吞吐量而进行优化。想要实现优化并确保数据中心稳定的吞吐量,不但需要极高的处理性能,而且需要采用诸多技术来确保 7*24 小时运行以及出色的扩展性能和管理功能。



TITAN X 是从 PC 的角度来进行设计和测试的。它与服务器 GPU 在设计上的区别,在于只需要考虑密度、气流、进气温度、正常运行时间以及管理功能等几个设计要点即可。

TITAN X: 专为 PC 平台而打造



NVIDIA TITAN X 旨在打造全球性能最强的 PC 计算平台,值得每一位开发者拥有。它基于全新的 Pascal™ 架构,让用户在 PC 端也能充分利用 GPU 强大的计算处理性能。



图片来源: Facebook Big Sur 开放式机架型服务器

Tesla GPU 为一些全球规模最大的数据中心提供计算动力。Tesla GPU 专为服务器而打造,已通过各种认证,可在机械、物理、管理功能、可靠性以及可用性等方面满足机架级部署需求。Tesla GPU 保修期限更长、包含企业级服务支持、具备更长的产品使用周期,是完美的数据中心解决方案。

	优势	TESLA GPU	消费级 GPU
 7*24 小时运行	服务器质量更有保障	服务器级设计 <ul style="list-style-type: none"> > 工作电压更低，稳定性更长久 > 以极高的工作频率仍可通过零故障测试 > 纠错码 (ECC) 功能可确保数据完整性 	PC 级设计
	更低的热应力可带来无法媲美的稳定性	强制风冷式设计 <ul style="list-style-type: none"> > 可在服务器内实现最大的空气流通 > 更低的 GPU 温度能够确保可靠性 > 功耗更低 	主动式风扇设计 <ul style="list-style-type: none"> > 风扇的运行会妨碍服务器内空气流通 > GPU 工作温度升高 30-40%，因而增加了故障率 > 功耗更高
	数据中心效率更高、更容易维护	动态页面引退 (DPR) <ul style="list-style-type: none"> > 可监控不良内存，简单地重启即可将之清除 	不适用：需要手动拆卸 GPU 才能清除不良内存
 扩展性能	节点间的应用程序性能	GPU Direct RDMA <ul style="list-style-type: none"> > GPU 之间直接传输 > 延迟降低 67% > GPU 之间的 MPI 带宽提升 5 倍 	不适用
	单节点扩展性能增强	NVIDIA NVLink™ <ul style="list-style-type: none"> > GPU 之间的带宽提升 5 倍 > 多颗 GPU 可带来性能的直线提升 	不适用
	处理大型模型	<ul style="list-style-type: none"> > Tesla M40 最多配备 24 GB GDDR5 内存 > Tesla P100 最多配备 16 GB HBM2 内存 	4 GB - 12 GB
 数据中心支持	简化数据中心的运行	数据中心 GPU 管理器 <ul style="list-style-type: none"> > 设备管理 > 系统诊断 > 集群管理 > 已融入到领先的 HPC 管理工具中 	不适用
	产品使用周期更长	<ul style="list-style-type: none"> > 3 年产品使用周期 > 3 年保修 > 10 个月产品终止通知期 	产品使用周期短且不批量销售
	快速解决问题以保证持续运行	企业级支持 <ul style="list-style-type: none"> > 长期的支持与维修服务 > 直接获得技术专家的支持服务 > 现场咨询 	不适用