



NVIDIA® TESLA® P100 GPU加速器

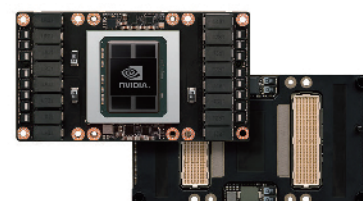
为现代数据中心提供无限的计算能力

更好地研发人工智能自动驾驶，更准确地预测气候变化，更快地研发出治疗癌症的新药，要应对这些严峻的挑战，我们需要进行大量的计算。但现有很多数据中心仍然依赖众多相互连接的计算节点，这在运行重要的高性能计算和超大规模运算时，却成为了限制。

NVIDIA Tesla P100 是迄今为止最先进的数据中心加速器，采用了突破性的 NVIDIA Pascal™ GPU 架构来打造全球最快的计算节点。该加速器融合了四项创新技术，在高性能计算和深度学习所需的性能方面实现了巨大的飞跃。

Tesla P100 还采用 NVIDIA NVLink™ 技术，为高性能计算和超大规模应用程序带来强大的扩展性能。每个节点可支持高达 8 个 Tesla P100 GPU 互联，其性能可媲美机架式标准 CPU 服务器。

TESLA P100 和 NVLINK 可为数据中心应用程序带来高达 50 倍的性能提升。

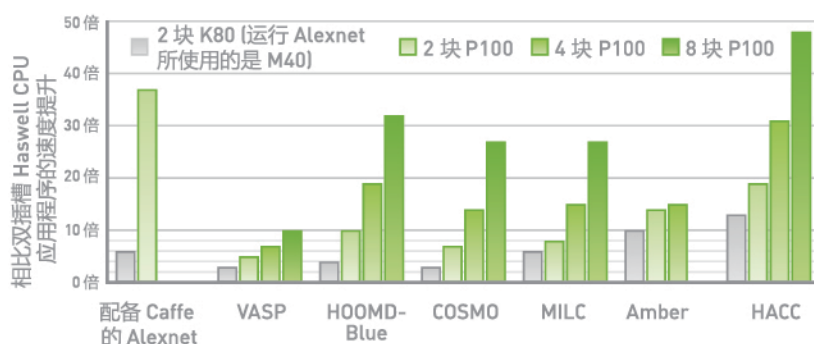


规格参数

GPU 架构	NVIDIA Pascal
NVIDIA CUDA® 核心数量	3584
双精度性能	5.3 TeraFLOPS
单精度性能	10.6 TeraFLOPS
半精度性能	21.2 TeraFLOPS
GPU 内存	16 GB CoWoS HBM2
内存带宽	720 GB/s
互联	NVIDIA NVLink
最大功耗	300 W
ECC	原生支持，无容量或性能占用
散热方案	被动式
造型规格	SXM2
计算 API	NVIDIA CUDA , DirectCompute, OpenCL™, OpenACC

TeraFLOPS 测量结果采用 NVIDIA GPU Boost™ 技术

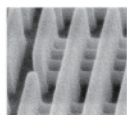
NVIDIA Tesla P100 性能



CPU: 16 核心, E5-2698v3@2.30 GHz. 256 GB 系统内存。Tesla K80 GPU: 2 块 K80 双 GPU 加速器, 预生产型 Tesla P100

性能实现全方位飞跃。

Tesla P100 从硬件到软件全部重新设计，在各个层面都融入了创新技术。每项突破性技术都实现了巨大的性能飞跃，助力打造全世界最快的计算节点。



Pascal™ 架构

超过 21 TeraFLOPS 的 FP16 半精度、10 TeraFLOPS 单精度和 5 TeraFLOPS 双精度的性能，可大幅提升在深度学习和高性能计算时的工作性能。



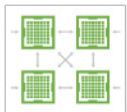
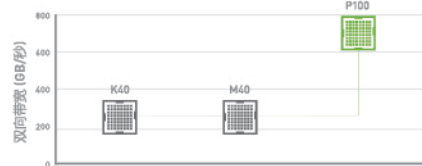
CoWoS HBM2

计算和数据采用 CoWoS (基板-晶圆-芯片封装) 通过 HBM2 技术封装在一起，在内存性能的提升上，可超过前代架构的 3 倍。

高性能计算和超大规模性能大幅提升

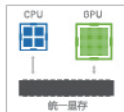


内存性能提升 3 倍



NVLink 互联

相比当前同级最佳的技术，这种高速双向互联技术，可将多 GPU 上应用程序的运行性能提升到 5 倍。



页迁移引擎

编程和计算性能调校更为简单，这意味着应用程序现在可以突破 GPU 的物理内存，几乎毫无限制。

互联性能提升到原先 5 倍



内存扩展能力大幅提升



如需了解更多 NVIDIA Tesla P100 的信息，敬请访问 www.nvidia.com/tesla

© 2016 NVIDIA 公司版权所有。保留所有权利。NVIDIA、NVIDIA 徽标、Tesla、NVIDIA GPU Boost、CUDA 以及 NVIDIA Pascal™ 均为 NVIDIA 公司在美国和其它国家的商标和/或注册商标。OpenCL 是苹果公司的商标，该商标的使用已经过 Khronos Group Inc. 的许可。所有其它商标和版权均为相应所有者的财产。2016 年 6 月

