



# NVIDIA TESLA V100 GPU 加速器

## 打造前所未有且最先進的資料中心 GPU

NVIDIA® Tesla® V100 是為加速 AI、HPC 及繪圖而建立，前所未有全球最先進的資料中心 GPU。Tesla V100 搭載最新的 GPU 架構 NVIDIA Volta™，在單一 GPU 中提供高達 100 個 CPU 的效能—使數據科學家、研究人員與工程師能夠因應先前無法解決的挑戰。

## 規格

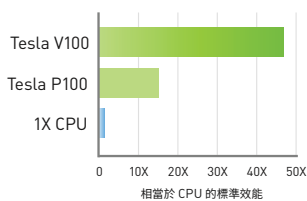


Tesla V100 PCIe

Tesla V100 SXM2

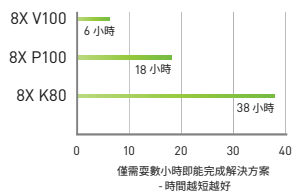
GPU 架構	NVIDIA Volta	
NVIDIA Tensor 核心	640	
NVIDIA CUDA® 核心	5,120	
雙精度浮點運算效能	7 TFLOPS	7.8 TFLOPS
單精度浮點運算效能	14 TFLOPS	15.7 TFLOPS
Tensor 效能	112 TFLOPS	125 TFLOPS
GPU 記憶體	16 GB HBM2	
記憶體頻寬	900 GB/秒	
ECC	支援	
互連頻寬	32 GB/秒	300 GB/秒
系統介面	PCIe Gen3	NVIDIA NVLink
尺寸	PCIe 全高度/長度	SXM2
最大功耗	250 W	300 W
散熱解決方案	被動式	
運算 APIs	NVIDIA CUDA、DirectCompute、OpenCL™、OpenACC	

在深度學習推論方面的輸送量較 CPU 伺服器高 47 倍



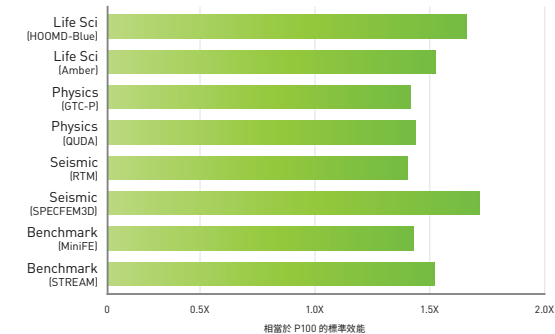
工作負載：ResNet-50 | CPU：2X Xeon E5-2690v4 @ 2.6GHz | GPU：新增 1X NVIDIA® Tesla® P100 或 V100 @ 150W | V100 是採用預先製造的硬體進行測量。

一個工作天即能完成深度學習訓練



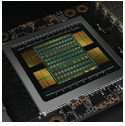
伺服器設定：Dual Xeon E5-2699 v4, 2.6GHz | 8x Tesla K80、Tesla P100 或 Tesla V100 | V100 效能是採用預先製造的硬體測量 | ResNet-50 是採用供 90 Epochs 使用的 Microsoft Cognitive Toolkit 進行訓練，並搭載 1.28M ImageNet 資料庫。

使用 NVIDIA Tesla V100，一年的 HPC 效能高達 P100 的 1.5 倍



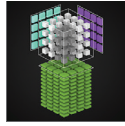
CPU 系統：2X Xeon E5-2690v4 @ 2.6GHz | GPU 系統：NVIDIA® Tesla® P100 或 V100 | V100 是採用預先製造的硬體進行測量。

## 突破性創新



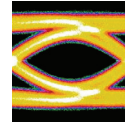
### VOLTA 架構

配備 Tesla V100 GPU 的單一伺服器，可透過在整合架構內將 CUDA 核心與 Tensor 核心配對後，取代上百部供傳統 HPC 和深度學習使用的市售 CPU 伺服器。



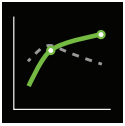
### TENSOR 核心

Tesla V100 配備了 640 個 Tensor 核心，可展現 120 TeraFLOPS 的深度學習效能。相較於 NVIDIA Pascal™ GPU，DL 訓練可達 12 倍 Tensor FLOPS，DL 推斷能力則可達 6 倍 Tensor FLOPS。



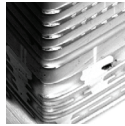
### 新一代 NVLINK

相較於舊世代，Tesla V100 的 NVIDIA NVLink 能展現高達 2 倍的輸送量。具備高達 300 GB/s 的速度，可與最高八部 Tesla V100 加速器互連，在單一伺服器上展現最高的應用效能。



### 最大效率模式

資料中心可利用全新的最大效率模式，在現有的功率分配中，達到每機架提高 40% 的運算容量。在此模式中，Tesla V100 會以尖峰處理效率運行，以減半功耗提供高達 80% 的效能。



### HBM2

Tesla V100 可在結合改良 900 GB/s 的原始頻寬和高達 95% 的 DRAM 使用效率後，如同於 STREAM 上測量般，展現高達 1.5 倍的 Pascal GPU 記憶體頻寬。



### 編程能力

Tesla V100 從初始即採用精簡編程能力的架構。全新獨立執行緒排程是藉由分享小型工作的資源，達到細粒度同步及改善 GPU 使用率。

Tesla V100 是 Tesla 資料中心運算平台進行深度學習、HPC 與繪圖的旗艦產品。Tesla 平台能為超過 450 種 HPC 應用項目及各種主要深度學習架構加速。其使用範圍從桌上型電腦到伺服器，再至雲端服務，能提供大幅提升效能與節省成本的機會。

### 各種深度學習架構



Caffe2



Microsoft

mxnet

PYTORCH

TensorFlow

theano

### 450+ GPU 加速應用



AMBER



ANSYS Fluent



GAUSSIAN



GROMACS



LS-DYNA



NAMD



OpenFOAM



Simulia Abaqus



VASP



WRF

想要進一步瞭解 Tesla V100，請造訪 [www.nvidia.com/v100](http://www.nvidia.com/v100)

© 2017 NVIDIA Corporation。保留所有權利。NVIDIA、NVIDIA 標誌、Tesla、NVIDIA GPU Boost、CUDA 及 NVIDIA Volta 皆是 NVIDIA 公司在美國及其他國家的商標及註冊商標。OpenCL 是 Apple Inc. 的商標，並授權給 Khronos Group Inc. 使用。所有其他商標及版權皆為個別擁有者所有之財產。7 月 17 日

