

# TESLA P100 效能指南

HPC 及深度學習應用



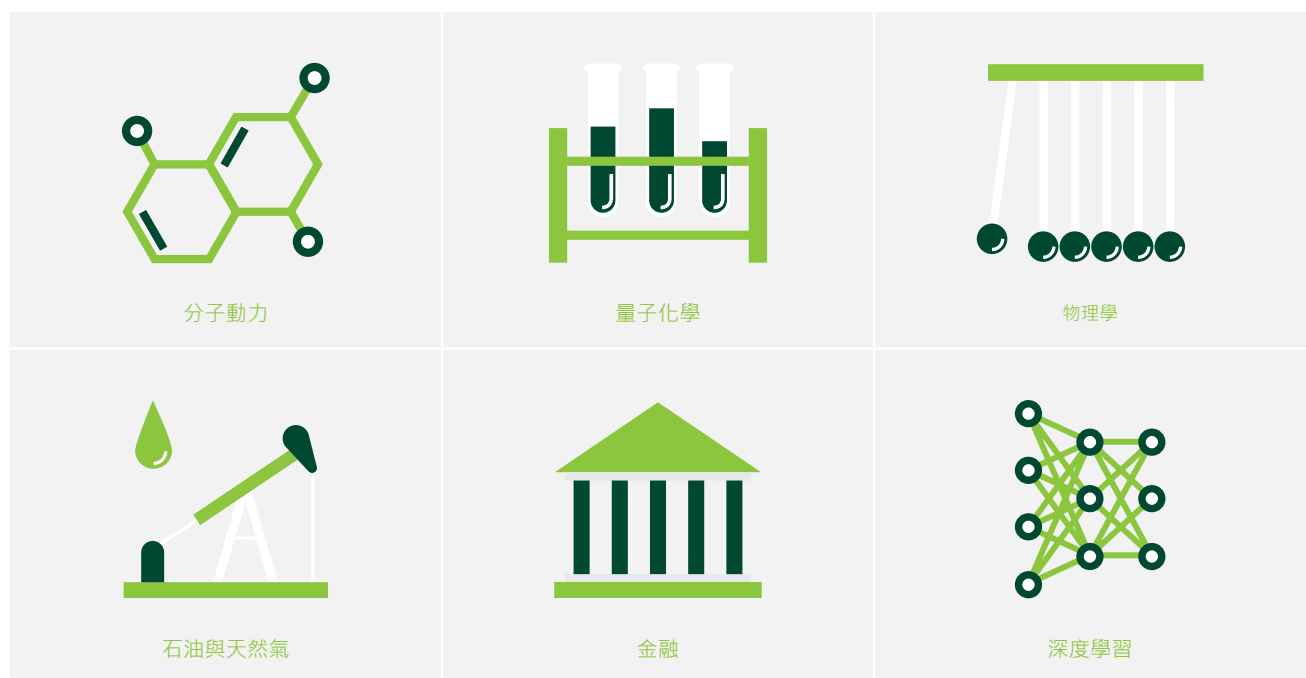
**nVIDIA.**

## TESLA P100 效能指南

現代的高效運算 ( HPC ) 資料中心是解決部分全球最重要之科學與工程挑戰的關鍵。 NVIDIA® Tesla® 加速運算平台利用領先業界的應用程式支援這些現代化資料中心，促進 HPC 與 AI 工作負載。 Tesla P100 GPU 是現代資料中心的引擎，能以更少的伺服器展現突破性效能，進而實現更快的解析能力，並大幅降低成本。

每一個 HPC 資料中心都能自 Tesla 平台獲益。在廣泛的領域中有超過 400 個 HPC 應用程式，採用 GPU 最佳化，包括所有前 10 大 HPC 應用程式和各種主要深度學習架構。

採用加速 GPU 應用程式的研究領域包括：



超過 400 個 HPC 應用及所有深度學習架構皆是採用加速 GPU。

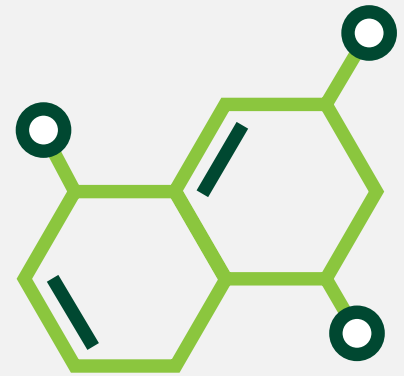
> 若想要取得最新 GPU 加速應用目錄，請造訪：

**[www.nvidia.com/teslaapps](http://www.nvidia.com/teslaapps)**

> 若想要立即在 GPU 上使用簡易指示，快速執行廣泛的加速應用，請造訪：

**[www.nvidia.com/gpu-ready-apps](http://www.nvidia.com/gpu-ready-apps)**

# 分子動力



分子動力 ( MD ) 代表 HPC 資料中心的大部分工作負載。100% 頂尖 MD 應用皆是採用 GPU 加速，以使科學家能進行從前僅有 CPU 版本之傳統應用項目無法執行的模擬工作。在執行 MD 應用時，配備 Tesla P100 GPU 的資料中心可節省高達 60% 的伺服器取得成本。

## TESLA 平台及適用 MD 的 P100 的關鍵功能

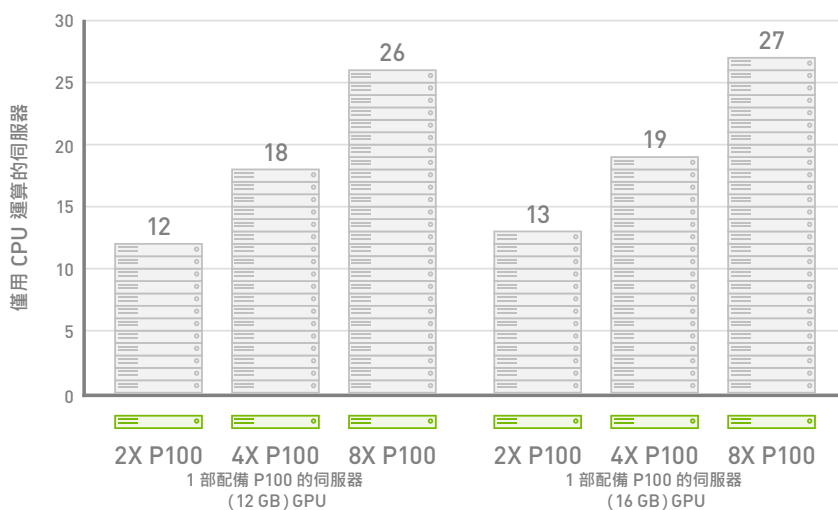
- > 搭載 P100 的伺服器，最多可取代 40 部適用 HOOMD-Blue、LAMMPS、AMBER、GROMACS 和 NAMD 等應用的 CPU 伺服器
- > 100% 頂尖 MD 應用項目皆採用加速 GPU
- > FFT 和 BLAS 等關鍵數學程式庫
- > 每一個 GPU 之單精度效能高達每秒 11 TFLOPS
- > 每一個 GPU 之記憶體頻寬高達每秒 732 GB

檢視所有相關的應用項目：

[www.nvidia.com/molecular-dynamics-apps](http://www.nvidia.com/molecular-dynamics-apps)

## HOO MD-Blue 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: Microsphere | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點 · 以達到 CPU 節點等效。

### HOO MD-BLUE

循序寫入 GPU 的粒子動力封裝

版本  
1.3.3

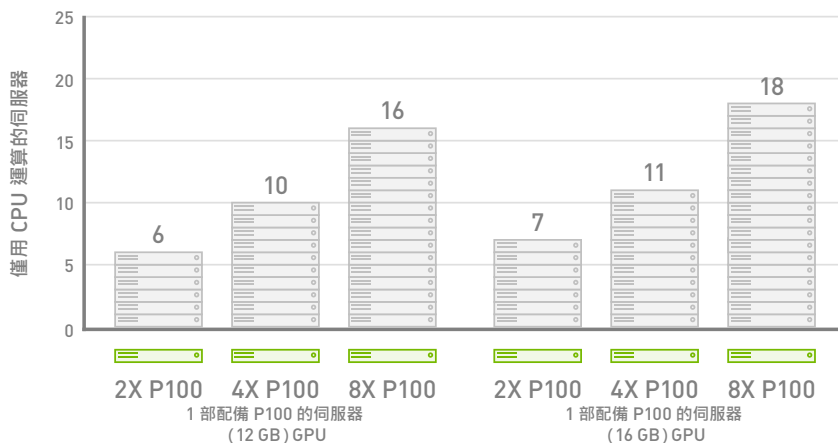
加速功能  
CPU 和 GPU 可用版本

延展性  
多 GPU 和多節點

更多資訊  
[www.nvidia.com/hoomd-blue](http://www.nvidia.com/hoomd-blue)

## LAMMPS 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: EAM | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點 · 以達到 CPU 節點等效。

### LAMMPS

典型粒子動力封裝

版本  
2016

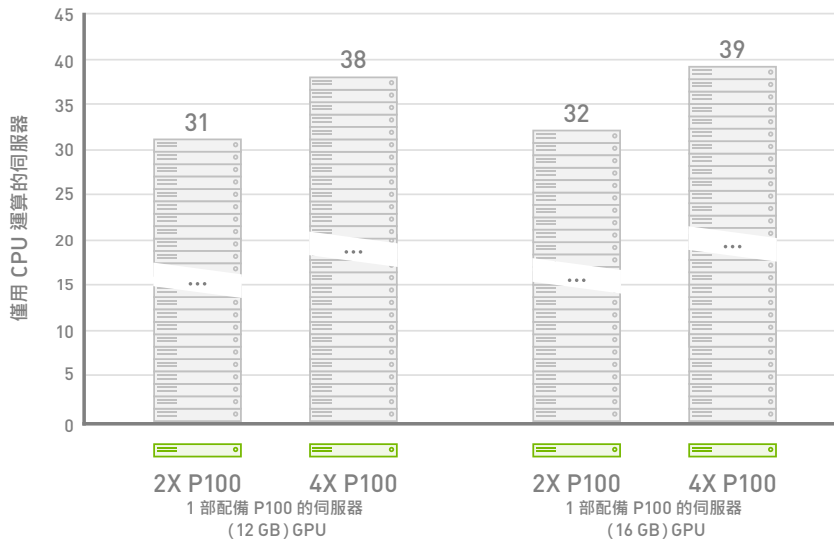
加速功能  
Lennard-Jones、Gay-Berne、Tersoff  
更多勢能

延展性  
多 GPU 和多節點

更多資訊  
[www.nvidia.com/lammps](http://www.nvidia.com/lammps)

## AMBER 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: GB-Myoglobin | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

黃色

在生物分子上模擬分子動力的程式套件

版本

16.3

加速功能

PMEMD 顯性溶劑和 GB、顯性及隱性溶劑、REMD、aMD

延展性

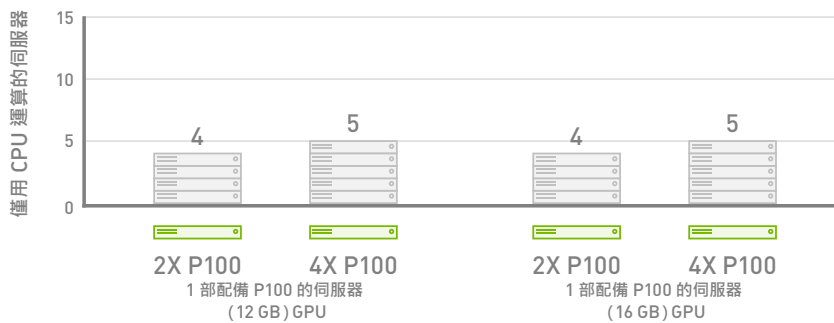
多 GPU 和多節點

更多資訊

[www.nvidia.com/amber](http://www.nvidia.com/amber)

## GROMACS 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: Water 3M | 我們使用高達 8 CPU 節點測量基準點, 以達到 CPU 節點等效。

GROMACS

模擬含複雜連結互動的生物模型分子

版本

5.1.2

加速功能

PME、顯性與隱性溶劑

延展性

多 GPU 和多節點

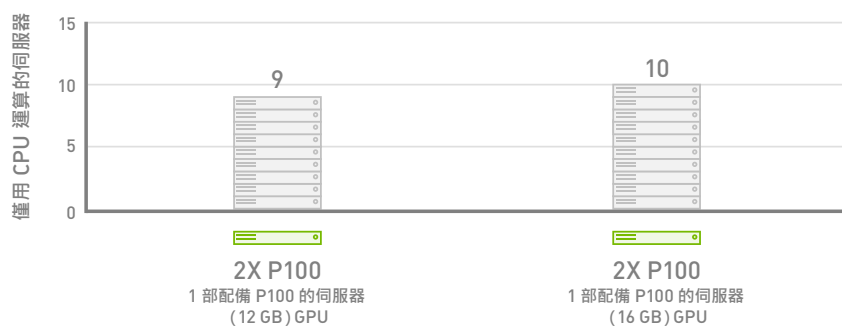
擴展至 4xP100

更多資訊

[www.nvidia.com/gromacs](http://www.nvidia.com/gromacs)

## NAMD 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: STVM | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### NAMD

專為高效模擬大分子系統而設計

版本

2.11

加速功能

PME 全靜電和眾多模擬功能

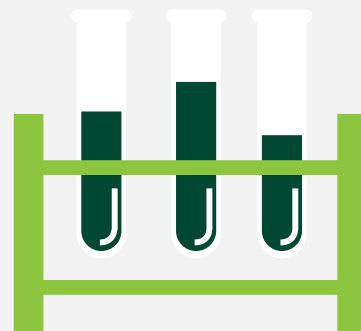
延展性

高達 100M 原子 · 多 GPU ·  
擴展為 2xP100

更多資訊

[www.nvidia.com/namd](http://www.nvidia.com/namd)

## 量子化學



量子化學 (QC) 模擬是探索新藥物與原料的關鍵，且會耗費大部分 HPC 資料中心的工作負載。目前，60% 的頂尖 QC 應用都採用加速 GPU。在執行 QC 應用時，配備 Tesla P100 GPU 的資料中心工作負載可節省高達 40% 的伺服器成本。

### TESLA 平台及適用 QC 的 P100 的關鍵功能

- > 搭載 P100 的伺服器，最多可取代 36 部適用 VASP 和 LSMS 等應用的 CPU 伺服器
- > 60% 的頂尖 QC 應用項目皆採用加速 GPU
- > FFT 和 BLAS 等關鍵數學程式庫
- > 每一個 GPU 之雙精度效能高達每秒 5.3 TFLOPS
- > 大資料集記憶體容量高達 16 GB

檢視所有相關的應用項目：

[www.nvidia.com/quantum-chemistry-apps](http://www.nvidia.com/quantum-chemistry-apps)

## VASP 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: B\_hr105 | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### VASP

從頭開始執行量子機制分子動力 ( MD ) 模擬的封裝

### 版本

5.4.1

### 加速功能

RMM-DIIS、Blocked Davidson、K-points 和正確交換

### 延展性

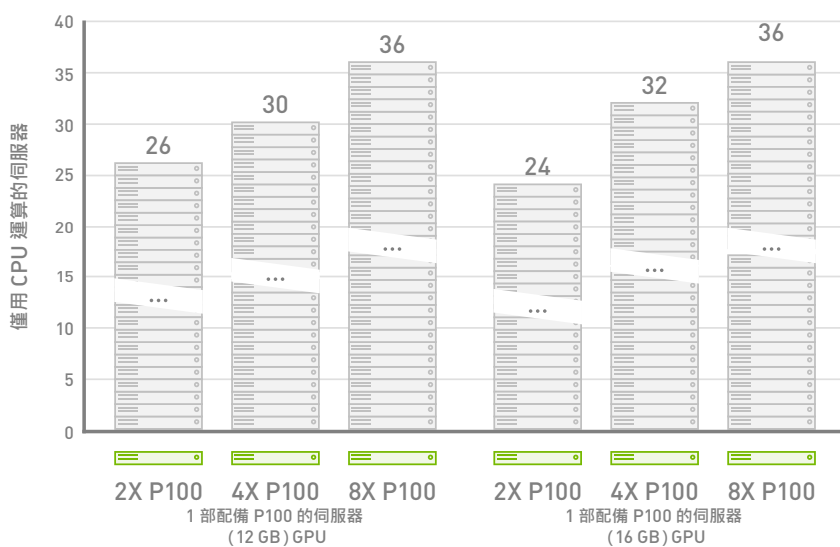
多 GPU 和多節點

### 更多資訊

[www.nvidia.com/vasp](http://www.nvidia.com/vasp)

## LSMS 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: Fe16 | 為達到 CPU 節點等效, 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮來測量基準點。

### LSMS

研究磁性溫度作用的材料代碼

### 版本

3

### 加速功能

廣義的 Wang-Landau 算法

### 延展性

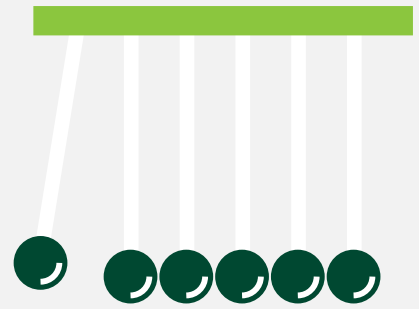
多 GPU

### 更多資訊

[www.nvidia.com/lsms](http://www.nvidia.com/lsms)



# 物理學



從聚變能量到高能粒子，HPC 資料中心的物理模擬涵蓋極廣泛的應用。多數頂尖物理應用皆是採用加速 GPU，解析從前無法實現的項目。在執行物理應用時，配備 Tesla P100 GPU 的資料中心可節省高達 70% 的伺服器取得成本。

## TESLA 平台及適用物理學的 P100 的關鍵功能

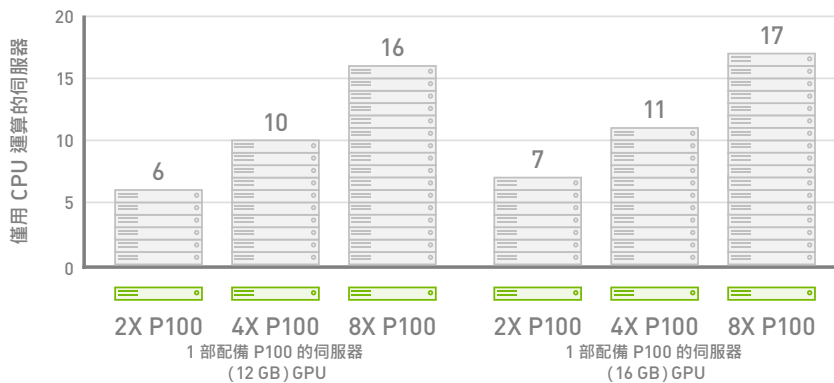
- > 搭載 P100 的伺服器，最多可取代 50 部適用 GTC-P、QUDA、MILC 和 Chroma 等應用的 CPU 伺服器
- > 絕大多數的頂尖物理學應用項目皆採用加速 GPU
- > 雙精度浮點效能高達 5.3 TFLOPS
- > 記憶體容量高達 16 GB 且記憶體頻寬高達 732 GB/s

檢視所有相關的應用項目：

[www.nvidia.com/physics-apps](http://www.nvidia.com/physics-apps)

## GTC-P 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.44 | 資料集: gtc. 輸入 | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### GTC-P

最佳化電漿物理的開發代碼

版本

2016

加速功能

推動、移動和碰撞

延展性

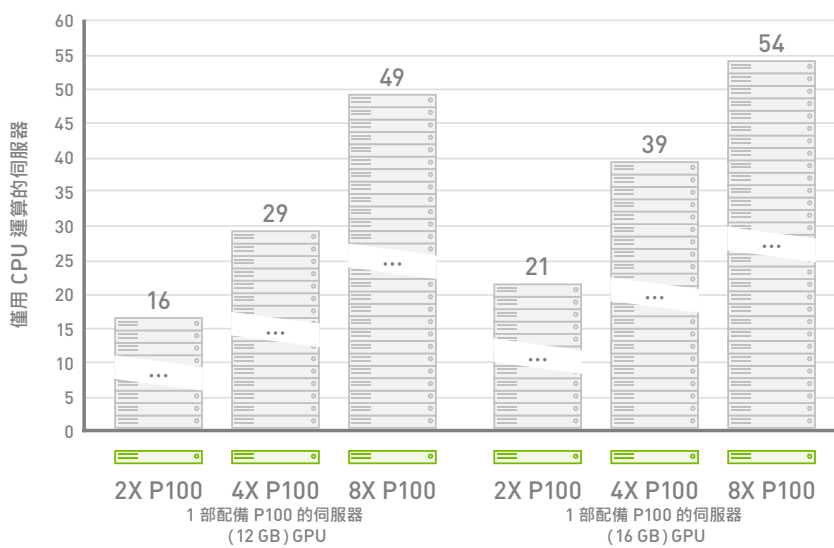
多 GPU

更多資訊

[www.nvidia.com/gtc-p](http://www.nvidia.com/gtc-p)

## QUDA 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.42 | 資料集: Glove Precision Single · Gauge Compression/ Recon: 12 · 問題規模 32x32x32x64 | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### QUDA

GPU 格點量子色動力學程式庫

版本

2017

加速功能

全部

延展性

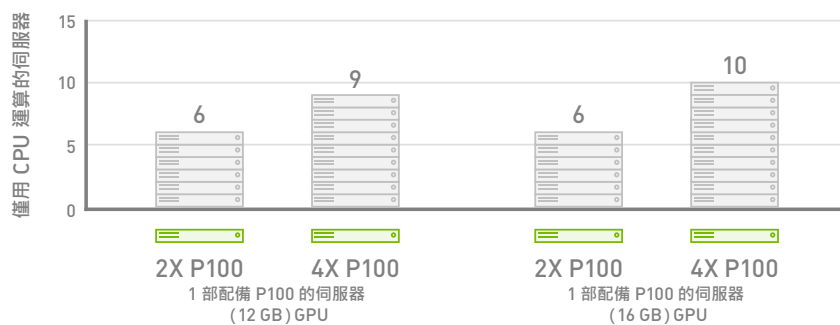
多 GPU 和多節點

更多資訊

[www.nvidia.com/quda](http://www.nvidia.com/quda)

## MILC 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.42 | 資料集: 雙倍精度 | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### MILC

格點量子色動力學 (LQCD) 代碼模擬元素微粒之形成方式, 以及藉由「強作用力」進行結合, 創建出質子和中子等更大微粒

版本  
7.8.0

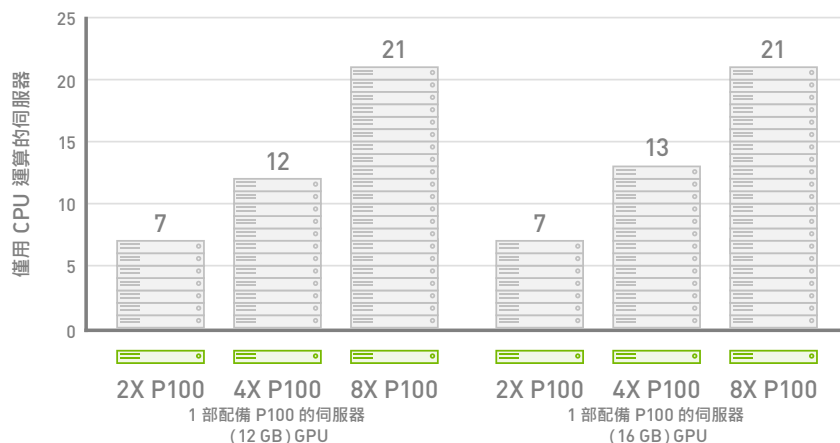
加速功能  
特性交錯費米子、Krylov 計算器和計量 - 鏈結厚化  
縮放為 4xP100

延展性  
多 GPU 和多節點

更多資訊  
[www.nvidia.com/milc](http://www.nvidia.com/milc)

## Chroma 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.42 | 資料集: szscl21\_24\_128 (總時間秒) | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### CHROMA

格點量子色動力學 (LQCD)

版本  
2016

加速功能  
Wilson-clover 費米子、Krylov 計算器和區域分解

延展性  
多 GPU

更多資訊  
[www.nvidia.com/chroma](http://www.nvidia.com/chroma)

# 石油與天然氣



地球科學模擬是探勘石油與天然氣和執行地質建模的關鍵。目前，多數頂尖的地球科學應用都採用加速 GPU。在執行地球科學應用時，配備 Tesla P100 GPU 的資料中心可節省高達 65% 的伺服器成本。

## TESLA 平台及適用地球科學的 P100 的關鍵功能

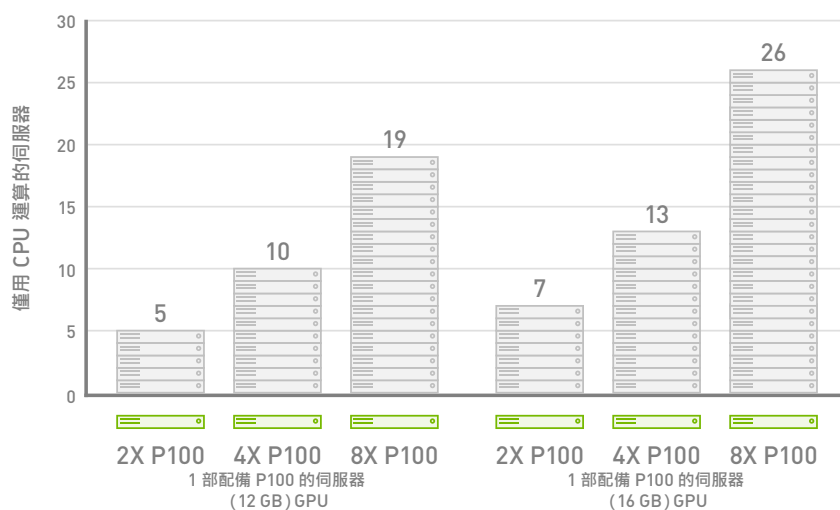
- > 搭載 P100 的伺服器，最多可取代 50 部適用 RTM 和 SPECFEM 3D 等應用的 CPU 伺服器
- > 頂尖石油與天然氣應用皆採用加速 GPU
- > 單精度浮點效能高達 10.6 TFLOPS
- > 記憶體容量高達 16 GB 且記憶體頻寬高達 732 GB/s

檢視所有相關的應用項目：

[www.nvidia.com/oil-and-gas-apps](http://www.nvidia.com/oil-and-gas-apps)

## RTM 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器：雙 Xeon E5-2690 v4 @ 2.6 GHz。GPU 伺服器：同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本：8.0.42 | 資料集：TTI R8 3 pass | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點，以達到 CPU 節點等效。

### RTM

逆時偏移 (RTM) 模型是石油與天然氣探勘地震處理流程的關鍵要素

### 版本

2016

### 加速功能

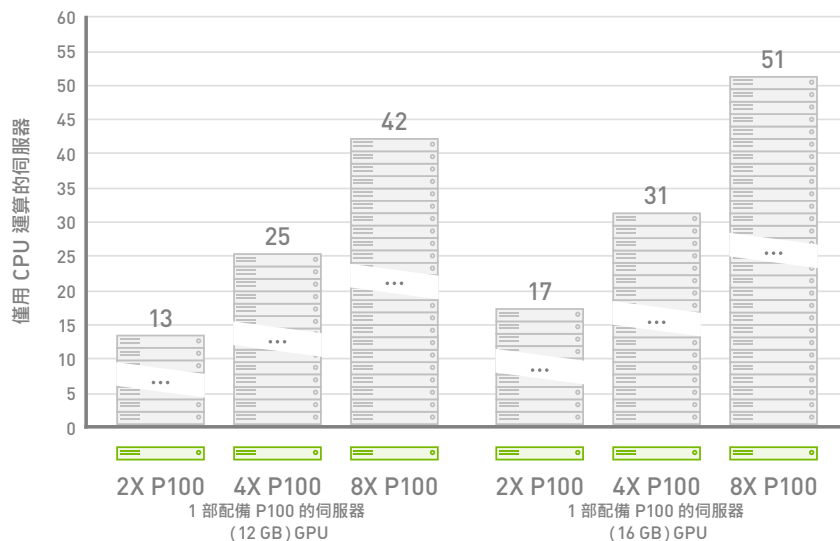
批次演算法

### 延展性

多 GPU 和多節點

## SPECFEM 3D 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器：雙 Xeon E5-2690 v4 @ 2.6 GHz。GPU 伺服器：同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本：8.0.42 | 資料集：全域 112x64 · 100 分 | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點，以達到 CPU 節點等效。

### SPECFEM 3D

模擬震波傳播

### 版本

7.0.0

### 加速功能

Wilson-clover 費米子、Krylov 計算器和區域分解

### 延展性

多 GPU 和多節點

### 更多資訊

[www.nvidia.com/specfem3d-globe](http://www.nvidia.com/specfem3d-globe)

# 金融



模擬是金融服務公司提供快速推動業務及平價優異分析的關鍵。頂尖金融應用皆是採用加速 GPU，並能為支援 Tesla P100 GPU 的資料中心節省高達 40% 的伺服器成本。

## TESLA 平台及適用金融的 P100 的關鍵功能

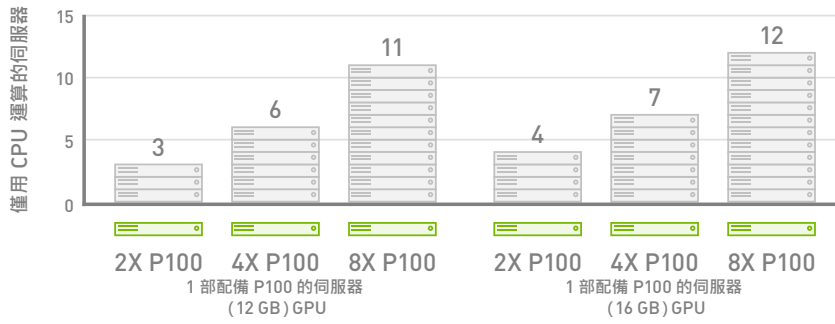
- > 搭載 P100 的伺服器，最多可取代 12 部適用 STAC A2 等應用的 CPU 伺服器
- > 頂尖金融應用皆採用加速 GPU
- > 雙精度浮點效能高達 5.3 TFLOPS
- > 記憶體容量高達 16 GB 且記憶體頻寬高達 732 GB/s

檢視所有相關的應用項目：

[www.nvidia.com/financial-apps](http://www.nvidia.com/financial-apps)

## STAC A2 效能等價

單台 GPU 運算伺服器與數台僅用 CPU 運算的伺服器比較



CPU 伺服器: 雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器: 同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | NVIDIA CUDA® 版本: 8.0.42 | 資料集: 10-100k-1260 (Warm Creek) | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點, 以達到 CPU 節點等效。

### STAC A2

Compute-intensive analytic workloads involved in pricing and risk management

### VERSION

2016

### ACCELERATED FEATURES

All

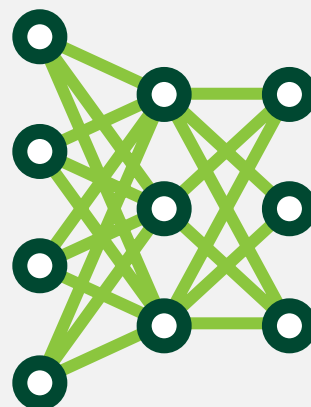
### SCALABILITY

Multi-GPU and Multi-Node

### MORE INFORMATION

[www.nvidia.com/stac-a2](http://www.nvidia.com/stac-a2)

# 深度學習



深度學習可解決數年前似乎已超越我們知識所及的重要科學、企業及消費問題。每一個主要的深度學習架構都是採用 NVIDIA GPU 最佳化，因此資料科學家與研究人員可運用人工智慧處理他們的工作。在執行深度學習架構時，配備 Tesla P100 GPU 的資料中心可節省高達 70% 的伺服器取得成本。

## TESLA 平台及適用深度學習訓練的 P100 的關鍵功能

- > 相較於 K80，配備 Tesla P100 的 Caffe、TensorFlow 和 CNTK 速度皆可高達 3x
- > 100% 頂尖深度學習架構項目皆採用加速 GPU
- > 原生半精度浮點高達 21.2 TFLOPS
- > 記憶體容量高達 16 GB 且記憶體頻寬高達 732 GB/s

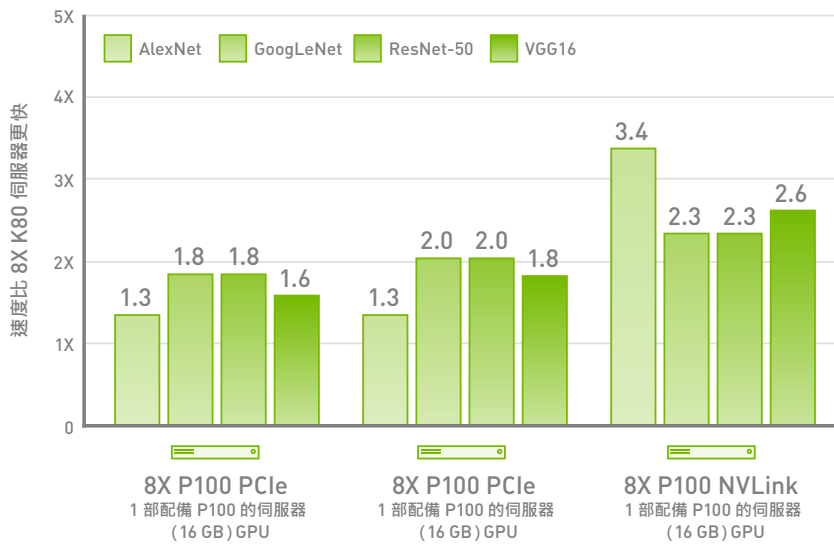
檢視所有相關的應用項目：

[www.nvidia.com/deep-learning-apps](http://www.nvidia.com/deep-learning-apps)



## Caffe 深度學習相對效能

P100 伺服器訓練與 K80 伺服器訓練的比較



CPU 伺服器：雙 Xeon E5-2690 v4 @ 2.6 GHz · GPU 伺服器：同樣搭載 NVIDIA® Tesla® P100 for PCIe (12 GB 或 16 GB) 的 CPU 伺服器 | Ubuntu: 14.04.5 | NVIDIA CUDA® 版本: 8.0.54 | cuDNN: 6.0.5 資料集: ImageNet | 批次規模: AlexNet (128) · GoogLeNet (256) · ResNet-50 (64) VGG-16 (32) | 我們使用高達 8 CPU 節點和超過 8 節點的線性伸縮測量基準點，以達到 CPU 節點等效。

### CAFFE

加州大學柏克萊分校開發出廣受歡迎的加速 GPU 深度學習架構

版本  
0.16

加速功能  
完整加速架構

延展性  
多 GPU

更多資訊  
[www.nvidia.com/caffe](http://www.nvidia.com/caffe)

## TESLA P100 產品規格



	NVIDIA Tesla P100 for PCIe 架構伺服器	NVIDIA Tesla P100 for NVLink 最佳化伺服器
雙精度浮點運算效能	高達 4.7 TFLOPS	高達 5.3 TFLOPS
單精度浮點運算效能	高達 9.3 TFLOPS	高達 10.6 TFLOPS
半精度浮點運算效能	高達 18.7 TFLOPS	高達 21.2 TFLOPS
NVIDIA NVLink™ 互連頻寬	-	160 GB/秒
PCIe x 16 互連頻寬	32 GB/秒	32 GB/秒
CoWoS HBM2 堆疊記憶體容量	16 GB 或 12 GB	16 GB
CoWoS HBM2 堆疊記憶體頻寬	732 GB/秒或 549 GB/秒	732 GB/秒

### 假設及免責聲明

採用加速 GPU 的頂尖應用百分比係取自 i360 報告中的 50 大應用清單。HPC 應用支援 GPU 運算。流量與節費計算均是假設在工作負載數據圖中，採用相同的運算循環檢測該領域之應用項目。