



# SDS vs 硬體 RAID vs 軟體 RAID

## 探索未來走向

在本白皮書中，我們將著眼於快速發展的 NVMe 技術以及它如何與 SATA/SAS 技術共存，又如何提供一系列令人驚豔的進階配置方案。

# 確認改變的步調：現在與未來觀點的儲存技術分析

過去十多年來，「軟體定義」一詞是一系列傳統網路架構 IT 服務虛擬化的詞頭。

最初是軟體定義網路 (SDN)，很快地，軟體定義運算 (SDC)、軟體定義儲存 (SDS)，甚至是軟體定義資料中心 (SDDC)，都被虛擬化為超融合式基礎架構 (HCI) 中的一個功能元素。

連接所有關係的推手是 PCIe (PCI Express) —從整合 CPU、GPU、儲存裝置和網路，進一步演進成支援 I/O 虛擬化的高速匯流排。PCIe 也搭載 NVMe (非揮發性記憶體通訊協定) 介面，這是專為多種不同外觀尺寸封裝的新型 SSD 固態硬碟所設計，支援最佳連接介面。

與 SAS 和 SATA 相比，NVMe SSD 固態硬碟對於頻寬、延遲和耗電量的表現都有顯著提升。NVMe 自從 2011 年建立標準化規格以來，其儲存介面不斷發展，以搭配 PCIe 新一代的處理能力，從 PCIe 2.0 的 5GT/s 攀升到 PCIe 4.0 的 16GT/s。PCIe 5.0 已經問世，提供 32GT/s 的傳輸速度，但要成為主流仍有一段路要走。

目前已確認 PCIe 6.0 的規格藍圖，將提供 64GT/s 的傳輸速度；另外 PCIe 7.0 規格將於 2025 年問世，預定提供 128GT/s 的傳輸速度，且向下相容。對於前述這些 PCIe 相關技術規格，NVMe 儲存裝置正好可運用這些所有優勢。

## 建立連接

目前支援 PCIe 3.0 和 PCIe 4.0 的硬體很多，單從表面來看，過渡到 NVMe SSD 固態硬碟的過程似乎相當容易。例如，可直接將擴充卡 (AIC) 插入 PCIe 插槽中使用。具備 U.3 三種連接器介面的 NVMe 磁碟能很容易地與 SATA 和 SAS 磁碟一同安裝在伺服器陣列中。比起 AIC 的可存取性缺點，U.2 磁碟架位於前端，維修方便，所以部署 U.2 是許多資料中心作業時的較佳實務選擇。

或者，如果主機板或 AIC 上具有合適的 M 型連接器鍵位，M.2 介面上的全部 PCIe x4 效能可與 NVMe 磁碟建立連接。B 型鍵位只能提供 SATA 3 或 PCIe x2 的傳輸速度。您可能看過具備 B+M 型鍵位的 SSD 固態硬碟，它們幾乎都支援 SATA 3 磁碟且具備這兩種插槽類型的相容性。

雖然以實際外觀尺寸來看 NVMe 儲存裝置的確能安裝在系統中，但真正重要的是背後的意義，故應該對如何妥善地管理 NVMe 生態系統中的儲存空間進行重新考量。

例如，為了獲得最佳處理能力效能，連接單個 U.2 或 M.2 NVMe SSD 固態硬碟將使用每個磁碟 4 PCIe 通道 (PCIe x4)，而舊系統的問題是一旦 GPU 佔滿 PCIe x16 通道，CPU 中只具備 24 個 PCIe 通道時很快就會被用完。只要使用最新的硬體，例如具備可支援最多 128 個 PCIe 通道的多核心，以及具備擴展通道數的 PCIe 切換器，這就不是什麼大問題。即便如此，在任何基礎架構元件升級計畫中仍須考慮到 PCIe 通道配置，才能確保配置達到最佳效能。

NVMe 儲存裝置的部署方式各有不同。縱使大家可能已熟悉透過軟體 RAID 和硬體 RAID 配置磁碟，但其利用率已經發展到 NVMe 儲存裝置的使用比例高於 SATA 和 SAS SSD 固態硬碟的情況。

# 軟體 RAID

NVMe 儲存裝置有個基本單純的優點：所有主要作業系統都具備支援的 NVMe 驅動程式。安裝 NVMe SSD 固態硬碟後，無論作業系統是 Windows、Linux、macOS 還是 Solaris，裝置都能順利存取。VMware 虛擬化環境也具備 NVMe 驅動程式支援功能，可支援更多不同選項設定，非常適合用於軟體定義的儲存應用。

NVMe 儲存裝置的這種現成可用性可與主流作業系統中標配的 RAID 軟體應用互補。這種簡單且有效的免費 RAID 軟體功能為消費者/終端使用者、遊戲玩家、內容創作者到成熟的企業部署或甚至所有人，提供一種強大儲存管理功能的便利方式。

基本 RAID 軟體應用可能僅提供 RAID 0 (分散連結) 和 RAID 1 (資料鏡射) 且分別用於增進效能和資料安全。**事實上，硬體 RAID 能提供的 RAID 級別，比軟體 RAID 替代方案多上許多。**即便如此，如 Linux 上預設的軟體 RAID 應用 mdraid 等應用程式，也有提供 RAID 4、5、6 和 10，這是能提供效能和資料安全的平衡組合。

SSD 固態硬碟尚未對應個別硬碟的儲存容量，故預定數量磁碟的整體儲存容量需求也是配置 RAID 陣列的主要考量因素。此外，使用軟體管理 RAID 儲存環境中的資料分布和同位元檢查功能，執行這些例行檢查會對主機 CPU 造成影響。演算法作業的複雜性各不相同，例如，寫入比讀取的運算密集度高，如果事務處理量很大，由於 RAID 配置的冗餘較高，這些任務可能會影響整體效能。

當軟體授權費用是依照核心數作為基準時，讓系統承擔儲存任務真的有意義嗎？長期以來這一直是硬體 RAID 的爭議點，但我們現在已不在 SATA/SAS 環境中。某種程度上，軟體 RAID 固有的效能損耗已被 NVMe 的延遲和處理能力改善，以及其直接存取 PCIe 匯流排的特性所抵銷。

## 較佳的設計

此外，SATA 介面是專為傳統硬碟所設計的，用在 SSD 固態硬碟上本就是一種妥協。與 HDD 傳統硬碟相比，SATA SSD 固態硬碟的速度表現相當優異，但這只是快閃儲存實際效能的一小部分而已。SATA 使用的進階主機控制器介面 (AHCI) 和其所對應的特性 (需要為旋轉磁盤型硬碟建立超過 120 個指令)，讓系統升級得以與快閃儲存相容，但終究還是有速度瓶頸。相較之下，NVMe 最少只需要 13 個指令，包括 10 個管理和 3 個 I/O：讀取、寫入、釋放。

而在指令列部分，AHCI/SATA 技術只具備一列，每列可傳送 32 個指令。相較之下，NVMe 具備 64,000 個 I/O 指令列，每列最多有 64,000 個指令，這表示 CPU 週期使用會顯著降低。

NVMe 儲存技術使用流線型的 PCIe 資料路徑，並具備極高的處理能力和效率，讓軟體 RAID 能在此領域中有不同以往的表現。這並非軟體 RAID 的侷限性，反而證明其在此領域中具一定效能表現。事實上，傳統意義上的硬體 RAID 不斷地發展以支援擴展 NVMe 儲存技術，所以對於大多數人而言，這一直是唯一選擇。

# 硬體 RAID

硬體 RAID PCIe 介面卡具備專用控制器晶片，可執行所有必要的運算功能，用來在目標儲存硬體上建立並管理 RAID 陣列。將處理作業全部卸載到 RAID 卡，硬體 RAID 可提供相當多具有不同複雜性的 RAID 級別，故主機平台就沒有這些處理作業負擔。

沒有佔用昂貴的主機 CPU 資源來處理 RAID 演算法，因此改善了讀取和寫入速度，也支援磁碟熱插拔功能。而使用軟體 RAID 時，缺乏專用處理元件會增加高儲存容量 SAS/SATA 環境中的延遲和所佔用的處理能力。這與硬體 RAID 不同，軟體 RAID 更換磁碟時通常需要移除先前執行的 RAID 管理程式，而且通常要重新啟動。

雖然要多花一些成本，但硬體 RAID PCIe 介面卡的低延遲、資料保護、快取功能，以及磁碟陣列擴展功能，使其在企業儲存管理領域中佔有一席之地。它也在進化發展。專用的 NVMe RAID 介面卡是市面上較新的產品，但 Broadcom、Marvell、Microchip 和其他供應商均提供支援結合 SATA、SAS 和 NVMe 功能的三模式 PCIe Gen 4 ROC (RAID-on-chip) 介面卡。

**這些硬體 RAID 介面卡能為 NVMe SSD 固態硬碟在混和儲存環境中共存的情況，提供一種簡化方式。透過基本的佈線方式，可將 U.2 背板配置成使用 U.2 外觀尺寸的 SATA/SAS 和 NVMe SSD 固態硬碟組合。**

U.3 規格標準的出現，又使得這種外觀尺寸往前邁一大步，藉由真正可容納統一佈線的三模式背板降低複雜性。但有一個問題：U.3 實體磁碟介面相同，但接腳配置有做過更動。因此，U.3 磁碟可用於 U.2 背板，但 U.2 磁碟無法相容 U.3 背板。

U.3 的混合及搭配功能似乎是有價值的發展目標，但這種配置能被多廣泛使用還是個問題。



## 磁碟架的觀點

當然，通用背板管理 (UBM) 標準的問世，進一步達成了混合儲存部署，以及相容 U.2 和 U.3 的設計。在超過 20 多家尖端儲存硬體供應商所組成的聯盟支援下，UBM 讓主機和磁碟控制器能偵測到背板功能，並支援檢測和監控不同磁碟類型 (SATA、SAS 和 NVMe)，即便單個磁碟槽也能支援。UBM 還能與 SATA/SAS 擴展卡和 PCIe 切換卡搭配使用，並提供一系列實用的背板管理功能，進一步強化 U.2 和 U.3 系統架構。

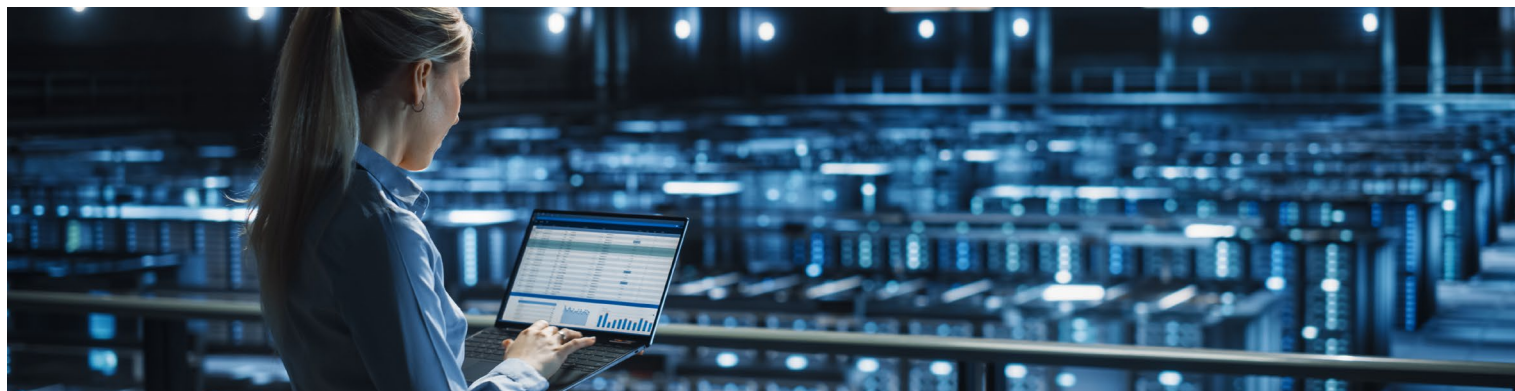
三模式 RAID 或 HBA (主機匯流排配接器) 卡將使用 x8 或 x16 PCIe 主機通道並具備 PCIe 切換功能，通道數量可倍增並有效增加頻寬。介面卡規格可能最多能引入 32 個 NVMe 裝置，但這和使用全部 PCIe x4 速度支援 8 個 NVMe SSD 固態硬碟不同，後者需要 32 個 PCIe 通道。理論上，這可以容納 32 個 x1 速度的實體 NVMe 磁碟，即使是在 PCIe 3.0 環境中，每個磁碟也能以 1000MB/s 的速度執行，這比 SATA 的 600MB/s 處理能力快上 2/3 倍。即便如此，有鑑於 NVMe SSD 儲存裝置的卓越效能可借助並行 PCIe 通道來達到效能最佳化，故這樣的配置只是 NVMe SSD 儲存裝置的其次運用選擇。在混合使用的環境中，三模式控制器可能只將 x8 或 x16 通道專用於 NVMe 儲存，這同樣會跟磁碟數選擇較少以及處理能力降低有關。

無縫整合在單一個背板中的不同磁碟類型，是否能激發出一個機殼內管理熱 (NVMe)、溫 (SAS/SATA) 和冷 (SATA/HDD) 儲存需求的精品級架構系統，尚有待觀察。

畢竟，劃分 PCIe 通道分配來維持與傳統儲存裝置相容，是一種妥協，即便支援 NVMe，但仍有其侷限性和成本。目前許多對現有 SAS/SATA 部署感到滿意的系統，可能只是關心磁碟更新，確保可靠性並提高儲存容量。U.2 儲存可能還會存在一段時間，但僅使用單一種類型的裝置通道

可能會變得很普遍，以利用現有 SAS/SATA 儲存裝置和低成本專用控制器和擴展卡。同樣地，為了最佳化效能和儲存容量，建議最好只使用 NVMe SSD 固態硬碟。

採用 NVMe 儲存的步調很大程度上取決於工作負載的強度，以及它能對現有系統效能強化到多少程度。投資大量純 NVMe 部署的雲端服務供應商已經感受到這些優點，因為頻寬帶來的巨大優勢，可推出新的服務模式，提供分級以滿足各種需求的客戶。



## 修正的期望

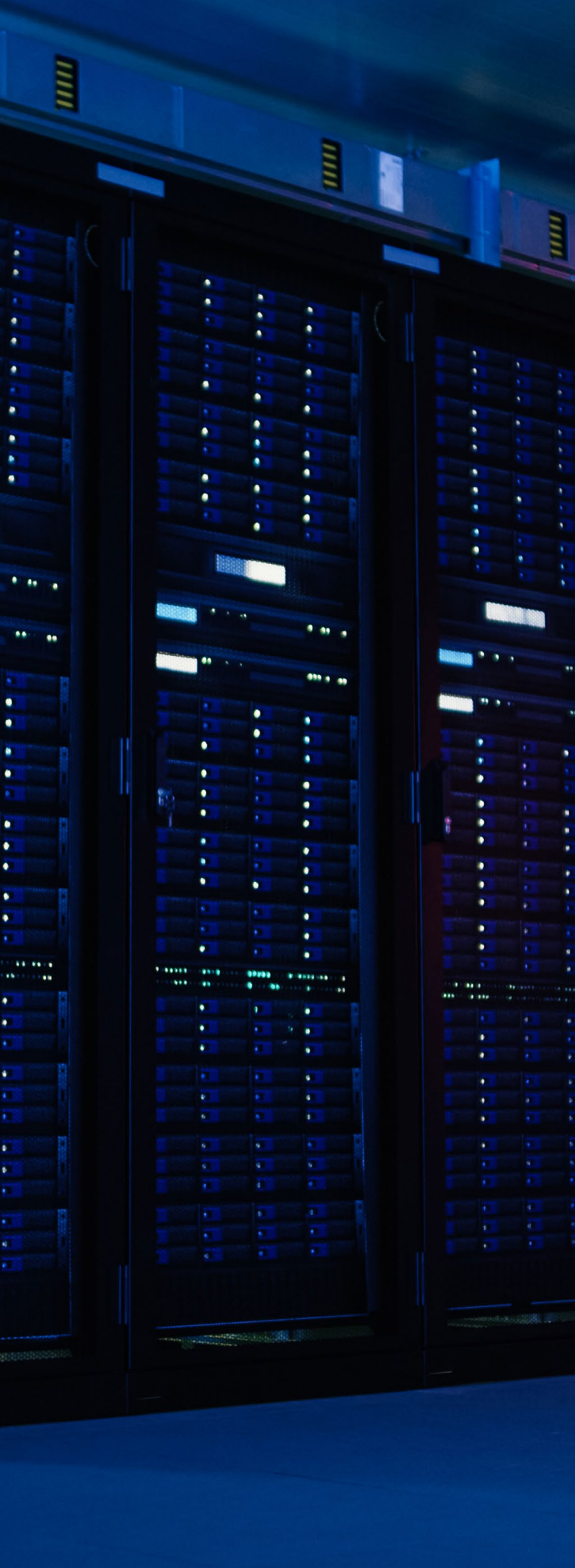
**企業介於靈活且著重 NVMe 的雲端服務供應商和傳統的資料中心這兩個極端之間，往往需要更多功能、更高效率和可擴展性。**

透過對成本、優勢、整合及最佳化的研究成果，正在擁抱 NVMe，但使用更針對性的方式，採用效果十分驚人。

當 NVMe 儲存用於快取時，很快就會曝露出系統缺陷，例如，低效率程式碼的應用系統會限制預期中的延遲和處理能力改善。其他瓶頸也會逐漸顯露出來，需要解決這些問題，才能達成 PCIe/NVMe 生態系統的效能優勢。

這並不是交換類似品，更像是自行車被子彈列車所取代。有鑑於此，也需要重新評估 SSD 固態硬碟規格，因為服務級別協議可能會堅持某些作業模式，但這些作業模式無法識別出您可以使用 NVMe 儲存做多少事情。

舉例來說，磁碟每日全碟寫入次數 (DWPD) 的資料是用來確認快閃儲存裝置的保固壽命耐用性。快閃儲存裝置很容易受到所謂寫入放大的影響，這是因為在儲存單元中儲存資料所採用的方法，容易增加 SSD 固態硬碟的耗損。從本質上來說，資料不能直接儲存在單元上，必須先清除資料才能進行覆蓋，而且隨著使用時間拉長，這個複雜的流程會導致儲存效能下降。超容量快取是一種 SSD 固態硬碟的儲存容量預備區，用來克服這些問題並執行磁碟例行處理流程，例如故障區塊管理。這是一個重新分配並儲存資料以釋放儲存區塊 (接著清除並準備寫入) 的流程，也是造成寫入放大的主要原因。



# 分區儲存

最近 NVMe 2.0 的一個補充規範是分區儲存 (ZNS)，其針對 NVMe SSD 固態硬碟的讀取/寫入程序提供了一種新方法。分區區塊管理介面位於主機和 NVMe SSD 固態硬碟之間。分區儲存和磁碟分區有點類似，但在主機應用程式中有不同級別。ZNS 讓 SSD 固態硬碟能與主機溝通、描述或提示執行效能，例如提供資料配置的最佳模式和佈局詳細資訊，作為 ZNS 寫入和清除作業依序執行的依據。

這種互相協作會將部分儲存管理功能卸載到主機應用系統，優點在於能降低超容量快取的需求，並且有機會增加多達 20% 的儲存空間容量。實行 ZNS 可改善 I/O 延遲，並將磁碟放大寫入問題降低 4 至 5 倍之間。此外，可以針對不同區域，分配特定工作負載或資料類型，來達到進一步可預見的效能模式。

ZNS 的採用目前仍處於起步階段，但 ZNS 已經是 Linux kernel 5.9 版本的一個功能。此外，對 ZNS 的研究是由 Microsoft、阿里巴巴和 NetApp 所贊助，著眼於其超大規模的作業研究，這表示 ZNS 被業界大規模採用只是時間問題而已。

隨著此功能集的發展，需要更新應用系統並加以充分利用，而且，目前越來越多 NVMe 磁碟採用 ZNS，在某些情況中使用現有 NVMe SSD 固態硬碟，可能只需要更新韌體。

系統架構師要留意這些嚴格規範。看來是重新編寫規則手冊來說明 DWPD 實際意義的時候了。執行 ZNS 後，寫入放大問題大幅降低，代表磁碟耐久度大幅提升。還需要多少磁碟？由於超容量快取大幅降低，磁碟的儲存容量因此顯著增加。展望資料管理的未來性，借助 NVMe SSD 固態硬碟和 ZNS 介面，確實可讓您事半功半。

# 軟體定義儲存

NVMe 帶來了不折不扣的混合採用方式，從 M.2 磁碟、PCIe 擴充卡，到 U.2 或 U.3 儲存裝置都有。新興的企業和資料中心 SSD 外觀尺寸 (EDSFF) 是另一種儲存裝置格式，專為 NVMe 生態系統的磁碟所設計，具有兩種寬度 (E.1 和 E.3)，以及兩種尺寸 (長/L 和短/S) 規格。E.1L 磁碟可在 1U 機殼中實現高密度儲存，而更靈活的 E.1S 尺寸則具有適用可擴展性的熱效率優勢。E.3 磁碟可作為 U.2 2.5 吋 SSD 固態硬碟的替代品，其適合較傳統的 2U 伺服器磁碟陣列機殼，這是設計來讓每個磁碟容納更多快閃晶片以提高儲存密度。

當然，將 NVMe 視作一般標準，並在所有主流作業系統中提供驅動程式支援，執行前述各種選項做法時都會不那麼麻煩。選擇時依照最符合工作負載和冗餘要求的儲存特性和配置。這可能和在邊緣伺服器上整合 NVMe 儲存裝置有關，是運用 SAS/SATA 硬體來執行密集度較低的作業。硬碟甚至磁帶備份也能作為儲存基礎架構中的一部分。由於企業儲存管理中不乏有專用平台，協調這相異的儲存系統只會迅速增加複雜性。此時就是軟體定義儲存 (SDS) 發揮作用的時刻，其可針對混合儲存區提供協調方式，並最佳化其運用模式。

在軟體定義儲存的領域中，是從儲存硬體中將可用儲存資源抽象化並虛擬化出來。使用產業標準通訊協定，甚至可以透過 SDS 虛擬化存取專用硬體，龐大的儲存裝置可不受限制地成為較大儲存池中的一部分，該儲存池可能還具備使用商用伺服器所建構的新型低成本可擴展儲存空間。這種分離的模式還能避免在更換、升級或擴展儲存硬體時造成作業中斷。

將所有可用儲存空間合併到虛擬儲存池中，需要決定某些配置決策，並具備多種功能來協助進行配置，包括自動化。在 SDS 控制面板中，會根據各儲存池中的儲存硬體



配置檔來識別出熱、溫和冷儲存空間。使用指令集來執行任務，分配和分發最適合這些儲存庫的資料負載。

**SDS 可藉由虛擬儲存層級來提供靈活性和可擴展性，管理適合企業需求和客戶可變動需求的儲存環境和部署，如暫存及配置虛擬主機，以處理鏡像和備份複製等。**

對於 NVMe SSD 固態硬碟，SDS 平台可使用被稱為穿透模式的功能，直接透過 PCIe 匯流排存取儲存裝置。例如，VMware 在其 ESXi/vSAN SDS 上有提供自己的 NVMe 儲存驅動程式，可使用被稱為 VMDirectPath I/O 的功能，將 NVMe 儲存空間直接分配給虛擬主機。根據主機 CPU 配置，每個虛擬主機最多支援 16 個穿透裝置。

整體而言，啟用 NVMe 穿透功能，來自主機的干擾可降到最低，藉此提高效能，並簡化針對虛擬主機實體和其他服務配置 NVMe SSD 固態硬碟的流程。既然如此，第三方軟體或硬體 RAID 控制器是否支援 NVMe RAID 功能，對於 SDS 來說就不再是個問題，因為其可直接配置 NVMe 軟體 RAID。

SDS 有機會成為資料管理的萬靈丹，但其成本和初期配置的複雜性，可能會讓有更多明確需求的企業停下來多考慮一下。不過，這就像儲存裝置本身一樣，這些成本是可以延伸擴展的，並且可使用不同版本來處理較小的硬體部署。



## 節奏的變化

儲存裝置正在不斷演化，雖然不太可能一夜之間發生驟變，但現有的資產很有可能已經準備報廢。因此，儲存領域的研發會繼續朝硬碟和 SATA SSD 固態硬碟等技術方向上邁進。這些產品都各有自己的角色定位，持續在儲存陣列中提供實際服務。例如，Kingston 的 [DC600M 2.5 吋混合用途企業級 SATA SSD 固態硬碟](#)，儲存容量倍增至 7.68TB。

**硬體 RAID 和 HBA (主機匯流排配接器) 仍在全球資料中心中佔主導地位，而供應商也不斷創新以滿足這些不斷擴張的 IT 產業需求。**

透過與 Broadcom 和 Microchip 合作，Kingston SSD 固態硬碟經過嚴格測試，確保其滿足目前資料推動技術的嚴苛要求。

搭配這些頂尖供應商的儲存裝置配接器，經極端工作負載和特殊配置等測試的 Kingston 企業級 SSD 固態硬碟，能確保提供良好的效能、耐久性和可靠度品質。當然，Kingston 的 [U.2 DC1500M PCIe NVMe Gen3x4 企業級 SSD 固態硬碟](#) 已完成上述測試。憑藉高達 7.68TB 的儲存容量和結合 1 DWPD，完全有資格成為最新一代伺服器儲存陣列中的重要角色。

也可針對消費型商品硬體去建構軟體定義儲存的裝置，但 SSD 的選擇在規模上就變得更加關鍵。消費級 SSD 固態硬碟的成本很具吸引力，若與針對耐久性和持續高頻寬負載所建構的企業級 SSD 固態硬碟相比，消費級 SSD 固態硬碟只是表面上省錢而已。我們提供超融合式基礎架構所需效能和可預測情況的服務，以便有效地管理工作負載，並滿足客戶期望。[Kingston 企業級 SSD 固態硬碟](#) 可與 VMWare 儲存應用程式配合使用，確保即便在軟體定義儲存的虛擬環境中，也能達成現實環境中的目標。

儲存配置正在發生變化，但改變的速度會因其適應的各種商業模式而異。從舊有介面的連接性到 NVMe 的創新，每個層面都有所改善。如果升級看起來是一項艱鉅的任務，Kingston [諮詢專家](#) 服務可為您提供協助。我們可免費協助您，決定適合您業務並符合預算的重要決策。無論這段升級旅程中您身在何處，Kingston 都會與您同在。

#KingstonIsWithYou

©2023 Kingston Technology Far East Corp. (Asia Headquarters), No. 1-5, Li-Hsin Rd. 1, Science Park, Hsin Chu, Taiwan  
版權所有，保留所有權利。所有商標及註冊商標係屬於各自所有者之智慧財產權。

