在 IBM Power Systems 部署 IBM DB2 pureScale 的功能

DB2 pureScale 共用磁碟架構的優勢

技術等級：中級

Miso Cilimdzic (cilimdzi@ca.ibm.com)
DB2 效能經理
IBM

Sanjeeva Kumar Ogirala (sogirala@in.ibm.com)
軟體工程師
IBM

2010 年 9 月 9 日

Enterprise Server Edition 的 DB2® pureScale™功能，是利用 IBM® DB2 for z/OS®資料庫軟體中常見的成熟設計功能所建立。本文將說明在 IBM Power Systems™ 上部署 DB2 pureScale 功能的各種不同方法，以及運用硬體零件元件來建立 pureScale 叢集的方法。

簡介

DB2(R) pureScale(TM)的功能可藉由提供近乎無限制的功能、連續的可用性，以及應用程式透明度，來協助降低企業成長的風險與成本。DB2 pureScale 會因為低延遲的互連(例如，InfiniBand)而受惠，而且會建立於共用磁碟架構的頂端之上。為了達成低延遲的目標，DB2 pureScale 會使用 Power Systems InfiniBand Host Channel Adapters (HCA)及交換器，而

光纖通道 SAN 可提供存取共用磁碟的能力。

本篇文章會解決下列問題：

- DB2 pureScale 的成員會以什麼方式連線？

- 成員及 PowerHA™ pureScale 伺服器會以什麼方式連線？
- 成員及 PowerHA pureScale 伺服器的 AIX® LPAR 在叢集中會以什麼方式連線？
- 一部主機的個別或每個 AIX LPAR 是否會連線到其他主機的個別或每個 LPAR？
- SAN 儲存體會以什麼方式連線到叢集？

本文將說明 DB2 pureScale 叢集硬體~~針對 DB2 pureScale 生產系統共同~~架構耦合的方法。本文也會澄清與安裝 DB2 pureScale 叢集相關的概念。安裝及設定叢集需要 UNIX(R)、InfiniBand 及 SAN 儲存體的專業能力。

瞭解 DB2 pureScale 的功能

DB2 pureScale 的功能會使用 IBM DB2 RDBMS 共用磁碟的技術。當您聽到 DB2 pureScale 時，通常是指利用以下數種緊密耦合的元件所組成的叢集架構解決方案：

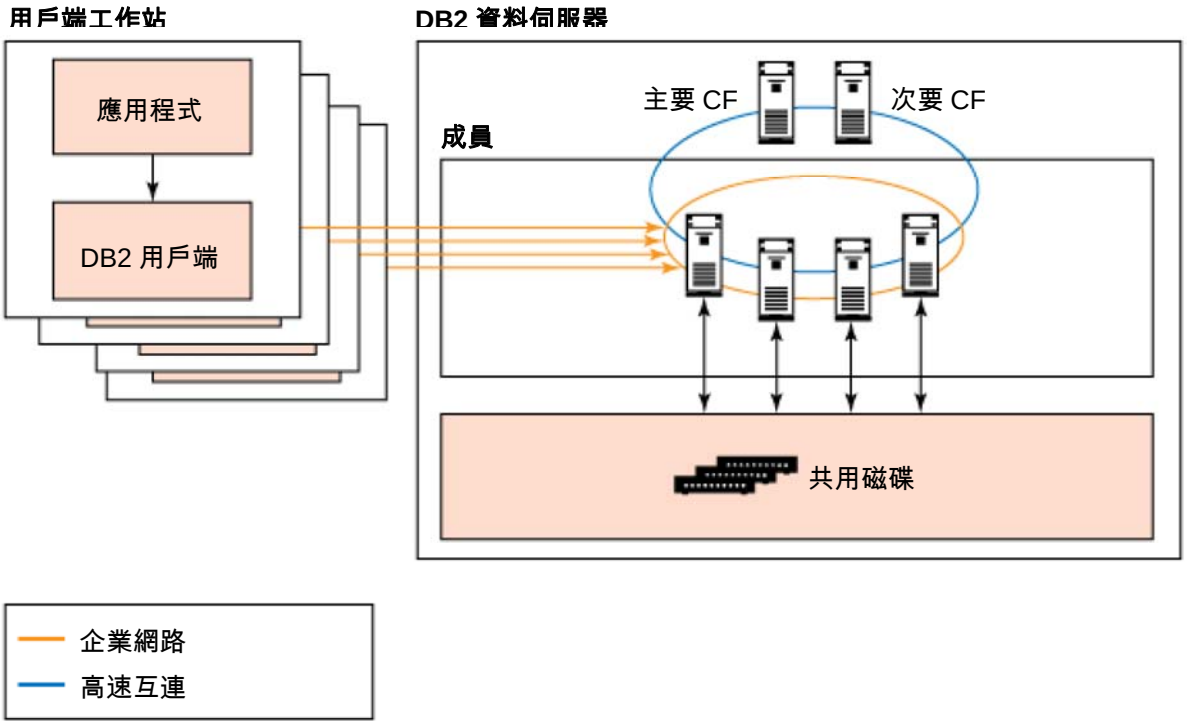
- 至少兩個 DB2 成員
- PowerHA pureScale 伺服器(CF)
- 高速通訊網路(例如 InfiniBand)
- IBM Tivoli® System Automation for Multiplatforms (Tivoli SA MP)軟體
- IBM Reliable Scalable Clustering Technology (RSCT)軟體
- IBM General Parallel File System (GPFS™)軟體

DB2 pureScale 會藉由提供更方便擴充或縮減的方法來解決容量及~~可用性~~可用度的問題，並確保整個資料庫隨時都可供使用。共用的磁碟讓所有成員都可以存取相同的資料集。任何成員故障或 CF 故障(~~假設出現雙工~~在兩組 CF 的情況)都不會影響資料庫的可用性。如果使用 DB2 pureScale，您只需將新成員加入現有的叢集，即可增加額外的容量。PowerHA pureScale 伺服器的 Global Buffer Manager (GBP)及 Global Lock Manager (GLM)提供集

中式資料存取同步處理的功能。

圖 1 顯示具有四個成員及兩個 CF 之 DB2 pureScale 實例的~~高階檢視圖~~示意圖。其中顯示 DB2 用戶端連線到資料伺服器~~的 DB2 用戶端~~。DB2 成員~~正在~~處理資料庫的要求，而 PowerHA pureScale 伺服器會提供集中式同步處理的服務。資料會儲存在所有成員都可以存取的共用磁碟儲存體中。

圖 1：DB2 pureScale 環境中的主要元件檢視圖



瞭解構成解決方案的硬體元件

以下是本篇文章描述 DB2 pureScale 環境中所需硬體的清單：

- 使用 AIX 的 IBM POWER6®或 POWER7®伺服器
- 光纖通道 SAN 儲存體、SAN 交換器及主機匯流排轉接器(HBA)
- InfiniBand 交換器、InfiniBand Host Channel Adapters (HCA)，以及纜線
- 乙太網路轉接器
- 硬體管理主控台(HMC)

以下章節將簡短說明解決方案的每個項目。

IBM POWER6 或 POWER7 伺服器

這些伺服器是~~部署~~安裝部屬 DB2 pureScale ~~二進位~~軟體的 POWER6 或 POWER7 電腦，使用的是 AIX Logical Partitions (LPAR)。建議使用至少兩個成員及兩部 PowerHA pureScale 伺服器。建議您將每個成員及 PowerHA pureScale 伺服器部署在其專屬的 LPAR 中，以及跨至少兩部 POWER6 或 POWER7 電腦。

目前支援以下的 POWER Systems：

- POWER6 550
- POWER6 595
- POWER7 710
- POWER7 720
- POWER7 730
- POWER7 740
- POWER7 750
- POWER7 755
- POWER7 770
- POWER7 780
- POWER7 795

光纖通道 SAN 儲存體、交換器及 HBA

所有~~的~~ DB2 成員會共用光纖通道連接~~的~~ SAN 儲存體。DB2 pureScale 會因為支援 SCSI3-Persistent Reserve 的儲存體受惠。當儲存體故障時，DB2 pureScale 會使用這項技術快速~~阻止~~隔離出現異狀~~異常~~的成員，如此可確定資料庫檔案能保持一致。~~如需要~~通過測試支援 SCSI3-PR 且獲得 GPFS 支援的儲存體清單，請參閱資源(Resources)部分中的線上 GPFS 常見問題(FAQ)。

因為共用的資料~~位於~~是 DB2 pureScale 系統的核心，所以建議您使用 RAID 組態以提供最高的備援能力及可用性。~~部分~~某些容錯性更高的 RAID 等級(例如 RAID10 及 RAID6)有助於提供額外的保障，讓儲存體子系統可以承受各種磁碟錯誤。

SAN 交換器通常用於連接伺服器和儲存體控制器。若要部署 DB2 pureScale，則 SAN 交換器必須要有備援，而且也要連接到不同的電源供應器，才能提供最高的可用性。

主機匯流排轉接器(HBA)通常會透過 SAN 交換器並利用光纖通道纜線連接伺服器和 SAN 儲存體。建議您在每個 DB2 成員上安裝備援 HBA，並使用多重路徑的軟體(例如 IBM AIX

MPIO)，或是支援多重路徑存取 LUNS 的裝置磁碟機。請注意，部分這種多重路徑的磁碟機可以使用負載平衡，如此可以在使用多個 HBA 時增加傳輸量。

InfiniBand 交換器、HCA 及纜線

InfiniBand 是一種用來在 DB2 成員及 PowerHA pureScale 伺服器間進行通訊的低延遲高頻寬互連。InfiniBand Host Channel Adapter (HCA)是一種可讓伺服器進行連線的裝置。HCA 會使用 InfiniBand 纜線連線到 InfiniBand 交換器，形成一個子網路。InfiniBand 連線將於下方的「使用 InfiniBand (IB)」做進一步的描述。

乙太網路轉接器

乙太網路轉接器通常會連接到企業網路，讓 DB2 用戶端能夠和 DB2 pureScale ~~實例~~資料庫連線(例如，EtherChannel 或 Network Interface Backup 技術)。DB2 pureScale ~~的功能~~會將連線~~要求以最低的工作負載自動路由傳送至成員~~導向工作負載最輕的成員。此外，您也可以指定 DB2 用戶端連線到 DB2 pureScale ~~實例~~中的特定~~使用中~~成員。

硬體管理主控台

IBM 硬體管理主控台(HMC)提供系統管理員一種用來規劃、部署及管理 IBM System p®伺服器的工具。HMC 提供伺服器硬體管理及虛擬化(磁碟分割)管理。

使用 InfiniBand (IB)

HCA、InfiniBand 纜線及 InfiniBand 交換器形成一個子網路。這個網路的效能很重要，因為該網路是用來在叢集中傳遞鎖定及快取的資訊。~~實例中的~~所有主機必須使用相同類型的互連。DB2 pureScale 會利用 InfiniBand 所提供的遠端直接記憶體存取(RDMA)~~支援的功能 InfiniBand~~。使用 RDMA 提供在成員主機記憶體中直接進行更新的能力，而不需要耗用成員處理器時間。~~每個~~ IB 元件及其零件編號，將於後續章節中說明。

Host Channel Adapters (HCA)

IBM GX++ HCA 安裝在作為部分 DB2 pureScale 叢集的 POWER System 伺服器中。DB2 pureScale 僅支援 GX++ HCA 轉接器。表 1 顯示擁有功能代碼的支援轉接器清單。

表 1：POWER System 伺服器型號及支援的 HCA 轉接器

POWER System 伺服器型號	HCA 功能代碼
500、750	5609
595、795	1816
710、730	5266
720、740	5615
770、780	1808

HCA 連接到 IB 交換器

HCA 會透過一條 12x 對 4x 的 IB 纜線(例如 FC 1854 下方的 10 公尺銅纜線)或 4x 對 4x 的 IB 纜線(如 FC 3246，這是僅用於 FC 5266 的 4x 對 4x 纜線)連接到 IB 交換器。

伺服器中多個 LPAR 連接到 IB 光纖

有多種方法可以連接 LPAR，而連接的方法則取決於 LPAR 的數量以及支援該伺服器型號的 HCA 數量。部分的選項包含下列項目：

具有一個 LPAR 的 POWER 750

HCA 會指派給 LPAR。一條 IB 纜線連接到 IB 交換器。

具有兩個 LPAR 的 POWER 750

HCA 會依邏輯使用 POWER 管理程序(hypervisor)進行分割，而每個 LPAR 會獲得指派的部分 HCA 頻寬與資源。一條 IB 纜線連接到 IB 交換器。

具有兩個 LPAR 的 POWER 770

安裝兩個 HCA，且每個 LPAR 都有專用的 HCA。兩條 IB 纜線連接到 IB 交換器。

具有多個 LPAR 的 POWER 770

安裝一或多個 HCA。每個 LPAR 都有專用的 HCA，或是部分或全部的 LPAR 共用 HCA。和 HCA 相同數量的 IB 纜線連接到 IB 交換器。

InfiniBand 交換器

IB 交換器位於 InfiniBand 光纖的中央，會將所有的 DB2 pureScale 伺服器連接到一個子網路。IBM 7874 IB 交換器產品線提供從 24 到 240 埠的各種交換器。

表 2 會列出支援的 IBM POWER Systems InfiniBand 交換器。

表 2：支援的 IBM POWER Systems InfiniBand 交換器

功能代碼	支援的交換器

7874-024	1U，24 埠 4x DDR IB Edge Switch (QLogic 9024CU)
7874-040	4U，48 埠 4x DDR IB Director Switch (QLogic 9040)
7874-120	7U，120 埠 4x DDR IB Director Switch (QLogic 9120)
7874-240	14U，240 埠 4x DDR IB Director Switch (QLogic 9240)

探索範例部署模式

有許多種組合方式可以部署 DB2 pureScale ~~功能的~~伺服器~~有各種組合方式~~。
本節將說明部分常見的部署模式。

- 部署兩部伺服器
- 部署三部伺服器
- 部署四部以上的伺服器

表 3 顯示這三種模式的組態。

表 3：三種組態模式

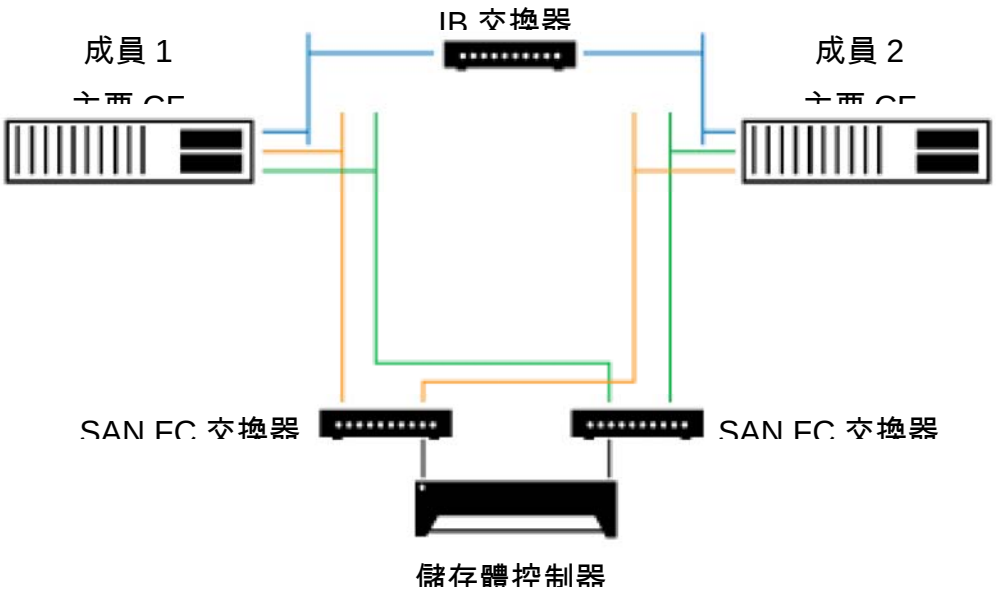
元件	伺服器數量	LPAR 數量	IBM IB 交換器	IBM IB HCA	IBM IB 纜線	FC SAN HBA	FC SAN 交換器	FC SAN 纜線	FC SAN 儲存體控制器
2 部伺服器的模式	2	4 (每部伺服器 2 個 LPAR)	~~強制性~~必要	最少 2 個	最少 2 個	最少 2 個雙埠	選擇性	4 條纜線 (每部伺服器 2 條)	~~強制性~~必要
3 部伺服器的模式	3	5 (2 個 LPAR 位於兩部伺服器上 , 1 個 LPAR 位於一部伺服器上)	~~強制性~~必要	最少 3 個	最少 3 個	最少 3 個雙埠	選擇性	最少 6 條纜線(每部伺服器 2 條)	~~強制性~~必要
4 部以上伺服器的模式	4 個以上	4 個以上	~~強制性~~必要	每部伺服器最少 1 個	每部伺服器最少 1 個	每部伺服器最少 2 個雙埠	選擇性	每部伺服器最少 2 個	~~強制性~~必要

部署兩部伺服器

為了保持高可用性(HA)的特性，兩部伺服器是最基本的組態。在這種組態中，每部伺服器會有兩個 LPAR(一個 DB2 LPAR、一個 PowerHA pureScale 伺服器 LPAR)。在這種組態中，即使遺失一部實體伺服器故障，DB2 pureScale 系統實例也能夠繼續使用，因為剩下的實體伺服器中還有一個 DB2 成員和一部 PowerHA pureScale 伺服器可以使用。

任何一部伺服器出現硬體故障後，或是處於在硬體維護的空窗期，這種組態都無法維持高可用性。IB 卡可以專屬於每個 LPAR (如果伺服器支援一個以上的 HCA)，或是由每個 LPAR 共用。同樣地，HBA 也可以專屬於每個 LPAR，或是透過虛擬 I/O 伺服器(VIOS)共用。每個 IB HCA 都會透過 IB 纜線連接到 IB 交換器。同樣地，HBA 轉接器也會利用 FC SAN 纜線連接到 FC SAN 交換器。圖 2 顯示這種組態。

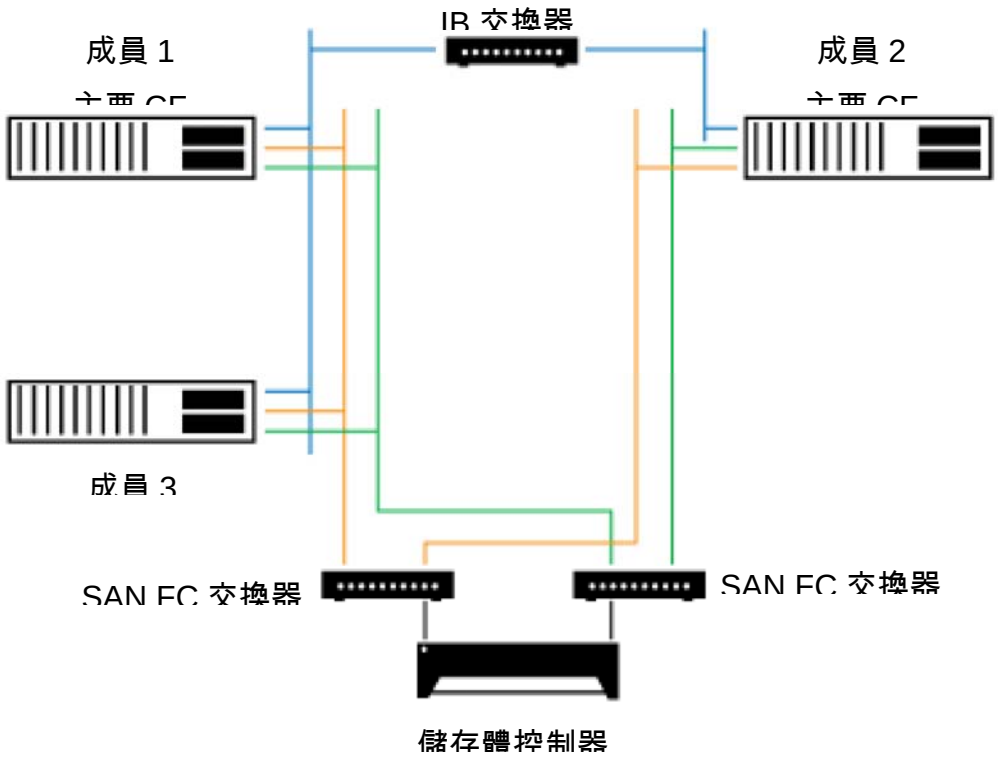
圖 2：透過纜線連接的四個 LPAR、兩部 POWER 伺服器的組態



部署三部伺服器

三部伺服器的部署模式，可在硬體故障或某部伺服器(例如，未搭載 PowerHA pureScale 伺服器 LPAR 的伺服器)的硬體維護期間保持高可用性。在這種組態中，兩個不同的伺服器各有一個成員 LPAR (總共三個成員)以及兩個 PowerHA pureScale 伺服器。除了伺服器唯一主控的成員 LPAR 擁有專用的 HCA 以外，IB 及 FC SAN 連線的說明，都和兩部伺服器的設定相同。圖 3 顯示此組態。

圖 3：使用纜線連接五個 LPAR 及三部 POWER 伺服器的組態

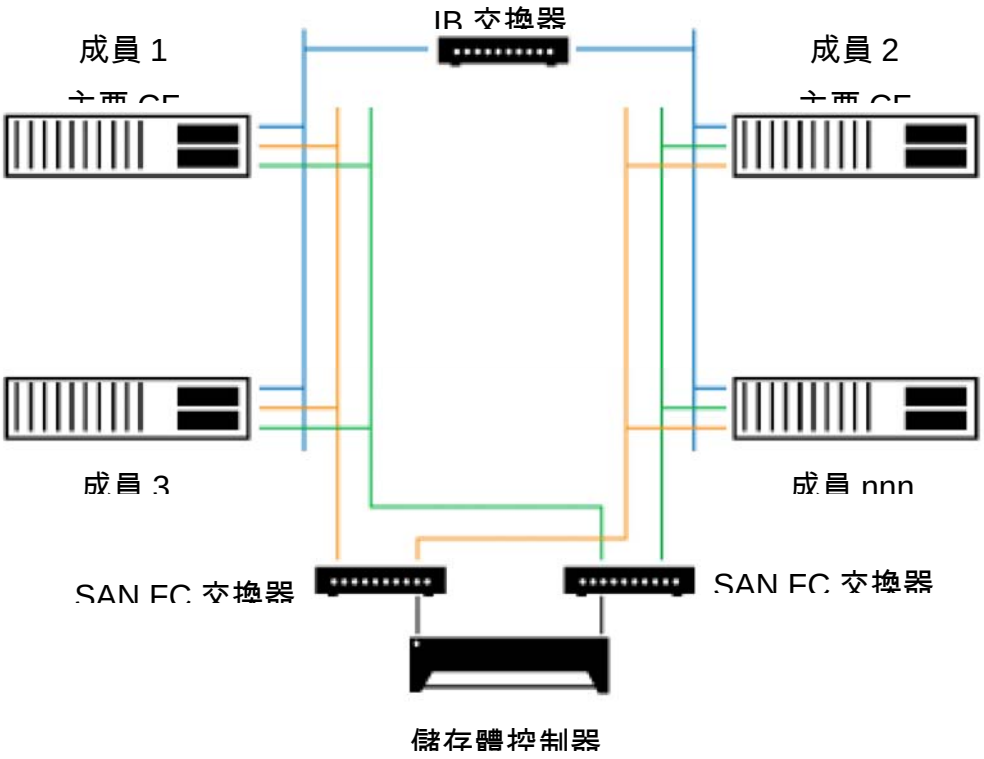


部署四部以上的伺服器

四部以上的伺服器部署模式可以增加額外的成員，並可選擇將 PowerHA pureScale 配置於~~隔離~~專用的伺服器上~~的 PowerHA pureScale 伺服器~~。您只要新增額外的伺服器，即可達到擴充叢集的目的，並可確保儲存體的輸入/輸出容量會按比例增加，而且 PowerHA pureScale 伺服器 LPAR 的容量也會逐步增加。

除了在額外的伺服器上新增額外的 LPAR 及成員，這個組態的部署方式，和部署三部伺服器的部署方式相同。當 DB2 pureScale 成員及 PowerHA pureScale 伺服器使用專用的 HCA/HBA 時，您也可以在每部伺服器上部署一個 LPAR。圖 4 顯示這種組態。

圖 4：透過纜線連接四部以上 POWER 伺服器的組態



結論

IBM DB2 pureScale 功能與 IBM POWER 伺服器提供緊密耦合的解決方案，可滿足業務成長及連續可用性的需求。本篇文章說明利用業界標準元件所建立的各種範例部署模式。各種這些部署模式顯示了保有說明了一個彈性的基礎建設架構，最少可以從基本的 2 個成員的叢集開始，最多可達 128 個成員的叢集，因此能滿足各種企業需求。

資源

瞭解

- 在 IBM 叢集資訊中心(Cluster Information Center)的一般平行檔案系統常見問題中獲得更多 GPFS 的相關資訊。
- 參閱 DB2 for Linux、UNIX 及 Windows Information Center 以獲得 DB2 pureScale 功能的相關資訊。
- 閱讀「InfiniBand 使用方式」以獲得更多在 IBM POWER 伺服器中使用 InfiniBand 的相關資訊。
- 瀏覽「IBM HMC」以獲得 IBM 硬體管理主控台(HMC)的完整資訊。
- 查看「IBM Qlogic」以獲得更多 IBM Qlogic IB 交換器的相關資訊。
- 於 developerWorks Information Management 專區進一步瞭解 Information Management。尋找技術文件、說明文章、教育、下載、產品資訊等。
- 隨時掌握最新的 developerWorks 技術活動及網路廣播。
- 上 Twitter 隨時瞭解 developerWorks。

取得產品與技術

- 利用 IBM 試用版軟體建立您的下一個開發專案(您可以直接從 developerWorks 下載試用版軟體)。

討論

- 參加與此內容相關的討論論壇。
- 查看 developerWorks 部落格並參與 developerWorks 網路社群。

作者簡介

Miso Cilimdzic

Miso 自 2000 年起即任職於 IBM , 負責處理各種 DB2 效能相關的活動 , 最近則專注在 DB2 pureScale 的事務。

Sanjeeva Kumar Ogirala

Sanjeeva Kumar Ogirala 是一位 DB2 效能團隊的軟體工程師。他是印度理工學院德里分校(IIT Delhi)的研究生，主要進行電力系統 M.Tech 的研究。他自 2007 年 7 月起即任職於 IBM，擁有 IBM 頒發的 DB2 for Linux、UNIX 及 Windows 資料庫管理員的認證。