Sun Cluster 3.0 概念

Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto,, CA 94303-4900
U.S.A. 650-960-1300

元件號碼： 806-6722
2000 年 11 月, Revision A

Adobe PostScript

內容

前言

Sun™ Cluster 3.0 概念包含 Sun Cluster 軟體的概念性與參考資訊。

本文件適合對於 Sun 軟體與硬體有廣泛瞭解的有經驗系統管理者。請不要將本文件當做規劃作業或售前指引。您應該已經先決定您的系統需求，並購買了適當的設備與軟體，再閱讀本文件。

要了解本書所說明的概念，您應該具備 Solaris™ 作業環境的知識，以及使用於 Sun Cluster 的容體管理者軟體的技術。

印刷習慣用法

字體或符號	意義	範例
`AaBbCc123`	指令、檔案和目錄的名稱；電腦螢幕的輸出	編輯您的 `.login` 檔案。 使用 `ls -a` 列出所有檔案。 `% You have mail.`
`AaBbCc123`	您鍵入的內容，與電腦螢幕上的輸出作爲對照	`% su` `Password:`

字體或符號	意義	範例
AaBbCc123	書名、新字或專有名詞，以及要強調的字	請閱讀使用手冊的第六章。 這些稱爲 *class* 選項。 您必須是高階使用者才能執行此項操作。
	指令行變數；以實際名稱或數值取代	若要刪除檔案，請鍵入 `rm` *filename*。

Shell 提示符號

Shell	提示符號
C shell	*machine_name*`%`
C shell 高階使用者	*machine_name*`#`
Bourne shell 和 Korn shell	`$`
Bourne shell 和 Korn shell 超級使用者	`#`

相關文件

主題	標題	組件號碼
安裝	*Sun Cluster 3.0* 安裝手冊	806-6728
硬體	*Sun Cluster 3.0 Hardware Guide*	806-1420
數據服務	*Sun Cluster 3.0 Data Services Installation and Configuration Guide*	806-1421

主題	標題	組件號碼
API 設計	*Sun Cluster 3.0 Data Services Developers' Guide*	806-1422
管理	*Sun Cluster 3.0* 系統管理手冊	806-6734
錯誤訊息與問題解決方法	*Sun Cluster 3.0 Error Messages Manual*	806-1426
版本注意事項	*Sun Cluster 3.0* 版次注意事項	806-6738

訂購 Sun 文件資料

Fatbrain.com 是一個 Internet 專業書店，其中備有精選之 Sun Microsystems, Inc. 產品文件資料。有關文件清單和訂購方式，可以在 Fatbrain.com 上的 Sun Documentation Center 取得說明，網址爲：

`http://www1.fatbrain.com/documentation/sun`

線上存取 Sun 文件資料

`docs.sun.com`SM 網站可讓您存取 Sun 在網路上的技術文件。您可以瀏覽 `docs.sun.com` 文件，或者搜尋特定的書名或主題，網址爲：

`http://docs.sun.com`

取得協助

如果在安裝或使用 Sun Cluster 上有問題，請聯絡您的服務供應商並提供下列資訊：

- 您的姓名和電子郵件地址 (如果有的話)
- 您的公司名稱、地址和電話號碼
- 您系統的機型和序號

- 作業環境的版次號碼 (例如，Solaris 8)
- Sun Cluster 的版次號碼 (例如，Sun Cluster 3.0)

使用下列指令收集您系統上每一個節點的相關資訊，提供給您的服務供應商：

指令	功能
`prtconf -v`	顯示系統記憶體的大小及報告周邊裝置的相關資訊
`psrinfo -v`	顯示處理器的相關資訊
`showrev --p`	報告安裝了哪些修補程式
`prtdiag -v`	顯示系統偵錯資訊
`scinstall -pv`	顯示 Sun Cluster 版次和套裝軟體版本資訊

並提供 `/var/adm/messages` 檔案的內容。

簡介與概觀

Sun Cluster 3.0 概念提供 Sun Cluster 文件主要讀者群所需具備的概念資訊。這些讀者包括：

- 安裝與維修叢集硬體的服務供應商
- 安裝、配置和管理 Sun Cluster 軟體的系統管理者
- 開發目前 Sun Cluster 產品所未包含的應用程式數據服務的應用程式開發人員

本書配合其餘的 Sun Cluster 文件集，提供 Sun Cluster 的完整概觀。

本章：

- 提供 Sun Cluster 的簡介和高層次的概觀
- 說明 Sun Cluster 讀者的各種觀點
- 指出在使用 Sun Cluster 之前需要瞭解的重要概念
- 對應重要概念至包括程序與相關資訊的 Sun Cluster 文件
- 對應叢集相關作業至包含用來完成那些作業之程序的文件

Sun Cluster 簡介

Sun Cluster 將 Solaris™ 作業環境延伸成爲叢集作業系統。叢集是一組鬆散式結合的運算節點，提供網路服務或應用程式的單一用戶端觀點，包括資料庫、網路服務和檔案服務。

每一個叢集節點均爲一個獨立的伺服器，可執行其本身的處理程序。這些處理程序可以互相通訊，形成如同（對一個網路用戶端）共同將應用程式、系統資源和資料提供給使用者的單一系統。

叢集可提供比傳統單一伺服器系統更多項的優點。這些優點包括支援可用性和可延伸性極高的應用程式、模組成長的能力，以及導入成本比傳統硬體容錯系統低。

Sun Cluster 的目標是：

- 減少或免除因爲軟體或硬體失效所造成的當機時間
- 確保對一般使用者的資料和應用程式可用性，不論是否出現一般會使單一伺服器系統當機的那種失效
- 增加節點至叢集，讓服務延伸至額外的處理器，而增加應用程式產量
- 讓您可以執行維護作業而不需要關閉整個系統，提供強化的系統可用性

Sun Cluster 的高度可用性

Sun Cluster 被設計成高可用性 (HA) 系統，亦即可提供幾近連續的資料和應用程式存取的系統。

相形之下，容錯*(fault-tolerant)* 硬體系統提供持續的資料和應用程式存取，但是因爲硬體特殊，所以成本較高。此外，容錯系統通常不會說明軟體失效。

Sun Cluster 透過硬體和軟體的結合來達到高可用性。多餘備用性的叢集交互連接、儲存體和 公用網路可防止發生單一失效點。叢集軟體持續監督成員節點的健康狀況，並阻止失效節點參與叢集， 以免資料遭到毀損。此外，叢集會監督應用程式與其相依系統資源，以及在發生失效時進行失效保護或重新啓動應用程式。

請參照 第57頁的「高可用性常問問題」 以取得關於高可用性的問題與解答。

Sun Cluster 的失效保護和延伸性

Sun Cluster 可讓您建立失效保護 *(failover)* 或延伸性 *(scalable)* 式的應用程式。 失效保護和可延伸式應用程式也可以並行於同一叢集上執行。一般而言，失效保護應用程式提供高可用性 (多餘備用性)，而可延伸式應用程式則提供高可用性以及增加效能。單一叢集可以同時支援失效保護和可延伸應用程式。

失效保護

失效保護是叢集將已失效之主要節點上的應用程式，自動重新放置於指定之次要節點的處理程序。利用失效保護，Sun Cluster 提供了高可用性。

當發生失效保護時，用戶端可能會看到短暫的服務中斷，以及可能需要在完成失效保護動作之後重新連線。然而，用戶端不會察覺提供應用程式和資料的實體伺服器。

延伸性

失效保護與多餘備用性有關，而延伸性則提供不變的回應時間或產量，與負載無關。可延伸應用程式調整叢集中的多個節點來並行執行應用程式，因此提供了較佳的效能。在可延伸配置中，叢集的每個節點均可提供資料和處理用戶端要求。

請參照 第46頁的「數據服務」 以取得有關失效保護和可延伸服務的更多詳細資訊。

三種 Sun Cluster 概觀

本節說明 Sun Cluster 的三種不同觀點和重要概念，以及每個觀點的相關文件。 這些觀點是來自：

- 硬體安裝與維修人員
- 系統管理者
- 應用程式程式設計師。Sun Cluster 提供一組高可用性的數據服務。這些服務是已配置成爲在叢集上執行的高可用性數據服務的應用程式，如 Oracle、Apache Web Server 和 DNS。其它的應用程式可以使用 Sun Cluster API 變成高可用性的數據服務。應用程式程式設計師可以撰寫使用 API 的 shell 指令集或 C 程式。

硬體安裝和維修觀點

對於硬體維修人員而言，Sun Cluster 就像是一組常用的硬體，包括伺服器、網路和儲存體。這些元件全部以電纜連接在一起，使得每一個元件均具有備份而不會有單一故障點。

重要概念 – 硬體

硬體維修人員需要瞭解下列的叢集概念。

- 叢集硬體配置和電纜安裝
- 安裝和維修 (新增，移除，更換)：
  - 網路介面元件 (配接卡，連接，電纜)
  - 磁碟介面卡
  - 磁碟陣列
  - 磁碟機
  - 管理主控台和主控台存取裝置
- 設定管理主控台和主控台存取裝置

建議的硬體概念參考文件

下列各節包含前述重要概念的相關資料：

- 第20頁的「叢集節點」
- 第22頁的「多主機磁碟」
- 第23頁的「區域磁碟」
- 第24頁的「叢集交互連接」
- 第24頁的「公用網路介面」
- 第25頁的「用戶端系統」
- 第25頁的「管理主控台」
- 第26頁的「主控台存取裝置」
- 第26頁的「叢集化配對拓樸架構」
- 第28頁的「N+1 (星狀) 拓樸」

相關的 Sun Cluster 文件

下列的 Sun Cluster 文件包括與硬體維修概念相關的程序和資訊：

- *Sun Cluster 3.0 Hardware Guide*

系統管理者觀點

對於系統管理者而言，Sun Cluster 就像是一群以電纜連接在一起的伺服器 (節點)，共用儲存裝置。系統管理者看見：

- 用以監督叢集節點之間的連接、與 Solaris 軟體整合的專用叢集軟體
- 用以監督執行於叢集節點的使用者應用程式執行狀況的專用軟體
- 設定和管理磁碟的容體管理軟體
- 讓所有節點可以存取所有儲存裝置 (即使是未直接連接的磁碟) 的專用叢集軟體
- 讓檔案以像是本端連接於該節點的方式，出現於每個節點上

重要概念 – 系統管理

系統管理者需要瞭解下列的概念和程序：

- 硬體和軟體元件之間的交談
- 如何安裝和配置叢集的一般流程，包括：
  - 安裝 Solaris 作業環境
  - 安裝和配置 Sun Cluster
  - 安裝和配置容體管理者
  - 安裝和配置應用軟體成爲具備叢集功能
  - 安裝和配置 Sun Cluster 數據服務軟體
- 新增、移除、更換和維修叢集硬體與軟體元件的叢集管理程序
- 修改配置以增進效能

建議的系統管理者概念參考文件

下列各節包含前述重要概念的相關資料：

- 第33頁的「管理介面」
- 第33頁的「高可用性的組織架構」
- 第36頁的「整體裝置」
- 第37頁的「磁碟裝置群組」
- 第38頁的「整體名稱空間」
- 第40頁的「叢集檔案系統」
- 第42頁的「法定人和法定裝置」
- 第45頁的「容體管理者」

- 第46頁的「數據服務」
- 第54頁的「資源與資源類型」
- 第55頁的「公用網路管理 (PNM) 和網路配接卡失效保護 (NAFO)」
- 第 4 章

相關的 Sun Cluster 文件 – 系統管理者

下列的 Sun Cluster 文件包含與系統管理概念相關的程序和資訊：

- *Sun Cluster 3.0* 安裝手冊
- *Sun Cluster 3.0* 系統管理手冊
- *Sun Cluster 3.0 Error Messages Manual*

應用程式程式設計師觀點

Sun Cluster 提供多種高可用性數據服務給應用程式，如 Oracle、NFS、DNS、iPlanet Web Server、Apache Web Server 與 Netscape Directory Server。如果某個站台必須使其它的應用程式在叢集上執行，可以使用「Sun Cluster 應用程式設計介面 (API)」和 Data Service Development Library API (DSDL API) 來開發必需的數據服務軟體，使其應用程式可以在叢集上執行，成爲高可用性的數據服務。

重要概念 – 應用程式程式設計師

應用程式程式設計師需要瞭解下列各項：

- 本身應用程式的性質，以判斷是否能夠執行爲具備高可用性或可延伸的數據服務。
- Sun Cluster API、DSDL API 與 “generic” 數據服務。程式設計師需要決定哪一種工具最適合用來撰寫程式或指令集，以配置其應用程式使用於叢集環境。

建議的應用程式程式設計師概念參考文件

下列各節包含前述重要概念的相關資料：

- 第46頁的「數據服務」
- 第54頁的「資源與資源類型」
- 第 4 章

相關的 Sun Cluster 文件 – 應用程式程式設計師

下列的 Sun Cluster 文件包含與應用程式程式設計師概念相關的程序和資訊：

- *Sun Cluster 3.0 Data Services Developers' Guide*
- *Sun Cluster 3.0 Data Services Installation and Configuration Guide*

Sun Cluster 作業

所有的概念對應至作業和所有作業需要一些概念知識背景。下列的表格提供了作業與說明作業步驟之文件的概觀。本書中的概念章節說明概念如何對應至這些作業。

表格1–1 工作對應：將使用者工作對應到文件

要執行此作業...	使用此文件...
安裝叢集硬體	*Sun Cluster 3.0 Hardware Guide*
安裝 Solaris 軟體於叢集	*Sun Cluster 3.0* 安裝手冊
安裝 Sun™ Management Center 軟體	*Sun Cluster 3.0* 安裝手冊
安裝和配置 Sun Cluster 軟體	*Sun Cluster 3.0* 安裝手冊
安裝和配置容體管理軟體	*Sun Cluster 3.0* 安裝手冊 您的容體管理文件
安裝和配置 Sun Cluster 數據服務	*Sun Cluster 3.0 Data Services Installation and Configuration Guide*
維修叢集硬體	*Sun Cluster 3.0 Hardware Guide*
管理 Sun Cluster 軟體	*Sun Cluster 3.0* 系統管理手冊
管理容體管理軟體	*Sun Cluster 3.0* 系統管理手冊 和您的容體管理文件
管理應用軟體	您的應用程式文件

表格1–1 工作對應：將使用者工作對應到文件 *(續上)*

要執行此作業...	使用此文件...
問題辨別與建議的使用者動作	*Sun Cluster 3.0 Error Messages Manual*
建立新的數據服務	*Sun Cluster 3.0 Data Services Developers' Guide*

重要概念 – 硬體服務供應商

本章說明有關 Sun Cluster 配置的硬體元件的重要概念。

Sun Cluster 硬體元件

本資訊主要是針對硬體服務供應商。這些概念可以協助服務供應商在安裝、配置或維修叢集硬體之前，瞭解各硬體元件之間的關係。叢集系統管理者也會發現，這項資訊對於安裝、配置和管理叢集軟體很有用。

叢集是許多硬體元件所組成，包括：

- 具有本端磁碟 (未共用) 的叢集節點
- 多重主機儲存 (節點之間共用磁碟)
- 抽換式媒體 (磁帶和 CD-ROM)
- 叢集交互連接
- 公用網路介面
- 用戶端系統
- 管理主控台
- 主控台存取裝置

Sun Cluster 可讓您將這些元件結合成各種的配置，說明於 第26頁的「Sun Cluster 拓樸」。

下圖顯示範例叢集配置。

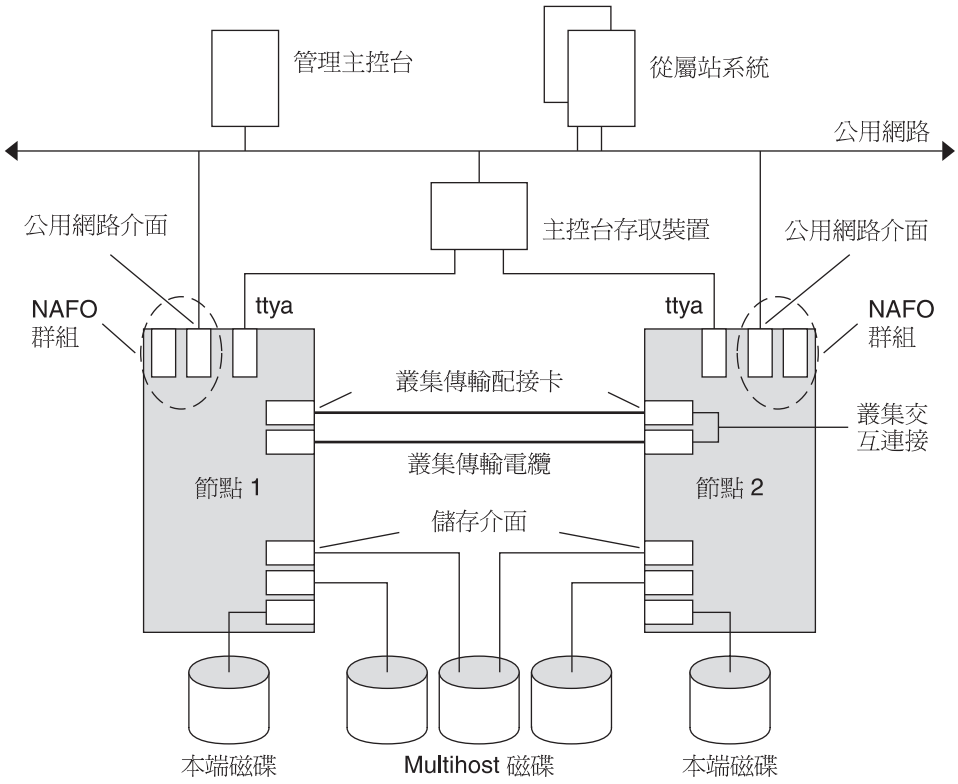

圖 2–1　兩個節點的叢集配置範例

叢集節點

叢集節點是執行 Solaris 作業環境和 Sun Cluster 軟體的機器，也是叢集的目前成員 (*cluster member*) 或潛在成員。 Sun Cluster 軟體可讓您在一個叢集中有二到八個節點。請參閱 第26頁的「Sun Cluster 拓樸」 以取得所支援的節點配置。

叢集節點均連接到一或多個多主機磁碟。可延伸的服務配置可讓節點能服務未直接連接到多重主電腦磁碟的要求。未連接到多重主電腦磁碟的節點，是使用檔案系統來存取多重主電腦磁碟。

在平行資料庫配置中，節點共用並行存取所有的磁碟。請參閱 第22頁的「多主機磁碟」 和 第 3 章 以取得平行資料庫配置的其他資訊。

叢集中的所有節點會依照一般名稱來分類—叢集名稱—用來存取和管理叢集。

公用網路配接卡連接節點到公用網路，提供用戶端存取叢集。

叢集成員透過一或多個實際獨立的網路 (稱爲 私有網絡) 與叢集的其它節點通訊。 這組叢集中的私有網路稱爲 *cluster interconnect*。

當另一個節點加入或離開叢集時，叢集中的每個節點都會知道。此外，叢集中的每個節點 也都知道本端執行的資源以及在其它叢集節點上執行的資源。

配置叢集成員的資源 (應用程式、磁碟儲存體等等)，使其能夠提供失效保護及 / 或可延伸功能。

確定相同叢集中的節點有類似的處理程序、記憶體和 I/O 能力，以便啓動失效保護，而不至於大幅降低效能。 因爲可能發生失效保護，請確定每個節點有足夠的額外容量，可以接管所有節點的工作負荷，作爲備份或次要。

每一個節點啓動其自己的個別 root (/) 檔案系統。

叢集成員的軟體元件

要作爲叢集成員，必須安裝下列的軟體：

- Solaris 作業環境
- Sun Cluster
- 容體管理 (Solstice DiskSuite™ 或 VERITAS 容體管理者)
- 數據服務應用程式

一種例外情形是在使用硬體多餘備用獨立磁碟陣列 (RAID) 的 Oracle Parallel Server(OPS) 配置中。這種配置不需要軟體容體管理者，如 Solstice DiskSuite 或 VERITAS 容體管理者 以便來管理 Oracle 資料。

請參閱 *Sun Cluster 3.0* 安裝手冊 以取得有關如何安裝 Solaris 作業環境、Sun Cluster 和 容體管理軟體的資訊。請參閱 *Sun Cluster 3.0 Data Services Installation and Configuration Guide* 以取得有關如何安裝和配置數據服務的資訊。

請參閱 第 3 章 以取得前述軟體元件的概念資訊。

下圖提供共同運作以建立 Sun Cluster 軟體環境之軟體元件的高階觀點。

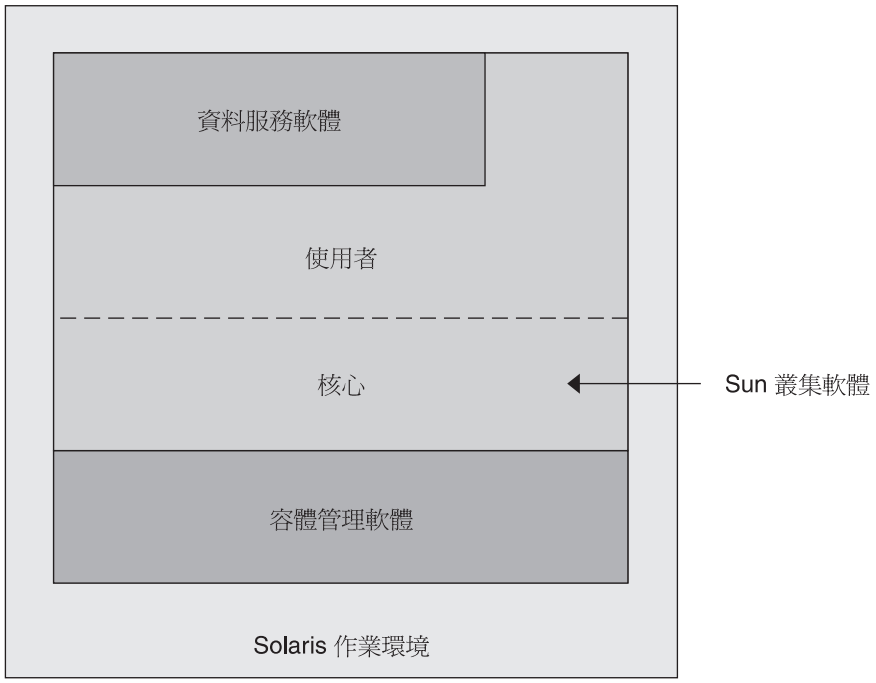

圖 2–2　Sun Cluster 軟體元件的高階關係

請參閱 第 4 章 以取得有關叢集成員的問題與解答。

多主機磁碟

Sun Cluster 需要多主機磁碟儲存體：可以一次連接至超過一個節點的磁碟。在 Sun Cluster 環境中，多重主電腦磁碟可讓磁碟裝置非常有用。位於多主機儲存體上的磁碟裝置可以承受單一節點失效。

多主機磁碟儲存應用資料，也可以儲存數據服務二進位檔案和配置檔。

多主機磁碟是透過「主控」磁碟的主要節點來全域存取，或透過區域路徑直接並行存取。目前使用直接並行存取的唯一應用程式是 OPS。

多主機磁碟可以防止節點失效。如果用戶端要求是透過某個節點來存取資料而該節點失效，這些要求會切換爲使用另一個可直接連接同一磁碟的節點。

容體管理者提供鏡映或 RAID-5 配置的多主機磁碟資料多餘備用性。目前，Sun Cluster 支援 Solstice DiskSuite 和 VERITAS 容體管理者 作爲容體管理者，以及 Sun StorEdge™ A3x00 儲存單位中的 RDAC RAID-5 硬體控制器。

結合多主機磁碟和磁碟映射與資料分置，可以防止節點失效和個別的磁碟失效。

請參閱 第 4 章 以取得有關多主機儲存體的問題與解答。

多重起始者 SCSI

本節僅適用於 SCSI 儲存裝置，不適用於多主機磁碟的「光纖通道 (Fibre Channel)」儲存體。

在獨立式伺服器中，伺服器節點是以連接此伺服器至特定 SCSI 匯流排的 SCSI 主機配接卡電路，來控制 SCSI 匯流排活動。此 SCSI 主機配接卡電路即爲 *SCSI initiator*。 這個電路起始此 SCSI 匯流排的所有匯流排活動。SCSI 主機配接卡的預設 SCSI 位址在 Sun 系統中是 7。

叢集配置在多重伺服器節點之間共用記憶體。當叢集儲存體是由單端或差動式 SCSI 裝置所組成時，該配置即爲多重起始者 SCSI。依照這個詞彙所衍生的意義，即 SCSI 匯流排上存在一個以上的 SCSI 起始者。

SCSI 規格需要 SCSI 匯流排上的每一個裝置均具有一個唯一的 SCSI 位址。 (主機配接卡也是 SCSI 匯流排上的一個裝置。) 在多重起始者環境中的預設硬體配置會導致衝突， 因爲所有的 SCSI 主機配接卡預設爲 7。

若要解決衝突，在每個 SCSI 匯流排上，留下其中一個 SCSI 主機配接卡的 SCSI 位址是 7， 並將其它的主機配接卡設定爲未用的 SCSI 位址。請適當地規劃指定這些「未用的」SCSI 位址，包括 目前和最後未使用的位址。未來未使用的位址範例，是安裝新磁碟到空磁碟插槽以便增加儲存體。 在大部份配置中，第二主機配接卡的可用 SCSI 位址是 6。

您可以藉由設定 `scsi-initiator-id` Open Boot PROM (OBP) 性質， 變更選取的主機配接卡的 SCSI 位址。您可以全域式或以個別主機配接卡的方式，來設定節點的這個性質。 設定每一個 SCSI 主機配接卡的唯一 `scsi-initiator-id`，其指示包含在 *Sun Cluster 3.0 Hardware Guide* 中各磁碟外殼的章節。

區域磁碟

區域磁碟是僅連接至單一節點的磁碟。因此，沒有節點失效的保護 (不具高可用性)。然而，所有的磁碟 (包括區域磁碟) 均含括於整體名稱空間中，並且配置爲 整體裝置。因此，從所有的叢集節點可以看到磁碟本身。您可以將這些磁碟上的檔案系統放在整體裝載點下，讓 其它節點使用。如果目前裝載這些整體檔案系統之其中一個檔案系統的節點失效， 所有節點均會遺失該檔案系統的存取。使用容體管理者可讓您鏡映這些磁碟，如此磁碟失效就不會導致 這些檔案系統變成無法存取，但是容體管理者不能防止節點失效。

抽換式媒體

叢集中支援如磁帶機和 CD-ROM 光碟機的抽換式媒體。一般而言，您安裝、配置和維修這些裝置的 方式與在非叢集環境的方式相同。這些裝置是配置爲 Sun Cluster 中的整體裝置，所以每一個裝置均可自叢集的任何節點來存取。請參照 *Sun Cluster 3.0 Hardware Guide* 以取得安裝和配置抽換式媒體的資訊。

叢集交互連接

此項 *cluster interconnect* 是用來傳輸叢集節點之間的叢集私有通訊與數據服務通訊的實體裝置配置。因爲交互連接廣泛使用於叢集私有通訊，所以會限制效能。

只有叢集節點可以連接至私有交互連接。Sun Cluster 安全性模型假設只有叢集節點具有實體存取私有交互連接。

必須透過至少兩個多餘備用私有網絡或路徑，藉由叢集交互連接來連接所有的節點，才能避免單一失效點的情形。任何兩個節點之間可以有多個私有網路 (二到六個)。叢集交互連接由三個硬體元件所組成：配接卡、接點和電纜。每一個私有網路的配置，會使其不會與任何其它私有網路共用共同的硬體元件。

下表說明各個硬體元件。

- 配接卡 – 位於每個叢集節點的實體網路卡。其名稱是來自產品的名稱，例如，qfe 代表 Quad FastEthernet。某些配接卡只有一個實體網路連接，但是有些配接卡 (如 qfe 卡) 則會有多重實體連線。部份網路卡還包含網路介面和儲存介面。

  具有多重介面的網路卡在整個卡失效時會變成單一失效點。爲了有最大的可用性，請規劃您的叢集，使兩個節點之間的唯一路徑不會依賴單一網路卡。

- 接點 – 位於叢集節點之外的轉換開關。執行透通和轉換功能，讓您將兩個以上的節點連接在一起。在兩個節點的叢集中，您不需要接點，因爲透過多餘備用實體電纜連接至每個節點上的多餘備用配接卡，節點可以直接彼此連接。大於兩個節點的配置一般會需要接點。

- 電纜 – 在兩個網路配接卡之間或配接卡與接點之間的實體連線。

請參閱 第 4 章 以取得有關叢集交互連接的問題與解答。

公用網路介面

用戶端透過公用網路介面連接至叢集。每一個網路配接卡可以連接至一或多個公用網路， 根據配接卡是否有多重硬體介面而定。您可以設定節點來包含多個配置的公用網

路介面卡，如此一來，一個介面卡在作用中，其它介面卡就作爲備用。Sun Cluster 軟體有一個子系統稱爲“「公用網路管理」” (PNM) ，可監督作用中的介面。如果作用中配接卡失效，會呼叫「網路配接卡失效保護 (NAFO)」軟體，將介面移轉至備用配接卡。

公用網路介面的叢集不需要特別的硬體注意事項。

請參閱 第 4 章 以取得有關公用網路的問題與解答。

用戶端系統

用戶端系統包括工作站或透過公用網路存取叢集的其它伺服器。用戶端程式使用由伺服器端應用程式執行於叢集所提供的資料或其它服務。

用戶端系統不具高可用性。叢集上資料和應用程式則具高可用性。

請參閱 第 4 章 以取得有關用戶端系統的問題與解答。

管理主控台

您可以使用專用的 SPARCstation™ 系統，即管理主控台，來管理作用中的叢集。 通常，您在管理主控台上所安裝和執行管理工具軟體，會像是 Sun Management Center 產品的「叢集控制台 (CCP)」和 Sun Cluster 模組。使用 CCP 下的 `cconsole` 可讓您一次連接一個以上的節點主控台。 如果需要使用 CCP 的其他資訊，請參閱 *Sun Cluster 3.0* 系統管理手冊。

管理主控台不是叢集節點。您使用管理主控台，透過公用網路或選擇透過網路型終端機集線器，來遠端存取叢集節點。如果您的叢集是由 Sun™ Enterprise E10000 平台所組成，您必須能夠從管理主控台登入「系統服務處理器 (SSP)」使用 `netcon(1M)` 指令連接。

一般而言，您配置沒有監視器的節點。然後，您透過管理主控台上的 `telnet` 階段作業 來存取節點的主控台，管理主控台連接至終端機集線器，以及從終端機集線器至節點的串列埠。 (如果是 Sun Enterprise E10000 server，您是從「系統服務處理器」連接。) 請參閱 第26頁的「主控台存取裝置」 以取得其他資訊。

Sun Cluster 不需要專用的管理主控台，但是使用專用主控台可以有以下優點：

- 在同一機器上將主控台和管理工具分組，以達到中央化叢集管理
- 由您的硬體服務供應商提供可較快速解決問題的方法

請參閱 第 4 章 以取得有關管理主控台的問題與解答。

主控台存取裝置

您必須可以主控台存取所有的叢集節點。要取得主控台存取，請使用向您的叢集硬體購買的 終端機集線器、Sun Enterprise E10000 server 伺服器上的「系統服務處理器 (SSP)」或是可以存取每個節點上 `ttya` 的其它裝置。

從 Sun 只能取得一個支援的終端機集線器。使用支援的 Sun 終端機集線器是可選用的。 終端機集線器允許使用 TCP/IP 網路來存取每一個節點上的 `ttya`。 結果是從網路上任意位置的遠端工作站，以主控台層次存取每一個節點。

「系統服務處理器 (SSP)」提供主控台存取 Sun Enterprise E10000 server。 SSP 是 Ethernet 上的 SPARCstation 系統，配置爲支援 Sun Enterprise E10000 server。 SSP 是 Sun Enterprise E10000 server 的管理主控台。使用「Sun Enterprise E10000 網路主控台」功能，網路上的任何工作站皆可開啓主機主控台階段作業。

其它的主控台存取方法包括其它終端機集線器，`tip(1)` 從另一個節點和沈默式終端機的串列埠存取。您可以使用 Sun™ 鍵盤和監視器，或其它串列埠裝置 (如果您的硬體服務供應商支援這些裝置)。

請參閱 第 4 章 以取得有關主控台裝置的問題與解答。

Sun Cluster 拓樸

拓樸是指連接叢集節點與叢集使用之儲存體平台的連接機制。

Sun Cluster 支援下列拓樸架構：

- 叢集化配對
- N+1 (星狀)

以下各節說明每一種拓樸架構。

叢集化配對拓樸架構

叢集化配對拓樸架構是二個或以上的節點配對，在單一叢集管理組織架構之下運作。在此配置中，失效保護僅發生於配對之間。然而，所有的節點以私有網路連接，並在 Sun Cluster 軟體控制下運作。您可能使用這種拓樸架構，在某個配對上執行平行資料庫應用程式， 而在另一個配對上執行高可用性應用程式。利用叢集檔案系統，您也可以讓兩個配對的配置， 其中有兩個以上的節點執行可延伸服務或平行資料庫，即使所有的節點均未直接連接儲存應用資料 的磁碟。

下圖說明叢集化配對配置。

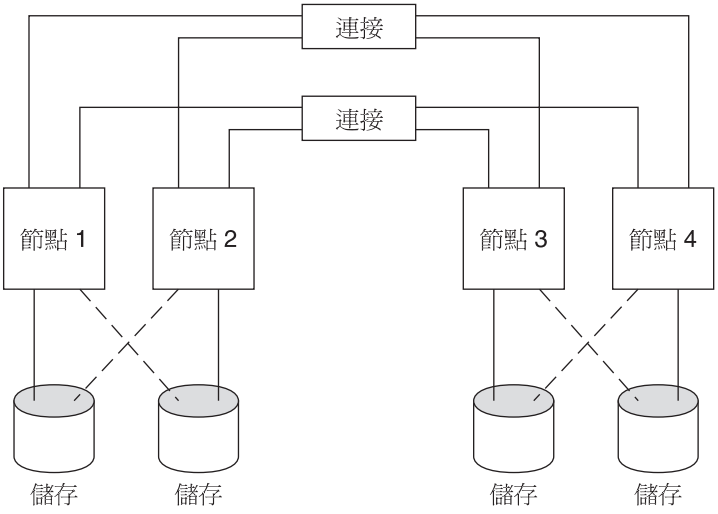

圖 2–3　叢集化配對拓樸架構

Pair+M 拓樸

此項 pair+M 拓樸中包含一對直接連接共用儲存體的節點與附加節點組，並使用叢集交互連接來存取共用儲存體 —其本身並未具備直接連接。在此配置中所有的節點仍然以容體管理者來加以配置。

下圖說明 pair+M 拓樸，其中四個節點的兩個 (節點 3 和節點 4) 使用叢集交互連接來存取儲存體。此項配置可加以擴展，以便納入其他並未具有可直接存取共用儲存體的節點。

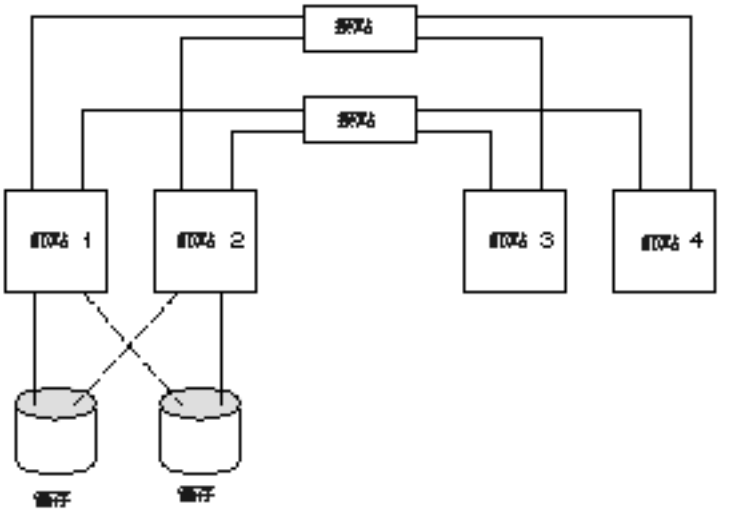

圖 2–4　Pair+M 拓樸

N+1 (星狀) 拓樸

N+1 拓樸架構包括一些主要節點和一個次要節點。您不需要配置相同的主要節點和次要節點。 主要節點主動地提供應用程式服務。等待主要節點失效時，次要節點不需要閒置。

次要節點在配置中是唯一實際連接至所有多主機儲存體的節點。

如果主要節點上發生失效，Sun Cluster 會移轉資源至次要節點繼續運作，直到轉換 (自動或手動) 回到主要節點爲止。

次要節點必須時常保有足夠的額外 CPU 容量，以便在主要節點之一失效時處理負載。

下圖說明 N+1 配置。

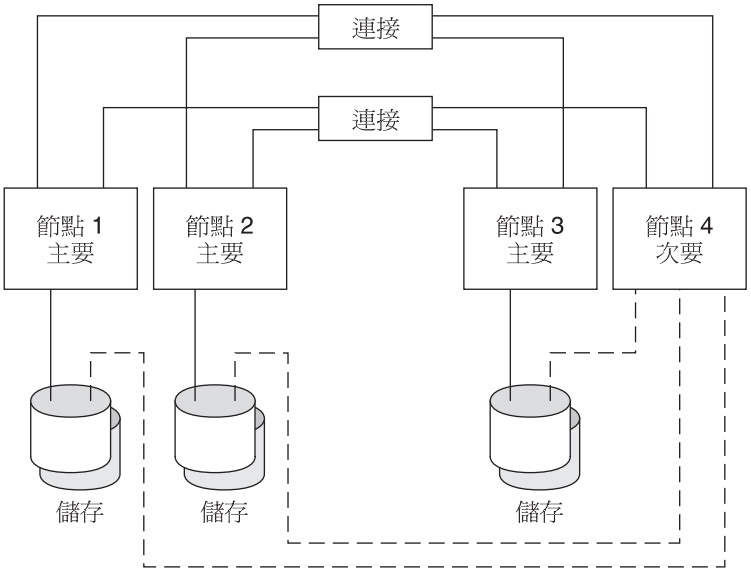

圖 2–5　N+1 拓樸架構

第 3 章

重要概念 – 管理和應用程式設計

本章說明有關 Sun Cluster 配置的軟體元件的重要概念。涵蓋的主題包含：

叢集管理與應用程式設計

這項資訊主要是給使用 Sun Cluster API 和 SDK 的系統管理者和應用程式開發人員參考。叢集系統管理者可以利用這些資訊來輔助安裝、配置和管理叢集軟體。應用程式開發人員可以使用這些資訊來瞭解將要利用的叢集環境。

下圖顯示了叢集管理概念如何對應至叢集架構的高階觀點。

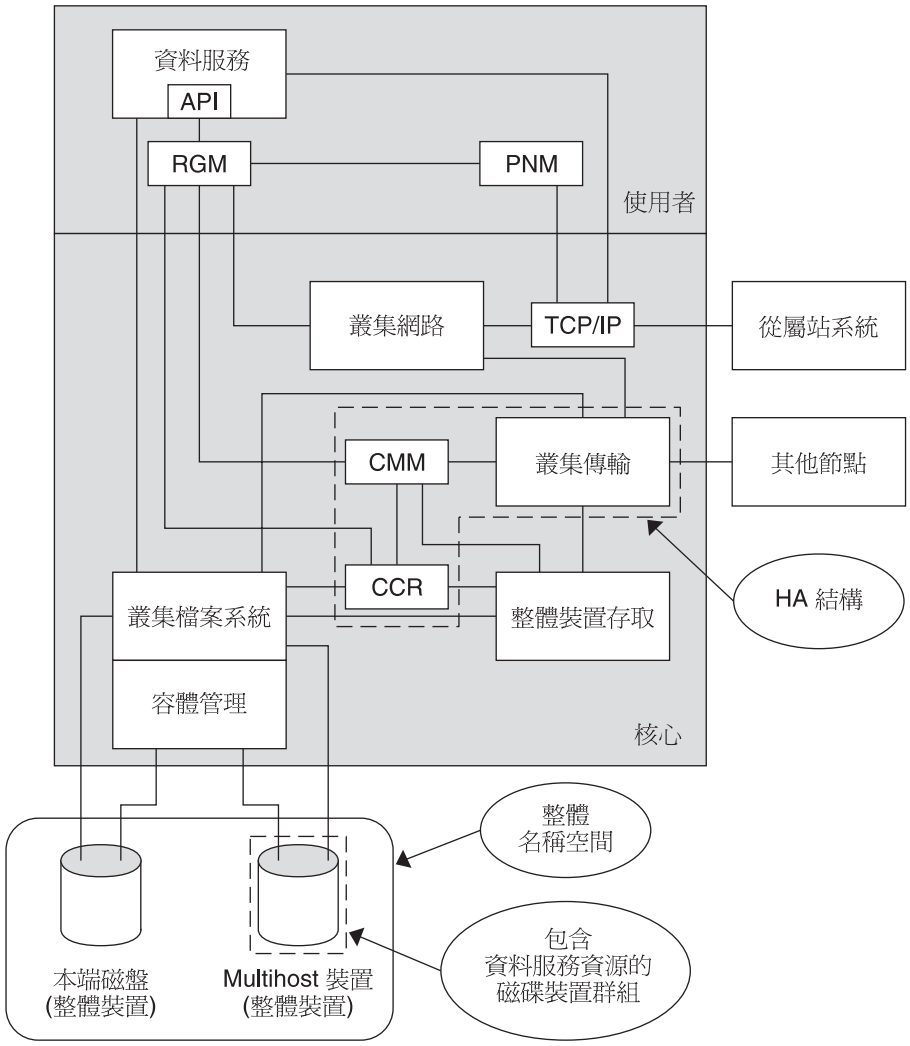

圖 3–1　Sun Cluster 軟體架構

管理介面

您可以從多種使用者介面選擇，以便安裝、配置和管理 Sun Cluster 與 Sun Cluster 數據服務。您可以透過具備說明的指令行介面來完成系統管理作業。在指令行介面頂端是可以簡化選取配置作業的部份公用程式集。Sun Cluster 也具有一個模組，此模組可作爲 Sun Management Center 的一部份來執行，以提供 GUI 給某些叢集作業。請參照 *Sun Cluster 3.0* 系統管理手冊 中的介紹章節以取得管理介面的完整說明。

叢集時間

叢集中所有節點的時間均必須同步。您是否以任何外來時間來源同步化叢集節點，對叢集作業並不重要。Sun Cluster 使用網絡時間協定 (NTP) 來同步化節點間的時鐘。

一般而言，系統時鐘在傾刻之間變更並不會造成問題。然而，如果您在作用中的叢集上執行 `date(1)`、`rdate(1M)`，或 `xntpdate(1M)`（交談式，或在 `cron` 指令集之內），您可以強制進行比傾刻更久的時間變更來同步化系統時鐘與時間來源。這種強制變更可能會導致檔案修改時間戳記有問題或混淆 NTP 服務。

當您在每一個叢集節點上安裝 Solaris 作業環境時，您有機會變更節點的預設時間及日期設定。一般而言，您可以接受出廠預設值。

當您使用 `scinstall(1M)` 來安裝 Sun Cluster 時，程序中的一個步驟是對叢集配置 NTP。 Sun Cluster 提供一個範本檔案，`ntp.cluster`（請參閱已安裝之叢集節點上的 `/etc/inet/ntp.cluster`），該檔案建立了所有叢集節點之間的對等關係，以某一個節點作爲“偏好的” 節點。由專用的主電腦名稱和跨叢集交互連接時發生的時間同步化來識別節點。關於如何配置 NTP 的叢集，已納入 *Sun Cluster 3.0* 安裝手冊。

另外一種方式是，您可以在叢集之外設定一或多部 NTP 伺服器，並變更 `ntp.conf` 檔案以反映該配置。

在正常作業中，您應該不會需要調整叢集的時間。然而，如果您安裝 Solaris 作業環境時未正確設定時間，而您想要變更時間，其執行程序就在 *Sun Cluster 3.0* 系統管理手冊 中。

高可用性的組織架構

Sun Cluster 讓使用者和資料間的“路徑”上所有元件具有高度的可用性，包括網路介面、應用程式本身、檔案系統和多重主機磁碟。一般而言，如果系統內有任何單一 (軟體或硬體) 失效，叢集元件就具有高度可用性。

下表顯示 Sun Cluster 元件失效的種類 (硬體和軟體)，以及內建於高可用性架構內的復原種類。

表格3–1 Sun Cluster 失效偵測與復原的層次

失效的叢集資源	軟體復原	硬體復原
數據服務	HA API，HA 組織架構	無
公用網路配接卡	網路配接卡失效保護 (NAFO)	多重公用網路配接卡
叢集檔案系統	主要與次要複製	多重主機磁碟
鏡映多重主機磁碟	容體管理 (Solstice DiskSuite 和 VERITAS 容體管理者)	硬體 RAID-5 (例如，Sun StorEdge A3x00)
整體裝置	主要與次要複製	多重裝置路徑，叢集傳輸接點
私有網路	HA 傳輸軟體	多重私有硬體獨立網路
節點	CMM，failfast 驅動程式	多重節點

Sun Cluster 高可用性組織架構快速地偵測到某個節點失效，並且建立一個新的相等伺服器給 叢集中剩餘節點上的組織架構資源。隨時皆可使用組織架構資源。未受故障節點影響的組織架構資源，在回復時完全可加以使用。此外，已失效節點的組織架構資源一經回復之後，便會成爲可使用。已回復的組織架構資源不必等待所有其他的組織架構資源完成回復。

大多數可用性頗高的組織架構資源會回復到使用此資源的應用程式（數據服務）。會在各項節點失效時完整保留組織架構資源存取的語義學。應用程式無法辨識出組織架構資源伺服器已移到另一個節點。只要從另一節點到磁碟存在著另一個替代的硬體路徑，對於在使用檔案、裝置以及連接到 此節點的磁碟容體上的程式而言，單一節點的失效便是完全的透通。其中的一項範例便是使用具有連到多重節點的連接埠的多重主機磁碟。

叢集成員監視器

「叢集成員監視器 (CMM)」是一組分散式的代理程式，每個叢集成員一個代理程式。代理程式透過叢集交互連接來交換訊息，達到：

- 強制對全部節點 (法定數目) 提供一致性的成員視區

- 回應成員變更的磁碟同步化重新配置，使用註冊的呼叫
- 處理叢集分割 (split brain, amnesia)
- 確保所有叢集成員之間的完整連接性

與先前的 Sun Cluster 版次不同，CMM 完全在核心程式中執行。

集成員

CMM 的主要功能，是建立在任何時候參與叢集之節點集合的全叢集協議。Sun Cluster 稱此限制 爲*cluster membership*。

若要決定叢集全體成員，並在最後確保資料完整性，CMM 會：

- 記錄叢集成員的變更，如節點結合或離開叢集
- 確保「錯誤」的節點會離開叢集
- 確保「錯誤」的節點會停留在叢集之外，直到修復爲止
- 防止叢集自行分割成節點子集

請參閱 第42頁的「法定人和法定裝置」 以取得有關叢集如何保護自，以免分割成多重個別叢集的其它資訊。

集成員監視器重新配置

爲了使資料免於毀損，所有的節點必須對叢集成員達成一致的協議。 必要時，CMM 會爲了回應失效而協調叢集服務 (應用程式) 的叢集重新配置。

CMM 從叢集傳輸層接收有關連接到其它節點的資訊。在重新配置期間，CMM 使用叢集交互連接來交換狀態資訊。

在偵測到叢集成員變更之後，CMM 會執行叢集的同步化配置，此時可能會根據新的叢集成員而重新分配叢集資源。

叢集配置儲存庫 (CCR)

「叢集配置儲存庫 (CCR)」是一個私有、全叢集式的資料庫，用來儲存專屬於叢集配置與狀態 的資訊。CCR 是分散式分散式資料庫。每一個節點保有一個完整的資料庫複製。CCR 確保所有的節點 均具有一致的叢集「世界」視區。爲了避免毀損資料，每一個節點都需要知道叢集資源的現行狀態。

CCR 是實作於核心程式中的一個高可用性服務。

CCR 對於更新作業是使用二階段式確定 (two-phase commit) 演算法：必須在所有的叢集成員均順利完成更新，否則更新就會被回復。CCR 使用叢集交互連接來應用分散式更新。

小心：雖然 CCR 是由文字檔所組成，請絕對不要手動編輯 CCR 檔案。每一個檔案均含有總和檢查紀錄，以確保一致性。手動更新 CCR 檔案會導致節點或整個叢集停止運作。

CCR 依賴 CMM 來保證叢集只有在到達法定數目時才能執行。CCR 負責驗證整個叢集的資料一致性、依需要執行復原，以及便利資料的更新。

整體裝置

Sun Cluster 使用整體裝置來提供全叢集、高可用性存取叢集中的任何裝置 (從任意節點)， 不管裝置是否爲實體連接。一般而言，如果節點是在提供整體裝置的存取時失效，Sun Cluster 自動探尋該裝置的其它路徑並將存取重新導向至該路徑。Sun Cluster 整體裝置包括磁碟、CD-ROM 和磁帶。然而，磁碟是唯一支援多埠的整體裝置。 這代表 CD-ROM 和磁帶裝置目前不是高可用性裝置。每部伺服器上的區域磁碟亦不是多埠式，因此不是高可用性裝置。

叢集自動指定唯一的 ID 給叢集中的每個磁碟、CD-ROM 和磁帶裝置。這項指定可以讓人從叢集的任何節點一致存取各個裝置。整體裝置名稱空間是保存於 `/dev/global` 目錄。請參閱 第38頁的「整體名稱空間」 以取得其他資訊。

多埠式整體裝置提供一條以上的裝置路徑。如果是多主機磁碟，因爲磁碟是由一個節點以上所共有之磁碟裝置群組的一部份，所以多主機磁碟具備高可用性。

裝置 ID (DID)

Sun Cluster 藉由建構裝置 ID (DID) 虛擬驅動程式來管理整體裝置。此驅動程式是用來自動指定唯一的 ID 給叢集中的每個裝置，包括多主機磁碟、磁帶機和 CD-ROM。

裝置 ID (DID) 虛擬驅動程式是叢集的整體裝置存取功能的主要部份。 DID 驅動程式會測試叢集的所有節點，並建置唯一磁碟裝置的清單，指定每個裝置唯一的主要號碼和次要號碼，在叢集的所有節點間是一致的。 整體裝置的存取是利用由 DID 驅動程式所指定的唯一裝置 ID 來執行的，而不是傳統的 Solaris 裝置 ID，如磁碟的 `c0t0d0`。

這種方式可以確保使用磁碟裝置的應用程式 (如容體管理者或使用原始裝置的應用程式) 可以使用一致的裝置存取路徑。這種一致性對多主機磁碟而言特別重要，因爲每個裝置的區域主要號碼和次要號碼會隨著節點不同而改變，因此也會變更 Solaris 裝置命

名慣例。 例如，node1 可能將多主機磁碟視爲 c1t2d0，node2 可能將同一磁碟視爲完全不同的其它名稱 c3t2d0。DID 驅動程式會指定一個整體名稱 (如 d10)，而節點則改用此名稱，提供了每個節點一致的多主機磁碟對應。

您是透過 scdidadm(1M) 和 scgdevs(1M) 來更新和管理裝置 ID。請參閱相關的線上援助頁，以取得其他資訊。

磁碟裝置群組

在 Sun Cluster 中，所有的多主機磁碟必須受 Sun Cluster 組織架構的控制。首先您在多主機磁碟上建立磁碟群組—或 Solstice DiskSuite 磁碟組或是 VERITAS 容體管理者磁碟群組。然後，註冊容體管理者磁碟群組爲 Sun Cluster 碟機裝置群組 *(disk device groups)*。磁碟裝置群組是一種整體裝置類型。 此外，Sun Cluster 會將各項個別的磁碟登錄爲磁碟裝置群組。

注意：磁碟裝置群組與資源群組無關。某個節點可以主控一個資源群組 (代表一群數據服務處理程序)，而另外一個節點則可以主控數據服務所存取的磁碟群組。 然而，最佳的方式是將儲存特定應用程式之資料的磁碟裝置群組，以及包含應用程式之資源 (應用程式常駐程式) 的資源群組保存在同一節點上。請參照 *Sun Cluster 3.0 Data Services Installation and Configuration Guide* 中的概觀章節，以取得有關磁碟裝置群組和資源群組關聯的資訊。

利用磁碟裝置群組，容體管理者磁碟群組會變成「整體」，因爲其提供了基礎磁碟的多重路徑支援。 實際連接置多主機磁碟的每一個叢集節點，均提供了一個磁碟裝置群組的路徑。

注意：整體裝置如果是由一個以上的叢集節點所共有之裝置群組的一部份，則具備高可用性。

磁碟裝置失效保護

因爲磁碟外殼連接至一個以上的節點，當目前主控裝置群組的節點失敗時，仍可透過替代路徑來存取該外殼中的所有磁碟裝置群組。 主控裝置群組的節點失效不會影響裝置群組的存取，但是在執行復原與一致性檢查的期間除外。 在這段期間內，所有的要求均會暫停執行 (對於應用程式爲透通的)，直到系統恢復使用裝置群組爲止。

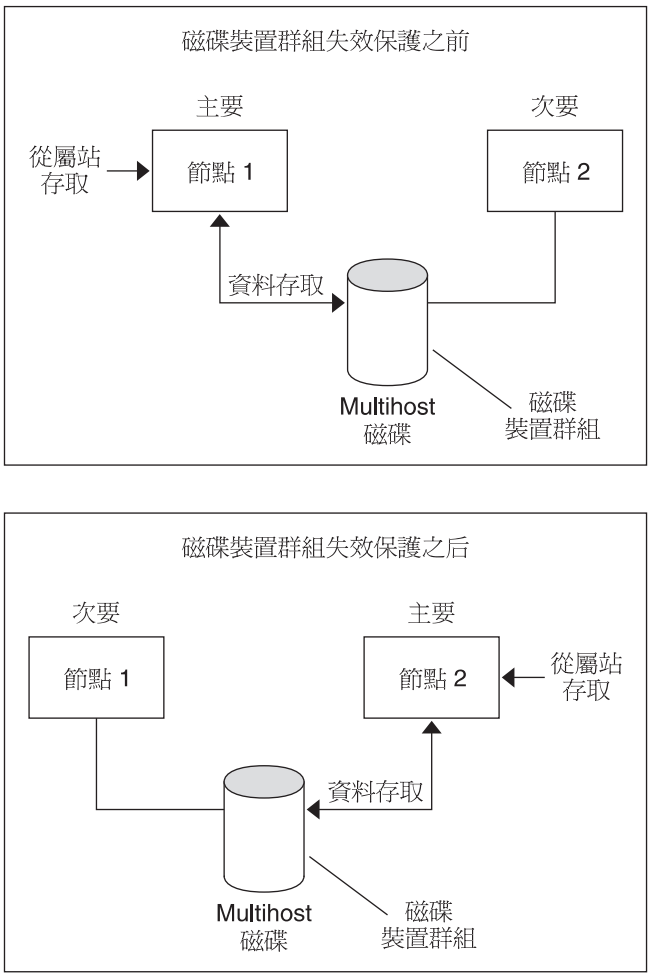

圖 3–2　磁碟裝置群組失效保護

整體名稱空間

讓整體裝置可行的 Sun Cluster 機制稱爲整體名稱空間 *(global namespace)*。 整體名稱空間包括 `/dev/global/` 階層以及容體管理者名稱空間。 整體名稱空間反映多主機磁碟和區域磁碟 (以及任何其它的叢集裝置，如 CD-ROM 和磁帶)，並提供 多主機磁碟的多重失效保護路徑。實際連接多主機磁碟的每一個節點，均提供了一條儲存體路徑給叢集中的任何節點。

一般而言，容體管理者名稱空間是位於 `/dev/md/`*diskset*`/dsk` (和 `rdsk`) 目錄 (Solstice DiskSuite)；以及 `/dev/vx/dsk/`*disk-group* 和 `/dev/vx/rdsk/`*disk-group*

目錄 (VxVM)。這些名稱空間是由各自在整個叢集匯入的每個 Solstice DiskSuite 磁碟組和每個 VxVM磁碟群組之目錄所組成。每個 目錄對該磁碟組或磁碟群組中的每個 metadevice 或容體均含一個裝置節點。

在 Sun Cluster 中，區域容體管理者名稱空間中的每個裝置節點均會被置換爲 /global/.devices/node@*nodeID* 檔案系統 中裝置節點的符號連接，其中 *nodeID* 是在叢集中代表節點的整數。Sun Cluster 仍繼續在其標準位置表示容體管理者裝置，如符號連結。整體名稱空間和標準容體管理者均可由任何叢集節點使用。

整體名稱空間的優點包括：

- 每個節點保持完全獨立，其中在裝置管理模型中有一點變更。
- 裝置可以選擇性地成爲整體。
- 協力廠商連結產生器繼續運作。
- 給定區域裝置名稱，有簡易的對應可以獲得其整體名稱。

區域和整體名稱空間範例

下表顯示多主機磁碟的區域和整體名稱空間之間的對應，c0t0d0s0。

表格3–2 區域和整體名稱空間對應

元件/路徑	本端節點名稱空間	整體名稱空間
Solaris 邏輯名稱	/dev/dsk/c0t0d0s0	/global/.devices/node@*ID*/dev/dsk/c0t0d0s0
DID 名稱	/dev/did/dsk/d0s0	/global/.devices/node@*ID*/dev/did/dsk/d0s0
Solstice DiskSuite	/dev/md/*diskset*/dsk/d0	/global/.devices/node@*ID*/dev/md/*diskset*/dsk/d0
VERITAS 容體管理者	/dev/vx/dsk/*disk-group*/v0	/global/.devices/node@*ID*/dev/vx/dsk/*disk-group*/v0

整體名稱空間是在安裝和更新的每次重新配置重新開機時自動產生。 您也可以執行 scgdevs(1M) 指令來產生整體名稱空間。

叢集檔案系統

叢集檔案系統是某個節點上的核心程式和基礎檔案系統，以及在擁有實體連接到磁碟之節點上的容體管理者間的代理。

叢集檔案系統是相依於與一或多個節點實體連線的整體裝置 (磁碟、磁帶 CD-ROM)。整體裝置 可以從叢集中的任何節點，透過相同的檔名來存取 (例如，`/dev/global/`)，不管 該節點是否實體連線儲存裝置。您可以像使用一般裝置一樣地使用整體裝置，亦即，您可以使用 `newfs` 及/或 `mkfs` 來建立檔案系統。

您可以使用 `mount -g` 以全域方式裝設檔案系統於整體裝置， 或 `mount` 以區域方式裝設。程式可以從叢集的任何節點，透過相同的檔名存取叢集檔案系統 中的檔案 (例如，`/global/foo`)。您可以裝設叢集檔案系統於所有的節點上。您不能裝設叢集檔案系統於叢集成員的子集上。

使用叢集檔案系統

在 Sun Cluster 中，所有的多主機磁碟均配置爲磁碟裝置群組，可以是 Solstice DiskSuite 磁碟組、VxVM 磁碟群組，或是不受軟體式容體管理者控制的個別磁碟。此外，也將區域磁碟配置爲磁碟裝置群組：路徑從每個節點通到每個區域磁碟。 這種設定不表示從所有的節點一定可以使用磁碟上的資料。當磁碟上的檔案系統已裝設爲叢集檔案系統時，才會將資料給所有節點使用。

被納入叢集檔案系統的區域檔案系統只有一個單一的磁碟儲存體連接。如果實體連接到磁碟儲存體的節點失效，其它的節點就無法再存取叢集檔案系統。您在無法從其它節點直接存取的單一節點上，可以擁有區域檔案系統。

設定 HA 數據服務，這樣一來，服務的資料就會儲存在叢集檔案系統中的磁碟裝置群組。 這種設定有許多優點。首先，資料具高可用性；亦即，因爲磁碟是多主機式，如果該節點的目前主要路徑失效，就會將存取切換至可以直接存取同一磁碟的另一個節點。第二，因爲資料是在叢集檔案系統上， 可以從任何的叢集節點直接檢視—您不需要登入目前主控磁碟裝置群組的節點，就可以 檢視資料。

代理檔案系統 (PXFS)

叢集檔案系統是根據具有下列特性的代理檔案系統 (PXFS)：

- PXFS 使檔案存取位置透通。處理程序可以開啓位於系統任何位置的檔案，而且所有節點上的處理程序均可使用相同的路徑名稱來尋找檔案。
- PXFS 使用一致的通信協定來保持 UNIX 檔案存取語意，即使檔案是從多個節點並行地被存取。

- PXFS 提供廣泛的快取功能，並提供「零複製」大量 I/O 移動，以有效地移動大型資料物件。
- PXFS 提供連續的資料存取，即使是在發生失效時。只要磁碟的路徑仍然是 作業中，應用程式不會偵測到失效。這項保證適用於原始磁碟存取和所有的檔案系統作業。
- PXFS 與基礎檔案系統和容體管理軟體無關。PXFS 可以讓任意所支援的「on-disk」檔案系統成爲整體性。
- 在 vnode 介面的現存 Solaris 檔案系統之上建立 vnode 介面。這個介面可讓 PXFS 的實作不需要太多的核心程式修改。

PXFS 不是另外的檔案系統類型。亦即，用戶端可以看見基礎檔案系統 (例如，UFS)。

叢集檔案系統獨立性

叢集檔案系統與基礎檔案系統和容體管理者無關。 目前，您可以使用 Solstice DiskSuite 或 VERITAS 容體管理者 在 UFS 上建置叢集檔案系統。

至於一般檔案系統，您可以用兩種方式裝設叢集檔案系統：

- **Manually**— 使用 `mount` 指令和 `-g` 選項從指令行來裝設叢集檔案系統， 例如：

```
# mount -g /dev/global/dsk/d0s0 /global/oracle/data
```

- **Automatically**— 在 `/etc/vfstab` 檔案中建立具有 `global` 裝設選項的登錄，於啓動時裝設叢集檔案系統。然後在所有節點的 `/global` 目錄下建立裝載點。`/global` 目錄是建議位置，不是基本要求。以下是從 `/etc/vfstab` 檔案之叢集檔案系統的範例行：

```
/dev/md/oracle/dsk/d1 /dev/md/oracle/rdsk/d1 /global/oracle/
data ufs 2 yes global,logging
```

注意：因爲 Sun Cluster 沒有強制叢集檔案系統的命名策略，您可以建立所有叢集檔案系統的裝載點在同一目錄下以簡化管理作業，如 `/global/`*disk-device-group*。 請參閱 *Sun Cluster 3.0* 安裝手冊 和 *Sun Cluster 3.0* 系統管理手冊 以取得其餘資訊。

Syncdir 裝設選項

syncdir 裝設選項可以用於叢集檔案系統。然而，如果您不指定 syncdir，效能就會明顯改善。如果您指定 syncdir，此項寫入便保證相容於 POSIX。如果沒有指定，您將會有與 UFS 檔案系統看到的相同行爲。例如，在某些情況下，沒有 syncdir，一直到關閉檔案，您才會發覺出現空間不足的狀況。利用 syncdir (和 POSIX 行爲)，在關閉之前，便可查覺空間不足的狀況。因爲您沒有指定 syncdir 而發生問題的機會非常小，所以我們建議您不要加以指定，以獲得效能上的益處。

請參閱 第58頁的「檔案系統 常問問題」 以取得有關整體裝置和叢集檔案系統的常見問題。

法定人和法定裝置

因爲叢集節點共用資料和資源，叢集必須採取步驟來維護資料和資源完整性。 當節點不符合叢集成員規則時，叢集必須拒絕節點參與叢集。

在 Sun Cluster 中，決定節點參與叢集的機制稱爲 *quorum*。 Sun Cluster 使用多票數演算法來實作法定程序。叢集節點和 *quorum devices*， (二或多個節點之間共用的磁碟)投票來形成法定人。在 quorum device 中可包含使用者資料。

法定程序演算法是動態運作：當叢集事件 觸發其計算時，計算的結果會變更變更的生命週期。法定人可以防止兩種潛在的演算法 問題—split brain 和 amnesia—這兩種均會造成用戶端取得不一致的資料。 下表說明這兩種問題以及法定人程序如何解決它們。

表格3–3 叢集法定人程序，Split-Brain 與 Amnesia 問題

問題	說明	法定人程序解決方案
Split brain	在節點之間的叢集交互連接遺失，而且叢集分割爲子叢集時發生，每個子叢集相信自己是唯一的分割區	只允許具有多數票的分割區 (子叢集) 作爲叢集執行 (在這樣的多數票情況下，最多只存在一個分割區)
Amnesia	發生時間，是在關機後叢集重新啓動，其中叢集資料比關機時還舊	保證在啓動叢集時，至少有一個節點是最近叢集全體成員中的成員之一 (因此具有最近的配置資料)

法定人票數

叢集節點和法定裝置 (二或多個節點之間共用的磁碟) 票選出法定人。依預設，當叢集節點 啓動和成爲叢集成員時，叢集節點會獲得一票的法定人票數。節點也可能會是零票，例如，當安裝節點或管理者將節點置於維護狀態時。

法定裝置根據節點與裝置的連接數會獲得法定人票數。當設定法定裝置時，它會獲得最大票數 *N*-1，其中 *N* 是非零票數、和以連接埠至 法定裝置之節點的數目。例如，連接至兩個非零票數之節點的法定裝置，擁有一票法定人票數 (二減一)。

您是在叢集安裝期間或者稍後，使用 *Sun Cluster 3.0* 系統管理手冊 中說明的程序來配置法定裝置。

注意：只有當目前連接的節點中至少有一個節點是叢集成員時，才會增加法定裝置票數。 此外，在叢集啓動期間，只有當目前連接的節點中至少有一個節點正在啓動中，而且在此節點上次 關機時是最近啓動之叢集的成員時，才會增加法定裝置票數。

法定人程序配置

法定人配置是根據叢集中的節點數而定：

- **Two-Node Clusters –** 兩個節點的叢集需要兩票法定人票數才能選出。這兩票可以來自兩個叢集節點，或一個節點和一個法定裝置。儘管如此，在兩個節點的叢集中必須配置一個法定裝置，以確保當某個節點失效時，單一節點可以繼續運作。
- **More Than Two-Node Clusters** – 您應該在共用存取 磁碟儲存體機殼的每對節點之間指定一個法定裝置。例如，假設您擁有一個三節點叢集，且與 圖 3–3 中所顯示的類似。在此配置中，nodeA 與 nodeB 共用存取相同的磁碟外殼，而 nodeB 與 nodeC 共用存取另一個磁碟外殼。總共會有五票法定人票，三票來自節點，兩票來自節點間共用的 法定裝置。叢集需要有多數法定人票數 (三票) 才能選出。

  在不需要共用存取磁碟 儲存體機殼或是由 Sun Cluster 執行的每對節點之間指定法定裝置。不過，這樣能提供此項案例必要的 quorum 投票，其中 N+1 配置可降級爲雙節點的叢集，接著具有同時存取兩個磁碟外殼的節點也會失敗。如果您在所有配對之間配置法定投票時，剩餘的節點仍能做爲叢集來運作。

  請參閱 圖 3–3 以取得這些配置的範例。

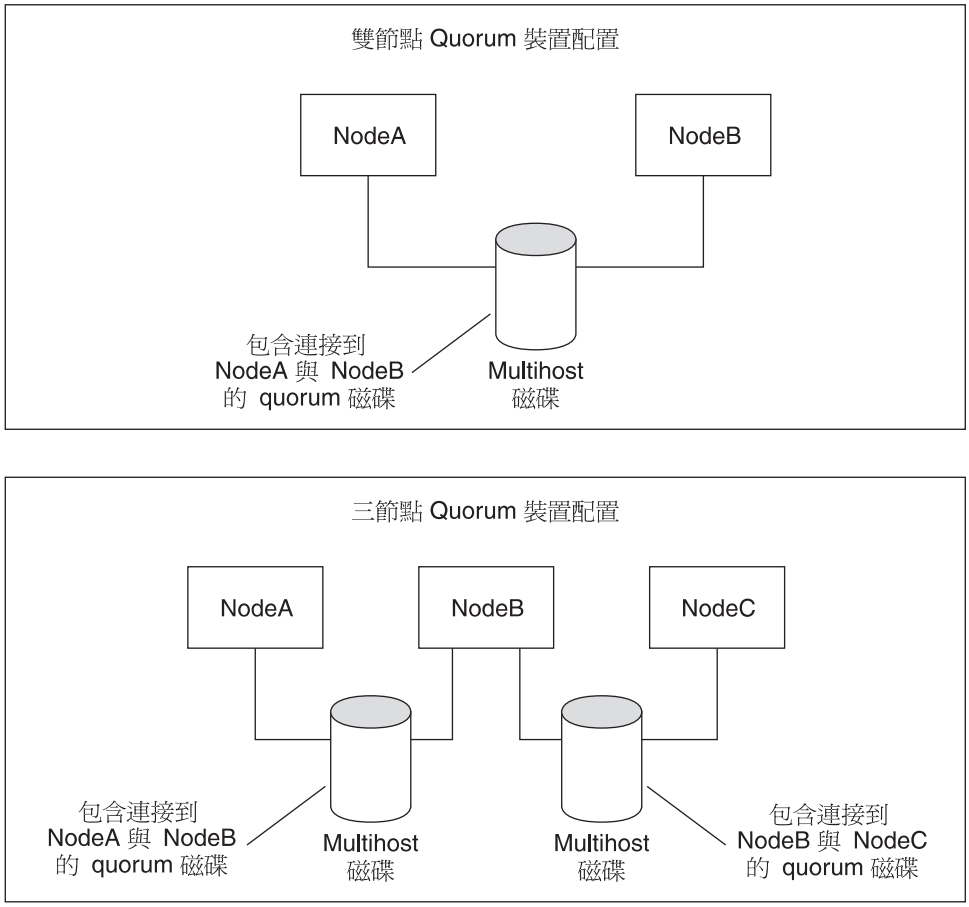

圖 3–3　法定裝置配置範例

法定人準則

設定法定裝置時請使用下列準則：

- 在連接到相同共用磁碟儲存體機殼的所有節點之間，建立法定裝置。 在共用機殼內增加一部磁碟作爲法定裝置，以確保如果有任何節點失效時，其它的節點可以維持法定人和主控共用機殼上的磁碟裝置群組。
- 您必須將法定裝置連接到至少兩個節點。
- 法定裝置可以是任何的 SCSI-2 或 SCSI-3 磁碟作爲雙埠連接的法定裝置。 連接至二個節點以上的磁碟必須支援 SCSI-3 Persistent Group Reservation (PGR)，不管磁碟是否 作爲法定裝置。請參閱 *Sun Cluster 3.0* 安裝手冊 中的規劃章節以取得其他資訊。

- 您可以使用包含使用者資料的磁碟來作爲法定裝置。

提示：在一組節點之間配置一個以上的法定裝置。使用來自不同機殼的磁碟，以及在每組節點 之間配置奇數個法定裝置。如此便可防止個別法定裝置的失效。

失效隔離

叢集的主要議題是造成叢集出現分割的失效 (稱爲 *split brain*)。發生此情形時，不是所有的節點均可通訊，所以個別節點或節點子集可能會嘗試形成個別或子集叢集。 每個子集或分割區可能相信，自己擁有唯一的多主機磁碟存取和所有權。 嘗試寫入磁碟的多個節點會導致資料毀損。

失效隔離藉由實際地防止磁碟存取，限制節點存取多主機磁碟。 當節點離開叢集時 (失效或被分割)，失效隔離可確保節點不會再存取碟。只有目前的成員可以存取 磁碟，因此維持了資料的完整性。

磁碟裝置服務提供失效保護功能給使用多主機磁碟的服務。當目前是磁碟裝置群組的主要 (所有者) 叢集成員失效或無法到達時，會選出新的主要成員，繼續提供磁碟裝置群組的存取，期間只出現輕微的中斷時間。 處理程序期間，在啓動新的主要成員之前，舊的主要成員會放棄存取裝置。然而，當成員退出叢集且接觸不到時，叢集就無法通知該主要節點釋放裝置。因此，您需要 一個方法讓存活的成員可以從失效的成員接手控制和存取整體裝置。

Sun Cluster 使用 SCSI 磁碟保留來實作失效隔離。使用 SCSI 保留，失效的節點會「隔離」多主機磁碟，以防止存取這些磁碟。

SCSI-2 磁碟保留支援一種保留形式，授與存取權給所有連接磁碟的節點 (沒有保留存在) 或限制單一節點的存取權 (握有保留的節點)。

當叢集成員偵測到另一個節點在叢集交互連接上已經不再進行通訊， 即會起始隔離程序來防止其它的節點存取共用磁碟。當發生此失效隔離時， 一般會令隔離節點混亂，並在其主控台上出現「保留衝突」訊息。

偵測到個節點不再是叢集成員時，會放置 SCSI 保留在此節點與其它節點之間共用的所有磁碟上，所以就發生保留衝突的狀況。隔離節點可能不知道，自己已被隔離，而且如果它嘗試存取其中一個共用磁碟，就會偵測到保留和混亂。

容體管理者

Sun Cluster 使用容體管理軟體，藉由鏡映和緊急備件磁碟來增加資料的可用性，以及處理磁碟失效和更換。

Sun Cluster 沒有自己的內部容體管理元件，但是依賴下列容體管理者：

- Solstice DiskSuite
- VERITAS 容體管理者

叢集中的容體管理軟體提供支援：

- 節點失效的失效保護處理
- 不同節點的多重路徑支援
- 遠程透通存取磁碟裝置群組

當設定容體管理者於 Sun Cluster 時，您配置多主機磁碟爲 Sun Cluster 磁碟裝置，容體管理者磁碟群組的外層。裝置可以是 Solstice DiskSuite 磁碟組或 VxVM 磁碟群組。

您必須將使用於數據服務的磁碟群組配置鏡映，才能讓磁碟於叢集內具備高可用性。

您可使用「metadevices」或「plexes」來做爲原始裝置（資料庫應用程式），或是用以保留 UFS 檔案系統。

容體管理物件—metadevices 和容體—受叢集的控制，因此變成磁碟裝置群組。 例如，在 Solstice DiskSuite 中，當您在叢集中建立磁碟組時 (使用 `metaset(1M)` 指令)，會建立相同名稱的對應磁碟裝置群組。然後，當您在該磁碟組中建立 metadevices 時，即變成整體裝置。因此，磁碟組是磁碟裝置 (DID 裝置) 和所有裝置連接之主機的集合。磁碟組中必須具有一部以上 的主機來達到 HA，才能建立叢集中所有的磁碟組。類似的狀況會發生於您使用 VERITAS 容體管理者 的時候。設定每一個容體管理者的詳細資訊，附於 *Sun Cluster 3.0* 安裝手冊 的附錄。

在規劃您的磁碟組或磁碟群組時，有一個重要考慮事項，就是瞭解相關的磁碟裝置群組 如何關聯叢集內的應用程式資源 (資料)。請參照 *Sun Cluster 3.0* 安裝手冊 和 *Sun Cluster 3.0 Data Services Installation and Configuration Guide* 以取得這些議題的討論資訊。

數據服務

數據服務一詞是用來描述已經配置在叢集上 (非單一伺服器) 執行的協力廠商應用程式。而數據服務包括應用軟體軟體，及可啓動、停止和監視此應用程式的 Sun Cluster 軟體。

Sun Cluster 提供數據服務方法來控制和監督叢集內的應用程式。這些方法在 Resource Group Manager (RGM) 的控制之下執行，用來啓動、停止和監督叢集節點

上的應用程式。這些方法配合叢集組織架構軟體和多主機磁碟，讓應用程式變成高可用性數據服務。 作爲高可用性的數據服務，它們可以在叢集內發生任意單一失效之後，防止應用程式明顯中斷。 失效可能是有關節點、介面元件或應用程式本身。

RGM 也會管理叢集內的資源，包括應用程式的實例和網路資源 (邏輯主機名稱和共用位址)。

Sun Cluster 亦提供 API 和數據服務設計工具，讓應用程式程式設計師開發使其它應用程式作爲高可用性數據服務來執行 Sun Cluster 所需要的數據服務方法。

Resource Group Manager (RGM)

Sun Cluster 提供環境讓應用程式具有高可用性或可延伸性。RGM 會影響 *resources*，這些邏輯元件可以被：

- 啓動爲線上或離線 (切換)
- 由 RGM 組織架構管理
- 放置於單一節點 (失效保護模式) 或多重節點 (可延伸模式)

RGM 控制數據服務 (應用程式) 如同資源，由 *resource type*施行管理。 這些施行是由 Sun 提供或由設計人員以一般數據服務範本、數據服務發展檔案庫 API (DSDL API)，或是 Sun Cluster 資源管理 API (RMAPI) 所建立的。叢集管理者建立和管理稱爲資源群組 *(resource groups)* 之容器中的資源， 形成失效保護和切換的基本單元。RGM 停止和啓動所選取節點上的資源群組，以回應叢集成員變更。

失效保護數據服務

如果正在執行數據服務的節點 (主要節點) 失效，該服務會移轉至其它運作中的節點，不需要使用者介入。失效保護服務利用失效保護資源群組 *(failover resource group)*，這是應用程式實例資源 和網路資源(邏輯主機名稱) 的儲存區。邏輯主機名稱是 IP 位址，可以在某個節點配置上線，稍後自動在原始節點配置下線，並在其它節點配置上線。

對於失效保護數據服務，應用程式實例僅在單一節點上執行。如果錯誤監視器偵測到錯誤， 則會嘗試於同一節點重新啓動實例，或於其它節點啓動實例 (失效保護)，視數據服務的配置方式 而定。

可延伸的數據服務

可延伸的數據服務具有在多重節點上的作用中實例之潛力。可延伸的服務利用 可延伸的資源群組 *(scalable resource group)* 來包含相關的應用程式資源，以及失效保護資源群

組 來包含相關的網路資源 (共享地址)。 可延伸資源群組可以在多重節點上成爲線上，所以即可一次執行多個服務實例。放置共用位址的失效保護資源群組一次只在一個節點上啓動線上。放置可延伸服務的所有節點，均使用相同的共用位址來放置服務。

服務要求經由單一網路介面 (*global interface* 或 GIF) 進入叢集， 並且根據平衡資料流量策略 *(load-balancing policy)* 所設定的預先定義演算法的其中一種演算法 來分配給各節點。叢集可以使用平衡資料流量策略，來均衡各個節點之間的服務負載。請注意，在不同的節點上可能有多重的 GIFs 在主控其他共用的位址。

對於可延伸服務，應用程式實例是同時執行於多個節點上。如果放置整體介面的節點失效， 該整體介面會轉移至另一個節點。如果此項應用程式實例失敗時，此實例會嘗試在同一節點上重新啓動。

如果無法在同一節點上重新啓動應用程式實例，就會配置另一個未使用的節點來執行此服務，該服務轉移至未使用的節點。否則，服務會繼續在剩餘的節點上執行，可能造成服務產量的降低。

注意：每個應用程式實例的 TCP 狀態是保存在具有該實例的節點上，而不是在 GIF 節點。 因此，GIF 節點的失效並不會影響連接。

圖 3–4 顯示，對於可延伸服務而言，失效保護和可延伸資源群組的範例， 以及兩者之間的相依關係。此範例顯示三個資源群組。失效保護資源群組包含高可用性 DNS 的應用程式資源，以及高可用性 DNS 和高可用性 Apache Web Server 所使用的網路資源。可延伸資源群組 僅包含 Apache Web Server 的應用程式實例。請注意，可延伸和失效保護資源群組之間的 相依關係 (實線)，以及所有的 Apache 應用程式資源，取決於網路資源 `schost-2`，它是共用位址 (虛線)。

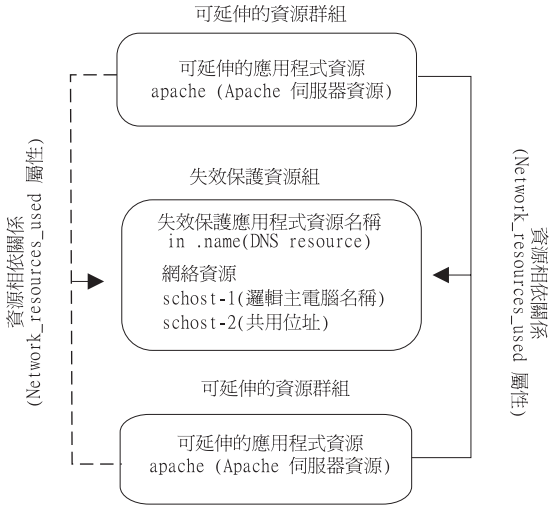

圖 3–4　失效保護和可延伸資源群組範例

可延伸服務架構

叢集網路的主要目標是提供數據服務的可延伸性。可延伸性表示，當服務的負載增加時，因爲將新的節點加入叢集，且執行新的伺服器實例， 所以數據服務在面臨這項增加的工作負擔，能維持不變的回應時間。 我們稱這樣的服務是可延伸數據服務。可延伸數據服務的典型範例是全球資訊網服務。 通常，可延伸數據服務是由許多實例所組成，每一個實例執行於叢集的不同節點上。 整合起來，這些實例便可作爲從此服務的遠端從屬站觀點而來的單一服務，並且建置此項服務的功能性。例如，我們可擁有由在不同節點上執行的數個 `httpd` 常駐程式所組成的可延伸性 Web 服務。任一個 `httpd` 常駐程式可能服務一項從屬站的要求。此項服務要求的常駐程式所依據的，便是平衡資料流量策略。 對用戶端的回答顯然是來自服務，不是服務該要求的特定常駐程式，因此保留了單一服務的外觀。

可延伸服務是由下列組成：

- 支援可延伸服務的網路基礎架構
- 平衡資料流量
- 網路和數據服務的 HA 支援 (使用 Resource Group Manager)

下圖說明了可延伸服務的架構。

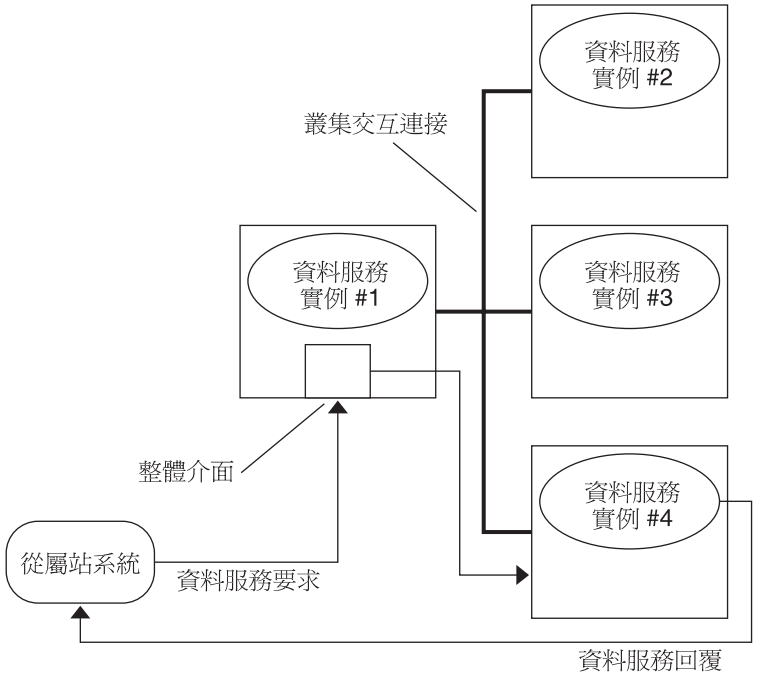

圖 3–5　可延伸服務的架構

沒有放置整體介面的節點 (代理節點) 將共用位址放在其迴圈介面。 進入 GIF 的封包會根據配置的平衡資料流量策略，來分送至其它叢集節點。 可能的平衡資料流量策略說明如後。

平衡資料流量策略

平衡資料流量可以在回應時間域產量上增進可延伸服務的效能。

有兩種可延伸數據服務的類別：*pure* 和 *sticky*。 Pure 服務是，它的任何實例均可回應用戶端要求。Sticky 服務是用戶端傳送要求給相同實例的服務。那些要求不會將方向轉至其它實例。

pure 服務使用加權平衡資料流量策略。在此平衡資料流量策略下，用戶端要求預設會平均地 分配給叢集中的伺服器實例。例如，在三個節點的叢集中，我們假設每一個節點的權重是 1。 每一個節點代表該服務，分別服務 1/3 的任何用戶端的要求。可以由管理者透過 `scrgadm(1M)` 指令介面隨時變更權重。

sticky 服務有兩種方式，*ordinary sticky* 和 *wildcard sticky*。 Sticky 服務允許在多個 TCP 連接上並行處理應用程式層次階段作業，以共用 in-state 記憶體 (應用程式階段作業狀態)。

Ordinary sticky 服務允許用戶端共用多個並行 TCP 連接之間的狀態。用戶端稱爲 “sticky” 是因爲該伺服器實例監聽單一埠。只要該實例維持啓動與可存取的狀態，且當此服務在上線時，載入平衡策略未曾改變，即可保證用戶端的所有要求均會到達相同的伺服器實例。

例如，用戶端上的全球資訊網瀏覽器使用三種不同的 TCP 連線連接到共用 IP 位址的 80 通訊埠，但是連線是在服務交換快取的階段作業資訊。

一般化的 sticky 策略擴展至多重可延伸服務，在相同實例上背景式交換階段作業資訊。 當這些服務在相同實例上於幕後交換階段作業資訊時，用戶端稱爲 “sticky” 是同一節點上的多個伺服器實例監聽不同的埠。

例如，電子商務網站上的客戶使用一般的 HTTP (80 通訊埠) 將物品填入其購物車， 但是會切換至 SSL (443 通訊埠) 傳送安全性資料，以使用信用卡付購物車中物品的帳款。

Wildcard sticky 服務使用動態指定的埠號，但是仍然希望用戶端要求會到達相同的節點。 此從屬站在相關的同一 IP 位址的連接埠上呈現 “sticky wildcard”。

這種策略的典型範例是被動模式 FTP。用戶端連接至 FTP 伺服器的 21 通訊埠， 然後被伺服器通知以動態埠範圍連接回至接收埠伺服器。對此 IP 位址的所有要求，均會轉遞至伺服器經由控制資訊通知用戶端的同一節點。

請注意，對此每一種 sticky 策略，預設都會使用加權平衡資料流量策略，因此用戶端的起始要求會被導向平衡資料流量程式所指定的實例。在用戶端建立與執行實例之節點的關係之後，只要該節爲可存取，且載入平衡策略未變更，則後續的 要求會被導向該實例。

特定平衡資料流量策略的其它明細討論如下。

- 加權式。這項載入會按照指定的加權值來分配到各種的節點。此策略是使用 `Load_balancing_weights` 性質的 `LB_WEIGHTED` 值。如果節點的權重尚未明顯設定時，則此節點的權重將預設爲「一」。

  請注意，此策略並非全體循環式。全體循環式策略一定會將用戶端的每個要求送至不同的節點： 第一個要求到節點 1，第二個要求到節點 2，以此類推。加權策略保證一定百分比的用戶端流量會被導向某個特定節點。本策略不針對個別的要求。

- Sticky。在此策略中，配置應用程式資源時會知道一組埠。此策略已使用 `Load_balancing_policy` 資源性質的 `LB_STICKY` 值來加以設定。

- Sticky-wildcard。此策略是一般 “sticky” 策略的超集。 對於以 IP 位址來識別的可延伸服務而言，是由伺服器來指定埠 (而且事先無法知道)。 埠可能會變更。此策略已使用 `Load_balancing_policy` 資源性質的 `LB_STICKY_WILD` 值來加以設定。

失效回復設定

資源群組因失效保護，從某個節點移轉至另一個節點。您可以指定，當發生資源群組移轉至另一個節點時，在先前執行資源群組的節點返回叢集之後，資源群組將會「失效回復」至原來的節點。這個 選項是使用 `Failback` 資源群組性質設定值來設定的。

在某些情況下，假如放置資源群組的原始節點重複失效和重新開機，設定失效回復可能會 造成資源群組的可用性降低。

數據服務錯誤監視器

每個 Sun Cluster 數據服務均提供了錯誤監視器，定期地測試數據服務以判斷其健康狀態。 錯誤監視器會驗證，應用程式常駐程式是否爲執行中，以及用戶端是否接受服務。根據測試所傳回的資訊， 可以起始預先定義的動作，如重新啓動常駐程式或進行失效保護。

開發新的數據服務

Sun 提供軟體，可讓您將各種應用程式當作叢集內可用性極高的數據服務來操作。 如果您要當作高可用性數據服務來執行的應用程式，目前不是由 Sun 所提供，您可以使用 API 或 DSDL API，將您的應用程式配置成爲高可用性數據服務。 數據服務有兩類，亦即失效保護和可延伸。有一組 基準規則可用來判斷您的應用程式是否可以使用這些數據服務種類。特定的基準規則在 Sun Cluster 文件中有說明，其中說明了可用於您的應用程式的 API。

在此，我們提供一些準則來協助您瞭解，您的服務是否可以利用可延伸數據服務架構。 檢閱 第47頁的「可延伸的數據服務」 節以取得可延伸服務的其它一般資訊。

滿足下列準則的新服務，則可以使用可延伸服務。如果現存的服務不完全符合這些準則， 可能需要改些某些部份，使服務能夠符合準則。

可延伸數據服務具有下列性質。首先，服務是由一或多個 伺服器 *instances* 所組成。每一個實例執行於不同的叢集節點上。同一節點無法執行相同服務的兩或多個實例。

第二，如果服務提供外部邏輯資料儲存處，從多部伺服器對此儲存處作並行存取時，必須同步化，以避免將之變更時遺失更新或讀取資料。請注意，我們強調「外部」，是爲了區分 in-memory state 的儲存處與「邏輯」，因爲儲存處是以單一實體呈現，雖然它本身可能會被複製。此外，此邏輯資料儲存處具有當任何伺服器實例更新儲存處時， 其它實例會立即看到更新的特性。

Sun Cluster 透過其叢集檔案系統與其整體原始分割區來提供這類的外部儲存體。例如，假設服務會寫入新的資料到外部登錄檔，或就地修改現存的資料。當執行此服務的多個實例時， 每個實例均存取此外部登錄，而且可能同時存取此登錄。每一個實例必須將此登錄的存取同步化， 否則實例會互相干擾。服務可以透過 `fcntl(2)` 和 `lockf(3C)` 的一般 Solaris 檔案鎖定，來達到所需的同步化。

這種儲存處的另一個範例是後端資料庫，例如高可用性的 Oracle 或 Oracle Parallel Server。請注意， 這種後端資料庫伺服器使用資料庫查詢或更新異動來提供內建的同步化，因此多重伺服器實例 不需要實作自己的同步化。

目前不是可延伸服務的服務範例，是 Sun 的 IMAP 伺服器。 服務會更新儲存處，但是該儲存處是私有的，而且當多個 IMAP 實例寫入此儲存處時，會因爲未同步化而彼此覆寫。 IMAP 伺服器必須要改寫以同步化並行存取。

最後請注意，實例可能會具有與其它實例的資料區隔的私有資料。在此情況下，服務不需要關心自己的同步化並行存取，因爲資料是私有的，而且只有該實例可以操作資料。因此，您必須愼防將此私有資料儲存在叢集檔案系統之下，因爲它可能會變成可全域存取。

數據服務 API 與數據服務檔案庫 API

Sun Cluster 提供下列項目，可以使應用程式具備高可用性：

- 提供爲 Sun Cluster 一部份的數據服務
- 數據服務 API
- 數據服務發展檔案庫 API
- 「一般」數據服務

Sun Cluster 3.0 Data Services Installation and Configuration Guide 說明如何安裝和配置 Sun Cluster 提供的數據服務。*Sun Cluster 3.0 Data Services Developers' Guide* 說明如何導入其它應用程式，以便在 Sun Cluster 組織架構下具備高可用性。

此項 Sun Cluster API 和「數據服務檔案庫 API」，可讓應用程式程式設計師開發錯誤監視器和啓動及停止數據服務 實例的指令集。利用這些工具，應用程式可以變成具備失效保護和可延伸數據服務。此外，Sun Cluster 可提供 " 一般" 數據服務，可用來快速產生應用程式需要的啓動和停止方法，使其執行爲高可用性數據服務。

資源與資源類型

數據服務利用數種 *resources* 類型：應用程式，如 Apache Web 伺服器或 iPlanet Web 伺服器，以及應用程式所仰賴的網路位址（邏輯名稱和共用位址）。應用程式和網路資源形成受 RGM 管理的基本單位。

資源是在全叢集式定義之 *resource type* 的個體化。 有數種已定義的資源類型。

數據服務式資源類型。例如，Sun Cluster HA for Oracle 是資源類型 `SUNW.oracle`， Sun Cluster HA for Apache 是資源類型 `SUNW.apache`。

網路資源爲 `SUNW.LogiclaHostname` 或 `SUNW.SharedAddress` 資源類型。有兩種資源類型是由 Sun Cluster 產品預先登錄。

此項 `SUNW.HAStorage` 資源類型是用於將資源的啓動與資源所仰賴的磁碟裝置群組進行同步化。它可確保在數據服務啓動之前，叢集裝載檔案系統點路徑，整體裝置，整體裝置名稱可獲得。

RGM 管理的資源會分成群組，稱爲資源群組，讓群組可以一個單位的方式來管理。 如果在資源群組起始了失效保護或切換保護移轉，則資源群組會被當作一個單位來移轉。

注意：當您將包含應用程式資源的資源群組啓動爲線上時，即會啓動應用程式。數據服務啓動方法會等到應用程式啓動並執行之後，才順利結束。 判斷應用程式何時啓動與執行的方式，與數據服務錯誤監視器判斷數據服務是否仍在服務用戶端的方式相同。 請參照 *Sun Cluster 3.0 Data Services Installation and Configuration Guide* 以取得此處理程序的其他資訊。

資源和資源群組性質

您可以配置您的 Sun Cluster 數據服務的資源和資源群組之性質值。有一組標準的性質值是適用所有的數據服務，另外一組延伸性質是專爲每一個數據服務。 部份標準和延伸性質是配置內定值，所以您不需要修改它們。其它性質則需要在建立和配置資源時設定。 每個數據服務的文件指定，資源類型使用哪些性質，以及應該如何加以配置。

標準性質是用來配置通常與任何特定數據服務無關的資源和資源群組性質。 *Sun Cluster 3.0 Data Services Installation and Configuration Guide* 的附錄中說明了這組標準性質。

延伸性質提供了諸如應用程式二進位檔案、配置檔和資源相依項目位置的資訊。您要依照數據服務的配置方式來修改性質。*Sun Cluster 3.0 Data Services Installation and Configuration Guide* 的個別數據服務章節說明了這組延伸性質。

公用網路管理 (PNM) 和網路配接卡失效保護 (NAFO)

用戶端透過公用網路來將要求送至叢集。每一個叢集節點透過公用網路配接卡至少連接到一個公用網路。

「Sun Cluster 公用網路管理 (PNM)」軟體提供監視公用網路配接卡、以及在偵測到失效時將 IP位址從某個配接卡移轉至另一個配接卡的基本機制。每一個叢集節點均擁有自己的 PNM 配置，這些配置可以和其它叢集節點上的 PNM 配置不同。

公用網路配接卡會組成 *Network Adapter Failover groups* (NAFO 群組)。 每一個 NAFO 群組均有一或多個公用網路配接卡。任何時候，針對指定的 NAFO 群組，只能一個配接卡爲作用中，相同群組內的其它配接卡，則作爲作用中配接卡上的 PNM 常駐程式偵測到錯誤而進行配接卡失效保護的備份配接卡。失效保護會令作用中配接卡相關的 IP 位址移到備份配接卡，因而保持了節點的公用網路連接性。因爲失效保護是發生在配接卡介面層次，所以較高層次的連接 (如 TCP) 不受影響， 但是在失效保護期間的短暫延遲除外。

注意：因爲 TCP 的壅塞回復特性，TCP 端點在失效保護成功之後可以承受更進一步的延遲，其中部份區段可能會在失效保護期間遺失，因而啓動 TCP 的壅塞控制機制。

NAFO 群組提供邏輯主機名稱和共用位址資源的建置區塊。如果有必要的話，`scrgadm(1M)` 指令會自動爲您建立 NAFO 群組。您也可以另外建立邏輯主機名稱和共用位址資源的 NAFO 群組來 監視叢集節點的公用網路連接性。節點上的相同 NAFO 群組可以擁有任意數目的邏輯主機名稱或共用位址資源。 有關邏輯主機名稱和共用位址資源的其他資訊，請參閱 *Sun Cluster 3.0 Data Services Installation and Configuration Guide*。

注意：NAFO 機制的設計是爲了偵測和遮罩配接卡失效。其設計目的不是爲了回復管理者使用 `ifconfig(1M)` 移除其中一個邏輯 (或共用) IP 位址的情形。Sun Cluster 設計將 邏輯和共用 IP 位址視爲受 RGM 管理的資源。管理者增加或移除 IP 位址的正確方式，是使用 `scrgadm(1M)` 來修改包含資源的資源群組。

PNM 錯誤偵測和失效保護處理程序

PNM 定期檢查作用中配接卡的封包計數器，假設正常配接卡的封包計數器將會 因爲正常網路流量通過配接卡而變更。如果封包計數器經過一段時間後並沒有變更， PNM 會進入 `ping` 序列，以強制流量通過作用中配接卡。 PNM 會在每次的序列結束時檢查封包計數器是否有任何變更，如果在重複幾次 ping 序列動作之後 封包計數器仍然

不變，則宣告配接卡故障。只要有一個備份配接卡可以使用，這些事件會觸發失效保護以備份配接卡。

PNM 會監視輸入和輸出封包計數器，所以當任一或兩者的計數器有一段時間沒有變更時， 即會起始 ping 序列。

ping 序列包含測試 ALL_ROUTER 廣播位址 (224.0.0.2)、ALL_HOST 廣播位址 (224.0.0.1) 和 區域子網路廣播位址。

Ping 的結構，是以花費最少爲優先考量的方式，所以如果有一個花費較少的 ping 成功時，花費較多的 ping 就不會執行。此外，ping 只是作爲在配接卡上產生流量的方法。其退出狀態不會 作爲配接卡是否爲可運作或故障的決策。

此演算法中有四個可調參數：inactive_time、ping_timeout、 repeat_test 和 slow_network。這些參數提供了錯誤偵測的速度和正確性之間的取捨選擇。請參照 *Sun Cluster 3.0* 系統管理手冊 中變更公用網路參數和變更方法的程序。

在 NAFO 群組的作用配接卡上偵測到錯誤之後，如果無法使用備份配接卡，群組會宣告爲 「當機 (DOWN)」，而所有其備份配接卡的測試會持續。否則，如果有備份配接卡可以使用，失效保護 會發生至備份配接卡。當故障的作用配接卡被關閉和停用時，邏輯位址與其關聯的旗號會「轉移」至備份配接卡。

當 IP 位址失效保護順利完成時，會送出無償式 ARP 廣播，所以可維持與遠程用戶端的連接。

第 4 章

常見問題

本章包含有關 Sun Cluster 最常見問題的解答。問題是依照主題來排列。

高可用性常問問題

- 倒底什麼是高可用性系統？

  Sun Cluster 將高可用性 (HA) 定義爲，即使發生一般會造成伺服器系統無法使用的故障，叢集仍可保持應用程式啓動並執行的能力。

- 叢集是利用何種處理程序來提供高可用性？

  藉由失效保護的處理程序，叢集組織架構提供高可用性的環境。失效保護是叢集執行的一系列步驟，可將應用程式從失效節點移轉至叢集中的另一個可作業節點上。

- 介於 **HA** 與可延伸的服務之間的差異是？

  HA 服務表示應用程式一次僅在叢集中的一個主要節點上執行。 其它的節點可能執行其它的應用程式，但是每個應用程式僅執行於單一節點上。 如果主要節點失效，於失效節點上執行的應用程式會移轉至另一個節點繼續執行。

  可延伸服務將應用程式分散在多個節點，以建立單一、邏輯的服務。可延伸服務會利用其執行所在的整個叢集中的節點與處理器數目。一個節點接收所有應用程式的要求，並將其分送到正在其上執行應用程式伺服器的節點。如果此節點失效 (稱爲「整體介面節點」或 GIF)，整體介面會移轉至存活的節點上。 如果應用程式所執行的任一節點失效，應用程式會繼續在其它的節點上執行，其中部份效能會降低，直到失效節點返回叢集之後才改善。

檔案系統 常問問題

- 用戶端是否可以執行含其它節點的一或多項叢集高可用性的 **NFS** 伺服器？

  不行。能夠除去並重新啓動 `lockd`（在 NFS 失效回復時會發生）的本端鎖定介面出現問題。而在除去和重新啓動期間，暫停執行的本端處理將被授予此項鎖定，此項鎖定可使擁有此項鎖定的從屬站系統免於在失效回復後加以收回。

- 是否可以使用不在 **Resource Group Manager** 控制下的應用程式的叢集檔案系統？

  可以。然而，沒有 RGM 的控制，應用程式無法在其執行的節點失效時存活。

- 是否所有的叢集檔案系統均必須具有一個位於 **/global/*device-group*** 目錄中的裝載點？

  不是。然而，將叢集檔案系統放在相同的裝載點之下 (如 `/global/`*device-group*)，會使這些檔案系統的組織和管理改善。

- 介於使用叢集檔案系統和匯出 **NFS** 檔案系統之間的差異是什麼？

  有許多不同：

  1. 叢集檔案系統支援整體裝置。NFS 不支援遠端存取裝置。
  2. 叢集檔案系統擁有整體名稱空間。只需要一個裝設指令。至於 NFS，您必須在每一個節點裝設檔案系統。
  3. 叢集檔案系統快取檔案的機會多於 NFS。例如，當某個檔案正在被多個節點存取進行讀取、寫入、檔案鎖定和非同步輸入/輸出。
  4. 如果有一個伺服器失敗，叢集檔案系統會支援緊密的失效保護。NFS 支援多重伺服器，但是失效保護只能針對唯讀檔案系統。
  5. 建置叢集檔案系統，是爲了利用提供遠程 DMA 和零複製功能的未來快速叢集交互連接。
  6. 如果您變更叢集檔案系統中某個檔案的屬性 (例如，使用 `chmod(1M)`)，此變更會立即反映到所有節點。 對於匯出式 NFS 檔案系統，此動作要花費較長時間。

容體管理常問問題

- 是否需要鏡映所有的磁碟裝置？

對於要作爲高可用性的磁碟裝置，必須要進行鏡映，或使用 RAID-5 硬體。所有的數據服務應該使用高可用性磁碟裝置，或裝設於高可用性磁碟裝置上的叢集檔案系統。這樣的配置可以容忍單一磁碟失效。

數據服務常問問題

- 可用的 **Sun Cluster** 數據服務是什麼呢？

  支援的數據服務清單包含於 *Sun Cluster 3.0* 版次注意事項。

- **Sun Cluster** 數據服務所支援的應用程式版本爲？

  支援的應用程式版本清單包含於 *Sun Cluster 3.0* 版次注意事項。

- 我是否可寫入自己的數據服務？

  可以。請參閱 *Sun Cluster 3.0 Data Services Developers' Guide* 及 Data Service Development Library API 所提供的「Data Service Enabling Technologies」文件，以取得其他資訊。

- 在建立網路資源時，我是否該指定數字型的 **IP** 位址或主機名稱？

  指定網路資源，最好是使用 UNIX 主機名稱，而非數字型 IP 位址。

- 在建立網路資源時，使用邏輯主機名稱（**LogicalHostname** 資源）或共用的位址（**SharedAddress** 資源）之間的差異是什麼？

  當文件提到在 `Failover` 模式資源群組中使用 `LogicalHostname` 資源 時，可能會交替使用 `SharedAddress` 資源或 `LogicalHostname` 資源。使用 `SharedAddress` 資源 會需要一些額外的負擔，因爲叢集網路軟體式是針對 `SharedAddress` 來配置，而不是 `LogicalHostname`。

  使用 `SharedAddress` 的優點，是當您同時配置可延伸和失效保護數據服務，而且要用戶端能夠使用相同的主機名稱來存取這兩種服務。在此情形下，`SharedAddress` 資源以及失效保護應用程式資源是包含於 一個資源群組中，而可延伸服務資源是包含於另外的資源群組，並且配置使用 `SharedAddress`。於是可延伸和失效保護服務均可使用 `SharedAddress` 資源中配置的同一組主機名稱/位址。

公用網路常問問題

- **Sun Cluster** 所支援的公用網路配接卡爲何？

  目前，Sun Cluster 支援 Ethernet (10/100BASE-T 和 1000BASE-SX Gb)公用網路配接卡。 因爲未來可能會支援新的介面，請洽詢您的 Sun 業務代表，以取得最新的資訊。

- 在失效保護中 **MAC** 位址扮演的角色是什麼？

  發生失效保護時，會產生新的「位址解析度通信協定 (ARP)」封包並廣播到網路上。 這些 ARP 封包包含新的 MAC 位址 (節點移轉後的新實體配接卡的位址) 和舊的 IP 位址。 當網路上的另一部機器收到這些封包時，會清除其 ARP 快取記憶體中的舊 MAC-IP 對應，並使用新的資訊。

- **Sun Cluster** 是否支援在主機配接卡的 **OpenBoot PROM** 中設定 **local-mac-address?=true**

  不是。不支援此變數。

叢集成員常問問題

- 所有的叢集成員是否需要相同的 **root** 密碼？

  每個叢集成員不需要有相同的 root 密碼。然而，所有的節點使用相同的 root 密碼可以簡化您的節點管理工作。

- 節點啓動的順序是否相當重要？

  在大部份的情況下並不會有影響。然而，啓動順序對防止 amnesia 是很重要的 (請參照 第42頁的「法定人和法定裝置」 以取得有關 amnesia 的詳細資訊)。例如，如果節點 2 是 quorum 裝置的所有者，而且節點 1 關機，接著您又將節點 2 關機，則您必須先啓動節點 2 再啓動節點 1。 這樣可以防止您意外啓動具有過時叢集配置資訊的節點。

- 我是否需要在叢集節點中鏡映本端磁碟？

  可以。雖然這種鏡映並非必要，但鏡映叢集節點的磁碟可以排除非鏡映磁碟失效而導致節點當機的情況。 鏡映叢集節點的區域磁碟的缺點，是需要較多的系統管理負擔。

- 叢集成員備份的問題有哪些？

您可以對叢集使用多種備份方法。其中一種方法是令某個節點連接磁帶機/磁帶庫作爲備份節點。 然後使用叢集檔案系統來備份資料。不要連接此節點至共用磁碟。

請參閱 *Sun Cluster 3.0* 系統管理手冊 以取得有關備份和復原程序的其餘資訊。

叢集儲存體常問問題

- 什麼原因讓多主機儲存體具備高可用性？

  多主機儲存體具備高可用性，是因爲有了鏡映 (或硬體式的 RAID-5 控制器) 而可以承受單一磁碟的遺失。 因爲多主機儲存裝置具有一個以上的主機連接，也可以承受失去它所連接的單一節點。

- 支援哪些多主機儲存體配置？

  目前，不支援大於兩節點的連接。單一機殼內的所有多主機磁碟必須連接至相同的兩個節點。 請參照 第26頁的「Sun Cluster 拓樸」 以取得其他資訊。

- 我可以使用對 **SCSI-3 PGR** 配置的磁碟作爲整體裝置嗎？

  目前，Sun Cluster 中不支援 SCSI-3 PGR。只有支援 Only SCSI-2 規格可作爲整體磁碟裝置。 因爲不支援 SCSI-3 磁碟，您必須對要作爲叢集之整體裝置的 SCSI-3 磁碟使用 `scdidadm(1M)` 的 `–R` 選項來設定正確的 SCSI 規格。

叢集交互連接常問問題

- **Sun Cluster** 支援哪些叢集交互連接？

  目前 Sun Cluster 支援 (100BASE-T Fast Ethernet 和 1000BASE-SX Gb) 叢集交互連接。 亦計劃支援 Scalable Coherent Interface (SCI)。

用戶端系統常問問題

- 使用叢集需要考慮任何特殊的用戶端需求或限制嗎？

  用戶端系統連接至叢集，與連接至任何其他伺服器相同。在某些情況下，視數據服務應用程式而定，您可能需要安裝用戶端軟體或執行其它配置變更，使得用戶端可

以連接至數據服務應用程式。請參閱 *Sun Cluster 3.0 Data Services Installation and Configuration Guide* 中的個別章節，以取得有關用戶-端配置需求的其他資訊。

管理主控台常問問題

- **Sun Cluster** 需要管理主控台嗎？

  是的。

- 管理主控台必須專屬於叢集，或者可以用於其它作業？

  Sun Cluster 不需要專用的管理主控台，但是使用專用主控台可以有以下優點：

  - 在同一機器上將主控台和管理工具分組，達到中央化叢集管理
  - 讓您的硬體服務供應商可較快速地解決問題

- 管理主控台位置必須 **"**靠近**"** 叢集本身，例如在同一房間中？

  請洽詢您的硬體服務供應商。供應商可能會要求主控台位置要靠近叢集本身。將主控台置於同一房間中，並無技術上的原因。

- 一部管理主控台在符合距離要求的前提下，可以服務一個以上的叢集嗎？

  可以。您可以從單一管理主控台來控制多個叢集。您也可以在叢集之間共用單一的終端機集線器。

終端機集線器和系統服務處理器常問問題

- **Sun Cluster** 需要終端機集線器嗎？

  執行 Sun Cluster 3.0 不需要終端機集線器。Sun Cluster 2.2 產品需要終端機集線器作爲失效隔離之用，Sun Cluster 3.0 並不依靠終端機集線器。

- 我發現到多數的 **Sun Cluster** 伺服器需要終端集線器，但是 **E10000** 則不用。這是什麼原因呢？

  終端機集線器對大部份的伺服器而言，實際上是一個串列對 Ethernet 轉換器。 其主控台是串列埠。Sun Enterprise E10000 server 沒有串列主控台。「系統服務處理器 (SSP)」是主控台，是透過 Ethernet 或 `jtag` 埠。對於 Sun Enterprise E10000 server，您一定要使用 SSP 於主控台。

- 使用終端集線器有些什麼樣的益處？

  使用終端機集線器可以提供，從網路上任何位置的遠端工作站以主控台層次來存取每一個節點，包括節點是在 OpenBoot PROM (OBP) 時。

- 如果我使用的並非 **Sun** 所支援的終端機集線器時，我該知道些什麼才能讓我想用的合乎標準呢？

  Sun 支援的終端機集線器與其它主控台裝置的主要差異，是 Sun 終端機集線器具有特殊的韌體可以防止終端機集線器在開機時送出中斷。請注意，如果您的主控台裝置會送出中斷，或可能會被解釋爲中斷的信號，它將會關閉節點。

- 我是否可以釋放在 **SUN** 所支援的終端機集線器上已鎖定的連接埠，而不需重新加以啓動？

  可以。請記下需要重設的埠號，並執行下列項目：

```
telnet tc
Enter Annex port name or number: cli
annex: su -
annex# admin
admin : reset port_number
admin : quit
annex# hangup
#
```

  請參照 *Sun Cluster 3.0* 系統管理手冊 以取得配置和管理 Sun 支援之終端機集線器的其他資訊。

- 如果終端機集線器本身失效怎麼辦？我需要有一個備用嗎？

  不需要。如果終端機集線器失效，您並不會失去任何叢集可用性。但是您會失去連接節點主控台的能力，直到集線器回復服務爲止。

- 如果我眞的使用終端機集線器，其安全性如何？

  一般而言，終端機集線器是連接至系統管理者所使用的小型網路，不是連接到其它用戶端存取的網路。 您可以藉由限制該特定網路的存取權來控制安全性。

術語匯編

這個詞彙的名詞解釋用於 Sun Cluster 3.0 文件。

A

管理主控台 (administrative console)
用來執行叢集管理軟體的工作站。

amnesia
在關機後，叢集以舊的叢集配置資料 (CCR) 重新啓動時的一種狀況。例如，在兩個節點叢集中，只有節點 1 可以運作，如果節點 1 發生叢集配置變更，則節點 2 的 CCR 即成爲舊的。如果叢集關機，然後於節點 2 重新啓動，就會因爲節點 2 的舊 CCR 而造成 amnesia 狀況。

自動失效復原 (automatic failback)
在主要節點失效稍後又重新啓動爲叢集成員時，讓資源群組或裝置返回其主要節點的處理程序。

B

備份群組 (backup group)
請參閱「網路配接卡失效保護群組」。

C

核對點 (checkpoint)
主要節點傳給次要節點，以保持兩者間的軟體狀態同步化之通知。亦請參閱「主要」和「次要」。

叢集 (cluster)
兩個以上交互連接的節點或領域共用叢集檔案系統，以及配置爲一起執行失效保護、平行或可延伸資源。

叢集配置儲存庫 (Cluster Configuration Repository，CCR)
Sun Cluster 軟體所使用的高可用性、複製資料儲存處，用以永久保存叢集配置資訊。

叢集檔案系統 **(cluster file system)**
提供全叢集、高可用性存取現存區域檔案系統的叢集服務。

叢集交互連接 **(cluster interconnect)**
包括電纜、叢集傳輸接點和傳輸配接卡的硬體網路基礎架構。Sun Cluster 和數據服務軟體使用此基礎架構進行叢集內的通訊。

叢集成員 **(cluster member)**
目前叢集實體的作用中成員。此成員可以與其他叢集成員共用資源，並提供服務給其他叢集成員和叢集的用戶端。亦請參閱「叢集節點」。

叢集成員監視器 **(Cluster Membership Monitor ，CMM)**
維護叢集登記表一致性的軟體。其餘叢集軟體會使用此成員資訊，來決定放置高可用性服務的位置。 CCM 確保非叢集成員不會毀損資料，以及傳輸毀損或不一致的資料給用戶端。

叢集節點 **(cluster node)**
配置爲叢集成員的節點。叢集節點可能是、也可能不是目前的成員。亦請參閱「叢集成員」。

叢集傳輸配接卡 **(cluster transport adapter)**
位於節點上的網路配接卡，連接節點至叢集交互連接。亦請參閱「叢集交互連接」。

叢集傳輸電纜 **(cluster transport cables)**
連接端點的網路連接。叢集傳輸配接卡和叢集接點、或兩個叢集傳輸配接卡之間的連接。 亦請參閱「叢集交互連接」。

叢集傳輸接點 **(cluster transport junction)**
作爲叢集交互連接之一部份的硬體開關。亦請參閱「叢集交互連接」。

排列 **(collocation)**
在同一節點上的特性。在叢集配置期間會使用這個概念來改進效能。

D

數據服務 **(data service)**
在「資源群組管理員 (RGM)」控制下，用來執行成爲高可用性資源的應用程式。

預設主控者 **(default master)**
失效保護資源類型啓動所在的預設叢集成員。

裝置群組 **(device group)**
使用者定義的裝置資源群組 (如磁碟)，可以從叢集 HA 配置中的不同節點來主控。 此群組可以包含磁碟、Solstice DiskSuite 磁碟組和 VERITAS 容體管理者 磁碟群組的裝置資源。

裝置 **ID (device id)**
透過 Solaris 識別可使用之裝置的機制。 `devid_get(3DEVID)` 線上援助頁中說明了裝置 ID。

Sun Cluster DID 驅動程式使用裝置 ID 來判斷不同叢集節點上 Solaris 邏輯名稱之間的相互關係。DID 驅動程式會測試每一個裝置的裝置 ID。如果該裝置 ID 符合在叢集其它位置的另一個裝置，這兩個裝置會指定相同的 DID 名稱。如果有先前未看過的裝置 ID，則會指定新的 DID 名稱。亦請參閱「Solaris 邏輯名稱」和「DID 驅動程式」。

DID 驅動程式 (DID driver) Sun Cluster 製作的驅動程式，用來提供叢集上的一致裝置名稱空間。亦請參閱「DID 名稱」。

DID 名稱 (DID name) 用來識別 Sun Cluster 中的整體裝置。這是叢集識別字，具有與 Solaris 邏輯名稱的一對一或一對多關係。 採用 d*X*s*Y* 的格式，其中 *X* 是整數，*Y* 是部份名稱。亦請參閱「Solaris 邏輯名稱」。

磁碟裝置群組 **(disk device group)** 請參閱「裝置群組」。

分散式鎖定管理員 **(Distributed Lock Manager ，DLM)** 共用磁碟 Oracle Parallel Server (OPS) 環境中使用的鎖定軟體。DLM 可啓動 於不同節點上執行的 Oracle 處理程序，以便將資料庫存取同步化。DLM 是爲了高可用性而設計。如果 處理程序或節點故障，其餘的節點不需要關機和重新啓動。會執行 DLM 快速重新配置，以復原這類失效。

磁碟組 **(diskset)** 請參閱「裝置群組」。

磁碟群組 **(disk group)** 請參閱「裝置群組」。

E

端點 **(endpoint)** 叢集傳輸配接卡或叢集傳輸接點上的實體通訊埠。

事件 **(event)** 受管理物件的狀態、支配、嚴重程度或說明有變更。

F

失效回復 **(failback)** 請參閱「自動失效回復」。

failfast 可以證明，依照順序關機並且從故障節點移除，然後再進行可能不正確的作業，將會造成損害。

失效保護 **(failover)** 發生失效之後，自動將資源群組或裝置群組從目前主要節點重新放置到新的主要節點。

失效保護資源 **(failover resource)**
一種資源，其中每一個資源一次只能由一個節點正確主控。 亦請參閱「單一實例資源」和「可延伸資源」。

錯誤監視器 **(fault monitor)**
用來測試數據服務的各個部份和採取動作的錯誤常駐程式與程式集。亦請參閱「資源監視器」。

G

一般資源類型 **(generic resource type)**
數據服務的範本。可以用一般資源類型將簡單的應用程式變成具失效保護的數據服務 (在某個節點停止時，會在另一個節點啓動)。這種類型不需要 Sun Cluster API 的程式設計。

一般資源 **(generic resource)**
作爲一般資源類型一部份，受「資源群組管理員」控制的應用程式常駐程式與其子程序。

整體裝置 **(global device)**
可以從所有叢集成員存取的裝置，如磁碟、CD-ROM 和磁帶。

整體裝置名稱空間 **(global device namespace)**
包含邏輯、全叢集的整體裝置名稱的名稱空間。Solaris 環境中的區域裝置是定義於 `/dev/dsk`、`/dev/rdsk` 和 `/dev/rmt` 目錄。 整體裝置名稱空間定義整體裝置於 `/dev/global/dsk`、`/dev/global/rdsk` 和 `/dev/global/rmt` 目錄。

整體介面 **(global interface)**
實際擁有共用位址的整體網路介面。亦請參閱「共用位址」。

整體介面節點 **(global interface node)**
放置整體介面的節點。

整體資源 **(global resource)**
在 Sun Cluster 軟體的核心程式層次提供的高可用性資源。整體資源可以包括 磁碟 (HA 裝置群組)、叢集檔案系統和整體網路。

H

HA 數據服務 **(HA data service)**
請參閱「數據服務」。

心跳 **(heartbeat)**
傳送到所有可用的叢集交互連接傳輸路徑的週期性訊息。 在經過了指定間隔和重試次數之後，沒有收到心跳信號，可能會觸發轉送通訊至另一個路徑的內部失效保護。 通往叢集成員的全部路徑失效時，會導致 CMM 重新評估叢集法定數目。

I

實例 **(instance)** 請參閱「資源呼叫」。

L

平衡資料流量 **(load balancing)** 僅適用可延伸服務。分散應用程式負載到叢集節點的處理程序，這樣可以及時服務用戶端的要求。請參照 第47頁的「可延伸的數據服務」以取得更多詳細資訊。

平衡資料流量策略 **(load-balancing policy)** 僅適用可延伸服務。應用程式要求負載分散至各節點的偏好方式。請參照 第47頁的「可延伸的數據服務」 以取得更多詳細資訊。

區域磁碟 **(local disk)** 實際專屬於某個指定叢集節點的磁碟。

邏輯主機 **(logical host)** 一個 Sun Cluster 2.0 (最小) 概念，包括應用程式，應用資料所在的磁碟組或磁碟群組，以及用來存取叢集的網路位址。這個概念在 Sun Cluster 3.0 已經不存在。請參照 第37頁的「磁碟裝置群組」 和 第54頁的「資源與資源類型」 以取得關於此概念如何在 Sun Cluster3.0 實作的說明。

邏輯主機名稱資源 **(logical hostname resource)** 包含代表網路位址之邏輯主機名稱集合的資源。邏輯主機名稱資源一次只能由一個節點所主控。亦請參閱「邏輯主機」。

邏輯網路介面 **(logical network interface)** 在 Internet 架構中，主機可以有一或多個 IP 位址。Sun Cluster 配置額外的邏輯網路介面，以建立多個邏輯網路介面和單一實體網路介面的對應。每一個邏輯網路介面皆有一個單一 IP 位址。這項對應可讓單一實體網路介面回應多個 IP 位址。這項對應也可以在發生接管或切換時，讓 IP 位址從某個叢集成員移到其它成員，而不需要額外的硬體介面。

M

主控者 **(master)** 請參閱「主要」。

metadevice 狀態資料庫抄本 **(metadevice state database replica，replica)** 儲存於磁碟上的資料庫，記錄所有 metadevice 的配置與狀態和錯誤狀況。這項資訊對於 Solstice DiskSuite 磁碟組的正確作業與其複製很重要。

multihomed host 位在一個以上的公用網路上的主機。

多主機磁碟 **(multihost disk)**
實際連接至多個節點的磁碟。

N

網路配接卡失效保護 **(NAFO)** 群組 **(Network Adapter Failover (NAFO) group)**
在相同節點和相同子網路上的一組一個或多個網路配接卡，爲了在配接卡失效時能夠彼此備份而配置。

網路位址資源 **(network address resource)**
請參閱 “網路資源。”

網路資源 **(network resource)**
包含一或多個邏輯主機名稱或共用位址的資源。亦請參閱「邏輯主機名稱資源」和 「共用位址資源」。

節點 **(node)**
可以成爲 Sun 叢集之一部份的實體機器或領域 (在 Sun Enterprise E10000 server內)。亦稱爲「主機」。

非叢集模式 **(non-cluster mode)**
以 –x 開機選項啓動叢集成員所達到的結果狀態。在此狀態下，節點不再是一個叢集成員，但仍是一個叢集節點。亦請參閱「叢集成員」和「叢集節點」。

P

平行資源類型 **(parallel resource type)**
引進在叢集環境中執行的一種資源類型 (如平行資料庫)，這樣一來，它就會同時由多個(二個或更多) 節點主控。

平行服務實例 **(parallel service instance)**
執行於個別節點上的平行資源類型的實例。

潛在主控者 **(potential master)**
請參閱「潛在主要」。

潛在主要 **(potential primary)**
在主要節點失效時，能夠主控失效保護資源類型的叢集成員。亦請參閱「預設主控者」。

主要 **(primary)**
資源群組或裝置群組目前爲線上狀態所在的節點。亦即，主要是目前放置或實作與資源 關聯之服務的節點。亦請參閱「次要」。

主要主機名稱 **(primary host name)**
主要公用網路上的節點名稱。這是在 `/etc/nodename` 中指定的節點名稱。亦請參閱「次要主機名稱」。

私有主機名稱 **(private hostname)**
用來透過叢集交互連接與節點通訊的主機名稱別名。

公用網路管理 **(Public Network Management ，PNM)**

使用錯誤監視器和失效保護，來防止因爲單一網路配接卡或電纜失效而造成的節點可用性遺失的軟體。 PNM 失效保護使用一組網路配接卡 (稱爲「網路配接卡失效保護」群組) 來提供叢集節點與公用網路之間的備用連接。錯誤監視器和失效保護功能一起確保資源的可用性。亦請參閱「網路配接卡失效保護群組」。

Q

quorum 裝置 **(quorum device)**

由兩個或更多節點所共享的磁碟，提供已使用的投票，以便建叢集的「quorum」來加以執行。只有在可使用投票的「quorum」之後，叢集方能運作。「quorum」裝置的使用時機，是在叢集劃分爲個別的節點集，以便建立由哪一個節點集投票給新的叢集。

R

資源 **(resource)**

資源類型的實例。相同類型的許多資源可能存在，每個資源擁有自己的名稱和一組屬性值，使得許多基礎應用程式的實例可以在叢集上執行。

資源群組 **(resource group)**

受 RGM 管理、視爲一個單元的資源集合。要由 RGM 管理的每一個資源都必須配置於資源群組中。一般而言，相關和獨立資源會被分組。

資源群組管理員 **(Resource Group Manager ，RGM)**

藉由自動啓動和停止所選取叢集節點上的叢集資源，使得這些資源具備高可用性和可延伸性的軟體設備。發生硬體或軟體失效或重新開機時，RGM 會依照預先配置的策略來運作。

資源群組狀態 **(resource group state)**

任意指定之節點上的資源群組狀態。

資源呼叫 **(resource invocation)**

執行於節點上的資源類型實例。代表啓動於節點上之資源的抽象概念。

資源管理 **API (Resource Management API ，RMAPI)**

Sun Cluster 內的應用程式設計介面，可以使應用程式在叢集環境中成爲具高可用性。

資源監視器 **(resource monitor)**

資源類型實作的選用部份，定期對資源執行錯誤測試，判斷是否正確執行它們以及其執行狀況。

資源狀態 **(resource state)**

指定節點上的 Resource Group Manager 資源狀態。

資源狀態 **(resource status)**
錯誤監視器所報告的資源狀況。

資源類型 **(resource type)**
指定給數據服務、LogicalHostname 或 SharedAddress 叢集物件的唯一名稱。 數據服務資源類型可以是失效保護類型或可延伸類型。亦請參閱「數據服務」，「失效保護資源」和「可延伸資源」。

資源類型性質 **(resource type property)**
一個鍵值配對，由 RGM 儲存爲資源類型的一部份，用來描述和管理指定類型的資源。

S

可延伸的一致性介面 **(Scalable Coherent Interface (SCI))**
作爲叢集交互連接的高速交互連接硬體。

可延伸資源 **(scalable resource)**
執行於多個節點的資源 (每個節點一個)，利用叢集交互連接將單一服務提供給該服務的遠程用戶端。

可延伸服務 **(scalable service)**
實作成爲可同時執行於多個節點的數據服務。

次要 **(secondary)**
發生主要節點失效時，可以主控磁碟裝置群組和資源服務的叢集成員。亦請參閱「主要」。

次要主機名稱 **(secondary host name)**
用來存取次要公用網路上之節點的名稱。亦請參閱「主要主機名稱」。

共用位址資源 **(shared address resource)**
網路位址，可由叢集中於節點上執行的所有可延伸服務來結合，以便使它們在那些節點上進行延伸。叢集可具有多個共用的位址，而且服務也可結合到多個共用的位址。

單一實例資源 **(single instance resource)**
叢集中最多只能有一個資源爲作用中的資源。

Solaris 邏輯名稱 **(Solaris logical name)**
一般用來管理 Solaris 裝置的名稱。對於磁碟而言，這些通常看來像是 `/dev/rdsk/c0t2d0s2`。對於 這些 Solaris 邏輯裝置名稱的每一個名稱而言，皆有一個基礎 Solaris 實體裝置名稱。亦請參閱「DID 名稱」和「Solaris 實體名稱」。

Solaris 實體名稱 **(Solaris physical name)**
由 Solaris 中的裝置驅動程式指定給該裝置的名稱。這個名稱在 Solaris 機器上顯示爲 `/devices` 目錄樹下的路徑。例如， 典型的 SCSI 磁碟的 Solaris 實體名稱類似：`/devices/sbus@1f,0/SUNW,fas@e,8800000/sd@6,0:c,raw`

亦請參閱「Solaris 邏輯名稱」。

Solstice DiskSuite Sun Cluster 所使用的容體管理。亦請參閱「容體管理者」。

split brain 叢集分裂成多個分割區的狀況，每個分割區在不知道其它分割區存在的情況下形成。

切換回復 **(switchback)** 請參閱「失效回復」。

切換 **(switchover)** 依照順序將資源群組或裝置群組自叢集中的某個主控者（節點）轉送至另一個主控者（或多個主控者，如果資源群組是配置給多個主要的話）。切換是由管理者使用 `scswitch(1M)` 指令所起始的。

系統服務處理器 **(System Service Processor ，SSP)** 在 Enterprise 10000 配置中，外接於叢集、特別用來與叢集成員通訊的裝置。

T

接管 **(takeover)** 請參閱「失效保護」。

終端機集線器 **(terminal concentrator)** 在非 Enterprise 10000 配置中，外接於叢集、特別用來與叢集成員通訊的裝置。

V

VERITAS 容體管理 Sun Cluster 所使用的容體管理。亦請參閱「容體管理者」。

容體管理者 **(volume manager)** 透過磁碟資料分置、接合、鏡映及 metadevice 或容體的動態成長來提供資料可靠性的軟體產品。