Sun Cluster 3.0 U1 概念

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

元件號碼： 816-1957-10
2001 年 8 月, Revision A

Adobe PostScript

內容

前言

Sun™ Cluster 3.0 U1 概念包含有關 SunPlex ™ 系統的概念性和參考資訊。SunPlex 系統包含構成 Sun 叢集解決方案的所有硬體和軟體元件。

此文件主要是針對在 Sun Cluster 軟體上接受過訓練且有經驗的系統管理員。請不要將本文件當做規劃作業或售前指引。您應該已經決定您的系統需求並購買了適當的設備與軟體之後再閱讀本文件。

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 SunPlex 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。 `%` 您有郵件。
`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。 您必須是超級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 超級使用者	*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 U1* 安裝手冊	806-7069
硬體	*Sun Cluster 3.0 U1 Hardware Guide*	806-7070
資料服務	*Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*	806-7071

主題	標題	組件號碼
API 設計	*Sun Cluster 3.0 U1 Data Services Developer's Guide*	806-7072
管理	*Sun Cluster 3.0 U1* 系統管理手冊	806-7073
錯誤訊息與問題解決方法	*Sun Cluster 3.0 U1 Error Messages Manual*	806-7076
版本注意事項	*Sun Cluster 3.0 U1* 版次注意事項	806-7078

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 Web 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 SunPlex 系統上有問題，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 軟體的版本號碼 (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
<command>prtconf -v</command>	顯示系統記憶體的大小及報告周邊裝置的相關資訊
<command>psrinfo -v</command>	顯示處理器的相關資訊
showrev --p	報告安裝了哪些修補程式
<command>prtdiag -v</command>	顯示系統偵錯資訊
scinstall -pv	顯示 Sun Cluster 軟體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

SunPlex 系統爲一整合的硬體與軟體解決方案，用於建立高可用性及可延伸的服務。

Sun Cluster 3.0 U1 概念 提供 SunPlex 文件主要讀者所需的概念性資訊。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 軟體的系統管理員
- 開發目前 Sun Cluster 產品所未包含的應用程式故障轉移和可延伸服務的應用程式開發人員

本書配合其餘的 SunPlex 文件集，提供 SunPlex 系統的完整概觀。

本章

- 提供 SunPlex 系統的簡介和進階概觀
- 說明 SunPlex 讀者的各種觀點
- 指出在使用 SunPlex 系統之前需要瞭解的重要概念
- 對應重要概念至包含程序與相關資訊的 SunPlex 文件
- 對應叢集相關作業至包含用來完成那些作業程序的文件

SunPlex 系統簡介

SunPlex 系統將 Solaris 作業環境延伸成爲叢集作業系統。叢集或檢視裝置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 (對網路用戶端) 協力將應用程式、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勢。這些優勢包括支援故障轉移和可延伸服務的支援、模組成長的能力，以及比傳統硬體容錯系統低的導入成本。

SunPlex 系統的目標是：

- 減少或免除因爲軟體或硬體故障所造成的當機時間
- 確保對一般使用者的資料和應用程式的可用性，不論是否爲一般使單一伺服器系統當機的那種故障
- 增加節點至叢集，讓服務延伸至額外的處理器，以增加應用程式的效率
- 讓您可以執行維護作業而不需要關閉整個系統，以提供強化的系統可用性

高可用性與容錯性

SunPlex 系統被設計成高可用性 (Highly Available，HA) 系統，亦即可幾近連續存取資料和應用程式的系統。

相形之下，容錯性硬體系統雖提供持續性的資料和應用程式存取，但因爲是特殊硬體，所以成本較高。此外，容錯性系統通常不會說明軟體故障。

SunPlex 系統透過硬體和軟體的結合來達到高可用性。多餘的叢集交互連接、儲存體和公用網路可防止發生單一點故障。叢集軟體會持續監督成員節點的運作狀況，並阻止故障節點參與叢集，以免資料受到毀損。此外，叢集會監督服務及其相依系統資源，並且在發生故障時進行故障轉移或重新啓動服務。

請參閱 第63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。

SunPlex 系統的故障轉移和延伸性

SunPlex 系統可以讓您使用故障轉移或可延伸的服務。一般而言，故障轉移服務僅提供高可用性 (冗餘)，而可延伸的服務則提供高可用性且增加效能。單一叢集可以同時支援故障轉移和可延伸的服務。

故障轉移服務

故障轉移是叢集自動將已發生故障之主要節點上的服務，重新放置到指定之次要節點的處理程序。有了故障轉移，Sun Cluster 軟體可提供高可用性。

當發生故障轉移時，用戶端可能會看到短暫的服務中斷，並且可能需要在完成故障轉移動作之後重新連線。然而，用戶端並不會意識到提供服務的實體伺服器的存在。

可延伸的服務

故障轉移與冗餘有關，而延伸性則提供不變的回應時間或產量，且與負載無關。可延伸的服務會調節叢集中的多個節點來同時執行應用程式，因此提供較佳的效能。在可延伸的配置中，叢集的每個節點均可提供資料並處理用戶端的要求。

請參閱 第48頁的「資料服務」，以取得關於故障轉移和可延伸服務的更多特定資訊。

SunPlex 系統的三個觀點

本節說明 SunPlex 系統的三個不同觀點和重要概念，以及每個觀點的相關文件。這些觀點來自：

- 硬體安裝與維修人員
- 系統管理員
- 應用程式設計師

硬體安裝與服務觀點

對於硬體維修人員而言，SunPlex 系統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，因此使得每一個元件均具有備份而不會有單點故障存在。

重要概念 – 硬體

硬體維修人員需要瞭解下列叢集概念。

- 叢集硬體配置和電纜佈線
- 安裝與維修 (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24頁的「本機磁碟」
- 第24頁的「叢集交互連接」
- 第25頁的「公用網路介面」
- 第25頁的「用戶端系統」
- 第26頁的「管理主控台」
- 第25頁的「主控台存取裝置」
- 第27頁的「叢集化配對拓樸架構」
- 第28頁的「N+1 (星狀) 拓樸」

相關的 SunPlex 文件

下列 SunPlex 文件包括與硬體維修概念相關的程序和資訊：

- *Sun Cluster 3.0 U1 Hardware Guide*

系統管理員觀點

對於系統管理員而言，SunPlex 系統就像是一組以電纜連接在一起的伺服器 (節點)，共用儲存裝置。系統管理員會看見：

- 與 Solaris 軟體整合的專用叢集軟體，用來監視叢集節點之間的連接
- 用來監視在叢集節點上執行的使用者應用程式運作狀況的專用軟體
- 設定和管理磁碟的容體管理軟體
- 可以讓所有節點存取所有儲存裝置 (即使未直接連接到磁碟) 的專用叢集軟體

- 可以讓檔案像是本機連接至該節點的方式出現於每個節點上的專用叢集軟體

重要概念 – 系統管理

系統管理員需要瞭解下列概念與程序：

- 硬體和軟體元件之間的相互作用
- 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 軟體
  - 安裝和配置容體管理者
  - 安裝和配置應用軟體使其具備叢集功能
  - 安裝和配置 Sun Cluster 資料服務軟體
- 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

建議的系統管理員概念參考文件

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34頁的「高可用性框架」
- 第37頁的「整體裝置」
- 第37頁的「磁碟裝置群組」
- 第39頁的「全域名稱空間」
- 第41頁的「叢集檔案系統」
- 第43頁的「法定數目和法定裝置」
- 第47頁的「容體管理者」
- 第48頁的「資料服務」
- 第58頁的「資源、資源群組與資源類型」
- 第60頁的「公用網路管理 (PNM) 和網路配接卡故障轉移 (NAFO)」
- 第 4 章

相關的 SunPlex 文件 – 系統管理員

下列 SunPlex 文件包含與系統管理概念相關的程序和資訊：

- *Sun Cluster 3.0 U1* 安裝手冊
- *Sun Cluster 3.0 U1* 系統管理手冊
- *Sun Cluster 3.0 U1 Error Messages Manual*

應用程式設計師觀點

SunPlex 系統提供用於諸如 Oracle、NFS、DNS、iPlanet Web Server、Apache Web Server 及 Netscape Directory Server 應用程式的 資料服務。資料服務是藉由配置 Sun Cluster 軟體控制下之常用應用程式而建立的。Sun Cluster 軟體提供啓動、停止與監視應用程式的配置檔案與管理方法。如果您需要建立新的故障轉移或可延伸的服務，您可以使用 SunPlex 應用程式設計介面 (Application Programming Interface，API) 與啓用資料服務技術 API (Data Service Enabling Technologies API，DSET API) 來發展配置檔案與管理方法，以使其應用程式以資料服務方式在叢集上執行。

重要概念 – 應用程式設計師

應用程式設計師需要瞭解下列各項：

- 應用程式的特性，以決定其是否可以被當作故障轉移或可延伸的資料服務來執行。
- Sun Cluster API、DSET API 及“一般”資料服務。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，以配置其用於叢集環境的應用程式。

建議的應用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48頁的「資料服務」
- 第58頁的「資源、資源群組與資源類型」
- 第 4 章

相關的 SunPlex 文件 – 應用程式設計師

下列 SunPlex 文件包含與應用程式設計師概念相關的程序和資訊：

- *Sun Cluster 3.0 U1 Data Services Developer's Guide*
- *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*

SunPlex 系統作業

所有 SunPlex 系統作業都需要具備某些概念背景。下列表格提供作業與說明作業步驟之文件的進階概觀。本書中有關的概念章節說明概念如何對應至這些作業。

表格1–1 對應作業：將使用者作業對應至文件

執行此作業...	使用此文件...
安裝叢集硬體	*Sun Cluster 3.0 U1 Hardware Guide*
將 Solaris 軟體安裝於叢集上	*Sun Cluster 3.0 U1* 安裝手冊
安裝 Sun™ Management Center 軟體	*Sun Cluster 3.0 U1* 安裝手冊
安裝和配置 Sun Cluster 軟體	*Sun Cluster 3.0 U1* 安裝手冊
安裝和配置容體管理軟體	*Sun Cluster 3.0 U1* 安裝手冊 您的容體管理文件
安裝和配置 Sun Cluster 資料服務	*Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*
維修叢集硬體	*Sun Cluster 3.0 U1 Hardware Guide*
管理 Sun Cluster 軟體	*Sun Cluster 3.0 U1* 系統管理手冊
管理容體管理軟體	*Sun Cluster 3.0 U1* 系統管理手冊 和您的容體管理文件
管理應用程式軟體	您的應用程式文件

表格1–1 對應作業：將使用者作業對應至文件 *(續上)*

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 U1 Error Messages Manual*
建立新的資料服務	*Sun Cluster 3.0 U1 Data Services Developer's Guide*

第 2 章

重要概念 – 硬體服務供應商

本章說明有關 SunPlex 系統配置的硬體元件的重要概念。

SunPlex 系統的硬體元件

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。叢集系統管理員可能也會發現，這項資訊對於安裝、配置和管理叢集軟體是很有用的。

叢集是由數個硬體元件所組成，包括：

- 具有本機磁碟 (未共用) 的叢集節點
- 多重主機儲存體 (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

SunPlex 系統可以讓您將這些元件結合成各種配置，請參閱 第26頁的「Sun Cluster 拓樸」 之說明。

下圖顯示範例叢集配置。

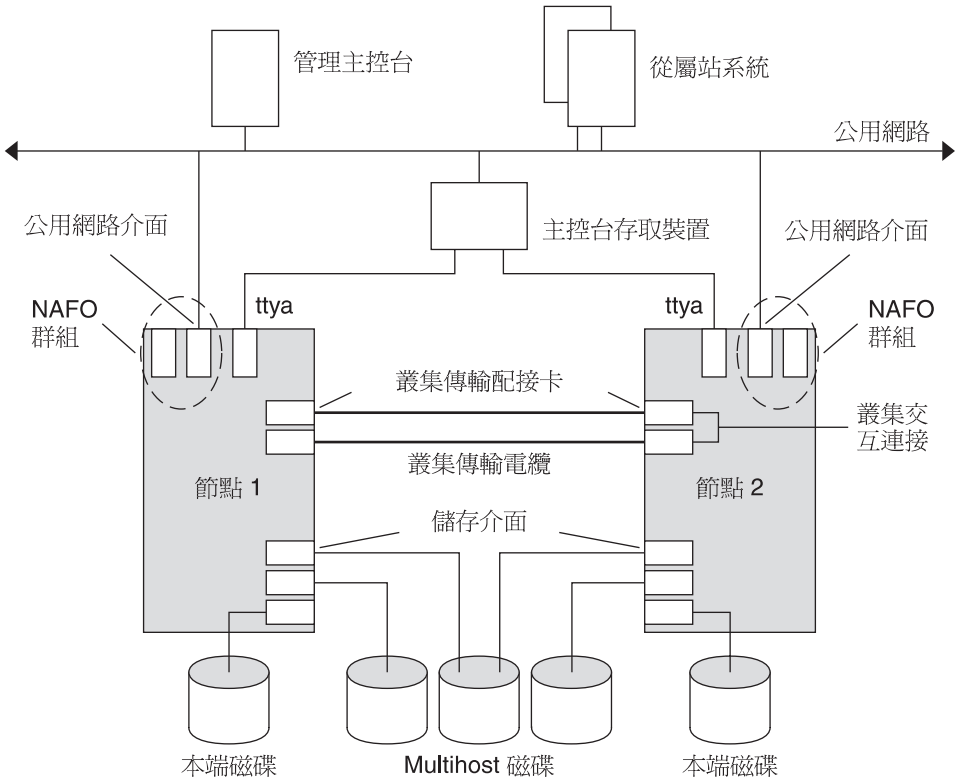

圖 2–1　兩個節點的叢集配置範例

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 (叢集成員 *IT-0144*) 或潛在成員。Sun Cluster 軟體可以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」，以取得支援的節點配置。

叢集節點一般是連接到一個或多個多重主機磁碟。未連接到多重主機磁碟的節點，是使用叢集檔案系統來存取多重主機磁碟。例如，一個可延伸的服務配置可以讓節點不需要直接連接到多重主機磁碟便可處理要求。

此外，平行資料庫配置中的節點會共用對於所有磁碟的並行存取。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詳細資訊。

叢集中的所有節點會依照一般名稱，即叢集名稱 (用來存取和管理叢集)，來加以分群。

公用網路配接卡會將節點連接到公用網路，以供用戶端存取叢集。

叢集成員是透過一或多個實體上獨立的網路來與叢集上的其他節點通訊。此組實體上獨立的網路是被視爲 叢集交互連接。

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機正在執行的資源，以及在其它叢集節點上執行的資源。

相同叢集中的節點必須有類似的處理程序、記憶體和 I/O 能力，以便啓動故障轉移，而不至於大幅降低效能。由於可能發生故障轉移，所以每個節點必須有足夠的額外容量，可以作爲備份或次要節點來接管所有節點的工作負荷。

每一個節點會啓動其個別的 root (/) 檔案系統。

叢集成員的軟體元件

若要作爲叢集成員，必須安裝下列軟體：

- Solaris 作業環境
- Sun Cluster 軟體
- 資料服務應用程式
- 容體管理 (Solstice DiskSuite™ 或 VERITAS Volume Manager)

  一種例外情形是在使用獨立磁碟 (RAID) 的硬體冗餘陣列的 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 Volume Manager，以便管理 Oracle 資料。

請參閱 *Sun Cluster 3.0 U1* 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 U1 Data Services Installation and Configuration Guide*，以取得有關如何安裝和配置資料服務的資訊。

請參閱 第 3 章，以取得前述軟體元件的概念資訊。

下圖提供軟體元件的高階觀點，那些軟體元件會共同運作以建立 Sun Cluster 軟體環境。

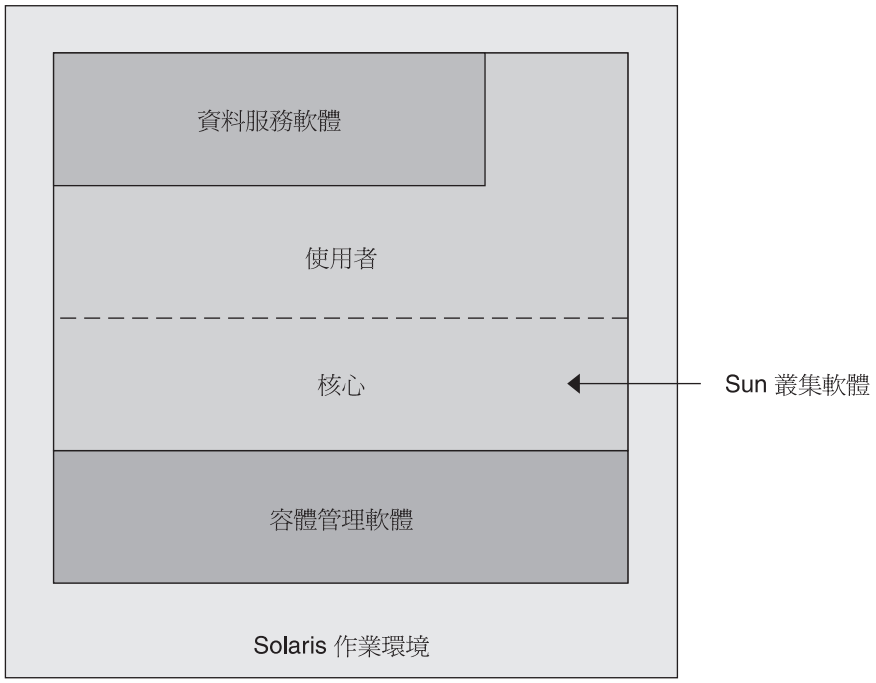

圖 2–2　Sun Cluster 軟體元件的高階關係

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 需要多重主機磁碟儲存體：可同時連接一個以上節點的磁碟。在 Sun Cluster 環境中，多重主機儲存體可讓磁碟非常有用。

多重主機磁碟有下列特性：

- 它們可容許單一節點故障。
- 它們儲存應用程式資料，也可儲存應用程式的二進位檔案與配置檔案。
- 它們對於節點故障做出保護。如果用戶端要求是透過某個節點來存取資料而該節點故障，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。
- 多主機磁碟是透過“主控”磁碟的主要節點來進行全域存取，或透過本機路徑直接進行並行存取。目前使用直接並行存取的唯一應用程式是 OPS。

容體管理者提供鏡像或 RAID-5 配置的多主機磁碟資料冗餘。目前，Sun Cluster 支援 Solstice DiskSuite 和 VERITAS Volume Manager 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鏡像與資料分置，可以防止節點故障和個別的磁碟故障。

請參閱 第 4 章，以取得有關多主機儲存體的問題與解答。

多重初始端 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 初始端 (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。

叢集配置利用多重主機磁碟在多重伺服器節點之間共用儲存體。當叢集儲存體是由單端或差動式 SCSI 裝置所組成時，該配置即爲多重初始端 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 初始端。

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。) 在多重初始端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 位址爲 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。將來不使用的位址範例，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) 屬性，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，來設定節點的這個屬性。有關設定每一個 SCSI 主機配接卡的唯一 `scsi-initiator-id` 的指示在 *Sun Cluster 3.0 U1 Hardware Guide* 中各磁碟機殼的章節中有所說明。

本機磁碟

本機磁碟是僅連接至單一節點的磁碟。因此，沒有節點故障的保護 (不具高可用性)。然而，所有的磁碟 (包括本機磁碟) 均含括於全域名稱空間中，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。

您可以將本機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故障，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像這些磁碟，如此磁碟故障就不會導致這些檔案系統變成無法存取，但是容體管理者無法防止節點故障。

請參閱 第37頁的「整體裝置」 一節，以取得有關整體裝置的詳細資訊。

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從叢集的任何節點來存取。請參照 *Sun Cluster 3.0 U1 Hardware Guide*，以取得安裝和配置抽換式媒體的資訊。

請參閱 第37頁的「整體裝置」 一節，以取得有關整體裝置的詳細資訊。

叢集交互連接

叢集交互連接是用來傳輸叢集節點之間的叢集私有通訊與資料服務通訊的實體裝置配置。由於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叢集交互連接。Sun Cluster安全性模型假設只有叢集節點具有叢集交互連接的實體存取權。

所有的節點必須透過至少兩個實體上多餘的獨立網絡或路徑，藉由叢集交互連接來連接，才能避免單點故障的情形。任何兩個節點之間可以有多個實體上獨立的網路 (二到六個)。叢集交互連接由三個硬體元件組成：配接卡、接點與電纜。

下表說明各個硬體元件。

- 配接卡 – 位於每個叢集節點的網路配接卡。它們的名稱是由後面緊接著實體單位號碼的裝置名稱 (例如 qfe2) 所組成。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故障時會變成單一故障點。爲擁有最大的可用性，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。執行通行和轉換功能，讓您將兩個以上的節點連接在一起。在雙節點的叢集中，您不需要接點，因爲透過多餘的實體電纜連接至每個節點上的冗餘配接卡，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，這要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，當一個介面卡爲作用中時，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」(Public Network Management，PNM)，可監視作用中的介面。如果作用中配接卡故障，會呼叫「網路配接卡故障轉移」(Network Adapter Failover，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。用戶端程式使用由執行於叢集上的伺服器端應用程式所提供的資料或其它服務。

用戶端系統不具高可用性。叢集上的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。

主控台存取裝置

對於所有的叢集節點，您必須擁有主控台存取權。要取得主控台存取，請使用與您叢集硬體一起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」(System Service Processor，SSP) 或是可以存取每個節點上 `ttya` 的其它裝置。

來自 Sun 之受支援的終端機集線器只有一個，而是否使用此支援的 Sun 終端機集線器是可選擇的。終端機集線器允許使用 TCP/IP 網路來存取每一個節點上的 `/dev/console`。結果是從網路上任意位置的遠端工作站，以主控台層次來存取每一個節點。

「系統服務處理器」(SSP) 提供 Sun Enterprise E10000 server 的主控台存取。SSP 是 Ethernet 網路上的機器，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，從另一個節點和無智型終端機的 `tip(1)` 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)。

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的管理工具軟體，會是像 Sun Management Center 產品的「叢集控制面板」(Cluster Control Panel，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。請參閱 *Sun Cluster 3.0 U1* 系統管理手冊，以取得有關使用 CCP 的詳細資訊。

管理主控台不是叢集節點。您是使用管理主控台，透過公用網路或選擇透過網路型終端機集線器，進行遠端存取叢集節點。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)，並使用 `netcon(1M)` 指令連接。

一般而言，您配置沒有監視器的節點。然後，透過管理主控台的 `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，並從終端機集線器連接至節點的串列埠。(如果是 Sun Enterprise E10000 server，您是從「系統服務處理器」連接。請參閱 第25頁的「主控台存取裝置」，以取得詳細資訊。

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，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

請參閱第 4 章，以取得有關管理主控台的問題與解答。

Sun Cluster 拓樸

拓樸是指連接叢集節點和叢集中所使用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，在單一叢集管理框架之下運作。在此配置中，故障轉移僅發生於配對之間。然而，所有的節點是由叢集交互連接來連接的，並且在 Sun Cluster 軟體控制下運作。您可能會使用這種拓樸架構，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故障轉移或可延伸的應用程式。

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。

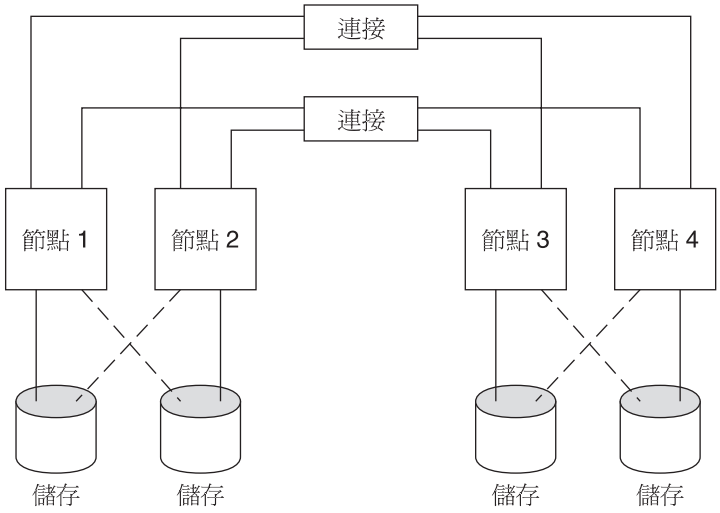

圖 2–3　叢集化配對拓樸架構

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。

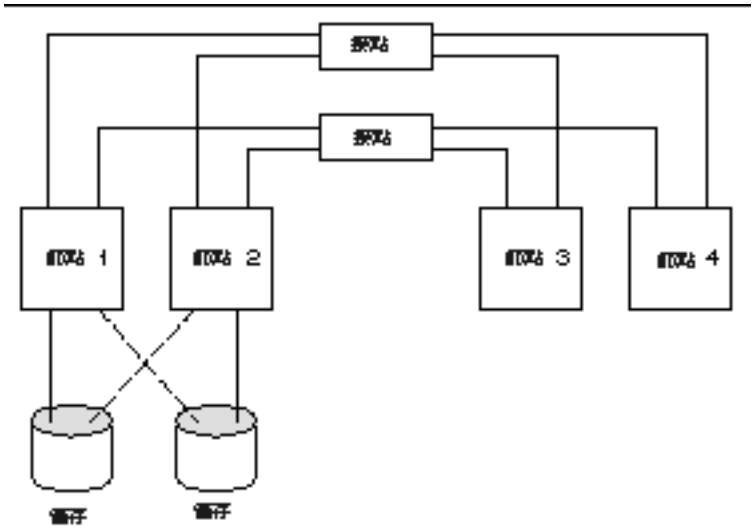

圖 2–4　Pair+M 拓樸

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。主要節點主動提供應用程式服務。在等待主要節點故障時，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故障，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故障時處理負載。

下圖說明 N+1 配置。

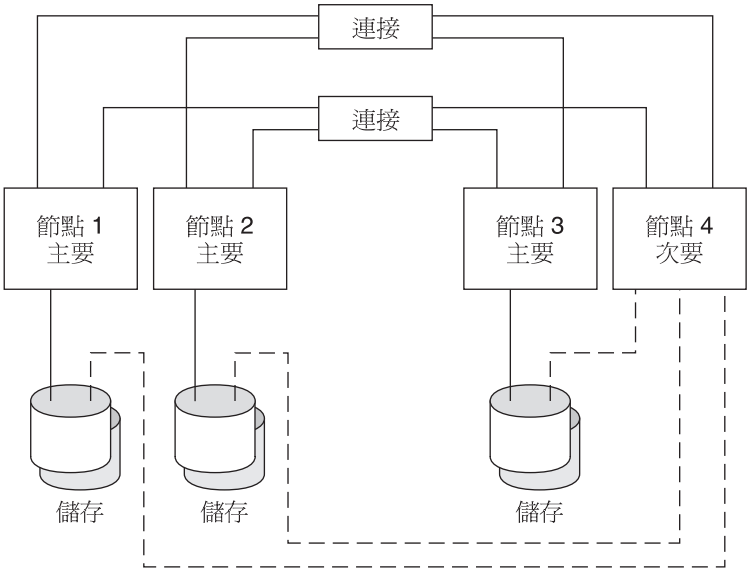

圖 2–5　N+1 拓樸架構

第 3 章

重要概念 – 管理和應用程式設計

本章說明有關 SunPlex 系統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 SunPlex API 和 SDK 的系統管理員和應用程式開發人員參考。叢集系統管理員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。

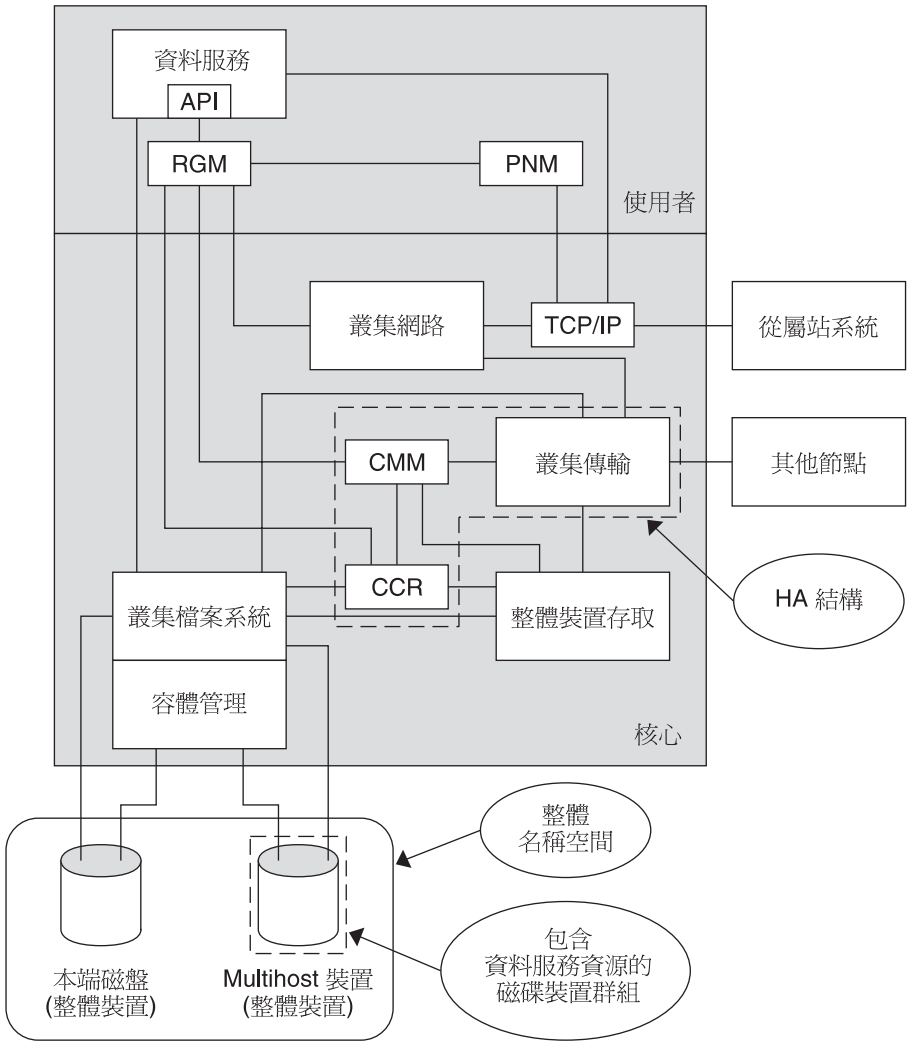

圖 3–1　Sun Cluster 軟體架構

管理介面

您可以選擇要如何從不同的使用者介面來安裝、配置與管理 SunPlex 系統。您可以透過具備說明的指令行介面來完成系統管理作業。在指令行介面的上層，尚有一些可簡化配置作業的公用程式。SunPlex 系統亦有一個模組，是當作 Sun Management Center 的一部分來執行，其提供 GUI 給某些叢集作業。請參照 *Sun Cluster 3.0 U1* 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。不論您是否將叢集節點與任何外在的時間來源同步化，對於叢集操作而言並不重要。SunPlex 系統使用「網路時間通訊協定」(Network Time Protocol，NTP) 來同步化各節點的時鐘。

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 軟體時，安裝程序中的一個步驟是配置叢集的 NTP。Sun Cluster 軟體提供範本檔案，即 `ntp.cluster` (請參閱已安裝叢集節點上的 `/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，已包含於 *Sun Cluster 3.0 U1 安裝手冊* 一書中。

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 U1 系統管理手冊* 中。

高可用性框架

SunPlex 系統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) 故障，叢集元件就具有高度可用性。

下表顯示 SunPlex 元件故障的種類 (硬體和軟體)，以及內建於高可用性框架內的復原種類。

表格3–1 SunPlex 故障偵測與復原的層次

故障的叢集元件	軟體復原	硬體復原
資料服務	HA API, HA 框架	N / A
公用網路配接卡	網路配接卡故障轉移 (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
鏡像的多重主機磁碟	容體管理 (Solstice DiskSuite 和 VERITAS Volume Manager)	硬體 RAID-5 (例如，Sun StorEdge A3x00)
整體裝置	主要與次要複製	至裝置的多重路徑，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 軟體的高可用性框架會快速地偵測到某個節點故障，並且建立一個相等的新伺服器給叢集中剩餘節點上的框架資源。框架資源隨時皆可使用。未受故障節點影響的框架資源，在回復時完全可加以使用。此外，已故障節點的框架資源一經回復之後，便會成爲可使用。已回復的框架資源不必等待所有其他的框架資源完成回復。

大多數可用性頗高的框架資源會回復到使用此資源的應用程式 (資料服務)。框架資源存取的語義學會在各項節點故障時被完整地保留。應用程式無法辨識出框架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，單一節點的故障便是完全透明。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」(Cluster Membership Monitor，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 的主要功能是在任何指定的時間參與叢集的一組節點上，建立全叢集的協議。此限制稱爲 叢集成員。

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43頁的「法定數目和法定裝置」，以取得有關叢集如何保護自己以免於分割成多個個別叢集的詳細資訊。

集成員監視器重新配置

爲了讓資料免於毀損，所有的節點必須對叢集成員達成一致的協議。必要時，CMM 會爲了回應故障而協調叢集服務 (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。

Failfast 機制

如果 CCM 偵測到節點的嚴重問題，它會呼叫叢集框架強制關掉 (混亂的) 節點，並將其從叢集成員中移除。 發生此情況的機制稱爲 *failfast*。Failfast 會導致節點以兩種方式關閉。

- 如果節點離開叢集，並嘗試在無法定數目的情況下開啓新的叢集，它就會被"隔離"而無法存取共用磁碟。請參閱 第46頁的「故障隔離」，以取得有關 failfast 此種用法的細節。
- 如果一或多個叢集特定的常駐程式掛掉 (die)(`clexecd`、 `rpc.pmfd`、`rgmd` 或 `rpc.ed`)，此故障可由 CMM 偵測出來，而節點也跟著混亂。

當由於叢集常駐程式掛掉而產生混亂時，類似下列訊息會顯示在該節點的主控台上。

```
panic[cpu0]/thread=40e60:Failfast:Aborting because "pmfd" died 35 seconds ago. (由於「pmfd」在 35 秒之前掛掉而中斷。)409b8 c
30, 70df54, 407acc, 0) %l0-7:1006c80 000000a 000000a 10093bc 406d3c80 7110340 0000000 4001 fbf0
```

混亂過後，節點可能重新啓動並嘗試重新連接叢集，或停留於 OpenBoot PROM (OBP) 提示處。所採取的行動取決於 OBP 中 `auto-boot?` 參數的設定。

Cluster Configuration Repository (CCR，叢集配置儲存庫)

「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。

OCR 使用兩階段確定演算法作爲更新之用：更新必須在所有叢集成員上成功完成，否則更新會轉返。CCR 使用叢集交互連接來套用分散式更新。

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記錄，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 ，必要時執行復原，以及促使資料的更新。

整體裝置

SunPlex 系統使用整體裝置來提供全叢集、高可用性來存取叢集中的任何裝置 (從任意節點)，並且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故障，Sun Cluster 軟體會自動探尋該裝置的其它路徑並將存取重新導向後的路徑。SunPlex 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。每部伺服器上的本機磁碟亦不是多埠式，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39頁的「全域名稱空間」，以取得詳細資訊。

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，因此多主機磁碟具備高可用性。

裝置 ID (DID)

Sun Cluster 軟體藉由建構裝置 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，而不是透過傳統的 Solaris 裝置 ID，如磁碟的 `c0t0d0`。

這種方式可以確保存取磁碟的任何應用程式 (如容體管理者或使用原始裝置的應用程式) 可以使用一致的叢集存取路徑。這種一致性對多主機磁碟而言特別重要，因爲每個裝置的本機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，以取得詳細資訊。

磁碟裝置群組

在 SunPlex 系統中，所有的多主機磁碟必須是在 Sun Cluster 軟體的控制之下。首先，您要在多主機磁碟上建立容體管理者磁碟群組，如 Solstice DiskSuite 磁碟組，或

是 VERITAS Volume Manager 磁碟群組。然後，將容體管理者磁碟群組註冊爲 碟機裝置群組。磁碟裝置群組是一種整體裝置類型。此外，Sun Cluster 軟體會將各個個別的磁碟註冊爲磁碟裝置群組。

註冊提供 SunPlex 系統資訊，是有關何種節點具有指向哪個容體管理者磁碟群組的路徑。在此，容體管理者磁碟群組會變成可由叢集內做全域存取。如果一個以上的節點可以寫至 (主控) 磁碟裝置群組，儲存在此磁碟裝置群組上的資料就變得高度可用了。高度可用的磁碟裝置群組可用來包含磁碟檔案系統。

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 (代表一群資料服務處理程序)，而另外一個節點則可以主控資料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 U1 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。

有了磁碟裝置群組，容體管理者磁碟群組即變成“全域式”，因爲 它對於基礎的磁碟提供多路徑的支援。實體連接到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。

磁碟裝置群組故障轉移

因爲磁碟機殼連接至一個以上的節點，當目前主控裝置群組的節點故障時，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故障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 (對於應用程式爲透明的)，直到系統恢復使用裝置群組爲止。

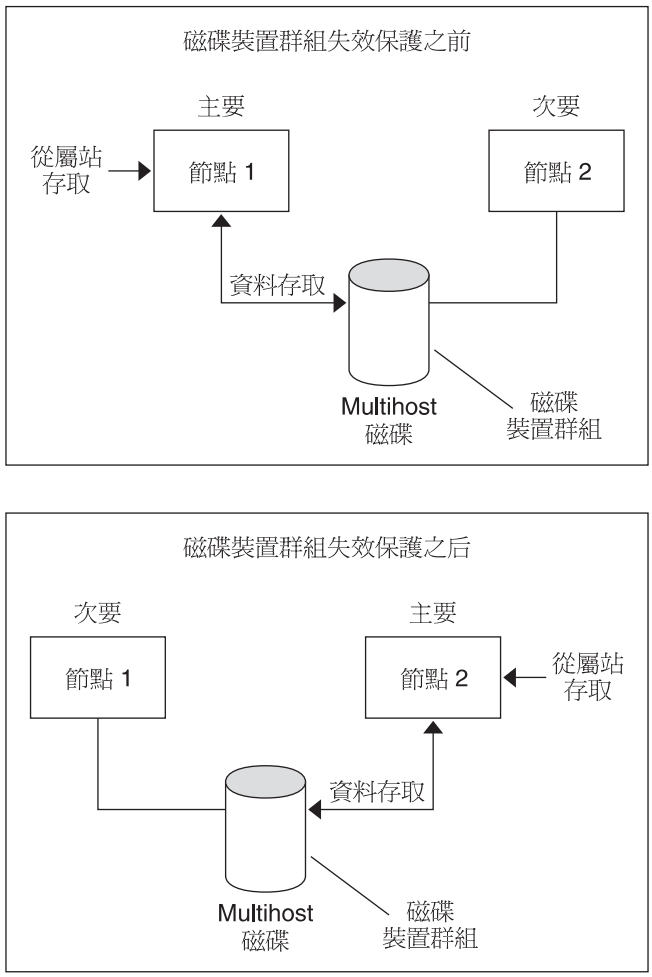

圖 3–2　磁碟裝置群組故障轉移

全域名稱空間

讓啓用整體裝置的 Sun Cluster 軟體機制稱爲全域名稱空間。全域名稱空間包括 `/dev/global/` 階層以及容體管理者名稱空間。全域名稱空間反映多主機磁碟和本機磁碟 (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故障轉移路徑。實際連接多主機磁碟的每一個節點，均提供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SunPlex 系統中，本機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 軟體仍繼續在其標準位置代表容體管理者裝置，例如符號連結。全域名稱空間和標準容體管理者均可由任何叢集節點使用。

全域名稱空間的優點包括：

- 每個節點保持完全獨立，而在裝置管理模型中可有一點變更。
- 裝置可以選擇性地成爲整體。
- 協力廠商連結產生器會繼續運作。
- 給定本機裝置名稱，提供簡易的對應，以獲得其整體名稱。

區域和全域名稱空間範例

下表顯示多主機磁碟 (c0t0d0s0) 的區域和全域名稱空間之間的對應，。

表格3–2 區域和全域名稱空間對應

元件/路徑	本機節點名稱空間	全域名稱空間
Solaris logical name (Solaris 邏輯名稱)	/dev/dsk/c0t0d0s0	/global/.devices/node@*ID* /dev/dsk/c0t0d0s0
DID name (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 Volume Manager	/dev/vx/dsk/*disk-group* /v0	/global/.devices/node@*ID* /dev/vx/dsk/*disk-group*/v0

全域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) 指令來產生全域名稱空間。

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/)，而不管該節點是否實體連接儲存裝置。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。

您可藉由 mount –g 或以 mount 做本機裝載。

應用程式可藉由相同檔名 (例如，/global/foo)，從叢集的任意節點存取叢集檔案系統上的檔案。

叢集檔案系統會裝載於所有叢集成員上。您不能將叢集檔案系統裝載於叢集成員的子集上。

叢集檔案系統並非不同的檔案系統類型。亦即，用戶端可以看見基礎檔案系統 (例如，UFS)。

使用叢集檔案系統

在 SunPlex 系統中，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。

要使叢集檔案系統爲高度可用，基礎的磁碟儲存體必須連結一個以上的節點。因此，納入叢集檔案系統中的本機檔案系統 (儲存在節點本機磁碟上的檔案系統) 就不是高度可用了。

至於一般檔案系統，您可以用兩種方式裝載叢集檔案系統：

- 手動方式—使用 mount 指令以及 –g 或 –o global 裝載選項，從指令行裝載叢集檔案系統，例如：

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

- 自動方式—在具有 global 裝載選項的 /etc/vfstab 檔案中建立項目，以便於啓動時裝載叢集檔案系統。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。以下是來自 /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 軟體對於叢集檔案系統沒有強制的命名策略，您可以建立所有叢集檔案系統的裝載點在同一目錄下，例如 `/global/`*disk-device-group*，以簡化管理作業。請參閱 *Sun Cluster 3.0 U1* 安裝手冊 和 *Sun Cluster 3.0 U1* 系統管理手冊以取得詳細資訊。

叢集檔案系統的功能

叢集檔案系統具備下述功能：

- 檔案存取位置是透明的。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。
- 使用一致的通訊協定來保持 UNIX 檔案存取語意，即使檔案是從多個節點並行地被存取。
- 廣泛的快取是與 zero-copy bulk I/O 移動一起使用，使檔案資料的移動更有效率。
- 叢集檔案系統藉由使用 `fcntl(2)` 介面來提供高度可用的建議檔鎖定功能。藉由使用叢集檔案系統檔案上的建議檔鎖定功能，在多重叢集節點上執行的應用程式便得以同步化資料的存取。檔案鎖可立即由離開叢集的節點，以及維持鎖定時故障的應用程式加以回復。
- 即使發生故障時，仍可確保資料的持續存取。只要磁碟的路徑仍然是作業中，應用程式不會受到故障的影響。這項保證適用於原始磁碟存取和所有的檔案系統作業。
- 叢集檔案系統與基礎檔案系統及容體管理軟體無關。叢集檔案系統使得任何在- 磁碟檔案系統上所支援的都是全域的。

Syncdir 裝載選項

`syncdir` 裝載選項可用於將 UFS 作爲基礎檔案系統的叢集檔案系統。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，您所看到的功能，將會與 UFS 檔案系統相同。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)，便可在寫入作業期間察覺空間不足的狀況。由於您未指定 `syncdir` 就會有問題的情況極少發生，因此我們建議您不要指定，並可享有效能上的利益。

請參閱 第64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。

法定數目和法定裝置

由於叢集節點共用資料與資源，因此很重要的一點是，叢集不可分成個別的且同時作用中的分割區。CMM 保證即使叢集交互連結已做了分割，在任何時刻只會有一個叢集在運作。

來自磁碟分割區的問題有兩種：Split Brain 與 AmnesiaSplit Brain 發生於節點之間的叢集交互連接遺失，以及叢集分割爲子叢集時，每個子叢集都相信自己是唯一的分割區這是叢集節點間的通訊問題所致。Amnesia 發生的時間是，在關機後叢集重新啓動，其中叢集資料比關機時還舊時。這會發生在有多重版本的框架資料儲存於磁碟上，而且新典型的叢集又在無最新版本的時候啓動之時。

藉由給每個節點一票，並強制給予運作中的叢集多數票，便得以避免 Split Brain 與 Amnesia 的狀況發生。有多數票的分割區具有 法定數目，並且是允許運作的。此多數投票機制只要在叢集中有二個以上的節點時，就會運作得極好。在兩個節點的叢集中，票數爲兩票。如果這樣的叢集被分割了，就需要進行外部投票以使其分割區取得法定數目。此外部投票乃由 法定裝置所提供。法定裝置可以是任何在兩節點之間共用的磁碟。用作法定裝置的磁碟可含有使用者資料。

表 3–3 說明 Sun Cluster 軟體如何使用法定數目來避免 Split Brain 與 Amnesia 的發生。

表格3–3 叢集法定數目以及 Split Brain 與 Amnesia 問題

分割區類型	法定數目解決方案
Split Brain	只允許具有多數票的分割區 (子叢集) 作爲叢集來執行 (在這樣的多數票情況下，最多只存在一個分割區)
Amnesia	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)

法定數目演算法是動態運作：當叢集事件觸發計算時，計算結果是可以變更叢集的生命週期。

法定票數

叢集節點與法定裝置會投票以形成法定數目。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來獲得法定票數。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連至法定裝置的節點數。例如，連接至兩個非零票數之節點的法定裝置，擁有一票法定票數 (二減一)。

您是在叢集安裝期間或者稍後，使用 *Sun Cluster 3.0 U1* 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。

法定數目配置

法定數目配置是根據叢集中的節點數而定：

- 雙節點的叢集– 兩個節點的叢集需要兩票法定票數才能選出。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，以確保當某個節點故障時，單一節點可以繼續運作。
- 兩個節點以上的叢集 –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機殼。總共會有五票法定票數，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 軟體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法定投票，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，剩餘的節點仍能作爲叢集來運作。

  請參閱圖 3–3，以取得這些配置的範例。

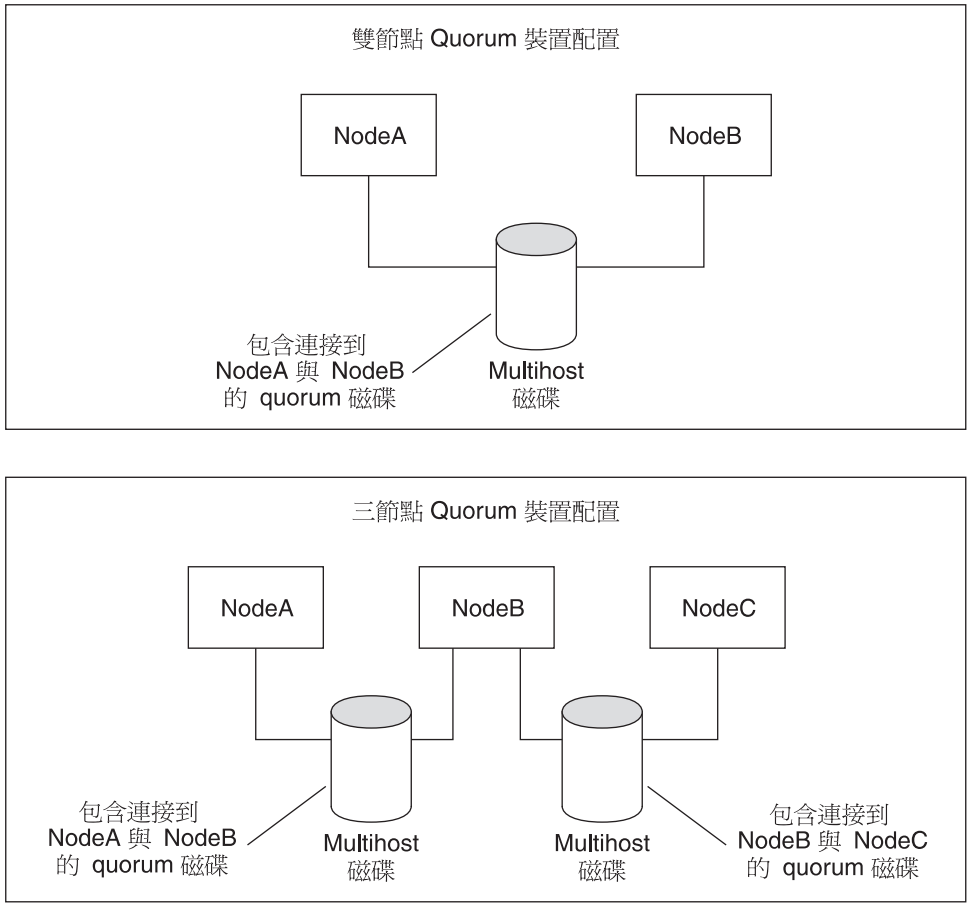

圖 3–3　法定裝置配置範例

法定數目準則

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故障時，其它的節點可以維持法定數目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 U1* 安裝手冊 中的規劃章節以取得詳細資訊。

- 您可以使用包含使用者資料的磁碟來作爲法定裝置。

提示：爲了保護個別法定裝置免於故障，請在各組節點間配置不只一個法定裝置。使用來自不同機殼的磁碟，以及在每組節點之間配置奇數的法定裝置。

故障隔離

叢集的主要問題是造成叢集出現分割的故障 (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。

故障隔離藉由實際防止磁碟存取，來限制節點存取多主機磁碟。當節點離開叢集時 (故障或被分割)，故障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。

磁碟裝置服務提供故障轉移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) 叢集成員故障或無法到達時，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情形。在此處理程序期間，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故障的成員接手控制和存取整體裝置。

SunPlex 系統使用 SCSI 磁碟保留來實作故障隔離。使用 SCSI 保留，故障的節點會“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節點存取共用磁碟。當發生此故障隔離時，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。

偵測到有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。

用於故障隔之 *Failfast* 機制

叢集框架用來確保故障的節點無法重新啓動，並開始寫入至共用儲存體的機制稱爲 *failfast*。

叢集成員的節點對於它們有存取權的磁碟，包括法定數目的磁碟，會連續啓用特定的 ioctl，也就是 MHIOCENFAILFAST。此 ioctl 爲磁碟驅動程式的指示詞，會讓節點在無法存取已被保留爲其它節點之用的磁碟時，有能力自我混亂。

MHIOCENFAILFAST ioctl 會使驅動程式檢查錯誤，而該錯誤是節點爲 `Reservation_Conflict` 錯誤碼讀寫至磁碟所傳回的錯誤。ioctl 會在背景中定期地對磁碟發出測試作業，以檢查 `Reservation_Conflict`。如果傳回 `Reservation_Conflict` 時，前景與背景的控制流程路徑都會混亂。

對於 SCSI-2 磁碟而言，保留並不持久，它們並不能在節點重新啓動時存活。對於具有 Persistent Group Reservation (PGR) 的 SCSI-3 磁碟而言，保留資訊是儲存在磁碟上，並且在節點重新啓動後仍會保留。不管您是否有 SCSI-2 磁碟或 SCSI-3 磁碟，failfast 機制的運作都一樣。

如果節點在叢集中失去與其他節點的連接，並且也不是可達法定容量的分割區，它會被其他節點強制從叢集中移除。。另一可達法定容量之分割區部分的節點，在共用磁碟上放置了保留，且當沒有法定容量的節點嘗試存取共用磁碟時，它會收到保留衝突並且由於 failfast 機制而混亂。

混亂過後，節點可能重新啓動並嘗試重新連接叢集，或停留於 OpenBoot PROM (OBP) 提示處。所採取的行動取決於 OBP 中 `auto-boot?`參數的設定。

容體管理者

SunPlex 系統使用容體管理軟體，藉由鏡像和緊急備用磁碟來增加資料的可用性，以及處理磁碟故障和更換。

SunPlex 系統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 Volume Manager

叢集中的容體管理軟體提供下述支援：

- 節點故障的故障轉移處理
- 不同節點的多重路徑支援
- 遠端透明存取磁碟裝置群組

當容體管理物件爲叢集所控制時，它們就成爲磁碟裝置群組。有關容體管理者的資訊，請參閱您的容體管理者軟體文件。

注意：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 U1 安裝手冊 和 Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide，以取得這些議題的討論資訊。

資料服務

資料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。資料服務是由應用程式、專用 Sun Cluster 配置檔及 Sun Cluster 控制應用程式下列動作的管理方法所組成。

- 啓動
- 停止
- 監控並採行校正措施

圖 3–4 比較在單一應用程式伺服器上執行的應用程式 (單一伺服器模型)，與在叢集上執行的同一應用程式 (叢集伺服器模型)。請注意，就使用者的觀點而言，這兩種配置除了叢集應用程式可能執行較快且較爲高度可用以外，並無任何差別。。

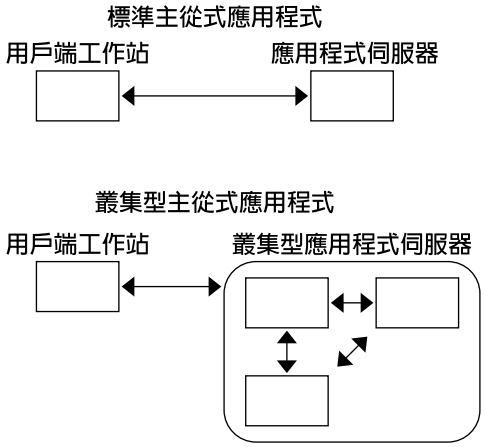

圖 3–4　標準與叢集用戶端 / 伺服器配置

在單一伺服器模型中，您會配置應用程式，以便透過特定的公用網路介面 (主機名稱) 來存取伺服器。主機名稱與實體伺服器有關。

在叢集伺服器模型中，公用網路介面爲邏輯主機名稱或 共用的位址。網路資源一詞是用來表示邏輯主機名稱與共用位址二者。

某些資料服務會要求您指定邏輯主機名稱或共用位址，來做爲網路介面，而它們是不能互相交換的。其他資料服務則容許您指定邏輯主機名稱或共用位址。請參閱各資料服務的安裝與配置檔，以取得您必須指定的介面類型的詳細資訊。

網路資源與特定的實體伺服器無關，它可在實體伺服器間遷移。

網路資源最初與一個稱爲主要的節點關連。如果主要節點故障，網路資源和應用程式資源就會發生故障轉移而移轉至不同的叢集節點 (稱爲次要節點)。當網路資源發生故障轉移時，只要稍有延誤，應用程式資源就繼續在次要節點上執行。

圖 3–5 比較單一伺服器模型與叢集伺服器模型。請注意，在叢集伺服器模型中，網路資源 (在此例中爲邏輯主機名稱) 可於兩或多個叢集節點間移動。應用程式被配置爲使用此邏輯主機名稱，而非與特定伺服器相關的主機名稱。

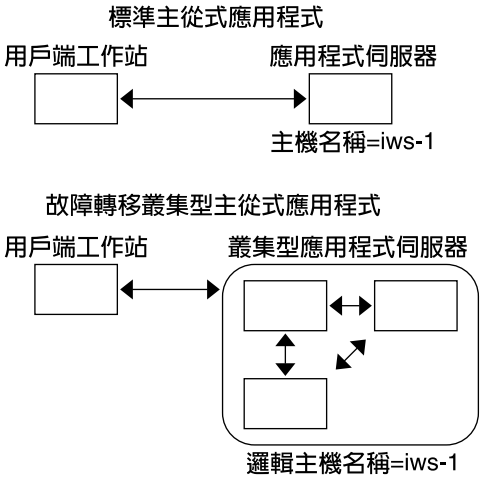

圖 3–5 固定主機名稱與邏輯主機名稱

共用位址最初也與一個節點關連。此節點稱爲「整體介面節點」(Global Interface Node，GIN)。共用位址被用作叢集的單一網路介面。稱之爲 整體介面 *(Global Interface)*。

邏輯主機名稱模型與可延伸服務模型的差異，在於後者的每個節點在其回送介面中亦主動配置有共用位址。此配置可使同時在數個節點上作用中的資料服務具有多重實例。“可延伸的服務”一詞表示，您可藉由新增附加的叢集節點來提供更多 CPU 的能力給應用程式，其效能也隨之延伸。

假如 GIN 故障，共用位址可攜至另一個也正在執行應用程式實例的節點 (因而使此另一節點成爲新的 GIN)。但共用位址也可能發生故障轉移而移轉至另一個先前未執行應用程式的節點。

圖 3–6 比較單一伺服器配置與叢集可延伸服務配置。請注意，在可延伸服務配置中，共用位址存在於所有節點上。與邏輯主機名稱用於移轉資料服務方式類似的是，應用程式被配置爲使用此共用位址，而不是與特定伺服器相關的主機名稱。

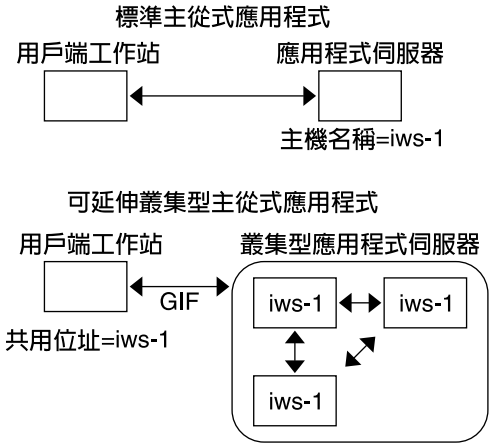

圖 3–6 固定主機名稱與共用位址

資料服務方法

此 Sun Cluster 軟體提供一套服務管理方法。這些方法在 Resource Group Manager (RGM) 的控制之下執行，用來啓動、停止和監視叢集節點上的應用程式。這些方法配合叢集框架軟體和多主機磁碟，讓應用程式變成高可用性資料服務。

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)。

除了 Sun Cluster 軟體提供的方法之外，此 SunPlex 系統亦提供 API 與數個資料服務發展工具。這些工具可以讓程式設計師以此 Sun Cluster 軟體發展出得以使其他應用程式如高可用資料服務般執行的資料服務方法。

Resource Group Manager (RGM)

RGM 控制資料服務 (應用程式) 如同資源，由資源類型施行管理。這些實作是由 Sun 提供或由開發人員以一般資料服務範本、「資料服務發展檔案庫 API」(Data Service Development Library API，DSDL API)，或是「資源管理 API」(Resource Management API，RMAPI) 所建立的。叢集管理者建立並管理在儲存區中的資源，稱爲 資源群組。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。

RGM 作用於資源及資源群組上。RGM 動作能使資源及資源群組在線上及離顯狀態間移動。可應用於資源及資源群組的狀態及設定的完整說明，請參閱 第59頁的「資源及資源群組狀態與設定值」 一節。

故障轉移資料服務

如果正在執行資料服務的節點 (主要節點) 故障，該服務會移轉至其它運作中的節點 不需要使用者介入。故障轉移服務利用 故障轉移資源群組 *(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。

對於故障轉移資料服務，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 (故障轉移)，視資料服務的配置方式而定。

可延伸的資料服務

可延伸的資料服務具有在多重節點上的作用中實例之潛力。可延伸的服務使用兩種資源群組：利用 可延伸的資源群組來包含相關的應用程式資源，以及故障轉移資源群組來包含與可延伸服務相關的網路資源 (共用地址)。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。放置共用位址的故障轉移資源群組一次只在一個節點上啓動成爲線上。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 (整體介面) 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整體介面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故障，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，該服務便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，而不是在整體介面節點上。因此，整體介面節點的故障並不會影響連接。

圖 3–7 顯示的範例是，可延伸服務的故障轉移和可延伸資源群組，以及兩者之間的相依關係。此範例顯示三個資源群組。故障轉移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，可延伸和故障轉移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)。

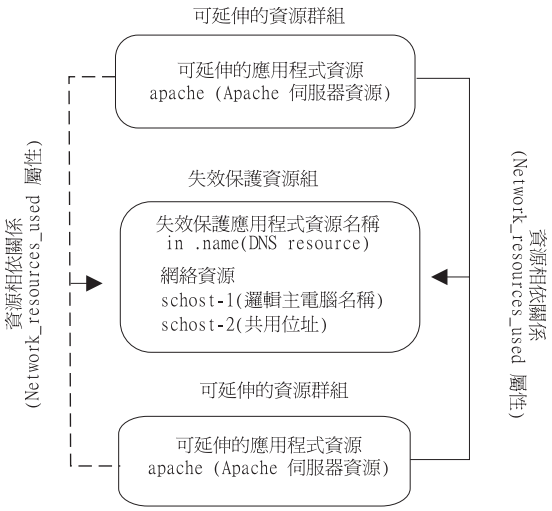

圖 3–7 故障轉移和可延伸資源群組範例

可延伸服務的架構

叢集網路的主要目標是提供資料服務的可延伸性。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，所以資料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資料服務。可延伸資料服務的典型範例是網際網路服務。通常，可延伸資料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，以此服務的遠端用戶端觀點來看，這些實例便可作爲單一的服務，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用戶端的要求。服務此項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，而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
- 網路和資料服務的支援 (使用 Resource Group Manager)

下圖說明了可延伸服務的架構。

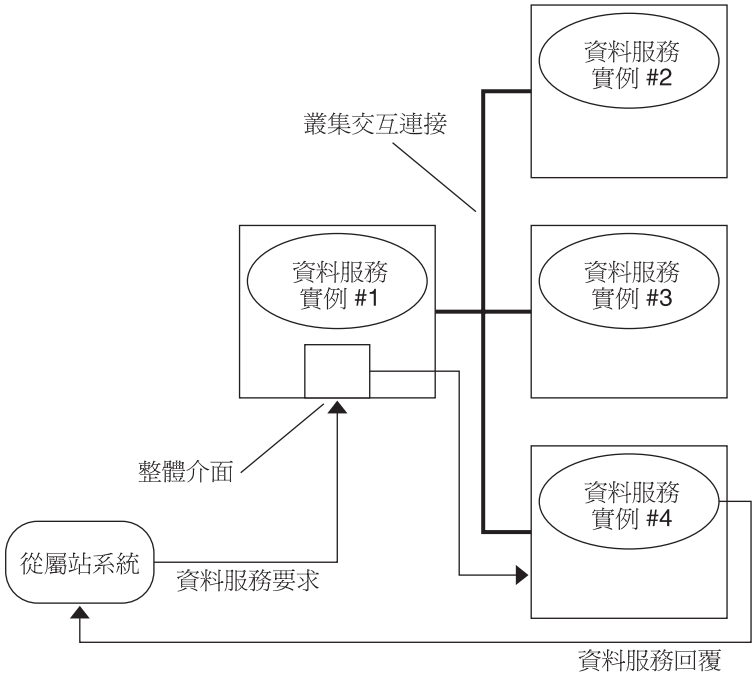

圖 3–8　可延伸服務的架構

沒有放置整體介面的節點 (代理節點) 將共用位址放在其回送介面上。進入整體介面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及產量上增進可延伸服務的效能。

可延伸資料服務的類別有兩種：*pure* 及 *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。那些要求不會重新導向至其它實例。

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，依預設用戶端要求會平均地分配給叢集中的伺服器實例。例如，在三節點的叢集中，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 的任何用戶端要求。權重隨時可以由管理者透過 `scrgadm(1M)` 指令介面或 SunPlex 管理者 GUI 來加以變更。

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。用戶端被稱爲 “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，且當此服務在線上時，平衡資料流量策略未曾改變，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，用戶端上的網際網路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，但是連線是在服務時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，在相同實例上進行幕後交換階段作業資訊。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。此用戶端在相關的同一 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，依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，只要該節點是可存取的，且平衡資料流量策略未變更，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詳細資訊如下。

- 加權式。這項載入會按照指定的加權值來分配到各種節點。此策略是使用 `Load_balancing_weights` 屬性的 `LB_WEIGHTED` 值。如果節點的權重未明確設定時，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：第一個要求送到節點 1，第二個要求送到節點 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。此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通訊埠。此策略已使用 `Load_balancing_policy` 資源屬性的 `LB_STICKY` 值來加以設定。
- Sticky-wildcard。此策略是一般 “Sticky” 策略的超集合。對於以 IP 位址來識別的可延伸服務而言，是由伺服器來指定通訊埠 (而且事先無法知道)。通訊埠可能會變更。此策略已使用 `Load_balancing_policy` 資源屬性的 `LB_STICKY_WILD` 值來加以設定。

故障回復設定

資源群組因故障轉移，從某個節點移轉至另一個節點。發生此情況時，原來的次要節點就成爲新的主要節點。故障回復設定指定當原來的主要節點回到線上時會採取的動作。此選項是要使原來的主要節點再次成爲主要節點 (故障回復) 或維持目前的主要節點。您可使用故障回復資源群組屬性設定來指定您要的選項。

在某些情況下，假如放置資源群組的原始節點重複故障和重新開機，設定故障回復可能會造成資源群組的可用性降低。

資料服務錯誤監視器

每個 SunPlex 資料服務均提供了錯誤監視器，定期地測試資料服務以判斷其運作狀況。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故障轉移。

開發新的資料服務

Sun 提供配置檔案與管理方法範本，讓您得以使各種應用程式在叢集中以故障轉移或可延伸的服務來運作。如果您要當作故障轉移或可延伸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故障轉移或可延伸的服務。

有一套準則可用來斷定應用程式是否可成爲故障轉移服務。特定的基準規則在 SunPlex 文件中有所說明，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資料服務架構。請參閱 第51頁的「可延伸的資料服務」 一節，以取得有關可延伸服務的一般資訊。

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，可能需要改寫些某些部份，使服務能夠符合準則。

可延伸資料服務具有下列特性。首先，服務是由一或多個伺服器實例所組成。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個或多個實例。

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做並行存取時，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 的儲存體與"邏輯"，因爲儲存體是以單一實體呈現，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屬性。

SunPlex 系統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。在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。

這類型的儲存體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。

資料服務 API 與資料服務檔案庫 API

SunPlex 系統提供下列項目，可以使應用程式具備高可用性：

- 資料服務是以作爲 SunPlex 系統的一部分提供
- 資料服務 API
- 資料服務發展檔案庫 API
- "一般"資料服務

Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide 說明如何安裝和配置SunPlex 系統提供的資料服務。 *Sun Cluster 3.0 U1 Data Services Developer's Guide* 說明如何導入其它應用程式，以便在 Sun Cluster 框架下具備高可用性。

這些 Sun Cluster API 讓應用程式設計師能發展啓動及停止資料服務實例的錯誤監視器及程序檔。利用這些工具，應用程式可以變成具備故障轉移和可延伸資料服務。此外，SunPlex 系統可提供“一般”資料服務，可用來快速產生應用程式需要的啓動和停止方法，使其作爲高可用性資料服務來執行。

使用資料服務通訊的叢集交互連接

叢集在節點之間必須具備多網路連接，以形成叢集交互連接。叢集軟體可使用多重交互連接來達到高可用性以及增進效能。對於內部通訊 (例如，檔案系統資料或可延伸服務資料)，訊息是以輪流的方式分送到所有可用的交互連接。

叢集交互連接也可以用於應用程式，以便在節點之間建立高可用性通訊。例如，分散式應用程式可能會有元件在多個需要通訊的節點上執行。如果使用叢集交互連接而不是公用傳輸，可以防制個別連結的故障。

要在節點之間使用叢集交互連接進行通訊，應用程式必須使用安裝叢集時配置的專用主機名稱。例如，如果節點 1 的專用主機名稱是 `clusternode1-priv`，請使用該名稱當作節點 1 的叢集交互連接的通訊。 使用這個名稱開啓的 TCP 插槽可在叢集交互連接中被傳遞 (route)，如果網路故障還可以再被傳遞。

請注意，由於專用主機名稱可以在安裝時配置，因此叢集交互連接可使用當時選取的任何名稱。可使用 `scha_privatelink_hostname_node` 引數來從 `scha_cluster_get`(3HA) 取得實際名稱。

在應用程式層次使用叢集交互連接時，每一對節點之間使用單一的交互連接，但若可能的話，不同的節點配對之間應使用個別的交互連接。例如，考慮到有應用程式在三個節點上執行，而且透過叢集交互連接來進行通訊的狀況。節點 1 與 2 之間的通訊可能透過 `hme0` 介面，節點 1 與 3 之間的通訊則可能透過介面 `qfe1`。也就是說，任意二個節點之間的應用程式通訊將限制於單一交互連接，內部叢集通訊則散置在所有的交互連接。

請注意，應用程式和內部叢集通訊共用交互連接，因此應用程式可用的頻寬是由其他叢集通訊所使用的頻寬來決定。在發生故障時，內部通訊可以在其餘交互連接中循環 (round-robin)，而在故障的交互連接上的應用程式也可以切換到運作的交互連接。

有兩種類型的位址支援叢集交互連接，而 `gethostbyname`(3N) 上的專用主機名稱通常會傳回兩個 IP 位址。第一個位址稱爲 邏輯 *pairwise* 位址，第二個位址稱爲邏輯 *pernode* 位址。

每一對節點會被指派個別的邏輯 pairwise 位址。這個小型邏輯網路支援連接的故障轉移。每一個節點還會被指派一個固定的 pernode 位址。也就是說，每一個節點上的 `clusternode1-priv` 的邏輯 pairwise 位址都不一樣，而每一個節點上的

clusternode1-priv 的邏輯 pernode 位址則都相同。節點本身沒有 pairwise 位址，因此節點 1 上的 gethostbyname (clusternode1-priv) 將只傳回邏輯 pernode 位址。

請注意，接受透過叢集交互連接之通訊，並依安全理由而驗證 IP 位址的應用程式，必須針對 gethostbyname 傳回的所有 IP 位址進行檢查，而不只是針對第一個 IP 位址。

如果您要求應用程式在各個點都是一致的 IP 位址，請將應用程式配置爲在用戶端以及伺服器都是連結到 pernode 位址，這樣所有的連接看起來都會是透過 pernode 位址往來。

資源、資源群組與資源類型

資料服務利用了數種類型的資源。應用程式諸如Apache Web Server或 iPlanet Web Server 使用應用程式所賴以運作的網路位址 (邏輯主機名稱及共用位址)。應用程式和網路資源形成受 RGM 管理的基本單位。

資料服務式資源類型。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。

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

網路資源爲 SUNW.LogicalHostname 或 SUNW.SharedAddress 資源類型。有兩種資源類型是由 Sun Cluster 軟體預先登錄。

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。它可確保在資料服務啓動之前，叢集檔案系統裝載點的路徑、整體裝置和裝置群組名稱是可用的。

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。如果在資源群組上啓動了故障轉移或切換保護移轉，則資源群組會被當作一個單位來遷移。

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。資料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，與資料服務錯誤監視器判斷資料服務是否仍在服務用戶端的方式相同。請參照 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide* 以取得此處理程序的詳細資訊。

資源及資源群組狀態與設定值

管理者將靜態設定值套用到資源與資源群組中。這些設定值只可經由管理動作來變更。RGM 在動態"狀態"間移動資源群組。這些設定值與狀態的說明列於下述清單中。

- 管理或不管理– 這些都是只套用在資源群組上的全叢集設定值。資源群組是由 RGM 所管理。 `scrgadm(1M)` 指令可讓 RGM 來管理或不管理資源群組。這些設定值不會隨著叢集再配置而變更。

  在建立第一個資源群組時，它是不被管理的。若要讓群組中的任何資源成爲作用的，它就必須是被管理的。

  在某些資料服務中，諸如可延伸式 Web 伺服器，在設定網路資源前與停止網路資源後都有工作要做。此工作是由 initialization (INIT) 及 finish (FINI) 資料服務方法來達成。INIT 方法只有在資源所在的資源群組在被管理狀態時才會執行。

  當資源群組由不管理移向管理的狀態時，任何用於群組已註冊的 INIT 方法都會在群組的資源上執行。

  當資源群組由管理移向不管理的狀態時，任何已註冊的 INIT 方法都會被呼叫以執行清除。

  INIT 及 FINI 方法的最常使用方式是用於可延伸服務的網路資源，但也可用於應用程式不做的初始化或清除工作。

- 啓用或停用–這些都是套用到資源的全叢集設定值。`scrgadm(1M)` 指令可用來啓用或停用資源。這些設定值不會隨著叢集再配置而變更。

  資源的正常設定值爲，它在系統中是啓用且主動執行的。

  如因某種原因，您要使資源在所有叢集節點上爲不可用的，您可停用此資源。停用的資源不作爲一般用途。

- 線上或離線–這些都是套用到資源與資源群組的動態狀態。

  在切換保護移轉或故障轉移時，當叢集由叢集在配置步驟間轉換時，會改變這些狀態。亦可經由管理動作來加以變更。 `scswitch(1M)` 可用來變更資源或資源群組的線上或離線狀態。

  在任何時間，故障轉移資源或資源群組只能在一個節點上爲線上。可延伸的資源或資源群組可在數個節點上爲線上，而在其他節點上爲離線。在切換保護移轉或故障轉移時，資源群組及其內的資源會在一個節點上離開線上，並在另一節點再次連到線上。

  假如一個資源群組爲離線的，則它所有的資源均爲離線的。假如一個資源群組爲線上的，則它所有的資源均爲線上的。

資源群組含有數種資源，在各資源間具有相依性。這些相依性要求資源要以特定次序連到線上及離開線上。連到線上及離開線上的方法，可能對於各個資源會花費不同的時間。因有資源相依性及開始與結束的時間差異，在叢集重新配置時單一資源群組內的資源會有不同的線上及離線狀態。

資源和資源群組屬性

您可以配置您的 SunPlex 資料服務的資源和資源群組之屬性值。標準屬性常見於所有資料服務中。延伸屬性則特定於個別的資料服務。部份標準和延伸屬性是以預設值配置的，所以您不需要修改它們。其它屬性則需要在建立和配置資源時加以設定。各資料服務的文件會指定可設定哪些資源屬性，及設定的方式。

標準屬性是用來配置通常與任何特定資料服務無關的資源和資源群組屬性。 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide* 的附錄中說明了這組標準屬性。

延伸屬性提供了諸如應用程式二進位檔案、配置檔案和資源相依項目位置的資訊。您要依照資料服務的配置方式來修改延伸屬性。 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide* 中有個別關於資料服務的章節說明這組延伸屬性。

公用網路管理 (PNM) 和網路配接卡故障轉移 (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、以及在偵測到故障時將 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網路配接卡故障轉移群組 (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故障轉移的備份配接卡。故障轉移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。因爲故障轉移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，但是在故障轉移期間的短暫延遲除外。

注意：因爲 TCP 的壅塞回復特性，TCP 端點在故障轉移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故障轉移期間遺失，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詳細資訊，請參閱 *Sun Cluster 3.0 U1 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故障。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。Sun Cluster 軟體檢視邏輯和共用 IP 位址，這些被視爲受 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 錯誤偵測和故障轉移處理程序

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，這個事件會觸發故障轉移以備份配接卡。

輸入與輸出封包計數器由 PNM 監督，因此 當任一或二者皆有一段時間維持不被更動時，就啓動了 ping 序列。

Ping 序列包含測試 `ALL_ROUTER` 廣播位址 `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，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 U1 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，故障轉移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。

當成功完成了 IP 位址的故障轉移時， 就會送出無償的 (gratuitous) ARP 廣播。也就維持了與遠端 用戶端的連接。

第 4 章

常見問題

本章包含有關 SunPlex 系統最常見問題的解答。問題是依照主題來排列。

高可用性常問問題

- 到底什麼是高可用性系統？

  SunPlex 系統將高可用性 (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？

  藉由故障轉移的處理程序，叢集框架提供高可用性的環境。故障轉移是叢集所執行的一系列步驟，可將應用程式從故障節點移轉至叢集中的另一個可作業節點上。

- 故障轉移與可延伸的資料服務之間的差異爲何？

  高可用性的資料服務有兩類，亦即故障轉移和可延伸。

  故障轉移資料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。如果主要節點故障，在故障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。

  對於各個應用程式，一個節點擁有叢集的實體介面。此節點稱爲「整體介面節點」(Global Interface Node，GIN) 。叢集中可有多重的 GIN。每一 GIN 皆擁有一或多個可供可延伸服務使用的邏輯介面。這些邏輯介面稱爲 整體介面。一個 GIN 擁有

對於特定應用程式所有要求的整體介面，並派送這些要求至應用程式伺服器正在執行的多重節點上。假如 GIN 故障，則整體介面發生故障轉移而移轉至存活節點上。

如果應用程式所執行的任一節點故障，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，直到故障節點返回叢集之後才改善。

檔案系統常問問題

- **<emphasis role = "strong">**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？**</emphasis>**

  不，不要做回送裝載。

- 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，應用程式需要在其執行的節點故障時以手動方式重新啓動。

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

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

- 使用叢集檔案系統和匯出 **NFS** 檔案系統之間的差異是什麼？

  有多處的差異：

  1. 叢集檔案系統支援整體裝置。NFS 不支援遠端存取裝置。
  2. 叢集檔案系統擁有全域名稱空間。只需要一個裝載指令。至於 NFS，您必須在每一個節點載設檔案系統。
  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故障轉移。NFS 支援多重伺服器，但是故障轉移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。

- **/global/.devices/<node>@<node ID>** 此檔案系統出現在我的叢集節點上。我可使用此系統檔，以儲存我想要讓其爲高可用及整體的資料嗎？

這些系統檔會儲存整體裝置的名稱空間。它們不供一般使用。當它們爲整體時，從不以整體方式存取，每一節點只存取自己的整體裝置的名稱空間。假如節點當機了，其他節點就無法存取當機節點的名稱空間。這些檔案系統不具高可用性。它們不應用來儲存需爲整體或高可用的資料

容體管理常問問題

- 是否需要鏡像所有的磁碟裝置？

  對於要作爲高可用性的磁碟裝置，必須要進行鏡像，或使用 RAID-5 硬體。所有的資料服務應該使用高可用性磁碟裝置，或裝載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故障。

- 我可對本機磁碟 **(**開機磁碟**)** 使用一個容體管理者，而對多重主機磁碟使用不同的容體管理者嗎？

  此配置乃由管理本機磁碟的 Solstice DiskSuite 軟體與管理多重主機磁碟的 VERITAS Volume Manager 所支援。但並不支援其他組合。

資料服務常問問題

- 可用的**SunPlex** 資料服務是什麼呢？

  支援的資料服務清單包含於 *Sun Cluster 3.0 U1* 版次注意事項。

- **SunPlex** 資料服務所支援的應用程式版本爲何？

  支援的應用程式版本清單包含於 *Sun Cluster 3.0 U1* 版次注意事項。

- 我是否可寫入自己的資料服務？

  可以。請參閱 *Sun Cluster 3.0 U1 Data Services Developer's Guide* 及 Data Service Development Library API 所提供的「Data Service Enabling Technologies」文件，以取得詳細資訊。

- 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？

除去 Sun Cluster HA for NFS 的情況外，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故障轉移資料服務，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 資源以及故障轉移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。於是可延伸和故障轉移服務均可使用 SharedAddress 資源中配置的同一組主機名稱/位址。

公用網路常問問題

- **SunPlex** 系統支援何種網路配接卡？

  目前，SunPlex 系統支援 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。

- 在故障轉移中 **MAC** 位址扮演的角色是什麼？

  發生故障轉移時，會產生新的「位址解析度通訊協定 (Address Resolution Protocol，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。

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

  不，不支援此變數。

- 當 **NAFO** 在作用中與備份的配接卡之間執行切換保護移轉時，能延遲多久？

  延遲可以達數分鐘。這是因爲當做了 NAFO 切換保護移轉時，牽涉到送出免費的 ARP。然而，並不保證用戶端和叢集間的路由器將使用免費的 ARP。因此，直到路由器上此 IP 位址的 ARP 快取項目逾時，還是有可能使用舊的 MAC 位址。延遲的第二個原因是兩個 NAFO 配接卡均連接到 Ethernet 切換器。當做了 NAFO 切換保護移轉時，NAFO 配接卡的其中一個在第二個配接卡爲開啓時會是關閉的。Ethernet 切換器現在必須停用連接埠，並啓用不同的連接埠，而這可能得花些時間。另外，有了 Ethernet，在切換器和新啓用的配接卡之間就有速度溝通的問題產

生。最後，在做了切換保護移轉之後，NAFO 會對啓用的配接卡做最小程度的檢查，以確定一切運作正常。

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43頁的「法定數目和法定裝置」，以取得 Amnesia 的詳細資訊)。例如，如果節點 2 是法定裝置的所有者，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像本機磁碟？

  可以。雖然這種鏡像並非必要，但鏡像叢集節點的磁碟可以排除非鏡像磁碟故障而導致節點當機的情況。鏡像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。請勿連接此節點至共用磁碟。

  請參閱 *Sun Cluster 3.0 U1* 系統管理手冊，以取得有關備份和復原程序的詳細資訊。

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，是因爲有了鏡像 (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。

叢集交互連接常問問題

- **SunPlex** 系統支援何種叢集交互連接？

  目前 SunPlex 系統支援 Ethernet (100BASE-T Fast Ethernet 和 1000BASE-SX Gb) 叢集交互連接。

- **“**電纜**”**和傳輸**“**路徑**”**有何不同？

  叢集傳輸電纜是使用傳輸配接卡和切換器來配置的。電纜是以元件對元件方式連接配接卡和切換器。叢集拓樸管理者使用可用的電纜來建立節點之間的點對點傳輸路徑。電纜並不會直接對應至傳輸路徑。

  電纜是由管理者做靜態的“啓用”和“停用”。電纜有“狀況，” (啓用或停用)，但非“狀態。”如果電纜是啓用的，其就如同尚未配置。停用的電纜無法用作傳輸路徑。由於電纜不是探測式的，所以無法得知它們的狀態。電纜的狀況可使用 `scconf –p` 來檢視.

  傳輸路徑並非由叢集拓樸管理者動態建立的。傳輸路徑的“狀態”是由拓樸管理者決定。路徑的狀態可以是“線上”或“離線”。傳輸路徑的狀態可使用 `scstat(1M)` 來檢視。

  請考慮下述具四條電纜的兩個節點叢集範例。

```
node1:adapter0     to switch1, port0 node1:adapter1     to switch2, port0 node2:adapter0     to
switch2, port1
```

  有兩個可能的傳輸路徑可由這四條電纜形成。

```
node1:adapter0    to node2:adapter0 node2:adapter1    to node2:adapter1
```

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，視資料服務應用程式而定，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資料服務應用程式。請參閱 *Sun Cluster 3.0 U1 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詳細資訊。

管理主控台常問問題

- **SunPlex** 系統需要管理主控台嗎？

  可以。

- 管理主控台必須專屬於叢集，或者可以用於其它作業嗎？

  SunPlex 系統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

- **SunPlex** 系統需要終端機集線器嗎？

  所有以 Sun Cluster 3.0 爲始的軟體版本不需要終端機集線器來執行。不似 Sun Cluster 產品需要終端機集線器作爲故障隔離之用，之後的產品並不依靠終端機集線器。

- 我發現多數的 **SunPlex** 伺服器需要終端機集線器，而 **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。

- 使用終端機集線器有些什麼樣的好處？

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，而不需重新將它啓動？

  可以。請注意，連接埠號碼需要重設並執行下述：

```
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 U1* 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詳細資訊。

- 萬一終端機集線器本身故障，要怎麼辦？我必須要有另一個備用的嗎？

  不需要。如果終端機集線器故障，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員所使用的小型網路，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

這個詞彙表的名詞解釋用於 SunPlex 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。

自動故障回復 (automatic failback)
在主要節點故障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「網路配接卡故障轉移群組」。

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故障轉移、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)**
提供全叢集及高可用性的叢集服務 存取現有的本機檔案系統。

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 和資料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)**
在同一節點上的屬性。在叢集配置期間會使用這個概念來改進效能。

D

資料服務 **(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)**
故障轉移資源類型啓動所在的預設叢集成員。

裝置群組 **(device group)**
裝置資源的使用者定義群組，例如磁碟， 可從叢集 HA 配置中不同節點來 控制。此群組可以包含磁碟、Solstice DiskSuite 磁碟組和 VERITAS Volume Manager 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 邏輯名稱之間的關聯性。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。假如以前在叢集中未曾看到裝置 ID，便會指定新的 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 軟體製作的驅動程式，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 SunPlex 系統中的整體裝置。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，以復原這類故障。

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、嚴重程程度或說明有變更。

F

故障回復 (failback) 請參閱「自動故障回復」。

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。

故障轉移 (failover) 發生故障之後，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。

故障轉移資源 **(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)**
用來測試資料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)**
資料服務的範本。可以用一般資源類型將簡單的應用程式變成具故障轉移的資料服務 (在某個節點停止時，會在另一個節點啓動)。這種類型不需要 SunPlex 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 資料服務 **(HA data service)**
請參閱「資料服務」。

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故障轉移。通往叢集成員的全部路徑故障時，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)** 僅適用可延伸的服務。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。請參閱 第51頁的「可延伸的資料服務」，以取得詳細資訊。

平衡資料流量策略 **(load-balancing policy)** 僅適用可延伸的服務。應用程式要求負載分散至各節點的偏好方式。請參閱 第51頁的「可延伸的資料服務」，以取得詳細資訊。

本機磁碟 **(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 SunPlex 系統已經不存在。請參閱 第37頁的「磁碟裝置群組」 和 第58頁的「資源、資源群組與資源類型」，以取得現在此概念在 SunPlex 系統中是如何實作的說明。

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 軟體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」。

複合裝置狀態資料庫複製 **(replica，metadevice state database replica)** 儲存於磁碟上的資料庫，記錄所有複合裝置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。

多重主目錄主機 **(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

網路配接卡故障轉移 **(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，爲了在配接卡故障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 SunPlex 系統之一部份的實體機器或網域 (在 Sun Enterprise E10000 server 內)。亦稱爲「主機」。

非叢集模式 **(non-cluster mode)**
結果狀態是透過啓動叢集成員 (具有 –x 啓動選項) 達成的。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)**
在主要節點故障時，能夠主控故障轉移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)**

使用錯誤監視器和故障轉移，來防止因爲單一網路配接卡或電纜故障而造成的節點可用性遺失的軟體。PNM 故障轉移使用一組網路配接卡 (稱爲「網路配接卡故障轉移」群組) 來提供叢集節點與公用網路之間的備用連接。錯誤監視器和故障轉移功能一起確保資源的可用性。亦請參閱「網路配接卡故障轉移群組」。

Q

法定裝置 **(quorum device)**

由兩個或更多節點所共用的磁碟，以那些節點所擁有的票數來建立叢集執行的法定數目。只有具有可用的法定票數，叢集方能運作。法定裝置的使用時機，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時。

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。

資源群組管理者 **(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。發生硬體或軟體故障或重新開機時，RGM 會依照預先配置的策略來運作。

資源群組狀態 **(resource group state)**

任何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)**

SunPlex 系統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。

資源狀況 **(resource state)**

指定節點上的 Resource Group Manager 資源狀況。

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)**
指定給資料服務、LogicalHostname 或 SharedAddress 叢集物件的唯一名稱。資料服務資源類型可以是故障轉移類型或可延伸類型。亦請參閱「資料服務」、「故障轉移資源」和「可延伸資源」。

資源類型屬性 **(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程用戶端。

可延伸的服務 **(scalable service)**
實作成爲可同時執行於多個節點的資料服務。

次要 **(secondary)**
發生主要節點故障時，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 名稱如： `/devices/sbus@1f,0/SUNW,fas@e,8800000/sd@6,0:c,raw`

亦請參閱「Solaris 邏輯名稱」。

Solstice DiskSuite SunPlex 系統所使用的容體管理者。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「故障回復」。

切換保護移轉 **(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)。切換保護移轉是由管理者使用 `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「故障轉移」。

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 Volume Manager SunPlex 系統所使用的容體管理者。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、鏡像及複合裝置或容體的動態成長來提供資料可靠性的軟體產品。