

SUSE Linux Enterprise High Availability Extension

11 SP1

www.novell.com

2010 年 4 月 22 日

High Availability 指南



High Availability 指南

版權所有 © 2006- 2010 Novell, Inc.

根據自由軟體基金會 (Free Software Foundation) 所發佈的 GNU 自由文件授權 (GNU Free Documentation License) 1.2 版或更新版本，使用者可以複製、散佈與/或修改本文件；「恆常章節」為此著作權聲明與授權。本節中包含授權「GNU 自由文件授權」的一份副本。

SUSE®、openSUSE®、openSUSE® 標幟、Novell®、Novell® 標幟、N® 標幟 是 Novell, Inc. 在美國和其他國家/地區的註冊商標。Linux* 是 Linus Torvalds 的註冊商標。所有其他協力廠商商標，為各所有人所有之財產。商標符號 (®、™ 等)代表 Novell 的商標；星號 (*) 代表協力廠商的商標。

本手冊中所有資訊在編輯時，都已全力注意各項細節。但這不保證百分之百的正確性。因此，Novell, Inc.、SUSE LINUX Products GMBH、作者或譯者都不需對任何錯誤或造成的結果負責。

目錄

關於本指南	ix
I 安裝與設定	1
1 產品綜覽	3
1.1 主要功能	3
1.2 優點	6
1.3 叢集組態：儲存	9
1.4 結構	11
2 開始使用	15
2.1 硬體要求	15
2.2 軟體需求	16
2.3 共享磁碟系統需求	16
2.4 準備	16
2.5 綜覽：安裝及設定叢集	17
3 使用 YaST 的安裝與基本設定	19
3.1 安裝 High Availability Extension	19
3.2 初始叢集設定	20
3.3 連線叢集	27
3.4 使用 AutoYaST 進行大規模部署	28

II	組態與管理	31
4	組態與管理基礎	33
4.1	全域叢集選項	33
4.2	叢集資源	35
4.3	資源監控	45
4.4	資源限制	46
4.5	如需更多資訊	51
5	設定和管理叢集資源 (GUI)	53
5.1	Pacemaker GUI — 綜覽	54
5.2	設定全域叢集選項	56
5.3	設定叢集資源	57
5.4	管理叢集資源	76
6	設定和管理叢集資源 (指令行)	83
6.1	crm 指令行工具 — 綜覽	83
6.2	設定全域叢集選項	89
6.3	設定叢集資源	90
6.4	管理叢集資源	101
7	使用 Web 介面管理叢集資源	105
7.1	啟動 HA Web Konsole 及登入	106
7.2	使用 HA Web Konsole	107
7.3	疑難排解	108
8	新增或修改資源代辦	109
8.1	STONITH 代辦	109
8.2	撰寫 OCF 資源代辦	110
8.3	OCF 傳回代碼與失敗復原	111
9	圍籬區隔與 STONITH	113
9.1	圍籬區隔的類別	113
9.2	節點層級圍籬區隔	114
9.3	STONITH 組態	116
9.4	監控圍籬區隔設備	120
9.5	特殊圍籬區隔設備	120
9.6	如需更多資訊	122

10	Linux Virtual Server 的負載平衡	123
10.1	概念綜覽	123
10.2	使用 YaST 設定 IP 負載平衡	125
10.3	更多設定	131
10.4	如需更多資訊	131
11	網路設備 Bonding	133
11.1	使用 YaST 設定 Bonding 設備	133
11.2	如需更多資訊	135
III	儲存與資料複製	137
12	Oracle Cluster File System 2	139
12.1	特點及優勢	139
12.2	OCFS2 套件與管理公用程式	140
12.3	設定 OCFS2 服務	141
12.4	建立 OCFS2 磁碟區	143
12.5	掛接 OCFS2 磁碟區	145
12.6	如需更多資訊	146
13	分散式複製區塊設備 (DRBD)	147
13.1	概念綜覽	147
13.2	安裝 DRBD 服務	149
13.3	設定 DRBD 服務	150
13.4	測試 DRBD 服務	153
13.5	調整 DRBD	155
13.6	DRBD 疑難排解	155
13.7	如需更多資訊	157
14	叢集 LVM	159
14.1	概念綜覽	159
14.2	cLVM 的組態	159
14.3	明確設定適合的 LVM2 設備	167
14.4	如需更多資訊	168
15	儲存保護	169
15.1	基於儲存區的圍籬區隔	169
15.2	確保啟動獨佔性儲存	174

16 Samba 叢集	177
16.1 概念綜覽	177
16.2 基本組態	178
16.3 對叢集化 Samba 進行除錯與測試	180
16.4 如需更多資訊	182
 IV 疑難排解與參考	 183
 17 疑難排解	 185
17.1 安裝問題	185
17.2 「除錯」HA 叢集	186
17.3 常見問題集	187
17.4 獲取詳細資訊	189
 18 叢集管理工具	 191
 19 HA OCF Agents	 243
 V 附錄	 335
 A 設定簡單測試資源的範例	 337
A.1 使用 GUI 設定資源	337
A.2 手動設定資源	339
 B 將叢集升級到產品的最新版本	 341
B.1 從 SLES 10 升級到 SLEHA 11	341
B.2 從 SLEHA 11 升級到 SLEHA 11 SP1	345
 C 最新功能	 347
C.1 版本 10 SP3 至版本 11	347
C.2 版本 11 至版本 11 SP1	350
 D GNU 授權	 353
D.1 GNU General Public License	353
D.2 GNU Free Documentation License	356

關於本指南

SUSE® Linux Enterprise High Availability Extension 是一個採用開放原始碼叢集技術的整合式套裝軟體，它可讓您實作高可用性的實體和虛擬 Linux 叢集。為進行快速有效的組態設定和管理，High Availability Extension 中提供有圖形使用者介面 (GUI) 和指令行介面 (CLI)。此外還隨附了 HA Web Konsole，可讓您透過 Web 介面管理 Linux 叢集。

本指南適用於需要安裝、設定及維護 High Availability (HA) 叢集的管理員。文中對兩種方法 (GUI 與 CLI) 均有詳細說明，以幫助管理員選擇滿足其執行主要任務需求的適當工具。

本指南分為以下幾個部分：

安裝與設定

在開始安裝及設定叢集之前，請先熟悉叢集的基礎知識與結構，瞭解一下其主要功能與優點。掌握須滿足的硬體與軟體要求，及在執行下一步操作之前需做的準備工作。使用 YaST 執行 HA 叢集的安裝與基本設定。

組態與管理

使用圖形使用者介面 (Pacemaker GUI) 或 `crm` 指令行介面新增、設定及管理資源。如果要透過 Web 介面監控叢集，請使用 HA Web Konsole。瞭解如何利用負載平衡與圍籬區隔如果您要撰寫自己的資源代辦或修改現存資源代辦，請取得有關如何建立不同類型資源代辦的背景資訊。

儲存與資料複製

SUSE Linux Enterprise High Availability Extension 隨附叢集感知檔案系統 (Oracle 叢集檔案系統 (OCFS2)) 與磁碟區管理員 (叢集邏輯磁碟區管理員 (cLVM))。複製資料時，會使用 DRBD (分散式複製區塊設備) 將 High Availability 服務的資料從叢集的主動節點鏡像複製到待機節點。此外，叢集化的 Samba 伺服器還針對異質環境提供了 High Availability 解決方案。

疑難排解與參考

若要管理自己的叢集，您需要執行一定的疑難排解作業。瞭解最常見的問題及其修復方法。請尋找 High Availability Extension 提供之指令行工具的完整參考，管理自己的叢集。

附錄

列出最新版 High Availability Extension 的新功能和行為變更。瞭解如何將叢集移轉到最新版本，看一下設定簡單測試資源的範例。

本手冊的許多章節包含連到其他文件資源的連結。包括系統和網際網路上所提供的其他文件。

如需適用於產品的文件與最新文件更新的綜覽，請參閱 <http://www.novell.com/documentation>。

1 意見反應

以下為可供使用的數種意見回應管道：

錯誤與增強功能要求

有關適用於產品的服務與支援選項，請參閱 <http://www.novell.com/services/>。

若要報告關於某個產品元件的錯誤，請使用 <http://support.novell.com/additional/bugreport.html>。

若要提交增強功能要求，請造訪 <https://secure-www.novell.com/rms/rmsTool?action=ReqActions.viewAddPage&return=www>。

使用者意見

我們希望得到您對本手冊以及本產品隨附之其他文件的意見和建議。請使用線上文件每頁底部的「使用者備註」功能，或造訪 <http://www.novell.com/documentation/feedback.html> 在其中輸入您的意見。

2 文件慣例

本手冊使用下列印刷慣例：

- `/etc/passwd`：目錄名稱與檔名
- `placeholder`：以實際的值來取代 `placeholder`

- **PATH**: 環境變數 **PATH**
- **ls**、**--help**: 指令、選項和參數
- **user**: 使用者或群組
- **Alt**、**Alt + F1**: 供人按下的按鍵或案件組合；顯示的按鍵與鍵盤上一樣為大寫
- 「檔案」、「檔案」>「另存新檔」: 功能表項目、按鈕
- ► **amd64 em64t**: 本段僅與指定的結構有關。箭頭標示了文字區塊的開頭與結尾。 ◀
- *Dancing Penguins* (章節 *Penguins*, ↑其他手冊): 這是對其他手冊某章節的參考。

I. 安裝與設定

產品綜覽

SUSE® Linux Enterprise High Availability Extension 是一個採用開放原始碼叢集技術的整合式套裝軟體，它可讓您實作高可用性的實體和虛擬 Linux 叢集，並可避免單一故障點。該套裝軟體可確保資料、應用程式和服務等重要網路資源的高可用性及其可管理性。因此，可協助您保持業務持續運作，保護資料完整性，並可降低關鍵任務 Linux 工作負載的意外停機時間。

它提供了基本的監控、訊息傳送和叢集資源管理功能，支援個別受管理叢集資源的容錯移轉、錯誤回復和移轉 (負載平衡)。High Availability Extension 做為 SUSE Linux Enterprise Server 11 SP1 的附加產品提供。

本章介紹 High Availability Extension 產品的主要功能以及優點。您會看到幾個範例叢集，並瞭解組成叢集的各個元件。最後一節所敘述的是該架構的綜覽，介紹了叢集內的個別架構層和程序。

High Availability 叢集內容中使用的一些常見詞彙，可以在術語 [第361頁] 中找到相關說明。

1.1 主要功能

SUSE® Linux Enterprise High Availability Extension 可協助您確保和管理網路資源的可用性。下面幾節重點介紹了部分主要功能：

1.1.1 眾多叢集情境

High Availability Extension 支援以下情境：

- 主動/主動組態
- 主動/被動組態：N+1、N+M、N 到 1、N 到 M
- 混合式實體和虛擬叢集，允許虛擬伺服器與實體伺服器叢集在一起。這能提高服務的可用性和資源的使用率。

叢集最多可以包含 16 個 Linux 伺服器。可以使用叢集內的任一伺服器重新啟動同一叢集中之失敗伺服器中的資源 (應用程式、服務、IP 位址和檔案系統)。

1.1.2 靈活性

High Availability Extension 附帶 Corosync/OpenAIS 訊息傳送與成員層及 Pacemaker 叢集資源管理員。使用 Pacemaker，管理員可以持續監控其資源的狀態，管理相依性，並能根據可靈活設定的多種規則自動停止和啟動服務。High Availability Extension 可讓您對叢集進行特定應用程式和硬體架構的調整以符合貴組織的需求。時間相關組態可讓服務在指定時間自動移轉回已修復節點。

1.1.3 儲存與資料複製

使用 High Availability Extension，您可以視需要動態指定和重新指定伺服器儲存。它支援光纖通道或 iSCSI 儲存區域網路 (SAN)。還支援共享磁碟系統，但這些系統並非必要系統。SUSE Linux Enterprise High Availability Extension 還隨附了叢集感知檔案系統 - Oracle 叢集檔案系統 (OCFS2) 和磁碟區管理員 - 叢集化邏輯磁碟區管理員 (cLVM)。若要複製您的資料，可以使用 DRBD (分散式複製區塊設備，Distributed Replicated Block Device) 將叢集主動節點上的 High Availability 服務資料鏡像複製到叢集的待機節點。此外，SUSE Linux Enterprise High Availability Extension 還支援一種用於 Samba 叢集的技術：CTDB (叢集化簡單資料庫，Clustered Trivial Database)。

1.1.4 支援虛擬化環境

SUSE Linux Enterprise High Availability Extension 支援包含實體與虛擬 Linux 伺服器的混合叢集。SUSE Linux Enterprise Server 11 SP1 附帶了 Xen (一種開放原始碼虛擬化監管程式) 和 KVM (核心虛擬機器, Kernel-based Virtual Machine), 後者是 Linux 系統的虛擬化軟體, 以硬體虛擬化延伸為基礎。High Availability Extension 中的叢集資源管理員可辨識、監控和管理虛擬伺服器以及實體伺服器中執行的服務。叢集可以將訪客系統做為服務進行管理。

1.1.5 資源代辦

SUSE Linux Enterprise High Availability Extension 含有大量資源代辦, 用以管理 Apache、IPv4、IPv6 等眾多資源。它還內建有適用於 IBM WebSphere Application Server 等廣受歡迎的協力廠商應用程式的資源代辦。如需產品隨附的開放叢集架構 (OCF) 資源代辦的清單, 請參閱第 19 章「*HA OCF Agents*」[第243頁]。

1.1.6 簡單易用的管理工具

High Availability Extension 附帶一組功能強大的工具, 可用於叢集的基本安裝和設定, 以及有效的組態設定和管理:

YaST

用於一般系統安裝和管理的圖形使用者介面。使用它可以在 SUSE Linux Enterprise Server 的基礎上安裝 High Availability Extension, 如第 3.1 節「安裝 High Availability Extension」[第19頁] 中所述。YaST 在 High Availability 類別中還包含以下模組, 可以幫助您設定叢集或個別元件:

- 叢集: 基本叢集設定。如需詳細資訊, 請參閱第 3.2 節「初始叢集設定」[第20頁]。
- DRBD: 分散式複製區塊設備的組態。
- IP 負載平衡: Linux Virtual Server 負載平衡的組態。如需詳細資訊, 請參閱第 10 章「*Linux Virtual Server 的負載平衡*」[第123頁]。

Pacemaker GUI

用於簡化叢集組態設定和管理的可安裝圖形使用者介面。可引導您完成資源建立和組態設定的整個過程，讓您執行啟動、停止或移轉資源等管理任務。如需詳細資訊，請參閱第 5 章「設定和管理叢集資源 (GUI)」[第53頁]。

HA Web Konsole

一款網路使用者介面，使用它您在非 Linux 機器上也能管理 Linux 叢集。如果您的系統不提供圖形使用者介面，它也是一個理想的解決方案。如需詳細資訊，請參閱第 7 章「使用 Web 介面管理叢集資源」[第105頁]。

crm

功能強大的統一指令行介面。協助您設定資源，以及執行所有監控或管理任務。如需詳細資訊，請參閱第 6 章「設定和管理叢集資源 (指令行)」[第83頁]。

1.2 優點

High Availability Extension 可讓您最多將 16 台 Linux 伺服器設定成一個 High Availability 叢集 (HA 叢集)，其中的資源可以動態切換或移動至叢集中的任一伺服器。資源可以設成在遇到伺服器故障時自動進行移轉，也可以選擇手動移動以對硬體進行疑難排解或平衡工作負載。

High Availability Extension 利用商用元件提供高可用性。透過將應用程式和作業整合到一個叢集可降低成本。High Availability Extension 還可讓您集中管理整個叢集並調整資源以滿足不斷變化的工作負載要求 (從而手動使叢集達到「負載平衡」)。允許具有兩個以上節點的叢集還可透過允許幾個節點共享一個「熱備用」來節省成本。

它還具有另一個同等重要的優點，就是可以潛在地縮短計畫外服務的中斷運作時間，以及為了執行軟體和硬體的維護與升級所需的計劃內中斷運作時間。

您希望實作叢集的理由包括：

- 增加可用性
- 改善效能
- 降低作業成本

- 可調適性
- 災害復原
- 資料保護
- 伺服器整合
- 儲存整合

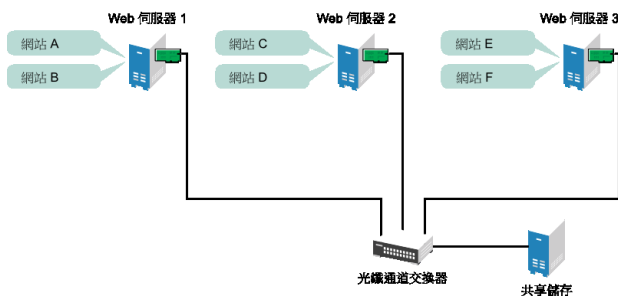
藉由在共享磁碟子系統上建置 RAID，可以達到磁碟容錯共享。

以下案例說明 High Availability Extension 具備的一些優點。

範例叢集案例

假設您已經設定一個含有三台伺服器的叢集，而這三台伺服器上都已安裝了 Web 伺服器。叢集中的每個伺服器都代管兩個網站。每個網站上的所有資料、圖形以及網頁內容都儲存在與叢集中的每一個伺服器相連接的共享磁碟子系統上。以下的圖解可以描繪這個設定的可能外觀。

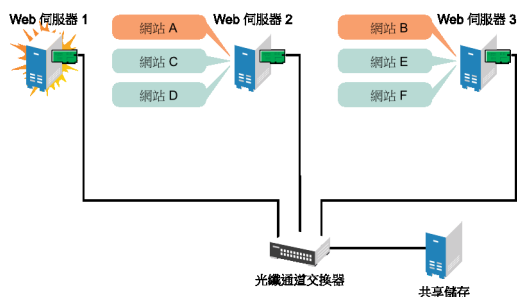
圖形 1.1 由三台伺服器所構成的磁簇



正常的磁簇作業期間，每一伺服器都會和磁簇中的其他伺服器保持通訊，並且定期檢查所有已登錄的資源以偵測是否有故障發生。

假設 Web 伺服器 1 發生硬體或軟體問題，且利用 Web 伺服器 1 進行網際網路存取、收發電子郵件和資訊的使用者失去連線。下圖說明當 Web Server 1 故障時，資源移動的情形。

圖形 1.2 一台伺服器故障後，由三台伺服器所構成的磁簇



網站 A 將移至 Web 伺服器 2，而網站 B 將移至 Web 伺服器 3。IP 位址和證書也會移至 Web 伺服器 2 和 Web 伺服器 3。

當您進行叢集設定時，可以決定發生故障時每一台 Web 伺服器上代管之網站的移動目的地。在前面的範例中，將網站 A 設定為移至 Web 伺服器 2，而將網站 B 設定為移至 Web 伺服器 3。這樣，之前由 Web 伺服器 1 處理的工作負載仍可用，並且會在所有正常運行的叢集成員之間平均分散。

Web 伺服器 1 失敗時，High Availability Extension 軟體將執行以下動作：

- 偵測故障並向 STONITH 確認 Web 伺服器 1 確實已停止運行。STONITH 是「Shoot The Other Node In The Head」的縮寫，它的用途是關閉行為異常的節點，以免在叢集中產生問題。
- 將先前掛接於 Web 伺服器 1 上的共享資料目錄重新掛接於 Web 伺服器 2 和 Web 伺服器 3。
- 在 Web 伺服器 2 和 Web 伺服器 3 上重新啟動先前於 Web 伺服器 1 上執行的應用程式。
- 將 IP 位址傳送至 Web 伺服器 2 和 Web 伺服器 3。

在此範例中，容錯移轉程序會快速完成，而使用者也將在數秒內重新恢復存取網站資訊。大部份的情況下，無需再次登入。

現在假設 Web 伺服器 1 所發生的問題已經解決，並且它已恢復到正常的作業狀態。此時，網站 A 和網站 B 可以自動錯誤回復 (移回) 到 Web 伺服器 1，也可以保留在現有伺服器上。這取決於您為它們設定資源的方式。將服務移轉回 Web 伺服器 1 會導致一定的停機時間，因此，High Availability Extension 也可讓您將

移轉延遲一段時間，等到不會使服務運行中斷或只會讓服務短時間中斷時再進行移轉。兩種備選方法各有優缺點。

High Availability Extension 還提供了資源移轉功能。您可以根據系統管理需要，將應用程式、網站等移轉至叢集中的其他伺服器。

例如，您可以手動將網站 A 或網站 B 從 Web 伺服器 1 移至叢集中的任一其他伺服器。您也許希望藉此對 Web Server 1 進行升級或執行排程維護的工作，或僅為了增加網站效能或連線能力。

1.3 叢集組態：儲存

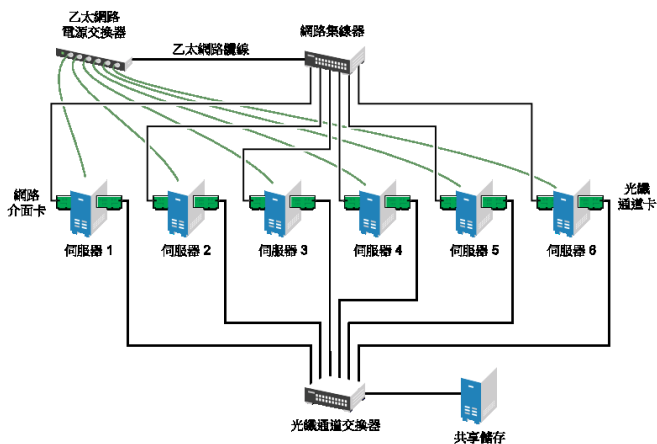
具有 High Availability Extension 的叢集組態可能包含，也可能不包含共享磁碟子系統。共享磁碟子系統可以透過高速光纖通道卡、纜線和交換器進行連接，也可設定為使用 iSCSI。如果有一台伺服器失敗，叢集中的另一台指定伺服器就會自動掛接先前掛接於失敗伺服器上的共享磁碟目錄。這使得網路使用者得以繼續存取共享磁碟子系統上的目錄。

重要：具有 cLVM 的共享磁碟子系統

使用具有 cLVM 的共享磁碟子系統時，必須將該子系統連接至叢集中需要從中存取它的所有伺服器。

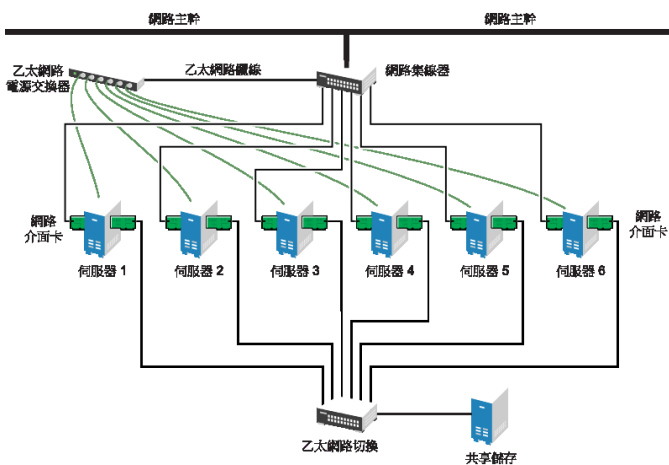
一般資源包括資料、應用程式和服務。下圖顯示一般光纖通道叢集組態的可能外觀。

圖形 1.3 一般光纖通道叢集組態



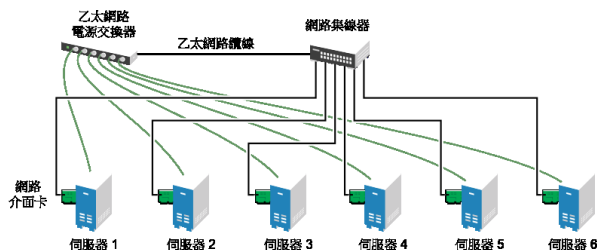
雖然光纖通道提供的效能最佳，但您仍可以將叢集設定為使用 iSCSI。iSCSI 是光纖通道的替代方案，可用於建立低成本的儲存區域網路 (SAN)。下圖顯示一般 iSCSI 叢集組態的可能外觀。

圖形 1.4 一般 iSCSI 叢集組態



雖然大部分叢集都包含共享磁碟子系統，但也可以建立不含共享磁碟子系統的叢集。下圖顯示不含共享磁碟子系統之叢集的可能外觀。

圖形 1.5 不含共享儲存的一般叢集組態



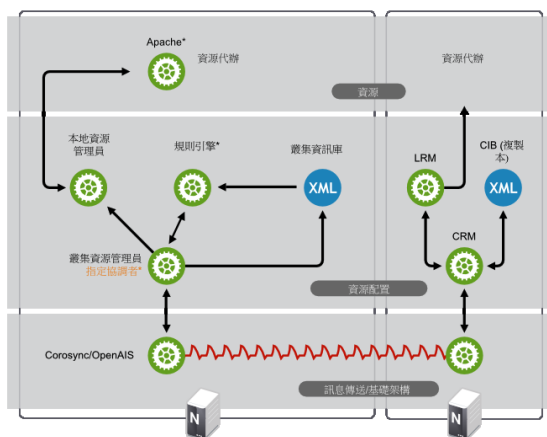
1.4 結構

本節提供 High Availability Extension 結構的簡要綜覽。它識別並提供有關結構元件的資訊，並說明這些元件如何相互操作。

1.4.1 結構層

High Availability Extension 採用分層結構。圖形 1.6 「結構」[第12頁] 說明不同的層及其相關的元件。

圖形 1.6 結構



訊息傳送與基礎架構層

主要層或第一層為訊息傳送/基礎架構層，也稱為 Corosync/OpenAIS 層。此層包含送出含有「I'm alive」(我存在)訊號及其他資訊之訊息的元件。High Availability Extension 的程式存放在訊息傳送/基礎架構層。

資源配置層

下一層為資源配置層。此層最為複雜，由下列元件組成：

叢集資源管理員 (CRM)

在資源配置層中執行的每個動作均透過叢集資源管理員進行傳遞。如果資源配置層的其他元件 (或更高一層中的元件) 需要進行通訊，通訊將透過本地 CRM 完成。

在每個節點上，CRM 維護叢集資訊庫 (CIB) [第13頁]，它包含所有叢集選項、節點、資源、它們的關係和目前狀態的定義。將會選出叢集中的一個 CRM 做為指定協調者 (DC)，這意味著它擁有主要 CIB。叢集中的所有其他 CIB 則擁有主要 CIB 的複製本。CIB 上的一般讀寫作業透過主要 CIB 序列

化。DC 是叢集中可以決定是否需要執行全叢集變更 (例如圍籬區隔節點或移動資源) 的唯一實體。

叢集資訊庫 (CIB)

叢集資訊庫是整個叢集組態和目前狀態的記憶體內部 XML 表示。它包含所有叢集選項、節點、資源、限制及相互關係的定義。CIB 還可同步更新至所有叢集節點。叢集中有一個主要 CIB，由 DC 負責維護。其他所有節點均包含一個 CIB 複製本。

規則引擎 (PE)

指定協調者需要進行整個叢集範圍的變更 (對新的 CIB 做出反應) 時，規則引擎會根據目前狀態和組態計算出叢集的下一個狀態。PE 還可產生轉換圖表，該圖表包含 (資源) 動作和相依性清單，用以取得下一個叢集狀態。PE 將在每個節點上執行，以提高 DC 容錯移轉速度。

本地資源管理員 (LRM)

LRM 代表 CRM 呼叫本地資源代辦 (請參閱章節「資源層」[第13頁])。因此，它可以執行啟動/停止/監控作業，並將結果報告給 CRM。它還可以隱藏資源代辦的受支援程序檔標準 (OCF、LSB、Heartbeat 版本 1) 之間的差異。LRM 是本地節點上所有資源相關資訊的管理來源。

資源層

最高層為資源層。資源層包含一或多個資源代辦 (RA)。資源代辦是用於啟動、停止和監控特定種類的服務 (資源) 的程序 (通常是外圍程序檔)。資源代辦僅可由 LRM 呼叫。協力廠商可以將他們自己的代辦包含在檔案系統中的已定義位置，從而為他們自己的軟體提供即裝即用的叢集整合功能。

1.4.2 程序流程

SUSE Linux Enterprise High Availability Extension 使用 Pacemaker 做為 CRM。CRM 被當做精靈 (crmd) 來實作，即在每個叢集節點上都有一個例項。Pacemaker 會選出一個 crmd 例項做為主要例項，以此來集中所有叢集決策。如果所選的 crmd 程序 (或它所在的節點) 失敗，則會建立一個新的 crmd 程序。

可反映出叢集組態及叢集中所有資源的目前狀態的 CIB 會保留在每個節點上。CIB 的內容將在整個叢集中自動保持同步。

在叢集中執行的許多動作都會導致全叢集發生變更。這些動作包括新增或移除叢集資源，或變更資源限制等。執行此類動作時，必須瞭解叢集中會發生什麼狀況。

例如，假設您要新增叢集 IP 位址資源。為此，您可以使用指令行工具或 GUI 其中之一來修改 CIB。不必在 DC 上執行動作，您可以使用叢集中的任何節點上的其中一種工具，動作即會轉送至 DC。然後，DC 會將 CIB 變更複製到所有叢集節點。

接著，PE 將根據 CIB 中的資訊計算出叢集的理想狀態，以及達到該狀態的方式，並會將一系列指示饋送至 DC。DC 透過訊息傳送/基礎架構層送出指令，而這些指令將由其他節點上的對等 crmd 接收。每個 crmd 皆使用其 LRM (當做 lrmd 實作) 執行資源修改。lrmd 為非支援叢集，它可直接與資源代辦 (程序檔) 互動。

對等節點都會將其作業結果回報給 DC。一旦 DC 確定已在叢集中成功執行了所有必要的作業，叢集就會轉回閒置狀態，等待接下來的事件。如果有任何作業未按計畫執行，則會用 CIB 中記錄的新資訊再次呼叫 PE。

在某些情況下，可能需要關閉節點以保護共享資料或完成資源復原。為執行此操作，Pacemaker 提供了圍籬區隔子系統 stonithd。STONITH 是「Shoot The Other Node In The Head」的縮寫，通常使用遠端電源交換器實作。在 Pacemaker 中，STONITH 設備會模型化為資源 (並在 CIB 中進行設定)，以便輕鬆地對它們進行故障監控。不過，stonithd 負責瞭解 STONITH 拓樸，因此，其用戶端只需要圍籬區隔一個節點，其餘的工作則由 stonithd 完成。

開始使用

在下文中，您將瞭解到有關系統要求的資訊以及在安裝 High Availability Extension 之前需要進行的準備工作，並可檢視安裝和設定叢集的基本步驟的簡要綜覽。

2.1 硬體要求

以下清單指定以 SUSE® Linux Enterprise High Availability Extension 為基礎之叢集的硬體要求。這些要求代表最低硬體組態。依您打算如何使用叢集而定，您可能還需要其他硬體。

- 1 至 16 部安裝了第 2.2 節「軟體需求」[第16頁] 中所指定之軟體的 Linux 伺服器。伺服器不需要配備完全一樣的硬體 (記憶體、磁碟空間等)。
- 至少兩個 TCP/IP 通訊媒體。叢集節點使用多路廣播進行通訊，因此，網路設備必須支援多路廣播。通訊媒體應支援 100 Mbit/s 或更高的資料速率。最好結合乙太網路通道。
- 選擇性：連接至叢集 (需要從該叢集存取磁碟子系統) 內所有伺服器的共享磁碟子系統。
- STONITH 機制。STONITH 是「Shoot the other node in the head」的縮略字。STONITH 設備是一種電源交換器，叢集可使用它重設被視為當機或運作有問題的節點。重設無活動訊號的節點是確保掛起節點或只是呈現為已停止動作的節點不會損毀資料的唯一可靠方法。

若需更多資訊，請參考第 9 章「圍籬區隔與 STONITH」[第113頁]。

2.2 軟體需求

確定下列軟體需求都已符合：

- 在將要屬於叢集的所有節點上安裝了 SUSE® Linux Enterprise Server 11 SP1 及所有可用的線上更新。
- 在將要屬於叢集的所有節點上安裝了 SUSE Linux Enterprise High Availability Extension 11 SP1，包括所有可用的線上更新。

2.3 共享磁碟系統需求

如果您希望資料具有高度可用性，建議您使用共享磁碟系統(儲存區域網路，即 SAN)。如果使用共享磁碟子系統，請注意下列事項：

- 已根據製造商的指示正確設定共用磁碟系統，而且可以正常運作。
- 共享磁碟系統中包含的磁碟應設定為使用鏡像複製或 RAID，以增強共享磁碟系統的容錯能力。建議使用硬體型 RAID。系統並不是對所有組態都支援主機型軟體 RAID。
- 如果使用 iSCSI 存取共用磁碟系統，必須確定已正確設定 iSCSI 啟動程式和目標。
- 當您使用 DRBD 來實作在兩部機器之間配送資料的鏡像 RAID 系統時，請確認僅存取複製的設備。請與叢集的其餘部分使用相同 (繫結) 的 NIC，以利用該處所提供的備援。

2.4 準備

安裝 High Availability Extension 之前，請執行下列準備步驟：

- 編輯叢集中每部伺服器上的 `/etc/hosts` 檔案，以設定主機名稱解析並使用靜態主機資訊。如需詳細資訊，請參閱 <http://www.novell.com/documentation> 上的「*ifup* 管理指南」。詳情位於「基本網路」>「設定主機名稱和 DNS」一章。

叢集的成員必須能夠藉由名稱相互找到對方，這是一項基本要求。如果名稱不可用，內部叢集通訊將會失敗。

- 透過將叢集節點與叢集外部的時間伺服器同步，設定時間同步。如需詳細資訊，請參閱 <http://www.novell.com/documentation> 上的「*ifup* 管理指南」。詳情位於「使用 NTP 進行時間同步化」一章。

叢集節點將使用時間伺服器做為其時間同步來源。

2.5 綜覽：安裝及設定叢集

完成準備工作後，需要執行下列基本步驟以便使用 SUSE® Linux Enterprise High Availability Extension 安裝並設定叢集：

1. 在 SUSE Linux Enterprise Server 的基礎上以附加產品的形式安裝 SUSE® Linux Enterprise Server 及 SUSE® Linux Enterprise High Availability Extension。如需詳細資訊，請參閱第 3.1 節「安裝 High Availability Extension」[第 19 頁]。
2. 初始叢集設定 [第 20 頁]
3. 連線叢集 [第 27 頁]
4. 設定全域叢集選項並新增叢集資源。

兩者均可透過圖形使用者介面 (GUI) 或指令行工具完成。如需詳細資訊，請參閱第 5 章「設定和管理叢集資源 (GUI)」[第 53 頁] 或第 6 章「設定和管理叢集資源 (指令行)」[第 83 頁]。

5. 若要透過圍籬區隔和 STONITH 保護資料，以避免可能的損毀，請確認將 STONITH 設備設定為資源。如需詳細資訊，請參閱第 9 章「圍籬區隔與 STONITH」[第 113 頁]。

根據您的要求，可能還需要為叢集設定下列檔案系統以及與儲存相關的元件：

- 在共享磁碟上建立檔案系統 (儲存區域網路, SAN)。如有必要，將那些檔案系統設定為叢集資源。
- 如果需要叢集感知的檔案系統，請使用 OCFS2。

- 若要讓叢集管理與邏輯磁碟區管理員共享的儲存區，請使用 **cLVM** (LVM 的一組叢集延伸功能)。
- 若要保護資料的完整性，可以透過使用圍籬區隔機制以及確保獨佔式儲存區存取的方式來執行儲存區保護。
- 如有需要，可以透過 **DRBD** 進行資料複製。

如需詳細資訊，請參閱第 III 部分「儲存與資料複製」[第137頁]。

使用 YaST 的安裝與基本設定

安裝 High Availability 叢集所需軟體的方法有兩種：使用 `zypper` 從指令行安裝，或使用提供圖形使用者介面的 YaST 進行安裝。在所有要納入叢集的節點上安裝軟體後，下一步就是對叢集進行初始設定(以便節點之間可以相互通訊)，以及開啟連線叢集所需的服務。叢集的初始設定可以透過編輯和複製組態檔案的方式手動進行，也可以使用 YaST 叢集模組來進行。

本章介紹如何重新安裝並設定 SUSE Linux Enterprise High Availability Extension 11 SP1。如果要移轉執行舊版 SUSE Linux Enterprise High Availability Extension 的現有叢集，或是更新現行叢集中節點上的軟體套件，請參閱附錄 B 將叢集升級到產品的最新版本 [第341頁] 一章。

3.1 安裝 High Availability Extension

High Availability 安裝模式中包含使用 High Availability Extension 設定與管理叢集所需的套件。此模式只在 SUSE® Linux Enterprise High Availability Extension 做為附加產品安裝後才能使用。如需有關安裝附加產品的資訊，請參閱 <http://www.novell.com/documentation> 中的《SUSE Linux Enterprise 11 SP1 部署指南》。詳情位於「安裝附加產品」一章。

注意：安裝軟體套件

High Availability 叢集所需的軟體套件不會自動複製到叢集節點。

如果不想在所有將納入叢集的節點上手動安裝 SUSE® Linux Enterprise Server 11 SP1 與 SUSE® Linux Enterprise High Availability Extension 11 SP1，請使用

AutoYaST 複製現有節點。如需詳細資訊，請參閱第 3.4 節「使用 AutoYaST 進行大規模部署」[第28頁]。

過程 3.1 安裝 *High Availability Extension* 模式

- 1 以 root 使用者身分啟動 YaST，然後選取「軟體」>「軟體管理」。

或者在指令行上使用 `yast2 sw_single` 以 root 身分啟動 YaST 套件管理員。

- 2 從「過濾器」清單中選取「模式」，然後啟用模式清單中的「*High Availability*」模式。
- 3 按一下「接受」開始安裝這些套件。

3.2 初始叢集設定

安裝 HA 套件後，可以繼續進行初始叢集設定。基本步驟包括以下幾項：

- 1 定義通訊通道 [第20頁]
- 2 定義驗證設定 [第23頁]
- 3 將組態傳輸至所有節點 [第24頁]

以下程序借助 YaST 叢集模組引導您完成每個步驟。若要存取叢集組態對話方塊，請以 root 身分啟動 YaST，然後選取「*High Availability*」>「叢集」。或者在指令行上使用 `yast2 cluster` 以 root 身分啟動 YaST 叢集模組。

第一次啟動叢集模組時會顯示精靈，指導您完成基本設定的所有步驟。如果並非第一次啟動，可以按一下左側面板上的類別，以存取每個步驟的組態選項。

3.2.1 定義通訊通道

若要讓叢集節點之間順利進行通訊，至少需定義一個通訊通道。不過，建議設定經由兩個或兩個以上備援路徑的通訊 (可以使用網路設備 **Bonding**，也可以借助 Corosync 新增第二個通訊通道)。對於每個通訊通道，您需要定義以下參數：

結合網路位址 (bindnetaddr)

要繫結到的網路位址。為方便在整個叢集共享組態檔案，OpenAIS 使用網路介面網路遮罩來僅遮罩用於路由網路的地址位元。將此值設定為要用於叢集多路廣播的子網路。

多路廣播位址 (mcastaddr)

可以是 IPv4 或 IPv6 位址。

多路廣播埠 (mcastport)

為 mcastaddr 指定的 UDP 埠。

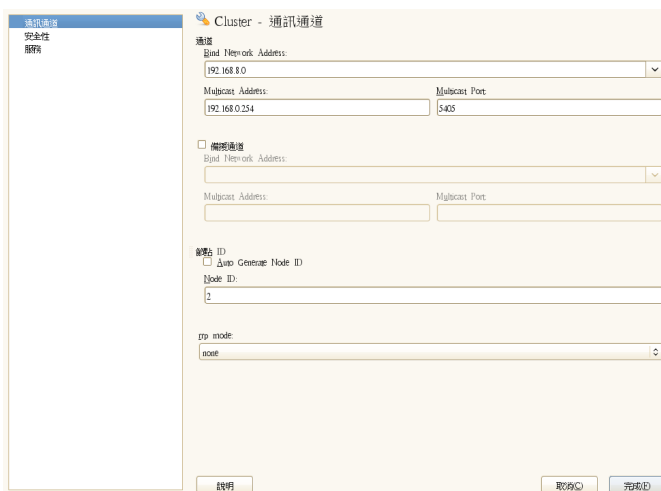
叢集中的所有節點透過使用相同的多路廣播位址及相同的埠號瞭解彼此的存在。對於不同的叢集，請使用不同的多路廣播位址。

若要使用 Corosync 設定備援通訊，需要定義 /etc/corosync/corosync.conf 中的多個介面區段，每個區段都使用不同的環狀網路編號。使用備援環狀網路協定 (RRP) 通知叢集如何使用這些介面。RRP 有三種模式 (rrp_mode)：如果設定為主動，Corosync 會使用所有介面。如果設定為被動，Corosync 只在第一個環狀網路出現故障時才使用第二個介面。如果 rrp_mode 設定為無，則會停用 RRP。借助 RRP，可以使用實體位置相互分隔的兩個網路進行通訊。如果一個網路發生故障，叢集節點仍然可以透過另一個網路進行通訊。

如果設定多個環狀網路，每個節點都可以擁有多個 IP 位址。一旦啟用 rrp_mode，便會依預設使用串流控制傳輸協定 (SCTP) 替代 TCP 進行節點間的通訊。

過程 3.2 定義通訊通道

- 1 在 YaST 叢集模組中，切換至「通訊通道」類別。
- 2 定義「結合網路位址」、「多路廣播位址」及「多路廣播埠」，以供所有叢集節點使用。



3 如果要定義第二個通道：

3a 啟動「備援通道」。

3b 為備援通道定義「結合網路位址」、「多路廣播位址」及「多路廣播埠」。

3c 選取要使用的「*rrp_mode*」。若要停用RRP，請選取「無」。如需有關各模式的詳細資訊，請按一下「說明」。

使用 RRP 後，主要的環狀網路 (設定的第一個通道) 會取得環狀網路編號 0，第二個環狀網路 (備援通道) 會取得編號 1 (位於 `/etc/corosync/corosync.conf` 中)。

4 啟動「自動產生節點 ID」為每個叢集節點自動產生唯一的 ID。

5 如果只想修改現有叢集的通訊通道，請按一下「完成」將組態寫入 `/etc/corosync/corosync.conf`，然後關閉 YaST 叢集模組。YaST 隨後還將自動調整防火牆設定，並開啟用於多路廣播的 UDP 埠。

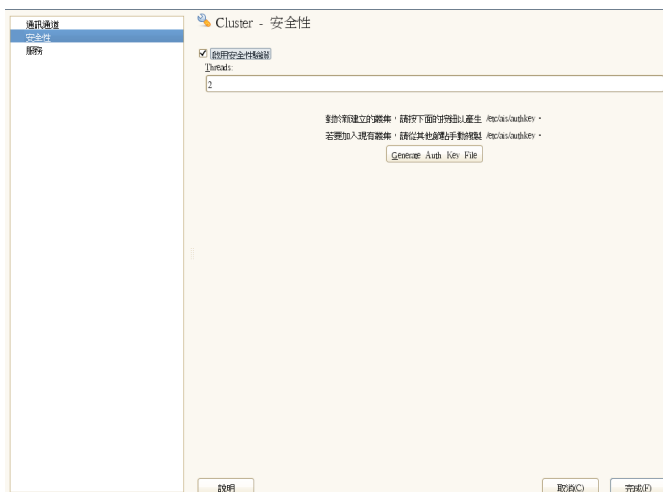
6 如需叢集組態的更多資訊，請前往過程 3.3「啟用安全驗證」[第23頁]。

3.2.2 定義驗證設定

下一步是定義叢集的驗證設定。您可以使用 HMAC/SHA1 驗證，這種驗證方式需要使用共享密碼來保護和驗證訊息。您指定的驗證金鑰 (密碼) 將用於叢集中的所有節點。

過程 3.3 啟用安全驗證

- 1 在 YaST 叢集模組中，切換至「安全性」類別。
- 2 啟動「啟用安全性驗證」。
- 3 對於新建立的叢集，請按一下「產生驗證金鑰檔案」。此時會建立一個驗證金鑰，並寫入 `/etc/corosync/authkey`。



- 4 如果只想修改驗證設定，請按一下「完成」將組態寫入 `/etc/corosync/corosync.conf`，然後關閉 YaST 叢集模組。
- 5 如需叢集組態的更多資訊，請參閱第 3.2.3 節「將組態傳輸至所有節點」[第24頁]。

3.2.3 將組態傳輸至所有節點

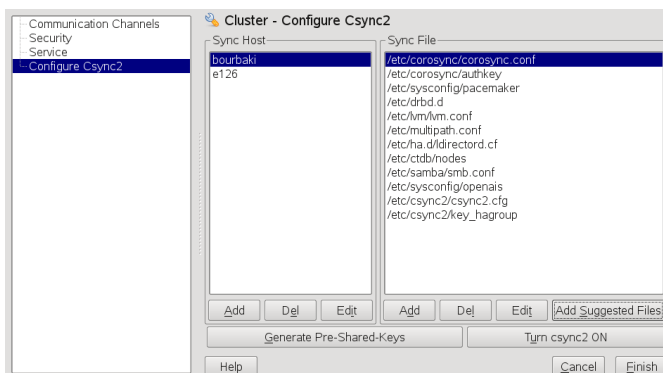
使用 `csync2` 工具完成叢集中所有節點上的複製，而不是將產生的組態檔案手動複製到所有節點。`Csync2` 可以處理任意數量的分類至同步化群組的主機。每個同步化群組有其自己的主機成員清單，以及定義同步化群組中應同步之檔案的包含/排除模式。群組、屬於各個群組的主機名稱以及各個群組的包含/排除規則會在 `Csync2` 組態檔案 `/etc/csync2/csync2.cfg` 中指定。

驗證時，`Csync2` 使用同步化群組內的 IP 位址和預先共用金鑰。您需要為每個同步化群組產生一個金鑰檔案，並將其複製到群組中的所有成員。

如需有關 `Csync2` 的詳細資訊，請參閱 <http://oss.linbit.com/csync2/paper.pdf>。

過程 3.4 使用 *YaST* 設定 *Csync2*

- 1 在 *YaST* 叢集模組中，切換至「*Csync2*」類別。
- 2 若要指定同步化群組，請按一下「*同步化主機*」群組中的「*新增*」，然後輸入叢集中所有節點的本地主機名稱。對於每個節點，都必須使用 `hostname` 指令傳回的字串。
- 3 按一下「*產生預先共用金鑰*」建立同步化群組的金鑰檔案。建立的金鑰檔案會寫入 `/etc/csync2/key_hagroup`。建立之後，必須手動將其複製到叢集的所有成員。
- 4 若要在「*同步化檔案*」清單中填入所有節點間執行同步化通常所需的檔案，請按一下「*新增建議的檔案*」。



- 5 若要在待同步檔案的清單中「編輯」、「新增」或「移除」檔案，則使用相應的按鈕。必須輸入各個檔案的絕對路徑名稱。
- 6 按一下「開啟 Csync2」以啟動 Csync2。這會讓 Csync2 在開機時自動啟動。
- 7 根據需要設定所有選項後，請按一下「完成」關閉 YaST 叢集模組。隨後，YaST 會將 Csync2 組態寫入 `/etc/csync2/csync2.cfg`。

設定 Csync2 之後，從指令行啟動同步程序，如下所述。

過程 3.5 使用 Csync2 同步組態檔案

若要使用 Csync2 成功同步檔案，請確定符合以下必要條件：

- 所有節點都能使用相同的 Csync2 組態。將 `/etc/csync2/csync2.cfg` 納入要使用 Csync2 進行同步的檔案清單中，或在設定檔案(如過程 3.4「使用 YaST 設定 Csync2」[第24頁]中所述)後手動將其複製到所有節點。
- 將步驟 3 [第24頁] 中於某個節點上產生的 `/etc/csync2/key_hagroup` 檔案複製到叢集中的所有節點，以供 Csync2 驗證之需。但是，不要在其他節點上重新產生此檔案，因為所有節點必須使用同一個檔案。
- 確定所有節點上都在執行 `xinetd`，因為 Csync2 要仰賴該精靈。使用以下列指令可以 `root` 身分啟動 `xinetd`：

```
rcxinetd start
```

注意：在開機時啟動服務

如果您希望 **Csync2** 和 **xinetd** 在開機時自動啟動，可以在所有節點上執行以下指令：

```
chkconfig csync2 on
chkconfig xinetd on
```

1 在某個節點上執行以下指令以啟動檔案同步化：

```
csync2 -xv
```

此操作將一次性同步所有檔案。如果能成功同步所有檔案，**Csync2** 便會完成並且不出現任何錯誤。

如果要同步的一或多個檔案在其他節點上(不僅僅是目前節點上)也進行了修改，**Csync2** 將報告有衝突。輸出內容與以下類似：

```
While syncing file /etc/corosync/corosync.conf:
ERROR from peer hex-14: File is also marked dirty here!
Finished with 1 errors.
```

2 如果您確定目前節點上的是檔案的「最佳」版本，可以透過強制使用此檔案並重新同步來解決衝突：

```
csync2 -f /etc/corosync/corosync.conf
csync2 -x
```

如需有關 **Csync2** 選項的詳細資訊，請執行 `csync2 -help`。

注意：觸發同步化

Csync2 不會持續在節點間同步檔案。每次對需要同步的檔案進行更新之後，都需要手動重新同步檔案。

同步叢集中所有節點的金鑰檔案後，啟動基本服務以連線叢集，如第3.3節「連線叢集」[第27頁]中所述。

3.2.4 啟動服務

YaST 叢集模組可讓您定義是否要在節點開機時啟動某些服務。如果您不想使用指令行啟動和停止服務，也可以使用此模組手動執行。為了連線叢集節點並啟動叢集資源管理員，必須啟動 OpenAIS 這項服務。

過程 3.6 啟動或停止服務

- 1 在 YaST 叢集模組中，切換至「服務」類別。
- 2 若要在每次開機此叢集節點時啟動 OpenAIS，則在「開機」群組中選取對應的選項。
- 3 如果要使用 Pacemaker GUI 設定、管理和監控叢集資源，則啟動「同時啟動 mgmt」。
- 4 若要立即啟動或停止 OpenAIS，則按一下相應的按鈕。
- 5 按一下「完成」關閉 YaST 叢集模組。

如果在「開機」群組中選取「關閉」，則每次節點開機時，必須手動啟動 OpenAIS。若要手動啟動 OpenAIS，請使用 `rcopenais start` 指令。

3.3 連線叢集

完成初始叢集組態設定後，可以立即啟動連線堆疊所需的服務。

過程 3.7 啟動 OpenAIS/Corosync 並檢查狀態

- 1 在每個叢集節點上執行以下指令，以啟動 OpenAIS/Corosync：

```
rcopenais start
```

- 2 在其中一個節點上，使用以下指令檢查叢集的狀態：

```
crm_mon
```

如果所有節點都已上線，則輸出應如下所示：

```
=====  
Last updated: Tue Mar  2 18:35:34 2010
```

```
Stack: openais
Current DC: e229 - partition with quorum
Version: 1.1.1-530add2a3721a0ecccb24660a97dbfdaa3e68f51
2 Nodes configured, 2 expected votes
0 Resources configured.
=====

Online: [ e231 e229 ]
```

此輸出表示叢集資源管理員已啟動，並可以管理資源。

完成基本組態設定並連線節點之後，便可開始設定叢集資源。此時，可使用 `crm` 指令行工具或圖形使用者介面。如需更多資訊，請參閱第 5 章「設定和管理叢集資源 (GUI)」[第53頁] 或第 6 章「設定和管理叢集資源 (指令行)」[第83頁]。

3.4 使用 AutoYaST 進行大規模部署

AutoYaST 這套系統可在不需要使用者介入的情況下自動安裝一或多個 SUSE Linux Enterprise 系統。SUSE Linux Enterprise 可讓您建立包含安裝和組態資料的 AutoYaST 設定檔。此設定檔會告訴 AutoYaST 要安裝什麼，及如何設定安裝的系統，以便最後獲得的系統萬事具備。隨後，可以借助多種方法使用此設定檔進行大規模部署。

如需有關在不同情境下使用 AutoYaST 的詳細指示，請參閱 <http://www.novell.com/documentation> 中的《SUSE Linux Enterprise 11 SP1 部署指南》。詳情位於「自動安裝」一章。

過程 3.8 使用 AutoYaST 複製叢集節點

部署現有節點之副本的叢集節點時可以採用以下程序。複製的節點將安裝相同的套件並具有相同的系統組態。

如果要在不相同的硬體上部署叢集節點，請參閱 <http://www.novell.com/documentation> 上《SUSE Linux Enterprise 11 SP1 部署指南》中的「以規則為基礎的自動安裝」一節。

重要：完全一樣的硬體

本案例假設您要將 SUSE Linux Enterprise High Availability Extension 11 SP1 部署到硬體組態完全一樣的一組機器上。

- 1 請確定已按照第 3.1 節「安裝 High Availability Extension」[第19頁] 和第 3.2 節「初始叢集設定」[第20頁] 中的說明正確安裝並設定要複製的節點。
- 2 按照《SUSE Linux Enterprise 11 SP1 部署指南》中的概要描述，進行簡單的大規模部署。基本步驟包括以下幾項：
 - 2a 建立 AutoYaST 設定檔。使用 AutoYaST GUI 可以在現有系統組態的基礎上建立和修改設定檔。在 AutoYaST 中，選擇「*High Availability*」模組，然後按一下「複製」按鈕。如有需要，調整其他模組中的組態，並將產生的控制檔案儲存為 XML 檔案。
 - 2b 指定 AutoYaST 設定檔及參數的來源，以便傳遞給其他節點的安裝常式。
 - 2c 指定 SUSE Linux Enterprise Server 及 SUSE Linux Enterprise High Availability Extension 安裝資料的來源。
 - 2d 指定及設定自動安裝的開機程序。
 - 2e 透過手動新增參數或建立 info 檔案的方式，將指令行傳遞給安裝常式。
 - 2f 啟動和監控自動安裝程序。

成功安裝副本後，請執行以下步驟將複製的節點加入叢集：

過程 3.9 連線複製的節點

- 1 使用 Csync2 將金鑰組態檔案從設定的節點傳輸到複製的節點，如第 3.2.3 節「將組態傳輸至所有節點」[第24頁] 中所述。
- 2 如第 3.3 節「連線叢集」[第27頁] 中所述啟動複製節點上的 OpenAIS 服務，以連線節點。

複製的節點將立即加入叢集，因為 `/etc/corosync/corosync.config` 檔案已透過 Csync2 套用至複製的節點。CIB 會自動在叢集節點間同步。

II. 組態與管理

組態與管理基礎

HA 叢集的主要目的是管理使用者服務。Apache Web 伺服器或資料庫便是使用者服務的典型範例。從使用者的角度來看，命令這些服務執某些特定作業，它們就會按要求執行。不過，對於叢集而言，它們只是可以啟動或停止的資源 — 服務的性質與叢集無關。

本章介紹一些設定資源和管理叢集時需瞭解的基本概念。後續章節將介紹如何使用 High Availability Extension 提供的各種管理工具執行主要的組態與管理任務。

4.1 全域叢集選項

全域叢集選項控制叢集在遇到特定情況時的運作方式。這些選項已進行分組，可以使用類似 Pacemaker GUI 與 `crm` 外圍程序的叢集管理工具進行檢視與修改。多數情況下，可以使用預先定義的值。但是，為了讓叢集的關鍵功能正常運作，還需要在執行基本叢集設定後調整以下參數：

- `no-quorum-policy` 選項 [第34頁]
- `stonith-enabled` 選項 [第35頁]

過程 5.1「修改全域叢集選項」[第57頁]中介紹了如何使用 GUI 調整這些參數。如果想要使用指令行方式，請參閱第 6.2 節「設定全域叢集選項」[第89頁]。

4.1.1 no-quorum-policy 選項

此全域選項定義在叢集沒有達到最低節點數 (節點多數不是分割區的一部分) 時如何運作。

允許的值：

`ignore`

最低節點數的狀態完全不會影響叢集的運作，資源管理將繼續。

此設定適合以下情境：

- 雙節點叢集：由於單個節點失敗後便無法滿足最低節點數規則，因此您通常會希望叢集忽略此限制，繼續執行。使用圍籬區隔可以保證資源的完整性，還可以防止出現電腦分裂的情況。
- 資源導向叢集：對於使用備援通訊通道的本地叢集，只有在特定情況下才會發生電腦分裂。因此，與一個節點通訊中斷很可能表示該節點當機，還在運作的節點應復原並重新為資源提供服務。

如果 `no-quorum-policy` 設定為 `ignore`，則對於一個 4 節點的叢集而言，即使有三個節點同時失敗，服務仍能保持運作；相反，如果使用其他設定，則有兩個節點同時失敗時，就會不滿足最低節點數規則。

`freeze`

如果不滿足最低節點數規則，叢集便停止。資源管理將繼續：執行中的資源不會停止 (但是可能會重新啟動，以便監控事件)，但是受影響的分割區內不會再啟動其他資源。

此設定適合某些資源依賴與其他節點之通訊的叢集 (例如，OCFS2 掛接)。在這種情況下，預設設定 `no-quorum-policy=stop` 不起作用，因為它會造成既無法停止資源，又無法連接對等節點的狀況。任何停止這些資源的嘗試最終都會逾時，並出現 `stop failure`，同時觸發升級式恢復與圍籬區隔。

`stop` (預設值)

如果不滿足最低節點數規則，受影響的叢集分割區中的所有資源都會依序停止。

suicide

圍籬區隔受影響的叢集分割區中的所有節點。

4.1.2 stonith-enabled 選項

此全域選項定義是否套用圍籬區隔，以便允許 STONITH 設備關閉失敗的節點以及有資源無法停止的節點。此全域選項預設為 `true`，因為正常的叢集運作需要使用 STONITH 設備。依據預設值，如果未定義 STONITH 資源，叢集將拒絕啟動任何資源。

如果出於某種原因需要停用圍籬區隔，請將 `stonith-enabled` 設定為 `false`。

若想瞭解所有全域叢集選項及其預設值，請參閱 <http://clusterlabs.org/wiki/Documentation> 上的《*Pacemaker 1.0—Configuration Explained*》(Pacemaker 1.0 — 組態說明)。詳情位於「*Available Cluster Options*」(可用的叢集選項)一節。

4.2 叢集資源

做為叢集管理員，您需要為您叢集中的伺服器上執行的所有資源或應用程式建立叢集資源。叢集資源可包括網站、電子郵件伺服器、資料庫、檔案系統、虛擬機器，以及其他您希望使用者隨時都可以存取的伺服器型應用程式或服務。

4.2.1 資源管理

若想在叢集中使用某項資源，必須先對其進行設定。例如，如果想要使用 Apache 伺服器做為叢集資源，請先設定 Apache 伺服器並完成 Apache 組態設定，然後在叢集中啟動各個資源。

如果資源具有特定的環境要求，請確保這些要求在所有叢集節點上均得到滿足並且一致。此類組態並非由 High Availability Extension 管理，您必須自行管理。

注意：不要對叢集管理的服務執行任何操作

使用 High Availability Extension 管理資源時，不能再啟動或停止相同的資源（例如在叢集之外手動開機或重新開機）。High Availability Extension 軟體負責所有服務的啟動或停止動作。

不過，如果您要檢查服務是否正確設定，請手動將其啟動，但請確保在 High Availability 接管之前將它再次停止。

在叢集中設定資源之後，可以使用叢集管理工具手動啟動、停止、清理、移除或移轉資源。如需執行此類操作的詳細資料，請參閱第 5 章「設定和管理叢集資源 (GUI)」[第53頁] 或第 6 章「設定和管理叢集資源 (指令行)」[第83頁]。

4.2.2 受支援的資源代辦類別

對於每個新增的叢集資源，都需要定義資源代辦所遵循的標準。資源代辦會提取它們所提供的服務並向叢集提供準確的狀態，這樣叢集便可不理會其所管理的資源。當接收到啟動、停止或監控指令時，叢集會依賴資源代辦來做出恰當的反應。

通常，資源代辦採用的是外圍程序檔的形式。High Availability Extension 支援以下類別的資源代辦：

舊版 Heartbeat 1 資源代辦

Heartbeat 版本 1 具有自己的資源代辦樣式。由於很多人已基於其慣例撰寫了自己的代辦，所以這些資源代辦仍受支援。但是，仍建議您在可能的情況下將您的組態移轉至 High Availability OCF RA。

Linux Standards Base (LSB) 程序檔

LSB 資源代辦通常由作業系統/套裝作業系統提供，位於 `/etc/init.d`。若要與叢集一起使用，它們必須符合 LSB init 程序檔規格。例如，它們必須執行幾個動作，至少包含 `start`、`stop`、`restart`、`reload`、`force-reload` 和 `status`。如需詳細資訊，請參閱 <http://ldn.linuxfoundation.org/lsb/lsb4-resource-page%23Specification>。

這些服務的組態尚未標準化。如果要將 LSB 程序檔與 High Availability 搭配使用，請確定您瞭解如何設定相關程序檔。相關資訊通常可以在 `/usr/share/doc/packages/套件名稱目錄` 中相關套件的文件內找到。

開放叢集架構 (OCF) 資源代辦

OCF RA 代辦最適合與 High Availability 搭配使用，特別是在您需要主要資源或特殊監控功能的情況下。代辦通常位於 `/usr/lib/ocf/resource.d/提供者/` 內。它們的功能類似於 LSB 程序檔。但是，組態始終使用環境變數進行設定，因此它們更容易接受並處理參數。OCF 規格 (與資源代辦相關時) 位於 <http://www.opencf.org/cgi-bin/viewcvs.cgi/specs/ra/resource-agent-api.txt?rev=HEAD&content-type=text/vnd.viewcvs-markup>。OCF 規格對於動作必須傳回的離開碼有著嚴格的定義，請參閱第 8.3 節「OCF 傳回代碼與失敗復原」[第111頁]。叢集完全遵循這些規格。如需所有可用 OCF RA 的詳細清單，請參閱第 19 章「*HA OCF Agents*」[第243頁]。

所有 OCF 資源代辦都必須至少含有動作 `start`、`stop`、`status`、`monitor` 和 `meta-data`。`meta-data` 動作可取回有關如何設定代辦的資訊。例如，如果您要詳細瞭解提供者 `heartbeat` 的 `IPaddr` 代辦，可以使用以下指令：

```
OCF_ROOT=/usr/lib/ocf /usr/lib/ocf/resource.d/heartbeat/IPaddr meta-data
```

輸出的是 XML 格式的資訊，分為多個區段，包括代辦的一般描述、可用參數和可用動作。

STONITH 資源代辦

此類別專用於圍籬區隔相關資源。如需詳細資訊，請參閱第 9 章「*圍籬區隔與 STONITH*」[第113頁]。

提供給 High Availability Extension 的代辦將寫入 OCF 規格。

4.2.3 資源類型

以下是可建立的資源類型：

原始資源

原始資源，是最基本的資源類型。

過程 5.2「新增原始資源」[第57頁] 介紹了如何使用 GUI 建立原始資源。如果想要使用指令行方式，請參閱第 6.3.1 節「建立叢集資源」[第90頁]。

群組

群組包含一組需放置在一起的資源，這些資源按順序啟動並以相反順序停止。若需更多資訊，請參考章節「群組」[第38頁]。

複製資源

複製資源是可在多個主機上處於使用中狀態的資源。任何資源都可複製，只要相應的資源代辦支援複製功能。若需更多資訊，請參考 章節「複製品」[第40頁]。

主要資源

主要資源是複製資源的一種特殊類型，它可以擁有多個模式。若需更多資訊，請參考章節「主要資源」[第40頁]。

4.2.4 進階資源類型

原始資源是最簡單的資源，因此設定簡單，不過您的叢集組態可能還需要更進階的資源類型 (例如群組、複製資源或主要資源)。

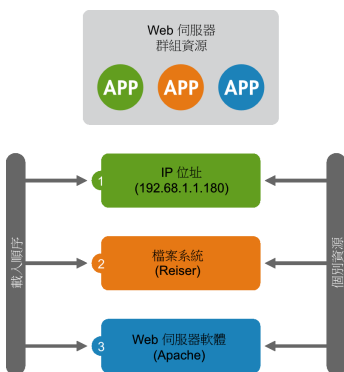
群組

某些叢集資源依賴於其他元件或資源，要求每個元件或資源以特定順序啟動，並在同一個伺服器上執行。若要簡化此組態，您可以使用群組。

範例 4.1 Web 伺服器的資源群組

需要 IP 位址與檔案系統的 Web 伺服器就是資源群組的一個範例。在此範例中，每個元件都是組合到叢集資源群組中的獨立叢集資源。資源群組會在一或多個伺服器上執行，若軟體或硬體出現異常，資源群組會容錯移轉至叢集中的其他伺服器，這一點與個別叢集資源相同。

圖形 4.1 群組資源



群組具有以下內容：

啟動和停止

資源以顯示的順序依次啟動，並以相反順序停止。

相依性

若群組中的某個資源在任何地方都無法執行，則群組中位於該資源之後的所有資源均不允許執行。

內容

群組僅會包含原始叢集資源的集合。群組至少須包含一個資源，否則組態視為無效。若要引用群組資源的子代，請使用子代 ID，而不要使用群組 ID。

限制

儘管您可以在限制中參考群組的子代，但一般最好使用群組的名稱。

相粘性

相粘性在群組中屬於加法類內容。群組中每個使用中成員的粘性值都會影響群組的總值。因此，若資源綁定的預設值為 100，並且群組有七個成員，其中五個處於使用中狀態，則整個群組偏向於其目前位置 (分數為 500)。

資源監控

若要對群組啟用資源監控，您必須為要監控的群組中的每個資源分別設定監控。

過程 5.12「新增資源群組」[第72頁]介紹了如何使用 GUI 建立群組。如果想要使用指令行方式，請參閱第 6.3.9 節「設定叢集資源群組」[第100頁]。

複製品

您可能要讓某些資源同時在叢集的多個節點上執行。若要實現此目的，您必須將資源設定為複製品。可以設定為複製品的資源範例包括 STONITH 以及叢集檔案系統，如 OCFS2。您可以複製所提供的任何資源。相關資源的資源代辦會為此操作提供支援。您甚至可以對複製品資源進行不同的設定，具體視代管它們的節點而定。

資源複製品分為三種類型：

匿名複製品

這是最簡單的一種複製品。無論在何處執行，它們的行為都相同。因此，每部機器上只能有一個匿名複製品例項處於使用中狀態。

全域唯一複製品

這些資源是不同的實體。在一個節點上執行的複製品例項與在另一個節點上執行的另一個例項不同；在相同節點上的任何兩個例項也不相同。

狀態複製品

這些資源的使用中例項分為兩種狀態：主動與被動。這兩種狀態有時也稱為主要與次要，或主要與從屬。狀態複製品可以是匿名複製品，也可以是全域唯一複製品。並請參閱章節「主要資源」[第40頁]。

複製資源只能包含一個群組或一個一般資源。

設定資源監控或限制時，主要資源的要求與簡單資源不同。如需詳細資訊，請參閱 <http://clusterlabs.org/wiki/Documentation> 上的《*Pacemaker 1.0—Configuration Explained*》(Pacemaker 1.0 — 組態說明)。詳情位於「*Clones - Resources That Should be Active on Multiple Hosts*」(複製資源 - 應在多個主機上啟動的資源)一節。

過程 5.14「新增或修改複製資源」[第76頁] 介紹了如何使用 GUI 建立複製資源。如果想要使用指令行方式，請參閱第 6.3.10 節「設定複製資源」[第100頁]。

主要資源

主要資源是一種特別的複製資源，允許例項處於主要或從屬操作模式。主要資源只能包含一個群組或一個一般資源。

設定資源監控或限制時，主要資源的要求與簡單資源不同。如需詳細資訊，請參閱 <http://clusterlabs.org/wiki/Documentation> 上的《*Pacemaker 1.0—Configuration Explained*》(Pacemaker 1.0 — 組態說明)。詳情位於「*Multi-state - Resources That Have Multiple Modes*」(多狀態 - 有多種模式的資源)一節。

4.2.5 資源選項 (中繼屬性)

對於新增的每個資源，您都可以定義選項。叢集使用選項來決定資源的行為方式 — 它們會告知 CRM 如何處理特定資源。資源選項可透過 `crm_resource --meta` 指令或 GUI (如過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁] 中所述) 來設定。

表格 4.1 原始資源選項

選項	描述
<code>priority</code>	如果並非所有資源都可啟動，則叢集將停止優先程度較低的資源，以便讓優先程度較高的資源啟動。
<code>target-role</code>	叢集應嘗試讓此資源保持什麼狀態？允許的值： <code>stopped</code> 、 <code>started</code> 。
<code>is-managed</code>	叢集是否允許啟動及停止資源？允許的值： <code>true</code> 、 <code>false</code> 。
<code>resource-stickiness</code>	資源希望留在原處的程度如何？預設值為 <code>default- resource-stickiness</code> 。
<code>migration-threshold</code>	此資源在節點上的失敗次數達到多少次才會取消該節點代管此資源的資格？預設值： <code>none</code> 。
<code>multiple-active</code>	如果叢集發現資源在多個節點上處於使用中狀態，應如何處理？允許的值： <code>block</code> (將資源標示為不受管理)、 <code>stop_only</code> 、 <code>stop_start</code> 。

選項	描述
<code>failure-timeout</code>	等待多少秒之後才會當做失敗未發生 (潛在含義為允許資源回到其失敗的節點上)? 預設值: <code>never</code> 。
<code>allow-migrate</code>	允許對支援 <code>migrate_to/migrate_from</code> 動作的資源執行資源移轉。

4.2.6 例項屬性

所有資源類別的程序檔均可接收參數，參數決定了資源的行為方式及其控制的服務例項。如果您的資源代辦支援參數，可以使用 `crm_resource` 指令或 GUI (如過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁]中所述) 新增參數。在 `crm` 指令行公用程式中，例項屬性稱為 `params`。您可以 `root` 身分執行以下指令來獲取受 OCF 程序檔支援的例項屬性清單：

```
crm ra info [class:[provider:]]resource_agent
```

或者更短的指令：

```
crm ra info resource_agent
```

輸出會列出所有受支援的屬性、其用途與預設值。

例如，指令

```
crm ra info Ipaddr
```

傳回下列輸出：

```
Manages virtual IPv4 addresses (portable version) (ocf:heartbeat:IPaddr)
```

```
This script manages IP alias IP addresses
It can add an IP alias, or remove one.
```

```
Parameters (* denotes required, [] the default):
```

```
ip* (string): IPv4 address
The IPv4 address to be configured in dotted quad notation, for example
"192.168.1.1".
```

```
nic (string, [eth0]): Network interface
The base network interface on which the IP address will be brought
```

online.

If left empty, the script will try and determine this from the routing table.

Do NOT specify an alias interface in the form eth0:1 or anything here; rather, specify the base interface only.

`cidr_netmask (string): Netmask`

The netmask for the interface in CIDR format. (ie, 24), or in dotted quad notation 255.255.255.0).

If unspecified, the script will also try to determine this from the routing table.

`broadcast (string): Broadcast address`

Broadcast address associated with the IP. If left empty, the script will determine this from the netmask.

`iflabel (string): Interface label`

You can specify an additional label for your IP address here.

`lvs_support (boolean, [false]): Enable support for LVS DR`

Enable support for LVS Direct Routing configurations. In case a IP address is stopped, only move it to the loopback device to allow the local node to continue to service requests, but no longer advertise it on the network.

`local_stop_script (string):`

Script called when the IP is released

`local_start_script (string):`

Script called when the IP is added

`ARP_INTERVAL_MS (integer, [500]): milliseconds between gratuitous ARPs`
milliseconds between ARPs

`ARP_REPEAT (integer, [10]): repeat count`

How many gratuitous ARPs to send out when bringing up a new address

`ARP_BACKGROUND (boolean, [yes]): run in background`

run in background (no longer any reason to do this)

`ARP_NETMASK (string, [ffffffffffff]): netmask for ARP`

netmask for ARP - in nonstandard hexadecimal format.

`Operations' defaults (advisory minimum):`

`start` `timeout=90`

`stop` `timeout=100`

`monitor_0` `interval=5s timeout=20s`

注意：群組、複製資源或主要資源的例項屬性

請注意，群組、複製資源與主要資源沒有例項屬性。但是，群組、複製資源或主要資源的子代會繼承所設定的所有例項屬性。

4.2.7 資源作業

叢集預設不會確保資源仍正常。若要指示叢集確保資源能正常運作，您需要在資源定義中新增監控作業。可為所有類別或資源代辦新增監控作業。若需更多資訊，請參考第 4.3 節「資源監控」[第45頁]。

表格 4.2 資源作業

操作	描述
id	動作執行人名稱。必須唯一。(ID 不顯示)。
name	執行的動作。通用值：monitor、start、stop。
interval	執行作業的頻率。單位：秒
timeout	宣告動作失敗之前等待多長時間？
requires	需滿足什麼條件才會執行此動作。允許的值： nothing、quorum、fencing。預設值取決於圍籬區隔是否啟用及資源類別是否為stonith。對於STONITH資源，預設值為nothing。
on-fail	此動作失敗時所採取的動作。允許的值： <ul style="list-style-type: none">• ignore：當做資源未失敗。• block：在資源上不執行任何進一步作業。• stop：停止資源且不在其他任何地方啟動。• restart：停止資源並再次啟動 (可能在其他節點上)。

操作	描述
	<ul style="list-style-type: none"> • <code>fence</code>: 將發生資源失敗的節點關閉 (STONITH)。 • <code>standby</code>: 將發生資源失敗之節點上的所有資源移出。
<code>enabled</code>	如果為 <code>false</code> , 則會將作業視做不存在。允許的值: <code>true</code> 、 <code>false</code> 。
<code>role</code>	只有在資源具有此角色時, 才執行操作。
<code>record-pending</code>	可以全域設定, 也可以針對個別資源設定。讓 CIB 反映對資源執行之「in-flight」操作的狀態。
<code>description</code>	操作的描述。

4.3 資源監控

若要確定資源是否正在執行, 必須設定針對該資源的資源監控。

若資源監控偵測到失敗, 系統將執行以下動作:

- 根據 `/etc/corosync/corosync.conf` 的 `logging` 區段中指定的組態產生記錄檔案訊息。記錄預設會寫入 `syslog`, 通常為 `/var/log/messages`。
- 在叢集管理工具 (Pacemaker GUI、HA Web Konsole、`crm_mon`) 及 CIB 狀態區段中反應出失敗。
- 叢集會啟動重要的復原動作, 其中可能包括停止資源以修復失敗狀態, 以及在本地或在其他節點上重新啟動資源。也可能根本不重新啟動資源, 具體視組態及叢集狀態而定。

若不設定資源監控, 則不會向您通知資源成功啟動後發生的失敗, 並且叢集會始終將資源顯示為處於正常狀態。

過程 5.11「新增或修改監控作業」[第70頁] 中介紹了如何使用 GUI 新增對資源的監控操作。如果想要使用指令行方式，請參閱第 6.3.8 節「設定資源監控」[第99頁]。

4.4 資源限制

設定所有資源只是工作的一部分。即使叢集瞭解所有必需的資源，可能仍然無法正確地對其進行處理。資源限制可讓您指定資源可在哪些叢集節點上執行，載入資源的順序以及特定資源所依賴的其他資源。

4.4.1 限制類型

系統中的限制分為三種類型：

資源位置

位置限制，定義資源可在、不可在或偏好在哪些節點上執行。

資源並存

並存限制，告訴叢集哪些資源可以或不可以同時在節點上執行。

資源順序

順序限制，定義動作執行的順序。

如需設定限制的詳細資訊，以及有關順序與並存基本概念的詳細背景資訊，請參閱 <http://clusterlabs.org/wiki/Documentation> 中提供的以下文件：

- 《*Pacemaker 1.0—Configuration Explained*》(Pacemaker 1.0 — 組態說明) 的「*Resource Constraints*」(資源限制) 一章
- 《*Collocation Explained*》(並存說明)
- 《*Ordering Explained*》(順序說明)

第 5.3.3 節「設定資源限制」[第62頁] 中介紹了如何使用 GUI 新增各種限制。如果想要使用指令行方式，請參閱第 6.3.4 節「設定資源限制」[第94頁]。

4.4.2 分數與無限大

定義限制時，還需要處理分數。所有類型的分數對於叢集的工作方式而言都是不可或缺的。實際上，從移轉資源到決定要將降級叢集中哪個資源停止的所有作業，都是透過某種方式對分數進行操作來實現。系統會對每個資源都計算分數，針對某個資源分數為負數的所有節點都不能執行該資源。計算完針對資源的分數之後，叢集會選擇分數最高的節點。

INFINITY 目前定義為 1,000,000。與其相關的加法或減法計算遵循以下三項基本規則：

- 任何值 + INFINITY = INFINITY
- 任何值 - INFINITY = -INFINITY
- INFINITY - INFINITY = -INFINITY

定義資源限制時，您還要指定每個限制的分數。分數表示您要指定給此資源限制的值。系統會先套用分數較高的限制，然後再套用分數較低的限制。透過為指定資源建立不同分數的其他位置限制，即可指定資源將容錯移轉至之目標節點的順序。

4.4.3 容錯移轉節點

若資源失敗，系統會自動將其重新啟動。若無法在目前節點上重新啟動，或資源已在目前節點上失敗 N 次，則會嘗試容錯移轉至其他節點。每次資源失敗，其 `failcount` 值都會增加。您可以定義一個數值，讓資源在失敗該次數 (`migration-threshold`) 之後移轉至新節點。若叢集中有兩個以上的節點，則特定資源容錯移轉所至的節點由 High Availability 軟體來選擇。

但是，您可以透過為該資源設定一或多個位置限制以及一個 `migration-threshold`，指定資源容錯移轉所至的節點。如需使用 GUI 執行此操作的詳細指示，請參閱第 5.3.4 節「指定資源容錯移轉節點」[第 65 頁]。如果想要使用指令行方式，請參閱第 6.3.5 節「指定資源容錯移轉節點」[第 96 頁]。

範例 4.2 移轉限定值 — 程序流程

例如，假設您已為資源 `r1` 設定位置限制，讓其偏向於在 `node1` 上執行。若資源在該節點上失敗，系統會檢查 `migration-threshold`，並將其與 `failcount` 進行比較。若 `failcount >= migration-threshold`，則將資源移轉至優先設定次佳的節點。

依預設，一旦達到限定值，便不再允許該節點執行失敗的資源，除非重設資源的 `failcount`。此操作可由叢集管理員手動執行，也可透過設定資源的 `failure-timeout` 選項來完成。

例如，設定 `migration-threshold=2` 及 `failure-timeout=60s` 會在失敗兩次後將資源移轉至新節點，並且允許其在一分鐘後移回原節點(具體視綁定與限制分數而定)。

移轉限定值概念有兩種例外情況，發生於資源無法啟動或無法停止之時：

- 啟動失敗會將 `failcount` 設定為 `INFINITY`，因此一旦發生便會立即移轉資源。
- 停止失敗會導致圍籬區隔(當 `stonith-enabled` 設定為預設值 `true` 時)。

若未定義任何 `STONITH` 資源(或將 `stonith-enabled` 設定為 `false`)，則資源一律不會移轉。

如需使用移轉限定值與重設 `failcount` 的詳細資訊，請參閱第 5.3.4 節「指定資源容錯移轉節點」[第65頁]。如果想要使用指令行方式，請參閱第 6.3.5 節「指定資源容錯移轉節點」[第96頁]。

4.4.4 錯誤回復節點

當原始節點恢復連線且位於叢集中時，資源可以錯誤回復至該節點。若不想讓資源錯誤回復至容錯移轉之前所處的節點，或要為資源指定另一個錯誤回復節點，必須變更其資源綁定的值。您可以在建立資源時或建立之後指定資源綁定。

指定資源綁定的值時，請考慮以下事項：

值為 0：

此為預設值。資源處於系統中的最佳位置。這表示當有「更佳」的或負載更低的節點可用時，則移動資源。此選項幾乎等同於自動錯誤回復，除了資源可能會移至原先節點 (資源之前於其上處於使用中狀態) 之外的節點這種情況。

值大於 0：

資源偏向於保留在其目前的位置，但當有更合適的節點時，則可能會移動。值越高表示越偏向於將資源保留在目前的位置。

值小於 0：

資源偏向於從其目前位置移開。絕對值越高表示越偏向於移動資源。

值為 INFINITY：

資源始終保留在其目前的位置，除非由於此節點不再符合執行該資源的條件 (節點關機、節點待機、達到 migration-threshold，或組態變更) 而強制關閉此選項。此選項幾乎等同於完全停用自動錯誤回復。

值為 -INFINITY：

資源始終從其目前位置移開。

4.4.5 依據負載影響放置資源

並非所有資源都相同。有些資源 (例如 Xen 訪客) 要求代管它們的節點滿足其容量要求。如果放置資源後其所需的容量之和超過了提供的容量，資源效能便會下降 (甚至無法執行)。

鑒於此，High Availability Extension 允許您指定以下參數：

1. 特定節點提供的容量。
2. 特定資源要求的容量。
3. 配置資源的整體策略。

目前這些設定都是靜態的，必須由管理員設定，無法動態探查或調整。

第 5.3.6 節「根據負載影響設定資源的配置」[第67頁] 中介紹了如何使用 GUI 設定這些設定。如果想要使用指令行方式，請參閱第 6.3.7 節「根據負載影響設定資源的配置」[第97頁]。

如果一個節點有足夠的可用容量來滿足資源的要求，則認為該節點符合資源要求。所要求或所提供容量的用意與 High Availability Extension 完全無關，只是為了確定節點滿足資源的全部容量要求之後，再移動資源。

若要設定資源的要求以及節點提供的容量，可以利用使用率屬性。您可以依據自己的偏好命名使用率屬性，依據組態需要定義任意數量的名稱/值對。但是，屬性的值必須是整數。

配置策略可以使用 placement-strategy 內容 (在全域叢集選項中) 指定。可用的值如下：

default (預設值)

完全不考量使用率值。資源依據位置分數配置。如果分數相同，則在各節點上平均分配資源。

utilization

判斷節點是否有足夠的可用容量來滿足資源的要求時，會考量使用率值。但是，依舊會依據配置給節點的資源數量完成負載平衡。

minimal

判斷節點是否有足夠的可用容量來滿足資源的要求時，會考量使用率值。系統會儘可能將資源集中到少量節點上，以節省其餘節點上的能耗。

balanced

判斷節點是否有足夠的可用容量來滿足資源的要求時，會考量使用率值。系統會嘗試將資源平均分配，以便最佳化資源效能。

注意：設定資源優先程度

可用的配置策略已經過最佳化，但未使用複雜的啟發式解析程序來確保始終達到最佳的配置結果。因此，您可以設定資源的優先程度，確保先排程最重要的資源。

範例 4.3 負載平衡放置的範例組態

以下範例中說明了一個含四台虛擬機器的三節點叢集 (每個節點相同)。

```
node node1 utilization memory="4000"
node node2 utilization memory="4000"
node node3 utilization memory="4000"
primitive xenA ocf:heartbeat:Xen utilization memory="3500" \
    meta priority="10"
primitive xenB ocf:heartbeat:Xen utilization memory="2000" \
    meta priority="1"
primitive xenC ocf:heartbeat:Xen utilization memory="2000" \
    meta priority="1"
primitive xenD ocf:heartbeat:Xen utilization memory="1000" \
    meta priority="5"
property placement-strategy="minimal"
```

三個節點都開啟時，資源 xenA 將首先放置到一個節點上，然後是 xenD。xenB 與 xenC 會同時隨 xenD 一起配置，或是其中一項隨 xenD 一起配置。

如果一個節點失敗，表示可用的總記憶體太少，無法代管全部資源。xenA 與 xenD 會保證得到配置。但剩餘的兩個資源 xenB 與 xenC 中僅有一個會予以放置。由於他們的優先程度相同，因此結果會有多種。若要解決這種不確定的狀況，您需要為其中一個設定較高的優先程度。

4.5 如需更多資訊

<http://clusterlabs.org/>

Pacemaker 首頁，High Availability Extension 隨附的叢集資源管理員。

<http://linux-ha.org>

The High Availability Linux Project 首頁。

<http://clusterlabs.org/wiki/Documentation>

CRM 指令行介面：crm 指令行工具的簡介。

<http://clusterlabs.org/wiki/Documentation>

《*Pacemaker 1.0—Configuration Explained*》(Pacemaker 1.0 — 組態說明)：說明設定 Pacemaker 時用到的概念。包含全面詳盡的資訊，以供參考。

設定和管理叢集資源 (GUI)

若要設定和管理叢集資源，請使用圖形使用者介面 (Pacemaker GUI) 或 `crm` 指令行公用程式。有關指令行這種方式，請參閱第 6 章「設定和管理叢集資源 (指令行)」[第 83 頁]。

本章對 Pacemaker GUI 做了介紹，說明設定和管理叢集資源時所需執行的基本任務：建立基本與進階類型的資源 (群組或複製資源)，設定限制，指定容錯移轉節點與錯誤回復節點，設定資源監控，啟動、清理或移除資源，以及手動移轉資源。

GUI 支援由兩個套件提供：`pacemaker-mgmt` 套件包含 GUI 的後端 (`mgmtd` 精靈)。該套件必須安裝在您要使用 GUI 連接的所有叢集節點上。在要執行 GUI 的每一台機器上安裝 `pacemaker-mgmt-client` 套件。

5.1 Pacemaker GUI — 綜覽

若要啟動 Pacemaker GUI，請在指令行中輸入 `crm_gui`。若要存取組態與管理選項，須登入叢集。

5.1.1 連接至叢集

注意：使用者認證

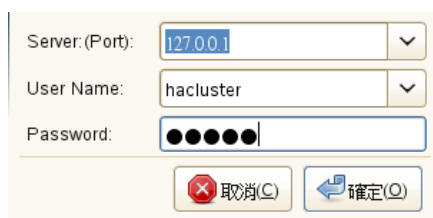
若要從 Pacemaker GUI 登入叢集，相關使用者必須是群組 `haclient` 的成員。安裝會建立名為 `hacluster` 的 Linux 使用者，其為群組 `haclient` 的成員。

在使用 Pacemaker GUI 之前，先為 `hacluster` 使用者設定密碼，或建立一個屬於群組 `haclient` 的新使用者。

在需要使用 Pacemaker GUI 連接的所有節點上執行此操作。

若要連接至叢集，請選取「連線」>「登入」。依預設，「伺服器」欄位會顯示 `localhost` 的 IP 位址，以及做為「使用者名稱」的 `hacluster`。輸入使用者密碼以繼續。

圖形 5.1 連接至叢集



The screenshot shows a login dialog box with the following fields and controls:

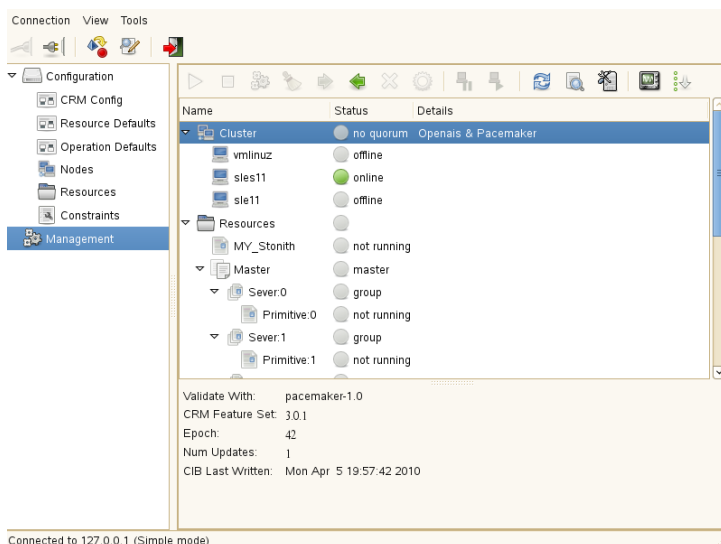
- Server (Port):** A dropdown menu showing `127.0.0.1`.
- User Name:** A dropdown menu showing `hacluster`.
- Password:** A text input field with six dots representing a masked password.
- Buttons:** At the bottom, there are two buttons: a red 'X' button labeled '取消' (Cancel) and a blue arrow button labeled '確定' (OK).

若要從遠端執行 Pacemaker GUI，請輸入做為「伺服器」之叢集節點的 IP 位址。對於「使用者名稱」，您也可以使用屬於 `haclient` 群組的任何其他使用者來連接叢集。

5.1.2 主視窗

連接之後，主視窗即會開啟：

圖形 5.2 Pacemaker GUI - 主視窗



若要檢視或修改 CRM、資源、節點或限制等叢集元件，請選取左側窗格中「組態」類別的相應子項目，然後使用右側窗格中可用的選項。此外，Pacemaker GUI 還可讓您輕鬆檢視、編輯、輸入和輸出 CIB 下列子項目的 XML 片段：「資源預設值」、「作業預設值」、「節點」、「資源」和「限制」。在視窗右上角選取任一「組態」子項目，然後選取「顯示」>「XML 模式」。

如果您已經設定了資源，按一下左側窗格中的「管理」類別可顯示叢集及其資源的狀態。此檢視窗還可讓您將節點設定為待機，以及修改節點的管理狀態(目前是否由叢集管理)。若要存取資源的主要功能(啟動、停止、清理或移轉資源)，請在右側窗格中選取資源，然後使用工具列中的圖示。您也可以在此資源上按一下滑鼠右鍵，然後從快顯功能表中選取相應的功能表項目。

Pacemaker GUI 還可讓您在不同檢視模式之間切換，以變更軟體的行為，隱藏或顯示某些元件：

簡單模式

可讓您在類似精靈的模式下新增資源。建立和修改資源時，會顯示子物件的常用索引標籤，您可以直接透過索引標籤新增該類型的物件。

可讓您在左側窗格中選取「*CRM 組態*」項目來檢視及變更所有可用的全域叢集選項。右側窗格隨後便會顯示目前設定的值。如果未對選項設定任何特定的值，便會顯示預設值。

進階模式

可讓您在類似精靈的模式或透過對話視窗新增資源。建立和修改資源時，如果 CIB 中已存在特定類型的子物件，則只會顯示相應的索引標籤。新增新的子物件時，將會提示您選取物件類型，這樣您便可新增所有受支援類型的子物件。

在左側窗格中選取「*CRM 組態*」項目時，僅顯示已實際設定之全域叢集選項的值。隱藏將自動使用預設值的所有叢集選項(因為尚未設定任何值)。在此模式下，全域叢集選項只能使用個別的組態對話方塊來修改。

Hack 模式

其功能與進階模式相同。允許您新增包含特定規則的其他屬性集，以使組態更為動態。例如，您可以根據資源所在的節點讓資源擁有不同的例項屬性。此外，您還可以為中繼屬性集新增以時間為基準的規則，以決定屬性生效的時間。

視窗的狀態列還會顯示目前處於使用中的模式。

以下幾節將引導您完成設定叢集選項與資源時需要執行的主要任務，並介紹使用 `hbgui` 管理資源的方式。除非另有說明，逐步指示反映的是在簡單模式下執行的程序。

5.2 設定全域叢集選項

全域叢集選項控制叢集在遇到特定情況時的運作方式。這些選項會分組到不同的集，可使用 `Pacemaker GUI` 等叢集管理工具和 `crm` 外圍程序來檢視和修改。多數情況下，可以使用預先定義的值。但是，為了讓叢集的關鍵功能正常運作，還需要在執行基本叢集設定後調整以下參數：

- `no-quorum-policy` 選項 [第34頁]

- stonith-enabled 選項 [第35頁]

過程 5.1 修改全域叢集選項

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 選取「檢視」>「簡單模式」。
- 3 在左側窗格中，選取「CRM 組態」以檢視全域叢集選項及其目前值。
- 4 根據叢集要求，將「無最低節點數規則」設定為適當值。
- 5 如果因某些原因需要停用圍籬區隔，請取消選取「已啟用 STONITH」。
- 6 按一下「套用」確認您的變更。

您可以隨時切換回所有選項的預設值，只需在左側窗格中選取「CRM 組態」，然後按一下「預設值」。

5.3 設定叢集資源

做為叢集管理員，您需要為您叢集中的伺服器上執行的所有資源或應用程式建立叢集資源。叢集資源可包括網站、電子郵件伺服器、資料庫、檔案系統、虛擬機器，以及其他您希望使用者隨時都可以存取的伺服器型應用程式或服務。

有關您可以建立之資源類型的綜覽，請參閱第 4.2.3 節「資源類型」[第37頁]。

5.3.1 建立簡單叢集資源

若要建立最基本的資源類型，請執行下列步驟：

過程 5.2 新增原始資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，選取「資源」並按一下「新增」>「原始資源」。

3 在下一個對話方塊中，設定資源的以下參數：

3a 輸入資源的唯一「ID」。

3b 在「類別」清單中，選取您要針對該資源使用的資源代辦類別：「*heartbeat*」、「*lsb*」、「*ocf*」或「*stonith*」。如需詳細資訊，請參閱第 4.2.2 節「受支援的資源代辦類別」[第36頁]。

3c 若選取了「*ocf*」類別，請另外指定 OCF 資源代辦的「提供者」。OCF 規格允許多個廠商提供相同的資源代辦。

3d 在「類型」清單中，選取要使用的資源代辦 (例如「*IPaddr*」或「*Filesystem*」)。此資源代辦的簡要描述會顯示在下方。

「類型」清單中顯示的選項取決於您所選的「類別」(對於 OCF 資源，還取決於「提供者」)。

3e 在「選項」下方，設定「資源的初始狀態」。

3f 若要让叢集監控資源是否正常，請啟用「新增監控作業」。

Add Primitive - Basic Settings

Required

ID:

Class:

Provider:

Type:

Description

Manages virtual IPv4 addresses.

This script manages IP alias IP addresses
It can add an IP alias, or remove one.

Options

Initial state of resource:

☒ Add monitor operation

- 4 按「下一步」。下一個視窗會顯示您已為該資源定義的參數摘要。其中會列出該資源所需的所有「例項屬性」。您需要對它們進行編輯，以將其設定為適當的值。依您的部署與設定而定，您可能還需要新增其他屬性。如需如何執行此操作的詳細資料，請參閱過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁]。
- 5 按需要設定了所有參數後，請按一下「套用」以完成該資源的組態設定。組態對話方塊會關閉，同時主視窗會顯示新增的資源。

建立資源期間或建立之後，可以為資源新增或修改以下參數：

- 例項屬性 — 決定資源控制的服務例項。若需更多資訊，請參考第 4.2.6 節「例項屬性」[第42頁]。
- 中繼屬性 — 告知 CRM 如何處理特定資源。若需更多資訊，請參考第 4.2.5 節「資源選項 (中繼屬性)」[第41頁]。
- 作業 — 需要使用它們來監控資源。若需更多資訊，請參考第 4.2.7 節「資源作業」[第44頁]。

過程 5.3 新增或修改中繼屬性與例項屬性

- 1 在 Pacemaker GUI 主視窗中，按一下左側窗格中的「資源」，以檢視已為叢集設定的資源。
- 2 在右側窗格中，選取要修改的資源，然後按一下「編輯」(或在該資源上連按兩下)。下一個視窗會顯示基本的資源參數，以及已為該資源定義的「中繼屬性」、「例項屬性」或「作業」。

Show: List Mode

Required

ID: my_primitive

Class: ocf

Provider: heartbeat

Type: IPAddr

Optional

Description

Manages virtual IPv4 addresses.

This script manages IP alias IP addresses
It can add an IP alias, or remove one.

Meta Attributes Instance Attributes Operations

Name	Value
ip	192.168.8.212

ID: nvpair-e2f36987-795f-459c-b445-7a3d7ba1924f

Name: ip

Value: 192.168.8.212

+ 加入(A)

編輯(E)

- 移除(R)

取消(C)

Reset

確定(O)

- 若要新增新的中繼屬性或例項屬性，請選取相應的索引標籤並按一下「新增」。
- 選取要新增之屬性的「名稱」。一個簡要「描述」即會顯示。
- 視需要指定屬性「值」。若不指定，系統將使用該屬性的預設值。
- 按一下「確定」以確認所做的變更。索引標籤上隨即會顯示新增的或修改的屬性。
- 按需要設定了所有參數後，請按一下「確定」以完成該資源的組態設定。組態對話方塊會關閉，同時主視窗會顯示修改的資源。

提示：資源的 XML 原始程式碼

Pacemaker GUI 可讓您檢視透過定義的參數產生的 XML 片段。如果是個別資源，請在資源組態對話方塊的右上角選取「顯示」>「XML 模式」。

若要存取所有已設定之資源的XML程式碼，請按一下左側窗格中的「資源」，然後在主視窗的右上角選取「顯示」>「XML 模式」。

顯示XML程式碼的編輯器可讓您「輸入」或「輸出」XML元素，或手動編輯XML程式碼。

5.3.2 建立 STONITH 資源

若要設定圍籬區隔，您需要設定一或多個 STONITH 資源。

過程 5.4 新增 STONITH 資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁]中所述登入叢集。
- 2 在左側窗格中，選取「資源」並按一下「新增」>「原始資源」。
- 3 在下一個對話方塊中，設定資源的以下參數：
 - 3a 輸入資源的唯一「ID」。
 - 3b 在「類別」清單中，選取資源代辦類別「*stonith*」。
 - 3c 從「類型」清單中，選取用於控制 STONITH 設備的 STONITH 外掛程式。此外掛程式的簡要描述會顯示在下方。
 - 3d 在「選項」下方，設定「資源的初始狀態」。
 - 3e 若要讓叢集監控圍籬區隔設備，請啟用「新增監控作業」。若需更多資訊，請參考第 9.4 節「監控圍籬區隔設備」[第120頁]。
- 4 按「下一步」。下一個視窗會顯示您已為該資源定義的參數摘要。其中會列出所選 STONITH 外掛程式需要的所有「例項屬性」。您需要對它們進行編輯，以將其設定為適當的值。依您的部署與設定而定，您可能還需要新增其他屬性或監控作業。如需如何執行此操作的詳細資料，請參閱過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁]與第 5.3.7 節「設定資源監控」[第70頁]。

- 5 按需要設定了所有參數後，請按一下「套用」以完成該資源的組態設定。組態對話方塊會關閉，同時主視窗會顯示新增的資源。

若要完成圍籬區隔組態，請新增限制或使用複製，或同時使用這兩種方式。如需詳細資訊，請參閱第 9 章「圍籬區隔與 STONITH」[第113頁]。

5.3.3 設定資源限制

設定所有資源只是工作的一部分。即使叢集瞭解所有必需的資源，可能仍然無法正確地對其進行處理。資源限制可讓您指定資源可在哪些叢集節點上執行，載入資源的順序以及特定資源所依賴的其他資源。

如需可用限制類型的綜覽，請參閱第 4.4.1 節「限制類型」[第46頁]。定義限制時，還需要指定分數。如需叢集中分數及其含義的詳細資訊，請參閱第 4.4.2 節「分數與無限大」[第47頁]。

下列程序將介紹如何建立不同類型的限制。

過程 5.5 新增或修改位置限制

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在 Pacemaker GUI 主視窗中，按一下左側窗格中的「限制」，檢視已為叢集設定的限制。
- 3 在左側窗格中，選取「限制」，然後按一下「新增」。
- 4 選取「資源位置」，然後按一下「確定」。
- 5 輸入限制的唯一「ID」。在您修改現有的限制時，其 ID 已定義好，會顯示在組態對話方塊中。
- 6 選取要對其定義限制的「資源」。清單會顯示已為叢集設定之所有資源的 ID。
- 7 為限制設定「分數」。正值表示資源可以在您於下方指定的「節點」上執行。負值表示資源不能在此節點上執行。+/- INFINITY 值則表示將「可以」變更為「必須」。

8 為限制選取「節點」。



Required

Show: List Mode

ID: my_location_constraint

Resource: MY_Stonith

Score: INFINITY

Node: sles11

+ 加入(A) 編輯(E) - 移除(R)

取消(C) Reset 確定(Q)

- 9 若將「節點」與「分數」欄位留為空白，您也可以透過按一下「新增」>「規則」來新增規則。若要新增生命期間，只需按一下「新增」>「生命期間」。
- 10 按需要設定了所有參數後，請按一下「確定」以完成該限制的組態設定。組態對話方塊會關閉，同時主視窗會顯示新增的或修改的限制。

過程 5.6 新增或修改並存限制

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在 Pacemaker GUI 主視窗中，按一下左側窗格中的「限制」，檢視已為叢集設定的限制。
- 3 在左側窗格中，選取「限制」，然後按一下「新增」。
- 4 選取「資源並存」，然後按一下「確定」。
- 5 輸入限制的唯一「ID」。在您修改現有的限制時，其 ID 已定義好，會顯示在組態對話方塊中。
- 6 選取做為並存來源的「資源」。清單會顯示已為叢集設定之所有資源的 ID。

若無法符合限制的要求，叢集可能會決定不允許執行該資源。

- 7 若將「資源」及「與資源」欄位留為空白，您也可以透過按一下「新增」>「資源集」來新增資源集。若要新增生命期間，只需按一下「新增」>「生命期間」。
- 8 在「與資源」中，定義並存目標。叢集會先決定放置此資源的位置，然後決定放置「資源」欄位中之資源的位置。
- 9 定義「分數」以決定兩個資源間的位置關係。正值表示資源應該在同一個節點上執行。負值表示資源不應該在同一個節點上執行。+/- INFINITY 值則表示將「應該」變更為「必須」。該分數會結合其他因素來決定將資源配置於何處。
- 10 如有需要，請指定其他參數，例如「資源角色」。

根據所選的參數與選項，系統會顯示一條簡要「描述」，說明要設定之並存限制的效果。

- 11 按需要設定了所有參數後，請按一下「確定」以完成該限制的組態設定。組態對話方塊會關閉，同時主視窗會顯示新增的或修改的限制。

過程 5.7 新增或修改順序限制

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在 Pacemaker GUI 主視窗中，按一下左側窗格中的「限制」，檢視已為叢集設定的限制。
- 3 在左側窗格中，選取「限制」，然後按一下「新增」。
- 4 選取「資源順序」，然後按一下「確定」。
- 5 輸入限制的唯一「ID」。在您修改現有的限制時，其 ID 已定義好，會顯示在組態對話方塊中。
- 6 使用「首先」，定義「接著」所指定之資源允許啟動之前必須先啟動的資源。
- 7 使用「接著」，定義在「首先」資源啟動之後將要啟動的資源。

根據所選的參數與選項，系統會顯示一條簡要「描述」，說明要設定之順序限制的效果。

8 視需要定義其他參數，例如：

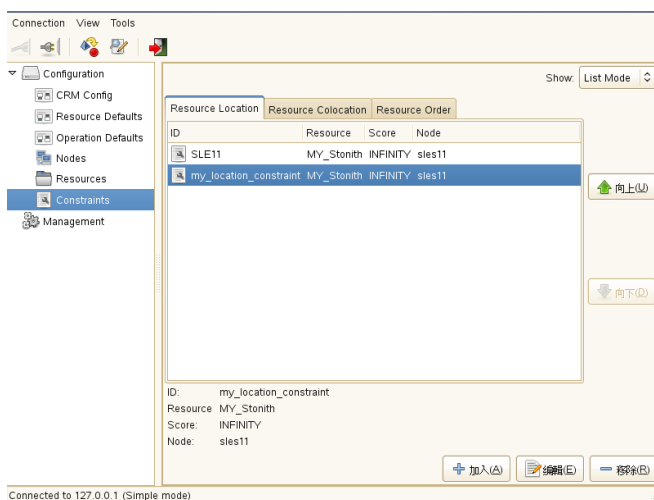
8a 指定「分數」。若大於零，則限制為強制性；否則僅做為建議。預設值為 INFINITY。

8b 指定「對稱式」的值。若為 true，則停止資源時使用相反順序。預設值為 true。

9 按需要設定了所有參數後，請按一下「確定」以完成該限制的組態設定。組態對話方塊會關閉，同時主視窗會顯示新增的或修改的限制。

您可在 Pacemaker GUI 的「限制」檢視窗中存取和修改已設定的所有限制。

圖形 5.3 Pacemaker GUI - 限制



5.3.4 指定資源容錯移轉節點

若資源失敗，系統會自動將其重新啟動。若在目前節點上無法將其重新啟動，或資源已在目前節點上失敗 N 次，則資源會嘗試容錯移轉至其他節點。您可以定義一個數值，讓資源在失敗該次數 (migration-threshold) 之後移轉至新

節點。若叢集中有兩個以上的節點，則特定資源容錯移轉所至的節點由 High Availability 軟體來選擇。

不過，您可以執行以下步驟來指定資源將容錯移轉所至的節點：

- 1 依過程 5.5「新增或修改位置限制」[第62頁]中所述為該資源設定位置限制。
- 2 依過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁]中所述將 `migration-threshold` 中繼屬性新增至該資源，並為該 `migration-threshold` 輸入「值」。該值應該為小於 INFINITY 的正數。
- 3 若要讓資源的 `failcount` 自動過期，請依過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁]中所述將 `failure-timeout` 中繼屬性新增至該資源，並為該 `failure-timeout` 輸入「值」。
- 4 若要指定具有資源優先設定的其他容錯移轉節點，請建立其他位置限制。

範例 4.2「移轉限定值 — 程序流程」[第48頁]中提供了一個範例，介紹叢集中有關移轉限定值和 `failcount` 的程序流。

您也可以隨時手動清理資源的 `failcount`，而不是等待資源的 `failcount` 自動過期。如需詳細資料，請參閱第 5.4.2 節「清理資源」[第78頁]。

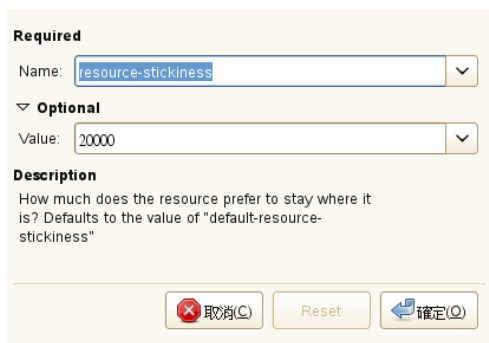
5.3.5 指定資源錯誤回復節點（資源相粘性）

當原始節點恢復連線且位於叢集中時，資源可以錯誤回復至該節點。若不想讓資源錯誤回復至其在容錯移轉之前所處的節點，或要為資源指定另一個要錯誤回復至的節點，您必須變更其資源相粘性的值。您可以在建立資源時或建立之後指定資源綁定。

有關不同資源相粘性值的含義，請閱第 4.4.4 節「錯誤回復節點」[第48頁]。

過程 5.8 指定資源相粘性

- 1 依過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁]中所述將 resource-stickness 中繼屬性新增至資源。



The screenshot shows a configuration window with a yellow background. It has three sections: 'Required' with a 'Name' dropdown set to 'resource-stickness'; 'Optional' with a 'Value' dropdown set to '20000'; and a 'Description' section containing the text: 'How much does the resource prefer to stay where it is? Defaults to the value of "default-resource-stickness"'. At the bottom are three buttons: a red 'X' button labeled '取消(C)', a 'Reset' button, and a blue checkmark button labeled '確定(O)'.

- 2 對於 resource-stickness 的「值」，請指定介於 $-\text{INFINITY}$ 與 INFINITY 之間的值。

5.3.6 根據負載影響設定資源的配置

並非所有資源都相同。有些資源 (例如 Xen 訪客) 要求代管它們的節點滿足其容量要求。如果放置資源後其所需的容量之和超過了提供的容量，資源效能便會下降 (甚至無法執行)。

鑒於此，High Availability Extension 允許您指定以下參數：

1. 特定節點提供的容量。
2. 特定資源要求的容量。
3. 配置資源的整體策略。

如需參數的詳細背景資訊及組態範例，請參閱第 4.4.5 節「依據負載影響放置資源」[第49頁]。

若要設定資源的要求和節點提供的容量，請依過程 5.9「新增或修改使用率屬性」[第68頁]中所述使用使用率屬性。您可以依據自己的偏好命名使用率屬性，依據組態需要定義任意數量的名稱/值對。

過程 5.9 新增或修改使用率屬性

在以下範例中，假設您已擁有叢集節點和資源的基本組態，現在還想要設定特定節點提供的容量和特定資源需要的容量。新增各使用率屬性的程序基本相同，只有在執行步驟 2 [第68頁] 和步驟 3 [第68頁] 時有些不同。

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」 [第54頁] 中所述登入叢集。
- 2 若要指定節點提供的容量：
 - 2a 在左側窗格中，按一下「節點」。
 - 2b 在右側窗格中，選取要設定其容量的節點，然後按一下「編輯」。
- 3 若要指定資源需要的容量：
 - 3a 在左側窗格中，按一下「資源」。
 - 3b 在右側窗格中，選取要設定其容量的資源，然後按一下「編輯」。
- 4 選取「使用率」索引標籤，然後按一下「新增」以新增使用率屬性。
- 5 輸入新屬性的「名稱」。您可以根據自己的偏好命名使用率屬性。
- 6 輸入屬性的「值」，然後按一下「確定」。屬性值必須為整數。
- 7 如果需要更多使用率屬性，請重複步驟 5 [第68頁] 至步驟 6 [第68頁]。
「使用率」索引標籤顯示您已為該節點或資源定義的使用率屬性摘要。
- 8 如果按需要設定了所有參數，請按一下「確定」關閉組態對話方塊。

圖形 5.4 「節點容量的範例組態」 [第69頁] 顯示了一個節點的組態，該節點為其上執行的資源提供 8 個 CPU 和 16 GB 記憶體：

圖形 5.4 節點容量的範例組態

Show: List Mode

Required

ID: bourbaki

Uname: bourbaki

Type: normal

Optional

Instance Attributes

Utilization

Name	Value
cpu	8
memory	16384

Up

Down

ID: nodes-bourbaki-cpu

Name: cpu

Value: 8

Add

Edit

Remove

Cancel

Reset

OK

其上一個資源需要 4096 KB 記憶體和 4 個 CPU 的節點的範例組態如下所示：

圖形 5.5 資源容量的範例組態

Show: List Mode

Required

ID: xen1

Class: ocf

Provider: heartbeat

Type: Xen

Optional

Description

Manages Xen unprivileged domains (DomUs).

Resource Agent for the Xen Hypervisor:

Manages Xen virtual machine instances hv_manning_cluster

Meta Attributes

Instance Attributes

Operations

Utilization

Name	Value
cpu	4
memory	4096

Up

Down

ID: xen1-utilization-cpu

Name: cpu

Value: 4

Add

Edit

Remove

Cancel

Reset

OK

設定完節點提供的容量和資源需要的容量後，須在全域叢集選項中設定配置策略，否則容量組態不會生效。可以使用幾個策略來排程負載：例如，您可將負載集中於最少的節點上，或在所有可用的節點上平均分攤。若需更多資訊，請參考第 4.4.5 節「依據負載影響放置資源」[第49頁]。

過程 5.10 設定配置策略

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 選取「檢視」>「簡單模式」。
- 3 在左側窗格中，選取「CRM 組態」以檢視全域叢集選項及其目前值。
- 4 根據要求將「配置策略」設定為適當的值。
- 5 如果因某些原因需要停用圍籬區隔，請取消選取「已啟用 STONITH」。
- 6 按一下「套用」確認您的變更。

5.3.7 設定資源監控


High Availability Extension 不僅可以偵測節點失敗，而且還可以偵測節點上個別資源失敗的時間。若要確定資源是否正在執行，必須設定針對該資源的資源監控。資源監控包括指定逾時與/或啟動延遲值以及間隔。該間隔會告知 CRM 應檢查資源狀態的頻率。您還可以設定特定參數，例如為 start 或 stop 作業設定 Timeout。

過程 5.11 新增或修改監控作業

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在 Pacemaker GUI 主視窗中，按一下左側窗格中的「資源」，以檢視已為叢集設定的資源。
- 3 在右側窗格中，選取要修改的資源，然後按一下「編輯」。下一個視窗會顯示基本的資源參數，以及已為該資源定義的中繼屬性、例項屬性及作業。
- 4 若要新增新的監控作業，請選取相應的索引標籤，然後按一下「新增」。
若要修改現有作業，請選取相應的項目，然後按一下「編輯」。

- 5 在「名稱」中，選取要執行的動作，例如「監控」、「啟動」，或「停止」。

下面顯示的參數視此處所作的選擇而定。



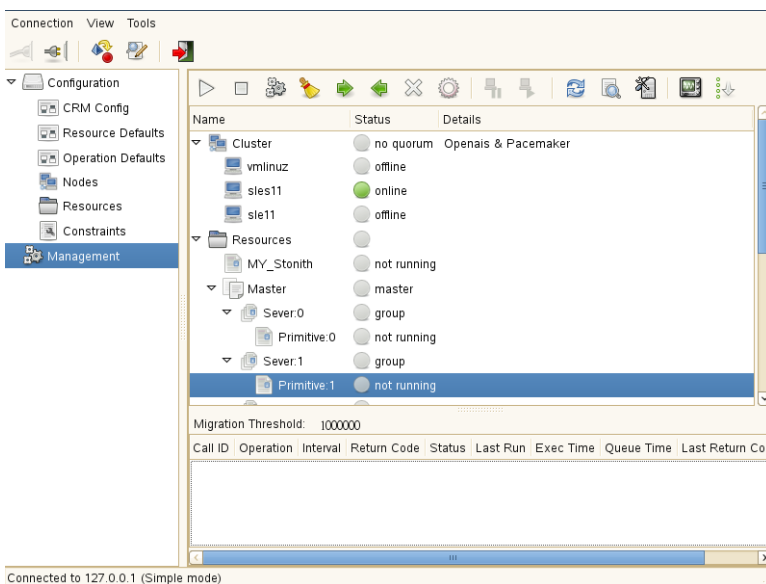
The screenshot shows a configuration window for a resource monitor. At the top, there is a 'Show:' dropdown menu set to 'List Mode'. Below this, the 'ID' is 'my_primitive-op-monitor-5s'. The 'Name' dropdown is set to 'monitor'. The 'Interval' dropdown is set to '5s'. The 'Timeout' dropdown is set to '20s'. Below these fields is an 'Optional' section with three buttons: '+ 加入(A)', '編輯(E)', and '- 移除(R)'. At the bottom, there are three buttons: '取消(Q)', 'Reset', and '確定(Q)'.

- 6 在「逾時」欄位中輸入值(以秒計)。作業在經過指定的逾時期間之後會被視為 failed。PE 將會決定需要採取的措施，或執行您在監控作業的「失敗時」欄位中指定的動作。
- 7 視需要展開「選擇性」區段，然後新增參數，例如「失敗時」(此動作失敗時應採取何措施?)或「必要」(執行此動作之前需要符合哪些條件?)。
- 8 按需要設定了所有參數後，請按一下「確定」以完成該資源的組態設定。組態對話方塊會關閉，同時主視窗會顯示修改的資源。

有關資源監控偵測到失敗時應執行的程序，請參閱第 4.3 節「資源監控」[第 45 頁]。

若要在 Pacemaker GUI 中檢視資源失敗，請在左側窗格中按一下「管理」，然後在右側窗格中選取要檢視其詳細資料的資源。對於失敗的資源，右側窗格的中間(位於「移轉限定值」項目下方)會顯示資源「失敗計數」和上次失敗時間。

圖形 5.6 檢視資源的 *failcount*



5.3.8 設定叢集資源群組

某些叢集資源依賴於其他元件或資源，要求每個元件或資源以特定順序啟動，並在同一個伺服器上執行。若要簡化此組態，您可以使用群組。

如需資源群組的範例和群組及其內容的詳細資訊，請參閱章節「群組」[第 38 頁]。

注意：空群組

群組至少須包含一個資源，否則組態視為無效。

過程 5.12 新增資源群組

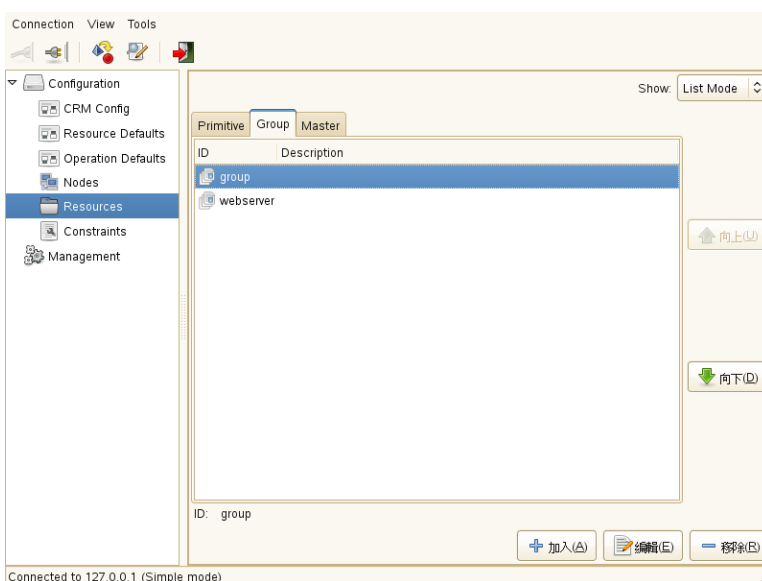
- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第 54 頁] 中所述登入叢集。
- 2 在左側窗格中，選取「資源」並按一下「新增」>「群組」。

- 3 輸入群組的唯一「*ID*」。
- 4 在「選項」下方，設定「資源的初始狀態」，然後按「下一步」。
- 5 在下一步中，您可以將原始資源新增為群組的子資源。建立這些資源的方式與過程 5.2「新增原始資源」[第57頁]中所述方式類似。
- 6 按需要設定了所有參數後，請按一下「套用」以完成該原始資源的組態設定。
- 7 在下一個視窗中，您可以再次選擇「原始資源」並按一下「確定」，繼續為群組新增子資源。

如果不想為群組新增更多原始資源，請按一下「取消」。下一個視窗會顯示已為該群組定義的參數摘要。其中會列出群組的「中繼屬性」與「原始資源」。「原始資源」索引標籤中資源的位置表示該資源在叢集中的啟動順序。

- 8 由於群組中資源的順序很重要，因此請使用「向上」與「向下」按鈕，對群組中的「原始資源」進行排序。
- 9 按需要設定了所有參數後，請按一下「確定」以完成該群組的組態設定。組態對話方塊會關閉，同時主視窗會顯示新建的或修改的群組。

圖形 5.7 Pacemaker GUI - 群組



假設您已依照過程 5.12「新增資源群組」[第72頁] 中的說明建立資源群組。以下程序描述如何修改群組以符合範例 4.1「Web 伺服器的資源群組」[第38頁]。

過程 5.13 將資源新增至現有群組

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，切換至「資源」檢視窗，然後在右側窗格中選取要修改的群組，並按一下「編輯」。下一個視窗會顯示基本的群組參數，以及已為該資源定義的中繼屬性與原始資源。
- 3 按一下「原始資源」索引標籤，然後按一下「新增」。
- 4 在下一個對話方塊中，設定以下參數以將 IP 位址新增為群組的子資源：
 - 4a 輸入唯一的「ID」，例如 my_ipaddress。
 - 4b 在「類別」清單中，選取「ocf」做為資源代辦類別。
 - 4c 對於 OCF 資源代辦的「提供者」，選取「heartbeat」。

- 4d** 在「類型」清單中，選取「*IPaddr*」做為資源代辦。
- 4e** 按「下一步」。
- 4f** 在「例項屬性」索引標籤中，選取「*IP*」項目並按一下「編輯」(或在「*IP*」項目上連按兩下)。
- 4g** 對於「值」，輸入所需的 IP 位址，例如 192.168.1.1。
- 4h** 按一下「確定」與「套用」。群組組態對話方塊會顯示新增的原始資源。
- 5** 再按一下「新增」以新增下一個子資源 (檔案系統與 Web 伺服器)。
- 6** 為每個子資源設定相應的參數 (類似於步驟步驟 4a [第74頁] 至步驟 4h [第75頁])，直到您已設定此群組的所有子資源。

Add Primitive - Basic Settings

Required

ID:

Class:

Provider:

Type:

Description

Manages virtual IPv4 addresses.

This script manages IP alias IP addresses
It can add an IP alias, or remove one.

Options

Initial state of resource:

☒ Add monitor operation

- 由於我們是依照子資源在叢集中的啟動順序對其進行設定，因此「原始資源」索引標籤中的順序已經是正確的。
- 7** 若需要變更群組的資源順序，請使用「向上」與「向下」按鈕對「原始資源」索引標籤中的資源進行排序。
- 8** 若要從群組中移除某個資源，請在「原始資源」索引標籤中選取該資源，然後按一下「移除」。

- 9 按一下「確定」以完成該群組的組態設定。組態對話方塊會關閉，同時主視窗會顯示修改的群組。

5.3.9 設定複製資源

您可能要讓某些資源同時在叢集的多個節點上執行。若要實現此目的，您必須將資源設定為複製資源。可以設定為複製資源的資源範例包括 STONITH 以及叢集檔案系統，如 OCFS2。您可以複製所提供的任何資源。相關資源的資源代辦會為此操作提供支援。您甚至可以對複製資源進行不同的設定，具體視代管它們的節點而定。

如需可用複製資源的綜覽，請參閱章節「複製品」[第40頁]。

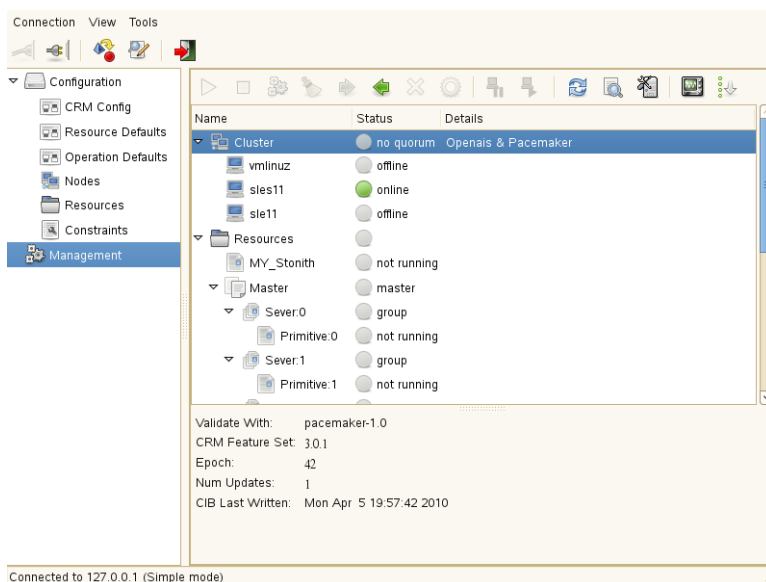
過程 5.14 新增或修改複製資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，選取「資源」並按一下「新增」>「複製」。
- 3 輸入複製資源的唯一「ID」。
- 4 在「選項」下方，設定「資源的初始狀態」。
- 5 啟用要為複製設定的相應選項，然後按「下一步」。
- 6 在下一步中，您可以新增「原始資源」或「群組」做為複製資源的子資源。建立這些資源的方式與過程 5.2「新增原始資源」[第57頁] 或過程 5.12「新增資源群組」[第72頁] 中所述方式類似。
- 7 按需要在複製資源組態對話方塊中設定了所有參數後，請按一下「套用」以完成該複製資源的組態設定。

5.4 管理叢集資源

除了能夠設定叢集資源外，Pacemaker GUI 還可讓您管理現有資源。若要切換至管理檢視窗並存取可用選項，請在左側窗格中按一下「管理」。

圖形 5.8 Pacemaker GUI - 管理



5.4.1 啟動資源

啟動叢集資源之前，請先確定該資源已正確設定。例如，如果想要使用 Apache 伺服器做為叢集資源，請先設定 Apache 伺服器並完成 Apache 組態設定，然後在叢集中啟動各個資源。

注意：不要對叢集管理的服務執行任何操作

使用 High Availability Extension 管理資源時，不能再啟動或停止相同的資源（例如在叢集之外手動開機或重新開機）。High Availability Extension 軟體負責所有服務的啟動或停止動作。

不過，如果您要檢查服務是否正確設定，請手動將其啟動，但請確保在 High Availability 接管之前將它再次停止。

若要對目前由叢集管理的資源進行管理，請先依第 5.4.5 節「變更資源的管理模式」[第81頁] 中所述將資源設定為不受管理模式。

使用 Pacemaker GUI 建立資源的過程中，您可以使用 `target-role` 中繼屬性設定資源的初始狀態。若將其值設定為 `stopped`，則資源建立後不會自動啟動。

過程 5.15 啟動新資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，按一下「管理」。
- 3 在右側窗格中的資源上按一下滑鼠右鍵，然後從快顯功能表中選取「啟動」(或使用工具列中的「啟動資源」圖示)。

5.4.2 清理資源

若資源失敗，系統會自動將其重新啟動，但每次失敗都會增加資源的 `failcount`。您可以使用 Pacemaker GUI 檢視資源的 `failcount`，方法是在左側窗格中按一下「管理」，然後在右側窗格中選取資源。如果資源失敗，右側窗格的中間(位於「移轉限定值」項目下方)會顯示其「失敗計數」。

如果已對該資源設定 `migration-threshold`，則一旦失敗次數達到該移轉限定值，就不再允許該節點執行該資源。

資源的 `failcount` 可自動重設(若為資源設定 `failure-timeout` 選項)，也可按如下方式手動重設。

過程 5.16 清理資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，按一下「管理」。
- 3 在右側窗格中的相應資源上按一下滑鼠右鍵，然後從快顯功能表中選取「清理資源」(或使用工具列中的「清理資源」圖示)。

如此會對指定節點上的指定資源執行指令 `crm_resource -C` 與 `crm_failcount -D`。

如需詳細資訊，另請參閱 `crm_resource(8)` [第217頁] 和 `crm_failcount(8)` [第208頁]。

5.4.3 移除叢集資源

如果需要從叢集移除資源，請按照下面的程序操作，以免出現組態錯誤：

注意：移除參考資源

若有限制正在參考叢集資源的 ID，則不能移除叢集資源。若無法刪除某個資源，請檢查參考該資源 ID 的位置，先從限制中移除該資源。

過程 5.17 移除叢集資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，按一下「管理」。
- 3 選取右側窗格中的相應資源。
- 4 依照過程 5.16「清理資源」[第78頁] 中的說明在所有節點上清理該資源。
- 5 「停止」資源。
- 6 移除與該資源相關的所有限制，否則將無法移除資源。

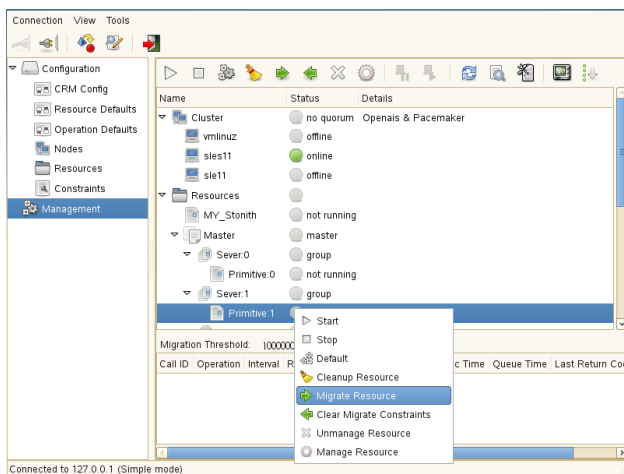
5.4.4 移轉叢集資源

如第 5.3.4 節「指定資源容錯移轉節點」[第65頁] 中所述，當軟體或硬體發生失敗時，叢集會自動對資源進行容錯移轉 (移轉) — 具體情況視您可以定義的某些參數 (例如移轉限定值或資源相粘性)。除此之外，您還可以手動將資源移轉至叢集資源中的其他節點。

過程 5.18 手動移轉資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。

- 2 在左側窗格中，按一下「管理」。
- 3 在右側窗格中的相應資源上按一下滑鼠右鍵，然後選取「移轉資源」。



- 4 在新視窗中，在「至節點」中選取要將資源移至的節點。如此會建立一個位置限制，其目的節點的分數為 INFINITY。
- 5 若只想暫時移轉資源，請啟用「持續時間」，並輸入資源移轉至新節點後應保留的時間。經過這段持續時間之後，資源可以移回其原始位置，也可以保留在目前的位置 (具體取決於資源相粘性)。
- 6 如果資源無法移轉 (若資源相粘性及限制總分大於目前節點上的 INFINITY)，請啟用「強制」選項。如此會為目前位置建立規則以及一個 -INFINITY 分數，以強制資源移動。

注意

這樣會禁止資源在此節點上執行，直到使用「清除移轉限制」移除限制或持續時間過期為止。

- 7 按一下「確定」以確認移轉。

若要讓資源回到原來的狀態，請執行下列步驟：

過程 5.19 清除移轉限制

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，按一下「管理」。
- 3 在右側窗格中的相應資源上按一下滑鼠右鍵，然後選取「清除移轉限制」。

此過程會使用 `crm_resource -U` 指令。資源可以移回其原始位置，也可以保留在目前的位置 (具體取決於資源相粘性)。

如需詳細資訊，請參閱 `crm_resource(8)` [第217頁] 或 <http://clusterlabs.org/wiki/Documentation> 上的《Pacemaker 1.0—Configuration Explained》(Pacemaker 1.0 — 組態說明)。可參閱其中的「Resource Migration」(資源移轉) 一節。

5.4.5 變更資源的管理模式

資源由叢集管理時，不能對其進行變更 (在叢集外部)。若要維護個別資源，可將相應資源設定成 不受管理模式，這樣您便可在叢集外部修改資源。

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第54頁] 中所述登入叢集。
- 2 在左側窗格中，按一下「管理」。
- 3 在右側窗格中的相應資源上按一下滑鼠右鍵，然後從快顯功能表中選取「不管理資源」。
- 4 完成該資源的維護任務之後，在右側窗格中的相應資源上再按一下滑鼠右鍵，然後選取「管理資源」。

從此時起，資源將重新由 High Availability Extension 軟體管理。

設定和管理叢集資源 (指令行)

若要設定和管理叢集資源，請使用圖形使用者介面 (Pacemaker GUI) 或 `crm` 指令行公用程式。有關 GUI 這種方式，請參閱第 5 章「設定和管理叢集資源 (GUI)」[第 53 頁]。

本章對指令行工具 `crm` 做了介紹，並對此工具、樣板的使用方式進行了概述，主要敘述了設定和管理叢集資源方面的資訊：建立基本與進階類型的資源 (群組與複製資源)，設定限制，指定容錯移轉節點與錯誤回復節點，設定資源監控，啟動、清理或移除資源，以及手動移轉資源。

6.1 `crm` 指令行工具 — 綜覽

安裝後，通常只需要 `crm` 指令。此指令有多個子指令，用於管理資源、CIB、節點、資源代辦及其他。執行 `crm help` 可取得所有可用指令的綜覽。該指令提供包含內嵌式範例的完整說明系統。

`crm` 指令的使用方式如下：

- **直接** 將所有子指令新增至 `crm`，再按 **Enter**，便會立即顯示輸出。例如，輸入 `crm help ra` 可獲取 `ra` 子指令 (資源代辦) 的相關資訊。
- **做為外圍程序檔** 使用 `crm` 和包含 `crm` 指令的程序檔。可以透過兩種方式來完成：

```
crm -f script.cli  
crm < script.cli
```

程序檔可以包含 `crm` 的任意指令。例如：

```
# A small example
status
node list
```

以井字號 (#) 開頭的任何行都是備註，系統會將其忽略。如果某行過長，則在結尾插入反斜線 (\)，然後換到下一行。

- **做為內部外圍程序進行互動** 輸入 `crm` 可進入內部外圍程序。提示變更為 `crm(live)#`。執行 `help` 可獲取可用子指令的綜覽。由於內部外圍程序有幾種不同層級的子指令，只需輸入一個子指令後按 **Enter** 便可「進入」此子指令。

例如，如果輸入 `resource`，便會進入資源管理層級。提示變更為 `crm(live)resource#`。若要離開內部外圍程序，請使用指令 `quit`、`bye` 或 `exit`。若需要回到上一個層級，請使用 `up`、`end` 或 `cd`。

如果輸入 `crm` 和相應的子指令 (不帶任何選項)，便可直接進入該層級。

內部外圍程序還支援 **Tab** 鍵對子指令和資源的補齊功能。輸入指令的開頭，然後按 `→|`，`crm` 便會補齊相應的物件。

注意：區分管理子指令與組態子指令

`crm` 工具具有管理功能 (子指令為 `resource` 和 `node`)，可用於進行組態設定 (`cib`、`configure`)。

以下小節概述了 `crm` 工具的一些重要方面。

6.1.1 顯示 OCF 資源代辦的相關資訊

由於您始終都需在叢集組態中處理資源代辦，因此 `crm` 工具包含了 `ra` 指令，用於獲取資源代辦的相關資訊並對其進行管理 (如需其他資訊，請參閱第 4.2.2 節「受支援的資源代辦類別」[第36頁])：

```
# crm ra
crm(live)ra#
```

指令 `classes` 可提供所有類別和提供者的清單：


```
crm(live)ra# classes
heartbeat
lsb
ocf / heartbeat linbit lvm2 ocfs2 pacemaker
stonith
```

若要取得某個類別(和提供者)的所有可用資源代辦的綜覽，請使用 `list` 指令：

```
crm(live)ra# list ocf
AoEtarget          AudibleAlarm       CTDB                ClusterMon
Delay              Dummy              EvmsSCC             Evmsd
Filesystem         HealthCPU          HealthSMART         ICP
IPaddr             IPaddr2            IPsrcaddr           IPv6addr
LVM                LinuxSCSI          MailTo              ManageRAID
ManageVE           Pure-FTPd          Raid1               Route
SAPDatabase        SAPInstance        SendArp             ServeRAID
...
```

若要檢視資源代辦的綜覽，請使用 `info`：

```
crm(live)ra# info ocf:drbd:linbit
This resource agent manages a DRBD resource
as a master/slave resource. DRBD is a shared-nothing replicated storage
device. (ocf:linbit:drbd)
```

Master/Slave OCF Resource Agent for DRBD

Parameters (* denotes required, [] the default):

```
drbd_resource* (string): drbd resource name
    The name of the drbd resource from the drbd.conf file.
```

```
drbdconf (string, [/etc/drbd.conf]): Path to drbd.conf
    Full path to the drbd.conf file.
```

Operations' defaults (advisory minimum):

```
start          timeout=240
promote        timeout=90
demote         timeout=90
notify        timeout=90
stop          timeout=100
monitor_Slave_0 interval=20 timeout=20 start-delay=1m
monitor_Master_0 interval=10 timeout=20 start-delay=1m
```

按 **Q** 鍵可離開檢視器。組態範例可在附錄 A 設定簡單測試資源的範例[第337頁]中找到。

提示：直接使用 **crm**

前面的範例中使用了 **crm** 指令的內部外圍程序。但是，您不一定要使用該外圍程序。如果將相應的子指令新增至 **crm**，也可獲得同樣的結果。例如，在外圍程序中輸入 **crm ra list ocf** 可列出所有 **OCF** 資源代辦。

6.1.2 使用範本

樣板是現有的叢集組態，只需略做調整就能符合特定使用者的需求。當某樣板建立組態時，便會有警告訊息發出提示，您稍後可以在進一步自定時編輯該提示。

以下程序說明如何建立一個簡單但功能齊備的 **Apache** 組態：

1 以 **root** 身分登入。

2 啟動 **crm** 工具：

```
# crm configure
```

3 從樣板建立新的組態：

3a 切換至 **template** 子指令：

```
crm(live)configure# template
```

3b 列出可用的樣板：

```
crm(live)configure template# list templates  
gfs2-base    filesystem  virtual-ip  apache      clvm        ocfs2       gfs2
```

3c 決定需要的樣板。由於現在需要 **Apache** 組態，因此請選擇 **apache** 樣板：

```
crm(live)configure template# new intranet apache  
INFO: pulling in template apache  
INFO: pulling in template virtual-ip
```

4 定義參數：

4a 列出剛才建立的組態：

```
crm(live)configure template# list
intranet
```

4b 顯示您必須填寫的必要變更：

```
crm(live)configure template# show
ERROR: 23: required parameter ip not set
ERROR: 61: required parameter id not set
ERROR: 65: required parameter configfile not set
```

4c 啟用偏好的文字編輯器，填寫步驟 4b [第87頁] 中顯示為錯誤的所有行：

```
crm(live)configure template# edit
```

5 顯示組態並檢查其是否有效 (視步驟 4c [第87頁] 中輸入的組態而定，可能會顯示粗體文字)：

```
crm(live)configure template# show
primitive virtual-ip ocf:heartbeat:IPaddr \
    params ip="192.168.1.101"
primitive apache ocf:heartbeat:apache \
    params configfile="/etc/apache2/httpd.conf"
monitor apache 120s:60s
group intranet \
    apache virtual-ip
```

6 套用組態：

```
crm(live)configure template# apply
crm(live)configure# cd ..
crm(live)configure# show
```

7 將變更提交至 CIB：

```
crm(live)configure# commit
```

如果您瞭解詳細資料，還可以使指令更簡單。您可以在外圍程序中使用以下指令來彙總上述程序：

```
crm configure template \
    new intranet apache params \
    configfile="/etc/apache2/httpd.conf" ip="192.168.1.101"
```

如果您在內部 `crm` 外圍程序中，請使用以下指令：

```
crm(live)configure template# new intranet apache params \  
configfile="/etc/apache2/httpd.conf" ip="192.168.1.101"
```

但是，之前的指令只是從樣板建立其組態，而不會套用或提交至 CIB。

6.1.3 使用非正式組態進行測試

非正式組態用於測試不同的組態案例。如果您已建立幾個非正式組態，則可以逐個進行測試以查看變更的效果。

一般程序如下：

- 1 開啟外圍程序，切換為 `root` 身分。

- 2 使用以下列指令啟動 `crm` 外圍程序：

```
crm configure
```

- 3 建立新的非正式組態：

```
crm(live)configure# cib new myNewConfig  
INFO: myNewConfig shadow CIB created
```

- 4 如果要將目前的即時組態複製至您的非正式組態，請使用以下指令，否則請跳過此步驟：

```
crm(myNewConfig)# cib reset myNewConfig
```

之前的指令可讓您以後修改任何現有資源時更為方便。

- 5 和平常一樣進行變更。建立非正式組態之後，所有變更即會套用至該組態。若要儲存所有變更，請使用以下指令：

```
crm(myNewConfig)#
```

- 6 如果重新需要使用即時叢集組態，請使用以下指令切換回來：

```
crm(myNewConfig)configure# cib use live  
crm(live)#
```

6.1.4 組態變更除錯

在將組態變更載入回叢集之前，建議您先使用 `pctest` 檢閱變更。`pctest` 可顯示提交變更後將發生之動作的圖表。您需要 `graphviz` 套件才能顯示圖表。以下範例是一份記錄，新增了監控作業：

```
# crm configure
crm(live)configure# show fence-node2
primitive fence-node2 stonith:apcsmart \
    params hostlist="node2"
crm(live)configure# monitor fence-node2 120m:60s
crm(live)configure# show changed
primitive fence-node2 stonith:apcsmart \
    params hostlist="node2" \
    op monitor interval="120m" timeout="60s"
crm(live)configure# pctest
crm(live)configure# commit
```

6.2 設定全域叢集選項

全域叢集選項控制叢集在遇到特定情況時的運作方式。您可以使用 `crm` 工具對它們進行檢視及修改。多數情況下，可以使用預先定義的值。但是，為了讓叢集的關鍵功能正常運作，還需要在執行基本叢集設定後調整以下參數：

- `no-quorum-policy` 選項 [第34頁]
- `stonith-enabled` 選項 [第35頁]

過程 6.1 使用 `crm` 修改全域叢集選項

- 1 開啟外圍程序，切換為 `root` 身分。
- 2 輸入 `crm configure` 開啟內部外圍程序。
- 3 使用以下指令，僅為只包含兩個節點的叢集設定選項：

```
crm(live)configure# property no-quorum-policy=ignore
crm(live)configure# property stonith-enabled=false
```

- 4 顯示您的變更：

```
crm(live)configure# show
property $id="cib-bootstrap-options" \
```

```
dc-version="1.1.1-530add2a3721a0ecccb24660a97dbfdaa3e68f51" \  
cluster-infrastructure="openais" \  
expected-quorum-votes="2" \  
no-quorum-policy="ignore" \  
stonith-enabled="false"
```

5 提交您的變更並離開：

```
crm(live)configure# commit  
crm(live)configure# exit
```

6.3 設定叢集資源

做為叢集管理員，您需要為您叢集中的伺服器上執行的所有資源或應用程式建立叢集資源。叢集資源可包括網站、電子郵件伺服器、資料庫、檔案系統、虛擬機器，以及其他您希望使用者隨時都可以存取的伺服器型應用程式或服務。

有關您可以建立之資源類型的綜覽，請參閱第 4.2.3 節「資源類型」[第37頁]。

6.3.1 建立叢集資源

系統提供了三種適用於叢集的 RA (資源代辦)，如需背景資訊，請參閱第 4.2.2 節「受支援的資源代辦類別」[第36頁]。若要建立叢集資源，請使用 `crm` 工具。若要將新資源新增至叢集，請執行以下步驟：

- 1 開啟外圍程序，切換為 `root` 身分。
- 2 輸入 `crm configure` 開啟內部外圍程序
- 3 設定原始 IP 位址：

```
crm(live)configure# primitive myIP ocf:heartbeat:IPaddr \  
    params ip=127.0.0.99 op monitor interval=60s
```

以上指令設定名為 `myIP` 的「原始」IP 位址。您需要選擇類別 (此處為 `ocf`)、提供者 (`heartbeat`) 和類型 (`IPaddr`)。此外，此原始資源還需要其他參數，例如 IP 位址。將位址變更為您的設定。

- 4 顯示並檢閱已進行的變更：

```
crm(live)configure# show
```

5 提交變更，使之生效：

```
crm(live)configure# commit
```

6.3.2 NFS 伺服器的範例組態

若要設定 NFS 伺服器，須完成以下操作：

- 1 設定 DRBD。
- 2 設定檔案系統資源。
- 3 設定 NFS 伺服器及 IP 位址。

以下小節中將介紹如何完成此作業。

設定 DRBD

在開始 DRBD High Availability 組態設定之前，請先手動設定 DRBD 設備。基本上，此操作就是設定 DRBD 以使其同步。具體的程序將在第 13 章「分散式複製區塊設備 (*DRBD*)」[第 147 頁] 中介紹。現在，假設您已設定資源 `r0`，它可從叢集的兩個節點上的設備 `/dev/drbd_r0` 來存取。

DRBD 資源屬於 OCF 主要/從屬資源。這一點可從 DRBD 資源代辦中繼資料的描述中看出。不過，重要的是中繼資料的 `actions` 區段中包含 `promote` 和 `demote` 動作。它們是主要/從屬資源的必要動作，通常不適用於其他資源。

對於 High Availability，主要/從屬資源可在不同節點上擁有多個主要資源。甚至可能在同一個節點上既有主要資源又有從屬資源。因此，將使用特定方式設定此資源，使得恰有一個主要資源和一個從屬資源分別在不同的節點上執行。要實現此目的，可使用主要資源的 `meta` 屬性。主要/從屬資源是 High Availability 中一種特殊的複製資源。每個主要資源和每個從屬資源皆計數為一個複製資源。

請執行下列步驟設定 DRBD 資源：

- 1 開啟外圍程序，切換為 `root` 身分。
- 2 輸入 `crm configure` 開啟內部外圍程序。

3 如果您的叢集只有兩個節點，請設定每個 ms 資源的以下內容：

```
crm(live)configure# primitive my-stonith stonith:external/ipmi ...
crm(live)configure# ms ms_drbd_r0 res_drbd_r0 meta \
    globally-unique=false ...
crm(live)configure# property no-quorum-policy=ignore
crm(live)configure# property stonith-enabled=true
```

4 建立原始 DRBD 資源：

```
crm(live)configure# primitive drbd_r0 ocf:linbit:drbd params \
    drbd drbd_resource=r0 op monitor interval="30s"
```

5 建立主要/從屬資源：

```
crm(live)configure# ms ms_drbd_r0 res_drbd_r0 meta master-max=1 \
    master-node-max=1 clone-max=2 clone-node-max=1 notify=true
```

6 指定並存和順序限制：

```
crm(live)configure# colocation fs_on_drbd_r0 inf: res_fs_r0
ms_drbd_r0:Master
crm(live)configure# order fs_after_drbd_r0 inf: ms_drbd_r0:promote
res_fs_r0:start
```

7 使用 show 指令顯示您的變更。

8 使用 commit 指令提交您的變更。

設定檔案系統資源

filesystem 資源設定為具有 DRBD 的 OCF 原始資源。當有啟動和停止要求時，它需要在目錄中掛接和卸載設備。在本範例中，設備為 /dev/drbd_r0，要用做掛接點的目錄為 /srv/failover。所使用的檔案系統為 xfs。

在 crm 外圍程序中使用以下指令來設定檔案系統資源：

```
crm(live)# configure
crm(live)configure# primitive filesystem_resource \
    ocf:linbit:drbd \
    params device=/dev/drbd_r0 directory=/srv/failover fstype=xfs
```


NFS 伺服器與 IP 位址

若要使 NFS 伺服器始終在同一個 IP 位址上可用，請使用一個額外的 IP 位址及機器執行常規作業所用的 IP 位址。然後，除系統的 IP 位址之外，還會將此 IP 位址指定給使用中的 NFS 伺服器。

NFS 伺服器與 NFS 伺服器的 IP 位址在同一台機器上應始終處於使用中狀態。在此情況下，啟動順序並不十分重要。甚至可以同時啟動它們。這些是群組資源的一般要求。

在開始 High Availability RA 組態設定之前，請使用 YaST 設定 NFS 伺服器。不要讓系統啟動 NFS 伺服器。只需設定組態檔案。如果要手動執行此操作，請參閱手冊頁 `exports(5)` (`man 5 exports`)。組態檔案為 `/etc/exports`。NFS 伺服器設定為 LSB 資源。

使用 High Availability RA 組態完整設定 IP 位址。不需要在系統中進行其他修改。IP 位址 RA 為 OCF RA。

```
crm(live)# configure
crm(live)configure# primitive nfs_resource ocf:nfsserver \
    params nfs_ip=10.10.0.1 nfs_shared_infodir=/shared
crm(live)configure# primitive ip_resource ocf:heartbeat:IPaddr \
    params ip=10.10.0.1
crm(live)configure# group nfs_group nfs_resource ip_resource
crm(live)configure# show
primitive ip_res ocf:heartbeat:IPaddr \
    params ip="192.168.1.10"
primitive nfs_res ocf:heartbeat:nfsserver \
    params nfs_ip="192.168.1.10" nfs_shared_infodir="/shared"
group nfs_group nfs_res ip_res
crm(live)configure# commit
crm(live)configure# end
crm(live)# quit
```

6.3.3 建立 STONITH 資源

從 `crm` 角度看，STONITH 設備只是另一個資源。若要建立 STONITH 資源，請執行下列步驟：

- 1 開啟外圍程序，切換為 `root` 身分。
- 2 輸入 `crm` 開啟內部外圍程序。

3 使用以下指令取得所有 STONITH 類型清單：

```
crm(live)# ra list stonith
apcmaster          apcsmart          baytech
cyclades           drac3            external/drac5
external/hmchttp   external/ibmrsa   external/ibmrsa-telnet
external/ipmi       external/kdumpcheck external/rackpdu
external/riloe      external/sbd      external/ssh
external/vmware     external/xen0     external/xen0-ha
ibmhmc             ipmilan          meatware
null               nw_rpc100s       rcd_serial
rps10              ssh              suicide
```

4 從上述清單中選擇一種 STONITH 類型並檢視可能的選項清單。使用以下指令：

```
crm(live)# ra info stonith:external/ipmi
IPMI STONITH external device (stonith:external/ipmi)
```

ipmitool based power management. Apparently, the power off method of ipmitool is intercepted by ACPI which then makes a regular shutdown. If case of a split brain on a two-node it may happen that no node survives. For two-node clusters use only the reset method.

Parameters (* denotes required, [] the default):

```
hostname (string): Hostname
    The name of the host to be managed by this STONITH device.
...
```

5 以 stonith 類別以及您在步驟 4 中所選的類型，然後根據需要設定相應參數來建立 STONITH 資源，例如：

```
crm(live)# configure
crm(live)configure# primitive my-stonith stonith:external/ipmi \
    params hostname="node1"
    ipaddr="192.168.1.221" \
    userid="admin" passwd="secret" \
    op monitor interval=60m timeout=120s
```

6.3.4 設定資源限制

設定所有資源只是工作的一部分。即使叢集瞭解所有必需的資源，可能仍然無法正確地對其進行處理。例如，嘗試不在 drbd 的從屬節點上掛接檔案系統（實際上，對 drbd 執行此操作將會失敗）。定義相關限制，讓此類資訊適用於叢集。

如需限制的詳細資訊，請參閱第 4.4 節「資源限制」[第46頁]。

位置限制

可以為每個資源多次新增此類限制。系統會對指定的資源評估所有位置限制。下面是一個簡單的範例，可將在名為 `earth` 的節點上執行 ID 為 `fs1-loc` 的資源的優先設定設為 100：

```
crm(live)configure# location fs1-loc fs1 100: earth
```

另外一個範例為包含 `pingd` 的位置：

```
crm(live)configure# primitive pingd pingd \  
    params name=pingd dampen=5s multiplier=100 host_list="r1 r2"  
crm(live)configure# location node_pref internal_www \  
    rule 50: #uname eq node1 \  
    rule pingd: defined pingd
```

並存限制

`collocation` 指令用於定義應在相同或不同主機上執行的資源。

您只能設定 `+inf` 或 `-inf` 範圍，即定義必須始終或永不在同一個節點上執行的資源。您也可以使用非 `inf` 範圍。在這情況下，並存限制只是一種建議，叢集可以決定不遵循該範圍，以便在有衝突發生時不停止其他資源。

例如，對於 ID 分別為 `filesystem_resource` 和 `nfs_group` 且始終位於同一個主機上的資源，請使用以下限制：

```
crm(live)configure# colocation nfs_on_filesystem inf: nfs_group  
filesystem_resource
```

對於主要從屬組態，除本地執行資源之外，還必須瞭解目前節點是否為主要節點。

順序限制

有時需要提供資源動作或操作的順序。例如，在設備可用於系統之前，不能掛接檔案系統。順序限制可用於在另一個資源符合特定條件(例如啟動、停止或升級為主要資源)的前後啟動或停止服務。在 `crm` 外圍程序中使用以下指令設定順序限制：

```
crm(live)configure# order nfs_after_filesystem mandatory: group_nfs
filesystem_resource
```

範例組態的限制

如果沒有其他限制，本章中所用的範例可能不會起作用。所有資源皆必須與 drbd 資源的主要資源在同一機器上執行，這是基本要求。在任何其他資源啟動之前，drbd 資源必須成為主要資源。若 DRBD 設備不是主要資源，嘗試掛接該設備時必定失敗。以下限制必須滿足：

- 檔案系統必須始終與 DRBD 資源的主要資源位於同一個節點上。

```
crm(live)configure# colocation filesystem_on_master inf: \
filesystem_resource drbd_resource:Master
```

- NFS 伺服器與 IP 位址必須與檔案系統位於同一個節點上。

```
crm(live)configure# colocation nfs_with_fs inf: \
nfs_group filesystem_resource
```

- NFS 伺服器與 IP 位址將在檔案系統完成掛接之後啟動：

```
crm(live)configure# order nfs_second mandatory: \
filesystem_resource:start nfs_group
```

- 在將 DRBD 資源升級為節點上的主要資源之後，才能在此節點上掛接檔案系統。

```
crm(live)configure# order drbd_first inf: \
drbd_resource:promote filesystem_resource
```

6.3.5 指定資源容錯移轉節點

若要判斷資源容錯移轉，請使用 meta 屬性 migration-threshold。例如：

```
crm(live)configure# location r1-node1 r1 100: node1
```

通常，r1 偏好在節點 node1 上執行。如果失敗，則會檢查 migration-threshold 並與 failcount 進行比較。若 failcount >= migration-threshold，則將資源移轉至優先設定次佳的節點。

根據 `start-failure-is-fatal` 選項的值，啟動失敗會將 `failcount` 設定為 `inf`。停止失敗將導致圍籬區隔。如果未定義 `STONITH`，則資源根本不會移轉。

如需綜覽，請參閱第 4.4.3 節「容錯移轉節點」[第47頁]。

6.3.6 指定資源錯誤回復節點 (資源相粘性)

當原始節點恢復連線且位於叢集中時，資源可以錯誤回復至該節點。若不想讓資源錯誤回復至其在容錯移轉之前所處的節點，或要為資源指定另一個要錯誤回復至的節點，您必須變更其資源相粘性的值。您可以在建立資源時或建立之後指定資源綁定。

如需綜覽，請參閱第 4.4.4 節「錯誤回復節點」[第48頁]。

6.3.7 根據負載影響設定資源的配置

根據負載影響設定資源的配置

並非所有資源都相同。有些資源 (例如 `Xen` 訪客) 要求代管它們的節點滿足其容量要求。如果放置資源後其所需的容量之和超過了提供的容量，資源效能便會下降 (甚至無法執行)。

鑒於此，`High Availability Extension` 允許您指定以下參數：

1. 特定節點提供的容量。
2. 特定資源要求的容量。
3. 配置資源的整體策略。

如需參數的詳細背景資訊及組態範例，請參閱第 4.4.5 節「依據負載影響放置資源」[第49頁]。

若要設定資源的要求和節點提供的容量，請依過程 5.9「新增或修改使用率屬性」[第68頁] 中所述使用使用率屬性。您可以依據自己的偏好命名使用率屬性，依據組態需要定義任意數量的名稱/值對。

在以下範例中，假設您已擁有叢集節點和資源的基本組態，現在還想要設定特定節點提供的容量和特定資源需要的容量。

過程 6.2 使用 *crm* 新增或修改使用率屬性

- 1 使用以下列指令啟動 *crm* 外圍程序：

```
crm configure
```

- 2 若要指定節點提供的容量，請使用以下指令，並以您的節點名稱取代佔位符 *NODE_1*：

```
crm(live)configure# node NODE_1 utilization memory=16384 cpu=8
```

透過設定以上的值，我們假設 *NODE_1* 為資源提供 16GB 的記憶體和 8 個 CPU 核心。

- 3 若要指定資源需要的容量，請使用：

```
crm(live)configure# primitive xen1 ocf:heartbeat:Xen ... \  
    utilization memory=4096 cpu=4
```

如此，資源會佔用節點 *nodeA* 的 4096 KB 記憶體和 4 個 *cpu*。

- 4 使用 *property* 指令設定配置策略：

```
crm(live)configure# property ...
```

以下四個值適用於配置策略：

```
propertyplacement-strategy=default
```

預設不會將使用率值納入考量。資源依據位置分數配置。如果分數相同，則在各節點上平均分配資源。

```
propertyplacement-strategy=utilization
```

若節點擁有的可用容量足以滿足資源的要求，則在確定該節點是否適合時，會將使用率值納入考量。但是，依舊會依據配置給節點的資源數量完成負載平衡。

```
propertyplacement-strategy=minimal
```

在確定節點是否有能力服務資源時，會將使用率值納入考量；會嘗試將資源集中在最少的節點上，以便為其餘節點節省耗電。

```
propertyplacement-strategy=balanced
```

在確定節點是否有能力服務資源時，會將使用率值納入考量；會嘗試將資源平均分佈，以優化資源效能。

配置策略的效能最佳，雖然不使用複雜的啟發式解析程式，卻總能獲得最佳的配置效果。請確保資源的優先程度已正確設定，以便先排程最重要的資源。

5 離開 crm 外圍程序之前，先提交您的變更：

```
crm(live)configure# commit
```

以下範例描述一個由三個同級別節點組成的叢集和 4 台虛擬機器：

```
crm(live)configure# node node1 utilization memory="4000"
crm(live)configure# node node2 utilization memory="4000"
crm(live)configure# node node3 utilization memory="4000"
crm(live)configure# primitive xenA ocf:heartbeat:Xen \
    utilization memory="3500" meta priority="10"
crm(live)configure# primitive xenB ocf:heartbeat:Xen \
    utilization memory="2000" meta priority="1"
crm(live)configure# primitive xenC ocf:heartbeat:Xen \
    utilization memory="2000" meta priority="1"
crm(live)configure# primitive xenD ocf:heartbeat:Xen \
    utilization memory="1000" meta priority="5"
crm(live)configure# property placement-strategy="minimal"
```

這三個節點啟動後，會先將 `xenA` 配置於一個節點上，接著是 `xenD`。`xenB` 和 `xenC` 會配置在一起，或其中一個與 `xenD` 配置在一起。

如果一個節點失敗，表示可用的總記憶體太少，無法代管全部資源。如此 `xenA` 及 `xenD` 都會予以配置，但 `xenB` 與 `xenC` 只有其中一個會予以配置，因為它們的優先程度相同，因此結果尚不確定。若要解決這種不確定的狀況，您需要為其中一個設定較高的優先程度。

6.3.8 設定資源監控

若要監控資源，可以使用兩種方法：使用 `op` 關鍵字定義監控操作或使用 `monitor` 指令。以下範例設定了一個 `Apache` 資源，並每隔 30 分鐘使用一次 `op` 關鍵字對其進行監控：

```
crm(live)configure# primitive apache apache \
    params ... \
    op monitor interval=60s timeout=30s
```

下列指令可達到相同目的：

```
crm(live)configure# primitive apache apache \  
    params ...  
crm(live)configure# monitor apache 60s:30s
```

如需綜覽，請參閱第 4.3 節「資源監控」[第45頁]。

6.3.9 設定叢集資源群組

叢集其中一個最常見的元素是需要存放在一起的一組資源。請按照順序啟動它們，停止時採用相反順序。若要簡化此組態，您可以使用群組。以下範例將建立兩個原始資源 (一個 IP 位址和一個電子郵件資源)：

1 以系統管理員身分執行 `crm` 指令。提示變更為 `crm(live)`。

2 設定原始資源：

```
crm(live)# configure  
crm(live)configure# primitive Public-IP ocf:IPaddr:heartbeat \  
    params ip=1.2.3.4  
crm(live)configure# primitive Email lsb:exim
```

3 以正確的順序按照相應的識別碼對原始資源分組：

```
crm(live)configure# group shortcut Public-IP Email
```

如需綜覽，請參閱章節「群組」[第38頁]。

6.3.10 設定複製資源

複製最初被認為是啟動 IP 資源的 N 個例項並在整個叢集進行分配以達到負載平衡的一種便利方法。經證明，複製還有其他各種各樣的用途，包括與 DLM 整合、圍籬區隔子系統和 OCFS2。只要資源代辦支援，就可以複製任何資源。

如需所複製之資源的詳細資訊，請參閱章節「複製品」[第40頁]。

建立匿名複製資源

若要建立匿名複製資源，首先要建立原始資源，然後使用 `clone` 指令參考它。請進行下列幾項操作：

- 1 開啟外圍程序，切換為 root 身分。
- 2 輸入 `crm configure` 開啟內部外圍程序。
- 3 設定原始資源，例如：

```
crm(live)configure# primitive Apache lsb:apache
```

- 4 複製原始資源：

```
crm(live)configure# clone apache-clone Apache
```

建立可設定狀態的/多狀態的複製資源

若要建立可設定狀態的複製資源，請先建立原始資源，然後建立主要/從屬資源。

- 1 開啟外圍程序，切換為 root 身分。
- 2 輸入 `crm configure` 開啟內部外圍程序。
- 3 設定原始資源。視需要變更間隔：

```
crm(live)configure# primitive myRSC ocf:myCorp:myAppl \  
    op monitor interval=60 \  
    op monitor interval=61 role=Master
```

- 4 建立主要從屬資源：

```
crm(live)configure# clone apache-clone Apache
```

6.4 管理叢集資源

除了能夠設定叢集資源外，`crm` 工具還可讓您管理現有資源。以下小節對此進行了概述。

6.4.1 啟動新的叢集資源

若要啟動新的叢集資源，您需要相應的識別碼。請執行下列步驟：

1 開啟外圍程序，切換為 `root` 身分。

2 輸入 `crm` 開啟內部外圍程序。

3 切換至資源層級：

```
crm(live)# resource
```

4 使用 `start` 啟動資源，然後按 `→|` 鍵顯示所有已知資源：

```
crm(live)resource# start start ID
```

6.4.2 清理資源

若資源失敗，系統會自動將其重新啟動，但每次失敗都會增加資源的 `failcount`。如果已對該資源設定 `migration-threshold`，則一旦失敗次數達到該移轉限定值，就不再允許該節點執行該資源。

1 開啟外圍程序並以 `root` 使用者身分登入。

2 取得所有資源的清單：

```
crm resource list
...
Resource Group: dlm-clvm:1
    dlm:1 (ocf::pacemaker:controld) Started
    clvm:1 (ocf::lvm2:clvmd) Started
    cmirrord:1 (ocf::lvm2:cmirrord) Started
```

3 如果資源正在執行中，必須先將其停止。以資源的名稱取代 `RSC`。

```
crm resource stop RSC
```

例如，如果要停止 `DLM` 資源，請在 `dlm-clvm` 資源群組中，以 `dlm` 取代 `RSC`。

4 刪除資源本身：

```
crm configure delete ID
```

6.4.3 移除叢集資源

若要移除叢集資源，您需要相應的識別碼。請執行下列步驟：

1 開啟外圍程序，切換為 `root` 身分。

2 執行下列指令以取得資源清單：

```
crm(live)# resource status
```

例如，輸出可能如下所示 (`myIP` 為資源的相應識別碼)：

```
myIP      (ocf::IPaddr:heartbeat) ...
```

3 刪除具有相應識別碼的資源 (此操作還隱含 `commit` 動作)：

```
crm(live)# configure delete YOUR_ID
```

4 提交變更：

```
crm(live)# configure commit
```

6.4.4 移轉叢集資源

雖然資源設定為在遇到硬體或軟體故障時自動容錯移轉 (或移轉) 至叢集的其他節點，您也可以使用 Pacemaker GUI 或指令行將資源手動移轉至叢集中的另一個節點。

1 開啟外圍程序，切換為 `root` 身分。

2 輸入 `crm` 開啟內部外圍程序。

3 若要將名為 `ipaddress1` 的資源移轉至名為 `node2` 的叢集節點，請輸入以下指令：

```
crm(live)# resource
crm(live)resource# migrate ipaddress1 node2
```


使用 Web 介面管理叢集資源

除 `crm` 指令行工具和 Pacemaker GUI 之外，High Availability Extension 還提供了用於管理任務的網路使用者介面 - HA Web Konsole。利用它，您在非 Linux 機器上也可以監控並管理 Linux 叢集。此外，如果您的系統不提供或不允許使用圖形使用者介面，它便是一個理想的解決方案。

該 Web 介面包含在 `hawk` 套件中。該套件必須安裝在您要使用 HA Web Konsole 連接的所有叢集節點上。而在要使用 HA Web Konsole 來存取叢集節點的機器上，您只需要一個啟用了 JavaScript 和 Cookie 的 (圖形) 網頁瀏覽器，就可以建立連線。

注意：使用者認證

若要從 HA Web Konsole 登入叢集，相關使用者必須是群組 `haclient` 的成員。安裝會建立名為 `hacluster` 的 Linux 使用者，其為群組 `haclient` 的成員。

在使用 HA Web Konsole 之前，先為 `hacluster` 使用者設定密碼，或建立一個屬於群組 `haclient` 的新使用者。

在您要使用 HA Web Konsole 連接的所有節點上執行此操作。

7.1 啟動 HA Web Konsole 及登入

過程 7.1 啟動 HA Web Konsole

若要使用 HA Web Konsole，必須在要使用該 Web 介面連接的節點上啟動相應的 Web 服務。如果要進行通訊，請使用標準 HTTP 通訊協定和連接埠 7630。

1 在要連接的節點上，開啟外圍程序並以 root 身分登入。

2 輸入以下指令檢查服務的狀態

```
rchawk status
```

3 如果服務沒有執行，則使用以下指令將其啟動

```
rchawk start
```

如果要在開機時自動啟動 HA Web Konsole，請執行以下指令：

```
chkconfig hawk on
```

4 在任一機器上，啟動網頁瀏覽器並確定已啟用 JavaScript 和 Cookie。

5 將機器指向任一叢集節點的 IP 位址或主機名稱，或已設定的任何 IPaddr(2) 資源的位址：

```
https://IPaddress:7630/main/status
```

注意：證書警告

首次嘗試存取該 URL 時，您可能會收到一則證書警告，具體視瀏覽器和瀏覽器選項而定。這是因為 HA Web Konsole 使用自行簽署的證書，該證書預設不被信任。

若仍要繼續，您可以在瀏覽器中新增一個例外以略過此警告。若要提前避免出現此警告，也可用官方證書管理中心簽署的證書取代自行簽署的證書。如需如何執行此操作的相關資訊，請參閱取代自行簽署的證書 [第108頁]。

- 6 在 HA Web Konsole 登入畫面中，輸入 `hacluster` 使用者 (或任何其他屬於群組 `haclient` 的使用者) 的「使用者名稱」和「密碼」，然後按一下「登入」。

隨即會出現「叢集狀態」畫面，顯示叢集節點和資源的狀態，類似於 `crm_mon` 的輸出。

7.2 使用 HA Web Konsole

登入後，HA Web Konsole 會顯示最重要的全域叢集參數及叢集節點和資源的狀態。以下色彩代碼可用於顯示狀態：

- 綠色：正常。例如，資源正在執行或節點正在線上。
- 紅色：錯誤，異常。例如，資源已失敗或節點未正常關閉。
- 黃色：轉換中。例如，節點目前即將關閉。
- 灰色：尚未執行，但叢集需要其執行。例如，管理員已停止或轉為待機模式的節點。另外，離線的節點也會顯示為灰色 (如果這些節點是正常關閉)。

圖形 7.1 HA Web Konsole — 叢集狀態

Cluster Status User: hacluster Log Out

Failed op: node hex-14 resource ctdb:1: call-id=46 operation=monitor rc-code=7

▽ 2 nodes configured

- hex-13: online
- hex-14: online

▽ 7 resources configured

- ▶ Clone Set: c-ocfs2-3
- ▽ Clone Set: ctdb-clone
 - ctdb:0: Started: hex-13
 - ctdb:1: Stopped
- ▶ Clone Set: dlm-clone
- ▶ Clone Set: o2cb-clone
- fencing-sbd: Started: hex-13
- ▶ Group: ga
- ▶ Clone Set: cg

Stack: openais
Version: 1.1.0-46679a8feec7
Current DC: hex-13
Stickiness: 1
STONITH: Disabled
Cluster is: Symmetric
No Quorum: stop

Copyright © 2009-2010 Novell, Inc. Host: hex-13

按一下「節點」和「資源」群組中的箭頭符號可展開和折疊樹狀目錄檢視。

如果資源已失敗，畫面頂部會以紅色顯示一則包含詳細資料的失敗訊息。

按一下節點或資源右側的扳手圖示可存取用於執行某些動作的快顯功能表，例如啟動、停止或清理資源 (將節點轉為線上或待機模式，或圍籬區隔節點)。

HA Web Konsole 目前只允許執行基本的操作員任務，以後將加入更多功能，例如設定資源與節點的功能。

7.3 疑難排解

HA Web Konsole 記錄檔案

HA Web Konsole 記錄檔案位於 `/srv/www/hawk/log` 中。如果您出於某些原因根本無法存取 HA Web Konsole，可以檢查這些記錄檔案來獲取幫助。

如果您在使用 HA Web Konsole 啟動或停止資源時出現問題，請檢查 Pacemaker 寫入 `/var/log/messages` (預設) 的記錄檔案。

驗證失敗

如果您無法使用新增至群組 `haclient` 的新使用者登入 HA Web Konsole (或 HA Web Konsole 允許此使用者登入前發生延遲)，請使用 `rcnsd stop` 停止 `rcnsd` 精靈，然後再試一次。

取代自行簽署的證書

若要在首次啟動 HA Web Konsole 時避免出現關於自行簽署的證書的警告，請以您自己的證書或官方證書管理中心 (CA) 簽署的證書取代自動建立的證書。

該證書儲存在 `/etc/lighttpd/certs/hawk-combined.pem` 中，其中包含金鑰和證書。在建立或收到新的金鑰和證書之後，執行以下指令將它們進行組合：

```
cat keyfile certificationfile > /etc/lighttpd/certs/hawk-combined.pem
```

變更許可權以使檔案只可由 `root` 存取：

```
chown root.root /etc/lighttpd/certs/hawk-combined.pem
chmod 600 /etc/lighttpd/certs/hawk-combined.pem
```


新增或修改資源代辦

需由叢集管理的所有任務都必須當成資源使用。其中，有兩個主要群組需要注意，即資源代辦和 STONITH 代辦。對於這兩種類別，您都可以新增自己的代辦，以延伸叢集的功能來滿足自己的需要。

8.1 STONITH 代辦

叢集有時會偵測到其中一個節點的行為不正常，需要移除。這稱為「圍籬區隔」，通常透過 STONITH 資源來執行。所有 STONITH 資源皆存放在每個節點的 `/usr/lib/stonith/plugins` 中。

警告：不支援 SSH 和 STONITH

我們無法得知 SSH 對其他系統問題會做出什麼反應。因此，SSH 和 STONITH 代辦不能用於線上環境。

若要取得目前所有可用 STONITH 設備的清單 (從軟體角度)，請使用 `stonith -L` 指令。

到目前為止，尚沒有關於撰寫 STONITH 代辦的文件。若要撰寫新的 STONITH 代辦，請參閱 `heartbeat-common` 套件原始碼中提供的範例。

8.2 撰寫 OCF 資源代辦

所有 OCF 資源代辦 (RA) 皆存放於 `/usr/lib/ocf/resource.d/` 中；如需詳細資訊，請參閱第 4.2.2 節「受支援的資源代辦類別」[第36頁]。每個資源代辦必須支援以下操作以便您對其進行控制：

`start`
啟動或啟用資源

`stop`
停止或停用資源

`status`
傳回資源的狀態

`monitor`
與 `status` 指令類似，但它還會檢查未預期狀態

`validate`
驗證資源的組態

`meta-data`
以 XML 格式傳回資源代辦的相關資訊

建立 OCF RA 的一般程序如下所示：

- 1 將 `/usr/lib/ocf/resource.d/pacemaker/Dummy` 檔案做為樣板載入。
- 2 為每個新的資源代辦新建子目錄，以避免命名衝突。例如，若您擁有資源群組 `kitchen` (內含資源 `coffee_machine`)，請將此資源新增至 `/usr/lib/ocf/resource.d/kitchen/` 目錄。若要存取此資源代辦，請執行 `crm` 指令：

```
configure
primitive coffee_1 ocf:coffee_machine:kitchen ...
```

- 3 執行不同的外圍程序函數，並以不同的名稱儲存檔案。

如需關於撰寫 OCF 資源代辦的更多詳細資料，請造訪 http://linux-ha.org/wiki/Resource_Agents。有關幾個概念的特殊資訊，請參閱第 1 章「產品綜覽」[第3頁]。

8.3 OCF 傳回代碼與失敗復原

根據 OCF 規格，對於動作必須返回的離開碼有著嚴格的定義。叢集會始終根據預期結果檢查傳回代碼。如果結果不符合預期值，則該作業將被視為失敗，並啟動復原動作。失敗復原有三種類型：

表格 8.1 失敗復原類型

復原類型	描述	叢集採取的動作
軟式	發生暫時錯誤。	重新啟動資源或將其移到新的位置。
硬式	發生非暫時錯誤。該錯誤可能與目前節點有關。	將資源移到別處並阻止其在目前節點上被重試。
嚴重錯誤	發生對於所有叢集節點均相同的非暫時錯誤。這表示指定了錯誤的組態。	停止資源並阻止其在任何叢集節點上啟動。

假設一個動作被視為已失敗，下表概述了不同的 OCF 傳回代碼以及接收到相應的錯誤碼時將啟動的叢集復原類型。

表格 8.2 OCF 傳回代碼

OCF 傳回代碼	OCF 別名	描述	復原類型
0	OCF_SUCCESS	成功。指令成功完成。這是所有 start、stop、promote 和 demote 指令的預期結果。	軟式
1	OCF_ERR_GENERIC	一般「發生問題」錯誤碼。	軟式

OCF 傳回代碼	OCF 別名	描述	復原類型
2	OCF_ERR_ARGS	此機器上的資源組態無效 (例如，它參考節點上找不到的位置/工具)。	硬式
3	OCF_ERR_UNIMPLEMENTED	未執行所需動作。	硬式
4	OCF_ERR_PERM	資源代辦沒有足夠的權限來完成任務。	硬式
5	OCF_ERR_INSTALLED	此機器上未安裝資源所需的工具。	硬式
6	OCF_ERR_CONFIGURED	資源的組態無效 (例如，缺少必需的參數)。	嚴重錯誤
7	OCF_NOT_RUNNING	資源未執行。叢集不會嘗試停止對任何動作返回此代碼的資源。 此 OCF 傳回代碼可能需要也可能不需要資源復原 — 具體取決於預期的資源狀態。如果不是預期情況，則進行軟式復原。	無
8	OCF_RUNNING_MASTER	資源正以主要模式執行。	軟式
9	OCF_FAILED_MASTER	資源在主要模式下執行，但已失敗。資源將被降級、停止，然後再次啟動 (還可能升級)。	軟式
其他	無	自定錯誤碼。	軟式

圍籬區隔與 STONITH

在 HA (High Availability) 的電腦叢集中，圍籬區隔是一個極其重要的概念。叢集有時會偵測到其中一個節點行為異常，需要將其移除。這稱為「圍籬區隔」，通常透過 STONITH 資源來執行。可將圍籬區隔定義為讓 HA 叢集處於已知狀態一種方法。

叢集中的每個資源都附有說明。例如：「資源 r1 已在節點 1 上啟動」。在 HA 叢集中，這樣的狀態隱含「資源 r1 已在除節點 1 之外的所有節點上停止」的含義，因為 HA 叢集必須確保每個資源最多只能在一個節點上啟動。每個節點都必須報告資源發生的每一項變更。因此，叢集狀態是資源狀態和節點狀態的集合。

如果無法明確確定某些節點或資源的狀態 (不論出於何種原因)，就會出現圍籬區隔。即使叢集未注意到指定節點上發生了狀況，圍籬區隔也可以確保該節點不會執行任何重要的資源。

9.1 圍籬區隔的類別

有兩類圍籬區隔：資源層級圍籬區隔和節點層級圍籬區隔。後者為本章的主要主題。

資源層級圍籬區隔

如果使用資源層級圍籬區隔，叢集可以確保節點無法存取一或多個資源。一個典型範例為 SAN，在此範例中，圍籬區隔作業會變更 SAN 交換器上的規則，以拒絕從節點進行存取。

資源層級圍籬區隔可利用要保護之資源所依賴的一般資源來實現。只需拒絕在此節點上啟動此資源，所依賴的資源亦不會在相同節點上執行。

節點層級圍籬區隔

節點層級圍籬區隔可確保節點不會執行任何資源。這通常是以一種極其簡單但也突然的方式實現，即使用電源交換器重設節點。當節點沒有回應時，必須執行此作業。

9.2 節點層級圍籬區隔

在 SUSE® Linux Enterprise High Availability Extension 中，圍籬區隔實作為 STONITH (Shoot The Other Node in the Head)。它提供節點層級圍籬區隔。High Availability Extension 包含 `stonith` 指令行工具，此為一個可擴充介面，用於從遠端關閉叢集中的節點。如需可用選項的綜覽，請執行 `stonith --help`，如需詳細資訊，請參閱 `stonith` 的 `man` 頁面。

9.2.1 STONITH 設備

若要使用節點層級圍籬區隔，首先需要擁有一個圍籬區隔設備。若要取得 High Availability Extension 支援的 STONITH 設備清單，請在任何節點上以 `root` 身分執行以下指令：

```
stonith -L
```

可將 STONITH 設備分成以下類別：

配電裝置 (PDU)

在管理重要網路、伺服器和資料中心設備的電源容量和功能方面，配電裝置扮演著至關重要的角色。它們可以提供已連接設備的遠端負載監控功能，並可進行個別插座電源控制以實現遠端電能回收。

不斷電供電系統 (UPS)

市電電源中斷時，透過其他來源供電的穩定供電系統可為連接的設備提供應急電能。

刀鋒電源控制設備

如果您在一組刀鋒上執行叢集，則刀鋒機箱中的電源控制設備是圍籬區隔唯一的候選設備。當然，此設備必須能夠管理單一刀鋒電腦。

無人職守設備

無人職守設備 (IBM RSA、HP iLO、Dell DRAC) 越來越受歡迎，將來它們甚至會成為現貨電腦的標準設備。不過，它們略遜於 UPS 設備，因為它們與其主機 (叢集節點) 共享電源。如果節點仍保持斷電狀態，讓設備進行控制亦無用。此時，CRM 仍會不斷嘗試圍籬區隔該節點，而所有其他資源作業都要等待該圍籬區隔/STONITH 作業完成。

測試設備

測試設備專用於測試用途。在硬體上，對測試設備的要求通常比較寬松。一旦叢集進入實際生產環境，它們必須由真正的圍籬區隔設備取代。

是否選擇 STONITH 設備主要取決於您的預算及所使用的硬體類型。

9.2.2 STONITH 實作

SUSE® Linux Enterprise High Availability Extension 的 STONITH 實作由兩個元件組成：

stonithd

stonithd 是一個可由本地程序或透過網路進行存取的精靈。它接受圍籬區隔作業的相應指令，包括重設、關機和開機。它還可以檢查圍籬區隔設備的狀態。

stonithd 精靈在 CRM HA 叢集中的每個節點上執行。在 DC 節點上執行的 stonithd 例項接收來自 CRM 的圍籬區隔要求。須由此程式及其他 stonithd 程式決定是否執行所需的圍籬區隔作業。

STONITH 外掛程式

每個受支援的圍籬區隔設備都有一個可控制該設備的 STONITH 外掛程式。STONITH 外掛程式是圍籬區隔設備的介面。所有 STONITH 外掛程式皆存放在每個節點上的 `/usr/lib/stonith/plugins` 中。所有 STONITH 外掛程式對於 stonithd 而言看起來都是一樣的，但在反映圍籬區隔設備性質的其他方面則大不相同。

某些外掛程式支援多個設備。典型範例為 `ipmilan` (或 `external/ipmi`)，它實作 IPMI 通訊協定並可控制支援此通訊協定的任何設備。

9.3 STONITH 組態

若要設定圍籬區隔，您需要設定一或多個 STONITH 資源 — stonithd 精靈不需要任何組態設定。所有組態皆儲存在 CIB 中。STONITH 資源即類別為 stonith 的資源 (請參閱第 4.2.2 節「受支援的資源代辦類別」[第36頁])。STONITH 資源是 STONITH 外掛程式在 CIB 中的表示。除圍籬區隔作業之外，還可對 STONITH 資源執行啟動、停止和監控作業，就如同任何其他資源一樣。在這種情況下，啟動和停止 STONITH 資源就表示啟用和停用 STONITH。因此，啟動和停止只是管理作業，不能轉換為圍籬區隔設備本身的任何作業。不過，監控會轉換為設備狀態。

可以對 STONITH 資源進行設定，就如同任何其他資源一樣。如需有關設定資源的詳細資訊，請參閱第 5.3.2 節「建立 STONITH 資源」[第61頁] 或第 6.3.3 節「建立 STONITH 資源」[第93頁]。

參數 (屬性) 清單取決於各 STONITH 類型。若要檢視特定設備的參數清單，請使用 stonith 指令：

```
stonith -t stonith-device-type -n
```

例如，若要檢視 ibmhmc 設備類型的參數，請輸入以下指令：

```
stonith -t ibmhmc -n
```

若要取得設備的簡短說明文字，請使用 -h 選項：

```
stonith -t stonith-device-type -h
```

9.3.1 範例 STONITH 資源組態

下面提供了使用 crm 指令行工具的語法撰寫的一些範例組態。若要套用它們，請將範例放入文字檔中 (例如，sample.txt)，然後執行以下指令：

```
crm < sample.txt
```

如需有關使用 crm 指令行工具設定資源的詳細資訊，請參閱第 6 章「設定和管理叢集資源 (指令行)」[第83頁]。

警告：測試組態

下面提供的一些範例僅用於演示和測試用途。切勿在現實叢集案例中使用任何測試組態範例。

範例 9.1 測試組態

```
configure
primitive st-null stonith:null \
params hostlist="node1 node2"
clone fencing st-null
commit
```

範例 9.2 測試組態

代用組態：

```
configure
primitive st-node1 stonith:null \
params hostlist="node1"
primitive st-node2 stonith:null \
params hostlist="node2"
location l-st-node1 st-node1 -inf: node1
location l-st-node2 st-node2 -inf: node2
commit
```

如果關注叢集軟體，則此組態範例極其適用。與現實組態的唯一區別在於不發生任何圍籬區隔作業。

範例 9.3 測試組態

下面的 `external/ssh` 組態是一個更貼近實際的範例 (但仍然只用於測試)：

```
configure
primitive st-ssh stonith:external/ssh \
params hostlist="node1 node2"
clone fencing st-ssh
commit
```

此組態也可重設節點。該組態與第一個組態非常相似，它具備空 STONITH 設備的功能。在此範例中，將使用複製。它們是 CRM/Pacemaker 功能。複製主要是一種捷徑，因此不需要定義 `n` 個名稱不同的相同資源，只需一個複製的資源就已足夠。複製最多見的無疑是與 STONITH 資源一起使用，但前提是所有節點都可存取 STONITH 設備。

範例 9.4 IBM RSA 無人職守設備組態

實際設備組態並沒有明顯差異，雖然某些設備可能需要更多屬性。可按如下所示設定 IBM RSA 無人職守設備：

```
configure
primitive st-ibmrsa-1 stonith:external/ibmrsa-telnet \
params nodename=node1 ipaddr=192.168.0.101 \
userid=USERID passwd=PASSWORD
primitive st-ibmrsa-2 stonith:external/ibmrsa-telnet \
params nodename=node2 ipaddr=192.168.0.102 \
userid=USERID passwd=PASSWORD
location l-st-node1 st-ibmrsa-1 -inf: node1
location l-st-node2 st-ibmrsa-2 -inf: node2
commit
```

由於 STONITH 作業失敗的可能性始終存在，因此範例中使用了位置限制。所以，在做為執行者的節點上的 STONITH 作業也不可靠。如果重設節點，它就不能傳送關於圍籬區隔作業結果的通知。想傳送通知的唯一方法是，假設作業即將成功，然後提前傳送通知。但如果作業失敗，就會出現問題。因此，stonithd 拒絕停止其主機。

範例 9.5 UPS 圍籬區隔設備組態

UPS 類型圍籬區隔設備的組態類似於上述範例。詳細情形留給讀者在練習中自行探究。所有 UPS 設備皆使用相同技術進行圍籬區隔，但存取設備本身的方法各不相同。舊式 UPS 設備通常只有一個序列埠，大多數情況下會使用特殊的序列纜線以 1200 鮑率進行連接。許多新型 UPS 設備仍然有一個序列埠，但它們通常還使用 USB 或乙太網路介面。可以使用的連線類型取決於外掛程式所支援的類型。

例如，使用 `stonith -t stonith 設備類型 -n` 指令比較 `apcmaster` 與 `apcsmart` 設備：

```
stonith -t apcmaster -h
```

傳回下列資訊：

```
STONITH Device: apcmaster - APC MasterSwitch (via telnet)
NOTE: The APC MasterSwitch accepts only one (telnet)
connection/session a time. When one session is active,
subsequent attempts to connect to the MasterSwitch will fail.
For more information see http://www.apc.com/
List of valid parameter names for apcmaster STONITH device:
ipaddr
login
password
```

使用

```
stonith -t apcsmart -h
```

您將取得下列輸出：

```
STONITH Device: apcsmart - APC Smart UPS
(via serial port - NOT USB!).
Works with higher-end APC UPSes, like
Back-UPS Pro, Smart-UPS, Matrix-UPS, etc.
(Smart-UPS may have to be >= Smart-UPS 700?).
See http://www.networkupstools.org/protocols/apcsmart.html
for protocol compatibility details.
For more information see http://www.apc.com/
List of valid parameter names for apcsmart STONITH device:
ttydev
hostlist
```

第一個外掛程式支援具有網路埠和 Telnet 通訊協定的 APC UPS。第二個外掛程式使用 APC SMART 通訊協定 (透過眾多 APC UPS 產品線皆支援的序列線)。

9.3.2 限制與複製

在第 9.3.1 節「範例 STONITH 資源組態」[第116頁] 中，您已經瞭解到設定 STONITH 資源有幾種方法：使用限制、複製或兩者。選擇使用何種建構進行組態設定取決於幾項因素，包括圍籬區隔設備的性質、設備管理的主機數目、叢集節點數目，或個人優先設定。

簡言之：如果可以放心地將複製與組態搭配使用，且如果它們確實可以減少組態設定工作，則使用複製的 STONITH 資源。

9.4 監控圍籬區隔設備

如同任何其他資源一樣，STONITH 類別代辦也支援用於檢查狀態的監控作業。

注意：監控 STONITH 資源

強烈建議您監控 STONITH 資源。應定期且謹慎地對它們進行監控。

圍籬區隔設備是 HA 叢集必不可少的一部分，但需要用到它們的情況卻是越少越好。眾所周知，電源管理設備在通訊端尤為脆弱。如果廣播流量過多，某些設備就會放棄處理。有些設備在每分鐘達到十個左右連線時便無法處理。如果兩個用戶端同時嘗試連接，有些設備還會發生混淆。大部分設備都不能同時處理多個工作階段。

在大多數情況下，只需每隔幾個小時檢查一次圍籬區隔設備即可。在這幾個小時內需要執行圍籬區隔作業以及電源交換器失效的可能性通常很低。

如需如何設定監控操作的詳細資訊，請參閱過程 5.3「新增或修改中繼屬性與例項屬性」[第59頁] (適用於 GUI 方法)，或參閱第 6.3.8 節「設定資源監控」[第99頁] (適用於指令行方法)。

9.5 特殊圍籬區隔設備

除了處理實際 STONITH 設備的外掛程式之外，某些 STONITH 外掛程式還需要其他說明。

警告：僅供測試之用

下面提到的一些 STONITH 外掛程式僅用於演示和測試之目的。請不要在實際情況中使用以下任一設備，否則可能會導致資料毀損以及無法預期的後果：

- `external/ssh`
- `ssh`
- `null`

`external/kdumpcheck`

此外掛程式可用於檢查節點上是否正在執行核心傾印。如果正在執行，此外掛程式將傳回 `true`，表示節點已圍籬區隔(此時該節點無法執行任何資源)。這樣可避免圍籬區隔已關閉但正在傾印的節點，省下相應的時間。外掛程式必須與另一個實際的 STONITH 設備搭配使用。如需詳細資料，請參閱 `/usr/share/doc/packages/cluster-glue/README_kdumpcheck.txt`。

`external/sbd`

這是一個自我圍籬區隔設備。它會對可插入到共享磁碟中的所謂「毒藥丸」做出反應。在共享儲存區連線中斷時，它還會讓節點停止作業。若要瞭解如何使用此 STONITH 代辦執行基於儲存區的圍籬區隔，請參閱第 15 章「儲存保護」[第169頁]。也可參閱 http://www.linux-ha.org/wiki/SBD_Fencing 以取得詳細資料。

`external/ssh`

另一個基於軟體的「圍籬區隔」機制。節點必須能夠以 `root` 身分登入彼此，且無需密碼。該機制需要一個參數 `hostlist`，指定它的目標節點。由於無法重設實際失敗的節點，因此該機制不能用於實際的叢集 — 只能用於測試和演示。如果將其用於共享儲存區，可能會導致資料毀損。

`meatware`

`meatware` 需要人員協助才能操作。呼叫 `meatware` 時，它會記錄 `CRIT` 嚴重程度訊息，該訊息會在節點的主控台上顯示。然後，操作員會確認該節點已關閉，並發出 `meatclient(8)` 指令。這將告訴 `meatware`，它可以通知叢集可以將節點視為已關閉。如需詳細資訊，請參閱 `/usr/share/doc/packages/cluster-glue/README.meatware`。

`null`

這是一個虛構的設備，用於各種測試案例。它總是表現為且聲稱自己關閉了一個節點，但其實不會執行任何動作。除非您完全瞭解所執行的操作，否則切勿使用它。

`suicide`

這是一個僅限軟體的設備，它可以使用 `reboot` 指令重新開機執行它所在的節點。這需要由節點的作業系統執行動作，在某些情況下可能會失敗。因此，應儘可能避免使用此設備。不過，在一個節點的叢集上使用還是安全的。

`suicide` 和 `null` 是「do not shoot my host」(不要關閉我的主機) 規則唯一的例外。

9.6 如需更多資訊

`/usr/share/doc/packages/cluster-glue`

在已安裝的系統中，此目錄存放眾多 STONITH 外掛程式和設備的讀我檔案。

<http://www.linux-ha.org/wiki/STONITH>

有關 STONITH 的資訊位於 The High Availability Linux Project 的首頁上。

http://www.clusterlabs.org/doc/crm_fencing.html

有關圍籬區隔的資訊位於 Pacemaker Project 的首頁上。

http://www.clusterlabs.org/doc/en-US/Pacemaker/1.0/html/Pacemaker_Explained

說明用於設定 Pacemaker 的概念。包含全面詳盡的資訊，以供參考。

http://techthoughts.typepad.com/managing_computers/2007/10/split-brain-quo.html

說明 HA 叢集中的電腦分裂、最低節點數和圍籬區隔之概念的文章。

Linux Virtual Server 的負載平衡

10

Linux Virtual Server (LVS) 的目標是提供一個基本架構，將網路連線導向至共享工作負載的多部伺服器。Linux Virtual Server 是一組伺服器叢集 (一或多個負載平衡器與幾部執行服務的實際伺服器)，對外部用戶端而言則是一個大型的高速伺服器。這種表面上的單個伺服器稱為*虛擬伺服器*。Linux Virtual Server 可用於構建擴充性強、可用性高的網路服務，例如 Web、快取、郵件、FTP、媒體和 VoIP 服務。

實際的伺服器與負載平衡器之間可透過高速 LAN 或地理位置分散的 WAN 來連接。負載平衡器可將請求發送到不同的伺服器。它們可以讓叢集的多個平行服務顯示為單一 IP 位址 (虛擬 IP 位址或 VIP) 上的一個虛擬服務。發送請求可以使用 IP 負載平衡技術或應用程式層級的負載平衡技術。以透明方式在叢集中新增或移除節點可以實現系統的延展性。透過偵測節點或精靈故障，然後重新正確地設定系統可以提供高可用性。

10.1 概念綜覽

以下幾節概要介紹了主要的 LVS 元件和概念。

10.1.1 導向器

LVS 的主要元件是 ip_vs (或 IPVS) 核心代碼。它會在 Linux 核心 (第 4 層交換) 內執行輸送層負載平衡。執行包含 IPVS 代碼之 Linux 核心的節點稱為*導向器*。在導向器上執行的 IPVS 代碼是 LVS 的必要特性。

當用戶端連接到導向器時，內送請求會在所有叢集節點上達到負載平衡：導向器會使用可使 LVS 工作的一組修改後路由規則，將封包轉遞到實際的伺服器。也即是，導向器不會發起或終止連線，也不會傳送確認通知等等。導向器類似特殊的路由器，它會將終端使用者的封包轉遞到實際的伺服器(執行處理請求之應用程式的主機)。

依預設，此核心不會安裝 IPVS 模組。IPVS 核心模組包含在 `cluster-network-kmp-default` 套件內。

10.1.2 使用者空間控制器和精靈

`ldirectord` 是一個使用者空間精靈，可在負載平衡虛擬伺服器的 LVS 叢集中管理 Linux Virtual Server 和監控實際的伺服器。組態檔案 `/etc/ha.d/ldirectord.cf` 指定虛擬服務及其關聯的實際伺服器，並告知 `ldirectord` 如何將伺服器設定為 LVS 重新導向器。啟始化精靈時，會為叢集建立虛擬服務。

`ldirectord` 精靈透過不時發出已知 URL 請求並檢查回應的方式，監控實際伺服器的狀態。如果實際伺服器出現故障，便會從負載平衡器的可用伺服器清單中移除。當服務監控器偵測到該停止的伺服器已恢復正常並可重新工作時，會重新將該伺服器新增到可用伺服器清單中。如果所有實際伺服器都當機，可以指定 Web 服務重新導向到的錯誤回復伺服器。錯誤回復伺服器通常是本地主機，它會顯示一個緊急頁面，告知 Web 服務暫時不可用。

10.1.3 封包轉遞

導向器將用戶端封包傳送至實際伺服器的方式有三種：

網址轉譯 (NAT)

內送請求會送達虛擬的 IP 位址，然後透過將目的地 IP 位址和連接埠變更為所選實際伺服器的 IP 位址和連接埠的方式，將此請求轉遞至實際的伺服器。實際伺服器會將回應傳送至負載平衡器，再由負載平衡器變更目的地 IP 位址，並將回應轉遞給用戶端，因此終端使用者會從預期的來源收到回覆。由於所有流量都會通過負載平衡器，因此它往往成為叢集的瓶頸。

IP 通道封裝 (IP-IP 封裝)

IP 通道封裝能將定址到某 IP 位址的封包重新導向至另一個位址(可能位於不同的網路)。LVS 透過 IP 通道將請求傳送至實際的伺服器(重新導向至不

同的 IP 位址)，然後實際伺服器使用自己的路由表直接回覆用戶端。叢集成員可以位於不同的子網路。

直接路由

直接將終端使用者的封包轉遞到實際的伺服器。IP 封包並未修改，因此必須將實際伺服器設定為接受送往虛擬伺服器 IP 位址的流量。實際伺服器的回應將直接傳送至用戶端。實際伺服器和負載平衡器必須位於相同的實體網路節區。

10.1.4 排程演算法

使用哪一部實際伺服器來處理用戶端請求的新連線，是由不同的演算法決定的。這些演算法以模組的形式提供，可根據特定需求進行調整。如需可用模組的綜覽，請參閱 `ipvsadm(8) man` 頁面。從用戶端接收到連線請求後，導向器會依據排程為該用戶端指定實際的伺服器。排程器是 IPVS 核心代碼的一部分，用於決定將獲取下一個新連線的實際伺服器。

10.2 使用 YaST 設定 IP 負載平衡

您可以使用 YaST `iplb` 模組設定基於核心的 IP 負載平衡。這是 `ldirectord` 的前端。

若要存取 IP 負載平衡對話方塊，請以 `root` 身分啟動 YaST，然後選取「*High Availability*」>「IP 負載平衡」。或者在指令行上使用 `yast2 iplb` 以 `root` 身分啟動 YaST 叢集模組。

然後 YaST 會將其組態寫入 `/etc/ha.d/ldirectord.cf`。YaST 模組中可用的索引標籤與 `/etc/ha.d/ldirectord.cf` 組態檔案的結構對應，用於定義全域選項和虛擬服務的選項。

如需組態以及所產生之負載平衡器與實際伺服器間程序的範例，請參閱範例 10.1「簡單的 `ldirectord` 組態」[第 130 頁]。

注意：全域參數和虛擬伺服器參數

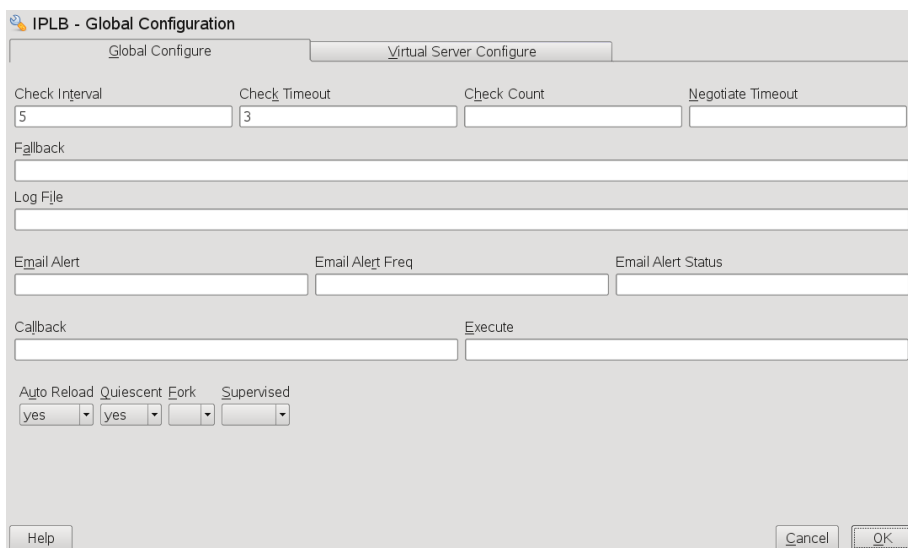
如果同時在虛擬伺服器區段和全域區段指定了某個參數，則虛擬伺服器區段中定義的值將覆寫全域區段中定義的值。

過程 10.1 設定全域參數

以下程序介紹如何設定最重要的全域參數。如需個別參數和此處未涉及之參數的詳細資料，請按一下「說明」或參閱 `ldirectord man` 頁面。

- 1 使用「檢查間隔時間」定義 `ldirectord` 連接各個實際伺服器以檢查它們是否連線的時間間隔。
- 2 使用「檢查逾時」設定實際伺服器應在上次檢查後的多少時間內給於回應。
- 3 使用「檢查次數」可以定義判定檢查失敗前，`ldirectord` 嘗試向實際伺服器發出請求的次數。
- 4 使用「協議逾時」可以定義協議檢查的逾時時間 (秒)。
- 5 在「錯誤回復」中，輸入當所有實際伺服器都當機時，Web 服務重新導向到之 Web 伺服器的主機名稱或 IP 位址。
- 6 如果要讓記錄使用不同的路徑，則在「記錄檔案」中指定記錄的路徑。依預設，`ldirectord` 會將其記錄寫入 `/var/log/ldirectord.log`。
- 7 如果希望系統在與實際伺服器的連線狀態發生變更時傳送警示，請在「電子郵件警示」中輸入有效的電子郵件地址。
- 8 使用「電子郵件警示頻率」可以定義當實際伺服器長時間無法存取時，重新傳送電子郵件警示的間隔時間 (秒)。
- 9 在「電子郵件警示狀態」中指定出現哪種伺服器狀態時傳送電子郵件警示。如果要定義多種狀態，可以逗號分隔。
- 10 使用「自動重新載入」定義 `ldirectord` 是否應持續監控組態檔案的修改情況。如果設定為是，便會在變更後自動重新載入組態。
- 11 使用「*Quiescent*」參數定義是否要從核心的 LVS 表格中移除出現故障的實際伺服器。如果設定為「是」，則不會移除出現故障的伺服器。而是將其權值設為 0，表示不接受任何新的連線。已建立的連線將保持，直到逾時。

圖形 10.1 YaST IP 負載平衡 — 全域參數



The image shows the 'IPLB - Global Configuration' window in YaST. It has two tabs: 'Global Configure' (selected) and 'Virtual Server Configure'. The 'Global Configure' tab contains several input fields and dropdown menus for configuring the load balancer. The fields are: 'Check Interval' (set to 5), 'Check Timeout' (set to 3), 'Check Count' (empty), 'Negotiate Timeout' (empty), 'Fallback' (empty), 'Log File' (empty), 'Email Alert' (empty), 'Email Alert Freq' (empty), 'Email Alert Status' (empty), 'Callback' (empty), and 'Execute' (empty). At the bottom, there are four dropdown menus for 'Auto Reload' (set to yes), 'Quiescent' (set to yes), 'Fork' (empty), and 'Supervised' (empty). At the bottom right, there are 'Help', 'Cancel', and 'OK' buttons.

過程 10.2 設定虛擬服務


您可以透過為各項虛擬服務定義一對參數的方式設定一或多個虛擬服務。以下程序介紹如何為虛擬服務設定最重要的全域參數。如需個別參數和此處未涉及之參數的詳細資料，請按一下「說明」或參閱 `ldirectord` **man** 頁面。

- 1 在 YaST `iplb` 模組中，切換至「**虛擬伺服器組態**」索引標籤。
- 2 「**新增**」新的虛擬伺服器，或「**編輯**」現有的虛擬伺服器。此時會出現一個新的對話方塊並顯示可用的選項。
- 3 在「**虛擬伺服器**」中輸入以 LVS 形式存取負載平衡器和實際伺服器時所使用的共享虛擬 IP 位址及連接埠。也可以指定主機名稱和服務來代替 IP 位址和連接埠名稱。或者，也可以使用防火牆標記。防火牆標記是將任意一組 VIP:連接埠服務集結到一個虛擬服務的一種方式。
- 4 若要指定「**實際伺服器**」，需要輸入伺服器的 IP 位址 (或主機名稱)、連接埠 (或服務名稱) 及轉遞方式。轉遞方式必須是 `gate`、`ipip` 或 `masq`，詳情請參閱第 10.1.3 節「封包轉遞」[第124頁]。

按一下「**新增**」按鈕並輸入各個實際伺服器所需的引數。

- 5 在「**檢查類型**」中，選取測試實際伺服器是否仍在工作時執行的檢查類型。例如，若要傳送請求並檢查回應中是否包含預期字串，請選取「**協議**」。
- 6 如果「**檢查類型**」已設定為「**協議**」，則還需要定義要監控之服務的類型。從「**服務**」下拉式清單中選取所需的服務類型。
- 7 在「**請求**」中輸入檢查間隔期間向每部實際伺服器所請求之物件的 URI。
- 8 如果要檢查實際伺服器的回應中是否包含某個字串（「**I'm alive**」訊息），請定義要比對的正規表示式。在「**接收**」中輸入正規表示式。如果實際伺服器的回應中包含此表示式，說明此實際伺服器正在運行。
- 9 您還需要根據在步驟6[第128頁]中所選的「**服務**」類型指定其他參數，如「**登入**」、「**密碼**」、「**資料庫**」或「**秘密**」。如需詳細資訊，請參閱 **YaST 說明**或 **ldirectord man** 頁面。
- 10 選取用於負載平衡的「**規劃程式**」。如需可用之規劃程式的相關資訊，請參閱 **ipvsadm(8) man** 頁面。
- 11 選取要使用的「**協定**」。如果虛擬服務指定為 **IP** 位址和連接埠，則必須是 **tcp** 或 **udp** 協定。如果虛擬服務指定為防火牆標記，則必須是 **fwm** 協定。
- 12 根據需要定義其他參數。按一下「**確定**」確認您的組態。**YaST** 會將此組態寫入 **/etc/ha.d/ldirectord.cf**。

圖形 10.2 YaST IP 負載平衡 — 虛擬服務

 **IPLB - Virtual Servers Configuration**

Virtual Server

Real Servers

192.168.0.110:80 gate
192.168.0.120:80 gate

Check Type Service Check Command Check Port

Request Receive Http Method Virtual Host

Login Password Database Name Radius Secret

Persistent Netmask Scheduler Protocol

Check Timeout Negotiate Timeout Check Count Email Alert

Email Alert Freq Email Alert Status Fallback Quiescent

範例 10.1 簡單的 *ldirectord* 組態

圖形 10.1 「YaST IP 負載平衡 — 全域參數」[第127頁] 和圖形 10.2 「YaST IP 負載平衡 — 虛擬服務」[第129頁] 中的顯示的值會產生以下組態(定義於 `/etc/ha.d/ldirectord.cf`)：

```
autoreload = yes ❶
checkinterval = 5 ❷
checktimeout = 3 ❸
quiescent = yes ❹
    virtual = 192.168.0.200:80 ❺
    checktype = negotiate ❻
    fallback = 127.0.0.1:80 ❼
    protocol = tcp ❸
    real = 192.168.0.110:80 gate ❾
    real = 192.168.0.120:80 gate ❾
    receive = "still alive" ❿
    request = "test.html" ⓫
    scheduler = wlc ⓬
    service = http ⓭
```

- ❶ 定義 *ldirectord* 應持續檢查組態檔案的修改情況。
- ❷ *ldirectord* 連接各個實際伺服器以檢查它們是否連線的時間間隔。
- ❸ 自上次檢查後，實際伺服器應做出回應的時間。
- ❹ 指定不要將出現故障的實際伺服器從核心的 LVS 表格中移除，但將其權值設為 0。
- ❺ LVS 的虛擬 IP 位址 (VIP)。LVS 可使用連接埠 80 進行連接。
- ❻ 為測試實際伺服器是否仍在工作而執行的檢查類型。
- ❼ 所有實際伺服器都當機時，將 Web 服務重新導向到的伺服器。
- ❸ 要使用的協定。
- ❾ 兩個已定義的實際伺服器，可使用連接埠 80 連接。封包轉遞方式是 *gate*，表示使用直接路由。
- ❿ 要在實際伺服器之回應字串中比對的正規表示式。
- ⓫ 檢查間隔期間向每部實際伺服器所請求之物件的 URI。
- ⓬ 所選的用於負載平衡的規劃程式。
- ⓭ 要監控的服務類型。

此組態會產生以下程序流：ldirectord 將每隔 5 秒連接一次各個實際伺服器 ❷，並請求 ❸ 和 ❹ 中指定的 192.168.0.110:80/test.html 或 192.168.0.120:80/test.html。如果在上次檢查後的 3 秒內 ❺ 沒有從實際伺服器收到預期的 still alive 字串 ❻，則會將實際伺服器從可用的池中移除。但是，由於設定了 quiescent=yes ❼，因此實際伺服器不會從 LVS 表格移除，但其權值會設為 0，這樣便不會接受與此實際伺服器的新連線。已建立的連線將保持，直到逾時。

10.3 更多設定

除了使用 YaST 進行的 ldirectord 組態設定外，您還需要確定符合以下條件，才能完成 LVS 的設定：

- 已正確設定實際伺服器以提供所需的服務。
- 一或多個負載平衡伺服器必須能夠採用 IP 轉遞方式將流量路由到實際伺服器。實際伺服器的網路組態取決於您選擇的封包轉遞方法。
- 為了防止一或多個負載平衡伺服器成為整個系統的單一故障點，需要設定一或多個負載平衡器的備份。在叢集組態中，設定 ldirectord 的原始資源，這樣在出現硬體故障時，ldirectord 便可容錯移轉到其他伺服器。
- 由於負載平衡器的備份也需要 ldirectord 組態檔案來完成此任務，因此請確定要用做負載平衡器之備份的所有伺服器上，/etc/ha.d/ldirectord.cf 均可用。您可以依照第 3.2.3 節「將組態傳輸至所有節點」[第24頁]中所述，使用 Csync2 同步組態檔案。

10.4 如需更多資訊

如需有關 Linux Virtual Server 的詳細資訊，請參閱專案首頁 <http://www.linuxvirtualserver.org/>。

如需有關 ldirectord 的資訊，請參閱完整的 man 頁面。

網路設備 Bonding

對於許多系統而言，實作的網路連線除了需要符合一般乙太網路設備的標準資料安全性或可用性要求之外，還需要符合其他要求。在這些情況下，數個乙太網路設備可以結集成單一的 bonding 設備。

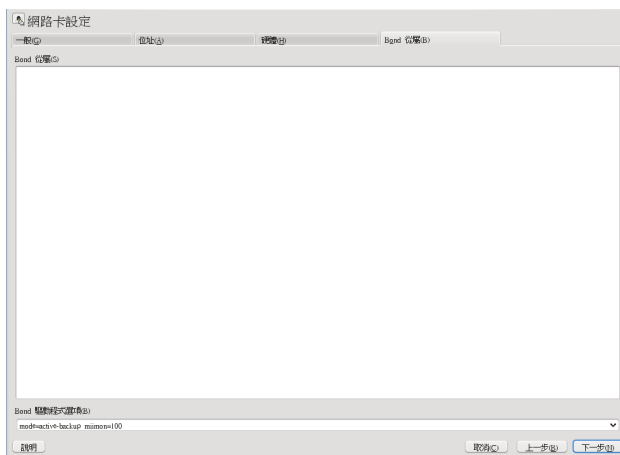
bonding 設備的組態是透過 bonding 模組選項來設定，而其行為由 bonding 設備的模式決定。該模式預設為 `mode=active-backup`，這表示如果使用中的從屬設備失敗，另一個從屬設備將變成使用中狀態。

使用 OpenAIS 時，bonding 設備不受叢集軟體的管理。因此，必須在可能需要存取 bonding 設備的每個叢集節點上設定該設備。

11.1 使用 YaST 設定 Bonding 設備

若要設定 bonding 設備，請執行以下程序：

- 1 以 root 身分啟動 YaST，然後選取「網路設備」>「網路設定」。
- 2 使用「新增」設定新的網路卡，然後將「設備類型」變更為「*Bond*」。按「下一步」繼續。



3 選取為 bonding 設備指定 IP 位址的方法。有三種方法可供您選擇：

- 無 IP 位址
- 動態位址 (透過 DHCP 或 Zeroconf)
- 靜態指定的 IP 位址

請使用適合您環境的方法。如果 OpenAIS 管理虛擬 IP 位址，請選取「靜態指定的 IP 位址」，然後為介面指定一個基本 IP 位址。

4 切換到「*Bond* 從屬」索引標籤。

5 若要選取需加入 Bond 的乙太網路設備，請選取相關「*Bond* 從屬」前面的核取方塊。

6 編輯「*Bond* 驅動程式選項」。可用模式如下：

balance-rr
提供負載平衡和容錯。

active-backup
提供容錯。

`balance-xor`
提供負載平衡和容錯。

`broadcast`
提供容錯

`802.3ad`
提供動態連結聚總 (若連接的交換器支援)。

`balance-tlb`
提供外送流量的負載平衡。

`balance-alb`
提供內送和外送流量的負載平衡 (若所用的網路設備允許修改使用中網路設備的硬體位址)。

- 7 確認參數 `miimon=100` 已新增至「*Bond 驅動程式選項*」。若沒有此參數，就無法定期檢查資料的完整性。
- 8 按「下一步」，然後按一下「確定」離開 YaST，以建立設備。

11.2 如需更多資訊

「*Linux Ethernet Bonding Driver HOWTO*」(Linux 乙太網路 Bonding 驅動程式 HOWTO) 對所有模式以及許多其他選項做了詳細說明，若您安裝了 `kernel-source` 套件，便可在 `/usr/src/linux/Documentation/networking/bonding.txt` 中找到該資訊。

III. 儲存與資料複製

Oracle Cluster File System 2

Oracle Cluster File System 2 (OCFS2) 是一般用途的日誌式檔案系統，已經完全整合到 Linux 2.6 核心及以上版本中。OCFS2 可讓您將應用程式二進位檔案、資料檔案和資料庫儲存於設備上的共享儲存中。叢集中所有節點均同時具有檔案系統的讀取與寫入權限。透過複製資源管理的使用者空間控制精靈可提供與 HA 堆疊的整合，特別是與 OpenAIS/Corosync 和分散式鎖定管理員 (DLM) 的整合。

12.1 特點及優勢

OCFS2 可用於下列儲存解決方案，例如：

- 一般應用程式與工作負載。
- 叢集中的 Xen 影像儲存。Xen 虛擬機器與虛擬伺服器可以儲存到由叢集伺服器掛接的 OCFS2 磁碟區。因此能夠便捷地實現伺服器之間 Xen 虛擬機器的可攜性。
- LAMP (Linux、Apache、MySQL 和 PHP | PERL | Python) 堆疊。

做為一款高效能、對稱、平行的叢集檔案系統，OCFS2 支援下列功能：

- 叢集上的所有節點均可使用應用程式的檔案。使用者只需在叢集上的 OCFS2 磁碟區安裝一次即可。
- 所有節點可直接透過標準檔案系統介面同時讀取及寫入儲存區，讓叢集上執行的應用程式更易於管理。

- 檔案存取透過 DLM 來協調。在多數情況下，DLM 控制都能起到很好的效果，但如果應用程式設計與 DLM 爭奪檔案存取協調權，則擴充性可能就會受到限制。
- 所有後端儲存區均可使用儲存區備份功能。您可輕鬆建立共享應用程式檔案的複本，以利於提供有效的災難復原。

OCFS2 亦提供下列功能：

- 中繼資料快取。
- 中繼資料日誌。
- 跨節點資料檔案一致性。
- 支援高達 4 KB 的多區塊大小 (各磁碟區可具有不同的區塊大小)，磁碟區的最大大小為 16 TB。
- 支援高達 16 個磁簇節點。
- 對資料庫檔案提供非同步且直接 I/O 支援，以加強資料庫效能。

12.2 OCFS2 套件與管理公用程式

OCFS2 核心模組 (ocfs2) 會自動安裝到 SUSE® Linux Enterprise Server 11 SP1 的 High Availability Extension 中。若要使用 OCFS2，請確定叢集中的每個節點均已安裝下列套件：ocfs2-tools 及符合核心的 ocfs2-kmp-* 套件。

ocfs2-tools 套件提供下列管理 OFS2 磁碟區的公用程式。如需有關語法的資訊，請參閱其 man 頁面。

表格 12.1 OCFS2 共用程式

OCFS2 共用程式	描述
debugfs.ocfs2	以偵錯為目的，檢驗 OCFS 檔案系統。
fsck.ocfs2	檢查檔案系統是否有錯誤，並選擇性修復錯誤。

OCFS2 共用程式	描述
mkfs.ocfs2	在設備上建立 OCFS2 檔案系統，通常是共用實體或邏輯磁碟上的分割區。
mounted.ocfs2	偵測並列出叢集系統上的所有 OCFS2 磁碟區。偵測並列出掛接 OCFS2 裝置的系統上之所有節點，或列出所有 OCFS2 裝置。
tunefs.ocfs2	變更 OCFS2 檔案系統參數，包括磁碟區標籤、節點插槽數目、所有節點插槽的日至大小，以及磁碟區大小。

12.3 設定 OCFS2 服務

建立 OCFS2 磁碟區之前，必須先在叢集中將資源 DLM 與 O2CB 設定為服務。OCFS2 會使用使用者空間中執行的 Pacemaker 所提供的叢集成員服務。因此，需要將 DLM 與 O2CB 設定為叢集中每個節點上都存在的複製資源。

過程 12.1 設定 DLM 與 O2CB 資源

下列程序使用 `crm` 外圍程序來設定叢集資源。對叢集中的某個節點，執行下列步驟。或者，也可以使用 `Heartbeat` 設定資源。

1 開啟終端機視窗，並以 `root` 或同等身分登入。

2 若要將 DLM (分散式鎖定管理員) 新增為資源：

2a 啟動 `crm` 外圍程序，並重新建立新的組態：

```
crm
cib new stack-glue
```

2b 建立 DLM 服務，並在叢集中的所有機器上執行該服務：

```
configure
primitive dlm ocf:pacemaker:controld op monitor interval=120s
clone dlm-clone dlm meta globally-unique=false interleave=true
end
```

dlm 複製資源控制分散式鎖定管理員服務，並確保此服務已在叢集中的所有節點上啟動。

2c 先驗證所做的變更，然後將其提交至 CIB：

```
cib diff
configure verify
```

2d 將組態上載至叢集，然後結束外圍程序：

```
cib commit stack-glue
quit
```

3 若要新增 O2CB 組態：

3a 啟動 crm 外圍程序，並重新建立新的組態：

```
crm
cib new oracle-glue
```

3b 確保 O2CB 服務已在叢集的每個節點上啟動：

```
configure
primitive o2cb ocf:ocfs2:o2cb op monitor interval=120s
clone o2cb-clone o2cb meta globally-unique=false interleave=true
```

3c 若要確保 O2CB 服務僅在同樣執行有 dlm 服務副本的節點上啟動，請新增並存限制：

```
colocation o2cb-with-dlm INFINITY: o2cb-clone dlm-clone
order start-o2cb-after-dlm mandatory: dlm-clone o2cb-clone
```

3d 將組態上載至叢集，然後結束外圍程序：

```
cib commit oracle-glue
quit
```

4 若要設定圍籬區隔設備：

4a 啟動 crm 外圍程序，並重新建立新的組態：

```
crm
cib new fencing
```

- 4b** 將 `external/sdb` 設定為圍籬區隔設備 (其 `/dev/sdb2` 用做共享儲存區上的專用分割區)，執行活動訊號與圍籬區隔：

```
configure
primitive sbd_stonith stonith:external/sbd \
meta target-role="Started"op monitor \
interval=15 timeout=15 start-delay=15 \
params sbd_device=/dev/sdb2
```

- 4c** 將組態上載至叢集，然後結束外圍程序：

```
cib commit fencing
quit
```

12.4 建立 OCFS2 磁碟區

按第 12.3 節「設定 OCFS2 服務」[第141頁]中所述將 DLM 和 O2CB 設定為叢集資源後，設定系統使用 OCFS2，並建立 OCFS2 磁碟區。

注意：應用程式與資料檔案使用的 **OCFS2** 磁碟區

一般情況下，建議您將應用程式檔案與資料檔案儲存在不同的 **OCFS2** 磁碟區中。如果您的應用程式磁碟區和資料磁碟區對於掛接有不同的要求，請務必將它們儲存到不同的磁碟區中。

開始之前，請先準備要用於 OCFS2 磁碟區的區塊設備。將裝置留為可用空間。

然後按過程 12.2「建立並格式化 OCFS2 磁碟區」[第145頁]中所述，使用 `mkfs.ocfs2` 建立並格式化 OCFS2 磁碟區。此指令最重要的參數列於表格 12.2「OCFS2 的重要參數」[第143頁]。如需詳細資訊及指令語法，請參閱 `mkfs.ocfs2 man` 頁面。

表格 12.2 *OCFS2 的重要參數*

OCFS2 參數	描述與建議
磁碟區標籤 (-L)	磁碟區的描述性名稱可讓其掛接於不同節點時易於辨識。使用 <code>tunefs.ocfs2</code> 公用程式依需要修改標籤。

OCFS2 參數	描述與建議
叢集大小 (-C)	叢集大小是配置給持有資料的檔案之空間最小單位。如需可用選項及建議，請參閱 <code>mkfs.ocfs2 man</code> 頁面。
節點插槽數目 (-N)	<p>可同時掛接磁碟區的最大節點數目。對於每個節點，OCFS2 會分別為其建立系統檔案，例如日誌。存取磁碟區的節點可以是小 endian 架構 (如 x86、x86-64 和 ia64) 和大 endian 架構 (如 ppc64 和 s390x) 的組合。</p> <p>節點特定的檔案會被視為本機檔案。節點插槽號碼會附加至本機檔案。例如：<code>journal:0000</code> 隸屬於指派至插槽 0 的任一節點。</p> <p>建立磁碟區時，請根據您希望同時掛接磁碟區的節點數量，設定節點插槽的最大磁碟區數目。使用 <code>tunefs.ocfs2</code> 公用程式可以視需要增加節點數。請注意，此值只能增加，不能減少。</p>
區塊大小 (-b)	檔案系統可定址的空間最小單位。請在建立磁碟區時指定區塊大小。如需可用選項及建議，請參閱 <code>mkfs.ocfs2 man</code> 頁面。
開啟/關閉特定功能 (--fs-features)	<p>可提供以逗號分隔的功能標籤清單，<code>mkfs.ocfs2</code> 會根據此清單嘗試建立具有這些功能集的檔案系統。若要開啟功能，請在清單中包含該功能。若要關閉功能，則在名稱前預增 <code>no</code>。</p> <p>如需所有可用標籤的綜覽，請參閱 <code>mkfs.ocfs2 man</code> 頁面。</p>
預定義功能 (--fs-feature-level)	<p>可讓您從一組預定義檔案系統功能中進行選擇。如需可用選項，請參閱 <code>mkfs.ocfs2 man</code> 頁面。</p>

使用 `mkfs.ocfs2` 建立並格式化磁碟區時，如果不指定任何具體功能，則預設會啟用下列功能：`backup-super`、`sparse`、`inline-data`、`unwritten`、`metaecc`、`indexed-dirs` 和 `xattr`。

過程 12.2 建立並格式化 OCFS2 磁碟區

在其中一個叢集節點上執行下列步驟。

- 1 開啟終端機視窗，並以 `root` 身分登入。
- 2 請使用指令 `crm_mon` 檢查叢集是否上線。
- 3 使用 `mkfs.ocfs2` 公用程式建立並格式化磁碟區。如需此指令的語法資訊，請參閱 `mkfs.ocfs2 man` 頁面。

例如，若要在最多支援 16 個叢集節點的 `/dev/sdb1` 上建立新的 OCFS2 檔案系統，請使用下列指令：

```
mkfs.ocfs2 -N 16 /dev/sdb1
```

12.5 掛接 OCFS2 磁碟區

您可以按過程 12.4「使用叢集管理員掛接 OCFS2 磁碟區」[第146頁] 中所述，手動或使用叢集管理員掛接 OCFS2 磁碟區。

過程 12.3 手動掛接 OCFS2 磁碟區

- 1 開啟終端機視窗，並以 `root` 身分登入。
- 2 請使用指令 `crm_mon` 檢查叢集是否上線。
- 3 從指令行掛接磁碟區，請使用 `mount` 指令。

警告：手動掛接的 OCFS2 設備

在測試以手動方式掛接 OCFS2 檔案系統之後，必須再次將其卸載，方可透過 OpenAIS 使用該檔案系統。

過程 12.4 使用叢集管理員掛接 OCFS2 磁碟區

若要使用 High Availability 軟體掛接 OCFS2 磁碟區，請在叢集中設定 ocf 檔案系統資源。下列程序使用 crm 外圍程序來設定叢集資源。或者，也可以使用 Heartbeat 設定資源。

- 1 啟動 crm 外圍程序，並重新建立新的組態：

```
crm
cib new filesystem
```

- 2 設定 Pacemaker 以在叢集內各節點上掛接檔案系統：

```
configure
primitive fs ocf:heartbeat:Filesystem \
    params device="/dev/sdb1" directory="/mnt/shared" fstype="ocfs2" \
    op monitor interval=120s
clone fs-clone fs meta interleave="true" ordered="true"
```

- 3 若要確保 Pacemaker 僅在同樣執行有 o2cb 資源之複製資源的節點上啟動 fs 複製資源，請新增並存限制：

```
colocation fs-with-o2cb INFINITY: fs-clone o2cb-clone
order start-fs-after-o2cb mandatory: o2cb-clone fs-clone
```

- 4 將組態上載至 CIB，並結束外圍程序：

```
cib commit filesystem
quit
```

12.6 如需更多資訊

如需有關 OCFS2 的詳細資訊，請參閱下列連結的內容：

<http://oss.oracle.com/projects/ocfs2/>
Oracle OCFS2 專案的首頁。

<http://oss.oracle.com/projects/ocfs2/documentation>
專案文件首頁中的《*OCFS2 User's Guide*》(OCFS2 使用者指南)。

分散式複製區塊設備 (DRBD)

DRBD 可讓您跨 IP 網路為位於兩個不同站台的兩個區塊設備建立鏡像。與 OpenAIS 搭配使用時，DRBD 支援分散式的 High Availability Linux 叢集。本章節將介紹如何安裝及設定 DRBD。

13.1 概念綜覽

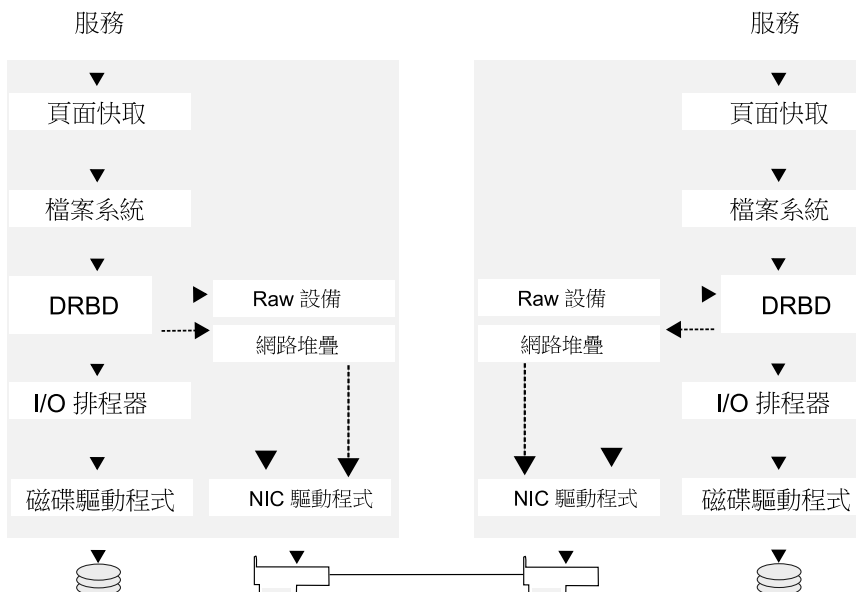
DRBD 能在將資料從主要設備複製到次要設備時，確保兩份資料保持一致。您可以將其視為網路 RAID 1。它會即時執行資料的鏡像複製，因此其複製會持續進行。應用程式無須知曉其資料實際上儲存在其他磁碟上。

重要：未加密資料

鏡像間的資料流量不會加密。為了保證資料交換的安全，您應該為連線部署虛擬私人網路 (VPN) 解決方案。

DRBD 是 Linux 核心模組，位於下層的 I/O 規劃程式和上層的檔案系統之間，請參閱圖形 13.1 「DRBD 在 Linux 中的位置」[第148頁]。若要與 DRBD 通訊，使用者需使用高階指令 `drbdadm`。為最大限度地提升靈活性，DRBD 隨附有低階工具 `drbdsetup`。

圖形 13.1 DRBD 在 Linux 中的位置



藉由 DRBD，使用者可以使用 Linux 支援的任何區塊設備，通常包括：

- 分割區或完整硬碟
- 軟體 RAID
- 邏輯磁碟區管理 (Logical Volume Manager, LVM)
- 企業磁碟區管理系統 (EVMS)

依預設，DRBD 使用 7780 及更高的 TCP 埠來處理 DRBD 節點之間的通訊。請確定您的防火牆不會阻止此埠上的通訊。

您必須先對 DRBD 設備進行設定，然後才能在其上建立檔案系統。所有與使用者資料相關的操作都會單獨透過 `/dev/drbd_R` 設備執行，不能在 Raw 設備上執行，因為 DRBD 會將 Raw 設備最後的 128 MB 容量用於儲存中繼資料。檔案系統只能在 `/dev/drbd<n>` 設備上建立，切勿在 Raw 設備上建立。

例如，如果 Raw 設備的大小為 1024 MB，則 DRBD 設備只能使用 896 MB 來儲存資料，另有 128 MB 隱藏留作儲存中繼資料之用。任何存取大小為 896 MB 至 1024 MB 之間的空間的嘗試都會失敗，因為這樣的空間不適用於使用者資料。

13.2 安裝 DRBD 服務

若要為 DRBD 安裝所需套件，請在網路叢集中的兩台 SUSE Linux Enterprise Server 機器上安裝 High Availability Extension 附加產品，如第 I 部分「安裝與設定」[第1頁]中所述。安裝 High Availability Extension 的同時也會安裝 DRBD 程式檔案。

若您不需要完整的叢集堆疊，只是想要使用 DRBD，請參閱表格 13.1 「DRBD RPM 套件」[第149頁]中列出之所有適用於 DRBD 的 RPM 套件。在最新版本中，drbd 套件已分割為多個獨立套件。

表格 13.1 DRBD RPM 套件

檔名	解釋
drbd	便利的套件，已分割成其他多個套件
drbd-bash-completion	drbdadm 的可程式化 Bash 補完支援
drbd-heartbeat	DRBD 的 Heartbeat 資源代辦(僅用於 Heartbeat)
drbd-kmp-default	DRBD 的核心模組(必需)
drbd-kmp-xen	DRBD 的 Xen 核心模組
drbd-udev	DRBD 的 udev 整合程序檔，用於管理 /dev/drbd/by-res 與 /dev/drbd/by-disk 中指向 DRBD 設備的符號連結
drbd-utils	DRBD 的管理公用程式(必需)
drbd-pacemaker	DRBD 的 Pacemaker 資源代辦

檔名	解釋
drbd-xen	DRBD 的 Xen 區塊設備管理程序檔
yast2-drbd	YaST DRBD 組態 (建議)

若要簡化 drbdadm 的使用，可以使用 RPM 套件 drbd-bash-completion 中的 Bash 補完支援。如果要在目前的外圍程序工作階段中將其啟用，請插入以下指令：

```
source /etc/bash_completion.d/drbdadm.sh
```

若要將其永久用於 root，請建立檔案 /root/.bashrc 並插入上述指令行。

13.3 設定 DRBD 服務

注意

下列步驟使用伺服器名稱 **jupiter** 與 **venus**，以及叢集資源名稱 **r0**，並將 **jupiter** 設定為主要節點。使用此指令時，請務必將節點與檔案名稱修改為您自己的節點與檔案名稱。

開始設定 DRBD 前，請確定 Linux 節點中的區塊設備已就緒並已分割 (如果需要)。下列步驟假設您擁有 **jupiter** 與 **venus** 兩個節點，並使用 TCP 埠 7780。請確定您的防火牆中已開啟此連接埠。

若要手動設定 DRBD，請執行下列步驟：

過程 13.1 手動設定 DRBD

1 以 root 身分登入。

2 變更 DRBD 的組態檔案：

2a 開啟檔案 /etc/drbd.conf 並插入以下幾行 (如果沒有的話)：

```
include "drbd.d/global_common.conf";
include "drbd.d/*.res";
```

從DRBD 8.3開始，組態檔案已分割為多個獨立檔案，位於目錄 `/etc/drbd.d` 下。

- 2b** 開啟檔案 `/etc/drbd.d/global_common.conf`。該檔案中包含部分預定義的值。轉至 `startup` 區段，並插入以下幾行：

```
startup {  
    # wfc-timeout degr-wfc-timeout outdated-wfc-timeout  
    # wait-after-sb;  
    wfc-timeout 1;  
    degr-wfc-timeout 1;  
}
```

這些選項可用於減少開機過程中的逾時，詳細資料請參閱 <http://www.drbd.org/users-guide-emb/re-drbdconf.html>。

- 2c** 建立檔案 `/etc/drbd.d/r0.res`，再根據具體情況變更以下幾行，然後儲存：

```
resource r0 { ❶  
    device /dev/drbd_r0 minor 0; ❷  
    disk /dev/sda1; ❸  
    meta-disk internal; ❹  
    on jupiter { ❺  
        address 192.168.1.10:7780; ❻  
    }  
    on venus { ❺ (page 151)  
        address 192.168.1.11:7780; ❻ (page 151)  
    }  
    syncer {  
        rate 7M; ❼  
    }  
}
```

- ❶ 資源的名稱。建議使用諸如 `r0`、`r1` 之類的資源名稱。
- ❷ DRBD 的設備名稱及其次要編號。建議以 `/dev/drbd` 開頭，後面加上資源名稱 (本例中為 `r0`)。
- ❸ 在兩個節點間複製的設備。請注意，本範例中兩個節點上的設備是相同的。如果需要不同的設備，請將 `disk` 參數移至 `on` 區段。
- ❹ `meta-disk` 參數通常包含值 `internal`，但您也可以指定具體的設備來儲存中繼資料。如需相關資訊，請參閱 <http://www.drbd.org/users-guide-emb/ch-internals.html#s-metadata>。

- ⑤ on 區段包含節點的主機名稱
- ⑥ 各節點的 IP 位址與埠號。每個資源都需要各自的連接埠，通常從 7780 開始。
- ⑦ 同步化速率。將其設定為頻寬的三分之一。此設定只能限制重新同步化操作，對鏡像複製不起作用。

3 檢查組態檔案的語法。若以下指令傳回錯誤，請檢查您的檔案：

```
drbdadm dump all
```

4 將 DRBD 組態檔案複製到其他節點：

```
scp /etc/drbd.conf venus:/etc/  
scp /etc/drbd.d/* venus:/etc/drbd.d/
```

5 在每個節點上分別輸入以下指令，啟始化兩個系統上的中繼資料。

```
drbdadm -- --ignore-sanity-checks create-md r0  
rcdrbd start
```

如果磁碟包含您不再需要的檔案系統，請使用以下指令破壞檔案系統結構，然後重複此步驟：

```
dd if=/dev/zero of=/dev/sdb1 count=10000
```

6 在每個節點上輸入以下指令，檢查 DRBD 狀態：

```
rcdrbd status
```

正常情況下會顯示如下內容：

```
drbd driver loaded OK; device status:  
version: 8.3.7 (api:88/proto:86-91)  
GIT-hash: ea9e28dbff98e331a62bcbcc63a6135808fe2917 build by phil@fat-tyre, 2010-01-13  
17:17:27  
m:res cs ro ds p mounted fstype  
0:r0 Connected Secondary/Secondary Inconsistent/Inconsistent C
```

7 在目標主要節點 (本範例中為 jupiter) 上啟動重新同步化程序：

```
drbdadm -- --overwrite-data-of-peer primary r0
```

8 使用 rcdrbd status 再次檢查狀態，您將看到：

```
...
m:res cs ro ds p mounted fstype
0:r0 Connected Primary/Secondary UpToDate/UpToDate C
```

兩個節點上 ds 列中的狀態 (磁碟狀態) 都必須為 UpToDate。

9 將 jupiter 設定為主要節點：

```
drbdadm primary r0
```

10 在您的 DRBD 設備上建立檔案系統，例如：

```
mkfs.ext3 /dev/drbd_r0
```

11 掛接檔案系統並投入使用：

```
mount /dev/drbd_r0 /mnt/
```

13.4 測試 DRBD 服務

如果安裝與組態程序按預期執行，您現在就可以執行基本的 DRBD 功能測試。此測試也有助於瞭解軟體的工作原理。

1 測試 jupiter 上的 DRBD 服務。

1a 開啟終端機主控台，然後以 root 身分登入。

1b 在 jupiter 上建立掛接點，如 /srv/r0mount：

```
mkdir -p /srv/r0mount
```

1c 掛接 drbd 設備：

```
mount -o rw /dev/drbd0 /srv/r0mount
```

1d 從主要節點建立檔案：

```
touch /srv/r0mount/from_node1
```

2 測試 venus 上的 DRBD 服務。

2a 開啟終端機主控台，然後以 `root` 身分登入。

2b 卸載 `jupiter` 上的磁碟：

```
umount /srv/r0mount
```

2c 在 `jupiter` 上輸入以下指令，降級 `jupiter` 上的 DRBD 服務：

```
drbdadm secondary r0
```

2d 在 `venus` 上，將 DRBD 服務升級為主要服務：

```
drbdadm primary r0
```

2e 在 `venus` 上檢查 `venus` 是否為主要節點：

```
rcdrbd status
```

2f 在 `venus` 上建立掛接點，如 `/srv/r0mount`：

```
mkdir /srv/r0mount
```

2g 在 `venus` 上掛接 DRBD 設備：

```
mount -o rw /dev/drbd0 /srv/r0mount
```

2h 驗證您在 `jupiter` 上建立的檔案是否可供檢視。

```
ls /srv/r0mount
```

此時應列出 `/srv/r0mount/from_node1` 檔案。

3 如果服務在兩個節點上都可執行，即表示 DRBD 設定已完成。

4 再次將 `jupiter` 設為主要節點。

4a 在 `venus` 上輸入以下指令，卸下 `venus` 上的磁碟：

```
umount /srv/r0mount
```

4b 在 `venus` 上輸入以下指令，降級 `venus` 上的 DRBD 服務：

```
drbdadm secondary r0
```

4c 在 `jupiter` 上，將 DRBD 服務升級為主要服務：

```
drbdadm primary r0
```

4d 在 `jupiter` 上檢查 `jupiter` 是否為主要節點：

```
rcdrbd status
```

- 5** 若要讓服務自動啟動並在伺服器出現問題時自動進行容錯移轉，您可以使用 OpenAIS 將 DRBD 設定為高可用性服務。如需安裝和設定適用於 SUSE Linux Enterprise 的 OpenAIS 的資訊，請參閱第 II 部分「組態與管理」[第 31 頁]。

13.5 調整 DRBD

調整 DRBD 的方法有多種：

1. 使用外部磁碟儲存中繼資料。這可以加快連線速度。
2. 建立 `udev` 規則，以便在 DRBD 設備之前變更讀取。將如下指令儲存到檔案 `/etc/udev/rules.d/82-dm-ra.rules` 中，並根據工作負載變更 `read_ahead_kb` 的值：

```
ACTION=="add", KERNEL=="dm-*", ATTR{bdi/read_ahead_kb}="4100"
```

此指令僅在您使用 LVM 時有效。

3. 啟動 Linux 軟體 RAID 系統上的 `bmbv`。在 DRBD 組態 (通常位於 `/etc/drbd.d/global_common.conf` 中) 的通用磁碟區段中使用以下指令：

```
disk {  
    use-bmbv;  
}
```

13.6 DRBD 疑難排解

`drbd` 設定涉及眾多不同元件以及產生原因不同的各種問題。以下各節將介紹幾種常見的情境並提供了多種解決方案。

13.6.1 組態

如果初始的 drbd 設定未按預期工作，則可能是組態有問題。

若要獲取有關組態的資訊，請執行下列步驟：

- 1 開啟終端機主控台，然後以 root 身分登入。
- 2 執行 drbdadm (含 -d 選項)，測試組態檔案。輸入下列指令：

```
drbdadm -d adjust r0
```

在 adjust 選項的試執行 (dry run) 期間，drbdadm 會將 DRBD 資源的實際組態與 DRBD 組態檔案進行比較，但不會執行呼叫。檢閱輸出，以確定您瞭解所有錯誤的來源及原因。

- 3 如果檔案 /etc/drbd.d/* 與 drbd.conf 中存在錯誤，請更正後再繼續。
- 4 如果分割區與設定均正確，請再次執行 drbdadm (不含 -d 選項)。

```
drbdadm adjust r0
```

此指令會將組態檔案套用於 DRBD 資源。

13.6.2 主機名稱

DRBD 的主機名稱區分大小寫，因此 Node0 與 node0 代表不同的主機。

如果您擁有多個網路設備，並希望使用專屬的網路設備，主機名稱可能不會解析成所使用的 IP 位址。在這種情況下，可以使用參數 `disable-ip-verification`。

13.6.3 TCP 埠 7788

如果您的系統無法連接至對等系統，可能是因為本地防火牆出現了問題。依預設，DRBD 使用 TCP 埠 7788 存取其他節點。請確定在兩個節點上均可存取此連接埠。

13.6.4 DRBD 設備在重新開機後損毀

如果 DRBD 不知道存放最新資料的真實設備，就會導致電腦分裂狀態。在這種情況下，各個 DRBD 子系統將會做為次要項目出現，並且不會相互連接。在這種情況下，會將以下訊息寫入到 `/var/log/messages`：

```
Split-Brain detected, dropping connection!
```

若要解決此問題，請在要丟棄資料的節點上輸入以下指令：

```
drbdadm secondary r0  
drbdadm -- --discard-my-data connect r0
```

在含有最新資料的節點上，輸入以下指令：

```
drbdadm connect r0
```

13.7 如需更多資訊

下列開放原始碼資源適用於 DRBD：

- 專案首頁 <http://www.drbd.org>。
- http://clusterlabs.org/wiki/DRBD_HowTo_1.0，由 Linux Pacemaker Cluster Stack Project 提供。
- 配送的產品中包含以下 DRBD 的 man 頁面：`drbd(8)`、`drbddisk(8)`、`drbdsetup(8)`、`drbdsetup(8)`、`drbdadm(8)`、`drbd.conf(5)`。
- `/usr/share/doc/packages/drbd/drbd.conf` 中可找到含備註的 DRBD 範例組態。

叢集 LVM

管理叢集上的共享儲存時，當儲存子系統發生變更時必須通知每個節點。廣泛用於管理本地儲存的 Linux Volume Manager 2 (LVM2) 已經過延伸，現可支援對整個叢集中磁碟區群組的透明管理。可使用與本地儲存相同的指令來管理叢集化磁碟區群組。

14.1 概念綜覽

叢集化 LVM 可與其他工具搭配使用：

分散式鎖定管理員 (DLM)

協調 cLVM 的磁碟存取。

邏輯磁碟區管理員 2 (LVM2)

能讓一個檔案系統靈活分散在多個磁碟上。LVM 可提供虛擬磁碟儲存區。

叢集邏輯磁碟區管理員 (cLVM)

協調對 LVM2 中繼資料的存取，讓每個節點瞭解相關的變更。cLVM 不會協調對共享資料本身的存取；若要讓其對此進行協調，必須在受 cLVM 管理的儲存區上設定 OCFS2 或其他叢集感知應用程式。

14.2 cLVM 的組態

某些情況下可以使用 cLVM 建立含以下幾層的 RAID 1 設備：

- **LVM** 如果您想要增加或減小檔案系統的大小，新增更多實體儲存或是建立檔案系統的快照，可以使用這項極為靈活的解決方案。有關此方法的介紹，請參閱第 14.2.1 節「案例：SAN 上 cLVM 與 iSCSI 搭配使用」[第 160 頁]。
- **DRBD** 此解決方案僅提供 RAID 0 (分割) 和 RAID 1 (鏡像)。有關最後一個方法的介紹，請參閱第 14.2.2 節「案例：cLVM 與 DRBD 搭配」[第 165 頁]。
- **MD 設備 (Linux 軟體 RAID 或 mdadm)** 雖然此解決方案提供所有 RAID 層級，但目前尚不支援叢集。

請確定您已滿足以下必要條件：

- 有透過光纖通道、FCoE、SCSI、iSCSI SAN 或 DRBD 等提供的共享儲存設備可用。
- 如果是 DRBD，兩個節點都必須是主要節點 (如下文程序中所述)。
- 檢查 LVM2 的鎖定類型是否能感知叢集。`/etc/lvm/lvm.conf` 中關鍵字 `locking_type` 的值必須包含 3 (應為預設值)。需要時，將組態複製到所有節點。

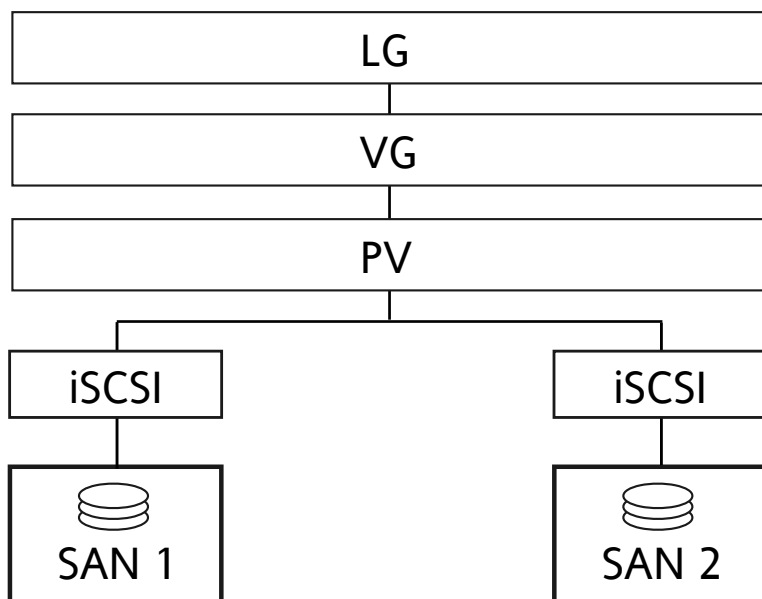
注意：先建立叢集資源

請先建立您的叢集資源，然後再建立 LVM 磁碟區。若不依照此順序，以後將無法移除這些磁碟區。

14.2.1 案例：SAN 上 cLVM 與 iSCSI 搭配使用

以下案例將使用兩個 SAN Box，它們會將 iSCSI 目標輸出至多個用戶端。圖形 14.1 「cLVM 搭配 iSCSI 的設定」[第 161 頁] 中說明了一般的情況。

圖形 14.1 cLVM 搭配 iSCSI 的設定



警告：資料損失

以下程序將損毀您磁碟上的資料！

開始請只設定一個 SAN Box。每個 SAN Box 必須輸出其自己的 iSCSI 目標。請執行下列步驟：

過程 14.1 設定 iSCSI 目標 (SAN)

- 1 執行 YaST 並按一下「網路服務」>「iSCSI 目標」，啟動 iSCSI 伺服器模組。
- 2 如果您希望電腦每次開機時啟動 iSCSI 目標，請選擇「開機時」，否則請選擇「手動」。
- 3 如果有防火牆在執行，則啟用「在防火牆中開啟埠」。
- 4 切換至「全域」索引標籤。如果需要驗證，請啟用內送驗證、外送驗證或兩者均啟用。在本例中，我們選取的是「無驗證」。

5 新增新的 iSCSI 目標：

5a 切換至「目標」索引標籤。

5b 按一下「新增」。

5c 輸入目標名稱。名稱必須採用以下格式：

`iqn.DATE.DOMAIN`

5d 如果您想要使用更具描述性的名稱，可以進行變更，前提是每個目標的識別碼都要唯一。

5e 按一下「新增」。

5f 在「路徑」中輸入設備名稱，並使用「*Scsiid*」。

5g 按兩次「下一步」。

6 出現警告對話方塊時，按一下「是」加以確認。

7 開啟組態檔案 `/etc/iscsi/iscsi.conf`，並將參數 `node.startup` 變更為 `automatic`。

接著，按如下所示設定 iSCSI 啟動器：

過程 14.2 設定 iSCSI 啟動器

1 執行 YaST，然後按一下「網路服務」>「iSCSI 啟動器」。

2 如果您要在電腦開機時啟動 iSCSI 啟動器，請選擇「開機時」，否則請設定「手動」。

3 切換至「探查」索引標籤，然後按一下「探查」按鈕。

4 新增 iSCSI 目標的 IP 位址和連接埠 (請參閱過程 14.1「設定 iSCSI 目標 (SAN)」[第161頁])。一般不用變更連接埠，使用其預設值即可。

5 如果使用驗證，請插入內送及外送的使用者名稱和密碼，否則請啟用「無驗證」。

6 選取「下一步」。清單中會顯示找到的連線。

7 按一下「完成」繼續。

8 開啟外圍程序，並以 root 身分登入。

9 測試 iSCSI 啟動器是否已正常啟動：

```
iscsiadm -m discovery -t st -p 192.168.3.100
192.168.3.100:3260,1 iqn.2010-03.de.jupiter:san1
```

10 建立工作階段：

```
iscsiadm -m node -l
Logging in to [iface: default, target: iqn.2010-03.de.jupiter:san2,
portal: 192.168.3.100,3260]
Logging in to [iface: default, target: iqn.2010-03.de.venus:san1,
portal: 192.168.3.101,3260]
Login to [iface: default, target: iqn.2010-03.de.jupiter:san2, portal:
192.168.3.100,3260]: successful
Login to [iface: default, target: iqn.2010-03.de.venus:san1, portal:
192.168.3.101,3260]: successful
```

使用 `ls SCSI` 查看設備名稱：

```
...
[4:0:0:2]    disk      IET        ...      0      /dev/sdd
[5:0:0:1]    disk      IET        ...      0      /dev/sde
```

尋找第三欄中含有 IET 的項目。在本例中，設備分別是 `/dev/sdd` 和 `/dev/sde`。

過程 14.3 建立 DLM 資源

1 啟動外圍程序，並以 root 身分登入。

2 執行 `crm configure`。

3 輸入下列指令：

```
primitive dlm ocf:pacemaker:controld
primitive clvm ocf:lvm2:clvmd \
    params daemon_timeout="30"
group dlm-clvm dlm clvm
clone dlm-clvm-clone dlm-clvm \
    meta interleave="true" ordered="true"
```

- 4 使用 `show` 指令檢閱所做的變更。
- 5 如果所有內容都正確無誤，請輸入 `commit`，然後使用 `exit` 指令離開 `crm`。

過程 14.4 建立 LVM 磁碟區群組

- 1 在過程 14.2「設定 iSCSI 啟動器」[第162頁]中執行了 iSCSI 啟動器的某個節點上開啟 `root` 外圍程序。
- 2 對磁碟 `/dev/sdd` 和 `/dev/sde` 使用指令 `pvccreate`，為 LVM 準備好實體磁碟區：

```
pvccreate /dev/sdd  
pvccreate /dev/sde
```

- 3 使用 `pvddisplay` 檢查是否所有內容正確無誤：

```
--- Physical volume ---  
PV Name                /dev/sdd  
VG Name                 clustervg  
PV Size                 509,88 MB / not usable 1,88 MB  
Allocatable             yes  
PE Size (KByte)         4096  
Total PE                127  
Free PE                 127  
Allocated PE            0  
PV UUID                 52okH4-nv3z-2AUL-GhAN-8DAZ-GMtU-Xrn9Kh  
  
--- Physical volume ---  
PV Name                /dev/sde  
VG Name                 clustervg  
PV Size                 509,84 MB / not usable 1,84 MB  
Allocatable             yes  
PE Size (KByte)         4096  
Total PE                127  
Free PE                 127  
Allocated PE            0  
PV UUID                 Ouj3Xm-AI58-lxB1-mWm2-xn51-agM2-0UuHFC
```

- 4 在兩個磁碟上建立叢集感知磁碟區群組：

```
vgcreate --clustered y clustervg /dev/sdd /dev/sde
```

- 5 使用 `vgdisplay` 檢查是否所有內容正確無誤：


```

--- Volume group ---
VG Name                clustervg
System ID
Format                 lvm2
Metadata Areas         2
Metadata Sequence No   1
VG Access               read/write
VG Status               resizable
Clustered              yes
Shared                 no
MAX LV                 0
Cur LV                 0
Open LV                 0
Max PV                 0
Cur PV                 2
Act PV                 2
VG Size                 1016,00 MB
PE Size                 4,00 MB
Total PE                254
Alloc PE / Size         0 / 0
Free PE / Size          254 / 1016,00 MB
VG UUID                 UCyWw8-2jqV-enuT-KH4d-NXQI-JhH3-J24anD

```

6 根據需要建立邏輯磁碟區：

```
lvcreate --name clusterlv --size 500M clustervg
```

建立磁碟區並啟動資源之後，會出現名為 `/dev/dm-0` 的新設備。建議您使用 LVM 資源上的叢集檔案系統，例如 OCFS。如需詳細資訊，請參閱第 12 章「*Oracle Cluster File System 2*」[第139頁]

14.2.2 案例：cLVM 與 DRBD 搭配

如果您的資料中心分佈在城市、國家甚至大陸的不同位置，可以參照以下案例。

過程 14.5 使用 DRBD 建立叢集感知磁碟區群組

1 建立主要/次要 DRBD 資源：

1a 首先，如過程 13.1「手動設定 DRBD」[第150頁]中所述將一部 DRBD 設備設定成主要或次要設備。確定兩個節點上的磁碟狀態均為最新。可以使用 `cat /proc/drbd` 或 `rcdrbd status` 指令來檢查。

1b 在組態檔案 (通常類似於 `/etc/drbd.d/r0.res`) 中新增以下選項：

```
resource r0 {
    startup {
        become-primary-on both;
    }

    net {
        allow-two-primaries;
    }
    ...
}
```

1c 將變更後的組態檔案複製到另一個節點，例如：

```
scp /etc/drbd.d/r0.res venus:/etc/drbd.d/
```

1d 在兩個節點上執行以下指令：

```
drbdadm disconnect r0
drbdadm connect r0
drbdadm primary r0
```

1e 檢查節點的狀態：

```
cat /proc/drbd
...
0: cs:Connected ro:Primary/Primary ds:UpToDate/UpToDate C r----
```

2 將 `clvm` 資源做為複製資源包括在 Pacemaker 組態中，並使之依賴於 DLM 複製資源。如需詳細指示，請參閱過程 14.3「建立 DLM 資源」[第163頁]。繼續之前，請先確定已在叢集中成功啟動這些資源。可以使用 `crm_mon` 或 GUI 檢查執行中的服務。

3 使用 `pvccreate` 指令為 LVM 備妥實體磁碟區。例如，在 `/dev/drbd_r0` 設備上使用如下指令：

```
pvccreate /dev/drbd_r0
```

4 建立叢集感知磁碟區群組：

```
vgcreate --clustered y myclusterfs /dev/drbd_r0
```

5 根據需要建立邏輯磁碟區。您有時可能需要變更邏輯磁碟區的大小。例如，使用以下指令建立 4 GB 的邏輯磁碟區：

```
lvcreate --name testlv -L 4G myclusterfs
```

- 6 為了確保在整個叢集範圍啟動磁碟區群組，請按如下方式設定 LVM 資源：

```
primitive vg1 ocf:heartbeat:LVM \  
    params volgrpname="myclusterfs"  
clone vg1-clone vg1 \  
    meta interleave="true" ordered="true"  
colocation colo-vg1 inf: vg1-clone dlm-clvm-clone  
order order-vg1 inf: dlm-clvm-clone vg1-clone
```

- 7 若要僅在一個節點上獨占啟動磁碟區群組，請使用下面的範例；在此範例中，由於針對非叢集化應用程式採用其他保護措施，cLVM 可避免在多個節點上啟動 VG 中的所有邏輯磁碟區：

```
primitive vg1 ocf:heartbeat:LVM \  
    params volgrpname="myclusterfs" exclusive="yes"  
colocation colo-vg1 inf: vg1 dlm-clvm-clone  
order order-vg1 inf: dlm-clvm-clone vg1
```

- 8 現在，VG 中的邏輯磁碟區可做為檔案系統掛接或 RAW 使用量使用。請確定使用它們的服務務必具有正確的相依性，這樣才能在啟動 VG 後對它們進行並存和排序處理。

完成這些組態設定步驟後，即可像在任何獨立工作站上一樣進行 LVM2 組態設定。

14.3 明確設定適合的 LVM2 設備

如果看似有多部設備共享同一個實體磁碟區簽名 (多重路徑設備或 DRBD 就有可能發生這種情況)，建議明確設定 LVM2 掃描 PV 的設備。

例如，如果 `vgcreate` 指令使用實體設備而非使用鏡像複製區塊設備，將使 DRBD 感到困惑，從而導致 DRBD 處於電腦分裂狀態。

若要停用 LVM2 的單一設備，請執行以下操作：

- 1 編輯 `/etc/lvm/lvm.conf` 檔案並搜尋以 `filter` 開頭的行。
- 2 該處的模式將被視為正規表示式進行處理。前置「a」表示接受要掃描的設備模式，前置「r」表示拒絕依照該設備模式的設備。
- 3 若要移除名為 `/dev/sdb1` 的設備，請將下列表示式新增至過濾器規則：

```
"r|^/dev/sdb1$|"
```

完整的過濾器行如下所示：

```
filter = [ "r|^/dev/sdb1$|", "r|/dev/.*/by-path/.*/",  
"r|/dev/.*/by-id/.*/", "a/.*/" ]
```

接受 DRBD 和 MPIO 設備但拒絕所有其他設備的過濾器行如下所示：

```
filter = [ "a|/dev/drbd.*|", "a|/dev/.*/by-id/dm-uuid-mpath-.*/", "r/.*/"  
]
```

4 寫入組態檔案並將其複製到所有叢集節點。

14.4 如需更多資訊

完整資訊可參閱 Pacemaker 郵寄清單 (網址為 <http://www.clusterlabs.org/wiki/Help:Contents>)。

官方 cLVM FAQ 可在 <http://sources.redhat.com/cluster/wiki/FAQ/CLVM> 中找到。

儲存保護

High Availability 叢集堆疊的首要任務是保護資料的完整性。所採用的方法是阻止在未經協調的情況下同時存取資料儲存區。舉例來說，叢集中只會掛接一次 ext3 檔案系統，以及只有在與其他叢集節點協調後才會掛接 OCFS2 磁碟區。在正常運作的叢集中，如果使用中的資源超出其同步限制，Pacemaker 會偵測到此情況，並啟動復原操作。而且，其規則引擎永遠不會超出這些限制。

但是，如果系統中有數個協調者，便可能導致網路分割區或軟體出現故障。如果系統允許出現這種所謂的電腦分裂情況，就有可能產生資料損毀。為此，叢集堆疊中新增了數道安全層以降低風險。

其中最關鍵的元件是 IO 圍籬區隔/STONITH，它可以確保在啟動儲存之前先終止其他所有存取。其他機制包括 cLVM2 獨佔式啟動或 OCFS2 檔案鎖定支援，它們可保護系統，避免管理或應用程式故障。如果再加以適當設定，這些安全措施就可以有效地防止電腦分裂現象，以免對系統造成損害。

本章先介紹可充分利用自身儲存區的IO圍籬區隔機制，然後介紹為確保獨佔式儲存區存取而增設的保護層。這兩項機制聯合可實現更高級別的保護。

15.1 基於儲存區的圍籬區隔

使用電腦分裂偵測器(SBD)、監視程式支援和external/sbd STONITH代辦，可以有效地避免出現電腦分裂的情況。

15.1.1 綜覽

如果某個環境中的所有節點都可以存取共享儲存區，系統會分配一個小小的分割區 (1MB) 用於 SBD。設定各自的精靈之後，系統會連接每個節點上 SBD，然後啟動其餘的叢集堆疊。當其他所有叢集元件都關閉後，SBD 才會終止，這樣便確保了只要叢集資源啟動，SBD 就會加以監督。

精靈會自動在分割區上為自己配置一個訊息槽，然後持續監控，查看是否有傳送給它的訊息。一旦收到訊息，精靈會立即回應請求，例如為圍籬區隔啟動關機或重新開機操作。

精靈還會持續監控與儲存設備的連線，如果無法存取分割區，就會自行終止。這樣可保證精靈不會錯過圍籬區隔訊息。如果叢集資料位於其他分割區的同一個邏輯單位，則一旦與儲存區失去連線，載入的工作便會終止，因此不會增加故障點。

監視程式支援進一步提升了安全性。最新的系統支援硬體監視程式，該監視程式必須透過軟體用戶端更新，否則硬體會強制系統重新啟動。這樣可以保障 SBD 程序自身不出現故障，例如沒有回應或陷入 IO 錯誤。

15.1.2 設定基於儲存區的保護

設定基於儲存區的保護時必須執行下列步驟：

- 1 建立 SBD 分割區 [第171頁]
- 2 設定軟體監視程式 [第172頁]
- 3 啟動 SBD 精靈 [第172頁]
- 4 測試 SBD [第173頁]
- 5 設定圍籬區隔資源 [第173頁]

下列所有程序都必須以 `root` 身分執行。在啟動之前，確定已符合下列要求：

重要：要求

- 環境中必須有所有節點都能存取的共享儲存區。
 - 共享儲存節區不得使用基於主機的 RAID、cLVM2 或 DRBD。
 - 但是，建議使用基於儲存區的 RAID 和多重路徑來提升可靠性。
-

建立 SBD 分割區

建議在設備啟動時建立 1MB 的分割區。如果 SBD 設備位於多重路徑群組中，MPIO 的向下路徑偵測可能會導致一定程度的延遲，因此需要調整 SBD 使用的逾時。達到 msgwait 逾時後，系統會認為訊息已傳送到節點。對多重路徑而言，這一時間也就是 MPIO 偵測到路徑故障並切換到下一個路徑所需的時間。您可能需要在自己的實際環境中對此進行測試。如果節點未能迅速及時地更新監視程式計時器，就會自行終止。監視程式的逾時必須短於 msgwait 的逾時，以一半為佳。

在下例中，SBD 分割區以 `/dev/SBD` 表示。請將其取代為實際的路徑名稱，例如：`/dev/sdc1`。

重要：覆寫現有資料

確定要用於 SBD 的設備沒有儲存任何資料。sdb 指令會直接覆寫設備，而不提出任何確認請求。

1 使用下列指令啟始化 SBD 設備：

```
sbd -d /dev/SBD create
```

此指令會在設備中寫入標頭，並為最多 255 個共享此設備的節點建立插槽 (採用預設時間設定)。

2 如果 SBD 設備位於多重路徑群組中，則調整 SBD 使用的逾時。逾時可以在啟始化 SBD 設備時指定 (所有逾時均以秒計)：

```
/usr/sbin/sbd -d /dev/SBD -4 $msgwait -1 $watchdogtimeout create
```

3 使用下列指令查看寫入設備的內容：

```
sbd -d /dev/SBD dump
Header version      : 2
Number of slots     : 255
Sector size         : 512
Timeout (watchdog)  : 5
Timeout (allocate)  : 2
Timeout (loop)      : 1
Timeout (msgwait)   : 10
```

如您所見，逾時也會存入標頭，以確保所有參與的節點在逾時上達成一致。

設定軟體監視程式

建議將 Linux 系統設定為使用監視程式，包括在系統開機時載入相應的監視程式驅動程式。

- 在 HP 硬體上即 hpwdt 模組。
- 如果系統配有 Intel TCO，可以使用 iTCO_wdt。softdog 是最常見的驅動程式，不過建議使用與實際硬體相整合的驅動程式。

請參閱核心套件之 `drivers/watchdog` 中列出的多種選擇。

啟動 SBD 精靈

SBD 精靈是叢集堆疊的重要組成部分。只要叢集堆疊正在執行，SBD 精靈就必須執行，即使部分出現故障也是如此，這樣才能加以圍籬區隔。

- 1 若要啟動 OpenAIS init 程序檔並停止 SBD，請在 `/etc/sysconfig/sbd` 中新增以下內容：

```
SBD_DEVICE="/dev/SBD"
# The next line enables the watchdog support:
SBD_OPTS="-W"
```

如果無法存取 SBD 設備，精靈將無法啟動並中斷 OpenAIS 的啟動。

注意

如果節點無法存取 SBD 設備，便有可能陷入重新開機的無限循環。這在技術上並沒有錯，但在某些管理規則下可能會帶來麻煩。出現這種情況時，可能需要系統在開機時不自動啟動 OpenAIS。

- 2 繼續下一步之前，請先執行 `rcopenais restart`，以確保所有節點上的 SBD 均已啟動。

測試 SBD

- 1 下列指令會從 SBD 設備傾印節點插槽及其目前的訊息：

```
sbd -d /dev/SBD list
```

現在您會看到，此處列出 SBD 的所有叢集節點均已啟動，訊息槽應顯示清除。

- 2 嘗試向其中一個節點傳送一則測試訊息：

```
sbd -d /dev/SBD message nodea test
```

- 3 節點會在系統記錄中確認收到訊息：

```
Aug 29 14:10:00 nodea sbd: [13412]: info: Received command test from nodeb
```

這証實了節點上的 SBD 的確已啟動並執行，並且已經收到訊息。

設定圍籬區隔資源

- 1 若要完成 SBD 設定，必須在 CIB 中將 SBD 做為 STONITH/圍籬區隔機制進行啟動，如下所示：

```
crm configure
crm(live)configure# property stonith-enabled="true"
crm(live)configure# property stonith-timeout="30s"
crm(live)configure# primitive stonith:external/sbd params
sbd_device="/dev/SBD"
crm(live)configure# commit
crm(live)configure# quit
```

由於系統會自動配置節點插槽，因此不需要手動定義主機清單。

- 2 因為現在透過 SBD 機制來實現圍籬區隔功能，所以請停用先前設定過的其他所有圍籬區隔設備。

只要啟動資源，叢集就會順利得到設定以提供共享儲存區圍籬區隔功能，並在需要對節點進行圍籬區隔時使用此方法。

15.2 確保啟動獨佔性儲存

本節介紹的 `sfex` 是一個附加的低層機制，用於將對共享儲存區的存取鎖定給某個節點。請注意，`sfex` 不會取代 `STONITH`。由於 `sfex` 需要使用共享儲存區，因此建議在儲存區的其他分割區上使用上述 `external/sbd` 圍籬區隔機制。

由於內在設計的原因，`sfex` 無法用於要求同時作業的工作負載 (如 `OCFS2`)，但可用作傳統容錯移轉方式之工作負載的保護層。其效果與保留 `SCSI-2` 類似，但更加常用。

15.2.1 綜覽

在共享儲存區環境中，會額外設定一個小分割區，用於儲存一或多個鎖定。

節點必須先取得保護鎖定，才能獲取受保護的資源。次序由 `Pacemaker` 強制設定，`sfex` 元件可確保即使 `Pacemaker` 遇到電腦分裂的狀況，系統也不會多次授予鎖定。

系統必須定期重新整理鎖定，這樣即使節點停止回應，也不會永久封鎖鎖定，其他節點仍能繼續處理。

15.2.2 設定

下例說明了如何建立用於 `sfex` 的共享分割區，以及如何在 `CIB` 中為 `sfex` 鎖定設定資源。一個 `sfex` 分割區可保存任意個鎖定 (預設為一個)，每個鎖定需要配置 1 KB 儲存空間。

重要：要求

- `sfex` 的共享分割區應與所要保護的資料位於同一個邏輯單位上。
 - 共享 `sfex` 分割區不得使用基於主機的 `RAID` 或 `DRBD`。
 - 可以使用 `cLVM2` 邏輯磁碟區。
-

過程 15.1 建立 sfex 分割區

- 1 建立一個共享分割區用於 sfex。記下此分割區的名稱，並用其取代下面的 /dev/sfex。
- 2 使用以下指令建立 sfex 中繼資料：

```
sfex_init -i 1 /dev/sfex
```

- 3 驗證該中繼資料已正確建立：

```
sfex_stats -i 1 /dev/sfex ; echo $?
```

此指令應傳回 2，因為目前並未鎖定。

過程 15.2 設定 sfex 鎖定的資源

- 1 sfex 鎖定透過 CIB 中的資源表示，設定如下所示：

```
primitive sfex_1 ocf:heartbeat:sfex \  
# params device="/dev/sfex" index="1" collision_timeout="1" \  
    lock_timeout="70" monitor_interval="10" \  
# op monitor interval="10s" timeout="30s" on_fail="fence"
```

- 2 若要透過 sfex 鎖定保護資源，則在保護對象和 sfex 資源之間建立強制性順序和配置限制。假設要保護之資源的 ID 為 filesystem1：

```
# order order-sfex-1 inf: sfex_1 filesystem1  
# colocation colo-sfex-1 inf: filesystem1 sfex_1
```

- 3 如果使用群組語法，請將 sfex 資源做為第一個資源新增到群組中：

```
# group LAMP sfex_1 filesystem1 apache ipaddr
```


Samba 叢集

叢集化的 Samba 伺服器能為您的異質網路提供 High Availability 解決方案。本章介紹了一些背景知識以及如何設定叢集化 Samba 伺服器。

16.1 概念綜覽

Samba 使用 Trivial Database (TDB) 已有多年。TDB 允許多個應用程式同時執行寫入操作。為了確保所有寫入操作成功執行且彼此不發生衝突，TDB 使用了內部鎖定機制。

Cluster Trivial Database (CTDB) 是現有 TDB 的一個小延伸功能。專案本身將 CTDB 描述為「Samba 和其他專案用於儲存暫存資料之 TDB 資料庫的叢集實作」。

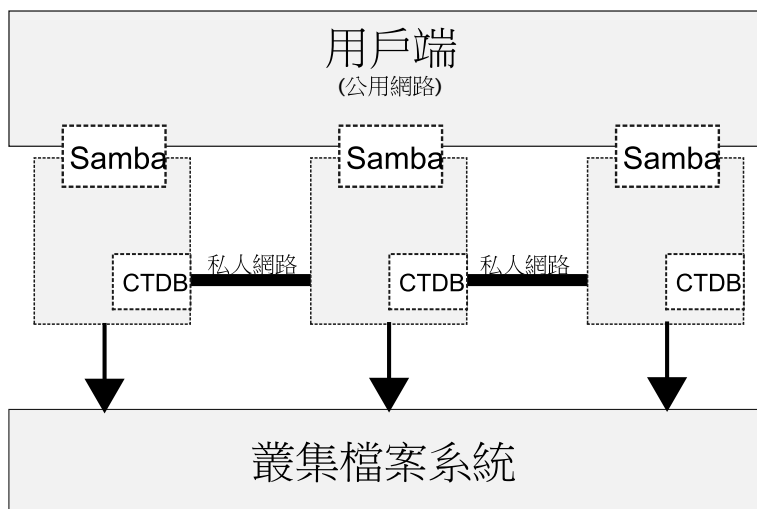
每個叢集節點都執行一個本地 CTDB 精靈。Samba 會與其本地 CTDB 精靈通訊，而不是直接寫入其 TDB。精靈透過網路交換中繼資料，但實際的讀寫操作是在一個具有快速儲存區的本地副本上執行的。CTDB 的概念請參閱圖形 16.1 「CTDB 叢集的結構」[第178頁]。

注意：CTDB 僅用於 Samba

目前實作的 CTDB 資源代辦將 CTDB 設定為僅管理 Samba。包括 IP 容錯移轉在內的其他內容應使用 Pacemaker 進行設定。

另外，僅完全同質的網路支援 CTDB。例如，叢集中的所有節點需要擁有相同的結構，i586 與 x86_64 不能合用。

圖形 16.1 CTDB 叢集的結構



叢集化的 Samba 伺服器必須共享某些資料：

- 將 Unix 使用者及群組 ID 與 Windows 使用者及群組相關聯的映射表。
- 使用者資料庫必須在所有節點之間同步。
- Windows 網域中成員伺服器的加入資訊必須在所有節點上都可用。
- 中繼資料必須在所有節點上都可用，例如活動的 SMB 工作階段、共享連線以及各類鎖定。

目的是讓擁有 $N+1$ 個節點的叢集化 Samba 伺服器快於僅有 N 個節點的 Samba 伺服器。一個節點不會慢於一部未叢集化的 Samba 伺服器。

16.2 基本組態

注意：變更的組態檔案

CTDB 資源代辦會自動變更 `/etc/sysconfig/ctdb` 與 `/etc/samba/smb.conf`。使用 `crminfo CTDB` 可以列出能夠為 CTDB 資源指定的所有參數。

若要設定叢集化的 Samba 伺服器，請執行下列步驟：

1 準備叢集：

1a 設定叢集 (OpenAIS、Pacemaker、OCFS2)，如本指南第 II 部分「組態與管理」[第31頁] 中所述。

1b 設定共享檔案系統 (例如 OCFS2) 並予以掛接 (例如，掛接至 /shared)。

1c 如果想要開啟 POSIX ACL，則將其啟用：

- 對於新 OCFS2 檔案系統，使用：

```
mkfs.ocfs2 --fs-features=xattr ...
```

- 對於現有的 OCFS2 檔案系統，使用：

```
tunefs.ocfs2 --fs-feature=xattr DEVICE
```

確定在檔案系統資源中指定了 `acl` 選項。按如下所示使用 `crm` 外圍程序：

```
crm(live)configure# primary ocfs2-3 ocf:heartbeat:Filesystem  
options="acl" ...
```

1d 確定 `ctdb`、`smb`、`nmb` 與 `winbind` 服務已停用：

```
chkconfig ctdb off  
chkconfig smb off  
chkconfig nmb off  
chkconfig winbind off
```

2 在共享檔案系統上為 CTDB 鎖定與 Samba 狀態建立目錄：

```
mkdir -p /shared/samba/private
```

3 在 /etc/ctdb/nodes 中插入包含叢集中每個節點之全部私人 IP 位址的所有節點：

```
192.168.1.10  
192.168.1.11
```

4 將 CTDB 資源新增至叢集：

```
crm configure
crm(live)configure# primitive ctdb ocf:heartbeat:CTDB params \
    ctdb_recovery_lock="/shared/samba/ctdb.lock" \
    smb_private_dir="/shared/samba/private" \
    op monitor timeout=20 interval=10
crm(live)configure# clone ctdb-clone ctdb \
    meta globally-unique="false" interleave="true"
crm(live)configure# colocation ctdb-with-fs inf: ctdb-clone fs-clone
crm(live)configure# order start-ctdb-after-fs inf: fs-clone ctdb-clone
crm(live)configure# commit
```

5 新增叢集化 IP 位址：

```
crm(live)configure# primitive ip ocf:heartbeat:IPaddr2 params
ip=192.168.2.222 \
    clusterip_hash="sourceip-sourceport" op monitor interval=60s
crm(live)configure# clone ip-clone ip meta globally-unique="true"
crm(live)configure# colocation ip-with-ctdb inf: ip-clone ctdb-clone
crm(live)configure# order start-ip-after-ctdb inf: ctdb-clone ip-clone
crm(live)configure# commit
```

6 檢查結果：

```
crm status
Clone Set: dlm-clone
    Started: [ hex-14 hex-13 ]
Clone Set: o2cb-clone
    Started: [ hex-14 hex-13 ]
Clone Set: c-ocfs2-3
    Started: [ hex-14 hex-13 ]
Clone Set: ctdb-clone
    Started: [ hex-14 hex-13 ]
Clone Set: ip-clone (unique)
    ip:0      (ocf::heartbeat:IPaddr2):      Started hex-13
    ip:1      (ocf::heartbeat:IPaddr2):      Started hex-14
```

7 從用戶端機器執行測試。在 Linux 用戶端上，執行以下指令，確定是否可以將檔案複製到系統或從系統複製檔案：

```
smbclient //192.168.2.222/myshare
```

16.3 對叢集化 Samba 進行除錯與測試

若要對叢集化 Samba 伺服器進行除錯，可以使用下列適用於不同層級的工具：

ctdb_diagnostics

執行此工具可以診斷叢集化 Samba 伺服器。該操作會提供大量除錯訊息，協助您追蹤可能會遇到的各種問題。

ctdb_diagnostics 指令會搜尋下列檔案，這些檔案必須在所有節點上均可用：

```
/etc/krb5.conf
/etc/hosts
/etc/ctdb/nodes
/etc/sysconfig/ctdb
/etc/resolv.conf
/etc/nsswitch.conf
/etc/sysctl.conf
/etc/samba/smb.conf
/etc/fstab
/etc/multipath.conf
/etc/pam.d/system-auth
/etc/sysconfig/nfs
/etc/exports
/etc/vsftpd/vsftpd.conf
```

如果存在 /etc/ctdb/public_addresses 與 /etc/ctdb/static-routes 檔案，則也會對其進行檢查。

ping_pong

使用 ping_pong 可以檢查您的檔案系統是否適用 CTDB。它會對叢集檔案系統執行某些測試，例如連貫性和效能測試 (請參閱 http://wiki.samba.org/index.php/Ping_pong)，讓您瞭解叢集在高負載下的表現。

若要對叢集檔案系統的某些方面進行測試，請執行下列步驟：

過程 16.1 測試叢集檔案系統的連貫性和效能

- 1 在一個節點上啟動指令 ping_pong 並將預留位址 *N* 取代為節點數量加 1。檔案名稱可以從共享儲存區獲得，因此可在所有節點上存取：

```
ping_pong data.txt N
```

僅執行一個節點時，鎖定率應該會很高。如果程式未顯示鎖定率，請更換您的叢集檔案系統。

- 2 在另一個節點上啟動第二個 ping_pong，並使用相同的參數。

鎖定率應該會有明顯的下降。如果您的叢集檔案系統出現以下任何一種情況，請予以更換：

- `ping_pong` 未輸出每秒的鎖定率
- 兩個例項中的鎖定率相差較大
- 啟動第二個例項後，鎖定率未下降

3 啟動第三個 `ping_pong`。新增其他節點並注意鎖定率的變化。

4 逐步停止各個 `ping_pong` 指令。在回到單一節點的情況之前，應會看到鎖定率不斷提高。如果未發生預期的情況，請更換您的叢集檔案系統。

16.4 如需更多資訊

- [http://linux-ha.org/wiki/CTDB_\(resource_agent\)](http://linux-ha.org/wiki/CTDB_(resource_agent))
- http://wiki.samba.org/index.php/CTDB_Setup
- <http://ctdb.samba.org>
- http://wiki.samba.org/index.php/Samba_%26_Clustering

IV. 疑難排解與參考

疑難排解

當出現奇怪的現象時，使用者往往會搞不清楚狀況，尤其是剛剛開始使用 High Availability 的使用者。不過，可以借助幾個公用程式進一步瞭解 High Availability 的內部程序。本章推薦了各種解決方案。

17.1 安裝問題

疑難排解安裝套件或連線叢集時所遇到的問題。

是否已安裝 HA 套件？

設定和管理叢集所需的套件包含於 High Availability 安裝模式中，由 High Availability Extension 提供。

請檢查 High Availability Extension 是否在各叢集節點上安裝為 SUSE Linux Enterprise Server 11 SP1 的附加產品，以及「*High Availability*」模式是否依第 3.1 節「安裝 High Availability Extension」[第19頁]中所述安裝於各機器上。

所有叢集節點的初始組態是否都相同？

若要在彼此之間進行通訊，則屬於相同叢集的所有節點均需依第 3.2 節「初始叢集設定」[第20頁]中所述使用相同的 `bindnetaddr`、`mcastaddr` 與 `mcastport`。

請檢查 `/etc/corosync/corosync.conf` 中為所有叢集節點設定的通訊通道與選項是否都相同。

若使用加密通訊，請檢查是否所有叢集節點上的 `/etc/corosync/authkey` 檔案都可使用。

所有 `corosync.conf` 設定及 `nodeid` 例外必須相同；所有節點上的 `authkey` 檔案必須一致。

防火牆允許透過 `mcastport` 進行通訊嗎？

若用於在各叢集節點間通訊的 `mcastport` 被防火牆阻擋，這些節點將無法相互查看。依第 3.1 節「安裝 High Availability Extension」[第19頁] 中所述使用 YaST 設定初始設定時，通常會自動調整防火牆設定。

若要確保 `mcastport` 不被防火牆阻擋，請檢查各節點上 `/etc/sysconfig/SuSEfirewall2` 中的設定。或者，啟動各叢集節點上的 YaST 防火牆模組。按一下「允許的服務」>「進階」之後，將 `mcastport` 新增至允許的「UDP 埠」清單，然後確認變更。

在各叢集節點上啟動了 OpenAIS 嗎？

使用 `/etc/init.d/openais status` 檢查各叢集節點上的 OpenAIS 狀態。若 OpenAIS 沒有執行，則透過執行 `/etc/init.d/openais start` 將其啟動。

17.2 「除錯」HA 叢集

以下指令將顯示資源作業歷程 (選項 `-o`) 和非使用中資源 (`-r`):

```
crm_mon -o -r
```

狀態變更後，系統會重新整理顯示的內容 (若要取消此功能，則按 `Ctrl + C`)。範例顯示如下：

範例 17.1 停止的資源

Refresh in 10s...

```
=====
Last updated: Mon Jan 19 08:56:14 2009
Current DC: d42 (d42)
3 Nodes configured.
3 Resources configured.
=====

Online: [ d230 d42 ]
OFFLINE: [ clusternode-1 ]

Full list of resources:

Clone Set: o2cb-clone
           Stopped: [ o2cb:0 o2cb:1o2cb:2 ]
Clone Set: dlm-clone
           Stopped [ dlm:0 dlm:1 dlm:2 ]
mySecondIP      (ocf::heartbeat:IPaddr):          Stopped

Operations:
* Node d230:
  aa: migration-threshold=1000000
    + (5) probe: rc=0 (ok)
    + (37) stop: rc=0 (ok)
    + (38) start: rc=0 (ok)
    + (39) monitor: interval=15000ms rc=0 (ok)
* Node d42:
  aa: migration-threshold=1000000
    + (3) probe: rc=0 (ok)
    + (12) stop: rc=0 (ok)
```

首先連線節點 (請參閱節 17.3 [第187頁])。然後檢查資源與作業。

<http://clusterlabs.org/wiki/Documentation> 中的《*Configuration Explained*》(組態說明, PDF 檔案) 在「*How Does the Cluster Interpret the OCF Return Codes?*」(叢集如何解譯 OCF 傳回代碼?) 一節介紹了三種不同的復原類型。

17.3 常見問題集

我的叢集狀態為何?

若要查看叢集目前的狀態, 請使用程式 `crm_mon` 或 `crm status`。此時會顯示目前 DC, 以及目前節點已知的所有節點與資源。

我叢集的一些節點相互間看不到。

導致的原因有多種：

- 請先查看組態檔案 `/etc/corosync/corosync.conf`，然後檢查叢集內各節點的多路廣播位址是否相同 (在 `interface` 區段中搜尋關鍵字 `mcastaddr`)。
- 檢查防火牆設定。
- 檢查交換器是否支援多路傳播位址
- 檢查節點之間的連線是否中斷。最常見的原因是防火牆設定不正確。這還可能是導致電腦分裂狀況的原因，其中的叢集被分區。

我要列出目前已知的資源。

使用指令 `crm_resource -L` 瞭解目前資源。

我已設定某個資源，但它總是失敗。

若要用 `ocf-tester` 檢查 OCF 程序檔，例如：

```
ocf-tester -n ipl -o ip=YOUR_IP_ADDRESS \  
/usr/lib/ocf/resource.d/heartbeat/IPaddr
```

若需納入多個參數，可以使用多個 `-o`。執行 `crmra info` 代辦，可顯示必需參數和可選參數的清單，指令範例如下：

```
crm ra info ocf:heartbeat:IPaddr
```

在執行 `ocf-tester` 之前，請確定該資源不受叢集管理。

我剛收到一條失敗訊息。能取得詳細資訊嗎？

您可以隨時在指令中新增 `--verbose` 參數。若多次執行此操作，將產生冗長的除錯輸出。請參閱 `/var/log/messages` 中提供的有用提示。

如何清理我的資源？

請使用以下指令：

```
crm resource list  
crm resource cleanup rscid [node]
```

如果不指定節點，系統會清理所有節點上的資源。如需詳細資訊，請參閱第 6.4.2 節「清理資源」[第102頁]。

無法掛接 ocfs2 設備。

檢查 `/var/log/message` 是否包含以下行：

```
Jan 12 09:58:55 clusternode2 lrmd: [3487]: info: RA output:
(o2cb:1:start:stderr) 2009/01/12_09:58:55
    ERROR: Could not load ocfs2_stackglue
Jan 12 16:04:22 clusternode2 modprobe: FATAL: Module ocfs2_stackglue not
found.
```

此範例中，缺少了核心模組 `ocfs2_stackglue.ko`。安裝套件 `ocfs2-kmp-default`、`ocfs2-kmp-pae` 或 `ocfs2-kmp-xen`，具體視所安裝的核心而定。

17.4 獲取詳細資訊

如需 Linux 和 Heartbeat 上有關高可用性的其他資訊，包括設定叢集資源及管理
和自定 Heartbeat 叢集，請參閱 <http://clusterlabs.org/wiki/Documentation>。

叢集管理工具

High Availability Extension 隨附可協助您利用指令行來管理叢集的多功能工具組。本章介紹管理 CIB 中的叢集組態與叢集資源所需的工具。第 17 章「疑難排解」[第185頁]中介紹了管理資源代辦的其他指令行工具和用於對設定進行除錯和疑難排解的工具。

以下清單提供了多項與叢集管理相關的任務，並簡要介紹了用於完成這些任務的工具：

監控叢集狀態

`crm_mon` 指令可讓您監控叢集狀態與組態。其輸出包括節點數、`uname`、`uuid`、狀態、叢集中設定的資源及其各自的目前狀態。`crm_mon` 的輸出可顯示在主控台中或列印成 HTML 檔案。如果提供無狀態區段的叢集組態檔案，`crm_mon` 就會建立節點及資源的綜覽 (如檔案中所指定)。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_mon(8)` [第213頁]。

管理 CIB

`cibadmin` 指令為低階管理指令，用於操作 Heartbeat CIB。它可用來傾印、更新及修改所有或部分 CIB，刪除整個 CIB，或執行其他 CIB 管理作業。如需此工具的使用與指令語法的詳細介紹，請參閱 `cibadmin(8)` [第193頁]。

管理組態變更

`crm_diff` 指令可協助您建立並套用 XML 修補程式。此作業有助於視覺化兩個叢集組態版本之間的變更，或儲存變更以便稍後使用 `cibadmin(8)` [第193頁] 套用變更。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_diff(8)` [第205頁]。

操作 CIB 屬性

`crm_attribute` 指令可讓您查詢並操作節點屬性和要在 CIB 中使用的叢集組態選項。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_attribute(8)` [第202頁]。

驗證叢集組態

`crm_verify` 指令可檢查組態資料庫 (CIB) 的一致性及其他問題。它可檢查檔案是否包含組態或連接正在執行的叢集。它可報告兩類問題。必須先修復錯誤，`Heartbeat` 才能正常工作，管理員負責解決警告問題。`crm_verify` 可協助建立新組態或修改的組態。您可在執行中的叢集中建立 CIB 的本地副本、對其進行編輯，並使用 `crm_verify` 進行驗證，然後使用 `cibadmin` 使新組態生效。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_verify(8)` [第241頁]。

管理資源組態

`crm_resource` 指令可在叢集上執行各種與資源相關的動作。它可讓您修改已設定資源的定義，啟動和停止資源，或在節點間刪除和移轉資源。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_resource(8)` [第217頁]。

管理資源失敗計數

`crm_failcount` 指令會查詢指定節點上每個資源的失敗次數。此工具還可用來重設 `failcount`，允許資源在失敗次數過多的節點上重新執行。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_failcount(8)` [第208頁]。

管理節點的待機狀態

`crm_standby` 指令可操作節點的待機屬性。待機模式下的任何節點都無法再代管資源，且其中的所有資源也都必須移出。對於執行核心更新等維護任務，待機模式十分有用。移除節點的待機屬性，可讓其再次成為叢集的完全使用中成員。如需此工具的使用與指令語法的詳細介紹，請參閱 `crm_standby(8)` [第238頁]。

cibadmin (8)

cibadmin — Provides direct access to the cluster configuration

Synopsis

Allows the configuration, or sections of it, to be queried, modified, replaced and/or deleted.

```
cibadmin (--query|-Q) -[Vrwlsmfbp] [-i xml-object-id|-o
    xml-object-type] [-t t-flag-whatever] [-h hostname]

cibadmin (--create|-C) -[Vrwlsmfbp] [-X xml-string]
    [-x xml- filename] [-t t-flag-whatever] [-h hostname]

cibadmin (--replace|-R) -[Vrwlsmfbp] [-i xml-object-id|
    -o xml-object-type] [-X xml-string] [-x xml-filename] [-t
    t-flag- whatever] [-h hostname]

cibadmin (--update|-U) -[Vrwlsmfbp] [-i xml-object-id|
    -o xml-object-type] [-X xml-string] [-x xml-filename] [-t
    t-flag- whatever] [-h hostname]

cibadmin (--modify|-M) -[Vrwlsmfbp] [-i xml-object-id|
    -o xml-object-type] [-X xml-string] [-x xml-filename] [-t
    t-flag- whatever] [-h hostname]

cibadmin (--delete|-D) -[Vrwlsmfbp] [-i xml-object-id|
    -o xml-object-type] [-t t-flag-whatever] [-h hostname]

cibadmin (--delete_alt|-d) -[Vrwlsmfbp] -o
    xml-object-type [-X xml-string|-x xml-filename]
    [-t t-flag-whatever] [-h hostname]

cibadmin --erase (-E)

cibadmin --bump (-B)

cibadmin --ismaster (-m)

cibadmin --master (-w)

cibadmin --slave (-r)

cibadmin --sync (-S)

cibadmin --help (-?)
```

Description

The `cibadmin` command is the low-level administrative command for manipulating the Heartbeat CIB. Use it to dump all or part of the CIB, update all or part of it, modify all or part of it, delete the entire CIB, or perform miscellaneous CIB administrative operations.

`cibadmin` operates on the XML trees of the CIB, largely without knowledge of the purpose of the updates or queries performed. This means that shortcuts that seem natural to users who understand the meaning of the elements in the XML tree are impossible to use with `cibadmin`. It requires a complete lack of ambiguity and can only deal with valid XML subtrees (tags and elements) for both input and output.

注意

`cibadmin` should always be used in preference to editing the `cib.xml` file by hand—especially if the cluster is active. The cluster goes to great lengths to detect and discourage this practice so that your data is not lost or corrupted.

Options

`--obj_type object-type, -o object-type`

Specify the type of object on which to operate. Valid values are `nodes`, `resources`, `constraints`, `crm_status`, and `status`.

`--verbose, -V`

Turn on debug mode. Additional `-V` options increase the detail and frequency of the output.

`--help, -?`

Obtain a help message from `cibadmin`.

`--xpath PATHSPEC, -A PATHSPEC`

Supply a valid XPath to use instead of an `obj_type`.

Commands

`--bump, -B`

Increase the `epoch` version counter in the CIB. Normally this value is increased automatically by the cluster when a new leader is elected. Manually increasing it can be useful if you want to make an older configuration obsolete (such as one stored on inactive cluster nodes).

`--create, -C`

Create a new CIB from the XML content of the argument.

`--delete, -D`

Delete the first object matching the supplied criteria, for example, `<op id="rsc1_op1" name="monitor"/>`. The tag name and all attributes must match in order for the element to be deleted

`--erase, -E`

Erase the contents of the entire CIB.

`--ismaster, -m`

Print a message indicating whether or not the local instance of the CIB software is the master instance or not. Exits with return code 0 if it is the master instance or 35 if not.

`--modify, -M`

Find the object somewhere in the CIB's XML tree and update it.

`--query, -Q`

Query a portion of the CIB.

`--replace, -R`

Recursively replace an XML object in the CIB.

`--sync, -S`

Force a resync of all nodes with the CIB on the specified host (if `-h` is used) or with the DC (if no `-h` option is used).

XML Data

`--xml-text string, -X string`

Specify an XML tag or fragment on which `crmadmin` should operate. It must be a complete tag or XML fragment.

`--xml-file filename, -x filename`

Specify the XML from a file on which `cibadmin` should operate. It must be a complete tag or an XML fragment.

`--xml_pipe, -p`

Specify that the XML on which `cibadmin` should operate comes from standard input. It must be a complete tag or an XML fragment.

Advanced Options

`--host hostname, -h hostname`

Send command to specified host. Applies to `query` and `sync` commands only.

`--local, -l`

Let a command take effect locally (rarely used, advanced option).

`--no-bcast, -b`

Command will not be broadcast even if it altered the CIB.

重要

Use this option with care to avoid ending up with a divergent cluster.

`--sync-call, -s`

Wait for call to complete before returning.

Examples

To get a copy of the entire active CIB (including status section, etc.) delivered to stdout, issue this command:

```
cibadmin -Q
```


To add an IPaddr2 resource to the *resources* section, first create a file `foo` with the following contents:

```
<primitive id="R_10.10.10.101" class="ocf" type="IPaddr2"
  provider="heartbeat">
  <instance_attributes id="RA_R_10.10.10.101">
    <attributes>
      <nvpair id="R_ip_P_ip" name="ip" value="10.10.10.101"/>
      <nvpair id="R_ip_P_nic" name="nic" value="eth0"/>
    </attributes>
  </instance_attributes>
</primitive>
```

Then issue the following command:

```
cibadmin --obj_type resources -U -x foo
```

To change the IP address of the IPaddr2 resource previously added, issue the command below:

```
cibadmin -M -X '<nvpair id="R_ip_P_ip" name="ip" value="10.10.10.102"/>'
```

注意

This does not change the resource name to match the new IP address. To do that, delete then re-add the resource with a new ID tag.

To stop (disable) the IP address resource added previously, and without removing it, create a file called `bar` with the following content in it:

```
<primitive id="R_10.10.10.101">
  <instance_attributes id="RA_R_10.10.10.101">
    <attributes>
      <nvpair id="stop_R_10.10.10.101" name="target-role" value="Stopped"/>
    </attributes>
  </instance_attributes>
</primitive>
```

Then issue the following command:

```
cibadmin --obj_type resources -U -x bar
```

To restart the IP address resource stopped by the previous step, issue:

```
cibadmin -D -X '<nvpair id="stop_R_10.10.10.101">'
```

To completely remove the IP address resource from the CIB, issue this command:

```
cibadmin -D -X '<primitive id="R_10.10.10.101"/>'
```

To replace the CIB with a new manually-edited version of the CIB, use the following command:

```
cibadmin -R -x $HOME/cib.xml
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk.

See Also

`crm_resource(8)` [第217頁], `crmadmin(8)` [第199頁], `lrmadmin(8)`, `heartbeat(8)`

Caveats

Avoid working on the automatically maintained copy of the CIB on the local disk. Whenever anything in the cluster changes, the CIB is updated. Therefore using an outdated backup copy of the CIB to propagate your configuration changes might result in an inconsistent cluster.

crmadmin (8)

crmadmin — controls the Cluster Resource Manager

Synopsis

```
crmadmin [-V|-q] [-i|-d|-K|-S|-E] node
crmadmin [-V|-q] -N -B
crmadmin [-V|-q] -D
crmadmin -v
crmadmin -?
```

Description

crmadmin was originally designed to control most of the actions of the CRM daemon. However, the largest part of its functionality has been made obsolete by other tools, such as `crm_attribute` and `crm_resource`. Its remaining functionality is mostly related to testing and the status of the crmd process.

警告

Some `crmadmin` options are geared towards testing and cause trouble if used incorrectly. In particular, do not use the `--kill` or `--election` options unless you know exactly what you are doing.

Options

`--help, -?`
Print the help text.

`--version, -v`
Print version details for HA, CRM, and CIB feature set.

`--verbose, -V`
Turn on command debug information.

注意

Increase the level of verbosity by providing additional instances.

`--quiet, -q`

Do not provide any debug information at all and reduce the output to a minimum.

`--bash-export, -B`

Create bash export entries of the form `export uname=uuid`. This applies only to the `crmadmin -N node` command.

注意

The `-B` functionality is rarely useful and may be removed in future versions.

Commands

`--debug_inc node, -i node`

Incrementally increase the CRM daemon's debug level on the specified node. This can also be achieved by sending the USR1 signal to the `crmd` process.

`--debug_dec node, -d node`

Incrementally decrease the CRM daemon's debug level on the specified node. This can also be achieved by sending the USR2 signal to the `crmd` process.

`--kill node, -K node`

Shut down the CRM daemon on the specified node.

警告

Use this with extreme caution. This action should normally only be issued by Heartbeat and may have unintended side effects.

`--status node, -S node`

Query the status of the CRM daemon on the specified node.

The output includes a general health indicator and the internal FSM state of the `crmd` process. This can be helpful when determining what the cluster is doing.

`--election node, -E node`

Initiate an election from the specified node.

警告

Use this with extreme caution. This action is normally initiated internally and may have unintended side effects.

`--dc_lookup, -D`

Query the uname of the current DC.

The location of the DC is only of significance to the `crmd` internally and is rarely useful to administrators except when deciding on which node to examine the logs.

`--nodes, -N`

Query the uname of all member nodes. The results of this query may include nodes in `offline` mode.

注意

The `-i`, `-d`, `-K`, and `-E` options are rarely used and may be removed in future versions.

See Also

`crm_attribute(8)` [第202頁], `crm_resource(8)` [第217頁]

crm_attribute (8)

`crm_attribute` — Allows node attributes and cluster options to be queried, modified and deleted

Synopsis

```
crm_attribute [options]
```

Description

The `crm_attribute` command queries and manipulates node attributes and cluster configuration options that are used in the CIB.

Options

`--help, -?`
Print a help message.

`--verbose, -V`
Turn on debug information.

注意

Increase the level of verbosity by providing additional instances.

`--quiet, -Q`
When doing an attribute query using `-G`, print just the value to stdout. Use this option with `-G`.

`--get-value, -G`
Retrieve, rather than set, the preference.

`--delete-attr, -D`
Delete, rather than set, the attribute.

`--attr-id string, -i string`
For advanced users only. Identifies the id attribute.

`--attr-value string, -v string`
Value to set. This is ignored when used with `-G`.

`--node node_name, -N node_name`
The uname of the node to change

`--set-name string, -s string`
Specify the set of attributes in which to read or write the attribute.

`--attr-name string, -n string`
Specify the attribute to set or query.

`--type string, -t type`
Determine to which section of the CIB the attribute should be set or to which section of the CIB the attribute that is queried belongs. Possible values are `nodes`, `status`, or `crm_config`.

Examples

Query the value of the `location` attribute in the `nodes` section for the host *myhost* in the CIB:

```
crm_attribute -G -t nodes -U myhost -n location
```

Query the value of the `cluster-delay` attribute in the `crm_config` section in the CIB:

```
crm_attribute -G -t crm_config -n cluster-delay
```

Query the value of the `cluster-delay` attribute in the `crm_config` section in the CIB. Print just the value:

```
crm_attribute -G -Q -t crm_config -n cluster-delay
```

Delete the `location` attribute for the host *myhost* from the `nodes` section of the CIB:

```
crm_attribute -D -t nodes -U myhost -n location
```

Add a new attribute called `location` with the value of `office` to the `set` subsection of the `nodes` section in the CIB (settings applied to the host *myhost*):

```
crm_attribute -t nodes -U myhost -s set -n location -v office
```

Change the value of the `location` attribute in the `nodes` section for the *myhost* host:

```
crm_attribute -t nodes -U myhost -n location -v backoffice
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk. Editing this file directly is strongly discouraged.

See Also

`cibadmin(8)` [第193頁]

crm_diff (8)

`crm_diff` — identify changes to the cluster configuration and apply patches to the configuration files

Synopsis

```
crm_diff [-?|-V] [-o filename] [-O string] [-p filename] [-n filename] [-N string]
```

Description

The `crm_diff` command assists in creating and applying XML patches. This can be useful for visualizing the changes between two versions of the cluster configuration or saving changes so they can be applied at a later time using `cibadmin`.

Options

`--help, -?`

Print a help message.

`--original filename, -o filename`

Specify the original file against which to diff or apply patches.

`--new filename, -n filename`

Specify the name of the new file.

`--original-string string, -O string`

Specify the original string against which to diff or apply patches.

`--new-string string, -N string`

Specify the new string.

`--patch filename, -p filename`

Apply a patch to the original XML. Always use with `-o`.

`--cib, -c`

Compare or patch the inputs as a CIB. Always specify the base version with `-o` and provide either the patch file or the second version with `-p` or `-n`, respectively.

`--stdin, -s`

Read the inputs from stdin.

Examples

Use `crm_diff` to determine the differences between various CIB configuration files and to create patches. By means of patches, easily reuse configuration parts without having to use the `cibadmin` command on every single one of them.

- 1 Obtain the two different configuration files by running `cibadmin` on the two cluster setups to compare:

```
cibadmin -Q > cib1.xml
cibadmin -Q > cib2.xml
```

- 2 Determine whether to diff the entire files against each other or compare just a subset of the configurations.

- 3 To print the difference between the files to stdout, use the following command:

```
crm_diff -o cib1.xml -n cib2.xml
```

- 4 To print the difference between the files to a file and create a patch, use the following command:

```
crm_diff -o cib1.xml -n cib2.xml > patch.xml
```

- 5 Apply the patch to the original file:

```
crm_diff -o cib1.xml -p patch.xml
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk. Editing this file directly is strongly discouraged.

See Also

cibadmin(8) [第193頁]

crm_failcount (8)

crm_failcount — Manage the counter recording each resource's failures

Synopsis

```
crm_failcount [-?|-V] -D -u|-U node -r resource
crm_failcount [-?|-V] -G -u|-U node -r resource
crm_failcount [-?|-V] -v string -u|-U node -r resource
```

Description

Heartbeat implements a sophisticated method to compute and force failover of a resource to another node in case that resource tends to fail on the current node. A resource carries a `resource-stickiness` attribute to determine how much it prefers to run on a certain node. It also carries a `migration-threshold` that determines the threshold at which the resource should failover to another node.

The `failcount` attribute is added to the resource and increased on resource monitoring failure. The value of `failcount` multiplied by the value of `migration-threshold` determines the *failover score* of this resource. If this number exceeds the preference set for this resource, the resource is moved to another node and not run again on the original node until the failure count is reset.

The `crm_failcount` command queries the number of failures per resource on a given node. This tool can also be used to reset the failcount, allowing the resource to run again on nodes where it had previously failed too many times.

Options

```
--help, -?
    Print a help message.

--verbose, -V
    Turn on debug information.
```

注意

Increase the level of verbosity by providing additional instances.

`--quiet, -Q`

When doing an attribute query using `-G`, print just the value to stdout. Use this option with `-G`.

`--get-value, -G`

Retrieve rather than set the preference.

`--delete-attr, -D`

Specify the attribute to delete.

`--attr-id string, -i string`

For advanced users only. Identifies the id attribute.

`--attr-value string, -v string`

Specify the value to use. This option is ignored when used with `-G`.

`--node node_uname, -U node_uname`

Specify the uname of the node to change.

`--resource-id resource name, -r resource name`

Specify the name of the resource on which to operate.

Examples

Reset the failcount for the resource `myrsc` on the node `node1`:

```
crm_failcount -D -U node1 -r my_rsc
```

Query the current failcount for the resource `myrsc` on the node `node1`:

```
crm_failcount -G -U node1 -r my_rsc
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk. Editing this file directly is strongly discouraged.

See Also

`crm_attribute(8)` [第202頁], `cibadmin(8)` [第193頁], and the Linux High Availability FAQ Web site [http://www.linux-ha.org/v2/faq/forced_failover]

crm_master (8)

`crm_master` — Manage a master/slave resource's preference for being promoted on a given node

Synopsis

```
crm_master [-V|-Q] -D [-l lifetime]  
crm_master [-V|-Q] -G [-l lifetime]  
crm_master [-V|-Q] -v string [-l string]
```

Description

`crm_master` is called from inside the resource agent scripts to determine which resource instance should be promoted to master mode. It should never be used from the command line and is just a helper utility for the resource agents. RAs use `crm_master` to promote a particular instance to master mode or to remove this preference from it. By assigning a lifetime, determine whether this setting should survive a reboot of the node (set lifetime to `forever`) or whether it should not survive a reboot (set lifetime to `reboot`).

A resource agent needs to determine on which resource `crm_master` should operate. These queries must be handled inside the resource agent script. The actual calls of `crm_master` follow a syntax similar to those of the `crm_attribute` command.

Options

`--help, -?`
Print a help message.

`--verbose, -V`
Turn on debug information.

注意

Increase the level of verbosity by providing additional instances.

`--quiet, -Q`

When doing an attribute query using `-G`, print just the value to stdout. Use this option with `-G`.

`--get-value, -G`

Retrieve rather than set the preference to be promoted.

`--delete-attr, -D`

Delete rather than set the attribute.

`--attr-id string, -i string`

For advanced users only. Identifies the id attribute.

`--attr-value string, -v string`

Value to set. This is ignored when used with `-G`.

`--lifetime string, -l string`

Specify how long the preference lasts. Possible values are `reboot` or `forever`.

Environment Variables

`OCF_RESOURCE_INSTANCE`—the name of the resource instance

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk.

See Also

`cibadmin(8)` [第193頁], `crm_attribute(8)` [第202頁]

crm_mon (8)

crm_mon — monitor the cluster's status

Synopsis

```
crm_mon [-V] -d -pfilename -h filename
crm_mon [-V] [-l|-n|-r] -h filename
crm_mon [-V] [-n|-r] -X filename
crm_mon [-V] [-n|-r] -c|-l
crm_mon [-V] -i interval
crm_mon -?
```

Description

The `crm_mon` command allows you to monitor your cluster's status and configuration. Its output includes the number of nodes, uname, uuid, status, the resources configured in your cluster, and the current status of each. The output of `crm_mon` can be displayed at the console or printed into an HTML file. When provided with a cluster configuration file without the status section, `crm_mon` creates an overview of nodes and resources as specified in the file.

Options

`--help, -?`

Provide help.

`--verbose, -V`

Increase the debug output.

`--interval seconds, -i seconds`

Determine the update frequency. If `-i` is not specified, the default of 15 seconds is assumed.

`--group-by-node, -n`
Group resources by node.

`--inactive, -r`
Display inactive resources.

`--simple-status, -s`
Display the cluster status once as a simple one line output (suitable for nagios).

`--one-shot, -l`
Display the cluster status once on the console then exit (does not use ncurses).

`--as-html filename, -h filename`
Write the cluster's status to the specified file.

`--web-cgi, -w`
Web mode with output suitable for CGI.

`--daemonize, -d`
Run in the background as a daemon.

`--pid-file filename, -p filename`
Specify the daemon's pid file.

Examples

Display your cluster's status and get an updated listing every 15 seconds:

```
crm_mon
```

Display your cluster's status and get an updated listing after an interval specified by `-i`. If `-i` is not given, the default refresh interval of 15 seconds is assumed:

```
crm_mon -i interval[s]
```

Display your cluster's status on the console:

```
crm_mon -c
```

Display your cluster's status on the console just once then exit:

```
crm_mon -l
```

Display your cluster's status and group resources by node:

```
crm_mon -n
```

Display your cluster's status, group resources by node, and include inactive resources in the list:

```
crm_mon -n -r
```

Write your cluster's status to an HTML file:

```
crm_mon -h filename
```

Run `crm_mon` as a daemon in the background, specify the daemon's pid file for easier control of the daemon process, and create HTML output. This option allows you to constantly create HTML output that can be easily processed by other monitoring applications:

```
crm_mon -d -p filename -h filename
```

Display the cluster configuration laid out in an existing cluster configuration file (*filename*), group the resources by node, and include inactive resources. This command can be used for dry runs of a cluster configuration before rolling it out to a live cluster.

```
crm_mon -r -n -X filename
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk. Editing this file directly is strongly discouraged.

crm_node (8)

crm_node — Lists the members of a cluster

Synopsis

```
crm_node [-V] [-p|-e|-q]
```

Description

Lists the members of a cluster.

Options

- V
be verbose
- partition, -p
print the members of this partition
- epoch, -e
print the epoch this node joined the partition
- quorum, -q
print a 1 if our partition has quorum

crm_resource (8)

crm_resource — Perform tasks related to cluster resources

Synopsis

```
crm_resource [-?|-V|-S] -L|-Q|-W|-D|-C|-P|-p [options]
```

Description

The `crm_resource` command performs various resource-related actions on the cluster. It can modify the definition of configured resources, start and stop resources, and delete and migrate resources between nodes.

`--help, -?`

Print the help message.

`--verbose, -V`

Turn on debug information.

注意

Increase the level of verbosity by providing additional instances.

`--quiet, -Q`

Print only the value on stdout (for use with `-W`).

Commands

`--list, -L`

List all resources.

`--query-xml, -x`

Query a resource.

Requires: `-r`

`--locate, -W`
Locate a resource.

Requires: `-r`

`--migrate, -M`
Migrate a resource from its current location. Use `-N` to specify a destination.

If `-N` is not specified, the resource is forced to move by creating a rule for the current location and a score of `-INFINITY`.

注意

This prevents the resource from running on this node until the constraint is removed with `-U`.

Requires: `-r`, Optional: `-N, -f`

`--un-migrate, -U`
Remove all constraints created by `-M`

Requires: `-r`

`--delete, -D`
Delete a resource from the CIB.

Requires: `-r, -t`

`--cleanup, -C`
Delete a resource from the LRM.

Requires: `-r`. Optional: `-H`

`--reprobe, -P`
Recheck for resources started outside the CRM.

Optional: `-H`

`--refresh, -R`
Refresh the CIB from the LRM.

Optional: -H

`--set-parameter string, -p string`

Set the named parameter for a resource.

Requires: -r, -v. Optional: -i, -s, and --meta

`--get-parameter string, -g string`

Get the named parameter for a resource.

Requires: -r. Optional: -i, -s, and --meta

`--delete-parameter string, -d string`

Delete the named parameter for a resource.

Requires: -r. Optional: -i, and --meta

`--list-operations string, -O string`

List the active resource operations. Optionally filtered by resource, node, or both.

Optional: -N, -r

`--list-all-operations string, -o string`

List all resource operations. Optionally filtered by resource, node, or both. Optional:

-N, -r

Options

`--resource string, -r string`

Specify the resource ID.

`--resource-type string, -t string`

Specify the resource type (primitive, clone, group, etc.).

`--property-value string, -v string`

Specify the property value.

`--node string, -N string`

Specify the hostname.

`--meta`

Modify a resource's configuration option rather than one which is passed to the resource agent script. For use with `-p`, `-g` and `-d`.

`--lifetime string, -u string`

Lifespan of migration constraints.

`--force, -f`

Force the resource to move by creating a rule for the current location and a score of `-INFINITY`

This should be used if the resource's stickiness and constraint scores total more than `INFINITY` (currently 100,000).

注意

This prevents the resource from running on this node until the constraint is removed with `-U`.

`-s string`

(Advanced Use Only) Specify the ID of the `instance_attributes` object to change.

`-i string`

(Advanced Use Only) Specify the ID of the `nvpair` object to change or delete.

Examples

Listing all resources:

```
crm_resource -L
```

Checking where a resource is running (and if it is):

```
crm_resource -W -r my_first_ip
```

If the `my_first_ip` resource is running, the output of this command reveals the node on which it is running. If it is not running, the output shows this.

Start or stop a resource:

```
crm_resource -r my_first_ip -p target_role -v started
crm_resource -r my_first_ip -p target_role -v stopped
```

Query the definition of a resource:

```
crm_resource -Q -r my_first_ip
```

Migrate a resource away from its current location:

```
crm_resource -M -r my_first_ip
```

Migrate a resource to a specific location:

```
crm_resource -M -r my_first_ip -H c001n02
```

Allow a resource to return to its normal location:

```
crm_resource -U -r my_first_ip
```

注意

The values of `resource_stickiness` and `default_resource_stickiness` may mean that it does not move back. In such cases, you should use `-M` to move it back before running this command.

Delete a resource from the CRM:

```
crm_resource -D -r my_first_ip -t primitive
```

Delete a resource group from the CRM:

```
crm_resource -D -r my_first_group -t group
```

Disable resource management for a resource in the CRM:

```
crm_resource -p is-managed -r my_first_ip -t primitive -v off
```

Enable resource management for a resource in the CRM:

```
crm_resource -p is-managed -r my_first_ip -t primitive -v on
```

Reset a failed resource after having been manually cleaned up:

```
crm_resource -C -H c001n02 -r my_first_ip
```

Recheck all nodes for resources started outside the CRM:

```
crm_resource -P
```

Recheck one node for resources started outside the CRM:

```
crm_resource -P -H c001n02
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk.
Editing this file directly is strongly discouraged.

See Also

`cibadmin(8)` [第193頁], `crmadmin(8)` [第199頁], `lrmadmin(8)`, `heartbeat(8)`

crm_shadow (8)

crm_shadow — Perform Configuration Changes in a Sandbox Before Updating The Live Cluster

Synopsis

```
crm_shadow [-V] [-p|-e|-q]
```

Description

Sets up an environment in which configuration tools (`cibadmin`, `crm_resource`, etc) work offline instead of against a live cluster, allowing changes to be previewed and tested for side-effects.

Options

- `--verbose, -V`
turn on debug info. additional instance increase verbosity
- `--which, -w`
indicate the active shadow copy
- `--display, -p`
display the contents of the shadow copy
- `--diff, -d`
display the changes in the shadow copy
- `--create-empty, -eNAME`
create the named shadow copy with an empty cluster configuration
- `--create, -cNAME`
create the named shadow copy of the active cluster configuration

`--reset, -rNAME`
 recreate the named shadow copy from the active cluster configuration

`--commit, -cNAME`
 upload the contents of the named shadow copy to the cluster

`--delete, -dNAME`
 delete the contents of the named shadow copy

`--edit, -eNAME`
 Edit the contents of the named shadow copy with your favorite editor

`--batch, -b`
 do not spawn a new shell

`--force, -f`
 do not spawn a new shell

`--switch, -s`
 switch to the named shadow copy

Internal Commands

To work with a shadow configuration, you need to create one first:

```
crm_shadow --create-empty YOUR_NAME
```

It gives you an internal shell like the one from the `crm` tool. Use `help` to get an overview of all internal commands, or `help subcommand` for a specific command.

表格 18.1 *Overview of Internal Commands*

Command	Syntax/Description
<code>alias</code>	<pre>alias [-p] [name[=value] ...]</pre> <p><code>alias</code> with no arguments or with the <code>-p</code> option prints the list of aliases in the form <code>alias NAME=VALUE</code> on standard output. Otherwise, an alias is defined for each <code>NAME</code> whose <code>VALUE</code> is given. A trailing space in <code>VALUE</code> causes the next word to be checked for alias</p>

Command	Syntax/Description
	substitution when the alias is expanded. Alias returns true unless a NAME is given for which no alias has been defined.
bg	bg [JOB_SPEC ...] Place each JOB_SPEC in the background, as if it had been started with &. If JOB_SPEC is not present, the shell's notion of the current job is used.
bind	bind [-lpvsPVS] [-m keymap] [-f filename] [-q name] [-u name] [-r keyseq] [-x keyseq:shell-command] [keyseq:readline-function or readline-command] Bind a key sequence to a Readline function or a macro, or set a Readline variable. The non-option argument syntax is equivalent to that found in ~/.inputrc, but must be passed as a single argument: bind "\C-x\C-r": re-read-init-file.
break	break [N] Exit from within a for, while or until loop. If N is specified, break N levels.
builtin	builtin [shell-builtin [arg ...]] Run a shell builtin. This is useful when you wish to rename a shell builtin to be a function, but need the functionality of the builtin within the function itself.
caller	caller [EXPR] Returns the context of the current subroutine call. Without EXPR, returns \$line \$filename. With EXPR, returns \$line \$subroutine \$filename; this extra information can be used to provide a stack trace.
case	case WORD in [PATTERN [PATTERN] [COMMANDS;;] ... esac

Command	Syntax/Description
	Selectively execute <i>COMMANDS</i> based upon <i>WORD</i> matching <i>PATTERN</i> . The ' ' is used to separate multiple patterns.
cd	cd [-L -P] [dir] Change the current directory to DIR.
command	command [-pVv] command [arg ...] Runs <i>COMMAND</i> with <i>ARGS</i> ignoring shell functions. If you have a shell function called 'ls', and you wish to call the command 'ls', you can say "command ls". If the -p option is given, a default value is used for PATH that is guaranteed to find all of the standard utilities. If the -V or -v option is given, a string is printed describing <i>COMMAND</i> . The -V option produces a more verbose description.
compgen	compgen [-abcdefgjkusv] [-o option] [-A action] [-G globpat] [-W wordlist] [-P prefix] [-S suffix] [-X filterpat] [-F function] [-C command] [WORD] Display the possible completions depending on the options. Intended to be used from within a shell function generating possible completions. If the optional <i>WORD</i> argument is supplied, matches against <i>WORD</i> are generated.
complete	complete [-abcdefgjkusv] [-pr] [-o option] [-A action] [-G globpat] [-W wordlist] [-P prefix] [-S suffix] [-X filterpat] [-F function] [-C command] [name ...] For each <i>NAME</i> , specify how arguments are to be completed. If the -p option is supplied, or if no options are supplied, existing completion specifications are printed in a way that allows them to be reused as input. The -r option removes a completion specification for each <i>NAME</i> , or, if no <i>NAMES</i> are supplied, all completion specifications.
continue	continue [N]

Command	Syntax/Description
	Resume the next iteration of the enclosing FOR, WHILE or UNTIL loop. If <i>N</i> is specified, resume at the <i>N</i> -th enclosing loop.
<code>declare</code>	<code>declare [-afFirtx] [-p] [name[=value] ...]</code> Declare variables and/or give them attributes. If no <i>NAMES</i> are given, then display the values of variables instead. The <code>-p</code> option will display the attributes and values of each <i>NAME</i> .
<code>dirs</code>	<code>dirs [-clpv] [+N] [-N]</code> Display the list of currently remembered directories. Directories find their way onto the list with the <code>pushd</code> command; you can get back up through the list with the <code>popd</code> command.
<code>disown</code>	<code>disown [-h] [-ar] [JOBSPEC ...]</code> By default, removes each <i>JOBSPEC</i> argument from the table of active jobs. If the <code>-h</code> option is given, the job is not removed from the table, but is marked so that SIGHUP is not sent to the job if the shell receives a SIGHUP. The <code>-a</code> option, when <i>JOBSPEC</i> is not supplied, means to remove all jobs from the job table; the <code>-r</code> option means to remove only running jobs.
<code>echo</code>	<code>echo [-neE] [arg ...]</code> Output the ARGs. If <code>-n</code> is specified, the trailing newline is suppressed. If the <code>-e</code> option is given, interpretation of the following backslash-escaped characters is turned on: \a (alert, bell) \b (backspace) \c (suppress trailing newline) \E (escape character) \f (form feed) \n (new line)

Command	Syntax/Description
	<p> <code>\r</code> (carriage return) <code>\t</code> (horizontal tab) <code>\v</code> (vertical tab) <code>\\</code> (backslash) <code>\0nnn</code> (the character whose ASCII code is NNN (octal). NNN can be 0 to 3 octal digits) </p> <p>You can turn off the interpretation of the above characters with the <code>-E</code> option.</p>
<code>enable</code>	<p> <code>enable [-pnds] [-a] [-f filename] [name...]</code> </p> <p>Enable and disable builtin shell commands. This allows you to use a disk command which has the same name as a shell builtin without specifying a full pathname. If <code>-n</code> is used, the <i>NAMES</i> become disabled; otherwise <i>NAMES</i> are enabled. For example, to use the <code>test</code> found in <code>\$PATH</code> instead of the shell builtin version, type <code>enable -n test</code>. On systems supporting dynamic loading, the <code>-f</code> option may be used to load new builtins from the shared object <i>FILENAME</i>. The <code>-d</code> option will delete a builtin previously loaded with <code>-f</code>. If no non-option names are given, or the <code>-p</code> option is supplied, a list of builtins is printed. The <code>-a</code> option means to print every builtin with an indication of whether or not it is enabled. The <code>-s</code> option restricts the output to the POSIX.2 'special' builtins. The <code>-n</code> option displays a list of all disabled builtins.</p>
<code>eval</code>	<p> <code>eval [ARG ...]</code> </p> <p>Read <i>ARGS</i> as input to the shell and execute the resulting command(s).</p>
<code>exec</code>	<p> <code>exec [-cl] [-a name] file [redirection ...]</code> </p> <p>Exec <i>FILE</i>, replacing this shell with the specified program. If <i>FILE</i> is not specified, the redirections take effect in this shell. If the first argument is <code>-l</code>, then place a dash in the zeroth arg passed to <i>FILE</i>, as <code>login</code> does. If the <code>-c</code> option is supplied, <i>FILE</i> is executed with a null environment. The <code>-a</code> option means to make <code>set argv[0]</code> of the</p>

Command	Syntax/Description
	executed process to <i>NAME</i> . If the file cannot be executed and the shell is not interactive, then the shell exits, unless the shell option <code>execfail</code> is set.
<code>exit</code>	<code>exit [N]</code> Exit the shell with a status of <i>N</i> . If <i>N</i> is omitted, the exit status is that of the last command executed.
<code>export</code>	<code>export [-nf] [NAME[=value] ...]</code> <code>export -p</code> <i>NAMES</i> are marked for automatic export to the environment of subsequently executed commands. If the <code>-f</code> option is given, the <i>NAMES</i> refer to functions. If no <i>NAMES</i> are given, or if <code>-p</code> is given, a list of all names that are exported in this shell is printed. An argument of <code>-n</code> says to remove the export property from subsequent <i>NAMES</i> . An argument of <code>--</code> disables further option processing.
<code>false</code>	<code>false</code> Return an unsuccessful result.
<code>fc</code>	<code>fc [-e ename] [-nlr] [FIRST] [LAST]</code> <code>fc -s [pat=rep] [cmd]</code> <code>fc</code> is used to list or edit and re-execute commands from the history list. <i>FIRST</i> and <i>LAST</i> can be numbers specifying the range, or <i>FIRST</i> can be a string, which means the most recent command beginning with that string.
<code>fg</code>	<code>fg [JOB_SPEC]</code> Place <i>JOB_SPEC</i> in the foreground, and make it the current job. If <i>JOB_SPEC</i> is not present, the shell's notion of the current job is used.
<code>for</code>	<code>for NAME [in WORDS ... ;] do COMMANDS; done</code>

Command	Syntax/Description
	<p>The <code>for</code> loop executes a sequence of commands for each member in a list of items. If <code>in WORDS ... ;</code> is not present, then <code>in "\$@"</code> is assumed. For each element in <i>WORDS</i>, <i>NAME</i> is set to that element, and the <i>COMMANDS</i> are executed.</p>
<code>function</code>	<pre>function NAME { COMMANDS ; } function NAME () { COMMANDS ; }</pre> <p>Create a simple command invoked by <i>NAME</i> which runs <i>COMMANDS</i>. Arguments on the command line along with <i>NAME</i> are passed to the function as <code>\$0 .. \$n</code>.</p>
<code>getopts</code>	<pre>getopts OPTSTRING NAME [arg]</pre> <p>Getopts is used by shell procedures to parse positional parameters.</p>
<code>hash</code>	<pre>hash [-lr] [-p PATHNAME] [-dt] [NAME...]</pre> <p>For each <i>NAME</i>, the full pathname of the command is determined and remembered. If the <code>-p</code> option is supplied, <i>PATHNAME</i> is used as the full pathname of <i>NAME</i>, and no path search is performed. The <code>-r</code> option causes the shell to forget all remembered locations. The <code>-d</code> option causes the shell to forget the remembered location of each <i>NAME</i>. If the <code>-t</code> option is supplied the full pathname to which each <i>NAME</i> corresponds is printed. If multiple <i>NAME</i> arguments are supplied with <code>-t</code>, the <i>NAME</i> is printed before the hashed full pathname. The <code>-l</code> option causes output to be displayed in a format that may be reused as input. If no arguments are given, information about remembered commands is displayed.</p>
<code>history</code>	<pre>history [-c] [-d OFFSET] [n] history -ps arg [arg...] history -awrm [filename]</pre> <p>Display the history list with line numbers. Lines listed with with a <code>*</code> have been modified. Argument of <i>N</i> says to list only the last <i>N</i> lines. The <code>-c</code> option causes the history list to be cleared by deleting all of</p>

Command	Syntax/Description
	<p>the entries. The <code>-d</code> option deletes the history entry at offset <i>OFFSET</i>. The <code>-w</code> option writes out the current history to the history file; <code>-r</code> means to read the file and append the contents to the history list instead. <code>-a</code> means to append history lines from this session to the history file. Argument <code>-n</code> means to read all history lines not already read from the history file and append them to the history list.</p>
<code>jobs</code>	<pre>jobs [-lnprs] [JOBSPEC ...] job -x COMMAND [ARGS]</pre> <p>Lists the active jobs. The <code>-l</code> option lists process id's in addition to the normal information; the <code>-p</code> option lists process id's only. If <code>-n</code> is given, only processes that have changed status since the last notification are printed. <i>JOBSPEC</i> restricts output to that job. The <code>-r</code> and <code>-s</code> options restrict output to running and stopped jobs only, respectively. Without options, the status of all active jobs is printed. If <code>-x</code> is given, <i>COMMAND</i> is run after all job specifications that appear in <i>ARGS</i> have been replaced with the process ID of that job's process group leader.</p>
<code>kill</code>	<pre>kill [-s sigspec -n signum -sigspec] pid JOBSPEC ... kill -l [sigspec]</pre> <p>Send the processes named by PID (or <i>JOBSPEC</i>) the signal <i>SIGSPEC</i>. If <i>SIGSPEC</i> is not present, then <i>SIGTERM</i> is assumed. An argument of <code>-l</code> lists the signal names; if arguments follow <code>-l</code> they are assumed to be signal numbers for which names should be listed. Kill is a shell builtin for two reasons: it allows job IDs to be used instead of process IDs, and, if you have reached the limit on processes that you can create, you don't have to start a process to kill another one.</p>
<code>let</code>	<pre>let ARG [ARG ...]</pre> <p>Each <i>ARG</i> is a mathematical expression to be evaluated. Evaluation is done in fixed-width integers with no check for overflow, though division by 0 is trapped and flagged as an error. The following list of operators is grouped into levels of equal-precedence operators. The levels are listed in order of decreasing precedence.</p>

Command	Syntax/Description
<code>local</code>	<pre>local NAME[=VALUE] ...</pre> <p>Create a local variable called <i>NAME</i>, and give it <i>VALUE</i>. <code>local</code> can only be used within a function; it makes the variable <i>NAME</i> have a visible scope restricted to that function and its children.</p>
<code>logout</code>	<pre>logout</pre> <p>Logout of a login shell.</p>
<code>popd</code>	<pre>popd [+N -N] [-n]</pre> <p>Removes entries from the directory stack. With no arguments, removes the top directory from the stack, and <code>cd</code>'s to the new top directory.</p>
<code>printf</code>	<pre>printf [-v var] format [ARGUMENTS]</pre> <p><code>printf</code> formats and prints <i>ARGUMENTS</i> under control of the <i>FORMAT</i>. <i>FORMAT</i> is a character string which contains three types of objects: plain characters, which are simply copied to standard output, character escape sequences which are converted and copied to the standard output, and format specifications, each of which causes printing of the next successive argument. In addition to the standard <code>printf(1)</code> formats, <code>%b</code> means to expand backslash escape sequences in the corresponding argument, and <code>%q</code> means to quote the argument in a way that can be reused as shell input. If the <code>-v</code> option is supplied, the output is placed into the value of the shell variable <i>VAR</i> rather than being sent to the standard output.</p>
<code>pushd</code>	<pre>pushd [dir +N -N] [-n]</pre> <p>Adds a directory to the top of the directory stack, or rotates the stack, making the new top of the stack the current working directory. With no arguments, exchanges the top two directories.</p>
<code>pwd</code>	<pre>pwd [-LP]</pre>

Command	Syntax/Description
	<p>Print the current working directory. With the <code>-P</code> option, <code>pwd</code> prints the physical directory, without any symbolic links; the <code>-L</code> option makes <code>pwd</code> follow symbolic links.</p>
<code>read</code>	<pre>read [-ers] [-u fd] [-t timeout] [-p prompt] [-a array] [-n nchars] [-d delim] [NAME ...]</pre> <p>The given <i>NAMES</i> are marked readonly and the values of these <i>NAMES</i> may not be changed by subsequent assignment. If the <code>-f</code> option is given, then functions corresponding to the <i>NAMES</i> are so marked. If no arguments are given, or if <code>-p</code> is given, a list of all readonly names is printed. The <code>-a</code> option means to treat each <i>NAME</i> as an array variable. An argument of <code>--</code> disables further option processing.</p>
<code>readonly</code>	<pre>readonly [-af] [NAME[=VALUE] ...] readonly -p</pre> <p>The given <i>NAMES</i> are marked readonly and the values of these <i>NAMES</i> may not be changed by subsequent assignment. If the <code>-f</code> option is given, then functions corresponding to the <i>NAMES</i> are so marked. If no arguments are given, or if <code>-p</code> is given, a list of all readonly names is printed. The <code>-a</code> option means to treat each <i>NAME</i> as an array variable. An argument of <code>--</code> disables further option processing.</p>
<code>return</code>	<pre>return [N]</pre> <p>Causes a function to exit with the return value specified by <i>N</i>. If <i>N</i> is omitted, the return status is that of the last command.</p>
<code>select</code>	<pre>select NAME [in WORDS ... ;] do COMMANDS; done</pre> <p>The <i>WORDS</i> are expanded, generating a list of words. The set of expanded words is printed on the standard error, each preceded by a number. If <code>in WORDS</code> is not present, <code>in "\$@"</code> is assumed. The PS3 prompt is then displayed and a line read from the standard input. If the line consists of the number corresponding to one of the displayed words, then <i>NAME</i> is set to that word. If the line is empty, <i>WORDS</i> and</p>

Command	Syntax/Description
	the prompt are redisplayed. If EOF is read, the command completes. Any other value read causes <i>NAME</i> to be set to null. The line read is saved in the variable <i>REPLY</i> . <i>COMMANDS</i> are executed after each selection until a break command is executed.
set	<pre>set [--abefhkmnptuvxBCHP] [-o OPTION] [ARG...]</pre> <p>Sets internal shell options.</p>
shift	<pre>shift [n]</pre> <p>The positional parameters from $\\$N+1$. . . are renamed to $\\$1$. . . If <i>N</i> is not given, it is assumed to be 1.</p>
shopt	<pre>shopt [-pqsu] [-o long-option] OPTNAME [OPTNAME...]</pre> <p>Toggle the values of variables controlling optional behavior. The <i>-s</i> flag means to enable (set) each <i>OPTNAME</i>; the <i>-u</i> flag unsets each <i>OPTNAME</i>. The <i>-q</i> flag suppresses output; the exit status indicates whether each <i>OPTNAME</i> is set or unset. The <i>-o</i> option restricts the <i>OPTNAME</i>s to those defined for use with <code>set -o</code>. With no options, or with the <i>-p</i> option, a list of all settable options is displayed, with an indication of whether or not each is set.</p>
source	<pre>source FILENAME [ARGS]</pre> <p>Read and execute commands from <i>FILENAME</i> and return. The pathnames in $\\$PATH$ are used to find the directory containing <i>FILENAME</i>. If any <i>ARGS</i> are supplied, they become the positional parameters when <i>FILENAME</i> is executed.</p>
suspend	<pre>suspend [-f]</pre> <p>Suspend the execution of this shell until it receives a SIGCONT signal. The <i>-f</i> if specified says not to complain about this being a login shell if it is; just suspend anyway.</p>

Command	Syntax/Description
<code>test</code>	<pre>test [expr]</pre> <p>Exits with a status of 0 (true) or 1 (false) depending on the evaluation of <i>EXPR</i>. Expressions may be unary or binary. Unary expressions are often used to examine the status of a file. There are string operators as well, and numeric comparison operators.</p>
<code>time</code>	<pre>time [-p] PIPELINE</pre> <p>Execute <i>PIPELINE</i> and print a summary of the real time, user CPU time, and system CPU time spent executing <i>PIPELINE</i> when it terminates. The return status is the return status of <i>PIPELINE</i>. The <code>-p</code> option prints the timing summary in a slightly different format. This uses the value of the <code>TIMEFORMAT</code> variable as the output format.</p>
<code>times</code>	<pre>times</pre> <p>Print the accumulated user and system times for processes run from the shell.</p>
<code>trap</code>	<pre>trap [-lp] [ARG SIGNAL_SPEC ...]</pre> <p>The command <i>ARG</i> is to be read and executed when the shell receives signal(s) <i>SIGNAL_SPEC</i>. If <i>ARG</i> is absent (and a single <i>SIGNAL_SPEC</i> is supplied) or <code>-</code>, each specified signal is reset to its original value. If <i>ARG</i> is the null string each <i>SIGNAL_SPEC</i> is ignored by the shell and by the commands it invokes. If a <i>SIGNAL_SPEC</i> is <code>EXIT</code> (0) the command <i>ARG</i> is executed on exit from the shell. If a <i>SIGNAL_SPEC</i> is <code>DEBUG</code>, <i>ARG</i> is executed after every simple command. If the <code>-p</code> option is supplied then the trap commands associated with each <i>SIGNAL_SPEC</i> are displayed. If no arguments are supplied or if only <code>-p</code> is given, trap prints the list of commands associated with each signal. Each <i>SIGNAL_SPEC</i> is either a signal name in <code>signal.h</code> or a signal number. Signal names are case insensitive and the <code>SIG</code> prefix is optional. <code>trap -l</code> prints a list of signal names and their corresponding numbers. Note that a signal can be sent to the shell with <code>kill -signal \$\$</code>.</p>

Command	Syntax/Description
true	<pre>true</pre> <p>Return a successful result.</p>
type	<pre>type [-afptP] NAME [NAME ...]</pre> <p>Obsolete, see declare.</p>
typeset	<pre>typeset [-afFirtx] [-p] name[=value]</pre> <p>Obsolete, see declare.</p>
ulimit	<pre>ulimit [-SHacdfilmpqstuvx] [limit]</pre> <p>Ulimit provides control over the resources available to processes started by the shell, on systems that allow such control.</p>
umask	<pre>umask [-p] [-S] [MODE]</pre> <p>The user file-creation mask is set to <i>MODE</i>. If <i>MODE</i> is omitted, or if <i>-S</i> is supplied, the current value of the mask is printed. The <i>-S</i> option makes the output symbolic; otherwise an octal number is output. If <i>-p</i> is supplied, and <i>MODE</i> is omitted, the output is in a form that may be used as input. If <i>MODE</i> begins with a digit, it is interpreted as an octal number, otherwise it is a symbolic mode string like that accepted by <code>chmod(1)</code>.</p>
unalias	<pre>unalias [-a] NAME [NAME ...]</pre> <p>Remove <i>NAMES</i> from the list of defined aliases. If the <i>-a</i> option is given, then remove all alias definitions.</p>
unset	<pre>unset [-f] [-v] [NAME ...]</pre> <p>For each <i>NAME</i>, remove the corresponding variable or function. Given the <i>-v</i>, unset will only act on variables. Given the <i>-f</i> flag, unset will only act on functions. With neither flag, unset first tries to unset a variable. If that fails, it then tries to unset a function. Some variables cannot be unset; also see <code>readonly</code>.</p>

Command	Syntax/Description
<code>until</code>	<pre>until COMMANDS; do COMMANDS; done</pre> <p>Expand and execute <i>COMMANDS</i> as long as the final command in the <code>until</code> <i>COMMANDS</i> has an exit status which is not zero.</p>
<code>wait</code>	<pre>wait [N]</pre> <p>Wait for the specified process and report its termination status. If <i>N</i> is not given, all currently active child processes are waited for, and the return code is zero. <i>N</i> may be a process ID or a job specification; if a job spec is given, all processes in the job's pipeline are waited for.</p>
<code>while</code>	<pre>while COMMANDS; do COMMANDS; done</pre> <p>Expand and execute <i>COMMANDS</i> as long as the final command in the <code>while</code> <i>COMMANDS</i> has an exit status of zero.</p>

crm_standby (8)

`crm_standby` — manipulate a node's standby attribute to determine whether resources can be run on this node

Synopsis

```
crm_standby [-?|-V] -D -u|-U node -r resource
crm_standby [-?|-V] -G -u|-U node -r resource
crm_standby [-?|-V] -v string -u|-U node -r resource [-l string]
```

Description

The `crm_standby` command manipulates a node's standby attribute. Any node in standby mode is no longer eligible to host resources and any resources that are there must be moved. Standby mode can be useful for performing maintenance tasks, such as kernel updates. Remove the standby attribute from the node when it needs to become a fully active member of the cluster again.

By assigning a lifetime to the `standby` attribute, determine whether the standby setting should survive a reboot of the node (set lifetime to `forever`) or should be reset with reboot (set lifetime to `reboot`). Alternatively, remove the `standby` attribute and bring the node back from standby manually.

Options

`--help, -?`

Print a help message.

`--verbose, -V`

Turn on debug information.

注意

Increase the level of verbosity by providing additional instances.

`--quiet, -Q`

When doing an attribute query using `-G`, print just the value to stdout. Use this option with `-G`.

`--get-value, -G`

Retrieve rather than set the preference.

`--delete-attr, -D`

Specify the attribute to delete.

`--attr-value string, -v string`

Specify the value to use. This option is ignored when used with `-G`.

`--attr-id string, -i string`

For advanced users only. Identifies the id attribute..

`--node node_undef, -u node_undef`

Specify the uname of the node to change.

`--lifetime string, -l string`

Determine how long this preference lasts. Possible values are `reboot` or `forever`.

注意

If a `forever` value exists, it is always used by the CRM instead of any `reboot` value.

Examples

Have a local node go to standby:

```
crm_standby -v true
```

Have a node (`node1`) go to standby:

```
crm_standby -v true -U node1
```

Query the standby status of a node:

```
crm_standby -G -U node1
```

Remove the standby property from a node:

```
crm_standby -D -U node1
```

Have a node go to standby for an indefinite period of time:

```
crm_standby -v true -l forever -U node1
```

Have a node go to standby until the next reboot of this node:

```
crm_standby -v true -l reboot -U node1
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk.
Editing this file directly is strongly discouraged.

See Also

`cibadmin(8)` [第193頁], `crm_attribute(8)` [第202頁]

crm_verify (8)

crm_verify — check the CIB for consistency

Synopsis

```
crm_verify [-V] -x file
crm_verify [-V] -X string
crm_verify [-V] -L|-p
crm_verify [-?]
```

Description

crm_verify checks the configuration database (CIB) for consistency and other problems. It can be used to check a file containing the configuration or can it can connect to a running cluster. It reports two classes of problems, errors and warnings. Errors must be fixed before High Availability can work properly. However, it is left up to the administrator to decide if the warnings should also be fixed.

crm_verify assists in creating new or modified configurations. You can take a local copy of a CIB in the running cluster, edit it, validate it using crm_verify, then put the new configuration into effect using cibadmin.

Options

--help, -h
Print a help message.

--verbose, -V
Turn on debug information.

注意

Increase the level of verbosity by providing additional instances.

`--live-check, -L`

Connect to the running cluster and check the CIB.

`--crm_xml string, -X string`

Check the configuration in the supplied string. Pass complete CIBs only.

`--xml-file file, -x file`

Check the configuration in the named file.

`--xml-pipe, -p`

Use the configuration piped in via stdin. Pass complete CIBs only.

Examples

Check the consistency of the configuration in the running cluster and produce verbose output:

```
crm_verify -VL
```

Check the consistency of the configuration in a given file and produce verbose output:

```
crm_verify -Vx file1
```

Pipe a configuration into `crm_verify` and produce verbose output:

```
cat file1.xml | crm_verify -Vp
```

Files

`/var/lib/heartbeat/crm/cib.xml`—the CIB (minus status section) on disk. Editing this file directly is strongly discouraged.

See Also

`cibadmin(8)` [第193頁]

HA OCF Agents

All OCF agents require several parameters to be set when they are started. The following overview shows how to manually operate these agents. The data that is available in this appendix is directly taken from the `meta-data` invocation of the respective RA. Find all these agents in `/usr/lib/ocf/resource.d/heartbeat/`.

When configuring an RA, omit the `OCF_RESKEY_` prefix to the parameter name. Parameters that are in square brackets may be omitted in the configuration.

ocf:anything (7)

ocf:anything — Manages an arbitrary service

Synopsis

```
OCF_RESKEY_binfile=string [OCF_RESKEY_cmdline_options=string]
[OCF_RESKEY_pidfile=string] [OCF_RESKEY_logfile=string]
[OCF_RESKEY_errlogfile=string] [OCF_RESKEY_user=string]
[OCF_RESKEY_monitor_hook=string] [OCF_RESKEY_stop_timeout=string]
anything [start | stop | monitor | meta-data | validate-all]
```

Description

This is a generic OCF RA to manage almost anything.

Supported Parameters

OCF_RESKEY_binfile=Full path name of the binary to be executed
The full name of the binary to be executed. This is expected to keep running with the same pid and not just do something and exit.

OCF_RESKEY_cmdline_options=Command line options
Command line options to pass to the binary

OCF_RESKEY_pidfile=File to write STDOUT to
File to read/write the PID from/to.

OCF_RESKEY_logfile=File to write STDOUT to
File to write STDOUT to

OCF_RESKEY_errlogfile=File to write STDERR to
File to write STDERR to

OCF_RESKEY_user=User to run the command as
User to run the command as

OCF_RESKEY_monitor_hook=Command to run in monitor operation
Command to run in monitor operation

OCF_RESKEY_stop_timeout=Seconds to wait after having sent SIGTERM before
sending SIGKILL in stop operation

In the stop operation: Seconds to wait for kill -TERM to succeed before sending
kill -SIGKILL. Defaults to 2/3 of the stop operation timeout.

ocf:AoEtarget (7)

ocf:AoEtarget — Manages ATA-over-Ethernet (AoE) target exports

Synopsis

```
[OCF_RESKEY_device=string] [OCF_RESKEY_nic=string]  
[OCF_RESKEY_shelf=integer] [OCF_RESKEY_slot=integer]  
OCF_RESKEY_pid=string [OCF_RESKEY_binary=string] AoEtarget [start |  
stop | monitor | reload | meta-data | validate-all]
```

Description

This resource agent manages an ATA-over-Ethernet (AoE) target using vblade. It exports any block device, or file, as an AoE target using the specified Ethernet device, shelf, and slot number.

Supported Parameters

OCF_RESKEY_device=Device to export

The local block device (or file) to export as an AoE target.

OCF_RESKEY_nic=Ethernet interface

The local Ethernet interface to use for exporting this AoE target.

OCF_RESKEY_shelf=AoE shelf number

The AoE shelf number to use when exporting this target.

OCF_RESKEY_slot=AoE slot number

The AoE slot number to use when exporting this target.

OCF_RESKEY_pid=Daemon pid file

The file to record the daemon pid to.

OCF_RESKEY_binary=vblade binary

Location of the vblade binary.

ocf:apache (7)

ocf:apache — Manages an Apache web server instance

Synopsis

```
OCF_RESKEY_configfile=string [OCF_RESKEY_httpd=string]
[OCF_RESKEY_port=integer] [OCF_RESKEY_statusurl=string]
[OCF_RESKEY_testregex=string] [OCF_RESKEY_client=string]
[OCF_RESKEY_testurl=string] [OCF_RESKEY_testregex10=string]
[OCF_RESKEY_testconf=string] [OCF_RESKEY_testname=string]
[OCF_RESKEY_options=string] [OCF_RESKEY_envfiles=string] apache
[start | stop | status | monitor | meta-data | validate-all]
```

Description

This is the resource agent for the Apache web server. This resource agent operates both version 1.x and version 2.x Apache servers. The start operation ends with a loop in which monitor is repeatedly called to make sure that the server started and that it is operational. Hence, if the monitor operation does not succeed within the start operation timeout, the apache resource will end with an error status. The monitor operation by default loads the server status page which depends on the mod_status module and the corresponding configuration file (usually /etc/apache2/mod_status.conf). Make sure that the server status page works and that the access is allowed **only** from localhost (address 127.0.0.1). See the statusurl and testregex attributes for more details. See also <http://httpd.apache.org/>

Supported Parameters

OCF_RESKEY_configfile=configuration file path

The full pathname of the Apache configuration file. This file is parsed to provide defaults for various other resource agent parameters.

OCF_RESKEY_httpd=httpd binary path

The full pathname of the httpd binary (optional).

`OCF_RESKEY_port=httpd port`

A port number that we can probe for status information using the `statusurl`. This will default to the port number found in the configuration file, or 80, if none can be found in the configuration file.

`OCF_RESKEY_statusurl=url name`

The URL to monitor (the apache server status page by default). If left unspecified, it will be inferred from the apache configuration file. If you set this, make sure that it succeeds **only** from the localhost (127.0.0.1). Otherwise, it may happen that the cluster complains about the resource being active on multiple nodes.

`OCF_RESKEY_testregex=monitor regular expression`

Regular expression to match in the output of `statusurl`. Case insensitive.

`OCF_RESKEY_client=http client`

Client to use to query to Apache. If not specified, the RA will try to find one on the system. Currently, `wget` and `curl` are supported. For example, you can set this parameter to "curl" if you prefer that to `wget`.

`OCF_RESKEY_testurl=test url`

URL to test. If it does not start with "http", then it's considered to be relative to the Listen address.

`OCF_RESKEY_testregex10=extended monitor regular expression`

Regular expression to match in the output of `testurl`. Case insensitive.

`OCF_RESKEY_testconf file=test configuration file`

A file which contains test configuration. Could be useful if you have to check more than one web application or in case sensitive info should be passed as arguments (passwords). Furthermore, using a config file is the only way to specify certain parameters. Please see `README.webapps` for examples and file description.

`OCF_RESKEY_testname=test name`

Name of the test within the test configuration file.

`OCF_RESKEY_options=command line options`

Extra options to apply when starting apache. See `man httpd(8)`.

OCF_RESKEY_envfiles=environment settings files

Files (one or more) which contain extra environment variables. If you want to prevent script from reading the default file, set this parameter to empty string.

ocf:AudibleAlarm (7)

ocf:AudibleAlarm — Emits audible beeps at a configurable interval

Synopsis

```
[OCF_RESKEY_nodelist=string] AudibleAlarm [start | stop | restart | status |  
monitor | meta-data | validate-all]
```

Description

Resource script for AudibleAlarm. It sets an audible alarm running by beeping at a set interval.

Supported Parameters

OCF_RESKEY_nodelist=Node list

The node list that should never sound the alarm.

ocf:ClusterMon (7)

ocf:ClusterMon — Runs `crm_mon` in the background, recording the cluster status to an HTML file

Synopsis

```
[OCF_RESKEY_user=string] [OCF_RESKEY_update=integer]  
[OCF_RESKEY_extra_options=string] OCF_RESKEY_pidfile=string  
OCF_RESKEY_htmlfile=string ClusterMon [start | stop | monitor | meta-data |  
validate-all]
```

Description

This is a ClusterMon Resource Agent. It outputs current cluster status to the html.

Supported Parameters

OCF_RESKEY_user=The user we want to run `crm_mon` as
The user we want to run `crm_mon` as

OCF_RESKEY_update=Update interval
How frequently should we update the cluster status

OCF_RESKEY_extra_options=Extra options
Additional options to pass to `crm_mon`. Eg. `-n -r`

OCF_RESKEY_pidfile=PID file
PID file location to ensure only one instance is running

OCF_RESKEY_htmlfile=HTML output
Location to write HTML output to.

ocf:CTDB (7)

ocf:CTDB — CTDB Resource Agent

Synopsis

```
OCF_RESKEY_ctdb_recovery_lock=string
OCF_RESKEY_smb_private_dir=string
[OCF_RESKEY_ctdb_config_dir=string]
[OCF_RESKEY_ctdb_binary=string] [OCF_RESKEY_ctdbd_binary=string]
[OCF_RESKEY_ctdb_socket=string] [OCF_RESKEY_ctdb_dbdir=string]
[OCF_RESKEY_ctdb_logfile=string]
[OCF_RESKEY_ctdb_debuglevel=integer] [OCF_RESKEY_smb_conf=string]
CTDB [start | stop | monitor | meta-data | validate-all]
```

Description

This resource agent manages CTDB, allowing one to use Clustered Samba in a Linux-HA/Pacemaker cluster. You need a shared filesystem (e.g. OCFS2) on which CTDB lock and Samba state will be stored. Configure shares in `smb.conf` on all nodes, and create `/etc/ctdb/nodes` containing a list of private IP addresses of each node in the cluster. Configure this RA as a clone, and it will take care of the rest. For more information see [http://linux-ha.org/wiki/CTDB_\(resource_agent\)](http://linux-ha.org/wiki/CTDB_(resource_agent))

Supported Parameters

`OCF_RESKEY_ctdb_recovery_lock=CTDB shared lock file`

The location of a shared lock file, common across all nodes. This must be on shared storage, e.g.: `/shared-fs/samba/ctdb.lock`

`OCF_RESKEY_smb_private_dir=Samba private dir`

The directory for `smbd` to use for storing such files as `smbpasswd` and `secrets.tdb`. This must be on shared storage, e.g.: `/shared-fs/samba/private`

`OCF_RESKEY_ctdb_config_dir`=CTDB config file directory

The directory containing various CTDB configuration files. The "nodes" and "notify.sh" scripts are expected to be in this directory, as is the "events.d" subdirectory.

`OCF_RESKEY_ctdb_binary`=CTDB binary path

Full path to the CTDB binary.

`OCF_RESKEY_ctdbd_binary`=CTDB Daemon binary path

Full path to the CTDB cluster daemon binary.

`OCF_RESKEY_ctdb_socket`=CTDB socket location

Full path to the domain socket that ctdbd will create, used for local clients to attach and communicate with the ctdb daemon.

`OCF_RESKEY_ctdb_dbdir`=CTDB database directory

The directory to put the local CTDB database files in. Persistent database files will be put in ctdb_dbdir/persistent.

`OCF_RESKEY_ctdb_logfile`=CTDB log file location

Full path to log file. To log to syslog instead, use the value "syslog".

`OCF_RESKEY_ctdb_debuglevel`=CTDB debug level

What debug level to run at (0-10). Higher means more verbose.

`OCF_RESKEY_smb_conf`=Path to smb.conf

Path to default samba config file.

ocf:db2 (7)

ocf:db2 — Manages an IBM DB2 Universal Database instance

Synopsis

```
[OCF_RESKEY_instance=string] [OCF_RESKEY_admin=string] db2 [start | stop  
| status | monitor | validate-all | meta-data | methods]
```

Description

Resource script for db2. It manages a DB2 Universal Database instance as an HA resource.

Supported Parameters

OCF_RESKEY_instance=instance
The instance of database.

OCF_RESKEY_admin=admin
The admin user of the instance.

ocf:Delay (7)

ocf:Delay — Waits for a defined timespan

Synopsis

```
[OCF_RESKEY_startdelay=integer] [OCF_RESKEY_stopdelay=integer]  
[OCF_RESKEY_mondelay=integer] Delay [start | stop | status | monitor | meta-data  
| validate-all]
```

Description

This script is a test resource for introducing delay.

Supported Parameters

OCF_RESKEY_startdelay=Start delay
How long in seconds to delay on start operation.

OCF_RESKEY_stopdelay=Stop delay
How long in seconds to delay on stop operation. Defaults to "startdelay" if unspecified.

OCF_RESKEY_mondelay=Monitor delay
How long in seconds to delay on monitor operation. Defaults to "startdelay" if unspecified.

ocf:drbd (7)

ocf:drbd — Manages a DRBD resource (deprecated)

Synopsis

```
OCF_RESKEY_drbd_resource=string [OCF_RESKEY_drbdconf=string]
[OCF_RESKEY_clone_overrides_hostname=boolean]
[OCF_RESKEY_ignore_deprecation=boolean] drbd [start | promote | demote
| notify | stop | monitor | monitor | meta-data | validate-all]
```

Description

Deprecation warning: This agent is deprecated and may be removed from a future release. See the ocf:linbit:drbd resource agent for a supported alternative. -- This resource agent manages a Distributed Replicated Block Device (DRBD) object as a master/slave resource. DRBD is a mechanism for replicating storage; please see the documentation for setup details.

Supported Parameters

OCF_RESKEY_drbd_resource=drbd resource name
The name of the drbd resource from the drbd.conf file.

OCF_RESKEY_drbdconf=Path to drbd.conf
Full path to the drbd.conf file.

OCF_RESKEY_clone_overrides_hostname=Override drbd hostname
Whether or not to override the hostname with the clone number. This can be used to create floating peer configurations; drbd will be told to use node_<cloneno> as the hostname instead of the real uname, which can then be used in drbd.conf.

OCF_RESKEY_ignore_deprecation=Suppress deprecation warning
If set to true, suppresses the deprecation warning for this agent.

ocf:Dummy (7)

ocf:Dummy — Example stateless resource agent

Synopsis

`OCF_RESKEY_state=string Dummy [start | stop | monitor | reload | migrate_to | migrate_from | meta-data | validate-all]`

Description

This is a Dummy Resource Agent. It does absolutely nothing except keep track of whether its running or not. Its purpose in life is for testing and to serve as a template for RA writers.

Supported Parameters

`OCF_RESKEY_state=State file`
Location to store the resource state in.

ocf:eDir88 (7)

ocf:eDir88 — Manages a Novell eDirectory directory server

Synopsis

```
OCF_RESKEY_eDir_config_file=string  
[OCF_RESKEY_eDir_monitor_ldap=boolean]  
[OCF_RESKEY_eDir_monitor_idm=boolean]  
[OCF_RESKEY_eDir_jvm_initial_heap=integer]  
[OCF_RESKEY_eDir_jvm_max_heap=integer]  
[OCF_RESKEY_eDir_jvm_options=string] eDir88 [start | stop | monitor | meta-  
data | validate-all]
```

Description

Resource script for managing an eDirectory instance. Manages a single instance of eDirectory as an HA resource. The "multiple instances" feature of eDirectory has been added in version 8.8. This script will not work for any version of eDirectory prior to 8.8. This RA can be used to load multiple eDirectory instances on the same host. It is very strongly recommended to put eDir configuration files (as per the eDir_config_file parameter) on local storage on each node. This is necessary for this RA to be able to handle situations where the shared storage has become unavailable. If the eDir configuration file is not available, this RA will fail, and heartbeat will be unable to manage the resource. Side effects include STONITH actions, unmanageable resources, etc... Setting a high action timeout value is very strongly recommended. eDir with IDM can take in excess of 10 minutes to start. If heartbeat times out before eDir has had a chance to start properly, mayhem WILL ENSUE. The LDAP module seems to be one of the very last to start. So this script will take even longer to start on installations with IDM and LDAP if the monitoring of IDM and/or LDAP is enabled, as the start command will wait for IDM and LDAP to be available.

Supported Parameters

OCF_RESKEY_eDir_config_file=eDir config file

Path to configuration file for eDirectory instance.

OCF_RESKEY_eDir_monitor_ldap=eDir monitor ldap

Should we monitor if LDAP is running for the eDirectory instance?

OCF_RESKEY_eDir_monitor_idm=eDir monitor IDM

Should we monitor if IDM is running for the eDirectory instance?

OCF_RESKEY_eDir_jvm_initial_heap=DHOST_INITIAL_HEAP value

Value for the DHOST_INITIAL_HEAP java environment variable. If unset, java defaults will be used.

OCF_RESKEY_eDir_jvm_max_heap=DHOST_MAX_HEAP value

Value for the DHOST_MAX_HEAP java environment variable. If unset, java defaults will be used.

OCF_RESKEY_eDir_jvm_options=DHOST_OPTIONS value

Value for the DHOST_OPTIONS java environment variable. If unset, original values will be used.

ocf:Evmsd (7)

ocf:Evmsd — Controls clustered EVMS volume management (deprecated)

Synopsis

[OCF_RESKEY_ignore_deprecation=boolean] Evmsd [start | stop | monitor | meta-data]

Description

Deprecation warning: EVMS is no longer actively maintained and should not be used. This agent is deprecated and may be removed from a future release. -- This is a Evmsd Resource Agent.

Supported Parameters

OCF_RESKEY_ignore_deprecation=Suppress deprecation warning
If set to true, suppresses the deprecation warning for this agent.

ocf:EvmsSCC (7)

ocf:EvmsSCC — Manages EVMS Shared Cluster Containers (SCCs) (deprecated)

Synopsis

```
[OCF_RESKEY_ignore_deprecation=boolean] EvmsSCC [start | stop | notify  
| status | monitor | meta-data]
```

Description

Deprecation warning: EVMS is no longer actively maintained and should not be used. This agent is deprecated and may be removed from a future release. -- Resource script for EVMS shared cluster container. It runs `evms_activate` on one node in the cluster.

Supported Parameters

`OCF_RESKEY_ignore_deprecation=Suppress deprecation warning`
If set to true, suppresses the deprecation warning for this agent.

ocf:Filesystem (7)

ocf:Filesystem — Manages filesystem mounts

Synopsis

```
[OCF_RESKEY_device=string] [OCF_RESKEY_directory=string]  
[OCF_RESKEY_fstype=string] [OCF_RESKEY_options=string]  
[OCF_RESKEY_statusfile_prefix=string] Filesystem [start | stop | notify  
| monitor | validate-all | meta-data]
```

Description

Resource script for Filesystem. It manages a Filesystem on a shared storage medium. The standard monitor operation of depth 0 (also known as probe) checks if the filesystem is mounted. If you want deeper tests, set `OCF_CHECK_LEVEL` to one of the following values: 10: read first 16 blocks of the device (raw read) This doesn't exercise the filesystem at all, but the device on which the filesystem lives. This is noop for non-block devices such as NFS, SMBFS, or bind mounts. 20: test if a status file can be written and read The status file must be writable by root. This is not always the case with an NFS mount, as NFS exports usually have the "root_squash" option set. In such a setup, you must either use read-only monitoring (depth=10), export with "no_root_squash" on your NFS server, or grant world write permissions on the directory where the status file is to be placed.

Supported Parameters

`OCF_RESKEY_device=block device`

The name of block device for the filesystem, or -U, -L options for mount, or NFS mount specification.

`OCF_RESKEY_directory=mount point`

The mount point for the filesystem.

OCF_RESKEY_fstype=filesystem type

The optional type of filesystem to be mounted.

OCF_RESKEY_options=options

Any extra options to be given as -o options to mount. For bind mounts, add "bind" here and set fstype to "none". We will do the right thing for options such as "bind,ro".

OCF_RESKEY_statusfile_prefix=status file prefix

The prefix to be used for a status file for resource monitoring with depth 20. If you don't specify this parameter, all status files will be created in a separate directory.

ocf:ICP (7)

ocf:ICP — Manages an ICP Vortex clustered host drive

Synopsis

```
[OCF_RESKEY_driveid=string] [OCF_RESKEY_device=string] ICP [start | stop  
| status | monitor | validate-all | meta-data]
```

Description

Resource script for ICP. It Manages an ICP Vortex clustered host drive as an HA resource.

Supported Parameters

OCF_RESKEY_driveid=ICP cluster drive ID
The ICP cluster drive ID.

OCF_RESKEY_device=device
The device name.

ocf:ids (7)

ocf:ids — Manages an Informix Dynamic Server (IDS) instance

Synopsis

```
[OCF_RESKEY_informixdir=string] [OCF_RESKEY_informixserver=string]  
[OCF_RESKEY_onconfig=string] [OCF_RESKEY_dbname=string]  
[OCF_RESKEY_sqltestquery=string] ids [start | stop | status | monitor | validate-  
all | meta-data | methods | usage]
```

Description

OCF resource agent to manage an IBM Informix Dynamic Server (IDS) instance as an High-Availability resource.

Supported Parameters

OCF_RESKEY_informixdir= INFORMIXDIR environment variable

The value the environment variable INFORMIXDIR has after a typical installation of IDS. Or in other words: the path (without trailing '/') where IDS was installed to. If this parameter is unspecified the script will try to get the value from the shell environment.

OCF_RESKEY_informixserver= INFORMIXSERVER environment variable

The value the environment variable INFORMIXSERVER has after a typical installation of IDS. Or in other words: the name of the IDS server instance to manage. If this parameter is unspecified the script will try to get the value from the shell environment.

OCF_RESKEY_onconfig= ONCONFIG environment variable

The value the environment variable ONCONFIG has after a typical installation of IDS. Or in other words: the name of the configuration file for the IDS instance specified in INFORMIXSERVER. The specified configuration file will be searched

at '/etc/'. If this parameter is unspecified the script will try to get the value from the shell environment.

OCF_RESKEY_dbname= database to use for monitoring, defaults to 'sysmaster'
This parameter defines which database to use in order to monitor the IDS instance. If this parameter is unspecified the script will use the 'sysmaster' database as a default.

OCF_RESKEY_sqltestquery= SQL test query to use for monitoring, defaults to 'SELECT COUNT(*) FROM systables;'
SQL test query to run on the database specified by the parameter 'dbname' in order to monitor the IDS instance and determine if it's functional or not. If this parameter is unspecified the script will use 'SELECT COUNT(*) FROM systables;' as a default.

ocf:IPAddr2 (7)

ocf:IPAddr2 — Manages virtual IPv4 addresses (Linux specific version)

Synopsis

```
OCF_RESKEY_ip=string [OCF_RESKEY_nic=string]
[OCF_RESKEY_cidr_netmask=string] [OCF_RESKEY_broadcast=string]
[OCF_RESKEY_iflabel=string] [OCF_RESKEY_lvs_support=boolean]
[OCF_RESKEY_mac=string] [OCF_RESKEY_clusterip_hash=string]
[OCF_RESKEY_unique_clone_address=boolean]
[OCF_RESKEY_arp_interval=integer] [OCF_RESKEY_arp_count=integer]
[OCF_RESKEY_arp_bg=string] [OCF_RESKEY_arp_mac=string] IPAddr2 [start
| stop | status | monitor | meta-data | validate-all]
```

Description

This Linux-specific resource manages IP alias IP addresses. It can add an IP alias, or remove one. In addition, it can implement Cluster Alias IP functionality if invoked as a clone resource.

Supported Parameters

OCF_RESKEY_ip=IPv4 address

The IPv4 address to be configured in dotted quad notation, for example "192.168.1.1".

OCF_RESKEY_nic=Network interface

The base network interface on which the IP address will be brought online. If left empty, the script will try and determine this from the routing table. Do NOT specify an alias interface in the form eth0:1 or anything here; rather, specify the base interface only.

OCF_RESKEY_cidr_netmask=CIDR netmask

The netmask for the interface in CIDR format (e.g., 24 and not 255.255.255.0) If unspecified, the script will also try to determine this from the routing table.

OCF_RESKEY_broadcast=Broadcast address

Broadcast address associated with the IP. If left empty, the script will determine this from the netmask.

OCF_RESKEY_iflabel=Interface label

You can specify an additional label for your IP address here. This label is appended to your interface name. If a label is specified in nic name, this parameter has no effect.

OCF_RESKEY_lvs_support=Enable support for LVS DR

Enable support for LVS Direct Routing configurations. In case a IP address is stopped, only move it to the loopback device to allow the local node to continue to service requests, but no longer advertise it on the network.

OCF_RESKEY_mac=Cluster IP MAC address

Set the interface MAC address explicitly. Currently only used in case of the Cluster IP Alias. Leave empty to chose automatically.

OCF_RESKEY_clusterip_hash=Cluster IP hashing function

Specify the hashing algorithm used for the Cluster IP functionality.

OCF_RESKEY_unique_clone_address=Create a unique address for cloned instances

If true, add the clone ID to the supplied value of ip to create a unique address to manage

OCF_RESKEY_arp_interval=ARP packet interval in ms

Specify the interval between unsolicited ARP packets in milliseconds.

OCF_RESKEY_arp_count=ARP packet count

Number of unsolicited ARP packets to send.

OCF_RESKEY_arp_bg=ARP from background

Whether or not to send the arp packets in the background.

OCF_RESKEY_arp_mac=ARP MAC

MAC address to send the ARP packets too. You really shouldn't be touching this.

ocf:IPaddr (7)

ocf:IPaddr — Manages virtual IPv4 addresses (portable version)

Synopsis

```
OCF_RESKEY_ip=string [OCF_RESKEY_nic=string]
[OCF_RESKEY_cidr_netmask=string] [OCF_RESKEY_broadcast=string]
[OCF_RESKEY_iflabel=string] [OCF_RESKEY_lvs_support=boolean]
[OCF_RESKEY_local_stop_script=string]
[OCF_RESKEY_local_start_script=string]
[OCF_RESKEY_ARP_INTERVAL_MS=integer]
[OCF_RESKEY_ARP_REPEAT=integer]
[OCF_RESKEY_ARP_BACKGROUND=boolean]
[OCF_RESKEY_ARP_NETMASK=string] IPaddr [start | stop | monitor | validate-all
| meta-data]
```

Description

This script manages IP alias IP addresses It can add an IP alias, or remove one.

Supported Parameters

OCF_RESKEY_ip=IPv4 address

The IPv4 address to be configured in dotted quad notation, for example "192.168.1.1".

OCF_RESKEY_nic=Network interface

The base network interface on which the IP address will be brought online. If left empty, the script will try and determine this from the routing table. Do NOT specify an alias interface in the form eth0:1 or anything here; rather, specify the base interface only.

OCF_RESKEY_cidr_netmask=Netmask

The netmask for the interface in CIDR format. (ie, 24), or in dotted quad notation 255.255.255.0). If unspecified, the script will also try to determine this from the routing table.

OCF_RESKEY_broadcast=Broadcast address

Broadcast address associated with the IP. If left empty, the script will determine this from the netmask.

OCF_RESKEY_iflabel=Interface label

You can specify an additional label for your IP address here.

OCF_RESKEY_lvs_support=Enable support for LVS DR

Enable support for LVS Direct Routing configurations. In case a IP address is stopped, only move it to the loopback device to allow the local node to continue to service requests, but no longer advertise it on the network.

OCF_RESKEY_local_stop_script=Script called when the IP is released

Script called when the IP is released

OCF_RESKEY_local_start_script=Script called when the IP is added

Script called when the IP is added

OCF_RESKEY_ARP_INTERVAL_MS=milliseconds between gratuitous ARPs

milliseconds between ARPs

OCF_RESKEY_ARP_REPEAT=repeat count

How many gratuitous ARPs to send out when bringing up a new address

OCF_RESKEY_ARP_BACKGROUND=run in background

run in background (no longer any reason to do this)

OCF_RESKEY_ARP_NETMASK=netmask for ARP

netmask for ARP - in nonstandard hexadecimal format.

ocf:IPsrcaddr (7)

ocf:IPsrcaddr — Manages the preferred source address for outgoing IP packets

Synopsis

```
[OCF_RESKEY_ipaddress=string] IPsrcaddr [start | stop | stop | monitor |  
validate-all | meta-data]
```

Description

Resource script for IPsrcaddr. It manages the preferred source address modification.

Supported Parameters

OCF_RESKEY_ipaddress=IP address
The IP address.

ocf:IPv6addr (7)

ocf:IPv6addr — Manages IPv6 aliases

Synopsis

```
[OCF_RESKEY_ipv6addr=string] [OCF_RESKEY_cidr_netmask=string]  
[OCF_RESKEY_nic=string] IPv6addr [start | stop | status | monitor | validate-all |  
meta-data]
```

Description

This script manages IPv6 alias IPv6 addresses, It can add an IP6 alias, or remove one.

Supported Parameters

OCF_RESKEY_ipv6addr=IPv6 address
The IPv6 address this RA will manage

OCF_RESKEY_cidr_netmask=Netmask
The netmask for the interface in CIDR format. (ie, 24). The value of this parameter overwrites the value of `_prefix_` of `ipv6addr` parameter.

OCF_RESKEY_nic=Network interface
The base network interface on which the IPv6 address will be brought online.

ocf:iSCSILogicalUnit (7)

ocf:iSCSILogicalUnit — Manages iSCSI Logical Units (LUs)

Synopsis

```
[OCF_RESKEY_implementation=string] [OCF_RESKEY_target_iqn=string]  
[OCF_RESKEY_lun=integer] [OCF_RESKEY_path=string]  
OCF_RESKEY_scsi_id=string OCF_RESKEY_scsi_sn=string  
[OCF_RESKEY_vendor_id=string] [OCF_RESKEY_product_id=string]  
[OCF_RESKEY_additional_parameters=string] iSCSILogicalUnit [start  
| stop | monitor | meta-data | validate-all]
```

Description

Manages iSCSI Logical Unit. An iSCSI Logical unit is a subdivision of an SCSI Target, exported via a daemon that speaks the iSCSI protocol.

Supported Parameters

OCF_RESKEY_implementation=iSCSI target daemon implementation

The iSCSI target daemon implementation. Must be one of "iet", "tgt", or "lio". If unspecified, an implementation is selected based on the availability of management utilities, with "iet" being tried first, then "tgt", then "lio".

OCF_RESKEY_target_iqn=iSCSI target IQN

The iSCSI Qualified Name (IQN) that this Logical Unit belongs to.

OCF_RESKEY_lun=Logical Unit number (LUN)

The Logical Unit number (LUN) exposed to initiators.

OCF_RESKEY_path=Block device (or file) path

The path to the block device exposed. Some implementations allow this to be a regular file, too.

OCF_RESKEY_scsi_id=SCSI ID

The SCSI ID to be configured for this Logical Unit. The default is the resource name, truncated to 24 bytes.

OCF_RESKEY_scsi_sn=SCSI serial number

The SCSI serial number to be configured for this Logical Unit. The default is a hash of the resource name, truncated to 8 bytes.

OCF_RESKEY_vendor_id=SCSI vendor ID

The SCSI vendor ID to be configured for this Logical Unit.

OCF_RESKEY_product_id=SCSI product ID

The SCSI product ID to be configured for this Logical Unit.

OCF_RESKEY_additional_parameters=List of iSCSI LU parameters

Additional LU parameters. A space-separated list of "name=value" pairs which will be passed through to the iSCSI daemon's management interface. The supported parameters are implementation dependent. Neither the name nor the value may contain whitespace.

ocf:iSCSITarget (7)

ocf:iSCSITarget — iSCSI target export agent

Synopsis

```
[OCF_RESKEY_implementation=string] OCF_RESKEY_iqn=string  
OCF_RESKEY_tid=integer [OCF_RESKEY_portals=string]  
[OCF_RESKEY_allowed_initiators=string]  
OCF_RESKEY_incoming_username=string  
[OCF_RESKEY_incoming_password=string]  
[OCF_RESKEY_additional_parameters=string] iSCSITarget [start | stop  
| monitor | meta-data | validate-all]
```

Description

Manages iSCSI targets. An iSCSI target is a collection of SCSI Logical Units (LUs) exported via a daemon that speaks the iSCSI protocol.

Supported Parameters

`OCF_RESKEY_implementation=`Manages an iSCSI target export

The iSCSI target daemon implementation. Must be one of "iet", "tgt", or "lio". If unspecified, an implementation is selected based on the availability of management utilities, with "iet" being tried first, then "tgt", then "lio".

`OCF_RESKEY_iqn=`iSCSI target IQN

The target iSCSI Qualified Name (IQN). Should follow the conventional "iqn.yyyy-mm.<reversed domain name>[:identifier]" syntax.

`OCF_RESKEY_tid=`iSCSI target ID

The iSCSI target ID. Required for tgt.

OCF_RESKEY_portals=iSCSI portal addresses

iSCSI network portal addresses. Not supported by all implementations. If unset, the default is to create one portal that listens on .

OCF_RESKEY_allowed_initiators=List of iSCSI initiators allowed to connect to this target

Allowed initiators. A space-separated list of initiators allowed to connect to this target. Initiators may be listed in any syntax the target implementation allows. If this parameter is empty or not set, access to this target will be allowed from any initiator.

OCF_RESKEY_incoming_username=Incoming account username

A username used for incoming initiator authentication. If unspecified, allowed initiators will be able to log in without authentication.

OCF_RESKEY_incoming_password=Incoming account password

A password used for incoming initiator authentication.

OCF_RESKEY_additional_parameters=List of iSCSI target parameters

Additional target parameters. A space-separated list of "name=value" pairs which will be passed through to the iSCSI daemon's management interface. The supported parameters are implementation dependent. Neither the name nor the value may contain whitespace.

ocf:iscsi (7)

ocf:iscsi — Manages a local iSCSI initiator and its connections to iSCSI targets

Synopsis

```
[OCF_RESKEY_portal=string] OCF_RESKEY_target=string  
[OCF_RESKEY_discovery_type=string] [OCF_RESKEY_iscsiadm=string]  
[OCF_RESKEY_udev=string] iscsi [start | stop | status | monitor | validate-all |  
methods | meta-data]
```

Description

OCF Resource Agent for iSCSI. Add (start) or remove (stop) iSCSI targets.

Supported Parameters

OCF_RESKEY_portal=portal

The iSCSI portal address in the form: {ip_address|hostname}[:"port"]

OCF_RESKEY_target=target

The iSCSI target.

OCF_RESKEY_discovery_type=discovery_type

Discovery type. Currently, with open-iscsi, only the sendtargets type is supported.

OCF_RESKEY_iscsiadm=iscsiadm

iscsiadm program path.

OCF_RESKEY_udev=udev

If the next resource depends on the udev creating a device then we wait until it is finished. On a normally loaded host this should be done quickly, but you may be unlucky. If you are not using udev set this to "no", otherwise we will spin in a loop until a timeout occurs.

ocf:ldirectord (7)

ocf:ldirectord — Wrapper OCF Resource Agent for ldirectord

Synopsis

```
OCF_RESKEY_configfile=string [OCF_RESKEY_ldirectord=string]  
ldirectord [start | stop | monitor | meta-data | validate-all]
```

Description

It's a simple OCF RA wrapper for ldirectord and uses the ldirectord interface to create the OCF compliant interface. You win monitoring of ldirectord. Be warned: Asking ldirectord status is an expensive action.

Supported Parameters

OCF_RESKEY_configfile=configuration file path
The full pathname of the ldirectord configuration file.

OCF_RESKEY_ldirectord=ldirectord binary path
The full pathname of the ldirectord.

ocf:LinuxSCSI (7)

ocf:LinuxSCSI — Enables and disables SCSI devices through the kernel SCSI hot-plug subsystem (deprecated)

Synopsis

```
[OCF_RESKEY_scsi=string] [OCF_RESKEY_ignore_deprecation=boolean]  
LinuxSCSI [start | stop | methods | status | monitor | meta-data | validate-all]
```

Description

Deprecation warning: This agent makes use of Linux SCSI hot-plug functionality which has been superseded by SCSI reservations. It is deprecated and may be removed from a future release. See the `scsi2reservation` and `sfex` agents for alternatives. -- This is a resource agent for LinuxSCSI. It manages the availability of a SCSI device from the point of view of the linux kernel. It make Linux believe the device has gone away, and it can make it come back again.

Supported Parameters

`OCF_RESKEY_scsi=SCSI instance`
The SCSI instance to be managed.

`OCF_RESKEY_ignore_deprecation=Suppress deprecation warning`
If set to true, suppresses the deprecation warning for this agent.

ocf:LVM (7)

ocf:LVM — Controls the availability of an LVM Volume Group

Synopsis

```
[OCF_RESKEY_volgrpname=string] [OCF_RESKEY_exclusive=string] LVM  
[start | stop | status | monitor | methods | meta-data | validate-all]
```

Description

Resource script for LVM. It manages an Linux Volume Manager volume (LVM) as an HA resource.

Supported Parameters

OCF_RESKEY_volgrpname=Volume group name
The name of volume group.

OCF_RESKEY_exclusive=Exclusive activation
If set, the volume group will be activated exclusively.

ocf:MailTo (7)

ocf:MailTo — Notifies recipients by email in the event of resource takeover

Synopsis

```
[OCF_RESKEY_email=string] [OCF_RESKEY_subject=string] MailTo [start |  
stop | status | monitor | meta-data | validate-all]
```

Description

This is a resource agent for MailTo. It sends email to a sysadmin whenever a takeover occurs.

Supported Parameters

OCF_RESKEY_email=Email address
The email address of sysadmin.

OCF_RESKEY_subject=Subject
The subject of the email.

ocf:ManageRAID (7)

ocf:ManageRAID — Manages RAID devices

Synopsis

[OCF_RESKEY_raidname=string] ManageRAID [start | stop | status | monitor |
validate-all | meta-data]

Description

Manages starting, stopping and monitoring of RAID devices which are preconfigured in /etc/conf.d/HB-ManageRAID.

Supported Parameters

OCF_RESKEY_raidname=RAID name

Name (case sensitive) of RAID to manage. (preconfigured in /etc/conf.d/HB-
ManageRAID)

ocf:ManageVE (7)

ocf:ManageVE — Manages an OpenVZ Virtual Environment (VE)

Synopsis

```
[OCF_RESKEY_veid=integer] ManageVE [start | stop | status | monitor | validate-all  
| meta-data]
```

Description

This OCF complaint resource agent manages OpenVZ VEs and thus requires a proper OpenVZ installation including a recent vzctl util.

Supported Parameters

OCF_RESKEY_veid=OpenVZ ID of VE

OpenVZ ID of virtual environment (see output of vzlist -a for all assigned IDs)

ocf:mysql-proxy (7)

ocf:mysql-proxy — Manages a MySQL Proxy daemon

Synopsis

```
[OCF_RESKEY_binary=string] OCF_RESKEY_defaults_file=string  
[OCF_RESKEY_proxy_backend_addresses=string]  
[OCF_RESKEY_proxy_read_only_backend_addresses=string]  
[OCF_RESKEY_proxy_address=string] [OCF_RESKEY_log_level=string]  
[OCF_RESKEY_heartbeat=string] [OCF_RESKEY_admin_address=string]  
[OCF_RESKEY_admin_username=string]  
[OCF_RESKEY_admin_password=string]  
[OCF_RESKEY_admin_lua_script=string]  
[OCF_RESKEY_parameters=string] OCF_RESKEY_pidfile=string  
mysql-proxy [start | stop | reload | monitor | validate-all | meta-data]
```

Description

This script manages MySQL Proxy as an OCF resource in a high-availability setup.
Tested with MySQL Proxy 0.7.0 on Debian 5.0.

Supported Parameters

OCF_RESKEY_binary=Full path to MySQL Proxy binary
Full path to the MySQL Proxy binary. For example, "/usr/sbin/mysql-proxy".

OCF_RESKEY_defaults_file=Full path to configuration file
Full path to a MySQL Proxy configuration file. For example, "/etc/mysql-proxy.conf".

OCF_RESKEY_proxy_backend_addresses=MySQL Proxy backend-servers
Address:port of the remote backend-servers (default: 127.0.0.1:3306).

OCF_RESKEY_proxy_read_only_backend_addresses=MySQL Proxy read only backend-servers

Address:port of the remote (read only) slave-server (default:).

OCF_RESKEY_proxy_address=MySQL Proxy listening address

Listening address:port of the proxy-server (default: :4040). You can also specify a socket like "/tmp/mysql-proxy.sock".

OCF_RESKEY_log_level=MySQL Proxy log level.

Log all messages of level (error|warning|info|message|debug) or higher. An empty value disables logging.

OCF_RESKEY_keepalive=Use keepalive option

Try to restart the proxy if it crashed (default:). Valid values: true or false. An empty value equals "false".

OCF_RESKEY_admin_address=MySQL Proxy admin-server address

Listening address:port of the admin-server (default: 127.0.0.1:4041).

OCF_RESKEY_admin_username=MySQL Proxy admin-server username

Username to allow to log in (default:).

OCF_RESKEY_admin_password=MySQL Proxy admin-server password

Password to allow to log in (default:).

OCF_RESKEY_admin_lua_script=MySQL Proxy admin-server lua script

Script to execute by the admin plugin.

OCF_RESKEY_parameters=MySQL Proxy additional parameters

The MySQL Proxy daemon may be called with additional parameters. Specify any of them here.

OCF_RESKEY_pidfile=PID file

PID file

ocf:mysql (7)

ocf:mysql — Manages a MySQL database instance

Synopsis

```
[OCF_RESKEY_binary=string] [OCF_RESKEY_config=string]
[OCF_RESKEY_datadir=string] [OCF_RESKEY_user=string]
[OCF_RESKEY_group=string] [OCF_RESKEY_log=string]
[OCF_RESKEY_pid=string] [OCF_RESKEY_socket=string]
[OCF_RESKEY_test_table=string] [OCF_RESKEY_test_user=string]
[OCF_RESKEY_test_passwd=string]
[OCF_RESKEY_enable_creation=integer]
[OCF_RESKEY_additional_parameters=string]
[OCF_RESKEY_replication_user=string]
[OCF_RESKEY_replication_passwd=string] mysql [start | stop | status | monitor
| monitor | monitor | notify | promote | demote | validate-all | meta-data]
```

Description

Resource script for MySQL. It manages a MySQL Database instance as an HA resource.

Supported Parameters

OCF_RESKEY_binary=MySQL binary
Location of the MySQL binary

OCF_RESKEY_config=MySQL config
Configuration file

OCF_RESKEY_datadir=MySQL datadir
Directory containing databases

OCF_RESKEY_user=MySQL user
User running MySQL daemon

OCF_RESKEY_group=MySQL group
Group running MySQL daemon (for logfile and directory permissions)

OCF_RESKEY_log=MySQL log file
The logfile to be used for mysqld.

OCF_RESKEY_pid=MySQL pid file
The pidfile to be used for mysqld.

OCF_RESKEY_socket=MySQL socket
The socket to be used for mysqld.

OCF_RESKEY_test_table=MySQL test table
Table to be tested in monitor statement (in database.table notation)

OCF_RESKEY_test_user=MySQL test user
MySQL test user

OCF_RESKEY_test_passwd=MySQL test user password
MySQL test user password

OCF_RESKEY_enable_creation=Create the database if it does not exist
If the MySQL database does not exist, it will be created

OCF_RESKEY_additional_parameters=Additional parameters to pass to mysqld
Additional parameters which are passed to the mysqld on startup. (e.g. --skip-external-locking or --skip-grant-tables)

OCF_RESKEY_replication_user=MySQL replication user
MySQL replication user. Used for replication client and slave.

OCF_RESKEY_replication_passwd=MySQL replication user password
MySQL replication password. Used for replication client and slave.

ocf:nfsserver (7)

ocf:nfsserver — Manages an NFS server

Synopsis

```
[OCF_RESKEY_nfs_init_script=string]  
[OCF_RESKEY_nfs_notify_cmd=string]  
[OCF_RESKEY_nfs_shared_infodir=string] [OCF_RESKEY_nfs_ip=string]  
nfsserver [start | stop | monitor | meta-data | validate-all]
```

Description

Nfsserver helps to manage the Linux nfs server as a failover-able resource in Linux-HA. It depends on Linux specific NFS implementation details, so is considered not portable to other platforms yet.

Supported Parameters

OCF_RESKEY_nfs_init_script= Init script for nfsserver

The default init script shipped with the Linux distro. The nfsserver resource agent offloads the start/stop/monitor work to the init script because the procedure to start/stop/monitor nfsserver varies on different Linux distro.

OCF_RESKEY_nfs_notify_cmd= The tool to send out notification.

The tool to send out NSM reboot notification. Failover of nfsserver can be considered as rebooting to different machines. The nfsserver resource agent use this command to notify all clients about the happening of failover.

OCF_RESKEY_nfs_shared_infodir= Directory to store nfs server related information.

The nfsserver resource agent will save nfs related information in this specific directory. And this directory must be able to fail-over before nfsserver itself.

OCF_RESKEY_nfs_ip= IP address.

The floating IP address used to access the nfs service

ocf:oracle (7)

ocf:oracle — Manages an Oracle Database instance

Synopsis

```
OCF_RESKEY_sid=string [OCF_RESKEY_home=string]
[OCF_RESKEY_user=string] [OCF_RESKEY_ipcrm=string]
[OCF_RESKEY_clear_backupmode=boolean]
[OCF_RESKEY_shutdown_method=string] oracle [start | stop | status | monitor
| validate-all | methods | meta-data]
```

Description

Resource script for oracle. Manages an Oracle Database instance as an HA resource.

Supported Parameters

`OCF_RESKEY_sid=sid`
The Oracle SID (aka `ORACLE_SID`).

`OCF_RESKEY_home=home`
The Oracle home directory (aka `ORACLE_HOME`). If not specified, then the SID along with its home should be listed in `/etc/oratab`.

`OCF_RESKEY_user=user`
The Oracle owner (aka `ORACLE_OWNER`). If not specified, then it is set to the owner of file `$ORACLE_HOME/dbs/*${ORACLE_SID}.ora`. If this does not work for you, just set it explicitly.

`OCF_RESKEY_ipcrm=ipcrm`
Sometimes IPC objects (shared memory segments and semaphores) belonging to an Oracle instance might be left behind which prevents the instance from starting. It is not easy to figure out which shared segments belong to which instance, in particular when more instances are running as same user. What we use here is the

"oradebug" feature and its "ipc" trace utility. It is not optimal to parse the debugging information, but I am not aware of any other way to find out about the IPC information. In case the format or wording of the trace report changes, parsing might fail. There are some precautions, however, to prevent stepping on other peoples toes. There is also a dumpinstipc option which will make us print the IPC objects which belong to the instance. Use it to see if we parse the trace file correctly. Three settings are possible: - none: don't mess with IPC and hope for the best (beware: you'll probably be out of luck, sooner or later) - instance: try to figure out the IPC stuff which belongs to the instance and remove only those (default; should be safe) - orauser: remove all IPC belonging to the user which runs the instance (don't use this if you run more than one instance as same user or if other apps running as this user use IPC) The default setting "instance" should be safe to use, but in that case we cannot guarantee that the instance will start. In case IPC objects were already left around, because, for instance, someone mercilessly killing Oracle processes, there is no way any more to find out which IPC objects should be removed. In that case, human intervention is necessary, and probably all instances running as same user will have to be stopped. The third setting, "orauser", guarantees IPC objects removal, but it does that based only on IPC objects ownership, so you should use that only if every instance runs as separate user. Please report any problems. Suggestions/fixes welcome.

```
OCF_RESKEY_clear_backupmode=clear_backupmode
```

The clear of the backup mode of ORACLE.

```
OCF_RESKEY_shutdown_method=shutdown_method
```

How to stop Oracle is a matter of taste it seems. The default method ("checkpoint/abort") is: alter system checkpoint; shutdown abort; This should be the fastest safe way bring the instance down. If you find "shutdown abort" distasteful, set this attribute to "immediate" in which case we will shutdown immediate; If you still think that there's even better way to shutdown an Oracle instance we are willing to listen.

ocf:oralsnr (7)

ocf:oralsnr — Manages an Oracle TNS listener

Synopsis

```
OCF_RESKEY_sid=string [OCF_RESKEY_home=string]  
[OCF_RESKEY_user=string] OCF_RESKEY_listener=string oralsnr [start |  
stop | status | monitor | validate-all | meta-data | methods]
```

Description

Resource script for Oracle Listener. It manages an Oracle Listener instance as an HA resource.

Supported Parameters

OCF_RESKEY_sid=sid

The Oracle SID (aka ORACLE_SID). Necessary for the monitor op, i.e. to do tnsping SID.

OCF_RESKEY_home=home

The Oracle home directory (aka ORACLE_HOME). If not specified, then the SID should be listed in /etc/oratab.

OCF_RESKEY_user=user

Run the listener as this user.

OCF_RESKEY_listener=listener

Listener instance to be started (as defined in listener.ora). Defaults to LISTENER.

ocf:pgsql (7)

ocf:pgsql — Manages a PostgreSQL database instance

Synopsis

```
[OCF_RESKEY_pgctl=string] [OCF_RESKEY_start_opt=string]
[OCF_RESKEY_ctl_opt=string] [OCF_RESKEY_psql=string]
[OCF_RESKEY_pgdata=string] [OCF_RESKEY_pgdba=string]
[OCF_RESKEY_pghost=string] [OCF_RESKEY_pgport=string]
[OCF_RESKEY_pgdb=string] [OCF_RESKEY_logfile=string]
[OCF_RESKEY_stop_escalate=string] psql [start | stop | status | monitor |
meta-data | validate-all | methods]
```

Description

Resource script for PostgreSQL. It manages a PostgreSQL as an HA resource.

Supported Parameters

OCF_RESKEY_pgctl=pgctl
Path to pg_ctl command.

OCF_RESKEY_start_opt=start_opt
Start options (-o start_opt in pgi_ctl). "-i -p 5432" for example.

OCF_RESKEY_ctl_opt=ctl_opt
Additional pg_ctl options (-w, -W etc..). Default is ""

OCF_RESKEY_psql=psql
Path to psql command.

OCF_RESKEY_pgdata=pgdata
Path PostgreSQL data directory.

OCF_RESKEY_pgdba=pgdba
User that owns PostgreSQL.

OCF_RESKEY_pghost=pghost
Hostname/IP Address where PostgreSQL is listening

OCF_RESKEY_pgport=pgport
Port where PostgreSQL is listening

OCF_RESKEY_pgdb=pgdb
Database that will be used for monitoring.

OCF_RESKEY_logfile=logfile
Path to PostgreSQL server log output file.

OCF_RESKEY_stop_escalate=stop escalation
Number of retries (using -m fast) before resorting to -m immediate

ocf:pingd (7)

ocf:pingd — Monitors connectivity to specific hosts or IP addresses ("ping nodes")
(deprecated)

Synopsis

```
[OCF_RESKEY_pidfile=string] [OCF_RESKEY_user=string]  
[OCF_RESKEY_dampen=integer] [OCF_RESKEY_set=integer]  
[OCF_RESKEY_name=integer] [OCF_RESKEY_section=integer]  
[OCF_RESKEY_multiplier=integer] [OCF_RESKEY_host_list=integer]  
[OCF_RESKEY_ignore_deprecation=boolean] pingd [start | stop | monitor |  
meta-data | validate-all]
```

Description

Deprecation warning: This agent is deprecated and may be removed from a future release. See the ocf:pacemaker:pingd resource agent for a supported alternative. -- This is a pingd Resource Agent. It records (in the CIB) the current number of ping nodes a node can connect to.

Supported Parameters

OCF_RESKEY_pidfile=PID file
PID file

OCF_RESKEY_user=The user we want to run pingd as
The user we want to run pingd as

OCF_RESKEY_dampen=Dampening interval
The time to wait (dampening) further changes occur

OCF_RESKEY_set=Set name
The name of the instance_attributes set to place the value in. Rarely needs to be specified.

OCF_RESKEY_name=Attribute name

The name of the attributes to set. This is the name to be used in the constraints.

OCF_RESKEY_section=Section name

The section place the value in. Rarely needs to be specified.

OCF_RESKEY_multiplier=Value multiplier

The number by which to multiply the number of connected ping nodes by

OCF_RESKEY_host_list=Host list

The list of ping nodes to count. Defaults to all configured ping nodes. Rarely needs to be specified.

OCF_RESKEY_ignore_deprecation=Suppress deprecation warning

If set to true, suppresses the deprecation warning for this agent.

ocf:portblock (7)

ocf:portblock — Block and unblocks access to TCP and UDP ports

Synopsis

```
[OCF_RESKEY_protocol=string] [OCF_RESKEY_portno=integer]
[OCF_RESKEY_action=string] [OCF_RESKEY_ip=string]
[OCF_RESKEY_tickle_dir=string] [OCF_RESKEY_sync_script=string]
portblock [start | stop | status | monitor | meta-data | validate-all]
```

Description

Resource script for portblock. It is used to temporarily block ports using iptables. In addition, it may allow for faster TCP reconnects for clients on failover. Use that if there are long lived TCP connections to an HA service. This feature is enabled by setting the tickle_dir parameter and only in concert with action set to unblock. Note that the tickle ACK function is new as of version 3.0.2 and hasn't yet seen widespread use.

Supported Parameters

OCF_RESKEY_protocol=protocol
The protocol used to be blocked/unblocked.

OCF_RESKEY_portno=portno
The port number used to be blocked/unblocked.

OCF_RESKEY_action=action
The action (block/unblock) to be done on the protocol::portno.

OCF_RESKEY_ip=ip
The IP address used to be blocked/unblocked.

OCF_RESKEY_tickle_dir=Tickle directory

The shared or local directory (must be absolute path) which stores the established TCP connections.

OCF_RESKEY_sync_script=Connection state file synchronization script

If the tickle_dir is a local directory, then the TCP connection state file has to be replicated to other nodes in the cluster. It can be csync2 (default), some wrapper of rsync, or whatever. It takes the file name as a single argument. For csync2, set it to "csync2 -xv".

ocf:proftpd (7)

ocf:proftpd — OCF Resource Agent compliant FTP script.

Synopsis

```
[OCF_RESKEY_binary=string] [OCF_RESKEY_confdir=string]
[OCF_RESKEY_pidfile=string] [OCF_RESKEY_curl_binary=string]
[OCF_RESKEY_curl_url=string] [OCF_RESKEY_test_user=string]
[OCF_RESKEY_test_pass=string] proftpd [start | stop | monitor | monitor |
validate-all | meta-data]
```

Description

This script manages Proftpd in an Active-Passive setup

Supported Parameters

OCF_RESKEY_binary=The Proftpd binary
The Proftpd binary

OCF_RESKEY_confdir=Configuration file name with full path
The Proftpd configuration file name with full path. For example, "/etc/proftpd.conf"

OCF_RESKEY_pidfile=PID file
The Proftpd PID file. The location of the PID file is configured in the Proftpd configuration file.

OCF_RESKEY_curl_binary=The absolut path to the curl binary
The absolut path to the curl binary for monitoring with OCF_CHECK_LEVEL greater zero.

OCF_RESKEY_curl_url=The URL which is checked by curl
The URL which is checked by curl with OCF_CHECK_LEVEL greater zero.

OCF_RESKEY_test_user=The name of the ftp user

The name of the ftp user for monitoring with OCF_CHECK_LEVEL greater zero.

OCF_RESKEY_test_pass=The password of the ftp user

The password of the ftp user for monitoring with OCF_CHECK_LEVEL greater zero.

ocf:Pure-FTPd (7)

ocf:Pure-FTPd — Manages a Pure-FTPd FTP server instance

Synopsis

```
OCF_RESKEY_script=string OCF_RESKEY_conf=string  
OCF_RESKEY_daemon_type=string [OCF_RESKEY_pidfile=string]  
Pure-FTPd [start | stop | monitor | validate-all | meta-data]
```

Description

This script manages Pure-FTPd in an Active-Passive setup

Supported Parameters

OCF_RESKEY_script=Script name with full path
The full path to the Pure-FTPd startup script. For example, "/sbin/pure-config.pl"

OCF_RESKEY_conf=Configuration file name with full path
The Pure-FTPd configuration file name with full path. For example, "/etc/pure-ftpd/pure-ftpd.conf"

OCF_RESKEY_daemon_type=Configuration file name with full path
The Pure-FTPd daemon to be called by pure-ftpd-wrapper. Valid options are "" for pure-ftpd, "mysql" for pure-ftpd-mysql, "postgresql" for pure-ftpd-postgresql and "ldap" for pure-ftpd-ldap

OCF_RESKEY_pidfile=PID file
PID file

ocf:Raid1 (7)

ocf:Raid1 — Manages a software RAID1 device on shared storage

Synopsis

```
[OCF_RESKEY_raidconf=string] [OCF_RESKEY_raiddev=string]  
[OCF_RESKEY_homehost=string] Raid1 [start | stop | status | monitor | validate-  
all | meta-data]
```

Description

Resource script for RAID1. It manages a software Raid1 device on a shared storage medium.

Supported Parameters

OCF_RESKEY_raidconf=RAID config file
The RAID configuration file. e.g. /etc/raidtab or /etc/mdadm.conf.

OCF_RESKEY_raiddev=block device
The block device to use.

OCF_RESKEY_homehost=Homehost for mdadm
The value for the homehost directive; this is an mdadm feature to protect RAIDs against being activated by accident. It is recommended to create RAIDs managed by the cluster with "homehost" set to a special value, so they are not accidentally auto-assembled by nodes not supposed to own them.

ocf:Route (7)

ocf:Route — Manages network routes

Synopsis

```
OCF_RESKEY_destination=string OCF_RESKEY_device=string  
OCF_RESKEY_gateway=string OCF_RESKEY_source=string  
[OCF_RESKEY_table=string] Route [start | stop | monitor | reload | meta-data |  
validate-all]
```

Description

Enables and disables network routes. Supports host and net routes, routes via a gateway address, and routes using specific source addresses. This resource agent is useful if a node's routing table needs to be manipulated based on node role assignment. Consider the following example use case: - One cluster node serves as an IPsec tunnel endpoint. - All other nodes use the IPsec tunnel to reach hosts in a specific remote network. Then, here is how you would implement this scheme making use of the Route resource agent: - Configure an ipsec LSB resource. - Configure a cloned Route OCF resource. - Create an order constraint to ensure that ipsec is started before Route. - Create a colocation constraint between the ipsec and Route resources, to make sure no instance of your cloned Route resource is started on the tunnel endpoint itself.

Supported Parameters

OCF_RESKEY_destination=Destination network

The destination network (or host) to be configured for the route. Specify the netmask suffix in CIDR notation (e.g. "/24"). If no suffix is given, a host route will be created. Specify "0.0.0.0/0" or "default" if you want this resource to set the system default route.

OCF_RESKEY_device=Outgoing network device

The outgoing network device to use for this route.

OCF_RESKEY_gateway=Gateway IP address
The gateway IP address to use for this route.

OCF_RESKEY_source=Source IP address
The source IP address to be configured for the route.

OCF_RESKEY_table=Routing table
The routing table to be configured for the route.

ocf:rsyncd (7)

ocf:rsyncd — Manages an rsync daemon

Synopsis

```
[OCF_RESKEY_binpath=string] [OCF_RESKEY_conf file=string]  
[OCF_RESKEY_bwlimit=string] rsyncd [start | stop | monitor | validate-all | meta-  
data]
```

Description

This script manages rsync daemon

Supported Parameters

OCF_RESKEY_binpath=Full path to the rsync binary
The rsync binary path. For example, "/usr/bin/rsync"

OCF_RESKEY_conf file=Configuration file name with full path
The rsync daemon configuration file name with full path. For example,
"/etc/rsyncd.conf"

OCF_RESKEY_bwlimit=limit I/O bandwidth, KBytes per second
This option allows you to specify a maximum transfer rate in kilobytes per second.
This option is most effective when using rsync with large files (several megabytes
and up). Due to the nature of rsync transfers, blocks of data are sent, then if rsync
determines the transfer was too fast, it will wait before sending the next data block.
The result is an average transfer rate equaling the specified limit. A value of zero
specifies no limit.

ocf:SAPDatabase (7)

ocf:SAPDatabase — Manages any SAP database (based on Oracle, MaxDB, or DB2)

Synopsis

```
OCF_RESKEY_SID=string OCF_RESKEY_DIR_EXECUTABLE=string
OCF_RESKEY_DBTYPE=string OCF_RESKEY_NETSERVICENAME=string
OCF_RESKEY_DBJ2EE_ONLY=boolean OCF_RESKEY_JAVA_HOME=string
OCF_RESKEY_STRICT_MONITORING=boolean
OCF_RESKEY_AUTOMATIC_RECOVER=boolean
OCF_RESKEY_DIR_BOOTSTRAP=string OCF_RESKEY_DIR_SECSTORE=string
OCF_RESKEY_DB_JARS=string OCF_RESKEY_PRE_START_USEREXIT=string
OCF_RESKEY_POST_START_USEREXIT=string
OCF_RESKEY_PRE_STOP_USEREXIT=string
OCF_RESKEY_POST_STOP_USEREXIT=string SAPDatabase [start | stop | status
| monitor | validate-all | meta-data | methods]
```

Description

Resource script for SAP databases. It manages a SAP database of any type as an HA resource.

Supported Parameters

OCF_RESKEY_SID=SAP system ID

The unique SAP system identifier. e.g. P01

OCF_RESKEY_DIR_EXECUTABLE=path of sapstartsrv and sapcontrol

The full qualified path where to find sapstartsrv and sapcontrol.

OCF_RESKEY_DBTYPE=database vendor

The name of the database vendor you use. Set either: ORA,DB6,ADA

OCF_RESKEY_NETSERVICENAME=listener name

The Oracle TNS listener name.

OCF_RESKEY_DBJ2EE_ONLY=only JAVA stack installed

If you do not have a ABAP stack installed in the SAP database, set this to TRUE

OCF_RESKEY_JAVA_HOME=Path to Java SDK

This is only needed if the DBJ2EE_ONLY parameter is set to true. Enter the path to the Java SDK which is used by the SAP WebAS Java

OCF_RESKEY_STRICT_MONITORING=Activates application level monitoring

This controls how the resource agent monitors the database. If set to true, it will use SAP tools to test the connect to the database. Do not use with Oracle, because it will result in unwanted failovers in case of an archiver stuck

OCF_RESKEY_AUTOMATIC_RECOVER=Enable or disable automatic startup recovery

The SAPDatabase resource agent tries to recover a failed start attempt automatically one time. This is done by running a forced abort of the RDBMS and/or executing recovery commands.

OCF_RESKEY_DIR_BOOTSTRAP=path to j2ee bootstrap directory

The full qualified path where to find the J2EE instance bootstrap directory. e.g.
/usr/sap/P01/J00/j2ee/cluster/bootstrap

OCF_RESKEY_DIR_SECSTORE=path to j2ee secure store directory

The full qualified path where to find the J2EE security store directory. e.g.
/usr/sap/P01/SYS/global/security/lib/tools

OCF_RESKEY_DB_JARS=file name of the jdbc driver

The full qualified filename of the jdbc driver for the database connection test. It will be automatically read from the bootstrap.properties file in Java engine 6.40 and 7.00. For Java engine 7.10 the parameter is mandatory.

OCF_RESKEY_PRE_START_USEREXIT=path to a pre-start script

The full qualified path where to find a script or program which should be executed before this resource gets started.

OCF_RESKEY_POST_START_USEREXIT=path to a post-start script

The full qualified path where to find a script or program which should be executed after this resource got started.

OCF_RESKEY_PRE_STOP_USEREXIT=path to a pre-start script

The full qualified path where to find a script or program which should be executed before this resource gets stopped.

OCF_RESKEY_POST_STOP_USEREXIT=path to a post-start script

The full qualified path where to find a script or program which should be executed after this resource got stopped.

ocf:SAPInstance (7)

ocf:SAPInstance — Manages a SAP instance

Synopsis

```
OCF_RESKEY_InstanceName=string OCF_RESKEY_DIR_EXECUTABLE=string
OCF_RESKEY_DIR_PROFILE=string OCF_RESKEY_START_PROFILE=string
OCF_RESKEY_START_WAITTIME=string
OCF_RESKEY_AUTOMATIC_RECOVER=boolean
OCF_RESKEY_MONITOR_SERVICES=string
OCF_RESKEY_ERS_InstanceName=string
OCF_RESKEY_ERS_START_PROFILE=string
OCF_RESKEY_PRE_START_USEREXIT=string
OCF_RESKEY_POST_START_USEREXIT=string
OCF_RESKEY_PRE_STOP_USEREXIT=string
OCF_RESKEY_POST_STOP_USEREXIT=string SAPInstance [start | stop | status
| monitor | promote | demote | validate-all | meta-data | methods]
```

Description

Resource script for SAP. It manages a SAP Instance as an HA resource.

Supported Parameters

OCF_RESKEY_InstanceName=**instance name: SID_INSTANCE_VIR-HOSTNAME**
The full qualified SAP instance name. e.g. P01_DVEBMGS00_sapp01ci

OCF_RESKEY_DIR_EXECUTABLE=**path of sapstartsrv and sapcontrol**
The full qualified path where to find sapstartsrv and sapcontrol.

OCF_RESKEY_DIR_PROFILE=**path of start profile**
The full qualified path where to find the SAP START profile.

OCF_RESKEY_START_PROFILE=start profile name

The name of the SAP START profile.

OCF_RESKEY_START_WAITTIME=Check the successful start after that time (do not wait for J2EE-Addin)

After that time in seconds a monitor operation is executed by the resource agent.

Does the monitor return SUCCESS, the start is handled as SUCCESS. This is useful to resolve timing problems with e.g. the J2EE-Addin instance.

OCF_RESKEY_AUTOMATIC_RECOVER=Enable or disable automatic startup recovery

The SAPInstance resource agent tries to recover a failed start attempt automatically one time. This is done by killing running instance processes and executing cleanipc.

OCF_RESKEY_MONITOR_SERVICES=

OCF_RESKEY_ERS_InstanceName=

OCF_RESKEY_ERS_START_PROFILE=

OCF_RESKEY_PRE_START_USEREXIT=path to a pre-start script

The full qualified path where to find a script or program which should be executed before this resource gets started.

OCF_RESKEY_POST_START_USEREXIT=path to a post-start script

The full qualified path where to find a script or program which should be executed after this resource got started.

OCF_RESKEY_PRE_STOP_USEREXIT=path to a pre-stop script

The full qualified path where to find a script or program which should be executed before this resource gets stopped.

OCF_RESKEY_POST_STOP_USEREXIT=path to a post-stop script

The full qualified path where to find a script or program which should be executed after this resource got stopped.

ocf:scsi2reservation (7)

ocf:scsi2reservation — scsi-2 reservation

Synopsis

```
[OCF_RESKEY_scsi_reserve=string] [OCF_RESKEY_sharedisk=string]  
[OCF_RESKEY_start_loop=string] scsi2reservation [start | stop | monitor  
| meta-data | validate-all]
```

Description

The scsi-2-reserve resource agent is a place holder for SCSI-2 reservation. A healthy instance of scsi-2-reserve resource, indicates the own of the specified SCSI device. This resource agent depends on the scsi_reserve from scsires package, which is Linux specific.

Supported Parameters

OCF_RESKEY_scsi_reserve=Manages exclusive access to shared storage media through SCSI-2 reservations

The `scsi_reserve` is a command from scsires package. It helps to issue SCSI-2 reservation on SCSI devices.

OCF_RESKEY_sharedisk= Shared disk.

The shared disk that can be reserved.

OCF_RESKEY_start_loop= Times to re-try before giving up.

We are going to try several times before giving up. `Start_loop` indicates how many times we are going to re-try.

ocf:SendArp (7)

ocf:SendArp — Broadcasts unsolicited ARP announcements

Synopsis

```
[OCF_RESKEY_ip=string] [OCF_RESKEY_nic=string] SendArp [start | stop |  
monitor | meta-data | validate-all]
```

Description

This script send out gratuitous Arp for an IP address

Supported Parameters

OCF_RESKEY_ip=IP address

The IP address for sending arp package.

OCF_RESKEY_nic=NIC

The nic for sending arp package.

ocf:ServeRAID (7)

ocf:ServeRAID — Enables and disables shared ServeRAID merge groups

Synopsis

```
[OCF_RESKEY_serveraid=integer] [OCF_RESKEY_mergegroup=integer]  
ServeRAID [start | stop | status | monitor | validate-all | meta-data | methods]
```

Description

Resource script for ServeRAID. It enables/disables shared ServeRAID merge groups.

Supported Parameters

OCF_RESKEY_serveraid=serveraid
The adapter number of the ServeRAID adapter.

OCF_RESKEY_mergegroup=mergegroup
The logical drive under consideration.

ocf:sfex (7)

ocf:sfex — Manages exclusive access to shared storage using Shared Disk File EXclusiveness (SF-EX)

Synopsis

```
[OCF_RESKEY_device=string] [OCF_RESKEY_index=integer]
[OCF_RESKEY_collision_timeout=integer]
[OCF_RESKEY_monitor_interval=integer]
[OCF_RESKEY_lock_timeout=integer] sfex [start | stop | monitor | meta-data]
```

Description

Resource script for SF-EX. It manages a shared storage medium exclusively .

Supported Parameters

OCF_RESKEY_device=block device

Block device path that stores exclusive control data.

OCF_RESKEY_index=index

Location in block device where exclusive control data is stored. 1 or more is specified. Default is 1.

OCF_RESKEY_collision_timeout=waiting time for lock acquisition

Waiting time when a collision of lock acquisition is detected. Default is 1 second.

OCF_RESKEY_monitor_interval=monitor interval

Monitor interval(sec). Default is 10 seconds

OCF_RESKEY_lock_timeout=Valid term of lock

Valid term of lock(sec). Default is 20 seconds.

ocf:SphinxSearchDaemon (7)

ocf:SphinxSearchDaemon — Manages the Sphinx search daemon.

Synopsis

```
OCF_RESKEY_config=string [OCF_RESKEY_searchd=string]  
[OCF_RESKEY_search=string] [OCF_RESKEY_testQuery=string]  
SphinxSearchDaemon [start | stop | monitor | meta-data | validate-all]
```

Description

This is a searchd Resource Agent. It manages the Sphinx Search Daemon.

Supported Parameters

OCF_RESKEY_config=Configuration file
searchd configuration file

OCF_RESKEY_searchd=searchd binary
searchd binary

OCF_RESKEY_search=search binary
Search binary for functional testing in the monitor action.

OCF_RESKEY_testQuery=test query
Test query for functional testing in the monitor action. The query does not need to match any documents in the index. The purpose is merely to test whether the search daemon is able to query its indices and respond properly.

ocf:Squid (7)

ocf:Squid — Manages a Squid proxy server instance

Synopsis

```
[OCF_RESKEY_squid_exe=string] OCF_RESKEY_squid_conf=string  
OCF_RESKEY_squid_pidfile=string OCF_RESKEY_squid_port=integer  
[OCF_RESKEY_squid_stop_timeout=integer]  
[OCF_RESKEY_debug_mode=string] [OCF_RESKEY_debug_log=string] Squid  
[start | stop | status | monitor | meta-data | validate-all]
```

Description

The resource agent of Squid. This manages a Squid instance as an HA resource.

Supported Parameters

OCF_RESKEY_squid_exe=Executable file

This is a required parameter. This parameter specifies squid's executable file.

OCF_RESKEY_squid_conf=Configuration file

This is a required parameter. This parameter specifies a configuration file for a squid instance managed by this RA.

OCF_RESKEY_squid_pidfile=Pidfile

This is a required parameter. This parameter specifies a process id file for a squid instance managed by this RA.

OCF_RESKEY_squid_port=Port number

This is a required parameter. This parameter specifies a port number for a squid instance managed by this RA. If plural ports are used, you must specify the only one of them.

OCF_RESKEY_squid_stop_timeout=Number of seconds to await to confirm a normal stop method

This is an omittable parameter. On a stop action, a normal stop method is firstly used. and then the confirmation of its completion is awaited for the specified seconds by this parameter. The default value is 10.

OCF_RESKEY_debug_mode=Debug mode

This is an optional parameter. This RA runs in debug mode when this parameter includes 'x' or 'v'. If 'x' is included, both of STDOUT and STDERR redirect to the logfile specified by "debug_log", and then the builtin shell option 'x' is turned on. It is similar about 'v'.

OCF_RESKEY_debug_log=A destination of the debug log

This is an optional and omittable parameter. This parameter specifies a destination file for debug logs and works only if this RA run in debug mode. Refer to "debug_mode" about debug mode. If no value is given but it's required, it's made by the following rules: "/var/log/" as a directory part, the basename of the configuration file given by "syslog_ng_conf" as a basename part, ".log" as a suffix.

ocf:Stateful (7)

ocf:Stateful — Example stateful resource agent

Synopsis

```
OCF_RESKEY_state=string Stateful [start | stop | monitor | meta-data | validate-  
all]
```

Description

This is an example resource agent that impliments two states

Supported Parameters

```
OCF_RESKEY_state=State file  
    Location to store the resource state in
```

ocf:SysInfo (7)

ocf:SysInfo — Records various node attributes in the CIB

Synopsis

```
[OCF_RESKEY_pidfile=string] [OCF_RESKEY_delay=string] SysInfo [start  
| stop | monitor | meta-data | validate-all]
```

Description

This is a SysInfo Resource Agent. It records (in the CIB) various attributes of a node
Sample Linux output: arch: i686 os: Linux-2.4.26-gentoo-r14 free_swap: 1999 cpu_info:
Intel(R) Celeron(R) CPU 2.40GHz cpu_speed: 4771.02 cpu_cores: 1 cpu_load: 0.00
ram_total: 513 ram_free: 117 root_free: 2.4 Sample Darwin output: arch: i386 os:
Darwin-8.6.2 cpu_info: Intel Core Duo cpu_speed: 2.16 cpu_cores: 2 cpu_load: 0.18
ram_total: 2016 ram_free: 787 root_free: 13 Units: free_swap: Mb ram_*: Mb root_free:
Gb cpu_speed (Linux): bogomips cpu_speed (Darwin): Ghz

Supported Parameters

OCF_RESKEY_pidfile=PID file
PID file

OCF_RESKEY_delay=Dampening Delay
Interval to allow values to stabilize

ocf:syslog-ng (7)

ocf:syslog-ng — Syslog-ng resource agent

Synopsis

```
[OCF_RESKEY_configfile=string]  
[OCF_RESKEY_syslog_ng_binary=string]  
[OCF_RESKEY_start_opts=string]  
[OCF_RESKEY_kill_term_timeout=integer] syslog-ng [start | stop | status  
| monitor | meta-data | validate-all]
```

Description

This script manages a syslog-ng instance as an HA resource.

Supported Parameters

`OCF_RESKEY_configfile=`Configuration file

This parameter specifies a configuration file for a syslog-ng instance managed by this RA.

`OCF_RESKEY_syslog_ng_binary=`syslog-ng executable

This parameter specifies syslog-ng's executable file.

`OCF_RESKEY_start_opts=`Start options

This parameter specifies startup options for a syslog-ng instance managed by this RA. When no value is given, no startup options is used. Don't use option '-F'. It causes a stuck of a start action.

`OCF_RESKEY_kill_term_timeout=`Number of seconds to await to confirm a normal stop method

On a stop action, a normal stop method(`pkill -TERM`) is firstly used. And then the confirmation of its completion is waited for the specified seconds by this parameter. The default value is 10.

ocf:tomcat (7)

ocf:tomcat — Manages a Tomcat servlet environment instance

Synopsis

```
OCF_RESKEY_tomcat_name=string OCF_RESKEY_script_log=string
[OCF_RESKEY_tomcat_stop_timeout=integer]
[OCF_RESKEY_tomcat_suspend_trialcount=integer]
[OCF_RESKEY_tomcat_user=string] [OCF_RESKEY_statusurl=string]
[OCF_RESKEY_java_home=string] OCF_RESKEY_catalina_home=string
OCF_RESKEY_catalina_pid=string
[OCF_RESKEY_tomcat_start_opts=string]
[OCF_RESKEY_catalina_opts=string]
[OCF_RESKEY_catalina_rotate_log=string]
[OCF_RESKEY_catalina_rotatetime=integer] tomcat [start | stop | status |
monitor | meta-data | validate-all]
```

Description

Resource script for tomcat. It manages a Tomcat instance as an HA resource.

Supported Parameters

OCF_RESKEY_tomcat_name=The name of the resource
The name of the resource

OCF_RESKEY_script_log=A destination of the log of this script
A destination of the log of this script

OCF_RESKEY_tomcat_stop_timeout=Time-out at the time of the stop
Time-out at the time of the stop

OCF_RESKEY_tomcat_suspend_trialcount=The re-try number of times awaiting a stop

The re-try number of times awaiting a stop

OCF_RESKEY_tomcat_user=A user name to start a resource

A user name to start a resource

OCF_RESKEY_statusurl=URL for state confirmation

URL for state confirmation

OCF_RESKEY_java_home=Home directory of the Java

Home directory of the Java

OCF_RESKEY_catalina_home=Home directory of Tomcat

Home directory of Tomcat

OCF_RESKEY_catalina_pid=A PID file name of Tomcat

A PID file name of Tomcat

OCF_RESKEY_tomcat_start_opts=Tomcat start options

Tomcat start options

OCF_RESKEY_catalina_opts=Catalina options

Catalina options

OCF_RESKEY_catalina_rotate_log=Rotate catalina.out flag

Rotate catalina.out flag

OCF_RESKEY_catalina_rotatetime=Time span of the rotate catalina.out

Time span of the rotate catalina.out

ocf:VIPArip (7)

ocf:VIPArip — Manages a virtual IP address through RIP2

Synopsis

```
OCF_RESKEY_ip=string [OCF_RESKEY_nic=string]  
[OCF_RESKEY_zebra_binary=string] [OCF_RESKEY_ripd_binary=string]  
VIPArip [start | stop | monitor | validate-all | meta-data]
```

Description

Virtual IP Address by RIP2 protocol. This script manages IP alias in different subnet with quagga/ripd. It can add an IP alias, or remove one.

Supported Parameters

OCF_RESKEY_ip=The IP address in different subnet
The IPv4 address in different subnet, for example "192.168.1.1".

OCF_RESKEY_nic=The nic for broadcast the route information
The nic for broadcast the route information. The ripd uses this nic to broadcast the route informaton to others

OCF_RESKEY_zebra_binary=zebra binary
Absolute path to the zebra binary.

OCF_RESKEY_ripd_binary=ripd binary
Absolute path to the ripd binary.

ocf:VirtualDomain (7)

ocf:VirtualDomain — Manages virtual domains through the libvirt virtualization framework

Synopsis

```
OCF_RESKEY_config=string [OCF_RESKEY_hypervisor=string]
[OCF_RESKEY_force_stop=boolean]
[OCF_RESKEY_migration_transport=string]
[OCF_RESKEY_monitor_scripts=string] VirtualDomain [start | stop | status
| monitor | migrate_from | migrate_to | meta-data | validate-all]
```

Description

Resource agent for a virtual domain (a.k.a. domU, virtual machine, virtual environment etc., depending on context) managed by libvirt.

Supported Parameters

OCF_RESKEY_config=Virtual domain configuration file
Absolute path to the libvirt configuration file, for this virtual domain.

OCF_RESKEY_hypervisor=Hypervisor URI
Hypervisor URI to connect to. See the libvirt documentation for details on supported URI formats. The default is system dependent.

OCF_RESKEY_force_stop=Always force shutdown on stop
Always forcefully shut down ("destroy") the domain on stop. The default behavior is to resort to a forceful shutdown only after a graceful shutdown attempt has failed. You should only set this to true if your virtual domain (or your virtualization backend) does not support graceful shutdown.

`OCF_RESKEY_migration_transport=Remote` hypervisor transport

Transport used to connect to the remote hypervisor while migrating. Please refer to the libvirt documentation for details on transports available. If this parameter is omitted, the resource will use libvirt's default transport to connect to the remote hypervisor.

`OCF_RESKEY_monitor_scripts=`space-separated list of monitor scripts

To additionally monitor services within the virtual domain, add this parameter with a list of scripts to monitor. Note: when monitor scripts are used, the start and migrate_from operations will complete only when all monitor scripts have completed successfully. Be sure to set the timeout of these operations to accommodate this delay.

ocf:vmware (7)

ocf:vmware — Manages VMWare Server 2.0 virtual machines

Synopsis

```
[OCF_RESKEY_vmxpath=string] [OCF_RESKEY_vimshbin=string] vmware  
[start | stop | monitor | meta-data]
```

Description

OCF compliant script to control vmware server 2.0 virtual machines.

Supported Parameters

OCF_RESKEY_vmxpath=VMX file path
VMX configuration file path

OCF_RESKEY_vimshbin=vmware-vim-cmd path
vmware-vim-cmd executable path

ocf:WAS6 (7)

ocf:WAS6 — Manages a WebSphere Application Server 6 instance

Synopsis

```
[OCF_RESKEY_profile=string] WAS6 [start | stop | status | monitor | validate-all |  
meta-data | methods]
```

Description

Resource script for WAS6. It manages a Websphere Application Server (WAS6) as an HA resource.

Supported Parameters

OCF_RESKEY_profile=profile name
The WAS profile name.

ocf:WAS (7)

ocf:WAS — Manages a WebSphere Application Server instance

Synopsis

[OCF_RESKEY_config=string] [OCF_RESKEY_port=integer] WAS [start | stop | status | monitor | validate-all | meta-data | methods]

Description

Resource script for WAS. It manages a Websphere Application Server (WAS) as an HA resource.

Supported Parameters

OCF_RESKEY_config=configuration file
The WAS-configuration file.

OCF_RESKEY_port=port
The WAS-(snoop)-port-number.

ocf:WinPopup (7)

ocf:WinPopup — Sends an SMB notification message to selected hosts

Synopsis

[OCF_RESKEY_hostfile=string] WinPopup [start | stop | status | monitor | validate-all | meta-data]

Description

Resource script for WinPopup. It sends WinPopups message to a sysadmin's workstation whenever a takeover occurs.

Supported Parameters

OCF_RESKEY_hostfile=Host file

The file containing the hosts to send WinPopup messages to.

ocf:Xen (7)

ocf:Xen — Manages Xen unprivileged domains (DomUs)

Synopsis

```
[OCF_RESKEY_xmfile=string] [OCF_RESKEY_name=string]  
[OCF_RESKEY_shutdown_timeout=boolean]  
[OCF_RESKEY_allow_mem_management=boolean]  
[OCF_RESKEY_reserved_Dom0_memory=string]  
[OCF_RESKEY_monitor_scripts=string] Xen [start | stop | migrate_from |  
migrate_to | monitor | meta-data | validate-all]
```

Description

Resource Agent for the Xen Hypervisor. Manages Xen virtual machine instances by mapping cluster resource start and stop, to Xen create and shutdown, respectively. A note on names We will try to extract the name from the config file (the xmfile attribute). If you use a simple assignment statement, then you should be fine. Otherwise, if there's some python acrobacy involved such as dynamically assigning names depending on other variables, and we will try to detect this, then please set the name attribute. You should also do that if there is any chance of a pathological situation where a config file might be missing, for example if it resides on a shared storage. If all fails, we finally fall back to the instance id to preserve backward compatibility. Para-virtualized guests can also be migrated by enabling the meta_attribute allow-migrate.

Supported Parameters

OCF_RESKEY_xmfile=Xen control file

Absolute path to the Xen control file, for this virtual machine.

OCF_RESKEY_name=Xen DomU name

Name of the virtual machine.

OCF_RESKEY_shutdown_timeout=Shutdown escalation timeout

The Xen agent will first try an orderly shutdown using `xm shutdown`. Should this not succeed within this timeout, the agent will escalate to `xm destroy`, forcibly killing the node. If this is not set, it will default to two-third of the stop action timeout. Setting this value to 0 forces an immediate destroy.

OCF_RESKEY_allow_mem_management=Use dynamic memory management

This parameter enables dynamic adjustment of memory for start and stop actions used for Dom0 and the DomUs. The default is to not adjust memory dynamically.

OCF_RESKEY_reserved_Dom0_memory=Minimum Dom0 memory

In case memory management is used, this parameter defines the minimum amount of memory to be reserved for the dom0. The default minimum memory is 512MB.

OCF_RESKEY_monitor_scripts=list of space separated monitor scripts

To additionally monitor services within the unprivileged domain, add this parameter with a list of scripts to monitor. NB: In this case make sure to set the start-delay of the monitor operation to at least the time it takes for the DomU to start all services.

ocf:Xinetd (7)

ocf:Xinetd — Manages an Xinetd service

Synopsis

```
[OCF_RESKEY_service=string] Xinetd [start | stop | restart | status | monitor |  
validate-all | meta-data]
```

Description

Resource script for Xinetd. It starts/stops services managed by xinetd. Note that the xinetd daemon itself must be running: we are not going to start it or stop it ourselves. Important: in case the services managed by the cluster are the only ones enabled, you should specify the -stayalive option for xinetd or it will exit on Heartbeat stop. Alternatively, you may enable some internal service such as echo.

Supported Parameters

OCF_RESKEY_service=service name
The service name managed by xinetd.

V. 附錄

設定簡單測試資源的範例

本章提供了設定簡單資源 (IP 位址) 組態的基本範例。示範了如何使用 Pacemaker GUI 或 `crm` 指令行工具兩種方法來執行此作業。

在以下範例中，我們假設您已依第 3 章「使用 *YaST* 的安裝與基本設定」[第 19 頁] 中所述設定叢集，且該叢集至少包含兩個節點。如需使用 Pacemaker GUI 與 `crm` 外圍程序設定叢集資源的介紹和綜覽，請參閱以下幾章：

- 設定和管理叢集資源 (GUI) [第 53 頁]
- 設定和管理叢集資源 (指令行) [第 83 頁]

A.1 使用 GUI 設定資源

建立範例叢集資源並將其移轉至其他伺服器，可協助您進行測試以確保叢集正常運作。設定和移轉的簡易資源為 IP 位址。

過程 A.1 建立 IP 位址叢集資源

- 1 啟動 Pacemaker GUI，並依第 5.1.1 節「連接至叢集」[第 54 頁] 中所述登入叢集。
- 2 在左側窗格中，切換至「資源」檢視窗，然後在右側窗格中選取要修改的群組，並按一下「編輯」。下一個視窗會顯示基本的群組參數，以及已為該資源定義的中繼屬性與原始資源。
- 3 按一下「原始資源」索引標籤，然後按一下「新增」。

4 在下一個對話方塊中，設定以下參數以將 IP 位址新增為群組的子資源：

4a 輸入唯一的 ID。例如 myIP。

4b 在「類別」清單中，選取「*ocf*」做為資源代辦類別。

4c 對於 OCF 資源代辦的「提供者」，選取「*heartbeat*」。

4d 在「類型」清單中，選取「*IPaddr*」做為資源代辦。

4e 按「下一步」。

4f 在「例項屬性」索引標籤中，選取「*IP*」項目並按一下「編輯」(或在「*IP*」項目上連按兩下)。

4g 對於「值」，請輸入所需的 IP 位址 (例如 10.10.0.1)，然後按一下「確定」。

4h 「新增」一個新的例項屬性，並將 *nic* 指定為「名稱」，將 *eth0* 指定為「值」，然後按一下「確定」。

名稱與值具體取決於您的硬體組態以及安裝 High Availability Extension 軟體期間為媒體組態所選選項。

5 按需要設定了所有參數後，按一下「確定」以完成該資源的組態設定。組態對話方塊會關閉，同時主視窗會顯示修改的資源。

若要使用 Pacemaker GUI 啟動資源，請選取左側窗格中的「管理」。在右側窗格中，在資源上按一下滑鼠右鍵，然後選取「啟動」(或從工具列中將資源啟動)。

若要將 IP 位址資源移轉至其他節點 (*saturn*)，請繼續執行以下步驟：

過程 A.2 將資源移轉至其他節點

1 切換至左側窗格中的「管理」檢視窗，然後在右側窗格中的 IP 位址資源上按一下滑鼠右鍵，並選取「移轉資源」。

2 在新視窗中，從「至節點」下拉式清單中選取 *saturn*，以將所選資源移至節點 *saturn*。

- 3 若只想暫時移轉資源，請啟用「*持續時間*」，並輸入資源移轉至新節點後應保留的時間。
- 4 按一下「*確定*」以確認移轉。

A.2 手動設定資源

資源是電腦提供之任何類型的服務。當由 LSB 程序檔、OCF 程序檔或舊版 Heartbeat 1 資源等 RA (資源代辦) 控制資源時，High Availability 才會識別這些資源。可以使用 `crm` 指令或以 `resources` 區段的 CIB (叢集資訊庫) 中的 XML 設定所有資源。如需可用資源的綜覽，請參閱第 19 章「*HA OCF Agents*」[第 243 頁]。

若要新增 IP 位址 10.10.0.1 做為目前組態的資源，請使用 `crm` 指令：

過程 A.3 建立 IP 位址叢集資源

- 1 開啟外圍程序，切換為 `root` 身分。
- 2 輸入 `crm configure` 開啟內部外圍程序。
- 3 建立 IP 位址資源：

```
crm(live)configure# resource
primitive myIP ocf:heartbeat:IPaddr params ip=10.10.0.1
```

注意

使用 High Availability 設定資源時，相同的資源不應由 `init` 啟始化。High Availability 負責所有服務的啟動或停止動作。

若組態設定成功，`crm_mon` 中會顯示一個新的資源，該資源將在叢集的一個隨機節點上啟動。

若要將資源移轉至其他節點，請執行以下步驟：

過程 A.4 將資源移轉至其他節點

- 1 啟動外圍程序，並成為使用者 `root`。

2 將資源 myip 移轉至節點 saturn:

```
crm resource migrate myIP saturn
```


B

將叢集升級到產品的最新版本

如果現有的叢集是基於 SUSE® Linux Enterprise Server 10，可以對其進行更新，讓其使用 High Availability Extension 在 SUSE Linux Enterprise Server 11 或 11 SP1 上執行。

若要從 SUSE Linux Enterprise Server 10 移轉到 SUSE Linux Enterprise Server 11 或 11 SP1，所有叢集節點都必須離線，且叢集必須做為一個整體移轉 — 不能同時使用 SUSE Linux Enterprise Server 10/SUSE Linux Enterprise Server 11 上執行的叢集。

B.1 從 SLES 10 升級到 SLEHA 11

為了方便起見，SUSE® Linux Enterprise High Availability Extension 中提供了一個 `hb2openais.sh` 程序檔，當您將資料從 Heartbeat 移至 OpenAIS 叢集堆疊期間，可用它來轉換資料。此程序檔可剖析儲存在 `/etc/ha.d/ha.cf` 中的組態，並產生適用於 OpenAIS 叢集堆疊的新組態檔案。此外，它還能調整 CIB 以符合 OpenAIS 慣例、轉換 OCFS2 檔案系統，以及用 cLVM 取代 EVMS。所有 EVMS2 容器都會轉換成 cLVM2 磁碟區。對於 CIB 中現有資源中參考的磁碟區群組，會為其建立新的 LVM 資源。

若要將叢集從 SUSE Linux Enterprise Server 10 SP3 成功移轉至 SUSE Linux Enterprise Server 11，需要執行以下步驟：

1. 準備 SUSE Linux Enterprise Server 10 SP3 叢集 [第342頁]
2. 更新至 SUSE Linux Enterprise 11 [第343頁]

3. 測試轉換 [第344頁]

4. 轉換資料 [第344頁]

成功完成轉換後，可以讓更新後的叢集重新恢復線上狀態。

注意：更新後回復

更新至 SUSE Linux Enterprise Server 11 後，不能重新回復到 SUSE Linux Enterprise Server 10。

B.1.1 準備與備份

在將叢集更新至產品的下一版本並相應地轉換資料之前，需要對目前的叢集做一些準備工作。

過程 B.1 準備 SUSE Linux Enterprise Server 10 SP3 叢集

- 1 登入叢集。
- 2 檢閱 Heartbeat 組態檔案 `/etc/ha.d/ha.cf`，並檢查所有通訊媒體是否支援多路廣播。
- 3 確認以下檔案在所有節點上都相同：`/etc/ha.d/ha.cf` 和 `/var/lib/heartbeat/crm/cib.xml`。
- 4 對每個節點執行 `rcheartbeat stop`，使它們離線。
- 5 在更新至最新版本之前，除了需要依建議執行一般系統備份之外，還需備份以下檔案，因為在更新至 SUSE Linux Enterprise Server 後需要使用它們來執行轉換程序檔：

- `/var/lib/heartbeat/crm/cib.xml`
- `/var/lib/heartbeat/hostcache`
- `/etc/ha.d/ha.cf`
- `/etc/logd.cf`

- 6 如果您擁有 EVMS2 資源，則將非 LVM EVMS2 磁碟區轉換為 SUSE Linux Enterprise Server 10 上相容的磁碟區。在轉換過程中(請參閱第 B.1.3 節「資料轉換」[第343頁])，這些磁碟區將轉變並歸入 LVM2 磁碟區群組。轉換之後，請務必使用 `vgchange -c y` 將各個磁碟區群組標示為 High Availability 叢集的成員。

B.1.2 更新/安裝

準備好叢集並備份檔案之後，便可開始將叢集節點更新至產品的下一版本。您也可以不執行更新，而是在叢集節點上全新安裝 SUSE Linux Enterprise 11。

過程 B.2 更新至 SUSE Linux Enterprise 11

- 1 在所有叢集節點上執行從 SUSE Linux Enterprise Server 10 SP3 至 SUSE Linux Enterprise Server 11 的更新。如需如何更新產品的相關資訊，請參閱《SUSE Linux Enterprise Server 11 部署指南》中的「更新 SUSE Linux Enterprise」一章。

或者，您也可以在所有叢集節點上全新安裝 SUSE Linux Enterprise Server 11。

- 2 在所有叢集節點上，以 SUSE Linux Enterprise Server 為基礎，將 SUSE Linux Enterprise High Availability Extension 11 安裝為附加產品。如需詳細資訊，請參閱第 3.1 節「安裝 High Availability Extension」[第19頁]。

B.1.3 資料轉換

安裝 SUSE Linux Enterprise Server 11 和 High Availability Extension 之後，便可開始轉換資料。High Availability Extension 隨附的轉換程序檔已經過周密設定，但它並不能在完全自動的模式下完成所有設定。它會警告您它所進行的變更，但仍需要您的互動和決定。您需要瞭解叢集的詳細情況，因為最終是由您來確認這些變更是否有意義。轉換程序檔位於 `/usr/lib/heartbeat` 中 (若您使用的是 64 位元系統，則位於 `/usr/lib64/heartbeat` 中)。

注意：執行測試回合

為了讓您熟悉轉換程序，我們強烈建議您先進行轉換測試 (不做任何變更)。您可以使用相同的測試目錄來執行重複的測試回合，但只需複製一次檔案。

過程 B.3 測試轉換

- 1 在其中一個節點上，建立測試目錄並將備份檔案複製到測試目錄中：

```
$ mkdir /tmp/hb2openais-testdir
$ cp /etc/ha.d/ha.cf /tmp/hb2openais-testdir
$ cp /var/lib/heartbeat/hostcache /tmp/hb2openais-testdir
$ cp /etc/logd.cf /tmp/hb2openais-testdir
$ sudo cp /var/lib/heartbeat/crm/cib.xml /tmp/hb2openais-testdir
```

- 2 使用以下指令開始進行測試回合

```
$ /usr/lib/heartbeat/hb2openais.sh -T /tmp/hb2openais-testdir -U
```

若您使用的是 64 位元系統，請使用以下指令執行測試：

```
$ /usr/lib64/heartbeat/hb2openais.sh -T /tmp/hb2openais-testdir -U
```

- 3 讀取並驗證產生的 openais.conf 和 cib-out.xml 檔案：

```
$ cd /tmp/hb2openais-testdir
$ less openais.conf
$ crm_verify -V -x cib-out.xml
```

如需轉換階段的詳細資訊，請參閱所安裝之 High Availability Extension 中的 /usr/share/doc/packages/pacemaker/README.hb2openais。

過程 B.4 轉換資料

執行測試回合並檢查輸出之後，便可以開始轉換資料。您只需在一個節點上執行轉換。主叢集組態(CIB)會自動複寫到其他節點中。轉換程序檔會自動複製需要複寫的所有其他檔案。

- 1 確定 root 可存取的所有節點上都在執行 sshd，以便轉換程序檔能成功將檔案複製到其他叢集節點。
- 2 確定所有 ocfs2 檔案系統已卸載。
- 3 High Availability Extension 隨附了預設的 OpenAIS 組態檔案。若在執行以下步驟時，不希望預設的組態被覆寫，請複製一份 /etc/ais/openais.conf 組態檔案。
- 4 以 root 身分啟動轉換程序檔。若正在使用 sudo，請使用 -u 選項指定有特權的使用者：

```
$ /usr/lib/heartbeat/hb2openais.sh -u root
```

程序檔會依據 `/etc/ha.d/ha.cf` 中儲存的組態，產生適用於 OpenAIS 叢集堆疊的新組態檔案 `/etc/ais/openais.conf`。程序檔還會分析 CIB 組態，並讓您知道您的叢集組態是否需要因從 Heartbeat 變為 OpenAIS 而變更。所有檔案處理都是在執行轉換的節點上完成，然後再複寫到其他節點。

5 按照螢幕上的指示執行操作。

成功完成轉換後，依第 3.3 節「連線叢集」[第27頁]中所述啟動新的叢集堆疊。

完成升級後，就不能重新回復到 SUSE Linux Enterprise Server 10。

B.1.4 如需更多資訊

如需轉換程序檔和轉換階段的更多詳細資料，請參閱所安裝之 High Availability Extension 中的 `/usr/share/doc/packages/pacemaker/README.hb2openais`。

B.2 從 SLEHA 11 升級到 SLEHA 11 SP1

若要將現有叢集從 SUSE Linux Enterprise High Availability Extension 11 成功移轉到 11 SP1，可以執行「滾存升級」，即按節點逐個升級。由於在 SUSE Linux Enterprise High Availability Extension 11 SP1 中，主要叢集組態檔案已從 `/etc/ais/openais.conf` 變更為 `/etc/corosync/corosync.conf`，因此會有一個程序檔負責必要的轉換。這些轉換會在更新 openais 套件之後自動進行。

過程 B.5 執行滾存升級

重要：更新軟體套件

如果要更新執行中叢集之節點上的軟體套件，應先停止該節點上的叢集堆疊，然後再開始更新軟體。若要停止叢集堆疊，請以 `root` 身分登入節點，然後輸入 `rcopenais stop`。

如果軟體更新期間 OpenAIS/Corosync 仍在執行，可能會出現無法預知的結果，如圍籬區隔使用中的節點。

- 1 以 root 身分登入要升級和停止 OpenAIS 的節點：

```
rcopenais stop
```

- 2 檢查您的系統備份是否最新、能否還原。
- 3 執行從 SUSE Linux Enterprise Server 11 到 SUSE Linux Enterprise Server 11 SP1 以及從 SUSE Linux Enterprise High Availability Extension 11 到 SUSE Linux Enterprise High Availability Extension 11 SP1 的升級。如需如何更新產品的相關資訊，請參閱《SUSE Linux Enterprise Server 11 SP1 部署指南》中的「更新 *SUSE Linux Enterprise*」一章。
- 4 在已升級的節點上重新啟動 OpenAIS/Corosync，讓節點重新加入叢集：

```
rcopenais start
```
- 5 讓下一個節點離線，並對其重複執行此程序。

最新功能

下面幾節將概要介紹版本與版本之間發生的軟體修改。文中彙總了基本設定是否已完全重新設定、組態檔案是否已移至他處，或者是否做了其他大幅變更等資訊。

C.1 版本 10 SP3 至版本 11

借助 SUSE Linux Enterprise Server 11，叢集堆疊已從 Heartbeat 變更為 OpenAIS。OpenAIS 實作工業標準 API，即 Service Availability Forum (服務可用性論壇) 所發佈的應用程式介面規範 (Application Interface Specification, AIS)。SUSE Linux Enterprise Server 10 中的叢集資源管理員被保留下來，但功能獲得顯著提升，且已移植到 OpenAIS，現在稱為 Pacemaker。

如需 SUSE® Linux Enterprise Server 10 SP3 升級至 SUSE Linux Enterprise Server 11 後，High Availability 元件發生之變更的詳細資料，請參閱下面幾節。

C.1.1 新特性和新功能

移轉限定值與故障逾時

現在，High Availability Extension 提出了移轉限定值與故障逾時的概念。您可定義大量資源故障，發生這些故障後會將資源移轉至新節點。預設情況下，在管理員手動重設資源的故障計數前，將不再允許節點執行失敗資源。不過，還可以透過設定資源的 `failure-timeout` 選項，讓資源過期。

資源和作業預設值

現在，您可為資源選項和作業設定全域預設值。

支援離線組態變更

通常，在進行重大組態更新前，需要先預覽一系列變更的效果。現在，您可以建立組態的「陰影」副本，並使用指令行介面對其進行編輯，然後再提交，以對使用中的叢集組態進行重大變更。

重複使用作業的規則、選項和設定

規則、`instance_attributes`、`meta_attributes` 和多組作業只需定義一次，便可在多處進行參考。

將 XPath 運算式用於 CIB 中的特定作業

現在，CIB 接受 XPath 式 `create`、`modify` 和 `delete` 作業。如需詳細資訊，請參閱 `cibadmin` 說明文字。

多維並存和順序限制

若要建立一組並存資源，以前可以定義一個資源群組 (但它並不總能準確反映出設計目地)，或將每個關係定義為個別限制，但隨著資源和組合數目的增長，此方法會導致限制爆炸。現在，您還可以透過定義 `resource_sets` 使用並存限制的替代形式。

從非叢集機器連接至 CIB

只要機器上安裝了 `Pacemaker`，即使機器本身不屬於叢集，也可能會連接該叢集。

在已知時間觸發週期性動作

預設會相對於資源啟動時間排定週期性動作，但並非所有情況都適合如此操作。若要指定作業應相對於的日期/時間，請設定作業的 `interval-origin` (間隔起始點)。叢集使用此時間點計算正確的 `start-delay` (啟動延遲時間)，即作業將在「起時點 + (間隔 * N)」時發生。

C.1.2 變更的特性與功能

資源和叢集選項的命名慣例

現在，所有資源和叢集選項皆使用破折號 (-) 取代底線 (_)。例如，`master_max` 中繼選項已重新命名為 `master-max`。

重新命名 master_slave 資源

master_slave 資源已重新命名為 master。主要資源是一種特殊類型的複製資源，可以兩種模式之一進行操作。

屬性的容器標籤

已移除 attributes 容器標籤。

先決條件的作業欄位

pre-req 作業欄位已重新命名為 requires。

作業間隔

所有作業都必須設定有間隔。對於啟動/停止動作，必須將間隔設定為 0 (零)。

並存和順序限制的屬性

為了清楚起見，重新命名了並存和順序限制的屬性。

因發生故障而進行移轉的叢集選項

resource-failure-stickiness 叢集選項已由 migration-threshold 叢集選項取代。並請參閱 移轉限定值與故障逾時 [第347頁]。

指令行工具的引數

指令行工具的引數已經過一致化處理。並請參閱 資源和叢集選項的命名慣例 [第348頁]。

驗證和剖析 XML

叢集組態以 XML 格式撰寫。現在已使用功能更為強大的 RELAX NG 綱要取代了「文件類型定義」(DTD)，用來定義結構和內容的模式。libxml2 用做剖析程式。

id 欄位

id 欄位現在為 XML ID，它們有以下限制：

- ID 不能含有冒號。
- ID 不能以數字開頭。
- ID 必須是全域唯一的 (而不僅僅對於標籤唯一)。

參考其他物件

有些欄位 (例如因需要參考資源而受到限制的欄位) 是 IDREF。這表示它們必須參考現有資源或物件才能使組態生效。因此，移除在其他地方參考之物件的動作將會失敗。

C.1.3 已移除的特性與功能

設定資源中繼選項

資源中繼選項不能再做為頂層屬性進行設定。請轉為使用中繼屬性。並請參閱 `crm_resource(8)` [第217頁]。

設定全域預設值

資源和作業預設值將不再從 `crm_config` 讀取。

C.2 版本 11 至版本 11 SP1

叢集組態檔案

主要叢集組態檔案已從 `/etc/ais/openais.conf` 變更至 `/etc/corosync/corosync.conf`。兩個檔案非常相似。從 SUSE Linux Enterprise High Availability Extension 11 升級到 SP1 後，會有一個程序檔負責處理兩個檔案間的細微差別。如需 OpenAIS 與 Corosync 間之關係的詳細資訊，請參閱 <http://www.corosync.org/doku.php?id=faq:why>。

滾存升級

為了在移轉現有叢集時儘可能減少停機時間，可以利用 SUSE Linux Enterprise High Availability Extension 執行從 SUSE Linux Enterprise High Availability Extension 11 到 11 SP1 的「滾存升級」。在您按節點逐一升級的過程中，叢集仍處於連線狀態。

自動叢集部署

為了簡化叢集的部署，AutoYaST 允許您複製現有的節點。AutoYaST 這套系統可借助包含安裝和組態資料的 AutoYaST 設定檔，在不需要使用者介入的情況下自動安裝一或多個 SUSE Linux Enterprise 系統。此設定檔會告訴 AutoYaST 要安裝什麼，及如何設定安裝的系統，以便最後獲得的系統萬事具備。進行大規模部署時，可以透過多種方式使用此設定檔。

傳輸組態檔案

SUSE Linux Enterprise High Availability Extension 隨附 Csync2 工具，可在叢集的所有節點間複製組態檔案。它可以處理任意數量的主機，也可以只在某些主機間同步檔案。使用 YaST 可以設定主機名稱以及要使用 Csync2 同步的檔案。

用於叢集管理的 Web 介面

High Availability Extension 現在還包含 HA Web Konsole，這是一個用於執行管理任務的 Web 使用者介面。利用它，您在非 Linux 機器上也可以監控並管理 Linux 叢集。如果您的系統不提供或不允許使用圖形使用者介面，這也是一個理想的解決方案。

資源組態的範本

現在，使用指令行介面建立和設定資源時，可以從各種資源範本中進行選擇，以更加便捷地完成組態設定。

資源的負載配置

透過定義某節點提供的容量和某資源需要的容量，並在叢集中選擇一個配置策略，系統可根據資源的載入影響來配置資源，防止叢集效能下降。

叢集感知的主動/主動 RAID1

使用 cmirrord 可以從兩個獨立的 SAN 建立災難恢復儲存組態。

唯讀 GFS2 支援

為了更輕鬆地從 GFS2 移轉到 OCFS2，您可以唯讀模式掛接 GFS2 檔案系統，以便將資料複製到 OCFS2 檔案系統。SUSE Linux Enterprise High Availability Extension 完全支援 OCFS2。

對 OCFS2 的 SCTP 支援

如果設定了備援環狀網路，則不必借助網路設備 Bonding，OCFS2 和 DLM 便可自動透過 SCTP 使用備援通訊路徑。

儲存保護

為了增強安全性以防止儲存區的資料被損毀，您可以結合使用 IO 圍籬區隔 (由 external/sbd 圍籬區隔設備提供) 和 sfex 資源代辦，以確保獨佔式資源存取。

Samba 叢集

High Availability Extension 現在支援 CTDB，即對 Trivial Database 的叢集實作。因此，您可以設定叢集化 Samba 伺服器 — 同時為異質環境提供 High Availability 解決方案。

用於 IP 負載平衡的 YaST 模組

新模組可以借助圖形使用者介面設定核心負載平衡的組態。它可以做為 `ldirectord` 使用者空間精靈 (用於管理 Linux Virtual Server 並監控實際的伺服器) 的前端。

GNU 授權

D

本附錄包含 GNU 通用公共授權 (General Public License) 和 GNU 自由文件授權 (Free Documentation License)。

GNU General Public License

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc. 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU GENERAL PUBLIC LICENSE TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The 「 Program 」, below, refers to any such program or work, and a 「 work based on the Program 」 means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term 「 modification 」.) Each licensee is addressed as 「 you 」.

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and 「any later version」, you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the 「copyright」 line and a pointer to where the full notice is found.

one line to give the program's name and an idea of what it does. Copyright (C) yyyy name of author

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

```
Gnomovision version 69, Copyright (C) year name of author
Gnomovision comes with ABSOLUTELY NO WARRANTY; for details
type `show w'. This is free software, and you are welcome
to redistribute it under certain conditions; type `show c'
for details.
```

The hypothetical commands 'show w' and 'show c' should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than 'show w' and 'show c'; they could even be mouse-clicks or menu items--whatever suits your program.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a 「copyright disclaimer」 for the program, if necessary. Here is a sample; alter the names:

```
Yoyodyne, Inc., hereby disclaims all copyright
interest in the program `Gnomovision'
(which makes passes at compilers) written
by James Hacker.
```

```
signature of Ty Coon, 1 April 1989
Ty Coon, President of Vice
```

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License [<http://www.fsf.org/licenses/lgpl.html>] instead of this License.

GNU Free Documentation License

Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc. 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of 「copyleft」, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The 「 Document 」, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as 「 you 」. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A 「 Modified Version 」 of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A 「 Secondary Section 」 is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The 「 Invariant Sections 」 are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The 「 Cover Texts 」 are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A 「 Transparent 」 copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not 「 Transparent 」 is called 「 Opaque 」.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The 「 Title Page 」 means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, 「 Title Page 」 means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section 「 Entitled XYZ 」 means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as 「 Acknowledgements 」, 「 Dedications 」, 「 Endorsements 」, or 「 History 」.) To 「 Preserve the Title 」 of such a section when you modify the Document means that it remains a section 「 Entitled XYZ 」 according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled 「History」, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled 「History」 in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the 「History」 section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled 「Acknowledgements」 or 「Dedications」, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled 「Endorsements」. Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled 「Endorsements」 or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled 「Endorsements」, provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled 「 History 」 in the various original documents, forming one section Entitled 「 History 」 ; likewise combine any sections Entitled 「 Acknowledgements 」 , and any sections Entitled 「 Dedications 」 . You must delete all sections Entitled 「 Endorsements 」 .

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled 「 Acknowledgements 」 , 「 Dedications 」 , or 「 History 」 , the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License 「 or any later version 」 applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
only as published by the Free Software Foundation;
with the Invariant Section being this copyright notice and license.
A copy of the license is included in the section entitled “GNU
Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with...Texts.” line with this:

with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

術語

主動/主動、主動/被動

關於服務如何在節點上執行的概念。主動-被動情境是指有一或多項服務正在主動節點上執行，而被動節點則等待主動節點失敗。主動-主動表示各節點同時處於主動與被動狀態。

磁簇(與硬碟，磁片有關時)

高效能叢集是指共同承擔應用程式負載以更快獲得結果的一個電腦 (真實或虛擬) 群組。高可用性叢集的主要用途是最大可能地確保服務的可用性。

叢集資訊庫 (CIB)

表示整個叢集組態與狀態(節點成員、資源、限制等)以 XML 編寫並存放在記憶體中。主要 CIB 會在指定協調者(DC)[第361頁]上保留並維護，並複製到其他節點。

叢集分割區

當一個或多個節點與叢集其餘節點間的通訊失敗時，就會出現叢集分割區。叢集分割區的節點仍為使用中狀態且能與彼此通訊，但它們無法察覺不能與之通訊的節點。由於無法確認其他分割區的遺失，將出現電腦分裂情境 (另請參閱電腦分裂 [第364頁])。

叢集資源管理員 (CRM)

負責協調所有非本地互動的主要管理實體。叢集的各節點均有各自的 CRM，但 DC 上所執行的是選出後負責將決策傳送到其他非本地 CRM 並處理其輸入的 CRM。一個 CRM 可與許多元件進行互動，包括所在節點及其他節點上的本地資源管理員、非本地 CRM、管理指令、圍籬區隔功能及成員層。

共識叢集成員 (CCM)

CCM 決定由哪些節點組成叢集並在叢集中共享此資訊。任何新的新增項及任何節點或最低節點數的遺失均透過 CCM 傳送。CCM 模組將在叢集的每個節點上執行。

指定協調者 (DC)

「主要」節點。此節點為保留 CIB 主要副本的所在。所有其他節點均從目前 DC 取得其組態和資源配置資訊。成員發生變更後將從叢集的所有節點中選出 DC。

分散式鎖定管理員 (DLM)

DLM 會協調叢集檔案系統的磁碟存取，並管理檔案鎖定以提高效能和可用性。

分散式複製區塊設備 (drbd)

DRBD 是專用於建立 High Availability 叢集的區塊設備。整個區塊設備透過專屬網路鏡像，並被視做網路 RAID-1。

容錯移轉

當一台機器上的資源或節點發生故障且受影響的資源將在其他節點上啟動時，會發生此情況。

圍籬區隔

描述阻止非叢集成員存取共享資源的概念。該功能可透過停止(關閉)「行為錯誤的」節點以防止它導致問題，從節點鎖定狀態未定的資源，或多種其他方式實現。此外，節點的圍籬區隔與資源的圍籬區隔有所不同。

Heartbeat 資源代辦

Heartbeat 資源代辦曾在 Heartbeat 版本 1 中廣泛使用，目前已廢棄，但在版本 2 中仍受支援。Heartbeat 資源代辦可執行 start、stop 與 status 作業，其位於 /etc/ha.d/resource.d 或 /etc/init.d。如需 Heartbeat 資源代辦的詳細資訊，請參閱 <http://www.linux-ha.org/HeartbeatResourceAgent> (另請參閱 OCF 資源代辦 [第363頁])。

本地資源管理員 (LRM)

本地資源管理員 (LRM) 負責在資源上執行作業。它使用資源代辦程序檔來執行工作。LRM 是「無用的」，自己並不知道任何規則，需要 DC 告訴它該做什麼。

LSB 資源代辦

LSB 資源代辦是標準 LSB init 程序檔。LSB init 程序檔不限用於高可用性網路位置。任何 LSB 相容的 Linux 系統均可使用 LSB init 程序檔來控制服務。任何 LSB 資源代辦均支援 start、stop、restart、status 與 force-reload 選項，並可選擇性地提供 try-restart 與 reload。LSB 資源代辦位於 /etc/init.d。如需更多有關 LSB 資源代辦及實際規格的詳細資訊，請參閱 <http://www.linux-ha.org/LSBResourceAgent> 與 http://www.linux-foundation.org/spec/refspecs/LSB_3.0.0/LSB-Core-generic/LSB-Core-generic/iniscriptact.html (另請參閱 OCF 資源代辦 [第363頁] 與 Heartbeat 資源代辦 [第362頁])。

節點

任何屬於叢集成員且對使用者不可見的電腦 (真實或虛擬)。

pingd

ping 精靈。它會使用 ICMP ping 持續聯絡叢集之外的一或多台伺服器。

規則引擎 (PE)

規則引擎會計算出在 CIB 中實作規則變更所需採取的動作。此資訊隨後將被傳遞到異動引擎，然後便會在叢集設定中實作規則變更。PE 始終在 DC 上執行。

OCF 資源代辦

OCF 資源代辦與 LSB 資源代辦 (init 程序檔) 類似。任何 OCF 資源代辦均必須支援 start、stop 與 status (有時稱為 monitor) 選項。此外，他們還支援 metadata 選項，該選項會以 XML 返回資源代辦類型的描述。亦可支援其他選項，但並非強制。OCF 資源代辦位於 /usr/lib/ocf/resource.d/ 提供者。如需有關 OCF 資源代辦及規格草稿的詳細資訊，請參閱 <http://www.linux-ha.org/OCFResourceAgent> 和 <http://www.openocf.org/cgi-bin/viewcvs.cgi/specs/ra/resource-agent-api.txt?rev=HEAD> (另請參閱 Heartbeat 資源代辦 [第 362 頁])。

最低節點數

在叢集中，如果叢集分割區擁有絕大多數節點 (或投票)，則其會被定義為具有最低節點數 (即「到達法定數目」)。最低節點數可準確辨識一個分割區。此部分演算法可阻止多個斷線分割區或節點繼續和導致資料與服務毀損 (電腦分裂)。最低節點數是圍籬區隔的先決條件，而圍籬區隔可確保最低節點數是唯一的。

資源

Heartbeat 已知的任何類型的服務或應用程式。範例包括 IP 位址、檔案系統或資料庫。

資源代辦 (RA)

資源代辦 (RA) 是一個程序檔，做為代理使用以管理資源。資源代辦有三種類型：OCF (開放叢集架構) 資源代辦、LSB 資源代辦 (標準 LSB init 程序檔) 與 Heartbeat 資源代辦 (Heartbeat v1 資源)。

單一故障點 (SPOF)

單一故障點 (SPOF) 是叢集的任一元件，若其發生故障，會觸發整個叢集故障。

電腦分裂

在該情境下，叢集節點將被分成兩個或多個彼此不知的群組 (透過軟體或硬體故障)。STONITH 可以防止電腦分裂狀況對整個叢集產生不良影響。也稱為「分割的叢集」情境。

DRBD 中也有「電腦分裂」一詞，但表示的是兩個節點包含不同的資料。

STONITH

「Shoot the other node in the head」的縮寫，基本上，它會關閉行為錯誤的節點以防止其在叢集中導致問題。

轉換引擎 (TE)

轉換引擎 (TE) 可從 PE 獲取規則指令，並加以執行。TE 始終在 DC 上執行。它會從該處指示其他節點上的本地資源管理員要採取的動作。