Red Hat Enterprise Linux 7 High Availability 外掛程式總覽

Red Hat Enterprise Linux 7 的 High Availability 外掛程式總覽

Red Hat Engineering Content Services

Red Hat Enterprise Linux 7 的 High Availability 外掛程式總覽

Red Hat Engineering Content Services
docs-need-a-fix@redhat.com

法律聲明

摘要

《Red Hat High Availability 外掛程式總覽》提供了有關於 Red Hat Enterprise Linux 7 上的 High Availability 外掛程式之總覽。 若要提升您建置 Red Hat High Availability 叢集的能力，您可參考〈Red Hat High Availability Clustering (RH436)〉培訓課程。

內容目錄

章 1. High Availability 外掛程式總覽

High Availability 外掛程式是個為重要生產服務提供高可靠性、延展性和可用性的叢集系統。下列部分提供了有關於 High Availability 外掛程式功能及元件的基本詳述。

1.1. 叢集的基礎資訊

叢集是指兩台以上的電腦（稱為 *節點（node）* 或*成員（member）*）共同運作來完成一項任務。叢集有四種主要類型：

- 儲存裝置（Storage）
- 高可用性（High Availability）
- 負載平衡（Load Balancing）
- 高效能（High Perfomance）

「儲存裝置」叢集會在叢集中的伺服器之間，提供一致的檔案系統映像，以讓伺服器同時讀取和寫入單一共享檔案系統。儲存裝置叢集會藉由將應用程式的安裝及升級限制在單一檔案系統中，以簡化儲存裝置上的管理。此外，當使用一個叢集全域的檔案系統時，儲存裝置叢集可省略複製應用程式資料，並簡化備份和災害復原（disaster recovery）。High Availability 外掛程式提供了儲存裝置叢集，並結合了 Red Hat GFS2（Resilient Storage 外掛程式的一部分）。

High availability 叢集藉由除去單一失敗點（single point of failure）並在節點失效時，將服務由一個叢集節點容錯移轉（fail over）至另一節點上，以提供高可用性的服務。一般來講，high availability 叢集中的服務會（透過讀寫掛載的檔案系統）讀取和寫入資料。因此，一個 high availability 叢集必須能在一個叢集節點由另一叢集節點接收服務控制權時，保留資料的完整性。一個 high availability 叢集中的節點失效，將不會被叢集外的客戶端看見。（high availability 叢集有時亦稱為容錯移轉叢集。）High Availability 外掛程式透過了其 High Availability Service Management 元件 **Pacemaker**，來提供了高可用性的叢集處理。

Load-balancing 叢集會將網路服務請求發送至數個叢集節點上，以平衡叢集節點之間的請求負載。負載平衡機制能提供高效益的延展性，因為您能夠根據負載需求來比對節點的數量。若一個 load-balancing 叢集中有個節點失效，load-balancing 軟體將會偵測到此失效情況，並將請求重新導向至其它叢集節點上。Load-balancing 叢集中的節點失效，不會被叢集外部的客戶端看見。負載平衡可藉由 Load Balancer 外掛程式提供。

High Performance 叢集使用叢集節點來進行同步計算。此叢集能讓應用程式同步運作，藉此增進應用程式的效能。（高效能叢集亦稱為「運算叢集」（computational cluster）或「網格運算」（grid computing）。）

注意

在本文一開始的部分中所概述的叢集類型反應了基本的設定；您的環境可能需要綜合這些叢集的觀念。

此外，Red Hat Enterprise Linux High Availability 外掛程式*僅*支援配置和管理 high availability 伺服器。它*不*支援 high-performance 叢集。

1.2. High Availability 外掛程式簡介

High Availability 外掛程式是個整合式的軟體元件集，它可藉由各種不同的配置方式來建置，以滿足您對於效能、高可用性、負載平衡、延展性、檔案共享與經濟效應上的需求。

High Availability 外掛程式包含了以下主要元件：

- Cluster infrastructure — 提供基礎功能，以讓節點成為叢集並協同運作：配置檔案管理、成員管理、鎖定管理，以及 fencing（隔離）。
- High availability Service Management — 當節點失效時，能將服務由一個叢集節點上，容錯移轉至另一個節點上。
- Cluster administration tools — 用來設定、配置和管理 High Availability 外掛程式的配置與管理工具。此工具可搭配叢集基礎結構（Cluster Infrastructure）元件、high availability 和 Service Management 元件，以及儲存裝置使用。

您可使用下列元件來補充 High Availability 外掛程式：

- Red Hat GFS2（Global File System 2）— Resilient Storage 外掛程式的一部分，它提供了一個叢集檔案系統，以搭配 High Availability 外掛程式使用。GFS2 能在區塊層級中，讓多個節點共享儲存裝置，就如儲存裝置已本地連上了各個叢集節點一般。GFS2 叢集檔案系統需要一個叢集基礎結構（cluster infrastructure）。
- Cluster Logical Volume Manager（CLVM）— 屬於 Resilient Storage 外掛程式的一部分，它提供了叢集儲存裝置的卷冊管理。CLVM 的支援亦需要叢集基礎結構。
- 負載平衡外掛程式（Load Balancer Add-On）— 這是個路由軟體，在網路通訊的第四層（TCP）與第七層（HTTPS）服務中，提供高可用性的負載平衡與容錯功能。負載平衡外掛程式會在冗餘的虛擬路由器叢集中執行，使用負載演算法則將客戶的需求分散到真實伺服器上，使這些真實伺服器以單一虛擬伺服器的形式運作。負載平衡外掛程式不一定要與 Pacemaker 結合使用。

1.3. Pacemaker 總覽

High Availability 外掛程式的叢集基礎結構，為一組電腦（亦稱為*節點*或是*成員*）提供了基礎功能，以讓它們作為叢集進行協作。當叢集透過使用叢集基礎結構形成之後，您便可使用其它元件來滿足您的叢集需求（比方說設定一個叢集，以在一個 GFS2 檔案系統上共享檔案，或是設定服務容錯移轉）。叢集基礎結構能進行下列功能：

- 叢集管理
- 鎖定管理
- 隔離
- 叢集配置管理

1.4. Pacemaker 架構元件

一個以 Pacemaker 配置的叢集，它包含了獨立元件的 daemon 以用來監控叢集成員、用來管理服務的 script，以及用來監控不同資源的資源管理子系統。以下元件形成了 Pacemaker 架構：

叢集資訊基礎（Cluster Information Base，CIB）

叢集資訊基礎（cluster information base，CIB）乃 Pacemaker 的資訊 daemon，它會使用內部 XML 來從 DC（Designated Co-ordinator，指定的協同者）分散和同步目前的配置與狀態資訊 — 這是個由 Pacemaker 所指定、透過 CIB 來將叢集狀態與動作儲存並分散至所有其它叢集節點上的

節點。

叢集資源管理 daemon（Cluster Resource Management Daemon，CRMd）

Pacemaker 的叢集資源動作會透過此 daemon 進行路由。CRMd 所管理的資源可以視需求，透過用戶端系統查詢、移動、列舉，和改變。

每個叢集節點亦包括了本地的資源管理 daemon（LRMd，local resource management daemon），作為 CRMd 與資源間的介面。LRMd 會將指令從 CRMd 傳遞到 agent 上，例如啟動與停止，和傳遞狀態資訊。

Shoot the Other Node in the Head（STONITH）

STONITH（Shoot the Other Node in the Head，直譯：將另一節點爆頭）通常會與電源交換器相結合，作為 Pacemaker 中的叢集資源，處理隔離的請求、強迫關掉節點、從叢集中移除節點以確保資料的完整性。STONITH 可在 CIB 中配置，也可以作為正常的叢集資源來監控。

1.5. Pacemaker 配置與管理工具

Pacemaker 的特性是擁有兩項配置工具，用來建置、監控，與管理叢集。

pcs

pcs 可以控制 Pacemaker 與 Corosync heartbeat daemon 的各方各面。**pcs** 乃指令列程式，用來進行以下管理工作：

- 建立、配置 Pacemaker/Corosync 叢集
- 在叢集執行時，修改叢集配置
- 遠端配置 Pacemaker 與 Corosync，並啟動、停止叢集；顯示叢集的資訊。

pcs-gui

圖形化介面，用來建立、配置 Pacemaker/Corosync 叢集，特性與能力與指令列工具 **pcs** 相同。

章 2. 叢集作業

本章提供了各項叢集功能與特性的相關摘要。從建立叢集仲裁至節點隔離，這些不同的特性構成了 High Availability Add-On 的核心功能。

2.1. 仲裁總覽

為了維持叢集完整性和可用性，叢集系統使用了一項名為 *quorum（仲裁）* 的概念，以避免資料損毀和遺失。當超過一半的叢集節點上線時，叢集便擁有仲裁。為了降低因為節點失效而造成資料損毀的機率，若叢集無仲裁的話，Pacemaker 就預設值會停下所有資源。

仲裁乃透過一項投票系統來建立的。當叢集節點未以正確的方式運作，或失去與整個叢集的通訊，多數的有效節點可進行投票並視需求將節點隔離以提供服務。

比方說，在一個包含了 6 個節點的叢集中，只有在至少 4 個叢集節點可運作的情況下，仲裁方可成立。若多數的節點離線或變得無法使用，叢集便無仲裁，而 Pacemaker 將會停下叢集服務。

Pacemaker 中的仲裁功能可避免 *split-brain（裂腦）* 情況發生，此情況代表當叢集失去通訊，但各部分卻依然作為獨立叢集繼續運作，這可能會使相同的資料被寫入並造成損毀或遺失。

High Availability Add-On 中的仲裁支援乃透過名為 **votequorum** 的 Corosync 外掛程式所提供的，它允許管理員為叢集中的各個系統配置指定給其的投票數，以確保只有在多數投票存在時，叢集作業才允許進行。

在無多數投票的情況下（比方說有個雙節點的叢集，當內部通訊網路失效時造成 50% 的叢集分裂），**votequorum** 能被配置為擁有一項 *tiebreaker* 政策，管理員可透過配置這項政策來藉由使用剩餘的叢集節點（此節點依然與擁有最低節點 ID 的叢集節點保持通訊）來繼續保有仲裁。

2.2. 隔離總覽

在一個叢集系統中能有多個節點負責處理重大的生產資料。在忙碌、多節點的叢集中，節點可能會運作異常或變得無法使用，並提示管理員進行動作。異常叢集節點所造成的問題可藉由建立一項 *fencing（隔離）* 政策來避免。

「隔離」代表將節點由叢集的共享儲存裝置中移除。隔離會切斷所有來自共享儲存裝置的 I/O，以確保資料的完整性。叢集基礎結構會透過 *STONITH* 來進行隔離。

當 Pacemaker 發現一組節點失效時，它會與其它叢集基礎結構的元件進行通訊，告知該節點已失效。當向 STONITH 通知了節點失效時，它會將失效的節點隔離。其它叢集基礎結構的元件會決定該進行哪些動作，亦即會進行任何必要的復原動作。舉例來說，DLM 與 GFS2 收到節點失效的通知時，會暫時停止活動，直到偵測到 STONITH 已完成隔離失效節點為止。確認失效節點已經隔離後，DLM 與 GFS2 就會開始進行復原。DLM 會解除對於失效節點的鎖定；GFS2 會復原失效節點的日誌檔。

透過 STONITH 進行的節點等級隔離可配置各種受支援的隔離裝置，包括：

- Uninterruptible Power Supply（UPS）— 一項含有電池的裝置，可被使用來在電源失效時隔離裝置
- Power Distribution Unit（PDU）— 一項包含了多重電源輸出的裝置，使用於資料中心以提供純淨的電源，以及阻斷服務與電源隔離服務。
- Blade power control devices — 安裝於資料中心裡的專門系統，配置來在失效事件發生時隔離叢集節點。
- Lights-out devices — 用來管理叢集節點可用性的網路連接裝置，並且可讓管理員本機或遠端執行隔離、開啟/關閉電源以及其它服務

章 3. Red Hat High Availability Add-On 資源

本章提供了下列資訊

3.1. Red Hat High Availability Add-On 資源總覽

「*叢集資源*」（cluster resource）是由叢集服務所管理的任何程式、資料或應用程式。這些資源會由提供標準介面的「*代理程式*」（agent）萃取，好在叢集環境中管理資源。這項標準根基於業界認證的架構與型別，這對叢集服務本身來說，可讓管理多種叢集資源的可用性更為通透。

3.2. Red Hat High Availability Add-On 資源類別

Red Hat High Availability Add-On 支援多種資源代理程式的類型：

- LSB — Linux 標準基礎（Linux Standards Base）代理程式會萃取 LSB 所支援的相容服務，亦即 **`/etc/init.d`** 中的服務與成功、失敗服務狀態（啟動、停止、執行中）的相關傳回碼。
- OCF — 開放叢集架構（Open Cluster Framework）是 LSB 上方的集合，設定了建立、執行伺服器啟始 script 的標準，使用環境變數為 script 輸入參數等等。
- Systemd — Systemd 是 Linux 系列最新的服務管理程式，使用了多組單元檔案，而不是 LSB 與 OCF 所使用的起始 script。這些單元檔案可以由系統管理員手動建立，或由服務自己來建立與管理。Pacemaker 管理單元檔案的方式，跟管理 OCF 或 LSB init script 的方式非常類似。
- Upstart — 與 systemd 非常類似，Upstart 是 Linux 的另一種系統起始管理程式。相對於 systemd 的單元檔案或 init script，Upstart 使用的是「工作」（job）。
- STONITH — 這是個專為隔離服務與使用 STONISH 的隔離代理程式所設計的資源代理程式。
- Nagios — 這是個為 Nagios 系統和基礎結構監控工具提取外掛程式的代理程式。

3.3. 監控資源

若要確保資源健全，您可加入一項監控作業至資源定義中。若您不為資源指定一項監控作業，就預設值，pcs 指令會建立一項監控作業，其間隔會由資源代理程式來判斷。若資源代理程式不提供預設的監控間隔，pcs 指令將會建立一項間隔為 60 秒的監控作業。

3.4. 資源限制式

您可藉由配置*限制式（constraint）*來判斷叢集中一項資源的行為。您可配置下列種類的限制式：

- location 限制式 — Location 限制式可決定資源能在哪個節點上執行。
- order 限制式 — order 限制式可決定資源的執行順序。
- colocation 限制式 — colocation 限制式可決定資源與其它資源相對之下應放置在什麼位置上。

Pacemaker 支援資源群組的概念，以簡化限制式的配置，它會將一組資源併在一起並確認資源會按照順序啟用，並以反向順序停下。

3.5. 資源群組

叢集最常見的要素之一就是一組需要位於相同位置、循序啟用，和相反順序停用的資源。為了簡化這項配置，Pacemaker 提供了*群組*概念上的支援。

您可透過 **pcs resource** 指令來建立資源群組，並指定欲包含在群組中的資源。若群組不存在的話，這項指令便會建立群組。若群組存在的話，這項指令便會將額外資源加入群組中。資源將會以您透過這項指令所指定的順序開始，並以其起始順序的相反順序停止。

附錄 A. 從 Red Hat Enterprise Linux High Availability Add-On 6 升級

本附錄提供了自Red Hat Enterprise Linux High Availability Add-On 第 6 版升級至第 7 版之概觀。

A.1. 不同版本間的差異一覽

Red Hat Enterprise Linux 7 High Availability Add-On 推介了新的、位於高可用性技術之下、根基於 Pacemaker 與 Corosync 的技術，完整取代了來自前一版 High Availability Add-On 的 CMAN 與 RGManager 技術。以下是兩個版本之間的部分差距。欲知版本間的完整差距，請參閱《*Red Hat Enterprise Linux High Availability Add-On Reference*・搭配 rgmanager 與 Pacemaker 建立叢集》。

- 配置檔案 — 先前，叢集配置可在 **/etc/cluster/cluster.conf** 檔案中找到；現在在第 7 版中，叢集配置之成員配置位於 **/etc/corosync/corosync.conf**，叢集節點與資源配置位於 **/var/lib/heartbeat/crm/cib.xml** 中。
- 執行檔 — 先前，叢集指令都可透過指令列的 **ccs** 與圖形介面的 **luci** 來完成。在 Red Hat Enterprise Linux 7 High Availability Add-On 中，配置可以透過指令列的 **pcs** 與位於桌面的網站圖形介面 **pcsd** 完成。
- 啟動服務 — 先前，High Availability Add-On 所包含的服務都是透過 **service** 指令來啟動，並透過 **chkconfig** 指令來配置開機時是否啟動服務。這必須針對所有叢集服務（**rgmanager**、**cman** 與 **ricci**）進行。例如：

  ```
  service rgmanager start
  chkconfig rgmanager on
  ```

  到了 Red Hat Enterprise Linux 7 High Availability Add-On，**systemctl** 會控制手動啟動、以及開機時自動啟動，同時所有叢集服務都集中在 **pcsd.service** 裡。例如：

  ```
  systemctl starat pcsd.service
  systemctl enable pcsd.service
  pcs cluster start -all
  ```

- 使用者存取 — 先前，root 使用者或擁有正確存取權限的使用者可以存取 **luci** 配置介面。所有存取行為都需要節點上的 **ricci** 之密碼。

  到了 Red Hat Enterprise Linux 7 High Availability Add-On，**pcsd** 網站圖形介面會要求使用者透過 **hacluster** 身份來進行認證，這是一般的系統使用者。**root** 使用者可以設定 **hacluster** 的密碼。

- 建立叢集、節點與資源 — 先前，建立節點可透過指令列的 **ccs** 或圖形介面的 **luci** 來完成。建立叢集、新增節點則是另外的工作。舉例來說，若要透過指令列建立叢集、新增節點，請執行以下指令：

  ```
  ccs -h node1.example.com --createcluster examplecluster
  ccs -h node1.example.com --addnode node2.example.com
  ```

  到了 Red Hat Enterprise Linux 7 High Availability Add-On，新增叢集、節點與資源皆會透過指令列的 **pcs**，或網站圖形介面 **pcsd** 來完成。舉例來說，若要透過指令列建立叢集，請執行以下指令：

  ```
  pcs cluster setup examplecluster node1 node2 ...
  ```

- 移除叢集 — 先前若要移除叢集，管理者必須從 **luci** 介面手動移除節點，或移除每個節點上的 **cluster.conf** 檔案。

到了 Red Hat Enterprise Linux 7 High Availability Add-On，管理者可以利用 **pcs cluster destroy** 指令來移除叢集。

附錄 B. 修訂記錄

修訂 2.1-5.2	**Wed Feb 3 2016**	**Terry Chuang**
翻譯、校閱完成		
修訂 2.1-5.1	**Wed Feb 3 2016**	**Terry Chuang**
讓翻譯檔案與 XML 來源 2.1-5 同步		
修訂 2.1-5	**Mon Nov 9 2015**	**Steven Levine**
準備文件，出版 7.2 GA		
修訂 2.1-4	**Mon Nov 9 2015**	**Steven Levine**
修正了 #1278841 新增了參照叢集相關培訓課程的註釋		
修訂 2.1-1	**Tue Aug 18 2015**	**Steven Levine**
準備文件，出版 7.2 Beta		
修訂 1.1-3	**Tue Feb 17 2015**	**Steven Levine**
7.1 GA 發行版本		
修訂 1.1-1	**Thu Dec 04 2014**	**Steven Levine**
7.1 Beta 發行版本		
修訂 0.1-9	**Tue Jun 03 2014**	**John Ha**
7.0 GA 發行版本		
修訂 0.1-8	**Tue May 13 2014**	**John Ha**
更新版本		
修訂 0.1-6	**Wed Mar 26 2014**	**John Ha**
最新的草稿		
修訂 0.1-4	**Wed Nov 27 2013**	**John Ha**
為 Red Hat Enterprise Linux 7 Beta 而建		
修訂 0.1-2	**Thu Jun 13 2013**	**John Ha**
Red Hat Enterprise Linux 7 的第一版本		
修訂 0.1-1	**Wed Jan 16 2013**	**Steven Levine**
Red Hat Enterprise Linux 7 的第一版本		

索引

符號

隔離, 隔離總覽

, 從 Red Hat Enterprise Linux High Availability Add-On 6 升級

H

High Availability Add-On