Red Hat Enterprise Linux 7

High Availability Add-On Overview

Overview of the components of the High Availability Add-On

Last Updated: 2021-05-26

Red Hat Enterprise Linux 7 High Availability Add-On Overview

Overview of the components of the High Availability Add-On

Steven Levine
Red Hat Customer Content Services
slevine@redhat.com

法律通告

Copyright © 2018 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at
http://creativecommons.org/licenses/by-sa/3.0/
. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux ® is the registered trademark of Linus Torvalds in the United States and other countries.

Java ® is a registered trademark of Oracle and/or its affiliates.

XFS ® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL ® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js ® is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack ® Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

摘要

Red Hat High Availability Add-On Overview provides an overview of the High Availability Add-On for Red Hat Enterprise Linux 7. To expand your expertise in deploying a Red Hat High Availability cluster, you may be interested in the Red Hat High Availability Clustering (RH436) training course.

目录

第 1 章 高可用性附加组件概述 3
1.1. 集群基础知识 3
1.2. HIGH AVAILABILITY ADD-ON 简介 3
1.3. PACEMAKER 概述 4
1.4. PACEMAKER 架构组件 4
1.5. PACEMAKER 配置和管理工具 5

第 2 章 CLUSTER OPERATION 6
2.1. 仲裁概述 6
2.2. 隔离概述 6

第 3 章 红帽高可用性附加组件资源 7
3.1. 红帽高可用性附加组件资源概述 7
3.2. RED HAT HIGH AVAILABILITY ADD-ON RESOURCE CLASSES 7
3.3. 监控资源 7
3.4. 资源约束 7
3.5. 资源组 7

附录 A. 从 RED HAT ENTERPRISE LINUX 6 HIGH AVAILABILITY ADD-ON 升级 9
A.1. 发行版本间差异概述 9

附录 B. 修订记录 11

索引 11

第 1 章 高可用性附加组件概述

高可用性附加组件是一个集群的系统，它为关键生产环境服务提供可靠性、可伸缩性以及高可用性。以下小节介绍了高可用性附加组件的组件和功能：

- 第 1.1 节 “集群基础知识”
- 第 1.2 节 “High Availability Add-On 简介”
- 第 1.4 节 “Pacemaker 架构组件”

1.1. 集群基础知识

集群由两个或者多个计算机（称为 *节点* (node)或 *成员*(member)）组成来一起执行任务。有四种主要类型的集群：

- 存储
- 高可用性
- 负载均衡
- High performance

存储集群在集群中的服务器间提供一致的文件系统镜像,允许服务器同时读和写入到一个共享文件系统。存储集群通过将应用程序安装和修补到一个文件系统来简化存储管理。另外,在集群范围的文件系统中,存储集群不再需要应用程序数据冗余副本,并简化备份和恢复过程。High Availability Add-On 与 Red Hat GFS2（Resilient Storage Add-On 的一部分）一起提供存储集群。

高可用性型机器通过消除单点失效以及在一个节点停止运作时将服务从一个群集节点切换到另外一个节点来提供高可用性服务。通常，高可用性群集中的服务会进行读写数据的操作（使用通过读写方式挂载的文件系统）。因此，高可用性集群必须维护数据的完整性，因为一个集群节点可能需要接管另外一个集群节点的任务。对于集群外的客户端，高可用性集群中出现的节点故障是不可见的。（高可用性机器有时候被称为故障转移（failover）集群。） High Availability Add-On 通过它的高可用性服务管理组件 **Pacemaker** 来提供高可用性集群。

负载均衡集群将网络服务请求发送到多个集群节点,以便在集群节点之间平衡请求负载。负载均衡提供了高性价比的可扩展性,因为您可以根据负载要求匹配节点数量。如果负载均衡集群中的节点无法工作,负载均衡软件会检测到故障,并将请求重定向到其他集群节点。集群外的客户端无法看到负载均衡集群中的节点故障。Load Balancer Add-On 提供了负载均衡。

高性能集群使用集群节点执行并发计算。通过使用高性能集群,应用程序可以并行工作,因此可以提高应用程序的性能。（高性能集群也称为计算集群或网格计算。）

注意

以上文本中总结的集群类型反映了基本配置,您的需要可能需要结合上述集群。

另外,Red Hat Enterprise Linux High Availability Add-On 包含了对 *只* 配置和管理高可用性服务器的支持。它 *不* 支持高性能集群。

1.2. HIGH AVAILABILITY ADD-ON 简介

High Availability Add-On 是一组集成的软件组件,可在各种配置中部署,以满足您对性能、高可用性、负载均衡、可扩展性、文件共享和功能的要求。

High Availability Add-On 主要由以下组件组成：

- 集群基础结构 - 为节点以集群方式运行提供基本功能：配置文件管理、成员资格管理、锁管理和保护。
- 高可用性服务管理 - 提供在节点无法操作时服务从一个集群节点切换到另一个节点的功能。
- 集群管理工具 - 用于设置、配置和管理高可用性附加组件的配置和管理工具。这些工具用于 Cluster Infrastructure 组件、高可用性和服务管理组件以及存储。

您还可以为高可用性附加组件添加以下组件：

- Red Hat GFS2 (Global File System 2) – Resilient Storage Add-On 的一部分，提供和高可用性附加组件一起使用的集群文件系统。GFS2 允许多个节点进行块级别的存储共享，就好像每个群集节点都连接至本地存储空间一样。GFS2 集群文件系统需要一个集群基础结构。
- Cluster Logical Volume Manager(CLVM)- Resilient Storage Add-On 的一部分,提供集群存储的卷管理。CLVM 支持还需要集群基础架构。
- Load Balancer Add-On – 在第 4 层（TCP）和第 7 层（HTTP, HTTPS）服务中提供高可用性负载平衡和故障转移的路由软件。Load Balancer Add-On 在冗余的虚拟路由器的集群中运行，使用负载算法将客户端请求发送到实际的服务器。Load Balancer Add-On 并不需要与 Pacemaker 一起使用。

1.3. PACEMAKER 概述

High Availability Add-On 集群基础结构为一组计算机（称为 *节点* 或 *成员*）作为一个集群一起工作提供基本功能。当集群使用集群基础架构后,您可以使用其他组件来满足集群需求（例如,设置集群以便在 GFS2 文件系统中共享文件或设置服务故障转移）。集群基础架构执行以下功能：

- 集群管理
- 锁定管理
- 隔离
- 集群配置管理

1.4. PACEMAKER 架构组件

使用 Pacemaker 配置的集群由独立的组件守护进程组成,这些守护进程监控集群成员资格、管理服务的脚本以及监控不同资源的管理子系统。以下组件组成 Pacemaker 架构：

Cluster Information Base（CIB）

Pacemaker 信息守护进程使用 XML 在内部向所有其他集群节点分发和同步当前配置和状态信息。这些信息来自于 DC（Designated Coordinator），它是由 Pacemaker 分配的用来存储和发布集群状态和动作的节点。

集群资源管理守护进程（CRMd）

Pacemaker 集群资源操作通过这个守护进程进行路由。由 CRMd 管理的资源可由客户端系统查询，并在需要时进行移动、实例化和更改。

每个节点还包括一个本地资源管理器守护进程（LRMd），它充当 CRMd 和资源间的接口。LRMd 将命令从 CRMd 传递给代理，如启动和停止状态信息。

STONITH（Shoot the Other Node in the Head）

通常与电源交换机一同部署,STONITH 在 Pacemaker 中作为集群资源使用,用于处理保护请求,强制关闭节点并从集群中移除它们以确保数据的完整性。STONITH 在 CIB 中配置,它可以作为普通的集群资源被监控。

corosync

corosync 是用于高可用性集群的核心成员资格和成员沟通需要的组件。它使用一个同名称的守护进程。它是 High Availability Add-On 正常运行所必需的功能。

除成员资格和消息功能外，**corosync** 还：

- 管理仲裁规则并进行裁定。
- 为在集群的多个成员间协调或操作的应用程序提供消息功能，因此必须在实例间进行有状态或其他信息通信。

1.5. PACEMAKER 配置和管理工具

Pacemaker 有两个配置工具用于集群部署、监控和管理。

pcs

pcs 可以控制 Pacemaker 和 Corosync heartbeat 守护进程的所有方面。基于命令行的程序 **pcs** 可以执行以下集群管理任务：

- 创建并配置 Pacemaker/Corosync 集群
- 在集群运行时修改集群配置
- 远程配置 Pacemaker 和 Corosync 以及启动、停止和显示集群状态信息

pcsd Web UI

用于创建和配置 Pacemaker/Corosync 集群的图形用户界面,其功能与基于命令行的 **pcs** 工具相同。

第 2 章 CLUSTER OPERATION

本章概述了各种集群功能和功能。从建立集群仲裁到隔离的节点隔离,这些不同功能就是高可用性附加组件的核心功能。

2.1. 仲裁概述

为了保持集群的完整性和可用性,集群系统使用一个称为 *仲裁* (quorum)的概念来防止数据崩溃和丢失。当超过一半的集群节点在线时，集群就已被"仲裁"。为减少由于失败造成数据崩溃的机会，在不满足仲裁数量的情况下，Pacemaker 会默认停止所有资源。

仲裁是通过一个投票（voting）系统来建立的。当一个集群节点工作不正常，或丢掉了与其他集群部分的通信，则大多数工作的节点可以通过投票来分离有问题的节点，如果需要，对节点进行隔离。

例如，在一个 6 个节点集群中，在至少有 4 个集群节点正常工作时就满足了仲裁。如果大多数节点离线或不可用,集群就不再具有仲裁的 Pacemaker 停止集群的服务。

Pacemaker 中的仲裁功能可以防止出现脑裂(*split-brain*)问题。当集群中出现无法相互通信的问题,但每个部分都可以作为一个独立的集群运行,则代表集群出现了脑裂的问题。这可能会导致出现数据被破坏的问题。

High Availability Add-On 中的仲裁支持由名为 **votequorum** 的 Corosync 插件提供,它允许管理员使用分配给集群中的每个系统的投票数来配置集群,并确保在大多数投票机制都存在时,集群操作才可以进行。

在没有大多数的情况下（例如,双节点集群,内部通信网络链接不可用,导致 50% 集群分割）, **votequorum** 可以配置为具有 *tiebreaker* 策略,管理员可以使用仍与具有最低节点 ID 的可用集群节点联系来继续仲裁。

2.2. 隔离概述

在集群系统中,可能有很多节点在几个重要生产数据上工作。忙碌的多节点集群中的节点可能会开始错误操作,或者变得不可用,并提示管理员动作。通过建立 *隔离* 策略可以缓解由错误集群节点造成的问题。

隔离(fencing)是节点与集群共享存储的断开连接。隔离会从共享存储中关闭 I/O,从而确保数据的完整性。集群基础架构通过 *STONITH* 工具执行隔离。

当 Pacemaker 确定节点失败时,它会与节点失败的其他集群基础架构组件通信。当出现故障时,STONITH 会隔离失败的节点。其他集群基础结构组件决定要采取什么操作,包括执行任何需要进行的恢复。例如,DLM 和 GFS2 当节点失败通知时,会挂起活动,直到它们检测到 STONITH 已经完成了对故障节点的隔离。确认有故障节点被隔离后,DLM 和 GFS2 执行恢复。DLM 释放失败节点的锁定 ; GFS2 恢复失败节点的日志。

通过 STONITH 进行节点级别的隔离可使用各种支持的隔离设备进行配置,其中包括：

- 不可破坏的电源电源(UPS)- 包含一个可以用来在电源失败时用来隔离设备的刻板设备
- 电源分布单元(PDU)- 在数据中心使用多个电源来提供干净的电源以及隔离和电源隔离服务的设备
- Blade power control 设备 - 在数据中心中安装的专用系统配置为在出现故障时隔离集群节点
- lights-out 设备 - 管理集群节点可用性的网络连接设备,并可由管理员在本地或远程执行隔离、电源/关闭和其他服务

第 3 章 红帽高可用性附加组件资源

本章概述了红帽高可用性资源及其操作。

3.1. 红帽高可用性附加组件资源概述

集群资源 是一个由集群服务管理的程序、数据或应用程序实例。这些资源通过 *代理(agent)*作为一个标准接口,用来在集群环境中管理资源。这个概念基于行业批准的框架和类,这些框架和类对集群服务本身而言透明管理各种集群资源的可用性。

3.2. RED HAT HIGH AVAILABILITY ADD-ON RESOURCE CLASSES

红帽高可用性附加组件支持多种资源代理：

- LSB - Linux 标准基础代理提取 LSB 支持的兼容服务,包括 **/etc/init.d** 中的服务以及关联的返回代码,以了解成功和失败的服务状态（启动、停止、运行状态）。
- OCF - Open Cluster Framework 是 LSB（Linux 标准基础）的超级集,用于创建和执行服务器初始化脚本、使用环境变量的脚本的输入参数等等。
- systemd - 基于 Linux 系统的最新的系统服务管理器, systemd 使用单元文件集合而不是初始化脚本,如 LSB 和 OCF 一样。这些单元可由管理员手动创建,甚至可由服务本身创建和管理。Pacemaker 以类似管理 OCF 或 LSB 初始化脚本的方式管理这些单元。
- Upstart - 与 systemd 类似, Upstart 是 Linux 的替代系统初始化管理器。Upstart 使用作业,而不是 systemd 或 init 脚本中的单元。
- STONITH - 专用于使用 STONITH 的隔离服务和隔离代理的资源代理。
- nagios - 为 Nagios 系统及基础架构监控工具提取插件的代理。

3.3. 监控资源

为确保资源健康,您可以在资源的定义中添加监控操作。如果您没有为资源指定监控操作,默认情况下, pcs 命令将创建一个监控操作,且间隔由资源代理决定。如果资源代理不提供默认的监控间隔， pcs 命令将创建监控操作，间隔为 60 秒。

3.4. 资源约束

您可以通过配置 *约束* 来决定集群中资源的行为。您可以配置以下约束类别：

- 位置约束 - 位置约束决定资源可在哪个节点上运行。
- 顺序约束 - 顺序约束决定资源运行的顺序。
- 共同位置约束 - 共同位置约束（colocation constraint）决定资源相对于其他资源的位置。

简而言之，配置一组限制会将一组资源放在一起，并确保资源按顺序启动并按相反顺序停止，Pacemaker 支持资源组的概念。

3.5. 资源组

集集的一个最常见的元素是一组资源，这些资源需要放置在一起，并按顺序启动并按反顺序停止。为简化此配置,Pacemaker 支持 *组* 的概念。

您可以使用 **pcs resource** 命令创建资源组,指定要包含在组中的资源。如果组不存在，这个命令会创建组。如果组存在，这个命令会向组群添加其他资源。这些资源将按您使用此命令指定的顺序启动，并以相反的顺序停止。

附录 A. 从 RED HAT ENTERPRISE LINUX 6 HIGH AVAILABILITY ADD-ON 升级

本附录概述了将 Red Hat Enterprise Linux High Availability Add-On 从版本 6 升级到版本 7。

A.1. 发行版本间差异概述

Red Hat Enterprise Linux 7 High Availability Add-On 引进了一组属于高可用性系统基础的新技术。这些技术基于 Pacemaker 和 Corosync,它们取代了之前的高可用性附加组件版本中的 CMAN 和 RNSanager 技术。以下是两个版本之间的一些区别。如需更全面地查看发行本之间的不同,请参阅 *Red Hat Enterprise Linux High Availability Add-On Reference* 中的名为"使用 rgmanager 和 Pacemaker 创建"的附录。

- 配置文件 - 之前,集群配置在 **/etc/cluster/cluster.conf** 文件中找到,版本 7 中的集群配置用于成员资格和仲裁配置, **/var/lib/pacemaker/cib/cib.xml** 用于集群节点和资源配置。**/etc/corosync/corosync.conf**

- 可执行文件 - 以前,集群命令位于 **ccs** 命令行,使用 **luci** 进行图形配置。在 Red Hat Enterprise Linux 7 High Availability Add-On 中,配置通过命令行 **pcs** 和桌面的 **pcsd** Web UI 配置完成。

- 启动服务 - 之前,包括高可用性附加组件的所有服务都使用 **service** 命令执行, **chkconfig** 命令配置服务在系统引导时启动。这必须为所有集群服务（**rgmanager**、**cman** 和 **ricci** ）单独配置。例如：

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

  对于 Red Hat Enterprise Linux 7 High Availability Add-On, **systemctl** 可以控制手动启动和自动引导时间启动,所有集群服务都分组到 **pcsd.service** 中。例如：

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

- User Access - 之前, root 用户或具有适当权限的用户可以访问 **luci** 配置界面。所有访问都需要节点的 **ricci** 密码。

  在 Red Hat Enterprise Linux 7 High Availability Add-On 中, **pcsd** Web UI 要求您验证为用户 **hacluster**,这是通用系统用户。**root** 用户可以为 **hacluster** 设置密码。

- 创建集群、节点和资源 - 以前,通过命令行或使用 **luci** 图形界面使用 **ccs** 执行节点创建。创建集群和添加节点是一个单独的流程。例如,要创建集群并使用命令行添加节点,请执行以下操作：

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

  在 Red Hat Enterprise Linux 7 High Availability Add-On 中,集群、节点和资源的添加是通过命令行的 **pcs** 或 **pcsd** Web UI 完成的。例如,要使用命令行创建集群,请执行以下操作：

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

- 集群删除 - 以前,管理员通过从 **luci** 接口手动删除节点或从每个节点中删除 **cluster.conf** 文件来删除集群

在 Red Hat Enterprise Linux 7 High Availability Add-On 中,管理员可以通过发出 **pcs cluster destroy** 命令删除集群。

附录 B. 修订记录

修订 7.1-1 为 7.7 GA 发布版本准备文件。	Wed Aug 7 2019	Steven Levine
修订 6.1-2 为 7.6 GA 发布准备文件。	Thu Oct 4 2018	Steven Levine
修订 5.1-2 为 7.5 GA 发布准备文件。	Wed Mar 14 2018	Steven Levine
修订 5.1-1 为 7.5 Beta 发布编写文件。	Wed Dec 13 2017	Steven Levine
修订 4.1-3 发布 7.4 GA 的文件版本。	Tue Aug 1 2017	Steven Levine
修订 4.1-1 为 7.4 发布版本准备文件。	Wed May 10 2017	Steven Levine
修订 3.1-2 7.3 GA 发布的版本。	Mon Oct 17 2016	Steven Levine
修订 3.1-1 为 7.3 Beta 发布准备文件。	Wed Aug 17 2016	Steven Levine
修订 2.1-5 为 7.2 GA 发布版本准备文件。	Mon Nov 9 2015	Steven Levine
修订 2.1-1 为 7.2 Beta 发布准备文件。	Tue Aug 18 2015	Steven Levine
修订 1.1-3 7.1 GA 版本	Tue Feb 17 2015	Steven Levine
修订 1.1-1 7.1 Beta 版本	Thu Dec 04 2014	Steven Levine
修订 0.1-9 7.0 GA 版本	Tue Jun 03 2014	John Ha
修订 0.1-4 Red Hat Enterprise Linux 7 Beta 构建	Wed Nov 27 2013	John Ha
修订 0.1-1 Red Hat Enterprise Linux 7 的第一个版本	Wed Jan 16 2013	Steven Levine

索引

符号

隔离， 隔离概述

，从 Red Hat Enterprise Linux 6 High Availability Add-On 升级

，Pacemaker 概述，Pacemaker 架构组件，Pacemaker 配置和管理工具，从 Red Hat Enterprise Linux 6 High Availability Add-On 升级

C

Cluster

Quorum（仲裁），仲裁概述

隔离，隔离概述

H

High Availability Add-On

版本 6 和 7 之间的区别，发行版本间差异概述

Q

Quorum（仲裁），仲裁概述