

Skybility HA Service V1.2 技术白皮书

软件简介

Skybility HA Service V1.2 是基于 Linux 的高可用性集群解决方案，能够保障企业的关键应用提供 24×7 的服务。通过 Skybility HA Service，您可以构建一个两节点的集群系统，于是，集群中的任一个节点就成为另一个节点的“热备份”，同时，每个节点又都能用作正常的 Linux 服务器。

在一个典型的配置中，Skybility HA Service 集群的硬件由两台连网的服务器。在系统运行过程中，服务器之间通过客户网络或独享网络（可选配）、RS232 连线（可选配）不断使用心跳机制相互监测，当发现其中一个节点由于某种原因发生故障时，无论软、硬件故障，另一个节点将接替故障节点的工作，从 IP 地址资源，存储资源到应用程序，从而保障了整个系统的高可用性。同时，Skybility HA Service 支持冗余的心跳路径和网络通讯路径，来保证故障检测和消息传递更安全可靠。

Skybility HA Service 支持 Linux 下所有常用服务的双机方案，从单纯的网络服务如 LVS，到 Web Server 和 Linux 下主流的数据库软件，Skybility HA Service 都能提供灵活的服务保护策略，包括对于故障服务的本地恢复和切换到备份节点。Skybility HA Service 支持 Active-Active 和 Active-Standby 等多种服务保护方式，能满足用户的多种要求。

同时 Skybility HA Service 支持大多数常用的网络环境和存储环境。他支持多网卡备份和负载均衡，使得服务器能持续，高性能的提供网络服务；另外，Skybility HA Service 支持大多数常用的存储设备，包括基于 SCSI 磁盘阵列柜，SAN, NAS，并内置了一些特殊存储设备（比如 IBM Exp300 系列磁盘阵列柜）的支持，同时支持多种日志文件系统和 RAW 文件系统，能满足所有高端用户和低端用户对存储的需求。

Skybility HA Service 提供了多个应用程序代理，如 Oracle、Sybase、IBM DB2 等，提供了对常用应用程序更好的支持。当然，您也可以通过 Skybility HA Service 的 Application Agent API 来定制自己的应用程序代理。Skybility HA Service 提供更丰富的配置模板，方便用户配置，避免因配置错误带来的问题。另外 Skybility HA Service 还提供开放的用户预警平台，用户可以自己定制或采用我们提供的解决方案来使用服务故障时的报警机制。

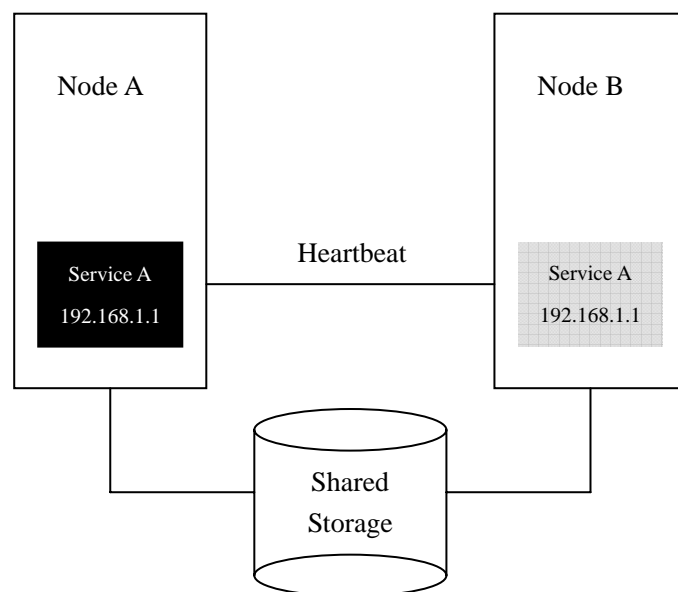
对于系统的安装和配置工作，Skybility HA Service 在设计时就充分考虑到了易用性。从本地应用程序到远程浏览器的 applet 的基于 java2 的图形化的配置界面，系统管理员可以根据喜好自由选择。同时，所有的系统监测和控制都可以远程进行，能够适应各种复杂环境。另外 Skybility HA Service 的可以在本地生成配置文件，然后拷贝到远程服务器上，就能实现服务器节点信息的配置，极大的方便了用户大规模的应用部署。

软件特点

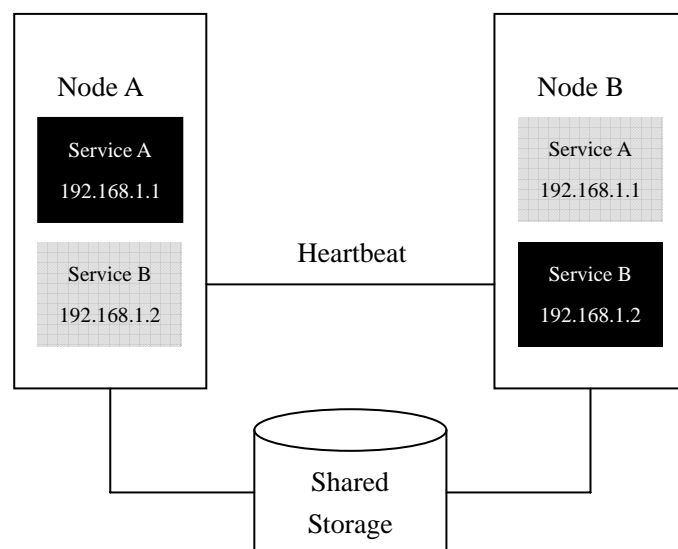
不间断的提供有效、准确的服务是高可用性集群软件的设计目标。在保证用户数据完整性的前提下，当系统或服务失效时，及时的将服务切换到正常节点，同时采取必要措施，帮助失败节点能够恢复正常，这就是 Skybility HA Service 作为优秀高可用性集群软件所提供的功能。

主机与服务

Skybility HA Service 设计为双节点集群系统，集群软件同时运行在两台主机上。对于主机上服务的配置，根据用户的需要，可以是一台主机提供服务，另一台主机待命的“主动—被动”模式，也可以是两台主机同时提供不同服务，并且互为备份的“主动—主动”模式。如果用户有两台同样高配置的服务器，并希望提供两种或两种以上的服务，则可以采用“主动—主动”模式以提高系统利用率；如果用户有一台高配置的服务器和一台较低配置的服务器，希望建立高可用性服务，则可以采用“主动—被动”模式，并把服务配置成“回切”型。在任意时刻，两台主机可以同时对外提供总和多达十几种的高可用性服务。当一个服务发生故障切换时，不会影响到其他服务的正常运行。



主动—被动模式



主动—主动模式

故障监测

Skybility HA Service 对于主机系统级的故障，两台主机间通讯的故障和所提供服务的故障都能进行准确的实时监测。

系统和通讯监测

任何操作系统，都有出现死机或系统挂起的可能。系统挂起和死机不同，系统挂起时对用户的输入不再有响应，好像被锁住一样，在有些情况下，系统挂起一段时间后，有可能重又继续工作。Skybility HA Service 可以准确的检测到一台主机系统挂起或死机的发生，并把服务切换到正常工作的主机上。

为了监测对等主机的状态，Skybility HA Service 集群在两台主机之间可以建立两种方式的任意多条连接通路，这也被称为“心跳”（Heartbeat）。“心跳”方式有 UDP/IP 连接和串行线连接两种。可以使用多块网卡，在两台主机间建立多条点对点的 UDP/IP 连接。如有多个串行口，则可以建立多条串行连接。

使用两种方式的多条连接，也就是利用冗余的硬件，提高主机间通讯的可靠性。只有当所有心跳通路全部失败时，才认为两主机在通讯上失效，此时 Skybility HA Service 会采取及时有效的应对措施。

服务检测

Skybility HA Service 对于服务的状态也会定时进行监测，监测的时间间隔可由用户指定。Skybility HA Service 提供一个“通用应用程序代理”，可以对各种服务进行一般性的监测。对于常用类型的服务，还有相应的“应用程序代理”可以实现具有针对性的服务监测功能。用户也可以自行编写应用程序代理，以满足特殊的需要。

Skybility HA Service 还对服务所使用的资源（包括存储资源和网络资源）进行准确的检测，通过底层的网卡数据流量的检测，Skybility HA Service 可以准确地获得网卡的连接状态，为服务的切换提供准确的参考。

故障的处理和服务的切换

通过的准确的故障监测手段，Skybility HA Service 能够准确智能的处理各种故障，以保护服务的不间断运行。

节点故障的处理

当 Skybility HA Service 中的节点 A 通过心跳发现节点 B 发生了故障，A 会通过网络或者电子开关的 STONITH(Shoot the other node in the head)的机制将 B 重起或者脱离集群。同时节点 B 也通过硬件级或系统级的 watchdog 机制，在系统崩溃时快速重起，从而能更好的释放服务占用的资源，以便节点 A 接管服务，并保障数据的完整性。

服务故障的处理

Skybility HA Service 通过应用程序代理和内置的服务资源监控策略，能准确地检测服务的故障。同时，用户可以自定义服务检测的间隔，避免频繁的服务检测带来的服务性能的损失。而且服务检测有一定的冗余度，避免某次突发的检测故障带来服务的误切换。

当节点 A 检测到应用程序自身的故障时，A 会先检测节点 B 的状态，如果 B 上的 HA 没有启动或者 B 上服务的启动条件不具备（比如提供网络服务的网络接口断开）时，A 将在本地恢复服务；如果 B 的节点状态正常且服务启动条件具备的时候 A 将服务切换到节点 B 上。

当节点 A 检测到应用程序所使用网络出现故障时，A 会先检测节点 B 的状态，如果 B 上的 HA 没有启动或者 B 上服务的启动条件不具备（比如提供网络服务的网络接口断开）时，A 对服务不做处理，并通过 emerge 的日志在控制台报警；如果 B 的节点状态正常且服务启动条件具备的时候 A 将服务切换到节点 B 上。

故障处理的日志和报警机制

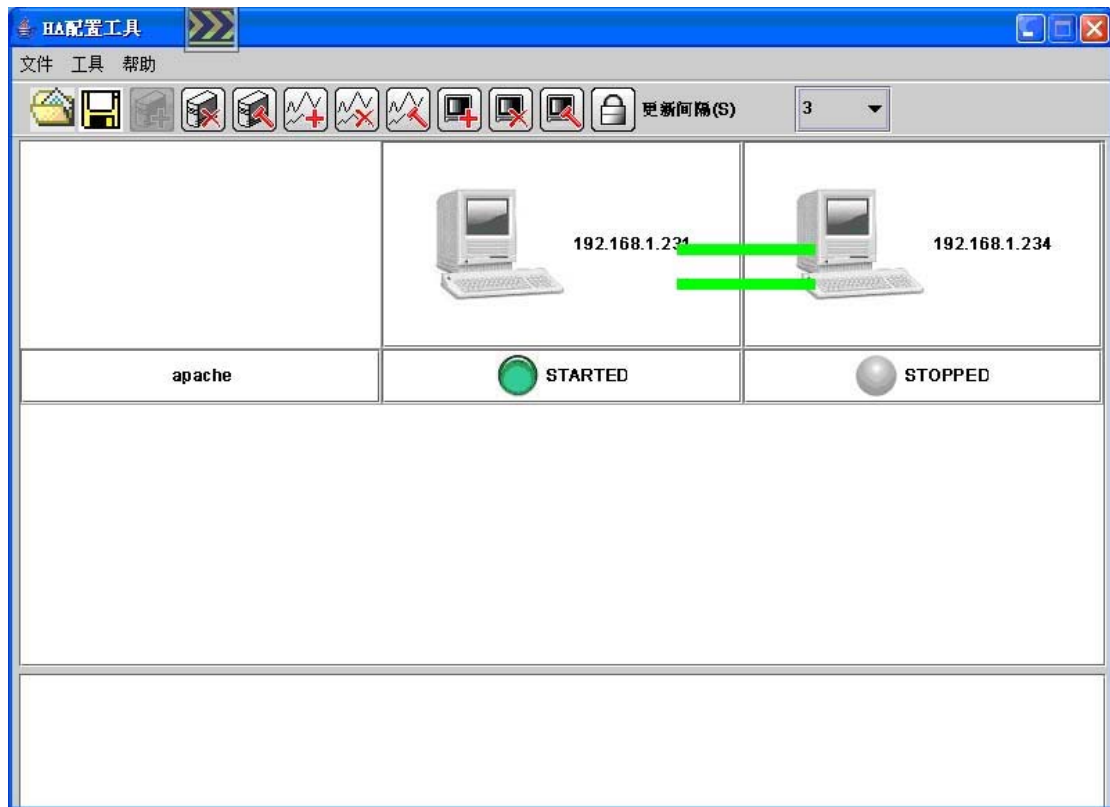
Skybility HA Service 通过系统的 syslog 接口来提供日志，您可以在/var/log/ha 中查看 HA 在运行中产生的日志。而且可以通过管理控制台修改记录日志的级别，使得用户能查看更进一步的日志信息，方便问题的发现和系统的诊断。同时，利用系统的 logrotate 工具，可以将日志自动的备份和清除，避免造成系统日志过多导致磁盘资源耗尽。

另外，Skybility HA Service 还预留用户报警的接口，用户可以自己添加在当有 emergency 日志时的报警事件，比如通过 mail，短信或者警报等手段来通知管理员。

配置和状态监控

Skybility HA Service 通过跨平台 java2 的图形界面来配置和监控节点和服务的状态。配置的客户端可以运行在 Linux 和 Windows 等不同的平台上，并支持浏览器访问，方便用户配置。而且配置界面支持 i18n 标准，在不同的字符集下显示不同的语言。配置界面简洁，易于操作，而且所有的服务控制都能在界面上完成。

通常的配置界面如下图。



软件技术指标

首先我们定义 HA 产品的技术指标如下：

- $Tr = Tc - Tf - (Tstart + Tstop)$ ，其中 Tr 为服务切换时间， Tc 为服务切换完成时刻， Tf 为服务发生故障时刻， $Tstart$ 为启动服务所用的时间， $Tstop$ 为停止服务所用的时间。
- 服务的可用性 = $MTTF / (MTTF + MTTR) * 100\%$ ，其中 $MTTF$ 为系统平均无故障时间，即计算机系统平均能够正常运行多长时间，才发生一次故障，系统的可靠性越高，平均无故障时间越长； $MTTR$ 为可维护性用平均维修时间，即系统发生故障后维修和重新恢复正常运行平均花费的时间，系统的可维护性越好，平均维修时间越短。

在配置良好的情况下，Skybility HA Service 的服务切换时间在 5s 到 10s 之间；而可用性可以达到 99.99%。