

# TurboHA6.5 技术白皮书

——企业级 Linux 高可用集群解决方案



北京拓林思软件有限公司  
2005 年 6 月



## 目录

概述 .....	1
TURBOHA6.5 简介 .....	1
应用支持 .....	2
TURBOHA6.5 技术特点 .....	2
支持磁盘镜像功能 .....	2
提供多种类型的磁盘阵列支持 .....	3
多种硬件心跳保证系统一致性 .....	3
支持 STONITH 技术 .....	3
支持 Watchdog Timers .....	3
智能的服务回迁以及多服务的负载分担 .....	4
可以检测更多的故障 .....	4
应用程序代理检查 .....	4
应用程序代理 API .....	5
图形管理工具 .....	5
更好的日志文件系统支持 .....	5
更详细的系统故障日志信息 .....	5
TURBOHA6.5 设计原理和使用 .....	6
设计原理 .....	6
主机及服务 .....	6
监测 .....	7
切换 .....	8
安装及使用 .....	9
系统安装和配置 .....	9
系统监测和管理 .....	9
远程维护 .....	10
客户端 .....	10

## 概述

TurboHA6.5 是北京拓林思软件有限公司推出的为满足 Linux 平台电信级和企业级应用的高可用产品。TurboHA6.5 提供的双机高可用方案能够更好的满足用户业务的连续性，并且能够满足不同应用对高可用的要求。

自 2001 年以来，北京拓林思软件有限公司先后推出了两个系列的 HA 产品：TurboHA6.0 和 TurboHA6.5。TurboHA6.5 是 TurboHA6.0 的升级版，于 2004 年 7 月发布。TurboHA6.5 提供更好的可靠性，更高的性价比，更好的易用性和可管理性，完全满足企业级应用所要求的 RASM (Reliability, Availability, Scalability, Manageability) 特性，同时 TurboHA6.5 支持磁盘镜像功能，使得用户应用和 TurboHA6.5 软件可以不需要磁盘阵列来提供高可用服务。

TurboHA6.5 可支持 Turbolinux 发布的所有操作系统平台（TDS7.0, TELS8.0, TDS7.1, TDS10/GTES10），适用于 i386、x86\_64、openpower 等主流的硬件平台。由于 TurboHA6.5 可以和 Turbolinux 操作系统更好的配合，使得从操作系统到 TurboHA6.5 软件构建的高可用解决方案更加的可靠，并广泛服务于电信、银行、政府等行业客户。

## TurboHA6.5 简介

TurboHA6.5 是专注于 Linux 上的高可用性产品，提供高可用性的双机集群系统。当集群中的某个节点由于软件或硬件原因发生故障时，集群会利用资源切换的方法保证整个系统继续对外提供服务，从而为企业 24x7 的关键业务应用提供了强大的保障。TurboHA6.5 提供对各种应用程序的支持，包括各种数据库应用、WEB 应用、Mail 应用等等，而其简便的安装和设置、详细的日志信息，减轻了用户日常的维护工作，其中跨平台的远程管理和监控使得系统具有更灵活的特性。此外，TurboHA6.5 同时提供图形化界面和命令行界面两种配置管理工具，使得系统管理员的操作和管理更加方便。

## 应用支持

当我们通过硬件（服务器、交换机、电子开关等）和软件（操作系统平台、HA系统软件、应用软件等）搭建一个高可用群集环境的时候，首先我们需要明确的是，高可用系统软件能否支持和管理我们的应用程序。TurboHA6.5 能够支持绝大多数的 Linux 环境下的应用程序，支持的典型应用程序类型如下：

- 通用的，无需修改的应用程序：TurboHA6.5支持大多数Linux平台的应用程序，这些应用大多数是能够接受几秒钟的停机时间的业务。
- 数据库应用：TurboHA6.5能够很好的支持各种数据库产品，包括Oracle 8i/9i/10g, MySQL, Sybase和IBM DB2数据库。
- 各种文件服务：TurboHA6.5能够为各种类型的文件服务提供高可用集群功能，如NFS和SMB/CIFS（使用Samba）。
- 主流的商业应用软件：TurboHA6.5能够很好的支持主流的商业应用软件，如SAP, Oracle Application Server和Tuxedo。
- 互联网和开放源代码的应用：TurboHA6.5可以很好的支持各种流行的互联网应用软件和各种开放源代码产品，如Apache, Wu-ftp, VSFTP等。
- 邮件服务软件：如Sendmail和Domino。

## TurboHA6.5 技术特点

### 支持磁盘镜像功能

磁盘镜像功能，是一种不需要磁盘阵列的双机数据共享方案。它的基本原理是通过对两个节点各自的本地磁盘分区进行实时镜像操作，使得这两个本地磁盘对双方节点而言，可以当作一个虚拟的共享磁盘设备来使用，这个虚拟的 RAID-1 级别的共享磁盘设备，能够作为应用的共享设备，既可以当作共享的裸设备来使用，也可以在其上创建各种 Linux 文件系统。TurboHA6.5 本身提供磁盘镜像功能，使得共享数据的应用不需要磁盘阵列也能够搭建双机高可用方案。

## 提供多种类型的磁盘阵列支持

对于需要磁盘阵列的一些应用，如数据库应用等，需要专业的硬件磁盘阵列来保证性能。而 TurboHA6.5 能够支持绝大多数的磁盘阵列设备。

目前业界使用的磁盘阵列时，一般分两种情况：第一种是带独立 RAID 处理器的磁盘阵列，主流厂商的 SCSI 或光纤磁盘阵列都可以适用于 TurboHA6.5 的要求。第二种是使用主机 RAID 卡和磁盘柜（磁盘柜是指不具备硬件 RAID 处理器的磁盘盒）的方式，这种磁盘设备通过双机的硬件 RAID 卡 clustering 技术和 RAID 的用户接口工具等，也可以满足 TurboHA6.5 的共享数据存取的需求，常见的这种类型的“磁盘阵列”的典型产品有 IBM EXP400，DELL 220S，HP MSA500-G2 等。

## 多种硬件心跳保证系统一致性

TurboHA6.5 同时支持直连网线和串口的方式来同步 HA 两个节点之间的心跳信息。可同时支持多条直连网线和串口线作为 Channel，提供更高可靠性的硬件冗余方式，以保证两个节点之间不会发生 Split-brain 现象，即使两节点之间的 Channel 都发生故障，TurboHA6.5 还可以通过第三方参考 IP 的方式，保证两个节点系统的一致性。

## 支持 STONITH 技术

Stonith(shut the other node in the head)，就是把故障节点重启，以保证资源被完全释放。Stonith 的方式有两种，一种是通过电子开关(power switch)来重启对方；（另外一种是通过网络发送命令来重启对方。但是后者通常不起作用）。有一些厂家的服务器有类似的管理界面（一种硬件设备），也可以用来作为 stonith 的工具，比如 IBM xServer 的 RSA(Remote Supervisor Adapter)和 Intel 的 IPMI，TurboHA6.5 也可以支持。

## 支持 Watchdog Timers

TurboHA 支持三种类型的看门狗定时器为系统提供了一个稳健的 I/O barrier。最简单的就是 Linux 内核自带的通过中断处理来实现的软件 softdog 定时器，它被 TurboHA 用来控制后台程序的执行。

Linux 内核也支持一种硬件 NMI (non-maskable interrupt) 看门狗，这种

硬件看门狗通常需要专门服务器硬件支持（常用的是主板上的 Intel 810 TCO 芯片组）。NMI 看门狗在没有检测到一个稳定正常的中断发生时就会触发节点服务器的重启动。

最后一种，就是传统的硬件看门狗定时器，它是一种 PCI 设备，在市场上很常见。当 PCI 设备的驱动没有正常的复位时，它就会强迫系统关闭或重启。

### 智能的服务回迁以及多服务的负载分担

TurboHA6.5 支持优先节点的设置，可以把一些服务设定到指定的优先节点，当优先节点故障时，服务切换到另一个节点，而当优先节点又恢复时，服务会自动迁移到优先节点。这样可以使多个服务分别运行在两个节点上，使得服务的负载可以分担到两个节点上。

### 可以检测更多的故障

TurboHA 6.5 能够检测更多的系统故障，从而增强了高可用性集群所提供的可靠性。

- 系统故障 —— 硬件错误
- 系统紊乱 —— 系统软件错误
- 存储不可访问 —— 存贮错误
- 网络断开 —— 网络错误
- 集群进程故障 —— 集群软件错误
- 服务故障 —— 服务应用程序错误。

### 应用程序代理检查

TurboHA 6.5 通过使用应用程序代理检查某一服务是否运行。应用程序代理用于定期检查某一服务是正常工作。如果服务没有正常运行，则相应地触发一次切换，使服务在另一节点被恢复。TurboHA 6.5 提供用于常用服务的应用程序代理，对于自身没有应用程序代理程序的服务则可以使用 TurboHA 6.5 提供的接口进行灵活的按需定制。请同时参见本文中的“应用程序代理 API”一节。

---

## 应用程序代理 API

应用程序代理 API 是一种在应用程序代理或服务检查程序和 TurboHA 服务进程之间的接口，在 TurboHA 6.5 用户手册中有详细介绍。按照此接口的规范，您可以为您的特定服务编写定制的应用程序代理。编写定制的应用程序代理的好处在于它可以为应用程序提供更精确的服务检查以及更快的切换。

## 图形管理工具

TurboHA 6.5 通过提供基于 Java 技术的图形管理工具而改善了集群的可管理性。支持本地和远程的监控和管理，支持 Linux/Windows 客户端的管理。

利用所提供的图形管理工具，可以方便地进行配置更改和状态监测。除了提供图形管理工具外，TurboHA 6.5 还提供有功能同样强大的命令行配置和监控管理工具。

## 更好的日志文件系统支持

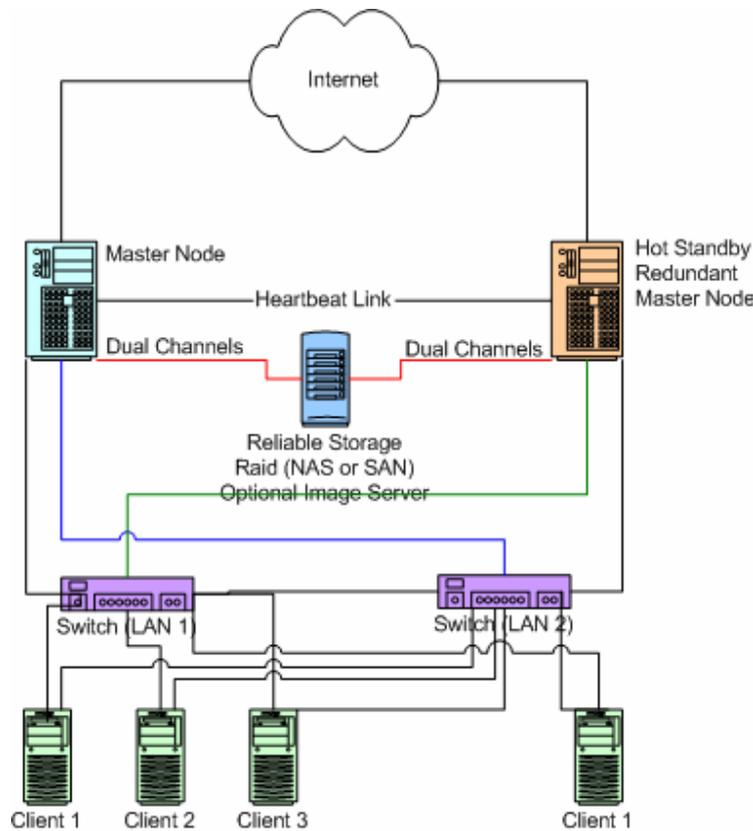
TurboHA 6.5 支持与日志文件系统诸如 Reiser 和 Ext3 等的协同工作。这些日志文件系统特别适用于 TurboHA 6.5，因为它们消除了诸如 Ext2 等文件系统中所化费的耗时的文件系统检查从而减少了切换时间。当系统装载时，日志文件系统仅仅要求恢复其日志。当一种日志文件系统被用于共享存贮时，TurboHA 6.5 能够自动地进行确认，跳过不需要的 FSCK 文件系统检查，并立即装载文件系统用于文件系统日志的恢复。

## 更详细的系统故障日志信息

TurboHA 6.5 采用的日志函数和 Linux 的 syslogd 是一样的方式，在两个节点均有记录，每个守护进程都有自己的日志级别，可以在配置文件中指定。每一条记录的信息，包括有时间、日志级别、进程名称、进程 id、消息等内容，这样可以方便用户进行应用故障现场的保护以及故障后的分析定位。同时日志的级别可以动态进行设置调整，以根据实际需要调整输出日志的信息内容。默认情况下，系统已经将日志级别设置成较为详细的信息输出，包括 HA 启动、停止过程，HA 事件触发原因，服务故障原因，服务切换过程，服务手动操作记录等。同时 TurboHA6.5 还提供日志收集工具，自动收集系统以及 HA 相关信息，以便于更方

便的进行故障定位。

## TurboHA6.5 设计原理和使用



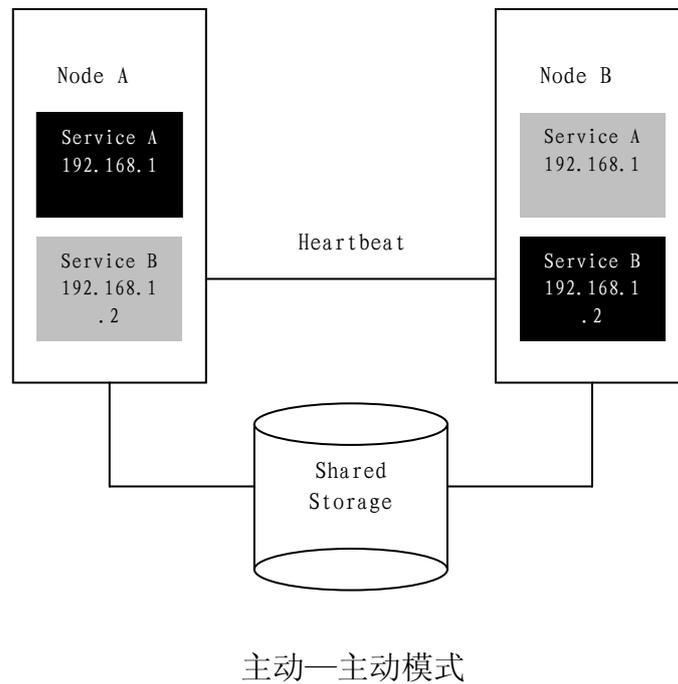
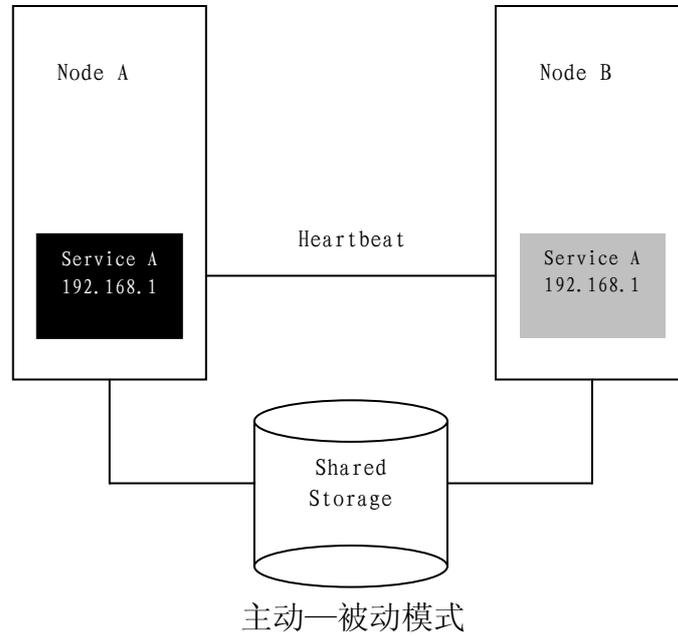
### 设计原理

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 TurboHA 6.5 作为优秀高可用性集群软件所提供的功能。

### 主机及服务

TurboHA 6.5 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。



## 监测

TurboHA 6.5 对于主机系统级的失败，两台主机间通讯的失败和所提供服务的失败都能进行准确的实时监测。

### ➤ 系统及通讯监测

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。TurboHA 6.5 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， TurboHA 6.5 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 TurboHA 6.5 会采取及时有效的应对措施。

### ➤ 服务监测

TurboHA 6.5 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。TurboHA 6.5 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。

## 切换

TurboHA 6.5 检测到一台主机上系统或服务的失败时，正常主机首先会建立 I/O 屏障，保护共用存储设备上的数据不被失败节点修改。然后，会把故障主机上的服务切换到正常主机上，继续对外提供服务。可以把一个或多个 IP 地址绑定在服务上，在服务切换时，IP 地址也随之切换到正常主机上，所以用户仅仅在服务切换的瞬间能感觉到极短时间的服务暂停。

在服务切换的同时，如果故障主机上集群软件仍在运行，则在检测到 I/O 屏障后，会把本机重新启动，如果重新启动后系统恢复正常，则重新加入集群，可以接管服务。这样，即使两台主机都发生故障，只要不在同一时刻发生，集群仍可保证提供服务。

## 安装及使用

### 系统安装和配置

TurboHA 6.5 需要安装在 Linux 系统之上，对于 Linux 的要求是 Kernel 2.2.18 以上，我们强烈推荐您使用 Turbolinux Server/DataServer 简体中文版 7.0 或更高版本，因为 TurboHA 6.5 已经在其上进行过严格的认证工作。整个安装和配置过程将十分轻松和简单。

通常来讲，整个系统的安装过程分为下面几个主要步骤：

- 安装及连接系统硬件，如连接共享存贮柜，连接独享网络（可选），连接 RS232 通讯连线（可选）等。
- 在两个节点上分别安装 Linux 操作系统，如 Turbolinux Server/DataServer。
- 在共享存贮柜上配置和初始化硬盘分区，并建立文件系统，推荐采用日志文件系统，Turbolinux Server/DataServer 推荐使用 EXT3 日志文件系统。
- 安装 TurboHA 6.5 软件。
- 使用命令行或图形化管理工具配置集群服务。

在 TurboHA 6.5 产品包装盒内附有快速安装指南为您快速导航，而每一步安装步骤都在用户手册中都有详细说明。

### 系统监测和管理

易管理性是衡量集群系统性能的一个重要指标。TurboHA 6.5 在保持 Linux 便捷的命令行控制方式的同时，提供了全功能的图形化控制界面来简化管理。系统管理员可以根据喜好选择下面两种方式的一种，甚至混合使用。

- Cluadmin 是命令行方式的工具，可以完成所有的集群配置和信息监控工作，对于习惯于传统 UNIX/Linux 管理的管理人员是最佳选择。
- Java 图形化的管理工具可以跨平台的在不同操作系统下运行。当图形化界面启动后，主窗口被分为下面三个部分：
  1. 配置：TurboHA 6.5 的所有配置工作。
  2. 状态：监测系统目前的工作状态。
  3. 服务管理：启动或停止特定的服务。

---

## 远程维护

远程维护是现代用来降低系统整体拥有成本的一个重要手段，系统管理员在异地维护系统可以节省大量的时间和资源。TurboHA 6.5 为您提供三种途径进行远程维护。

- 远程对主机登录后使用 Cluadmin 命令进程管理，所有操作与本地一样。
- 直接通过 Java 客户端管理工具，此种方式更适用于跨平台管理和远程监测。

## 客户端

正如所有其他高可用性系统一样，TurboHA 6.5 对于客户端是透明的，即客户端并不了解也无需了解集群系统。当集群内部发生服务切换操作时，对于面向非连接的服务，如 WEB 服务，客户端完全感觉不到切换动作的发生；而对于面向连接的服务，如数据库服务，客户端会接到警报，在以秒为时间单位的切换动作完成后，如果客户端有自动重连功能，将自动重新连接到服务器，如果没有，用户也只需重新登录即可继续正常工作。