

易腾纯软群集

技术白皮书

目录

| | |
|----------------------------------|-----------|
| 第一章 信息高可用性概述..... | 4 |
| 一、 高可用性需求..... | 4 |
| 二、 服务宕机的必然性..... | 4 |
| 三、 高可用性解决方案..... | 5 |
| 四、 纯软群集的简介..... | 5 |
| (一)、 高可用硬件配置..... | 5 |
| (二)、 数据镜像特性..... | 6 |
| (三)、 群集特性..... | 6 |
| (四)、 镜像群集的优点..... | 7 |
| 第二章 易腾纯软群集概述..... | 8 |
| 一、 软件组成..... | 8 |
| (一)、 易腾纯软群集操作界面..... | 8 |
| (二)、 易腾纯软群集操..... | 8 |
| (三)、 EterneTarget | 8 |
| (四)、 EterneDisk..... | 9 |
| 二、 功能特点..... | 9 |
| (一)、 分为两大产品..... | 9 |
| (二)、 应用高可用性..... | 9 |
| (三)、 自我修复能力..... | 10 |
| (四)、 支持多条心跳路径..... | 10 |
| (五)、 资源组管理..... | 10 |
| (六)、 单机热备高可用..... | 10 |
| (七)、 实时的块镜像..... | 11 |
| (八)、 磁盘/分区镜像..... | 错误！未定义书签。 |
| (九)、 真正的数据同步镜像..... | 错误！未定义书签。 |
| (十)、 快速重新同步..... | 13 |
| (十一)、 稳定的数据镜像效率..... | 14 |
| (十二)、 纯软双机平滑过渡到纯软群集..... | 14 |
| (十三)、 支持多种群集工作模式..... | 14 |
| (十四)、 采用C/S架构，管理员可以远程管理双机资源..... | 14 |
| (十五)、 支持自动迁回..... | 14 |
| (十六)、 支持无为ARP..... | 15 |
| (十七)、 直观的GUI用户界面..... | 15 |
| (十八)、 支持多种应用保护..... | 15 |
| (十九)、 全面支持用户自己开发的应用程序和服务..... | 15 |
| 第三章 易腾纯软群集软件应用规格..... | 16 |
| 一、 软硬件规格..... | 16 |
| (一)、 支持各种类型服务器..... | 16 |

- (二)、 支持的操作系统平台 16
- (三)、 提供丰富的应用软件支持 16
- 二、 群集规格配置..... 16
 - (一)、 EterneMirrorHA双机热备 16
 - (二)、 EterneMirrorCluster 19

第一章 信息高可用性概述

一、高可用性需求

随着信息化建设的不断推进，企业的信息化已经非常普遍，各个企事业单位的活动越来越多的依赖于其关键的业务信息系统。而对于日益承担企事业单位核心业务来说，数据的高可用性和系统的连续运转能力极其重要，服务器是企事业单位存放数据的重要设备，如果一旦因为服务器的故障而无法正常运转，就会造成业务停顿，导致不可挽回的损失。

服务器的故障可能由各种原因引起，如设备故障、操作系统故障、软件系统故障等等。一般地讲，在技术人员在现场的情况下，恢复服务器正常可能需要 10 分钟、几小时甚至几天。从实际经验上看，除非是简单地重启服务器（可能隐患仍然存在），否则往往需要几个小时以上。而如果技术人员不在现场，则恢复服务的时间就更长了。而对于一些重要系统而言，用户是很难忍受这样长时间的服务中断的。

随着企事业不断完善和投入运营，如何管好用企事业，充分发挥其现代化基础设施的作用，成为最迫切的问题。数据畅通，访问及时是各企事业运营管理的重要任务之一，工作正常运行直接影响到管理运作质量和管理的经济效益。可见，对那些需要保障信息安全和提供不间断的信息服务的企业而言，业务系统的容错性和不间断性显得尤为重要。如何保障各种关键应用持续运营，达到永续经营的良性循环，已成为当今企事业单位和 IT 领域急需解决的关键问题。

二、服务宕机的必然性

通常企业构造服务的硬件环境为单一服务器配置模式，一旦因为计划中宕机和计划外宕机，便会引起服务终止。

计划中的宕机通常指完成系统升级、应用迁移、部件更换等操作引起的宕机，这些计划中的宕机是企业维护服务正常工作不可避免的工作项目，虽然可以人为地采取一些措施减少损失，但依然会造成服务的终止。

计划外宕机通常是由于系统出错造成的。错误包括硬件、软件、系统和网络，或是系统运行外部环境原因等，这些情况包含了：硬件故障、操作系统崩溃死机、硬盘满、电压不稳、跳电、电源损坏、网络故障、软件漏洞（Bug）、应用出错。同样人为失误造成的故障也称为出错。当然，并非所有的出错会造成宕机，也不是所有的意外宕机都是由于部件出错造成的，灾难或其他意外情况同样会造成服务终止。计划外宕机通常会导致非常严重的后果，比如酒店的入店和结帐系统故障，所有的客户将无法 check-in 和 check-out；医院的电脑系统出错，将导致患者无法结账、医生无法得到患者信息，甚至无法进行手术；银行的记帐系统因电脑出错，客户将无法提现和存款。总之，计划外宕机所造成的损害将非常大。有些环境下，系统停止服务将导致事务处理无法进行，必将导致客户对企业信任度降低，甚至会给客户带来不可挽回的损失。

三、高可用性解决方案

通常，对于服务因错误造成的不可避免终止，企业都会安排相关的管理人员进行定时的监控，一旦发现服务不可用，便会立即进行处理排错。这是非常原始的高可用解决方案，存在诸多问题，譬如及时性问题，一旦错误在深夜发生，而人工不可能 24 小时监控；再譬如处理时间长短不可控，一旦错误造成的原因非常复杂，排错非常困难，需要非常长的时间，那么服务的终止将非常长。在排错过程中，人为操作失误将可能导致系统的永久瘫痪。另一方面，为了维持系统的高可用性而配备操作人员，劳务费用也非常昂贵。

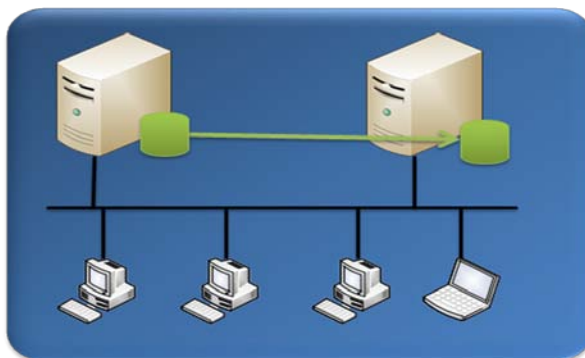
目前，市场上提供多种高可用性解决的软件解决方案。按照服务器分：主要分为双机和多机的解决方案。其主要原理是在提供硬件冗余的系统基础之上，运行高可靠性软件而构成。高可用性软件自动检测服务的运行状态，一旦服务在主机上出现故障，高可用性软件将自动地把服务转到另一台服务器上，从而让服务持续可用。对系统应用状态的监控和服务故障的恢复，由高可用软件自动完成，减少了人员的负担。按照共享磁盘分类：主要为镜像群集和共享磁盘群集。其主要原理是数据只存一份在磁盘阵列中，而磁盘阵列成群集内的服务共享，如果不是，那么数据在群集中服务间进行镜像，并且保证服务器间的数据完全一致性，为热备方案提供数据可靠性。

四、纯软群集的简介

（一）、高可用硬件配置

高可用解决方案的硬件通常有 2 种硬件配置方案，针对不同的硬件配置提供的高可用软件的也不相同。一种需要共享磁盘的高可用解决方案，即服务器中的数据都放到共享磁盘中，所有的服务其都能访问共享存储的数据，但服务运行时，为了数据的安全性，该数据只能被启动业务的服务器所访问，其它服务器是不能访问的；一种是不需要共享磁盘的高可用解决方案，服务器的数据进行实时镜像，保证服务器数据的一致性，因而不需要共享磁盘阵列，从而减少硬件的投入成本，纯软群集属于第二种群集模式。

纯软群集硬件配置



群集软件的解决方案就是避免单点故障出现，因而所有硬件都需要冗余的，包括两台服务器（我们称这种环境为双机容错），而部署多机高可用集群环境时，服务器则是多台。每

台服务器拥有足够的 I/O 能力、内存容量、系统硬盘空间、网卡部件，使得配置能够在使用中符合系统的需求，并能够最大限度的降低系统停机时间，同时为满足镜像需要，镜像的网线需要千兆网卡以上。

单独的镜像软件，无需较贵的 RAID 磁盘子系统，无需考虑小型计算机系统接口 (SCSI) 对缆线长度的限制，去除了单点故障 (磁盘阵列子系统)。集群计算机环境所要考虑的主要问题是系统总开销，而使用镜像软件的热备容错软件是解决这个问题的首选方案。公司只需最小的硬件集成便可以得到低价、高可用性的方案。

机镜像软件在局域网的服务器之间提供完全基于软件的镜像。一个服务器被指定为主服务，另一个为从服务器。客户只能对主服务器上的镜像盘或分区，从服务器上相应的磁盘或分区锁定以防对数据的存取，除非在主服务器上检测到故障状态。这是实现一个高性能方案要考虑的一个重要问题，因为数据的完整性是主要因素。尽管镜像软件可以单独使用，但是与热备和容错软件配合使用会显著增强它的功能。

(二)、数据镜像特性

在群集镜像环境下，一个镜像为主服务器上的整块硬盘或分区，其内容与从服务器上的相关磁盘通过 LAN 同步。在磁盘镜像建立后，主服务器和从服务器上的数据已经同步，并且两个服务器都已启动运行，会有下面的特征：

1. 最初的镜像建立后，系统禁止所有用户存取从服务器的镜像数据，不允许对从服务器上的镜像数据读和写。主服务器镜像则可以进行读和写操作。
2. 无论何时主服务器收到一个写请求，系统首先决定这个请求是否针对镜像磁盘数据。如果不是，写操作可以完全正常的完成。如果主服务器写请求是针对镜像卷的，那么请求被同时送到主从镜像卷中去。如果是同步镜像，则只有当主从镜像卷上执行的写请求都完成后，才继续下一个写操作；如果是异步镜像，则当主镜像卷上写请求都完成后，就可以继续下一个写操作。
3. 如果执行镜像数据写操作时从系统磁盘发生错误，那么从系统上的写操作将被中止，主系统继续自己的镜像数据写请求；如果是主系统磁盘发生错误，则会根据用户设定策略，继续用从系统磁盘进行读写，或者将应用切换到从系统。

(三)、群集特性

- 群集提供的全系统检测功能与镜像的数据镜像功能相辅相成，使群集系统无需昂贵的磁盘阵列子系统。
- 使用镜像功能将数据由一个主机系统镜像到另一个主机系统，并与群集相结合，可有效的避免磁盘阵列子系统中单点故障的可能性。
- 当主备服务器发生切换时。群集系统开机后并不是将备机数据分区或硬盘全部复制到主服务器，而是将备机中变化的数据部分复制到主服务器，从而减少了数据复制的时间，同时也增加了数据的安全性。
- 具备远程的灾难复原能力
 - 镜像群集在服务器端的数据复制能力。
 - 镜像群集有在局域网和广域网的数据复制能力。
 - 所有镜像群集都可以配合双机热备软件，实现异地的数据备份功能。

（四）、镜像群集的优点

- 避免了磁盘阵列的单点故障：对于需要共享磁盘阵列的群集，本身即是防范由于单个设备的故障导致服务中断，但磁盘阵列恰恰又形成了一个新的单点。（比如，服务器的可靠系数是 99.9%，磁盘阵列的可靠系数是 99.95%，则纯软群集的可靠系数是 $1-99.9%\times 99.95%=99.99\%$ ，而基于磁盘阵列的群集可靠系数则会略低于 99.95%
- 节约投资：不需购买昂贵的磁盘阵列。
- 不受距离的限制：两台服务器不需受 SCSI 电缆的长度限制（光纤通道的磁盘阵列也不受距离限制，但投资会大得多）。这样，可以更灵活地部署服务器，包括通过物理位置的距离来提高安全性。

第二章 易腾纯软群集概述

易腾纯软群集产品是来自中国本土的，拥有自主知识产权的纯软群集软件产品，其分为双机纯软（EterneMirrorHA），群集纯软（EterneMirrorCluster）两款软件。支持 Windows 操作系统平台，具有稳定可靠，真正数据镜像同步、数据镜像效率高、易于管理、备援方式灵活多样、节点和任务伸缩性强、保护用户投资、整合和优化用户系统环境、支持多种群集环境等特点。

一、软件组成

软件主要分为 4 大部分构成：易腾纯软群集操作界面、易腾纯软群集服务、EterneTarget、EterneDisk。

（一）、易腾纯软群集操作界面

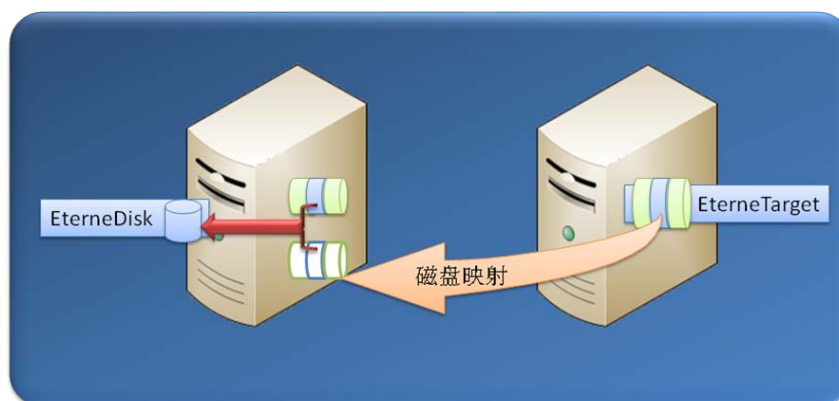
易腾纯软群集操作界面是由是 JAVA 编写的用户界面，通过该界面可以管理整个群集的所有节点和服务，同时直观地表示群集的各个任务运行状态和细节。

（二）、易腾纯软群集服务

易腾纯软群集服务是 Cluster 的核心服务程序。其负责执行来自管理界面的命令，协调各个节点的任务分配，传输和接收各个节点的心跳信息，切换和接管任务的管理中心，同时接收来自 Monitor 定时发送的资源状态信息，并进行解析和进行相关的操作。

（三）、EterneTarget

EterneTarget 负责在网络上（TCP/IP）输出本地磁盘的服务程序，其通过 iSCSI 协议命令把本地盘输出到群集中的其它服务器中，其它服务器通过 iSCSI Initiator 将其输出的磁盘作为本地盘来使用。



（四）、EterneDisk

EterneDisk 为块级别的实时镜像驱动程序，该驱动在文件系统之下，卷驱动之上的过滤驱动系统。其把系统的两块盘或分区合成一个虚拟的磁盘（EterneDisk），操作系统的读写都是针对该虚拟磁盘（EterneDisk）操作。如果是采用同步复制协议，一旦数据有写的操作，EterneDisk 会把该写操作同时写入两个磁盘或分区中，并且两个磁盘的数据都写入完成后，下一个写 IO 才能顺序写，从而达到两个磁盘的数据完全一致，并且是真正实时同步，而非异步实时同步解决方案。如果是采用异步复制协议，则会在主卷写入完成后自动进行写同步。目前市场上较流行的数据同步解决方案为文件级别的镜像方案，其提供的是异步实时同步镜像方案，该解决方案在某些意外情况，譬如网络上或主服务器的 buffer 中还有数据没有更新到服务器时突然主服务器宕机，那么便会出现从服务器的数据和主服务器数据不一致的情况。

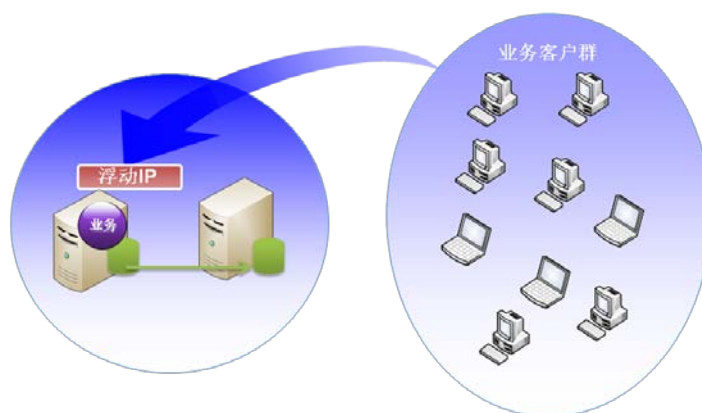
二、功能特点

（一）、分为两大产品

易腾纯软群集软件分为两大产品：双机纯软群集（EterneMirrorHA），群集纯软群集（EterneMirrorCluster）。EterneMirrorHA 针对两个服务器进行纯软双机热备的解决方案，其即支持数据镜像高可用，也支持共享磁盘高可用解决方案。EterneMirrorCluster 支持两个或两个以上节点的纯软群集软件，其功能涵盖 EterneMirrorHA 所有功能，同时支持更为复杂的多机环境。

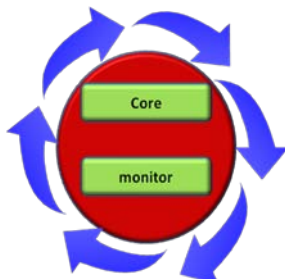
（二）、应用高可用性

提供高可用性解决方案是易腾纯软群集软件的核心功能，该解决方案提供双机热备系统，也提供多机的高可用解决方案。其原理都是当一台服务器出现故障时，可以由另一台服务器承担服务任务，从而在不需要人工干预的情况下，自动保证系统能持续提供服务。



（三）、自我修复能力

在集群系统的每一台服务器内，易腾纯软群集软件具有两个核心进程：**Service** 和 **Monitor**，他们互相监控（如：系统宕机、网络、**SCSI** 线缆是否断开），如果其中一个进程失败，另一个进程会立即进行恢复，最大程度保证双机热备系统自身健壮，避免双机丢失。



（四）、支持多条心跳路径

支持网卡(私网/公网)、磁盘心跳等多种心跳方式, 多重保障心跳检测途径, 配置多条心跳路径可以避免系统的单点故障, 进一步提高了系统的可靠性。

（五）、资源组管理

无论易腾纯软群集软件的安装还是管理都设计的十分简单和直观, 包括辨别关键的资源 (例如 IP 地址) 和指定这些资源到一个管理组 (**Resources Group**)。一旦被指定到一个管理组之后, 这些资源就被确认为被管理的资源。当其中一个服务器失败时, 这些资源将迁移到另一个服务器。易腾纯软群集软件可以单独地在每台服务器上安装和卸载, 用户可以自由地从一个组中删除和增加一个服务器, 无需全部卸载每台服务器。

（六）、单机热备高可用

在应用软件的检测时间间隔执行检测, 以察看资源本身是否失效。如果快速检查和深入检查均局部告失败, 系统将根据用户设定的策略尝试局部恢复资源, 实现单机的自动修复, 从而达到用户单机系统同样得到高可用级别。

✓ 本地磁盘失败转移

数据镜像卷中, 因为磁盘分区是以数据块级的镜像被复制到两个磁盘驱动器中, 它被操作系统当作一个分区对待。如果一个服务器上的 I/O 卡或磁盘驱动器失效, 对于数据的访问不会发生任何事情。因为在镜像卷中还有一个激活的 I/O 卡或磁盘驱动器存在。就像在单个服务器的内部进行磁盘复制一样, 其中一个失效不会有任何影响, 用户还可以从镜像卷中存活的驱动器上访问数据。(类似于 RAID 1)。

✓ 本地网卡失败转移

在主服务器上，可以加上另外一个 NIC 作为主服务器的备份 NIC。当主 NIC 出现故障时，EterneMirrorCluster 将可切换的 IP 地址让由本地的备份 NIC 接管，本地资源不需要切换，有效的减少的切换时间。

✓ 核心业务失败

当核心业务出现问题时，EterneMirrorCluster 会及时侦测到，并重新启动核心业务，从而不需要异地切换。

✓ 浮动 IP 失败

当浮动 IP 被释放或出现问题时，EterneMirrorCluster 会及时重新带入该浮动 IP，实现单机的高可用。



（七）、分区镜像

易腾纯软群集软件能够对磁盘的某个分区进行块镜像，每一个在被管理分区上的数据块都被镜像/复制到镜像对的另一个磁盘设备上。这一过程很像在一个服务器内部的磁盘复制，只不过镜像卷中的磁盘分别在不同的服务器上。

（八）、复制模式

同步镜像，即主服务器上的一个写操作必须在本地和远端磁盘都确定完成时，主服务器才认为写完成。这样，任何一个结点故障不会导致任何数据丢失，从而保证任何时候两台服务器的数据完全一致。只有两个结点同时发生故障时，才可能导致未完成的 IO 数据丢失。

异步镜像，即只要主服务器完居本地磁盘写操作，就认为写操作完成，需要复制的数据会在随后被复制到远端磁盘。当主结点发生故障切换时，最近更新的数据可能会丢失。

目前文件系统一级的镜像软件并不真正支持实时同步（即使支持其效率特别低并且大大降低操作系统的性能），其通常采用异步复制模式进行实时同步，一旦出现意外宕机，就可

能造成主从服务器的数据不一致。

（九）、智能在线初始化

EterneDisk 镜像卷在第一次进行初始化或重构时，系统会自动识别数据源和目标卷的文件系统数据，只将有用的数据进行镜像，不会对磁盘所有扇区进行同步，大大减少了初始化时间。比如分区中只有几百 M 数据的情况下，对 100GB 的分区进行镜像，初始化只需要很短时间即可完成。

EterneDisk 镜像卷可以进行在线初始化。在初始化期间，应用程序就可以对镜像卷进行在线读写操作，系统会自动对 IO 进行调度以保证数据的一致性，初始化期间所有的读写请求都会自动转移数据源卷上进行响应，数据源卷可以不一定是当前生产主机的本地磁盘。在初始化期间，镜像卷收到生产主机的读请求，若要读入位置的数据块正在拷贝中，等待该数据块拷贝完成后再进行读。在初始化期间，镜像卷收到生产主机的写请求，若要写入位置的数据尚未拷贝到目标卷，只需要写数据源卷即可返回主机成功，稍后同步任务会将整个数据块同步到目标卷，若要写入位置的数据块已经拷贝到目标卷，需要分别写入数据源卷和目标卷，若要写入位置的数据块正在拷贝中，等待该数据块拷贝完成后再分别写入数据源卷和目标卷。

（十）、自动处理磁盘错误

EterneDisk 可以自动、透明地对磁盘 I/O 错误进行处理。对于写操作，若底层逻辑卷 I/O 错误，则将错误传递给上层文件系统处理；如果是设备错误，则将其逻辑卷进行隔离，其后的读写操作只发生另一逻辑卷上。对于读操作，若发生底层逻辑卷设备错误，EterneDisk 对上层屏蔽 I/O 错误，透明地从另一逻辑卷读取所需的数据块，将其逻辑卷进行隔离，并透明地处理所有相应的 I/O 操作。若发生的是磁盘扇区错误，则会对其自动进行恢复；从另一逻辑卷读取相应数据块并写到相应扇区，并读取进行检验确认是恢复成功；如果失败，则对该数据块的扇区进行重新映射，将 EterneDisk 底层逻辑卷的相应坏扇区重新映射到空闲的扇区。

（十一）、脑裂(split-brain)恢复

Split-brain 发生在集群节点间所有网络连接同时发生故障，所有节点在连接断开时都会尝试切换为生产状态。这是一个潜在的故障状态，由于此时数据的修改可能会发生在任何节点上，并且没有复制给对端。这样，由于两个不一致的数据集被创建，将不能按常规方式合并。当发生 split-brain 的情况，建议采用手工解决 split-brain 方法，然后消除导致 split-brain 的原因。当然在某些情况下，也可以对 split-brain 进行自动故障恢复。EterneDisk 提供了以下解决 split-brain 的方法：

- 放弃较年轻生产节点的修改数据：这种设置模式，在网络连接重新建立，并且 EterneDisk 发现了 split-brain，则会丢弃掉故障发生后切换为生产节点的所有新数据变化（也就是故障发生前处于生产状态的节点的更新数据）。
- 放弃较老生产节点的修改数据：这种设置模式，与上一种相反，会放弃掉故障发生前处于生产状态节点的所有新数据变化。

- 放弃没有数据变化的节点：这种设置模式，EterneDisk 会自动从有数据变化的节点同步到在 split-brain 期间没有产生任何数据变化的结点。
- 放弃数据变化较少节点的数据变化：这种设置模式，EterneDisk 会自动放弃掉 split-brain 期间产生数据变化较少的结点数据更新，从数据变化较多的节点同步到数据变化的结点。
- 放弃本地节点的数据变化：这种设置模式，EterneDisk 会自动放弃掉最后拥有资源的节点的数据变化（即 split-brain 仲裁后为生产节点的数据变化）。
- 放弃远程节点的数据变化：这种设置模式，EterneDisk 会自动以最后拥有资源的节点的数据为最新数据（即 split-brain 仲裁后为生产节点的数据）。

（十二）、读负载均衡

EterneDisk 镜像卷中每个磁盘设备上所有的数据块同步之后，该镜像卷被认为是健康的（Health）。当 EterneDisk 镜像卷是健康状态时，对于关键任务数据的访问可以有两途径，进行负载均衡读，使得对镜像卷的读性能超过单盘的读性能，大大提高了应用系统的 IO 性能。读负载均衡只能适用于采用同步复制模式，否则可能导致读写数据不一致。

（十三）、快速重新同步

易腾纯软群集软件第一次进行初始化完全同步镜像后，即使服务器重新启动或易腾纯软群集软件服务重启把磁盘或分区重新同步，因为 EterneDisk 采用了专有的数据版本 UUID 技术、活动区域日志以及快速同步 Bitmap 技术，便能够快速进行重新同步，而不用再进行数据的比较校验，大大提高效率减少不必要的校验。该技术是目前文件系统级别镜像系统无法提供的，其服务或服务器一旦重新启动，为了保证数据完全一致性其必须进行数据集的数据校验，但如果数据集中的数据非常庞大或是小文件，那么校验的时间便非常长，在校验过程中服务是无法进行切换的。

（十四）、在线设备校验

在线设备校验确保用户以一个非常高效的方式进行不同节点的块数据一致性校验。对于同步复制，一般网络带宽比较高，为减少 CPU 使用和负载，在线校验不计算数据块的签名，而只直接依次比较数据验证源和验证目标每个数据块，如果不一致，则这个块被标记为需要同步，并在以后进行同步。如果是异步复制，在线数据校验是在一个节点（验证目标）存储设备上顺序计算每个数据块的数字签名。然后传输这个数字签名到另一节点（验证源），比较对应块的数字签名。如果数字签名不一致，则这个块被标记为需要同步，并在以后进行同步。由于传输只是一个数字签名，而不是完整数据块，所以在异步复制情况下，在线校验对网络带宽的使用是非常高效的。

对于一般情况，用户只需要进行快速在线设备校验就可以，因为 EterneDisk 底层存储设备对用户来说是透明的，没法进行直接修改，所以只要利用 EterneDisk 专有的 Bitmap 技术就可以快速比较数据块的一致性。

（十五）、稳定的数据镜像效率

因为易腾纯软群集软件采用的是块级别数据镜像，只镜像不同的数据块，因而镜像网络传输都以数据块为单位，因而有稳定的数据镜像效率。文件系统的过滤系统，其复制效率随文件的大小变化其镜像效率将有所不同，如果是批量的小文件，其镜像效率将非常低，因为其 IO 瓶颈出在从服务器上的写入操作上。从服务器在写文件时，其必须先打开这个文件，写完后需要关闭这个文件，同时还需要修改文件属性等操作，这样大大增加文件系统的 IO 量，和从服务器操作系统的系统资源。块级别数据镜像的写操作是直接对磁盘进行裸操作，并不增加操作的负担，从而大大的提高效率。

（十六）、纯软双机平滑过渡到纯软群集

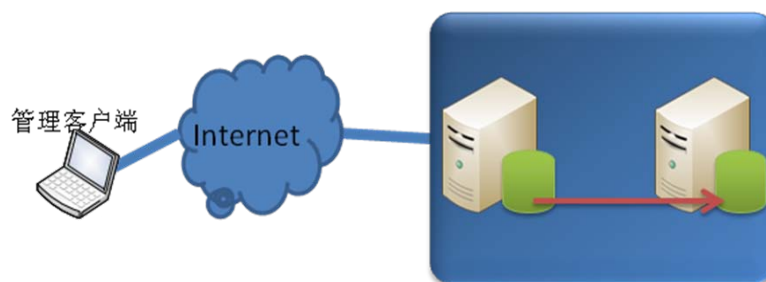
客户在装上了 EterneMirrorHA 纯软双机软件后，如果因环境的更改或需求的变化，想升级为群集软件，易腾公司支持其平滑过渡到 EterneMirrorCluster 软件，不用重新再进行配置客户的应用环境。

（十七）、支持多种群集工作模式

不仅支持传统的单活(Active/Passive)、双活(Active/Active)，还支持多对一、多对多、双机双柜等多种集群备份。目前 EterneMirrorCluster 最多支持 32 个节点的群集工作模式。

（十八）、采用 C/S 架构，管理员可以远程管理双机资源

易腾纯软群集软件客户端界面可以装载在远程客户端，不用装载到服务端。如果服务端装载了客户端软件，管理员可以在任何远程客户端（该客户端不用安装客户端软件）通过 IE 进行远程管理群集环境，进行资源组的操作。



（十九）、支持自动迁回

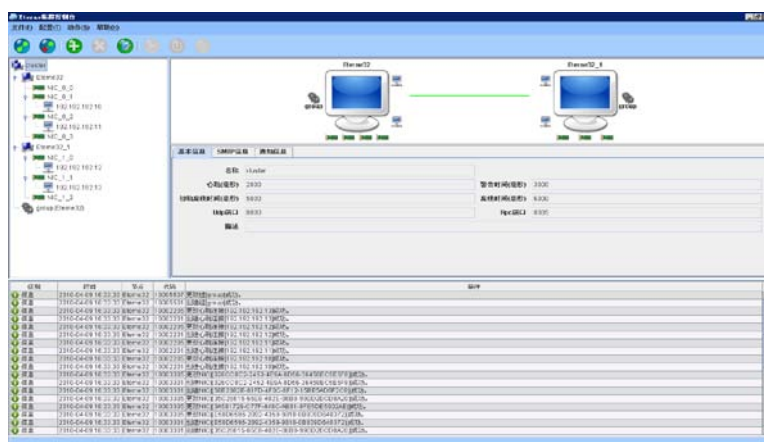
当首选节点发生故障后又恢复正常，如果允许自动迁回，备份节点在指定的恢复时间窗口内将资源组切换回首选节点运行；禁止自动迁回时，资源组将保留在备份结点上，管理员可在合适的时间手工切换。

（二十）、支持无为 ARP

当浮动 IP 切换到其它网卡上挂载时，Cluster 会主动进行 ARP 广播通知客户端该 IP 的 MAC 地址发生改变，让客户端主动更新映射关系，这样客户端便会连接到相关的服务上，避免因浮动 IP 切换造成客户端无法访问的情况。

（二十一）、直观的 GUI 用户界面

提供直观的 GUI 用户界面，用户可以交互式地对不同的集群系统进行配置、监控和远程管理，实时地显示出主机系统及服务的状态，大大减少了系统因用户人为的错误而引起的损失，同时基于向导形式的设定界面简化了系统安装，配置工作。



（二十二）、支持多种应用保护

可以保护大量的应用程序和服务，几乎涵盖了目前 Windows 平台上所有主流的应用和服务，如 Oracle、SQLServer、DB2、Sybase、Apache、Samba 等,除了这些主流的应用和服务。

（二十三）、全面支持用户自己开发的应用程序和服务

用户自定义的程序不用任何改变，可以直接进行配置进行监控。用户可以根据需要指定自定义服务或应用程序的运行身份，运行目录，甚至是否是 agent 方式进行检测运行。方式灵活多样，完全适应用户的各种需求。

第三章 易腾纯软群集软件应用规格

一、软硬件规格

(一)、支持各种类型服务器

支持 DELL, HP, PowerLeader, IBM, 浪潮, 联想等 X86 架构的所有服务器, 各个节点的服务器类型和硬件配置可以不一致。

(二)、支持的操作系统平台

目前 EterneCluster 支持 Window 系列和 Linux 系列的操作系统平台, 下面是具体支持的操作系统平台:

- Windows 系列操作系统
 - Windows 2000 系列
 - Windows 2003 系列
 - Windows 2008 系列

(三)、提供丰富的应用软件支持

支持多种应用的高可用, 并能与重要的应用如数据库: Microsoft SQL Server、Exchange 2000/2003、Oracle、文件服务器等紧密配合。EterneCluster 支持的应用包括:

- ✓ 数据库: MS SQL Server、Oracle, Sybase 等。
- ✓ 文件服务器:NFS, Samba 等。
- ✓ WEB 应用服务:IIS, Apache 等。
- ✓ 用户自定义。

二、群集规格配置

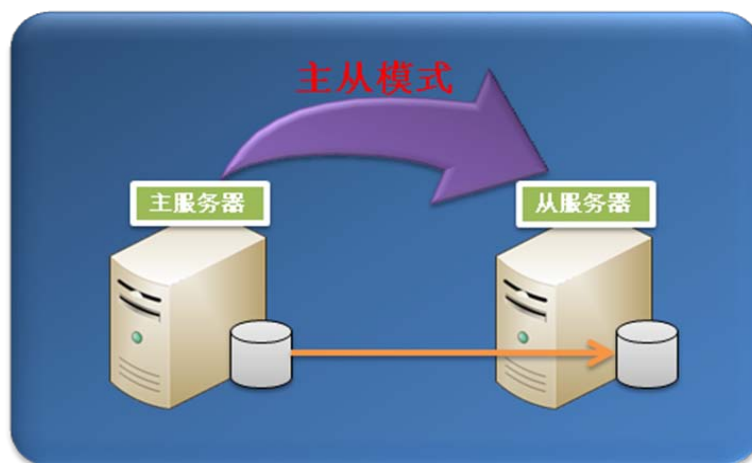
易腾纯软群集软件分支持多种类型的配置, 主要分为 2 大类: 双机纯软集群 EterneMirrorHA 和多机纯软集群 EterneMirrorCluster。

(一)、EterneMirrorHA 双机热备

纯软双机热备是目前大多用户采用的高可用环境, 简单的说就是两台服务器, 通过

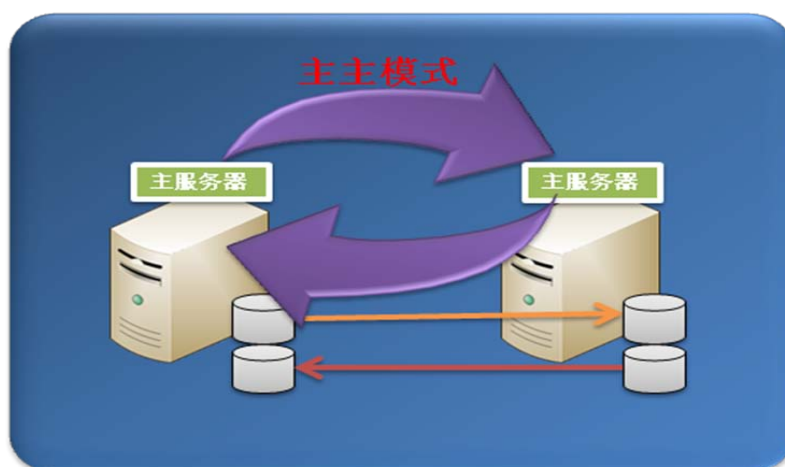
EterneMirrorHA 软件实现主从工作方式的双机环境。主从就是 Active-Standby 工作模式，容错软件作为不可缺少部分起到监控系统状态并在系统故障时，自动做出相应的反应，保证整个系统提供服务的不间断。同样双机集群可以实现 Active-Active 工作模式，即两台服务器分别运行不同的服务，并且互相作为备份，一旦某一台的系统出现故障，另外一台便会自动进行服务的接管。

(1) 主从配置



主从模式是纯软双机的典型配置模式，一般用于以前企业只有一台服务器，该服务器完全能够提供企业核心业务的服务，为了提升服务的可用性，便再购置一台服务器作为从服务器，新购置的服务硬件配置可用不同于以前的服务器。正常工作时，企业的核心业务在主服务器上运行，数据通过镜像实时保持主从服务数据完全一致，当主服务器出现故障时，企业的核心业务自动切换到从服务器上，保证业务不间断运行。

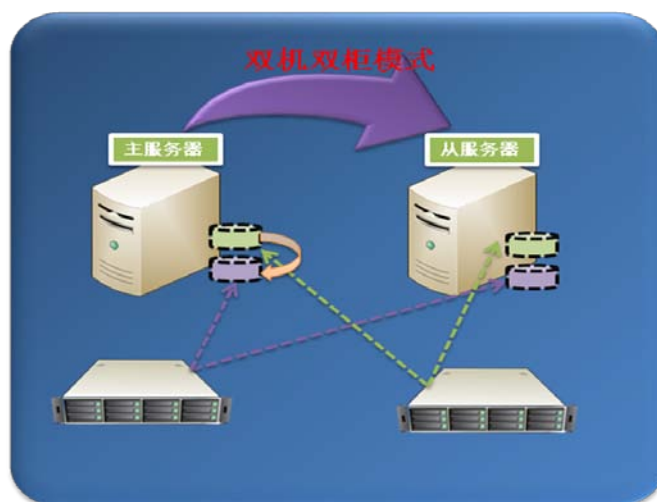
(2) 主主配置



这个环境下，有两套不同的应用运行在集群环境中，每台服务器运行各自的应用，在其中一台出现故障时，另外服务器将接管其服务。这种环境，在磁盘或分区的分配时，两个业

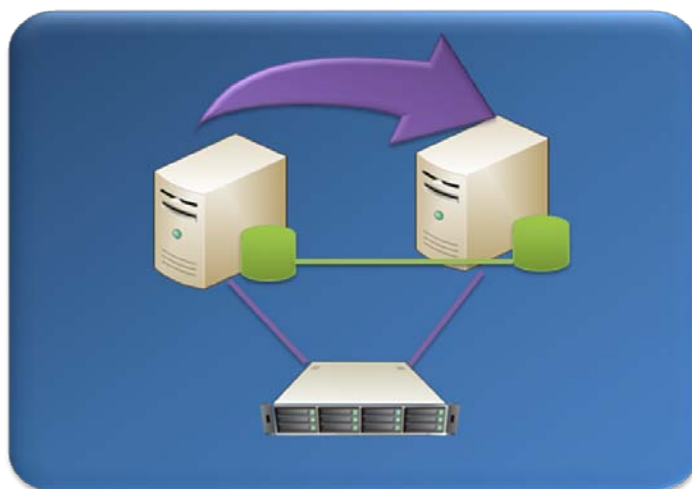
务应用的数据一定在不同的磁盘或分区上，否则无法配置该环境。这种配置可以大大提高设备的利用率，缺点是增加了系统的复杂度，而且对于某些特殊应用环境可能无法实施。双主方式是真正的双机互备，要求服务器具有较强的处理能力，来满足两个应用的需求。

(3) 双机双柜



双机容错系统其核心理念是防止单点失败的可能，譬如双机本身，心跳至少有 2 条，双机系统服务是双进程等。但一般的双机系统的磁盘阵列只有一个，一旦这个磁盘阵列出现问题，整个双机容错系统将出现崩溃。为了避免这个情况出现，易腾公司提供双机双柜的解决方案，让系统没有物理上的单点失败的故障，让任何情况下都能保障系统的不间断运行。EterneMirrorHA 提供双机双柜完整的解决方案，其不需要第三方的软件组合，其 EterneDisk 镜像驱动部分可以把两个磁盘阵列的 LUN 合并为一个磁盘，做类似于 RAID1 的工作，即使其中一个磁盘阵列坏掉了，核心业务也不会有任何影响，也不会发生任务切换。

(4) 镜像与磁盘共享型



该模式可以实现共享存储和数据镜像共存于一个双机热备的环境中。该模式适用于两种

情况：第一种是建立 2 个资源组，一个资源组是基于共享磁盘的资源组，另外一个资源组需要数据冗余，因而建立数据镜像组的高可用配置；第二种只建立一个资源组，该资源组的一部分数据使用共享磁盘，另外一部分数据使用磁盘镜像来保持一致的方式建立核心业务的高可用性。

(5) 全面支持虚拟机系统

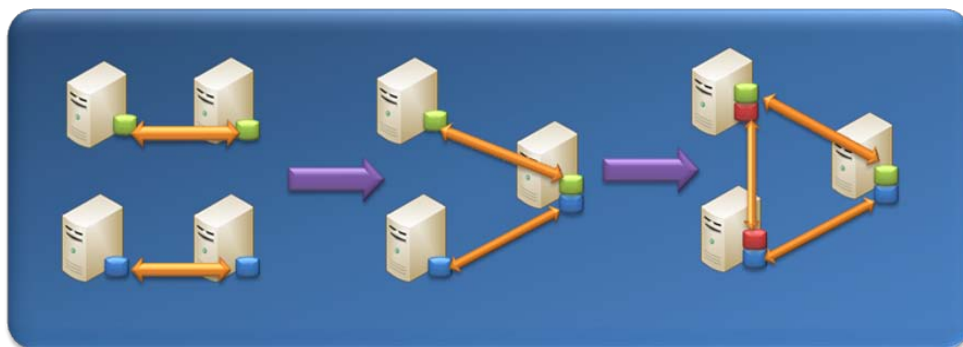


EterneMirrorHA 全面支持虚拟集群，它既可以做虚拟机之间做高可用解决方案，也可以对虚拟机上的业务之间构建高可用，同样也支持虚拟机和真实服务器之间做高可用解决方案。即当虚拟计算机上运行业务发生故障时，EterneMirrorHA 可以将其切换到另外一个虚拟计算机。如果虚拟机软件本身发生故障，或者物理服务器发生故障，EterneMirrorHA 能够将虚拟计算机整个切换到另外一台物理服务器上。此外，还可以利用虚拟计算机，对物理服务器进行备份。

(二)、EterneMirrorCluster

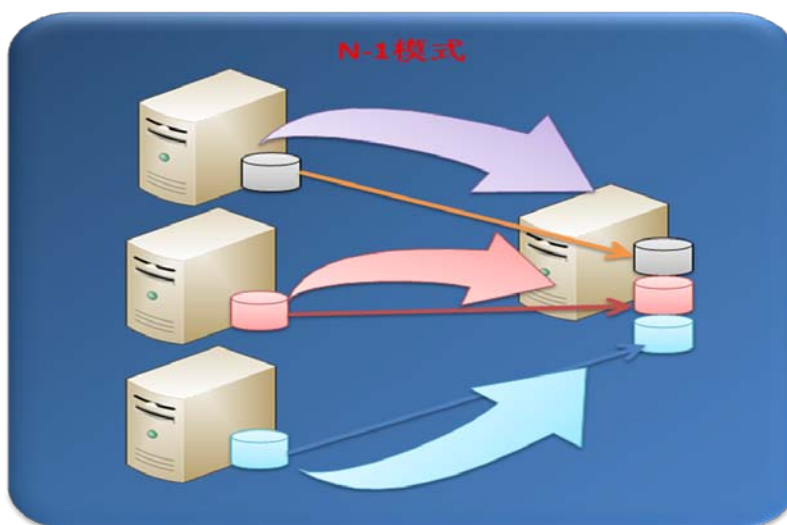
多节点集群在未来会有大量需求，大家对集群的认识目前还在双机环境下。而实际上，在有多个关键应用都具有高可用需求，在一个机房内建立多个双机系统显然是很浪费的，管理的复杂度又高，多节点集群可以有效的解决这个问题。对于多节点高可用，可以理解多机互备，同时运行多个不同应用，在其中任意服务器出现故障时，其它服务器根据备援策略进行接管服务，保证整个集群中的服务都能高可用。EterneMirrorCluster 是 EterneMirrorHA 的扩展模式，其适用于多节点模式，即包含了 EterneMirrorHA 的所有功能，又支持比较常用的多机集群配置：双机的扩展模式，N+1、M+N，容灾模式等。

(1) 双机扩展模式



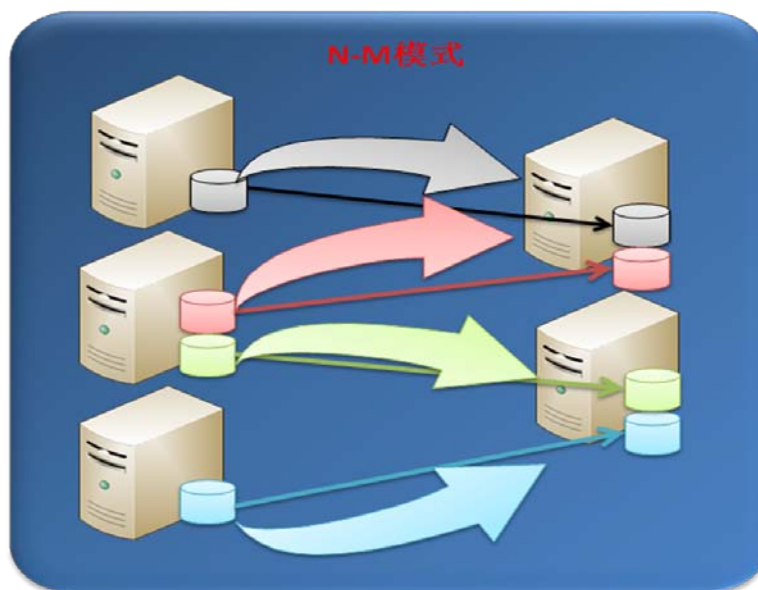
对于多个服务需要双机备份模式，可用利用多节点群集技术，用一台服务器同时对多个节点进行镜像复制，或者互为备份模式，实现多节点之间互为备份的构筑方式，从而节省硬件投入并且布局方式更加灵活。利用 EterneMirrorCluster 软件实现的多节点镜像技术，可以在不牺牲系统可靠性和性能的前提下，对目前现有的双机备份进行改造升级，进一步节省硬件投资，实现集中管理，进一步减少管理成本，实现更高更灵活的硬件利用率。

(2) N+1 配置



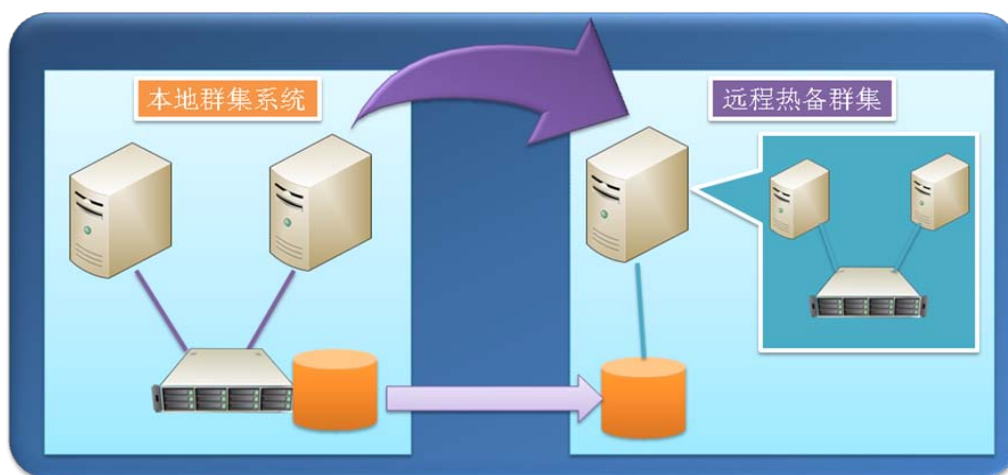
这个配置方式是所有的服务器都在一个 cluster 中，其中一个服务器作为 cluster 中其它服务的备份服务器，一旦某个主服务器系统出现故障，任务会按照策略把该服务器的任务切换到该备份服务器。某种意义上，其为双机系统的扩展版本，但这种体系结构配置对客户有多个服务的环境是非常优化的系统配置，同时能够让整个集群集中管理并且能够达到高可用。在配置磁盘卷时，不同的服务器对应的备份服务器必须是不同的磁盘或分区。

(3) M+N 配置



这个方式中多个生产服务器工作在一个集群中，通过合理定义备援策略，使环境中服务器相互备援，任何一个任务停止工作都将由其它服务器接管任务，是一种设备利用率最佳的部署方式。该部署方式同样有效的优化了系统结构，需要注意的是在定义备援策略时，仔细分析各服务器处理能力、任务的兼容性问题。同样该 M-N 的配置也是服务两两配对，对应的磁盘或分区也是独立的，不能互相有交集，否则数据会出现混乱。

(4) 容灾模式



一般群集系统都用于本地服务的高可用性，但对于一些非常关心数据和服务可靠性的客户，其担心本地因为火灾，地震等不可意料的情况造成本地数据和业务都不可恢复情况，希望要求远程有一个群集，当本地群集不可用时，远程的备份群集能够替代本地群集的工作，达到容灾效果。通常的远程容灾解决方案，对网络环境要求高、实施复杂、实现成本也比较昂贵。EterneMirrorCluster 将镜像型集群应用于远程容灾解决方案中，不但能够异地备份

业务数据，还可以对硬件资源进行监视，在本地群集发生故障时，还能够自动在备份群集中恢复业务，将业务服务的停止控制在最小限度。