# 集群脑裂

## 简介

在“双机热备”高可用（HA）系统中，当联系2个节点的“心跳线”断开时，本来为一个整体、动作协调的HA系统，就分裂为2个独立的个体。由于相互失去了联系，都以为是对方出现了故障，2个节点上的HA软件像“裂脑人”一样，本能地争抢共享资源、争起应用服务，这样就会发生严重后果：或者共享资源被瓜分、两边服务都起不来了；或者两边服务都起来了，但同时读写共享内存，导致数据损坏（常见的如数据库轮询着的联机日志出错）。

## HeartBeat

运行于备用主机上的HeartBeat可以通过以太网连接检测主服务器的运行状态，一旦其取法检测到主服务器的“心跳”则自动接管主服务器的资源。通常情况下，注、备服务器间的心跳连接是一个独立的物理连接，这个连接可以是串行线缆、一个由“交叉线”实现的以太网连接。HeartBeat甚至可同时通过多个物理连接检测主服务器的工作状态，而其只要能通过其中一个连接收到主服务器处于活动状态的信息，就会认为主服务器处于正常状态。

从实践经验的角度来看，建议为HeartBeat配置多条独立的物理连接，以避免HeartBeat通信线路本身存在单点故障。

## 连接方式

1. 串行电缆：被认为是比以太网连接安全性稍好些的连接方式，因为hacker无法通过串行连接运行诸如telnet、ssh或者rsh类的程序，从而可以降低其通过已劫持的服务器再次侵入备份服务器的几率。但串行线缆受限于可用长度，因此主、备服务器的举例必须非常短；
2. 以太网连接：使用此方式可以消除串行线缆在长度方面的限制，因此可以通过此连接主、备服务器间同步文件系统，从而减少了从正常通信连接带宽的占用

基于冗余的角度考虑，应该在主、备服务器使用两个物理连接传输HeartBeat的控制信息，这样可以避免在一个网络或者线缆故障时导致两个节点同时认为自己是唯一处于活动状态的服务器，从而出现争用资源的情况，这样争用资源的场景就是所谓的“脑裂”（split-brain）或“partitioned luster”。在两个节点共享同一个物理设备资源的情况下，脑裂会产生相当可怕的后果。

## 预防措施

为了避免出现脑裂，可采用如下的预防措施：

1. 添加冗余的心跳线，例如双心跳线，尽量减少“脑裂”发生几率；
2. 启动磁盘锁，正在服务一方锁住共享磁盘，“脑裂”发生时，让对方完全“抢不走”共享磁盘资源。但使用磁盘锁也会有一个问题，如果占用共享盘一方不主动“解锁”，另一方就永远得不到共享盘。现实中假如服务节点突然死机或崩溃，就不可能执行解锁命令。后备节点也就接管不了共享资源和应用服务。于是有人在HA中设计了“智能锁”，即正在服务的一方只要发现心跳线全部断开（觉察不到对端）时才启用磁盘锁，平时就不上锁。
3. 设置仲裁机制，例如设置参考IP（例如网关IP），当心跳完全断开时，2个节点都各自ping一下参考IP，不通则表明断点就出在本端，不仅“心跳”、还兼对外“服务”的本端网络链路断了，即使启动（或继续）应用服务也没有用了，那就主动放弃竞争，让能够ping通参考IP的一端去启动服务。更保险一些，ping不通参考IP的一方干脆自动重启，以彻底释放可能还占用着的那些共享资源。