**关于网络心跳与磁盘心跳超时多久会引发节点重启的问题**

原创 2015年10月09日 09:53:26

标签：

[heartbeat fatal](http://so.csdn.net/so/search/s.do?q=heartbeat%20fatal&t=blog) /

[网络心跳](http://so.csdn.net/so/search/s.do?q=%E7%BD%91%E7%BB%9C%E5%BF%83%E8%B7%B3&t=blog) /

[磁盘心跳](http://so.csdn.net/so/search/s.do?q=%E7%A3%81%E7%9B%98%E5%BF%83%E8%B7%B3&t=blog) /

[misscount](http://so.csdn.net/so/search/s.do?q=misscount&t=blog) /

[disktimeout](http://so.csdn.net/so/search/s.do?q=disktimeout&t=blog)

      最近时间，一从事系统运维的同事问我，说，因为需要调整网络，需要对一套生产中的RAC的期中一台主机的网络心跳线插拨一次，因为走停机检修手续流程太长，想利用oracle的心跳可延迟时间内插拨完成，所以想搞清楚心跳的延时时间。

      遇到这个问题，首先，我们要知道ORACLE是由OCSSD守护进程在负责心跳机制的，并且OCSSD维护的是网络和磁盘两种心跳同步机制。以及，不能对上述问题一概而论，因为在不同OS平台，不同cluster版本，采用的不同集群软件一环境下，默认的最大延时时间是不同的。并且，该时间是可以调整的，并不能确保他要操作的系统使用的就是默认值，在cluster上决定节点分离的参数主要有misscount（简称MC，定义网络心跳的最大缺失时间）、disktimeout（简称DTO，定义磁盘心跳的最大缺失时间）、reboottime（重启时间）三个参数所决定，三个参数的单位均为秒，并且具有关联性，期中参数值的修改会关联影响整体的节点分离重启时间。

      对于因为网络心跳和磁盘心跳超时最经典的日志信息如下：

网络心跳：

|  |
| --- |
| CRS-1612:node XXdb1 (1) at 50% heartbeat fatal, eviction in 29.118 seconds |

     --上面这条信息，看到缺失约30秒时，超时的比例是50%，就可以猜出，在这台机器上，网络心跳的最大缺失时间应该是60秒，该机器是linux平台oracle10gR2.

磁盘心跳：

|  |
| --- |
| [cssd(7361)]CRS-1613:No I/O has completed after 90% of the maximum interval. Voting file /dev/asm-ocr will be considered not functional in 19270 milliseconds |

      对于网络心跳与磁盘心跳最大缺失时间的权威说明，在metalink上，有一篇**文档 ID 294430.1**的文章，详细的说出了上面的问题。

**1、  不同平台与使用oracle cluster与在不同oracle版本下的默认最大网络延时**

|  |  |  |
| --- | --- | --- |
| **OS** | **10g (R1 &R2)** | **11g** |
| Linux | 60 | 30 |
| Unix | 30 | 30 |
| VMS | 30 | 30 |
| Windows | 30 | 30 |

**2、  使用第三方集群软件的最大网络延迟时间**

      使用第三方集群软件，默认为600秒的延迟时间，并且，决定脑裂的工作由第三方集群软件所决定

    对于在AIX平台下，数据库版本为10.2.0.3的环境下，网络超时时间却只有30秒

**3、  CSS心跳机制及相互的关系**

      主要涉及关于网络心跳、磁盘心跳及rebootime的关系

      The synchronization services component (CSS) ofthe Oracle Clusterware maintains two heartbeat mechanisms 1.) the diskheartbeat to the voting device and 2.) the network heartbeat  across theinterconnect which establish and confirm valid node membership in the cluster.Both of these heartbeat mechanisms have an associated timeout value. The diskheartbeat has an internal i/o timeout interval (DTO Disk TimeOut), in seconds,where an i/o to the voting disk must complete. The misscount parameter (MC), asstated above, is the maximum time, in seconds, that a network heartbeat can be missed. The disk heartbeat i/o timeout interval is directly related tothe misscount parameter setting. There has been some variation in thisrelationship

between versions asdescribed below:

|  |  |
| --- | --- |
| 9.x.x.x | NOTE, MISSCOUNT WAS A  DIFFERENT ENTITY IN THIS RELEASE |
| 10.1.0.2 | No one should be on this version |
| 10.1.0.3 | DTO = MC - 15 seconds |
| 10.1.0.4 | DTO = MC - 15 seconds |
| 10.1.0.4+Unpublished Bug 3306964 | DTO = MC - 3 seconds |
| 10.1.0.4 with CRS II Merge patch | DTO =Disktimeout (Defaults to 200 seconds) Normally OR Misscount seconds only during initial Cluster formation or Slightly before reconfiguration |
| 10.1.0.5 | IOT = MC - 3 seconds |
| 10.2.0.1 +Fix for unpublished Bug 4896338 | IOT=Disktimeout (Defaults to 200 seconds) Normally OR Misscount seconds only during initial Cluster formation or Slightly before reconfiguration |
| 10.2.0.2 | Same as above (10.2.0.1 with Patch [Bug:4896338](https://support.oracle.com/epmos/faces/BugDisplay?parent=DOCUMENT&sourceId=294430.1&id=4896338) |
| 10.1 - 11.2 | During node join and leave (reconfiguration) in a cluster we need to reconfigure, in that particular case we use Short Disk TimeOut (SDTO) which is in all versions SDTO = MC – reboottime (usually 3 seconds) |

    从上面来看，MC时间的变化，会影响DTO的值，10.1.0.3和10.1.0.4安装3306964补丁后的DTO时间是不一样的，在10.1.0.5以后，分裂时间的概念也生了变化，以由以的DTO，变成了IOT（i/o timeoutinterval）。

**4、  misscount、disktimeout、reboottime三个参数的修改**

      通过crsctlget css misscount、crsctlget css disktimeout、crsctlget css reboottime可

      以分别查出其设置值，如：

|  |
| --- |
| [grid@rac1 ~]$ crsctl get css misscount  CRS-4678: Successful get misscount 30 for Cluster Synchronization Services.  [grid@rac1 ~]$ crsctl get css disktimeout  CRS-4678: Successful get disktimeout 200 for Cluster Synchronization Services.  [grid@rac1 ~]$ crsctl get css reboottime  CRS-4678: Successful get reboottime 3 for Cluster Synchronization Services. |

 修改三个参数值的方法则是：

|  |
| --- |
| $CRS\_HOME/bin/crsctl set css misscount <n>    #### where <n> is the maximum private network latency in seconds  $CRS\_HOME/bin/crsctl set css reboottime <r> [-force]  #### (<r> is seconds)  $CRS\_HOME/bin/crsctl set css disktimeout <d> [-force] #### (<d> is seconds) |

**5、  小结**

      从上面信息来看，在ORACLE RAC在生产情况下拨插网线，是有一定风险的，具体需要根据实际生产运行环境而决定，并且，在超时时没有到达到重启条件时，数据库为了保证数据的一致性，也是处于HANG状态的。致于能不能在线操作，需要三思。

本文作者：黎俊杰（网名：踩点），从事”系统架构、操作系统、存储设备、数据库、中间件、应用程序“六个层面系统性的性能优化工作

欢迎加入 系统性能优化专业群 ，共同探讨性能优化技术。群号：258187244