|  |  |  |
| --- | --- | --- |
| **Oracle Grid Infrastructure: Understanding Split-Brain Node Eviction (Doc ID 1546004.1)** | [To Bottom](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\o%20To%20Bottom) |  |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **In this Document**   |  |  | | --- | --- | |  | [Purpose](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20REF_PURPOSE) | |  | [Scope](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20REF_SCOPE) |  |  |  | | --- | --- | |  | [Details](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20REF_TEXT) | |  | [What does "split brain" mean?](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section31) |  |  |  | | --- | --- | |  | [Why is this a problem?](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section32) | |  | [How does the clusterware resolve a "split brain" situation?](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section33) |  |  |  | | --- | --- | |  | [Identifying a split-brain eviction](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section34) | |  | [Finding the cohort](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section35) |  |  |  | | --- | --- | |  | [Understanding the cohort message](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section36) | |  | [Using the cohort message to identify interconnect network issues](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section37) |  |  |  | | --- | --- | |  | [**Follow-up Action**](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section38) | |  | [Community Discussions](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20aref_section39) |  |  |  | | --- | --- | |  | [References](https://support.oracle.com/epmos/faces/DocumentDisplay?_afrLoop=344915570894569&id=1546004.1&_adf.ctrl-state=18w595m6et_104%20\l%20REF) |     **APPLIES TO:**  Oracle Database - Enterprise Edition - Version 11.2.0.1 and later  Information in this document applies to any platform.  **PURPOSE**  The purpose of this note is to explain split-brain node evictions in Oracle Clusterware release 11.2  **SCOPE**  The intended audience of this note is Oracle Clusterware 11.2 administrators at any level of expertise. As written, this note applies only to 11.2.  **DETAILS**  Missed network heartbeat (NHB) evictions happen when ocssd of the surviving node loses contact with the evicted node over the interconnect. The nodes must be able to communicate over the interconnect to avoid a "split brain" situation. In the case of a "split brain" node eviction, one node aborted itself to avoid "split brain" when communication over the interconnect was compromised.  **What does "split brain" mean?**  "Split brain" means that there are 2 or more distinct sets of nodes, or "cohorts", with no communication between the two cohorts.  For example:  Suppose there are 4 nodes named A, B, C, D, in the following situation  \* Nodes A,B can talk to each other; nodes C,D can talk to each other  \* But A and B cannot talk to C or D, and vice versa  Then there are two cohorts: {A, B} and {C, D}.  **Why is this a problem?**  In a split-brain situation, there are in a sense two (or more) separate clusters working on the same shared storage. This has the potential for data corruption. So the split-brain must be resolved.  **How does the clusterware resolve a "split brain" situation?**  Oracle Clusterware handles the split-brain by terminating all the nodes in the SMALLER cohort.  If both of the cohorts are the same size, the cohort with the lowest numbered node in it survives.  The clusterware identifies the LARGEST cohort, and aborts all the nodes which do NOT belong to that cohort.  **Identifying a split-brain eviction**  In a split-brain node eviction, the following message is present in the ocssd log ($GRID\_HOME/log/<hostname>/ocssd/ocssd.log) of the evicted node:  clssnmCheckDskInfo: Aborting local node to avoid splitbrain.  And earlier in the same log, within 10 minutes prior to "clssnmCheckDskInfo: Aborting local node" message:  clssnmPollingThread: node %s (%n) at <X>% heartbeat fatal, removal in...    **Finding the cohort**  The split-brain message in the ocssd.log will show "cohort" information. For example:  2012-12-28 20:26:25.803: [    CSSD][1111296320]clssnmCheckDskInfo: My cohort: 1  2012-12-28 20:26:25.803: [    CSSD][1111296320]clssnmCheckDskInfo: Surviving cohort: 2,3,4  2012-12-28 20:26:25.803: [    CSSD][1111296320](:CSSNM00008:)clssnmCheckDskInfo: Aborting local node to avoid splitbrain. Cohort of 1 nodes with leader 1, sprora01, is smaller than cohort of 3 nodes led by node 2, sprora02, based on map type 2  **Understanding the cohort message**  In a split-brain situation, ocssd on each node records on the voting disk the set of nodes it can communicate with. Each set is known as a "cohort". When there are two (or more) mutually non-intersecting sets, we have a "split-brain" situation. It means that there are two (or more) separate sets of nodes which cannot talk to each other over the interconnect.  For example, in the above quote  **My cohort: 1**  **Surviving cohort: 2,3,4**  The meaning of these messages is  \* "My cohort: 1" => The list of nodes I can communicate with: 1  \* "Surviving cohort: 2,3,4" => From the voting disk, I know that nodes 2,3,4 can all communicate with each other.  \* "Cohort of 1 nodes with leader 1, sprora01, is smaller than cohort of 3 nodes led by node 2, sprora02"  => Oracle Clusterware has identified that the cohort {1} is smaller than the cohort {2,3,4}.  Oracle Clusterware handles the split-brain by terminating all the nodes in the SMALLER cohort. In this case, the smaller cohort is {1}. Therefore, ocssd on node {1} aborts the node.  **Using the cohort message to identify interconnect network issues**  The cohort message describes which nodes can communicate with each other.  Each cohort is a set of nodes that can talk to each other, and cannot talk to the nodes NOT in the cohort.  In the above example, the cohort message tells us that nodes {2,3,4} are all in communication; node 1 is not in communication with any of them.    **Follow-up Action**  The private network between node 1 and the other 3 nodes should be checked.  Please refer to the following note to check private interconnect network:  [Document 1534949.1](https://support.oracle.com/epmos/faces/DocumentDisplay?parent=DOCUMENT&sourceId=1546004.1&id=1534949.1) - Oracle Grid Infrastructure: How to Troubleshoot Missed Network Heartbeat Evictions    **Community Discussions**  Still have questions? Use the communities window below to search for similar discussions or start a new discussion on this subject.  Note: Window is the **LIVE** community not a screenshot.  Click [here](https://community.oracle.com/community/support/oracle_database/database_-_rac_scalability%20/o%20Database%20-%20RAC/Scalability%20Community) to open in main browser window. |