Overview of the Tiebreaker software

ONTAP MetroCluster

Zachary Wambold
April 06, 2021

Table of Contents

Overview of the Tiebreaker software

It is helpful to understand what the NetApp MetroCluster Tiebreaker software is and how it distinguishes between types of failures so that you can monitor your MetroCluster configurations efficiently. You use the Tiebreaker CLI to manage settings and monitor the status and operations of MetroCluster configurations.

Detecting failures with NetApp MetroCluster Tiebreaker software

The Tiebreaker software resides on a Linux host. You need the Tiebreaker software only if you want to monitor two clusters and the connectivity status between them from a third site. Doing so enables each partner in a cluster to distinguish between an ISL failure, when inter-site links are down, from a site failure.

After you install the Tiebreaker software on a Linux host, you can configure the clusters in a MetroCluster configuration to monitor for disaster conditions.

How the Tiebreaker software detects site failures

The NetApp MetroCluster Tiebreaker software checks the reachability of the nodes in a MetroCluster configuration and the cluster to determine whether a site failure has occurred. The Tiebreaker software also triggers an alert under certain conditions.

Components monitored by the Tiebreaker software

The Tiebreaker software monitors each controller in the MetroCluster configuration by establishing redundant connections through multiple paths to a node management LIF and to the cluster management LIF, both hosted on the IP network.

The Tiebreaker software monitors the following components in the MetroCluster configuration:

- Nodes through local node interfaces
- Cluster through the cluster-designated interfaces
- Surviving cluster to evaluate whether it has connectivity to the disaster site (NV interconnect, storage, and intercluster peering)

When there is a loss of connection between the Tiebreaker software and all of the nodes in the cluster and to the cluster itself, the cluster will be declared as “not reachable” by the Tiebreaker software. It takes around three to five seconds to detect a connection failure. If a cluster is unreachable from the Tiebreaker software, the surviving cluster (the cluster that is still reachable) must indicate that all of the links to the partner cluster are severed before the Tiebreaker software triggers an alert.

All of the links are severed if the surviving cluster can no longer communicate with the cluster at the disaster site through FC (NV interconnect and storage) and intercluster peering.

Failure scenarios during which Tiebreaker software triggers an alert

The Tiebreaker software triggers an alert when the cluster (all of the nodes) at the disaster site is down or unreachable and the cluster at the surviving site indicates the “AllLinksSevered” status.

The Tiebreaker software does not trigger an alert (or the alert is vetoed) in the following scenarios:

- In an eight-node MetroCluster configuration, if one HA pair at the disaster site is down
- In a cluster with all of the nodes at the disaster site down, one HA pair at the surviving site down, and the cluster at the surviving site indicates the “AllLinksSevered” status

  The Tiebreaker software triggers an alert, but ONTAP vetoes that alert. In this situation, a manual switchover is also vetoed

- Any scenario in which the Tiebreaker software can either reach at least one node or the cluster interface at the disaster site, or the surviving site still can reach either node at the disaster site through either FC (NV interconnect and storage) or intercluster peering

Related information

Risks and limitations of using MetroCluster Tiebreaker in active mode

How the Tiebreaker software detects intersite connectivity failures

The MetroCluster Tiebreaker software alerts you if all connectivity between the sites is lost.

Types of network paths

Depending on the configuration, there are three types of network paths between the two clusters in a MetroCluster configuration:

- **FC network (present in fabric-attached MetroCluster configurations)**

  This type of network is composed of two redundant FC switch fabrics. Each switch fabric has two FC switches, with one switch of each switch fabric co-located with a cluster. Each cluster has two FC switches, one from each switch fabric. All of the nodes have FC (NV interconnect and FCP initiator) connectivity to each of the co-located FC switches. Data is replicated from cluster to cluster over the ISL.

- **Intercluster peering network**

  This type of network is composed of a redundant IP network path between the two clusters. The cluster peering network provides the connectivity that is required to mirror the storage virtual machine (SVM) configuration. The configuration of all of the SVMs on one cluster is mirrored by the partner cluster.

- **IP network (present in MetroCluster IP configurations)**

  This type of network is composed of two redundant IP switch networks. Each network has two IP switches, with one switch of each switch fabric co-located with a cluster. Each cluster has two IP switches, one from each switch fabric. All of the nodes have connectivity to each of the co-located FC switches. Data is replicated from cluster to cluster over the ISL.

Monitoring intersite connectivity

The Tiebreaker software regularly retrieves the status of intersite connectivity from the nodes. If NV interconnect connectivity is lost and the intercluster peering does not respond to pings, then the clusters assume that the sites are isolated and the Tiebreaker software triggers an alert as “AllLinksSevered”. If a cluster identifies the “AllLinksSevered” status and the other cluster is not reachable through the network, then the Tiebreaker software triggers an alert as “disaster”.

How different disaster types affect Tiebreaker software detection time

For better disaster recovery planning, the MetroCluster Tiebreaker software takes some time in detecting a disaster. This time spent is the “disaster detection time”. The MetroCluster Tiebreaker software detects the site disaster within 30 seconds from the time of occurrence of the disaster and triggers the disaster recovery operation to notify you about the disaster.

The detection time also depends on the type of disaster and might exceed 30 seconds in some scenarios, mostly known as “rolling disasters”. The main types of rolling disaster are as follows:

- Power loss
- Panic
- Halt or reboot
- Loss of FC switches at the disaster site

Power loss

The Tiebreaker software immediately triggers an alert when the node stops operating. When there is a power loss, all connections and updates, such as intercluster peering, NV interconnect, and MailBox disk, stop. The time taken between the cluster becoming unreachable, the detection of the disaster, and the trigger, including the default silent time of 5 seconds, should not exceed 30 seconds.

Panic

The Tiebreaker software triggers an alert when the NV interconnect connection between the sites is down and the surviving site indicates the “AllLinksSevered” status. This only happens after the coredump process is complete. In this scenario, the time taken between the cluster becoming unreachable and the detection of a disaster might be longer or approximately equal to the time taken for the coredump process. In many cases, the detection time is more than 30 seconds.

If a node stops operating but does not generate a file for the coredump process, then the detection time should not be longer than 30 seconds.

Halt or reboot

The Tiebreaker software triggers an alert only when the node is down and the surviving site indicates the “AllLinksSevered” status. The time taken between the cluster becoming unreachable and the detection of a disaster might be longer than 30 seconds. In this scenario, the time taken to detect a disaster depends on how long it takes for the nodes at the disaster site to be shut down.

Loss of FC switches at the disaster site (fabric-attached MetroCluster configuration)

The Tiebreaker software triggers an alert when a node stops operating. If FC switches are lost, then the node tries to recover the path to a disk for about 30 seconds. During this time, the node is up and responding on the peering network. When both of the FC switches are down and the path to a disk cannot be recovered, the node produces a MultiDiskFailure error and halts. The time taken between the FC switch failure and the number of times the nodes produced MultiDiskFailure errors is about 30 seconds longer. This additional 30 seconds must be added to the disaster detection time.

About the Tiebreaker CLI and man pages

The Tiebreaker CLI provides commands that enable you to remotely configure the Tiebreaker software and monitor the MetroCluster configurations.

The CLI command prompt is represented as NetApp MetroCluster Tiebreaker::>.

The man pages are available in the CLI by entering the applicable command name at the prompt.