Windows Failover Cluster Quorum Modes in SQL Server Always On Availability Groups

https://www.sqlshack.com/windows-failover-cluster-quorum-modes-in-sql-server-always-on-availability-groups/

This article gives an overview of Windows Failover Cluster Quorum modes that is necessary for SQL Server Always on Availability Groups.

Introduction

SQL Server Always On is a flexible design solution to provide high availability (HA) and disaster recovery (DR). It is built upon the Windows Failover Cluster, but we do not require the shared storage between the failover cluster nodes. Each replica maintains its storage, system databases, agent jobs, login, linked servers, credentials and integration service packages. Before SQL Server 2012, we use database mirroring and the log shipping for HADR solutions.

SQL Server Always On replaces the database mirroring and log shipping for the HADR solution. We define availability groups as a single unit of failure to the secondary replica. We also define a SQL listener to redirect all application connections to the primary replica, regardless of which replica is having a primary role.

The basic requirements for configuring SQL Server Always On is the failover clustering. All the participating nodes should be part of the failover cluster.

- **Note:** Starting from SQL 2017, we can configure clusterless AG groups as well. You can refer to the article, Read Scale Availability Group in a clusterless availability group for more details. This article assumes the SQL Server Always on with failover clustering.

Once you configure a new availability group, it validates the Windows Failover Cluster quorum vote configuration.

In this article, we will discuss the quorum in the Windows cluster and various methods to configure the nodes voting.

Overview of Quorum in Windows Failover Cluster

In the Windows Failover Cluster, we have two or more cluster nodes, and these nodes communicate with each other on the UDP port 3343. Suppose due to some unforeseen circumstances; nodes cannot communicate with each other. In this case, all cluster nodes assume that they need to bring the resources online. Multiple nodes might try to bring the same resource online at the same time, but it is not possible to do so. This situation is called a Split-Brain situation, and we never want to be in this situation. We might face data corruption issues in this case.

We can avoid a split-brain scenario in Windows Failover Cluster by bringing the resources on a single node. Windows cluster achieve this by implementing a voting mechanism. Quorum checks for the minimum number of votes required to have a majority and own the resources.

Each cluster node is allowed to cast its single vote. In each scenario, more than half of the voters must be online, and they should be able to communicate with each other. If the cluster does not achieve the majority, it fails to start the cluster services and the entire cluster is down.

- The cluster node in the Windows Failover Cluster tries to begin a communication with all other nodes in the cluster group
- Once communication is set up between the available nodes, they talk to each other to achieve a majority
- If after voting, a majority is established, the cluster resources become online
- In the case of no majority, cluster services are set to offline status for a temporary basis, and nodes wait for other nodes to come online so that the majority can be established
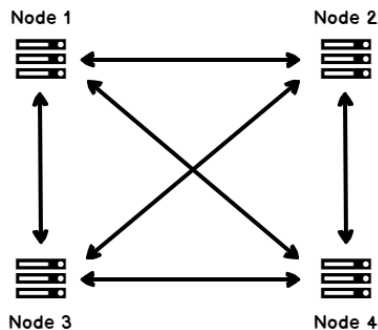
Quorum Modes for Windows Failover Cluster

In the previous section, we overviewed the Windows Cluster Quorum tool. Based on the information, we can set the following formula for the majority in a cluster.

Quorum = FLOOR (N/2+1)

Where N=number of nodes.

Suppose we have four nodes Cluster and each cluster nodes runs a SQL Server instance.

As per the above formula, we require three votes for the cluster to remain online.

Quorum= 4/2 +1 = 3

Now imagine the scenario

- Node 1 and Node 2 cannot communicate between Node 3 and Node 4. Node 1 and Node 2 can communicate with each other
- Node 3 and Node 4 cannot communicate between Node 1 and Node 2. Node 3 and Node 4 can communicate with each other
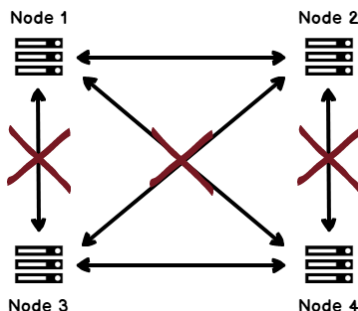
In this scenario, Node 1 and Node 2 tries to bring the resources online that is owned by Node 3 and Node 4.

Similarly, Node 3 and Node 4 tries to bring the resources online that are owned by Node 1 and Node 2.

To avoid these scenarios, we have various Quorum configurations, and we should set the Quorum configuration carefully.

Let's discuss these Quorum configuration modes.

**Node Majority Quorum configuration mode in Windows Failover Cluster**

- In the Node Majority, each cluster node contributes a Vote.
- This Quorum configuration handles a loss of half of the cluster nodes after rounding off to a lower number.
- It is a recommended method to use with an odd number of cluster nodes.
- Let's say we have three cluster nodes, and one node is not able to communicate with the other nodes.
- In this case, as shown in the following image, we have the node majority with two online nodes.
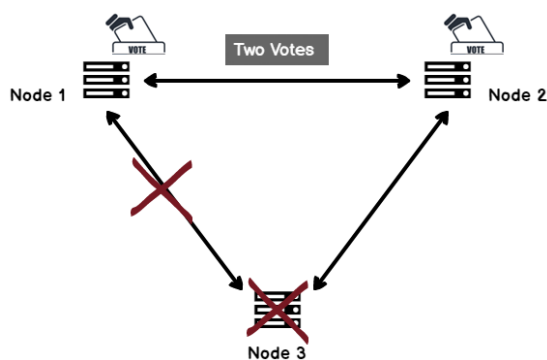

Similarly, in the case of five nodes cluster, it can handle up to two nodes failures.

Node & Disk Majority Quorum configuration mode in Windows Failover Cluster

- In this Quorum configuration mode, each cluster node gets one vote, and additionally, one vote comes from clustered disk (quorum disk).
- This disk can be a small size high available disk and should be part of the cluster group.
- It is suitable for a cluster with even nodes.
- Quorum disk also stores the cluster configuration data.
- Usually, Windows administrators set the size of the Quorum disk as 1 GB.
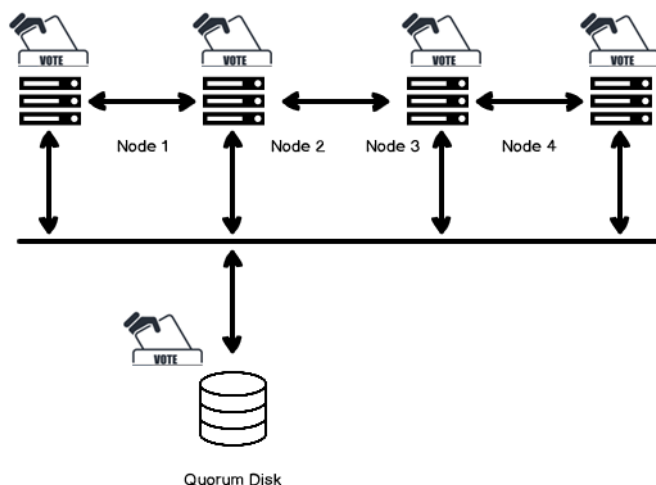
- Let's say Node 1 and Node 2 are down.
- We still have three votes to bring the cluster services online.
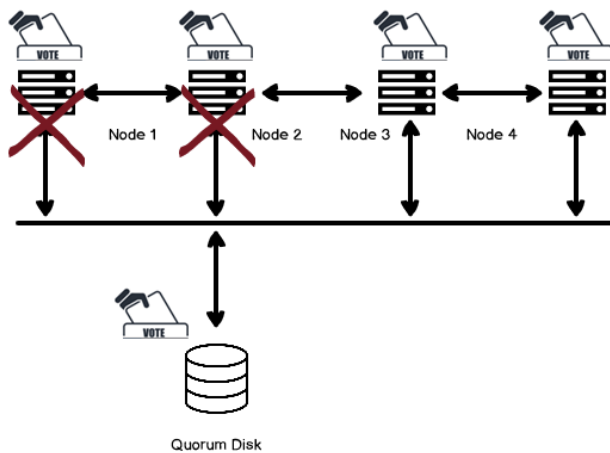
Quorum Disk

- If one more node also fails and we have only one node and one quorum disk to Vote, cluster services are offline in this case.

- Node and File Share Majority is similar to the previous Node and Disk Majority quorum modes.
- In this case, we replace the quorum disk with a file share witness (FSW).
- FSW also cast a single vote similar to a quorum disk.
- We might have windows cluster in different regions and might not have shared storage between them.
- File Share witness is a file share, and we can configure it on any server present in the active directory, and all cluster nodes should have access to that network share.
- It does not contain any configuration data.
- The active cluster node places a lock on the file share, and the file share witness contains the information about the owner.
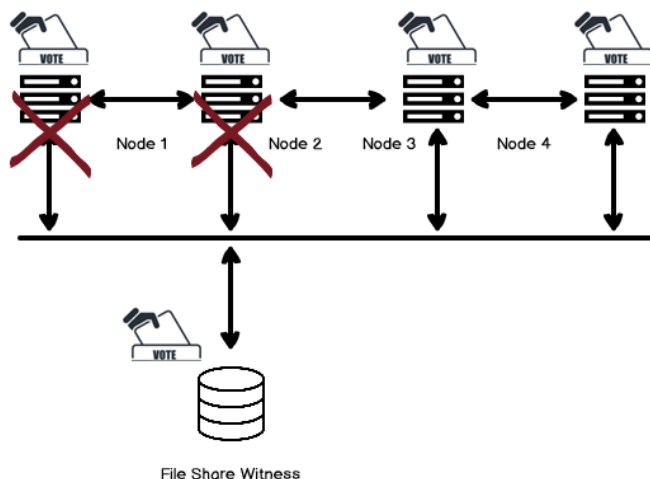

File Share Witness

- If the File share Witness is not available, the cluster remains online because of the majority of votes.
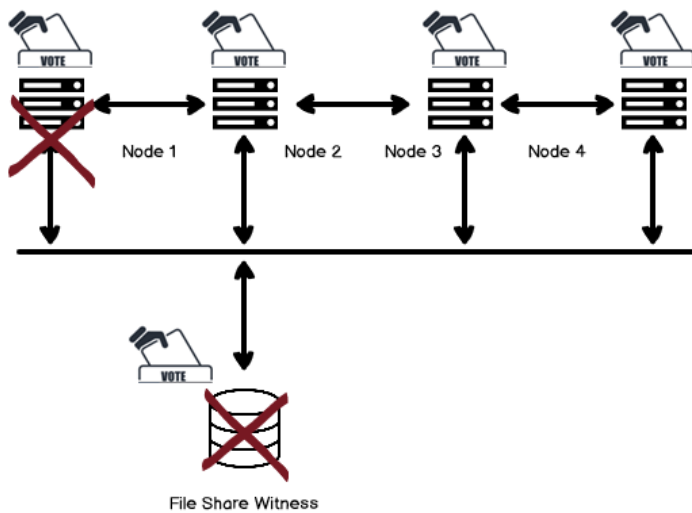
File Share Witness

- In this configuration, only witness disk has a vote, and cluster nodes cannot act as a voter.
- In this case, the witness disk acts as a single point of failure, and if it is not available or corrupted, the entire cluster is down.
- It is not a recommended quorum configuration.
- In the following image, you can see once the Witness disk is not available, the entire cluster is down.

Witness Disk
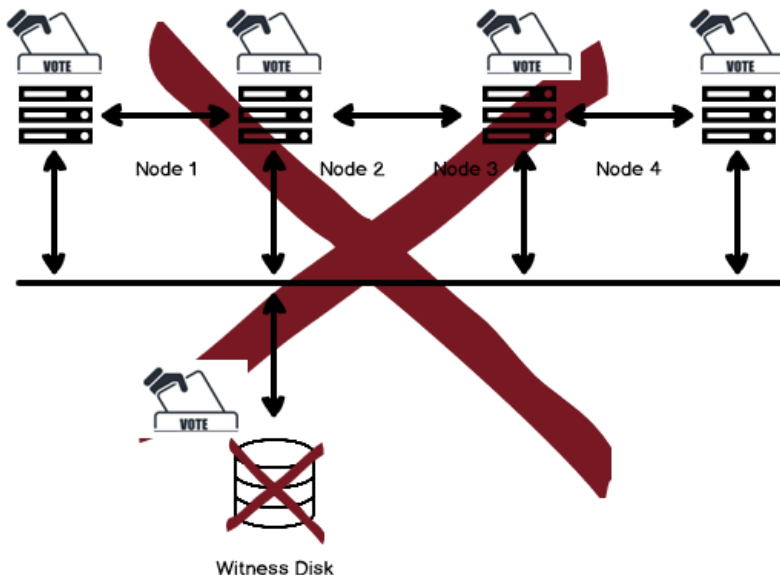
**Dynamic Quorum configuration mode in Windows Failover Cluster**

- Windows Server 2012 introduced a new feature Dynamic Quorum. It is enabled by default.
- In this configuration, the Windows Failover Cluster dynamically calculates the quorum as per the state of the active voters.
- If a node is down, the cluster group does not count its vote for the quorum calculation.
- Once the node rejoins the cluster, it is assigned a quorum vote.
- We can increase or decrease the quorum votes by using the dynamic quorum to keep the cluster running.
- Quorum votes are dynamically adjusted if the sum of votes is less than three in a multi-node cluster configuration.
- We should add a witness to the even number of nodes.

- If a particular node is down, you might leave with the even number of nodes that might be a problematic situation.

- Windows Server 2012 R2 also enables the Dynamic Witness once the dynamic quorum is configured.
- By using this feature, the Witness vote is dynamically adjusted.
- If we have an odd number of quorum votes in the cluster configuration, it does not use dynamic quorum witness. In the case of even number of votes, dynamic quorum also cast a vote.
- We can configure either File Share Witness or Disk witness as part of dynamic witness configuration.
- We can also use Cloud witness starting from Windows Server 2016.

It works on the following concepts.

- For the even number of nodes, each node has a vote and the witness dynamic vote
- For the odd number of nodes, each node has a vote, but the witness dynamic vote does not count. (Witness dynamic Vote= 0)

Let's understand dynamic quorum using an example. Suppose we have four nodes in a failover cluster.
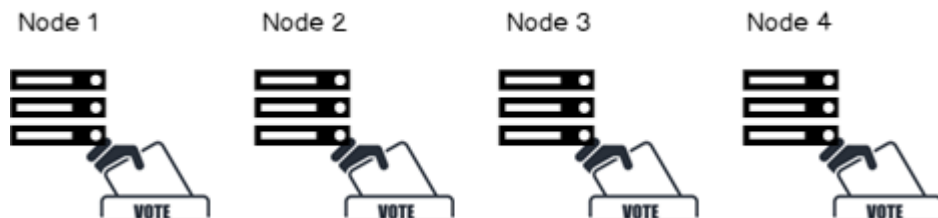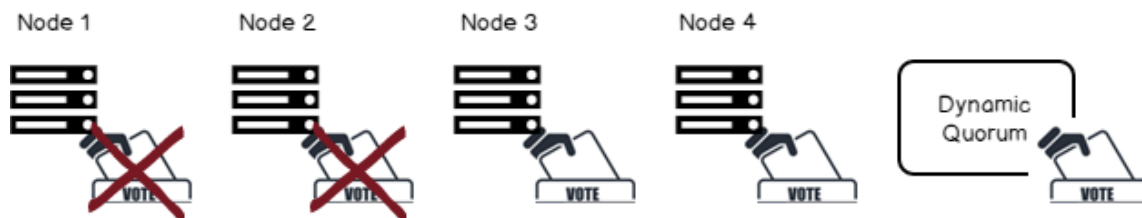

Each node can cast a Vote, and we require (4/2+1) three votes for the majority.

Let's say Node 1 is down. The cluster will be up as we still have three votes that are required for the majority.
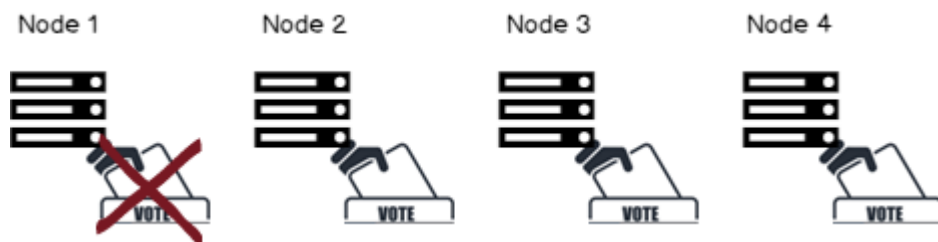

Now, Node 2 is down as well. In this case as well, the cluster will remain online because of the dynamic quorum casted one vote.

At this time, Node 3 is down as well. Using the dynamic quorum, we determine the majority based on the available nodes. Due to this, we still can have cluster service up.

The *total number of votes* required for a quorum is now determined based on the number of nodes available. Therefore, with a dynamic quorum, the cluster will stay up even if you lose three nodes. This situation is called as last-man-standing in which cluster works with a single node as well.
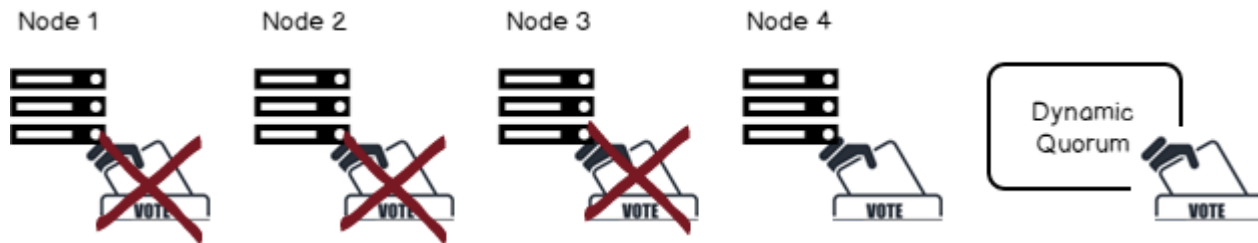
Conclusion

- In this article, we explored the concepts of Windows Failover Cluster Quorum modes that are necessary for a SQL Server Always on Availability Group configuration.
- You should consider the pros and cons of each approach as per your configuration. Y
- ou should select the quorum modes as per the requirements and should test failover to avoid issues in the production environment.

https://www.sqlshack.com/windows-failover-cluster-quorum-modes-in-sql-server-always-on-availability-groups/