SQLskills Immersion Event

IECAG: Clustering and Availability Groups

Module 1: Clustering Essentials

Jonathan Kehayias

Jonathan@SQLskills.com

Overview

- Planning for failover clustering
- Differences in Windows Server versions
- Planning storage requirements
- Active Directory disconnected clusters
- Quorum configuration
- Cluster Validation Tests
- How failover works
- Troubleshooting cluster issues

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

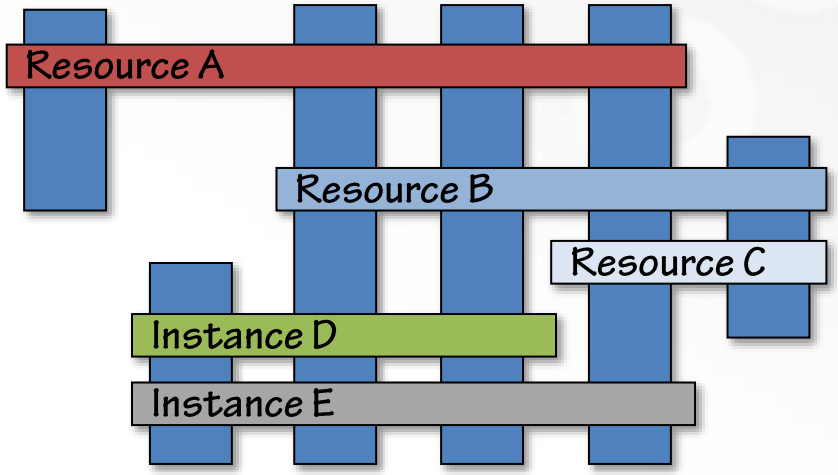
What is Failover Clustering?

- **High-availability solution built on top of Windows Clustering that provides automatic detection and failover at the instance level**
- **Only one node can own the instance or service and its dependent resources such as shared disks, IP address, and MSDTC (if configured as a cluster resource)**
- **Windows Server Standard Edition supports 2 nodes, Datacenter Edition supports the OS-Edition maximum for nodes**
  - Up to 16 nodes can participate (Windows Server 2008/2008R2)
  - Up to 64 nodes can participate (Windows Server 2012/2012R2/2016)

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

Multi-Node Clusters



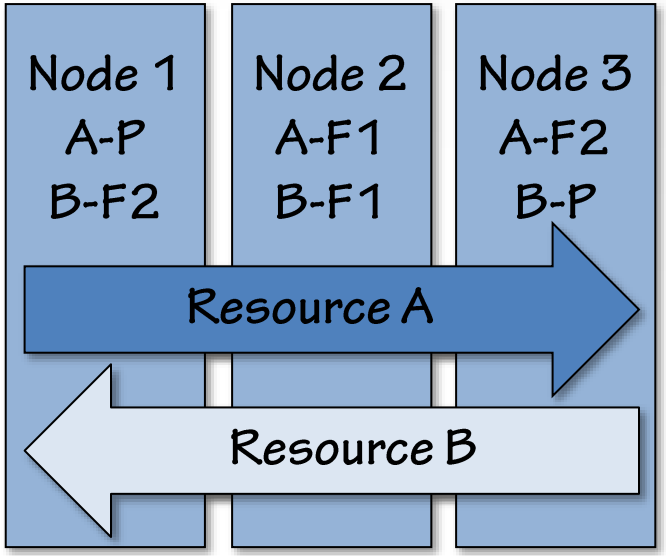
- **Multi-instance clusters have multiple instances of SQL Server installed**
  - Each instance requires its own cluster resource group
  - Each instance should be assigned a dedicated TCP port to avoid firewall and client connectivity issues from dynamic ports if the Browser service is disabled
  - Nodes must be able to handle the workload of all instances that can failover to the node
- **Multi-node clusters provide added flexibility to the configuration**
  - N+1: N active nodes with 1 passive standby server
  - N+K: N active nodes with K passive standby servers

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

Multi-Instance Preferred Ownership

- **Cluster resource groups should have preferred owners and failure paths configured for the instances to ensure that SLAs continue to be met in the event of a failure**
- **Example three-node cluster with two resources installed**
  - Resource Group A has Node 1 as its preferred owner and a defined failure path to Node 2 and then Node 3
  - Resource Group B has Node 3 as its preferred owner and a defined failure path to Node 2 and then Node 1



© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts
Understanding Quorum

- **Quorum determines the number of failures that a cluster can sustain and still remain online**
- **Quorum exists to handle scenarios where communication between cluster nodes has failed to prevent multiple nodes from trying to host the same resource group(s) simultaneously resulting in a 'split-brain'**
- **Voting towards quorum**
  - The cluster has quorum if more than half of the voters are online and communicating with each other
  - Each cluster node has 1 vote
  - A disk, file share, or cloud witness can be configured for 1 vote
  - Always configure quorum so that an odd number of votes exists in the cluster
    - 5-node cluster needs 3 nodes online to run, so does a 4-node cluster without a disk or file share witness

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

Quorum Types

- **Node Majority**
  - Default for odd number of nodes
- **Node and Disk Majority**
  - Recommended for even number of nodes
  - Requires an additional small clustered disk that can failover between the nodes as a witness and additional voter for quorum
- **Node and File Share Majority**
  - Required for multi-site or geo-clustering
  - Requires a file share in the same AD forest as the cluster nodes
- **Legacy: Disk Only**
  - Is a single point of failure for the cluster quorum and is not recommended for use
  - Maintained only for backwards-compatibility reasons
- **Cloud Witness (Windows Server 2016)**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

Voting Towards Quorum: Dynamic Quorum

- **Dynamic Quorum**
  - Provides the ability of the cluster to recalculate quorum on the fly to maintain a working cluster
  - Provides the ability to continue to run a cluster even if less than 50% of the nodes remain available
  - Allows the cluster to be reduced down to the last node (known as last man standing) and still maintain quorum
- **Dynamic Quorum requires:**
  - The cluster has already achieved quorum
  - Sequential failures of nodes occurs
- **If multiple nodes in a cluster go down simultaneously, dynamic quorum will not recalculate the number of votes to maintain quorum**
  - When this occurs, a regroup must occur to determine quorum can be maintained, and then dynamic quorum will resume when a subsequent node failure occurs

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

Voting Towards Quorum: Witness

- **Dynamic Witness (Windows Server 2012 R2+)**
  - Microsoft recommendation is to always configure a witness on Windows Server 2012 R2+
  - Provides the ability of the cluster to change the witness vote on the fly
    - Odd number of nodes online = no witness vote
    - Failed witness = no witness vote
  - Prevents a witness failure from causing the cluster to go down
  - A cluster with a witness requires at least 3 votes and majority for quorum
- **Cloud Witness (Windows Server 2016)**
  - Eliminates the need to manage a 3rd datacenter for a witness
  - Does not require a cloud based VM as a witness
  - Leverages Microsoft Azure Blob Storage node failure occurs
    - Same storage account can support multiple clusters
    - Low cost of maintenance since writes only occur when cluster node state changes

© SQLskills, All rights reserved.
https://www.SQLskills.com

Clustering Concepts

Site-aware Failover Clusters in Windows 2016+

- **Groups nodes in stretched clusters based on their physical location using cluster fault domains**
  - Supports up to four levels (Site, Rack, Chassis, and Node)
  - Nodes are automatic, other levels are optionally configured; i.e. you wouldn't use Chassis if you don't have blade servers
- **Cluster resource groups (Roles) failover between nodes in the same fault domain before failing to another fault domain**
- **Dynamic Quorum drops votes from the Disaster Recovery Site first during a regroup ensuring the Preferred Site survives**
- **Fault domains support optional Location and Description metadata attributes that are automatically included in alerting from the Cluster Health Service**
- **Provides storage affinity for shared storage and S2D maximizing performance and resiliency**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Server Hardware

- **Traditionally failover clustering required identical servers from the Hardware Compatibility List (HCL)**
- **Windows Server Cluster Validation Tool (CVT)**
  - Any hardware that passes the configuration tests run by the CVT can be used for failover clustering
  - Heterogeneous cluster nodes are supported as long as the hardware and OS architecture (x86/x64/IA64) matches
- **Separate public and private networks are no longer required, but still recommended as a best practice**
  - The cluster heartbeat can now communicate across any network available to the cluster

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Storage

- **The shared disk arrays must be presented to each of the cluster nodes and pass the CVT Disk tests before creating failover cluster**
- **Determines the disks available for clustering, validates arbitration, failover ability, and multipath configuration**
- **SCSI-3 Persistent Reservation Requirement**
  - Avoid SCSI bus resets, much less disruptive than the older reserve/release commands
- **All tests must pass in order for the cluster configuration to be supported for SQL Server failover clustering**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Storage Spaces Direct (S2D)

- **Storage Spaces Direct uses local-attached drives to create a converged or hyper-converged environment that supports failover clustering**
  - Converged = SMB3
  - Hyper-converged = CSVs
- **Hyper-converged configurations require 4GB RAM per TB of drive capacity for cache and S2D internal usage**
- **Requires Windows Server 2016+ Datacenter and a minimum of 2 nodes with a maximum of 16 nodes**
  - All servers must have the same drive types
  - 2 drive minimum for cache per server if using caching
  - 4 drive minimum for capacity per server (2 for virtual machines)

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Storage: Mount Points or Drive Letters

- **Mount points provide a way to exceed the 26-volume limit imposed by using drive letters**
  - Additional volumes can be mounted as folders under a drive letter
  - May be problematic for space monitoring
- **Each instance requires at least one disk that has a drive letter assigned to it unless using file shares or CSVs**
  - Analysis Services cannot be clustered without a physical drive letter

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Storage: Virtual Machines

- **Failover clustering is supported using VM guests with shared storage**
  - Cluster must pass the Cluster Validation Tests to be supported
  - Support for in-guest iSCSI initiator with MPIO
  - VMware RDM shared disks
    - Limited to a single path in some versions, which imposes performance limitations that have to be considered
    - VM admins generally hate managing RDMs
  - Hyper-V Server allows a shared VHDx file to serve as a cluster shared disk between VMs simplifying the configuration
- **Reference:**
  - https://sqlskills.com/help/VMClusterReqs

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Active Directory/DNS

- **Prior to Windows Server 2016, Active Directory is required to configure failover clustering**
  - Windows Server 2016 allows single-domain, multi-domain, and workgroup clusters
- **For best security a separate domain service account should be used for each service with the minimum permissions required**
- **DNS name resolution must be working properly using fully-qualified domain names (FQDN)**
- **If using multiple NICs, ensure that the domain-name suffix is configured appropriately**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Workgroup and Multi-domain Clusters

- **All servers must be running Windows Server 2016**
- **A local account must be created on all nodes of the cluster**
  - Username and Password must be the same on all nodes
  - Must be a member of the Local Administrators group
  - The LocalAccountTokenFilterPolicy registry policy must be set to 1
    - HKLM:\SOFTWARE\Microsoft\Windows\CurrentVersion\Policies\System
- **The cluster must be an Active Directory Detached cluster with a Cluster Network Name of type DNS**
  - DNS must be working properly and all nodes must have a primary DNS suffix
  - Multi-Domain cluster nodes should have the DNS suffix of all domains in the cluster on all nodes

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering

Networking

- **Separate private network is not required for heartbeat but is still recommended**
- **Redundant NICs connected to separate switches should be used to remove single point of failure**
  - For 2 node clusters a x-over cable can be used for the private network
- **NIC Teaming can be used to overcome network bottleneck if necessary**
  - Allows multiple network cards to share a common IP address and balance load across their connections
- **Configure all adapters for full-duplex operation**
- **Public network required for client access to clustered service**
  - Important for Availability Groups for private network for log traffic

© SQLskills, All rights reserved.
https://www.SQLskills.com

Planning for Failover Clustering
MSDTC

- **MSDTC is not required in failover clustering**
- **If you need MSDTC for distributed transactions, it must be configured as a cluster resource using the Cluster Administrator**
- **MSDTC can be installed as a resource in an existing cluster resource group**
  - Allows multi-instances of DTCs to be associated in the cluster to specific resource groups
  - Ensures that the DTC resource runs on the same cluster node as the service that depends on it and that they failover together
- **Blog by Cindy Gross:**
  - https://sqlskills.com/help/msdtc

© SQLskills, All rights reserved.
https://www.SQLskills.com

How Failover Works

Failure Detection

- **Node failure is detected by the cluster service when a node misses 6 consecutive heartbeats**
- **During failover, clustered resource groups and all associated resources are moved from one node to another**
- **Failover occurs when:**
  - The node that currently hosts the resource becomes unavailable for any reason
  - One of the resources in the group fails based on failover policy conditions
  - A windows administrator manually initiates a failover/failure

© SQLskills, All rights reserved.
https://www.SQLskills.com

How Failover Works

Resource Failover

- When a manual failover occurs, the resource group is taken offline on the current owning node
- The Failover Manager performs arbitration to locate a new owner for the group and brings the resources online on the new node in dependency order

© SQLskills, All rights reserved.
https://www.SQLskills.com

How Failover Works

Failback

- Cluster resource groups can be assigned a 'preferred owner'
- When the preferred owner node comes back online, the cluster resource group can automatically failback to the node based on its configuration
- Failback is not required and should only be configured in multi-node/multi-instance clusters where specific instances should ideally be running on specific nodes
- Cluster resource groups should be configured to perform automatic failbacks outside of normal operating hours only
- Failback is the same as a failover and all client application connections are broken impacting end users

© SQLskills, All rights reserved.
https://www.SQLskills.com

Troubleshooting Cluster Issues

- **Windows event logs**
- **Cluster log**
  - cluster.exe /Cluster:ClusterName log /gen /copy “c:\”
- **SQL Server error log**
- **SQL Server setup log**
- **Connection failures**
  - Ensure that each instance is configured to use a fixed TCP/IP port in the cluster
  - Ensure that the appropriate firewall rules have been implemented in each of the cluster nodes for the fixed ports
  - Ensure there are no IP conflicts in the network

© SQLskills, All rights reserved.
https://www.SQLskills.com

Troubleshooting Clusters Issues

Recovering from Cluster Hardware Failure

- **Evict the failed node from the cluster using the Cluster Administrator**
- **Verify that the node has been removed from the cluster**
- **Repair or replace the failed server hardware and reinstall Windows Server**
- **Add the node to the cluster using the Cluster Administrator**
  - Plan a downtime to re-run Cluster Validation Tests
- **Install/configure cluster services required**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Forcing Quorum
The Last Ditch Effort

- **Overrides the safety controls that took the cluster offline**
  - Loss of multiple cluster nodes causing quorum failure
    - Hardware failures
    - Networking/software failures
- **Start the WSFC on a single node that can communicate with a majority of remaining nodes using forced quorum**
  - FCM – Actions > Force Cluster Start (Windows 2012 R2+)
  - Powershell - Start-ClusterNode –FixQuorum from elevated window
  - Command Prompt - net.exe start clussvc /forcequorum
- **Start the cluster service on remaining nodes one at a time and wait for them to join into the cluster**
- **Adjust cluster quorum configuration and voter node weights as required for the new configuration**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Windows 2012 R2+ Benefits

- **Dynamic Quorum**
- **Dynamic Witness**
- **Force quorum resiliency**
- **Tie breaker for 50% node split**
  - LowerQuorumPriorityNodeID can be set to predetermine what nodes lose votes under Dynamic Quorum
- **Cluster operating system rolling upgrades from Windows Server 2012 R2 to Windows Server 2016 without downtime**

© SQLskills, All rights reserved.
https://www.SQLskills.com

Key Takeaways

- Use Windows Server 2016+ for new installations, especially in multi-site geo-clusters, for dynamic quorum to allow a last-man-standing configuration
- Plan for and configure quorum settings to provide a odd number of votes and to allow a majority of voters available in the event of a failure
- During rolling updates to SQL Server, while node(s) are being patched remove them from possible ownership to prevent a failover attempting to start the service during patching

© SQLskills, All rights reserved.
https://www.SQLskills.com

Review

- Planning for failover clustering
- Differences in Windows Server versions
- Planning storage requirements
- Active Directory disconnected clusters
- Quorum configuration
- Cluster Validation Tests
- How failover works
- Troubleshooting cluster issues

© SQLskills, All rights reserved.
https://www.SQLskills.com

Questions?

Clustering Concepts

Voting Towards Quorum: Windows 2008 R2 and earlier

- **The cluster has quorum if more than half of the voters are online and communicating with each other**
- **Each cluster node has 1 vote**
  - Based on its NodeWeight configuration in Windows Server 2008 onwards
  - http://support.microsoft.com/kb/2494036
- **A disk or file share witness can be configured for 1 vote**
- **Always configure quorum so that an odd number of votes exists in the cluster**
  - 5-node cluster needs 3 nodes online to run, so does a 4-node cluster without a disk or file share witness
  - Plan for servers participating in the cluster that are in a remote data center when working with geo-clusters

© SQLskills, All rights reserved.
https://www.SQLskills.com