Failover Clustering:
What's new in Windows Server 2012 R2

Didier Van Hoye
Technical Architect &
Technology Strategist

MVP Microsoft® Most Valuable Professional

meet Microsoft Extended Experts Team

Dell TechCenter
ROCKSTAR ★

W http://workinghardinit.wordpress.com

@workinghardinit

Microsoft

Technical Summit 2014
Pure Technology.

Die Microsoft Konferenz
für Entwickler und IT Pros vom
11.–13. November in Berlin

Objectives

➢ Overview of new features & enhancements in Windows Server 2012 R2 Failover Clustering

➢ Motivate you to investigate Windows 2012 R2 & vNext

Microsoft

Guest Clustering Options

Guest Clustering shared storage deployment options

	Windows Server 2008 (R2)	Windows Server 2012	Windows Server 2012 R2
Fibre Channel		✓	✓
iSCSI	✓	✓	✓
File (SMB)		✓	✓
Shared VHDX			✓

Microsoft

Guest Clustering with Shared VHDX

Guest Clustering

- Guest Clustering with commodity storage
- Sharing VHDX files provides shared storage for Hyper-V guest Failover Clustering
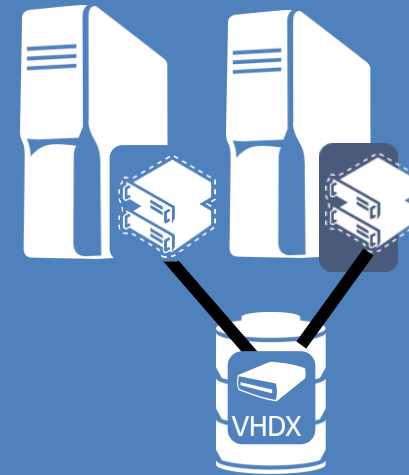- Maintains separation between infrastructure and tenants: security, operational, organizational issues

Virtual SAS

- VM presented a shared virtual SAS disk
- Utilizes SCSI Persistent Reservations
- Used for data disk only
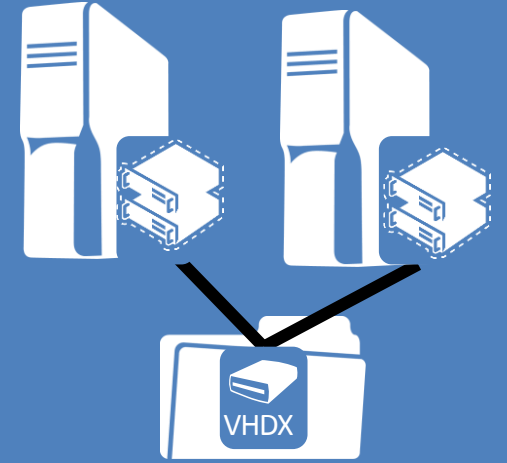- Supports both fixed & dynamically expanding VHDX

Cluster Shared Volumes (CSV) on block storage

VHDX

Block Storage

Scale-Out File Server for file based storage

VHDX

File Based Storage

➢ Hosts have to be W2K12R2
➢ Guests W2K12(R2)* with IC installed. Nothing architecturally prevents the use of older versions, but they do not have test coverage which makes it unsupported

Microsoft

Shared VHDX Limitations

1. You cannot do host-level backups of the guest cluster

   ➤ You do in guest backups as you do with physical clusters

2. You cannot on line resize the shared VHDX

   ➤ But you can hot add or remove them.

3. You cannot Storage Live Migrate the shared VHDX file

   ➤ But you can move the other VM files and perform normal Live Migration

Microsoft

Creating Shared VHDX Example

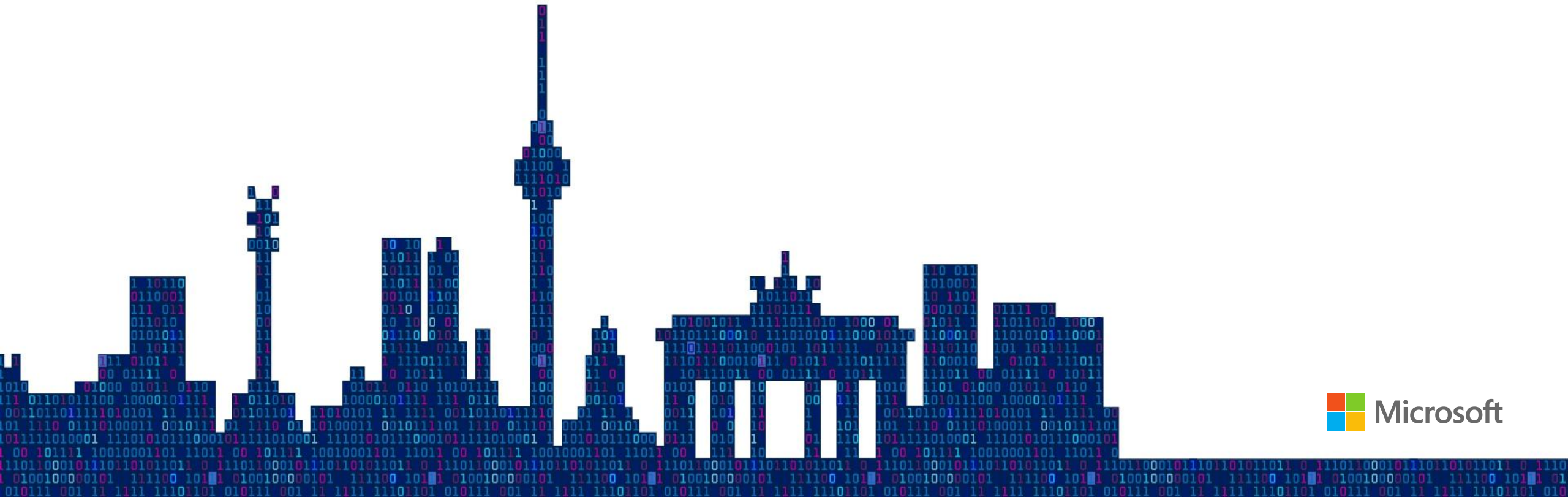Example of creating and attaching a shared VHDX to two existing VMs

```
PS C:\> New-VHD -Path C:\ClusterStorage\Volume1\Shared.VHDX -Fixed -SizeBytes 30GB

PS C:\> Add-VMHardDiskDrive -VMName Node1 -Path C:\ClusterStorage\Volume1\Shared.VHDX -SupportPersistentReservations

PS C:\> Add-VMHardDiskDrive -VMName Node2 -Path C:\ClusterStorage\Volume1\Shared.VHDX -ShareVirtualDisk
```

Shared VHDX
DEMO

Microsoft
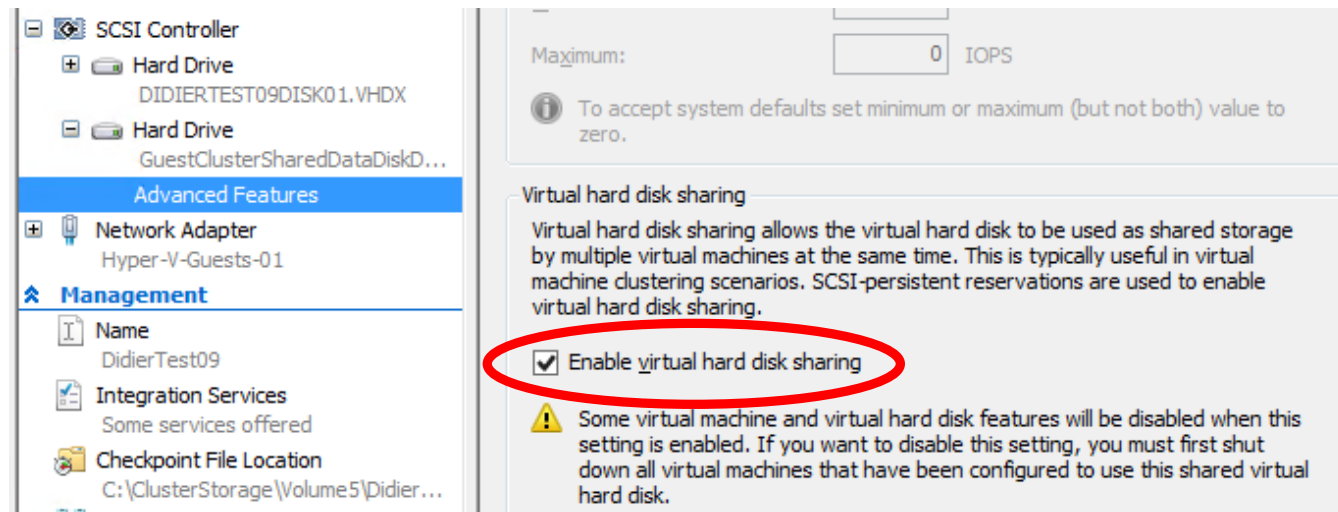
VM Drain on Shutdown

- VMs live migrated to another node during shutdown
- Protects from downtime for VMs due to unknowingly or mistakenly rebooting or shutting down the wrong host
- VMs moved to "Best Available Node" (most free memory)
- Honors VM prioritization
- Enabled/Disabled via **DrainOnShutdown** cluster common property

Sleep

Shut down

Restart

Power
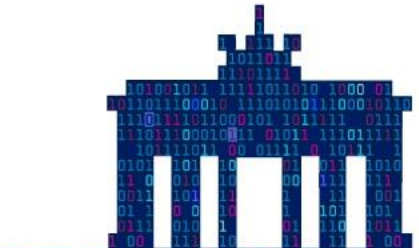
Recommended

Still recommended to drain nodes before shutdown
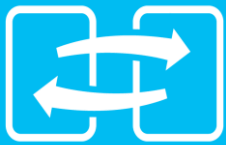
Microsoft

Network Failure Detection for VMs

Health Monitoring

Component level VM network health detection

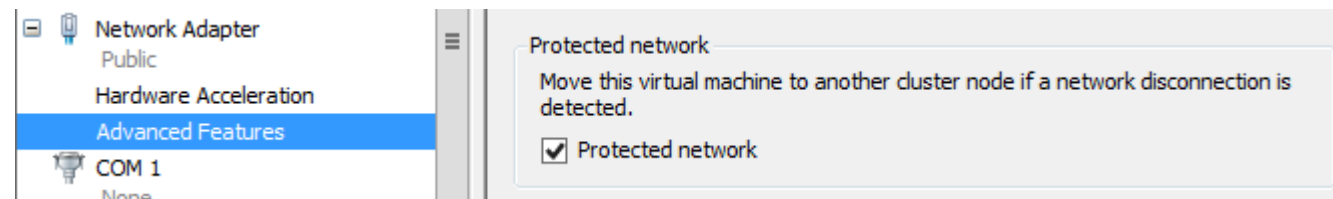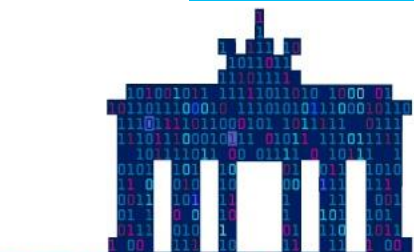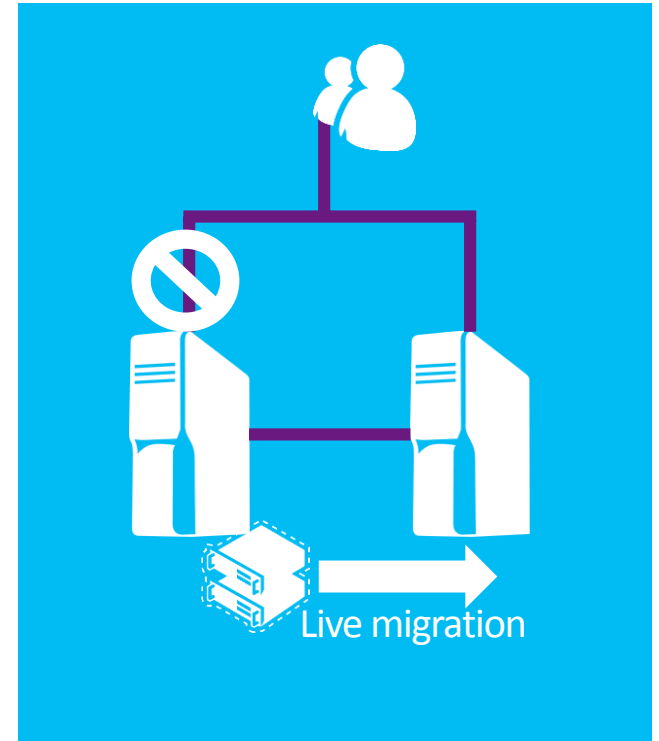Media sense loss detection of network disconnects

Recovery

Live migrates VMs off hosts with no external network connectivity for the VM

Verifies destination node has the network available

Configurable per network

Live migration

Network Adapter
Public
Hardware Acceleration
Advanced Features
COM 1
None

Protected network
Move this virtual machine to another cluster node if a network disconnection is detected.
☑ Protected network

Blog

Microsoft

Storage Failure Detection for VMs

Health Monitoring

Windows Server 2012 R2 Failover Clustering now detects physical storage failures on storage devices that are not managed by the failover cluster (for example, SMB 3 shares).
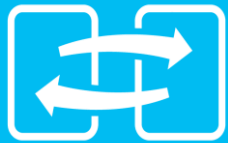
Recovery

If such an event occurs, Failover Clustering ensures that the virtual machine is relocated and restarted on another node in the cluster. This eliminates situations where unmanaged storage failures would not be detected and where virtual machines resources may become unavailable.

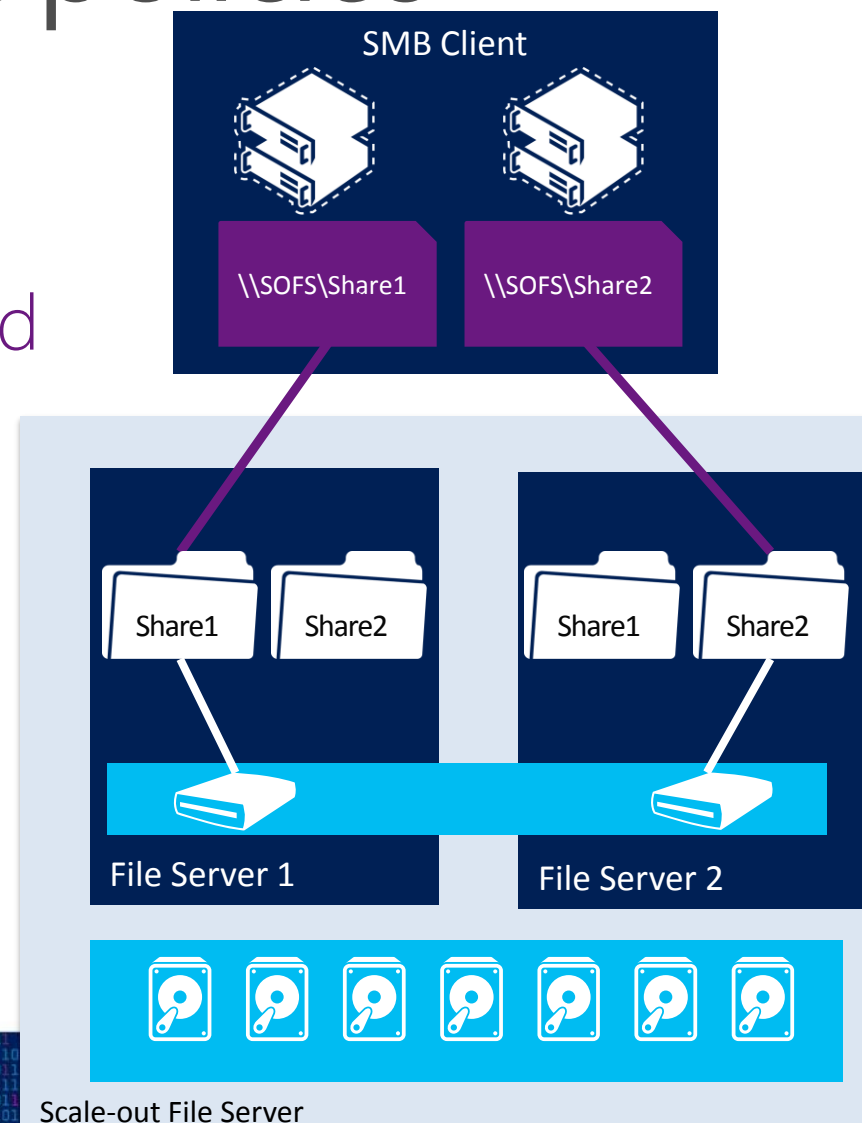VHDX on SOFS Share

Move & restart

Blog

Microsoft
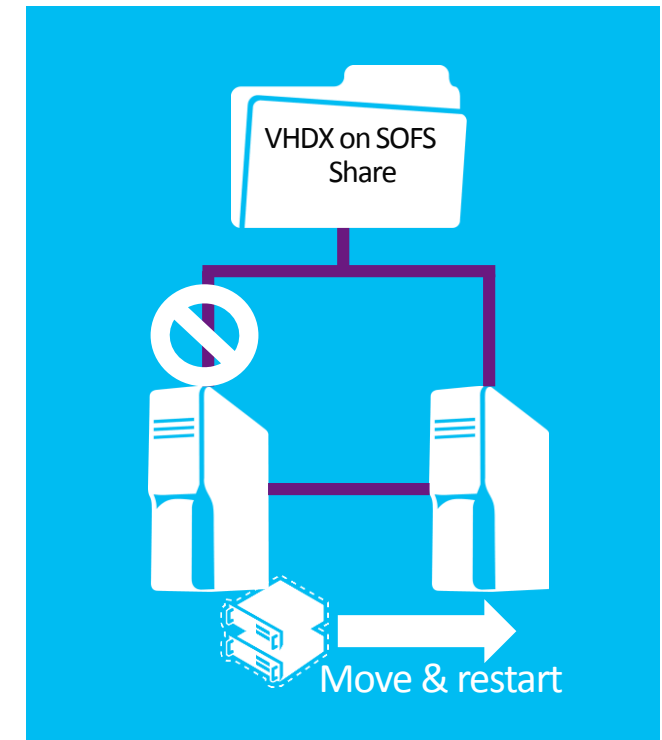
Optimized CSV placement policies

Optimized failover placement policies to spread CSV ownership across the cluster
Ensures CSV volume ownership is distributed across the cluster

Placement Scenarios:
- CSV Failover
- Node rejoining the cluster
- New node added to the cluster
- Cold start of cluster

Technet

SMB Client

\\SOFS\Share1   \\SOFS\Share2

Share1   Share2   Share1   Share2

File Server 1   File Server 2
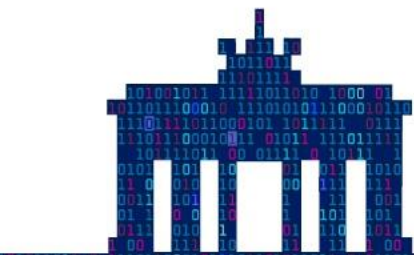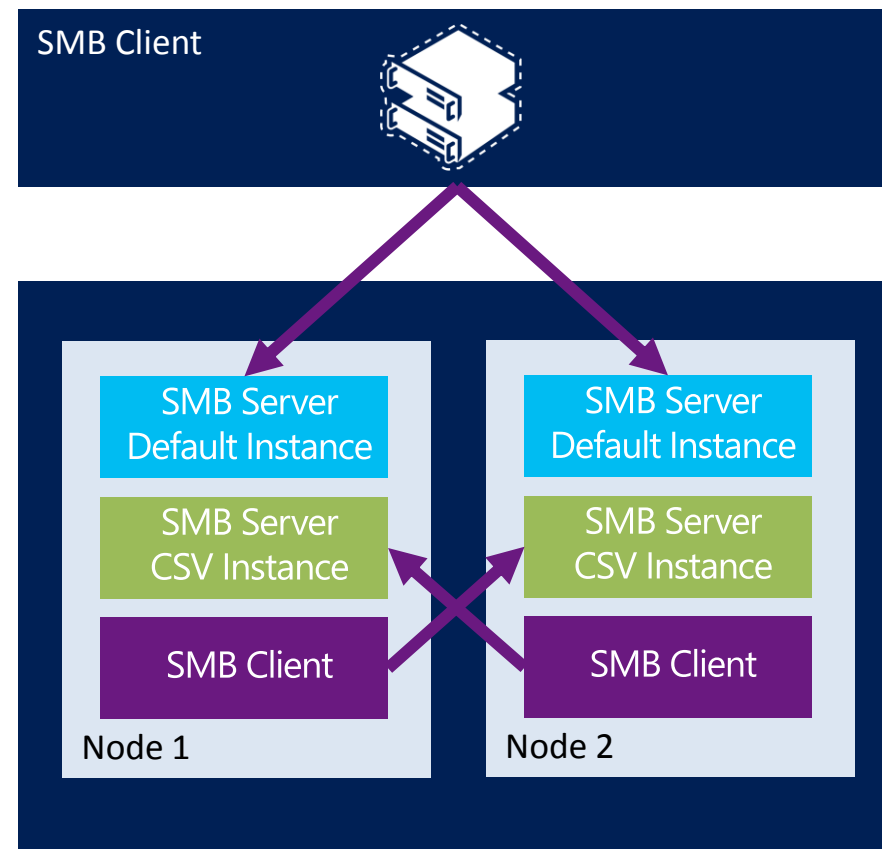
Scale-out File Server

Increased CSV Resiliency

Isolated SMB Instances

- Separation for improved resiliency
- Independent processes and queues for regular SMB client traffic and inter-node SMB traffic
- Improves scale of inter-node SMB traffic between CSV nodes

Multiple SMB Server instances per Node

- Default Instance – Handles incoming traffic from SMB clients accessing regular file shares
- CSV Instance – Handles only inter-node CSV traffic (metadata access or redirected traffic)


SMB Client

SMB Server Default Instance

SMB Server CSV Instance

SMB Client

Node 1

SMB Server Default Instance

SMB Server CSV Instance

SMB Client

Node 2

CSV Dependency Health Detection

Cluster Shared Volumes (CSV) health monitoring of the Server Service

➤ CSV leverages SMB as a transport protocol for orchestrating and forwarding I/O between nodes

If the Server Service becomes unhealthy on a node, CSV will designate a new coordinator node
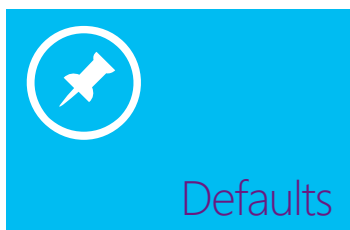

Cluster Shared Volumes

Server Service

Microsoft

Technet

CSV Cache

IMPROVED
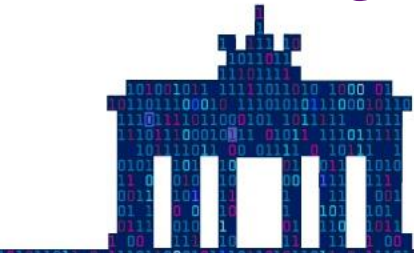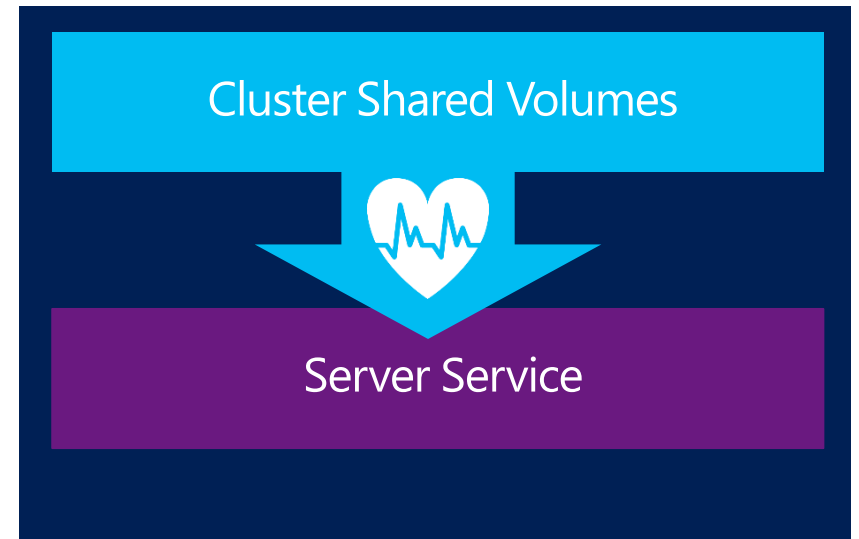
Scalability	CSV cache supports allocating larger amounts of memory (up to 80% of total memory) Most interesting for Scale-out File Servers
Defaults	Physical Disk resource EnableBlockCache enabled by default Still need to enable via BlockCacheSize to allocate memory
Recommended	Recommended to enable CSV Cache for all Hyper-V and Scale-out File Server deployments Greater allocation for Scale-out File Server as they are not memory bound

Blog

Microsoft

Improved Diagnosability

Deployment Logging

Improved logging when creating a cluster and adding a node

Enables pinpointing root cause quickly and easily

State Logging

Increased logging to Operational channel for cluster resource state changes

For example: When cluster resources move from one node to another
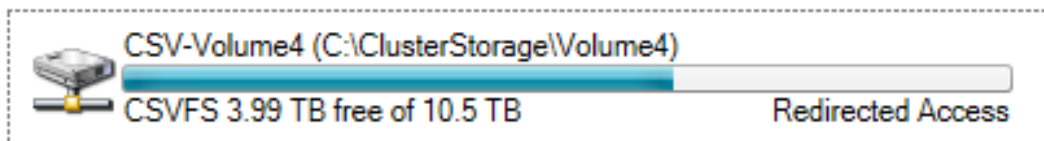
Microsoft

Improved CSV Diagnosability

Displays on a per-node basis the CSV I/O mode and the reason if in a redirected mode
Direct I/O, Block level redirected mode, File system level redirected mode

Get-ClusterSharedVolumeState PowerShell cmdlet

CSV State

Volumes (1) ────────────────────

CSV-Volume4 (C:\ClusterStorage\Volume4)

CSVFS 3.99 TB free of 10.5 TB

Volumes (1) ────────────────────

CSV-Volume4 (C:\ClusterStorage\Volume4)

CSVFS 3.99 TB free of 10.5 TB   Redirected Access

```
PS C:\Windows\system32> get-clustersharedvolumestate

Name                       : Cluster Disk 1
VolumeName                 : \\?\Volume{380133d9-8118-44a8-b721-e36c065fbf94}\
Node                       : NODE-A
StateInfo                  : BlockRedirected
VolumeFriendlyName         : Volume1
FileSystemRedirectedIOReason : NotFileSystemRedirected
BlockRedirectedIOReason    : NoDiskConnectivity

Name                       : Cluster Disk 1
VolumeName                 : \\?\Volume{380133d9-8118-44a8-b721-e36c065fbf94}\
Node                       : NODE-B
StateInfo                  : Direct
VolumeFriendlyName         : Volume1
FileSystemRedirectedIOReason : NotFileSystemRedirected
BlockRedirectedIOReason    : NotBlockRedirected

Name                       : Cluster Disk 2
VolumeName                 : \\?\Volume{ee4380f4-f88e-4777-9bed-489b01adde90}\
Node                       : NODE-A
StateInfo                  : BlockRedirected
VolumeFriendlyName         : Volume2
FileSystemRedirectedIOReason : NotFileSystemRedirected
BlockRedirectedIOReason    : NoDiskConnectivity

Name                       : Cluster Disk 2
VolumeName                 : \\?\Volume{ee4380f4-f88e-4777-9bed-489b01adde90}\
Node                       : NODE-B
StateInfo                  : Direct
VolumeFriendlyName         : Volume2
FileSystemRedirectedIOReason : NotFileSystemRedirected
BlockRedirectedIOReason    : NotBlockRedirected
```
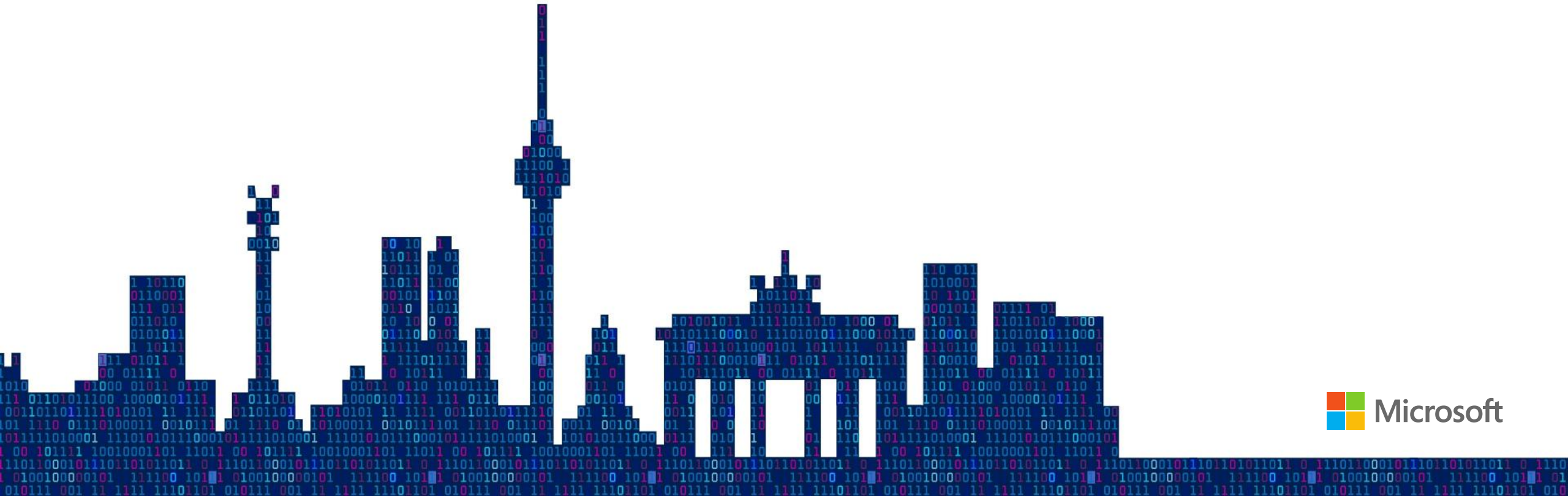
DEMOS
CSV Reliability, Improved CSV Cache, CSV Diagnosability, VM Priority / Drain on shutdown, CSV Placement

Microsoft

Extended CSV Interoperability

IMPROVED

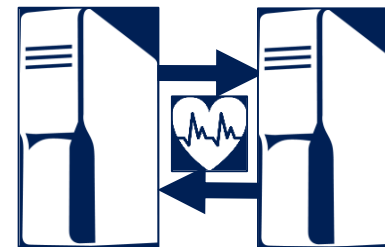| Interoperability | ReFS with CSV | Tiered Spaces with CSV | Spaces Write-Back Caching with CSV | Parity Spaces with CSV | Dedup with CSV |

Technet

Microsoft

Node Health Detection over Network

Cluster Health monitoring:
- Nodes exchange heartbeats every 1 second (configurable)
- Nodes considered down if it does not respond to 5 heartbeats (configurable)

Default settings are fairly aggressive to deliver the highest levels of availability
- However are fully configurable to meet your business needs

For a Hyper-V deployment slightly more relaxed settings may make sense
- Traditionally the definition of down, is when clients cannot connect to an app in the VM
- In general, TCP defines recoverable network errors for applications
  - **Recommended for cluster heartbeats not to exceed 20 seconds**

Greater resiliency to transient network failures with Windows Server 2012 R2
- Heartbeat thresholds increased by default for Hyper-V Clusters
- Cluster heartbeating improved for increased resiliency to packet loss

Cluster Property	Default	Hyper-V Default
SameSubnetThreshold	5	10
CrossSubnetThreshold	5	20

Blog

Microsoft

Cluster Global Updates

For state change or database update nodes must be notified and updated

- Leverages clusters Global Update Manager (GUM)

Two GUM modes now in Windows Server 2012 R2

Majority

Greater than 50% of nodes must acknowledge before considered committed

Toggled on by default for Hyper-V clusters

Enables the cluster to move forward if there is a node with latency

All

All nodes must acknowledge before considered committed

Default mode for all other workloads (SQL, Exchange, etc...)

Configurable with DatabaseReadWriteMode cluster common property

Cluster

Update

Network Validation Improvements

- Network validation improved from simple PING to using NetFT

- Verifies port 3343 and full cluster network connectivity requirements

- Provides better diagnosability & pre-identifies cluster configuration problems

Network

Name
List Network Binding Order
Validate Cluster Network Configuration
Validate IP Configuration
Validate Network Communication
Validate Windows Firewall Configuration
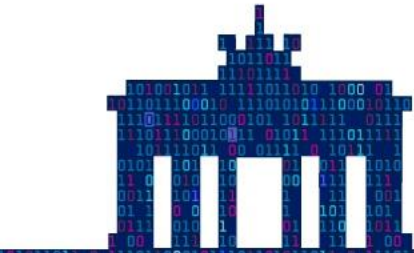
Cluster Dashboard

- New Cluster Dashboard for Status at a Glance

- Focused at multi-cluster management

Failover Cluster Manager

Create failover clusters, validate hardware for potential failover clusters, and perform configuration changes to your failover clusters.

^ Overview

A failover cluster is a set of independent computers that work together to increase the availability of server roles. The clustered servers (called nodes) are connected by physical cables and by software. If one of the nodes fails, another node begins to provide services. This process is known as failover.
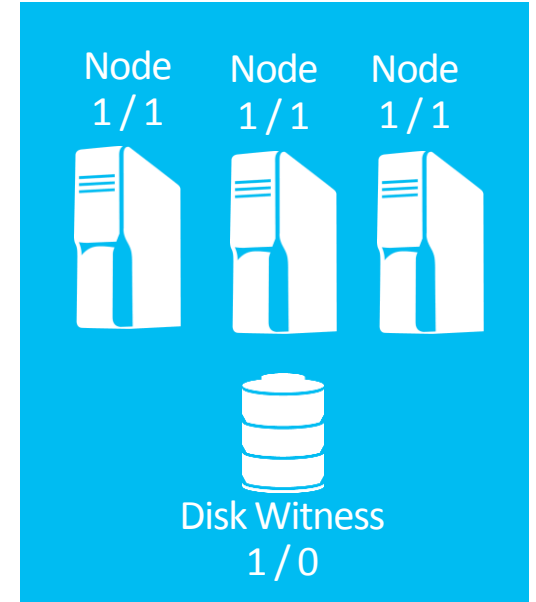
^ Clusters

Name	Role Status	Node Status	Event Status
eldenc-cl58.redmond.corp.microsoft.com	20 total	⊗ 1 down, 2 total	⚠ Error: 4
SQLCSV0.redmond.corp.microsoft.com	0 total	2 total	None in the last 24 hours
SQLCSV1.redmond.corp.microsoft.com	1 total	2 total	None in the last 24 hours
ekhostclu429.redmond.corp.microsoft.com	1 total	2 total	None in the last 24 hours

Microsoft

Dynamic Witness

Witness vote dynamically/automatically adjusted

based on cluster membership with dynamic quorum

- Odd node votes (3) + no witness vote (0) = 3
- Even node votes (2) + witness vote (1) = 3

Automatic functionality based on WitnessDynamicWeight property.

Node
1 / 1

Node
1 / 1

Node
1 / 1

Disk Witness
1 / 0

New
Recommendation

Always configure a witness with Windows Server 2012 R2

Clustering will determine when it is best to use the Witness

Configure Disk Witness if shared storage, otherwise FSW

Microsoft

Intuitive Quorum Configuration UI

- Node vote weights and dynamic quorum status easy & quick to view

- Removed legacy concept of 'quorum modes'

  ~~Node Majority~~

  ~~Node and Disk Majority~~

  ~~Node and File Share Witness Majority~~

http://blogs.msdn.com/b/microsoft_press/archive/2014/04/28/from-the-mvps-understanding-the-windows-server-failover-cluster-quorum-in-windows-server-2012-r2.aspx

Tie Breaker

Cluster will survive simultaneous 50% loss of votes
- Balanced multi-site clusters with complete site partition

One site automatically elected to win
- Site without LowerQuorumPriorityNodeID cluster common property wins
- Nodes in the other site drop out of the cluster

```
1  #Get the IDs from all the nodes in the cluster
2  Get-ClusterNode  | ft -AutoSize Name, ID
3
4  #Set the LowerQuorumPriorityNodeID for all ID of the nodes
5  #in the site you want to sacrifice itself
6  (Get-Cluster).LowerQuorumPriorityNodeID = 5
7  (Get-Cluster).LowerQuorumPriorityNodeID = 6
8  (Get-Cluster).LowerQuorumPriorityNodeID = 7
9  (Get-Cluster).LowerQuorumPriorityNodeID = 8
```
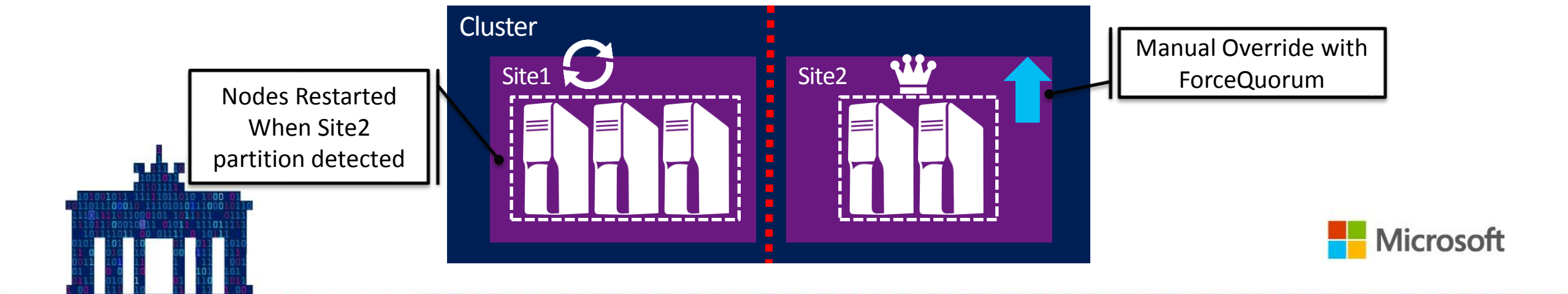
Cluster

Site1

Site2
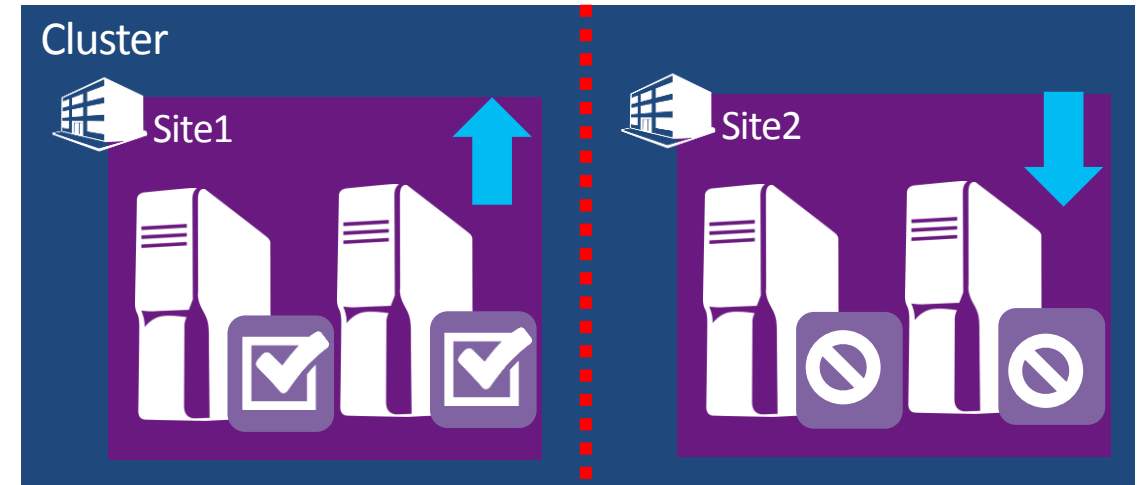
Force Quorum Resiliency

Cluster detects partitions after a manual ForceQuorum

ForceQuorum'd partition is deemed authoritative

Partitioned nodes restarted and re-joined

- Cluster brought back into a single view of membership

Cluster

Site1

Site2

Nodes Restarted When Site2 partition detected

Manual Override with ForceQuorum

Dynamic Witness, Quorum, Tie Breaker A hands on look

Microsoft

Reducing Cluster Dependencies

Active Directory-detached cluster Network Names
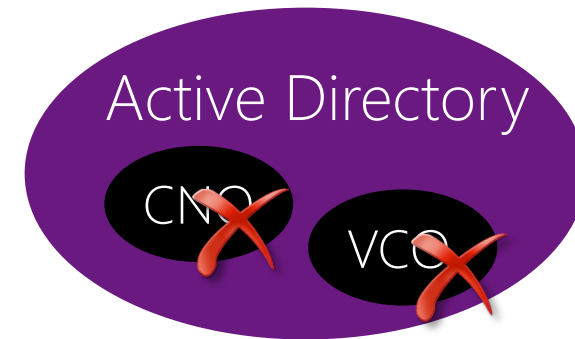- Enables creating a cluster without computer objects
  New-Cluster –AdministrativeAccessPoint DNS

Simplifies cluster deployments
- Best fit for SQL Server Clusters

Flexibility to create clusters with
or without Active Directory integration
- Still required that Nodes are domain joined

Active Directory

CNO

VCO

Cluster

Technet

Microsoft

AD-detached Clusters Support

What to consider when choosing a model:

- No computer objects, means no Kerberos authentication to the name
- NTLM only client authentication against cluster names
  - **Intra-cluster authentication still leverages Kerberos**

Role	Position	Notes
MSMQ Clusters	Not Supported	MSMQ stores properties in AD
File Server Clusters	Not Recommended	Kerberos for SMB preferred
Hyper-V Clusters	Not Recommended	No live migration support
SQL Server Clusters	Supported	Best fit if using SQL Authentication

Microsoft

Sneak Peak at vNext 1/2

Rolling cluster upgrades are here

- From W2K12R2 to vNext (W2K15?)
- Cluster Functional Level 8 ➔ 9
- Virtual Machine version 5 ➔ 6

Storage Replication

- Synchronously & Asynchronously
- Server to server, cluster to cluster for now.

Microsoft

Sneak Peak at vNext 2/2

Node isolation & quarantine

Dealing elegantly with transient issues

Microsoft

SQL Server 2014 Supported on CSV

Consolidation

Simplified storage management with consolidation of multiple instances onto a single LUN
Better capacity planning and storage utilization
Addresses drive letter limitations allowing more than 24 SQL instances on a single cluster

Availability

Databases can be moved without requiring any drive ownership changes
Faster Failover times with no dismounting and remounting of volumes required
Increased resiliency to storage faults with CSV I/O redirection
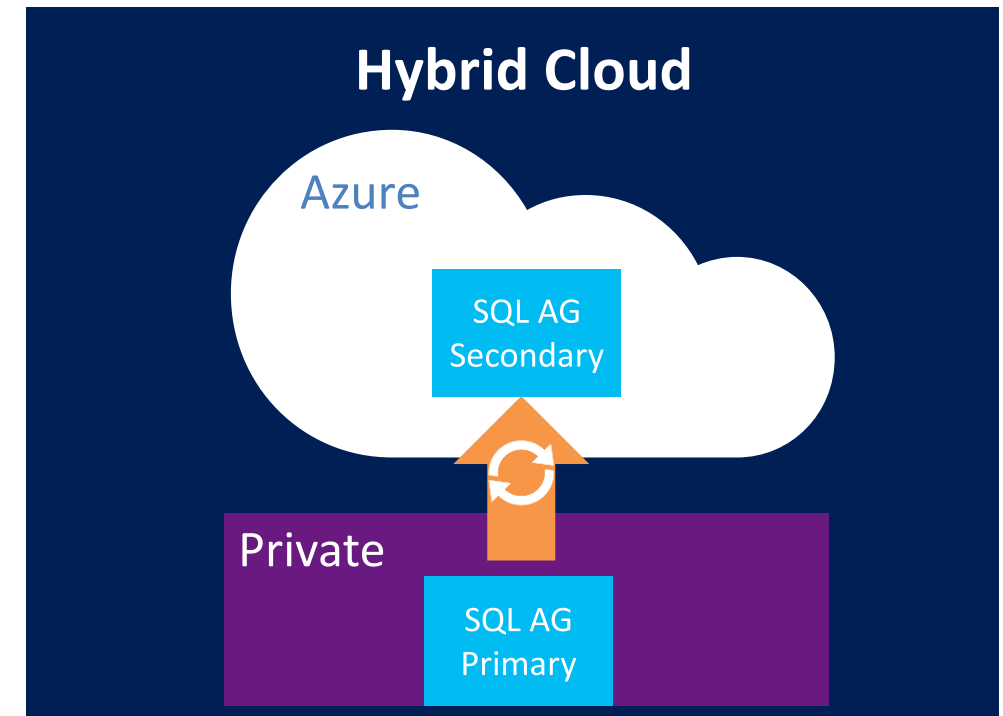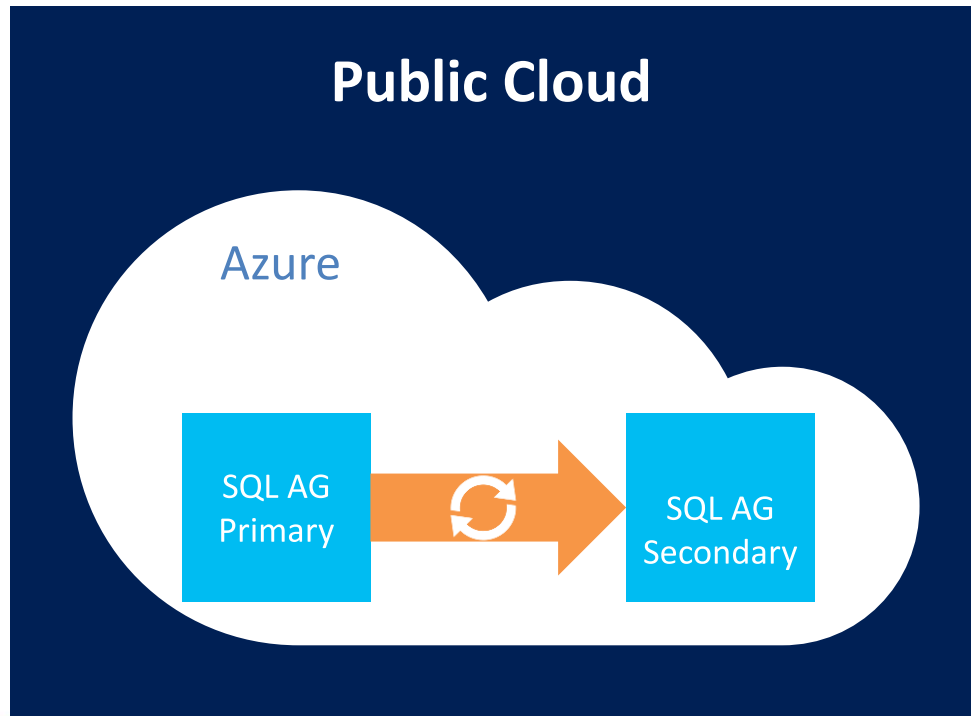Chkdsk integration for file system correction with zero downtime

Performance

CSV block level read-only cache for unbuffered I/O to SQL databases
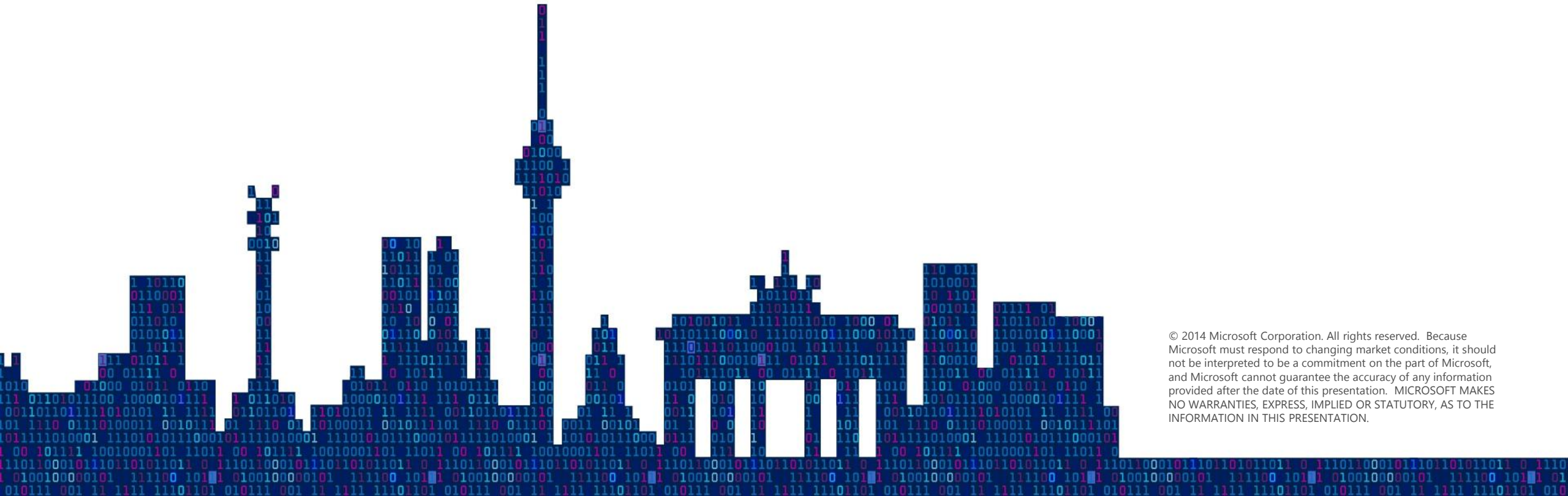
New

Failover Clusters on Windows Azure

SQL Server Availability Groups are now supported on an Azure VM

- http://msdn.microsoft.com/en-us/library/jj870962.aspx

Azure has no shared storage, so clustering requires data replication

Vielen Dank

Microsoft