Microsoft Service

Failover Clustering: Quorum Model

- Premier Field Engineer
- Windows Server Clustering

April 2014

Session Objectives And Takeaways

Session Objective(s):

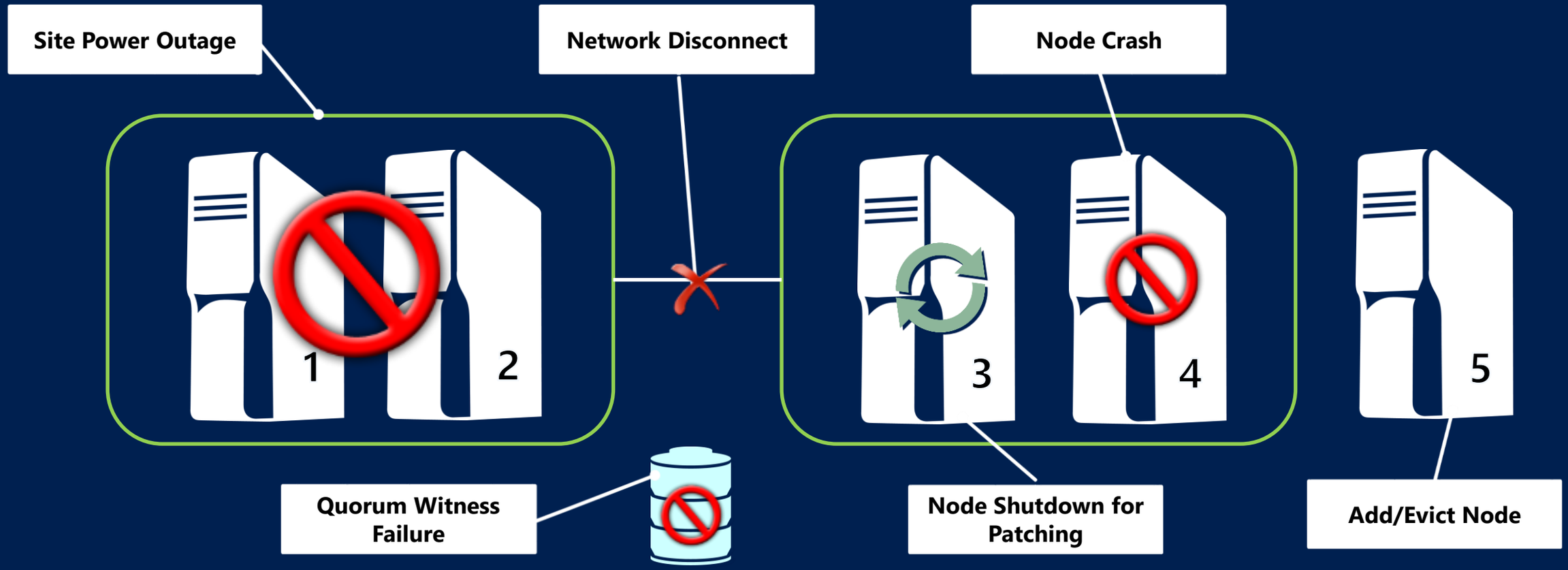
Walk-through Cluster Quorum Fundamentals
New Quorum Features in Windows Server 2012 & R2
Configuration of cluster quorum
Insight into disaster recovery multi-site quorum

Key Takeaway(s):

“Simplified” Cluster quorum configuration
Dynamic Quorum – Increases availability of cluster
Step by step configuration of DR multi-site quorum

Quorum Basics

Cluster challenges



How do I make sure my Cluster stays up ??…

Why Quorum

Faster Start & Recovery of Cluster

Effective quorum policy helps faster start of cluster

Determines the set of nodes that have latest cluster database

Identifying point when to start workload

Determines the point when cluster can host applications

Effective quorum policy prevents unnecessary downtime

Addressing split-brain

Prevent two disjointed instances of the same cluster

Windows Server 2012+R2: Quorum Goals

Simplify Quorum Configuration

Quorum shouldn't affect number of nodes in cluster

Simplified quorum witness selection

Updated wizard for quorum configuration

Increase Cluster High Availability

Cluster more resilient to node/witness failures

Cluster can now survive with <50% majority nodes with Dynamic Quorum

Cluster can now survive even split 50% nodes

Enable more disaster recovery quorum scenarios

Voting Elements in Quorum

Nodes

- Every cluster node has 1 vote
- User configurable per node

Witness

- Witness has 1 vote
  - Disk Witness
  - File Share Witness
- User configurable
  - Single witness per cluster

Cluster needs majority of participating votes to survive

More about this in later slides…

Disk Witness Considerations

Dedicated LUN for internal cluster use

Quorum Disk

Used as arbitration point

Stores a copy of cluster database

Recommendations:

Small disk at least 512 MB in size

Dedicated LUN

NTFS or ReFS formatted

No need for drive letter

File Share Witness Considerations

Simple Windows File Server

Easy to deploy
Single File Server can be used for multiple clusters
Unique File Share per clusters
CNO requires write permissions on the File Share

File Server Location

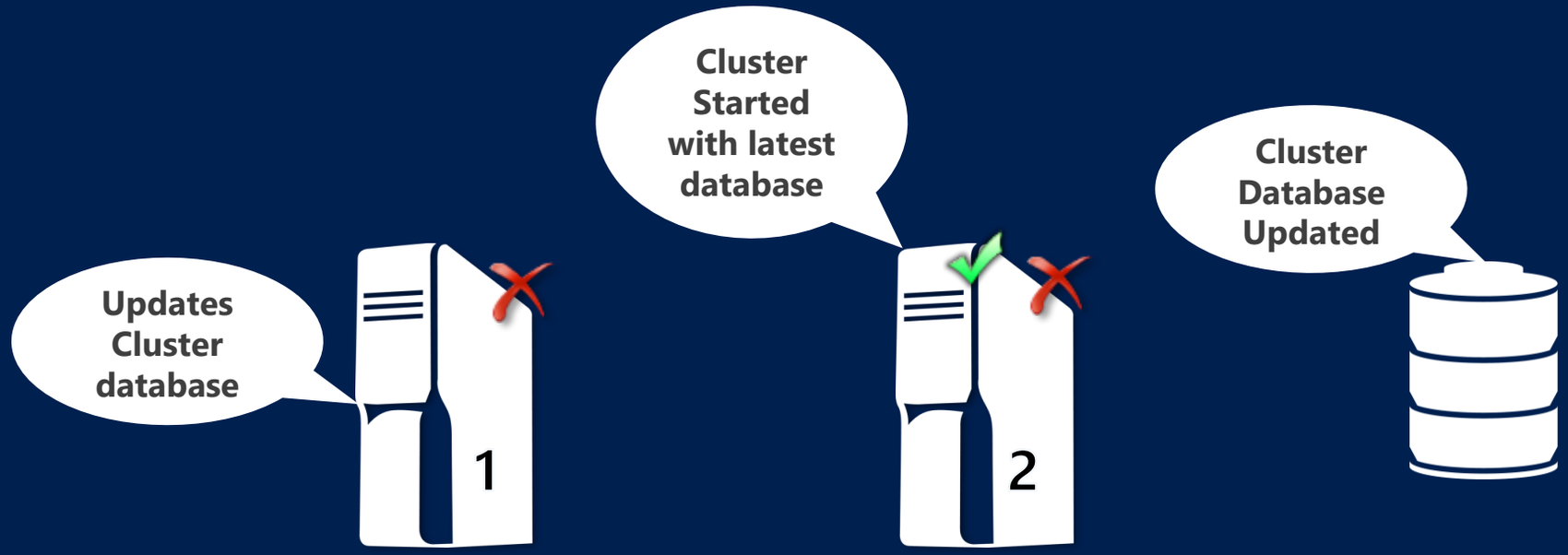
Recommended at 3rd separate site
Not on a node in the same cluster
Not inside VM running in the same cluster
HA File Server configured in a separate cluster

File Share Witness

No copy of cluster database
Minimal network traffic – Cluster membership change only

Partition In Time: Disk Witness

Latest cluster database copy on Disk Witness



Partition In Time: File Share Witness

Prevents node with stale database from forming cluster

Deciding Which Witness to Use

Witness: Disk vs. File Share

	Disk	File Share
Prevents Split-Brain	✓	✓
Prevents Partition-in-Time	✓	✓
Solves Partition-in-Time	✓	
Arbitration Type	SCSI Persistent Reservation	Witness.log file on SMB Share

Recommended: Use Disk Witness if you have shared storage

Key Points to Remember

Quorum enables cluster to survive

Determines the point at which cluster is successfully formed

Voting Elements

Each node has 1 vote and (if configured) witness has 1 vote

Look for updated guidance with Dynamic Witness

Witness selection: Disk or File Share

Disk Witness (recommended) – Stores Cluster DB

File Share Witness – Multisite cluster with replicated storage

Node Vote Weights

Node Vote Weights

Granular control of which nodes have votes

Directly affects quorum calculations

Limit impact on cluster quorum

Cluster quorum does not change if nodes with no vote go down

Nodes with No-Vote continue to be part of the cluster

Receive cluster database updates

Ability to host applications

Why modify Node Vote?

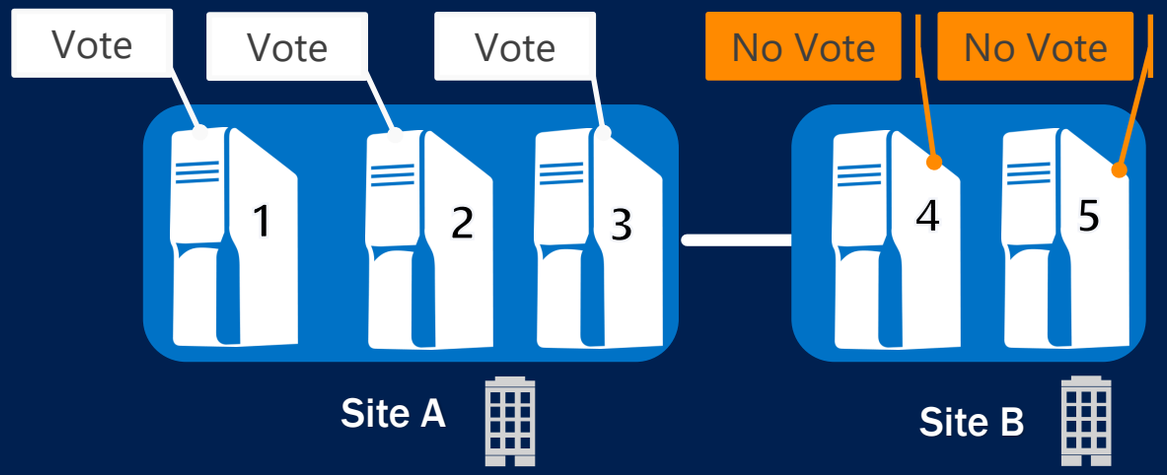
Not all nodes in your cluster are equally important

Typically nodes from Disaster Recovery Backup site

Primarily used for multi-site clusters

Recommended only for manual failover across sites

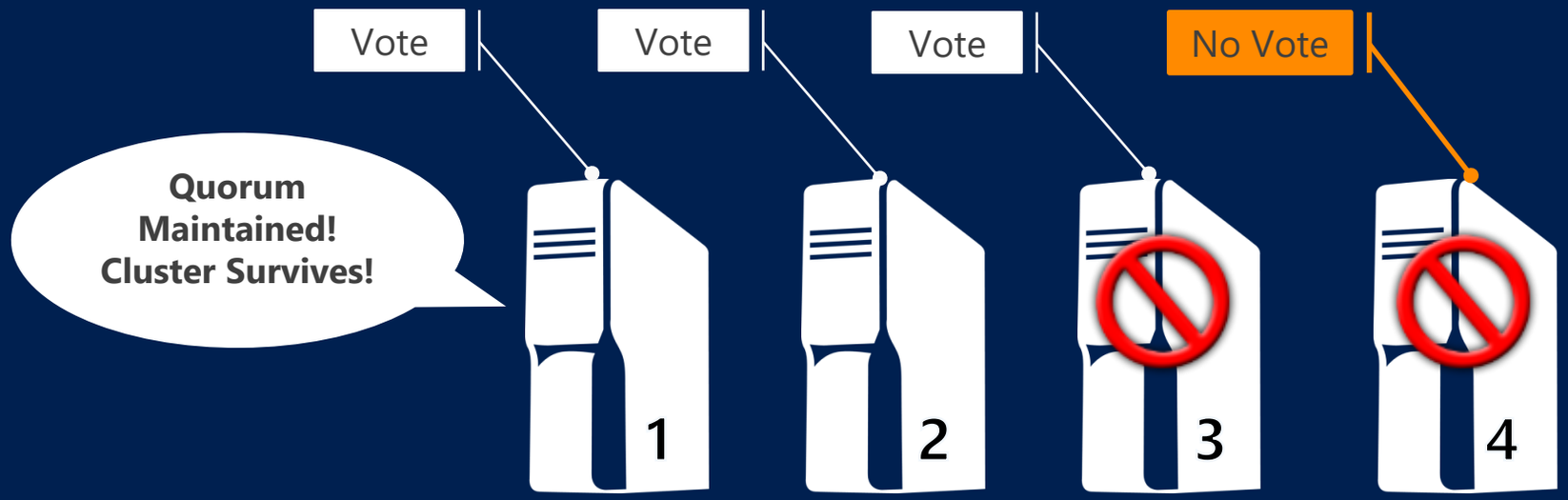
More about this in later slides …



Adjusting majority votes using Node Votes

Original: Total Votes = 4 Majority Votes = 3

Updated: Total Votes = 3 Majority Votes = 2



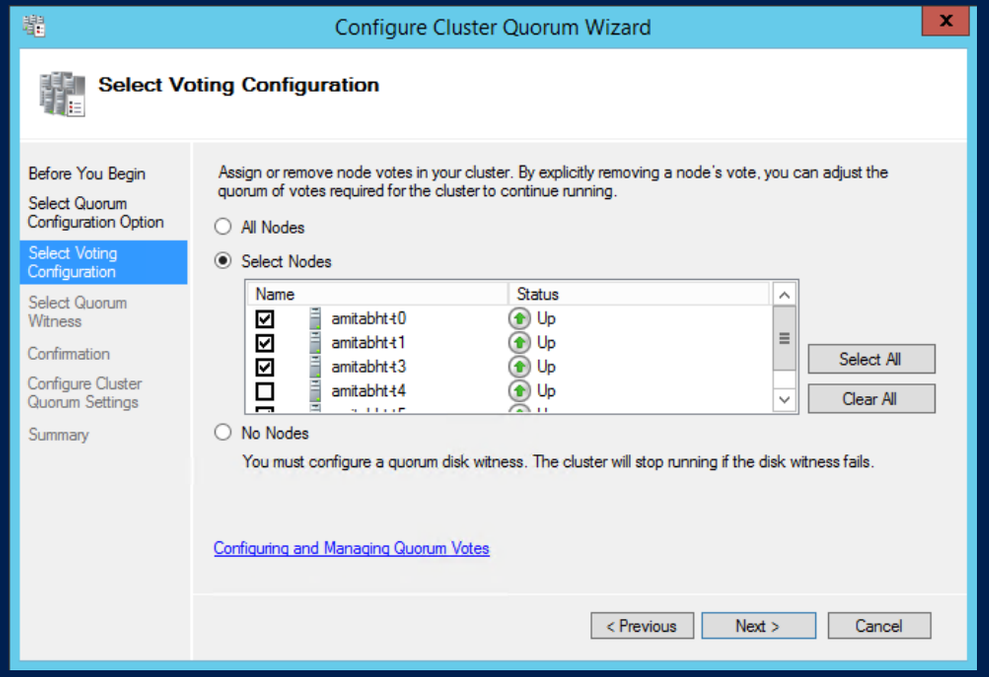
Adjusting Node Vote Weights

Granular control of which nodes have votes

Configurable per cluster node

Can be modified with no downtime

Use PowerShell or Configure Quorum Wizard



NodeWeight
`Default = 1`
`Remove Vote = 0` `Cluster Assigned = 1`
`(Get-ClusterNode <name>).NodeWeight = 0`

UI: Viewing Node Vote Weights

Updated Nodes Page For Easy Viewing

User configured node vote weights in “Assigned Vote” column

Cluster assigned dynamic vote weights in “Current Vote” column

Nodes (5)

Search

Name	Status	Assigned Vote	Current Vote	Information
amitabht-t0	Up	1	1	
amitabht-t1	Up	1	1	
amitabht-t3	Up	1	1	
amitabht-t4	Up	1	1	
amitabht-t5	Up	1	1	

Validate Quorum Configuration

Description: Validate that the current quorum configuration is optimal for the cluster.

Validating cluster quorum settings.

Witness Type: Disk Witness

Witness Resource: Cluster Disk 1

Cluster managed voting: Enabled

Voter Name	State	Assigned Vote	Current Vote
Cluster Disk 1	Online	1	0
amitabht-t0	Up	1	1
amitabht-t1	Up	1	1
amitabht-t3	Up	1	1
amitabht-t4	Up	1	1
amitabht-t5	Up	1	1

This quorum model will be able to sustain failures of 2 node(s) with the disk witness online and 1 node(s) when the disk witness goes offline or fails.

Dynamic Quorum

Dynamic Quorum

Automatic Node Vote Adjustment

Automatic adjustment of Node Vote based on node' state

- Active Node : Dynamic Vote = 1
- Down Node : Dynamic Vote = 0

No change for node with no assigned vote

Dynamic Quorum Majority

Quorum majority is dynamically determined by active cluster nodes

Increase High Availability of Cluster Itself

Sustain sequential node failures or shutdowns

Enables cluster to survive with <50% active nodes

Dynamic Quorum Functionality

Last Man Standing

Cluster can now survive with only 1 node
64-node cluster all the way down to 1 node

Seamless Integration

With existing cluster quorum features & configurations
With multisite disaster recovery deployments

Enabled By Default

Configurable via PowerShell

Dynamic Quorum for Witness

Automatic Witness Vote Adjustment

Automatic adjustment of Witness Vote based on active cluster membership

- Even Active Nodes with Dynamic Vote of 1 : Witness Dynamic Vote = 1
- Odd Active Nodes with Dynamic Vote of 1 : Witness Dynamic Vote = 0

Cluster now has the smarts to determine when to use Witness Vote!

State of Witness

Witness Offline or Failed will automatically make Witness Dynamic Vote = 0

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

User Configurable Quorum Properties

DynamicQuorum

Cluster Common Prop
Default: Enabled

```
1: Enabled
0: Disabled
```

NodeWeight

Node Common Prop
Default: Vote assigned

```
1: Cluster Managed
0: Disable Vote
```

PowerShell
(Get-Cluster).DynamicQuorum = 1
(Get-ClusterNode “name”).NodeWeight = 1

Cluster Managed Quorum Properties

DynamicWeight

Node Common Prop
Value Adjusted by Cluster

```
1: Node Has Vote
0: Node Has No Vote
```

WitnessDynamicWeight

Cluster Common Prop
Value Adjusted By Cluster

```
1: Witness Has Vote
0: Witness Has No Vote
```

PowerShell
(Get-ClusterNode “name”).DynamicWeight (read only)
(Get-Cluster).WitnessDynamicWeight (read only)

Dynamic Quorum : Node Scenarios

Node Shutdown

Node removes its own vote

Node Crash

Remaining active nodes remove vote of the downed node

Node Join

On successful join the node gets its vote back

Dynamic Quorum : Witness Scenarios

Witness Offline

Witness vote gets removed by the cluster

Witness Failure

Witness vote gets removed by the cluster

Witness Online

If necessary, Witness vote is added back by the cluster

Tie Breaker

Cluster will survive simultaneous loss of 50% votes

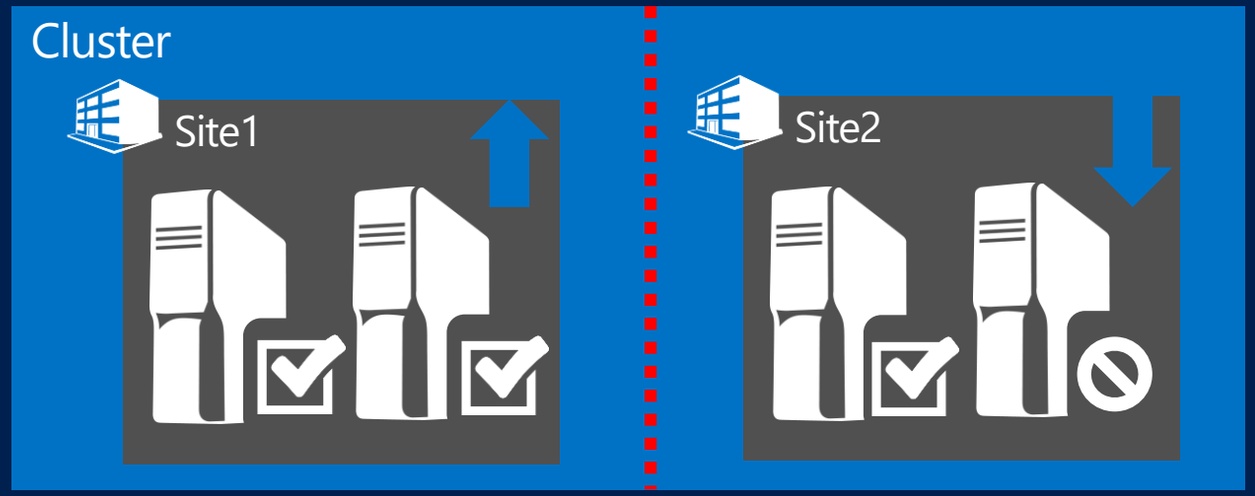
Especially useful in multi-site DR scenarios with even split

Cluster always ensures total number of votes are Odd

One site automatically elected to win

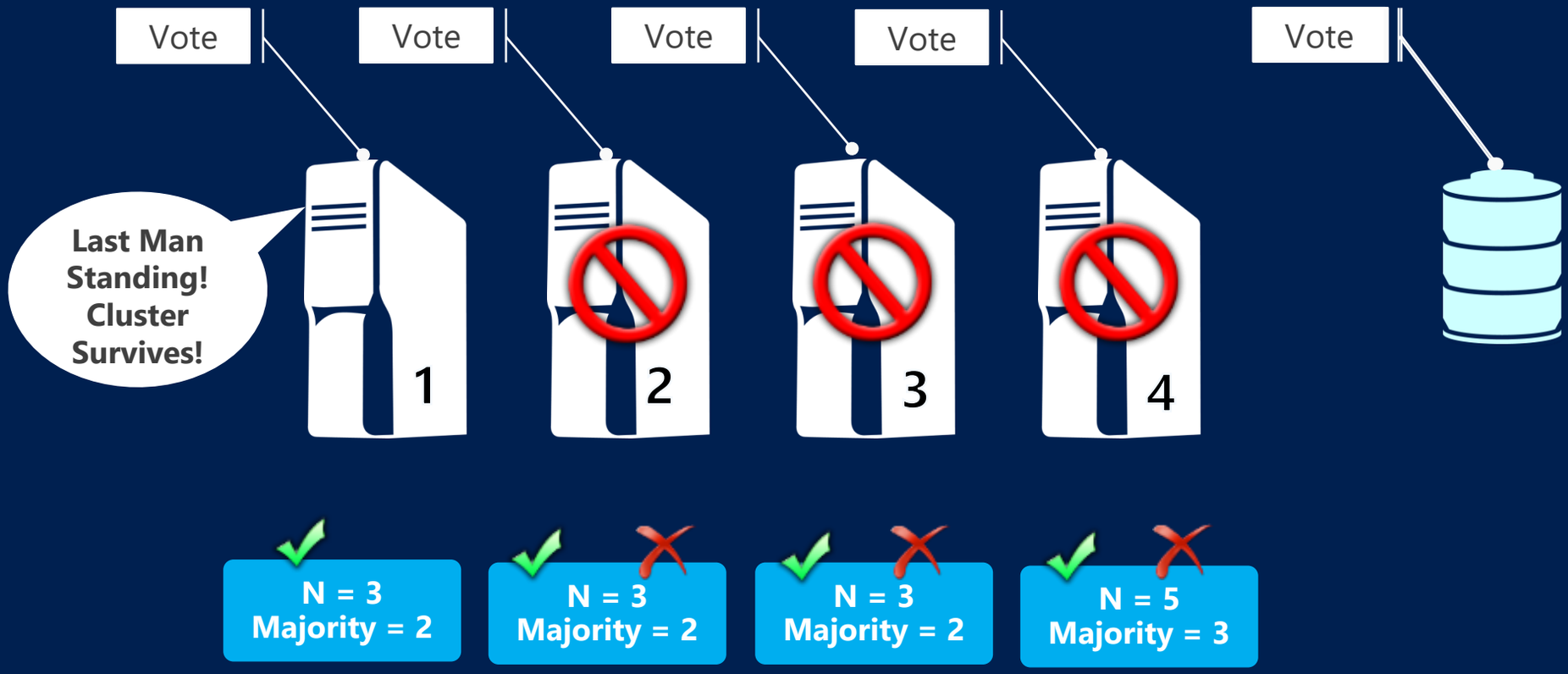
By default, cluster randomly selects a node to take its vote out

`LowerQuorumPriorityNodeID` cluster common property identifies a node to take its vote out



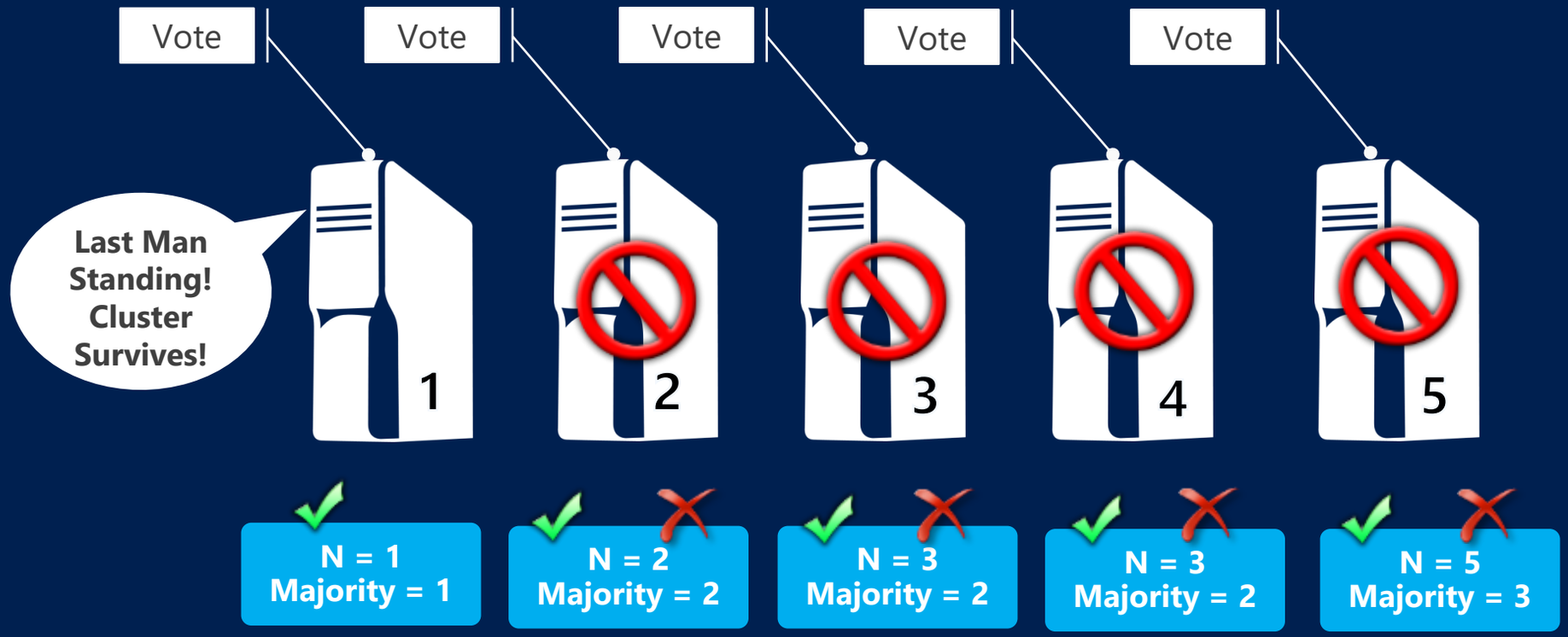
Last Man Standing: Witness Configured

4 Nodes + Witness Configured (N = Number of Votes)



Last Man Standing: No Witness

5 Nodes + No Witness Configured (N = Number of Votes)



No Witness: Last Two Active Nodes

Cluster dynamically removes one node's vote

Cluster can sustain communication loss between the last two nodes

Cluster can sustain crash of node with no vote

Random selection of the node whose vote gets removed

Cluster survives graceful shutdown of either node

	Node 1	Node 2
State	UP	UP
NodeWeight	1	1
DynamicWeight	1	0

Dynamic Quorum

DEMO

Dynamic Quorum Considerations

Simultaneous Loss of Majority Nodes

Need existing majority votes to update new majority votes

Custer cannot sustain simultaneous loss of majority nodes

Always Configure Witness

Witness helps cluster to sustain one extra node failure

Witness helps in giving equal opportunity to survive in DR scenarios (more details later)

Cluster running with <50% majority nodes

The remaining <50% nodes become more important

“Last Man Standing” node becomes necessary for cluster start

Helps prevent partition in time

Dynamic Quorum vs. Disk Only Quorum

Dynamic Quorum

Helps achieve true “Last Man Standing”

Increases cluster availability by making cluster resilient

Disk Only Quorum

No flexibility around vote adjustment (1 vote of disk witness)

Disk Witness is single point of failure

With Dynamic Quorum, no need for Disk Only Quorum

Why lose the cluster when storage is lost?

Key Points to Remember

Dynamic Quorum increase Availability of Cluster

Automatic adjustment of dynamic vote of nodes & witness

Dynamic Quorum enables "Last Man Standing"

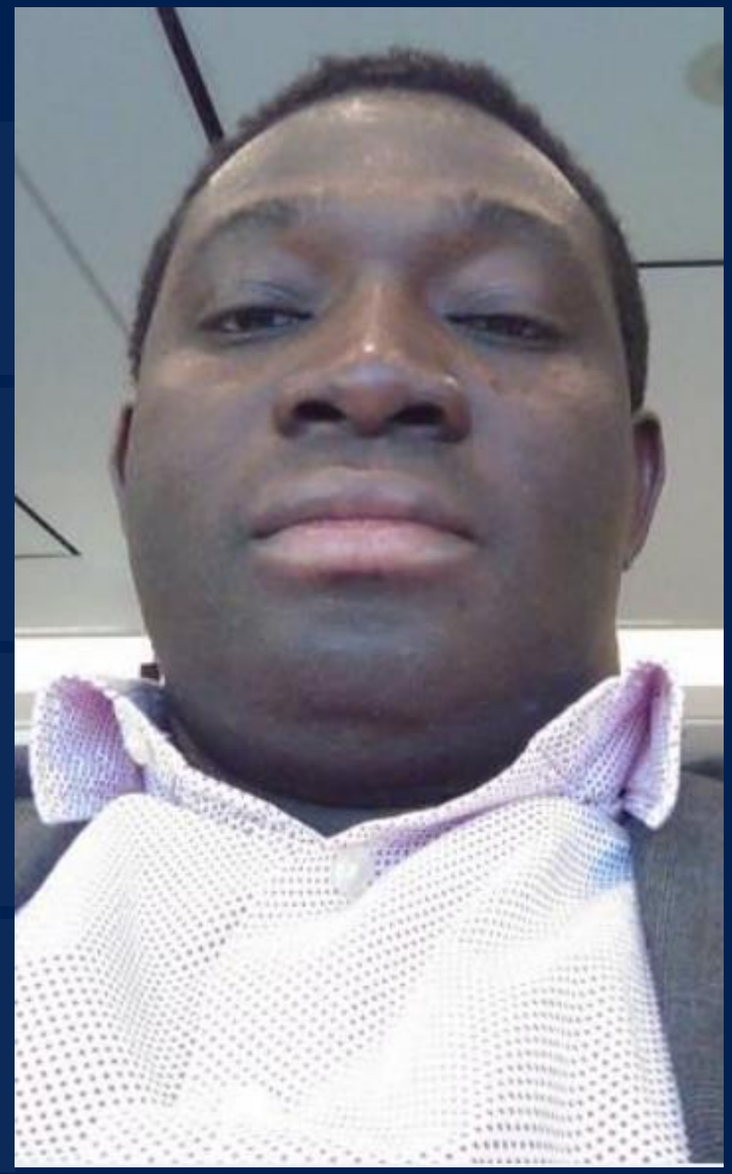
Cluster can survive with only 1 node remaining

Node Vote Adjustment

Only with Manual Failover to DR site; Remove vote of nodes from DR site

Simplified witness selection with Dynamic Witness

Best practice guidelines to always configure quorum witness



Configuring Cluster Quorum

Intuitive Quorum Configuration

Updated Cluster UI Experience

Simplified quorum configuration with updated quorum wizard

Updated Nodes Page

Ability to view node's user configured vote & cluster managed vote

Updated Quorum Validation

Simplified guidance & warning text

Nodes & witness vote information is captured in detail

Simplified Terminology

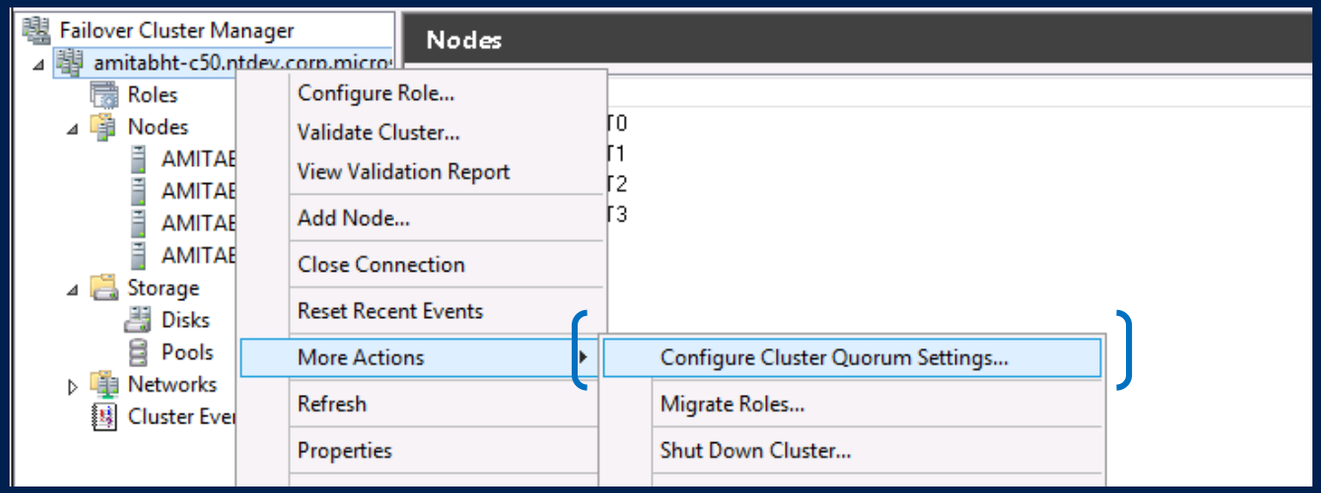
Removed legacy concepts of 'quorum modes'

It is all about witness selection:

"File Share Witness" or "Disk Witness" or "No Witness"

Configured via Cluster Manager GUI and PowerShell

Cluster Quorum Wizard



Updated PowerShell

PowerShell
`Set-ClusterQuorum -NoWitness`
`Set-ClusterQuorum -DiskWitness "DiskResourceName"`
`Set-ClusterQuorum -FileShareWitness "FileShareName"`
`Set-ClusterQuorum -DiskOnly "DiskResourceName"`

New Quorum Wizard

DEMO

Recovery Actions

Force Quorum

Manual Override

Allows to start cluster without majority votes

Cluster starts in a special “forced quorum” mode

Remains in this mode till majority votes achieved

Cluster automatically switches to normal functioning

Caution

Always understand why quorum was lost

Split-brain between nodes possible

You are now in control!

Prevent Quorum Flag
Command Line: net start clussvc /ForceQuorum
PowerShell: Start-ClusterNode –ForceQuorum

Prevent Quorum

Helps prevent nodes with vote to form cluster

Nodes started with ‘Prevent Quorum’ always join existing cluster

Applicable to cluster in “Force Quorum”

Always start remaining nodes with ‘Prevent Quorum’

Helps prevent overwriting of latest cluster database

Forward progress made by nodes in ‘Force Quorum’ is not lost

Most applicable in multisite DR setup

Prevent Quorum Flag
Command Line: net start clussvc /PQ
PowerShell: Start-ClusterNode –PreventQuorum

Force Quorum Resiliency

Cluster detects partitions after a manual Force Quorum

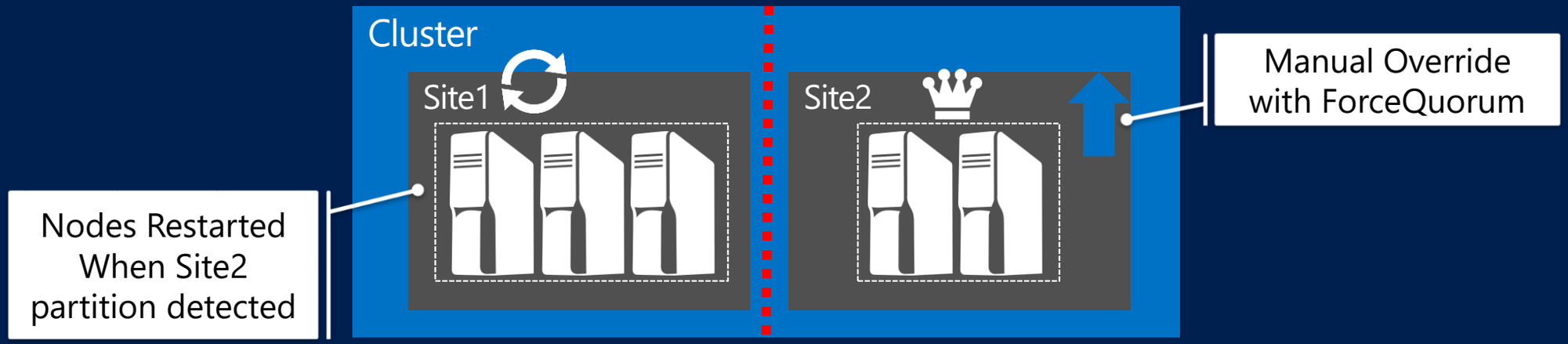
Cluster has the built-in logic to track Force Quorum started partition

Partition started with Force Quorum is deemed authoritative

Other partitions automatically restart up on detecting a FQ cluster

Restarted nodes in other partition join the FQ cluster

Cluster automatically restarts the nodes with Prevent Quorum



Multi-Site DR Quorum

Considerations of Quorum with DR solutions

Types of Multi-Site DR Configurations

What are you Service Level Agreements (SLA's)?

In the event of a disaster, how do you want to switch to your DR site?

Automatic Failover

- Services automatically failover to recovery site in the event of a disaster
- All sites equal

Manual Failover

- Services manually failover to recovery site in the event of a disaster
- Primary & Backup (DR) sites

Automatic Failover Considerations

All Sites Equal

Allow cluster to sustain failure of any one site

Allow automatic failover of workload to the surviving site

Node Vote Weight Adjustments

All nodes equally important

No need to modify node vote weights

Number of Nodes per Site

Keep equal number of nodes in both sites

Helps cluster sustain failure of any site

- Otherwise the site with more nodes would become Primary site

Automatic Failover: Witness Considerations

Always Configure File Share Witness (recommended)

File Server running at a separate site

The separate site must be accessible from the workload sites

- Allows cluster to sustain communication loss between sites

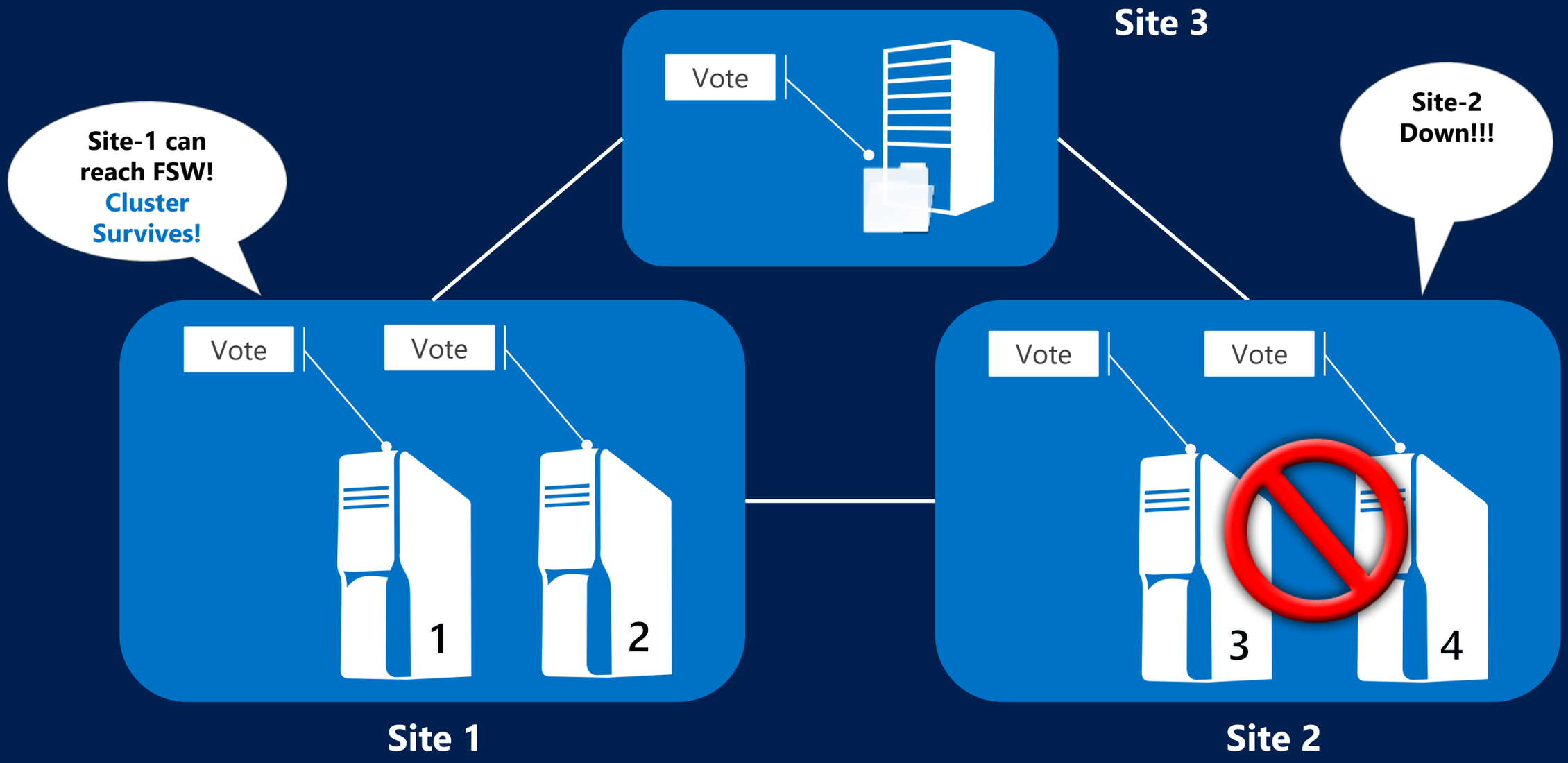
Witness Selection

Highly Available File Server, for witness, in a separate cluster

Disk Witness can be used as directed by storage vendor

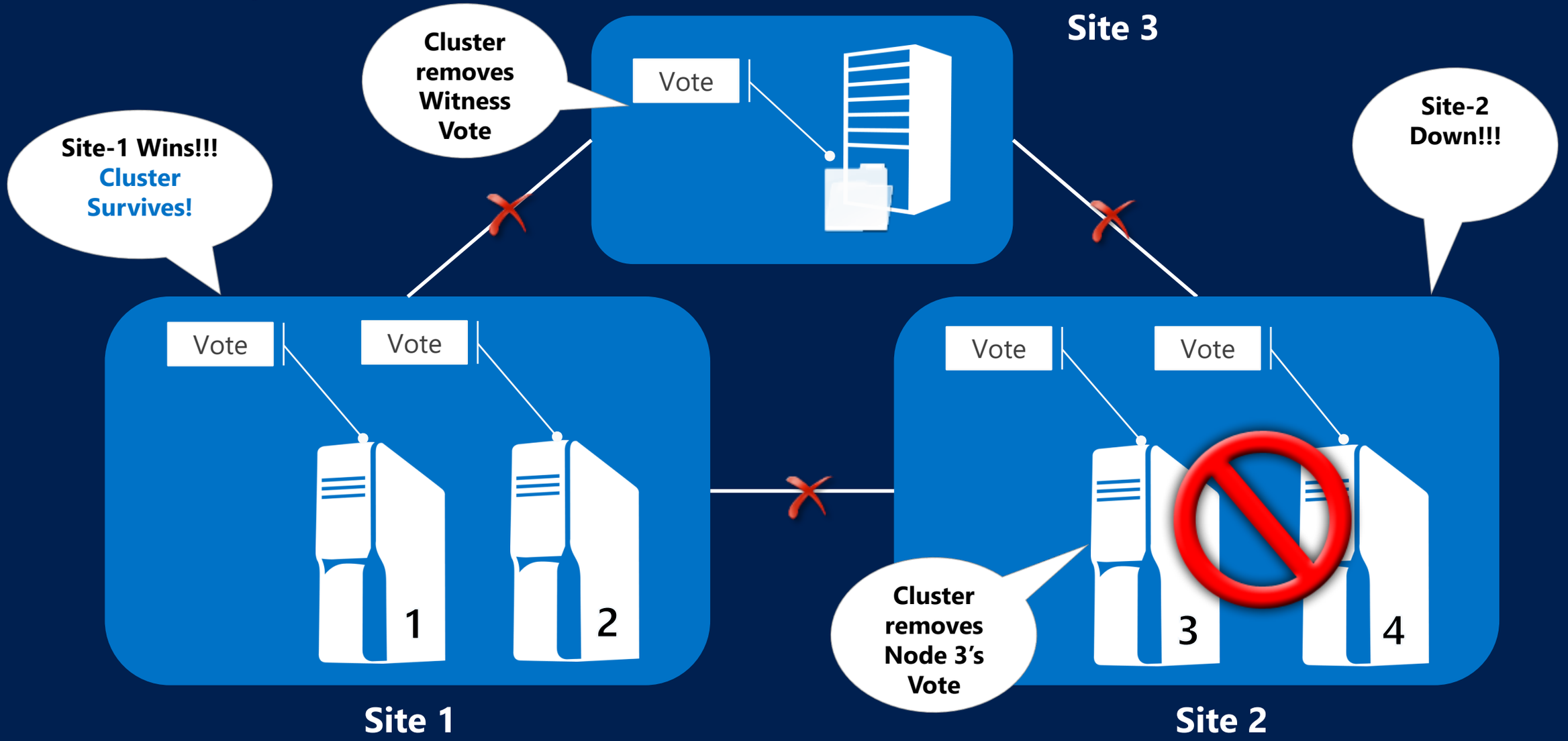
Automatic Failover: 2-Site Cluster

Failover Example



Automatic Failover: WAN Link Issues

Witness Dynamic Vote & Tie Breaker



Manual Failover Considerations

All Sites Not Equal

Cluster cannot sustain failure of Primary site
Allow cluster to sustain failure of the Backup site

Node Vote Weight Adjustments

Disallow nodes in Backup site in affecting cluster quorum
Remove node vote weight of nodes in Backup site

Number of Nodes per Site

No requirement to keep equal number of nodes in both sites

Manual Failover: Workload Considerations

Workload Management

Use Preferred Owners to prioritize keeping workload on Primary site

Recovery Actions

Primary site failure would require “Force Quorum” on Backup site

Recover Primary site nodes using “Prevent Quorum”

Manual Failover: Witness Considerations

Always Configure Witness

File Server running at a separate site (recommended)
File Server running local in Primary Site may be Ok (consider recovery scenarios)

Witness Selection

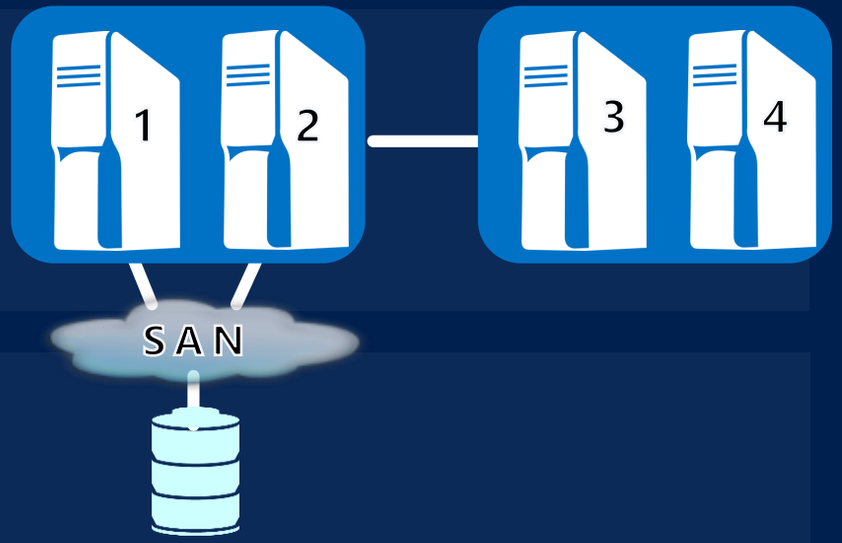
Highly Available File Server, for witness, in a separate cluster
Asymmetric Disk Witness can be used as well (consider recovery scenarios)

Asymmetric Disk Witness

Disk Witness accessibility

Subset of nodes can access the disk

Witness can come online only on subset of nodes



Most applicable in multi-site clusters

Disk only seen by primary site

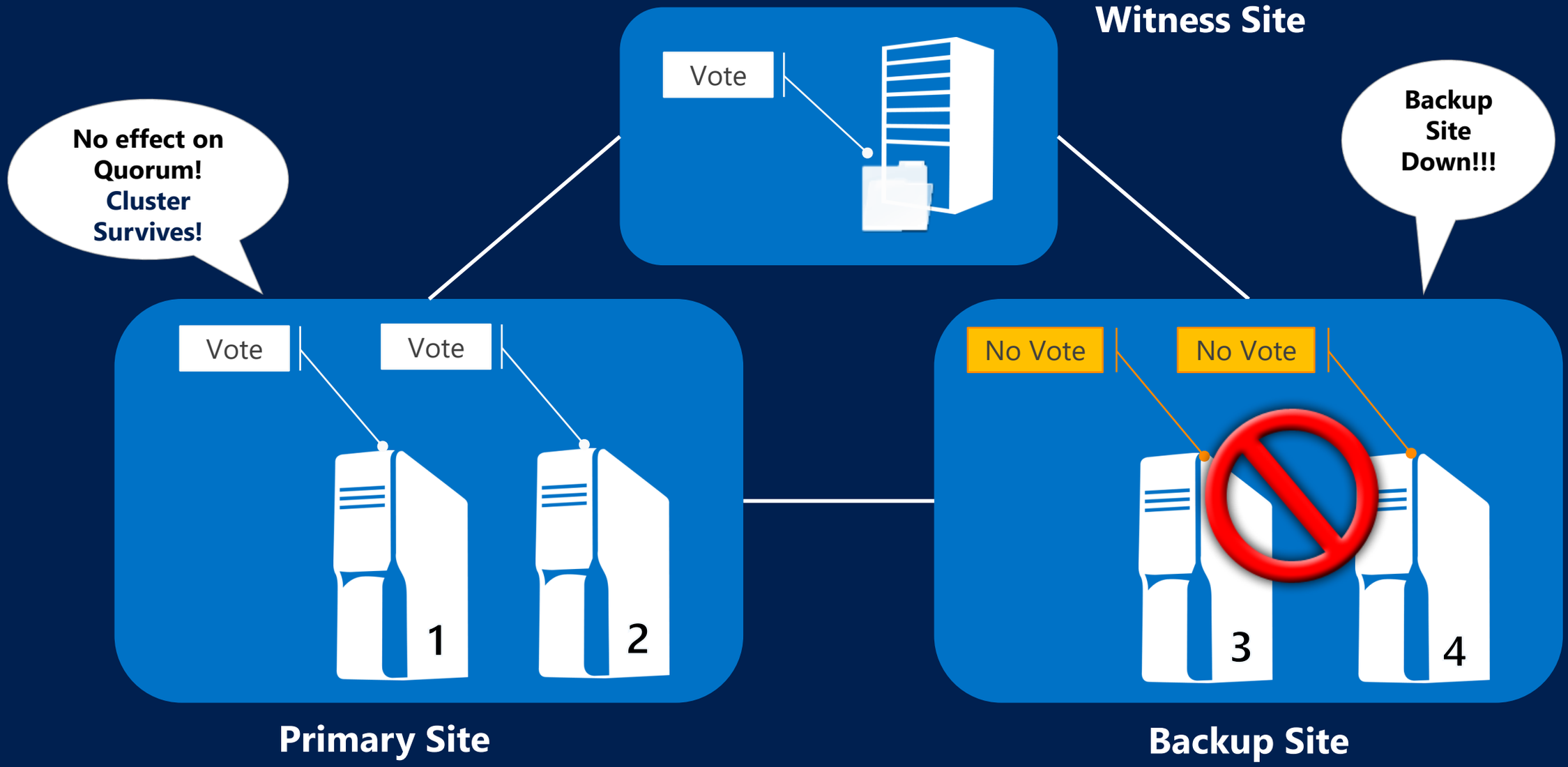
Witness can come online only on primary site

Cluster recognizes asymmetric storage topology

Uses this to place cluster quorum group

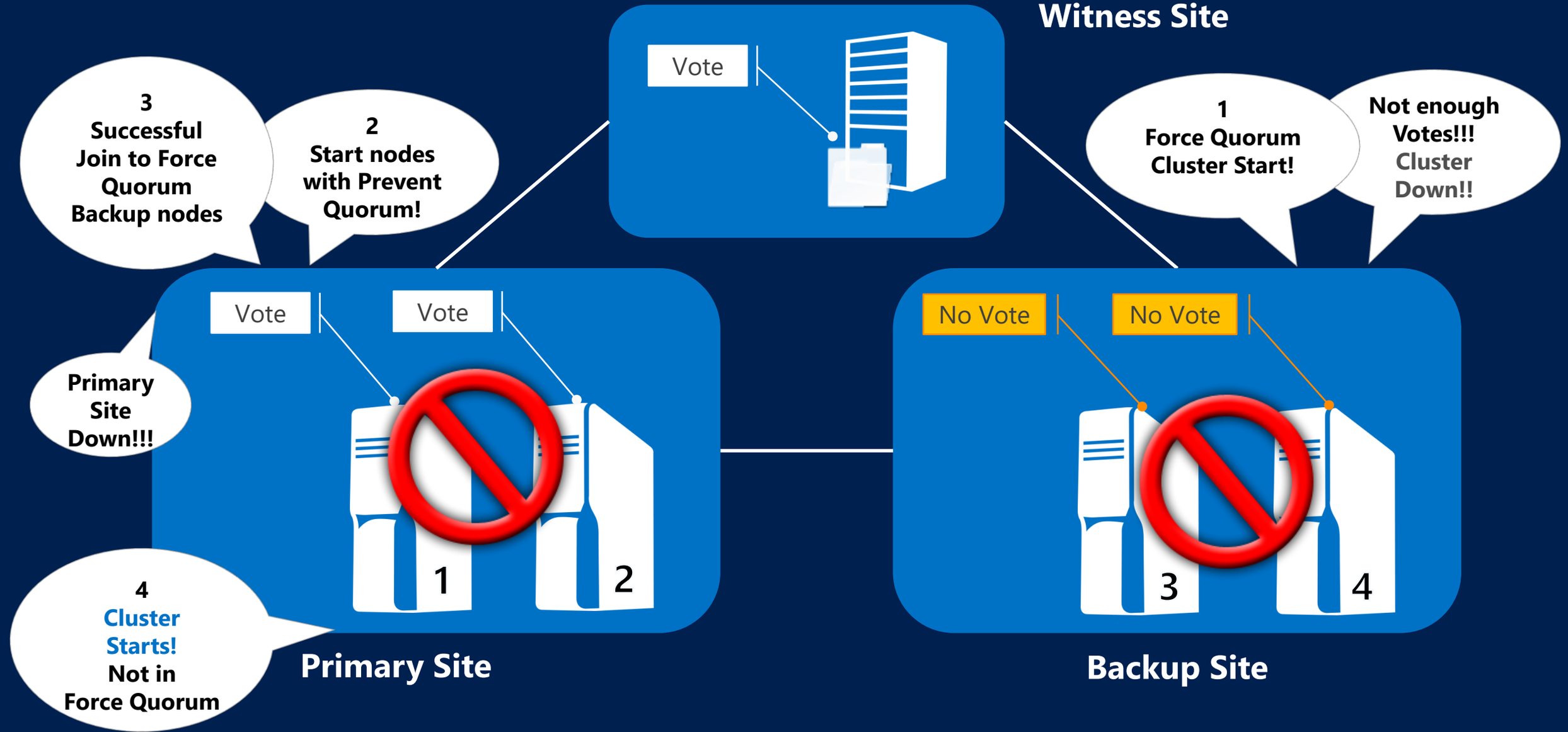
Manual Failover: 2-Site Cluster

Backup Site Down



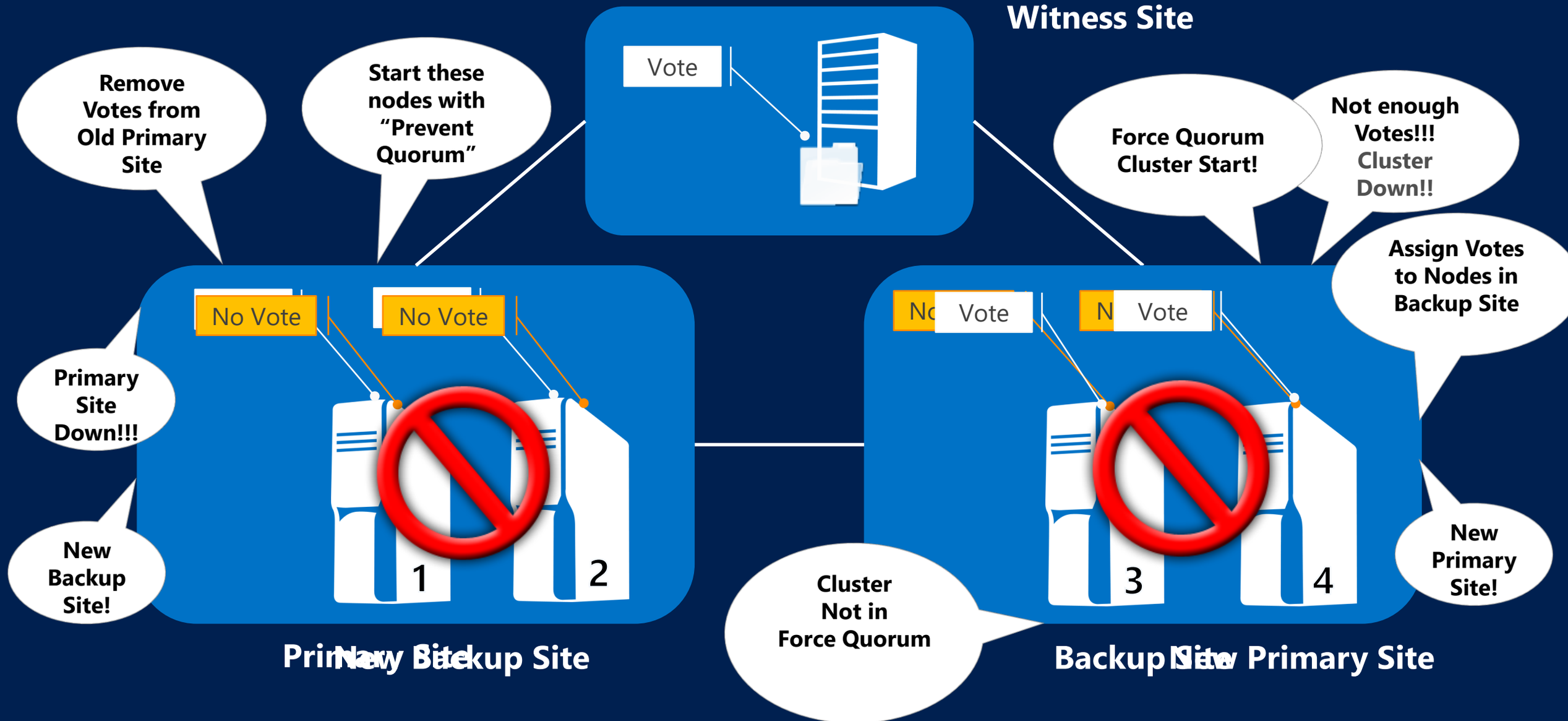
Manual Failover: Temporary Outage

Recommended Recovery



Manual Failover: Long Term Outage

Recommended Recovery



Key Points to Remember

Identify your SLA's for multisite clusters

Automatic vs. Manual Failover

Automatic Failover

Keep nodes equal in both sites

Configure File Share Witness at separate site

Manual Failover

Remove votes of nodes in DR site

Remember the order of recovery actions

Configure asymmetric disk witness or FSW as per votes

In Review: Session Objectives And Takeaways

Session Objective(s):

Walk-through Cluster Quorum Fundamentals
New Quorum Features in Windows Server 2012
Configuration of cluster quorum
Insight into disaster recovery multi-site quorum

Key Takeaway(s):

“Simplified” Cluster quorum configuration
Dynamic Quorum – Increases availability of cluster
Step by step configuration of DR multi-site quorum