Real-Life SQL Server 2012: AlwaysOn Lessons Learned

Brent Ozar: MCM, MVP, Consultant, BrentOzar.com
Jason Hall: Manager, Systems Consulting, Dell

Quest Software
is now a part of Dell

BRENT OZAR
UNLIMITED

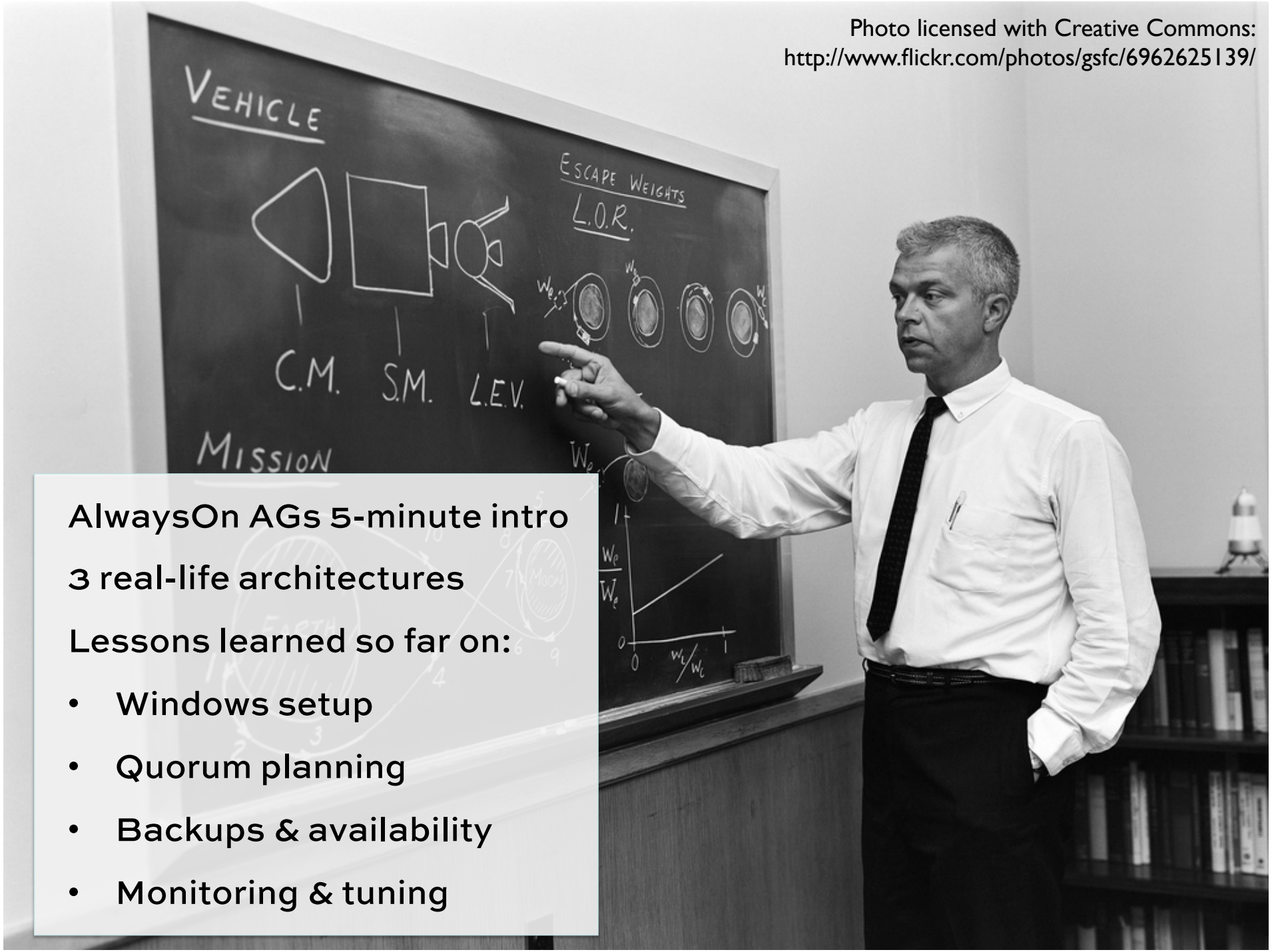
SQL Server 2012 AlwaysOn Availability Groups: Real-Life Lessons Learned

From BrentOzar.com/go/alwayson, it's our

SQL Server 2012 AlwaysOn Availability Groups Checklist

BEFORE YOU RUN THE WIZARD...

The power of Microsoft SQL Server 2012's new AlwaysOn Availability Groups can give you both high availability and disaster recovery, but with power comes...well, complex configuration. We'll help you get started.



AlwaysOn AGs 5-minute intro

3 real-life architectures

Lessons learned so far on:

- Windows setup
- Quorum planning
- Backups & availability
- Monitoring & tuning

5-Minute Intro to AlwaysOn Availability Groups

What we need

High Availability:
BSODs, hardware failure

Disaster Recovery:
the smoking crater scenario

Scale Out:
spread work across multiple servers

What we had

Clustering: tricky, required shared drives, still in one city, single point of failure with shared drives

Mirroring: two servers max, 2nd box unreadable, databases fail over individually

Log Shipping: worked, but kicked people out when restoring

Replication: #%&! unreliable

2012 AlwaysOn Features

	AlwaysOn Clustering	AlwaysOn Availability Groups
Failover unit	Entire instance	Groups of databases
Shared storage required	Yes	No
Automatic failover	Yes	Yes
Query and back up from secondaries	No	Yes

Databases:
Sales
SalesDelivery
SalesWebSite

Microsoft® SQL Server™

SQLPROD1

Chicago, Illinois

Databases:
Sales
SalesDelivery
SalesWebSite

Microsoft® SQL Server™

SQLPROD2

Chicago, Illinois

Databases:
Sales
SalesDelivery
SalesWebSite

Microsoft® SQL Server™

SQLDR1

Portland, Oregon

AlwaysOn Availability Group SQLPROD_Sales
IP Address: 172.16.226.139

Databases:
Sales
SalesDelivery
SalesWebSite

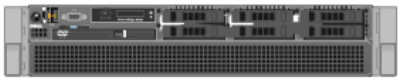
Microsoft® SQL Server®

SQLPROD1

Chicago, Illinois

Databases:
Sales
SalesDelivery
SalesWebSite

Microsoft® SQL Server®



SQLPROD2

Chicago, Illinois

Databases:
Sales
SalesDelivery
SalesWebSite

Microsoft® SQL Server®

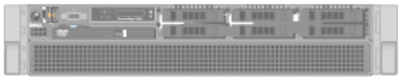
SQLDR1

Portland, Oregon

AlwaysOn Availability Group SQLPROD_Sales
IP Address: 172.16.226.139

Databases:
Sales
SalesDelivery
SalesWebSite

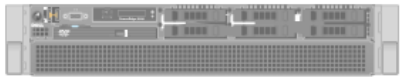
Microsoft® SQL Server®



SQLPROD1
Chicago, Illinois

Databases:
Sales
SalesDelivery
SalesWebSite

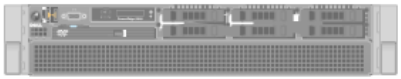
Microsoft® SQL Server®



SQLPROD2
Chicago, Illinois

Databases:
Sales
SalesDelivery
SalesWebSite

Microsoft® SQL Server®



SQLDR1
Portland, Oregon

Windows Cluster SQLPROD-C1
IP Address: 172.16.226.138

AlwaysOn Availability Groups

Built atop Windows Clustering
(but not shared storage)

Like database mirroring, 1 primary server for writes

Up to 4 more replicas for reads, backups, HA, DR

Up to 3 of those replicas can be synchronous

Any 2 of the synch replicas can do automatic failover

Replicas can be in different cities, different hardware

A.G. Drawbacks

Licensing: SQL Enterprise Edition ($7k/core)

All servers must be on same domain, in one cluster

Doesn't provide HA for replication distributor

No cross-database transactional consistency on failover (or for readable replicas)

Version 1.0

Pushes Windows clustering harder than any other app, and is uncovering breaking points

Real-Life Examples

3 sample deployments

Small: standalone SQL Servers, no SAN

Medium: failover cluster with shared storage

Most complex: multiple failover clusters, SANs, AGs

Small: StackExchange

Web site network for questions & answers:
StackOverflow.com, DBA.StackExchange.com, more

High traffic: >8mm visitors and >75mm pages per month

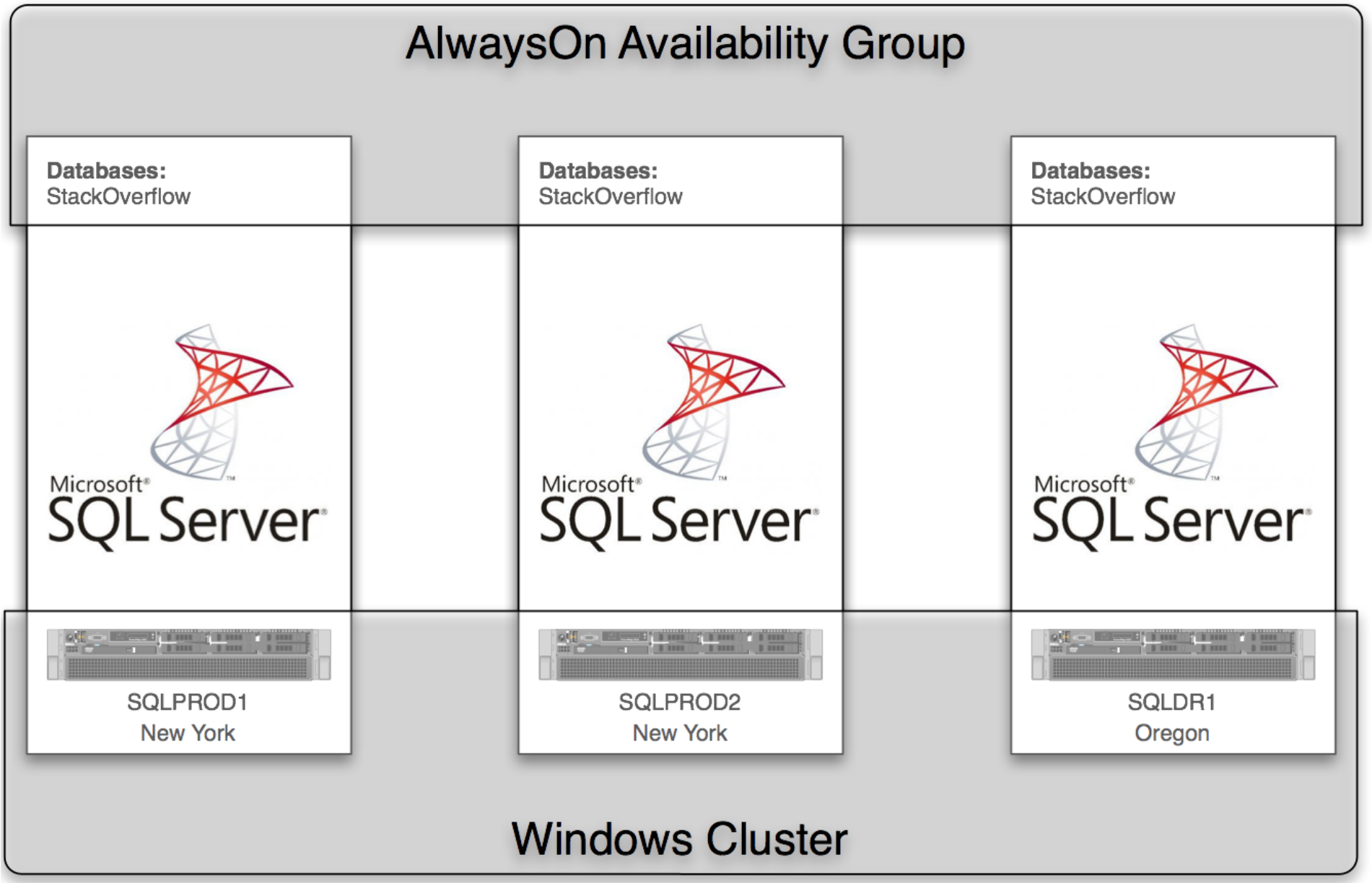
Microsoft SQL Server 2008/2012 back end

No dedicated full time DBAs

Requirements:

- Easy high availability and disaster recovery
- Some data loss and downtime tolerable
- Avoid expensive shared storage
- Scale-out is a bonus for things like API read requests

StackExchange Architecture



Medium: Allrecipes.com

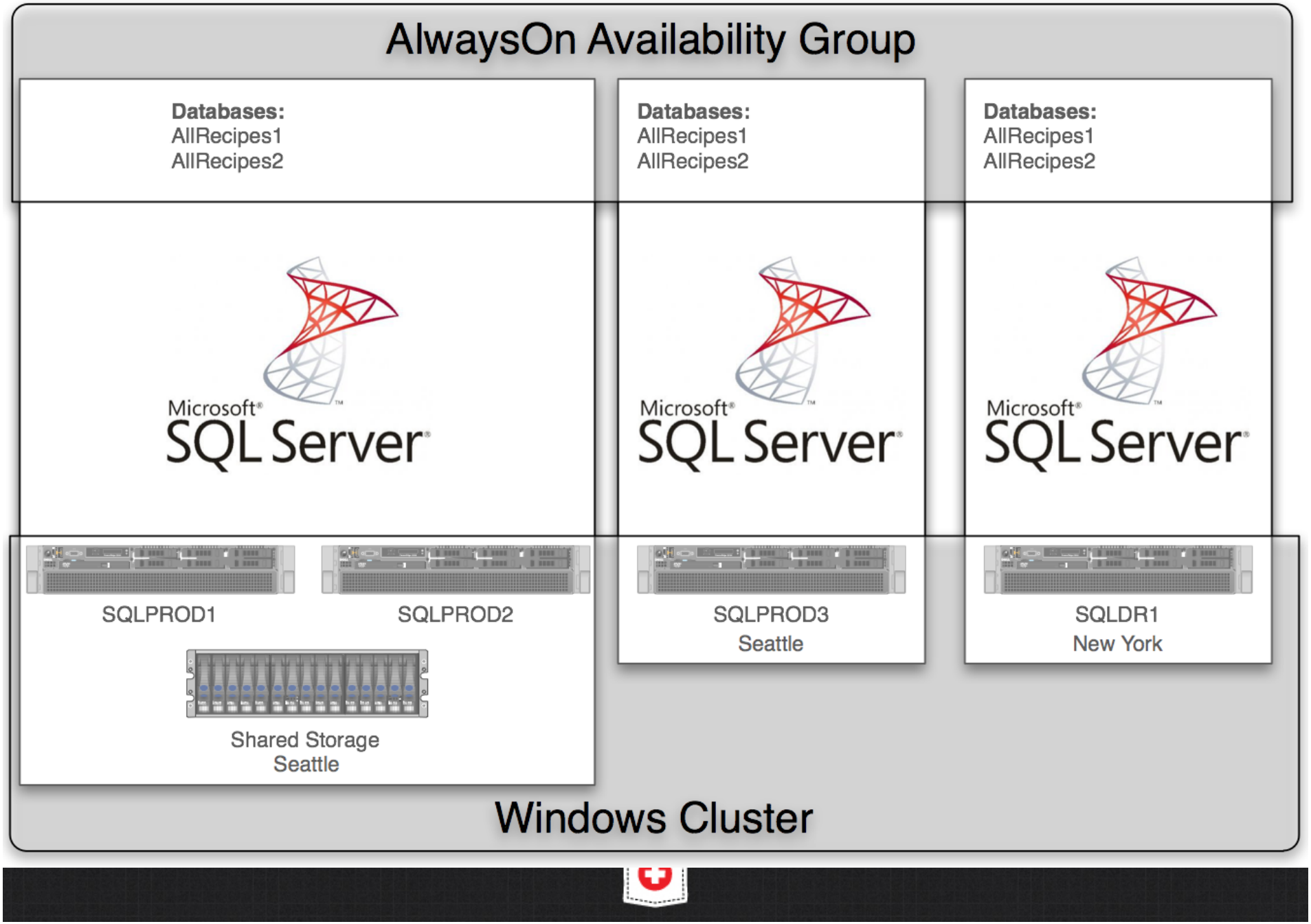
High traffic: >10mm visitors/month during holiday season

Microsoft SQL Server 2008/2012 back end

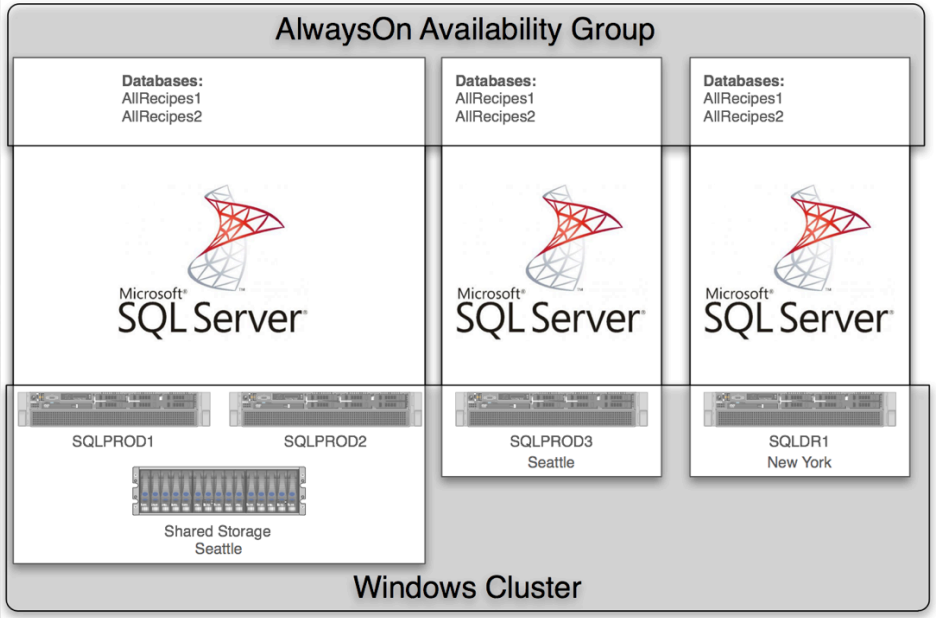
Team of dedicated DBAs comfortable with failover clustering and shared storage

Requirements:

- Automatic failover between servers with zero data loss and no performance degradation
- Manual failover betw datacenters with some data loss
- Need readable replica for near-real-time business reports
- Require the ability to serve traffic from two datacenters


AlwaysOn Availability Group
Databases:
AllRecipes1
AllRecipes2
Microsoft
SQL Server
SQLPROD1
SQLPROD2
Shared Storage
Seattle
Databases:
AllRecipes1
AllRecipes2
Microsoft
SQL Server
SQLPROD3
Seattle
Databases:
AllRecipes1
AllRecipes2
Microsoft
SQL Server
SQLDR1
New York
Windows Cluster

Allrecipes.com



Failover clustering for automatic failover with zero data loss

Seattle read replica for BI

If the Seattle datacenter fails, can serve traffic out of NY instead (but reports will be slower)

Can serve read-only web pages out of NY, do writes across the WAN to Seattle

More complex than standalones: involves a SAN

Large: Discovery Education

Complex 24/7 web application

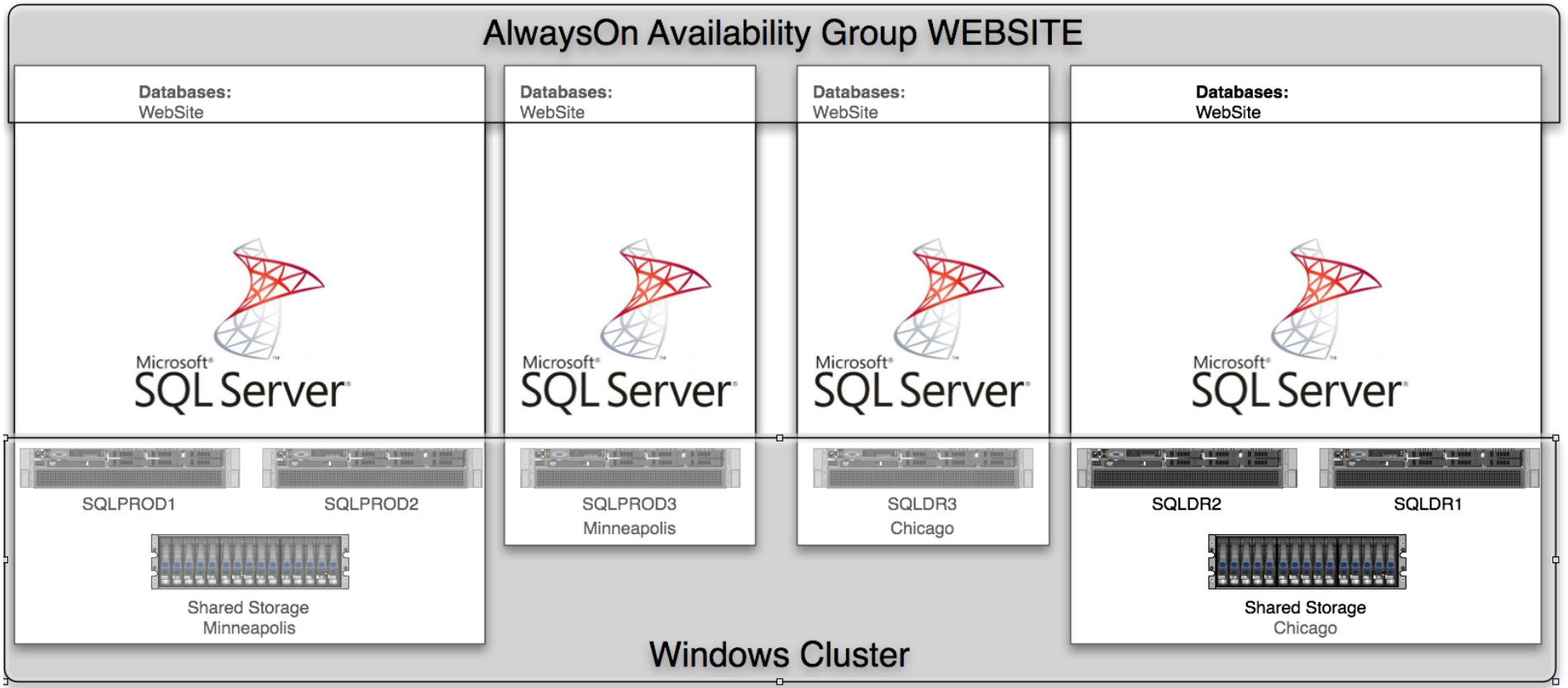
Microsoft SQL Server, MySQL, other technologies

Team of dedicated DBAs comfortable with failover clustering and shared storage

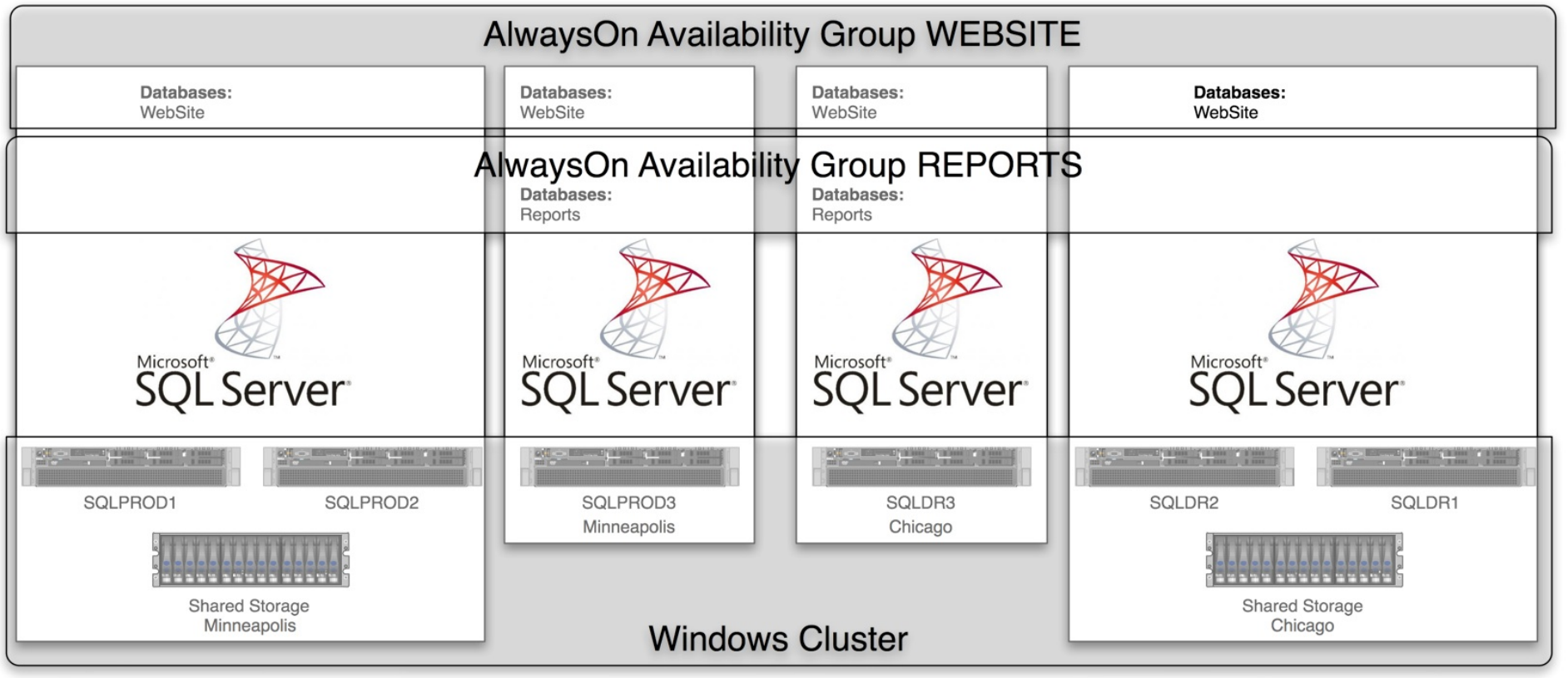
Requirements:

- Automatic failover between servers
- Manual failover between datacenters with some data loss, but need full power in either datacenter
- Require the ability to serve traffic from two datacenters
- Wanted simpler reporting processes

The “simple” version



Closer to the real thing



Lessons Learned: Windows Setup

OS prep: apply hotfixes

Microsoft maintains a list of some* of the hotfixes recommended for Windows clusters:

- Win2008R2 SP1 – KB 2545685
- Win2012 – KB 2784261 (note that the new "cumulative updates" don't list what hotfixes are included – we're running on blind faith)

It's up to you to follow up on these, and they do change: consider using WatchThatPage.com

Find the new links when a new Win SP comes out

* Some – not all, as we'll see next

Win2012 Cumulative Update Mystery

Summary

This article describes the cumulative update for Windows 8 and for Windows Server 2012 that is dated November 2012. This cumulative update package includes performance and reliability improvements for Windows 8 and for Windows Server 2012. We recommend that you apply this cumulative update as part of your regular maintenance routines.

↑ Back to the top | Give Feedback

Introduction

This cumulative update includes the following performance and reliability improvements:

- Enable enterprise customers to customize the default lock screen.
- Improves the performance when you wake the computer and when the computer is asleep, in order to improve battery life
- Resolves an issue that may prevent Windows Store Apps from being installed fully
- Other software updates and performance improvements

Note If the Nalpeiron Licensing Service is running on the computer, when you install this update, the Configuring Updates stage stops at 15 percent completion.
For more information about how to work around this issue, click the following article number to view the article in the Microsoft Knowledge Base:
2787757 Configuring Updates stage stops at 15 percent completion when you try to install update 2770917 in Windows 8

↑ Back to the top | Give Feedback

KB 2777201 for Win2008R2

SQL Server 2012 service crashes when a replica SQL Server 2012 instance goes offline on a Windows Server 2008 R2-based failover cluster

Article ID: 2777201 - View products that this article applies to.

Hotfix Download Available

Expand all | Collapse all

⊕ On This Page

⊖ Symptoms

Assume that you enable the AlwaysOn Availability Groups feature in Microsoft SQL Server 2012 on a Windows Server 2008 R2-based failover cluster. In this situation, when a replica SQL Server 2012 instance goes offline, the SQL Server service on the replica crashes.

↑ Back to the top | Give Feedback

⊖ Cause

This issue occurs because a *double free* situation occurs in the Clusapi.dll file.

KB 2779069 for Win2008R2 (Note: doesn't mention SQL.)

A hotfix is available that adds two new cluster control codes to help you determine which cluster node is blocking a GUM update in Windows Server 2008 R2

Article ID: 2779069 - View products that this article applies to.

Hotfix Download Available

Expand all | Collapse all

On This Page

INTRODUCTION

This article describes a hotfix that introduces a design-level change for Windows Server 2008 R2. The hotfix adds two new cluster control codes that you can use to determine which cluster node is blocking a Global Update manager (GUM) update.

Currently, when Microsoft Exchange Online encounters one or more cluster nodes that have storage problems, the cluster node is shut down.

For example, a write request that is sent to the cluster database becomes stuck in the storage stack. Then, the cluster node that holds the GUM lock sends a Multicast Request Reply (MRR) message to request a cluster database update. However, the node does not receive all the message receipt confirmations from the nodes that receive the message. This behavior occurs because one or more nodes are stuck in the writing process. Therefore, the cluster is shut down.

(Did I mention that using Win2008R2 is a bad idea?)

Need great support processes

Start support cases with Microsoft early and often

Challenge what MS engineers tell you

Read all new KB articles as they come out at http://support.microsoft.com/select/default.aspx?target=rss&c1=508&

Use a logically identical staging environment to replicate and test hotfixes

- Doesn't have to be the same physical hardware, but the more identical it is, the better your chances are at replicating issues
- Needs to be the same network and SAN hardware
- Ideally, include load testing if load is a problem

Multi-subnet clusters

Multi-subnet can mean

Cluster nodes have multiple network adapters, like regular networking plus iSCSI or heartbeats

Cluster nodes are located in multiple different subnets, like multiple datacenters

Multiple subnet gotcha

Typical multi-subnet config across datacenters:

- Chicago-Server1:
  regular: 192.168.100.1
  iSCSI: 192.168.150.1
- NewYork-Server2:
  regular: 192.168.200.1
  iSCSI: 192.168.250.1

Servers have no subnets in common, so Windows will try to reach every IP from every subnet.

Read that again, because it's crazy and important.

http://blogs.msdn.com/b/clustering/archive/2009/02/23/9441604.aspx

Implications

This means if you don't have any subnets in common, you have to route every IP across every subnet. This means routing traffic across your iSCSI and heartbeat networks, which ordinarily sounds like the dumbest idea in the history of networking, and it probably is, but hey, I told you this stuff was going to be painfully complicated.

Or, you can take the easy button:

- Use fiber channel storage, not iSCSI
- Don't bother with heartbeat networks
- Use just one teamed network per node

Listener IPs and DNS

In multi-datacenter clusters, all possible listener IPs are in DNS at all times.

To handle this, all of your apps have to use:

- Latest .NET frameworks (v4.5, v4.02, v3.5 SP1 KB 2654347)
- Connection string parameter to try all IPs simultaneously instead of serially: MultiSubnetFailover = True

This is even true with things like SQLCMD and SSIS.

Lessons Learned: Quorum Planning

Plan quorum votes & reboots

By default, each node has a vote

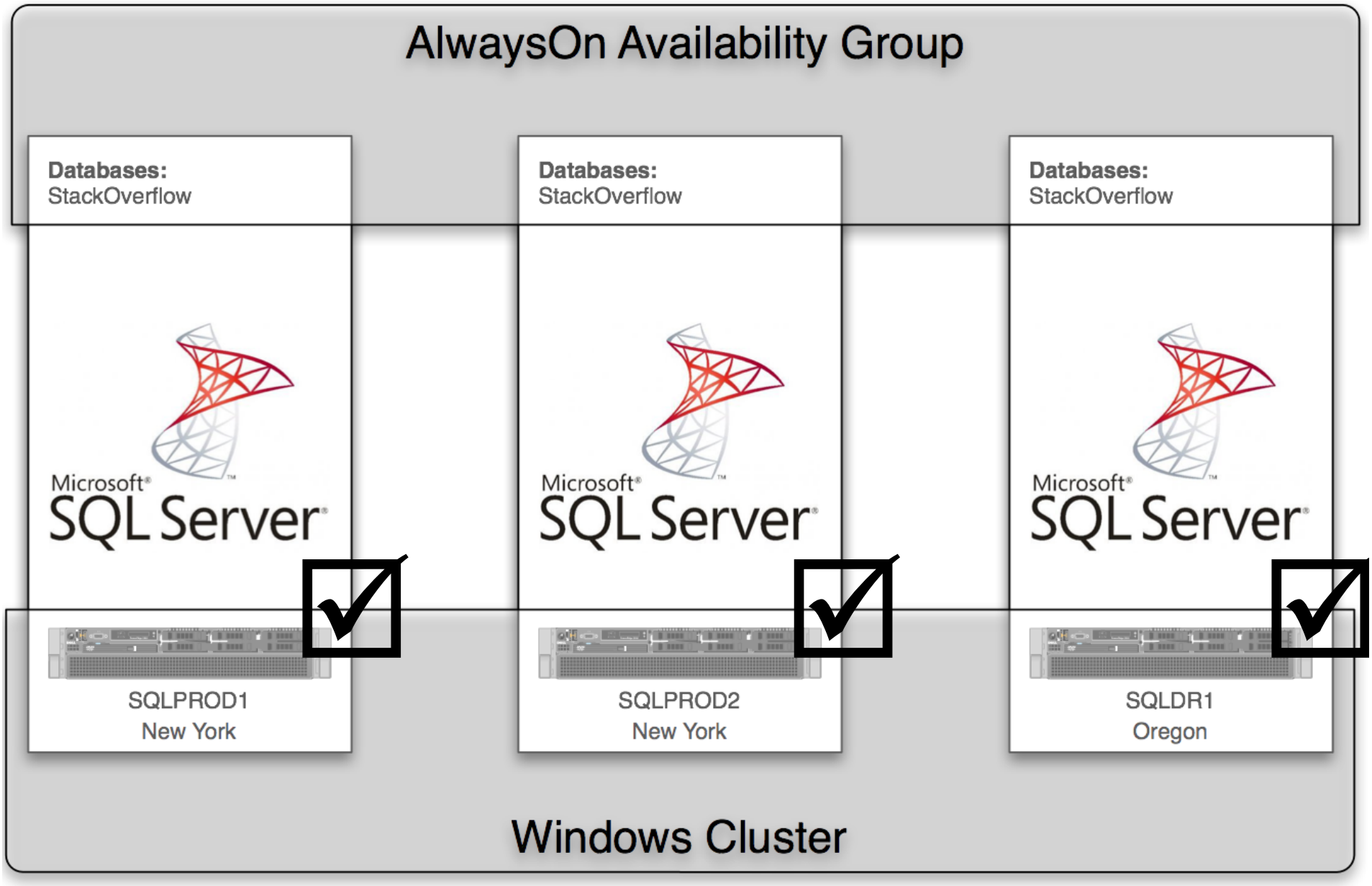
As long as a majority of voters can see each other, the cluster stays up and running

When a minority of voters can't see anyone else, their clustered services shut down

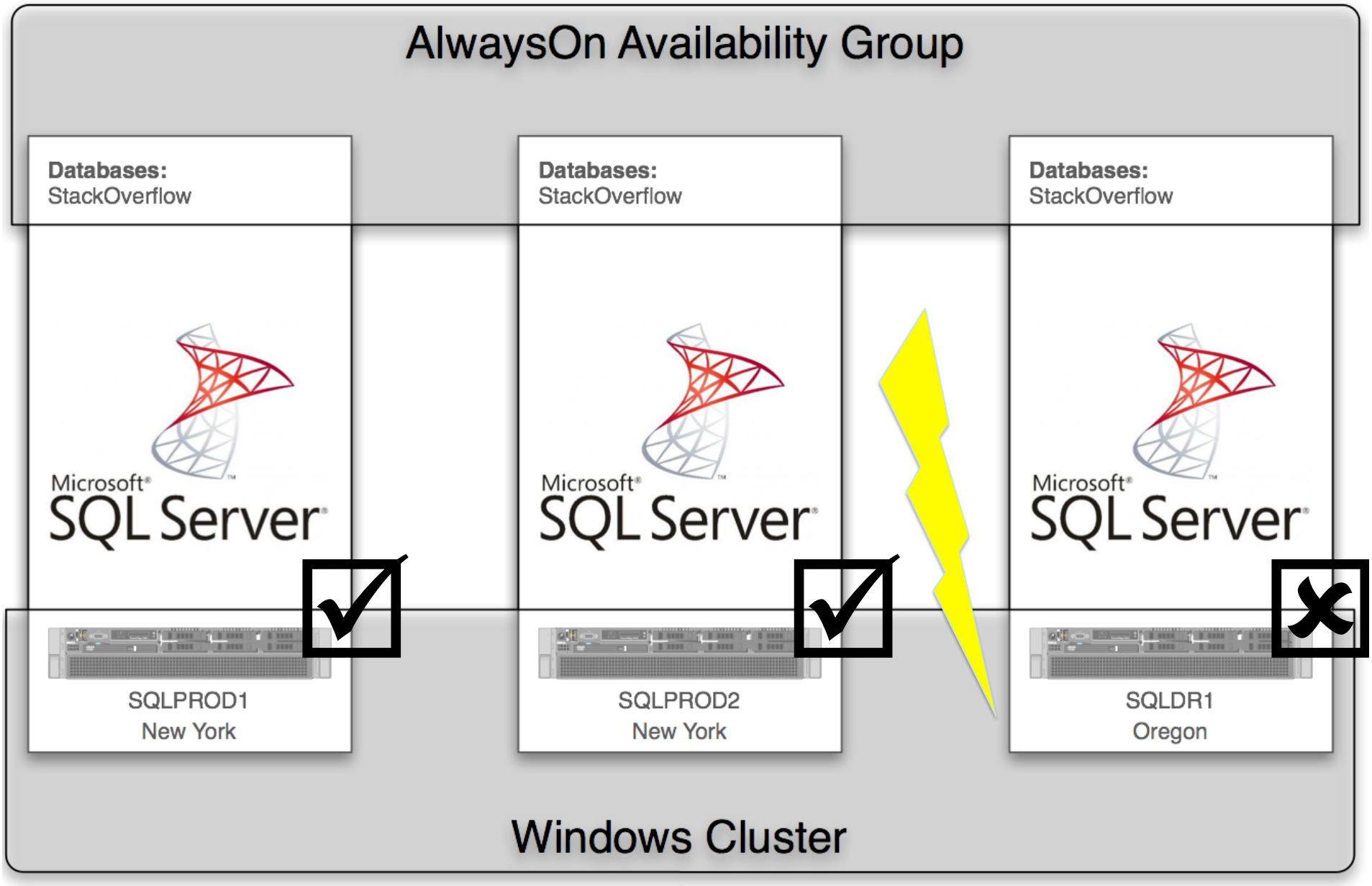
Differences in Windows versions:

- 2008R2 – number of quorum voters required is calculated once, and Windows doesn't recalc
- 2012 – Windows recalculates it for you every few seconds

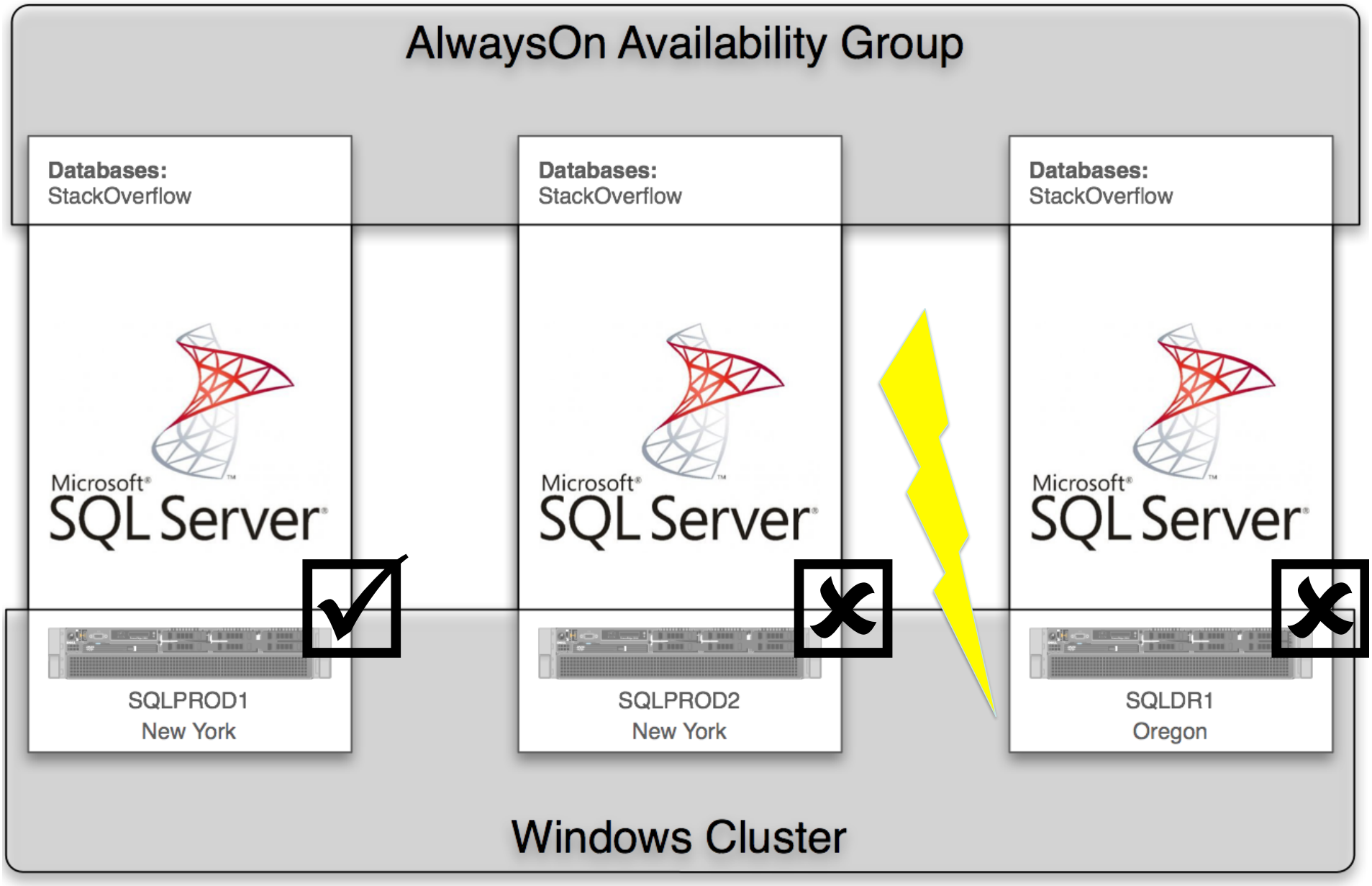
3 Voters, All Online



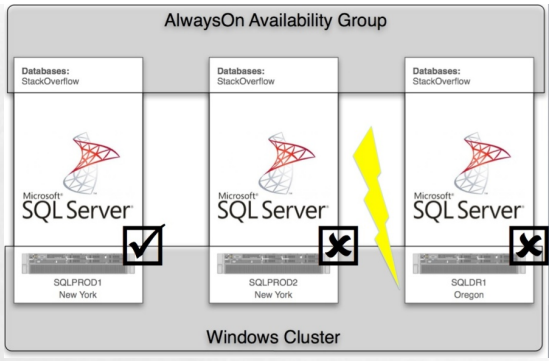
WAN Link Drops, 2/3 Voters Up



Server Rebooted, 1/3 Up



Depends on Windows



Win2012: last man standing stays up because quorum is recalculated every few seconds

Win2008R2: SQLPROD1 would go down because quorum isn't recalculated. 2 votes of 3 are required to stay up, and he's only got 1 vote out of 3.

Plan your voters carefully

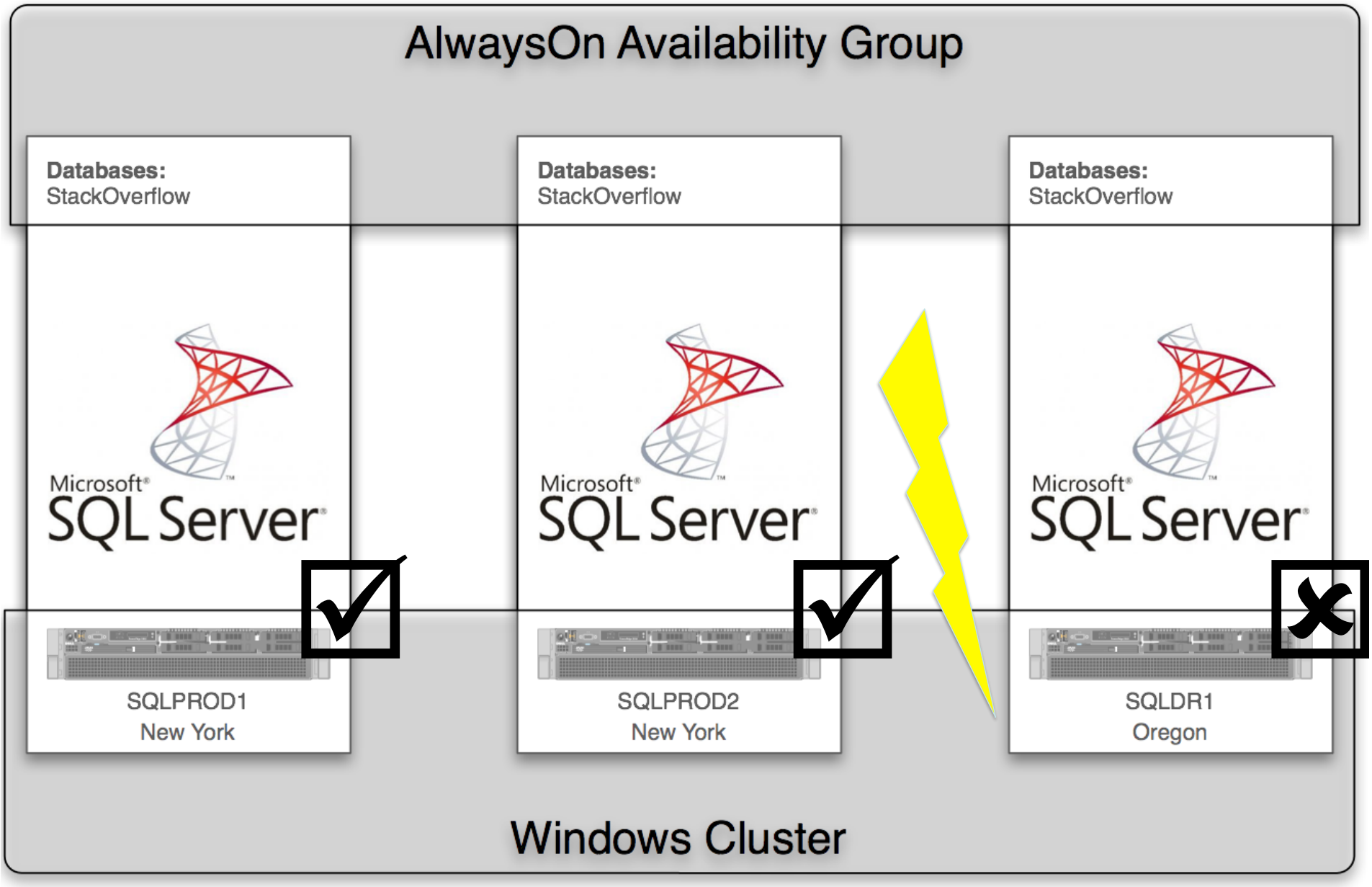
A file share or a shared disk can be a voter too

Nodes can have zero votes
(requires Windows 2008 hotfix KB2494036)

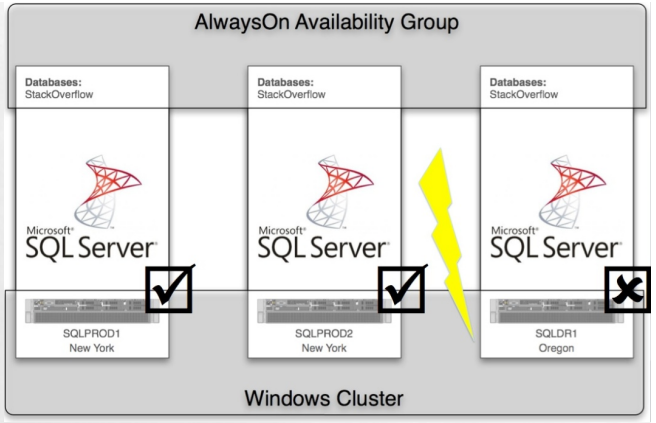
Know how many network links, switches, and nodes can be down before the cluster dies

Plan for datacenter failure with minority of votes remaining

Alternate scenario: NY link dies



Network complications



New York's network connection fails, severing them from the world

Oregon is still connected to the world, but it had a minority of votes, so SQL went down

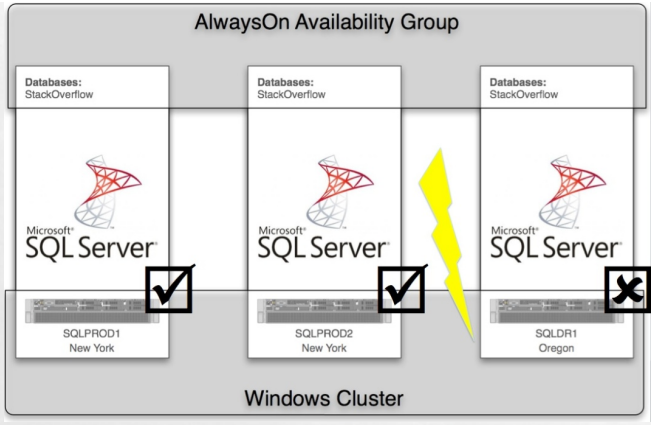
We want to force Oregon online – but New York is still online by itself too

We might have running processes in New York still adding data to the database there

If we force Oregon online, now we have a two-headed cluster: both sets of databases are up, and they have different transaction logs.

Where are backups running? Now need both.

The aftermath



The instant the network comes back online, web servers in Oregon will start getting packets through to New York – whose SQL Servers are still online.

Worst case scenario: both sets of SQL Servers have live data, and our database has parent/child tables with identity fields. Conflicts everywhere.

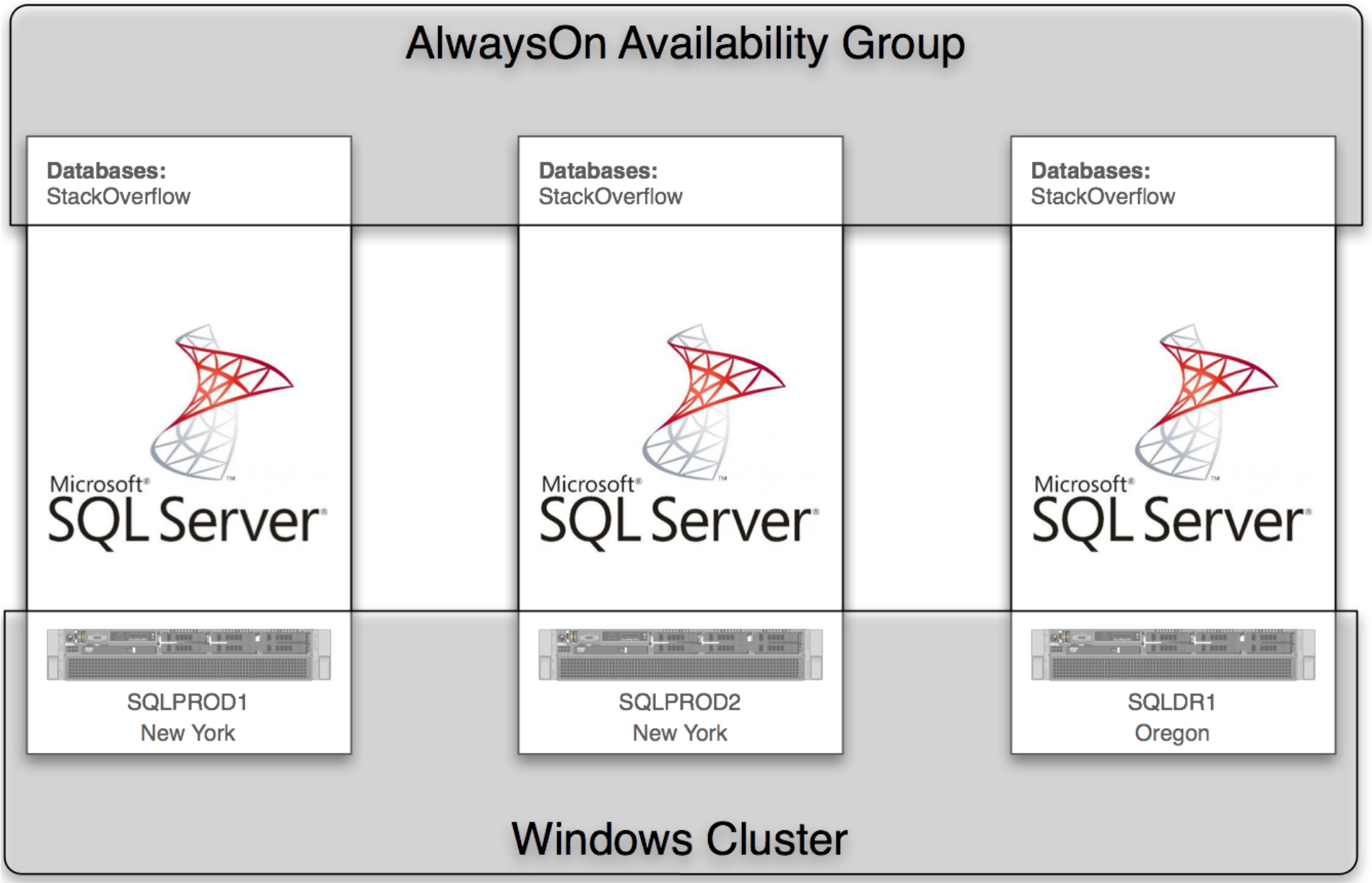
More Setup & Planning Lessons

Validate the bejeezus out of the cluster

Learn PowerShell basics, GUI doesn't cover 100%

Lessons Learned: Backups & Availability

Where Do You Run & Write Backups?



AlwaysOn Availability Group AskUbuntu

Databases: (PRIMARY)
AskUbuntu

Databases:
AskUbuntu

Databases:
AskUbuntu

AlwaysOn Availability Group SuperUser

Databases: (PRIMARY)
SuperUser

Databases:
SuperUser

Databases:
SuperUser

AlwaysOn Availability Group OtherSites

Databases:
(many small databases)

Databases: (PRIMARY)
(many small databases)

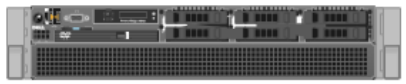
Databases:
(many small databases)

SQLPROD1

Chicago, Illinois

SQLPROD2

Chicago, Illinois



SQLDR1

Portland, Oregon

Backup Considerations

Where will you run backup jobs?

Where will you write the backups?

Do you need both onsite and offsite backups?

How do you plan to do restores?

Do you need to restore any databases regularly?

How will you track backup success?

Backups Working?

Set one server as the preferred replica for backups

DBA disabled jobs on that server to do maintenance

DBA forgot to enable the jobs again

Result:

- No job failure alarms
- T-log backup jobs ran successfully on non-preferred replicas (but didn't know there was a problem)
- No transaction log backups
- Transaction log grew to 2x the database size

More Backup & HA Lessons

Got SQL Agent jobs that run stored procs?
Put them on a separate job server pointed at the listener

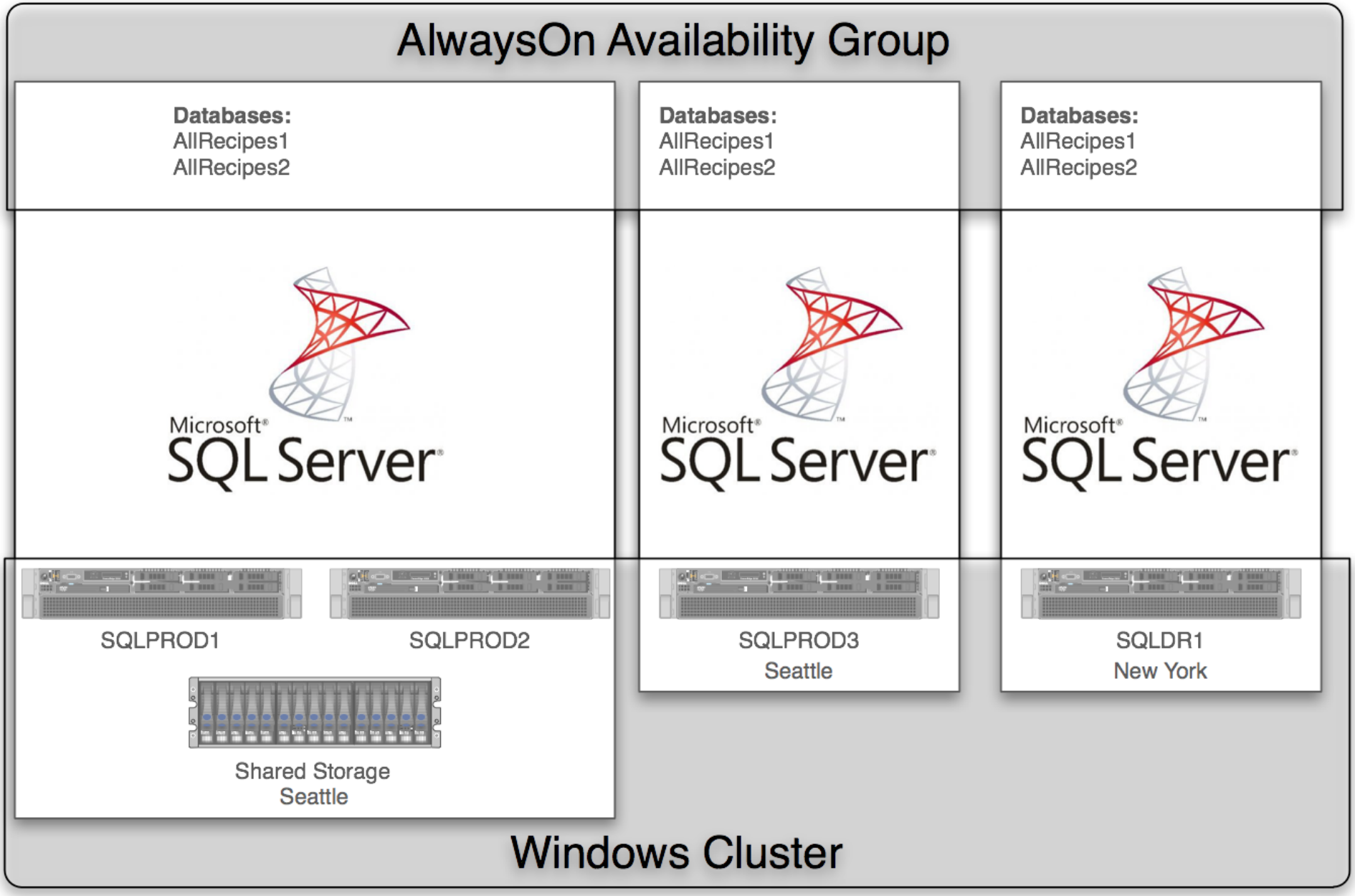
Using Ola.Hallengren.com's maintenance scripts?
Understand the copy_only implications.

Test a forced failover and fail-back before you go live.
Script it, and agree on who's allowed to perform it.

Plan for Windows Server 2012 upgrades:
a new cluster is required

Lessons Learned: Monitoring & Tuning

What do we monitor?



Our monitoring needs

Is the listener accepting writeable connections?

Is the listener routing read-only requests to other servers?

Are all of the read-only replicas up and running?

Is load distributed between replicas?

Are any of the replicas running behind?

Nothing is really ready for this complexity:
we're still building best practices for processes.

For tuning, each replica has its own:

DMV data

- Index usage
- Plan cache

Statistics

Query plans

Hardware

Different databases

And combining & interpreting them is up to you.

5 Key Lessons for Availability Groups

1. Use Windows Server 2012, not 2008R2
2. Patch management is a new important duty
3. Understand your quorum config's side effects
4. Monitoring, tuning require disciplined processes
5. Some parts make your job easier, but overall, you're going to be doing more work on more technologies: clustering, networking, PowerShell

Learn more at BrentOzar.com/go/alwayson

Closing Remarks

- Please attend any of our Performance Management Product Demo's
  - Invites will be sent to all attendee's
- Download a Free Trial of any of our Award Winning Solutions
  - **http://www.quest.com/performance-monitoring/database-performance-monitoring.aspx**

Thank you