

## Continuous Availability: A Scenario Validation Approach

Aniket Malatpure Ningyu He Microsoft

### Agenda

- Continuous Availability
- Cluster-in-a-Box
- Test Scenarios
- Validation Toolkit
- Key Takeaways



# **Continuous Availability (CA)**

#### Continuous Availability

- Transparent failover of application data storage
- Application sees contained IO delay

#### Value propositions

- Servicing without downtime
- Reliable, low-cost, file-based storage

#### Deployment Scenarios

- Server application storage platform
- File Server consolidation
- Virtual Desktop Infrastructure

#### Deployment Variations

- Multiple customer segments (small businesses to enterprises & hosted clouds)
- Multiple networking configurations (Ethernet, Infiniband, RDMA etc.)
- Multiple storage options (JBOD, RAID, SAS, SATA, FC, IP SAN etc.)



# Cluster-in-a-Box (CiB)

- Volume platform for Continuous Availability
- Design Considerations
  - At least one node and storage always available, despite failure or replacement of any component
  - Dual power domains
  - Internal interconnect between nodes, controllers
  - Flexible PCIe slot for LAN options
  - External SAS ports for JBOD expansion
  - Office-level power and acoustics for entry-level NAS



Additional JBODs ...



# What are the Requirements ?

#### GOALS ARE FUNCTIONALITY AND CUSTOMER EXPERIENCE, NOT SPECIFIC TECHNOLOGY OR IMPLEMENTATIONS

Requirement Area	Goals	Requirement Summary
RAID	<ul> <li>Survive single-component failures</li> <li>Reliable handling of node failure by storage systems without data loss</li> </ul>	<ul> <li>RAID levels 1, 5, 6 or 10 (or equivalent)</li> <li>RAID "write hole" solution</li> </ul>
Windows Failover Clustering	Needed to survive node failure	<ul> <li>SPC-3 Persistent Reservation support</li> <li>Shared LUNs accessible from all nodes</li> </ul>
LUN access after failover	Transparent failover at the application layer requires • Data Integrity • Prevention of read/write timeouts	<ul> <li>Preserve data for all acknowledged writes</li> <li>5 seconds maximum blocking time for up to 100 LUNs</li> <li>All I/O requests must complete within 25s</li> </ul>
Minimum active mode	Customer expectation is active use of ALL nodes	Dual-Active (non-shared LUN access)
Recovery Processing	Transparent failover should not require IT admin attention to continue normal operation	Recovery processing resulting from node failure completes automatically
Performance	Customer performance expectations must be met with Windows in Write-Through cache mode	System builders are permitted to publish performance measurements for the device in clustered configurations



# Test Scenario Design



# **E2E Test Scenarios to Validate CA**

#### CA HARDWARE REQUIREMENT

- Define scenarios which validate compliance
- How to generate IO load? How long? How much?
- Which faults to trigger? How?
- What is the success criteria?

#### CA TEST SELECTION

- Does it Validate data integrity and inconsistency?
- Perform variety of IO types
- Does it monitor IO timeouts and errors?
- Represent relevant synthetic workload

#### **E2E CA Scenario**

#### CA FAULT SELECTION

- Partner feedback? Relevant to requirement?
- Which faults cause delays in storage recovery?
- What are the top 10 faults seen in storage?
- Storage controller WBC state? Cascading failures?

#### **MEASURE SUCCESS CRITERIA**

- Impact on data availability?
- Did the failover happen? Resources available?
- Storage recovery within time bounds to deliver CA?
- LUN recovery automatic? Within 5 seconds?



# **Test Selection for CA Validation**

#### CA.Test.SysCache

- Data Verification Logo Supplemental Test
- Variety of IOs (seq., async, mapped, buffered, truncation, file pointer move, scatter-gather, read/write attributes etc.). Ability to detect and halt after detecting corruption

#### CA.Test.RapidFile

- Data Verification Logo Supplemental Test
- Multithreaded async IOs to different regions of file
- Verify files for data corruption, IO errors or timeouts

#### CA.Test.SQLIOSim

- Generates IO Patterns similar to IO activity of SQL Server
- Load generator stressing the file system, Verification of data integrity
- Lazywriter, logwriter, checkpoint, grow/shrink, read-ahead, random access
- Publically available and also ships with SQL Server



## **Fault Selection for CA Validation**

#### Goal: Simulate failures that cause resources to move between cluster nodes

'Planned' Faults	'Unplanned' Faults
Cluster Move-Group OS-Reboot	Power Supply Failures OS-Kernel Crash HBA Hang or Reset Cascading HBA Failures
Online RAID Migration	Failures during RAID migration
Dual Active Planned FailOver	Failures during Reconstruction Failures during Capacity Expansion
Physical Drive Hot Swap	Physical Drive Link Failure Physical Drive Failure



### How do the tests map to CA Requirements?

<b>Requirement Area</b>	Requirement Summary	Test Scenarios / Workflows
RAID	<ul> <li>RAID levels 1, 5, 6 or 10 (or equivalent)</li> <li>RAID "write hole" solution</li> </ul>	<ul> <li>Physical Drive Failure</li> <li>Hot swap of Physical Drive</li> <li>Online RAID Migration</li> <li>Online Capacity Expansion</li> </ul>
Windows Failover Clustering	<ul> <li>SPC-3 Persistent Reservation support</li> <li>Shared LUNs accessible from all nodes</li> </ul>	- FailoverClustering Cluster Validation
LUN access after failover	<ul> <li>Preserve data for all acknowledged writes</li> <li>5 seconds maximum blocking time for up to 100 LUNs</li> <li>All I/O requests must complete within 25 seconds</li> </ul>	<ul> <li>Unplanned OS Kernel Crash</li> <li>Physical Drive to Storage Link Failure</li> <li>HBA Power Failure</li> <li>Controller Hang or Reset</li> <li>Maximum LUNs Failover</li> <li>Cascading HBA Failures</li> <li>Planned Cluster Move-Group</li> <li>Planned OS Reboot</li> </ul>
Minimum active mode	Dual-Active (non-shared LUN access)	- Dual-Active Cluster Planned Failovers
Recovery Processing	Recovery processing resulting from node failure completes automatically	<ul> <li>Failover During Reconstruction</li> <li>Failover During RAID Migration</li> <li>Failover During Capacity Expansion</li> </ul>
Performance	System builders are permitted to publish performance measurements for the device in clustered configurations	- Report generation



# Validation Toolkit Overview



### **Key Features**

- Tests performed end to end with clients applying load to a clustered server
  - Runs multiple data integrity tests and synthetic I/O workloads from clients
  - Triggers planned and unplanned cluster failovers and verifies:
    - □ No loss of data and no data inconsistency observed from the clients
    - No impact on data availability (IO timeouts or failures) is observed
    - Volumes always accessible and LUN downtime is within the required bounds
    - □ Failover happens within time limit to deliver CA
- Fully automated software faults and extensible plugin framework to enable IHVs simulating hardware faults
- Support for WS2012 and WS2012r2 with iSCSI Target, SMB and NFS CA



### **Failover Time Measurements**



#### Total Failover Time (IO Brownout)

- Failover time as seen by the Client (SMB, NFS, iSCSI) application
- Measured as max IO latency from client during failover
- Tool: Canary
- COMPONENTS RESPONSIBLE: SMB, NFS, iSCSI, NTFS, Cluster, and Storage Controller

#### **Storage Recovery**

- Time to **recover storage** post cluster failover action
- Measured as time between first disk online call issued on surviving node and last disk coming online on surviving node
- Tool: LUNAccessTime.ps1
- COMPONENTS RESPONSIBLE: Cluster and Storage Controller

#### **Disk Online Latency (LUN Recovery)**

- **Time to online a volume** (LUN) on surviving node post cluster failover
- Measured as time between PR Arbitrate and volumes arriving on surviving node as seen by MountMgr & Cluster
- Tool: LUNAccessTime.ps I
- COMPONENTS RESPONSIBLE: Only Storage Controller



### **Test Framework Architecture**

SD @



### **Test Framework - Workflow**





### **Example Test Scenario – OS Kernel Crash**

<ul> <li>Review CA Requirement</li> <li>LUNs to remain accessible p failover</li> <li>Data integrity and prevention IO timeouts or errors</li> <li>5 seconds LUN online laten up to 100 LUNs or max support of the second s</li></ul>	<ul> <li>CA Test (IO Workload)</li> <li>Data integrity test, synthetic I/O workloads</li> <li>End to end tests with client applying load</li> <li>Tests with ability to detect corruption , inconsistency or timeouts</li> </ul>	<pre>************************************</pre>			
CA.LUNA	Measure Success Criteria Failover of resources is observed	<ul> <li>B0/26/2014-18:36:46 : Started LM downtime calculator</li> <li>B0/26/2014-18:38:48 : Started LM downtime calculator</li> <li>B0/26/2014-18:38:48 : Started LM downtime calculator</li> <li>B0/26/2014-18:38:48 : Canary:Spawned all the threads, waiting for all to complete within 300 seconds</li> <li>B0/26/2014-18:39:36 : Finished LUM downtime calculator. exit code 0</li> <li>B0/26/2014-18:39:56 : Finished LUM downtime calculator. exit code 0</li> <li>B0/26/2014-18:39:56 : Waiting for canary to exit</li> <li>B0/26/2014-18:30:57 : canary exits</li> <li>B0/26/2014-18:40:33 : All instances of CA.Test.SysCache SysCache Data Integrity Test Suite finished</li> <li>B0/26/2014-18:40:33 : TestInstance::PerformTestInstance - Waiting for end of iteration</li> <li>B0/26/2014-18:40:42 : Finished KernelCrash Node software failure via 0S kernel crash on active node. exit cod</li> <li>B0/26/2014-18:40:42 : Finished KernelCrash Node software failure via 0S kernel crash on active node. exit cod</li> <li>B0/26/2014-18:40:42 : Finished KernelCrash Node software failure via 0S kernel crash on active node. exit cod</li> <li>B0/26/2014-18:40:42 : Finished iteration index 1. Total to complete 1</li> <li>B0/26/2014-18:40:42 : Signal End of iteration</li> </ul>			
<ul> <li>OS kernel crash of active clunode</li> <li>No further interaction from the Storage Controller</li> </ul>	OS to Overall requirement result	Report: Requirement = LUMAccess Report: Test Scenario = KernelCrash Node software failure via OS kernel crash Report: Criteria - Failover Ubserved = Yes, Failover Expected = Yes Report: Criteria - Max LUN Downtime = 0.581 seconds, pass Report: Criteria - Data Availability Impact = Pass Report: Overall Result = Pass ***********************************			

#### Report

iSCSI Target Resource	Group		$\bigcirc$				$\bigcirc$		$\bigcirc$	
CA Hardware	OverallResult	Failure Action	CA-Criteria-01		Failover Time Seen by	CA-Criteria-02		CA-Criteria-03		
Requirement	(4)		Result	Failover Expected	Detected	(Worst)	Result	LUN Downtime (Worst)	Result	Impact on Availability
LUNAccess	PASS	Node software failure via OS kernel crash	PASS	true	1 of 1	10.80	PASS	0.58	PASS	5 of 5

SD @













# **Key Takeaways**

- Validate Continuous Availability by using customer scenario-focused testing
- Create validation requirements based on application expectations in customer deployments
- Simulate both 'Planned' and 'Unplanned' faults during typical application usage patterns
- Measure success based on transparent failure recovery during application usage (rather than implementation specific details)



### Q & A



# Appendix



### Abstract

CONTINUOUS AVAILABILITY: A SCENARIO VALIDATION APPROACH

#### □ Abstract:

Systems designed for 'Continuous Availability' functionality need to satisfy strict failure resiliency requirements from scenario, performance and reliability perspective. Such systems normally incorporate a wide variety of hardware-software combinations to perform transparent failover and accomplish continuous availability for end applications. Building a common validation strategy for diversified software and hardware solution mix needs focus on the end-user scenarios for which customers would deploy these systems. We developed the 'Cluster In a Box' toolkit to validate such 'Continuous Availability' compliant systems. In this presentation, we examine the test strategy behind this validation. We focus on end-to-end scenarios, discuss different user workloads, potential fault inducers and the resiliency criteria that has to be met in the above deployment environment.

#### Learning Objectives

- **1**. Continuous Availability and Transparent Failover
- **2**. End-to-end scenario testing strategy
- □ 3. User workload simulation
- **4**. Environment fault injection
- **5**. System resiliency SLA/criteria measurement



## **Test Configuration File Example**

1	)vml version_"1 0" encoding_"utf 2")
2 0	TANE VERSION IN COURSE OF STATES
	- Cluther Config
	(onfig Name="ClusterName", Value="CRCCU12" />
-	
2	<pre><comtig <="" pre="" value="20-1305c00000,20-1305c000000" walle="wolevalles"></comtig></pre>
6	<pre><contig name="Resourcedroup" value="CBCCU12-15CS1"></contig> </pre>
/	<pre><contig name="Resourcedroup" value="CBCCLU12-F5"></contig> </pre>
8	<contig name="ResourceGroup" value="CBCCLU12-S0FS"></contig>
9	
10 -	<testconfig></testconfig>
11 -	<test <="" id="CA.Test.SysCache" th=""></test>
12	Name="SysCache Data Integrity Test Suite"
13	Command="readwrit.exe -S -N -B -c -v 0"
14	Active="Yes"
15	WarmUpTimeInSeconds="15"
16	RunTimeInMinutes="30" />
17	<test <="" id="CA.Test.RapidFile" th=""></test>
18	Name="Rapid File Data Availability Test Suite"
19	Command="rapidfile.exe"
20	Active="Yes"
21	WarmUpTimeInSeconds="15"
22	RunTimeInMinutes="30" />
23 📄	<test <="" id="CA.Test.SQLIOSim" th=""></test>
24	Name="SQL IO Simulation Test Suite"
25	Command="sqliosim.com -cfg sqliosim.default.cfg.ini -size 1"
26	Active="No"
27	WarmUpTimeInSeconds="15"
28	RunTimeInMinutes="30" />
29	
30 E	<failurescenarios failureactionstore="FailureActions"></failurescenarios>
31 🖻	<failurescenario <="" id="KernelCrash" th=""></failurescenario>
32	Requirement="LUNAccess"
33	Type="Software"
34	Name="Node software failure via OS kernel crash"
35	FailoverExpected="Yes"
36	Iterations="1">
37 🖻	<preaction <="" command="Identifies and performs kernel crash failure on the active cluster node" th=""></preaction>
38	WaitPeriodInSeconds="10" />
39 🚊	<pre><failureaction 300"="" command="FailOverSimulator.ps1 -FailType CrashMachine -ResourceGroup \$ResourceGroup\$&lt;/pre&gt;&lt;/th&gt;&lt;/tr&gt;&lt;tr&gt;&lt;th&gt;40&lt;/th&gt;&lt;th&gt;TimeoutInSeconds="></failureaction></pre>
41 🚊	<postaction <="" command="" th=""></postaction>
42	WaitPeriodInSeconds="10" />
43	
44	
45	/THVSuite>

#### □ Supported Resource Group

- □ Singleton File Server + SMB/NFS shares
- □ Scale-out File Server + CSV + SMB shares
- iSCSI Target

#### Multiple IO Workloads

- SysCache
- RapidFile
- SQLIOSim

#### Pluggable Fault Generator

- Automated fault action
- Manual fault action

# **Fault Action Configuration**

33 🛓	<failurescenarios failureactionstore="FailureActions"></failurescenarios>	
34 E	<failurescenario <="" id="KernelCrash" td=""><td></td></failurescenario>	
35	Requirement="LUNAccess"	
36	Type="Software"	
37	Name="Node software failure via OS kernel crash"	
38	FailoverExpected="Yes"	
39	Iterations="1">	
40 🖻	<preaction <="" command="Identifies and performs kernel crash failure on the active cluster node" td=""><td></td></preaction>	
41	WaitPeriodInSeconds="10" />	
42 📮	<failureaction <="" command="FailOverSimulator.ps1 -FailType CrashMachine -ResourceGroup \$ResourceGroup\$ " td=""><td></td></failureaction>	
43	TimeoutInSeconds="300" />	
44 E	<postaction <="" command="" td=""><td></td></postaction>	
45	WaitPeriodInSeconds="10" />	
46		
47 🚊	<failurescenario <="" id="HBA-Reset" td=""><td></td></failurescenario>	
48	Requirement="LUNAccess"	
49	Type="Hardware"	
50	Name="HBA controller reset or hung"	
51	FailoverExpected="Yes"	_
52	Iterations="1">	
53 🚊	<pre><preaction <="" command="" pre=""></preaction></pre>	
54	WaitPeriodInSeconds="10" />	
55 🚊	<failureaction <="" command="ccu.exe cli controller local lockup this_controller " td=""><td></td></failureaction>	
56	TimeoutInSeconds="1000" />	
57 🚊	<postaction <="" command="" td=""><td></td></postaction>	
58	WaitPeriodInSeconds="300" />	
59		
60 🚊	<failurescenario <="" id="DriveHotSwap" td=""><td></td></failurescenario>	
61	Requirement="RAID"	
62	Type="Hardware"	
63	Name="Single physical drive hot swap in RAID set"	
64	FailoverExpected="No"	
65	Iterations="2">	
66 🚊	<preaction <="" command="" td=""><td></td></preaction>	
67	WaitPeriodInSeconds="0" />	
68 🚊	<failureaction <="" command="Please perform single physical drive Hot Swap in the RAID set\n&lt;/td&gt;&lt;td&gt;&lt;/td&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;69&lt;/td&gt;&lt;td&gt;Refer to user guide for more detailed steps" td=""><td></td></failureaction>	
70	<pre>TimeoutInSeconds="-1" /&gt;</pre>	
71 🗐	<postaction <="" command="" td=""><td></td></postaction>	
72	WaitPeriodInSeconds="300" />	
73		
74		

#### PreAction, FailureAction & PostAction

- Generic Commands
- Plain text, displays message box
- □ Can execute .cmd, .ps1, .exe, .\*

#### Inbuilt Software Fault Simulator:

#### FailOverSimulator.ps1

 Software failure simulator capable of triggering move resource group, reboot node, crash node

#### IHV Example:

- Model enables IHVs to integrate their automation to trigger h/w faults
- ccu.exe cli controller local lockup this\_controller

#### Manual Example:

- Plain text for the FailureAction Command
- Manually trigger the failure action
- Close the message box



# **Cluster Log - LUN Recovery Time**



SD (14