Cameron McAllister

IBM SAN Volume Controller Standby Storage Engine

This IBM® Redpaper™ describes the procedure for making use of an IBM SAN Volume Controller (SVC) standby storage engine in a SAN which is not a part of any cluster, but is a candidate and able to be quickly switched in to production when needed. The purpose of this is to reduce the impact of having an SVC Storage Engine in a cluster go offline for an extended period.

This paper is aimed at experienced system administrators who understand SVC/Storwize thoroughly, its jargon and terminology, and have experience of manually replacing nodes.

This paper applies to all SVCs that are capable of running version 6.3.0 or higher, namely:

- 2145-DH8
- 2145-CG8
- 2145-CF8
- 2145-8A4
- 2145-8G4

If you are unsure about any of this, then speak to your IBM representative.

Overview

This procedure is intended to allow a spare SVC standby node to exist on the storage area network (SAN) such that it can be swapped in to the cluster should any node fail for one reason or another.

The reason for wanting to do this is to minimise the time that a cluster is exposed to only having a single node in the IOgroup. If a node fails due to a hardware issue, waiting for spare parts could possibly take a long time. If a spare node is ready and waiting, it can be quickly switched in to production when needed. The node that is removed can then be repaired in a more leisurely fashion, and once repaired can become the new standby node.

Provided all of the requirements in , “Requirements” are met, the precise ratio of standby nodes to SVC clusters does not need to be fixed at 1:1, although with more standby nodes and clusters, careful management of which standby nodes can be used with which clusters will be needed.

Requirements

All of these requirements must be met:

- Service IP access to all the nodes in the cluster
- An additional node not part of any cluster (also known as the standby node) which must be:
  - The same node type as the nodes in the cluster
    - The standby node should also have the same hardware feature configuration as the nodes in the cluster.
    - Switching between different node types will require additional steps. Check with the node replacement procedures for the particular node types to understand it. Refer to this for more information:

      https://ibm.biz/BdEquP
  - Directly connected to the same switches as the other nodes in the cluster
    - Connected to the same SAN, VSAN or LSAN
    - Connected to the same physical ethernet and VLAN networks.
  - Can be logged into remotely to query it via its service IP
  - It must be capable of running the code level that is on the cluster
    - To speed up the node addition, it should be kept in sync with the cluster code level outside of this procedure.
  - The node should be powered up at all times waiting to be used.
- Remote access into the SAN switches that the cluster (and standby) nodes are connected to - so that port changes can be made
- A mechanism for getting the WWPNs from a node that is currently offline
  - Such as a well defined, up-to-date, and documented naming convention, or by querying the Switch
- A node naming structure that will not be compromised by adding and removing nodes.
- A temporary WWNN that can be used on the fabric (makes the switching procedure easier) – such as “**FFFFFFFFFFFFFFFF**”

- The zoning for the cluster is implemented using WWPNs, not ports.

Notes:

- These instructions do not cover systems with iSCSI configurations present – additional work will be needed – based on the node replacement procedure in the infocenter.
- No additional steps are required if Remote Copy or IP replication functions are enabled.
- Stretched systems (clusters) will require standby nodes at each site.

Risks

This procedure is very low risk provided that it is carried out correctly.

If the WWNN changing is not done correctly, and results in duplicate WWNNs, both nodes with the same WWNN will assert and turn off because this violates the Fibre Channel Specification.

A node that is not connected to a cluster will not call home in event of a hardware failure. If the standby nodes are not explicitly monitored, they could have failed silently and not be available when needed. To mitigate this risk, regular explicit checks of any spare nodes should be performed and any errors fixed.

Systems running AIX® will need to perform a rescan of the Fibre Channel paths to pick up paths that have moved to the standby node. This will need to be coordinated by the administrators.

Automating this Procedure

Due to the complexity and repeatability of these steps, it is attractive to automate this procedure with scripts set up for your specific SVC cluster and SAN configurations.

Although the creation of an automated procedure is the responsibility of the end user, SVC Development has the following recommendations:

- The procedure is thoroughly tested on a test cluster before it is implemented on a production system.
- The checks to perform are also automated, and the script will not proceed if any checks fail.
- The script is always triggered manually – there should always be an administrator overseeing this, and actively making the decision to switch a standby node into the cluster.

Checks to perform before switching nodes

These checks *must* be performed in advance of running this procedure. If *any* of these fail, the procedure should not be run. Make sure that:

- The cluster is not performing an upgrade.
- All of the ports connected to the node that has failed on the SAN switches can be found or discovered.
- The standby node is in a fit state to be used:

- It has WWPNs that are appropriate to be used (not duplicates to any other on the SAN)
- It is powered up
- The spare node WWPNs and “Temporary WWPNs” are zoned into the cluster – but they do not have to be zoned into the back end.
- It appears as a candidate node to the cluster
- It is running the same code level as the cluster.

- The node that is being replaced is not the only node in the IOgroup
- The node that is being replaced has no volumes (vdisks) dependent on it:

  Run the command:

  `lsdependentvdisks -node node_id | node_name`

  and make sure no volumes are listed. If volumes are listed then removing this node will cause an IO outage. You will need to correct this before proceding.

  - If the node being removed has Ethernet ports configured for iSCSI then additional work will be needed. Refer to this for more information:

    https://ibm.biz/BdEquP

- Check for the iSCSI configuration

  Run the command:

  `lsportip -filtervalue node_id=<node id>`

  if any ports are not “unconfigured” additional work will be needed to prevent duplicate IP addresses on the network.

Procedure details

In the sections that follow we provide the sequential steps that must be taken:

1. Store the WWNNs
   a. Get the WWNNs of the outgoing node directly from the node using the command:

      `$ svcinfo lsnode 1`

      look for the line starting with **`WWNN`** e.g.:

      `WWNN 500507680105597E`

      this will be referred to as the “outgoing node WWNN”

   b. Get the WWNNs of the standby node using the command:

      `$ sainfo lsservicestatus`

      Look for the line starting with **`node_WWNN`** e.g.:

      `node_WWNN 5005076801xxxxxx`

   c. Get the temporary WWNN (from wherever you have it recorded) and make sure nothing on the fabric is using this WWNN
2. Get the IOgroup of the outgoing node so that we know which iogrp we are adding the node to later using the command:

   `svcinfo lsnode <outgoing node ID / name>`

   look for the line starting with **`IO_group_id`** e.g.:

   `IO_group_id 1`

3. Remove the outgoing node from the cluster using the SVC command:

   svctask rmnode <node id>

   wait until the node is no longer a member of the cluster and poll the view using the command:

   svcinfo lsnode

   until the removed node has disappeared from the view

   Note: If the node was the active config node – there may be a delay while the cluster IP fails over.

4. If the outgoing node is running and can be accessed via its service interface, change its WWNN to the temporary one using the service CLI:

   For SAN Volume Controller models 2145-DH8 use:

   satask chvpd –wwnn <temporary WWNN>

   For SAN Volume Controller models 2145-CG8, 2145-CF8, 2145-8A4 and 2145-8G4 use:

   satask chwwnn -wwnnsuffix <temporary WWNN suffix>

   Note that the <temporary WWNN suffix> *is the last 5 digits of the temporary WWNN*

   Wait 30 seconds for the node to warmstart and verify that its WWNN is now set to the temporary one. (if this check fails then you must stop immediately, retry, and contact IBM Support if it continually fails)

   sainfo lsservicestatus

   Look for the line starting with **node_WWNN** and verify that it is set to the <temporary WWNN>

5. If step 4 could not be run because the node is inaccessible:

   Record somewhere safe that the WWNNs will need to be changed to the spare WWNN manually (by the front panel or the service interface) before it can be reconnected to the fabric.

   Disable the ports on the switch that the offline node is connected to in order to isolate it from the rest of the fabric (so that when it is brought online with the same WWNN as the node that replaces it, it can be serviced, and its WWNN changed without causing a duplicate WWNN error)

   Note: When disabling ports, make sure that the right number of ports are found and disabled. If the right number is not found, this offline node is not safe to be powered up.

6. Change the WWNN of the incoming standby node to look like the outgoing node using the service CLI:

   For SAN Volume Controller models 2145-DH8 use:

   satask chvpd –wwnn <outgoing node WWNN >

   For SAN Volume Controller models 2145-CG8, 2145-CF8, 2145-8A4 and 2145-8G4 use:

   satask chwwnn -wwnnsuffix <outgoing node WWNN suffix>

   Note that the <outgoing node suffix> is the last 5 digits of the WWNN

   Wait 30 seconds for the node to warmstart and verify that its WWNN is now the spare one. (if this check fails then you must stop immediately)

   sainfo lsservicestatus

look for the line starting with “**node_WWNN**” and verify that it is set to the WWNN of the <outgoing node WWNN>

7. Add the incoming node to the cluster in the IOgroup that the outgoing node was in using the CLI:

   Svctask addnode –wwnodename <outgoing node WWNN > -iogrp <iogrp ID>

   If you use a specific naming convention for node names, make sure that you conform to it here, wait until the node has fully joined the cluster, and then use the command:

   svcinfo lsnode

   Wait until the “**status**” for the node is “**online**”

8. If the outgoing node is running and can be accessed via its service interface, change its WWNN to the standby one using the service CLI.

   For SAN Volume Controller models 2145-DH8 use:

   satask chvpd –wwnn <standby WWNN>

   For SAN Volume Controller models 2145-CG8, 2145-CF8, 2145-8A4 and 2145-8G4 use:

   satask chwwnn -wwnnsuffix <standby WWNN suffix>

   Note that the <outgoing node suffix> is the last 5 digits of the WWNN

   Wait 30 seconds for the node to warmstart, and verify that its WWNN is now set to the temporary one. (if this check fails then you must stop immediately)

   sainfo lsservicestatus

   Look for the line starting with “node_WWNN” and verify that it is set to the <standby WWNN>

This procedure is now complete, and the outgoing node is now safely removed from the cluster and can be repaired at your leisure.

Author

This paper was produced by the International Technical Support Organization, at Hursley Labs, UK.

Cameron McAllister joined IBM in 2002 as a developer in IBM Software Group. He moved to Storage in 2005 and spent six years working on SVC culminating in the release of the Storwize V7000 before moving onto Storwize V7000 Unified as Test Architect in 2010. Now he has expanded his role further as an architect of Elastic Storage, but always keeps his ties with SVC and Storwize V7000 close, and is the Lab Advocate for several SVC customers including Sprint.

Special thanks go to Dave Lounsberry from Sprint Corporation who performed most of the testing and refining of this procedure.

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your

network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks
- Follow us on Twitter:

  http://twitter.com/ibmredbooks
- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806
- Explore new Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm
- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document REDP-5162-00 was created or updated on December 19, 2014.

Send us your comments in one of the following ways:
- Use the online **Contact us** review Redbooks form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbooks@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

Trademarks