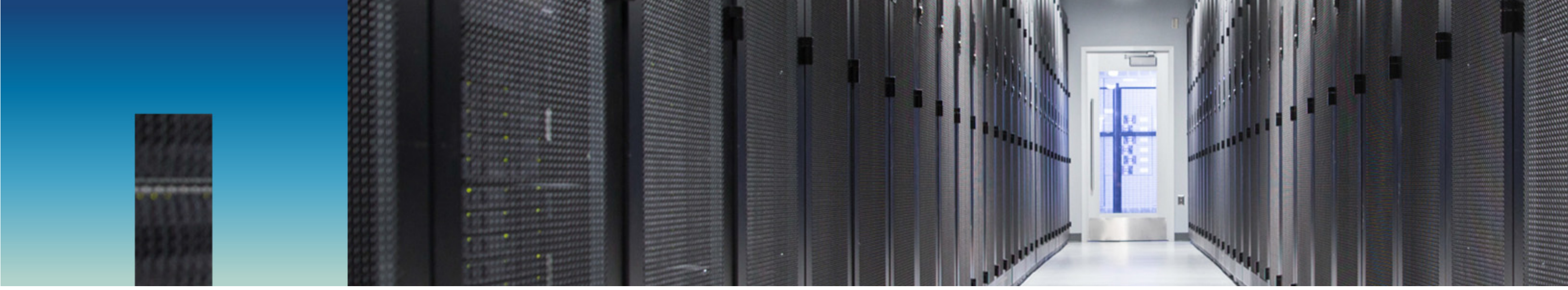


Technical Report

Best Practices for Networking and Network Maintenance on NetApp SolidFire Storage Systems

Goutham Baru, NetApp

April 2019 | TR-4763

Abstract

Best practices to help provide a reliable, high-performance SAN deployment.

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

This report provides guidelines for a reliable, high-performance SAN deployment. Some, or all, of the items mentioned might apply to your environment. You must consult your network and host systems vendor for detailed implementation and configuration guidelines that apply to your specific equipment.

2 General Network Recommendations

Consider implementing the following general network recommendations:

- A redundant, fault-tolerant network infrastructure.
- A dedicated 10Gb storage network.
- A dedicated, or shared network for NetApp® SolidFire® storage node management and intracluster communications.
- End-to-end jumbo frames support within the switching infrastructure on all NetApp SolidFire storage nodes and all servers connected to the storage network.
- Redundant connectivity for all storage nodes.
- Redundant network interface card (NIC) connectivity for all server-side connections, using NIC switch fault-tolerance or NIC bonding (where supported), for optimal end-to-end availability and enhanced performance.

2.1 Fully Redundant Network Infrastructure

You should implement a fully redundant, fault-tolerant network infrastructure for performance, uptime, and availability of the NetApp SolidFire storage system.

For higher availability, connect each individual management NIC (1GbE) and each individual storage NIC (10GbE) to separate redundant switches. If you are using a highly resilient switch chassis, connect to redundant switch modules.

By default, the management NICs and storage NICs use active-passive bonding because it requires no configuration on the switch side, while facilitating 10GbE bandwidth, and switch fault tolerance.

2.2 SFP+ Connectivity

NetApp SolidFire supports most industry standard SFP+ transceivers, either in the form of direct-attach cables (DAC) or optical (fiber) transceivers.

Although NetApp SolidFire supports third-party cables, you should follow your network switch vendor's recommendations for cable selection. Many vendors have restrictions on which cables they support.

2.3 Nonblocking Switch Architecture

It is optimal for the storage-side switches to have a nonblocking backplane architecture, and support full-duplex, line-rate forwarding on all ports. The backplane capacity should be greater than or equal to the following value:

```
((the  total  number of switch  ports  *  10  Gb/s)  *  2)
```

2.4 Switch Port Buffers

All the storage network switches should support a minimum of 512K packet buffers per port. Some switches, by design, share the buffer allocation either between a group of ports or among all ports on the switch. You should use a switch that implements per-port packet buffers for dedicated high-bandwidth storage networks.

If the switch does not support per-port packet buffers, try to configure adequate buffer space for the entire group of ports to minimize the possibility of packet loss.

2.5 Flow Control

Ethernet flow control must be implemented consistently across the network infrastructure to handle network oversubscriptions that can affect performance. Failure to do this can cause packet loss and degraded network throughput.

You must consider the following for consistent flow control:

- Verify that switches are configured to only receive pause frames (rx).
- Verify that flow control is enabled only on edge ports connected to NetApp SolidFire storage nodes and iSCSI hosts.
- Do not enable flow control within the network core or between Inter-Switch Links (ISLs). Enabling flow control can negatively affect all traffic, instead only throttle back certain hosts.
- You do not need flow control between switches because most switches forward at line-rate speeds. However, you should have adequate bandwidth between ISLs so that the links are not oversubscribed.

2.6 Jumbo Frames Support

Jumbo frames should be enabled end-to-end between hosts and storage nodes, and across all network paths on the storage network.

By default, all storage node 10GbE interfaces are enabled with an maximum transmission unit (MTU) of 9000. For optimal performance, all server-side storage interfaces should be configured with the same MTU as the NetApp SolidFire storage nodes.

Network switches should be configured to support an MTU of at least 9016 or more to account for jumbo frame overhead and for proper forwarding through the network. If you want to change this configuration to support a lower MTU setting, you should contact NetApp SolidFire support.

2.7 Storage Node Network Management

You should verify that the following network services are available through the 1GbE management interfaces and network for proper network functionality:

- DNS: Outbound access required for client lookups
- lSNMP: Inbound access to SNMP queries
- Syslog: Outbound access for logging to an external log server
- Secure Shell (SSH): Inbound access required for node-level management
- HTTPS: Inbound access required for cluster-level management and API communications
- Management node: NetApp SolidFire management node communications
- A virtual machine connected to a cluster of nodes that enables remote support, long-term performance and capacity statistics collection through NetApp Active IQ®, and the CLI.

2.8 Network Interface Configuration

You must consider the following network configuration requirements:

- Management 1GbE interfaces require an IP address, network mask, and gateway for proper operation of the storage network.
- Storage-side 10GbE interfaces require at least an IP address and network mask. Storage-side interfaces normally do not require any default gateway unless there is a requirement to access the storage node 10GbE interfaces from a network or subnet other than the storage network itself.
- Bond1G and Bond10G interfaces should be configured with IP addresses that reside on separate, distinct subnets, and that do not overlap. Configuring both interfaces with IP addresses out of the same IP subnet is not supported.

  Note: If you configure the node to use Bond10G for the management interface instead of the default Bond1G, you must disable Bond1G on all the nodes.

2.9 Considerations for Spanning-Tree Protocol

You must enable Portfast or similar functionality, or else disable Spanning Tree Protocol (STP) on all host and storage node ports.

Network topologies with multiple switches and multiple ISLs are vulnerable to loops, which can severely disrupt network traffic. STP was implemented decades ago but was slow during topology changes. Modern switch architectures are more likely to use rapid spanning tree.

When a switch comes up, or there is a topology change, an election occurs to determine which paths should be blocked and which paths should pass traffic. A topology change could be a switch port link that goes down or comes up. During the election process, traffic might be blocked until this completes.

Note: Storage nodes and clients cannot participate in this election. They must be configured as edge-ports for spanning tree.

2.10 Dedicated Management and Storage Networks

You should create separate network segments to isolate storage traffic from management traffic and any other network traffic.

You can implement separate networks either physically or logically using virtual LANs (VLANs).

2.11 Unicast Storm Control

Many switches misdiagnose iSCSI traffic as a packet storm of traffic with malicious intent and block this traffic unnecessarily.

Because the SAN should be isolated from all other general traffic, the possibility of malicious traffic is minor. Storage-node switches must pass Ethernet packets regardless of the traffic profile.

2.12 Gratuitous Address Resolution Protocol

You must enable gratuitous Address Resolution Protocol (ARP) on the switches connected to the storage side of NetApp Element® software. Gratuitous ARP is an ARP request where the source and target IP address on the IP packet are same. The gratuitous ARP request is sent as a broadcast. The gratuitous ARP packets are used to address IP address conflicts and discover the mandatory access control (MAC) address of the hosts.

Gratuitous ARP packets are used by NetApp Element software during the upgrade process.

2.13 More Network Recommendations

You should also consider the following network recommendations:

- Disable TCP delayed acknowledgment (ACK) because it is for low-bandwidth environments.
- Disable Large Receive Offload (LRO) on network cards because it is associated with TCP Offload Engine (TOE) and is known to cause issues with iSCSI traffic.

3 Best Practices for Common Network Maintenance Tasks

To perform network maintenance tasks, you must be an experienced user with full cluster administrator rights. Consider the following before performing any network maintenance task:

- Verify that the cluster has sufficient capacity to handle one storage node being down.
- Verify that there are no running tasks by clicking Reporting > Running Tasks from the Element software web UI.
- Check for any alerts on the cluster by clicking Reporting > Alerts from the Element software web UI.
- Work on one storage node at a time and wait for five to ten minutes before moving to the next storage node. This delay allows any outstanding jobs running in the background to complete. Check that no alerts were reported.
- (Optional) Confirm that every storage node is configured with a dedicated integrated Dell Remote Access Controller (iDRAC) port to provide remote access at any time. For details about iDRAC configuration, contact NetApp SolidFire support.

  Note: You cannot change the cluster management virtual IP address (MVIP address), storage virtual IP address (SVIP address), or cluster name at any time.

3.1 Removing a Storage Node

You can remove storage nodes from a cluster for maintenance or replacement. To remove storage nodes before taking them offline, you must use the Element software web UI, or an API command.

Prerequisites

- The cluster has sufficient capacity to safely permit a storage node removal.
- There are no existing alerts, verified by clicking Reporting > Alerts from the Element software web UI.

Procedure

1. Remove drives from the storage node that you are removing by doing the following:
   a. Click Drives > Active Drives and select all block drives for the storage node that you are removing.
   b. At the bottom of the page, click Remove.

      Data migrates off the drives. The time the process takes to complete depends on the amount of data being migrated, and how busy the cluster is.
   c. To monitor the progress, click Reporting > Running Tasks.

      Wait an extra five minutes for any outstanding jobs to complete.
   d. After the process is complete, click Drives > Available Drives to verify that the block drives you selected in Step 1 are listed.
2. To remove slice or metadata drives from the storage node that you are removing, repeat Step 1.

   Note: Your storage node can have multiple slice or metadata drives if it was configured as a Multiple Disk Slice Service (MDSS) storage node. If your storage node has multiple slice or metadata drives, remove both slice and metadata drives at the same time.

   Caution: Element software does not support removing a drive if it might result in an insufficient amount of storage to migrate data.

Note: If there is not enough capacity to remove active drives before removing a storage node, you see the `Not Enough Space` error message.

3. Remove the storage node from the cluster by doing the following:
    a. Click Nodes > Active Nodes and select the storage node you want to remove.
    b. Click Remove.

Note: If a storage node is offline for more than five and a half minutes, Element software determines that the storage node is not coming back to join the cluster and begins to write single replicated blocks to another available storage node in the cluster. The drives in the offline node are changed to Failed status. After the storage node is back online, the status of the drive's changes to Available.

Note: If you attempt to remove a storage node that has active drives, you see an error message indicating that there are registered drives with data that are preventing the node from being removed. You must remove all active drives for the storage node before it can be removed from the cluster.

3.2 Restarting a Storage Node

To restart a storage node, you can use the `Shutdown` API command by entering the following into a web browser:

https://<clusterMVIP>/json-rpc/8.0?method=Shutdown&nodes=[X]&option=restart

Note: X is the storage node ID. You can find the storage node ID from Nodes > Active Nodes in the Element software web UI.

For more information about the `Shutdown` API command, see the SolidFire API Reference Guide.

3.3 Turning off a Storage Node

Turning off storage nodes and clusters involves risks if not performed properly. To safely complete these procedures, you can use the `Shutdown` API command.

NetApp recommends that you use a third-party tool such as Postman or Windows PowerShell, and not the web browser, to run API calls. By doing so, you can avoid potential mistakes caused by cached commands in your web browser.

Procedure

To turn off a storage node, enter the following in a web browser:

https://<clusterMVIP>/json-rpc/8.0?method=Shutdown&nodes=[X]&option=halt

Note: X is the storage node ID. You can find the storage node ID from Nodes > Active Nodes.

For more information about the `Shutdown` API command, see the SolidFire API Reference Guide.

If a storage node has been offline for more than five and a half minutes, the Element software determines that the storage node is not coming back to the cluster. Double Helix™ data protection starts writing single replicated blocks to another location to re-replicate the data. For technical assistance, contact NetApp SolidFire support.

3.4 Turning on a Storage Node

If a storage node has lost communication, or is in an offline state, contact NetApp SolidFire support before you bring it back online. NetApp SolidFire support investigates why the storage node went offline and assists you with the necessary recovery steps.

After the storage node is back online, you must add the available drives to the cluster, depending on how long it was offline.

3.5 Considerations for Upgrading Switch Firmware

You must implement a fully redundant, fault-tolerant network infrastructure for performance, uptime, and availability of the SAN. For high-availability (HA), verify that each individual management NIC (1GbE), and each individual storage NIC (10GbE), is diversely connected to separate, redundant switches, or redundant switch modules, if using a highly resilient switch chassis.

- You must perform a switch upgrade procedure on only one switch at a time. Redundant interfaces on the nodes should be able to communicate with the second switch at any time.
- Before you upgrade the switch firmware, check the running configuration, and check that all routes are in place.
- Verify that your running configuration is saved to memory before restarting any switch.
- You do not have to perform any special actions on your storage nodes for upgrading switch firmware.

Note: You might see BlockServiceUnhealthy and SliceServiceUnhealthy alerts if your storage nodes are configured in active-passive mode.

3.6 Performing Switch Migration

If your cluster is in a fully redundant network, the switch migration procedure is not disruptive.

Caution: You must perform this procedure carefully, because your data becomes unavailable if multiple nodes lose connectivity among them while you are performing the procedure.

Prerequisites

- The cluster has no unresolved faults, verified by clicking Reporting > Alerts from the Element software web UI.
- All jobs are complete, verified by clicking Reporting > Running Tasks from the Element software web UI.
- All the storage nodes and switches are connected.
- The storage NICs (10GbE) are connected to two redundant storage switches, and the network NICs (1GbE) are also connected to two redundant management switches.
- The bonding method on the cluster or nodes and switches are the same.

Procedure

1. Unplug the cables from the ports on the switch or the storage nodes, considering the following:
   a. To maintain connectivity between the nodes in the cluster, verify that that one NIC (eth0 or eth1 for storage network (10GbE) and eth2 or eth3 for management (1GbE)) is connected to a redundant switch.
   b. Unplug cables from one node at a time and wait 30 seconds before proceeding to the next node.
2. Check the cluster alerts on the Reports > Alerts page for any unexpected alerts.

   Note: If you see any alerts, stop the switch migration procedure, and contact NetApp SolidFire support.
3. After the switch migration is complete, wait for five to ten minutes to confirm that there are no unresolved cluster faults on the Reports > Alerts page.

3.7 Changing VLAN Tags on a Switch or a Storage Node

The procedure to change VLAN tags can cause data unavailability issues. You must time the procedure to coincide with changes on the switches. To avoid a data unavailability issue, you should configure your switch ports to allow traffic from both the current and future VLAN tags.

Before Element software 9.0, you could not tag primary networks (1GbE or 10GbE) or route (layer 3) storage traffic to VLAN interfaces. The following recommendations are based on the Element software 9.0 release. For recommendations for previous releases, contact NetApp SolidFire support.

Prerequisites

- The cluster has no unresolved faults, verified by clicking Reporting > Alerts in the Element software web UI.
- All jobs are complete, verified by clicking Reporting > Running Tasks in the Element OS Web UI.
- All iSCSI client I/O is stopped and all iSCSI clients from the cluster are disconnected.
- All remote replicas are fully synchronized, verified by clicking Volumes > Replicating.

Procedure

1. Choose a nonensemble storage node, and on the Nodes > Active Nodes page, change the VLAN tag to the new value.
2. Verify that the change succeeded by accessing the storage node UI from the new VLAN (https://<NodeIP>:442). The change can take up to a minute to take effect.
3. Wait for five minutes.
4. Repeat Steps 1 to 3 for all nonensemble storage nodes.
5. Repeat Steps 1 to 3 for the remaining storage nodes.
6. Access the Element software web UI from the new VLAN and wait for all cluster faults to clear.
7. Add any Available drives back into the cluster.

3.8 Changing Ports, Shutting Down a Port or Interface, or Changing a Cable

You must work on only one port on the switch at a time. You must also confirm that at least one NIC, eth0 or eth1 (10GbE), for the storage network is up and running.

3.9 Considerations for Bonding Methods

NetApp SolidFire supports three bonding methods. They are adaptive load-balancing, active-passive, and Link Aggregation Control Protocol (LACP).

If your environment has only one switch, it is better to use adaptive load-balancing. If your environment has two nontrunked switches, it is better to use active-passive bonding.

If your environment has two trunked switches, where the switches appear as one, you must use LACP (802.3ad) bonding. The two switches must appear as one switch, either by being different switch blades that share a back-plane, or have software installed to make it appear as a stacked switch. The two ports on either switch must be in a LACP trunk to enable the failover from one port to the next to happen successfully. If you want to use LACP bonding, you must confirm that the switch ports between both switches are enabled for trunking at the specific port level.

If you have questions, contact NetApp SolidFire support.

3.10 Changing the Bonding Method and MTU Setting

You can change network settings, such as the subnet mask, bonding method, and MTU value from the storage node UI. Consider the following recommendations before making changes:

- Change one storage node at a time.
- Change the cluster master node only after you finish updating the storage nodes.
- Wait for the block data synchronization to complete between the updates. You can view the synchronization status by clicking Reporting > Running Tasks in the Element software web UI.

Procedure

1. Select the storage node to update.
2. Review the Element software web UI to confirm that there are no active alerts.
3. Update the storage node UI at https://<NodeIP>:442.
4. View the UI for any faults.
5. Wait for faults, if any, to clear.
6. Navigated to the System Utilities tab on the storage node UI, and restart services.
7. Verify that the change succeeded by accessing the storage node UI from the new VLAN (https://<NodeIP>:442). The change can take up to a minute to take effect.
8. Wait for faults, if any, to clear.
9. Repeat the above steps for each storage node, as needed.
10. Repeat the above steps for the cluster master node.

3.11 Changing the IP Address on a Storage Node

You can change the IP address of a storage node to a new IP address within the same range of IP addresses.

Prerequisites

You have removed the storage node from the cluster. See Removing a Storage Node.

Procedure

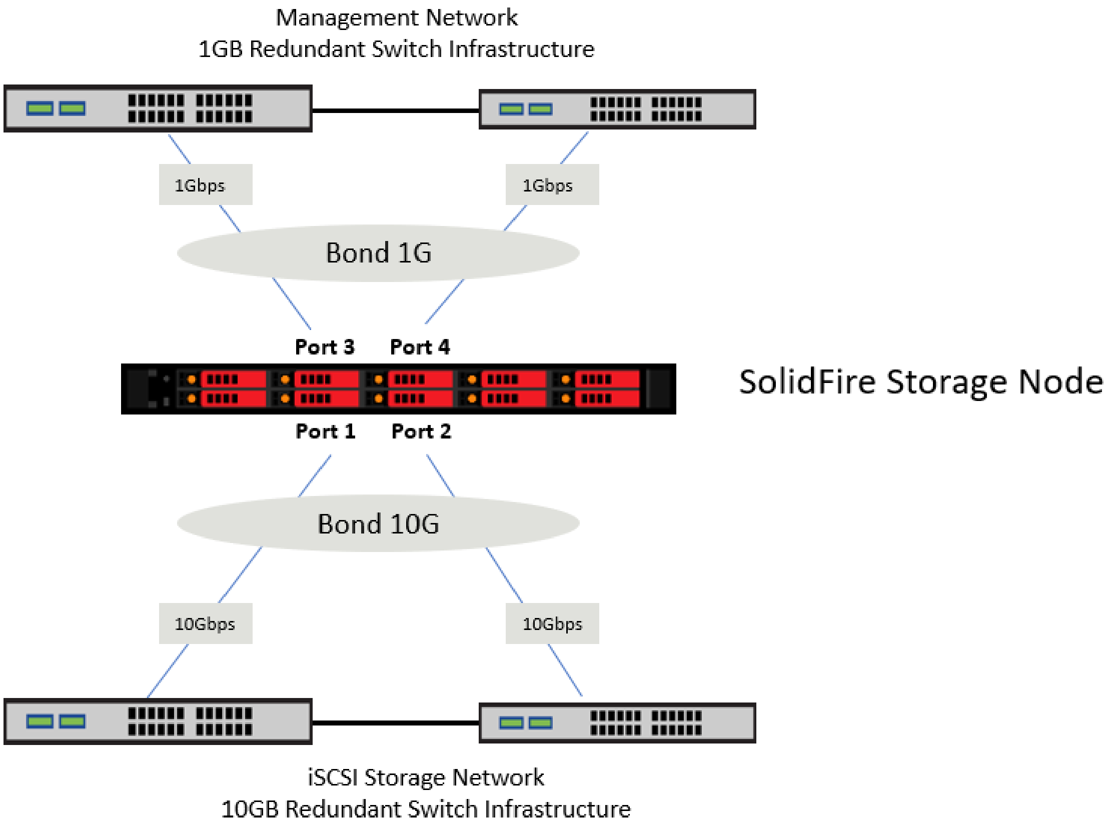
1. Log in to the storage node UI at https://<NodeIP>:442.
2. Change the IP address for the 1GbE or 10GbE interfaces.
3. Save the changes.
4. In the Element software web UI, go to Cluster > Nodes > Pending to view the list of pending nodes.
5. To add individual nodes, click the Actions button for the node you want to add, and click Add.
6. Go to Cluster > Drives.
7. Click Available to view the list of available drives.
8. Select the available drives and add them to the cluster.
9. Click Reporting > Running Tasks to view the block and metadata synchronization status and wait for the process to complete.
10. Repeat the above steps to change more IP addresses.

    Note: You cannot change the management virtual IP address (MVIP address) and storage virtual IP address (SVIP address).

4 NetApp SolidFire Network

Figure 1 shows the node configuration in a NetApp SolidFire storage network.

Figure 1) NetApp SolidFire network diagram.



Where to Find Additional Information

To learn more about the information that is described in this document, review the following documents and/or websites:

- NetApp Product Documentation
  https://docs.netapp.com

Contacting SolidFire Active Support

If you have any questions or comments about SolidFire documents or products in general, contact SolidFire support or email support@solidfire.com.

Version History

Version	Date	Document Version History
Version 1.0	April 2019	Initial release

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

Data contained herein pertains to a commercial item (as defined in FAR 2.101) and is proprietary to NetApp, Inc. The U.S. Government has a non-exclusive, non-transferrable, non-sublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b).

Trademark Information