Technical Report

Best Practices for Clustered Data ONTAP Network Configurations

Mike Worthen, NetApp
June 2013 | TR-4847

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

This technical report describes the implementation of clustered Data ONTAP® network configurations. It provides common clustered Data ONTAP network deployment scenarios and networking best practice recommendations as they pertain to a clustered Data ONTAP environment. A thorough understanding of the networking components of a clustered Data ONTAP environment is vital to successful implementations.

This report should be used as a reference guide ONLY. It is NOT a replacement for product documentation or specific clustered Data ONTAP technical reports or end-to-end clustered Data ONTAP operational recommendations or cluster planning guides.

2 OVERVIEW

There are different types of cluster configurations that can be implemented, all of which utilize various networking concepts and features.

- Single-node cluster (Figure 1): In a single-node cluster, settings such as flow control and tcp options still need to be configured appropriately. Also, if the configuration will be upgraded from a single node, the best practice recommendation is to install the necessary components for the upgrade and expansion during the initial implementation. For example, install NICs for cluster interconnectivity. This could save a reboot or two when the need to move to a highly available solution presents itself.
- Two-node switchless cluster (Figure 2): In a two-node switchless cluster configuration, settings such as flow control, tcp options, and MTU also need to be configured appropriately. Set the expectation that this configuration can be moved to a multinode switched configuration nondisruptively.
- Multinode switched cluster (Figure 3): In a multinode switched cluster, the customer will gain all the benefits that clustered Data ONTAP offers: the nondisruptive capabilities, the highly available capabilities, and the performance capabilities. This is the more complex solution of the three to configure, but it is also the one with the greatest return.

Figure 1) Single node cluster.

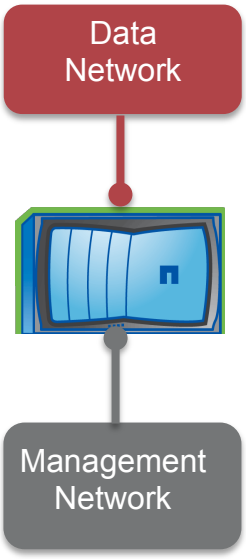


Figure 2) Two-node switchless cluster.

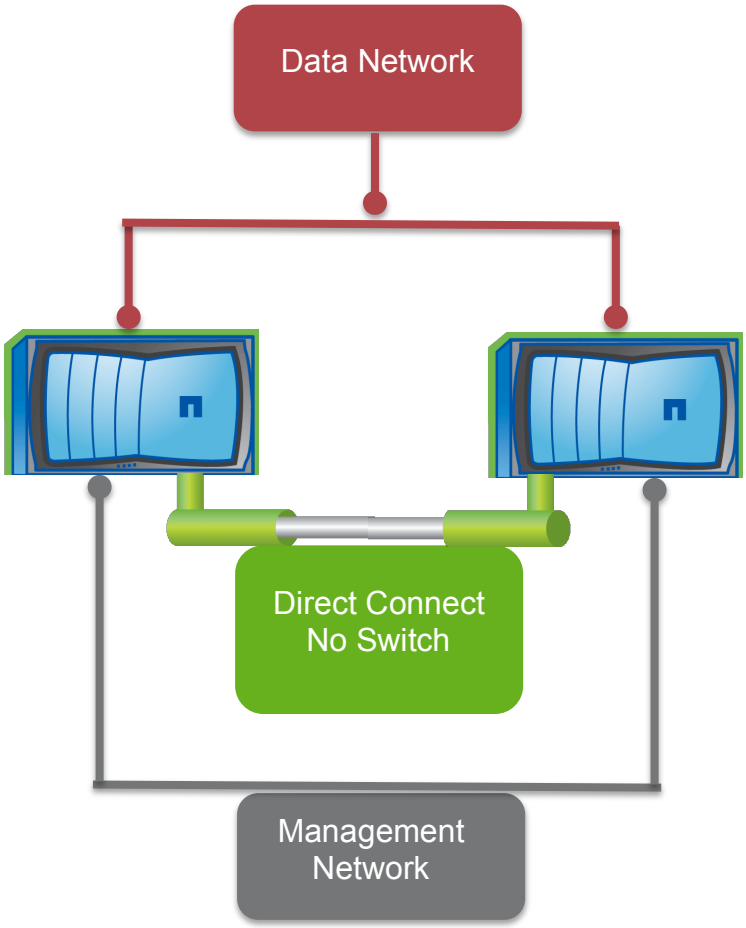
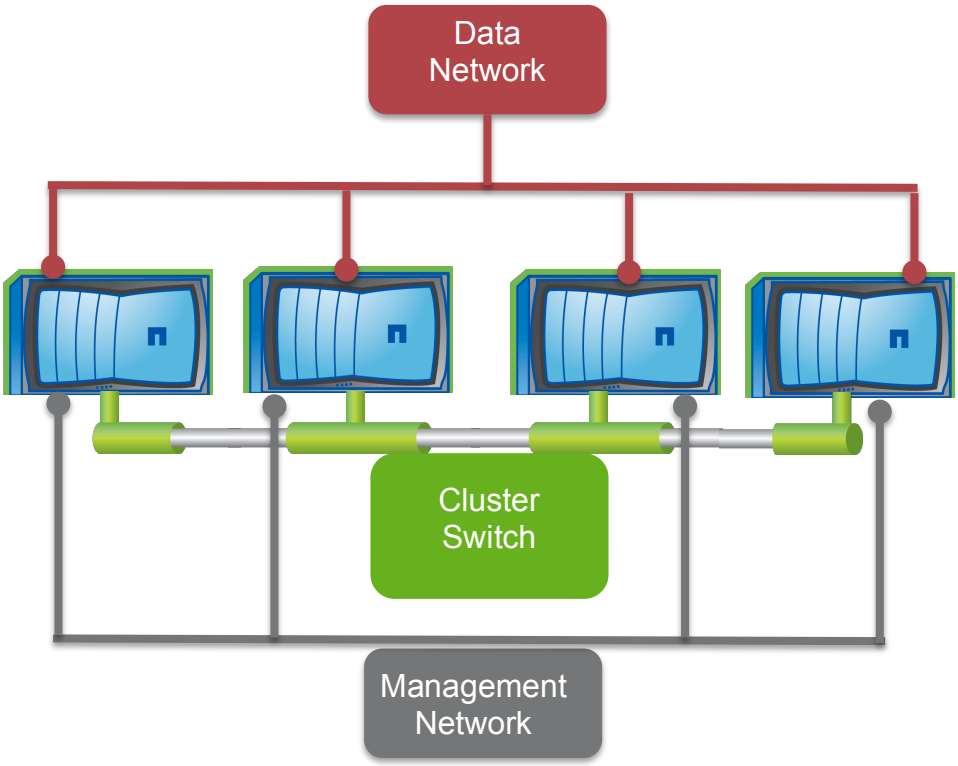


Figure 3) Multinode switched cluster



2.1 Setting Up the Cluster

Several software and hardware prerequisites are required for initially setting up and configuring a clustered Data ONTAP implementation.

- Ports
  - Physical ports: Will be used for various functions and can have different types of configurations.
  - Virtual ports: VLANs and interface groups (IFGRPs) make up the options for virtual ports.
- Logical Interfaces
  - Data
  - Cluster
  - Cluster management
  - Node management
  - Intercluster

2.2 Ports

There are different types of ports in clustered Data ONTAP: physical and virtual VLANs and interface groups. The different port types are used throughout the cluster in different configurations but are the building blocks that logical interfaces (LIFs, which are described in more detail in section 2.3) will use to allow the sending and receiving of data.

2.2.1 Physical Ports

Physical ports can be used individually or in combination when configuring virtual ports. If you use several physical ports together for an interface group it is important to remember to configure all relevant port settings in the same way (including MTU size and flow control). However, these settings will also be relevant if you exclusively use physical ports in a configuration in which only failover groups are in play; consistency in the settings and consistency in the configurations is needed.

2.2.2 VLANs

A Virtual Local Area Network (VLAN) subdivides a physical network into distinct broadcast domains. As a result, traffic is completely isolated between VLANs unless a router (Layer 3) is used to connect the networks. Complete isolation is one of the primary reasons to use VLANs from a security perspective. In clustered Data ONTAP VLANs subdivide a physical port into several separate virtual ports.

Using VLANs in a clustered Data ONTAP networking environment complements the resilient and flexible nature of clustered Data ONTAP. VLANs provide allowances by not requiring all physical equipment to be on the same subnet and not having to be located together. If resources within the cluster needed to be moved (logically or physically), VLANs would help reduce the number of changes needed from the switch and controller perspective. Due to the flexible and secure aspects of VLANs it is a best practice recommendation that they be used throughout the environment (specific examples of how to configure VLANs in a clustered Data ONTAP environment from a best practices perspective are listed throughout this document).

2.2.3 Interface Groups

Interface groups (IFGRPs) can be configured to add an additional layer of redundancy and functionality to a clustered Data ONTAP environment. They can also be used in conjunction with a failover group, which would help protect against Layer 2 and Layer 3 Ethernet failures. Below are characteristics listed for each type as well as situations in which it would be advisable to use each.

Note: Common to all three types.

- When creating an interface group, the best practice recommendation, if it is physically possible (if there are enough slots, NICs, and so on), is to create the interface group using ports from different NICs but verify that they are the same model and have the same speed, functionality, and so on. This is critical in maintaining consistency in the ifgrp in the event of a port failure. By maintaining consistency with port aggregation and by spreading the ifgrp over NICs in different slots you decrease the chances of a slot being responsible for taking offline all the ports in an ifgrp.
- The network interfaces and the switch ports that are members of the IFGRP **MUST** be set to use the same speed, duplex, and flow control settings.

Different Types of Interface Groups

2.2.3.1 Single Mode

A single mode interface group is an active-passive configuration (one port will sit idly waiting for the active port to fail) and it cannot aggregate bandwidth. Due to its limited capabilities, as a best practice recommendation NetApp advises not using this type. To achieve the same level of redundancy, instead use failover groups (see section 3.1).

2.2.3.2 Static Multimode

A static multimode interface group might be used if you want to utilize all the ports in the group to simultaneously service connections. It does differ from the type of aggregation that happens in a dynamic multimode interface group (described in section 2.2.3.3) in that no negotiation or autodetection happens within the group in regard to the ports.

2.2.3.3 Dynamic Multimode (LACP)

A dynamic multimode interface group might be used to aggregate bandwidth of more than one port. LACP monitors the ports on an ongoing basis to determine the aggregation capability of the various ports and continuously provides the maximum level of aggregation capability achievable between a given pair of devices.

However, all the interfaces in the group will be active, will share the same MAC address, and will handle load balancing outbound traffic. But this does not mean a single host will achieve larger bandwidth, exceeding the capabilities of any of the constituent connections. For example, adding four 10GbE ports to a dynamic multimode interface group will not result in one 40GbE link for one host. This is due to the way the aggregation of the ports in the interface group is handled by both the switch and the node. A recommended best practice is to use this type of interface group so that you are able to take advantage of all the performance and resiliency functionality the interface group algorithm has to offer.

2.2.3.4 Load Balancing for Multimode IFGRPs

Four distinct load-balancing modes are available.

- MAC: Only useful when the IFGRP shares the same VLAN with the clients having access to the storage. If any storage traffic traverses a router or firewall, do not use this type of load balancing.
- IP: Second-best load distribution method, since the IP addresses of both sender (LIF) and client are used to deterministically select the particular physical link that a packet traverses. Although deterministic in the selection of a port, the balancing is performed using an advanced hash function. This has been found to work under a wide variety of circumstances, but particular selections of IP addresses might still lead to unequal load distribution.
- Sequential: Nondeterministic load balancing. Under specific circumstances, this type of load balancing can cause performance issues due to high overhead to the switch (potential constant remapping of MAC/IP/Port) or out-of-order delivery of individual packets destined for a client.
- Port: Use this distribution method for best load balancing results. However, it lends itself less well to troubleshooting, since the TCP/UDP port of a packet is also used to determine the physical port used to send a particular packet. It has also been reported that switches operating in particular modes (mapping MAC/IP/Port) may exhibit lower than expected performance in this mode.

2.3 Logical Interfaces

Logical interfaces (LIFs) are created as an abstraction on top of the physical (physical ports) or virtual interface (VLANs or IFGRPs) layer. IP-based LIFs for NAS or iSCSI are assigned IP addresses, and FC-based LIFs are assigned WWPNs.

Note: Each node can support 256 NAS LIFs. However, if HA failover is enabled, only a maximum of 128 NAS LIFs should be configured per node. Example diagrams are listed in Figure 4 and Figure 5, below.

Figure 4) **Failover is enabled and 256 LIFs (per node) are configured. No apparent issues will be seen.**

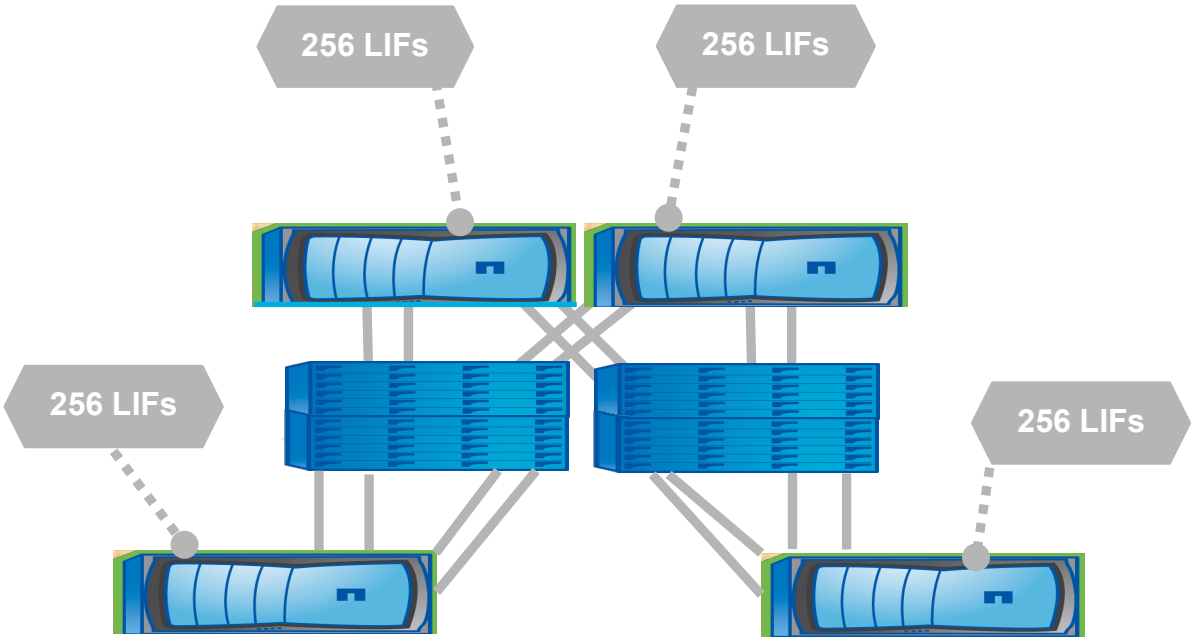
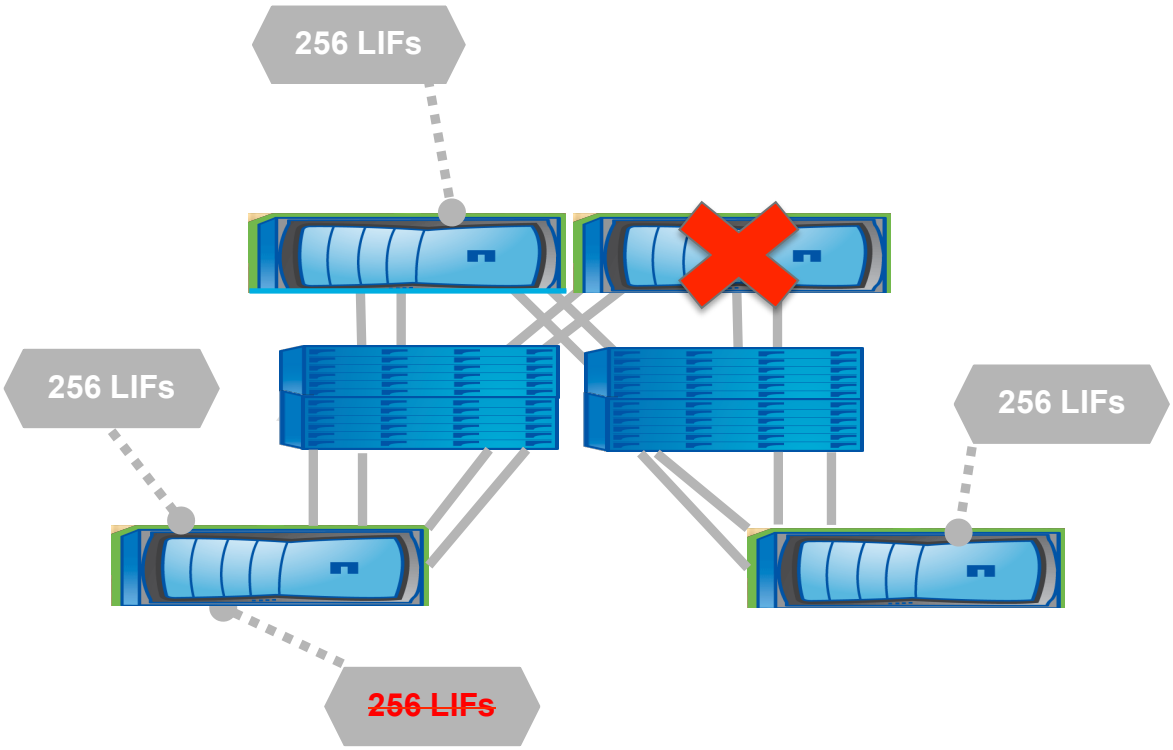


Figure 5) **When one node in the cluster fails, its 256 LIFs fail over to another node in the cluster but none of the failed node LIFs come online.**



2.3.1 Data

The data LIF is used for data traffic (NFS, CIFS, FC, iSCSI). Although you use a data LIF for either NAS or SAN traffic, you cannot use the same data LIF for both. The data LIF can fail over or migrate to other data ports throughout the cluster if configured to do so via a failover group.

Also, as a very important distinction, NAS data LIFs will migrate; SAN data LIFs **(including iSCSI)** do not migrate but will instead use ALUA and mpio processes on the initiators to handle path failures.

Note: Make certain that failover groups and the LIFs residing in them are configured correctly, meaning that you should configure the failover groups to use ports in the same subnet and verify that LIFs are assigned to the correct failover groups. If ports from different subnets are used in the same failover group or if LIFs aren't assigned to the correct failover groups and a failover occurs, it will result in loss of network connectivity that will result in the loss of data availability.

2.3.2 Cluster

The cluster LIF can only be configured on 10GbE cluster ports and can only fail over to cluster ports on the same node.

- It is used for operations such as:
  - Volume moves
  - To synchronize cluster/node configuration and metadata among the nodes in the cluster (this is a very important communication aspect since it keeps nodes in the cluster in quorum)
  - Access data that is remote to an interface

Visit the Cluster Management and Interconnect Switches link below for additional information:

http://support.netapp.com/documentation/productlibrary/index.html?productID=61470

2.3.3 Cluster Management

The cluster management LIF is used to manage the cluster. It can only reside on and fail over to data ports but can fail over to any data port on any of the nodes in the cluster.

2.3.4 Node Management

The node management LIF can be used to manage the node directly in the cluster for system maintenance. It can fail over to other data or node management ports on the same node only.

2.3.5 Intercluster

The intercluster LIF is used for peering from cluster to cluster. These are node specific; they can only use or fail over to intercluster or data ports on the same node. At least one intercluster LIF is required per node for replication between clusters. However, for the sake of redundancy the best practice recommendation is to either have two dedicated per node or configure a failover group for the intercluster LIF. Maintain consistent settings between the intercluster LIFs (same MTUs, flow control, tcp options, and so on).

Table 1) Additional information regarding LIFs

LIF Type	Function	Minimum Required	Recommended Number	Maximum Allowed
Node management	Used for system maintenance of a specific node, SNMP, NTP, and ASUP™ tool	1 per node		1 per port/subnet
Cluster management	Management interface for the entire cluster	1 per cluster		N/A
Cluster	Used for intracluster traffic	2 per node		2 per node
Data	Associated with a Storage Virtual Machine and used for data protocols and protocol services (NIS, LDAP, AD, WINS, DNS)	1 per Storage Virtual Machine	1 per Storage Virtual Machine	128 per node in HA configuration 256 per node in non-HA
Intercluster	Used for intercluster communication, such as setting up cluster peers and SnapMirror® traffic	1 per node if cluster peering is enabled	At least one; scale as the number of SnapMirror relationships increases	N/A

3 Storage Virtual Machine Networking

3.1 Failover Groups

In the event of a failure, LIFs need to be migrated in a coordinated manner. When a LIF is created it is assigned to a system-defined failover group by default. However, the behavior of the default failover group may not be sufficient for every different type of environment.

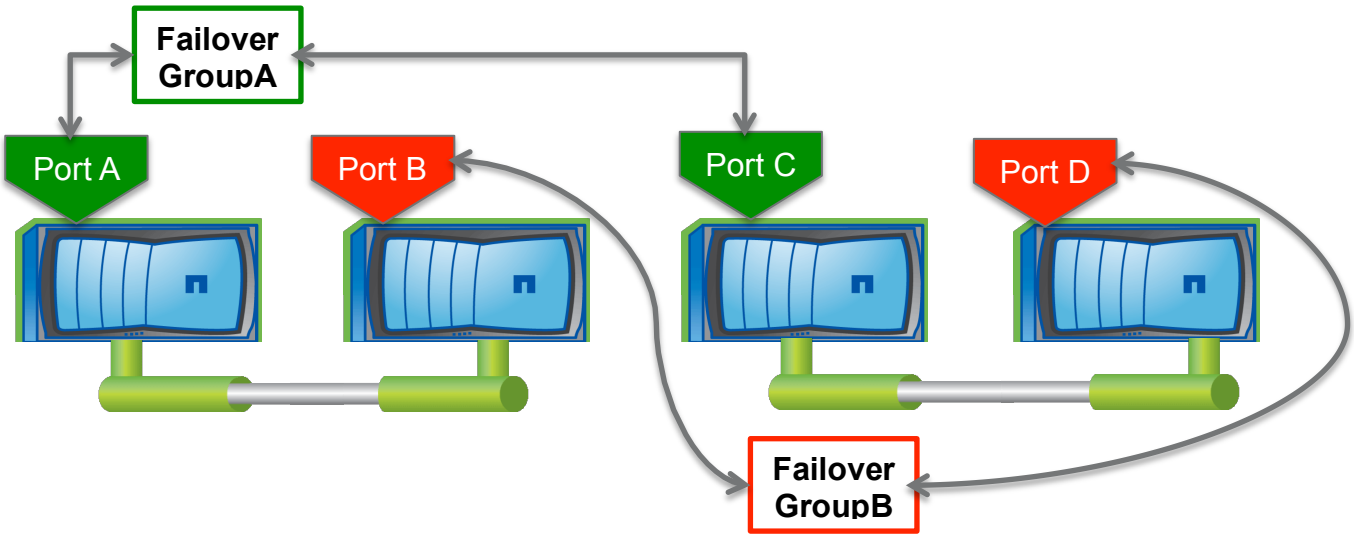
A failover group contains a set of network ports on one or more nodes. A failover group can have cluster management, node management, intercluster, and NAS data LIFs assigned to it. As mentioned previously in this document, SAN LIFs don't fail over so they don't utilize failover groups. The network ports that are present in the failover group define the failover targets (failover targets are considered the set of nodes and ports that will make up the parameters in a failover group) for the LIF. The best practice recommendation for LIFs capable of utilizing failover groups should always be to assign those LIFs to an appropriate failover group. Also, make double-checking a best practice by making certain all ports in the failover group are part of the same subnet. Failure to determine that ports are in the same subnet and failure to assign LIFs to appropriate failover groups will result in loss of connectivity to data.

There are currently three different types of failover groups. Below, each is described with details on when you might want to use it.

- Cluster-Wide—This type is automatically created during setup and it cannot be modified. It includes all data ports and cluster management LIFs by default. As long as the network is flat (that is, there are no subnets), it will successfully control failover of the LIFs that are assigned to it.
- System Defined—This type is also automatically created during setup, cannot be modified, and will control failing over all data LIFs by default. As with the cluster-wide type of group, system defined is useful as long as the network is flat.

  Note (see Figure 6): System-defined groups will only contain ports from a maximum of two nodes: ports from one node of an HA pair combined with ports from a node of a different HA pair. This decreases the chance of complete loss of connectivity in the event one node fails in an HA pair followed by the second node in the same HA pair.

Figure 6) Example of the system-defined failover group.

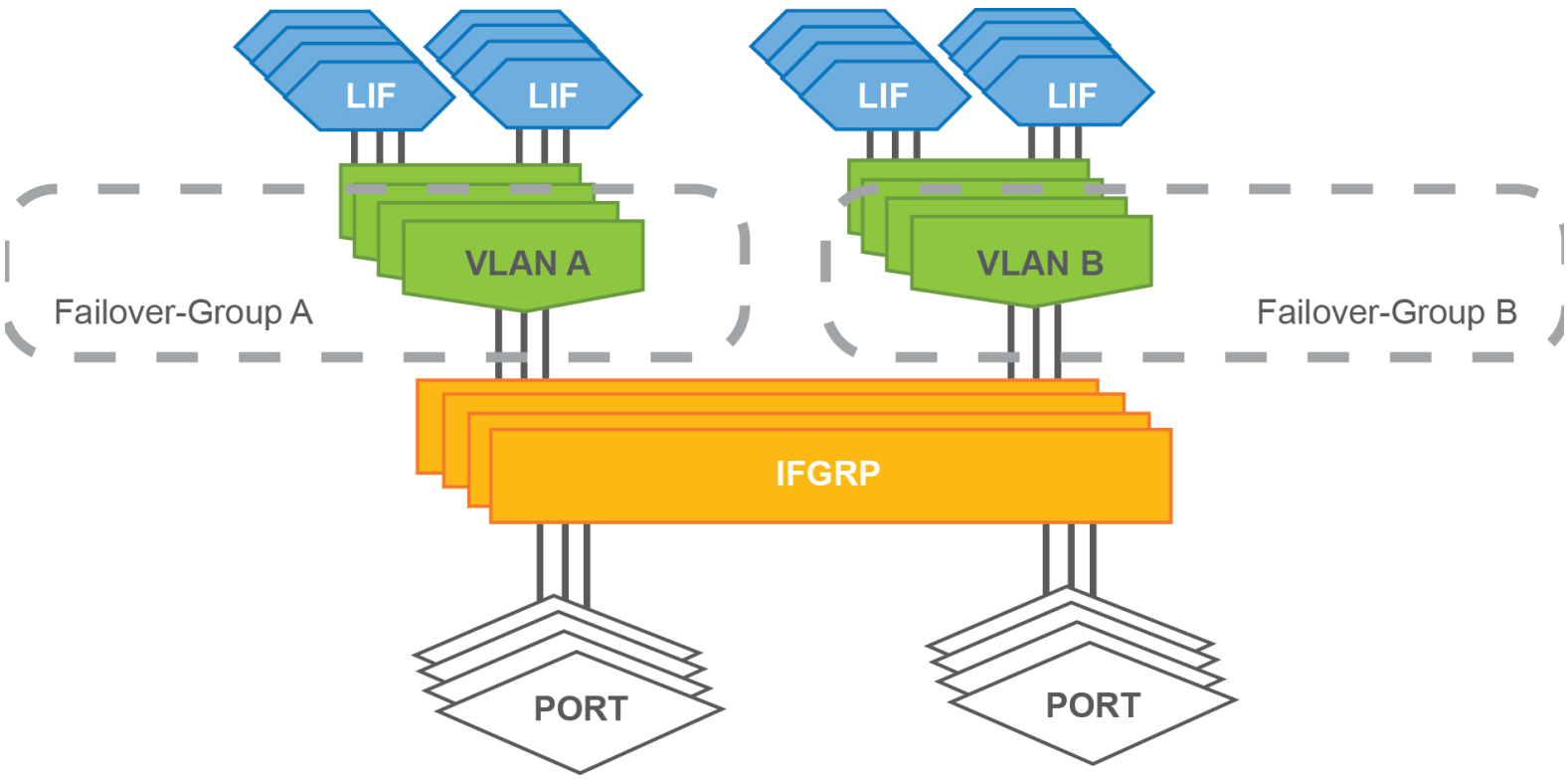


- User Defined—This option can be used if the system-defined failover group does not meet your needs. Several reasons to consider using it are:
  - If multiple subnets exist—System-defined groups could fail LIFs over to a port that resides in a different subnet.
  - You want to logically group a certain type of interface (for example, 10GbE-based LIFs only fail over to 10GbE ports).
  - LIFs are configured on top of VLAN ports and you want to be certain the LIFs move to port(s) that can communicate with the other devices that are members of that same VLAN configuration.

  This is the type of failover group to use from a recommended best practice perspective. You can configure it to provide all configuration requirements of any environment due to its very flexible functionality.

There are different LIF/failover/VLAN/IFGRP configurations possible in a clustered Data ONTAP environment (for more examples refer to the “Clustered Data ONTAP 8.2 Network Management Guide”). The best practice recommendation is to utilize the configuration in Figure 7. This configuration takes advantage of the cluster-wide failover capabilities of failover groups, the port aggregation functionality of interface groups, and the security aspects of VLANs.

Figure 7) Best practice LIF/failover/VLAN/IFGRP configuration.



4 Performance Considerations

4.1 Flow Control

Flow control from a NetApp perspective can be thought of as the mechanics that allow the receiving party of a connection to control the rate of the sending party. With that said, due to limitations with buffer designs on switches in the industry today, the best practice recommendation is to disable flow control throughout the network (including host ports, switch ports, and all node ports). Allow the upper layer protocols to handle congestion control as needed.

Note: When creating or configuring an interface on a NetApp® controller, the default for flow control settings should be "on" for both send and receive. You will need to change the settings "send off" and "receive off" using the `network port modify` command.

4.2 Jumbo Frames

Jumbo frames are Ethernet frames with more than 1,500 maximum transmission units (MTUs) of payload. In a clustered Data ONTAP environment, ports with a role type of cluster **must** be set to an MTU size of 9,000. The cluster may operate but will do so at a suboptimal level if the cluster ports are set to a 1,500 MTU size. Also, keep in mind that if the MTU is changed while the cluster port is active, the NIC will reset and connections will be dropped; this could have a very detrimental effect on the cluster.

Note: When a port is configured as a cluster, during initial setup the clustered Data ONTAP setup wizard will set it to 9,000 MTUs.

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.

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

Go further, faster®

www.netapp.com