GUIDE

Optimal Network Designs with Cohesity

TABLE OF CONTENTS

FIGURES

Introduction

Cohesity provides a hyper-converged secondary storage solution that converges backups, file and object services, test/dev, and analytics into a single, web-scale platform. The Cohesity platform can span both private and public clouds and enables replication and archival across enterprise LANs and WANs. As with any hyper-converged platform, Cohesity achieves highly efficient storage by integrating network, storage, and computing into a single data platform. However, integrating into a production network can be very involved and complex, requiring changes to routers, switches, and configurations.

A seamless integration into such a complex environment requires careful consideration for networking plans, to ensure the performance and availability of the Cohesity cluster. In this paper, we explore the most common network configurations and present the best practices to design an optimal Cohesity cluster.

This document covers:

- Key Concepts
- Five Common Configurations
- Terminology

Key Concepts

Before we explore the different configurations, there are several core concepts we should review in the Cohesity context:

CONCEPT	COHESITY CONTEXT
Ethernet	A Cohesity cluster relies on a high-bandwidth, high-throughput Ethernet network that is resilient against bursts, congestion, and failures from network ports, switches, and the corresponding links.
Connectivity	The ideal connectivity provides a reliable, single hop between nodes without introducing latencies or bottlenecks due to oversubscription (for example, when multiple links fan-in into a single link).
Network	All nodes of a Cohesity cluster need to be part of a single, layer-2 network. The IP addresses need to belong to a single subnet.
Bond Mode	Bonding is the creation of a single bonded interface by combining two or more Ethernet interfaces for high availability or performance reasons. Bond Mode 1 is configured by default on the Cohesity nodes.
IP Address Allocation	Configuration of a Cohesity cluster requires careful consideration of IP address allocations. IP addresses are required for each node, and optionally, an IP address is required for VIP. There must be at least one VIP for the entire cluster and one IPMI IP per node. The best practice calls for a VIP corresponding to each Cohesity node. Cohesity does not support DHCP and requires statically configured IP addresses at Cluster creation and Node addition.
Nodes	Cohesity clusters are deployed as a minimum of 3 nodes. In the current generation of Cohesity hardware appliances, like the C2000 and the C3000 series, there are two 10GbE interfaces available on each node for networking. One 10Gb interface should be connected to switch A and another 10Gb interface to switch B.
East-West Traffic	In a Cohesity cluster, node-to-node traffic involves traffic traversing between the Cohesity nodes. This traffic includes: node rebalancing, auto-healing operations, metadata replication, and replication for the purpose of data redundancy — due either to Replication Factors (RF) or Erasure Coding (EC).
North-South Traffic	Traffic in a Cohesity cluster also involves, for example, the actual backup, file services, or replication traffic from outside the cluster (North-South traffic).
TCP/IP Connections	TCP/IP connections are established between all nodes to facilitate replication and metadata transfers. TCP/IP provides protocol-level support in case of switch overloads (due to oversubscription) and retransmissions (due to packet drops).

Five Common Configurations

The following sections provide the most common configurations encountered in production network environments.

3.1 SIMPLE TOPOLOGY

A simple topology is common for proof-of-concept testing and other non-production use cases.

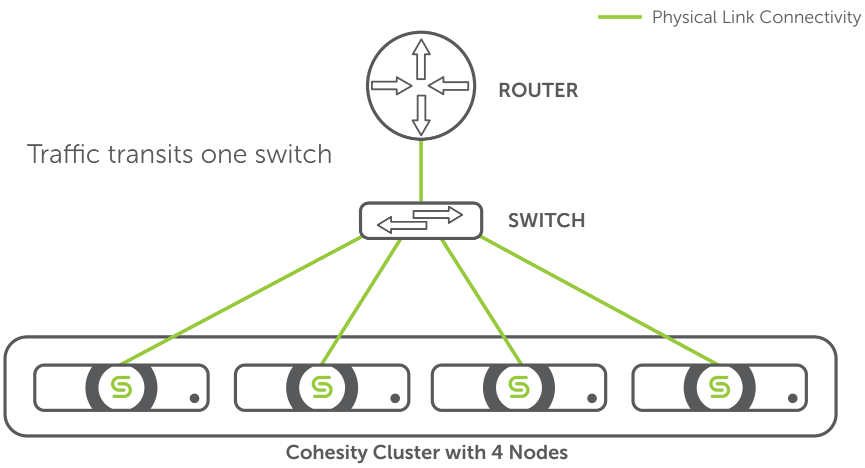

Figure 1: *Cohesity on a Simple Topology*

NOTE: The two parallel lines in the diagram represent two network links per node that are bonded.

Characteristics of a simple network topology:

- A single path through a single switch from one Node to another Node.
- A single, separate path from every node to the router (North-South traffic).
- This enables a line-rate performance (for example, 10GbE) from one node to another.
- There is no inherent choke point, and all traffic is non-blocking.
- East-West inter-node traffic is deterministic and constant.
- North-South traffic is deterministic and constant.
- Ethernet communication happens directly between two MAC addresses and does not change.

Before choosing a simple topology, consider:

- No NIC or Switch redundancy. A NIC failure can lead to the node becoming unreachable. A switch failure can lead to catastrophic failure that can make the whole cluster becoming unreachable
- There is possibility of contention with other traffic. If other devices on the production network are connected to the same switch, it can lead to the switch becoming loaded and can lead to Ethernet drops and TCP retransmissions between nodes.

BEST PRACTICE

Default Bond Mode 1 enables active-backup so that only one interface is in use, and the second interface is used when the first fails.

3.2 STANDARD TOPOLOGY

A standard topology introduces a second switch for redundancy, which is achieved by connecting the two Ethernet interfaces to two separate switches.

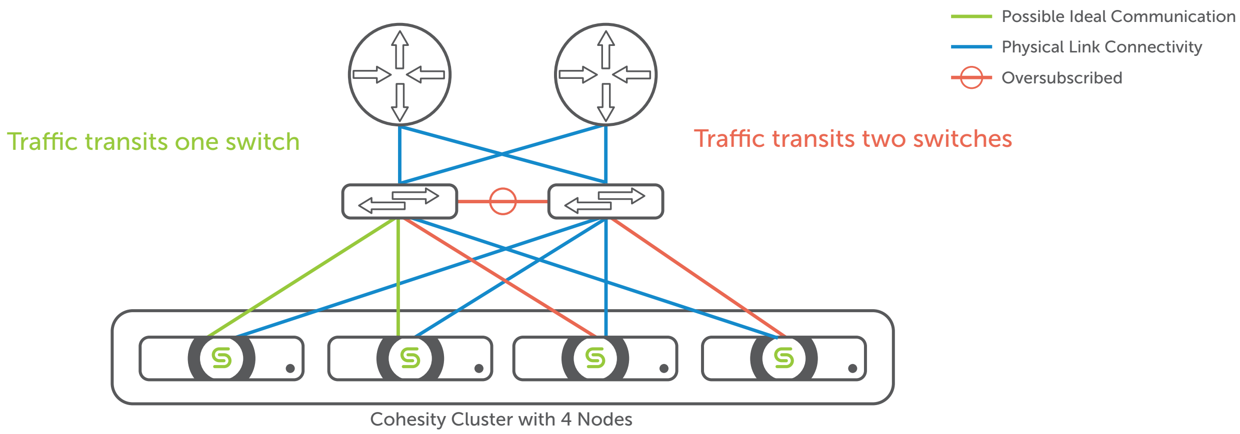

Figure 2: *Cohesity on a Standard Topology*

Characteristics of a standard network topology:

- Provides Node NIC and network switch redundancy.
- All nodes might not forward traffic to the same switch, therefore forcing some traffic to transit the peer link between switches. This might introduce drops, delay, and latency.
- No special configuration is required on the network switches. (All node interfaces are required to be in the same VLAN)

Before choosing a standard topology, consider:

- Oversubscription or head-of-line blocking can occur where multiple links have to traverse through the single inter-switch peer link. For example, multiple 10GbE links could go into a single 10GbE link.
- Oversubscription can lead to frame drops. Frame drops are expensive and can lead to performance degradation because of packet retransmissions.
- Scaling becomes a problem as more switches are added.

BEST PRACTICE

For best results, configure Bond Mode 4 on the nodes with MLAG/VPC at the cluster switch layer.

3.3 LAYERED TOPOLOGY

Some production environments have a layered topology to support a large number of devices. This is similar to the standard topology, except that nodes may communicate over multiple switch layers to reach other nodes within the cluster.

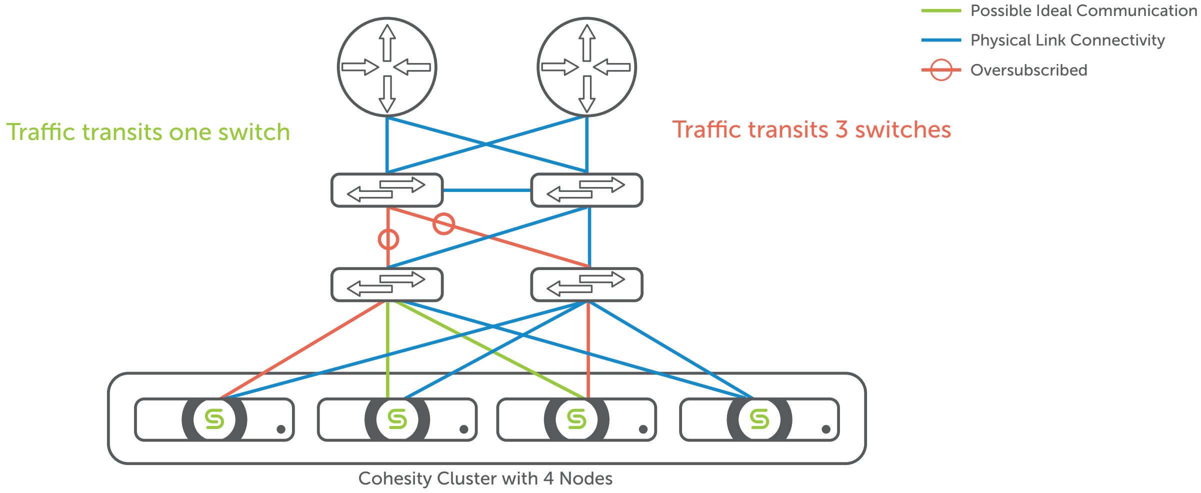

Figure 3: *Cohesity on a Layered Topology*

Characteristics of a layered network topology:

- The peer link might be established only on the top-most layered switches, which means some traffic might have to flow through all the layers, and back down, before reaching its destination.
- Helps with redundancy because multiple switches are used.
- The peer links at the first-hop switches are sometimes removed to simplify the spanning tree topology.

Before choosing a layered topology, consider:

- This can lead to oversubscription if multiple paths traverse the same inter-switch links.
- This can lead to some nodes using a single hop and some using multiple hops to reach the other nodes, resulting in increased latencies.

BEST PRACTICE

Connect all Cohesity nodes directly to a pair of switches. The most ideal configuration is Bond Mode 4 with MLAG/VPC at the cluster switch layer.

3.4 CISCO FEX TOPOLOGY

Production environments can include Cisco Fabric Extender Technology (FEX) switches that have already been installed for a legacy environment involving many network devices.

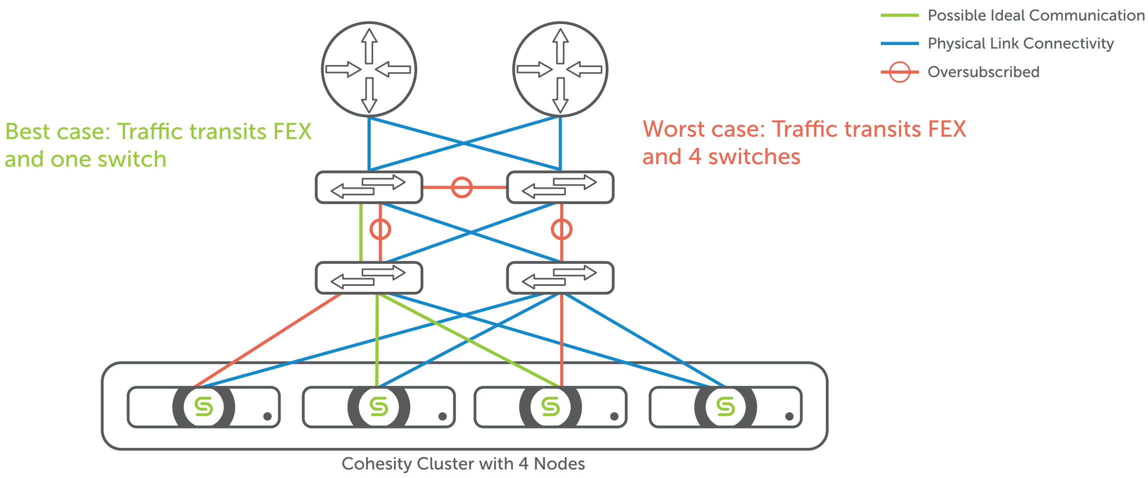

Figure 4: *Cohesity on a Cisco FEX Topology*

Characteristics of an FEX network topology:

- A Cisco FEX topology involves using intermediate FEX devices to increase port counts and improve fabric reach.

Before choosing an FEX topology, consider:

- An FEX is designed for north south traffic that must transit the uplink anyways.
- An FEX lacks the capability to perform any local switching.
- Can lead to traversing multiple hops between nodes.
- Even the use of Bond Mode 4 might not reduce oversubscription.
- For east-west traffic, as the number of paths, latency, and delays increase, the possibility of packet drops increases.
- Increased traffic congestion at the switch and router layer.

BEST PRACTICE

Cohesity supports Bond Modes 1 (active-backup, the default) and 4 (802.3ad, aka LACP). Bond Mode 4 is useful with the LACP use case, where multiple ports on the switch might be configured to do LACP, while Bond Mode 1 can be used in the other use cases.

3.5 LAYERED BOND MODE 4 (LACP) TOPOLOGY

A Bond Mode 4 topology currently provides the optimal solution for cluster connectivity. It solves the problem of oversubscription by having the traffic directly traverse only one intermediate switch instead of several intermediate switches. This topology involves configuring nodes for LACP, and a special configuration on the switches.

Administrators need to configure a pair of switches to perform either MLAG (Multi-Chassis Link Aggregation) or VPC (Virtual Port Channel).

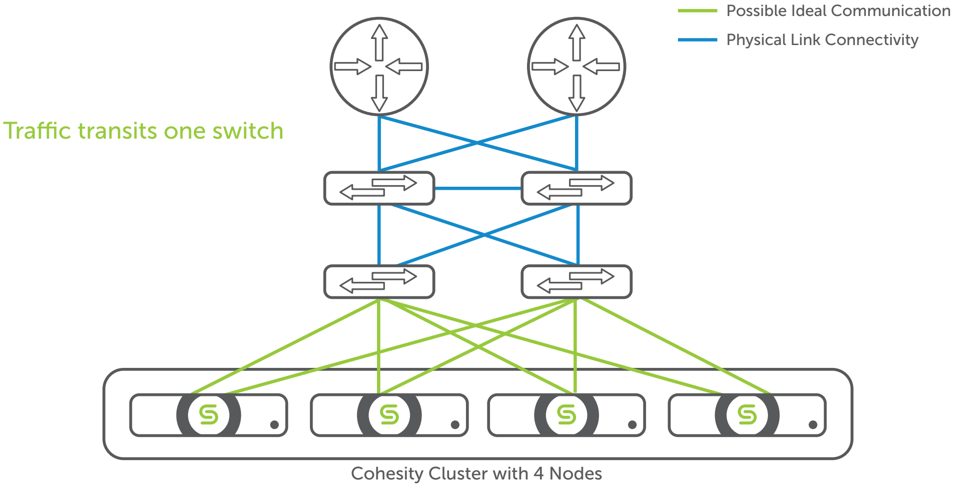

Figure 5: *Cohesity on a Layered Bond Mode 4 (LACP) Topology*

Characteristics of a Bond Mode 4 (LACP) network topology:

- Cohesity nodes must be configured for Bond Mode 4 (802.3ad).
- Connected network switches must support LACP.
- Connection to two different switches requires software support from the switch vendor. For example, Cisco provides Virtual Port Channel (VPC) while Arista provides Multi-Chassis Link Aggregation (MLAG).

Before choosing a Bond Mode 4 topology, consider:

- Requires careful configuration and testing from the network team.

BEST PRACTICE

This topology represents the best configuration. Configure Bond Mode 4 on the nodes, connecting each interface on a node to a different switch, and configure the switches to enable VPC or MLAG.

Terminology

Terms, protocols, and concepts discussed in this document include:

TERM	DEFINITION
Bond	A Linux virtual network interface used to aggregate physical network interfaces for the purpose of failover, increased throughput, or both. Bonds are numbered such that each Bond Mode <number> implies a certain way to aggregate network ports. Some bond types require special configurations on the switch that involves some initial negotiation on connectivity to make the bonding functionality successful.
Bond Mode 1	Combines network interfaces in an active-backup configuration, which can be useful in cases where the active one fails. In those cases, the backup interface can take over and continue serving traffic.
Bond Mode 4	Combines network interfaces in an active-active configuration, in accordance with the IEEE 802.3ad specification.
Cluster	A 3-node (minimum) deployment of Cohesity physical nodes.
FEX	A Cisco Nexus 2000 Series Fabric Extender is like a switch that helps extend the network fabric.
LACP	Link Aggregation Control Protocol, as defined in the IEEE 802.3ad specification.
MLAG and VPC	Multi Chassis Link Aggregation (Arista) and Virtual Port Channel (Cisco) enable the switches to learn about the interfaces that the other connects to.
NIC	Network Interface Card. Indicates a single network link on the node used to connect to other devices.
Node	A server running Cohesity software.
Virtual IPs (VIPs)	We recommend you use one virtual IP for each cluster node.

Conclusion

Cohesity provides a few different options for network topologies. It is possible to configure the internal networking for optimal performance without sacrificing efficiency. While configuring the network, administrators need to give careful consideration to both the existing environment and their future needs.

We recommend that you deploy a layered topology using multiple switches and multiple NIC ports going from each node to each switch. If MLAG or VPC support is available on the switches, a Layered Bond Mode 4 topology represents the best configuration and solves the network oversubscription problem, while simultaneously leveraging both the switches to provide additional redundancy.

Document History

DATE	VERSION	AUTHORS
November 2018	1.0	Arvind Jagannath Barnaby DiAnni

2000001-004-EN-US 11-16-2018