Networking Guide for Lenovo Converged HX Series Nutanix Appliances

Last Update: 9 January 2017

Describes the recommended networking topology using Lenovo RackSwitch top-of-rack switches

Shows how to configure the Lenovo RackSwitches in a VLAG peer configuration

Explains methods to connect the cluster switches to core network switches

Describes vSwitch settings for VMware vSphere

William Lloyd Scull

LENOVO PRESS

Abstract

Lenovo® Converged HX Series Nutanix Appliances are designed to help you simplify IT infrastructure, reduce costs, and accelerate time-to-value. These hyperconverged appliances from Lenovo combine industry-leading hyperconvergence software from Nutanix with industry-leading Lenovo enterprise platforms.

This paper is a best practices guideline for how to interconnect HX Series appliances running VMware vSphere and using Lenovo RackSwitch™ networking switches. It also recommends how to connect the completed cluster up to a client external network. The paper is not intended to be a detailed guide for all possible configurations, but rather will describe the most common configuration covering the majority of installations, with further information on selected variants in the appendices.

This paper is for anyone involved in installation and implementation of HX nodes and in possession of basic networking skills, including Lenovo technical colleagues, business partners and customers.

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

Do you have the latest version? We update our papers from time to time, so check whether you have the latest version of this document by clicking the **Check for Updates** button on the front page of the PDF. Pressing this button will take you to a web page that will tell you if you are reading the latest version of the document and give you a link to the latest if needed. While you're there, you can also sign up to get notified via email whenever we make an update.

Contents

Introduction

Lenovo Converged HX Series is a highly resilient Nutanix software-based hyperconverged compute and storage platform, designed to support virtual environments, such as VMware vSphere, Microsoft Hyper-V, and AHV. AHV (Acropolis Hypervisor) is the virtualization solution built into Nutanix Acropolis, based on KVM, Linux's Kernel-based Virtual Machine.

Figure 1 The Lenovo Converged HX5510 Nutanix appliance

The partnership announced between Lenovo and Nutanix for the Lenovo HX systems will bring reduced complexity in server, storage, networking and virtualization in data centers of all sizes. By combining the industry-leading reliability of Lenovo enterprise hardware systems with Nutanix, the software market-leader in hyperconvergence, enterprises will be able to bring greater efficiency and agility to their data centers.

The Lenovo Converged HX Series portfolio will enable you to:

- Simplify IT Infrastructure by integrating server, storage, networking, and virtualization in a centrally managed appliance, with an intuitive, consumer-grade friendly interface.
- Reduce costs and gain quicker ROI by breaking down silos enabling customers to easily add capacity and scale as future needs demand.
- Deliver greater reliability and service with Lenovo enterprise server innovation rated #1 in reliability and customer satisfaction.

Lenovo Converged HX Series hyperconverged architecture

Lenovo Converged HX Series appliances are available in a range of hardware platforms from remote office/branch office (ROBO) to compute-intensive, based on the application requirements of customers.

The HX Series architecture includes a storage controller, running in a virtual machine, called the Controller VM (CVM). This VM is run on every HX Series node in a cluster to form a highly distributed, shared-nothing converged infrastructure. All CVMs actively work together to aggregate storage resources into a single global pool that can be leveraged by user virtual machines running on the HX Series nodes.

The storage resources are managed by the Nutanix Distributed File System (NDFS) to ensure that data and system integrity is preserved in the event of node, disk, application, or hypervisor software failure. NDFS also delivers data protection and high availability functionality that keeps critical data and VMs protected.

Complete hyperconvergence includes the interconnection of the cluster nodes as a basic element of the complete solution – and is critical to both the performance and the availability of the HX Series appliances.

Networking and network design are critical parts of any distributed system in the data center. A resilient network design is important to ensure connectivity between HX CVMs, for virtual machine traffic, and for vSphere management functions, such as ESXi management and vMotion. HX Series appliances come standard with redundant 10 GbE NICs which can be used by the virtualized OS for resilient virtual networking.

Lenovo RackSwitch Ethernet switches are recommended for the interconnection of HX Series cluster nodes. Most of the 10 GbE SFP+ ports of the switches can be run in either 10 GbE or 1 GbE mode using either fiber or copper cabling. 40 GbE ports can run as four 10 GbE links.

These switches are designed for the data center and have redundancy in both power and cooling. They provide low latency, and implement required data center features, such MC-LAG (or Lenovo Networking's Virtual Link Aggregation Group - VLAG) for the creation of a single logical switch from a pair of physical switches.

The basic design overview of the switches for interconnecting HX Series nodes is shown in Figure 2.

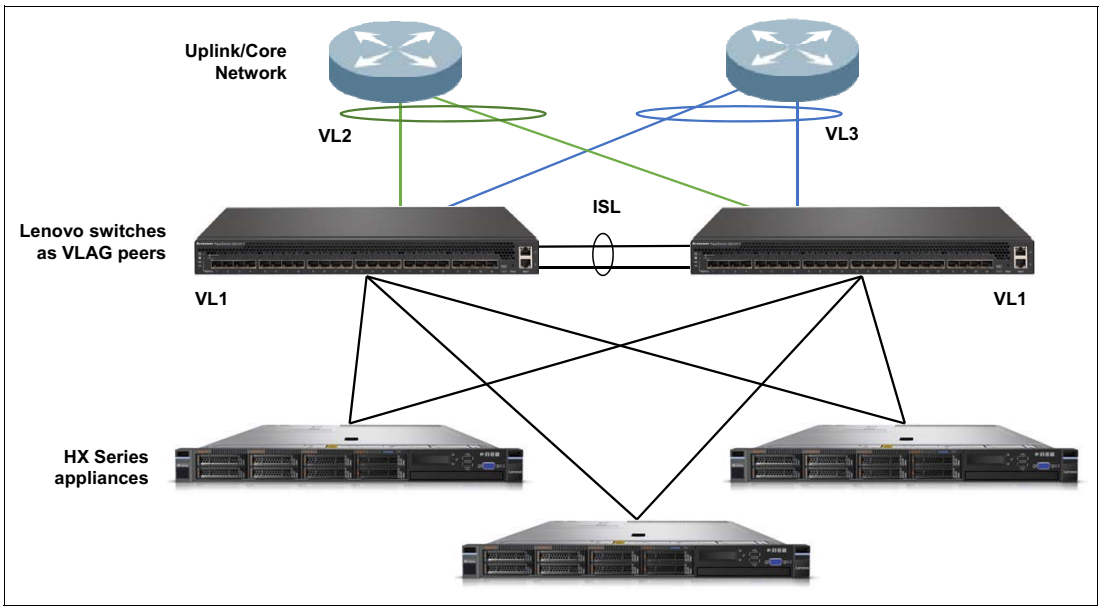

Figure 2 Basic design overview

For the basic interconnect, two Lenovo switches are placed in an MC-LAG configuration (called VLAG, similar to Cisco Nexus VPC) which enables this switch pair to act as a single logical switch over which single link aggregations can be formed between ports on both hardware switches. The HX Series appliances are connected redundantly to each of the VLAG peer switches and rely on VMware vSwitch features to spread the traffic over the two VLAG peers.

Connection to the uplink core network is facilitated by the VLAG peers, which present a logical switch to the uplink network, enabling connectivity with all links active and without a hard requirement for spanning-tree protocol (STP). The link between the two VLAG peers is an inter-switch link (ISL) and provides excellent support of east-west cluster traffic between HX nodes. The VLAG presents a flexible basis for interconnecting to the uplink/core network,

ensures the active usage of all available links, and provides high availability in case of a switch failure or a required maintenance outage. (The remaining switch carries all the traffic.)

About this paper

This paper is intended to describe the best practices for interconnection of HX nodes running VMware vSphere as the operating system. While other variants are available on the HX Series appliances, such as Hyper-V and Acropolis AHV, vSphere is still the most common OS in the data center and thus will be the focus of this paper. Considerations for Hyper-V and AHV are included in the appendices.

The paper takes you through the following steps:

1. "Preparing the switches"

   Verify that the prerequisites for network connectivity are in place, including identifying required existing infrastructure servers, preparing IP addresses, and verifying switch firmware levels and proper vSwitch configuration.

2. "Configuring the switches as VLAG peers" on page 11

   Configure the pair of identical Lenovo RackSwitch switches (G8124-E or G8272) as MC-LAG/VLAG peers such that they behave as a single logical switch for the purposes of external connections, and provide an inter-switch link (ISL) which can also be used for intra-cluster communication.

3. "Connect the VLAG to your existing core network" on page 16

   In this section, we provide considerations and best practices for connecting the interconnected HX cluster to an existing customer network infrastructure.

This document provides a number of design examples. Find the design which best fits your environment, fill in the locally relevant parameters (such as IP addresses, system names according to local naming rules, and relevant VLANs, if required), if necessary with the cooperation of your networking administrator, and follow the configuration steps described for the appropriate design.

Commands with blue text: Some of the commands have **`text in blue`**. This refers to configuration variables which need to be configured according to your local environment.

Preparing the switches

To connect the HX Series nodes in a local cluster using the Lenovo Networking switches, the following preparation steps are required:

1. Ensure you have a minimum of three HX Series appliances running VMware ESXi with minimum 2x10 GbE NIC ports.
2. Ensure you have the required physical and virtual IP address information and additional IP addresses for each node
3. Configure the vSwitches on the cluster connections as "Route based on originating virtual port" and connect them to the vmnics of the 10 GbE network
4. Ensure you have two of the same model of Lenovo Networking switches with Lenovo Networking firmware 8.1 or later. Required OS and Boot image can be downloaded from:

   `https://www-945.ibm.com/support/fixcentral`

The following subsections take you through the details of each of these steps.

Lenovo Converged HX Series appliances imaged with VMware ESXi

The following versions of VMware ESXi are supported:

- VMware ESXi 5.5 U2
- VMware ESXi 5.5 U3
- VMware ESXi 6.0 U1

Lenovo installs the Acropolis hypervisor (AHV) and the Nutanix Controller VM at the factory before shipping a node to a customer. To use a different hypervisor (ESXi or Hyper-V) on factory nodes or to use any hypervisor on bare metal nodes, the nodes must be imaged in the field.

For information about imaging, see the following documents:

- Reference Architecture for Workloads using the Lenovo Converged HX Series Nutanix Appliances

  https://lenovopress.com/lp0084

- Nutanix Field Installation Guide

  https://portal.nutanix.com/#/page/docs/details?targetId=Field_Installation_Guide-v3_3:Field_Installation_Guide-v3_3

Acropolis Hypervisor or Hyper-V? This document will concentrate on the interconnection of VMware vSphere nodes. For AHV, see “Appendix B: Acropolis Hypervisor considerations” on page 22.

Obtain the required IP address information

The appropriate network (gateway and DNS server IP addresses), cluster (name, virtual IP address), and node (Controller VM, hypervisor, and IPMI IP address ranges) parameter values needed for installation.

Network addresses

You will need the following existing information for the network during the cluster configuration:

- Default gateway
- Network mask
- DNS server
- NTP server

You should also check whether a proxy server is in place in the network. If so, you will need the IP address and port number of that server when enabling Nutanix support on the cluster.

Node IP addresses

Each node in the cluster requires three IP addresses, one for each of the following components:

- IPMI/IMM interface (recommended to be a separate physical network and IP range)
- Hypervisor host (same layer 2 network as the controller VM)

- Controller VM (same layer 2 network as the hypervisor host)

Hypervisor host and CVM IP addresses need to be in the same Layer 2 network. IPMI/IMM should be in a separate management network, if possible.

All Controller VMs and hypervisor hosts must be on the same subnet. No systems other than the Controller VMs and hypervisor hosts can be on this network, which must be isolated and protected.

System management

The HX Series appliances also require an IMM/IPMI connection to the management network. It is recommended to connect the switch management ports to this same network, which ideally is physically separated from the 10 GbE networking production traffic (at a minimum in a different IP range).

You will need to provide two switch IP addresses:

- Switch 1 management IP address
- Switch 2 management IP address

It is recommended that you access the switches over the out-of-band (OOB) management ports from a separate management network address range. Each switch can have up to two management addresses on the OOB management ports A and B, however at least one is required. Normally, these addresses will be given in the management network to which the HX node IMM and other management systems are connected.

Configure the vSwitches on the cluster connections

There is no specific VMware ESXi NIC bonding requirement for the NICs on the HX Series appliances in this configuration. Instead, the NICs are left unbonded and the traffic distribution over the NICs is controlled by the vSwitches connected to the vmnics configured as described in this section.

The available options in ESXi for NIC load balancing are as follows:

- Route based on originating virtual port (default, recommended)
- Route based on IP hash
- Route based on source MAC hash
- Route based on physical NIC load – called load-based teaming or LBT (VDS only)
- Explicit failover order

The recommended option for vSwitch VSS is “Route based on originating virtual port”.

The route based on originating virtual port option is the default load balancing policy and has no requirement for advanced switching configuration, such as LACP. These attributes make it simple to implement, maintain, and troubleshoot. A route based on the originating virtual port requires 802.1q VLAN tagging for secure separation of traffic types.

The main disadvantage of this option is that there is no load balancing based on network load. This results in traffic from a single VM always being sent to the same physical NIC, unless there is a failover event caused by a NIC or upstream link failure. This is less of an issue with the high throughput 10GbE network interfaces of the HX Series platform.

Figure 3 on page 8 shows the default NIC configuration in ESXi:

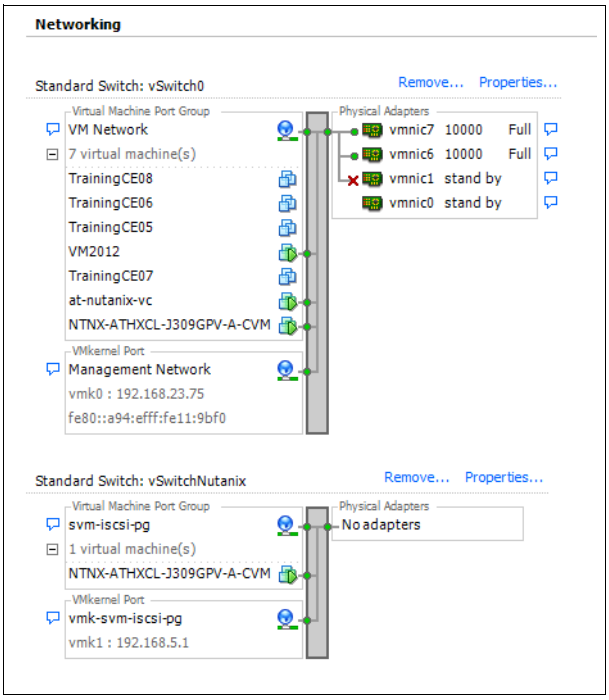

Figure 3 ESXi default NIC configuration

Figure 4 shows the physical 10 GbE NICs connected to vSwitch0.

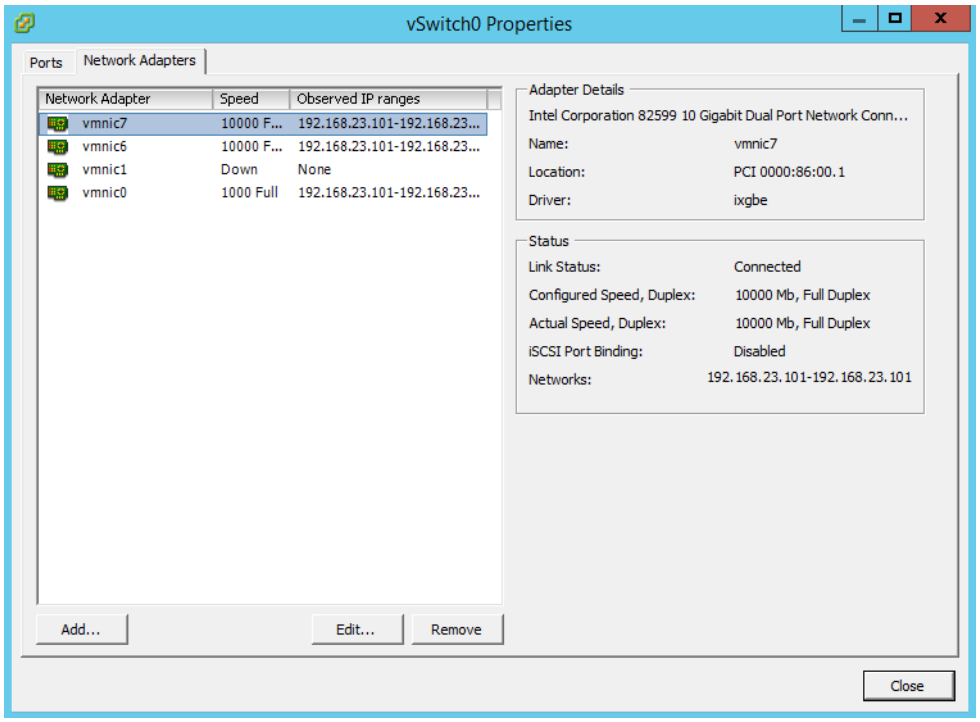

Figure 4 vSwitch0 properties showing connected network adapters

Figure 5 on page 9 shows the required load balancing option for vSwitch0, to which the HX Series appliances are connected to.

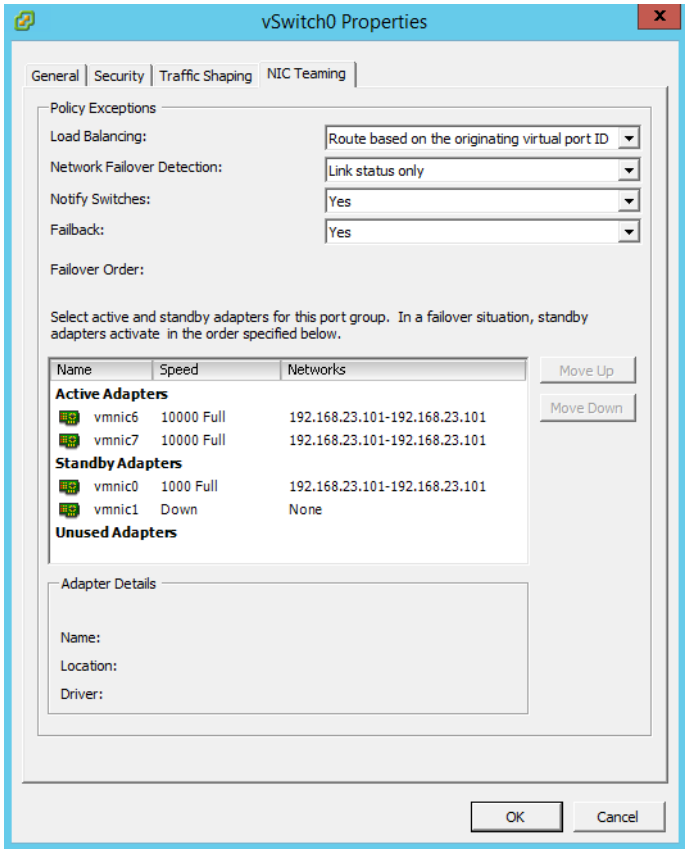

Figure 5 vSwitch0 NIC teaming

For configuration using the VMware Virtual Distributed Switch (VDS), see “Appendix C: VMware Virtual Distributed Switch considerations” on page 22.

Prepare the switches for implementation

Lenovo recommends the following switches for interconnection of HX Series appliances:

- For smaller clusters of up to 14 HX Series nodes:

  Two Lenovo RackSwitch G8124E switches running the latest Lenovo Networking OS (a minimum of version 8.1)

Figure 6 Lenovo RackSwitch G8124E

The RackSwitch G8124E is a datacenter-class 24-port SFP+ switch/router which can support either 1 GbE or 10 GbE links. Its available with either rear-to-front airflow (model 7159BR6 – when the switch is mounted in the back of the rack in the opposite direction of the servers – the most common scenario) or with front-to-rear airflow (model 7159BF7 – when the switch is mounted with the SFP+ ports visible on the front of the rack – uncommon due to the NIC cabling of the servers which is from the back of the servers). It has redundant power supplies and fans by default and low power usage.

- For larger clusters from 15 nodes to 35 nodes:

  Two Lenovo RackSwitch G8272 switches running the latest Lenovo Networking OS (a minimum of version 8.1).

Figure 7 Lenovo RackSwitch G8272

The RackSwitch G8272 is a datacenter-class 72-port SFP+ switch/router with 6 QSFP+ 40GE ports. It can support either 1 GbE or 10 GbE links on the SFP+ ports and 40GbE or 4x 10 GbE on the QSFP+ ports. Its also available in rear-to-front and front-to-rear airflow models as with the G8124E. It has hot-swappable redundant power supplies and fans by default, and boosts power usage.

- Clusters beyond 35 nodes

  Even larger HX cluster configurations can be supported using larger products, such as the RackSwitch G8296 and/or combinations of Lenovo switches in scalable spine-leaf configurations. Such configurations are supported but are beyond the scope of this document. Please contact your local Lenovo networking team for support.

Figure 8 Lenovo RackSwitch G8296

Figure 9 shows an example of a spine-leaf configuration, which - thanks to the high port-density of Lenovo RackSwitch switches - enables low-latency interconnect of a large number of HX Series appliances using a minimal number of switches.

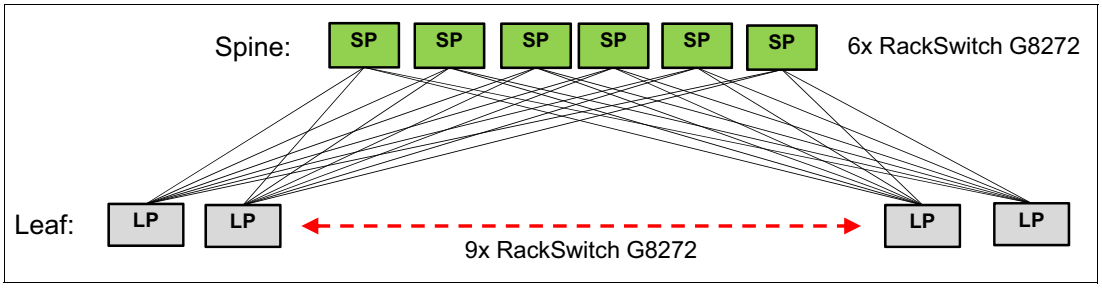

Figure 9 Example spine-leaf configuration

First-time access of the Lenovo switches requires initial setup. The factory default settings permit initial switch administration through only the built-in serial port. All other forms of access require additional switch configuration before they can be used. Remote access using the network requires the accessing terminal to have a valid, routable connection to the switch interface. The client IP address may be configured manually, or an IPv4 address can be provided automatically through the switch, using a service such as DHCP or BOOTP relay.

To manage the switch through the management ports, you must configure an IP interface for each management interface, as follows:

1. Connect to the switch via the serial port and log in.
2. Enter Global Configuration mode.

```
RS G8124E> enable
RS G8124E# configure terminal
```

3. Configure a management IP address and mask from the management network range

```
RS G8124E(config)# interface ip [127|128]
RS G8124E(configipif)# ip address <management interface IPv4 address>
RS G8124E(configipif)# ip netmask <IPv4 subnet mask>
RS G8124E(configipif)#enable
RS G8124E(configipif)#exit
```

Interface 127|128 refers to the interface designation used for the management interface; which one to use depends upon whether physical port A or B is used for the 1 GbE management link

- IF 127 supports IPv4 management port A and uses IPv4 default gateway 3.
- IF 128 supports IPv4 management port B and uses IPv4 default gateway 4.

Configuring the switches as VLAG peers

Both the clustering of HX Series nodes (server-to-server communication, including the storage networking and vSphere functions) and the connection of the clustered HX Series nodes and the customer uplink/core network can take place on the same switches.

While 1 GbE cluster connections are officially supported, Lenovo recommends the use of dual-port 10 GbE or 40GE NIC bandwidth for the interconnection of HX Series nodes. The examples in this document show single 10 GbE links to the HX Series nodes, but in fact multiple 10 GbE connections can be implemented using the same design.

The central element of the interconnected HX Series cluster are two Lenovo 10 or 40 GbE Ethernet switches set up to behave as a single logical switch, using the Virtual Link Aggregation protocol. For a physically separate management network, Lenovo offers a number of datacenter-class 1 GbE switches, such as the G8052 or G7028/52, which can implement the 1 GbE IMM/IPMI server links and the 1 GbE management links of the Lenovo 10 GbE switches.

Virtual Link Aggregation Group (VLAG) is Lenovo Networking's proprietary implementation of the MC-LAG protocol, similar to VPC on Cisco Nexus switches.

A full VLAG implementation consists of the following three elements:

- A pair of the same type of switches joined with a redundant inter-switch link (ISL) and configured with the VLAG protocol to form a single logical switch, on which aggregations can be implemented over ports on both physical switches.

  The formation of a single logical switch from two physical ones using VLAG is done using the MAC address tables of both switches being synchronized over the ISL. The two members of this pair of switches actively running the VLAG protocol are referred to as *VLAG peers*.

- A third element which supports link aggregation protocol such as LACP.

This device need not support the VLAG protocol directly, but only link aggregation. An aggregation is formed on this device by taking at least two ports of the NIC or switch and “bundling” them together locally. The remote end of the link aggregation is then “split” over the two VLAG peers. This device connecting to our VLAG peers using link aggregation is referred to as a *VLAG client*.

Note: It is recommended that you *do not* connect the HX Series nodes running vSphere as VLAG clients. This is because such a configuration requires the use of the VMware distributed virtual switch DVS to configure LACP on the NICs, and the standard vSwitch cannot do LACP.

The VLAG is, however, important for the connectivity to the uplink/core network, as we will see in “Connect the VLAG to your existing core network” on page 16

VLAG ensures that all links are active and that the traffic is balanced over the links using link aggregation features, without the need for implementing the spanning-tree protocol (which would block any links causing a loop). As a result, while every VLAG implementation is in fact a triangle, and thus a loop, the VLAG protocol ensures that a loop will not occur, thus making the use of spanning-tree optional in such environments.

Figure 10 shows an example of 2 VLAG instances with two other switches as the VLAG clients (on the bottom).

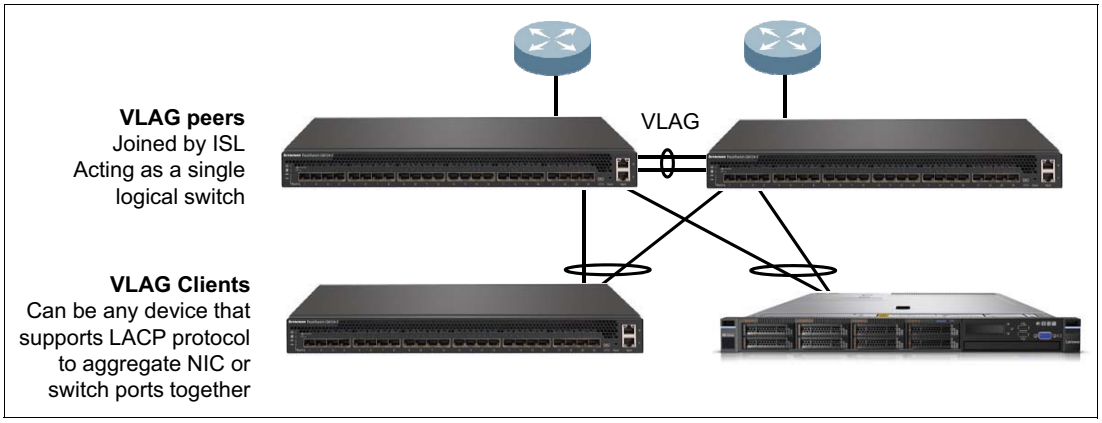

Figure 10 VLAG example topologies to both servers and switches

Configure the host side of the switch configuration

For our HX Series cluster connection we will configure our two Ethernet switches as VLAG peers to ensure redundant connection points from the HX host, but not place the NIC ports in a team or bond, as is shown in Figure 11 on page 13.

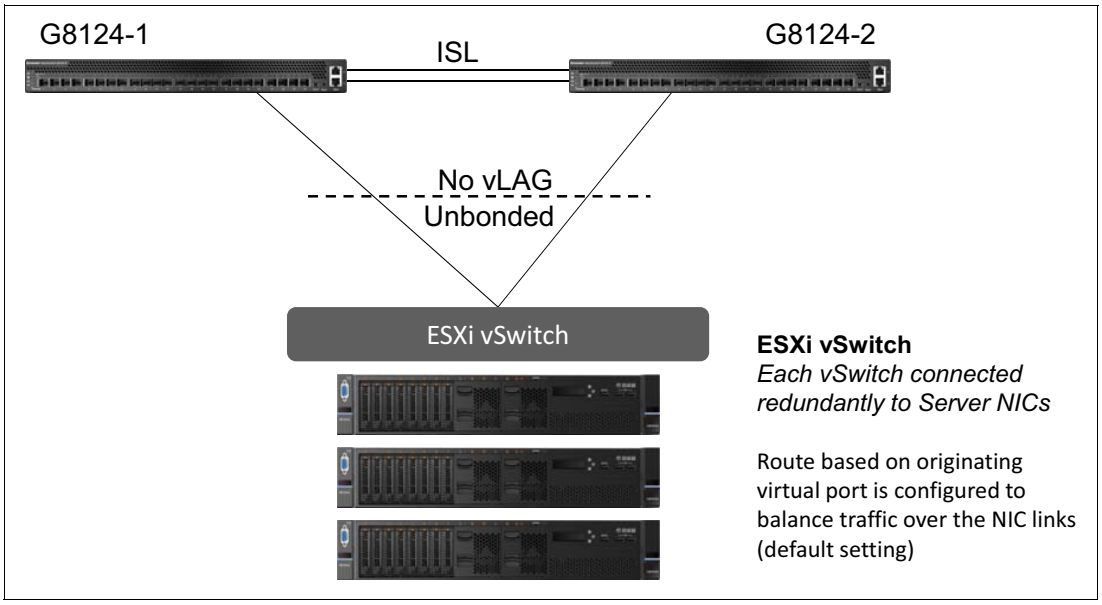

Figure 11 VLAG peers providing redundant uplinks to the HX Series host

The first step is to configure the pair of Lenovo switches as VLAG (MC-LAG) peers so that both switches are interconnected and act as a logical single switch for the purposes of connecting other switches. This will also provide a direct connection between the switches to ensure the traffic can efficiently be transferred between HX Series cluster nodes ("east-west" traffic), and well as enabling the ability to do a link aggregation across physical switches, which is a great advantage for the uplink connection to the core network. With VLAG, the loss of one of the switches and/or the ISL will force all HX traffic over the remaining switch in standalone mode.

The steps to set up VLAG between the two identical Lenovo switches are as follows:

Commands with blue text: Some of the commands have **`text in blue`**. This refers to configuration variables which need to be configured according to your local environment.

1. Choose a minimum of 2x 10 GbE (or 40 GbE) SFP+ ports on each switch for the redundant inter-switch link (ISL).

   Sizing tip: In normal VLAG operation, the ISL is used only for synchronization of MAC address tables of the VLAG peers. However, in case of the loss of an uplink to the uplink/core network, the ISL can in some cases be used to carry the production traffic, and thus needs to be sized accordingly. Therefore it is recommended that you size the ISL according to the number of HX nodes connected to the VLAG peers. A good rule of thumb is half the total downlink bandwidth for the ISL. For example, if there are eight HX nodes connected (totaling 80 Gb/sec) then the ISL should have at least 40 Gb/sec of bandwidth. Remember that all VLANs configured toward the HX nodes also must be added to the ISL ports, in case of redirection of uplink traffic over the ISL.

2. Connect these ports in parallel, using copper or fiber components. For example, connect port 3 on switch1 to port 3 on switch2, and port 4 on switch1 to port 4 on switch2.

For details on the cabling and distance requirements, please see the Lenovo Networking Products and Options Catalog:

http://www.lenovo.com/images/products/system-x/pdfs/datasheets/lenovo_networking_catalog_ds.pdf

3. Access the switch and log in.
4. Access Enable Privilege:

```
G8124E-1 # enable
```

5. Enter Global Configuration Mode:

```
G8124E-1 # conf t
```

6. Disable spanning tree protocol *globally* on both switches (not on a per-port basis)[1]

```
G8124E-1 (config)# no spanning-tree enable
```

7. To test that STP is disabled, run:

```
G8124E-1 # show span
```

The STP BPDUs from the uplink core network will continue to be forwarded through the Lenovo switches. In this case we are STP passive. The VLAG protocol will ensure the absence of STP loops as long as it is configured correctly.

8. Enable Link-layer Discovery Protocol on both the switches. This is useful for verifying device connectivity using the `show lldp remote-device` command.

```
G8124E-1 (conf)#  lldp enable
```

9. Create ISL LACP aggregation Switch 1

The VLAG peer switches share a dedicated ISL for synchronizing VLAG information and cross-connect. Use of a single ISL link is possible, but not recommended. The use of the VLAG health check over the management network is optional, but recommended.

```
G8124E-1(config)# interface port 3-4
G8124E-1 (config-if)# switchport mode trunk
G8124E-1 (config-if)# lacp mode active
G8124E-1 (config-if)# lacp key 200
G8124E-1 (config-if)# exit
```

10. Create VLAG tier and health check

```
G8124E-1 (config)# vlag tier-id 1
G8124E-1 (config)# vlag hlthchk peer-ip 10.0.0.2
G8124E-1 (config)# vlag isl adminkey 200
G8124E-1 (config)# exit
```

11. Create ISL LACP aggregation Switch 2

```
G8124E-2 (config)# interface port 3-4
G8124E-2(config-if)# switchport mode trunk
G8124E-2(config-if)# lacp mode active
G8124E-2(config-if)# lacp key 200
G8124E-2(config-if)# exit
```

[1] In normal VLAG operation, the spanning-tree protocol is not required, because the connection is a logical point-to-point connection of two elements. However, without spanning tree there is no protection in case of misconfigured aggregations or an incomplete VLAG configuration which can lead to a loop and possible broadcast storm and network outage. Please proceed with caution and follow local networking policies. Spanning-tree is compatible with VLAG and can be configured if network security policies require it – please see the application guide for the switch for further details.

12. Create VLAG tier and health check

```
G8124E-2 (config)# vlag tier-id 1
G8124E-2 (config)# vlag hlthchk-peer-ip 10.0.0.1
G8124E-2 (config)# vlag isl adminkey 200
G8124E-2 (config)# vlag enable
G8124E-2 (config)# exit
```

Now the two switches are configured as VLAG peers, and are thus acting as a single logical switch. This enables high-availability active-active connections between the switches and other devices, while maintaining two separate control instances (in contrast with switch stacking, in which a single control instance is used over all participating switches). This means that when one of the switch peers is lost (for example, during a reboot after a firmware update), the other will continue to forward all the remaining traffic until the peer returns to the VLAG.

13. Verify that the VLAG peers are configured:

```
G8124E-1 # show vlag info
```

The output should resemble Figure 12.

```
vLAG status: enabled
vLAG Tier ID: 1
vLAG system MAC: 08:17:f4:c3:dd:09
Local Priority: 0
ISL Information: Trunk 0, LACP Key 200
Health check Peer IP Address: 10.0.0.2
Health check connection retry interval: 30 seconds
Health check number of keepalive attempts: 3
Health check keepalive interval: 5 seconds
vLAG Auto Recovery Interval: 300  seconds
vLAG Startup Delay Interval: 120  seconds
```

Figure 12 Output from show vlag info

We strongly recommend that you configure the Lenovo switch VLAG to check the health status of its VLAG peer (step 12, above). Although the operational status of the VLAG peer is generally determined via the ISL connection, configuring a network health check provides an alternate means to check peer status in case the ISL links fail. Use an independent link (normally over the management network) between the VLAG switches to configure health check.

Because the HX host NIC connections are unbonded and use the vSwitch setting for traffic load balancing, the HX nodes are not fully participating in VLAG as VLAG clients. But as we will see in the next section, the VLAG is of great use for connectivity of our HX cluster to the network core by our VLAG peer switches.

At this time, the HX node NIC ports can be connected to the configured ports on the switches and the HX nodes will have connectivity to each other – this can be verified over the Nutanix PRISM interface.

Connect the VLAG to your existing core network

Now that the switches are set up as VLAG peers, and our HX Series cluster is up and running over the server-side ports, we can turn our attention to the connection of our cluster switch to the uplink core network. There are two scenarios regarding connectivity to the core network:

- Scenario 1: A pair of switches acting as a single logical switch (as with our VLAG peers)
- Scenario 2: A pair of switches that are not virtualized (two standalone switches with no direct connection to each other)

In both cases, the connection needs to be approached with care, as this represents an activity that can affect the production network. Please work with the responsible core network administrator to ensure that they are aware of, and involved with, this connection of the HX cluster. It could, for example, cause a spanning-tree recalculation and thus affect production traffic. Also, the core network administrator needs to be aware of any VLANs configured on the HX system and match the port aggregation components (ports, port speeds, copper/fiber, etc).

In general, we recommend using fiber transceiver technology between Lenovo switches and the uplink/core switches to ensure interoperability. Copper DAC cables can also be used; this will reduce the connection cost but can sometimes lead to compatibility issues, due to support of active/passive DACs and possible blocking of certain DACs due to quality or testing issues.

For more details on cabling options for Lenovo switches, please refer to the Lenovo Networking Products and Options Catalog:.

http://www.lenovo.com/images/products/system-x/pdfs/datasheets/lenovo_networking_catalog_ds.pdf

Scenario 1: The core uplink switches are virtualized

This is a common scenario, becausemany switches today implement a proprietary version of the MC-LAG feature. For Lenovo switches its called VLAG, for Cisco Nexus: VPC, Cisco Catalyst: VSS, Arista and Mellanox: MLAG, Juniper: MC-LAG. These function the same way as VLAG – a pair of vendor switches of the same type are connected using an ISL and the MAC address tables of both systems are synchronized to enable the formation of a single LACP aggregation over ports on both switch peers.

In this case we can implement MC-LAG on both sides of our interconnect: our VLAG facing the MC-LAG of the other vendor - see Figure 13 on page 17. In this section, we use Cisco Nexus VPC as the example.

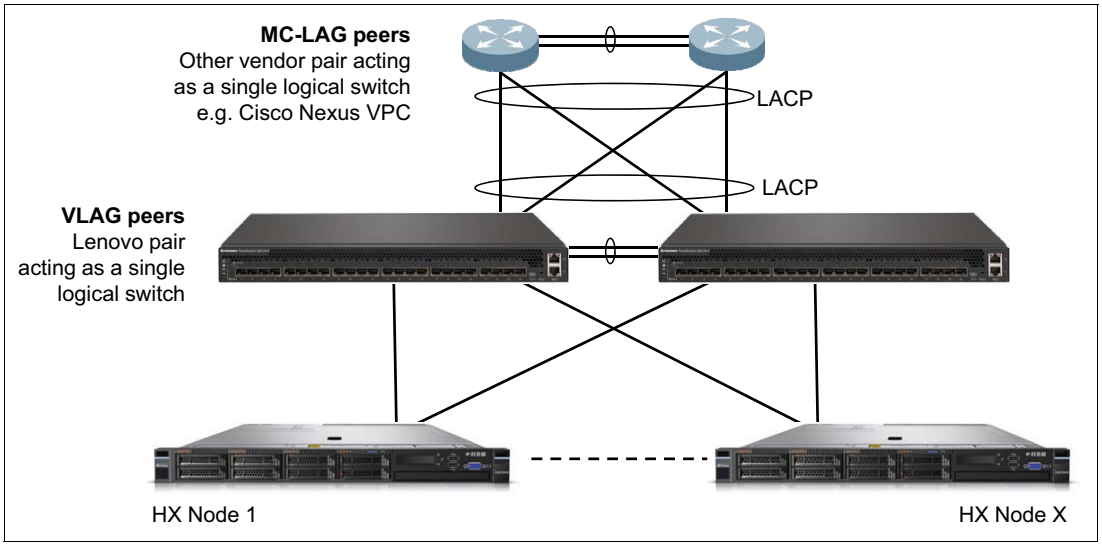

Figure 13 Scenario 1 design

On the Nexus side we will create an aggregation over the VPC peers and we will connect that directly to an aggregation formed over our VLAG peers as shown in Figure 14.

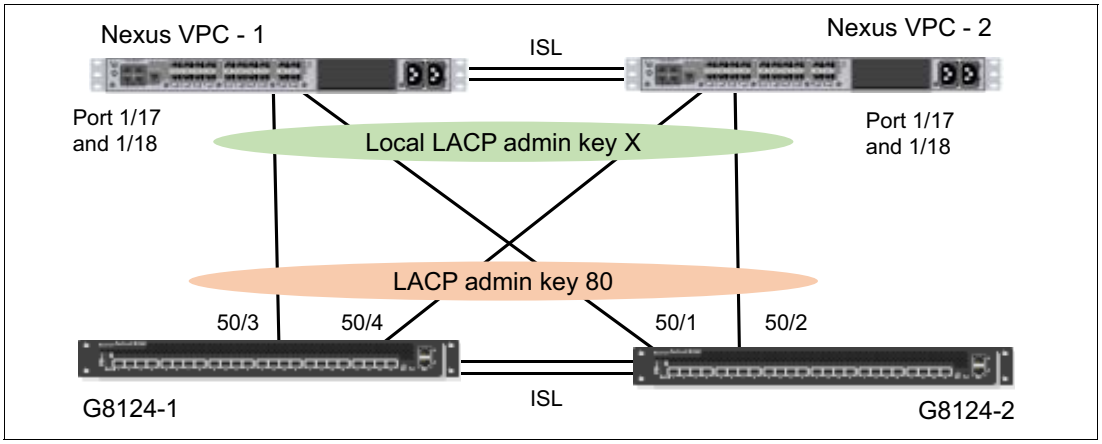

Figure 14 Interoperability between uplink MC-ALG instance, in this case VPC, and VLAG

The steps to configure the switches are as follows:

1. Issue the following commands on the Lenovo switches are as follows:

```
G8124E-1#( (config)# int port 50/3-50/4
G8124E-1#( (config-if)# lacp mode active
G8124E-1#( (config-if)# lacp key 80
G8124E-1#( (config)#
G8124E-1#( (config)# sho lacp info
```

   The output of the `sho lacp info` command will be similar to Figure 15.

```
port    mode    adminkey  operkey   selected   prio  aggr  trunk  status  minlinks
--------------------------------------------------------------------------------
.
.
.
50/1    off          53       53     no      32768    --     --    --        1
50/2    off          54       54     no      32768    --     --    --        1
50/3    active       80       80    yes      32768    55     74    up        1
50/4    active       80       80    yes      32768    55     74    up        1

.
.
.
 (*) LACP PortChannel is statically bound to the admin key
```

Figure 15 Output from sho lacp info

2. Configure the same on the corresponding ports on the other VLAG peer, this time ports 50/1 and 50/2. The `sho acp info` command result is shown in Figure 16.

   `G8124E-2#( (config)#` **`sho lacp info`**

```
port    mode    adminkey  operkey   selected   prio  aggr  trunk  status  minlinks
--------------------------------------------------------------------------------
.
.
.
50/1    active       80       80    yes      32768    53     74    up        1
50/2    active       80       80    yes      32768    53     74    up        1
.
.
.
 (*) LACP PortChannel is statically bound to the admin key
```

Figure 16 Output from sho lacp info

3. Define this LACP as a VLAG aggregation extending over both Lenovo switches by entering the following command on both switches.

   On Switch 1:

   `G8124E-1#( (config-if)#vlag adminkey 80 enable`

   On Switch 2:

   `G8124E-2#( (config-if)#vlag adminkey 80 enable`

   As can be seen using the following command that our VLAG for LACP 80 is up and running, and ready to be connected to the VPC LACP on the NEXUS switch:

   `G8124E-2#( (config)#` **`sho vlag info`**

```
vLAG Tier ID: 10
vLAG system MAC: 08:17:f4:c3:dd:09
Local MAC a4:8c:db:33:57:00 Priority 0 Admin Role PRIMARY (Operational Role PRIMARY)
Peer MAC a4:8c:db:33:7d:00 Priority 0
Health local 192.168.22.250 peer 192.168.22.251 State UP
ISL trunk id 73
ISL state    Up
Auto Recovery Interval:  300s (Finished)
Startup Delay Interval:  120s (Finished)
vLAG 73: config with admin key 80, associated trunk 74, state formed
```

Figure 17 sho vlag info command output

4. On the Cisco Nexus side, as with VLAG, an identical configuration is required on both VPC peers. The commands enable LACP and VPC.

```
switch# configure terminal
switch# configure terminal
switch(config)# feature lacp
switch(config)# show feature
switch(config)# feature vpc
switch(config)# show feature
switch(config)# vpc domain domain-id
switch(config)# interface port-channel channel-id
switch(config-if)# vpc peer-link
switch(config)# interface port-channel channel-id
switch(config-if)# vpc domain-id
```

5. Verify the vpc with the following command

```
switch# show vpc brief
```

For more information, see the chapter “Configuring Virtual Port Channels” in the Cisco NX-OS Layer 2 Switching Configuration Guide:

http://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus5000/sw/layer2/421_n1_1/b_Cisco_n5k_layer2_config_gd_rel_421_n1_1/Cisco_n5k_layer2_config_gd_rel_421_n1_1_chapter8.html

Scenario 2: The core uplink switches are not virtualized

This scenario means connecting our VLAG peer LACP aggregation to two separate switches on the far end. These switches are most likely connected somehow in the core network (not shown in Figure 18 on page 20), but are neither directly connected using an ISL nor using a synchronizing MAC table to behave as a single logical switch. In order to prevent loops, the use of spanning-tree protocol on either the core/uplink switches or on both sets of switches is recommended, but not required.

The design is shown in Figure 18 on page 20.

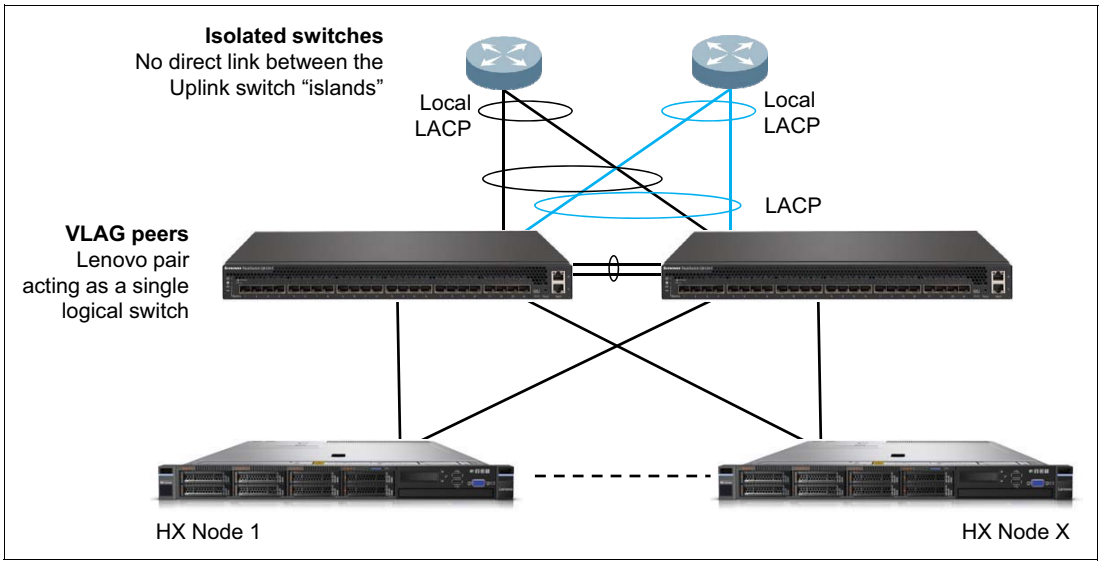

Figure 18 Scenario 2 design

The steps to configure the switches are as follows:

First, form two different VLAG aggregations on our Lenovo switch VLAG peers, in the same manner as in “Scenario 1: The core uplink switches are virtualized” on page 16. On each of the uplink switch “islands” we will create a simple LACP aggregation over 2 (or more) ports, which we will terminate on the aggregation split over our VLAG peers.

On the core network “island” switches, only a simple LACP aggregation is required (not active participation in VLAG/MC-LAG). These LACP are terminated over our local VLAG peers.

On the Lenovo VLAG peers, the participating ports in the aggregation need to be added on both switches. For example, if our ports 18 on both switches are joined together in a 2-port, cross-VLAG aggregation using the LACP admin key 80, the following configuration is required on *both* switches as shown in Figure 19 on page 20.

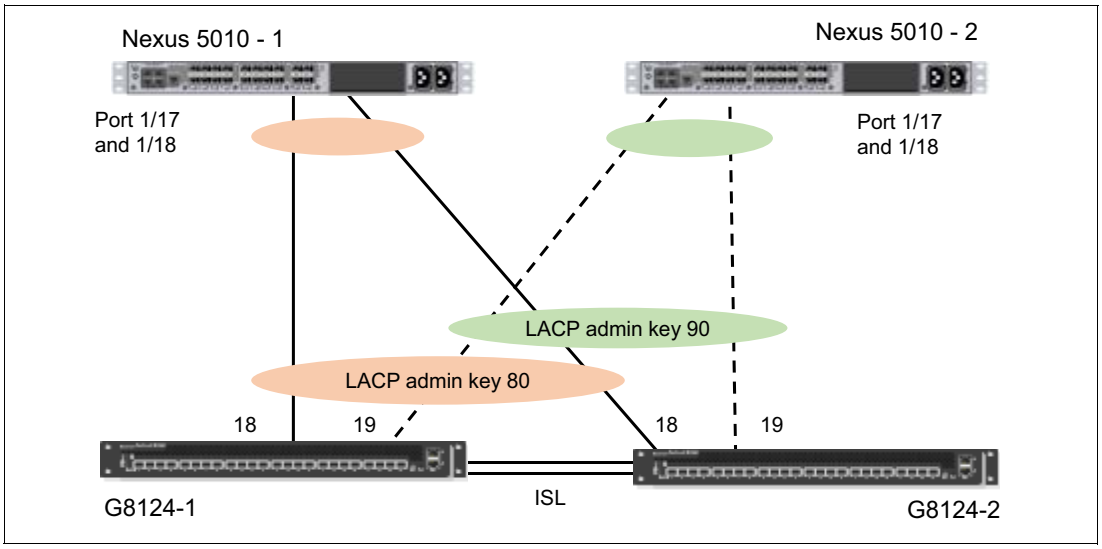

Figure 19 Connection of VLAG peers to non-virtualized uplink switches

The steps to configure the switches are as follows:

1. Issue the following commands on the Lenovo switches are as follows:

```
G8124E-1#(config)# int port 18
G8124E-1#(config)# lacp mode active
G8124E-1#(config)# lacp adminkey 80
G8124E-1#( (config-if)# vlag adminkey 80 enable

G8124E-2#(config)# int port 18
G8124E-2#(config)# lacp mode active
G8124E-2#(config)# lacp adminkey 80
G8124E-2#( (config-if)# vlag adminkey 80 enable
```

   This aggregation can be connected to a two-port local LACP aggregation on one of the core/uplink switches

2. On the core/uplink switch side, only a simple aggregation is required – in this example with 2 ports on each uplink switch on Cisco Nexus. The commands enable LACP on the Cisco switches:

```
switch# configure terminal
switch(config)# feature lacp
switch(config)# show feature
switch (config)# interface ethernet 1/17-1/18
switch(config-if)# channel-group 5 mode active
```

   For more information, see the “Configuring Port Channels” chapter of the Cisco NX-OS Software Configuration Guide:

   http://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus5000/sw/configuration/guide/cli/CLIConfigurationGuide/EtherChannel.html

3. For the aggregation to the other core/uplink switch, another port pair must be configured, as above, in this case with port 19 on each VLAG peer:

```
G8124E-1#(config)#int port 19
G8124E-1#(config)#lacp mode active
G8124E-1#(config)#lacp adminkey 90
G8124E-1#( (config-if)#vlag adminkey 90 enable

G8124E-2#(config)#int port 19
G8124E-2#(config)#lacp mode active
G8124E-2#(config)#lacp adminkey 90
G8124E-2#( (config-if)#vlag adminkey 90 enable
```

4. Check the status of the aggregation:

```
G8124E-1#sho lacp
G8124E-1#sho vlag info
```

Appendix A: VLANs and TCP Port reference

You may want to segregate Hypervisor/ESX internal traffic from HX cluster traffic through the use of VLANs, which in turn can be propagated to the customer uplink network in line with the customer's own VLAN design/policy.

Figure 20 on page 22 shows the ports that must be open for supported hypervisors. The diagram also shows ports that must be opened for infrastructure services.

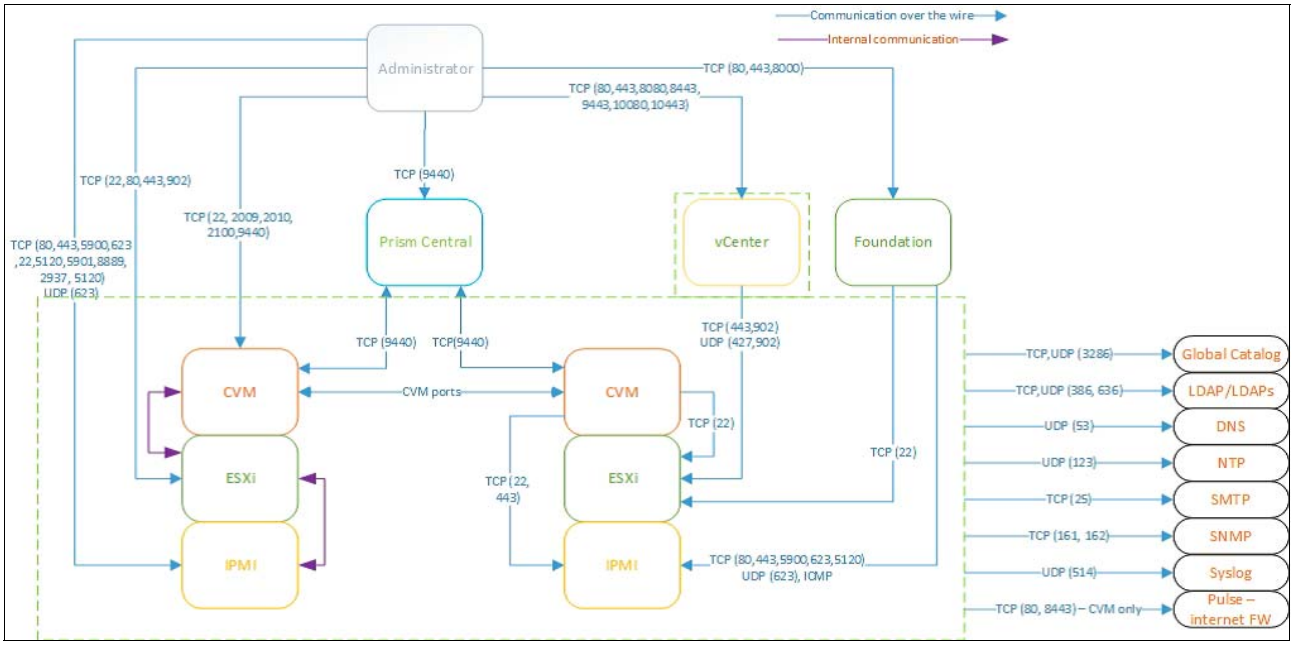

Figure 20 Nutanix Network Port Diagram for VMware ESXi

Appendix B: Acropolis Hypervisor considerations

The corresponding virtual switch setting to vSphere's "Route based on Originating Port" in the Acropolis hypervisor's OpenVswitch (OVS) is called "Balance-slb".

For details about Balance-slb, see the heading "Balance-slb" in the following blog post:

https://next.nutanix.com/t5/Nutanix-Connect-Blog/Network-Load-Balancing-with-Acropolis-Hypervisor/ba-p/6463

Appendix C: VMware Virtual Distributed Switch considerations

For considerations regarding implementing the Virtual Distributed Switch (VDS), which is an advanced feature of the VMware Enterprise+ license, please refer to:

http://www.mikelaverick.com/2014/02/back-to-basics-migrating-to-enhanced-lacp-part-3-of-5/

References and further reading

- Lenovo Networking and Options Catalog:

  http://www.lenovo.com/images/products/system-x/pdfs/datasheets/lenovo_networking_catalog_ds.pdf?menu-id=networking_options
- Nutanix TechNote – VMware vSphere Networking with Nutanix:

  http://go.nutanix.com/rs/nutanix/images/Nutanix_TechNote-VMware_vSphere_Networking_with_Nutanix.pdf
- Nutanix TechNote – Network Load Balancing with Acropolis Hypervisor:

  https://next.nutanix.com/t5/Nutanix-Connect-Blog/Network-Load-Balancing-with-Acropolis-Hypervisor/ba-p/6463
- Nutanix Field Installation Guide (imaging to VMware or Hyper-V from AHV)

  https://portal.nutanix.com/#/page/docs/details?targetId=Field_Installation_Guide-v3_1:Field_Installation_Guide-v3_1
- Lenovo Press – Lenovo Networking Best Practices

  https://lenovopress.com/sg248245
- Switch Application Guide and Switch Command Reference are two essential documents for Lenovo switches and can be found here, for each of the major firmware versions:

  http://publib.boulder.ibm.com/infocenter/systemx/documentation/index.jsp
- More general information on Lenovo Networking Switches:

  http://shop.lenovo.com/us/en/systems/networking/ethernet-rackswitch/

Change history

9 January 2017:

- Clarified that the Acropolis hypervisor is AHV and not KVM (AHV is based on KVM, however).
- Minor grammar corrections

7 December 2016:

- Corrections to the commands in step 10 on page 14 and step 12 on page 15

Authors

This paper was written by the following subject matter expert:

William Scull is a 30-year veteran of the IT industry and has been supporting x86 server connectivity since before IBM acquisition of BNT in 2010. He is currently the product manager for Lenovo networking in EMEA, based in Munich, Germany.

Thanks to the following people for their contributions to this project:

- Marijn Joosten, Nutanix
- Silvio Erdenberger, Lenovo
- David Watts, Lenovo Press

- Mark T. Chapman, Lenovo

Notices

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

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO 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 jurisdictions 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. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

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

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

Any performance data contained herein was determined in a controlled environment. Therefore, the result 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.

This document was created or updated on January 9, 2017.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp0546

Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at http://www.lenovo.com/legal/copytrade.html.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo(logo)® Lenovo® RackSwitch™

The following terms are trademarks of other companies:

Hyper-V, Windows, Windows Server, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.