VMware HCX Availability Guide

You can find the most up-to-date technical documentation on the VMware website at:

https://docs.vmware.com/

VMware, Inc.
3401 Hillview Ave. Palo
Alto, CA 94304
www.vmware.com

Contents

About This Document

The *VMware HCX Availability Guide* provides information to help HCX users understand the configurations and scenarios that affect the availability of virtual machines, virtual machine networks and VMware® HCX systems.

The purpose of this document is to analyze HCX service behavior and to explore configurations and architecture practices to maximize availability during planned and unplanned events. This information is for HCX Architects and Administrators and Consultants. It is assumed that readers have familiarity with VMware HCX, vSphere and NSX, and have basic knowledge of the systems underpinning HCX services.

Intended Audience

This information is intended for users who want to better understand and configure HCX availability. To provide considerations and guidance for configuring HCX. The information is written for systems administrators who are familiar with virtual machine technology and datacenter operations.

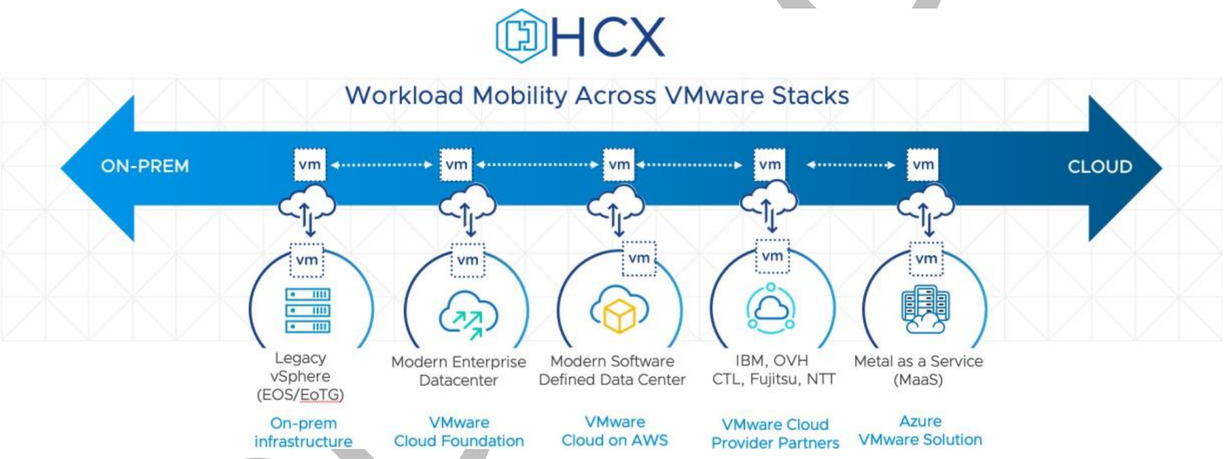
VMware Technical Publications Glossary

VMware Technical Publications provides a glossary of terms that might be unfamiliar to you. For definitions of terms as they are used in VMware technical documentation, go to http://www.vmware.com/support/pubs.

About VMware HCX

VMware HCX streamlines application migration, workload rebalancing and business continuity across data centers and clouds.

VMware HCX delivers secure and seamless application mobility and infrastructure hybridity across vSphere environments both on-premises and in the cloud. HCX abstracts on-premises and cloud resources and presents them as one continuous hybrid environment, enabling users to connect infrastructure and consume adopt a hybrid cloud vision, or a full migration to cloud as a consistent experience.


HCX
Workload Mobility Across VMware Stacks
ON-PREM
CLOUD
vm
Legacy vSphere (EOS/EoTG)
On-prem infrastructure
Modern Enterprise Datacenter
VMware Cloud Foundation
Modern Software Defined Data Center
VMware Cloud on AWS
IBM, OVH CTL, Fujitsu, NTT
VMware Cloud Provider Partners
Metal as a Service (MaaS)
Azure VMware Solution

Updated Information

This *VMware HCX Availability Guide* is updated with each release of the product or when necessary.

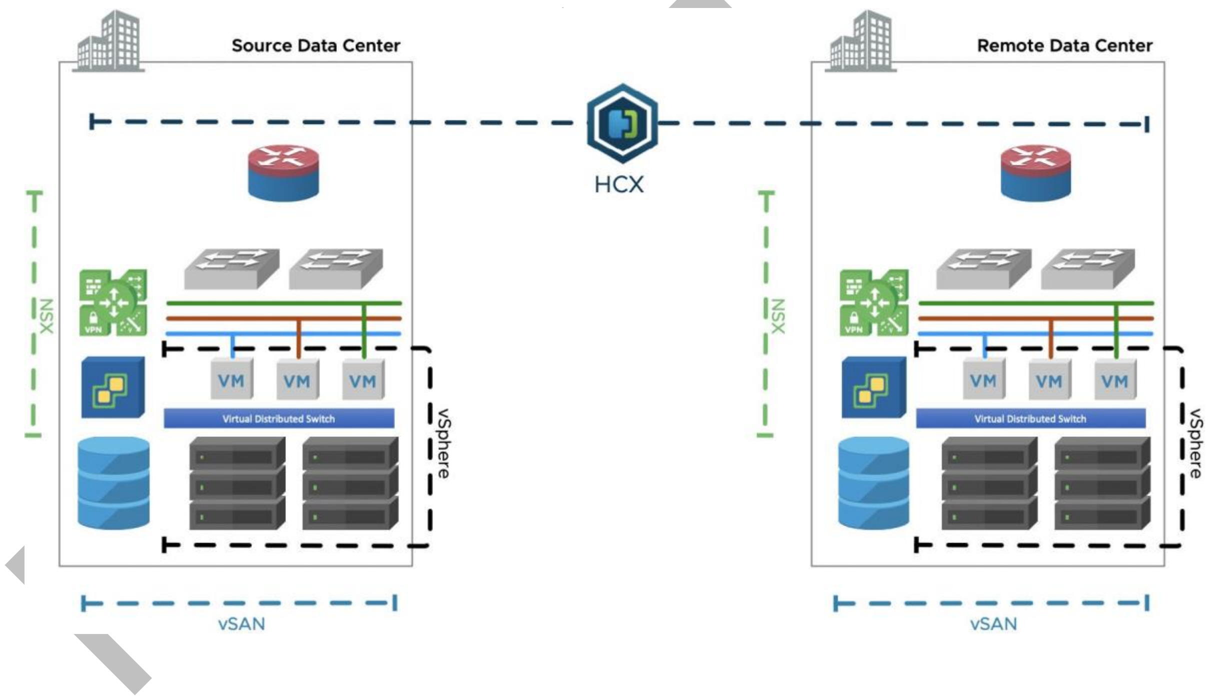
This table provides the update history of the *VMware HCX Availability Guide*.

Revision	Description
30 Jan 2021	HCX Availability Guide released. HCX Service Availability & Resiliency deprecated.

HCX Availability Configurations and Best Practices

This section includes known best practices for vSphere, NSX and HCX to improve the availability of HCX services and migrated workloads.

VMware HCX provides workload connectivity and mobility functionality in a proxy model, on behalf of distinct environments tethered by HCX. The HCX implementation generally presupposes that at least one source and destination environment is implemented following existing availability and business continuity best practices, and does not aim to dwell on the topics that will make the underlying vSphere and NSX components highly available.



Instead, throughout this document, we explore configurations in vSphere, NSX and HCX that directly contribute to the availability condition of HCX services (workload migrations in-flight, and any virtual machines relying on HCX Network Extension based connectivity

This chapter includes the following topics:

- vSphere Best Practices for HCX Availability
- NSX Best Practices for HCX Availability

- HCX Best Practices for Availability

vSphere Best Practices for HCX Availability

vSphere related considerations and best practices when designing vSphere environments to be used with highly available HCX deployments.

vSphere Distributed Resource Scheduler

DRS spreads the virtual machine workloads across vSphere hosts inside a cluster and monitors available resources. Fully automated DRS is a powerful vSphere feature that allows a cluster to automatically re-balance virtual machines across the cluster hosts.

Best practices:

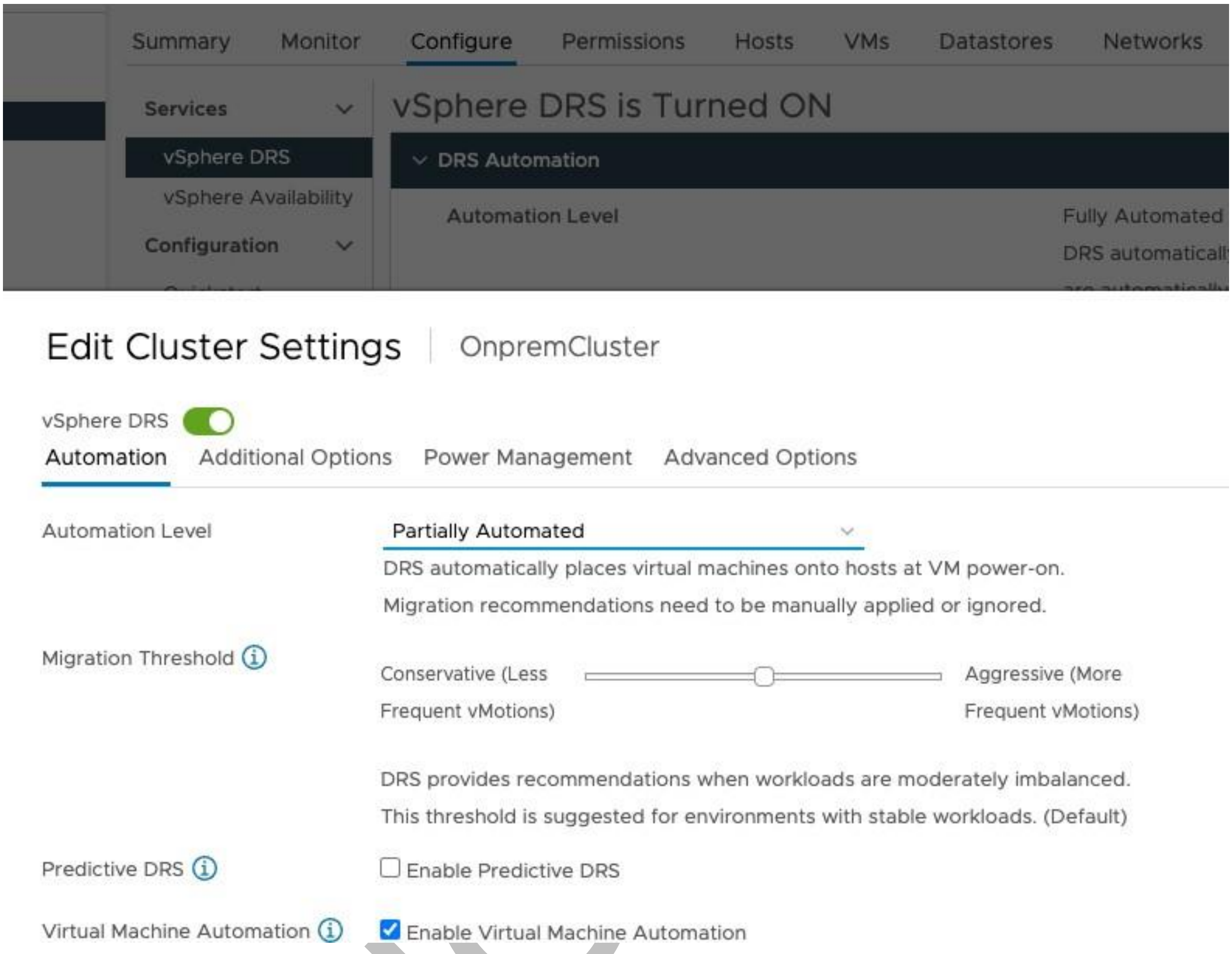
- Use **Partially Automated DRS** mode in the workload clusters being migrated with HCX. In this mode, DRS automatically places powered up virtual machines on a host that is guaranteed to provide the required resources, and make recommendations for rebalancing the cluster resources.

- Avoid the use of **Fully Automated DRS** mode. In this mode, DRS automatically applies rebalancing recommendations on powered on VMs. This mode can cause the following effects on HCX services:
  - Aggressive rebalancing "thrashing" of HCX migration and network extension appliances due to their smaller disk footprint.
  - Contention over vMotion resources between the DRS intra-cluster relocation and the HCX inter-cluster relocation.
  - Replication reconfigurations halt migration progress, disrupt checksum operations. A full synchronization may be triggered, delaying migration progress.
  - Excessive Network Extension flooding and active path changes.

- Using **DRS Affinity Rules** for HCX:
  - In deployments with multiple migration (HCX-IX) appliances, use host anti-affinity rules to allow parallel relocations.
- In deployments with multiple WAN Optimization appliances, use datastore anti-affinity, or ensure the datastores are able to meet the WAN Optimization IOPs requirements.
- Network Extension High Availability configures host anti-affinity rules automatically.

Note Virtual Machine Affinity Rules are not currently persistent to HCX lifecycle operations like redeployments and upgrades.

Figure 2-1. Configuring DRS Automation in the vCenter Server



vSphere DRS is Turned ON

Automation Level: Fully Automated

DRS automaticall

Edit Cluster Settings | OnpremCluster

vSphere DRS: ON

Automation Level: Partially Automated

DRS automatically places virtual machines onto hosts at VM power-on. Migration recommendations need to be manually applied or ignored.

Migration Threshold: Conservative (Less Frequent vMotions) — Aggressive (More Frequent vMotions)

DRS provides recommendations when workloads are moderately imbalanced. This threshold is suggested for environments with stable workloads. (Default)

Predictive DRS: ☐ Enable Predictive DRS

Virtual Machine Automation: ☑ Enable Virtual Machine Automation

ESXi Maintenance Mode

You place a host in maintenance mode when you need to service it, for example, to install more memory. A host enters or leaves maintenance mode only as the result of a user request.

Best practices:

- Perform required ESXi host maintenace after current replication based migrations have completed.
  - Replication reconfigurations halt migration progress, disrupt checksum operations. A full synchronization may be triggered, delaying migration progress.
  - A host cannot relocate a virtual machine during HCX Cold Migration or vMotion.
- Perform manual relocation of the HCX Nework Extension appliances during a maintenance window.

vSphere Cluster High Availability

vSphere Cluster HA provides high availability for virtual machines by pooling the virtual machines and the hosts they reside on into a cluster. Hosts in the cluster are monitored and in the event of a failure, the virtual machines on a failed host are restarted on alternate hosts.

Best practice:

- Use **vSphere HA** enabled clusters to deploy HCX components. See the vSphere Availability guide for more information.
- Configure HCX Network Extension appliances with **High HA Restart Priority**.

Fault Tolerance

FT provides continuous availability for such a virtual machine by creating and maintaining another VM that is identical and continuously available to replace it in the event of a failover situation.

Best practice:

- Do not use vSphere Fault Tolerance with HCX appliances, it is not supported for use with HCX Service Mesh components.

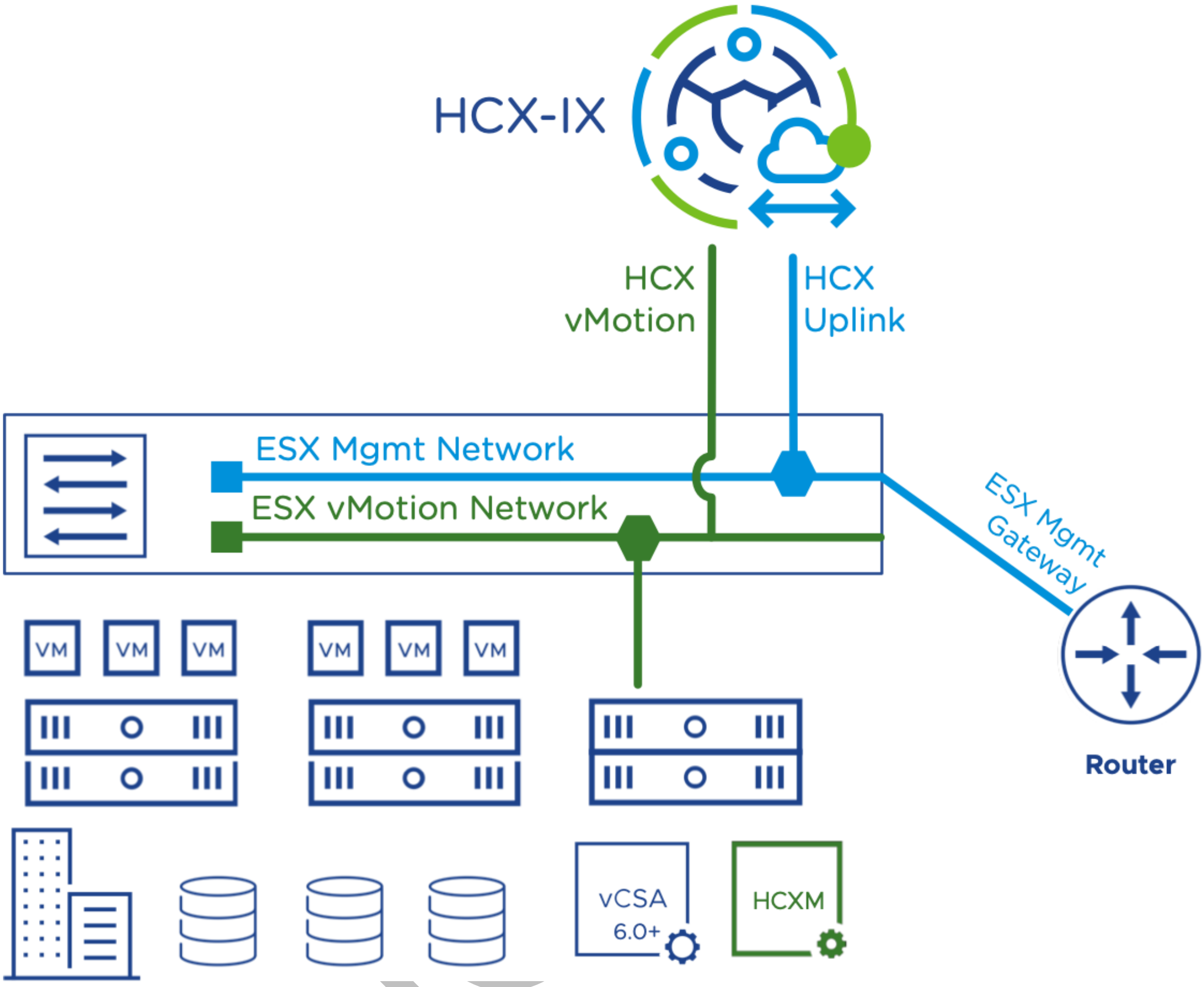
Workload VMkernel vMotion Network

HCX vMotion and Replication Assisted vMotion migrations use vMotion between the HCX migration appliance and the workload cluster vMotion network.

Best practice configuration for HCX deployments:

- Allocate IP addresses for the HCX migration appliance (HCX-IX) from existing vMotion networks to optimize the data path and simplify troubleshooting.
- When working with virtual machines connected to a VMware Standard Switch, verify that a vMotion Standard Port Group is created consistently on all hosts in the cluster.
- HCX operates at maximum availability when the underlying cluster vMotion configuration is implemented according to best practices. Please see networking best practices for vMotion for more information.

Figure 2-2. HCX vMotion vNIC connected to the ESXi vMotion Network



Workload VMkernel Replication Network

HCX Bulk Migration and Replication Assisted vMotion migrations use Host Based Replication between the HCX migration appliance and the workload cluster vmkernel management or replication network. It is common for the workload management vmkernel interface for replication traffic.

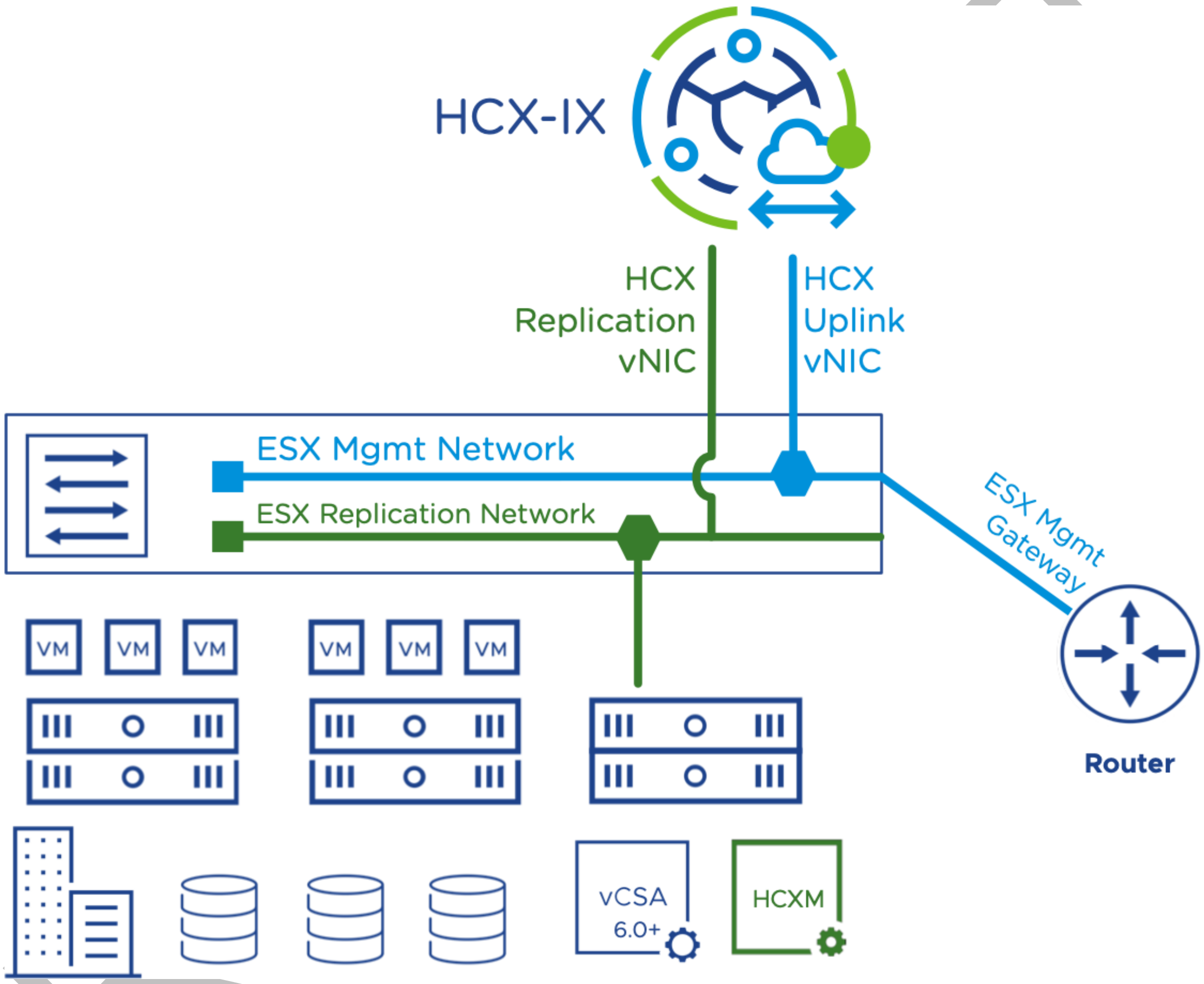
Best practices:

- Allocate IP addresses for the HCX migration appliance (HCX-IX) from existing Replication networks to optimize the data path and simplify troubleshooting.

- When designing cluster networking, use a dedicated cluster replication network . A workload replication vmkernel network can be designed using the same principles as the networking best practices for vMotion. This allows data transfer traffic to be separated from management workflows.

Note When a dedicated Replication VMkernel network is not present, the cluster will use the management network for replication traffic.

Figure 2-3. HCX Replication vNIC connected to the ESXi Replication Network



Using Dedicated HCX Mobility Clusters

This refers to the practice of implementing HCX with dedicated cluster hosts for the migration and network extension services. Mobility clusters do not contend with the workload virtual machines that will be migrated. This configuration maximizes overall resiliency and recovery by eliminating contention with workloads and optimizes traffic distribution by separating migration and network extension traffic flows from non-migrated workload traffic.

- Clusters designed specifically for HCX mobility should meet requirements and best practices for vSphere High Availability. For more information See the vSphere Availability Guide.
- Deployment requirements for mobility clusters:
  - A mobility cluster and the workload clusters must be managed by the same vCenter Server.
  - The mobility cluster is selected as the deployment cluster the Compute Profile configuration.
  - The workload clusters are selected as the service clusters in the Compute Profile configuration.
- Requirements for HCX Network Extension with mobility clusters:
  - The mobility cluster hosts must have access to the workload VLANs by association to the existing workload clusters either joined to existing workload Distributed Switches, or by ensuring the new Distributed Switch and Distributed Portgroups can connect to the same underlying VLANs.
- Requirements for HCX Bulk migration:
  - The migration appliance on the mobility cluster hosts must have IP reachability to the workload cluster vmkernel interfaces for Replication/NFC.
- For Cold Migration, HCX vMotion and Replication-Assisted vMotion:
  - The migration appliance on the mobility cluster hosts must have routed reachability to the workload cluster management and vMotion vmkernel interfaces.
  - The mobility cluster host must belong to the workload Distributed Switch.
- In scaleout deployments where multiple service meshes share one mobility cluster:
  - Each workload cluster can only have one migration appliance, multiple workload clusters can share a mobility cluster, but he migration appliances must run on different hosts to provide increased relocation (HCX vMotion/RAV) concurrency.

NSX Best Practices for HCX Availability

VMware NSX-T related considerations and best practices when designing for highly available HCX deployments.

Working with NSX Distributed Firewall

Distributed firewall monitors all the East-West traffic on your virtual machines. Best

practices:

- Use permisive firewall policies to decouple the workload migration from the firewall policy migration.

- In migrations with strict zero-trust microsegmentation, or other restrictive firewall objectives, apply the security polices prior to migrating the workloads.

- Use Security Tags for dynamic Security Group membership. In NSX to NSX migrations, HCX will migrate the NSX Security Tag.

NSX Edge Transport Nodes

The NSX Edge provides routing services and connectivity to network NSX Edges that are external to the NSX-T Data Center deployment.

Best practices:

- HCX migrations and network extensions can add multi-Gbps traffic over a WAN. Separate HCX traffic from application traffic whenever possible to reduce network contention scenarios.

Segment Policies

When migrating virtual machines that rely on Ensure that NSX Segment Policies are not blocking the DHCP requests.

Best practices:

- When migrating virtual machines that depend on DHCP (without Mobility Optimized Networking), use NSX segment configurations that allow DHCP requests to be forwarded to the source gateway.

- When migrating virtual machines that depend on DHCP (with Mobility Optimized Networking), use NSX segment configurations that provide DHCP services. See Configure DHCP on a segment.

Route Advertisements

A tier-1 gateway is typically connected to a tier-0 gateway in the northbound direction and to segments in the southbound direction. Route advertisements can optionally be configured on the tier-1 gateway.

Best practices:

- When using Mobility Optimized Networking, verify that the **All Static Routes** in the Tier-1's **Route Advertisement** is configured to advertise or not advertise according to the design (MON migrations will always add virtual machine static routes). For more information, see HCX Network Extension with Mobility Optimized Networking for NSX-T in the HCX User Guide.

HCX Best Practices for Availability

VMware HCX configurations and best practices for highly available HCX deployments.

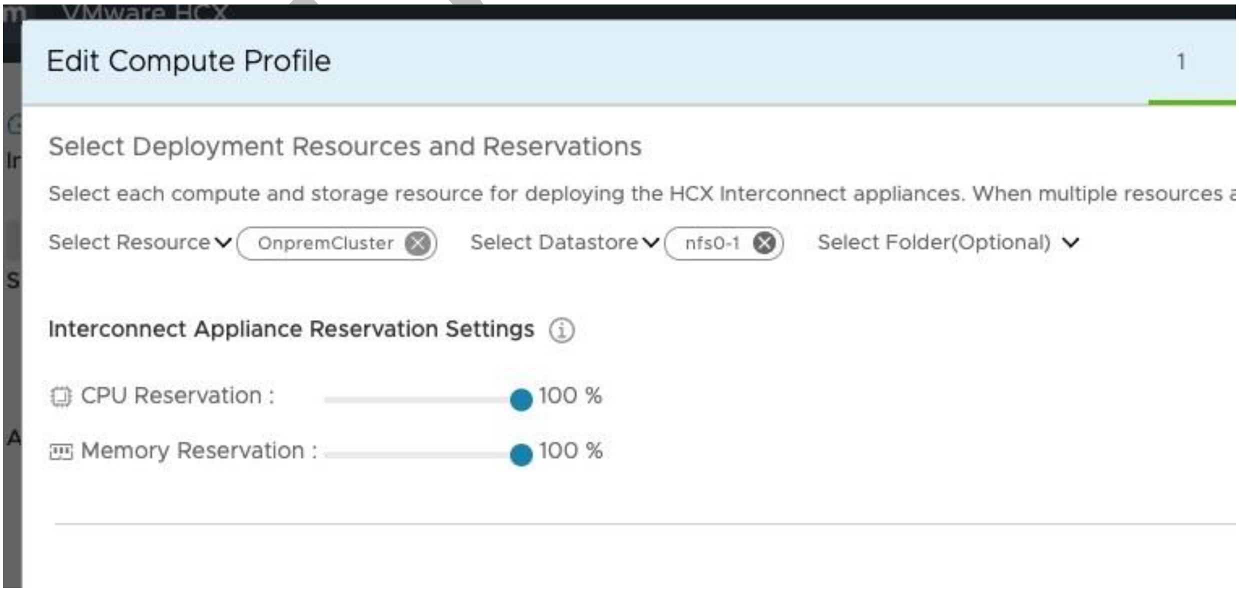
Compute Profile Resource Reservations

A **Compute Profile** contains the compute, storage, and network settings that HCX uses on this site to deploy the Interconnect-dedicated virtual appliances when a Service Mesh is added. A Compute Profile can be used to define CPU and Memory.

Best practices:

- Apply 100% CPU and Memory resouce reservations when Network Extension appliances are sharing resources with workload virtual machines.
- Use the HCX compute profile to configure CPU and Memory reservations. Resource reservations configured directly in vCenter Server are not persistent to HCX lifecycle operations.

Figure 2-4. Configuring CPU and Memory Reservations in the HCX Compute Profile



Edit Compute Profile

Select Deployment Resources and Reservations

Select each compute and storage resource for deploying the HCX Interconnect appliances. When multiple resources a

Select Resource: OnpremCluster

Select Datastore: nfs0-1

Select Folder(Optional)

Interconnect Appliance Reservation Settings

CPU Reservation : 100 %

Memory Reservation : 100 %

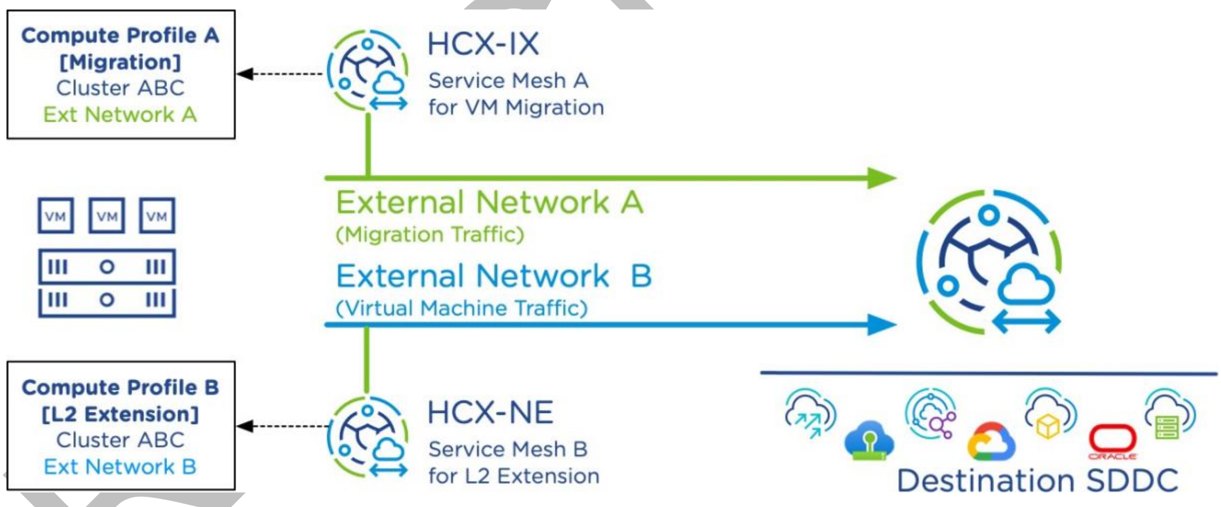
Configuring Dedicated Service Mesh

Typically a single **Service Mesh** is configured for both migration and network extension using a single Compute Profile configuration for both services. HCX allows Network Extension services to be separated into a second Service Mesh, with different uplinks and resource assignments.

Best Practices:

- Design for Migration Data traffic to be separate from Virtual Machine traffic Port Groups.
    - Using distinct vSphere Networks for migration egress enables configurations where network extension traffic is prioritized, and where migration traffic and resource is isolated and appropriate network policies can be implemented.
    - Create the first compute profile and service mesh to enable migration services only.
    - Create a second compute profile and service mesh to enable Network Extension services.
    - This configuration allows the deployed appliances to use network and compute resources and network policies to be tailored to the service mesh function.

Figure 2-5. Dedicated Service Mesh for HCX Extension and HCX Migration



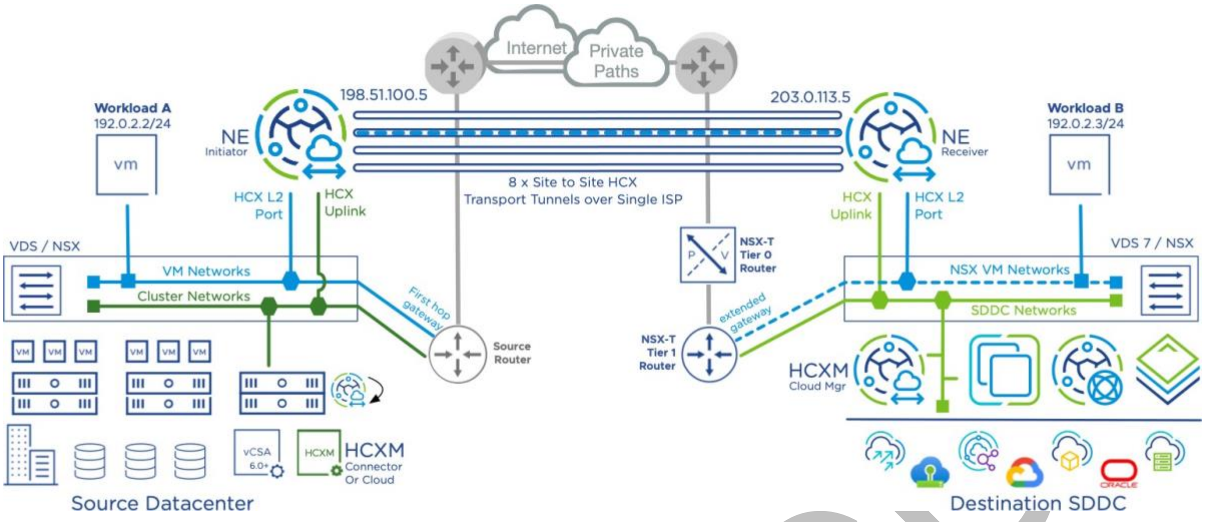
Application Path Resiliency

Network path resiliency for single vNIC HCX Service Mesh deployments.

Application Path Resilency (APR) is an HCX Enterprise capability that creates additional site to site tunnels per uplink. HCX tracks the condition of each tunnel and avoids sending traffic over degraded or black hole paths. When APR is enabled, eight tunnels are used in every migration appliance (HCX-IX) and network extension (HCX-

NE) appliance in the service mesh. This enables environments that do not have access to multiple network providers or paths an option to achieve additional uplink path resiliency.

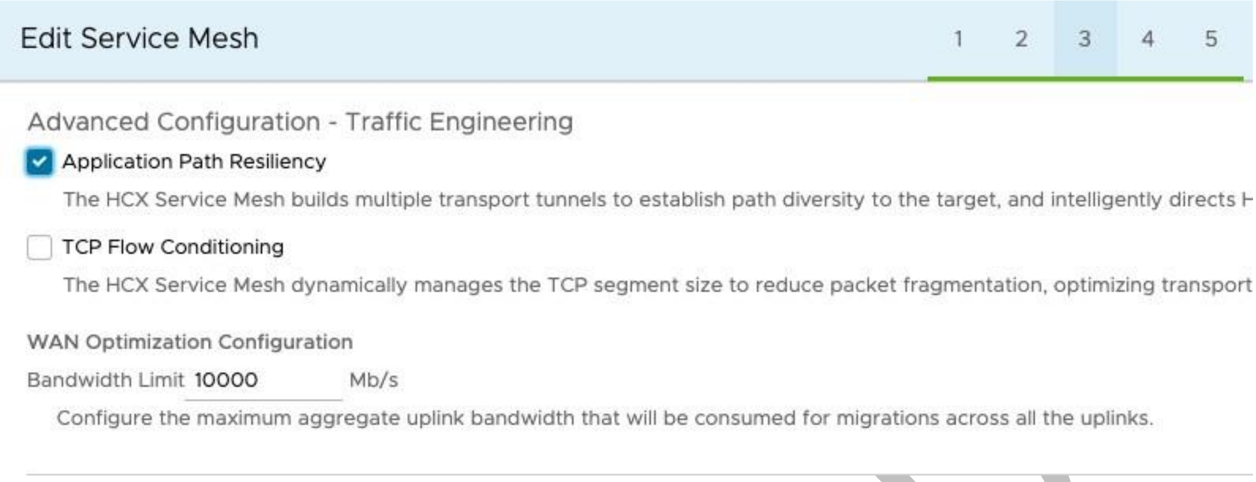
Figure 2-6. Single Uplink Multi-Tunneling for Network Path Resiliency



Best Practices:

- Enable APR in single uplink HCX Service Mesh configurations.
- Enable APR in single ISP configurations.
- Enable APR in single Direct Connect/Private Path configurations.
- Enable APR for network path resiliency. APR is not designed to provide link aggregation or load balancing functionality.
- Using APR may not be optimal, or may be excessive in deployments where HCX is connected to diverse network paths for resiliency using multiple uplinks.
  - As an example, in deployments with 3 uplinks (e.g. Direct Connect 1, Direct Connect 2 & Internet), using APR will result in maintaining 24 service mesh tunnels.
- Firewalls typically do not need to be adjusted for APR. The following characteristics apply:
  - Eight tunnels created per HCX Uplink.
  - Same destination IP address and destination port for the eight APR tunnels.
  - Same source IP address and variable port for the eight APR tunnels.

Figure 2-7. Configuring Application Path Resiliency in the Service Mesh



Edit Service Mesh 1 2 3 4 5

Advanced Configuration - Traffic Engineering

☑ Application Path Resiliency

The HCX Service Mesh builds multiple transport tunnels to establish path diversity to the target, and intelligently directs H

☐ TCP Flow Conditioning

The HCX Service Mesh dynamically manages the TCP segment size to reduce packet fragmentation, optimizing transport

WAN Optimization Configuration

Bandwidth Limit 10000 Mb/s

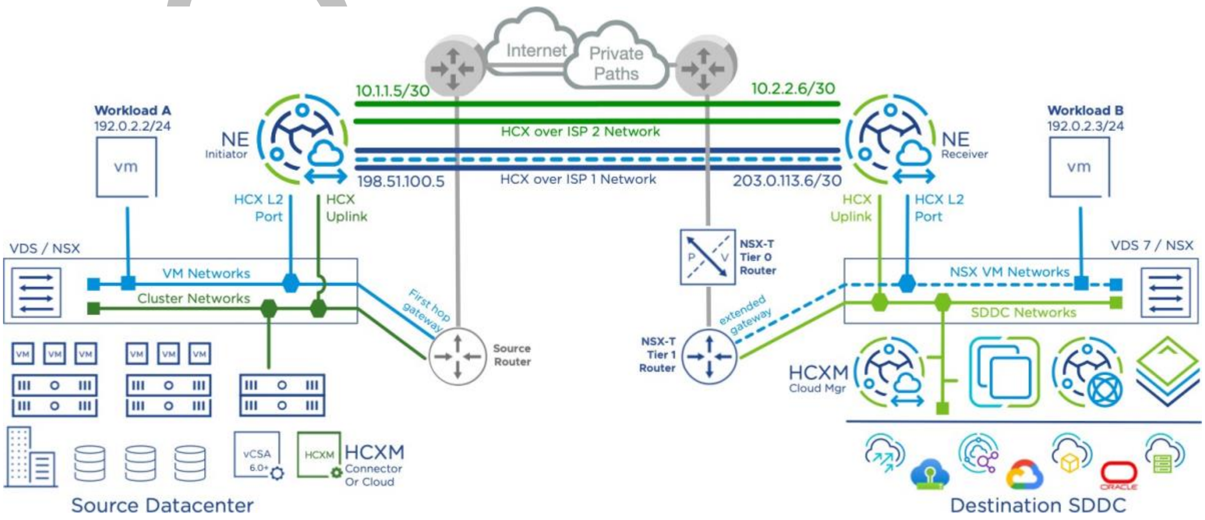
Configure the maximum aggregate uplink bandwidth that will be consumed for migrations across all the uplinks.

Multi-Uplink Service Mesh Resiliency

Network path resiliency with multi vNIC HCX Service Mesh deployments.

A multi-uplink service mesh deployment refers to a configuration that uses two or a maximum of three uplink vNICs for site to site traffic for network path resiliency. Additional uplinks are automatically added to every migration appliance (HCX-IX) and network extension appliance (HCX-NE) in the service mesh. This enables resilient connectivity in environments with multiple network underlays or private lines, or even fallback to internet to be used for resilient site to site deployments.

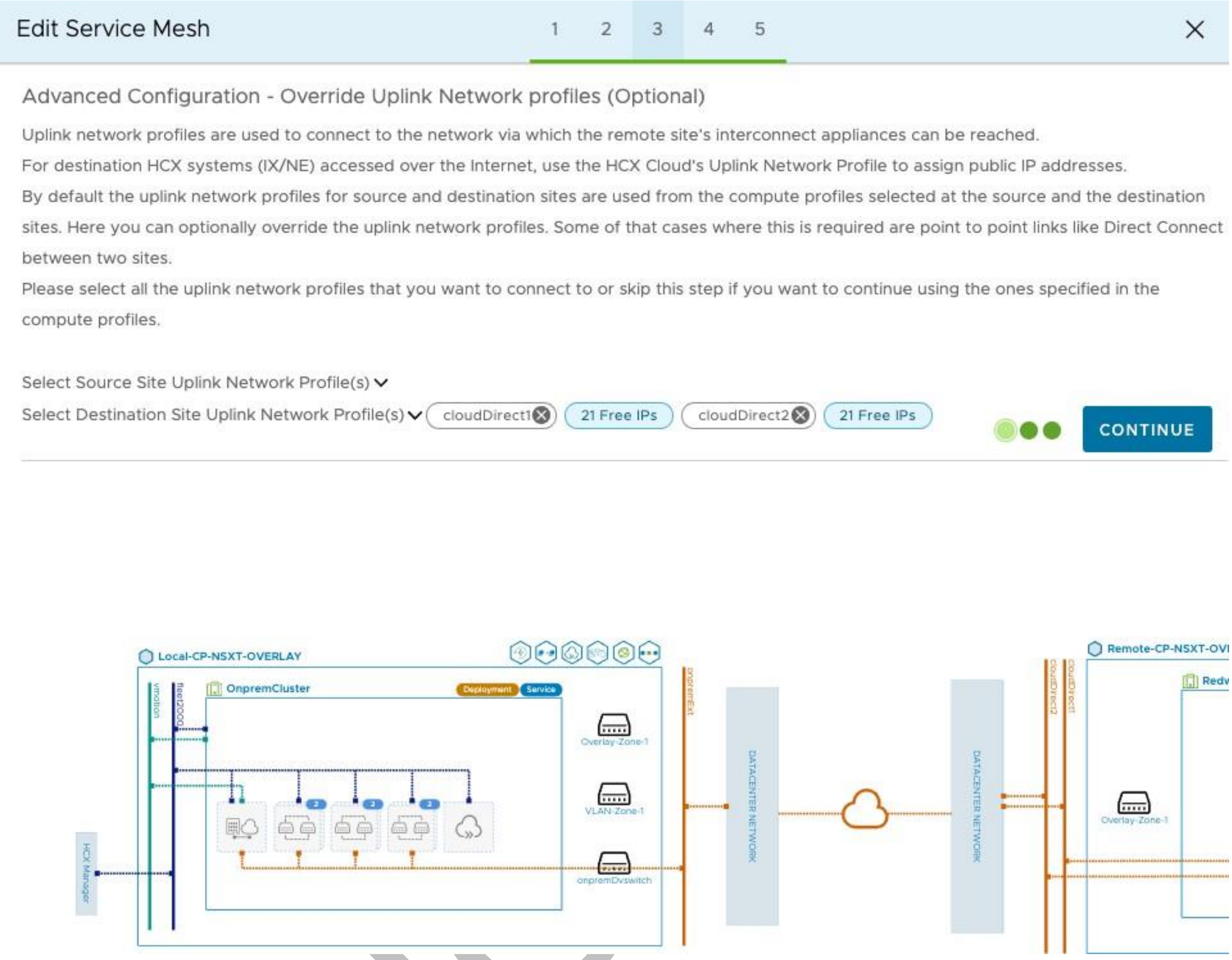
Figure 2-8. Multi-Uplink Service Mesh for Network Path Resiliency



Best Practices:

- Use a multi-uplink configuration for environments with access to different network underlays between the environments (up to 3).
- Use a multi-uplink configuration to configuring connections on private network underlays with fallback to the public internet.
- In multi-uplink configurations, every HCX Uplink network must connect to a unique VLAN/VNI and a IP subnet.
- Multiple uplinks can be assigned in either the source or destination environment with the same resiliency outcome. For example:
    - Assigning two uplinks at the source and one at the destination will yield two service mesh tunnels.
    - Assigning one uplink at the source and two at the destination will also yield two service mesh tunnels.
- A tunnel will be created for every unique combination of uplinks. As an example, assigning two uplinks at the source and two uplinks at the destination will yield four tunnels:
    - Source Uplink 1 <--> Destination Uplink 1
    - Source Uplink 1 <--> Destination Uplink 2
    - Source Uplink 2 <--> Destination Uplink 1
    - Source Uplink 2 <--> Destination Uplink 2
- Using Application Path Resiliency may be inefficient with multi-uplink configurations, 8 tunnels will be created per unique uplink set. In the previous example with 4 unique uplink combinations, enabling APR will result in maintaining 32 transport tunnels for resiliency.

Figure 2-9. Configuring a Service Mesh with Multiple Uplinks



Edit Service Mesh

Advanced Configuration - Override Uplink Network profiles (Optional)

Uplink network profiles are used to connect to the network via which the remote site's interconnect appliances can be reached.

For destination HCX systems (IX/NE) accessed over the Internet, use the HCX Cloud's Uplink Network Profile to assign public IP addresses.

By default the uplink network profiles for source and destination sites are used from the compute profiles selected at the source and the destination sites. Here you can optionally override the uplink network profiles. Some of that cases where this is required are point to point links like Direct Connect between two sites.

Please select all the uplink network profiles that you want to connect to or skip this step if you want to continue using the ones specified in the compute profiles.

Select Source Site Uplink Network Profile(s)

Select Destination Site Uplink Network Profile(s): cloudDirect1 ⊗ 21 Free IPs, cloudDirect2 ⊗ 21 Free IPs

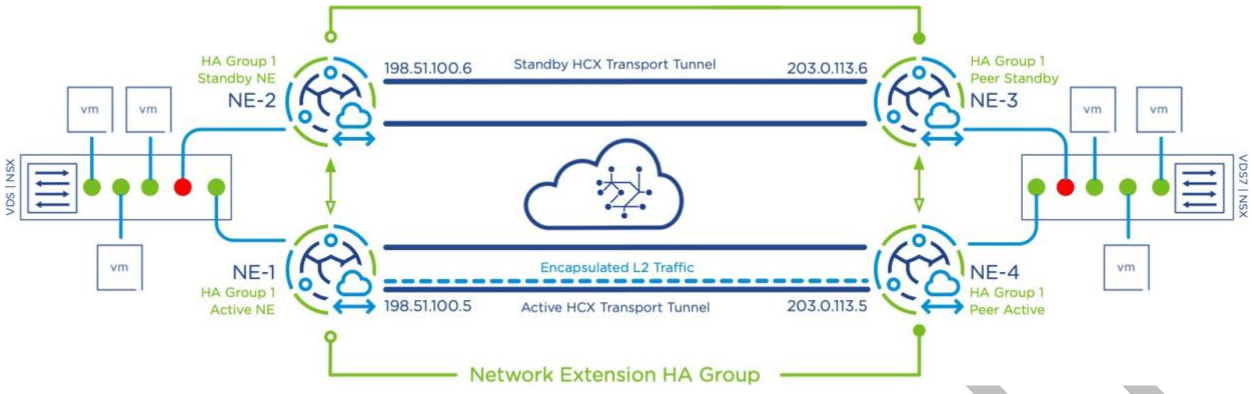
Network Extension High Availability

Appliance failure tolerance for HCX Network Extension.

Network Extension High Availability is an HCX Enterprise capability that groups Network Extension appliances into resilient pairs (called an HA Group). When Network Extension High Availability is enabled for a selected appliance, HCX will pair it with an eligible appliance and enable an Active Standby resiliency configuration. This enables highly available configurations that can remain in-service in the event of an unplanned appliance level failure.

When either of the HA Actives fail, both standby appliances take over. The Network Extension High Availability is designed to recover within a few seconds after a single appliance has failed.

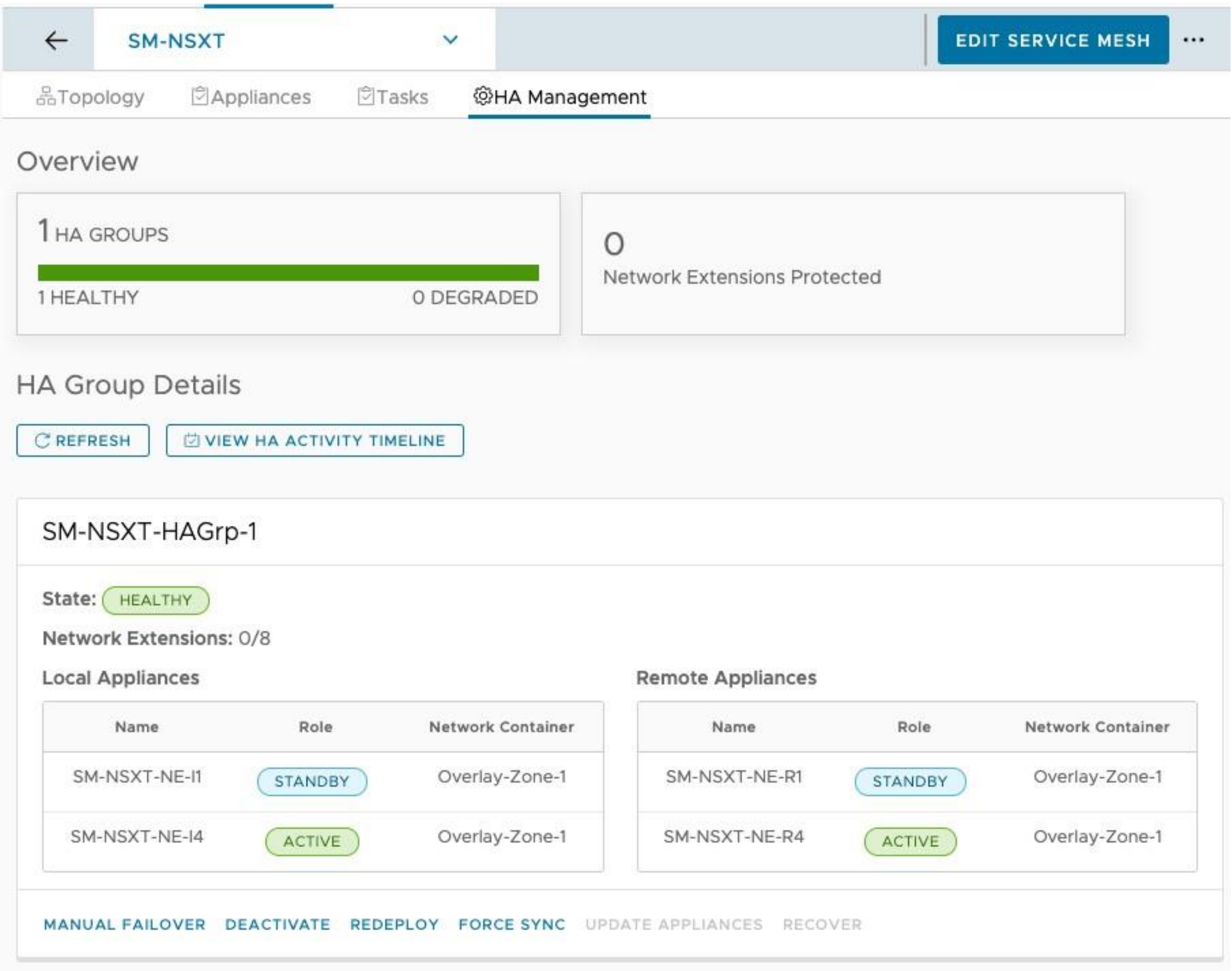
Figure 2-10. Network Extension in a High Available Active Standby Configuration



Best practices:

- Plan for additional provisioning requirements:
  - Plan for increased appliance count. Each site will have a local HA Active and Standby. Four appliances per group.
  - Plan IP address consumption in the context of HA Groups. Both the HA Active and HA Standby appliances will establish service mesh transport tunnels with their remote peers.
- Plan for additional ESXi Host requirements:
  - Network Extension HA creates host anti-affinity rules. The selected ESXi cluster should be licensed for DRS capabilities.
  - The ESXi host running the HA Standby appliance should have enough capacity reserved to take over the forwarding load of the HA Active appliance.
- Always use an uplink redundancy strategy for complete redundancy (e.g. Application Path Resiliency or Multi-Uplink), HA heartbeats will be delivered using all existing networks. A failed uplink will not result in declaring an HA node failure if the heartbeats are received on one of the other interfaces.

Figure 2-11. Active Network Extension HA Configuration



Overview

1 HA GROUPS

1 HEALTHY 0 DEGRADED

0
Network Extensions Protected

HA Group Details

SM-NSXT-HAGrp-1

State: HEALTHY

Network Extensions: 0/8

Local Appliances

Name	Role	Network Container
SM-NSXT-NE-I1	STANDBY	Overlay-Zone-1
SM-NSXT-NE-I4	ACTIVE	Overlay-Zone-1

Remote Appliances

Name	Role	Network Container
SM-NSXT-NE-R1	STANDBY	Overlay-Zone-1
SM-NSXT-NE-R4	ACTIVE	Overlay-Zone-1

Network Extension In-Service Upgrade

Minimal downtime for Network Extension lifecycle operations.

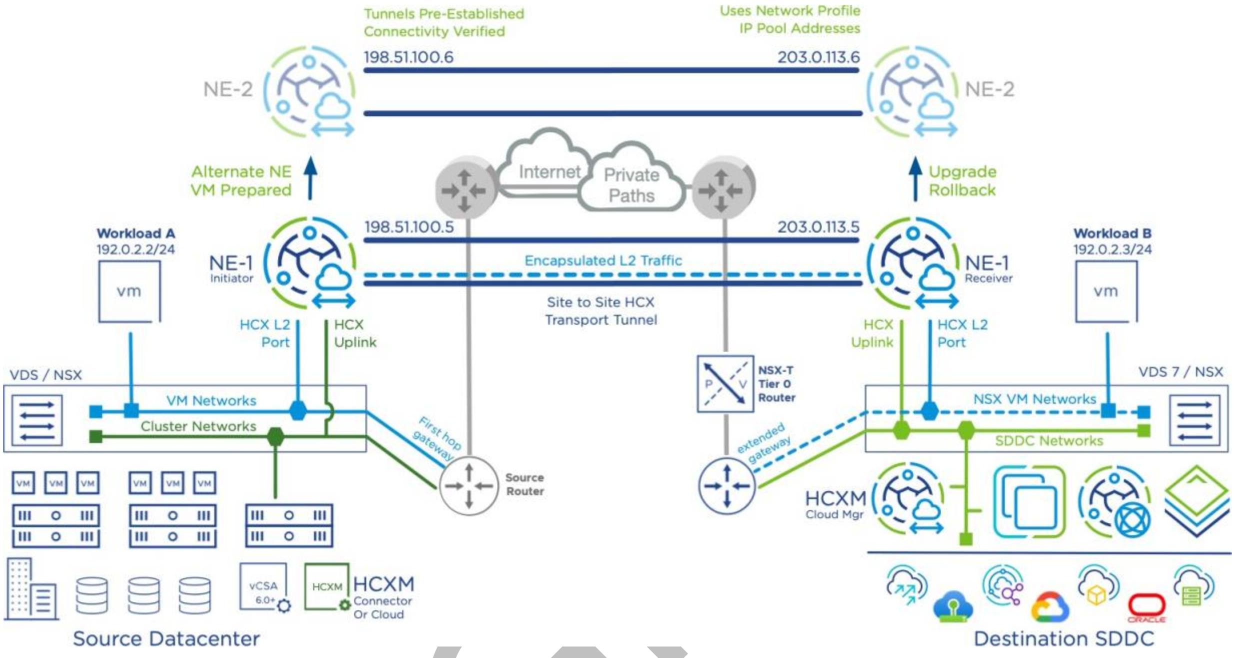
The Network Extension appliance is a critical component of many HCX deployments, providing workload connectivity during a migration and in multi-cloud deployments. The **Network Extension In-Service Upgrade** (NE ISSU) capability enables the Network Extension component to receive critical patches and capability upgrades while minimizing impact to connected virtual machines. Network Extension ISSU is designed to recover Network Extension within a second or a few seconds maximum.

When In-Service upgrade or redeployment is selected, the following applies:

- A new NE appliance is provisioned at the source and destination site.
- New Uplink and Management IP addresses are assigned for each new Network Extension appliance.
- The NICs on the new appliances are connected, including bridge NICs for extended networks (flagged down)

- Secure tunnel connections are established and verified between the sites.
- The old appliance Bridge NICs are disconnected. And new Appliances Bridge NICs are connected.
- The old appliance is deleted. The IP addresses used for the old appliance are released back into the IP Pool.

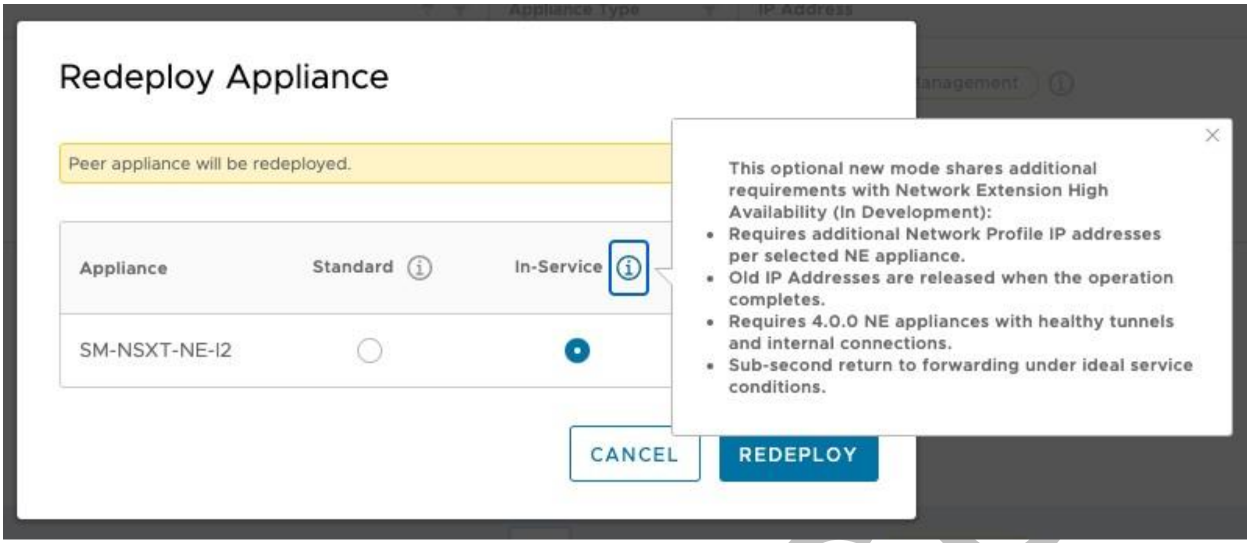
Figure 2-12. ISSU New Appliances are Connected and Verified



Best practices:

- Always use the ISSU option for Network Extension upgrade and redeploy operations to minimize downtime.
- Plan for additional IP addresses in the Network Profile IP Pools to acommodate the In-Service operations.
- Verify that HCX related firewall rules are configured for the full IP pool ranges.

Figure 2-13. Selecting the In-Service Operation



Redeploy Appliance

Peer appliance will be redeployed.

Appliance	Standard	In-Service
SM-NSXT-NE-I2	○	●

This optional new mode shares additional requirements with Network Extension High Availability (In Development):

- Requires additional Network Profile IP addresses per selected NE appliance.
- Old IP Addresses are released when the operation completes.
- Requires 4.0.0 NE appliances with healthy tunnels and internal connections.
- Sub-second return to forwarding under ideal service conditions.

About Standard Upgrades

The **Standard** option creates new Service Mesh appliances using the current IP addresses. It requires the new Service Mesh appliances to be fully disconnected until the final steps and requires tunnel negotiation to delay until the new appliances are connected.

When this option is used for Network Extension upgrades, virtual machine network connections will time out.

Best Practice:

- Use ISSU for Network Extension upgrades.
- Plan for a few minutes VM downtime when Network Extension appliances will be upgraded using the Standard option.

About the Author

Authors

Gabe Rosas is a Staff Technical Product Manager for VMware HCX in the Networking and Advanced Security Business Group, at VMware. He is experienced in multi-cloud, workload migration, designing and optimizing classic and software defined infrastructure.