APRIL 2022

VMWARE NSX-T ® MULTI-LOCATIONS REFERENCE DESIGN GUIDE

Software Version 4.1

Table of Contents

1 Introduction 8

2 What are the NSX-T solutions for Multi-Locations 13

2.1 NSX-T Multisite 13

2.2 NSX-T Federation 14

2.3 Summary of Pros/Cons of NSX-T Multi-Locations solutions 15

3 NSX-T Multisite 16

3.1 Architecture components 17

3.1.1 Management Plane 17

3.1.1.1 Management Cluster Deployment Mode1: Metropolitan Region 17

3.1.1.1.1 Management Cluster Deployment Mode1 - Option1: NSX-T Managers VMs deployed in two locations 17

3.1.1.1.2 Management Cluster Deployment Mode1 – Option2: NSX-T Managers VMs deployed in three locations 18

3.1.1.2 Management Cluster Deployment Mode2: Large Distance Region 18

3.1.2 Data Plane 19

3.2 Network & Security services supported 21

3.2.1 Multisite Network Services 21

3.2.2 Multisite Security Services 23

3.3 Best Practice Design 26

3.3.1 Management Plane 26

3.3.1.1 Management Cluster Deployment Mode1: Metropolitan Region 26

3.3.1.1.1 Management Cluster Deployment Mode1 - Option1: NSX-T Managers VMs deployed in two locations 26

3.3.1.1.2 Management Cluster Deployment Mode1 – Option2: NSX-T Managers VMs deployed in three locations 27

3.3.1.2 Management Cluster Deployment Mode2: Large Distance Region 28

3.3.1.2.1 Management Cluster Deployment Mode2 – Option1: L2-VLAN Management stretch 29

3.3.1.2.2 Management Cluster Deployment Mode2 – Option2: No L2-VLAN Management stretch 30

3.3.2 Data Plane 31

VMware, Inc. – www.vmware.com
Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.3.2.1 Edge Cluster Deployment Mode1: Stretched Edge Clusters with Edge Nodes deployed in different failure domains 31

3.3.2.2 Edge Cluster Deployment Mode2: Non-Stretched Edge Clusters with Edge Nodes deployed in no failure domain 33

3.3.2.3 (Special Use Case) Edge Cluster Deployment Mode3: Stretched Edge Cluster Cross Locations 35

3.4 Disaster Recovery 40

3.4.1 Management Plane 40

3.4.1.1 Management Cluster Deployment Mode1: Metropolitan Region 40

3.4.1.1.1 Management Cluster Deployment Mode1 - Option1: Two locations or more with recovery via vSphere HA or SRM 40

3.4.1.1.2 Management Cluster Deployment Mode1 – Option2: Two locations with recovery via Manager cluster deactivation 45

3.4.1.1.3 Management Cluster Deployment Mode1 – Option3: Three locations with recovery via Managers distributed across all 47

3.4.1.2 Management Cluster Deployment Mode2: Large Distance Region 48

3.4.1.2.1 Management Cluster Deployment Mode2 - Option1: Two locations or more with recovery via SRM 49

3.4.1.2.2 Management Cluster Deployment Mode2 – Option2: Two locations or more with recovery via FQDN + backup/restore 51

3.4.2 Data Plane 53

3.4.2.1 Edge Cluster Deployment Mode1: Stretched Edge Clusters with Edge Nodes deployed in different failure domains 53

3.4.2.2 Edge Cluster Deployment Mode2: Non-Stretched Edge Clusters with Edge Nodes deployed in no failure domain 58

3.4.2.3 (Special Use Case) Edge Cluster Deployment Mode3: Stretched Edge Cluster Cross Locations 62

3.4.2.4 Compute VMs Recovery 68

3.4.2.5 (Special Use Case) Network Introspection and Endpoint Protection 71

3.4.3 What about GSLB option 72

3.5 Requirements and Limitations 76

3.6 Orchestration / Eco-System 77

3.7 Scale and Performance guidance 77

4 NSX-T Federation 78

4.1 Architecture components 79

4.1.1 Management Plane 79

4.1.1.1 GM Cluster Deployments 79

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.1.1.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations 79
4.1.1.1.2 GM Cluster Deployment Mode2: NSX-T GM-Active and GM-Standby 80
4.1.1.2 GM, LM, Edge Node Communication Flows 81
4.1.1.2.1 GM-Active to GM-Standby Communication Flow 82
4.1.1.2.2 GM to LM Communication Flow 83
4.1.1.2.3 LM to LM Communication Flow 84
4.1.1.2.4 Edge Node to Edge Node Communication Flow 86
4.1.1.3 LM Registration and LM Onboarding 87
4.1.1.3.1 LM Registration (Addition) 87
4.1.1.3.2 (Optional) LM Onboarding (Import) 91
4.1.1.4 Federation Regions 96
4.1.1.5 Logical Configuration and Infrastructure Ownership 97
4.1.1.5.1 Logical Configuration Ownership 97
4.1.1.5.2 Infrastructure Ownership 103
4.1.1.6 Federation API 104
4.1.2 Data Plane 107
4.2 Network & Security services supported 109
4.2.1 GM Network Services 110
4.2.1.1 Network Objects Span 110
4.2.1.2 L2 Overlay Switching Service 112
4.2.1.2.1 GM Segment Configuration Options 112
4.2.1.2.2 GM Segment Data Plane 112
4.2.1.3 L3 Routing Service 114
4.2.1.3.1 GM Tier-0 and Tier-1 Gateway Configuration Options 114
4.2.1.3.2 Tier-0 Data Plane (South/North) 119
4.2.1.3.3 Tier-1 with Service Data Plane (South/North) 125
4.2.1.3.4 East/West with Service Data Plane 127
4.2.1.3.5 Routing Protocols 130
4.2.1.4 Stateful NAT Service 141
4.2.1.5 DHCP (Relay and Static Binding) and DNS 143
4.2.1.5.1 DHCP Server with DHCP Relay 143
4.2.1.5.2 DHCP Server with DCHP Static Bindings 144
4.2.1.5.3 DNS Service 145
4.2.1.6 Load Balancing service (Avi) 146

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.2.2 GM Security Services 155
4.2.2.1 GM Groups 155
4.2.2.2 GM Distributed Firewall (DFW) 156
4.2.2.3 GM Gateway Stateful Firewall 161
4.3 Best Practice Design 164
4.3.1 Management Plane 164
4.3.1.1 GM Cluster VMs deployment 164
4.3.1.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations 164
4.3.1.1.2 GM Cluster Deployment Mode2: NSX-T GM-Active and GM-Standby 167
4.3.1.2 Security configuration for best scale 167
4.3.1.3 Upgrade Federation 168
4.3.1.3.1 From NSX-T 3.2.x (prior to 3.2.2) to NSX 4.1.0 168
4.3.1.3.1 From NSX-T 3.2.2+ to NSX 4.1.x 169
4.3.1.3.2 From NSX 4.0.x to NSX 4.1.0 170
4.3.2 Data Plane 171
4.3.2.1 Edge Node configuration for optimal performance 171
4.3.2.1.1 Recommended design for Edge Node VM 172
4.3.2.1.2 Recommended design for Edge Bare Metal 174
4.3.2.2 Tier-0 options for most services and best performance 178
4.4 Disaster Recovery 180
4.4.1 Management Plane 180
4.4.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations 180
4.4.1.2 GM Cluster Deployment Mode2: NSX-T GM-Active and GM-Standby 182
4.4.2 Data Plane 188
4.4.2.1 Automatic Network Data Plane Recovery 189
4.4.2.1.1 Any Stretched Tier-0 with BGP and without services (No NAT / No GW-FW) + Stretched Distributed Tier-1 189
4.4.2.1.2 Stretched Tier-0 Active/Standby Location Primary/Secondary with Static Routes + Stretched Distributed Tier-1 196
4.4.2.1.3 Stretched Tier-0 Active/Standby Location Primary/Secondary with Static Routes + Stretched Distributed Tier-1 with Local-Egress 200
4.4.2.2 Manual Network Data Plane Recovery 204
4.4.2.2.1 Tier-0 or Tier-1 stretched with services (NAT / GW-FW) 204
4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery 220

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.4.2.3 Compute VMs Recovery 222
4.4.2.4 Load Balancing Data Plane Recovery 228
4.4.2.4.1 LB Disaster Recovery with GSLB 228
4.4.2.4.2 LB Disaster Recovery without GSLB 230
4.5 Requirements and Limitations 234
4.6 Orchestration / Eco-System 237
4.7 Scale and Performance guidance 239
5 Migration to Multi-Locations 240
5.1 From “Single Location” To “NSX-T Multisite” 240
5.2 From “Single Location” To “NSX-T Federation” 242
5.3 From “NSX-T Multisite” To “NSX-T Federation” 245
5.4 Mixing “NSX-T Multisite” and “NSX-T Federation” 251
6 Federation Support within VCF + VVD 252

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Intended Audience

This document is targeted toward virtualization and network architects with NSX-T Data Center good technical background who are interested in deploying VMware NSX-T Data Center in a variety of on-premise Locations.

Locations term here is used in a very broad way. It can be considered **Sites** in different cities, also **Buildings** within a campus, or even **Racks** within a Data Center.

Revision History

Version	Updates	Comments
1.0	NSX-T 4.1.0 updates (Support of L2-Bridge).	Updated chapter 4.2, and 4.5.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

1 Introduction

This document provides guidance and best practices for designing network and security services in multiple locations based on NSX-T Data Center 4.0 capabilities.
Locations being Sites (data centers in different cities), or Buildings (different data centers in a campus), or even Racks (different racks in a single data center).

Figure 1-1: NSX-T Multi-Locations Use Case

Enterprise use multiple locations for three main reasons:

- Mobility of workload
  You might have applications are deployed in different Locations, and they can be moved to different locations for use cases such as Data Center migration, or Disaster Recovery tests.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 1-2: Multi-Locations for mobility

Each application is deployed in a specific location. Then Application App1 has to be moved (compute, storage, network and security) from Location-A to Location-B.

- Growth
  One location (rack, building, site) does not offer enough capacity to host all its applications. Capacity can be of different types, like compute (servers), and/or storage, and/or network (bandwidth).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 1-3: Multi-Locations for growth

Each application can be deployed within one specific location or can be stretched across locations. The deployment mode is an important choice as it has impact on the network and security constructs. The first mode (applications deployed in a single location) does not require network and security constructs to be known in all locations. However, the second mode (applications deployed in multiple locations) does require the network and security constructs to stretched and known in all locations to allow the cross-location communication.

- Disaster Avoidance / Disaster Recovery
  Even in case of one location full failure (rack, building, site), the application services have to remain available. The maximum time requirement to recover the application service varies greatly between customers and their applications. It's important to note the application recovery time is not only based on network recovery, but also compute and storage recovery solutions.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 1-4: Multi-Locations for location DR

There are many ways to recover the application service. It can be with the recovery of the compute, storage, network, and security in another location. It can also be the deployment of the same application in multiple locations and use of GSLB technology to load balance users across the different locations in an active/active or active/standby mode.

This document is organized into several chapters.

- Chapter 2 will present the different NSX-T solutions offered for Multi-Locations use case and the sweet spot of each.
- Chapter 3 and 4 will go deep on each of the NSX-T Multi-Locations solutions. And for each solution, it will start with its architecture, and its supported network and security services. Then it will go over its best practice design and present its disaster recovery solution. Finally, it will highlight the solution requirements and limitations, orchestration and eco-system, as well as its scale and performance guidance.
- Chapter 5 will present how to upgrade from a single Location to a multi locations, as well as how to move from 1 Multi-Location solution to another.
- Chapter 6 will be dedicated to the popular integrated stack VMware Cloud Foundation (VCF) and how VCF supports NSX-T Multi-Locations use case.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

This document does not cover installation, operational monitoring, and troubleshooting. For further details, review the complete NSX-T installation and administration guides.

This document does not explain the architectural building blocks of NSX-T as full stack solution, nor detail functioning of NSX-T components, features and scope either. For further details, review the complete VMware NSX-T Reference Design Guide.

Finally starting with this design guide, readers are encouraged to send a feedback to NSX Design Feedback NSXDesignFeedback@groups.vmware.com.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

2 What are the NSX-T solutions for Multi-Locations

NSX-T offers two distinct solutions for the use case on premise Multi-Locations.
Each has its own pros and sweet spots.

2.1 NSX-T Multisite

The first solution is called **NSX-T Multisite**.
That solution has been introduced with NSX-T 2.3 and has been enhanced since.
At a high-level this solution is one single NSX-T Manager Cluster managing Transport Nodes (hypervisors and Edge nodes) physically in different locations. And that single NSX-T Manager Cluster configures the network and security services centrally for those locations.

Figure 2-1: NSX-T Multisite high-level solution

This NSX-T Multisite solution offers the smallest footprint with only 3 NSX-T Managers VMs, and a large set of networking and security features (see chapter "3.2 Network & Security services supported").
Disaster Recovery in case of Location failure is also offered in a simple way with some requirements on the fabric and vCenter are met with limited downtime (see chapter "3.4 Disaster Recovery").

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

However, the Management plane is only in one single location and this can be a blocker for customer who have policy requirements like General Data Protection Regulation (GDPR).
Also, its scale is limited to the scale of one single NSX-T Manager Cluster capability (NSX-T scale information available on configmax.vmware.com). At last, large MTU is required over the WAN/DCI to allow cross location East/West NSX-T data plane communication.

2.2 NSX-T Federation

The second solution is called **NSX-T Federation**.
That solution has been introduced with NSX-T 3.0, and at least NSX-T 3.1 is recommended for Production.
At a high-level this solution is one central NSX-T Global Manager Cluster (GM) offering central configuration of the network and security services for all locations, and one NSX-T Manager Cluster per location called here Local Manager (LM), managing Transport Nodes for that location (hypervisors and Edge nodes). The GM pushes the network and security configuration to the different LM, which implements it locally.

Figure 2-2: NSX-T Federation high-level solution

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

This NSX-T Federation solution offers local Management plane and so replies to policy requirements like GDPR. Also, it does support fragmentation over the WAN for cross location East/West NSX-T data plane communication and so no large MTU is required over the WAN/DCI. Also, NSX-T Federation offers a simplified Disaster Recovery mode from NSX-T 3.1 with no requirements on the fabric nor vCenter with the Standby Global Manager Cluster (GM-Standby) deployed in any other location. The current Disaster Recovery solution is detailed in the chapter "4.4 Disaster Recovery".

However, currently the networking and security features centrally managed from GM are limited (see chapter "4.2 Network & Security services supported").
Also, it has a larger footprint with 3 Management VMs per location (NSX-T Local Manager Cluster - LM), and another 6 Management VMs for the federation (2 NSX-T Global Manager Cluster – GM).

Note: NSX-T Federation future releases will also offer scale above the scale of one NSX-T Manager Cluster (latest NSX-T Federation scale information available on configmax.vmware.com).

2.3 Summary of Pros/Cons of NSX-T Multi-Locations solutions

	Pros	Cons
NSX-T Multisite	• Large set of Networking & Security features • Smallest solution footprint • Simple DR solution with some strict requirements on Fabric and vCenter	• Scale limited to the scale of 1 NSX-T Manager Cluster • Large MTU requirement on the WAN • Management Plane centralized in 1 location • Complex DR solution if some strict requirements not met
NSX-T Federation	• No large MTU requirement on the WAN (for performance avoid fragmentation) • Management Plane distributed in each location (GDPR) • Simple DR solution with no fabric nor vCenter requirements	• Limited set of Networking & Security features • Large solution footprint • Current scale limited

Each NSX-T Multi-Locations solution has its own architecture, supported services, best practice design, DR solution, requirements and limitations, as well as scale and performance guidance.

The following chapters will detail all those for each solution.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3 NSX-T Multisite

At a high-level this solution is one single NSX-T Manager Cluster managing Transport Nodes (hypervisors and Edge nodes) physically in different locations. And that single NSX-T Manager Cluster configures the network and security services centrally for those locations.

Figure 3-1: NSX-T Multisite Use Case

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.1 Architecture components

This chapter will detail first the Management Plane architecture, then the Data Plane architecture of the NSX-T Multisite solution.

3.1.1 Management Plane

On the Management Plane, the NSX-T Multisite solution is composed of one single NSX-T Manager Cluster. So, three NSX-T Manager VMs.
As explained in the VMware NSX-T Reference Design Guide, the NSX-T Manager VMs can be on the same subnet or different subnets. Also, the maximum latency (RTT) between any of its NSX-T Manager VMs is 10 milliseconds. At last, to operate, the NSX-T Manager Cluster can handle the loss of one of its NSX-T Manager VMs.

There are two modes of NSX-T Manager deployments with each two options.

3.1.1.1 Management Cluster Deployment Mode1: Metropolitan Region

The typical use cases would be different racks / buildings in metropolitan region.

3.1.1.1.1 Management Cluster Deployment Mode1 - Option1: NSX-T Managers VMs deployed in two locations

For the Multi-Locations use case with 2 locations, <u>and</u> latency (RTT) below 10 milliseconds, <u>and</u> no congestion between those locations; it is recommended to have the three NSX-T Manager VMs split in the two locations.

Figure 3-2: NSX-T Multisite Manager Cluster– Use-case 2 locations in metropolitan region (< 10ms latency)

It's important to highlight in this NSX-T Manager Cluster deployment, the loss of one location (Location-A in the figure above) stops the Management Plane service, since the NSX-T Manager Cluster requires at least two valid members.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The recovery of Management Plane is detailed in the chapter “3.4 Disaster Recovery”.

3.1.1.1.2 Management Cluster Deployment Mode1 – Option2: NSX-T Managers VMs deployed in three locations

For the Multi-Locations use case with 3 locations or more, <u>and</u> latency (RTT) below 10 milliseconds between each, <u>and</u> no congestion between those locations; it is recommended to have one NSX-T Manager VM in three different locations.

Figure 3-3: NSX-T Multisite Manager Cluster– Use-case 3 locations+ in metropolitan region (< 10ms latency)

It’s important to highlight in this NSX-T Manager Cluster deployment, the loss of one location does not stop the Management Plane service, since the cluster has still 2 valid members.
The recovery of Management Plane is detailed in the chapter “3.4 Disaster Recovery”.

3.1.1.2 Management Cluster Deployment Mode2: Large Distance Region

The typical use cases would be two Data Centers or more in large distance regions.

For the Multi-Locations use case with latency (RTT) above 10ms across the locations; it is recommended to have all three NSX-T Manager VMs in one single location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-4: NSX-T Multisite Manager Cluster– Use-case two Locations only and/or Data Centers far apart (> 10 ms latency)

It's important to highlight in this NSX-T Manager Cluster deployment, the loss of the location hosting the NSX-T Managers VMs does stop the Management Plane and Control Plane services. The recovery of Management Plane is detailed in the chapter "3.4 Disaster Recovery".

3.1.2 Data Plane

On the Data Plane, the NSX-T Multisite solution is composed of Edge Nodes and hypervisors. Two options are available for the grouping of Edge Nodes in Edge Cluster(s) based on the latency across the different locations and the Tier-0 mode Active/Standby or Active/Active (ECMP). Those will be detailed in the Best Practice Design chapter "3.3.2 Data Plane".

Figure 3-5: NSX-T Multisite Edge Nodes and hypervisors

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.2 Network & Security services supported

NSX-T Data Center offers a very large number of Network & Security services and the NSX-T Multisite solution supports the vast majority of those.

On the Network side, all features are supported: Switching (Overlay and VLAN), IPAM (DHCP and DNS), Routing (VRF, EVPN, NAT and route redistribution), Layer4+ services (Load Balancing, VPN).

On the Security side, most features are supported: Distributed Firewall, Gateway Firewall, FQDN Filtering, URL Filtering, L7 App ID context support, Time-Based Firewall, Identity Firewall, Distributed IDS, Gateway IDS/IPS, Distributed Security only for vCenter VDS Port Group, and TLS inspection.
Network Introspection (Host-Based deployment only) and Endpoint Protection are also supported with some limitations (see chapter 3.2.2 Multisite Security Services).
The features not supported are Malware Prevention, Network Detection and Response.

Figure 3-6: NSX-T Multisite Network and Security services

3.2.1 Multisite Network Services

For the Network services across locations, **all switches (Segments) are stretched**.
The central Network services (NAT, load balancing, VPN) are individually active only in one location. Then in case of that location failure, those services recover in another location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-7: NSX-T Multisite Network services across locations

In the figure above, you can see all Segments "Blue" and "Green" are available to all locations. The East/West routing is offered in a distributed fashion locally through the Tier-0 and Tier-1 Distributed Routers. The North/South routing is offered by Tier-0 that are active in one location, "Blue" in Location-A and "Green" in Location-C.

Then in case of that location failure, the North/South routing failover to the other location. If those locations have a latency below 10 milliseconds, the failover can be automatic. If those locations

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

have a latency above 10 milliseconds, then then failover has to be manual or scripted. This will be detailed in the chapter "3.4 Disaster Recovery".

3.2.2 Multisite Security Services

For the Security services across locations, **all NSX Groups and NSX DFW Sections/Rules are stretched**. In other words, Groups can use any static or dynamic membership and have members from any location; and DFW Sections/Rules are pushed to all locations.

Network Introspection Host-Based (previously named Service Insertion) is supported.
There are a couple of points to keep in mind:

- On NSX side
  - Host-Based deployment only (no Network Introspection Cluster-based support)
  - In case of Hosted Partner SVM failure on an ESXi, that ESXi redirects traffic to any ESXi Partner SVM. That redirected traffic may go to a remote ESXi.

- on Partner side (such as Palo Alto, CheckPoint, Fortinet, Netscout, etc):
  - Partner must validate NSX-T Multisite support on their side, especially
    - Partner Console communication to remote Partner SVM
  - The Partner Console must offer a DR solution with IP preservation (with SRM or vSphere-HA or other)

Endpoint Protection (previously named Guest Introspection) is supported.
There are a couple of points to keep in mind:

- On NSX side
  - In case of Hosted Partner SVM failure on an ESXi, that ESXi does not redirect traffic to another ESXi Partner SVM. The traffic goes through without protection.

- on Partner side (such as Bitdefender, Trend Micro, etc):
  - Partner must validate NSX-T Multisite support on their side, especially
    - Partner Console communication to remote Partner SVM
    - Partner Console DR with IP preservation
  - The Partner Console must offer a DR solution with IP preservation (with SRM or vSphere-HA or other)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-8: NSX-T Multisite Network Introspection and Endpoint Protection across locations

In the figure above, on the Management Plane, the Partner Console receives the NSX Groups and pushes its configuration to the different Host-Based Partner SVM.

On the Data Plane, each ESXi redirects the VM traffic to its hosted Partner SVM.

Figure 3-9: NSX-T Multisite Network Introspection and Endpoint Protection redirection

For Network Introspection, in case of hosted Partner SVM failure in one of the ESXi, that ESXi will redirect its VM traffic new flows to another ESXi Partner SVM (existing flows will be dropped). That other ESXi could be remote, and in such case the redirected traffic will go cross-location, as shown in the figure below.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-10: NSX-T Multisite Network Introspection redirection – After Partner SVM failure

For Guest Introspection, in case of hosted Partner SVM failure in one of the ESXi, new files won't be inspected for the VMs on that ESXi until the recovery of its Partner SVM.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.3 Best Practice Design

The NSX-T Multisite solution offers different options for its Management Plane architecture and its Data Plane architecture.
This section will detail the best practice design for each architecture based on the different cases such as the latency between locations, numbers of location, and Tier-0 deployment mode.
It's important to note the choice of Management Plane and Data Plane architectures are independent. All Management Plane modes can work with all Data Plane modes.

3.3.1 Management Plane

As presented in the chapter "3.1.1 Management Plane", the Management Plane is offered by one single NSX-T Manager Cluster, so three NSX-T Managers VMs. And there are 2 possible deployment modes based on the number of locations and the latency (RTT) across those different locations**.**

3.3.1.1 Management Cluster Deployment Mode1: Metropolitan Region

The typical use cases would be different racks / buildings in metropolitan region.

3.3.1.1.1 Management Cluster Deployment Mode1 - Option1: NSX-T Managers VMs deployed in two locations

For the Multi-Locations use case with 2 locations, <u>and</u> latency (RTT) below 10 milliseconds, <u>and</u> no congestion between those locations; it is recommended to have the three NSX-T Manager VMs split in the two locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-11: NSX-T Multisite Manager Cluster– Management Deployment Mode1 Option1

NSX-T Manager VMs can be hosted on one single vCenter or dedicated local vCenter in each location (as represented in the figure above). There is no requirement for stretch vCenter-Cluster, nor vCenters in linked mode in this Management Cluster deployment mode1.

There is also no requirement for L2-VLAN Management stretch across the different locations. Each NSX-T Manager can be on a different L2-VLAN Management subnet. However, if an NSX-T Manager Cluster VIP is needed, then L2-VLAN Management across the locations is required. If no L2-VLAN Management across locations is available, an external load balancer VIP has to be configured.

In this Management Deployment Mode1 – Option1, the loss of one location (Location-A in the figure above) stops the Management Plane service, since the NSX-T Manager Cluster requires at least two valid members.
The recovery of Management Plane is detailed in the chapter "3.4 Disaster Recovery".

3.3.1.1.2 Management Cluster Deployment Mode1 – Option2: NSX-T Managers VMs deployed in three locations

For the Multi-Locations use case with 3 locations or more, and latency (RTT) below 10 milliseconds between each, and no congestion between those locations; it is recommended to have one NSX-T Manager VM in three locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-12: NSX-T Multisite Manager Cluster– Management Deployment Mode1 Option2

NSX-T Manager VMs can be hosted on one single vCenter or dedicated local vCenter in each location (as represented in the figure above). There is no requirement for stretch vCenter-Cluster, nor vCenters in linked mode in this Management Cluster deployment mode1.

There is also no requirement for L2-VLAN Management stretch across the different locations. Each NSX-T Manager can be on a different L2-VLAN Management subnet. However, if an NSX-T Manager Cluster VIP is needed, then L2-VLAN Management across the locations is required. If no L2-VLAN Management across locations is available, an external load balancer VIP has to be configured.

In this Management Deployment Mode1 – Option2, the loss of one location does not stop the Management Plane service since the NSX-T Manager cluster still has 2 valid members.

3.3.1.2 Management Cluster Deployment Mode2: Large Distance Region

The typical use cases would be two Data Centers or more in large distance regions.

For the Multi-Locations use case with latency (RTT) above 10ms across the locations; it is recommended to have all three NSX-T Manager VMs in one single location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

There are 2 deployment options for this use case.

3.3.1.2.1 Management Cluster Deployment Mode2 – Option1: L2-VLAN Management stretch

The option1 requires one L2-VLAN Management stretch across the different locations.

Figure 3-13: NSX-T Multisite Manager Cluster– Management Deployment Mode2 Option1

With this option1, the three NSX-T Manager VMs are hosted on the same location and connected on one Management VLAN(s) stretched across the two locations. Since all three NSX-T Manager VMs are on the same subnet / VLAN an NSX-T Manager Cluster VIP can be used here; or an external load balancer VIP can also be configured.

In this Management Deployment Mode2 – Option1, the loss of the location hosting the three NSX-T Managers stops the Management Plane service (Location-A in the figure above).
The NSX-T Manager service VMs can be recovered with 2 methods:
vCenter vSphere-HA offers an automatic Management service recovery. This method requires a vCenter stretched cluster and synchronized datastore across the 2 locations. The NSX-T Management Plane service outage will be around 15 minutes.
VMware Site Recovery Manager (SRM) with a manual or scripted NSX-T Management service recovery. Attention, this method may lose the most recent NSX-T Manager configuration; the configuration done between the last SRM synchronization (recovery point) and location failure. The Management Plane service outage will be around 20 minutes after the start of the SRM recovery.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Disaster recovery for this Management Deployment Mode2 – Option1 will be detailed in the chapter "3.4 Disaster Recovery".

3.3.1.2.2 Management Cluster Deployment Mode2 – Option2: No L2-VLAN Management stretch

The option2 requires No L2-VLAN Management stretch across the different locations:

Figure 3-14: NSX-T Multisite Manager Cluster– Management Deployment Mode2 Option2

With this option2, the three NSX-T Manager VMs are hosted on the same location and connected to a Management VLAN(s) not stretched across the two locations. All three NSX-T Manager VMs don't need to be on the same subnet / VLAN. If they are, an NSX-T Manager Cluster VIP can be used here; otherwise an external load balancer VIP has to be configured.

At last in this Management Deployment Mode2 – Option2, the loss of the location hosting the three NSX-T Managers stops the Management Plane service (Location-A in the figure above). The Management Plane service is recovered via the usage of FQDN for the three NSX-T Manager VMs, and the NSX-T backup/restore. The Management Plane service outage will be around 1 hour.
Disaster recovery for this Management Deployment Mode2 – Option2 will be detailed in the chapter "3.4 Disaster Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.3.2 Data Plane

The Data Plane service is offered by hypervisors and Edge Nodes, and each location must host hypervisors and Edge Nodes.
The NSX-T configuration of the hypervisors is always the same for all the locations. They are all NSX prepared from the NSX-T Manager Cluster.
However, there are two modes for the NSX-T configuration Edge Nodes based on the latency across the different locations and the Tier-0 mode Active/Standby or Active/Active (ECMP).

3.3.2.1 Edge Cluster Deployment Mode1: Stretched Edge Clusters with Edge Nodes deployed in different failure domains

For this use case with latency below 10 milliseconds across locations, and Tier-0 in Active/Standby mode; it is recommended to create Edge Clusters with Edge Nodes in two locations with the use of Edge Failure Domains.
The typical use cases would be customers with different racks / buildings in metropolitan region and North/South throughput need below the performance of one Edge Node.

VMware, Inc. – www.vmware.com
Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-15: NSX-T Multisite Edges - Edge Node Deployment Mode1

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

There is an Edge Cluster per location pair Active/Standby. In the figure above, there is an Edge Cluster "Blue" for the pair Location-A/Location-B; and an Edge Cluster "Green" for the pair Location-B/Location-C.
Each Edge Cluster has at least one Edge Node in each of its location. In the figure above, there are two Edge Nodes in each location per Edge Cluster. Edge Nodes can be from any form factor, VM or Bare Metal. In case of Edge Node VM, they can be installed in different vCenters or not. There is also no requirement for vSphere-HA.
Then Edge Nodes are added to a Failure Domain with a preference option. The failure domain preferred option allows the automatic deployment of Tier-1 active in the preferred location. For Tier-0, their active location is part of the Tier-0 configuration. In the figure above, the Edge Cluster "Blue" has Edge Nodes EN1A + EN2A (in Location-A) in a failure domain preferred, and Edge Nodes EN1B + EN2B (in Location-B) in a failure domain not preferred. Also, the Edge Cluster "Green" has Edge Nodes EN1C + EN2C (in Location-C) in a failure domain preferred, and Edge Nodes EN3B + EN4B (in Location-B) in a failure domain not preferred. Then Tier-0 / Tier-1 "Blue" / "Green" deployed in the Edge Cluster "Blue" / "Green" makes the North/South "Blue" / "Green" Data Plane in in Location-A / Location-C, with a recovery automatically in Location-B.
In addition, to allow an automatic route redistribution of the internal "Blue" / "Green" networks, NAT, and load balancer VIP to the external world, BGP is configured between the ToR and the Tier-0 active and standby. In the figure above, the "Blue" networks are advertised via Location-A and Location-B, but with a worst cost from Location-B (AS Prepend). The "Green" networks are advertised via Location-C and Location-B, but with a worst cost from Location-B (AS Prepend).

It's worth keeping in mind in this Edge Node Deployment Mode1, the loss of one Edge Node hosting an active Tier-0 or active Tier-1 would generate an automatic move of that Tier-0 / Tier-1 service to the other location hosting the standby Tier-0 or standby Tier-1 and so the North/South data plane would cross locations. The Data Plane service outage will be around 1 or 3 seconds based on the Edge Node form factor (Edge Node Bare Metal or Edge Node VM). The same will also happen during an NSX upgrade when the Edges are upgraded, the data plane traffic will be slightly interrupted (1 or 3 seconds), and the North/South data plane would cross locations.
Now in case of full location failure would stop all central network services (for instance the loss of Location-A in the figure above would stop the "Blue" North/South central network services). However, thanks to the Tier-0 / Tier-1 standby hosted in another location, those Tier-0 / Tier-1 turns again automatically active in that other location with an outage of 1 or 3 seconds based on the Edge Node form factor.
Disaster recovery for this Edge Node Deployment Mode1 will be detailed in the chapter "3.4 Disaster Recovery".

3.3.2.2 Edge Cluster Deployment Mode2: Non-Stretched Edge Clusters with Edge Nodes deployed in no failure domain

For this use case with latency above 10 milliseconds across locations, <u>or</u> Tier-0 in Active/Active (ECMP) mode; it is recommended to create Edge Clusters with Edge Nodes in one single location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The typical use cases would be Data Centers in large distance region or North/South throughput need above the performance of one Edge Node.

Figure 3-16: NSX-T Multisite Edges - Edge Node Deployment Mode2

There is an Edge Cluster per location. Each Edge Cluster has at least two Edge Nodes.
Edge Nodes can be from any form factor, VM or Bare Metal. In case of Edge Node VM, they can be installed in different vCenters or not. There is also no requirement for vSphere-HA. In the figure above, there is an Edge Cluster LocA, Edge Cluster LocB, and Edge Cluster LocC.
Each Edge Cluster has a dedicated Tier-0. In the figure above, there is a "T0 GW LocA" in Location-A, "T0 GW LocB" in Location-B, and "T0 GW LocC" in Location-C.
Then Tier-1 are deployed in the specific location where you want your North/South and connected to that location Tier-0 GW. In the figure above, the "T1 GW Blue" is deployed in Edge Cluster LocA + connected to "T0 GW LocA", and the "T1 GW Green" is deployed in Edge Cluster LocC + connected to "T0 GW LocC".
In addition, to allow an automatic route redistribution of the internal "Blue" / "Green" networks, NAT, and load balancer VIP to the external world, BGP is configured between the ToR and the

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Tier-0 active and standby. In the figure above, the "Blue" networks is advertised via Location-A, and the "Green" networks is advertised via Location-C.

At last in this Edge Node Deployment Mode2, the loss of a location hosting a Tier-0 / Tier-1 stops the networking service for those routers (the loss of Location-A in the figure above would stop the "Blue" North/South network services). The failover of this "Blue" North/South" network services to another location, like Location-B is done manually or via script. This will be in the chapter "3.4 Disaster Recovery".
Disaster recovery for this Edge Node Deployment Mode2 will be detailed in the chapter "3.4 Disaster Recovery".

3.3.2.3 (Special Use Case) Edge Cluster Deployment Mode3: Stretched Edge Cluster Cross Locations

First and foremost, this deployment mode replies to a specific use case where increase cross-location traffic is acceptable and asymmetric routing is not a concern.
The typical use cases would be customers with different racks / buildings in metropolitan region and no firewall cross-locations, with automatic failover and high North/South bandwidth.

This mode has one Edge Cluster stretched across all locations.

Heavy Traffic Cross-Location:
This deployment mode generates a large amount of cross-location traffic.
The main reason is Transport Nodes don't have location information and forward South/North traffic evenly (based on ECMP hash) between the different Edge Nodes.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-17: NSX-T Multisite Edges - Edge Node Deployment Mode3 with Tier1 DR_Only

That's the case with Tier-1 DR_Only gateways (see figure above), where each hypervisor sends traffic evenly to all Edge Nodes in all locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-18: NSX-T Multisite Edges - Edge Node Deployment Mode3 with Tier1 with Services

That's also the case with Tier-1 SR gateways, where each hypervisor sends traffic to the Edge Node hosting the Tier-1 SR Active.
Then assuming that Edge Node hosting the Tier-1 SR Active is also hosting the Tier-0 Active/Active, it forwards all South/North traffic to itself and then its ToR (see figure above).
Otherwise if that Edge Node Tier-1 SR Active is not also hosting the Tier-0 Active/Active, as with a dedicated stretched Edge cluster for Tier-1, then it would forward all South/North traffic to all Edge Nodes in the different locations (not shown in the picture above).

Asymmetric Routing:

This deployment mode generates asymmetric routing, so be sure North/South communication does not cross a physical firewall (like a physical firewall above the ToR).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-19: NSX-T Multisite Edges - Edge Node Deployment Mode3 with Tier1 with Services

That's the case with Tier-1 DR_Only gateways (see figure above), where Tier-0 sends traffic to its local ToR, but external response may come via another location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-20: NSX-T Multisite Edges - Edge Node Deployment Mode3 with Tier1 with Services

That's also the case with Tier-1 SR gateways (see figure above), where Tier-0 sends traffic to its local ToR, but external response may come via another location.

It's important to note the Edge Node selection of the Tier-1 with Services has to be done manually to assure the Active and Standby are in different locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.4 Disaster Recovery

The different NSX-T Multisite deployments modes of Management Plane and Data Plane have been detailed in the chapter above.
This chapter reviews in detail each deployment mode and its Disaster Recovery steps.
At last, the final section of this chapter will detail Disaster Recovery with the pre-deployment of the same application in multiple locations and the use of GSLB technology.

3.4.1 Management Plane

3.4.1.1 Management Cluster Deployment Mode1: Metropolitan Region

The typical use cases would be different racks / buildings in metropolitan region.

3.4.1.1.1 Management Cluster Deployment Mode1 - Option1: Two locations or more with recovery via vSphere HA or SRM

This mode has all NSX-T Manager VMs in a single location with L2-VLAN Management stretch across the 2 locations.

The first recovery method is based on vSphere-HA.
The typical use cases would be two locations only at proximity (rack / building).
It requires a stretched vCenter Cluster and synchronized datastore across the 2 locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-21: Management Plane Deployment Mode1 Option1 with vSphere-HA – Before location failure

Before any location failure, the NSX-T Manager Cluster is up and running with 3 members in single location (Location-A in the figure above).
Also, vCenter has a Management stretched cluster with ESXi in both locations.
At last, vSphere-HA is configured on that vCenter stretched cluster to have those NSX-T Manager VMs on ESXi primary location (Location-A in the figure above) and restarted on ESXi secondary locations only if no ESXi resources are up in primary location.
Note: Such vSphere-HA configuration is done on vCenter Cluster under:

- *"Configure – VM/Host Groups" with*
  Two groups "ESXi Location-A" and "ESXi Location-B", each containing the ESXi member of those locations.
  One group "NSX-T Manager VMs" containing the 3 NSX-T Manager VMs.
- *"Configure – VM/Host Rules" with*
  One rule "NSX-T Manager VMs primary in Location-A" of type "Virtual Machines to Hosts" and configured with "NSX-T Manager VMs" "Should run on hosts in group" "ESXi Location-A"

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-22: Management Plane Deployment Mode1 Option1 with vSphere-HA – After location failure

After the primary location failure (Location-A in the figure above), there is no more running NSX-T Manager VMs running and the NSX-T Management Plane is down.
vSphere-HA notices the loss of the different ESXi hosting those NSX-T Manager VMs, and restart those on ESXi host located in location B (see figure above).
Once the NSX-T Managers VMs restarted their services and rejoined in their NSX-T Manager Cluster, then the NSX-T Management Plane is back operational.
In case of NSX-T Manager Cluster VIP configured, this one is still working. In case of external load balancer VIP configured for NSX-T Managers, that service must have a Disaster Recovery solution in pace (like GSLB or vSphere-HA).

The Management Plane service outage will be around 15 minutes.
During the outage, new workload deployed doesn't have network connectivity, vMotion of existing VMs is prevented by vCenter, vCenter stops the redistribution of VMs among hypervisors with DRS.

The second recovery method is based on VMware Site Recovery Manager (SRM).
The typical use cases would be two locations in a metropolitan area.
Attention, this method may lose the most recent NSX-T Manager configuration; the configuration done between the last SRM synchronization (recovery point) and location failure.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-23: Management Plane Deployment Mode1 Option1 with SRM – Before location failure

Before any location failure, the NSX-T Manager Cluster is up and running with 3 members in single location (Location-A in the figure above).
SRM, or more accurately vSphere Replication, replicates it to the secondary location at specific intervals (default 1 hour and minimum is 5 minutes).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-24: Management Plane Deployment Mode1 Option1 with SRM – After location failure

After the primary location failure (Location-A in the figure above), there is no more running NSX-T Manager VMs running and the NSX-T Management Plane is down.
SRM recovery has to be started to recover the NSX-T Manager in the secondary location (Location-B in the figure above). The NSX-T Manager VMs are automatically plugged on the same VLAN, with their same IP addresses, and with their configuration from the last replication.
Once the NSX-T Managers VMs restarted their services and rejoined in their NSX-T Manager Cluster, then the NSX-T Management Plane is back operational.
In case of NSX-T Manager Cluster VIP configured, this one is still working. In case of external load balancer VIP configured for NSX-T Managers, that service must have a Disaster Recovery solution in pace (like GSLB or vSphere-HA).

The Management Plane outage will be around 20 minutes after the start of the SRM recovery.
During the outage, new workload deployed doesn't have network connectivity, vMotion of existing VMs is prevented by vCenter, vCenter stops the redistribution of VMs among hypervisors with DRS.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.4.1.1.2 Management Cluster Deployment Mode1 – Option2: Two locations with recovery via Manager cluster deactivation

This mode has the three NSX-T Manager VMs split in the two locations.

Figure 3-25: Management Plane Deployment Mode1 Option2 – Before location failure

Before any location failure, the NSX-T Manager Cluster is up and running with the three NSX-T Manager VMs split in the two locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-26: Management Plane Deployment Mode1 Option2 – After location failure

After the failure of the location hosting two NSX-T Manager VMs, there are only 1 running NSX-T Manager member (VM3) in the Cluster and so the Management Plane service is no more working.
The remaining NSX-T Manager VM3 is no more reachable via UI nor API. Its only access it via Console or SSH. From that CLI access it is possible to recover the Management Plane service, making that NSX-T Manager VM a standalone manager. The CLI command is "deactivate cluster". When the NSX-T Manager VM3 is changed to standalone, it updates all Transport Nodes (hypervisors and Edge Nodes) to use only itself as Manager.
Then it is recommended to rebuild high-availability on the NSX-T Manager Cluster adding two new NSX-T Manager VMs in the cluster.

In case of NSX-T Manager Cluster VIP configured, this one is still working after the recovery of the Management Plane. In case of external load balancer VIP configured for NSX-T Managers, that service must have a Disaster Recovery solution in place (like GSLB).

The Management Plane outage will be around 10 minutes after the deactivation of the NSX-T Manager Cluster.
During the outage, new workload deployed doesn't have network connectivity, vMotion of existing VMs is prevented by vCenter, vCenter stops the redistribution of VMs among hypervisors with DRS.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of recovery of the Location-A; the two NSX-T Managers VM1 + VM2 regain connectivity. However, Transport Nodes configuration has been updated to use only the NSX-T Manager VM3 and so it has no impact on their Management Plane and Control Plane. Those NSX-T Manager VMs VM1 + VM2 can't rejoin the NSX-T Manager Cluster, so delete them and rebuild a new NSX-T Cluster with new VMs if not already done.

3.4.1.1.3 Management Cluster Deployment Mode1 – Option3: Three locations with recovery via Managers distributed across all

This mode has one NSX-T Manager VM in three different locations.

Figure 3-27: Management Plane Deployment Mode1 Option3 – Before location failure

Before the any location failure, the NSX-T Manager Cluster is up and running with 1 member in different locations.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-28: Management Plane Deployment Mode1 Option3 – After location failure

After one location failure, there are still 2 running NSX-T Managers members in the Cluster and so the Management Plane service is still working.
In case of NSX-T Manager Cluster VIP configured and L2-VLAN Management stretch, this one is still working. In case of external load balancer VIP configured for NSX-T Managers, that service must have a Disaster Recovery solution in place (like GSLB).

Since the Management Plane service is not impacted: new workload can be deployed, existing VMs can be vMotioned, vCenter can redistribute automatically VMs among hypervisors with DRS.

In case the lost location can't be recovered for a long time, it is recommended to rebuild high-availability on the NSX-T Manager Cluster.
For that, you have two options. First option is to detach from the NSX-T Manager Cluster the NSX-T Manager member lost and re-add a new NSX-T Manager member. Second option in you have a single vCenter Management with stretched vCenter-Cluster across locations, is to recover the lost NSX-T Manager member via vSphere-HA.

3.4.1.2 Management Cluster Deployment Mode2: Large Distance Region

The typical use cases would be two Data Centers or more in large distance regions.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

For the Multi-Locations use case with latency (RTT) above 10ms across the locations; it is recommended to have all three NSX-T Manager VMs in one single location.

There are 2 deployment options for this use case.

3.4.1.2.1 Management Cluster Deployment Mode2 - Option1: Two locations or more with recovery via SRM

This mode has all NSX-T Manager VMs in a single location with L2-VLAN Management stretch across the 2 locations.

Attention, this method may lose the most recent NSX-T Manager configuration; the configuration done between the last SRM synchronization (recovery point) and location failure.

Figure 3-29: Management Plane Deployment Mode2 Option1 with SRM – Before location failure

Before any location failure, the NSX-T Manager Cluster is up and running with 3 members in single location (Location-A in the figure above).
SRM, or more accurately vSphere Replication, replicates it to the secondary location at specific intervals (default 1 hour and minimum is 5 minutes).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-30: Management Plane Deployment Mode2 Option1 with SRM – After location failure

After the primary location failure (Location-A in the figure above), there is no more running NSX-T Manager VMs running and the NSX-T Management Plane is down.
SRM recovery has to be started to recover the NSX-T Manager in the secondary location (Location-B in the figure above). The NSX-T Manager VMs are automatically plugged on the same VLAN, with their same IP addresses, and with their configuration from the last replication.
Once the NSX-T Managers VMs restarted their services and rejoined in their NSX-T Manager Cluster, then the NSX-T Management Plane is back operational.
In case of NSX-T Manager Cluster VIP configured, this one is still working. In case of external load balancer VIP configured for NSX-T Managers, that service must have a Disaster Recovery solution in pace (like GSLB or vSphere-HA).

The Management Plane outage will be around 20 minutes after the start of the SRM recovery.
During the outage, new workload deployed doesn't have network connectivity, vMotion of existing VMs is prevented by vCenter, vCenter stops the redistribution of VMs among hypervisors with DRS.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.4.1.2.2 Management Cluster Deployment Mode2 – Option2: Two locations or more with recovery via FQDN + backup/restore

This mode has all NSX-T Manager VMs in a single location, and without L2-VLAN Management stretch across the different locations nor stretched vCenter Cluster.

The typical use cases would be Data Centers in large distance region.

Figure 3-31: Management Plane Deployment Mode2 Option2 – Before location failure

Before the any location failure, the NSX-T Manager Cluster is up and running with 3 members in single location (Location-A in the figure above).

DNS names are created for the three NSX-T Managers (step1 in the figure above) with a short TTL. We suggest 5 minutes TTL.
NSX-T Managers are configured with FQDN option. This configuration option automatically asks all the Transport Nodes (Edge Nodes and hypervisors) to communicate with the NSX-T Managers via their DNS name (step 2 in the figure above).
At last, a continuous backup is configured on one of the three NSX-T Managers with the location of the backup in the secondary location (step 3 in the figure above).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-32: Management Plane Deployment Mode2 Option2 – After location failure

After the primary location failure (Location-A in the figure above), there is no more running NSX-T Manager VMs running and the NSX-T Management Plane is down.
The recovery of the Management Plane is done in three steps:

1. Update of the DNS entries of the NSX-T Manager with their new IP in secondary location
2. Restore of the NSX-T Manager backup
   All the Transport Nodes (Edge Nodes and hypervisors) continuously try to reconnect to the new NSX-T Managers IP received via DNS. Once the NSX-T Manager cluster is up and running, they reconnect to it (step 2' in the figure above).
3. In case of NSX-T Manager Cluster VIP, modify it with its new IP from the Management Location-B subnet. In case of NSX-T Manager external load balancer VIP, modify its pool members with the new NSX-T Manager IP.

The step 1, 2, and 3 can be manual or scripted.

The Management Plane service outage will be around 1 hour once the NSX-T Manager restore started.
During the Management Plane service outage, new workload deployed doesn't have network connectivity, vMotion of existing VMs is prevented by vCenter, vCenter stops the redistribution of VMs among hypervisors with DRS.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.4.2 Data Plane

3.4.2.1 Edge Cluster Deployment Mode1: Stretched Edge Clusters with Edge Nodes deployed in different failure domains

This mode has Edge Clusters with Edge Nodes in two locations with the use of Edge Failure Domains.
The typical use cases would be customers with different racks / buildings in metropolitan region and North/South throughput need below the performance of one Edge Node.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-33: Data Plane Deployment Mode1 – Before location failure

Before any location failure, the Tier-0 / Tier-1 "Blue" / "Green" are created in the Edge Cluster "Blue" / "Green". With the usage of Edge Failure Domains, automatically the Tier-1 "Blue" /

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

"Green" active are in Location-A / Location-C, and standby in Location-B. For Tier-0, the active location is part of the Tier-0 configuration.
So, the North/South "Blue" / "Green" Data Plane is via Location-A / Location-C.

Figure 3-34: Data Plane Deployment Mode1 – After location failure

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

After the loss of a Location-A hosting the Tier-0 / Tier-1 “Blue” active, the North/South “Blue” Data Plane is stopped. The East/West “Blue” Data Plane in other locations is not impacted, as long as it doesn’t need to cross the Tier-1 Service Router (SR) component.
No impact for the North/South and East/West “Green” Data Plane.
Then the Tier-0 / Tier-1 “Blue” standby hosted in Location-B turn automatically active in Location-B and the North/South “Blue” Data Plane is recovered.

The Data Plane service outage will be around 1 or 3 seconds based on the Edge Node form factor (Edge Node Bare Metal or Edge Node VM).
There is no need to wait for the BGP update. The “Blue” networks, NAT, and load balancer VIP were already advertised via Location-B at a worst cost (AS Prepend). Now that location is the only one advertising some subnets and the external world goes automatically via this location to reach those.

Special Case: Split-Brain Scenario

This special failure case covers the loss of cross-location communication between Location-A and Location-B, but its Internet communication is still working.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-35: Data Plane Deployment Mode1 – After cross-location failure

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In the case of Location-A cross-location only failure, communication between Edge Nodes Loc-A and Edge-Nodes Loc-B is disconnected. So Tier-0 / Tier-1 “Blue” are active in both locations Loc-A and Loc-B.
Tier-0 “Blue” Loc-A still active, still advertise the “Blue” segments.
Tier-0 “Blue” Loc-B now active, still advertise the “Blue” segments at the same worst cost (AS Preprend).
So the “Blue” North/South is still via Loc-A and there is no impact for the North/South Data Plane for the communication to “Blue” Compute (VMs) hosted in Loc-A.
However, the “Blue” North/South to “Blue” Compute (VMs) hosted in Loc-B will be interrupted, as well as “Blue” cross-location East/West communication.

3.4.2.2 Edge Cluster Deployment Mode2: Non-Stretched Edge Clusters with Edge Nodes deployed in no failure domain

This mode has one Edge Cluster per location.
The typical use cases would be Data Centers in large distance region or North/South throughput need above the performance of one Edge Node.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-36: Data Plane Deployment Mode2 – Before location failure

Before any location failure, the Tier-1 "Blue" / "Green" are created in the Edge Cluster LocA / LocC and are connected to the local Tier-0 GW LocA / LocC.
So, the North/South "Blue" / "Green" Data Plane is via Location-A / Location-C.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-37: Data Plane Deployment Mode2 – After location failure

After the loss of a Location-A hosting the Tier-1 "Blue" active, the North/South "Blue" Data Plane is stopped. The East/West "Blue" Data Plane in other locations is not impacted, as long as it doesn't need to cross the Tier-1 Service Router (SR) component.
No impact for the North/South and East/West "Green" Data Plane.
To recover the North/South "Blue" Data Plane, the Tier-1 "Blue" must be attached to Edge Cluster LocB (instead of Edge Cluster LocA). In addition, the Tier-1 "Blue" has to be connected to Tier-0 GW LocB (instead of Tier-0 GW LocA). Those two configurations can be done directly from the NSX-T Manager UI, or can be scripted via API and part of the recovery plan.

The Data Plane service outage will last up to the Tier-1 configuration change. Once the configuration change is configured on the NSX-T Manager, the Data Plane service is immediately recovered.
The Tier-1 configuration change can be done in manually or can be scripted. Example of scripts are available on https://github.com/dcoghlan/NSX-T-MultiSite-Example-Scripts.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Special Case: Split-Brain Scenario

This special failure case covers the loss of cross-location communication between Location-A and Location-B, but its Internet communication is still working.

Figure 3-38: Data Plane Deployment Mode2 – After cross-location failure

In the case of Location-A cross-location only failure, communication between hypervisors Loc-A and Loc-B are interrupted.

So the "Blue" North/South is still via Loc-A and there is no impact for the North/South Data Plane for the communication to "Blue" Compute (VMs) hosted in Loc-A.

However the "Blue" North/South to "Blue" Compute (VMs) hosted in Loc-B will be interrupted, as well as "Blue" cross-location East/West communication.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.4.2.3 (Special Use Case) Edge Cluster Deployment Mode3: Stretched Edge Cluster Cross Locations

First and foremost, this deployment mode replies to a specific use case where increase cross-location traffic is acceptable and asymmetric routing is not a concern.
The typical use cases would be customers with different racks / buildings in metropolitan region and no firewall cross locations.

This mode has one Edge Cluster stretched across all locations.

Figure 3-39: Data Plane Deployment Mode3 with Tier1 DR_Only – Before location failure

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-40: Data Plane Deployment Mode3 with Tier1 with Services – Before location failure

Before any location failure, the Tier-0 are Active/Active across locations and traffic is distributed across them with the Tier-1 DR_Only, and locally with the Tier-1 with Services.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-41: Data Plane Deployment Mode3 with Tier1 DR_Only – After location failure

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-42: Data Plane Deployment Mode3 with Tier1 with Services – After location failure

After the loss of Location-A, Tier-0 and Tier-1 DR_Only services is still offered by the remaining two locations. Tier-1 with Service automatically failovers to another location (Location-B).

No impact for the North/South and East/West Data Plane.

The Data Plane service outage will be around 1 or 3 seconds based on the Edge Node form factor (Edge Node Bare Metal or Edge Node VM).

<u>Special Case: Split-Brain Scenario</u>

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

This special failure case covers the loss of cross-location communication between Location-A and Location-B, but its Internet communication is still working.

Figure 3-43: Data Plane Deployment Mode3 with Tier1 DR_Only – After cross-location failure

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-44: Data Plane Deployment Mode3 with Tier1 with Services – After cross-location failure

In the case of Location-A cross-location only failure, the "Blue" Segments are still advertised by Tier-0 "Blue" from all 3 locations.
However, communication between Transport Nodes (Edge Nodes and hypervisors) Loc-A and Loc-B + Loc-C are interrupted.

And because of asymmetric data plane of that deployment mode, traffic from a VM Loc-A:

- its South/North exit via Loc-A
- its North/South may enter via Loc-C (or Loc-B or Loc-A)

So split-brain scenario breaks the data plane of this deployment mode "Edge Cluster Deployment Mode3: Stretched Edge Cluster Cross Locations".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.4.2.4 Compute VMs Recovery

In case of location failure, all compute VMs hosted in that location are also lost.
VMware offers a solution to replicate compute to another location and recover them in case of location failure: VMware Site Recovery Manager (SRM).

In the figure below, I have a Tier-0 / Tier-1 / Segment stretched topology with SRM used to replicate VMs from Location-A to Location-B. Any other topology with a stretched Segment could be used.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-45: Compute VMs before location failure

SRM replicates the VMs from one location to another. The replicated VMs are powered off, unplugged, with different IDs (BIOS ID, Instance ID, External ID, and Attachment ID), and with no TAG.

For your information, the BIOS ID is viewable from LM under "Inventory / Virtual Machines", and the Attachment ID is viewable from GM or LM under "Networking / Segment / Ports".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of location failure, the NSX Management Service must be recovered first, as discussed in the chapter "3.4.1 Management Plane".
Then for the Compute VMs recovery, as represented in the figure below, SRM powers on the replicated VMs and plug them to the appropriate segments. SRM also updates the IDs (BIOS ID, Instance ID, External ID, and Attachment ID) of the replicated started VMs to match the original ones. And NSX-T Manager has those VMs IDs in its Inventory database with Tags.

Figure 3-46: Compute VMs after location failure

Important Note for "SRM Planned Migration":
"SRM Planned Migration" has the following workflow:

1. SRM asks the source Location-A ESXi to power off and unplug the VMs
   So the source Location-A ESXi update the NSX-T Manager about those VMs unplugged.
   NSX-T Manager keeps those VMs and their NSX VM Tags information for 30 minutes.
2. SRM asks the destination Location-B ESXi to power on the replicated VMs and plug the, to Segments

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

So the destination Location-B ESXi update their NSX-T Manager about those new VMs. NSX-T Manager adds those VMs in its inventory.
If those VMs are powered on in less than 30 minutes after they were unplugged from the source Location-A ESXi, NSX-T Manager will have those with the NSX VM Tags information since they have the same VM Instance-ID as the VMs in the source Location-A ESXi.
In the unlikely case those VMs are powered on in more than 30 minutes after they were unplugged from the source Location-A ESXi, NSX-T admin will have to recreate the NSX VM Tags information for those VMs.

"SRM Disaster Recovery" and "SRM Test" do not lose NSX VM Tags information, even if the recovery takes more than 30 Minutes.
In the "SRM Disaster Recovery" use case, the source Location-A ESXi are dead and do not update their NSX-T Manager about its VMs lost. So, the NSX-T Manager keeps those VMs and their NSX Tags information in its inventory forever, even when it detected those ESXi are disconnected.
And in the "SRM Test" use case, the source Location-A ESXi keep their VMs. So, the NSX-T Manager keeps those VMs and their NSX Tags information in its inventory, and when the Test VMs are plugged on the destination Location-B ESXi, they will get the NSX VM Tags.

3.4.2.5 (Special Use Case) Network Introspection and Endpoint Protection

As presented in chapter 3.2.2 Multisite Security Services, Network Introspection Host-Based and Endpoint Protections are supported with NSX-T Multisite.
However, the Partner Console must offer a DR solution with IP preservation (with SRM or vSphere-HA or other).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-47: NSX-T Multisite Network Introspection and Endpoint Protection across locations – After location failure

After the loss of Location-A, NSX-T Manager Cluster and Partner Console are lost.
The different options of NSX-T Manager Cluster recovery are detailed in the chapter 3.4.1 Management Plane.
The recovery of the Partner console is the responsibility of the partner (via SRM or vSphere-HA or other). The requirement on the NSX side is the Partner Console must offer a DR solution with IP preservation. Indeed, the recovered NSX-T Manager and Location-B Partner SVM will still talk to the Partner Console via its IP address configured in NSX-T Manager.

3.4.3 What about GSLB option

GSLB is a popular option for Disaster Recovery.
This option is for the incoming traffic only of the data plane (traffic from External to Inside) and is based on DNS. It does not cover the Management Plane recovery.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-48: Multisite with GSLB option - Before location failure

The same application is deployed in in different locations. In the figure above, the application in Location-A is reachable via IP1, in Location-B is reachable via IP2, and in Location-C is reachable via IP3.

Those applications are completely isolated and don't need any cross-location communication.

Users access that application via its FQDN (app1.xyz.com). The DNS in charge of resolving that FQDN is a GSLB solution. That GSLB solution is configured with all location IP: IP1 + IP2 + IP3 and continuously validates the application is running in each location.

The GSLB solution can be configured to resolve the FQDN with only one IP (active/standby) or multiple IP (active/active). In the figure above, the GSLB solution will resolve app1.xyz.com with IP1 only if configured in active/standby or resolve app1.xyz.com with IP1+IP2+IP3 if configured in active/active.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 3-49: Multisite with GSLB option - After location failure

After the loss of a Location-A, the GSLB solution will detect its failure and stop using it for its FQDN resolution.
In the figure above, now the GSLB solution will resolve app1.xyz.com with IP2 only if configured in active/standby or resolve app1.xyz.com with IP2+IP3 if configured in active/active.

The Data Plane service outage varies based on the GSLB location healthcheck interval, and the Time-To-Live (TTL) of the FQDN entry. It is usually around 5 minutes.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

More information on VMware GSLB solution Avi on https://avinetworks.com/docs/18.2/avi-gslb-overview/.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

3.5 Requirements and Limitations

The different requirements and limitations of the NSX-T Multisite solution have been detailed in the different chapters above. This chapter summarize them all.

NSX-T Multisite requirements:

- WAN
  - Maximum 150 milliseconds latency (RTT) between locations
  - Bandwidth large enough to accommodate cross location Management Plane + Data Plane
    - The Management Plane traffic is minimal with few Mbps at peak
    - The Data Plane traffic varies greatly between customers
    - In case of possible congestion cross location, it is recommended to configure QoS to prioritize NSX Management Plane traffic: NSX-T Managers to Transport Nodes (Edge Nodes and hypervisors)
  - MTU at least 1700 bytes (for TEP traffic)
    - Recommended 9000
  - IP connectivity
    - For Management Plane between NSX-T Managers and Transport Nodes (Edge Nodes and hypervisors)
    - For Data Plane between Transport Nodes.
  - Public IP@ (Segments, NAT, Load Balancer VIP) must be advertisable from both locations
    - In case of different Internet Providers (Verizon in Site-A and Orange in Site-B), both will advertise the public IP@ when then turn Active. In such case, be sure the public IP@ belong to the customer and not the Internet Provider.

- Automatic Disaster Recovery
  - Maximum 10 milliseconds latency (RTT) between locations
  - L2-VLAN Management stretch across the different locations
  - vCenter Cluster Stretched between locations with vSphere-HA

- Manual or Scripted Disaster Recovery
  - DNS name resolution for NSX-T Managers

NSX-T Multisite limitations:

- Networking
  - All Networking features are supported, but
    - No Local-Egress support
  - Each Tier-0 has all its North/South traffic via one single location
    - Different Tier-0 can have their North/South traffic via different locations though
  - No Tier-0 Active/Active (ECMP) with Automatic DR (see Note below)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

        - Automatic DR supports only Tier-0 Active/Standby
        - *Note: Automatic DR also supports Tier-0 Active/Active but only for metropolitan data centers (< 10 ms latency) and where asymmetric traffic is possible (no physical firewall cross data centers)*

- Security
    - All Security features are supported, but
        - Malware Prevention (NAPP does not support multi-location)
        - Network Detection and Response (NAPP does not support multi-location)
        - Network Introspection Cluster-based support
    - The below features are supported on NSX-T, but require partner support:
        - Network Introspection Host-based support
        - Endpoint Protection support

- Workload
    - Supported
        - Virtual Machines on ESXi
        - Physical Servers NSX prepared
        - Containers

3.6 Orchestration / Eco-System

NSX-T Multisite solution is based on NSX-T Data Center.
So, all orchestration tools (vRA, Terraform, Ansible, etc) and 3rd party solution (Skybox, Tufin, etc) supports this NSX-T Multisite solution.

3.7 Scale and Performance guidance

NSX-T Multisite solution is based on one NSX-T Manager Cluster.

So the scale of that solution is limited to the scale of one single NSX-T Manager Cluster capability.
NSX-T scale information is available on configmax.vmware.com.

And all performance considerations detailed in the complete VMware NSX-T Reference Design Guide apply for that solution too.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4 NSX-T Federation

At a high-level this solution is:

- one central NSX-T Global Manager Cluster (GM) offering central configuration of the network and security services for all locations
- one NSX-T Manager Cluster per location called here Local Manager (LM), managing Transport Nodes for that location (hypervisors and Edge nodes)

The GM pushes the network and security configuration to the different LM, which implements it locally.

Figure 4-1: NSX-T Federation Use Case

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.1 Architecture components

In addition to the traditional NSX-T architecture components: NSX-T Manager Cluster (LM), Edge Nodes, and hypervisors; NSX-T Federation solution relies on one extra component: NSX-T Global Manager Cluster (GM).

This chapter will detail first the Management Plane architecture, then the Data Plane architecture focusing on the changes brought by NSX-T Federation.

4.1.1 Management Plane

On the Management Plane, the NSX-T Federation solution is composed of two central NSX-T Global Manager Clusters (GM-Active and GM-Standby), and one NSX-T Manager Cluster per location called here Local Manager (LM).

Each GM Cluster is composed of three NSX-T Manager VMs of type NSX Global Manager.
The LM Cluster is composed of three NSX-T Manager VMs of type NSX-T Manager.
As explained in the VMware NSX-T Reference Design Guide, the NSX-T Manager VMs members of a cluster (LM or GM) can be on the same subnet or different subnets. Also, the maximum latency between any of the NSX-T Manager VMs is 10 milliseconds. At last, to operate, the NSX-T Manager Cluster can handle the loss of one of its NSX-T Manager VMs.

4.1.1.1 GM Cluster Deployments

There are two modes of NSX-T Federation deployments.

4.1.1.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations

For the Federation use case with 3 locations or more, and latency (RTT) below 10 milliseconds between each, and no congestion between those locations; it is recommended to have each location with its own LM Cluster, and one GM NSX-T VM in three different locations.
The typical use cases would be different buildings in metropolitan region where each require a dedicated Management Plane to be able to fully operate even in case of any other location failure. And at the same time, a central configuration is required for ease of operation.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-2: NSX-T Federation Management Plane – Use-case buildings in metropolitan region (< 10ms latency)

It's important to highlight in this NSX-T GM Cluster deployment, the loss of one location does not stop the GM Management Plane service, since the GM cluster has still 2 valid members. That's why there is no deployment of GM-Standby in this GM Cluster Deployment Mode1.
In the case of loss of one location, the LM Management Plane service does not stop on the other locations either, since the LM cluster has still 2 valid members

More information on that GM cluster deployment on "4.3.1.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations"

4.1.1.1.2 GM Cluster Deployment Mode2: NSX-T GM-Active and GM-Standby

For the Federation use case with 2 Locations, <u>or</u> latency (RTT) above 10ms across the locations; it is recommended to have all three NSX-T GM VMs in one single location.
The typical use cases would be two Data Centers only or Data Centers in large distance region (as represented in the figure below).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-3: NSX-T Multisite Manager Cluster– Use-case two Locations only and/or Data Centers far apart (> 10 ms latency)

It's important to highlight in this GM Cluster Deployment Mode2, the loss of the location hosting the GM-Active does stop the GM Management Plane service until the GM-Standby is made Active, however the LM Management Plane service does not stop on the different locations. The recovery of GM Management Plane is detailed in the chapter "4.4 Disaster Recovery".

More information on that GM cluster deployment on "4.3.1.1.2 GM Cluster Deployment Mode2: NSX-T "

<u>Note about maximum latency:</u>
The maximum latency (RTT) between GM-Active/GM-Standby, GM/LM, and LM/LM moves up from 150 milliseconds to 500 milliseconds. However, the maximum latency between Edge Nodes cross location remains at 150 milliseconds.
So, the maximum latency between locations is 500 milliseconds or 150 milliseconds latency if stretched T0/T1/Segments are configured. More information on stretched networks in chapter 4.2.1.1 Network Objects Span.

4.1.1.2 GM, LM, Edge Node Communication Flows

There are GM-GM, GM-LM, LM-LM, and EdgeNode–EdgeNode communication flows. NAT is not supported for those communication, and in case of firewall between GMs, LMs, and Edge Nodes specific ports have to be open (see https://ports.vmware.com/home/NSX-T-Data-Center).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Technical Note:
The GM to GM, GM to LM and LM to LM management plane and control plane synchronization is offered by the Async Replicator (AR) service. AR runs on port 1236 opened by the Application Proxy (APH) Service.
The APH connectivity is established between each GM-Active VMs to GM-Standby VMs.
The APH connectivity is also established between each GM-Active and Standby VMs to every location LM VMs.
Finally, the APH connectivity is also established between each location LM VM to every other location LM VMs.
But there is no APH connectivity between GMs inside a GM Cluster, nor LMs inside a LM Cluster.

The GM-Active also gets NSX objects status from registered LM IP/FQDN via 443. That's why registration with the LM Cluster VIP is strongly recommended for that channel availability even in the case of one LM VM loss.

4.1.1.2.1 GM-Active to GM-Standby Communication Flow

GM-Active synchronizes all received configuration to its GM-Standby.

Figure 4-4: GM-Active to GM-Standby communication flow

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In the figure above, you can see an example of Network configuration done on the GM-Active. GM-Active synchronizes this configuration to the GM-Standby.
The same synchronization would occur for any other configuration like Security configuration, or Principal Identity users, or vIDM users.
Important Note: vIDM/LDAP configuration is not synchronized to GM Standby (vIDM/LDAP users are).

For more information about the vIDM deployment in multiple data centers use case, see the white paper "Configuring VMware Identity Manager For Multiple Data Centers" https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/vmware-identity-manager-multiple-data-center-configuration.pdf.

4.1.1.2.2 GM to LM Communication Flow

In both GM cluster deployment modes, network and security is centrally configured and managed through the GM, which pushes the relevant network and security objects to the different LMs.

Figure 4-5: GM to LM communication flow

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In the figure above, you can see one “Tier-0 + Tier-1 + Segment stretched between Location-A and Location-B” configuration done on GM-Active. GM-Active pushes those GM objects to Location-A LM + Location-B LM. This configuration is not pushed to Location-C LM, since none of those objects are relevant to Location-C.
The GM objects received by LM are tagged as GM and are in read-only mode in LM. Only very specific fields of some specific objects can be edited on LM. Those are “Tier-0 interfaces + BGP”, and “Segment Ports” configuration.

It’s important to also note each LM can also receive network and security configuration directly. LM direct configuration is useful for specific network or security features currently not supported from GM, and/or to accept orchestration from tools not enhanced to talk to GM yet.
More information on supported LM configuration in chapter 4.1.1.5.1 Logical Configuration Ownership.

4.1.1.2.3 LM to LM Communication Flow

There are two cases where LM will synchronize information with other LMs: stretched NSX Group and stretched Segment. More information on Stretched Networking in chapter 4.2.1.1 Network Objects Span.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-6: LM to LM communication Flow – stretched NSX Group

In the figure above, one GM-Group1 stretched to all locations is created. Since that GM-Group1 is Global, it is pushed to all LMs.

Each LM receives that GM-Group1 and resolves its local membership: LM-A = VM1, LM-B = VM2, and LM-C = VM3. Then since that GM-Group1 is Global, each LM synchronizes its local membership with the other LMs. At the end of the synchronization, all LMs know about all remote LMs memberships.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-7: LM to LM communication Flow – stretched Segment

In the figure above, one GM-Segment stretched to all locations is created. Since that GM-Segment is Global, it is pushed to all LMs.
Each LM receives that GM-Segment and fills up its local Segment table: LM-A = Mac1/IP1, LM-B = Mac2/IP2, and LM-C = Mac3/IP3. Then since that GM-Segment is Global, each LM will synchronize its local Segment table with the other LMs. At the end of the synchronization, all LMs know about all remote LMs Segment table.

4.1.1.2.4 Edge Node to Edge Node Communication Flow

There is one use case where Edge Nodes will talk to remote Edge Nodes: when stretched networks are configured (Stretched T0/T1/Segment).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-8: Edge Node to Edge Node communication Flow

In the figure above, stretched networks are configured and East/West cross-location communication is done through Edge Nodes RTEP interface.

More information on Stretched Networks in chapter 4.2.1.1 Network Objects Span.

Note: In case of stretched networks, the maximum latency cross-locations is 150 milliseconds.

4.1.1.3 LM Registration and LM Onboarding

4.1.1.3.1 LM Registration (Addition)

For GM to centrally manage the network and security services of the different locations, each LM must be registered to GM.
LM registration is done from GM-Active, where the LM IP (or FQDN) and admin credentials are asked.
Attention, currently the LM Cluster VIP (or FQDN LM Cluster VIP) must be provided to allow the GM to LM communication to keep on even after one LM Management VM failure.
The IP or FQDN of any of the 3 LM must not be used.

Note: An external load balancer VIP for NSX-T Managers could also be used.
But each LM VM would need to be updated with the same node API certificate for the GM to accept communication to any of the LM VMs.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Once the LM Cluster VIP (or FQDN) and admin credentials are provided, **the GM will validate the LM license is Enterprise Plus and perform the LM registration**.
Then a final configuration step is required to finalize the registration: the LM RTEP configuration of I Edge Nodes (see chapter "4.1.2 Data Plane for more information on RTEP").

Figure 4-9: LM Registration

In the figure above, Location-A LM-A and Location-B LM-B are already registered. GM registers Location-C LM-C using its Cluster VIP + admin credentials. Once done, LM-C registration ends with Edge Nodes RTEP configuration.

From that point, GM can be used to centrally configure the network and security services of that location.
On the Network side from GM, new Tier-0/Tier-1/Segments can be created and pushed to that new registered LM. Also, existing Tier-0/Tier-1/Segments can be edited to be stretched to that new registered LM.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-10: Example of Network configuration on new LM registered

In the figure above, existing GM-Tier-0, GM-Tier-1, and GM-Segment are stretched to new Location-C after LM-C registration.

On the Security side from GM, all existing Groups and DFW Sections with a Global span are immediately pushed to that new registered LM. Also new Groups and DFW Sections for that sole location can be created and pushed to that new registered LM.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-11: Example of GM Global Security configuration on new LM registered

In the figure above, registered LM-A and LM-B use GM only for central security configuration and one GM-Group1 with span global is configured. GM-Group1 membership can be static or dynamic, and here its members are VM1 + VM2 + VM3 + VM4.
In the figure also, the network configuration is done on LMs with one LM-Tier-0, one LM-Tier-1, and one LM-Segment configured on each LM.
When LM-C is registered, it automatically receives all Groups and DFW Sections with a span global. So LM-C immediately receives the GM-Group1 with its member information.
Not represented in the figure above, if GM-Group1 were configured with dynamic membership (e.g. VM Tags) matching LM-C objects, then automatically those LM-C objects would be part of GM-Group1, and their IP would be synchronized between all LMs.

More information on the GM network and security services in the chapter "4.2 Network & Security services supported".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.1.1.3.2 (Optional) LM Onboarding (Import)

Once LM registration to GM has been performed and RTEP configuration has been done, GM can start creating network and security objects for that location.

If that LM already had some existing network and security configuration prior to its GM registration, then those objects remain on LM and GM does not have knowledge of those.

Before LM-C Registration – LM-C has local network and security configuration

After LM-C Registration – LM-C keeps its local network and security objects + gets global objects

Figure 4-12: Example of Security configuration on new LM registered

In the figure above, prior to LM-C registration, LM-C has some local network (LM-Tier-0, LM-Tier-1, and LM_Segment) and security configuration (LM-Group2 + LM-DFW). After its registration on GM, those network and security objects don't move up to GM.
And as discussed in the chapter above "4.1.1.3.1 LM Registration", the GM global security objects (GM-Group1) are automatically stretched to LM-C. The GM global network objects (GM-Tier-0,

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

GM-Tier-1, and GM-Segment) are not automatically stretched to LM-C. They can become extended to LM-C with a GM configuration update on those objects though.

However, there is an option to onboard LM objects to GM.
This onboarding option is transparent to the data plane, as it only changes the ownership of the network and security objects. So onboarding does not require a maintenance window.

This option has some strict requirements:

- It's a one-time option
  It can be called only once.

- It's a full LM configuration onboarding
  All the Network and Security objects will be moved to GM. It's not possible to do a partial configuration onboarding.

- Only Policy objects can be onboarded.
  All MP only objects will remain on LM.

- Onboarding is possible only if 100% of the existing LM Policy network and security configuration is supported by GM
  Onboarding will be prevented if any LM object is using a feature not supported by GM, such as VPN or Network Introspection.
  There is only one exception for specific Load Balancing configuration:
  - LB service must be on 1-arm load balancer (T1 1-arm attached with LB service)
  - If LM Pool is using Groups, then those groups must not be used in DFW sections

- All Segments and Edge nodes must be configured with a single TZ-Overlay which is defined as default (the default TZ-Overlay can be any TZ "nsx-overlay-transportzone" or other).
  No support of LM designs with multiple TZ-Overlays.

- In case of LM objects created by Principal Identity (PI) users, those PI users must be created on GM prior to the onboarding.

- In case of LM objects created by vIDM users, those objects will be onboarded to GM and the API information on the user who created the object ("_create_user") is kept.
  Even if that object is then updated by a GM user (admin / LDAP user / vIDM user); the original vIDM user information of who created the object is kept (the information of the user who modified it is saved in the API field "_last_modified_user").

- In case of LM objects created by LDAP users, those objects will be onboarded to GM and the API information on the user who created the object (API field "_create_user") is kept.
  Even if that object is then updated by a GM user (admin / LDAP user/ vIDM user); the original LDAP user information of who created the object is kept (the information of the user who modified it is saved in the API field "_last_modified_user").

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-13: Example of LM Onboarding

In the figure above, prior to LM-C onboarding, LM-C has some local network (LM-Tier-0, LM-Tier-1, LM_Segment, and one 1-arm T1-LB), and security configuration (LM-Group2 + LM-DFW).
After its onboarding to GM, those networks and security objects move up to GM with the exception of the 1-arm T1-LB.

Note: In case of the LM onboarding of objects with names already existing on GM, it's possible to imported objects and prepended or appended them with a suffix (configuration option).

Then once the LM configuration has been onboarded, the original LM-C networks onboarded to GM can be extended.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-14: Example of LM Onboarded Networks Extended

In the figure above, the LM-C onboarded networks Tier-0 and Tier-1 are edited and other locations are added to them. The Segments automatically get the new span of their attached Tier-0 or Tier-1.

Attention, the span of GM Groups and GM DFW Sections can not be changed. So those remain with their onboarded span: Location-C.
However, VMware offers a python script to update the GM Groups and GM DFW Sections span LM to make them span Global. That script is available on https://github.com/vmware-samples/nsx-t/blob/master/helper-scripts/Multi-Location/Federation/change_dfw_global.py

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

```
root@lab2-jumphost:~# python3.5 change_dfw_global.py

Connecting to Global Manager: 10.114.218.181
Got the following items:
  Domain: LM-London
      Group: London-Group
      SecurityPolicy: London-Section
  Domain: LM-Paris
      Group: Paris-Group1
      SecurityPolicy: Paris-Section
Writing existing config with above Groups + DFW-Policies (and its DFW-Rules) with existing span in file: fwll_backup.json
Writing new config with above Groups + DFW-Policies (and its DFW-Rules) with span Global in file: fwll_new.json
Continue with changing those objects span to Global in GM configuration ? (y/n): y
Applying new configuration. Please wait ...
Changes applied Successfully!
```

Figure 4-15: Python script to change GM Groups and GM DFW Sections to span Global

With the example above, the python script would change the Groups + DFW Sections with span LM-C to a span Global (DFW Section not represented in the figure below):

After After python script on Groups + DFW Sections with span LM-C => They move to a Span Global

Figure 4-16: Groups + DFW Sections with Span LM or Regional move to Span Global after running python script

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.1.1.4 Federation Regions

By default, GM creates one Global region which contains all the LMs and one region per LM which contains only the individual LM.
Then it's possible to configure extra regions which will contain one or multiple LM.
One specific LM can be only in one extra region in addition to its LM region and the Global region.

Figure 4-17: Example of Federation regions

In the figure above, there are 4 LMs: LM-Paris, LM-London, LM-Boston, and LM-Seattle.
The default GM regions are: Global, Paris, London, Boston, and Seattle.
Then the following two extra regions were created: EMEA region which contains LM-Paris and LM-London, and USA region which contains LM-Boston and LM-Seattle.

Specific regions allow you to create specific security policies (Groups and DFW Sections) that will be applied only to those specific regions.
Regions don't apply to network constructs (Tier-0, Tier-1, Segment, NAT, GW-FW), where each LM has to be individually selected on those stretch network elements.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.1.1.5 Logical Configuration and Infrastructure Ownership

4.1.1.5.1 Logical Configuration Ownership

As discussed in the chapter "4.1.1.2.2 GM to LM Communication Flow", network and security objects can be created on GM or LM.

Figure 4-18: Logical configuration ownership

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

As represented in the top part of the figure above, the GM created objects are pushed to LM. But those can only be modified/deleted by GM and are read-only on LM.
The only two exceptions are on:

- GM-Tier-0 where LM can edit its interface and BGP configuration
  It's to allow LM admin to be able to those changes quickly in case of emergency and GM admin is not available.
- GM-Segment-Ports where LM can edit TAG
  The goal is to allow orchestration tools, like vRA, to deploy applications (VMs) on existing GM-Segments and add TAG on those created LM-Segment-Ports. Then if GM Security Groups are based on Segment-Port-TAGs, those applications will automatically receive their appropriate security with no extra configuration.

As represented in the bottom part of the figure above, once LM has been registered to GM, most Network and Security objects can still be configured from LM directly. In such configuration, those LM created objects can only be modified/deleted by LM. And those LM objects are not seen by GM.

It's also possible in some cases to link those LM objects to GM objects as represented in the figure below where:

- LM Tier-1 is connected to a GM Tier-0
- LM DFW rule is created with LM and GM objects

Figure 4-19: LM configuration consuming GM objects

Warning: Since GM is not aware of the consumption of its objects by LM, GM will allow the deleting of those objects. However, LM won't allow their deletion and those GM objects will remain in LM configuration (in deleting greyed-out state) until their consumption is removed from its LM objects.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Below is the exhaustive list of supported LM Network and Security features configured from LM once registered by GM, and the ability to link those LM objects with GM objects:

LM configuration	Object	Support (Yes / No)
Networking		
	LM T1	• LM config consuming LM objects (LM_T1 connected to LM_T0) • LM config consuming GM objects (LM_T1 connected to GM_T0)
	LM Segment	• LM config consuming LM objects (LM_Segment connected to LM_T1) • No LM config consuming GM objects (LM_Segment connected to GM_T1), ability to ability create/update a Segment Port on GM_Segment
	LM L2-Bridge	• LM config consuming LM objects (LM_Segment) • No LM config consuming GM objects (LM_Segment)
	LM Edge NAT	• LM config consuming LM objects (LM_NAT on LM_T0/T1) • No LM config consuming GM objects (LM_NAT on GM_T1)
	LM LB	• LM config consuming LM objects (LM_LB on LM_T1) • No LM config consuming GM objects (LM_LB on GM_T1)
	LM VPN	• LM config consuming LM objects (LM_VPN on LM_T0/T1) • No LM config consuming GM objects (LM_VPN on GM_T0/T1)
Security		
	LM Group	• LM config consuming LM objects (LM_Group with LM_Members) • No LM config consuming GM objects (LM_Group with GM_Members), only exception LM_Group with Static Member = GM_Segment
	LM DFW	• LM config consuming LM objects (LM_DFW on LM_Group) • LM config consuming GM objects (LM_DFW on GM_Group)
	LM Firewall IPFIX	• LM config consuming LM objects (LM_FWIPFIX on LM_Group) • No LM config consuming GM objects (LM_FWIPFIX on GM_Group)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

	LM GWFW	• LM config consuming LM objects (LM_GWFW on LM_T0/T1) • LM config consuming GM objects (LM_GWFW on GM_T0/T1)
	LM IDFW	• LM config consuming LM objects (LM_IDFW on LM_Group) • LM config consuming GM objects (LM_GWFW on GM_Group)
	LM Security Profile (Session Time/ DNS/Flood)	• LM config consuming LM objects (LM_SecProf on LM_Group) • No LM config consuming GM objects (LM_SecProf on GM_Group)
	LM IDS/IPS	• LM config consuming LM objects (LM_IDS/IPS with LM_Group) • No LM config consuming GM objects (LM_IDS/IPS with Source=GM_Group), only exception LM_IDS/IPS with "LM_Group = Global_Segment Member" + "Service = GM_Service"
	LM Network Introspection*	• See below*
	LM Endpoint Protection*	• See below*
	LM Segment Security	• LM config consuming LM objects (LM_Segment with LM_SegSec) • LM config consuming GM objects (LM_Segment with GM_SegSec)
	LM Malware Prevention	Not supported.
	LM Network Detection and Response	Not supported.
	LM TLS Inspection	Not supported.
	NAPP	• NAPP can be deployed on LM • NAPP can not be deployed on GM
Monitoring		

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

	LM Switch IPFIX	• LM config consuming LM objects (LM_SwitchIPFIX on LM_Segment / LM_Groups / etc) • No LM config consuming GM objects (LM_SwitchIPFIX on GM_Segment / GM_Groups / etc)
	LM Port Mirroring	• LM config consuming LM objects (LM_PortMirror on LM_Segment / LM_Groups / etc) • No LM config consuming GM objects (LM_PortMirror on GM_Segment / GM_Groups / etc)

* **Network Introspection Host-Based and Cluster-Based** (previously named Service Insertion) are supported.
There are a couple of points to keep in mind:

- On NSX side
  - Usage of LM_Segment is required (no GM_Segment)
  - Usage of LM_Groups is required (no GM_Group as Partner Console does not collect/receive the GM Groups)
  - No Network Introspection on Edges T0/T1
- on Partner side (such as Palo Alto, CheckPoint, Fortinet, Netscout, etc):
  - Partner must validate NSX-T Federation support on their side

* **Endpoint Protection** (previously named Guest Introspection) is supported.
There are a couple of points to keep in mind:

- On NSX side
  - In case of Hosted Partner SVM failure on an ESXi, that ESXi does not redirect traffic to another ESXi Partner SVM. The traffic goes through without protection.
- on Partner side (such as Bitdefender, Trend Micro, etc):
  - Partner must validate NSX-T Federation support on their side

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-20: NSX-T Federation Network Introspection and Endpoint Protection

In the figure above, on the location Management Plane, each Partner Console receives its local LM Groups and pushes its configuration to the different Host-Based Partner SVM in its location (Partner Console does not receive GM Groups).
On the Data Plane, each ESXi redirects the VM traffic to its hosted Partner SVM.

For Network Introspection, in case of hosted Partner SVM failure in one of the ESXi, that ESXi will redirect its VM traffic new flows to another ESXi Partner SVM (existing flows will be dropped). That other ESXi will always be local as shown in the figure below.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-21: NSX-T Federation Network Introspection redirection – After Partner SVM failure

For Guest Introspection, in case of hosted Partner SVM failure in one of the ESXi, new files won't be inspected for the VMs on that ESXi until the recovery of its Partner SVM.

4.1.1.5.2 Infrastructure Ownership

Infrastructure objects (Edge Nodes, Edge Clusters, Hypervisors and Physical Servers preparation, Transport Zones, etc) are always created and managed by LM.

GM only consume those when needed for the GM-object creation, like Tier-1 Edge Cluster information for a specific LM location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-22: Infrastructure management

4.1.1.6 Federation API

NSX-T Federation API (GM API) is available on https://developer.vmware.com/apis/1230/nsx-t-global-manager.

NSX-T Federation API is very much like NSX-T LM API.
However, it has two key differences:

API URL Paths on GM:

GM API	LM API

`/global-manager/api/v1/global-infra/`
 	
Example:
`GET /global-manager/api/v1/global-infra/segments`
	`/policy/api/v1/infra/`
	
	Example:
	`GET /policy/api/v1/infra/segments`

New Tree "Global-Infra" on LM:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-23: GM and LM API trees

In the figure above, you can see an example of NSX-T GM tree structure: "/global-infra". The NSX-T LM has the tree structure "/infra" for configuration done directly on LM. And it has the tree structure "/global-infra" for configuration done from GM and pushed to LM.

It's important to note LM Intra tree objects can link to Global-Infra tree objects. This allows, for instance configuration such as LM 1-arm T1-LB to be connected to a GM Segment.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-24: LM trees example

<u>Note about the tree object "Domain":</u>
Domain is a root object on both trees "/infra" and "/global-infra".
Under the tree "/infra", there is only one domain "default". No other domain is allowed.
Under the tree "/global-infra", there is one domain "default", and one domain per LM "LM_name".
Then each region created will create a new domain "Region_name".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.1.2 Data Plane

GM works only on the Management Plane, pushing configuration to the different LMs.
LMs are in charge of the local Management Plane (configuration) and local Control Plane (Mac/IP/TEP tables).
Data plane elements are the Edge Nodes and hypervisors in the different locations.

Figure 4-25: NSX-T Federation Data Plane

North/South traffic is still processed by the Edge Nodes in the different locations.
East/West overlay traffic within a location is still processed by the hypervisors using their TEP interfaces. However unlike NSX-T Multisite, East/West overlay traffic cross-location is not processed between hypervisors TEP interfaces but by the Edge Nodes and their Remote TEP (RTEP) interfaces.

So with NSX-T Federation, Edge Nodes have 2 Overlay interfaces:

- TEP interfaces for the Overlay communication within a location to other local Edge Nodes and local hypervisors
  Each Edge Node can have multiple TEP IP (see VMware NSX-T Reference Design Guide).
  And fragmentation is not supported on TEP traffic.
- RTEP interface for the Overlay communication cross-locations to remote Edge Nodes
  Each Edge Node can have a single RTEP IP.
  And fragmentation is supported on RTEP traffic. However for best performance, it's recommended to avoid fragmentation (RTEP MTU greater or equal at 1700).

Note about RTEP Fragmentation:
In case of TCP traffic, the Edge Nodes will change the MSS and so the endpoints will notice they can't send full TCP packets:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-26: RTEP traffic for TCP applications

In case of Non-TCP traffic (such as UDP), there is no MSS mechanism and fragmentation will occur:

Figure 4-27: RTEP traffic for Non-TCP applications

More information on RTEP in the chapter “4.2.1.2 L2 Overlay Switching Service”.
And more information on North/South and East/West routing in the chapter “4.2.1.3 L3 Routing Service”.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.2 Network & Security services supported

NSX-T Data Center offers a very large number of Network & Security services and the NSX-T Federation solution currently support some of those.

On the Network side, features currently supported on GM are: Switching (Overlay and VLAN), IPAM (DHCP Relay and static binding, and DNS), Routing (NAT and route redistribution), Routing protocols (BGP, Static), and L2-Bridge.
The features not supported on GM are: T0-VRF, DHCP dynamic binding, Routing protocols (OSPF), Routing VPN and EVPN, Load Balancing, and Tier-0 and Tier-1 Active/Active with stateful services.

On the Security side, features currently supported on GM are: Distributed Firewall, Gateway Firewall, FQDN Filtering, L7 App ID context support, Time-Based Firewall, Exclusion list, and overall Enable/Disable of Location DFW.
The features not supported on GM are: URL Filtering, Identity Firewall, Distributed IDS, Gateway IDS/IPS, Malware Prevention, Network Detection and Response, Network Introspection, Endpoint Protection, Distributed Security only for vCenter VDS Port Group (*), and TLS inspection.

*: GM does not see the vCenter VDS Port Groups to assign them in Security Groups. However, GM can use Dynamic Membership in Groups based on vCenter VDS Port Groups Tags; those vCenter VDS Port Groups Tags being added by LMs.

Figure 4-28: NSX-T Federation Network and Security services

You can see NSX-T Administrator Guide here for more information on supported features at which NSX-T release.

You can also see the exhaustive list of supported LM Network and Security features configured from LM once registered by GM in the chapter 4.1.1.5.1 Logical Configuration Ownership.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.2.1 GM Network Services

This chapter will detail two new options brought by Federation: Network objects span and Tier-0/Tier-1 gateway primary and secondary locations.
Then all the supported network services will be detailed.

4.2.1.1 Network Objects Span

GM Tier-0, Tier-1, and Segment-Overlay objects can be defined as stretched (multi locations) or not stretched (single location).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-29: NSX-T Federation different supported Tier-0 / Tier-1 / Segments topologies

In the figure above, you can see different Tier-0, Tier-1, and Segment-Overlay with different span. There are few rules to keep in mind though:

- GM Tier-1 DR_Only span equals to attached T0 span (see bottom Tier-1)
- GM Tier-1 with SR spans is equal or a subset of T0 span (see all Tier-1 with services)
- GM Segment-Overlay span is always equal to its attached Tier-0 and Tier-1 span (see all Segments)
- GM Segment-Overlay is realized only when attached to Tier-0 or Tier-1 (see all Segments)

Note: In case of stretched networks, the maximum latency cross-locations is 150 milliseconds.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.2.1.2 L2 Overlay Switching Service

4.2.1.2.1 GM Segment Configuration Options

As described in chapter “4.2.1.1 Network Objects Span”, GM Segment-Overlay span is always equal to its attached Tier-0 and Tier-1 span. Also, GM Segments are realized only when attached to Tier-0 or Tier-1.

Figure 4-30: NSX-T Federation Segments topologies

4.2.1.2.2 GM Segment Data Plane

The L2 cross-location traffic is handled by the Edge Nodes. This is to avoid the management of many Tunnels/BFD between all hosts cross sites.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-31: NSX-T Federation cross-location logical and physical packet walk

Each stretched Segment cross-location communication is offered by different Edge Nodes in Active/Standby mode to offer Edge Node load distribution.
The selection of the pair of Edge Nodes is based on the segment attachment:

- Segments attached to T1_DR
  All segments connected to same T1_DR will use the same 2 Edge Nodes (some segments active on one, and some segments active on the other one).
  The 2 Edge Nodes selected for the segments of that T1_DR, are 2 Edge Nodes from the Edge Cluster used by the T0 connected to that T1_DR.
  Different T1_DR connected to the same T0 will use different 2 Edge Nodes if the Edge Cluster used by the T0 has more than 2 Edge Nodes.
- Segments attached to T1_SR
  All segments connected to same T1_SR will use the same 2 Edge Nodes (some segments active on one, and some segments active on the other one).
  The 2 Edge Nodes selected for the segments of that T1_SR, are the2 Edge Nodes hosting that T1_SR.
- Segments on T0 A/S
  All segments connected to same T0 A/S will use the same 2 Edge Nodes (some segments active on one, and some segments active on the other one).
  The 2 Edge Nodes selected for the segments of that T0 A/S, are 2 the Edge Nodes hosting that T0_SR.
- Segments on T0 A/A
  All segments connected to same T0 A/A will use the same 2 Edge Nodes (some segments active on one, and some segments active on the other one).
  The 2 Edge Nodes selected for the segments of that T0 A/A, are 2 Edge Nodes out of Edges Nodes used by that T0 A/A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The cross-location communication between Edge Nodes is using Remote TEP (RTEP) using Geneve encapsulation. Unlike TEP, only one RTEP IP is supported by Edge Node. Also, fragmentation on RTEP is allowed, but for best performance fragmentation should be avoided.
Selection of VNI-TEP is done by LM (not GM).
Selection of VNI-RTEP is done by GM and told to each LM. There is no incidence if one LM already uses the same VNI for its TEP, since TEP and RTEP VNI are in different zones.

For more information on Edge Node RTEP configuration, see chapter “4.3.2.1 Edge Node configuration for optimal performance”.

4.2.1.3 L3 Routing Service

4.2.1.3.1 GM Tier-0 and Tier-1 Gateway Configuration Options

As discussed in the chapter “4.2.1.1 Network Objects Span”, Tier-0 and Tier-1 gateways can be stretched across multiple locations.
When a Tier-0 gateway is stretched, the different locations of that gateway can be configured as “Primary” or “Secondary”. Same thing for Tier-1 gateway with services (Tier-1 SR+DR).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-32: NSX-T Federation cross-location Primary and Secondary Tier-0

Figure 4-33: NSX-T Federation cross-location Primary and Secondary Tier-1

In the two figures above, you can see different stretched Tier-0 and Tier-1 routing options.

The first routing option is “All Primary”. This option is only available for Tier-0 without Services. VMs South to North traffic is sent to their default gateway which is hosted by their local Edge Nodes. Last each Edge Node forwards it locally to the fabric.

The second routing option is “Primary/Secondary”. This option is available for Tier-0 and Tier-1 SR+DR. VMs South to North traffic is sent to their default gateway which is hosted by their local Edge Node. Then all Edge Nodes forward it to the Edge Nodes location hosting the Tier-0/Tier-1 Primary. Last those Edge Nodes forwards it locally to the fabric.

The last routing option is “T1-Stretched DR_Only”. This option is only available for Tier-1 without Services. VMs South to North traffic is sent to their default gateway which is hosted by their hypervisor. Last each hypervisor forwards it locally Tier-0 (not represented in the figure).

Within a location, the Gateways can be in Active/Active or Active/Standby mode. And in the specific case of T1-Stretched DR_Only, the router is only distributed and on the hypervisors.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-34: NSX-T Federation within a location different T0/T1 options.

In the figure above, you can see different availability options for Tier-0/Tier-1 within a location.

The first availability option is "Active/Active" (top-left figure). This option is only available for Tier-0 without Services. Within a location South to North traffic is sent evenly across all Edge Nodes part of the Tier-0. Last each Edge Node forwards it locally to the fabric.

The second availability option is "Active/Standby" (both right figures). This option is available for Tier-0 and Tier-1 with Services. Within a location South to North traffic is sent only to the Edge Node hosting the Tier-0 or Tier-1 active. Last this Edge Node forwards it locally to the fabric.

The last availability option is "T1 DR_Only" (bottom-left figure). This option is only available for Tier-1 without Services. Within a location South to North traffic is sent directly by all hypervisors. Last each hypervisor forwards it locally to the Tier-0 (not represented in the figure).

And combining the routing and availability options, you have the following Tier-0 and Tier-1 configuration options:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-35: NSX-T Federation all possible Tier-0 configuration options

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In the figure above, you can see different stretched Tier-0 configuration options.

The first configuration option is “Active/Standby Location Primary/Secondary”. This option is available for Tier-0 with Services. VMs South to North traffic is sent to their local Edge Node hosting the Tier-0-Active. Then those Edge Nodes forward it to the Edge Node hosting the Tier-0 Primary-Active. Last this Edge Node forwards it locally to the fabric.

The second configuration option is “Active/Active Location Primary/Secondary”. This option is available for Tier-0 without services. VMs South to North traffic is sent to their local Edge Nodes hosting the Tier-0-Active. Then those Edge Nodes forward it to the Edge Nodes hosting the Tier-0 Primary-Active. Last those Edge Nodes forwards it locally to the fabric.

The third configuration option is “Active/Active Location All Primary”. This option is available for Tier-0 only without services. VMs South to North traffic is sent to their local Edge Nodes hosting the Tier-0-Active. Then those Edge Nodes forward it locally to the fabric.

The fourth configuration option is “Active/Standby Location All Primary”. This option is not offered today.

Figure 4-36: NSX-T Federation all possible Tier-1 configuration options

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In the figure above, you can see different stretched Tier-1 configuration options.

The first configuration option is “Active/Standby Location Primary/Secondary”. This option is available for Tier-1 with Services. VMs South to North traffic is sent to their local Edge Node hosting the Tier-1-Active. Then those Edge Nodes forward it to the Edge Node hosting the Tier-1 Primary-Active. Last this Edge Node forwards it locally to the fabric.

The second configuration option is “T1 DR_Only”. This option is only available for Tier-1 without Services. VMs South to North traffic is sent directly by all hypervisors. Then those hypervisors forwards it locally to the Tier-0 (not represented in the figure).

Note: IPv4 and IPv6 routing are supported

4.2.1.3.2 Tier-0 Data Plane (South/North)

In the figures below, I have two locations with Internet access via Location-A and Storage network access via Location-B.
One stretched Tier-0 with different configuration, Tier-1 DR_Only, and Segment are configured across those two locations.
At last, two VMs are connected to that stretched Segment; VM1 is in Location-A and VM2 is in Location-B.

T0 Active/Standby Location Primary/Secondary
The different figures represent the South/North packet walk of the different use cases:
South/North traffic to Internet is always processed via the Edge Node hosting the Tier-0 Primary Active (EN2).
South/North traffic to Storage is always processed via the Edge Node hosting the Tier-0 Secondary Active (EN3).
South/North traffic that has to cross locations is always processed via the Edge Nodes hosting the Tier-0 Active (EN2/EN3) RTEP tunnels.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-37: NSX-T Federation T0 A/S LocP_S packet walk1

Figure 4-38: NSX-T Federation T0 A/S LocP_S packet walk2

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-39: NSX-T Federation T0 A/S LocP_S packet walk3

Figure 4-40: NSX-T Federation T0 A/S LocP_S packet walk4

T0 Active/Active Location Primary/Secondary

The different figures represent the South/North packet walk of the different use cases:
South/North traffic to Internet is always processed via the Edge Nodes hosting the Tier-0 Primary Active (EN1+EN2).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

South/North traffic to Storage is always processed via the Edge Nodes hosting the Tier-0 Secondary Active (EN3+EN4).
South/North traffic that has to cross locations is always processed via the Edge Nodes hosting the Tier-0 Active (EN1+EN2/EN3+EN4) RTEP tunnels.

Figure 4-41: NSX-T Federation T0 A/A LocP_S packet walk1

Figure 4-42: NSX-T Federation T0 A/A LocP_S packet walk2

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-43: NSX-T Federation T0 A/A LocP_S packet walk3

Figure 4-44: NSX-T Federation T0 A/A LocP_S packet walk4

T0 Active/Active Location All Primaries

In this Tier-0 configuration the two locations offer Internet access.
The different figures represent the South/North packet walk of the different use cases:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

South/North traffic to Internet is always processed via the local Edge Nodes hosting the Tier-0 Primary Active (EN1+EN2/EN3+EN4).

Figure 4-45: NSX-T Federation T0 A/A Loc_AllP packet walk1

Figure 4-46: NSX-T Federation T0 A/A Loc_AllP packet walk2

Attention, you have possible asymmetric routing with Tier-0 A/A Loc_AllP!

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-47: NSX-T Federation T0 A/A Loc_AllP asymmetric routing

In the figure above, the whole segment subnet (20.20.20.0/24) is advertised from both locations. The external App receives route from both locations and in this example selects Location-A for its response. The firewall in Location-A didn't see the initiated traffic from VM2 and so drops the App response.

Asymmetric routing for this Tier-0 A/A Loc_P/S is detailed in the chapter "4.2.1.3.5 Routing Protocols".

4.2.1.3.3 Tier-1 with Service Data Plane (South/North)

In the figures below, I have two locations with Internet access via Location-A and Storage network access via Location-B.
One stretched Tier-0 A/S Loc_P/S, T1 with SR, and Segment are configured across those two locations.
At last, two VMs are connected to that stretched Segment; VM1 is in Location-A and VM2 is in Location-B.

Other Tier-0 configurations could be used, such as Tier-0 A/A Loc_P/S or Tier-0 A/A Loc_AllP. The Tier-0 A/S Loc_P_S is chosen here to simplify the packet walks explanation.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The different figures represent the South/North packet walk of the different use cases:
Tier-1 South/North traffic to Internet is always processed via the Edge Node hosting the Tier-1 Primary Active (EN2).
Tier-1 South/North traffic that has to cross locations is always processed via the Edge Nodes Active (EN2/EN3) RTEP tunnels.
Then packet walk through Tier-0 follows the path explained in the previous section.

Figure 4-48: NSX-T Federation T1 with SR packet walk1

Figure 4-49: NSX-T Federation T1 with SR packet walk2

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-50: NSX-T Federation T1 with SR packet walk3

Figure 4-51: NSX-T Federation T1 with SR packet walk4

4.2.1.3.4 East/West with Service Data Plane

In the figures below, I have two locations with a Tier-0 A/S Loc_P/S.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Then each figure as a different type of Tier-1 or Tier-1s connected to Segments.
At last, VMs are connected to the different Segments.
Other Tier-0 configurations could be used, such as Tier-0 A/A Loc_P/S or Tier-0 A/A Loc_AllP.
The Tier-0 A/S Loc_P/S is chosen here to simplify the packet walks explanation.

The different figures represent the East/West packet walk of the different use cases:
Tier-1 East/West cross locations traffic is always processed by the Edge Node hosting the stretched Segment active or Tier-0/Tier-1 active.

Figure 4-52: NSX-T Federation T1 with SR packet walk1

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-53: NSX-T Federation T1 with SR packet walk2

Figure 4-54: NSX-T Federation T1 with SR packet walk3

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-55: NSX-T Federation T1 with SR packet walk4

4.2.1.3.5 Routing Protocols

Let's start with a recap of the different possible topologies with Stretched Tier-0 and Tier-1 gateways:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-56: NSX-T Federation T0 topologies

Stretched Tier-0 can be:

- Tier-0 Active/Standby Location Primary/Secondary (T0 A/S Loc_P/S)
- Tier-0 Active/Active Location Primary/Secondary (T0 A/A Loc_P/S)
- Tier-0 Active/Active Location All Primaries (T0 A/A Loc_AllP)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-57: NSX-T Federation T1 topologies

Stretched Tier-1 can be:

- Tier-1 Stretched DR_Only
- Tier-1 Stretched SR+DR with a specific location primary
- Tier-1 SR+DR Not-Stretched

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Tier-0 Stretched and Tier-1 Stretched use routing to exchange routes:

Figure 4-58: NSX-T Federation Routing Options

iBGP is internally used to exchange routes within the locations of the stretch Tier-0 gateways. eBGP and/or static routes are used to exchange routes between the Tier-0 gateway and physical fabric.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The Tier-0 routes exchange varies based on its configuration and is summarized in the tables below:

T0 A/S Loc_P/S	
eBGP receive	EN Primary: Yes EN Secondaries: Don't learn subnets received

eBGP advertise	EN Primary:
	• T1 DR_Only routes
	• T1 SR+DR routes
	  • Stretched Primary Local
	  • Stretched Primary Remote
	  • Not_Stretched Local
	  • Not_Stretched Remote
	• iBGP routes
	  • All
	
	EN Secondaries:
	• T1 DR_Only routes
	  • No
	• T1 SR+DR routes
	  • No
	• iBGP routes
	  • No

iBGP receive*	Edge Node Primary: Yes Edge Node Secondaries: Yes
iBGP advertise*	Edge Node Primary: Yes Edge Node Secondaries: Yes
FIB Order	EN Primary: • Static • iBGP • eBGP EN Secondaries: • iBGP • Static • eBGP

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/A Loc_P/S	
eBGP receive	EN Primary: Yes EN Secondaries: Yes

eBGP advertise	EN Primary:
	• T1 DR_Only routes
	• T1 SR+DR routes
	  • Stretched Primary Local
	  • Stretched Primary Remote
	  • Not_Stretched Local
	  • Not_Stretched Remote
	• iBGP routes
	  • All
	
	EN Secondaries:
	• T1 DR_Only routes
	  • *(User should add cost ++ to avoid asymmetric routing)*
	• T1 SR+DR routes
	  • Stretched Primary Local
	  • Not_Stretched Local
	  • *(User should add cost ++ to avoid asymmetric routing)*
	• iBGP routes
	  • No

iBGP receive*	Edge Node Primary: Yes Edge Node Secondaries: Yes
iBGP advertise*	Edge Node Primary: Yes Edge Node Secondaries: Yes
FIB Order	EN Primary and Secondaries: • Static • iBGP • eBGP *Note for Secondaries:* *Unlike usual routing, if same route received from multiple locations, then route from EN-Primary is kept (instead of eBGP route).*

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/A Loc_All_P	
eBGP receive	Edge Node Primaries: Yes

eBGP advertise	EN Primaries:
	• T1 DR_Only routes
	• T1 SR+DR routes
	  • Stretched Active Local
	  • Not_Stretched Local
	• iBGP routes
	  • No

iBGP receive*	Edge Node Primaries: Yes
iBGP advertise*	Edge Node Primaries: Yes
FIB Order	EN Primaries: • Static • eBGP • iBGP

*: *iBGP is automatically configured within stretched Tier-0 and can not be tuned (like set up filters).*

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

And to clarify it all, let's finish with the figures below to illustrate the routing in the different supported topology options:

T0 Active/Standby Location Primary/Secondary

*: To exit via Location-B for subnet 32.0.0.0/24, one Static Route must be manually configured on EN-LocB (same for Location-C).
Note: With the Static Route on EN-LocB, EN-LocC automatically learns the subnet 32.0.0.0/24 via EN-LocC. (same for 33.0.0.0/24 via Location-C).

If you want Location-B to access internal subnets directly via its local Edge Nodes LocC (not via LocA), you need to add routes on Location-B Fabric (same for Location-C).
The diagram below doesn't show those fabric routes.

Figure 4-59: T0 Active/Standby Location Primary/Secondary with T1 DR_Only

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

*: To exit via Location-B for subnet 32.0.0.0/24, one Static Route must be manually configured on EN-LocB (same for Location-C).
Note: With the Static Route on EN-LocB, EN-LocA automatically learns the subnet 32.0.0.0/24 via EN-Loc-B. (same for 33.0.0.0/24 via Location-C).

If you want Location-B to access internal subnets directly via its local Edge Nodes LocB (not via LocA), you need to add routes on Location-B Fabric (same for Location-C).
The diagram below doesn't show those fabric routes.

Figure 4-60: T0 Active/Standby Location Primary/Secondary with T1 SR+DR

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 Active/Active Location Primary/Secondary

Figure 4-61: T0 Active/Active Location Primary/Secondary with T1 DR_Only

Figure 4-62: T0 Active/Active Location Primary/Secondary with T1 SR+DR

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 Active/Active Location All Primaries

Figure 4-63: T0 Active/Active Location All Primaries with T1 DR_Only

Figure 4-64: T0 Active/Active Location All Primaries with T1 SR+DR

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.2.1.4 Stateful NAT Service

As explained in the VMware NSX-T Reference Design Guide, NAT is available on Tier-0 and Tier-1 Active/Standby, and the NAT function is offered by the Active element.
And all NAT sessions processed by the Active element are synchronized between the Active and Standby. So, in case of Active Edge Node failure, there is no data plane impact.

Figure 4-65: NSX Federation – NAT service

In case of stretched Tier-0, GM Stretched-T0 NAT configuration is pushed to the different LMs and is applied on the Edge Nodes Tier-0 based on the "Apply To" field:

GM NAT Apply To	Pushed to LM	Applied on Tier-0
All LM	All LM receives NAT configuration	Only T0-Primary gets the NAT configuration
Specific LM hosting the T0-Primary (*)	Specific LM receives NAT configuration	This location T0-Primary gets the NAT configuration
Specific LM hosting the T0-Secondary (*)	Specific LM receives NAT configuration	This location T0-Secondary gets the NAT configuration

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

*: The use case for specific GM Tier-0 NAT configuration per location is for the topologies where North/South is delivered through different locations, such as:

- T0 A/S Loc P/S with different Location exits for Internet and Storage:

Figure 4-66: NSX Federation – NAT service on T0 A/S Loc P/S with different location exits

In case of stretched Tier-1, GM Stretched-T1 NAT configuration is pushed to the different LMs and is applied on the Edge Nodes Tier-1 based on the "Apply To" field:

GM NAT Apply To	Pushed to LM	Applied on Tier-0
All LM	All LM receives NAT configuration	Only T1-Primary gets the NAT configuration
Specific LM hosting the T1-Primary (*)	Specific LM receives NAT configuration	This location T1-Primary gets the NAT configuration
Specific LM hosting the T1-Secondary (*)	Specific LM receives NAT configuration	This location T1-Secondary does NOT get the NAT configuration

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

*: The use case for specific GM Tier-1 NAT configuration per location is in case of DR and each location needs to do NAT with specific IPs:

Before Location Failure

After Location Failure and change T0/T1 Primary from Loc-A to Loc-B

Figure 4-67: NSX Federation – NAT service on T1 A/S Loc P/S with different NAT per location exit

4.2.1.5 DHCP (Relay and Static Binding) and DNS

DHCP Relay and static binding, as well as DNS are also supported from GM.

4.2.1.5.1 DHCP Server with DHCP Relay

The NSX DHCP Relay configuration is attached to a Tier-1.
GM DHCP Relay configuration is pushed to the different LMs hosting that Tier-1, but only the Edge Node hosting the Tier-1 Primary-Active has the DHCP service running (EN2 in the figure below).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-68: NSX-T Federation DHCP Server with DCHP Relay

The figure above shows the logical and physical packet walk of clients DHCP requests.
In both cases VM1_Location-A, and VM2_Location-B, the DHCP response is offered by the Edge Node hosting the Tier-1 Primary-Active (EN2). For the VM2_Location-B, the cross-location traffic is done between the Edge Node hosting Tier-1 Secondary-Active (EN3) and the Tier-1 Primary-Active (EN2).
Then the Tier-1 forwards clients DHCP requests to its configured external DHCP server.

4.2.1.5.2 DHCP Server with DCHP Static Bindings

The NSX DHCP Service configuration is attached to a Tier-1.
GM DHCP Service configuration is pushed to the different LMs hosting that Tier-1, but only the Edge Node hosting the Tier-1 Primary-Active has the DHCP service running (EN2 in the figure below).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-69: NSX-T Federation DHCP Server with DCHP static bindings

The figure above shows the logical and physical packet walk of clients DHCP requests and server DHCP response.
In both cases VM1_Location-A, and VM2_Location-B, the DHCP response is offered by the Edge Node hosting the Tier-1 Primary-Active (EN2). For the VM2_Location-B, the cross-location traffic is done between the Edge Node hosting Tier-1 Secondary-Active (EN3) and the Tier-1 Primary-Active (EN2).

4.2.1.5.3 DNS Service

The NSX DNS Service configuration is attached to a Tier-1.
GM DNS Service configuration is pushed to the different LMs hosting that Tier-1, but only the Edge Node hosting the Tier-1 Primary-Active has the DNS service running (EN2 in the figure below).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-70: NSX-T Federation DNS service

The figure above shows the logical and physical packet walk of clients DNS requests.
In both cases VM1_Location-A, and VM2_Location-B, the DNS request is forwarded to the Edge Node hosting the Tier-1 Primary-Active (EN2). For the VM2_Location-B, the cross-location traffic is done between the Edge Node hosting Tier-1 Secondary-Active (EN3) and the Tier-1 Primary-Active (EN2).
Then the Tier-1 forwards clients DNS requests to its configured external DNS server.

4.2.1.6 Load Balancing service (Avi)

As mentioned in the NSX-T 3.2.0 release notes here, NSX recommends the use of VMware NSX Advanced Load Balancer (Avi) and no more NSX-T native Load Balancer service.

The chapter goes over the design of our solution "Federation + Avi" to add load balancing service in a Federation environment.
Note: This chapter will not cover the basics of VMware NSX Advanced Load Balancer (Avi) and Avi technical knowledge is a pre-requirement.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In a Federation + Avi design, one Avi Controller Cluster per LM Cluster is deployed:

Figure 4-71: Federation with Advanced Load Balancing (Avi)

In each location, one Avi Controller cluster is deployed, and the local LM is configured as Avi NSX-T Cloud.
Each Avi Controller cluster offers the load balancing service for its location.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Load Balancing configuration is done by the Avi administrator individually on each Avi Controller:

Figure 4-72: Federation with Advanced Load Balancing (Avi) – LB Configuration

Each Avi Controller:

- creates the local load balancers Avi-SEs with Management interface and 1 or many Data interfaces
- pushes the load balancer configuration to Avi-SEs (VIPs, Pools, etc) through its Management interface

The Avi-SE Management interface can be connected to GM or LM Segment.
Each Avi-SE receive an IP from the Avi-Controller. The communication between the Avi-Controller and Ave-SE Management must be done without NAT.

The Avi-SE Data interface can be only connected to LM Segment.
So each location must have one LM_T1-Avi with 1 LM-Segment for the Avi Controller to connect its Avi-SEs Data interface.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

There are a couple of LM and Avi specific configurations / requirements in "Federation + Avi" environments:

Figure 4-73: Federation with Advanced Load Balancing (Avi) – Avi-SE / VIP / Pools

<u>LM_T1:</u>

LM_T1-AVI can be configured as T1_A/S or T1_DR modes.

Each LM_T1-Avi must be configured with "LB VIP redistribution" for the VIP to be automatically distributed to physical and logical networks, and with "Connected Segments redistribution" for the Pool Members to reply to the LB Avi-SE.

Figure 4-74: LM_T1-Avi Route Advertisement

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0:
T0 can be LM_T0 (A/A or A/S) or GM_T0 (any mode).
T0 must be configured with "LB VIP redistribution" for the VIP to be automatically distributed to the physical fabric.

Figure 4-75: T0 Route Advertisement

Optionally if external clients need to talk directly to VMs on the GM_Segment Orange (see Figure 4-73), then the T0 needs to redistribute internal T1 Segments as configured in the Figure 4-75 with "Advertised Tier-1 Subnets – Connected Interfaces & Segments".
This advertises all T1-Segments below the T0 which are configured with "Route Advertisement – All Connected Segments & Service Ports", so T1-Avi Segments too. If those subnets should be internal only and you don't want those to be advertised to the physical fabric, T0 needs to filter out the T1-Avi Segments:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-76: T0 Route Filtering

<u>Avi-SEs:</u>

Avi load balancers can be deployed in any mode: Active/Active or N+M or Active/Standby.

Figure 4-77: Avi-SE modes

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Avi VIPs:
VIPs must be configured with LB-SNAT (no Preservice Client IP).

Figure 4-78: Avi Application Profile without Preserve Client IP Address

In case of Disaster Recovery need, 2 options are possible:

- GSLB
  The same application runs in different locations behind different VIPs.
  This option is detailed in chapter 4.4.2.4.1 LB Disaster Recovery with GSLB.
- Non-GSLB
  The same application runs in different locations behind the same VIP.
  See Figure 4-73: Federation with Advanced Load Balancing (Avi) – Avi-SE / VIP / Pools with VIP1 in Location-A and Location-B.
  This option is detailed in the chapter 4.4.2.4.2 LB Disaster Recovery without GSLB.

In the Non-GSLB option, the VIP is active only in the primary location with the Virtual Service configured with Traffic Enabled, and other locations have the Virtual Service configured with Traffic Disabled:

Figure 4-79: Avi Virtual Service configured as "Traffic Disabled" in the secondary locations

Note: We recommend using "Traffic Disabled" instead of "overall Disabled" for faster Disaster Recovery.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

With the Virtual Service "overall disabled", Avi-SE load balancers are not pre-deployed in that location. So after the failure of the primary location and enabling the Virtual Service in the secondary location, the recovery of the load balancing service will take an extra 5 minutes for the Avi-SE VM to be deployed.

More information on Federation + Avi Disaster Recovery in chapter 4.4.2.4 Load Balancing Data Plane Recovery.

<u>Avi Pool Members:</u>
Pool members can be with Local VMs only and/or Remote VMs.
Pool members can use LM or GM Groups.

Figure 4-80: Example of Avi Pool Members based on GM Groups

Note: The Avi Controller collects groups from LM, the LM created groups and the GM created Groups. Currently, GM Groups collected from LM (such as "Web-VM-Group") are listed twice on Avi Controller "GM:Web-VM-Group" and "LM-Web-VM-Group".
Avi Pools can consume LM created Groups and GM Groups received, but GM Groups must be selected using the entry "LM-Web-VM-Group" (not "GM:Web-VM-Group") as shown in the figure above.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Load balancing traffic automatically enters where the Virtual Service is enabled:

Figure 4-81: Federation with Advanced Load Balancing (Avi) – Traffic Flow

Once the Virtual Service is enabled, the Avi Controller automatically configures the LM_T1 with a Static Route "VIP next hop = Avi-SE(s)".
Note: If the Avi-SE are deployed in Active/Standby; then only one next-hop is configured = Avi-SE Active. If the Avi-SE are deployed in Active/Active or N+M; then multiple next-hops are configured = Avi-SEs Active.

Then the LM_T1 redistribute its static route to the GM_T0, which redistributes it to the physical fabric in that location.

So traffic to the VIP enters via the GM_T0 (T0-Stretched-Slice-LocationA for VIP1 and T0-NonStretched for VIP2) to the LM_T1 to the Avi-SE(s).

At last the load balancer distributes the traffic to the pool members (local and/or remote).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.2.2 GM Security Services

This chapter will detail all the security services supported within Federation.

4.2.2.1 GM Groups

GM Groups configuration is very similar to LM Groups configuration.
As described in the chapter "4.1.1.4 Federation Regions", Federation brings a new option for Groups though: Region.
GM Groups can be defined as Global (all locations), Regional (multiple locations), or Local (single location).

Figure 4-82: NSX-T Federation Group Span

The Group1span is Global and is pushed to all LMs (LM-A + LM-B + LM-C).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The Group2 span is Regional (LM-A + LM-B).
The Group3 + Group4 + Group5 span are Local (each in a specific LM Location).
The membership of each group can be static or dynamic. The figure above shows dynamic membership with groups membership based on VM Tags.

And for each Global or Regional Group, each LM synchronizes its local members with the other LMs in the Group span. With the example of the figure above:

	LM-A	LM-B	LM-C
Group1 *(Loc ABC)*	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6
Group2 *(Loc AB)*	VM7 + VM8 + VM9 + VM10	VM7 + VM8 + VM9 + VM10	Don't have the Group
Group3 *(Loc A)*	VM11 + VM12	Don't have the Group	Don't have the Group
Group4 *(Loc A)*	VM13 + VM14	Don't have the Group	Don't have the Group
Group5 *(Loc B)*	Don't have the Group	VM15 + VM16	Don't have the Group
Group6 *(Loc C)*	Don't have the Group	Don't have the Group	VM17 + VM18

4.2.2.2 GM Distributed Firewall (DFW)

GM DFW configuration is very similar to LM DFW configuration.
Federation brings a new option though for DFW Sections: Region. Region can be considered as the "span" of the DFW Section (called DFW Policy in UI).
GM DFW Sections can be defined as Global (all locations), Regional (multiple locations), or Local (single location). And the DFW Rules within a section have the same span as the DFW Section.

Figure 4-83: NSX-T Federation DFW Span

All other GM DFW configuration options are the same as LM DFW options.
Each section has an "Applied-To" to define the scope of that section (which VMs will receive those section rules). If no specific "Applied-To" is configured at the section level, an "Applied-To" can be defined at the rules level to define the scope of that rule.
For further details, review the complete [VMware NSX-T Reference Design Guide.]

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

However, there are few constraints in the creation of GM DFW Sections and Rules.

- GM can create FW in all categories but “Ethernet” and “Emergency” (only LM can create FW in “Ethernet” and “Emergency” categories)
- GM can not see nor edit the LM “Default Layer3 Section” in section “Application”
- GM DFW Rules
    - in DFW Section Local
      must have Source or Destination equals to ANY or Groups which belong to the same location.
    - in DFW Section Region
      must have Source or Destination equals to ANY or Groups which belong to the same region or location within that region.
    - in DFW Section Global
      Source or Destination can be anything.

The created GM DFW Sections and Rules are pushed to the LM associated to the span of the GM DFW Section. For instance, the DFW Sections Global are pushed to all LMs, and the DFW Sections Paris are pushed only to LM-Paris.
LM receives the GM DFW Sections in each category (Infrastructure, Environment, Application) and always place them on top of its LM DFW Sections within each category. To have an LM DFW rule above any GM DFW Section, then LM can create DFW Rules under “Emergency” category”.

NSX-T 3.2 also brings in GM Firewall draft to save (auto or manually), and immediately publish or save for publishing at a later date.

GM DFW Examples

Different examples of DFW Rules are detailed in the examples below.
There are all based on the following VM locations and groups:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-84: NSX-T Federation Security examples

Each DFW rule can be "Correct" ✔, "Not Optimal / Maybe Improper" ⚠, or "Incorrect (rejected by GM)" ✖.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

DFW Section Span Global:

Section	Region: Global - Applied To: DFW						
		Source	Destination	Service	Profile	Applied To	Action
✓	Rule11	Group1*(Loc ABC)*	Group1*(Loc ABC)*	HTTP	None	Group1*(Loc ABC)*	Reject
⚠	Rule12	Group1*(Loc ABC)*	Group1*(Loc ABC)*	HTTPS	None	DFW	Deny
✓	Rule13	Group3*(Loc A)*	Group5*(Loc B)*	SMTP	None	Group3*(Loc A)* + Group5*(Loc B)*	Drop
⚠	Rule14	Group1*(Loc ABC)*	Group2*(Loc AB)*	HTTPS	None	DFW	Allow

The rule12 is not optimal for scale. This rule is applied to all VMs in all three locations, and not only the VMs of Group1 (such rule should have an "Applied To = Group1" to be optimal for scale).

The rule 14 is also not optimal for scale. This rule is applied to all VMs in all three locations, and not only the VMs of Group1 + Group2 (such rule should have an "Applied To = Group1 + Group2" to be optimal for scale).

The rules 13 and 14 will change the span of Group2, Group3 and Group5.
The span of those groups is not Global, but since these groups are now consumed in a DFW Section Global, GM updates their span to Global.

	LM-A	LM-B	LM-C
Group1*(Loc ABC)*	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6
Group2*(Loc AB)*	VM7 + VM8 + VM9 + VM10	VM7 + VM8 + VM9 + VM10	VM7 + VM8 + VM9 + VM10 (added with rule 13)
Group3*(Loc A)*	VM11 + VM12	VM11 + VM12 (added with rule 13)	VM11 + VM12 (added with rule 13)
Group4*(Loc A)*	VM13 + VM14	Don't have the Group	Don't have the Group
Group5*(Loc B)*	VM15 + VM16 (added with rule 14)	VM15 + VM16	VM15 + VM16 (added with rule 14)
Group6*(Loc C)*	Don't have the Group	Don't have the Group	VM17 + VM18

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

DFW Section Span Region:

Section	Region: Region1 (Loc A + Loc B) - Applied To: DFW						
		Source	Destination	Service	Profile	Applied To	Action
✓	Rule21	Group2*(Loc AB)*	Group2*(Loc AB)*	LDAP	None	Group2*(Loc AB)*	Reject
⚠	Rule22	Group3*(Loc A)*	Group5*(Loc B)*	LDAPS	None	DFW	Allow
⚠	Rule23	Group2*(Loc AB)*	Group1*(Loc ABC)*	IMAP	None	DFW	Drop
✓	Rule24	Group5*(Loc B)*	Group2*(Loc AB)*	NTP	None	Group5*(Loc B)* + Group2*(Loc AB)*	Allow
⚠	Rule25	Any	Group6*(LocC)*	HTTP	None	DFW	Allow

The rule 22 is not optimal for scale. This rule is applied to all VMs in Location A + Location B, and not only the VMs of Group3 + Group5 (such rule should have an "Applied To = Group3 + Group5" to be optimal for scale).

The rule 23 may be Improper. This rule is applied to all VMs in Location A + Location B and is missing VM5 + VM6. VM5 + VM6 may block that traffic (such rule should be in a section Global to be able to be pushed to all VMs in Group1).

The rule 25 may be improper. This rule is applied to all VMs in Location A + Location B and is missing Group6. VM17 + VM18 may block that traffic (such rule should be in a section Location C or Global to be able to be pushed to VMs in Group6).

The rules 22, 24 and 25 will change the span of Group3, Group5 and Group6.
The span of those groups is not in this Region, but since these groups are now consumed in a DFW Section with this Region, GM updates its span to this Region.

	LM-A	LM-B	LM-C
Group1*(Loc ABC)*	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6
Group2*(Loc AB)*	VM7 + VM8 + VM9 + VM10	VM7 + VM8 + VM9 + VM10	Don't have the Group
Group3*(Loc A)*	VM11 + VM12	VM11 + VM12 (added with rule 22)	Don't have the Group
Group4*(Loc A)*	VM13 + VM14	Don't have the Group	Don't have the Group
Group5*(Loc B)*	VM15 + VM16 (added with rule 24)	VM15 + VM16	Don't have the Group
Group6*(Loc C)*	VM17 + VM18 (added with rule 25)	VM17 + VM18 (added with rule 25)	VM17 + VM18

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

DFW Section Span Local:

Section	Region: Loc A - Applied To: DFW					
	Source	Destination	Service	Profile	Applied To	Action
Rule31	Group3*(Loc A)*	Group3*(Loc A)*	POP3	None	Group3*(Loc A)*	Reject
Rule32	Group3*(Loc A)*	Group3*(Loc A)*	POP3S	None	DFW	Allow
Rule33	Group4*(Loc A)*	Group5*(Loc B)*	HTTP	None	DFW	Allow
Rule34	Group5*(Loc B)*	Group6*(Loc C)*	SSH	None	DFW	Drop
Rule35	Any	Group6*(Loc C)*	DNS	None	DFW	Allow

The rule 32 is not optimal for scale. This rule is applied to all VMs in Location A, and not only the VMs of Group3 (such rule should have an "Applied To = Group3" to be optimal for scale).

The rule 33 may be improper. This rule is applied to all VMs in Location A and is missing Group5. VM15 + VM16 may block that traffic (such rule must be in a section Region1 or Global to be able to be pushed to VMs in Group5).

The rule 34 is improper. Source AND Destination do not belong to Location A.

The rule 35 may be improper. This rule is applied to all VMs in Location A and is missing Group6. VM17 + VM18 may block that traffic (such rule must be in a section Global to be able to be pushed to VMs in Group6).

The rules 33 and 35 will change the span of Group5 and Group6.
The span of those groups is not in this Location, but since these groups are now consumed in a DFW Section with this Location, GM updates its span to this Location.

	LM-A	LM-B	LM-C
Group1*(Loc ABC)*	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6	VM1 + VM2 + VM3 + VM4 + VM5 + VM6
Group2*(Loc AB)*	VM7 + VM8 + VM9 + VM10	VM7 + VM8 + VM9 + VM10	Don't have the Group
Group3*(Loc A)*	VM11 + VM12	Don't have the Group	Don't have the Group
Group4*(Loc A)*	VM13 + VM14	Don't have the Group	Don't have the Group
Group5*(Loc B)*	VM15 + VM16 (added with rule 33 + 34)	VM15 + VM16	Don't have the Group
Group6*(Loc C)*	VM17 + VM18 (added with rule 35)	Don't have the Group	VM17 + VM18

4.2.2.3 GM Gateway Stateful Firewall

As explained in the VMware NSX-T Reference Design Guide, Gateway Firewall is available on Tier-0 and Tier-1 Active/Standby, and the Gateway Firewall function is offered by the Active element.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-85: NSX Federation – Gateway Firewall service

In case of stretched Tier-0 and Tier-1, Gateway Firewall is available on Tier-0 and Tier-1 Active/Standby Location Primary/Secondary, and the Gateway Firewall function is offered by the Primary-Active element.

All Gateway Firewall sessions are synchronized between the primary-active and primary-standby. So, in case of primary-active Edge Node failure (EN2 in the figure above), there is no data plane impact.
Loss of a location (loss of EN1 + EN2 in the figure above), and data plane recovery is covered in the section 4.4.2.1 Automatic Network Data Plane Recovery.

<u>Note about the Region:</u>
Region construct is only used for DFW. The span of Tier-0 and Tier-1 is defined by its selected locations.

<u>Note about the Groups: used in Gateway Firewall Rules:</u>
Unlike DFW, Groups used in Gateway Firewall Rules must have the same span as the Gateway.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T1GW Firewall Configuration:

	Source	Destination	Service	Profile	Applied To	Action
Rule11	any	Group2(Loc B)	HTTPS	None	T1	Allow

T1GW Firewall Implementation on Active EN1/EN2:

	Source	Destination	Service	Profile	Applied To	Action
Rule11	any	No IP	HTTPS	None	T1	Allow

Figure 4-86: NSX Federation – Gateway Firewall with rule using a group of different span

In the figure above, one Tier-1 Gateway is stretched across Location-A / Location-B / Location-C and has a Gateway Firewall rule using the Group2 that has a span of Location-B only.
Since the Group2 has a span of Location-B only, this group and its members are known only by LM Location-B. So LM Location-A does not know that group and cannot push its IP members to its EN1/EN2 hosting the active Tier-1.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.3 Best Practice Design

Federation does not change the LM best practice design, as detailed in the VMware NSX-T Reference Design Guide.

This chapter focuses only on the Federation Best Practice Design.
The first section will cover the Management side with the best GM deployment options and Security configuration for best scale.
The second section will cover the Data Plane side with the Edge Node and Tier-0 configuration for most services and best performance.

4.3.1 Management Plane

4.3.1.1 GM Cluster VMs deployment

As presented in the chapter "4.1.1 Management Plane", Federation solution offers two options for the GM Management Plane architecture based on the different cases such as the latency (RTT) between locations, numbers of location.

4.3.1.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations

In case of 3 locations or more in a metropolitan region (< 10 ms RTT) with no congestion, it is recommended to have GM VMs in different locations.
This GM Cluster Deployment Mode1 does not need GM-Standby since the loss of one location does not stop the GM Management Plane service.

Use Case1: Stretched VLAN Management across locations
If the multiple locations offer stretched VLAN Management, then the GM Cluster VIP can be used.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-87: NSX-T Federation Management Plane – Use-case metropolitan region with stretched Management VLAN

In the figure above, the 3 Global Managers VMs are physically in different locations but connected to the same Management VLAN.
They have a Cluster VIP configured, and that's how the GM service is accessed.

This mode offers automatic GM Service recovery against location failure.
Disaster recovery will be detailed in the chapter "4.4 Disaster Recovery".

Note: An external load balancer could also be used in that case to offer high availability of the GM service, as well as load balancing of its service (not represented in the figure above). The active and standby external load balancers would be both connected to the same VLAN Management but in different locations.

<u>Use Case2: No Stretched VLAN Management across locations</u>

If the multiple locations do not offer stretched VLAN Management, then a Global Server Load Balancing service (GSLB) has to be configured for the GM service.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-88: NSX-T Federation Management Plane – Use-case metropolitan region without stretched Management VLAN

In the figure above, the 3 Global Managers VMs are physically in different locations and connected to different Management VLANs.
The Global Managers can't get a Cluster VIP configured since they are on different subnets. The GM FQDN (gm.nsx.xyz.com) is resolved by the GSLB service. This one validates the availability of each GM VM and resolves its FQDN with the IP of GM VMs running. The IP resolution can be done in an active/standby/standby mode or active/active/active mode.

This mode offers automatic GM Service recovery against location failure.
Disaster recovery will be detailed in the chapter "4.4 Disaster Recovery".

Note: An external load balancer cannot be used in that case. Indeed, its external load balancer VIP would belong to one of the locations Management subnet and its active and standby appliances would be in that specific location. So in case of loss of that location, the GM access via this external load balancer VIP would be down.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.3.1.1.2 GM Cluster Deployment Mode2: NSX-T GM-Active and GM-Standby

In the use case with 2 Locations, or latency (RTT) above 10ms across the locations; it is recommended to have Global Manager Cluster Active and Global Manager Cluster Standby in dedicated locations.

Figure 4-89: NSX-T Multisite Manager Cluster– Use-case two Locations only and/or Data Centers far apart (> 10 ms latency)

In the figure above, the Global Managers Active VMs are physically in one single location and the Global Managers Standby VMs are physically in another location.

This mode does offer manual GM Service recovery against location failure.
Disaster recovery will be detailed in the chapter "4.4 Disaster Recovery".

4.3.1.2 Security configuration for best scale

For best scale and performance, it is recommended to have the "Section Span", "Source / Destination", and "Applied To" matching the need; so rules are pushed only to relevant LMs and to all relevant VMs.

As presented in the chapter "4.2.2.2 GM Distributed Firewall (DFW)", NSX-T allows DFW not optimal or improper configuration.

For instance, the rule12 below is not optimal for scale since this rule is applied to all VMs in all 3 locations, and not only the VMs of Group1.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Section	Region: Global - Applied To: DFW						
		Source	Destination	Service	Profile	Applied To	Action
⚠	Rule12	Group1*(Loc ABC)*	Group1*(Loc ABC)*	HTTPS	None	DFW	Reject

The optimal configuration of this rule is:

Section	Region: Global - Applied To: DFW						
		Source	Destination	Service	Profile	Applied To	Action
✓	Rule12	Group1*(Loc ABC)*	Group1*(Loc ABC)*	HTTPS	None	Group1*(Loc ABC)*	Reject

And the rule 23 is improper since this rule is applied to all VMs in Location A + Location B and is missing VM5 + VM6, and not optimal for scale since this rule is applied to all VMs in Location A and Location B

Section	Region: Region1 (Loc A + Loc B) - Applied To: DFW						
		Source	Destination	Service	Profile	Applied To	Action
⚠	Rule23	Group2*(Loc AB)*	Group1*(Loc ABC)*	IMAP	None	DFW	Drop

The optimal configuration of this rule is:

Section	Region: Global - Applied To: DFW						
		Source	Destination	Service	Profile	Applied To	Action
✓	Rule23	Group2*(Loc AB)*	Group1*(Loc ABC)*	IMAP	None	Group1*(Loc ABC)* + Group2*(Loc AB)*	Drop

4.3.1.3 Upgrade Federation

This section describes the steps to upgrade a Federated environment from 3.2/4.0 to 4.1.0

4.3.1.3.1 From NSX-T 3.2.x (prior to 3.2.2) to NSX 4.1.0

This section is for NSX-T 3.2 releases prior to NSX-T 3.2.2.

Day0: Federation environment running 3.2.x

- LM-LocationA: NSX-T 3.2.x
- LM-LocationB: NSX-T 3.2.x
- LM-LocationC: NSX-T 3.2.x
- GM-Active: NSX-T 3.2.x
- GM-Standby: NSX-T 3.2.x

At Day0, all NSX elements (LM and GM) are running 3.2.x.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Day1: Upgrade GM to 4.0.x

- LM-LocationA: NSX-T 3.2.x
- LM-LocationB: NSX-T 3.2.x
- LM-LocationC: NSX-T 3.2.x
- **GM-Active: NSX-T 4.0.x**
- **GM-Standby: NSX-T 4.0.x**

At Day1, the GM Active and Standby are upgraded to 4.0.x, starting with the GM-Standby.
There is no data plane interruption during each GM upgrade.
There is no management plane interruption in each location during the GM upgrade (new VMs can't be connected to Segments, existing VMs can't be vmotioned, etc).

Day2: Upgrade LM to 4.1.x

- **LM-LocationA: NSX-T 4.1.0**
- **LM-LocationB: NSX-T 4.1.0**
- **LM-LocationC: NSX-T 4.1.0**
- GM-Active: NSX-T 4.0.x
- GM-Standby: NSX-T 4.0.x

At Day2, all the LM are upgraded to 4.1.0.
The LM upgrades can be done on any order, and at any pace.
There is no data plane interruption during each LM upgrade.
Each location will have a management plane interruption during the LM Manager Cluster upgrade (new VMs can't be connected to Segments, existing VMs can't be vmotioned, etc). However, configuration for that location on GM can still be done even during the LM Manager Cluster upgrade. GM will re-synchronize the full configuration after the LM upgrade.

Day3: Upgrade GM to 4.1.0

- LM-LocationA: NSX-T 4.1.0
- LM-LocationB: NSX-T 4.1.0
- LM-LocationC: NSX-T 4.1.0
- **GM-Active: NSX-T 4.1.0**
- **GM-Standby: NSX-T 4.1.0**

At Day3, the GM Active and Standby are upgraded to 4.1.0, starting with the GM-Standby.
There is no data plane interruption during each GM upgrade.
There is no management plane interruption in each location during the GM upgrade (new VMs can't be connected to Segments, existing VMs can't be vmotioned, etc).

4.3.1.3.1 From NSX-T 3.2.2+ to NSX 4.1.x

This section is for NSX-T 3.2 releases from NSX-T 3.2.2.
The release NSX-T 3.2.2 has been released after NSX 4.0, and so it's not possible to upgrade a Federation environment to NSX 4.1 with the GM transition step (upgrade to NSX 4.0) described in the chapter above.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The release NSX-T 3.2.2 will support upgrades directly to NSX 4.1.1, which is the first NSX release that will support N+/-2 releases.

4.3.1.3.2 From NSX 4.0.x to NSX 4.1.0

This section is for NSX-T 4.0.x releases.

Day0: Federation environment running 4.0.x

- LM-LocationA: NSX-T 4.0.x
- LM-LocationB: NSX-T 4.0.x
- LM-LocationC: NSX-T 4.0.x
- GM-Active: NSX-T 4.0.x
- GM-Standby: NSX-T 4.0.x

At Day0, all NSX elements (LM and GM) are running 4.0.x.

Day2: Upgrade LM to 4.1.0

- **LM-LocationA: NSX-T 4.1.0**
- **LM-LocationB: NSX-T 4.1.0**
- **LM-LocationC: NSX-T 4.1.0**
- GM-Active: NSX-T 4.0.x
- GM-Standby: NSX-T 4.0.x

At Day1, all the LM are upgraded to 4.1.0.
The LM upgrades can be done on any order, and at any pace.
There is no data plane interruption during each LM upgrade.
Each location will have a management plane interruption during the LM Manager Cluster upgrade (new VMs can't be connected to Segments, existing VMs can't be vmotioned, etc). However, configuration for that location on GM can still be done even during the LM Manager Cluster upgrade. GM will re-synchronize the full configuration after the LM upgrade.

Day2: Upgrade GM to 4.1.0

- LM-LocationA: NSX-T 4.1.0
- LM-LocationB: NSX-T 4.1.0
- LM-LocationC: NSX-T 4.1.0
- **GM-Active: NSX-T 4.1.0**
- **GM-Standby: NSX-T 4.1.0**

At Day2, the GM Active and Standby are upgraded to 4.1.0, starting with the GM-Standby.
There is no data plane interruption during each GM upgrade.
There is no management plane interruption in each location during the GM upgrade (new VMs can't be connected to Segments, existing VMs can't be vmotioned, etc).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.3.2 Data Plane

4.3.2.1 Edge Node configuration for optimal performance

As presented in the chapter "4.1.2 Data Plane", Federation Data Plane architecture is very similar to the Local Data Plane architecture. The main addition is the Cross-Location traffic provided via Edge Node RTEP.

Figure 4-90: NSX-T Federation Data Plane

So with Federation, the different Edge Node traffic types are:

- TEP for the intra-site overlay traffic
- RTEP for the inter-site overlay traffic
- VLAN for the T0 North/South traffic
- and VLAN for L2-Bridging traffic.

For best performance, it is recommended to have dedicated interface(s) for each type of traffic. However since this would require a large number of interfaces (and Edge Node VM supports only 4 vNICs), other designs are also valid options.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.3.2.1.1 Recommended design for Edge Node VM

The designs below are based on the recommended design for Edge Node VM in the VMware NSX-T Reference Design Guide.

Edge Node VM on ESXi host with 2 pNICs:

Figure 4-91: NSX-T Federation Edge VM with ESXi 2 pNICS recommended configuration

As documented in the VMware NSX-T Reference Design Guide, the figure above shows two Edge Nodes on the same ESXi. Obviously for high availability one Tier-0 should always be deployed on different Edge Nodes hosted on different ESXi. It is assumed in the figure above that Tier-0 is actually deployed on 4 Edge Nodes hosted on 2 ESXi (with only 1 ESXi represented here).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Also as documented in the VMware NSX-T Reference Design Guide, one N-VDS (N-VDS1) is created with 3 Uplink Teaming Policies.

- The "Default" teaming policy with 2 active uplinks (Load Balance Source) associated to vNIC2 + vNIC3
  It is used by TEP. Each Edge Node receives 1 TEP-IP for each uplink.
  It is also used by RTEP. Each Edge Node receives 1 RTEP-IP for the first uplink interface.
  Note: In a future release, each Edge Node will receive 1 RTEP-IP for each uplink.
- The "Left" teaming policy with 1 single active uplink associated to vNIC2
  It is used by one Segment-VLAN (light blue) connected to Tier-0 uplink.
- The "Right" teaming policy with 1 single active uplink associated to vNIC3
  It is used by one Segment-VLAN (dark blue) connected to Tier-0 uplink.

The ESXi hosting the Edge Node VMs is configured with different Port Groups to send TEP-1 / RTEP / first Segment-VLAN primary via its pNIC1, and to send TEP-2 / second Segment-VLAN primary via its pNIC2.

The ToR are configured with all the VLANs on all interfaces, but VLAN 300 and 400 (T0 uplinks) which are configured respectively only on interface to ESXi-P1 and ESXi-P2.
ToR router IP of VLAN TEP and VLAN RTEP are active/standby on ToR-Left/ToR-Right, Segment VLAN Light Blue is active on ToR-Left, and Segment VLAN Dark Blue is active on ToR-Right.

And in case of bridging directly configured on the LM (Overlay-VLAN):

- Edge Node: A second N-VDS (N-VDS2) is created for that service and associated to Edge Node vNIC4
- ESXi: A new VDS Port Group is created with load balance based on source mac address teaming policy and enabled with mac-learning
- ToR: VLANs Bridged are on all interfaces. ToR router IP of those VLANs Bridged are active/standby on ToR-Right/ToR-Left (to balance traffic with TEP and RTEP active/standby on ToR-Left/ToR-Right)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Edge Node VM on ESXi host with 4 pNICs:

Figure 4-92: NSX-T Federation Edge VM with ESXi 4 pNICS recommended configuration

This design is very similar to previous.
The change with this one is each Edge Node VM has its own Port Group connected to its own pNICs.

4.3.2.1.2 Recommended design for Edge Bare Metal

The designs below are based on the recommended design for Edge bare metal in the VMware NSX-T Reference Design Guide.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Edge Bare Metal with 2 pNICs:

Figure 4-93: NSX-T Federation Edge BareMetal with 2 pNICS recommended configuration

As documented in the [VMware NSX-T Reference Design Guide](), one N-VDS (N-VDS1) is created with 3 Uplink Teaming Policies.

- The "Default" teaming policy with 2 active uplinks (Load Balance Source) associated to pNIC1 + pNIC2
  It is used by TEP. Each Edge Node receives 1 TEP-IP for each uplink.
  It is also used by RTEP. Each Edge Node receives 1 RTEP-IP for the first uplink interface.
  Note: In a future release, each Edge Node will receive 1 RTEP-IP for each uplink.
- The "Left" teaming policy with 1 single active uplink associated to pNIC1
  It is used by one Segment-VLAN (light blue) connected to Tier-0 uplink.
- The "Right" teaming policy with 1 single active uplink associated to pNIC2
  It is used by one Segment-VLAN (dark blue) connected to Tier-0 uplink.

The ToR are configured with all the VLANs on all interfaces, but VLAN 300 and 400 (T0 uplinks) which are configured respectively only on interface to EdgeNode-P1 and EdgeNode-P2.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

ToR router IP of VLAN TEP and VLAN RTEP are active/standby on ToR-Left/ToR-Right, Segment VLAN Light Blue is active on ToR-Left, and Segment VLAN Dark Blue is active on ToR-Right.

And in case of bridging (Overlay-VLAN):

- Edge Node: the VLANs Bridged are all created on the “Right” teaming policy (associated to pNIC2)
- ToR: VLAN Bridged are on all interfaces. ToR router IP of those VLAN Bridged are active/standby on ToR-Right/ToR-Left (to balance traffic with TEP and RTEP active/standby on ToR-Left/ToR-Right)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Edge Bare Metal with 4 pNICs:

Figure 4-94: NSX-T Federation Edge BareMetal with 4 pNICS recommended configuration

This design is very similar to previous.
The change with this one is each Edge Node has its own pNICs for RTEP traffic. Each Edge Node receives 1 RTEP-IP for the first uplink interface.
Note: In a future release, each Edge Node will receive 1 RTEP-IP for each uplink and that's why RTEP is configured with the teaming "P3+P4".

And in case of bridging (Overlay-VLAN):

- Edge Node: Some of the VLANs Bridged are created on the "P1-Left" teaming policy (associated to pNIC1), and some are created on the "P2-Right" teaming policy (associated to pNIC2)
- ToR: VLAN Bridged are on all interfaces. ToR router IP of those VLAN Bridged are active/standby on ToR-Left/ToR-Right for the VLANs Bridged on the "P1-Left" teaming

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

policy, and ToR-Right /ToR-Left for the VLANs Bridged on the "P2-Right" teaming policy

4.3.2.2 Tier-0 options for most services and best performance

As presented in the chapter "4.2.1.3.1 GM Tier-0 and Tier-1 Gateway Configuration Options", Tier-0 can be deployed in different modes.

Figure 4-95: NSX-T Federation T0 topologies

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Tier-0 Active/Standby Location Primary/Secondary is the mode that offers the maximum of services on the Tier0. In this mode the Tier-0 can host central services, such as Gateway Firewalling, NAT, DHCP.

Tier-0 Active/Active Location Primary/Secondary is the mode that offers the best performance without asymmetric routing challenges.
In case of central services (Gateway Firewalling, NAT, DHCP), those can be deployed on Tier-1.

Tier-0 Active/Active Location All Primaries is the mode that offers the best performance but with possible asymmetric routing.

Figure 4-96: NSX-T Federation T0 A/A Loc_AllP asymmetric routing

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.4 Disaster Recovery

4.4.1 Management Plane

With NSX-T Federation the Management Plane is composed of GM and LM services.
In case of the loss of a location, the recovery of that LM location is not needed since it was only in charge of the Network and Security services for that location that has been lost.
However, if that location was also hosting the GM service, that one needs to be recovered.

4.4.1.1 GM Cluster Deployment Mode1: NSX-T GM-Active VMs deployed in 3 different locations

This mode offers automatic GM Service recovery against location failure.

<u>Use Case1: Stretched VLAN Management across locations</u>

Figure 4-97: NSX-T Federation Management Plane – Use-case metropolitan region with stretched Management VLAN

In case of failure of the location hosting the GM Cluster VIP Active, automatically another GM will take ownership of the GM Cluster VIP.

The GM Management Plane service outage will be around 1 minute.

During the GM service outage, the LM Management Plane is still operational in the different locations. So new workload VM can be deployed in those locations, as well as vMotion/DRS of

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

existing VMs is offered. DFW security will also be enforced on new VMs even when configured with dynamic group membership such as VM names or VM Tags.
What is not offered is the ability to modify GM network and security configuration. However, it's always possible to create on LM DFW Rules in the Emergency section which run before any GM DFW Rules.

During the GM service outage, the Data Plane is also operational in the different locations.
But Tier-0 / Tier-1 Primary in the failed location need to be moved to another location to recover the North/South traffic of their connected Segments. This is covered in the section 4.4.2 Data Plane.

Use Case2: No Stretched VLAN Management across locations

Figure 4-98: NSX-T Federation Management Plane – Use-case metropolitan region with stretched Management VLAN

In case of failure of the location of one of the GM VM, automatically the GSLB service will detect its failure and will stop resolving with its IP address.

The GM Management Plane service outage varies based on the TTL of the GM DNS entry. Assuming a TTL of 5 minutes, the outage will be 5 minutes.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

During the GM service outage, the LM Management Plane is still operational in the different locations. So new workload VM can be deployed in those locations, as well as vMotion/DRS of existing VMs is offered. DFW security will also be enforced on new VMs even when configured with dynamic group membership such as VM names or VM Tags.
What is not offered is the ability to modify GM network and security configuration. However, it's always possible to create on LM DFW Rules in the Emergency section which run before any GM DFW Rules.

During the GM service outage, the Data Plane is also operational in the different locations.
But Tier-0 / Tier-1 Primary in the failed location need to be moved to another location to recover the North/South traffic of their connected Segments. This is covered in the section 4.4.2 Data Plane.

4.4.1.2 GM Cluster Deployment Mode2: NSX-T GM-Active and GM-Standby

This mode offers manual/scripted GM Service recovery against location failure.

Before any location failure, GM-Active synchronizes all its network and security configuration to the GM-Standby. And each LM has a persistent connection to the GM-Active and GM-Standby, the GM-Active connection being the one used to receive any configuration update from GM.

Figure 4-99: NSX-T Federation GM service before location failure.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of a failure of the location hosting the GM service, the GM service is restored in another location activating the GM-Standby.
Once restored, each LM automatically learns the connection to the new GM-Active is the one to receive any configuration update from GM.

Figure 4-100: NSX-T Federation GM service after location failure hosting the GM-Active service + GM failover

The GM service failover will be around 5 minutes after the start of the failover.

During the GM service outage, the LM Management Plane is still operational in the different locations. So new workload VM can be deployed in those locations, as well as vMotion/DRS of existing VMs is offered. DFW security will also be enforced on new VMs even when configured with dynamic group membership such as VM names or VM Tags.
What is not offered is the ability to modify GM network and security configuration. However, it's always possible to create on LM DFW Rules in the Emergency section which run before any GM DFW Rules.

During the GM service outage, the Data Plane is also operational in the different locations.
But Tier-0 / Tier-1 Primary with services (GW-NAT or GW-FW) in the failed location need to be moved to another location to recover the North/South traffic of their connected Segments. This is covered in the section 4.4.2 Data Plane.

<u>Special Case: Split-Brain Scenario</u>
This special failure case covers the loss of cross-location communication from the location hosting the GM-Active (Location-A in the figure below), but its Internet communication is still working.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-101: NSX-T Federation GM service after cross-location failure hosting the GM-Active service

In the case of Location-A cross-location only failure:

- GM-Active to LM communication is only working with LM-A
- GM-Standby to LM communication is working with LM-B and LM-C
- LM to LM communication is working between the LMs other than LM-A
- Edge Nodes to Edge Nodes communication is working between the Edge Nodes other than Edge Nodes in Location-A

So the GM Management Plane is only operational with LM-A.
The LM Management Plane is still operational in all locations. So new workload VM can be deployed in those locations, as well as vMotion/DRS of existing VMs is offered.
DFW security will also be enforced on new VMs even when configured with dynamic group membership such as VM names or VM Tags. But LM-B / LM-C won't be able to update LM-A with their new Group members, and same for LM-A to LM-B / LM-C.

During this cross-location communication outage, the Data Plane is also operational in the different locations.
But cross-location traffic from or to Location-A will be blocked. This is covered in the section 4.4.2 Data Plane.

Note in case the cross-location communication is recovered:
The GM-Active and GM-Standby constantly try to reconnect. When connection happens, the GM-Active automatically does a full configuration sync to GM-Standby.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

At some point the NSX Admin can decide to make Active the GM-Standby in Location-B to restore GM service for LM-B and LM-C.

Figure 4-102: NSX-T Federation GM service after cross-location failure hosting the GM-Active service+ GM failover

LM-B and LM-C automatically recovers their GM Management Plane with the GM Location-B now Active. LM-A keeps its GM Management Plane with the GM Location-A.
From that point, it is strongly recommended to update GM configuration only from the new GM-Active: GM Location-B, as from that point GM configuration is no more synchronized between GM Location-A and GM Location-B. In case of configuration done on the GM Location-A, this configuration will never be synched back to GM Location-B when the cross-location failure is recovered.

Later in the case of recovery of the Location-A cross-location communication, GM Location-A will re-establish its communication with GM-Location-B. Both GM notice they are in Active mode, but since GM Location-B turned Active the most recently then GM Location-A will turn itself in "None" mode.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-103: NSX-T Federation GM service after cross-location recovery

GM Location-B remains in GM-Active mode, and GM Location-A turns automatically to GM-None mode.
LM-A learns GM Location-A is no more Active and the new GM-Active is in Location-B. So LM-A automatically connects to the GM-Active Location-B and get its GM configuration fully synchronized.

At last, from that point, GM Location-A can be manually added as GM-Standby to recover GM Service high availability.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-104: NSX-T Federation GM service after cross-location recovery+ recovery of GM Standby

From GM-Active (Location-B) GM Location-A is configured as Standby.
And GM-Active pushes a full synchronization to the new GM-Standby.

Note: In NSX-T 3.1.0, the action "Make GM Location-A GM-Standby" in API only.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.4.2 Data Plane

As presented in the chapter "4.2.1.3 L3 Routing Service" Federation Data Plane offers a mix of network topologies:

Figure 4-105: Different Tier0/Tier-1 topologies

The figure above represents the different Tier-0 and Tier-1 options and the traffic path for each.

The loss of one location does not inevitably generates a data plane failure.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Some topologies will offer automatic data plane recovery:

- Any Stretched Tier-0 with BGP and without stateful services (no NAT and no GW-FW) + Stretched Distributed Tier-1
- Stretched Tier-0 Active/Standby Location Primary/Secondary with Static Routes + Stretched Distributed Tier-1

Some topologies will require a manual data plane recovery:

- Stretched Tier-0 or Stretched Tier-1 with services (NAT / GW-FW)

4.4.2.1 Automatic Network Data Plane Recovery

This chapter details the network topologies with automatic plane recovery.

4.4.2.1.1 Any Stretched Tier-0 with BGP and without services (No NAT / No GW-FW) + Stretched Distributed Tier-1

There are two Tier-0 / Tier-1 stretched network topologies with BGP and without services (no NAT / No GW-FW) which offer automatic data plane recovery in case of any location loss.

T0 A/A Loc P/S with BGP + T1-Streched DR_Only:
This topology is with stretched Tier-0 Active/Active Location Primary/Secondary with BGP connected to a Tier-1 DR_Only.

Figure 4-106: T0 A/A Loc P/S with BGP + T1-Stretched DR_Only

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of a Tier-0 Secondary location, the data plane is always automatically recovered:

T0 A/A Loc P/S with BGP + T1 with No Service – Loss of T0 Secondary Location

Figure 4-107: T0 A/A Loc P/S with BGP + T1-Stretched DR_Only – Loss of T0 Secondary Location

The loss of Location-C (Tier-0 Secondary location) makes the whole Location-C down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

VMware, Inc. – www.vmware.com
Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of the Tier-0 Primary location, the data plane is also always automatically recovered:

Figure 4-108: T0 A/A Loc P/S with BGP + T1-Stretched DR_Only – Loss of T0 Primary Location

The loss of Location-A (Tier-0 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South now via Location-B and South/North exit locally.

Attention:
With 3 locations, as shown in the figure above, the North/South traffic will be asymmetric for the 3rd location. Any stateful service in the physical fabric (like physical firewall) would block North/South traffic from Location-C.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 Primary, the North/South data plane with Location-B and Location-C is down:

Figure 4-109: T0 A/A Loc P/S with BGP + T1-Stretched DR_Only – Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B and Location-C is also down.

It's possible to recover North/South traffic for Location-B and Location-C making the Tier-0 Location-B with the best BGP advertisements. However, that means North/South to Location-A will then be down.
This is done directly editing the Tier-0 configuration on GM-Active.

Note: The last failure does not offer automatic recovery of the Data Plane for Loc-B and Loc-C.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/A Loc All_P with BGP + T1-Streched DR_Only:
This topology is with stretched Tier-0 Active/Active Location All Primary with BGP connected to a Tier-1 DR_Only.

Figure 4-110: T0 A/A Loc All_P with BGP + T1-Stretched DR_Only

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of a location not best BGP advertisement, the data plane is always automatically recovered:

Figure 4-111: T0 A/A Loc All_P with BGP + T1-Stretched DR_Only – Loss of not best BGP advertisement

The loss of Location-C makes the whole Location-C down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A and asymmetric routing remains for traffic in Location-B.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of the location best BGP advertisement, the data plane is also always automatically recovered:

Figure 4-112: T0 A/A Loc All_P with BGP + T1-Stretched DR_Only – Loss of best BGP advertisement

The loss of Location-A makes the whole Location-A down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with now the North/South via Location-B and asymmetric routing remains for traffic in Location-C.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Finally, in the case of loss of cross-location communication from the location hosting the best BGP advertisement, the North/South data plane with Location-B and Location-C is down:

Figure 4-113: T0 A/A Loc All_P with BGP + T1-Stretched DR_Only – Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B and Location-C is also down.

It's possible to recover North/South traffic for Location-B and Location-C making the Tier-0 Location-B with the best BGP advertisements. However, that means North/South to Location-A will then be down.
This is done directly editing the Tier-0 configuration on GM-Active.

Note: The last failure does not offer automatic recovery of the Data Plane for Loc-B and Loc-C.

4.4.2.1.2 Stretched Tier-0 Active/Standby Location Primary/Secondary with Static Routes + Stretched Distributed Tier-1

This Tier-0 / Tier-1 stretched network topologies with static routes services offers automatic data plane recovery in case of any location loss.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/S Loc P/S with static routes + T1-Streched DR_Only (+ Optionally NAT):
This topology is with stretched Tier-0 Active/Standby Location Primary/Secondary with static routes connected to a Tier-1 DR_Only.
There is also below NAT configured on the T0 because the Segment is with a private subnet. NAT would not be required if the Segment were with a public subnet.

Physical advertisement:

- Location-A
  - SNAT (21.0.0.0/24) and DNAT (31.0.0.0/24) subnets
- Location-B
  - SNAT (21.0.0.0/24) and DNAT (31.0.0.0/24) subnets with cost +

Physical static route:

- ToR-LocA:
  - SNAT (21.0.0.0/24) and DNAT (31.0.0.0/24) subnets via T0-LocA-HA_VIP
- ToR-LocB:
  - SNAT (21.0.0.0/24) and DNAT (31.0.0.0/24) subnets via T0-LocB-HA_VIP

Tier-0 is configured with:

- Static routes:
  - T0 Primary
    - 0.0.0.0/0 via ToR-LocA (only for LM LocA + do not enable on secondary)
  - T0 Secondary
    - 0.0.0.0/0 via ToR-Loc2 (only for LM LocB + enable on secondary)

    Note: T0 Secondary FIB will choose the route 0.0.0.0/0 received from T0 Primary over its static route.

- NAT
  - Both T0 Primary and Secondary
    - SNAT for internal subnets to External
      10.1.1.0/24 to 21.0.0.6
    - DNAT for External to internal subnets
      31.0.0.11 to 10.1.1.11
      31.0.0.12 to 10.1.1.12
      31.0.0.13 to 10.1.1.13
      31.0.0.14 to 10.1.1.14

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-114: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only

In case of loss of a Tier-0 Secondary location, the data plane is always automatically recovered:

Figure 4-115: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of T0 Secondary Location

The loss of Location-B (Tier-0 Secondary location) makes the whole Location-B down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of the Tier-0 Primary location, the data plane is also always automatically recovered:

T0 A/S Loc P/S with static routes and NAT + T1 with No Service – Loss of T0 Primary Location

Figure 4-116: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of T0 Primary Location

The loss of Location-A (Tier-0 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
The Tier-0 Secondary (T0-LocB) does not receive its default gateway via Tier-0 Primary (T0-LocA) and so its static route is now pushed to its FIB.
So the Data Plane is still working though, with the North/South now via Location-B and South/North exit locally.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 Primary, the North/South data plane with Location-B is down:

T0 A/S Loc P/S with static routes and NAT + T1 with No Service – Loss of cross-location communication

Figure 4-117: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B is also down.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

It's possible to recover North/South traffic for Location-B making better route advertisements from Location-B. However, that means North/South to Location-A will then be down.
This is done directly editing the route advertisements configuration on physical routers.

4.4.2.1.3 Stretched Tier-0 Active/Standby Location Primary/Secondary with Static Routes + Stretched Distributed Tier-1 with Local-Egress

This other Tier-0 / Tier-1 stretched network topologies with static routes services offers automatic data plane recovery in case of any location loss; but this time with Local-Egress and no asymmetric routing with NAT configuration.

T0 A/S Loc P/S with static routes + T1-Streched DR_Only + NAT:
This topology is with stretched Tier-0 Active/Standby Location Primary/Secondary with static routes connected to a Tier-1 DR_Only.
In this topology NAT is also required for symmetric routing with different SNAT subnets per location.

Physical advertisement:
- Location-A
  - SNAT (21.0.0.0/24) and DNAT (31.0.0.0/24) subnets
- Location-B
  - SNAT (22.0.0.0/24) and DNAT (31.0.0.0/24) subnets with cost +

Physical static route:
- ToR-LocA:
  - SNAT (21.0.0.0/24) and DNAT (31.0.0.0/24) subnets via T0-LocA-HA_VIP
- ToR-LocB:
  - SNAT (22.0.0.0/24) and DNAT (31.0.0.0/24) subnets via T0-LocB-HA_VIP

Tier-0 is configured with:
- Static routes:
  - T0 Primary
    - 0.0.0.0/0 via ToR-LocA (only for LM LocA + do not enable on secondary)
    - 0.0.0.0/0 via ToR-LocA (only for LM LocA + do not enable on secondary)
  - T0 Secondary
    - 0.0.0.0/0 via ToR-Loc2 (only for LM LocB + enable on secondary)

    Note: T0 Secondary FIB will choose the route 0.0.0.0/0 received from T0 Primary over its static route.

- NAT
  - Both T0 Primary and Secondary
    - SNAT for internal subnets to External
      10.1.1.0/24 to 21.0.0.6

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

- DNAT for External to internal subnets
  31.0.0.11 to 10.1.1.11
  31.0.0.12 to 10.1.1.12
  31.0.0.13 to 10.1.1.13
  31.0.0.14 to 10.1.1.14

Figure 4-118: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only

South/North traffic is symmetric thanks to specific SNAT subnets in each location:

Figure 4-119: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only – Traffic initiated from Internal to External

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

North/South traffic is single location with the same DNAT subnet in each location, but advertised better from Location-A:

Figure 4-120: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only – Traffic initiated from External to Internal

In case of loss of a Tier-0 Secondary location, the data plane is always automatically recovered:

Figure 4-121: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of T0 Secondary Location

The loss of Location-B (Tier-0 Secondary location) makes the whole Location-B down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of the Tier-0 Primary location, the data plane is also always automatically recovered:

T0 A/S Loc P/S with static routes and NAT + T1 with No Service – Loss of T0 Primary Location

Figure 4-122: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of T0 Primary Location

The loss of Location-A (Tier-0 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
The Tier-0 Secondary (T0-LocB) does not receive its default gateway via Tier-0 Primary (T0-LocA) and so its static route is now pushed to its FIB.
So the Data Plane is still working though, with the North/South now via Location-B and South/North exit locally.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 Primary, the only impact is for North/South traffic to Location-B and the East/West traffic:

T0 A/S Loc P/S with static routes and NAT + T1 with No Service – Loss of cross-location communication – Traffic initiated from Internal to External

Figure 4-123: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of cross-location communication – Traffic initiated from Internal to External

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-124: T0 A/S Loc P/S with static routes + T1-Stretched DR_Only - Loss of cross-location communication – Traffic initiated from External to Internal

4.4.2.2 Manual Network Data Plane Recovery

This chapter details the network topologies with manual plane recovery.

4.4.2.2.1 Tier-0 or Tier-1 stretched with services (NAT / GW-FW)

There are four Tier-0 / Tier-1 stretched network topologies with services (NAT / GW-FW). Those which offer automatic data plane recovery in case of any location loss.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/S Loc P/S with BGP + T1-Streched DR_Only:
This topology is with stretched Tier-0 Active/Standby Location Primary/Secondary connected to a Tier-1 DR_Only.

Figure 4-125: T0 A/S Loc P/S with BGP + T1-Stretched DR_Only

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of a Tier-0 Secondary location, the data plane is always automatically recovered:

Figure 4-126: T0 A/S Loc P/S with BGP + T1-Stretched DR_Only – Loss of T0 Secondary Location

The loss of Location-C (Tier-0 Secondary location) makes the whole Location-C down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

However, in case of loss of the Tier-0 Primary location, the data plane requires manually recovery:

Figure 4-127: T0 A/S Loc P/S with BGP + T1-Stretched DR_Only – Loss of T0 Primary Location

The loss of Location-A (Tier-0 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
Also there is no more advertisement on the internal subnet to the physical fabric, so North/South traffic is broken. East/West intra-location or cross-location is still working though.

To recover North/South traffic, the Tier-0 Primary location configuration must be changed from Location-A to Location-B or Location-C.
This can be done directly editing the Tier-0 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 Primary, the North/South data plane with Location-B and Location-C is down:

Figure 4-128: T0 A/S Loc P/S with BGP + T1-Stretched DR_Only – Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B and Location-C is also down.

It's possible to recover North/South traffic for Location-B and Location-C making the Tier-0 Primary location configuration from Location-A to Location-B or Location-C. However, that means North/South to Location-A will then be down.
This can be done directly editing the Tier-0 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/S Loc P/S with BGP + T1 A/S Loc P/S:
This topology is with stretched Tier-0 Active/Standby Location Primary/Secondary connected to a Tier-1 Active/Standby Location Primary/Secondary.

Figure 4-129: T0 A/S Loc P/S with BGP + T1 A/S Loc P/S

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of a Tier-0 and Tier-1 Secondary location, the data plane is always automatically recovered:

T0 A/S Loc P/S with BGP + T1 with Service – Loss of T0/T1 Secondary Location

Figure 4-130: T0 A/S Loc P/S with BGP + T1 A/S Loc P/S – Loss of T0/T1 Secondary Location

The loss of Location-C (Tier-0 and Tier-1 Secondary location) makes the whole Location-C down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

However, in case of loss of the Tier-0 and Tier-1 Primary location, the data plane requires manually recovery:

Figure 4-131: T0 A/S Loc P/S with BGP + T1 A/S Loc P/S – Loss of T0/T1 Primary Location

The loss of Location-A (Tier-0 and Tier-1 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
Also, there is no more advertisement of the internal subnet to the physical fabric, so North/South traffic is broken. East/West intra-location or cross-location is still working though.

To recover North/South traffic, the Tier-0 and Tier-1 Primary location configuration must be changed from Location-A to Location-B or Location-C.
This can be done directly editing the Tier-0 and Tier-1 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 and Tier-1 Primary, the North/South data plane with Location-B and Location-C is down:

Figure 4-132: T0 A/S Loc P/S with BGP + T1 A/S Loc P/S – Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B and Location-C is also down.

It's possible to recover North/South traffic for Location-B and Location-C making the Tier-0 and Tier-1 Primary location configuration from Location-A to Location-B or Location-C. However, that means North/South to Location-A will then be down.
This can be done directly editing the Tier-0 and Tier-1 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/A Loc P/S with BGP + T1 A/S Loc P/S:
This topology is with stretched Tier-0 Active/Active Location Primary/Secondary connected to a Tier-1 Active/Standby Location Primary/Secondary.

Figure 4-133: T0 A/A Loc P/S with BGP + T1 A/S Loc P/S

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of a Tier-0 and Tier-1 Secondary location, the data plane is always automatically recovered:

T0 A/A Loc P/S with BGP + T1 with Service – Loss of T0/T1 Secondary Location

Figure 4-134: T0 A/A Loc P/S with BGP + T1 A/S Loc P/S – Loss of T0/T1 Secondary Location

The loss of Location-C (Tier-0 and Tier-1 Secondary location) makes the whole Location-C down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

However, in case of loss of the Tier-0 and Tier-1 Primary location, the data plane requires manually recovery:

Figure 4-135: T0 A/A Loc P/S with BGP + T1 A/S Loc P/S – Loss of T0/T1 Primary Location

The loss of Location-A (Tier-0 and Tier-1 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
The best advertisement of the internal subnet is now via Location-B. The Tier-0 in Location-B will receive it but then it will forward it to the Tier-1 Primary location (Location-A) breaking North/South traffic. East/West intra-location or cross-location is still working though.

To recover North/South traffic, the Tier-0 and Tier-1 Primary location configuration must be changed from Location-A to Location-B or Location-C.
This can be done directly editing the Tier-0 and Tier-1 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 and Tier-1 Primary, the North/South data plane with Location-B and Location-C is down:

Figure 4-136: T0 A/A Loc P/S with BGP + T1 A/S Loc P/S – Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B and Location-C is also down.

It's possible to recover North/South traffic for Location-B and Location-C making the Tier-0 and Tier-1 Primary location configuration from Location-A to Location-B or Location-C. However, that means North/South to Location-A will then be down.
This can be done directly editing the Tier-0 and Tier-1 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

T0 A/A Loc All_P with BGP + T1 A/S Loc P/S:
This topology is with stretched Tier-0 Active/Active Location All Primary connected to a Tier-1 Active/Standby Location Primary/Secondary.

Figure 4-137: T0 A/A Loc All_P with BGP + T1 A/S Loc P/S

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In case of loss of a Tier-1 Secondary location, the data plane is always automatically recovered:

Figure 4-138: T0 A/A Loc All_P with BGP + T1 A/S Loc P/S – Loss of T1 Secondary Location

The loss of Location-C (Tier-1 Secondary location) makes the whole Location-C down: its Compute, its Network and Security, and its LM.
The Data Plane is still working though, with the North/South still via Location-A.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

However, in case of loss of the Tier-1 Primary location, the data plane requires manually recovery:

Figure 4-139: T0 A/A Loc All_P with BGP + T1 A/S Loc P/S – Loss of T0/T1 Primary Location

The loss of Location-A (Tier-1 Primary location) makes the whole Location-A down: its Compute, its Network and Security, and its LM.
The subnets behind the Tier-1 are no more advertised breaking North/South traffic. East/West intra-location or cross-location is still working though.

To recover North/South traffic, the Tier-1 Primary location configuration must be changed from Location-A to Location-B or Location-C.
This can be done directly editing the Tier-1 configuration on GM-Active, or with the Network Recovery process detailed in chapter "4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Finally, in the case of loss of cross-location communication from the location hosting the Tier-0 and Tier-1 Primary, the North/South data plane with Location-B and Location-C is down:

Figure 4-140: T0 A/A Loc All_P with BGP + T1 A/S Loc P/S – Loss of cross-location communication

The loss of cross-location communication makes all East/West communication from Location-A down and so North/South traffic for Location-B and Location-C is also down.

It's possible to recover North/South traffic for Location-B and Location-C making the Tier-0 Location-B with the best BGP advertisements and Tier-1 Primary location configuration from Location-A to Location-B or Location-C. However, that means North/South to Location-A will then be down.
This is done directly editing the Tier-0 and Tier-1 configuration on GM-Active.

4.4.2.2.2 Fully Orchestrated Network Data Plane Network Recovery

As presented in the previous chapter 4.4.2.2 Manual Network Data Plane Recovery, the recovery of each Tier-0 / Tier-1 can be done individually via a configuration change on GM.
But, when GM detects the loss of the LM, GM also offers the ability to fully move all the Tier-0 / Tier-1 primary of that location to another location.

Important Note:
This GM fully orchestrated move of all the Tier-0 / Tier-1 primary of that location, is a manual GM NSX Admin action (UI or API). It is strongly recommended to validate that location is really

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

down before running it. Indeed, the data plane of that location could be fine but only the GM-LM communication went down.

Figure 4-141: NSX-T Federation Fully Orchestrated Data Plane Recovery

Each Tier-0 is configured with a fallback preference. For instance, the T0_A can be configured with a fallback preference of Priority 1 for Location-B and Priority 2 for Location-C.
Using the GM orchestrated data plane recovery, all Tier-0 primary in a failed location will be moved automatically to their Priority 1 location. Then by default all the Tier-1 primary in that failed location will follow the same location as their connected primary Tier-0.

It's worth highlighting prior to executing the orchestrated recovery, GM offers a review of each Tier-0 / Tier-1 new primary location; as well as the ability to update their new location if wanted.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-142: NSX-T Federation Fully Orchestrated Data Plane Recovery – Manual Overwrite of New Active Location

4.4.2.3 Compute VMs Recovery

In case of location failure, all compute VMs hosted in that location are also lost.
VMware offers a solution to replicate compute to another location and recover them in case of location failure: VMware Site Recovery Manager (SRM).

In the figure below, a Tier-0 and Tier-1 are stretched between Location-B and Location-C. The compute VMs and routing services are primary in Location-B.

Figure 4-143: Compute VMs before location failure

SRM is configured to replicate the compute VMs from Location-B to Location-C.

As discussed in the section "4.4.2.1 Automatic Network Data Plane Recovery", in case of failure of Location-B the Tier-0 primary is moved from Location-B to Location-C.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-144: Compute VMs after location failure

And SRM recovers the compute VMs in Location-C.
Those VMs now running in Location-C are the very same, and so also have the same IP addresses, default gateway and DFW Rules.

Important Note:
Recovered VMs via SRM to a new location lose their NSX VM Tags in the new LM. Also, recovered VMs will receive new Segment Ports on the new LM.
So, if the Federation Security is based on VM Tags, or Segment Ports, or Segment Ports Tags; then the recovered compute VMs in Location-C do not have their DFW Rules.

In the figure below, you can see the VM replicated by SRM from Location-B to Location-C. In LM-C the inventory shows the VM is powered off, unplugged, with different BIOS ID / Attachment ID, and with no TAG.
For your information, the BIOS ID is viewable from LM under “Inventory / Virtual Machines”, and the Attachment ID is viewable from GM or LM under “Networking / Segment / Ports”.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-145: VM inventory information in different LMs prior to SRM recovery without GM Sync VM Tag configuration

Now in case of DR, SRM powers on the replicated VMs and plug them to the appropriate segments. SRM also updates the IDs (BIOS ID, Instance ID, External ID, and Attachment ID) of the replicated started VMs to match the original ones. But the LM on the recovered site is still not aware of any Tags for those VMs, as you can see in the figure below:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-146: VM inventory information in different LMs after SRM recovery without GM Sync VM Tag configuration

That's why **prior to NSX-T 3.2, SRM is not supported with Federation with VM Tags, or Segment Ports, or Segment Ports Tags**.

However, since NSX-T 3.2, NSX offers the ability to synchronize the VM Tags across LMs (not Segment Ports MP ID nor Segment Ports Tags).
This is done via GM VM Tag Replication Policy (currently API only):

GM API

```
PATCH https://<GM>/global-manager/api/v1/global-infra/vm-tag-
replication-policies/<name>
      {
        "display_name": "<name>",
        "protected_site": "<site>",
        "recovery_sites": [
            "<site>"
          ],
        "vm_match_criteria":       "[MATCH_NSX_ATTACHMENT_ID       |
      MATCH_BIOS_UUID_NAME]*",
        "groups": [
            "<group_id>",
            "<group_id>"
          ]
```


Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

```
        }
```

*: MATCH_NSX_ATTACHMENT_ID or MATCH_BIOS_UUID_NAME can be used for replicating the VM Tags via SRM recovery. Other recovery tool may work with one or the other. This will have to be validated by the recovery tool vendor.

Example:

```
PATCH https://gm-paris.corp.com/global-manager/api/v1/global-
infra/vm-tag-replication-policies/policyparislondon
        {
          "display_name": "policyparislondon",
          "protected_site": "/global-infra/sites/LM_Paris",
          "recovery_sites": [
        d       "/global-infra/sites/LM_London"
          ],
          "vm_match_criteria": "MATCH_BIOS_UUID_NAME",
          "groups": [
              "/global-infra/domains/default/groups/gr1",
              "/global-infra/domains/default/groups/gr2"
          ]
        }
```

This GM VM Tag Replication Policy copies the VM information for the VMs in those groups from "protected site LM" to "recovery site LM".
In the figure below, you can see the VM inventory information is copied from LM Location-B to LM Location-C internal database. However, that information is not viewable from LM Location-C UI nor API, as LM lists VM entries only for the VMs known by vCenter Location-C.

Figure 4-147: VM inventory information in different LMs with GM Sync VM Tag configuration

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Now in case of DR, SRM powers on the replicated VMs and plug them to the appropriate segment. SRM also updates the IDs of the replicated started VMs to match the original ones. LM Location-C match those VMs IDs with the entries in its internal database and those entries have the Tag information. So now those powered on VMs have the Tag information in LM Location-C, as you can see in the figure below:

Figure 4-148: VM inventory information in different LMs after SRM recovery with GM Sync VM Tag configuration

Conclusion **in NSX-T 3.1 SRM is not currently supported with Federation based on VM Tags, or Segment Ports, or Segment Ports Tags**.
From **NSX-T 3.2 SRM is supported Federation based on VM Tags** (requires specific GM-API configuration). **But not supported with Federation with Segment Ports, or Segment Ports Tags** (it would require Segment Ports MP ID and Segment Ports Tags to be synchronized between LM).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.4.2.4 Load Balancing Data Plane Recovery

As described in the chapter 4.2.1.6 Load Balancing service (Avi), 2 Disaster Recovery options are possible:
In case of Disaster Recovery need, 2 options are possible:

- GSLB
  The same application runs in different locations behind different VIPs.
  This option is detailed in chapter 4.4.2.4.1 LB Disaster Recovery with GSLB.
- Non-GSLB
  The same application runs in different locations behind the same VIP.
  This option is detailed in the chapter 4.4.2.4.2 LB Disaster Recovery without GSLB.

4.4.2.4.1 LB Disaster Recovery with GSLB

GSLB is a popular option for Disaster Recovery.
This option is for the incoming traffic only of the data plane (traffic from External to Inside) and is based on DNS. It does not cover the Management Plane recovery.

Figure 4-149: Federation with GSLB option - Before location failure

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The same application is deployed in in different locations. In the figure above, the application in Location-A is reachable via VIP1, in Location-B via VIP2, and in Location-C via VIP3.
Those applications are completely isolated and don't need any cross-location communication.
Users access that application via its FQDN (app1.xyz.com). The DNS in charge of resolving that FQDN is a GSLB solution. That GSLB solution is configured with all location IP: VIP1 + VIP2 + VIP3 and continuously validates the application is running in each location.
The GSLB solution can be configured to resolve the FQDN with only one IP (active/standby) or multiple IP (active/active). In the figure above, the GSLB solution will resolve app1.xyz.com with VIP1 only if configured in active/standby or resolve app1.xyz.com with VIP1+VIP2+VIP3 if configured in active/active.

Figure 4-150: Federation with GSLB option - After location failure

After the loss of a Location-A, the GSLB solution will detect its failure and stop using it for its FQDN resolution.
In the figure above, now the GSLB solution will resolve app1.xyz.com with VIP2 only if configured in active/standby or resolve app1.xyz.com with VIP2+VIP3 if configured in active/active.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The Data Plane service outage varies based on the GSLB location healthcheck interval, and the Time-To-Live (TTL) of the FQDN entry. It is usually around 5 minutes.

More information on VMware GSLB solution Avi on https://avinetworks.com/docs/18.2/avi-gslb-overview/.

4.4.2.4.2 LB Disaster Recovery without GSLB

As presented in the chapter 4.2.1.6 Load Balancing service (Avi), the load balancing service can be added in a Federation deployment with VMware NSX Advanced Load Balancer (Avi).

Figure 4-151: Federation without GSLB option - Before location failure

Disaster Recovery can be offered without GSLB.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

In that case, the same application runs in different locations behind the same VIP, as represented in the figure above with the VIP1 configured in Location-A and Location-B.

In the Non-GSLB option, the VIP is active only in the primary location (VIP1 in Location-A in the figure above) with the Virtual Service configured with Traffic Enabled, and other locations have the Virtual Service configured with Traffic Disabled (VIP1 in Location-B in the figure above).

Once the Virtual Service is enabled, the Avi Controller automatically configures the LM_T1 with a Static Route "VIP next hop = Avi-SE(s)".
Note: If the Avi-SE are deployed in Active/Standby; then only one next-hop is configured = Avi-SE Active. If the Avi-SE are deployed in Active/Active or N+M; then multiple next-hops are configured = Avi-SEs Active.
Then the LM_T1 redistribute its static route to the GM_T0, which redistributes it to the physical fabric in that location.
So traffic to the VIP enters via the GM_T0 (T0-Stretched-Slice-LocationA for VIP1) to the LM_T1 to the Avi-SE(s).
At last the load balancer distributes the traffic to the pool members (local and/or remote).

In case of loss of the primary location (Location-A in the figure above), all services are down: Compute, Network and Security, Avi-SEs load balancers.
Also there is no more advertisement of the VIP1 to the physical fabric and access to VIP1 is down:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-152: Federation without GSLB option - After location failure

To recover the VIP1, the load balancing data plane requires manually recovery:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 4-153: Federation with Advanced Load Balancing (Avi) – Disaster Recovery

The Virtual Service is enabled on the secondary location (Location-B in the figure above).
The Avi Controller Location-B automatically configures the LM Location-B T1 with a Static Route "VIP next hop = Avi-SE(s)".

Then the LM_T1 redistribute its static route to the GM_T0, which redistributes it to the physical fabric in that location.

So traffic to the VIP enters via the secondary location (T0-Slice-LocationB in the figure above) to the LM_T1 to the Avi-SE(s).

At last the load balancer distributes the traffic to the pool members (local and/or remote).

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.5 Requirements and Limitations

The different requirements and limitations of the NSX-T Federation solution have been detailed in the different chapters above. This chapter summarize them all.

NSX-T Federation requirements:

- Management
  - NSX-T Release
    - N+/-1 release (https://docs.vmware.com/en/VMware-NSX/4.0/upgrade/GUID-B828A8FC-DD9F-486E-A616-9531D9F2E9C1.html)
    - License Enterprise+ on each LM
  - LM Cluster VIP configuration required
    - Currently the LM Cluster VIP (or FQDN LM Cluster VIP) must be provided to allow the GM to LM communication to keep on even after one LM Management VM failure

- WAN
  - Maximum 500 milliseconds latency (RTT) between locations or 150 milliseconds latency between locations if stretched T0/T1/Segments.
  - Bandwidth large enough to accommodate cross location Management Plane + Data Plane
    - The Management Plane traffic is minimal with few Mbps at peak
    - The Data Plane traffic varies greatly between customers
    - In case of possible congestion cross location, it is recommended to configure QoS to prioritize NSX Management Plane traffic: GM-GM traffic, GM-LM traffic and NSX Data Plane traffic: BGP and BFD traffic encapsulated in Geneve Overlay tunnels exchanged between the RTEP on the Edge Nodes, those packets could still be classified based on DSCP value CS6.
  - MTU at least 1500 bytes
    - Recommended 9000
  - IP connectivity
    - GM-LM and LM-LM: Connectivity without NAT + Allow Management traffic
    - Edge Nodes cross-locations: Connectivity without NAT + Allow Data Plane (RTEP traffic)
  - Public IP@ (Segments, NAT, Load Balancer VIP) must be advertisable from both locations
    - In case of different Internet Providers (Verizon in Site-A and Orange in Site-B), both will advertise the public IP@ when then turn Active. In such case, be sure the public IP@ belong to the customer and not the Internet Provider.

NSX-T Federation limitations:

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

- Management
  - GM-Active does not synchronize vIDM/LDAP configuration to GM-Standby
  - Port Mirroring
  - IPFIX
  - Live Traffic Analysis
  - Consolidated Capacity
  - No Multi-Tenancy / Projects support

- Networking
  - All Networking features are supported from GM, but
    - DHCP dynamic binding
    - Routing VRF and EVPN
    - Layer4+ services NSX-T Load Balancing and VPN
    - OSPF
    - Multicast
    - Tier-0 and Tier-1 Active/Active with stateful services
  - Federation + LB Service is supported with VMware NSX Advanced Load Balancer (Avi) – see chapter 4.2.1.6 Load Balancing service (Avi)
  - List of supported LM Network features configured from LM once registered by GM in the chapter 4.1.1.5.1 Logical Configuration Ownership
  - No Tier-0/Tier-1 Automatic DR
    - Network DR requires GM Tier-0/Tier-1 primary location configuration change

- Security
  - Stretched Groups based on Segment Ports or Segment Ports Tag do not support VM cold vMotion / SRM across locations
  - All Security features are supported from GM, but
    - URL Filtering
    - Identity Firewall
    - Distributed IDS
    - Gateway IDS/IPS
    - Malware Prevention
    - Network Detection and Response
    - Network Introspection
    - Endpoint Protection
    - Distributed Security for vCenter VDS Port Group (using GM dynamic group membership based on LM VDS Port Group Tags)
    - TLS inspection
  - List of supported LM Security features configured from LM once registered by GM in the chapter 4.1.1.5.1 Logical Configuration Ownership

- Workload
  - Supported
    - Virtual Machines on ESXi
    - Physical Servers NSX prepared

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

- Not Supported
  - Containers (Containers have no concept of Locations. Network and Security services for Containers can be offered directly by LM; even with LM registered to GM)

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.6 Orchestration / Eco-System

NSX-T Federation solution is based on NSX-T Data Center but still NSX-T GM API is slightly different than NSX-T LM API. See section "4.1.1.6 Federation API" for more information.
So all orchestration tools (NSX Intelligence, vRA, Terraform, Ansible, etc) and 3rd party solution (Skybox, Tufin, etc) requires an NSX-T Federation plugin.

Currently NSX-T Federation is supported with:

- **Terraform**
  See https://registry.terraform.io/providers/vmware/nsxt/latest/docs/guides/federation for more information.
  Examples of Terraform on:
  https://github.com/vmware-samples/nsx-t/tree/master/helper-scripts/Multi-Location/Federation/End2End

- **PowerCLI**
  Examples of PowerCLI on:
  https://github.com/vmware-samples/nsx-t/tree/master/helper-scripts/Multi-Location/Federation/End2End

- **vRA** (from vRA 8.5)
  vRA can consume Existing Network, Existing Security Group, Tagging:
  https://blogs.vmware.com/networkvirtualization/2022/01/vmware-network-automation-with-nsx-t-3-2-and-vrealize-automation.html/

- **vRNI** (from vRNI 6.4)
  vRNI offers visibility into global VMware NSX-T entities and cross-site VM-VM path:
  https://docs.vmware.com/en/VMware-vRealize-Network-Insight/6.4/rn/vrealize-network-insight-64-release-notes.html

- **NSX Intelligence** (from NSX Intelligence 3.2)
  NSX Intelligence offers visibility into flows going through your Compute VMs. It can also assist you with planning micro-segmentation by making firewall rule recommendation.
  Note: Each LM will have its own NAPP + NSX Intelligence deployment. Each NSX Intelligence will have visibility of its LM Group and also GM Groups. However, the visualization will not reflect specifics from other locations. NSX Intelligence recommendations will also not function across locations sites because NSX Intelligence does not integrate with the Global Manager of NSX Data Center.
  https://docs.vmware.com/en/VMware-NSX-Intelligence/4.0.1/rn/vmware-nsx-intelligence-401-release-notes/index.html

And the following NSX-T Federation orchestration tools are in beta:

- **Ansible**
  Beta code available on: https://github.com/vmware/ansible-for-nsxt

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Other orchestration tools and 3rd party tools (NSX Intelligence, Skybox, Tufin, etc) are not supported on GM nor LM once registered to GM. Please contact their sales representative for their latest road-maps.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

4.7 Scale and Performance guidance

Federation scale is detailed under https://configmax.vmware.com/.

NSX 4.1.0 Configuration Limits

EXPORT ALL LIMITS TO PDF | COLLAPSE ALL

	Limits	Description
Federation (General)		
General Locations	8	
General Hypervisor Hosts Across all Locations	1,024	
General Network Latency between Global Manager Active Cluster and Global Manager Standby Cluster	500ms	Round-trip time
General Network Latency between Global Manager Active Cluster and Local Manager Cluster	500ms	Round-trip time
General Network Latency between Local Manager Clusters across Different Locations	500ms	Round-trip time
General Network Latency between Remote TEPs across Different Locations	500ms	Round-trip time
General Physical Servers	500	Non-hypervisor and non-container host machines with at least 16Gb of RAM. Windows Servers can have a maximum of 100 firewall rules each.
Federation (Networking)		
Federation (Layer 2)		
Federation (Layer 3)		
Federation (DHCP)		
Federation (Grouping and Tagging)		
Federation (Global Firewall)		
Federation (Distributed Firewall)		
Federation (Gateway Firewall)		

Figure 4-154: NSX-T 4.1.0 configuration limits

Performance:

All performance considerations detailed in the complete VMware NSX-T Reference Design Guide apply for the Federation solution too.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

5 Migration to Multi-Locations

5.1 From "Single Location" To "NSX-T Multisite"

NSX-T Multisite can be implemented on Green Field (new NSX-T deployment for all locations), or Brown Field (addition of NSX-T in a new location).

In the case of Brown Field, there is one single NSX-T Manager Cluster with one single location prepared:

Figure 5-1: NSX-T Brown Field Deployment Prior to NSX-T Multisite

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The addition of a second location is done in 1 step and 1 optional step:

Figure 5-2: NSX-T Brown Field Deployment After NSX-T Multisite

First, prepare the hypervisors and deploy Edge Nodes in the second location. From that point, all the Segments are stretched to the second location.

Second, move the Tier-0 and Tier-1 standby to the second location. This step is optional and is detailed in the chapter "3.3.2 Data Plane".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

5.2 From "Single Location" To "NSX-T Federation"

NSX-T Federation can be implemented on Green Field (new NSX-T deployment for all locations), or Brown Field (addition of Federation into existing NSX-T deployments).

In the case of Brown Field, there is one independent NSX-T Manager Cluster (LM) for each location:

Figure 5-3: NSX-T Brown Field Deployment Prior to NSX-T Federation

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The addition of Federation is done in 2 steps and 1 optional step:

Figure 5-4: NSX-T Brown Field Deployment After NSX-T Federation

First, deploy the Global Manager Cluster (GM). The Global Manager Cluster can be installed in any location, or it can be stretched across multiple locations as detailed in the chapter "4.1.1.1 GM Cluster Deployments". The only requirement is GM to LM and LM to LM connectivity as detailed in the chapter "4.1.1.2 GM, LM, Edge Node Communication Flows".

Second, register each Local Manager Cluster (LM) and config RTEP on their Edge Nodes. Each LM must be running at least NSX-T released 3.1.0. This step is detailed in the chapter "4.1.1.3.1 LM Registration".

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Third, onboard LM existing Network and Security configuration to GM. This step is optional and is detailed in the chapter “4.1.1.3.2 (Optional) LM Onboarding”.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

5.3 From "NSX-T Multisite" To "NSX-T Federation"

The transition from NSX-T Multisite to NSX-T Federation is complex, but still can be accomplished.

With NSX-T Multisite, both locations transport nodes (hypervisors and Edge Nodes) are registered to Location-A NSX- T Manager.

Figure 5-5: NSX-T Multisite Prior to NSX-T Federation

Registered transport nodes cannot change NSX-T Managers on the fly.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Figure 5-6: Location-B Transport Nodes can't Register to NSX-T Manager Location-B

The transport nodes move from one NSX-T Manager is a multiple step process, then the addition of NSX-T Federation can be done.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The first step is to remove NSX-T from Location-B. For that, all Location-B Tier-0, Tier-1, and compute VMs to Location-A (step1a), and NSX-T can be uninstalled from Location-B (step 1b).

Figure 5-7: Remove NSX-T from Location-B

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The second step is to register Location-B transport nodes to NSX-T Manager B (step 2).

Figure 5-8: Register Location-B Transport Nodes to NSX-T Manager B

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The third step is to deploy GM. For that, one Global Manager Cluster is deployed in one location (step 3a), each Location Local Manager is registered (step 3b), and LM configuration is onboarded (step 3c). For more information, see chapter "4.1.1 Management Plane".

Figure 5-9: Add NSX-T Federation

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

The fourth and final step is to stretch Network (step 4a) and Security (step 4b) to Location-B.

Figure 5-10: Stretch Network and Security to Location-B

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

5.4 Mixing "NSX-T Multisite" and "NSX-T Federation"

NSX-T Multisite and NSX-T Federation can be mixed.

A typical example would be multiple metropolitan datacenters.
NSX-T Multisite is used within metropolitan datacenters, and NSX-T Federation is used across metropolitan datacenters.

Figure 5-11: Mix of NSX-T Multisite and NSX-T Federation

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

6 Federation Support within VCF + VVD

VCF supports NSX-T Multi-Site with its design described in the VMworld Session HC2519 (https://www.vmworld.com/en/video-library/search.html#text=%22HCI2519%22&year=2020).

VCF starts the support of NSX-T Federation from its VCF release 4.2.
https://blogs.vmware.com/cloud-foundation/2021/02/09/introducing-nsx-t-federation-support-in-vmware-cloud-foundation/.
For more information on NSX-T Federation in VCF:

- VMware Cloud Foundation 4.3 FAQ
  https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/datasheet/products/vmware-cloud-foundation-faq.pdf
- Contact your VMware Sales Representative.

Copyright © 2022 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents. VMware is a registered trademark or trademark of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.