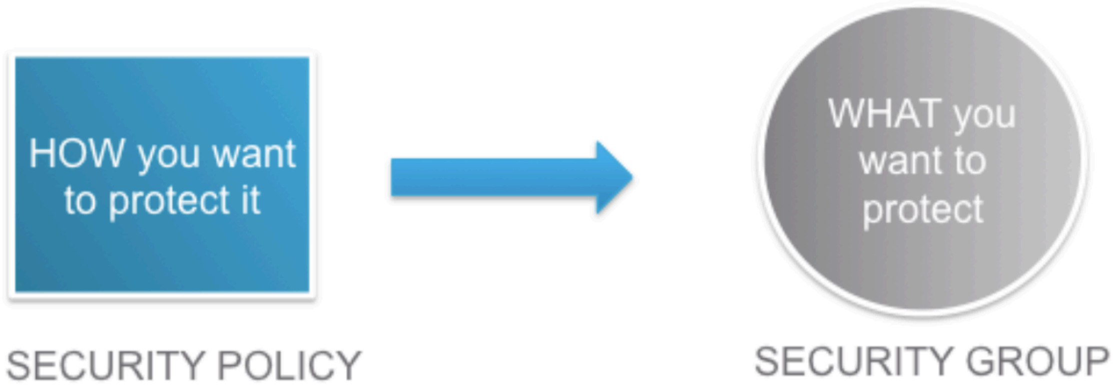
Daniel Paluszek - Clouds and Cartridges

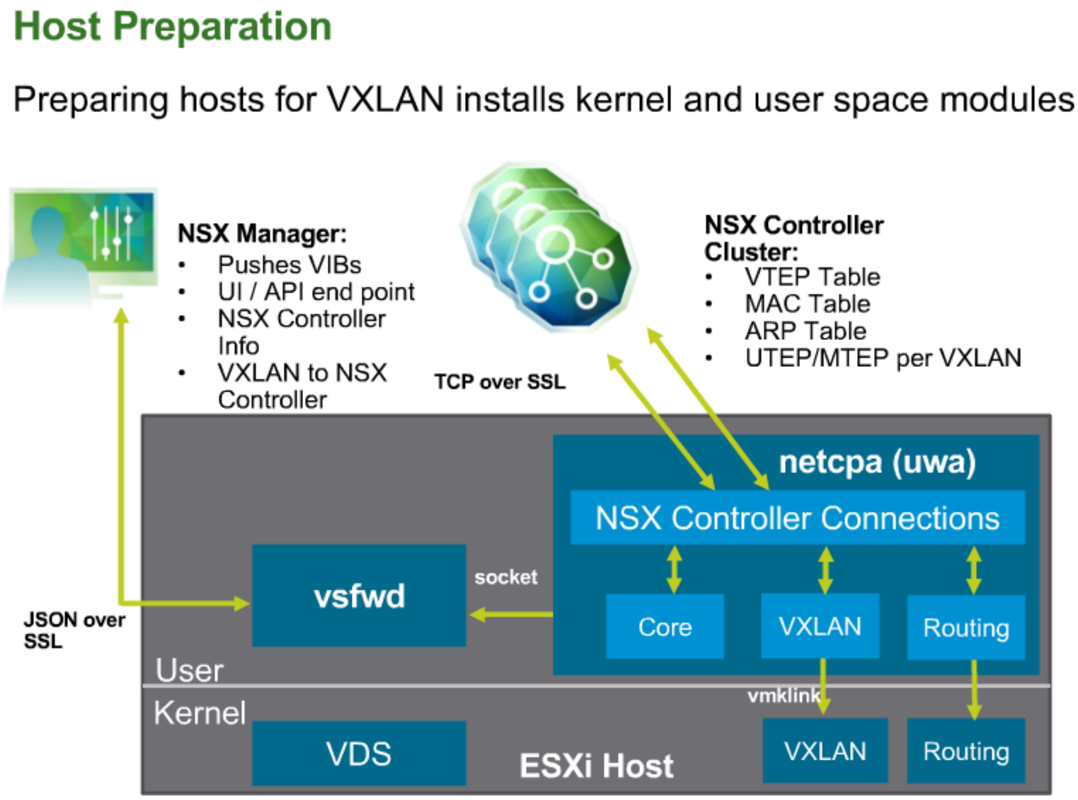
UNOFFICIAL NSX Study Notes for 2v0-642 - 30June2017

1. The DLR kernel modules are equipped with logical interfaces (LIFs) connecting to the different logical switches. Each LIF has an IP address representing the default IP gateway for its logical L2 segment as well as a vMAC address. The IP address is unique per LIF and remains same where the logical switch exists. The vMAC associated with each LIF remains the consistent in each hypervisor as well and thus during the vMotion, the default gateway and MAC remains the same.



2. You can run a IPv6 network over NSX on IPv4
3. Which features are supported on the vDS only
    1. Network I/O Control
    2. DSCP Marking (CoS)
    3. Traffic Visibility and Monitoring
    4. Scalability
        1. Enhanced LACP
        2. Enhanced SR-IOV
        3. 40GigE support
    5. Port Mirroring
4. Network services such as load balancers and firewalls are attached to the L2/L3 boundary between the core and distribution layer in a 3-tier design
5. No connectivity between spine switches
6. Always 3 controllers deployed
7. User World Agent is referred to as the NETCPA service
8. UAW - mediates between NSX Controller and the hypervisor kernel modules, EXCEPT THE DISTRIBUTED FIREWALL
9. Each role requires a master controller
10. Masters for different roles can sit on different nodes

11. Physical switches must support 1600 MTU
12. NSX Controller Cluster should be deployed as a dedicated management cluster if possible
13. Have to have correct DNS for ESXi hosts + user permissions to add and power on VM's
14. Ports are 443, 902, 903 between web client and ESXi hosts.
15. 4 vCPUs, 4GB memory controller node
16. Controller HA requires anti-affinity rules with a minimum of 3 hosts
17. NSX Controllers eliminate the need for multicast network
18. There's hardware offloading to network adapters - creates overhead and using hardware offloading helps with utilization
19. VSFWD runs on the ESXI hosts too, communicates with NSX Manager to make distributed firewall changes



Host Preparation

Preparing hosts for VXLAN installs kernel and user space modules.

NSX Manager:
- Pushes VIBs
- UI / API end point
- NSX Controller Info
- VXLAN to NSX Controller

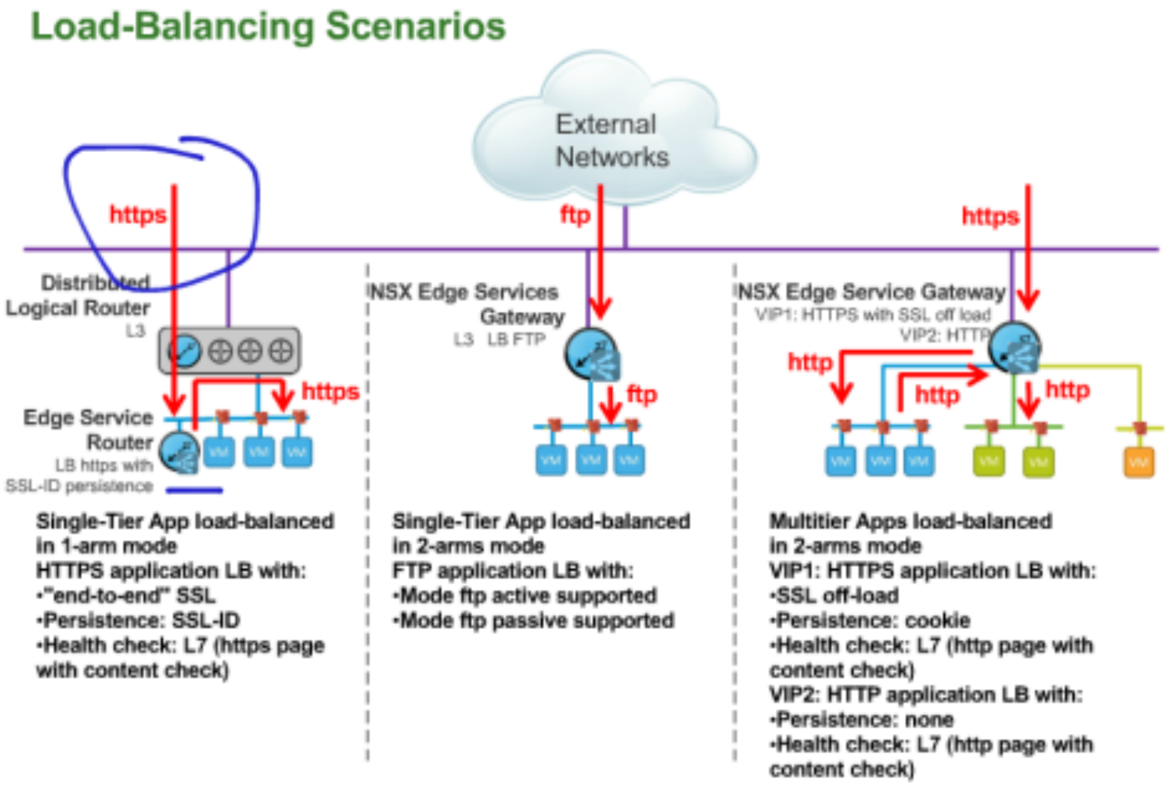
NSX Controller Cluster:
- VTEP Table
- MAC Table
- ARP Table
- UTEP/MTEP per VXLAN

20. Broadcast - flood it everywhere, all ports on a layer 2 network
21. Unknown unicast -
    1. One machine trying to communicate to another machine but doesn't know where it is or how to get there
    2. MAC addresses table maps shows what MAC goes to what port
    3. Floods out all ports until destination is found
    4. Then table is updated with correct location
22. Multicast
    1. Think of a Pay-Per-View analogy - one senders to many recipients

    2. Machine1 is the multicast source, sending out something that others have to view
    3. Will transmit data once to the switch, then the switch will send it to the joined ports in that multicast group
23. VTEP Proxy is a VTEP that forwards VXLAN traffic to its local segment from another VTEP in a remote segment. In unicast mode, this is known as the UTEP - VTEP designated as a proxy
24. Unicast and Hybrid are based on the NSX Controller
    1. Unicast mode is all replication using unicast
    2. Hybrid mode is local replication that is offloaded to the physical network and remote replication through unicast
    3. Hybrid requires Layer 2 multicast, but not Layer 3 multicast
25. Multicast mode requires Internet Group Management Protocol (IGMP)
    1. It also requires layer 2 and layer 3 multicast (IGMP and PIM) to delivery traffic
    2. The physical network handles all of the replication of the traffic - source VTEP just sends it once
    3. Proxy mode does not exist, no reliance on the NSX Controllers
26. Hybrid Mode
    1. All local VTEPs are delivered by Multicast via layer 2
    2. Packet is then unicasted to the remote UTEP - each remote transport subnet
    3. Does not need PIM in the router
    4. Proxy UTEP then uses layer 2 multicast to replicate locally on the network
    5. In hybrid mode, a VTEP designated to be a proxy for traffic originating in other segments. This proxy is known as a MTEP because it uses multicast to replicate traffic within a segment
27. MTEP IS ONLY USED FOR HYBRID MODE!
28. NSX Controller Cluster has three different tables - MAC, ARP, and VTEP
29. Jobs of the VTEP table - inform the NSX Controllers about the VNI on who's a member of
30. NSX Controller builds the VTEP table, then sends a copy to all VTEPs
31. IP Report is used for ARP Suppression - minimize on L2 flooding
32. Host / VTEP must query the controller for ARP/MAC addresses every time
33. If the NSX Controller is down..
    1. Sends the ARP request - controller doesn't answer because its down
    2. So replicates unicast copy to all VTEPs to see if they can answer
    3. All ARP requests are sent to local VTEPs and proxy VTEPs
    4. A ton of unknown broadcast traffic will be high when the controllers are down
    5. This is when you are in unicast mode - if using Hybrid/Multicast mode - problem isn't compounded as bad bc the physical network manages the multicast groups for offloading

34. VXLAN adds 50 bytes to a standard Ethernet frame
35. Proxy mode in unicast - UTEP
36. Proxy mode in hybrid - MTEP
37. DLR Control VM does not communicate directly with the kernel modules, uses the controller cluster to push it down using the UAW
38. Point is - traffic never flows through the Control VM, goes to the Edge or next hop GW
39. If the distributed logical router connects to a logical switch, the interface is called VXLAN LIF
40. If the DLR connects to a vDS Port group, it's called a VLAN LIF
41. IP addresses are assigned on the LIFs
42. Multiple LIFs can be configured one one distributed logical router instance.
43. An ARP table is maintained per LIF
44. The virtual MAC (vMAC) is the MAC address of the LIF
    1. vMAC is the same across all the hosts and it is never seen by the physical network. The vMAC is seen only by VMs and never seen by the physical network
    2. VMs use the vMAC as their default gateway MAC address
45. The physical MAC (pMAC) is the MAC address of the uplink through which the traffic flows to the physical network
    1. For vLAN LIFs, the pMAC is seen by the physical network
    2. You need a designated instance of a DLR for a VLAN LIF
    3. VLAN LIFs make things more complicated
46. If you are going to use a VLAN LIF, you MUST HAVE A DESIGNATED INSTANCE
47. Only one VXLAN LIF can connect to a logical switch
48. VLAN LIFs can only span one distributed switch
49. Edge Sizing
    1. Compact - 1vCPU / 512MB
    2. Large - 2 vCPU / 1024MB
    3. Quad-Large - 4 vCPU / 1024MB - Suitable for high-performance firewall and routing
    4. X-Large - 6 vCPU / 8192MB - Suitable for high-performance Layer 7 load balancer
50. Load Balancing Modes
    1. One-Arm Load Balancer - also called proxy mode. The NSX Edge gateway uses one interface to advertise the vIP address and to connect to the web servers.
        1. Considerations - increases the number of NSX Edge appliances that are deployed + Client IP address is not preserved
    2. Inline Load Balancer - also called transparent mode. Uses the following distinct interfaces
        1. An interface to advertise the vIP address

    2. An interface to connect to the web servers
    3. Considerations - client IP address is preserved + Edge gateway must exist and the web servers must point to this Edge as the default gateway
    4. Traffic flow is like a traditional firewall design



Single-Tier App load-balanced in 1-arm mode
HTTPS application LB with:
•"end-to-end" SSL
•Persistence: SSL-ID
•Health check: L7 (https page with content check)

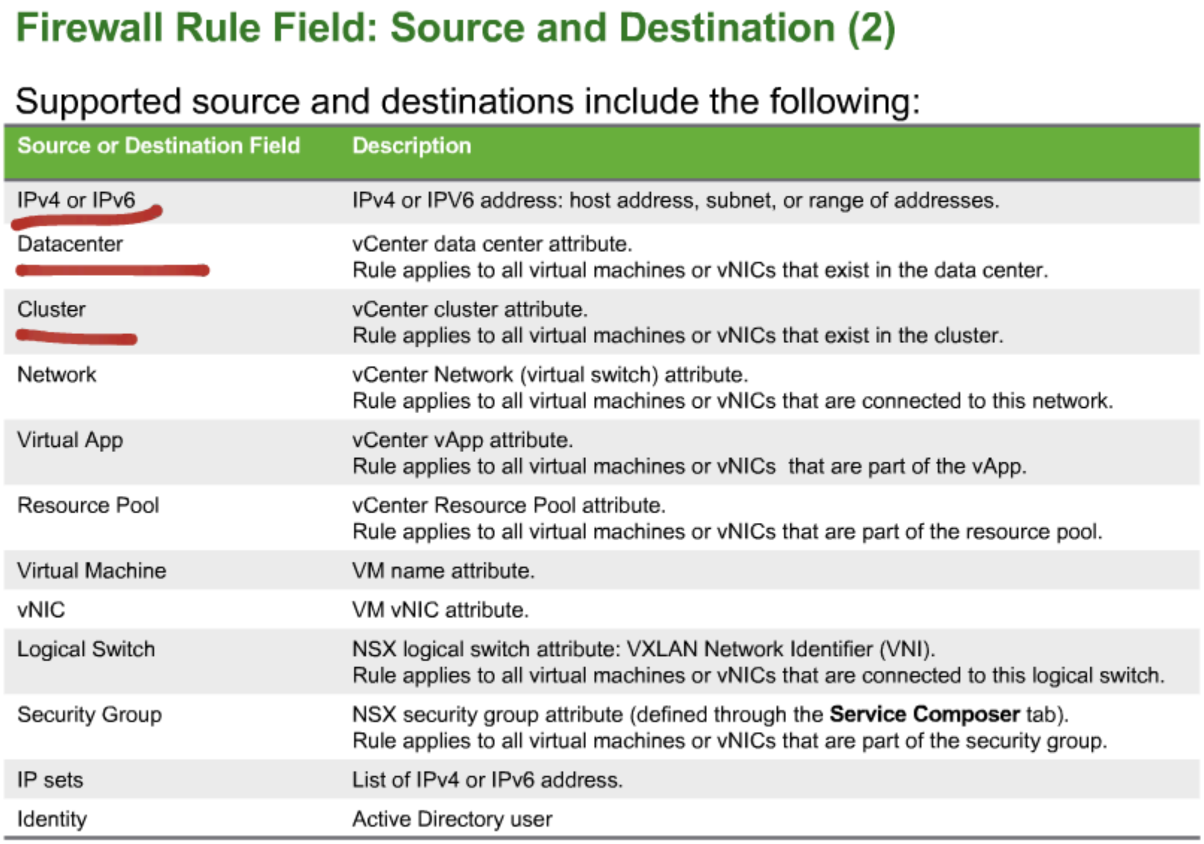
Single-Tier App load-balanced in 2-arms mode
FTP application LB with:
•Mode ftp active supported
•Mode ftp passive supported

Multitier Apps load-balanced in 2-arms mode
VIP1: HTTPS application LB with:
•SSL off-load
•Persistence: cookie
•Health check: L7 (http page with content check)
VIP2: HTTP application LB with:
•Persistence: none
•Health check: L7 (http page with content check)

51. Edge HA - can be deployed in pairs for a HA, network-services solution provider
    1. Active/standby Edges are placed in different hosts
    2. Heartbeat and sync packets are send over the same internal vNIC - internal subnet
    3. On host failure, an attempt is made to maintain the Edge gateways in separate hosts
    4. Called Stateful HA - kind of stateful, not stateful for VPN connections
    5. Default dead timer is 15 seconds and can be brought down to 6
52. ECMP - provides a STATEFUL FIREWALL
    1. N/S traffic is handled by all active NSX Edge instances
    2. Multiple equal cost paths can be programmed
    3. ECMP can be enabled on both DLR and ESG
    4. Traffic is based across equal cost paths based on SRC/DST IP address values
    5. When ECMP is enabled, the NSX Edge firewall is disabled
    6. Recommended to tune the hello and hold time timers to speed up traffic

recovery - can be to 1/3 second

53. VPNs
    1. Layer 2 VPN - join layer networks
        1. Use case - Layer 2 Cloud Onboarding -> same addresses in the cloud
        2. Do not need a full NSX Deployment at client location for this to work
    2. IPSec VPN - used for site to site connectivity
        1. Uses Internet Key Exchange v1 and use NON-OVERLAPPING networks
        2. Supports AES, AES256, and TripleDES
    3. SSL VPN-Plus - Enables remote users to connect to a private network behind an Edge
        1. SSL VPN - supports 3DES, AES128, AES256
        2. Encrypted with AES256 and SHA
54. Bridging - instance can now do bridge and router mode at the same time.
55. Firewall Capabilities
    1. Compact - vCPU / 512MB - 64k connections, 2k rules
    2. Large - 2 vCPU / 1GB - 1MM connections, 2k rules
    3. Quad Large - 4vCPU / 1GB - 1MM connections, 2k rules
    4. Extra-Large - 6 vCPU 8GB - 1MM connections, 2k rules
56. NSX Manager is going to communicate directly with the VFSWD daemon as you make rule changes. NSX MANAGER PUSHES directly to this daemon on the host - NOT THROUGH CONTROLLERS!
57. First rule that is match is enforced - goal is to have the specific rules at the top

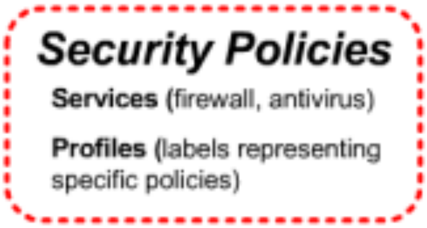


Firewall Rule Field: Source and Destination (2)

Supported source and destinations include the following:

Source or Destination Field	Description
IPv4 or IPv6	IPv4 or IPV6 address: host address, subnet, or range of addresses.
Datacenter	vCenter data center attribute. Rule applies to all virtual machines or vNICs that exist in the data center.
Cluster	vCenter cluster attribute. Rule applies to all virtual machines or vNICs that exist in the cluster.
Network	vCenter Network (virtual switch) attribute. Rule applies to all virtual machines or vNICs that are connected to this network.
Virtual App	vCenter vApp attribute. Rule applies to all virtual machines or vNICs that are part of the vApp.
Resource Pool	vCenter Resource Pool attribute. Rule applies to all virtual machines or vNICs that are part of the resource pool.
Virtual Machine	VM name attribute.
vNIC	VM vNIC attribute.
Logical Switch	NSX logical switch attribute: VXLAN Network Identifier (VNI). Rule applies to all virtual machines or vNICs that are connected to this logical switch.
Security Group	NSX security group attribute (defined through the **Service Composer** tab). Rule applies to all virtual machines or vNICs that are part of the security group.
IP sets	List of IPv4 or IPv6 address.
Identity	Active Directory user

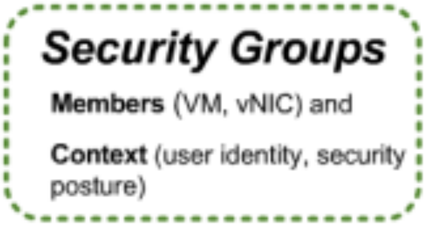
58. Actions: Deny - black holed, Accept - allowed, Reject - received response saying prohibited (unreachable with network administratively prohibited code for UDP, ICMP, and other IP connections)
59. Service Composer
    1. Static exclusion supersedes everything
    2. (Dynamic Inclusion + Static Inclusion) - Static Exclusion = Security Group Members
    3. Policy can contain Guest introspection services (Antivirus, vulnerability mgmt, data security/loss prevention), network introspection services, firewall services



Security Policies

Services (firewall, antivirus)

Profiles (labels representing specific policies)



Security Groups

Members (VM, vNIC) and

Context (user identity, security posture)

4. Weight as a tie breaker in these policies. Applied according to their respective weight while the highest weight wins -> most priority

60. Flow Monitoring

    1. Traffic analysis tool that provides a detailed view of the traffic to and from the protected VMs
    2. Configures the DFW to capture flows and send them to the NSX Manager for retention
61. NAT
    1. Source NAT - translates source IP address of outbound packets so that packets appear as originating from a different network
        1. Example: translate private (internal) IP addresses into a public IP (globally routable) for all the traffic going outbound from the private addresses
    2. Destination NAT - translates the destination IP addresses of inbound packets so that packets are delivered to a target address into another network
        1. Example: make a private (internal) service available (published) from the outside on a publicly accessible IP address
62. RBAC
    1. Identity sources from SSO
        1. Microsoft AD
        2. NIS
        3. LDAP
    2. Default NSX admin user cannot be disabled
    3. NSX has predefined roles
    4. Roles
        1. Enterprise Administrator - has read/write access to all areas in NSX - do everything, everywhere!
        2. NSX Administrator - R/W access to NSX opes such as installing virtual appliances and configuring Pos, but only read-only to other areas
        3. Security Administrator - R/W to NSX Security area such as data security policies, creating port groups and creating reports for NSX modules. Read-only for other others
        4. Auditor - has read only to all areas
        5. New roles cannot be created
    5. Scopes
        1. Global - user has access to all areas in NSX
        2. Limited Access - user has access to the NSX areas defined in the user profile
    6. NSX permissions are independent of vCenter Server permissions
63. Monitoring Tools
    1. Rules messages are in a dedicated file - /var/log/dfwpktlogs.log
64. Cross-vCenter NSX
    1. NSX Controller is only present with one NSX instance
    2. Universal functions - transport zone, logical switches, etc are created
    3. Components

   1. Universal controller cluster
   2. Universal transport zone
   3. Universal logical switch
   4. Universal distributed logical router
   5. Universal IP Set
   6. Universal MAC set
   7. Universal security group
4. NSX Manager will have the following roles
   1. Standalone - newly installed or upgraded instance and default. Not part of cross-vCenter setup
   2. Primary - designated Manager instance as the primary in a cross-center setup. Has controllers installed and all universal objects are created, modified, or deleted on the primary Manager instance
   3. Secondary - comes secondary when it is added to the primary Manager instance. All universal objects are read/only. Cannot have its own controllers.
   4. Transit - can be made standalone from either the primary or secondary instances. However, if the NSX manager instance has universal objects, it cannot be assigned the standalone role by definition. In such cases, the NSX Manager instance is assigned the transit role. In the transit role, a universal object can only be deleted. An NSX Manager instance can be assigned the secondary role after all the universal objects are deleted.
5. Egress-optimized routing includes: Locale ID - metadata that describes location
6. Universal Synchronization Service -
   1. Responsible for synchronizing configuration changes from the primary instance to all secondary instances
   2. Service is part of the NSX Manager appliance. However, only runs on the primary instance.
   3. Service uses the REST API protocol to perform the sync.
   4. Service uses a special replicator user to execute the REST APIs. This user is created when the secondary NSX manager instance is registered with the primary. This user exists on all the secondary instances.
   5. The replicator user has limited privileges on the secondary instances. It can only create, update, or delete universal objects on the secondary instances.
7. Establish a pool of Universal VNIs - do not overlap with local VNIs
8. Enabling Local Egress
   1. Local egress must be enabled when an admin creates the universal logical router. Local egress cannot be changed after creation.
   2. Logical egress controls the routes provides to the ESXi hosts based

        on the locale ID. When the local egress is enabled, the NSX Controller instance sends routes only to hosts with a matching local ID. Local egress is useful when the universal LR extends across multiple sites and customized routes are required.
    3. Each NSX manager instance is assigned a locale ID. Default local ID is the NSX Manager UUID
    4. If the same NSX manager instance is used across multiple sites, an admin can define a locale ID at the universal logical router, cluster, or at the host level.
    5. Using a site-specific uplink, each site can have a local routing configuration.
    6. The locale ID is ignored when local egress is not enabled.
    7. Up to 8 sites

65. Design Considerations
    1. Failover as the VMKNIC teaming policy - single VTEP
    2. Configured with Source MAC or sport port - multiple VTEPs
    3. Multicast - PIM and IGMP Snooping. If using Hybrid, PIM Is not required.