Virtual Design Master

Challenge 1 – Datacenter Design

Prepare a Multi-Site environment for the world's new Infrastructure

Submitted by-Harshvardhan Gupta

6/28/2016

Table of Contents

1. Executive Summary

1.1.1 Project Synopsis

Our billionaire friend has been recruiting to re-build his empire, and his first order of business is **datacentre** hardware. Prepare a multisite environment for the world's new infrastructure. **Primary site must be on Earth,** wherever you would like it, and Secondary site must be on the moon. The most critical application is the Humanity Link software suite, which consists of three front end web servers, one database, and two application servers. Performance of this software is paramount. The environment must also support 25 web servers, 5 databases, and 10 application servers.

1.1.2 Intended Viewers

This document is specifically written for technical people responsible for deploying Datacentre in Earth and Moon.

1.1.3 Project Vision

Prepare a Multi-Site environment for the world's new Infrastructure. Primary Site will be established at North Virginia (U.S.) and Secondary Site will exist on the moon. Will use State-of the art Hyper-converged infrastructure (HCI) comprises of All-Flash Configuration (NVMe SSD), Infiniband 40 Gbps Dual port Nic connectivity option and Xeon E7-8880 v4 (22 cores/L3 cache 55 mb/150 W TDP) with DDR4-1866 RAM.

1.1.4 Project Requirements

Reference	Description
R001	Multisite Datacentre with Primary site on earth & Secondary on Moon.
R002	Humanity Link Software must be Highly Available with no performance lag.
R003	Must run 25 Web Srv, 5 DB and 10 App Srv.
R004	

1.1.5 Project Constraints

Reference	Description
C001	Primary Site on Earth and Secondary on the Moon
C002	Humanity Link Software is paramount
C003	Must use currently Available Cloud Suite.
C004	

1.1.6 Project Assumptions

Reference	Description
A001	Existing WAN Link provides less than 150 ms RTT between Sites.
A002	Technical people are trained on VMware SDDC software suite
A003	Enough license are available for Software being used.
A004	Datacentre in both location must have 3 diff. power sources and sufficient cooling
A005	Sufficient Copper/Fibre connectivity is available and infiniband switch is available
A006	NTP exist for time sync between equipment's
A007	AD/DNS Infrastructure exist for SSO/Cert/etc.
A008	Hardened MS Win 2012 R2 template will be available for OS deployment

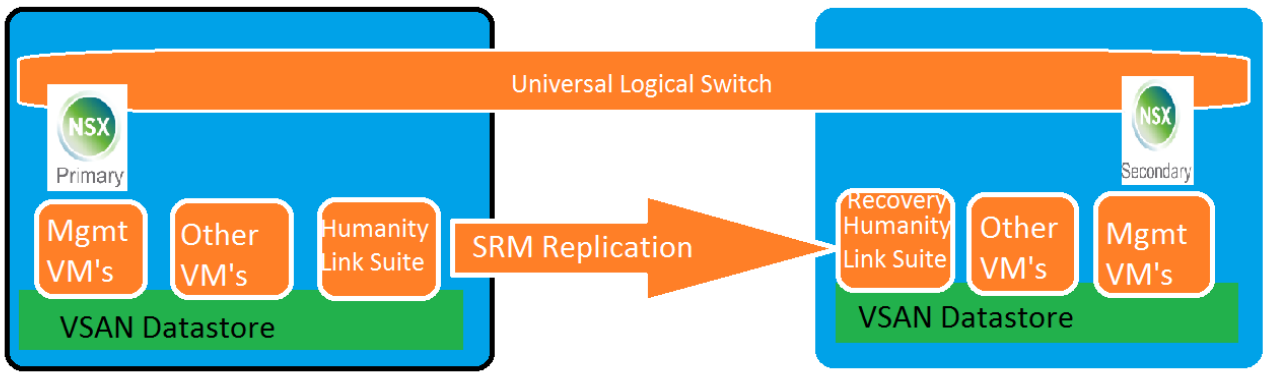
1.1.7 Project Risks

Reference	Description
I001	Obtaining spare parts is difficult\impossible\expensive on Moon
I002	No protection shield - fully exposed to Solar flare radiation
I003	

2. Conceptual Design

The Environment will use two datacentre one in North Virginia (US) and another at VDM Season 2 Site on the moon.
Both datacentre will utilize 3-Node HCI Servers with State-of-the-art H/W for high durability, optimum efficiency and low power consumption (Spec shared above under 1.1.3 Project Vision).

1) North Virginia Datacentre
   a) Management VM's (VCSA, NSX Mgr, Controller Cluster, VUM, SRM, PSC etc.)
   b) Other VM's (Database, Application and Web Servers of different Multi-tier apps)
   c) Humanity Link Software Suite (3 FE Web Srv, 2 App Srv & 1 DB srv.)
2) VDM Season 2 Site Datacentre
   a) Management VM's (VCSA, NSX Mgr, Controller Cluster, VUM, SRM, PSC etc.)
   b) Other VM's (Database, Application and Web Servers of different Multi-tier apps)
   c) DR for Humanity Link Software Suite (3 FE Web Srv, 2 App Srv & 1 DB srv.)

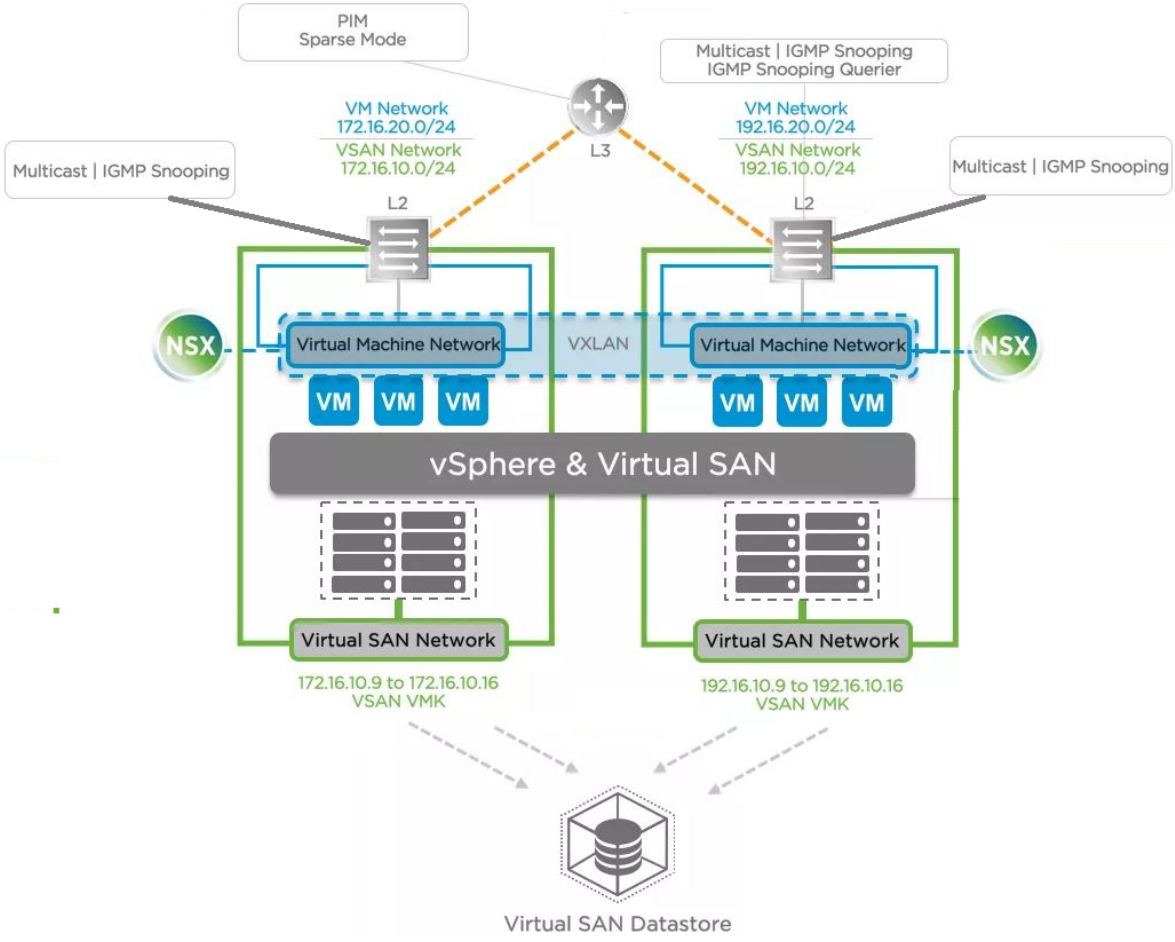


Advantages of above approach-

a) "Universal Logical Switches" (ULS) – This allows for the creation of Layer-2 networks that span vCenter boundaries. This means that when utilizing ULS with NSX there will be a virtual port group at both the protected and recovery site that connect to the same Layer-2 network. When virtual machines are connected to port groups that are backed by ULS, SRM implicitly creates a network mapping, without requiring the admin to configure it. Providing seamless network services portability and synchronization automatically reconnects virtual machines connected to a ULS to the same logical switch on the other vCenter.
b) Cross vCenter Networking and Security enables key use cases such as:
   - Resource pooling, virtual machine mobility, multi-site and disaster recovery
   - Cross-vCenter NSX eliminates the need for guest customization of IP addresses
   - and management of port group mappings, two large SRM pain points today
c) Centralized management of universal objects, reducing administration effort Increased mobility of workloads; virtual machines can be "vMotioned" across vCenter Servers without having to reconfigure the virtual machine or making changes to firewall rules

2.1.1 Architecture Design

Both Datacenters will be setup with identical configuration to provide standardization and Keep it simple for implementation teams. 10,000 foot view of data centre layout is illustrated below.



2.1.2 Software Defined Storage – VSAN 6.2

Design choices- following design choices were made

Virtual SAN- Shared storage for this environment will be provided by Virtual SAN (SDS offering from VMware), VSAN has got below advantage compared to traditional/legacy SAN-

- Saves on power, cooling and space when compared to any standalone shared storage solution.
- Reduces the need for additional storage related components.
- Saves on shipping\transport costs.
- Smaller learning curve for Datacenters support staff.

All flash drive (NVMe) Configuration chosen over regular magnetic drives for the following reasons:

- Flash drives are significantly lighter to transport assuming that they can be shipped from Earth.
- SSDs consume much less power than magnetic drives.
- SSDs run cooler than magnetic drives.
- SSDs provide more IOPS per disk which lowers the number of disks needed.

VSAN Disk Groups (DG)

2 disk groups per host is my default choice with VSAN to increase the overall availability.

Smaller disk groups would also mean that it is more likely that the VSAN cluster will find the spare capacity on another host\disk to reconfigure the storage objects.

Failures to Tolerate (FTT)

FTT will be set to 1 to retain maximum usable capacity while still providing resiliency.

Number of hosts

With FTT set to 1 the minimum number of hosts would be 3, however following the standard N+2 model and taking into account possible failures during maintenances 4 hosts would be needed at a minimum.

Capacity

Based on the server choice which allows for a maximum of 6 disks per server and the design choices above, the vSAN layout for each datacenter will be as follows:

Hosts	ESXi A		ESXi B		ESXi C		ESXi D	
Disk Groups	DG1	DG2	DG1	DG2	DG1	DG2	DG1	DG2
Caching Layer	2 TB	2 TB	2 TB	2 TB	2 TB	2 TB	2 TB	2 TB
Capacity Layer	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB
	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB	1.6 TB
Per Host (RAW)	6.4 TB		6.4 TB		6.4 TB		6.4 TB	
Per Datacentre (RAW)	25.6 TB							
Usable Capacity (FTT=1)	12.8 TB							
Usable Capacity (FTT=1) across environment	38.4 TB							

2.1.3 HCI Physical Host

Hyper convergence is 22% to 49% cheaper in comparison to any Rack or blade architecture. HCI also requires less power/cooling and rack space compared to traditional Rack or Blade server.

Based on this I've zeroed down to SuperMicro 4U FatTwin, this houses 4 blade server with NVMe enabled Virtual SAN Solution.

Processor	**8 x 22-core Intel Xeon E7-8880 v4/L3 Cache 55 mb**
Memory	**64 x 16 GB PC3-14900 1866 MHz DDR4-ECC DIMM**
Hard Drive	**16 x 1.6 TB Intel SSD 8 x 2 TB Intel NVMe SSD**
Network Card	**4 x Mellanox 40- Gigabit Ethernet Adapter**
Power	**1280 W high-efficiency Redundant PSU**

Like SCSI and SATA, NVMe is designed to take advantage of the unique properties of pipeline-rich, random access, memory-based storage. The spec also reflects improvements in methods to lower data latency since SATA and AHCI were introduced. Advances include requiring only a single message for 4KB transfers as opposed to two, and the ability to process multiple queues instead of only one. By multiple, I mean a whopping 65,536 of them. That's going to speed things up a lot for servers processing lots of simultaneous disk I/O requests.

2.1.4 vSphere SDDC design

VMware vSphere SDDC Suite 6.2 will be used as preferred Operating System of choice for Bare metal H/W. Enterprise Plus Edition will be installed for SDS and SDN implementation.

SD-CARD with Raid-1 will be used for Booting ESXi software.

Compute Capacity-

	Per Host	**Per Data Center**
CPU Sockets	**2**	**8**
CPU Cores (Physical)	**22**	**88**
CPU Cores (Logical with HT Enabled)	**44**	**176**
RAM	**1 TB**	**4 TB**

BIOS Recommended Settings-

- Enable all populated processor sockets
- Enable all cores

- Enable Hyper threading
- Disable Node interleaving
- Enable VT-x
- Set Power management to ' OS controlled'
- Enable C-states

vSphere Design includes-

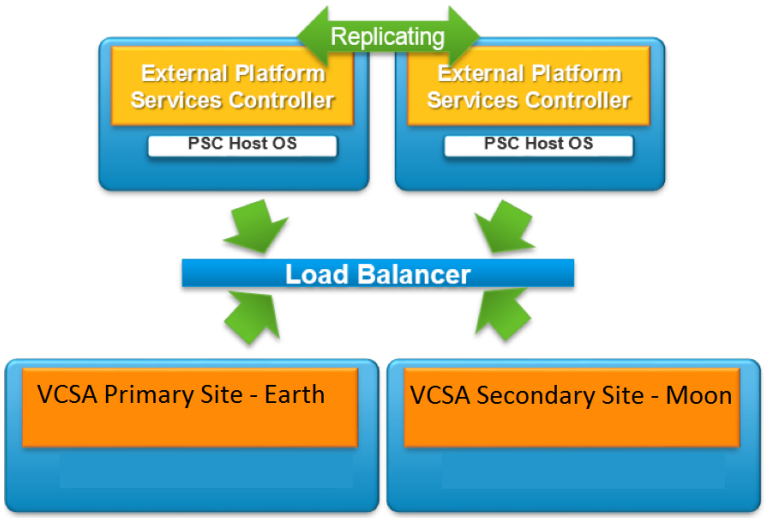
Each Data center will run a separate Management Folder comprises of VCSA 6.2, VUM, SRM, NSX Manager, Controller Cluster (3 VM's), AD, DNS etc. there will be 1 Cluster of 4 ESXi hosts in each Datacenter.

Cluster Configuration-

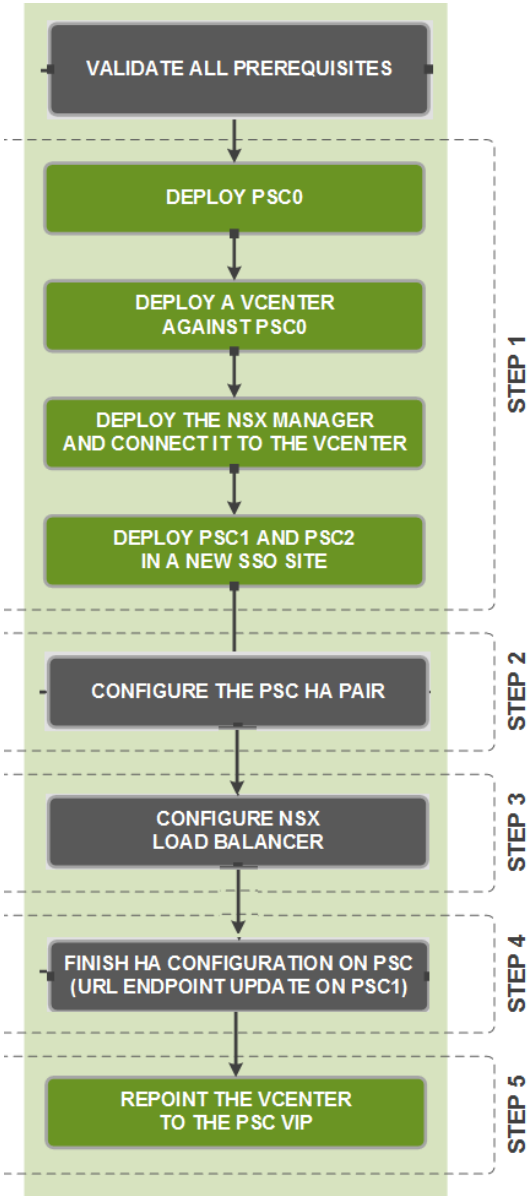
- The cluster will be enabled for vSAN.
- HA and vMotion on each host will be enabled once the VSAN cluster has been created.
- DRS will be enabled (fully automated) once the VSAN cluster has been created.
- As there will be only 1 VSAN datastore available in the cluster Datastore heartbeating will be disabled.
- VMware EVC mode will be enabled and set to Broadwell.
- VSphere HA admission policy will be set to tolerate 1 host failure to match the VSAN FTT level.
- VSphere HA isolation response will be set to 'Power off' to prevent the possibility of a FDM restart resulting in 2 instances of the same VM.

VMware vCenter Design-

- Provisioned as Virtual Appliance in each data center.
- An embedded vPostgresql Database will be used and External PSC will be used, load balanced with NSX v-Load Balancer
- Platform Service Controller (PSC) provides a set of common infrastructure services
    - Single Sign-On (SSO)
    - Licensing
    - Certificate Authority
- The following functions will be installed as a separate Windows Server 2012 R2 VM:
    - VMware Update manager
    - VMware Syslog Collector
    - vSphere ESXi Dump collector
- VCenter server will be accessible via the web client as well as a locally installed vSphere client on the additional VM.



The deployment process would be to:



Virtual Machine Environments- The VM in each environment can be segregated into two logical groups (folders)

Management VM's-

Virtual Machine	vCPU	vRAM (GB)	Storage (GB)
vCSA	8	16	120
PSC appliance x2	2	2	30
VUM/Syslog/Dump collector/SRM	4	8	250
AD+DNS	2	4	80
SRM replication appliance x2	2	4	20
NSX Mgr	4	12	60
Controller VM x3	4	4	25
NSX Edge	2 (Large)	1 (Large)	512 Mb

Production VM's-

This will host all VM's related to Humanity Link software Suite (3 FE Web Srv, 2 App Srv & 1 DB) and 25 Web Srv, 5 Database Srv and 10 Application Srv. No Sizing Information is Available at this stage.

Guest OS/deployments-

All VM will be deployed from Hardened MS Win 2012 R2 Template. KMS server will be used for License Activation across datacenter. Appliance will be installed manually from .ova with their default VM H/w version.

Network Design-

Each ESXi host comes equipped with Mellanox 40 Gigabit (2x SFP) Ethernet Adapter i.e. infiniband.

Following Network Setup will be used with 2 uplinks per host-

- A single virtual distributed switch with 2 uplinks.
- Ports groups will be created for:
  - Management (VMkernel)
  - vMotion (VMkernel)
  - vSAN (VMkernel)
  - VM traffic
- Vlans will be used to segregate traffic.
- Explicit Failover order will be defined.
- NIOC v3 will be enabled and shares will be assigned to ensure appropriate bandwidth in case of contention.
- Each uplink will be connected to a separate Voltaire infiniband switch.

Port Group Name	Explicit Failover Order	
	Primary Uplink	Secondary Uplink
Management	dvUplink 1	dvUplink 2
vMotion	dvUplink 1	dvUplink 2
vSAN	dvUplink 1	dvUplink 2
VM Traffic	dvUplink 2	dvUplink 1

Site Recovery Design-

The disaster recovery planning, maintenance and testing process becomes much simpler, with the integration of NSX and SRM.

- The ability to create a network that spans vCenter boundaries creates a cross-site Layer-2 network, which means that after failover, it is no longer necessary to re-configure IP addresses. Not having to re-IP recovered virtual machines can further reduce recovery time by up to 40 percent.
- There is more automation with networking and security objects. Logical switching, logical routing, security policies (such as security groups), firewall settings and edge configurations are also preserved on recovered virtual machines, further decreasing the need for manual configurations post-recovery.
- Making an isolated test network with all the same capabilities identical to a production environment becomes much easier.
- This lowers operational expenses, increases testing capabilities and reduces recovery times.

The deployment process would be to:

- Configure Master NSX Manager at primary site and Secondary NSX Manager at recovery site
- Configure Universal Distributed Logical Router between primary and secondary site
- Deploy Universal Logical Switch between primary and recovery site and connect it to Universal Distributed Logical Router
- Finally map SRM network resources between primary and recovery sites

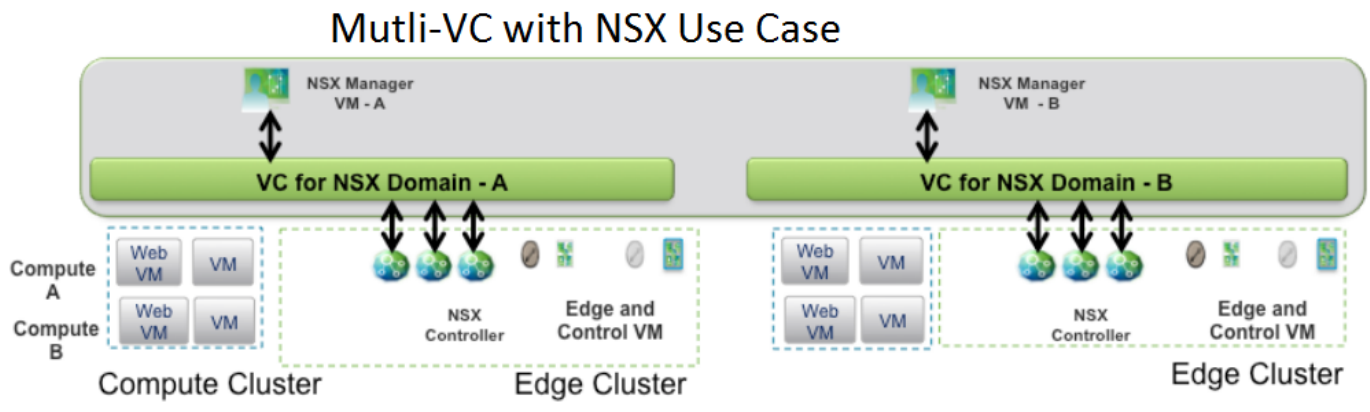
2.1.5 Software Defined Network

VMware Network Virtualization Platform (NSX) is the network virtualization platform for the software-defined datacenter (SDDC). Network virtualization using VMware NSX enables virtual networks to be created as software entities, saved and restored, and deleted on demand without requiring any reconfiguration of the physical network. Logical network entities like logical switch, logical routers, security objects, logical load balancers, distributed firewall rules and service composer rules are created as part of virtualizing the network.

Advantages of NSX-

- Moves Networking to Software
- Cuts Network Provisioning time
- Let's move workloads seamlessly across vCenter
- Enables Network Micro-segmentation
- Integrates with 3rd Party products

Multi vCenter with NSX Design-



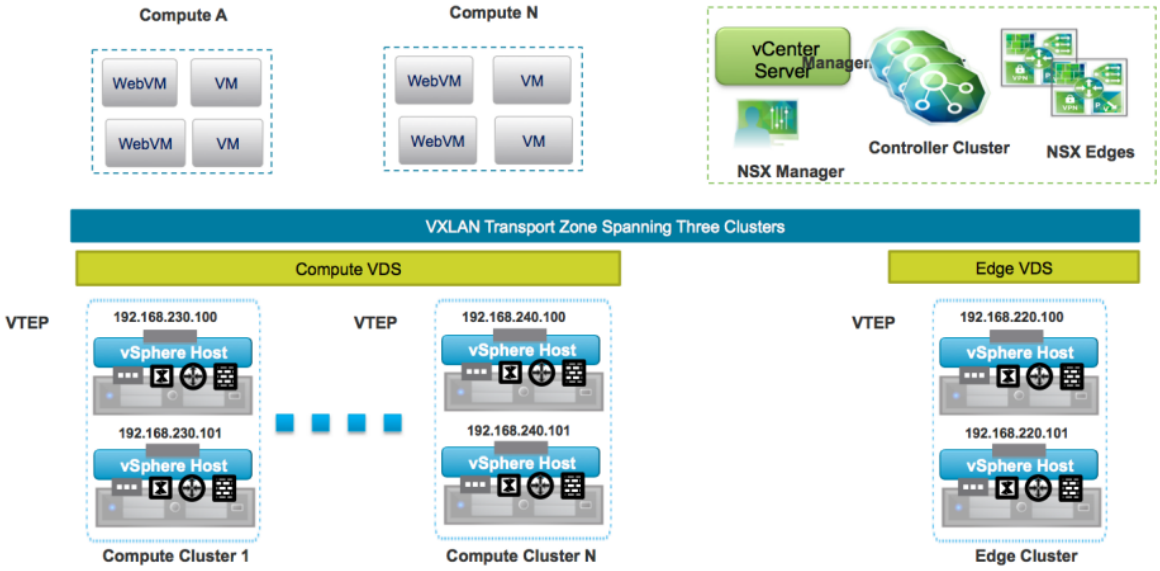
There are several advantages in adopting such approach:

- Avoid circular dependencies – the management cluster does not need to be prepared for VXLAN and normally resides outside of the domain it manages
- Mobility of management cluster for remote datacenter operation.
- Integration with an existing vCenter.
- Ability to deploy more than one NSX domain and upgrade independent of each other.
- Upgrade of management vCenter does not directly impact the NSX domains.
- SRM and other explicit state management are possible.

Multi vCenter with Universal Distributed Firewall design-

With UDFW, rules can be created and synced between vCenter. This is in addition to the local distributed rules that are allowed per vCenter and helps isolate Web/App and DB VM's.

VDS Design with NSX-



Advantages in keeping separate VDS for compute and edge:

- Flexibility in span of operational control. Compute/virtual infrastructure administration and network administration are typically separate groups, allowing each to manage the cluster specific tasks.
- Flexibility in managing uplink connectivity on computes and edge clusters.
- The VDS boundary is typically aligned with the transport zone, allowing VMs connected to logical switches to span the transport zone.
- Flexibility in managing VTEP configuration.

3. References

1. http://blogs.vmware.com/consulting/2015/11/configuring-nsx-v-load-balancer-for-use-with-vsphere-platform-services-controller-psc-6-0.html
2. http://www.pcworld.com/article/2899351/everything-you-need-to-know-about-nvme.html
3. http://blogs.vmware.com/vsphere/2015/03/vcenter-server-6-topology-ha.html

4. http://www.vmwarearena.com/vmware-site-recovery-manager-srm-6-0-part-2-vmware-srm-6-0-installation/
5. https://www.vmware.com/files/pdf/products/vsan/vmware-vsan-disaster-recovery-solution-overview.pdf
6. https://blogs.vmware.com/virtualblocks/2016/01/18/vmware-virtual-san-and-vmware-nsx-compatibility/