Infrastructure design

CHALLENGE 1: TAKING BACK THE EARTH!

Katarina Wagnerova
katkaaw@gmail.com
28.06.2016

Table of Contents

Table of Contents 1
1 Mission objective 2
1.1 Design qualities 2
1.1.1 Design quality ranking 2
1.2 Requirements 3
1.3 Constrains 3
1.4 Assumptions 3
1.5 Risks 3
2 Conceptual design 3
2.1 Datacenter location 4
2.1.1 Earth 4
2.1.2 Moon 4
3 Infrastructure design 5
3.1 Host hardware 5
3.2 Cluster design 6
4 Network design 7
4.1 Physical layer 7
4.2 Logical layer 7
4.2.1 NSX Components 7
4.2.2 Distributed switch configuration 8
4.2.3 VLANs 8
4.2.4 DNS and DHCP 9
4.3 Storage design 9
4.3.1 VSAN 9
4.4 Virtual Machine design 10
4.4.1 Naming convention 11
4.4.2 HumanityLink Application 11
4.5 Management Infrastructure 12
4.5.1 Active Directory 12
4.5.2 vCenter server 12
4.6 Backups 13
5 References 14
6 Appendix A: VM list 15

1 Mission objective

We're coming home!

Now that we have finally defeated the zombies it is time for the mankind to return back to Earth. Before we do that, we have to prepare an infrastructure to run all of our important applications, especially the HumanityLink. Luckily, hardware and software vendors continued their operations even during the apocalypse so we will be able to purchase what we need to start building the infrastructure.

All of us, who have been stocking supplies that we'd hoped to trade for needed items, were surprised to find out that money still rules the universe. Well, they may come in handy later, nobody really knows what's going to happen next. For now, we have a billionaire who will pay for everything so we don't have to worry about that (let's just hope his name is not Bobby Axelrod).

We still have some people living on the Moon so we will have to support them as well. Fortunately, given what we went through in past years, we are pros at space travelling and shipping of equipment.

Mission objective is clear now, let's start!

1.1 Design qualities

Availability

We have to make sure that our infrastructure provides a highly available solution. We cannot afford long lasting downtimes as our applications are necessary for keeping humans alive. We need to avoid single points of failures and ensure that we can sustain operations during unexpected interruptions, such as hardware failures.

Manageability

Our IT admins have been through a lot; our infrastructure should be easy to manage. We don't want to make their lives harder than it already is. At this stage we don't know how it will grow in the future; however, scalability is a factor we'll keep in mind.

Performance

HumanityLink, as well as the other applications, need to be performing well at all costs to support humans. Lag deaths are the worst.

Recoverability

Mankind depends on our ability to build a stable infrastructure; we need to be able to recover from unexpected events. When disaster strikes, we'll be ready.

Security

Most of the zombies have been wiped out but there still may be some running around. While we do not expect them to be particularly skilled hackers, we still need to make sure that our infrastructure is secure and controlled. This applies to both software part as well as our datacenter premises.

1.1.1 Design quality ranking

Based on the definitions we have prioritized the design qualities based on their importance as follows:

Rank	Design quality
1	Availability
2	Performance
3	Recoverability
4	Manageability
5	Security

1.2 Requirements

Following requirements have been derived from the mission objective:

#	Description	Link to Design qualities
R01	Build multi-site environment	AMPRS
R02	Ensure high performance of HumanityLink application	AP
R03	Provide high availability for all applications	AR
R04	Support future workloads	APM
R05	Create easy-to-use solution	M
R06	Keep zombies away	S

1.3 Constrains

#	Description
C01	Site locations specified as Earth and Moon
C02	Use existing software solutions
C04	Limited number of people back on Earth

1.4 Assumptions

#	Description
A01	Vendors have hardware available in stock
A02	Abandoned datacenters are still usable
A03	Power and cooling systems are in place
A04	Technology has advanced to provide high speed connection between sites
A05	Latency between Earth and Moon is less than 4.9ms
A06	Software installation binaries are available
A07	Licenses for all products are available
A08	Sufficient budget is available
A09	Zombies cannot travel long distances
A10	There are no zombies on the Moon
A11	Infrastructure still exists on the Moon

1.5 Risks

#	Description	Risk mitigation
RI01	Future workloads are unknown	Build scalable infrastructure
RI02	Some zombies are still living on Earth	Select secure location for datacenter
RI03	Vendor support is non-existent	Hire skilled professionals and train others
RI04	Hardware failures may occur on both locations	Implement HA solution
RI05	OS and Application failures may occur	Implement Backup solution

2 Conceptual design

Infrastructure will be spread among two interconnected locations; Earth and Moon. Earth datacenter will be used as a primary site running 80% of the workloads, including HumanityLink application and all management systems. 20% of the remaining workloads will be running in Moon datacenter to provide more convenient access to humans still residing on the Moon and to reduce the latencies when accessing applications.

Both datacenters will be equipped with the same set of hardware to allow for a full recovery in case of a disaster in one of the locations. Configuration will be identical on both sites in order to ease the management of the environment.

2.1 Datacenter location

2.1.1 Earth

Earth datacenter will be located in Bratislava, Slovakia. Location has been selected mainly due to security reasons. Nobody really knows where Slovakia is so it is safe to assume that there will be very limited zombie activity in the area. Also, zombies are known to reside mainly in the US; therefore, being slow and all, it would take them a very long time to reach the location even if they could find it on the map. In most cases it would get confused with Slovenia anyway. [R01, R06]

There is a sufficient infrastructure available to easily transfer all of the equipment and datacenters are in a good shape with available racks. Power and cooling systems are intact and ready to use, there is also a UPS generator available to provide redundancy in case of a temporary power failure.

To further improve the security, access to the datacenter will be granted to a limited list of authorized people based on a DNA sample taken and evaluated at the entrance. If any DNA mutation will be detected, access won't be granted to the zombie-free environment. Armed security guards (aged less than 60) will be available at location at all times and prepared to react in case of a security breach. We will also hang a sign *"Zombies keep out"* at the entrance gate, that should help. [R06]

2.1.2 Moon

Second datacenter will be located on the bright side of the Moon. We will utilize the existing infrastructure to provide racking, power and cooling for our systems. [R01]

There are still remaining humans occupying the Moon so we will be able to recruit them for onsite support.

While there are no zombies present on the Moon, we have to keep the location secure from other forces in the universe. Therefore, we will put up a big illuminated sign saying *"Moon Datacenter – do not destroy"* on the roof to avoid confusion with potential Death Star.

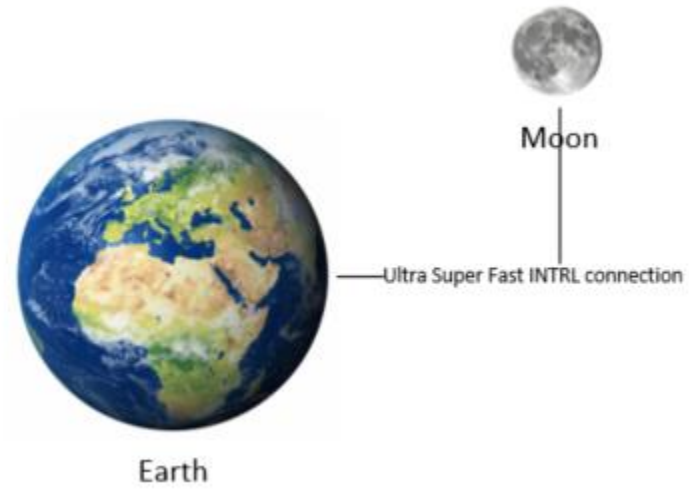

FIGURE 1 - EARTH - MOON INTERSITE CONNECTION (NOT TO SCALE)

3 Infrastructure design

3.1 Host hardware

We will utilize HPE ProLiant 380 Gen9 servers with following configuration per host:

Processor	2 x 22 core
Memory	8 x 64 GB (512 GB)
Hard Drive	1 x 120 GB SSD
	2 x 1.6 TB SSD
Network Card	2 x 10GB
MicroSD	16 GB
Remote Management	iLO 4 Advanced
Form Factor	2U

Based on the Virtual Machine list provided in Appendix A we will require total amount of 274 GB of RAM and 109 CPUs. 5 hosts per site will be needed to provide enough resources to accommodate the workloads and to follow the n+2 principle. This will allow us to perform maintenance on the hosts and provide enough capacity in case of a hardware failure.

	RAM	CPU
Resources required	698	225
Host configuration	512	88
# of hosts required	1.36	2.56
n+2	3.36	4.56
Total hosts required	**5**	**5**

Additional HPE ProLiant 380 Gen9 will be installed and act as a VSAN witness with following configuration:

Processor	1 x 2 core
Memory	16 GB
Hard Drive	1 x 10 GB SSD
	1 x 350 GB SSD
Network Card	2 x 10GB
MicroSD	16 GB
Remote Management	iLO 4 Advanced
Form Factor	2U

All hosts will be configured with the following:

- Hyper threading will be enabled on all hosts
- Installed with vSphere ESXi 6.0 (Build number 3825889)
- Installed on a local SD card
- vSphere Enterprise Plus license
- SPP 2016.04 installed
- NTP set to synchronize with PDC
- Host configuration will be stored in a Host profile

Justification

HPE hardware models are well spread and known among IT professionals, most of them will have had previous experience with such hardware, therefore it will be easier to manage.

Due to the number of needed hosts Rack Mount servers have been selected over Blade systems, to eliminate the need of purchasing additional chassis. Blade systems would also call for spreading into two chassis to eliminate a SPOF in case of a chassis failure, increasing the complexity of the infrastructure.

NTP synchronization with Primary Domain Controller will prevent time discrepancies.

Host profile will be used by Auto Deploy to configure all hosts. This will ensure a consistent configuration across the environment, as well as provide an option to easily check for compliancy.

Design qualities impacted

Availability, Performance, Manageability, Recoverability, Security

3.2 Cluster design

One stretched cluster will be spread across both sites and use the following configuration:

# of clusters	1
# of hosts	10
vSphere HA	Enabled
Admission Control	50% reserved
Host Isolation Response	Power off and restart VMs
Host Monitoring	Enabled
das.usedefaultisolationaddress	False
das.isolationaddress0	IP address on VSAN network on Earth
das.isolationaddress1	IP address on VSAN network on Moon
Distributed Resource Scheduler	Enabled – Fully Automatic

VM Group	Host Group		
BTS VMs	BTS Hosts	VM to Host Affinity	Should run
MOO VMs	MOO Hosts	VM to Host Affinity	Should run
Database VMs		Virtual Machine Rule	Separate VMs
Application VMs		Virtual Machine Rule	Separate VMs
Web Server VMs		Virtual Machine Rule	Separate VMs

Justification

The intention was to have two active sites with some kind of disaster recovery solution; however, bidirectional SRM was rapidly increasing complexity of every aspect of the design. Stretched cluster will provide disaster and downtime avoidance while keeping the environment easier to manage. We have focused on providing redundancy for applications to survive host and OS failures, which are more likely to happen than complete site failures. Also, using VSAN stretched cluster we will be able to recover quickly even from potential site disasters. [R01, R03, R04, R05]

Design qualities impacted

Availability, Performance, Manageability, Recoverability

4 Network design

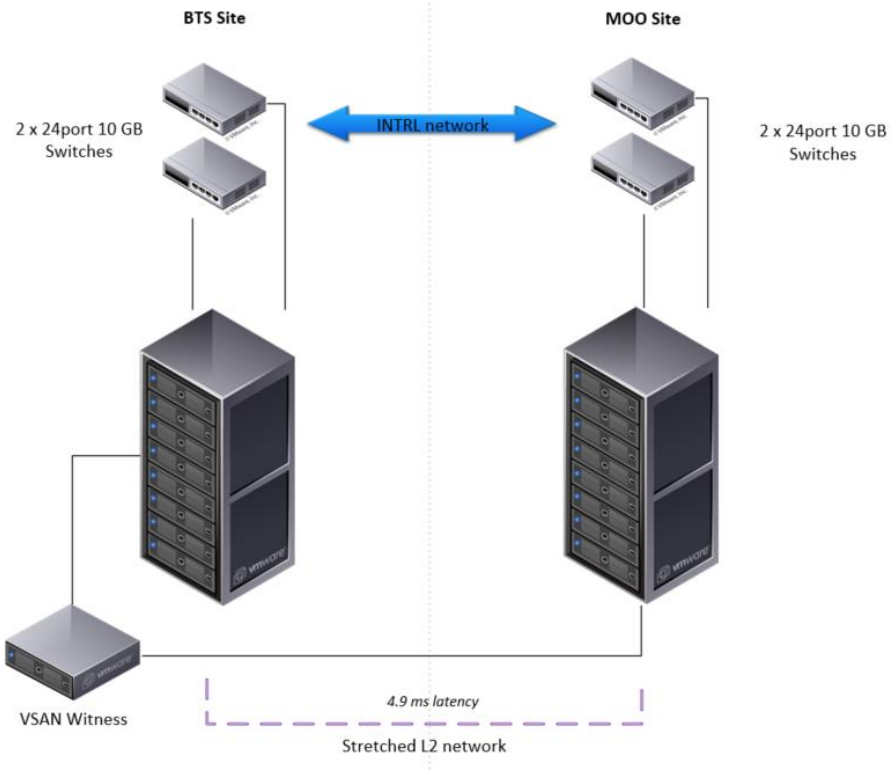

FIGURE 2 - PHYSICAL NETWORK

4.1 Physical layer

Stretched L2 network will be used to connect the two sites.

Each site will be equipped with two 24 port 10G switches.

Justification

Two physical switches have been selected to provide redundancy. 24 ports per switch will be sufficient to accommodate current connections as well as offer room for expansion in the future. [R01, R04]

Design qualities impacted

Availability

4.2 Logical layer

VMware NSX 6.2 will be used to create and operate the logical layer.

4.2.1 NSX Components

NSX infrastructure will consist of one NSX manager and three NSX controllers which will provide redundancy and prevent split brain scenarios.

VM name	Description	RAM	CPU	Disk
btsmgnm001	NSX manager	16	4	60
btsmgnc001	NSX controller 1	4	4	60
btsmgnc002	NSX controller 2	4	4	60

moomgnc001	NSX controller 3	4	4	60

Justification

NSX provides additional security by utilizing features such as Distributed Firewall. It also provides a centralized management and future scalability. [R04]

Design qualities impacted

Security, Manageability

4.2.2 Distributed switch configuration

Traffic will be separated to following port groups:

- Management traffic
- vMotion
- VSAN
- VM traffic

Four uplinks will be available on each host.

Portgroup	NIC
Management Traffic	Vmnic0 (active)
	Vmnic2 (passive)
vMotion	Vmnic0 (active)
	Vmnic2 (passive)
VSAN	Vmnic1 (active)
	Vmnic3 (active)
VM traffic	Vmnic2 (active)
	Vmnic0 (passive)

Justification

Uplinks will be assigned to individual port groups to maximize performance and provide failover capabilities. [R02, R03]

Design qualities impacted

Performance, Recoverability

4.2.3 VLANs

Network traffic will be separated into VLANs as follows:

Management	100
vMotion	200
VSAN	300
	400
VM quests	500

VM traffic will be further segmented using Distributed Firewall provided by NSX.

Justification

VLANs will be used to segregate traffic. Two VLANs will be used by VSAN in order to allow for two uplinks to be used in active mode to increase performance. [R02]

Design qualities impacted

Security

4.2.4 DNS and DHCP

DNS and DHCP server roles will be configured on the AD servers.

All hosts and VMs will have A records and PTR records created.

IP reservations will be created for all host.

DHCP scope will be created for host IP range, this will be later used for Auto Deploy.

4.3 Storage design

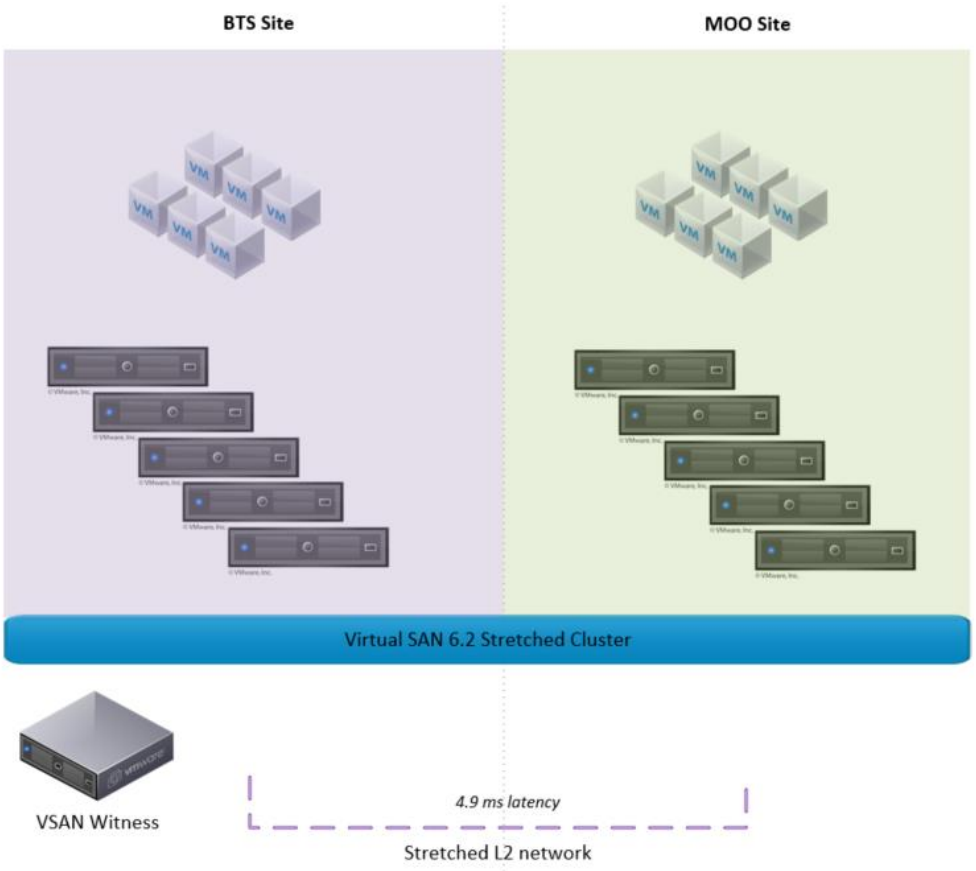

FIGURE 3 - VSAN STRETCHED CLUSTER

4.3.1 VSAN

Storage will be provided by implementing Virtual SAN 6.2 Stretched Cluster using all flash SSDs in order to provide maximum performance.

Cluster will be installed in 5+5+1 configuration, with Witness ESXi standalone host which will be located in BTS datacenter but not part of the cluster.

4.3.1.1 VSAN sizing

Raw Storage Requirements (without FTT)	10 680 GB
Cache required (10%)	1 068 GB
FTT	1 (Raid-5/6, 1.33x Overhead)

Raw Storage Requirements (with FTT)	14 204.40 GB
Required slack space	30%
Raw Formatted Storage Capacity	20 292 GB
On disk format overhead	1%
Raw Unformatted Storage Capacity	20 496.70 GB
# of Hosts	10
Cache required per host	106.8 GB
Capacity required per host	2 049.70 GB

Based on the calculation, each host will be equipped with one 120 GB SSD for Cache and two 1.6 TB SSDs for Capacity. Configuration will allow for keeping the utilization under the defined threshold of 80% and allow for future growth.

	Per Host	Per Cluster
Cache	120 GB	1.2 TB
Capacity	3.6 TB	32 TB

Two Fault domains will be available:

BTS	MOO
Preferred	Secondary
BTS Hosts	MOO Hosts

Standalone ESXi host will be deployed and act as a witness to the two fault domains.

Justification

VSAN stretched cluster has been selected to provide the shared storage foundation for the infrastructure. [R01]

All flash storage was selected to maximize performance. [R02]

Version 6.2 has been selected due to enhanced read locality which will help to avoid operations across sites. [R02]

Design qualities impacted

Performance, Manageability

4.4 Virtual Machine design

Infrastructure will be able to host both Windows based and Linux based guest operating systems.

VM templates will be utilized to allow for quick deployments and to ensure consistent configuration across the environment.

Justification

Infrastructure provides quick scale out possibilities. [R04, R05]

Design qualities impacted

Manageability

4.4.1 Naming convention

In order to quickly identify VMs, following naming convention has been put in place:

(location)+(application)+(server_type)+(ID)

In example, HumanityLink Web Server running in Bratislava (Earth) would be named *btshlws001*.

All codes are listed in the table below:

Location	
Bratislava	BTS
Moon	MOO
Application	
HumanityLink	HL
Generic Application	AP
Management	MG
Server Type	
Web Server	WS
Application Server	AP
Database Server	DB
Domain Controller	DC
vCenter Server	VC
Veeam Server	VE
NSX Manager	NM
NSX Controller	NC

Justification

Naming convention will allow for quick identification of VM in case of issues and improve overall management of the environment. [R05]

Design qualities impacted

Manageability

4.4.2 HumanityLink Application

HumanityLink application will be running on 7 virtual machines which will be distributed across the two sites as follows. This will help us to provide increased redundancy and ensure that application will be available even in case of host or OS failures.

VM	Description	RAM (GB)	CPU	Disk (GB)
btshlws001	HL Web Server 1	16	8	40
btshlws002	HL Web Server 2	16	8	40
moohlws001	HL Web Server 3	16	8	40
btshldb001	HL Database Server 1	128	12	500
moohldb001	HL Database Server 2	128	12	500
btshlap001	HL Application Server 1	64	22	50
moohlap001	HL Application Server 2	64	22	50

To further increase the availability, database will be hosted on two Windows 2012 R2 servers, running in SQL 2012 Always On mode.

Justification

HumanityLink servers have been sized to allow for high performance. Database has been clustered in order to provide additional level of redundancy to increase the overall availability of the application. [R02, R03]

Design qualities impacted

Availability, Performance

4.5 Management Infrastructure

4.5.1 Active Directory

Active Directory will be running on two Windows 2012 R2 VMs:

VM	Description	RAM (GB)	CPU	Disk (GB)
btsmgdc001	Domain controller 1	4	1	50
moomgdc001	Domain controller 2	4	1	50

AD servers will be separated to provide redundancy with btsmgdc001 being the Primary DC.

DNS and DHCP roles will be enabled on both servers.

4.5.2 vCenter server

vCenter server will be running on a Windows 2012 R2 VM:

VM	Description	RAM (GB)	CPU	Disk (GB)
btsmgvc001	vCenter server	16	4	120

vCenter will be installed with the embedded PostgreSQL database as well as the embedded Platform Services Controller.

Following components will be installed on the vCenter server:

- Webclient
- PowerCLI
- Putty
- AutoDeploy
- Update Manager

Justification

Windows based vCenter has been selected over VCSA due to the intention to install additional tools on one server. As the plan was to use Auto Deploy and Update Manager, additional VMs would have to be deployed. [R05]

Design qualities impacted

Manageability

4.5.2.1 Auto Deploy

Auto Deploy will be configured for host installation by executing following steps:

- Adding installation binary to software depot
- Creating DeployRule to install the image on all hosts in defined IP range, add them to the cluster and apply host profile
- Adding the new rule to the active ruleset

Justification

Auto Deploy will help us to quickly deploy all of the hosts by simply PXE booting them. It will also allow for quick scale out in the future in a consistent manner. [R04, R05]

Design qualities impacted

Manageability

4.5.2.2 Update Manager

Even though the patch release may be limited due to the ongoing apocalypse, we still need to ensure that our infrastructure will be secure and if any patches are released.

Update manager will be used to manage patches and extensions for ESXi hosts.

Justification

Update manager provides an easy to manage framework for patch remediation and also provides a compliancy overview. [R04]

Design qualities impacted

Security

4.6 Backups

Veeam Backup & Replication 9 will be installed on a Windows 2012 R2 server with following specification:

VM	Description	RAM (GB)	CPU	Disk (GB)
btsmgve001	Veeam Server	4	1	5000

Name	VM list	Schedule
HumanityLink	All HL VMs	Every day – 00:00
Production DB	All remaining production database VMs	Every day – after HL
Production AP	All remaining production application VMs	Every day – after Prod DB
Production WS	All remaining production web server VMs	Every day – after Prod AP
Management	All management VMs	Every day – after Prod WS

Full Backup will be scheduled to run on Sunday, Reverse Incremental Backups will be running every day.

Retention policy will be set to 1 week.

Justification

Veeam software has been selected due to its reliability and simplicity. Features, such as Instant VM Recovery, will be very useful to restore VMs back to production and minimize downtime. [R05]

Design qualities impacted

Recovery

5 References

Arrasjid, J. Y., Gabryjelski, M., & McCain, C. (2016). *IT Architect: Foundation in the Art of Infrastructure Design.*

Arrasjid, J. Y., Lin, B., & Khalil, M. (2013). *VCDX Boot Camp.*

HP. (2016). *HPE ProLiant DL380 Generation9 (Gen9) Quick Specs*. Retrieved from www8.hp.com: http://www8.hp.com/h20195/v2/getpdf.aspx/c04346247.pdf

McCarty, J. (2016). *VMware® Virtual SAN™ 6.2 Stretched Cluster*. Retrieved from www.vmware.com: http://www.vmware.com/files/pdf/products/vsan/VMware-Virtual-SAN-6.2-Stretched-Cluster-Guide.pdf

Nicholson, John;. (2016). *VMware® Virtual SAN™ 6.2 Design and Sizing Guide*. Retrieved from www.vmware.com: http://www.vmware.com/files/pdf/products/vsan/virtual-san-6.2-design-and-sizing-guide.pdf

Veeam. (2016). *VEEAM BACKUP & REPLICATION 9.0 U1 RELEASE NOTES*. Retrieved from www.veeam.com: https://www.veeam.com/pdf/release_notes/veeam_backup_9_0_release_notes_en.pdf

VMware. (2015). *Knowledge Base*. Retrieved from kb.vmware.com: https://kb.vmware.com/

VMware. (2016). *Reference Design: VMware® NSX for vSphere (NSX) Network Virtualization Design Guide*. Retrieved from www.vmware.com: https://www.vmware.com/files/pdf/products/nsx/vmw-nsx-network-virtualization-design-guide.pdf

6 Appendix A: VM list

#	VM	Description	RAM (GB)	CPU	Disk (GB)
1	btshlws001	HL Web Server 1	16	8	40
2	btshlws002	HL Web Server 2	16	8	40
3	moohlws001	HL Web Server 3	16	8	40
4	btshldb001	HL Database Server 1	128	12	500
5	moohldb001	HL Database Server 2	128	12	500
6	btshlap001	HL Application Server 1	64	22	50
7	moohlap001	HL Application Server 2	64	22	50
8	btsapws001	Web server 1	2	2	40
9	btsapws002	Web server 2	2	2	40
10	btsapws003	Web server 3	2	2	40
11	btsapws004	Web server 4	2	2	40
12	btsapws005	Web server 5	2	2	40
13	btsapws006	Web server 6	2	2	40
14	btsapws007	Web server 7	2	2	40
15	btsapws008	Web server 8	2	2	40
16	btsapws009	Web server 9	2	2	40
17	btsapws010	Web server 10	2	2	40
18	btsapws011	Web server 11	2	2	40
19	btsapws012	Web server 12	2	2	40
20	btsapws013	Web server 13	2	2	40
21	btsapws014	Web server 14	2	2	40
22	btsapws015	Web server 15	2	2	40
23	btsapws016	Web server 16	2	2	40
24	btsapws017	Web server 17	2	2	40
25	btsapws018	Web server 18	2	2	40
26	btsapws019	Web server 19	2	2	40
27	btsapws020	Web server 20	2	2	40
28	mooapws021	Web server 21	2	2	40
29	mooapws022	Web server 22	2	2	40
30	mooapws023	Web server 23	2	2	40
31	mooapws024	Web server 24	2	2	40
32	mooapws025	Web server 25	2	2	40
33	btsapdb001	Database server 1	16	4	500
34	btsapdb002	Database server 2	16	4	500
35	btsapdb003	Database server 3	16	4	500
36	btsapdb004	Database server 4	16	4	500
37	mooapdb005	Database server 5	16	4	500
38	btsapap001	Application server 1	8	4	50
39	btsapap002	Application server 2	8	4	50
40	btsapap003	Application server 3	8	4	50
41	btsapap004	Application server 4	8	4	50
42	btsapap005	Application server 5	8	4	50
43	btsapap006	Application server 6	8	4	50
44	btsapap007	Application server 7	8	4	50
45	btsapap008	Application server 8	8	4	50
46	mooapap009	Application server 9	8	4	50
47	mooapap010	Application server 10	8	4	50
48	btsmgdc001	AD 1	4	1	50
49	moomgdc001	AD 2	4	1	50
50	btsmgvc001	vCenter server	16	4	120
51	btsmgnm001	NSX manager	16	4	60
52	btsmgnc001	NSX controller 1	4	4	60
53	btsmgnc002	NSX controller 2	4	4	60
54	moomgnc001	NSX controller 3	4	4	60

55	btsmgve001	Veeam Server	4	1	5000
			698	225	10680