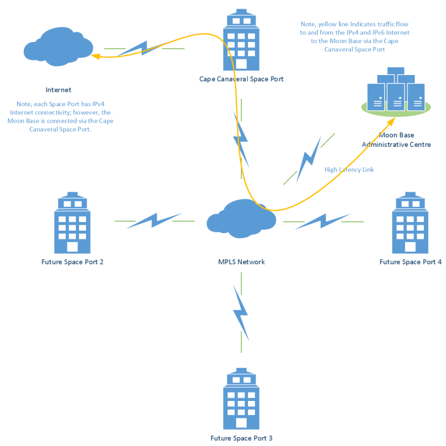
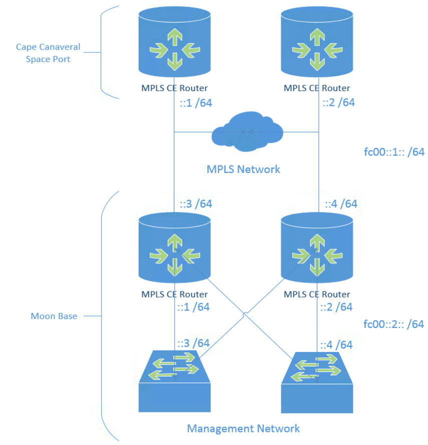
JOEL GIBSON





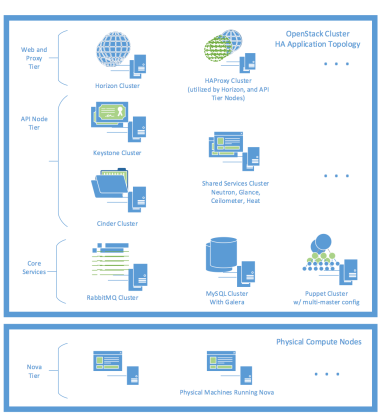
Design Proposal - Challenge 2

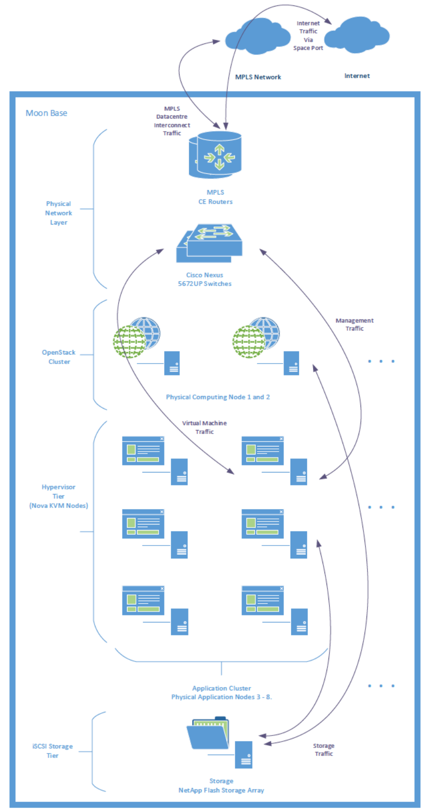
Prepared for: VirtualDesignMaster.com

Prepared by: Joel Gibson

July 21, 2014

Proposal number: Challenge 2





EXECUTIVE SUMMARY

Objective

The objective of this design is to support the mission critical infrastructure on the moon base. Based on the description provided, the moon base will be used to temporarily house humans until the Mars colony has been completed.

Assumptions

It is assumed that the moon base is not a manufacturing depot that produces space ships, as was the case on Earth. The documentation provided would indicate that the moon base is a piece of supporting infrastructure; therefore, for the purpose of this design, it is assumed that the main purpose of these systems will be of an operational nature to ensure availability of administrative, security, logistics, and other services.

Goals

This solution should be reliable, easily deployable, and be conscious of power, cooling, and space limitations. Downtime is to be minimized.

Constraints

- Due to severe power, cooling, and space limitations, the infrastructure must fit into a 21U rack space.
- The solution must be built utilizing components from one or more of the following vendors: VMware, Cisco, NetApp, Synology, RedHat, and Puppet.
- The design must utilize IPv6 only.

Solution

With the above goals in mind, the solution has been designed around the following key elements:

- All supporting elements must be highly available.
- The physical and logical infrastructure must be able to scale easily, and quickly.
- The application framework must be able to scale automatically.

The physical infrastructure has been kept simple, and relies heavily on converged systems. It is easy to rack, stack, and cable this design.

DESIGN OVERVIEW

Background

The Cape Canaveral Space Port is the first, of at least four, critical production facilities. The facilities in the Netherlands, Australia, and New Zealand will be ready soon. In addition, there is a base located on the moon which will be used as a temporary colony until the human race can be moved to Mars.

The infrastructure design for the moon base has been configured in a highly available manner. It is the goal of this design to eliminate any potential single points of failure, such that it is able to support the mission critical systems.

The physical design of the infrastructure has been kept relatively simple, so that scarce resources can be best utilized. Due to severe power, cooling, and space limitations, the infrastructure must fit into a 21U rack space. In addition, only IPv6 infrastructure is available on the moon base; therefore, all supporting infrastructure must be compatible.

Objective

The objective of this design is to support the mission critical infrastructure on the moon base. Based on the description provided, the moon base will be used to temporarily house humans until the Mars colony has been completed.

Constraints

- Due to severe power, cooling, and space limitations, the infrastructure must fit into a 21U rack space.
- The solution must be built utilizing components from one or more of the following vendors: VMware, Cisco, NetApp, Synology, RedHat, and Puppet. (based on the challenge framework provided)
- The design must utilize IPv6 only.

Assumptions

It is assumed that the moon base is not a manufacturing depot that produces space ships, as was the case on Earth. The documentation provided would indicate that the moon base is a piece of supporting infrastructure; therefore, for the purpose of this design, it is assumed that the main purpose of these systems will be of an operational nature to ensure availability of administrative, security, logistics, and other services. It is also assumed that while the design must stick to the vendors noted above, that VMware is not a requirement.

Goals

This solution should be reliable, easily deployable, and be conscious of power, cooling, and space limitations. Downtime is to be minimized.

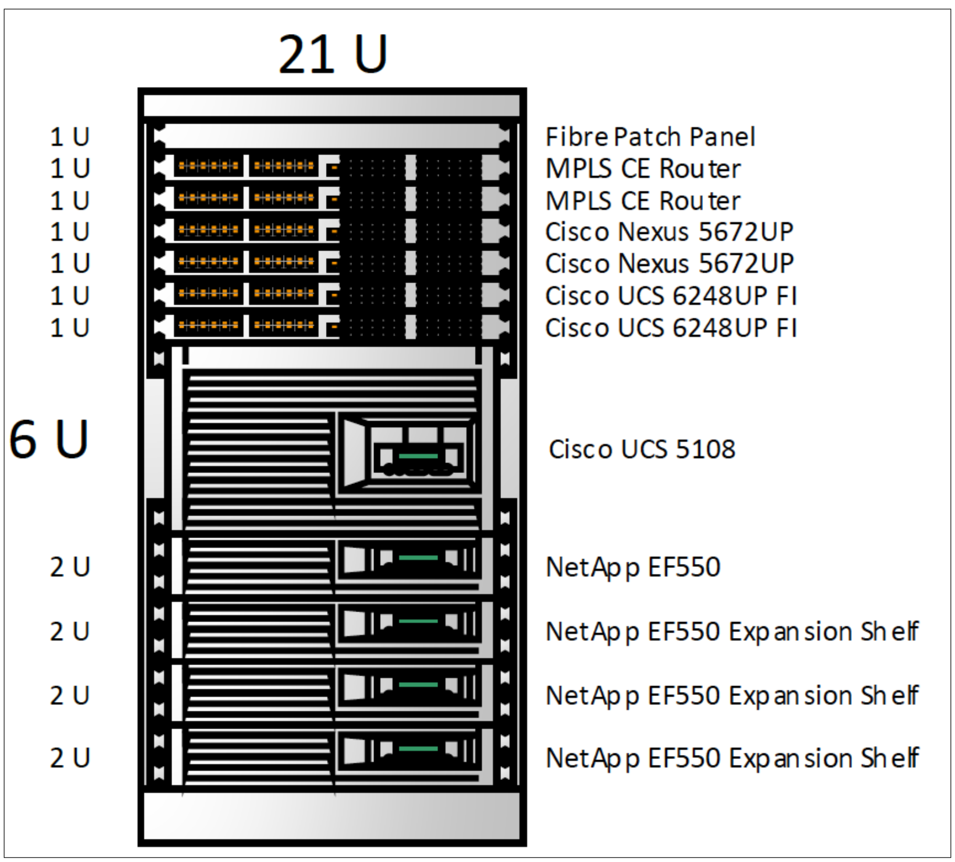
PHYSICAL DESIGN

Infrastructure

As noted earlier, the physical infrastructure has been kept simple, and relies heavily on converged systems. It is easy to rack and stack this design.

The physical systems will reside in a single 21U rack space, supplied by two UPS systems. The rack will contain the required network gear to provide moon-spaceport MPLS connectivity, top-of-rack network, compute, and storage.

Figure 1 - Physical Rack Layout



Network

The Cisco Nexus 5672UP was chosen as the top-of-rack switch, which provides ample bandwidth and ports for the existing infrastructure, along with room to grow. It also supports network overlay technology, such as VxLAN should the future need arise.

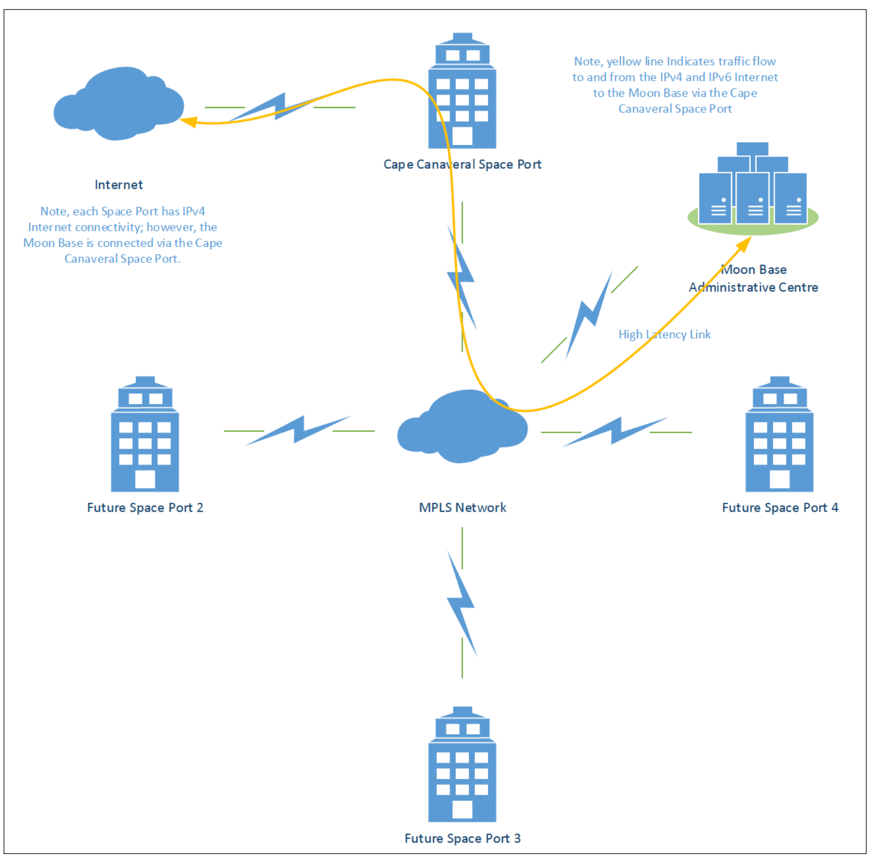
Moon-to-spaceport connectivity will be established using a specialized high latency, high bandwidth connection to the MPLS network via 3rd party Earth stations.

The design will utilize IPv6 only.

Internet

No direct connectivity to the Internet will be available from the moon base. If traffic to the IPv6 Internet is required, traffic will be routed from the moon base to the Internet via the Cape Canaveral Space Port. If a connection to the IPv4 Internet is required from the moon base, and future provisions in the design requirements are allowed to do so, a tunnel will be created between two compatible routers to encapsulate IPv4 traffic within an IPv6 packet.

Figure 2 - High Level Overview of Inter-Facility Connectivity



Compute

The Cisco UCS 5108 blade-system was chosen for scalable compute, instead of the C460 M4 rack mount servers that are utilized in the space ports. The main reason for this was to *increase density* of the compute nodes. The C460 M4s contained several disk trays which were not being utilized at the space ports, and were seen as utilizing excessive space, based on the amount of compute capability per rack unit (RU).

Based on the estimates noted in tables one and two below, it is believed that by utilizing the UCS 5108 blade system the following efficiencies can be gained:

- 50% reduction in physical rack space
- roughly 87% more compute
- approximately 3-5% gain in power and cooling efficiency based on the compute density

In addition to the efficiencies gained through the increased density, this design enables a certain amount of flexibility to scale up/down the compute infrastructure. In the event that power availability becomes constrained, one or more blades can be shutdown while maintaining core management and application availability. If more compute resources are required, and the power is available, scaling up can be accomplished by replacing the CPUs, and/or adding memory.

These statements, and the physical characteristics described are estimates, and for learning purposes only.

Table 1 - Estimation of Physical Characteristics (Cisco C460 M4 Rack Mount)

							75% Utilization		
	Description	CPU	Clock Rate	Cores	Memory	QTY	Power Consumption (Watts)	Cooling (BTU)	Rack Space
C460 M4	Rack Mount Server	4x E7-4809	1.9	24	128	4	532	1814	4
Total M4			182.4	96	512	4	2128	7256	16

Table 2 - Estimation of Physical Characteristics (Cisco 5108 with B230 M2 Blades)

							75% Utilization		
	Description	CPU	Clock Rate	Cores	Memory	QTY	Power Consumption (Watts)	Cooling (BTU)	Rack Space
B230 M2	Cisco 5108 with Blade Server	2x EL-8867L	2.13	20	128	8	410	1397	0.75
6248UP FI	Fabric Interconnect					2	256	875	1
Subtotal M2			340.8	160	1024	8	3280	11176	6
Subtotal FI							512	1750	2
Total							3792	12926	8

Storage

The NetApp EF550 flash storage array was chosen, instead of the E2624 which is utilized in the space ports. The main reason for this was to *increase power and cooling efficiency.*

Based on the estimates noted in tables three and four below, it is believed that by utilizing the flash storage array the following efficiencies can be gained:

- approximate reduction in power and cooling requirements by 33-34%
- approximate increase in available raw disk capacity by 33%

These statements, and the physical characteristics described are estimates, and for learning purposes only.

In addition, some of the savings in power and cooling could be utilized to maintain a higher density of compute nodes for longer periods.

Table 3 - Estimation of Physical Characteristics (NetApp E2624 Storage Array Config)

			75% Utilization			
	Description	QTY	Power Consumption (Max Watts)	Cooling (Max BTU)	Rack Space	Raw Capacity (TB)
E2624	NetApp Storage Array Populated with 120 x 1.2 TB SAS 10k Drives	2	482	1644	2	
Expansion		2	371	1267	2	
Total			**1706**	**5822**	**8**	**144**

Table 4 - Estimation of Physical Characteristics (NetApp EF550 Flash Storage Array Config)

			75% Utilization			
	Description	QTY	Power Consumption (Max Watts)	Cooling (Max BTU)	Rack Space	Raw Capacity (TB)
EF550	NetApp Flash Array Populated with 120 x 1.6 TB SSD Drives	1	498	1630	2	38
Expansion		3	216	738	2	
Total			**1146**	**3844**	**8**	**192**

Physical Configuration Maximums

The following configuration maximums are based on the physical constraint of 21U of available rack space.

- Cisco UCS 5108 Chassis (1)
- Cisco B230 M2 Blades (8)
  - Intel CPUs (2)
  - Memory (512 GB)
  - Local Disk Space SSD (800 GB)
- Cisco 6248UP Fabric Interconnects
  - 10 GB Ports (48 each)
  - Expansion Ports (2)
- NetApp EF550 Flash Storage Array
  - SSD Drives (120)
  - Size of SSD Disk (1.6 TB)
  - Raw Capacity (192 TB)
- Cisco 5672UP Switch
  - 10 GB Ports (48 each)

Risks

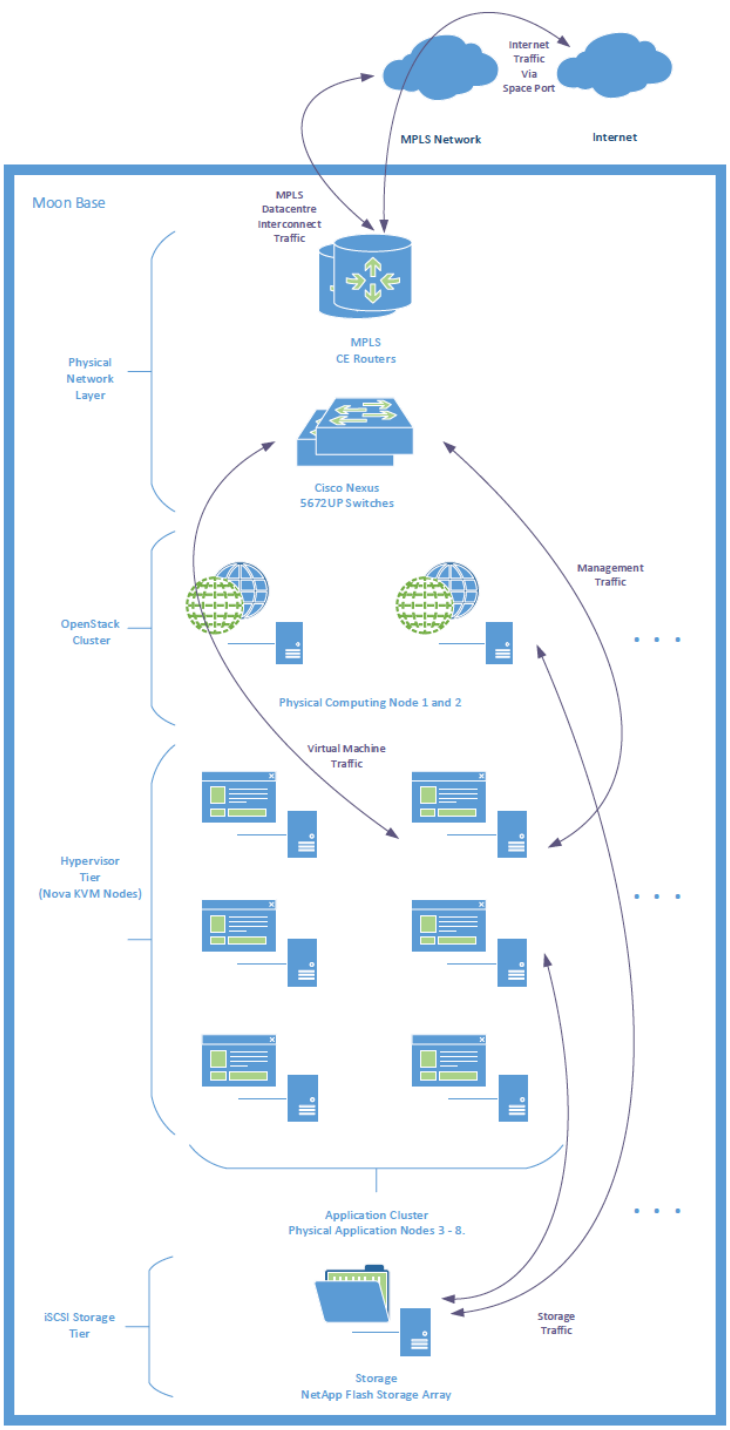
Due to the physical power, space, and cooling constraints, there is no multi-site, multi-room, multi-rack redundancy. The solution is highly available, but within a single rack.

If the data centre is physically compromised, the mission critical services provided by the moon base infrastructure could be at significant risk.

Assumptions

It is assumed that the Cape Canaveral Space Port has been designed and built to Tier 3 or above standards according to the Uptime Institute. In addition, there is enough physical space, and power and cooling capacity available to scale the physical equipment, if required.

Figure 3 - High Level Overview of Virtualization Infrastructure



LOGICAL DESIGN

Overview

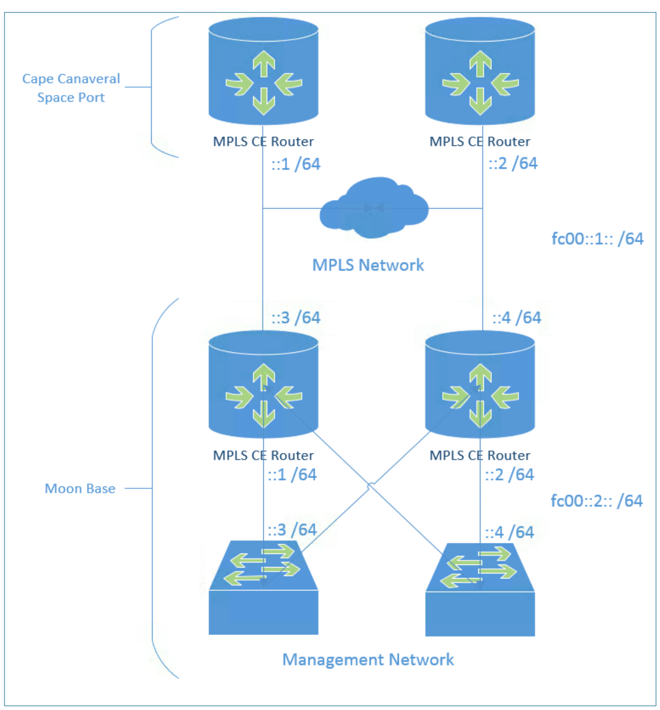
The premise for this design is based on high availability and orchestration. The purpose of this is to protect the critical components (i.e. application workload, and data), and ensure that the infrastructure can scale quickly and efficiently within the physical power, space, and cooling constraints.

The virtualization and private cloud infrastructure will be based on RedHat supported OpenStack framework, and will include a highly available management cluster. The virtual servers supporting the operational applications will reside on a separate cluster of nodes managed by OpenStack.

Network

For the IP network, IPv6 will be utilized. The network addresses will be based on the Unique Local Address (ULA) range fc00::/7, which is not Internet routable. The moon base infrastructure will utilize the following /64 subnets as shown in the diagram below. In addition, most addresses (with the exception of core physical equipment) will be assigned using DHCPv6.

Figure 4 - High Level Overview of IPv6 Network Addressing



OpenStack Cluster

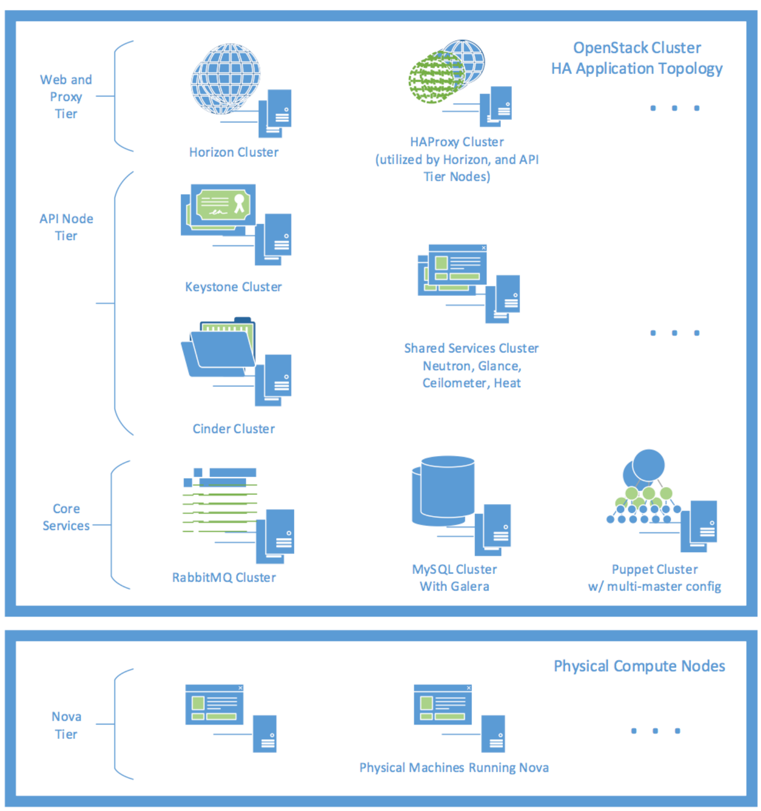
The management cluster will reside on two physical nodes (UCS Blades 1 and 2) and is based on a highly available application topology.

The chosen operating system for the bare-metal install, as well as the supporting virtual servers is RedHat Enterprise Linux 6.x with OpenStack. The bare-metal install will be installed directly onto the two SSD drives installed in each of the UCS B-series blades. The two SSDs will be configured using RAID 1. By utilizing the drive capacity in the blades for the bare-metal instal (versus a boot-on-SAN configuration), valuable space on the flash storage array is conserved. The management components within the cluster will reside on virtual machines, running on top of the KVM hypervisor.

The underlying core services supporting the OpenStack cluster are messaging (RabbitMQ), databases (MySQL with Galera), and orchestration (Puppet with a multi-master config). HAProxy has been chosen to act as a highly available load-balancer for the web tier (Horizon), as well as the OpenStack API nodes.

Puppet will be used to orchestrate and maintain the state and consistency of the OpenStack cluster. It can easily integrate with, and maintain the desired state of the OpenStack projects, HAProxy, the supporting core services.

Figure 5- High Level Overview of OpenStack Cluster



Application Cluster

The chosen operating system for the bare-metal install (Cisco UCS Blades 3 - 8) is RedHat Enterprise Linux 6.x with OpenStack. The type two hypervisor to be used is KVM, and will be managed by OpenStack Nova.

The virtual servers which form the application framework will utilize RedHat Enterprise Linux 6.x for their guest operating system.

As part of this design, it was mentioned earlier that the application framework would have the ability to scale automatically. This requirement will be satisfied utilizing OpenStack Heat and Ceilometer to spin-up/down instances based on workload, and Puppet to set the desired state of the application configuration.

It was decided that a combination approach to orchestration would be best suited to this design to enable both automatic scaling, while maintaining desired state.

Storage

The storage utilized by the physical nodes will be presented by the NetApp flash storage array in the form of iSCSI LUNs. The management and application servers will reside on separate LUNs, as well as any additional special configuration required for the highly available databases.

Logical Configuration Maximums

The following configuration maximums are based on the physical equipment.

- Storage (~188,880 GB based on Raid 6)
- Virtual Machines (based on Table 5)
  - *Management Cluster*
    - m1.tiny (256)
    - m1.small (448)
    - m1.medium (224)
    - m1.large (112)
    - m1.xlarge (56)

- *Application Cluster*
  - m1.tiny (1280 or 256 per blade)
  - m1.small (320 or 64 per blade)
  - m1.medium (224 or 32 per blade)
  - m1.large (160 or 16 per blade)
  - m1.xlarge (40 or 8 per blade)

  … or a combination of the totals above, depending on instance size.

Assumptions on VM Maximums

- Please note, the VM quantities noted above are estimates only, and are based on the assumption that the largest bottleneck is memory.
- It is also assumed that physical CPU cores will be oversubscribed, and that it is not desirable to do likewise with memory.
- These numbers do not factor in memory compression, swapping, deduplication, or other optimization methods.
- In addition, the calculations were based on the physical blade characteristics described in Table 2, and that the resources equivalent to at least one blade per cluster would be reserved for failover in the event of hardware failure.

Table 5 - OpenStack Default Flavors

Flavors	Memory (MB)	Disk (GB)	Ephemeral (GB)	VCPUs
m1.tiny	512	1	0	1
m1.small	2048	10	20	1
m1.medium	4096	10	40	3
m1.large	8192	10	80	4
m1.xlarge	16384	104	160	8

Source: http://docs.openstack.org/openstack-ops/content/flavors.html