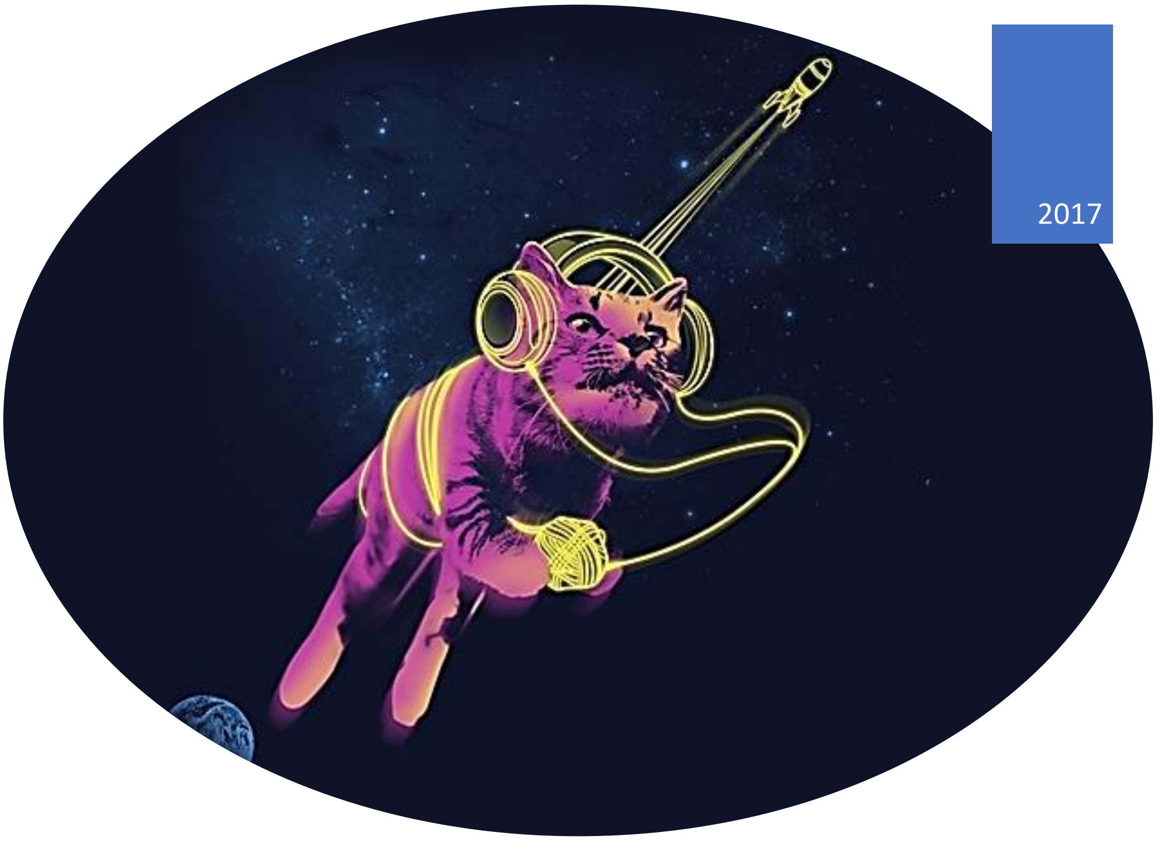


2017

Challenge 1

HIGH LEVEL DESIGN

KYLE JENNER

TERRAFORMING SPACE AGENCY

1 Contents

2 Contact Information & Document Control

The Primary contacts for questions and discussions regarding this proposal are:

2.1 Document Information

Title	**High Level Design**
Version	1.0
Author	Jenner, Kyle
Distribution Date	01/07/2017
Number of Pages (Excluding Cover)	23

3 Executive Summary

3.1 Document Purpose

The purpose of this document is to outline the key elements and design decisions which make up the proposed infrastructure design.

3.2 Project Overview

Following Earths effort to recolonize it is now at the stage to begin the terraforming effort, a special task force has been assembled to take on this important venture. The task force is made up of infrastructure, security and software architects, developers and scientists that is called Terraforming Space Agency hereinafter known as TSA.

TSA have been tasked to design a new version of the HumanityLink application, the application must be able to assist human scientist with their efforts in the terraforming process. The design must include a 3-site architecture and must be scalable and resilient.

The known billionaire investor has agreed to fully back TSA in their effort with no constraints on budget.

3.3 Design Qualities

The following design qualities will be referenced.

Qualities	Ref	Example
Availability	DQA	System up time to deliver SLAs.
Manageability	DQM	Simplified management layer to reduce overall efficiency.
Performance	DQP	Ensure system performance to meet project requirements.
Recoverability	DQR	Ability to recover from a failure.
Security	DQS	Authorization and access to the system.

4 Requirements

Requirement	Ref
The proposed new solution must be a 3-site architecture.	REQ001
The infrastructure must support a new HumanityLink application.	REQ002
The solution must be scalable.	REQ003
The solution must be resilient as possible.	REQ004
New HumanityLink application must be able to support the army of deployed robots.	REQ005
All maintenance on robots must be performed whilst the robots are off shift.	REQ006
New HumanityLink application must be able to be replicated easily to any other infrastructure on Earth or elsewhere.	REQ007

5 Constraints

Constraint	Ref
Initial deployment is on Earth.	CON001
No RPOs and RTOs have been specified in the project requirements for critical systems.	CON002

6 Risks

Risk	Impact	Mitigation	Ref
No application sizing details or expected workload details.	High	The application and supporting infrastructure must be scalable.	RIS001
Hardware failure may occur in one or more locations.	High	Implement a recovery method for hardware within and including a single site.	RIS002
Application failures can occur.	High	Implement an auto scale out application.	RIS003
Faulty code updates to HumanLink application in production systems could affect Terraforming effort.	High	Application updated must be tested in an isolated environment before deployed into production.	RIS004

7 Assumptions

Assumption	Ref	Additional Information
AWS still exists in post-apocalyptic world.	ASU001	The billionaire investor also provided the funds to get AWS back up and running on the return to earth.
VMware on AWS is GA.	ASU002	VMware on AWS is GA.
VMware vendor support still exists.	ASU003	VMware vendor support exists in some shape or form as part of the VMware on AWS service.
VMware on AWS automatically deploys on multiple AZ within a region	ASU004	When VMware on AWS went GA, this information was released.

Internet comms between sites is managed by AWS.	ASU005	AWS will manage the comms link between sites and ensure performance and resiliency.
All current information known on VMware on AWS made it into GA.	ASU006	Details pre-GA made it into the GA product.
Human staff are required to access this infrastructure.	ASU007	Human staff such as scientists and developers must be able to access this infrastructure not just the application.
Connectivity between datacentres are still intake and are resilient.	ASU008	All connectivity between datacentres are still in place and resilient.
The Internet still exists.	ASU009	The internet still exists in this post-apocalyptic world.
Each terraforming site is capable of an internet connection.	ASU010	Any terraforming site can create an internet connection for devices on the site.
All zombies are gone	ASU011	All zombies have been wiped out and will not come back.
Robot devices have already been built.	ASU012	The robot army is ready to be deployed, the robot spec is not a part of this project.
Human scientists and engineers will be mobile and work from a terraforming site.	ASU013	Human scientist and engineers are required at each terraforming site.
Human scientists and engineers have internet connected devices.	ASU014	Human scientists and engineers have internet connected devices to work on such as tablets
Developers work from a remote location.	ASU015	Developers must work from a remote location connected to the internet.
Developers do not require a highly scalable solution	ASU016	Develops do not require the same auto scale function as production.
Assumed acceptable up time is 99.9%	ASU017	No specific uptime has been stipulated.
Licenses are not restricted	ASU018	Quantities of licenses required are not restricted.
3 site architecture will need to be configured as an active site and DR.	ASU019	3 site architecture will need to be configured as an active site but also be capable of handling another site failure.
No limit on CAPEX or OPEX costs.	ASU020	No CAPEX or OPEX limits have been specified.

8 Evaluation and Measurement

The proposed solution is a “green field” solution, no existing workloads will be migrated to this lab environment.

9 Proposed Solution

9.1 Conceptual Design

The following conceptual design will be a very high level view on what the solution will look like after completion. The conceptual design does not include any sizing figure, vendors or product specifications.

9.1.1 Datacentre Conceptual Design

The design includes 3 regions for resiliency and availability. Each region is located on different continents on Earth so the terraforming process is not disrupted, a 'follow the sun' premise. Each region will comprise of 3 availability zones for further resilience.

9.1.2 Compute Conceptual Design

An auto scaling hypervisor solution has been proposed for each environment.

9.1.3 Management Conceptual Design

The solution provides all 3-site infrastructure through a single plane of glass whilst providing a portable application deployment.

9.1.4 Storage Conceptual Design

Scalable hyper converged storage has been selected for this design to provide the scale and agility.

9.1.5 Network and Security Conceptual Design

Each site will share the same virtual network constructs and security configurations.

9.1.6 Application Conceptual Design

The new HumanLink application will be developed on separate infrastructure as production to prevent any updates affecting the terraform effort on Earth.

The new application will enable the robot army to collect data and report back to a centralised database. Critical sensors at the terraforming sites that record information such as O2 levels are also uploaded to the same database once connected to the internet, they connect and upload their data once an internet connected robot passes by.

The application processes this information and through machine learning the application can predict robot patterns and manage shifts to make the work more efficient. Scientists also use this application to pull the information gathered for research while engineers use this application to be able to update robots during a predicted shift schedule.

9.1.7 Operational Conceptual Design

Monitoring of all infrastructure is included to predict and action when the infrastructure demand dictates.

9.1.8 Availability Conceptual Design

Each site will be configured in a high availability way whilst a site to site replication solution has been proposed to cover for a full site failure.

9.1.9 End User Conceptual Design

Scientists and Engineers will connect to central collaboration and admin tools using a remote tablet device, the HumanLink application will be accessed over the internet. Developers will connect to a dedicated developer environment.

9.2 Logical Design

The following logical design will take the conceptual design and put together a solution that will use technologies to meet the requirements. Each section will list a design decision and link it to the requirement for reference. Each design decision has considered the risks and constraints highlighted above.

9.2.1 Logical Design Overview

The premise of this design is to provide an agile and scalable 'cloud native' application whilst also providing the required infrastructure for the scientists, engineers and developers that are critical to the terraforming effort on Earth.

The application will be able to scale up as workload dictates and scale back down when not. This solution also provides a saleable server infrastructure in case the terraforming process expands beyond current workload predictions.

The server infrastructure will be configured across 3 AWS regions and 3 availability zones within those regions. The same infrastructure will be used by developers to test any new releases and deploy into production.

The server infrastructure will be deployed on VMware on AWS service across 3 regions. Native AWS tools will also be leveraged for the HumanLink application to run in whilst the VMware infrastructure will be used for development.

The production HumanLink application will be a containerized application using Docker running on Amazon EC2 Container Service (ECS) that will be auto scaled within a region using Amazon CloudWatch. Terraform (the product) will be used to script the creation of the infrastructure to make it portable across any provider and duplicate to other regions consistently. For further details see section 9.2.7 of the logical design.

VMware Horizon will be deployed for scientists, engineers and developers to work from a centralized secure environment. Scientists and engineers will use this infrastructure for collaboration tools, admin task and shared data access.

Each site will be configured with vCenter and NSX providing easy manageability with network and security consistency. VMware vSAN will be used for the storage at each site and VMware vRealize Operations will provide monitoring and remediation of the infrastructure.

Rubrik has been selected to provide backup and replication between sites with the view to leverage the RESTful APIs to orchestrate any required site failover to another location.

9.2.2 Datacentre Logical Design

VMware on AWS has been chosen as the main infrastructure for scientists, engineers and developers due to the elastic scaling abilities.

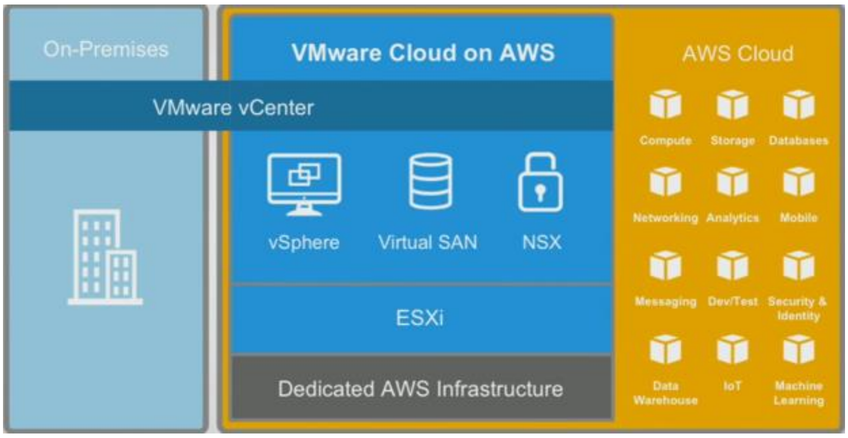
DDN001	Decision – VMware on AWS will be deployed.
	Justification – AWS provides auto scaling of ESXi through elastic DRS.
	Type – DQA, DQM, DQP
	Impact – vSphere hypervisor must be used.
	Associated Risk – The HumanLink application must not be deployed on ESXi as it is required to be portable.
	Risk Mitigation – The HumanLink application will be developed in a 'Cloud Native' manner removing the requirement of any particular hypervisor.
	Reference – REQ001, REQ003, REQ004 CON001 ASU001, ASU002, ASU003, ASU006

VMware infrastructure will be deployed across 3 AWS regions with at least 3 availability zones within the region.

DDN002	Decision – 3 AWS regions we selected in different continents.
	Justification – To provide around the clock access to the new HumanityLink application. Regions = US East, Ireland, Sydney. Each region has 3 AZs.
	Type – DQA, DQM, DQP, DQR
	Impact – Connectivity between sites is global.
	Associated Risk – Extra latency accessing the different locations.
	Risk Mitigation – Configure Amazon Route 53s Latency Based Routing.
	Reference – REQ001, REQ003, REQ004 ASU001, ASU002, ASU005, ASU007, ASU008, ASU009

Connectivity between AWS Virtual Private Cloud (VPC) will be enabled to allow the use of AWS native features, AWS native features will be leveraged for the new HumanityLink application.

Figure 1 – VMware on AWS with AWS native services



DDN003	Decision – Connectivity between VMware on AWS and native Amazon services will be enabled.
	Justification – Connectivity between VPC will be enabled to provide developers access to Amazon DynamoDB for the collected data using VMware Cloud Endpoint.
	Type – DQA, DQP
	Impact – Developers running on a vSphere environment can still access the required data from the database.
	Associated Risk – Developers accessing live data could affect production.
	Risk Mitigation – Backups of the database will be required.
	Reference – REQ003, REQ004, REQ005, REQ006 ASU001, ASU002, ASU009, ASU010, ASU013, ASU014

9.2.3 Compute Logical Design

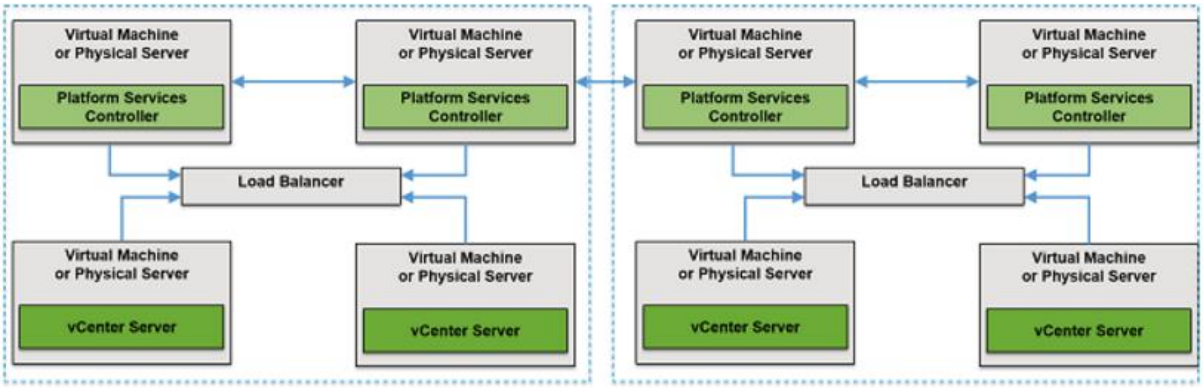
Hypervisor for the server workload will be vSphere ESXi.

DDN004	Decision – Hypervisor for server workloads will be ESXi.
	Justification – VMware Horizon required vCenter and ESXi.
	Type - DQM
	Impact – Workloads on ESXi can't migrate to other hypervisors without conversion.
	Associated Risk – The HumanLink application must not be deployed on ESXi as it is required to be portable.
	Risk Mitigation – The HumanLink application will be developed in a 'Cloud Native' manner removing the requirement of any particular hypervisor.
	Reference – REQ001 ASU001, ASU003, ASU016

9.2.4 Management Logical Design

Each site will be deployed with external Platform Service Controller (PSC) for all sites to share SSO domain.

Figure 2 – PSC topology


Virtual Machine or Physical Server
Platform Services Controller
Virtual Machine or Physical Server
Platform Services Controller
Load Balancer
Virtual Machine or Physical Server
vCenter Server
Virtual Machine or Physical Server
vCenter Server
Virtual Machine or Physical Server
Platform Services Controller
Virtual Machine or Physical Server
Platform Services Controller
Load Balancer
Virtual Machine or Physical Server
vCenter Server
Virtual Machine or Physical Server
vCenter Server

DDN005	Decision – Each site will have an external Platform Service Controller (PSC)
	Justification – PSC at each site will be configured in external mode to be able to share the same SSO domain. Each site will be configured as a SSO site. A single vCenter will be deployed at each site.
	Type – DQM, DQR, DQS
	Impact – Each site can be configured and managed from a single plane of glass using Enhanced Linked Mode. Workloads can be migrated between sites using Cross vCentre vMotion.
	Associated Risk – Each site's PSC represents a SPOF.
	Risk Mitigation – NSX Edge Load Balancer will be deployed at each site.
	Reference – REQ004 ASU005, ASU008, ASU009, ASU017

DDN006	Decision – Each site will be configured with a HA and DRS enabled cluster.
	Justification – To mitigate against a host failure and to distribute workloads across the cluster.
	Type – DQA, DQP, DQR
	Impact – Each cluster in each site will be configured for HA and DRS.
	Associated Risk – Failure to replaced failed ESXi hosts in the cluster could result in performance issues in the cluster.
	Risk Mitigation – Leverage AWS elastic capabilities and VMware's managed service to replace hardware.
	Reference – REQ004 ASU001, ASU002, ASU003, ASU006, ASU017

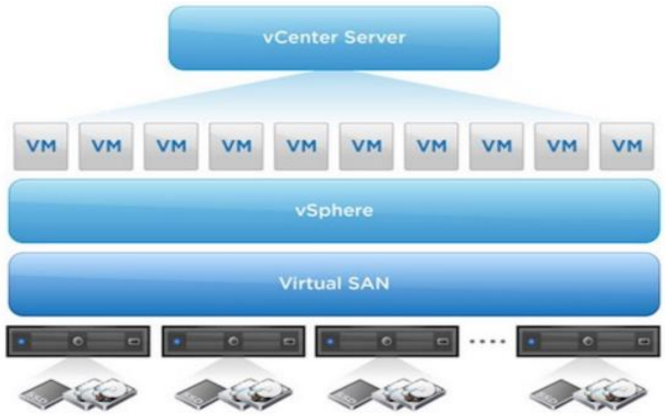
Management workloads and production server workloads will be combined into the same cluster at each site.

DDN007	Decision – Management and compute cluster will be combined in each site.
	Justification – By using Elastic DRS resources can be shared.
	Type – DQA, DQM, DQP
	Impact – Hosts at each site will share a cluster and workloads will share resources.
	Associated Risk – Maximum number of hosts per cluster is 64.
	Risk Mitigation – Deploy new cluster manually once limit has been reached.
	Reference – REQ003 ASU001, ASU002, ASU006, ASU018

9.2.5 Storage Logical Design

VMware vSAN has been picked for the shared storage at each site with each site configured for their own vSAN cluster.

Figure 3 – VMware vSAN



DDN008	Decision – VMware vSAN has been picked for the shared storage.
	Justification – vSAN allows for an easy scalable storage solution and configured using All-Flash for the performance. vSAN provides a Docker Volume Driver to assist the developers providing API for provisioning and policy configuration.
	Type – DQA, DQP
	Impact – Each compute host will participate in the vSAN cluster.
	Associated Risk – Maximum number of vSAN nodes per cluster is 64. Maximum VMs on a vSAN node is 200 and 6000 per cluster. Each cluster can only have 1 vSAN datastore.
	Risk Mitigation – Deploy new cluster manually once limits have been reached.
	Reference – REQ003, REQ004 ASU017, ASU018

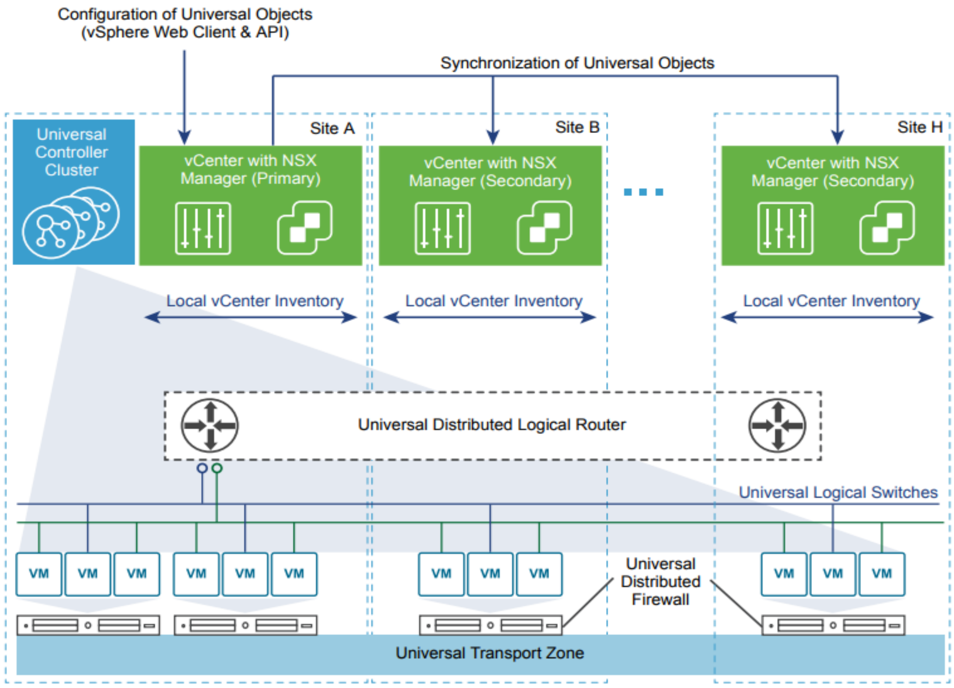
DDN009	Decision – Each site will be configured in their own vSAN cluster – no stretched cluster will be configured.
	Justification – Stretched cluster requires another site to host a witness, our 3 sites are designed to be live and require vSAN at each site.
	Type – DQM, DQP
	Impact – Each site vSAN cluster is independent from each other. Stretched cluster will not be enabled.
	Associated Risk – vSAN objects will not be accessible across sites.
	Risk Mitigation – Leverage Rubrik to protect VMs against a full site failure.
	Reference – REQ003, REQ004 ASU017, ASU018

DDN010	Decision – Workloads will be assigned Storage Policy of Fault Tolerance Method (FTM) and Failure to tolerate (FTT) will be set to FDM=RAID1 FTT=1
	Justification – Enabled fault tolerance for objects within the vSAN cluster.
	Type – DQA, DQR
	Impact – Additional storage space is required.
	Associated Risk – RAID1 only protects against one failure.
	Risk Mitigation – VMware on AWS remediation ensures vSAN operations following a host failure.
	Reference – REQ003, REQ004 ASU001, ASU002, ASU003, ASU006, ASU017, ASU018

9.2.6 Network and Security Logical Design

Cross vCenter NSX deployment has been proposed to span 3 sites, this will be used to span logical virtual networks across sites and for consistent distributed firewall polices across sites.

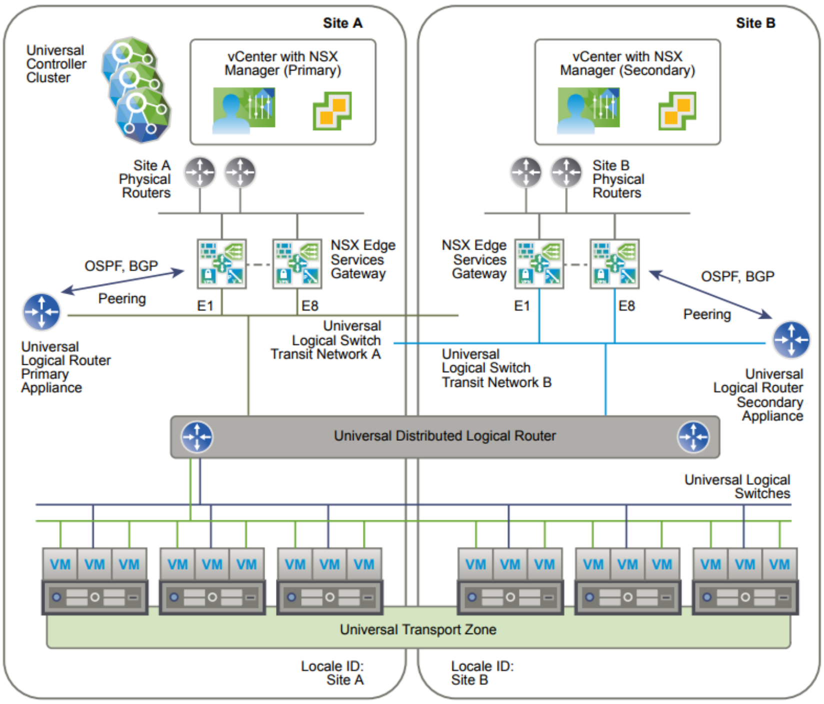
Figure 4 – Cross vCenter NSX


Configuration of Universal Objects
(vSphere Web Client & API)
Synchronization of Universal Objects
Site A
Site B
Site H
Universal Controller Cluster
vCenter with NSX Manager (Primary)
vCenter with NSX Manager (Secondary)
vCenter with NSX Manager (Secondary)
Local vCenter Inventory
Local vCenter Inventory
Local vCenter Inventory
Universal Distributed Logical Router
Universal Logical Switches
VM
Universal Distributed Firewall
Universal Transport Zone

DDN011	Decision – NSX Manager will be deployed at each site and configured for Cross vCenter.
	Justification – To allow for vMotion across vCenter's without having to reconfigure the VM or change security rules. This also centralises the security policy management across sites.
	Type – DQA, DQM, DQR, DQS
	Impact – Universal objects can be created on the primary NSX manager that is then synchronised across all sites. Primary and Secondary NSX Managers are deployed. Universal controller cluster is only deployed in the primary site.
	Associated Risk – Only the primary NSX manager running from one site can create and edit universal objects such as universal logical switches.
	Risk Mitigation – Secondary NSX Manager must be promoted if the primary NSX Manager fails. Controller cluster must also be redeployed.
	Reference – REQ004 ASU005, ASU008, ASU009, ASU018, ASU019

Each site will have universal objects deployed across sites with Edge Service Gateways (ESG) deployed at each site.

Figure 5 – Cross vCenter NSX ESG


Site A
Site B
Universal Controller Cluster
vCenter with NSX Manager (Primary)
vCenter with NSX Manager (Secondary)
Site A Physical Routers
Site B Physical Routers
NSX Edge Services Gateway
NSX Edge Services Gateway
OSPF, BGP
Peering
OSPF, BGP
Peering
E1
E8
E1
E8
Universal Logical Router Primary Appliance
Universal Logical Switch Transit Network A
Universal Logical Switch Transit Network B
Universal Logical Router Secondary Appliance
Universal Distributed Logical Router
Universal Logical Switches
VM
Universal Transport Zone
Locale ID: Site A
Locale ID: Site B

DDN012	Decision – Universal Distributed Logical Routers (UDLR) and Universal Logical Switches are to be deployed.
	Justification – To centralise routing configuration admin and to allow layer 2 networks to span multiple sites for VM migrations.
	Type – DQA, DQM
	Impact – UDLR and Universal Logical Switches are to be deployed, local objects can still be created if required.
	Associated Risk – Local egress to be enabled on UDLR to provide egress at each site.
	Risk Mitigation –Local egress is enabled on deployment, local control VM at each site.
	Reference – REQ004 ASU005, ASU008, ASU009, ASU018, ASU019

DDN013	Decision – Edge Service Gateway (ESG) to be deployed in each site.
	Justification – To pair with external routers.
	Type – DQS
	Impact – ESG will be deployed in each site, these cannot be synchronised universally.
	Associated Risk – A single ESG represents a SPOF.
	Risk Mitigation – Each ESG will be deployed in HA mode.

	Reference – REQ004 ASU005, ASU008, ASU009, ASU019

DDN014	Decision – Universal Transport Zone deployed across 3 sites
	Justification – Clusters that need to participate in universal logical networks must be added to the universal transport zone.
	Type – DQA, DQM
	Impact – Universal Transport Zone is created on the primary NSX Manager and synchronized to the secondary.
	Associated Risk – Only one transport zone can be created, maximum hosts per transport zone is 256.
	Risk Mitigation – Local transport zone to be used once maximum is reached.
	Reference – REQ004 ASU005, ASU008, ASU009, ASU018, ASU019

DDN015	Decision – Universal Distributed Firewall rules will be enabled.
	Justification – DFW support for cross vCenter vMotion to enable secure mobility of VMs.
	Type – DQS
	Impact – Universal rule sets will need to be configured and maintained, these are created on the primary NSX Manager and synchronised to the secondary. Local rules remain local.
	Associated Risk – Not all services are available universally. Universal network and security objects must be created on the primary NSX Manager. Universal security groups cannot be created from Service Composer.
	Risk Mitigation – Use the primary NSX Manager to create only the supported objects.
	Reference – REQ004 ASU007, ASU018

Each site will have virtual Distributed Switches deployed.

DDN016	Decision – Virtual Distributed Switches (vDS) will be deployed.
	Justification – vDS are required for NSX. Network IO Control will also be configured on the vDS and shares configured to ensure services such as vSAN and VM traffic are not contended.
	Type – DQP

	Impact – All hosts in the cluster will be configured with vDS. Shares will be configured for services such as vSAN traffic and VM traffic. The default shares will be enabled.
	Associated Risk – The default shares could affect performance or certain services. AWS do not allow direct vDS configuration.
	Risk Mitigation – Look to balance workloads manually.
	Reference – REQ003, REQ004 ASU007

9.2.7 Application Logical Design

The HumanLink application will be developed on the VMware on AWS infrastructure but deployed on native AWS.

DDN017	Decision – HumanLink application will be deployed into production using native AWS services.
	Justification – AWS services provide the required auto scalability and resilience.
	Type – DQA, DQP
	Impact – The application must be deployed into AWS services and managed through AWS. The application will be deployed across 3 regions.
	Associated Risk – The deployment could differ from each deployment in each region. Changes to the deployment could affect availability.
	Risk Mitigation – The application will be developed on a separate platform to production (DDNX21) and the deployment will be scripted (DDN018).
	Reference – REQ001, REQ002, REQ003, REQ004, REQ005 ASU001, AUS008, ASU010, ASU015, ASU019, ASU020

Terraform (the application) will be used to deploy the production infrastructure at each AWS region providing an 'infrastructure as code' approach to ensure each deployment is consistent.

DDN018	Decision – Terraform (the application) will be used to deploy the production infrastructure across regions and to update configs.
	Justification – Terraform will provide the infrastructure as reusable code to ensure consistency and protect against accidental or unregulated changes.
	Type – DQA, DQM, DQR, DQS
	Impact – Terraform must be used to deploy the infrastructure by code.
	Associated Risk – Provider change could affect future deployment.
	Risk Mitigation – If new platform is not supported deploy manually or look at another product.
	Reference – REQ007 ASU001, AUS008, ASU010, ASU015

Amazon Container Services (ACS) will be used for the containerized application running Docker. Amazon Elastic Files System (EFS) will be used to persist data as a temporary scratch space to process the data. The Docker images in Amazon Container Registry. The application analyses information gathered by robots and IOT devices in the terraforming sites.

DDN019	Decision – Amazon Container Services is to be used to run Docker, using Amazon Container Registry for the image and EFS to store temporary data.
	Justification – Amazon ACS provides the auto scaling and orchestration required.
	Type – DQA, DQP, DQR
	Impact – The application must be developed using Docker and deployed in AWS.
	Associated Risk – Docker wont scale automatically or across regions.
	Risk Mitigation – Amazon CloudWatch will be used to instigate the scaling. Terraform will be used to deploy in each region.
	Reference – REQ002, REQ003, REQ004 ASU001, AUS008, ASU010, ASU015

IOT devices sends data to Amazon Simple Queue Service (SQS) that integrates direct with Amazon DynamoDB. Robot will also upload all data they receive this data will be used by Scientists and Engineers in the Terraforming effort.

DDN020	Decision – Amazon DynamoDB will be deployed to be used by robots and IOT devices.
	Justification – Amazon DynamoDB provides fast and scalable database solution.
	Type – DQA, DQP
	Impact – The application must be programmed to use DynamoDB.
	Associated Risk – Scale only applied to one region up to 3 AZs.
	Risk Mitigation – To scale across regions the developers will need to pull the data across DBs.
	Reference – REQ002, REQ003, REQ004, REQ005, REQ006 ASU001, AUS008, ASU010, ASU015, ASU020

The server VMware infrastructure will be used by developers to test new versions of the application before rolling it out into production. Terraform will be used to script the deployment on the required infrastructure.

DDN021	Decision – Developers will use the server VMware cluster to test and update the HumanLink application.
	Justification – To mitigate against changes to the production application.
	Type – DQA, DQR, DQS
	Impact – The application needs to be developed on a different platform as production.
	Associated Risk – Application code or deployment is not consistent with production.

	Risk Mitigation – Run development using separate AWS accounts on separate infrastructure.
	Reference – REQ002, REQ007 ASU001, AUS002, ASU006, ASU007, AUS008, ASU015, ASU020

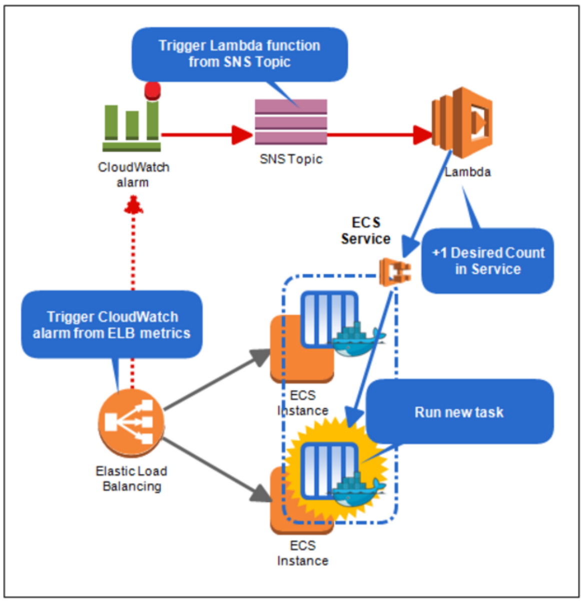
9.2.8 Operational Logical Design

For monitoring vRealize Operations Manager will be deployed to monitor the vSphere infrastructure, vSAN, NSX and AWS.

DDN022	Decision – vRealize Operations Manager (vROps) will be deployed.
	Justification – To monitor the full vSphere environment including vSAN and NSX. Remediation can be configured to respond to resource alerts.
	Type – DQM
	Impact – vROps will be deployed at each site and configure to work with vSAN and NSX.
	Associated Risk – vROps appliance represents a SPOF.
	Risk Mitigation – Deploy vROps in a cluster.
	Reference – REQ004 ASU008, ASU018

Additional monitoring will be deployed for the production HumanLink application using AWS Cloud Watch to send an alert that will trigger a AWS Lambda function to scale the ECS service.

Figure 6 – Docker ECS Auto Scale


Trigger Lambda function from SNS Topic
CloudWatch alarm
SNS Topic
Lambda
ECS Service
+1 Desired Count in Service
Trigger CloudWatch alarm from ELB metrics
ECS Instance
Run new task
Elastic Load Balancing
ECS Instance

DDN023	Decision – Configure Amazon Cloud Watch
	Justification – Amazon CloudWatch can be used to auto scale Amazon Container Service (ECS) by sending an Amazon SNS alert to AWS Lambda function.
	Type – DQA, DQP
	Impact – An alarm is to be created on the Elastic Load Balancer.
	Associated Risk – Configuration miss match across different regions.
	Risk Mitigation – Use Terraform to standardise the deployment.
	Reference – REQ001, REQ002, REQ003, REQ004, REQ007 ASU001, ASU009, ASU010, ASU015

9.2.9 Availability Logical Design

DDN024	Decision – Rubrik will be deployed at each site.
	Justification – Rubrik provides a scalable backup and replication device that can leverage both vSphere and AWS. Rubrik will be used to replicate VMs across sites to provide failover in the event of a full site failure.
	Type – DQA, DQM, DQR
	Impact – A Rubrik brick will be deployed at each site and scaled as required.
	Associated Risk – One of the judges works for Rubrik. *Real risk* - Replicated VMs need to be manually recovered in the event of a full site failure.
	Risk Mitigation – Learn Rubrik fast ☺ *Real mitigation* - Leverage the RESTful APIs to automate.
	Reference – REQ001, REQ003, REQ004 ASU001, ASU008, ASU017, ASU018, ASU019, AUS20

9.2.10 End User Logical Design

VMware Horizon will be configured for human scientists, engineers and developers to connect remotely and work on collaboration tools, admin applications and for the developers to test new versions of the HumanLink application. VDI pools will be deployed with NSX DFW integration for a secure workplace. Instant Clones will be configured for a 'just in time' scalable end user experience.

Human staff will connect to their VDI environment across the internet using tablet devices. The HumanLink application will also be accessed over the internet using the virtual desktop or published app.

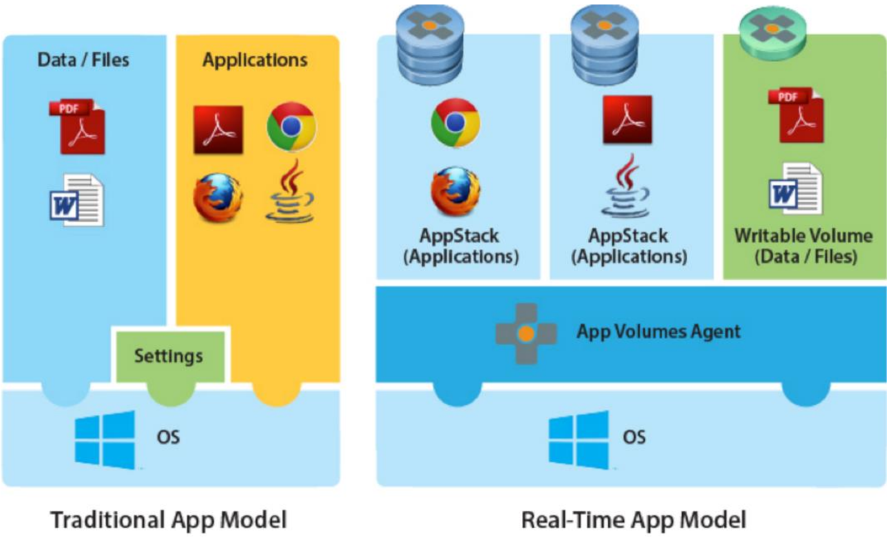
DDN025	Decision – VMware Horizon will be configured.
	Justification – To provide secure highly scalable desktops for human staff.
	Type – DQM, DQS

	Impact – Horizon connection servers will be configured internally and Access Points externally to connect to the users desktop. Horizon will integrate with vCenter to provision virtual desktops.
	Associated Risk – Connection servers and Access Points represent SPOFs. Limited scale. No shared storage between sites for the image.
	Risk Mitigation – Implement NSX Edge Load Balancer and implement CPA (DDN026). Desktop image needs to be manually copied across sites.
	Reference – REQ003, REQ004, REQ006 ASU007, ASU008, ASU009, ASU010, ASU013, ASU014, AUS015, ASU018

DDN026	Decision – Cloud Pod Architecture (CPA) will be configured.
	Justification – To provide users access to their closest site when working at a terraform site.
	Type – DQA, DQM
	Impact – Each site must be configured with CPA with assigned local VDI pools.
	Associated Risk – Users connecting to any site rather than their local.
	Risk Mitigation – Configure Amazon Route 53s Latency Based Routing.
	Reference – REQ001, REQ003, REQ004 ASU001, ASU007, ASU008, ASU009, ASU013, ASU014, AUS015, ASU018

Each application a human would use within the VDI deployment will be virtualised and only snapped into the users session at boot, images are read only and centrally managed.

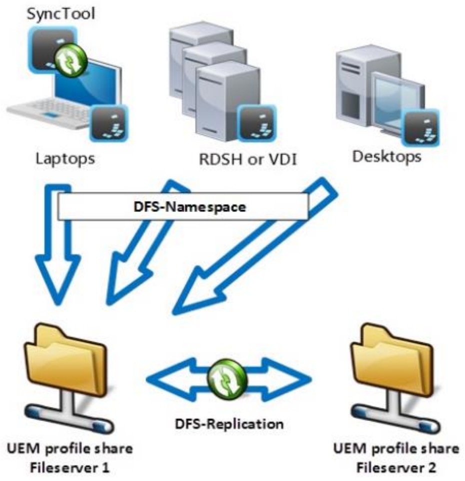
Figure 7 – App Volumes



DDN027	Decision – Collaboration tools and admin applications are to be virtualised using App Volumes.
	Justification – To provide scalable instant deployments of applications.
	Type – DQM, DQP
	Impact – Each application is virtualised and attached on deployment.
	Associated Risk – No shared storage between sites for the image.
	Risk Mitigation – The image needs to be manually copied between sites.
	Reference – REQ003 ASU007, ASU008, ASU013, ASU014, AUS015, ASU018

Users persona will be stored centrally allowing the user to log onto any pool within the CPA. DFS shares will be used to replicate the data but in an Active-Passive topology to prevent corruption.

Figure 8 – UEM with DFS topology

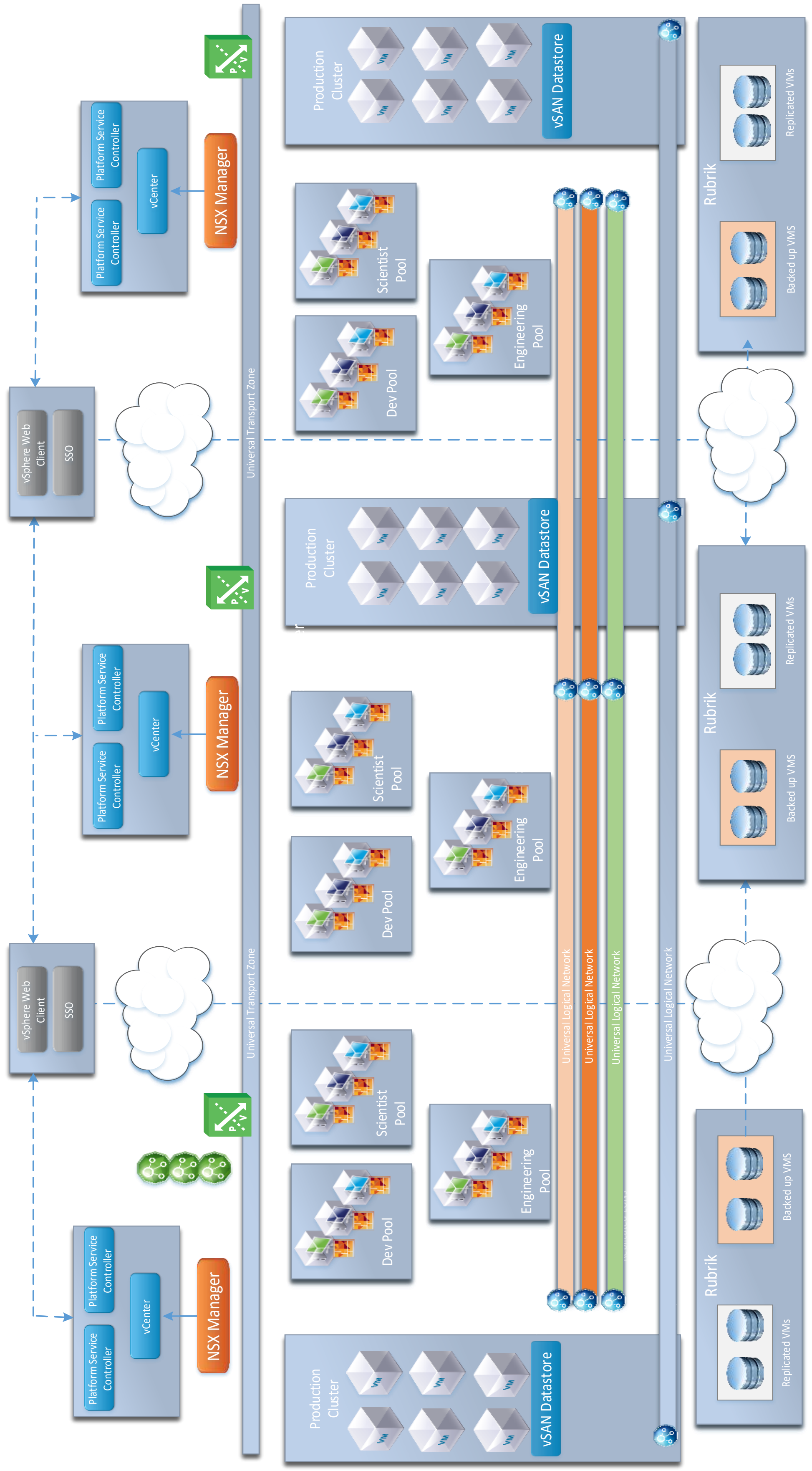


DDN028	Decision – Users personal data will be centralised using VMware User Environment Manager.
	Justification – To keep a user persistence across sites and to be able to share data among fellow staff.
	Type – DQA, DQM
	Impact – Users persona data will be redirected centrally.
	Associated Risk – No shared storage between sites for the data.
	Risk Mitigation – DFS will be configured in Active-Standby mode.
	Reference – REQ003, REQ004 ASU007, ASU008, ASU013, ASU014, AUS015, ASU018

10 Estimated Duration and Timetable

As soon as humanely possible, the human race is relying on it.

Figure 9 – VMware Infrastructure



11 References

AWS Global Infrastructure - https://aws.amazon.com/about-aws/global-infrastructure/

VMware on AWS - https://blogs.vmware.com/vsphere/2016/10/vmware-cloud-on-aws-a-closer-look.html

Supported vCenter topologies - https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2147672

Using NSX Load Balancer for PSC - https://docs.vmware.com/en/VMware-NSX-for-vSphere/6.2/com.vmware.nsx.admin.doc/GUID-CD26BE40-EFD0-4098-A1E3-FFDDCC162E2D.html

Cross vCenter vMotion Requirements - https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2106952

NSX cross vCenter - https://docs.vmware.com/en/VMware-NSX-for-vSphere/6.2/nsx_62_cross_vc_install.pdf

ESG HA - https://pubs.vmware.com/NSX-6/index.jsp#com.vmware.nsx.admin.doc/GUID-6C4F0C33-C6DD-432B-AA91-10AD6B449125.html

L2VPN NSX - http://pubs.vmware.com/NSX-61/index.jsp?topic=%2Fcom.vmware.nsx.admin.doc%2FGUID-7977542E-C4BA-484A-9811-1233E2401A23.html

AWS CloudWatch - https://aws.amazon.com/blogs/compute/scaling-amazon-ecs-services-automatically-using-amazon-cloudwatch-and-aws-lambda/

UEM supported topologies - https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2137300

Terraform on VMware - https://www.terraform.io/docs/providers/vsphere/r/virtual_machine.html

Amazon ECS persist data on EFS - https://aws.amazon.com/blogs/compute/using-amazon-efs-to-persist-data-from-amazon-ecs-containers/

Amazon ECS Auto Scaling - http://docs.aws.amazon.com/AmazonECS/latest/developerguide/service-auto-scaling.html

NoSQL and IOT white paper - http://s3.amazonaws.com/info-mongodb-com/2014-04-10_machina_research_databases_and_the_iot.pdf

Amazon SQS - https://aws.amazon.com/sqs/

Amazon DynamoDB - https://aws.amazon.com/dynamodb/details/

Terraform - https://www.terraform.io/intro/index.html