GLCG Proposal for Infrastructure Choices (master copy)

Revision history

Revision	Author	Date	Description
0.75	Jyoti Ranjan	15-June-2019	Investigation and evaluation based recommendation for Agena Cloud Gateway Storage System. It is assumed that KVM hypervisor cluster is used to realize the gateway.
0.85	Jyoti Ranjan	18-June-2019	Updated with vSAN details if we chose ESX for Agena Cloud Gateway. Also, indicated the recommended configuration can be used for compute infrastructure as well.
0.96	Jyoti Ranjan	20-June-2019	Used point based system to objectify evaluation of Agena Cloud Gateway. The scope is enlarged to consider generic gateway's requirement instead of focusing only on storage aspect.
0.99	Jyoti Ranjan	23-June-2019	Refactored the content for better representation of information.
1.0	Jyoti Ranjan	25-June-2019	Replace reference of storage to platform as the scope of the proposal is wider now.

Overview

GreenLake Hybrid Cloud (GLHC) cloud platform is a flagship cloud solution provided by HPE. Gemini is one of the product line up which aims to support **VMaaS** (VM as a Service) and **CaaS** (Container as a Service) in private cloud space with the robustness, simplicity, scalability and pay-as-you go model. In other words, the goal of the “VMaaS” offering is to provide customers a public cloud like experience on premise. The solution will allow customers to deploy various “services” (virtual machines, networks, storage etc.) through a self-service portal. The services will run on infrastructure residing on the customer premise or in a co-location facility. In order to deliver a public cloud like experience HPE will need to manage and operate certain aspects of the solution. Customers will no longer have to deal with the complexity of managing and integrating a heterogeneous solution

Managing and operating the VMaaS cloud life-cycle remotely requires a set of services running on-premises and these set of services form the "Green Lake Cloud Gateway (GLCG)". In order to run these set of services as part of the GLCG, we need to come up with the choice of physical infra definition, hypervisor stack including compute, network and storage options. Consideration to be given for optimization in terms of number of nodes / resources used within the rack, cost of the option chosen, license requirement etc, as it is mainly the control / management plane. Aim is to design the GLCG to run on set of management applications layered over very small foot print of physical hardware with required resiliency, high availability and robustness to run and control the required set of services.

Requirement(s), Constraint(s) and Assumption(s)

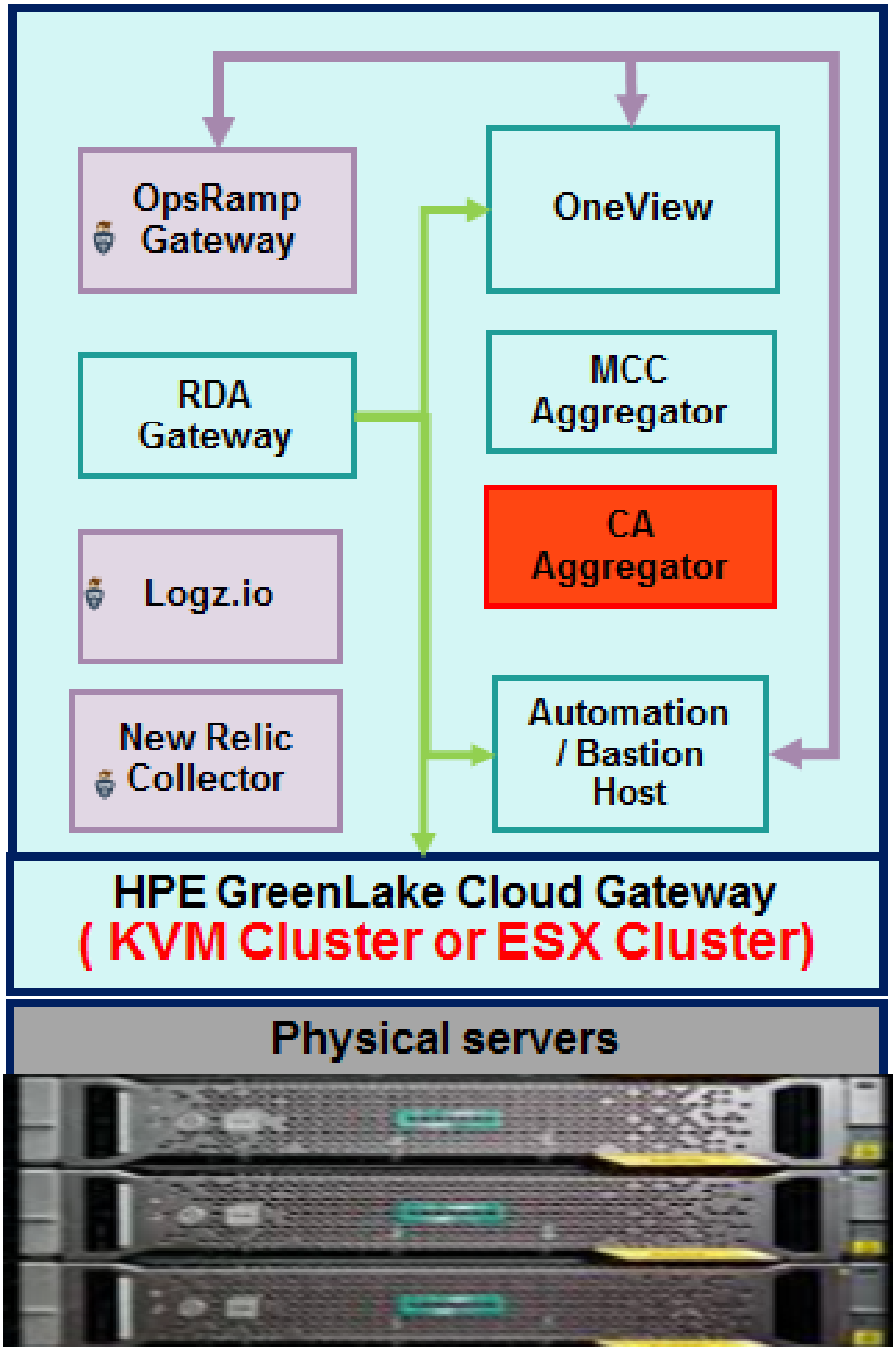
Requirements	• Need a shared storage system for Agena cloud gateway: VMaaS and CaaS (preferred) • Low foot print in terms of physical hardware. • Limited capacity in stead of ever-growing storage capacity. The required capacity can be reasonably predicted with reasonable accuracy. • Lower total cost of ownership (TCO)
Constraint(s)	• Leverage physical server used to form a compute cluster in stead of independent need of storage system (preferred)

Assumptions	1. For the design of Green Lake Cloud Gateway following assumptions have been made.
	  • Available choices of hypervisor: KVM, ESX and HyperV (no Xen hypervisor etc).
	  • Management applications will be deployed as virtual machine or appliances.
	  • Some services deployed as appliance might have its own HA while some might rely on high availability of VM.
	2. Only block storage is needed. There is no need of supportability of shared file system and object storage.

Green Lake Cloud Gateway and Core Services

Core responsibilities	Pictorial view

- Hosting core Services required for VMaaS MSP in customer Site
  - OpsRamp Gateway, RDA CAS,
  - OneView Appliance, MCC Aggregator, Logz.io,
  - Automation Host
- Acts as bastion Host
- Provides the SaaS connectivity to on-prem VMaaS cloud



Recommended design choices (Hypervisor / Infrastructure) for cloud gateway

Considering the various factor including the cost, appliances needing shared storage, low maintainability, ease of operation etc, the following choices of hypervisor stack with HPE specific hardware is considered with server choice of HPE Proliant DL380 Gen 10. More details can be found at choice under consideration and evaluation tree.

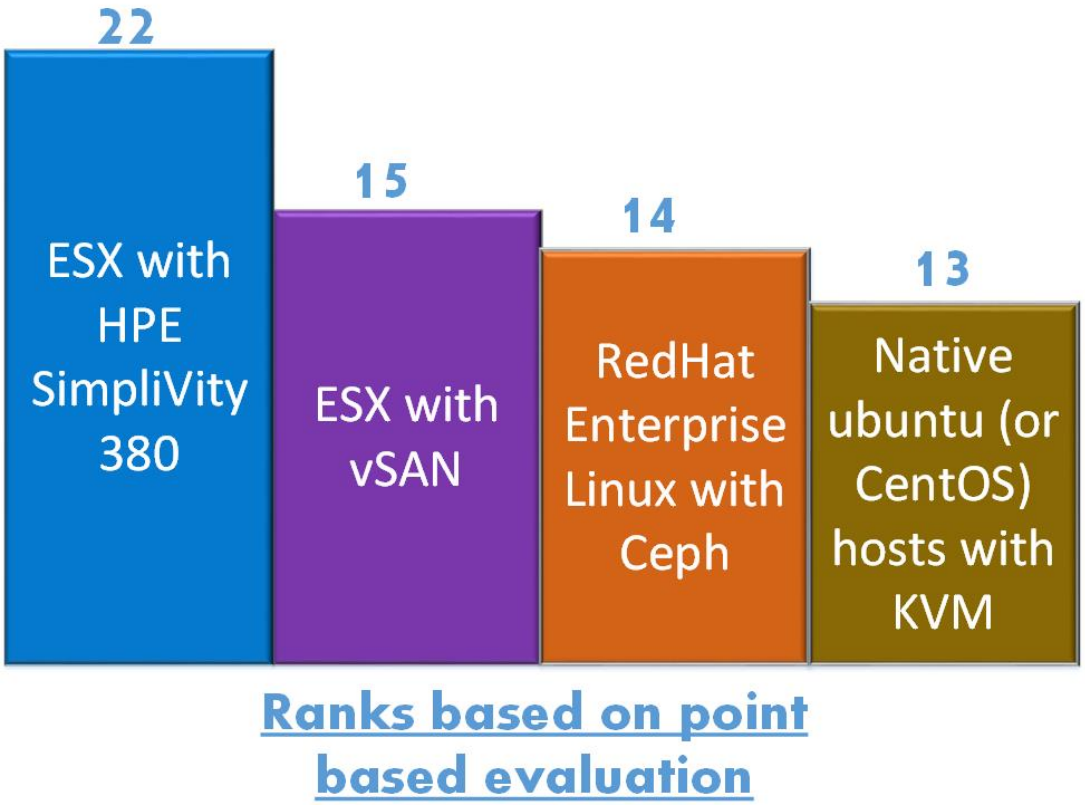
- ESX Cluster based on HPE SimpliVity 380
    - Proven solution
    - Complete automated system
    - Support shared storage
    - Support high availability of virtual appliances
    - Benefit of highly efficient storage system
    - HPE product
- ESX with vSAN
    - Proven solution
    - Easy automation because of availability of vSphere API
    - Support shared storage
    - Support high availability of virtual appliances
    - Possiblities of automation leverage from Composable Cloud (Partial)
- RHEL KVM Cluster with Ceph Storage
    - Support shared storage
    - Support high availability of virtual appliances
    - Possibilities of automation leverage from Telco blueprint
- Standalone KVM based on CentOS / Ubuntu
    - Lowest cost
    - Minimal hardware footprint and configuration
    - No license cost (if used CentOS)

To make the objective discussion, the point based system is used to rank the above 4 choices based on the parameter which is of importance for decision making. These parameters does not belong to technical points but is based on other aspects of ease of use, business sense, branding etc. The 3 point of rating system is used: 0 - No value, 1 - Low value, 2 - Good value and 3 - High value, X - Important but not being considered at this point of time. . Here is evaluation matrix-based on 3 point rating system:

	ESX with vSAN on DL380	**ESX with HPE SimpliVity 380**	**RHEL KVM Cluster with Ceph storage on DL380**	**Standalone KVM based on CentOS / Ubuntu on DL380**
Belong to HPE or Opensource family of product(s) - helps in communicating better proposition instead of telling ourselves as VMware shop	0	3	0	3
Recommended number of minimum physical servers	3	3	3	3 Note: The gateway can be realized using two physical servers hosting 3 VMs if they service level HA using Keepalived and HAproxy. The configuration can support single point of failure at each level but we are recommending to appliance to be evenly distributed i.e. one per physical server
Simplified high availability of hypervisor cluster See GLCG Proposal for service's high availability for more detail on high availability of services in association with gateway infrastructure	3	3	2	0
Built in support for intelligent backup, move, clone and disaster recovery feature	1	3	0	0
Ease of day-zero deployment	1	3	2	3
Ease of maintainability (day-n) post deployment including bug resolution, customer support etc	1	3	2	1
Available shared storage capacity with minimal configuration	3	3	3	0
Cost including physical infrastructure and licensing	2	1	2	3

Supports native services running on baremetal as well as appliances (do we need?)* ⭐ At this point of time this aspect is marked as don't care as it is not certain its need at this point of time. So, the X (don't care) is assigned to it.	X	X	X	X
Total	**15**	**22**	**14**	**13**

So the winner is **ESX with HPE SimpliVity 380** by marginal points as depicted below.



System configuration

Here is recommended hardware specification for each choices of Agena cloud gateway.

Storage system	**ESX with vSAN**	**ESX with HPE SimpliVity**	**RHEL KVM Cluster with Ceph Storage**	**Standalone KVM based on CentOS / Ubuntu**

Hardware configuration	• 3 DL 380 Gen 10 Performance 5 servers • 192 GB (6 x 32GB) Memory. • 1 NVMe disk per DL 380 Gen 10 server • 8-12 LFF SAS /SATA disk per DL 380 Gen 10 server (varies on specific storage capacity requirement) • HPE Smart Array P408i-a SR Ge 10 Ctrlr • 2 NIC to separate management and storage traffic	• 2-4 DL 380 Gen 10 Performance 5 servers based on who is responsible for bringing vCenter and Arbiter for core OmniStack cluster. See GLCG Deployment Architecture using HPE SimpliVity 380 (master copy) for more details. • 192 GB (6 x 32GB) Memory. • 8-12 SSD disk per DL 380 Gen 10 server (varies on specific storage capacity requirement) • HPE Smart Array P408i-a SR Ge 10 Ctrlr • 3 NIC to separate management, federation and storage traffic	• 3 DL 380 Gen 10 Performance 5 servers • 192 GB (6 x 32GB) Memory. • 2 NVMe disk per DL 380 Gen 10 server • 8-12 LFF SAS /SATA disk per DL 380 Gen 10 server (varies on specific storage capacity requirement) • HPE Smart Array P408i-a SR Ge 10 Ctrlr • 2 NIC to separate public and cluster network	• 2 DL 380 Gen 10 Performance 5 servers • 192 GB (6 x 32GB) Memory. • Set of 8-12 LFF SAS/SATA disk per DL 360 Gen 10 server (varies on specific storage capacity requirement) • HPE Smart Array P408i-a SR Ge 10 Ctrlr • 2 NIC to separate public and cluster network
Operating system and hypervisor	ESX	ESX	• Operating system : RHEL • Hypervisor: KVM	• Operating system : CentOS / ubuntu • Hypervisor: KVM
Raw storage capacity	• Varies with the number of disk we use but will be sufficient for gateway requirement.	• Varies with the number of disk we use but will be sufficient for gateway requirement.	• Varies with the number of disk we use for OSD.	• No shared storage, only local storage • Varies with number of disk
Usable storage capacity	• 50% because of RAID 10	• 50% because of RAID 10	• 1/3 of raw storage capacity	• Depends upon RAID configuration. Recommended is to use RAID 5. In that case, it will vary based on number of disks for RAID lun,
Storage configuration	• vSAN provides shared shared data store which can be used to present disk to VMs.	• SimpliVity SDS provides shared shared data store which can be used to present disk to VMs.	• Use RBD to present volume to Virtual Machine • Use thin provisioning • Use R10 volume created from two NVMe disk for journaling • Use ceph tool for manageability • Create separate pool for specific storage purpose • Do not use to launch instance directly on Ceph cluster (not recommended because of performance)	• Local file system laid over RAID luns created from local disks
Licensing	• Need vCenter license • Need ESX hypervisor license • Need vSAN license	• Need vCenter license • Need ESX hypervisor license • Need HPE SimpliVity license	• Need RedHat Linux Enterprise edition license • Need RHEA or RedHat Enterprise Linux High Availability license	• Open Source license with CentOS (if we do not consider ubuntu)

Note:

In This proposal GLCG infrastructure choice of DL 380 Gen 10 being primarily considered. But the design proposal can be used with 3 DL 360 Gen10 servers to serve both compute and storage requirement if capacity and network requirements are acceptable. This goes with all choices where we go for hyper-converged model for cloud gateway. There is one exception if we decide to use Nimble. In that case, our gateway platform will be composed of following entities

- *3 DL Gen 10 servers to host compute infrastructure*
- *Nimble array for shared storage system*
- *iSCSI network to connect compute with storage system*

See critical specification differences between DL 360 and DL 380 here, which can result in reduced storage and network capabilities.

Conclusion

Based on the above analysis, Its recommended to go with **ESX with HPE SimpliVity** as the platform choice hosting On-Prem core service(s). Salient points on the proposed platform choice are Hypervisor HA capability, Life Cycle Management automation availability, Efficient Shared Storage and has the advantage being HPE Solution / Product.

Appendix

Choices under consideration

The section evaluates the Agena's cloud gateway. The prime consideration of gateway choices are:

1. Hypervisor
2. Shared storage system for hypervisor
3. Networking support

Among the three, networking support is very much tied to hypervisor selection and henceforth it is not discussed in expanded manner. On the other hand, there can be multiple shared storage system choices for a given hypervisor. For e.g HPE SimpliVity, vSAN and Nimble can be storage system choices for ESX hypervisor based cloud gateway. And hence, a detailed diagnosis of storage configuration is needed. It is worth to mention that the parameter will vary greatly if the storage system needs to be built of PB and is meant for user applications or for other purposes like backup etc. So, the reader is requested to keep this view point in mind. So, consider this section as why some of the combination of hypervisor and storage system are being considered and which parameters have critical influence in decision making system

Storage aspect	**HPE Nimble**	**HPE SimpliVity**	**Ceph**	**vSAN**
Compute hardware requirement	No extra hardware, comes with built-in hardware box	• 2-4 Physical servers with plenty of storage disk based on who is responsible for bringing vCenter and Arbiter. • DL 380 should serve the purpose	• 3 physical servers with plenty of storage disks • DL 380 should serve the purpose	• 3 physical servers with plenty of storage disks. Recommendation is for 4 servers for better manageability and performance. • DL 380 should serve the purpose
Minimal compute hardware requirement	No extra hardware, comes with built-in hardware box	• One physical server (not recommended, only for PoC)	• One physical server (not recommended, only for PoC)	• One physical server (not recommended, only for PoC)
Specific host hardware requirement	No specific requirement	• PCIe Accelerator Card	• At least SSD per compute node for journal disk (recommended)	• At least ne SSD per compute node for caching disk

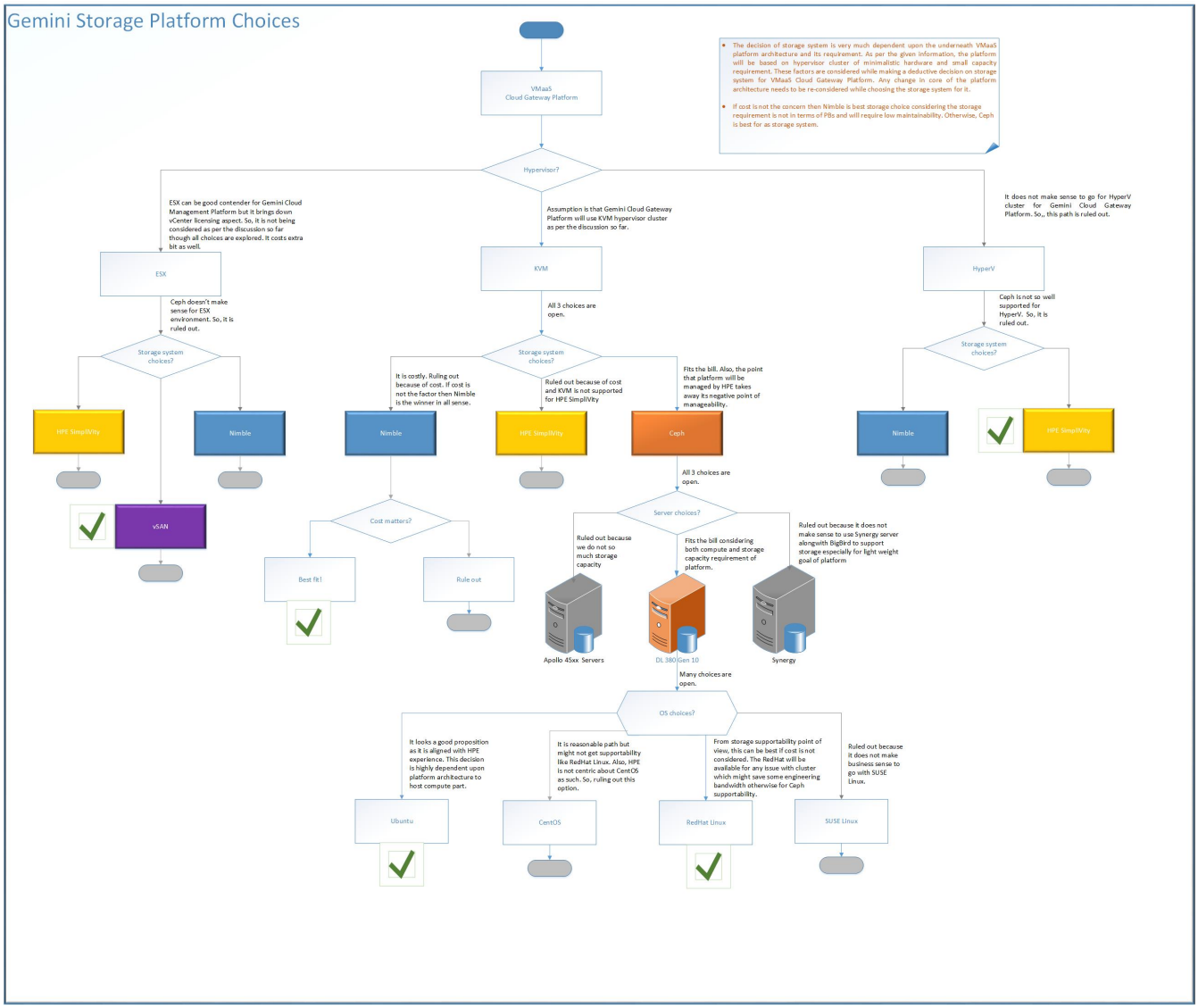
Network requirement	• Separate network for control and data path • At least two NIC (recommended) for storage. Each NIC forms its own VLAN.	• 3 NIC to separate management, federation and storage traffic	• Depends • Separate public and cluster network (recommended). It will require at two NIC used to create two different VLAN.	• Depends • Separate public and cluster network (recommended). It will require at two NIC used to create two different VLAN.
Disk storage requirement	No specific requirement	• Depends upon storage requirement. More disks per server we have, more disk storage we have. All disks have to be SSD.	• At least one SSD per compute • Depends upon storage requirement. More disks per server we have, more disk storage we have	• At least one SSD per compute • Depends upon storage requirement. More disks per server we have, more disk storage we have
Hyper-visor supportability	Hyper-visor agnotics	• ESX • KVM	• KVM	• ESX
Operating system supportability	• Linux flavors • Windows	• ESX • Window	• Ubuntu • RedHat CentOS • RedtHat Linux	• Build component of hypervisor. Tenant VM can have any OS.
Specific host software requirement other than what is needed to implement storage system	• Multi-path software if we are connecting device through more than one path	No	No	No
Can be used as hyper-converged storage system?	No	Yes	Yes	Yes
Storage capacity	Need to be figure out for minimal cost Nimble storage system	Depends upon storage configuration	Depends upon storage configuration	Depends upon storage configuration
Fault tolerance	RAID	RAID 10 mirroring	Replicas	RAID 10 mirroring
Storage protocol meaningful for gateway	iSCSI	NFS	RBD	Native VMDK storage protocol
Thin provisioning to leverage consume bits as you need	Yes	Yes	Yes	Native VMDK storage protocol
Availability	99.9999%	Need to be discovered	Not published. It significantly depends upon how well we manage the storage system.	TBD
Scalabiltiy	Vertically scalable	Vertically and Horizontally scalable	Vertically and Horizontally scalable	Vertically and Horizontally scalable
Performance	Very good	Very good	Good	Very good
Manageability	Easy	Easy	Moderate	Easy
Telemetry	Very good	Very good	Good	Easy
OpenStack compatibility	No	Yes	Yes (Very good)	Yes
Maintainability / Operation cost	Low	Low	Medium	Low

Purchase cost	$40K	• Varies based on the configuration we choose. Proposed model as per the author of the page 'DL 380 Gen 10 Performance 5', which costs around 5000$ per unit. • Extra cost for PCIe card • Extra cost for more number of disks to be attached to physical server • HPE SimpliVity (and vCenter) license. There is no point in going for HPE SimpliVity for HyperV support.	• Varies based on the configuration we choose as well as the deployment strategy we adopt for. For e.g. if we take support from RedHat then we might have to pay. As far as physical server alone is concerned, recommendation is to use 'DL 380 Gen 10 Performance 5', which costs around 5000$ per unit (by author of the page). • Extra cost for SSD disk • Extra cost for more number of disks to be attached to Ceph physical server	• Varies based on the configuration we choose. Proposed model as per the author of the page 'DL 380 Gen 10 Performance 5', which costs around 5000$ per unit. • Extra cost for more number of disks to be attached to physical server • Extra license for vSAN

Evaluation tree

The below diagram illustrates the various choices and reasoning behind it for its positioning for Agena cloud gateway's infrastructure. As mentioned above, three aspects were considered: compute, storage and networking. Compute and networking requirement being constant for all choices, the variance is because primarily storage system. So, the below depicted decision tree is focused around storage system to see which fits the bill more closely. As part of evaluation the following storage system were considered:

- vSAN
- HPE Nimble
- HPE SimpliVity
- Ceph



DL 380 vs. DL 360 Gen 10 servers

Above recommendation is made using DL 380 servers because of its proven and availability of technical references. However, this section captures its comparison with DL 360 if cost and real state consumption is a critical aspect to be looked after for. Based on below points, it is recommended to go for DL 380 Gen 10 servers.

Aspects	DL360 Gen 10	DL380 Gen 10	Authors comment
CPU	4-28 core	4-28 core	-
Storage capacity	Good	Very good	Having a higher storage capacity per node gives us ability to vertically scale storage requirement if we add more and more services.
Network controller	1 x 4 Gb	1 x 4 Gb	it is one critical aspects as high network bandwidth will help to deal with heavy traffic system as we are going to use hyper converged storage system - typically more chattier over network. Also, physical isolation of network can be achieved
Form factor	1 U	2 U	Considering that we are recommending for 3 physical servers, we are not going to get much saving because of low form factor of DL 360. We can place at max one extra DL 380s.
Disks	4 LFF drive	8-12 LFF drive	It is recommended to use DL380 as it gives us flexibility to full fill the storage requirement for day zero and as well as for day n requirement.