

# Cloud Infrastructure Planning and Design



## Welcome to Cloud Infrastructure Planning and Design.

Copyright ©2015 EMC Corporation. All Rights Reserved. Published in the USA. EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license. The trademarks, logos, and service marks (collectively "Trademarks") appearing in this publication are the property of EMC Corporation and other parties. Nothing contained in this publication should be construed as granting any license or right to use any Trademark without the prior written permission of the party that owns the Trademark.

EMC, EMC<sup>2</sup> AccessAnywhere Access Logix, AdvantEdge, AlphaStor, AppSync ApplicationXtender, ArchiveXtender, Atmos, Authentica, Authentic Problems, Automated Resource Manager, AutoStart, AutoSwap, AVALONidm, Avamar, Bus-Tech, Captiva, Catalog Solution, C-Clip, Celerra, Celerra Replicator, Centera, CenterStage, CentraStar, EMC CertTracker, CIO Connect, ClaimPack, ClaimsEditor, Cloralert, cLARIION, ClientPak, CloudArray, Codebook Correlation Technology, Common Information Model, Compuset, Compute Anywhere, Configuration Intelligence, Configuresoft, Connectrix, Constellation Computing, EMC ControlCenter, CopyCross, CopyPoint, CX, DataBridge, Data Protection Suite, Data Protection Advisor, DBClassify, DD Boost, Dantz, DatabaseXtender, Data Domain, Direct Matrix Architecture, DiskXtender, DiskXtender 2000, DLS ECO, Document Sciences, Documentum, DR Anywhere, ECS, elnput, E-Lab, Elastic Cloud Storage, EmailXaminer, EmailXtender, EMC Centera, EMC ControlCenter, EMC LifeLine, EMCTV, Enginuity, EPFM, eRoom, Event Explorer, FAST, FarPoint, FirstPass, FLARE, FormWare, Geosynchrony, Global File Virtualization, Graphic Visualization, Greenplum, HighRoad, HomeBase, Illuminator, InfoArchive, InfoMover, Infoscape, Infra, InputAcce/, InputAcce/ Express, Invista, Ionix, ISIS, Kazeon, EMC LifeLine, Mainframe Appliance for Storage, Mainframe Data Library, Max Retriever, MCx, MediaStor, Metro, MetroPoint, MirrorView, Multi-Band Deduplication, Navisphere, Netstorage, NetWorker, nLayers, EMC OnCourse, OnAlert, OpenScale, Petrocloud, PixTools, Powerlink, PowerPath, PowerSnap, ProSphere, ProtectEverywhere, ProtectPoint, EMC Proven, EMC Proven Professional, QuickScan, RAPIDPath, EMC RecoverPoint, Rainfinity, RepliCare, RepliStor, ResourcePak, Retrospect, RSA, the RSA logo, SafeLine, SAN Advisor, SAN Copy, SAN Manager, ScaleIO Smarts, EMC Snap, SnapImage, SnapSure, SnapView, SourceOne, SRDF, EMC Storage Administrator, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix, Symmetrix DMX, Symmetrix VMAX, TimeFinder, TwinStrata, UltraFlex, UltraPoint, UltraScale, Unisphere, Universal Data Consistency, Vblock, Velocity, Viewlets, ViPR, Virtual Matrix, Virtual Matrix Architecture, Virtual Provisioning, Virtualize Everything, Compromise Nothing, Virtuent, VMAX, VMAXe, VNX, VNXe, Voyence, VPLEX, VSAM-Assist, VSAM I/O PLUS, VSET, VSPEX, Watch4net, WebXtender, xPression, xPresso, Xtrem, XtremCache, XtremSF, XtremSW, XtremIO, YottaYotta, Zero-Friction Enterprise Storage.

## Course Overview

|                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Description</b> | This course provides the knowledge and skills to successfully design a cloud infrastructure. It uses an open approach to planning and design, focuses on core concepts and practices, and highlights the benefits, challenges, and considerations of design choices. This course provides students with an opportunity to practice these new skills during a collaborative cloud design project.                                                                                                                                                                                                                                                                                                                                                      |
| <b>Audience</b>    | This course is intended for architects, designers, and consultants who are involved in the planning and design of cloud infrastructure components.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Objectives</b>  | <p>Upon completion of this course, you should be able to:</p> <ul style="list-style-type: none"><li>• Apply relevant design processes and techniques when producing an effective cloud infrastructure design.</li><li>• Create a cloud infrastructure design that addresses business requirements and constraints that were identified during an assessment.</li><li>• Communicate the benefits, challenges, and considerations of various cloud-enabling technologies as well as their relevance in meeting a business requirement or constraint.</li><li>• Develop a cloud infrastructure design that includes components such as cloud management, consumer resources, elasticity, metering, hybrid capabilities, and disaster recovery.</li></ul> |

*EMC believes the information in this course is accurate as of its publication date.*

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



2

This course covers the basic knowledge and skills that are required in the cloud design process.

# Agenda

| Module/Lessons                                                                                                                                                                                                                                                                                                                                                                                        | Labs                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| <b>Day 1</b> <ul style="list-style-type: none"><li>• Introduction to Cloud Design<ul style="list-style-type: none"><li>• Technologies and terms</li><li>• Planning and Design Overview</li></ul></li><li>• Cloud Management<ul style="list-style-type: none"><li>• Cloud Management Platforms</li><li>• Cloud Management Infrastructure</li></ul></li></ul>                                           | <ul style="list-style-type: none"><li>• Introduction to Cloud Design</li><li>• Cloud Management Design</li></ul> |
| <b>Day 2</b> <ul style="list-style-type: none"><li>• Compute Resources<ul style="list-style-type: none"><li>• Pools and Physical Servers</li><li>• Host Software and Functionality</li></ul></li><li>• Storage Resources<ul style="list-style-type: none"><li>• Storage Requirements and Architectures</li><li>• Storage Types and Pools</li><li>• Advanced Storage Functionality</li></ul></li></ul> | <ul style="list-style-type: none"><li>• Compute Resource Design</li><li>• Storage Resource Design</li></ul>      |
| <b>Day 3</b> <ul style="list-style-type: none"><li>• Network Resources<ul style="list-style-type: none"><li>• Local Area Networks</li><li>• Storage Networks</li></ul></li></ul>                                                                                                                                                                                                                      | <ul style="list-style-type: none"><li>• Network Resource Design</li></ul>                                        |



# Agenda

| Module/Lessons                                                                                                                                                                                                                                                          | Labs                                                                                                                                           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Day 4</b> <ul style="list-style-type: none"><li>• Elasticity, Monitoring, and Metering<ul style="list-style-type: none"><li>• Elasticity</li><li>• Monitoring</li><li>• Metering</li></ul></li><li>• Hybrid Cloud Capabilities</li><li>• Disaster Recovery</li></ul> | <ul style="list-style-type: none"><li>• Elasticity, Monitoring, and Metering</li><li>• Hybrid Cloud Design</li><li>• Service Backups</li></ul> |
| <b>Day 5</b> <ul style="list-style-type: none"><li>• Planning and Design Lab Presentation</li></ul>                                                                                                                                                                     | <ul style="list-style-type: none"><li>• Final Lab Design Presentation</li></ul>                                                                |





This slide intentionally left blank.





# Module 1

## Introduction to Cloud Design



Upon completion of this module, you should be able to:

- Describe the concepts and terminology associated with cloud characteristics, deployment models, services, and infrastructure.
- Apply key processes such as structured assessment, requirements gathering, and goal definition to a cloud design project.
- Articulate the importance of governance, risk, and compliance requirements within the assessment process.
- Produce logical and physical designs and explain the appropriate uses for each.



Cloud Infrastructure Planning and Design

1

This module focuses on the concepts, technologies, and processes used in the cloud and in cloud design.

# Lesson 1

## Technologies and Terms

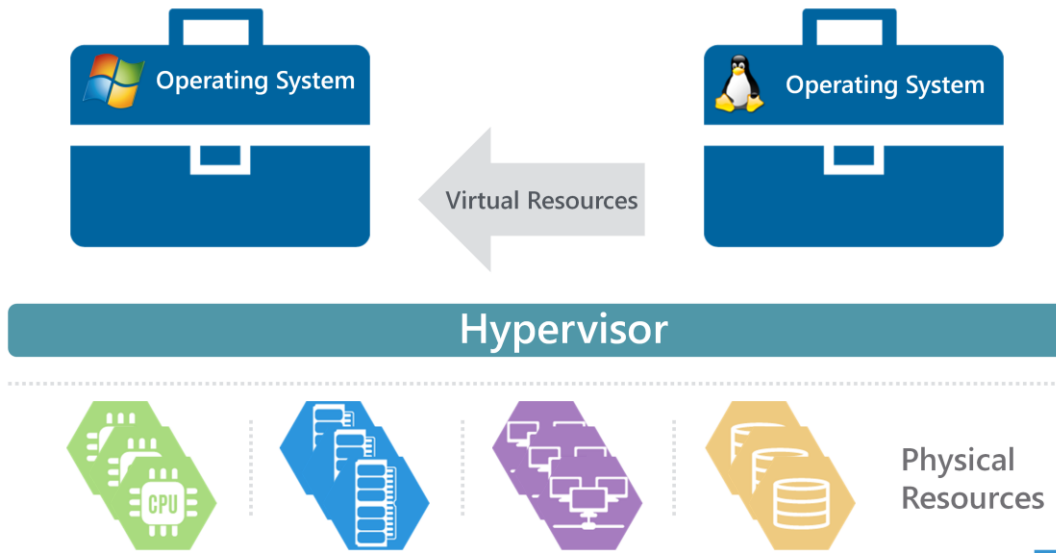
This lesson covers the following topics:

- Review of general cloud concepts
- Descriptions of cloud infrastructure technologies



This lesson reviews the terms used to describe general cloud concepts and cloud infrastructure technologies. Its purpose is to provide a baseline set of knowledge which is used throughout this course.

## Server Virtualization Overview



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

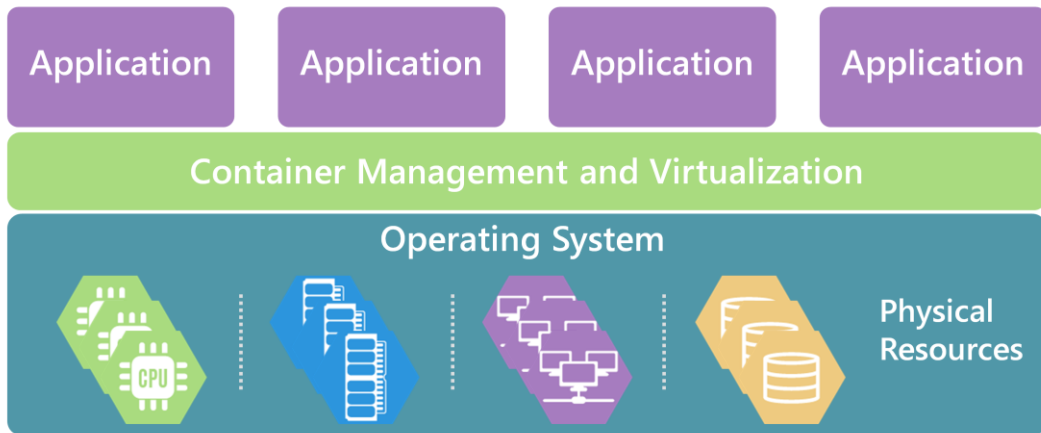


3

Virtualization is the application of different technologies that provide an abstract or logical representation of physical resources which can then be used by other services. In server virtualization, for example, the technology that provides the abstracted or logical resources is called a hypervisor. A hypervisor is software that is installed on a computer system. It provides abstracted versions of resources such as CPU, memory, network interface cards, storage interface cards, and disks to a special container called a virtual machine. When you install an operating system within this virtual machine, the operating system recognizes these abstracted resources as if they were physical resources. Usually the abstracted resources represent only a portion of the actual physical resources, which allows multiple virtual machines to share the resources of a single computer system. Once a virtual machine is created, consumers can install applications on the operating system.

Although virtualization is not a requirement for cloud computing, it is used in most cloud infrastructures for more efficient resource sharing, rapid scaling, and cost control.

## Container-Based Virtualization Overview



© Copyright 2015 EMC Corporation. All rights reserved.

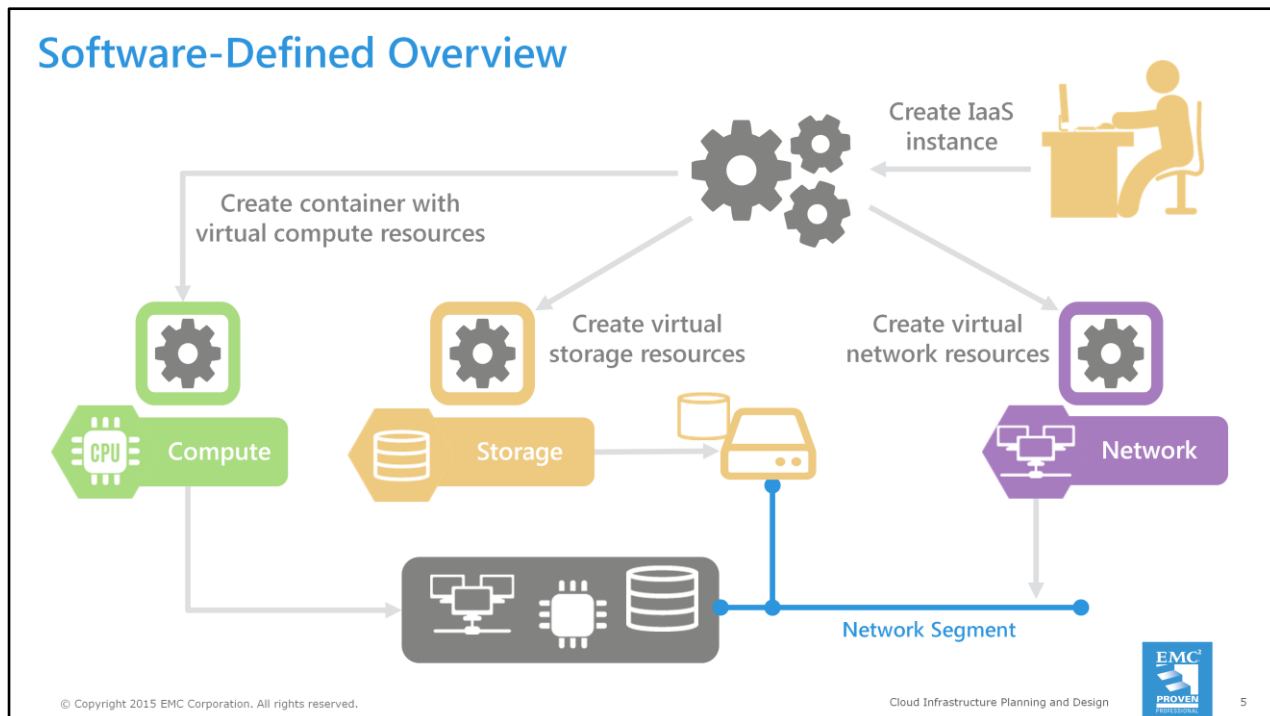
Cloud Infrastructure Planning and Design



4

Another option for deploying applications at scale is to use containers. Containers still reside on a physical server with a base operating system but there is no hypervisor and therefore no hardware virtualization. Containers share the same physical hardware and base operating system. A management and virtualization layer is added to the host OS which is responsible for deployment, resource provisioning, and logical separation of the containers. Applications run in their own confined area just like in a virtual machine but they share the same kernel with the host and there is no physical device abstraction. Containers can be deployed more quickly than virtual machines and because of the reduced overhead, containers have better consolidation ratios.

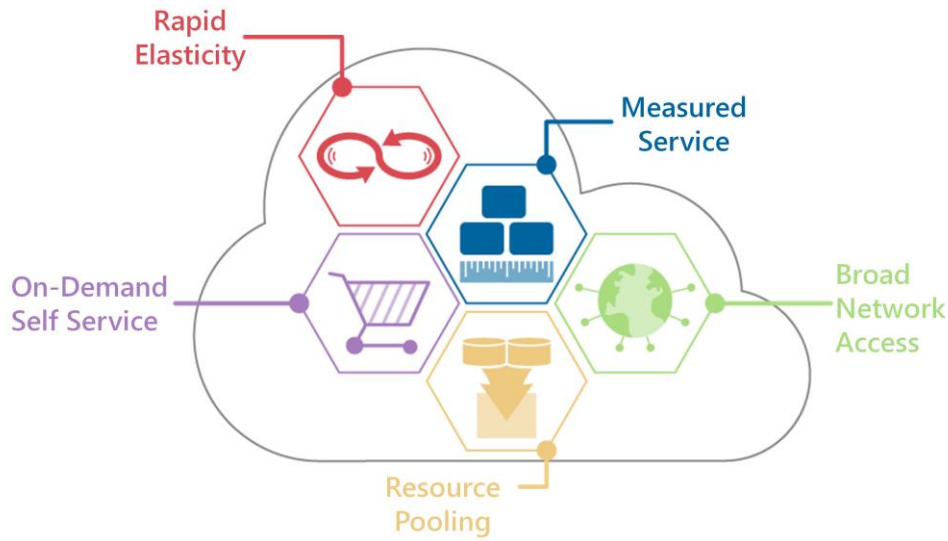
## Software-Defined Overview



The term “software-defined” has been used to describe compute, network, storage, infrastructure, data centers, and other domains. Although the definition of the term may vary across these domains and across vendors, generally speaking it is a method of delivering consumable resources using software, standards, policies and protocols that allows enterprises to deliver services across multiple hardware platforms in a scalable and automated fashion. Going beyond just hardware virtualization, software-defined environments enable the programmatic provisioning, control, and configuration of these abstracted resources which are then presented to services as virtualized resources.

An example of software-defined capabilities can be seen when a consumer requests an infrastructure instance through a service catalog. The orchestration component of the service catalog initiates processes that use standard protocols and policies to communicate with various controllers. The controllers, in turn, use various protocols and policies to create the abstracted resources in their respective domains. For instance, the network controller communicates with various hardware and software components to provision the necessary networks segments. This could involve creating VLANs on physical switches, overlay networks on virtual switches, or enabling interfaces on routers or firewalls. The storage controller communicates with storage components and creates the necessary storage that matches the policies defined in the service. This might include providing storage that has certain performance or availability characteristics. Finally, the compute controller allocates the required CPU and memory resources that are defined in the policy associated with the service. Once all of the required pieces are created, the resources are combined from the different domains and an infrastructure instance is created.

## Cloud Overview



Based on: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



6

The term cloud computing has different meanings to different people. Many experts view the cloud computing definition by the National Institute of Standards and Technology (NIST), a US Government standards authority, as a good working definition. A cloud is an environment in which resources such as compute, network, storage, and applications can be rapidly provisioned by consumers with little to no interaction from the cloud provider. These resources are accessible over the network and are drawn from central pools as required and returned to these pools when no longer needed.

NIST defines five key characteristics for cloud:

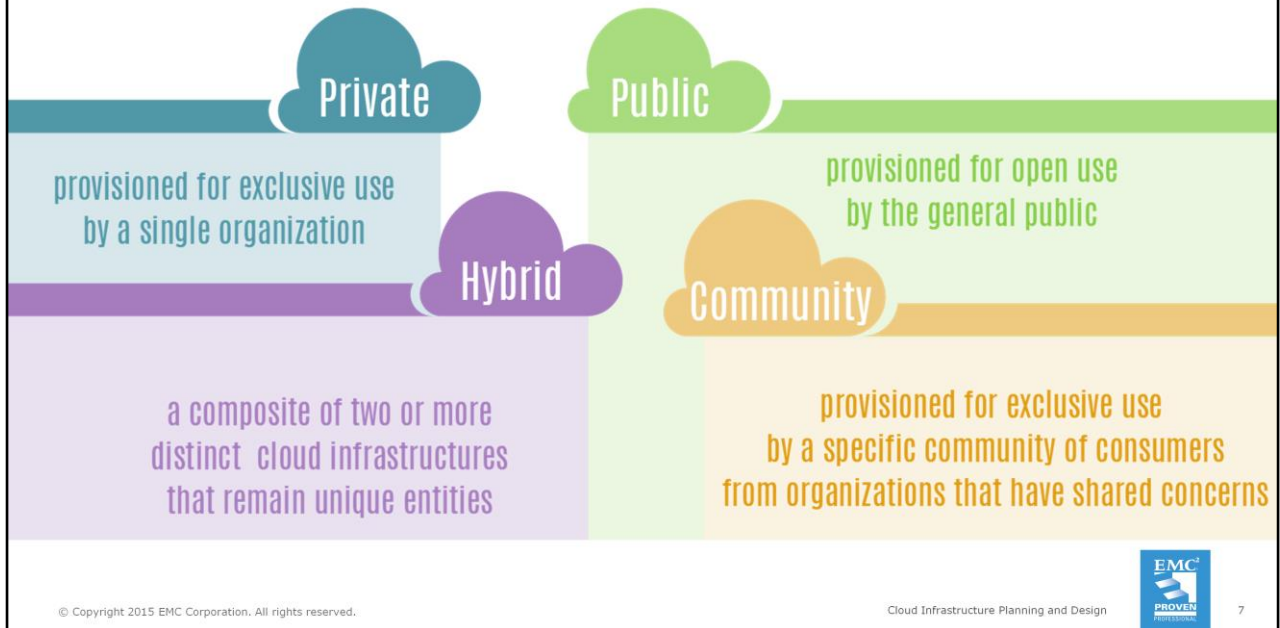
- On-demand self-service – Consumers provision resources or services as required without intervention from the cloud service provider.
- Broad network access – Consumers can access resources and cloud capabilities over the network using standard protocols, processes, and applications.
- Resource pooling – Resources are shared by all of the cloud consumers but may be separated to support a multi-tenant model. Consumers may not be aware of where these resources are located once provisioned.
- Rapid elasticity – Resources are provisioned and released, in some cases automatically, as demand for those resources increases or decreases. From the consumer's perspective, these resources appear to be unlimited and available at any time.
- Measured service – Provides monitoring and reporting mechanisms that show the consumer how much of a resource they are consuming and the associated cost.

The five characteristics listed here are what make a cloud infrastructure design different from other infrastructure designs. Your design requires all of these characteristics to be included and must align with the organizations requirements.

Source: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>



# Cloud Deployment Models



The National Institute of Standards and Technology (NIST) created a formal definition for cloud deployment models, which are shown here.

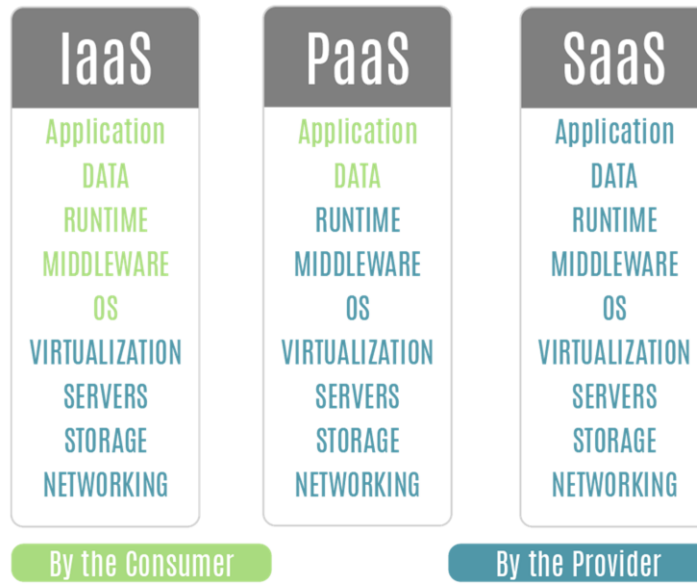
- In a Private Cloud model, the cloud infrastructure is provisioned for exclusive use by a single organization such as an enterprise or business. It may be owned, managed, and operated by the organization or an outside entity, and it may exist on or off premises. The capabilities provided by the private cloud are for the exclusive use of the organization, but may be subdivided to support multiple tenants such as business units.
- In the Public Cloud model, the cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization. It exists on the premises of the cloud provider and is subdivided to support multiple tenants. In a public cloud, the tenants could be large organizations, enterprises, or businesses, as well as individual consumers. Amazon, Rackspace, Microsoft and VMware all offer Public Cloud services.
- In a Community Cloud model, the cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns. For example, a community cloud may be provisioned to address security requirements or compliance considerations for an organization in a specific industry such as healthcare. It may be owned, managed, and operated by one or more of the organizations in the community or a third party, and it may exist on or off premises.
- In a Hybrid Cloud model, the cloud infrastructure is a composite of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability.

Source: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>

Each of the cloud models has a different impact on the cloud design in areas such as security,

compliance, availability, access, and networking.

## Cloud Service Models



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

There are three basic types of services that consumers can provision in the cloud. NIST refers to these as Cloud Service Models.

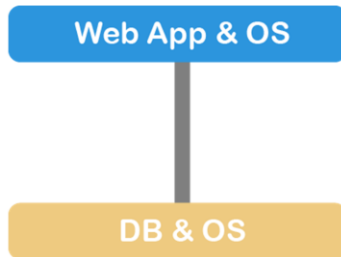
The service models are:

- Infrastructure as a Service – Ability to provision fundamental infrastructure resources such as compute, network, and storage. Although consumers are provided with access to these resources, the provider maintains control of the underlying infrastructure that supports it. Consumers typically have control of the operating system and applications that are deployed within this service.
- Platform as a Service – A collection of consumer-created or acquired applications created using programming languages, tools, libraries, and other services supported by the provider and that are deployed onto cloud infrastructure. Consumers have control over the deployed applications and configurations and providers maintain control over the infrastructure and operation systems.
- Software as a Service – An application deployed on cloud infrastructure which is controlled by the provider. Consumers are granted access to use the application.

Source: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>

During the requirements gathering phase, it is important to determine which service models are planned by the organization since each service model has a different impact on a cloud design. However, it is also important to note that when designing a private cloud environment, the organization consists of both the consumers and the provider, so for all three service models, the organization is still responsible for the entire stack.

## Traditional Workloads



- Persistent
- Stateful
- Monolithic
- Scale-up

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



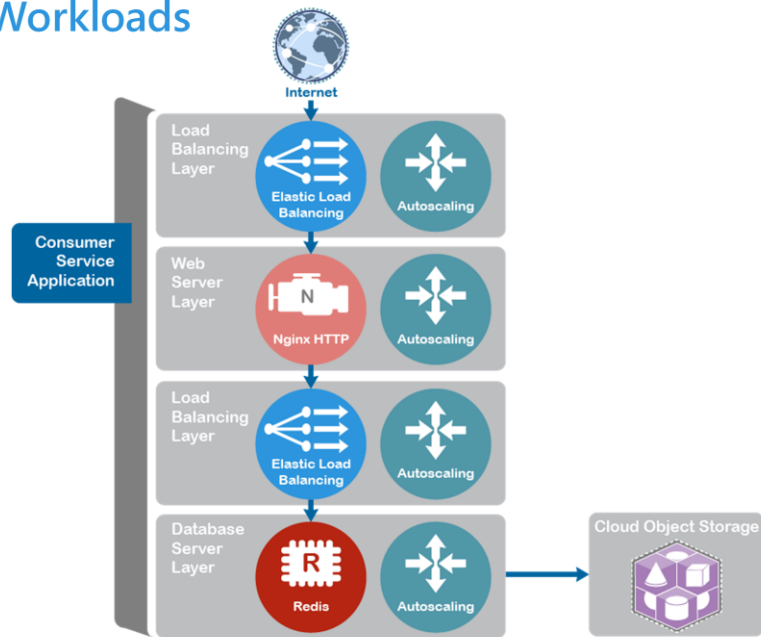
9

### Traditional

Traditional workloads are persistent, stateful, and monolithic. Although this workload runs in a virtual machine, it still maintains a client or application execution state in local memory or disk. These virtual machines must remain in an always-on state for the application to work properly. If a host running this virtual machine fails, the application must be repaired and may take time to recover. When traditional style workloads require better performance or capacity, the only option tends to be added resources to the existing virtual machine. This ultimately means that the infrastructure supporting this type of workload must also be able to scale up to support additional resources.

## Cloud-Native Workloads

- Disposable
- Stateless
- Modular
- Scale-out



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



10

Cloud-native workloads are disposable, stateless, and modular. These workloads do not maintain a local state. If a host fails, it may be easier to turn on another copy or redeploy the workload on a different host. When cloud-native workloads require more resources, they are usually scaled out horizontally by adding additional instances and include a load-balancing solution at the front-end. This means that the underlying infrastructure must also support a scale-out concept.

Many organizations are moving to cloud-native applications but this effort will take time. Therefore, a cloud design may need to support both types of workloads or it may focus on the cloud-native type. It is important to note though that public cloud providers may not provide the right infrastructure for traditional workloads and they may be confined to the private cloud.

## Discussion: Multi-Tenancy



A method for allowing one or more services to consume shared resources while maintaining a logical separation from other services.

.....

? Why does multi-tenancy matter to a cloud architect?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



11

One of the characteristics of a cloud is the enablement of pooled resources which promotes efficient and cost-effective use of infrastructure. What if those applications and services are consumed by different users or groups of users who don't wish to share their applications and data with each other? Multi-tenancy is a method for allowing one or more services to consume shared resources while maintaining a logical separation from other services. Multi-tenancy allows different consumers or groups of consumers to use services in a cloud but apply their own security, compliance, and business policies to their respective services.

For example, if two businesses wish to use services from a public cloud provider, these services would be logically isolated from each other so that neither has access to the other's applications or data. Each can use its own authentication mechanism, security controls, and compliance controls to secure and maintain their services but still share the underlying physical sources.

Why is this important to a cloud architect?

Maintaining isolation happens at all levels of the infrastructure. Here are examples of design considerations that relate to multi-tenancy:

- Compliance regulations or business requirements that may demand physical isolation
- Strict control of access the virtualization layer
- Strict control of access to orchestration tools
- Separation of services within the service catalog
- Implementation of network segmentation
- Implementation of firewalls

# Traditional Infrastructure

## Scaling



© Copyright 2015 EMC Corporation. All rights reserved.

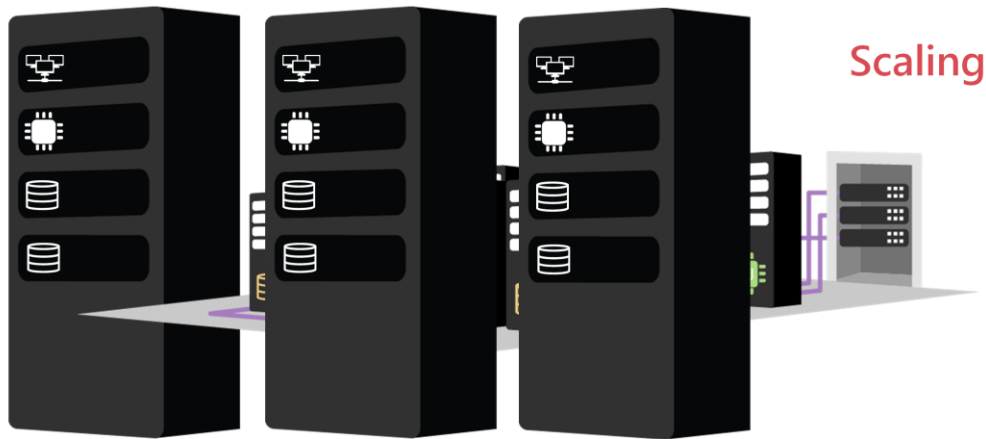
Cloud Infrastructure Planning and Design



12

Traditional infrastructure refers to an environment in which compute, network, storage, and even software are treated as separate entities. Each has its own way of being managed, has its own refresh cycle, upgrade path, and may have dedicated staff. Although the components connect to and work with each other, they are often treated in silos. For example, a datacenter may have racks of network equipment in one area, racks of servers in another, and storage arrays in a third. Each is managed, monitored, and refreshed separately. To add or expand resources, capacity is added to each of the resource types.

## Converged Infrastructure



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



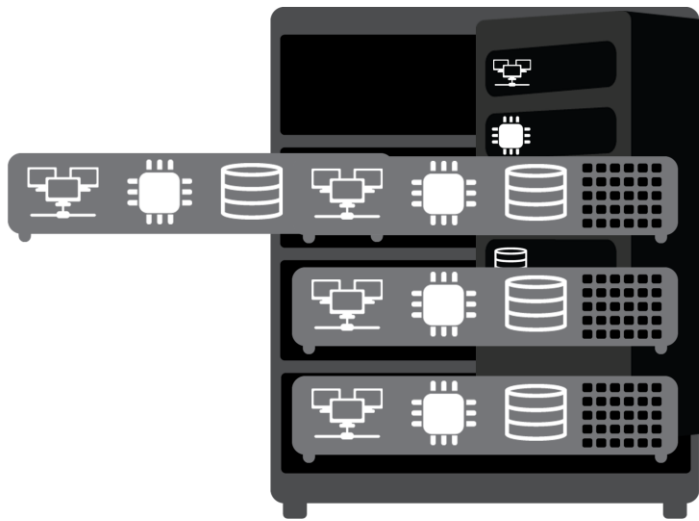
13

Converged infrastructure logically works like traditional infrastructure but the difference is that the physical components are more closely integrated or bundled. Compute, network, storage, and certain software are combined, pre-configured, occupy a more confined space (typically a small number of racks), and are treated as one entity. Converged infrastructure has a more rigid design, usually consists of equipment from a single or small set of vendors, and may have a refresh cycle that includes all of the components. The VCE Vblock is an example of converged infrastructure in which the components come from select vendors, and are assembled, pre-configured, and shipped as a unit. Capacity can be added in a modular fashion by adding more units of converged infrastructure as required.



# Hyper-Converged Infrastructure

## Scaling



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



14

Unlike the traditional and converged model in which separate compute, network, and storage nodes exist, in a hyper-converged infrastructure these functions are all physically located on each node but are more abstracted through software. Nodes are combined to present a distributed pool of compute, network, and storage and are operated as one entity. To add capacity in any one resource type, a node is added providing resources in all types but at smaller increments than in converged. VSPEX BLUE is an example of a hyper-converged infrastructure that combines technologies from EMC and VMware to provide linear scale-out, software-defined building blocks that can be used for cloud infrastructure.

## Greenfield vs. Brownfield Infrastructure

### Brownfield

- Reuse existing equipment
- Requires upgrades to existing firmware, software, and so on
- More upfront planning to shift existing workloads
- Staff familiarity with technology
- In-place support contracts

### Greenfield

- New equipment
- Avoids legacy impediments and constraints
- Enables migration to new technologies
- Higher cost
- Longer initial acquisition time



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



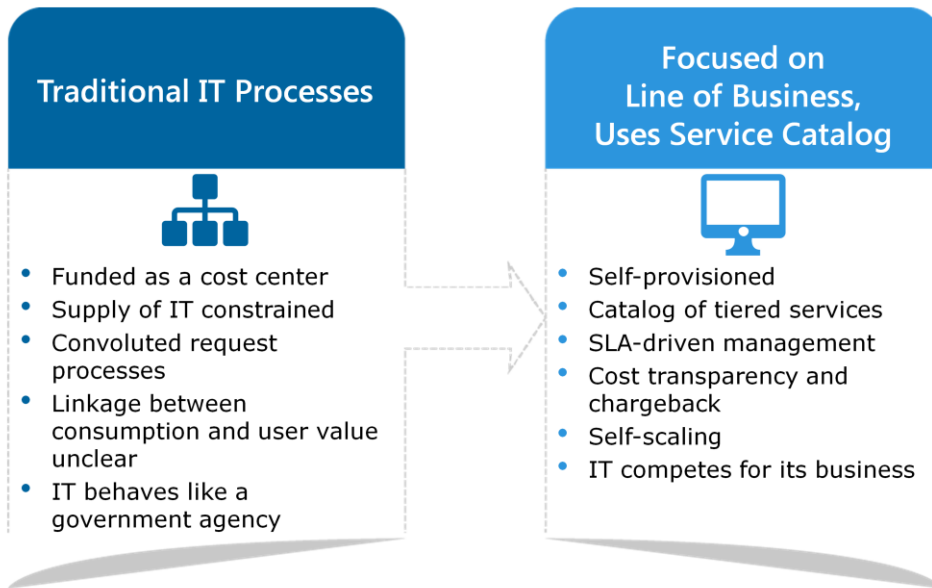
15

When used to describe cloud infrastructure, greenfield refers to building something new and brownfield refers to reusing existing infrastructure. Both options have their pros and cons.

Greenfield environments allow architects to design exactly what is required to meet the business needs using new infrastructure that is built specifically for a purpose. Greenfield environments can avoid some of the older and less efficient processes, rules, methods, misconfigurations, constraints, and bottlenecks that exist in the current environment. Greenfield environments also have the added benefit of allowing a business to migrate infrastructure to a different technology or vendor and to build in technologies that help avoid future lock-in. But greenfield environments also have some downsides, such as higher cost, lack of staff expertise, and possibly increased implementation time.

In a brownfield environment, architects can still design what is required to meet the business needs but using existing infrastructure. A benefit of using a brownfield environment is that existing staff most likely have the required expertise to support the environment and to implement it quickly. Brownfield environments usually cost less because the business is not buying as much equipment to support the new initiative. Brownfield environments come with downsides as well. For instance, existing infrastructure or processes may place extra constraints on the architect's design, and that may negatively impact performance or functionality. Another drawback to using a brownfield environment is that more effort may be required to upgrade existing infrastructure or migrate existing workloads so that infrastructure can be repurposed.

As you can see, making the choice to use a greenfield or brownfield environment takes some thought. The major factors that impact the selection are cost, time to implement, and even a business' willingness to adapt or change.



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

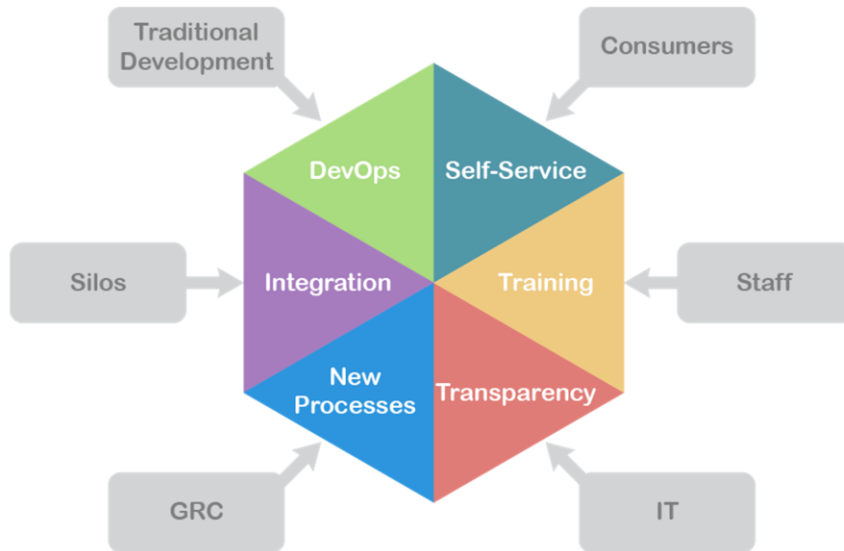


16

IT as a Service is not like other cloud service models in that you can't request a specific instance of it from a service catalog. ITaaS is a business-centric, transformational approach to providing IT services to an organization. It focuses on business outcomes such as operational efficiency, competitiveness, and rapid response which may improve costs as well as time-to-market goals. ITaaS shifts the role of IT from a cost center to a broker of strategic business value.

ITaaS defines and provides many services to the consumers of IT and uses the cloud infrastructure to help deliver these services. ITaaS changes the way in which IT consumers demand and utilize IT resources.

## Organizational Changes



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



17

Transforming from a traditional IT delivery model to a cloud-based service delivery model has a significant impact on an organization. The consumers are impacted because they are not only empowered to provision their own services but also are responsible for justifying their consumption. IT is impacted by the introduction of new technologies, processes, and procedures. They are forced to support these technologies in a more integrated manner, eliminating the legacy siloed approach. IT is no longer seen as a cost center and the focus of customer support changes since the organization moves to a self-service model. The entire organization is impacted because IT services are provisioned faster, enabling improved responsiveness to business challenges. These are just some of the changes that will be realized.

Even if your focus is strictly on the technical design, as an architect, you will influence the business transformation activities as well as be impacted by them. As the technical expert in the cloud platforms and tools, you will influence the organization through education and persuasion. During the design process, as you identify business-related dependencies or constraints, such as training needs or process changes, you will influence organizational change. However, you will also experience some of the downside impacts. For instance, during the assessment phase, you will interview stakeholders and SMEs who will be resistant to change, worried about their job, or just unfamiliar with the proposed environment. It will be up to you to educate and sell the cloud technologies and design in order to obtain a complete set of requirements.



GRC is an integrated approach or framework for addressing governance, risk management, and compliance issues within an organization. This framework helps ensure that an organization acts ethically and in accordance with its risk appetite, internal policies, and external regulations. A GRC framework should be integrated, holistic, and implemented organization-wide in order to effectively support business operations.

## Governance

- Organization
  - System of process and control
  - Determines direction
  - Defines goals
  - Ensures stakeholder best interests
- IT
  - Ensures alignment with organization
  - Addresses governance technology
- Source of requirements for cloud design



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



19

Organization governance is a combination of elements, such as policies, procedures, organizational design, roles, responsibilities, relationships, and so on, that defines how an organization should be managed and controlled, and determines direction and performance. Governance ensures that management makes decisions that are in the best interest of the organization's stakeholders. Stakeholders may include management, employees, suppliers, partners, investors, auditors and customers. Decision-making rights and accountability are distributed throughout an organization and the rules and procedures for making and monitoring those decisions are clearly defined.

IT governance is the subset of elements that guide management to monitor, evaluate, and direct IT operations to ensure organizational alignment, operational effectiveness, accountability, and compliance. IT governance also addresses the technology required to measure performance, mitigate risk, improve service delivery, and efficiently manage resources. IT governance does not involve just the IT department but also includes management and stakeholders from other groups within the organization.

Architects should be aware of an organization's governance system. When it comes to IT, governance bodies typically have concerns that include availability, security, continuity, recovery, compliance, quality, cost, and responsiveness. IT governance policies and guidelines will contain some of the root requirements for a cloud design and may offer more insight into areas such as performance measurement, auditing, security, procedures, and approval processes. The IT governance team members will also most likely be stakeholders that you need to interview during the assessment.

An architect also needs to understanding the concerns of the governance body and be able to communicate how a cloud infrastructure aligns with governance requirements.

## Risk Management



- Identify, control, and minimize undesirable outcomes
- Cloud technology and process introduces new risks
- Architect helps to identify risks
- Design includes mitigation strategy

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



20

Risk is the potential that a specific action or activity (including no action) will lead to an undesirable outcome for an organization. Risk management is the process used to identify, control, and minimize risks in an organization. Examples of risks related to IT include security breaches, loss of system or application availability, loss of data, or project failure.

An architect understands that even though an organization may have a fairly complete risk analysis and risk management system in place, the cloud introduces new potential risks that the organization must identify and address. Once risks are identified, the architect will work with the organization to create mitigation strategies and include these in the design.

The following list illustrates how a new cloud infrastructure can impact an organization's strategies for controlling risk and the questions that may need to be addressed in the design.

- Introducing multi-tenancy means that data may no longer be kept on a system dedicated to just one business unit or customer. How will we maintain data separation and protection?
- Introducing application auto-scaling means that application and infrastructure instances can be added and removed programmatically. What is the risk on performance or availability? What is the risk to exposing residual data when an instance is removed?
- Adding hybrid cloud capabilities means that data may flow across boundaries. Does the environment have adequate controls to assure that the data won't cross borders if it is not allowed to due to rules of law or regulation?
- Using a public cloud provider extends controls beyond the internal IT department. What policies and procedures does the organization have for this type of relationship? Does the provider also have policies and procedures that support the organization's goals for risk control?

# Compliance

- Conforms to regulations
- Requires ongoing feedback
- Affects IT operations
- Cloud design will include solutions
  - Security
  - Data recovery
  - Encryption
  - Monitoring



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



21

Compliance is a state of being in accordance with established guidelines, regulations, or legislation. Compliance involves identifying the proper regulations that apply to the organization as well as developing a system to ensure conformance and alert when something goes wrong.

Many times, these regulations that affect the organization will have a direct impact on IT operations. An architect may not need to be an expert in all of these rules and regulations but must take steps during the assessment process to understand which regulations do apply and develop a plan to address these.

Here are examples of where a cloud infrastructure may be impacted by regulations for compliance:

- Personal data
- Financial reporting
- Intellectual property
- Electronic discovery
- Document and record retention
- Sector or entity specific regulations such as
  - Payment Card Industry Data Security Standard (PCI DSS)
  - Sarbanes-Oxley Act (SOX)
  - Health Insurance Portability and Accountability Act (HIPAA)
  - EU General Data Protection Regulation (GDPR)



# Lesson 1

## Summary

Key points covered in this lesson:

- The cloud is defined by certain characteristics as defined by NIST.
- Services in the cloud support infrastructure, platform, and software.
- Cloud infrastructure will support traditional and cloud-native applications.
- Organizations have processes in place and will be undergoing change to provide better and more aligned services.



This lesson reviewed the baseline concepts for cloud and infrastructure technologies that are used throughout this course.

# Lesson 2

## Planning and Design Overview

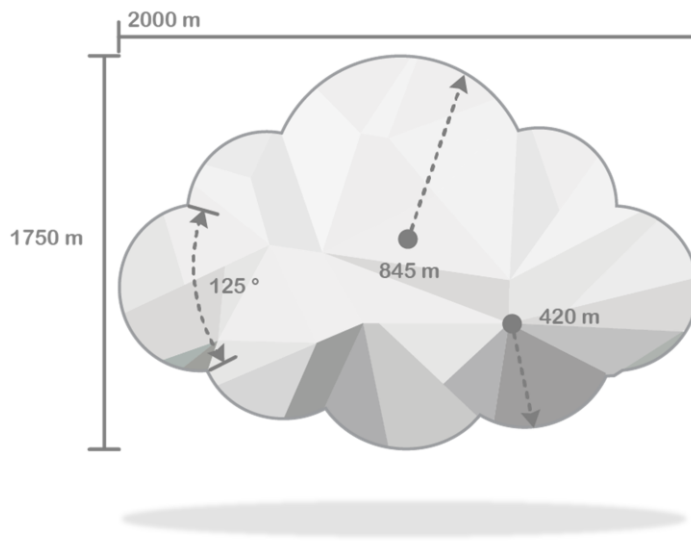
This lesson covers the following topics:

- Overview of the cloud design process
- Importance of requirements gathering
- Relevance of GRC process in requirements gathering



This lesson provides an overview of the design process and emphasizes the importance of requirements gathering to create a successful cloud design.

# Cloud Infrastructure Design



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



24

Cloud infrastructure design is the development of a set of guides that allows an organization to implement a cloud solution that meets its needs. Cloud designs are produced by architects who have the ability to gather and understand requirements, apply technical solutions that solve the requirements, and then document and communicate the entire solution.

Designing cloud infrastructure is not like designing the infrastructure for a traditional data center environment. Technology is constantly changing but more importantly, cloud computing changes the way that we deploy services, and a cloud architect must understand how these changes affect the outcome of a design. For instance, availability options are shifting from the underlying infrastructure to the application layer.

A good design ensures that a cloud is being created that addresses the needs of the organization. It also ensures that the resulting cloud infrastructure continues to be scalable, supportable, and functional after it is built.

## Design Approach

- Use standard framework and methodology
- Define goals and scope
- Assessment
  - Identify requirements, constraints, dependencies, and assumptions
- Research and select technical solutions
- Communicate
  - Assessment, reviews, education, and presentations
- Produce deliverables

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



25

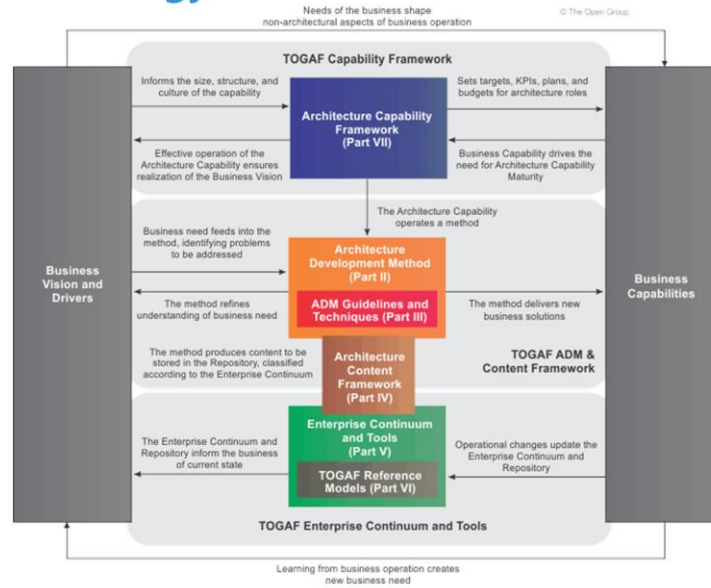
Creating a cloud infrastructure design is not just about drawing a diagram and creating a list of materials. It is a process. The process includes using a standardized methodology that can be applied across many projects and helps ensure consistent deliverables. The process also includes defining the end goals of the cloud environment and using that as a guide to create the design. In order to define the end goals, you must conduct a proper assessment of the business and environment to identify requirements and constraints that must be included in the design in order for it to be successful. The design process also includes researching and identifying technical solutions, understanding the benefits and considerations of implementing these solutions, and balancing the business requirements against each of these solutions.

Effective and frequent communication is important. During the assessment phase, you meet with stakeholders and subject matter experts. Here you need to know how to ask the right questions, explain your reasoning for asking the questions, and educate others about terminologies and technologies. The meetings continue into the design process as you need to review design decisions to gain validation and acceptance or to adjust requirements due to certain constraints. Finally, you present your final solution and defend your design decisions.

The design process includes producing certain deliverables. This includes documentation outlining requirements, constraints, and assumptions. It also includes the design and plan for implementing a cloud environment. All documents should be easy to understand, include relevant best practices, and define standards so that others can build and operate the solution. The final deliverables may also include a test and validation plan to ensure that the design meets the needs of the business.

# Framework and Methodology

## Example: TOGAF



<http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



26

A framework is a loosely-defined structure or set of guidelines that states how to do something. A framework may define a goal and the overall steps needed to produce an end result. Methodologies are a more rigid set of actions, processes, and rules that defines how to accomplish a task. They include a specific set of reproducible tasks that must be accomplished to obtain an end result. Frameworks may include methodologies.

Many consulting firms, large enterprises, and government agencies use their own standard frameworks and methodologies for architecture and design projects. Some generally accepted and open frameworks and methodologies are also available and even have their own training curriculum. The example shown here is called TOGAF (The Open Group Architecture Forum) and includes an open framework and methodology which can be applied to many enterprise architecture projects. It can be used directly or may be modified or combined with our frameworks and methodologies to address the needs of your organization.

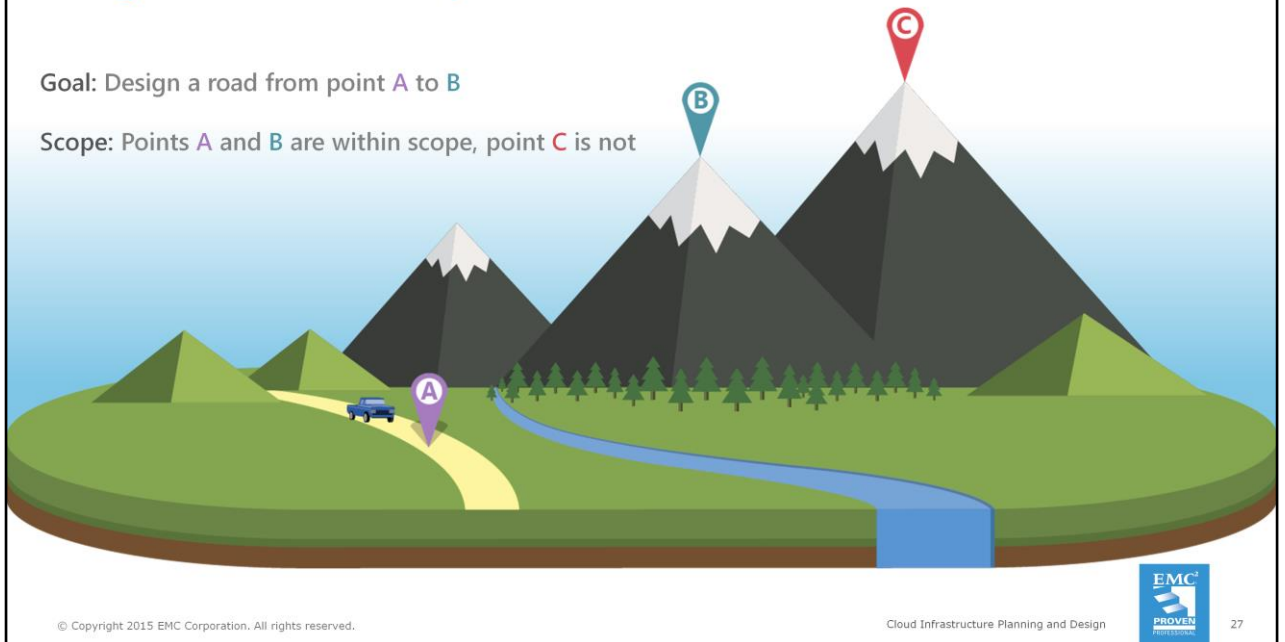
Whether you are an architect for a consulting firm or an individual organization, it is important to use a standard approach to guide you through the design process. A benefit of following a standard approach is the ability to produce consistent deliverables across all projects. Standard approaches provide guidance for assessments and requirements gathering to ensure that you identify the business needs completely and avoid the risks of a prolonged development time or an inadequate design. Using a standard approach also enables you to present a more organized and professional approach with customers and stakeholders.

Image Source: <http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>

## Design Goals and Scope

Goal: Design a road from point A to B

Scope: Points A and B are within scope, point C is not



Design goals describe the desired end result of a design and are used as guides during the design process. They represent the high level or general requirements of a design. When designing a road in the mountains, an example of a goal would be that the road must go from point A to point B.

Examples of design goals for a cloud would be:

- We would like to provide our developers the ability to provision infrastructure on demand.
- We wish to reduce the datacenter footprint.
- We would like to offer all of our customers the ability to use our application in a secure and cost effective manner.

Goal statements offer a broad direction for the project and provide the architect with the knowledge of where the organization would like to go. Goals are the starting point for gathering more specific requirements, and all of this information is used to ensure the final design meets the organization's needs.

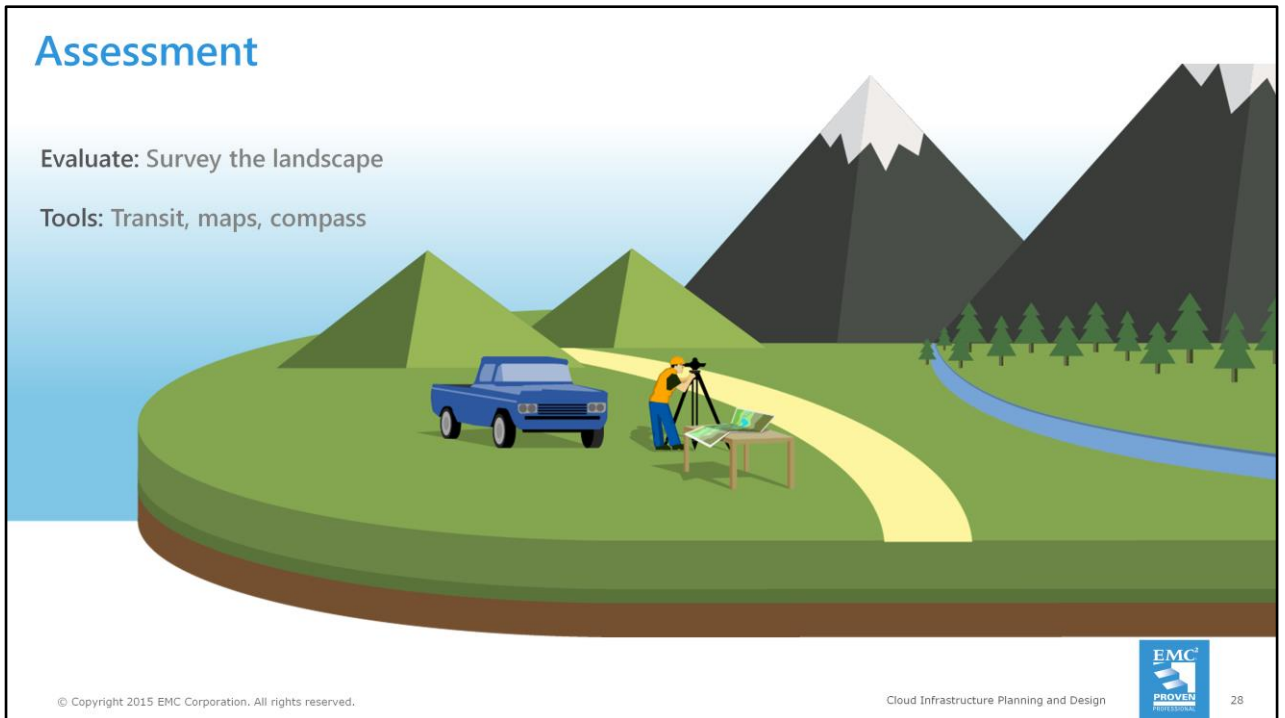
The scope of a design is the boundary that defines what should be included in the design and what should not. In the mountain road design, we see a statement that point C should not be located on the road.

In a cloud infrastructure design, you may define the scope where you may state that the design includes the compute, network, and storage required to build it. However, it does not include the design of a new datacenter that is needed to house it, nor does it address the organizational change required to support it.

## Assessment

Evaluate: Survey the landscape

Tools: Transit, maps, compass



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



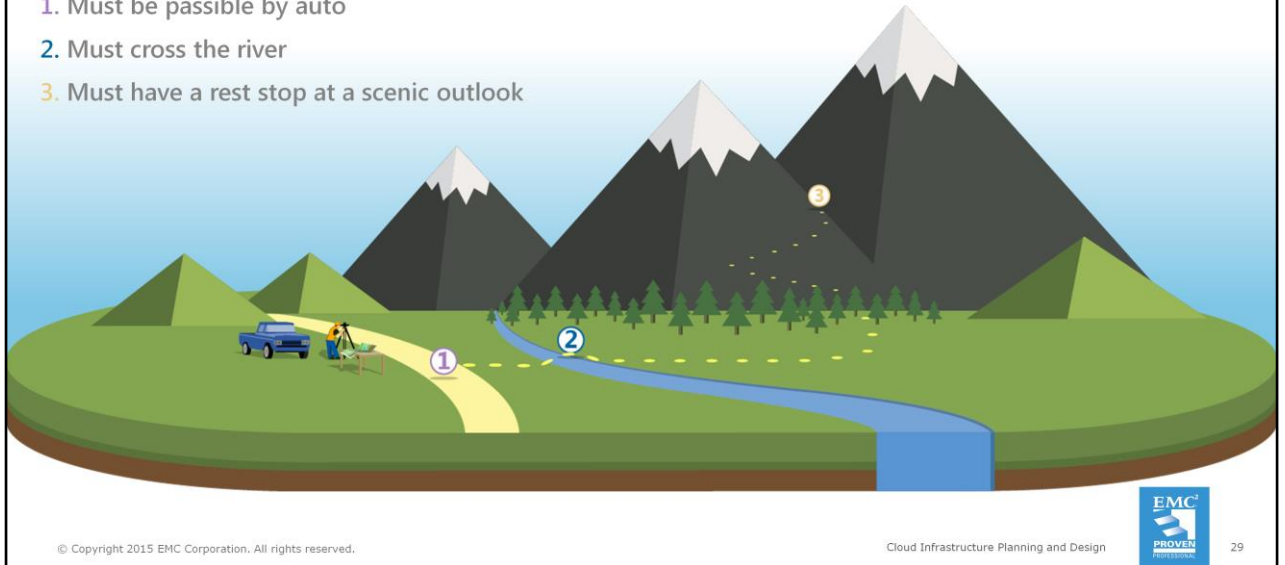
28

An assessment is an in-depth evaluation of something. In our mountain road example, it could include surveying the landscape with purpose built tools. In a cloud design process, an assessment is an evaluation of the organization and its assets as they relate to the cloud design project. It is during the assessment that you clarify and define the project goals, requirements, and constraints that are used to guide your design to meet the business' needs. You examine business goals, processes and policies, project goals and requirements, IT processes and abilities, and the current infrastructure design and capabilities.

The assessment will require gathering information from existing documentation such as business policies, operational procedures, and architecture designs as well as through interviews and meetings with subject matter experts and stakeholders.

# Requirements

1. Must be passible by auto
2. Must cross the river
3. Must have a rest stop at a scenic outlook



Requirements are desired characteristics or behaviors that the cloud environment should possess as dictated by an organization. They may include specific functionality or performance characteristics to be addressed in a design. In the mountain road example, requirements dictate the functions and layout of the road. For example, it must support usage by a normal passenger automobile. Another example is that the road must cross the river rather than go around it.

Examples of cloud infrastructure design requirements are:

- The cloud should provide 99.99% uptime for “critical” services
- Consumers should be able to access the cloud using their Active Directory credentials
- Sarbanes-Oxley compliant services must be supported

In order for a design to be considered successful, the requirements must be included and addressed. Otherwise the design will not meet the business’s goals.



## Requirements Gathering

- Listen
- Ask questions
- Prioritize requirements
- Review requirements before starting the design
- Obtain acceptance
- Continue to include stakeholders throughout the design process

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



30

Requirements gathering is a critical step of the design process. An important part of proper requirements gathering is to listen to the stakeholders and SMEs because they are responsible for creating the requirements to be used in your design. While you are listening, ask questions. Ask for further clarification and don't always assume that you know what the organization wants. Collect and review the requirements with the organization and ensure that they provide acceptance or approval before you start the design. As the design progresses, you may hit constraints that prevent you from fully meeting a requirement. This is normal but it is important to include the organization in the process when this happens so that they have input into the design decisions or changes. Prioritize the requirements so that you know what is an absolute must-have and what is a nice-to-have. This will also help if you run into any constraints.

## Requirements Focus for Cloud

- Access
  - Authentication
  - Security
  - Performance
  - Availability
  - Compliance
  - Applications
- Self-service
  - Service types
  - Orchestration
  - Automation
  - Business processes
  - Multi-tenancy

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



31

Listed here are some of the focus areas for the requirements gathering process. Although many of these are not specific to cloud, some are, and others have new meaning in cloud environments. Many of these areas are addressed throughout the course.

## Discussion: Gathering Access Requirements



An organization is planning to implement Infrastructure as a Service in a cloud. The stakeholders have stated that all users of the new cloud infrastructure will be internal employees of the company. Developers will use the IaaS instances to deploy Web services.

- .....
- ❓ What questions should you ask for clarification about network access requirements?
- ❓ How would the answers impact a design?



## Review: Network Access Requirements



An organization is planning to implement Infrastructure as a Service in a cloud. The stakeholders have stated that all users of the new cloud infrastructure will be internal employees of the company. Developers will use the IaaS instances to deploy Web services.

### Possible Questions:

- ❓ How will consumers access the IaaS instances?
- ❓ Where are the employees located?
- ❓ Who will access the Web services?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



33

The scenario presented here provides some detail and exposes some business requirements. However, in order to create a cloud design to support the scenario, you need more information.

For further clarification on the first statement, as an architect you might ask, “How will consumers access the IaaS instances?” This is important because once you deploy an IaaS instance, there can be multiple methods to access the instance. Consumers could use the service catalog capabilities, SSH, Remote Desktop, X-Windows, and so on, and each of these may require changes to the design to support them.

The scenario also states that IaaS consumers are internal employees. One question to ask is, “where are the employees located?” This is important to know because your design may require an access solution such as VPN for remote employees or maybe there is already a remote access solution in place that may cause a constraint on your design.

The third part of the scenario needs further clarification because it isn’t stated who will access the web services. If the web services will be accessed by external customers, then this will require a solution that grants access from the internet and may involve network segmentation and firewall technologies.

## Discussion: Gathering Security Requirements



An organization is planning to implement Infrastructure as a Service in a cloud. The stakeholders have stated that all users of the new cloud infrastructure will be internal employees of the company. Developers will use the IaaS instances to deploy Web services.

- .....
- ❓ What questions should you ask for clarification about security requirements?
- ❓ How would the answers impact a design?



## Review: Security Requirements



An organization is planning to implement Infrastructure as a Service in a cloud. The stakeholders have stated that all users of the new cloud infrastructure will be internal employees of the company. Developers will use the IaaS instances to deploy Web services.

### Possible Questions:

- ❓ Do you require multi-tenancy?
- ❓ How will employees authenticate?
- ❓ What information is stored on the Web services?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



35

The scenario presented here provides some detail and exposes some business requirements. However, in order to create a cloud design to support the scenario, you need more information.

For further clarification on the first statement, you should ask whether the IaaS instances will be used by one group or many. As the architect, you need to understand whether a multi-tenant solution is required. Although resources will be shared across the organization, a logical separation may be needed for each business group or business function. If multi-tenancy is required you may need to enable separate catalog instances or authentication mechanisms.

The scenario also states that IaaS consumers will be internal employees. You might ask how employees are expected to authenticate with the operating systems within the instances? From a design standpoint, this means not only configuring authentication for the service catalog but also enabling the proper authentication mechanisms for the IaaS instances as well.

Again, the final part of the scenario needs further clarification because it does not state what information will be stored there. From a security standpoint, this is important because if the web services contain customer data or regulated data, this may involve creating a separate trust zone with isolated resources, separate authentication mechanisms, or possibly even implementing compliance enforcement capabilities.

## Discussion: GRC Requirements



- Line of business stakeholders need to be accountable for IT resources that they consume and they need to understand associated costs.
- The organization will lose millions of dollars per day if customers are unable to place orders online.
- Only authorized persons may access electronic protected health information.

.....

 How could a cloud infrastructure design address these requirements?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



36

Implementing an integrated GRC program is beyond the scope of this course. However, a newly planned or existing GRC program will be a source for requirements when designing a cloud architecture. These requirements are not usually about user functionality but are related more to availability, business continuity, security and business operations.

Listed on this slide are requirements relating to GRC. Below are some possible solutions that address these requirements in a cloud design.

### Governance Requirement

In the design, include a service catalog that supports multi-tenancy and self-service capabilities. Include components that present cost information, usage statistics, reporting capabilities, approval processing and alerting.

### Risk Requirement

In the design, include orchestration capabilities which allow applications to scale out. Distribute consumer resources so that instances do not all exist on the same infrastructure.

### Compliance Requirement

Create a design that includes proper authentication methods, isolation mechanisms, and encryption technology for infrastructure that contains protected data.

# Application Requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

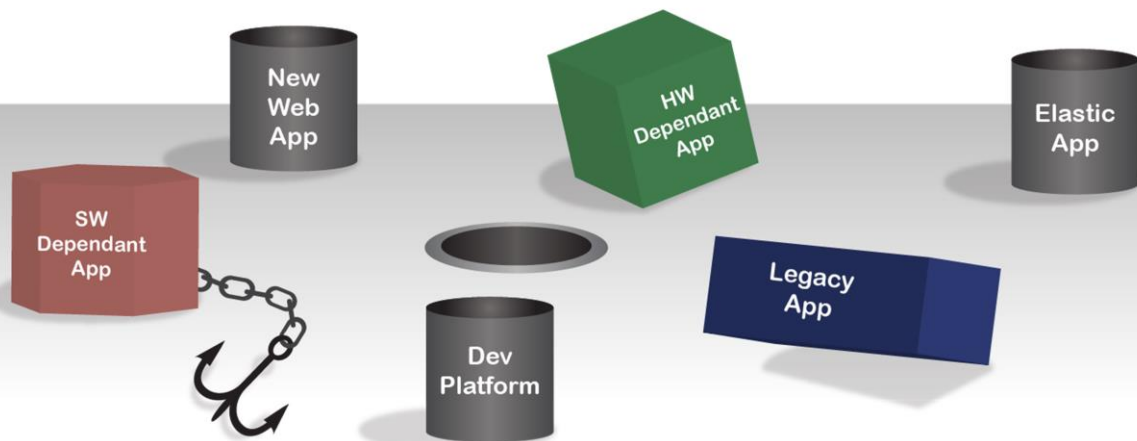


37

Application assessment is one of the most critical steps of the cloud design process. With the help of the architect, the organization has to identify which applications will be made available in the cloud. Like in a traditional environment, the architect determines the performance characteristics, availability requirements, security requirements, and so on to design infrastructure that will support the applications. But unlike the traditional environment, the cloud introduces functionality that will require additional considerations for applications. For instance, since the cloud provides a self-service model, does it make sense to include the application? If so, does it need to be redesigned or enhanced to support self-service? Cloud also supports multi-tenancy, so will the design require additional solutions to protect access to the applications and their data? If a hybrid cloud solution is planned, can the application be moved between two cloud infrastructures, and what is required to maintain performance in this distributed model? Understanding applications and how they relate to cloud models and technology is a critical skill for a cloud architect.



## Application "Fit"



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



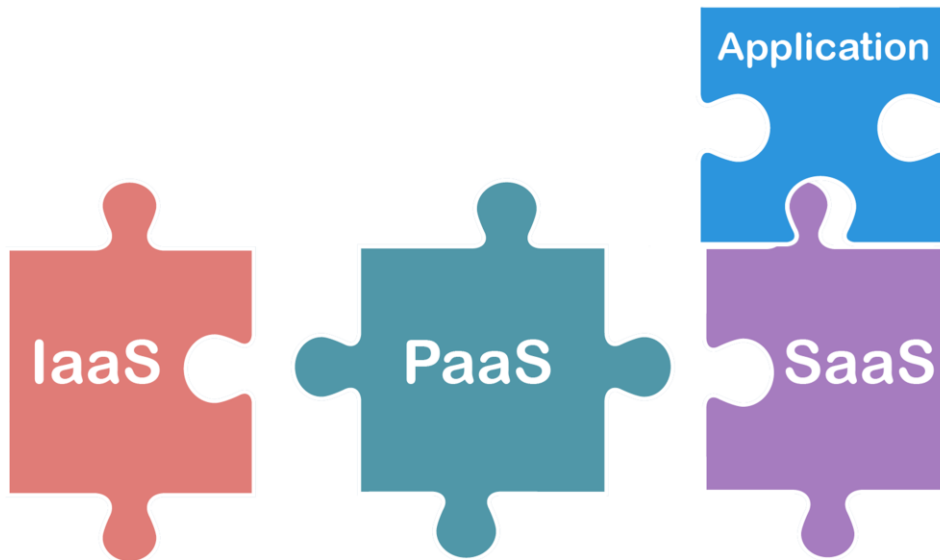
38

Not all applications fit or will work in a cloud model. If the organization requires that an existing application be placed in a cloud infrastructure, the architect needs to look at the application and determine whether the cloud design can accommodate the application. If not, the requirement needs to be reconsidered or the application may need to be redesigned. An example of an application that does not work well in a cloud infrastructure is a large, monolithic, legacy application. This typically only scales well by running on bigger hardware or has a low tolerance for component outages. Another example of a cloud "unfriendly" application is one that depends on specific hardware, operating system, or a driver that can't run in a cloud infrastructure. A third example would be a case in which an application requires a connection to another application that is run in an environment unavailable to the cloud infrastructure. It may not make sense to move it to a cloud, or it may require additional design additions that enable it to work.

Many applications in the cloud are designed with the assumption that things will break. Instances of an application that fail may be removed and replaced with a new instance. Applications are also designed to horizontally scale-out and scale-back as needed. Helping the organization understand these concepts is part of the architect's job. It also means that the architect needs to understand that some of the older ways to build infrastructure are no longer required when creating a cloud design.

A final issue to consider is whether it even makes sense to move an application into a cloud. If an organization uses an email system such as Microsoft Exchange to run the business, it would not make sense to move the Exchange servers into a private cloud since the users will never be requesting exchange servers from a catalog. However, it may make sense to use a public cloud provider to house the Exchange servers if the organization wishes to maintain a small footprint in their datacenter. In that case, the organization becomes a consumer of IaaS instances from a public cloud catalog. An alternative could also be to purchase email as a service from a public SaaS provider.

## Service Requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



39

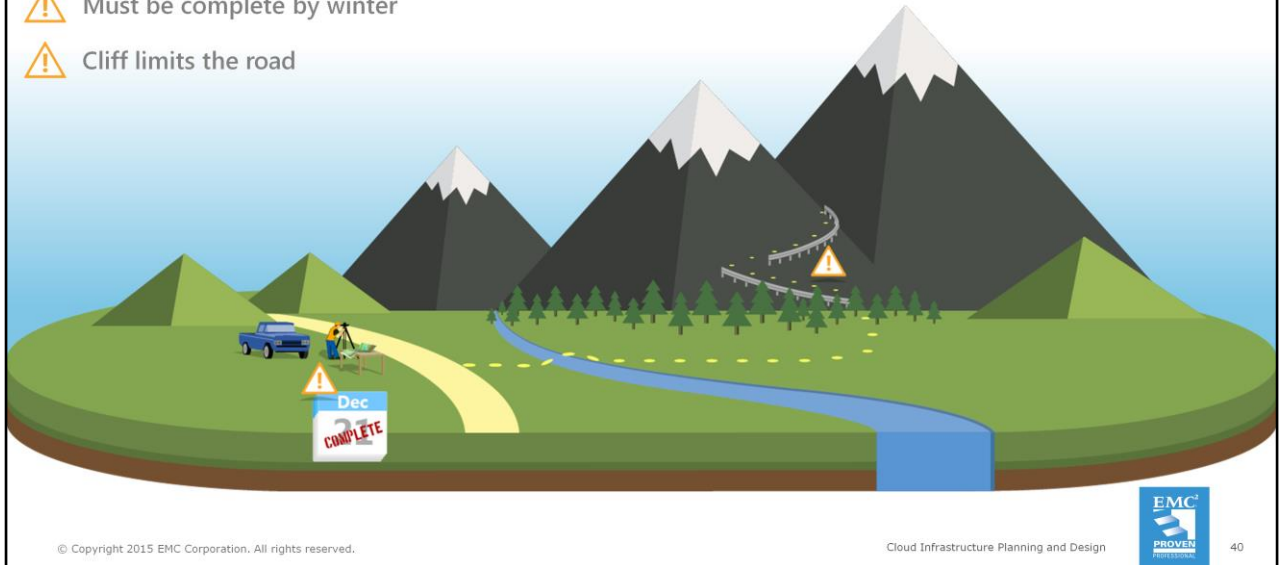
Understanding the nature of the different cloud service models helps an architect design the appropriate cloud resources to support them. But just as important is applying the correct service model to a planned service. For example, an organization may state that it wishes to implement a PaaS solution that allows the organization to develop and deliver web applications. As an architect you should consider the use of a public PaaS provider. Using a public PaaS provider may give the organization the flexibility it needs to develop and test the application, but when deploying an application that can be resource intensive or latency sensitive, the organization may not have the ability to tweak infrastructure components to maintain performance. An alternative may be to use a public IaaS provider and layer PaaS on top so that the organization has more control of the entire stack. Deploying PaaS in a private cloud could also be a solution since the organization will control the entire stack as well.

When gathering requirements and selecting solutions to support cloud service models, consider the following:

- Select a PaaS solution that supports the programming language and tools that match the organization's requirements
- Select a PaaS or IaaS solution that supports the operating system the organization plans to use
- Identify opportunities to use an outside provider for certain service models
- If a hybrid cloud model is anticipated, determine whether a service model solution can be used to deploy services in both a public and private cloud
- Know the requirements for the final services as well as requirements throughout the service delivery lifecycle

## Design Constraints

- ⚠ Must be complete by winter
- ⚠ Cliff limits the road



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



40

Design constraints are items or conditions that may limit your design choices. In a mountain road design, the entity responsible for building the road may put a time constraint on the completion of the project. Physical limitations, such as a cliff can also place constraints on where the road must be placed.

In a cloud design, constraints may be related to a business process or policy, budget, physical space, project timeframe or a technical limitation. Design constraints may be provided upfront by stakeholders or may be derived during the assessment process.

Examples of cloud infrastructure design constraints are:

- The business has a relationship with a specific vendor and their products must be used
- The overall budget for this project is \$1 million
- The datacenter has only 20 sq. meters available for new infrastructure

## Design Assumptions

1. The road will not have sidewalks
2. Road will not be plowed in the winter



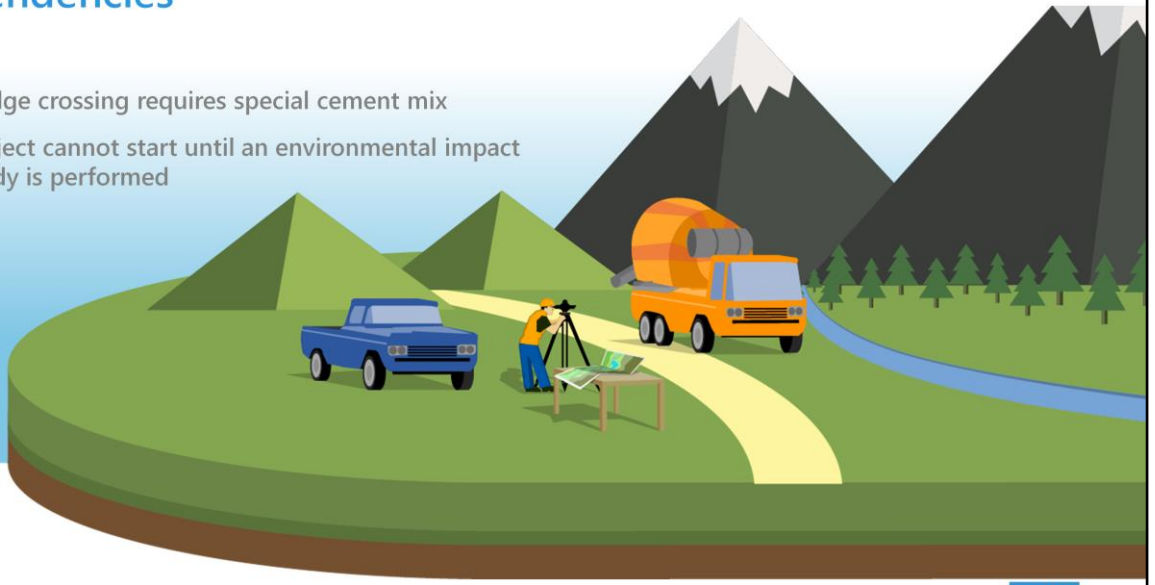
Assumptions are beliefs or expectations that an architect uses while developing a design. They may be used in the absence of requirements and constraints to clarify or identify design choices. However, it is a good idea to review assumptions with stakeholders to ensure that everyone is in agreement. Since the road in our example is being built in the wilderness, the architect may assume that sidewalks are not required nor will it be plowed, since all other roads in the area are not plowed.

Examples of cloud infrastructure design assumptions are:

- Sufficient power will be available in the data center for the new cloud infrastructure
- IT staff will receive relevant training before the infrastructure goes into production
- The latest versions of infrastructure software will be used

## Dependencies

1. Bridge crossing requires special cement mix
2. Project cannot start until an environmental impact study is performed



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



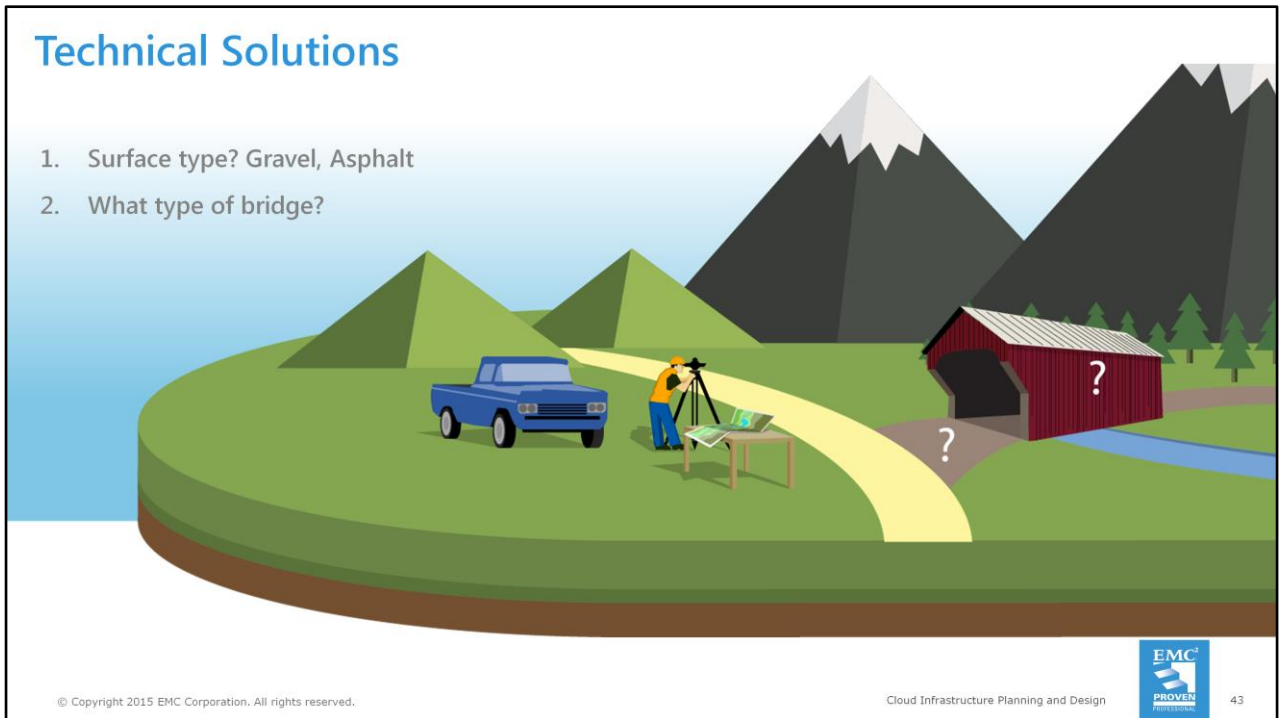
42

Identifying dependencies is an important part of gathering requirements. Dependencies are technologies or processes that a solution or project relies on to work fully. For instance, in order to place a bridge foundation in water, it requires a special cement mixture to keep it from crumbling. A project dependency may be that an environmental study must be performed before construction begins.

In a cloud design example, let's say that an organization wants to deploy an elastic web application on the cloud infrastructure. It should be obvious to you that this may require orchestration capabilities and a network load balancer. However, some questions you also need to ask are: how will users authenticate? Will these web servers need to connect to a common backend database? For authentication, the web servers may need access to one or many active directory or LDAP services. This, like the database dependency, may require adding firewall capabilities, VPN capabilities, network segments, or additional bandwidth.

## Technical Solutions

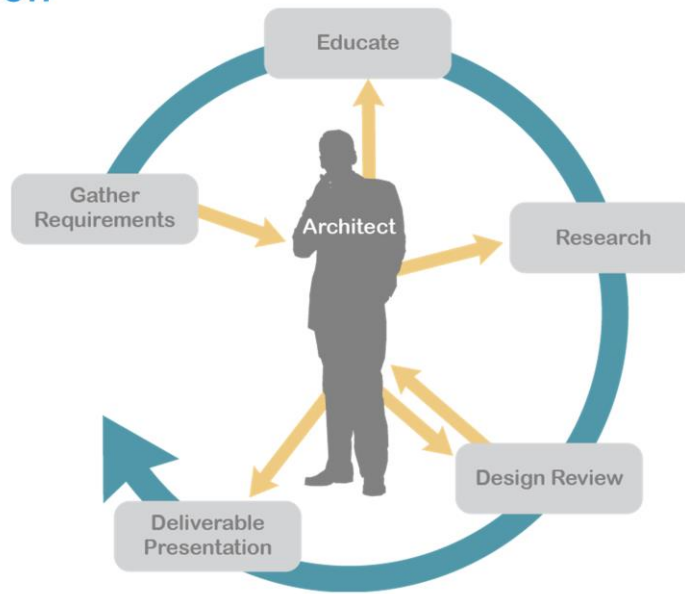
1. Surface type? Gravel, Asphalt
2. What type of bridge?



An architect must understand the various technical solutions and options that are available to create a design. For road design, the road can be made of many materials and the bridge may have various support types. An architect will match the technical solution with the requirements to determine what goes into the design.

Although this course will introduce some technical solutions, as an architect, you will be required to maintain a working knowledge of cloud infrastructure technologies. Since this environment changes frequently and business requirements will not all be the same, you should expect to do additional research during the design process. Another job of the architect is to explain to stakeholders how the technical solutions work and meet the requirements. Discovering how technical solutions can fulfill the business requirements may require discussions with vendors, reading product documentation, conducting internet research, or reaching out to others in your professional network.

## Communication



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

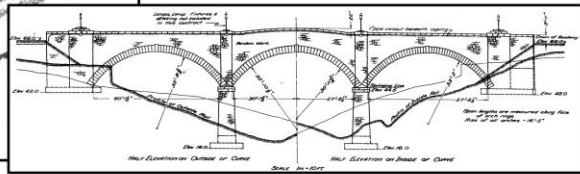


44

The ability to effectively communicate in a variety of modes is a key quality of a cloud architect. At the beginning of the process, you employ effective listening and questioning skills to gather requirements. During the assessment phase, you find yourself teaching stakeholders about cloud methods and technologies. Once it is time to research technologies, you begin to communicate with different vendors and their professional networks. Design review is an iterative process in which you explain design decisions, clarify requirements, and teach terminology and technology. Finally, you deliver the final design as well as an overview presentation at which you defend decisions and answer questions. During this entire process you work with all levels of stakeholders from management to staff, technical to non-technical, and accepting to skeptical.



## Deliverables



Cloud Infrastructure Planning and Design



45

The final step in the design process is to produce documents that support your design. These documents can include logical designs, physical designs, bills of materials, operational procedures, and recommended standards. In a road design, this may include overhead maps, cross-sectional views of the surface, bridge diagrams and lists of building materials or equipment required.

In a cloud design, these will include not only documentation about the infrastructure but also definitions for cloud internals. For instance, the portal and service catalog should support multi-tenancy and role-based access. As the cloud architect, you will need to create standards for naming roles, defining the roles capabilities, assigning or mapping roles to individuals, applying roles to tenants, assigning tenants, and so on. Another example would be naming conventions used for hosts, virtual machine instances and storage pools.

Deliverables can be very large and very detailed documents. Exactly what is to be delivered and what level of detail is needed is defined during the requirements gathering phase. Additionally, you should expect that the organization will want to meet to discuss your design and you will likely create a presentation to be used during this meeting to help explain your ideas.



# Logical vs. Physical Designs

Click within the play area to launch the video

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



46

# Lesson 2

### Summary

Key points covered in this lesson:

- The design process uses standard methodologies to define goals and scope, conduct research and assessments, and create design documents.
- Requirements come from many sources including an organization's GRC processes and procedures.
- The architect must understand the entire design process and understand the importance of requirements gathering.



This lesson covered an overview of the design process and the importance of requirements gathering.

## Lab 1: Introduction to Cloud Design

### Purpose:

Discuss and apply the concepts and definitions as presented in the lecture portion of Module 1: Introduction to Cloud Design.

### Tasks:

- Define key concepts for cloud technologies and terminologies.
- Define key concepts in the cloud design process.
- Apply assessment concepts discussed during lecture to the lab case study.

## Lab 1 Debrief

### Introduction to Cloud Design

**Purpose:**

Discuss and apply the concepts and definitions as presented in the lecture portion of Module 1: Introduction to Cloud Design.

# Module 1 Introduction to Cloud Design



Key points covered in this module:

- Services in the cloud support infrastructure, platform, and software.
- Cloud infrastructure can support traditional and cloud-native applications.
- The design process uses standard methodologies to define goals and scope, conduct research and assessments, and create design documents.
- Requirements come from many sources, including an organization's GRC processes and procedures.



Cloud Infrastructure Planning and Design

50

This module covered cloud introductory topics as well as an overview of the cloud design process.

This slide intentionally left blank.





# Module 2

## Cloud Management



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges, and considerations of implementing various cloud management platforms and the underlying technologies.
- Align business requirements when selecting cloud management technologies.
- Develop a design and plan for deploying components that support cloud management and operations.

Cloud Infrastructure Planning and Design



1

This module focuses on the design decisions and considerations for building a cloud management platform.



# Lesson 1

## Cloud Management Platforms

This lesson covers the following topics:

- The cloud management platform and the components that may be included
- Key requirements for selecting a cloud management platform



This lesson covers components of a cloud management platform and the requirements that contribute to the selection process of these components.

# Cloud Management Platform



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

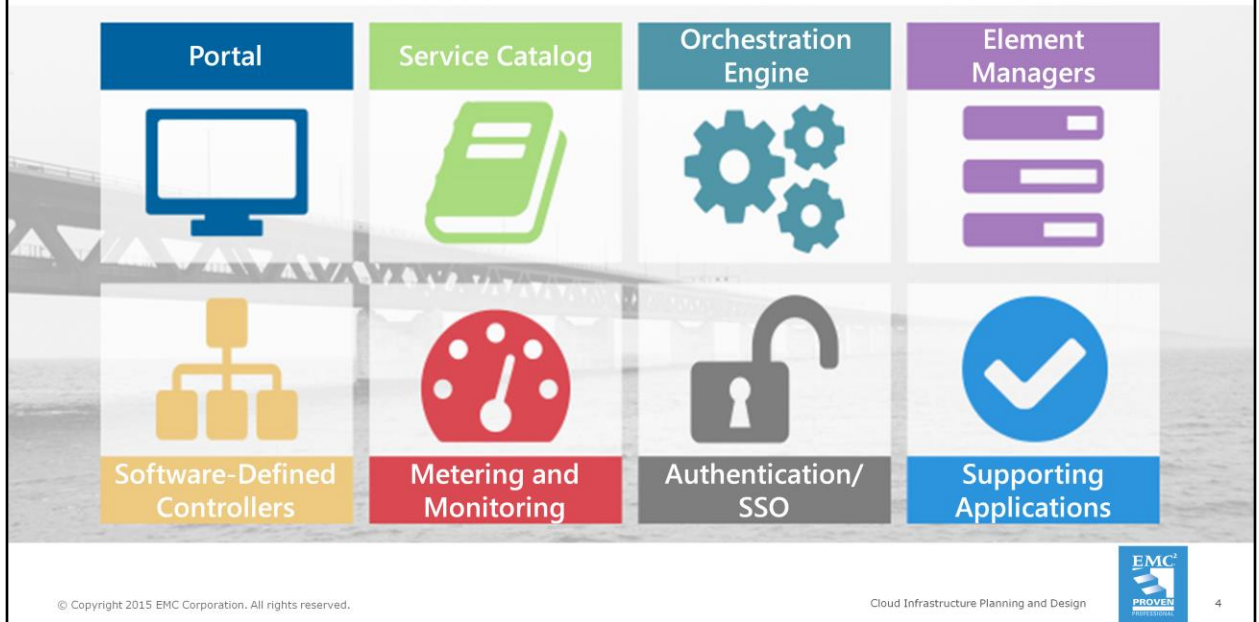


3

A cloud management platform (CMP) is a set of integrated components that are used to manage your cloud. The CMP may contain a portal, service catalog, orchestration engine, metering capabilities, and authentication mechanisms. The cloud management platform may also include other capabilities that support integration with software-defined controllers, element managers, enterprise systems, and other cloud platforms. Since definitions surrounding cloud management can vary, for the purposes of this course all of these components are considered to be part of the cloud management platform.

Cloud management platforms come in many flavors and have various capabilities. Some are vendor-specific such as VMware's vCloud Director Suite, or Microsoft's System Center Suite while others are more open such as OpenStack or Apache CloudStack. CMP components are used to deliver Infrastructure as a Service, Platform as a Service, and Software as a Service capabilities for consumers. OpenStack, as an example, provides Infrastructure as a Service but may be combined with Pivotal Cloud Foundry to provide Platform as a Service as well.

## Components



For the purposes of this course, a cloud management platform includes the following components:

- Portal
- Service catalog
- Orchestration engine
- Element managers
- Metering and monitoring
- Authentication/SSO
- Software-defined controllers
- Supporting applications

# Portal and Service Catalog



Federation Enterprise Hybrid Cloud self-service portal using VMware vRealize Automation

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

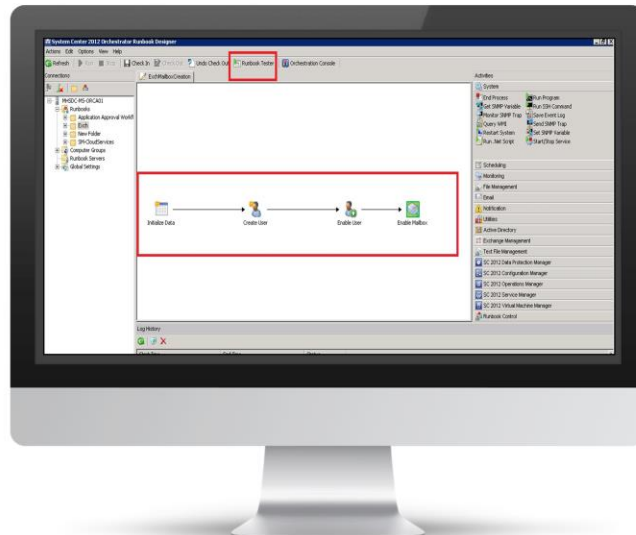


5

The portal is the primary entry point for cloud consumers, and the service catalog presents a list of available services. Together, these components enable on-demand, self-service capabilities for a cloud. They promote the alignment of IT services with business goals and requirements. Through multi-tenancy and role-based access, you can allow groups of consumers to view one set of services while other groups see a different set. The service catalog is also a vehicle that is used to control costs and manage demand for services. Rather than treating every IT request individually, a service catalog allows IT to provide standardized, best-fit services that reduce the time it takes to deliver the service. Because the catalog includes pricing information and chargeback integration, consumers are influenced to purchase only the services that are necessary. The service catalog also enables the application of quotas on resources, preventing overconsumption. Additionally, the use of service catalogs helps ensure that services meet the governance policies and standards set by the business, as well as IT best practices.

Shown here is an example of a portal displaying the service catalog of a Federation Enterprise Hybrid Cloud which uses the VMware product call vRealize Automation.

# Orchestration



Microsoft System Center 2012 Orchestrator Runbook Designer

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



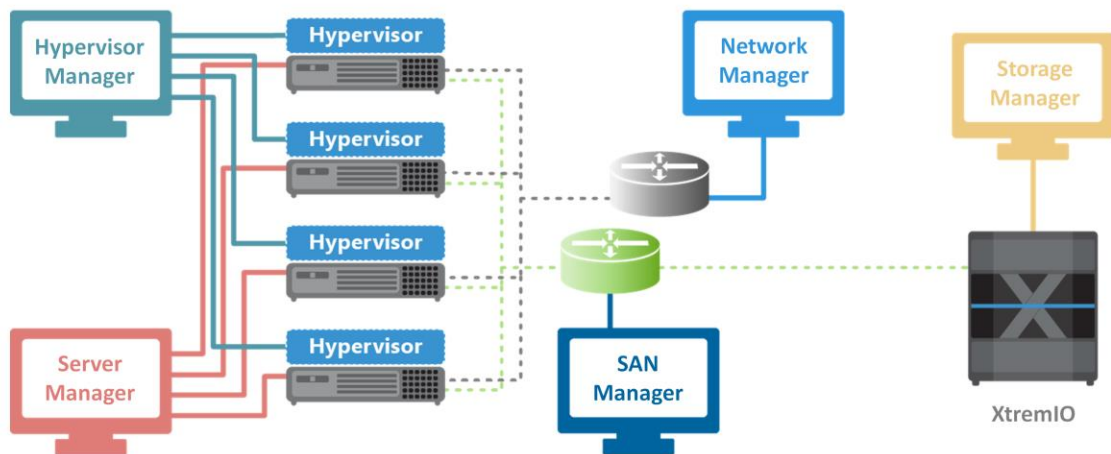
6

In the cloud, orchestration is the planned automation of tasks which are defined by workflows and business rules and are used to deliver services. A service catalog without orchestration is just a list of services. To implement an on-demand self-service model, you need to implement a set of tools that can process the actionable requests from consumers. The orchestration layer is where rules are defined in the form of workflows and it uses APIs to talk to various components to execute commands or run automated processes. Depending on the cloud management platform, orchestration can be initiated by various components within the stacks. For instance, in the EMC Federated Hybrid Cloud solution, IaaS capabilities are orchestrated using VMware vRealize Automation but other services can be orchestrated by integrating VMware vRealize Orchestrator. Shown here is Microsoft System Center 2012 Orchestrator which is used in the deployment of Microsoft clouds.

Common orchestration tasks include:

- Provisioning (VMs, storage, accounts, and so on)
- Configuration management
- IT automation
- Reporting

## Element Managers



© Copyright 2015 EMC Corporation. All rights reserved.

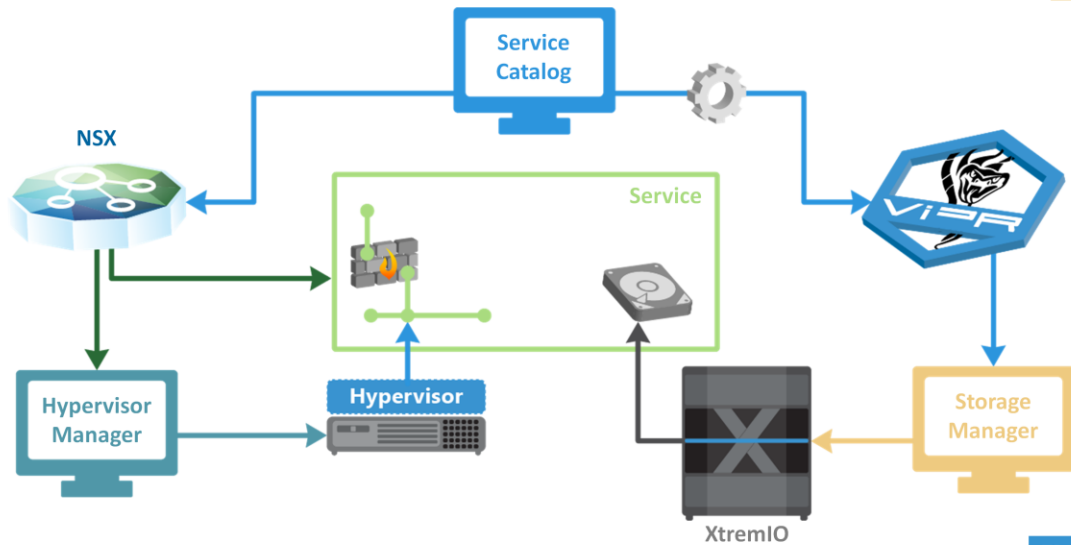
Cloud Infrastructure Planning and Design



7

Element managers are the management interfaces for infrastructure services. They may consist of a separate or integrated web interface or a Command Line Interface (CLI) used to manage virtualization, storage, compute, or networking components. Element managers are used to configure physical infrastructure, enable interfaces, manage user access, define policies, assign IP addresses, and so on. The element manager interface is usually proprietary or at least unique to a specific vendor's infrastructure components.

# Software-Defined Controllers



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

Software-defined controllers are used for provisioning infrastructure resources but function more universally than element managers. They are applications that may have a web interface for users, but definitely have an API that can be used by cloud orchestration processes. Many times software-defined controllers have plug-in capabilities that enable provisioning across multiple infrastructure components and multiple vendors. The software-defined controllers most likely run in the infrastructure dedicated to your cloud management platform since they are a critical part of infrastructure provisioning.

An example of a software-defined controller in the storage domain is the EMC ViPR Controller. It can be used to programmatically provision storage on multiple EMC solutions, other vendor solutions, various SAN switches, multiple storage types, and multiple storage protocols. ViPR provides one set of APIs to provision storage while using various plug-ins of the different storage platforms to complete tasks. In the Federation Enterprise Hybrid Cloud, VMware's vRealize Orchestrator integrates with ViPR to provision cloud storage.

Another example of a software-defined controller in the network domain is the VMware NSX controller. It can programmatically provision virtual network segments and services such as firewalls, load balancers, and routers. NSX also works with services from different vendors. It is also found in the Federation Enterprise Hybrid Cloud solution.



- Chargeback
  - Resource usage
  - Billing
- Performance
  - Services and infrastructure
  - Alerting
- Logging
  - Infrastructure events
  - Security events
  - Problems and errors

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



9

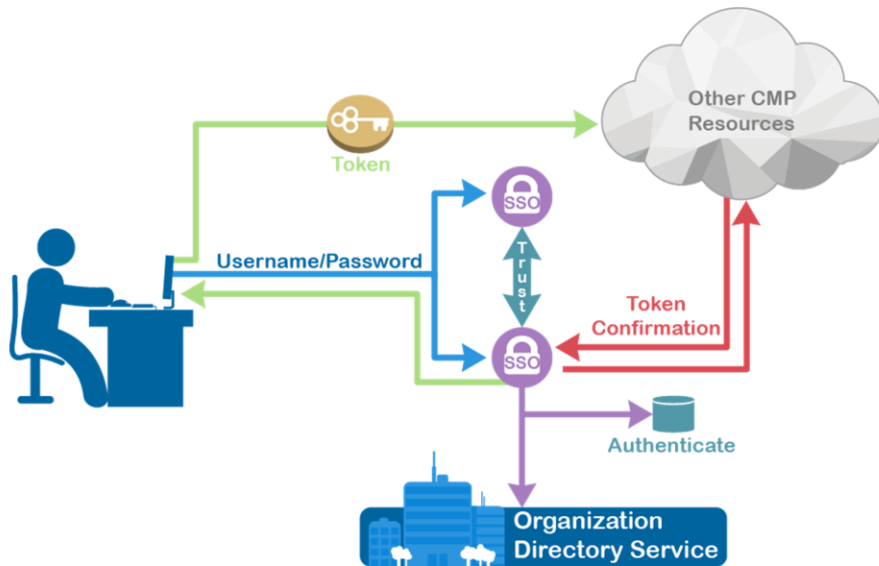
In a cloud infrastructure, resource usage must be monitored, controlled, and reported in order to provide transparency for both the provider and consumer. The cloud management platform includes a chargeback component which will be deployed to collect usage statistics and provide billing reports to consumers. The CMP also includes infrastructure monitoring capabilities which will alert the provider when a problem exists and collect performance metrics to ensure service levels are being met.

A cloud infrastructure design should also include a centralized logging capability. You should collect log information from all of the CMP components including account access, actions taken, and infrastructure errors. Also include an alerting capability so that cloud administrators know when there is an issue such as a component failure or security breach. Use an organization's business requirements, governance rules, and compliance policies as a guide for determining which information must be stored and for how long.

In a multi-tenant solution like a public cloud, it may not be feasible to grant the tenants access to the log information since the cloud provider must maintain tenant separation at all levels. In this case, a process must be established to produce relevant log information for a tenant when required such as in the case of a security breach. When possible, locate the logging server on an isolated network and minimize the number of accounts granted access.



# Authentication and Single Sign-on



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

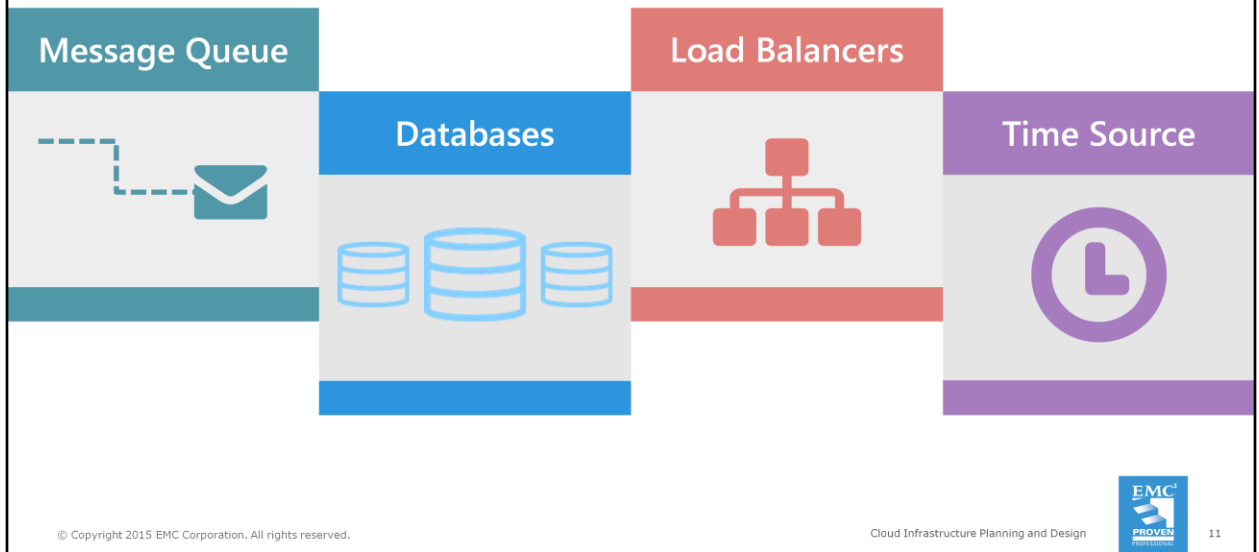


10

A cloud management platform requires an authentication mechanism that can be used to access the portal, control services, and support multi-tenancy. CMPs can support local user accounts, external directory services, and federated authentication models. Many cloud management platforms use a single sign-on mechanism to authenticate a user once and grant access to the various components.

Clouds should have the ability to integrate with an existing authentication service to facilitate a single sign-on capability. This improves customer satisfaction by removing the need for multiple sign-ons. It also simplifies user and group management for IT. For public cloud providers, the ability to integrate with multiple existing authentication services is critical in order to support multi-tenancy.

## Supporting Applications

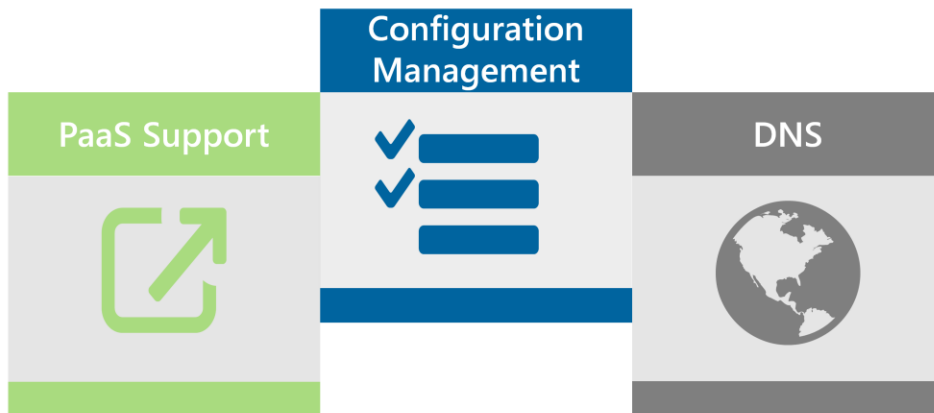


Cloud Management Platforms are made up of various components or applications which must be able to communicate with each other. In some CMPs, components communicate through an intermediary service such as a message queue. CMP components place requests or data in the queue for other components to eventually read and process. OpenStack, for example, uses RabbitMQ or Qpid which are Advanced Message Queuing Protocol (AMQP) frameworks. Queue implementations are typically deployed as a centralized or decentralized pool of queue servers.

Components of a CMP also need to store information such as state, configuration, permissions, roles, statistics, and so on. This is accomplished using database servers and like the message queues, can be deployed in a centralized or decentralized manner. The Federation Enterprise Hybrid Cloud solution is an example of a CMP that uses various database servers to store catalog information, access information, and log information and metrics.

Another example of a supporting application is a load balancer. Load balancers can be used to distribute network traffic across multiple instances of CMP components which improves performance. If the load balancer has the capability to detect a failed instance of a component, then by redirecting traffic to the remaining instances, the load balancer helps maintain cloud availability. Load balancers may be deployed as physical or virtual appliances within your CMP infrastructure.

Because of integration between components and single sign-on capabilities within the CMP, a single and reliable source for time should be configured within all components to ensure proper functionality.



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



12

Most Cloud Management Platforms have service catalog and orchestration components that support IaaS. Adding Platform as a Service (PaaS) capabilities will most likely require additional applications used to deploy development software and tools on top of the IaaS instances. This PaaS deployment suite of applications will most likely reside on the infrastructure that supports the CMP and will have its own integration, performance, and availability requirements. Cloud Foundry is an example of an open PaaS solution which can be run on multiple IaaS infrastructures.

In conjunction with the PaaS deployment tools or as a separate entity, your cloud design may need a configuration management tool that can be used to manage the configuration of the CMP servers and also the configuration of any services that are deployed in the cloud. Your design may include one configuration management environment for the entire cloud infrastructure, or the requirements may dictate that the tenants be responsible for their own instances. Examples of the configuration management tools are Puppet or Chef. The primary controllers for these environments will also reside on the CMP infrastructure in the form of one to many servers and will also have its own integration, performance and availability requirements.

DNS is required to support services and the CMP. To support elastic capabilities within the cloud, DNS services will need to be identified for tenants, and orchestration workflows and secured accounts will be needed for updates.

## Selecting a CMP



What requirements influence CMP selection?

© Copyright 2015 EMC Corporation. All rights reserved.

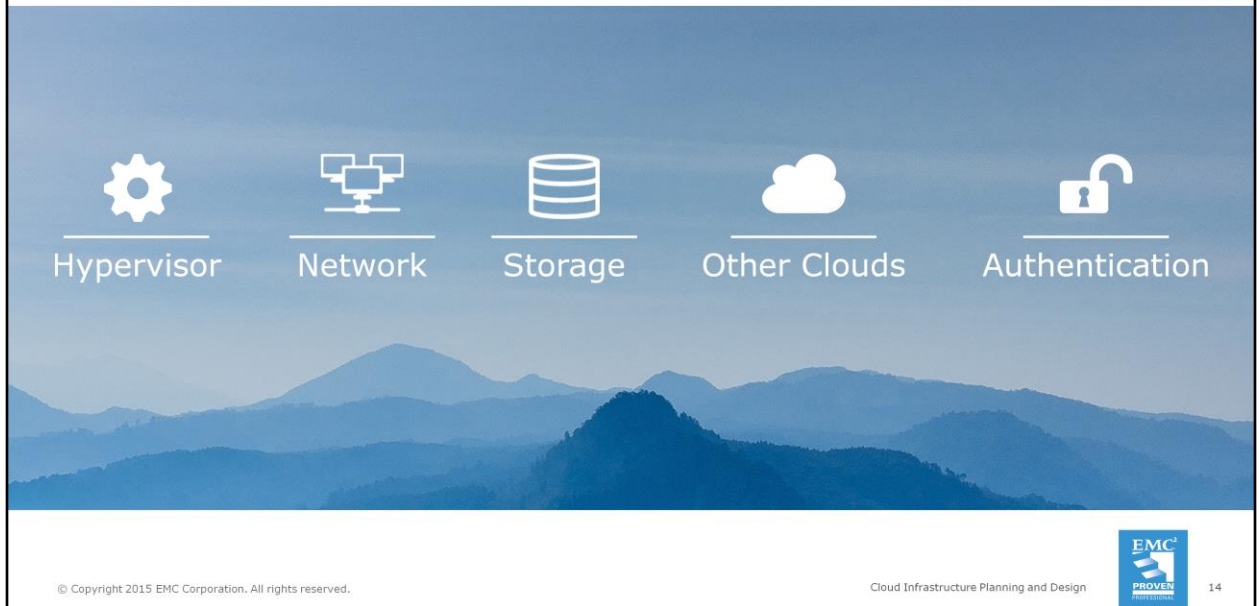
Cloud Infrastructure Planning and Design



13

When designing a cloud infrastructure, an architect must select a cloud management platform that aligns with the business needs. The requirements gathering phase helps you select the correct platform. Selecting the right cloud management platform is all about understanding the requirements and then finding the solution that meets the needs of the organization.

# Infrastructure Integration



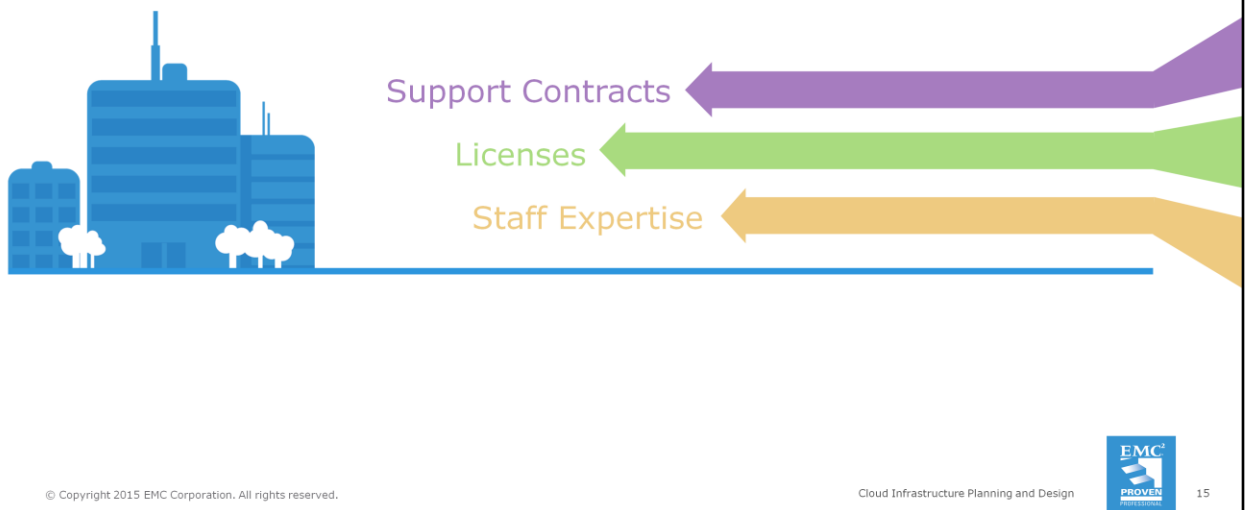
Many cloud management platforms integrate with a variety of infrastructure components. However, not all CMPs do integrate with everything and some support is limited in functionality. In some cases, a plug-in or adapter may be necessary for integration and it may be supported by a separate vendor or community. The cloud architect will need to help select infrastructure components for the cloud that best match the requirements of the organization. The architect will also need to select a CMP which not only meets organizations requirements for providing services, but also can be integrated with the selected cloud infrastructure.

An example that may influence choices for cloud management components is if an organization wishes to use a specific hypervisor to support IaaS instances. Although many cloud management platforms support many hypervisors, the support might be limited for the planned hypervisor.

Another example that will influence choices for cloud management platform components would be the selection of either network or storage infrastructure. In some cases, additional services or plug-ins may be required to use some of the advanced features of these infrastructure components.

A final example that will influence the CMP selection is the organization's desire to integrate a private cloud with a public cloud. The architect will need to understand the API capabilities, hypervisor support, connectivity requirements, and many other items before selecting CMP components that can support this integration.

## Organizational Readiness



An organization's readiness for supporting a cloud infrastructure will most likely impact the selection of a cloud management platform. For instance, if an organization already has support and licensing agreements in place that enable the deployment of a specific vendor's solution at low to no additional cost, then the choice may be limited to that vendor. Or if the organization has expertise in specific programming languages, this may influence some of the orchestration and automation capabilities that are added to the design. Finally, if a PaaS solution is being considered, then the selection of the PaaS platform may depend on this programming expertise.

## Proprietary vs. Open Source CMP

- Proprietary
  - VMware, Microsoft
- Open source
  - OpenStack, CloudStack
- Vendor-supported distribution of open source
  - Red Hat, Mirantis, Citrix, VMware, Canonical, SUSE

vmware®



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



16

One of the many design decisions that an architect makes is whether to use proprietary or open source components in the CMP. When it comes to open source, an additional decision must be made as to whether to go the pure open source route or to adopt a vendor-supported distribution of an open source solution. As with most decisions, the organization's requirements guide the architect when selecting the components for the cloud management platform.

Considerations for selecting a proprietary solution:

- Enhanced integration with existing vendor solution
- Existing support agreements or licensing agreements
- Advanced out-of-the-box capabilities or features
- Paid support options available
- Simpler integration with other cloud platforms that use the same vendor solution
- Higher software costs

Considerations for selecting a completely open source CMP solution:

- Lower software costs
- Simpler integration with other open cloud platforms
- Requires in-house development expertise or reliance on open community to add functionality
- Limited or crowd-sourced support options

Considerations for selecting a vendor supported distribution of open source CMP solution:

- Lower cost for software
- Some compatibility testing to ensure functionality
- Paid support options for certain versions
- Simpler integration with other open cloud platforms
- Supported release cycle may lag behind open source release cycle

## Services Supported by CMP



- Solution support for single service only
- Infrastructure as a Service
- Platform as a Service
- Software as a Service
- Multiple solutions required
- Build your own

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

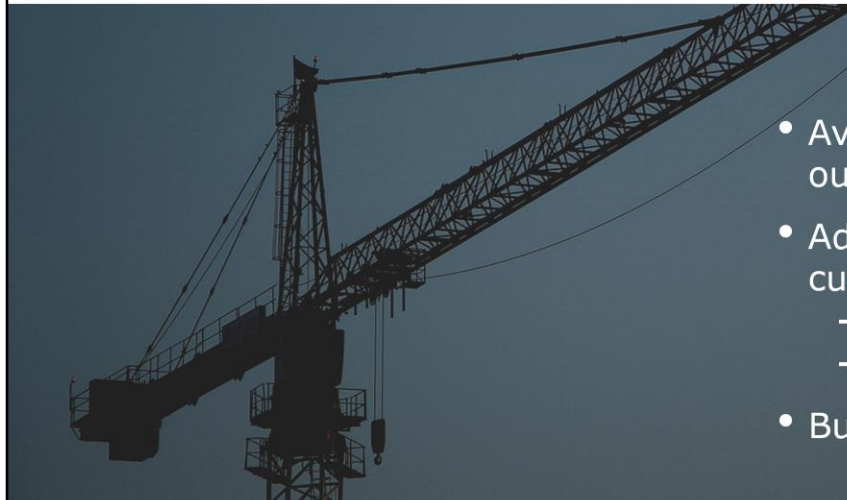


17

Whether the organization plans to purchase or implement an open solution, the architect must realize that not all cloud management solutions will provide all types of services. Many solutions only provide infrastructure as a service. Others may provide platform as a service. To meet a full set of ITaaS requirements, an organization may deploy multiple service catalogs and integrate these into a centralized portal. It is also possible that an organization will develop and build its own service catalog.



## Customization



- Available options on out-of-the-box solution
- Additional customizations
  - Integration
  - Full service catalog
- Build your own

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



18

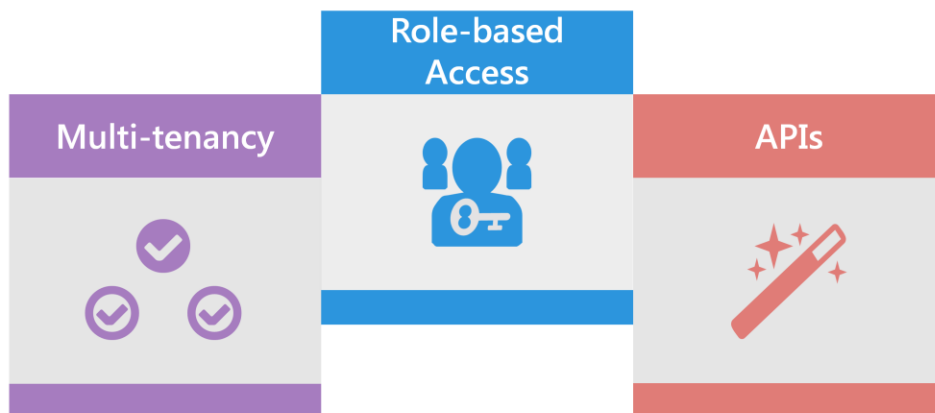
Most cloud management solutions offer basic IaaS functionality with some ability to customize the user interface and orchestration capabilities. Other solutions provide deeper customization capabilities as well as additional integrated components that provide functionality beyond IaaS. Many organizations will implement these solutions and use the out-of-the-box functionality.

However, there are more alternatives for customizing cloud capabilities. Third party products are available that layer above these other solutions and provide enhanced features, customizations, and integration capabilities. They provide a full IT as a Service capability for the organization.

Another solution is for the organization to build its own portal and catalog capability that either integrates with the lower layer functions or build the complete solution encompassing the full functionality of a CMP. These options require a strong development team to create and maintain the solution.

The cloud architect needs to understand the requirements, research the available options, and recommend the best option for the organization.

## Features



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



19

Some of the other features that will influence the selection process are:

**Multi-tenancy** – Support for multi-tenancy is expected in a cloud but understanding how multi-tenancy is supported is important. Some clouds require that each tenant has its own authentication source, so the selected solution must support this. Other requirements may be that some tenants' data must be physically isolated from others. Having a solution that allows for this capability is also important.

**Role-Based Access** – Most solutions provide this capability, but each solution may implement this in a different way. The mechanisms supporting role-based access should align with requirements.

**APIs** – The cloud management platform is made up of many different components. As a cloud architect you must understand how the APIs are implemented and secured so that you can create a proper design. If a component does not have a properly secure API, then either the design will need modification or another solution may need to be investigated.

# Lesson 1

### Summary

Key points covered in this lesson:

- The cloud management platform includes the components that are necessary to support self-service, elasticity, metering, and other cloud characteristics.
- Selecting a CMP requires an understanding of requirements, as well as the integration capabilities with the underlying infrastructure.



This lesson covered cloud components and the requirements used in selecting them.

# Lesson 2

## Cloud Management Infrastructure

This lesson covers the following topics:

- Requirements for implementing a cloud management infrastructure
- Design considerations for implementing a cloud management infrastructure



This lesson covers the requirements and design consideration that contribute to the design of a cloud management infrastructure.

# Cloud Management Infrastructure

Click within the play area to launch the video

© Copyright 2015 EMC Corporation. All rights reserved.

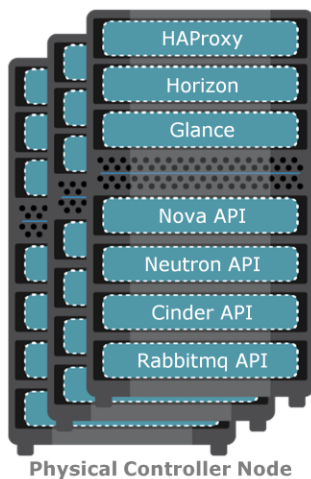
Cloud Infrastructure Planning and Design



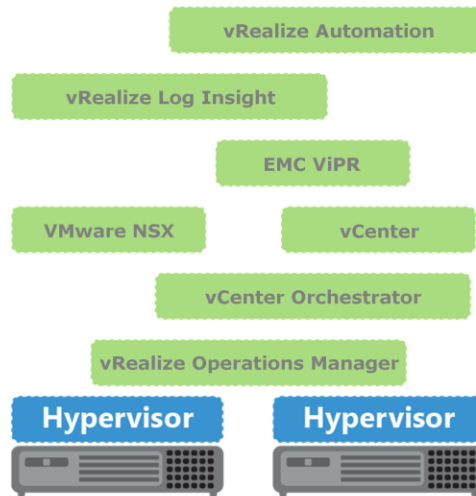
22

## Physical or Virtual

### OpenStack Example



### Federation Enterprise Hybrid Cloud Example



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



23

A cloud management platform is made up of multiple applications. These applications will all run on a server with an operating system. An architect must decide whether the server should be a physical box or a virtual machine running on a hypervisor. As with everything else, the requirements are a guide in deciding which to choose but there are some general factors to consider.

Whether the application is on a physical or virtual server, you may have performance concerns. However, virtual servers can be scaled up more quickly and may also be moved around within the CMP infrastructure to balance load. Another point to consider is availability. A hypervisor like VMWare's ESXi has the ability to restart an instance if a server crashes. Although there is an outage of the instance, the recovery time can be minimal. Because of expertise or cost, an organization may desire this availability option over configuring an application availability option such as database-clustering. Another consideration is recovery options. Entire virtual machines can be backed up more easily or can even have snapshots taken periodically. While you may not consider this to be too important with servers that only have running applications with minimal data, it makes recovery easier after a security breach or datacenter disaster.

# Cloud Management Infrastructure Sizing

| Example Requirement                   | Possible Sizing Impact |                                                     |      |                                                                                                  |
|---------------------------------------|------------------------|-----------------------------------------------------|------|--------------------------------------------------------------------------------------------------|
| Expected Number of Users              | 500                    | 2 Portal Servers @ (2 vCPU, 4 GB RAM, 50 GB Disk)   | 1000 | 3 Portal Servers @ (2 vCPU, 4 GB RAM, 50 GB Disk)                                                |
| Expected Number of Services           | 1000                   | 1 Monitoring Server @ (4 vCPU, 8 GB RAM, 1 TB Disk) | 3000 | 1 Monitoring Server @ (8 vCPU, 16 GB RAM, 3 TB Disk)                                             |
| IaaS with SDN                         | No                     | 3 hosts                                             | Yes  | 6 hosts                                                                                          |
| Infrastructure Host High Availability | No                     | 3 hosts                                             | Yes  | 4 hosts (cluster)                                                                                |
| Backend Database Redundancy           | No                     | 1 DB Server @ (2 vCPU, 8 GB RAM, 200 GB Disk)       | Yes  | 1 DB Server @ (2 vCPU, 8 GB RAM, 200 GB Disk) Additional CPU for clustering and synchronization. |

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



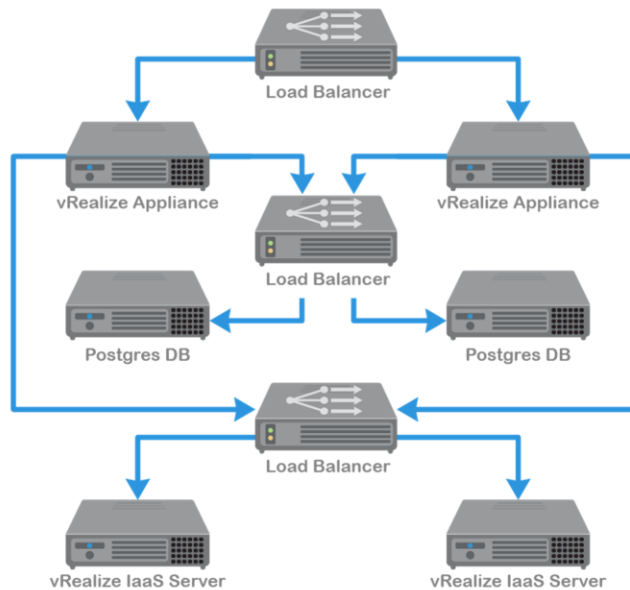
24

Sizing the cloud management infrastructure is somewhat easier than sizing the infrastructure to support consumer services. This is because once the cloud management components are selected, documentation will be available explaining the sizing requirements. Some of the requirements that will influence the sizing of the cloud management infrastructure are:

- Expected number of users
- Expected number of services
- Types of services
- Component redundancy
- High availability

Displayed here are examples of how some requirements will impact the sizing of the cloud management infrastructure.

## Availability



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



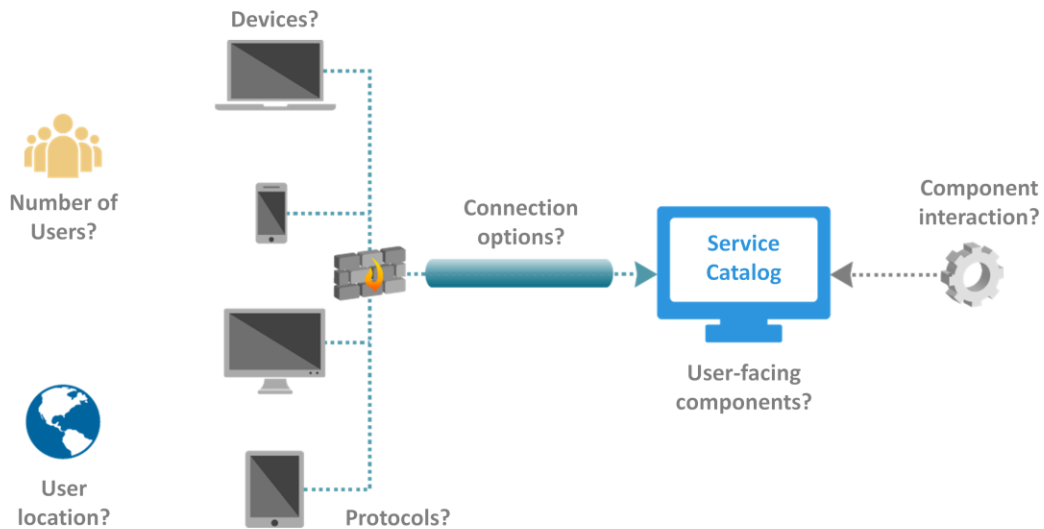
25

The idea of an on-demand, self-service model is to provide consumers with necessary IT resources when they need them. Because an organization will come to rely on a cloud as a critical part of the business, the cloud should be in an “always on” state. Although it is very likely that the responsibility for implementing high availability for consumer services will be tasked to the developers, it is the architect’s responsibility to design a highly available CMP to support those services. This may include designing a CMP infrastructure that has more redundancy than the consumer resource pools. It will also include ensuring that enough underlying resources are available to support the needs of the various components. It will most likely include the deployment of redundant management components with load balancers to distribute traffic across components and failover traffic if components go offline. Designing backend database redundancy using something like replication is also a good practice for improved availability. The design may include Quality of Service (QoS) policies for LANs and SANs that will give priority to traffic that flows between CMP components or to the underlying infrastructure.

Shown here is an example design for a highly available implementation of a VMware cloud platform using vRealize Automation (From: VMware vRealize Automation Reference Architecture)



## Connectivity Considerations



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



26

When selecting a cloud management platform, it is important to understand the connectivity requirements of the organization as well as understand the connectivity requirements for the components within the platform. Examples of questions that should be asked are:

- What devices and protocols does the organization expect to use to access the service catalog?
- How many consumers are expected to use the cloud at any given time?
- Where are the consumers located?
- What components of the CMP need to be accessed by the consumers?
- How do the CMP components talk to each other?
- Are the CMP components running on virtual machines or physical?
- What connection options are available from the Internet into the datacenter where the CMP is located?

Let's examine the impact on the design from a network connectivity point of view.

## Users, Locations, Devices, and Protocols

- Bandwidth requirements
- Number and sizing of load balancers
- Firewall configuration and sizing
- Static content served via CDN

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



27

The number of users, protocols, and device types influences design decisions in multiple areas. For example, knowing this information helps determine the bandwidth requirements for accessing the CMP. Bandwidth calculations must be applied across all of the physical components such as switches, routers, external network connections, and host networking. If the cloud will be accessible from the Internet, then the connection to the Internet will need to be examined to ensure that it supports the bandwidth, availability, and redundancy requirements of the organization. If load balancers will be deployed, then this information is useful for configuring the load balancer properly as well as determining the proper number and sizing. Finally, the protocol information is necessary for properly configuring and sizing any firewalls that may exist between the consumers and the CMP components.

Knowing the consumer location may be useful information in a solution in which the catalog or desired services involves static content such as images. The use of a content delivery network (CDN) will improve response time at the consumer's device when downloading static content. A CDN is a system of distributed servers that delivers content to consumers based on the geographic location.

## User-Facing Components

- Multiple points of access
  - UI, API, multiples of each
  - Integrated components may just look integrated
- Important for firewall and load balancer configuration

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



28

The choice of a CMP also influences the connection requirements for consumers. Each CMP solution has different access points, and if users need to access various UI's or API's then each one will require configuration. Some integrated components appear to be served from the same place but may in fact be multiple services running on different servers. For example, if you choose a vRealize Automation solution like the one used in the Federated Enterprise Hybrid Cloud, then consumers need access to vRealize Automation portal server in order to deploy an IaaS instance. However, in an OpenStack solution, the consumer not only requires access to the portal server (Horizon) but may require access to any server running the Compute (Nova) API. Knowing the consumer touchpoints is critical for firewall and load balancer configurations.

## Component Interaction



- Isolated networks vs. publicly accessible
- Virtual machine connectivity

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



29

It is important not only to know how or where the consumers interact with components, but also to understand how the components interact with each other. You can secure component interaction by using an isolated network, but in some cases, components may require a connection to publicly-available interfaces. For instance, in an OpenStack environment, the Nova Compute controller API must be available to consumers and backend processes alike, so the cloud design needs to enable these connections. However, the message queue for Nova does not need to be accessible by the consumers and it can be located in an isolated network.

If CMP resources are running on virtual machines, then the cloud design needs to include network bandwidth on physical hosts that not only support multiple components, but also the potential movement of components.

## Security Zones

- Also known as trust zones
- Groups of users, servers, applications, or objects with similar security requirements
- Supported by:
  - Authentication mechanism
  - Physical or virtual border controls
- Some services may span zones



© Copyright 2015 EMC Corporation. All rights reserved.

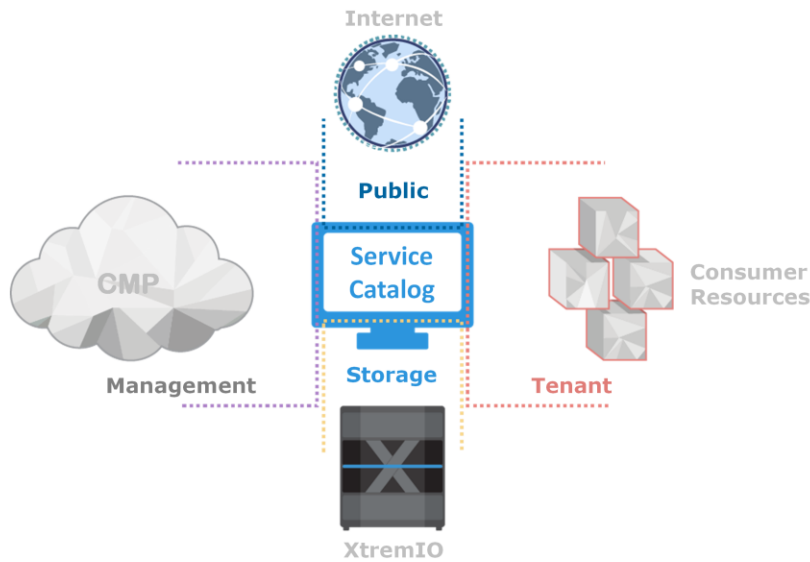
Cloud Infrastructure Planning and Design



30

It may be useful to define security zones within your cloud design. Security zones also known as trust zones are groupings of users, servers, applications, or other objects that have similar security and trust requirements. The objects within a security zone may share the same authentication and authorization mechanisms which could differ from other security zones. Security zones may also have their own physical or logical controls placed at their borders to restrict unauthorized access to internal objects. Some CMP components may require access to multiple zones. The cloud architect must identify these exceptions and create a plan for accommodating and securing these components.

## Security Zones Example



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



31

Here we see an example of security zones that can be defined for a cloud environment. This list is meant to be an example and a more accurate list of security zones will be defined by the requirements of the organization.

**Public:** This zone contains public facing objects such as the portal servers, SSOs, load balancers, and networks (including the Internet). It is considered to be the least trusted zone and because of this, it requires a solid security design and should contain the least number of CMP components as possible. Examples of security mechanisms that should be considered for this zone would be secure authentication, network traffic encryption, and publicly trusted Public Key Infrastructure (PKI).

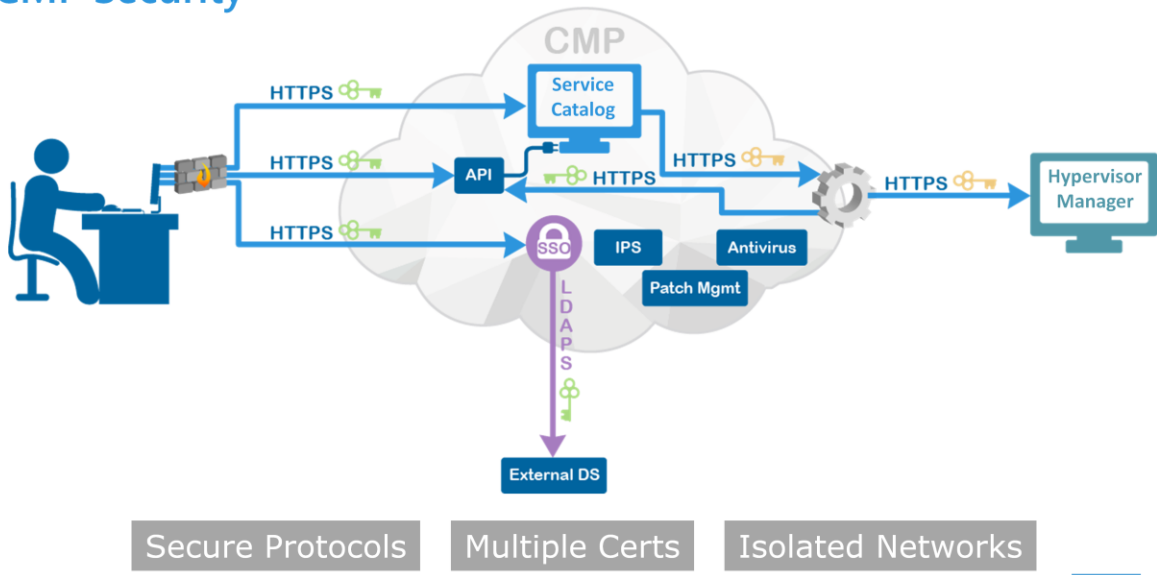
**Tenant:** This zone contains the services that are provisioned by tenants. If multiple tenants exist, then each will have its own zone with its own network segments, firewalls, services, and possibly authentication mechanism. Examples of security mechanisms that should be considered for this zone would be tenant-managed authentication system, firewalls, VPNs, network segmentation, and portal/service catalog multi-tenancy support. Since the portal enforces multi-tenancy, authenticates tenants, and provides the API for public and other CMP components, it is an example of one CMP component that “bridges” multiple zones.

**Management:** This zone contains the majority of CMP resources and other applications used to manage the cloud infrastructure. Network access to this zone should be extremely limited to authorized cloud provider employees. This zone would be considered one of the most trusted zones but still should have all of the proper security processes in place. Examples of security mechanisms that should be considered for this zone are isolated networks, separate authentication system from public zones, separate PKI from public zone, and network traffic encryption.

**Storage:** This zone is similar to the management zone in its trust level. It contains the storage traffic (data) that is transmitted across the network as well as the data repositories. Examples of security mechanisms that should be considered for this zone would be isolated networks (LAN or SAN), network traffic encryption, and data-at-rest encryption.



## CMP Security



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



32

The Cloud Management Platform contains the primary entry point into your cloud infrastructure. CMP components are also granted the authority to trigger the creation and deletion of services as well as to control existing services. Allowing unauthorized access to your CMP components can cripple an entire infrastructure.

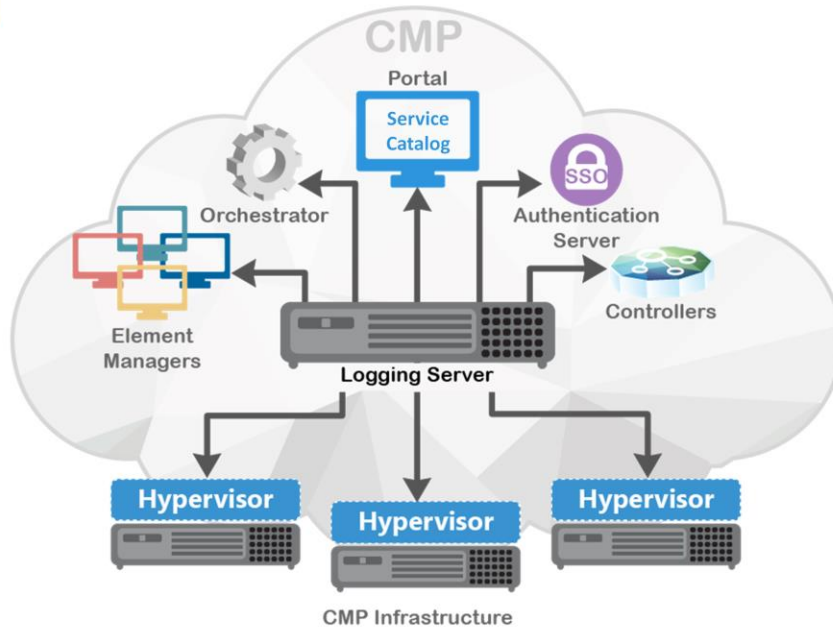
In protecting the cloud, the first consideration is to implement a strong authentication system using standard secure protocols, such as HTTPS and LDAPS. Additionally, consider implementing a PKI solution with known trusted certificates that are used during the traffic encryption process. This includes using these methods to protect the cloud portal (human interface) and the application programming interfaces (APIs) as well.

If some CMP resources require access from an external network such as the Internet, place a firewall between the CMP and consumer devices to ensure only the appropriate traffic is allowed. The cloud design should also include any existing or new standards or procedures that are required to protect the underlying servers such as intrusion prevention services, antivirus, and OS patching.

If implementing an open or proprietary solution, ensure that you include or reference vendor or community guidelines for securing the product and applying regular security updates and patches to the CMP components.

Element managers should be accessed by a limited number of entities such as Cloud Administrators, the service catalog, metering applications, or orchestration engine. They should not be accessed directly by the users and the interfaces should be well protected. As part of the cloud design, consider placing the element manager interfaces on separate, isolated networks. If the public-facing interfaces are necessary, make sure that routing and firewall rules are in place to block inappropriate access. The element managers should also have a limited number of user accounts set up that can be used by other CMP components. These accounts should be configured with the least amount of privileges needed to accomplish a task. The account information should be transmitted using trusted encryption methods.

## Logging



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



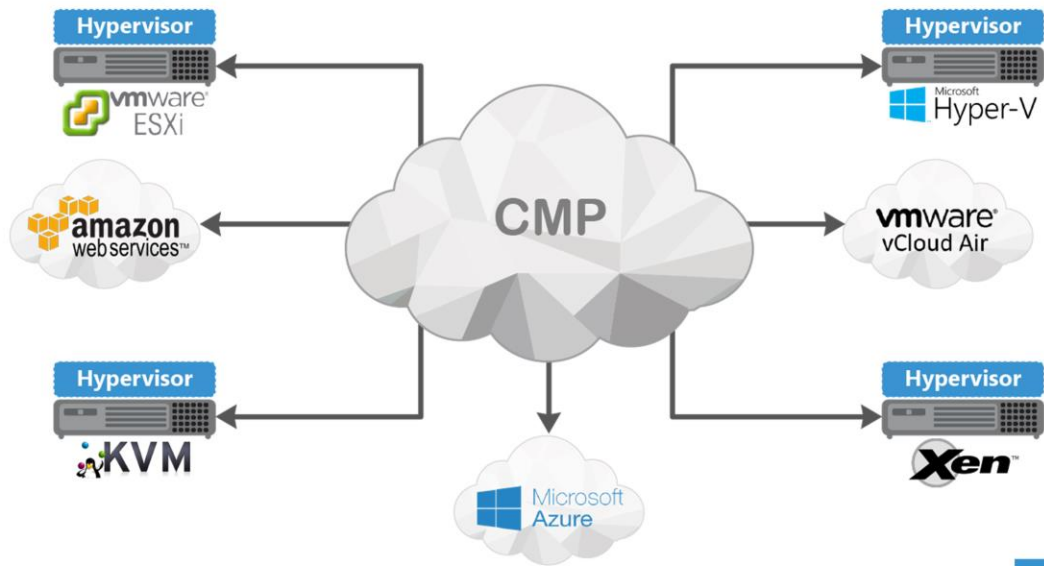
33

A cloud infrastructure design should also include a centralized logging capability. Log information should be collected from all of the CMP components and include account access, actions taken, and infrastructure errors. Also, include an alerting capability so that cloud administrators know when there is an issue such as a component failure or security breach. Use an organization's business requirements, governance rules, and compliance policies as a guide for determining which information must be stored and for how long.

In a multi-tenant solution like a public cloud, it may not be feasible to grant the tenants access to the log information since the cloud provider must maintain tenant separation at all levels. In this case, a process must be established to produce relevant log information for a tenant when required, such as in the case of a security breach. When possible, locate the logging server on an isolated network and minimize the number of accounts granted access.



## Provisioning Capabilities



© Copyright 2015 EMC Corporation. All rights reserved.

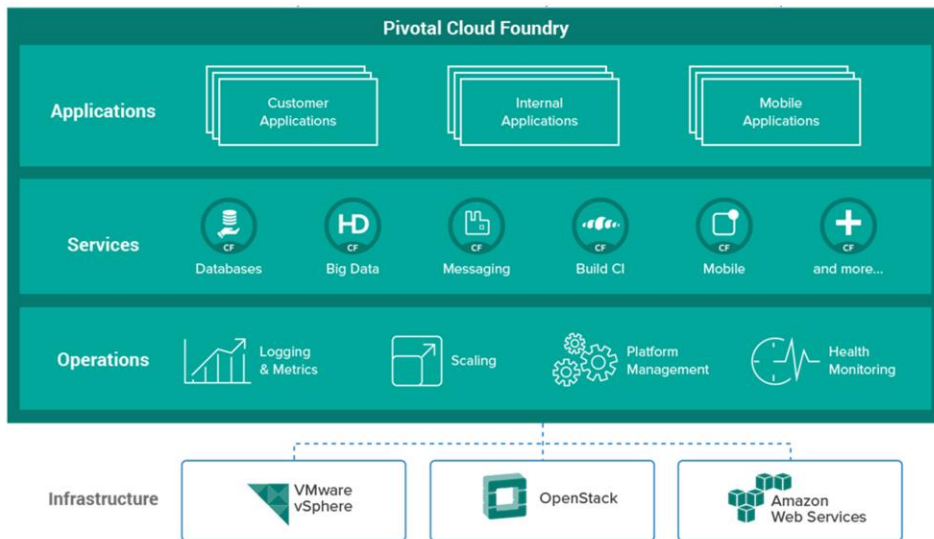
Cloud Infrastructure Planning and Design



34

The CMP platform will have capabilities to deploy services in a cloud, but what infrastructure is supported for provisioning? Using IaaS as an example, an architect needs to understand which hypervisors and public cloud platforms are supported for provisioning and how deployment is accomplished. OpenStack, for instance, requires that the Nova-Compute service is run on the nodes where KVM, Xen, and Hyper-V hypervisors are running, but the VMware integration consists of a driver used by Nova-Compute running outside the VMware environment that talks to the vCenter API. In another case, vRealize Automation can be used to deploy IaaS instances using endpoints. The VMware endpoint connects to a vCenter server, Hyper-V endpoint connects to a System Center Virtual Machine Manager server, AWS endpoint connects to the Amazon APIs, and the Cisco UCS endpoint connects to UCS manager to provision physical servers. In all cases, physical connections should exist between the CMP and the provisioning sources as well as the proper credentials for creating services.

# PaaS



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



35

Platform as a Service is quickly becoming a more integral part of cloud infrastructures, but deploying a PaaS solution may not be that simple. There are many technologies and methodologies for deploying PaaS solutions and many of these are still maturing.

Many organizations have deployed PaaS solutions by first deploying an IaaS instance and then installing and maintaining the development applications using orchestration and configuration management tools. In this scenario, the developer is granted access to the development tools but may also have access to the operating system environment in the IaaS instance. This may include deploying an IaaS instance from a template that includes a configuration tool like Puppet and a continuous integration tool such as Jenkins.

An alternative solution is to deploy a full PaaS suite such as Cloud Foundry that is essentially a self-contained environment. This solution contains a control plane, services that control and manage the development environment, and a data plane, another set of services that executes the applications that are being developed. In this solution, development services are pre-deployed in the infrastructure, and the developer interfaces with a single front-end for code check-in, development, test, and production. The developer has no involvement with any infrastructure. The PaaS provider may need to monitor the environment to ensure adequate resources are available. Rather than the organization maintaining and scaling a single PaaS environment, it may wish to add the PaaS suite solution to the catalog. This would enable individual tenants or projects to deploy instances of the entire suite so that each tenant could have their own environment.

Selecting and designing a PaaS solution is not part of the scope of this course. However, the cloud architect needs to understand the requirements for this solution so that adequate resources are available. For instance, the CMP infrastructure may need to support a central configuration management server like Puppet to deploy and control development environments. If a single PaaS environment will be deployed, some services may reside on the CMP infrastructure and other services within the consumer resource pools. If the PaaS environment is offered as a service via a catalog, then the entire solution may be deployed in the consumer resource pools.

# Lesson 2

## Summary

Key points covered in this lesson:

- Cloud management components should reside on separate infrastructure from the consumer resources.
- Securing cloud management interfaces is critical for maintaining multi-tenancy, data security, privacy, and availability in the cloud.



This lesson covered the requirements and design consideration that contribute to the design of a cloud management infrastructure.

## Lab 2: Cloud Management Design

### Purpose:

Develop a design and plan for deploying components that support cloud management and operations.

### Tasks:

- Identify cloud management components that are relevant to a NanCo cloud design.
- Provide a logical diagram that describes the cloud management infrastructure and is based on the NanCo requirements.

## Lab 2: Debrief

### Cloud Management Design

**Purpose:**

Develop a design and plan for deploying components that support cloud management and operations.

# Module 2

## Cloud Management



Key points covered in this module:

- The CMP is composed of many different components that require their own infrastructure.
- Requirements gathering is key to selecting CMP components and designing the underlying infrastructure.
- CMP infrastructure design deliverables include considerations for sizing, availability, and security.



This module covered the design decisions and considerations for building a cloud management platform.



# Module 3

## Compute Resources



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges, and considerations of implementing various compute technologies for consumers in a cloud infrastructure.
- Align business requirements when selecting technologies that support compute resources in a cloud infrastructure.
- Develop a design and plan for deploying consumer compute resources in a cloud infrastructure.



Cloud Infrastructure Planning and Design

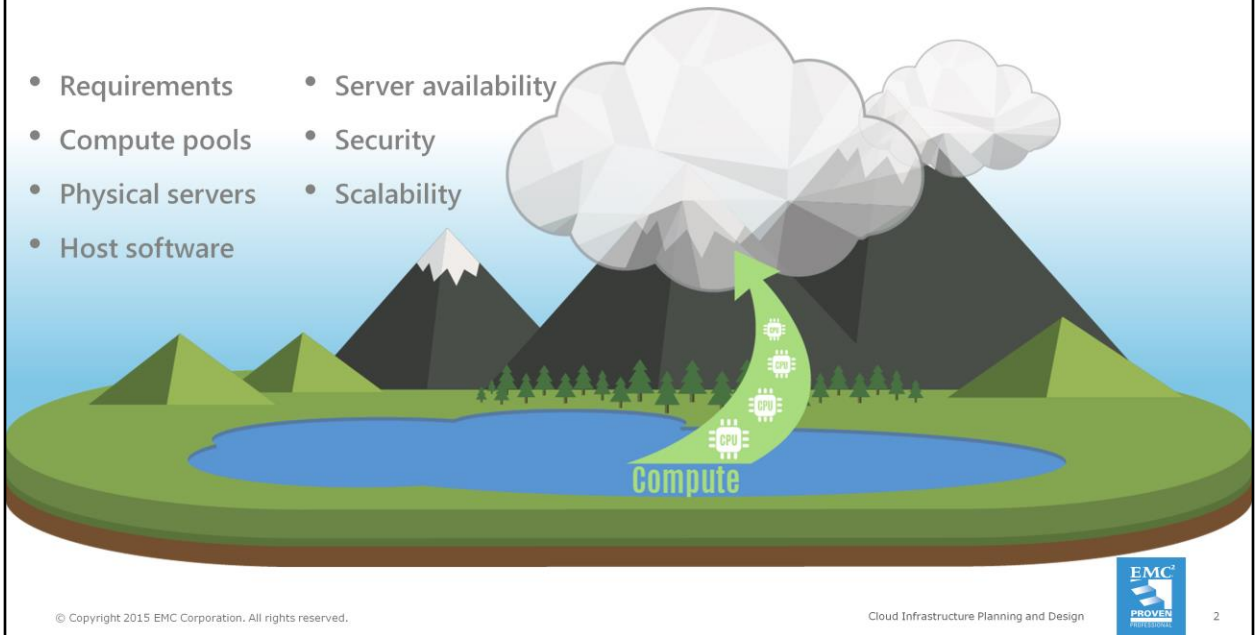
1

This module focuses on technologies and considerations used in designing compute resources for cloud consumers.



## What's This About?

- Requirements
- Compute pools
- Physical servers
- Host software
- Server availability
- Security
- Scalability



A cloud must have pools of resources allocated in order to support the planned services. This module highlights the technologies, options, and choices as they relate to compute infrastructure that may be included in a cloud design.

Listed here are the high level topics covered in this module.

# Lesson 1

## Pools and Physical Servers

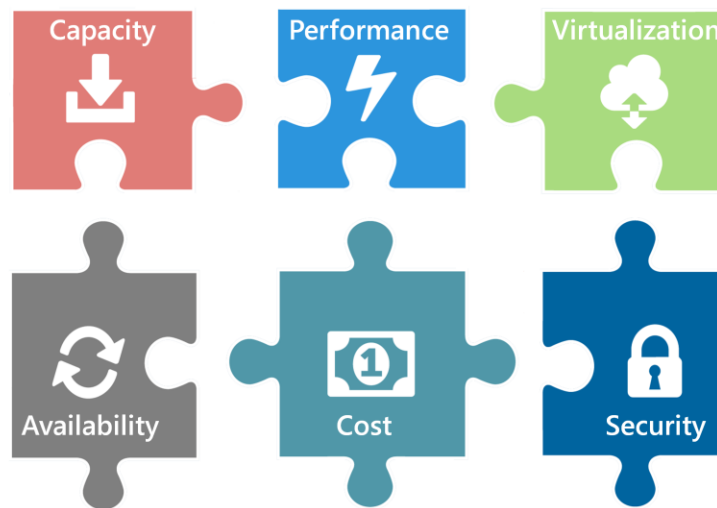
This lesson covers the following topics:

- Requirements for implementing compute resources in a cloud
- Design considerations for segregating and sizing servers in a cloud environment



This lesson covers the requirements and considerations for sizing servers and designing compute pools.

## Understand Compute Requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



4

The compute component of a cloud infrastructure design is driven by requirements. Compute requirements fall into categories such as those listed here.

**Capacity** – The compute resources requires enough CPU and memory to run the planned services as well as overhead associated to an underlying OS or hypervisor, management processes, advanced features, and reserved capacity.

**Virtualization** – Cloud infrastructure works best using a virtual infrastructure. Many decisions exist around the type of virtualization to use and the appropriate vendor or open source solution.

**Performance** – Application performance depends on the proper selection of CPU and RAM speed and architecture, number of servers, number of I/O slots, choice of I/O cards, and hypervisor/OS capabilities.

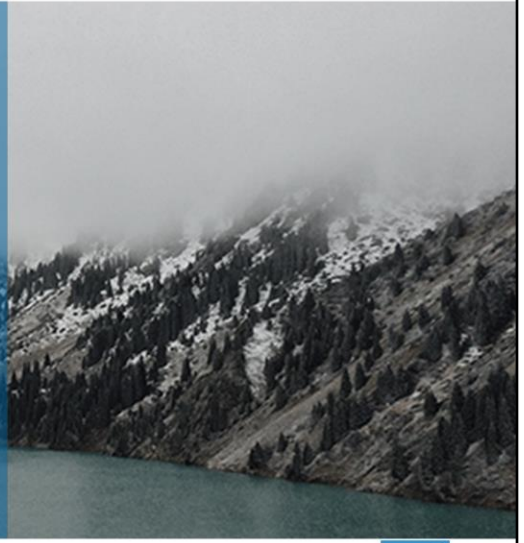
**Cost** – The cost of the compute infrastructure is influenced by the number and types of servers, software licenses associated with hypervisors/operating systems, support contracts, and datacenter elements such as power and cooling.

**Security** - Since most designs include shared compute infrastructure, hypervisor or virtual container security is critical to maintaining tenant and application separation.

As we discuss different technologies and design considerations throughout this lesson, we will see how these technologies help to address the different requirements.

## Compute Pools

- Logical grouping of compute resources
- Cloud design may contain multiple compute pools
- Organized by characteristics
  - Performance
  - Security
  - Isolation
  - Location
- Defined by the requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



5

In cloud computing, server infrastructure is usually presented as pools, which are logical groups of compute resources that are organized by their characteristics. These characteristics may include performance, security options, availability, location, configuration or even tenant ownership. Pools promote sharing of resources and hide unnecessary details of physical infrastructure from the cloud consumer.

Physical servers may be grouped together in one pool of resources or into many different pools. If the cloud design is intended to support a generic workload with a “best effort” policy for supporting performance and availability, then a single pool may be the option. It is also possible that each tenant is provided with its own pool of resources, which may or may not be delineated by physical servers. If the design needs to support varying performance levels, availability options, or service levels, then multiple pools may be necessary. Another example of a cloud requiring multiple pools would be one designed to support IaaS using virtual machines and PaaS using containers. The underlying physical servers in this scenario may require two different base operating system/hypervisor requirements and therefore require two compute pools. Finally, it is possible that the cloud management platform controls compute infrastructure in multiple sites and each site could represent a pool.

If multiple pools are required, then the cloud management platform must support this capability. This may include different element managers for different hypervisors or operating systems which would require that more resources be allocated in the cloud management infrastructure.

A consistent hardware design within a pool eases virtual machine movement, support, and lifecycle management.

## Physical Separation

- Support compliance requirements
- Implement a DMZ
- Support for different application types
  - Traditional
  - Cloud-native
- Tenant separation requirements

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



6

Although the goal of cloud computing is to pool resources and share them among the tenants, it is sometimes necessary to segregate hosts in a cloud environment. In a private cloud for instance, an organization may have a compliance requirement to keep certain data and services separate from the other infrastructure. An organization may also decide to implement a design that includes a DMZ for hosting public facing applications. Although this can be accomplished using software and virtualization, the requirement may still exist for physical isolation.

It is very possible that an organization that is implementing a cloud design is doing so for the very first time and plans to transition to cloud-native applications. These organizations may still wish to provision more traditional services while redesigning the these to become cloud-native. Alternatively, an organization may have a requirements to deploy more traditional services that support a PaaS or SaaS function. In both cases, the design may require that two different types of compute resources be implemented: One using enterprise class hypervisors enabling redundancy and hypervisor HA capabilities, and the other using a more commodity or open infrastructure supporting cloud-native applications.

Finally, host pool separation may occur if a particular tenant wishes to have a dedicated set of infrastructure. If this requirement is to be supported, then the design will need to be adapted to support this hosting type model.

## Physical Server Type

- Reduced footprint
- Easy installation
- Reduced cabling



- More I/O ports
- More local disk

© Copyright 2015 EMC Corporation. All rights reserved.

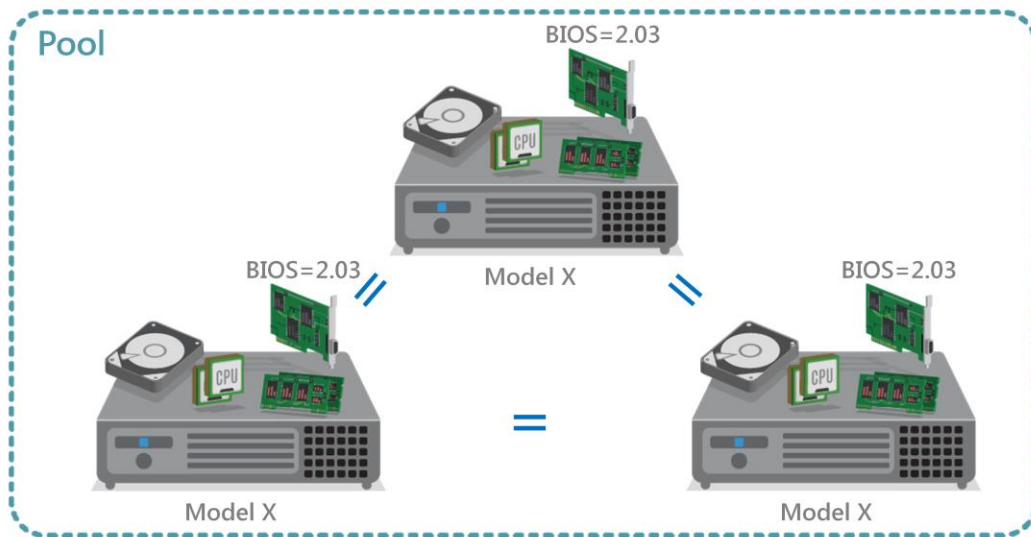
Cloud Infrastructure Planning and Design



7

Many factors can influence the decision to include blade or rack servers in your design including organization preference, availability of I/O ports, availability of PCI slots, and datacenter footprint. Blade servers allow for a reduced footprint but may limit you in terms of available PCI slots. Rack servers take up more space but may provide more options in terms of expandability and port options. The desired type of backend storage also plays a role in the decision process, since a popular option is to implement software-defined storage solutions in which the local disks that reside on individual hosts are pooled. In this instance, rack servers have the advantage since they have more space for local disk.

## Physical Server Consistency



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

Once you have decided between rack and blade servers, it is a good practice to implement a consistent server configuration and vendor throughout the environment. Consider obtaining servers with the same CPU type, memory configuration, firmware versions, and adapters. This minimizes compatibility issues if instances need to be migrated between hosts, reduces the need to manage multiple versions of drivers and patches, makes troubleshooting easier, and provides a known performance profile as new hosts are added to the environment.

# Physical Servers Compatibility

QME8362 10GigE Network Adapter by QLogic Corporation

**UCS B230 M2**

**Certifications**

| Product                                              |
|------------------------------------------------------|
| Red Hat Enterprise Linux (x86_64)                    |
| Red Hat Enterprise Linux (x86_64)                    |
| Red Hat Enterprise Linux OpenStack Platform (x86_64) |
| Red Hat Enterprise Linux OpenStack Platform (x86_64) |
| Red Hat Storage Server (x86_64)                      |
| Red Hat Enterprise Linux OpenStack Platform (x86_64) |
| Red Hat Enterprise Linux (x86_64)                    |
| Red Hat Enterprise Linux OpenStack Platform (x86_64) |
| Red Hat Enterprise Linux OpenStack Platform (x86_64) |
| Red Hat Enterprise Linux OpenStack Platform (x86_64) |

Compatible with the following versions:

- Windows Server 2012 R2 x64
- Windows Server 2012 x64
- Windows Server 2008 R2 x64

## VMware Compatibility Guide

[Back to Search Results](#)

**Model Details**

Model: Phoenix  
Partner: EMC  
CPU Series: Intel Xeon E5-2600-v2 Series  
System Type: Mother Board  
Number of Sockets: 2  
Max Cores Per Socket: 12

**Model Release Details**

VMware Product Name: ESXi 6.0

BIOS: Intel Corp. SE5C600.86B.02.01.0002.082220131453

| Feature Category | Feature   |
|------------------|-----------|
| 7.0 - 7.x        | Certified |
| 5.0              | Certified |
| 6.0              | Certified |
| 7.0              | Certified |

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



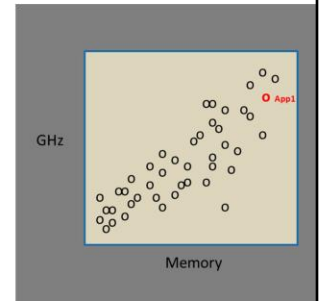
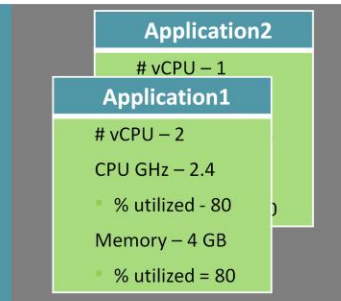
9

No matter which physical platform is selected, as an architect, you will ensure that it is compatible with other hardware and software used throughout the infrastructure. Most hardware and software vendors supply compatibility guides for this purpose.



## CPU and Memory Requirements

- Planned services
  - Assessment
    - Current environment
    - Stakeholders and subject matter experts
    - Pilot
  - Vendor/community documentation
- Reserved capacity
  - Usage spike
  - Overhead
- Future consumption



### Planned New Service Sizes

- Small = 0-3 GHz, 0-1 GB
  - 1vCPU, 1GB
- Medium = 3-6 GHz, 2-8 GB
  - 2vCPU, 4GB
- Large = 6-12 GHz, 8-16 GB
  - 4vCPU, 8 GB

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



10

Multiple sources are available to help determine CPU and memory sizing for the consumer resources. As part of the assessment, you gathered information about the services that were planned for your cloud environment. This information should include the number of services expected, the planned architecture of each service, the operating systems and applications that will be deployed with each service, the expected usage pattern for the services, and so on.

Once you have an idea of the planned service consumption, you then need to equate this to the expected CPU and memory consumption. For this, you can use other information from the assessment. For instance, if the organization already has a version of this application running in their current datacenter, you can use monitoring tools to measure CPU and memory usage and use that information to estimate planned service consumption. If the organization is running the service in a pilot version of a cloud, then this could provide more accurate consumption information. Asking members of the development staff will also provide information about expected consumption. Another source for information, although less accurate, is the documentation that is provided by the operating system or application vendor or community.

It is good practice to reserve CPU and memory capacity on a per-host basis for short-term usage spikes, infrastructure overhead, and similar issues. Infrastructure overhead can include the hypervisor or operating system, deployment agents, monitoring agents, device drivers, and advanced capability applications. For example, if you planned to deploy services with OpenStack on KVM hypervisor, then each host would need additional CPU and memory capacity to run the hypervisor plus the Nova compute and Neutron networking agents. Vendor or community documentation should provide CPU and memory resource consumption requirements for overhead components.

Finally, if the design does not initially include enough resources for the expected growth, then it must include a plan for adding compute capacity and in a non-disruptive manner.

## Compute Capacity Calculations

- Overcommitment
  - vCPU to CPU ratio
  - Memory sharing
  - Seek vendor guidance for various workloads
- Planned services
  - Concurrent number
  - Size

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



11

Once you understand the requirements for CPU and memory, you then need to determine the size and number of servers that are required to support the expected usage. You need to include enough hosts to meet the overall requirements, but additional factors influence these calculations as well.

It is important to consider the amount of CPU and memory resource sharing that is allowed in the environment when you calculate the number and size of hosts. This is influenced by the hypervisor that is in use and its efficiency in supporting CPU and memory overcommitment. For instance, the architect must determine the overall vCPU/CPU ratio that will be allowed in this environment. This is not always easy to determine, but guidance is available from hypervisor or CMP vendors on the best ratio for different workload types. The various hypervisors also have different methods to manage memory both inside and outside the virtual machine instances. Consult the vendor documentation to understand configuration options and best practices. Once you determine the overcommitment ratios, you can apply this to the overall service capacity requirements to determine overall physical server capacity requirements.

The planned services also influence the server sizing exercise. For instance, in an IaaS or PaaS model, implementing a large number of smaller-sized instances may lead to a design with many servers with a high CPU core count and moderate memory capacity. On the other hand, if the plan is for fewer instances, each with a large memory footprint, the design may have fewer servers but with more memory added to each. The important thing to consider is that even though you have calculated the overall physical CPU and memory consumption, deployed instances can only exist on one server at a time. Each server must have enough resources to accommodate a certain number of complete instances as well as spare capacity.

## Compute Capacity Calculations Cont.

- High availability
  - Host failure, maintenance
- Server isolation
  - DMZ
  - Compliance
- Organization constraints
  - Vendor preferences or relationships
    - Preferred pricing

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



12

High availability requirements influence the server size and number decisions as well. If high availability is implemented at the hypervisor layer such as in a VMware environment, then additional capacity will be required for failover. This will most likely include additional servers being added to the design and additional capacity per server for overhead. If high availability is designed into the application layer, then additional servers may be required to house duplicate instances or even additional services such as load balancers. It is also important to remember not only that hosts fail, but also they will require maintenance so the design must have proper capacity and procedures to accommodate planned outages.

Although a goal of cloud computing is to share resources more efficiently, there may be times when you want to isolate some servers from others. There are many reasons for this, including a tenant's demand for isolated resources, compliance requirements, or a physically isolated DMZ requirement for public-facing services. Any hardware isolation requirements may add to the overall number of servers in the design.

Finally, the organization may have a relationship with certain vendors and want to use these vendors' products in the design. Using the preferred vendor may place constraints on the design because of limited server configuration options or a preferred pricing promotion. This could impact the number and sizing of servers in the design.

## Discussion: Sizing Servers



Many small servers



Fewer large servers

? What are the advantages and disadvantages of these approaches?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



13

## Review: Sizing Servers



Many small servers



Fewer large servers

### Areas of impact:

- Maintenance
- Cost
- Licensing
- Consolidation

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

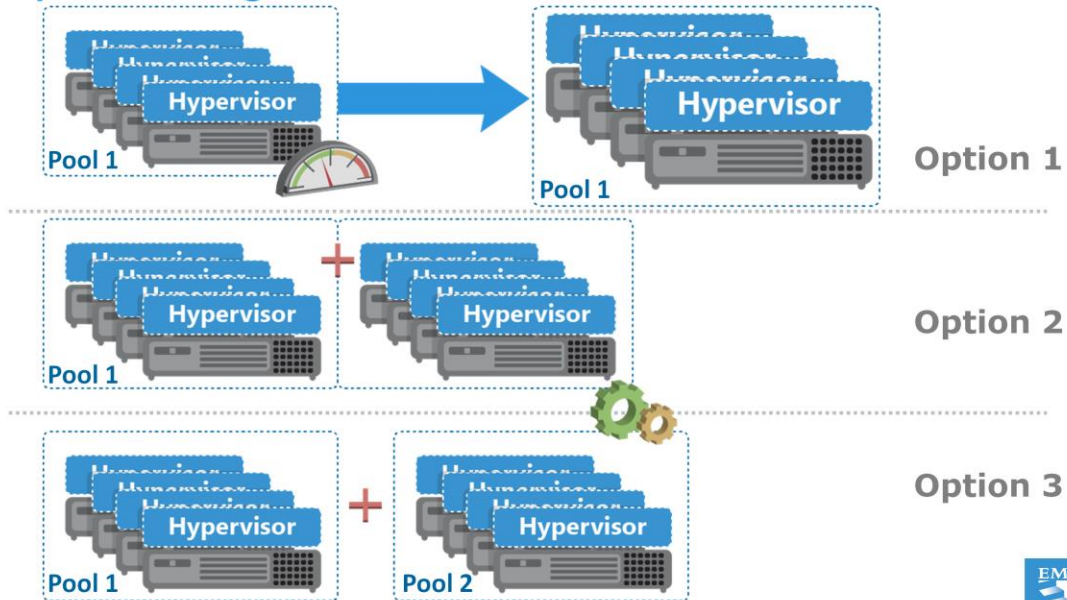


14

### Items to consider when sizing a server:

- Maintenance – Upgrading and patching hypervisors is simpler with fewer servers, as is maintaining firmware and BIOS on the servers and associated components. However, more servers may be required to accommodate moving resources for planned outages.
- Hardware Cost – Two smaller servers are often less expensive than a comparable larger server, when looking at base hardware cost (that is, chassis, CPU, memory, and so on). However, you must also factor in the number of I/O cards in the equation. Assume that each server requires a minimum of two of each I/O card (NIC, CNA, HBA) for redundancy. A larger server that can consolidate the memory requirements of two smaller servers may only require two I/O cards. If you were to use the two smaller servers, that would translate to four I/O cards. Each I/O card also represents an associated switch port, cable, optic, and so on.
- Licensing – How is the hypervisor licensed? If it's licensed per physical host, then larger servers may be more cost effective. Are there other factors to licensing, such as CPU and memory?
- Level of consolidation – Larger servers typically allow you to consolidate more virtual machine instances onto a single hypervisor. This can significantly reduce the amount of space utilized, as well as power connections, cooling, and other operational expenses.

## Compute Scaling



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



15

From the consumer's perspective, a cloud should appear to have unlimited resources. Services may consist of traditional workloads that require scaling up or cloud-native workloads that may require scaling out. Most likely the cloud infrastructure will not be static since requirements change and growth should be expected. These three statements help to enforce the importance of creating a design, and processes and procedures that address how and when cloud resources will grow. Expect cloud adoption to be successful in the organization.

Unlike the services themselves, the physical servers supporting the infrastructure need time to procure, upgrade, and build. This means understanding the options and including the necessary components to support growth. At a minimum, the design needs to have the right monitoring tools included so that the organization stakeholders are able to monitor resource consumption, and also have time to react when server purchases are required.

The first option is to scale-up by adding resources to the existing compute pool. This option is mostly a manual process and requires that processes be defined to move services between hosts to avoid outages.

The second option is to add more servers to the existing pool. Completely documenting the standard server configuration is necessary to ensure that any additionally purchased servers have similar performance characteristics. Although manual processes are involved, some of the work can be controlled and automated using orchestration and configuration management tools to more rapidly deploy resources and to ensure a consistent configuration across hosts.

The final option is to define the compute pool as a known unit of consumption. Again, a documented, standard server configuration is necessary to ensure similar performance characteristics. As additional resources are necessary, additional pools can be added to the infrastructure that will have a known performance and capacity characteristics. This option too, will require orchestration and configuration management tools for a more efficient and consistent deployment process.

# Lesson 1

## Summary

Key points covered in this lesson:

- Compute pools are defined by requirements and group by characteristics.
- Compute sizing is calculated using planned service consumption, availability options, overhead and over-commitment.
- Compute pools can be scaled up or scaled out.



This lesson covered the requirements and considerations for sizing servers and designing compute pools.

## Lesson 2

### Host Software and Functionality

This lesson covers the following topics:

- Requirements for selecting host software
- Design considerations for enabling compute high availability
- Design consideration for host security



This lesson covers the requirements and considerations for selecting host software, and supporting high availability and security.



## Hypervisors



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



18

A hypervisor is software that is installed on a computer system which provides abstracted versions of resources such as CPU, memory, network interface cards, storage interface cards, and disks to a special container called a virtual machine. Although using a hypervisor is not a requirement for cloud computing, it is used in most cloud infrastructures for more efficient resource sharing, rapid scaling, and cost control. There are multiple hypervisors available today and they can be either proprietary or open. Proprietary hypervisors include VMware ESXi and Microsoft Hyper-V. Open hypervisors include KVM (Kernel-based Virtual Machine) and XEN. Similar to the CMP arena, vendors such as Red Hat, Canonical, Oracle, and Citrix offer distributions and support for open hypervisors.

## Hypervisor Considerations



- Hypervisor must be supported by CMP
  - May be additional layers or configuration to enable this
- Physical server compatibility and support
  - Includes components installed within server
- Requirements may dictate usage of advanced hypervisor features
- Understand licensing costs
- Organization expertise
  - May add additional design constraints or costs

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



19

Though the requirements guide the selection of a hypervisor, the following items influence the decision regarding which hypervisor to choose:

**CMP support** – Not all CMPs support all hypervisors, and in some cases the hypervisor support may be limited or may need additional configuration or application layers.

**Physical infrastructure** – A hypervisor must have the proper drivers and plug-ins to support the underlying physical components such as network adaptors, storage adaptors, storage devices or arrays, and CPU architecture.

**Advanced features** – Some hypervisors such as ESXi and Hyper-V natively support features like host clustering, and this may be desirable. Other examples of advanced features are automatic load distribution, memory management, and network/storage path traffic distribution.

**Licensing** – Some hypervisor require the purchase of licenses while other do not.

**Organizational comfort or expertise** – If an organization has developed expertise or a relationship with a particular vendor, then it may be difficult to switch to a different hypervisor technology. If the organization does not have expertise in the proposed hypervisor technology, then this will mean additional costs for support contracts and training.

## Discussion: Implementing Multiple Hypervisors



An organization is currently using the VMware ESXi hypervisor in its datacenter. They are considering using the KVM hypervisor on Red Hat Linux for their cloud.

.....

- ? What are the advantages and disadvantages of deploying a second hypervisor?



## Review: Implementing Multiple Hypervisors



An organization is currently using the VMware ESXi hypervisor in its datacenter. They are considering using the KVM hypervisor on Red Hat Linux for their cloud.

### Areas of impact:

- Management
- Flexibility
- BC/DR
- Performance
- Compatibility
- Cost

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



21

Listed are some of the benefits and tradeoffs between using single and multiple hypervisors:

- Simplified management – a single hypervisor environment can be managed with a single tool, giving a global view of the entire environment. A multi-hypervisor environment may require multiple tools to manage the environment, providing a more limited view, as well as requiring administrators to understand multiple systems, configurations, and so on.
- Flexibility in allocating resources – a single hypervisor environment allows administrators to reallocate resources anywhere within the environment.
- BC and DR planning – each hypervisor platform must be treated independently from a recovery perspective, since not all hypervisors have the same options available. This can also complicate the storage and network recovery options, as there may need to be multiple options available within those layers as well.
- Performance and compatibility – a single hypervisor may not be able to run all of the required operating systems within the data center. In addition, optimal performance might only be possible for a particular application if it is paired with a specific hypervisor. Depending on the criticality of the application and the performance improvement, this alone may be a valid reason to deploy a second hypervisor. Also, if the eventual plan is to move to a hybrid or cloud environment, ensure that the hypervisors are compatible with the provider's systems.
- Cost – a single hypervisor generally costs less in terms of operational expenses. However, capital expenses may or may not be higher, depending upon factors such as licensing, hardware costs, and so on.

## Container Hosts



- Base OS
- Less overhead than hypervisor
- Types
  - LXC
  - LXD
  - Libcontainer
  - OpenVZ

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



22

If the organization is using containers, then hosts will be configured using a base operating system rather than a hypervisor. Containers allow for process or application isolation without the overhead of a hypervisor or additional operating system instances. Containers are implemented using an additional virtualization layer on to the OS which is responsible for deployment, resource provisioning, and logical separation of the containers. There are multiple container virtualization environments available such as:

**LXC** – LXC is a userspace interface for Linux kernel container virtualization. It is freely available and managed as a project under LinuxContainers.org. Support relies on the various Linux distributions and associated vendors. LXC gains much of its support through Ubuntu and Canonical.

**LXD** – Also managed under LinuxContainers.org, LXD builds on the LXC platform with interface improvements including an OpenStack Nova plugin. It is also free with support relying on the various Linux distributions.

**Libcontainer** – Libcontainer is a container virtualization product from Docker that is open source and runs on Linux.

**OpenVZ** – OpenVZ is an open sourced, container-based virtualization project for Linux. It is the basis for Virtuozzo which is a commercially available product from Odin.

## Container Hosts Considerations



- May require additional component in CMP
- Ensure base OS compatibility with physical infrastructure when installing on bare metal
- Solutions may require additional software installs on the OS to support advanced features
- Ensure OS compatibility with containerization solution and deployed applications

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



23

The requirements guide the selection of a virtual container solution as well as some of the items below listed here:

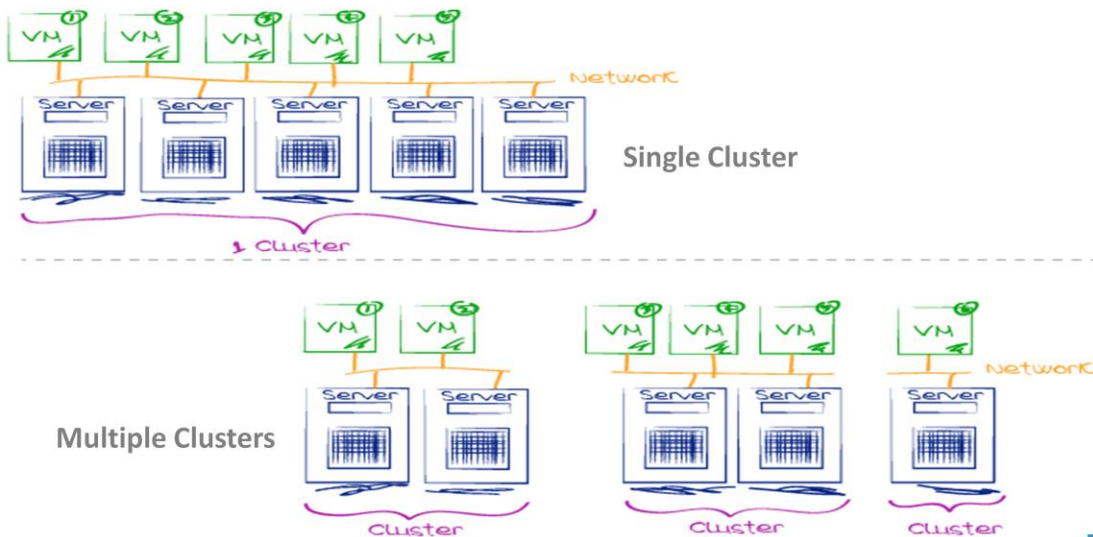
**CMP – Container virtualization** may not be supported by all of the components within the CMP. Additional components may be required to implement a container solution. For instance, an additional element manager may be required to manage the host operating system that is different than the hypervisor manager. The container solution may have additional applications that are required to manage virtual container lifecycles.

**Physical infrastructure** – A base operating system must have the proper drivers and plug-ins to support the underlying physical components such as network adaptors, storage adaptors, storage devices or arrays, and CPU architecture.

**Advanced features** – Some container solutions may require additional layers of application to be installed on the host to support advanced features such as clustering. Consider these additional processes and applications when you are sizing container hosts.

**OS compatibility** – Since virtual containers run on a base operating system it is important to select an OS that is supported by the container management applications. It is also important to realize that since the containers share the same operating system kernel, some applications cannot be deployed across all hosts such as Windows applications on a Linux OS.

## Host Clusters



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



24

Hosts can be clustered for many reasons. In enterprise environments running VMware vSphere it is common to see clusters used for high availability and resource distribution. However, in a public cloud, hosts may not be configured with clustering, as it is expected that availability will be handled at application level and resource distribution will be handled through other means. The cloud design may require a clustering solution for some workloads.

Some of the key considerations for designing clusters include:

**Single clusters** – If clustering is required, the design may include a single cluster that is targeted to handle general stateful workloads.

**Multiple clusters** – Multiple clusters may be required to support different performance tiers.

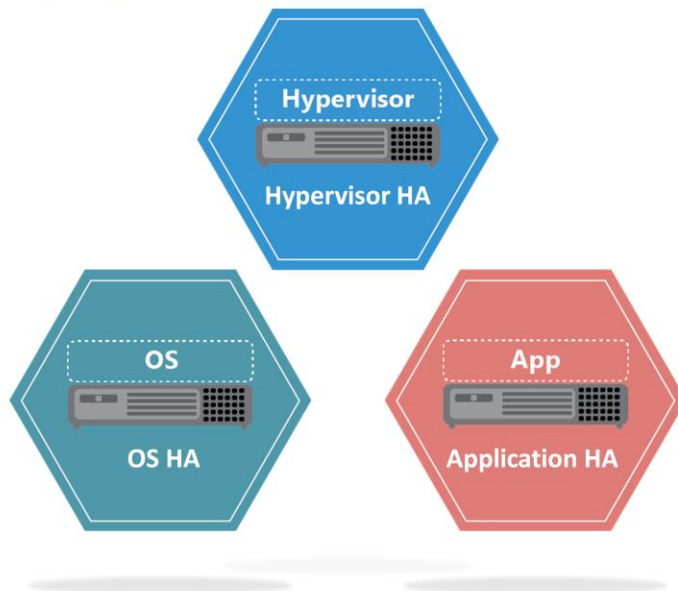
**Capacity planning** – It is simpler to plan for growth with a single general purpose cluster rather than multiple clusters based on performance levels. However, capacity restrictions on the number of servers per cluster, number of VMs per cluster, and so forth, may require multiple clusters even in the general purpose scenario.

**Hardware cost** – Since each cluster requires a planned amount of spare capacity to accommodate failures, having multiple clusters can result in a larger amount of hardware being unused for overhead and failover capacity.

**Security** – Physical isolation for tenants or applications may be a security or compliance requirement and may require multiple clusters.

**Performance** – Separating tenants or applications into separate clusters provides a mechanism to guarantee that resources are available for those tenants or applications.

## High Availability Options For Host Failure



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



25

Some options for maintaining application availability during a host failure include:

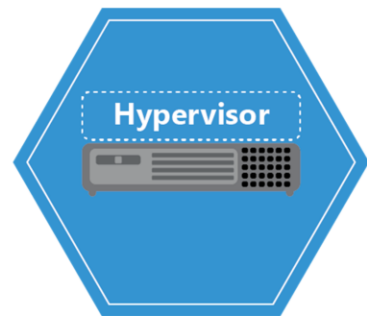
- Hypervisor HA
- Operating System HA
- Application HA

We will review these in the following slides.



## Hypervisor HA

- Implemented at the hypervisor layer
- VM availability
  - Restart or hot standby
- Application recovery
- Considerations
  - Simpler to deploy
  - Time to restart entire VM
  - Additional VM or application awareness



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



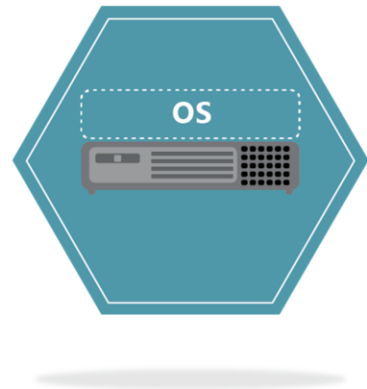
26

One option to provide high availability for services in the event that a host fails is to use advanced capabilities of a hypervisor. VMware and Microsoft, for example, provide a function to restart virtual machine instances on other hosts when the original host fails. Although this is a somewhat easy feature to enable, you should understand that when the host crashes the virtual machine will also crash and must be restarted, and the applications within the instance will be unavailable for a time. The application within the instance must be able to recover from this crash as well.

In many cases, hypervisor high availability is easier to deploy since it is just be a matter clicking a box to enable the option. However, various hypervisors have different high availability capabilities, so enabling this might take more work. Many hypervisors support a virtual machine restart capability after a restart which means that the application does go down for a period of time. Some hypervisors, such as VMware's vSphere, have the capability to keep a virtual machine available during a host failure. This VMware feature is called Fault Tolerance (FT) which provides continuous availability by creating a live shadow instance of a virtual machine that is always up-to-date with the primary virtual machine. In the event of a host failure, FT automatically triggers failover, ensuring zero downtime and preventing data loss. Some hypervisors also have the capability to detect that an application within the VM has failed and can restart the VM for this event. Using the high availability options of a hypervisor would normally happen in a private cloud model, as it would be up to the provider to maintain this capability in the cloud infrastructure.

## Operating System HA

- Clustering capabilities of the OS layer
- Service availability
- Considerations
  - More complex configuration
    - Provider may need to be involved
  - Time to restart services on another cluster node
  - Limited to maximum cluster nodes
  - Failback configuration



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

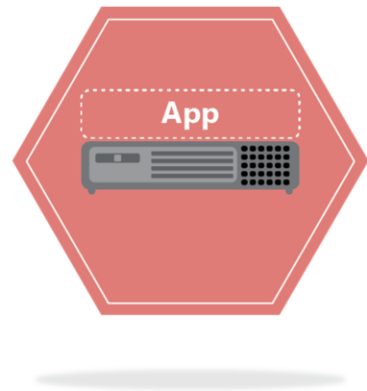


27

Another option for availability is to use clustering capabilities of the operating system within a virtual machine instance. Microsoft has the capability to failover individual services from one instance to another. In this case, at least two instances would be deployed on different hosts. Deploying an OS cluster within a hypervisor environment is generally a bit more complex, as there are certain disk configurations that are required for clusters. Since the hypervisor typically masks much of that configuration from the VM, special considerations must be taken into account. Also, deploying services into a cluster usually requires additional IP addresses, DNS names, and other network components to allow that service to float between the cluster nodes. The time to restart the services on another node should be considered against the amount of time to restart the entire VM. However, the cluster is natively able to detect a service failure and trigger a restart. Examine the maximum number of cluster nodes that the OS supports. You may need to have multiple clusters configured for your services, which can complicate the environment even further. Finally, be aware of how failback is configured, especially if you are using both hypervisor HA and OS clusters. If the cluster is configured to fail the resource back immediately, then when the VM is restarted, the service will experience another outage to return to normal operating status. Because of the complexity of this solution, it most likely will be up to the provider to maintain this capability in the cloud infrastructure.

## Application HA

- Availability options are designed into the application
- Requires supporting technologies
  - Load balancers
  - Replication technologies
- May require more but simpler infrastructure
- Ensure redundant copies are stored on different hosts
  - Impact on pool naming



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



28

Hypervisor and OS availability options are not generally used in public clouds. In the public cloud, hardware is expected to fail and applications are expected to be stateless or redundant. Developers are expected to build high availability into the applications. At a high level, this may mean including load balancers between consumers and application front-ends, using replication technologies for back-end databases or data stores, and redeploying instances upon failure. From an infrastructure perspective, this makes the design simpler. It is true that there may be additional requirements for more virtual machines instances, virtual load balancers, an orchestration engine, and so on, but this means additional host capacity and proper network connectivity. It does not require advanced infrastructure capabilities, licensing, processes, or monitoring. And now, the cost and support associated for high availability options becomes directly linked to the application and business or consumer.

For application high availability, developers must ensure that the redundant copies are stored on different hypervisors so that a server failure doesn't impact the entire application. Since using compute pools masks the details of which hosts are running applications within a pool, it would be a good practice to create multiple pools identified by a failure domain to ensure a host or domain failure does not impact the application.

## Host Accounts

- Minimal number of local accounts
  - Physical server management
  - Hypervisor management
- Deploy separate authentication source
- Deploy separate public key infrastructure

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

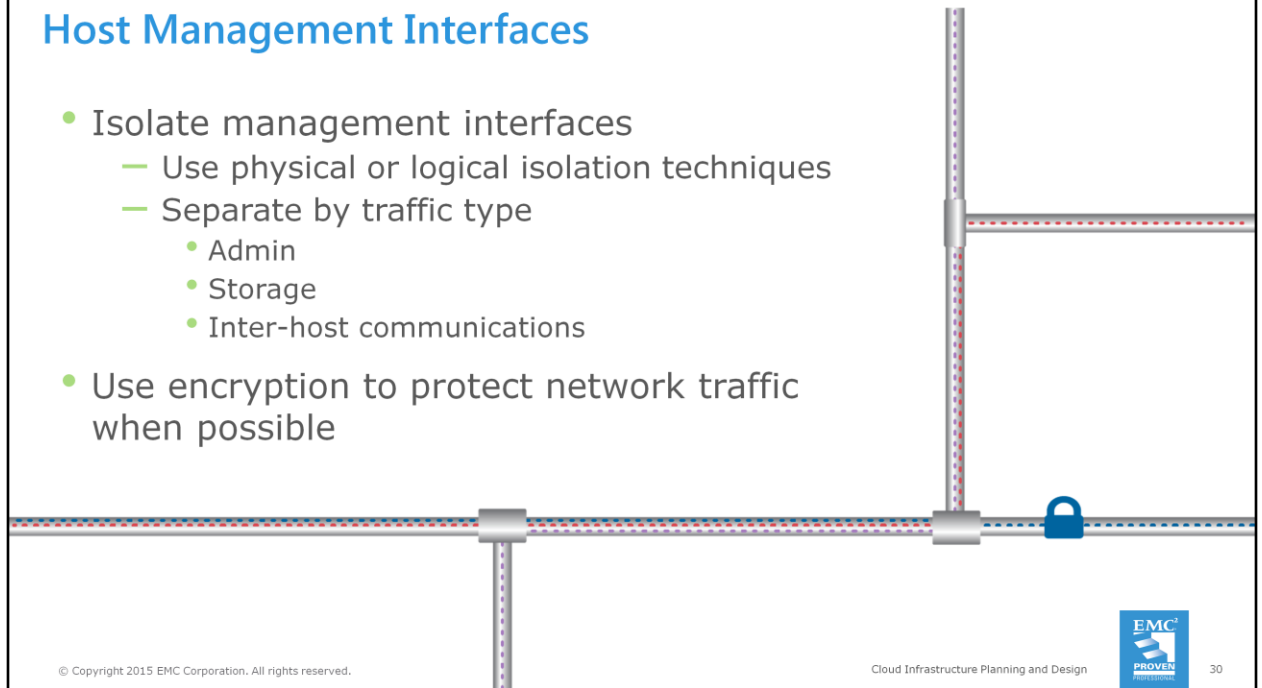


29

As mentioned previously, identifying trust zones is an important part of cloud design. Most likely, hosts will reside in multiple zones since they have administrative interfaces that only privileged accounts should access, and they also contain the services for consumers. The administrative interfaces may include physical server remote management, hypervisor management GUI, CLI and API, hypervisor storage interfaces, CMP component APIs, and compute element managers or controllers. Consider using a separate PKI for each trust zone. Having a separate, internal PKI for all administrative interfaces adds an extra layer of protection since only certificates using the internal authority will be trusted. Deploying a separate, internal authentication service with limited accounts and strong username/password combinations helps secure connections to administrative interfaces as well. If local accounts are necessary, then they should be created sparingly and should also use strong username/password combinations. In either case, the accounts used to access administrative interfaces should be given only the minimum level of permission that is necessary to accomplish the required tasks.

## Host Management Interfaces

- Isolate management interfaces
  - Use physical or logical isolation techniques
  - Separate by traffic type
    - Admin
    - Storage
    - Inter-host communications
- Use encryption to protect network traffic when possible



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



30

Physically or logically separating management and infrastructure traffic from general tenant traffic is recommended and can be accomplished by connecting the management interfaces mentioned previously to a separate physical network or VLAN. The selected hypervisor may support migrating virtual machine instances between hosts, or having storage located on remote servers or arrays. The networks that carry this traffic should also be isolated and encrypted if possible. Only trusted employees of the provider or organization should be given access to any these networks. It is good practice to encrypt all network traffic that connects to the administrative interfaces.

## Host Software

- Software updates
  - Design deliverables should include processes for
    - Testing updates
    - Applying updates
    - Validating hosts after updates
    - Rollback if problems occur
  - Use vendor or community guidance
- Use configuration management tools
  - Ensure consistency
  - Alert on changes



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



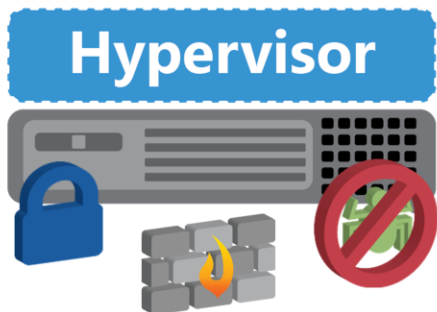
31

The physical host may have a hypervisor or base OS installed which plays an important part in securing consumers' services by isolating them from other services. In addition to securing access to this base software, the design should also include plans and procedures for patching and maintaining anti-virus or intrusion detection applications if possible. This also applies to any element managers or controllers used to centrally manage hosts. Your design security can be hardened further by applying recommended configuration settings to minimize the number of running services. Most vendors and support communities provide guidance for security hardening.

The cloud design deliverables should include procedures for performing upgrades. These procedures should define, when and if upgrades are installed, how services are vacated from hosts, how will the upgrade be tested, how will it be validated and finally what is the rollback procedure.

In addition to base OS hardening, consider implementing a configuration management tool to ensure that consistent host configurations are maintained. Some hypervisor managers such as VMware vCenter natively provide such capabilities. Other third party solutions exist and can extend configuration management capabilities to the individual servers or virtual machines running the management infrastructure.

## Host Protection



- Include host-based antivirus applications to prevent malware attacks
- Include host-based intrusion prevention or detection systems
- May not be supported by hypervisor but other alternatives may exist
- Minimize attack surface by deploying host-based firewall

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



32

Many hypervisors are designed with a minimal footprint and don't allow for installing too many external options. This can be beneficial because it limits the ability for inserting malware into the hypervisor. Some hypervisors may have more exposure to attack and support more options for software installation. This flexibility to install options also applies to container hosts that have a base OS installed.

The architect should consider recommending host-based intrusion prevention and detection and antivirus applications to protect hosts in environments in which malware insertion is possible. Additionally, deploying a hypervisor-based IPS or IDS or antivirus solution throughout the infrastructure will minimize the spread of malware among all of the services and reduce the compute overhead on the individual services.

Another method for minimizing the attack surface of the host is to enable the internal firewall and reduce the number of available ports on the network. Consider using a central configuration management tool to maintain firewall rule settings. Again, most vendors as well as public organizations provide guidance for security hardening.

## Event Management

- Enable host logging for security and system event notification
  - Use central logging for improved security
  - Restrict access
- Enable event alerting capabilities
- Use a tool that can analyze and correlate events
  - Streamlines alerting
  - Improves troubleshooting and security breach identification

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



33

Host logging should be enabled and access to logs should be restricted. Ensure that the design includes enough resources per host to support logging. A central logging tool should be deployed to collect information from infrastructure hosts. Event logs for items such as host access, system events, failures, and errors should be included in the log information to aid in troubleshooting or to satisfy audit requirements. It may be difficult to provide access to this log information to the individual tenants; processes may be required to provide this information for individual tenant audits.

To aid cloud provider staff with responding to problems, use a tool that can not only collect information from hosts but can provide alerts when certain conditions occur. The capability to correlate events across hosts is also useful in troubleshooting or identifying a security breach.



## Lesson 2

### Summary

Key points covered in this lesson :

- Compute pools are made up of hypervisor or container hosts.
- High availability is accomplished using hypervisor, OS, or application functionality.
- Host security is implemented by securing accounts, management APIs, and host software.



This lesson covered the requirements and considerations for selecting host software, and supporting high availability and security.

## Lab 3: Compute Resources Design

### Purpose:

Develop a design and plan for deploying consumer compute resources in a cloud infrastructure.

### Tasks:

- Identify consumer compute requirements that are relevant to a NanCo cloud design.
- Provide a logical diagram that describes the consumer compute resources and is based on the NanCo requirements.

## Lab 3: Debrief

### Compute Resources Design

**Purpose:**  
Develop a design and plan for deploying consumer compute resources in a cloud infrastructure.

# Module 3

## Compute Resources Summary



Key points covered in this module:

- Compute pools are server resources that are logically grouped based on various characteristics.
- Compute pool sizing is determined by application requirements, installed software overhead, oversubscription ratios, and advanced host capabilities.
- Host security is accomplished by protecting management interfaces, installing patches and layered security applications, and through effective event monitoring.

Cloud Infrastructure Planning and Design



37

This module covered requirements and considerations that relate to the design of consumer compute resources in a cloud design.



# Module 4

## Storage Resources



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges, and considerations of implementing various storage technologies for consumers in a cloud infrastructure.
- Align business requirements when selecting technologies that support storage resources in a cloud infrastructure.
- Develop a design and plan for deploying consumer storage resources in a cloud infrastructure.

Cloud Infrastructure Planning and Design

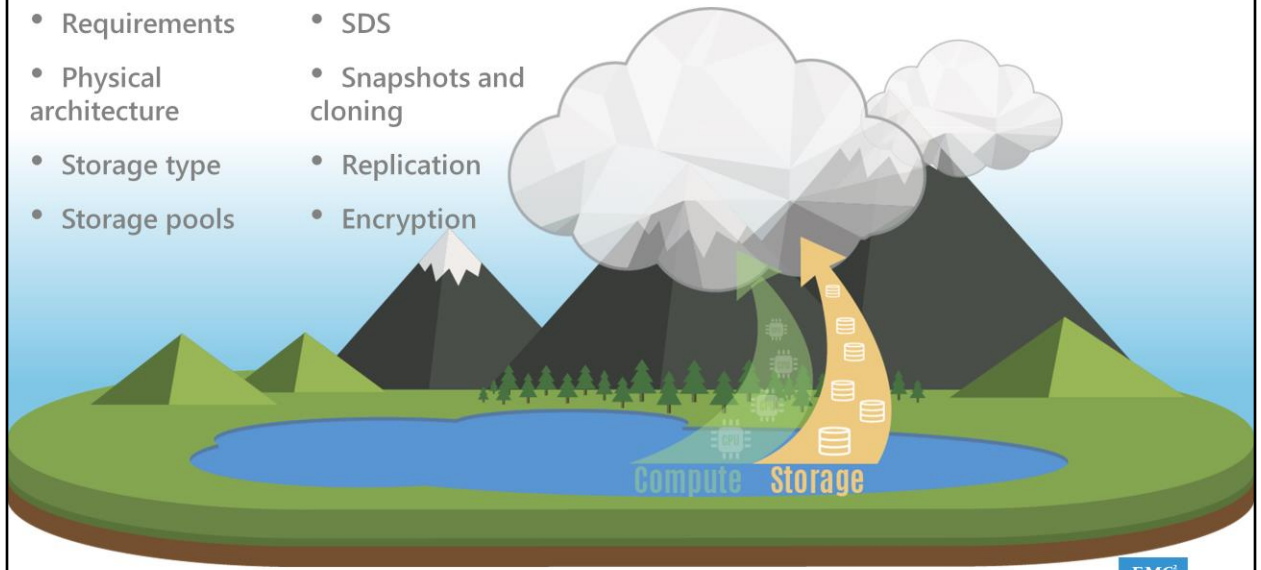


1

This module focuses on technologies and considerations used in designing storage resources for cloud consumers.

## What's This About?

- Requirements
- Physical architecture
- Storage type
- Storage pools
- SDS
- Snapshots and cloning
- Replication
- Encryption



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



2

# Lesson 1

## Storage requirements and architectures

This lesson covers the following topics:

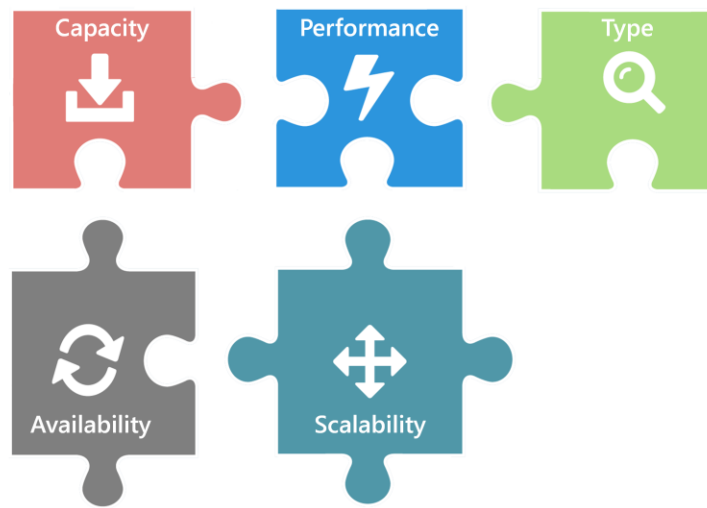
- Requirements for implementing storage resources in a cloud
- Design considerations and features for the various storage architectures



This lesson covers the requirements and considerations for implementing various storage architectures.



## Understand Requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



4

The storage component of a cloud infrastructure design is driven by requirements. Storage requirements fall into categories such as those listed here.

**Capacity** – Capacity requirements for storage are influenced by the amount of live data to be stored, backup or snapshot capabilities, and expected retention periods for data.

**Performance** – Application performance depends on understanding the I/O and latency requirements of the application that are planned for the cloud. Multiple pools of storage may be necessary to address various application performance characteristics.

**Type** – Like datacenter environments, cloud environments most likely will require various types of storage such as block, file, and object.

**Availability** – Data availability can be handled by designing redundant hardware or software infrastructure or will be provided at the application level. Understanding the service requirements is critical because either of these solutions impacts all of the other requirements mentioned here.

**Scalability** – The cloud infrastructure should be designed to support current and future capacity and performance requirements. The service provider needs an easily scalable solution to address future requirements and provide seamless upgrades.

## Understand Requirements Cont.



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



5

The storage component of a cloud infrastructure design is driven by requirements. Storage requirements fall into categories such as those listed here.

**Security** – Data security in the cloud includes data encryption technologies and data separation. Storage security decisions are influenced by compliance and multi-tenancy requirements in the cloud.

**Interoperability** – At a minimum, storage solutions must be able to function with the chosen hypervisor and hosts to support the deployment of services. Additional considerations exist for cloud-enabled storage solutions such as integration with the service catalog, orchestration components, and metering and monitoring capabilities.

**Cost** – Cost is always a requirement and although implementing a cloud may be an expensive endeavor, the cloud architect must constantly balance cost with the organization's requirements and the received benefits.

**Data Integrity** - The primary function of a storage system is to reliably store user data. The accuracy of the data must be maintained from the receiving data on a host write operation, storing the data on the backend media, and then providing that data back to the host on the next read operation.

As we discuss different technologies and design considerations throughout this lesson, we will see how these technologies help to address the different requirements.

## Physical Architecture

- Local Storage
- Distributed Storage
- Central Storage



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



6

Pools of storage can be made up from different physical architectures. Shown here are the three storage architectures that are highlighted in this module.

## Local Storage

- Local physical disks
- Used for
  - Hypervisor or OS installation
  - Non-persistent data
  - Used for applications that are designed to be highly available



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



7

The simplest storage architecture supported in a cloud is local storage. Local storage means that every host that supports services has directly attached storage that can only be used by that host. Local storage is used for host operating systems, hypervisors, and non-persistent data such as temporary or swap files.

Local storage can also be used to support cloud services. In a cloud environment, this may not seem to be the most efficient use of resources. However, with cloud-native applications that are designed to address failure by being redundant, local storage can be used to store operating system images and non-persistent data.

## Local Storage Considerations

- Server form factor
- Hardware
  - Controllers
  - Disks
- Scaling
- Data protection
- Mobility



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

**Server form factor** – In the modern datacenter, there are primarily two server form factors used; blade and rack servers. If hosts will only be using local storage for a hypervisor or OS installation, then either of these form factors will support this since very little capacity is required. However, if the local storage will also support the storage requirements of some services as well, then a rack server will most likely be required to support additional disk capacity, RAID protection, and performance.

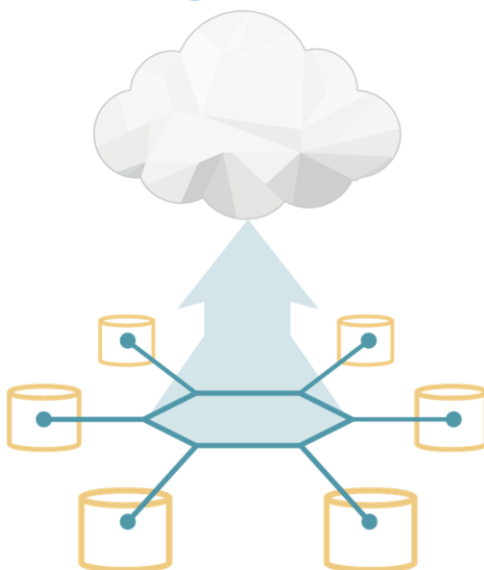
**Hardware** – The number and type of disk controllers depends on the requirements. A standard SCSI controller can provide capacity up to the number of available drive slots, but RAID controllers provide increased protection levels and better performance if drives are aggregated. The requirements also dictate the capacity and speed of the disk drives that are required. In the case of a host using local storage for a hypervisor and temp file system only, a pair of mirrored disks with moderate speed and low capacity will suffice.

**Scaling** – Once local drives have been configured, the only scaling possibilities may be to add additional disks and/or controllers up to the maximum supported configuration of the server or hypervisor/OS. Replacing existing drives for a larger version is impractical.

**Data protection** – In a local disk architecture, data protection is fairly limited and relies upon a RAID configuration which protects only against a drive failure and not a controller failure. In the case of a host using local storage for a hypervisor, if the controller malfunctions or if data is corrupted, the host will fail and a reinstallation may be necessary. If the host is supporting storage for services as well, then application redundancy should be used to maintain service availability.

**Mobility** – If services use local storage, then migration of the service to a different host will rely on support of the installed hypervisor.

## Distributed Storage



- Aggregated local storage from multiple nodes
- Software-defined
- Hyperscale
- Hyper-converged

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



9

Distributed storage is an option that is used in many production cloud environments. The distributed model can use commodity hardware, made up of multiple, inexpensive servers or nodes running specialized storage software. Each node contains local hard disks and is connected by a network to other nodes. This software-defined solution presents the local disks as an aggregate pool(s) of storage to other hosts running services. The nodes can have SSD, SAS, or NL-SAS/SATA disks installed to support any workload performance requirements. When additional capacity or performance is required, more nodes can be added to the solution. Software is responsible for maintaining data integrity and redundancy by placing multiple copies of data on different disks and nodes within the environment. Distributed storage solutions can support block, file and object storage.

Examples of software based storage solutions that can be installed on commodity servers are OpenStack Swift (object), Ceph (Object and block), and EMC ScaleIO (Block).

Hyperscale storage solutions are meant to support data storage only on a set of nodes and compute only on a different set of nodes. Hyper-converged solutions are meant to support the compute capabilities as well as the storage on each node.

Some software-defined storage solutions can be used in either environment.

## Distributed Storage Considerations



- Rack servers provide more options and capacity
- Disk and controller hardware
  - RAID may be used for redundancy or performance
  - Disk speed defined in requirements
- Software support for
  - Underlying hardware
  - Pool configurations
- Distribute storage nodes across underlying fault domains
- Identify scaling limitations and document scaling procedures as part of the deliverables

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



10

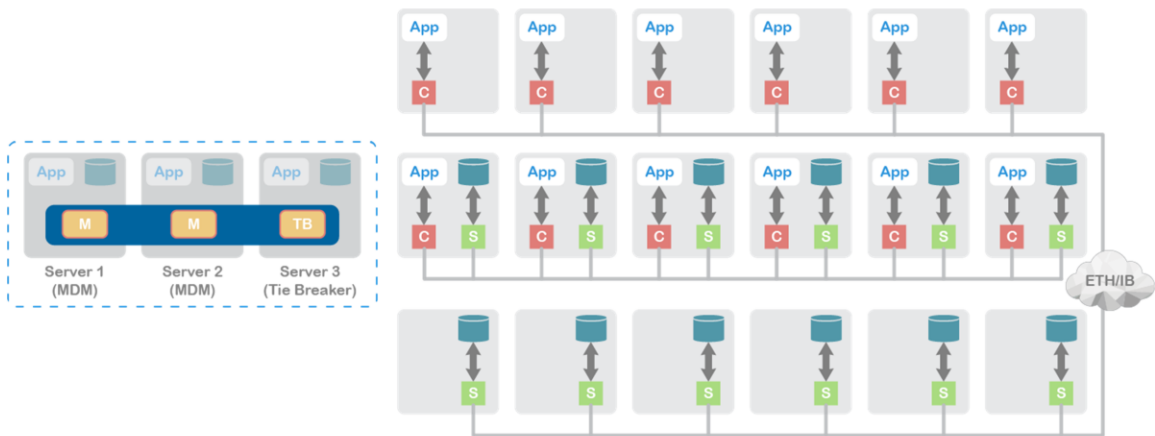
**Server form factor** – A distributed storage architecture usually includes rack servers since they generally support more disk capacity than blade servers. Rack servers come in various sizes and configurations, and choosing the proper configuration depends on the requirements for capacity, performance, scalability, and availability.

**Hardware** – When designing nodes for storage only, more emphasis is placed on throughput so nodes require less CPU and memory than in a hyper-converged model. However, design decisions about types of drives and drive controllers depends on requirements for redundancy, type of storage, and performance. For instance, if distributed block storage is desired, then implementing high speed disk with RAID controllers may satisfy requirements for speed and redundancy. **Software** – Different software-defined storage solutions support different types of storage. Some solutions may support more than one type of storage in the same infrastructure. However, if more than one type is required, then the design may require two separate distributed infrastructures to support performance requirements.

**Fault domains** – Identifying single points of failure within an environment such as individual servers, racks, power supplies, or even datacenters is important. Associating these with a fault domain and then designing storage systems to span these domains helps to ensure data availability. For example, place storage nodes in multiple racks with multiples power supplies and then configure the storage software so that data replicas are placed on multiple racks.

**Scaling** – Deciding whether to scale up by adding more drives to a node vs. scaling out by adding more nodes depends on many factors. First is whether or not the servers have the capacity for more drives and whether adding these drives will reduce performance. Other considerations are the limits and recommended practices for the storage software being used. Physical restrictions such as rack space, floor space, power, and cooling can influence scaling decisions. Whatever the decisions, the design must include a documented description of the scaling procedure, and the procedure must attempt to minimize downtime.

## Distributed Block Storage – EMC ScaleIO



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

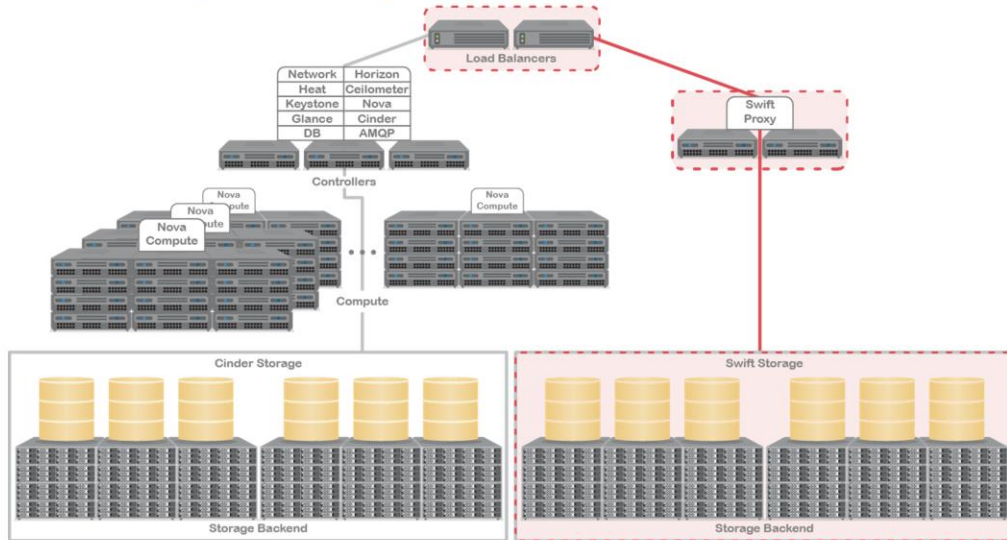


11

Shown here is an example of a distributed block storage solution in a hyper-converged architecture.



# Distributed Object Storage – OpenStack Swift



Based on: OpenStack Architecture Design Guide

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

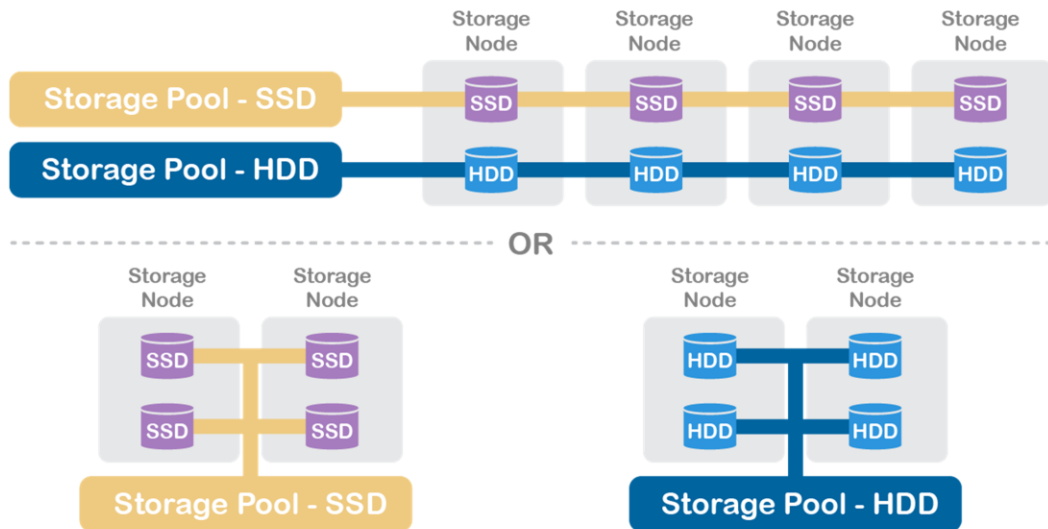


12

Shown here is an example of a distributed object storage solution in a hyperscale architecture.

Source: OpenStack Architecture Design Guide

## Distributed Storage Pools



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



13

In a distributed architecture, storage pools can be defined by aggregating similar storage types that exist across nodes. As in the first example shown here, high performance drives from across all nodes are combined in one pool and the slower speed disk drives are combined into a second pool. Alternately, as shown in the second example, different storage types may be collocated on nodes and the pools can be defined by the groups of nodes containing that storage type.

## Distributed Storage Scalability

|                                      | Scale-up                                      | Scale-out                                  |
|--------------------------------------|-----------------------------------------------|--------------------------------------------|
| <b>Hardware Scaling</b>              | Add more components to existing storage nodes | Add more storage nodes                     |
| <b>Hardware Limits</b>               | Expand to single storage node limits          | Expand to aggregated storage system limits |
| <b>Availability, Resiliency</b>      | Maintained by software                        | Maintained by software                     |
| <b>Storage Management Complexity</b> | Single management interface                   | Single management interface                |

Source: Enterprise Strategy Group, 2010

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



14

The terms scale-up and scale-out are used very often in the technology industry. When applied to a distributed storage architectures, they refer to specific attributes.

Scaling up refers to expanding a storage node by adding more internal components. Some examples include:

- Adding larger or faster drives to increase storage capacity or IOPS
- Adding RAM to improve performance
- Adding additional network interface cards to increases throughput

Adding components to existing storage nodes is limited by the configuration and form-factor of the node. In general for a distributed system, a good practice is to maintain a consistent configuration across like nodes so a scale-up process may require an upgrade of all nodes. Although it may be supported to upgrade some nodes, it could lead to performance hotspots in the future.

Distributed storage solutions are usually designed to scale-out when additional capacity is required. Scaling out involves adding nodes, which contributes additional processing, storage capacity, and network bandwidth to the existing environment. Some solutions rebalance data across the new nodes proactively to maintain consistent performance. With a distributed system, adding additional nodes does not require additional management components. With the proper product and processes in place, scaling out distributed storage should have minimal impact to production activities. Again, it is important to include the proper scale-out process in the cloud design deliverables to ensure maximum uptime.

Understanding current and future capacity requirements, performance requirements, and cost constraints impacts design decisions for scaling.

## Central Storage

- Purpose-built arrays
- Provide for storage consolidation
- Implemented with component redundancy
- May be part of a dedicated storage network
- Provide advanced features

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



15

The central storage architecture consists of multiple hosts sharing a common storage infrastructure, such as one or more storage arrays over a network. Central storage systems are the traditional storage arrays that have been deployed in datacenters for decades. They are purpose-built storage solutions that offer component redundancy, speed optimizations, central management, and capacity efficiencies. Storage arrays provide consolidated storage for the environment and may also provide advanced features such as deduplication, snapshot capabilities, tiering, and redundancy within the array.

## Virtual Provisioning



- Best/most predictable performance
- Precise data placement
- Less concerned about space efficiency



- Space efficiency is needed
- Minimal host impact
- Best energy and capital savings
- For applications where space consumption is difficult to forecast

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



16

Traditional or thick provisioning is best suited for applications that cannot tolerate performance variations, or that require the highest levels of performance. With thick provisioning, the entire capacity is allocated, so generally the smallest possible capacity is allocated to prevent wasted space.

Virtual or thin provisioning is best suited for situations where space efficiency is paramount, or when host disruption cannot be tolerated. With thin provisioning, an enormous amount of capacity can be allocated initially and only a fraction is actually consumed.

## Storage Tiering

- Tiers classified by specific performance characteristics
- Manual tiering method
  - provides control to the cloud consumer or administrator
- Automated tiering method
  - Requires policies
  - Aligns to organization requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



17

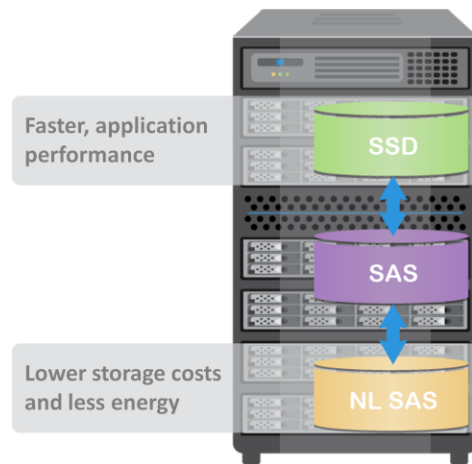
Storage tiering classifies different categories of data and assigns these to various types of storage media. Each tier is defined by performance. Storage is usually laid out in a hierarchical structure with the high performance and high cost storage at the top, and the lower performance and lower cost storage at the bottom.

One method to implement storage tiering is through a manual process. In a traditional datacenter, this meant that the storage administrators would be responsible for initial placement and then would monitor storage workloads periodically and move the data between the tiers. In a cloud environment, the process of manual placement falls to the consumer who requests the required storage tier and pays for its consumption.

An alternative tiering method occurs through the use of automation and policies. This tiering method may or may not allow an initial placement decision to be made by a consumer. After the data is placed, it may be automatically migrated to a different storage tier based on activity levels noted in a policy. For example, frequently accessed data may move to a higher performance storage tier while infrequently accessed data is moved to slower, low cost storage. Policies may also refer to other data classifications or characteristics such as file type, retention period, or performance classification which would then be used for tier placement. Automated tiering helps to align data storage to organizational requirements for performance, cost, data retention and data availability while minimizing errors from human intervention.

## Automated Storage Tiering – Intra Array

- Enables efficient use drives within an array
- Implementations vary by vendor
  - Sub-LUN movement
  - Policy and schedule driven



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

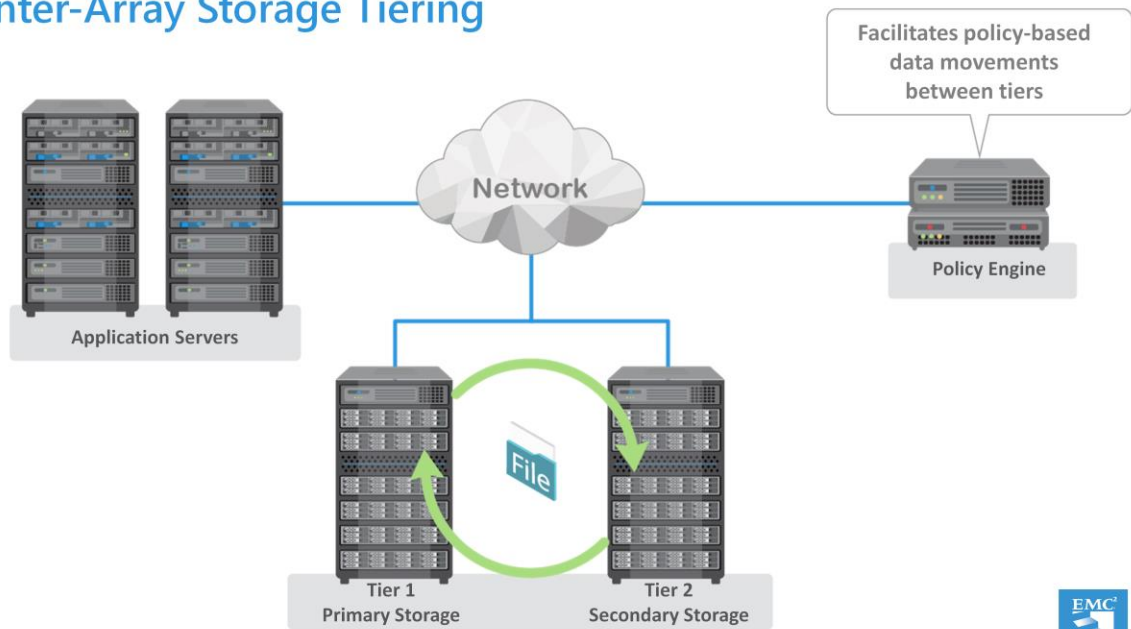


18

Some array vendors have implemented automated storage tiering within their storage array. This enables efficient use of SSDs and NL SAS drive technologies and provides performance and cost optimization. Automated storage tiering proactively monitors application workload and automatically moves the active data to a higher performing SSD tier, and inactive data to a higher capacity, lower performance NL SAS drive tier. The goal is to keep the SSDs busy by storing the most frequently accessed data on them, while moving out less frequently accessed data to NL SAS drives.

Data movements executed between tiers can be performed at the sub-LUN level. This means that only the portion or chunk of the LUN is moved to a different tier but the overall appearance or structure of the LUN is maintained and presented to hosts.

## Inter-Array Storage Tiering



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

19

Storage tiering may also occur automatically across different arrays each with its own performance and capacity classifications. This slide shows an example of a two-tiered storage environment. This environment employs a policy engine, which may reside externally to the arrays, and facilitates moving inactive or infrequently accessed data from the primary to secondary storage. Some of the prevalent reasons to tier data across arrays is for archival or compliance requirements. As an example, the policy engine may be configured to locate all files in the primary storage that have not been accessed in one month, and archive those files to the secondary storage. For each file it archives, the policy engine leaves behind a small space-saving stub file that points to the real data on the secondary storage. When a user tries to access the file at its original location on the primary storage, the user is transparently provided with the actual file to which the stub points from the secondary storage.



# Storage Tiering Considerations

## Intra-Array Tiering

- Tier everything on the array or certain pools?
- Performance considerations?
- Management overhead? Is it automated?

## Inter-Array Tiering

- What do you tier?
- Performance considerations?
- Single management interface or multiple?
- How is data transmitted between systems?
- How does this work with a multi-vendor environment?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



20

When designing intra-array tiering, you first need to determine what you plan to tier. Will you tier all data on the system, or only certain pools? Applications that are highly transactional, or that do not store long-term data, may not be good candidates for tiering as data will always be active.

Performance may also be a consideration. Even if data is being moved to a higher performing level of storage on access, that data has to be accessed at least once on the slower disk, which may cause latency for the application.

Is the tiering automated? Meaning it is performed automatically on the array based on policies? Or, is it initiated manually or via a set schedule? In addition, if the tiering is automated, and it is being performed at all times, is there a performance impact on your storage array?

For inter-array tiering, the same considerations apply in terms of determining what data to tier and performance considerations of accessing data on slower storage. However, there are additional considerations for a multi-array tiering solution. If the arrays are not the same, do you require multiple management tools to configure them? Is the functionality embedded into the array or do you require a software component to be installed in addition to the hardware?

How is data transferred between the systems? Is it done using the front-end (host) ports and network? Does that impact host performance? What protocol(s) are used? Are there security concerns? Or is the data sent across an isolated back-end network? While that can improve performance and security, it can increase management complexity or require additional infrastructure to accommodate it.

What if the systems are from different vendors? Can it be implemented? How is it managed?

## Central Storage Considerations



- Align requirements for performance and capacity with proper hardware
- Identify fault domains to minimize application disruption
- Software
  - Align array software features to the requirements
    - Replication
    - Deduplication
    - Snapshots
    - Tiering
  - Document procedures for upgrades

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



21

**Hardware** – Centralized storage solutions tend to be shared by many hosts and virtual machine instances and have been optimized for this type of usage. Still it is important to understand the requirements for IOPS, capacity, bandwidth and latency so that this can be aligned with a proper array. These requirements impact decisions for disk sizing, disk type, disk number, cache, and front-end processor speed and size. The design also needs to address hardware upgrades and hardware failures within the array.

**Fault Domain** – A storage array is a fault domain. Vendors know that arrays are critical pieces of infrastructure and build in component redundancy to avoid failure. However there is still a possibility that an entire array can have an outage. Availability requirements may dictate that more than one array be implemented and the underlying pools be spread across the arrays to minimize outages. Many array vendor also include replication capabilities so that data can be made available across arrays.

**Software** – Storage arrays also have software installed that provides for management capabilities, data placement, cache control, quality of service capabilities, alerting and other supported enhancements. Cloud design decisions will be based on how software functionality meets the requirements of the organization. The design also needs to address software upgrades and software induced outages within the array.

**Advanced Features** – Central storage arrays can offer a very large pool of storage that is shared by multiple hosts. Performance can become an issue in such environments so consider using storage arrays that offer enhancements such as QOS capabilities, cache, and tiering to maintain performance. Since an array can become a single point of failure, consider arrays with remote replication capabilities that can be used for disaster recovery. To aid in capacity management consider using an array with deduplication capabilities.

## Central Storage Considerations Cont.



- Ensure high availability
  - Component redundancy
  - Path redundancy
  - Controller redundancy
    - Active/active
    - Active/passive

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



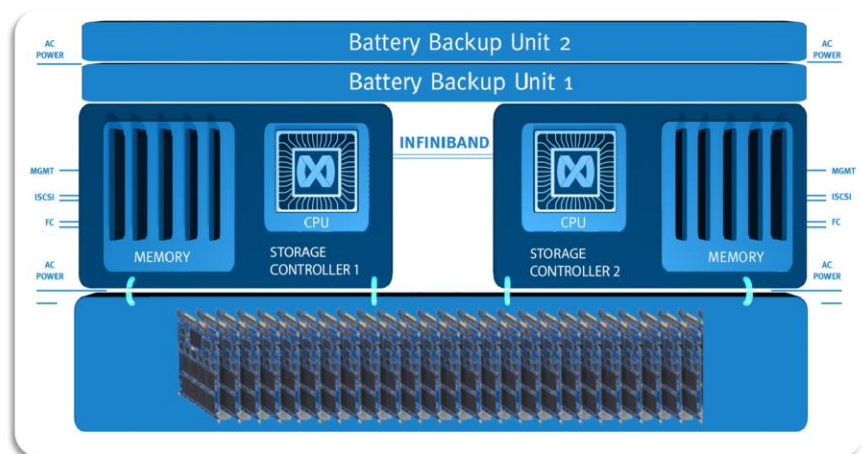
22

In a central storage environment, the design should include an array that contains enough redundancy to prevent a single point of failure to cause data to be unavailable. The central storage solution should also maintain redundancy to protect against multiple concurrent failures. Component failures must be prevented from causing data loss or corruption as well. Most enterprise class storage arrays have high availability functions built in, but it is up to the cloud architect to ensure that the capabilities meet the requirements of the organization.

Storage arrays may offer one of two types of redundancy: passive or active. Passive redundancy includes components that remain idle until needed. An example of this would be active/passive storage controller configuration where only one controller actively serves I/O until a failure occurs. Another example would be designating a hot spare disk drive in an array. In an active redundancy scenario, all components support I/O simultaneously. Active design is the most cost effective; however, careful consideration should be made to ensure that during a component failure, enough resources remain to support the workloads.

## Central Storage – EMC XtremIO

- Active/Active
- Full redundancy
- All flash
- Block storage
- Separate control and data



Source: <https://inside.emc.com/docs/DOC-129803>

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



23

Shown here is an example of a central block storage array. EMC XtremIO is a active/active array with fully redundant hardware and interfaces that support block protocols.

Picture source: <https://inside.emc.com/docs/DOC-129803>

## Central Storage Scalability

|                                      | Scale-up                                      | Scale-out                                   |
|--------------------------------------|-----------------------------------------------|---------------------------------------------|
| <b>Hardware Scaling</b>              | Add more components to existing storage nodes | Add more storage systems                    |
| <b>Hardware Limits</b>               | Expand to single storage system limits        | Expand to aggregated storage systems limits |
| <b>Availability, Resiliency</b>      | Protects against component failure            | Protects against storage system failure     |
| <b>Storage Management Complexity</b> | Single storage system to manage               | Multiple storage systems to manage          |

*Source: Enterprise Strategy Group, 2010*

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



24

Generally, a central storage architecture can be scaled up or scaled out like a distributed storage system.

Scaling up refers to expanding a storage array by adding more internal components. Some examples of scaling up include:

- Adding larger or faster drives to increase storage capacity or IOPS
- Adding cache to improve performance
- Adding storage controllers to increases throughput and IOPS
- Adding more or faster front-end (host) ports to increases throughput

A benefit of a scale-up situation is that a single set of management tools can still be used. If the cloud design supports a scalable storage array, then ensure that the proper processes are documented to scale up the array and avoid an outage.

In some cases, a single storage array cannot scale to the needed capacity, IOPS, throughput, and so on, or an array may have a fixed size. In these situations, it is necessary to scale out by adding more storage arrays. Expanding the number of arrays, however, may add management challenges, especially if the arrays are from different vendors. When scaling out the number of storage arrays, it is also important to understand that the arrays will most likely still function individually, unlike in a distributed system. Scaling out central storage should be less intrusive to production activities as the original array should have minimal interaction. Again, it is important to include the proper scale-out process in the cloud design deliverables to ensure maximum uptime.

As more storage resources are required, solutions must scale-up or scale-out with minimal impact to services. As a part of scalable design, consider using a "building block" approach that allows more resources to be easily deployed as known units of capacity and performance. Understanding current and future capacity requirements, performance requirements, and cost constraints impacts design decisions for scaling.

# Lesson 1

## Summary

Key points covered in this lesson:

- Local storage is used for non-persistent data and host software installations.
- Distributed storage is software defined and can be used in a hyper-scale or hyper-converged environment.
- Central storage provides consolidated storage using purpose-built arrays.



This lesson covered the design considerations and options for various storage architectures.

## Lesson 2

### Storage Types and Pools

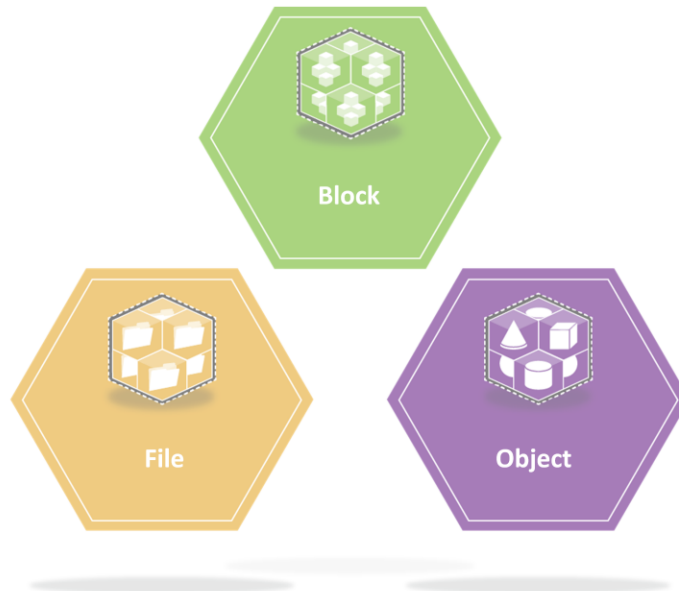
This lesson covers the following topics:

- Design considerations for deploying various storage types
- Requirements and characteristics for defining storage pools



This lesson covers the requirements and considerations for implementing various storage types and pools.

## Storage Types



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



27

Storage can be classified into three different types: block, file, and object. The following slides discuss these types.



## Block Storage



- Physical device
- Data stored in chunks or blocks
- Use block protocols
  - SCSI
  - iSCSI
  - FC



© Copyright 2015 EMC Corporation. All rights reserved.

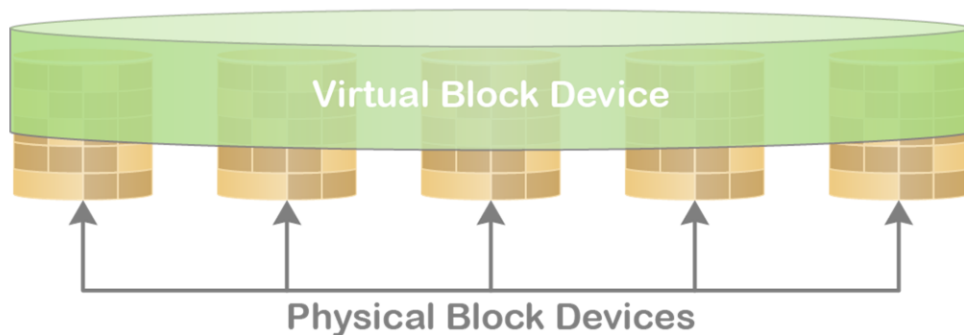
Cloud Infrastructure Planning and Design



28

Block storage is storage that is presented in the form of a physical disk drive where data is stored and managed in chunks called blocks and is accessed using block protocols such as SCSI, iSCSI, and Fibre Channel. Block devices are flexible and can be used as boot devices for an operating system and for storing data. Since the supporting infrastructure and protocols are optimized and are highly efficient, block devices tend to be used for high performance and low latency applications. In most cases, once a block device is presented to a host, partitions, volumes, and files systems are created on these devices so that they can be used by applications.

# Block Storage Virtualization



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

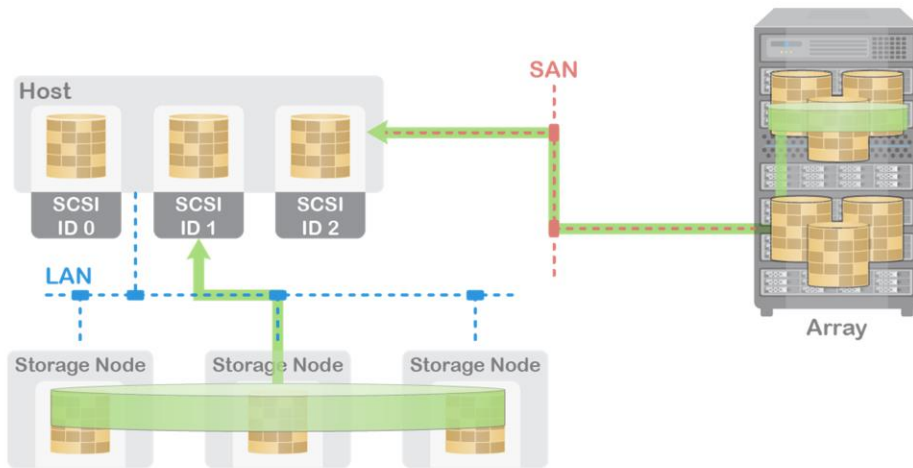


29

Physical disk drives are not always accessed individually to obtain block storage. Physical disk drives can also be aggregated and then presented to a host as a logical or virtual block device. This virtual block device behaves just like a physical device, responding to block level commands and protocols.

Physical disk drives can be aggregated for various reasons. One reason is to provide additional layers of data protection such as RAID. In this instance multiple physical disks are combined to provide a specific level of protection and then sliced into LUNs or volumes and presented to hosts as virtual block devices. Another reason to aggregate disks is to provide better performance. Adding more disks to an aggregate increases available bandwidth and reduces response time for data retrieval.

## Block Storage for Hosts



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



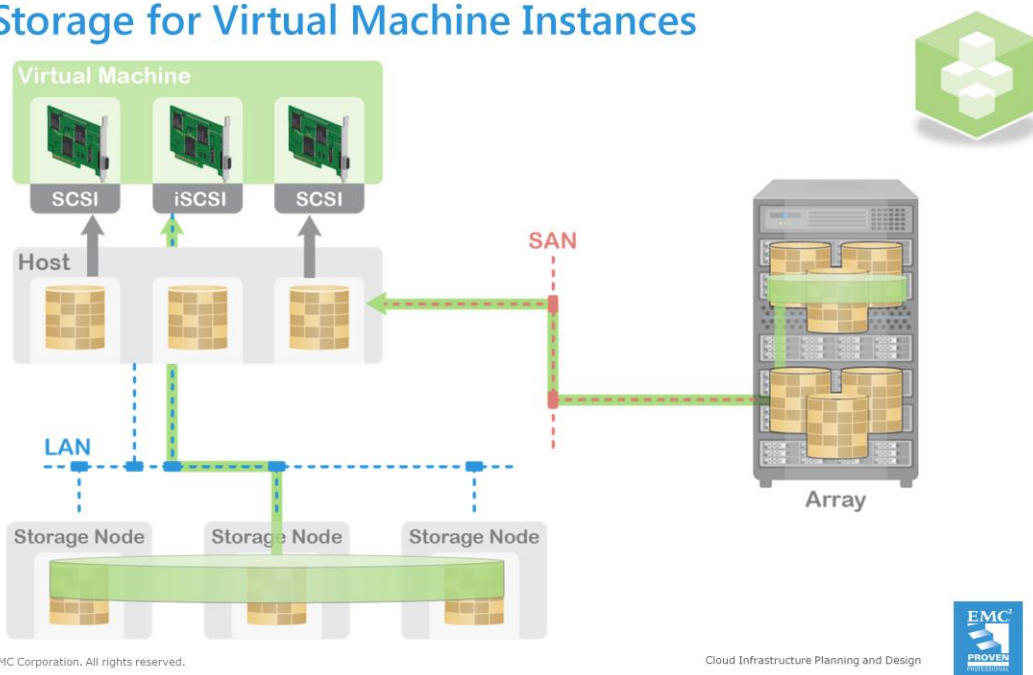
30

A local hard drive is a block device that can be accessed by a physical host.

When using storage arrays, virtual LUNS or volumes are created on the array and presented to hosts and treated as if they were physical drives. These LUNS are made available to a host over a storage area network.

In a distributed storage environment, a virtual LUN or volume is created using the local disk of multiple hosts and presented as a block device to a host. These LUNS are made available over a local area network. EMC ScaleIO and the open source Ceph project are examples of distributed block storage solutions.

## Block Storage for Virtual Machine Instances



In virtualized environments, a virtual disk can be created and presented to a virtual machine which is then treated as a physical block device. Virtual disks must be attached to a block disk controller in order to be recognized as a block device by the operating system. Unlike in the physical environment, virtual disks are not always backed by a physical or virtual block device. Many hypervisors, for example, create a file on a file system, which emulates a block device and presents the file to the virtual machine.

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



31



Cloud Infrastructure Planning and Design

32

In the case where hosts with installed hypervisors are configured with block devices, the hypervisor controls virtual machine access to these devices. Consumers accessing the virtual disks of a virtual machine are prevented access to the virtual disks of other virtual machines unless specifically configured by an administrator. This occurs even if the underlying storage is shared by multiple hosts, virtual machines, and tenants.

## Additional Block Storage Considerations



- Redundancy
  - Cloud-native applications
  - Redundant local or array components
  - Distributed storage using software
- Performance



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



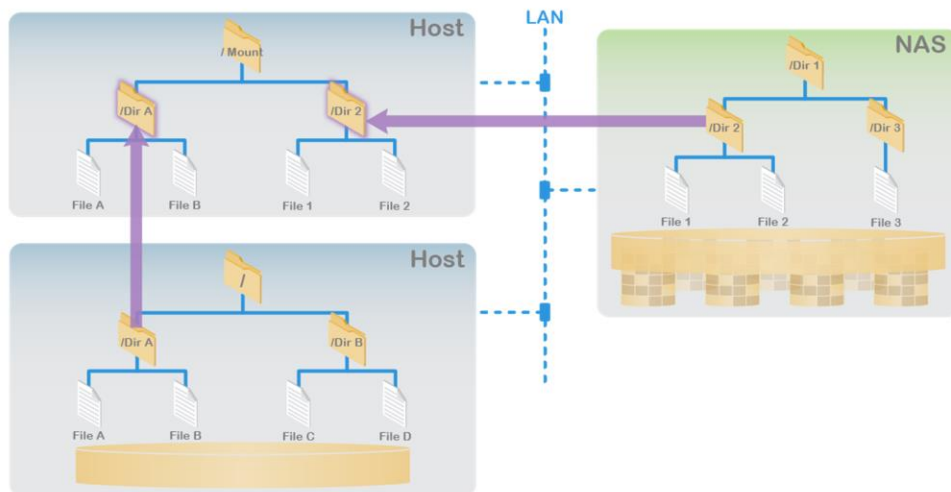
33

**Redundancy** – One method for handling data redundancy of block devices is to deploy cloud-native applications and services where redundancy is built into the service. In this case, storage redundancy is not required and service outages can still be minimized. However, application redundancy may not reduce the storage capacity required since multiple copies of data must still be maintained. Deploying only cloud-native applications may not be possible and the cloud design may require a combination of redundancy solutions.

If data redundancy is required in the storage infrastructure, for instance if more traditionally-architected services are required, then designing for block data redundancy depends on the type of back-end storage selected. Local block storage redundancy can be implemented by using a RAID controller on the host. Distributed storage systems use software to distribute multiple copies of data across storage nodes. Central storage arrays use redundant hardware and RAID to ensure block data redundancy.

**Performance** - Block storage is usually deployed for high performance and low latency applications. Network considerations that affect performance are discussed elsewhere. When designing pools of block storage, it is important to identify the requirements of the planned services. In many cases, performance can be improved by increasing the number of drives or storage nodes in a pool or spreading workloads across pools. Matching the right type of media to a block pool is also important. For example, if services will require extremely high performance, then SSD or fast SAS drives should be considered for the pool and NL-SAS drives should be avoided. Also note that when the pools are added to the catalog for consumers, they should be labeled appropriately to identify performance characteristics.

## File Storage for Hosts



© Copyright 2015 EMC Corporation. All rights reserved.

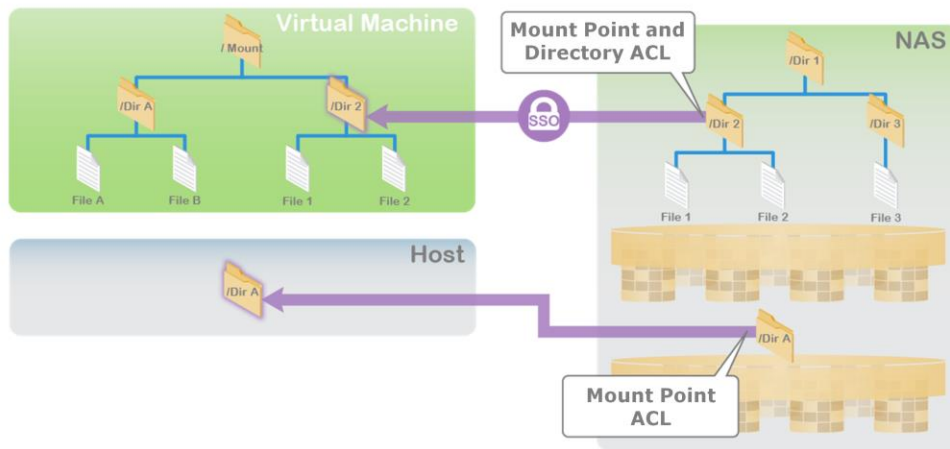
Cloud Infrastructure Planning and Design



34

File storage is storage that is presented to a host as a hierarchical structure of directories and files. In this type of system, data is stored and managed as complete files. Technically, file storage is what you get once you create a file system on a block device. However, file storage can be shared with other hosts across networks and is accessed as if the storage was local to those hosts. Hosts use special protocols to access and share these remote file systems. Remote file systems can also be shared from a central array referred to as a Network Attached Storage (NAS) device. As the name implies, file storage is used to store files and is not used as a boot device.

## File Access and Multi-tenancy



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



35

As with block storage, access to file storage can be controlled to support multi-tenancy. A common way to accomplish this is by creating an access control list that determines which hosts are allowed to access a particular directory. Access can also be controlled through directory and file permissions inside the file system. To support file and directory permissions, a common authentication mechanism is required that is accessible by tenants.



## Additional File Storage Considerations



- Redundancy
  - NAS array with redundant hardware
  - Scale-out NAS array using software
  - Distributed file systems
  - File server clusters with shared storage
- Performance
  - Similar to block storage factors



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

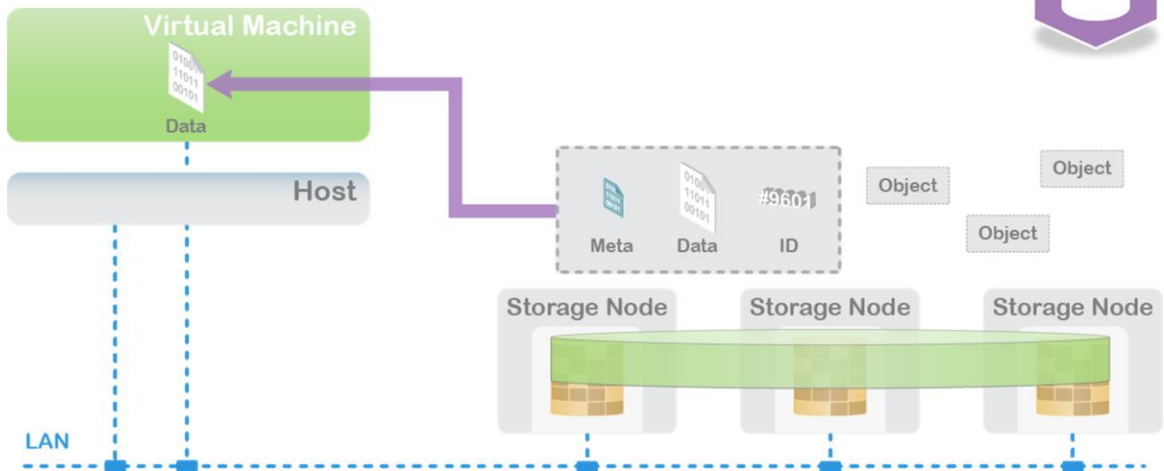


36

**Redundancy** – File storage redundancy can be accomplished in many ways. If the file storage is being offered through a NAS array, then redundancy occurs similarly to a block storage array by eliminating single points of failure using redundant components such as disks and controllers. Scale-out NAS arrays are made up of multiple storage nodes with local storage and use software to create multiple copies of data to maintain redundancy. If a distributed file system is deployed, then like the scale-out NAS, multiple copies of data are distributed across nodes. Finally, if file servers are used to offer file storage then using clustering software for the file services along with shared storage can offer a layer of redundancy.

**Performance** – File storage is usually deployed when the performance requirements are not too demanding. However, file system performance is still tunable by using the proper disk drives and other hardware components to match requirements. Since file storage resides on block storage, many of the factors that affect block performance also affect file storage performance. Again, when designing pools of file storage, it is important to identify the requirements of the planned services.

# Object Storage



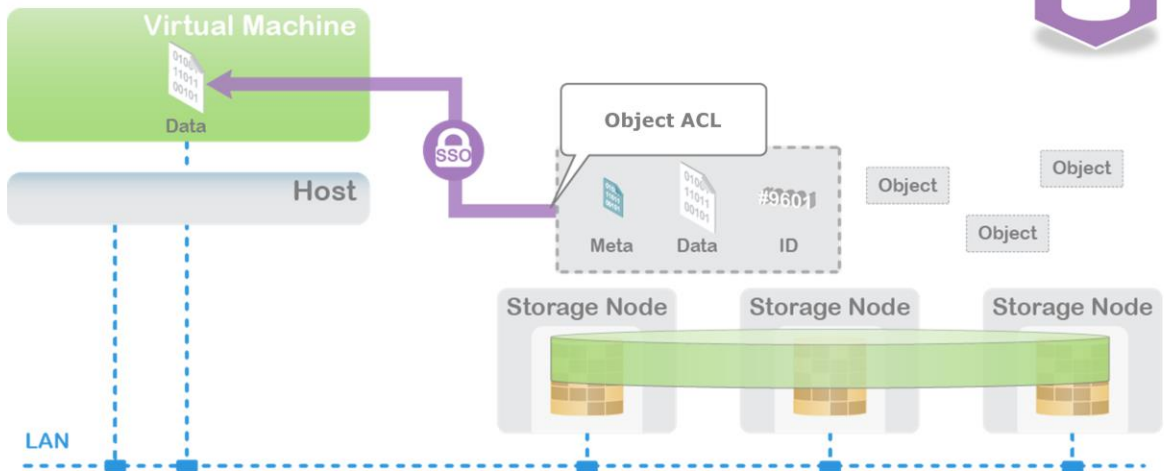
© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

37

Object storage consists of data that is stored and managed as objects in a flat repository. Each object represents the data itself, associated metadata and a unique identifier. Where block and file storage systems tend to be confined within a physical set of devices or volumes, object storage systems are usually designed to span multiple sets of hardware. Object storage systems also have no hierarchical structures so a catalog of objects must be maintained so that data can be located and retrieved when necessary. Object storage is used for data storage and is not used for boot devices.

## Object Access and Multi-Tenancy



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



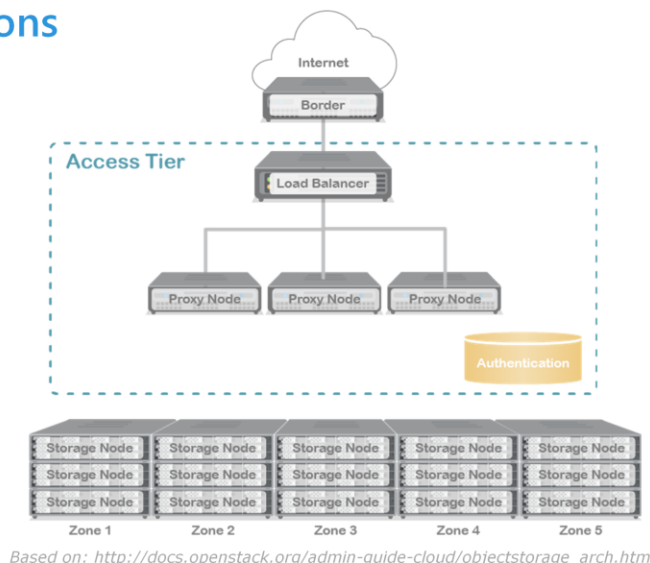
38

Object storage supports multi-tenancy through object permissions. To support object permissions, a common authentication mechanism is required that is accessible by tenants.

Some solutions may offer additional multi-tenancy capabilities. OpenStack Swift, for example, allows for the configuration of tenant-specific image locations.

## Example: Swift Considerations

- Front-end connectivity
- Redundancy
- Performance



Based on: [http://docs.openstack.org/admin-guide-cloud/objectstorage\\_arch.html](http://docs.openstack.org/admin-guide-cloud/objectstorage_arch.html)

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



39

**Front-end** – Some object storage solutions have separate front-end servers that are used to access the storage. From a network perspective, these frontend servers should be located close to the backend storage nodes to ensure performance. If supported, multiple front-end servers should be deployed using load balancers to distribute traffic and redirect during a node failure. The front-end servers should have enough compute resources to handle the additional overhead for network traffic encryption. If possible, consider using multiple network interfaces on each node, one used for public facing API traffic and the other for backend storage node traffic.

**Redundancy** – Data redundancy is handled by software where multiple replicas of data are spread across nodes. The number of replicas is configurable and should match the requirements for data redundancy.

**Performance** – Although object storage may not be used for fast access applications, an acceptable level of performance should be maintained according to the requirements for the design. At the node level, using a single high bandwidth NIC or multiple bonded NICs improves performance. Object storage software is designed to support concurrent requests across multiple spindles; however, single threaded requests are impacted by the speed of a single disk. If necessary, using higher speed disks can improve performance as well as adding additional disk per node to spread the concurrent load. At the pool level, adding additional nodes to the storage pool increases overall performance.

Picture Source: [http://docs.openstack.org/admin-guide-cloud/objectstorage\\_arch.html](http://docs.openstack.org/admin-guide-cloud/objectstorage_arch.html)

## Example: EMC ECS Considerations



- Connectivity
- Redundancy
- Performance

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



40

EMC Elastic Cloud Storage (ECS) is a complete, software-defined cloud storage platform that supports the storage, manipulation, and analysis of unstructured data on a massive scale on commodity hardware. ECS is specifically designed to support mobile, cloud, big data, and social networking applications. It can be deployed as a turnkey storage appliance or as a software product that can be installed on a set of qualified commodity servers and disks.

**Connectivity** - Within the ECS appliance, multiple nodes are connected together via top of rack switches. These switches must be connected to upstream switches for access to the management interface and APIs. As appliances are added to scale out within a datacenter, they must be interconnected via a private management network that supports inter-service communications. In a multi-site configuration, the maximum latency of 1000 ms is allowed between sites.

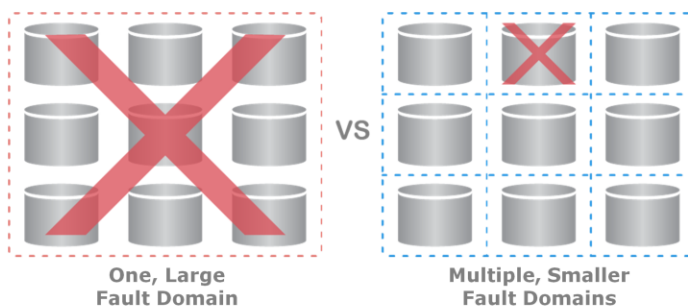
For front-end connectivity the use of a load balancer or global load balancer should be used for performance and availability.

**Redundancy** - ECS uses a replication group construct that defines where storage pool content is protected. Replication groups can be local or global. Local replication groups protect objects within the same VDC against disk or node failures. Global replication groups protect objects against disk, node, and site failures.

**Performance** - ECS is a scale out solution. Adding more appliances not only increases capacity but also provides additional compute, network, and storage resources to maintain performance.

## Multiprotocol Solution Considerations

- Simpler management
- Efficient utilization
- Single vendor support
- Single point of failure
- Performance



*Based on: The Case for Tiered Storage in Private Clouds – Cloudscaling.com*

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



41

Having a single storage solution that can provide block, file, and object storage may be a valid option for some organizations. A multi-protocol solution can offer simpler management, better storage efficiencies, and a single vendor support solution. However, adding this to a cloud design may also introduce other considerations. A single multi-protocol solution may be a single point of failure and additional redundancy may need to be added to the design. Another consideration is performance. Some solutions are essentially an implementation of a single protocol solution with additional layers added to support the other protocols. An example is Ceph, which essentially is object storage with additional block and file interfaces layered on top. It is a scale-out solution but an organization may not be able to obtain the performance required for block level applications.

An alternative to a single multiprotocol solution is to deploy a limited number of solutions, each supporting different performance characteristics to ensure certain performance guarantees. This also increases the number of fault domains.

Picture Source: The Case for Tiered Storage in Private Clouds – Randy Bias. (2014, Feb 23). Retrieved from [www.cloudscaling.com](http://www.cloudscaling.com)

## Storage Pools

- Aggregate of storage
- Presented to consumers
- Associated with characteristics
  - Determined from requirements
- May include single or multiple arrays, single or multiple hosts



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



42

In a cloud environment, a storage pool is an aggregate of media used to store data. Storage pools are presented to the consumer and can be dedicated to a single tenant or may be shared across all tenants in a cloud. Usually the pool is associated with a set of characteristics that is common across all of the media within that pool. For instance, you may define a storage pool to meet specific performance characteristics such as high performance and low latency. Another pool characteristic may be the level of protection offered for the data. For example, a cloud could offer two pools of object storage, one configured with a higher level of redundancy than the other pool.

The number of pools and the characteristics that define the pools are determined from the requirements gathered during the assessment. The number of pools varies on cloud designs. The goal should be to define the smallest number of pools that can support the requirements.

In many traditional datacenters, storage pools existed as collections of disks within an enterprise class storage array. In a cloud environment, a storage pool may include storage from a single array, multiple arrays or even multiple hosts with local storage. The point of storage pools is to present tenants with a collection of storage that performs or functions consistently and meets a certain set of requirements.

## Storage Pool Examples

| Pool                    | Characteristics                                          | Usage                      |
|-------------------------|----------------------------------------------------------|----------------------------|
| <b>Mission Critical</b> | high performance, high availability, block, replicated   | Databases, Enterprise Apps |
| <b>General</b>          | Moderate performance, block, moderate recovery time      | General Apps               |
| <b>Near-line</b>        | Low performance, high volume, file, longer recovery time | General Data Repository    |

| Pool               | Characteristics                                         | Usage                   |
|--------------------|---------------------------------------------------------|-------------------------|
| <b>Production</b>  | high performance, high availability, block, recoverable | Production Apps         |
| <b>Development</b> | Moderate performance, block                             | Development Apps        |
| <b>Data</b>        | Low performance, high volume, object                    | General Data Repository |

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



43

Shown here are just two examples of storage pool layouts.



## Pools in the Cloud

|                               | Non-Cloud     |   | Cloud                     |
|-------------------------------|---------------|---|---------------------------|
| <b>Tier 1<br/>Performance</b> | SAN, some NAS | → | Elastic Block Storage     |
| <b>Tier 2<br/>Balance</b>     | NAS, some SAN | → | Distributed File Systems  |
| <b>Tier 3<br/>Capacity</b>    | Tape Backup   | → | Object Storage, some Tape |

*Picture Source: The Case for Tiered Storage in Private Clouds - Cloudscaling*

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

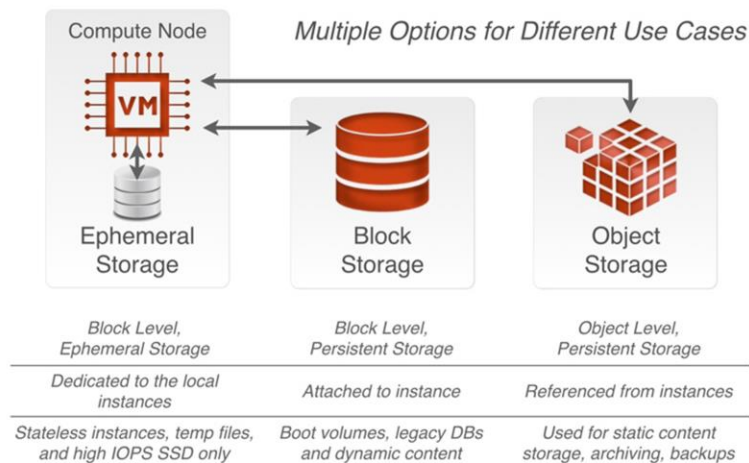


44

The cloud provides an alternative view of tiered storage. Rather than using backend methods to tier storage, in the cloud we can leave tiering decisions to the developers and consumers. In this scenario, the cloud infrastructure supplies the individual storage pools that represent the tiers and the applications and tools use the tiers as required.

Picture Source: The Case for Tiered Storage in Private Clouds - Cloudscaling

# OpenStack Deployment Discussion



Picture Source: The Case for Tiered Storage in Private Clouds - Cloudscaling

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



45

In an OpenStack environment, there are generally three different types of storage: ephemeral, block, and object. Ephemeral storage is dedicated to an instance and contains stateless data that only exists as long as the instance exists. Block storage contains persistent data which can be attached to any instance and is usually used to support higher performance requirements of services. Object storage also contains persistent data and is referenced by instances or used by the CMP itself. Object storage supports slower performance and higher capacity requirements such as static content, archives, or backups.

Picture Source: The Case for Tiered Storage in Private Clouds - Cloudscaling

## Lesson 2

### Summary

Key points covered in this lesson:

- Block storage supports high performance requirements and is accessed using block protocols.
- File storage supports general performance and user access requirements.
- Object storage is a scalable storage solution that is accessed through a RESTful API.



This lesson covered the design considerations and options for various storage types.

# Lesson 3

## Advanced Storage Functionality

This lesson covers the following topics:

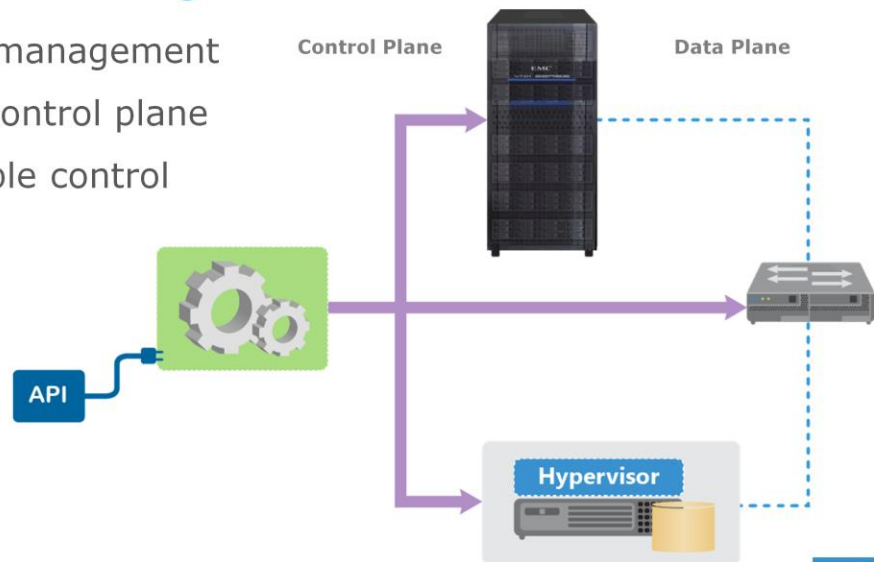
- Design considerations for deploying software-defined storage
- Options and design considerations for deploying storage snapshots, replication, and encryption



This lesson covers the requirements and considerations for implementing advanced storage functionality.

## Software-Defined Storage

- Centralized management
- Abstracted control plane
- Programmable control
- Agile



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



48

Software-Defined Storage (SDS) is a solution that decouples the storage control and data planes, enabling directly programmable storage provisioning and the abstraction or virtualization of the underlying storage infrastructure. In this solution, an SDS controller is used as a central point of entry for deploying and managing storage which is virtualized and presented to hosts or services for consumption. A software-defined storage solution is hardware agnostic and may interface with centralized storage arrays, decentralized storage hosts, hosts supporting services, and the underlying storage network.

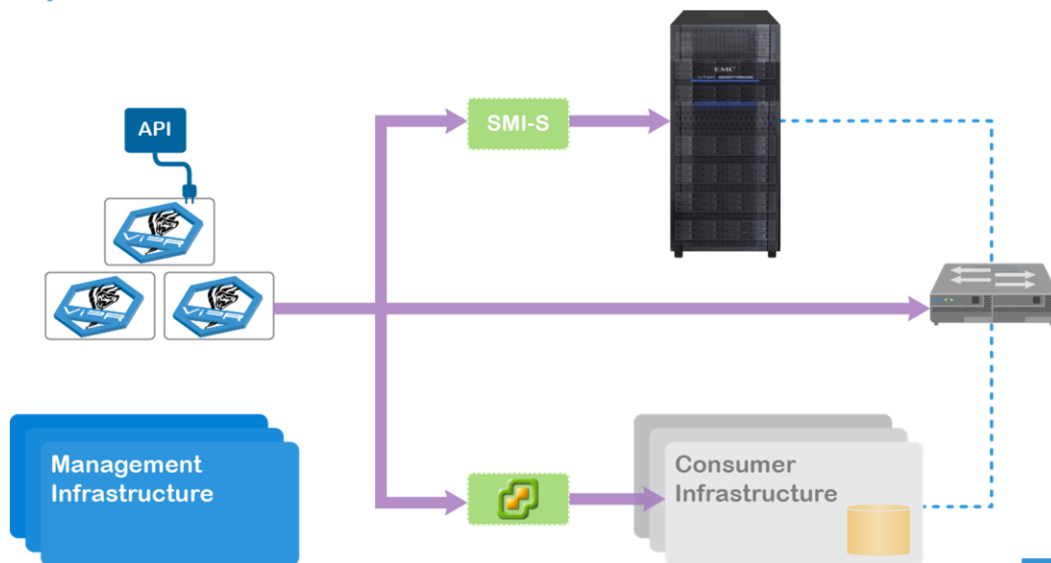
When implementing a software-defined storage solution, some key considerations are listed here:

The controller will most likely be deployed in the infrastructure supporting the cloud management components. If the SDS solution requires direct access to manipulate or control physical arrays or storage network resources, then the design needs to include access through an isolated or secured network.

The controller may also need access to the hypervisor or hypervisor manager to enable access to storage.

The controller API may require access from the cloud management platform to support deploying and controlling cloud services. The organization may also have a requirement to enable API access from deployed services in the cloud.

## Example: ViPR Controller Architecture



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



49

Here we will use EMC ViPR Controllers deployment to illustrate considerations that may be included in the cloud design to support a software-defined solution. The first design decision would be the size of the environment to be provisioned. In the case of ViPR there are two different controller configurations: 3 and 5 nodes. These controllers are virtual machines and will exist in the cloud management infrastructure. The management infrastructure must have enough resources to support whichever configuration is selected.

To manage and provision block storage, ViPR controllers must have access to the management interfaces of the block storage solutions. If switch configuration such as zoning is required, then the controller needs access to the management interfaces of the switches. ViPR can provision storage to hypervisors such as VMware vSphere. The controller will also need access to the hypervisor management interface.

## Snapshots and Cloning

- Snapshot
  - Point-in-time reference to data
  - Usually maintained by pointers
  - Created at storage system or hypervisors
- Clone
  - Point-in-time copy of data
- Definitions may vary



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



50

A snapshot is a point-in-time reference to data. It is usually represented by freezing current data and using pointers to reference current and future data. This means that the original volume cannot be removed if a snapshot exists. Snapshots can be created by underlying storage infrastructure or at the hypervisor layer. From a storage infrastructure, a snapshot can be taken of a virtual block device or a file system. From the hypervisor, snapshots can be taken of virtual machines. The technologies for creating snapshots vary and will impact capacity and performance of the cloud infrastructure. If snapshot capabilities are required, then the architect will need to research the various solutions to ensure they meet the requirements of the organization.

Clones are also a point-in-time representation of data but usually involve creating a completely separate copy of the data.

Snapshots and clones may have different meanings in different environments. In OpenStack, for example, when creating an instance snapshot, a complete copy of the instance is created and stored in object storage as an image. However, a block volume snapshot is just as described above. It is a point-in-time reference to the volume and is stored within the block storage.

## Snapshots and Clones Considerations



- Block array-based snapshots and clones are performed at the LUN level
- File array-based snapshots may be performed at the file level as well as the entire file system
- Hypervisor-based snapshots are performed at the VM level
- Not a replacement for backups

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



51

Using snapshots and clones in a classic environment provides a mechanism to very quickly (nearly instantaneously) restore data to a previous point in time configuration. While this does not provide a solution for disaster recovery, it does provide an effective solution to counter data loss or corruption.

In a cloud environment, however, care must be taken when using snapshots and clones on central storage arrays. Snapshots and clones are taken at the LUN level, and if they are reverted, this too will be at the LUN level. Because there are often multiple VMs storing their virtual disks on the same LUN, it is very possible to inadvertently roll back all of VMs during a recovery event and not just the desired VM.

To avoid this scenario, a solution could be to deploy LUNs for each virtual machine directly from block storage. This may be a feasible solution in an elastic block storage environment. Another alternative is to use the hypervisor capability to create a snapshot or clone of the individual virtual machines.

For file-level storage, in addition to the full file system recovery (which has the same issues as the block level copies), there is often a mechanism to recover individual files using snapshots. This can be a simple or complex process, depending on the hypervisor and the storage capabilities. The process may be lengthy as well, as it may need to recover data from a number of disparate snapshots and reassemble them into the final state of the file.

It is important to understand that snapshots and clones may exist in the same pool or same array as the original disk. If that pool or array becomes inaccessible, then so do the clones and backups. Also, snapshots use pointers to the original disk, so if the disk is corrupted, then so is the snapshot.



## Replication

- Synchronizing of data between two locations
  - Are requirements for asynchronous or synchronous replication?
- Implemented at the application layer
  - Design requires additional storage capacity and network bandwidth
- Implemented at storage infrastructure layer
  - Will storage solution support it?
  - Will secondary site have same performance?
  - Is source and target same vendor?
  - What is required to manage source and target?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



52

Replication is the synchronizing of data either synchronously or asynchronously between two locations. Replication is generally used to support some type of disaster recovery capabilities.

With cloud-native applications, data replication may be built into the application and from an infrastructure standpoint; the design needs only to support the additional storage and the network bandwidth requirements for the second copy. It is assumed that if one data instance becomes unavailable, the application will failover to the replicated instance.

However, if replication is desired within the storage infrastructure, this adds complexity to the design. Before exploring design considerations for replication, you should understand the requirements.

Some questions for requirements gathering are:

- What data needs to be replicated?
- Does the storage at the second site have the same performance requirements?
- Is the storage from both sites from the same vendor?
- What are the options for management capabilities at both sites?

## Replication Considerations



- Bandwidth and latency
  - Synchronous
    - Latency must be very low
  - Asynchronous
    - Bandwidth must support established RPO
- Mirrors expect to have identical storage configuration
- Partial or degraded performance could possibly use a different storage system
- Management capabilities need to exist in both locations

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



53

One of the most critical aspects of replication is the amount of bandwidth available between the sites, and the latency associated with the connectivity. Using synchronous replication provides the benefit of a full mirror of your data at all times. Since synchronous replication requires data to be written to both arrays simultaneously, the data will always be identical. However, synchronous replication requires enough bandwidth to transmit the changing data in real time and a very low latency—essentially treating the storage as if it were locally connected. If the connectivity does not meet this requirement, application performance may suffer, and the replication processes may actually fail.

For asynchronous replication, the bandwidth and latency requirements are not as stringent, since data is not being written simultaneously. However, the connectivity must be able to support the established RPO. In either case, if the available connectivity cannot meet the requirements, an acceleration product may be used to increase throughput and reduce latency.

Since the replication process is directly impacted by the amount of data that is being transmitted between sites, if you are expecting to have a full mirrored environment at the recovery site, you should expect to have an identically-configured storage system there. This not only provides the same performance in the event of a failover, but also provides the necessary performance during the replication process. If you are only planning to replicate a portion of the environment, or expect to run in a degraded state when failed over, you can potentially save on capital expenses by using a smaller storage system, or a system from a different vendor. This also introduces the potential requirement for an external replication product; however, that can accommodate replication across heterogeneous storage devices.

If storage arrays are located on multiple sites and a single site goes down, then the cloud design needs to include additional components located in the secondary site to support storage management and provisioning.

## Storage System Data Encryption

- Data-at-rest
- Privacy
- Loss of media control
- Requirements
  - Strength
  - Compliance
  - Performance
  - Complexity
  - Key management
  - Cost



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

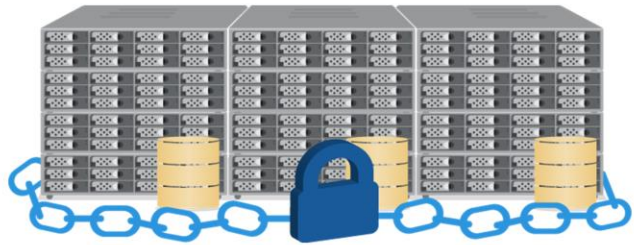


54

Cloud computing is based on a shared, multi-tenant architecture where traditional security controls are not sufficient and data owners are responsible for securing sensitive data. Data security solutions should address requirements for privacy, regulatory compliance, and loss of control of media. Data that is stored on a storage array, or even backup media, can be encrypted for protection. Before selecting a mechanism for encrypting data, the assessment should provide the requirements for which data needs to be encrypted and the strength of encryption. If the organization falls under specific compliance regulations, those will be a source for encryption requirements as well. Architects must also take into consideration factors such as performance, complexity, key management, and cost.

## Storage System Data Encryption Options

- Application
- File system
- Network-based
- Storage array
- Storage device



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



55

Data encryption can be implemented at the application layer. One advantage to this is that only the data that truly needs to be encrypted can be identified and addressed. This method can minimize cost, performance impact and complexity of the solution. However, this also means that users are responsible for properly identifying the data that must be encrypted and that the application must support encryption.

Encryption can also occur at the individual file level or for an entire file system. This form of encryption can be executed by many modern operating systems or third party applications running on a server. Encrypting individual files again leaves the security burden on the user where encrypting the entire file system removes them from the process. File system encryption requires that additional overhead resources be added to the cloud design. Key management can be complex for this option, especially when implementing hundreds or thousands of virtual machine instances.

Network-based solutions allow for the encryption of data as it enters the network at the switch port all the to the storage array. Some solutions integrate with storage arrays and tape units to offer encryption of the stored data as well. These solutions offer advantages such as centralized management, centralized key management, and minimal server overhead. However, implementing this type of solution takes careful planning to ensure proper performance and identify compatible components.

Some storage array vendors also offer encryption capabilities which are usually executed at the storage processor. Storage processor encryption adds more overhead requirements and this should be considered in the design. Storage array solutions also offer centralized management and key management.

Self-encrypting storage devices provide encryption embedded in the storage device. With this solution, key management is unnecessary since each device has its own key which is not accessible outside the device. This solution offers negligible overhead during data reads and writes. Another advantage is that the process for deleting disk data is as simple as deleting the key. Without the key, the data cannot be retrieved.

## Data Encryption Considerations

- Key management
- Virtual machines

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

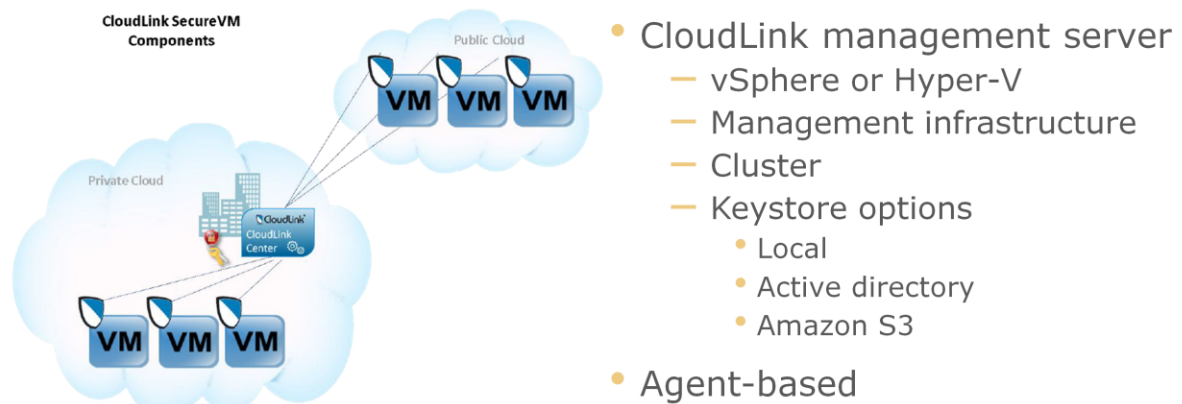


56

Data encryption requires keys that are used to encrypt and decrypt the data. The keys must be secured but must also be available to the process performing the encryption and decryption. Any passwords or key recovery strings should also be secured and not stored in any clear text format. Access to the keys should be strictly controlled. Secure backups of the keys should be maintained because loss of a key can result in loss of data. Keys must be kept for as long as the encrypted data exists; however, a process should be put in place to create new keys, change or upgrade existing keys, and destroy keys. This last process is important when an employee leaves the organization and keys need to be replaced.

If a requirement is to encrypt the entire virtual disk of virtual machine instances, then the solution needs to integrate with multiple operating systems and manage keys across these instances. If a single solution is not workable, then it may be necessary for the organization to manage multiple solutions for each type of operating system deployed. Virtual machines, like physical servers, can have different disk combinations such as boot disks and data disks or a combination. The organization may only wish to encrypt data disks to minimize performance impact but this needs to be a configurable option in the solution. Additionally, the solution should be capable of automatically protecting any disks added in the future to avoid accidental data exposure.

## Example: EMC CloudLink SecureVM Architecture



Picture Source: CloudLink SecureVM Version 4.0 Deployment Guide for Enterprise

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



57

An example of a cloud-enabled encryption solution is CloudLink SecureVM, which secures sensitive information within virtual machines (VMs) across both public and private clouds. This solution provides boot partition (sometimes referred to as the “boot volume” in Windows environments) and additional disk encryption with pre-startup authorization for virtual machines hosted in the cloud by using native operating system encryption features: Microsoft BitLocker for Windows and eCryptfs for Linux. BitLocker and eCryptfs are proven and high performance volume encryption solutions widely implemented for physical machines.

The CloudLink Center is the management interface for the solution. It can be deployed as a virtual machine on VMware vSphere or Microsoft Hyper-V environments and should be deployed within the management infrastructure. This server can be deployed as a cluster to maintain availability. It provides key management services as well and the keys can be stored in a local repository, within Microsoft Active Directory or on AmazonS3. Agents are deployed on the virtual machine instances, and these agents require access to the CloudLink Center over a specific TCP port.

Picture Source: CloudLink SecureVM Version 4.0 Deployment Guide for Enterprise

# Lesson 3

### Summary

Key points covered in this lesson:

- Software-defined storage solutions support programmable storage provisioning.
- Snapshots are implemented at the storage and hypervisor layer and create a point-in-time reference to data.
- Data encryption is used to support compliance and privacy requirements and can be implemented at the application layer through the storage infrastructure layer.



This lesson covered the requirements and considerations for implementing advanced storage functionality.

## Lab 4: Storage Resource Design

### Purpose:

Develop a design and plan for deploying consumer storage resources in a cloud infrastructure.

### Tasks:

- Identify consumer storage requirements that are relevant to NanCo cloud design.
- Provide a logical diagram that describes the consumer storage resources and is based on the NanCo requirements.



## Lab 4: Debrief

### Storage Resource Design

**Purpose:**

Develop a design and plan for deploying consumer storage resources in a cloud infrastructure.

# Module 4

## Storage Resources Summary



Key points covered in this module:

- Storage pools are storage resources that are logically grouped based on various characteristics.
- Storage pools can be built with block storage, file storage, or object storage. Each of these storage types has different use cases in the cloud.
- Storage pools can be built on different architectures such as local disks, distributed storage solutions, or central storage arrays.

Cloud Infrastructure Planning and Design



61

This module covered requirements and considerations that relate to the design of consumer storage resources in a cloud design.



# Module 5

## Network Resources



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges and considerations of implementing various network technologies for consumers in a cloud infrastructure.
- Align business requirements when selecting technologies that support network resources in a cloud infrastructure.
- Develop a design and plan for deploying consumer network resources in a cloud infrastructure.



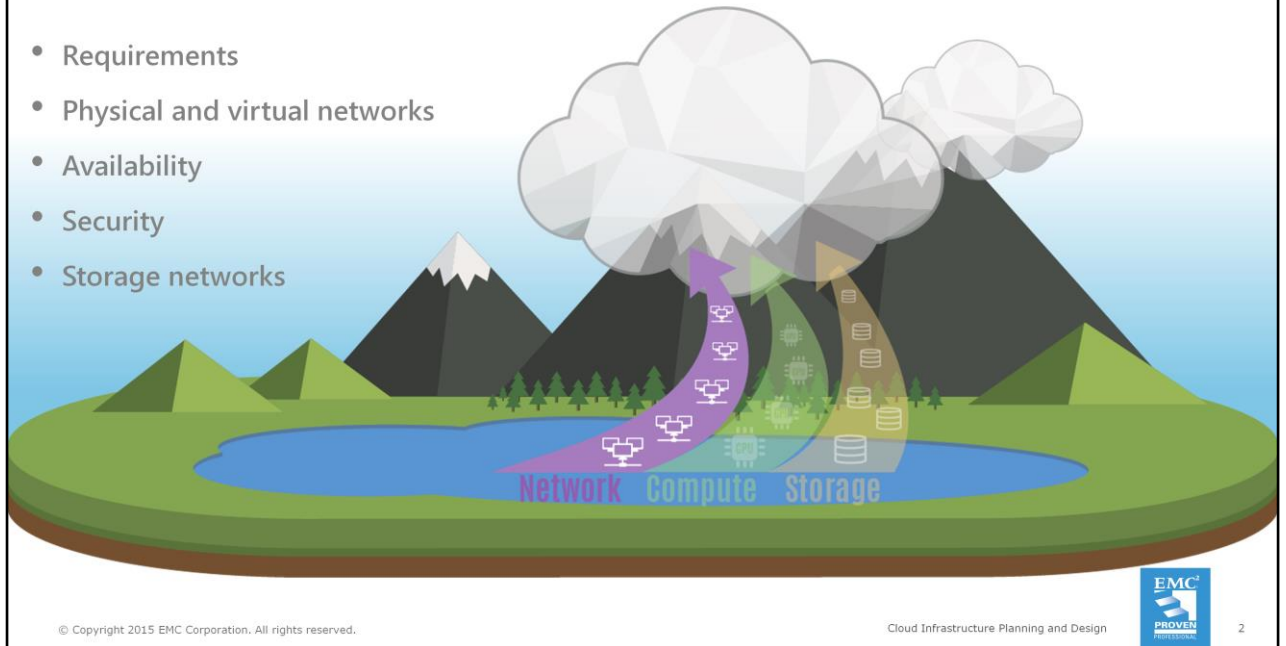
Cloud Infrastructure Planning and Design

1

This module focuses on technologies and considerations used in designing network resources for cloud consumers.

## What's This About?

- Requirements
- Physical and virtual networks
- Availability
- Security
- Storage networks



A cloud must have pools of resources allocated in order to support the planned services. This module highlights the technologies, options, and choices as they relate to network infrastructure that may be included in a cloud design.

Listed here are the high level topics covered in this module.

# Lesson 1

## Local Area Networks

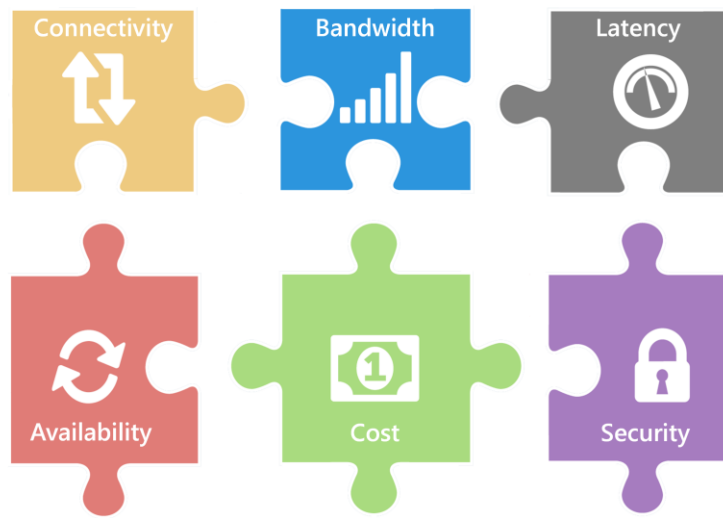
This lesson covers the following topics:

- Requirements for implementing local area network technologies in a cloud
- Design considerations for implementing networks in a cloud environment



This lesson covers the requirements and design consideration for implementing local area networks in a cloud.

## Understand Network Requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



4

The networking component of a cloud infrastructure design is driven by requirements. Network requirements fall into categories such as those listed here.

**Connectivity** – The infrastructure design must support consumer connectivity to services, connectivity between hosts and management elements, connectivity options to link tenant networks to public-facing networks, and the capability to deploy these networks on demand.

**Bandwidth** – The infrastructure design must include sufficient network capacity to support consumer traffic, management traffic, and infrastructure traffic, for example storage. It must also be expandable for future needs.

**Latency** – To reduce network latency, the infrastructure design may include Quality of Service policies, minimal hop counts, and dedicated infrastructure supporting network function components such as load balancers and firewalls.

**Availability** – Physical and virtual network infrastructure requires redundancy to support availability. Also, since a cloud environment should contain services that are designed for failure, the cloud infrastructure needs additional resources to support orchestration, load balancers, and even redundant service instances.

**Cost** – Organizational cost requirements guide decision points and you need to balance cost, performance, and capabilities throughout your design.

**Security** – Security is managed and controlled at many levels. The cloud networking infrastructure requires network isolation, firewalls, and secure access to network applications and components to assist in enforcing security zone boundaries.

As we discuss different technologies and design considerations throughout this lesson, we will see how these technologies help to address the different requirements.

## General Network Design Considerations for Cloud



- Modular
- Scalable
- Supports requirements for bandwidth and latency
- Automated
- Redundant
- Interactions with
  - Virtualization layer
  - SDN
  - NFV

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



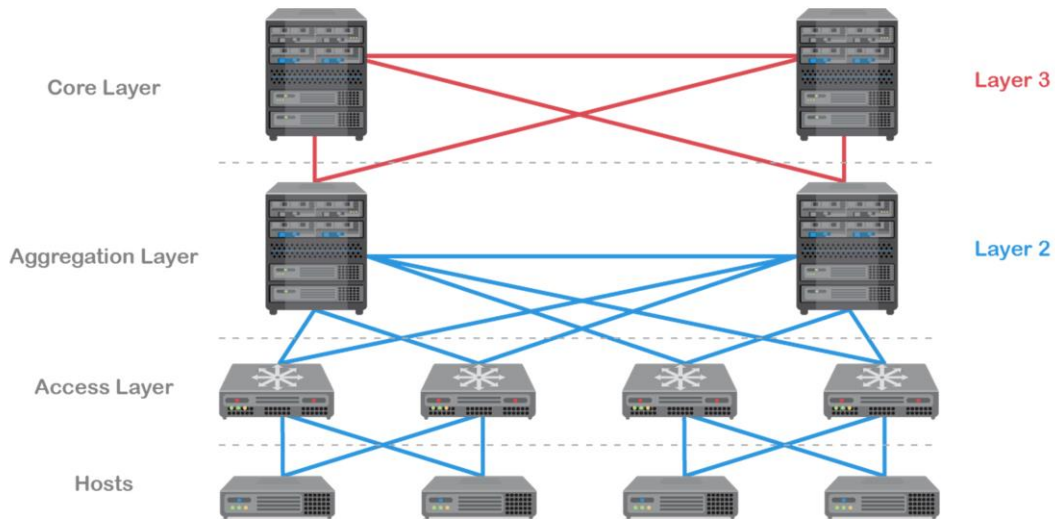
5

A cloud infrastructure design may include a detailed datacenter network design, or the network may already be in place. At the very least, the cloud architect must understand the type of network infrastructure that will support the cloud. The network infrastructure should be modular in design, be able to scale out as more compute resources are added, provide adequate bandwidth to support planned services and include redundant paths and hardware. As the cloud environment grows, so too will the network infrastructure. Consider using automation for network configuration and service deployment to promote consistency and reduce the chance of outage caused by operator error.

Even though services will be designed to minimize the effect of infrastructure failure, it is still important to maintain network redundancy within a single datacenter. Ensure that servers are connected to multiple switches and can failover automatically, that all network paths have some type of redundancy, and that if a path fails, the remaining paths will not become a bottleneck. Researching and understanding how the underlying network infrastructure supports or interacts with other cloud technologies such as network virtualization, software-defined networking (SDN), and network function virtualization (NFV) is an important part of the design process.



## Hierarchical Physical Networks



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

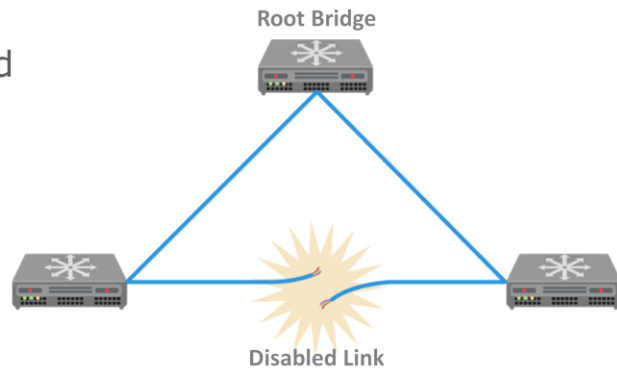


6

Pictured here is an example of a hierarchical design that consists of three layers: core, aggregation, and access. The access layer provides connectivity to the hosts within the data center. It is made up of switches that define Layer 2 networks or IP subnets and the servers or other devices that need to communicate on these networks. The aggregation layer is a combination of Layer 2 and Layer 3 connectivity. Layer 2 connectivity exists between the access layer switches and the aggregation layer switches, as well as within the aggregation layer. Layer 3 connectivity from the aggregation layer to the core layer. The core layer is designed to be a high-speed routing environment that transfers data between Layer 2 networks and to and from external networks.

# Spanning Tree Protocol

- Loop-free topology
- Redundant links are blocked
- Enabled per VLAN



© Copyright 2015 EMC Corporation. All rights reserved.

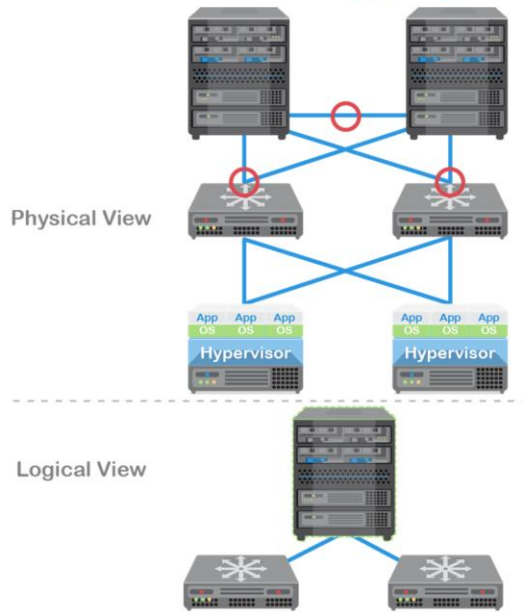
Cloud Infrastructure Planning and Design



7

The Spanning Tree Protocol (STP) is responsible for maintaining a loop-free topology in a bridged Ethernet network. Specifically, STP examines the topology of a network and ensures that there is only one path between any switch and the root bridge. STP prevents loops by blocking redundant links in an Ethernet network during normal operation. If an active link fails, blocked links can be enabled. This means that although you can connect switches together with multiple links for high availability, only one of those links will be used. This would most likely mean more switches will be added to a design to handle bandwidth requirements. Newer versions of STP enable the protocol to be enabled per VLAN. If multiple VLANs are in use in an environment, then traffic from some VLANs can traverse one physical link and traffic from other VLANs can traverse another physical link. This would help somewhat with bandwidth requirements but requires more planning.

## Multi-Chassis Link Aggregation



- Improves high availability
- Multiple active paths
- Simplifies network design
- No STP concerns
- Improves throughput

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



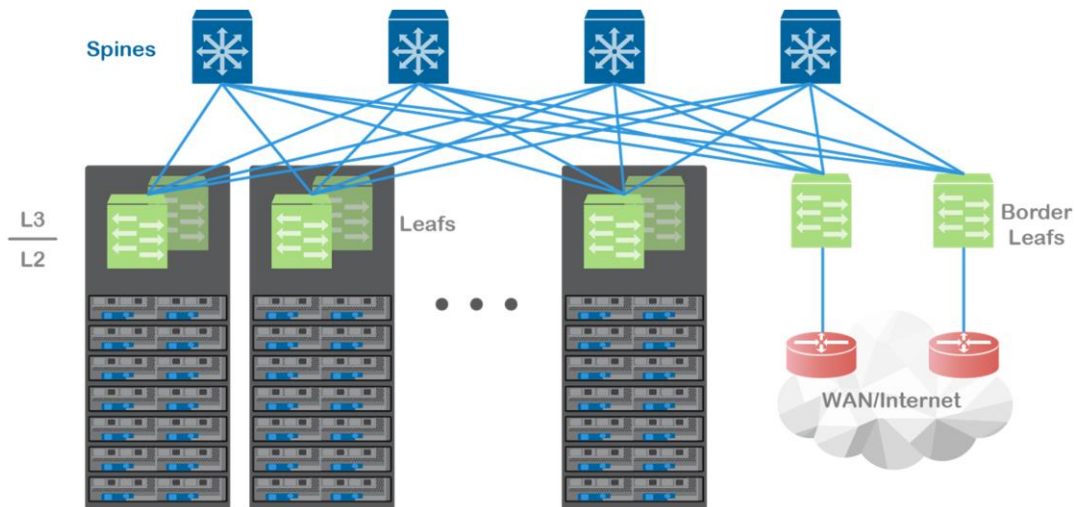
8

Link aggregation (LAG) is a method of combining parallel physical links on a switch into a single virtual link. This configuration allows for full use of the physical links and STP does not detect a loop condition. Depending upon the vendor solution, LAG is known by other names such as port channel, Etherchannel, link bonding or multi-link trunking.

Multi-chassis link aggregation (MLAG) makes a pair of physical switches appear as a single logical switch. While the implementation may differ among vendors (or even product lines from a single vendor), where some switches are still managed as two distinct entities and others are managed as a single unit, the ultimate goal is to provide a mechanism to improve availability and performance.

By aggregating physical switches together into a single logical switch, it allows you to design a network with multiple active paths through different switches. This not only prevents disruptions by avoiding convergence events, but it also provides alternate pathways under normal conditions to forward data, avoiding performance issues due to bottlenecks in the network. Cisco calls this technology Virtual Port Channels (vPC) for their Nexus switch line, and Virtual Switching System (VSS) for the Catalyst switch line. Brocade uses the term Multi-Chassis Trunking (MCT), and Juniper calls it Virtual Chassis. Spanning Tree Protocol can still be used in conjunction with MLAG to detect and handle any misconfigurations.

## Spine/Leaf Physical Networks – Example



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



9

An alternative example to the hierarchical design, shown here, is the Spine and Leaf design. This design consists of two layers and can be configured for a combined Layer 2/3 environment or Layer 2 only.

In a Layer 2 only configuration, each leaf switch is connected to two spine switches and both paths are active using Multi-chassis Link Aggregation (MLAG). Spine switches are also interconnected as MLAG peers. Leaf switches will also be connected together as MLAG peers to support connecting hosts to two leaf switches for redundancy and load balancing.

In a Layer 2/3 configuration, as shown in this diagram, the leaf switches use Layer 2 switching with hosts as described above but connect to the spine switches via Layer 3 protocols. This design uses a mesh, where leaf switches connect to all spine switches but spine switches are not connected to each other. The Layer 3 version of this design employs Equal-Cost Multi-Path routing (ECMP), which is a routing strategy in which multiple next-hop paths exist to a destination, and any one of those paths may be used to transfer packets on a per-flow basis.

Layer 2/3 configuration limits Layer 2 reachability which may impact certain infrastructure application or hypervisor functionality, such as virtual machine migration or clustering capabilities. Although an virtual overlay fabric may overcome this impact, it may be harder to implement for infrastructure components.

Both versions of the Spine and Leaf design uses multiple paths to distribute traffic for performance as well as maintain connectivity when a spine switch fails.

## Physical Network Considerations



- Switch port configuration
  - Will the host require access or trunk ports
  - The number of access ports may be limited on a host
- Design consistent network configurations across all hosts in a pool
- Avoid or minimize STP configuration issues in the design
- Using link and switch aggregation for better redundancy and improved bandwidth

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



10

Since hypervisors have multiple VMs connecting to the network over the same physical infrastructure, you need to determine how to configure the switch ports on the access layer, as well as the networking on the hypervisor. If all of the VMs on a particular hypervisor reside on a single VLAN, then you can configure the switch ports as access ports. Since an access port only supports a single VLAN, a physical connection from each host is needed for each required VLAN. This configuration is viable if the number of VLANs is less than the maximum number of NICs that your host can support and will be reduced even further with NIC HA.

If, however, you must support multiple VLANs, you need to use either multiple access ports or a trunk port. Trunk ports allow you to have a large number of VLANs configured on a single interface and provides a more flexible configuration for hosts. This approach works best with high bandwidth network connections.

When placing services in a large compute pool, all of the hosts' network configurations should be identical. This configuration enables the deployment of like services anywhere in the pool and the ability to migrate services as needed.

Avoid connecting access layer switches together since they will complicate the STP environment, and ultimately result in most, if not all, of the links being disabled by STP. If multiple connections exist between an access layer and aggregation layer switches, they maintain an Active/Standby configuration. If you want to utilize the full bandwidth, you need to manually configure one link as primary for certain VLANs, and the other link as primary for other VLANs.

By aggregating two physical switches into a virtual switch at the aggregation layer, you can now utilize all the interconnects between the access layer and aggregation layer. This allows you to increase the port density at the access layer, possibly reducing the number of switches that are required. However, unlike a non-virtualized switch environment, if an aggregation layer switch fails, the bandwidth will be reduced by 50% and you may experience degraded performance.

## Cloud Impact

- Increased east-west traffic
- Cloud-native application design
  - Database synchronization
  - Distributed services
  - Load balancers
  - Message queues
- Distributed storage

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



11

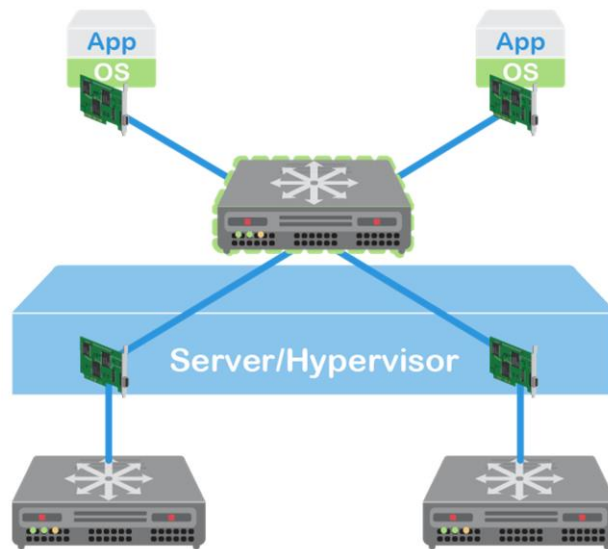
Network traffic patterns within the data center are changing because of the cloud. With cloud-native application design, services are distributed across physical nodes in order to avoid outages. This introduces more service traffic across the internal network in the forms of data access, database synchronization, message queue communication, and load balancing. This increases the east-west or inter-server traffic that was not seen in the days of monolithic applications where everything stayed on one server and availability came from the underlying infrastructure.

Another contributor to this traffic pattern change is the increased usage of distributed storage systems. Distributed storage uses the LAN infrastructure for both host to storage traffic and also inter-node traffic of the storage system. This traffic includes the writing of redundant copies of data across nodes, data synchronization, error checking, and node heartbeats. Distributed storage also adds to the east-west traffic within a data center. Cloud infrastructure designs must account for this changing pattern and have network components with enough bandwidth and low latency to support these newer traffic patterns.

A Spine and Leaf design may be better in a cloud environment where services are disputed across hosts since there will be fewer hops between services.



## Hypervisor Virtual Switches



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



12

A virtual switch is software that runs on a hypervisor and provides functionality similar to a Layer 2 physical switch. Virtual machines are provided with a virtual network adapter which is then connected to a virtual switch port. The virtual switch may also be connected to physical network adapters from the host which provide an uplink to the physical network infrastructure. Virtual switch technologies come from many vendors or communities and have many different features. Besides the normal packet forwarding, virtual switches can support VLAN tagging, uplink aggregation, port mirroring, and Quality of Service (QoS) capabilities. ESXi provides a proprietary virtual switch with its hypervisor but also supports others. KVM and other Linux-based hypervisors support internal virtual switch capabilities as well as open solutions such as Open vSwitch.

When the classic hosts are replaced with hypervisors, another layer of switching is added. The hypervisor virtual switches extend the access layer and perform Layer 2 switching within the hypervisor itself. So traffic traveling between two VMs that are on the same hypervisor and VLAN will have traffic switched locally. Traffic traveling between two VMs on different hypervisors or on different VLANs require data to be sent to the access or aggregation layer before being sent back to the destination VM.

Requirements guide the choice of the virtual switch selection within your hypervisor. Listed below are just some of the items that impact the design.

**CMP interoperability** – What virtual switches are supported by the chosen hypervisor, service catalog, orchestration engine or SDN controller? Are plug-ins or add-on required?

**Physical network compatibility** – Will the virtual switch support link aggregation or failover across a certain vendor's switches? What configurations are required?

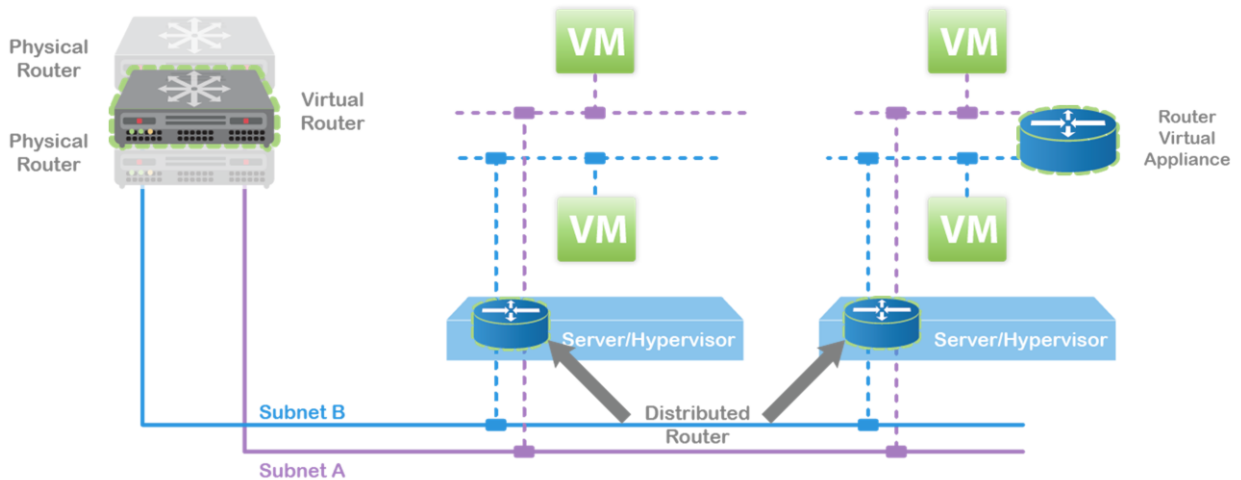
**Management** – How will the virtual switch be managed? Individually? Collectively? By whom?

**Monitoring** – Will the monitoring tool have insight into the virtual switch performance?

**Cost** – Is there an additional cost or license required to run the virtual switch? Is support available?

**VLANs** – Are VLANs required and will the virtual switch support them? What configuration is required at the virtual and physical layer?

## Routers



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



13

Routers are Layer 3 devices that are used to connect multiple Layer 2 networks or IP subnets. A router can be a physical device that has connections to all subnets. Having a single router on a network infrastructure creates a single point of failure. To alleviate this problem, multiple physical routers can be deployed and using specialized routing protocols, can act as one virtual router. In this case, hosts and virtual machines would have a default gateway that is set to the virtual router interface and not be bound to a specific physical device. This provides redundancy and increased throughput.

If physical routing devices are implemented in a virtualized or cloud environment, this means that all network traffic from a virtual machine that is destined for a different subnet must leave the physical host and traverse the network, even if the destination VM resides on the same host. This places increased network load on the infrastructure. To simplify things, router virtual appliances can be deployed within the cloud or virtual environment but this can also present problems since traffic still needs to leave the host if the appliance is not collocated with the communicating VMs. Since VMs can move between hosts, changing gateway assignments would make this solution problematic.

Appliances may also have additional functionality such as firewalls, VPN, and load balancing capabilities, which make them good solutions for north-south routing in the cloud. An appliance model supports multi-tenancy, where each tenant can have control of their own appliance.

With some network/hypervisor combinations such as VMware NSX, routing between subnets can be done without traffic going out to the physical router using processes in the hypervisor kernel. This is referred to as a distributed logical router. A distributed routing solution can help minimize network bandwidth consumption for east-west traffic and should decrease latency in a cloud. This solution will cause increased CPU and memory consumption on the host.



## Bandwidth



- Capacity of a network connection
- Calculating the proper bandwidth requirements include:
  - Expected service usage
  - Overhead
    - Protocols
    - Hypervisors
    - Storage
    - Management
  - Service placement
  - Inter-service communication

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



14

Bandwidth represents the capacity or the amount of packet data that can be transferred over a network connection. It is usually measured in bits per second. Calculating expected bandwidth to be used by services is not always an easy task. Bandwidth requirement calculations start with understanding the requirements for each service. Then you add to this the overhead for the transport protocols, hypervisor, storage, and other dependencies. Additionally, you need to factor in the expected placement of services and distribution across hosts to add any intra-host traffic. You also need to understand the requirements for services for network bandwidth between services and any source external to the cloud. Once the requirements are understood, bandwidth requirements must be applied throughout the design.

## Bandwidth Considerations



- Combine host ports to increase bandwidth
  - Services, management, storage, and so on
- Avoid oversubscription issues
  - Access layer to aggregation layer
  - Aggregation layer to core
- Use host affinity to minimize bandwidth
- Implement supported network management features to control bandwidth allocation

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



15

Hosts require enough bandwidth to support the planned services, hypervisors, storage, management, and other advanced services. High bandwidth connections, such as 10Gb links, are recommended when supporting converged network traffic, many VLANs, or a dense service allocation. When using multiple NICs for redundancy, plan for sufficient bandwidth if a path fails. Host traffic can be aggregated across multiple ports using teaming or bonding to increase bandwidth capabilities at the host. Understanding the supported configurations will require research.

At the access layer, you most likely will not have a 1:1 ratio of host ports to uplink ports, and in a hierarchical design you may not have a 1:1 ratio of access uplinks to core uplinks. Understanding the amount of traffic being sent and received by each host, as well as the network path of that traffic is important. Traffic that is generated from the hosts may be aggregated into a small number of switch uplinks and could lead to oversubscription and bottlenecks. For instance, if a switch supports sixteen 10Gb host connections and four 10Gb uplinks connections, you will have a 4:1 oversubscription ( $16 \times 10 / 4 \times 10$ ). This may be acceptable if the planned workload from these 16 hosts does not exceed 40Gb, but this cannot be determined from the requirements gathering process. The design should also account for switch failure since all traffic will be diverted to the remaining paths and this could have a negative affect on performance until the switch is replaced.

Virtual machines located on the same host may communicate without the traffic ever leaving the host. If related systems (for example application and database server) are strategically placed, and a well-defined process is in place to maintain that placement, a significant amount of traffic may be removed from the physical switch infrastructure.

Some hypervisors have mechanisms to control or limit bandwidth for specific types of traffic. Although this is not a guarantee that there is enough bandwidth for all traffic, it does provide guarantees for traffic that may be considered higher priority. Setting a bandwidth priority for storage traffic can ensure that storage performance does not suffer. Research is required to determine which hypervisors and physical infrastructure supports this, and what the available options and limitations are.

## Availability

- NIC teaming
- Host connection to multiple switches
- Redundant components
- Identify fault domains



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



16

NIC Teaming or bonding not only supports increased bandwidth through aggregation but also can provide redundancy for host connections to the network. The cloud architect evaluates the different options available at the hypervisor and the support through the network stack and then documents the best solution. High availability decisions will include whether to run in active/active configurations, how path failures or upstream failures will be handled, and whether or not failover is handled at the hardware or software layer.

When supported, NIC teams should be spread across multiple switches to guard against a single point of failure in switch infrastructure. As mentioned previously, remember to consider the impact on available bandwidth if a path fails.

Avoid any and all single points of failure.

If the organization deploys true cloud-native services which are designed for failure, then an alternative method of design is to create fault domains and distribute redundant services across the fault domains. In a single data center model as an example, this may mean that each rack has its own leaf switch and if anything causes a network outage in one rack, the redundant service remains running in a different rack. Even though services can be distributed across a fault domain for redundancy, having an entire domain go down may still be taxing on the business. A hybrid approach of building some redundancy into a fault domain, such as deploying leaf switches in pairs within a rack, may still be beneficial.

## Load Balancers

- Purpose-built devices that distribute processing load across several nodes
- Physical appliance
  - Dedicated hardware and software
  - Limited scalability
- Virtual appliance
  - Virtual machine and software
  - Scalable
  - Better multi-tenancy support



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

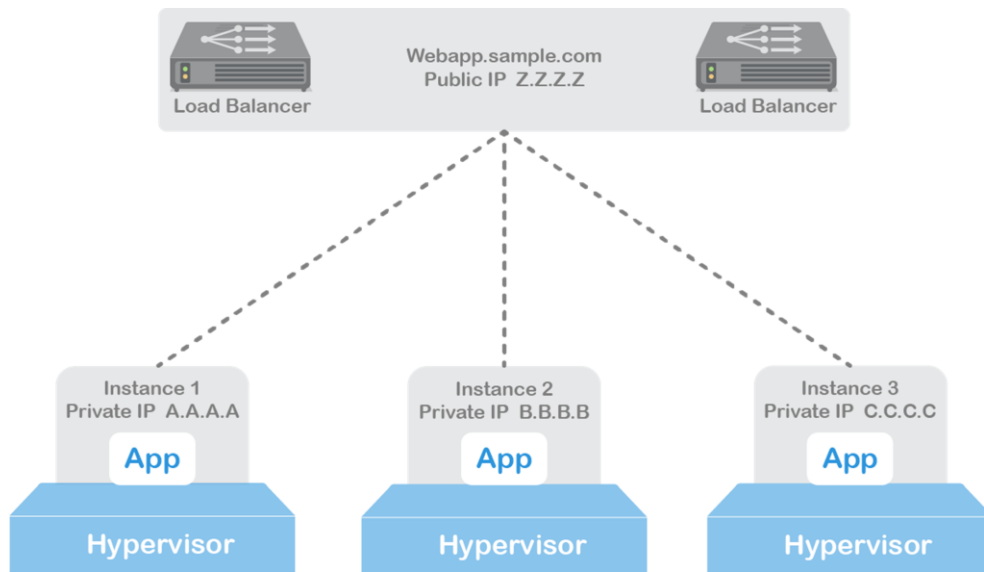


17

Load balancers are purpose-built devices (physical or virtual) that distribute processing load across several nodes. The physical appliance is a standalone unit or switch module that is dedicated to performing load balancing functions. These standalone units can be custom-built devices or software that is installed on a general purpose server. While these solutions are generally more expensive as they require both hardware and software, they do not have to contend with other service resources and can be scaled to the limitations of the hardware. These devices have known performance capabilities and limitations, and often come in different sizes (custom appliance) or have hardware requirements for specific configurations (general-purpose server). To scale these devices, you can add more units, deploy a larger unit, or in some cases, bond multiple units together.

Virtual load balancers are instantiated within a hypervisor environment. While this option is typically less expensive than its physical counterpart, it is limited to the resources of the hypervisor and is subject to resource contention with other VMs residing on the same hypervisor. Scaling for virtual load balancers is generally done by deploying multiple appliances across multiple hypervisors. However, if the environment is large, this can become complex to manage. Cloud environments enable the use of network segments. If a load balancer is required within a segment, for example a tenant network, virtual load balancers may be the only solution to support this.

## Example: Load Balancing



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



18

In this example, there are three application servers, each with a private IP address. A pair of load balancers is placed in the network before the application servers, and provides a publicly accessible IP address and DNS name. When users access the application located at `webapp.sample.com`, they are directed to one of load balancers, and are then redirected to one of the application servers.

## Network Pools

- Predefined groups of network segments
  - Not always pre-allocated
- Consumable
- Various types
  - VLAN-backed
  - Overlay technologies
- Network isolation requirements
  - Firewall
  - NAT

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

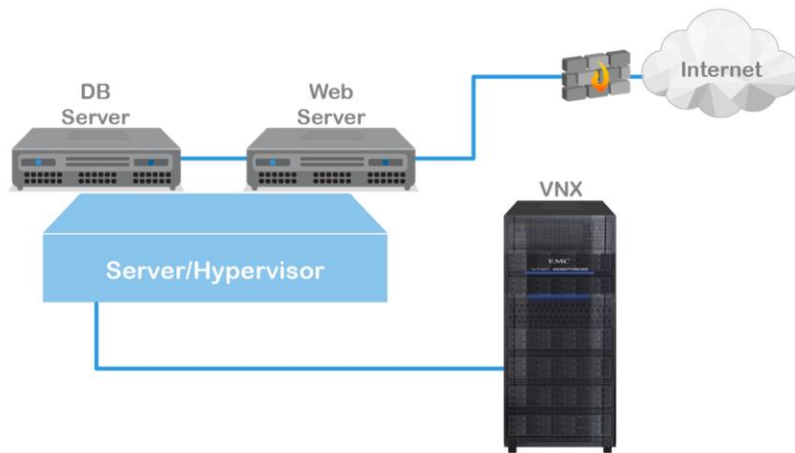


19

Virtual Data Centers use network segmentation at the virtual and physical layers to isolate some workloads. Multi-tenancy, security, and individual service requirements of cloud environments increase the need to create network segments. Cloud environments provide network pools to tenants and consumers. Network pools are comprised of network segments used to facilitate communications between VMs. However, unlike compute resource pools, network pools may not contain preexisting networks that can just be attached to services. Some network solutions provide the network segments as needed, and in this case the pool is really just definitions and rules for how the segments are implemented and used. Segmentation is provided by the underlying technologies, such as VLANs or network overlay protocols. A pool of VLANs may be preconfigured and ready to use, but overlay networks may be predefined but not implemented until the service is implemented. Since the segments are isolated, they have their own address space and can be interconnected using virtual firewalls and NAT-capable technologies. All of these technologies together allow for clouds to provide network isolation and still maintain a level of secure connectivity with the outside environment.

## Segmentation

- Physical
- VLANs
- Overlays



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



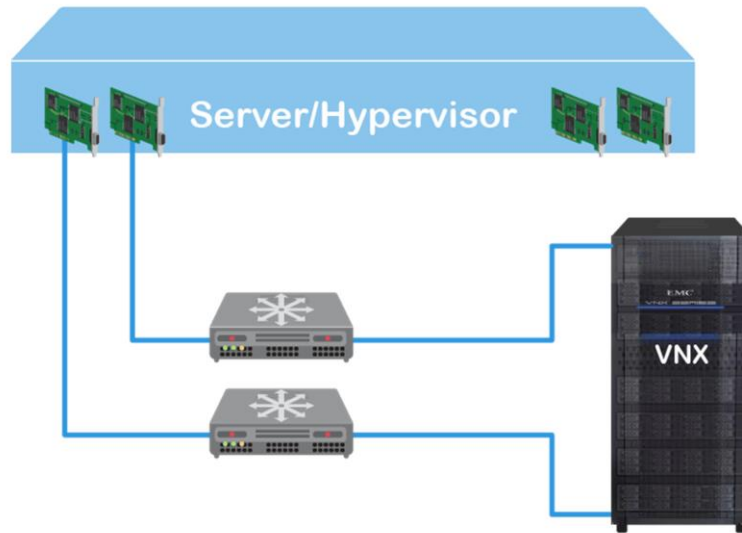
20

Network segmentation is the act of isolating networks from each other and is an important concept in cloud computing. Cloud environments are made up of tenants, and in many cases those tenants are from different organizations. Unless a tenant only maintains an Internet-facing service, that tenant expects that its network segments will be separated from other tenants. Segmentation is also used to isolate management traffic from tenant traffic. Network segmentation is a form of security and is part of trust zone definition in a cloud design.

Segmentation can also be used to improve performance by reducing contention and latency. By separating different types of traffic into different physical networks, you can reduce the number of nodes trying to access those networks and reduce the number of protocols being used. However, if these networks are logically separated, such as by using VLANs, then further planning is required to ensure that sufficient bandwidth exists to support all traffic.

Network segmentation can be implemented through complete physical or logical isolation of networks or network segments may be interconnected using firewalls or routers. The method of segmentation is driven by the requirements. Complete isolation may be required for the network that is used to connect hosts to a iSCSI storage arrays and may be implemented through a isolated physical network infrastructure. In a cloud environment, consumers may wish to use isolated network segments for some backend database services but may use other network segments that have access to the Internet through a firewall to host frontend web services.

## Physical Isolation



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



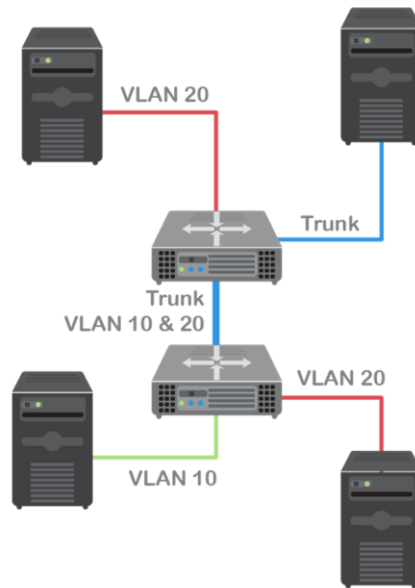
21

In a cloud infrastructure design, physical network isolation may not be practical except in the case of separating infrastructure traffic, such as storage, from the consumer traffic. Requirements may dictate physical isolation for enhanced security or performance. Physical separation may not mean a complete set of different physical infrastructure. In the case of storage, for instance, it may just mean dedicating host network adapters and storage array ports to dedicated ports on network switches. This may be sufficient as long as the network switches have the capacity to support all of the traffic in the environment. However, in some cases, requirements may dictate complete physical isolation with dedicated network switches for storage traffic. In other cases, such as converged infrastructure environments, physical isolation may be impossible and logical separation may be necessary.



## Virtual Local Area Network (VLAN)

- A VLAN has all of the attributes of a physical LAN
- Broadcasts are limited to the VLAN
- A server can be connected to multiple VLANs
- VLAN tagging enables traffic for multiple VLANs to pass through a trunk link



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



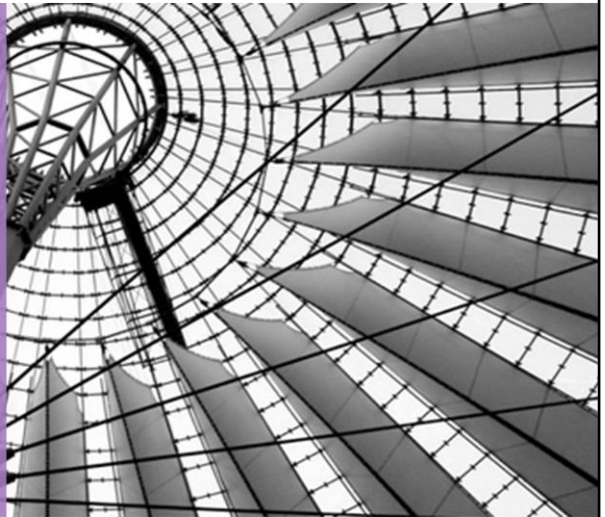
22

A virtual LAN, or VLAN, is a Layer 2 network segment isolation technique defined in the IEEE 802.1Q standard. A VLAN is a group of devices with a common set of network requirements, which communicates as if it were in the same broadcast domain, regardless of physical location. Essentially, a VLAN has all the same attributes of a physical LAN (broadcast domain, security, and so on), but is not restricted to a physical location or switch. By using VLANs, it is possible to distribute devices across switches, improving fault tolerance and reducing hardware costs.

A server can be connected to multiple VLANs, either by having multiple physical NICs which are connected to a switch port configured for the appropriate VLAN, or by having a single NIC that uses VLAN tags and trunking. A VLAN tag is an identifier that is inserted into the Ethernet frame header that identifies which VLAN the traffic is associated with. When a switch port is configured as a trunk port, it allows traffic from multiple VLANs to cross that port. Ports that can only support a single VLAN are known as access ports.

## VLAN Considerations

- Configured throughout the infrastructure
- Limited to 4094 networks
- Easily added to an existing trunk port
- Useful on blade servers



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



23

In order to use VLANs, they must be configured and supported throughout the network infrastructure. This means that in a virtual infrastructure connected by a physical network for example, all physical switches, hypervisors, virtual switches, and even virtual machine NICs must be configured properly to use VLANs. This also means that once VLANs are configured, adding new VLANs to the environment requires configuration changes throughout the stack. Although this could be automated, it becomes very hard to support and will not be a fast process in a large infrastructure.

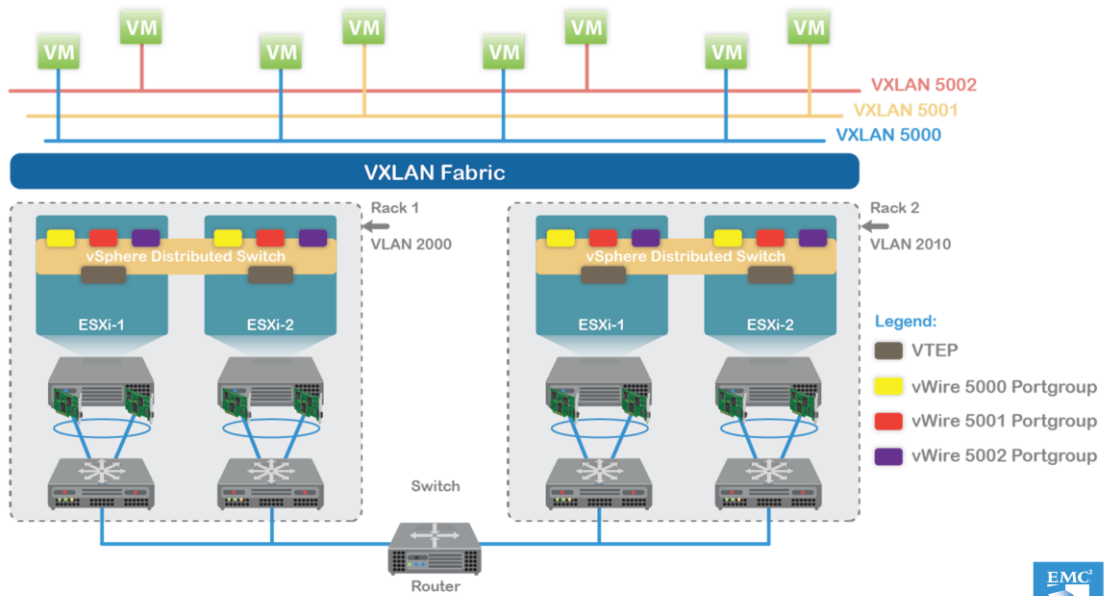
Unless you are willing to reconfigure a virtual machine instance, mobility of that instance is limited to hosts that have access to the same VLANs.

VLAN identifiers are implemented by adding a 4-byte tag field in the original Ethernet frame but the actual VLAN address portion is 12 bits. Because of the limited number of address bits defined in the 802.1Q standard, VLANs can only support up to 4094 virtual networks. In a cloud where many network segments will be deployed for services or tenants, this could be a limiting factor.

Although VLANs have some limitations, they are easier and faster to deploy than additional physical network connections. If bandwidth is available, such as when using 10G network connections, then using trunk ports with VLANs to separate infrastructure and management traffic may help to reduce hardware costs but still maintain security.

In converged infrastructure or blade server enclosures, where network interface cards can be limited in number, VLANs are useful for separating network traffic types.

## Overlay



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



24

Overlay networks use tunneling and encapsulation techniques to build Layer 2 virtual networks that are decoupled from physical infrastructure. Overlay network packets contain an encapsulated header that identifies a unique virtual network identifier. Encapsulation is performed at the network edge, which could be a physical or virtual switch, and the packet is then transferred using standard Layer 2 and Layer 3 network protocols. Common examples of overlay network protocols are VXLAN, NVGRE, STT and Geneve. Geneve (Generic Network Virtualization Encapsulation) is a proposed standard meant to merge features from all of the popular protocols.

An advantage to using an overlay solution is that it can be implemented and scaled over existing networks with little to no changes required in most of the underlying infrastructure. Another advantage is that overlay network technologies can support over 16 million virtual networks. Overlay networks support rapid deployment via software which means that service orchestration technologies can deploy these virtual networks on-demand.

## Overlay Considerations



- MTU Size
- Hypervisor support
  - Identify configuration limitations
- Additional overhead
- Ensure sufficient bandwidth
  - Monitoring tools

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



25

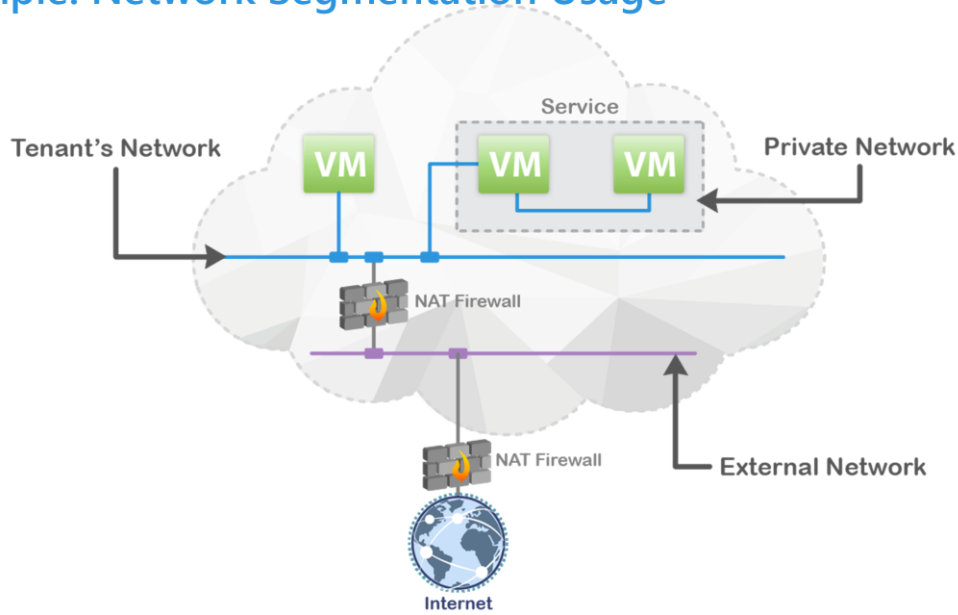
An overlay technology encapsulates the original Ethernet frame that is generated by a workload and a VXLAN header into a new Ethernet frame. The overall size of the frame will be bigger than a standard Ethernet frame and MTU size must be adjusted accordingly. In a VMware NSX environment using VXLAN, it is recommended to set the maximum MTU size to at least 1600 bytes. This MTU size must be documented on the design and set throughout the network infrastructure.

Alternatively, some hypervisor environments may handle the MTU sizing in a different manner, and you will need to include this information in the design documentation. For example, in certain circumstances with Microsoft Hyper-V and NVGRE, the MTU size of a vmNIC (virtual adapter of the virtual machine) gets lowered automatically during the initialization of the vmNIC.

When choosing an overlay network protocol, ensure that the protocol selected is supported by the underlying hypervisor and cloud management platform. Also ensure that cloud design has enough compute overhead to handle the encapsulation process on the hypervisor. Research the supported configuration settings as well since some may not be supported for overlay protocols. As an example, although VMware NSX supports NIC teaming, not all configurations are supported.

As the number of virtual networks increases, so too will the traffic. The underlying physical network must be designed in a way that it can handle the increased usage over time and should be easily upgradeable. Managing overlay networks could become challenging since the control for deployment is passed to the consumer. The cloud design needs to include monitoring tools that provide insight into performance, as well as the relationship between virtual and physical networks.

## Example: Network Segmentation Usage



© Copyright 2015 EMC Corporation. All rights reserved.

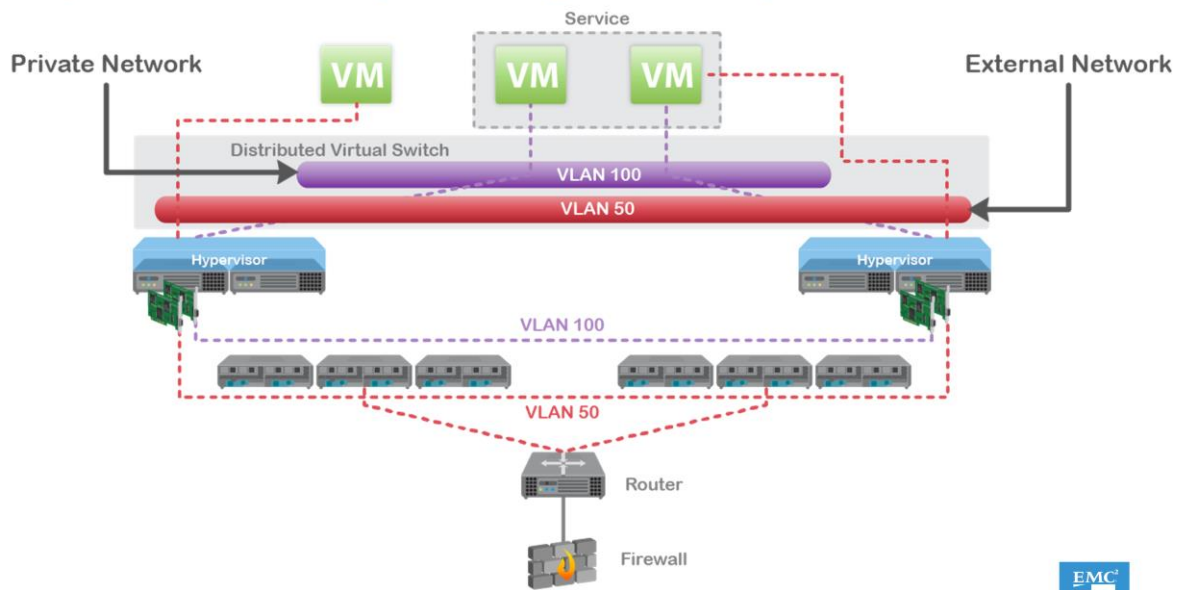
Cloud Infrastructure Planning and Design



26

This diagram shows three examples of different uses for network segments that can be deployed in a cloud environment. Private networks are assigned to individual tenants and used for isolated VM to VM communications. They are made up of virtual Layer 2 network segments in combination with a non-routable network address range. A tenant network is similar to a private network, but it is connected externally through a NAT capable firewall device. External networks are similar to the tenant network except that the NAT capable firewall is connected to a public-facing network and is shared by multiple tenants.

## Example: Network Segment Physical Design



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

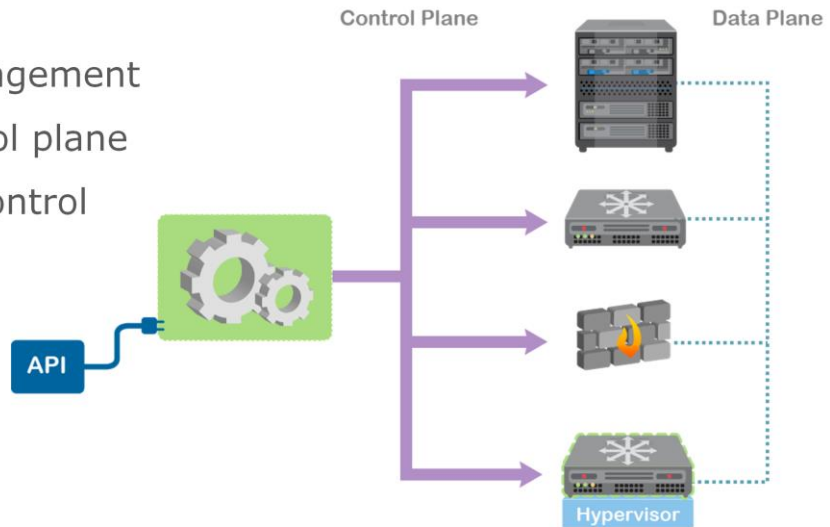


27

This diagram shows two of the example network segments viewed from a physical design perspective. Note how the different types of network pools are mapped to a specific VLAN which are made available throughout the virtual and physical infrastructure.

# Software-Defined Networks

- Centralized management
- Abstracted control plane
- Programmable control
- Agile



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



28

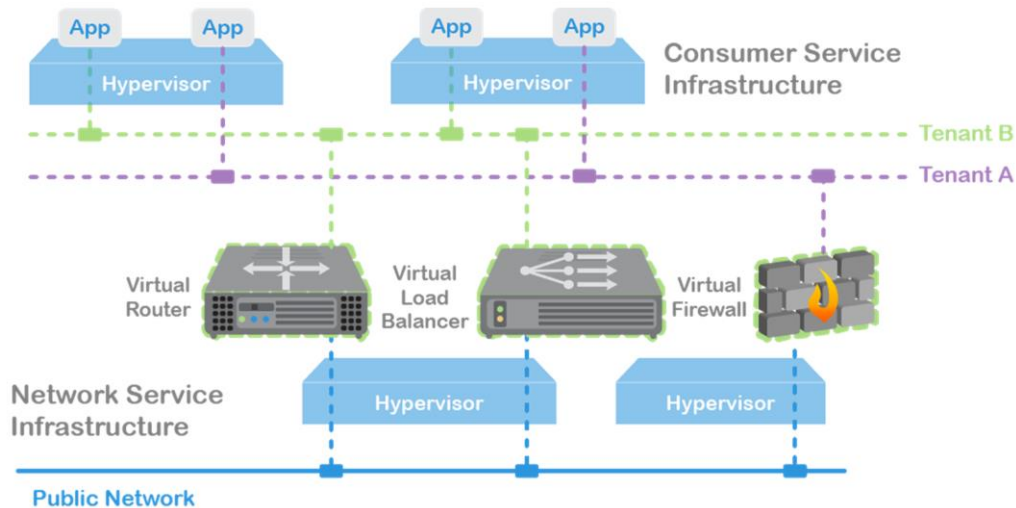
Software-Defined Networking (SDN) is an architecture that decouples the network control and data, enabling the network control to become directly programmable, and the underlying infrastructure to be abstracted for applications and network services. A SDN controller centralizes the network intelligence, provides a global view of the network, and provides an interface for people, applications, or orchestration engines to manage and deploy resources. An organization can programmatically alter network behavior in real-time and deploy new applications and network services faster. In addition to abstracting the network, SDN architectures support deploying network services, including routing, multicast, security, access control, bandwidth management, traffic engineering, quality of service, and so on.

When implementing a software-defined network solution, some key considerations are listed here:

- The controller most likely will be deployed in the infrastructure supporting the cloud management components. If the SDN solution requires direct access to manipulate or control physical or virtual switches, then the design needs to include access through an isolated or secured network.
- The controller may also need access to the hypervisor or hypervisor manager to manipulate virtual switch configuration or deploy virtual network resources such as firewalls.
- The controller API may require access from the cloud management platform to support deploying and controlling cloud services. The API must also be accessible to applications running within the cloud. An example of this would be if an application needs to contact the controller to set a QoS policy. With a true software-defined network, this capability could exist but may be very hard to deploy in a multi-tenant environment.



## Network Service Infrastructure



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



29

Designing a separate pool of infrastructure to support network functions and services may have benefits as described below.

- Reduced time to locate and troubleshoot network services since they are located in one area.
- Dedicated compute resources for running network services which reduces load on consumer infrastructure
- Eliminates need for physical, public-facing network connections to the consumer infrastructure
- Allows for a separate authentication mechanism and security policies to be applied to the network service infrastructure, creating an additional layer of security

In the separate network service infrastructure, network connections (physical or virtual) may be required on the hypervisor host to support the additional networks.



## Quality of Service

- Various traffic types have different network requirements
  - Tenant, management, and storage
- Alternative to physical network separation
- Ensure preferential delivery service
- Assign bandwidth, control latency, reduce data loss
- Must be supported throughout the network

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



30

Cloud environments carry various types of network traffic such as tenant, storage, and management. Each traffic type has different characteristics and requirements on the physical switching infrastructure. For instance, management traffic typically is low in volume but is critical for controlling physical and virtual infrastructure. Storage traffic is typically high in volume and is also critical for maintaining service availability and functionality. The cloud may also offering various levels of service for tenants.

The cloud design can include various options to support these requirements such as dedicated networks and high bandwidth components. But this may not be enough to support al of the requirements for performance and availability. Implementing Quality of Service (QoS) capabilities can help to ensure certain types of traffic get the required priority and bandwidth.

QoS is a method that provides the ability to set network characteristics to ensure preferential delivery service for certain applications. Administrators use it to assign bandwidth, control latency and jitter, and reduce data loss. Providing high priority network transmission for individual tenants may not be feasible, but using QoS to ensure management and storage traffic has guaranteed performance levels which will help to ensure cloud availability.

An important concept to remember for the cloud design is that QoS capabilities must be supported throughout the network. All network elements, such as network interface cards, physical and virtual switches, and routers must support QoS.

## Performance Considerations



- TCP Offload
  - Requires supported components
  - May not always be beneficial
- Jumbo Frames
  - Increased Ethernet frame size
  - Must be configured throughout the local network
- SSL offload
  - Separate encryption processes from services

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



31

TCP offload engine (TOE) is a technology used in network interface cards (NIC) to offload TCP/IP processing from the computer CPU to the NIC. It reduces the CPU overhead allowing more resources to be available for services. Various pieces of the TCP/IP stack can be offloaded such as checksums, segmentation, and large receive. If you are considering using TCP offload capabilities, the design must include appropriate network adapters, supported operating systems, supported network drivers, and supported hypervisors. It is also important to understand that although it may be beneficial to enable TOE, in some cases, performance of virtual machines may be negatively impacted and therefore further research will be required for the design.

Jumbo frame support is when the underlying network is capable of transmitting Ethernet frames that are larger than the standard size of 1500 bytes. Jumbo frames improve performance because more data is transmitted per frame and fewer frames can be used to transmit large amounts of data. Fewer frames means less overhead. Enabling jumbo frames is very beneficial for environments with large file transfers and those that support IP-based storage solutions. In a cloud design, jumbo frames must be enabled and the size must be properly configured throughout the entire infrastructure. If any component within the data path is improperly configured, performance and packet loss issues may result.

Secure Sockets Layer (SSL) is a standard security technology for establishing an encrypted link between two points on a network. The process of encryption can add overhead to servers, and performance can be impacted significantly. SSL offload moves the SSL encryption and decryption process off of the primary service, for example a web server, and moves it to a separate device such as a firewall or load balancer. Deploying specialized SSL endpoint devices can remove overhead from applications and spread the load across the infrastructure.

## Security

- Network isolation
- Firewalls
- Traffic encryption
- VPN
- IDS/IPS



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



32

As discussed previously, network segmentation is used prevalently in a cloud design. Restricting traffic to specific network segments and minimizing or even blocking access to those segments is an effective mechanism for securing network traffic.

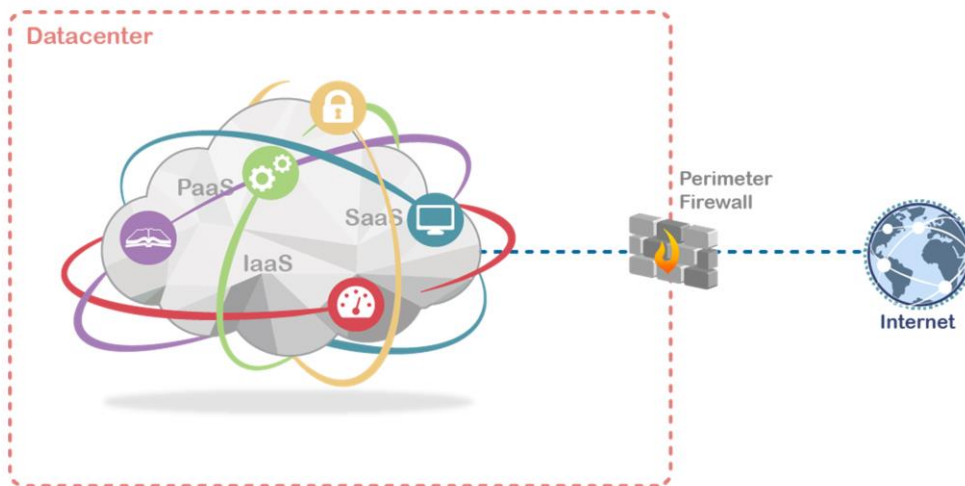
Firewalls are an effective technology for controlling traffic to and from a network. Modern firewall technologies can detect various security threats through all layers of the network stack.

Traffic encryption is a useful method for obscuring the data contained within network packets. This method of security is commonly used for north-south bound traffic where data is exposed to other parties on publicly-available networks.

Virtual Private Networks are a method used to secure traffic to and from a particular network segment.

Intrusion Detection and Intrusion Prevention Services are technologies that seek out malicious activity on the network and either alert or stop the threat.

## Perimeter Firewalls



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



33

In a traditional datacenter, physical firewalls are high-speed, optimized devices that are designed to perform packet inspection.

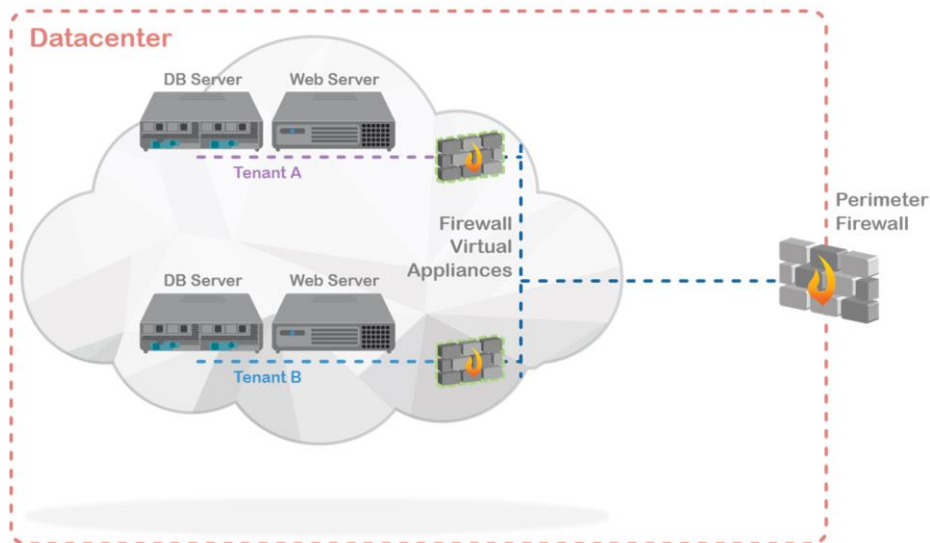
When designing a cloud in an existing datacenter, you will most likely place a physical firewall at the outermost perimeter between the Internet and the public-facing access points of the cloud. One function of this firewall is to protect the cloud management platform. Certain components of the CMP, such as the portal, require access from the outside and it will take the form of web client and API calls mostly using HTTP or HTTPs protocols. The perimeter firewall must guard these access points as well as block access to any other ports on CMP servers.

Consumers will also need access into their services and tenant networks. If the services are Internet-facing and use standard protocols such as HTTP, then the perimeter firewall will allow this traffic to pass. If services are located on protected or isolated tenant segments, even though this traffic most likely will be encrypted and may terminate at another virtual firewall or VPN endpoint, the perimeter firewall will need to be configured to enable this access.

Traditional firewalls protect data through Layer 4 of the OSI model and can filter using combinations of source or destination MAC address, IP address, or TCP or UDP port. Newer, next generation firewalls, operate through Layer 7 of the OSI model and can inspect and apply rules based on the data content of the packet. Including a next generation firewall into the design satisfies any requirements for advanced functions such as stateful inspection, application awareness, blocking denial-of-service attacks, or intrusion detection or prevention.

The perimeter firewall must handle all traffic in and out of the cloud infrastructure, so it must be sized to support the bandwidth and latency requirements determined during the assessment. Since it will also be a single point of failure, consider designing a redundant solution that can handle traffic during a firewall failure.

## Virtual Firewalls



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



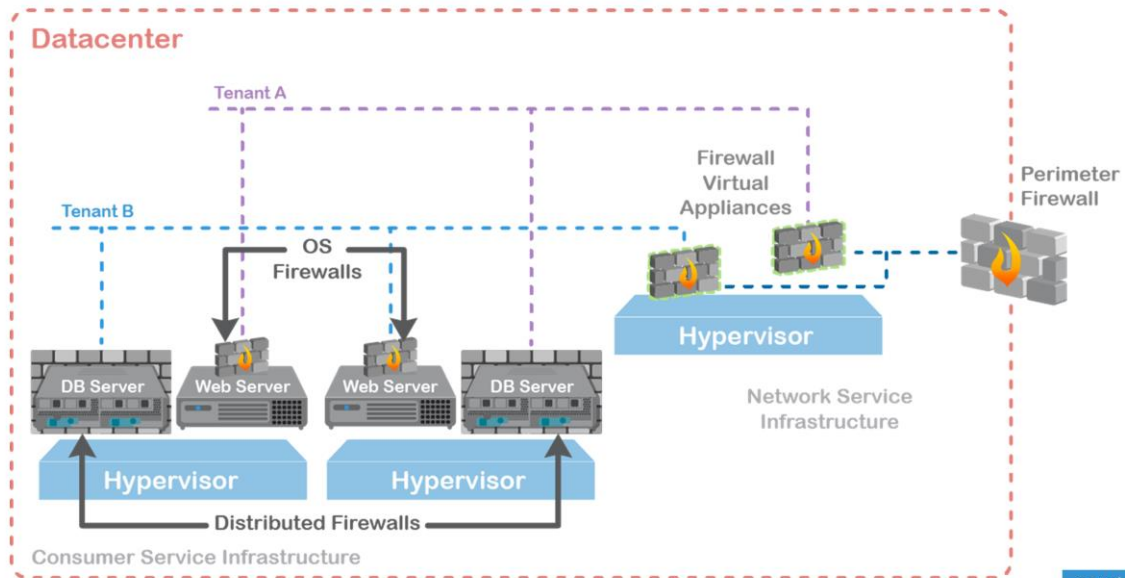
34

The networking environment in a cloud does not lend itself to using a physical firewall. Because VMs are mobile, and can be located on different physical hypervisors at any time, designing a physical network infrastructure that forces inter-VM communications to pass through a physical firewall is almost impossible. In addition, since VMs located on the same hypervisor can have their communication switched within the hypervisor itself, physical firewalls are unable to filter this traffic. In a private cloud scenario, virtual machine instances may be accessible from the organization's internal or corporate network and may require dedicated firewall rules and services. Using a single (or redundant) perimeter firewall will not support these requirements, nor will deploying multiple physical firewalls within the infrastructure. Including virtual firewalls in the cloud design adds a second layer of protection and better supports the scenarios described above.

Virtual firewalls can be integrated with the hypervisor platform, or a 3<sup>rd</sup> party offering, but essentially are instantiated into the virtual environment, and perform the same functionality as a physical firewall. By establishing rules within the virtual environment, VMs are placed into "trust zones," which are boundaries of access. VMs within a trust zone can communicate freely with one another, but VMs outside of the trust zone are not allowed to interact. Deploying virtual firewalls directly in front of services or tenant networks can provide security for both of these scenarios. This also provides consumers or tenants with the ability to deploy firewalls as needed and customize them.

Selecting the proper virtual firewall solution depends on support for the cloud management platform and underlying hypervisor.

## Other Firewall Solutions



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



35

### Distributed Firewalls

Some vendors have distributed firewall solutions. An example of this is VMware NSX, which provides a central management capability for defining rules but is enforced by the ESXi hypervisor. Firewall policies are based on the virtual machine or its membership to a policy set and not its IP address. This provides the virtual machine with the proper protection even if it is moved between hosts without the need to move a virtual firewall or maintain a proper network connection.

### OS Firewalls

Operating systems have firewall services as well. These should be used as an extra layer of protection especially on infrastructure components. Maintaining the various OS firewalls can be challenging, and your design should not only include documentation on the rule sets to be used for these firewalls but also a configuration management tool that can set and maintain these rules.

### Integrated Solutions

Some vendors provide integrated platforms that offer a single pain of glass that can manage perimeter and virtual firewalls as well as other security services. These solutions may even be integrated with public cloud providers and can be used to manage firewalls throughout a hybrid cloud environment.



# Encryption



- Method for securing data in transit
- SSL is standard protocol for network encryption
  - Uses keys which must be protected
- Additional overhead
  - Requires additional CPU and network resources
  - Encryption processes can be implemented on network devices to offload overhead on the application

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



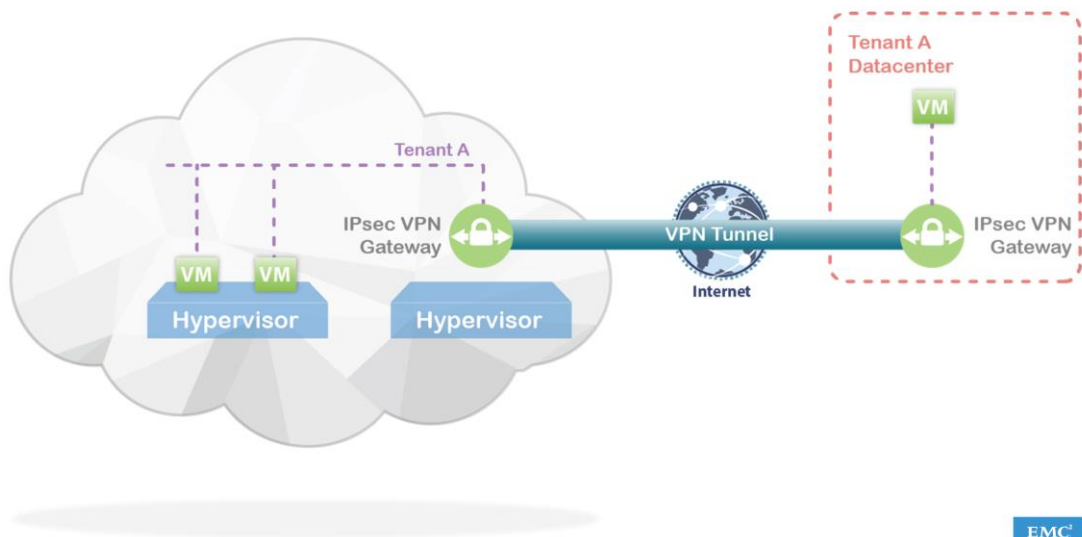
36

Network encryption is a method to mask transmitted data using a standard algorithm, protocols, and keys. SSL (Secure Socket Layer) encryption is a standard technology used to secure data in transit. In the cloud, a user can communicate with a web service, for example, using SSL encryption and the standard HTTPS protocol. As depicted in this simplified example, this communication requires that each side have the appropriate keys and follow the exact protocols. In this example, as in typical Internet use cases, secure sessions must be created for every service that the users wishes to communicate with.

From a cloud design perspective, the underlying network must have the additional resources, such as additional CPU capacity, that are required to encrypt and decrypt data. Also, SSL connections use specific TCP ports, and these ports must be known and traffic must be allowed to pass on these ports in various places within the network infrastructure. This information must be collected and documented in the design for future audit, operations, and support reasons.

One method of reducing, or at least moving some of the overhead involved with the encryption or decryption process, is to implement SSL termination on an edge device such as a firewall or load balancer. If this type of solution meets requirements, it is important to realize that since the additional resource demand will be on these edge devices, additional CPU may be required on the separate network service infrastructure rather than the consumer resource infrastructure.

## IPsec VPN



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



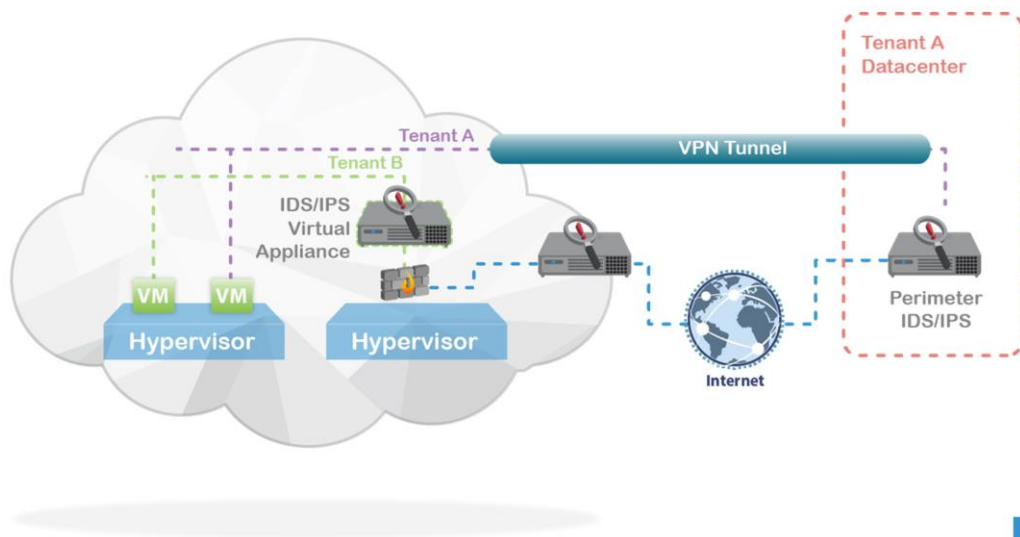
37

An IPsec Virtual Private Network (VPN) is a networking technology that allows you to connect two networks via an Internet connection. With VPN tunneling technology, one endpoint device is placed on each network segment and a secure tunnel is created between them. This tunnel can connect the two segments using Layer 2 protocols. The tunnel uses encryption to secure traffic, but unlike a client server connection, the session is maintained and clients or servers on either end of the tunnel have access to the entire network segment on the other end.

If requirements dictate that tenants require VPN connections into the cloud, then the cloud will require certain items to support this. VPN gateways will need to be implemented and configured. Endpoint firewalls may support VPN termination and must be configured appropriately. Ports may need to be opened on any firewalls that pass the VPN traffic. Since encryption is still required, sufficient resources must be included in the design to support this.



# Intrusion Prevention/Detection



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



38

Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) solutions are critical for any environments that are connected to the Internet. IDS/IPS solutions come in different types. The first is a device that sits on the network and watches for malicious activity. The second type is software that resides on a host or virtual machine and watches for malicious activity. Here we discuss network devices.

While the terms are often used interchangeably, Intrusion Prevention and Intrusion Detection are different functions.

A network IDS is a passive monitoring system. Its role is to monitor networks for any suspicious activities and generate an alert when it discovers an abnormality. Often, an IDS uses a network replication function, such as SPAN, to examine a mirror of the data that is being transmitted.

A network IPS is designed to stop an attack from reaching an internal network or device. This is done by placing one or more devices at entry points into the network, such as Internet or WAN connections. These devices scan inbound packets looking for suspicious contents. If a packet arrives that appears suspect, the device will drop the packet.

Protecting cloud services is still possible with a network-based IDS/IPS solution. For one thing, some firewall solutions have IPS/IDS functionality, so these would be implemented as discussed in the previous firewall discussion. There are IPS/IDS solutions that are packaged as a virtual appliance, and with special configuration, they can be placed on the appropriate network segments and perform their role.

Another option is to extend a tenant's private network into a cloud environment using something like a VPN solution rather than connecting it through an edge device in the cloud. If the private network has an IDS/IPS solution deployed and a VPN linking the internal and cloud network, then all traffic from the Internet will flow through the IPS/IDS before reaching the cloud segment.

# Lesson 1

## Summary

Key points covered in this lesson:

- LAN infrastructure must be able to support more east-west traffic in a cloud.
- Requirements for virtualized network devices such as routers, load balancers and firewalls will influence network design.
- Network security is critical for protecting infrastructure components and enforcing multi-tenancy requirements.



This lesson covered the requirements and design consideration for implementing local area networks in a cloud.

## Lesson 2

### Storage Networks

This lesson covers the following topics:

- Requirements for implementing storage networking technologies in a cloud
- Design considerations for implementing storage networks in a cloud environment

This lesson covers the requirements and design consideration for implementing storage networks in a cloud.

## Understand Storage Network Requirements



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



41

The storage networking component of cloud infrastructure is driven by requirements. Storage network requirements fall into categories such as those listed here.

**Protocols** – Various storage protocols are used in the cloud and the design identifies the usage areas for these protocols and configuration requirements.

**Connectivity** – The infrastructure design must support access to storage from the hosts that will provide storage, such as a boot drive or block device, to services. Other hosts may provide file or object storage to services across the network infrastructure.

**Bandwidth** – The design must include sufficient network capacity to support storage traffic. It must also be expandable for future needs.

**Latency** – To maintain a low network latency, the infrastructure design may include Quality of Service policies, fewer hop counts, and dedicated infrastructure supporting network function components such as load balancers and firewalls.

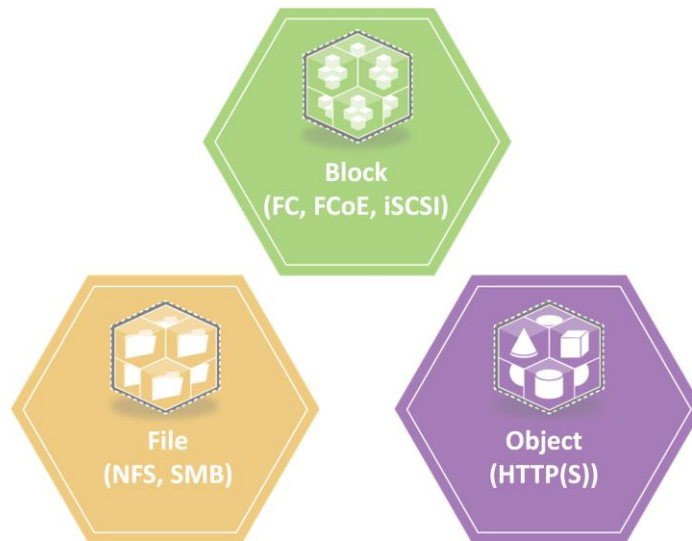
**Availability** – Storage network infrastructure requires redundancy to support availability.

**Cost** – Organizational cost requirements will guide decision points, and you will need to balance cost, performance, and capabilities throughout your design.

**Security** – Security is managed and controlled at many levels. Storage networking infrastructure may require network isolation and encryption techniques to secure traffic.

As we discuss different technologies and design considerations throughout this lesson, we will see how these technologies help to address the different requirements.

## Protocols



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



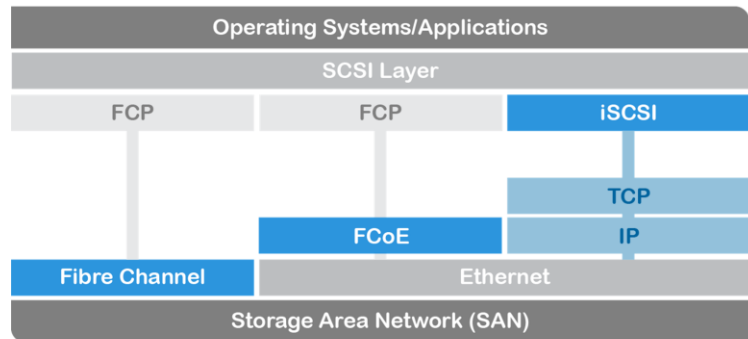
42

When designing a storage environment for the cloud, it's critical to understand how it will impact your network design. The cloud design needs to reflect the organization's storage requirements. One step in the design process is to determine which storage types (block, file, or object) will be used in the environment. Each of these types has its own set of protocols that are used to access the storage across the network. Block storage can be accessed using the Fibre Channel (FC), Fibre Channel over Ethernet (FCoE), and Internet Small Computer System Interface (iSCSI) protocols. File storage systems can be accessed through the Network File System (NFS) and Server Message Block (SMB) network protocols. Object storage is accessed through a RESTful API using the (Secure) Hypertext Transfer Protocol (HTTP(S)). Each protocol has an impact on the design. For example, if using the Fibre Channel (FC) protocol, then storage will be accessed over a dedicated and isolated network which must be accessible by all hosts that will host consumer services.

# Block Storage Protocols



- FC
  - Separate from LAN
  - Traffic isolation
  - High performance
- iSCSI
  - Uses existing LAN
  - VLAN or physical separation
- FCoE
  - Reduced hardware
  - VLAN separation



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



43

Fibre Channel is an established block protocol that provides high performance, and typically runs at speeds of 4, 8, or 16 Gbps. While Fibre Channel provides a solid foundation for storage access, it requires a separate network infrastructure designed specifically for the FC protocol. This is typically deployed in a redundant configuration which can be costly, and adds complexity to the overall management of a data center. However, a Fibre Channel network does provide high performance and additional security by virtue of the network being a completely independent and isolated network.

iSCSI is also an established block protocol that can be implemented using existing network infrastructure, making it an inexpensive option for scalable storage. Like Fibre Channel, it can also be implemented using a independent and isolated network. If you are implementing iSCSI in an existing network, performance may be a concern due to bandwidth limitations , and network connections of 10G and higher should be considered. Performance can also be improved by using specialized hardware to offload processing from the host CPU, but this increases the cost of deploying iSCSI, especially if a large number of hypervisors are involved. For improved security, consider using VLAN separation to logically isolate iSCSI traffic from other networks. An encryption mechanism such as IPsec is an alternative for securing traffic if this is supported by the storage vendor, but this will add additional bandwidth requirements to the design.

FCoE brings together the flexibility of Ethernet and the reliability of Fibre Channel into a single network. This can reduce the number of I/O cards, cables, and switches in a data center by up to 50%. While this does provide a cost savings, the required switches are not standard Ethernet switches, negating the ability to reuse existing infrastructure. In an FCoE environment, storage traffic will be sharing the same network infrastructure with other types of traffic which means the design must include support for the additional bandwidth. FCoE traffic can use VLAN separation to logically isolate the traffic and QoS capabilities to ensure performance.

# File Storage Protocols



- NFS
  - Primarily used in Linux/UNIX environments
  - Supports encryption
  - Supports multipathing
- SMB/CIFS
  - Primarily used in Windows environments
  - Supports encryption
  - Supports multipathing

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



44

NFS is an established protocol that has been used in Linux and UNIX environments for decades. NFS can be used over existing Ethernet networks and is routable, allowing an NFS server to be accessed from any location. Recent versions of NFS provide secure authentication and data encryption using Kerberos v5 with privacy. Encryption adds overhead to the communication channel and may not be supported by all vendors. NFS functions over a 1Gb network connection, however performance may be better over multiple teamed connections or a larger 10Gb or greater connection. Newer versions of NFS support multipathing and load balancing as well. Parallel NFS is a recent enhancement to the NFS protocol that provides performance improvement. The pNFS separates metadata (onto a metadata server) and data (onto a storage device) for a file.

SMB is a general-purpose network file system protocol that functions similarly to NFS but is used predominantly in Microsoft server environments. The latest version of SMB supports encryption of data across the network as well as multipathing for enhanced throughput and failover.

# Object Storage Protocols



- HTTP(s)
  - Used to access storage solution API
  - Supports encryption

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

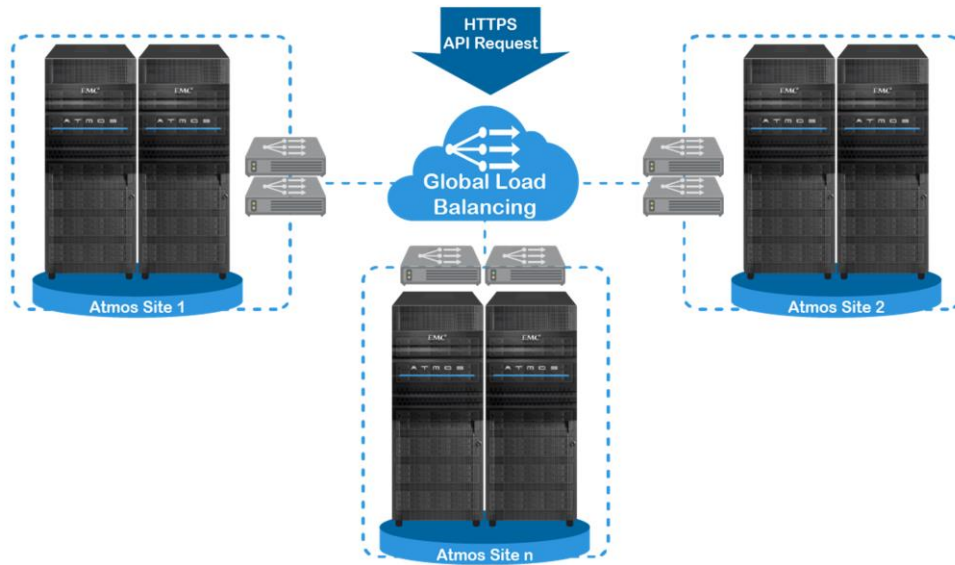


45

Object storage is accessed using the API for the storage solution. The API is accessible over the network using either HTTP or HTTPS protocols.



## Example: EMC Atmos Considerations



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



46

EMC Atmos is an object-based cloud storage platform to store, archive, and access unstructured content at scale. EMC Atmos can be deployed in a single site having multiple nodes. Load balancers are used to distribute traffic across nodes. In a multi-site scenario, global load balancers are deployed to direct users to the closest Atmos nodes for improved performance. In a multi-site design, sufficient bandwidth must exist to support geoparity (multi-site erasure coding capability) and user access.

## Storage Protocol Impact on Cloud Design



? What impact does the storage protocol have on cloud design?

.....



## Storage Protocol Impact on Cloud Design



? What impact does the storage protocol have on cloud design?

- Host support
- Virtual machine support
- Isolation
- Encryption
- Overhead
- Performance

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



48

The first and foremost design consideration is whether the host OS or hypervisor supports the storage protocol. An example of this is VMware ESXi which does not support using SMB for accessing storage.

If the protocol is supported, then the next thing to confirm is whether the latest version of the protocol is supported.

If the plan includes the use of virtual machines that require access to storage directly, then the OS of the virtual machine must support the storage protocol as well. Consideration should also include which protocols make sense for the OS. For example, if the virtual machines deployed within the cloud are running Microsoft Windows and require access to a file share then it would make sense to use the SMB protocol on the storage resource.

Traffic isolation techniques improve security by separating storage traffic from access by users. However, this may also present a problem if both hosts and virtual machines require access to storage.

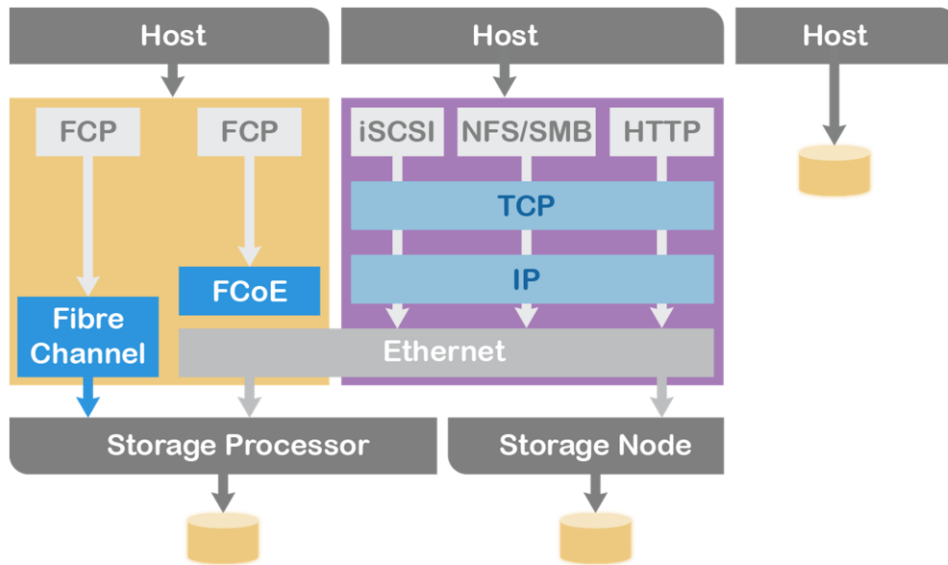
Some storage protocols support end-to-end traffic encryption, but this may produce more overhead than desired. Using network isolation may meet the requirements and reduce bandwidth requirements.

Some protocols such as NFS and SMB may have more overhead associated with them and may impact performance when used with low latency applications. Implementing a block protocol on a dedicated network may be best for this type of application.

In most cases, implementing Jumbo Frame support within the Ethernet infrastructure will improve performance. In fact, FCoE requires a larger frame than the standard Ethernet frame size.

Multipathing capabilities, if supported, requires additional hardware but is important for maintaining throughput and availability in the storage infrastructure.

## Storage Connectivity



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



49

Hosts in a cloud will require access to storage for multiple reasons. One method of accessing block storage is to use local disks connected to the hosts. This configuration is referred to as direct-attached storage (DAS). Since DAS is local and does not use a network, it is covered in the storage module.

Another method for hosts to access storage is across a network infrastructure that has storage devices directly attached, such as disk arrays or tape drives. A Storage Area Network (SAN) is a dedicated communication path between hosts and consolidated block storage devices which supports block storage protocols.

An FC SAN is a Fibre Channel Protocol (FCP) based network dedicated to storage. FC SANs are implemented using a separate infrastructure and encapsulate SCSI commands within Fibre Channel frames.

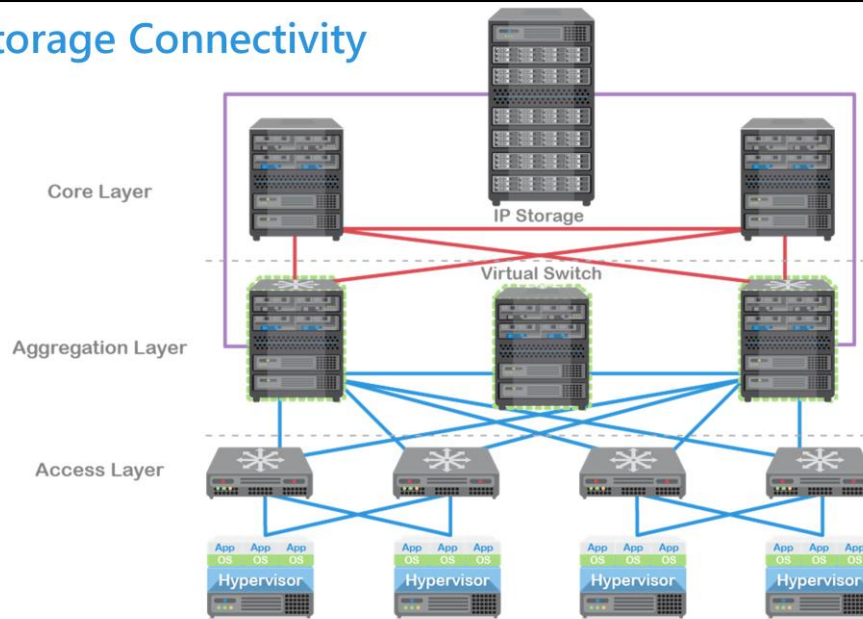
Fibre Channel over Ethernet (FCoE) uses Ethernet networks for connectivity between hosts and storage arrays and layers FCP over Ethernet Frames rather than FC frames.

An IP SAN is a network dedicated to storage traffic that transmits SCSI commands over an Internet Protocol based network. An IP SAN is implemented across existing network infrastructure or dedicated infrastructure. IP SANs support protocols such as Internet Small Computer System Interface (iSCSI).

Network Attached Storage (NAS) also uses an IP based network for connectivity between the hosts and storage arrays. NAS environments use file system protocols such as NFS or SMB for storage access.

Object based storage systems are also accessed over a standard IP based network using HTTP.

# IP Storage Connectivity



? What are considerations for using IP Storage?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



50

## IP Storage Connectivity Considerations



- Connect IP storage to the aggregation layer
- iSCSI
  - Keep iSCSI traffic on an isolated network
  - Use link aggregation on the storage system to increase throughput
  - Use multiple iSCSI targets to distribute load across interfaces
- NFS/SMB
  - Keep NFS and SMB traffic (between hypervisor and storage) on a dedicated VLAN
  - Create publicly accessible mounts or shares on different interfaces

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



51

Implementing an IP SAN follows most of the same guidelines as with standard IP communications. For iSCSI connectivity, create one or more non-routed VLANs and place the hypervisor (initiator) and storage (target) NICs on that VLAN. That avoids performance degradation caused by routing, as well as protects the data being transmitted via iSCSI if it is unencrypted.

If the storage system supports link aggregation, you can utilize that to aggregate connections across aggregation layer switches for redundancy and increased throughput. If the storage system does not support link aggregation, then configure multiple iSCSI targets on the different host NICs and manually load balance the initiators across the available NICs, or use a portal group to associate multiple targets into a single logical entity.

The guidelines for NFS and SMB are very similar. One exception, however, could be that the network between hosts and storage cannot be isolated (non-routable) because the storage interface may need to interact with an external authentication service (Active Directory, LDAP, or NIS). However, it may be possible to place ACLs on the VLAN to prevent traffic initiated externally from accessing those interfaces. If you are also using the storage system as a file sharing device, create additional shares and mounts that are exported via publicly accessible interfaces.

## Fibre Channel SAN

- High speed network of hosts, storage and switches
- Support on fibre channel protocol
- Hosts connect with HBAs
- Storage is centralized and shared
- Fabric is the networking components
- SANs have multiple fabrics for redundancy

© Copyright 2015 EMC Corporation. All rights reserved.

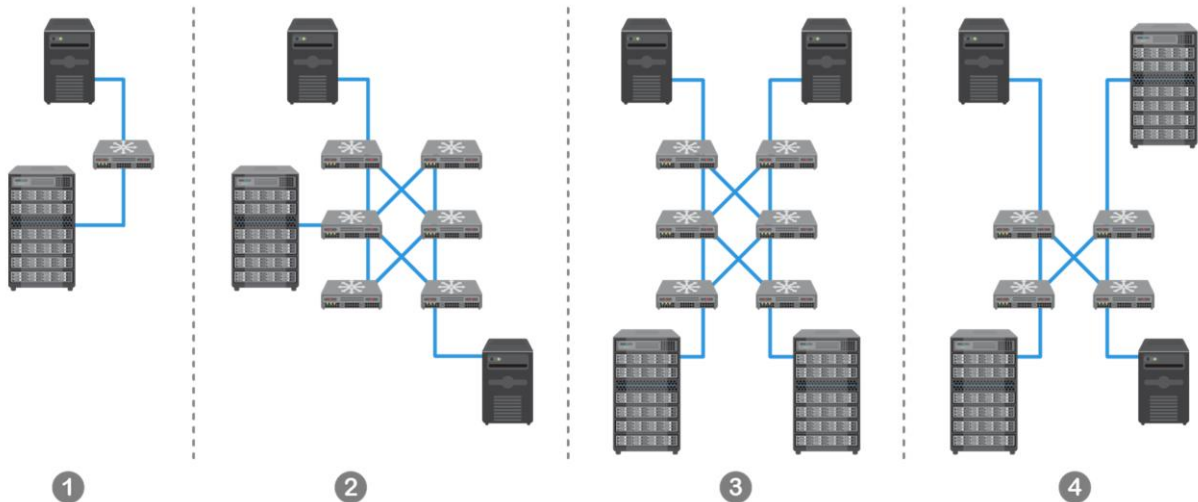
Cloud Infrastructure Planning and Design



52

A Fibre Channel (FC) SAN is a specialized high-speed network of storage devices, switches, and servers/hosts. FC SANs are specifically designed to use only the fibre channel protocol for transporting SCSI commands between host and storage devices. Hosts use specialized devices to attach to a FC SAN called host bus adapters (HBAs). Connecting storage arrays to a FC SAN provides centralized, shared storage to all connected hosts. The network components that connect hosts to storage in a FC SAN is called a fabric. SAN can have multiple fabrics for redundancy.

## FC SAN Topologies



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



53

FC SANs can have many topologies. Listed here are four examples:

Topology one is a small scale single switch fabric with hosts and storage connected to the same switch. This design provides low latency because there are no inter-switch hops but may be difficult to scale.

Topology two is called a core-edge topology, where hosts are connected to edge switches and storage is connected to the core switches. This design has slightly more latency due to the single inter-switch hop between host and storage. It is easier to scale at the host level but may be difficult at the core level.

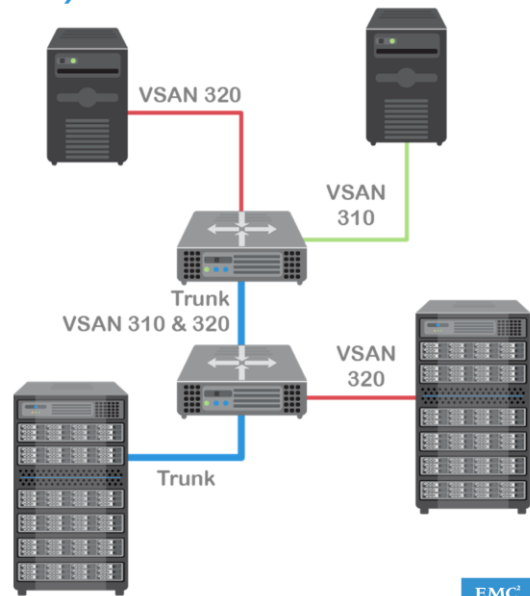
Topology three is called the edge-core-edge topology, where hosts and storage are connected to edge switches which are then connected through core switches. This design provides the most scalability, but adds additional latency due to the second inter-switch hop between hosts and storage.

Topology four is called a full mesh topology, where the FC switches are all interconnected and hosts and storage can be connected anywhere. This design is scalable and latency will never be more than one inter-switch hop.



## Virtual Storage Area Network (VSAN)

- Contains its own
  - Fabric services (name server, zoning)
  - Configuration
  - Set of FC addresses
- VSAN benefits are similar to VLAN
- Tagging enables traffic for multiple VSANs to pass through a trunk link



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



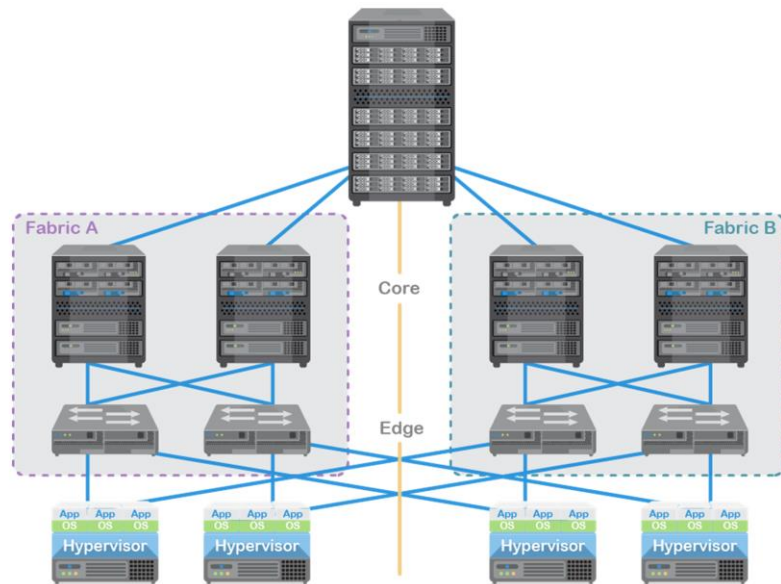
54

A virtual SAN or VSAN is a logical fabric, created on a physical FC SAN. A VSAN enables communication among a group of nodes (physical servers and storage systems) with a common set of requirements, regardless of their physical location in the fabric. A VSAN conceptually functions in the same way as a VLAN.

Each VSAN acts as an independent fabric and is managed independently. Each VSAN has its own fabric services (name server, zoning), configuration, and set of FC addresses. Fabric-related configurations in one VSAN do not affect the traffic in another VSAN. The events causing traffic disruptions in one VSAN are contained within that VSAN and are not propagated to other VSANs.

Similar to VLAN tagging, a VSAN has its own tagging mechanism. The purpose of VSAN tagging is similar to VLAN tagging in LAN. The diagram displayed on this slide shows the assignment of VSAN ID and the frame-forwarding process.

# FC SAN



What are considerations for designing a SAN?

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



55

## Design Considerations: FC SAN



- Select topology to match needs of environment
- Size inter-switch link appropriately
- Use NPV and NPIV to reduce domain count
- Build redundancy into fabric
- Redundant fabrics
  - Servers connected to each
  - Use MPIO
  - Architecturally similar

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



56

The topology that you choose should match your requirements. A core-edge model provides a simple design with moderate scalability for host connectivity. Storage is connected to the core. If you need to add more hosts than a single core switch can support, or need more ports for storage connectivity, you can add more core switches. For the largest environments, an edge-core-edge design allows you to scale up the host ports and storage ports independently. In this model, the core layer acts only as a connectivity layer.

Regardless of the model, it is critical to understand the paths that communications will take between host and storage, and size the Inter-Switch Links (ISLs) appropriately to meet the requirements.

For large environments, especially those that use blade servers or virtualization technologies, using N-Port Virtualization (NPV) and N-Port ID Virtualization (NPIV) allows you to scale the environment further. Consider designing redundancy into fabrics to avoid a single point of failure. Connect servers to multiple fabrics for redundancy and use Multi-Path I/O (MPIO) failover solutions from server to storage. Ensure that redundant fabrics are designed with similar architectures for consistency and ease of management.

## SAN Security

- Zoning controls access between devices on a SAN
- LUN masking controls access to individual LUNs or volumes
- ACLs and RBAC for administrator authentication
- FC Security Protocol (FC-SP) for device authentication
- Port security
- Traffic encryption

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



57

Zoning is a method that controls which devices in a fabric should be allowed to communicate with each other. When zoning is used, devices that are not in the same zone cannot communicate. Zoning is only used to control device communication on a SAN and other security mechanisms, such as LUN masking, may be necessary to limit which specific LUN or volume is visible to the specific host. Additionally, zoning provides protection from fabric disruption by limiting the scope of change notifications to devices within a particular zone when a change occurs.

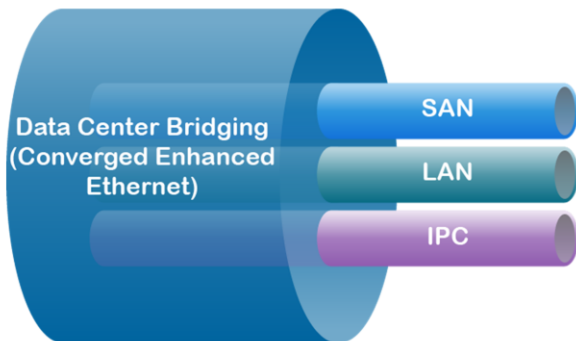
Access control lists and role-based access control provide protection by defining which users have access to switches or storage devices and which functions are available to these accounts. Use encrypted communication protocols to access all component management interfaces.

The Fibre Channel Security Protocol (FC-SP) can be used to force switches to authenticate ISL partners and hosts that are connecting to switches.

Port security allows you to specify which WWN is allowed to connect to a specific switch port, preventing rogue devices from being connected to a switch. Disabling inactive ports also prevents unauthorized devices from connecting to the fabric.

To secure data transmissions, consider implementing encryption methods for all data traffic transferred between devices.

## Converged Networks: SAN and LAN



### Benefits

- Reduce infrastructure costs
- Reduce operational costs
  - Decrease power, cooling, and space
  - Reduce complexity
  - Fewer components to maintain
- Encapsulate FC frames over Ethernet
- Enables FC to run on a lossless Ethernet

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



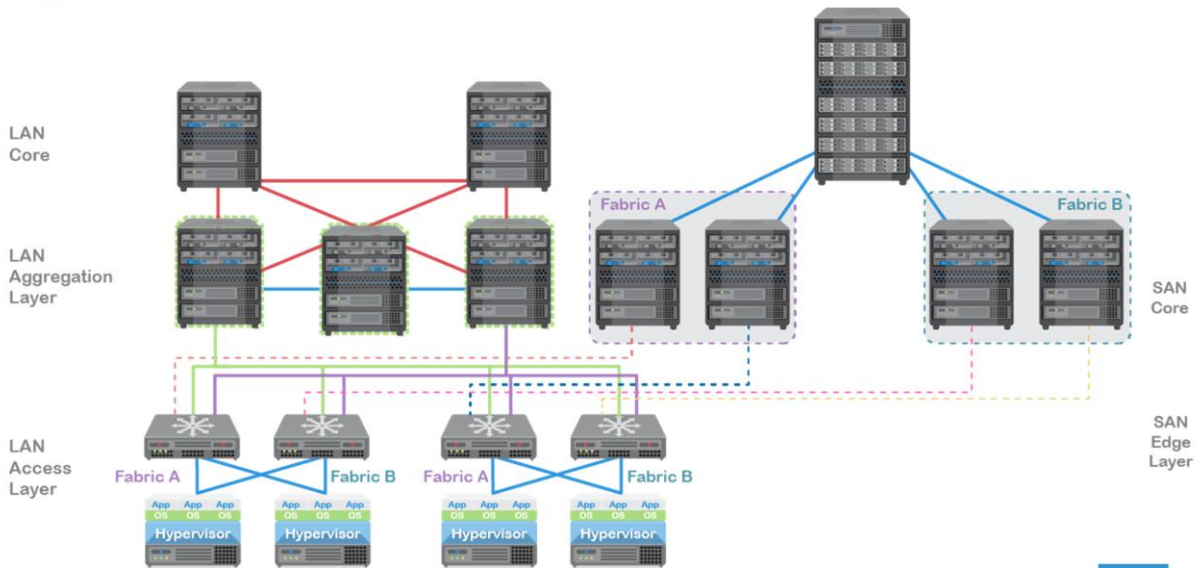
58

Classic LAN and WAN environments provided the infrastructure for IP Storage networks. Storage protocols (for example iSCSI, FCIP, NFS, and CIFS) use TCP or UDP over IP-based networks. IP, a Layer 3 protocol, can similarly operate over a multitude of Layer 2 network protocols, such as Ethernet, ATM, and others. This flexibility made these protocols an attractive option when looking to deploy a low-cost storage network (in comparison to a Fibre Channel SAN). However, performance for these protocols was always a concern, as the limitations of Gigabit Ethernet became apparent when comparing performance to 4 or 8 Gbps Fibre Channel environments.

As 10 Gigabit Ethernet grows in popularity, these IP Storage technologies become more attractive as the performance limitations no longer exist. However, companies that have invested in Fibre Channel implementations are hesitant to discard proven solutions and adopt technologies that use a very different management model.

Data Center Bridging (DCB), also known as Converged Enhanced Ethernet (CEE), bridges the gap between traditional Fibre Channel and Ethernet/IP networks. DCB utilizes a 10 Gigabit Ethernet infrastructure, with additional functionality incorporated. While standard 10 Gigabit Ethernet can support the traditional IP SAN technologies, DCB can also support the Fibre Channel over Ethernet (FCoE) protocol over the same physical infrastructure. FCoE combines the functionality and reliability of Fibre Channel with the flexibility of Ethernet. In addition, with the features that are incorporated into DCB, other data protocols can also be consolidated onto the same infrastructure.

## Edge FCoE



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



59

In an Edge FCoE design, the access layer switches must support FCoE, FC, and Ethernet. A single connection from the host to the access layer switch provides converged connectivity.

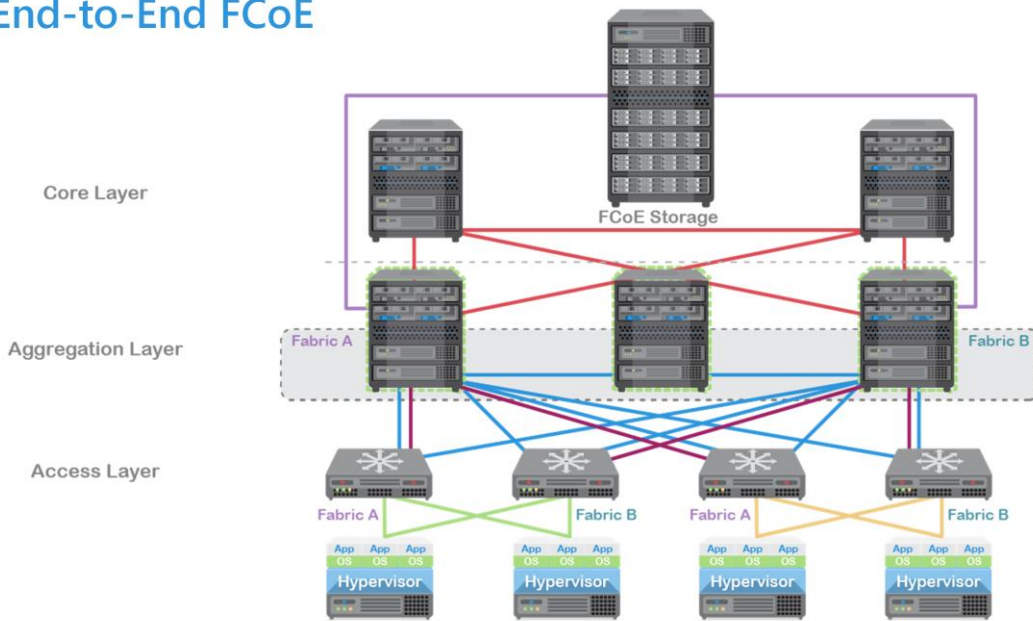
The connectivity between the access and aggregation layers needs to only support traditional Ethernet, so the aggregation layer switches do not need to be converged. Connectivity from the access layer to the SAN is accomplished using FC ISLs from the access layer to the SAN switches.

Each access layer switch is only connected to a single FC fabric, A or B. Connecting them to both fabrics can cause the fabrics to merge (if not configured properly), and does not provide any benefit because of zoning and masking.

This model is ideal for environments that want to migrate to FCoE over time. It provides a hybrid environment that supports both protocols, and hosts and storage can be migrated to FCoE over time.

For environments that have a large number of converged access layer switches, you can similarly connect the two environments from the LAN aggregation layer. If you choose this model, however, you will need to deploy converged switches at both the access and aggregation layers.

## End-to-End FCoE



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



60

In an End-to-End FCoE design, the converged network is extended to the aggregation layer. However, parallel Ethernet and FCoE connectivity exists between access and aggregation layers. Because each switch is a member of only a single FCoE fabric, you do not want to cross-connect the switches as you do for the Ethernet environment. Instead, you want to only connect each access layer switch to a single aggregation layer switch, or to multiple switches in the same fabric.

Storage is connected at the aggregation layer and not at the core. Since there is no Layer 2 connectivity between the aggregation layer and core, you cannot have FCoE traverse that link because it is not routable.

## Lesson 2

### Summary

Key points covered in this lesson:

- Storage can be accessed directly or through an IP or FC storage network topology.
- The selected storage topology depends on storage type, performance, security, and other requirements.



This lesson covers the requirements and design consideration for implementing storage networks in a cloud.



## Lab 5: Network Resource Design

### Purpose:

Develop a design and plan for deploying consumer network resources in a cloud infrastructure.

### Tasks:

- Identify consumer network requirements that are relevant to NanCo cloud design.
- Provide a logical diagram that describes the consumer network resources and is based on the NanCo requirements.

## Lab 5: Debrief

### Network Resource Design

**Purpose:**

Develop a design and plan for deploying consumer network resources in a cloud infrastructure.

# Module 5

## Network Resources Summary



Key points covered in this module:

- Network pools are network resources that are logically grouped based on various characteristics.
- Network segmentation is a key component in a cloud design and can be implemented using physical and virtual network separation methods.
- Storage networks can be implemented over LANs and SANs and support the various types of storage.



Cloud Infrastructure Planning and Design

64

This module covered requirements and considerations that relate to the design of consumer network resources in a cloud design.

This slide intentionally left blank.





# Module 6

## Elasticity, Monitoring, and Metering



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges, and considerations of implementing technologies that support elasticity, monitoring, and metering in a cloud infrastructure.
- Align business requirements when designing infrastructure to support these capabilities.
- Develop a design and plan for these capabilities in a cloud infrastructure.



Cloud Infrastructure Planning and Design

1

This module focuses on requirements and design considerations when supporting application elasticity, and implementing monitoring and metering in a cloud.

# Lesson 1

## Elasticity

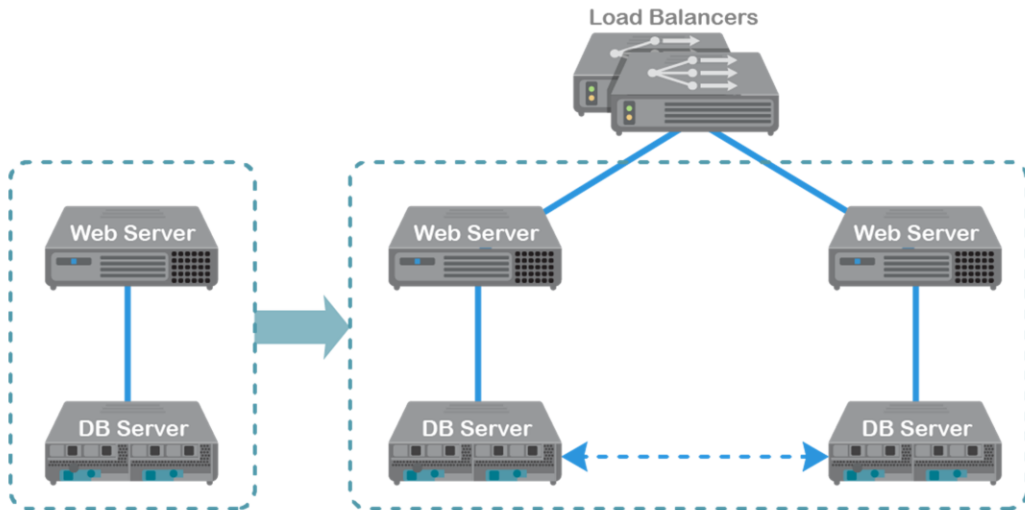
This lesson covers the following topics:

- Resource requirements for elastic applications
- CMP components that enable elastic capabilities in a cloud



This lesson covers the resource and CMP requirements for enabling elastic applications.

## Application Elasticity



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



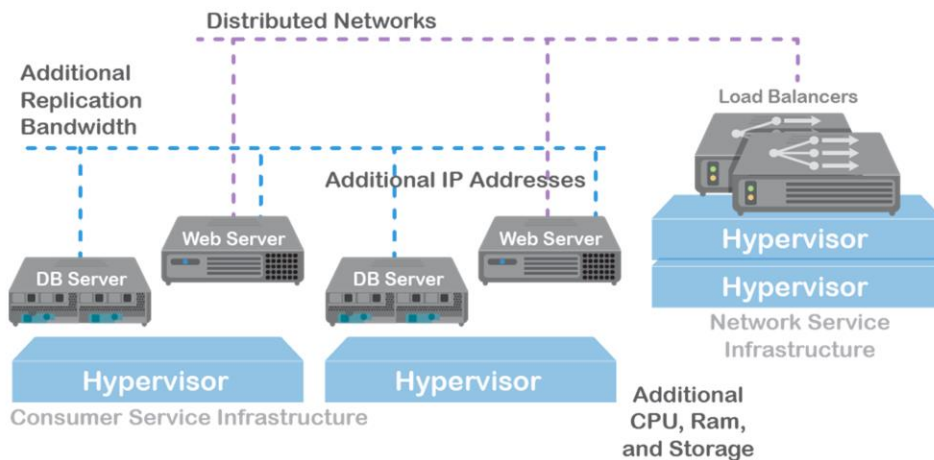
3

The NIST definition for Rapid Elasticity is "Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time." From a cloud infrastructure design perspective, rapid elasticity is not about allocating resources to the cloud infrastructure. This may never really be rapid since it takes time to procure hardware, software, licenses, and so on. From a consumer's point of view though, rapid elasticity means that when demand increases or decreases, services can be scaled out or scaled back to meet the demand. This viewpoint of rapid elasticity does have impact on the cloud design. If an organization has a requirement for elastic services, then the infrastructure must have certain components in place to support this. Some of these components have been discussed, but this module focuses on the infrastructure needed to support application elasticity.

For illustration purposes, we'll use this simple example of scaling out an application consisting of a web server and database server. To scale-out the web tier, we'll add additional instances as needed and use a load balancer. To scale-out the database tier we'll add an additional instance of a database server and use replication to maintain an active/active configuration.



## Resources



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



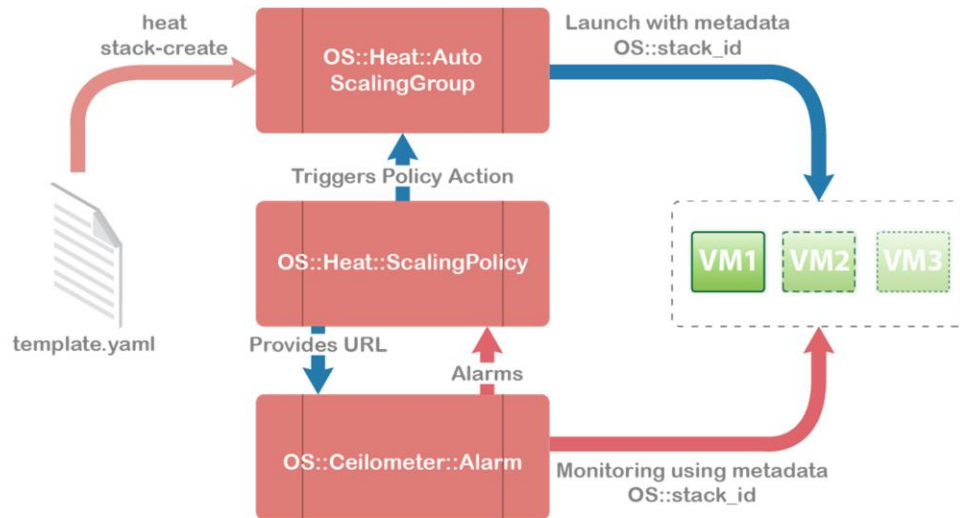
4

Designing a cloud to support elastic applications means understanding the impact on infrastructure resources. In this example, consumer services and network services have their own infrastructure. Generally speaking, the redundant instances will be spread across multiple servers to maintain availability in the event of a server failure.

The addition of a second database means that additional CPU, memory and storage are required to support the second instance and also both instances will require more CPU and memory to support the replication between the instances. From a network perspective, the design must take into account additional IP addresses for the database server, a network segment that spans across all hosts and bandwidth to support the additional replication traffic.

The addition of more web front-end servers also means additional CPU, RAM, and storage to support instances. The web servers will also require additional IP addresses, and a network segment that spans the consumer service infrastructure and is isolated from the database server network segment. Load balancers can be used to balance the web traffic across front-end servers. These will require additional CPU, memory, and storage resources on the network service infrastructure as well as access to the network segment connecting the web servers.

## Manual vs. Automatic



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

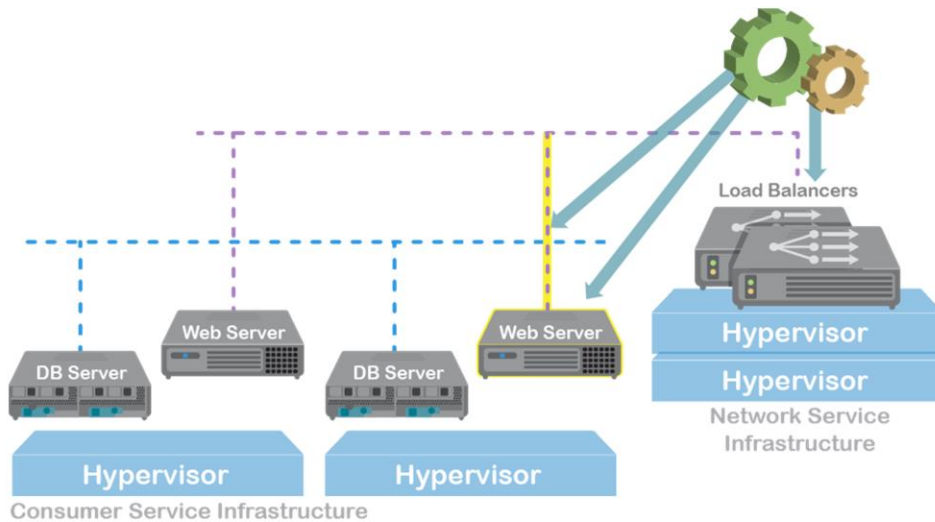


5

There are two methods to scale applications in and out: manual and automatic. Applications can be scaled manually through the use of the service catalog. Consumers can use the management interface to access an existing application and change the number of instances for the web server. This functionality must be supported in the portal and service catalog selected for the cloud.

Automatic scaling happens when a certain threshold is reached, such as CPU usage, that is captured by a monitoring mechanism and then triggers a change using orchestration capabilities. As in the OpenStack example shown here, the environment could be implemented using an auto scaling group and scaling policy and the design would need to include Heat services for orchestration, Ceilometer for monitoring and threshold triggering.

## Orchestration



© Copyright 2015 EMC Corporation. All rights reserved.

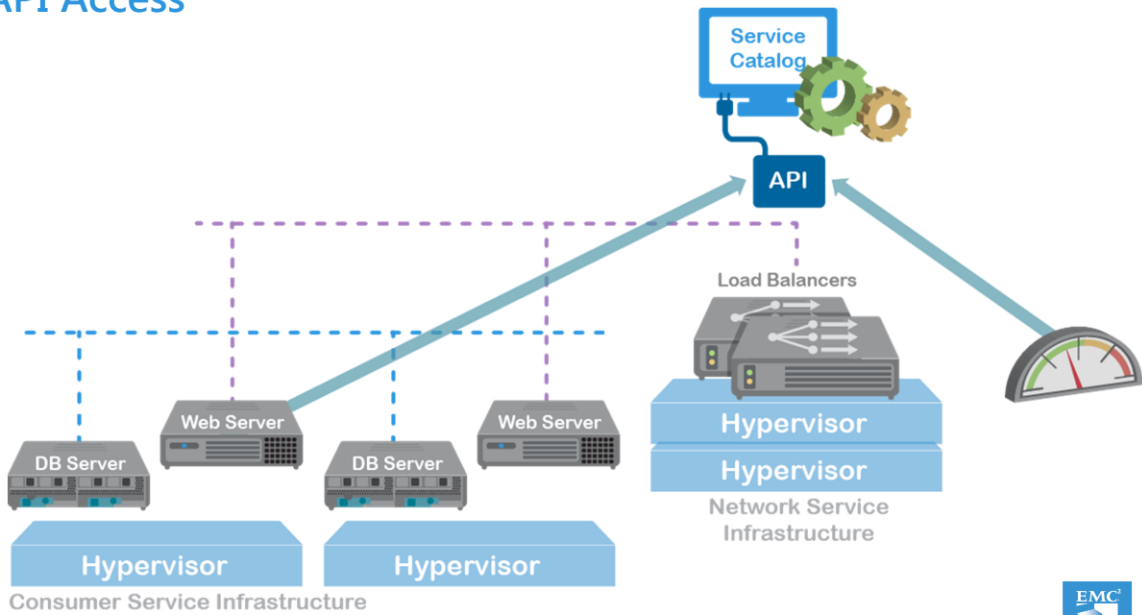
Cloud Infrastructure Planning and Design



6

An orchestration engine is required that can deploy additional virtual machine instances and connect them to the proper network with a new IP address. The orchestration engine may also need to update the load balancers with any IP addresses of new instances. A configuration management application would be useful for ensuring that the web server application and IaaS instance is up-to-date and match the already deployed web services.

## API Access



© Copyright 2015 EMC Corporation. All rights reserved.

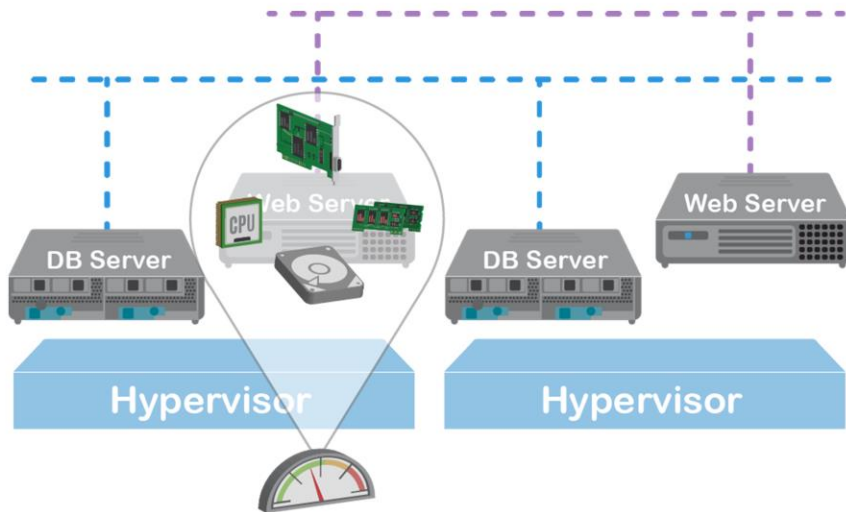
Cloud Infrastructure Planning and Design



7

In some solutions, the monitoring component directly interfaces with the API for the portal/service catalog in order to trigger an additional instance creation. This API may also be addressable from the services themselves to trigger additional resources since it is possible that the application can self-monitor for a condition that will trigger additional instances. In either case, the API must be made available and secured to ensure scaling capabilities.

## Monitoring



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

The monitoring system used to support application scaling must be able to monitor all of the infrastructure such as CPU, memory, network and storage consumption. Although the monitoring solution will be able to provide insight into the entire infrastructure, it must be able to also provide this information as it relates to the service being measured.

## Other Supporting Services

- Chargeback system
- Load balancers
  - Virtual: Sufficient network infrastructure
  - Physical: Orchestrate changes
- IP Addresses
  - Automation
  - DNS

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



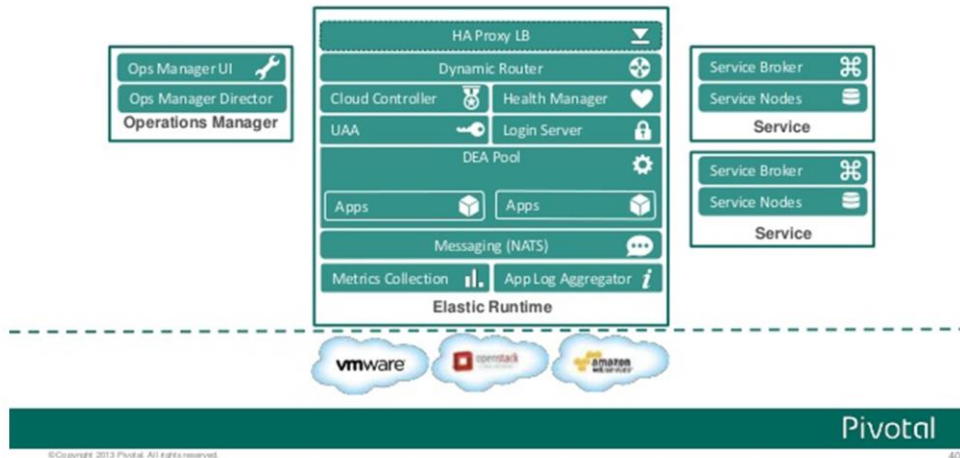
9

An integrated chargeback system must be in place to capture newly-created instances and associate them with the proper tenant.

As mentioned previously, load balancers will be needed to support front-end web servers. However, the requirements may dictate the use of virtual load balancers and the need to be deployed on a per-application basis. If this is the case, then the network service infrastructure must have sufficient resources for all planned services. This may also require that additional network segments be made available to the network service infrastructure. If this is a small, single-tenant, private cloud environment that is supported by enterprise class hardware load balancers, then an orchestration capability must be developed to make load balancer configuration changes.

Basic network capabilities such as IP address assignment and DNS entry creation may need to be automated to support the addition of services. Many of these considerations can be addressed by deploying a software-defined networking solution such as VMware NSX. This solution can be used to programmatically deploy and manage load balancers, attach these to new or existing network segments and assign IP addresses via DHCP services.

## Pivotal CF Architecture



If you are deploying a PaaS solution, elastic application capabilities may be included in the platform and may be as simple as selecting the number of instances to deploy and programmatically adjust. However, you need to research the particular PaaS platform to understand resource requirements to enable elastic capabilities. Pivotal Cloud Foundry is one example where it is important to understand the solution's capabilities. By default, this solution includes a single instance of HAProxy load balancer which may be adequate for test and development, but Pivotal recommends deploying a more robust load balancing solution for production environments.

# Lesson 1

## Summary

Key points covered in this lesson:

- Elastic applications need additional infrastructure resources such as CPU, storage, bandwidth, and IP addresses.
- The orchestration engine is used to scale application instances.
- The monitoring tools are used to trigger application scaling.



This lesson covered the resources and CMP components that are necessary to scale applications elastically.



# Lesson 2

## Monitoring

This lesson covers the following topics:

- Identifying reasons for implementing monitoring tools
- Design considerations for the selection and placement of monitoring tools in a cloud



This lesson covered the purpose and design considerations for monitoring tools in a cloud.

## Monitoring

- Service availability and performance
- Infrastructure availability and performance
- Ensure requirements and SLAs are met
- Capacity planning
- Security issue
- Aligns to the type of service provided

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



13

A cloud monitoring solution ensures that your services and infrastructure remain up and perform appropriately. You can monitor for application availability and responsiveness as well as for infrastructure availability, performance, security issues, and usage. There are many reasons why monitoring is necessary in a cloud environment. The first is to ensure that the services are performing to the requirements gathered in the assessment and specified in service level agreements. Monitoring is also used for capacity planning to ensure that the cloud does not run out of resources. Another reason for monitoring is to collect security-related information to notify about breaches or to support audits. A final reason for implementing a monitoring tool is to support the automation of elastic applications.

From a cloud provider's perspective, tools that are used to ensure consumer application availability are more focused on Software as a Service offerings. This is because cloud service providers are not responsible for applications developed on IaaS or PaaS instances—the consumer has responsibility for these applications. The cloud service provider is responsible for the underlying infrastructure and the development platforms. As an example, the service provider can monitor the infrastructure that supports an IaaS instance to ensure adequate capacity and performance of resources. Since it has no control over what is deployed within the instance or whether the instance is even turned on, the cloud provider has no need to monitor at this level.

However, consumers require insight as to how their services are performing or are being used. Cloud service providers offer this information to consumers for planning purposes and to identify unnecessary services. Consumers may wish to obtain performance data or consumption information for an application that they developed and are testing. Consumers or developers may also wish to have access to monitoring tools to take advantage of elastic provisioning capabilities.

## What to Monitor



- Cloud management components
- Infrastructure components
- Compute, network, and storage resources
- Hypervisors
- PaaS infrastructure
- SaaS
- Consumer applications

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



14

The cloud provider requires tools to monitor the various components that are part of their responsibility. Components that should be monitored include:

- Cloud management component availability and responsiveness
- Cloud management supporting applications such as databases or message queues
- Infrastructure account access
- Infrastructure component failures
- Infrastructure errors and events
- Infrastructure configuration changes
- CPU and memory resource utilization
- Network consumption and response times
- Storage consumption and response times
- Hypervisor utilization and response times
- PaaS development environment utilization, response, and events
- SaaS availability and responsiveness

For various reasons, the cloud provider may also provide monitoring capabilities for consumers to monitor their own services.

## Choosing Monitoring Tools

- Single or multiple tools
- Flexible plug-in architecture
  - Standard format, protocols, and APIs
- Agent(less)
- Accounts
- Multiple clouds



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



15

A cloud environment is a complex blend of various management solutions, hardware components, and software, and each of these will most likely have its own monitoring tools. It is possible to implement a cloud and use the multiple monitoring tools, but this may not present a holistic view of the environment or any issues that arise. The key to having an effective monitoring solution in the cloud is to select tools that can pull all of this information together and present information as simply as possible. This means selecting a tool that can either collect or unify statistics and events from all of the underlying components or consolidate the information from the different component monitoring tools.

Some monitoring solutions are extensible through a flexible plug-in architecture. These plug-ins may directly poll devices through an API or CLI, or using the Simple Network Management Protocol (SNMP) to gather information. Monitoring tools can also passively stand-by waiting for information such as through SNMP traps or SYSLOG events. With direct polling, monitoring tools will require a username and password to access components. Select a monitoring tool that properly secures this login information on the monitoring server and over the network.

Monitoring tools may also collect information through agents deployed throughout the infrastructure, and if this is the case, additional resources may be required to support these agents on the target components. Although agents may require additional resources they also have a security benefit. By polling the agent directly, monitoring tools do not require a username and password to be stored centrally or transmitted over the network.

Monitoring tools are of very little use if they can not notify cloud provider staff of problems. Select a monitoring tool that can provide alerts through email, SMS, or any other mechanism required by the organization. Support for an external alert capability, such as direct dial to a paging service, should be considered in the selection process.

A final consideration when selecting a monitoring tool is to include capabilities to connect and monitor other clouds such as on a public cloud provider. Multiple methods for accomplishing this are covered in the hybrid cloud design module.

## Analysis and Alerts

- Analysis
  - Identify root cause or dependency events
  - Future prediction
    - Aids in timely upgrades
- Alerts
  - Email, SMS, or other mechanism
  - Multiple message delivery systems
    - One external system

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



16

Selecting a tool with an analysis engine or that can identify dependencies will help to avoid unnecessary alerts. In a cloud environment, components are interrelated in many ways and if something fails, events may be triggered at multiple levels within the infrastructure. A monitoring tool should have the capability to identify component dependencies and report on the lowest layer or root cause failure. An analysis capability will also be able to more accurately predict future trends for consumption or performance which aids in planning for upgrades in a timely fashion.

Monitoring tools are of very little use if they cannot notify cloud provider staff of problems. Select a monitoring tool that can provide alerts through email, SMS, or any other mechanism required by the organization. Remember that if the email system resides in the same infrastructure or datacenter as the cloud, support for an external alert capability, such as direct dial to a paging service, should be considered for the design.



## Monitoring Tool Placement

- Located in cloud management infrastructure
- Access to all components
- Secure inbound access
- Secure storage
- Access to message delivery systems
- May be deployed as a service from an external vendor

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



17

Monitoring tools are an important part of cloud management and will therefore be deployed in the cloud management infrastructure. Monitoring tools capture data that prove that the cloud provider is meeting service levels and complying with regulations. They are also used to support audits and for other business critical or legal purposes. This data must be protected from unauthorized access, deletion, and corruption, and should be placed on highly secure, available, and reliable storage with a solid backup plan. Since so much data is collected, the storage and network infrastructure should have high bandwidth and low latency. The monitoring solution will be the primary alert mechanism for the cloud and as such will require direct access to email, SMS, or other messaging systems to send alerts. You may wish to consider a secondary monitoring system outside the cloud management infrastructure that send alerts in the event the primary monitoring solution becomes unavailable.

An alternative to deploying an in-house monitoring tool is to implement an external monitoring service. Some vendors offer application and server monitoring as a service which has the benefit of removing the monitoring tool from the infrastructure being monitored. The implementation details vary, but using an external provider requires enabling network access to infrastructure resources, providing user accounts to those resources, and supplying enough bandwidth to collect data.

## Monitoring Tool Sizing



- Expect high activity and volume
- Follow vendor or community recommendations
- Factors
  - Number of targets
  - Number of metrics
  - Data retention
  - Performance
- Future growth and scalability

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



18

Your monitoring solution will most likely include a component for data collection reporting, analysis, and database. Because of the multiple components, high activity, and high volume of data, expect that the monitoring solution will consume the greatest amount of resources in the cloud management infrastructure. The cloud design should follow vendor and community recommendations for sizing and placement. One of the sizing components will be the number of targets being monitored. Include future growth requirements in the sizing especially if the monitoring solution architecture needs to change as the environment becomes larger. For example, if the solution requires that the database server is run on its own instance once the number of targets reaches a threshold and the future plans for the cloud will exceed that threshold, it may be easier to plan for a separate database server in the initial design.

The amount of storage required for the monitoring solution depends on requirements for the number of targets to be monitored, number of metrics to be monitored on each target, and the retention time for the data. Retention time may be dictated by legal requirements, regulatory compliance, or business policy. It is important to determine the correct requirements for retention during the assessment phase since an incorrect assessment may require a future design modification and expose the business to legal issues. Some tools provide an archiving mechanism that can be used to move or export older data to a less expensive and slower storage tier which helps to maintain performance and meet retention requirements.

## Monitoring Access and Security

- Two audiences
  - Consumers
    - Service specific metrics and statistics
    - Multi-tenancy support
  - Cloud Provider
    - Infrastructure and service metrics and statistics
- Secure network traffic
- Secure account info when using automation

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



19

Two audiences may need access to the monitoring tools, consumers, and cloud provider staff. These two audiences, however, should not have access to the same set of information. Since a cloud is a multi-tenant environment, consumers should only have access to the metrics and statistics that pertain to the service instances for which they have permissions. Consumers should not see metrics or statistics for other service instances nor the underlying infrastructure. Cloud provider staff, on the other hand, will need access to the metrics and statistics for the infrastructure and may need access to some or all of the service instance information. The design needs to either include a monitoring tool that supports multi-tenancy, or to include two tools: one to be used for infrastructure monitoring and the other for application monitoring that can ideally be integrated with the cloud portal.

Monitoring tools can be used to support service automation as described previously in the elasticity lesson. To do this, the API for the monitoring tool may need to be accessible by consumers or their applications to support programmatic polling of the monitoring tool or setting alert thresholds. Consumers may wish to implement their own monitoring tool which monitors multiple cloud sources and will require API access to centrally collect information. This is similar to how public cloud providers expose some monitoring information to consumers through a publicly documented API.

Whether access is granted through a UI or API, the account information should be encrypted over the network. The monitoring tool should also have integration capabilities with the organization's authentication mechanisms to ensure proper access control and provide a single sign-on capability.



## Logging

- Centrally collect logs
- Detect errors, failures, general events, security events
- Integrate with entire infrastructure
- Accessed by cloud provider



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



20

Collecting logs is an important part of the monitoring process. Logs should be gathered in a central location to ensure retention, minimize tampering, and improve analysis. Log collection and analysis can be used to identify component failures and errors, security events, application events, and other information. The log collection tool should have search capability and an analysis engine to help identify issues that are relayed across the environment or that have dependencies on other issues. Ideally, the tool should be able to collect logs from all infrastructure components rather than implement separate log collection tools for each component.

Generally, infrastructure logs are accessed only by the cloud provider since granting accesses to tenants creates privacy concerns with other tenants. If information must be provided to tenants or auditors, a report generation tool and process is needed to ensure information that is shared does not violate privacy rules or policy.

## Lesson 2

### Summary

Key points covered in this lesson:

- Monitoring tools are used for capacity planning, event notification, and performance and availability troubleshooting.
- Monitoring tools must have secure access to many infrastructure components.
- Monitoring tool sizing depends on the number of targets and metrics to be monitored.



This lesson covered the reasons and design considerations for deploying monitoring tools in a cloud.

# Lesson 3

## Metering

This lesson covers the following topics:

- Identifying reasons for implementing metering tools
- Design considerations for the selection and placement of metering tools in a cloud



This lesson covered the purpose and design considerations for metering tools in a cloud.

## Metering

- Measuring and recording usage of services and resources
- Used for billing
  - Chargeback
  - Showback



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



23

Metering is the process of measuring and recording usage of services and their underlying resources, usually with the intention of providing billing information for that usage. Examples of service measurements include service uptime, CPU consumption, network bandwidth consumption, and storage consumption. Metering tools can be integrated with other monitoring tools that collect the usage statistics, and then the metering tool aggregates the information and generates billing information. Chargeback is the process of reporting service usage and collecting money for that usage. Showback is the process of reporting service usage without collecting revenue.

## Integration

- Infrastructure components
- Portal and service catalog
  - Multi-tenancy
  - Cost information

© Copyright 2015 EMC Corporation. All rights reserved.

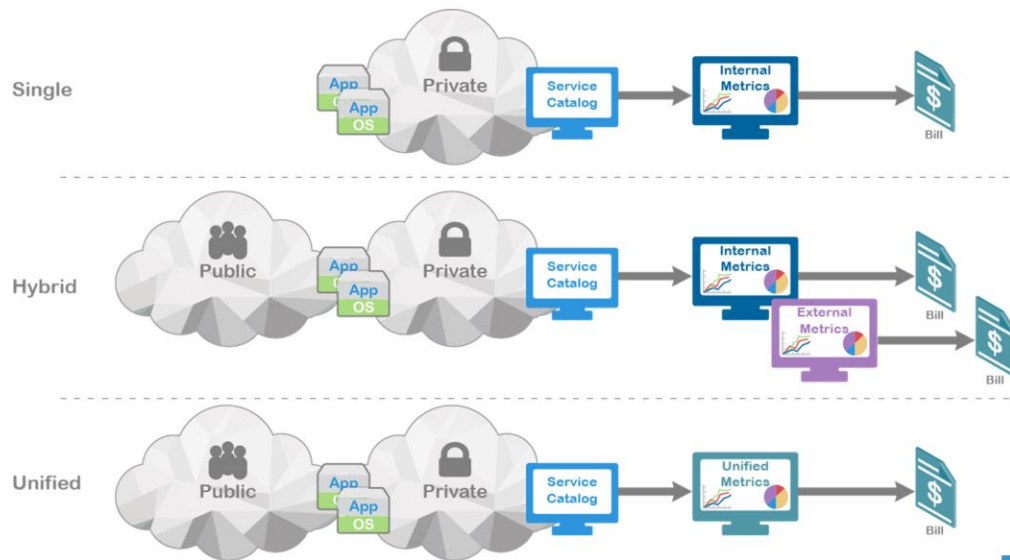
Cloud Infrastructure Planning and Design



24

There are many integration points to consider when you select a metering tool. The tool must be able to log in to various infrastructure components to collect usage metrics. However, this information is only partially useful since there is no relationship between the service and the tenant. Therefore the tool must also integrate with the cloud portal and catalog so that the metering tool is aware of services being instantiated, and usage and billing information can be mapped to the proper tenant. Catalog integration is also important so that consumers can visually see the costs of services that will be deployed and so that cost information flows through to the billing system.

## Metering Options



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



25

The effectiveness of a metering solution relies on the visibility into how cloud resources are being used and where services are deployed. A single metering and billing solution can be used when services are deployed in a single private or public cloud. However, if services are deployed across multiple clouds, then a hybrid or unified metering and billing solution is required. A hybrid solution calls for metering services separately in each cloud and then providing billing or reports from each. A unified solution is a single tool that can provide metering and billing for both clouds.

## Metering Tool Sizing

- Fewer resources than a monitoring tool
- Follow vendor or community recommendations
- Factors
  - Number of services
  - Number of metrics
  - Data retention
    - Pricing model
- Future growth and scalability



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



26

The sizing considerations for metering tools and monitoring tools are similar. Because the metering tool may not have as frequent of a polling interval and will collect only a subset of the metrics that a monitoring tool does, its requirements for resources may be lower. In fact, if the metering tool integrates with the monitoring tool, then resource consumption will be drastically less, since the monitoring tool will do most of the work. As with the monitoring tool, the cloud design should follow vendor and community recommendations for sizing and placement. Metering is concerned with services only, so the number of services to be metered and the polling frequency affects sizing. Include future growth of services in the sizing calculations as well.

The amount of storage required for the metering solution depends on requirements for the number of targets to be monitored, number of metrics to be monitored on each target, and the retention time for the data. Retention time may be dictated by the metrics needed to support the pricing model. For instance, a two year subscription for a service may store key metrics for the entire lease period, whereas a pay-as-you-go model with a monthly billing cycle may retain metrics for the billing cycle only.

## Metering Tool Placement

- Located in cloud management infrastructure
- Access to all components
- Secure inbound access
- May be deployed as service from external vendor



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



27

Metering tools are part of cloud management, and are deployed in the cloud management infrastructure. Like monitoring tools, the metering tools require access to infrastructure components. Because of this, access to the metering servers should be limited. Any accounts used by the metering solution should have minimal rights, and account information should be stored in an encrypted format. If the tool does support multi-tenancy and integrates with the cloud portal, then public access to servers may be necessary and security will be addressed using accounts and role-based access.

An alternative to deploying an in-house metering tool is to implement an external metering service. The implementation details will vary, but using an external provider will require enabling network access to infrastructure resources, providing user accounts to those resources, and supplying enough bandwidth to collect metrics.



## Metering Access and Security

- Consumers
  - Service usage and billing
  - Multi-tenancy support
- Secure network traffic
- Integration with internal authorization sources

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



28

Since a cloud is a multi-tenant environment, consumers should only have access to the metrics and billing information that pertain to the service instances for which they have permissions. Consumers should not see information for other tenants. Access to the UI should require encryption over the network to protect account information. The metering tool should also have integration capabilities with the organization's authentication mechanisms to ensure proper access control and provide a single sign-on capability.

## Lesson 3

### Summary

Key points covered in this lesson:

- Metering tools are used for recording service usage and for billing.
- Metering tool sizing depends on the billing cycle and number of service and metrics to be monitored.



This lesson covered the reasons and design considerations for deploying metering tools in a cloud.

## Lab 6: Elasticity, Monitoring, and Metering

### Purpose:

Understand the role of elasticity, monitoring, and metering and its impact on a cloud design.

### Tasks:

- Identify relevant elasticity, monitoring, and metering requirements for the NanCo cloud design.
- Amend the NanCo cloud design to reflect these requirements.

## Lab 6: Debrief

### Elasticity, Monitoring, and Metering

**Purpose:**

Understand the role of elasticity, monitoring, and metering and its impact on a cloud design.

# Module 6 Elasticity, Monitoring, and Metering Summary



Key points covered in this module:

- Elastic applications require additional infrastructure resources which are provided using various CMP capabilities.
- Monitoring and metering tools must have secure access to many infrastructure components.
- Monitoring and metering tools provide feedback about resource consumption, performance, events, availability, and capacity.

Cloud Infrastructure Planning and Design



32

This module covered requirements and considerations that relate to the application elasticity, monitoring and metering in a cloud design.

This slide intentionally left blank.





# Module 7

## Hybrid Cloud Capabilities



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges, and considerations of implementing a hybrid cloud infrastructure.
- Align business requirements when selecting technologies to support hybrid cloud capabilities.
- Develop a design and plan for a hybrid cloud infrastructure.

Cloud Infrastructure Planning and Design

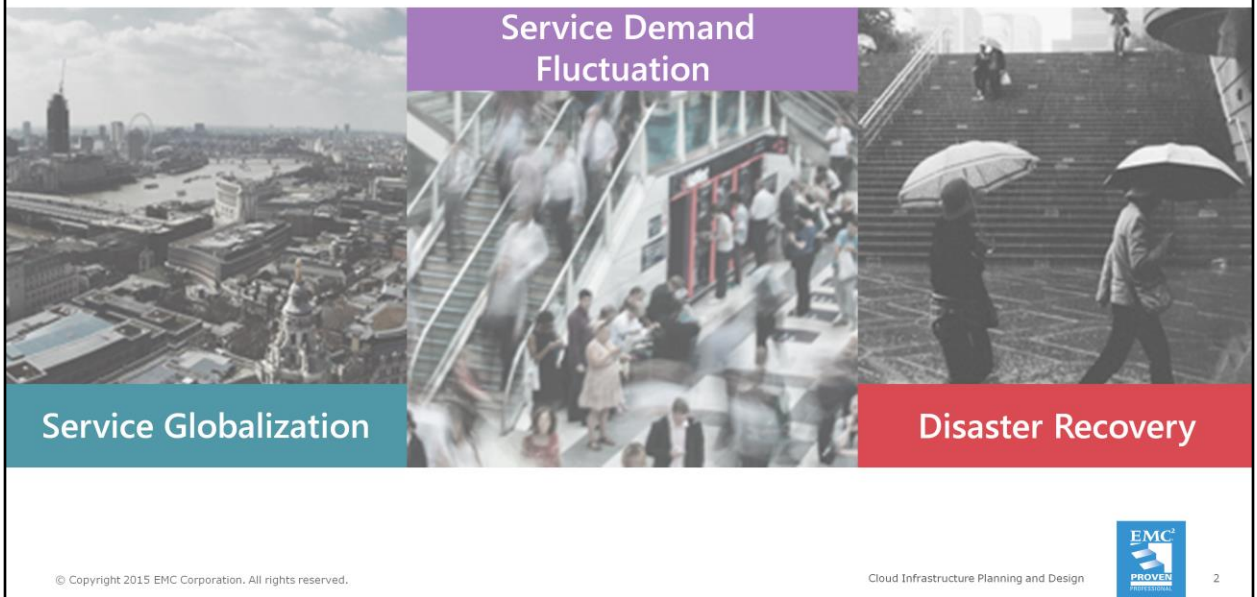


1

This module focuses on requirements and design considerations that support hybrid cloud capabilities.



## Hybrid Cloud Use Cases



**Service Globalization** – For many organizations using the private cloud model, cloud infrastructure exists in only one or possibly a small number of locations. If consumers are spread across the world then service performance may suffer from network latency. One way to reduce latency and maintain acceptable performance is to distribute services globally as well. Organizations can adopt a hybrid cloud model, using internal cloud infrastructure as well as that from a public cloud provider to deploy services around the world and bring services closer to consumers. To support this, organizations will require connectivity between clouds, cloud management integration, global load balancers, and distributed supporting services (for example, authentication or software patching).

**Service Demand Fluctuation** – Another use case for the hybrid cloud is to support fluctuation of demand for services. Rather than purchase enough private cloud infrastructure to support worse case demand scenarios, organizations can use public clouds to augment internal resources when demand is high and then scale back public cloud usage when demand returns to normal. This scenario requires similar functionality as described in service globalization described above.

**Disaster Recovery** – A third use case for hybrid cloud is disaster recovery. There are many ways to accomplish this. One would be to distribute services across multiple clouds in an active-active configuration where services in one cloud can maintain availability if the other cloud experiences an outage. A less costly alternative could be to use a public cloud as a cold standby where services are duplicated from the private cloud but are not turned on unless the private cloud has an outage.

## Hybrid Cloud Requirements

- Management integration
- Connectivity
- Service mobility
- Monitoring
- Security

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



3

Shown here are some of the requirements that will dictate which hybrid cloud capabilities should be included in the design.

## Management Integration

- Unified front-end
- Deployment and management capabilities in public cloud
  - Insight into capabilities
  - Control service instances
- Billing system integration
- Ability to enforce policies consistently



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

4

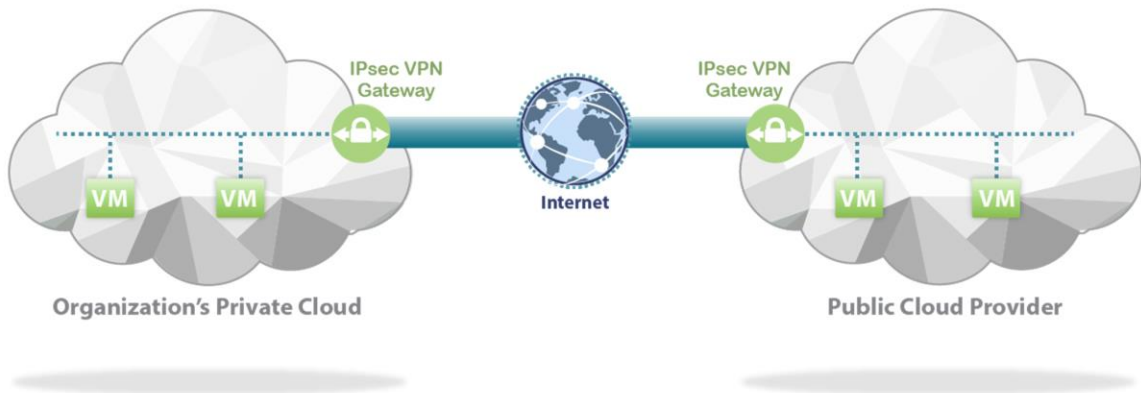
An essential part of adding hybrid capabilities to the cloud design is to select a cloud management tools that will integrate with the public cloud provider. One capability to consider is providing a unified front-end management interface that allows consumers to deploy and manage services in either private or public cloud resource pools. The front-end should provide insight into the public cloud provider capabilities where applicable. For example, it may be useful to know which availability zones exist for deploying distributed applications.

Billing system integration is also useful for consumers so that they can make decisions about the costs of deploying services in the private or public clouds.

Many cloud management platforms provide policy-driven deployment capabilities that support governance, compliance, and other business requirements. Selecting a solution that can enforce these policies across the private or public cloud resources pools is also very important. For instance if a policy defines that the data for a particular service must remain within the organization's datacenter, the management tools should enforce this policy by preventing a deployment into the public cloud resource pool.

Some same-vendor solutions may offer better integration or advanced functionality when creating hybrid cloud solutions. As an alternative, you could deploy a third party, unified cloud management solution that can provide a consolidated view of services and resources, consolidated billing, and can streamline processes.

## Inter-Cloud Connectivity: IPsec VPN



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

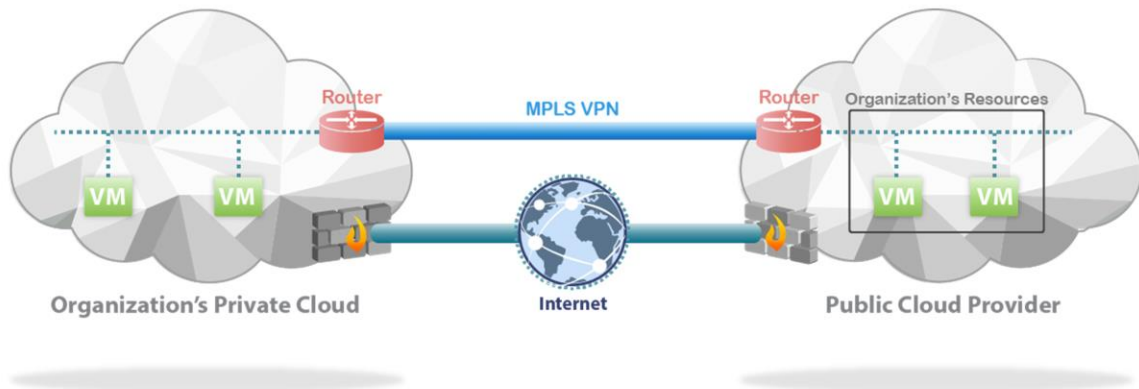


5

As described previously, an IPsec Virtual Private Network (VPN) is a networking technology that allows you to connect two networks via an Internet connection. With VPN tunneling technology, one endpoint device is placed on each network segment and a secure tunnel is created between them. This tunnel can connect the two segments using layer 2 protocols. The tunnel uses encryption to secure traffic but unlike a client server connection, the session is maintained and clients or servers on either end of the tunnel have access to the entire network segment on the other end.

This is one of the most common scenarios for extending an organization's on-premises network into the organization's network that has been deployed on a public cloud. The VPN gateway deployed in the on-premises network could be a physical or virtual appliance and may be a capability of a firewall appliance. The VPN gateway at the public cloud provider will most likely be a virtual appliance. This solution enables communication between services in both cloud environments.

## Inter-Cloud Connectivity: Direct Connect



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



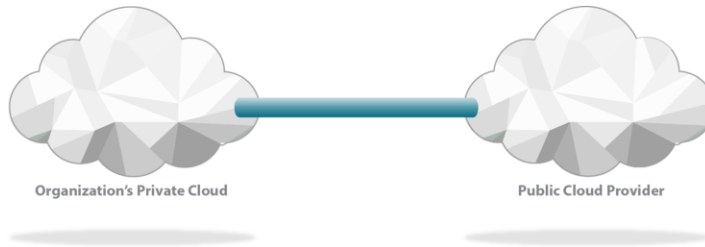
6

Some public cloud service providers offer organizations the option to directly connect their on-premises network to the public cloud provider. This option avoids performance and security issues that may be experienced using an Internet connection. The direct connections can provide 1Gb or 10Gb virtual private network connections between clouds using routers located at both ends of the connection. Some public cloud providers will not only provide secure connectivity to the organizations network segments but also to the SPs management interfaces, allowing internal users to access public-facing resources without traversing the internet.

## Discussion: Inter-Cloud Links



? What are general design considerations for inter-cloud links?



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



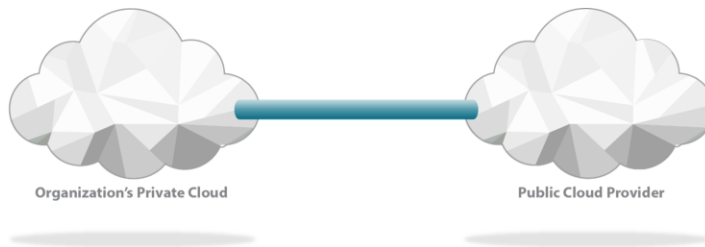
7

What are general design considerations for inter-cloud links?

## Discussion: Inter-Cloud Links Cont.



? What are general design considerations for inter-cloud links?



- Internet connection
  - Type
  - Redundancy
  - Bandwidth
- Network service provider
- Cloud provider
- Application
  - Minimize crosstalk
    - Internet connection points
    - CDN
  - Placement

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

The first step in designing an inter-cloud solution is to understand the network requirements for the services deployed in the public cloud. The assessment provides guidance in selecting the IPsec VPN solution or direct connection, and in determining the correct bandwidth. Although an organization may have a high-bandwidth connection to the network service provider, it has no control over what happens beyond that connection. A general increase in Internet traffic could impact performance through a IPsec VPN link. Any network service provider along the path to the public cloud provider could have an outage. To avoid this problem, the organization may wish to install multiple connections using different network service providers or at least multiple paths of entry into a single network service provider. Establishing inter-cloud links using as many redundant paths as possible reduces the possibility of an outage and increases bandwidth if the links are used in an active/active state.

Another solution for reducing negative network impacts is to choose a network service provider that also provides network connectivity for the public cloud provider. If this is impractical, then the organization may be able to request that a peering connection be established with the network service provider for the public cloud. Also, the public cloud provider will most likely have multiple data center locations with multiple network termination points. Selecting the location that is nearest to the organization's datacenter (from a network viewpoint), will also reduce the potential number problems experienced in the inter-cloud link.

If using a public cloud provider, some of the design decisions may be limited to what the provider supports. If the design is linking to private clouds, then this may also be the case if your private cloud is hosted by a provider.



## Applications



- Minimize crosstalk
  - Internet connection points
  - CDN
- Placement

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



9

Minimizing application crosstalk helps maintain performance across the inter-cloud link. One method of reducing the amount of data that must traverse the inter-cloud link, is to ensure that an Internet connection exists at both ends of the tunnel when public-facing services are deployed. If the Internet connection is at one end only, any public requests from services on the opposite end will traverse the tunnel. Also, for public-facing services, using an external content delivery network for static content can reduce traffic across the inter-cloud link.

Planning the proper service placement of services is also key to designing a hybrid cloud solution and ensuring performance. For example, placing backend database servers in the private cloud, and web frontend servers in the public cloud, may be a requirement to ensure data security or compliance. In this case, a high bandwidth, low latency solution may be required to maintain adequate performance. However, if compliance or data security are not an issue, then placing the backend database servers in the same cloud as the frontend servers will ensure performance and reduce load on the inter-cloud link. Placing support services such as DHCP, DNS, and authentication in both clouds will reduce bandwidth on the inter-cloud link and improve resiliency in the event one cloud environment experiences outages. However, it is important to realize that these support services may also require synchronization which may add load to the link.



## Service Mobility

- May not be possible without downtime
- Tools and procedures for import and export
- Hypervisor and virtual machine format compatibility
- Network
  - Bandwidth
  - Encryption



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



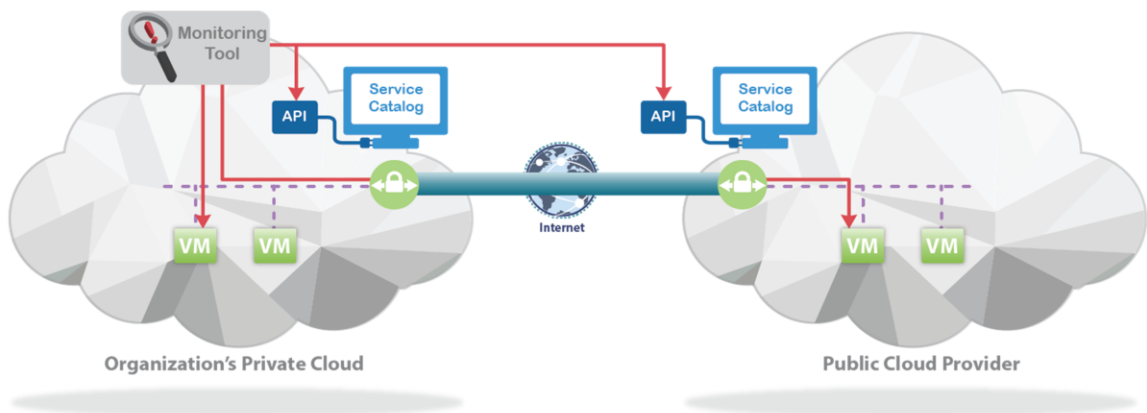
10

Many organizations think that cloud bursting means migrating services or instances from one cloud to another. This is not usually the case and organizations may find it very difficult to migrate a service. If the organization has requirements to move services between clouds, then the design will require certain options.

One thing that will certainly prevent a direct migration of a service is the hypervisor used on both clouds. Various hypervisors support the different virtual machine formats. As an example, if an organization uses a VMware hypervisor in their private cloud and the public cloud provider does not, then the only way to migrate would be to use tools to export and import the service between clouds. If hypervisors differ, then the cloud design should define the correct tool and process for moving a service. However, selecting a public cloud provider that supports a similar infrastructure and tool set as your private cloud can offer many advantages and should also be considered for the design.

The cloud design should also include network connections with enough bandwidth to support migrations. Some IaaS instances can be large and may take time to transfer across the network between clouds. This link should also be secured using network encryption so that no internal data will be compromised.

## Monitoring



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



11

Monitoring of public cloud resources can be accomplished in two different ways. The first is to use the internally deployed monitoring solution to monitor the individual services. In this case, the internal monitoring system will use the secured connection between clouds to collect statistics, logs, and events from the services. Include the additional bandwidth to support the information collection process in your design.

Additional monitoring capabilities may be provided by the public cloud provider through an external API. If this additional information is required, the monitoring solution will require the capability to connect to this API. This will include bandwidth, access through firewalls and the proper public cloud credentials for accessing the information.

## Security

- Compliance
  - Proper workload placement
- Data encryption
  - In-flight
  - At-rest

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

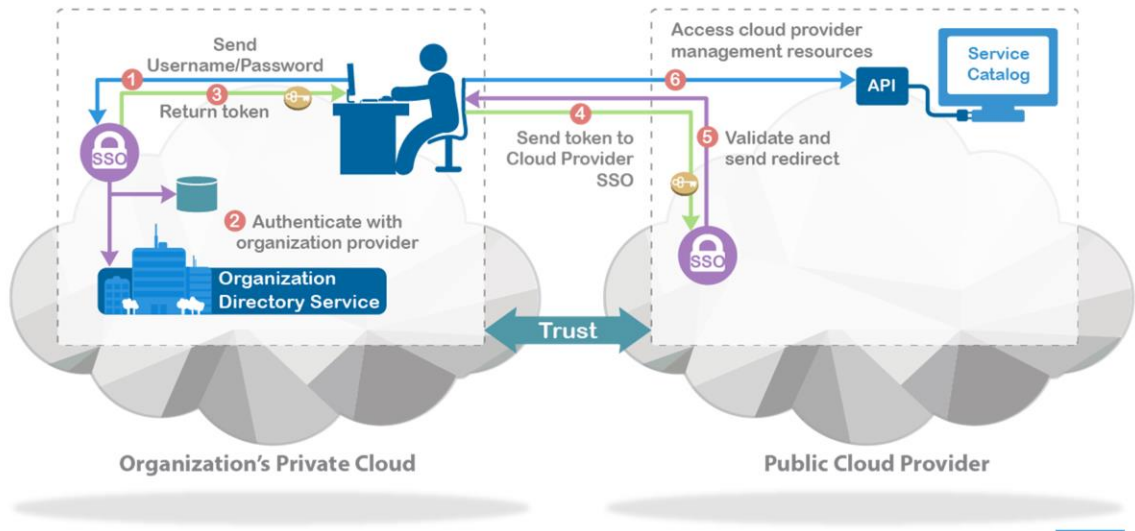


12

Enforcing compliance may be a requirement in the cloud design. When resources are available from both a private and public cloud provider, mechanisms may be needed to ensure that data is not placed in a noncompliant location. These mechanisms will most likely be embedded in the portal, catalog, and orchestration capabilities of the cloud. Workflows can be created that will prevent incorrect service placement based on a classification in the catalog, an orchestrated approval process or the consumer's role or permissions. It is also possible to build compliance into the application.

Encryption should be used for protecting data that is being transmitted across the network as well as stored in the public cloud. This will address requirements for privacy, regulatory compliance, and loss of control of media.

# Federated Identity Management



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

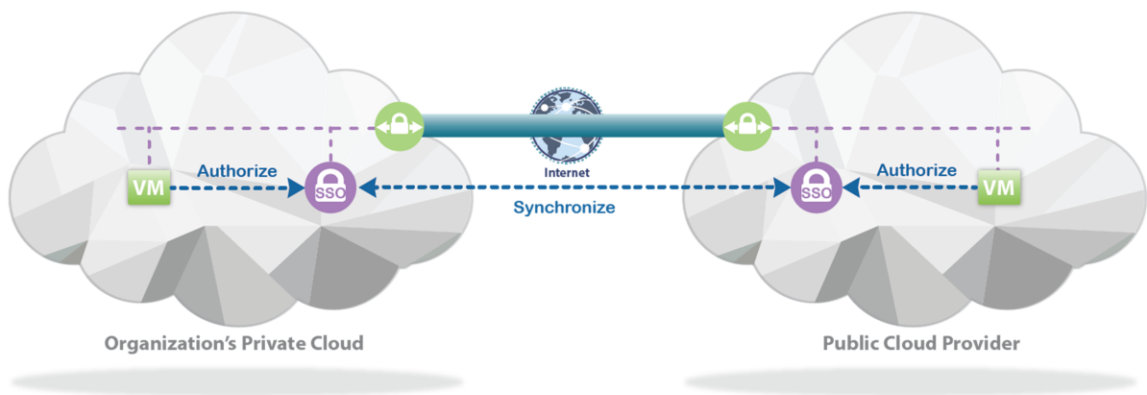


13

In a hybrid cloud model, one option for authenticating with the public cloud provider's management interface is to use the internal credentials given by the provider. However, consumers often prefer to log on to a system once and be authenticated automatically by services as they are accessed. This single-sign-on experience can be implemented using a federated identity management solution.

A federated solution uses a trust system between different entities, in which one entity authenticates a user and then shares the authentication information with other entities in the form of a token. The user provides a username and password to the first entity, and when the user tries to access a service in another entity, the token containing the user's credentials (no password) are passed on for authentication purposes. The non-authenticating entity must be configured to trust the authenticating entity and they must use the same identity protocol. Of course, the federated user must be given the proper privileges to use the service as well. Standard identity protocols include SAML, OpenID, WS-Trust, WS-Federation, and OAuth. The benefit to federation is security and authentication into both on-premises and cloud applications.

## Service Authentication



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



14

In addition to enabling authentication to the public cloud provider resources, it is also necessary to enable access to the services themselves. For instance, if deploying IaaS services in the public and private cloud, the organization will want consumers to authenticate to these services using the same identity provider. One way to accomplish this is to implement an inter-cloud link with enough bandwidth to allow public cloud consumers to authenticate with the identity provider located in their datacenter. This is not ideal because of increased bandwidth requirements and the possibility of a link failure.

Deploying instances of the identity provider service into the public cloud is a better solution. As an example, many organizations use Microsoft Active Directory as their identity provider. Deploying additional domain controllers from the organizations existing Active Directory Forest will enable a single-sign on capability across the hybrid cloud, improve logon performance, and minimize impact from a inter-cloud link failure. The inter-cloud link would only need to support the synchronization traffic between the domain controllers.

This design concept also applies to using different fault domains or datacenters in the private or public cloud. If services will be deployed in different fault domains or datacenters, then deploying identity provider instances in each domain will improve performance and help ensure availability if a datacenter experiences an outage.

## Lab 7: Hybrid Cloud Design

### Purpose:

Understand the impact of hybrid cloud capabilities on a cloud design.

### Tasks:

- Identify relevant hybrid cloud requirements for the NanCo cloud design.
- Amend the NanCo cloud design to highlight the impact of implementing hybrid cloud capabilities.

## Lab 7: Debrief

### Hybrid Cloud Design

**Purpose:**  
Understand the impact of hybrid cloud capabilities on a cloud design.

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



16

# Module 7

## Hybrid Cloud Capabilities Summary



Key points covered in this module:

- CMP component integration is necessary to support a hybrid cloud model.
- Inter-cloud links require additional security and bandwidth to support applications and dependencies.
- CMP and application security can be achieved using a federated identity management solution.

Cloud Infrastructure Planning and Design



17

This module covered requirements and design considerations that support hybrid cloud capabilities.





# Module 8

## Disaster Recovery



Upon completion of this module, you should be able to:

- Communicate the benefits, challenges, and considerations of implementing technologies that support disaster recovery in a cloud infrastructure.
- Align business requirements when designing disaster recovery capabilities.
- Develop a design and plan for disaster recovery capabilities in a cloud infrastructure.

Cloud Infrastructure Planning and Design



1

This module focuses on technologies and considerations used in designing for disaster recovery in a cloud.

## Cloud Disaster Recovery Overview

- Design applications for failure
  - Provide multiple fault domains
  - Multiple datacenters
  - Multiple cloud
- Provide backup or replication capabilities
  - Support more traditional use cases
  - Protects data from deletion and corruption
  - Provide a backup solution that stores data outside the fault domain



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



2

In a cloud environment, one way to reduce the impact of infrastructure failure is to enable your developers to design applications for failure. To support this, the cloud design must include multiple fault domains. These can be separate datacenters or even different cloud infrastructures. By providing multiple fault domains, developers can create applications that span the domains and by including services such as load balancers and replication, these applications will be available if a domain becomes experiences a failure.

Designing applications for failure is not always possible and backup capabilities may still be required. For instance, if the cloud provides virtual machine instances with off-the-shelf applications, consumers may wish to have backup capabilities for these instances. Also, even with applications designed for failure and infrastructure that is highly reliable, backups may still be required to protect data from deletion or corruption. When backups are required, it is always good practice to store the backup data in a location outside the fault domain.

## Backup Infrastructure Deployment

| Backup Deployment        | Description                                                                                |
|--------------------------|--------------------------------------------------------------------------------------------|
| Local                    | Backup managed and data stored at the local site                                           |
| Local with replication   | Backup managed locally. Data stored at the local site and also replicated to a second site |
| Local with cloud gateway | Backup managed locally. Backup data stored at offsite cloud storage using a gateway device |
| Remote                   | Backup managed and data stored at offsite cloud storage                                    |

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



3

Listed here are examples of backup deployment models.

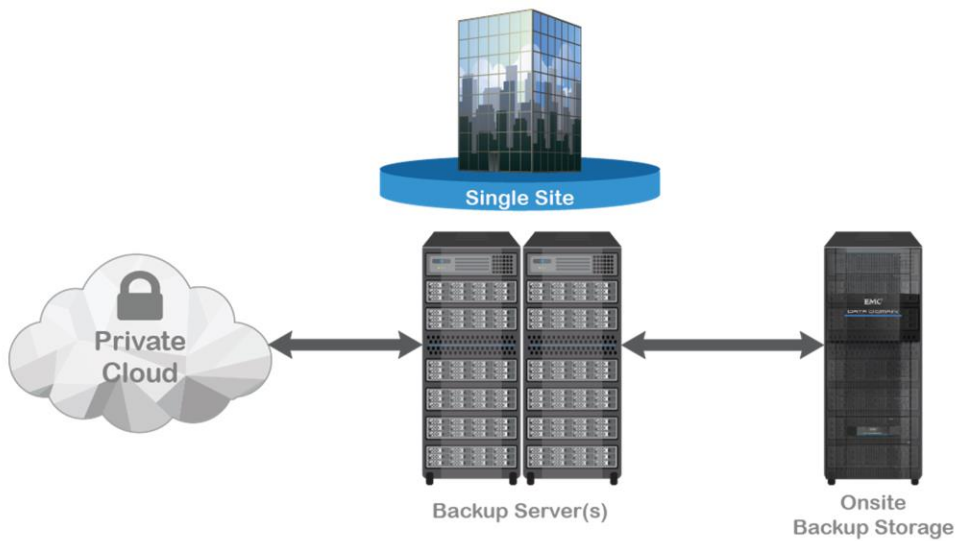
**Local** – In a local model, backups are performed at the local site and the data is also stored at the same site.

**Local with replication** – This model functions similarly to the local model except that not only is data stored locally but it is also replicated to the remote site.

**Remote with a cloud gateway** – In a remote solution using a cloud gateway, the backup is managed locally but data is directed through a local gateway device and stored in a remote cloud.

**Remote** – This solution is a Backup as a Service model where a remote provider provides the management and storage for backups.

## Local Deployment Overview



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



4

In a local model, backups are performed at the local site and the data is also stored at the same site.

## Local Deployment Considerations

- Provider retains control
- Integration with selected CMP
- Requires additional storage and network capacity
- Requires additional components throughout infrastructure
- Additional maintenance and management of backup infrastructure
- Additional costs to acquire backup infrastructure

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



5

In a local backup deployment model, the provider maintains control of the backup application, data, and backup infrastructure. When choosing a backup application to support this, it is important to find one that will integrate with CMP components. The cloud design will require additional storage and network capacity to support the additional load of the backups. Additional components may be necessary within both the cloud management infrastructure and consumer resource infrastructure to support backup. Since data is placed in the same site as the cloud, security requirements can be enforced using some of the techniques used to address security throughout the cloud. Once the backup infrastructure is designed and implemented, the cloud provider will need staff to maintain and manage the infrastructure.

## Additional Local Deployment Considerations



- Which backup or restore types will be performed?
  - Image: requires proxy server deployments
  - Agent: can restore directly to VM
- Use multiple backup servers for HA
- Deploy enough proxy servers to handle (load balance) the backup and restore load traffic for image-based backups
- Use a separate backup network so that backup traffic does not impact production traffic

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

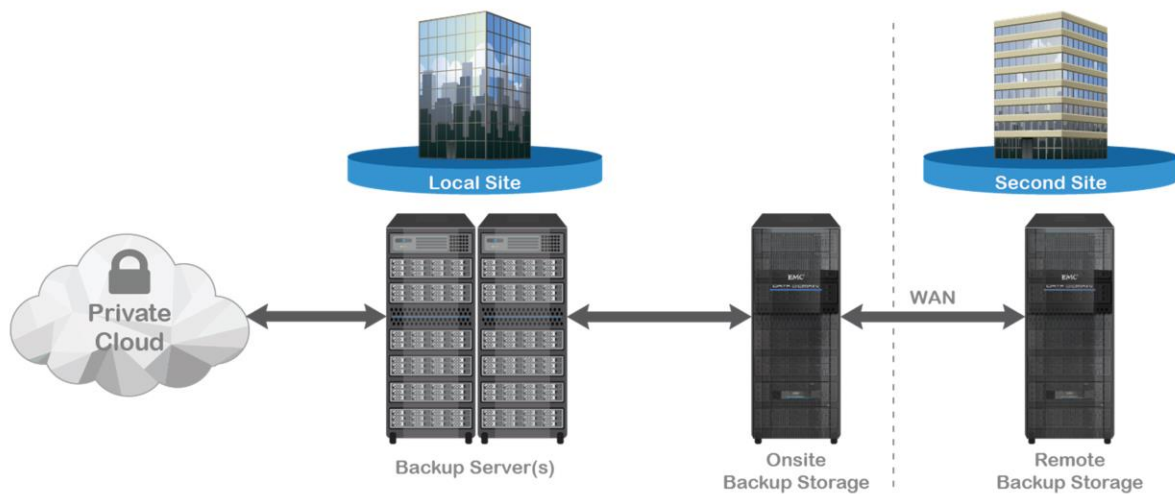


6

One decision that will impact your design is which type of backup to implement: image or agent. Agent-based backups require overhead on individual VMs where image-based backups require the deployment of proxy servers within the infrastructure. In addition, they each have different requirements for how backend storage is accessed.

Deploy multiple backup servers to maintain high availability of the management interfaces for the backup infrastructure. If image-based backups are used, deploy sufficient backup nodes and proxies to support the current and expected VM workload. Use a separate network for backup traffic to enhance security and performance.

## Local Deployment with Replication Overview



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



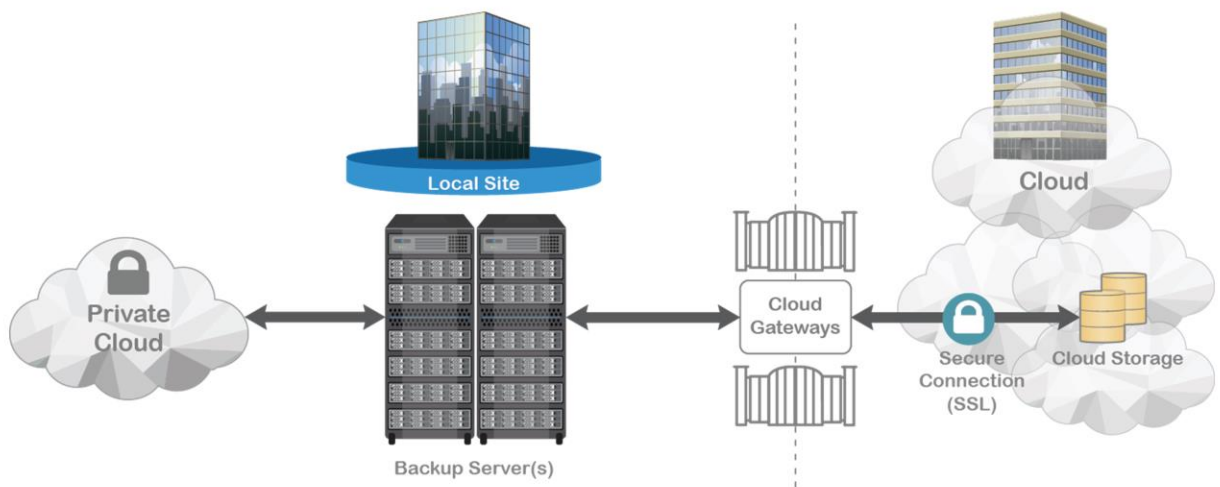
7

The local with replication deployment functions similarly to the local model except that not only is data stored locally but it is also replicated to the remote site.

This design has additional considerations. For instance, additional bandwidth is required to replicate the data offsite. Encryption may be necessary to secure the replicated data both in-flight and at-rest. Replication may be synchronous or asynchronous, and the settings for this will be driven by the organization's requirements around recovery point objectives (RPO).



## Local with Cloud Gateway Deployment Overview



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



8

In a remote with cloud gateway model the cloud is backed up by onsite backup servers. These servers direct the backup data to cloud storage through one or more cloud gateways that are deployed at the local site. Gateways normally do not include any backup software and in most cases appear as a storage location which the customer's backup software can write to. Backup servers never communicate with the cloud storage interfaces directly and must go through the gateway, providing a translation capability between local protocols and object storage protocols in the cloud.

## Local with Cloud Gateway Design Considerations



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



9

**Seeding** - One common issue in using remote backups is how the initial backup is delivered to the remote cloud. If the 1<sup>st</sup> backup is sent across the WAN, it could take hours, days, or even weeks to fully complete the backup, depending upon the data set size. An alternative way to populate the cloud storage may be possible to minimize bandwidth consumption.

**Recovery** - Once the data is stored in the cloud, it is stored as objects. To recover this data in the event of a site disaster, your design will require policies and procedures to build a recovery platform that will be able to translate the object-based storage back into a format that is recognizable and accessible to the backup application.

**Segregation** - Some type of logical data separation may be necessary to maintain the multi-tenancy support.

**Local Data Protection Laws** - If a cloud storage provider is used, the design must still address the organization's data protection laws or cross-border compliance laws requirements.

**Availability** - Using multiple cloud gateways allows for multiple redundant paths to the cloud backup. It is best practice to use multiple cloud gateways for continued access to be able to create and send backups to the cloud in the face of a cloud gateway failure.

## Optimizing Performance Cloud Gateway

- On cloud gateway use
  - More local disk capacity to reduce latency when storing or fetching remote data
  - Faster disks to enable faster delivery of backup data
  - More CPU and memory for faster execution for deduplication, compression, encryption, and WAN optimization
- Between cloud gateway and cloud use
  - More bandwidth between gateway and cloud to reduce RTO
  - Compression to reduce amount of data stored on cloud gateway and amount of data that must be transferred to the cloud
  - WAN optimization to reduce the amount of bandwidth required between the gateway and the cloud

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



10

When determining the performance requirements of the cloud gateway, it is important to identify the maximum allowable backup window as this will influence the requirements. This is also influenced by the cloud gateway hardware. Increasing the performance between the gateway and the cloud helps meet desired RTO and RPO objectives.

## Cloud Gateway Security Considerations

- What data must be encrypted?
  - All data
  - Or only private and sensitive data
- Is in-transit encryption used?
- Is at-rest encryption used?
- Who is in charge of generating encryption keys and where are they stored?
- Are encryption keys rotated?

© Copyright 2015 EMC Corporation. All rights reserved.

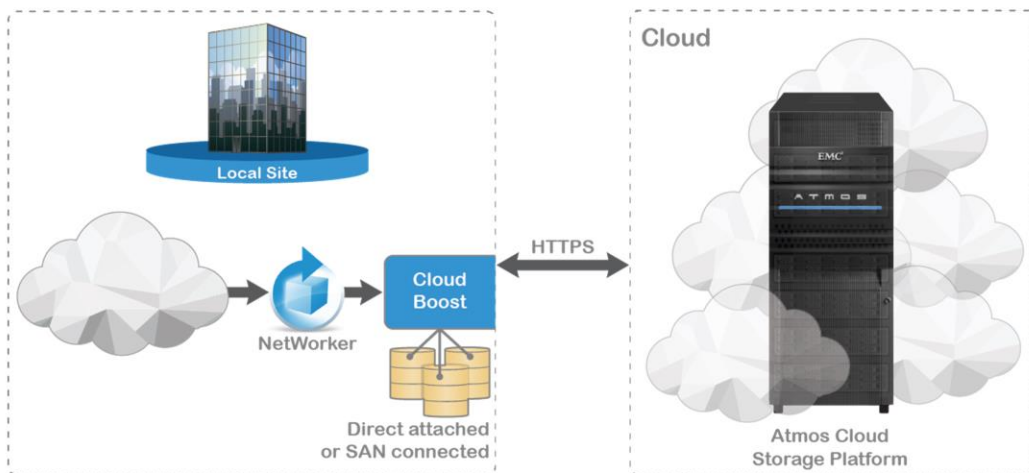
Cloud Infrastructure Planning and Design



11

When sending backup data to the cloud storage from the local site, it is important to understand the data security requirements so that you can identify the level of encryption needed for backups. Encryption can occur in-flight or at rest. If it is deemed that data must be encrypted, is it equally important to know who is responsible for generating encryption keys to be used to encrypt the data?

## Cloud Gateway Example: NetWorker and CloudBoost



© Copyright 2015 EMC Corporation. All rights reserved.

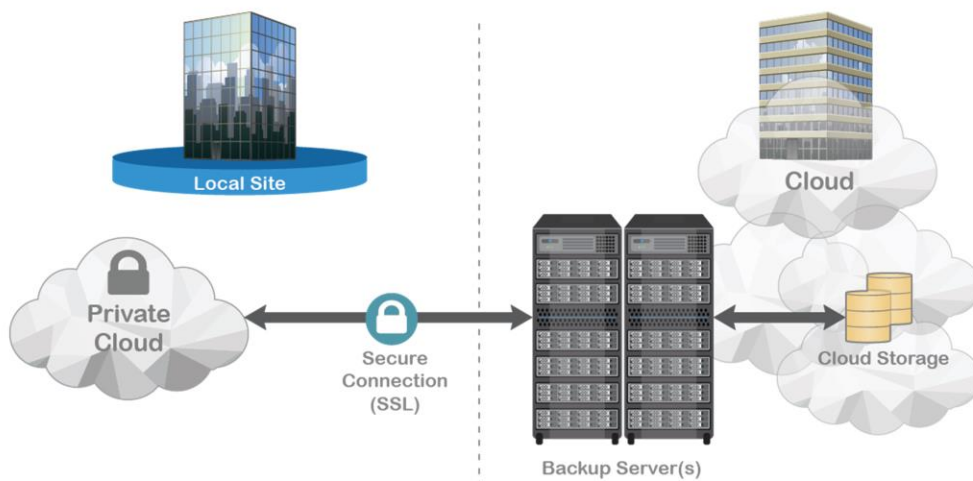
Cloud Infrastructure Planning and Design



12

NetWorker with CloudBoost enables long term storage provisioning via the cloud. NetWorker sends a backup clone to the CloudBoost virtual appliance. The CloudBoost virtual appliance translates these into generic objects which are sent to an object store, which can be public, private, or hybrid. The CloudBoost virtual appliance presents itself as a NetWorker Advanced File Type Device. The enabled workflow is a clone operation to the cloud; it is not a backup to the cloud. With this low cost tape replacement solution, each CloudBoost virtual appliance can support up to 400 TB of addressable back end storage.

## Remote Deployment



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



13

This solution is an outsourced model where a remote provider provides the management and storage for backups. Data is backed up directly to a remote site from the private cloud. Agents must be installed on the services requiring backup services. This model can support both file, image, and application level backups. Some remote backup providers support writing backups to both local storage cache as well as cloud storage. Encryption is normally configured between the device being backed up and the cloud.

## Backup Types

- Agent-based
  - Targets application or service data
  - Required in non-virtualized environment
  - May require additional OS or software installation during recovery
- Image-based
  - Captures entire virtual machine
  - Full OS, application, and data recovery of virtual machine

© Copyright 2015 EMC Corporation. All rights reserved.

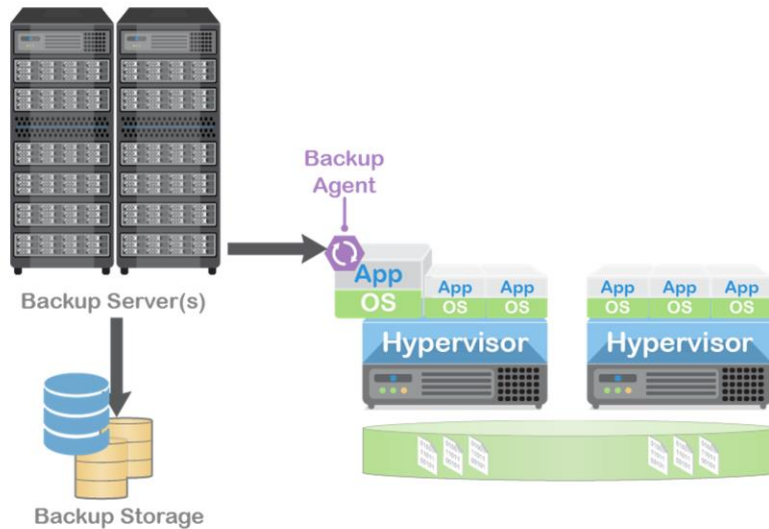
Cloud Infrastructure Planning and Design



14

Listed here are types of backups that are used in a cloud.

## Agent-Based Backups Overview



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design

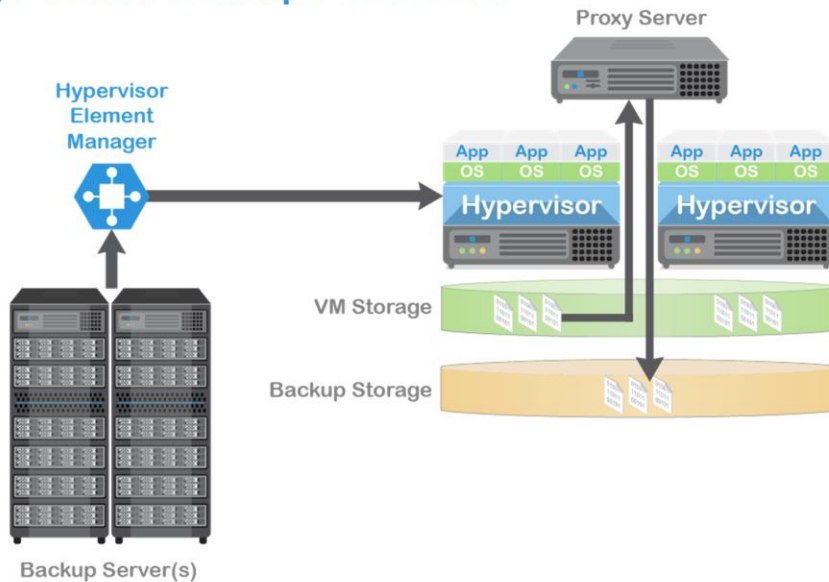


15

With agent-based backups, an agent is installed on the virtual machine. The backup server and agent communicate with each other and data is transferred to the backup server. From there it is stored in the backup storage device.



## VM Image-Based Backups Overview



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



16

With image-based backups, the backup server works with the hypervisor element manager and a proxy server. The proxy server may be in the form of a virtual machine appliance. When a virtual machine (VM) is scheduled to be backed up, the backup server contacts the hypervisor element manager to create a snapshot of the VM. After this, the backup server directs the proxy server to backup the VM files to the backend storage. Once the backup is complete, the snapshot is released.

## CMP Backup

- Traditional backup processes and procedures still apply
- Challenges
  - Potentially large number of components
  - Component integration
  - Consistency across CMP
  - Quiescing
  - Freezing data



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



17

Many of the traditional procedures and best practices for backups still apply when dealing with the components of the cloud management infrastructure. The number and the extensive integration of these components may present challenges when performing backups. In an ideal situation you will want to maintain a consistently timed backup across all of the components in the management infrastructure.

However, this may not be practical from an implementation standpoint for various reasons. For one, it may be costly to implement a solution that can freeze and backup every component, all at the same time. Secondly, some components, such as databases, need to be quiesced in order to flush data to permanent storage to provide an accurate backup, and this process can take time. Finally, since it takes time to perform a backup, if there is no method to freeze the state of a components such as with using snapshot capability, then all of the components would need to be brought offline to perform a consistent backup. This is not practical since it would cause an outage for the entire cloud.

## CMP Backup Considerations



- Follow vendor or community recommendations
  - Order or component backup
  - Identify backup dependencies
  - Identify specific data for backups
  - Recommend methods for backup
- Use image snapshots
  - Supports higher component uptime
  - Integrated quiescing
- Identify times of least activity
- Design deliverables

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



18

There are some things that can be done to ensure a successful backup of the cloud management infrastructure. First, you should follow the guidelines or best practices provided by the CMP vendors or communities. They may provide suggestions for a specific order of backups or may identify dependencies where certain components that should be backed up together. Vendors may also help to identify which data should be backed up and possibly suggest the best method for the backup.

If cloud management services are running in a virtual environment, use image-based snapshot capabilities to freeze data and also allow components to remain active during the backup window. This does not alleviate a requirement for quiescing services but some snapshot mechanisms may have integrated quiescing capabilities.

Finally, as has always been a good practice, try to identify a time when activity is light since this will reduce the amount of data that must be flushed to permanent storage. However, some clouds may support consumer activity all over the globe and a single cloud instance may few inactive times.

The completed cloud design should include all processes and procedures that are required to restore CMP functionality. In addition, the design should include a validation and verification process to confirm a successful recovery.

## Additional CMP DR Option

- Replication
  - Does not guarantee protection for data deletion or corruption
  - Use secondary site for backup source
  - Use secondary site for recovery

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



19

One way to enhance recoverability of the cloud management platform is to replicate it to another site. This site could be another datacenter or another cloud. Replication protects a site from a disaster that makes it unavailable but does not protect against data deletion or corruption. Backups are still necessary to protect against this. However, replication can also aid you with backups, since the secondary site can be used for the source of backups which can help to maintain CMP availability during quiesce and snapshot activities. Using the replicated CMP for recovery is possible, but the cloud design may require modifications to support this as well as documentation describing the processes and procedures that are involved.

## Service Backup Overview



- Backup as a Service
  - Enables backups of cloud services
- Why?
  - Consumer control and accountability
  - Provides a flexible and on-demand backup infrastructure
  - Reduces costs through sharing of resources
  - Standardizes IT processes

© Copyright 2015 EMC Corporation. All rights reserved.

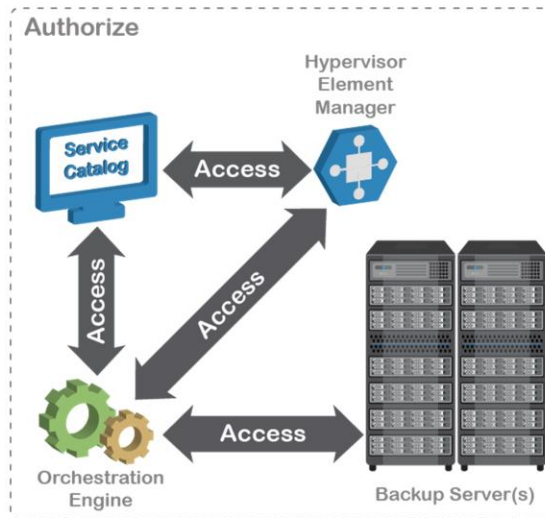
Cloud Infrastructure Planning and Design



20

Backup as a Service (BaaS) is a service that allows cloud providers the ability to offer backups to their consumers. BaaS places control of backups into the hands of consumers who become accountable for the data that is backed up and the storage required for the backups. Offering BaaS also streamlines and standardizes the process for backups in a private cloud.

## BaaS Requirements: Catalog and Orchestrator



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



21

To support Backup as a Service, the service catalog and orchestration engine requires access to the hypervisor element manager so that it can deploy and manipulate VMs and other resources. They will also require access to the backup application. The service catalog will require customization to support the selection and control of backups by users. Workflows will need to be created to support the creation of backup schedules, addition of backup targets, execution of backups, and so on. The cloud design will need to define the connection points for all of these components as well as the accounts and privileges required to execute all of the tasks involved. Some of these accounts may not be the same as the accounts used by consumers within the cloud and should therefore be secured in accordance to the organizations requirements and policies.

## Recovering VMs



### Catalog

- Provided restore choices



### Orchestration

- Delete existing VM if present
- Restore VM from image
- Power on VM
- Update CMP components

© Copyright 2015 EMC Corporation. All rights reserved.

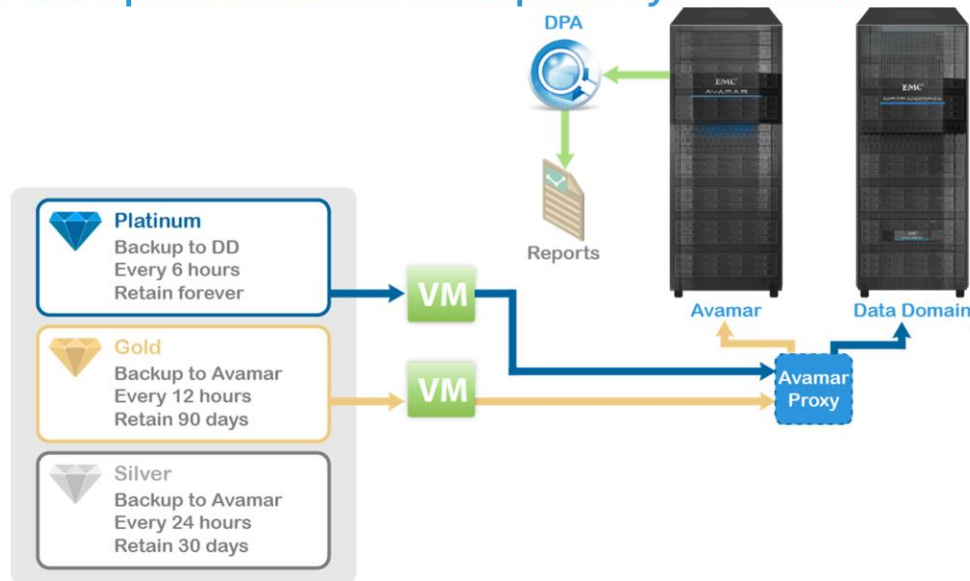
Cloud Infrastructure Planning and Design



22

If a tenant VM needs to be restored for any reason, the tenant will browse the catalog and select a previous backup to used for a restore. Once a restore point is selected, orchestration workflows are activated that will first shutdown the VM using the hypervisor element manager. The next steps in the workflow may be to delete the VM, signal the backup server to execute a restore, and then power on the restored VM. If the backup application is completely integrated with the CMP, this may be the end of the process. However, additional workflows may be required to properly update CMP information to reflect the change.

## BaaS Example: Federation Enterprise Hybrid Cloud



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



23

Avamar is the data protection management and scheduling engine in the EMC Enterprise Hybrid Cloud. It manages the backup policies, retention policies, and backup schedules. Optionally, Data Domain can be integrated with Avamar to provide larger-scale backup storage and target-based deduplication.

Backup policies are managed by delegated individuals using the service catalog. This means that authorized individuals can configure the appropriate policies and schedules from within the self-service portal, and then all consumers can choose those policies when they provision virtual machines.

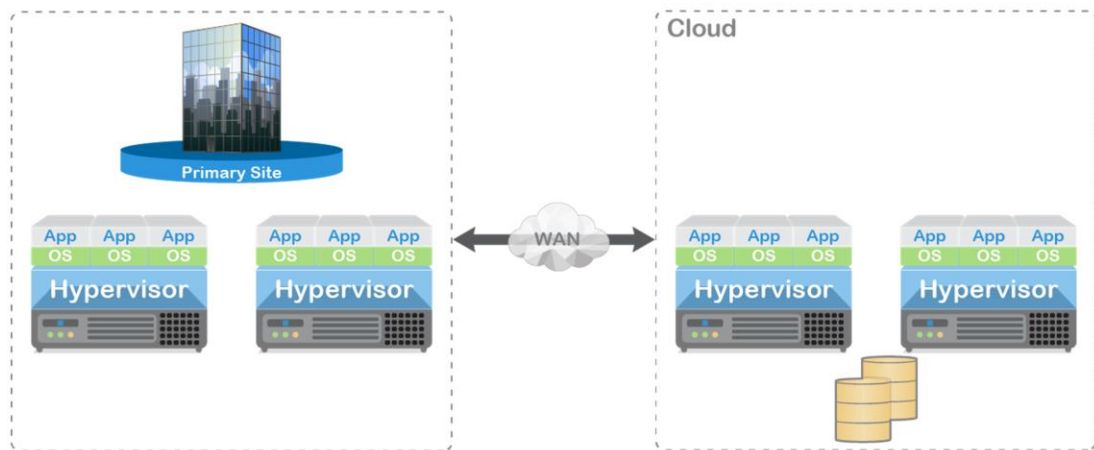
Data Protection Advisor (DPA) provides chargeback reporting capabilities for backup and restore activities.

Within the FEHC Solution, BaaS is implemented in the following way:

- Backup policy and backup schedule must be created on the Avamar backup server
- Folders are created in vCenter
- Backup policies and schedules are associated with the folders
- vCenter and Proxy servers are registered with the Avamar backup server
- vRealize Orchestration workflows are created to manage and execute backups
- vRealize Automation is enhanced with additional BaaS options and capabilities
- Data Domain can be added for storing backup data



## Additional Service DR Option



© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



24

One way to enhance recoverability of the cloud services is to replicate services and data to another site. This site could be another datacenter or another cloud. Replication will protect a site from a disaster that makes it unavailable but will not protect against data deletion or corruption. Backups are still necessary to protect against this. Replicating applications and virtual machine instances to another site will require that infrastructure exists in the secondary site to activate these services when a disaster arises. The cloud design will require modifications to support this as well documentation describing the processes and procedures that are involved.

## Backups for DR: Pros and Cons

- Pros
  - Cost effective
- Cons
  - Higher RTO than other solutions
  - RTO depends on time to bring up DR infrastructure and restore backups
- Considerations
  - Use automation tools to improve RTO (Restore, load, and start VMs)
  - Test DR plan frequently

© Copyright 2015 EMC Corporation. All rights reserved.

Cloud Infrastructure Planning and Design



25

Backups are a cost effective solution that requires only backup storage and software. However, the organization's requirements may dictate a recovery time objective (RTO) that cannot be met by just using backups. During a disaster, the backup data must be transferred to the DR site, the DR infrastructure must be brought up, the environment must be restored using the backup data, and the infrastructure may need to be modified to use the disaster recovery environment. This process can take hours, days, or even weeks to complete depending upon the amount of backup data. Tools can be used to automate some of the process but ultimately it still takes time to restore data. A comprehensive plan must be created and periodically tested to ensure recovery from a disaster.

## Lab 8: Service Backups

### Purpose:

Develop a design and plan for deploying resources that support backing up services in the cloud.

### Tasks:

- Identify service backup requirements that are relevant to NanCo cloud design.
- Provide a high level overview slide summarizing the required design changes for supporting service backups.

## Lab 8: Debrief

### Service Backups

**Purpose:**

Develop a design and plan for deploying resources that support backing up services in the cloud.

# Module 8

## Disaster Recovery Summary



Key points covered in this module:

- Designing applications for failure should be the main goal.
- Offering Backup as a Service fulfills requirements for protecting data and more traditional applications.
- Backups can be stored locally and in the cloud.
- Backing up the CMP components presents challenges and requires careful planning.



Cloud Infrastructure Planning and Design

28

This module covered the options and considerations for disaster recovery in the cloud.

# Cloud Infrastructure Planning and Design Course Summary



## Key points covered in this course:

- A successful cloud design ensures that all requirements, constraints, and goals are addressed in the deliverables.
- Because the security, availability, and performance requirements are dissimilar, cloud management components exist on different infrastructure than consumer resources.
- Cloud infrastructure designs support elasticity, self-service, multi-tenancy and hybrid cloud connectivity, in addition to traditional infrastructure design concepts.
- Cloud-native applications introduce different traffic flows and availability requirements for cloud infrastructure.



This course covered the design process for cloud infrastructure, and the characteristics, requirements, and technologies that influence the design decisions.



# Module 9

## Planning and Design Lab Presentation



Upon completion of this module, you should be able to:

- Align requirements from the NanCo case study to design decisions
- Create a high level presentation of a NanCo cloud design.
- Explain reasons for cloud design decisions

Cloud Infrastructure Planning and Design



1

This module focuses on the presentation of a high level cloud design that meets the requirements of the NanCo case study.



## Lab 9

### Final Lab Design Presentation

**Purpose:**

The purpose of this lab is to summarize your design proposal for NanCo.

**Tasks:**

- Prepare a final presentation
- Present your team's design
- Articulate all aspects of your design in a 20 minute or less presentation.

# Module 9

## Planning and Design Lab Presentation



Key points covered in this module:

- Understanding requirements is a critical part to cloud design.
- A cloud infrastructure contains an integrated collection of components that are required to deliver services.
- A cloud infrastructure design requires knowledge of many different technologies used in building IT infrastructure.



Cloud Infrastructure Planning and Design

3

This module covered a presentation of a high level cloud design that meets the requirements of the NanCo case study.

