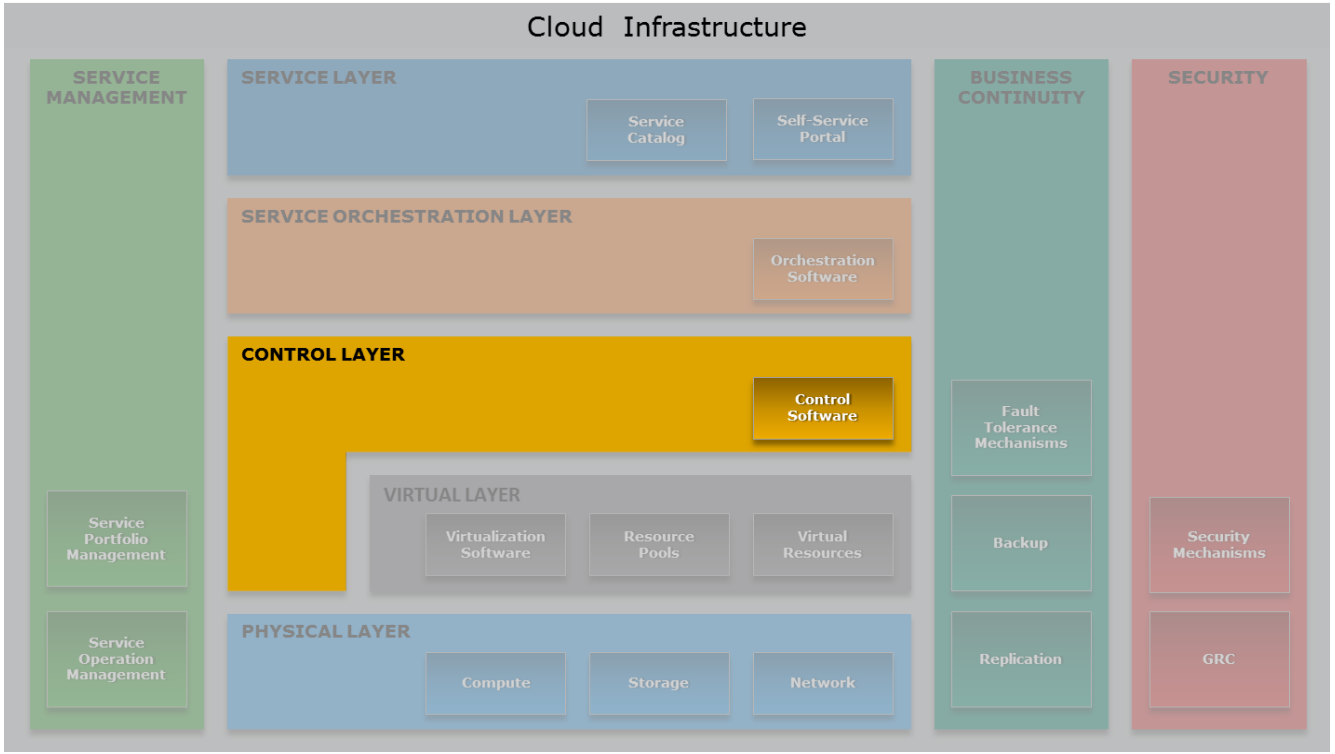
Module: Control Layer

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

- Describe the control layer and its key functions
- Describe control software and its types
- Describe the software-defined approach for managing IT resources
- Describe the key resource management techniques

Cloud Computing Reference Model

Control Layer



Lesson: Control Layer Overview

This lesson covers the following topics:

- Control layer and its functions
- Control software and its types
- Key phases for provisioning resources using unified manager

Introduction to Control Layer

Control Layer

Includes software tools that are responsible for managing and controlling the underlying cloud infrastructure and enables provisioning of IT resources for creating cloud services.

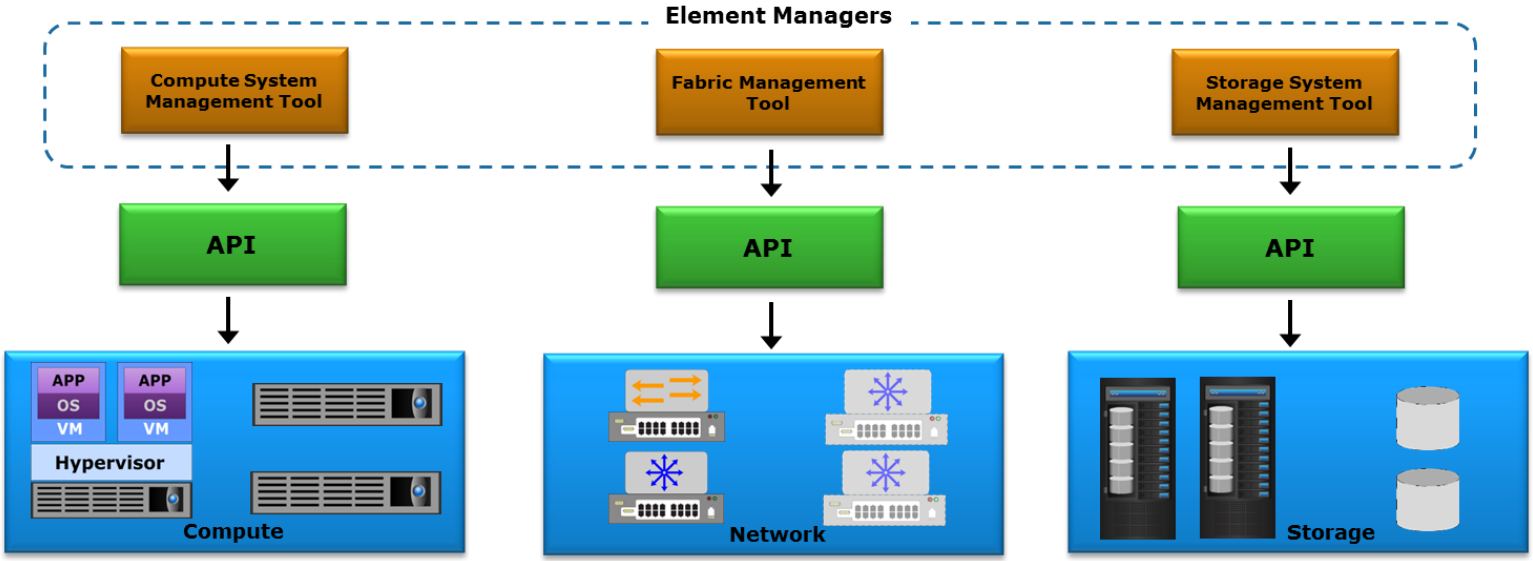
- Control layer can be deployed on top of the virtual layer or on top of the physical layer
- Receives request from the service and orchestration layers
  - Provisions the required resources to fulfill the service request
- Key functions of the control layer are resource configuration, resource provisioning, and monitoring resources

Control Software

- Ties together the underlying resources and works in conjunction with virtualization software to enable
  - Resource pooling
  - Dynamic allocation of resources for services
  - Optimizing utilization of resources
- Provides a complete view of all the resources in the cloud environment
  - Enables to centralize management of IT resources
- Two types of control software:
  - Element manager
  - Unified manager

Element Manager

- Infrastructure component vendors may provide the element managers as built-in or external software
- Required to manage infrastructure components independently

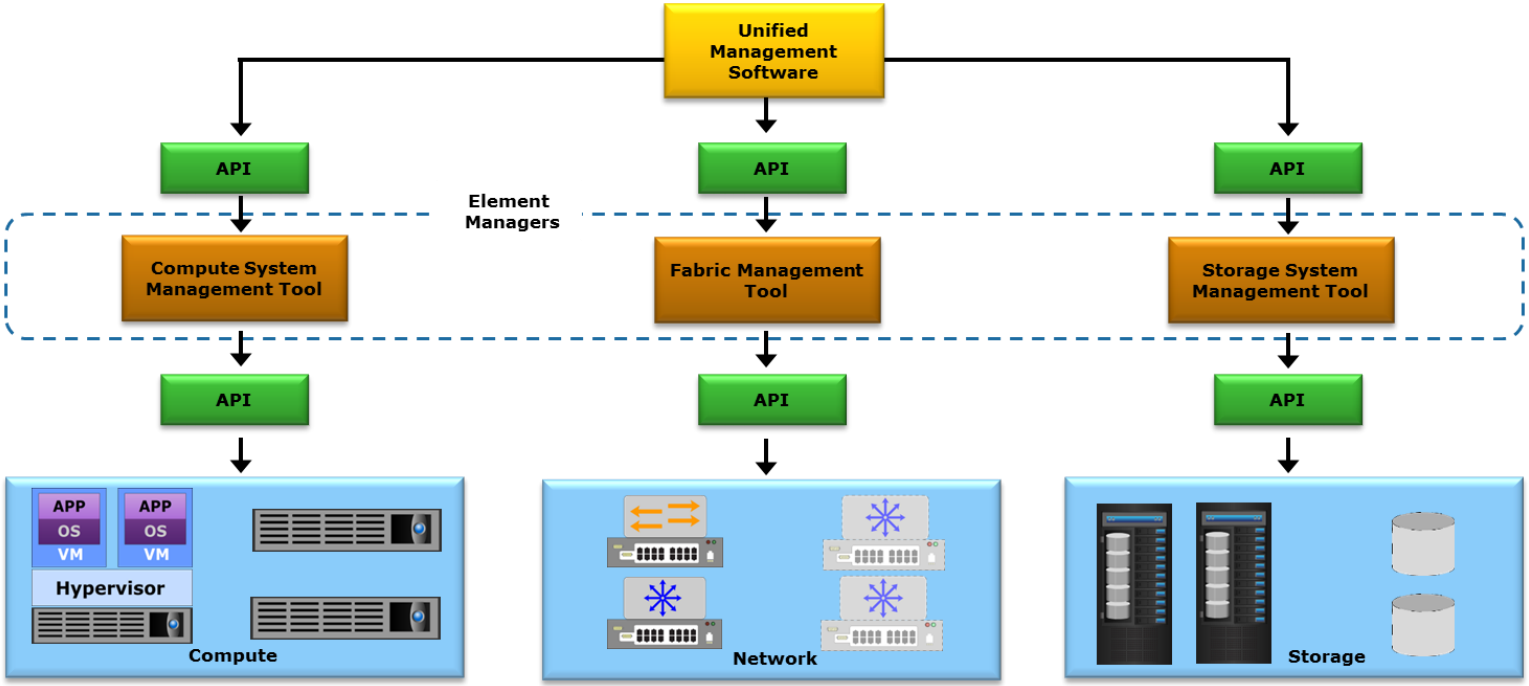


Key Tasks Performed by Element Manager

- Enables to perform initial component configurations and allows to modify it
  - Installing guest OS, configuring zoning, security settings, VLANs, RAID, and LUN masking
- Allows to expand resource capacity
  - Detects the newly added resources and adds them to an existing pool
- Enables to identify the problem and performs troubleshooting
- Monitors the infrastructure component for performance, availability, capacity, and security

Unified Manager

- Provides a single management interface for configuring and provisioning resources for applications and services



Unified Manager (Cont'd)

- Exposes APIs that can be integrated with the orchestration layer to automate service provisioning
- Enables adding or removing infrastructure resources to an already provisioned service
- Performs compliance check during resource configuration
- Provides a dashboard showing resource configurations and utilization
  - Allows administrator to perform monitoring, reporting, and root cause analysis

Key Phases for Provisioning Resources

Resource Discovery

- Enables unified manager to learn about resources that are available for service deployment
  - Provides visibility to each resource
  - Enables to manage cloud infrastructure resources centrally

Compute Systems	Network Components	Storage Systems
• Number of blade servers • Slot location • Blade model • CPU speed, memory capacity, CPU, and memory pools • Physical-to-virtual compute mapping	• Switch model • Network adapters • VLAN IDs • VSAN IDs • Physical-to-virtual network mapping • QoS • Zones	• Type of storage system, drive type • Total capacity, free capacity, used capacity • RAID level, storage pools • Physical-to-virtual storage mapping

Resource Pool Management – Grading Pools

- Unified manager allows to grade pools
  - Categorizes resources and identity pools based on predefined criteria
  - Helps creating variety of services, providing choices to consumers
- Multiple grade levels (e.g. 'Gold', 'Silver', 'Bronze') may be defined for each type of pool
- Costs/prices of resource pools differ depending on grade level

Grading Storage Pools

Grade 'Gold': Includes Flash, FC, and SATA drives, supports automated storage tiering, capacity 3 TB (Flash 1TB, FC 1TB, SATA 1TB), and RAID level 5

Grade 'Silver': Includes Flash, FC, and SATA drives, supports automated storage tiering, capacity 3 TB (Flash 0.5TB, FC 1TB, SATA 1.5TB), and RAID level 1+0

Grade 'Bronze': Includes FC drives, capacity 2TB, RAID level 5, and does not support automated storage tiering

Resource Provisioning

Resource Provisioning

Involves allocating resources from graded resource pools to the service instances.

- Provisioning commences when consumers select cloud services from the service catalog
- A service template defined in a service catalog facilitates consumers to understand service capabilities
  - Resources are allocated and configured as per service template to create an instance of a service

Lesson Summary

During this lesson the following topics were covered:

- Control layer and its functions
- Control software and its types
- Key phases for provisioning resources using unified manager

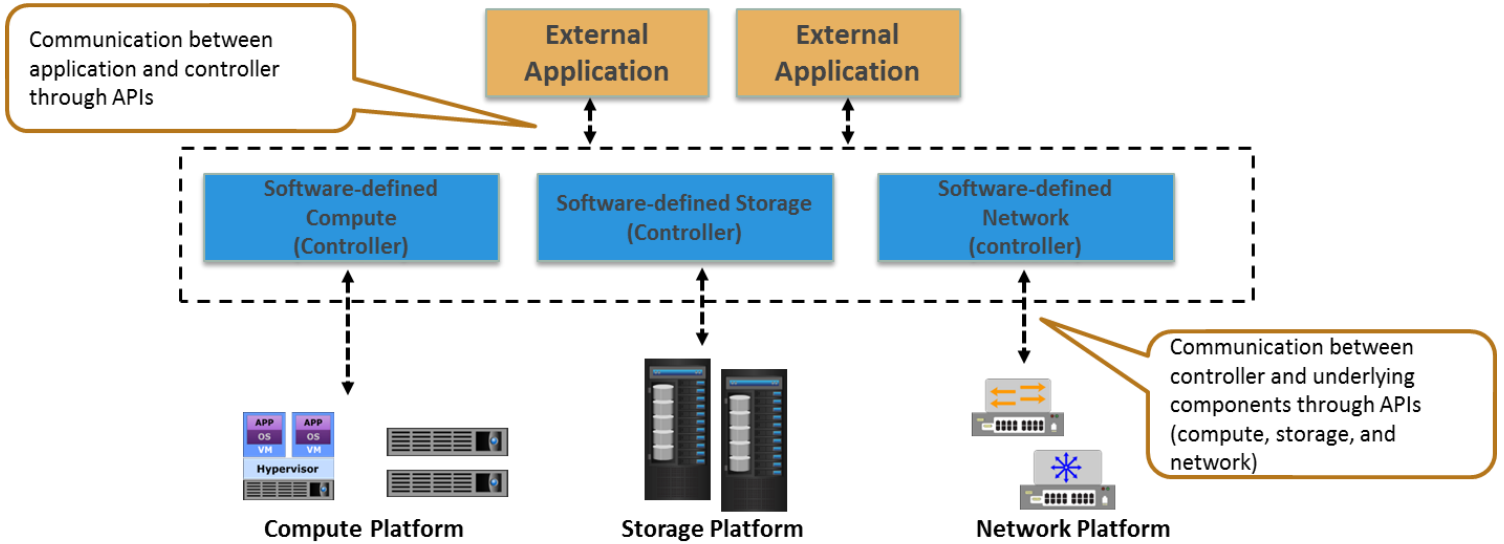
Lesson: Software-defined Approach

This lesson covers the following topics:

- Introduction to software-defined approach
- Key functions of software-defined controller
- Benefits of software-defined approach

Software-defined Approach – A New Model For Managing Resources

- Abstracts the underlying infrastructure components
- Separates the management functions from the infrastructure components to the external software that runs on a controller
  - Enables controlling IT infrastructures centrally



Key Functions of Software-defined Controller

- Discovers underlying resources and provides an aggregated view of resources
  - Abstracts the underlying hardware resources (compute, storage, and network) and pools them
- Enables the rapid provisioning of resources based on pre-defined policies
- Enables to apply policies uniformly across the infrastructure components, all from a software interface
- Provides interfaces that enable applications external to the controller to request resources and access them as services

Benefits of Software-defined Approach

- Improves business agility
  - Minimizes resource provisioning time to get new services up and running
- Provides cost efficiency
  - Enables to effectively use the existing infrastructure and low-cost commodity hardware to lower CAPEX
- Enables to achieve scale-out architecture
- Provides a central point of access to all management functions
- Allows to create new innovative services using the underlying resources

Lesson Summary

During this lesson the following topics were covered:

- Functions of software-defined controller
- Benefits of software-defined approach

Lesson: Resource Management Technique – 1

This lesson covers the following topics:

- Resource management aspect of cloud infrastructure
- Resource allocation model
- Key compute resource management techniques

Introduction to Resource Management

Resource Management

Process of allocating resources effectively to a service instance from a pool of resources and monitoring the resources that help in maintaining service levels.

- Key goals of resource management
  - Controls utilization of resources
  - Prevents service instances from monopolizing resources
- Management server is used to centrally manage the resources
  - Enables defining policies
  - Configures and monitors the resources
  - Provides the ability to pool the resources

Resource Allocation Models

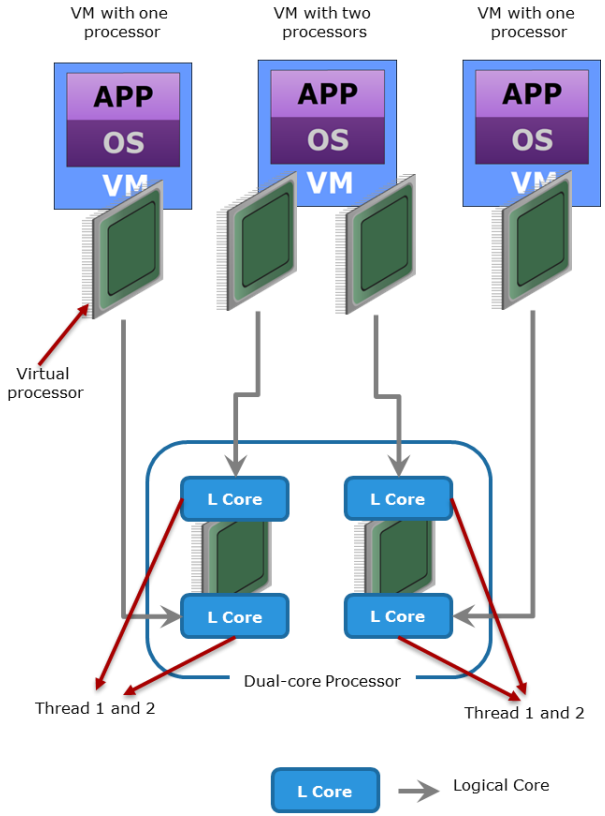
- Relative resource allocation
  - Resource allocation to a service instance is defined proportionally relative to the resource allocated to other service instances
- Absolute resource allocation
  - Resource allocation for a service instance is based on defining a quantitative bound
    - Lower bound guarantees minimum amount of resources
    - Upper bound limits a service instance from consuming resources beyond the defined maximum level

Key Resource Management Techniques

Compute	Storage	Network
• Hyper-threading • Memory page sharing • Dynamic memory allocation • VM load balancing across hypervisors • Server flash-cache	• Virtual storage provisioning • Storage pool rebalancing • Storage space reclamation • Automated storage tiering • Cache tiering • Dynamic VM load balancing across storage volumes	• Balancing client workload across nodes • Network storm control • Quality of Service (QoS) • Traffic shaping • Link aggregation • NIC teaming • Multipathing

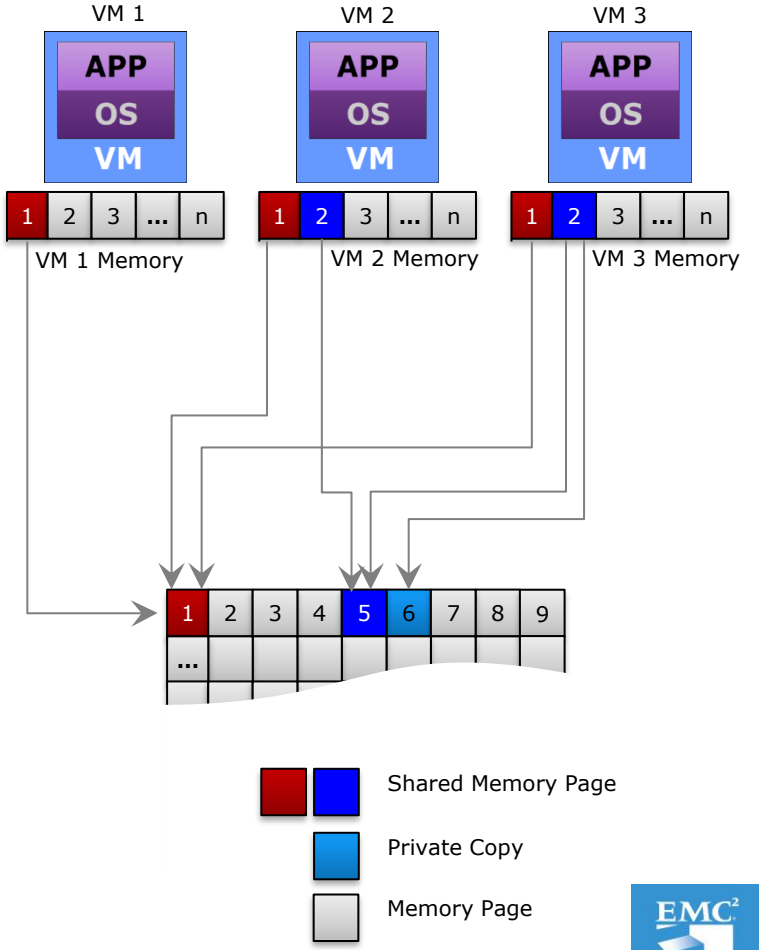
Hyper-threading

- Makes a processor appear as two logical processor cores
  - Enables an OS to schedule two threads simultaneously
- Two logical processor cores share the same physical resources
  - While the current thread is stalled, processor can execute another thread
- Provides improved performance and utilization



Memory Page Sharing

- Eliminates redundant copies of memory pages
- Allows a greater degree of memory over-commitment
- Hypervisor identifies redundant pages
  - VM memory pointer is updated to point to shared location
  - Redundant memory pages are reclaimed



Dynamic Memory Allocation

- A memory optimization technique that reclaims memory pages
- VMs have agent installed in guest OS that communicates with hypervisor
- When memory become scarce:
  - Agent in a VM demands memory from their guest OS
  - Guest OS allocates memory pages to the agent
  - Agent reserves the memory and puts it back into memory pool
  - Hypervisor then assigns the relinquished memory pages to other VMs that require memory

VM Load Balancing Across Hypervisors

- Provides uniform distribution of load across hypervisors
- Key process involved in balancing load of VMs:
  - Management server checks the availability of resources on all hypervisors when a new VM is powered-on
  - Management server places the VM on a hypervisor with sufficient resources and ensures that the load is balanced
  - Management server monitors the load across hypervisors
    - If there is any imbalance, then the server balances the load by migrating the VMs from over-utilized to underutilized hypervisors

Server Flash-cache Technology

- Uses intelligent caching software and a flash card on the compute system
  - Cache software places the most frequently referenced data on the flash card
- Dramatically improves the application performance
  - Provides performance acceleration for read-intensive workloads
  - Avoids network latencies associated with I/O access to the storage system
- Requires “warm-up” time before significant performance improvement is realized

Lesson Summary

During this lesson the following topics were covered:

- Resource allocation models
- Hyper-threading
- Memory page sharing
- Dynamic memory allocation
- VM load balancing across hypervisors
- Server flash-cache

Lesson: Resource Management Technique – 2

This lesson covers the following topics:

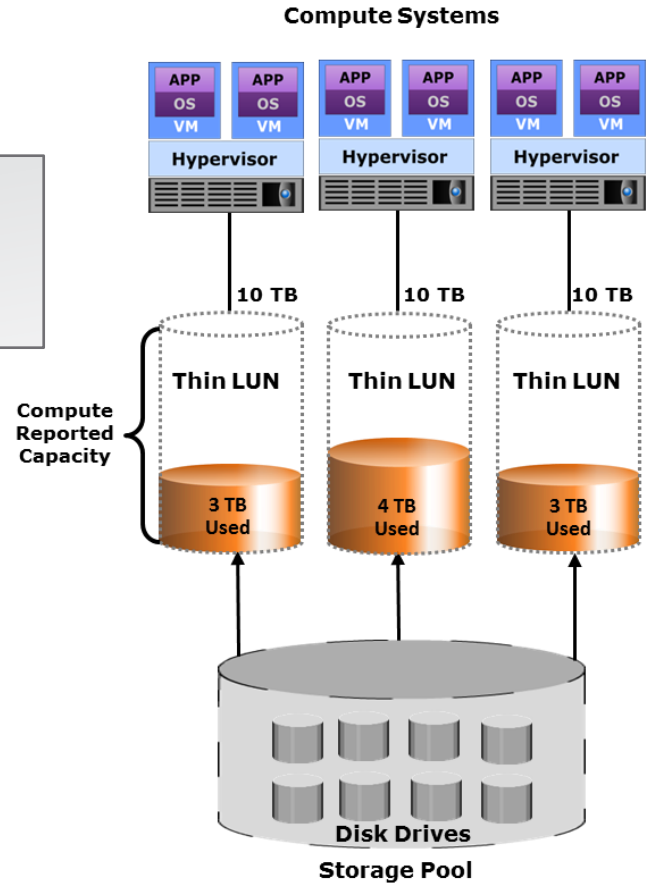
- Virtual storage provisioning
- Storage pool rebalancing
- Thin LUN storage space reclamation
- Automated storage tiering
- Cache tiering
- Dynamic VM load balancing across storage volumes

Virtual Storage Provisioning

Virtual Storage Provisioning

It enables to present a LUN to an application with more capacity than is physically allocated to it on the storage system.

- Physical storage is allocated to the application on-demand
  - Provides more efficient utilization of storage and reduces storage cost
  - Simplified storage management
- Two types of LUNs can be created:
  - Thin LUN
  - Thick LUN



Storage Pool Rebalancing

- Provides the ability to rebalance allocated extents on physical disk drives over the pool when new drives are added
- Restripes data across all disk drives in the shared storage pool
  - Helps in achieving higher overall pool performance
- Enables spreading out the data equally on all the drives within the pool
  - Ensures that the used capacity of each drive is uniform across the pool

Storage Space Reclamation

- Identifies unused space in thin LUNs and re-assigns it to the storage pool
  - Provides cost savings
- Options to reclaim the unused space on a thin LUN
  - Zero extent reclamation
    - De-allocate storage extents that contain all zeroes in a thin LUN
    - De-allocated extents are added back to the pool
  - API-based reclamation
    - API communicate the location of all the identified unused space on the LUN to the storage system to reclaim all unused space to the pool

Automated Storage Tiering

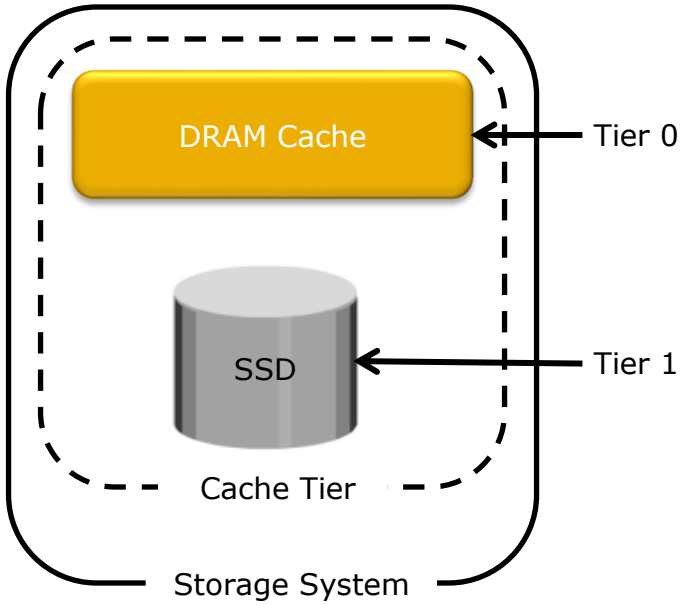
Automated Storage Tiering

A technique of establishing a hierarchy of different storage types for different categories of data that enables storing the right data automatically to the right tier, to meet the service level requirements.

- Each tier has different levels of protection, performance, and cost
- Data is moved between tiers based on defined tiering policies
  - Tiering policy is usually based on parameters such as file type, frequency of access, and so on
- Data movement occurs between tiers
  - Within a storage array (Intra-array)
  - Between storage arrays (Inter-array)

Cache Tiering

- Enables creation of a large capacity secondary cache using SSDs
- Enables tiering between DRAM cache and SSDs (secondary cache)
- Most reads are served directly from high performance tiered cache
- Key benefits:
  - Enhances performance during peak workload
  - Non-disruptive and transparent to applications



Dynamic VM Load Balancing Across Storage Volumes

- Enables intelligent placement of VMs during creation, based on the I/O load and available storage capacity on the volume
  - Improves the performance
- Management server performs ongoing load balancing within a cluster of volumes
  - Cluster volume is a collection or pool of volumes that are aggregated as a single volume
    - Enables efficient and rapid placement of new VMs
- User-configurable space utilization or I/O latency thresholds are defined to ensure space efficiency

Lesson Summary

During this lesson the following topics were covered:

- Virtual storage provisioning
- Storage pool rebalancing
- Storage space reclamation
- Automated storage tiering
- Cache tiering
- Dynamic VM load balancing across storage volumes

Lesson: Resource Management Technique – 3

This lesson covers the following topics:

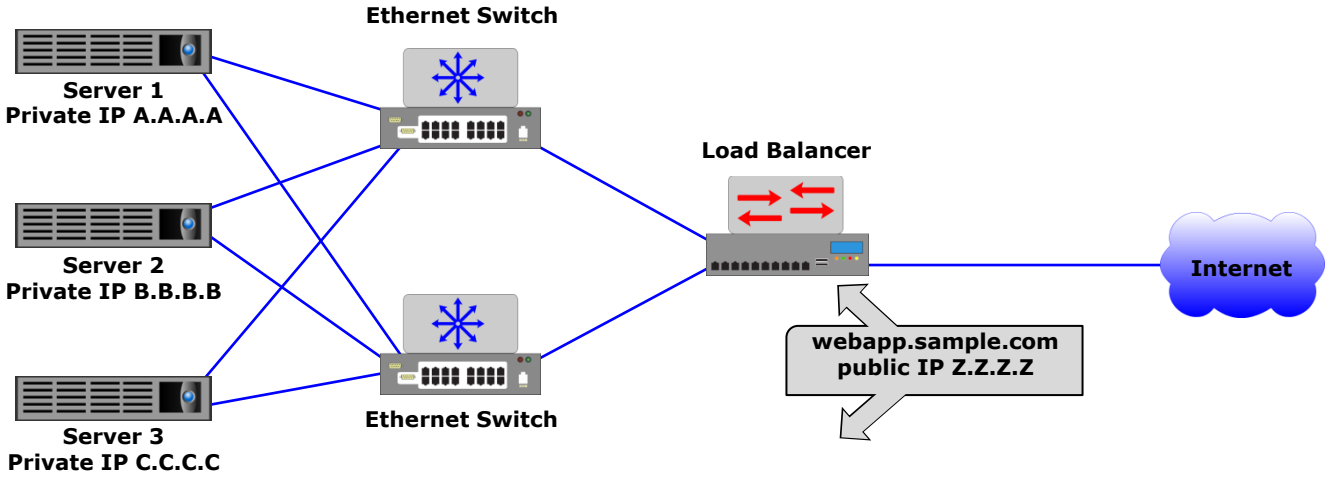
- Balancing client workload across nodes
- Network storm control
- Quality of Service
- Traffic shaping
- Link aggregation, NIC teaming, and multipathing

Introduction to Network Traffic Management

- Network traffic flow is controlled and managed to optimize the performance and availability of cloud services
- Key network traffic management techniques are:
  - Balancing client workload across nodes
  - Network storm control
  - Quality of Service (QoS)
  - Traffic shaping
  - Link aggregation
  - NIC teaming
  - Multipathing

Balancing Client Workload across Nodes

- Splits client workload across multiple nodes
  - Usually performed by a purpose-built device called load balancer
- Load balancer is placed between node cluster and Internet
  - Load balancer decides where to forward each request



Network Storm Control

Storm Control

A networking technique that prevents regular network traffic on a LAN or VLAN from being disrupted by a network storm. A network storm occurs due to flooding of frames on a LAN or VLAN, creating excessive traffic and resulting in degraded network performance.

- Switch monitors incoming frames to switch ports over specific time interval
- Switch counts frames of specific type over the interval
- Switch compares the count with pre-configured threshold
  - Switch port blocks traffic when threshold is reached and filters out subsequent frames until the interval ends

Quality of Service (QoS)

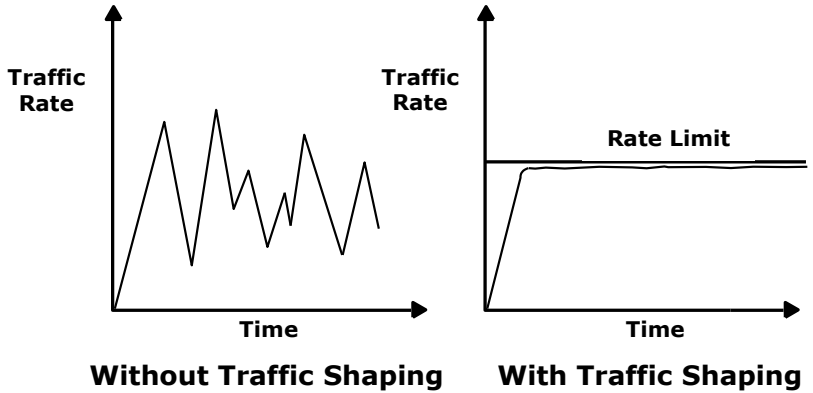
Quality of Service

Capability of a network to prioritize business critical and latency-sensitive network traffic and to provide better service to such traffic over less critical traffic. QoS enables applications to obtain consistent service levels in terms of network bandwidth, latency variations, and delay.

Approach	Description
Integrated Services	• Application signals the network to inform network components about required QoS • Application can transmit data through network only after receiving confirmation from the network
Differentiated Services	• Priority specification to network packets are inserted by applications or by switches or routers • Network uses priority specification to classify traffic and then manage network bandwidth based on the traffic class

Traffic Shaping

- Limits the traffic rate at a network interface such as a node port or a router port
- If traffic rate exceeds the pre-configured limit, traffic shaping queues excess packets for later transmission
  - Ensures required service level for business-critical applications
  - Controls traffic rate per client/tenant to avoid network congestion



Link Aggregation, NIC Teaming, and Multipathing

- Link aggregation
  - Combines links between two switches and between a switch and a node
  - Enables distribution of network traffic across links in the aggregation
- NIC Teaming
  - Distributes network traffic across NICs
  - Provides network traffic failover in the event of a NIC/link failure
- Multipathing
  - Provides load balancing and path failover
  - Improves I/O performance and data path utilization

Lesson Summary

During this lesson the following topics were covered:

- Balancing client workload across nodes
- Network storm control
- Quality of Service
- Traffic shaping
- Link aggregation, NIC teaming, and multipathing

Concepts in Practice

- EMC Unisphere
- EMC Unified Infrastructure Manager (UIM)
- EMC ViPR and ViPR SRM
- EMC FAST VP
- EMC XtremSF
- EMC PowerPath/VE

EMC Unisphere and EMC UIM

Unisphere	UIM
• Provides unified management for file-based, block-based, and object-based storage • Enables to monitor health, alerts, and performance of large numbers of VNX storage systems • Enables to create storage pool and to configure tiering • Enables administrators to drill down and troubleshoot the issues	• Unified management solution for Vblock and VSPEX systems • Enables configuring and provisioning resources for services • Provides a dashboard showing infrastructure configurations and utilization • Performs compliance check during resource configurations • Rapidly performs root cause analysis

EMC ViPR and EMC FAST VP

ViPR/ViPR SRM

- EMC ViPR
  - A software-defined storage solution
  - Separates the management functions from the storage
  - Centralizes storage management centrally through software
- EMC ViPR SRM
  - Enables management of physical infrastructure under ViPR control and helps associate relationships between physical and software-defined resources

FAST VP

- Performs storage tiering at sub-LUN level
- Data movement between tiers are based on user-defined policies
- Optimizes performance and cost
- Increases storage efficiency
- Supported on both VMAX and VNX storage arrays

EMC XtremSF and PowerPath/VE

XtremSF	PowerPath/VE
• A server-flash cache solution • A PCIe flash card deployed in the compute system, which improves performance • XtremSW Cache accelerates reads and protects data using write-through cache	• Provides multipathing solution for VMware ESX/ESXi and Microsoft Hyper-V • Provides advanced multipathing with path failover and load balancing across FC, iSCSI, or FCoE I/O paths

Module Summary

Key points covered in this module:

- Control layer and its key functions
- Element manager and unified manager
- Software-defined approach for managing IT resources
- Key resource management techniques