DIGITAL NOTES ON

CLOUD COMPUTING(R20A0521)

B.TECH IV Year - I Sem
(2023-24)

PREPARED BY
Dr.A.V.H.SAI PRASAD
A.VIJETHA
T.SRINIDHI

DEPARTMENT OF INFORMATION TECHNOLOGY

MALLA REDDY COLLEGE OF ENGINEERING AND TECHNOLOGY
(Autonomous Institution – UGC, Govt. of India)
Sponsored by CMR Educational Society
(Affiliated to JNTU, Hyderabad, Approved by AICTE- Accredited by NBA& NAAC–'A'Grade-ISO9001:2008Certified) Maisammaguda,Dhulapally(PostViaHakimpet),Secunderabad–500100,TelanganaState,India.
www.mrcet.ac.in

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

Department of Information Technology

IV Year B.Tech IT – I Sem L/T/P/C

3/-/-/3

(R20A0521) CLOUD COMPUTING

Objectives:

- To understand the various distributed system models and evolving computing paradigms
- To gain knowledge in virtualization of computer resources
- To realize the reasons for migrating into cloud
- To introduce the various levels of services that can be achieved by a cloud.
- To describe the security aspects in cloud and the services offered by a cloud.

UNIT-I

Cloud Computing Fundamentals :Definition of Cloud computing , Roots of Cloud Computing , Layers and Types of Clouds, Desired Features of a Cloud, Cloud Infrastructure Management , Infrastructure as a Service Providers , Platform as a Service Providers

Computing Paradigms: High Performance Computing, Parallel Computing, Distributed Computing, Cluster Computing, Grid Computing.

UNIT-II

Migrating into a cloud: Introduction, Broad Approaches to Migrating into the cloud, the seven-step Model of Migration into a cloud.

Virtualization: Virtual Machines and Virtualization of clusters and data centers-Implementation Levels of Virtualization -Virtualization Structures/Tools and Mechanisms-Virtualization of CPU, Memory, and I/O Devices-Virtual Clusters and Data Centers.

UNIT-III

InfraStruture as a service (IAAS) & Platform(PAAS): Virtual machines provisioning and Migration services – Virtual Machines Provisioning and Manageability – Virtual machine Migration Services – VM Provisioning and Migration in Action.On the Management of Virtual machines for cloud Infrastructures. Aneka – Integration of Private and Public Clouds.

UNIT- IV

Software as a Service (SAAS) & Data Security in the Cloud: Software as a Service (SAAS),Google App Engine-Centralizing Email Communications- Collaborating via Web-Based Communication Tools-An Introduction to the idea of Data Security.The Current State of Data Security in the Cloud-Cloud Computing and Data Security Risk – Cloud Computing and Identity.

UNIT-V

SLA Management in cloud computing: Traditional Approaches to SLO Management, Types of SLA, Life Cycle of SLA, SLA Management in Cloud.

TEXT BOOK

1. Distributed and Cloud Computing, Kaittwang Geoffrey C.Fox and Jack J Dongrra, ElsevierIndia 2012

REFERENCE BOOKS

1. Mastering Cloud Computing- Raj Kumar Buyya, Christian Vecchiola and S.TanuraiSelvi , 2.TMH, 2012.

2. 3. Michael Miller, Cloud Computing: Web-Based Applications That Change the Way YouWork and Collaborate Online, Que Publishing, August 2008.

COURSE OUTCOMES:

Students will be able to:

1. To distinguish the different models and computing paradigms.
2. To explain the levels of virtualization and resources virtualization
3. To analyze the reasons for migrating into cloud
4. To effectively use the cloud services in terms of infrastructure and operating platforms.
5. To apply the services in the cloud for real world scenarios

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY
DEPARTMENT OF INFORMATION TECHNOLOGY

INDEX

UNIT-I

Cloud Computing Fundamentals: Definition of Cloud computing, Roots of Cloud Computing , Layers and Types of Clouds, Desired Features of a Cloud, Cloud Infrastructure Management, Infrastructure as a Service Providers, Platform as a Service Providers.

Computing Paradigms: High Performance Computing, Parallel Computing, Distributed Computing, Cluster Computing, Grid Computing.

Introduction to Cloud Computing:

Cloud is a parallel and is tribute computing system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements (SLA) established through negotiation between the service provider and consumers.

Clouds are a large pool of easily usable and accessible virtualized resources(such as hardware, development platforms and/or services).These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization‖

This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized Service Level Agreements.

Clouds are hardware based services offering compute , network , and storage capacity where Hardware management is highly abstracted from the buyer, buyers incur infrastructure costs as variable OPEX, and infrastructure capacity is highly elastic.

Key characteristicsofcloudcomputing

1. Theillusionofinfinitecomputingresources;
2. Theeliminationofan up-frontcommitmentbycloudusers;
3. Theabilitytopayforuse…asneeded

The National Institute of Standards and Technology (NIST) characterizes **cloud computing** as-...a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g.networks ,servers ,storage ,applications , services) that can be rapidly provisioned and released with minimal management effortor service provider interaction.

Most common characteristics which a cloud should have:

(i) pay-per-use (no on going commitment, utility prices); (ii) elastic capacity and the illusion of infinite resources; (iii) self-service interface; and (iv) resources that is abstracted or virtualized.

Roots of Cloud Computing:

The roots of clouds computing can be tracked by observing the advancement of several technologies, especially in hardware (virtualization, multi core chips), Internet technologies (Web services ,service-oriented architectures,Web2.0),distributed computing (clusters, grids), and systems management (autonomic computing ,data center automation).

Figure 1.1 shows the convergence of technology fields that significantly advanced and contributed tot head vent of cloud computing.

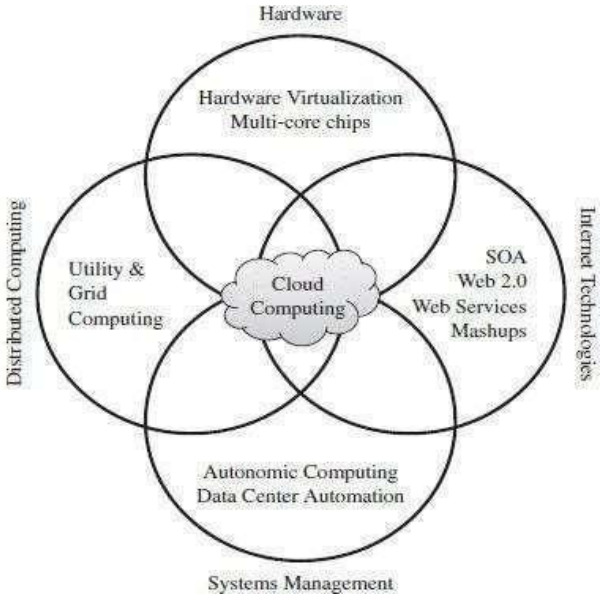


FIGURE 1.1. Convergence of various advances leading to the advent of cloud computing.

The IT world is currently experiencing a switch from in-house generated computing power into utility-supplied computing resources delivered over the Internet as Web services.

Computing delivered as a utility can be defined as—on demand delivery of infrastructure, applications, and business processes in a security-rich, shared, scalable ,and based computer environment over the Internet for a fee‖.

This model brings benefits to both consumers and providers of IT services. Consumers can attain reduction on IT-related costs by choosing to obtain cheaper services from external providers as opposed to heavily investing on IT infrastructure and personnel hiring. The—on- demand component of this model allows consumers to adapt their IT usage to rapidly increasing or unpredictable computing needs.

Providers of IT services achieve better operational costs; hardware and software infrastructures are built to provide multiple solutions and serve many users, thus increasing efficiency and ultimately leading of asterre turn on investment(ROI)as Well as lower total cost of ownership(TCO).

In the 1970s, companies who offered common data processing tasks, such as payroll automation, operated time-shared main frames as utilities, which coulds erve dozens of applications and often operated closeto100%oftheircapacity.

The mainframe era collapsed with the advent off a stand inexpensive microprocessors and IT data centers moved to collections of commodity servers. A part from its clear advantages, this new mode line vitably led to isolation of workload into dedicated servers, mainly due to incompatibilities between software stacks and operating systems.

Inaddition, the unavailability of efficient computer networks meant that IT infrastructure should be hosted inproximity to where it would be consumed. Altogether, these facts have prevented the utility computing reality of taking place on modern computer systems. These facts reveal the potential of delivering computing services with the speed and reliability that businesses enjoy with their local machines. The benefits of economies of scale and high utilization allow providers to offer computing services for a fraction of what it costs for a typical company that generates its own computing power.

SOA, WebServices,Web2.0, and Mashups

The emergence of Webservices (WS) open standards has significantly contributed to advances in the domain of software integration. Webservices can combine together applications running on different messaging product platforms, enabling information from one application to be made available to others, and enabling internal applications to be made available over the Internet.

WS standards have been created on top of existing ubiquitous technologies such as HTTP and XML, thus providing a common mechanism for delivering services, making them ideal for implementing a service-oriented architecture (SOA).The purpose of a SOA is to address requirements of loosely coupled, standards-based, and protocol-independent distributed computing. In a SOA, software resources arepackagedas‐services,‖whicharewell-defined,self-containedmodulesthat provide standard business functionality and are independent of the state or context of other services.

Services are described in a standard definition language and have a published interface. The maturity of WS has enabled the creation of powerful services that can be accessed on-demand, in a uniform way. An enterprise application that follows the SOA paradigm is a collection of services that together perform complex business logic.

In the consumer Web, information and services may be programmatically aggregated, acting as building blocks of complex compositions, called service mashups. Many service providers, such as Amazon, del.icio.us, Facebook, and Google, make their service APIs publicly accessible

using standard protocols such as SOAP and REST. Consequently, one can put an idea of a fully functional Web application into practice just by gluing pieces with few lines of code.

In the Software as a Service(SaaS) domain, cloud applications can be built as Compositions of other services from the same or different providers. Services such as user authentication, e-mail, payroll management, and calendars are examples of building blocks that can be reused and combined in a business solution in case a single, ready-made system does not provide all those features.

LAYERS AND TYPES OF CLOUDS

Cloud computing services are divided into three classes

(1) Infrastructure as a Service, (2)Platform as a Service, and(3)Software as a Service

Figure1.3depicts the layered organization of the cloud stack from physical infrastructure to applications.

These abstraction levels can also be viewed as a layered architecture where services of a higher layer can be composed from services of the underlying layerA core middleware manages physical resources and the VMs deployed on top of them; in addition, it provides the required features(e.g., accounting and billing) to offer multi-tenantpay-as-you-go services.

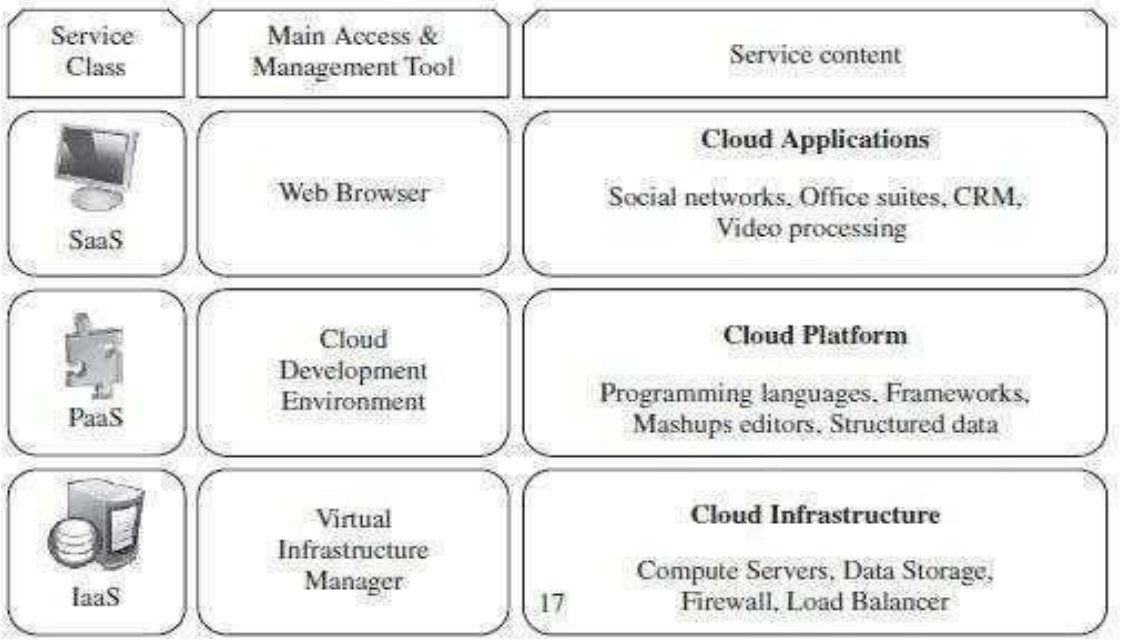


FIGURE 1.3 The cloud computing stack Infrastructure as a Service

Offering virtualized resources (computation,storage,andcommunication)on demand is known asInfrastructure as a Service(IaaS).A cloud infrastructure enables on demand provisioning of servers runnings everal choices of operating systems and a customizeds of twarestack. Infrastructure services are considered to be the bottom layer of cloud computingsystems.

Platform as a Service

A cloud platform offers an environment on which developers create and deploy applications and do not necessarily need to know how many processors or how much memory that applications will be using. In addition, multiple programming models and specialized services(e.g.,dataaccess,authentication,andpayments)are offered as building block stone wapplications.

Google App Engine, an example of Platform as a Service, offers a scalable environment for developing and hosting Web applications, which should be written in specific programming languages such as Python or Java, and use the services 'own proprietary structured objectd atastore.

Software as a Service

Applications reside on the top of the cloud stack. Services provided by this layer can be accessed by end users through Web portals. Therefore, consumers are increasingly shifting from locally installed computer programs to on-lines of tware services that offer the same functionally. Traditional desktop applications such as word processing and spread sheet can now be accessed as a service in the Web. This model of delivering applications, known as Software as aService(SaaS), alleviates the burden of software maintenance for customers and simplifies development and testing for providers.

Sales force.com, which relies on the SaaS model, offers business productivity applications (CRM) that reside completely on their servers, allowing customers to customize and access applications on demand.

Deployment Models

A cloud can be classified as public, private, community, or hybrid based on model of deployment as showninFigure 1.4.

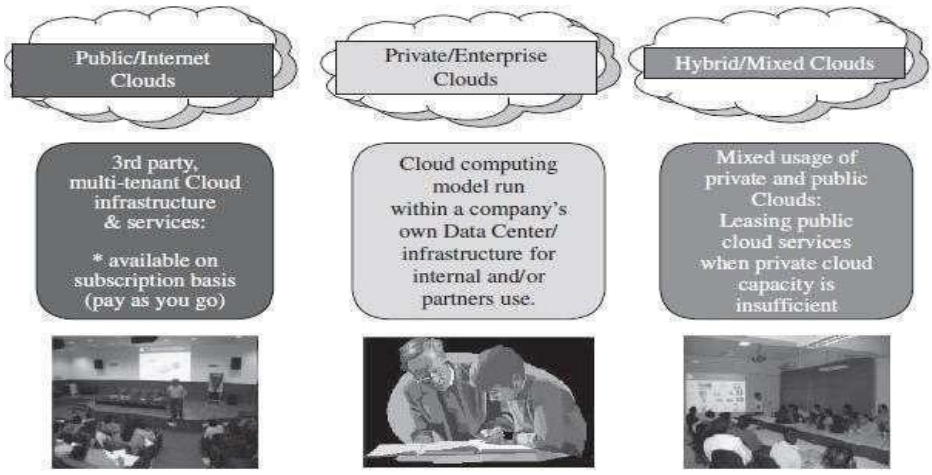


FIGURE 1.4. Types of clouds based on deployment models.

Public cloud: —cloud made available in a pay-as-you-go manner to the general public

Private cloud:-internal data center of a business or other organization, not made available to the general public.

Community cloud: shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations)

Hybrid cloud takes shape when a private cloud is supplemented with computing capacity from public clouds.

The approach of temporarily renting capacity to handle spikes in load is known as **"cloud-bursting"**

Desired Features of a Cloud

Certain features of a cloud are essential to enable services that truly represent the cloud computing model and satisfy expectations of consumers, and cloud offerings must be

(i) self-service, (ii)per-usage metered and billed, (iii)elastic, and (iv)Customizable

Self-Service: clouds must allow self-service accesss othat customers can request, customize ,pay ,and use services without intervention of human operators

Per-UsageMetering and Billing: Cloud computing eliminates up-front commitment by users, allowing them to request and use only the necessary amount. Services must be priced on a short term basis (e.g., by the hour), allowing users to release (andnotpayfor)resources as soon as they are not needed

Elasticity: Cloud computing gives the illusion of infinite computing resources available on demand. Therefore users expect clouds to rapidly provide resources in any quantity at any time. In particular, it is expected that the additional resources can be (a) provisioned, possibly automatically, when an application load increases and(b)released when load decreases(scale up and down)

Customization: resources rented from the cloud must be highly customizable .customization means allowing users to deploy specialized virtual appliances and to be given privileged(root)access to the virtual servers.

CLOUD INFRASTRUCTURE MANAGEMENT

A key challenge IaaS providers face when building a cloud infrastructure is managing physical and virtual resources, namely servers, storage, and networks, in a holistic fashion. The orchestration of resources must be performed in a way to rapidly and dynamically provision resources to applications.

The software tool kit responsible forth is orchestration is called a virtual infrastructure emanager(VIM).This type of software resemblesa traditional operating system— but instead of dealing with a single computer, it aggregates resources from multiple computers ,present in gauni form view touser and applications.

Features Virtualization Support:

The multi-tenancy aspect of clouds requires multiple customers with disparate requirements to be served by a single hardware infrastructure. Virtualized resources (CPUs, memory, etc.) can be sized and Resized with certain flexibility. These features make hardware virtualization, the ideal technology to create a virtual infrastructure that partitions a data center among multiple tenants.

Multiple Backend Hypervisors: Different virtualization models and tools offer different benefits, drawbacks, and limitations. Thus, some VI managers provide auniform management layer regardless of the virtualization technology used. This characteristic is more visible in open- source VI managers, which usually providepluggable drivers to interact with multiple hypervisors.

Storage Virtualization: Virtualizing storage means abstracting logical storage from physical storage.By consolidating all available storage devices in a data center, it allows creating virtual disks independent from device and location.

Interface to Public Clouds: Extending the capacity of a local in-house computing infrastructure by borrowing resources from public clouds is advantageous. In this fashion, institutions can make good use of their available resources and, in case of spikes in demand, extra load can be offloaded to rented resources. A VI manager can be used in a hybrid cloud setup if it offers a driver to manage the life cycle of virtualized resources obtained from external cloud providers. To the applications, the use of leased resources must ideally be transparent.

Virtual Networking: Virtual networks allow creating an isolated network on top of a physical infrastructure independently from physical topology and locations.A virtua lLAN(VLAN)allows isolating traffic that shares a switched network, allowing VMs to be grouped into the same broadcast domain. Additionally, aVLAN can be configured to block traffic originated from VMs from other networks.

Virtual Clusters: Several VI managers can holistically manage groups of VMs.This feature is useful for provisioning computing virtual clusters on demand, and interconnected VMs formulti-tier Internet applications.

INFRASTRUCTURE AS A SERVICE PROVIDERS

Public Infrastructure as a Service providers commonly offer virtual servers containing one or more CPUs, running several choices of operating systems and acustomized software stack. In addition, storage space and communication facilities are often provided.

Features

IaaS offerings can be distinguished by the availability of specialized features that influence the cost benefit ratio to be experienced by user applications when moved to the cloud.The most relevant features are: (i)geographic distribution of datacenters; (ii)variety of user interfaces and API t to access the system; (iii)specialized components and services that aid particular applications(e.g.,load balancers,firewalls.

Geographic Presence: To improve availability and responsiveness, a provider of world wide services would typically build several data centers distributed around the world. For example, Amazon Web Services presents the concept of-availabilityzones‖and-regions‖foritsEC2

User Interfaces and Access to Servers: Ideally, a public IaaS provider must provide multiple access means to its cloud, thus catering for various users and their preferences. Different types of user interfaces(UI)provide different levels of abstraction, the most common being graphical user interfaces(GUI).

PLATFORM AS A SERVICE PROVIDERS

Public Platform as a Serviceproviders commonly offer a development and deployment environment that allow users to create and run their applicationswith little or no concern to low-level details of the platform.In addition,specific programming languages and frameworks are made available in the platform,as well as other services such as persistent datastorage and in memory caches.

Features

Programming Models, Languages, and Frameworks:

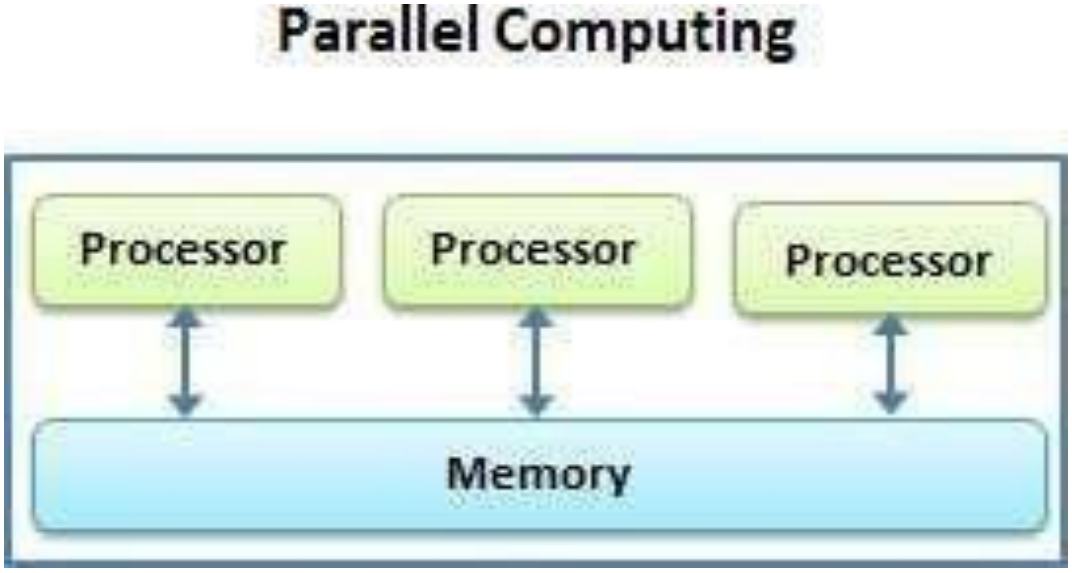
Programming models made available by IaaS providers define how users can express their applications using higher levels of abstraction and efficiently run the month ecloud platform. Each model aims at efficiently solving a particular problem.

Persistence Options: A persistence layer is essential to allowapplications torecord their state and recover it in case of crashes, as well as to store user data.Web and enterprise application developers have chosen relational databases as thepreferred persistence method. These databases offer fast and reliable structureddata storage and transaction processing, but may lack scalability to handle severalpetabytes of data stored in commodity computers.

Computing Paradigm

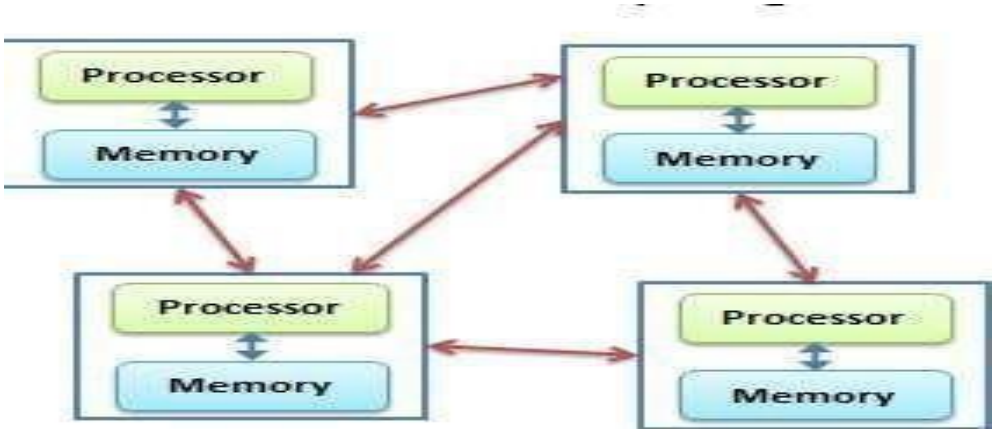
Parallel computing:

- In parallel computing, all processors are either tightly coupled with centralized shared memory or loosely coupled with distributed memory.
- Interprocessor communication is accomplished through shared memory or via message passing.
- A computer system capable of parallel computing is commonly known as a **parallel computer.**
- Programs running in a parallel computer are called **parallel programs.**The process of writing parallel programs is often referred to as **parallel programming.**



Distributed computing:

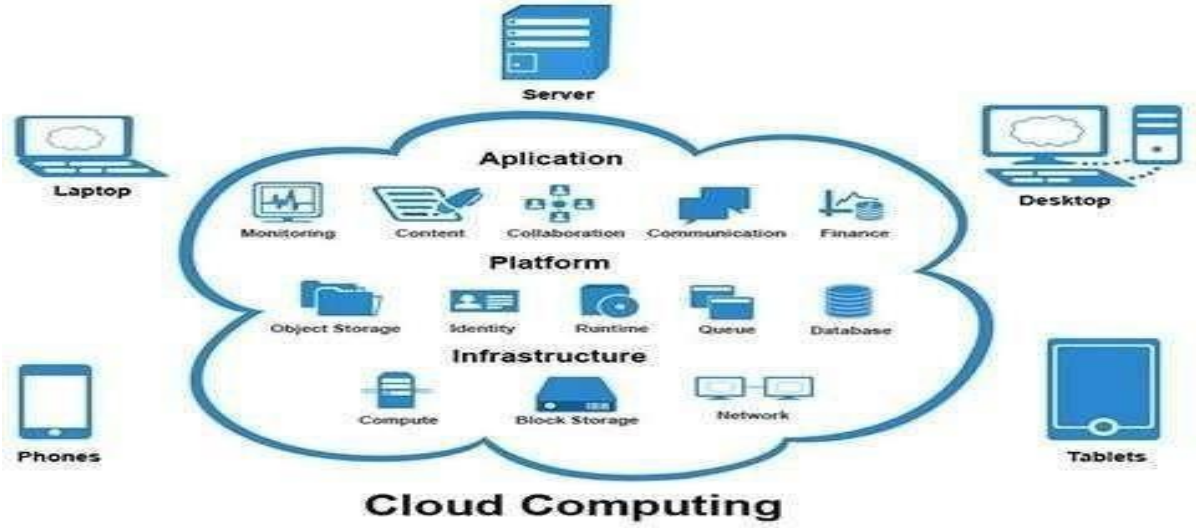
- A distributed system consists of multiple **autonomous computers**,each having it own private memory, communicating through a computernetwork.
- Information exchange in a distributed system is accomplishedthrough message passing.
- A computer program that runs in a distributed system is known as a **Distributed program.**



Cloud computing:

- An Internet cloud of resources can be either a centralized or a distributedcomputing system.

- The cloud applies parallel or distributed computing, or both. Clouds can be built with physical or virtualized resources over large datacenters that are centralized or distribute



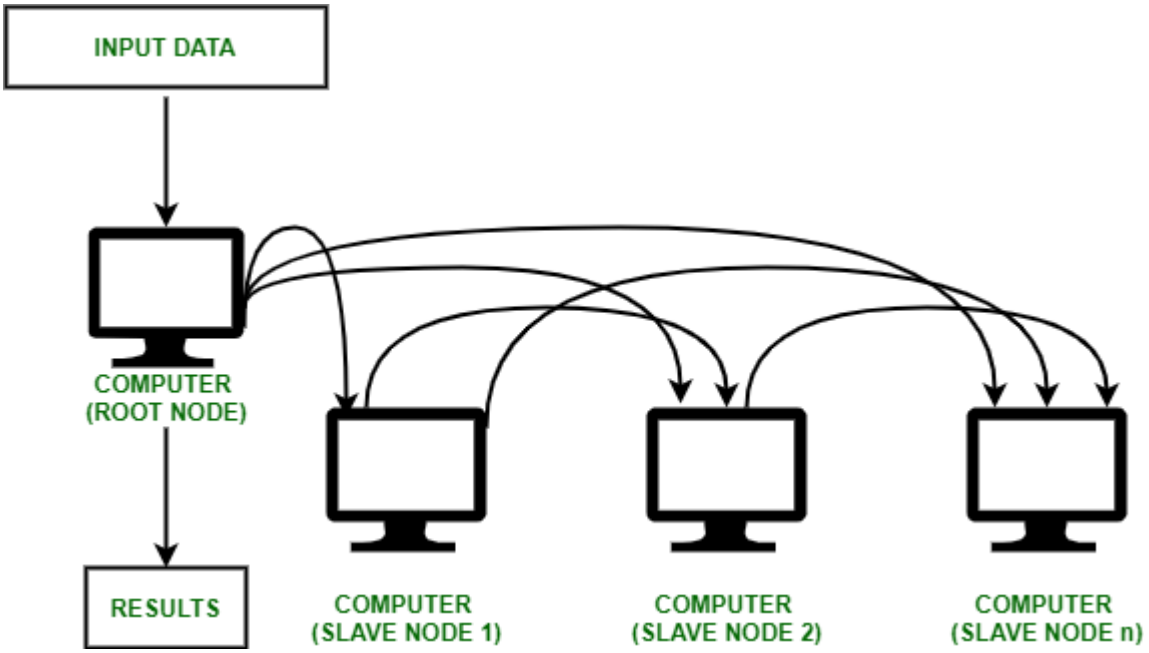
Grid Computing

Grid computing enables aggregation of distributed resources and transparently access to them. Most production grids such as Tera Grid and EGEE seek to share compute and storage resources distributed across different administrative domains, with their main focus being speeding up a broad range of scientific applications, such asclimate modeling, drug design, and protein analysis.

A key aspect of the grid vision realization has been building standard Web services -basedprotocolsthatallowdistributedresourcestobe-discovered,accessed,allocated,monitored, accounted for, and billed for, etc., and in general managed as a single virtual system.The Open Grid Services Architecture (OGSA) addresses this need for standardization by defining a set of core capabilities an d behaviors that address key concerns in grid systems.

Globus Toolkit is a middleware that implements several standard Grid services.

Cluster computing



Quantum computing

Quantum computing is a multidisciplinary field comprising aspects of computer science, physics, andmathematics that utilizes quantum mechanics to solve complex problems faster than on classical computers. The field of quantum computing includes hardware research and application development. Quantum computers are able to solve certain types of problems faster than classical computers by taking advantage of quantum mechanical effects, such as superposition and quantum interference. Some applications where quantum computers can provide such a speed boost include machine learning (ML), optimization, and simulation of physical systems. Eventual use cases could beportfolio optimization in finance or the simulation of chemical systems, solving problems that are currently impossible for even the most powerful supercomputers on the market.

UNIT-II

Migrating into a cloud: Introduction, Broad Approaches to Migrating into the cloud, the seven-step Model of Migration into a cloud.

Virtualization: Virtual Machines and Virtualization of clusters and data centers-Implementation Levels of Virtualization -Virtualization Structures/Tools and Mechanisms-Virtualization of CPU, Memory, and I/O Devices-VirtualClusters and DataCenters.

Migrating into a cloud

Introduction:

Cloud migration is the procedure of transferring **applications, data**, and **other types of business components** to any cloud computing platform. There are several parts of cloud migration an organization can perform. The most used model is the **applications** and **data transfer** through an on-premises and local data center to any public cloud.

But, a cloud migration can also entail transferring applications and data from a single cloud environment or facilitate them to another- a model called **cloud-to-cloud migration**. The other type of cloud migration is reverse cloud migration, cloud exit, and cloud repatriation where applications or data are transferred and back to the local data center.

Broad Approaches to Migrating into the cloud:

There are different strategies when it comes to the migration of applications to the cloud. Multiple factors influence the strategy, and it could vary from the portfolio of applications to individual ones. Enterprises should examine all the factors for their application while picking up the most suitable strategy. Brief descriptions of effective migration strategies are as follows:

Strategy or Approaches	Description
Re-hosting	• Also known as “Lift and shift”. • Redeploy your existing data and applications on the cloud server as is. • It is most suited for companies new to the cloud and cases where modifying the application code is complex. • This help companies to reduce their on-premise infrastructure expenses Immediately

Strategy or Approaches	Description
Re-platform	• Also known as "Lift, Modify, and Shift". • With this strategy, the application will be tweaked and optimised for the cloud. • The core architecture of the application will remain intact; however, adjustments will be made to enable leveraging of cloud-based tools. • This approach is most suited for companies who have a conservative approach to the cloud and want to achieve benefits of cloud-like improved system performance.
Revise	• It is very similar to the previous platform strategy. However, it is called revising when more changes are required to the architecture and the codebase while moving to the cloud.
Rebuild	• The rebuilding strategy takes the revised strategy further by discarding the existing code base and replace with a new set of code. • This strategy will require more time and cost. It is considered when the existing applications do not meet their current business needs.
Replace	• Replacing with another solution will address the challenges associated with the rebuild approach. • Replacing involves migration to a third-party, pre-built application. • Here, migration activities mainly restrict data migration from the current environment, and everything else will be new. • This is driven by businesses need to leverage cloud capabilities that are not available now.
Retire	• While assessing the application's readiness to move to the cloud, applications could no longer be helpful to the business. In this case, retire those applications by simply turning them off.
Retain	• There could be applications due to various factors not get prioritised for moving to the cloud. • The application may have gone through an upgrade recently etc. • In this case, retain the application as-is and can revisit for cloud migration later.

The seven-step Model of Migration into a cloud:

Migrating a model to a cloud can help in several ways, such as improving scalability, flexibility, and accessibility. There are seven steps to follow when migrating a model to the cloud:

Step 1: Choose the right cloud provider (Assessment step)

The first step in migrating your model to the cloud is to choose a cloud provider that aligns with your needs, budget, and model requirement. consider the factors such as compliance, privacy, and security.

Step 2: Prepare your data (Isolation step)

Before migrating to your cloud, you need to prepare your data. for that ensure your data is clean and well organized, and in a format that is compatible with your chosen cloud provider.

Step3:Choose your cloud storage (Mappingstep) Once your data is prepared, you need to choose your cloud storage. This is where your data is stored in the cloud. there are many cloud storage services such as GCP Cloud Storage, AWS S3, or Azure Blob Storage.

Step 4: Set up your cloud computing resources and deploy your model (Re- architect step) If you want to run a model in the cloud, you will need to set up your cloud computing resources. This includes selecting the appropriate instance type and setting up a virtual machine(VM) or container for your model. After setting up your computing resource, it is time to deploy your model to the cloud. This includes packaging your model into a container or virtual machine image and deploying it to your cloud computing resource. and while deploying it may be possible that some functionality gets lost so due to this some parts of the application need to be re-architect.

Step-5: Augmentation step

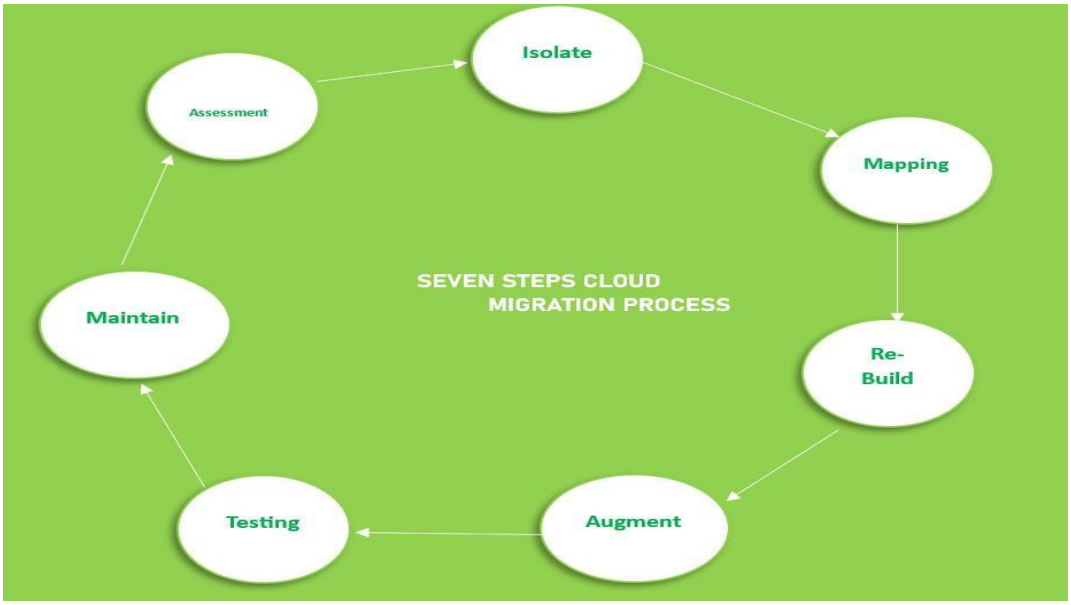
It is the most important step for our business for which we migrate to the cloud in this step by taking leverage of the internal features of cloud computing service we augment our enterprise.

Step 6: Test your Model

Once your model is deployed, we need to test it to ensure that it is working or not. That involves running test data through your model and comparing the results with your expected output.

Step 7: Monitor and maintain your Model

After the model is deployed and tested, it is important to monitor and maintain it. That includes monitoring the performance, updating the model as needed, and need to ensure your data stays up-to-date. Migrating your machine learning model to the cloud can be a complex process, but above 7 steps, you can help ensure a smooth and successful migration, ensuring that your model is scalable and accessible.

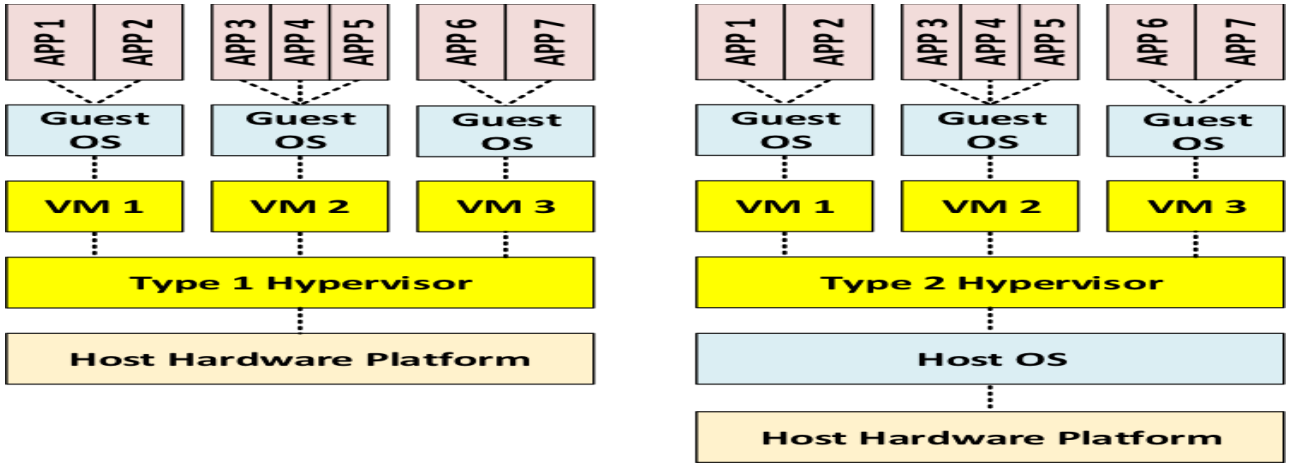


Virtualization:

In layman words, Virtualization enables users to disjoint operating systems from the underlying hardware, i.e., users can run multiple operating systems such as Windows, Linux, on a single physical machine at the same time. Such operating systems are known as guest Oses (operating systems).Virtualization deploys software that makes an abstraction layer across computer hardware, letting the hardware components such as processors, memory, storage etc of a particular computer tobe segmented into several virtual elements (also known as virtual machines).

Moreover, in today's' time, virtualization is globally adopted in enterprise IT architecture and drives cloud computing economics. Essentially, Virtualization allows cloud providers to deliver users alongwith existing physical computer hardware.

As a simple process, it enables cloud users to purchase only necessary computing resources when they actually need it, and to sustain those resources cost-effectively when the workload expands.



Some terminologies associated with Virtualization

1. **Hypervisor:** It is an operating system, performing on the actual hardware, the virtual counterpart is a subpart of this operating system in the form of a running process. Hypervisors are observed as Domain 0 or Dom0.
2. **Virtual Machine (VM):**It is a virtual computer, executing underneath a hypervisor.
3. **Container:** Some light-weighted VMs that are subpart of the same operating system instance as its hypervisor are known as containers. They are a group of processes that runs along withtheir corresponding namespace for process identifiers.
4. **Virtualization Software:** Either be a piece of a software application package or an operating system or a specific version of that operating system, this is the software that assists in deploying the virtualization on any computer device.
5. **Virtual Network:** It is a logically separated network inside the servers that could be expanded across multiple servers.

Virtual Machines and Virtualization of clusters and data centers:

Levels of Virtualization Implementation:

- Virtualization is acomputer architecture technology by which multiple virtual machines(VMs) are multiplexed in the same hardware machine.
- After virtualization, different user applications managed by their own operating

systems(guestOS) can run on the same hardware independent of the hostOS done by adding

additional software, called a virtualization layer.

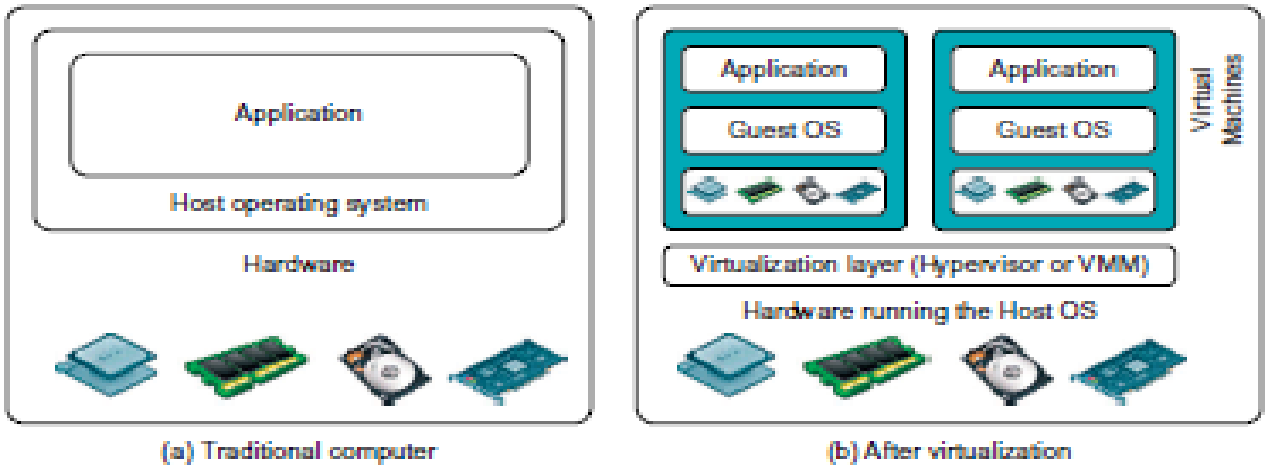


FIGURE 3.1

The architecture of a computer system before and after virtualization, where VMM stands for virtual machine monitor.

- This virtualization layer is known as hyper visoror virtual machine monitor(VMM)
- The main function of the software layer for virtualization is to virtualize the physical hardware of a osmachine into virtual resources to be used by the VMs.
- Common virtualization layers include the instruction set architecture(ISA) level, hardware level, operating system level, library support level, and application level.

Instruction Set Architecture Level

- At the ISA level, virtualization is performed by emulating a given ISA by the ISA of the host machine.
- For example,MIPS binary code can run on an x86-based hostmachine with the help of ISA emulation. With this approach, it is possible to run a large amount of legacy binarycode written for various processors on any given new hardware hostmachine Instruction set emulation leads to virtual ISAs created on any hardware machine.
- Basic emulation method is through code interpretation. An interpreter program interprets the source instructions to target instructions one by one. One source instruction may require tens or hundreds of native target instructions to perform itsfunction. Obviously, this process is relatively

slow. For better performance, dynamic binary translation is desired.

- This approach translates basic blocks of dynamic source instructions to target instructions. The basic block scan also be extended to program traces or super blocks to increase translation

  efficiency.
- Instruction set emulation requires binary translation and optimization. A virtual instruction set architecture (V-ISA) thus requires adding a processor-specific software translation layer to the compiler.

Hardware Abstraction Level

- Hardware-level virtualization is performed right ontop of the bare hardware.
- This approach generates a virtual hardware environment for aVM.
- The process manages the underlying hardware through virtualization. The idea is to virtualize a computer's resources, such as its processors, memory, and I/O devices.
- The intention is to upgrade the hardware utilization rate by multiple users concurrently. The idea was implemented in the IBM VM/370inthe1960s.
- More recently, the Xen hypervisor has been applied to virtualize x86-based machines to run Linux orother guest OS applications.

Operating System Level

- This refers to an abstraction layer between traditional OS and user applications.
- OS-level virtualization creates isolated container so a single physical server and the OS instances to utilize the hardware and software in data centers.
- The containers behave like real servers.
- OS level virtualization is commonly used in creating virtual hosting environments to allocate hardware resources among a large number of mutually distrusting users.
- It is also used, to a lesser extent, in consolidating server hardware by moving services on separate hosts into containers or VM so none server.

Library Support Level

- Most applications use API sex ported by user-level libraries rather than using lengthy system

calls by the OS.

- Since most systems provide well-documented APIs, such an interface becomes another candidate for virtualization.
- Virtualization with library interfaces is possible by controlling the communication link between applications and there to system through API hooks.

User-Application Level

- Virtualization at the application level virtualizes an application as a VM.
- On a traditional OS, an application often runs as a process. Therefore, application-level virtualization is also known as process-level virtualization.
- The most popular approach is to deploy high level language (HLL)VMs. In this scenario, the virtualization layer sits as an application program on top of the operating system,
- The layer exports an abstraction of a VM that can run programs written and compiled to a particular abstract machine definition.
- Any program written in the HLL and compiled for this VM will be able to run on it. The Microsoft .NET CLR and Java Virtual Machine (JVM) are two good examples of this class of VM.

Virtualization Structures/Tools and Mechanisms

- The layer between real hardware and traditional operating systems. This layer is commonly called the Virtual Machine Monitor (VMM) three requirements for a VMM
- a VMM should provide an environment for programs which is essentially identical to the original machine programs run in this environment should show, at worst, only minor decreases in speed
- VMM should be in complete control of the system resources.
- VMM includes the following aspects:

(1) The VMM is responsible for allocating hardware resources for programs;

(2) It is not possible for a program to access any resource not explicitly allocated to it;

(3) It is possible under certain circumstances for a VMM to regain control of resources already allocated.

Virtualization of CPU

- A CPU architecture is virtualizable if it supports the ability to run the VM's privileged and unprivileged instructions in the CPU's user mode while the VMM runs in supervisor mode.
- Hardware-Assisted CPU Virtualization: This technique attempts to simplify virtualization because full or para virtualization is complicate.

Memory Virtualization

- **Memory Virtualization** :the operating system maintains mappings of virtual memory to machine memory using page table
- All modern x86 CPUs include a memory management unit (MMU) and a translation look aside buffer (TLB) to optimize virtual memory performance

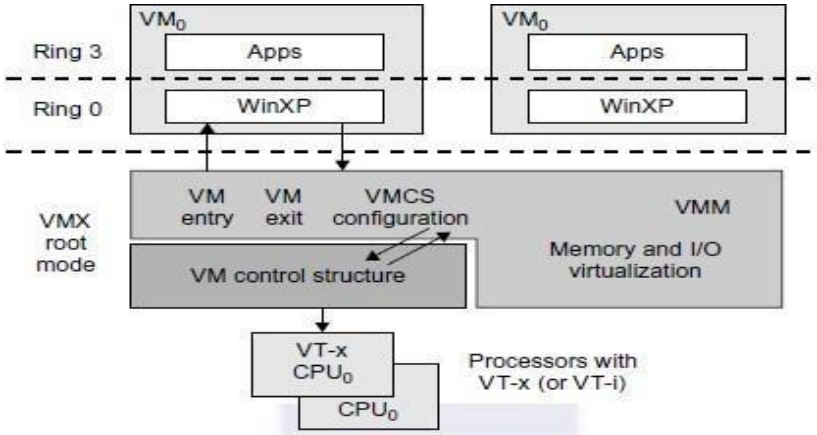


FIGURE 3.11

Intel hardware-assisted CPU virtualization.

(Modified from [68], Courtesy of Lizhong Chen, USC)

- Two-stage mapping process should be maintained by the guest OS and the VMM, respectively: virtual memory to physical memory and physical memory to machine memory.

- The VMM is responsible for mapping the guest physical memory to the actual machine memory.

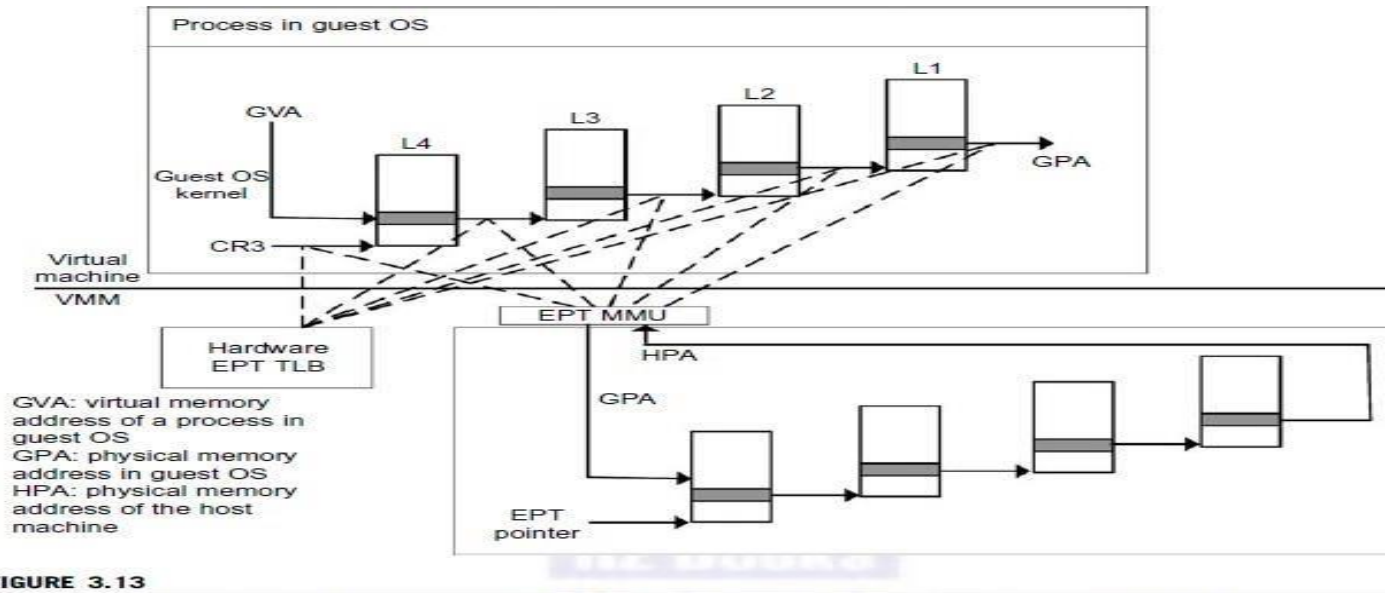


FIGURE 3.13
Memory virtualization using EPT by Intel (the EPT is also known as the shadow page table [68]).

I/O Devices-Virtual Clusters and Data Centers

- I/O Virtualization managing the routing of I/O requests between virtual devices and the shared physical hardware
- managing the routing of I/O requests between virtual devices and the shared physical hardware
- Full device emulation emulates well-known, real-world devices All the functions of a device or bus infrastructure, such as device enumeration, identification, interrupts, and DMA, are replicated in software. This software is located in the VMM and acts as a virtual device
- Two-stage mapping process should be maintained by the guest OS and the VMM, respectively: virtual memory to physical memory and physical memory to machinememory.
- The VMM is responsible for mapping the guest physical memory to the actual machine memory.

Virtualization in Multi-Core Processors

- **Muti-core virtualization** has raised some new **challenges**
- **Two difficulties**: Application programs must be parallelized to use all cores fully, andsoftware must explicitly
- Assign tasks to the cores, which is a very complex problem

- The **first challenge**, new programming models, languages, and libraries are needed to make parallel programming easier.
- The **second challenge** has spawned research involving scheduling algorithms and resource management policies
- **Dynamic heterogeneity** is emerging to mix the fat CPU core and thin GPU cores on thesame chip

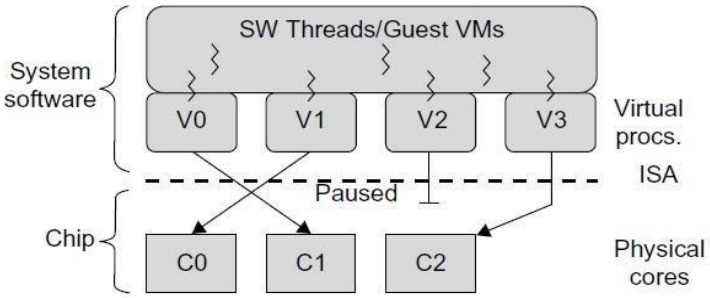


FIGURE 3.16

Multicore virtualization method that exposes four VCPUs to the software, when only three cores are actually present.

(Courtesy of Wells, et al. [74])

- In many-core chip multiprocessors (CMPs), Instead of supporting time-sharing jobs on one or a few cores, use the abundant cores space-sharing, where single-threaded or multithreaded jobs are simultaneouslyassigned to separate groups of cores.

Physical versus Virtual Clusters

- Virtual clusters are built with VMs installed at distributed servers from one or more physical clusters.
- Assign tasks to the cores, which is a very complex problem
- Fast deployment
- High-Performance Virtual Storage
- reduce duplicated blocks

Virtual Clusters and Resource Management

- **Four ways to manage a virtual cluster**.
- First, you can use a **guest-based manager**, by which the cluster manager resides on a guestsystem.

- The **host-based manager** supervises the guest systems and can restart the guest system on another physical machine
- Third way to manage a virtual cluster is to use an **independent cluster manager** on both the host and guest systems.
- Finally, use an **integrated cluster** on the guest and host systems.
- This means the manager must be designed to distinguish between virtualized resources and physical resources

Virtualization for data-center automation

- **Data-center automation** means that huge volumes of hardware, software, and database resources in these data centers can be allocated dynamically to millions of Internet users simultaneously, with guaranteed QoS and cost-effectiveness.
- This **automation** process is triggered by the growth of virtualization products and cloud computing services.
- The latest virtualization development highlights high availability (HA), backup services, workload balancing, and further increases in client bases.

Server Consolidation in Data Centers

- Heterogeneous workloads -chatty workloads and non interactive workloads
- **Server consolidation** is an approach to improve the low utility ratio of hardware resources by reducing the number of physical servers

Virtual Storage Management

storage virtualization has a different meaning in a system virtualization environment system virtualization, virtual storage includes the storage managed by VMMs and guest OSes data stored in this environment can be classified into two categories: VM images and application data.

Cloud OS for Virtualized Data Centers

- Data centers must be virtualized to serve as cloud providers
- Eucalyptus for Virtual Networking of Private Cloud :
- Eucalyptus is an **open source software system** intended mainly for supporting

Infrastructure as a Service (IaaS) clouds

- The system primarily supports **virtual networking** and the management of **VMs**;
- virtual storage is not supported.
- Its purpose is to build **private clouds**

Three resource managers

- Instance Manager
- Group Manager
- Cloud Manager

UNIT-III

InfraStruture as a service (IAAS) &Platform(PAAS): Virtual machines provisioning and Migrationservices – Virtual Machines Provisioning and Manageability – Virtual machine Migration Services – VM Provisioning and Migration in Action.On the Management of Virtual machines for cloud Infrastructures. Aneka – Integration of Private and Public Clouds.

INFRASTRUCTURE AS SERVICE PROVIDERS (IAAS)

- Public Infrastructure as a Service providers commonly offer virtual servers containingone or more CPUs, OS, software stack, storage space and communication facilities
- The most relevant features are:

(i) Geographic distribution of data centers.

(ii) Variety of user interfaces and APIs to access the system.

(iii) Specialized components and services that aid particular applications. (e.g.,load balancers)

(iv) Choice of virtualization platform and operating systems.

(v) Different billing methods and period.

Geographic Presence: To improve availability and responsiveness, a provider of worldwide services would typically build several data centers distributed around the world.

- Availability zones are "distinct locations that are engineered to be insulated from failures in other availability zones and provide inexpensive, low-latency network connectivity to other availability zones in the same region."

User Interfaces and Access to Servers:

- Ideally, a public IaaS provider must provide multiple access means to its cloud, thus catering for various users and their preferences.
- Different types of user interfaces (UI) provide different levels of abstraction, the most common being
    - graphical user interfaces (GUI),
    - command-line tools (CLI), and
    - Web service (WS) APIs.

Advance Reservation of Capacity:

- **Advance reservations** allow users to request for an IaaS provider to reserve resourcesfor a specific time frame in the future, thus ensuring that cloud resources will be availableat that time

- Amazon Reserved Instances is a form of advance reservation of capacity, allowing users to pay a fixed amount of money in advance to guarantee resource availability Automatic Scaling and Load Balancing.

- As mentioned earlier in this chapter, elasticity is a key characteristic of the cloudcomputing model.
- Applications often need to scale up and down to meet varying load conditions.

Service-Level Agreement:

- Service-level agreements (SLAs) are offered by IaaS providers to express theircommitment to delivery of a certain QoS.
- To customers it serves as a warranty.
  - Amazon EC2 states that "if the annual uptime Percentage for a customer drops below 99.95% for the service year, that customer is eligible to receive a service credit equal to 10% of their bill.3"

Hypervisor and Operating System Choice:

- Traditionally, IaaS offerings have been based on heavily customized open-source Xen deployments.
- IaaS providers needed expertise in Linux, networking, virtualization, metering, resource management, and many other low-level aspects to successfully deploy and maintain their cloud offerings.

Public Cloud and Infrastructure Services

- **Public cloud or external cloud**
- Describes cloud computing - resources are dynamically provisioned via publicly accessible Web applications/Web services (SOAP or RESTful interfaces) from an off-site third-party provider,
- Who shares resources and bills on a fine-grained utility computing basis,
- The user pays only for the capacity of the provisioned resources at a particular time.

Amazon Elastic Compute Cloud (EC2) is an IaaS service that provides elastic compute capacity in the cloud

- A private cloud aims at providing public cloud functionality, but on private resources, while maintaining control over an organization's data and resources to meet security and governance's requirements in an organization.
- Private clouds exhibit the following characteristics:
  - Allow service provisioning and compute capability for an organization's users in a self-service manner.
  - Automate and provide well-managed virtualized environments.
  - Optimize computing resources, and servers' utilization.
  - Support specific workloads.
- Examples are Eucalyptus and OpenNebula.

"Hybrid cloud"

- In which a combination of private/internal and external cloud resources exist together by enabling outsourcing of noncritical services and functions in public cloud and keeping the critical ones internal.
- Release resources from a public cloud and to handle sudden demand usage, which called "cloud bursting".
- Cloud and Virtualization Standardization Efforts
- Standardization is important to ensure interoperability between virtualization mangement vendors, the virtual machines produced by each one of them, and cloud computing
- Distributed Management Task Force(DMTF) initiated the VMAN (Virtualization Management Initiative), delivers broadly supported interoperability and portability standards for managing the virtual computing lifecycle.

OVF (Open Virtualization Format)

- VMAN's OVF (Open Virtualization Format) in a collaboration between industry key players: Dell, HP, IBM, Microsoft, Xen Source, and Vmware.
- OVF specification provides a common format to package and securely distribute virtual appliances across multiple virtualization platforms.
- VMAN profiles define a consistent way of managing a heterogeneous virtualized

environment.

- Standardization effort has been initiated by Open Grid Forum (OGF) through organizing an official new working group to deliver a standard API for cloud IaaS, the Open Cloud Computing Interface Working Group (OCCIWG).

VIRTUAL MACHINES PROVISIONING:

- Typical life cycle of VM and its major possible states of operation, which make the management and automation of VMs in virtual and cloud environments easier

Process:

- Steps to Provision VM. Here, we describe the common and normal steps of provisioning a virtual server:
- Firstly, you need to select a server from a pool of available servers (physical servers with enough capacity) along with the appropriate OS template you need to provision the virtual machine.
- Secondly, you need to load the appropriate software (operating system you selected in the previous step, device drivers, middleware, and the needed applications for the service required).

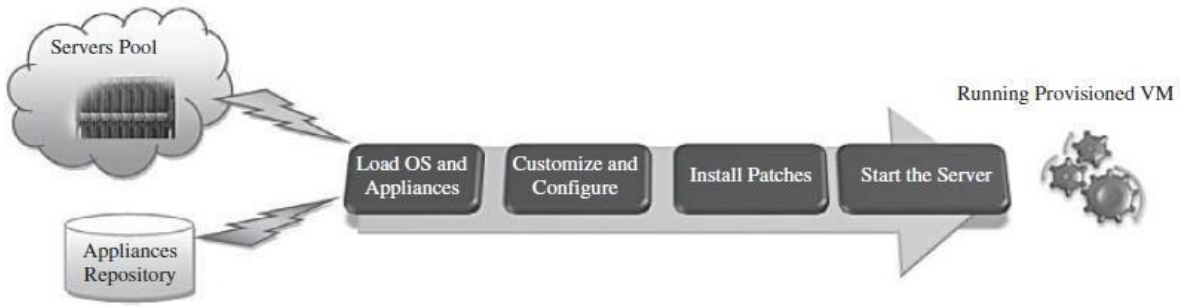


FIGURE 5.4. Virtual machine provision process.

- Thirdly, you need to customize and configure the machine (e.g., IP address, Gateway) to configure an associated network and storage resources.
- Finally, the virtual server is ready to start with its newly loaded software

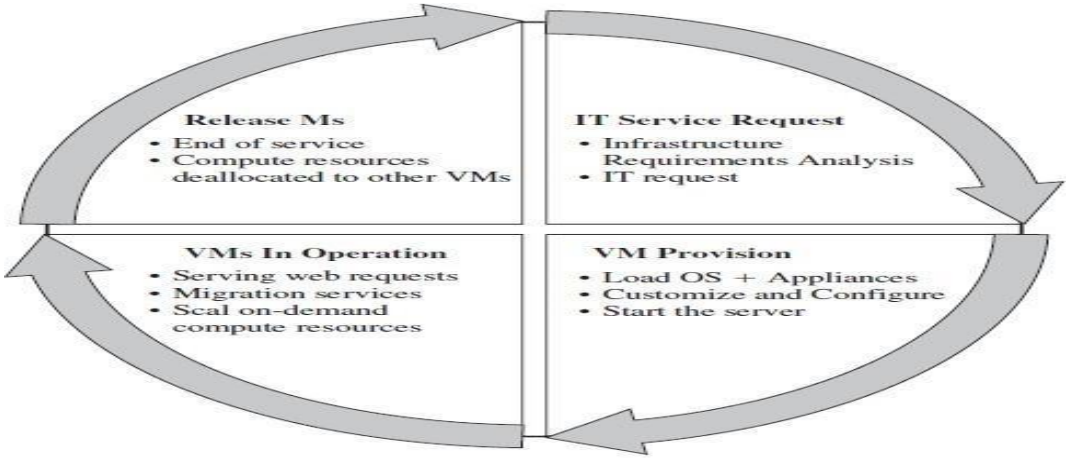


FIGURE 5.3. Virtual machine life cycle.

VIRTUAL MACHINE MIGRATION SERVICES:

Migration service:

- In the context of virtual machines, is the process of moving a virtual machine from onehost server or storage location to another.
- There are different techniques of VM migration, hot/life migration, cold/regular migration, and live storage migration of a virtual machine.

VM Migration, SLA and On-Demand Computing:

- virtual machines' migration plays an important role in data centers once it has been detected that a particular VM is consuming more than its fair share ofresources at the expense of other VMs on the same host,
- It will be eligible, for this machine, to either be moved to another underutilized host or assign more resources for it.
- There should be an integration between virtualization's management tools (with its migrations and performance's monitoring capabilities), and SLA's management tools to achieve balance in resources by migrating and monitoring the workloads, and accordingly, meeting the SLA.

On the Management of Virtual Machines for Cloud Infrastructures:

- Advantages of having facility in data center's technologies is to have the **ability to migrate virtual machines** from one platform to another

- For example, the **VMware converter that handles migrations** between ESX hosts;

- The VMware server; and the VMware workstation.
- The VMware converter can also import from other virtualization platforms, such as Microsoft virtual server machines

Deployment Scenario:

- ConVirt deployment consists of at least one ConVirt workstation,Where ConVirt is installed and ran, which provides the main console for managing the VM life cycle, managing images, provisioning new VMs, monitoring machine resources, and so on.
- There are two essential deployment scenarios for ConVirt:
- A, **basic configuration** in which the Xen or KVM virtualization platform is on the local machine, where ConVirt is already installed; B,
- An **advanced configuration** in which the Xen or KVM is on one or more remote servers.

Installation:

The installation process involves the following:

- Installing ConVirt on at least one computer.
- Preparing each managed server to be managed by ConVirt.
- We have two managing servers with the following Ips(managed server 1, IP:172.16.2.22; and managed server 2, IP:172.16.2.25) as shown in the deployment diagram

Environment, Software, and Hardware. ConVirt 1.1, Linux Ubuntu 8.10, three machines, Dell core 2 due processor, 4G RAM.

- Adding Managed Servers and Provisioning VM.
- Once the installation is done and you are ready to manage your virtual infrastructure, then you can start the ConVirt management console :
- Select any of servers' pools existing (QA Lab in our scenario) and on its context menu,select "Add Server."
- You will be faced with a message asking about the virtualization platform you want to manage (Xen or KVM), as shown in Figure
- Choose KVM, and then enter the managed server information and credentials (IP,username, and password) as shown in Figure
- Once the server is synchronized and authenticated with the management console, it will appear in the left pane/of the ConVirt, **Live Migration Effect** on a Running Web Server.
- Clark et al. did evaluate the above migration on an Apache 1.3 Web server; this servedstatic content at a high rate, as illustrated in Figure 5.6.

- The throughput is achieved when continuously serving a single 512-kB file to a set of one hundred concurrent clients.
- This simple example demonstrates that a highly loaded server can be migrated with both controlled impact on live services and a short downtime

VMware Vmotion.

- This allows users to

(a) automatically optimize and allocate an entire pool of resources for maximum hardware utilization, flexibility, and availability and

(b) perform hardware's maintenance without scheduled downtime along with migrating virtual machines away from failing or underperforming servers.

Citrix XenServer XenMotion.

- This is a nice feature of the Citrix XenServer product, inherited from the Xen live migrate utility, which provides the IT administrator with the facility to move a running VM from one XenServer to another in the same pool without interrupting the service

Regular/Cold Migration.

Cold migration is the migration of a powered-off virtual machine.

- Main differences between live migration and cold migration are that
- 1) live migration needs a shared storage for virtual machines in the server's pool, butcold migration does not;
- 2) live migration for a virtual machine between two hosts, there would be certain CPU compatibility checks to be applied; while in cold migration this checks do not apply
- The cold migration process (VMware) can be summarized as follows:
    - The configuration files, including the NVRAM file (BIOS settings), log files, as well as the disks of the virtual machine, are moved from the source host to the destination host's associated storage area.
    - The virtual machine is registered with the new host.
    - After the migration is completed, the old version of the virtual machine is deleted from the source host.

Aneka – Integration of Private and Public clouds

- Manjrasoft Aneka is a .NET-based platform and framework designed for building and deploying distributed applications on clouds.

It provides a set of APIs for transparently exploiting distributed resources and expressing the business logic of applications by using the preferred programming abstractions.

- Aneka also provides support for deploying and managing clouds.
- By using its Management Studio and a set of Web interfaces, it is possible to set up either public or private clouds, monitor their status, update their configuration, and perform the basic management operations.

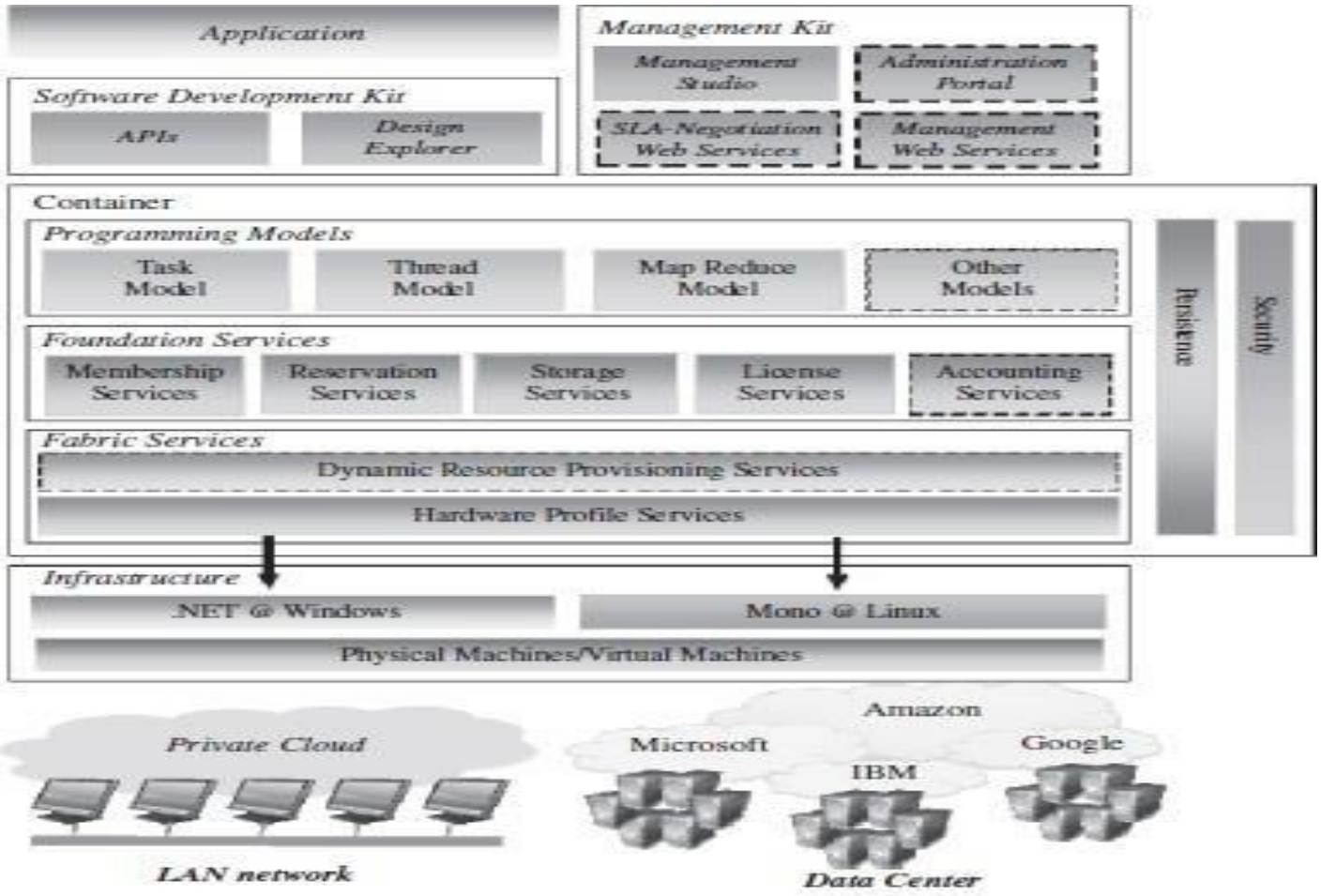


FIGURE 5.22. Manjras oft Aneka layered architecture [10].

UNIT- IV

Software as a Service (SAAS) & Data Security in the Cloud: Software as a Service (SAAS),Google App Engine-Centralizing Email Communications- Collaborating via Web-Based Communication Tools-An Introduction to the idea of Data Security.The Current State of Data Security in the Cloud-Cloud Computing and Data Security Risk – Cloud Computing and Identity

Software as a Service (SAAS) & Data Security in the Cloud:

Software as a Service (SAAS):

- SaaS is a model of software deployment where an application is hosted as a service provided to customers across the Internet.
- Saas alleviates the burden of software maintenance/support but users relinquish control over software versions and requirements

SaaS Maturity Model

Level 1: Ad-Hoc/Custom – One Instance per customer

Level 2: Configurable per customer

Level 3: configurable & Multi-Tenant-Efficient

Level 4: Scalable, Configurable & Multi-Tenant-Efficient

SaaS INTEGRATION PRODUCTS AND PLATFORMS

- Cloud-centric integration solutions are being developed and demonstrated for showcasing their capabilities for integrating enterprise and cloud applications.

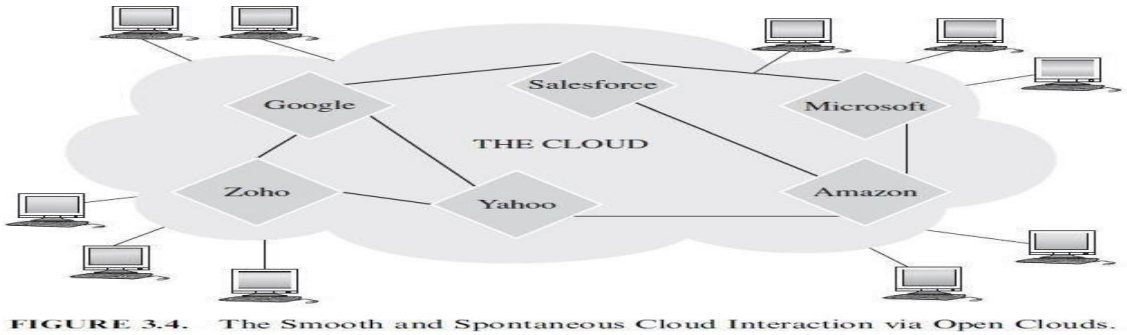


FIGURE 3.4. The Smooth and Spontaneous Cloud Interaction via Open Clouds.

- Composition and collaboration will become critical and crucial for the mass adoption of clouds

Jitterbit:

- Jitterbit is a fully graphical integration solution that provides users a versatile platform suite of productivity tools to reduce the integration efforts sharply.
- Jitterbit can be used standalone or with existing EAI infrastructures Help us quickly design, implement, test, deploy, and manage the integration projectsTwo major components :
- Jitterbit Integration Environment
- An intuitive point-and-click graphical UI that enables to quickly configure, test, deployand manage integration projects on the Jitterbit server.
- Jitterbit Integration Server
- A powerful and scalable run-time engine that processes all the integration operations, fully configurable and manageable from the Jitterbit application.

Linkage with On premise and on demand Applications

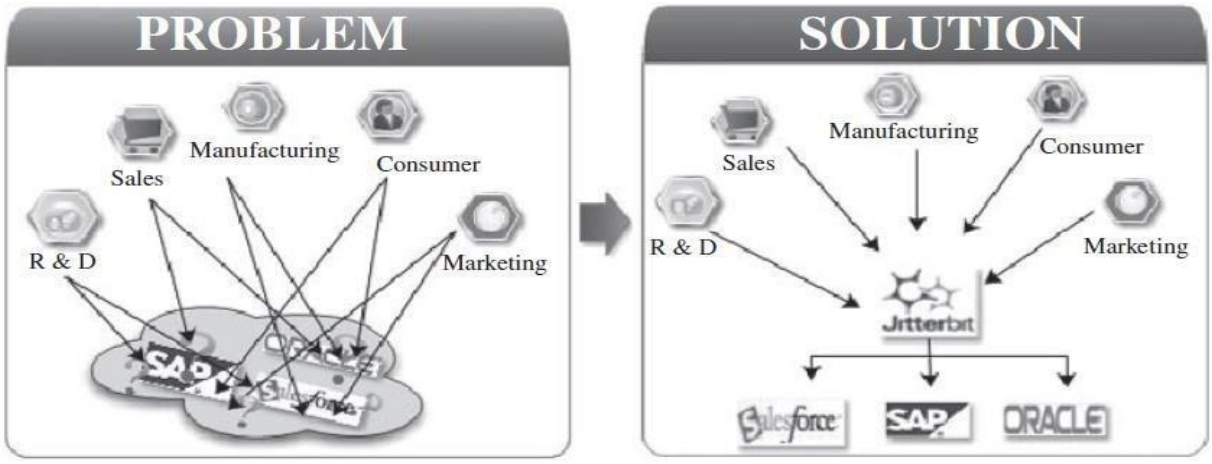


FIGURE 3.5. Linkage of On-Premise with Online and On-Demand Applications.

Google APP Engine:

- The app engine is a Cloud-based platform, is quite comprehensive and combines infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS).
- The app engine supports the delivery, testing and development of software on demand ina Cloud computing environment that supports millions of users and is highly scalable.
- The company extends its platform and infrastructure to the Cloud through its app engine. It presents the platform to those who want to develop SaaS solutions at competitive costs.

Google is a leader in web-based applications,

so it's not surprising that the company also offers cloud development services.

- These services come in the form of the Google App Engine, which enables developers to build their own web applications utilizing the same infrastructure that powers Google's powerful applications.
- The Google App Engine provides a fully integrated application environment. Using Google's development tools and computing cloud, App Engine applications are easy to build, easy to maintain, and easy to scale.

Features of App Engine

- These are covered by the depreciation policy and the service-level agreement of the app engine. Any changes made to such a feature are backward-compatible and implementation of such a feature is usually stable. These include data storage, retrieval, and search; communications; process management; computation; app configuration and management.
- Data storage, retrieval, and search include features such as HRD migration tool, Google CloudSQL, logs, datastore, dedicated Memcache, blobstore, Memcache and search.
- Communications include features such as XMPP. channel, URL fetch, mail, and Google CloudEndpoints.
- Process management includes features like scheduled tasks and task queue
- Computation includes images.
- App management and configuration cover app identity, users, capabilities, traffic splitting, modules, SSL for custom domains, modules, remote access, and multi tenancy.

Centralizing email Communications:

- The key here is to enable anywhere/anytime access to email.
- Pre-cloud computing, your email access was via a single computer, which also stored all your email messages. For this purpose, you probably used a program like Microsoft Outlook or Outlook Express, installed on your home computer.
- To check your home email from work, it took a bit of juggling and perhaps the use of your ISP's email access web page. That web page was never in sync with the messages

  on your home PC, of course, which is just the start of the problems with trying to communicate in this fashion.

- A better approach is to use a web-based email service, such as Google's Gmail (mail.google.com), Microsoft's Windows Live Hotmail (mail.live.com), or Yahoo! Mail (mail.yahoo.com). These services place your email inbox in the cloud; you can access it from any computer connected to the Internet.

Collaborating via Web-Based Communication Tools:

GMAIL

- Gmail offers a few unique features that set it apart from the web-based email crowd.
- First, Gmail doesn't use folders. With Gmail you can't organize your mail into folders, as you can with the other services.
- Instead, Gmail pushes the search paradigm as the way to find the messages you want— not a surprise, given Google's search-centric business model.
- Gmail does, however, let you "tag" each message with one or more labels. This has the effect of creating virtual folders, as you can search and sort your messages by any of their labels.
- In addition, Gmail groups together related email messages in what Google calls conversations.

Yahoo! Mail Yahoo! Mail (mail.yahoo.com)

- is another web mail service, provided by the popular Yahoo! search site.
- The basic Yahoo! Mail is free and can be accessed from any PC, using any web browser.
- Yahoo! also offers a paid service called Yahoo! Mail Plus that lets you send larger messages and offers offline access to your messages via POP email clients

Web Mail Services

- AOL Mail (mail.aol.com)
- BigString (www.bigstring.com) E
- xcite Mail (mail.excite.com)
- FlashMail (www.flashmail.com)
- GMX Mail (www.gmx.com)
- Inbox.com (www.inbox.com)
- Lycos Mail (mail.lycos.com)
- Mail.com (www.mail.com)

- Zoho Mail (zoho.mail.com)

An Introduction to the idea of Data Security:

- Information in a cloud environment has much more dynamism and fluidity than information that is static on a desktop or in a network folder
- Nature of cloud computing dictates that data are fluid objects, accessible from a multitude of nodes and geographic locations and, as such, must have a data security methodology that takes this into account while ensuring that this fluidity is not compromised
- The idea of content-centric or information-centric protection, being an inherent part of a data object is a development out of the idea of the "de-perimerization" of the enterprise.
- This idea was put forward by a group of Chief Information Officers (CIOs) who formed an organization called the Jericho Forum

The Current State of data Security in the Cloud:

When it comes to data, the cloud poses a variety of risks that the enterprise must address as part of its security strategy. The biggest risks—as organizations increasingly rely on the cloud for collecting, storing, and processing critical data—are cyberattacks and data breaches.

A SailPoint survey, for example, found that 45% of companies that have implemented IaaS have experienced cyberattacks and 25% have experienced a data breach. Other research found that IT security professionals cite the proliferation of cloud services as the second-biggest barrier to their ability to respond to a data breach, and this challenge has grown in recent years.

Some of the common cloud-related risks that organizations face include:

- **Regulatory noncompliance**—whether it's the General Protection Data Regulation (GDPR) or the Healthcare Insurance Portability and Accountability Act (HIPAA), cloud computing adds complexity to satisfying compliance requirements.
- **Data loss and data leaks—**data loss and data leaks can result from poor security practices such as misconfigurations of cloud systems or threats such as insiders.
- **Loss of customer trust and brand reputation**—customers trust organizations to safeguard their personally identifiable information (PII) and when a security incident leads to data compromise, companies lose customer goodwill.
- **Business interruption—**risk professionals around the globe identified business disruption caused by failure of cloud technology / platforms or supply chains as one of their top five cyber exposure concerns.

- **Financial losses**—the costs of incident mitigation, data breaches, business disruption, and other consequences of cloud security incidents can add up to hundreds of millions of dollars.

Cloud Computing and Data Security Risk:

- Cloud computing is a development that is meant to allow more open accessibility and easier and improved data sharing.

- Data are uploaded into a cloud and stored in a data center, for access by users from that data center; or in a more fully cloud-based model, the data themselves are created in the cloud and stored and accessed from the cloud (again via a data center).

- The most obvious risk in this scenario is that associated with the storage of that data. A user uploading or creating cloud-based data include those data that are stored and maintained by a third-party cloud provider such as Google, Amazon, Microsoft, and so on.

This action has several risks associated with it:

- Firstly, it is necessary to protect the data during upload into the data center to ensure that the data do not get hijacked on the way into the database.
- Secondly, it is necessary to the stores the data in the data center to ensure that they are encrypted at all times.
- Thirdly, and perhaps less obvious, the access to those data need to be controlled; this control should also be applied to the hosting company, including the administrators of the data center.
- In addition, an area often forgotten in the application of security to a data resource is the protection of that resource during its use Data security risks are compounded by the open nature of cloud computing.

- Access control becomes a much more fundamental issue in cloud-based systems because of the accessibility of the data

- Information-centric access control (as opposed to access control lists) can help to balance improved accessibility with risk, by associating access rules with different data objects within an open and accessible platform, without losing the Inherent usability of that platform

- A further area of risk associated not only with cloud computing, but also with traditional

network computing, is the use of content after access.

- The risk is potentially higher in a cloud network, for the simple reason that the information is outside of your corporate walls.

CLOUD COMPUTING AND IDENTITY:

Digital identity

- holds the key to flexible data security within a cloud Environment.
- A digital identity represents who we are and how we interact with others on-line.
- **Access, identity, and risk** are three variables that can become inherently connected when applied to the security of data, because access and risk are directly proportional: As access increases, so then risk to the security of the data increases.
- Access controlled by identifying the actor attempting the access is the most logical manner of performing this operation.
- Ultimately, digital identity holds the key to securing data, if that digital identity can be programmatically linked to security policies controlling the post-access usage of data.

Identity, Reputation, and Trust

- Reputation is a real-world commodity; that is a basic requirement of human-to-human relationships
- Our basic societal communication structure is built upon the idea of reputation and trust.
- Reputation and its counter value, trust, is easily transferable to a digital realm:eBay, for example, having partly built a successful business model on the strength of a ratings system, builds up the reputation of its buyers and sellers through successful (or unsuccessful) transactions.
- These types of reputation systems can be extremely useful when used with a digital identity.
- They can be used to associate varying levels of trust with that identity, which in turn can be used to define the level (granular variations) of security policy applied to data.

User-Centric Identity:

- Digital identities are a mechanism for identifying an individual, particularly within a cloud environment, identity ownership being placed upon the individual is known as user- centric identity.

- It allows users to consent and control how their identity (and the individual identifiersmaking up the identity, the claims) is used.
- This reversal of ownership away from centrally managed identity platforms(enterprise-centric) has many advantages.
- This includes the potential to improve the privacy aspects of a digital identity, by giving an individual the ability to apply permission policies based on their identity and to control which aspects of that identity are divulged.
- An identity may be controllable by the end user, to the extent that the user can thendecide what information is given to the party relying on the identity.

Information Card:

- Information cards permit a user to present to a Web site or other service (relying party) one or more claims, in the form of a software token, which may be used to uniquely identify that user.
- They can be used in place of user name/ passwords, digital certificates, and other identification systems, when user identity needs to be established to control access to a Web site or other resource, or to permit digital signing

Information cards are part of an identity meta-system consisting of:

1. **Identity providers (IdP**), who provision and manage information cards,with specificclaims, to users.
2. **Users** who own and utilize the cards to gain access to Web sites and other resourcesthat support information cards.
3. **An identity selector/service**, which is a piece of software on the user's desktop or in the cloud that allows a user to select and manage their cards.
4. **Relying parties.** These are the applications, services, and so on, that can use an Information card to authenticate a person and to then authorize an action such as loggingonto a Web site, accessing a document, signing content, and so on.

Each information card is associated with a set of claims which can be used to identify the user.These claims include identifiers such as name, email address, post code

- Information cards are built around a set of open standards devised by a consortium that includes Microsoft, IBM, Novell, and so on.
- The original remit of the cards was to create a type of single sign on system for the Internet, to help users to move away from the need to remember multiple passwords.
- However, the information card system can be used in many more ways.
- Because an information card is a type of digital identity, it can be used in the same waythat other digital identities can be used.

For example, an information card can be used to digitally sign data and content and to control access to data and content. One of the more sophisticated uses of an information card is the advantage given to the cards by way of the claims system.

Data-centric mashups :

- Those are used to perform business processes around data creation and dissemination by their very nature, can be used to hijack data, leaking sensitive information and/or affecting integrity of that data

- Cloud computing, more than any other form of digital communication technology, has created a need to ensure that protection is applied at the inception of the information, in a content centric manner, ensuring that a security policy becomes an integral part of that data throughout its life cycle.

Encryption

- It is a vital component of the protection policy, but further controls over the access of that data and on the use of the data must be met.

- In the case of mashups, the controlling of access to data resources, can help to all eviate the security concerns by ensuring that mashup access is authenticated.

- Linking security policies, as applied to the use of content, to the access control method offer a way of continuing protection of data, post access and throughout the life cycle; this type of data security philosophy must be incorporated into the use of cloud computing to alleviate security risks.

UNIT-V

SLA Management in cloud computing: Traditional Approaches to SLO Management, Types of SLA, Life Cycle of SLA, SLA Management in Cloud.

SLA MANAGEMENT IN CLOUD COMPUTING:

In the early days of web-application deployment, performance of the application at peak load was a single important criterion for provisioning server resources [1]. Provisioning in those days involved deciding hardware configuration, determining the number of physical machines, and acquiring them upfront so that the overall business objectives could be achieved. The web applications were hosted on these dedicated individual servers within enterprises‘own server rooms. These web applications were used to provide different kinds of e-services to various clients. Typically, the service-level objectives (SLOs) for these applications were response time and throughput of the application end-user requests. The capacity buildup was to cater to the estimated peak load experienced by the application. The activity of determining the number of servers and their capacity that could satisfactorily serve the application end-user requests at peak loads is called capacity planning [1]. An example scenario where two web applications, application A and application B, are hosted on a separate set of dedicated servers within the enterprise-owned server rooms is shown in Figure16.1. The planned capacity for each of the applications to run successfully is three servers. As the number of web applications grew, the server rooms in the organization became large and such server rooms were known as data centers. These data centers were owned and managed by the enterprises themselves.

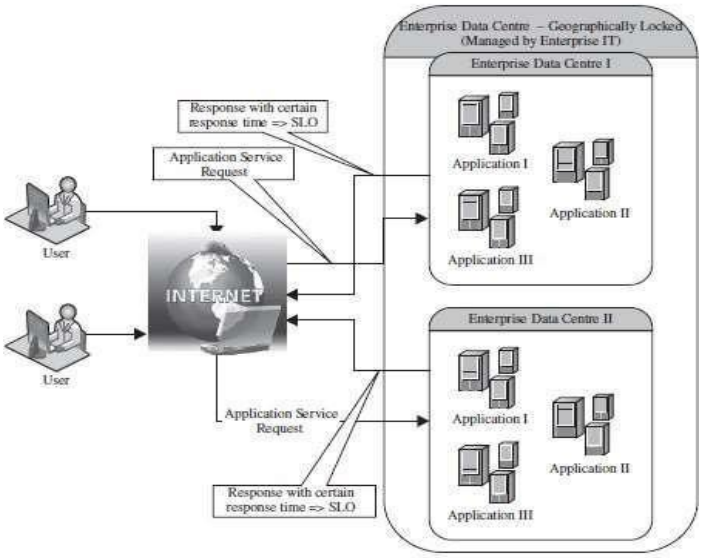


FIGURE 16.1. Hosting of applications on servers within enterprise's data centers.

TRADITIONAL APPROACHES TO SLO MANAGEMENT

Traditionally, load balancing techniques and admission control mechanisms have been used to provide guaranteed quality of service (QoS) for hosted web applications. These mechanisms can be viewed as the first attempt towards managing the SLOs. In the following subsections we discuss the existing approaches for load balancing and admission control for ensuring QoS.16.2.1 Load Balancing The objective of a load balancing is to distribute the incoming requests onto a set of physical machines, each hosting a replica of an application, so that the load on the machines is equally distributed [4]. The load balancing algorithm executes on a physical machine that interfaces with the clients. This physical machine, also called the front- end node, receives the incoming requests and distributes these requests to different physical machines for further execution.

This set of physical machines is responsible for serving the incoming requests and are known as the back-end nodes.

TYPES OF SLA:

Service-level agreement provides a framework within which both seller and buyer of a service can pursue a profitable service business relationship. It outlines the broad understanding between the service provider and the service consumer for conducting business and forms the basis for maintaining a mutually beneficial relationship. From a legal perspective, the necessary terms and conditions that bind the service provider to provide services continually to the service consumer are formally defined in SLA.

SLA can be modeled using web service-level agreement (WSLA) language specification [7]. Although WSLA is intended for web-service-based applications, it is equally applicable for hosting of applications. Service-level parameter, metric, function, measurement directive, service-level objective, and penalty are some of the important components of WSLA and are described in Table 16.1.

TABLE 16.1. Key Components of a Service-Level Agreement

Service-Level Parameter	Describes an observable property of a service whose value is measurable.
Metrics	These are definitions of values of service properties that are measured from a service-providing system or computed from other metrics and constants. Metrics are the key instrument to describe exactly what SLA parameters mean by specifying how to measure or compute the parameter values.
Function	A function specifies how to compute a metric's value from the values of other metrics and constants. Functions are central to describing exactly how SLA parameters are computed from resource metrics.
Measurement directives	These specify how to measure a metric.

.

Infrastructure SLA. The infrastructure provider manages and offers guarantees on availability of the infrastructure, namely, server machine, power, network connectivity, and so on. Enterprises manage themselves, their applications that are deployed on these servermachines. The machines are leased to the customers and are isolated from machines of other customers. In such dedicated hosting environments, a practical example of service-level guarantees offered by infrastructure providers is shown in Table 16.2.

Application SLA. In the application co-location hosting model, the server capacity is available to the applications based solely on their resource demands. Hence, the service providers are flexible in allocating and de-allocating computing resources among the co- located applications.

Therefore, the service providers are also responsible for ensuring to meet their customer's application SLOs. For example, an enterprise can have the following application SLA with a service provider for one of its application.

LIFE CYCLE OF SLA:

Each SLA goes through a sequence of steps starting from identification of terms and conditions, activation and monitoring of the stated terms and conditions, and eventual termination of contract once the hosting relationship ceases to exist. Such a sequence of steps is called SLA life cycle and consists of the following five phases:

1. Contract definition
2. Publishing and discovery
3. Negotiation
4. Operationalization
5. De-commissioning

Here, we explain in detail each of these phases of SLA life cycle.

Contract Definition. Generally, service providers define a set of service offerings and corresponding SLAs using standard templates. These service offerings form a catalog. Individual SLAs for enterprises can be derived by customizing these base SLA templates.

Publication and Discovery. Service provider advertises these base service offerings through standard publication media, and the customers should be able to locate the service provider by searching the catalog. The customers can search different competitive offerings and shortlist a few that fulfill their requirements for further negotiation.

Operationalization. SLA operation consists of SLA monitoring, SLA accounting, and SLA enforcement. SLA monitoring involves measuring parameter values and calculating the metrics defined as a part of SLA and determining the deviations. On identifying the deviations, the concerned parties are notified. SLA accounting involves capturing and archiving the SLA adherence for compliance.

As part of accounting, the application‘s actual performance and the performance guaranteed as a part of SLA is reported. Apart from the frequency and the duration of the SLA breach, it should also provide the penalties paid for each SLA violation. SLA enforcement involves taking appropriate action when the runtime monitoring detects a SLA violation. Such actions could be notifying the concerned parties, charging the penalties besides other things. The different policies can be expressed using a subset of the Common Information Model (CIM) [9]. The CIM model is an open standard that allows expressing managed elements of data center via relationships and common objects.

De-commissioning. SLA decommissioning involves termination of all activities performed under a

particular SLA when the hosting relationship between the service provider and the service consumer has ended. SLA specifies the terms and conditions of contract termination and specifies situations under which the relationship between a service provider and a service consumer can be considered to be legally ended.

SLA MANAGEMENT IN CLOUD:

SLA management of applications hosted on cloud platforms involves five phases.

1. Feasibility
2. On-boarding
3. Pre-production
4. Production
5. Termination

Different activities performed under each of these phases are shown in Figure 16.7. These activities are explained in detail in the following subsections.

Feasibility Analysis

MSP conducts the feasibility study of hosting an application on their cloud platforms. This study involves three kinds of feasibility: (1) technical feasibility, (2) infrastructure feasibility, and (3) financial feasibility. The technical feasibility of an application implies determining the following:

1. Ability of an application to scale out.
2. Compatibility of the application with the cloud platform being used within the MSP‘s datacenter.
3. The need and availability of a specific hardware and software required for hosting andrunning of the application.

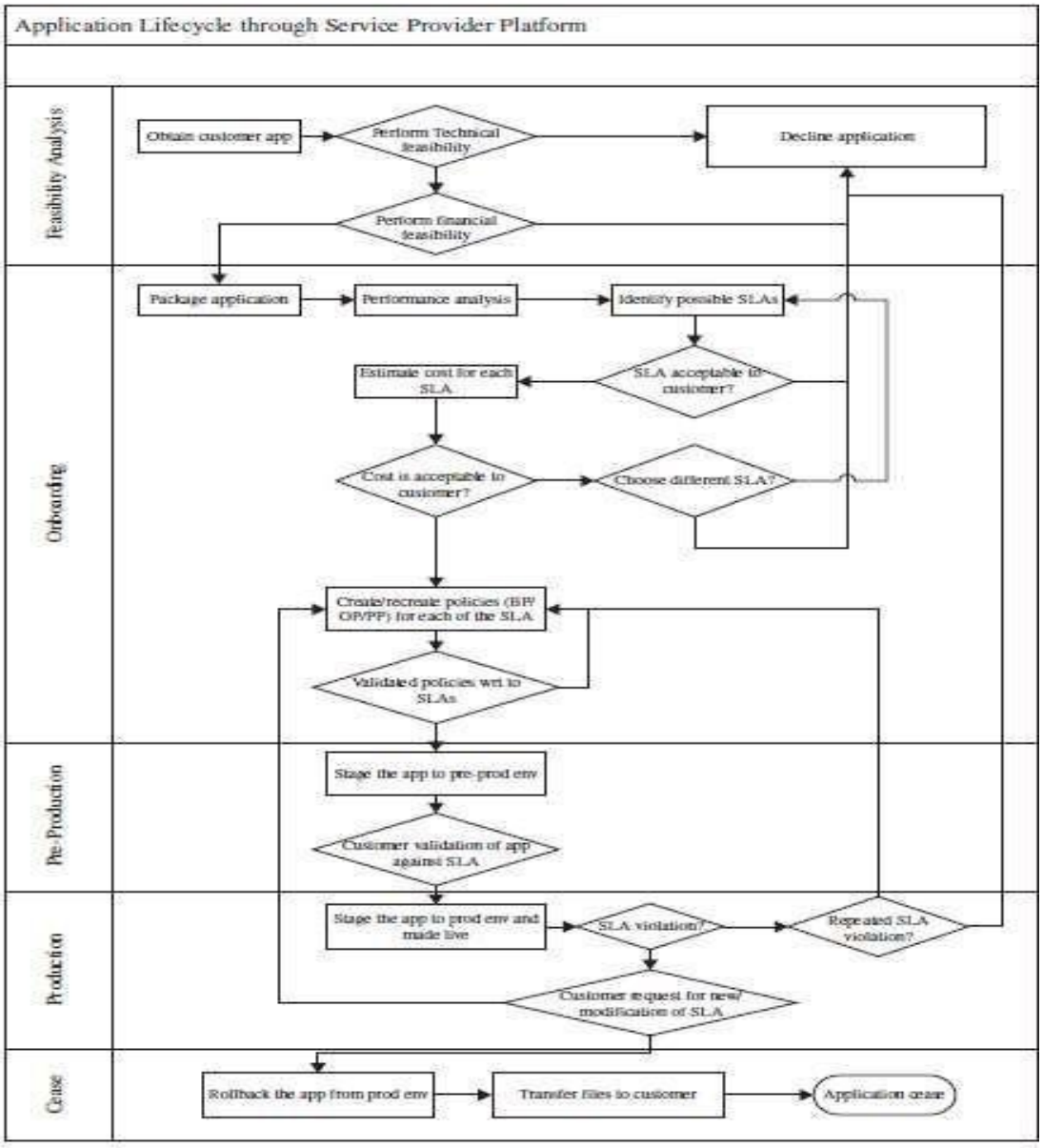


FIGURE 16.7. Flowchart of the SLA management in cloud.

Once the customer and the MSP agree in principle to host the application based on the findings of the feasibility study, the application is moved from the customer servers to the hosting platform. Moving an application to the MSP's hosting platform is called on-boarding [10]. As part of the on-boarding activity, the MSP understands the application runtime characteristics using runtime profilers. This helps the MSP to identify the possible SLAs that can be offered to the customer for that application. This also helps in creation of the necessary policies (also called rule sets) required to guarantee the SLOs mentioned in the application SLA. The application is accessible to its end users only after the on-boarding activity is completed.

Preproduction

Once the determination of policies is completed as discussed in previous phase, the application is hosted in a simulated production environment. It facilitates the customer to verify and validate the MSP's findings on application's runtime characteristics and agree on the defined SLA. Once both parties agree on the cost and the terms and conditions of the SLA, the customer sign-off is obtained. On successful completion of this phase the MSP allows the application to go on-live.

Production

In this phase, the application is made accessible to its end users under the agreed SLA. However, there could be situations when the managed application tends to behave differently in a production environment compared to the preproduction environment. This in turn may cause sustained breach of the terms and conditions mentioned in the SLA. Additionally, customer may request the MSP for inclusion of new terms and conditions in the SLA. If the application SLA is breached frequently or if the customer requests for a new non-agreed SLA, the on-boarding process is performed again. In the case of the former, on-boarding activity is repeated to analyze the application and its policies with respect to SLA fulfillment. In case of the latter, a new set of policies are formulated to meet the fresh terms and conditions of the SLA.

Termination

When the customer wishes to withdraw the hosted application and does not wish to continue to avail the services of the MSP for managing the hosting of its application, the termination activity is initiated. On initiation of termination, all data related to the application are transferred to the customer and only the essential information is retained for legal compliance. This ends the hosting relationship between the two parties for that application, and the customer sign-off is obtained.