Student Manual

STL academy

Subject

Cloud Computing

Vol.01

Empowering Youth!

ISO 21001 | Guinness World Records Record Holder | Asia Book of Records

Skill & Assessment Partner NASSCOM®

stl.tech | stlacad.tech

Cloud Computing

Submitted to :- **Bihar Skill Development Mission, Labour Resources Department, GoB**	Submitted By :- **Sterlite Technologies Ltd**
	Session : 2022-23

Course name:

- Course Id-
- Candidate Eligibility : Diploma/ Graduate
- Course Duration: (In hours) 550

CONTACT DETAILS OF THE BODY SUBMITTING THE QUALIFICATION FILE

Name and address of submitting body:

Sterlite Technologies Ltd

Name and contact details of individual dealing with the submission

Name : Mrs./Mr. Srikant Pattnaik

Position in the organization : Manager

Tel number(s) (Mobile no.) : 9702048264

Website : www.stlacad.tech

E-mail address : srikant.pattnaik@stl.tech

CLOUD COMPUTING

STUDENT GUIDE

About the Student Guide

The student guide contains modules which will help you to acquire relevant knowledge and skills (generic and domain-specific skills) related to the '**Cloud Architect**' job role. Knowledge in each module is easily understood and grasped by you before you move on to the next module. Comprehensible diagrams & images from world of work have been included to bring about visual appeal and to make the text lively and interactive for you. You can also try to create your own illustrations using your imagination or taking the help of your trainer.

Let us now see what the sections in the modules have for you.

Section 1: Learning Outcome

This section introduces you to the learning objectives and knowledge criteria covered in the module. It also tells you what you will learn through the various topics covered in the module.

Section 2: Relevant Knowledge

This section provides you with the knowledge to achieve relevant skill and proficiency to perform tasks of the **Cloud Architect**. The knowledge developed through the module will enable you to perform certain activities related to the job market. You should read through the textual information to develop an understanding on the various aspects of the module before you complete the exercise(s).

Section 3: Exercises

Each module has exercises, which you should practice on completion of the learning sessions of the module. You will perform the activities in the classroom, at home or at the workplace. The activities included in this section will help you to develop necessary knowledge, skills and attitude that you need for becoming competent in performing the tasks at workplace. The activities should be done under the supervision of your trainer who will guide you in completing the tasks and also provide feedback to you for improving your performance.

Section 4: Assessment Questionnaire

The review questions included in this section will help you to check your progress. You must be able to answer all the questions before you proceed to the next module.

CONTENTS

MODULE 1

Introduction to Cloud Computing

Section 1: Learning Outcomes

After completing this module, you will be able to:

- Describe Concept of Cloud Computing
- Explain various layers and types of Cloud Computing
- Differentiate between Cloud Computing and Cloud Services
- Tell the New Technologies that enabled Cloud Computing
- Use Cloud Computing Features and Standards
- Resolve Security Issues
- Describe Key Cloud Computing Platforms
- Explain Cloud Computing Challenges
- Throw a light on Future of Cloud Computing
- Describe the Components of Cloud Computing
- Define Cloud Computing
- Categorize different types of Clouds and Services

Section 2: Relevant Knowledge

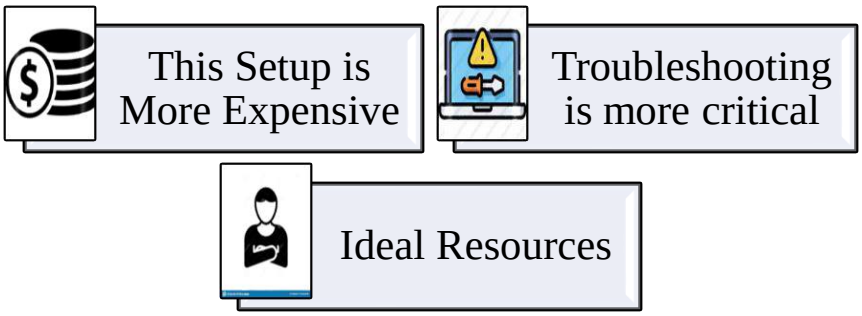
1.1 Introduction to Cloud Computing

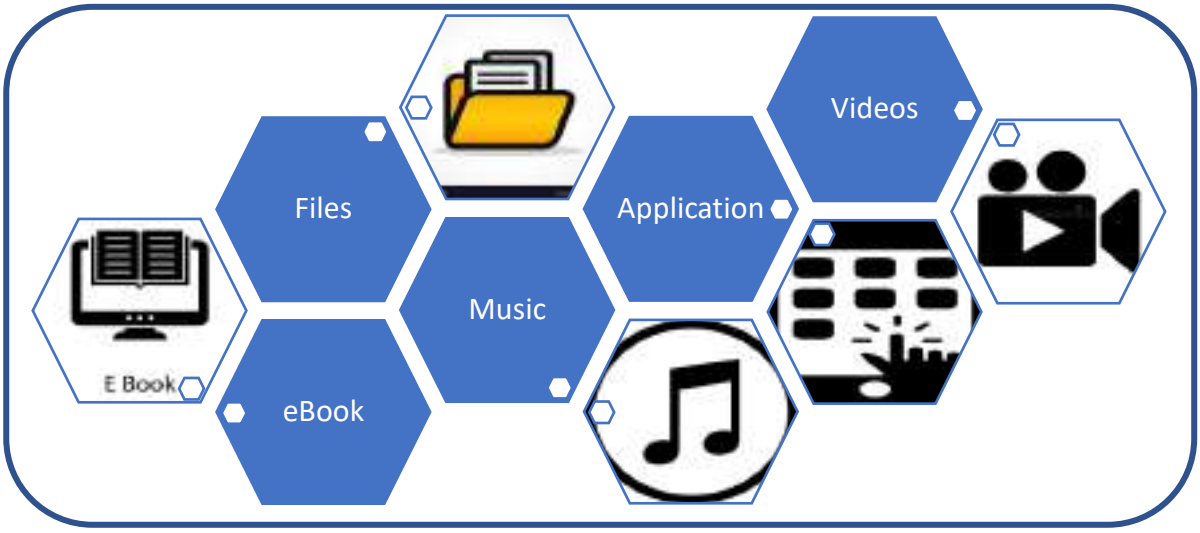
Why Cloud?

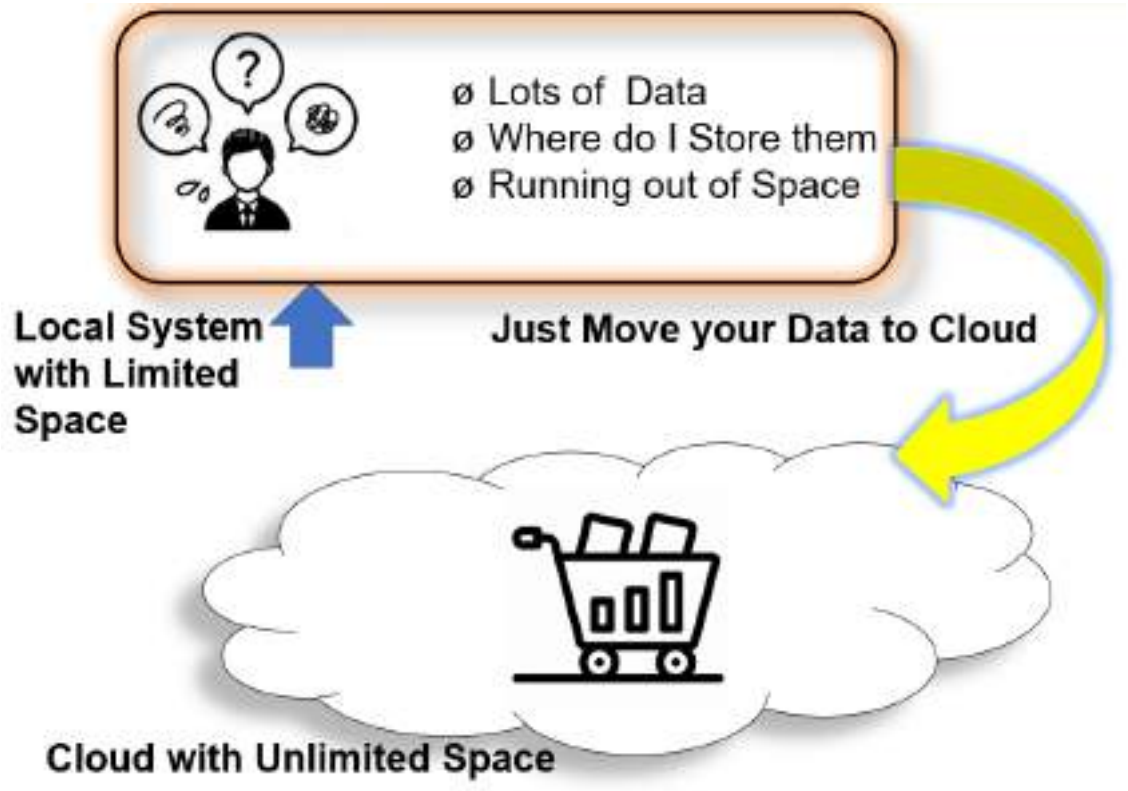
Before Cloud Computing:

If you want to Host a Website, following things are required.

Before Cloud Computing: Disadvantages







1.2 Defining Cloud Computing

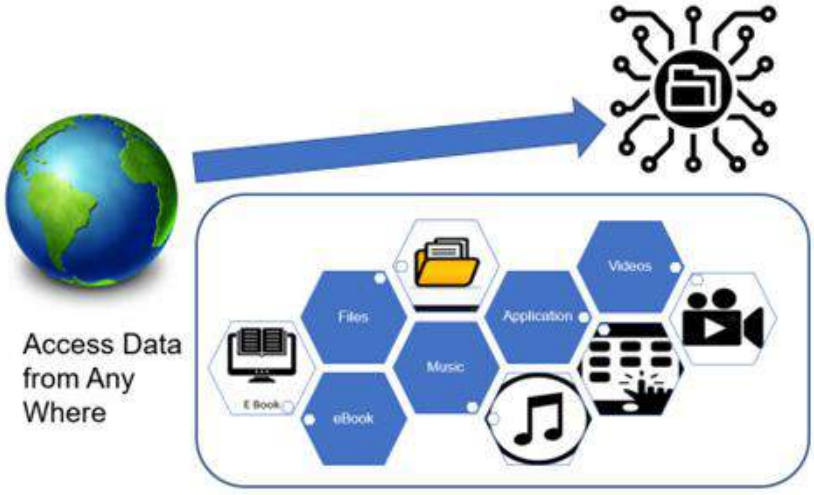
Concept of Cloud Computing

Cloud computing is the delivery of hosting services that are provided to a client over the Internet. - Enable large-scale services without up-front investment.

What is Cloud Computing?

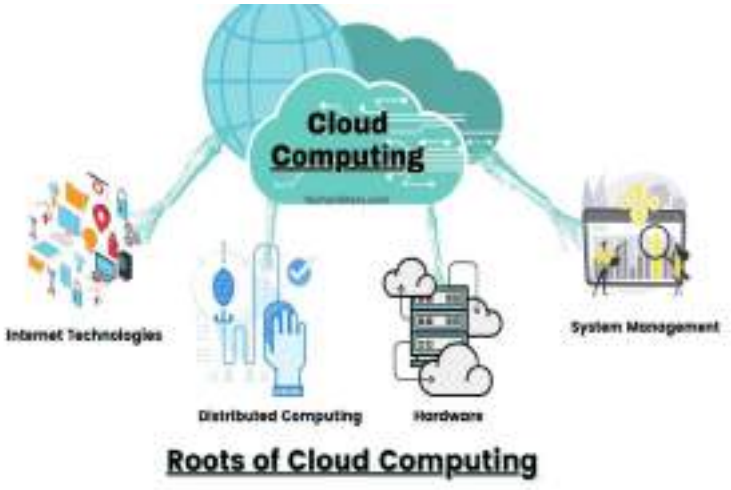
Cloud Computing is:

- Storing Data/ Application on Remote Servers
- Processing Data/Application from Servers
- Accessing Data/Application from Internet



Roots of Cloud Computing

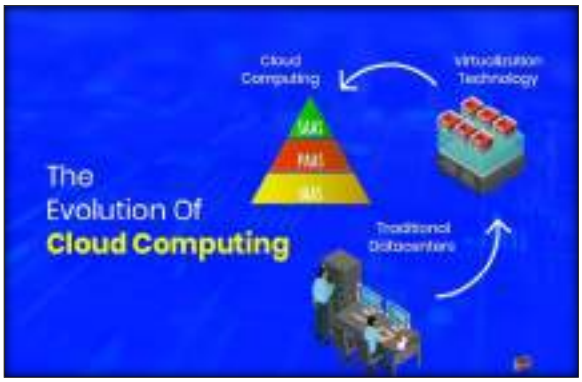
- The roots of clouds computing by observing the advancement of several technologies, especially in:
- Hardware (virtualization, multi-core chips)
- Internet technologies (Web services, service-oriented architectures, Web 2.0)
- Distributed computing (clusters, grids)
- Systems management (autonomic computing, data center automation).

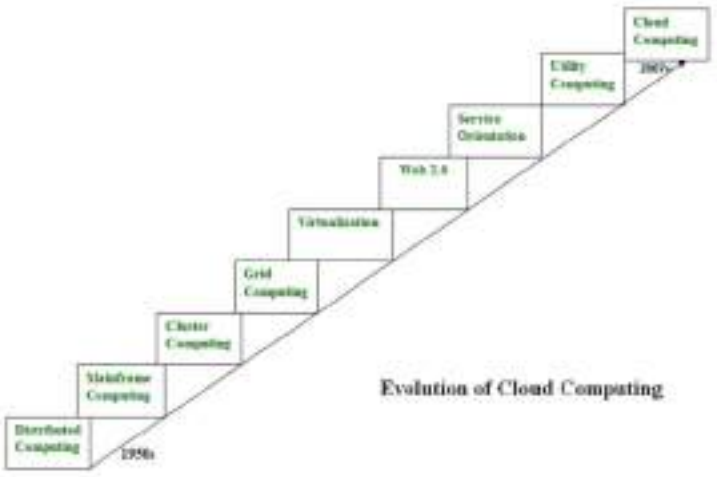


Roots of Cloud Computing

Evolution of Cloud Computing

We are currently experiencing a switch in the IT world, from in-house generated computing power into utility-supplied computing resources delivered over the Internet as Web services. This trend is similar to what occurred about a century ago when factories, which used to generate their own electric power, realized that it is was cheaper just plugging their machines into the newly formed electric power grid.





From Mainframes to Clouds

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.

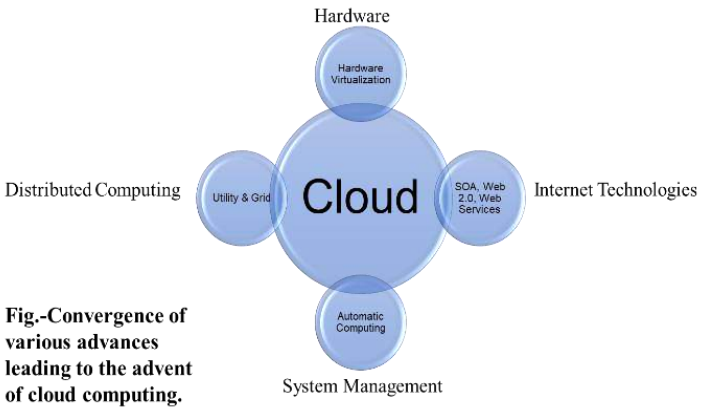
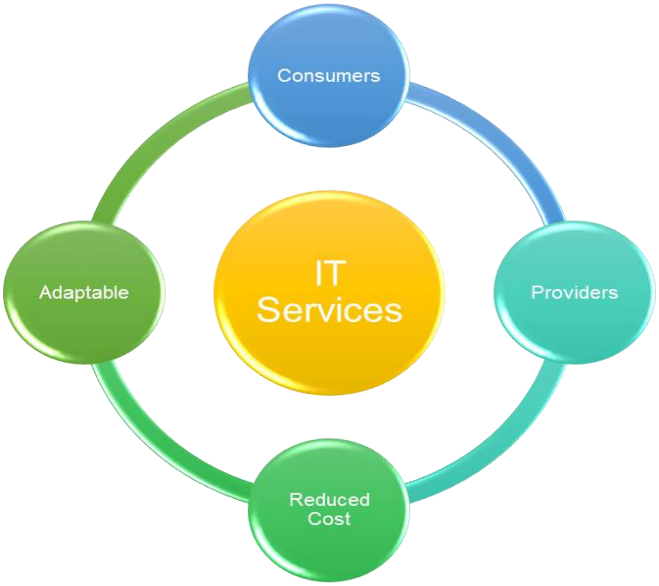
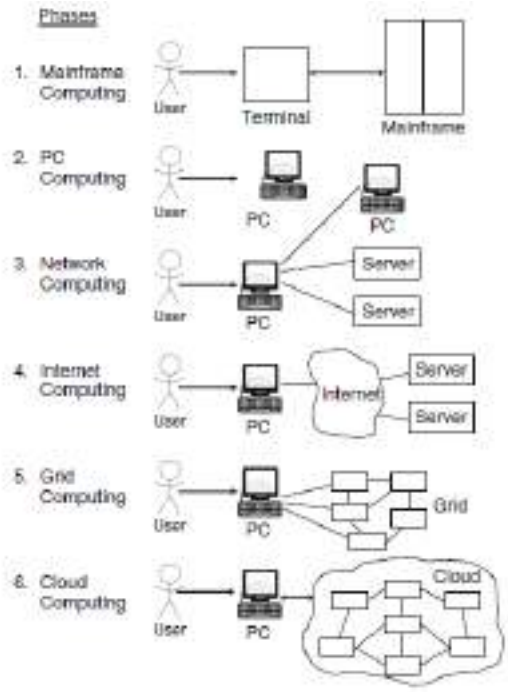


Fig.-Convergence of various advances leading to the advent of cloud computing.

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.



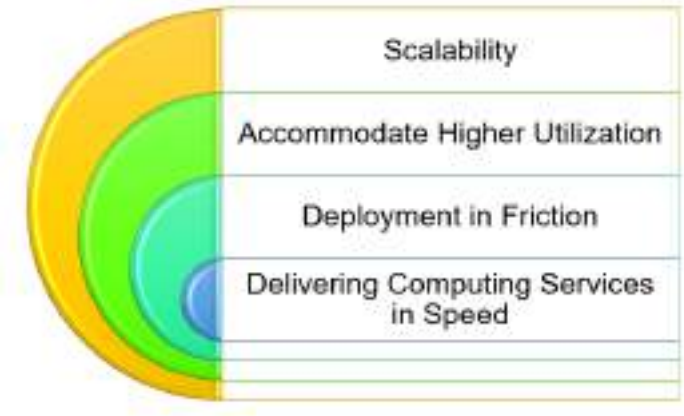


Providers of IT services achieve better operational costs; hardware and software infrastructures are built to provide multiple solutions and serve many users, thus increasing efficiency and ultimately leading to faster return on investment (ROI) as well as lower total cost of ownership (TCO).

The mainframe era collapsed with the advent of fast and inexpensive microprocessors and IT data centres moved to collections of commodity servers. Apart from its clear advantages, this new model inevitably led to isolation of workload into dedicated servers, mainly due to incompatibilities

Between software stacks and operating systems

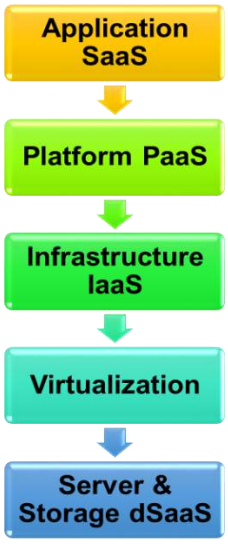
These facts reveal the potential of delivering computing services with the speed and reliability that businesses enjoy with their local machines. The benefits of economies of scale and high utilization allow providers to offer computing services for a fraction of what it costs for a typical company that generates its own computing power.



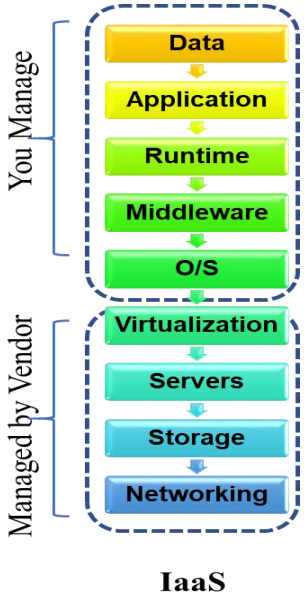
1.3 Cloud Computing: Service Models

Layered Architecture of Cloud Computing

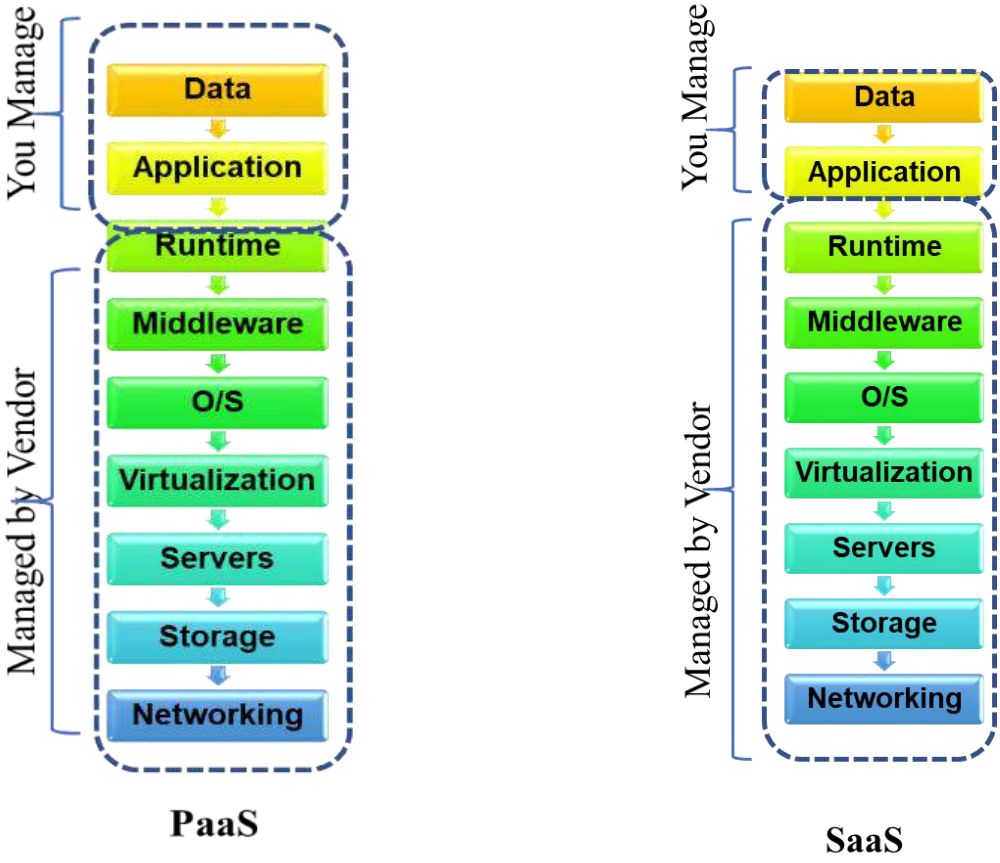
Cloud computing can be viewed as a collection of services, which can be presented as a layered cloud computing architecture, as shown in Fig.



Service Models of Cloud Computing



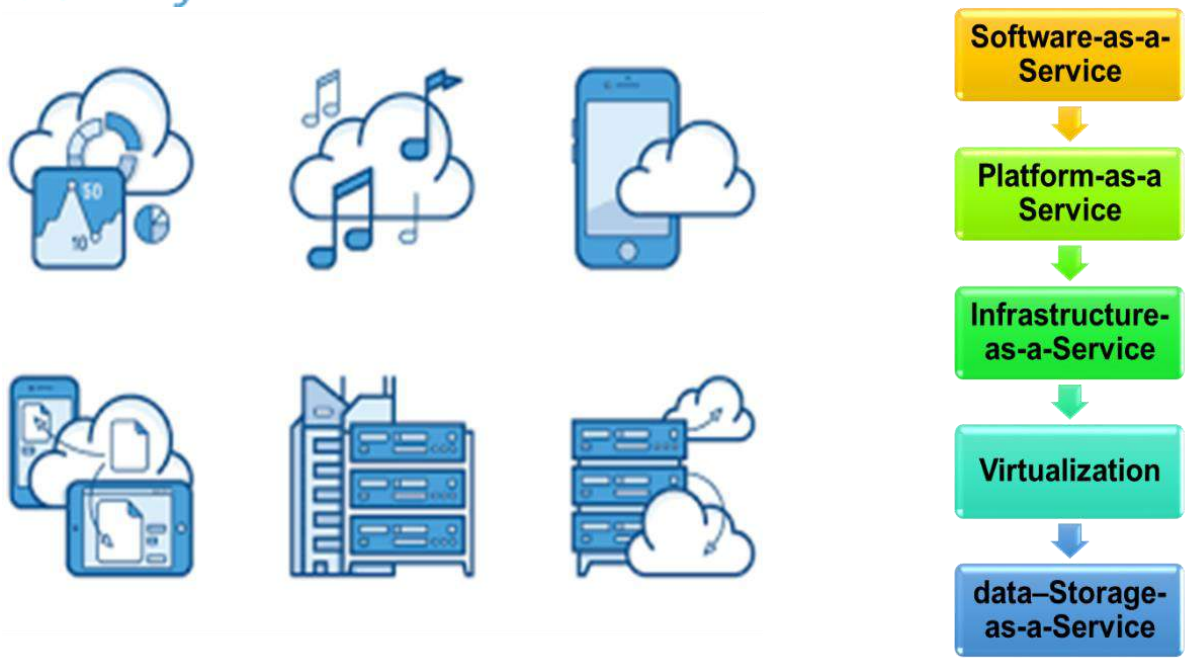
IaaS



The services offered through cloud computing usually include IT services referred as to SaaS (Software-as-a-Service), which is shown on top of the stack. SaaS allows users to run applications remotely from the cloud.

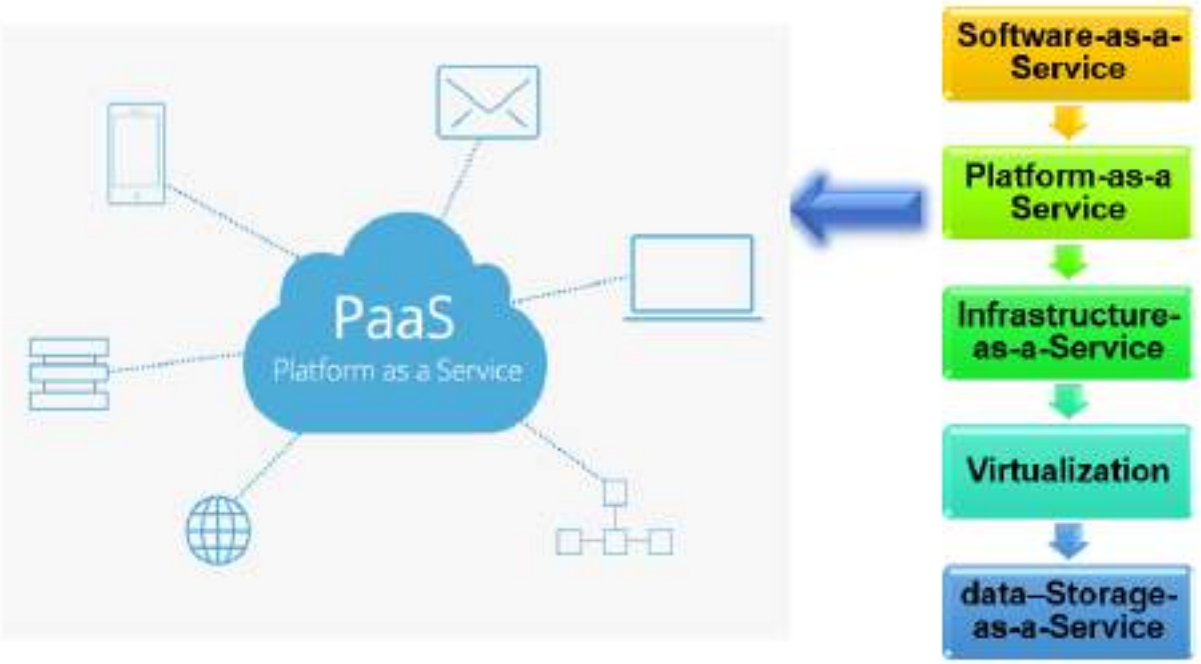
Infrastructure-as-a-service (IaaS) refers to computing resources as a service.

This includes virtualized computers with guaranteed processing power and reserved bandwidth for storage and Internet access.

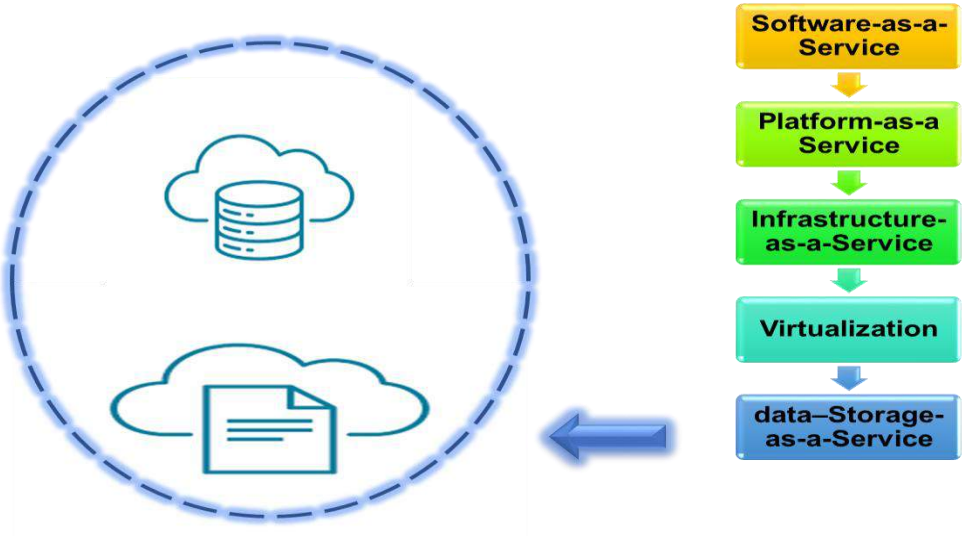


Infrastructure-as-a-Service

Platform-as-a-Service (PaaS) is similar to IaaS, but also includes operating systems and required services for a particular application. In other words, PaaS is IaaS with a custom software stack for the given application.



The **data-Storage-as-a-Service (dSaaS)** provides storage that the consumer is used including bandwidth requirements for the storage.



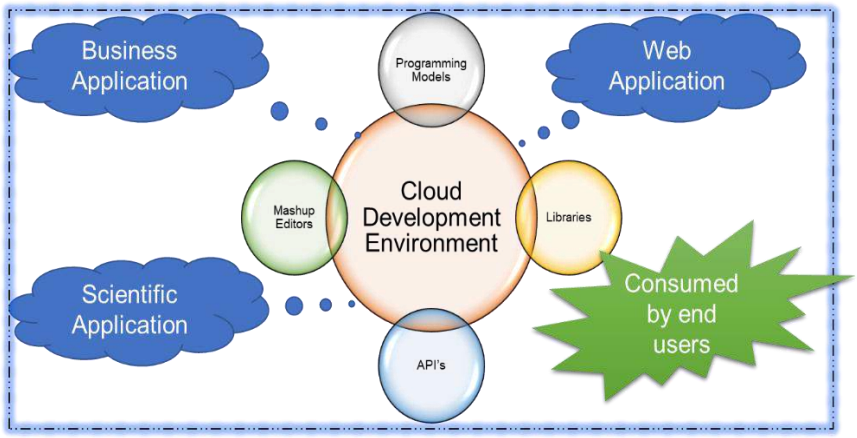
Layers and Types of Clouds

Cloud computing services are divided into three classes, according to the abstraction level of the capability provided and the service model of providers, namely:

1. Infrastructure as a Service
2. Platform as a Service and
3. Software as a Service.

- 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 layer.
- The reference model explains the role of each layer in an integrated architecture.
- A 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-tenant pay-as-you-go services.

Cloud development environments are built on top of infrastructure services to offer application development and deployment capabilities; in this level, various programming models, libraries, APIs, and mashup editors enable the creation of a range of business, Web, and scientific applications. Once deployed in the cloud, these applications can be consumed by end users.



Cloud Computing Versus Cloud Services

Cloud computing is the IT foundation for cloud services and it consists of technologies that enable cloud services. The key attributes of cloud computing are shown in Table.

Key Cloud Computing Attributes

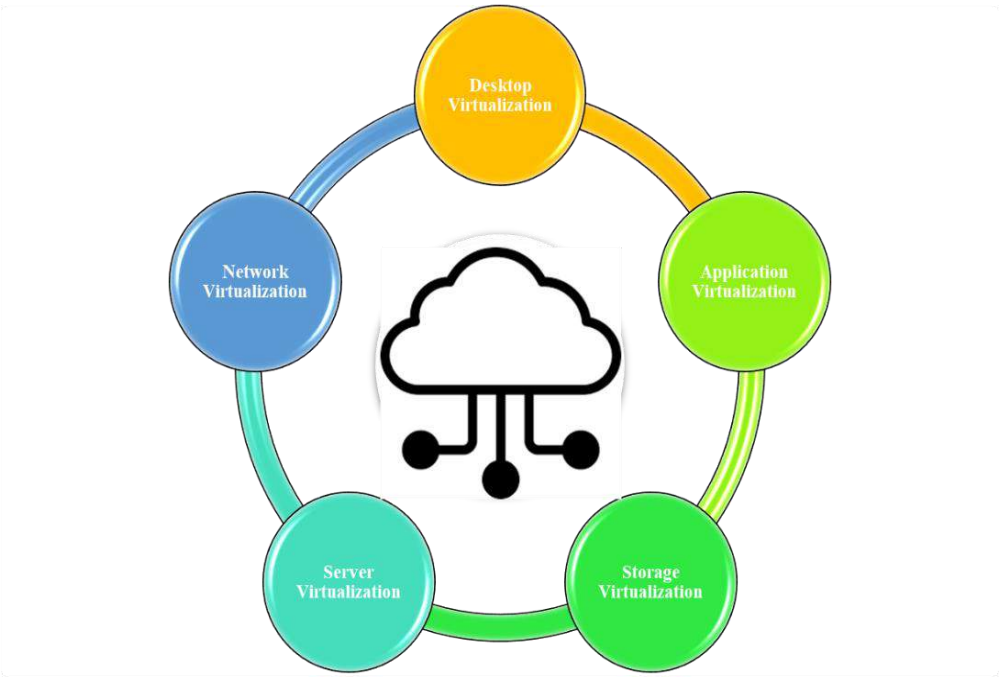
Attributes	Description
Infrastructure System	It includes servers, storage, and networks that can scale as per user demand.
Application software	It provides Web-based user interface, Web services APIs, and a rich variety of configurations.
Application development and deployment software	It supports the development and integration of cloud application software.
System and application management software	It supports rapid self-service provisioning and configuration and usage monitoring.
IP networks	They connect end users to the cloud and the infrastructure components.
Offsite. Third-party provider	In the cloud execution, it is assumed that third-party provides services. There is also a possibility of in-house cloud service delivery.
Accessed via the Internet	Services are accessed via standard-based, universal network access. It can also include security and quality-of-service options.
Minimal or no IT skill required Provisioning	There is a simplified specification of requirements. It includes self-service requesting, near real-time deployment, and dynamic and fine-grained scaling.
Pricing	Pricing is based on usage-based capability and it is fine-grained.
User interface	User interface include browsers for a variety of devices and with rich capabilities.
System interface	System interfaces are based on Web services APIs providing a standard framework for accessing and integrating among cloud services.
Shared resources	Resources are shared among cloud services users; however via configuration options with the service, there is the ability to customize.

Enabling Technologies

Key technologies that enabled cloud computing are described in this section; they include virtualization, Web service and service-oriented architecture, service flows and workflows, and Web 2.0 and mashup.

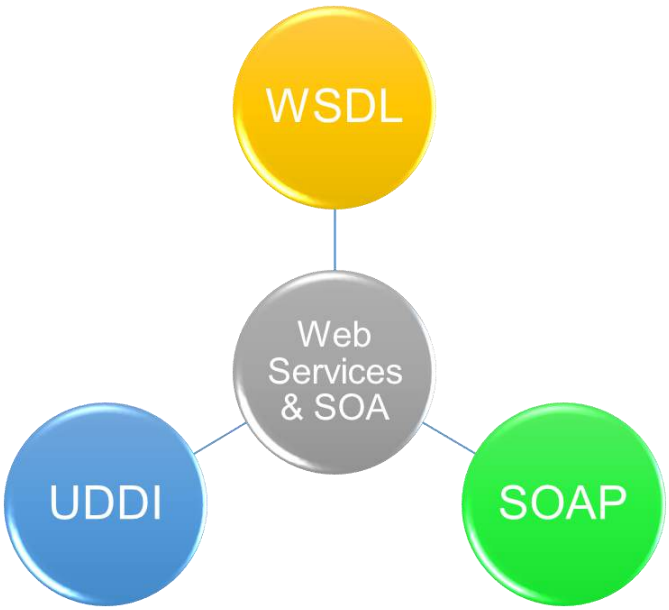
Virtualization

The advantage of cloud computing is the ability to virtualize and share resources among different applications with the objective for better server utilization.



Web Service and Service Oriented Architecture

Web Services and Service Oriented Architecture (SOA) are not new concepts; however, they represent the base technologies for cloud computing. Cloud services are typically designed as Web services, which follow industry standards including WSDL, SOAP, and UDDI.

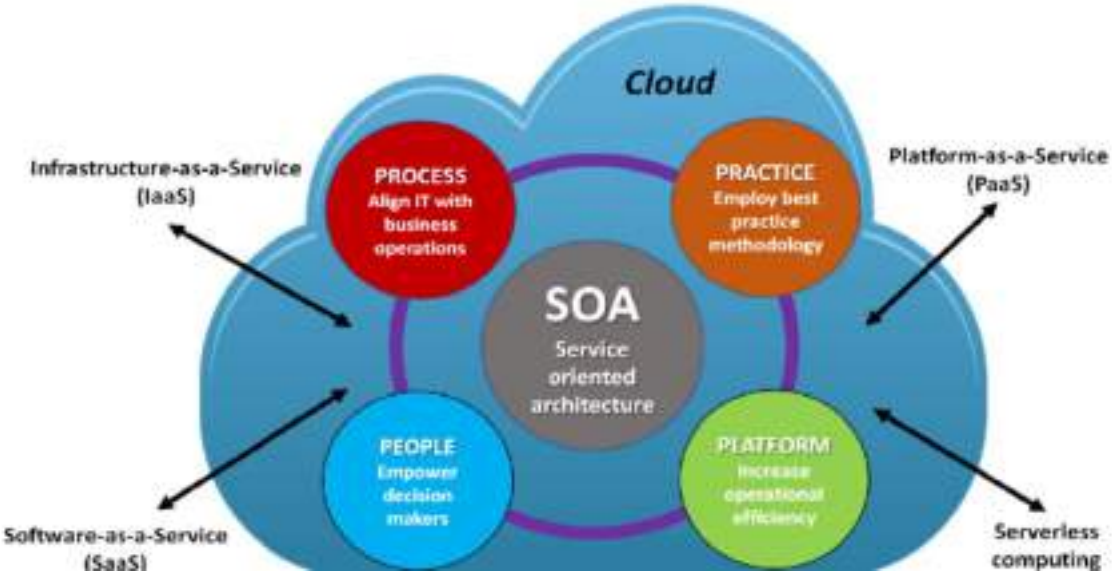


Service Flow and Workflows

- The concept of service flow and workflow refers to an integrated view of service-based activities provided in clouds.
- Workflows have become one of the important areas of research in the field of database and information systems.

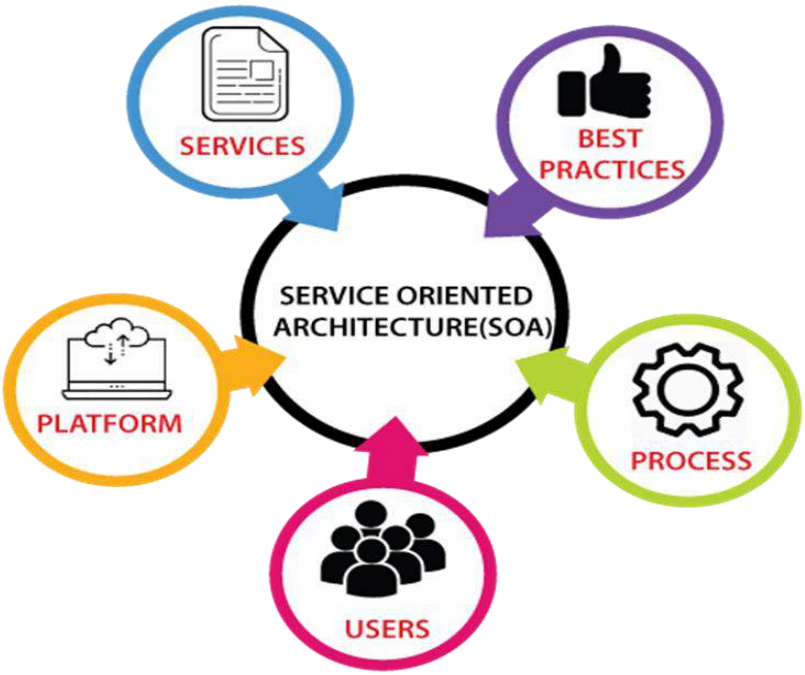
SOA, WEB SERVICES, WEB 2.0, and MASHUPS

- The emergence of Web services (WS) open standards has significantly con-tributed to advances in the domain of software integration.
- Web services can glue 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.
- Over the years a rich WS software stack has been specified and standardized, resulting in a multitude of technologies to describe, compose, and orchestrate services, package and transport messages between services, publish and dis- cover services, represent quality of service (QoS) parameters, and ensure security in service access.
- 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 are packaged as services, which are well-defined, self- contained modules that 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. While some WS are published with the intent of serving end-user applications, their true power resides in its interface being accessible by other services. An enterprise application that follows the SOA paradigm is a collection of services that together perform complex business logic.



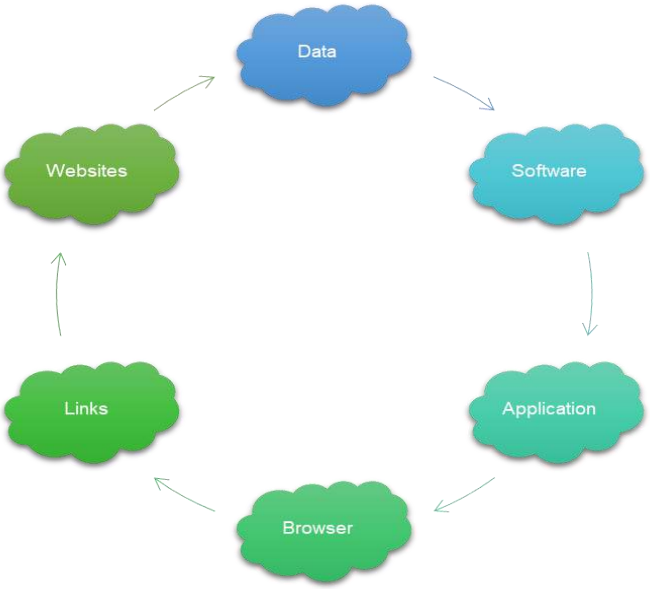
- In the consumer Web, information and services may be programmatically aggregated, acting as building blocks of complex compositions, called service mashups.

- 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.



- 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 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. Many building blocks and solutions are now available in public marketplaces.



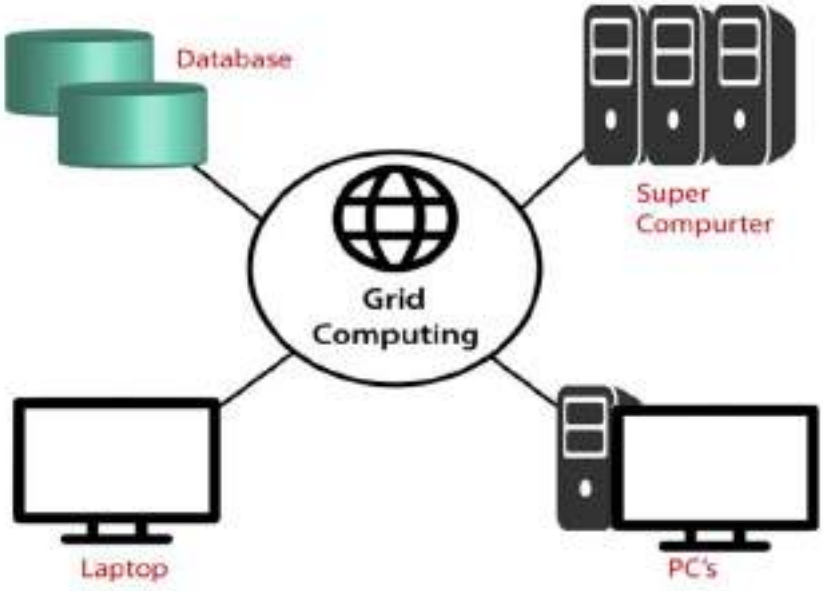
For example, Programmable Web is a public repository of service APIs and mashups currently listing thousands of APIs and mashups. Popular APIs such as Google Maps, Flickr, YouTube, Amazon ecommerce, and Twitter, when combined, produce a variety of interesting solutions, from finding video game retailers to weather maps. Similarly, Salesforce.com's offers AppExchange, which enables the sharing of solutions developed by third-party developers on top of Salesforce.com components.

Grid Computing

Grid computing enables aggregation of distributed resources and transparently access to them. Most production grids such as TeraGrid 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 as climate modelling, drug design, and protein analysis.

- A key aspect of the grid vision realization has been building standard Web services-based protocols that allow distributed resources to be discovered, accessed, allocated, monitored, accounted for, and billed for, etc., and in general managed as a single virtual system.

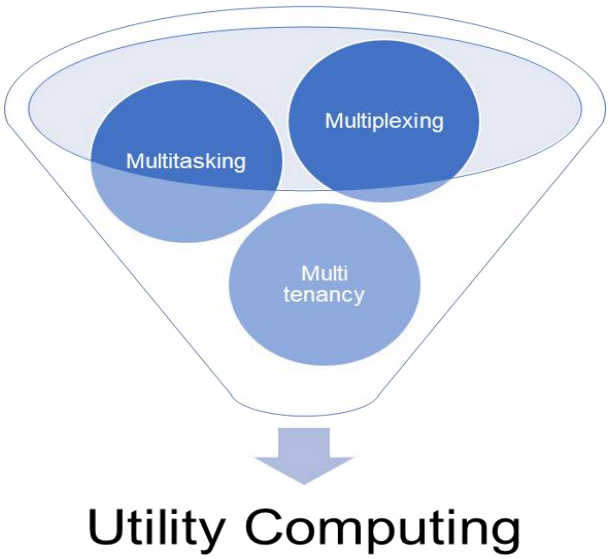
- The Open Grid Services Architecture (OGSA) addresses this need for standardization by defining a set of core capabilities and behaviors that address key concerns in grid systems.



Utility Computing

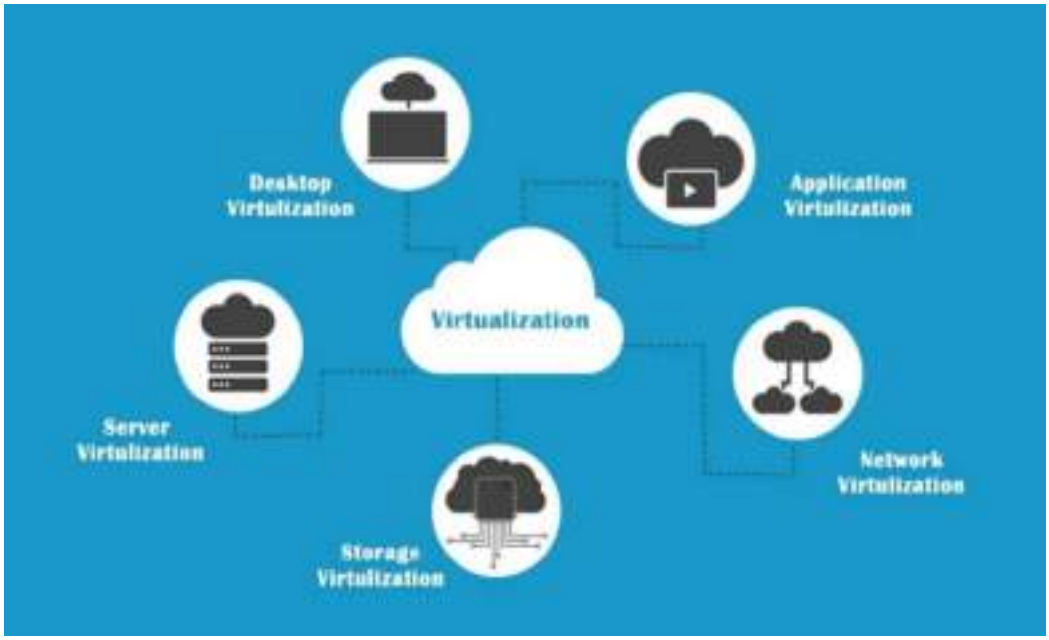
- In utility computing environments, users assign a —utility‖ value to their jobs, where utility is a fixed or time-varying valuation that captures various QoS constraints (deadline, importance, satisfaction).

- The valuation is the amount they are willing to pay a service provider to satisfy their demands. The service providers then attempt to maximize their own utility, where said utility may directly correlate with their profit.



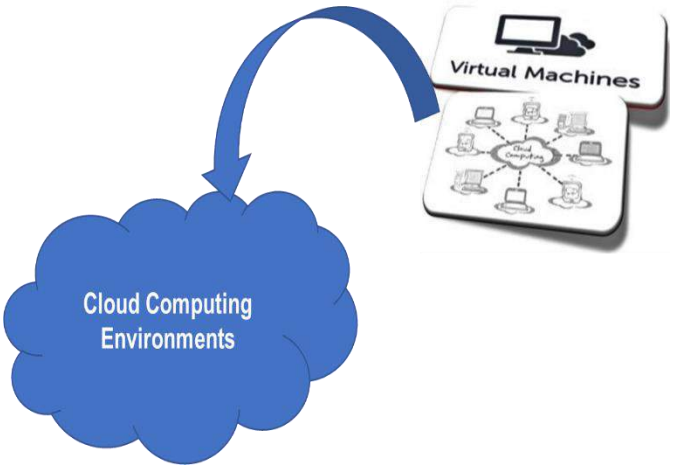
Hardware Virtualization

- Cloud computing services are usually backed by large-scale data centers composed of thousands of computers. Such data centers are built to serve many users and host many disparate applications.
- For this purpose, hardware virtualization can be considered as a perfect fit to overcome most operational issues of data center building and maintenance.
- The idea of virtualizing a computer system resources, including processors, memory, and I/O devices, has been well established for decades, aiming at improving sharing and utilization of computer systems.
- Hardware virtualization allows running multiple operating systems and software stacks on a single physical platform. As depicted in Figure, a software layer, the virtual machine monitor (VMM), also called a hypervisor, mediates access to the physical hardware presenting to each guest operating system a virtual machine (VM), which is a set of virtual platform interfaces.
- The advent of several innovative technologies—multi-core chips, paravirtualization, hardware-assisted virtualization, and live migration of VMs—has contributed to an increasing adoption of virtualization on server systems. Traditionally, perceived benefits were improvements on sharing and utilization, better manageability, and higher reliability.
- Management of workload in a virtualized system, namely isolation, consolidation, and migration. Workload isolation is achieved since all program instructions are fully confined inside a VM, which leads to improvements in security. Better reliability is also achieved because software failures inside one VM do not affect others.



- Workload migration, also referred to as application mobility, targets at facilitating hardware maintenance, load balancing, and disaster recovery. It is done by encapsulating a guest OS state within a VM and allowing it to be suspended, fully serialized, migrated to a different platform, and resumed immediately or preserved to be restored at a later date. A VM's state includes a full disk or partition image, configuration files, and an image of its RAM.

- A number of VMM platforms exist that are the basis of many utility or cloud computing environments. The most notable ones, VMWare, Xen, and KVM.

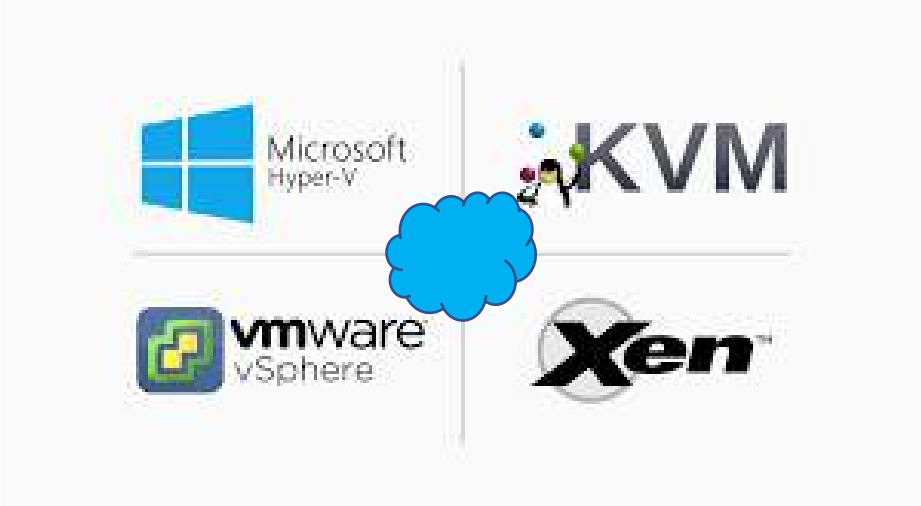


Virtual Appliances and the Open Virtualization Format

- An application combined with the environment needed to run it (operating system, libraries, compilers, databases, application containers, and so forth) is referred to as a —virtual appliance.
- Packaging application environments in the shape of virtual appliances eases software customization, configuration, and patching and improves portability.

- Most commonly, an appliance is shaped as a VM disk image associated with hardware requirements, and it can be readily deployed in a hypervisor.
- On-line marketplaces have been set up to allow the exchange of ready-made appliances containing popular operating systems and useful software combinations, both commercial and open-source.
- Most notably, the VMWare virtual appliance marketplace allows users to deploy appliances on VMWare hypervisors or on partners public clouds, and Amazon allows developers to share specialized Amazon Machine Images (AMI) and monetize their usage on Amazon EC2.
- In a multitude of hypervisors, where each one supports a different VM image format and the formats are incompatible with one another, a great deal of interoperability issues arises. For instance, Amazon has its Amazon machine image (AMI) format, made popular on the Amazon EC2 public cloud.

Other formats are used by Citrix XenServer, several Linux distributions that ship with KVM, Microsoft Hyper-V, and VMware ESX.

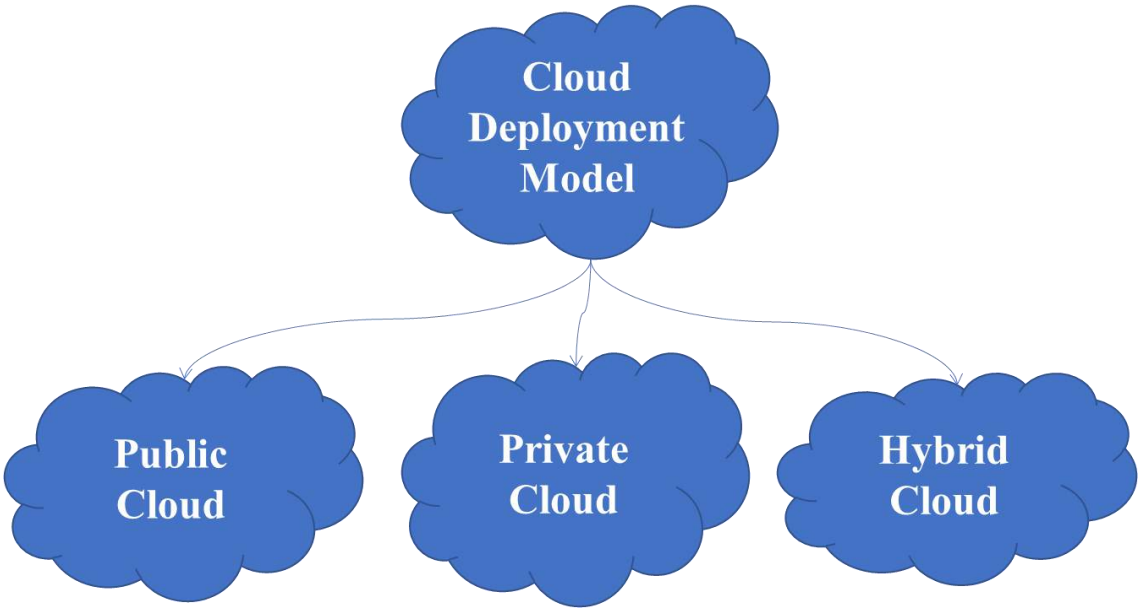


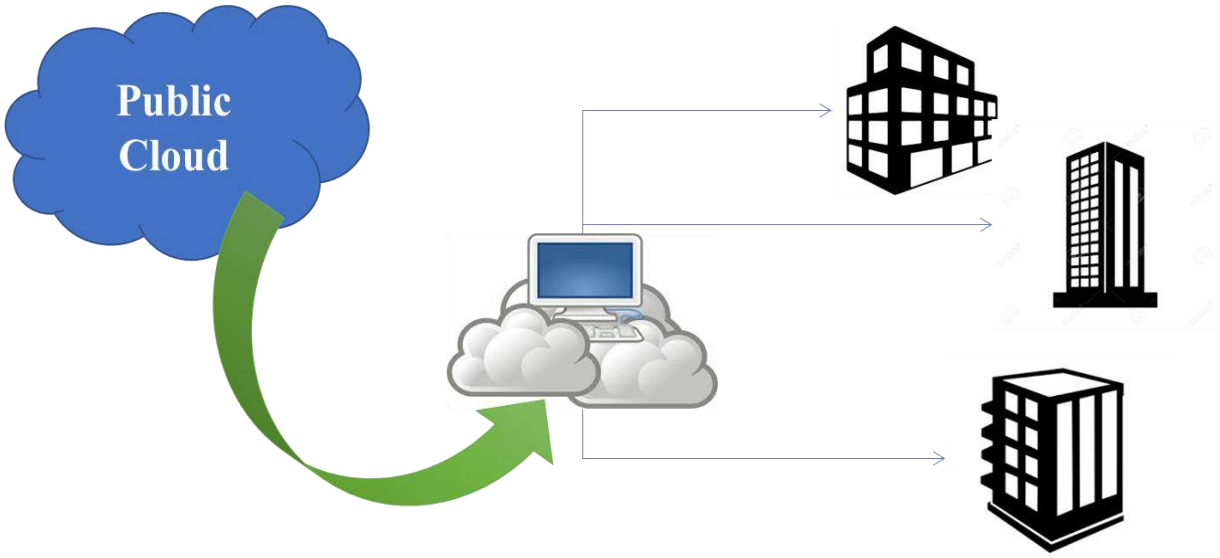
Autonomic Computing

- The increasing complexity of computing systems has motivated research on autonomic computing, which seeks to improve systems by decreasing human involvement in their operation. In other words, systems should manage themselves, with high-level guidance from humans.
- Autonomic, or self-managing, systems rely on monitoring probes and gauges (sensors), on an adaptation engine (autonomic manager) for computing optimizations based on monitoring data, and on effectors to carry out changes on the system.
- IBM's Autonomic Computing Initiative has contributed to define the four properties of autonomic systems: self-configuration, self- optimization, self-healing, and self-protection.

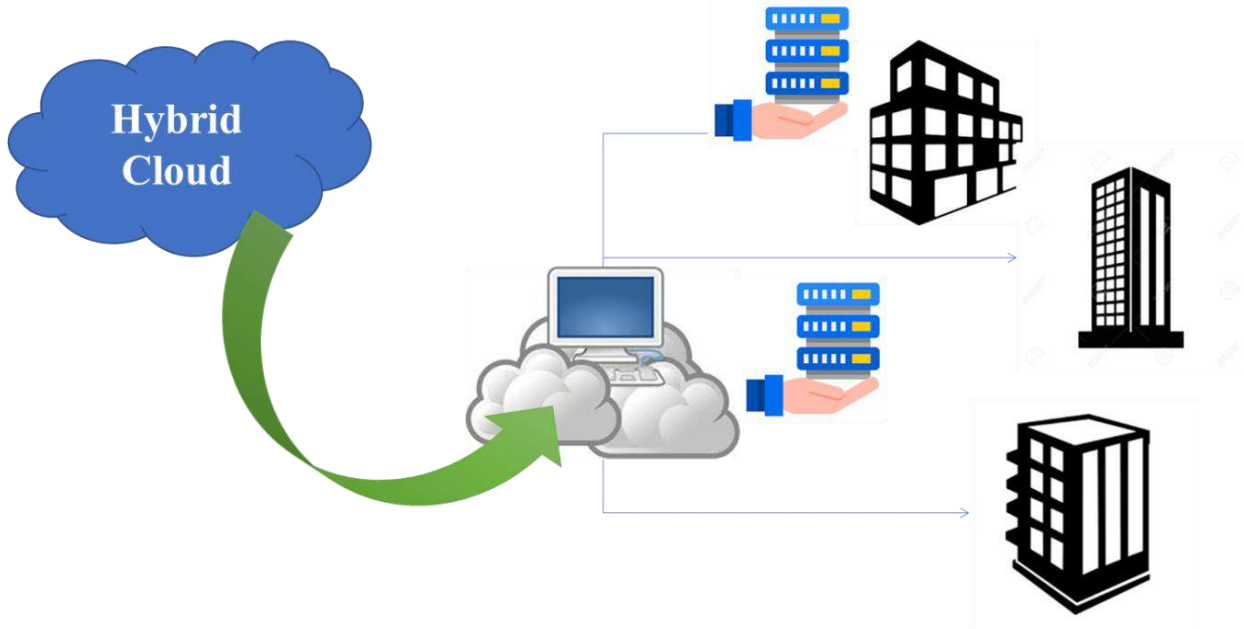
1.4 Delivering Services from the Cloud

Deployment Models



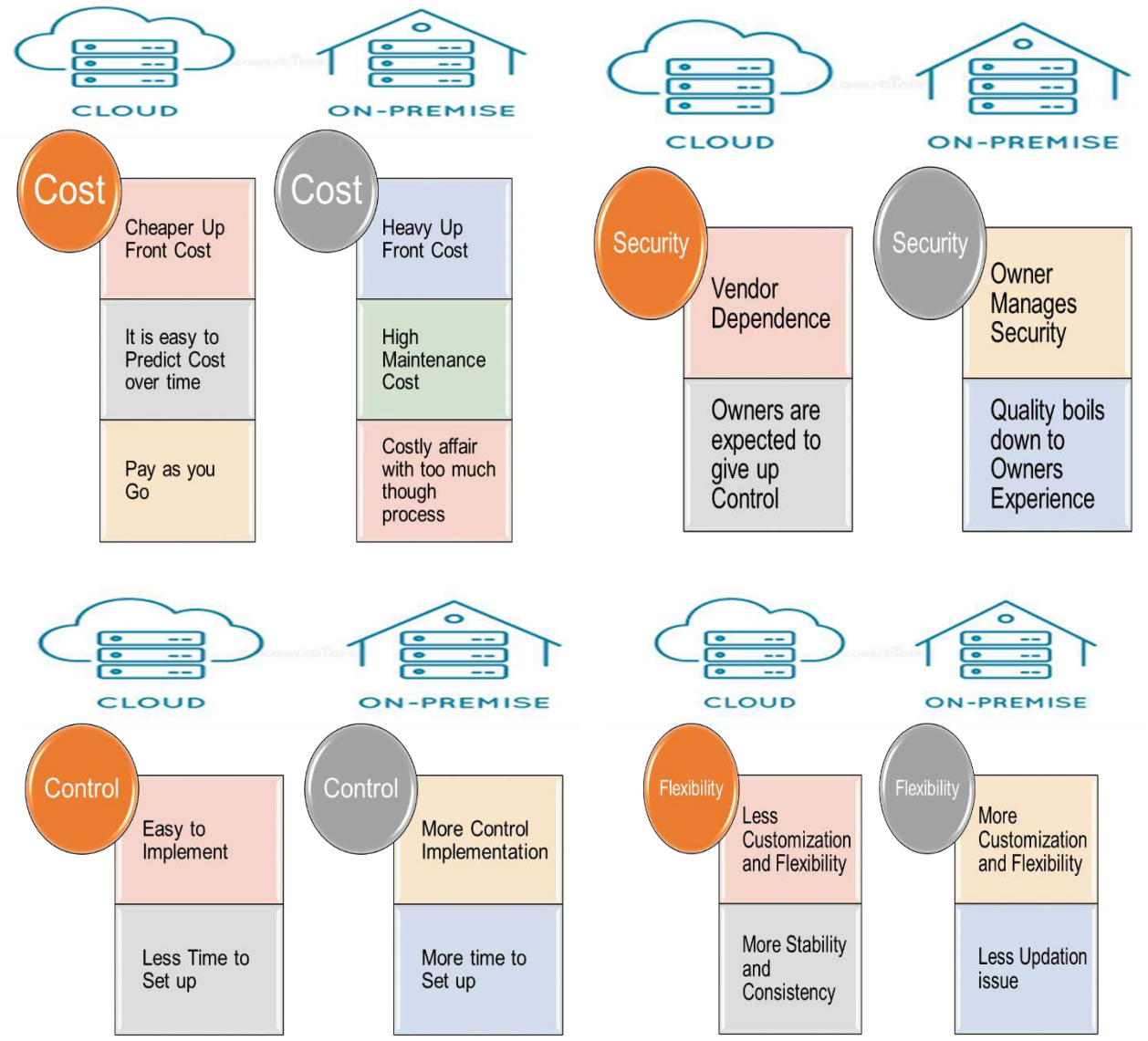


- A Service Provider Makes Resources, Such as applications and storage available to the general public over the internet.
- Easy and inexpensive set up because hardware, application and bandwidth costs are covered by the provider.
- No wasted resources because you pay for what you use.
- Offers Hosted Service to a limited number of people behind the firewall, so it minimizes the security concerns.
- Private Cloud gives companies direct control over their data.



- A Cloud Computing environment which uses a mix of on-premises, private cloud and third-party, public cloud services.
- It helps you leverage the best of both worlds

Cloud Computing Vs On-Premise Computing



Cloud Computing Features

Cloud computing brings a number of new features compared to other computing.

- **Scalability and on-demand services**
  Cloud computing provides resources and services for users on demand. The resources are scalable over several data centers.

- **User-centric interface**
  Cloud interfaces are location independent and can be accesses by well-established interfaces such as Web services and Internet browsers.

- **Guaranteed Quality of Service (QoS)**
  Cloud computed can guarantee QoS for users in terms of hardware/CPU performance, bandwidth, and memory capacity.

- **Autonomous System**
  The cloud computing systems are autonomous systems managed transparently to users. However, software and data inside clouds can be automatically reconfigured and consolidated to a simple platform depending on user's needs.

- **Pricing**
  Cloud computing does not require up-from investment. No capital expenditure is required. Users pay for services and capacity as they need them.

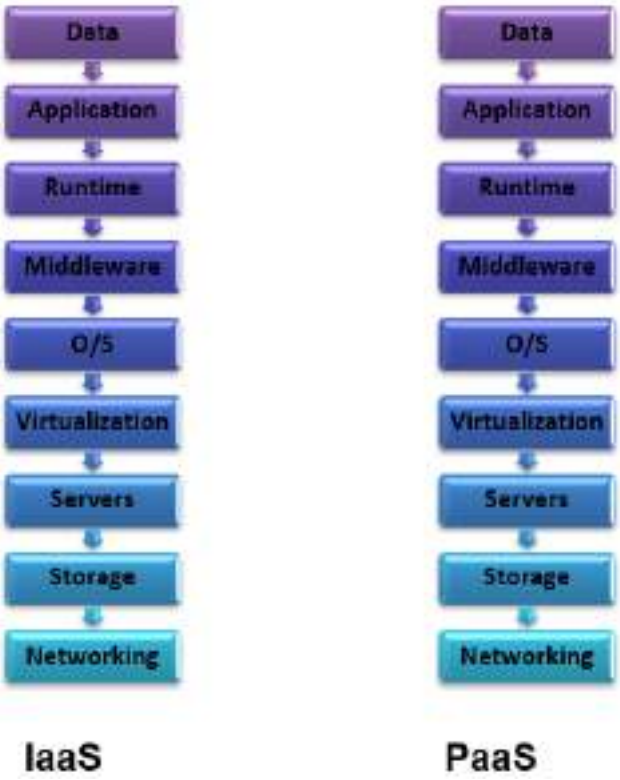
Cloud Providers

Cloud Computing Platforms

Company	Cloud Computing Platform	Year of Lunch	Key Offerings
Amazon. com	AWS (Amazon Web Services)	2006	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)
Microsoft	Azure	2009	Application platform as a service (.Net, SQL data services)
Google	Google App. Engine	2008	Web Application Platform as a service (Python run time environment)
IBM	Blue Cloud	2008	Virtualized Blue cloud data center

Section 3: Exercises

Exercise 1: Mark the Things Managed by You and Vendor in below Layered Architecture of Cloud Computing.



Exercise 2: Participate in a group discussion on following topics:

a) Layers and types of Cloud Computing
b) Difference between Cloud Computing and Cloud Services
c) New Technologies that enabled Cloud Computing
d) Cloud Computing Features and Standards
e) Cloud Computing Platforms
f) Cloud Computing Challenges
g) Components of Cloud Computing
h) Different types of Clouds and Services

Section 4: Assessment Questionnaire

1. What Is Cloud Computing?
2. What is Cloud?
3. Explain Benefits of Cloud Computing?
4. Write Notes on Origin of Cloud Computing?
5. Explain SPI?
6. What are the different data types used in cloud computing?
7. What are the different layers in cloud computing? Explain working of them?
8. What is the difference between cloud computing and mobile computing?
9. Which are different layers are used by cloud architecture?

Multiple choice Questions:

1. Who is the father of cloud computing?

a. Sharon B. Codd

b. Edgar Frank Codd
c. J.C.R. Licklider
d. Charles Bachman

2. Which of the following are the features of cloud computing?
a. Security
b. Availability
c. Large Network Access
d. All of the mentioned

3. Applications and services that run on a distributed network using virtualized resources is known as:
a. Parallel computing
b. Soft computing
c. Distributed computing
d. Cloud computing

4. Which architectural layer is used as a backend in cloud computing?
a. Cloud
b. Soft
c. Client
d. All of the mentioned

5. Which of the following is the correct statement?
a. Cloud computing presents new opportunities to users and developers
b. Service Level Agreements (SLAs) is small aspect of cloud computing
c. Cloud computing does not have impact on software licensing
d. All of the mentioned

6. Which of the following is the Cloud Platform provided by Amazon?
a. AWS
b. Cloudera
c. Azure
d. All of the mentioned

----------End of Module----------

MODULE 2

Adopting the Cloud

Section 1: Learning Outcomes

After completing this module, you will be able to:

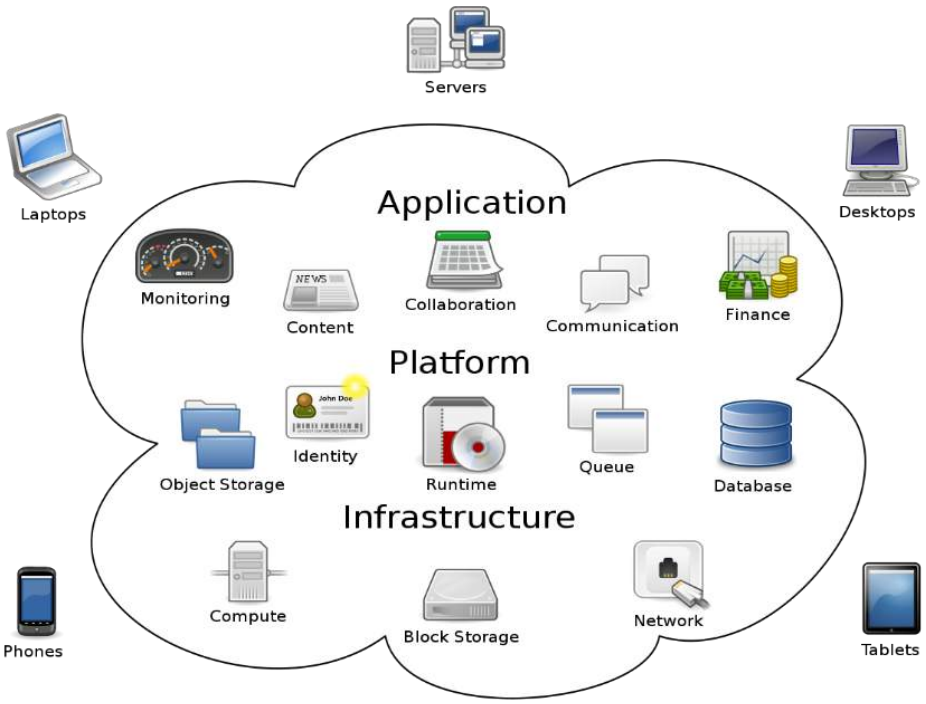
- Explain the facts of adopting the cloud
- Describe Key Drivers of Cloud Computing Solutions
- Use Self –Service Feature of Cloud Computing
- Explain the term 'Pre-Usage Metered' and Billed Feature
- Define the Elastic feature

Section 2: Relevant Knowledge

2.1 Adopting the Cloud

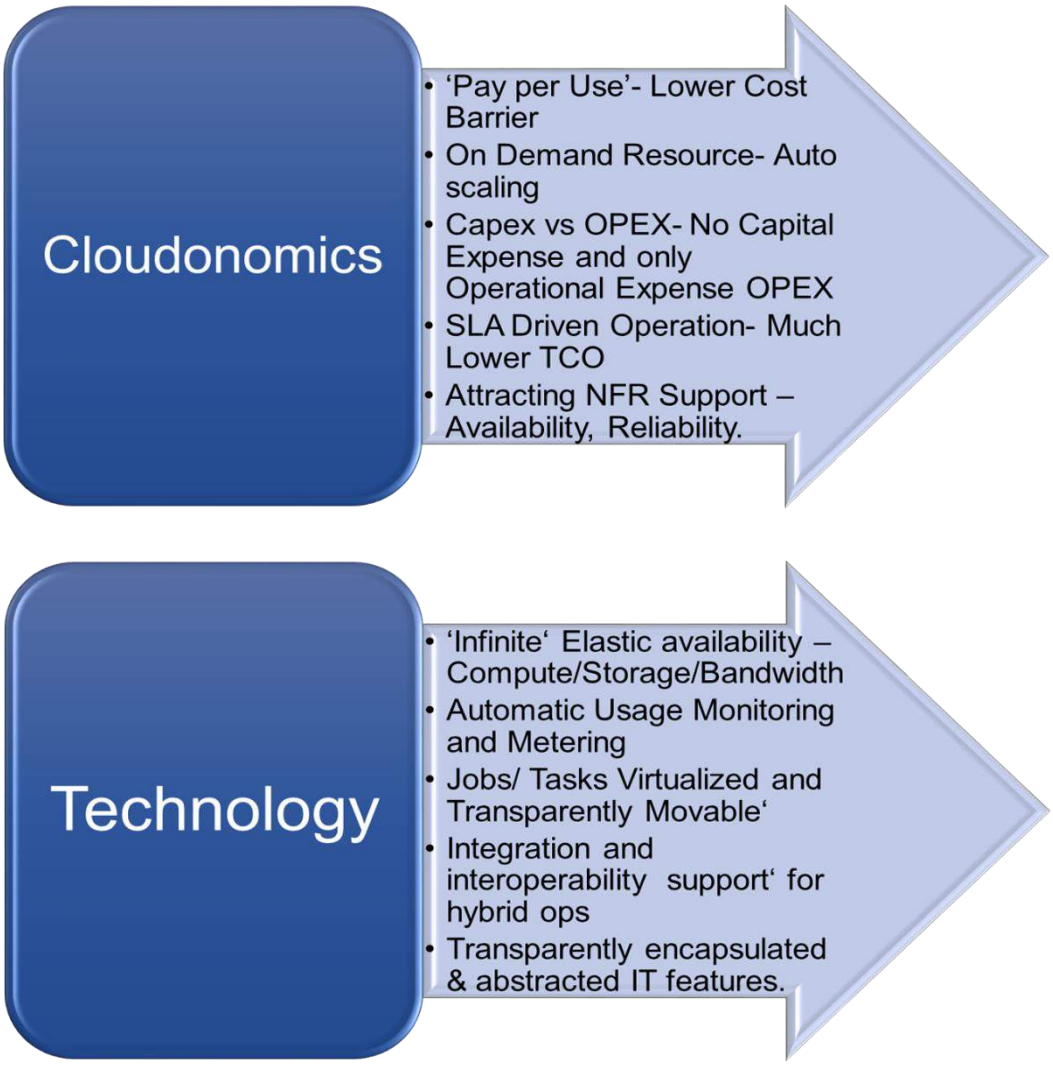
Interesting Facts about Cloud Computing

- Over half of US Government Agencies Depends upon Cloud.
- Most Cloud Computing activities involve the Banking Sector.
- Cloud Market is expected to Reach over $650 billion in under 3 Years.
- It is One of the Fastest Growing IT Sector.



The Promise of Cloud Computing

The promise of cloud computing has raised the IT expectations of small and medium enterprises beyond measure.

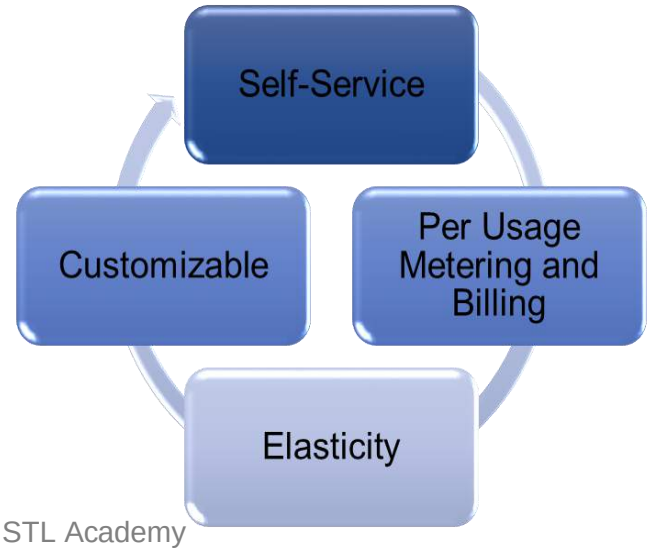


2.2 Key Drivers of Cloud Computing Solutions

Key Features which Drive Cloud Adoption

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
(iv) customizable



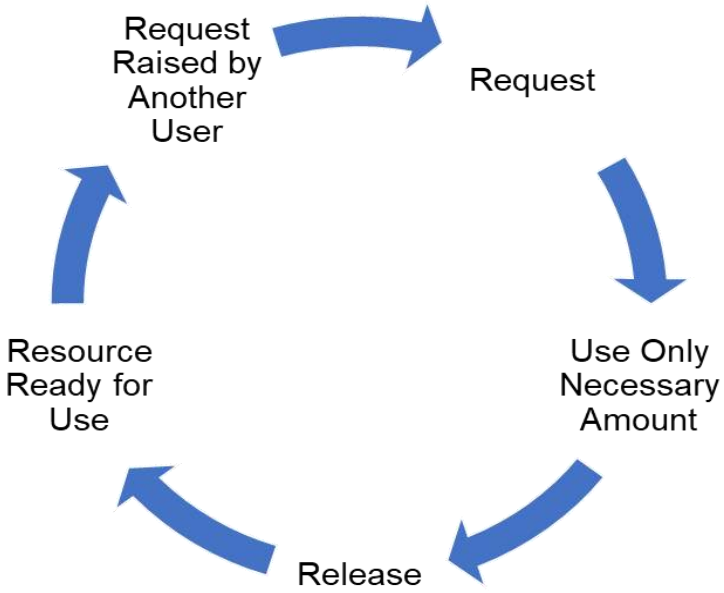
Self-Service

- Consumers of cloud computing services expect on-demand, nearly instant access to resources.
- To support this expectation, clouds must allow self-service access so that customers can request, customize, pay, and use services without intervention of human operators.

On-Demand Instant Access to Resources

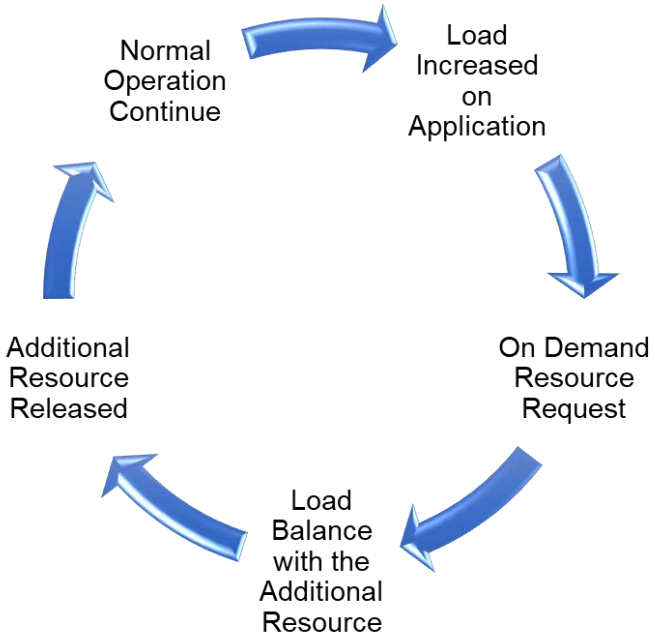
Per-Usage Metering and Billing

- 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 (and not pay for) 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
  (b) released when load decreases (scale up and down)



Customization

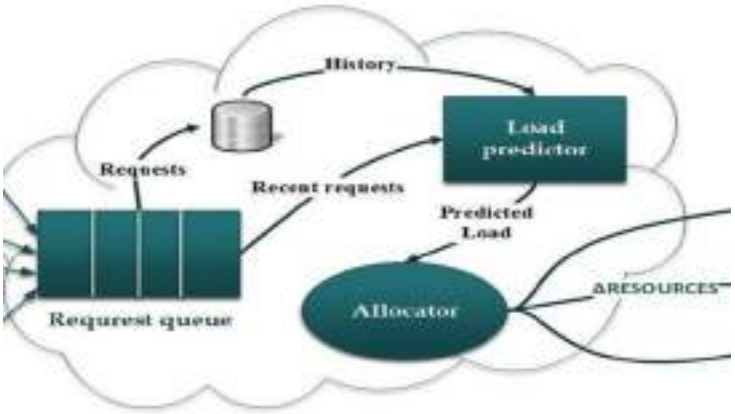
- In a multi-tenant cloud a great disparity between user needs is often the case. Thus, resources rented from the cloud must be highly customizable.
- In the case of infrastructure services, customization means allowing users to deploy specialized virtual appliances and to be given privileged (root) access to the virtual servers.
- Other service classes (PaaS and SaaS) offer less flexibility and are not suitable for general-purpose computing, but still are expected to provide a certain level of customization.

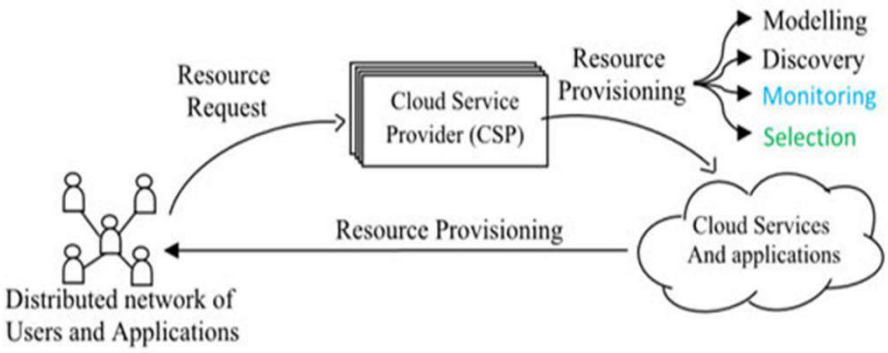
2.3 Instantaneous Provisioning of Computing Resources

Provisioning of Cloud Computing Services

Cloud computing is currently emerging as an ever-changing, growing paradigm that models "everything-as-a-service." Virtualised physical resources, infrastructure, and applications are supplied by service provisioning in the cloud.

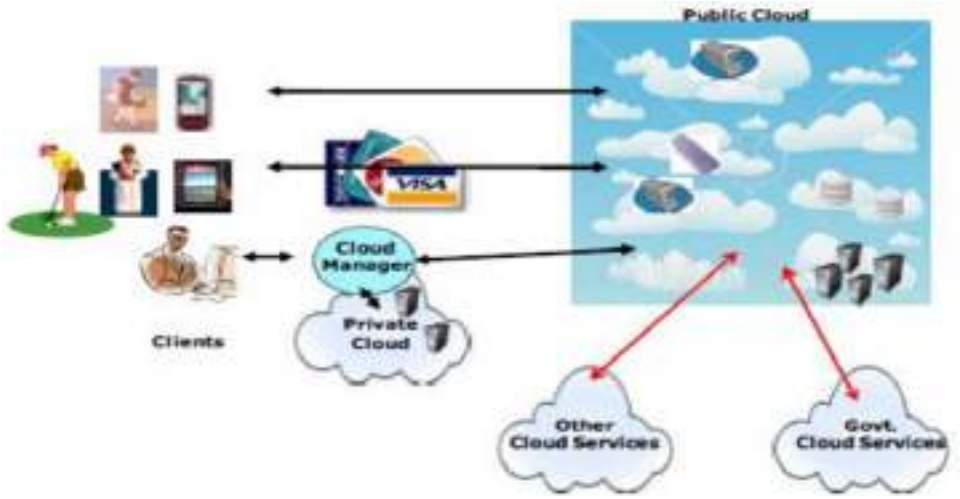
The evolution in the adoption of cloud computing is driven by clear and distinct promising features for both cloud users and cloud providers. However, the increasing number of cloud providers and the variety of service offerings have made it difficult for the customers to choose the best services.



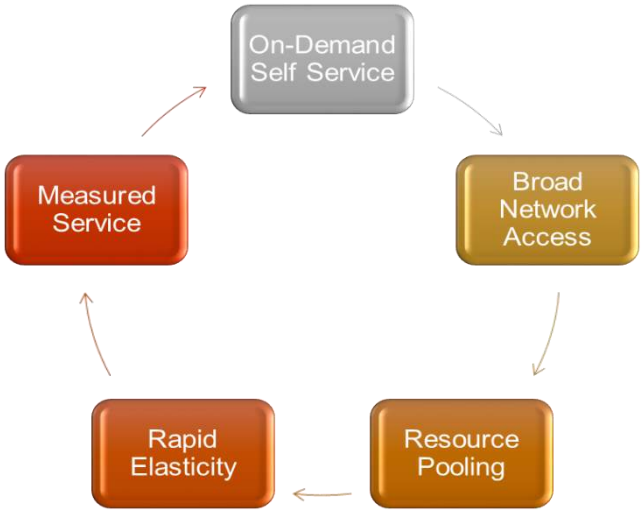


Resource Provisioning System in Cloud Computing (RPS)

- By employing successful service provisioning, the essential services required by customers, such as agility and availability, pricing, security and trust, and user metrics can be guaranteed by service provisioning.
- Hence, continuous service provisioning that satisfies the user requirements is a mandatory feature for the cloud user and vitally important in cloud computing service offerings.
- Therefore, we aim to review the state-of-the-art service provisioning objectives, essential services, topologies, user requirements, necessary metrics, and pricing mechanisms.



Cloud computing is the distributed computing model that provides computing facilities and resources to users in an on-demand, pay-as-you-go model.

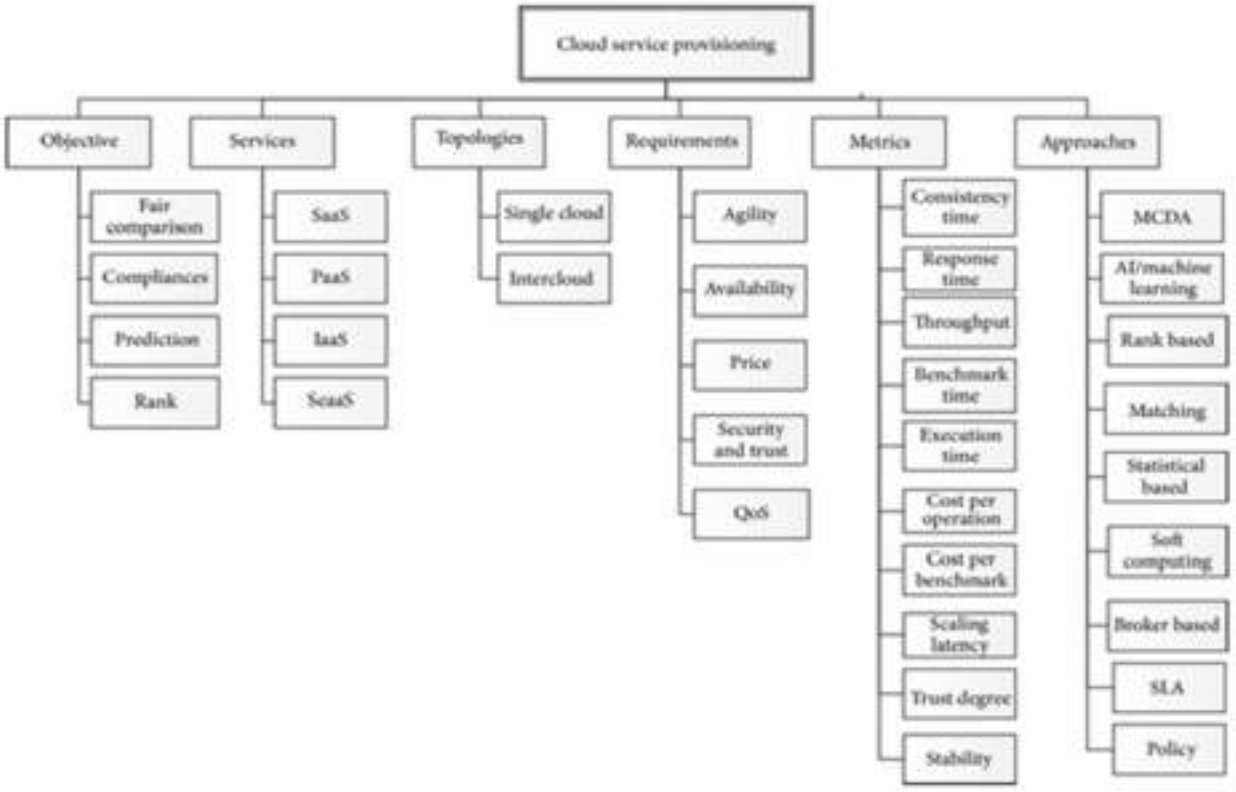


The aim of the cloud computing model is to increase the opportunities for cloud users by accessing leased infrastructure and software applications anywhere and anytime. Therefore, cloud computing offers a new type of information and services that broadens the brand-new vision of information technology (IT) services.

Service Provisioning Definition

- The aim of the cloud computing model is to increase the opportunities for cloud users by accessing leased infrastructure and software applications anywhere and anytime. Therefore, cloud computing offers a new type of information and services that broadens the brand-new vision of information technology (IT) services.
- Cloud service provisioning is a manner of providing customers access to resources to complete the desired tasks required by the customer. The hardware, software, or computational tasks can be the form of provisioned resources.
- The state-of-the-art thematic taxonomy of service provisioning is presented by classifying several vital key issues for further discussion. Figure shows the taxonomy of service provisioning selection, comprising approaches, objectives, requirements, metrics, techniques, services, and topologies.

Service Provisioning Taxonomy



Service Provisioning Topology

- In topological perspective, service provisioning is divided into two parts: single cloud and intercloud.
- A single cloud computing data center is used by the client who brings several challenges. The unavailability of cloud service can leave thousands of customers relying solely on limited essential and paid resources.
- Grozev and Buyya introduce and present taxonomies of federated cloud architectures, mechanism of application brokering, and the current environments.
- Formally, intercloud computing is defined as in “a cloud model that, for the purpose of guaranteeing service quality, such as the performance and availability of each service, allows on-demand reassignment of resources and transfer of workload through an interworking of cloud systems of different cloud providers based on coordination of each consumer’s requirements for service quality providers SLA and use of standard interfaces.”

Objective of Service Provisioning

The strategic objectives of provisioning cloud services have a paramount importance. Major objectives are as follows.

Fair Comparison

- One of the objectives of service provisioning is the fair comparison among the available services or with the CSP. Generally, users compare different cloud offerings according to their priorities and along several dimensions to select whatever is appropriate to their needs.

- It is a difficult task to perform an unbiased comparison and evaluation of all services. Several challenges must be addressed to develop an evaluation model that precisely measures the service level of each cloud provider. This study aims to provide a comparable service analysis for the cloud user to choose among desired services.

Compliance

- Service provisioning should comply with appropriate policies.
- The assurance of service compliance comes from the service providers.
- The CSP assures the customer of their compliance policies such as data protection, data confidentiality, and necessary data security by complying with the international compliance authority.
- NIST, ENISA, HIPAA, ISO 27001, and CSA are several compliance authorities who provide guidelines to establish the current cloud compliance security standards for the industry.

Prediction

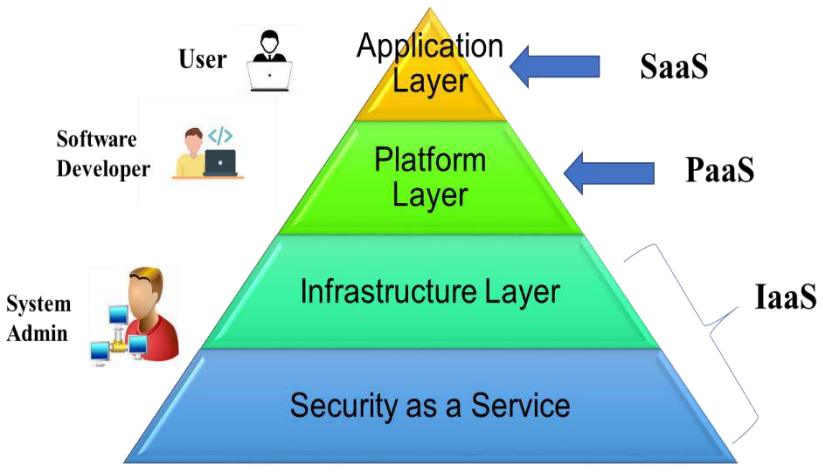
- Prediction is important in cloud service provisioning.
- A service user should be ensured of the elasticity and scalability of the services, even during peak hours or when the user suddenly makes an unusually high demand on the resources.
- In this situation, one of the objectives of the service provisioning selection is that the request should be instantly fulfilled by the service provider.
- Therefore, the user should be assured of the available required resources on demand with the predictable elastic and scalable services.

Rank

- Selecting the best and most appropriate service is a vital factor for the cloud service user.
- Selecting services depends on comparing and ranking them suitably.
- A reasonable and acceptable ranking system helps the cloud customer to make decisions about service selection.
- Therefore, the cloud service ranking system is an important aspect of a fair cloud service comparison and selection process.
- However, there is a lack of comparison of services across providers due to a lack of common comparable criteria or attributes.

Major Services of Service Provisioning

- In cloud computing, in the perspective of resource allocation and service provisioning, the services layers are divided into several working layers. There are then four service layers:
  - The application layer (SaaS)
  - The platform layer (PaaS)
  - The infrastructure layer (IaaS)
  - Security as a service (SecaaS)
- Each of these layers provides a specific service for users, which are explained as follows.

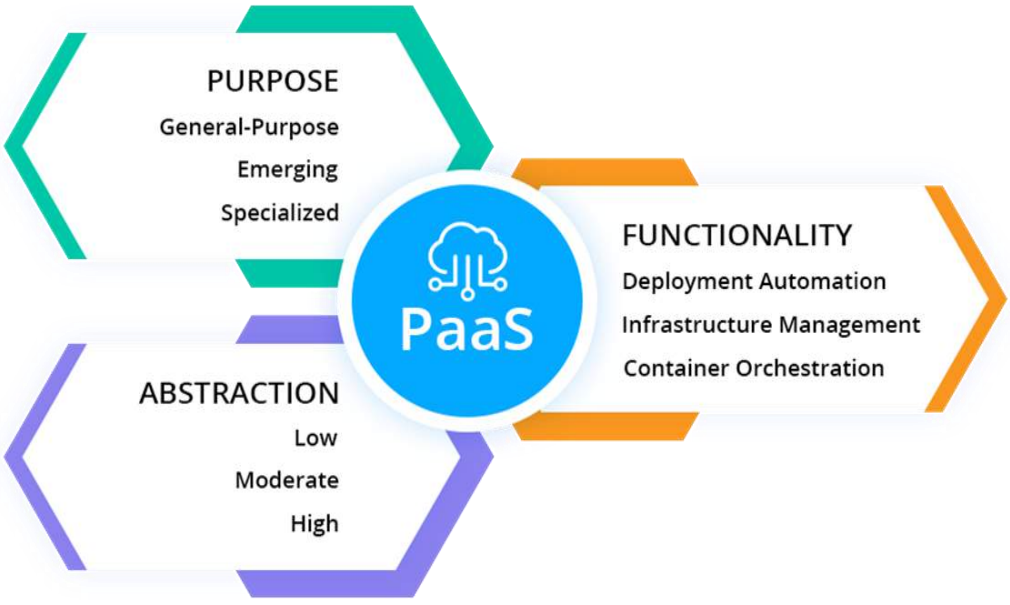


Infrastructure as a Service (IaaS)

- Infrastructure as a service, defined as providers who offer computing and storage resource capacity via vitalization, allowing physical resources to be assigned and split dynamically.
- A typical application could be an on-line alternative to a word processor or spreadsheet. Several types of virtualizations occur in this layer. Along with other resources, it includes computing, network, hardware, and storage.
- At the bottom layer of the framework, infrastructure devices and hardware are virtualised and provided as a service to users to install the operating system (OS) and to operate software applications.
- Therefore, this layer is called infrastructure as a service (IaaS). The Elastic Computing Cloud of Amazon (Amazon EC2) and storage by both Elastic Book Store (EBS) and Simple Storage Services (S3) are typical services of this layer.

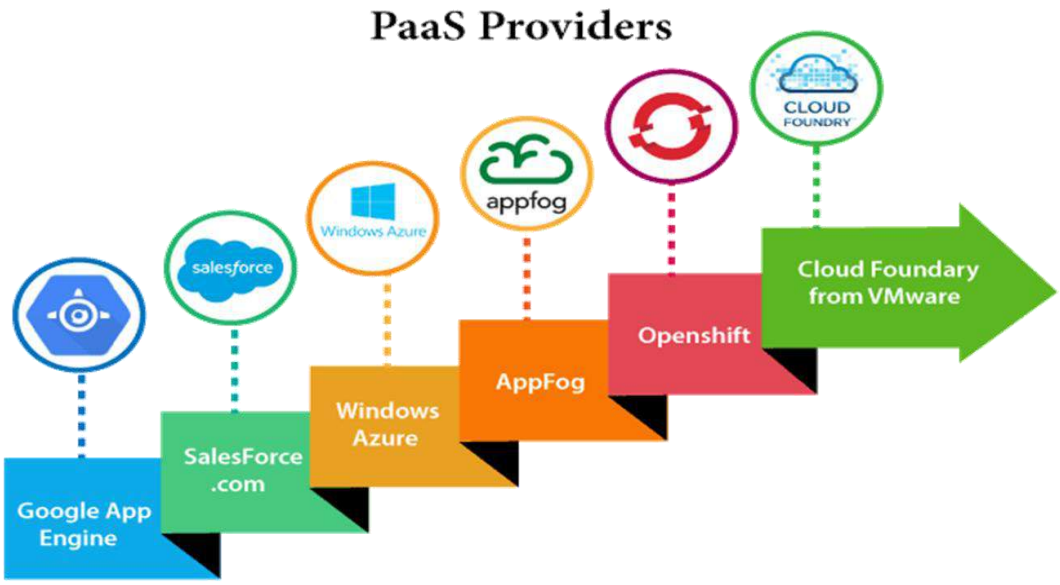
Platform as a Service (PaaS)

- Platform as a service, defined as a provider that offers an additional layer of abstraction above the virtualised infrastructure.
- The provided software platform trades off restrictions on the type of software that can be deployed for built-in scalability PaaS including mobile operating systems such as Android, iPhone, Symbian, and other OSs, as well as database management and IMS.
- This layer contains the environment for distributing storage, parallel programming design, the management system for organising distributed file systems, and other system management tools for cloud computing.



Platform as a Service (PaaS)

- Program developers are the primary clients of this platform layer. Entire platform resources such as program testing, running, maintaining, and debugging are delivered by the platform directly from this layer.
- Hence, this form of services in the platform layer is termed platform as a service (PaaS). Classic examples of these services include Google App Engine and Microsoft Azure.



Software as a Service (SaaS)

- Software as a service, defined as a provider who supplies remotely run software packages to consumers via the Internet on a utility-based pricing model.
- Analytical, interactive, transaction, and browsing facilities are included in the application layer.

- SaaS delivers several simple software programs and applications as well as customer interfaces to the end users. Thus, in the application layer, this type of service is called software as a service (SaaS).
- By using the client software or browser, the user can connect to services from providers via the Internet and pay fees according to the services consumed, in a pay-as-you-go model.
- Customer relationship management (CRM) from Salesforce is one of the early SaaS applications. Among other services, Google provides online office tools such as documentation, presentations, and spreadsheets, which are all part of SaaS.

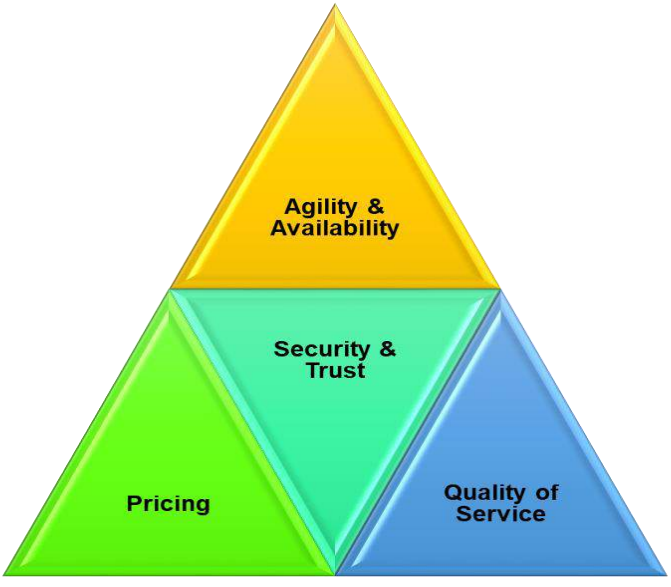
Security as a Service (SecaaS)

- The agility offered by the on-demand provisioning of computing resources and the ability to align information technology with business demands are valuable; however, clients are also very anxious about the security risks of cloud computing and the cost of direct control over the security of systems.
- Although vendors have attempted to satisfy this demand for security by offering security services in a cloud platform, the selection process is still completed.
- These issues have led to the restricted adoption of cloud-based security services, but the future looks bright for SecaaS, with Gartner predicting that cloud-based security service will be more than triple in many segments.
- To support both cloud customers and cloud providers, CSA has adopted a new research project to provide greater clarity in the area of SecaaS.
- It refers to the provision of security applications and services from the cloud to cloud-based infrastructure and software or from the cloud to the customers on premise systems.
- SecaaS will allow enterprises to make use of security services in new ways that would be more costly if provisioned locally.

Service Provisioning Requirements

There are several types of service provisioning from which we can make need-based selections, as discussed below.

- Agility & Availability
- Pricing
- Security & Trust
- Quality of Service



2.4 Tapping into an Infinite Storage Capacity

What is Cloud Storage?

- Cloud Storage is a cloud computing model that stores data on the Internet through cloud computing provider who manages and operate data storage as a service.
- It's delivered on demand with just in time capacity and costs and eliminates buying and managing your own data Storage Infrastructure.
- This gives Agility, Global Scale and Durability, with "anytime, anywhere" data access.

How does Cloud Storage Works?

- Cloud Storage is purchased from a third-party cloud vendor who owns and operates data storage capacity and delivers it over the Internet in a pay-as-you-go model.
- These cloud storage vendors manage capacity, security and durability to make a data accessible to your applications all around the world.
- Application access cloud storage through traditional storage protocols or directly via an API.
- Many vendors offer complementary services designed to help collect manage, secure and analyse data at massive scale.

Survey of Storage

- Companies such as Google, Amazon and Microsoft have been building massive data centers over the past few years.
- Spanning geographic and administrative domains, these data centers tend to be built out of commodity desktops with the total number of computers managed by these companies being in the order of millions.
- Additionally, the use of virtualization allows a physical node to be presented as a set of virtual nodes resulting in a seemingly inexhaustible set of computational resources.
- By leveraging economies of scale, these data centers can provision cpu, networking, and storage at substantially reduced prices which in turn underpins the move by many institutions to host their services in the cloud.
- Let's see what are the most dominant storage strategies that are currently being used in cloud computing settings. There are several unifying themes that underlie the systems.

Theme 1: Voluminous Data

- The datasets managed by these systems tend to be extremely voluminous. It is not unusual for these datasets to be several terabytes.
- The datasets also tend to be generated by programs, services and devices as opposed to being created by a user one character at a time.
- The amount of data being generated has been growing on an exponential scale there are growing challenges not only in how to effectively process this data, but also with basic storage.

Theme 2: Commodity Hardware

- The storage infrastructure for these datasets tend to rely on commodity hard drives that have rotating disks. This mechanical nature of the disk drives limits their performance.

- While processor speeds have grown exponentially disk access times have not kept pace. The performance disparity between processor and disk access times is in the order of 14,000,000:1 and continues to grow

Theme 3: Distributed Data

- A given dataset is seldom stored on a given node, and is typically distributed over a set of available nodes.
- This is done because a single commodity hard drive typically cannot hold the entire dataset.
- Scattering the dataset on a set of available nodes is also a precursor for subsequent concurrent processing being performed on the dataset.

Theme 4: Expect Failures

- Since the storage infrastructure relies on commodity components, failures should be expected.
- The systems thus need to have a failure model in place that can ensure continued progress and acceptable response times despite any failures that might have taken place.
- Often these datasets are replicated, and individual slices of these datasets have checksums associated with them to detect bit-flips and the concomitant data corruptions that often taken place in commodity hardware.

Theme 5: Tune for Access by Applications

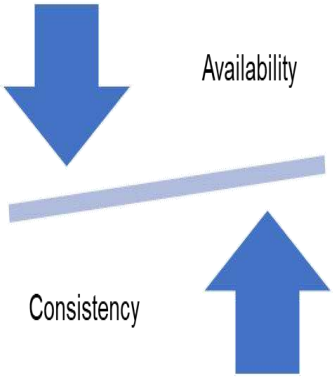
- Though these storage frameworks are built on top of existing file systems, the stored datasets are intended to be processed by applications and not humans. Since the dataset is scattered on a large number of machines, reconstructing the dataset requires processing the metadata (data describing the data) to identify the precise location of specific portions of the datasets.
- Manually accessing any of the nodes to look for a portion of the dataset is futile since these portions have themselves been modified to include checksum information.

Theme 6: Optimize for Dominant Usage

Another important consideration in these storage frameworks is optimizing the most general access patterns for these datasets. In some cases, this would mean optimizing for long, sequential reads that puts a premium on conserving bandwidth while in others it would involve optimizing small, continuous updates to the managed datasets.

Theme 7: Tradeoff Between Consistency and Availability

- Since these datasets are dispersed (and replicated) on a large number of machines accounting for these failures entails a trade-off between consistency and availability.
- Most of these storage frameworks opt for availability and rely on eventual consistency.



2.5 Cost-effective Pay-as-You-Use Billing Models

Pay-as-you-use

- **Pay-as-you-use** (or pay-per-use) is a payment model in cloud computing those charges based on resource usage. The practice is similar to the utility bills (e.g. electricity), where only actually consumed resources are charged.
- One major benefit of the pay-as-you-use method is that there are no wasted resources (that were reserved, but not consumed), which can be a source of significant losses for the companies. Users only pay for utilized capacities, rather than provisioning a chunk of resources that may or may not be used.

Payment Model Concept Evolution

Cost efficiency is one of the most distinctive and advertised benefits of cloud computing alongside the ease of use. Due to cloud computing rapid development, the utilized payment model is also evolving.

Payment Model	Charged when server is stopped?	Charged for Unused Resource?	Billing Period	Summary
Subscription	⬆	⬆	Long(Month/Year)	Charged full price at once. Similar to purchasing a server
Pay-as-You-Go	⬇	⬆	Short(Second/Hour)	Charged for the full capacity when the server is running. Similar to renting a server.
Pay-as-You-Stay	⬇	⬆	Short(Minutes/Hour)	Charged for the consumed hours. Similar to renting a car.
Pay-as-You-Use	⬇	⬇	Short(Second/Hour)	Charged for the consumed resources. Similar to renting resources.

Role in Solving the Right-Sizing Problem

- **Right-sizing** is a process of reserving the cloud computing instances (containers, VMs, or bare metal) with enough resources (RAM, CPU, storage, network) to achieve a sufficient performance at the lowest cost possible.
- Right-sizing aims to solve two problems in cloud computing:
  - **Over allocation**, which leads to inefficient utilization of the cloud infrastructure and overpayment for resources that are not actually used.
  - **Under allocation**, which results in resource shortage that causes performance issues or even downtime of the hosted projects, leading to the poor end-user experience, missed clients, and revenue losses.
- Currently, the pay-per-use model is the most efficient answer to the right-sizing problem.
- It allows avoiding manual prediction on the required server size by shifting this responsibility to the precise tools offered by modern cloud hosting providers.
- As a result, applications are automatically provided with the exact amount of resources to serve the on-going load.

2.6 Evaluating Barriers to Cloud Computing

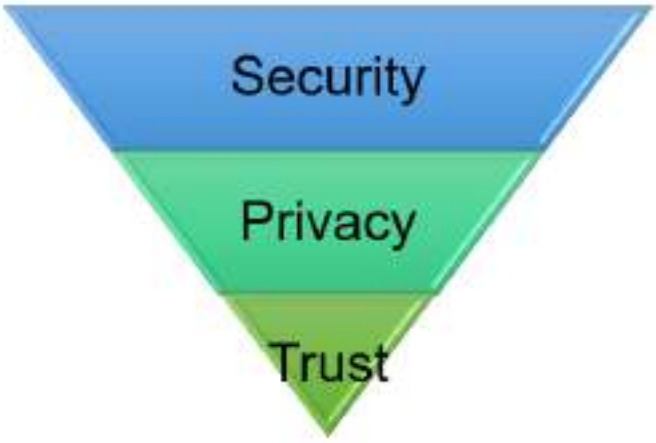
Challenges and Risks

- Despite the initial success and popularity of the cloud computing paradigm and the extensive availability of providers and tools, a significant number of challenges and risks are inherent to this new model of computing.
- Providers, developers, and end users must consider these challenges and risks to take good advantage of cloud computing.

- Security, Privacy, and Trust
- Data Lock-In and Standardization
- Availability, Fault-Tolerance, and Disaster Recovery
- Resource Management and Energy-Efficiency

Security, Privacy, and Trust

- "Current cloud offerings are essentially public exposing the system to more attacks." For this reason, there are potentially additional challenges to make cloud computing environments as secure as in-house IT systems.
- At the same time, existing, well understood technologies can be leveraged, such as data encryption, VLANs, and firewalls.



Data Lock-In and Standardization

- A major concern of cloud computing users is about having their data locked-in by a certain provider. Users may want to move data and applications out from a provider that does not meet their requirements.
- In their current form, cloud computing infrastructures and platforms do not employ standard methods of storing user data and applications. Consequently, they do not interoperate and user data are not portable.

Availability, Fault-Tolerance, and Disaster Recovery

- It is expected that users will have certain expectations about the service level to be provided once their applications are moved to the cloud.
- These expectations include availability of the service, its overall performance, and what measures are to be taken when something goes wrong in the system or its components.
- In summary, users seek for a warranty before they can comfortably move their business to the cloud.

Resource Management and Energy-Efficiency

- One important challenge faced by providers of cloud computing services is the efficient management of virtualized resource pools.
- Physical resources such as CPU cores, disk space, and network bandwidth must be sliced and shared among virtual machines running potentially heterogeneous workloads.
- Another challenge concerns the outstanding amount of data to be managed in various VM management activities. Such data amount is a result of particular abilities of virtual machines, including the ability of traveling through space (i.e., migration) and time (i.e., check pointing and rewinding), operations that may be required in load balancing, backup, and recovery scenarios.
- In addition, dynamic provisioning of new VMs and replicating existing VMs require efficient mechanisms to make VM block storage devices (e.g., image files) quickly available at selected hosts.

2.7 Handling Sensitive Data

Data Security in Cloud

Introduction to the Idea of Data Security

- Taking information and making it secure, so that only yourself or certain others can see it, is obviously not a new concept.
- It is one that we have struggled with in both the real world and the digital world. In the real world, even information under lock and key, is subject to theft and is certainly open to accidental or malicious misuse.
- In the digital world, this analogy of lock-and-key protection of information has persisted, most often in the form of container-based encryption.
- But even our digital attempt at protecting information has proved less than robust, because of the limitations inherent in protecting a container rather than in the content of that container.
- This limitation has become more evident as we move into the era of cloud computing: Information in a cloud environment has much more dynamism and fluidity than information that is static on a desktop or in a network folder, so we now need to start to think of a new way to protect information.
- Before we embark on how to move our data protection methodologies into the era of the cloud, perhaps we should stop, think, and consider the true applicability of information security and its value and scope.
- Perhaps we should be viewing the application of data security as less of a walled and impassable fortress and more of a sliding series of options that are more appropriately termed —risk mitigation.

- In a typical organization, the need for data security has a very wide scope, varying from information that is set as public domain, through to information that needs some protection (perhaps access control), through data that are highly sensitive, which, if leaked, could cause catastrophic damage, but nevertheless need to be accessed and used by selected users.
- Computer technology is the most modern form of the toolkit that we have developed since human prehistory to help us improve our lifestyle.
- From a human need perspective, arguably, computing is no better or worse than a simple stone tool, and similarly, it must be built to fit the hand of its user.
- Technology built without considering the human impact is bound to fail. This is particularly true for security technology, which is renowned for failing at the point of human error.
- If we can start off our view of data security as more of a risk mitigation exercise and build systems that will work with humans (i.e., human-centric), then perhaps the software we design for securing data in the cloud will be successful.

The Current State of Data Security in the Cloud

- Cloud computing has many arguing for its use because of the improved interoperability and cost savings it offers.
- On the other side of the argument are those who are saying that cloud computing cannot be used in any type of pervasive manner until we resolve the security issues inherent when we allow a third party to control our information.
- These security issues began life by focusing on the securing of access to the data centers that cloud-based information resides in. However, it is quickly becoming apparent in the industry that this does not cover the vast majority of instances of data that are outside of the confines of the data center, bringing us full circle to the problems of having a container-based view of securing data.

- This is not to say that data-center security is obsolete. Security, after all, must be viewed as a series of concentric circles emanating from a resource and touching the various places that the data go to and reside.
- The very 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.
- This apparent dichotomy data security with open movement of data is not as just a posed as it first seems.
- Security is better described as risk mitigation, we can then begin to look at securing data as a continuum of choice in terms of levels of accessibility and content restrictions: This continuum

allows us to choose to apply the right level of protection, ensuring that the flexibility bestowed by cloud computing onto the whole area of data communication is retained.

- The IT industry is beginning to wake up to the idea of content- centric or information-centric protection, being an inherent part of a data object.
- This new view of data security has not developed out of cloud computing, but instead is a development out of the idea of the deperimerization 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 Jericho Forum was founded in 2004 because of the increasing need for data exchange between companies and external parties for example: employees using remote computers; partner companies; customers; and so on.
- The old way of securing information behind an organization‘s perimeter wall prevented this type of data exchange in a secure manner. However, the ideas forwarded by the Jericho Forum are also applicable to cloud computing.

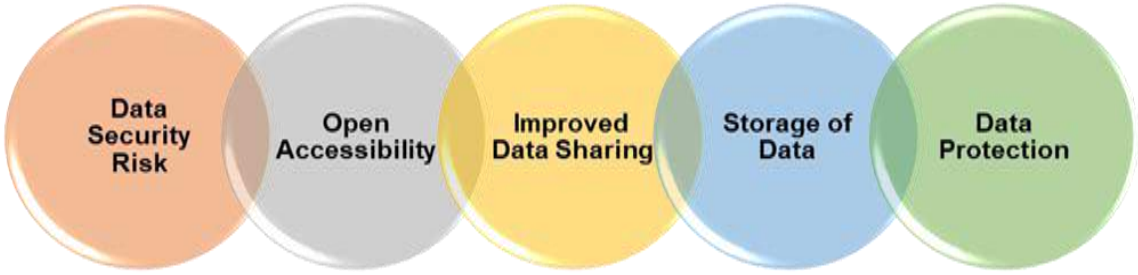
- The idea of creating, essentially, de-centralized perimeters, where the perimeters are created by the data object itself, allows the security to move with the data, as opposed to retaining the data within a secured and static wall.
- This simple but revolutionary change in mind set of how to secure data is the ground stone of securing information within a cloud and will be the basis of this discussion on securing data in the cloud.

2.8 Aspect of Cloud Security

Cloud Computing and Data Security Risk

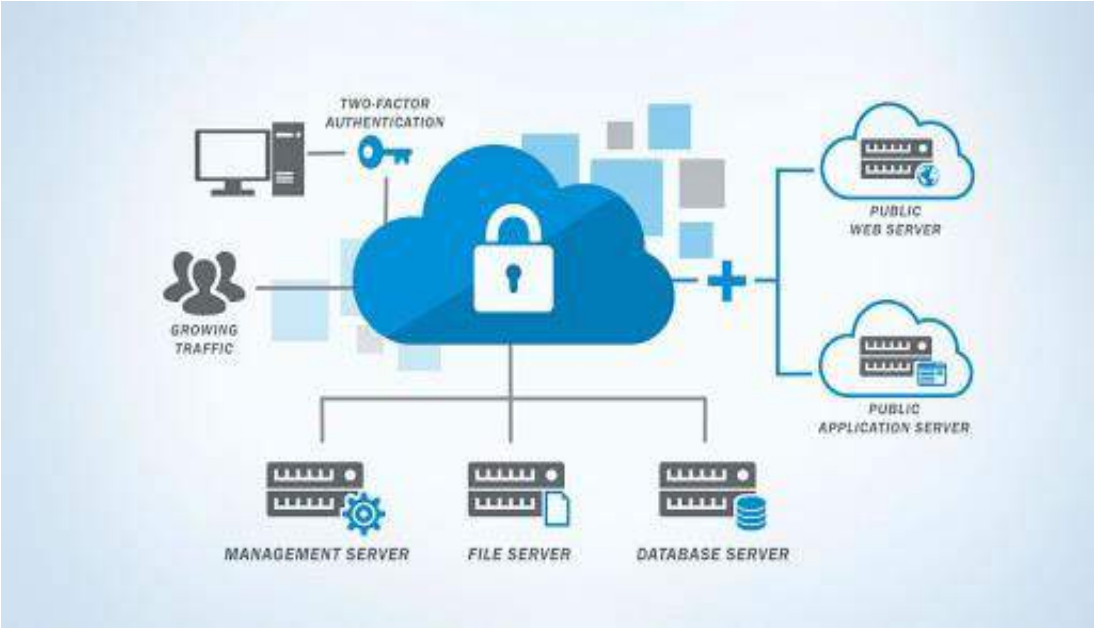
- Cloud computing is a development that is meant to allow more open accessibility and easier and improved data sharing.
- Data are uploaded into a cloud and stored in a data center, for access by users from that data center; or in a more fully cloud-based model, the data themselves are created in the cloud and stored and accessed from the cloud (again via a data center).
- 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 needs 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 that is, during a collaboration step as part of a document workflow process.
- Other issues that complicate the area of hosted data include ensuring that the various data security acts and rules are adhered to; this becomes particularly complicated when you consider the cross-border implications of cloud computing and the hosting of data in a country other than that originating the data.



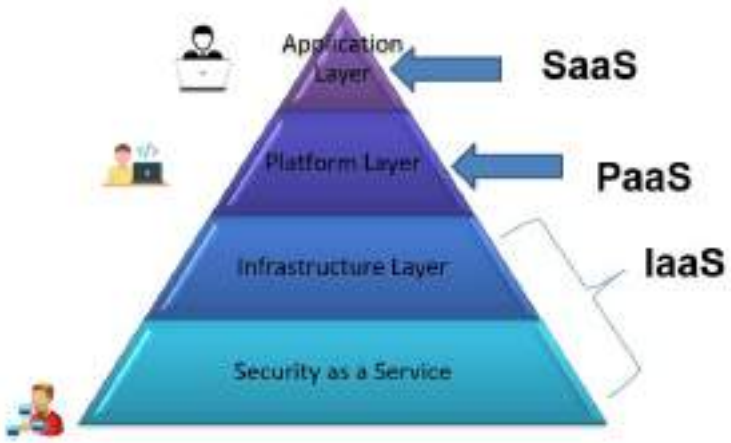
2.9 Assessing Governance Solutions

Governance in Cloud Computing

- The adaptation of cloud computing has forced many companies to recognize that clarity of ownership of the data is of paramount importance.
- The protection of intellectual property (IP) and other copyright issues is of big concern and needs to be addressed carefully.

Section 3: Exercises

Exercise 1: Write down the Respective Customer in-front of Respective Cloud Service.



Exercise 2: Participate in a group discussion on following topics:

a) Key Drivers of Cloud Computing Solutions
b) Self –Service Feature of Cloud Computing
c) Elastic feature
d) Customizable Nature of Cloud

Section 4: Assessment Questionnaire

1. What are the challenges in cloud adopting?
2. Why is adoption important?
3. What are the key features which drive cloud adoption?
4. Define Service provisioning?
5. What re the key objectives of Service provisioning?
6. Explain Infrastructure as a service (IaaS)?

7. Explain Platform as a service (PaaS)?
8. What are the layers of PAAS and give some examples of PAAS?
9. What is Software as a service (SaaS)?
10. What is cloud Storage?
11. How does cloud storage works?
12. What is Service Oriented Architecture (SOA)
13. What is Data security?

----------End of Module----------

MODULE 3

Software As A Service (SaaS) in Cloud Computing

Section 1: Learning Outcomes

After completing this module, you will be able to:

- Explain the concept of Software as a Service (SaaS)
- State Pros and Cons of SaaS
- Describe the SaaS Challenges
- Explain Integration Approach for SaaS
- Tell the Characteristics of SaaS

Section 2: Relevant Knowledge

3.1 Exploiting Software as a Service (SaaS)

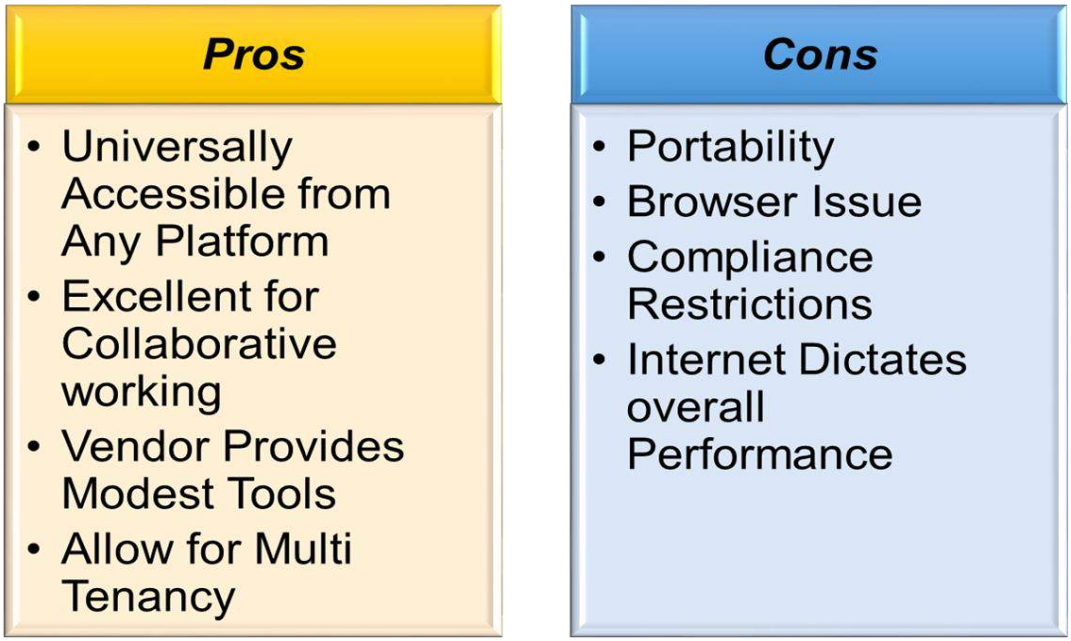
SaaS: Software as a Service

It Stands for "**Software as a Service**"

- On Demand Service
- Independent Platform
- No Need to install on PC
- Resource Management by Vendor

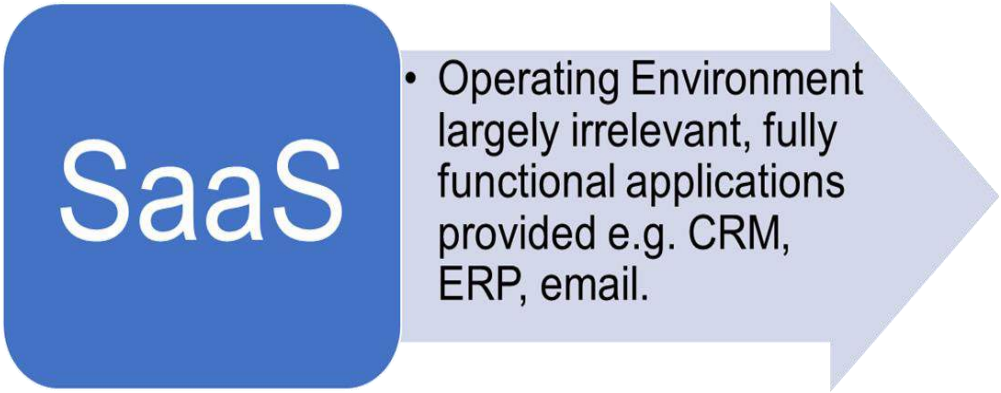
Who Use it? : - End Customers

Pros & Cons of SaaS



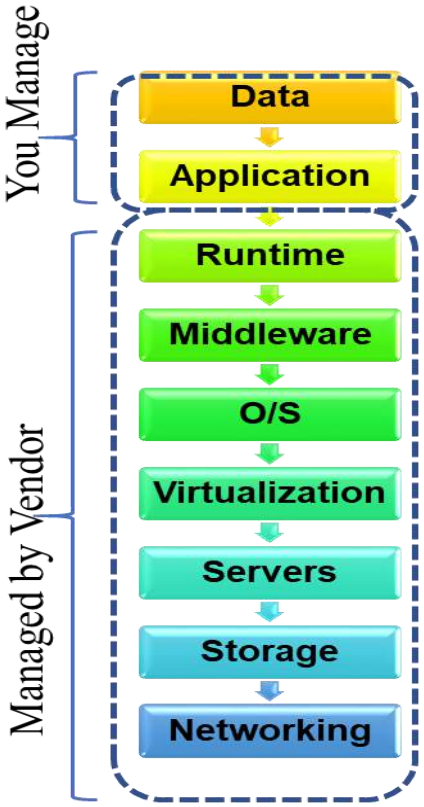
The Evolution of SaaS

- SaaS paradigm is on fast track due to its innate powers and potentials.
- Executives, entrepreneurs, and end-users are ecstatic about the tactic as well as strategic success of the emerging and evolving SaaS paradigm.
- A number of positive and progressive developments started to grip this model.
- Newer resources and activities are being consistently readied to be delivered as a IT as a Service (ITaaS) is the most recent and efficient delivery method in the decisive IT landscape.
- With the meteoric and mesmerizing rise of the service orientation principles, every single IT resource, activity and infrastructure is being viewed and visualized as a service that sets the tone for the grand unfolding of the dreamt service era. This is accentuated due to the pervasive Internet.



Software as a Service

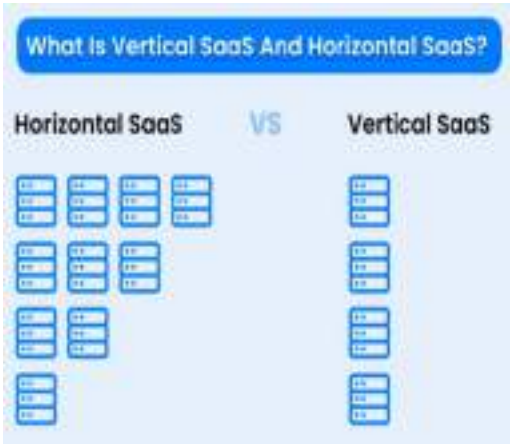
- Applications reside on the top of the cloud stack. Services provided by this layer can be accessed by end users through Web portals.
- Therefore, consumers are increasingly shifting from locally installed computer programs to on-line software services that offer the same functionally.
- Traditional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



Salesforce.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.

Vertical vs Horizontal SaaS

- Horizontal SaaS and vertical SaaS are different models of cloud computing services.
- Horizontal SaaS targets a broad variety of customers, generally without regard to their industry. Some popular examples of horizontal SaaS vendors are Salesforce and HubSpot.
- Vertical SaaS, on the other hand, refers to a niche market targeting a narrower variety of customers to meet their specific requirements.



1. Business Model

Business models make a business successful, valid for SaaS business models. A business model is how a company will profit from its products and services or why it believes it can charge customers.

- A vertical Saas business model focuses on solving the needs of one particular industry, such as real estate or healthcare. It provides an end-to-end solution for the needs of a specific sector. Companies like Zillow and ZocDoc are good examples of vertical SaaS companies. They have created solutions for the real estate and healthcare industries.
- Horizontal SaaS business models provide solutions to everyday needs across different industries. They offer services that can be used by many kinds of businesses in any industry, such as accounting or employee scheduling software. Some examples include QuickBooks and WhenIWork.

2. **Target Market**

A target market is the consumers most likely to buy what you sell. When creating a target market analysis, you define the ideal customers for your product or service. Defining a target market is not about limiting your customer base. It's about identifying who will buy from you and why.

- A vertical SaaS is an application aimed at a specific industry, such as health care or banking. A vertical SaaS has lower marketing costs since the target audience is limited. The key here is to find a niche market where your software solves a big problem.
- Horizontal SaaS companies offer a generic product or service that serves various industries. They cater to a diverse customer base and typically have a lower barrier to entry for new customers.

3. **Competitive Landscape**

Competitive analysis is an essential part of your overall product strategy for SaaS companies. One can do competitive analysis relatively quickly for a horizontal SaaS company. There are standard tools that most companies use, such as Google Alerts and SEMRush. These tools allow you to track competitors' rankings, keywords, and traffic volume over time.

- Horizontal SaaS companies need to monitor the traditional factors that impact their industry: product features, pricing, brand value, and customer reviews.
- Vertical Saas companies need to analyze the competitive landscape of their niche market. For example, if you're in real estate software, you'll want to understand the search queries used by potential buyers and sellers. You'll also want to understand the advertising options available on relevant feeds, i.e., Facebook and Instagram.

4. **Marketing**

When it comes to marketing strategies, horizontal SaaS is focused on user acquisition, and vertical SaaS concentrates on customer retention.

- The goal of horizontal SaaS is to get as many users as possible using their software. So, they have a high market share and can eventually charge more for their product once they've established themselves. They often offer their product for free or at a low cost and then set a premium for the extra features users need to pay more to access.
- They rely heavily on user feedback and use it to adjust the features they provide and how they market their product. Their key metric is user adoption and how often users opt-in to use the software.
- Vertical SaaS products are built with a specific industry or group of people (i.e., real estate agents, hospitals). So the goal is not necessarily to attract as many new users as possible but rather to establish trust with existing customers so that those customers stay loyal to their service over time. To keep those customers reliable, vertical SaaS companies will sometimes offer free trials. Thereby allowing potential customers to try out the software before buying it.

5. Capital Efficiency for IPO

- Horizontal SaaS companies can have lower customer acquisition costs (CAC) than vertical because they take advantage of economies of scale. It means that horizontal companies can typically spend less money on marketing per client.
- On the other hand, Vertical SaaS companies may have much higher CACs when going public because they are not able to take advantage of economies of scale in the same way as horizontal companies.

6. Growth Prospects

When it comes to growth prospects, a horizontal SaaS company wins out. Here's why:

- Horizontal SaaS companies sell their product to everyone in the industry. Specific client needs don't limit them; they can meet any demand. It means that there's no limit on the number of potential customers!
- On the other hand, Vertical SaaS companies specialize in one area of their industry, selling only to those with particular needs. It limits the number of potential customers they can have.

B2B vs B2C SaaS Products

The final differentiator between these cloud services is their intended target audience. The functionality, design, and even pricing model is vastly different between B2B and B2C products.

What Is a B2B SaaS Product?

- B2B or business-to-business SaaS products are hosted software solutions designed to solve business problems.
- Think software solutions like CRM, ecommerce platforms, analytics, and more.

What Is a B2C SaaS Product?

- B2C or business-to-consumer SaaS products are cloud-based software solutions designed to solve individual problems.
- Think online editors, file sharing, website builders, streaming services, and even social media networks.

The Challenges of SaaS Paradigm

- As with any new technology, SaaS and cloud concepts too suffer a number of limitations.
- These technologies are being diligently examined for specific situations and scenarios.
- The prickling and tricky issues in different layers and levels are being looked into.
- The overall views are listed out below. Loss or lack of the following features deters the massive adoption of clouds

Integration Conundrum: While SaaS applications offer outstanding value in terms of features and functionalities relative to cost, they have introduced several challenges specific to integration. The first issue is that the majority of SaaS applications are point solutions and service one line of business.

APIs are Insufficient: Many SaaS providers have responded to the integration challenge by developing application programming interfaces (APIs). Unfortunately, accessing and managing data via an API requires a significant amount of coding as well as maintenance due to frequent API modifications and updates.

Data Transmission Security: SaaS providers go to great length to ensure that customer data is secure within the hosted environment. However, the need to transfer data from on-premise systems or applications behind the firewall with SaaS applications hosted outside of the client‘s data center poses new challenges that need to be addressed by the integration solution of choice.

The Impacts of Cloud:

- On the infrastructural front, in the recent past, the clouds have arrived onto the scene powerfully and have extended the horizon and the boundary of business applications, events and data.
- That is, business applications, development platforms etc. are getting moved to elastic, online and on-demand cloud infrastructures.
- Increasingly for business, technical, financial and green reasons, applications and services are being readied and relocated to highly scalable and available clouds.

Important factors for good design of SAAS model

- Three distinct points that separates a well-design from a poorly designed SAAS application:
  - Scalability
  - Multi-tenant efficient
  - Configurable
- **Scalability-** maximizing concurrency, and efficient use of resources i.e. optimizing locking duration, statelessness, sharing pooled resources such as threads and network connections, caching reference data, and partitioning large databases

- **Configurable -** a single application instance on a single server has to accommodate users from several different companies. Customizing the application for one customer will change the application for other customers as well.
- Traditionally customizing an application would mean changes in the code.
- Each customer must use metadata to configure the way the application appears and behaves for its users.
- Customers configuring applications must be simple and easy without any extra development or operation costs.

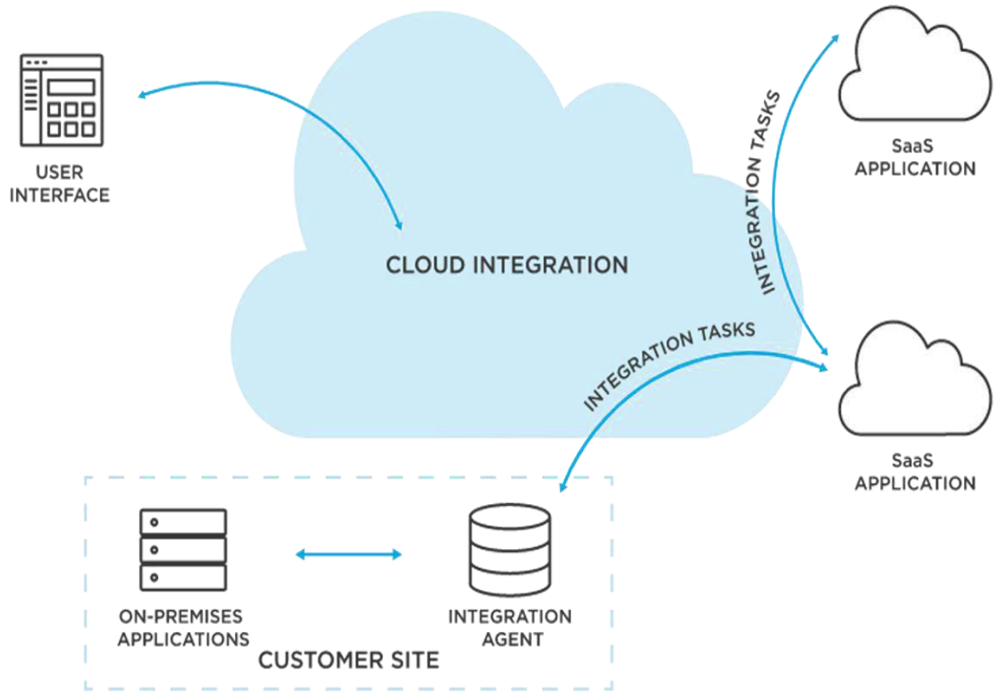
3.2 Streamlining Administration with Centralized Installation

Desktop as a Service

- Desktop as a Service is a special variant of Software as a Service that provides a virtualized desktop-like personal workspace, and sends its image to the user's real desktop.
- Instead of a local desktop, the user can access their own desktop-on-the cloud from different places for convenience, and receive the benefit of SaaS at same time.

Approaching the SaaS Integration Enigma

- Integration as a Service (IaaS) is all about the migration of the functionality of a typical enterprise application integration (EAI) hub / enterprise service bus (ESB) into the cloud for providing for smooth data transport between any enterprise and SaaS applications. Users subscribe to IaaS as they would do for any other SaaS application.



The Integration Methodologies

Excluding the custom integration through hand-coding, there are three types for cloud integration:

- **Traditional Enterprise Integration Tools can be empowered with special connectors to access Cloud-located Applications**—This is the most likely approach for IT organizations, which have already invested a lot in integration suite for their application integration needs.

- **Traditional Enterprise Integration Tools are hosted in the Cloud**—This approach is similar to the first option except that the integration software suite is now hosted in any third-party cloud infrastructures so that the enterprise does not worry about procuring and managing the hardware or installing the integration software.
- **Integration-as-a-Service (IaaS) or On-Demand Integration Offerings**— These are SaaS applications that are designed to deliver the integration service securely over the Internet and are able to integrate cloud applications with the on-premise systems, cloud-to-cloud applications.

SaaS Integration Services

- There are fresh endeavours in order to achieve service composition in cloud ecosystem.
- Existing frameworks such as service component architecture (SCA) are being revitalized for making it fit for cloud environments.
- Composite applications, services, data, views and processes will be become cloud-centric and hosted in order to support spatially separated and heterogeneous systems.
  - ø Informatica On-Demand
  - ø Microsoft Internet Service Bus (ISB)

3.3 Optimizing Cost and Performance with Scale on Demand

Services Provided by SaaS Providers

- There are the following services provided by SaaS providers -

Business Services - SaaS Provider provides various business services to start-up the business. The SaaS business services include ERP (Enterprise Resource Planning), CRM (Customer Relationship Management), billing, and sales.

Document Management - SaaS document management is a software application offered by a third party (SaaS providers) to create, manage, and track electronic documents.
Example: Slack, Samepage, Box, and Zoho Forms.

Social Networks - As we all know, social networking sites are used by the general public, so social networking service providers use SaaS for their convenience and handle the general public's information.

Mail Services - To handle the unpredictable number of users and load on e-mail services, many e-mail providers offering their services using SaaS.

Google Apps

- Google Apps (2010) is a typical SaaS implementation.
- It provides several Web applications with similar functionality to traditional office software (word processing, spreadsheets etc.), but also enables users to communicate, create and collaborate easily and efficiently.
- Since all the applications are kept online and are accessed through a web browser, users can access their accounts from any internet-connected computer, and there is no need to install anything extra locally.

- Google Apps has several components.
- The communication components consist of:

Google Mail

Google Talk

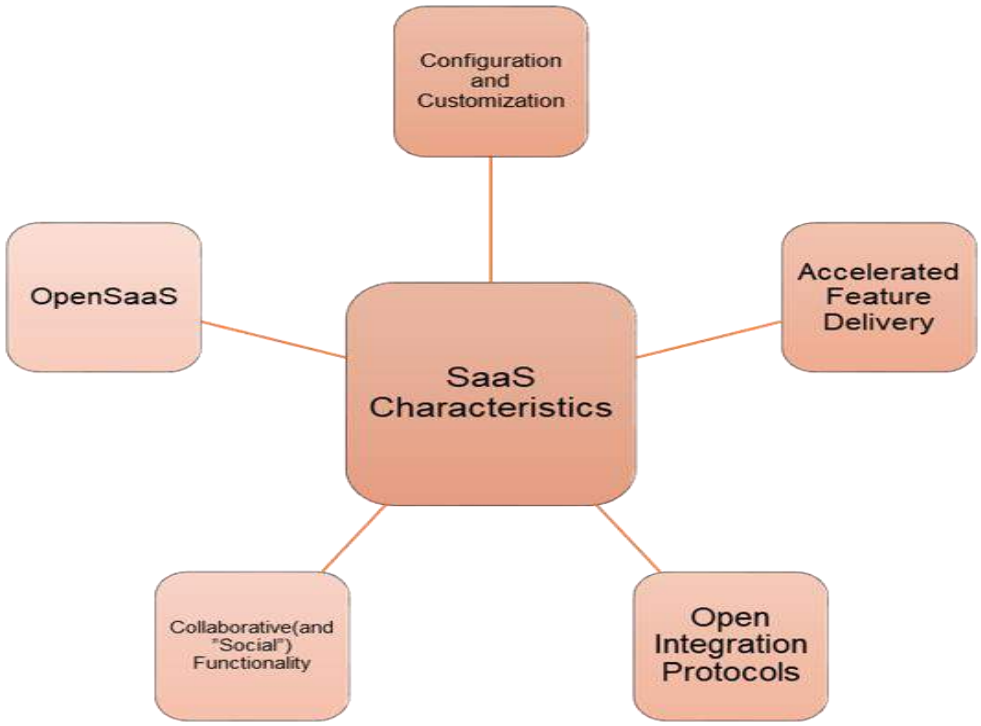
- These components which allow for communication through email, instant messaging and voice calls.
- The office components include **docs** and **spreadsheets**, through which users can create online documents that also facilitate searching and collaboration.
- **Google Calendar** is a flexible calendar application for organizing meetings and events. With Google's "Web Pages", administrators can easily publish web pages, while "Start Pages" provide users with a rich array of content and applications that can be personalized.

3.4 Characterizing SaaS

Software as a Service - Characteristics

Although not all software-as-a-service applications share all the following traits, the characteristics below are common among many of them:

- Configuration and customization
- Accelerated Feature Delivery
- Open Integration Protocols
- Collaborative (and “Social”) Functionality
- OpenSaaS



Configuration and Customization

- SaaS applications similarly support what is traditionally known as application configuration. In other words.
- Like traditional enterprise software, a single customer can alter the set of configuration options (a.k.a. parameters) that affect its functionality and look-and-feel. Each customer may have its own settings (or: parameter values) for the configuration options.
- The application can be customized to the degree it was designed for based on a set of predefined configuration options.
- To support customers' common need to change an application's look-and-feel so that the application appears to be having the customer's brand (or if so desired co-branded), many SaaS applications let customers provide (through a self-service interface or by working with application provider staff) a custom logo and sometimes a set of custom colors.
- The customer cannot, however, change the page layout unless such an option was designed.

Accelerated Feature Delivery

- SaaS applications are often updated more frequently than traditional software, in many cases on a weekly or monthly basis. This is enabled by several factors:
  - The application is hosted centrally, so an update is decided and executed by the provider, not by customers.
  - The application only has a single configuration, making development testing faster.
  - The application vendor does not have to expend resources updating and maintaining backdated versions of the software, because there is only a single version.
  - The application vendor has access to all customer data, expediting design and regression testing.
  - The service provider has access to user behavior within the application (usually via web analytics), making it easier to identify areas worthy of improvement.
- Accelerated feature delivery is further enabled by agile software development methodologies. Such methodologies, which have evolved in the mid-1990s, provide a set of software development tools and practices to support frequent software releases.

Open Integration Protocols

- Because SaaS applications cannot access a company's internal systems (databases or internal services), they predominantly offer integration protocols and application programming interfaces (APIs) that operate over a wide area network.

- The ubiquity of SaaS applications and other Internet services and the standardization of their API technology has spawned the development of mashups, which are lightweight applications that combine data, presentation, and functionality from multiple services, creating a compound service.
- Mashups further differentiate SaaS applications from on-premises software as the latter cannot be easily integrated outside a company's firewall.

Collaborative (and "Social") Functionality

- Inspired by the development of the different internet networking services and the so-called web 2.0 functionality, many SaaS applications offer features that let their users collaborate and share information.
- For example, many project management applications delivered in the SaaS model offer—in addition to traditional project planning functionality—collaboration features letting users comment on tasks and plans and share documents within and outside an organization.
- Several other SaaS applications let users vote on and offer new feature ideas.
- Although some collaboration-related functionality is also integrated into on-premises software, (implicit or explicit) collaboration between users or different customers is only possible with centrally hosted software.

OpenSaaS

- OpenSaaS refers to software as a service (SaaS) based on open-source code. Similar to SaaS applications, Open SaaS is a web-based application that is hosted, supported, and maintained by a service provider.
- While the roadmap for Open SaaS applications is defined by its community of users, upgrades and product enhancements are managed by a central provider. The term was coined in 2011 by Dries Buytaert, creator of the Drupal content management framework.

3.5 Comparing Service Scenarios

Advantages of SaaS

1) SaaS is easy to buy

- SaaS pricing is based on a monthly fee or annual fee subscription, so it allows organizations to access business functionality at a low cost, which is less than licensed applications.

- Unlike traditional software, which is sold as a licensed based with an up-front cost (and often an optional ongoing support fee), SaaS providers are generally pricing the applications using a subscription fee, most commonly a monthly or annually fee.

2) One to Many

- SaaS services are offered as a one-to-many model means a single instance of the application is shared by multiple users.

3) Less hardware required for SaaS

- The software is hosted remotely, so organizations do not need to invest in additional hardware.

4) Low maintenance required for SaaS

- Software as a service removes the need for installation, set-up, and daily maintenance for the organizations. The initial set-up cost for SaaS is typically less than the enterprise software. SaaS vendors are pricing their applications based on some usage parameters, such as a number of users using the application. So SaaS does easy to monitor and automatic updates.

5) No special software or hardware versions required

- All users will have the same version of the software and typically access it through the web browser. SaaS reduces IT support costs by outsourcing hardware and software maintenance and support to the IaaS provider.

6) Multidevice support

- SaaS services can be accessed from any device such as desktops, laptops, tablets, phones, and thin clients.

7) API Integration

- SaaS services easily integrate with other software or services through standard APIs.

8) No client-side installation

- SaaS services are accessed directly from the service provider using the internet connection, so do not need to require any software installation.

Disadvantages of SaaS

1) Security

Actually, data is stored in the cloud, so security may be an issue for some users. However, cloud computing is not more secure than in-house deployment.

2) Latency issue

Since data and applications are stored in the cloud at a variable distance from the end-user, there is a possibility that there may be greater latency when interacting with the application compared to

local deployment. Therefore, the SaaS model is not suitable for applications whose demand response time is in milliseconds.

3) Total Dependency on Internet
Without an internet connection, most SaaS applications are not usable.

4) Switching between SaaS vendors is difficult
Switching SaaS vendors involves the difficult and slow task of transferring the very large data files over the internet and then converting and importing them into another SaaS also.

Applicability of SAAS

- Enterprise Software application
- Sharing of data between internal and external users e.g.
  - Salesforce CRM application
  - Single user Software application
- Runs on single user computer and serves 1 user at a time e.g. : Microsoft office
- Business Utility SaaS - Applications like Salesforce automation are used by businesses and individuals for managing and collecting data, streamlining collaborative processes and providing actionable analysis. Popular use cases are Customer Relationship Management (CRM), Human Resources and Accounting.
- Social Networking SaaS - Applications like Facebook are used by individuals for networking and sharing information, photos, videos, etc.

Consideration for SAAS Application development

Other Software as Service Examples

SaaS Provider	Important Services
A2Zapps.com(2010)	Marketing Automation, School Automation(ERP)
Envysion.com(2010)	Video Management
Learn.com(2010)	Training, HR, Online Courses
Microsoft(2010)	Office Live Meeting, Dynamics CRM, SharePoint
OpenID(2010)	Log in Identificatoin
Zoho(2010)	Mail, Docs, Wiki, CRM, Meeting, Business

3.6 Inspecting SaaS technologies

SaaS-Adoption drivers

Several important changes to the software market and technology landscape have facilitated the acceptance and growth of SaaS:

- The growing use of web-based user interfaces by applications, along with the proliferation of associated practices (e.g., web design), continuously decreased the need for traditional client-server applications.
- Consequently, traditional software vendor's investment in software based on fat clients has become a disadvantage (mandating ongoing support), opening the door for new software vendors' offering a user experience perceived as more "modern".
- The standardization of web page technologies (HTML, JavaScript, CSS), the increasing popularity of web development as a practice, and the introduction and ubiquity of web application frameworks like Ruby on Rails or Laravel (PHP) gradually reduced the cost of developing new software services and enabled new providers to challenge traditional vendors.
- The increasing penetration of broadband Internet access enabled remote centrally hosted applications to offer speed comparable to on-premises software.
- The standardization of the HTTPS protocol as part of the web stack provided universally available lightweight security that is sufficient for most everyday applications.
- The introduction and wide acceptance of lightweight integration protocols such as Representational State Transfer (REST) and SOAP enabled affordable integration between SaaS applications (residing in the cloud) with internal applications over wide area networks and with other SaaS applications.

SaaS-Adoption challenges

Some limitations slow down the acceptance of SaaS and prohibit it from being used in some cases:

- Because data is stored on the vendor's servers, data security becomes an issue.
- SaaS applications are hosted in the cloud, far away from the application users. This introduces latency into the environment; for example, the SaaS model is not suitable for applications that demand response times in milliseconds (OLTP).

- Multi-tenant architectures, which drive cost efficiency for service providers, limit customization of applications for large clients, inhibiting such applications from being used in scenarios (applicable mostly to large enterprises) for which such customization is necessary.
- Some business applications require access to or integration with customers' current data. When such data are large in volume or sensitive (e.g. end-user's personal information), integrating them with remotely hosted software can be costly or risky, or can conflict with data governance regulations.
- Constitutional search/seizure warrant laws do not protect all forms of SaaS dynamically stored data. The result is that a link is added to the chain of security where access to the data, and, by extension, misuse of these data, are limited only by the assumed honesty of third parties or government agencies able to access the data on their recognizance.
- Switching SaaS vendors may involve the slow and difficult task of transferring very large data files over the Internet.
- Organizations that adopt SaaS may find they are forced into adopting new versions, which might result in unforeseen training costs, an increase in the probability that a user might make an error or instability from bugs in the newer software.
- Should the vendor of the software go out of business or suddenly EOL the software, the user may lose access to their software unexpectedly, which could destabilize their organization's current and future projects, as well as leave the user with older data they can no longer access or modify.
- Relying on an Internet connection means that data is transferred to and from a SaaS firm at Internet speeds, rather than the potentially higher speeds of a firm's internal network.
- The Ability of the SaaS hosting company to guarantee the uptime level agreed in the SLA (Service Level Agreement)
- The reliance on SaaS applications and services can lead to SaaS sprawl within enterprises. These disparate applications and services can become challenging to maintain technically and administratively, leading to the proliferation of shadow IT.

The standard model also has limitations:

- Compatibility with hardware, other software, and operating systems.
- Licensing and compliance problems (unauthorized copies of the software program putting the organization at risk of fines or litigation).
- Maintenance, support, and patch revision processes.

Popular SaaS Providers

Provider	Services
Salseforce.com	On-demand CRM solutions
Microsoft Office 365	Online office suite
Google Apps	Gmail, Google Calendar, Docs, and sites
NetSuite	ERP, accounting, order management, CRM, Professionals Services Automation (PSA), and e-commerce applications.
GoToMeeting	Online meeting and video-conferencing software
Constant Contact	E-mail marketing, online survey, and event marketing
Oracle CRM	CRM applications
Workday, Inc	Human capital management, payroll, and financial management.

Section 3: Exercises

Exercise 1: Write down Characteristics of Software as a Service in below diagram.

Exercise 2: Write Examples of Horizontal and Vertical SaaS Service Provider.

Horizontal SaaS	Vertical SaaS

Exercise 3: Participate in a group discussion on following topics:

a) Concept of Software as a Service (SaaS)
b) Pros and Cons of SaaS
c) SaaS Challenges
d) Characteristics of SaaS

Section 4: Assessment Questionnaire

1. Define SaaS?
2. What are the Benefits of SaaS?
3. Which are different types of SaaS?
4. Give 2-3 Examples of SaaS.
5. What Is a B2B SaaS Product?
6. What Is a B2C SaaS Product?
7. What are the Characteristics of SaaS?
8. List Few of the SaaS Provider.
9. List Fes of Limitation of SaaS.
10. What are the Disadvantages of SaaS?

----------End of Module----------

MODULE 4

Delivering Platform as a Service (PaaS)

Section 1: Learning Outcomes

After completing this module, you will be able to:

- Explain the concept of Platform as a Service (PaaS)
- State Pros and Cons of PaaS
- Define PaaS Architecture
- Describe PaaS and Its Services
- Explain the PaaS Monitoring
- Tell the Benefits of Cloud Monitoring

Section 2: Relevant Knowledge

4.1 Delivering Platform as a Service (PaaS)

PaaS: Platform as a Service

It Stands for **"Platform as a Service"**

- Programming Language + OS + Server + Database
- Provides Encapsulation
- Build, Compile & Run Programs
- Users Manage Data & Application Resources

Who Use it? : - Developers

Pros & Cons of PaaS

Pros	Cons
• Scalable & Cost Effective • Faster Market for Developers • Easy Development for Web Application • Private & Public Deployment	• Provider Language Only • Developers Limited • Migration Issues • Vendor Lock-in

PaaS Market Size, Share, and Leading Vendors

- The PaaS market's reported size and how it compares to other cloud services depend on the source.
- For example, according to Gartner, PaaS will be dwarfed by IaaS in 2021, with $27.5 billion vs. $61.9 billion in revenue, respectively.

Worldwide Public Cloud Service Revenue Forecast (Billions of U.S. Dollars)

	2018	2019	2020	2021	2022
Cloud Business Process Services (BPaaS)	45.8	49.3	53.1	57.0	61.1
Cloud Application Infrastructure Services (PaaS)	15.6	19.0	23.0	27.5	31.8
Cloud Application Services (SaaS)	80.0	94.8	110.5	126.7	143.7
Cloud Management and Security Services	10.5	12.2	14.1	16.0	17.9
Cloud System Infrastructure Services (IaaS)	30.5	38.9	49.1	61.9	76.6
Total Market	**182.4**	**214.3**	**249.8**	**289.1**	**331.2**

BPaaS = business process as a service; IaaS = infrastructure as a service; PaaS = platform as a service; SaaS = software as a service
Note: Totals may not add up due to rounding.

PaaS- Delivery Ways

PaaS can be delivered in three ways:

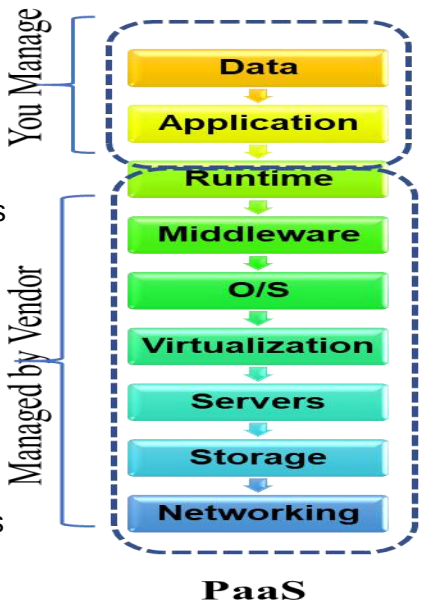
- As a public cloud service from a provider, where the consumer controls software deployment with minimal configuration options, and the provider provides the networks, servers, storage, operating system (OS), middleware (e.g. Java runtime, .NET runtime, integration, etc.), database and other services to host the consumer's application.
- As a private service (software or appliance) behind a firewall.
- As software deployed on public infrastructure as a service.

How does PaaS work?

- PaaS does not replace a company's entire IT infrastructure for software development. It is provided through a cloud service provider's hosted infrastructure.
- Users most frequently access the offerings through a web browser.
- PaaS can be delivered through public, private and hybrid clouds to deliver services such as application hosting and Java development.
- Other PaaS services include the following:
  - Development team collaboration
  - Application design and development
  - Application testing and deployment
  - Web service integration
  - Information security
  - Database integration
- Users will normally have to pay for PaaS on a per-use basis. However, some providers charge a flat monthly fee for access to the platform and its applications.

PaaS Architecture

- PaaS enables developers to develop, test, and deploy in the same environment. A typical PaaS architecture consists of the following categories:
  - Integration and Middleware: It refers to the software that offers runtime services.
  - API: It implies Application Platform Interface, which acts as a communication between client and server that offers abstraction (running the details in the background) and core connectivity.
  - Hardware: It comprises of all hard requirements to handle the resources.
- This facilitates and allows the users to build and run applications without the complexity of constructing and maintaining the infrastructure as the PaaS architecture covers the requirements.



Understanding PaaS with Types of Services

PaaS leads to faster development as there is no need for the user to worry about setting and maintaining the infrastructure. PaaS services are available in 3 types:

- Public
- Private
- Hybrid

What Services Does PaaS Include?

Although the most common use case of PaaS is web app deployment, many other cloud services also fall under it.

- Database as a Service (DBaaS)
- Internet of Things (IoT) Platforms
- Mobile Services (APIs)
- Push Notification APIs
- Machine Learning
- Hadoop, Spark, & Other Data Processing Frameworks

Database as a Service (DBaaS)

- A cloud-hosted database that you manually install on a virtual machine is only an implementation of IaaS.
- To be considered a PaaS offering, it needs to be an integrated solution that offers storage, computing power, and relational database capabilities.
- An example of this is the Azure SQL Database service, which offers a fully managed database with automated updates, scalability, smart threat protection, and AI-powered search

Internet of Things (IoT) Platforms

- More items are powered by computers and connected to the internet than ever before.
- The new HTTP/3 standard will only accelerate that further.
- Connected devices now include lights, thermostats, ovens, washing machines, locks, and even truck engines.
- The bare bones of connectivity to the internet could be considered IaaS, but complex APIs for controlling and sharing data across devices and apps fall under PaaS.

Mobile Services (APIs)

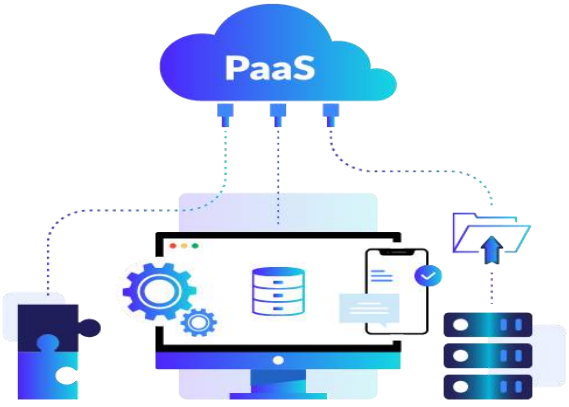
- Companies are no longer settling for email when sending notifications and marketing campaigns to their customers.
- They also use automated SMS messages at scale.
- With SMS APIs, companies can build automated messages into their applications.
- For example, they can text customers to:
  - Remind them of scheduled calls or meetings.
  - Promote a new related product or service.
  - Ask for feedback on a recent customer service encounter.
  - Recruit customers to join a case study or survey.
  - These services are sometimes categorized separately as Communications Platform as a Service (CPaaS), a PaaS subcategory.

Machine Learning

- If you genuinely want to take advantage of your data, it's not enough to just store it in the cloud. The data is still just sitting around, only in a new location.
- You need to set up algorithms to sift through your data and find meaningful insights and actionable steps.
- With cloud-based machine learning platforms, you can easily create models (from templates), apply them to your databases, and scale your computing power as needed.

Public PaaS

- Public Platform as a Service runs on the public cloud the user have to focus on building application.
- It helps developers to be more agile, which helps them to develop and deliver faster. And the vendor manages and maintains the infrastructure.



Private PaaS

- A private Platform as a Service is a good choice for companies that wish to maintain some of their own hardware.
- It's also a good alternative for companies who wish to maintain part of their information, in some cases sensitive, in their own data centers.

Hybrid PaaS

- Hybrid Platform as a Service offers flexibility to choose what percent of the user's infrastructure in his control.
- Private PaaS provides scalability for hybrid PaaS. Well, a hybrid is a combination of a bit of private and public.
- These platforms reduce the time taken to develop and deploy, increase flexibility, help users achieve performance and better results, and maintain control over the cost.

Serverless vs PaaS

- Both serverless and PaaS provide the same facilities, as they both are backend architectures that hide the backend from the developers.
- They only differ in scalability, timing, start-up time and tools, and deployment process.
- Differences are:
  - The pricing of serverless is exact as it charges developers for the time the application utilizes. On the other hand, PaaS pricing is not as precise as serverless, as PaaS vendors charge a monthly fee for the services offered.
  - PaaS provides more control over the deployment environment, while on the other hand, serverless provides less control over the environment.
  - Serverless applications are active most of the time. The built-in PaaS applications can be up and run quickly, but they are not as lightweight as serverless. Serverless provides agility to its built-in applications makes it more suitable for web applications.
- It is not that serverless services are more affordable.
- It depends on the type of application we are developing and the facilities and services we require.
- We have to choose between PaaS and serverless according to the project requirements.

Common PaaS scenarios

Organisations typically use PaaS for these scenarios:

Development framework

- PaaS provides a framework that developers can build upon to develop or customise cloud-based applications.
- Similar to the way you create an Excel macro, PaaS lets developers create applications using built-in software components.
- Cloud features such as scalability, high-availability and multi-tenant capability are included, reducing the amount of coding that developers must do.

Analytics or business intelligence

Tools provided as a service with PaaS allow organisations to analyse and mine their data, finding insights and patterns and predicting outcomes to improve forecasting, product design decisions, investment returns and other business decisions.

Additional services

PaaS providers may offer other services that enhance applications, such as workflow, directory, security and scheduling.

PaaS Feature

- **Programming Models, Languages, and Frameworks**

Programming mod-els made available by IaaS providers define how users can express their applications using higher levels of abstraction and efficiently run them on the cloud platform.

- **Persistence Options**

A persistence layer is essential to allow applications to record their state and recover it in case of crashes, as well as to store user data.

Security, Privacy and Trust

- Security and privacy affect the entire cloud computing stack, since there is a massive use of third-party services and infrastructures that are used to host important data or to perform critical operations.
- In this scenario, the trust toward providers is fundamental to ensure the desired level of privacy for applications hosted in the cloud.
- When data are moved into the Cloud, providers may choose to locate them anywhere on the planet.
- The physical location of data centers determines the set of laws that can be applied to the management of data.

Data Lock-In and Standardization

- The Cloud Computing Interoperability Forum (CCIF) was formed by organizations such as Intel, Sun, and Cisco in order to —enable a global cloud computing ecosystem whereby organizations are able to seamlessly work together for the purposes for wider industry adoption of cloud computing technology.
- The development of the Unified Cloud Interface (UCI) by CCIF aims at creating a standard programmatic point of access to an entire cloud infrastructure.
- In the hardware virtualization sphere, the Open Virtual Format (OVF) aims at facilitating packing and distribution of software to be run on VMs so that virtual appliances can be made portable.

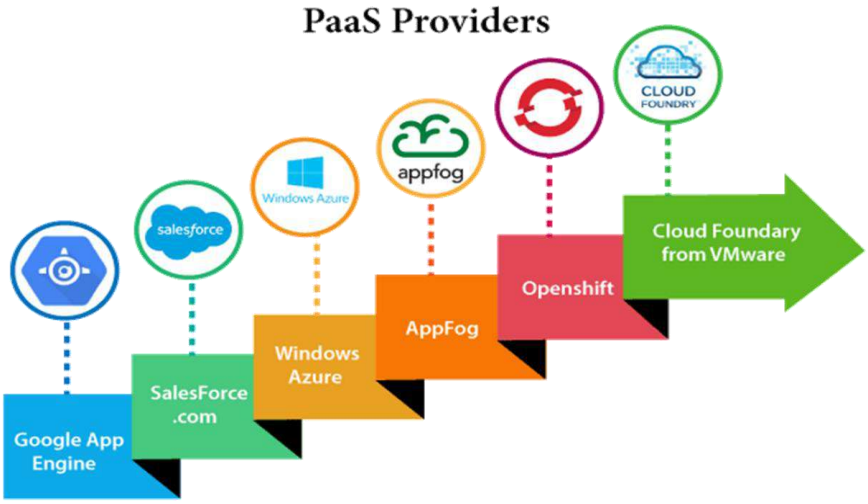
Availability, Fault-Tolerance and Disaster Recovery

- It is expected that users will have certain expectations about the service level to be provided once their applications are moved to the cloud.
- These expectations include availability of the service, its overall performance, and what measures are to be taken when something goes wrong in the system or its components.
- In summary, users seek for a warranty before they can comfortably move their business to the cloud.
- SLAs, which include QoS requirements, must be ideally set up between customers and cloud computing providers to act as warranty.
- An SLA specifies the details of the service to be provided, including availability and performance guarantees.
- Additionally, metrics must be agreed upon by all parties, and penalties for violating the expectations must also be approved.

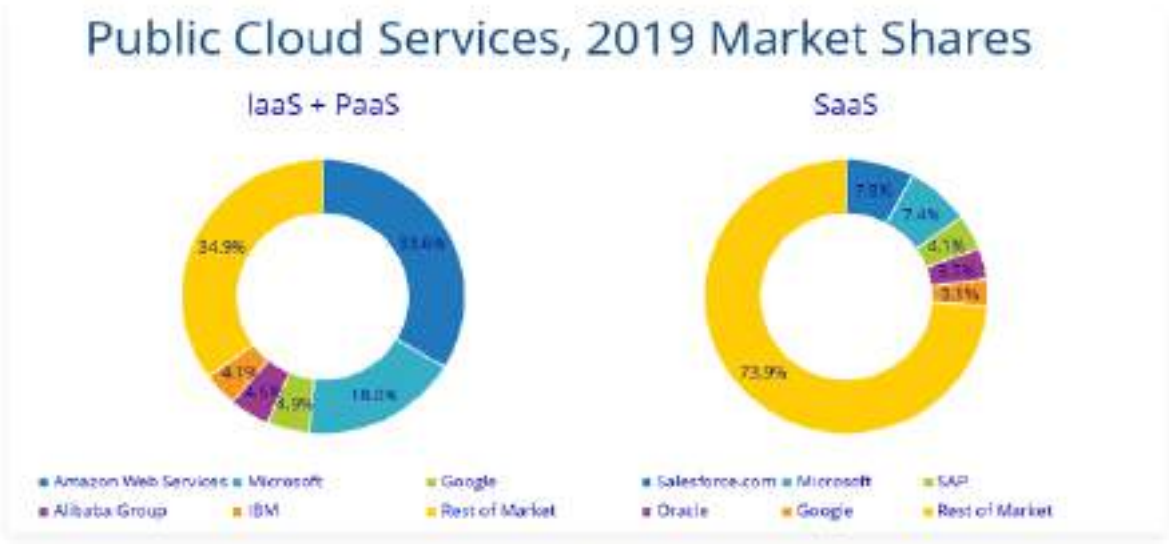
Resource Management and Energy-Efficiency

- The multi-dimensional nature of virtual machines complicates the activity of finding a good mapping of VMs onto available physical hosts while maximizing user utility.
- Dimensions to be considered include:
  - Number of CPUs
  - Amount of memory
  - Size of virtual disks
  - Network bandwidth
- Dynamic VM mapping policies may leverage the ability to suspend, migrate, and resume VMs as an easy way of pre-empting low-priority allocations in favour of higher-priority ones.
- Migration of VMs also brings additional challenges such as detecting when to initiate a migration, which VM to migrate, and where to migrate.
- In addition, policies may take advantage of live migration of virtual machines to relocate data center load without significantly disrupting running services.

Popular PaaS Providers



Leading Vendors and Their Market Share



— Public Cloud Services Market Share – 2019 (Source: IDC)

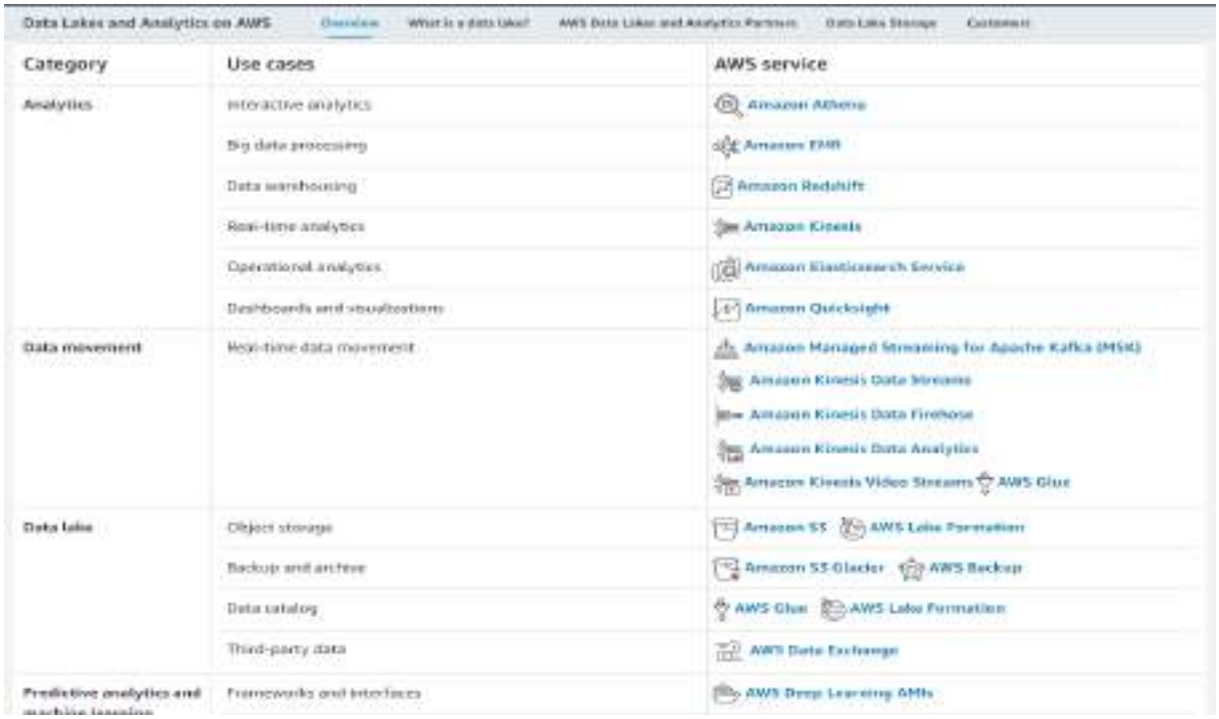
Popular PaaS Providers

Providers	Services
Google App Engine (GAE)	App Identity, URL Fetch, Cloud storage client library, Logservice
Salesforce.com	Faster implementation, Rapid scalability, CRM Services, Sales cloud, Mobile connectivity, Chatter.
Windows Azure	Compute, security, IoT, Data Storage.
AppFog	Justcloud.com, SkyDrive, GoogleDocs
Openshift	RedHat, Microsoft Azure.
Cloud Foundry from VMware	Data, Messaging, and other services.

The 4 Leading PaaS Providers:

AWS

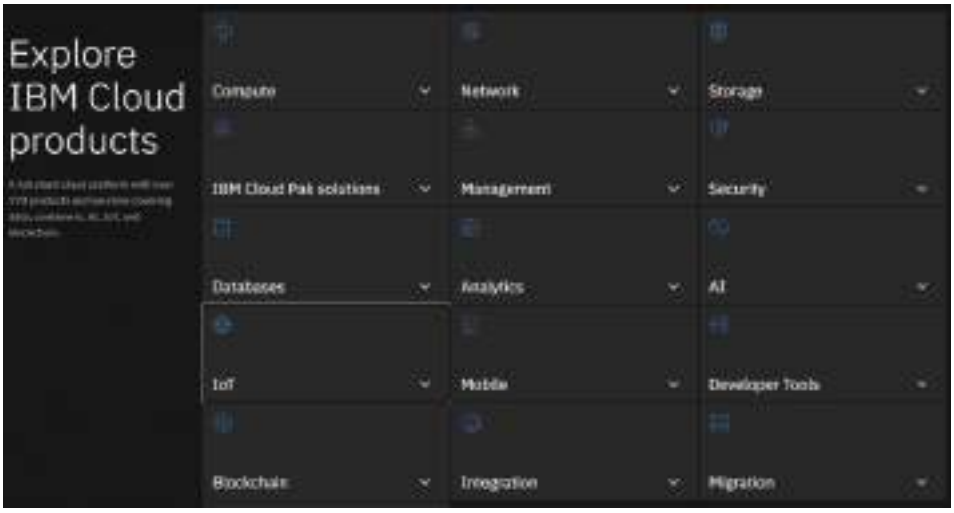
- AWS is the original cloud computing provider, having launched the revolution with its primary EC2 product in 2006.
- The head start cemented them as the clear market leader, and it's still the largest cloud services company in the world. But for PaaS specifically, what does it bring to the table?
- A quick look at Amazon's services overview will tell you everything you need to know.



IBM Cloud

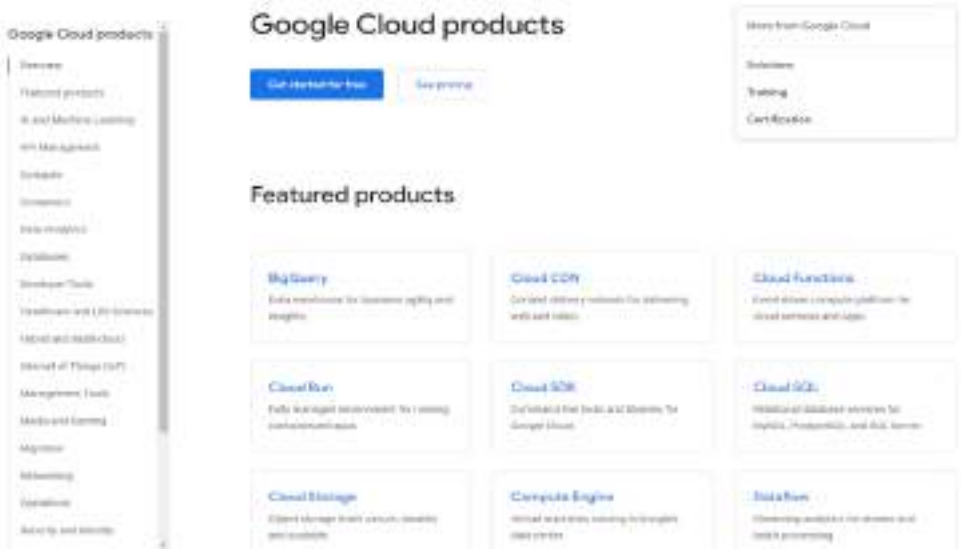
- An early innovator in computing, IBM has put a lot of money and effort into developing its cloud services suite.
- IBM first launched its PaaS services as IBM Bluemix in 2014.

- In 2017, IBM dropped the Bluemix brand and grouped its PaaS, IaaS, and private cloud offerings under the IBM Cloud umbrella.
- With a wide range of enterprise clients, IBM Cloud has quickly grown to become one of the leading PaaS providers since its launch in 2011 and that shows in its range of services:



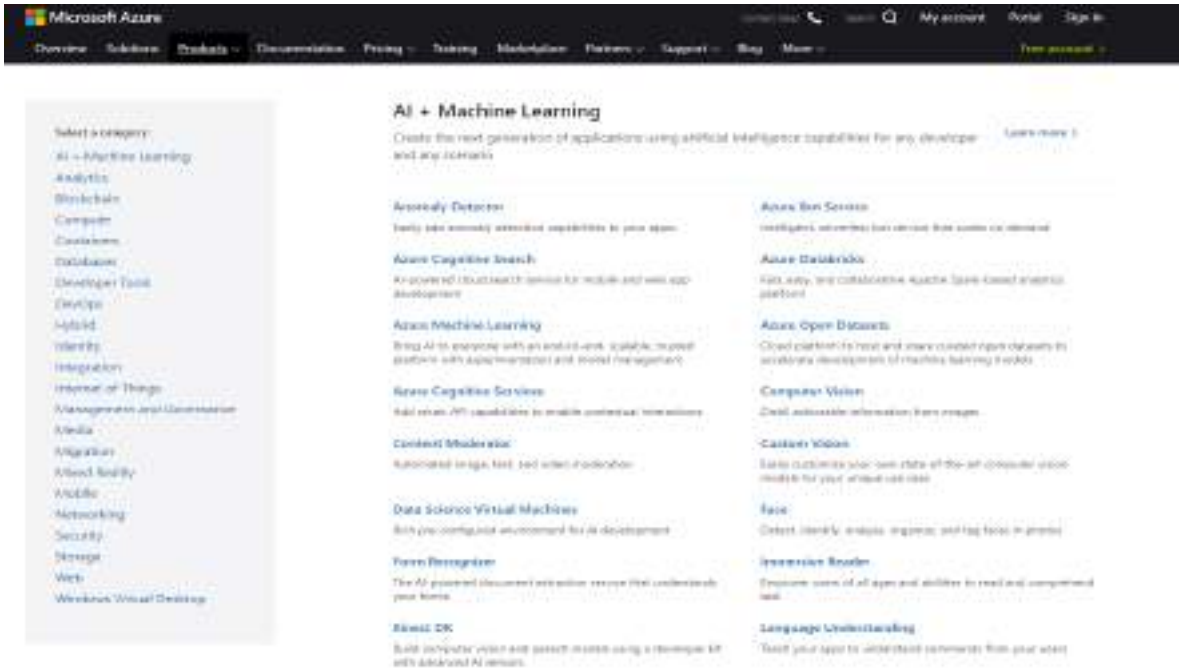
Google Cloud

- Google isn't just a search engine. It's also one of the leading SaaS companies, with Google Docs, Drive, Gmail, and the entire Google Workspace.
- Google also lets you rent the infrastructure and platforms that make it possible to handle billions of visitors every month.
- Launched in 2008, Google Cloud was the second major player to enter the market. Its extensive list of products shows why it's still one of the market leaders.



Microsoft Azure

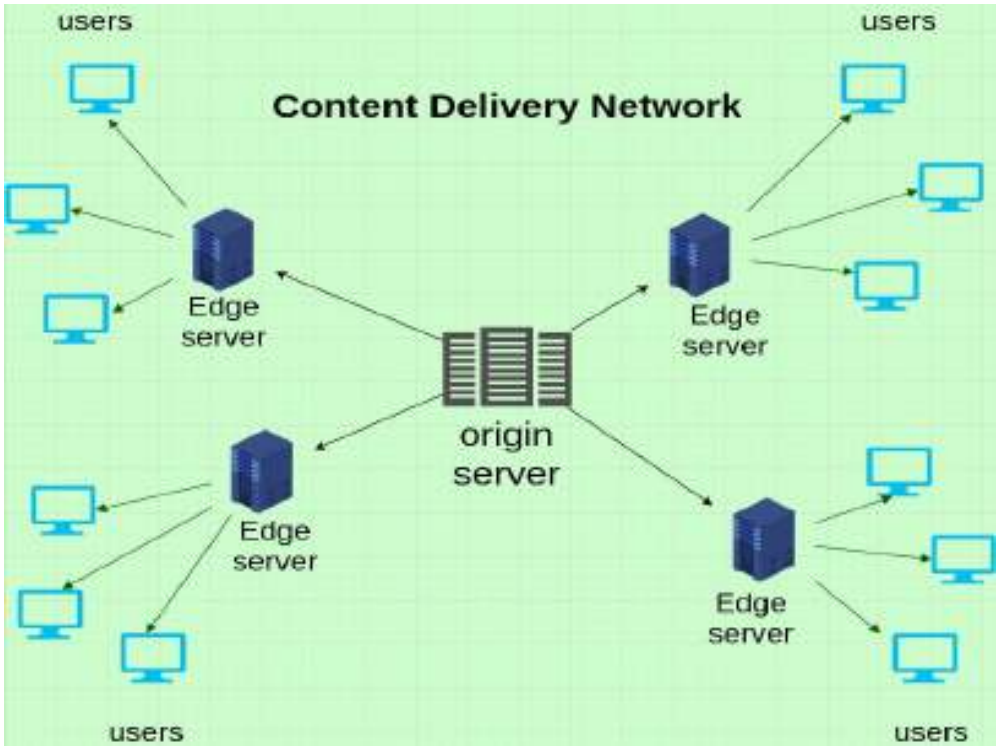
- Microsoft isn't just responsible for the operating systems on most desktop and laptop computers around the world.
- It also has one of the largest public cloud services collections, including Office 365, Microsoft Teams (SaaS), and Azure (IaaS & PaaS).
- The Azure cloud platform includes a range of services from AI and machine learning to analytics, development tools, data processing, and more.



4.2 Managing Cloud Storage

Building Content Delivery Networks Using Clouds

- Numerous “storage cloud” providers (or “Storage as a Service”) have recently emerged that can provide Internet-enabled content storage and delivery capabilities in several continents, offering service-level agreement (SLA)- backed performance and uptime promises for their services.
- Customers are charged only for their utilization of storage and transfer of content (i.e., a utility computing model), which is typically on the order of cents per gigabyte.



Following are the some of the Popular Content Delivery Network Providers

- Microsoft Azure
- Amazon S3 and CloudFront
- Nirvanix SDN
- Rackspace Cloud Files

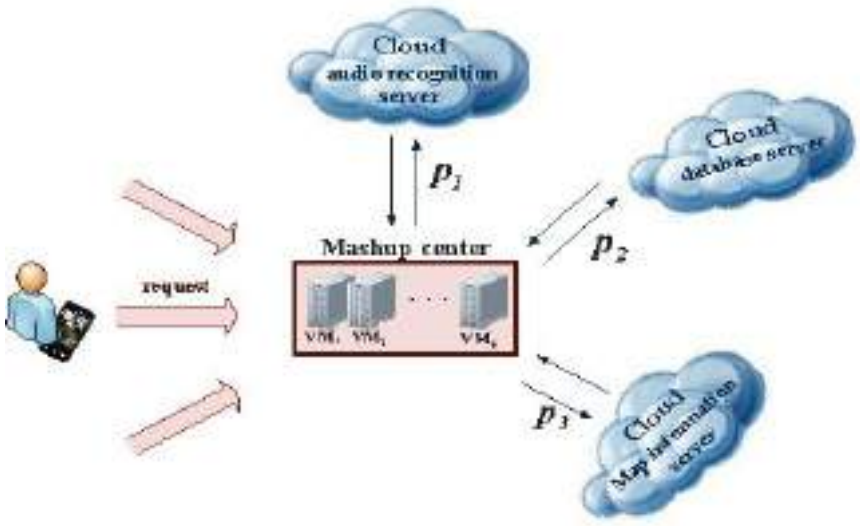
- This represents a large paradigm shift away from typical hosting arrangements that were prevalent in the past, where average customers were locked into hosting contracts (with set monthly/yearly fees and excess data charges) on shared hosting services like DreamHost.

- Larger enterprise customers typically utilized pervasive and high-performing Content Delivery Networks (CDNs), who operate extensive networks of “edge” servers that deliver content across the globe.

4.3 Employing Support Services

Resource Cloud Mashups

Outsourcing computation and/or storage away from the local infrastructure is not a new concept itself: Already the grid and Web service domain presented (and uses) concepts that allow integration of remote resource for seemingly local usage.



Interoperability and Vendor Lock-In

- Since most cloud offerings are proprietary, customers adopting the according services or adapting their respective applications to these environments are implicitly bound to the respective provider.

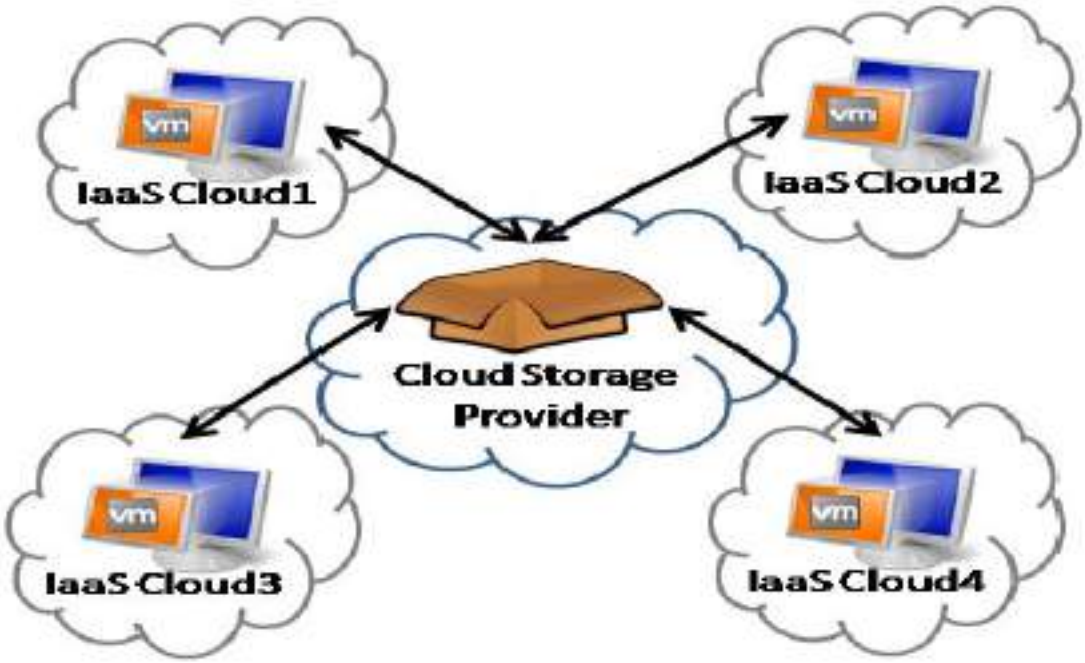
- Movement between providers is restricted by the effort the user wants to vest into porting the capabilities to another environment, implying in most cases reprogramming of the according applications.

- This makes the user dependent not only on the provider’s decisions, but also on his/her failures: As the example of the Google crash on the May 14, 2009 showed, relying too much on a specific provider can lead to serious problems with service consumption.

The Problem of Interoperability

The Web service domain has already shown that interoperability cannot be readily achieved through the definition of common interfaces or specifications:

- The standardization process is too slow to capture the development in academy and industry.
- Specifications (as predecessors to standards) tend to diverge quickly with the standardization process being too slow.
- "Competing" standardization bodies with different opinions prefer different specifications.



A Need for Cloud Mashups

- By integrating multiple cloud infrastructures into a single platform, reliability and scalability is extended by the degree of the added system(s).
- Platform as a Service (PaaS) providers often offer specialized capabilities to their users via a dedicated API, such as Google App Engine providing additional features for handling (Google) documents, and MS Azure is focusing particularly on deployment and provisioning of Web services, and so on.
- Through aggregation of these special features, additional, extended capabilities can be achieved (given a certain degree of interoperability), ranging from extended storage and computation facilities (IaaS) to combined functions, such as analytics and functionalities.
- The Cloud Computing Expert Working Group refers to such integrated cloud systems with aggregated capabilities across the individual infrastructures as Meta-Clouds and Meta-Services, respectively.

With the main focus of cloud-based services being “underneath” the typical Web service level that is, more related to resources and platforms key interoperability issues relate to compatible data structures, related programming models, interoperable operating images, and so on. Thus, to realize a mashup requires at least:

- A compatible API/programming model, respectively an engine that can parse the APIs of the cloud platforms to be combined (PaaS).
- A compatible virtual machine, respectively an image format that all according cloud infrastructures can host (IaaS).
- Interoperable or transferrable data structures that can be interpreted by all engines and read by all virtual machines involved. This comes as a side effect to the compatibility aspects mentioned above.

Why do developers use PaaS?

Faster time to market

- PaaS is used to build applications more quickly than would be possible if developers had to worry about building, configuring, and provisioning their own platforms and backend infrastructure.
- With PaaS, all they need to do is write the code and test the application, and the vendor handles the rest

Why do developers use PaaS?

One environment from start to finish

- PaaS permits developers to build, test, debug, deploy, host, and update their applications all in the same environment.
- This enables developers to be sure a web application will function properly as hosted before they release, and it simplifies the application development lifecycle.

Price

- PaaS is more cost-effective than leveraging IaaS in many cases. Overhead is reduced because PaaS customers don't need to manage and provision virtual machines.
- In addition, some providers have a pay-as-you-go pricing structure, in which the vendor only charges for the computing resources used by the application, usually saving customers money. However, each vendor has a slightly different pricing structure, and some platform providers charge a flat fee per month.

Ease of licensing

- PaaS providers handle all licensing for operating systems, development tools, and everything else included in their platform.

Future of the PaaS market and business model

- PaaS has emerged as a cost-effective and capable cloud platform for developing, running and managing applications -- and the PaaS market is expected to gain popularity and grow through 2027. As an example, IDC predicted that the cloud and PaaS market should see a compound annual growth rate of 28.8 percent in 2021 through 2025.
- Such expectations are based on the need for businesses to accelerate application time to market, reduce complexity, shed local infrastructure, build collaboration -- especially for remote and geographically distributed teams -- and streamline application management tasks.
- PaaS expansion and growth are also being driven by cloud migration and cloud-first or cloud-native application development efforts in concert with other emerging cloud technologies, such as IoT.
- The role of iPaaS is also expected to make considerable gains by 2027 as businesses of all sizes seek to modernize, connect and share data between disparate software applications and deliver unified tools across the business and their customer base.

4.4 Monitoring Cloud-Based Services

What is Cloud Monitoring?

- Cloud monitoring is a method of reviewing, observing, and managing the operational workflow in a cloud-based IT infrastructure.
- Manual or automated management techniques confirm the availability and performance of websites, servers, applications, and other cloud infrastructure.
- This continuous evaluation of resource levels, server response times, and speed predicts possible vulnerability to future issues before they arise.

Types of Cloud Monitoring

The main types of cloud monitoring are:

- Database Monitoring
- Website Monitoring
- Virtual Network Monitoring
- Cloud Storage Monitoring
- Virtual Machine Monitoring

Database Monitoring

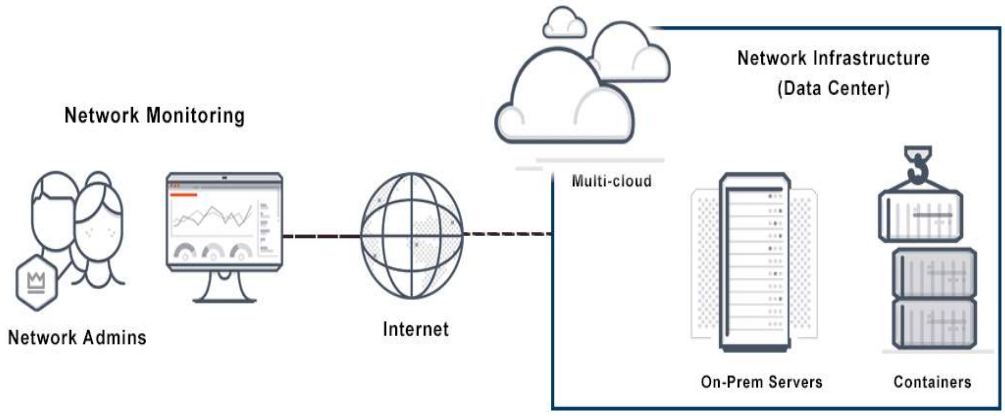
- Because most cloud applications rely on databases, this technique reviews processes, queries, availability, and consumption of cloud database resources.
- This technique can also track queries and data integrity, monitoring connections to show real-time usage data.
- For security purposes, access requests can be tracked as well. For example, an uptime detector can alert if there's database instability and can help improve resolution response time from the precise moment that a database goes down.

Website Monitoring

A website is a set of files that is stored locally, which, in turn, sends those files to other computers over a network. This monitoring technique tracks processes, traffic, availability, and resource utilization of cloud-hosted sites.

Virtual Network Monitoring

- This monitoring type creates software versions of network technology such as firewalls, routers, and load balancers. Because they're designed with software, these integrated tools can give you a wealth of data about their operation.
- If one virtual router is endlessly overcome with traffic, for example, the network adjusts to compensate. Therefore, instead of swapping hardware, virtualization infrastructure quickly adjusts to optimize the flow of data.



Cloud Storage Monitoring

- This technique tracks multiple analytics simultaneously, monitoring storage resources and processes that are provisioned to virtual machines, services, databases, and applications.
- This technique is often used to host infrastructure-as-a-service (IaaS) and software-as-a-service (SaaS) solutions.
- For these applications, you can configure monitoring to track performance metrics, processes, users, databases, and available storage.
- It provides data to help you focus on useful features or to fix bugs that disrupt functionality.

Virtual Machine Monitoring

- This technique is a simulation of a computer within a computer; that is, virtualization infrastructure and virtual machines.
- It's usually scaled out in IaaS as a virtual server that hosts several virtual desktops.
- A monitoring application can track the users, traffic, and status of each machine. You get the benefits of traditional IT infrastructure monitoring with the added benefit of cloud monitoring solutions.

Benefits of Cloud Monitoring

Monitoring is a skill, not a full-time job

- In today's world of cloud-based architectures that are implemented through DevOps projects, developers, site reliability engineers (SREs), and operations staff must collectively define an effective cloud monitoring strategy.

- Such a strategy should focus on identifying when service-level objectives (SLOs) are not being met, likely negatively affecting the user experience. So, then what are the benefits of leveraging cloud monitoring tools? With cloud monitoring:
  - Scaling for increased activity is seamless and works in organizations of any size
  - Dedicated tools (and hardware) are maintained by the host
  - Tools are used across several types of devices, including desktop computers, tablets, and phones, so your organization can monitor apps from any location
  - Installation is simple because infrastructure and configurations are already in place
  - Your system doesn't suffer interruptions when local problems emerge, because resources aren't part of your organization's servers and workstations
  - Subscription-based solutions can keep your costs low

Monitoring in Public, Private and Hybrid Clouds

- A private cloud gives you extensive control and visibility. Because systems and the software stack are fully accessible, cloud monitoring is relaxed when it's operated in a private cloud.
- Monitoring in public or hybrid clouds, however, can be tough.
- Let's review the focal points:
  - Because the data exists between private and public clouds, a hybrid cloud environment presents curious challenges. Limited security and compliance create problems for data access. Your administrator can solve these issues by deciding which data to store in various clouds and which data to asynchronously update.
  - A private cloud gives you more control, but to promote optimal performance, it's still wise to monitor workloads. Without a clear picture of workload and network performance, it's nearly impossible to justify configuration or architectural changes or to quantify quality-of-service implementations.

Cloud Monitoring Best Practices

- Observe your cloud service usage and fees. Increased costs can be triggered when scaling kicks in to meet demand. Strong monitoring solutions should track how much activity is on the cloud and its associated cost.
- Identify metrics and events that affect your bottom line. Not everything that can be measured needs to be reported.
- Use a single platform to report all data. You need solutions that can report data from different sources to a single platform. This consolidated information enables you to calculate uniform metrics and results in a complete performance view.
- Trigger rules with data. If activity surpasses or drops below certain levels, the right solution should be to add or subtract servers to maintain efficiency and performance.
- Separate your centralized data. Your organization must store your monitoring data separately from your proprietary apps, but the information should still be centralized for easy access.
- Monitor the user experience. To get the full picture of performance, review metrics such as response times and frequency of use.
- Try failure. Test tools to see what happens when an outage or a data breach occurs. This evaluation can create new standards for the alert system.

- Cloud Testing refers to the verification of software quality on the cloud. Essentially, this translates to running manual and automation testing on a cloud computing environment with the requisite infrastructure.
- Cloud Testing (also termed as Cloud based testing) takes the entire testing process online, sparing QAs the hassle of limited device/browsers/OS coverage, geographical limitations, extensive setup and maintenance processes, and the like.
- With Cloud Testing, testing becomes faster, easier, and infinitely more manageable.

Why is Cloud Testing needed?

- Automated testing is almost always more complicated to set up and execute than manual testing.
- With cloud automated testing, the process is simplified for the following reasons:
  - Cloud testing platforms are set up to facilitate tests for multiple users and teams on multiple devices simultaneously. That means QA teams won't have to share test environments with other teams/projects.
  - Even if some tests have to be queued, a cloud-based testing environment worth the cost is schematized to expedite tests without compromising accuracy.
  - Efficient cloud testing platforms also possess features to accelerate and enhance collaboration between teams or members of the same teams. This helps monitor all team members' progress and keeps everyone on the same page about project direction and achievements.
- This is also available for cloud-based manual testing platforms, such as BrowserStack's Live for Teams, but this is much rarer than the former.

What are the Benefits of Cloud Testing?

- Generally, in-house labs for most organizations do not possess the infrastructure necessary to replicate real-world devices and software usage.
- Due to rapidly changing user expectations and standards, organizations will have to continually update their labs, demanding constant money and human resources.
- Cloud testing tools solve this by providing a real-made testing environment that mirrors the production environment quite closely.
- Testers simply have to sign up, select the real devices they want to start tests on, and start flagging bugs.
- As explained previously, setting up on-premise device labs incurs high costs.

- Not only does the organization have to keep purchasing new devices hitting the market, but they also have to upgrade frameworks, testing software, renew licenses, pay maintenance costs and ensure device security.
- It is far cheaper to leave all of that to a third-party platform and only pay for access to devices and session time.

- Cloud testing tools offer optimized test environments with all requisite software-hardware configuration in place.
- With platforms like BrowserStack, testers can be assured that every device on the real device cloud is pristine. Every device offered is calibrated to factory settings. Once a test is complete, every last bit of data is destroyed.
- With automated testing and parallel testing, testing in the cloud allows QAs to accelerate test execution and results significantly. Faster results can also be achieved by virtue of features that allow for improved collaboration and project management.
- Leading cloud testing platforms like BrowserStack offer 99% uptime. That means testers can access real desktop and mobile devices for testing anytime, from anywhere.
- Cloud-based testing on platforms like BrowserStack offers integrations with numerous tools that assist with implementing DevOps and CI/CD workflows. This allows for a more streamlined, result-oriented software development pipeline.

Best Practices of Cloud Testing

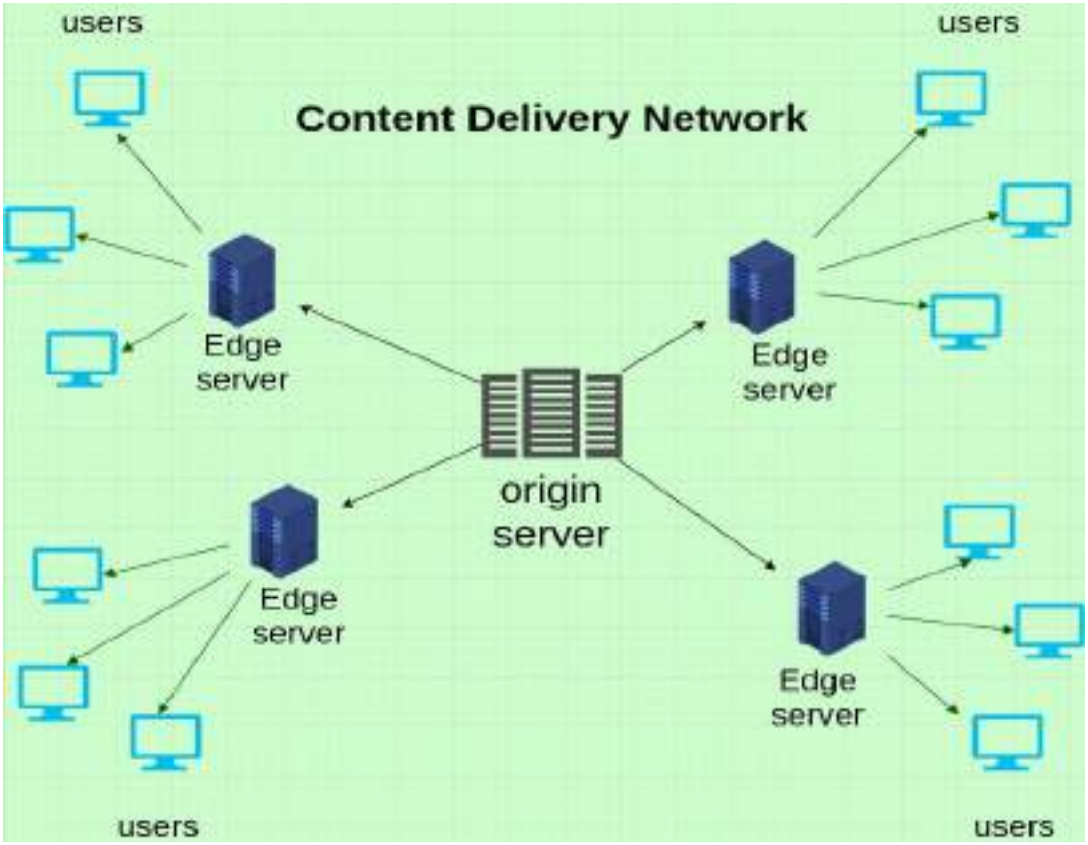
- Look for a cloud testing platform that offers the devices and browsers the target audience is likely to use while using the software in question. For instance, BrowserStack offers 2000+ real browsers and devices. Chances are that users of this cloud will have access to the devices their potential customers would prefer.
- Before choosing a platform, put in the research. The ideal cloud should offer high-security levels, reasonably consistent tech support, and ensure that wait times for queued tests are not too long. The point of moving tests to the cloud is to speed them up without compromising on quality or security.
- A cloud testing platform worth the cost should cater not just to individual testers but to WA managers as well. Especially in these times of remote testing, the cloud should provide team-wide testing on a single plan, as well as features designed to help QA managers keep track of project progress and individual activity of each team member.

Section 3: Exercises

Exercise 1: Write down the services provided by PaaS service providers in below Table.

Providers	Services
Google App Engine (GAE)	
Salesforce.com	
Windows Azure	
AppFog	
Openshift	
Cloud Foundry from VMware	

Exercise 2: Write Down all the Parts of Content Delivery Network in below Diagram.



Exercise 3: Participate in a group discussion on following topics:

a) Concept of Platform as a Service (PaaS)

b) Pros and Cons of PaaS
c) PaaS Architecture
d) PaaS and Its Services
e) PaaS Monitoring
f) Benefits of Cloud Monitoring

Section 4: Assessment Questionnaire

1. What is PaaS?
2. Who is the End Customer of PaaS?
3. What are ways to Deliver PaaS?
4. What Services Does PaaS Includes?
5. What is Public PaaS?
6. What is Hybrid PaaS?
7. What are the Common PaaS Scenarios?
8. What are the Features of PaaS?
9. What are Popular PaaS Providers?
10. List few of the Content Delivery Network Using Cloud.

----------End of Module----------

MODULE 5

Deploying Infrastructure as a Service (IaaS)

Section 1: Learning Outcomes

After completing this module, you will be able to:

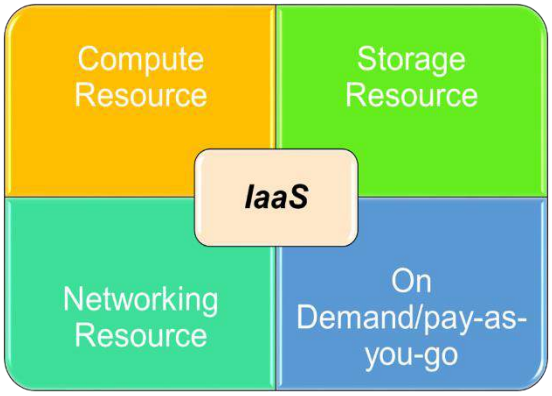
- Explain the Concept of Infrastructure as a Service (IaaS)
- Differentiate between various Enabling Technologies
- Define Scalable Server Clusters
- Achieve Transparency with Platform Virtualization
- Describe the Elastic Storage Devices Information
- Show How to Access IaaS
- Provision Servers on Demand
- Enlist Tools and Support for Management and Monitoring

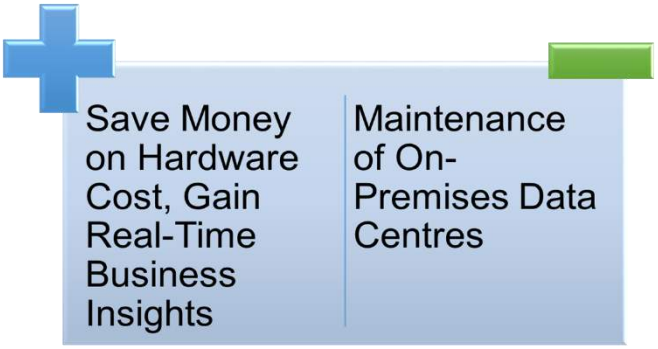
Section 2: Relevant Knowledge

5.1 Deploying Infrastructure as a Service (IaaS)

Infrastructure as a Service

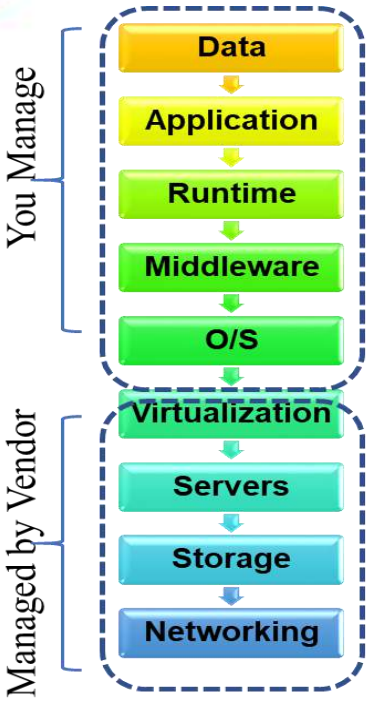
- Infrastructure as a service (IaaS) is a type of cloud computing service that offers essential compute, storage and networking resources on demand, on a pay-as-you-go basis.
- IaaS is one of the four types of cloud services, along with software as a service (SaaS), platform as a service (PaaS) and serverless.
- Migrating your organisation's infrastructure to an IaaS solution helps you reduce maintenance of on-premises data centres, save money on hardware costs and gain real-time business insights.
- IaaS solutions give you the flexibility to scale your IT resources up and down with demand. They also help you quickly provision new applications and increase the reliability of your underlying infrastructure.





- IaaS lets you bypass the cost and complexity of buying and managing physical servers and datacentre infrastructure.
- Each resource is offered as a separate service component and you only pay for a particular resource for as long as you need it.
- A cloud computing service provider like Azure manages the infrastructure, while you purchase, install, configure and manage your own software including operating systems, middleware and applications.

A *cloud infrastructure* enables on-demand provisioning of servers running several choices of operating systems and a customized software stack. Infrastructure services are considered to be the bottom layer of cloud computing systems.



IaaS

Common IaaS use cases include:

Test and Development

Teams can rapidly create development and test environments to bring new applications to market faster. IaaS enables teams to create test and development environments automatically, as part of their development pipeline.

Web Apps

- IaaS provides all the infrastructure needed to run large scale web applications, including storage, web servers, and networking.
- Organizations can quickly deploy web applications using IaaS services, and easily scale their infrastructure when application requirements increase or decrease.

Storage, Backup and Recovery

- Organizations can avoid the high upfront cost of storage and the complexity of storage management.
- Leveraging cloud storage services eliminates the need for trained personnel to manage data and comply with legal and regulatory requirements, and helps organizations respond to storage requirements on-demand.
- It also simplifies the planning and management of backup and recovery systems.

High-Performance Computing

- High-performance computing (HPC) can help solve complex and complex problems with millions of variables and calculations, by running them on supercomputers or large clusters of computers.

- The major IaaS providers offer services that place HPC within the reach of ordinary businesses, allowing them to use HPC on demand instead of making a huge investment in HPC infrastructure.

Big Data Analytics

- Big data processing and analysis is critical in today's economy, and requires complex infrastructure including large-scale storage systems, distributed processing engines, and high-speed databases.
- IaaS providers provide all this infrastructure as a managed service, and most of them also offer PaaS services that can perform the actual analytics, including machine learning and AI.

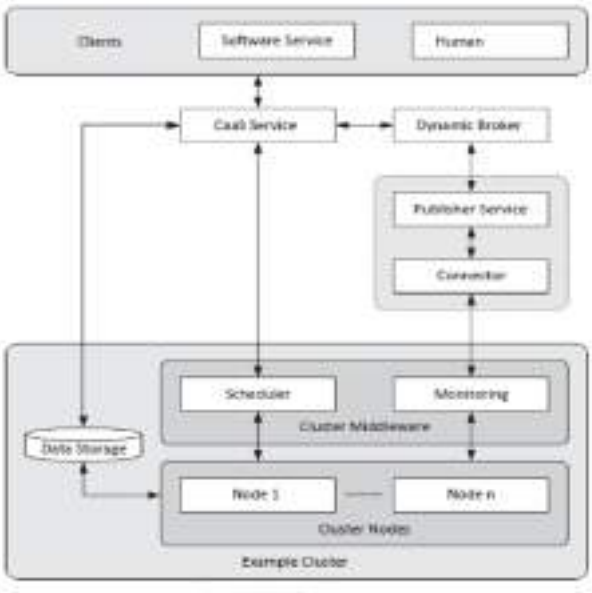
5.2 Scalable Server Clusters

Cluster As a Service: The Logical Design

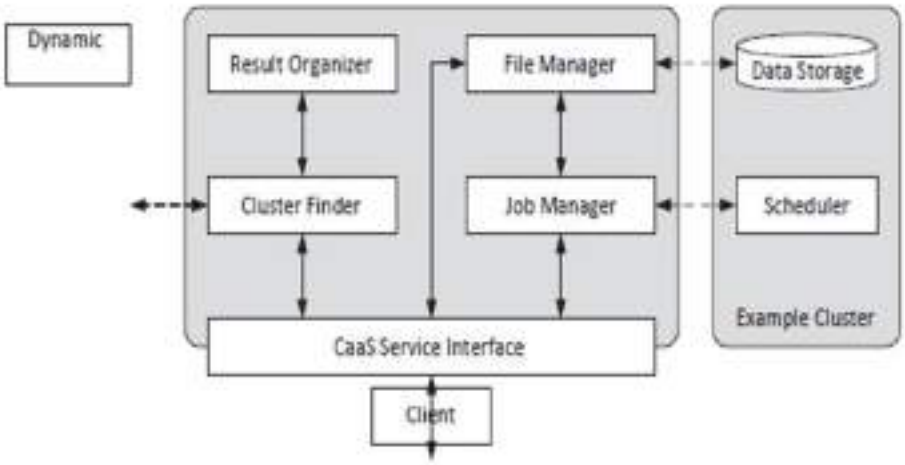
- Simplification of the use of clusters could only be achieved through higher layer abstraction that is proposed here to be implemented using the service-based Cluster as a Service (CaaS) Technology.
- The purpose of the CaaS Technology is to ease the publication, discovery, selection, and use of existing computational clusters.

CaaS Overview

- The exposure of a cluster via a Web service is intricate and comprises several services running on top of a physical cluster. Figure shows the complete CaaS technology.
- A typical cluster is comprised of three elements:
  - Nodes
  - Data storage
  - Middleware
- The middleware virtualizes the cluster into a single system image; thus, resources such as the CPU can be used without knowing the organization of the cluster.
- As time progresses, the amount of free memory, disk space, and CPU usage of each cluster node changes. Information about how quickly the scheduler can take a job and start it on the cluster also is vital in choosing a cluster.



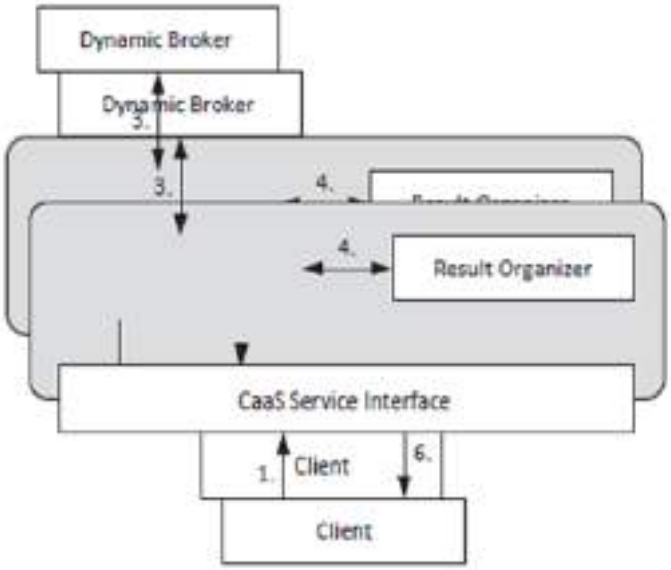
Complete CaaS system



CaaS Service design

Cluster Discovery

- Before a client uses a cluster, a cluster must be discovered and selected first. Figure 3.5 shows the workflow on finding a required cluster.
- To start, clients submit cluster requirements in the form of attribute values to the CaaS Service Interface:
  (1) The requirements range from the number of nodes in the cluster to the installed software (both operating systems and software APIs). The CaaS Service Interface invokes the Cluster Finder module.
  (2) that communicates with the Dynamic Broker
  (3) returns service matches (if any). To address the detailed results from the Broker, the Cluster Finder module invokes the Results Organizer module
  (4) that takes the Broker results and returns an organized version that is returned to the client

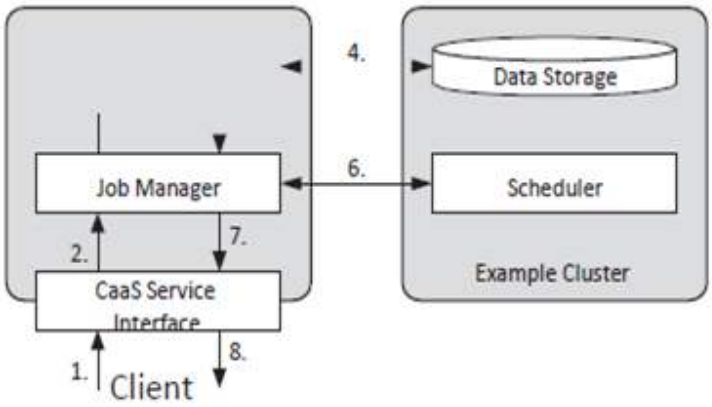


Cluster discovery

Job Submission

- After selecting a required cluster, all executables and data files have to be transferred to the cluster and the job submitted to the scheduler for execution.

- As clusters vary significantly in the software middleware used to create them, it can be difficult to place jobs on the cluster.
- To do so requires knowing how jobs are stored and how they are queued for execution on the cluster.



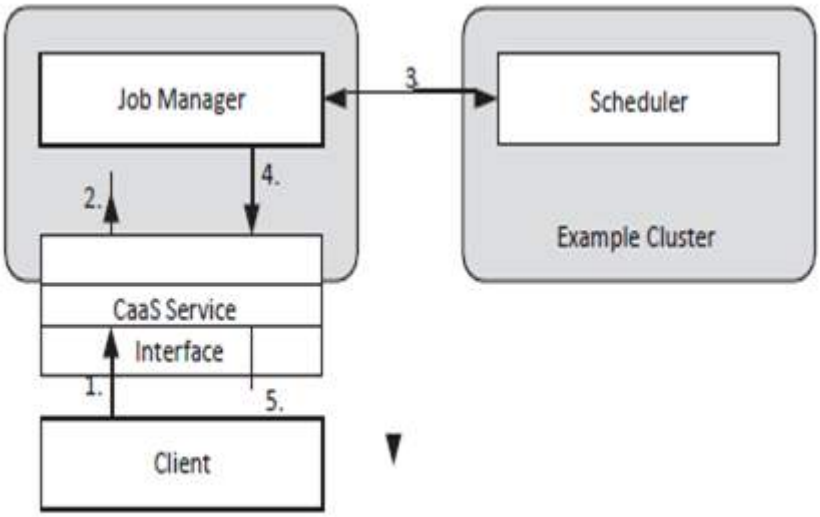
Job submission

Job Monitoring

- During execution, clients should be able to view the execution progress of their jobs. Even though the cluster is not the owned by the client, the job is.
- It is the right of the client to see how the job is progressing and (if the client decides) terminate the job and remove it from the cluster.

Result Collection

- The final role of the CaaS Service is addressing jobs that have terminated or completed their execution successfully.
- In Both cases, error or data files need to be transferred to the client. Figure presents the workflow and CaaS Service modules used to retrieve error or result files from the cluster.



Job result collection

5.3 Achieving Transparency with Platform Virtualization

What Is Virtualization?

- The basic concept of virtualization is that a piece of software will function as a physical object, that is, it will “look” and “behave” like hardware. Thus, it will perform all of the functions that a piece of hardware performs without the hardware in place. As such, the software emulates a desktop PC or other equipment on a server.
- This, in fact, is what cloud-based IT service provides – a place where business functions can occur and be stored without the need for in-house hardware.

How Virtualization is different from Cloud Computing?

- Virtualization software allows multiple operating systems and applications to run on the same server at the same time, and, as a result, lowers costs and increases efficiency of a company’s existing hardware.
- It’s a fundamental technology that powers Everything-as-a-Service model of computing.
- Virtualization decouples software and physical machines to construct numerous virtual machines running on the same server.
- While the principle behind the cloud computing is the same, it is more complex and includes the creation of multiple virtual infrastructures.

The Main Types of Virtualizations

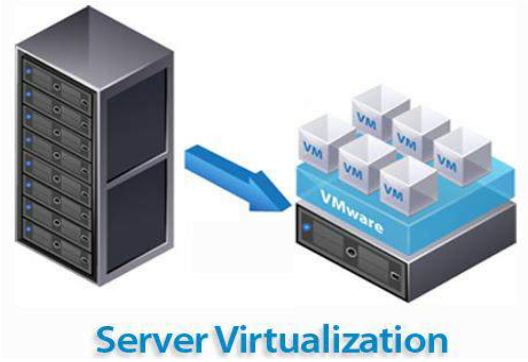
There are several types of virtualizations categorized according to the elements they are used on.

- Server Virtualization
- Storage Virtualization
- Network Virtualization

The Main Types of Virtualizations

1. Server Virtualization

- Server space is conserved by consolidating multiple machines into a single server that then runs several virtual environments.
- It is a method by which businesses can run the same applications on various servers, so that there is a "fail-safe" position.
- A fail-safe system design allows for automatic failure mitigations based on the anticipated scenarios. Because each server is independent, running software on one will not affect the other.

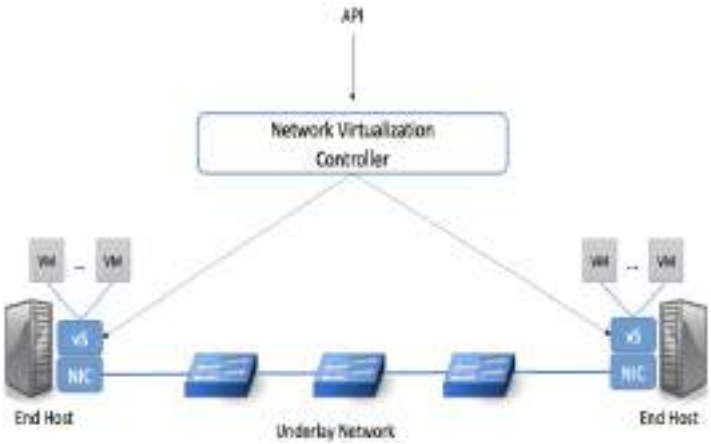


2. Storage Virtualization

- Disk storage used to be a simple matter. If a business needed more, it simply purchased a larger disk drive.
- While storages continue to grow to handle all the data, they become much harder to manage.
- According to Statista, the global estimated enterprise data volume can be anything from 1 petabyte (PB) to 2.02 petabytes in 2022.

3. Network Virtualization

- This type of virtualization allows management and monitoring of an entire network as a single entity.
- Primarily, it is designed to automate administrative tasks, disguising the complexity of the network.
- Each server (and service) is considered a part of one pool of resources to be used without worry about its physical components.



Private Cloud Virtualization: Advantages and Disadvantages

Advantages

- Businesses that operate in a regulatory environment, such as financial services or health, have critical data and protection responsibilities. Building virtualization infrastructures themselves rather than sharing them with others in a public cloud, can address potential compliance issues.
- Likewise, companies that have data which they wish to remain confidential, e.g., research, can feel a bit better about in-house virtualization, in which they can protect that data. No other company has access to that infrastructure.
- Virtualization in the cloud has greater reliability. When public clouds are considered, potential users must conduct solid research to determine if the server they select can provide premiere performance for the types of applications and services they need. In building a private cloud, predictable and reliable service for businesses is generally most assured.
- Cost and Flexibility. There are always trade-offs when implementing new hardware and software. In the case of a private cloud, the initial expense of installing servers and storage can be high. On the other hand, great flexibility can be built in so that workloads can easily be shifted during peak usage spikes and when new applications are deployed. There is no need to make a request of a cloud service provider, before changes can be accomplished.

Disadvantages

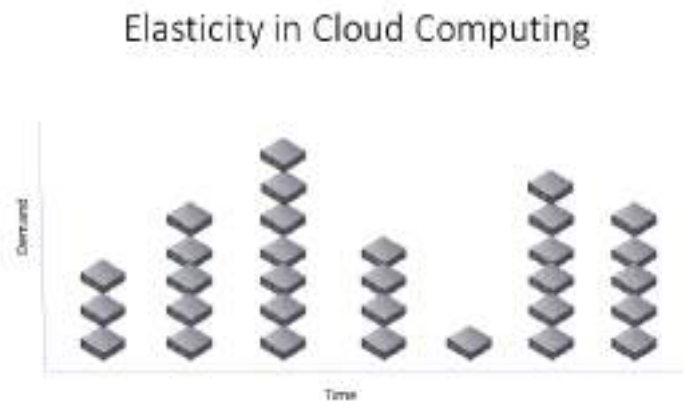
- No software or hardware solution is perfect, and that is certainly the case with private cloud virtualization. Before building and deploying one, organizations have to consider its disadvantages:
  - Integration with other in-house systems can be an issue.
  - Managing and supporting virtualization will often require dedicated IT staff, and that may bring costs up, if there is already not a good-sized department. This is the primary reason why smaller businesses opt for external cloud services.
  - Scaling and security will require specific expertise.

5.4 Elastic Storage Devices

What is Elasticity?

Elasticity

- One of the main advantages of cloud computing is the capability to provide, or release, resources on-demand.
- These elasticity capabilities should be enacted automatically by cloud computing providers to meet demand variations, just as electrical companies are able (under normal operational circumstances) to automatically deal with variances in electricity consumption levels.
- Clearly the behavior and limits of automatic growth and shrinking should be driven by contracts and rules agreed on between cloud computing providers and consumers.



Why is Cloud Elasticity Important?

- Without Cloud Elasticity, organizations would have to pay for capacity that remained unused for most of the time, as well as manage and maintain that capacity with OS upgrades, patches, and component failures.
- It is Cloud Elasticity that in many ways defines cloud computing and differentiates it from other computing models such as client-server, grid computing, or legacy infrastructure.
- Cloud Elasticity helps businesses avoid either over-provisioning (deploying and allocating more IT resources than needed to serve current demands) or under-provisioning (not allocating enough IT resources to meet existing or imminent demand).
- Organizations that over-provision spend more than is necessary to meet their needs, wasting valuable capital which could be applied elsewhere. Even if an organization is already utilizing public cloud, without elasticity, thousands of dollars could be wasted on unused VMs every year.
- Under-provisioning can lead to the inability to serve existing demand, which could lead to unacceptable latency, user dissatisfaction, and ultimately loss of business as customers abandon long online and take their business to more responsive organizations. In this way the lack of Cloud Elasticity can lead to lost business and severe bottom-line impacts.

How does Cloud Elasticity Work?

- Cloud Elasticity enables organizations to rapidly scale capacity up or down, either automatically or manually.
- Cloud Elasticity can refer to ‘cloudbursting’ from on-premises infrastructure into the public cloud for example to meet a sudden or seasonal demand.
- Cloud Elasticity can also refer to the ability to grow or shrink the resources used by a cloud-based application.
- Cloud Elasticity can be triggered and executed automatically based on workload trends, or can be manually instantiated, often in minutes. Before organizations had the ability to leverage Cloud Elasticity, they would have to either have additional stand-by capacity already on hand or would need to order, configure, and install additional capacity, a process which could take weeks or months.
- If and when demand eases, capacity can be removed in minutes.
- In this manner organizations pay only for the number of resources in use at any given time, without the need to acquire or retire on-premises infrastructure to meet elastic demand.

Use Cases of Cloud Elasticity

Typical use cases for Cloud Elasticity include:

- Retail or e-tail holiday seasonal demand, in which demand increases dramatically from Black Friday shopping specials until the end of the holiday season in early January
- School district registration which spikes in demand during the spring and wanes after the school term begins
- Businesses that see a sudden spike in demand due to a popular product introduction or social media boost, such as a streaming service like Netflix adding VMs and storage to meet demand for a new release or positive review.
- Disaster Recovery and Business Continuity (DR/BC). Organizations can leverage public cloud capabilities to provide off-site snapshots or backups of critical data and applications, and spin up VMs in the cloud if on-premises infrastructure suffers an outage or loss.

- Scale virtual desktop infrastructure in the cloud for temporary workers or contractors or for applications such as remote learning
- Scale infrastructure into the cloud for test and development activities and tear it down once test/dev work is complete.
- Unplanned projects with short timelines
- Temporary projects like data analytics, batch processing, media rendering, etc.

What are the Benefits of Cloud Elasticity?

The benefits of cloud elasticity include:

Agility

By eliminating the need to purchase, configure, and install new infrastructure when demand changes, Cloud Elasticity prevents the need to plan for such unexpected demand spikes, and enables organizations to meet any unexpected demand, whether due to seasonal spike, mention on Reddit, or selection by Oprah's book club.

Pay-as-needed pricing

- Rather than paying for infrastructure whether or not is being used, Cloud Elasticity enables organizations to pay only for the resources that are in use at any given point tin time, closely tracking IT expenditures to the actual demand in real-time.
- In this way, although spending may fluctuate, organizations can 'right-size' their infrastructure as elasticity automatically allocates or deallocates resources on the basis of real-time demand.
- Amazon has stated that organizations that adopt its instance scheduler with their EC2 cloud service can achieve savings of over 60 percent versus organizations that do not.

High Availability

- Cloud elasticity facilitates both high availability and fault tolerance, since VMs or containers can be replicated if they appear to be failing, helping to ensure that business services are uninterrupted and that users do not experience downtime.
- This helps ensure that users perceive a consistent and predictable experience, even as resources are provisioned or deprovisioned automatically and without impact on operations.

Efficiency

As with most automations, the ability to autonomously adjust cloud resources as needed enables IT staff to shift their focus away from provisioning and onto projects that are more beneficial to the organization.

Speed/Time-to-market

Organizations have access to capacity in minutes instead of the weeks or months it may take through a traditional procurement process.

Secure Distributed Data Storage in Cloud Computing

Cloud Storage: From LANs to WANs

- Cloud computing will be a revolutionary change in computing services.
- Users will be allowed to purchase CPU cycles, memory utilities, and information storage services conveniently just like how we pay our monthly water and electricity bills.

Existing Commercial Cloud Services

- In normal network-based applications, user authentication, data confidentiality, and data integrity can be solved through IPSec proxy using encryption and digital signature.
- The key exchanging issues can be solved by SSL proxy. These methods have been applied to today's cloud computing to secure the data on the cloud and also secure the communication of data to and from the cloud.
- The service providers claim that their services are secure.

5.5 Enabling Technologies

AWS IaaS Services

Amazon S3

- Amazon Simple Storage Service (S3) is the first and most popular Amazon service, which provides object storage at unlimited scale.
- S3 is easy to access via the Internet and programmatically via API, and is integrated into a wide range of applications.
- It provides 11 9's of durability (99.999999999%), and offers several storage tiers, allowing users to move data that is used less frequently into a low-cost archive tier within S3.

AWS EC2

- Amazon Elastic Compute Cloud (Amazon EC2) offers scalable computing resources.
- It lets you run as many virtual servers as you want, configure your network and security, and manage storage.
- You can increase or decrease resources on-demand according to changing business requirements, and set up auto scaling to scale resources up and down according to actual workloads.

AWS EBS

- Amazon Elastic Block Store (Amazon EBS) is a block-level storage service for use with Amazon EC2 instances.
- When mounted on an Amazon EC2 instance, you can use Amazon EBS volumes like any other raw block storage device.
- It can be formatted and mirrored for specific file systems, host operating systems, and applications.

AWS Lambda

- AWS Lambda is a serverless, on-demand IT service that provides developers with a fully managed, event-driven cloud system that executes code.
- AWS Lambda uses Lambda functions anonymous functions that are not associated with identifiers enabling users to package any code into a function and run it, independently of other infrastructure

Azure IaaS Services

Linux Virtual Machines in Azure

- Traditionally Azure focused on Windows virtual machines, but now has a robust offering for Linux users as well.
- Azure virtual machines (VMs) are scalable on-demand compute resources provided by Azure. Microsoft Azure supports popular Linux distributions deployed and managed by multiple partners.
- Linux machine images are available in the Azure Marketplace for the following Linux distributions (more distributions are added on an ongoing basis):
  - FreeBSD
  - Red Hat Enterprise
  - CentOS
  - SUSE Linux Enterprise
  - Debian
  - Ubuntu
  - CoreOS
  - RancherOS

Azure Managed Disk

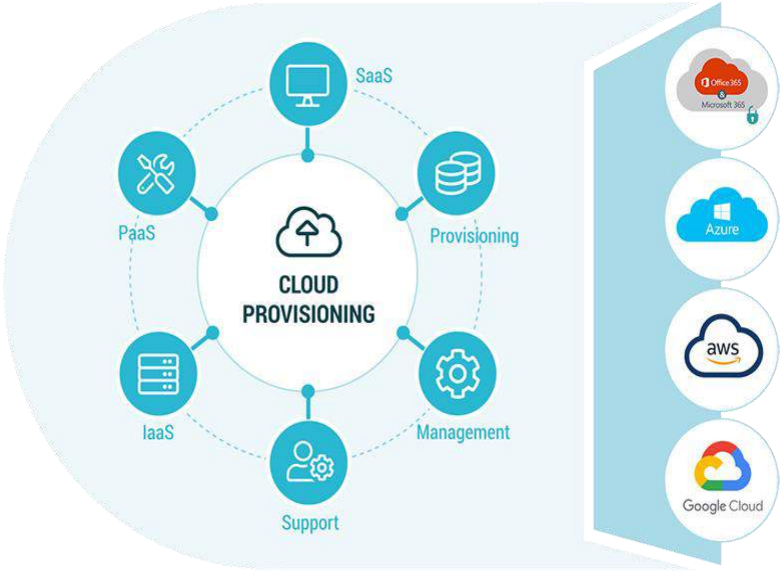
- Azure managed disks are block-level storage volumes managed by Azure and used by Azure virtual machines.
- A managed disk is similar to a physical disk on a local server, but it is virtualized.
- For managed disks, you only need to specify the disk size and disk type, and provision—Azure does the rest. The available hard drive types are:
  - Standard hard disks (HDD)
  - Standard SSD
  - Premium SSDs
  - Ultra-disks—optimized for sub-millisecond latency

5.6 Accessing IaaS

- Cloud provisioning means allocating a cloud service provider's resource to a customer.
- It is a key feature of cloud computing. It refers to how a client gets cloud services and resources from a provider.
- The cloud services that customers can subscribe to include infrastructure-as-a-service (IaaS), software-as-a-service (SaaS), and platform-as-a-service (PaaS) in public or private environments.

Types of Cloud Provisioning

- There are various cloud provisioning delivery models.
- Each model depends on the types of resources or services an organization purchases, how and when the cloud service provider delivers them, and how customers pay for them.
- The three models
  - Advanced
  - Dynamic
  - User self-provisioning



Advanced Cloud Provisioning

- Also known as “post-sales cloud provisioning,” customers get the resources upon contract or service signup. They sign formal contracts with the cloud service provider. The provider then prepares and delivers the agreed-upon resources or services.
- The customers are charged a flat fee or billed every month.

Dynamic Cloud Provisioning

- Also referred to as “on-demand cloud provisioning,” customers are provided with resources on runtime.
- In this delivery model, cloud resources are deployed to match customers’ fluctuating demands. Deployments can scale up to accommodate spikes in usage and down when demands decrease.
- Customers are billed on a pay-per-use basis. When this model is used to create a hybrid cloud environment, it is sometimes called “cloud bursting.”

User Cloud Provisioning

- In this delivery model, customers add a cloud device themselves. Also known as “cloud self-service,” clients buy resources from the cloud service provider through a web interface or portal.
- The model usually involves creating a user account and paying for resources with a credit card. The resources are quickly spun up and made available for use within hours, if not minutes.
- An example of this includes an employee purchasing cloud-based productivity applications via Microsoft 365 or G Suite.

Cloud Provisioning Benefits

Cloud provisioning has several benefits that are not available with traditional provisioning approaches, such as:

Scalability

- The traditional information technology (IT) provisioning model requires organizations to make large investments in their on-premises infrastructure. That needs extensive preparation and

forecasting of infrastructure needs since on-premises infrastructures are often set up to last for many years.
- The cloud provisioning model, meanwhile, lets companies simply scale up and down their cloud resources depending on their short-term usage requirements.

Speed

Organizations' developers can quickly spin up several workloads on-demand, so the companies no longer require IT administrators to provide and manage computing resources.

Cost Savings

While traditional on-premises technology requires large upfront investments, many cloud service providers let their customers pay for only what they consume. But the attractive economics of cloud services presents challenges, too, which may require organizations to develop a cloud management strategy.

Cloud Provisioning Challenges

Like any other technology, cloud provisioning also presents several challenges, including:

Complex management and monitoring

- Organizations may need several provisioning tools to customize their cloud resources.
- Many also deploy workloads on more than one cloud platform, making viewing everything on a central console more challenging.

Resource and service dependencies

- Cloud applications and workloads often tap into basic infrastructure resources, such as computing, networking, and storage. But public cloud service providers offer higher-level ancillary services like serverless functions and machine learning (ML) and big data capabilities.
- Such services may carry dependencies that can lead to unexpected overuse and surprise costs.

Policy enforcement

- User cloud provisioning helps streamline requests and manage resources but requires strict rules to make sure unnecessary resources are not provided. That is time-consuming since different users require varying levels of access and frequency.
- Setting up rules to know who can provide which resources, for how long, and with what budgetary controls can be difficult.

Adopting IaaS: Cloud Migration Strategies

- Following are the most common approaches to cloud migration, taken from the influential "5 R's" model proposed by Gartner.
  - Rehosting
  - Replatforming, Refactoring, or Re-architecture
  - Repurchasing
  - Retire
  - Retain

Rehosting

- Re-hosting (also known as "lift and shift") is the fastest way to move your application to the cloud.
- This is usually the first approach taken in a cloud migration project because it allows moving the application to the cloud without any changes.
- Both physical and virtual servers are migrated to infrastructure as a service (IaaS).
- Lift and shift are commonly used to improve performance and reliability for legacy applications.

Replatforming, Refactoring, or Re-architecture

- This migration strategy involves detailed planning and a high investment, but it is the only strategy that can help you get the most out of the cloud.
- Applications that undergo replatforming or re-architecture are completely rebuilt on cloud-native infrastructure.
- They scale up and down on-demand, are portable between cloud resources and even between different cloud providers.

Repurchasing

- In most cases, repurchasing is as easy as moving from an on-premise application to a SaaS platform.
- Typical examples are switching from internal CRM to Salesforce.com, or switching from internal email server to Google's G Suite.
- It is a simple license change, which can reduce labor, maintenance, and storage costs for the organization.

Retire

- When planning a move to the cloud, it often turns out that part of the company's IT product portfolio is no longer useful and can be decommissioned.
- Removing old applications allows you to focus time and budget on high priority applications and improve overall productivity.

Retain

- Moving to the cloud doesn't make sense for all applications. You need a strong business model to justify migration costs and downtime.
- Additionally, some industries require strict compliance with laws that prevent data migration to the cloud. Some on-premises solutions should be kept on-premises, and can be supported in a hybrid cloud migration model.

What is Cloud Management?

- Cloud management refers to the exercise of control over public, private or hybrid cloud infrastructure resources and services.
- A well-designed cloud management strategy can help IT pros control those dynamic and scalable computing environments.
- Cloud management can also help organizations achieve three goals:
  - *Self-service* refers to the flexibility achieved when IT pros access cloud resources, create new ones, monitor usage and cost, and adjust resource allocations.
  - *Workflow automation* lets operations teams manage cloud instances without human intervention.
  - *Cloud analysis* helps track cloud workloads and user experiences.
- Without a competent IT staff in place, it's difficult for any cloud management strategy to succeed.

- These individuals must possess knowledge of the proper tools and best practices while they keep in mind the cloud management goals of the business.
- Companies are more likely to improve cloud computing performance, reliability, cost containment and environmental sustainability when they adhere to tried-and-true cloud optimization practices.
- There are many ways to approach cloud management, and they are ideally implemented in concert.
- Cost-monitoring tools can help IT shops navigate complex vendor pricing models. Applications run more efficiently when they use performance optimization tools and with architectures designed with proven methodologies.
- Many of these tools and strategies dovetail with environmentally sustainable architectural strategies to lower energy consumption.
- Cloud management decisions must ultimately hinge on individual corporate priorities and objectives, as there is no single approach.

What is Cloud Monitoring?

- Cloud monitoring measures the conditions of a workload and the various quantifiable parameters that relate to overall cloud operations.
- Results are monitored in specific, granular data, but that data often lacks context.
- Cloud observability is a process similar to cloud monitoring in that it helps assess cloud health. Observability is less about metrics than what can be gleaned from a workload based on its externally visible properties.
- There are two aspects of cloud observability: methodology and operating state. Methodology focuses on specifics, such as metrics, tracing and log analysis.
- Operating state relies on tracking and addresses state identification and event relationships, the latter of which is a part of DevOps.

Why Cloud Monitoring?

Cloud monitoring can perform the following capabilities:

- Monitoring cloud data across distributed locations
- Eliminating potential breaches by providing visibility into files, applications, and users
- Continually monitoring the cloud to ensure real-time file scans
- Regular auditing and reporting to ensure security standards
- Merging monitoring tools with different cloud providers

AWS First-Party Monitoring Tools

- There are multiple services and utilities available from AWS that you can use to monitor your systems and access.
- Some of these tools are included in existing services, while others are available for additional costs.
  - AWS CloudTrail
  - AWS CloudWatch
  - AWS Certificate Manager
  - Amazon EC2 Dashboard

AWS CloudTrail

- CloudTrail is a service that you can use to track events across your account.

- The service automatically records event logs and activity logs for your services and stores the data in S3.
- Collected data includes user identities, traffic origin IPs, and timestamps.
- You can view all management events for free for the most recent 90 days. Data events and insights based on your data are also available for an additional fee.

AWS CloudWatch

- CloudWatch is a service you can use to aggregate, visualize, and respond to service metrics.
- CloudWatch has two main components: alarms, which create alerts according to thresholds for single metrics, and events, which can automate responses to metric values or system changes.

AWS Certificate Manager

- Certificate Manager is a tool you can use to provision, manage, and apply transport layer security (TLS) and secure sockets layer (SSL) certificates.
- These certificates are used to prove your services or devices' authenticity and enable you to secure network connections.

Amazon EC2 Dashboard

- EC2 Dashboard is a monitoring tool for the Amazon EC2 virtual machine service.
- You can use this dashboard to monitor and maintain your EC2 instances and infrastructure.
- The dashboard lets you view instance states and service health, manage alarms and status reports, view scheduled events, and assess volume and instance metrics

Google Cloud Platform Monitoring Tools

Azure Monitoring Tools

Benefits of Cloud Monitoring

There are innumerable benefits cloud monitoring provides. Even businesses that solely rely on a private cloud architecture can enjoy key cloud monitoring deliverables, including:

- Improving the security of cloud applications and networks
- Simplifying the implementation of continuity plans, enabling proactive (rather than reactive) risk remediation
- Achieving and maintaining ideal application performance
- Optimizing service availability thanks to rapid issue reporting and rapid resolutions
- Reduction of surprise cloud cost leaks thanks to complete architecture visibility
- Simple scaling in the event cloud activity increases
- Usability on multiple devices, ensuring cloud awareness at all times

Cloud Monitoring Best Practices

As you implement a cloud monitoring service, keep the following best practices in mind to ensure you experience the full benefits.

- Decide which activity(ies) need to be monitored. Choose the metrics that matter the most to your bottom line.
- Consolidate report data onto a single platform to eliminate confusion and complexity that arises from juggling multiple cloud services and infrastructures. Your solution should report data from various sources and present them in one platform, enabling you to calculate metrics comprehensively.
- Keep track of subscription and service fees. The more you use your cloud monitoring service, the pricier it will be to use. Choosing a more advanced service can track how much activity is occurring on the cloud and determine costs from there.
- Be aware of which users are using which cloud applications to track accountability. You'll also need to know what these users see when they're using certain applications, and you'll want to monitor response time, frequency of use, and other metrics overall.

- Automate rules with the appropriate data to account for activities that go over or below your thresholds, ensuring you're able to add or remove servers to maintain consistent performance.
- Separate your monitoring data from your applications and services, and centralize this information to ensure your stakeholders have easy access.
- Always test your cloud monitoring tools at a regular cadence. While a service may seem operational, an outage or breach will truly put it to the test, so test your tools to ensure there are no surprises.

Section 3: Exercises

Exercise 1: Mention the Virtualization Elements in below Diagram.

Exercise 2: Participate in the group discussion on following topics:

a) Concept of Infrastructure as a Service (IaaS)
b) Various Enabling Technologies
c) Elastic Storage Devices Information
d) How to Access IaaS
e) Provision Servers on Demand
f) Tools and Support for Management and Monitoring

Section 4: Assessment Questionnaire

1. What is IaaS?
2. What is Virtualization?
3. How Virtualization is Different from Cloud Computing?
4. What are the Main Types of Virtualizations?
5. What is Elasticity?
6. What are the Use Cases of Cloud Elasticity?
7. What is Cloud Provisioning?
8. What are the types of Cloud Provisioning?
9. What is 5R Approach of Cloud Migration?
10. List few of the AWS Monitoring Tools.

----------End of Module----------

MODULE 6
Building a Business Case

Section 1: Learning Outcomes

After completing this module, you will be able to:

- Explain the Business Case planning for Cloud Adoption
- Calculate the Financial Implications
- Compare In-house Facilities to the Cloud
- Estimate Economic Factors Downstream
- Select appropriate Service-Level Agreements
- Safeguard access to Assets in the cloud
- Describe Security, Availability and Disaster Recovery Strategies

Section 2: Relevant Knowledge

6.1 Re-architecting Applications for the Cloud

What is a business case?

- Your organization depends on information technology (IT) for its operations, and probably for creating and supplying its products as well. It's a significant expense. For these reasons, a move to the cloud must be carefully considered and planned.
- A business case provides a view of the technical and financial timeline of your environment and can represent the opportunities for reinvestment into further modernization.
- Developing a business case includes building a financial plan that takes technical considerations into account and aligns with business outcomes.

- It helps you foster support from your Finance team and other areas of the business, helps accelerate cloud migration, and enables business agility.

Key Components of a Business Case

When you're planning your business case to migrate to the cloud, there are several key components to consider.

Environment scope**, technical and financial**

- As you build out the on-premises view of your environment, think about how your environment scope, from both a technical and financial perspective, is aligned.
- You want to be sure the technical environment you're using for your plan matches up to the financial data.

Baseline financial data: Cost to run today

- When you build out your business case, it's important to pull your baseline financial data. Common questions you can ask to gather the financial data needed are:

- How much does it cost to run my environment today?
- What am I spending on servers in an average year?
- What am I spending in my data center operations categories, for example, power or lease costs?
- When is the next hardware refresh?

Three Types of Cloud Compared

Infrastructure as a Service (IaaS)

- The CPU, data storage, bandwidth and an operating system delivered as a flexible service (ordered and provisioned incrementally).
- Each customer can then load their application software stack on top, taking advantage of this easily-resized Cloud Infrastructure (CI) on demand.
- This allows for scaling up or down as needed, providing a huge advantage to businesses in terms of flexibility and preservation of capital.

Platform as a Service (PaaS)

- Simply, this adds to CI a full software stack; for example, Linux.
- Each customer is then able to write or load applications into this environment, with the provider responsible for expanding or contracting all elements to adapt to the changing requirements of the users.
- Where a high degree of availability or service is required, this provides an impressive advantage for businesses looking to react and scale rapidly.

customization or control.

Software as a Service (SaaS)

- This takes CI and Cloud Platform (CP) and adds a fully managed application, with examples of this being salesforce.com, dropbox.com or facebook.com.
- The user will consume these apps, incrementally and usually on a per-user basis, with very little long-term commitment.

- This does provide advantages to businesses needing an application for a short-term or test basis, or in a business where pure applications and dedicated staff are scarce or expensive. But, this format also provides very little customization or control.

Approach to Building a Business Case

- In building a business case, ultimately, we are driving an expression of ROI. And, eventually, this leads into comparison between the deployment of capital (CapEx) vs. the preservation of capital - aka the deployment of operational expense (OpEx).
- This is purely a traditional “buy vs build” analysis and mostly straight forward and we'll present examples of this type of analysis and calculation within this series. However, when it comes to the cloud, there are different approaches to building the business case to ensure all unique costs are truly identified:
  - Infrastructure business case
  - Applications business case
  - Talent business case

Infrastructure Case

- Usually, there is some compelling event to initiate a move to the cloud, such as a compute upgrade due to increased demands from users or applications, end-of-life of data centre facility assets or a facility move where everything needs to be built again.
- The initial and most significant savings are usually found when abandoning infrastructure in favour of the cloud, with infrastructure savings being the most significant part of the business case in terms of cost savings.
- The reason why is that in-house IT is typically under-utilised, because when infrastructure purchases are being considered, not all applications that will be deployed on it are known and so a margin is added for this misunderstood capacity requirement.
- Additionally, over-deployment is a result of companies configuring infrastructure for peak loads.
- We add to these the other fixed and variable costs we identified previously: cost of specialised data centre assets to house, power and cool servers, the cost of real estate which includes carrying finance charges, lease costs and other terms, the skilled staffing costs of maintaining the data centre and the systems within it.
- Other costs include back-ups, redundancy at a second facility, certifications, security and decommissioning costs when moving to the cloud.

Applications case

- With the three main types of cloud there are a number of options of what to do with applications, but this requires a look at what the main drivers are.

- This includes a major fork-lift to update an in-house custom application, a shift to a new application requiring higher availability and performance, and scarcity of maintenance resources such as talent and quality control. Therefore, there's many options for then dealing with the applications.
- Each of these reasons will have different cost implications, which will need to be outlined for the business case. For example, leaving one application where it is, but moving others, places a greater share of the remaining infrastructure costs on that application.

Reasons to Move Applications to the Cloud

6.2 Calculating the Financial Implication

How to Calculate Cloud Computing Cost

- Determining the potential costs of cloud computing can be as complex as the technology itself.
- Not only are you faced with the different pricing structures of cloud providers, you must find a reasonable way to estimate the resources that you will need in the future.
- Despite your best efforts, there is no guarantee that your calculations will be correct. That's because of the variable cost architecture of the cloud.

From Capex to Opex

- The traditional computing model was dependent on significant capital expenditures (CAPEX).
- The one-time purchase of hardware, software and licenses meant that companies had to squeeze as much work out of these resources as possible throughout the life cycle of these platforms.
- There was always a strong focus on the maintenance and configuration of proprietary machines. This meant that vendor support was essential to keeping systems current and healthy.

Cost Centers in the Cloud

The three cost centres of the cloud

- Compute
- Storage

- Network

These provide the outline for the calculation of cloud costs. But these broad areas don't account for everything. The options and variables related to cloud usage can make predicting costs something of a guessing game. But the main components give us something to work with.

Compute

- The costs for compute depend entirely on what you're going to do with it. How much processing power is required for your computing projects? A common way to deal with computing capability has been to purchase more than you need. Better to have too much than too little, as they say. But the scalability of cloud computing means that you can take a much different approach.
- If you have a temporary project that is entirely contingent upon the response of website visitors, you can use cloud computing to scale up automatically and scale back down just as quickly. The computing systems that you set up with cloud providers can purchase for on-demand usage or for a fixed period of time.
- The parameters involved in selecting a CPU include the operating system and the expected usage (in percent). Cloud providers will then calculate the cost of CPU based on their cost per gigabyte (GB) of virtual RAM.

Cost Centers in the Cloud

Storage

- Advancements in storage technology have brought down prices considerably.
- It is no longer necessary to have dedicated hardware for each client or project.
- Virtual disks have replaced their physical counterparts.
- The same scalability in the compute sphere applies to storage as well. Storage is calculated in units of GB of virtual disk.

Network

- This area is generally measured in GB of data transfer. But bandwidth is also calculated in terabytes (TB) or petabytes (PB).
- While these are the main cost centers, not everything that a cloud provider offers will fit neatly in these three categories.
- Each provider packages their offerings in different ways. It might even seem like comparing apples to oranges when putting one service up against another

6.3 Comparing in-house facilities to the cloud

Adoption and Consumption Strategies

The selection of strategies for enterprise cloud computing is critical for IT capability as well as for the earnings and costs the organization experiences, motivating efforts toward convergence of business strategies and IT. Some critical questions toward this convergence in the enterprise cloud paradigm are as follows:

- Will an enterprise cloud strategy increase overall business value?
- Are the effort and risks associated with transitioning to an enterprise cloud strategy worth it?
- Which areas of business and IT capability should be considered for the enterprise cloud?
- Which cloud offerings are relevant for the purposes of an organization?
- How can the process of transitioning to an enterprise cloud strategy be piloted and systematically executed?

These questions are addressed from two strategic perspectives:

(1) Adoption

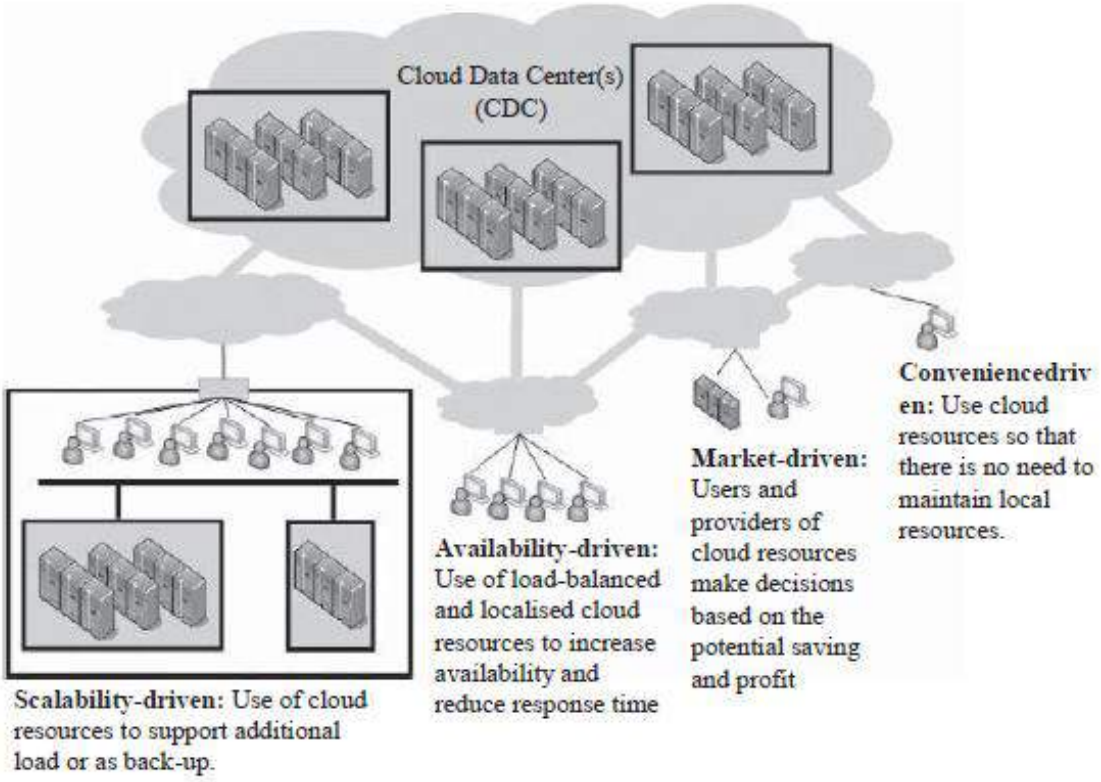
(2) Consumption

where an organization makes a decision to adopt a cloud computing model based on fundamental drivers for cloud computing— scalability, availability, cost and convenience.

Enterprise cloud adoption strategies

1. **Scalability-Driven Strategy.** The objective is to support increasing workloads of the organization without investment and expenses exceeding returns.
2. **Availability-Driven Strategy.** Availability has close relations to scalability but is more concerned with the assurance that IT capabilities and functions are accessible, usable and acceptable by the standards of users.
3. **Market-Driven Strategy.** This strategy is more attractive and viable for small, agile organizations that do not have (or wish to have) massive investments in their IT infrastructure on their profiles and requests service requirements.
4. **Convenience-Driven Strategy.** The objective is to reduce the load and need for dedicated system administrators and to make access to IT capabilities by users easier, regardless of their location and connectivity (e.g., over the Internet)

Enterprise cloud adoption strategies



Enterprise cloud adoption strategies using fundamental cloud drivers

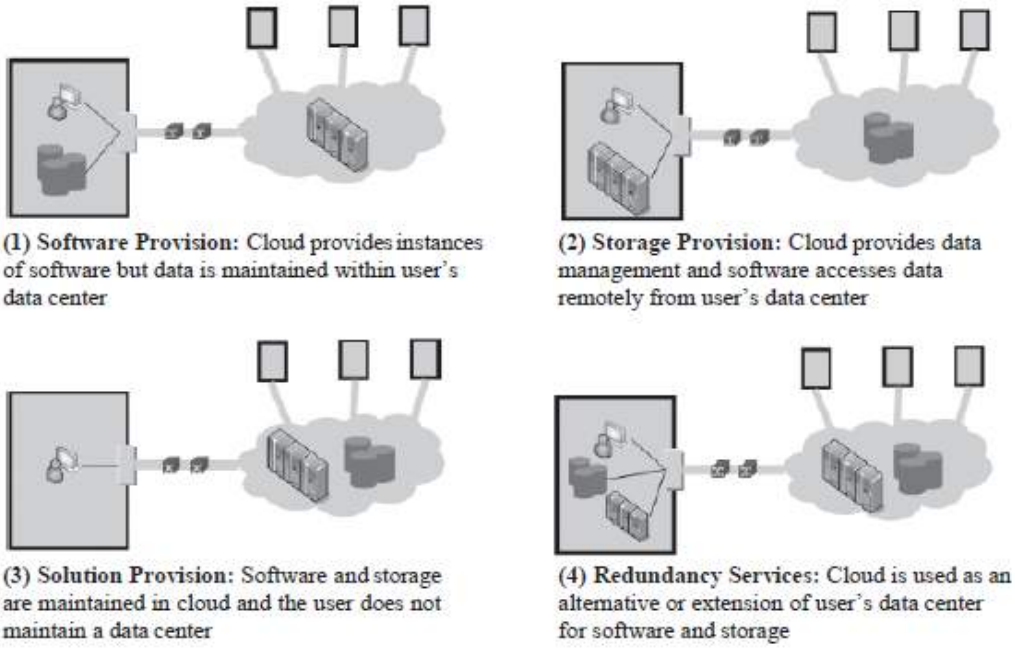
There are four consumptions strategies identified, where the differences in objectives, conditions and actions reflect the decision of an organization to trade-off hosting costs, controllability and resource elasticity of IT resources for software and data. These are discussed in the following:

1. Software Provision. This strategy is relevant when the elasticity requirement is high for software and low for data, the controllability concerns are low for software and high for data, and the cost reduction concerns for software are high, while cost reduction is not a priority for data, given the high controllability concerns for data, that is, data are highly sensitive.

2. Storage Provision. This strategy is relevant when the elasticity requirements is high for data and low for software, while the controllability of software is more critical than for data. This can be the case for data intensive applications, where the results from processing in the application are more critical and sensitive than the data itself.

3. Solution Provision. This strategy is relevant when the elasticity and cost reduction requirements are high for software and data, but the controllability requirements can be entrusted to the CDC.

4. Redundancy Services. This strategy can be considered as a hybrid enterprise cloud strategy, where the organization switches between traditional, software, storage or solution management based on changes in its operational conditions and business demands.



Enterprise cloud consumption strategies

Business Benefits of Cloud Computing

There are some clear business benefits to building applications in the cloud. A few of these are listed here:

***Almost Zero Upfront Infrastructure Investment*.**

If you have to build a large- scale system, it may cost a fortune to invest in real estate, physical security, hardware (racks, servers, routers, backup power supplies), hardware management (power management, cooling), and operations personnel. Because of the high upfront costs, the project

would typically require several rounds of management approvals before the project could even get started. Now, with utility-style cloud computing, there is no fixed cost or start-up cost

Just-in-Time Infrastructure

- In the past, if your application became popular and your systems or your infrastructure did not scale, you became a victim of your own success.
- By deploying applications in-the-cloud with just-in-time self-provisioning, you do not have to worry about pre-procuring capacity for large-scale systems.
- This increases agility, lowers risk, and lowers operational cost because you scale only as you grow and only pay for what you use.

More Efficient Resource Utilization

- System administrators usually worry about procuring hardware (when they run out of capacity) and higher infrastructure utilization (when they have excess and idle capacity).
- With the cloud, they can manage resources more effectively and efficiently by having the applications request and relinquish resources on-demand.

Usage-Based Costing

- With utility-style pricing, you are billed only for the infrastructure that has been used. You are not paying for allocated infrastructure but instead for unused infrastructure. This adds a new dimension to cost savings.

- You can see immediate cost savings (some- times as early as your next month's bill) when you deploy an optimization patch to update your cloud application.
- For example, if a caching layer can reduce your data requests by 70%, the savings begin to accrue immediately and you see the reward right in the next bill. Moreover, if you are building platforms on the top of the cloud, you can pass on the same flexible, variable usage-based cost structure to your own customers.

Reduced Time to Market

- Parallelization is one of the great ways to speed up processing.
- If one compute-intensive or data-intensive job that can be run in parallel takes 500 hours to process on one machine, with cloud architectures, it would be possible to spawn and launch 500 instances and process the same job in 1 hour.
- Having available an elastic infrastructure provides the application with the ability to exploit parallelization in a cost-effective manner reducing time to market.

Technical Benefits of Cloud Computing

Some of the technical benefits of cloud computing includes:

Automation— "Scriptable Infrastructure"

- You can create repeatable build and deployment systems by leveraging programmable (API-driven) infrastructure.

- *Auto-scaling:* You can scale your applications up and down to match your unexpected demand without any human intervention. Auto-scaling encourages automation and drives more efficiency.

Proactive Scaling

Scale your application up and down to meet your anticipated demand with proper planning understanding of your traffic patterns so that you keep your costs low while scaling.

More Efficient Development Life Cycle

Production systems may be easily cloned for use as development and test environments. Staging environments may be easily promoted to production.

Improved Testability

Never run out of hardware for testing. Inject and automate testing at every stage during the development process. You can spawn up an "instant test lab" with preconfigured environments only for the duration of testing phase.

Disaster Recovery and Business Continuity

The cloud provides a lower cost option for maintaining a fleet of DR servers and data storage. With the cloud, you can take advantage of geo-distribution and replicate the environment in other location within minutes.

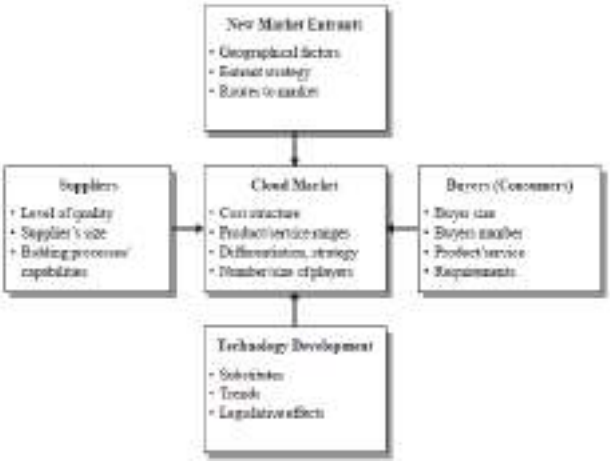
"Overflow" the Traffic to the Cloud

With a few clicks and effective load balancing tactics, you can create a complete overflow-proof application by routing excess traffic to the cloud.

6.4 Estimating Economic Factors Downstream

Business Drivers Toward a Marketplace

- In order to create an overview of offerings and consuming players on the market, it is important to understand the forces on the market and motivations of each player.
- The Porter model consists of five influencing factors/views (forces) on the market. The intensity of rivalry on the market is traditionally influenced by industry-specific characteristics:



Porter's five forces market model (adjusted for the cloud market)

6.5 Selecting Appropriate Service-Level Agreements

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.
- The capacity build up 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.
- Enterprises developed the web applications and deployed on the infrastructure of the third-party service providers.
- These providers get the required hardware and make it available for application hosting. Typically, the QoS parameters are related to the availability of the system CPU, data storage, and network for efficient execution of the application at peak loads. This legal agreement is known as the service-level agreement (SLA).

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.

There are two types of SLAs from the perspective of application hosting. These are described in detail here.

Infrastructure SLA. The infrastructure provider manages and offers guarantees on availability of the infrastructure, namely, server machine, power, network connectivity, and so on.

Application SLA. In the application co-location hosting model, the server capacity is available to the applications based solely on their resource demands. Therefore, the service

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.

Hardware availability	99% uptime in a calendar month
Power availability	99.99% of the time in a calendar month
Data centre network availability	99.99% of the time in a calendar month
Backbone network availability	99.999% of the time in a calendar month
Service credit for unavailability	Refund of service credit prorated on downtime period
Outage notification guarantee	Notification of customer within 1 hr of complete downtime
Internet latency guarantee	When latency is measured at 5-min intervals to an upstream provider, the average doesn't exceed 60 msec
Packet loss guarantee	Shall not exceed 1% in a calendar month

Key contractual components of an application SLA

Service-level parameter metric	• Web site response time (e.g., max of 3.5 sec per user request)
	• Latency of web server (WS) (e.g., max of 0.2 sec per request)
	• Latency of DB (e.g., max of 0.5 sec per query)
Function	• Average latency of WS = (latency of web server 1 +latency of web server 2) /2
	• Websiteresponsetime= Averagelatencyofwebserver+ latency ofdatabase
Measurement directive	• DB latency available via http://mgmtserver/em/latency
	• WS latency available viahttp://mgmtserver/ws/instanceno/ latency
Service-level objective	• Service assurance
	• website latency , 1 sec when concurrent connection , 1000
Penalty	• 1000 USD for every minute while the SLO was breached

Challenges for provisioning the infrastructure on demand

From the SLA perspective there are multiple challenges for provisioning the infrastructure on demand. These challenges are as follows:

a. The application is a black box to the MSP and the MSP has virtually no knowledge about the application runtime characteristics.

b. The MSP needs to understand the performance bottlenecks and the scalability of the application.

c. The MSP analyses the application before it goes on-live. However, subsequent operations/enhancements by the customers to their applications or auto updates beside others can impact the performance of the applications, thereby making the application SLA at risk.

d. The risk of capacity planning is with the service provider instead of the customer.

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.

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 catalogue.

Negotiation

Once the customer has discovered a service provider who can meet their application hosting need, the SLA terms and conditions needs to be mutually agreed upon before signing the agreement for hosting the application.

Operationalization

SLA operation consists of SLA monitoring, SLA ac- counting, and SLA enforcement. SLA monitoring involves measuring parameter values and calculating the metrics defined as a part of SLA and determining the deviations.

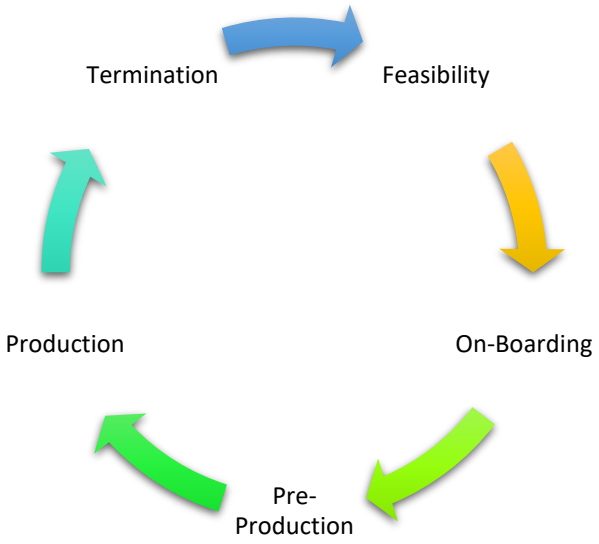
De-commissioning

SLA decommissioning involves termination of all activ- ities performed under a particular SLA when the hosting relationship between the service provider and the service consumer has 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



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
(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 data center.
3. The need and availability of a specific hardware and software required for hosting and running of the application.
4. Preliminary information about the application performance and whether they can be met by the MSP.

Performing the infrastructure feasibility involves determining the availability of infrastructural resources in sufficient quantity so that the projected demands of the application can be met.

On-Boarding of Application

- 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.
- The application is accessible to its end users only after the on- boarding activity is completed.

On-boarding activity consists of the following steps:

a. Packing of the application for deploying on physical or virtual environments. Application packaging is the process of creating deployable components on the hosting platform (could be

physical or virtual). Open Virtualization Format (OVF) standard is used for packaging the application for cloud platform.

b. The packaged application is executed directly on the physical servers to capture and analyse the application performance characteristics.

c. The application is executed on a virtualized platform and the application performance characteristics are noted again.

d. Based on the measured performance characteristics; different possible SLAs are identified. The resources required and the costs involved for each SLA are also computed.

e. Once the customer agrees to the set of SLOs and the cost, the MSP starts creating different policies required by the data center for automated management of the application. These policies are of three types:

(1) Business

(2) Operational

(3) Provisioning

Business policies help prioritize access to the resources in case of contentions.

Preproduction

- Once the determination of policies is completed as discussed in previous phase, the application is hosted in a simulated production environment.
- 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 applica- tion to go on-live.

Production

- In this phase, the application is made accessible to its end users under the agreed SLA.
- In the case of the former, on-boarding activity is repeated to analyse the application and its policies with respect to SLA fulfilment. 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.

6.6 Safeguarding Access to Assets in the Cloud

Security Best Practices

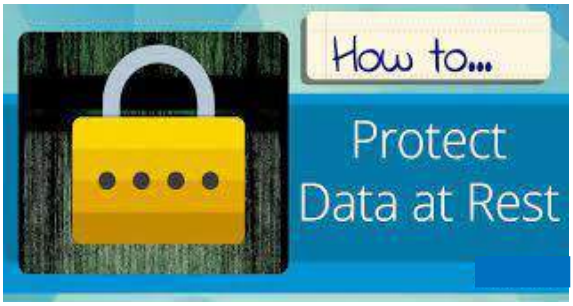
- In a multi-tenant environment, cloud architects often express concerns about security. Security should be implemented in every layer of the cloud application architecture.
- Physical security is typically handled by your service provider (Security Whitepaper, which is an additional benefit of using the cloud. Network and application-level security is your responsibility, and you should implement the best practices as applicable to your business.
- It is recommended to take advantage of these tools and features mentioned to implement basic security and then implement additional security best practices using standard methods as appropriate or as they see fit.

Protect Your Data in Transit

- If you need to exchange sensitive or confidential information between a browser and a Web server, configure SSL on your server instance. You'll need a certificate from an external certification authority like VeriSign or Entrust.
- The public key included in the certificate authenticates your server to the browser and serves as the basis for creating the shared session key used to encrypt the data in both directions.
- Create a virtual private cloud by making a few commands line calls (using Amazon VPC). This will enable you to use your own logically isolated resources within the AWS cloud, and then connect those resources directly to your own data center using industry-standard encrypted IPSec VPN connections.
- You can also set up an OpenVPN server on an Amazon EC2 instance and install the OpenVPN client on all user PCs.

Protect your Data at Rest

- If you are concerned about storing sensitive and confidential data in the cloud, you should encrypt the data (individual files) before uploading it to the cloud.
- For example, encrypt the data using any open source or commercial PGP-based tools before storing it as Amazon S3 objects and decrypt it after download.
- This is often a good practice when building HIPPA-compliant applications that need to store protected health information (PHI).
- On Amazon EC2, file encryption depends on the operating system. Amazon EC2 instances running Windows can use the built-in Encrypting File System (EFS) feature available in Windows. This feature will handle the encryption and decryption of files and folders automatically and make the process transparent to the users.



Secure Your Application

- Every Amazon EC2 instance is protected by one or more security groups that is, named sets of rules that specify which ingress (i.e., incoming) network traffic should be delivered to your instance.
- You can specify TCP and UDP ports, ICMP types and codes, and source addresses. Security groups give you basic firewall-like protection for running instances.

6.7 Security, Availability and Disaster Recovery Strategies

What is disaster recovery in cloud computing?

- Disaster recovery (DR) is the process that goes into preparing for and recovering from a disaster.
- This disaster could take one of a number of forms, but they all end up in the same result: the prevention of a system from functioning as it normally does, preventing a business from completing its daily objectives.

What kind of disasters should you prepare for?

There are three main categories of disaster that can affect businesses:

Natural disasters: Natural disasters such as floods or earthquakes are rarer but not infrequent. If a disaster strikes an area that contains a server that hosts the cloud service you're using, this could disrupt services and require disaster recovery operations.

Technical disasters: Perhaps the most obvious of the three, technical disasters encompass anything that could go wrong with the cloud technology. This could include power failures or a loss of network connectivity.

Human disasters:

- Human failures are a common occurrence and are usually accidents that happen whilst using the cloud services. These could include inadvertent misconfiguration or even malicious third-party access to the cloud service.

- The cloud providers are responsible for everything they have direct control over. This includes the resiliency of the general infrastructure such as the hardware, software, network and facilities. You, the customer, are usually responsible for areas such as the cloud configuration, secure data backups, the workload architecture and the availability.

Why is disaster recovery important?

- Creating protocols and contingencies for disaster recovery is vital for the smooth operation of business. In the event of a disaster, a company with disaster recovery protocols and options can minimize the disruption to their services and reduce the overall impact on business performance.
- Minimal service interruption means a reduced loss of revenue which, in turn, means user dissatisfaction is also minimised.
- Having plans for disaster in place also means your company can define its Recovery Time Objective (RTO) and its Recovery Point Objective (RPO). The RTO is the maximum acceptable delay between the interruption and continuation of the service and the RPO is the maximum amount of time between data recovery points.
- Quantifying these areas can help your company identify its optimal protection level for disaster recovery and choose the right protocols to implement such as backups and multiple servers.

Some examples of cloud computing disasters

Although uncommon, disasters in cloud computing have occurred in the past and even to some of the largest cloud providers such as AWS.

OVHCloud

A data centre run by OVHCloud was destroyed in early 2021 by a fire. All four data centres had been too close, and it took over six hours for firefighters at the scene to put out the blaze. **This severely affected the cloud services** run by OVHCloud and spelt disaster for companies whose entire assets were hosted on those servers.

AWS

In June 2016, storms in Sydney battered the electrical infrastructure and caused an extensive power outage. This led to **the failure of a number of Elastic Compute Cloud instances and Elastic Block Store volumes** which hosted critical workloads for a number of large companies. This meant that some heavily trafficked websites and the online presence of some of the biggest brands was decimated for over ten hours on a weekend, severely affecting business.

Amazon

In February 2017 an Amazon employee was attempting to debug an issue with the billing system when they accidentally took more servers offline than they needed to. This started **a domino effect that removed two other server subsystems** which then snowballed to other subsystems. This meant that thousands of people were unable to access Amazon servers for a few hours.

What are the benefits of cloud disaster recovery in the cloud?

- **Using** the cloud for cloud disaster recovery means that data backups don't have to be maintained by the customer on disks or physical hard drives.
- The distributed nature of the cloud means that services can be spread out to different servers in different geographical locations, essentially providing complete protection against local natural disasters.
- Some of the responsibility can be offloaded onto the cloud provider.
- The cloud provider is responsible for the core resilience of the infrastructure of the cloud, removing this worry from the customer.
- Cloud disaster recovery using the cloud also proves to be cost-effective. Because cloud providers only charge for the services that they use, your business can pick and choose which services it wants from the provider.
- This leads to a huge cost reduction by increasing the personalization of the package that your business pays for.

How should you prepare your recovery plans, step by step?

Here are 5 steps that can help you prepare a recovery plan:

1. Your disaster recovery plan should be part of your business continuity plan

This should involve definitions of RTO and RPO to help you decide which cloud services you'll need and improve cost efficiency.

2. If you haven't done so already, define the RTO and RPO for your disaster recovery

This forms the basis of your disaster recovery plan and, in turn, the kinds of disaster recovery services you'll need.

3. Design your plan with your recovery goals in mind

This involves looking at your RTO and RPO points to decide which disaster recovery pattern you'll need to meet those criteria. Your recovery goals should outline the maximum and minimum affects to your services

4. Design for end-to-end recovery

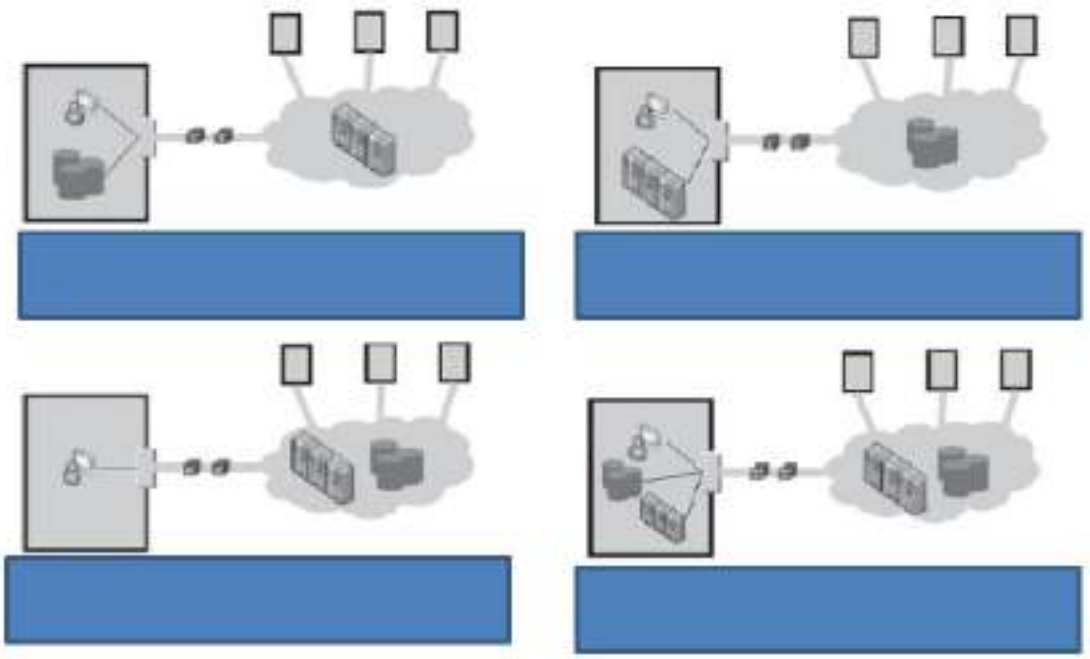
Your plan should include recovery for every aspect of your business that needs to be operational.

5. Create specific tasks to ensure a smooth-running process

The more specific your tasks are, the easier the recovery process will be and the fewer chances there will be of deviating from the plan.

Section 3: Exercises

Exercise 1: Write down the name of all enterprise cloud consumption strategies in below diagram**.**



Exercise 2: Participate in the group discussion on following topics:

a) Business Case planning for Cloud Adoption
b) Compare In-house Facilities to the Cloud
c) Estimate Economic Factors Downstream
d) Service-Level Agreements
e) Safeguard access to Assets in the cloud
f) Security, Availability and Disaster Recovery Strategies

Section 4: Assessment Questionnaire

1. What are the business benefits of cloud computing?
2. Explain SLA in Cloud computing?
3. What are the types of SLA?
4. List the Key Components of SLA?
5. Explain Phases Life cycle of SLA?
6. List out the Security practices?
7. What kind of Disaster happen? How to recover disaster in Cloud Computing?

----------End of Module----------

MODULE 7
Migrating to Cloud

Section 1: Learning Outcomes

After completing this module, you will be able to:

- Explain about the Technical Consideration for Cloud Migration
- Define the term 'Cloud Migration'
- Re-architect applications for the cloud
- Integrate the cloud with existing applications
- Avoid vendor lock-in
- Plan the migration and selecting a vendor

Section 2: Relevant Knowledge

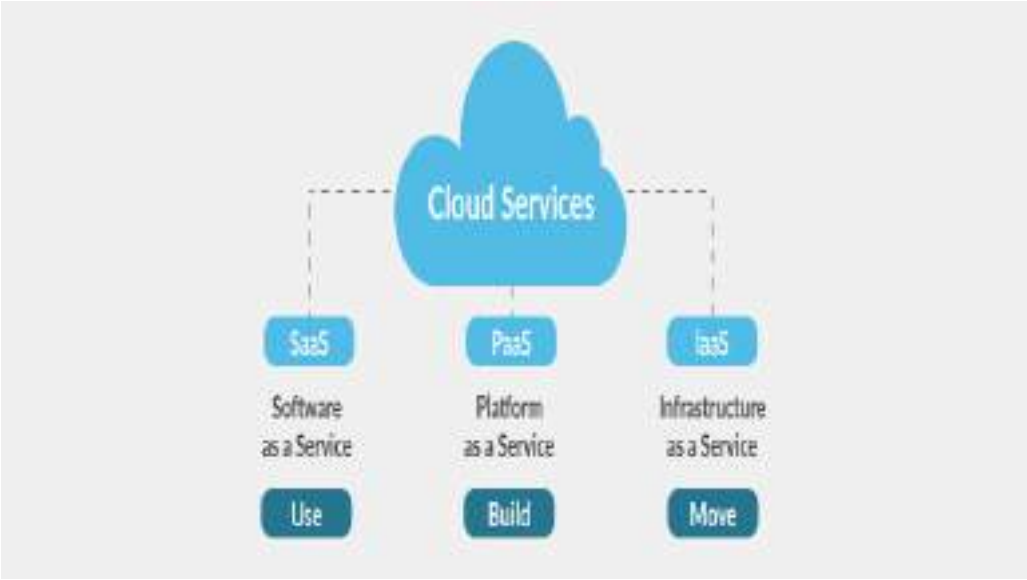
7.1 Re-architecting Applications for the Cloud

What Is Cloud Application Migration?

- The term "application migration" refers to the process of shifting software applications between computing environments.
- The process may apply to moving applications between a public cloud to a private cloud or moving applications from a local server to a cloud environment.
- Cloud migration helps organizations leverage the advantages of the cloud for their applications, including cost reduction, a higher level of scalability, and quick application updates.

What Are Your Cloud Migration Options?

- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS)



Software Migration Challenges to Overcome

Here are some of the main challenges involved in software migration.

Unexpected Costs

- When you migrate an application, you could face unexpected costs resulting from the complexity of the migration process.
- For example, you may have to train staff in using the new system or toolset, requiring extra hours and expenses. For your migration to be successful, you need to assess the expected costs realistically, considering potential complications.

Disruptions and Downtime

- Migration can impact processes that are critical to your business functions. If you experience an unanticipated outage, you may lose customers and revenue.
- To reduce unexpected downtime, you should consider the potential issues that may affect performance so you can address them in advance.

Maintaining Privacy

- It is essential to protect the privacy of your business operations and data when migrating to a third-party system, such as a cloud server. Whenever you work with a third-party vendor, you need to carefully oversee the migration process and ensure the proper SLAs are in place.

Maintaining Compliance

- You need to ensure that the new environment is compliant with regulations such as HIPAA.
- It is important to have a compliance strategy in place before you begin the migration process to find suitable vendors and solutions.

Stakeholder Commitment Issues

- Migration projects often take a long time to complete, testing the commitment of key stakeholders.
- You need to have a clearly defined long-term with measurable targets to help keep team leaders and department heads on board.

Using Different Systems Simultaneously

- Organizations typically migrate applications gradually to maintain business continuity, resulting in a period of overlap between the data and functions of the old and new environments.
- This overlap can create confusion as to which system should be used for each task. You need to have a clear plan outlining the data storage requirements of each migration phase.

Application Migration Plan Stages

- Your application migration plan is key to making the process manageable. While the specifics of a migration plan differ for each organization, any application migration plan should address the following basic elements.

Identify and Assess Your Applications

- First, you need to discover and audit all applications used in your enterprise environment. You should assess the importance and complexity of your applications, categorizing them as business-critical or non-critical.

- An application assessment should include any requirements for modifications or re-coding, helping you decide whether to migrate or replace the application.

Determine Which Legacy Applications to Migrate

- Most organizations continue to use legacy applications long after the introduction of new technologies.
- You might want to keep your legacy applications to avoid the expense or disruption of acquiring a replacement—as long as they perform adequately.

Application Migration Plan Stages

- When migrating to a new environment, especially in the cloud, legacy applications can be difficult to migrate or maintain.
- You can migrate some applications unchanged or with minor alterations but replacing other applications with cloud-compatible alternatives could be cheaper.

Calculate Your TCO

- Software migration carries a significant risk of unanticipated costs. Review your application migration plan to evaluate the total cost of ownership (TCO).
- You can compare various scenarios to see which options strike an acceptable balance between cost savings and performance. Consider factors such as the maintenance costs, the cost of replacing or acquiring new applications, and training.

Assess the Project Duration and Identify Potential Risks

Do your best to forecast the likely duration of your migration project and consider the risks of unexpected hurdles. Your forecast will not be perfect, but it should help reduce the risk of overblown costs and disruptions.

Managed Application Migration

Cloud providers offer managed services that can make it easier to migrate your applications to the cloud. Here are a few types of application migration services you can use to plan, execute, and automate an application migration.

Migration Blueprint

- In a complete blueprint service offer, your vendor helps you define your migration objectives and strategy by recognizing your users' needs and your organizational requirements.
- They also collect details about your environment and applications, developing a complete action plan for the migration process.

Migration Deployment

- If you select a managed deployment, your vendor helps you strategize and plan your migration.
- They also help you manage the migration and any related troubleshooting and testing. This method is typically a turn-key option that features full-scale and end-to-end support.

Cloud Managed Services

- A managed cloud service option provides observation and maintenance of your cloud-based IT environment. Your managed cloud service provider takes responsibility for functions, including acquiring as-a-service providing on your behalf.

- They also manage cloud security. Application migration may also be part of the packaged service.

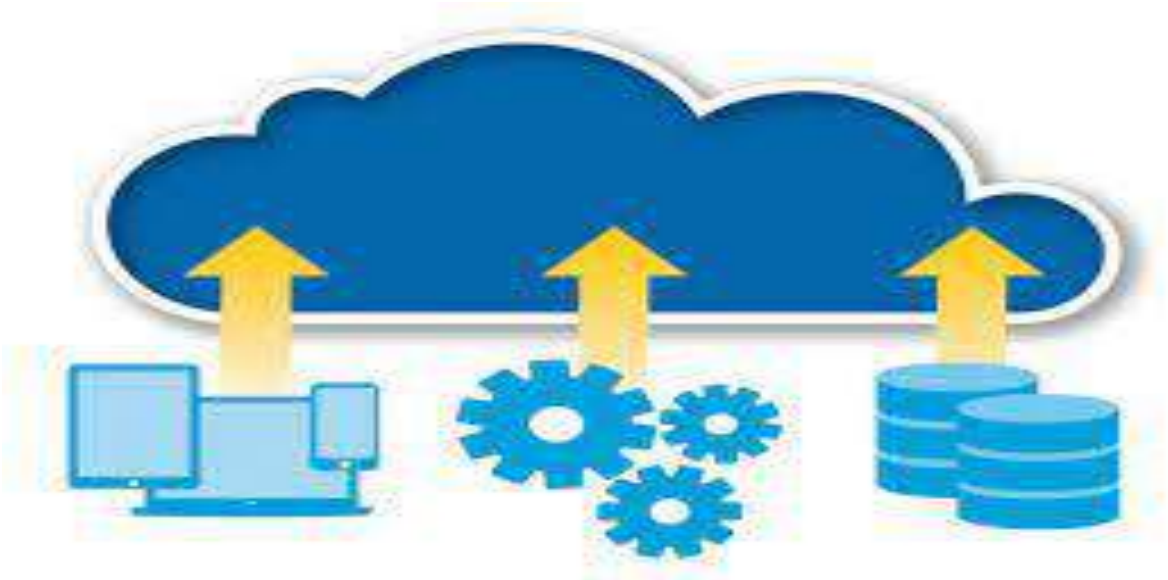
Application Modernization

- Application modernization services provide custom development services.
- They can help you prepare legacy applications for utilization in the cloud, by adapting them to run in virtualized environments or containers.

7.2 Migrating to the Cloud

What is Cloud Migration?

- Cloud migration is the process of moving data, applications or other business elements to a cloud computing environment.
- There are various types of cloud migrations an enterprise can perform. One common model is to transfer data and applications from a local on-premises data center to the public cloud.
- However, a cloud migration could also entail moving data and applications from one cloud platform or provider to another; this model is known as cloud-to-cloud migration.
- A third type of migration is a reverse cloud migration, cloud repatriation or cloud exit, where data or applications are moved off of the cloud and back to a local data center.



What are the Key Benefits of Cloud Migration?

Scalability

- Cloud computing can scale to support larger workloads and more users, much more easily than on-premises infrastructure.
- In traditional IT environments, companies had to purchase and set up physical servers, software licenses, storage and network equipment to scale up business services.

Cost

- Cloud providers take over maintenance and upgrades, companies migrating to the cloud can spend significantly less on IT operations.

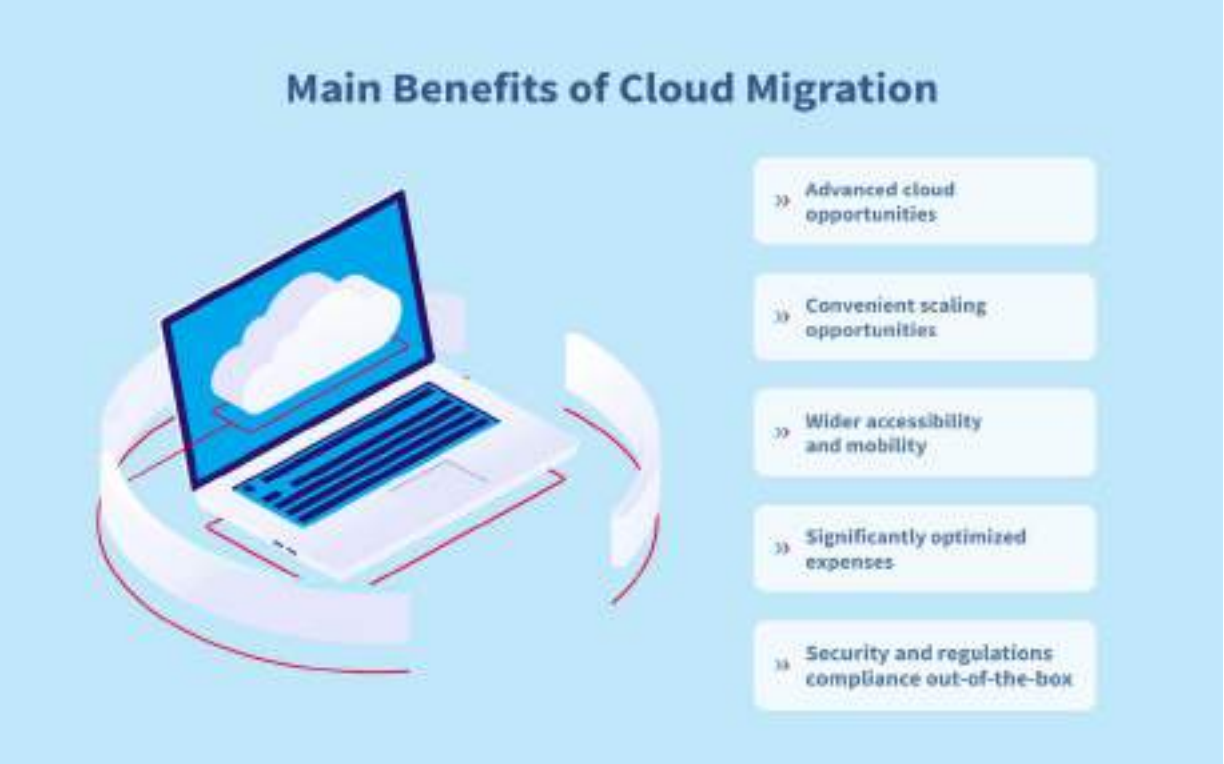
- They can devote more resources to innovation - developing new products or improving existing products.

Performance

Migrating to the cloud can improve performance and end-user experience. Applications and websites hosted in the cloud can easily scale to serve more users or higher throughput, and can run in geographical locations near to end-users, to reduce network latency.

Digital experience

Users can access cloud services and data from anywhere, whether they are employees or customers. This contributes to digital transformation, enables an improved experience for customers, and provides employees with modern, flexible tools.



What are Common Cloud Migration Challenges?

Cloud migrations can be complex and risky. Here are some of the major challenges facing many organizations as they transition resources to the cloud.

Lack of Strategy

- Many organizations start migrating to the cloud without devoting sufficient time and attention to their strategy.
- Successful cloud adoption and implementation requires rigorous end-to-end cloud migration planning.
- Each application and dataset may have different requirements and considerations, and may require a different approach to cloud migration.
- The organization must have a clear business case for each workload it migrates to the cloud.

Cost Management

- When migrating to the cloud, many organizations have not set clear KPIs to understand what they plan to spend or save after migration.
- This makes it difficult to understand if migration was successful, from an economic point of view. In addition, cloud environments are dynamic and costs can change rapidly as new services are adopted and application usage grows.

Vendor Lock-In

- Vendor lock-in is a common problem for adopters of cloud technology. Cloud providers offer a large variety of services, but many of them cannot be extended to other cloud platforms.
- Migrating workloads from one cloud to another is a lengthy and costly process. Many organizations start using cloud services, and later find it difficult to switch providers if the current provider doesn't suit their requirements.

Data Security and Compliance

- One of the major obstacles to cloud migration is data security and compliance. Cloud services use a shared responsibility model, where they take responsibility for securing the infrastructure, and the customer is responsible for securing data and workloads.
- So, while the cloud provider may provide robust security measures, it is your organization's responsibility to configure them correctly and ensure that all services and applications have the appropriate security controls.
- The migration process itself presents security risks. Transferring large volumes of data, which may be sensitive, and configuring access controls for applications across different environments, creates significant exposure.

Migrating into a Cloud

- The promise of cloud computing has raised the IT expectations of small and medium enterprises beyond measure. Large companies are deeply debating it.
- Cloud computing is a disruptive model of IT whose innovation is part technology and part business model in short, a disruptive techno-commercial model of IT.
- We propose the following definition of cloud computing: —It is a techno-business disruptive model of using distributed large-scale data centers either private or public or hybrid offering customers a scalable virtualized infrastructure or an abstracted set of services qualified by service-level agreements (SLAs) and charged only by the abstracted IT resources consumed
- Several small and medium business enterprises, however, leveraged the cloud much beyond the cautious user. Many start-ups opened their IT departments exclusively using cloud services very successfully and with high ROI. Having observed these successes, several large enterprises have started successfully running pilots for leveraging the cloud.
- Many large enterprises run SAP to manage their operations. SAP itself is experimenting with running its suite of products: SAP Business One as well as SAP Netweaver on Amazon cloud offerings.

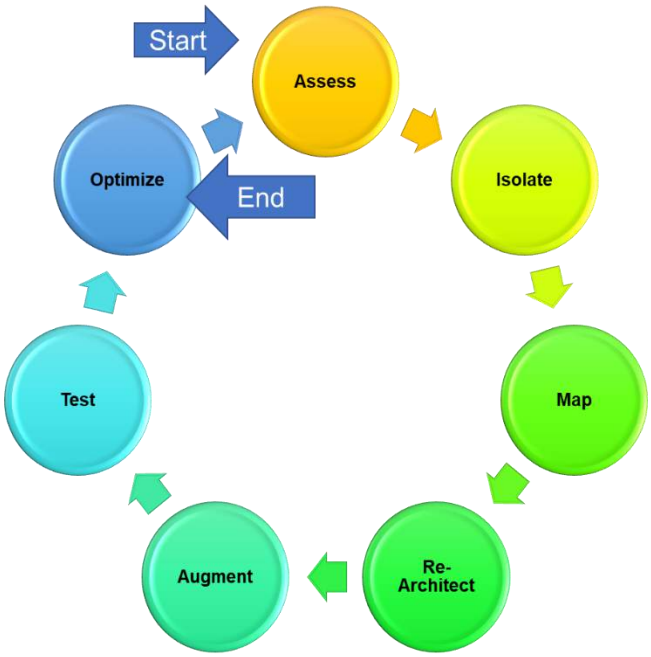
Broad Approaches to Migrating into the Cloud

- Cloud Economics deals with the economic rationale for leveraging the cloud and is central to the success of cloud-based enterprise usage. Decision-makers, IT managers, and software architects are faced with several dilemmas when planning for new Enterprise IT initiatives.

The Seven-Step Model of Migration into a Cloud

Typically, migration initiatives into the cloud are implemented in phases or in stages. A structured and process-oriented approach to migration into a cloud has several advantages of capturing within itself the best practices of many migration projects.

1. Conduct Cloud Migration Assessments
2. Isolate the Dependencies
3. Map the Messaging & Environment
4. Re-architect & Implement the lost Functionalities
5. Leverage Cloud Functionalities & Features
6. Test the Migration
7. Iterate and Optimize



Migration Risks and Mitigation

- The biggest challenge to any cloud migration project is how effectively the migration risks are identified and mitigated.
- In the Seven-Step Model of Migration into the Cloud, the process step of testing and validating includes efforts to identify the key migration risks.
- In the optimization step, we address various approaches to mitigate the identified migration risks.
- Migration risks for migrating into the cloud fall under two broad categories:
  - General migration risks
  - Security-related migration risks

- In the former we address several issues including:
  - Performance monitoring and tuning essentially identifying all possible production level deviants
  - The business continuity and disaster recovery in the world of cloud computing service
  - The compliance with standards and governance issues; the IP and licensing issues
  - The quality of service (QoS) parameters as well as the corresponding SLAs committed to
  - The ownership, transfer, and storage of data in the application; the portability and interoperability issues which could help mitigate potential vendor lock-ins
  - The issues that result in trivializing and non-comprehending the complexities of migration that results in migration failure and loss of senior management's business confidence in these efforts.

AWS Migration Best Practices

Leverage AWS Tools

AWS offers a wide range of tools designed for the migration process, from the initial planning phase to features for post-migration. Here are several useful tools to consider:

AWS Migration Hub

- A dashboard that centralizes data and helps you monitor and track the progress of migration.
- AWS Application Discovery - collects data needed for pre-migration due diligence.
- TSO Logic - offers data-driven recommendations based on predictive analytics. The recommendations are tailored to help during the planning and strategizing phase.

AWS Server Migration Service

- Provides automation, scheduling, and tracking capabilities for incremental migrations.
- AWS Database Migration Service - keeps the source data store fully-operational while the migration is in process, to minimize downtime.
- Amazon S3 Transfer Acceleration - improves the speed of data transfers made to Amazon S3, to maximize available bandwidth.

Automate Repetitive Tasks

- The migration process typically involves many repetitive tasks. You can perform these tasks manually, and you can automate them.
- The main purpose of automation is to enable you to achieve a higher level of efficiency while reducing costs. In many cases, automation can also help you complete tasks much faster than manually possible.

Outline and Share a Clear Cloud Governance Model

- A cloud governance model defines and specifies the practices, roles, responsibilities, tools, and procedures involved in the governance of your cloud environments.
- Your model needs to be as clear as possible, to ensure all relevant stakeholders understand how cloud resources should be managed and used. Ideally, you should define this information before migrating.
- Here are several questions your cloud governance model should answer:
  - What controls are set in place to meet security and privacy requirements?
  - How many AWS accounts are maintained?
  - What privileges are enabled for each role?

There are many more considerations to address in your cloud governance model, depending on your industry and business needs. Be sure to keep your documentation flexible to allow for change and optimization after the migration process is completed and your workloads settle in the new cloud environment.

Azure Migration Best Practices

Azure Migration Tools

Azure offers several migration tools designed to simplify and automate the migration process. Here are three commonly used Azure migration tools:

- **Azure Migrate** —helps you to assess your local workloads, determine the required size of cloud resources, and estimate cloud costs.
- **Microsoft Assessment and Planning**—helps you discover your servers and applications and build an inventory. Additionally, this tool can create reports that determine whether Azure can support your workloads.
- **Azure Database Migration Service**—helps you migrate on-premise SQL Server workloads to Azure.

Cost Management in Azure

Cloud resources are highly accessible and flexible, but costs can quickly skyrocket if you don't have a cost management strategy in place.

Here are several tools and techniques you can use to manage your cloud costs:

Tag your resources - to manage costs, you need visibility into cloud resource consumption. You can set this up by tagging resources and monitoring them. Be sure to use standard tags and keep this organized.

Use policies - to automate tagging and monitoring.

- Cloud resources are highly scalable and this can make manual tagging and monitoring incredibly time consuming.
- Use policies to standardize the process and automation to enforce these rules.
- You can leverage either third-party and first-party tools for tagging.
- There are also tools dedicated to cost management and optimization and monitoring. In addition, you can set up role-based access control (RBAC) to ensure resources are properly used by authorized users, and set up several resource groups.

Review Every Policy and Procedure

- Policies and procedures are a foundational component of the migration process and heavily impact the success of the implementation.
- To ensure your migration runs smoothly, you should define and review all policies and then apply them in a cohesive and standardized manner.
- Properly implementing security can ensure all required security measures are set in place. Policies are not only responsible for enforcing security, but also help you achieve and maintain compliance. Data encryption, for example, is a component you can enforce using a policy.
- Once you define your policies and procedures, you should test them before running in production.
- You can automate this process using several tools. Azure Migrate, for example, can help you automatically identify, assess, and migrate your local VMs to the Azure cloud.

Google Cloud Migration Best Practices

- Moving **Data**

Here are several aspects to consider when migrating to Google Cloud:

- **Move your data first** - and then move the rest of the application. This is recommended by Google.

- **Choose the relevant storage** - Google Cloud offers several tiers for hot and warm storage, as well as several archiving options. You can also leverage SSDs and hard disks, or choose a cloud-based database service, such as Bigtable, Datastore, and Google Cloud SQL.

- **Plan the data transfer process** - determine and define how to physically move your data. You can, for example, send your offline disk to a Google data center or opt to stream to persistent disks.

- Moving **Applications**

There are several ways to migrate applications, depending on the application's suitability to the cloud. In some cases, you might need to re-architect the entire application before it can be moved to the cloud. In other cases, you might need to do light modification before the migration. Ideally, when possible, your application can be lifted and shifted to the cloud.

A lift and shift migration means you do not need to make any changes to your application. You can lift it and move it directly to the new cloud environment. For example, you can create a local VM within your on-premise center, and then import it as a Google VM. Alternatively, you can back-up your application to GCP - this option lets you automatically create a cloud copy.

- Optimize

After the migration process is complete and your application is safely hosted in the cloud, you need to set up measures that help you continuously optimize your cloud environment. Here are several tools offered by Google:

- **Google Cloud operations suite (Stackdriver)** - provides features that enable full observability into your Google cloud environment. The information is centralized in a single database that lets you run queries and leverage root-cause analysis to gain detailed insights.

- **Google Cloud Pub/Sub** - helps you set up communication between any independent applications. You can use Pub/Sub to rapidly scale, decouple applications, and improve performance.

- **Google Cloud Deployment Manager** - lets you automate the configuration of your applications. You specify the requirements and Deployment Manager automatically initiates the deployments.

What Is a Cloud First Strategy?

- White House CIO Vivek Kundra coined the term “cloud-first”, referring to the practice of preferring the cloud as a first option for building programs and applications.
- A cloud-first strategy promotes building software directly in the cloud rather than building on-premises and migrating to the cloud. The goal is to help you create software faster and reduce the overhead associated with on-premises resources and cloud migration.

Why Should a Cloud-First Approach be Considered?

Here are key advantages of a cloud-first approach:

Flexibility- build your systems piece by piece according to business needs.

Less overhead- a cloud-first strategy lowers or eliminates the overhead associated with equipment and maintenance costs incurred when using on-premises server solutions.

More resources- cloud vendors provide access to additional services, which typically require lower or no initial investment.

Cost-effective upgrades- cloud vendors offer various pricing options you can leverage to reduce the costs of upgrades on-demand.

Support- cloud service providers offer support for their services, provided by experts.

Quick release- working directly in the cloud can help you achieve a faster speed of delivery for repairs, improvements, and updates.

Collaboration- cloud services often provide collaboration tools that enable you to work remotely, using numerous device types to access tools, storage, and data from any location.

Cloud First Challenges and Considerations

Cloud-First Security Challenges

- Many organizations continue to rely on legacy security protocols established in pre-cloud or sometimes pre-web times. These legacy systems are complex or sometimes impossible to implement successfully in the cloud.
- There are steps your organization can adopt to ensure your cloud-first strategy prioritizes cloud security. Central to these strategies is a focused DevSecOps approach, uniting development securities and operations into a collaborative team to improve testing and efficiency and reduce time-to-market.
- Here are steps you can adopt to secure critical data and resources when using a cloud-first approach:
  - **Foster organizational alignments**- protecting cloud native applications should be the shared responsibility of all project teams and departments.
  - **Secure the application lifecycle**- build security into the integration and deployment stages using practice, including vulnerability remediation and code scanning. You should also automatically apply runtime management with integrations.
  - **Limit privileges**- use a policy of least privilege of your most employees and users and only give access when necessary. This approach will reduce perimeter data leaks caused by human error.
  - **Deploy runtime protection**- next generation firewalls (NGFW) and web application firewalls (WAF) can help monitor request traffic and compare it to normal behavior to identify anomalies and block threats.

End-to-End Application Performance

- In recent years, the cloud has been approaching the edge. Certain use cases demand stringent measures regarding application performance, making it difficult for cloud-based solutions to meet latency demands for some critical applications.
- Storage-intensive applications responsible for processing hundreds of TBs of data every day are an example of performance limitations that affect the suitability of cloud-based solutions.

Vendor Lock-In

- Even when organizations gain effective control over cloud deployments, there are hidden costs of vendor lock-in. Enterprise-grade commercial agreements with cloud providers are rigid and difficult to change over time, as an organization's requirements change.
- While the market is heading in a good direction, customer protections in cloud agreements are not comparable to those offered by other IT outsourcing contracts. Without good commercial protection, organizations can unknowingly give away future flexibility.

Business Continuity and Disaster Recovery

- Even before the pandemic and the world's mass adoption of remote infrastructure, cloud-based solutions demanded that the industry rethink traditional BC/DR approaches. Cloud providers typically provide data solutions stored and backed up in several locations.

- Cloud-based failover protection is not guaranteed. For example, global-scale cyber-attacks can affect multiple cloud data centers, and locations with a high concentration of cloud data centers can be severely impacted by natural disaster, which can cause ripple effects worldwide.

How to Adopt a Cloud-First Strategy Approach

Learn from Your Peers

- A helpful step in creating a cloud-first strategy is learning from others' experiences. Look towards organizations that have effectively navigated the cloud migration process.
- You can ask questions about how they achieve their goals and their long-term aims for their solution.

Build a Cloud-First Culture

- The success of your organization's cloud-first strategy depends on cooperation from the top down. To make this possible, you will need to initiate a culture shift to the cloud-first approach—emphasizing transparency.
- Don't shy away from employees' apprehensions from the onset. Be approachable so that employees can come to you with questions before, during, and after implementation. Also, it helps employees understand how cloud migration will make their roles simpler.
- Many organizations approach a cloud-first culture shift through educational initiatives and employee engagement. For instance, an organization could create a cloud training program for technical and non-technical employees. Such a program could help employees understand how the technology works and the impact that it will have on their jobs.

Create a Cloud-First Migration Roadmap

- Like any major project, having a cloud migration plan is key. Create a roadmap specific to your organization that has all your solutions. Outline each step in your cloud migration approach.
- Establish a migration path for each application you have, from your most recent applications to your legacy applications: select private, public or hybrid cloud deployment.

7.3 Planning the migration and selecting a vendor

What is a Cloud Strategy Roadmap?

- A cloud strategy roadmap is a visual communication tool that describes how your organization will migrate to the cloud. It includes key tasks, deliverables, and deadlines.
- IT teams use roadmaps to put their cloud migration strategy on track and hold all stakeholders accountable.
- According to Gartner, cloud strategy roadmaps should have at least five parts: aligning objectives, planning, preparing for execution, governance, optimization, and collaboration.

- Because cloud migration projects are complex and involve multiple parts of an organization, developing a cloud strategy roadmap is not a simple task.
- You should follow a structured process to ensure all relevant stakeholders are on board and align your roadmap with available resources and operational considerations.

What Questions Should You Ask When Developing a Cloud Roadmap?

Addressing the "why"

The best way to start building your cloud roadmap is to start with the why. Why are you migrating? What are the benefits? Why should other members of your organization join your cloud vision?

Addressing the "how"

When you approach a cloud roadmap, you'll need concrete answers to the technical challenges of migration. Ask yourself how you'll migrate workloads to the cloud, how they will operate in a hybrid environment, and which cloud migration method is the most appropriate—lift-and-shift, refactoring, or rebuilding.

Addressing cultural factors

- Another category of questions involves the people in your organization. Don't underestimate the importance of culture. Technology is often the easier part of cloud transformation; changing workflows people are accustomed to is more challenging.
- How will you encourage people in your organization to cooperate with the migration, and what will you do to ensure it impacts them positively? Empathy, professional development, and support are as important as choosing between cloud providers or cloud-native technologies.

Addressing the "what"

- An essential part of your roadmap is what you will migrate.
- Ask yourself which workloads will move to the cloud, which are easier to migrate, and which are more challenging.
- Define datasets that will move to the cloud and critical aspects like data sensitivity and availability requirements.

Define success

Ask yourself what will make your cloud migration a success. Are you aiming to shut down the on-premises data center or move all new development to the cloud? Define the organization's ultimate goal with specific metrics to measure migration success.

Cloud Roadmap Strategy: Five Must-Have Stages

According to research by Gartner, the ideal cloud migration roadmap consists of five steps.

Align Objectives

- Organizations should create a cloud migration value proposition for business and IT early in the cloud migration roadmap. Start by conducting a survey to understand the use cases for cloud adoption, aligning cloud strategy with IT goals, and defining action steps to achieve your goals.
- Another important aspect is to define migration principles based on application and team readiness, business priorities, and vendor capabilities. Use data available in the organization to define the metrics and key performance indicators (KPIs) for a successful migration.

Develop a Plan of Action

- Choose the right cloud provider and negotiate a successful contract. At this stage, you should build cloud capabilities across the organization, assess alternative service providers, and prepare to mitigate cloud-related risks. Identify the necessary investments in your network, security, identity architecture, and other tools.
- At this stage, determine whether to migrate your entire environment to the cloud, or one workload at a time. Identify if your organization requires a multi-cloud environment or a single cloud provider will suffice. Think about the long term—will the capabilities of your cloud provider fit your needs in the future, and how will costs grow over time given your future growth?

Prepare for Execution

- At this stage, you deploy and optimize workloads in the cloud. Deployment involves identifying workloads for migration, defining your cloud management workflow, adopting implementation best practices, and analyzing how workloads perform in the cloud.
- Managing cloud migration as a structured, well-defined process can help an organization significantly improve the efficiency and effectiveness of cloud workload management.

Establish Governance While Mitigating Risk

- The goal of a successful cloud migration includes setting up robust processes to minimize disruption to your workflow. To be successful, you must discover, analyze and monitor sensitive data throughout your cloud deployment. Set up a security control plane using a third-party tool with suitable functionality.
- Take a lifecycle approach to governance, it is important to realize that you must continuously maintain governance to be effective. Governance and compliance feedback should be an integral part of your workflow, leveraging automation.

Optimize and Scale

- At this stage, workloads are already running successfully in the cloud. Consider investments that can improve existing use of the cloud and address operational challenges. Define customer-centric goals, communicating to teams how improved cloud use can benefit the organization. Align all stakeholders around the need to continuously develop and optimize your cloud presence.

Collaborate

Cloud migration processes can only succeed by achieving cooperation between cross-departmental teams. The following roles should be included in your roadmap and in relevant planning stages:

- **CIO**—provides strategic and planning guidance and can help define the goals of cloud migration. The CIO can help communicate progress to other stakeholders.

- **Development leaders and teams**—provide technical advice and can help establish a vision. They can work with other IT leaders to define specific cloud migration plans using up-to-date progress and planning information.

- **Operations leaders and teams**—provide insight into the infrastructure and operations requirements of cloud migration and determine activities required to implement the strategy. They will typically manage the operational mechanisms needed to enable the migration.

- **Cloud experts**—any cloud migration program will benefit from a team of cloud experts, either in-house or outsourced, who can provide architectural and process plans for the project. They can help evaluate and select the best tools and processes for migrating and refactoring systems and help build the required skills among other teams.

Cloud Migration Strategies

Gartner has identified five cloud migration techniques, known as the “5 Rs”. Organizations looking to migrate to the cloud should consider which migration strategy best answers their needs. The following is a brief description of each:

Rehost
Rehosting, or 'lift and shift,' involves using infrastructure-as-a-service (IaaS). You simply redeploy your existing data and applications on the cloud server. This is easy to do and is thus suited for organizations less familiar with cloud environments. It is also a good option for cases where it is difficult to modify the code, and you want to migrate your applications intact.

Refactor
Refactoring, or 'lift, tinker, and shift,' is when you tweak and optimize your applications for the cloud. In this case, a platform-as-a-service (PaaS) model is employed. The core architecture of the applications remains unchanged, but adjustments are made to enable the better use of cloud-based tools.

Revise
Revising builds upon the previous strategies, requiring more significant changes to the architecture and code of the systems being moved to the cloud. This is done to enable applications to take full advantage of the services available in the cloud, which may require introducing major code changes. This strategy requires foreplaning and advanced knowledge.
Rebuild. Rebuilding takes the Revise approach even further by discarding the existing code base and replacing it with a new one. This process takes a lot of time and is only considered when companies decide that their existing solutions don't meet current business needs.

Replace
Replacing is another solution to the challenges that inform the Rebuild approach. The difference here is that the company doesn't redevelop its own native application from scratch. This involves migrating to a third-party, prebuilt application provided by the vendor. The only thing that you migrate from your existing application is the data, while everything else about the system is new.

Cloud Migration Strategic Process
The cloud migration steps or processes an enterprise follows will vary based on factors such as the type of migration it wants to perform and the specific resources it wants to move. That said, common elements of a cloud migration strategy include the following:

- Evaluation of performance and security requirements
- Selection of a cloud provider
- Calculation of costs
- Any reorganization deemed necessary

At the same time, be prepared to address several common challenges during a cloud migration:

- Interoperability
- Data and application portability
- Data integrity and security
- Business continuity

Without proper planning, a migration could degrade workload performance and lead to higher IT costs -- thereby negating some of the main benefits of cloud computing.

A 4-Step Cloud Migration Process

1. Cloud Migration Planning

- One of the first steps to consider before migrating data to the cloud is to determine the use case that the public cloud will serve. Will it be used for disaster recovery? DevOps? Hosting enterprise workloads by completely shifting to the cloud? Or will a hybrid approach work best for your deployment.
- In this stage it is important to assess your environment and determine the factors that will govern the migration, such as critical application data, legacy data, and application interoperability.
- It is also necessary to determine your reliance on data: do you have data that needs to be resynced regularly, data compliance requirements to meet, or non-critical data that can possibly be migrated during the first few passes of the migration?
- Determining these requirements will help you charter a solid plan for the tools you'll need during migration, identifying which data needs to be migrated and when, if the data needs any scrubbing, the kind of destination volumes to use, and whether you'll need encryption of the data both at rest and in transit.

2. Migration Business Case

- Once you have determined your business requirements, understand the relevant services offered by cloud providers and other partners and their costs.
- Determine the expected benefits of cloud migration along three dimensions: operational benefits, cost savings, and architectural improvements.
- Build a business case for every application you plan to migrate to the cloud, showing an expected total cost of ownership (TCO) on the cloud, compared to current TCO.
- Use cloud cost calculators to estimate future cloud costs, using realistic assumptions - including the amount and nature of storage used, computing resources, taking into account instance types, operating systems, and specific performance and networking requirements.
- Work with cloud providers to understand the options for cost savings, given your proposed cloud deployment.
- Cloud providers offer multiple pricing models, and provide deep discounts in exchange for long-term commitment to cloud resources (reserved instances) or a commitment to a certain level of cloud spend (savings plans). These discounts must be factored into your business plan, to understand the true long-term cost of your cloud migration.

3. Cloud Data Migration Execution

- Once your environment has been assessed and a plan has been mapped out, it's necessary to execute your migration. The main challenge here is carrying out your migration with minimal disruption to normal operation, at the lowest cost, and over the shortest period of time.
- If your data becomes inaccessible to users during a migration, you risk impacting your business operations. The same is true as you continue to sync and update your systems after the initial migration takes place. Every workload element individually migrated should be proven to work in the new environment before migrating another element.

- You'll also need to find a way to synchronize changes that are made to the source data while the migration is ongoing. Both AWS and Azure provide built-in tools that aid in AWS cloud migration and in Azure data migration, and later in this article we'll see how NetApp users benefit from migrating with services and features that come with Cloud Volumes ONTAP.

4. Ongoing Upkeep

- Once that data has been migrated to the cloud, it is important to ensure that it is optimized, secure, and easily retrievable moving forward. It also helps to monitor for real-time changes to critical infrastructure and predict workload contentions.
- Apart from real-time monitoring, you should also assess the security of the data at rest to ensure that working in your new environment meets regulatory compliance laws such as HIPAA and GDPR.
- Another consideration to keep in mind is meeting ongoing performance and availability benchmarks to ensure your RPO and RTO objectives should they change.

Cloud migration deployment models

- Enterprises today have more than one cloud scenario from which to choose:
  - The public cloud lets many users access compute resources through the internet or dedicated connections.
  - A private cloud keeps data within the data center and uses a proprietary architecture.
  - The hybrid cloud model mixes public and private cloud models and transfers data between the two.
  - In a multi-cloud scenario, a business uses IaaS options from more than one public cloud provider.
- As you consider where the application should live, consider how well it will perform once it's migrated. Ensure there is adequate bandwidth for optimal application performance. Also, determine whether an application's dependencies may complicate a migration.
- Review what's in the stack of the application that will make the move.
- Local applications may contain a lot of features that go unused, and it is wasteful to pay to migrate and support those nonessential items.
- Stale data is another concern with cloud migration. Without a good reason, it's probably unwise to move historical data to the cloud, which typically incurs costs for retrieval.
- As you examine the application, it may be prudent to reconsider its strategic architecture to set it up for what could potentially be a longer life.
- A handful of platforms support hybrid and multi-cloud environments, including the following:
  - Microsoft Azure Stack;
  - Google Cloud Anthos;
  - AWS Outposts;
  - VMware Cloud on AWS; and
  - a container-based PaaS, such as Cloud Foundry or Red Hat OpenShift.

Best Practices to ensure Cloud Migration Success

- There are many reasons why an organization chooses to migrate an app or workload to the cloud, and each project will be unique depending on resource allocations, integrations with other services and multiple other factors.
- Here are some general guidelines for a cloud migration that streamline the process and improve changes for success:

Get organizational buy-in
The transition is much smoother when all stakeholders are on board and know their roles, from management to technical practitioners to end users.

Define cloud roles and ownership
Determine right upfront who is responsible to manage various aspects of the cloud workload. Is it a shared environment? How is identity confirmed and access granted, or limited? This includes proper documentation of setups and processes.

Pick the right cloud services
Cloud providers have a vast menu of services to pick from. Be clear with which ones your workload will tap into, or you risk running extraneous services some of which may be interdependent and become problematic to manage.

Understand security risks
Cloud environments can be susceptible to mischief from internet attacks. Misconfigurations are arguably a bigger problem, given the complexity of cloud environments.

Calculate cloud costs
The cloud's pay-as-you-go model may seem attractive and simpler to organizations used to large infrastructure investments. But it's a double-edged sword: Pay close attention to service selections and usage, or you'll get a shock at the end of the month.

Devise a long-term cloud roadmap
If a cloud migration is successful, organizations likely will look to replicate that success for other workloads. Identify the criteria to follow, from project timelines to different deployment options, such as a hybrid cloud setup.

Cloud Migration Tools and Services

The big IaaS providers -- AWS, Microsoft and Google -- offer various cloud migration services as well as free tiers. Here are a few examples:

	AWS	Azure	Google Cloud
Database migration	AWS Database Migration Service	Azure Database Migration Service	Database Migration Service (preview)
Data transfer appliance	Snow Family	Data Box	Transfer Appliance
Disaster recovery	CloudEndure Disaster Recovery	Azure Site Recovery	N/A
Online data transfer	AWS DataSync, AWS Transfer Family	Azure File Sync	BigQuery Data Transfer Service, Cloud Data Transfer
On-premises application analysis	AWS Application Discovery Service, Migration Evaluator	Azure Migrate, Movere, Azure Resource Mover	N/A

	AWS	Azure	Google Cloud
On-premises and cloud storage integration	Storage Gateway	StorSimple	N/A (offered by partner Cloudian)
Migration tracker	AWS Migration Hub	Azure Migrate	N/A
Server migration	AWS App2Container, AWS Server Migration Service, CloudEndure Migration	Azure Migrate	Migrate for Anthos, Migrate for Compute Engine, VM migration

Ready to migrate to the cloud? Answer these questions

Ready to migrate to the cloud? Answer these questions

Cloud computing ultimately frees an enterprise IT team from the burden of managing uptime. Placing an application in the cloud is often the most logical step for growth. A positive answer to some or all of these questions may indicate your company's readiness to move an app to the cloud.

Should your application stay or go?

Legacy applications, or workloads that require low latency or higher security and control, probably should stay on premises or move to a private cloud.

What's the cost to run an application in the cloud?

One of the primary benefits of a cloud migration is workload flexibility. If a workload suddenly needs more resources to maintain performance, its cost to run may escalate quickly.

Which cloud model fits best?

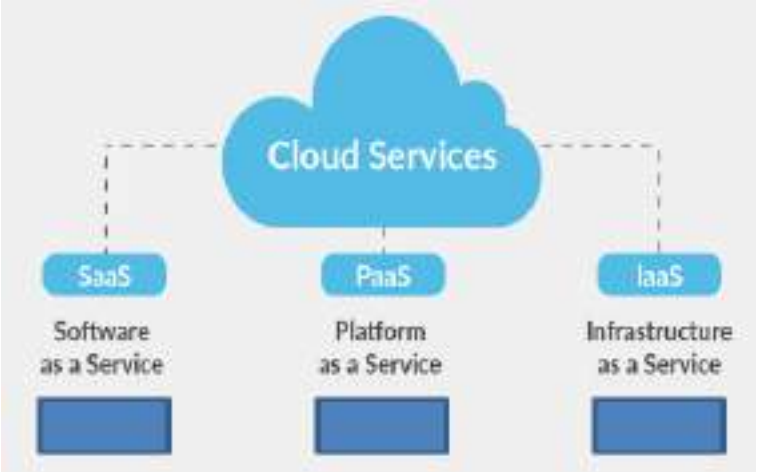
Public cloud provides scalability through a pay-per-usage model. Private or on-premises cloud provides extra control and security. A hybrid cloud model provides the best of both, although performance and connectivity may suffer.

How do I choose the right cloud provider?

The top three cloud providers -- AWS, Microsoft and Google -- generally offer comparable services to run all kinds of workloads in the cloud, as well as tools to help you efficiently move apps there. Gauge your specific needs for availability, support, security and compliance, and pricing to find the best fit.

Section 3: Exercises

Exercise 1: Write down the Purpose of Use in Single word for Different Cloud Services in below Diagram.



Exercise 2: Write Down the Main Benefits of Cloud Migration in below.

Exercise 3: Participate in group discussion on following topics:

a) Technical Consideration for Cloud Migration
b) Cloud Migration
c) Planning the migration and selecting a vendor

Section 4: Assessment Questionnaire

1. Explain Cloud migration?
2. List few challenges while migrating to cloud?
3. List some advantages and disadvantages of Cloud Migration?
4. What are the tools for cloud migration services?

----------End of the Module----------

MODULE 8

BASICS OF AMAZON WEB SERVICES (AWS)

Section 1: Learning Outcomes

After completing this module, you will be able to:

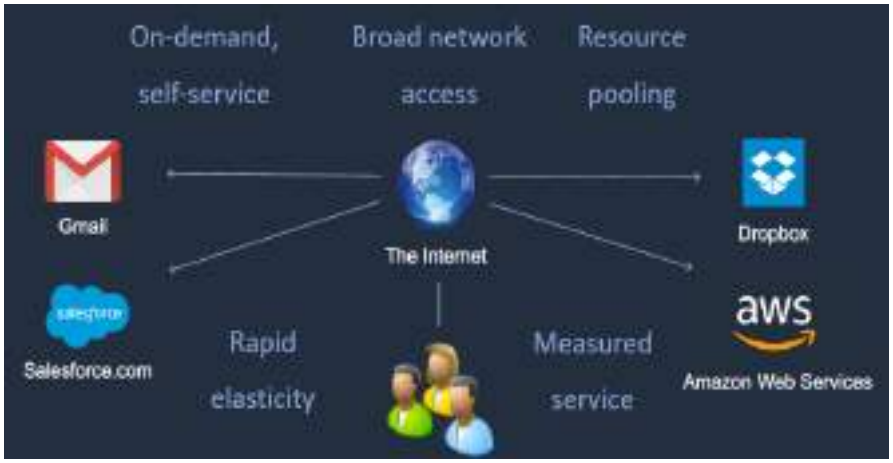
- Ø Explain functions of AWS (AMAZON WEB SERVICES)
- Ø Give examples and benefits of Cloud Computing
- Ø Tell types of Cloud Service and Deployment
- Ø Draw AWS Global Infrastructure
- Ø Explain AWS Shared Responsibility Model
- Ø Describe Application Programming Interfaces (APIs)
- Ø Launch Cloud Services
- Ø Describe the Identity and Access Management (AWS IAM)
- Ø Explain AWS Compute Services
- Ø Describe Server Virtualization
- Ø Work on Amazon Elastic Compute Cloud (EC2), Amazon Elastic Container Service (ECS) and Amazon Elastic Block Store (EBS)
- Ø Explain functions of Amazon Machine Images (AMI), Amazon Elastic File System (EFS) and Amazon Simple Storage Service (S3)
- Ø Use AWS Lambda Functions
- Ø Execute Amazon Step Functions and Services
- Ø Explain Amazon EventBridge / CloudWatch Events and API Gateway
- Ø Describe the concept of Virtual Private Cloud (VPC), Security Groups and Network ACLs
- Ø Work with IP Addresses
- Ø Use Amazon VPN, Direct Connect, Gateway and Outposts
- Ø Adopt CloudFront, Global Accelerator and Cloud Formation techniques
- Ø Work on AWS Cloud Development Kit and Elastic Beanstalk
- Ø Use AWS Developer Tools (Code*), AWS X-Ray and OpsWorks
- Ø Differentiate between various types of databases
- Ø Describe the concepts of Amazon Aurora, Dynamo DB, Redshift, Elastic Map Reduce and ElastiCache

Section 2: Relevant Knowledge

8.1 Cloud Computing & AWS (AMAZON WEB SERVICES)

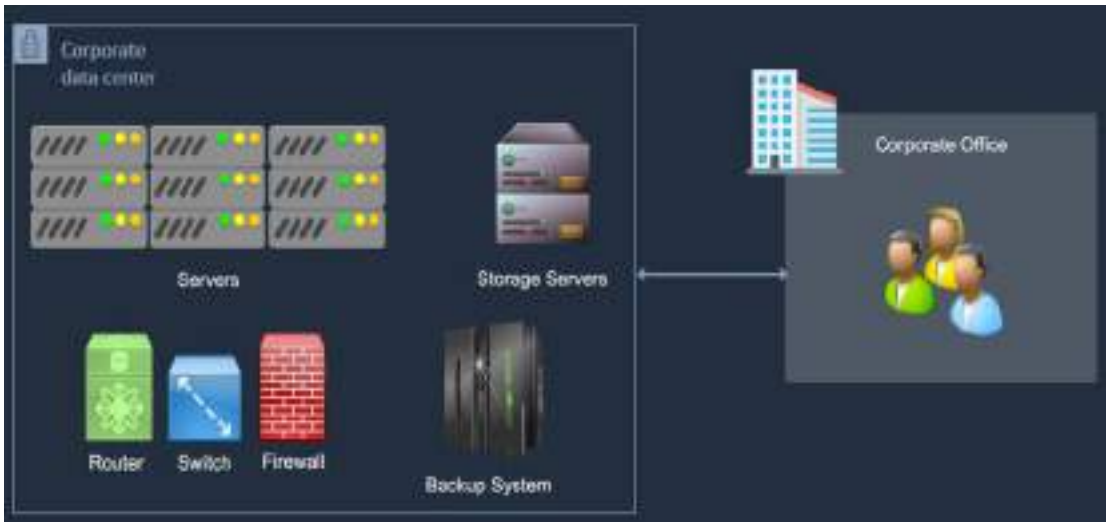
Traditional IT and Cloud Computing

- The term cloud is the articulation for the Internet, and this is the greatest distinction between cloud computing vs traditional computing.
- Cloud computing runs on outsider servers facilitated by third-party hosting organizations, while traditional computing happens on website servers and physical hard drives. Organizations can get to these servers on the web.



- The IT equipment is owned by the company.
- A company leases space in a data center or may own the whole building.

- IT staff must design, build, operate and manage equipment.



Expenses:

- By and large, traditional computing costs are higher than cloud computing costs.
- This is chiefly since the maintenance and operation of the server is divided among a few distinct gatherings, which lessens the expense of public services.
- Organizations can save money on investment costs by not accepting costly equipment.
- Costs:
  - Data Center Building
  - Data Center Security
  - Physical IT Hardware
  - Software Licensing Costs
  - Maintenance Contracts
  - Power
  - Internet Connectivity
  - Staff Wages

Architecture:

- For any computing, architecture is an imperative viewpoint as it encourages you to comprehend the application's design.
- Cloud-based applications are made for infrastructure development.
- They work on theories of automation and user interface, while Traditional applications are made on three essential levels known as:
  - App logic tier
  - Presentation tier
  - Database tier

Operating system dependency:

- Operating system reliance is a critical perspective on which a cloud-based application and traditional computing can be recognized.
- Cloud-based applications are autonomous, as cloud computing technology is inescapable; on the other hand, a Traditional application is consistently reliant on a particular operating system for working appropriately.
- It is likewise dependent on hardware, storage, and backing services.

The convenience of collaboration:

- As of now, an efficient and easy coordinated effort is crucial to maintain a business on the advanced stage.
- Cloud-based computing permits simple coordinated effort, and the developers can finish codes efficiently.

- Since cloud computing is service-oriented, you can guarantee appropriate creation in your organization. Besides, the business area has turned data-centric. In such scenery, traditional computing requires completed codes and regularly leads to an inside clash in an organization.

Security:

- Security is one of the crucial requirements to maintain a business appropriately.
- Both traditional computing and cloud computing have distributable highlights regarding security.
- You can have numerous layers of security while utilizing cloud-based computing.
- Cyberattacks are uncommon on account of cloud-based computing because of the presence of various hosts.
- On the off chance that you intend to start a business and extend it, the assistance situated cloud computing can be your best support network, while traditional computing is made statically, it just gives a solitary security layer to the business-related data.

Backup and Recovery:

- Cloud-based computing has all around planned design that guarantees legitimate backup for all the data.
- Also, the DRaaS application can help you access the backups on the off chance that it gets erased out of sudden, on the opposite side, with regards to recovery and backup, there is no computerized highlight present in traditional computing.
- Nor is there a disaster recovery administration. Thus, as a business person, you can run over a few issues while utilizing traditional computing.

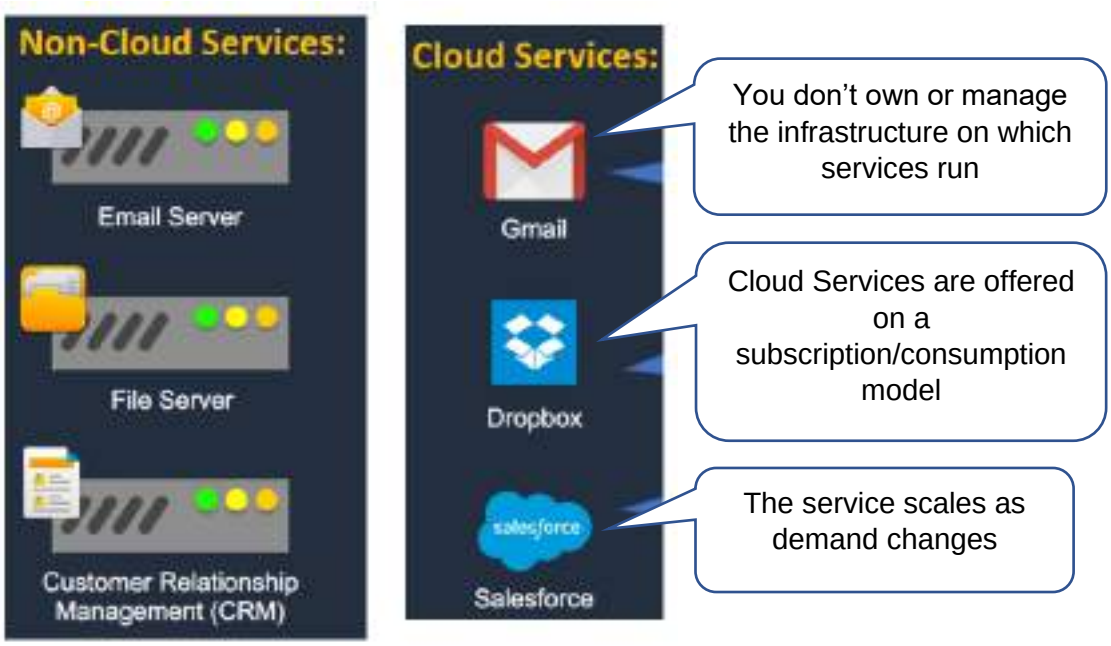
Availability:

- Cloud-based computing is unique. On account of this computing, you can get customary updates and improved highlights. Therefore, maintaining your business will turn out to be substantially more reasonable. Indeed, even if there should arise an occurrence of certain escape clauses, the IT group works dedicatedly to quick eradication.
- This sensibility helps your business run at a decent speed, and you can procure a decent benefit, while the IT heads discharge customary applications over long stretches, frequently half a months or weeks.
- It happens as the traditional applications need manual scripting. Also, it can’t be delivered except if all the parts of coding are finished.

Cloud Computing	Traditional IT
On-demand, self-service	Requires human involvement
Broad network access	Internal accessibility, limited public presence
Resource pooling	Single-tenant, can be virtualized
Rapid elasticity	Limited scalability
Measured service	Usage is not typically measured

8.2 Examples and Benefits of Cloud Computing

Examples of Cloud Computing



Deploying a Website On-Premises

Activity	Timeline
Purchase Hardware	4-12 Weeks
Install and Build	4-8 Weeks
Acceptance Testing	2-4 Weeks
Handover to operations	1-2 Weeks

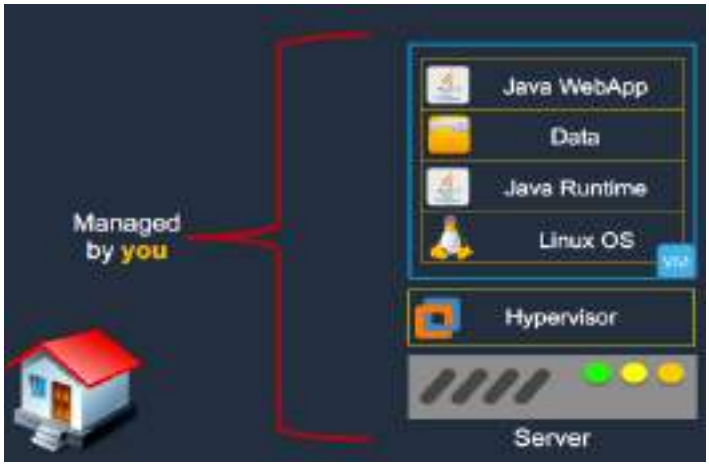
Deploying a Website in the Cloud

- Customers connect over the internet to place orders.
- Admin uses a browser or command line to deploy website and database

8.3 Types of Cloud Service and Deployment

Cloud Service Models: Private Cloud

- A private cloud must also include self-service, multi-tenancy, meeting and elasticity



Cloud Service Models: Infrastructure as a Service (IaaS)

Examples:

- Amazon Elastic Compute Cloud (EC2)
- Azure Virtual Machine
- Google Compute Engine

Cloud Service Models: Platform as a Service (PaaS)

Examples:

- AWS Elastic Beanstalk
- Azure WebApps
- Compute App Engine

Cloud Service Models: Software as a Service (SaaS)

Examples:

- Google Apps
- Salesforce.com
- Zoom

Cloud Service Models: Comparison

- **Private Cloud:** You manage everything – greater responsibility + greater control
- **IaaS:** You manage from the virtual server upwards
- **PaaS:** You simply upload your code/data to create your application
- **SaaS:** You simply consume the service – little responsibility + little control

Private Cloud

Benefits

- Compare control of the entire stack
- Security – in a few cases, organization may need to keep all or some of their applications and data in house.

Public Cloud

Examples

- AWS
- Microsoft Azure
- Google Cloud Platform

Benefits

- Variable Expense instead of capital expense
- Economies of Scale
- Massive Elasticity

Hybrid Cloud

Benefits

- Allows companies to keep the critical applications and sensitive data in a traditional data center environment or private cloud.
- Take advantage of the public cloud resources like SaaS, for the latest applications and IaaS for elastic virtual resources.
- Facilitates probability of data, apps, and services and more choices for deployment models.

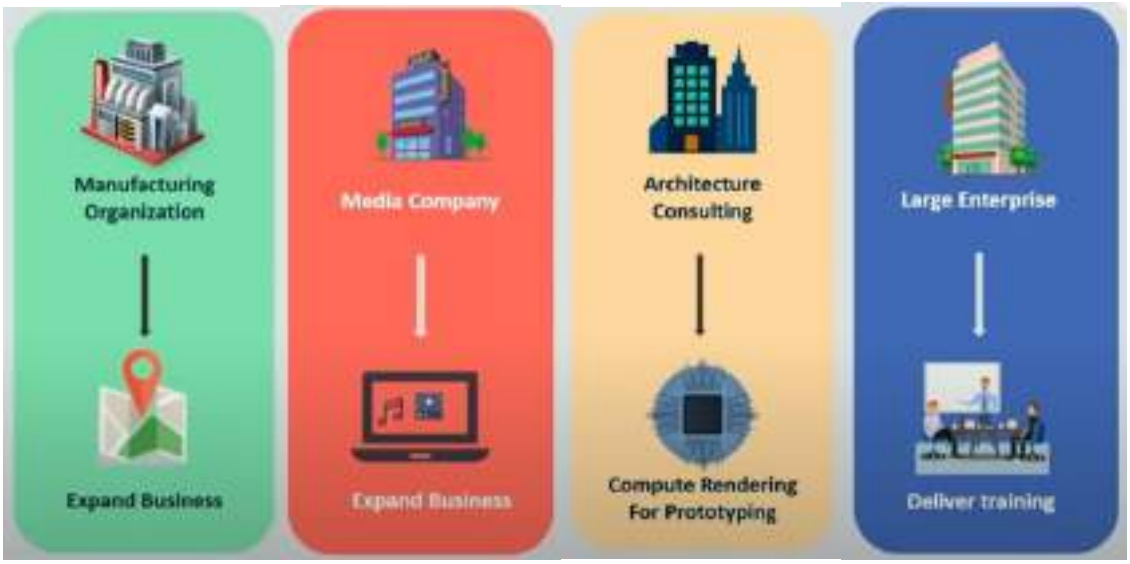
Multi-cloud

8.4 Overview of Amazon Web Services (AWS)

What is AWS?

Amazon Web Service (AWS) is a secure cloud services, platform, offering compute power, data storage, content delivery and other functionality to help business scale and grow.

Use Cases



Advantages of AWS

Easy to use

AWS is designed to allow application providers, ISVs, and vendors to quickly and securely host your applications – whether an existing application or a new SaaS-based application.

Flexible

AWS enables you to select the operating system, programming language, web application platform, database, and other services you need.

Cost-Effective

You pay only for the compute power, storage, and other resources you use, with no long-term contracts or up-front commitments.

Reliable

With AWS, you take advantage of a scalable, reliable, and secure global computing infrastructure, the virtual backbone of Amazon.com's multi-billion-dollar online business that has been honed for over a decade.

Scalable and high-performance

Using AWS tools, Auto Scaling, and Elastic Load Balancing, your application can scale up or down based on demand.

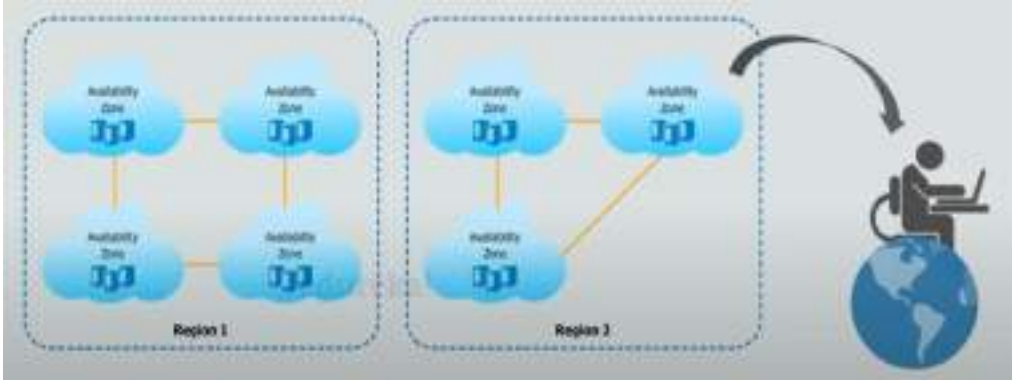
Secure

AWS utilizes an end-to-end approach to secure and harden our infrastructure, including physical, operational, and software measures.

AWS Architecture

Amazon Infrastructure is divided into following categories:

- Regions
- Availability Zones



Sign In Process

How to sign in to the AWS Management Console?

Signing in as the AWS account root user

- If you're a root user, open the https://signin.aws.amazon.com/, select Root user, and sign in using your AWS account root user credentials.

Signing in as the AWS Identity and Access Management (IAM) user with a custom URL

- Sign in using a custom URL https://account_alias_or_id.signin.aws.amazon.com/console/
- You must replace account_alias_or_id with the account alias or account ID provided by the root user.

Signing in as the IAM user on the Sign-in page

- If you have previously signed in as the IAM user on the browser, you might see the Sign in as IAM user page when you open the https://signin.aws.amazon.com.
- Your account ID or account alias might already be saved. In that case, just enter your IAM user credentials, and then choose Sign in.
- If you are signing in on the browser for the first time, open the https://signin.aws.amazon.com, select IAM user, and then enter the 12-digit AWS account ID or account alias.
- Choose Next.
- In the Sign in as IAM user page, enter your IAM user credentials, and then choose Sign in.

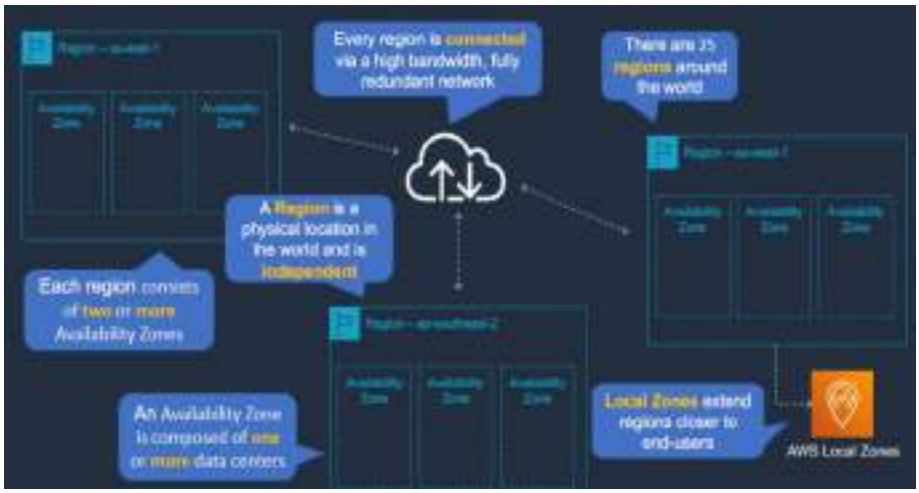
AWS Service Categories (a few examples)

AWS Pricing Fundamentals

- There are three fundamental drivers of cost with AWS: compute, storage, and outbound data transfer.
- These characteristics vary somewhat, depending on the AWS product and pricing model you choose.

8.5 The AWS Global Infrastructure

- The AWS Global Cloud Infrastructure is the most secure, extensive, and reliable cloud platform, offering over 200 fully featured services from data centers globally.
- Whether you need to deploy your application workloads across the globe in a single click, or you want to build and deploy specific applications closer to your end-users with single-digit millisecond latency, AWS provides you the cloud infrastructure where and when you need it.
- With millions of active customers and tens of thousands of partners globally, AWS has the largest and most dynamic ecosystem.
- Customers across virtually every industry and of every size, including start-ups, enterprises, and public sector organizations, are running every imaginable use case on AWS.

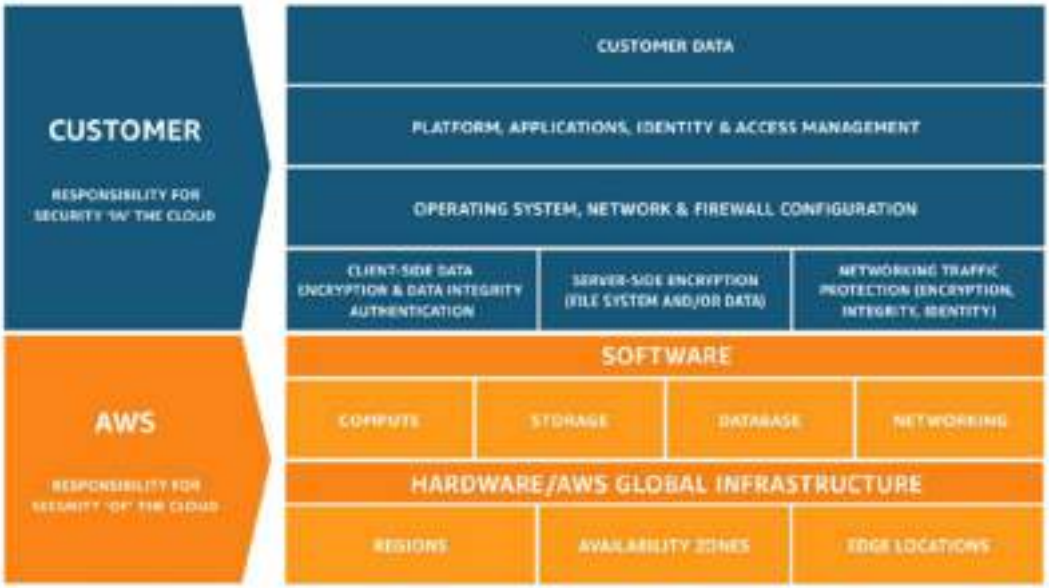


Deploying Services Globally

- As of Aug-2022, The AWS Cloud spans 84 Availability Zones within 26 geographic regions around the world, with announced plans for 24 more Availability Zones and 8 more AWS Regions in:
  - Australia
  - India
  - Indonesia
  - Israel
  - New Zealand
  - Spain
  - Switzerland
  - United Arab Emirates (UAE)

8.6 The AWS Shared Responsibility Model

The AWS Shared Responsibility Model



- AWS responsibility “Security of the Cloud” - AWS is responsible for protecting the infrastructure that runs all of the services offered in the AWS Cloud.
- This infrastructure is composed of the hardware, software, networking, and facilities that run AWS Cloud services.
- A shared responsibility model is a cloud security framework that dictates the security obligations of a cloud computing provider and its users to ensure accountability.

8.7 Application Programming Interfaces (APIs)

Application Programming Interfaces (APIs) – Building API a house analogy

- Builder provides set of standard and options

- The Builder gives instructions to the workers in a language they understand

Application Programming API Interfaces (APIs)

- The API provides the instructions developers use in their code
- Instructions are sent to the API using the HTTP protocol

Flight API Aggregator Example

8.8 Launching Cloud Services

Launching Cloud Services: Management Console

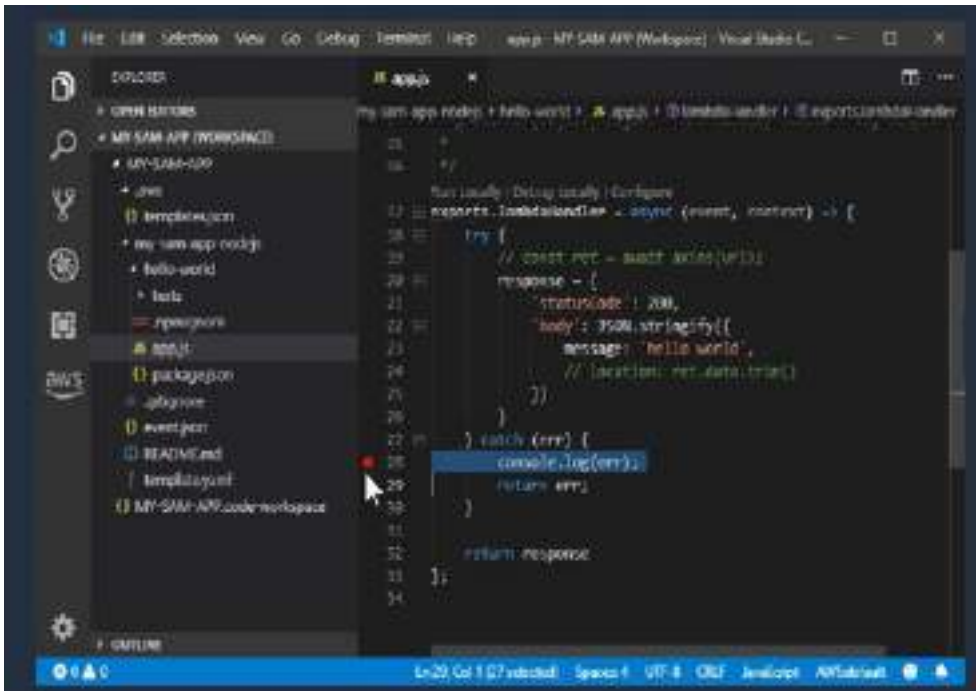
The AWS Management Console is a web application that comprises and refers to a broad collection of service consoles for managing AWS resources.

Launching Cloud Services: Command Line

- The AWS Command Line Interface (AWS CLI) is a unified tool to manage your AWS services.
- With just one tool to download and configure, you can control multiple AWS services from the command line and automate them through scripts.

Launching Cloud Services: Software Development Kit

- A developer writes the code in an integrated development environment (IDE).
- The code leverages the SDK to work with cloud services.



AWS Public and Private Services

- The instances in the public subnet can send outbound traffic directly to the internet, whereas the instances in the private subnet can't.
- Instead, the instances in the private subnet can access the internet by using a network address translation (NAT) gateway that resides in the public subnet.

8.9 The Advantages of Cloud Computing

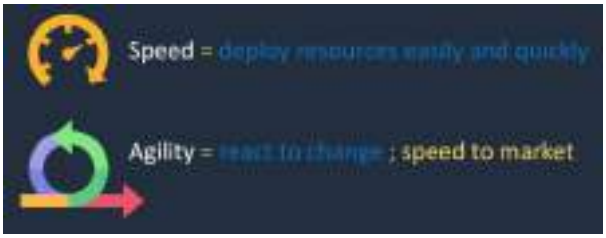
1. Trade Capital expense for variable expense

2. Benefit from massive economies of scale

Aggregated usage across hundreds of thousands of customers = lower variable costs for customers

3. Stop guessing capacity

4. Increase speed and agility



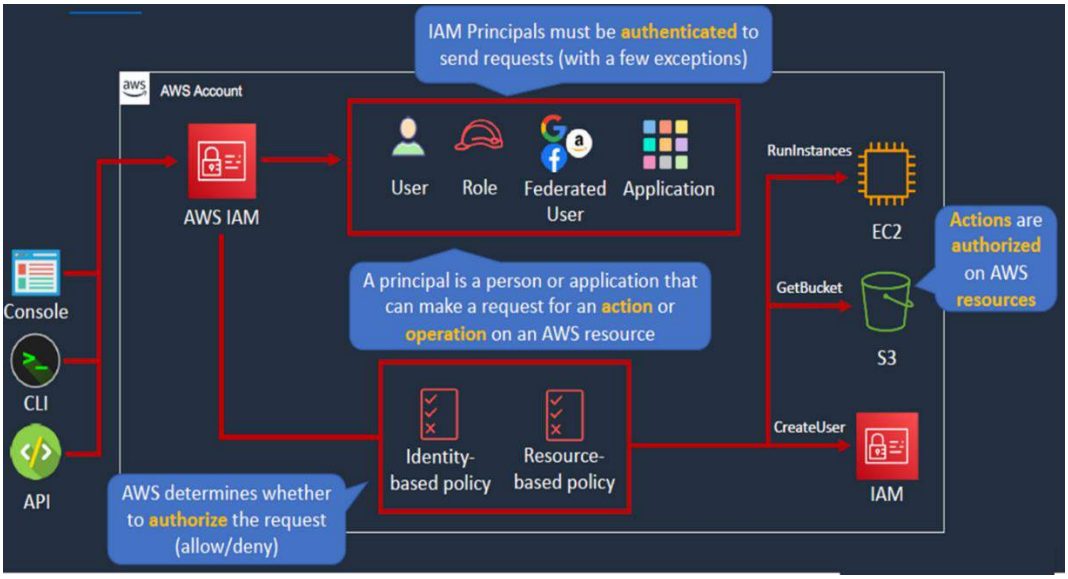
5. Stop spending money running and maintaining data centers

6. Go global in minutes

8.10 Identity and Access Management (AWS IAM)

AWS Identity and Access Management (IAM)

- With AWS Identity and Access Management (IAM), you can specify who or what can access services and resources in AWS, centrally manage fine-grained permissions, and analyze access to refine permissions across AWS.

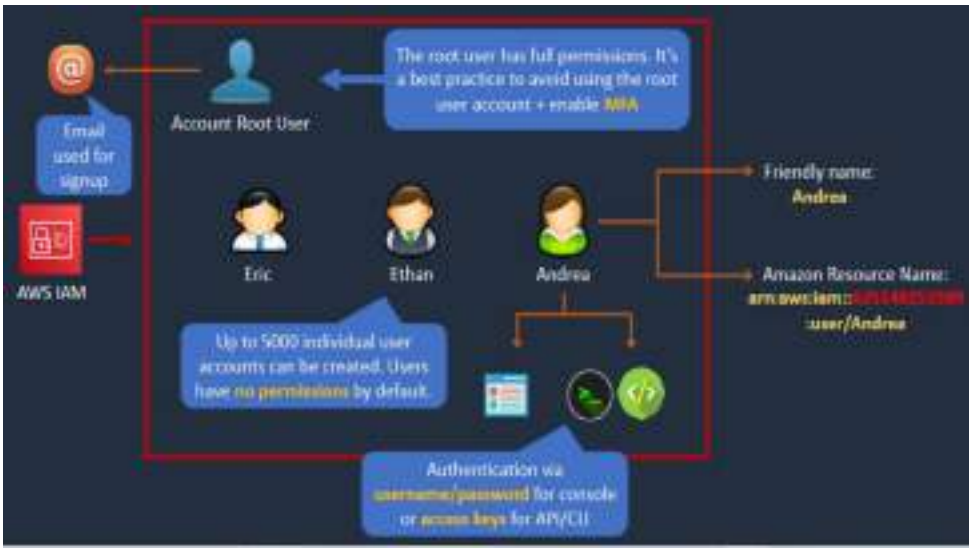


Users, Groups, Roles and Policies

AWS account root user

- When you first create an Amazon Web Services (AWS) account, you begin with a single sign-in identity that has complete access to all AWS services and resources in the account.
- This identity is called the AWS account root user and is accessed by signing in with the email address and password that you used to create the account.

IAM Users



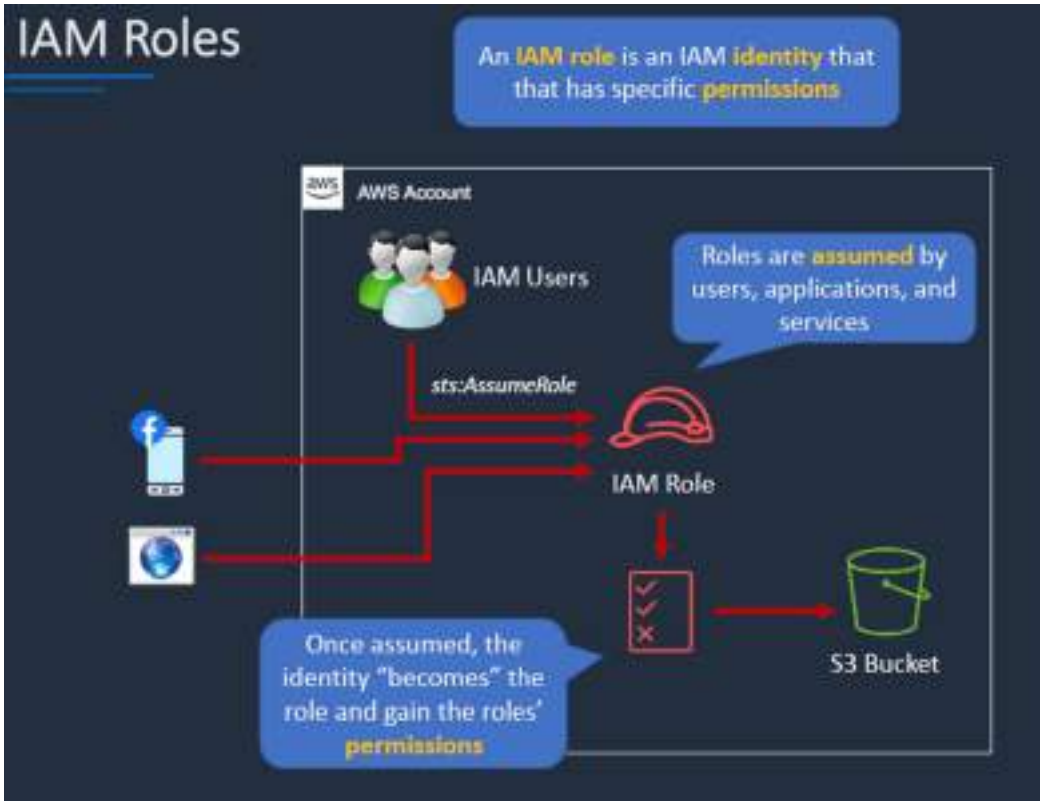
- An IAM user is an entity that you create in AWS.
- The IAM user represents the person or service who uses the IAM user to interact with AWS.
- A primary use for IAM users is to give people the ability to sign in to the AWS Management Console for interactive tasks and to make programmatic requests to AWS services using the API or CLI.
- A user in AWS consists of a name, a password to sign into the AWS Management Console, and up to two access keys that can be used with the API or CLI.
- When you create an IAM user, you grant it permissions by making it a member of a user group that has appropriate permission policies attached (recommended), or by directly attaching policies to the user.
- You can also clone the permissions of an existing IAM user, which automatically makes the new user a member of the same user groups and attaches all the same policies.

IAM Groups

- An IAM user group is a collection of IAM users.
- You can use user groups to specify permissions for a collection of users, which can make those permissions easier to manage for those users.

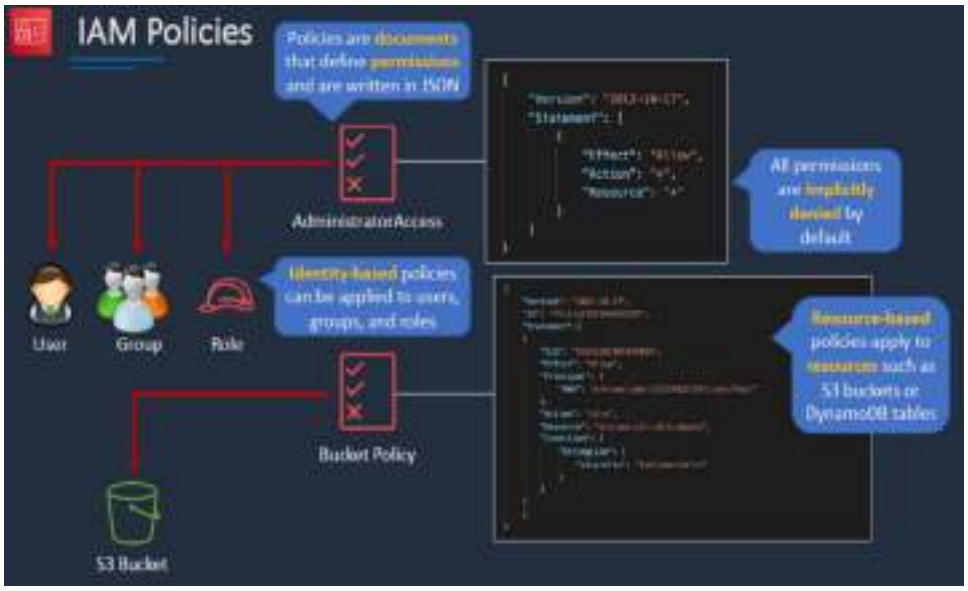
- For example, you could have a user group called Admins and give that user group the types of permissions that administrators typically need. Any user in that user group automatically has the permissions that are assigned to the user group.
- If a new user joins your organization and should have administrator privileges, you can assign the appropriate permissions by adding the user to that user group.
- Similarly, if a person changes jobs in your organization, instead of editing that user's permissions, you can remove him or her from the old user groups and add him or her to the appropriate new user groups.
- A user group cannot be identified as a principal in a resource-based policy.
- A user group is a way to attach policies to multiple users at one time.
- When you attach an identity-based policy to a user group, all of the users in the user group receive the permissions from the user group.

IAM Roles



- An IAM role is very similar to a user, in that it is an identity with permission policies that determine what the identity can and cannot do in AWS.
- However, a role does not have any credentials (password or access keys) associated with it. Instead of being uniquely associated with one person, a role is intended to be assumable by anyone who needs it.
- An IAM user can assume a role to temporarily take on different permissions for a specific task.
- A role can be assigned to a federated user who signs in by using an external identity provider instead of IAM.
- AWS uses details passed by the identity provider to determine which role is mapped to the federated user.

IAM Policies



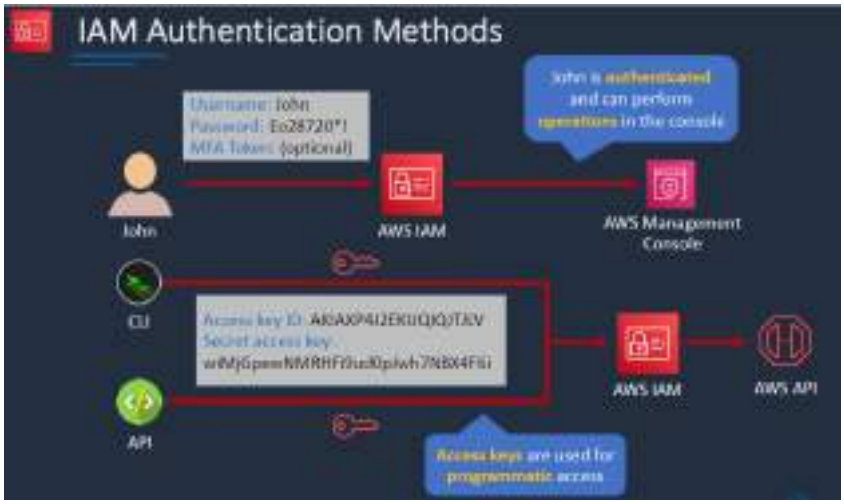
When to create an IAM user (instead of a role)

- Because an IAM user is just an identity with specific permissions in your account, you might not need to create an IAM user for every occasion on which you need credentials. In many cases, you can take advantage of IAM roles and their temporary security credentials instead of using the long-term credentials associated with an IAM user.
- You created an AWS account and you're the only person who works in your account.
- Other people in your user group need to work in your AWS account, and your user group is using no other identity mechanism.

When to create an IAM role (instead of a user)

- You're creating an application that runs on an Amazon Elastic Compute Cloud (Amazon EC2) instance and that application makes requests to AWS.
- You're creating an app that runs on a mobile phone and that makes requests to AWS.

IAM Authentication



- Authentication occurs whenever a user attempts to access your organization's network and downstream resources.
- The user must verify their identity before being granted entry for security.

- Entering credentials at a login prompt remains the most common authentication method.

Multi-Factor Authentication

- Multi-factor authentication (MFA) in AWS is a simple best practice that adds an extra layer of protection on top of your username and password.
- As a Security Best Practice, we should always require IAM Users to have Multi-Factor Authentication (MFA) enabled when accessing the AWS Console.
- A hardware device that generates a six-digit numeric code based upon a time-synchronized one-time password algorithm. The user must type a valid code from the device on a second webpage during sign-in. Each MFA device assigned to a user must be unique.

To enable a virtual MFA device for an IAM user (console)

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose **Users**.
3. In the **User Name** list, choose the name of the intended MFA user.
4. Choose the **Security credentials** tab. Next to **Assigned MFA device**, choose **Manage**.
5. In the **Manage MFA Device** wizard, choose **Virtual MFA device**, and then choose **Continue**. IAM generates and displays configuration information for the virtual MFA device, including a QR code graphic. The graphic is a representation of the "secret configuration key" that is available for manual entry on devices that do not support QR codes.

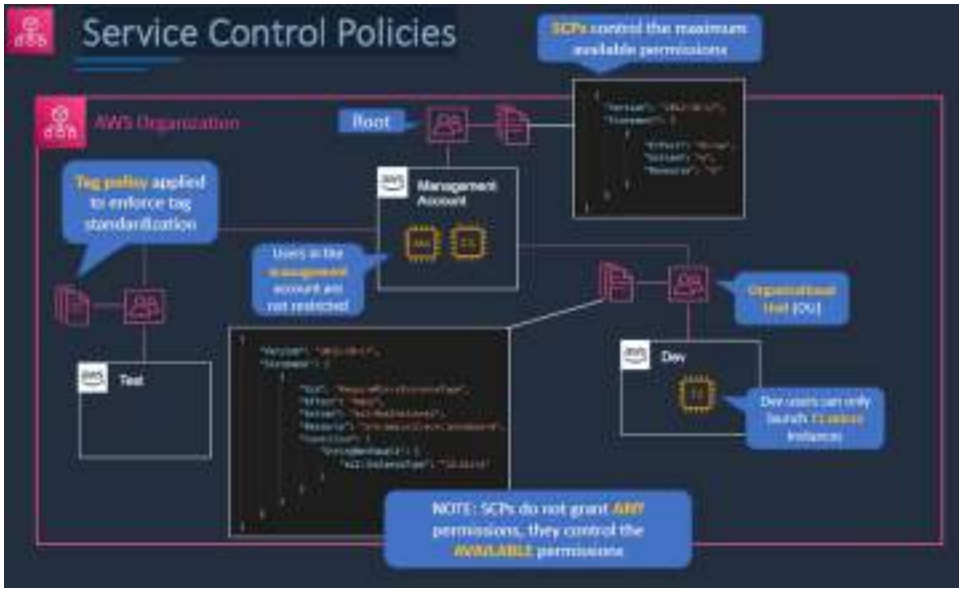
6. Open your **virtual MFA app**. For a list of apps that you can use for hosting virtual MFA devices, see Multi-Factor Authentication. If the virtual MFA app supports multiple virtual MFA devices or accounts, choose the option to create a new virtual MFA device or account.
7. Determine whether the MFA app supports QR codes, and then do one of the following:

- From the wizard, choose **Show QR code**, and then use the app to scan the QR code. For example, you might choose the camera icon or choose an option similar to **Scan code**, and then use the device's camera to scan the code.
- In the **Manage MFA Device** wizard, choose **Show secret key**, and then type the secret key into your MFA app.
  When you are finished, the virtual MFA device starts generating one-time passwords.

8. In the **Manage MFA Device** wizard, in the **MFA code 1** box, type the one-time password that currently appears in the virtual MFA device. Wait up to 30 seconds for the device to generate a new one-time password. Then type the second one-time password into the **MFA code 2** box. Choose **Assign MFA**.
   The virtual MFA device is now ready for use with AWS.

Service Control Policies

- Service control policies (SCPs) are a type of organization policy that you can use to manage permissions in your organization.
- SCPs offer central control over the maximum available permissions for all accounts in your organization.
- SCPs help you to ensure your accounts stay within your organization's access control guidelines.
- SCPs are available only in an organization that has all features enabled.



- SCPs aren't available if your organization has enabled only the consolidated billing features.
- SCPs alone are not sufficient to granting permissions to the accounts in your organization.
- No permissions are granted by an SCP.
- An SCP defines a guardrail, or sets limits, on the actions that the account's administrator can delegate to the IAM users and roles in the affected accounts.
- The administrator must still attach identity-based or resource-based policies to IAM users or roles, or to the resources in your accounts to actually grant permissions.

- The effective permissions are the logical intersection between what is allowed by the SCP and what is allowed by the IAM and resource-based policies.

AWS IAM Best Practices

- Lock away your AWS account root user access keys
- Create individual IAM users
- User groups to assign permissions to IAM users
- Grant least privilege
- Get started using permissions with AWS managed policies
- Use customer managed policies instead of inline policies
- Use access levels to review IAM permissions
- Configure a strong password policy for your users
- Enable MFA
- Use roles for applications that run on Amazon EC2 instances
- Use roles to delegate permissions
- Do not share access keys
- Rotate credentials regularly
- Remove unnecessary credentials
- Use policy conditions for extra security
- Monitor activity in your AWS account

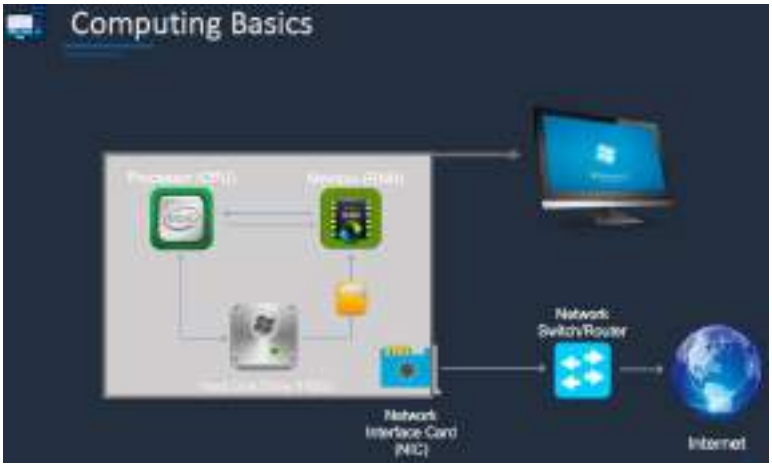
8.11 AWS Compute Services

Computing Basics

- AWS offers the broadest and deepest functionality for compute.
- Amazon Elastic Cloud Compute (EC2) offers granular control for managing your infrastructure with the choice of processors, storage, and networking.
- AWS container services offer the best choice and flexibility of services to run your containers.

Measurements:

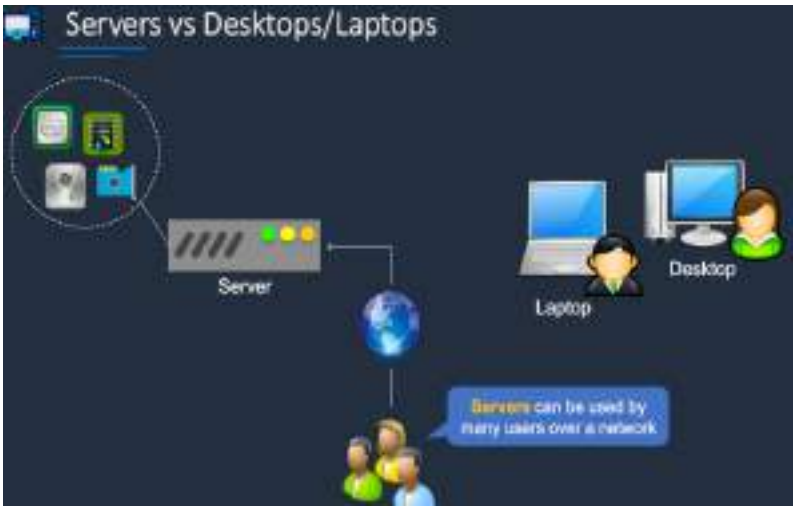
- CPU is measured in Gigahertz (Ghz)
- RAM is measured in Gigabyte (GB)
- HDD is measured in Gigabyte (GB)
- NIC is measured in Megabits per second (Mbps) or Gigabits per second (Gbps)



Servers' vs Desktops/Laptops

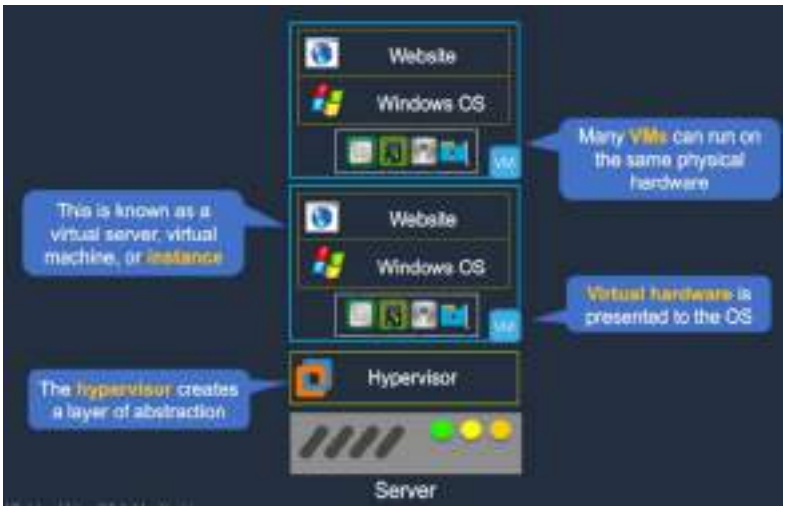
Server Hardware Build:

- Hardware is more specialized
- Much higher prices compared to desktops / laptops
- Includes redundancy

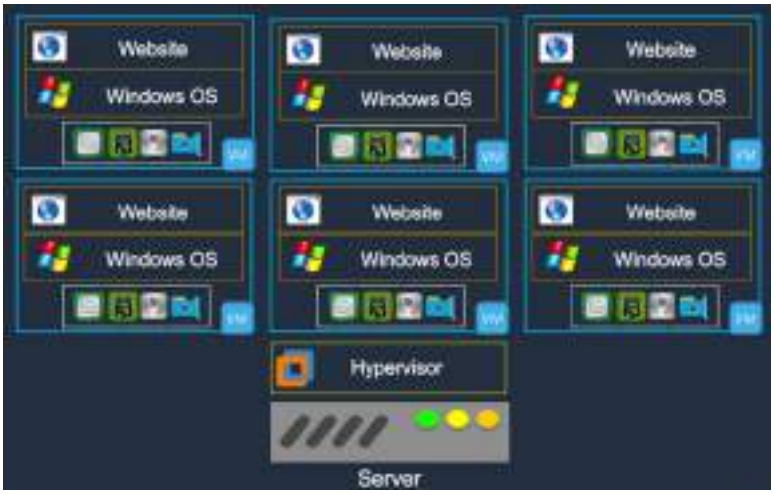


8.12 Server Virtualization

Server Virtualization



- This virtualization type provides the ability to run an operating system directly on top of a virtual machine without any modification, as if it were run on the bare-metal hardware.
- The Amazon EC2 host system emulates some or all of the underlying hardware that is presented to the guest.

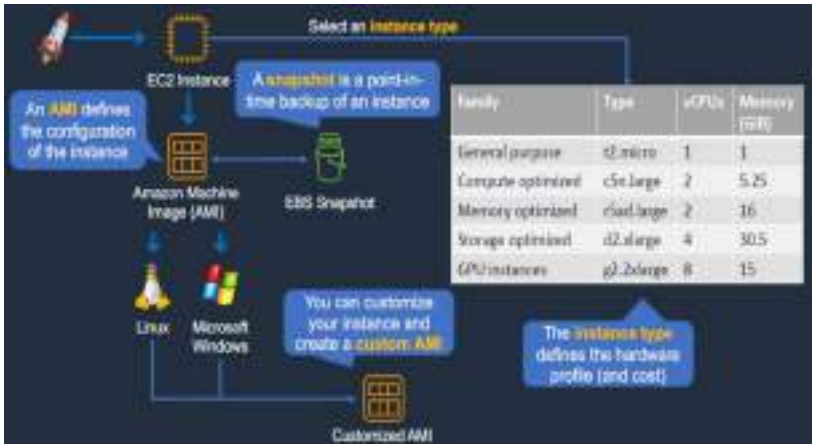


8.13 Amazon Elastic Compute Cloud (EC2)

Amazon EC2

- Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud.
- It is designed to make web-scale computing easier for developers.

Launching an EC2 Instance



To launch a new EC2 instance from an AMI, do the following:

1. Open the EC2 console.
   Note: Be sure to select the AWS Region that you want to launch the instance in.
2. From the navigation bar, choose AMIs.
3. Find the AMI that you want to use to launch a new instance. To begin, open the menu next to the search bar, and then choose one of the following:
   - If the AMI that you're using is one that you created, select Owned by me.
   - If the AMI that you're using is a public AMI, select Public images.
   - If the AMI that you're using is a private image that someone else shared with you, select Private images.

   Note: The search bar automatically provides filtering options as well as automatically matching AMI IDs.
4. Select the AMI, and then choose Launch.
5. Choose an instance type, and then choose Next: Configure Instance Details. Optionally select configuration details, such as associating an IAM role with the instance.
6. Select Next: Add Storage. You can use the default root volume type, or select a new type from the Volume Type drop down. Select Add New Volume if you want to add additional storage to your instance.
7. Select Next: Add Tags. You can add custom tags to your instance to help you categorize your resources.
8. Select Next: Configure Security Group. You can associate a security group with your instance to allow or block traffic to the instance.
9. Select Review and Launch. Review the instance details.
10. Select Previous to return to a previous screen to make changes. Select Launch when you are ready to launch the instance.
11. Select an existing key pair or create a new key pair, select the acknowledge agreement box, and then choose Launch Instances.
12. Choose View Instances to check the status of your instance.

Benefits of Amazon EC2

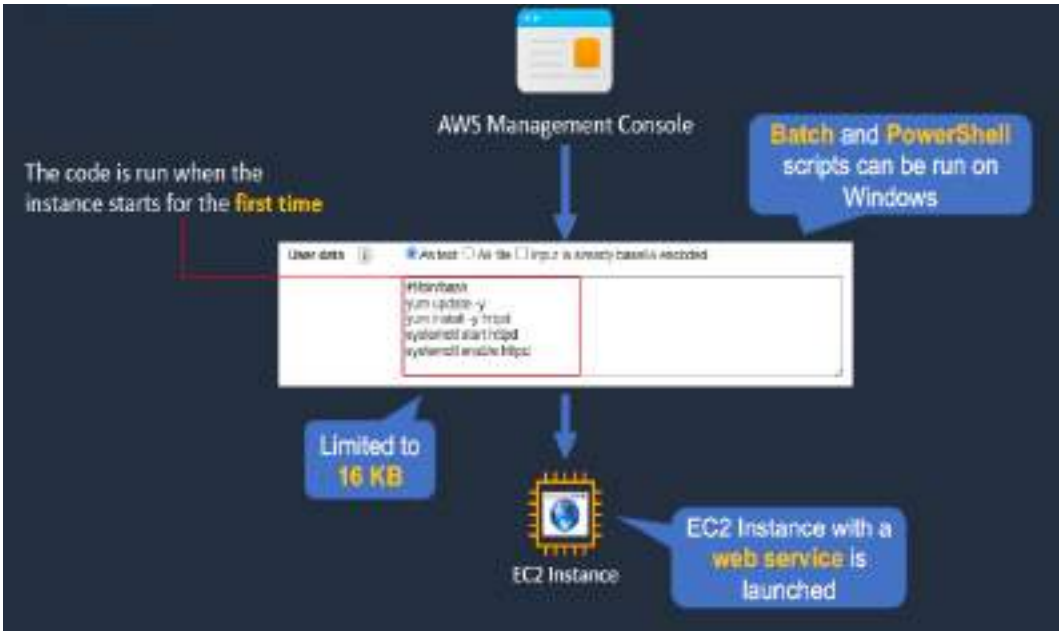
- **Elastic Computing:** easily launch hundreds to thousands of EC2 instances within minutes
- **Complete Control:** you control the EC2 instances with full root/administrative access
- **Flexible:** Choice of instance types, operating systems, and software packages
- **Reliable:** EC2 offers very high levels of availability and instances can be rapidly commissioned and replaced
- **Secure:** Fully integrated with Amazon VPC and security features
- **Inexpensive**: Low cost, pay for what you use

Amazon EC2 Instance in a Public Subnet

- EC2 instance is in a public subnet (defined as a subnet with a Route Table pointing to an Internet Gateway).
- EC2 instance has a public IP address.

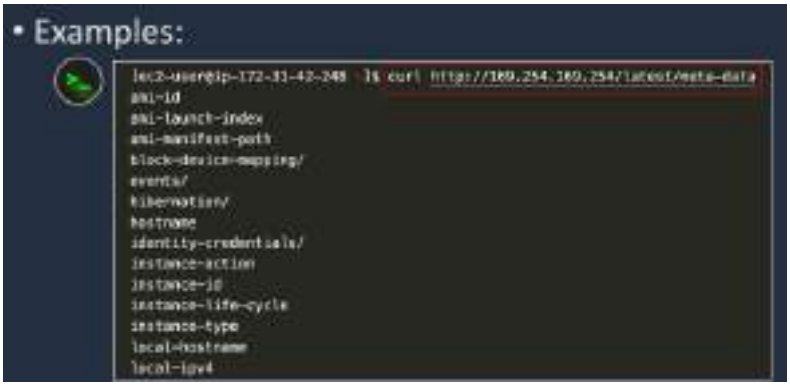
Amazon EC2 User Data

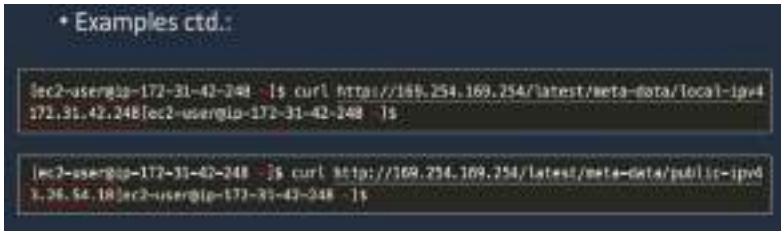
- When you launch an instance in Amazon EC2, you have the option of passing user data to the instance that can be used to perform common automated configuration tasks and even run scripts after the instance starts.
- You can pass two types of user data to Amazon EC2: shell scripts and cloud-init directives.



Amazon EC2 Metadata

- Instance metadata is data about your EC2 instance
- Instance metadata is available at http://169.254.169.254/latest/meta-data





Access Keys

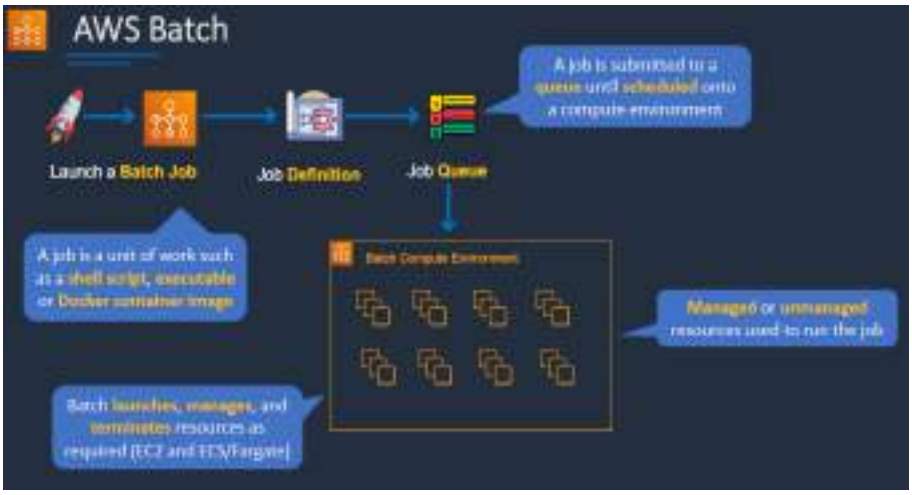
- An access key grants programmatic access to your resources.
- Access keys are long-term credentials for an IAM user or the AWS account root user.
- This means that you must guard the access key as carefully as the AWS account root user sign-in credentials.

Amazon EC2 Instance Profiles (IAM Roles for EC2)

- You can use IAM roles to grant permissions to applications running on your instances that need to use a bucket in Amazon S3.
- You can specify permissions for IAM roles by creating a policy in JSON format.
- These are similar to the policies that you create for IAM users.

AWS Batch

AWS Batch is a set of batch management capabilities that enables developers, scientists, and engineers to easily and efficiently run hundreds of thousands of batch computing jobs on AWS.

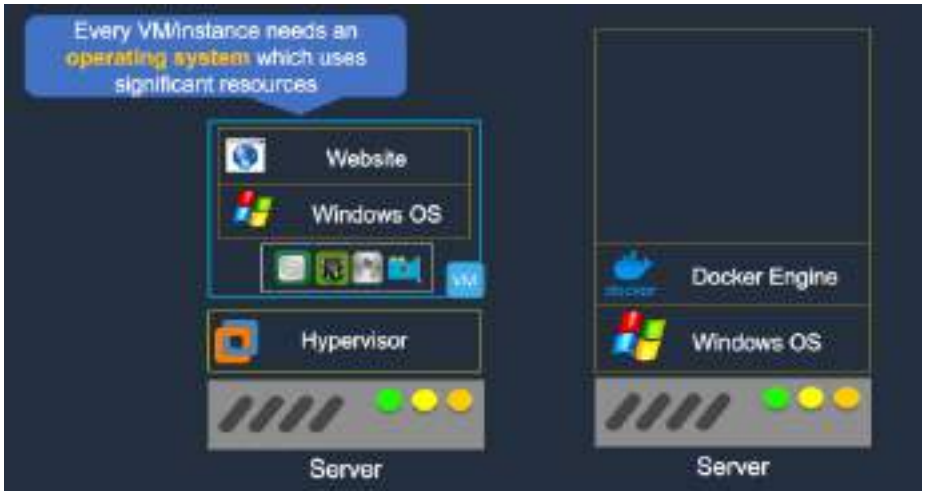


Amazon LightSail

- Low cost and ideal for users with less technical expertise.
- Compute, storage and network
- Preconfigures virtual servers
- Virtual servers. Databases and load balancers
- SSH and RDP access
- Can access Amazon VPC

Server Virtualization vs Containers

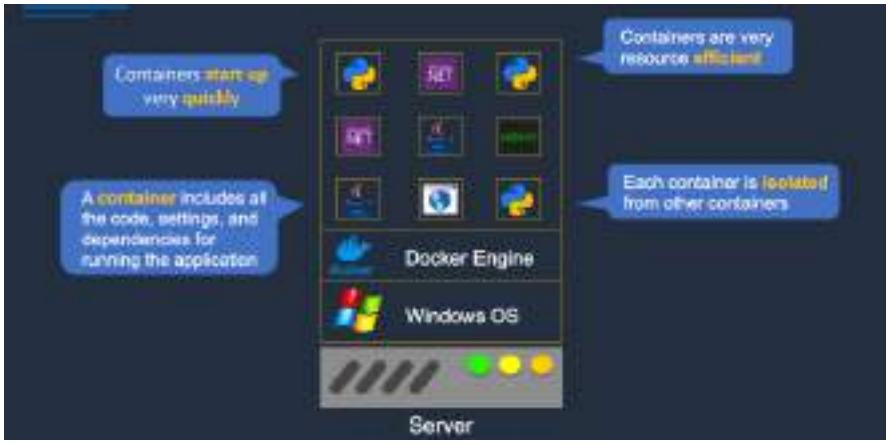
Virtualization enables you to run multiple operating systems on the hardware of a single physical server, while containerization enables you to deploy multiple applications using the same operating system on a single virtual machine or server.



Docker Containers

- Docker is a software platform that allows you to build, test, and deploy applications quickly.

- Docker packages software into standardized units called containers that have everything the software needs to run including libraries, system tools, code, and runtime.
- Using Docker, you can quickly deploy and scale applications into any environment and know your code will run.
- Running Docker on AWS provides developers and admins a highly reliable, low-cost way to build, ship, and run distributed applications at any scale.



Monolithic Application

- A monolithic application is built as a single unit. Enterprise applications are built in three parts:
  - A database: consisting of many tables usually in a relational database management system.
  - A client-side user interface: consisting of HTML pages
  - JavaScript: running in a browser
- They're typically complex applications that encompass several tightly coupled functions.
- For example, consider a monolithic ecommerce SaaS application. It might contain a web server, a load balancer, a catalogue service that services up product images, an ordering system, a payment function, and a shipping component.

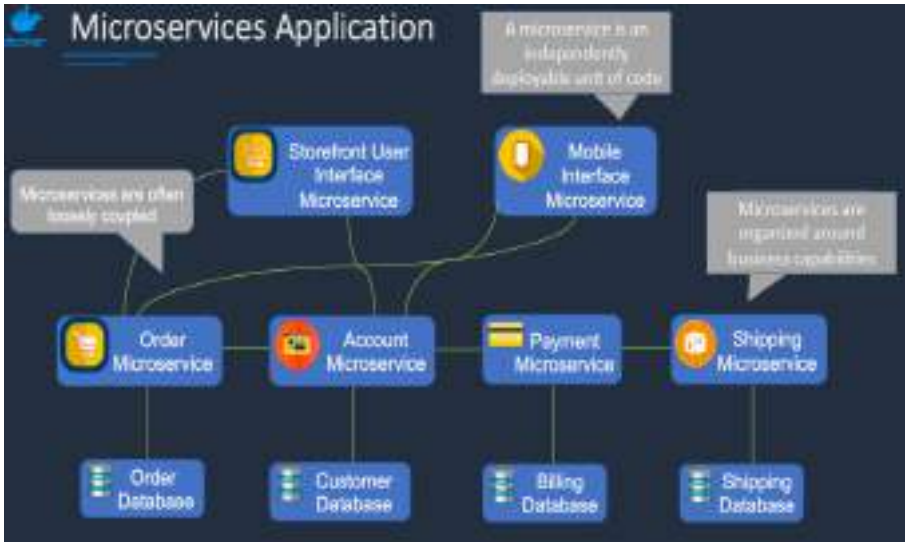
How do you deploy monolithic applications in AWS?

1. Launch an ECS Cluster using AWS CloudFormation
2. Check your Cluster is Running
3. Write a Task Definition

4. Configure the Application Load Balancer: Target Group
5. Configure the Application Load Balancer: Listener
6. Deploy the Monolith as a Service
7. Test your Monolith

Microservices Application

- Microservices are an architectural and organizational approach to software development where software is composed of small independent services that communicate over well-defined APIs.
- These services are owned by small, self-contained teams.
- Microservices architectures make applications easier to scale and faster to develop, enabling innovation and accelerating time-to-market for new features.



- With a microservices architecture, an application is built as independent components that run each application process as a service.
- These services communicate via a well-defined interface using lightweight APIs.
- Services are built for business capabilities and each service performs a single function. Because they are independently run, each service can be updated, deployed, and scaled to meet demand for specific functions of an application.

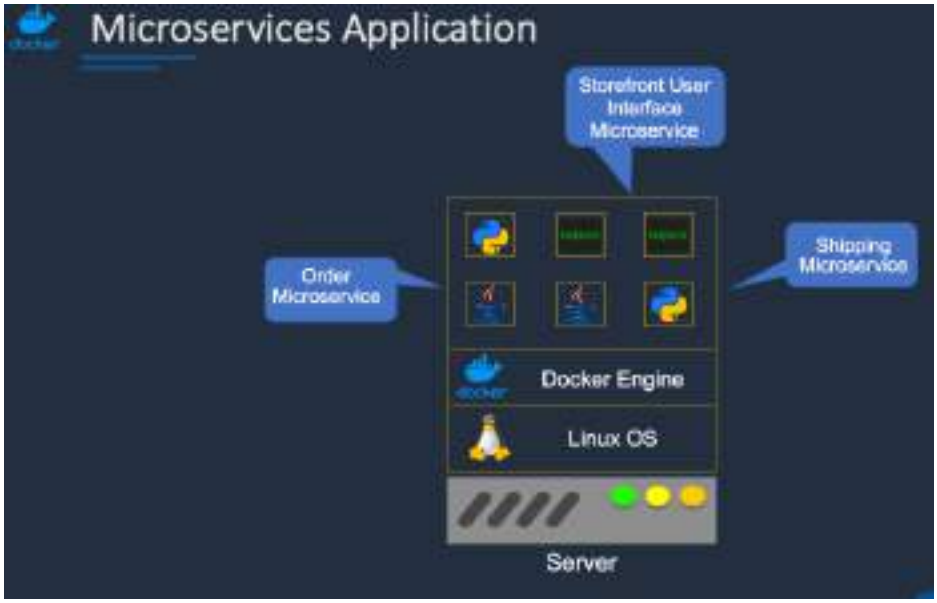
Characteristics of Microservices

Autonomous

- Each component service in a microservices architecture can be developed, deployed, operated, and scaled without affecting the functioning of other services.
- Services do not need to share any of their code or implementation with other services.
- Any communication between individual components happens via well-defined APIs.

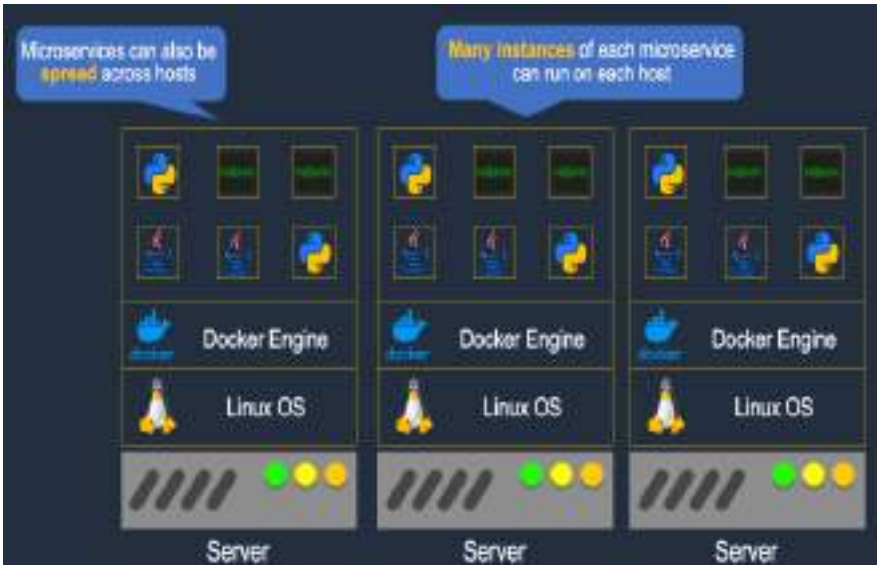
Specialized

- Each service is designed for a set of capabilities and focuses on solving a specific problem.
- If developers contribute more code to a service over time and the service becomes complex, it can be broken into smaller services.



Benefits of Microservices

- Agility
- Flexible Scaling
- Easy Deployment
- Technological Freedom
- Reusable Code
- Resilience

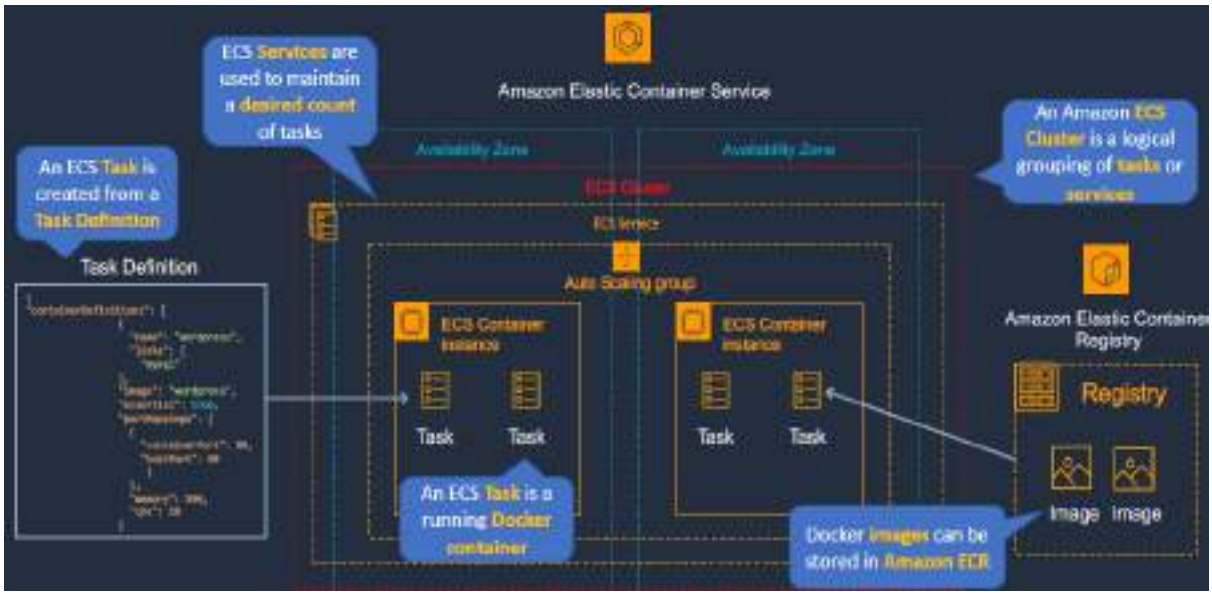


8.14 Amazon Elastic Container Service (ECS)

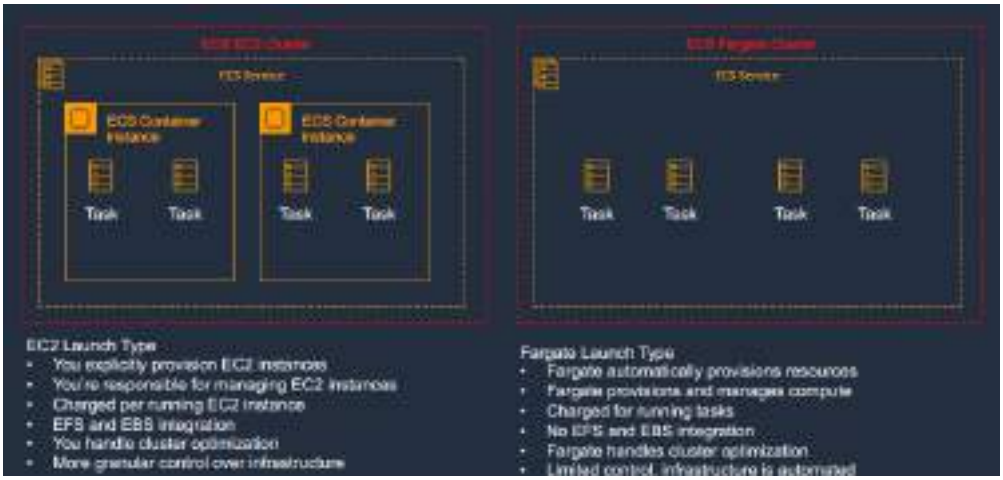
Amazon ECS

- Amazon Elastic Container Service (Amazon ECS) is a highly scalable and fast container management service.
- You can use it to run, stop, and manage containers on a cluster.
- With Amazon ECS, your containers are defined in a task definition that you use to run an individual task or task within a service.

- Amazon Elastic Container Service (ECS) is a cloud computing service in Amazon Web Services (AWS) that manages containers and allows developers to run applications in the cloud without having to configure an environment for the code to run in.



- An Amazon ECS cluster is a logical grouping of tasks or services.
- Your tasks and services are run on infrastructure that is registered to a cluster.



AWS Storage Services

Block Storage

- Other enterprise applications like databases or ERP systems often require dedicated, low latency storage for each host.
- This is analogous to direct-attached storage (DAS) or a Storage Area Network (SAN).
- Block-based cloud storage solutions like Amazon Elastic Block Store (EBS) are provisioned with each virtual server and offer the ultra-low latency required for high performance workloads.

File Storage

- Some applications need to access shared files and require a file system.
- This type of storage is often supported with a Network Attached Storage (NAS) server.
- File storage solutions like Amazon Elastic File System (EFS) are ideal for use cases like large content repositories, development environments, media stores, or user home directories.

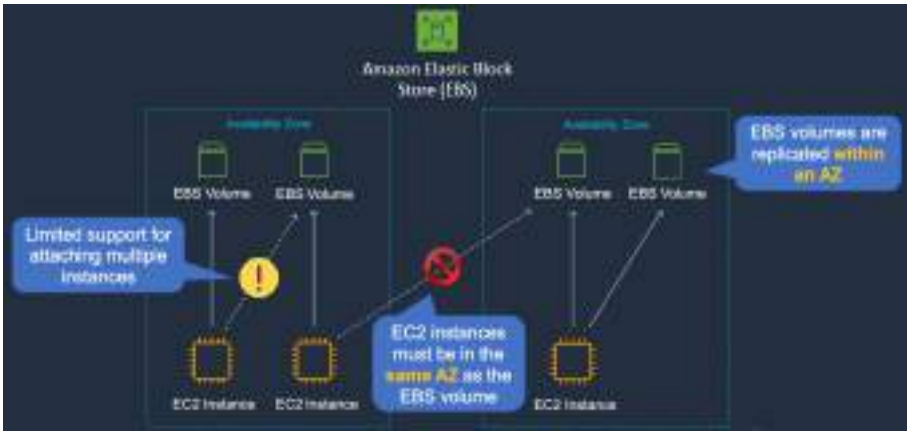
Object Storage

- Applications developed in the cloud often take advantage of object storage's vast scalability and metadata characteristics.
- Object storage solutions like Amazon Simple Storage Service (S3) are ideal for building modern applications from scratch that require scale and flexibility, and can also be used to import existing data stores for analytics, backup, or archive.

8.15 Amazon Elastic Block Store (EBS)

Amazon EBS

- Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with EC2 instances.
- EBS volumes behave like raw, unformatted block devices. You can mount these volumes as devices on your instances.

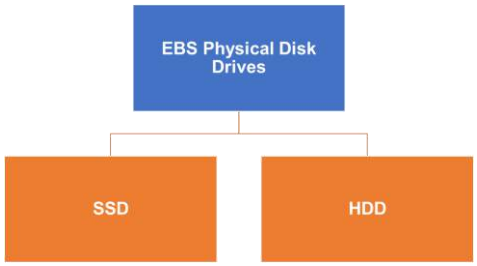


- EBS volume data persists independently of the life of the instance
- EBS volumes do not need to be attached to an instance

- You can attach multiple EBS volumes to an instance
- You can use multi-attach to attach a volume to multiple instances but with some constraints
- EBS volumes must be in the same AZ as the instances they are attached to
- Root EBS volumes are deleted on the termination by default
- Extra non-boot volumes are not deleted on termination by default

AWS EBS uses two categories of physical disk drives:

- These are Solid State Drive (SSD) and Hard Disk Drives (HDD) drives which can be selected upon provisioning the EBS volume based on the use case.

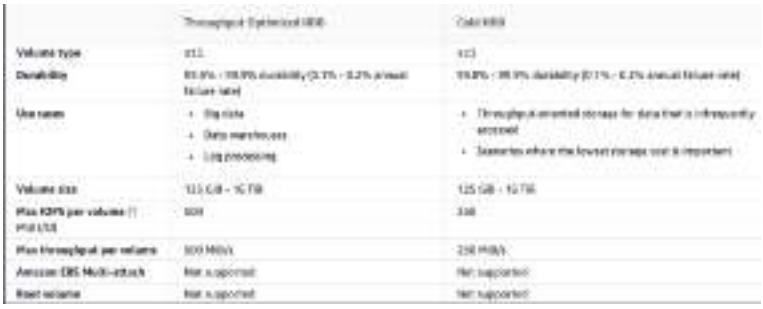


Amazon EBS SSD-Backed Volumes

SSD-backed storage is for transactional workloads (performance depends primarily on IOPS, latency, and durability).

Amazon EBS HDD-Backed Volumes

HDD-backed storage for throughput workloads (performance depends primarily on throughput, measured in MB/s)



Amazon Data Lifecycle Manager (DLM)

- You can use Amazon Data Lifecycle Manager to automate the creation, retention, and deletion of EBS snapshots and EBS-backed AMIs.
- When you automate snapshot and AMI management, it helps you to:
  - Protect valuable data by enforcing a regular backup schedule.
  - Create standardized AMIs that can be refreshed at regular intervals.
  - Retain backups as required by auditors or internal compliance.
  - Reduce storage costs by deleting outdated backups.
  - Create disaster recovery backup policies that back up data to isolated accounts.
- When combined with the monitoring features of Amazon CloudWatch Events and AWS CloudTrail, Amazon Data Lifecycle Manager provides a complete backup solution for Amazon EC2 instances and individual EBS volumes at no additional cost.

Elements

The following are the key elements of Amazon Data Lifecycle Manager.

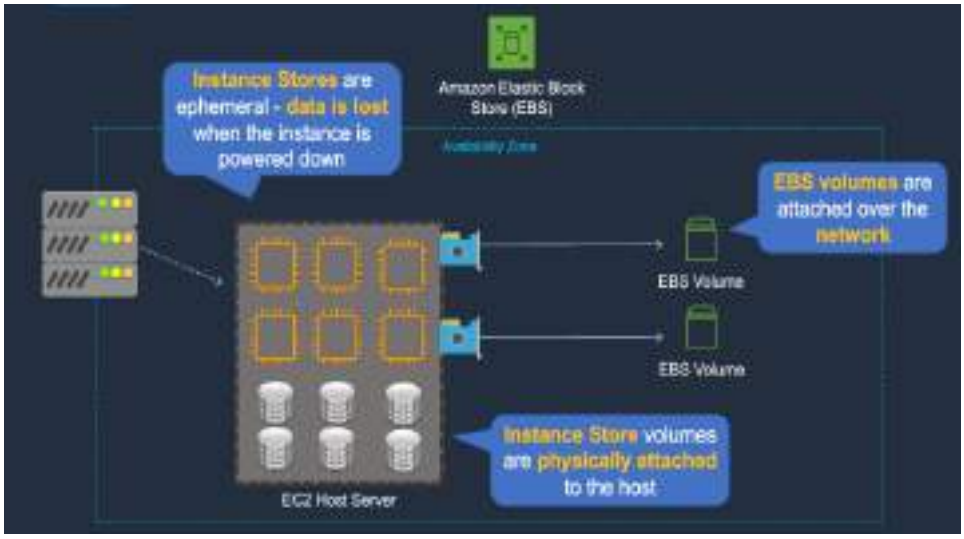
- Snapshots
- EBS-backed AMIs
- Target resource tags
- Amazon Data Lifecycle Manager tags
- Lifecycle policies
- Policy schedules

Amazon EBS Snapshots and DLM

EBS vs Instance Store

- Some Amazon Elastic Compute Cloud (Amazon EC2) instance types come with a form of directly attached, block-device storage known as the instance store.
- The instance store is ideal for temporary storage, because the data stored in instance store volumes is not persistent through instance stops, terminations, or hardware failures.
- For data you want to retain longer, or if you want to encrypt the data, use Amazon Elastic Block Store (Amazon EBS) volumes instead.
- EBS volumes preserve their data through instance stops and terminations, can be easily backed up with EBS snapshots, can be removed from one instance and reattached to another, and support full-volume encryption.

- To prevent unintentional changes or data loss, it's a best practice to perform regular snapshots, which can be automated with AWS Backup.



8.16 Amazon Machine Images (AMI)

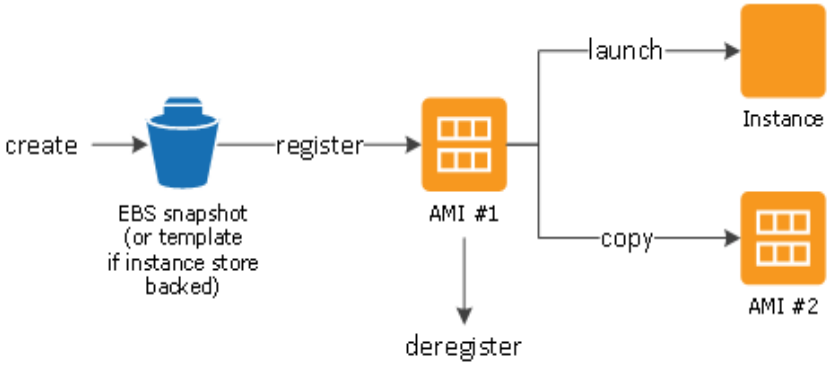
Amazon Machine Images (AMIs)

An Amazon Machine Image (AMI) provides the information required to launch an instance.
An AMI includes the following:

- One or more EBS snapshots, or, for instance-store-backed AMIs, a template for the root volume of the instance (for example, an operating system, an application server, and applications)
- Launch permissions that control which AWS accounts can use the AMI to launch instances
- A block device mapping that specifies the volumes to attach to the instance when it's launched

AMIs come in three main categories:

- **Community AMIs** - free to use, generally you just select the operating system you want
- **AWS Marketplace AMIS** - pay to use, generally come packaged with additional, licensed software
- **My AMIS** - AMIs that you create yourself



8.17 Amazon Elastic File System (EFS)

Amazon EFS

- Amazon Elastic File System is a cloud storage service provided by Amazon Web Services designed to provide scalable, elastic, concurrent with some restrictions, and encrypted file storage for use with both AWS cloud services and on-premises resources.
- Amazon EFS automatically grows and shrinks as you add and remove files with no need for management or provisioning.

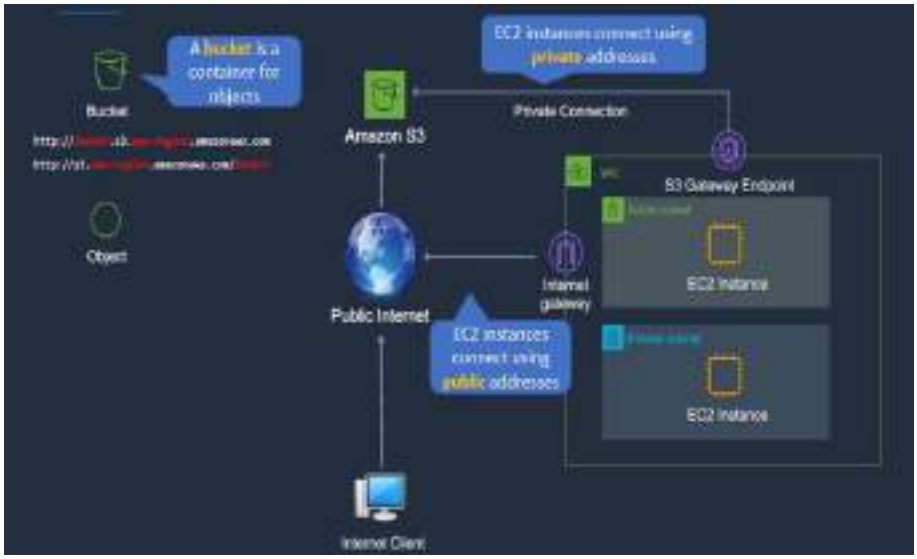
Amazon EFS features

- Serverless
- Storage classes and lifecycle management
- Security and compliance
- Scalable performance
- Shared file system with NFS v4
- Performance modes
- Containers and serverless file storage

8.18 Amazon Simple Storage Service (S3)

Amazon S3

- Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance.
- You can use Amazon S3 to store and retrieve any amount of data at any time, from anywhere.
- Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its e-commerce network.
- You can store any type of file in S3
- Files can be anywhere from 0 bytes to 5 TB
- There is unlimited storage available



An object consists of:

- Key (name of objects)
- Version ID
- Value (actual Date)
- Metadata
- Subresources
- Access Control Information

- S3 is a universal namespace so bucket names must be unique globally
- You create your buckets within a REGION
- It is a best practice to create buckets in regions that are physically closest to your users to reduce latency

Additional Features

S3 Capability	What it Does
Transfer	Acceleration Speed up data uploads using CloudFront in reverse
Requester Pays	The requester rather than the bucket owner pays for requests and data transfer
Events	Events Trigger notifications to sms, sas, or Lambda when certain events happen in your bucket
Static Web Hosting	Simple and massively scalable static website hosting
Versioning and Replication	Retain versions of objects and replicate objects within and across AWS Regions

Amazon S3 Availability and Durability

Availability	Durability
Measures how readily available the service is	Measures the likelihood of data loss
Measured as a percentage	All storage classes offer 99.999999999% durability
S3 availability SLA varies between storage classes	This means that if you store 100 billion objects in S3, you will lose one object at most

Amazon S3 Storage Classes

Amazon S3 Versioning

- Versioning is a means of keeping multiple variants of an object in the same bucket
- Use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket
- Versioning-enabled buckets enable you to recover objects from accidental deletion or overwrite

Amazon S3 Replication

Amazon S3 Glacier

- Extremely low cost and you pay only for what you need with no commitments of upfront fees
- Two classes Glacier and Glacier Deep Archive
- Three options for access to archives, listed in the table below:

	Expedited	Standard	Bulk
Data access time (Glacier)	1-5 minutes	3-5 hours	5-12 hours
Data access time (Deep Archive)	N/A	12 hours	48 hours

Object Lock and Glacier Vault Lock

S3 Object Lock

- Store objects using a write-once-read-many (WORM) model
- Prevent objects from being deleted or overwritten for a fixed time or indefinitely

S3 Glacier Vault Lock

- Also used to enforce a WORM model
- Can apply a policy and lock the policy from future edits
- Use for compliance objectives and data retention

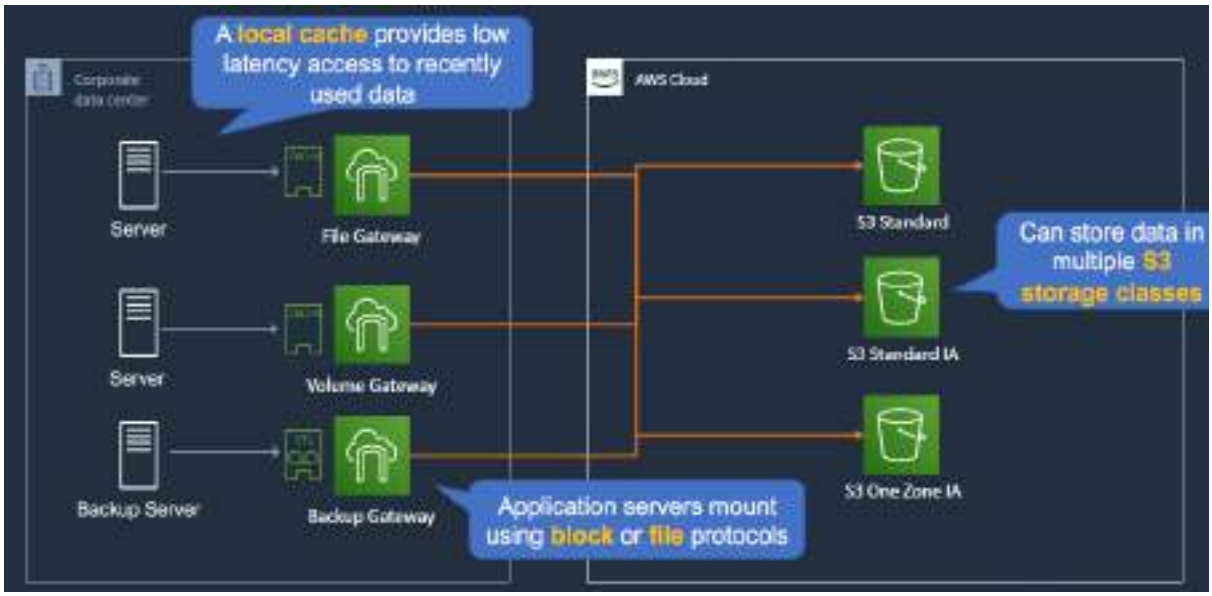
AWS Storage Gateway

- Hybrid cloud storage service
- Access cloud storage from on-premises applications
- Enables access to proprietary object storage (53) using standard protocols

Use cases:

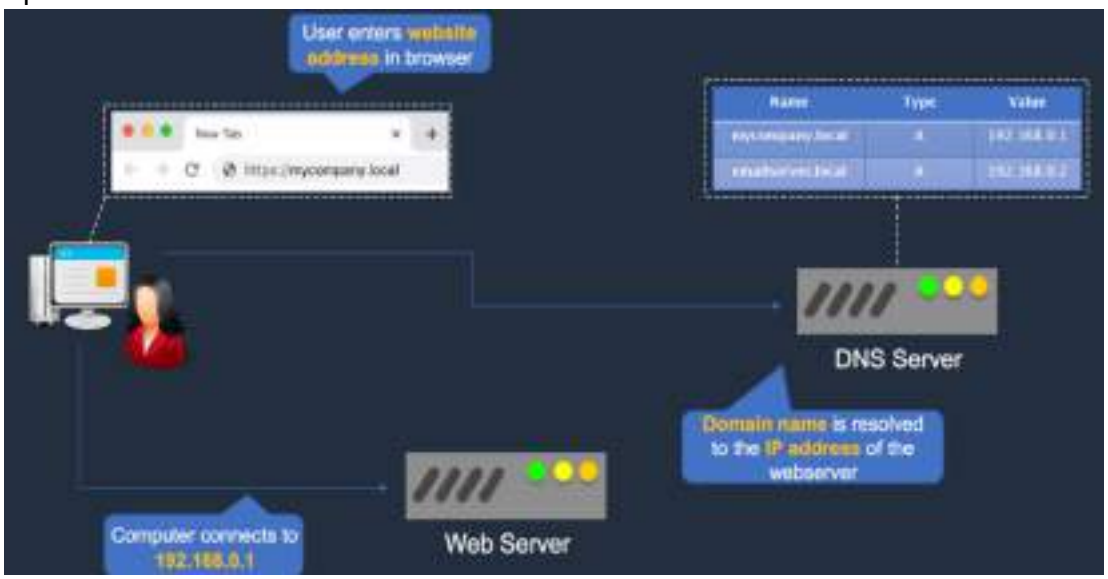
- Moving backups to the cloud

- Using on-premises file shares backed by cloud storage
- Low latency access to data in AWS for on-premises applications
- Disaster recovery



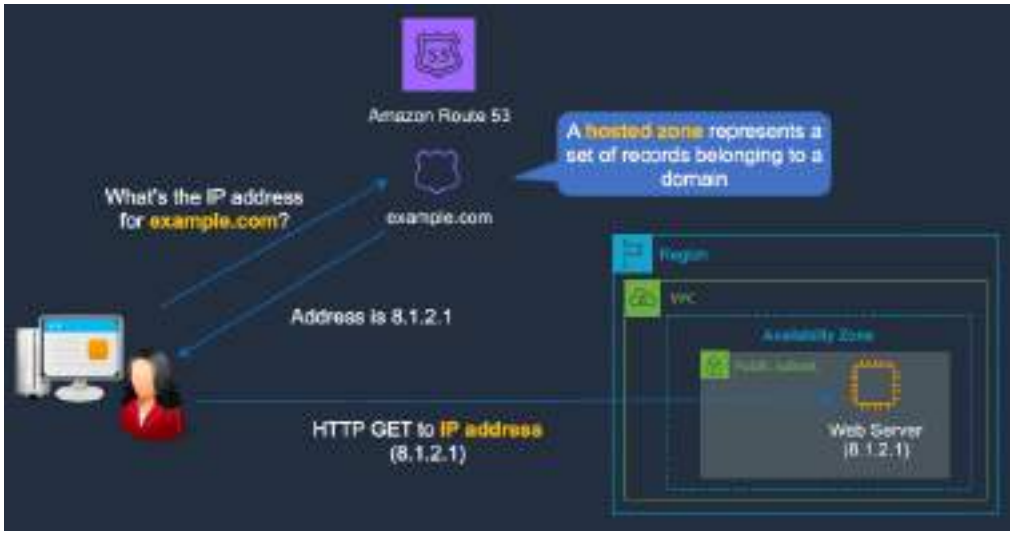
The Domain Name System (DNS)

- The domain name system (DNS) is a naming database in which internet domain names are located and translated into Internet Protocol (IP) addresses.
- The domain name system maps the name people use to locate a website to the IP address that a computer uses to locate that website.



Amazon Route 53

- Amazon Route 53 connects user requests to internet applications running on AWS or on-premises.
- You can use Amazon Route 53 as the DNS service for your domain, such as example.com.
- When Route 53 is your DNS service, it routes internet traffic to your website by translating friendly domain names like www.example.com into numeric IP addresses, like 192.0. 2.1, that computers use to connect to each other.



- The name for our service (Route 53) comes from the fact that DNS servers respond to queries on port 53 and provide answers that route end users to your applications on the Internet.

Amazon Route 53 Routing Policies

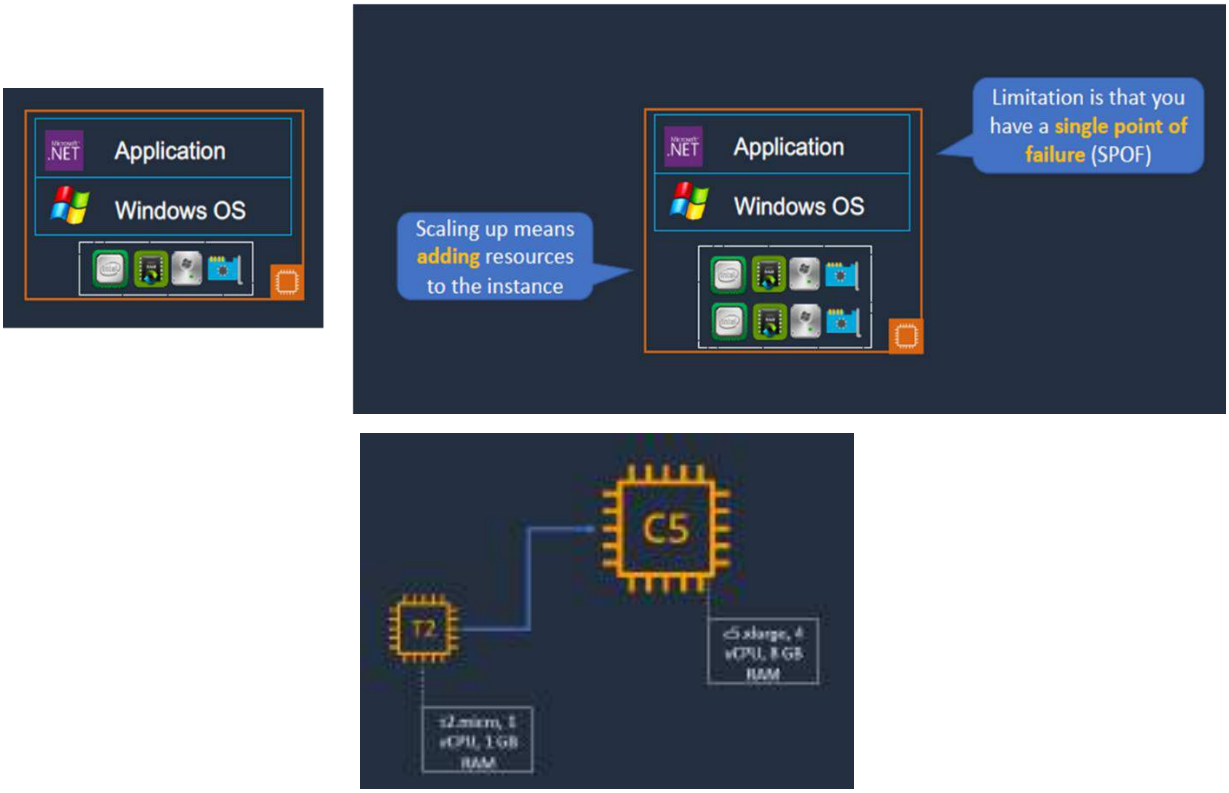
Routing Policy	What it does
Simple	Simple DNS response providing the IP address associated with a name
Failover	If primary is down (based on health checks), routes to secondary destination
Geolocation	Uses geographic location you're in (e.g. Europe) to route you to the closest region
Geoproximity	Routes you to the closest region within a geographic area
Latency	Directs you based on the lowest latency route to resources
Multivalue answer	Returns several IP addresses and functions as a basic load balancer
Weighted	Uses the relative weights assigned to resources to determine which to route to

Amazon Route Features

Scaling Up (Vertical Scaling)

- Vertical scaling is about changing the instance up and down
- The new version of the AWS features vertical scaling for Amazon EC2 instances.

- With vertical scaling, the solution automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost.



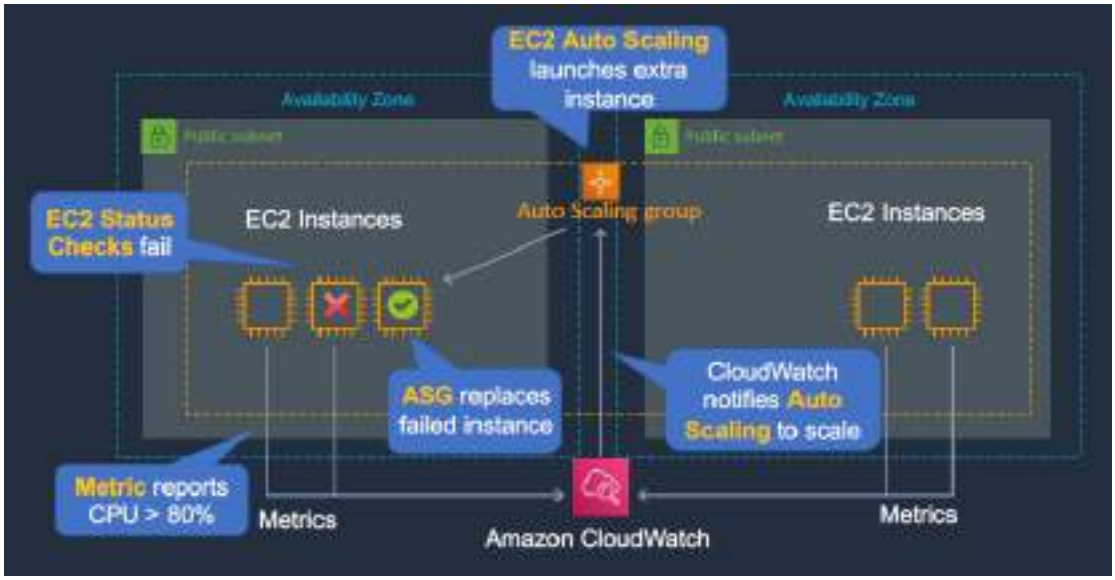
Scaling Out (Horizontal Scaling)

- Horizontal scaling is about adding more machines of similar capacity to the infrastructure.
- A simple example of horizontal scaling in AWS Cloud is adding/removing Amazon EC2 instances from your application architecture behind Elastic Load Balancer.

Amazon EC2 Auto Scaling

- Amazon EC2 Auto Scaling helps you maintain application availability and allows you to automatically add or remove EC2 instances according to conditions you define.

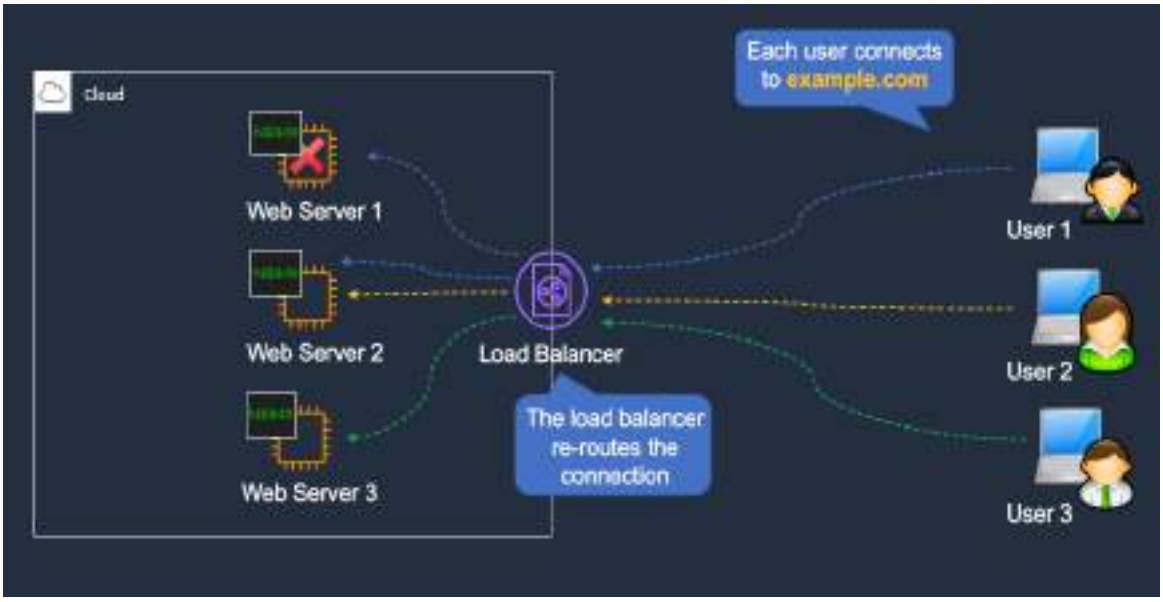
- You can use the fleet management features of EC2 Auto Scaling to maintain the health and availability of your fleet.



- EC2 Auto Scaling launches and terminates instances dynamically
- Scaling is horizontal (scales out)
- Provides elasticity and scalability
- Responds to EC2 status checks and CloudWatch metrics
- Can scale based on demand (performance) or on a schedule
- Scaling policies define how to respond to changes in demand

Load Balancing and High Availability

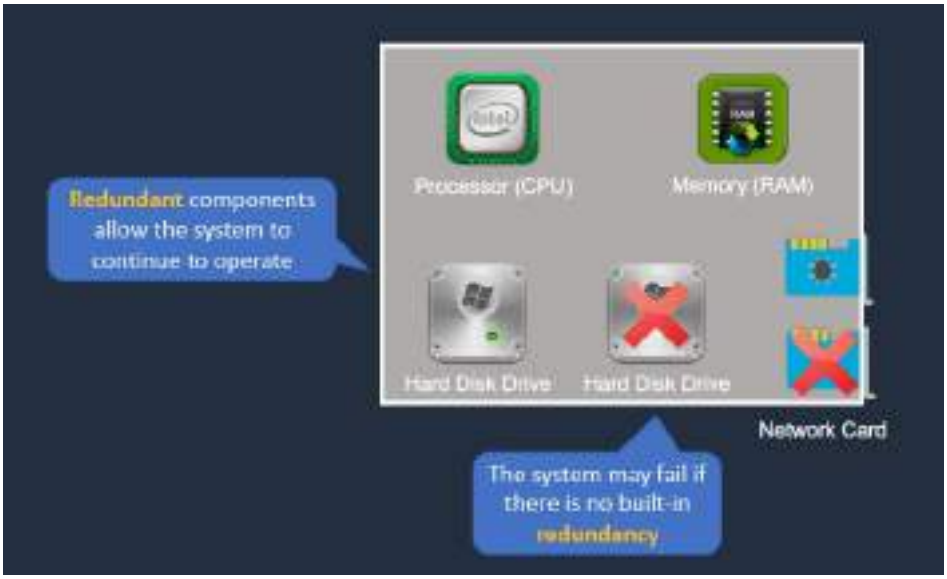
Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple targets and virtual appliances in one or more Availability Zones (AZs).



Fault Tolerance

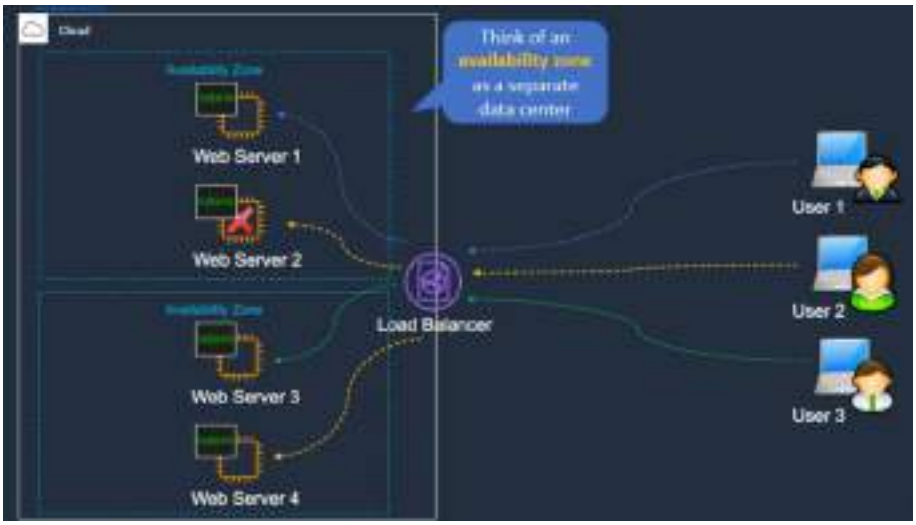
- Fault tolerance is the ability of a workload to remain operational with zero downtime or data loss in the event of a disruption.

- In a fault-tolerant environment, instances of the same workload are typically hosted on two or more independent sets of servers.



High Availability

- It is the ability of a workload to remain operational, with minimal downtime, in the event of a disruption. Disruptions include hardware failure, networking problems or security events, such as DDoS attacks.
- In a highly available system, workloads are spread across a cluster of servers. If one server fails, the workloads running on it automatically move to other servers.



High Availability Vs Fault Tolerance

Level of disruption:

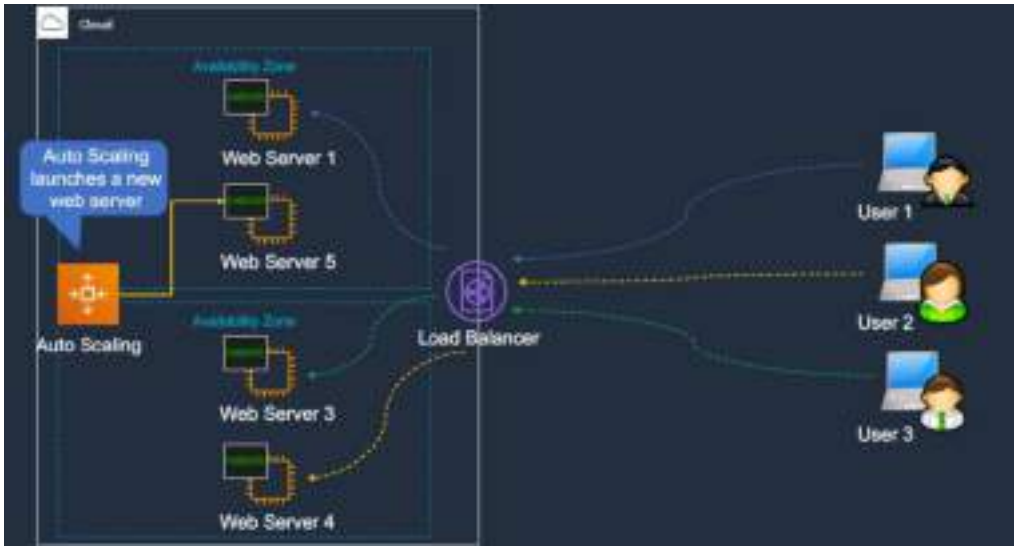
- With high availability, a workload usually experiences some level of disruption when a failure occurs.
- It may take seconds or minutes to migrate a workload from a failed server to a new one in a highly available cluster. Because the new server likely does not have identical copies of the failed server's data, there may be some permanent data loss.
- In contrast, a successful fault-tolerant environment provides zero downtime and no data loss because both instances maintain identical copies of the data.

Infrastructure requirements:

- Fault-tolerant environments require IT organizations to mirror workloads on dedicated infrastructure.
- As a result, these environments double an organization's infrastructure footprint, in the cloud or on premises.
- In either deployment scenario, expect twice the hosting costs of a non-fault tolerant workload.
- Highly available environments are not as demanding, but they do require some extra infrastructure capacity.
- This makes highly available environments less expensive to operate than fault-tolerant ones.

Management:

- Fault-tolerant workloads are more challenging to set up and administer.
- To ensure fault tolerance, admins must keep two or more workload instances in sync.
- This mean that changes in one instance are implemented in the other instance instantaneously.
- In contrast, high-availability workloads are less complex to set up and manage.



Types of Elastic Load Balancer (ELB)

Application Load Balancer

- Operates at the request level
- Routes based on the content of the request (layer 7)
- Supports advanced routing

Network Load Balancer

- Operates at the connection level
- Routes connections based on the IP protocol data (layer 4)
- Offers ultra-high performance, low latency and TLS offloading at scale

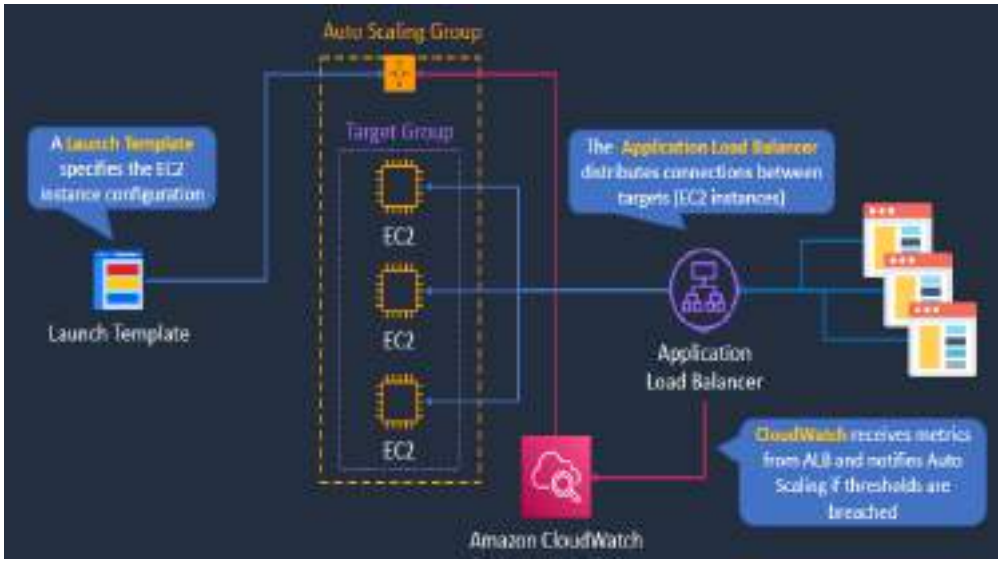
Classic Load Balancer

- Old generation: not recommended for new application
- Performs routing a Layer 4 and Layer 7
- Use of existing applications running in EC2-Classic

Gateway Load Balancer

- Used in front of virtual appliances such as firewalls, IDS/IPS, and deep packet inspection systems.

Elastically Scale the Application



Scaling Policies

- **Target Tracking -** Attempts to keep the group at or close to the metric

- **Simple Scaling -** Adjust group size based on a metric
- **Step Scaling -** Adjust group size based on a metric - adjustments vary based on the size of the alarm breach
- **Scheduled Scaling -** Adjust the group size at a specific time

Serverless Services

- With serverless there are no instances to manage
- You don't need to provision hardware
- There is no management of operating systems or software
- Capacity provisioning and patching is handled automatically
- Provides automatic scaling and high availability
- Can be very cheap

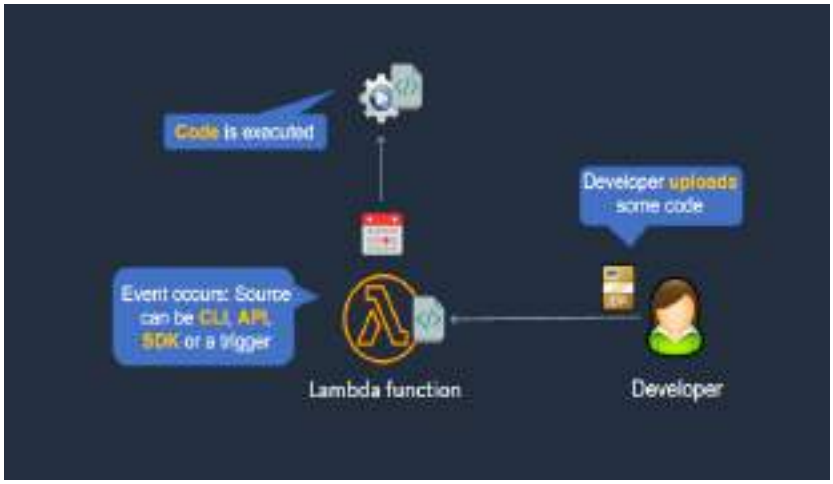
Serverless services include

- AWS Lambda
- AWS Fargate
- Amazon EventBridge
- AWS Step Functions
- Amazon SQS
- Amazon SNS
- Amazon API Gateway
- Amazon S3
- Amazon DynamoDB

8.19 AWS Lambda Functions

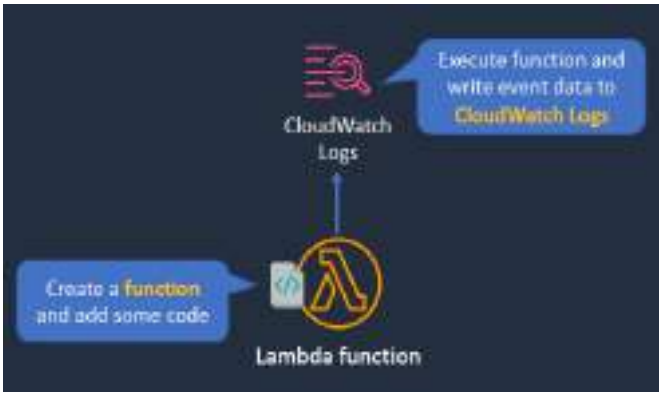
AWS Lambda Functions

- AWS Lambda executes code only when needed and scales automatically
- You pay only for the compute time you consume (you pay nothing when your code is not running)
- Benefits of AWS Lambda:
  - No servers to manage
  - Continuous scaling
  - Millisecond billing
  - Integrates with almost all other AWS services



- Primary use cases for AWS Lambda:
  - Data processing
  - Real-time file processing
  - Real-time stream processing
  - Build serverless backends for web, mobile, IOT, and 3rd party API requests

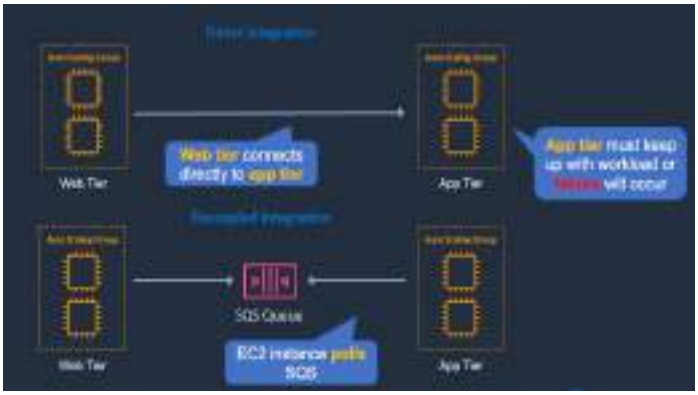
Create a Simple Lambda Function



8.20 Amazon Step Functions and Services

Amazon Simple Queue Service (SQS)

- SQS offers a reliable, highly-scalable, hosted queue for storing messages in transit between computers
- SQS is used for distributed/decoupled applications
- SQS uses a message-oriented API
- SQS uses pull based (polling) not push based



Amazon MQ

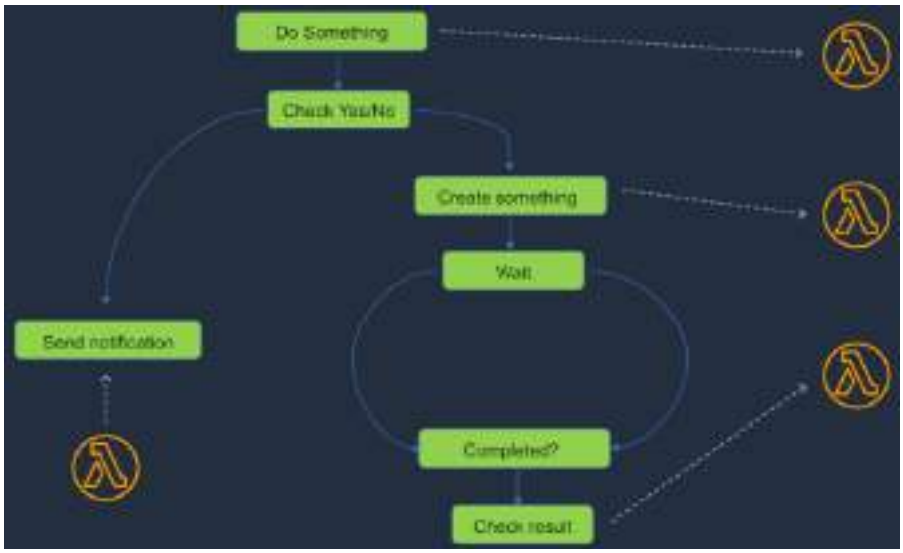
- Amazon MQ is a managed message broker service for Apache ActiveMQ and RabbitMQ that makes it easy to set up and operate message brokers on AWS.
- Message broker service
- Similar to Amazon SQS
- Based on Apache Active MQ and RabbitMQ
- Used when customers require industry standard APIs and protocols
- Useful when migrating existing queue-based applications into the cloud

Amazon Simple Notification Service (SNS)

- Amazon SNS is used for building and integrating loosely coupled, distributed applications
- Provides instantaneous, push-based delivery (no polling)
- Uses simple APIs and easy integration with applications
- Offered under an inexpensive, pay-as-you-go model with no up-front costs

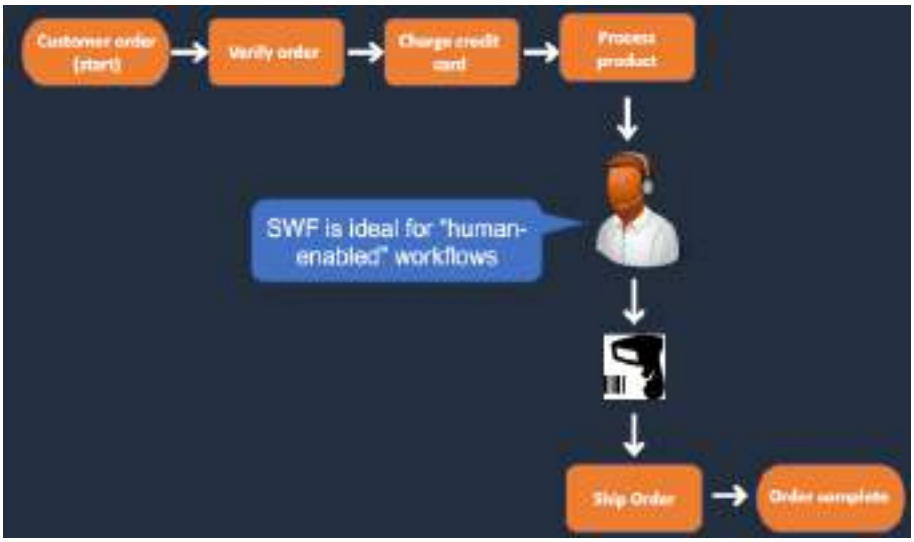
AWS Step Functions

- AWS Step Functions makes it easy to coordinate the components of distributed applications as a series of steps in a visual workflow
- You can quickly build and run state machines to execute the steps of your application in a reliable and scalable fashion



AWS Simple Workflow Service (SWF)

- Amazon Simple Workflow Service (SWF) is a web service that makes it easy to coordinate work across distributed application components.
- Create distributed asynchronous systems as workflows.
- Best suited for human-enabled workflows like an order fulfilment system or for procedural requests.
- AWS recommends that for new applications customers consider Step Functions instead of SWE.



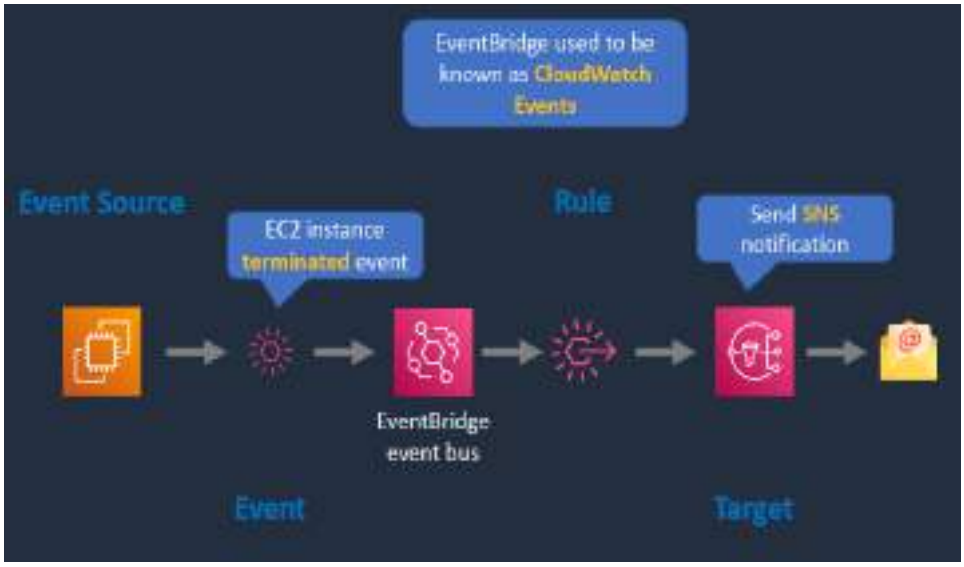
Application Integration Services Comparison

Service	What it does	Example use cases
Simple Queue Service	Messaging queue; store and forward patterns	Building distributed / decoupled applications
Simple Notification Service	Set up, operate, and send notifications from the cloud	Send email notification when CloudWatch alarm is triggered
Step Functions	Out-of-the-box coordination of AWS service components with visual workflow	Order processing workflow
Simple Workflow Service	Need to support external processes or specialized execution logic	Human-enabled workflows like an order fulfilment system or for procedural requests Note: AWS recommends that for new applications customers consider Step Functions instead of SWF
Amazon MQ	Message broker service for Apache Active MQ and RabbitMQ	Need a message queue that supports industry standard APIs and protocols; migrate queues to AWS

8.21 Amazon EventBridge

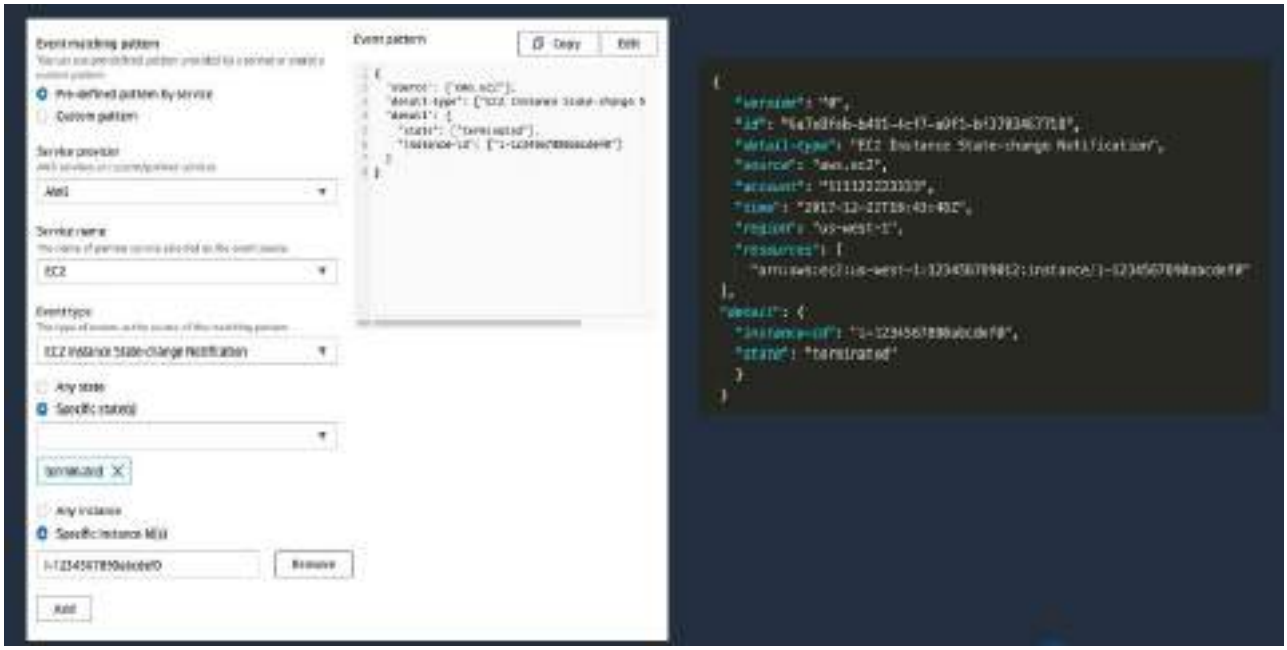
Amazon EventBridge

- Amazon EventBridge is a serverless event bus that makes it easier to build event-driven applications at scale using events generated from your applications, integrated Software-as-a-Service (SaaS) applications, and AWS services.

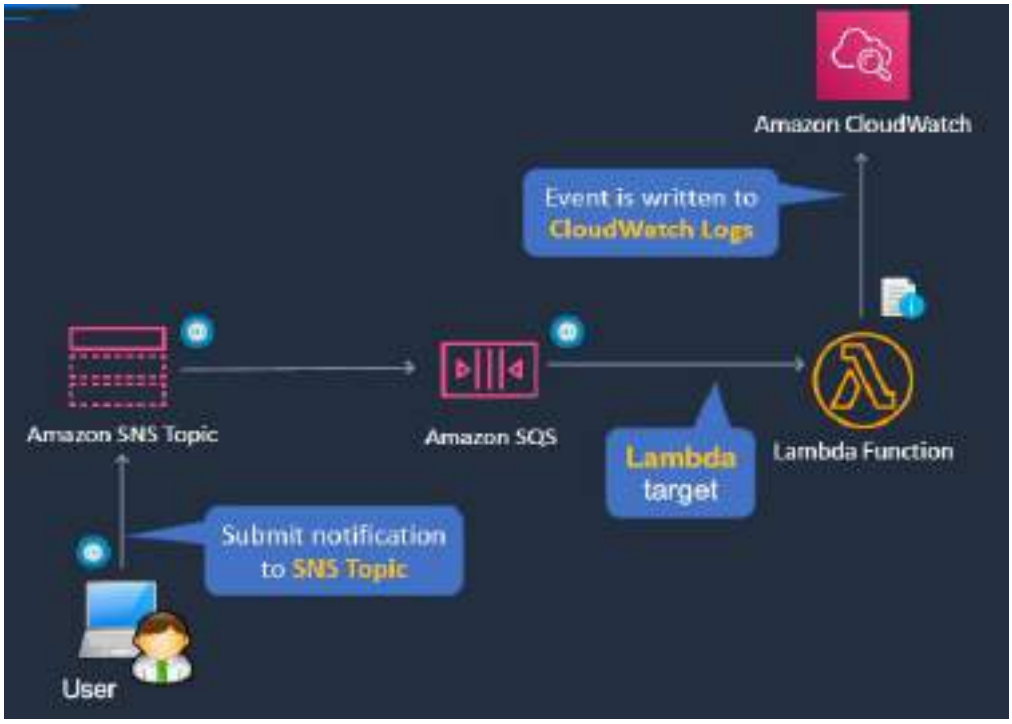


To create a custom event bus:

1. Open the Amazon EventBridge console at https://console.aws.amazon.com/events/.
2. In the navigation pane, choose Event buses.
3. Choose Create event bus.
4. Enter a name for the new event bus.
5. Do one of the following:
   - Enter the policy that includes the permissions to grant for the event bus. You can paste in a policy from another source or enter the JSON for the policy. You can use one of the example policies and modify it for your environment.
   - To use a template for the policy, choose Load template. Modify the policy as appropriate for your environment, including adding additional actions that you authorize the principal in the policy to use
6. Choose Create.

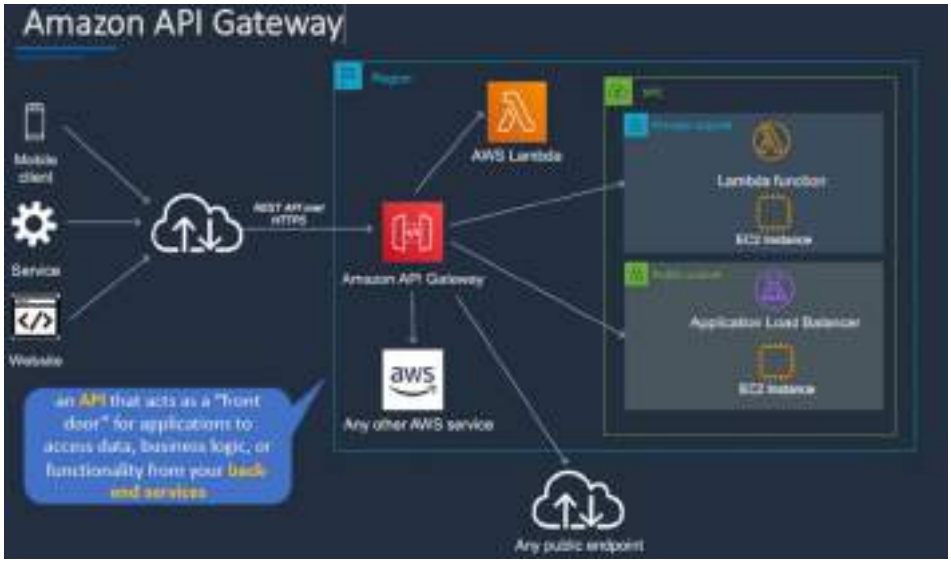


Simple Event-Driven Application



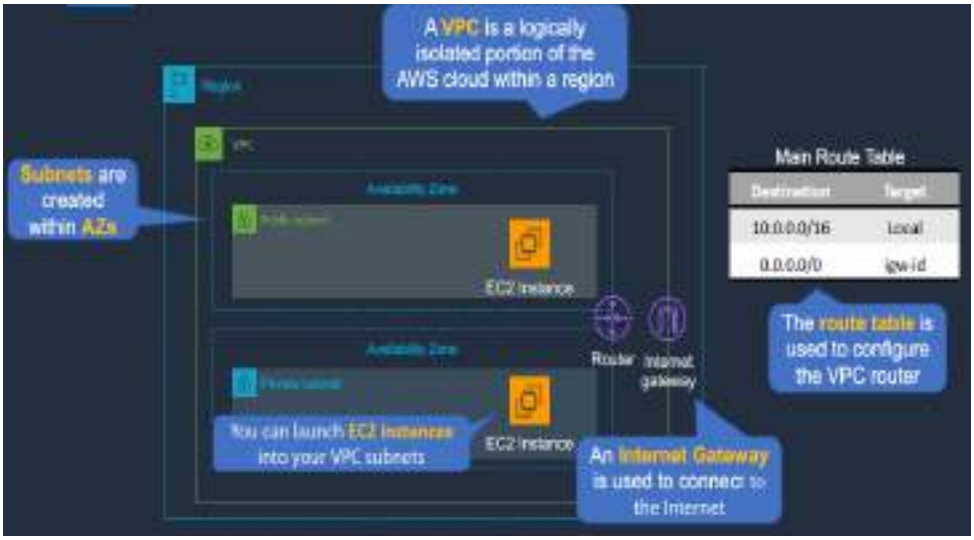
8.22 Amazon API Gateway

- Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale.
- APIs act as the "front door" for applications to access data, business logic, or functionality from your backend services.
- Amazon API Gateway is a closed-source software-as-a-service (SaaS) product written in Node.



8.23 Amazon Virtual Private Cloud (VPC)

- Amazon Virtual Private Cloud (VPC) is a service that lets you launch AWS resources in a logically isolated virtual network that you define.



- Amazon VPC enables you to build a virtual network in the AWS cloud - no VPNs, hardware, or physical datacenters required.
- You can define your own network space, and control how your network and the Amazon EC2 resources inside your network are exposed to the Internet.

- A virtual private cloud (VPC) is a virtual network dedicated to your AWS account
- Analogous to having your own DC inside AWS
- It is logically isolated from other virtual networks in the AWS Cloud
- Provides complete control over the virtual networking environment including selection of IP ranges, creation of subnets, and configuration of route tables and gateways
- You can launch your AWS resources, such as Amazon EC2 instances, into your VPC

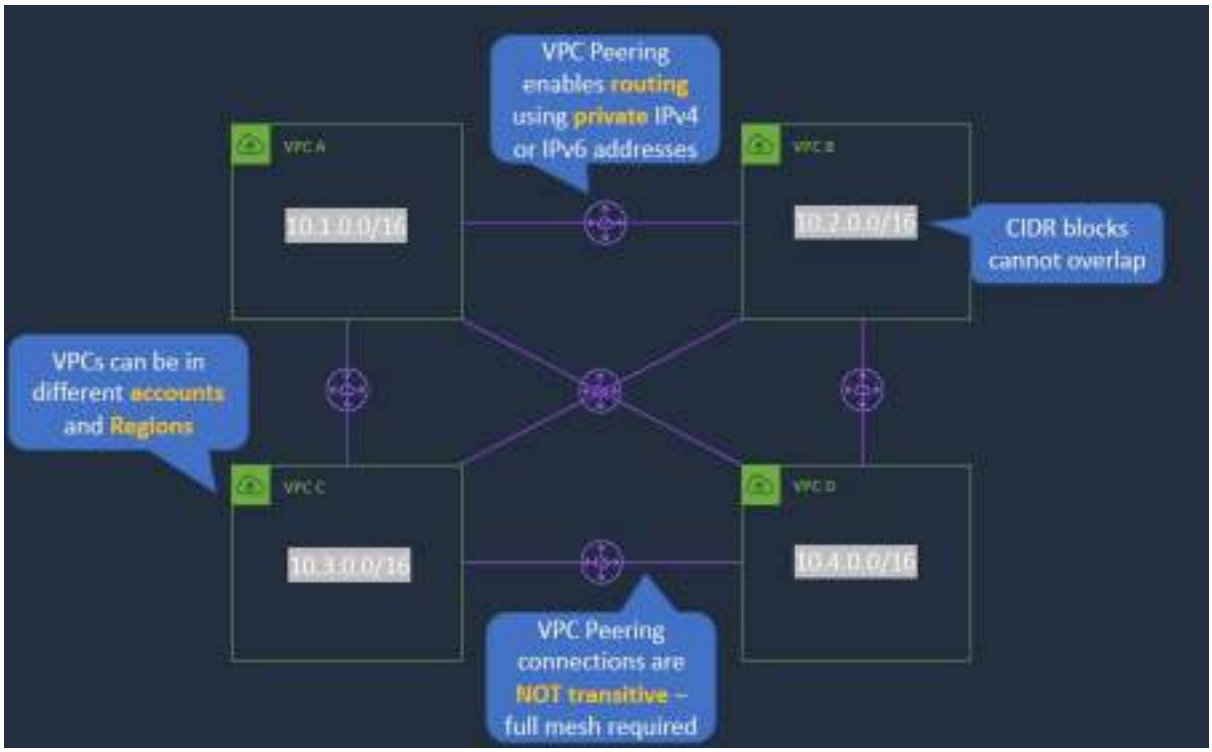
VPC Component	What it is
Virtual Private Cloud (VPC)	A logically isolated virtual network in the AWS cloud
Subnet	A segment of a VPC's IP address range where you can place groups of isolated resources
Internet Gateway/Egress-only Internet Gateway	The Amazon VPC side of a connection to the public Internet for IPv4/IPv6
Router	Routers interconnect subnets and direct traffic between Internet gateways, virtual private gateways, NAT gateways, and subnets
Peering Connection	Direct connection between two VPCs
VPC Endpoints	Private connection to public AWS services
NAT Instance	Enables Internet access for EC2 instances in private subnets (managed by you)
NAT Gateway	Enables Internet access for EC2 instances in private subnets (managed by AWS)
Virtual Private Gateway	The Amazon VPC side of a Virtual Private Network (VPN) connection
Customer Gateway	Customer side of a VPN connection
AWS Direct Connect	High speed, high bandwidth, private network connection from customer to aws
Security Group	Instance-level firewall
Network ACL	Subnet-level firewall

- When you create a VPC, you must specify a range of IPv4 addresses for the VPC in the form of a Classless Inter-Domain Routing (CIDR) block; for example, 10.0.0.0/16
- A VPC spans all the Availability Zones in the region
- You have full control over who has access to the AWS resources inside your VPC
- By default, you can create up to 5 VPCs per region
- A default VPC is created in each region with a subnet in each AZ

Create a Custom VPC

VPC Peering

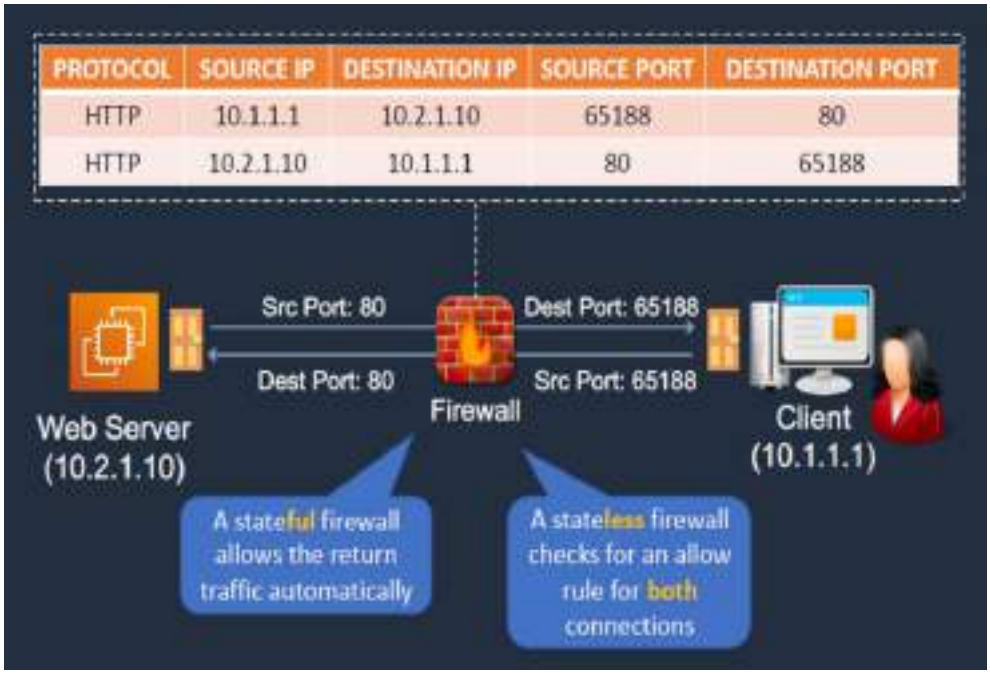
A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses.



8.24 Security Groups and Network ACLs

Stateful vs Stateless Firewalls

- Stateful firewalls are capable of monitoring and detecting states of all traffic on a network to track and defend based on traffic patterns and flows.
- Stateless firewalls, however, only focus on individual packets, using preset rules to filter traffic.

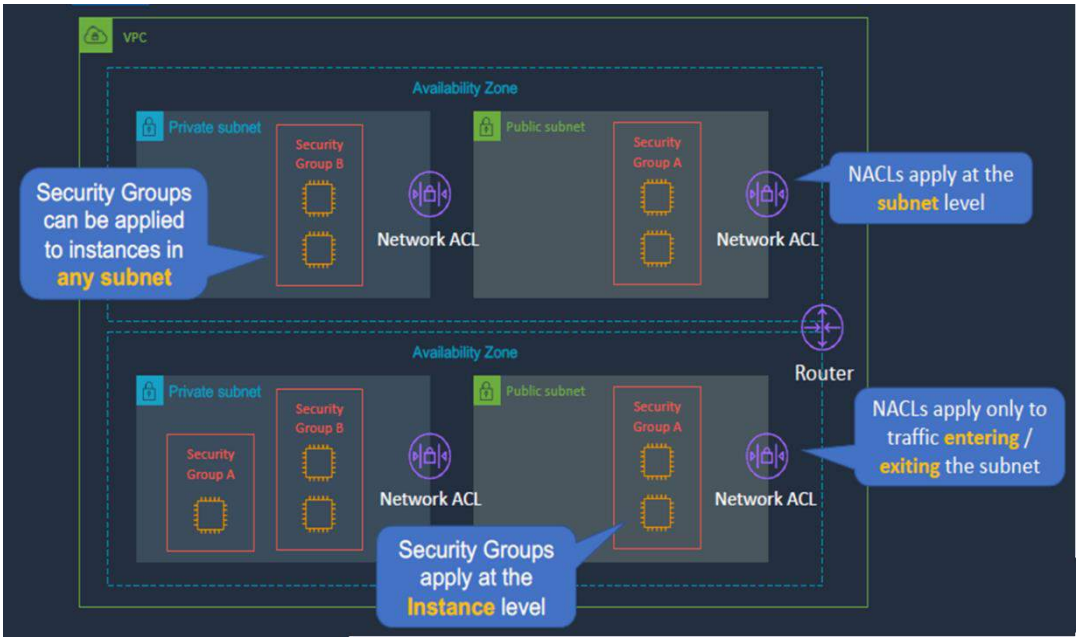


Network ACLs

- An optional layer of security that acts as a firewall for controlling traffic in and out of a subnet.
- You can associate multiple subnets with a single network ACL, but a subnet can be associated with only one network ACL at a time.

Security Groups and Network ACLs

- Security groups are tied to an instance whereas Network ACLs are tied to the subnet.
- Network ACLs are applicable at the subnet level, so any instance in the subnet with an associated NACL will follow rules of NACL.
- That's not the case with security groups, security groups has to be assigned explicitly to the instance.



Security Group Rules

- Security group rules enable you to filter traffic based on protocols and port numbers.
- Security groups are stateful—if you send a request from your instance, the response traffic for that request is allowed to flow in regardless of inbound security group rules.

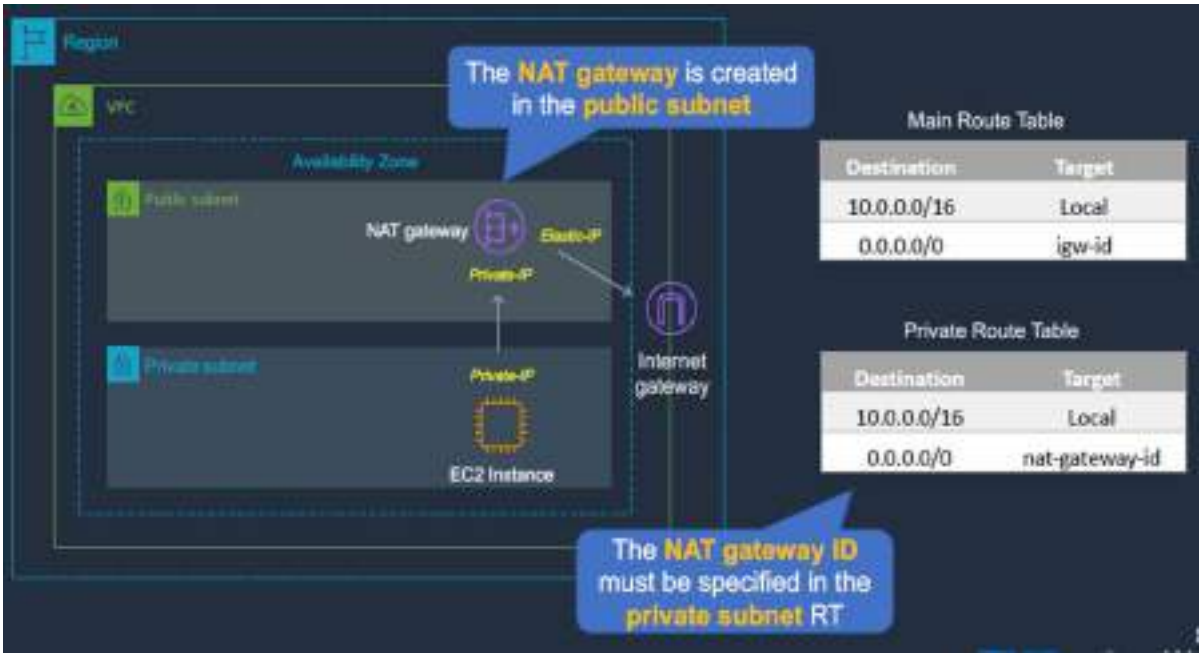
8.25 Working with IP Addresses

Public, Private and Elastic IP addresses

Name	Description
Public IP address	Lost when the instance is stopped Used in Public Subnets No charge Associated with a private IP address on the instance Cannot be moved between instances
Private IP address	Retained when the instance is stopped Used in Public and Private Subnets
Elastic IP address	Static Public IP address You are charged if not used Associated with a private IP address on the instance Can be moved between instances and Elastic Network Adapters

NAT Gateways

A NAT gateway is a Network Address Translation (NAT) service. You can use a NAT gateway so that instances in a private subnet can connect to services outside your VPC but external services cannot initiate a connection with those instances.



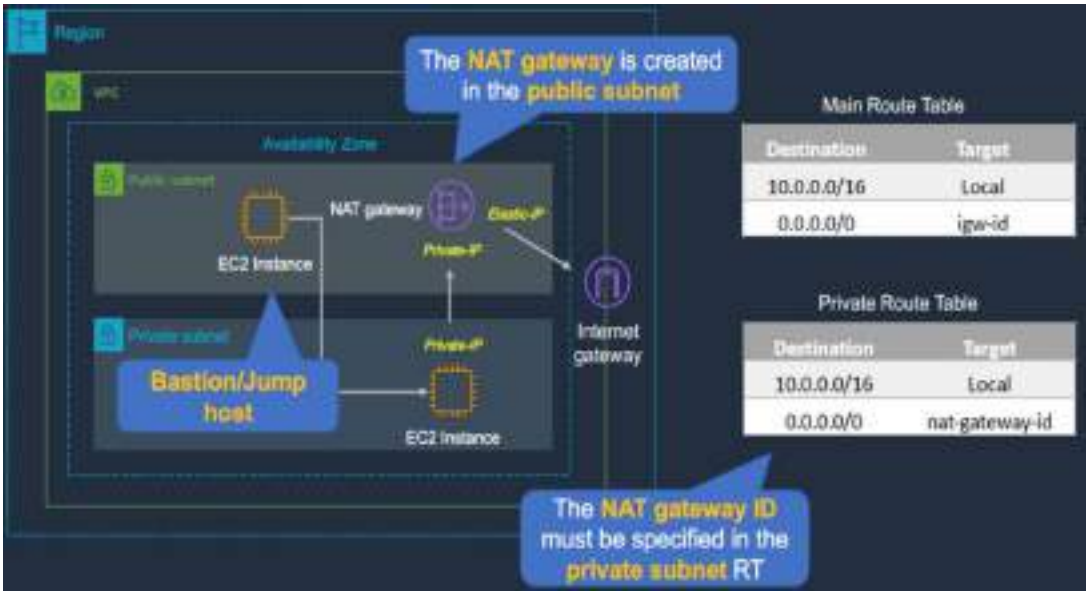
NAT Instances

A NAT (Network Address Translation) instance is, like a bastion host, an EC2 instance that lives in your public subnet. A NAT instance, however, allows your private instances outgoing connectivity to the internet while at the same time blocking inbound traffic from the internet.

NAT Instance vs NAT Gateway

NAT Instance	NAT Gateway
Managed by you (e.g. software updates)	Managed by AWS
Scale up (instance type) manually and use enhanced networking	Elastic scalability up to 45 Gbps
No high availability – scripted/auto-scaled HA possible using multiple NATs in multiple subnets	Provides automatic high availability within an AZ and can be placed in multiple AZs

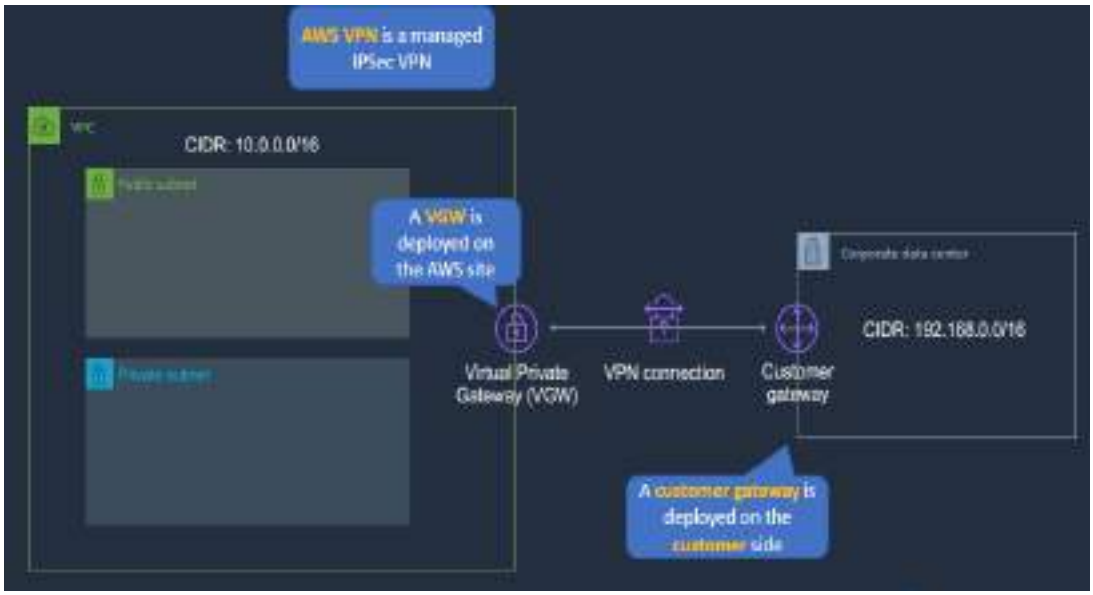
Deploy a NAT Gateway



8.26 Amazon VPN, Direct Connect, Gateway and Outposts

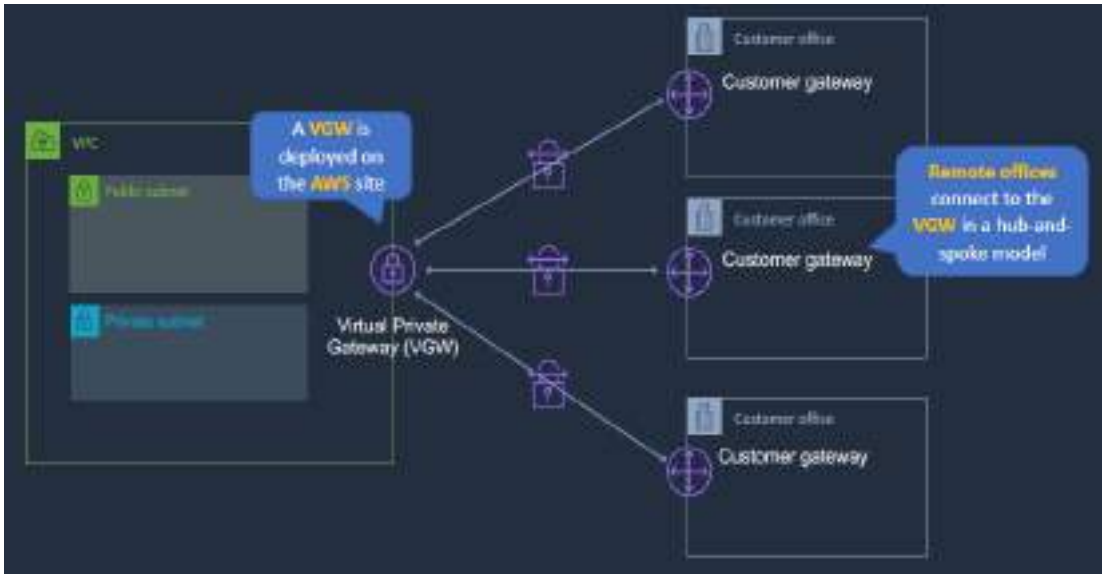
AWS Site-to-Site VPN

AWS Site-to-Site VPN is a fully-managed service that creates a secure connection between your data center or branch office and your AWS resources using IP Security (IPSec) tunnels.



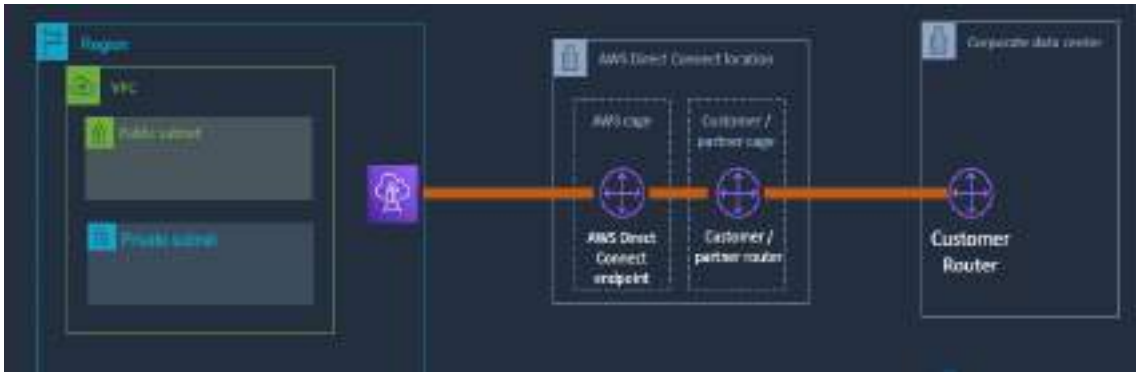
AWS VPN CloudHub

- AWS VPN CloudHub is a hub-and-spoke VPN technology offered by AWS.
- CloudHub allows your remote sites to communicate with one another over VPN tunnels that are created between your AWS Virtual Private Gateway (VPG) and your remote sites.



AWS Direct Connect

- Private connectivity between AWS and your data center/office.
- Consistent network experience – increased speed/latency & bandwidth/throughput
- Lower costs for organizations that transfer large volumes of data

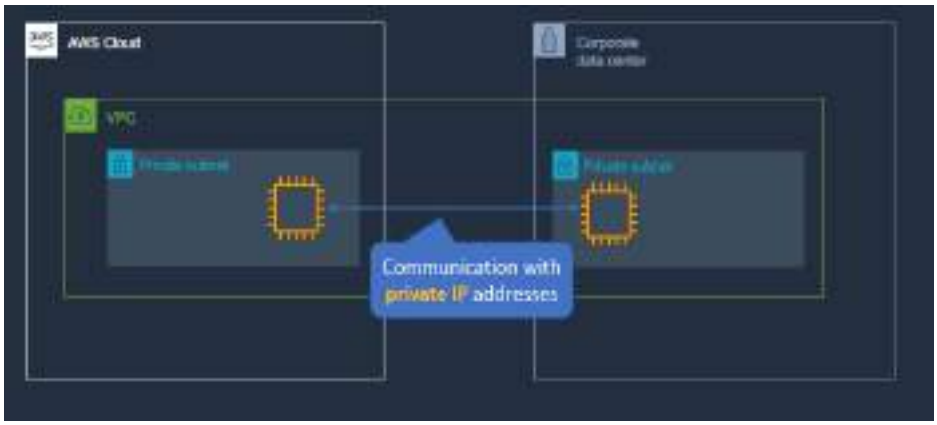


AWS Transit Gateway

- AWS Transit Gateway connects your Amazon Virtual Private Clouds (VPCs) and on-premises networks through a central hub.
- This simplifies your network and puts an end to complex peering relationships.
- It acts as a cloud router – each new connection is only made once.

AWS Outposts

- An Outpost is a pool of AWS compute and storage capacity deployed at a customer site.
- AWS operates, monitors, and manages this capacity as part of an AWS Region.
- You can create subnets on your Outpost and specify them when you create AWS resources such as EC2 instances, EBS volumes, ECS clusters, and RDS instances.



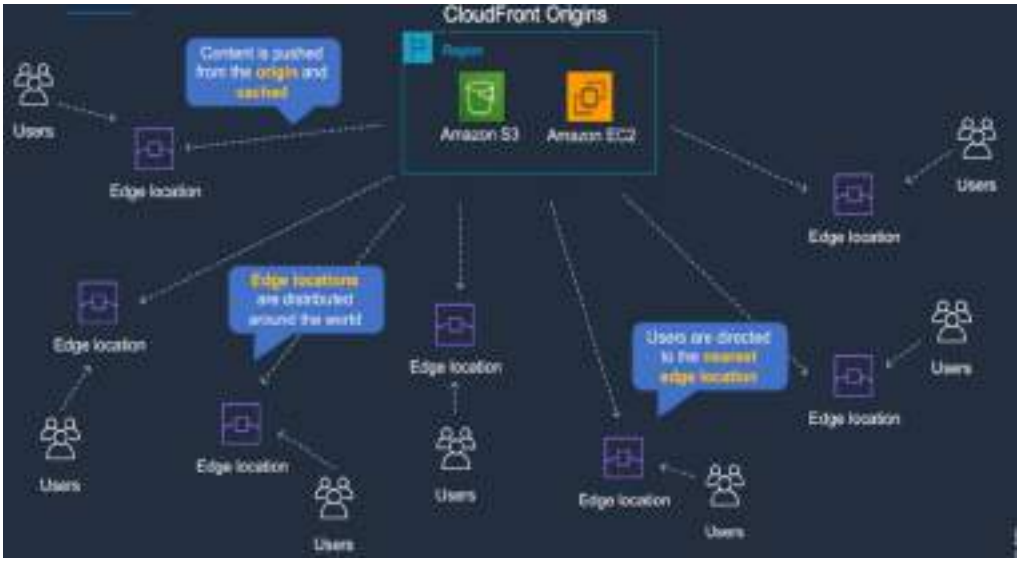
Services you can run on AWS Outposts include:

- Amazon EC2
- Amazon EBS
- Amazon S3
- Amazon VPC
- Amazon ECS/EKS
- Amazon RDS
- Amazon EMR

8.27 CloudFront, Global Accelerator and Cloud Formation

Amazon CloudFront

- Amazon CloudFront is a content delivery network operated by Amazon Web Services.
- Content delivery networks provide a globally-distributed network of proxy servers that cache content, such as web videos or other bulky media, more locally to consumers, thus improving access speed for downloading the content.
- This web service speeds up distribution of your static and dynamic web content, such as html, css, js, and image files, to your users.



AWS Global Accelerator

- AWS Global Accelerator combines advanced networking features with the dedicated AWS Global Network to improve your application network performance by up to 60%.
- AWS Global Accelerator simplifies global traffic management by providing 2 static anycast IP addresses that only need to be configured by users once.
- Behind these IP address you can add or remove AWS origins, opening up uses such as endpoint failover, scaling, or testing without any user-side changes.

AWS Global Accelerator vs CloudFront

- Both use the AWS global network and edge locations
- CloudFront improves performance for cacheable content and dynamic content
- GA improves performance for a wide range of applications over TCP and UDP
- GA proxy's connections to applications in one or more AWS Regions
- GA provides failover between AWS Regions

AWS CloudFormation

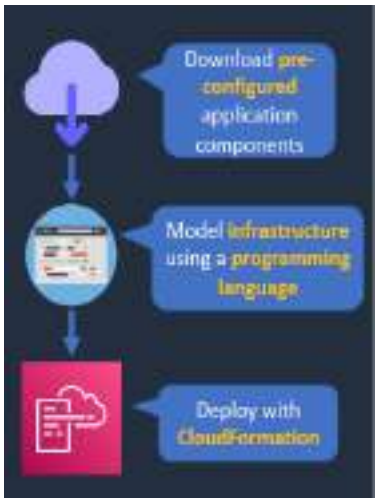
- AWS CloudFormation is an infrastructure as code (IaC) service that allows you to easily model, provision, and manage AWS and third-party resources.
- It gives developers and businesses an easy way to create a collection of related AWS and third-party resources, and provision and manage them in an orderly and predictable fashion.

- Infrastructure is provisioned consistently, with fewer mistakes (human error)
- Less time and effort than configuring resources manually
- Free to use (you're only charged for the resources provisioned)
- A template is a YAML or JSON template used to describe the endstate of the infrastructure you are either provisioning or changing
- CloudFormation creates a Stack based on the template
- Can easily rollback and delete the entire stack as well

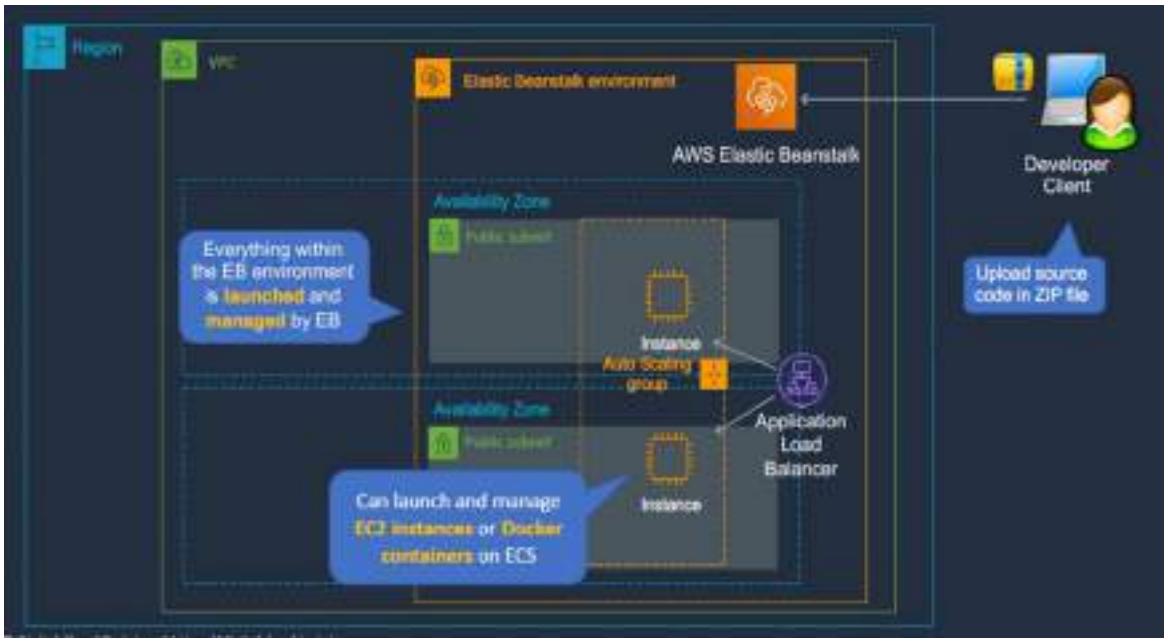
8.28 AWS Cloud Development Kit and Elastic Beanstalk

AWS Cloud Development Kit (CDK)

- Open-source software development framework to define your cloud application resources using familiar programming languages
- Preconfigures cloud resources with proven defaults using constructs
- Provisions your resources using AWS CloudFormation
- Enables you to model application infrastructure using TypeScript, Python, Java, and .NET
- Use existing IDE, testing tools, and workflow patterns



AWS Elastic Beanstalk



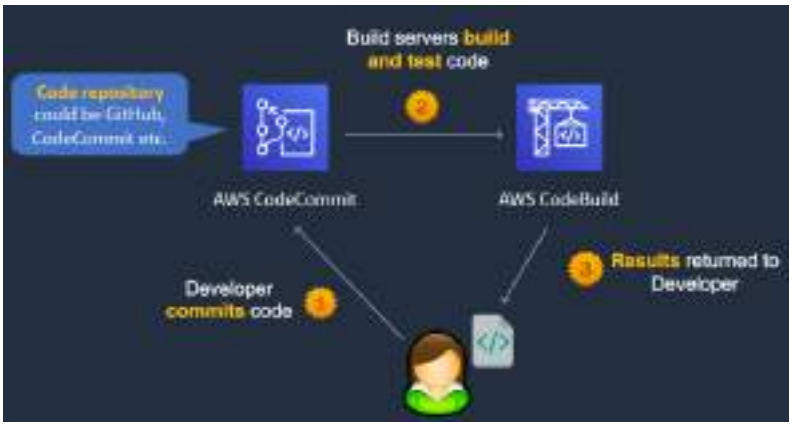
- Supports Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker web applications
- Integrates with VPC
- Integrates with IAM
- Can provision most database instances
- Allows full control of the underlying resources
- Code is deployed using a WAR file or Git repository

CloudFormation	Elastic Beanstalk
"Template-driven provisioning"	"Web apps made easy"
Deploys infrastructure using code	Deploys applications on EC2 (PaaS)
Can be used to deploy almost any AWS service	Deploys web applications based on Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker
Uses JSON or YAML template files	Uses ZIP or WAR files (or Git)
Similar to Terraform	Similar to Google App Engine

8.29 AWS Developer Tools (Code*)

Continuous Integration

- Continuous integration is a DevOps software development practice where developers regularly merge their code changes into a central repository, after which automated builds and tests are run.
- Continuous integration most often refers to the build or integration stage of the software release process and entails both an automation component (e.g. a CI or build service) and a cultural component (e.g. learning to integrate frequently).
- The key goals of continuous integration are to find and address bugs quicker, improve software quality, and reduce the time it takes to validate and release new software updates.

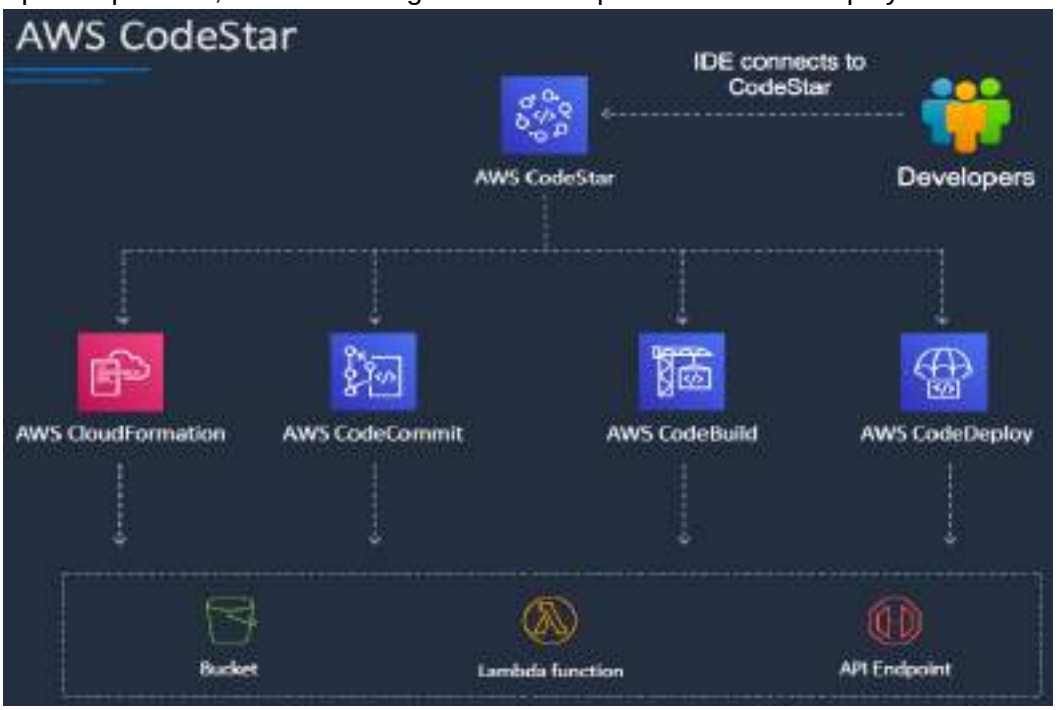


Continuous Delivery and Continuous Deployment

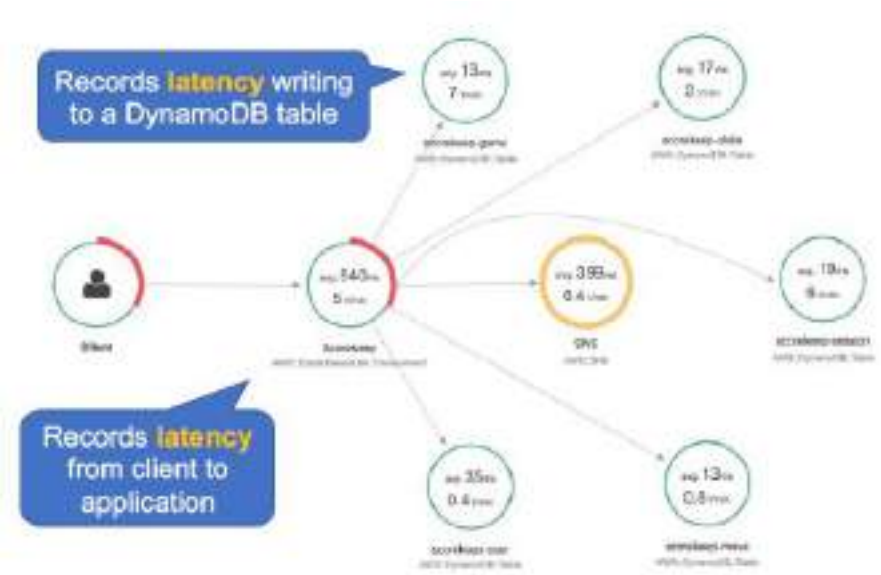
- With continuous delivery, code changes are automatically built, tested, and prepared for a release to production.
- Continuous delivery is an extension of continuous integration since it automatically deploys all code changes to a testing and/or production environment after the build stage.
- AWS CodePipeline is a fully managed continuous delivery service that helps you automate your release pipelines for fast and reliable application and infrastructure updates.
- The difference between continuous delivery and continuous deployment is the presence of a manual approval to update to production.
- With continuous deployment, production happens automatically without explicit approval.

AWS CodeStar

- AWS CodeStar provides the tools you need to quickly develop, build, and deploy applications on AWS.
- With AWS CodeStar, you can set up your entire continuous delivery toolchain in minutes, allowing you to start releasing code faster.
- AWS CodeStar makes it easy for your whole team to work together securely, with built-in role-based policies that allow you to easily manage access and add owners, contributors, and viewers to your projects.
- With the AWS CodeStar project dashboard, you can easily track your entire software development process, from a backlog work item to production code deployment.



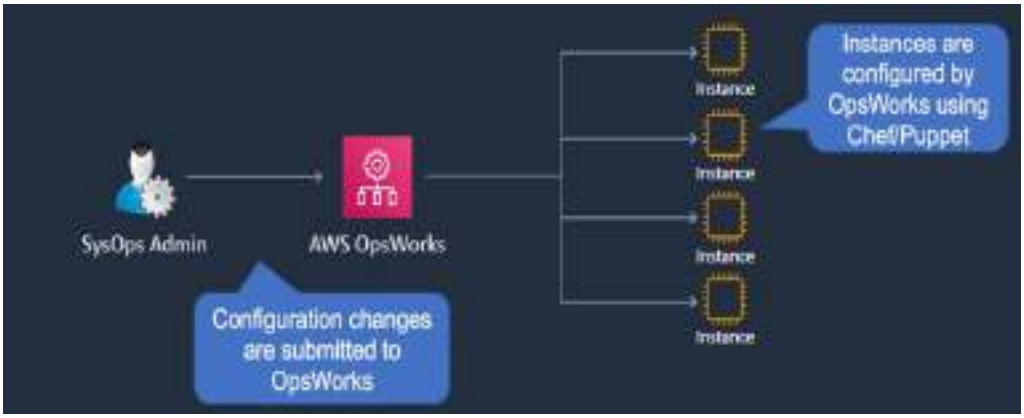
8.30 AWS X-Ray and OpsWorks



- AWS X-Ray helps developers analyze and debug production, distributed applications, such as those built using a microservices architecture.
- AWS X-Ray supports applications running on:
  - Amazon EC2
  - Amazon ECS
  - AWS Lambda
  - AWS Elastic Beanstalk
- Need to integrate the X-Ray SDK with your application and install the X-Ray agent.

AWS OpsWorks

- AWS OpsWorks is a configuration management service that provides managed instances of Chef and Puppet.
- Updates include patching, updating, backup, configuration and compliance management.



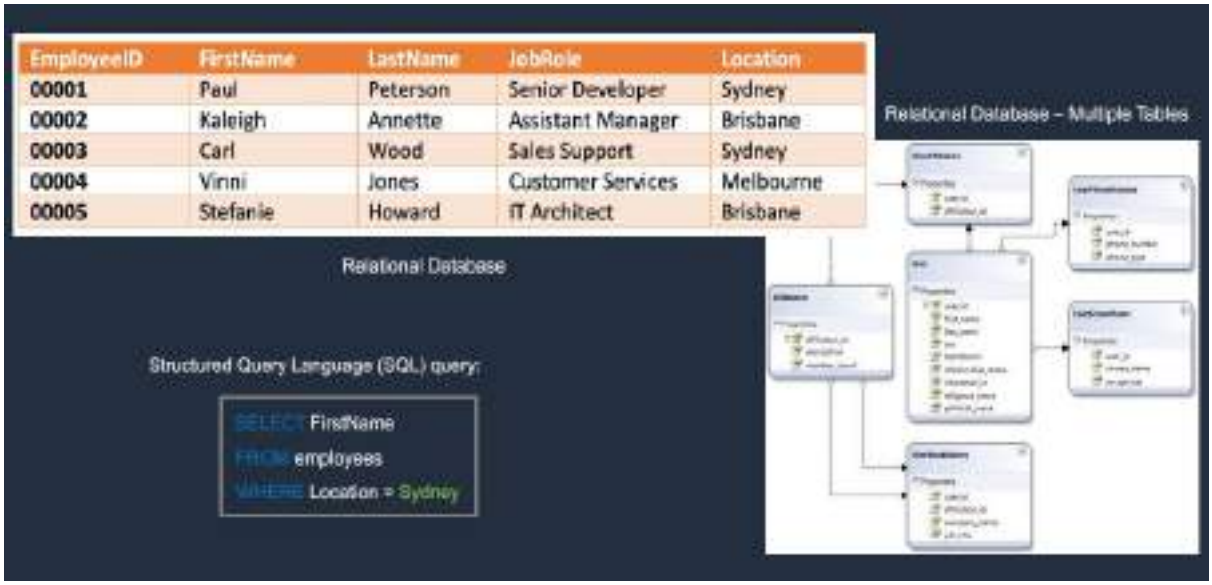
8.31 Types of Databases

Relational vs non-Relational

- Key differences are how data are managed and how data are stored

Relational	Non-Relational
Organized by tables, rows and columns	Varied data storage models
Rigid schema (SQL)	Flexible schema (NoSQL) - data stored in key-value pairs, columns, documents or graphs
Rules enforced within database	Rules can be defined in application code (outside database)
Typically scaled vertically	Scales horizontally
Supports complex queries and joins	Unstructured, simple language that supports any kind of schema
Amazon RDS, Oracle, MySQL, IBM DB2, PostgreSQL	Amazon DynamoDB, MongoDB, Redis, Neo4j

Relational Database



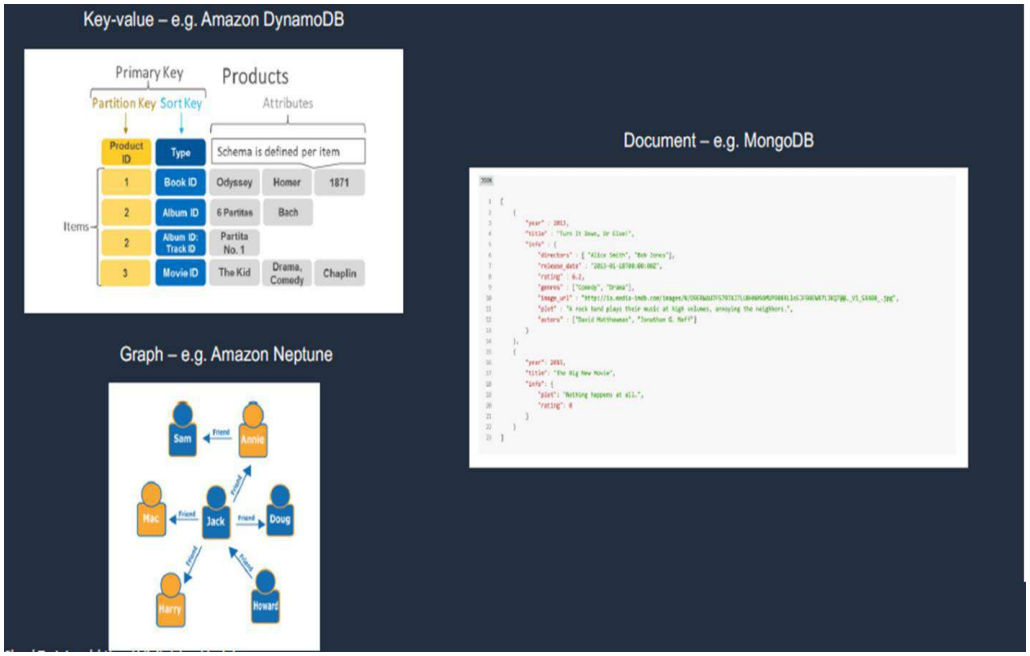
Types of Non-Relational DB (NoSQL)

Key-value stores

- The simplest type of NoSQL database is a key-value store. Every data element in the database is stored as a key value pair consisting of an attribute name (or "key") and a value.
- In a sense, a key-value store is like a relational database with only two columns: the key or attribute name (such as "state") and the value (such as "Alaska").
- Use cases include shopping carts, user preferences, and user profiles.

Document

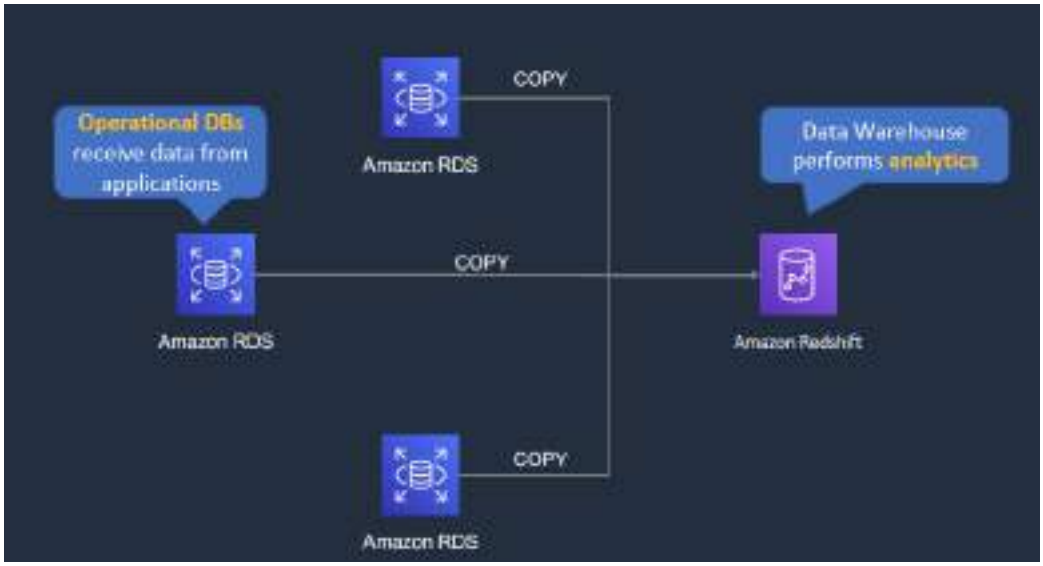
- A document database stores data in JSON, BSON, or XML documents (not Word documents or Google Docs, of course).
- In a document database, documents can be nested. Particular elements can be indexed for faster querying.



Operational vs Analytical

Key differences are use cases and how the database is optimized

Operational / transactional	Analytical
Online Transaction Processing (OLTP)	Online Analytics Processing (OLAP) – the source data comes from OLTP DBS
Production DBs that process transactions. E.g. adding customer records, checking stock availability (INSERT, UPDATE, DELETE)	Data warehouse. Typicaly, separated from the customer facing DBs. Data is extracted for decision making
Short transactions and simple queries	Long transactions and complex queries
Relational examples: Amazon RDS, Oracle, IBM DB2, MySQL	Relational examples: Amazon RedShift, Teradata, HP Vertica
Non-relational examples: MongoDB, Cassandra, Neo4j, HBase	Non-relational examples: Amazon EMR, MapReduce

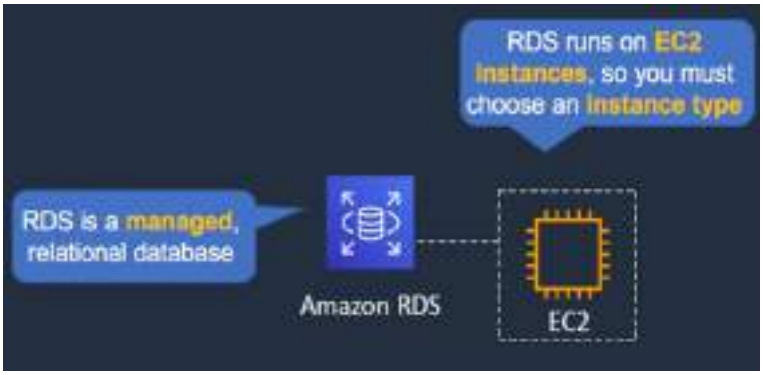


AWS Databases

Data Store	Use Case
Database on EC2	▪ Need full control over instance and database. ▪ Third-party database engine (not available in RDS)
Amazon RDS	▪ Need traditional relational database ▪ e.g. Oracle, PostgreSQL, Microsoft SQL, MariaDB, MySQL ▪ Data is well-formed and structured
Amazon DynamoDB	▪ NoSQL database ▪ In-memory performance ▪ High 1/0 needs High ▪ Dynamic scaling
Amazon Redshift	▪ Data warehouse for large volumes of aggregated data
Amazon ElastiCache	▪ Fast temporary storage for small amounts of data ▪ In-memory database
Amazon EMR	▪ Analytics workloads using the Hadoop framework

Amazon Relational Database Service (RDS)

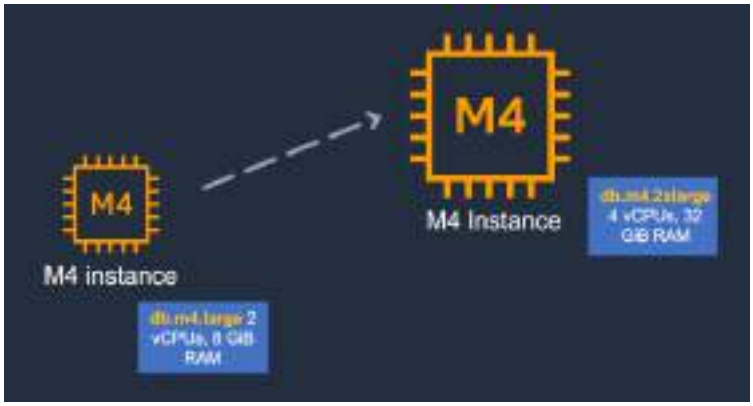
- Amazon Relational Database Service (Amazon RDS) is a collection of managed services that makes it simple to set up, operate, and scale databases in the cloud.
- RDS supports the following database engines:
  - Amazon Aurora
  - MySQL
  - MariaDB
  - Oracle
  - Microsoft SQL Server
  - PostgreSQL



- Scales up by increasing instance size (Compute and Storage)
- Disaster recovery with multi-AZ option.
- RDS uses EC2 instances, so you must choose an instance family/type
- Relational databases are known as Structured Query Language (SQL) databases
- RDS is an Online Transaction Processing (OLTP) type of database
- Easy to setup, highly available, fault tolerant, and scalable
- Common use cases include online stores and banking systems
- You can encrypt your Amazon RDS instances and snapshots at rest by enabling the encryption option for your Amazon RDS DB instance (during creation)
- Encryption uses AWS Key Management Service (KMS)

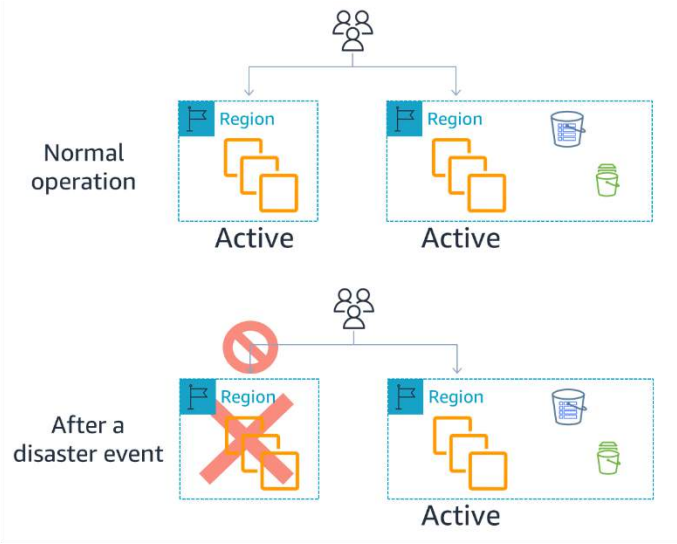
Amazon RDS Scaling Up (vertically)

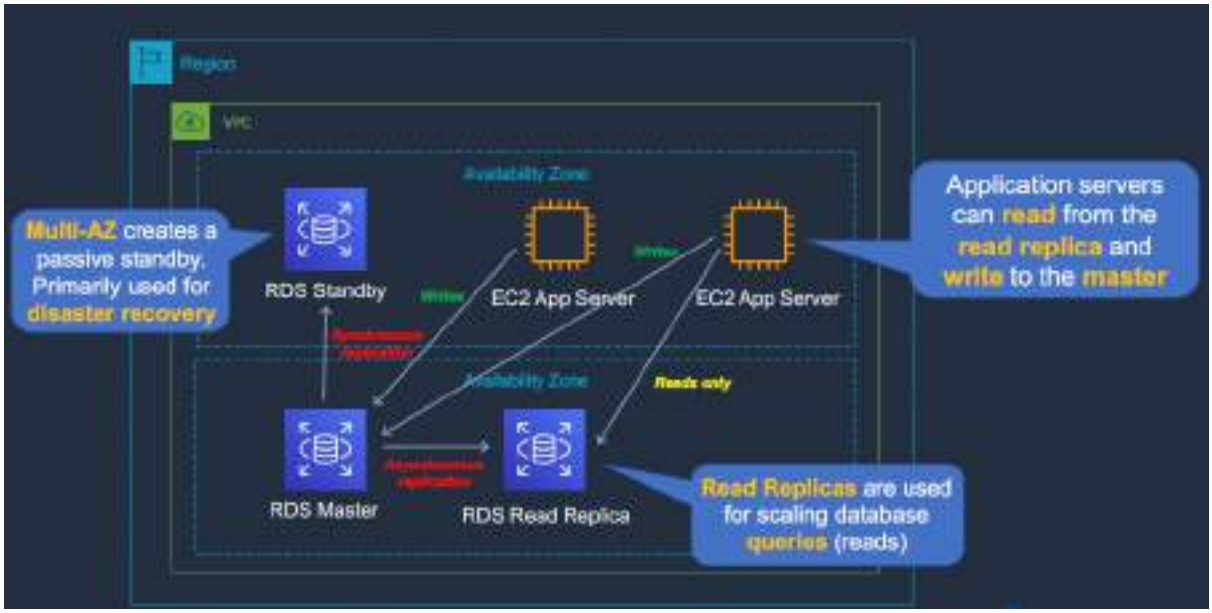
- You can vertically scale up your RDS instance with a click of a button.
- Several instance sizes are available, from general purpose to CPU and memory optimized, when resizing in Amazon RDS for MySQL, Amazon RDS for PostgreSQL, Amazon RDS for Maria DB, Amazon RDS for Oracle, or Amazon RDS for SQL Server.



Disaster Recovery (DR) and Scaling Out (Horizontally)

- Disaster recovery is the process by which an organization anticipates and addresses technology-related disasters.
- IT systems in any company can go down unexpectedly due to unforeseen circumstances, such as power outages, natural events, or security issues.

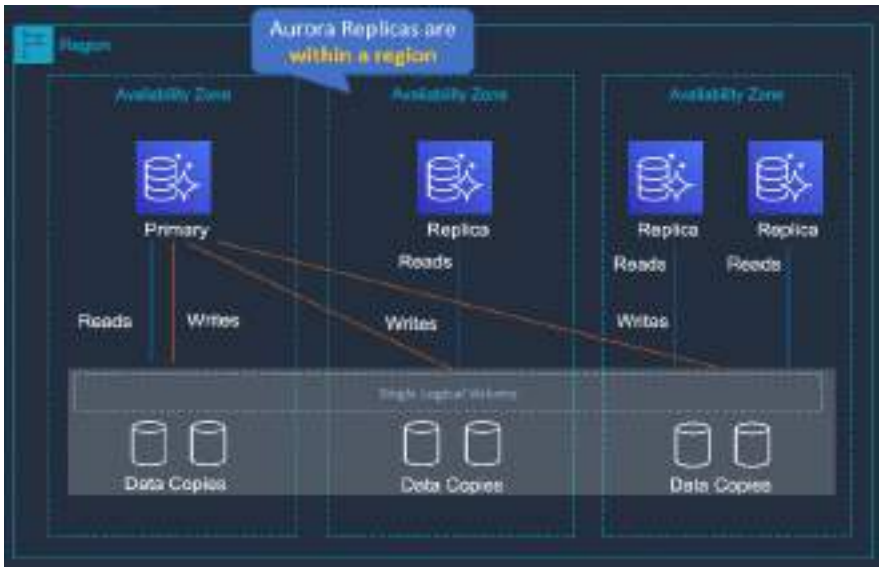




8.32 Amazon Aurora, Dynamo DB, Redshift, Elastic Map Reduce and ElastiCache

Amazon Aurora

- Amazon Aurora is an AWS database offering in the RDS family.
- Amazon Aurora is a MySQL and PostgreSQL-compatible relational database built for the cloud.
- Amazon Aurora is up to five times faster than standard MySQL databases and three times faster than standard PostgreSQL databases.
- Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 64TB per database instance.



Aurora Fault Tolerance

- Fault tolerance across 3 AZs
- Single logical volume
- Aurora Replicas scale-out read requests
- Can promote Aurora Replica to be a new primary or create new primary
- Can use Auto Scaling to add replicas
- Aurora Replicas are within a region

Amazon Aurora Key Features

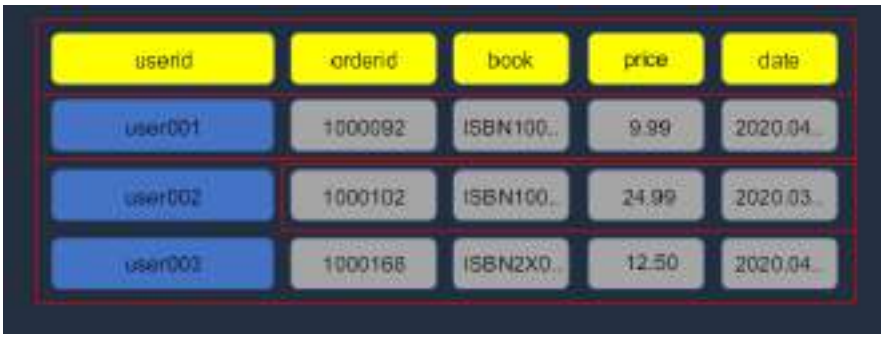
Aurora Feature	Benefit
High performance and scalability	Offers high performance, self-healing storage that scales up to 64TB, point-in-time recovery and continuous backup to $3
DB compatibility	Compatible with existing MySQL and PostgreSQL open source databases
Aurora Replicas	In-region read scaling and failover target - up to 15 (can use Auto Scaling)
MySQL Read Replicas	Cross-region cluster with read scaling and failover target - up to 5 (each can have up to 15 Aurora Replicas)
Global Database	Cross-region cluster with read scaling (fast replication/low latency reads). Can remove secondary and promote
Multi-Master	Scales out writes within a region. In preview currently and will not appear on the exam
Serverless	On-demand, autoscaling configuration for Amazon Aurora - does not support read replicas or public IPs (can only access through VPC or Direct Connect - not VPN)

Amazon DynamoDB

- Fully managed NoSQL database service
- Key/value store and document store
- It is a non-relational, key-value type of database
- Fully serverless service
- Push button scaling

Dynamo DB is made up of:

- Tables
- Items
- Attributes

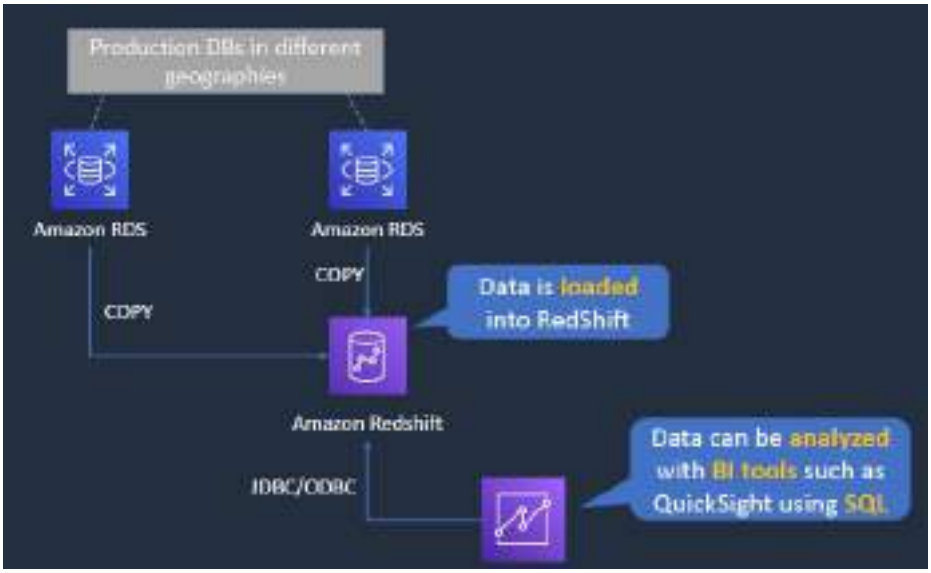


Amazon DynamoDB Key Features

DynamoDB Feature	Benefit
Serverless	Fully managed, fault tolerant, service
Highly available	99.99% availability SLA-99.999% for Global Tables!
NoSQL type of database with Name / Value Structure	Flexible schema, good for when data is not well structured or unpredictable
Horizontal scaling	Seamless scalability to any scale with push button scaling or Auto Scaling
Dynamo DB Accelerator (DAX)	Fully managed in-memory cache for DynamoDB that increases performance (microsecond latency)
Backup	Point-in-time recovery down to the second in last 35 days; On-demand backup and restore
Global Tables	Fully managed multi-region, multi-master solution

Amazon RedShift

- Amazon Redshift is a data warehouse product which forms part of the larger cloud-computing platform Amazon Web Services.
- It is built on top of technology from the massive parallel processing data warehouse company ParAccel, to handle large scale data sets and database migrations.
- Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud.
- You can start with just a few hundred gigabytes of data and scale to a petabyte or more. This enables you to use your data to acquire new insights for your business and customers.



AWS Lambda Functions

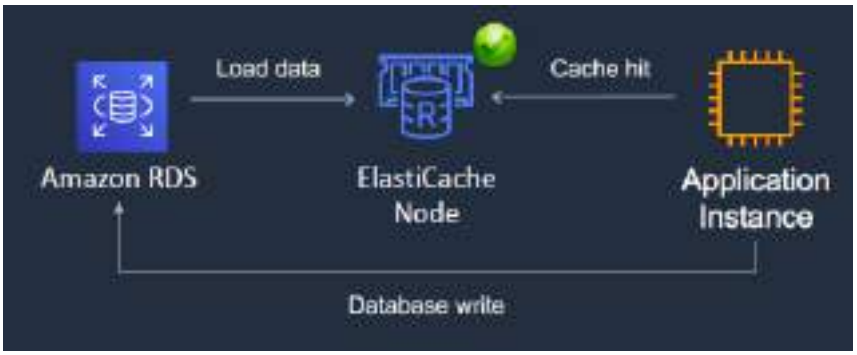
- Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost-effective to analyze all your data using standard SQL and existing Business Intelligence (BI) tools
- RedShift is a SQL based data warehouse used for analytics applications
- RedShift is a relational database that is used for Online Analytics Processing (OLAP) use cases
- RedShift uses Amazon EC2 instances, so you must choose an instance family/type
- RedShift always keeps three copies of your data
- RedShift provides continuous/incremental backups

Amazon Elastic Map Reduce (EMR)

- Managed cluster platform that simplifies running big data frameworks including Apache Hadoop and Apache Spark
- Used for processing data for analytics and business intelligence
- Can also be used for transforming and moving large amounts of data
- Performs extract, transform, and load (ETL) functions

Amazon ElastiCache

- Fully managed implementations Redis and Memcached
- ElastiCache is a key/value store
- In-memory database offering high performance and low latency
- Can be put in front of databases such as RDS and DynamoDB Application



- ElastiCache nodes run on Amazon EC2 instances, so you must choose an instance family/type

Use Case	Benefit
Web session store	In cases with load balanced web servers, store web session information in Redis so if a server is lost, the session info is not lost, and another web server can pick it up
Database caching	Use Memcached in front of AWS RDS to cache popular queries to offload work from RDS and return results faster to users
Leaderboards	Use Redis to provide a live leaderboard for millions of users of your mobile app
Streaming data dashboards	Provide a landing spot for streaming sensor data on the factory floor, providing live real-time dashboard displays

Section 3: Exercises

Exercise 1: Tabulate difference between Traditional IT and Cloud Computing.

Exercise 2: Fill the following table.

S3 Capability	What it Does
Transfer	
Requester Pays	
Events	
Static Web Hosting	
Versioning and Replication	

Exercise 3: Draw mechanism of various types of Load Balancers.

Exercise 4: Tabulate Difference between NAT Instance and NAT Gateway.

Exercise 5: Participate in a group discussion on following topics:

a) Functions of AWS (AMAZON WEB SERVICES)
b) AWS Global Infrastructure
c) Explain AWS Shared Responsibility Model
d) Identity and Access Management (AWS IAM)
e) Explain AWS Compute Services
f) Amazon Simple Storage Service (S3)
g) Types of databases

Section 4: Assessment Questionnaire

1. What are the advantages of AWS?
2. The instances in the _____subnet can send outbound traffic directly to the internet, whereas the instances in the _____ subnet can't.
3. What are the advantages of cloud computing?
4. What is the use of AWS Identity and Access Management (IAM)?
5. The _____ represents the person or service who uses the IAM user to interact with AWS.
6. An IAM user group is a collection of IAM users. (True/False)
7. What is IAM role?
8. Policies are documents that define permissions and are written in:
9. Authentication occurs whenever a user attempts to access your organization's network and downstream resources. (True/False)
10. What is multi-factor authentication (MFA) in AWS?
11. Fill in the blanks:
    a) CPU is measured in _____
    b) RAM is measured in _____
    c) HDD is measured in _____
    d) NIC is measured in _____
12. What are the methods to launch an EC2 instance?
13. What are the benefits of EC2?
14. What are the two types of user data to Amazon EC2?
15. Docker packages software into standardized units called _____ that have everything the software needs to run including libraries, system tools, code, and runtime.
16. In what three parts Enterprise applications are built?
17. What are the Benefits of Microservices?
18. What is Amazon ECS?
19. With Amazon ECS, your containers are defined in a task definition that you use to run an individual task or task within a service. (True/False)
20. What are the three storage services in AWS?
21. _____provides block level storage volumes for use with EC2 instances.
22. HDD-backed storage is for transactional workloads and DDD-backed storage for throughput workloads. (True/False)
23. What is the use of Amazon Data Lifecycle Manager?
24. What are the key elements of Amazon Data Lifecycle Manager.
25. An _____ provides the information required to launch an instance.
26. What is Amazon Elastic File System (EFS)?

27. Amazon _____ is an object storage service that offers industry-leading scalability, data availability, security, and performance.
28. The domain name system (DNS) is a naming database in which internet domain names are located and translated into_____.
29. Vertical scaling is about changing the instance up and down and Horizontal scaling is about adding more machines of similar capacity to the infrastructure. (True/False)
30. Amazon EC2 ____ Scaling helps you maintain application availability and allows you to automatically add or remove EC2 instances according to conditions you define.
31. What are the types of Elastic Load Balancer?
32. _____ is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale.
33. What is AWS Site-to-Site VPN?
34. _____ integration is a DevOps software development practice where developers regularly merge their code changes into a central repository, after which automated builds and tests are run.
35. With the AWS CodeStar project dashboard, you cannot track your entire software development process, from a backlog work item to production code deployment. (True/False)
36. AWS X-Ray supports applications running on:
37. What is Disaster Recovery?
38. What Dynamo DB is made up of?
39. _____is a fast, fully managed data warehouse that makes it simple and cost-effective to analyze all your data using standard SQL and existing Business Intelligence (BI) tools.
40. ElastiCache nodes run on Amazon EC2 instances. (True/False)

----------End of the Module----------

MODULE 9
HANDS-ON AWS

Section 1: Learning Outcomes

After completing this module, you will be able to:
Execute following operations on AWS platform:

- Launching a Linux EC2 Instance
- Connecting to your EC2 Instance
- Transferring files to your Amazon Instance
- Stopping and Restarting an Instance
- Creating Snapshots
- Converting Snapshot to EBS Volume
- Launching and using Amazon RDS Instance

Section 2: Relevant Knowledge

9.1 Pre-Requisites

Check if following pre-requisites are met:

- You have an Amazon account. Else create one
- In case you are using Windows, ensure you have downloaded the following software (all of them are free)
  - PuTTY
  - PuttyGen
  - FileZilla
- In case you are using a Linux system, ensure the SSH and FTP services are running

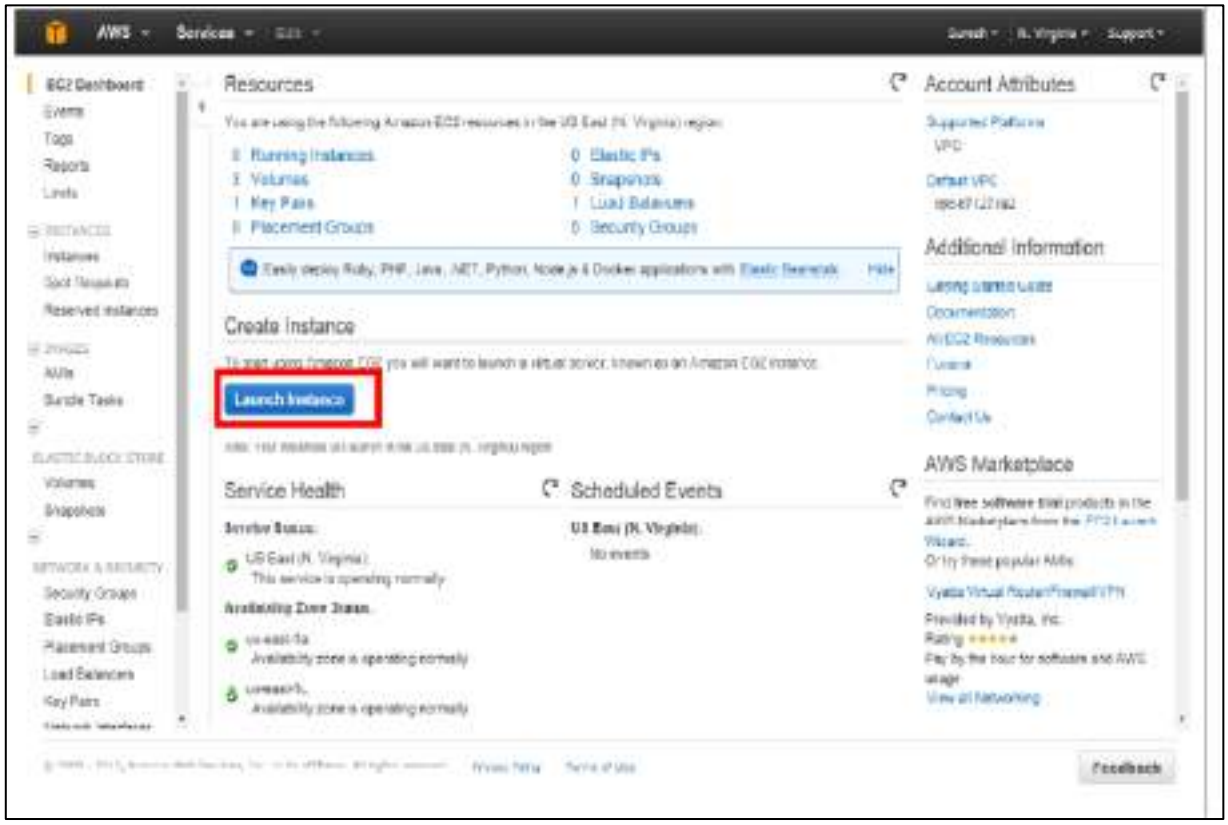
Important Note (Disclaimer):

We have tried to design the labs in such a way that we make use of the free tier and the participant incurs no charge. Yet it is possible that due to extra work done by the participants on their own or not terminating some instances, not releasing storage or due to change in Amazon pricing policy or any other factor some fee may be incurred. We are not responsible for any charge that gets levied by Amazon for usage as part of this lab.

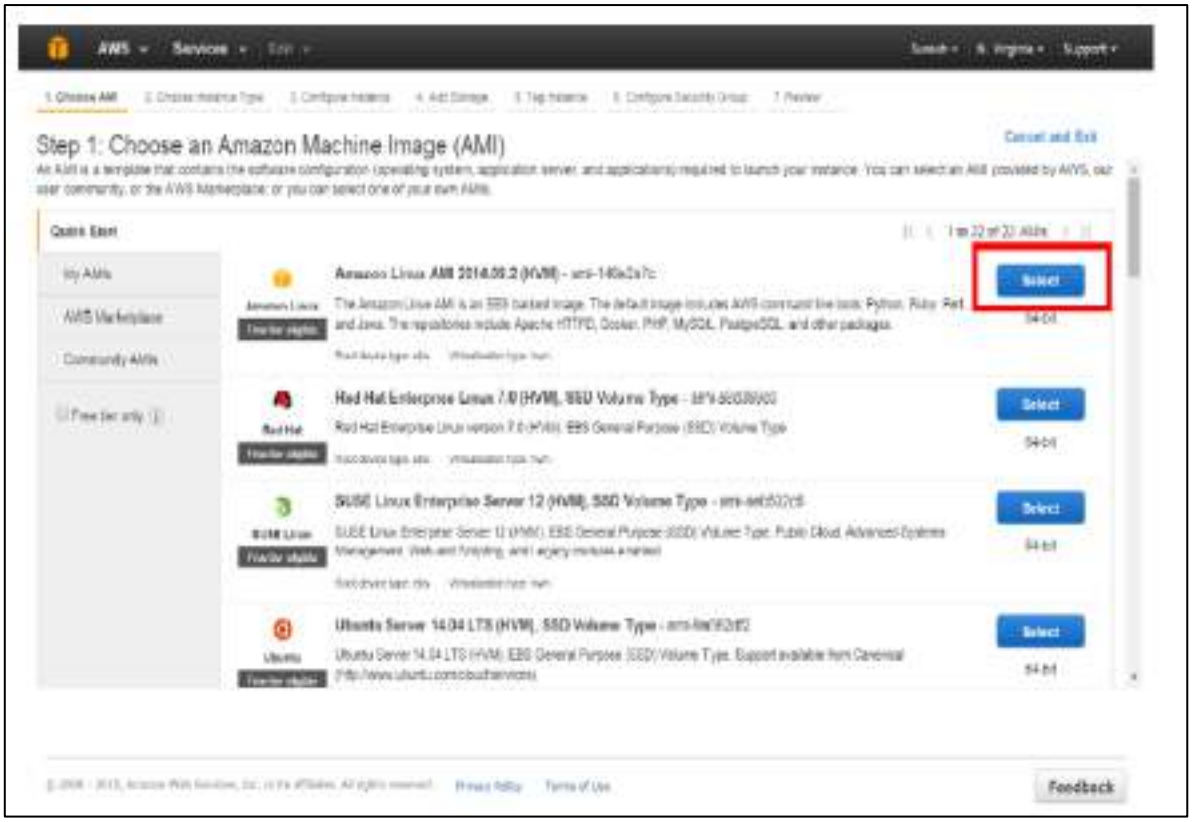
9.2 Exercise 1 - Launching a Linux EC2 Instance

- EC2 is the basic compute instance on the Cloud and forms the foundation for many of the features in Amazon.
- This exercise will provide step-by-step instruction on how to launch an EC2 instance.
  - Login into console
  - http://console.aws.amazon.com
  - Login using your Amazon account credentials
  - On the AWS Console, click on EC2
- This is the AWS Dashboard.
- Clicking on EC2 will take you to the EC2 Dashboard.

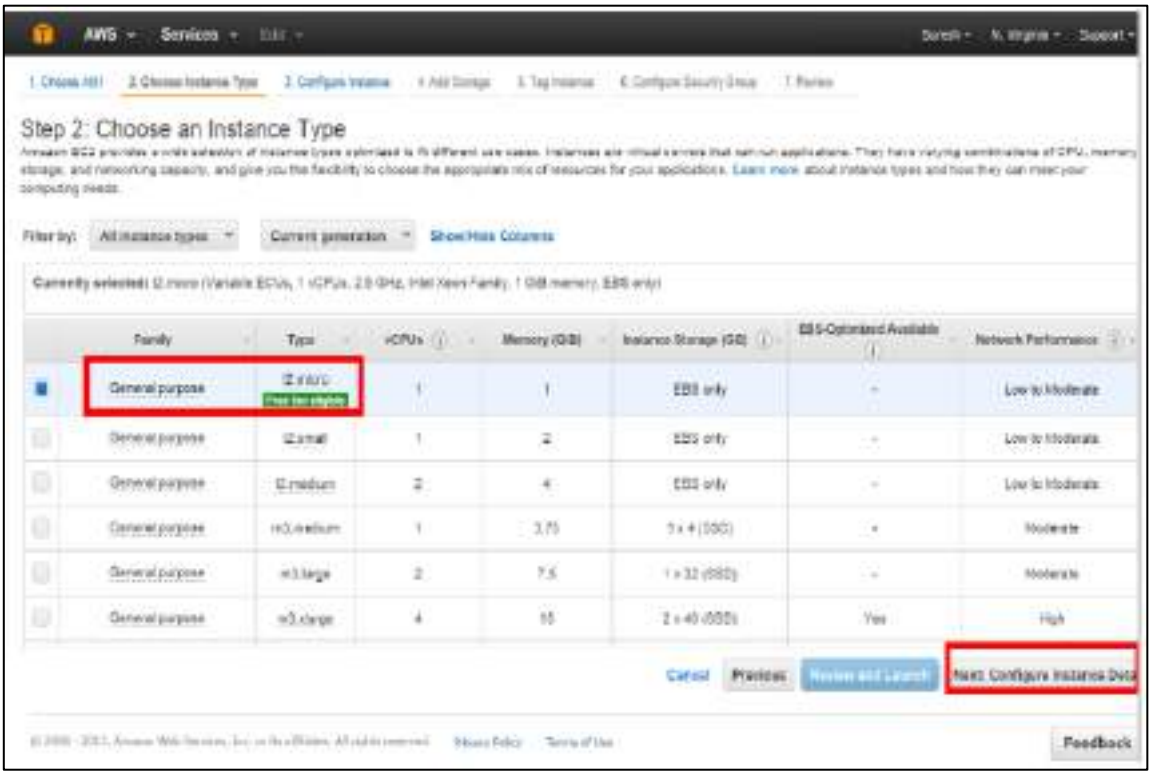
- Following is the EC2 Dashboard. When you start enter this menu for the very first time all the Resources must be zero. Click on the blue ‘Launch Instance’ button to start the launch process.



- The first screen that will appear when the ‘Launch Instance’ button is pressed is the AMI selection screen.
- This is where you need to select an AMI of your choice.
- In our case we will choose the Amazon Linux AMI, which is free.

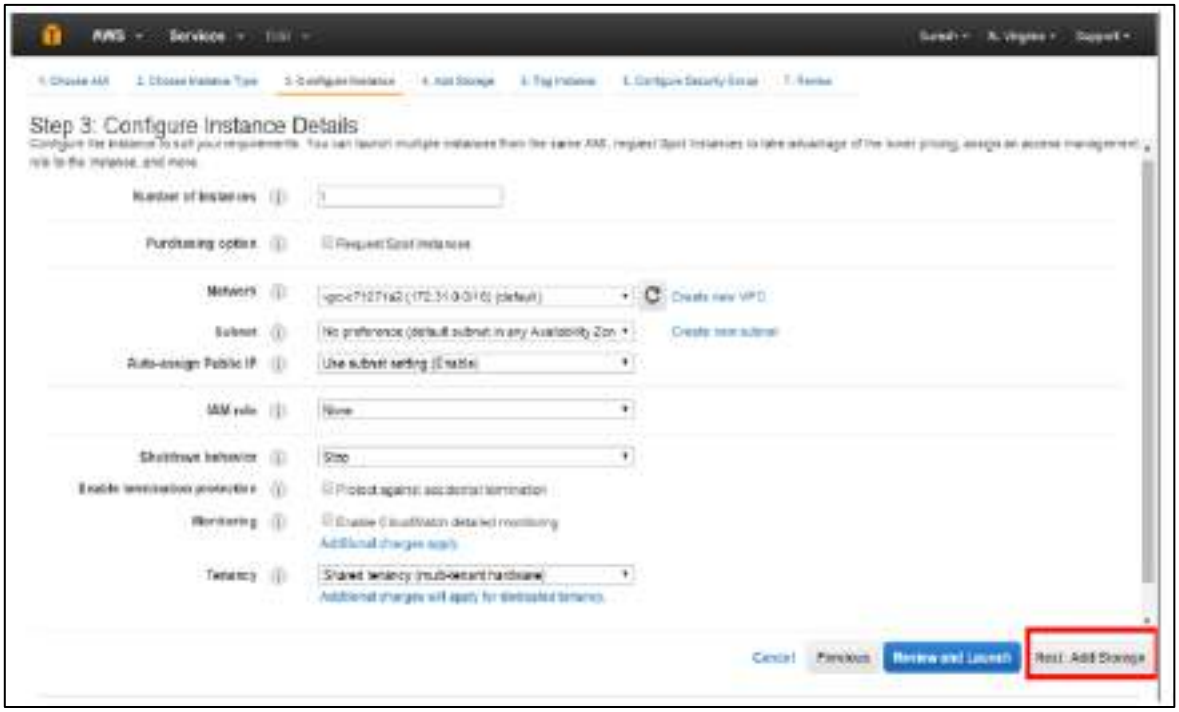


- Selecting the AMI will lead you to the Instance Type screen. Select the t2.micro instance which is the free instance.
- After choosing this instance click on ‘Next: Configure Instance Details’

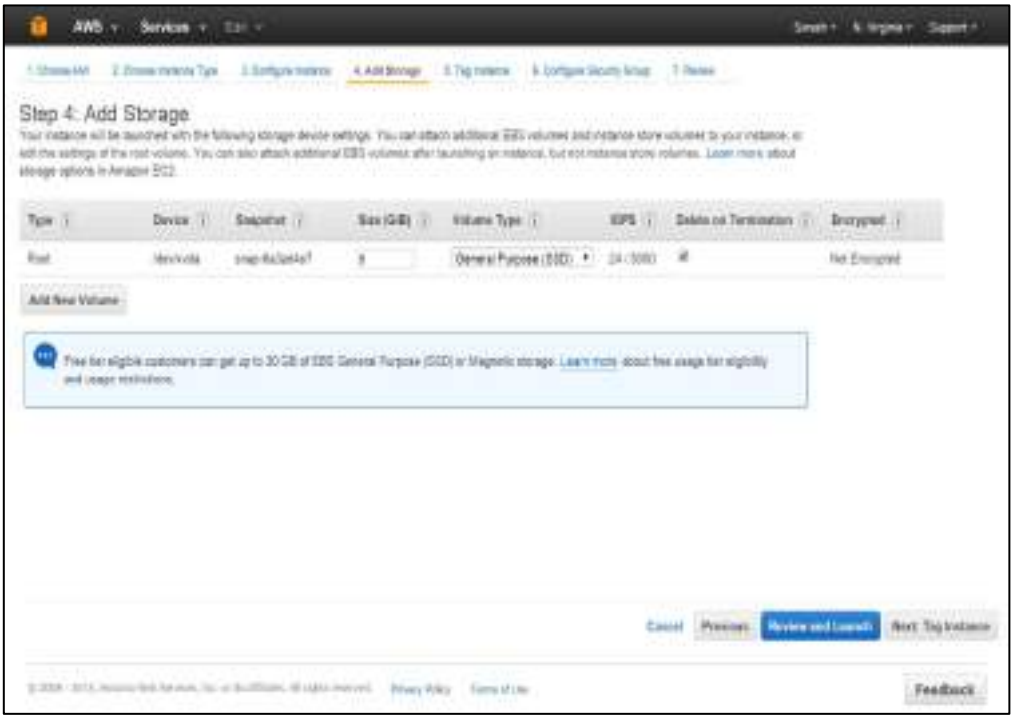


- This leads to the ‘Configure Instance Details Screen’.
- You can launch multiple instances from the same AMI by specifying the number of instances.
- You can also create a new VPC or assign an IAM role here.

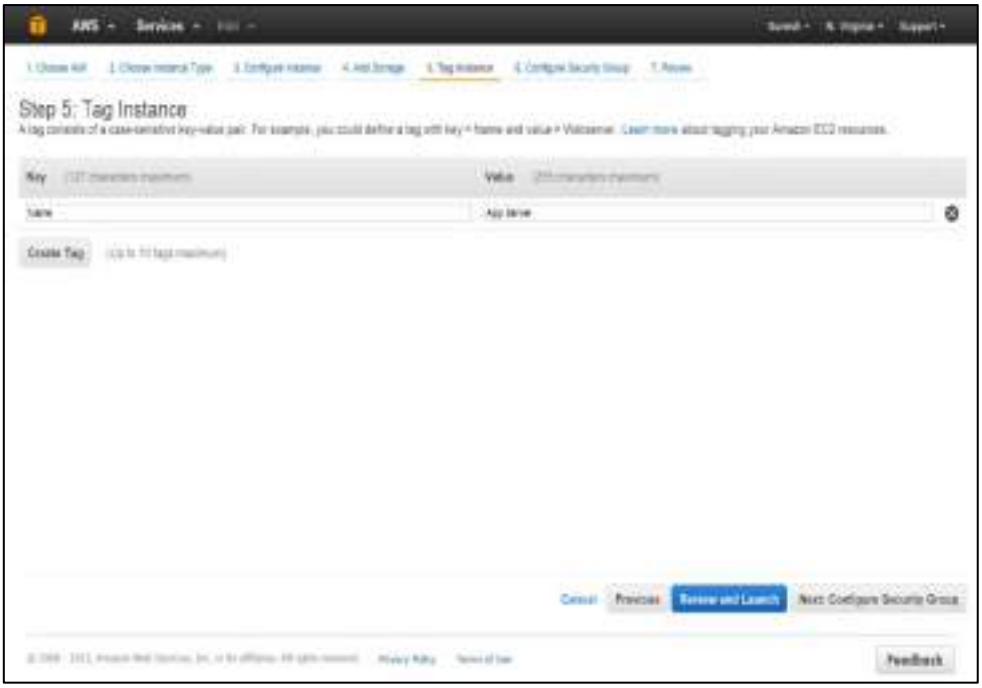
- For our exercise, we will not change anything in this screen. Press ‘Next: Add Storage’



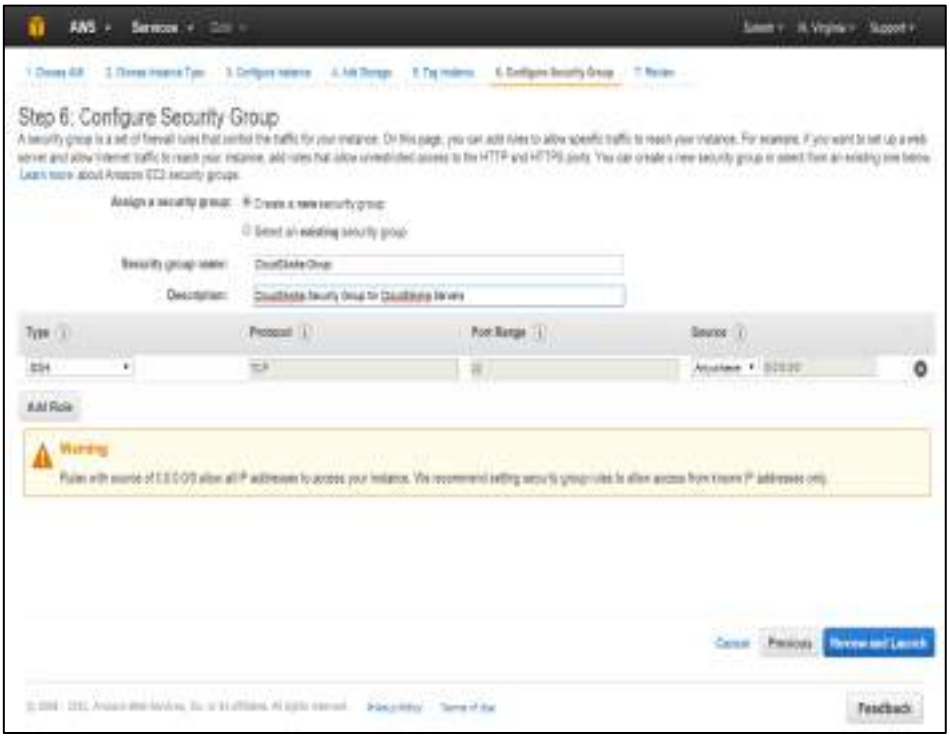
- This will bring up the ‘Add Storage’ screen.
- Select ‘General Purpose SSD’ as Volume Type. Let other parameters be at their default value.
- Then Click ‘Next: Tag Instance’.



- In ‘Tag Instance’ screen you can assign any key-value combination that you want so that you can identify the instance easily.
- For our exercise let the Key be ‘Name’ and the Value can be ‘App Server’ or ‘Web Server’
- Click on ‘Next: Configure Security Group’

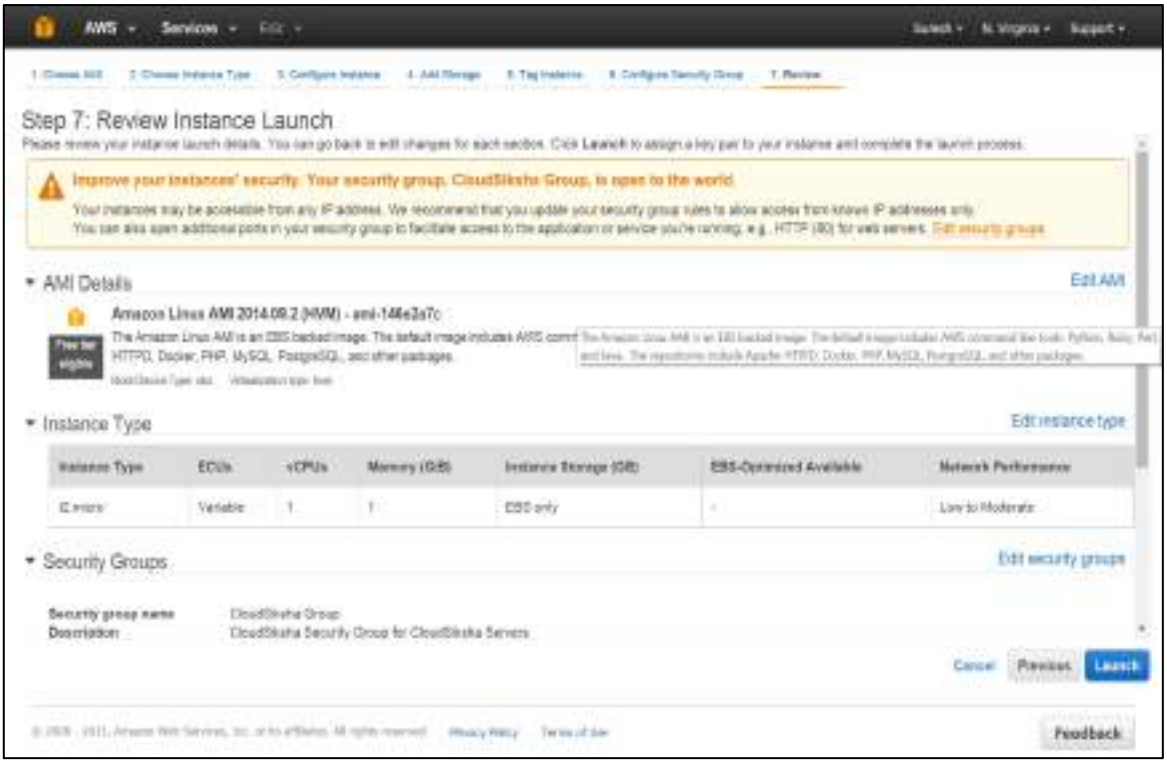


- In the ‘Configure Security’ screen choose ‘Create a New Security Group’.
- Give a name to the Security Group and write a description for the Group.
- SSH is enabled by default. Click on ‘Add Rule’ and allow HTTP and HTTPS for everyone.
- When all rules are added, click on ‘Review and Launch’ button

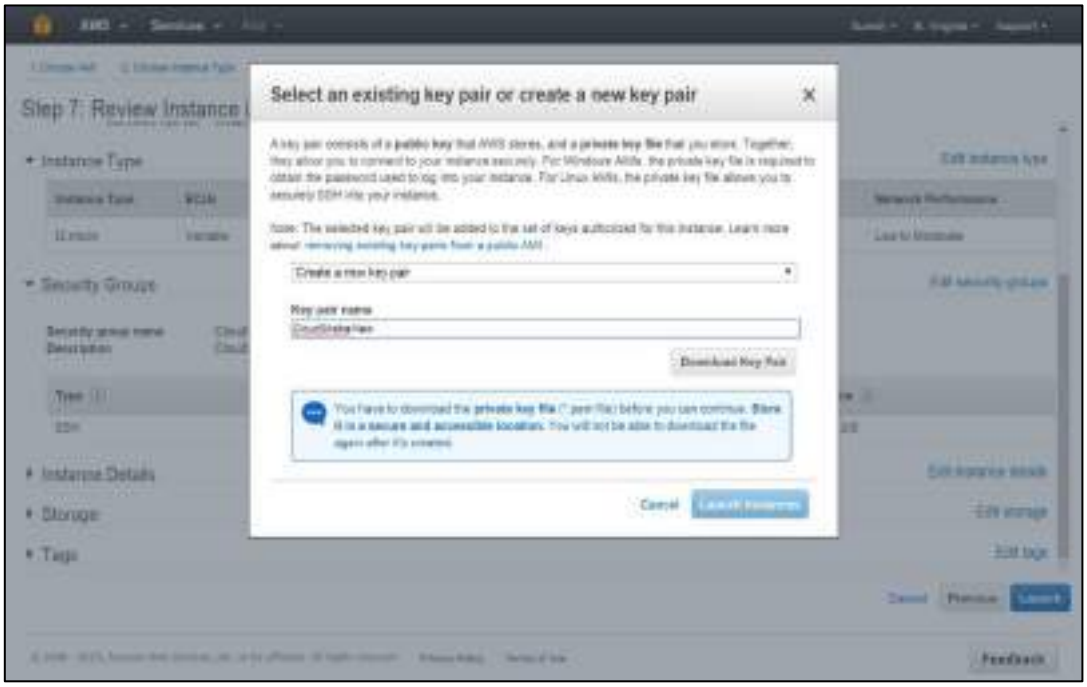


- The ‘Review Instance Launch’ page provides all the options we have chosen in a single page. We can review the options and then click ‘Launch’

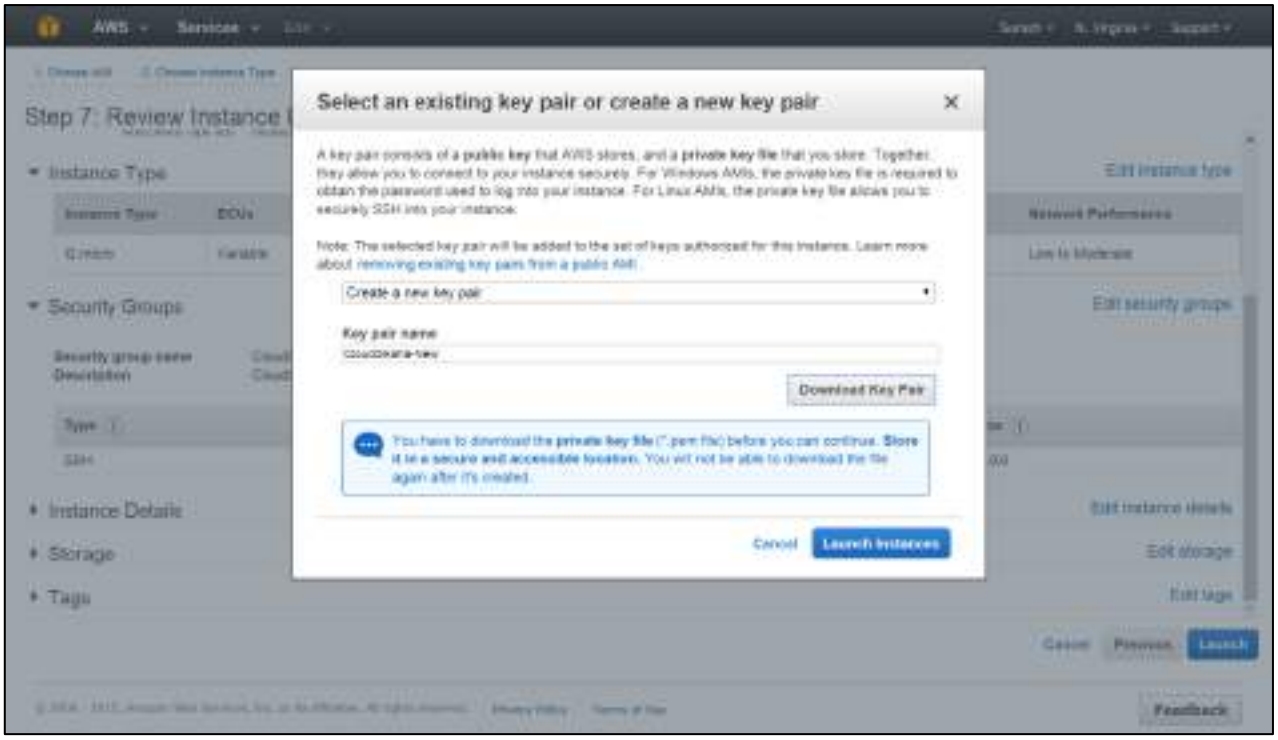
Note: All options are not seen in the screenshot. In actual case you need to scroll to see all the options



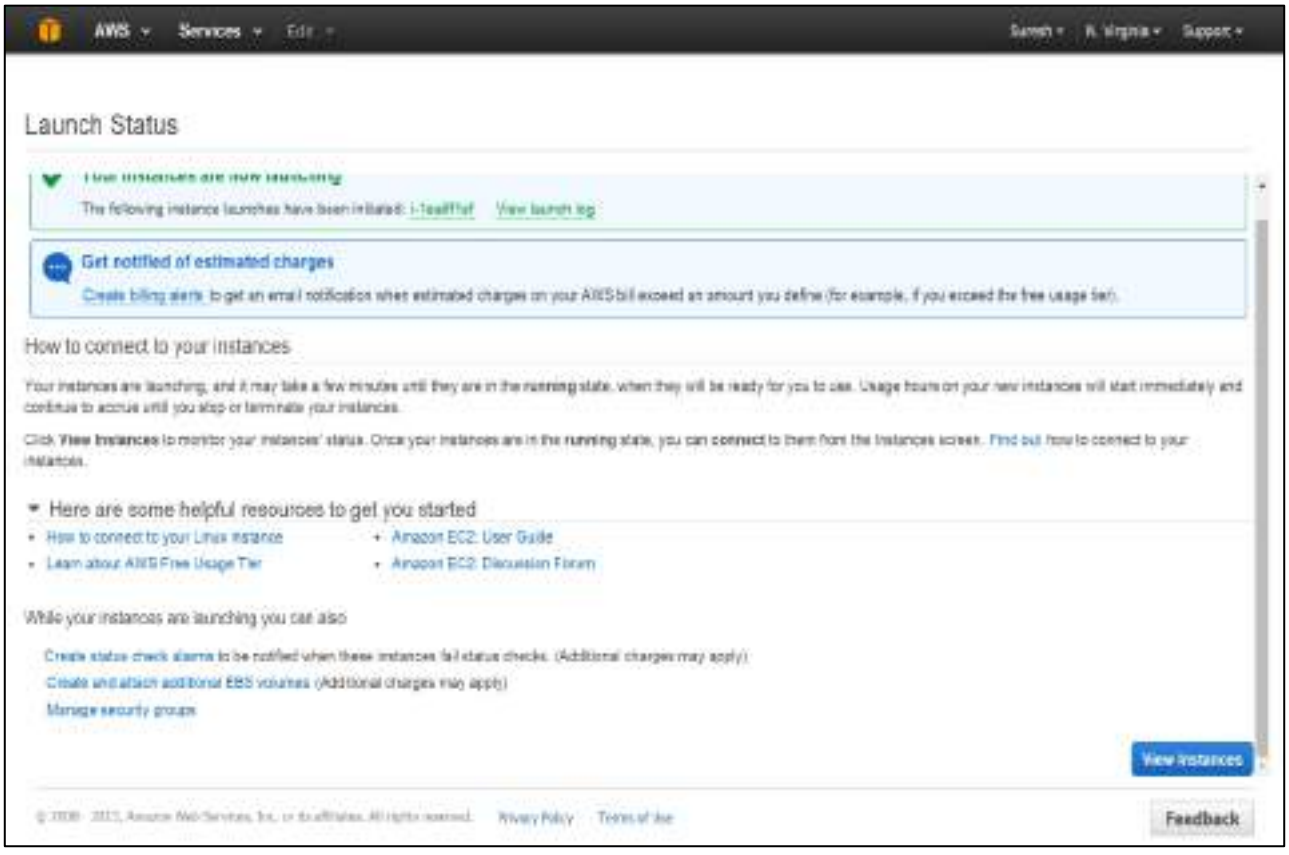
- When the ‘Launch’ button is clicked, a pop-up screen appears asking for a Key Pair.
- This is the Public-Private key that is used for security purposes.
- Since we do not have a key yet choose ‘Create a New Key Pair’ from the drop down, give a name to key pair and click on ‘Download Key Pair’



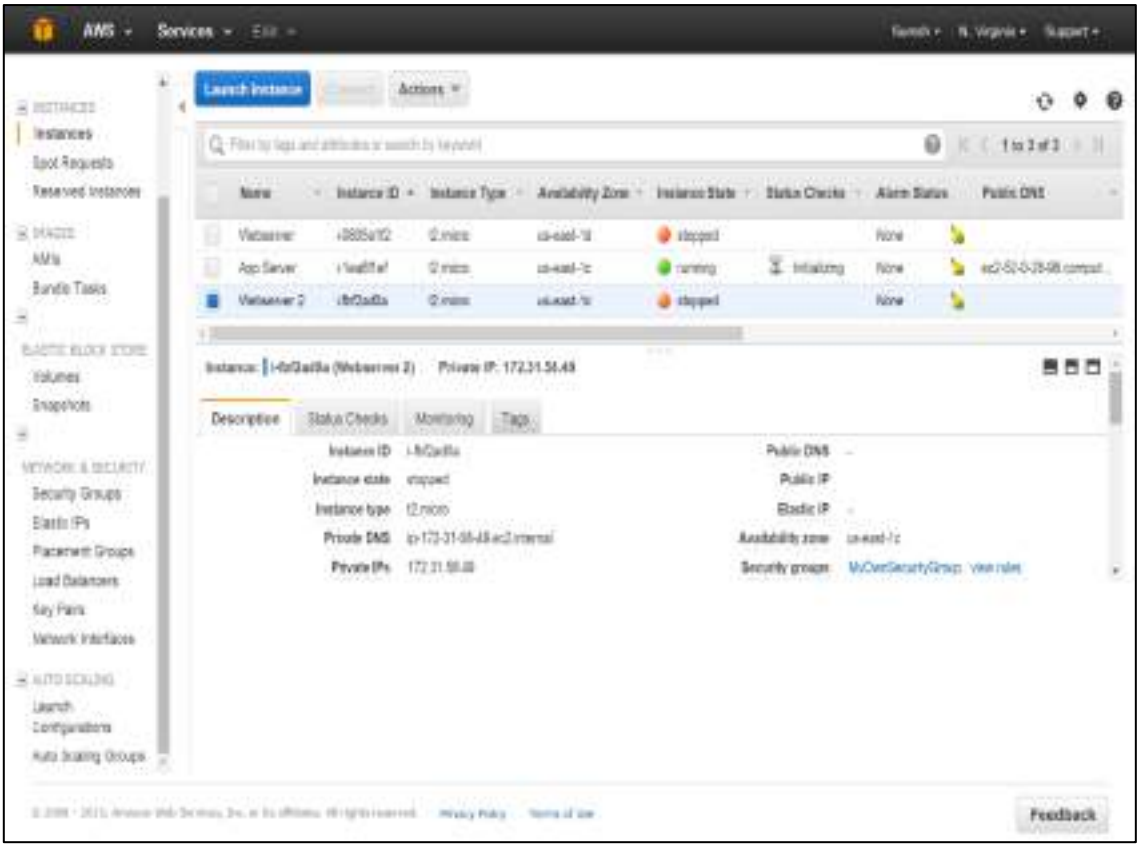
- Save the key safely in a directory of your choice.
- Once you have downloaded the key (which is a *.pem file) select ‘Launch Instance’



- The instance would now be launched.
- From the Launch Status screen, scroll down and click ‘View Instances’



- This will take you to the instance dashboard and you will see the list of all instances (in our case there must be only one instance)



You have now successfully launched an instance

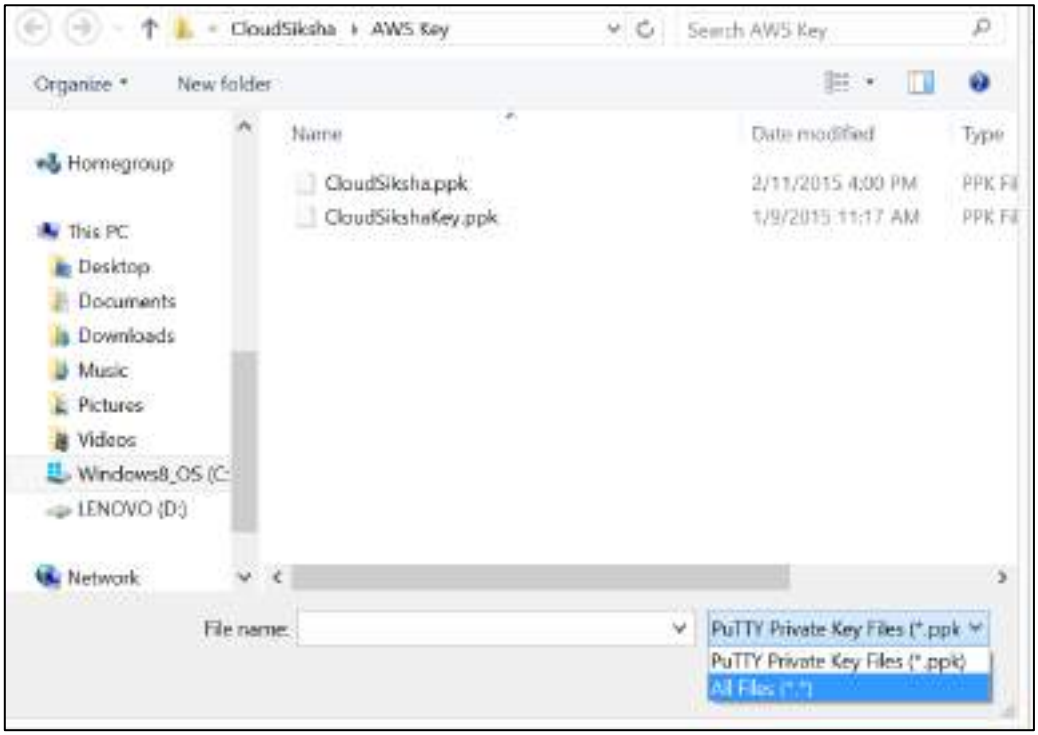
9.3 Exercise 2 - Connecting to your EC2 Instance

- Once you have launched an instance, you need to connect to it.
- In all these exercises we will be dealing with Linux Instances.
- We can login to the instance from Windows, Linux or using Browser (requires Java).
- In this exercise we will login into the instance using PuTTY on Windows and using SSH from Linux.
- There are three methods to connect to your EC2 instance:

➢ Using SSH from your Linux /Mac system
➢ Using PuttY from a Windows system
➢ Using Browser based SSH (Java required)

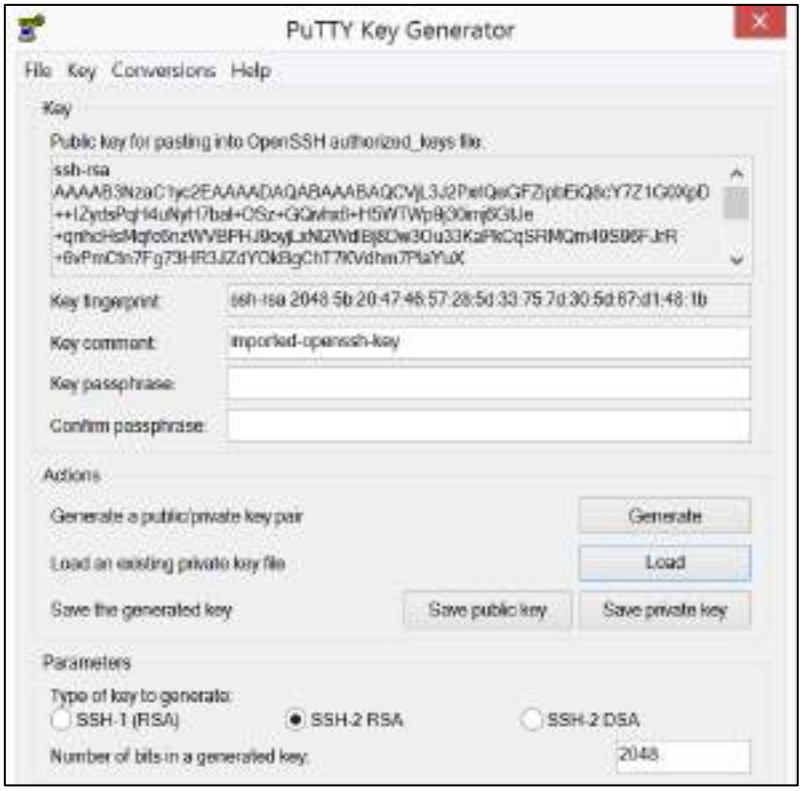
- In our lab we will do the first two methods
- We start with the Windows method as it involves a few additional steps in order to connect to the instance.
- Generating .ppk file using PuttyGen
- The key pair file that was download in is the .pem format. For using PuttY we must convert this to .ppkformat. This is done using Puttygen software
- Launch Puttygen and you will see this screen.

- Ensure that the type of key to generate is SSH-2RSA and click on the Load button
- In the File Dialog box initially, you can only see the *.ppk files (we will not have any initially. So, the screen will be blank)
- Choose ‘All Files’ in order to view the *.pem files.



- Click ‘OK’ on the pop up that appears

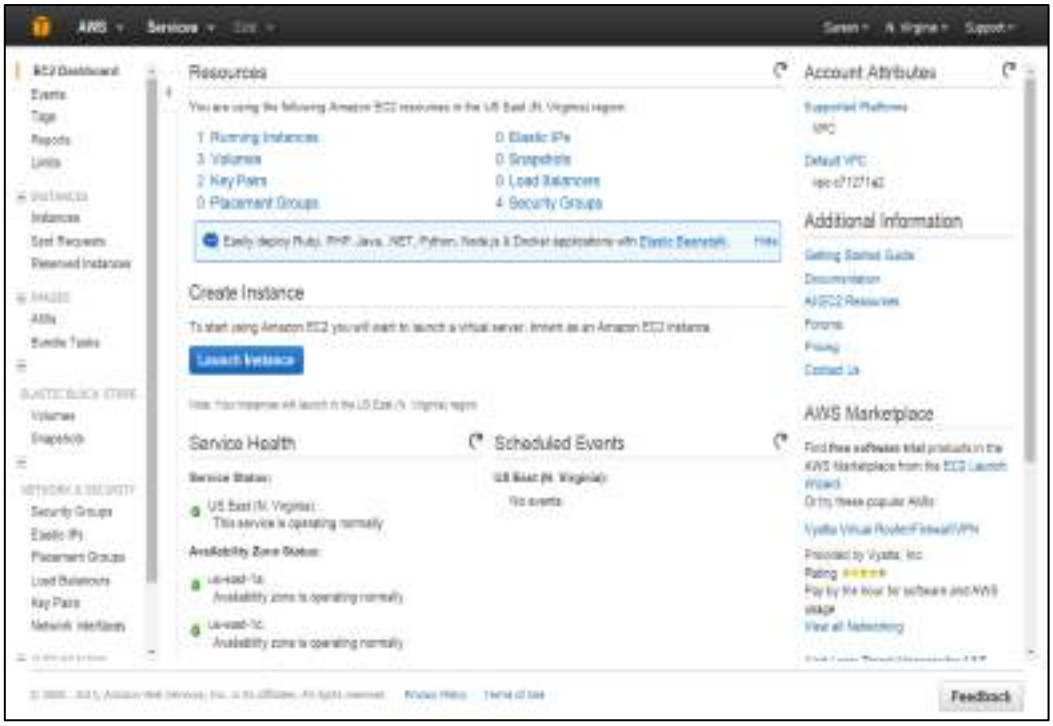
- Now the .pem file is loaded. Click on ‘Save Private Key’



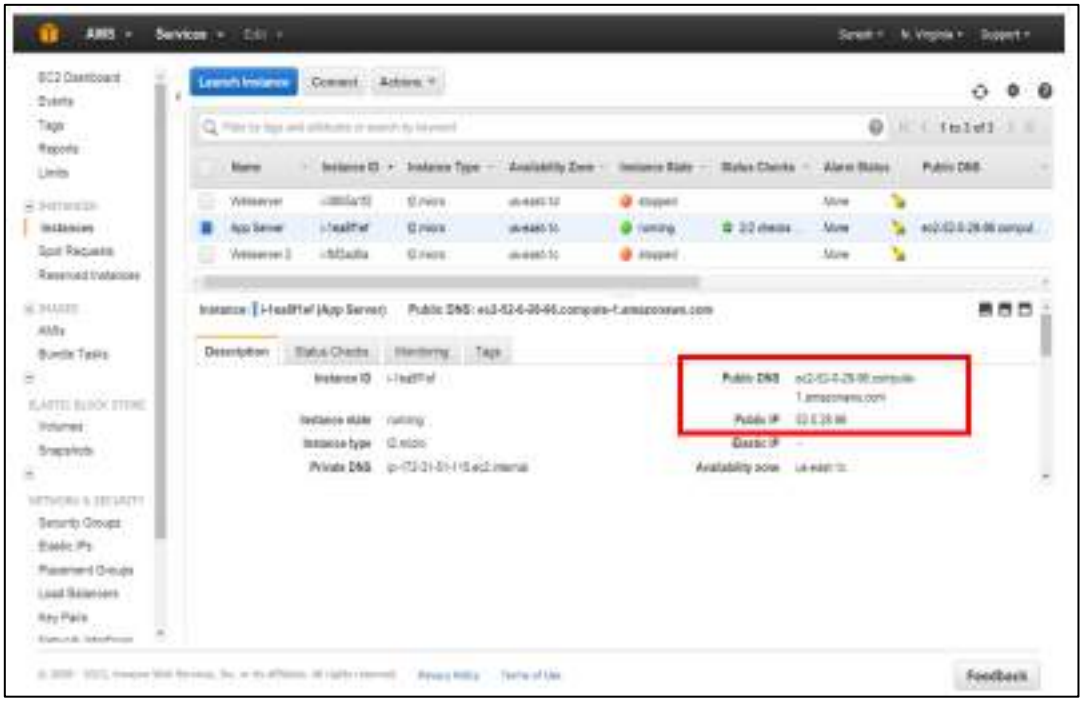
- Say ‘Yes’ to the warning that appears
- Save the key with the same name as the .pem file (If you have saved .pem file as CloudSiksha.pem, this key must be saved as CloudSiksha. You will now have a CloudSiksha.ppk file)
- Exit PuttyGen by closing the window
- Connecting to the EC2 Instance using PuttY

Note down the IP Address or the Public DNS of the instance

- Go to EC2 Dashboard and Click on Running Instances to get to the Instance Screen



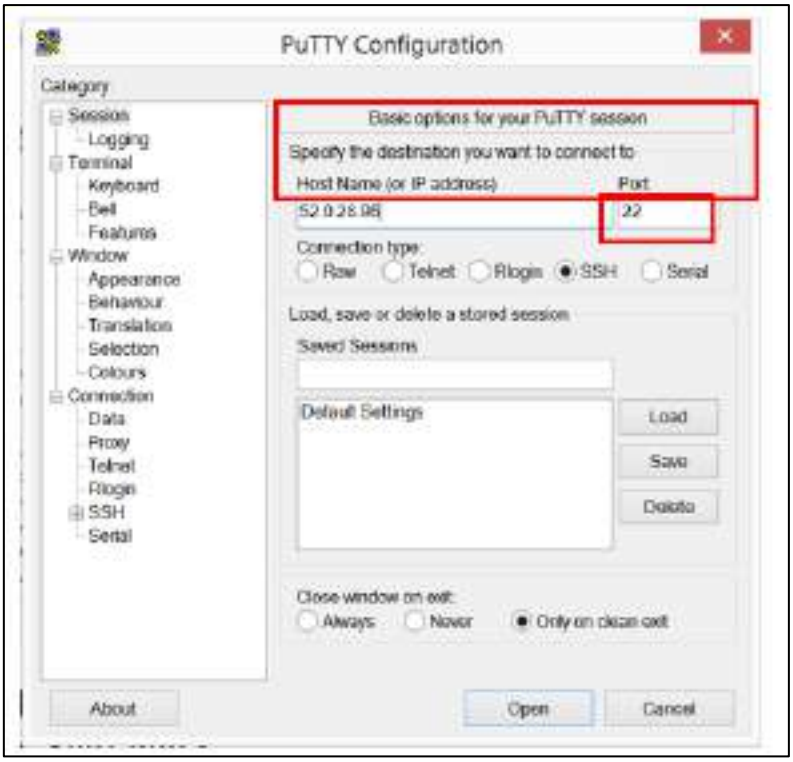
- This will give us a list of running instances.
- Select the instance you need and you will be able to see the IP Address and Public DNS of the instance.



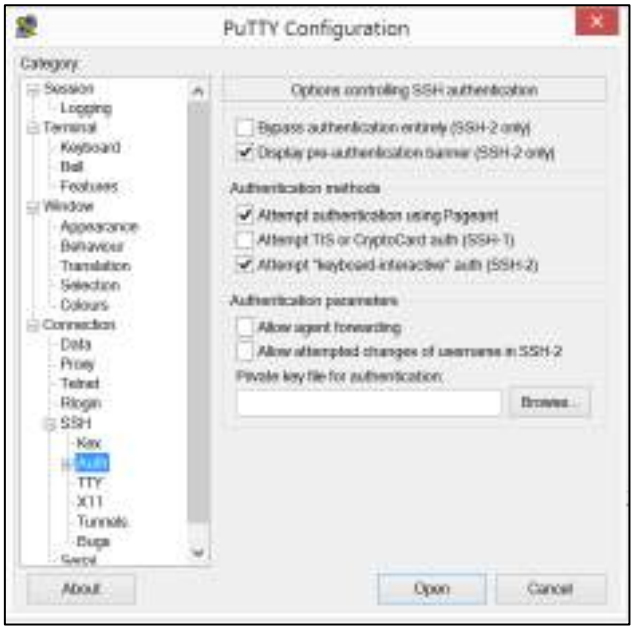
Note down the Public IP and Public DNS Name

PuTTY settings and Login to the EC2 Instance

- Start PuTTY.
- In the Host Name, enter either the IP Address or the DNS Name.
- Ensure ‘Connection Type’ is SSH and Port is 22



Open the SSH on the Category column and Click on ‘Auth’



- Click on the ‘Browse’button. From the file menu choose the .ppk file that you have saved earlier
- Click ‘Open’ and click ‘Yes’ against the PuTTY security warning pop-up
- This will lead you to the Login screen

- The Login name is ec2-user.

- There is no password required
- You have now successfully logged in. You can issue any Linux command like ‘df –h’ to test your instance

Logging from a Linux System

- Logging from a Linux system is a simple one step process using SSH. Issue the following command to login to the instance

$ ssh –i <path/to/*.pem file> ec2-user@<Public IP of instance>

- For example, if you pem file is name CloudSiksha.pem and is available in your current directory, and the Public IP is 52.1.159.87, then your command will be:

$ ssh –iCloudSiksha.pemec2-user@172.31.151.15

9.4 Exercise 3 - Transferring files to your Amazon Instance

- One of the tasks which you definitely have to perform after starting an instance is to transfer files to the instance.
- This can be done using multiple FTP Client software in Windows.
- We will use FileZilla for our exercise.
- This is a free software. (You can use any other FTP software like CoreFTP etc. if you so desire).
- In case of Linux, we will use the Secure Copy, scp, command to transfer files.

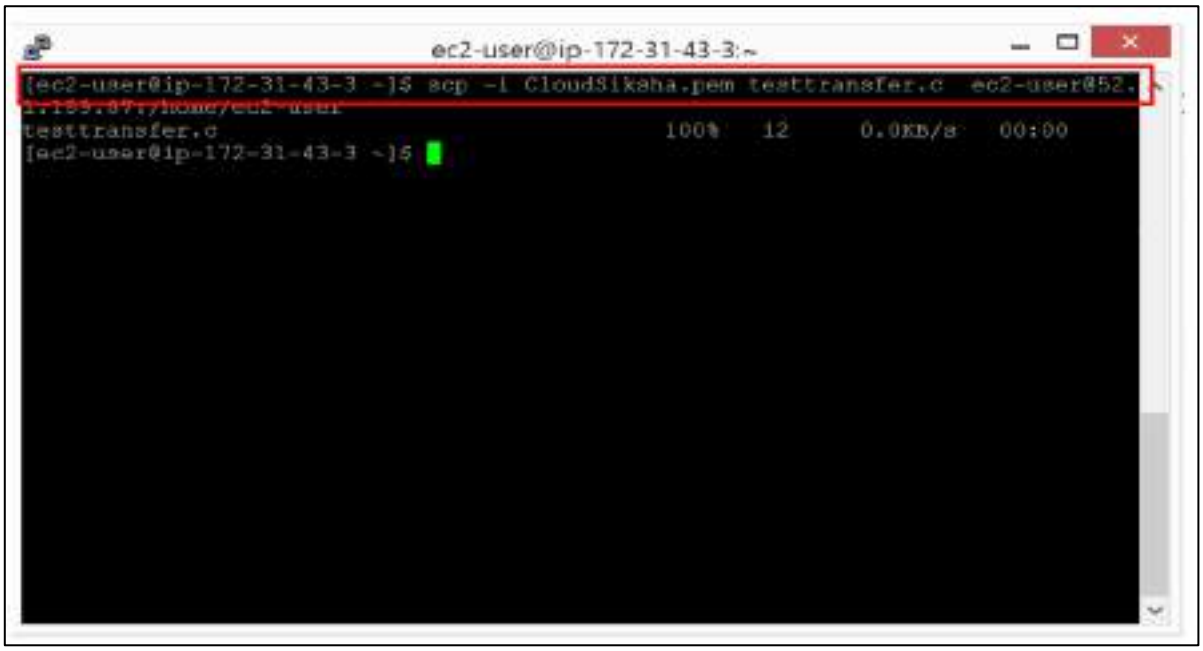
Transferring Files to Amazon Instance from a Linux System

- Files can be transferred from a Linux system to an Amazon Instance using the Secure Copy (scp) command. Issue the following command:

$ scp –i<*.pem file><sourcefile> ec2-user@<destIP>:/<path to file>

Ex :

$ scp –iCloudSiksha.pemtesttransfer.cec2-user@52.159.87:/home/ec2-user

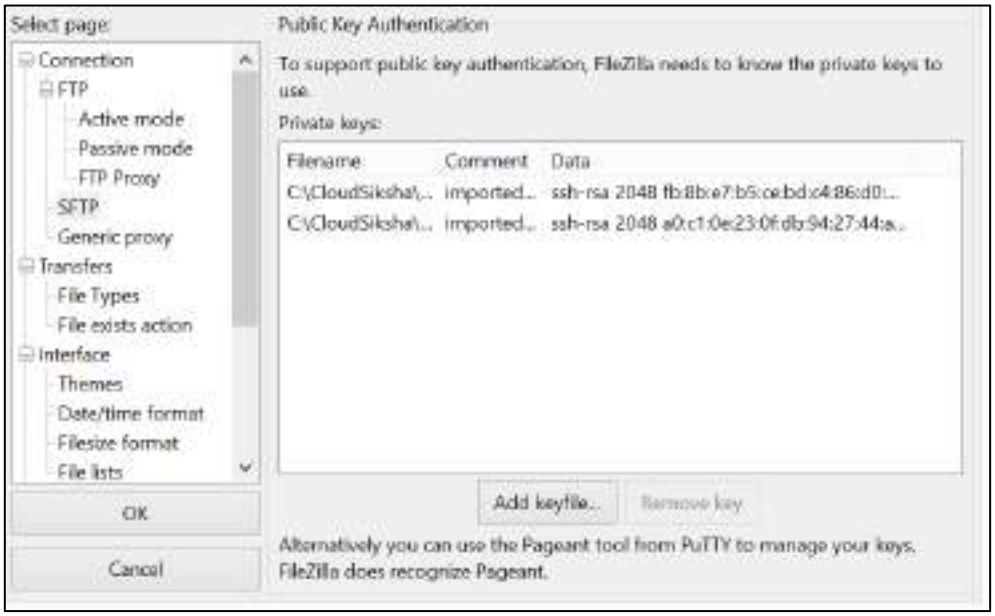


Transferring files to Amazon Instance from Windows

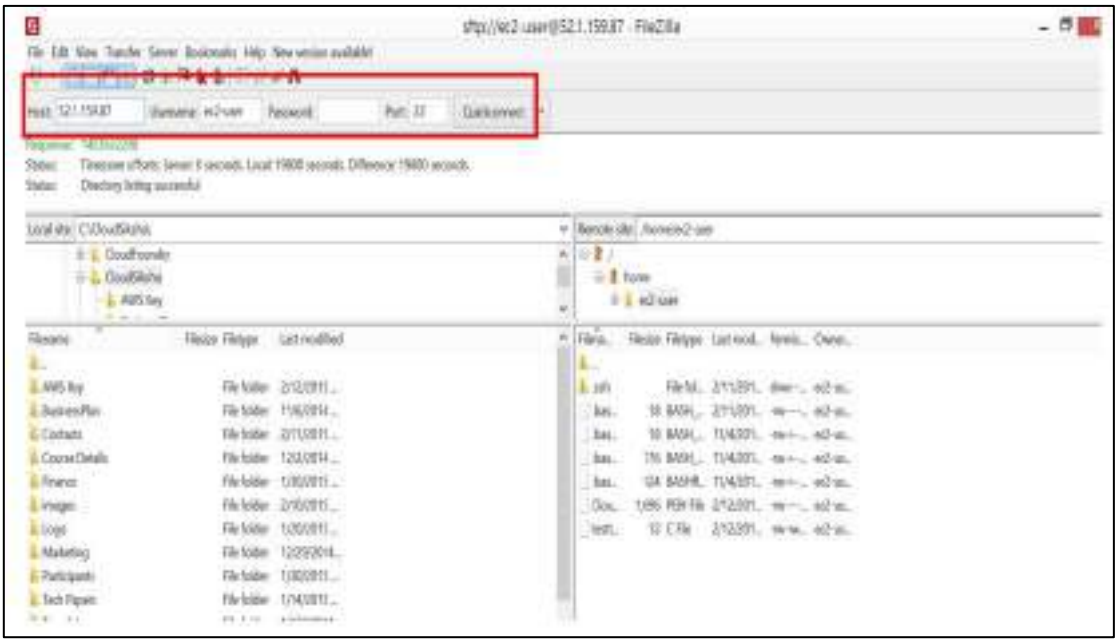
- We can use any FTP client to transfer files from Windows.
- We will use the FileZilla client to transfer files to the Amazon Instance
- The FileZilla Client can be downloaded from:

https://filezilla-project.org/download.php?type=client

- Open FileZilla Client and go to ‘Edit’ →‘Settings’. Select ‘SFTP’ under ‘FTP’ and click on ‘Add KeyFile’



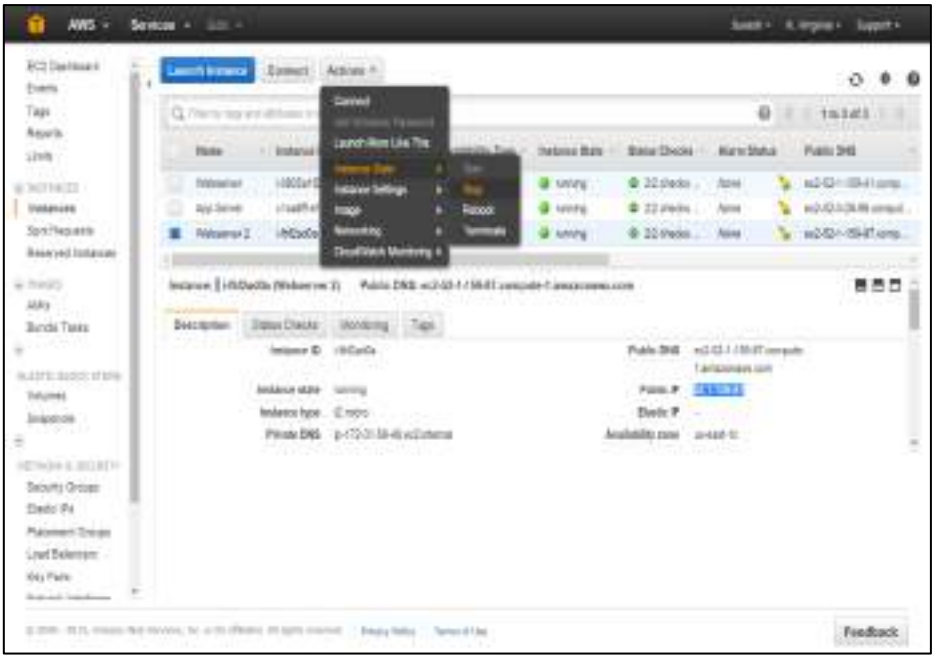
- Input the *.ppk file you had saved earlier.
  (Remember: Whenever it is Windows use the *.ppk file. User *.pem in Linux)
- Click ‘OK’ after your select your keyfile
- In the FileZilla main screen, input the IP Address of the instance, User is ‘ec2-user’ and Port ‘22’ and press ‘Quick Connect’



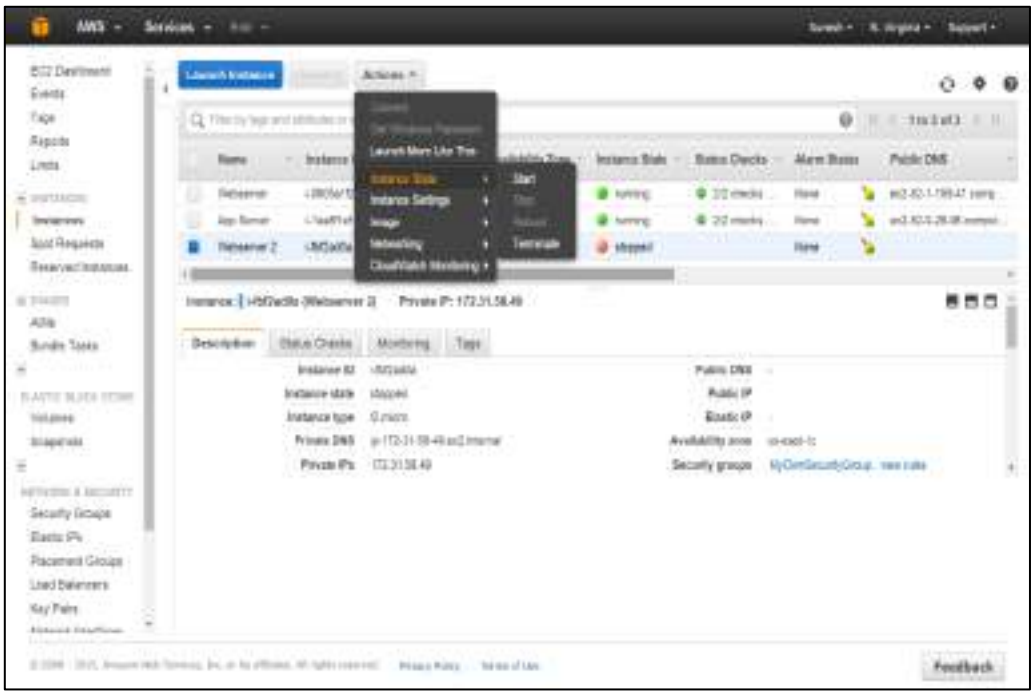
- The local directories are listed on the left and the Amazon Instance files are listed on the right.
- You can now drag and drop files into the Amazon Instance

9.5 Exercise 4 - Stopping and Restarting an Instance

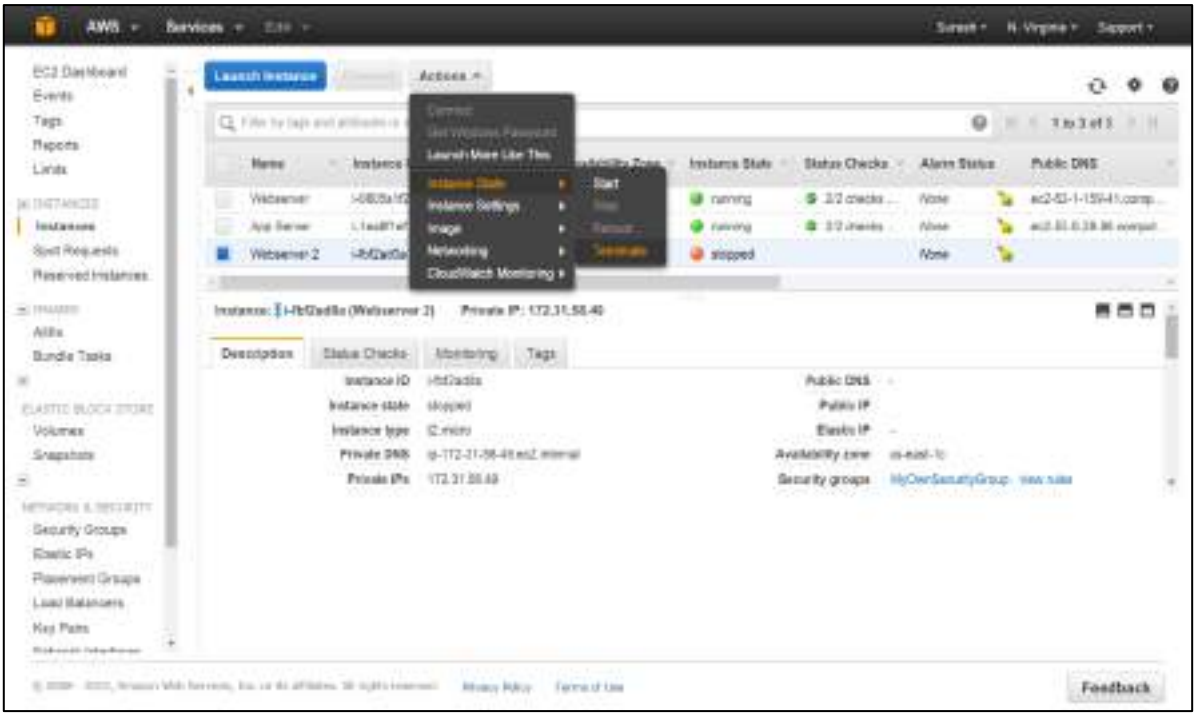
- This is a simple exercise which teaches the user how to start, stop and terminate an EC2 instance
- From EC2 Dashboard, click on ‘Running Instance’ to bring up the Instance Dashboard. Select the Instance you want to Stop and Click on ‘Actions’→’Instance State’→’Stop’



- Press Yes, Stop button when the warning comes up. The instance will shut down
- To start an instance, select the instance you want to Start and Click on ‘Actions’→’Instance State’→’Start’



- To terminate an instance, select the instance you want to terminate and click on ‘Actions’→’Instance State’→’Terminate’

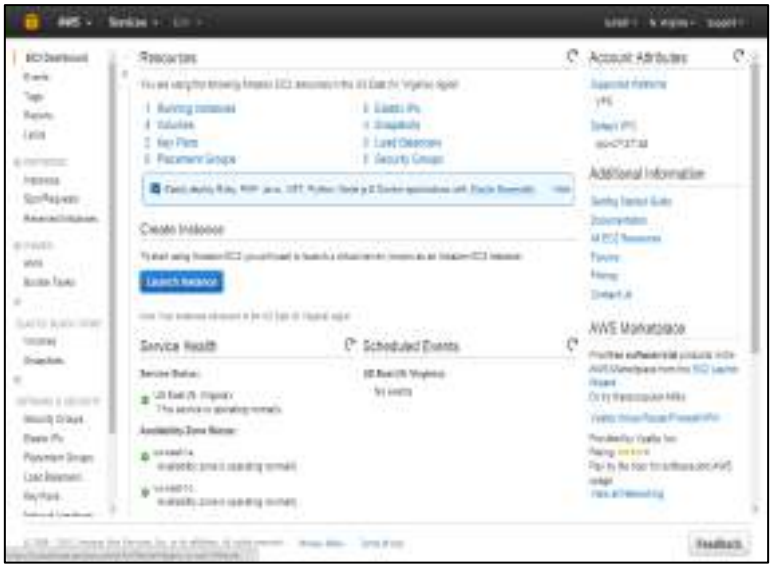


9.6 Exercise 5 - Creating Snapshots

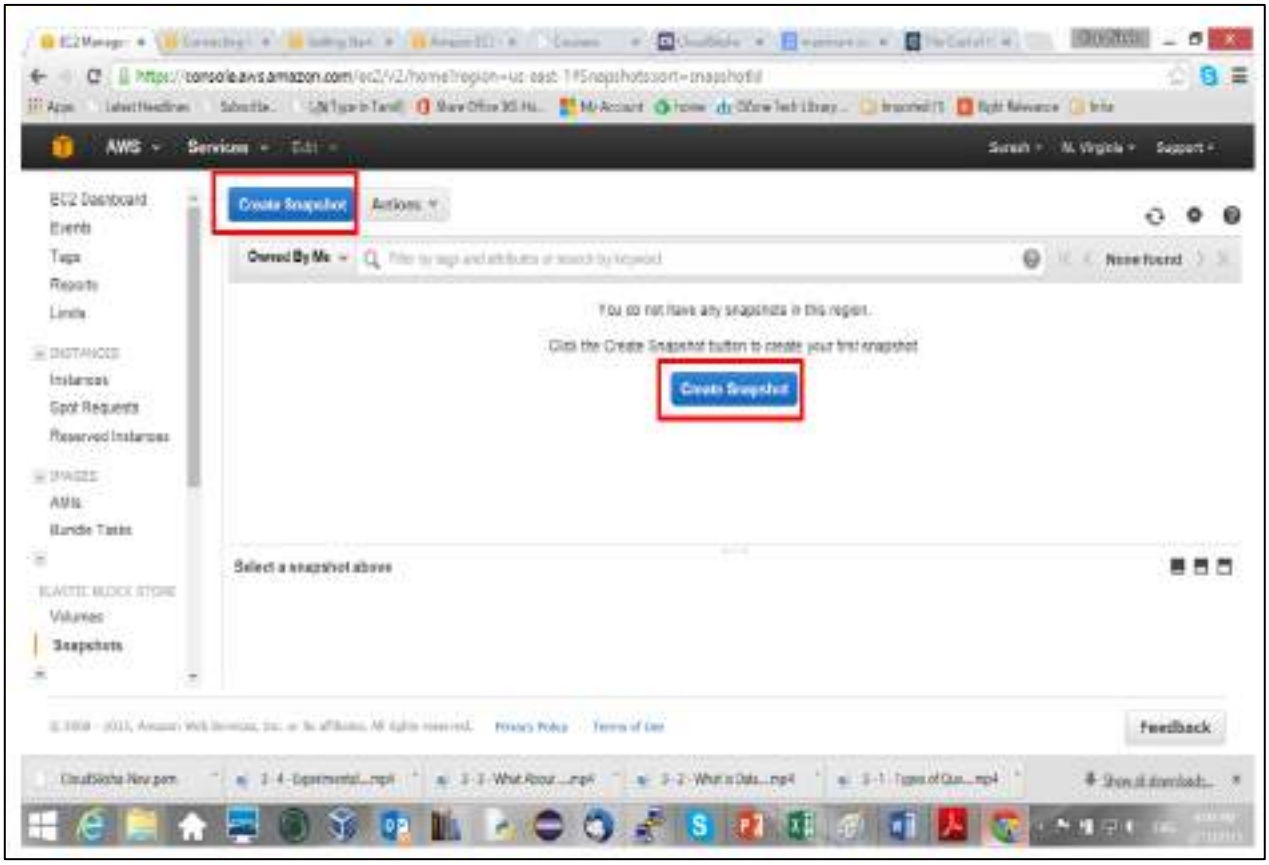
- Snapshots are point-in-time copies.
- They are used extensively to protect data from unexpected deletion or virus attacks.
- When needed we can always fall back on a Snapshot.
- Additionally, Snapshots of root volumes can be used to build AMIs.
- Snapshots can be used to build EBS Volumes as well.
- This exercise details how Snapshots are created.

Creating Snapshots: Method 1

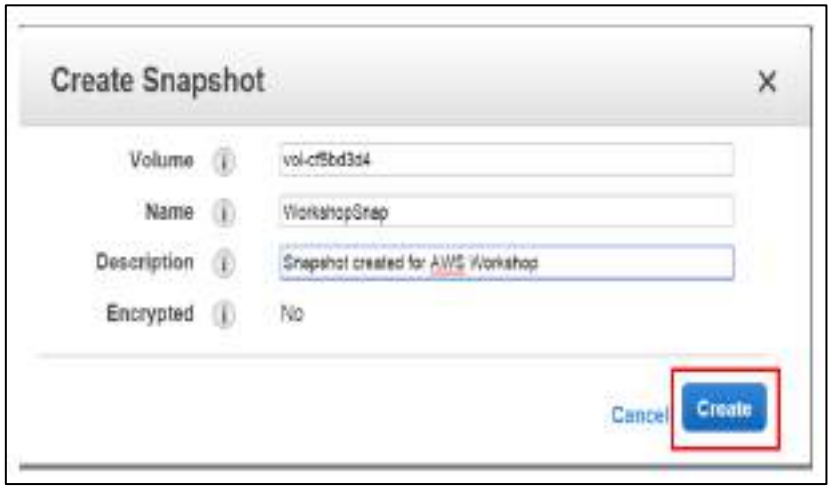
- You can create a snapshot from any EBS volume.
- From the EC2 Dashboard select ‘Snapshots’ either under ‘Resources’ or from the left side menu



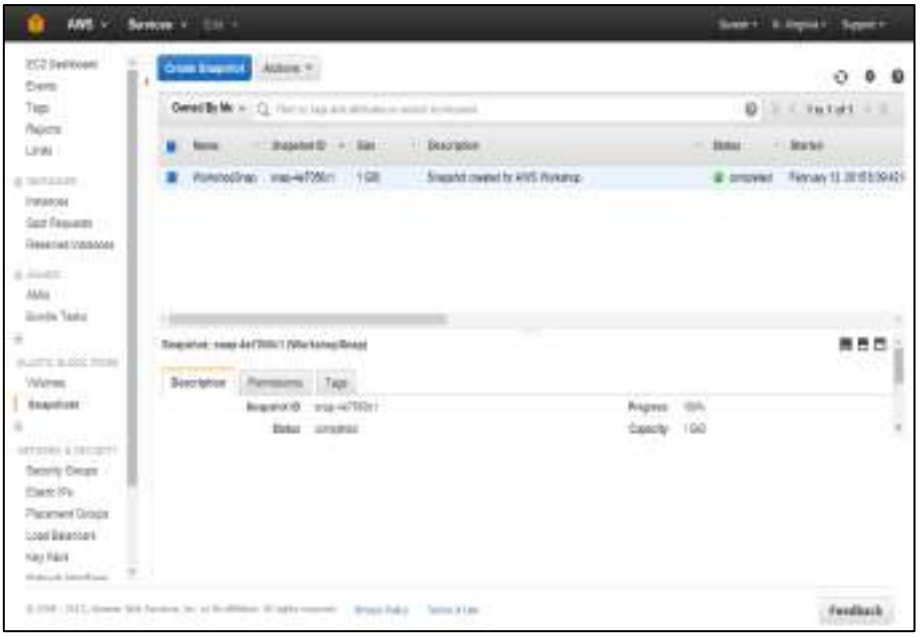
- (Before pressing the ‘Create Snapshot’ button make sure you note down the name of the volume for which you want to create the snapshot.) Press the ‘Create Snapshot’ button.



- In the pop-up box, select the volume, provide a name for the Snapshot and give a description so you remember why you created the snapshot. Press ‘Create’

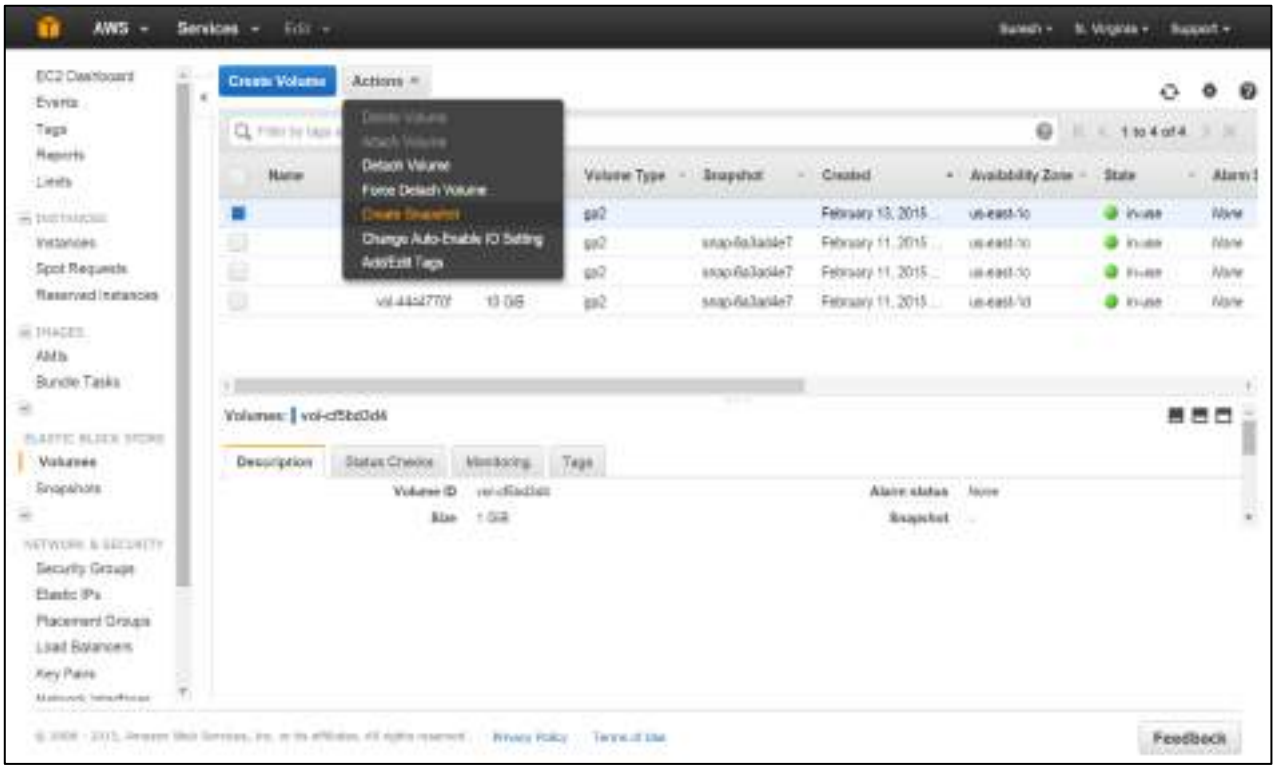


- The snapshot is created



Creating Snapshots: Method 2

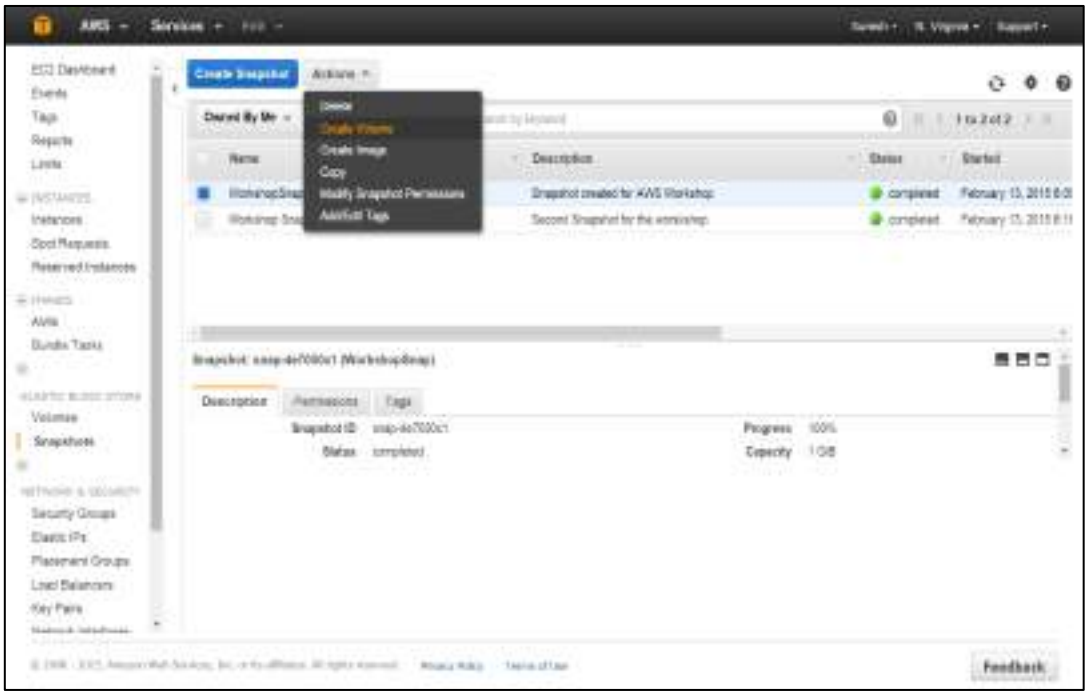
- You can create a snapshot starting from the volume menu. Go to Ec2 Dashboard and from there go to the Volume Dashboard.
- Select the Volume for which you want a snapshot and go to ‘Actions’→’Create Snapshot’



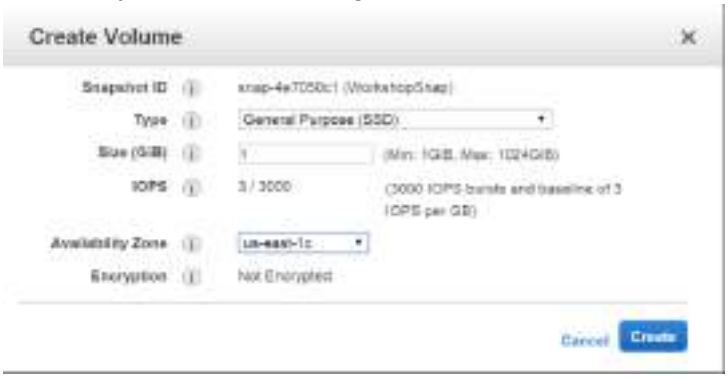
- The snapshot is created

9.7 Exercise 6 - Converting Snapshot to EBS Volume

- This exercise details how a Snapshot created earlier can be converted into an EBS Volume.
- From EC2 Dashboard, go to the Snapshot Dashboard, select the Snapshot from which you want to make a volume. Then select ‘Actions’→’Create Volume’.



- The Create Volume menu appears
- Let us build a 1GB volume with General Purpose SSD.
- You can choose in which Availability Zone you want to create the volume.
- Once created, the volume can only be used in that Availability Zone.
- So, choose a zone in which you have a running instance and then press ‘Create’.

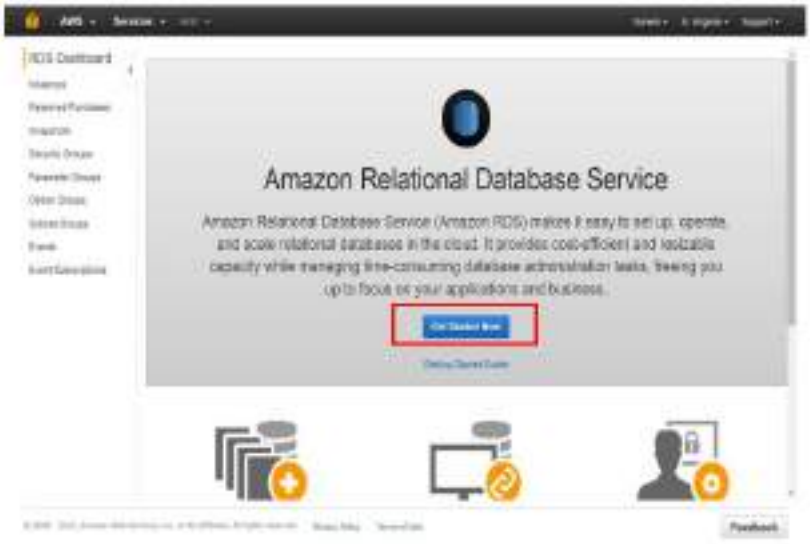


- The volume will be successfully created.

9.8 Exercise 7 - Launching and using Amazon RDS Instance

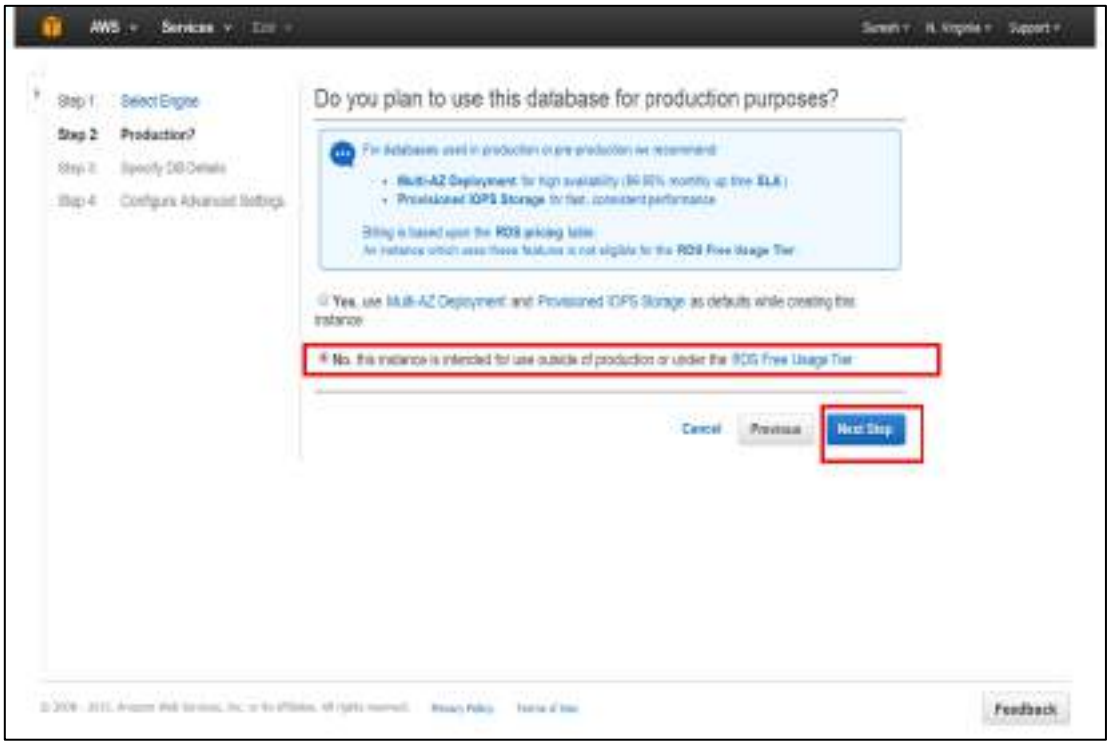
- This exercise will show you how to use the Amazon Relational Database Service (RDS).
- You will launch and RDS instance and connect to it from a Linux instance.
- Starting an Amazon RDS Instance
- Go to the AWS console (console.aws.amazon.com). Click on ‘RDS’.

- Since we are entering RDS service for the first time, we will see this screen. Press on ‘Get Started Now’



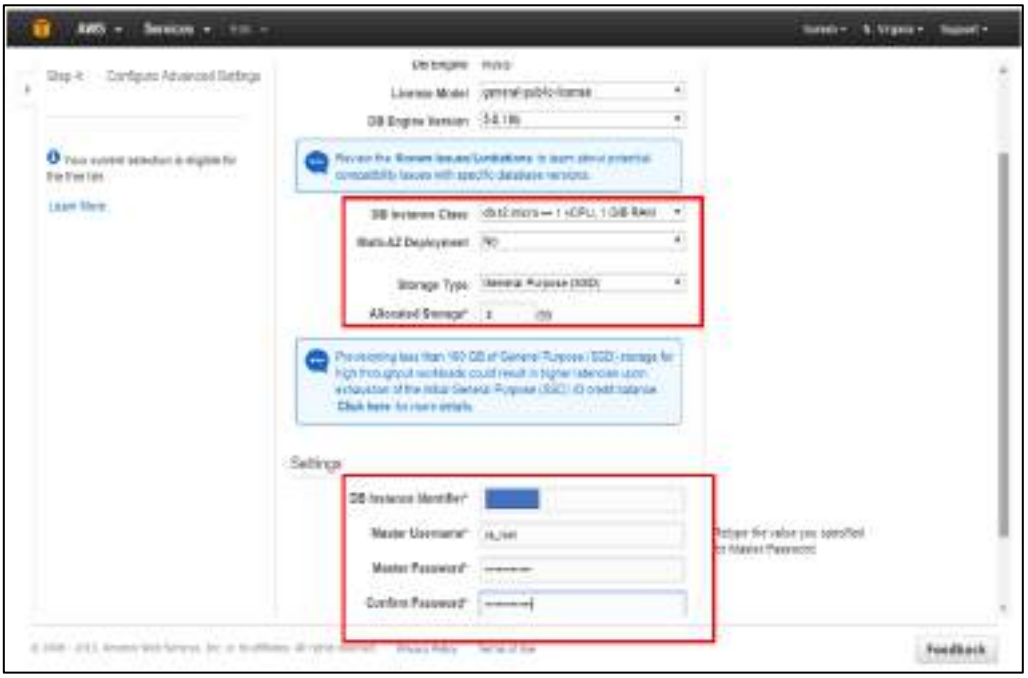
- This will lead us to ‘Select Engine’ screen.
- The Databases supported by Amazon are listed here. Select ‘MySQL’

- The next screen asks us if we are going to use this Database in Production.
- If you are going to use Database in a production environment, you need to select a Multi Availability Zone setup so that you uptime is high.
- You should also select Provisioned IOPS for high performance.
- In our case as we are not going to use this in production environment select, ‘No’ and click on ‘Next Step’.

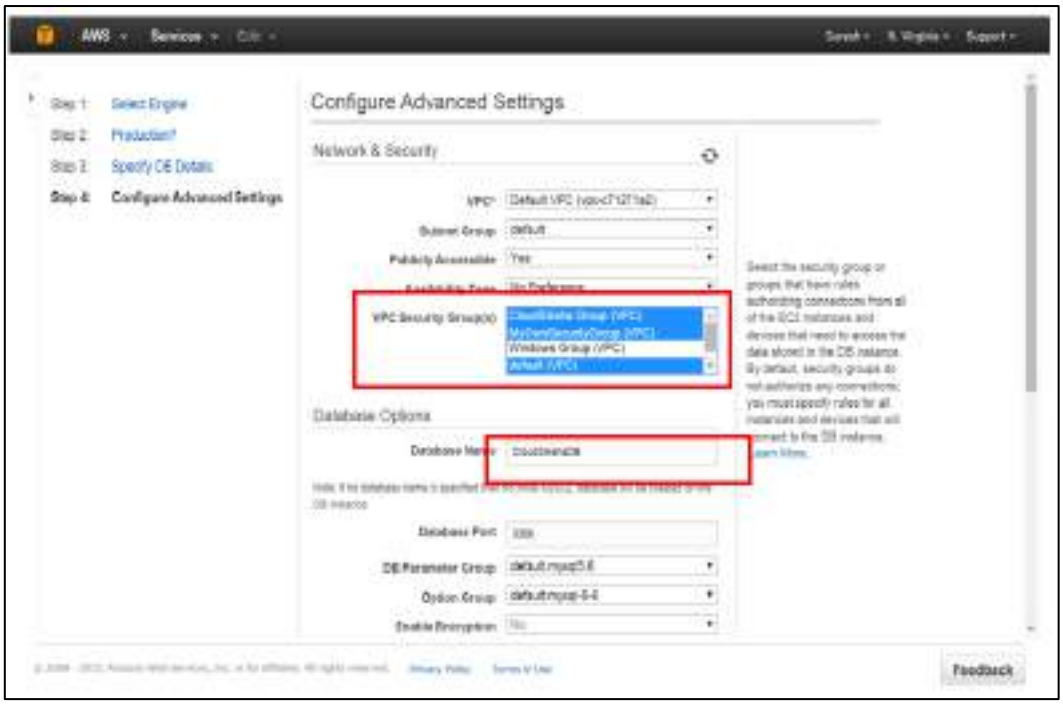


- In the DB Details screen, input the following:
  - Instance: db.t2.micro instance
  - Multi-AZ Deployment: No
  - Storage Type: General Purpose SSD

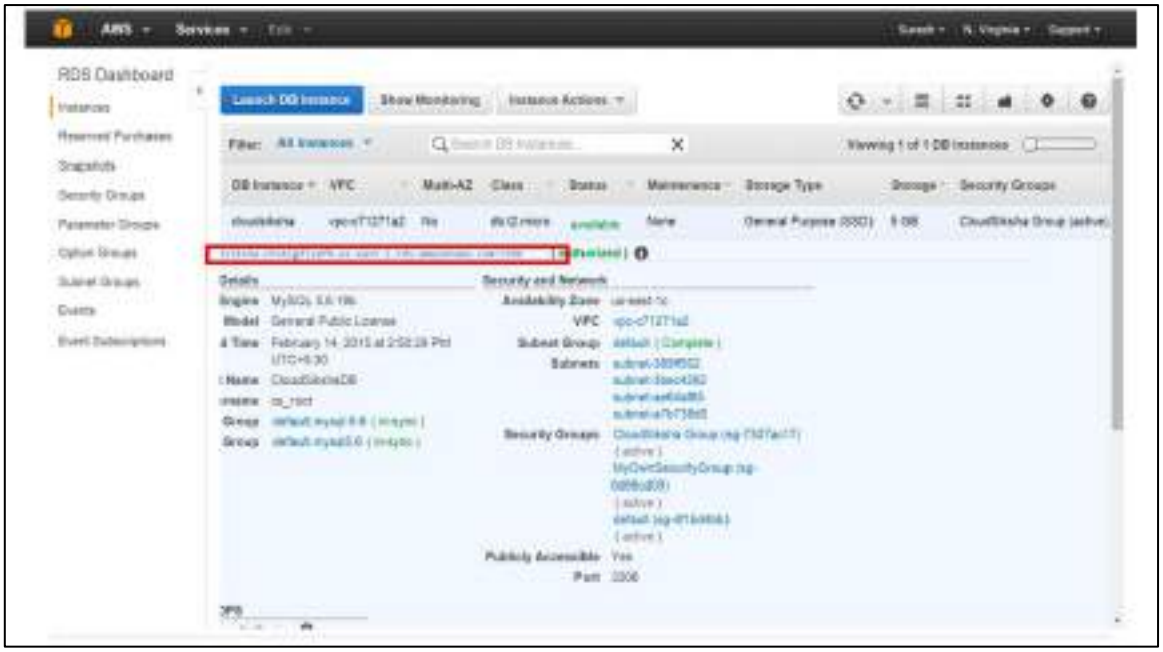
- Allocated Storage: 5GB
- DB Instance Identifier: CloudSiksha (you can give any name you want)
- Input a login id of your choice and a password of your choice
- Leave the others as default



- Press the ‘Next Step’ button (not seen in this screenshot)
- In the next page, select the Security groups which can access the DB. Also input a Database name of your choice. RDS will create the database when the instance comes up.



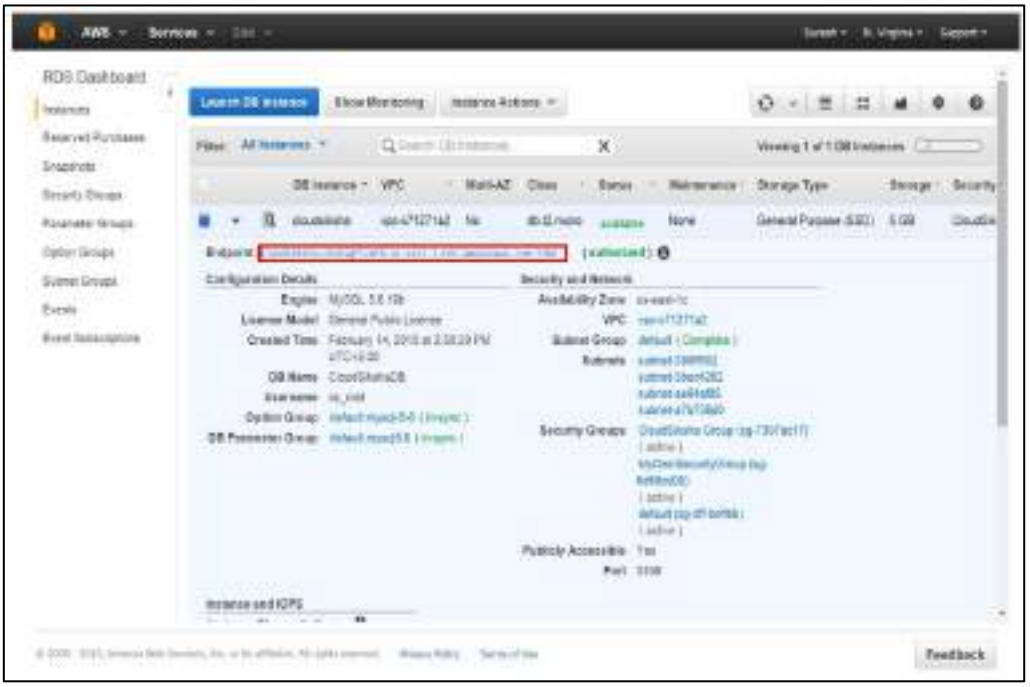
- In the Backup and Maintenance segment (not shown here) leave the defaults as they are and click ‘Launch Database’
- After the Database is launched, click on ‘View your DB Instance’.
- This will show all your DB instances. Click on the DB Instance to get the details.



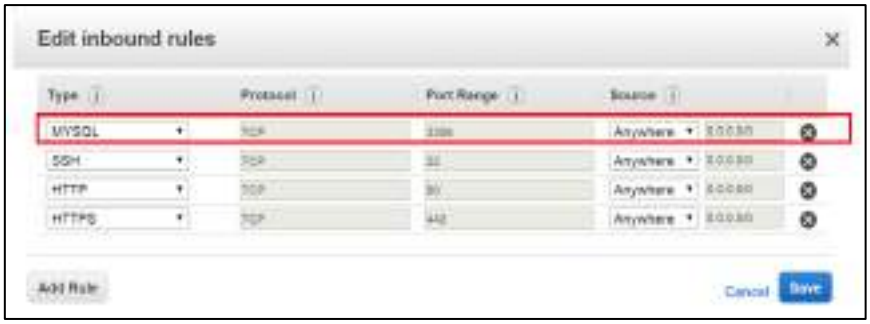
Note down the DNS address which we will use to connect to the DB

- Connecting to the Database
- Login into your instance. We will need mysql in order to access the database. We will install Apache Web Server, MySQL and PHP in our instance. Use this command to install these packages:

  $ sudo yum groupinstall -y "Web Server" "MySQL Database" "PHP Support" (please type)
- Check if all services are running by using $sudo service <service_name> status
- Find out the mysql end point from the RDS Instance screen. Note the name of the endpoint. The port number at the end is not required.



- Next, ensure that your security groups have given permission for the mysql port which is 3306.



- From your instance terminal issue the following command in order to connect to the database
- $ mysql –u<user_name> -p –h<endpoint> (example is given below)
- $mysql -ucs_root -p -hcloudsiksha.cbskigfizwfb.us-east-1.rds.amazonaws.com

- Enter your password, you will be logged into your database.

- Issue ‘show databases’ command and you will see that the database name we had given during RDS instance creation has been created.
- You can now start using the database.

----------End of the Module----------

MODULE 10

MULTIPLE CHOICE QUESTIONS

1. Which cloud is deployed when there is a budget constraint but business autonomy is most essential?

1. Private cloud
2. Public cloud
3. Hybrid cloud
4. Community cloud

2. What is Cloud Computing replacing?

1. Corporate data centers
2. Expensive personal computer hardware
3. Expensive software upgrades
4. All of the above

3. _____as a Service is a cloud computing infrastructure that creates a development environment upon which applications may be built.

1. Infrastructure
2. Service
3. Platform
4. All of the mentioned

4. How many phases are present in Cloud Computing Planning?

1. 2
2. 3
3. 4
4. 5

5. What is the number one concern about cloud computing?

1. Too expensive
2. Security concerns
3. Too many platforms
4. Accessibility

6. What type of computing technology refers to services and applications that typically run on a distributed network through virtualized resources?

1. Distributed Computing
2. Cloud Computing
3. Soft Computing
4. Parallel Computing

7. Which of the following can be considered PaaS offering?

1. Google Maps
2. Gmail
3. Google Earth
4. All of the mentioned

8. Which of these companies is not a leader in cloud computing?

1. Google
2. Amazon
3. Blackboard
4. Microsoft

9. Point out the wrong statement:

1. in cloud computing user don't have to worry about data backup and recovery
2. cloud computing can be used by small as well as big organisation
3. Cloud offer almost unlimited storage capacity
4. All applications benefit from deployment in the cloud

10. An internal cloud is…

1. An overhanging threat
2. A career risk for a CIO
3. A cloud that sits behind a corporate firewall
4. The group of knowledge workers who use a social network for water-cooler gossip

11. Cloud computing architecture is a combination of?

1. service-oriented architecture and grid computing
2. Utility computing and event-driven architecture.
3. Service-oriented architecture and event-driven architecture
4. Virtualization and event-driven architecture

12. Which of the following is most important area of concern in cloud computing?

1. Security
2. Storage
3. Scalability
4. All of the mentioned

13. Match the provider with the cloud-based service.

1. Amazon1. Azure
2. IBM2. Elastic Compute Cloud
3. EMC3. Decho
4. Microsoft4. Cloudburst

14. Which one of the following cloud concepts is related to sharing and pooling the resources?

1. Polymorphism
2. Virtualization
3. Abstraction
4. None of the mentioned

15. All cloud computing applications suffer from the inherent that is intrinsic in their WAN connectivity.

1. propagation
2. latency
3. noise
4. None of the mentioned

16. What August event was widely seen as an example of the risky nature of cloud computing?

1. Spread of Conficker virus
2. Gmail outage for more than an hour
3. Theft of identities over the Internet
4. Power outages in the Midwest

17. Which one of the following can be considered as a utility is a dream that dates from the beginning of the computing industry itself?

1. Computing

2. Model
3. Software
4. All of the mentioned

18. You can't count on a cloud provider maintaining your in the face of government actions.

1. scalability
2. reliability
3. privacy
4. none of the mentioned

19. Which of the following is true of cloud computing?

1. It's always going to be less expensive and more secure than local computing.
2. You can access your data from any computer in the world, as long as you have an Internet connection.
3. Only a few small companies are investing in the technology, making it a risky venture.

20. Which of the architectural layer is used as backend in cloud computing?

1. client
2. cloud
3. software
4. Network

21. Which of the following is an essential concept related to Cloud?

1. Reliability
2. Abstraction
3. Productivity
4. All of the mentioned

22. What is private cloud?

1. A standard cloud service offered via the Internet
2. A cloud architecture maintained within an enterprise data center.
3. A cloud service inaccessible to anyone but the cultural elite

23. Which one of the following is Cloud Platform by Amazon?

1. Azure
2. AWS
3. Cloudera
4. All of the mentioned

24. What is Cloud Foundry?

1. A factory that produces cloud components
2. An industry wide PaaS initiative
3. VMware-led open-source PaaS

25. Which of the following statement is not true?

1. Through cloud computing, one can begin with very small and become big in a rapid manner.
2. All applications benefit from deployment in the Cloud
3. Cloud computing is revolutionary, even though the technology it is built on is evolutionary.
4. None of the mentioned

26. Point out the wrong statement:

1. The vendor is responsible for all the operational aspects of the service
2. The customer is responsible only for his interaction with the platform
3. Google's App Engine platform is PaaS offering
4. SaaS require specific application to be accessed globally over the internet

27. Which of the following is specified parameter of SLA?

1. Response times
2. Responsibilities of each party
3. Warranties
4. All of the mentioned

28. This is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet.

1. Platform as a Service (PaaS)
2. Infrastructure as a Service (IaaS)
3. Software as a Service (SaaS)

29. Cloud computing shifts capital expenditures into expenditures.

1. operating
2. service
3. local
4. none of the mentioned

30. In the Planning Phase, which of the following is the correct step for performing the analysis?

1. Cloud Computing Value Proposition
2. Cloud Computing Strategy Planning
3. Both A and B
4. Business Architecture Development

31. "Cloud" in cloud computing represents what?

1. Wireless
2. Hard drives
3. People
4. Internet

32. What is the biggest disadvantage of community cloud?

1. Collaboration has to be maintained with other participants
2. Less security features
3. Cloud is used by many organisations for different purposes
4. Organisation losses business autonomy

33. Which of the following is Cloud Platform by Amazon?

1. Azure
2. AWS
3. Cloudera
4. All of the mentioned

34. In which one of the following, a strategy record or Document is created respectively to the events, conditions a user may face while applying cloud computing mode.

1. Cloud Computing Value Proposition
2. Cloud Computing Strategy Planning
3. Planning Phase
4. Business Architecture Development

35. What second programming language did Google add for App Engine development?

1. C++
2. Flash
3. Java
4. Visual Basic

36. enables batch processing, which greatly speeds up high-processing applications.

1. Scalability
2. Reliability
3. Elasticity
4. Utility

37. What is Business Architecture Development?

1. We recognize the risks that might be caused by cloud computing application from a business perspective.
2. We identify the applications that support the business processes and the technologies required to support enterprise applications and data systems.
3. We formulate all kinds of plans that are required to transform the current business to cloud computing modes.
4. None of the above

38. What facet of cloud computing helps to guard against downtime and determines costs?

1. Service-level agreements
2. Application programming interfaces
3. Virtual private networks
4. Bandwidth fees

39. Which one of the following refers to the non-functional requirements like disaster recovery, security, reliability, et

1. Service Development
2. Quality of service
3. Plan Development
4. Technical Service

40. Which of these is not a major type of cloud computing usage?

1. Hardware as a Service
2. Platform as a Service
3. Software as a Service
4. Infrastructure as a Service

41. Which one of the following is related to the services provided by Cloud?

1. Sourcing
2. Ownership
3. Reliability
4. PaaS

42. SaaS supports multiple users and provides a shared data model through model.

1. single-tenancy
2. multi-tenancy
3. multiple-instance
4. all of the mentioned

43. Cloud Services have a_____ relationship with their customers.

1. Many-to-many
2. One-to-many
3. One-to-one

44. Which of the following is best known service model ?

1. SaaS
2. IaaS

3. PaaS
4. All of the mentioned

45. Which one of the following refers to the user's part of the Cloud Computing system?

1. Back End
2. Management
3. Infrastructure
4. Front End

46. What is the name of Rackspace's cloud service?

1. Cloud On-Demand
2. Cloud Servers
3. EC2

47. provides virtual machines, virtual storage, virtual infrastructure, and other hardware assets.

1. IaaS
2. SaaS
3. PaaS
4. All of the mentioned

48. Which one of the following can be considered as the example of the Front-end?

1. Web Browser
2. Google Compute Engine
3. Cisco Metapod
4. Amazon Web Services

49. Cloud Service consists of

1. Platform, Software, Infrastructure
2. Software, Hardware, Infrastructure
3. Platform, Hardware, Infrastructure

50. Cloud computing is also a good option when the cost of infrastructure and management is

1. Low
2. High
3. Moderate
4. None of the mentioned

51. AWS Stands for Amazon Web Services.

A. True
B. False

52. _____ is a billing and account management service.

A. Amazon Mechanical Turk
B. Amazon Elastic MapReduce
C. Amazon DevPay
D. Multi-Factor Authentication

53. _____ is the central application in the AWS portfolio.

A. Amazon Simple Queue Service

B. Amazon Elastic Compute Cloud
C. Amazon Simple Notification Service
D. All of the above

54. Amazon Web Services falls into which of the following cloud-computing category?
A. Platform as a Service
B. Software as a Service
C. Infrastructure as a Service
D. Back-end as a Service

55. AWS reaches customers in _____countries.
A. 137
B. 182
C. 190
D. 86

56. S3 stands for Simple Storage Service
A. True
B. False

57. How many buckets can you create in AWS by default?
A. 100 buckets
B. 200 buckets
C. 110buckets
D. 125 buckets

58. What are the different types of instances?
A. General purpose
B. Computer Optimized
C. Storage Optimized
D. All of the above

59. What are the advantages of auto-scaling?
A. Better availability
B. Offers fault tolerance
C. Better cost management
D. All of the above

60. The types of AMI provided by AWS are:
A. Instance store backed
B. EBS backed
C. Both 1 & 2
D. None of the above

61. Storage classes available with Amazon S3 are –
A. Amazon S3 standard
B. Amazon S3 standard-infrequent Access
C. Amazon Glacier

D. All of the above

62. Amazon Web Services supports which Type II Audits?
A. SAS20
B. SAS70
C. SAS702
D. None of the above

63. What is the Authentication in AWS?
A. Username/Password
B. Access Key
C. Access Key/ Session Token
D. All of the above

64. Which of the following is a message queue or transaction system for distributed Internet-based applications?
A. Amazon Simple Notification Service
B. Amazon Elastic Compute Cloud
C. Amazon Simple Queue Service
D. Amazon Simple Storage System

65. Which of the following is an online backup and storage system?
A. Amazon Simple Queue Service
B. Amazon Elastic Compute Cloud
C. Amazon Simple Notification Service
D. Amazon Simple Storage System

66. Which of the following statement is wrong about Amazon S3?
A. Amazon S3 provides large quantities of reliable storage that is highly protected
B. Amazon S3 is highly available
C. Amazon S3 is highly reliable
D. All of the above

67. Which service performs the function that when an instance is healthy it is terminated and replaced with a new one?
A. Sticky Sessions
B. Fault Tolerance
C. Connection Draining
D. None of the above

68. Amazon S3 is which type of storage service?
A. Block
B. Object
C. Simple
D. Secure

69. Amazon S3 offers encryption services for:
A. Data in Flight

B. Data in Motion
C. Data in Rest
D. Both 1 & 2

70. A virtual CloudFront user is called an OAI. This stands for what?
A. Origin Archive Initiative
B. Origin Access Identity
C. Original Archive Identity
D. Original Accessible Initiative

71. What is Cloud Computing?
a) Cloud Computing means providing services like storage, servers, database, networking, etc
b) Cloud Computing means storing data in a database
c) Cloud Computing is a tool used to create an application
d) None of the mentioned

72. Who is the father of cloud computing?
a) Sharon B. Codd
b) Edgar Frank Codd
c) J.C.R. Licklider
d) Charles Bachman

73. Which of the following is not a type of cloud server?
a) Public Cloud Servers
b) Private Cloud Servers
c) Dedicated Cloud Servers
d) Merged Cloud Servers

74. Which of the following are the features of cloud computing?
a) Security
b) Availability
c) Large Network Access
d) All of the mentioned

75. Which of the following is a type of cloud computing service?
a) Service-as-a-Software (SaaS)
b) Software-and-a-Server (SaaS)
c) Software-as-a-Service (SaaS)
d) Software-as-a-Server (SaaS)

76. Which of the following is the application of cloud computing?
a) Adobe
b) Paypal
c) Google G Suite
d) All of the above

77. Which of the following is an example of the cloud?
a) Amazon Web Services (AWS)

b) Dropbox
c) Cisco WebEx
d) All of the above

78. Applications and services that run on a distributed network using virtualized resources is known as _____
a) Parallel computing
b) Soft computing
c) Distributed computing
d) Cloud computing

79. Which of the following is an example of a PaaS cloud service?
a) Heroku
b) AWS Elastic Beanstalk
c) Windows Azure
d) All of the above

80. Which of the following is an example of an IaaS Cloud service?
a) DigitalOcean
b) Linode
c) Rackspace
d) All of the above

81. Which of the following is the correct statement about cloud computing?
a) Cloud computing abstracts systems by pooling and sharing resources
b) Cloud computing is nothing more than the Internet
c) The use of the word “cloud” makes reference to the two essential concepts
d) All of the mentioned

82. Point out the wrong statement.
a) Azure enables .NET Framework applications to run over the Internet
b) Cloud Computing has two distinct sets of models
c) Amazon has built a worldwide network of data centers to service its search engine
d) None of the mentioned

83. Which of the following model attempts to categorize a cloud network based on four dimensional factors?
a) Cloud Cube
b) Cloud Square
c) Cloud Service
d) All of the mentioned

84. Which of the following is the correct statement about cloud types?
a) Cloud Square Model is meant to show is that the traditional notion of a network boundary being the network’s firewall no longer applies in cloud computing
b) A deployment model defines the purpose of the cloud and the nature of how the cloud is located
c) Service model defines the purpose of the cloud and the nature of how the cloud is located

d) All of the mentioned

85. Which architectural layer is used as a backend in cloud computing?
a) cloud
b) soft
c) client
d) all of the mentioned

86. All cloud computing applications suffer from the inherent _____ that is intrinsic in their WAN connectivity.
a) noise
b) propagation
c) latency
d) all of the mentioned

87. Which of the following architectural standards is working with cloud computing industry?
a) Web-application frameworks
b) Service-oriented architecture
c) Standardized Web services
d) All of the mentioned

88. Which of the following is the correct statement?
a) Cloud computing presents new opportunities to users and developers
b) Service Level Agreements (SLAs) is small aspect of cloud computing
c) Cloud computing does not have impact on software licensing
d) All of the mentioned

89. What is the correct formula to calculate the cost of a cloud computing deployment?
a) CostCLOUD = Σ(UnitCostCLOUD / (Revenue + CostCLOUD))
b) CostCLOUD = Σ(UnitCostCLOUD / (Revenue – CostCLOUD))
c) CostCLOUD = Σ(UnitCostCLOUD x (Revenue – CostCLOUD))
d) None of the mentioned

90. Which of the following is the wrong statement about cloud computing?
a) Private cloud doesn’t employ the same level of virtualization
b) Data center operates under average loads
c) Private cloud doesn’t pooling of resources that a cloud computing provider can achieve
d) Abstraction enables the key benefit of cloud computing: shared, ubiquitous access

91. Identify the wrong statement about cloud computing.
a) Virtualization assigns a logical name for a physical resource and then provides a pointer to that physical resource when a request is made
b) Virtual appliances are becoming a very important standard cloud computing deployment object
c) Cloud computing requires some standard protocols
d) None of the mentioned

92. Identify the correct statement about cloud computing.
a) Cloud computing relies on a set of protocols needed to manage inter-process communications
b) Platforms are used to create more complex software
c) Cloud architecture can couple software running on virtualized hardware in multiple locations to provide an on-demand service
d) All of the mentioned

93. Point out the wrong statement.
a) Cloud services span the gamut of computer applications
b) The impact of cloud computing on network communication is to discourage the use of open-source network protocols in place of proprietary protocol
c) Atom is a syndication format that allows for HTTP protocols to create and update information
d) None of the mentioned

94. Which of the following is required by Cloud Computing?
a) That the identity be authenticated
b) That the authentication be portable
c) That you establish an identity
d) All of the mentioned

95. Cloud computing is a concept that involves pooling physical resources and offering them as which sort of resource?
a) cloud
b) real
c) virtual
d) none of the mentioned

96. Which of the following is the Cloud Platform provided by Amazon?
a) AWS
b) Cloudera
c) Azure
d) All of the mentioned

97. Into which expenditures does Cloud computing shifts capital expenditures?
a) local
b) operating
c) service
d) none of the mentioned

98. Point out the wrong statement.
a) With a pay-as-you-go, endlessly expandable, and universally available system, cloud computing realises the long-held goal of utility computing
b) The widespread use of the Internet enables the huge size of cloud computing systems
c) Soft computing represents a significant change in the way computers are delivered
d) All of the mentioned

99. Which of the following is the most essential element in cloud computing by CSA?
a) Virtualization

b) multi-tenancy
c) Identity and access management
d) All of the mentioned

100. Which of the following monitors the performance of the major cloud-based services in real time in Cloud Commons?
a) CloudWatch
b) CloudSensor
c) CloudMetrics
d) All of the mentioned

101. Which of the following model consists of the service that you can access on a cloud computing platform?
a) Deployment
b) Service
c) Application
d) None of the mentioned

102. Which of the following is the most important area of concern in cloud computing?
a) Scalability
b) Storage
c) Security
d) All of the mentioned

103. Which of the following is the most refined and restrictive cloud service model?
a) PaaS
b) IaaS
c) SaaS
d) CaaS

104. Which of the following is not a property of cloud computing?
a) virtualization
b) composability
c) scalability
d) all of the above

105. How many phases are there in Cloud Computing Planning?
a) 1
b) 5
c) 3
d) 6

106. Which of the following is an example of a SaaS cloud service?
a) Google Workspace
b) Dropbox
c) Salesforce
d) All of the above

107. Which is the most essential concept related to Cloud computing?
a) Abstraction
b) Reliability
c) Productivity
d) All of the mentioned

108. In which of the following service models the hardware is virtualized in the cloud?
a) NaaS
b) PaaS
c) CaaS
d) IaaS

109. Which of the following is the Virtual machine conversion cloud?
a) Amazon CloudWatch
b) AbiCloud
c) BMC Cloud Computing Initiative
d) None of the mentioned

110. Which of the following is a workflow control and policy-based automation service by CA?
a) CA Cloud Compose
b) CA Cloud Insight
c) CA Cloud Optimize
d) CA Cloud Orchestrate

111. An application that provides for transaction overflow in a reservation system is an example of:
a) Cloud bursting
b) Cloud provisioning
c) Cloud servicing
d) All of the mentioned

NOTES

Thank You!

STL academy

empower yourself

STL academy

Empowering Youth!

STL is one of the industry's leading integrators of digital networks providing All-in 5G solutions. Our capabilities across optical networking, services, software, and wireless connectivity place us amongst the top optical players in the world. These capabilities are built on converged architectures helping telcos, cloud companies, citizen networks, and large enterprises deliver next-gen experiences to their customers. STL collaborates with service providers globally in achieving a green and sustainable digital future in alignment with UN SDG goals. STL has a global presence in India, Italy, the UK, the US, China, and Brazil

ISO 21001 | Guinness World Records Record Holder | Asia Book of Records

Skill & Assessment Partner
NASSCOM®

stl.tech | stlacad.tech