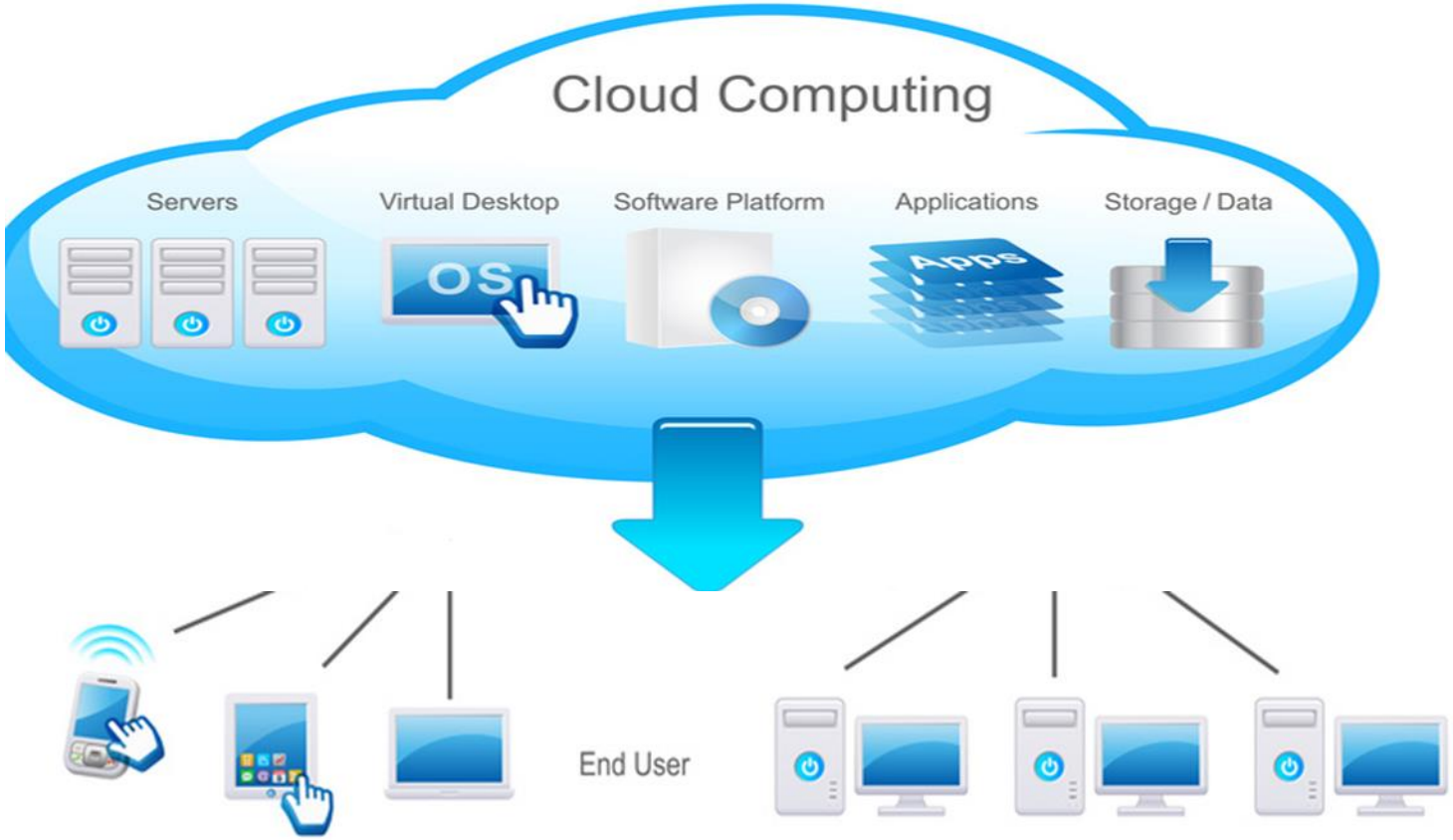
CLOUD COMPUTING

- Cloud computing is a **service provisioning technique** where **computing resources** like **hardware such as servers and storage devices**, **software's** and **complete platform for developing applications** are provided **as a service by the cloud providers to the customers.**



CLOUD COMPUTING (Cont…)

- Customers **can use these resources as and when needed**, **can increase or decrease resource capacities dynamically** according to their requirements and **pay only for how much the resource were used**.

- Customers **no need to invest money to purchase**, **manage** and **scale infrastructures**, **software upgradation** and **software licensing.**

Cloud Service Models

- The services that are provided by the cloud providers are broadly classified into three categories:

  - **Infrastructure-as-a-Service (IaaS)**

  - **Platform-as-a-Service (PaaS)**

  - **Software-as-a-Service (SaaS)**

Cloud Service Models

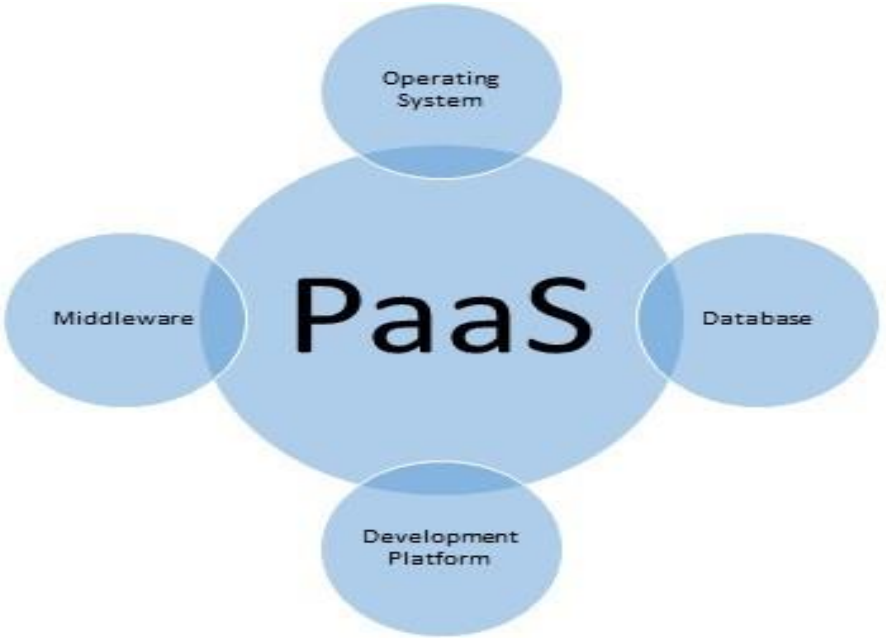
- **Infrastructure-as-a-Service (IaaS):** In Infrastructure-as-a-Service model, the service provider owns the hardware equipment's such as **Servers, Storage**, **Network** and is **provided as services** to the clients. The **client uses these equipment's and pays on per-use basis.**

- E.g. **Amazon Elastic Compute** (EC2) and **Simple Storage Service** (S3).

Cloud Service Models

- **Platform-as-a-Service (PaaS):** In Platform-as-a-Service model, **complete resources** needed to **Design**, **Develop**, **Testing**, **Deploy** and **Hosting** an application are provided as services **without spending money for purchasing and maintaining the servers, storage and software**.

- **PaaS is an extension of IaaS**. **In addition to the fundamental computing resource supplied by the hardware** in an IaaS offering, **PaaS models also include the software and configuration required to create an applications**.



- E.g. **Google App Engine**.

Cloud Service Models

- **Software-as-a-Service (SaaS):** In Software-as-a-Service model, the service provider provides **software's** as a service over the Internet, eliminating the need to **buy**, **install**, **maintain**, **upgradation and licensing** on their local machine.

- E.g. **Accounting, CRM, Google Docs** are all popular examples of SaaS.

Cloud Deployment Models

- Mainly there are four cloud deployment models (4 ways we can create/organize a cloud)
  - **Public Cloud**
  - **Private Cloud**
  - **Community Cloud**
  - **Hybrid Cloud**

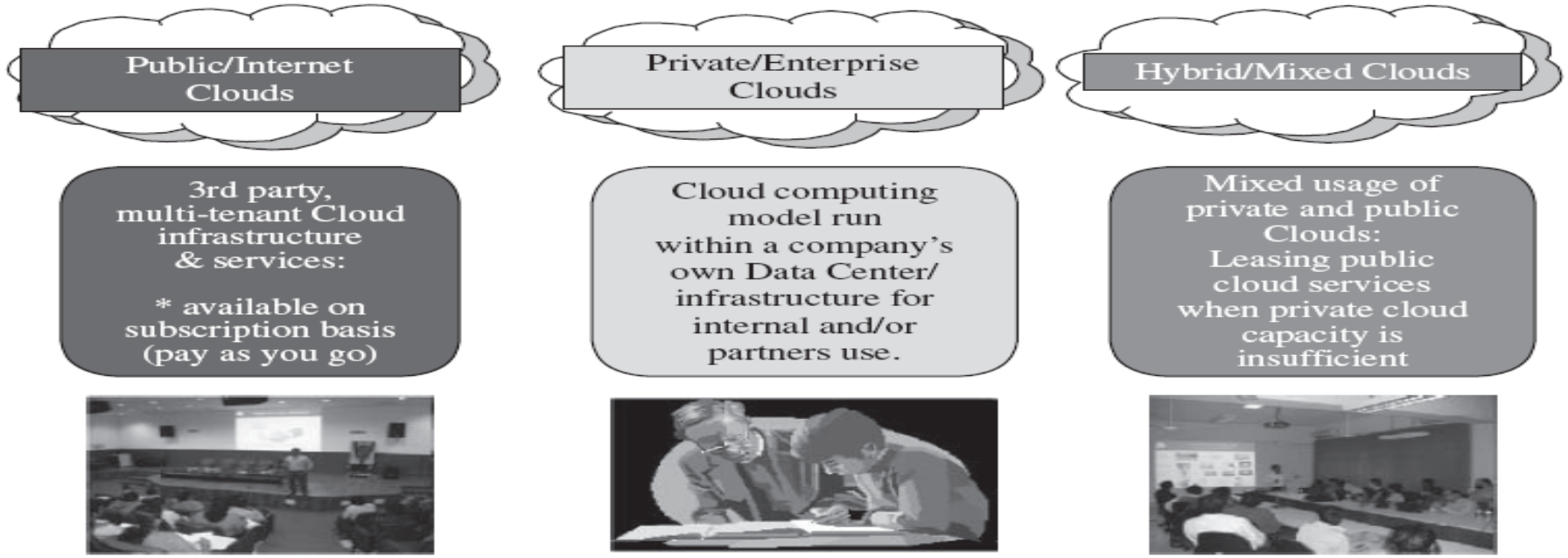


FIGURE 1.4. Types of clouds based on deployment models.

Cloud Deployment Models (Cont…)

- **Public Cloud**: A public cloud is a cloud in which **services and infrastructure are hosted off-site by a cloud provider** (owned by an organization selling cloud services) and easily **accessible to general public via internet**.

- **Private Cloud:** Private Cloud is a cloud where **services and infrastructure are operated for a single operation accessible via private network**, managed internally or by a third party. It is greater level of security.

Cloud Deployment Models (Cont…)

- **Community Cloud:** Community Cloud is a cloud where **services and infrastructure are accessible by a group of organizations**.

- **Hybrid Cloud:** Hybrid Cloud is a cloud which is a **mixture of private and public cloud**. In this type of cloud **all critical and sensitive applications and data are stored in private cloud** and **non critical and non sensitive applications and data are stored in public cloud**.

Features of Cloud Computing

- **It is elastic:** Cloud computing is flexible in nature, where users can **scale up** and **scale down** the resources as needed.
- **Pay per use:** Usage is metered and user **pays only for how much the resources were used**.
- **Operation:** The services are completely **handled by the provider**.
- **Reduce capital cost:** No need to **invest money** on purchasing and maintaining of hardware and software, software licensing, training required for IT staff.
- **Remote accessibility:** Users can access **applications and data stored on cloud** from anywhere any time worldwide through a device with internet connection.
- **Better use of IT staff:** Staff with in enterprise need not worry on purchasing and maintaining of servers, softwares, up gradation of servers and softwares, software licensing etc., instead they can concentrate more on work.

Cloud Services Examples:
IaaS-Amazon EC2, Google Compute Engine, Azure VMs

Amazon EC2

- **Amazon Elastic Compute Cloud** is an Infrastructure as a Service offering from Amazon.
- EC2 is a web service that **provides a computing capacity in the form of virtual machines**.
- Amazon EC2 allows **users to launch instances on demand using a simple web based interface**.
- Amazon provides **pre-configured Amazon Machine Images** (AMIs) which are templates of cloud instances.

	Small	Large	Extra Large	High CPU-Medium
Compute unit	1	4	8	5
Memory	1.7 GB	7.5 GB	15 GB	1.7 GB
Storage	160 GB	850 GB	1690 GB	350 GB
Platform	32 bit	64 bit	64 bit	32 bit

- Users can also **create their own AMIs with custom applications, libraries and data**.

Cloud Services Examples:
IaaS-Amazon EC2, Google Compute Engine, Azure VMs

Amazon EC2

- Amazon EC2 also provides **instances with high memory, high CPU resources, Cluster Compute instances, Cluster Graphical Processor unit (GPU)instances and high Input/Output instances**.

- Instances can be launched with a **variety of operating systems**.

- Users can **load their applications on running instances** and rapidly and easily **increase or decrease capacity to meet dynamic application performance requirements**.

- With EC2, **users can even provision hundreds or thousands of server instances simultaneously**, manage network access permissions and monitor usage resources through web interface. (Create 2 VMs at 3 different places and running applications, creating network among 3 and allowing data transfer among 3 VMs and monitoring resource usage)

Cloud Services Examples:
IaaS-Amazon EC2, Google Compute Engine, Azure VMs

Amazon EC2

- Amazon EC2 provides **instances of various computing capacities ranging from Small Instances** (Eg: 1 Virtual core with 1 EC2 compute unit, 1.7 GB memory and 160 GB instance storage) to **Extra Large Instances** (Eg: 4 Virtual cores with 2 EC@ compute unit each with 15GB memory and 1690 GB instance storage).

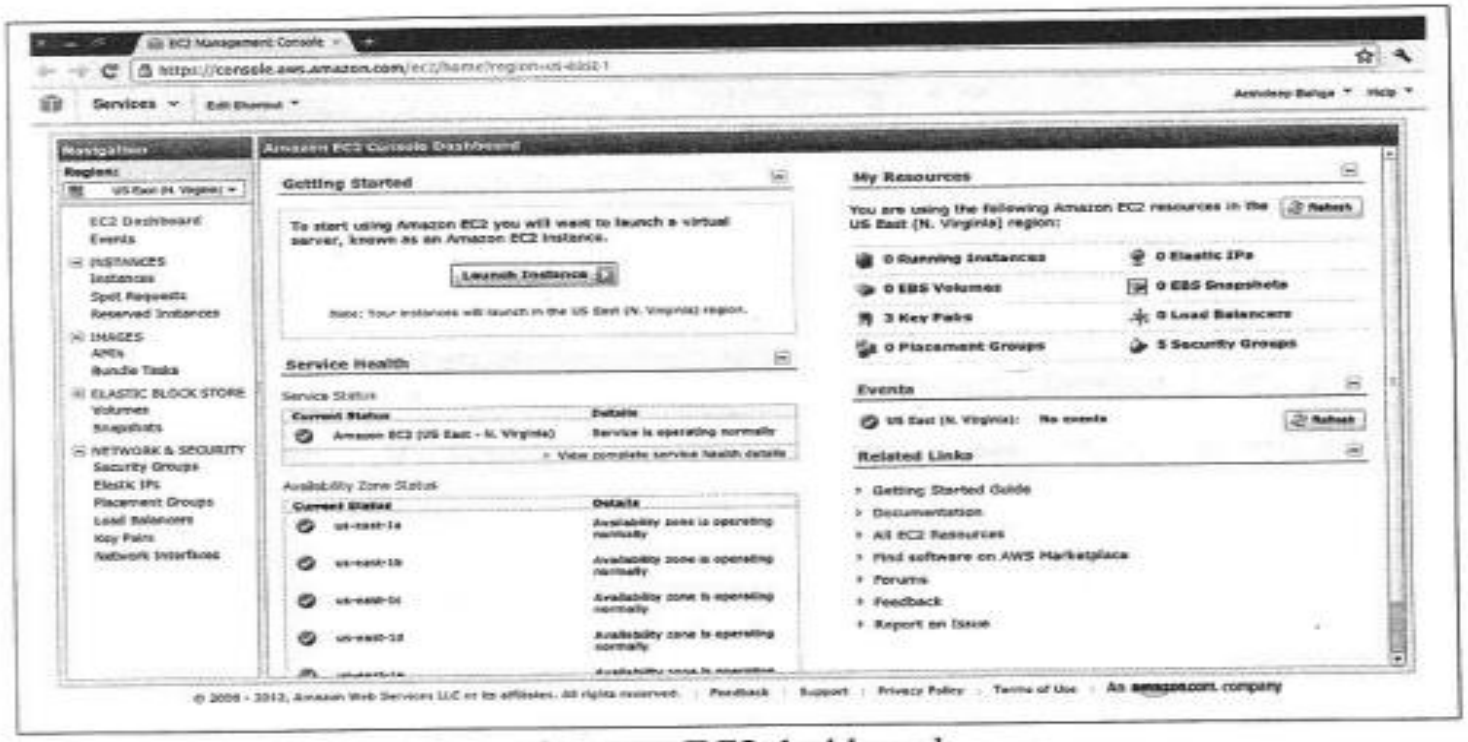
- The **pricing model for EC2 instances is based on Pay-Per Use model**. Users are billed based on the number on instance hours used for on demand instances.

- EC2 also provides **spot instances that allow users to bid on unused Amazon EC2 capacity and run those instances for as long as their bid exceeds the current spot price**.

Cloud Services Examples:
IaaS-Amazon EC2, Google Compute Engine, Azure VMs

Amazon EC2

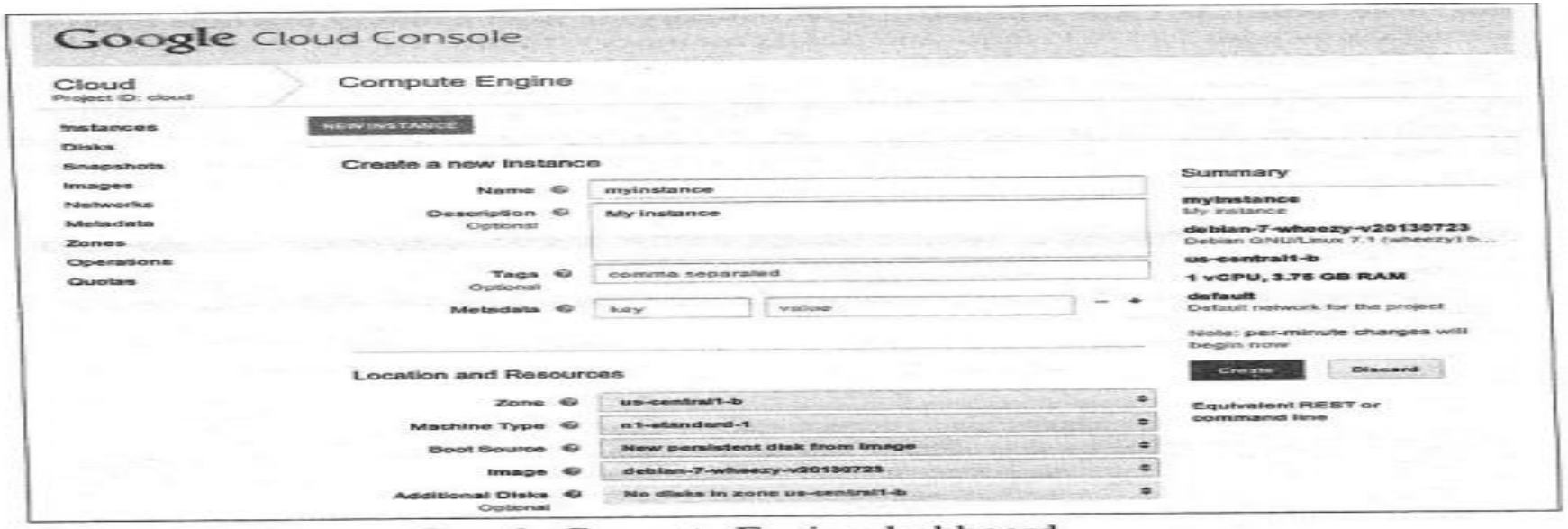
- The below figure shows screenshot of Amazon EC2 dashboard



Amazon EC2 dashboard

Cloud Services Examples:
IaaS-Amazon EC2, Google Compute Engine, Azure VMs
Google Compute Engine

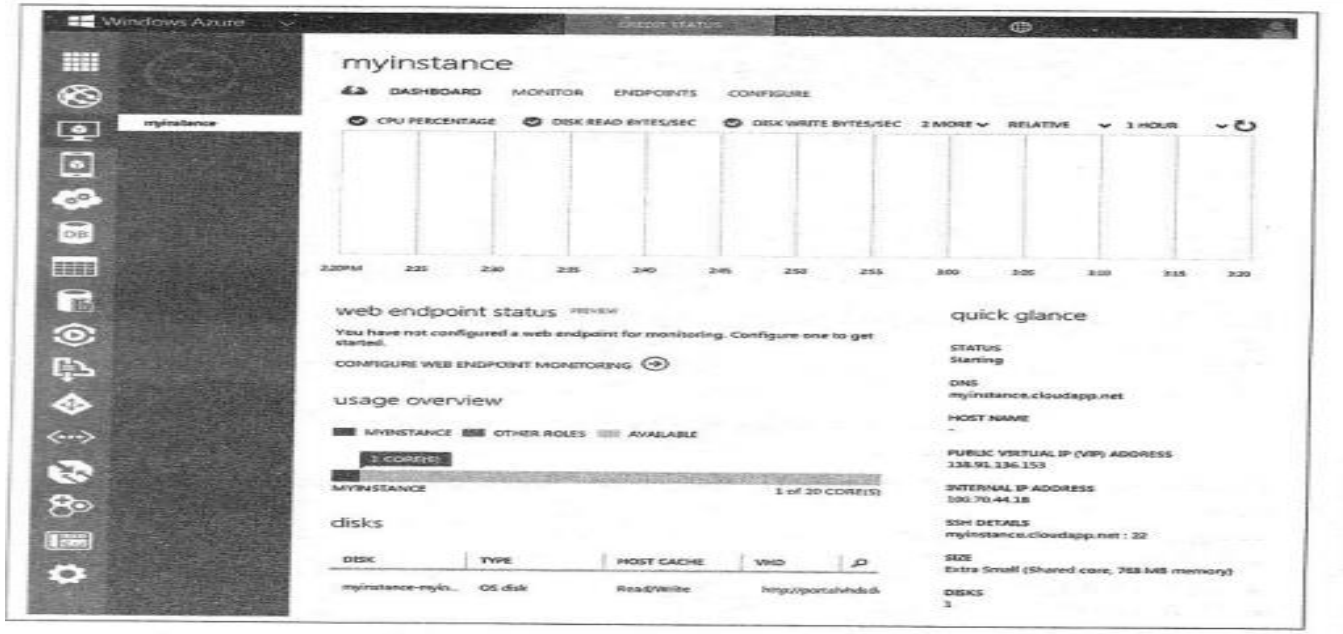
- **Google Compute Engine (GCE) is an IaaS offering from Google**.
- GCE provides **virtual machines of various computing capacities ranging from small instances** (Eg: Virtual core with 1.38 GCE unit and 1.7 GB memory) **to high memory machine types** (8 virtual cores with 22 GCE unit and 52 GB memory.
- The below figure shows screenshot of Google Compute Engine dashboard



Google Compute Engine dashboard

Cloud Services Examples:
IaaS-Amazon EC2, Google Compute Engine, Azure VMs
Windows Azure

- **Windows Azure Virtual Machine is an IaaS offering from Microsoft**.
- Azure VMs provides **virtual machines of various computing capacities ranging from small instances** (1 virtual core with 1.75GB memory) to **memory intensive machine types** (8 virtual cores with 56GB memory).
- The below figure shows screenshot of Google Compute Engine dashboard.



Windows Azure Virtual Machines dashboard

Cloud Services Examples:
PaaS-Google App Engine

- **Google App Engine (GAE) is a Platform as a Service offering from Google**.

- GAE is a cloud based **web service for hosting web applications and storing data**.

- GAE allows users to **build scalable and reliable applications that run on the same systems that power Google's own applications**.

- GAE provides a **Software Development Kit (SDK) for developing web applications software that can be deployed on GAE.**

Cloud Services Examples: PaaS-Google App Engine

- Developers can **develop and test their applications with GAE SDK on a local machine** and **then upload it to GAE** with a simple click of a button

- Applications hosted in GAE are **easy to build, maintain and scale**. Users **don't need to worry about launching additional computing instances when the applications load increases**.

- GAE provides **automatic scaling and load balancing capability**.

- GAE **supports applications written in several programming language**.

- With Java runtime environment developers can **build applications using Java programming language and standard Java technologies such as Java Servlets**. GAE also **provides runtime environment for Python programming languages**.

Cloud Services Examples: PaaS-Google App Engine

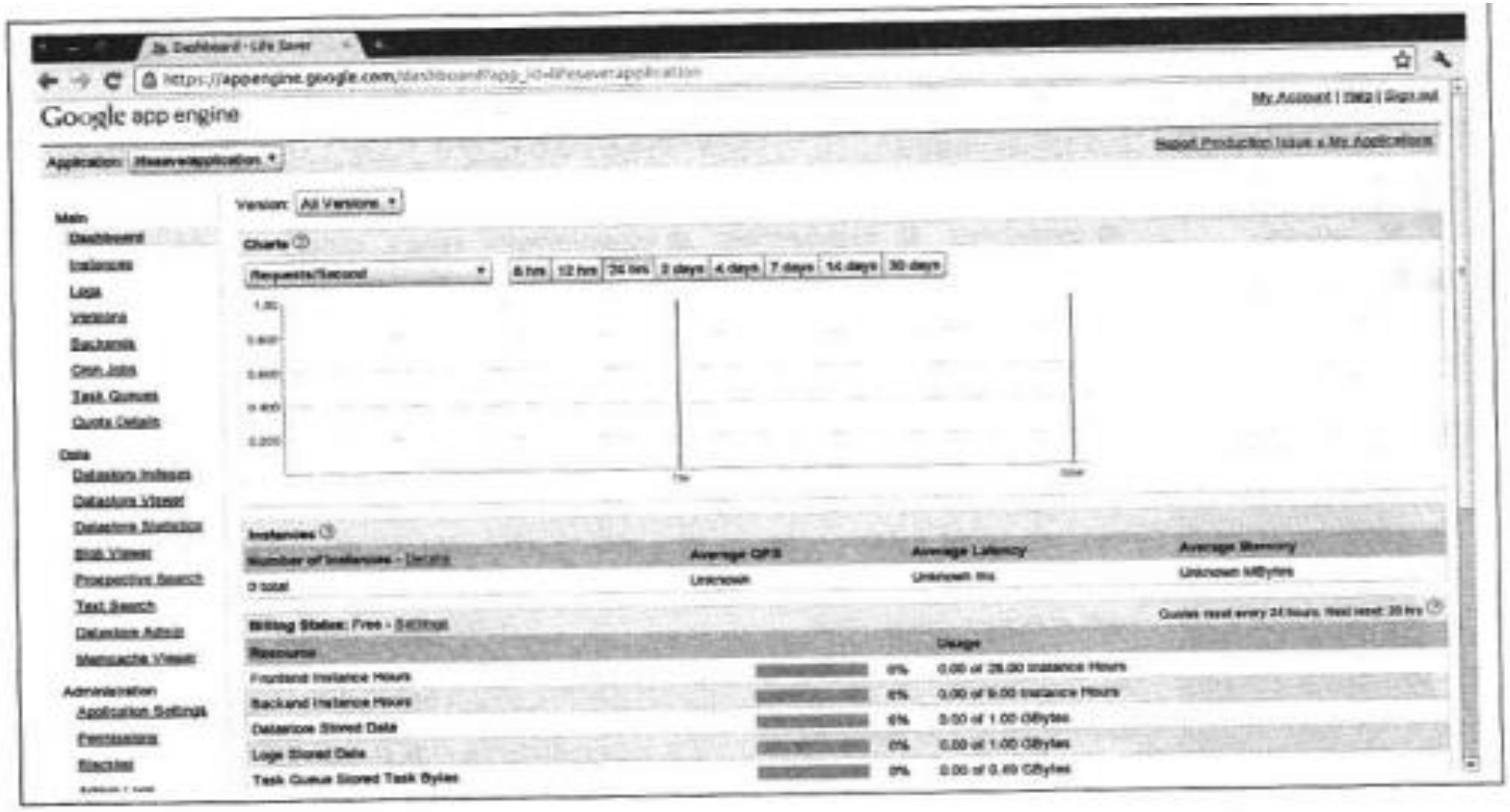
- **Applications hosted in GAE run in secure sandbox** with limited access to the underlying operating system and hardware.

- The **pricing model for GAE is based on the amount of computing resources used**.

- GAE provides **free computing resources for applications up to a certain limit**. **Beyond that limit, users are billed based on the amount of computing resources used such as amount of bandwidth consumed, number of resource instance hours, amount of data stored**.

Cloud Services Examples: PaaS-Google App Engine

- The below figure shows the screenshot of GAE dashboard.



Google App Engine dashboard

Cloud Services Examples: SaaS-Salesforce

Salesforce Sales Cloud

- Salesforce **Sales Cloud** is a cloud based **Customer Relationship Management (CRM) SaaS offering**.

- Users can **access CRM application from anywhere through internet enabled devices such as workstations, laptops, tablets and smartphones.**

- Sales Cloud allows **sales representatives to manage customer profiles**, **track opportunities**, **optimize campaigns from lead to close** and monitor the impact of campaigns. (A lead can be a company or an individual who has expressed interest in a company's product and/or service).

Cloud Services Examples: SaaS-Salesforce

Salesforce Service Cloud

- **Salesforce Service Cloud is a cloud based Customer Service Management SaaS**.

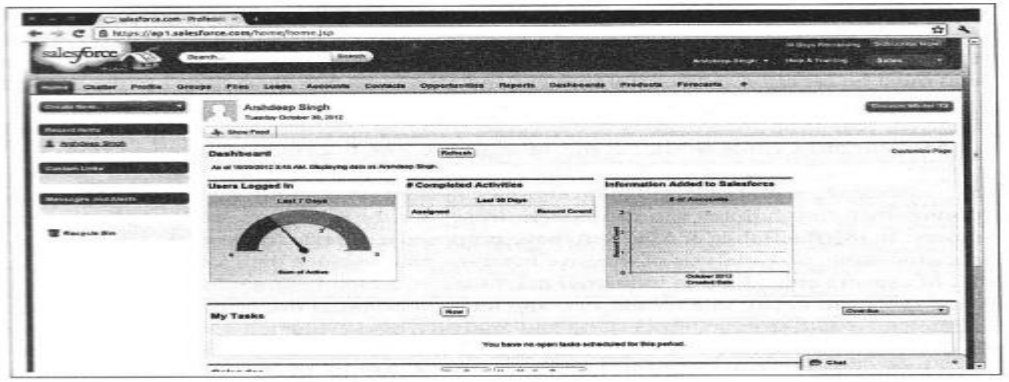
- Service cloud provides companies a call center like view and allows creating, tracking, routing and escalating cases.

- Service cloud includes a **social networking plug-in** that enables social customer service **where comments from social media channels can be used to answer customer questions**.

Cloud Services Examples: SaaS-Salesforce

Salesforce Marketing Cloud

- **Salesforce Marketing Cloud is cloud based social marketing SaaS**.
- Marketing cloud allows companies to **identify sales leads from social media**, discover advocates, identify most trending information on any topic.
- Marketing cloud **allows companies to pro-actively engage with customers**, manage **social advertisement campaigns** and track the performance of social campaigns.
- The below figure shows a screenshot of Salesforce dashboard



Salesforce dashboard

Cloud Services Examples: SaaS-Salesforce

Salesforce Marketing Cloud

- **Some of the tools included in the Salesforce Sales, Service and Marketing Clouds include**
  - **Accounts and Contacts**
  - **Leads**
  - **Opportunities**
  - **Campaigns**
  - **Chatter**
  - **Analytics an Forecasts**

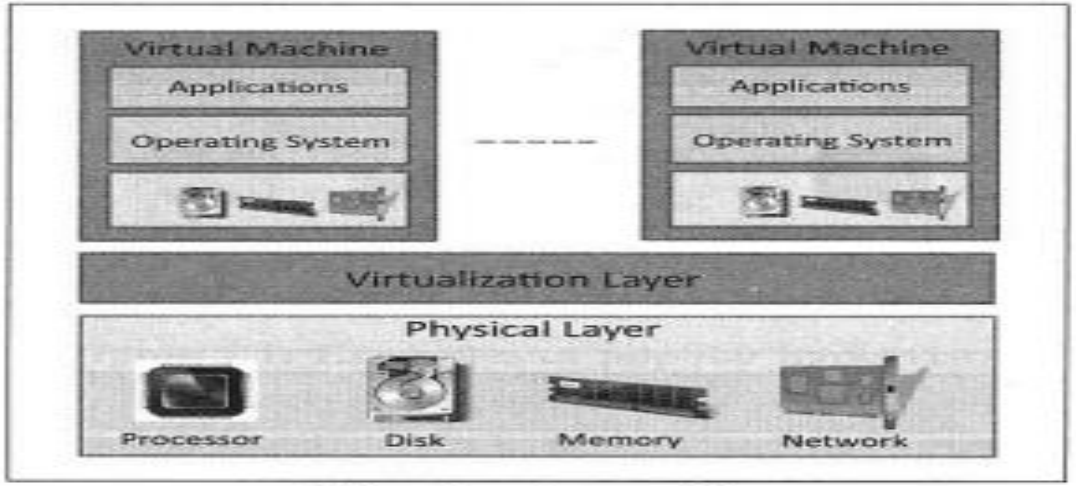
Cloud Concepts and Technologies

- **Virtualization**
- **Load Balancing**
- **Scalability and Elasticity**
- **Deployment**
- **Replication**
- **Monitoring**
- **Software Defined Networking**
- **MapReduce**
- **Identity and Access Management**
- **Service Level Agreements**
- **Billing**

Cloud Concepts and Technologies

Virtualization

- Virtualization refers to **the partitioning the resources of a physical system** (such as computing, Storage, Network and Memory) **into multiple virtual resources**.
- In cloud computing, **resources are pooled to serve multiple users using Multi-Tenancy**.
- Multi-Tenant aspects of the cloud **allow multiple users to be served by the same physical hardware**.
- The below figure shows the architecture of a virtualization technology in cloud computing.
- The physical resources such as **computing, storage, memory and network resources are virtualized**.
- The **virtualization layer partitions the physical resources into multiple virtual machines**.

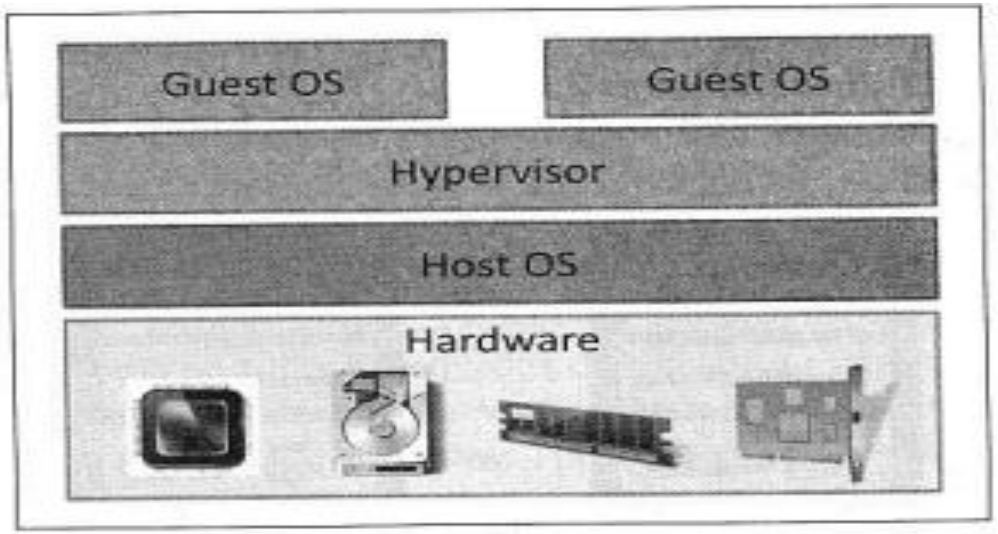


Virtualization architecture

Cloud Concepts and Technologies

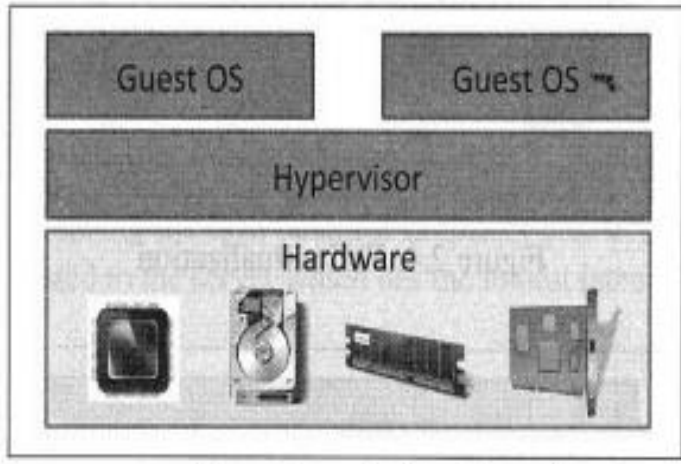
Virtualization: Guest Operating System

- A **guest OS is an operating system that is installed in a virtual machine** in addition to the host OS.
- In virtualization, the **guest OS can be different from the host OS**.

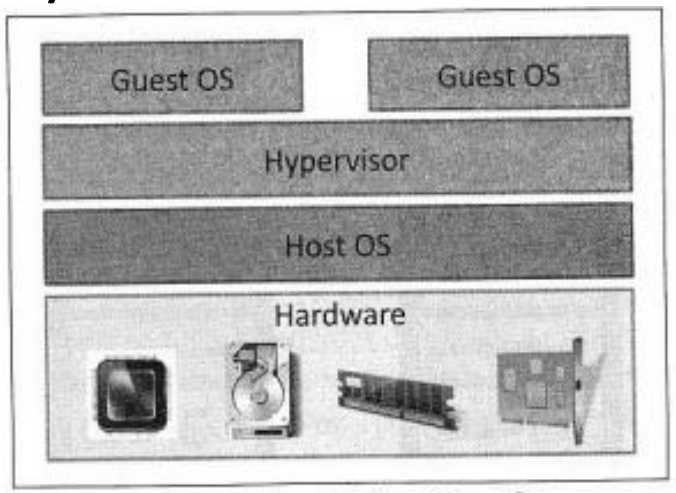


Cloud Concepts and Technologies
Virtualization: Hypervisor

- The **virtualization layer consists of a hypervisor or a Virtual Machine Monitor (VMM)**.
- There are two types of hypervisors
  - **Type-1 Hypervisors or Native Hypervisors**
  - **Type-2 Hypervisors or Hosted Hypervisors**



Hypervisor design: Type-1



Hypervisor design: Type-2

Type-1 Hypervisors or Native Hypervisors

- **Type-1 Hypervisors or Native Hypervisors run directly on the host hardware** and **control the hardware** and **monitor the guest operating system**.

Type 2 Hypervisors or Hosted Hypervisors

- **Type 2 Hypervisors or Hosted Hypervisors run on top of a conventional** (main or Host) **operating system** and **monitor the guest operation systems.**

Cloud Concepts and Technologies

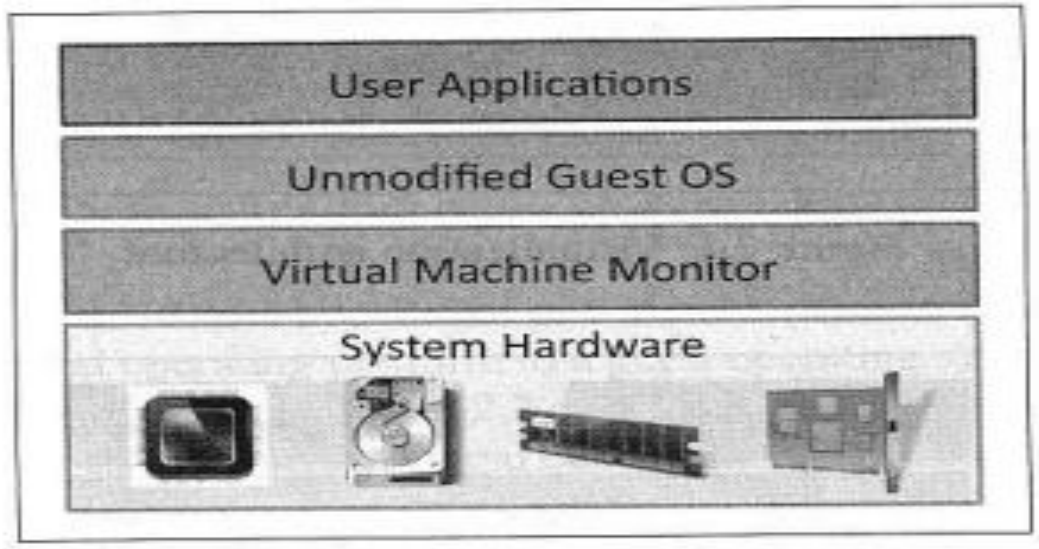
Virtualization: Guest Operating System

- Various forms of virtualization approaches exist:
  - **Full Virtualization**
  - **Para-Virtualization**
  - **Hardware Virtualization**

Cloud Concepts and Technologies

Virtualization: Full Virtualization

- In Full Virtualization, **the guest OS requires no modification** and **is not aware that it is being virtualized**.
- **Full virtualization is enabled by direct execution of user requests** and **binary translation of OS requests**.
- **The below figure shows the Full virtualization approach**



Full virtualization

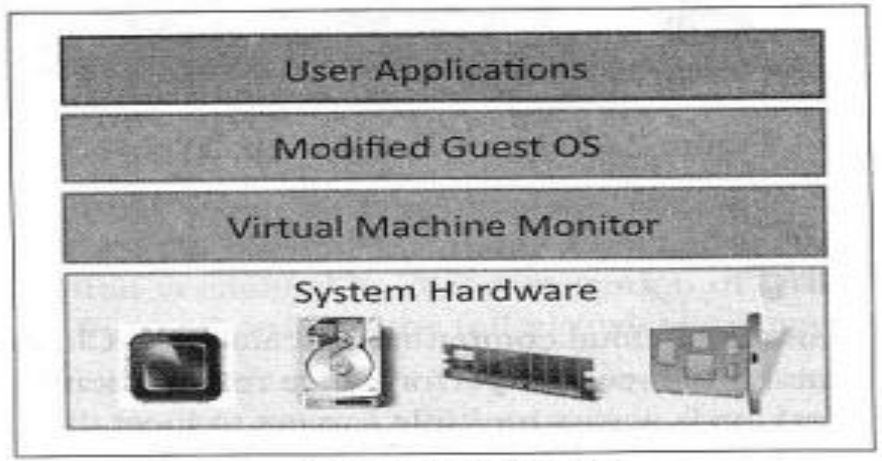
Cloud Concepts and Technologies

Virtualization: Para Virtualization

- In Para virtualization, **the guest OS is modified to enable communication with the hypervisor** to improve performance and efficiency.

- The **guest OS kernel is modified to replace non virtualizable instructions with hypercalls that communicate directly with the virtualization layer hypervisor.**

- The **below figure shows the para virtualization approach**



Para-virtualization

Cloud Concepts and Technologies

Virtualization: Hardware Assisted Virtualization

- Hardware Assisted virtualization **is enabled by hardware features such as Intel's Virtualization technology** (VT-x) and AMD's **AMD-V**. In hardware virtualization, privileged and sensitive calls are set to automatically trap to the hypervisor.

- Thus, **there is no needed for either binary translation or Para virtualization**. Hardware-assisted full virtualization **eliminates the binary translation** and it **directly interrupts with hardware** using the virtualization technology which has been integrated on X86 processors since 2005 (Intel VT-x and AMD-V).

- **Guest OS's instructions might allow a virtual context execute privileged instructions directly on the processor, even though it is virtualized**.

Cloud Concepts and Technologies

Load Balancing Technique

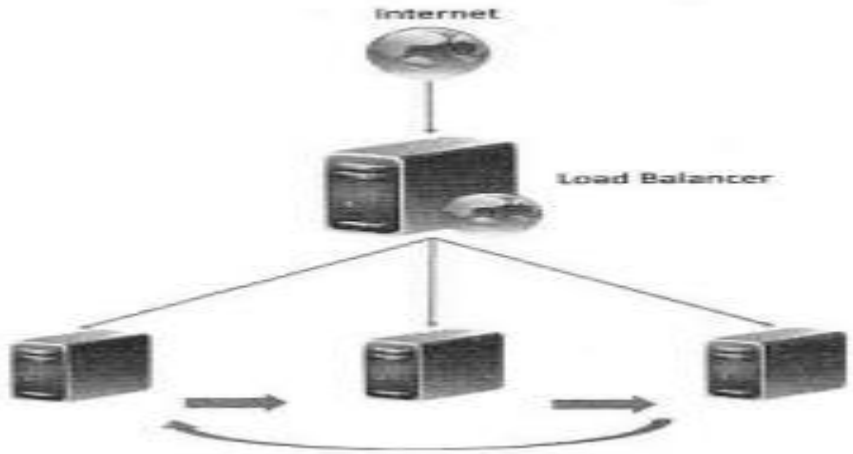
- One of the important features of cloud computing is **scalability**.
- Cloud resources can be **scaled up on demand to meet the performance requirements of applications**.
- **Load balancing distributes workload across multiple servers to meet the application workloads**.
- The **goal of load balancing techniques are to achieve maximum utilization of resources, minimize the response times, maximizing throughput**.
- Since multiple resources under a load balancer are used to serve the user requests, **in the event of failure of one or more of the resources**, the **load balancer can automatically reroute the user traffic to the healthy resources**.

Cloud Concepts and Technologies

Load Balancing Techniques

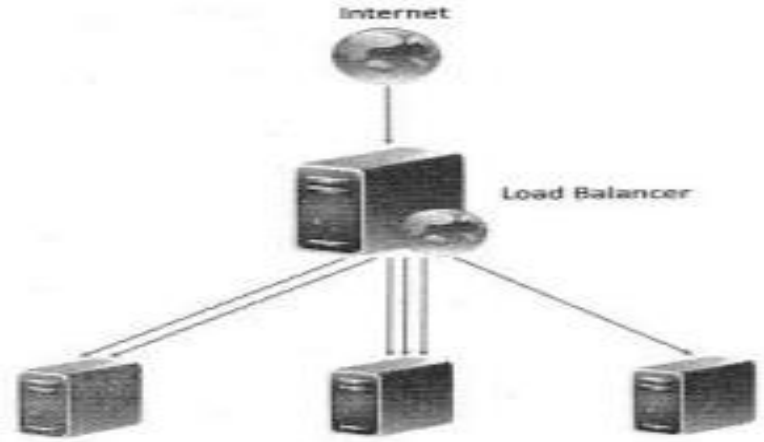
Round Robin

- **In round robin load balancing, the servers are selected one by one in a circular fashion to server the incoming requests from the user.**



Weighted Round Robin

- **In Weighted round robin load balancing, servers are assigned some weights. The incoming requests are proportionally routed to a server based on its weight.**

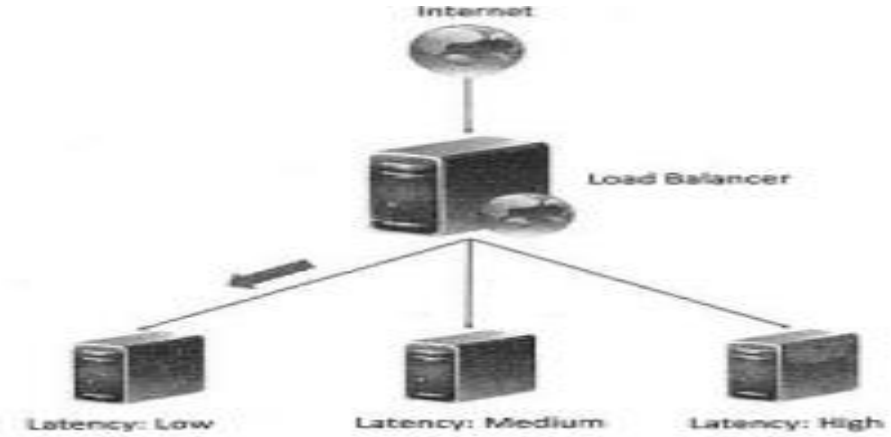


Cloud Concepts and Technologies

Load Balancing techniques

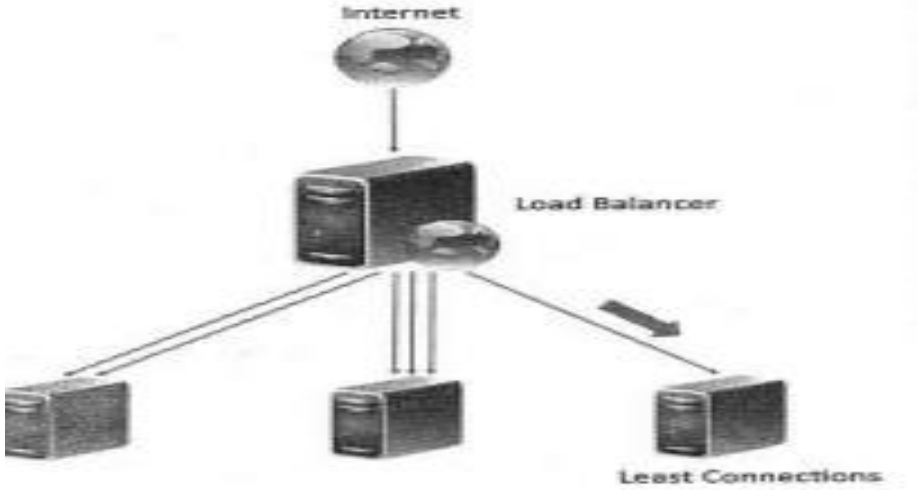
Low Latency

- **In low latency load balancing, the load balancer monitors the latency of each server and request is routed to a server which has lowest latency.**



Least Connections

- **In least connection load balancing, the incoming requests are routed to the server with least number of connections.**

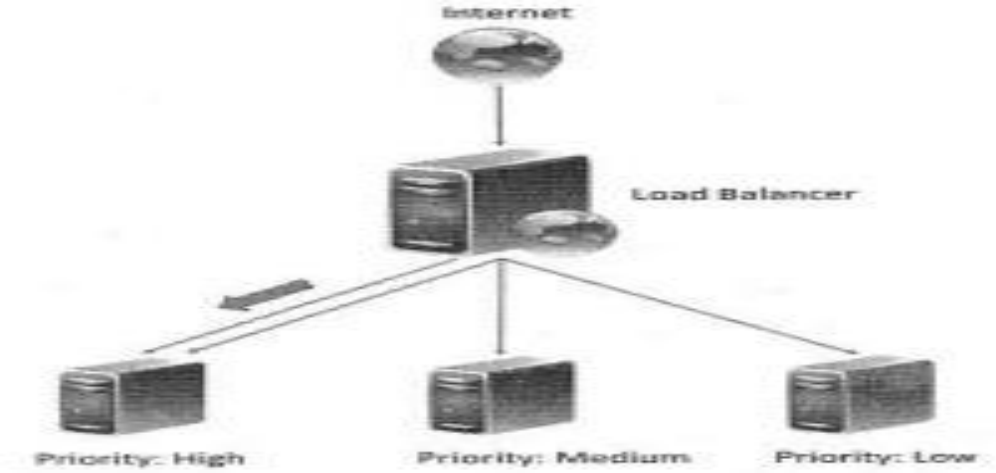


Cloud Concepts and Technologies

Load Balancing techniques

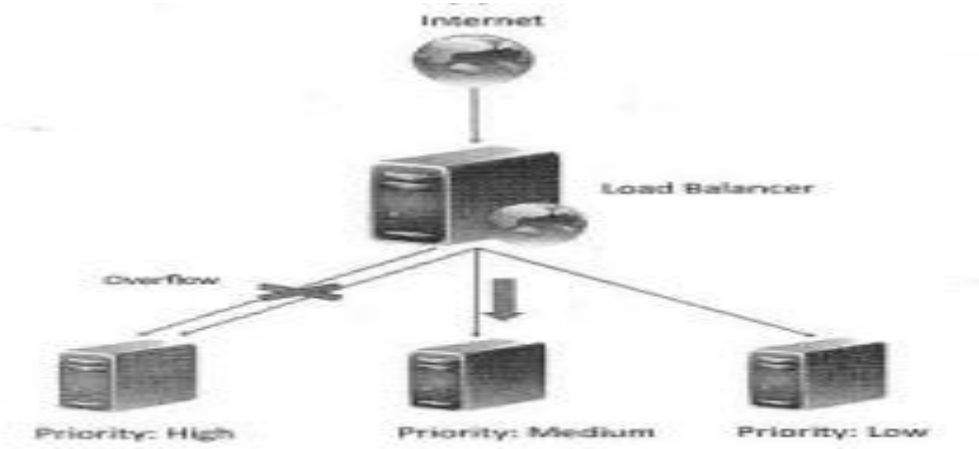
Priority

- **In priority load balancing, each server is assigned a priority. The incoming traffic is routed to the highest priority servers as long as the server is available. When the highest priority server fails, the incoming traffic is routed to a server with a lower priority.**



Overflow

- **Overflow load balancing is similar to priority load balancing. When the incoming request to high priority servers overflow, the requests are routed to a lower priority server.**



Cloud Concepts and Technologies

Load Balancing Techniques

- **A session is defined as a series of related browser requests that come from the same client during a certain time period.** Session tracking ties together a series of browser requests—think of these requests as pages—that may have some meaning as a whole, **such as a shopping cart application**.

- **A web session is a series of contiguous actions by a visitor on an individual website within a given time frame.** This could include your **search engine searches**, **filling out a form to receive content**, **scrolling on a website page**, **adding items to a shopping cart**, researching airfare, or which pages you viewed on a single website. **Any interaction that you have with a single website is recorded as a web session** to that website property.

- **To track sessions, a web session ID is stored in a visitor's browser**. This **session ID is passed along with any HTTP requests that the visitor makes while on the site** (e.g., clicking a link).

- ("Session" is the term used to refer to a visitor's time browsing a web site. It's meant to represent the **time between a visitor's first arrival at a page on the site and the time they stop using the site**.

- A **cookie is a small piece of data from a web site that is stored on a visitor's browser to help the website track the visitor's activity on the web site**.)

Cloud Concepts and Technologies

Load Balancing techniques

- For the **session based applications, an important issue to handle during load balancing is the persistence of multiple requests from a particular user session.** (Because you may go from current page to next page and back to previous page)
- **Since load balancing can route successive requests from a user session to different servers, maintain the state or the information of the session is important**. Four persistence approaches are:
  - **Sticky Sessions**
  - **Session Database**
  - **Browser Cookies**
  - **URL Re-Writing**

Cloud Concepts and Technologies

Load Balancing Techniques

Sticky Sessions

- In this approach, **all the requests belonging to a user session are routed to the same server**.
- **Theses sessions are called Sticky Sessions**.
- **The benefit of this approach is that it makes session management simple**.
- However, **a drawback of this approach is that if a server fails all the sessions belonging to that server are lost**, since there is **no automatic failover possible**.

Session Database

- In this approach, **the session information is stored externally in a separate session database**, which is often replicated **to avoid a single point of failure**.
- **Though, this approach involves additional overhead of storing the session information**, however unlike the sticky session approach, **this approach allows automatic failover**.

Cloud Concepts and Technologies

Load Balancing Techniques

Browser Cookies

- In this approach, **the session information is stored on the client side in the form of browser cookies**.
- **The benefit of this approach is that it makes the session management easy** and has the **least amount of overhead for the load balancer.**

URL Re-Writing

- In this approach, a **URL re-write engine stores the session information by modifying the URL's on the client side**.
- Though this approach avoids overhead on the load balancer, **a drawback is that the amount of session information that can be stored is limited.** (Modifying each URL→ Shortening URL→ storing in engine→ degrades performance)
- **For applications that require larger amounts of session information, this approach does not work.**

(Changing a URL to the required format. URL rewriting allows URLs to be more easily remembered by the user. When the URL is entered into the Web server, the URL rewrite engine modifies the syntax behind the scenes to enable the appropriate Web page or database item to be retrieved. For example, to look up the definition for "path," a user friendly URL might look like computerlanguage.com/path. The rewrite engine could turn it into the following syntax: computerlanguage.com/results.php?definition=path. So, the URL rewrite function simply puts a layer on top of the original address and turns it into something easy to find and that makes sense. Thus, turning https://wiredelta.com/?page_id=16825 into wiredelta.com/url-rewrite, for example. From a user perspective, when a URL rewrite occurs the URL of the website remains the same in the browser and they are none the wiser. But behind the scenes, the browser rewrites the URL back into that complicated mess and sends a query to the servers.)

Cloud Concepts and Technologies

Load Balancing Techniques

- **Load balancing can be implemented in software or hardware.**
- **Software based load balancers run on standard operating systems** and like other cloud resources.
- **Hardware load balancers implement load balancing algorithms in Application Specific Integrated Circuits** (ASICs).

Cloud Concepts and Technologies

Scalability and Elasticity

- **Muti-tier applications such as e-Commerce**, **Social networking**, **business-to-business** etc. **can experience rapid changes in their traffic.**

- **Each website has a different traffic pattern** which is determined by a number of factors that are **generally hard to predict beforehand.**

- **Capacity planning involves determining the right sizing of each tier of the deployment of an application in terms of number of resources and capacity of each resource**.

- **Capacity planning may be for computing, storage, memory or network resources**.

- Traditional approaches for capacity planning are based on **predicted demands for applications and account for worst case peak load as of applications.**

Cloud Concepts and Technologies

Scalability and Elasticity

- **When the workloads of applications increase, the traditional approaches have been either to scale up or scale out.**

- **Scaling up involves upgrading the hardware resources** (adding computing, memory, storage or network resources. **Scaling out involves addition of more resources of the same type**.

- Traditional scaling up and scaling out approaches are based on **demand forecasts at regular intervals of time**.

- When variations in workload are rapid, traditional approaches are **unable to keep track with the demand and either overprovisioning or under provisioning of resources**.

- **Over provisioning of resources of resources leads to higher capital expenditure** and **under provisioning of resources leads to traffic overloads, slow response time, low throughput.**