# UNIT- I

## CLOUD ARCHITECTURE AND MODEL

## ⇒ SCALABLE COMPUTING OVER THE INTERNET:

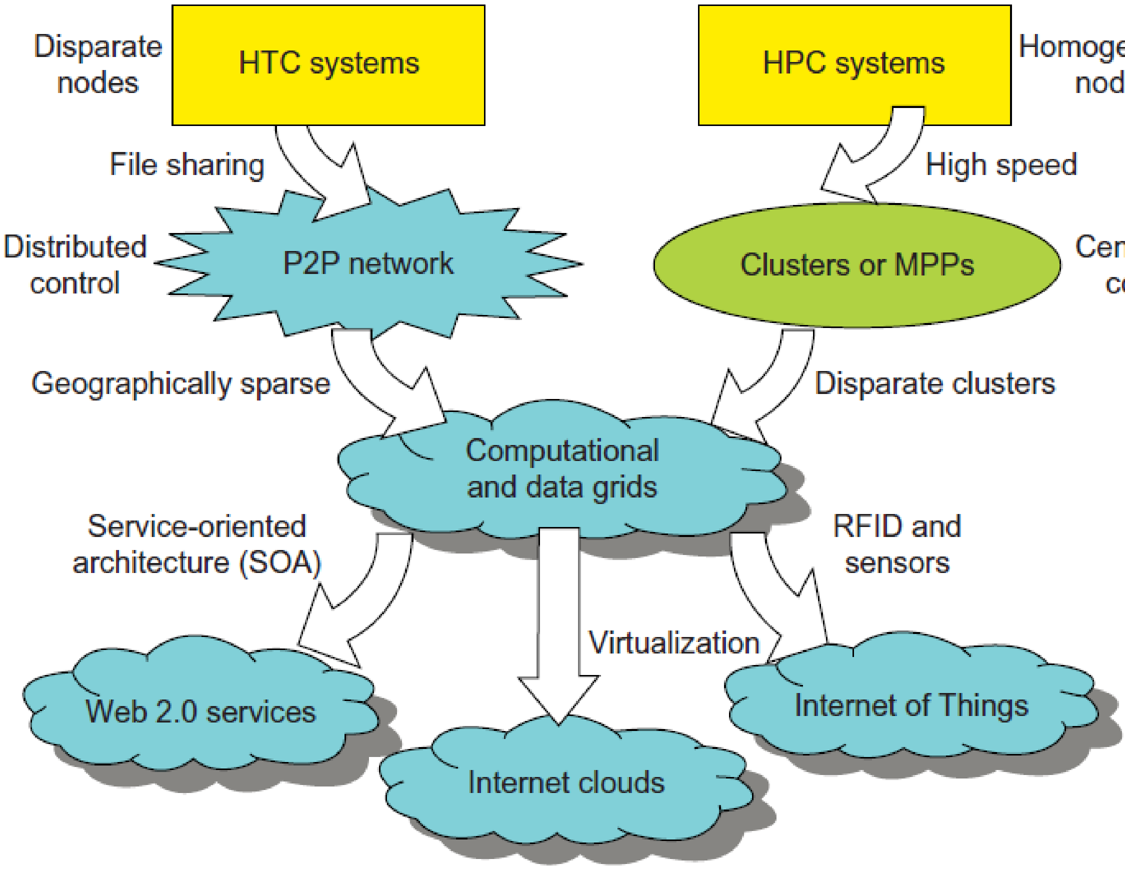
Over the past 60 years, computing technology has undergone a series of platform and environment changes. Instead of using a centralized computer to solve computational problems, a parallel and distributed computing system uses multiple computers to solve large-scale problems over the Internet. Thus, distributed computing becomes data-intensive and network-centric. These large-scale Internet applications have significantly enhanced the quality of life and information services in society today.

### ➢ The Age of Internet Computing

- Billions of people use the Internet every day. As a result, supercomputer sites and large data centers must provide high-performance computing services to huge numbers of Internet users concurrently.
- Because of this high demand, the benchmark for high-performance computing (HPC) applications is no longer optimal for measuring system performance.
- The emergence of computing clouds instead demands high-throughput computing (HTC) systems built with parallel and distributed computing technologies.

### ➢ The Platform Evolution

- Since 1990, the use of both HPC and HTC systems hidden in clusters, grids, or Internet clouds has proliferated.
- These systems are employed by both consumers and high-end web-scale computing and information services.
- The general computing trend is to leverage shared web resources and massive amounts of data over the Internet. Figure 1.1 illustrates the evolution of HPC and HTC systems.
- On the HPC side, supercomputers (massively parallel processors or MPPs) are gradually replaced by clusters of cooperative compute rs out of a desire to share computing resources.
- The cluster is often a collection of homogeneous compute nodes that are physically connected in close range to one another.
- On the HTC side, peer-to-peer (P2P) networks are formed for distributed file sharing and content delivery applications.
- A P2P system is built over many clients. Peer machines are globally distributed in nature. P2P, cloud computing, and web service platforms are more focused on HTC applications than on HPC applications.
- Clustering and P2P technologies lead to the development of computational grids or data grids.




## ➢ High-Performance Computing

For many years, HPC systems emphasize the raw speed performance. The speed of HPC systems as increased from Gflops in the early 1990s to now Pflops in 2010. This improvement was driven mainly by the demands from scientific, engineering, and manufacturing communities.

## ➢ High-Throughput Computing

HTC technology needs to not only improve in terms of batch processing speed, but also address the acute problems of cost, energy savings, security, and reliability at many data and enterprise computing centers.

## ➢ Three New Computing Paradigms

As Figure 1.1 illustrates, with the introduction of SOA, Web 2.0 services become available. Advances in virtualization make it possible to see the growth of Internet clouds as a new computing paradigm.
The maturity of radio-frequency identification (RFID), Global Positioning System (GPS), and sensor technologies has triggered the development of the Internet of Things (IoT).

## ➢ Computing Paradigm Distinctions

The high-technology community has argued for many years about the precise definitions of centralized computing, parallel computing, distributed computing, and cloud computing. In general, distributed computing is the opposite of centralized computing. The field of parallel computing overlaps with distributed computing to a great extent, and cloud computing overlaps with distributed, centralized, and parallel computing.

- **Centralized computing** This is a computing paradigm by which all computer resources are centralized in one physical system. All resources (processors, memory, and storage) are fully shared and tightly coupled within one integrated OS. Many data centers and supercomputers are centralized systems, but they are used in parallel, distributed, and cloud computing applications.

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

- **Distributed computing** This is a field of computer science/engineering that studies distributed systems. A distributed system consists of multiple autonomous computers, each having its own private memory, communicating through a computer network. Information exchange in a distributed system is accomplished through message passing. A computer program that runs in a distributed system is known as a distributed program. The process of writing distributed programs is referred to as distributed programming.

- **Cloud computing** An Internet cloud of resources can be either a centralized or a distributed computing system. The cloud applies parallel or distributed computing, or both. Clouds can be built with physical or virtualized resources over large data centers that are centralized or distributed. Some authors consider cloud computing to be a form of utility computing or service computing.

- **Distributed System Families**

Technologies for building P2P networks and networks of clusters have been consolidated into many national projects designed to establish wide area computing infrastructures, known as computational grids or data grids.

In the future, both HPC and HTC systems will demand multicore or many-core processors that can handle large numbers of computing threads per core. Both HPC and HTC systems emphasize parallelism and distributed computing. Future HPC and HTC systems must be able to satisfy this huge demand in computing power in terms of throughput, efficiency, scalability, and reliability. Meeting these goals requires to yield the following design objectives:

- **Efficiency** measures the utilization rate of resources in an execution model by exploiting massive parallelism in HPC. For HTC, efficiency is more closely related to job throughput, data access, storage, and power efficiency.

- **Dependability** measures the reliability and self-management from the chip to the system and application levels. The purpose is to provide high-throughput service with Quality of Service (QOS) assurance, even under failure conditions.

- **Adaptation in the programming model** measures the ability to support billions of job requests over massive data sets and virtualized cloud resources under various workload and service models.

- **Flexibility in application deployment** measures the ability of distributed systems to run well in both HPC (science and engineering) and HTC (business) applications.

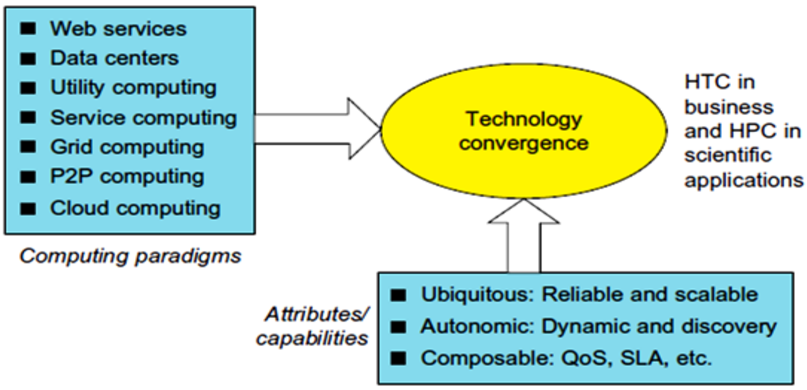
Both HPC and HTC systems desire transparency in many application aspects. For example, data access, resource allocation, process location, concurrency in execution, job replication, and failure recovery should be made transparent to both users and system management.

**Table 1.1** Applications of High-Performance and High-Throughput Systems

| Domain | Specific Applications |
|---|---|
| Science and engineering | Scientific simulations, genomic analysis, etc.<br>Earthquake prediction, global warming, weather forecasting, etc. |
| Business, education, services industry, and health care | Telecommunication, content delivery, e-commerce, etc.<br>Banking, stock exchanges, transaction processing, etc.<br>Air traffic control, electric power grids, distance education, etc.<br>Health care, hospital automation, telemedicine, etc. |
| Internet and web services, and government applications | Internet search, data centers, decision-making systems, etc.<br>Traffic monitoring, worm containment, cyber security, etc.<br>Digital government, online tax return processing, social networking, etc. |
| Mission-critical applications | Military command and control, intelligent systems, crisis management, etc. |

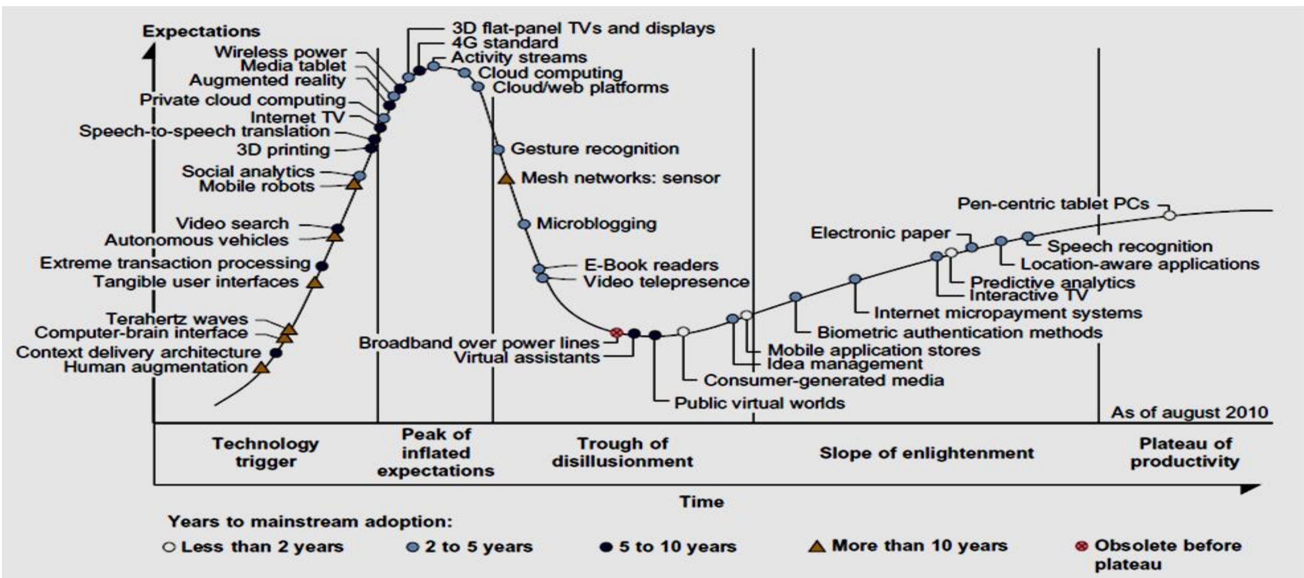
- **The Trend toward Utility Computing**

Figure identifies major computing paradigms to facilitate the study of distributed systems and their applications. These paradigms are composable with QOS and SLAs (service-level agreements). These paradigms and their attributes realize the computer utility vision. Utility computing focuses on a business model in which customers receive computing resources from a paid service provider. All grid/cloud platforms are regarded as utility service providers.




- **The Hype Cycle of New Technologies**

Any new and emerging computing and information technology may go through a hype cycle, shows the expectations for the technology at five different stages. The expectations rise sharply from the trigger period to a high peak of inflated expectations.



- **Cyber-Physical Systems**

  A cyber-physical system (CPS) is the result of interaction between computational processes and the physical world. A CPS integrates "cyber" (heterogeneous, asynchronous) with "physical" (concurrent and information-dense) objects. A CPS merges the "3C" technologies of computation, communication, and control into an intelligent closed feedback system between the physical world and the information world.

## ⇒ TECHNOLOGIES FOR NETWORK-BASED SYSTEMS

With the concept of scalable computing under our belt, it's time to explore hardware, software, and network technologies for distributed computing system design and applications.

- **Multicore CPUs and Multithreading Technologies**

Consider the growth of component and network technologies over the past 30 years. They are crucial to the development of HPC and HTC systems. The processor speed is measured in millions of instructions per second (MIPS) and network bandwidth is measured in megabits per second (Mbps) or gigabits per second (Gbps). The unit GE refers to 1 Gbps Ethernet bandwidth.

- **Multicore CPU and Many-Core GPU Architectures**

Multicore CPUs may increase from the tens of cores to hundreds or more in the future. But the CPU has reached its limit in terms of exploiting massive DLP due to the aforementioned memory wall problem.

- **Multithreading Technology**

Consider in **Figure** the dispatch of five independent threads of instructions to four pipelined data paths (functional units) in each of the following five processor categories, from left to right: a four-issue superscalar processor, a fine-grain multithreaded processor, a coarse-grain multithreaded processor, a two-core CMP, and a simultaneous multithreaded (SMT) processor. The superscalar processor is single-threaded with four functional units. Each of the three multithreaded processors is four-way multithreaded over four functional data paths. In the dual-core processor, assume two processing cores, each a single-threaded two-way superscalar processor.

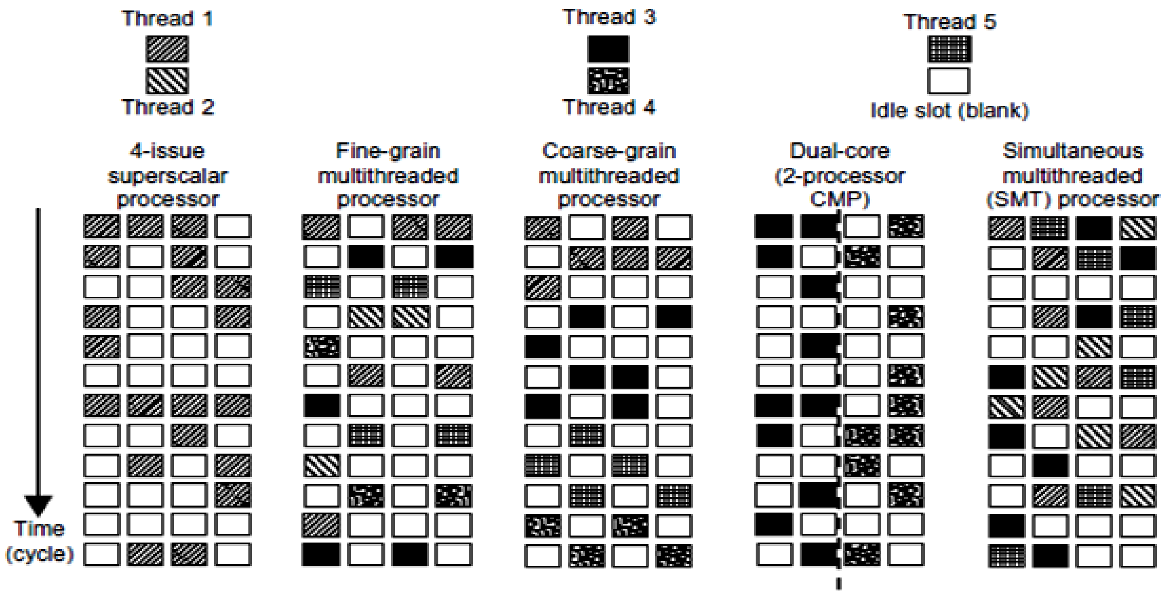
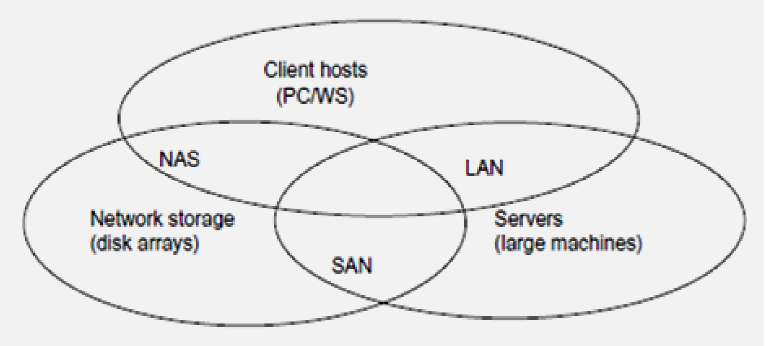



**FIGURE 1.6**

Five micro-architectures in modern CPU processors, that exploit ILP and TLP supported by multicore and multithreading technologies.

## ➢ Memory, Storage, and Wide-Area Networking

The growth of DRAM chip capacity from 16 KB in 1976 to 64 GB in 2011. For hard drives, capacity increased from 260 MB in 1981 to 250 GB in 2004. The disks or disk arrays have exceeded 3 TB in capacity. A storage area network (SAN) connects servers to network storage such as disk arrays. Network attached storage (NAS) connects client hosts directly to the disk arrays.




## ➢ Virtual Machines

The VM is built with virtual resources managed by a guest OS to run a specific application. Between the VMs and the host platform, one needs to deploy a middleware layer called a virtual machine monitor (VMM). Virtual machines (VMs) offer novel solutions to underutilized resources, application inflexibility, software manageability, and security concerns in existing physical machines.

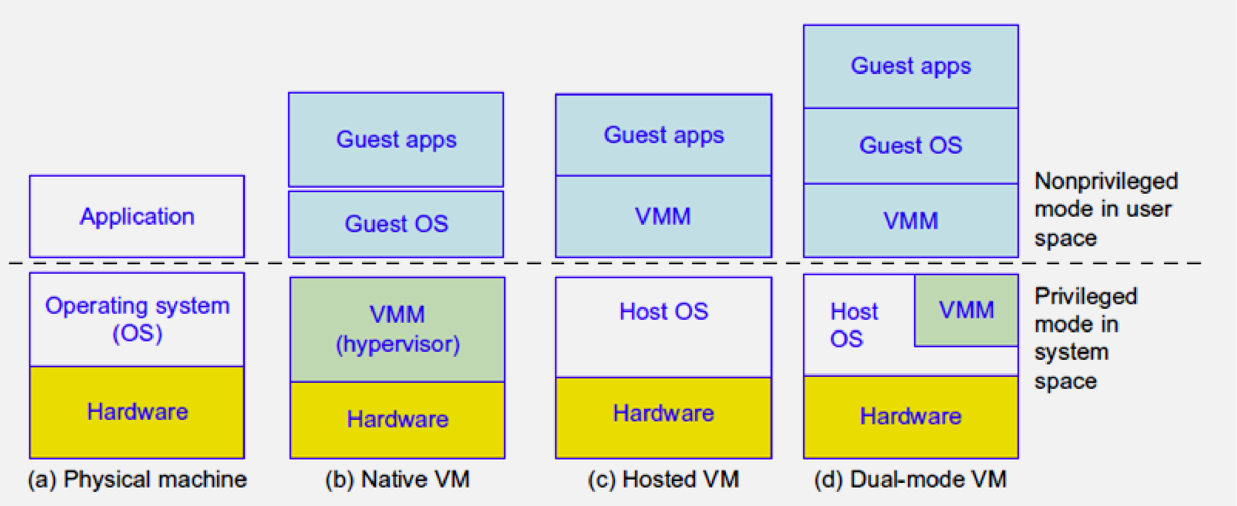



**FIGURE 1.12**

Three VM architectures in (b), (c), and (d), compared with the traditional physical machine shown in (a).

## ➢ GPU Computing to Exascale and Beyond

A GPU is a graphics coprocessor or accelerator mounted on a computer's graphics card or video card. A GPU offloads the CPU from tedious graphics tasks in video editing applications.

## ➢ GPU Programming Model

Figure shows the interaction between a CPU and GPU in performing parallel execution of floating-point operations concurrently. The CPU is the conventional multicore processor with limited parallelism to exploit. The GPU has a many-core architecture that has hundreds of simple processing cores organized as multiprocessors.

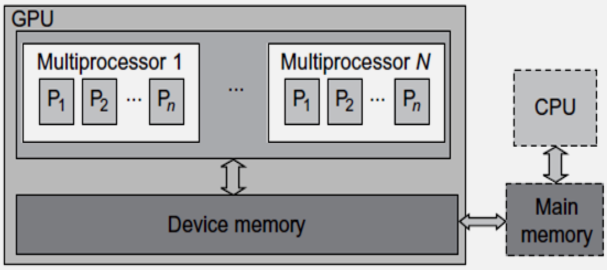



**FIGURE 1.7**

The use of a GPU along with a CPU for massively parallel execution in hundreds or thousands of processing cores.

## ➢ Data Center Virtualization for Cloud Computing

- Cloud architecture is built with commodity hardware and network devices. Almost all cloud platforms choose the popular x86 processors. Low-cost terabyte disks and Gigabit Ethernet are used to build data centers. Data center design emphasizes the performance/price ratio over speed performance alone.
- A large data center may be built with thousands of servers. Smaller data centers are typically built with hundreds of servers. The cost to build and maintain data center servers has increased over the years.
- High-end switches or routers may be too cost-prohibitive for building data centers. Thus, using high-bandwidth networks may not fit the economics of cloud computing.

## ⇒ SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- Distributed and cloud computing systems are built over a large number of autonomous computer nodes. These node machines are interconnected by SANs, LANs, or WANs in a hierarchical manner. With today's networking technology, a few LAN switches can easily connect hundreds of machines as a working cluster. A WAN can connect many local clusters to form a very large cluster of clusters. In this sense, one can build a massive system with millions of computers connected to edge networks.

**Table 1.2** Classification of Parallel and Distributed Computing Systems

| Functionality, Applications | Computer Clusters [10,28,38] | Peer-to-Peer Networks [34,46] | Data/ Computational Grids [6,18,51] | Cloud Platforms [1,9,11,12,30] |
|---|---|---|---|---|
| Architecture, Network Connectivity, and Size | Network of compute nodes interconnected by SAN, LAN, or WAN hierarchically | Flexible network of client machines logically connected by an overlay network | Heterogeneous clusters interconnected by high-speed network links over selected resource sites | Virtualized cluster of servers over data centers via SLA |
| Control and Resources Management | Homogeneous nodes with distributed control, running UNIX or Linux | Autonomous client nodes, free in and out, with self-organization | Centralized control, server-oriented with authenticated security | Dynamic resource provisioning of servers, storage, and networks |
| Applications and Network-centric Services | High-performance computing, search engines, and web services, etc. | Most appealing to business file sharing, content delivery, and social networking | Distributed supercomputing, global problem solving, and data center services | Upgraded web search, utility computing, and outsourced computing services |
| Representative Operational Systems | Google search engine, SunBlade, IBM Road Runner, Cray XT4, etc. | Gnutella, eMule, BitTorrent, Napster, KaZaA, Skype, JXTA | TeraGrid, GriPhyN, UK EGEE, D-Grid, ChinaGrid, etc. | Google App Engine, IBM Bluecloud, AWS, and Microsoft Azure |

Massive systems are considered highly scalable, and can reach web-scale connectivity, either physically or logically. In the below Table, massive systems are classified into four groups: clusters, P2P networks, computing grids, and Internet clouds over huge data centers.

- **Clusters of Cooperative Computers**

A computing cluster consists of interconnected stand-alone computers which work cooperatively as a single integrated computing resource.

- **Cluster Architecture**

Figure shows the architecture of a typical server cluster built around a low-latency, high bandwidth interconnection network. This network can be as simple as a SAN (e.g., Myrinet) or a LAN (e.g., Ethernet). To build a larger cluster with more nodes, the interconnection network can be built with multiple levels of Gigabit Ethernet, Myrinet, or InfiniBand switches. Through hierarchical construction using a SAN, LAN, or WAN, one can build scalable clusters with an increasing number of nodes. The cluster is connected to the Internet via a virtual private network (VPN) gateway. The gateway IP address locates the cluster. The system image of a computer is decided by the way the OS manages the shared cluster resources.

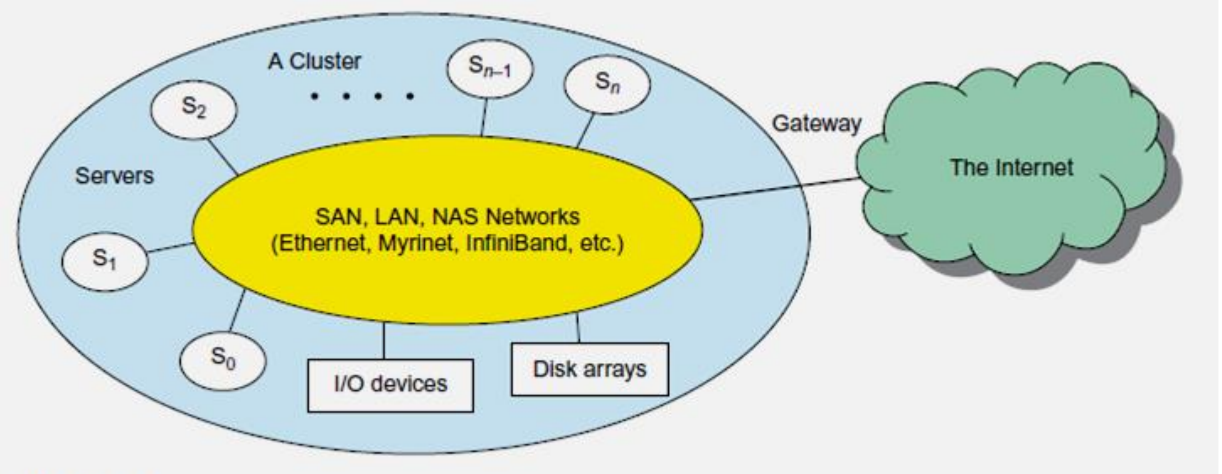



**FIGURE 1.15**

A cluster of servers interconnected by a high-bandwidth SAN or LAN with shared I/O devices and disk arrays; the cluster acts as a single computer attached to the Internet.

- **Single-System Image**

An ideal cluster should merge multiple system images into a single-system image (SSI). Cluster designers desire a cluster operating system or some middleware to support SSI at various levels, including the sharing of CPUs, memory, and I/O across all cluster nodes.

A cluster-wide OS for complete resource sharing is not available yet. Middleware or OS extensions were developed at the user space to achieve SSI at selected functional levels. Without this middleware, cluster nodes cannot work together effectively to achieve cooperative computing.

- **Grid Computing Infrastructures**

In the past 30 years, users have experienced a natural growth path from Internet to web and grid computing services. Internet services such as the Telnet command enables a local computer to connect to a remote computer. A web service such as HTTP enables remote access of remote web pages. Grid computing is envisioned to allow close interaction among applications running on distant computers simultaneously.

- **Computational Grids**

  A computing grid offers an infrastructure that couples computers, software/middleware, special instruments, and people and sensors together. The grid is often constructed across LAN, WAN, or Internet backbone networks at a regional, national, or global scale. Enterprises or organizations present grids as integrated computing resources.

- **Grid Families**

  New grid service providers (GSPs) and new grid applications have emerged rapidly, similar to the growth of Internet and web services in the past two decades. In below Table, grid systems are classified in essentially two categories: computational or data grids and P2P grids. Computing or data grids are built primarily at the national level.

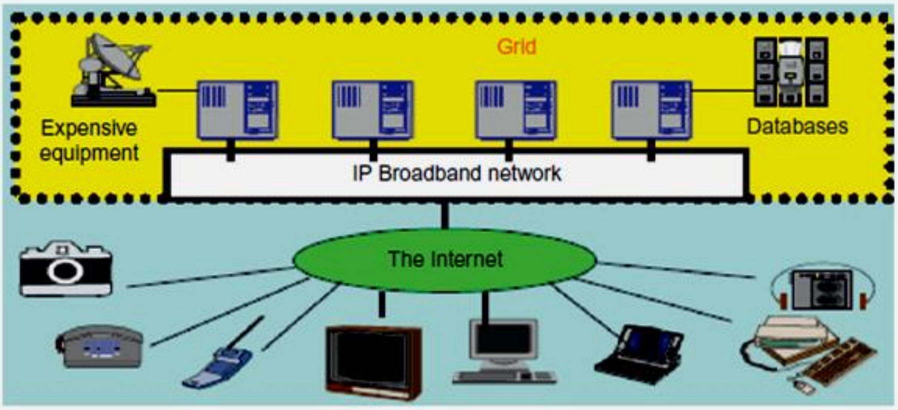



**FIGURE 1.16**

Computational grid or data grid providing computing utility, data, and information services through resource sharing and cooperation among participating organizations.

- **Peer-to-Peer Network Families**

An example of a well-established distributed system is the client-server architecture. The P2P architecture offers a distributed model of networked systems. First, a P2P network is client-oriented instead of server-oriented. In a P2P system, every node acts as both a client and a server, providing part of the system resources. Peer machines are simply client computers connected to the Internet.

Figure 1.17 shows the architecture of a P2P network at two abstraction levels. Each peer machine joins or leaves the P2P network voluntarily. Only the participating peers form the physical network at any time. Data items or files are distributed in the participating peers. Based on communication or file-sharing needs, the peer IDs form an overlay network at the logical level. This overlay is a virtual network formed by mapping each physical machine with its ID, logically, through a virtual mapping as shown in Figure 1.17.

There are two types of overlay networks: unstructured and structured. An unstructured overlay network is characterized by a random graph. There is no fixed route to send messages or files among the nodes. Structured overlay networks follow certain connectivity topology and rules for inserting and removing nodes (peer IDs) from the overlay graph.
Based on application, P2P networks are classified into four groups, as shown in Table 1.5.

**Table 1.5** Major Categories of P2P Network Families [46]

| System Features | Distributed File Sharing | Collaborative Platform | Distributed P2P Computing | P2P Platform |
|---|---|---|---|---|
| Attractive Applications | Content distribution of MP3 music, video, open software, etc. | Instant messaging, collaborative design and gaming | Scientific exploration and social networking | Open networks for public resources |
| Operational Problems | Loose security and serious online copyright violations | Lack of trust, disturbed by spam, privacy, and peer collusion | Security holes, selfish partners, and peer collusion | Lack of standards or protection protocols |
| Example Systems | Gnutella, Napster, eMule, BitTorrent, Aimster, KaZaA, etc. | ICQ, AIM, Groove, Magi, Multiplayer Games, Skype, etc. | SETI@home, Geonome@home, etc. | JXTA, .NET, FightingAid@home, etc. |

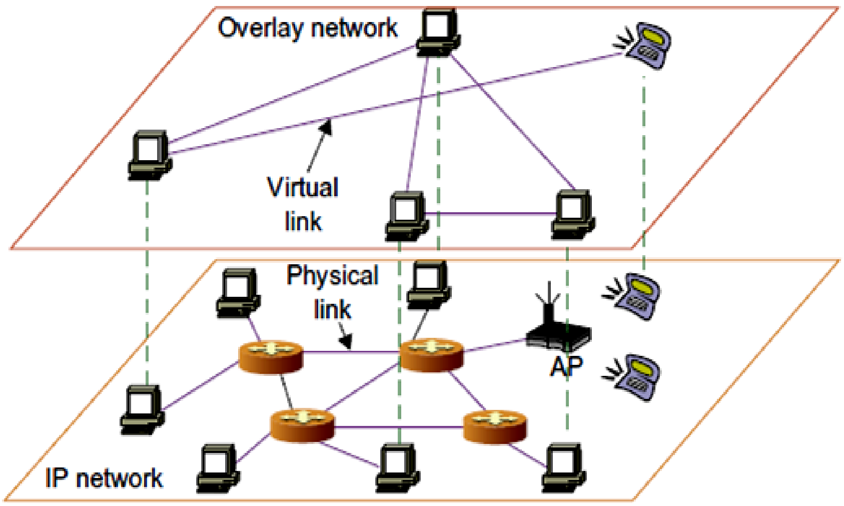



**FIGURE 1.17**

The structure of a P2P system by mapping a physical IP network to an overlay network built with virtual links.

- **Cloud Computing over the Internet**

In the future, working with large data sets will typically mean sending the computations (programs) to the data, rather than copying the data to the workstations. This reflects the trend in IT of moving computing and data from desktops to large data centers, where there is on-demand provision of software, hardware, and data as a service. This data explosion has promoted the idea of cloud computing.

“A cloud is a pool of virtualized computer resources. A cloud can host a variety of different workloads, including batch-style backend jobs and interactive and user-facing applications.”

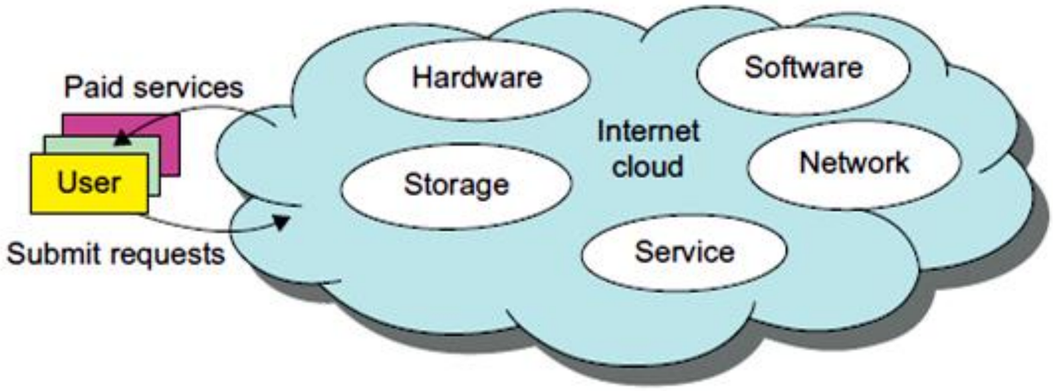



FIGURE Virtualized resources from data centers to form an Internet cloud, provisioned with hardware, software, storage, network, and services for paid users to run their applications.

**The Cloud Landscape:** Figure 1.19 depicts the cloud landscape and major cloud players, based on three cloud service models.

- **Infrastructure as a Service (IaaS)** This model puts together infrastructures demanded by users—namely servers, storage, networks, and the data center fabric. The user can deploy and run on multiple VMs running guest OSes on specific applications. The user does not manage or control the underlying cloud infrastructure, but can specify when to request and release the needed resources.

- **Platform as a Service (PaaS)** This model enables the user to deploy user-built applications onto a virtualized cloud platform. PaaS includes middleware, databases, development tools, and some runtime support such as Web 2.0 and Java. The platform includes both hardware and software integrated with specific programming interfaces.

- **Software as a Service (SaaS)** This refers to browser-initiated application software over thousands of paid cloud customers. The SaaS model applies to business processes, industry applications, consumer relationship management (CRM), enterprise resources planning (ERP), human resources (HR), and collaborative applications.

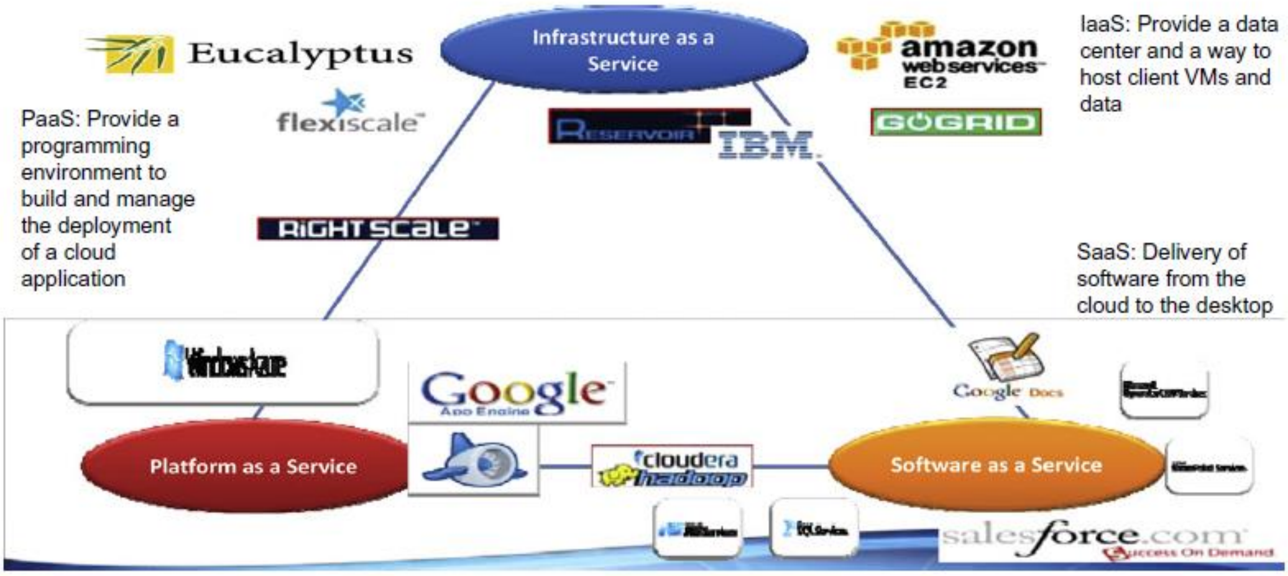



**FIGURE 1.19**

Three cloud service models in a cloud landscape of major providers.

## ⇒ NIST CLOUD COMPUTING REFERENCE ARCHITECTURE

- ✓ **The National Institute of Standards and Technology** (NIST) has been designated by the Federal Chief Information Officer (CIO) to accelerate the federal government's secure adoption of cloud computing by leading efforts to identify existing standards and guidelines.
- ✓ Consistent with NIST's mission, 1 the NIST Cloud Computing Program has developed a USG Cloud Computing Technology Roadmap, as one of many mechanisms in support of United States Government (USG) secure and effective adoption of the Cloud Computing model2 to reduce costs and improve services.
- ✓ The NIST Cloud Computing Standards Roadmap Working Group has surveyed the existing standards landscape for interoperability, performance, portability, security, and accessibility standards/models/studies/use cases/conformity assessment programs, etc., relevant to cloud computing.
- ✓ The NIST Definition of Cloud Computing identified cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Using the taxonomy developed by the NIST Cloud Computing Reference Architecture and Taxonomy Working Group, cloud computing relevant standards have been mapped to the requirements of accessibility, interoperability, performance, portability, and security.

Present areas with standardization gaps include: SaaS (Software as a Service) functional interfaces; SaaS self-service management interfaces; PaaS (Platform as a Service) functional interfaces; business support / provisioning / configuration; security; and privacy. Present standardization areas of priority to the federal government include: security auditing and compliance; identity and access management; SaaS application specific data and metadata; and resource description and discovery.

At the beginning of 2011, NIST created the following public working groups in order to provide a technically oriented strategy and standards-based guidance for the federal cloud computing implementation effort:

- Cloud Computing Reference Architecture and Taxonomy Working Group
- Cloud Computing Standards Acceleration to Jumpstart Adoption of Cloud Computing
- (SAJACC) Working Group  Cloud Computing Security Working Group
- Cloud Computing Standards Roadmap Working Group
- Cloud Computing Target Business Use Cases Working Group

## THE NIST DEFINITION OF CLOUD COMPUTING

**“**Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.**”** This cloud model is composed of **five essential characteristics, three service models, and four deployment models.**

### Essential Characteristics:

- **On-demand self-service -** A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
- **Broad network access -** Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
- **Resource pooling -** The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.
- **Rapid elasticity -** Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.
- **Measured service -** Cloud systems automatically control and optimize resource use by leveraging a metering capability10 at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, active user accounts). Resource usage can be monitored, controlled, audited, and reported, providing transparency for both the provider and consumer of the utilized service.

### Service Models:

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

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email) or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited userspecific application configuration settings.

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

The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. 12 The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

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

The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

- **<u>Deployment Models:</u>**

- **Private cloud.**

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

- **Community cloud.**

The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

- **Public cloud.**

The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

- **Hybrid cloud.**

The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

## CLOUD COMPUTING REFERENCE ARCHITECTURE

The NIST cloud computing reference architecture is a generic high-level conceptual model that is a powerful tool for discussing the requirements, structures, and operations of cloud computing. It defines a set of actors, activities, and functions that can be used in the process of developing cloud computing architectures, and relates to a companion cloud computing taxonomy.

The NIST cloud computing reference architecture defines five major actors: **cloud consumer, cloud provider, cloud auditor, cloud broker, and cloud carrier**. Each actor is an entity (a person or an organization) that participates in a transaction or process and/or performs tasks in cloud computing.

For example, a **Cloud Consumer** is an individual or organization that acquires and uses cloud products and services. The purveyor of products and services is the **Cloud Provider**. Because of the possible service offerings (Software, Platform or Infrastructure) allowed for by the cloud provider, there will be a shift in the level of responsibilities for some aspects of the scope of control, security and configuration.

The **Cloud Broker** acts as the intermediary between consumer and provider and will help consumers through the complexity of cloud service offerings and may also create value-added cloud services. The **Cloud Auditor** provides a valuable inherent function for the government by conducting the independent performance and security monitoring of cloud services. The **Cloud Carrier** is the organization which has the responsibility of transferring the data, somewhat akin to the power distributor for the electric grid.

**Figure 1** – Cloud Actors briefly lists the five major actors defined in the NIST cloud computing reference architecture.

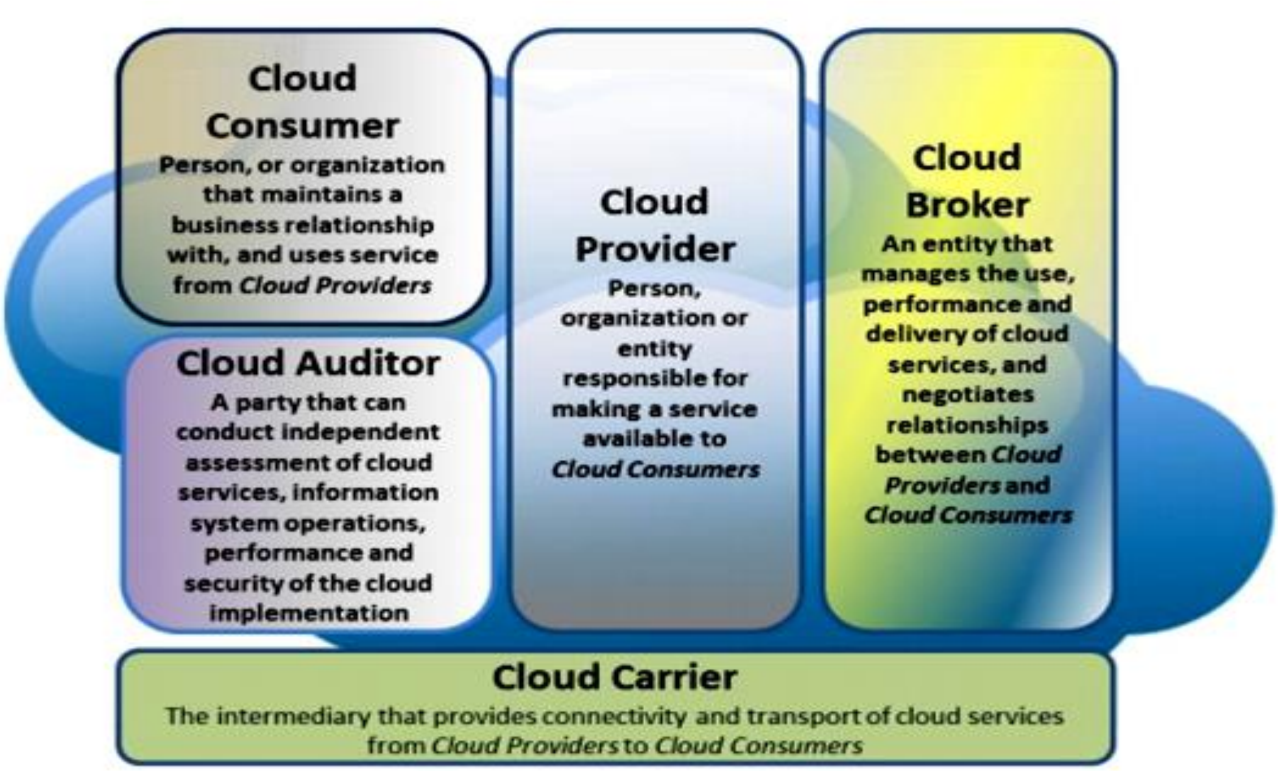
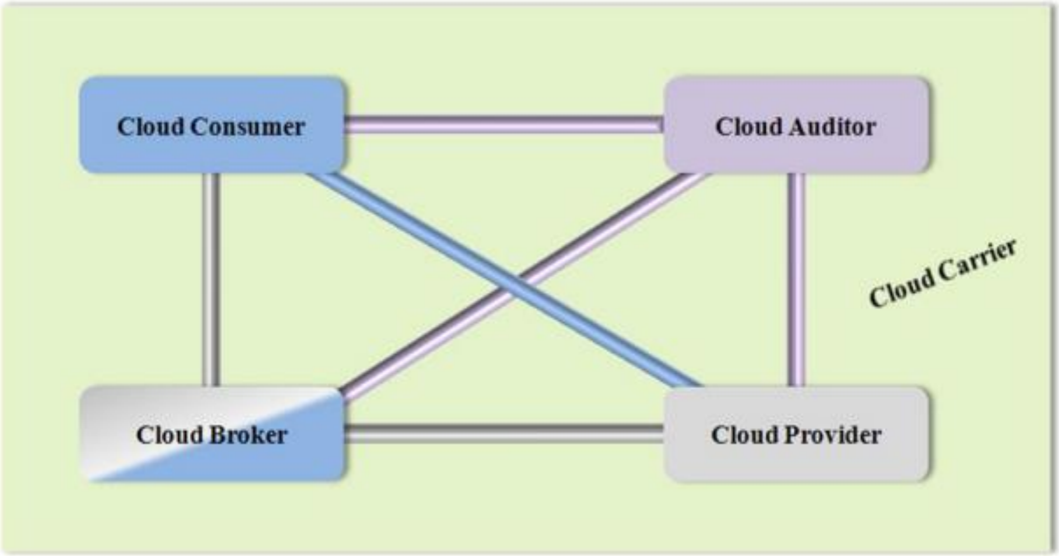



**Figure 2** – Interactions between the Actors in Cloud Computing shows the interactions among the actors in the NIST cloud computing reference architecture. A cloud consumer may request cloud services from a cloud provider directly or via a cloud broker. A cloud auditor conducts independent

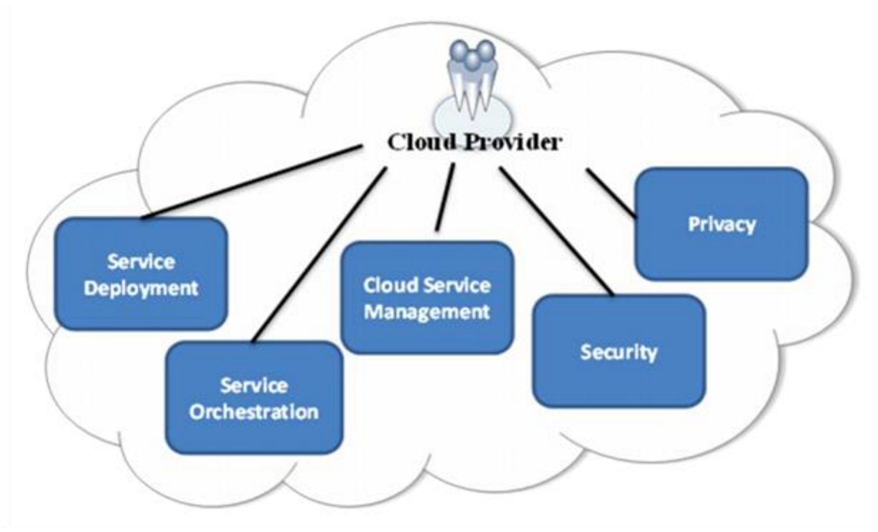
audits and may contact the others to collect necessary information. The details will be discussed in the following sections and be presented as successive diagrams in increasing levels of detail.




**1) <u>CLOUD CONSUMER</u> -** The cloud consumer is the ultimate stakeholder that the cloud computing service is created to support. A cloud consumer represents a person or organization that maintains a business relationship with, and uses the service from, a cloud provider.

| Service Models | Consumer Activities | Provider Activities |
|---|---|---|
| SaaS | Uses application/service for business process operations. | Installs, manages, maintains, and supports the software application on a cloud infrastructure. |
| PaaS | Develops, tests, deploys, and manages applications hosted in a cloud system. | Provisions and manages cloud infrastructure and middleware for the platform consumers; provides development, deployment, and administration tools to platform consumers. |
| IaaS | Creates/installs, manages, and monitors services for IT infrastructure operations. | Provisions and manages the physical processing, storage, networking, and the hosting environment and cloud infrastructure for IaaS consumers. |

**2) <u>CLOUD PROVIDER</u> -** A cloud provider can be a person, an organization, or an entity responsible for making a service available to cloud consumers. A cloud provider builds the requested software/platform/ infrastructure services, manages the technical infrastructure required for providing the services, provisions the services at agreed-upon service levels, and protects the security and privacy of the services.




**For SaaS,** the cloud provider deploys, configures, maintains, and updates the operation of the software applications on a cloud infrastructure so that the services are provisioned at the expected service levels to cloud consumers.

**For PaaS,** the cloud provider manages the cloud infrastructure for the platform, and provisions tools and execution resources for the platform consumers to develop, test, deploy, and administer applications.

**For IaaS,** the cloud provider provisions the physical processing, storage, networking, and other fundamental computing resources, as well as manages the hosting environment and cloud infrastructure for IaaS consumers.

The activities of cloud providers can be discussed in greater detail from the perspectives of Service Deployment, Service Orchestration, Cloud Service Management, Security and Privacy.

## SERVICE DEPLOYMENT

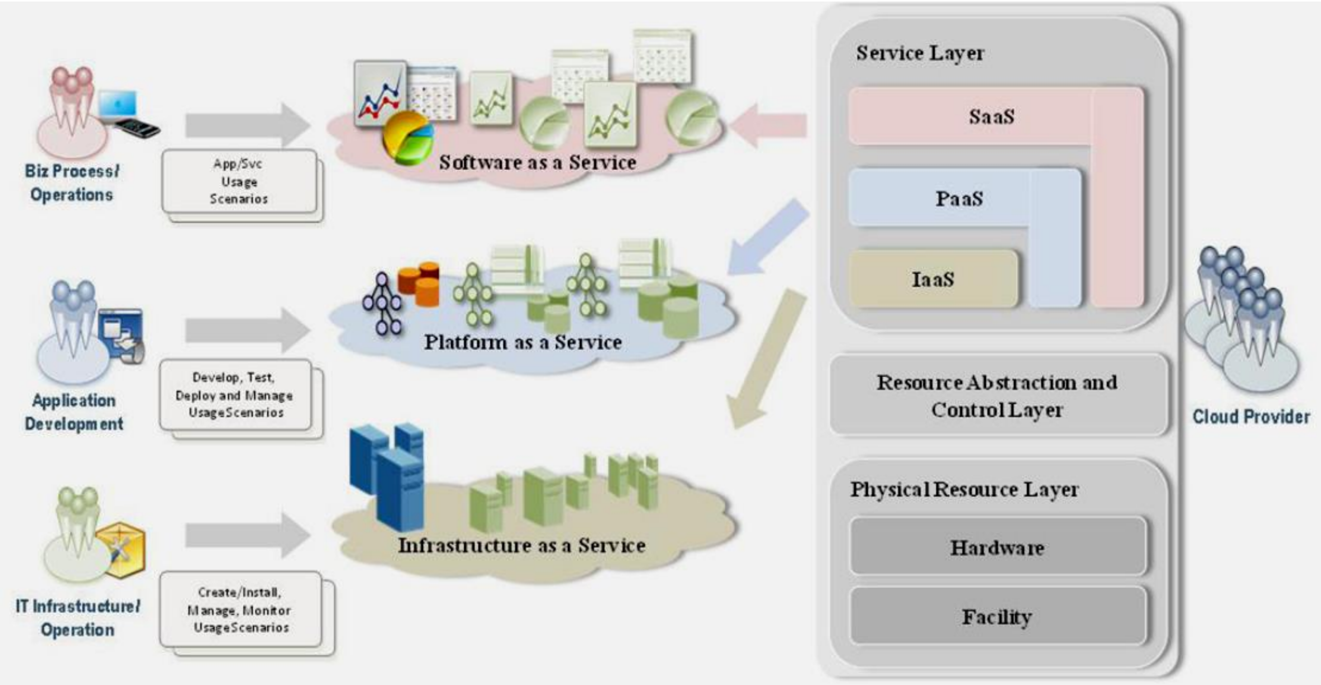
As identified in the NIST cloud computing definition, a cloud infrastructure may be operated in one of the following deployment models: public cloud, private cloud, community cloud, or hybrid cloud.

- A **public cloud** is one in which the cloud infrastructure and computing resources are made available to the general public over a public network. A public cloud is owned by an organization selling cloud services and serves a diverse pool of clients.
- For **private clouds**, the cloud infrastructure is operated exclusively for a single organization. A private cloud gives the organization exclusive access to and usage of the infrastructure and computational resources. It may be managed either by the organization or by a third party, and may be implemented at the organization's premise (i.e., on-site private clouds) or outsourced to a hosting company (i.e., outsourced private clouds).
- Similar to private clouds, a **community cloud** may be managed by the organizations or by a third party, and may be implemented at the customer's location (i.e., on-site community cloud) or outsourced to a hosting company (i.e., outsourced community cloud). However, a community cloud serves a set of organizations that have common security, privacy, and compliance considerations, rather than serving a single organization as does a private cloud.

- A hybrid cloud is a composition of two or more cloud deployment models (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability. As discussed in this section, both private clouds and community clouds can be either implemented on-site or outsourced to a third party. Therefore, each constituent cloud of a hybrid cloud can be one of the five variants.

### SERVICE ORCHESTRATION

Service orchestration refers to the arrangement, coordination, and management of cloud infrastructure to provide the optimizing capabilities of cloud services, as a cost-effective way of managing IT resources, as dictated by strategic business requirements. Figure 5 shows the general requirements and processes for cloud providers to build each of the three service models.

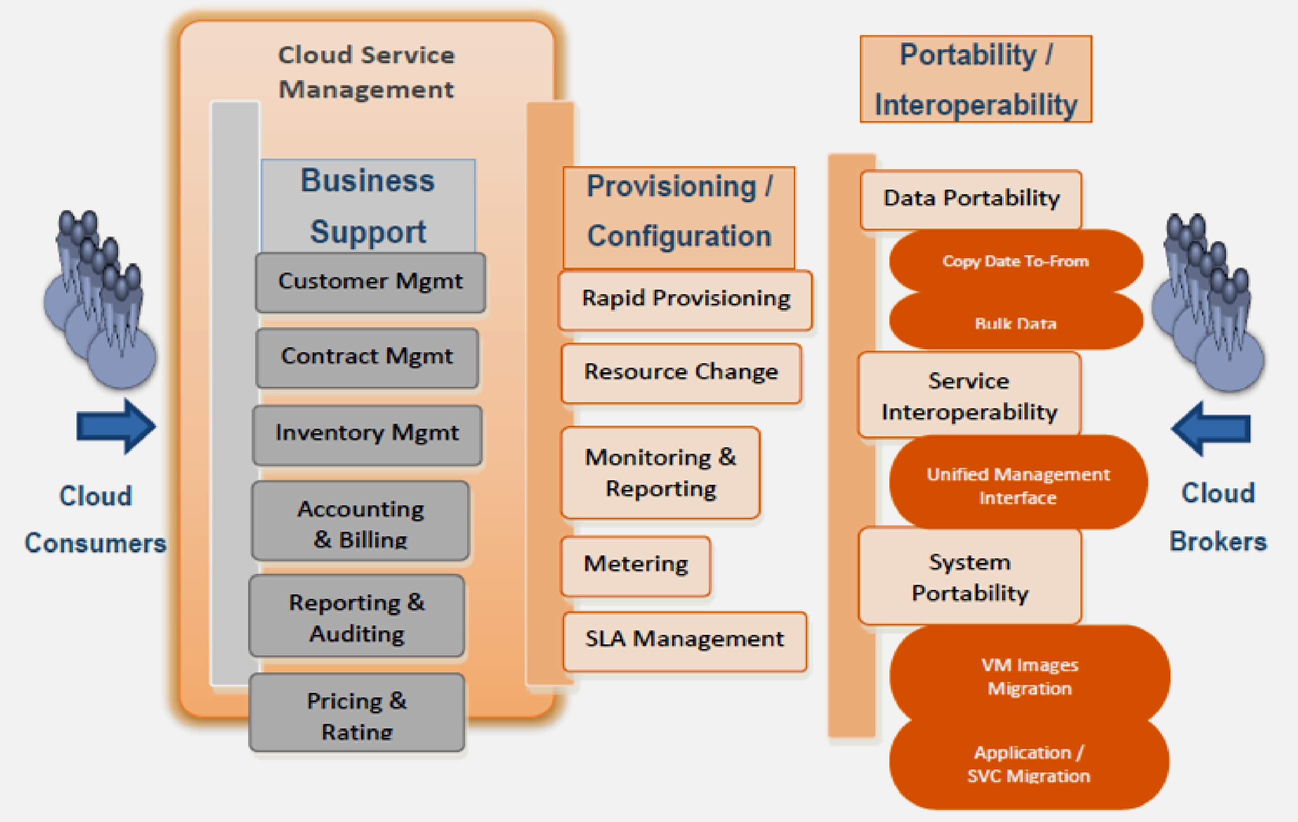



- A three-layered framework is identified for a generalized cloud system in Figure 5. The top layer is the service layer, where a cloud provider defines and provisions each of the three service models.
- The middle layer is the resource abstraction and control layer. This layer contains the system components that a cloud provider uses to provide and manage access to the physical computing resources through software abstraction.
- The lowest layer in the framework is the physical resource layer, which includes all the physical computing resources. This layer includes hardware resources, such as computers (CPU and memory), networks (routers, firewalls, switches, network links, and interfaces), storage components (hard disks), and other physical computing infrastructure elements.

### CLOUD SERVICE MANAGEMENT

*Cloud Service Management* includes all of the service-related functions that are necessary for the management and operation of those services required by or proposed to cloud consumers. As illustrated in Figure 6, cloud service management can be described from the perspective of *business support, provisioning and configuration,* and from the perspective of *portability and interoperability* requirements.

**Figure 6 – Cloud Provider: Cloud Service Management**




## SECURITY

**"**As the Federal Government moves to the cloud, it must be vigilant to ensure the security and proper management of government information to protect the privacy of citizens and national security**".**

Security is a cross-cutting function that spans all layers of the reference architecture involving end-to-end security that ranges from physical security to application security, and in general, the responsibility is shared between cloud provider and federal cloud consumer.

Cloud Providers should ensure that the facility hosting cloud services is secure and that the staff has proper background checks. When data or applications are moved to a cloud, Cloud Consumers ensure that the cloud offering satisfies the security requirements and enforces the compliance rules.

## PRIVACY

Cloud providers should protect the assured, proper, and consistent collection, processing, communication, use, and disposition of personal information (PI) and personally identifiable information (PII) in the cloud system.

**3) CLOUD AUDITOR -** A cloud auditor is a party that can conduct independent assessment of cloud services, information system operations, performance, and the security of a cloud computing implementation. A cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance, and adherence to service level agreement parameters.

**4) CLOUD BROKER -** The NIST Reference Architecture defines a Cloud Broker as an entity that manages the use, performance, and delivery of cloud services, and negotiates relationships between

Cloud Providers and Cloud Consumers. As cloud computing evolves, the integration of cloud services may become too complex for cloud Consumers to manage. In such cases, a Cloud Consumer may request cloud services from a Cloud Broker instead of directly contacting a Cloud Provider. Cloud Brokers provide a single point of entry for managing multiple cloud services. In general, Cloud Brokers provide services in three categories:

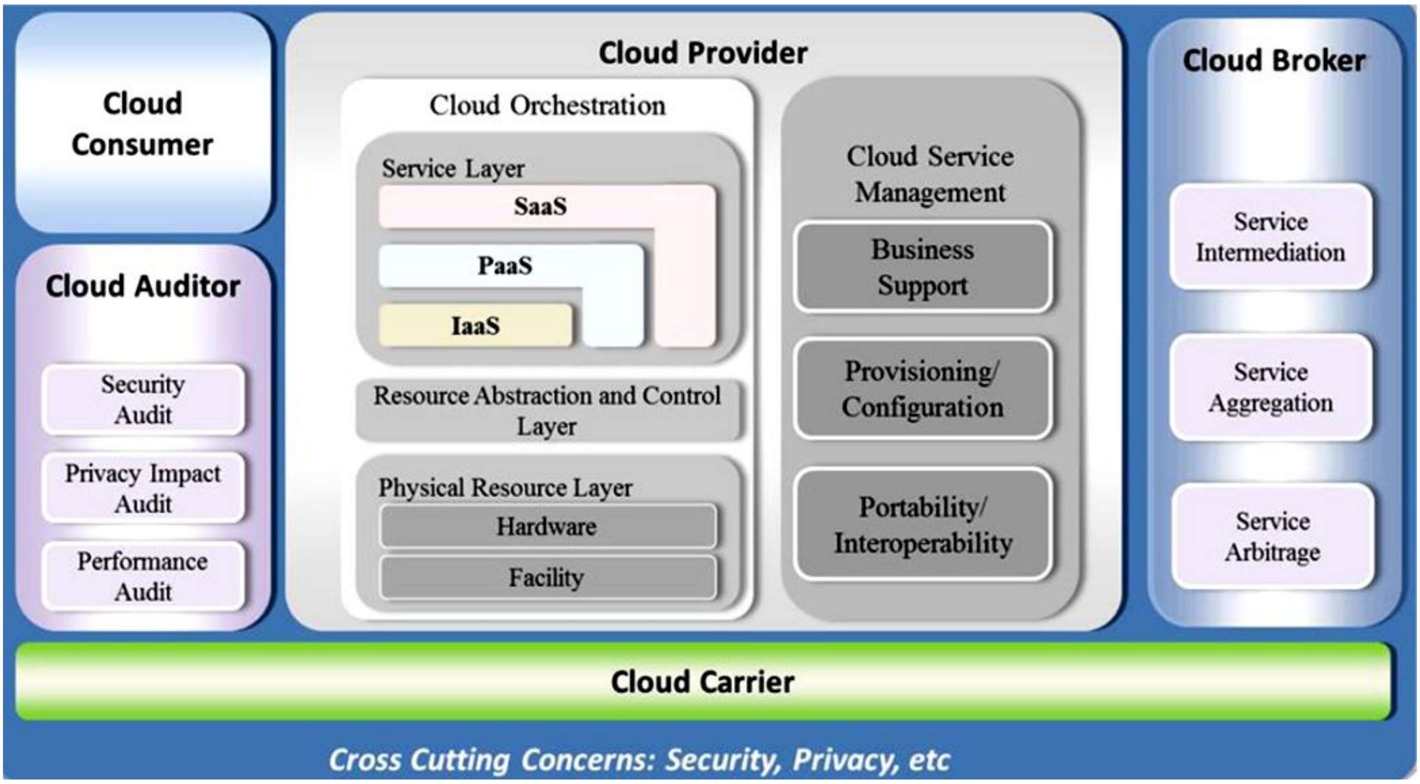
**Intermediation:** A Cloud Broker enhances a given service by improving some specific capability and providing value-added services to cloud Consumers. The improvement can be managing access to cloud services, identity management, performance reporting, enhanced security, etc.

**Aggregation:** A Cloud Broker combines and integrates multiple services into one or more new services. The Broker provides data and service integration and ensures the secure data movement between the cloud Consumer and multiple cloud Providers.

**Arbitrage:** Service arbitrage is similar to service aggregation except that the services being combined/consolidated are not fixed. Service arbitrage means a Broker has the flexibility to choose services from multiple service Providers.

**5) CLOUD CARRIER -** A cloud carrier acts as an intermediary that provides connectivity and transport of cloud services between cloud consumers and cloud providers. Cloud carriers provide access to consumers through network, telecommunication, and other access devices. For example, cloud consumers can obtain cloud services through network access devices, such as computers, laptops, mobile phones, mobile Internet devices (MIDs), etc.

The distribution of cloud services is normally provided by network and telecommunication carriers or a transport agent, where a transport agent refers to a business organization that provides physical transport of storage media such as high-capacity hard drives. Note that a cloud provider will set up service level agreements (SLAs) with a cloud carrier to provide services consistent with the level of SLAs offered to cloud consumers, and may require the cloud carrier to provide dedicated and encrypted connections between cloud consumers and cloud providers.




.

**Figure – The Combined Conceptual Reference Diagram**

## ⇒ CLOUD MODELS

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

### ➤ The 5 Essential Characteristics of Cloud Computing:

**1) *On-demand self-service.***

A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

**2) *Broad network access.***

Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

**3) *Resource pooling.***

The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.

4) ***Rapid elasticity.***

Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

5) ***Measured service.***

Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

## Service Models:

### 1) Software as a service (SaaS)

A service provider delivers software and applications through the internet. Users subscribe to the software and access it via the web or vendor APIs.
Software as a service (SaaS) is a cloud computing offering that provides users with access to a vendor's cloud-based software. Users do not install applications on their local devices. Instead, the applications reside on a remote cloud network accessed through the web or an API. Through the application, users can store and analyze data and collaborate on projects.

- SaaS vendors provide users with software and applications via a subscription model.
- Users do not have to manage, install or upgrade software; SaaS providers manage this.
- Data is secure in the cloud; equipment failure does not result in loss of data.
- Use of resources can be scaled depending on service needs.
- Applications are accessible from almost any internet-connected device, from virtually anywhere in the world.

### 2) Platform as a service (PaaS)

A service provider offers access to a cloud-based environment in which users can build and deliver applications. The provider supplies underlying infrastructure.

Platform as a service (PaaS) is a cloud computing offering that provides users with a cloud environment in which they can develop, manage and deliver applications. In addition to storage and other computing resources, users are able to use a suite of prebuilt tools to develop, customize and test their own applications.

**Key features**

- PaaS provides a platform with tools to test, develop and host applications in the same environment.
- Enables organizations to focus on development without having to worry about underlying

infrastructure.

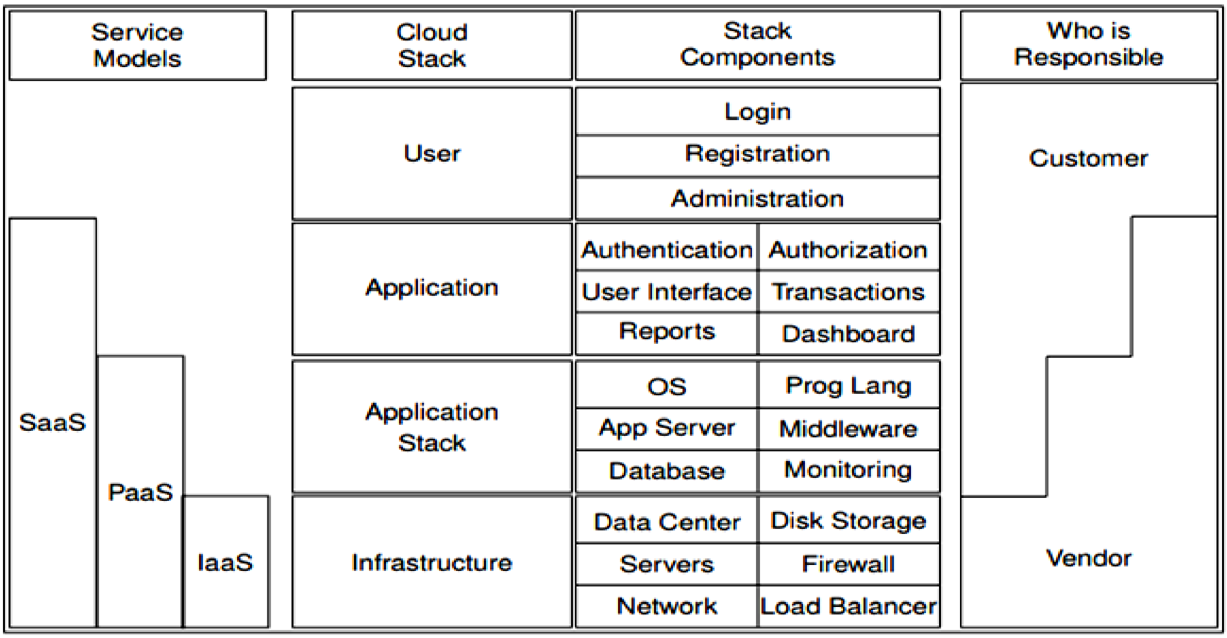
- Providers manage security, operating systems, server software and backups.
- Facilitates collaborative work even if teams work remotely.

### 3) <u>Infrastructure as a service (IaaS)</u>

A vendor provides clients pay-as-you-go access to storage, networking, servers and other computing resources in the cloud.
Infrastructure as a service (IaaS) is a cloud computing offering in which a vendor provides users access to computing resources such as servers, storage and networking. Organizations use their own platforms and applications within a service provider's infrastructure.

**Key features:**

- Instead of purchasing hardware outright, users pay for IaaS on demand.
- Infrastructure is scalable depending on processing and storage needs.
- Saves enterprises the costs of buying and maintaining their own hardware.
- Because data is on the cloud, there can be no single point of failure.
- Enables the virtualization of administrative tasks, freeing up time for other work.




## ⇒ CLOUD MODELS: Public Vs Private Cloud

### ❖ Public Cloud

The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

– Available to general public

– Owned and provisioned by an organization selling cloud services

– NIST definition

∗ Cloud infrastructure provisioned for open use by general public

∗ May be owned, managed, and operated by some combination of a business, academic, or government organization

– Multitenant environment

– Resources used from a shared grid of commodity resources

– Users unaware of physical location of resources or data center

∗ Users access resources through an abstraction layer on top of the physical hardware

∗ Virtual compute resources created by APIs in the abstraction

– Advantages of public clouds

∗ Utility pricing

- Pay for the resources consumed
- Scale up or scale down as per need
- No procurement of physical hardware, except for the hardware to connect to the cloud
- No wasted compute cycles

∗ Elasticity

- Endless pool of resources
- Configure software solutions to dynamically increase/decrease resources to handle peak loads
- React to traffic spikes in real-time

∗ Core competency

- Outsourced data center and infrastructure management
- More time on core competence

– Risks of public clouds

∗ Control

- Reliance on vendor for performance and uptime
- Outage at cloud vendor could adversely affect services

∗ Regulatory issues

- PCI DSS – Payment Card Industry Data Security Standard
- HIPAA – Health Information Portability and Accountability Act
- Data privacy issues
- May be solved by leveraging certified SaaS solutions for components that are hadr to audit in public cloud

∗ Limited configurations

- May not be able to access specific hardware to solve intensive computational problems

## ❖ **Private Cloud**

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

– Hosted within an organization's firewall, managed by the organization

– Created and controlled by the enterprise

– NIST definition

∗ Cloud infrastructure provisioned for exclusive use by a single organization with multiple consumers or business units

∗ May be owned, managed, and operated by the organization, a third party, or some combination

∗ May exist on or off premises

– Deploy in a single-tenant environment and not comingled with other customers

– Costs more than sharing in public cloud environment, but more control and security

– Reduce regulatory risks

## ❖ **Hybrid Cloud**

The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

– Combination of public and private clouds

– Management responsibilities divided between public cloud provider and the business renting it

– NIST definition

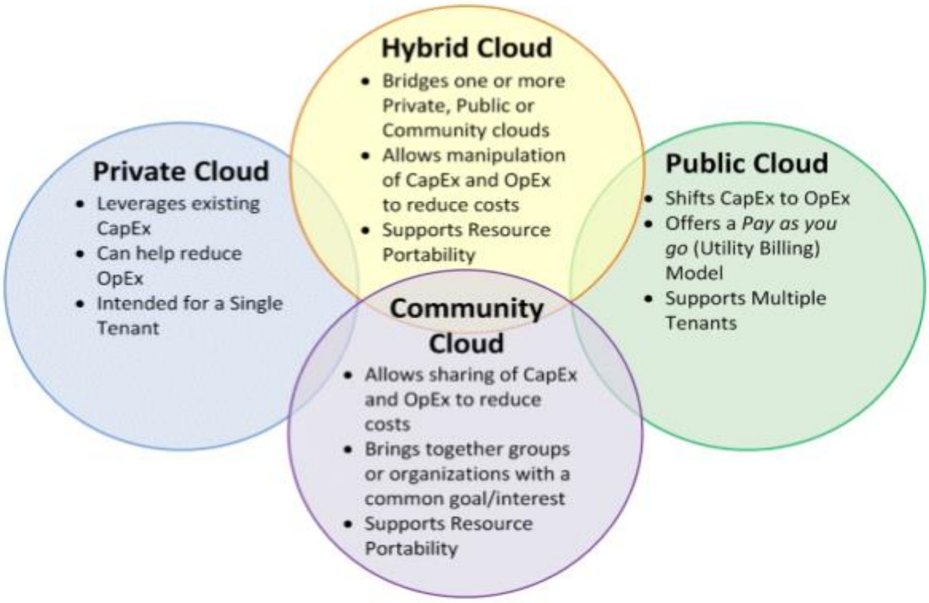
∗ Composition of two or more distinct cloud infrastructures (private, community, or public)

∗ The clouds remain unique entities but bound together by standardized or proprietary technologies to enable data/application portability

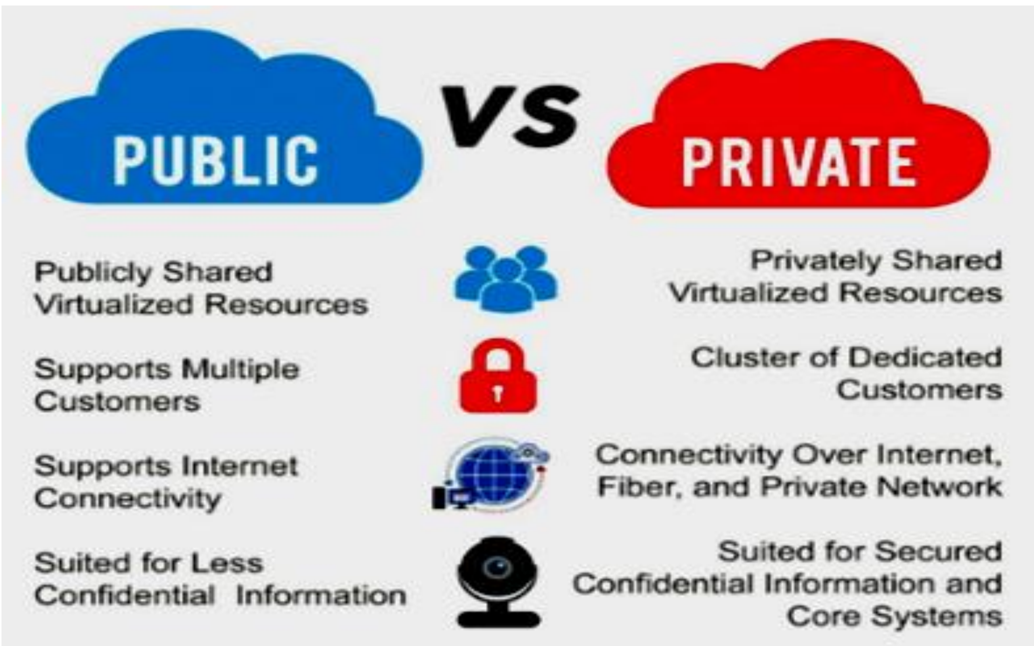
∗ Cloud bursting for load balancing between clouds

❖ **Community Cloud.**

The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the

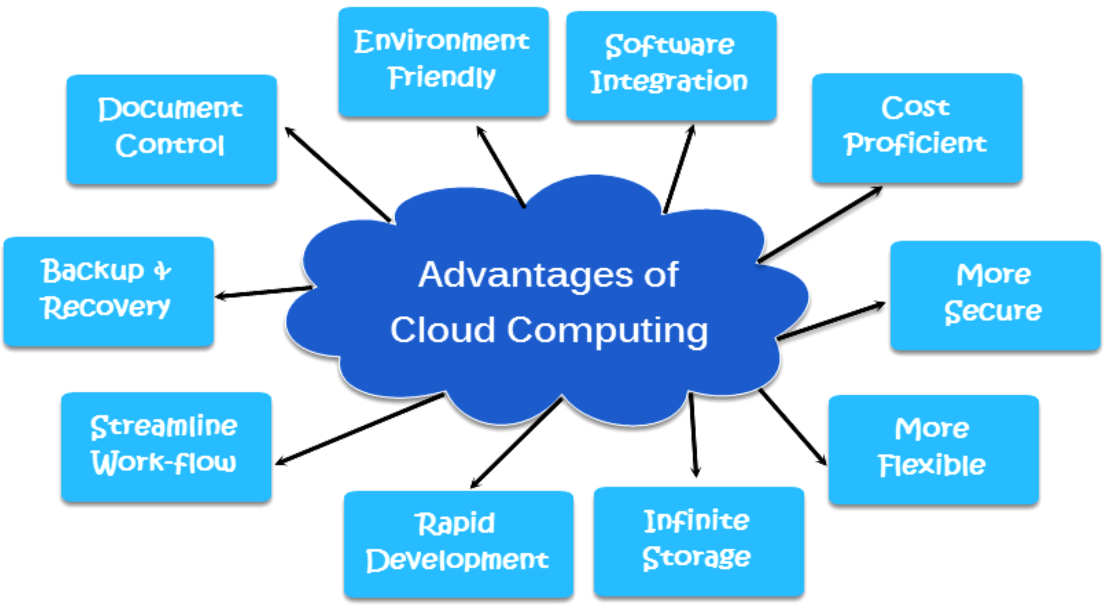



community, a third party, or some combination of them, and it may exist on or off premises.



## ⇒ CLOUD SOLUTIONS

- A **cloud based solution** refers to on-demand services, computer networks, storage, applications or resources accessed via the internet and through another provider's shared **cloud** computing infrastructure.
- The benefits of cloud based solutions to end users and businesses include increased capacity, scalability, functionality, and reduced maintenance and cost for computer infrastructure or in-house staff. Additionally, cloud-based solutions can enable companies to focus on revenue driving initiatives rather than time-consuming, non-core business tasks.
- The ability to access cloud-based solutions from anywhere with an internet connection paired with the widespread adoption of smart phones and faster mobile networks have given users the ability to access cloud-based solutions from anywhere and anytime.

- Sparkle Cloud Solutions allow customers to enjoy computing resources on a pay as you grow model. They can access data, applications and services from any location.

## ⇒ CLOUD ECOSYSTEM

✓ A **cloud ecosystem** is a complex system of interdependent components that all work together to enable **cloud** services. ... In **cloud** computing, the **ecosystem** consists of hardware and software as well as **cloud** customers, **cloud** engineers, consultants, integrators and partners.

### ➢ The concept of a cloud ecosystem

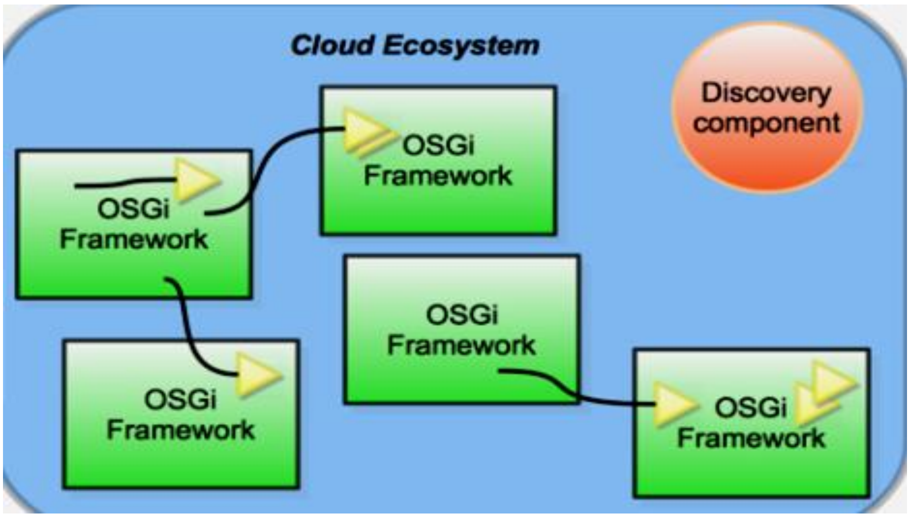
Cloud is motivating business and IT to emerge from their respective silos and forge partnerships that apply new technologies in innovative ways. But converting cloud's considerable benefits into business opportunities requires an astute understanding of your cloud ecosystem, including:

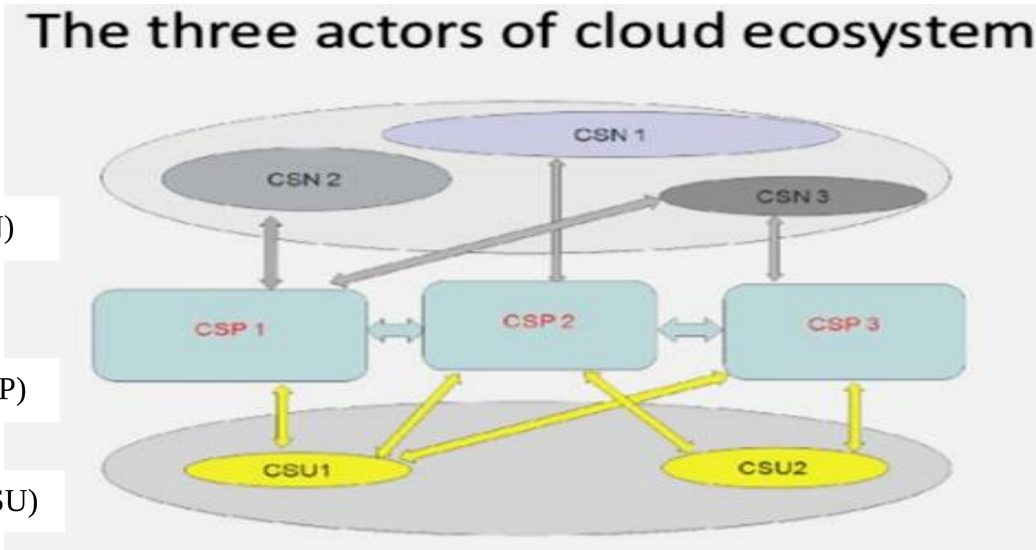
- Determining the type of cloud environment (public, private or hybrid) best suited for your organization.
- Developing your cloud adoption vision, including governance strategy, business outcomes, and project benefits.
- Establishing use cases and a detailed plan.
- Understanding the implications of adopting specific Cloud Service Layers, including business process as a service (BPaaS), software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).

### ➢ The benefits of a Cloud Ecosystem

- Companies can use a cloud ecosystem to build new business models. It becomes relatively easy for a medical device manufacturer, for example, to launch a heart-monitoring service on its cloud service provider's cloud infrastructure and then sell the service alongside its main business of manufacturing heart monitors for hospitals.
- In a cloud ecosystem, it is also easier to aggregate data and analyze how each part of the system affects the other parts. For example, if an ecosystem consists of patient records, smart device logs and healthcare provider records, it becomes possible to analyze patterns across an entire patient population.
- Data Migration provides an automated mechanism to transfer data from one **cloud** environment to another or to other computing systems. Data Synchronization is the mechanism to ensure consistency across the **Cloud Ecosystem** to a single set of source data for all duplicated target data storage and vice versa.


- ✓ The center of a cloud ecosystem is a public cloud provider. It might be an IaaS provider such as Amazon Web Services (AWS) or a SaaS vendor such as Sales force. ... For example, AWS is the center of its own ecosystem, but it's also a part of the Sales force ecosystem.




Cloud Service Partner (CSN)

Cloud Service Provider (CSP)

Cloud service User (CSU)

## ⇒ <u>**CLOUD SERVICE MANAGEMENT**</u>

- Cloud service management and operations refer to all the activities that an organization does to plan, design, deliver, operate, and control the IT and cloud services that it offers to customers.

- Service management includes the operational aspects of your applications and services. After an application is pushed to production, it must be managed. Applications are monitored to ensure availability and performance according to service level agreements (SLAs) or service level objectives (SLOs).

- The delivery of dynamic, cloud-based infrastructure, platform and application services doesn't occur in a vacuum. In addition to best practices for effective administration of all the elements associated with cloud service delivery, cloud service management and cloud monitoring tools enable providers to keep up with the continually shifting capacity demands of a highly elastic environment.

- Cloud monitoring and cloud service management tools allow cloud providers to ensure optimal performance, continuity and efficiency in virtualized, on-demand environments. These tools -- software that manages and monitors networks, systems and applications -- enable cloud providers not just to guarantee performance, but also to better orchestrate and automate provisioning of resources.

- Cloud monitoring tools, specifically, enable cloud providers to track the performance, continuity and security of all of the components that support service delivery: the hardware, software and services in the data center and throughout the network infrastructure.

## Aspects of service management

- Cloud service management and operations is redefining traditional service management to better fit the needs for cloud and DevOps patterns. At the same time, it bridges traditional approaches to service management, such as the IT Infrastructure Library (ITIL).
- The service management reference architecture includes incident management, problem management, change management, and operations.

### Incident management

Incident management aims to restore the service as quickly as possible by using a first-responder team that is equipped with automation and well-defined runbooks. To maintain the best possible levels of service quality and availability, the incident management team performs sophisticated monitoring to detect issues early, before the service is affected.

### Problem management

Problem management aims to resolve the root causes of incidents to minimize their adverse impact and prevent recurrence. Capabilities include root-cause analysis, incident analysis, dashboards, and collaboration.

- **Change management**

The purpose of change management is to achieve the successful introduction of changes to an IT system or environment. Success is measured as a balance of the timeliness and completeness of change implementation, the cost of implementation, and the minimization of disruption that is caused in the target system or environment.

- **Operations**

Operations describes the activities to deliver the right quantity of the right set of services at competitive costs for customers. IT operations management runs daily routine tasks that are related to the operation of infrastructure components and applications. These tasks include application and systems availability, health checks, compliance checks, performance monitoring, backups, and capacity monitoring and management.

➢ **What are the key processes associated with cloud service management?**

✓ Cloud service management shares some basic principles with traditional IT service management (ITSM). Cloud management tools help providers administrate the systems and applications that facilitate the on-demand service delivery model. The goal of these practices is to improve the efficiency of the cloud environment and achieve a high level of customer satisfaction.

✓ Essentially, cloud service management takes the customer perspective as the measure of service assurance and manages all the individual IT resources in a way that will support that. This involves adjusting the operations and policies, as necessary, of all the assets in the virtual environment that support and affect the on-demand service delivery model. Such assets include servers, software and services that provide access and connectivity to these cloud services.

✓ The core elements of cloud service management mirror those of traditional ITSM -- including cloud service-level agreement (SLA) management, cloud capacity management, availability management and billing -- and are applied to administrate a cloud delivery environment in a systemic way. These processes are supported with tools that track provisioning and change management, configuration management, release management, incident management, performance management and service continuity.

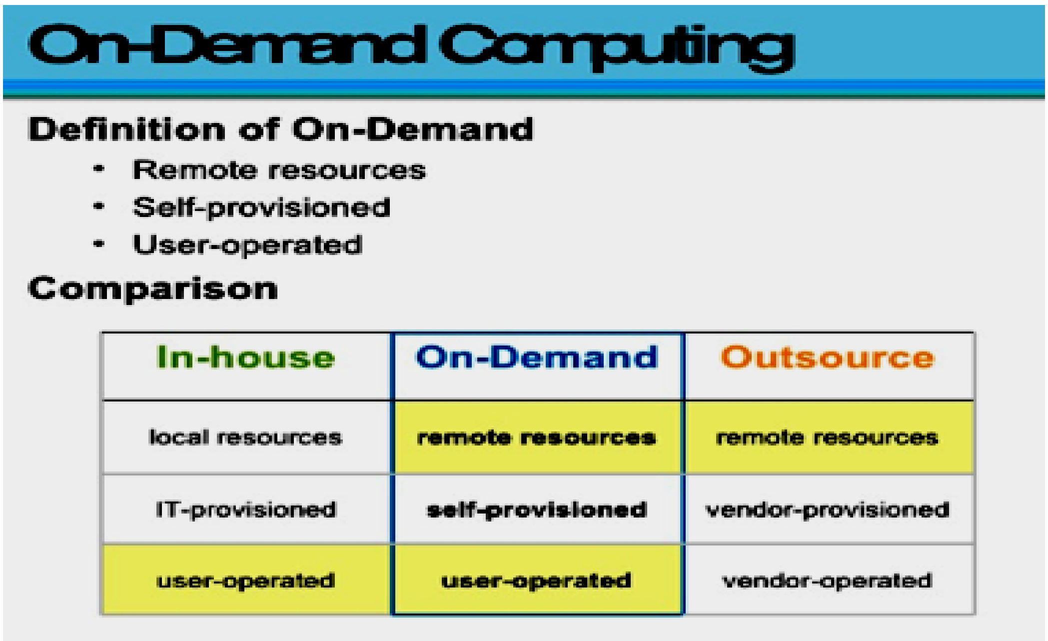
**Cloud service management platforms: Build or buy?**

Although vendors have developed many cloud service management and monitoring tools for enterprises that build and manage their own private clouds, there are far fewer tools that meet the scale, security and performance requirements of cloud providers. Beyond that, there are even fewer solutions that provide the comprehensive capabilities associated with the entire ITSM process for cloud providers, namely orchestration.


## ⇒ COMPUTING ON DEMAND

- On-**demand computing** is a business **computing** model in which **computing** resources are made available to the user on an "as needed" basis. Rather than all at once, on-**demand computing** allows **cloud** hosting companies to provide their clients with access to **computing** resources as they become necessary.

- On-**demand computing** is a delivery model in which **computing** resources are made available to the user as needed. The resources may be maintained within the user's enterprise, or made available by a cloud service provider.
- On-demand computing (ODC) is an enterprise-level model of technology and computing in which resources are provided on an as-needed and when-needed basis. ODC make computing resources such as storage capacity, computational speed and software applications available to users as and when needed for specific temporary projects, known or unexpected workloads, routine work, or long-term technological and computing requirements. Web services and other specialized tasks are sometimes referenced as types of ODC.



**Advantages of On-Demand Computing:**

The on-demand computing model was developed to overcome the common challenge that enterprises encountered of not being able to meet unpredictable, fluctuating computing demands in an efficient manner. Businesses today need to be agile and need the ability to scale resources easily and quickly based on rapidly changing market needs.

Industry experts predict on-demand computing to soon be the most widely used computing model for enterprises.

## What is Cloud Computing?

**"Cloud refers to Internet or Network". "Cloud is an off-premise form of computing that stores data on the Internet."** Cloud can provide services over network i.e. public or private networks such as LAN, WAN, VPN.

- ❖ Cloud computing is defined to be Internet based computing technology, where the term 'cloud' simply means Internet – and cloud computing refers to services that are accessed directly over the

Internet. A 'cloud' data server or cluster is a collection of computer servers maintained by a cloud provider to provide computing services on a massive scale.

- ❖ Cloud computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. This is an **internet** definition of **Cloud computing**.
- ❖ Cloud computing is a general term for the delivery of hosted services over the internet.
- ❖ Cloud computing enables companies to consume a compute resource, such as a virtual machine (VM), storage or an application, as a utility -- just like electricity -- rather than having to build and maintain computing infrastructures in house.
- ❖ **Cloud computing** is the use of various services, such as software development platforms, servers, storage and software, over the internet, often referred to as the "**cloud**.. A user only pays for services used (memory, processing time and bandwidth, etc.).

**1) Cloud computing** is a computing term or metaphor that evolved in the late 2000s, based on utility and consumption of computing resources. Cloud computing involves deploying groups of remote servers and software networks that allow centralized data storage and online access to computer services or resources;

**2) Cloud computing** a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service **(IaaS)**, Platform-as-a-Service **(PaaS)** and Software-as-a-Service **(SaaS)**.

**3) Gartner group** 'A forecaster of Information technology' defines **cloud computing** as a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using Internet technologies.

**4) National Institute of Standards and Technology (NIST),** which defines the standard, defines **cloud computing** as, a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.