UNIT – I

Introduction:

- Computing is being transformed into a model consisting of services that are commoditized and Delivered in a manner similar to utilities such as water, electricity, gas, and telephony.
- In such a model, users access services based on their requirements, regardless of where the services are hosted.
- Several computing paradigms, such as grid-computing, have promised to deliver this utility computing vision.
- Cloud computing is the most recent emerging paradigm promising to turn the vision of “computing utilities” into a reality.
- Cloud computing is a technological advancement that focuses on the way we design computing systems, develop applications, and leverage existing services for building software.
- It is based on the concept of dynamic provisioning, which is applied not only to services but also to compute capability, storage, networking, and Information Technology(IT) infrastructure in general.
- Resources are made available through the Internet and offered on a pay-per-use basis from cloud Computing vendors.
- Today, anyone with a credit card can subscribe to cloud services and deploy and configure servers for an application based on their requirements and paying only for the time these resources have been used.

Cloud computing at a glance:

- In 1969, Leonard Kleinrock, one of the chief scientists of the original Advanced Research Projects Agency Network (ARPANET), which seeded the Internet, said.
  - As of now, computer networks are still in their beginning, but as they grow up and become sophisticated, we will probably see the spread of ‘computer utilities’ which, like present electric and telephone utilities, will service individual homes and offices across the country.
- This vision of computing utilities based on a service-provisioning model anticipated the massive transformation of the entire computing industry in the 21st century, where by computing services will be readily available on demand, just as other utility services such as water, electricity, telephone.
- In such a model, users access services based on their requirements without regard to where the services are hosted.

- This model has been referred to as utility computing or, recently(since2007), as cloud computing.
- The latter term often denotes the infrastructure as a "cloud" from which businesses and users can access applications as services from anywhere in the world and on demand.
- Cloud computing allows renting infrastructure, runtime environments, and services on a pay-per-use basis. This principle finds several practical applications and then gives different images of Cloud computing to different people.
- One of the most diffuse views of cloud computing can be summarized as follows:
  - I don't care where my servers are, who manages them, where my documents are stored, or where my applications are hosted. I just want them always available and access them from any device connected through Internet. And I am willing to pay for this service for as long as I need it.
- They have transformed the Internet into a rich application and service delivery platform, mature enough to serve complex needs.
  - Service orientation allows cloud computing to deliver its capabilities with familiar abstractions,
  - while virtualization deliberates on cloud computing the necessary degree of customization, control, and flexibility for building production and enterprise systems.
- Besides being an extremely flexible environment for building new systems and applications, cloud computing also provides an opportunity for integrating additional capacity or new features into existing systems.
- The use of dynamically provisioned IT resources constitutes a more attractive opportunity than buying additional infrastructure and software, the sizing of which can be difficult to estimate and the needs of which are limited in time.

The vision of cloud computing:

- Cloud computing allows anyone with a credit card to provision virtual hardware, runtime environments, and services. These are used for as long as needed, with no up-front commitments required.
- Cloud computing provides the facility to provision virtual hardware, runtime environment and services to a person having money.
- These all things can be used as long as they are needed by the user, there is no requirement for the upfront commitment.
- The whole collection of computing system is transformed into a collection of utilities, which can be provisioned and composed together to deploy systems in hours rather than days, with no maintenance costs.

- The long term vision of a cloud computing is that IT services are traded as utilities in an open market without technological and legal barriers.
- In the near future we can imagine that it will be possible to find the solution that matches with our requirements by simply entering our request in a global digital market that trades with cloud computing services.
- The existence of such market will enable the automation of the discovery process and its integration into its existing software systems.
- Due to the existence of a global platform for trading cloud services will also help service providers to potentially increase their revenue.
- A cloud provider can also become a consumer of a competitor service in order to fulfill its promises to customers.

Origins and Influences:

A Brief History:

- The idea of computing in a "cloud" trace back to the origins of utility computing, a concept that computer scientist John McCarthy publicly proposed in 1961.
- "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility".
- In 1969, Leonard Kleinrock, a chief scientist of the Advanced Research Projects Agency Network or ARPANET project that seeded the Internet, stated.
- "As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of 'computer utilities'.

- The general public has been leveraging forms of Internet-based computer utilities since the mid-1990s through various incarnations of search engines, email services, open publishing platforms, and other types of social media.
- Though consumer centric, these services popularized and validated core concepts that form the basis of modern-day cloud computing.
- In 1999, Salesforce.com pioneered the notion of bringing remotely provisioned services into the enterprise.
- In 2006, Amazon.com launched the Amazon Web Services (AWS) platform, a suite of enterprise-oriented services that provide remotely provisioned storage, computing resources, and business functionality.
- A slightly different elicitation of the term "Network Cloud" or "Cloud" was introduced in the early 1990s throughout the networking industry.
- It **referred to an abstraction layer** derived in the delivery methods of data across heterogeneous public and semi-public networks.
- The networking method at this point supported the transmission of data from one end-point (local network) to the "Cloud" (wide area network) and then further decomposed to another intended end-point.

Defining a cloud:

- Cloud computing has become a popular buzzword; it has been widely used to refer to different technologies, services, and concepts.
- It is often associated with virtualized infrastructure or hardware on demand, utility computing, IT outsourcing, platform and software as a service, and many other things that now are the focus of the IT industry.
- Forrester defines cloud computing as:
  - "A pool of abstracted, highly scalable, and managed compute infrastructure capable of hosting end-customer applications and billed by consumption.".
- Figure 1.2 depicts the plethora of different notions included in current definitions of cloud computing.

The NIST definition of Cloud:

- 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.

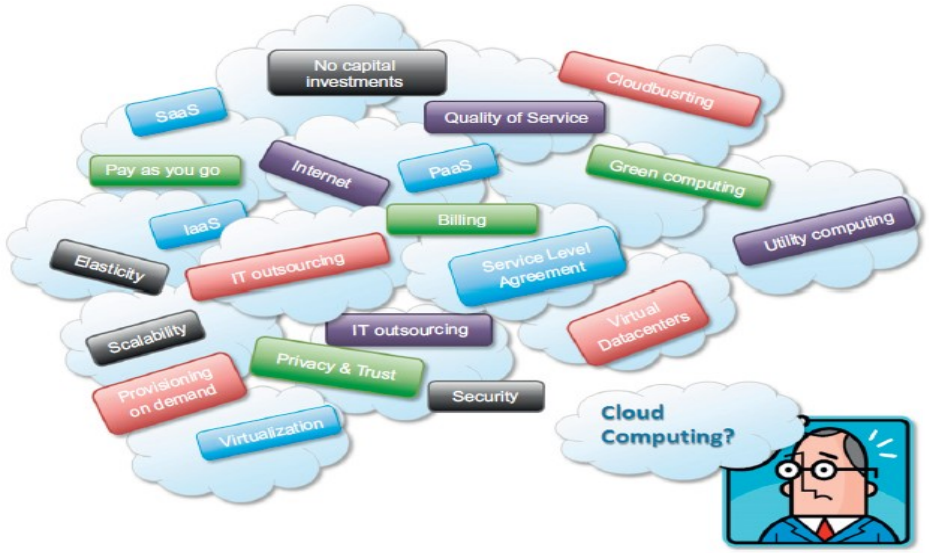


FIGURE 1.2
Cloud computing technologies, concepts, and ideas.

- We can define three criteria to discriminate whether a service is delivered in the cloud computing style:
  - The service is accessible via a Web browser (nonproprietary) or a Web services application programming interface (API).
  - Zero capital expenditure is necessary to get started.
  - You pay only for what you use as you use it.
- Even though many cloud computing services are freely available for single users, enterprise- class services are delivered according a specific pricing scheme.
- In this case users subscribe to the service and establish with the service provider a service-level agreement (SLA) defining the quality-of-service parameters under which the service is delivered.
- The utility-oriented nature of cloud computing is clearly expressed by Buyyaetal.[30]:
  - A cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers.
- Cloud computing takes the technology, services, and applications that are similar to those on the Internet and turns them into a self-service utility.
- The use of the word "cloud" makes reference to the two essential concepts……

1. **Abstraction:** Cloud computing abstracts the details of system implementation from users and developers.

    - Applications run on physical systems that aren't specified, data is stored in locations that are unknown, administration of systems is outsourced to others, and access by users is ubiquitous.

2. **Virtualization:** Cloud computing virtualizes systems by pooling and sharing resources.
    - Systems and storage can be provisioned as needed from a centralized infrastructure, costs are assessed on a metered basis, multi-tenancy and scalability is enabled.

- To help clarify how cloud computing has changed the nature of commercial system deployment, consider these three examples…
- **Google:**
    - In the last decade, Google has built a worldwide network of datacenters to service its search engine. In doing so Google has captured a substantial portion of the world's advertising revenue.
    - That revenue has enabled Google to offer free software to users based on that infrastructure and has changed the market for user-facing software.
- **Azure Platform:**
    - By contrast, Microsoft is creating the Azure Platform. It enables .NET Framework applications to run over the Internet.
- **Amazon Web Services:**
    - One of the most successful cloud-based businesses is Amazon Web Services, which is an Infrastructure as a Service.
- Most people separate cloud computing into two distinct sets of models…
    - **Deployment models:** This refers to the location and management of the cloud's infrastructure.
    - **Service models:** This consists of the particular types of services that you can access on a cloud computing platform.

Business Drivers:

- Before exploring into the layers of technologies that underlie clouds, the motivations that led to their creation by industry leaders must first be understood.
- Several of the primary business drivers that fostered modern cloud-based technology.

1. Capacity Planning:

- Capacity planning is the process of determining and fulfilling future demands of an organization's IT resources, products, and services.

- Within this context, capacity represents the maximum amount of work that an IT resource is capable of delivering in a given period of time.
- A discrepancy between the capacity of an IT resource and its demand can result in a system becoming either inefficient (over-provisioning) or unable to fulfill user needs (under-provisioning).
- Capacity planning is focused on minimizing this discrepancy to achieve predictable efficiency and performance.
- Different capacity planning strategies exist……
  - Lead Strategy – adding capacity to an IT resource in anticipation of demand.
  - Lag Strategy – adding capacity when the IT resource reaches its full capacity.
  - Match Strategy – adding IT resource capacity in small increments, as demand increases.
- Planning for capacity can be challenging because it requires estimating usage load fluctuations.
- There is a constant need to balance peak usage requirements without unnecessary over-expenditure on infrastructure.
- An example is outfitting IT infrastructure to accommodate maximum usage loads which can impose unreasonable financial investments.
- In such cases, moderating investments can result in under-provisioning, leading to transaction losses and other usage limitations from lowered usage thresholds.

2. Cost Reduction:

- A direct alignment between IT costs and business performance can be difficult to maintain.
- The growth of IT environments often corresponds to the assessment of their maximum usage requirements, this can make the support of new and expanded business automations an ever-increasing investment.
- Much of this required investment is focused into infrastructure expansion because the usage potential of a given automation solution will always be limited by the processing power of its underlying infrastructure.
- Two costs need to be accounted for: the cost of acquiring new infrastructure, and the cost of its ongoing ownership.
- Operational overhead represents a considerable share of IT budgets, often exceeding up-front investment costs.
- Common forms of infrastructure-related operating overhead include the following:

- Technical personnel required to keep the environment operational
- Upgrades and patches that introduce additional testing and deployment cycles.
- Utility bills and capital expense investments for power and cooling.
- Security and access control measures that need to be maintained and enforced to protect infrastructure resources.
- Administrative and accounts staff that may be required to keep track of licenses and support arrangements.

- The on-going ownership of internal technology infrastructure can include burdensome responsibilities that impose compound impacts on corporate budgets.

3. Organizational Agility:

- Businesses need the ability to adapt and evolve to successfully face change caused by both internal and external factors.
- Organizational agility is the measure of an organization's responsiveness to change.
- An IT enterprise often needs to respond to business change by scaling its IT resources beyond the scope of what was previously predicted or planned for.
  - For eg., changing business needs and priorities may require IT resources to be more available and reliable than before.
  - Even if sufficient infrastructure is in place for an organization to support anticipated usage volumes, the nature of the usage may generate runtime exceptions that bring down hosting servers.
- On a broader scale, the up-front investments and infrastructure ownership costs that are required to enable new or expanded business automation solutions may themselves be prohibitive enough for a business.
- Worse yet, the business may decide against proceeding with an automation solution altogether upon review of its infrastructure budget, because it simply cannot afford to.
- This form of inability to respond can inhibit an organization from keeping up with market demands, competitive pressures, and its own strategy.

Technology Innovations:

- Established technologies are often used as inspiration and, at times, the actual foundations upon which new technology innovations are derived and built.

- Here we describe the pre-existing technologies considered to be the primary influences on cloud computing.

1. Clustering:

- A cluster is a group of independent IT resources that are interconnected and work as a single system.
- System failure rates are reduced while availability and reliability are increased, since redundancy and failover features are inherent to the cluster.
- A general prerequisite of hardware clustering is that its component systems have reasonably identical hardware and operating systems to provide similar performance levels when one failed component is to be replaced by another.
- Component devices that form a cluster are kept in synchronization through dedicated, high-speed communication links.
- The basic concept of built-in redundancy and failover is core to cloud platforms.

2. Grid Computing:

- A computing grid (or "computational grid") provides a platform in which computing resources are organized into one or more logical pools.
- These pools are collectively coordinated to provide a high-performance distributed grid, sometimes referred to as a "super virtual computer."
- Grid computing differs from clustering in that grid systems are much more loosely coupled and distributed.
- As a result, grid computing systems can involve computing resources that are heterogeneous and geographically dispersed.
- The technological advancements achieved by grid computing projects have influenced various aspects of cloud computing platforms and mechanisms such as networked access, resource pooling, and scalability and resiliency.
- These types of features can be established by both grid computing and cloud computing, in their own distinctive approaches.
    - For example, grid computing is based on a middleware layer that is deployed on computing resources.
    - The middle tier can contain load balancing logic, failover controls, and autonomic configuration management.
- It is for this reason that some classify cloud computing as a successor of earlier grid computing initiatives.

3. Virtualization:

- Virtualization represents a technology platform used for the creation of virtual instances of IT resources.

- A layer of virtualization software allows physical IT resources to provide multiple virtual images of themselves so that their underlying processing capabilities can be shared by multiple users.
- Prior to the advent of virtualization technologies, software was limited to residing on and being coupled with static hardware environments.
- The virtualization process severs this software-hardware dependency, as hardware requirements can be simulated by emulation software.

Basic Concepts and Terminology:

- This section represents the fundamental concepts and aspects pertaining to the notion of a cloud and its most primitive artifacts.
- **Cloud:**
  - A cloud refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources.
  - The term originated as a metaphor for the Internet which is, in essence, a network of networks providing remote access to a set of decentralized IT resources.
  - The symbol of a cloud was commonly used to represent the Internet in a variety of specifications and mainstream documentation of Web-based architectures.
  - The Internet provides open access to many Web-based IT resources, a cloud is typically privately owned and offers access to IT resources that is metered.
- **IT Resource:**
  - An IT resource is a physical or virtual IT-related artifact that can be either software-based, such as a virtual server or a custom software program, or hardware-based, such as a physical server or a network device (Figure 3.2).

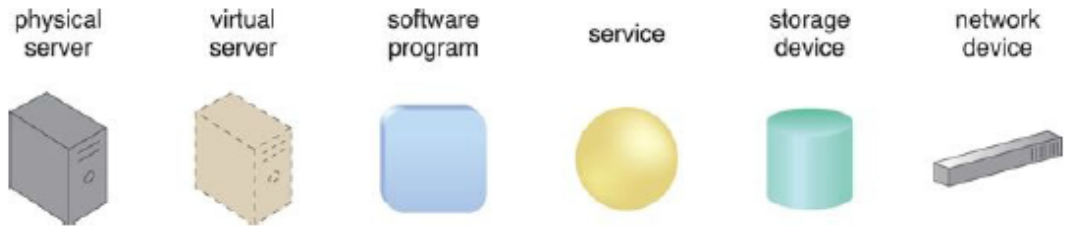


Figure 3.2. Examples of common IT resources and their corresponding symbols.

- **On-Premise:**
  - As a distinct and remotely accessible environment, a cloud represents an option for the deployment of IT resources.
  - An IT resource that is hosted in a conventional IT enterprise within an organizational boundary (that does not specifically

represent a cloud) is considered to be located on the premises of the IT enterprise, or on-premise.

- Note the following key points:
  - An on-premise IT resource can access and interact with a cloud-based IT resource.
  - An on-premise IT resource can be moved to a cloud, thereby changing it to a cloud-based IT resource.
  - Redundant deployments of an IT resource can exist in both on-premise and cloud-based environments.

- **Cloud Consumers and Cloud Providers:**
  - The party that provides cloud-based IT resources is the cloud provider. The party that uses cloud-based IT resources is the cloud consumer.
  - These terms represent roles usually assumed by organizations in relation to clouds and corresponding cloud provisioning contracts.

- **Scaling:**
  - Scaling, from an IT resource perspective, represents the ability of the IT resource to handle increased or decreased usage demands.
  - The following are types of scaling:
    - Horizontal Scaling – scaling out and scaling in
    - Vertical Scaling – scaling up and scaling down.

Horizontal Scaling:

- The allocating or releasing of IT resources that are of the same type is referred to as horizontal scaling (Figure 3.4).
- The horizontal allocation of resources is referred to as scaling out and the horizontal releasing of resources is referred to as scaling in.

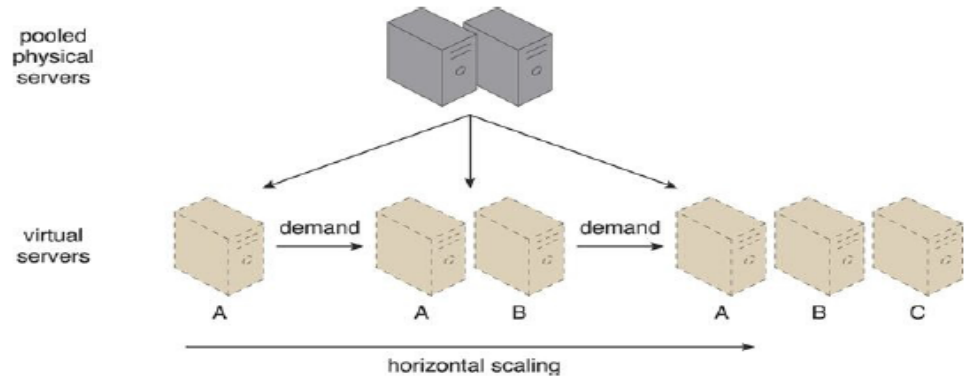


Figure 3.4. An IT resource (Virtual Server A) is scaled out by adding more of the same IT resources (Virtual Servers B and C).

Vertical Scaling:

- When an existing IT resource is replaced by another with higher or lower capacity, vertical scaling is considered to have occurred (Figure 3.5).

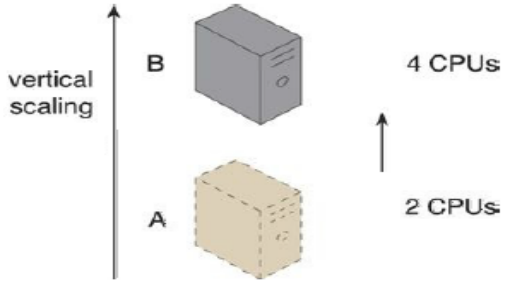


Figure 3.5. An IT resource (a virtual server with two CPUs) is scaled up by replacing it with a more powerful IT resource with increased capacity for data storage (a physical server with four CPUs).

- Specifically, the replacing of an IT resource with another that has a higher capacity is referred to as scaling up and the replacing an IT resource with another that has a lower capacity is considered scaling down.

- **Cloud Service:**
  - Although a cloud is a remotely accessible environment, not all IT resources residing within a cloud can be made available for remote access.
  - For example, a database or a physical server deployed within a cloud may only be accessible by other IT resources that are within the same cloud.
  - A software program with a published API may be deployed specifically to enable access by remote clients.
  - A cloud service is any IT resource that is made remotely accessible via a cloud.
  - A cloud service can exist as a simple Web-based software program with a technical interface invoked via the use of a messaging protocol, or as a remote access point.
  - In Figure 3.6, the yellow circle symbol is used to represent the cloud service as a simple Web-based software program.

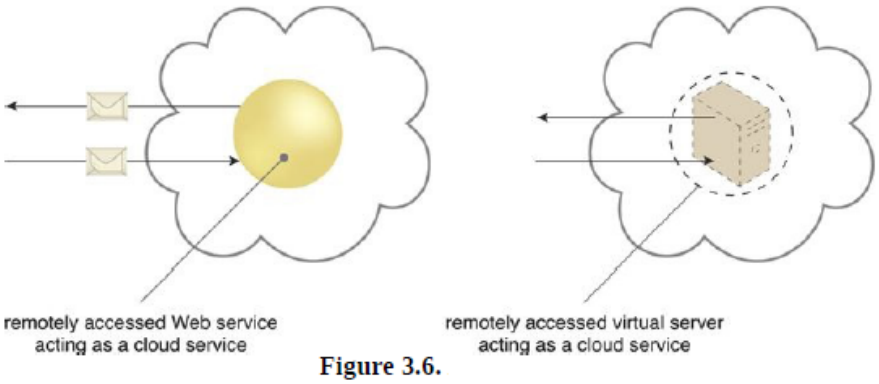


Figure 3.6.

- **Cloud Service Consumer:**
  - The cloud service consumer is a temporary runtime role assumed by a software program when it accesses a cloud service.
  - As shown in Figure 3.7, common types of cloud service consumers can include software programs and services capable of remotely accessing cloud services with published service contracts.

Figure 3.7. Examples of cloud service consumers. Depending on the nature of a given diagram, an artifact labeled as a cloud service consumer may be a software program or a hardware device (in which case it is implied that it is running a software program capable of acting as a cloud service consumer).

Goals and Benefits:

- The **common goals** associated with adopting cloud computing are……
- **Increased Responsiveness:**
  - Cloud computing plays an essential role in enhancing an organization's business agility by empowering it to be more responsive to business and usage scenarios that can be more effectively addressed by leveraging native cloud capabilities.
  - Such as scalability on-demand, data availability, reduced infrastructure maintenance, reduced business complexity, automation, and increased up-time.
  - For example, greater data availability can enable employees to more easily work remotely, thereby providing staff with increased flexibility and productivity.
  - Utilizing platforms maintained by cloud providers, frees organizations from the responsibilities they would normally have for administering them internally.
- **Reduced Investments and Proportional Costs:**
  - Similar to a product wholesaler that purchases goods in bulk for lower price points, public cloud providers base their business model on the mass-acquisition of IT resources that are then made available to cloud consumers via attractively priced leasing packages.
  - This opens the door for organizations to gain access to powerful infrastructure without having to purchase it themselves.
  - The most common economic rationale for investing in cloud-based IT resources is in the reduction or outright elimination of up-front IT investments, namely hardware and software purchases and ownership costs.
  - A cloud's Measured Usage characteristic represents a feature-set that allows measured operational expenditures to replace anticipated capital expenditures.

Common benefits to cloud consumers include:

- On-demand access to pay-as-you-go computing resources on a short-term basis (such as processors by the hour), and the ability to release these computing resources when they are no longer needed.

- The perception of having unlimited computing resources that are available on-demand, thereby reducing the need to prepare for provisioning.
- The ability to add or remove IT resources at a fine-grained level, such as modifying available storage disk space by single gigabyte increments.
- Abstraction of the infrastructure so applications are not locked into devices or locations and can be easily moved if needed.
- For example, a company with sizable batch-centric tasks can complete them as quickly as their application software can scale.
- **Increased Scalability:**
  - By providing pools of IT resources, along with tools and technologies designed to leverage them collectively, clouds can instantly and dynamically allocate IT resources to cloud consumers, on-demand.
- **Increased Availability and Reliability:** The availability and reliability of IT resources are directly associated with tangible business benefits.
  - Outages limit the time an IT resource can be "open for business" for its customers, thereby limiting its usage and revenue generating potential.
  - Runtime failures that are not immediately corrected can have a more significant impact during high-volume usage periods.

Risks and Challenges:

- Several of the most critical cloud computing challenges are……
- **Increased Vulnerability Due to Overlapping Trust Boundaries:**
  - Moving business data to the cloud means that the responsibility over data security is shared with the cloud provider.
  - The remote usage of IT resources requires an expansion of trust boundaries by the cloud consumer to include the cloud, external to the organization.
  - It can be difficult to establish a security architecture that spans such a trust boundary without introducing vulnerabilities.
  - Unless cloud consumers and cloud providers happen to support the same or compatible security frameworks, which is doubtful with public clouds.
  - Another consequence of overlapping trust boundaries relates to the cloud provider's privileged access to cloud consumer data.
  - The extent to which the data is secure is now limited to the security controls and policies applied by both the cloud consumer and cloud provider.

- The overlapping of trust boundaries and the increased exposure of data can provide malicious cloud consumers (human and automated) with greater opportunities to attack IT resources and steal or damage business data.

- **Increased Vulnerability Due to Shared Security Responsibility:**
  - Information security related to on-premise resources is clearly the responsibility of the organization that owns those resources.
  - However, information security related to cloud-based resources is not the sole responsibility of the cloud provider, even if the cloud-based resources are owned by the cloud provider.
  - This is because the information stored and processed in them is owned by the cloud consumer.
  - As a result, information security in the cloud is a shared responsibility, with both the cloud provider and the cloud consumer having a role to play in securing the cloud environment.
  - It is important to be able to understand and identify where the responsibility for each role begins and ends, as well as knowing how to address the security requirements that correspond to the cloud consumer.
  - A cloud provider will typically propose a cloud shared responsibility model as part of the SLA, which essentially outlines the respective responsibilities.
- **Increased Exposure to Cyber Threats:**
  - The increased adoption of contemporary digital technologies and digital transformation practices has led organizations to move more IT resources toward and build more solutions within cloud environments.
  - This has opened the door to cybersecurity threats and risks that may be new to organizations and for which they need to be prepared.
- **Reduced Operational Governance Control:**
  - Cloud consumers are usually allotted a level of governance control that is lower than that over on-premise IT resources.
  - This can introduce risks associated with how the cloud provider operates its cloud, as well as the external connections that are required for communication between the cloud and the cloud consumer.
- **Limited Portability Between Cloud Providers:**
  - Due to a lack of established industry standards within the cloud computing industry, public clouds are commonly proprietary to various extents.

- For cloud consumers that have custom-built solutions with dependencies on these proprietary environments, it can be challenging to move from one cloud provider to another.
- Portability is a measure used to determine the impact of moving cloud consumer IT resources and data between clouds.

- **Multi-Regional Compliance and Legal Issues:**
  - Third-party cloud providers will frequently establish data centers in affordable or convenient geographical locations.
  - Cloud consumers will often not be aware of the physical location of their IT resources and data when hosted by public clouds.
  - For some organizations, this can pose serious legal concerns pertaining to industry or government regulations that specify data privacy and storage policies.
  - Another potential legal issue pertains to the accessibility and disclosure of data.
  - Countries have laws that require some types of data to be disclosed to certain government agencies or to the subject of the data.
  - Most regulatory frameworks recognize that cloud consumer organizations are ultimately responsible for the security, integrity, and storage of their own data, even when it is held by an external cloud provider.

Roles and Boundaries:

- Organizations and humans can assume different types of predefined roles depending on how they relate to and/or interact with a cloud and its hosted IT resources.
- Each of the upcoming roles participates in and carries out responsibilities in relation to cloud-based activity.
- **Cloud Provider:**
  - The organization that provides cloud-based IT resources is the cloud provider.
  - When assuming the role of cloud provider, an organization is responsible for making cloud services available to cloud consumers, as per agreed upon SLA guarantees.
  - The cloud provider is further tasked with any required management and administrative duties to ensure the on-going operation of the overall cloud infrastructure.
  - Cloud providers normally own the IT resources that are made available for lease by cloud consumers; however, some cloud providers also “resell” IT resources leased from other cloud providers.

- **Cloud Consumer:**
  - A cloud consumer is an organization (or a human) that has a formal contract or arrangement with a cloud provider to use IT resources made available by the cloud provider.
  - Specifically, the cloud consumer uses a cloud service consumer to access a cloud service.
- **Cloud Broker:**
  - A third-party organization that assumes the responsibility of negotiating, managing and operating cloud services on behalf of a cloud consumer is assuming the role of cloud broker.
  - Cloud brokers can provide mediation services between cloud consumers and cloud providers, including intermediation, aggregation, arbitrage, and others.
  - A cloud broker commonly provides these services for multiple cloud consumers facing multiple cloud providers alternatively or simultaneously, acting as an integrator of cloud services and an aggregator of cloud consumers.
- **Cloud Service Owner:**
  - The person or organization that legally owns a cloud service is called a cloud service owner.
  - The cloud service owner can be the cloud consumer, or the cloud provider that owns the cloud within which the cloud service resides.
  - A cloud consumer can be a cloud service owner when it deploys its own service in a cloud.
- **Cloud Resource Administrator:**
  - A cloud resource administrator is the person or organization responsible for administering a cloud-based IT resource (including cloud services).
  - The cloud resource administrator can be (or belong to) the cloud consumer or cloud provider of the cloud within which the cloud service resides.
  - Alternatively, it can be (or belong to) a third-party organization contracted to administer the cloud-based IT resource.
  - A cloud resource administrator can be with a cloud consumer organization and administer remotely accessible IT resources that belong to the cloud consumer.
  - A cloud resource administrator can be with a cloud provider organization for which it can administer the cloud provider's internally and externally available IT resources.
- The NIST Cloud Computing Reference Architecture defines the following supplementary roles……

- **Cloud Auditor:**
  - A third-party (often accredited) that conducts independent assessments of cloud environments assumes the role of the cloud auditor.
  - The typical responsibilities associated with this role include the evaluation of security controls, privacy impacts, and performance.
  - The main purpose of the cloud auditor role is to provide an unbiased assessment (and possible endorsement) of a cloud environment to help strengthen the trust.
- **Cloud Broker:**
  - This role is a party that assumes the responsibility of managing and negotiating the usage of cloud services between cloud consumers and cloud providers.

Characteristics and benefits:

- Cloud computing has some interesting characteristics that bring benefits to both cloud service consumer's(CSCs) and cloud service providers(CSPs). These characteristics are as following…
  - No up-front commitments
  - On-demand access
  - Nice pricing
  - Simplified application acceleration and scalability.
  - Efficient resource allocation
  - Energy efficiency
  - Seamless creation and use of third-party services.
- The most evident benefit from the use of cloud computing systems and technologies is the increased economical return due to the reduced maintenance costs and operational costs related to IT software and infrastructure.
- This is mainly because IT assets, namely software and infrastructure, are turned into utility costs, which are paid for as long as they are used, not paid for up front.
- Before cloud computing, IT infrastructure and software generated capital costs, since they were paid up front so that business start-ups could afford a computing infrastructure, enabling the business activities of the organization.
- The revenue of the business is then utilized to compensate over time for these costs. Organizations always minimize capital costs, since they are often associated with depreciable values.
- Minimizing capital costs, then, is fundamental. Cloud computing transforms IT infrastructure and software into utilities, thus significantly contributing to increasing a company's net gain.

- Moreover, cloud computing also provides an opportunity for small organizations and start-ups: these do not need large investments to start their business.
- Finally, maintenance costs are significantly reduced: by renting the infrastructure and the application services, organizations are no longer responsible for their maintenance.
- Increased agility in defining and structuring software systems is another significant benefit of cloud computing.
- Since organizations rent IT services, they can more dynamically and flexibly compose their software systems, without being constrained by capital costs for IT assets.
- There is a reduced need for capacity planning, since cloud computing allows organizations to react to unplanned surges in demand quite rapidly.
- For example, organizations can add more servers to process workload spikes and dismiss them when they are no longer needed.
- Ease of scalability is another advantage. By leveraging the potentially huge capacity of cloud computing, organizations can extend their IT capability more easily.
- Infrastructure providers offer simple methods to provision customized hardware and integrate it into existing systems.
    - Platform-as-a-Service providers offer runtime environment and programming models that are designed to scale applications.
    - Software-as-a-Service offerings can be elastically sized on demand without requiring users to provision hardware or to program application for scalability.
- End users can benefit from cloud computing by having their data and the capability of operating on it always available, from anywhere, at any time, and through multiple devices.
    - Information and services stored in the cloud are exposed to users by Web-based interfaces that make them accessible from portable devices as well as desktops at home.
- Since the processing capabilities (that is, office automation) also reside in the cloud, end users can perform the same tasks that previously were carried out through considerable software investments.
- The cost for such opportunities is generally very limited, since the cloud service provider shares its costs across all the tenants that he is servicing.
    - Multitenancy allows for better utilization of the shared infrastructure that is kept operational and fully active.
- The concentration of IT infrastructure and services into large datacenters also provides opportunity for considerable optimization in terms of resource allocation and energy

efficiency, which eventually can lead to a less impacting approach on the environment.

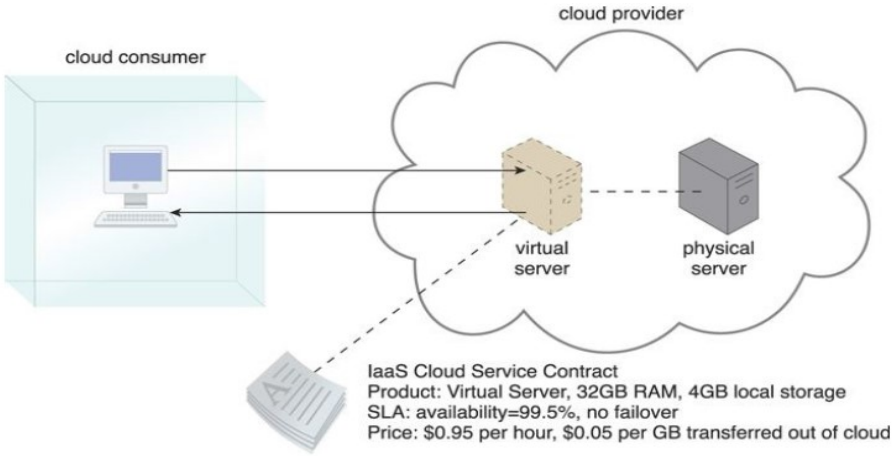
- Finally, service orientation and on-demand access create new opportunities for composing systems and applications with a flexibility not possible before cloud computing.
- New service offerings can be created by combining together existing services and concentrating on added value.
- Since it is possible to provision on demand any component of the computing stack, it is easier to turn ideas into products with limited costs.

Cloud delivery Models:

- A cloud delivery model represents a specific, pre-packaged combination of IT resources offered by a cloud provider.
- Three common cloud delivery models have become widely established and formalized……
  - Infrastructure-as-a-Service (IaaS)
  - Platform-as-a-Service (PaaS)
  - Software-as-a-Service (SaaS)

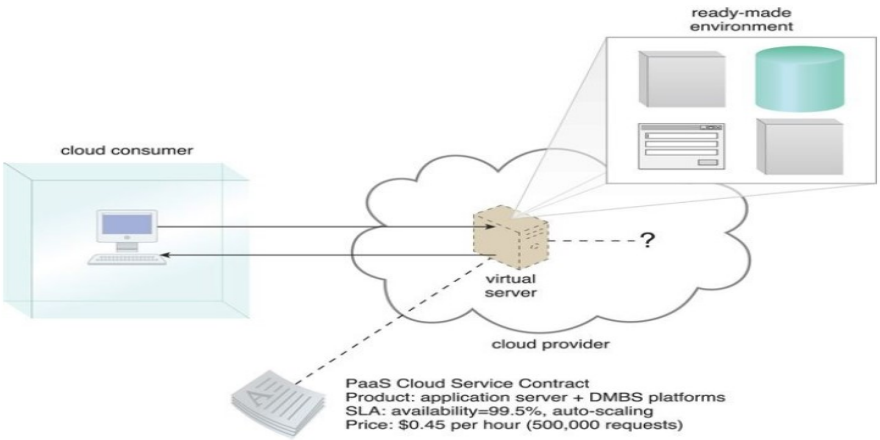
Infrastructure-as-a-Service (IaaS):

- The IaaS delivery model represents a self-contained IT environment comprised of infrastructure-centric IT resources that can be accessed and managed via cloud service-based interfaces and tools.
- This environment can include hardware, network, connectivity, operating systems, and other "raw" IT resources.
- In contrast to traditional hosting or outsourcing environments, with IaaS, IT resources are typically virtualized and packaged into bundles that simplify up-front runtime scaling and customization of the infrastructure.
- The general purpose of an IaaS environment is to provide cloud consumers with a high level of control and responsibility over its configuration and utilization.
- The IT resources provided by IaaS are generally not pre-configured, placing the administrative responsibility directly upon the cloud consumer.
- This model is therefore used by cloud consumers that require a high level of control over the cloud-based environment they intend to create.
- A central and primary IT resource within a typical IaaS environment is the virtual server.
- Virtual servers are leased by specifying server hardware requirements, such as processor capacity, memory, and local storage space, as shown in Figure 4.11.



Platform-as-a-Service (PaaS):

- The PaaS delivery model represents a pre-defined "ready-to-use" environment typically comprised of already deployed and configured IT resources.
- Specifically, PaaS relies on the usage of a readymade environment that establishes a set of pre-packaged products and tools used to support the entire delivery lifecycle of custom applications.
- Common reasons a cloud consumer would use and invest in a PaaS environment include:
  - The cloud consumer wants to extend on-premise environments into the cloud for scalability and economic purposes.
  - The cloud consumer uses the ready-made environment to entirely substitute an on-premise environment.
  - The cloud consumer wants to become a cloud provider and deploys its own cloud services to be made available to other external cloud consumers.
- By working with PaaS, the cloud consumer is spared the administrative burden of setting up and maintaining the bare infrastructure IT resources.
- Conversely, the cloud consumer is granted a lower level of control over the underlying IT resources that host and provision the platform (Figure 4.12).



- PaaS products are available with different development stacks. For example, Google App Engine offers a Java and Python-based environment.

Software-as-a-Service (SaaS):

- A software program positioned as a shared cloud service and made available as a "product" or generic utility represents the typical profile of a SaaS offering.
- The SaaS delivery model is typically used to make a reusable cloud service widely available (often commercially) to a range of cloud consumers.
- An entire marketplace exists around SaaS products that can be leased and used for different purposes and via different terms (Figure 4.13).

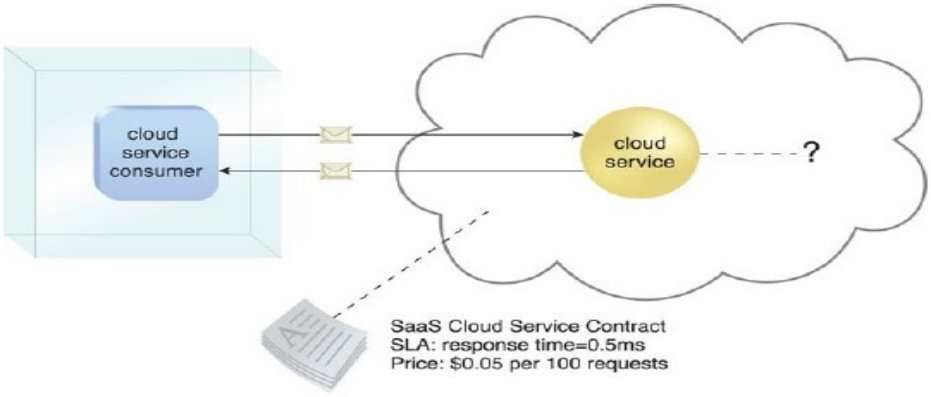


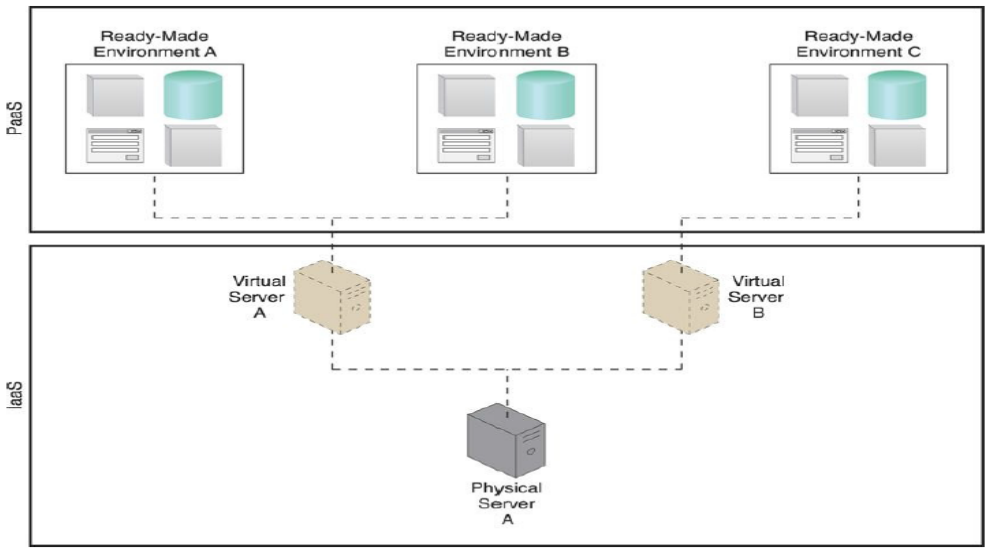
Figure 4.13. The cloud service consumer is given access the cloud service contract, but not to any underlying IT resources or implementation details.

- A cloud consumer is generally granted very limited administrative control over a SaaS implementation.
- It is most often provisioned by the cloud provider, but it can be legally owned by whichever entity assumes the cloud service owner role.
- For example:
  - An organization acting as a cloud consumer while using and working with a PaaS environment can build a cloud service that it decides to deploy in that same environment as a SaaS offering.
  - The same organization then effectively assumes the cloud provider role as the SaaS-based cloud service is made available to other organizations that act as cloud consumers when using that cloud service.
- Table 4.1. A comparison of typical cloud delivery model control levels.

Cloud Delivery Model	Typical Level of Control Granted to Cloud Consumer	Typical Functionality Made Available to Cloud Consumer
SaaS	usage and usage-related configuration	access to front-end user-interface
PaaS	limited administrative	moderate level of administrative control over IT resources relevant to cloud consumer's usage of platform
IaaS	full administrative	full access to virtualized infra-structure-related IT resources and, possibly, to underlying physical IT resources

Combining Cloud Delivery Models:

- The three base cloud delivery models comprise a natural provisioning hierarchy, allowing for opportunities for the combined application of the models to be explored.
- IaaS + PaaS: A PaaS environment will be built upon an underlying infrastructure comparable to the physical and virtual servers and other IT resources provided in an IaaS environment.

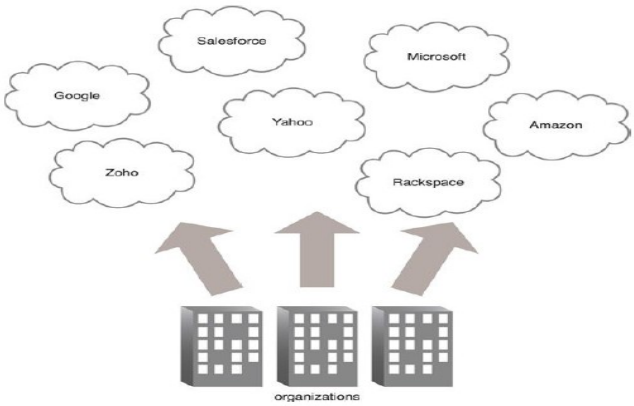


Cloud Deployment Models:

- A cloud deployment model represents a specific type of cloud environment, primarily distinguished by ownership, size, and access. There are four common cloud deployment models……

Public Clouds:

- A public cloud is a publicly accessible cloud environment owned by a third-party cloud provider.
- The IT resources on public clouds are usually provisioned via the previously described cloud delivery models and are generally offered to cloud consumers at a cost or are commercialized.
- The cloud provider is responsible for the creation and on-going maintenance of the public cloud and its IT resources.
- Figure 4.17 shows a partial view of the public cloud landscape, highlighting some of the primary vendors in the marketplace.



Community Clouds:

- A community cloud is similar to a public cloud except that its access is limited to a specific community of cloud consumers.
- The community cloud may be jointly owned by the community members or by a third-party cloud provider that provisions a public cloud with limited access.
- The member cloud consumers of the community typically share the responsibility for defining and evolving the community cloud.
- Membership in the community does not necessarily guarantee access to or control of all the cloud's IT resources.
- Parties outside the community are generally not granted access unless allowed by the community.

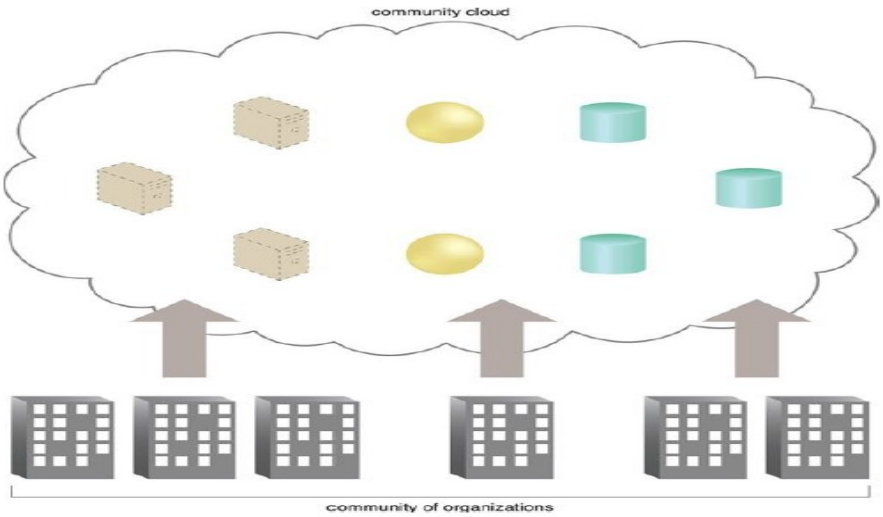


Figure 4.18. An example of a "community" of organizations accessing IT resources from a community cloud.

Private Clouds:

- A private cloud is owned by a single organization. Private clouds enable an organization to use cloud computing technology as a means of centralizing access to IT resources by different parts, locations, or departments of the organization.
- When a private cloud exists as a controlled environment, the problems such as Risks and Challenges do not tend to apply.
- The use of a private cloud can change how organizational and trust boundaries are defined and applied.
- The actual administration of a private cloud environment may be carried out by internal or outsourced staff.
- With a private cloud, the same organization is technically both the cloud consumer and cloud provider (Figure 4.19).
- In order to differentiate these roles:
    - A separate organizational department typically assumes the responsibility for provisioning the cloud (and therefore assumes the cloud provider role)
    - Departments requiring access to the private cloud assume the cloud consumer role.

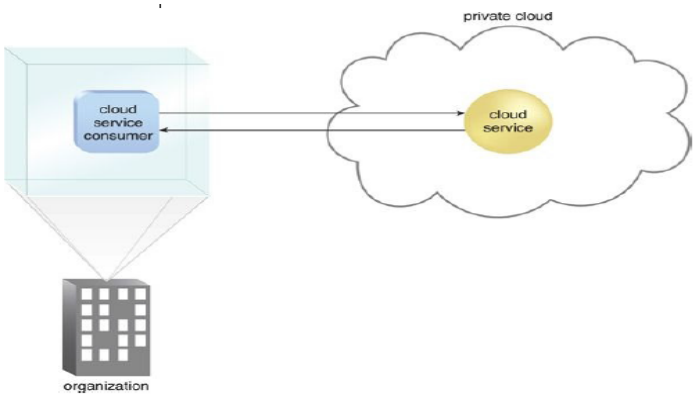


Figure 4.19. A cloud service consumer in the organization's on-premise environment accesses a cloud service hosted on the same organization's private cloud via a virtual private network.

- It is important to use the terms "on-premise" and "cloud-based" correctly within the context of a private cloud.
- Even though the private cloud may physically reside on the organization's premises, IT resources it hosts are still considered "cloud-based" as long as they are made remotely accessible to cloud consumers.
- IT resources hosted outside of the private cloud by the departments acting as cloud consumers are therefore considered "on-premise".

Hybrid Clouds:

- A hybrid cloud is a cloud environment comprised of two or more different cloud deployment models.
    - For example, a cloud consumer may choose to deploy cloud services processing sensitive data to a private cloud and other, less sensitive cloud services to a public cloud.
- The result of this combination is a hybrid deployment model (Figure 4.20).
- Hybrid deployment architectures can be complex and challenging to create and maintain due to the potential disparity in cloud environments and
- In fact, the management responsibilities are typically split between the private cloud provider organization and the public cloud provider.

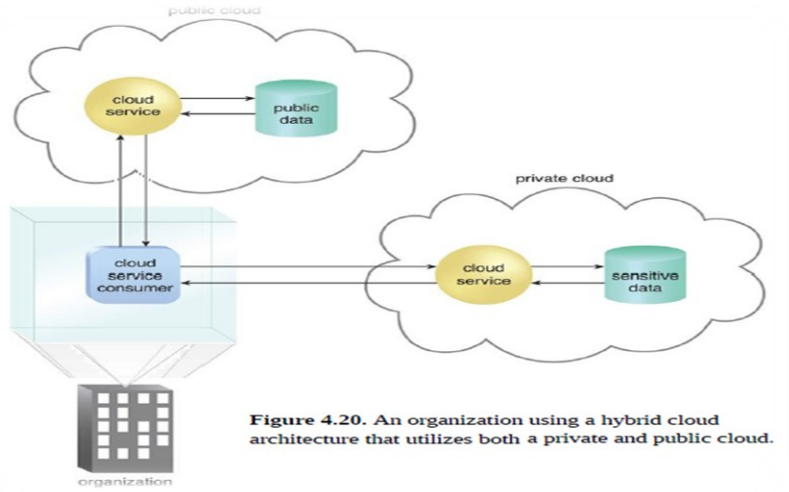


Figure 4.20. An organization using a hybrid cloud architecture that utilizes both a private and public cloud.

Other Cloud Deployment Models:

- Additional variations of the four base cloud deployment models can exist, for examples……
- Virtual Private Cloud – Also known as a "dedicated cloud" or "hosted cloud," this model results in a self-contained cloud environment hosted and managed by a public cloud provider, and made available to a cloud consumer.
- Inter-Cloud – This model is based on an architecture comprised of two or more inter-connected clouds.

Virtualization:

- Virtualization technology is one of the fundamental components of cloud computing, especially in regard to infrastructure-based services.
- Virtualization allows the creation of a secure, customizable, and isolated execution environment for running applications, even if they are untrusted, without affecting other users' applications.
- The basis of this technology is the ability of a computer program—or a combination of software and hardware— to emulate an executing environment separate from the one that hosts such programs.
- For example, we can run Windows OS on top of a virtual machine, which itself is running on Linux OS.
- Virtualization provides a great opportunity to build elastically scalable systems that can provision additional capability with minimum costs.
- Virtualization is a large umbrella of technologies and concepts that are meant to provide an abstract environment—whether virtual hardware or an operating system—to run applications.
- The term virtualization is often synonymous with hardware virtualization, which plays a fundamental role in efficiently delivering Infrastructure-as-a-Service (IaaS) solutions for cloud computing.
- Moreover, virtualization technologies provide a virtual environment for not only executing applications but also for storage, memory, and networking.
- Virtualization technologies have gained renewed interested recently due to the meeting of several phenomena's:
- **Increased performance and computing capacity.**
    - Nowadays, the average end-user desktop PC is powerful enough to meet almost all the needs of everyday computing, with extra capacity that is rarely used.

- Almost all these PCs have resources enough to host a virtual machine manager and execute a virtual machine with by far acceptable performance.
- The supercomputers can provide immense compute power that can accommodate the execution of hundreds or thousands of virtual machines.

- **Underutilized hardware and software resources.**
  - Hardware and software underutilization is occurring due to……
  - (1) increased performance and computing capacity, and (2) the effect of limited or irregular use of resources.
  - Computers today are so powerful that in most cases only a fraction of their capacity is used by an application or the system.
  - For example, desktop PCs mostly devoted to office automation tasks and used by administrative staff are only used during work hours, remaining completely unused overnight.
  - Using these resources for other purposes after hours could improve the efficiency of the IT infrastructure.
  - To transparently provide such a service, it would be necessary to deploy a completely separate environment, which can be achieved through virtualization.
- **Lack of space.**
  - The continuous need for additional capacity, whether storage or compute power, makes data centers grow quickly.
  - Companies such as Google and Microsoft expand their infrastructures by building data centers as large as football fields that are able to host thousands of nodes.
  - Although this is viable for IT giants, in most cases enterprises cannot afford to build another data center to accommodate additional resource capacity.
  - This condition, along with hardware underutilization, has led to the diffusion of a technique called server consolidation, for which virtualization technologies are fundamental.
- **Greening initiatives.**
  - Recently, companies are increasingly looking for ways to reduce the amount of energy they consume and to reduce their carbon footprint.
  - Data centers are one of the major power consumers; they contribute consistently to the impact that a company has on the environment.
  - Maintaining a data center operation not only involves keeping servers on, but a great deal of energy is also consumed in keeping them cool.

- Infrastructures for cooling have a significant impact on the carbon footprint of a data center.
- Hence, reducing the number of servers through server consolidation will definitely reduce the impact of cooling and power consumption of a data center.

- **Rise of administrative costs.**
  - Power consumption and cooling costs have now become higher than the cost of IT equipment.
  - Moreover, the increased demand for additional capacity, which translates into more servers in a data center, is also responsible for a significant increment in administrative costs.
  - Computers—in particular, servers—do not operate all on their own, but they require care and feeding from system administrators.
  - Virtualization can help reduce the number of required servers for a given workload, thus reducing the cost of the administrative personnel.

Characteristics of virtualized environments:

- Virtualization is a broad concept that refers to the creation of a virtual version of something, whether hardware, a software environment, storage, or a network.
- In a virtualized environment there are three major components: guest, host, and virtualization layer.
  - The guest represents the system component that interacts with the virtualization layer rather than with the host, as would normally happen.
  - The host represents the original environment where the guest is supposed to be managed.
  - The virtualization layer is responsible for recreating the same or a different environment where the guest will operate (see Figure 3.1).

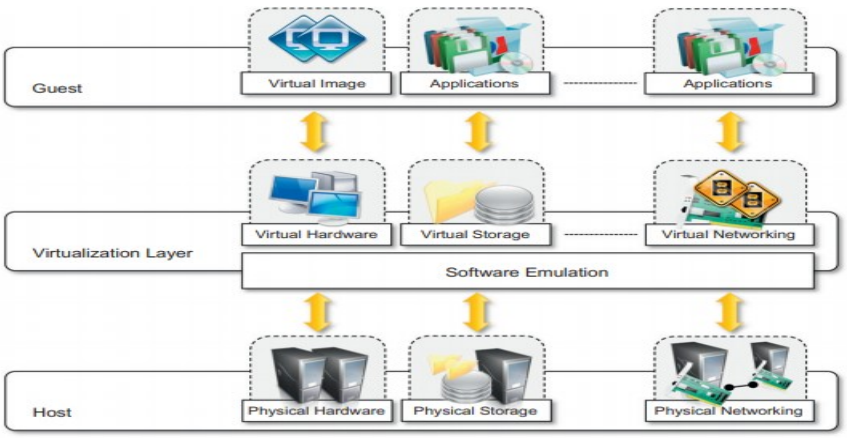


FIGURE 3.1
The virtualization reference model.

- Such a general abstraction finds different applications and then implementations of the virtualization technology.
- The most instinctive and popular is represented by **hardware virtualization**, which also constitutes the original realization of the virtualization concept.
- In the case of hardware virtualization, the guest is represented by a system image comprising an operating system and installed applications.
- These are installed on top of virtual hardware that is controlled and managed by the virtualization layer, also called the virtual machine manager(VMM).
- The host is instead represented by the physical hardware, and in some cases the operating system, that defines the environment where the virtual machine manager is running.
- In the case of virtual storage, the guest might be client applications or users that interact with the virtual storage management software deployed on top of the real storage system.
- The case of **virtual networking** is also similar.
  - The guest—applications and users—interacts with a virtual network, such as a virtual private network (VPN), which is managed by specific software (VPN client) using the physical network available on the node.
  - VPNs are useful for creating the illusion of being within a different physical network and thus accessing the resources in it, which would otherwise not be available.
- The technologies of today allow profitable use of virtualization and make it possible to fully exploit the advantages that come with it. Such advantages have always been characteristics of virtualized solutions.

Increased security:

- The ability to control the execution of a guest in a completely transparent manner opens new possibilities for delivering a secure, controlled execution environment.
  - The virtual machine represents an emulated environment in which the guest is executed.
  - All the operations of the guest are generally performed against the virtual machine, which then translates and applies them to the host.
  - This level of indirection allows the virtual machine manager to control and filter the activity of the guest, thus preventing some harmful operations from being performed.
  - Resources exposed by the host can then be hidden or simply protected from the guest.

- Moreover, sensitive information that is contained in the host can be naturally hidden without the need to install complex security policies.
- Increased security is a requirement when dealing with untrusted code.

Managed execution:

- Virtualization of the execution environment not only allows increased security, but a wider range of features also can be implemented.
- In particular, sharing, aggregation, emulation, and isolation are the most relevant features (see Figure 3.2).

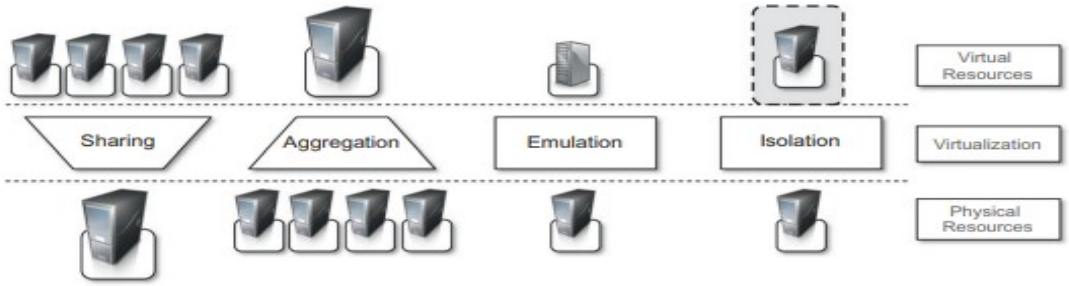


FIGURE 3.2
Functions enabled by managed execution.

- **Sharing:**
  - Virtualization allows the creation of a separate computing environments within the same host.
  - In this way it is possible to fully exploit the capabilities of a powerful guest, which would otherwise be underutilized.
  - Sharing is a particularly important feature in virtualized data centers, where this basic feature is used to reduce the number of active servers and limit power consumption.
- **Aggregation.**
  - Not only is it possible to share physical resource among several guests, but virtualization also allows aggregation, which is the opposite process.
  - A group of separate hosts can be tied together and represented to guests as a single virtual host.
- **Emulation.**
  - Guest programs are executed within an environment that is controlled by the virtualization layer, which ultimately is a program.
  - This allows for controlling and tuning the environment that is exposed to guests.
  - For instance, a completely different environment with respect to the host can be emulated, thus allowing the execution of

guest programs requiring specific characteristics that are not present in the physical host.

- **Isolation:**
  - Virtualization allows providing guests—whether they are operating systems, applications, or other entities—with a completely separate environment, in which they are executed.
  - The guest program performs its activity by interacting with an abstraction layer, which provides access to the underlying resources.
  - Isolation brings several benefits; for example, it allows multiple guests to run on the same host without interfering with each other.
  - Second, it provides a separation between the host and the guest.

Portability:

- The concept of portability applies in different ways according to the specific type of virtualization considered.
- In the case of a hardware virtualization solution, the guest is packaged into a virtual image that, in most cases, can be safely moved and executed on top of different virtual machines.
- Virtual images are generally proprietary formats that require a specific virtual machine manager to be executed.
- This makes the application development cycle more flexible and application deployment very straightforward: One version of the application, in most cases, is able to run on different platforms with no changes.
- Finally, portability allows having your own system always with you and ready to use as long as the required virtual machine manager is available.

Taxonomy of virtualization techniques:

- Virtualization covers a wide range of emulation techniques that are applied to different areas of computing.
- A classification of these techniques helps us better understand their characteristics and use (see Figure 3.3).
- The first classification distinguishes against the service or entity that is being emulated.
- Virtualization is mainly used to emulate execution environments, storage, and networks.
- In particular, we can divide these execution virtualization techniques into two major categories by considering the type of host they require.

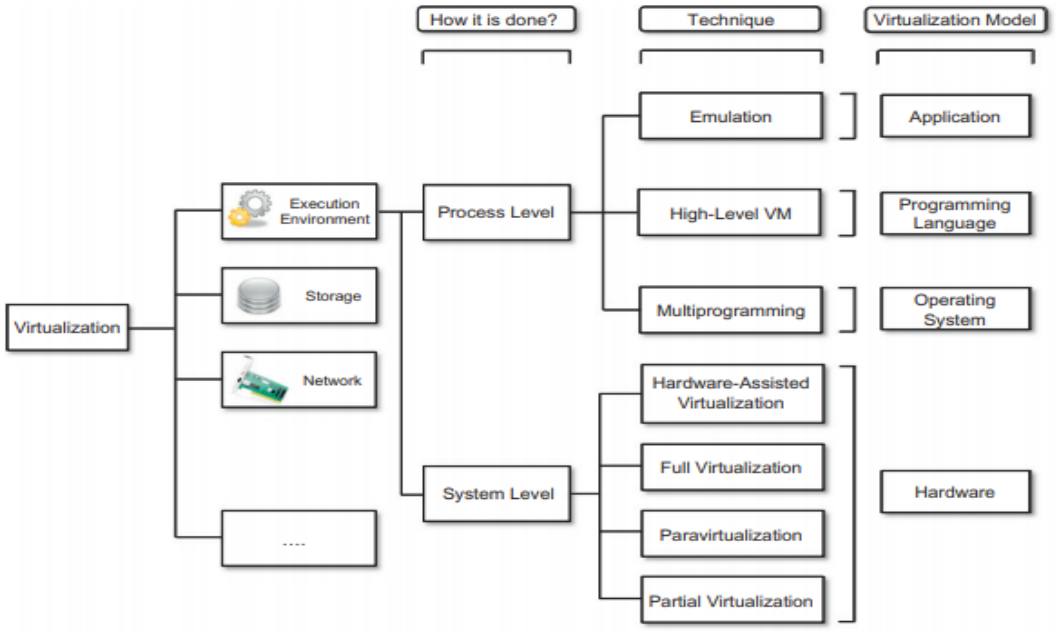


FIGURE 3.3 A taxonomy of virtualization techniques.

- **Process-level techniques** are implemented on top of an existing operating system, which has full control of the hardware.
- **System-level techniques** are implemented directly on hardware and do not require—or require a minimum of support from—an existing operating system.

- Within these two categories we can list various techniques that offer the guest a different type of virtual computation environment:
- plain hardware, operating system resources, low-level programming language, and application libraries.

Execution virtualization:

- Execution virtualization includes all techniques that aim to emulate an execution environment that is separate from the one hosting the virtualization layer.
- All these techniques concentrate their interest on providing support for the execution of programs, whether these are the operating system, a binary specification of a program compiled against an abstract machine model, or an application.
- Therefore, execution virtualization can be implemented directly on top of the hardware by the operating system, an application, or libraries dynamically or statically linked to an application image.

Machine reference model:

- Virtualizing an execution environment at different levels of the computing stack requires a reference model that defines the interfaces between the levels of abstractions, which hide implementation details.

- From this perspective, virtualization techniques actually replace one of the layers and intercept the calls that are directed toward it.
- Therefore, a clear separation between layers simplifies their implementation, which only requires the emulation of the interfaces and a proper interaction with the underlying layer.
- Modern computing systems can be expressed in terms of the reference model described in Figure 3.4.

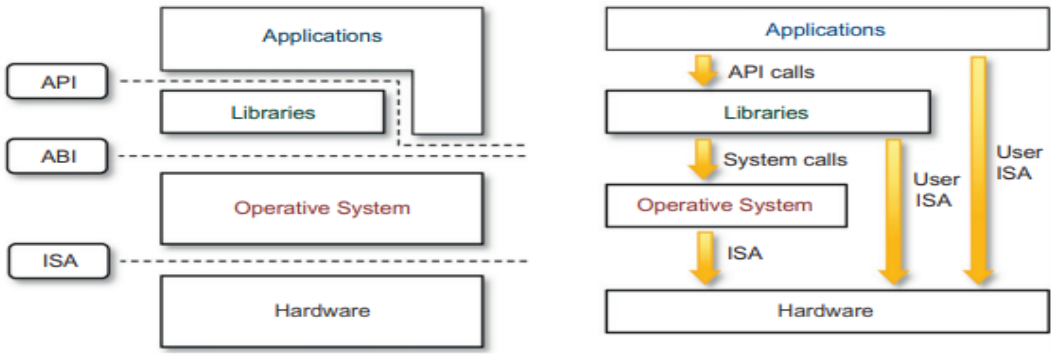


FIGURE 3.4 A machine reference model.

- At the bottom layer, the model for the hardware is expressed in terms of the Instruction Set Architecture (ISA),
  - It defines the instruction set for the processor, registers, memory, and interrupt management.
  - ISA is the interface between hardware and software, and it is important to the operating system (OS) developer (System ISA) and developers of applications that directly manage the underlying hardware (User ISA).
  - The application binary interface (ABI) separates the operating system layer from the applications and libraries, which are managed by the OS.
  - ABI covers details such as low-level data types, alignment, and call conventions and defines a format for executable programs.
  - System calls are defined at this level. This interface allows portability of applications and libraries across operating systems that implement the same ABI.
  - The highest level of abstraction is represented by the application programming interface (API), which interfaces applications to libraries and/or the underlying operating system.
- For any operation to be performed in the application level API, ABI and ISA are responsible for making it happen.
- The high-level abstraction is converted into machine-level instructions to perform the actual operations supported by the processor.

- The machine-level resources, such as processor registers and main memory capacities, are used to perform the operation at the hardware level of the central processing unit (CPU).
- This layered approach simplifies the development and implementation of computing systems and simplifies the implementation of multitasking and the coexistence of multiple executing environments.
- In fact, such a model not only requires limited knowledge of the entire computing stack, but it also provides ways to implement a minimal security model for managing and accessing shared resources.
- For this purpose, the instruction set exposed by the hardware has been divided into different security classes that define who can operate with them.
- The first distinction can be made between privileged and non-privileged instructions.
  - Non-privileged instructions are those instructions that can be used without interfering with other tasks because they do not access shared resources.
  - Privileged instructions are those that are executed under specific restrictions and are mostly used for sensitive operations, which expose (behavior-sensitive) or modify (control-sensitive) the privileged state.

Hardware-level virtualization:

- Hardware-level virtualization is a virtualization technique that provides an abstract execution environment in terms of computer hardware on top of which a guest operating system can be run.
- In this model, the guest is represented by the operating system, the host by the physical computer hardware, the virtual machine by its emulation, and the virtual machine manager by the hypervisor (see Figure 3.6).

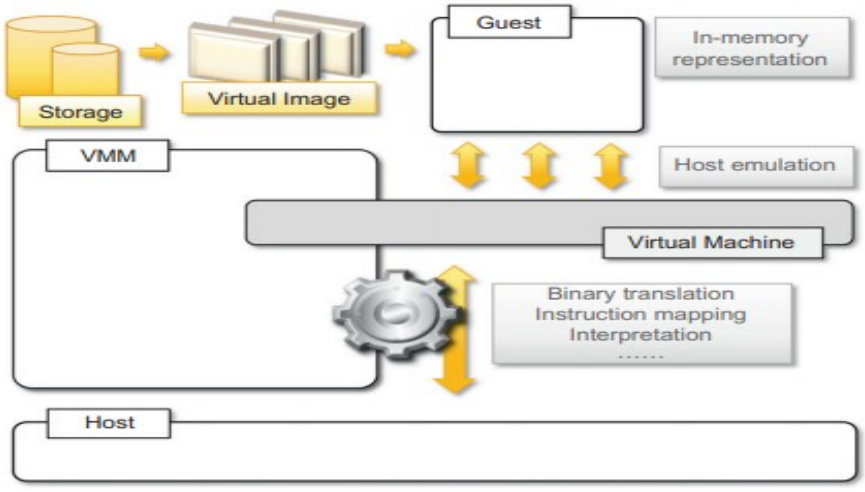


FIGURE 3.6 A hardware virtualization reference model.

- The hypervisor is generally a program or a combination of software and hardware that allows the abstraction of the underlying physical hardware.
- Hardware-level virtualization is also called system virtualization, since it provides ISA to virtual machines, which is the representation of the hardware interface of a system.

Hypervisors:

- A fundamental element of hardware virtualization is the hypervisor, or virtual machine manager (VMM). It recreates a hardware environment in which guest operating systems are installed.
- There are two major types of hypervisor: Type I and Type II (see Figure 3.7).

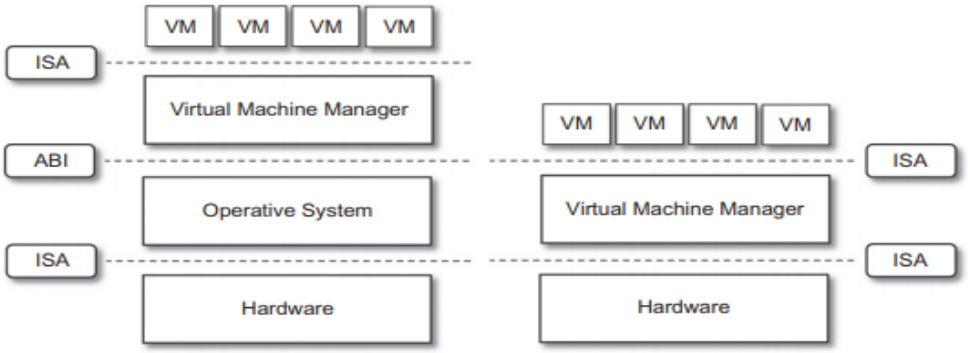


FIGURE 3.7

Hosted (left) and native (right) virtual machines. This figure provides a graphical representation of the two types of hypervisors.

  - Type I hypervisors run directly on top of the hardware.
    - Therefore, they take the place of the operating systems and interact directly with the ISA interface exposed by the underlying hardware, and they emulate this interface in order to allow the management of guest operating systems.
    - This type of hypervisor is also called a native virtual machine since it runs natively on hardware.
  - Type II hypervisors require the support of an operating system to provide virtualization services.
    - This means that they are programs managed by the operating system, which interact with it through the ABI and emulate the ISA of virtual hardware for guest operating systems.
    - This type of hypervisor is also called a hosted virtual machine since it is hosted within an operating system.
  - Conceptually, a virtual machine manager is internally organized as described in Figure 3.8.
- Three main modules dispatcher, allocator, and interpreter, coordinate their activity in order to emulate the underlying hardware.

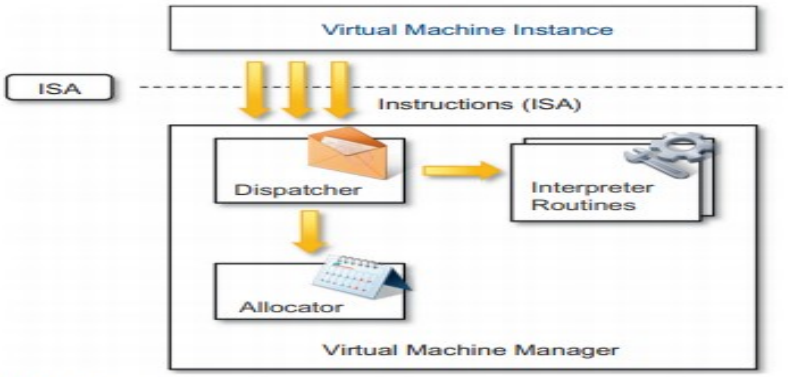


FIGURE 3.8 A hypervisor reference architecture.

- The dispatcher constitutes the entry point of the monitor and reroutes the instructions issued by the virtual machine instance to one of the two other modules.
- The allocator is responsible for deciding the system resources to be provided to the VM: whenever a virtual machine tries to execute an instruction that results in changing the machine resources associated with that VM, the allocator is invoked by the dispatcher.
- The interpreter module consists of interpreter routines. These are executed whenever a virtual machine executes a privileged instruction: a trap is triggered and the corresponding routine is executed.

- The design and architecture of a virtual machine manager, together with the underlying hardware design of the host machine, determine the full realization of hardware virtualization.
- The criteria that need to be met by a virtual machine manager to efficiently support virtualization were established by Goldberg and Popek in 1974 [23]. Three properties have to be satisfied……
  - **Equivalence.** A guest running under the control of a virtual machine manager should exhibit the same behavior as when it is executed directly on the physical host.
  - **Resource control**. The virtual machine manager should be in complete control of virtualized resources.
  - **Efficiency.** A statistically dominant fraction of the machine instructions should be executed without intervention from the virtual machine manager.
- The major factor that determines whether these properties are satisfied is represented by the layout of the ISA of the host running a virtual machine manager.

Hardware virtualization techniques:

1. Hardware-assisted virtualization.

- This term refers to a scenario in which the hardware provides architectural support for building a virtual machine manager able to run a guest operating system in complete isolation.

- This technique was originally introduced in the IBM System/370.

2. Full virtualization.

- Full virtualization refers to the ability to run a program, most likely an operating system, directly on top of a virtual machine and without any modification, as though it were run on the raw hardware.
- To make this possible, virtual machine manager's are required to provide a complete emulation of the entire underlying hardware.
- The principal advantage of full virtualization is complete isolation, which leads to enhanced security, ease of emulation of different architectures, and coexistence of different systems on the same platform.
- A successful and efficient implementation of full virtualization is obtained with a combination of hardware and software, not allowing potentially harmful instructions to be executed directly on the host.

3. Paravirtualization.

- This is a not-transparent virtualization solution that allows implementing thin virtual machine managers.
- Paravirtualization techniques expose a software interface to the virtual machine that is slightly modified from the host and, as a consequence, guests need to be modified.
- The aim is to provide the capability to demand the execution of performance-critical operations.

4. Partial virtualization.

- Partial virtualization provides a partial emulation of the underlying hardware, thus not allowing the complete execution of the guest operating system in complete isolation.
- Partial virtualization allows many applications to run transparently, but not all the features of the operating system can be supported, as happens with full virtualization.
- An example of partial virtualization is address space virtualization used in time-sharing systems.

Operating system-level virtualization:

- Operating system-level virtualization offers the opportunity to create different and separated execution environments for applications that are managed concurrently.
- Differently from hardware virtualization, there is no virtual machine manager or hypervisor, and the virtualization is done within a single operating system, where the OS kernel allows for multiple isolated user space instances.

- The kernel is also responsible for sharing the system resources among instances and for limiting the impact of instances on each other.
- A user space instance in general contains a proper view of the file system, which is completely isolated, and separate IP addresses, software configurations, and access to devices.
- It is an efficient solution for server consolidation scenarios in which multiple application servers share the same technology: operating system, application server framework, and other components.
- Examples of operating system-level virtualizations are FreeBSD Jails, IBM Logical Partition (LPAR).

Programming language-level virtualization:

- Programming language-level virtualization is mostly used to achieve ease of deployment of applications, managed execution, and portability across different platforms and operating systems.
- It consists of a virtual machine executing the byte code of a program, which is the result of the compilation process.
- Generally, these virtual machines constitute a simplification of the underlying hardware instruction set and provide some high-level instructions that map some of the features of the languages compiled for them.
- At runtime, the byte code can be either interpreted or compiled on the fly against the underlying hardware instruction set.
- The main advantage of programming-level virtual machines, also called process virtual machines, is the ability to provide a uniform execution environment across different platforms through bytecodes.
- From a development lifecycle point of view, this simplifies the development and deployment efforts since it is not necessary to provide different versions of the same code.
- The implementation of the virtual machine for different platforms is still a costly task, but it is done once and not for any application.
- Moreover, process virtual machines allow for more control over the execution of programs since they do not provide direct access to the memory.
- Security is another advantage of managed programming languages; by filtering the I/O operations, the process virtual machine can easily support sandboxing of applications.

Application-level virtualization:

- Application-level virtualization is a technique allowing applications to be run in runtime environments that do not natively support all the features required by such applications.

- In this scenario, applications are not installed in the expected runtime environment but are run as though they were.
- In general, these techniques are mostly concerned with partial file systems, libraries, and operating system component emulation.
  - Such emulation is performed by a thin layer—a program or an operating system component—that is in charge of executing the application.
- Emulation can also be used to execute program binaries compiled for different hardware architectures. In this case, one of the following strategies can be implemented……
  - Interpretation. In this technique every source instruction is interpreted by an emulator for executing native ISA instructions, leading to poor performance.
  - Binary translation. Here every source instruction is converted to native instructions with equivalent functions. After a block of instructions is translated, it is cached and reused.
- Application virtualization is a good solution in the case of missing libraries in the host operating system; in this case a replacement library can be linked with the application, or library calls can be remapped to existing functions available in the host system.
- Another advantage is that in this case the virtual machine manager is much lighter since it provides a partial emulation of the runtime environment compared to hardware virtualization.
- Moreover, this technique allows incompatible applications to run together.

Other types of Virtualization:

- Other than execution virtualization, other types of virtualization provide an abstract environment to interact with. These mainly cover storage, networking, and client/server interaction.
- **Storage virtualization:**
  - Storage virtualization is a system administration practice that allows decoupling the physical organization of the hardware from its logical representation.
  - Using this technique, users do not have to be worried about the specific location of their data, which can be identified using a logical path.
  - It allows us to bind a wide range of storage facilities and represent them under a single logical file system.
  - There are different techniques for storage virtualization, one of the most popular being network-based virtualization by means of storage area networks (SANs).

- SANs use a network-accessible device through a large bandwidth connection to provide storage facilities.

- **Network virtualization:**
  - Network virtualization combines hardware appliances and specific software for the creation and management of a virtual network.
  - Network virtualization can aggregate different physical networks into a single logical network (External) or provide network-like functionality to an operating system partition (Internal).
  - The result of external network virtualization is generally a virtual LAN (VLAN).
  - Internal network virtualization is generally applied together with hardware and operating system-level virtualization, in which the guests obtain a virtual network interface to communicate with.
  - There are several options for implementing internal network virtualization:
  - The guest can share the same network interface of the host and use Network Address Translation (NAT) to access the network;
  - The virtual machine manager can emulate, and install on the host, an additional network device, together with the driver; or
  - The guest can have a private network only with the guest.
- **Desktop virtualization:**
  - Desktop virtualization abstracts the desktop environment available on a personal computer in order to provide access to it using a client/server approach.
  - Desktop virtualization provides the same outcome of hardware virtualization but serves a different purpose.
  - Similarly, to hardware virtualization, desktop virtualization makes accessible a different system as though it were natively installed on the host, but this system is remotely stored on a different host and accessed through a network connection.
  - Moreover, desktop virtualization addresses the problem of making the same desktop environment accessible from everywhere.
  - Generally, the desktop environment is stored in a remote server or a data center that provides a high-availability infrastructure and ensures the accessibility and persistence of the data. In this scenario,
    - An infrastructure supporting hardware virtualization is fundamental to provide access to multiple desktop environments hosted on the same server;

    - A specific desktop environment is stored in a virtual machine image that is loaded and started on demand when a client connects to the desktop environment.
  - The advantages of desktop virtualization are high availability, persistence, accessibility, and ease of management.
  - Desktop environment are implemented in software components such as Windows Remote Services, VNC, and X Server.
- **Application server virtualization:**
  - Application server virtualization abstracts a collection of application servers that provide the same services as a single virtual application server by using load-balancing strategies and providing a high-availability infrastructure for the services hosted in the application server.
  - This is a particular form of virtualization and serves the same purpose of storage virtualization: providing a better quality of service rather than emulating a different environment.

Virtualization and cloud computing:

- Virtualization plays an important role in cloud computing since it allows for the appropriate degree of customization, security, isolation, and manageability that are fundamental for delivering IT services on demand.
- Virtualization technologies are primarily used to offer configurable computing environments and storage.
- Particularly important is the role of virtual computing environment and execution virtualization techniques.
- Among these, hardware and programming language virtualization are the techniques adopted in cloud computing systems.
  - Hardware virtualization is an enabling factor for solutions in the Infrastructure-as-a-Service (IaaS) market segment,
  - while programming language virtualization is a technology leveraged in Platform-as-a-Service (PaaS) offerings.
- In both cases, the capability of offering a customizable and sandboxed environment constituted an attractive business opportunity for companies featuring a large computing infrastructure that was able to sustain and process huge workloads.
- Moreover, virtualization also allows isolation and a finer control, thus simplifying the leasing of services and their accountability on the vendor side.
- Besides computation on demand, virtualization also gives the opportunity to design more efficient computing systems by means of consolidation.

- Since virtualization allows us to create isolated and controllable environments, it is possible to serve these environments with the same resource without them interfering with each other.
- If the underlying resources are capable enough, there will be no evidence of such sharing. This opportunity is particularly attractive when resources are underutilized,
- Because it allows reducing the number of active resources by aggregating virtual machines over a smaller number of resources that become fully utilized.
- This practice is also known as server consolidation, while the movement of virtual machine instances is called virtual machine migration (see Figure 3.10).
- Storage virtualization constitutes an interesting opportunity given by virtualization technologies, often complementary to the execution of virtualization.
- Even in this case, vendors backed by large computing infrastructures featuring huge storage facilities can attach these facilities into a virtual storage service, easily partitionable into slices, which can be dynamic and offered as a service.

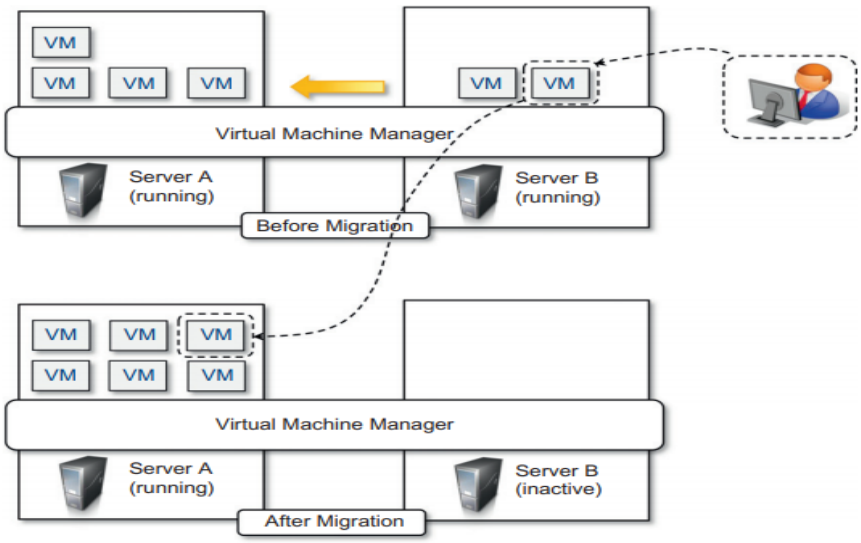


FIGURE 3.10
Live migration and server consolidation.

- Finally, cloud computing revamps the concept of desktop virtualization, initially introduced in the mainframe era.

Pros and cons of virtualization:

- Virtualization has now become extremely popular and widely used, especially in cloud computing.
- The primary reason for its wide success is the elimination of technology barriers that prevented virtualization from being an effective and viable solution in the past.

- The advancements in computing technology have made virtualization an interesting opportunity to deliver on-demand IT infrastructure and services.

- **Advantages of virtualization:**

1. Managed execution and isolation are perhaps the most important advantages of virtualization.
   - These two characteristics allow building secure and controllable computing environments.
   - A virtual execution environment can be configured as a sandbox, thus preventing any harmful operation to cross the borders of the virtual host.
   - Moreover, allocation of resources and their partitioning among different guests is simplified, being the virtual host controlled by a program.
   - This enables fine-tuning of resources, which is very important in a server consolidation scenario and is also a requirement for effective quality of service.
2. Portability is another advantage of virtualization, especially for execution virtualization techniques.
   - Virtual machine instances are normally represented by one or more files that can be easily transported with respect to physical systems.
   - Moreover, they also tend to be self-contained since they do not have other dependencies besides the virtual machine manager for their use.
   - It is in fact possible to build our own operating environment within a virtual machine instance and bring it with us wherever we go, as though we had our own laptop.
   - This concept is also an enabler for migration techniques in a server consolidation scenario.
   - Portability and self-containment also contribute to reducing the costs of maintenance, since the number of hosts is expected to be lower than the number of virtual machine instances.
3. Finally, by means of virtualization it is possible to achieve a more efficient use of resources.
   - Multiple systems can securely coexist and share the resources of the underlying host, without interfering with each other.
   - It allows adjusting the number of active physical resources dynamically according to the current load of the system.

- **Disadvantages:**

- The most evident is represented by a performance decrease of guest systems as a result of the intermediation performed by the virtualization layer.

- In addition, suboptimal use of the host because of the abstraction layer introduced by virtualization management software can lead to a very inefficient utilization of the host or a degraded user experience.
- Less evident, but perhaps more dangerous, are the implications for security, which are mostly due to the ability to emulate a different execution environment.

1. Performance degradation:

- Performance is definitely one of the major concerns in using virtualization technology.
- Since virtualization interposes an abstraction layer between the guest and the host, the guest can experience increased latencies.
- For instance, in the case of hardware virtualization, where the intermediate emulates a bare machine on top of which an entire system can be installed, the causes of performance degradation can be traced back to the overhead introduced by the following activities:
  - Maintaining the status of virtual processors.
  - Support of privileged instructions (trap and simulate privileged instructions).
  - Support of paging within VM.
  - Console functions.
- These concerns are becoming less and less important thanks to technology advancements and the ever-increasing computational power available today.

2. Inefficiency and degraded user experience:

- Virtualization can sometime lead to an inefficient use of the host.
- In particular, some of the specific features of the host cannot be exposed by the abstraction layer and then become inaccessible.
- In the case of hardware virtualization, this could happen for device drivers: The virtual machine can sometime simply provide a default graphic card that maps only a subset of the features available in the host.

3. Security holes and new threats:

- Virtualization opens the door to a new and unexpected form of phishing.
- The capability of emulating a host in a completely transparent manner led the way to malicious programs that are designed to extract sensitive information from the guest.

- In the case of hardware virtualization, malicious programs can preload themselves before the operating system and act as a thin virtual machine manager toward it.
- The operating system is then controlled and can be manipulated to extract sensitive information of interest to third parties.
- The same considerations can be made for programming-level virtual machines:
- Modified versions of the runtime environment can access sensitive information or monitor the memory locations utilized by guest applications while these are executed.

References:

- Cloud Computing - Concepts, Technology, Security, and Architecture - Second Edition - Thomas Erl, Eric Barcelo - Pearson
- Mastering Cloud Computing: Foundations and Applications Programming – Rajkumar Buyya – Tata Mcgraw Hill publishing.

****