

Cloud Computing – CSE4001

Course Instructor: Dr. Arunkumar Gopu

Senior Assistant Professor – Grade I

School of Computer Science and Engineering (SCOPE)

VIT – AP University, Amaravati, Andhra Pradesh.

Email-id: arunkumar.gopu@vitap.ac.in

Course Objectives

1. Understanding of cloud computing and a systematic knowledge of the fundamental technologies, architecture, and security.

2. To expose frontier areas of Cloud Computing while providing sufficient foundations to enable further study and research.

Expected Outcome

1. Articulate the main concepts, key technologies, strengths, and limitations of cloud computing and the possible applications for state-of-the-art cloud computing
2. Identify the architecture and infrastructure of cloud computing, including SaaS, PaaS, IaaS, public cloud, private cloud, hybrid cloud.
3. Explain the core issues of cloud computing such as security, privacy, and interoperability.
4. Choose the appropriate technologies, algorithms, and approaches for the related issues.
5. Identify problems, and explain, analyze, and evaluate various cloud computing solutions.

#	Modules and Course Content	Hours
1	Understanding Cloud Computing - Cloud origins and influences, basic concepts and terminology, goals and benefits, risks and challenges. Fundamental Concepts and Models: Roles and boundaries, cloud characteristics, cloud delivery models, cloud deployment models.	6
2	Cloud Enabling Technology - Data center technology, virtualization technology, web technology, multitenant technology, service technology.	6
3	Cloud Infrastructure Mechanisms - Network perimeter, virtual server, cloud storage device, cloud usage monitor, resource replication.	6
4	Fundamental Cloud Architectures - Workload distribution architecture, resource pooling architecture, dynamic scalability architecture, elastic resource capacity architecture, service load balancing architecture, cloud bursting architecture, elastic disk provisioning architecture, redundant storage architecture, Cloud operations: Migration, static and dynamic Scheduling	10
5	Cloud Delivery Model Considerations - Cloud Delivery Model Considerations: The cloud provider perspective- Building IaaS environments, equipping PaaS environments, optimizing SaaS environments, the cloud consumer perspective, working with IaaS environments, working with PaaS environments, working with SaaS services.	8
6	Fundamental Cloud Security and Mechanisms - Basic terms and concepts, Threat agents, Cloud security threats, Encryption, Hashing, Digital Signature, Public Key Infrastructure(PKI), Identity and Access Management(IAM), Single Sign-On(SSO), Cloud Based Security Groups, Handled Virtual Server Machines	9

Textbooks and Reference Books

1. Thomas Erl, Ricardo Puttini, Zaigham Mahmood, “Cloud Computing: Concepts, Technology & Architecture”, PHI Publications, 2013.

1. John W. Rittinghouse, James F.Ransome, “Cloud Computing: Implementation, Management and Security”, CRC Press, 2017.
2. Sandeep Bhowmik, “Cloud Computing” , Cambridge University Press, publishers 2017.
3. Meikang Qiu, Keke Gai, “Mobile Cloud computing : models, implementations and security ”, CRC Press, 2017.



Mode of Evaluation

Evaluation Components	Weightage
Continuous Assessment Test-1	20%
Continuous Assessment Test-2	20%
Final Assessment Test	20%
Practical Assessment (Mini Project)	25%
Digital Assignment	15%

Module 1

Cloud origins and influences, basic concepts and terminology, goals and benefits, risks and challenges. Fundamental Concepts and Models: Roles and boundaries, cloud characteristics, cloud delivery models, cloud deployment models.

Cloud Origin and Influences

- The idea of computing in a “cloud” traces back to the origins of utility computing.
- 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. ... The computer utility could become the basis of a new and important industry.”

Cloud Origin and Influences

- 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’ ...”.
- In the late 1990s, Salesforce.com pioneered the notion of bringing remotely provisioned services into the enterprise.
- In 2002, 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.

Cloud Origin and Influences

- “Network Cloud” or “Cloud” was introduced in the early 1990s.
- The transmission of data from one end-point (local network) to the “Cloud” (wide area network).
- It wasn’t until 2006 that the term “cloud computing” emerged in the commercial arena.
- Amazon launched its Elastic Compute Cloud (EC2) and Google Apps also began providing browser-based enterprise applications.

Definitions

- Garter - *a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.”*
- Forrester Research - *a standardized IT capability (services, software, or infrastructure) delivered via Internet technologies in a pay-per-use, self-service way.*

Definitions

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

Business Drivers

- Business Drivers – Why we need Cloud Computing?
- **Capacity planning** is the process of determining and fulfilling future demands of an organization's IT resources, products, and services.
- Over-provisioning or under-provisioning.
- 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.

Business Drivers

- **Cost Reduction**
- Two costs need to be accounted for:
 - The cost of acquiring new infrastructure
 - The cost of its ongoing ownership.
- Operational overhead represents a considerable share of IT budgets, often exceeding up-front investment costs.

Business Drivers

- 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

Business Drivers

- **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.
- Changing business needs and priorities may require IT resources to be more available and reliable than before.
- This inability to respond can inhibit an organization from keeping up with market demands, competitive pressures, and its own strategic business goals.

Technology Innovations

- Pre-existing technologies considered to be the primary influences on cloud computing.
- **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.
- 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.

Technology Innovations

- **Grid Computing** - A computing 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.
- Common feature-sets such as networked access, resource pooling, and scalability and resiliency are also an enabling factor in cloud computing.

Technology Innovations

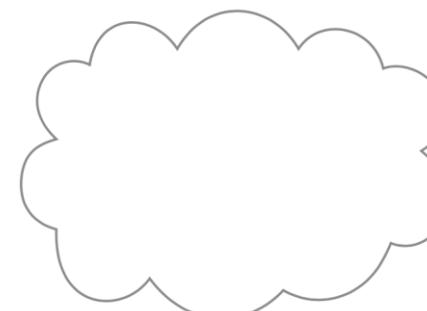
- **Virtualization** - Virtualization represents a technology platform used for the creation of virtual instances of IT resources.
- A layer of virtualization software allows physical IT resources to provide multiple virtual images of themselves so that their underlying processing capabilities can be shared by multiple users.
- The virtualization process severs this software-hardware dependency, as hardware requirements can be simulated by emulation software running in virtualized environments.

Technology Innovations vs. Enabling Technologies

- Broadband Networks and Internet Architecture
- Data Centre Technology
- (Modern)Virtualization Technology
- Web Technology
- Multitenant Technology
- Service Technology

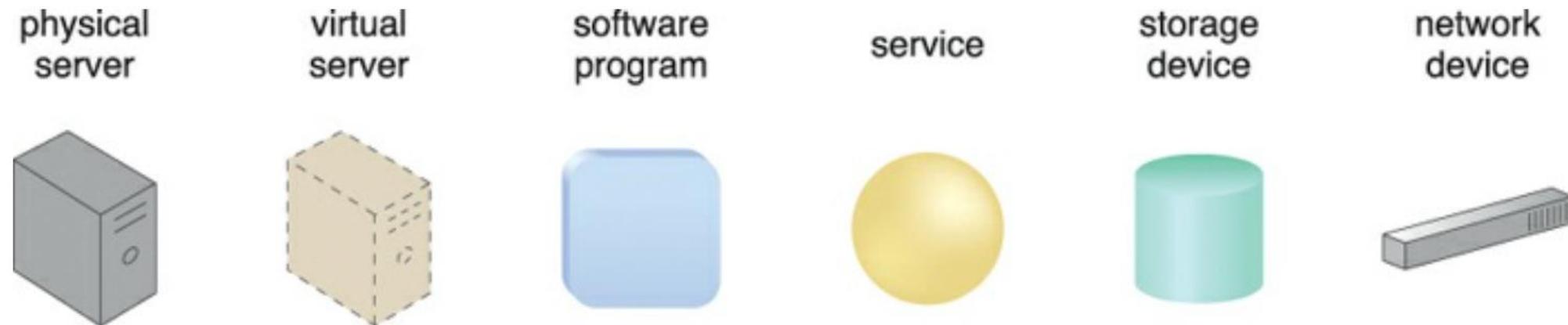
Basic Concepts and Terminology

- Cloud refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources.
- This same symbol is now used to specifically represent the boundary of a cloud environment.
- A cloud can be based on the use of any protocols that allow for the remote access to its IT resources.

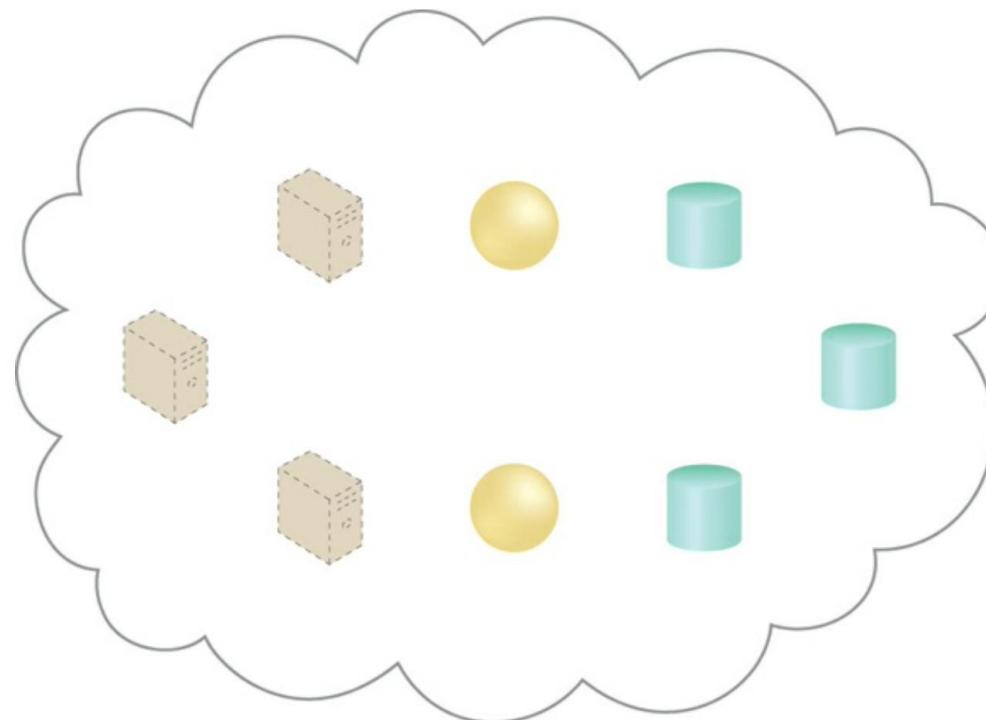


IT Resources

- An **IT resource** is a physical or virtual IT-related artifact
 - software-based, such as a virtual server or a custom software program.
 - hardware-based, such as a physical server or a network device.



Cloud-based IT resources



Cloud-based IT resources

On Premise

- An IT resource that is hosted in a conventional IT enterprise within an organizational boundary
- 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.

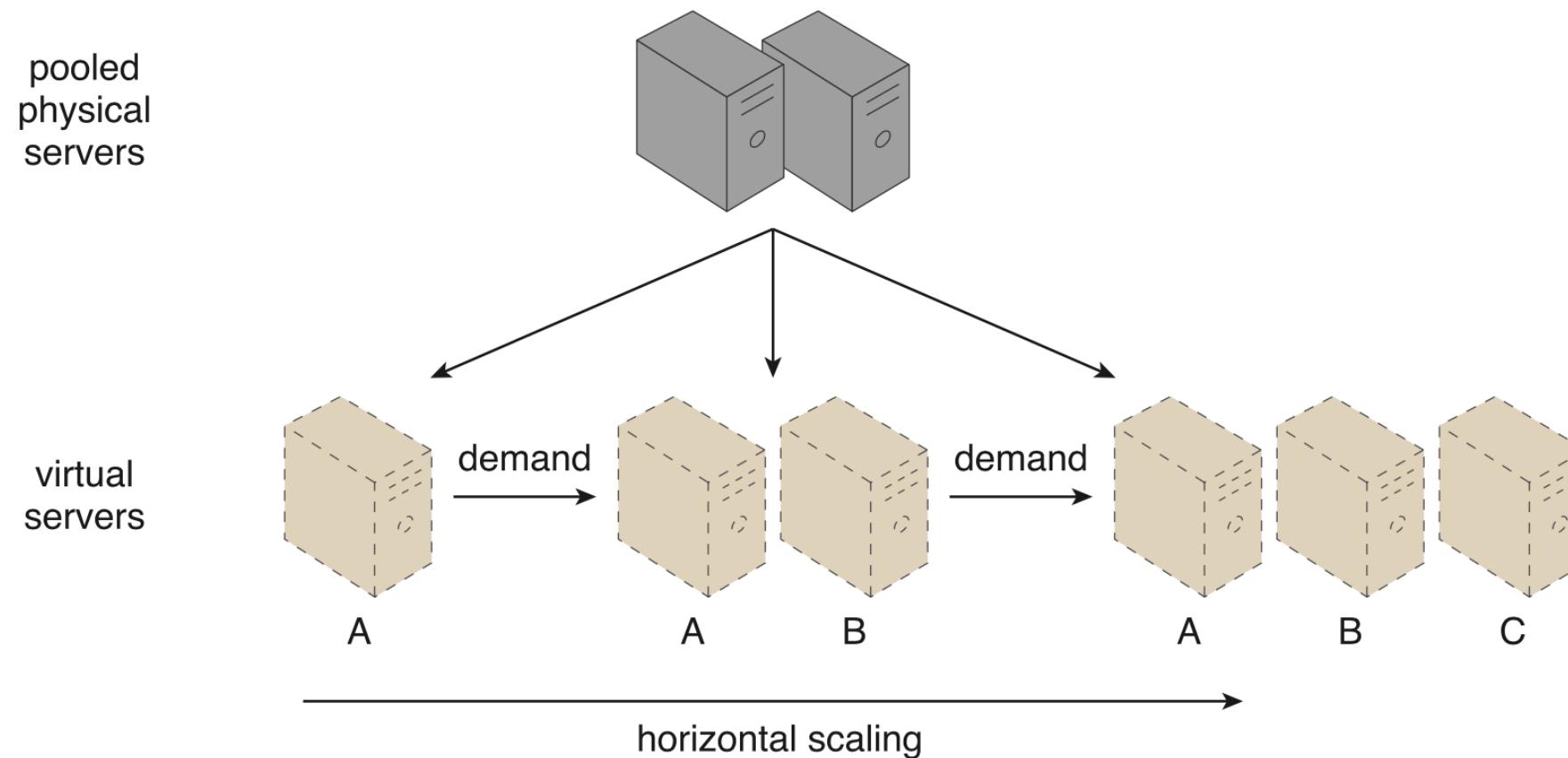
Cloud Consumers and Cloud Providers

- The party that provides cloud-based IT resources is the cloud provider.
- The party that uses cloud-based IT resources is the cloud consumer.

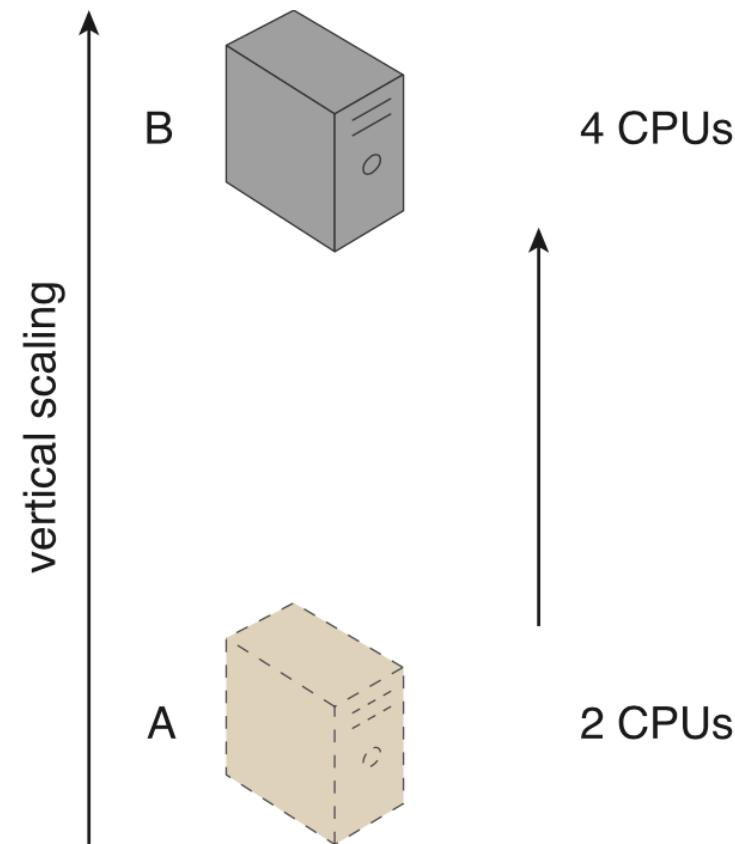
Scaling

- 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



Vertical Scaling

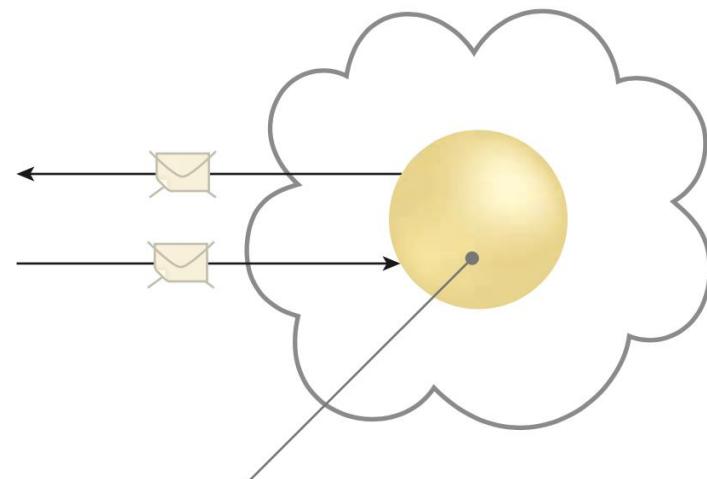


Horizontal vs. Vertical Scaling

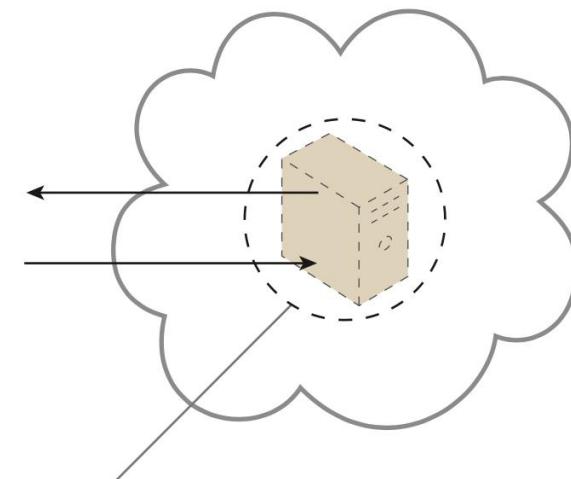
Horizontal Scaling	Vertical Scaling
less expensive (through commodity hardware components)	more expensive (specialized servers)
IT resources instantly available	IT resources normally instantly available
resource replication and automated scaling	additional setup is normally needed
additional IT resources needed	no additional IT resources needed
not limited by hardware capacity	limited by maximum hardware capacity

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.
- A cloud service can be anything.



remotely accessed Web service
acting as a cloud service

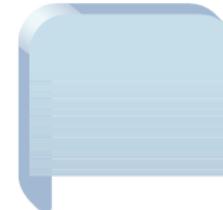


remotely accessed virtual server
acting as a cloud service

Cloud Service Consumer

- The cloud service consumer is a temporary runtime role assumed by a software program when it accesses a cloud service.

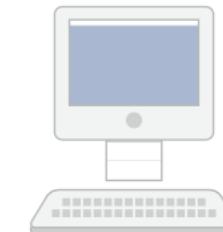
software
program



service



workstation



laptop



mobile
device



Goals and Benefits

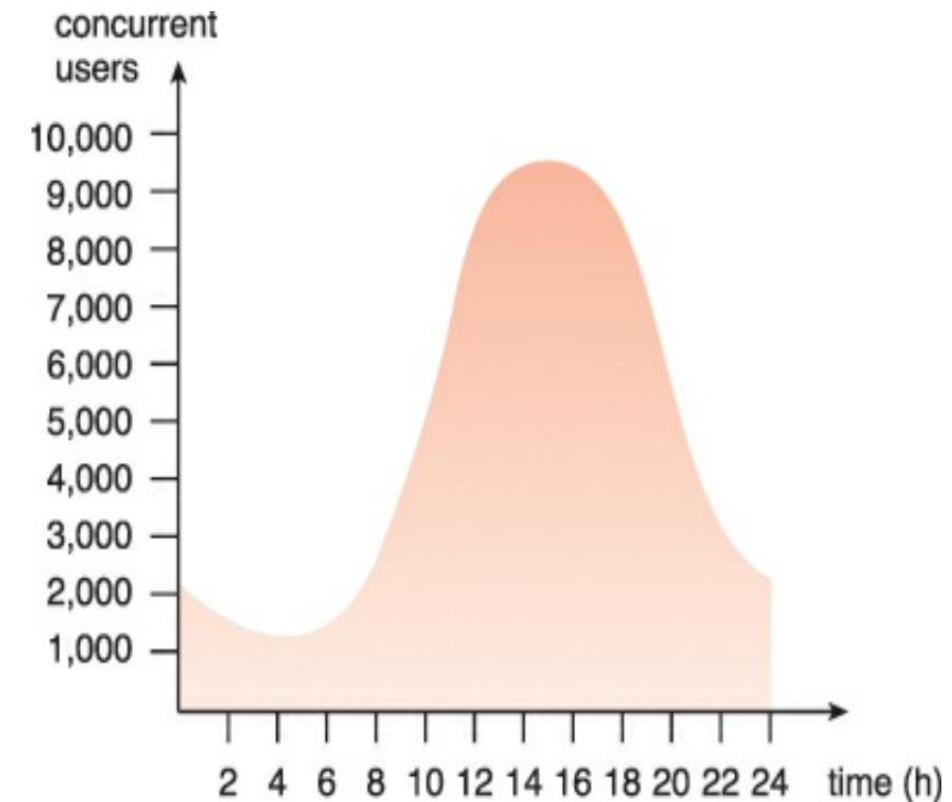
- **Reduced Investments and Proportional Costs**
- Elimination of up-front IT investments, namely hardware and software purchases and ownership costs.
- Measured operational expenditures.
- This elimination or minimization of up-front financial commitments allows enterprises to focus on business objectives.
- Cloud data centers are commonly located in destinations where real estate, IT professionals, and network bandwidth can be obtained at lower costs, resulting in both capital and operational savings.

Goals and Benefits

- Pooled IT resources are made available to and shared by multiple cloud consumers, resulting in increased or even maximum possible utilization.
 - On-demand access to pay-as-you-go computing resources.
 - Unlimited computing resources are available on demand.
 - The ability to add or remove IT resources at a fine-grained level.
 - Applications are not locked into devices or locations and can be easily moved if needed.
- Using 100 servers for one hour costs the same as using one server for 100 hours.

Goals and Benefits

- **Increased Scalability**
- Clouds can instantly and dynamically allocate IT resources to cloud consumers.
- IT resources to always meet and fulfill unpredictable usage demands avoids potential loss of business that can occur when usage thresholds are met.



Goals and Benefits

- **Increased Availability and Reliability**
- IT resource unable to respond to customer requests, its unexpected failure can decrease overall customer confidence.
- Cloud-based IT resource to minimize or even eliminate outages, and for increasing its reliability to minimize the impact of run-time failure conditions.
- Cloud providers generally offer “resilient” IT resources for which they can guarantee high levels of availability.

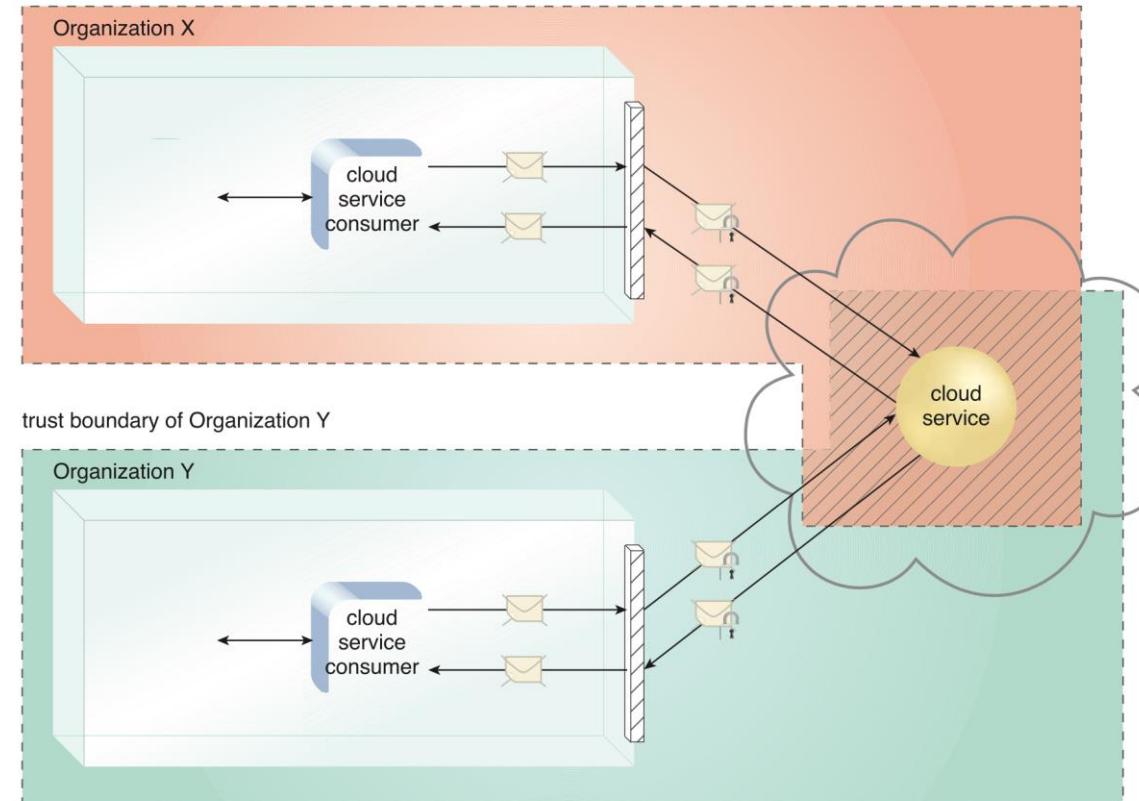
Goals and Benefits

- The modular architecture of cloud environments provides extensive failover support that increases reliability.
- It is important that organizations carefully examine the SLAs offered by cloud providers when considering the leasing of cloud-based services and IT resources.

Risks and Challenges

- **Increased Security Vulnerabilities**
- The moving of business data to the cloud means that the responsibility over data security becomes shared with the cloud provider.
- Another consequence of overlapping trust boundaries relates to the cloud provider's privileged access to cloud consumer data.
- 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.

Risks and Challenges

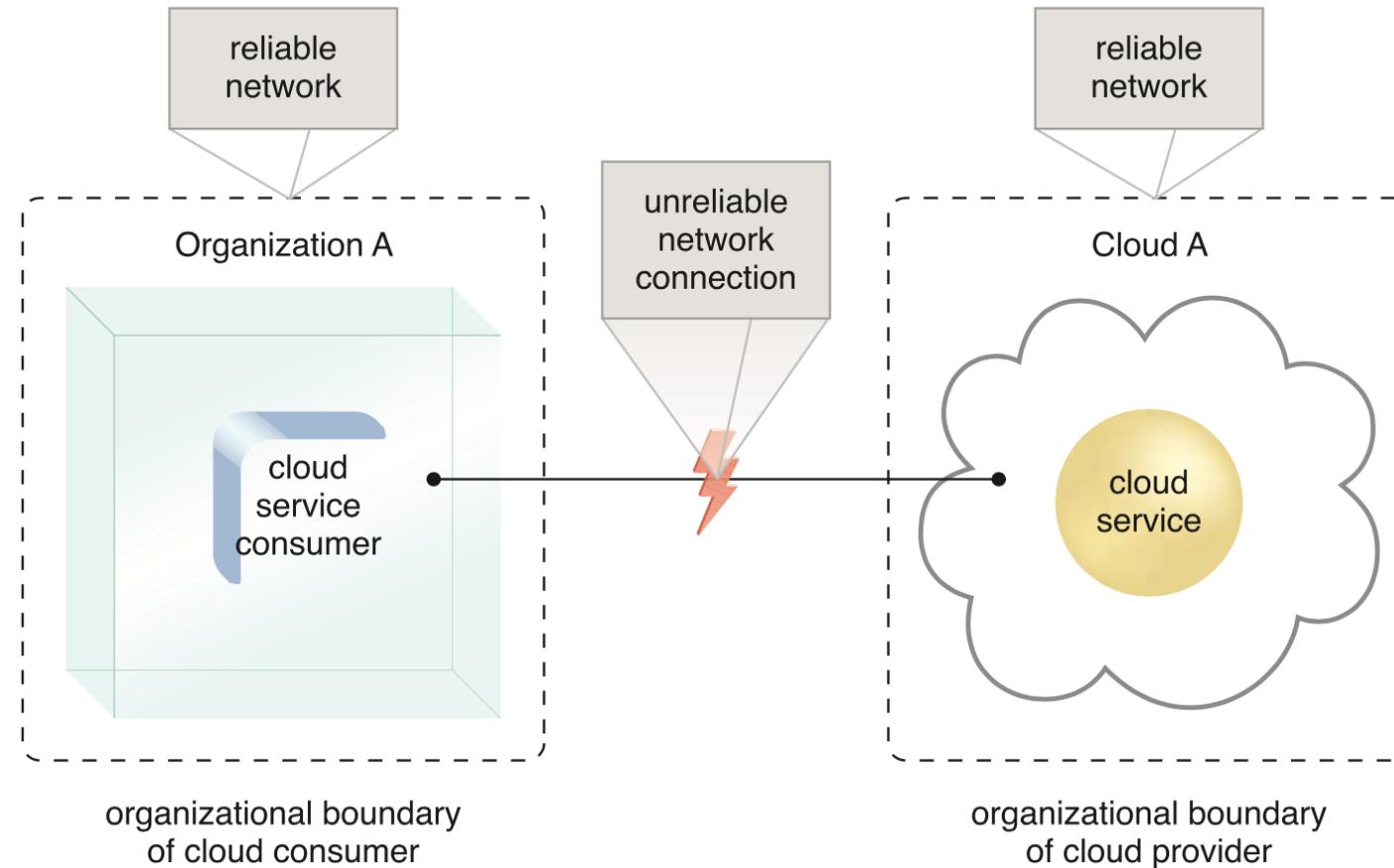


Overlapping trust boundaries

Risks and Challenges

- **Reduced Operational Governance Control**
- Cloud consumers are usually allotted a level of governance control that is lower than that over on-premise IT resources.
- An unreliable cloud provider may not maintain the guarantees it makes in the SLAs that were published for its cloud services.
- This can jeopardize the quality of the cloud consumer solutions that rely on these cloud services.
- Longer geographic distances between the cloud consumer and cloud provider can require additional network latency.

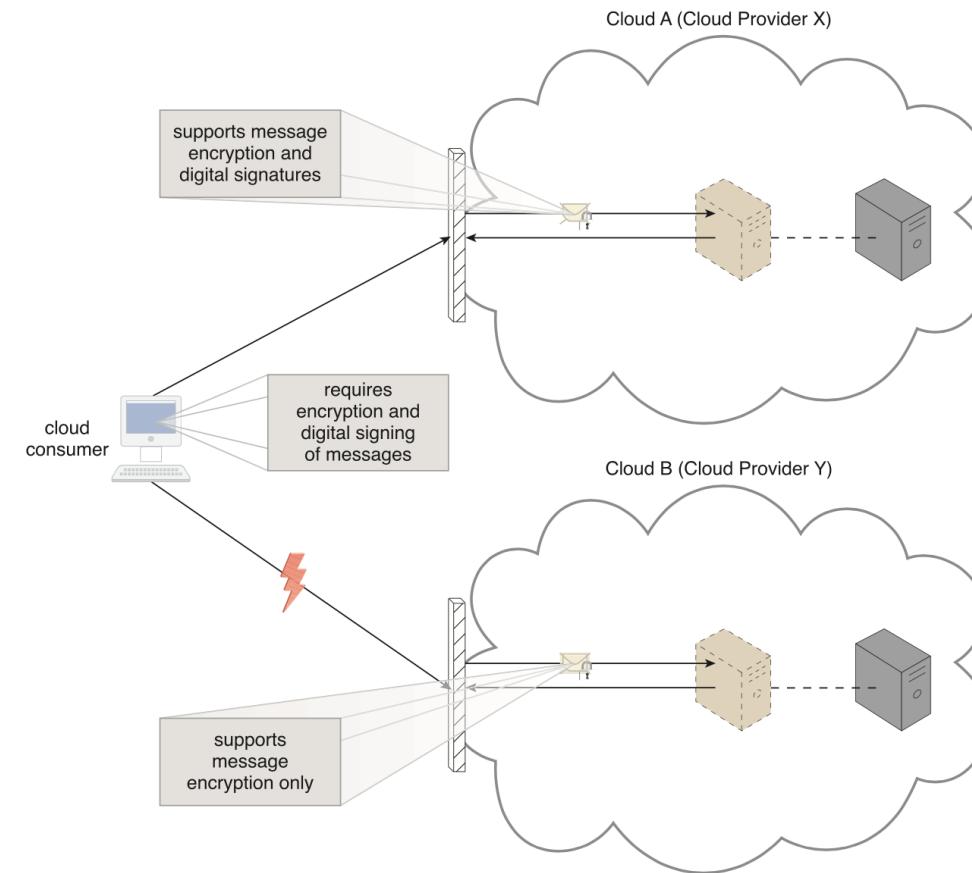
Risks and Challenges



Risks and Challenges

- **Limited Portability Between Cloud Providers**
- Due to a lack of industry standards in 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

Risks and Challenges



Risks and Challenges

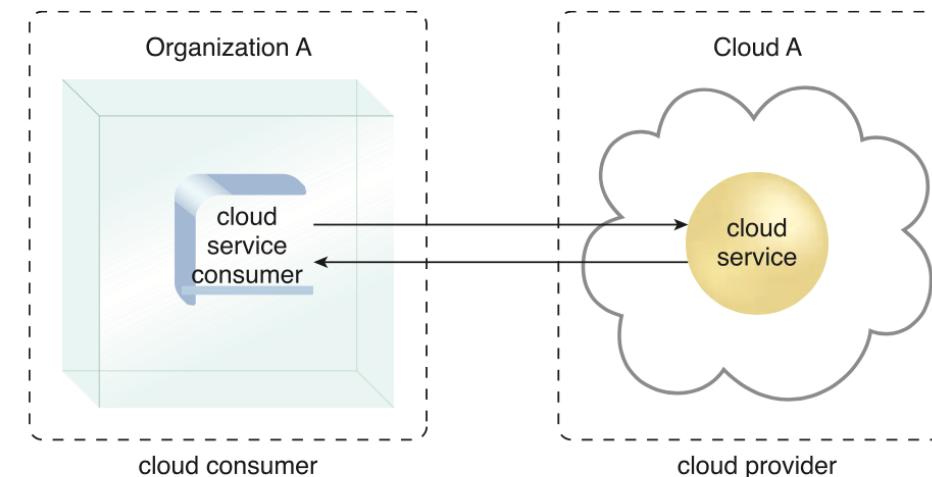
- **Multi-Regional Compliance and Legal Issues**
- Cloud providers will frequently establish data centers in affordable or convenient geographical locations.
- Cloud consumers will often not be aware of the physical location of their IT resources and data when hosted by public clouds.
- Countries have laws that require some types of data to be disclosed to certain government agencies or to the subject of the data.

Roles and Boundaries

- Organizations and humans can play different roles based on how they relate to and/or interact with a cloud and its hosted IT resources.
- **Cloud Provider**
- The organization that provides cloud-based IT resources is the *cloud provider*.
- Cloud provider is responsible for making cloud services available to cloud consumers, as per agreed SLA guarantees.
- Cloud Providers does required management and administrative tasks.

Roles and Boundaries

- **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.



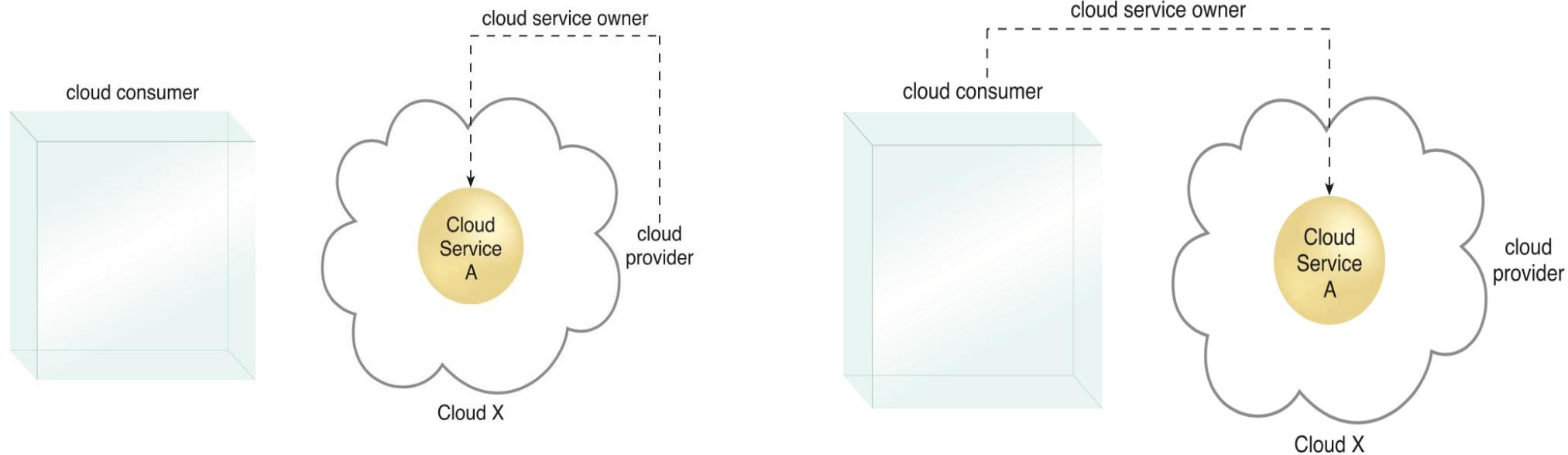
Roles and Boundaries

- **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.
- Example 1: Consider the Spotify application is hosted in amazon web services (AWS) and the end users are accessing the application.
- Here the cloud service owner is Spotify and cloud provider is AWS.

Roles and Boundaries

- **Example 2:** Consider the Amazon Prime video streaming platform is hosted in Amazon Web Services (AWS) and the end users are accessing the platform.
- Here both the cloud service owner and cloud provider is AWS.
- **Example 3:** Consider Bob is having an application that converts the pdf files to word file. The application is hosted in AWS.
- Here the cloud service owner is Bob, cloud provider is AWS, and the users accessing the conversion service is cloud consumers.

Roles and Boundaries



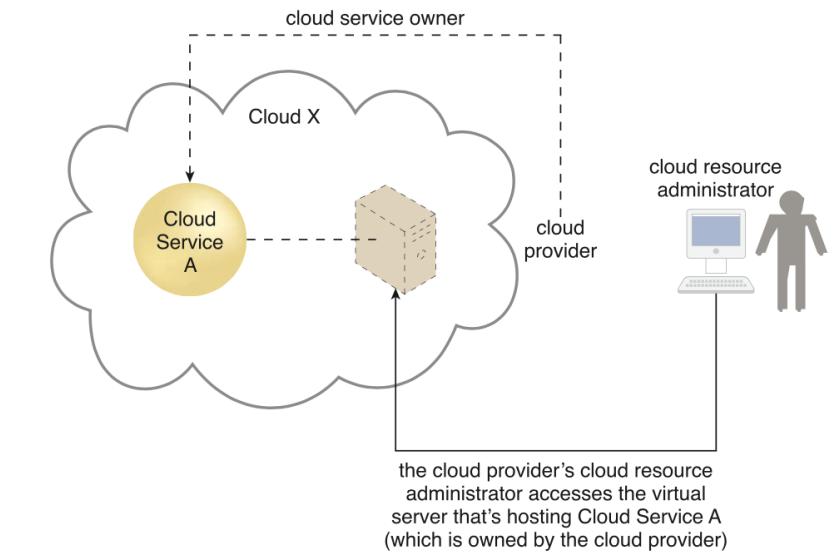
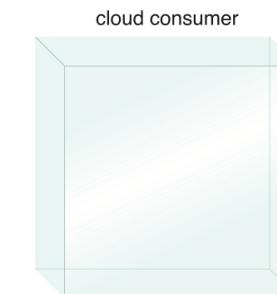
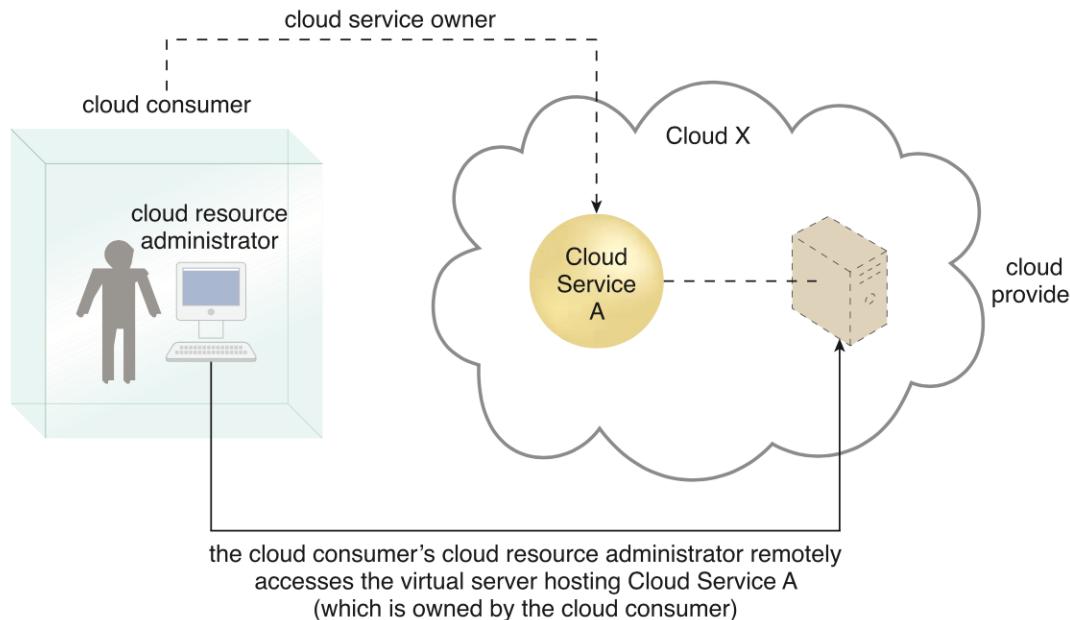
A cloud provider becomes a cloud service owner if it deploys its own cloud service, typically for other cloud consumers to use.

A cloud consumer can be a cloud service owner when it deploys its own service in a cloud.

Roles and Boundaries

- **Cloud Resource Administrators**
- A *cloud resource administrator* is the person or organization responsible for administering a cloud-based IT resource
- The cloud resource administrator can be the cloud consumer or cloud provider of the cloud.
- You can also hire a third party to manage your cloud resources.

Roles and Boundaries



Additional Roles

- *Cloud Auditor*—A third-party (often accredited) that conducts independent assessments of cloud environments assumes the role of the *cloud auditor*.
- Cloud auditor role is to provide an unbiased assessment of a cloud environment to help strengthen the trust relationship between cloud consumers and cloud providers.
- *Cloud Broker*—is a party that assumes the responsibility of managing and negotiating the usage of cloud services between cloud consumers and cloud providers.

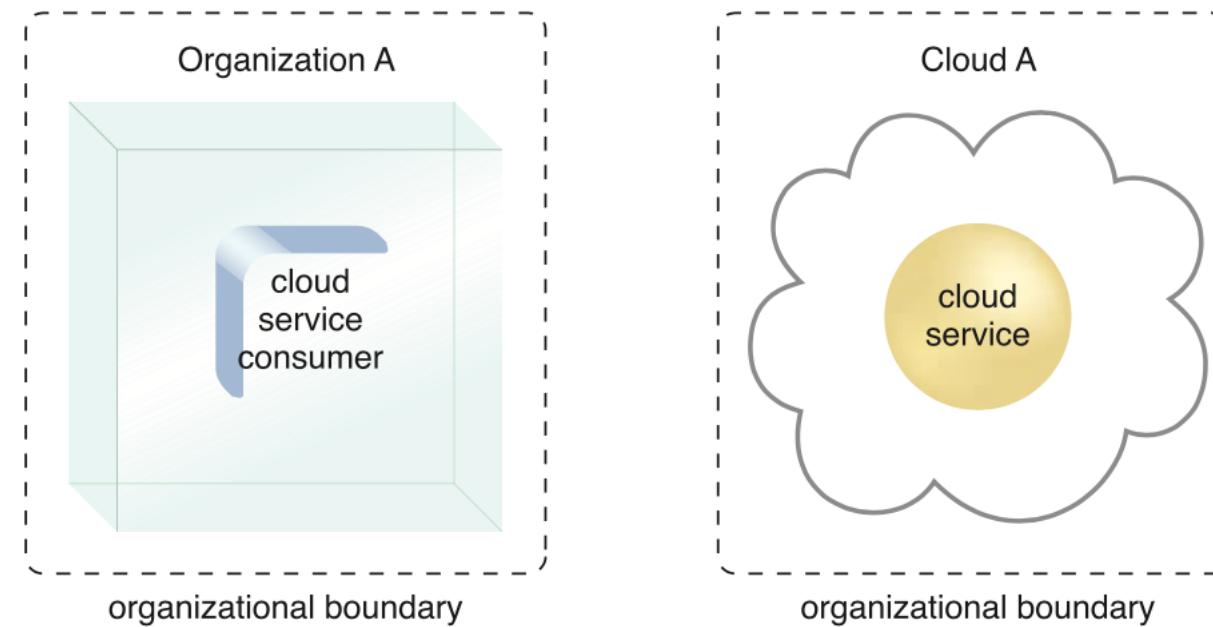


VIT-AP
UNIVERSITY

Additional Roles

- *Cloud Carrier*—The party responsible for providing the wire-level connectivity between cloud consumers and cloud providers assumes the role of the *cloud carrier*.

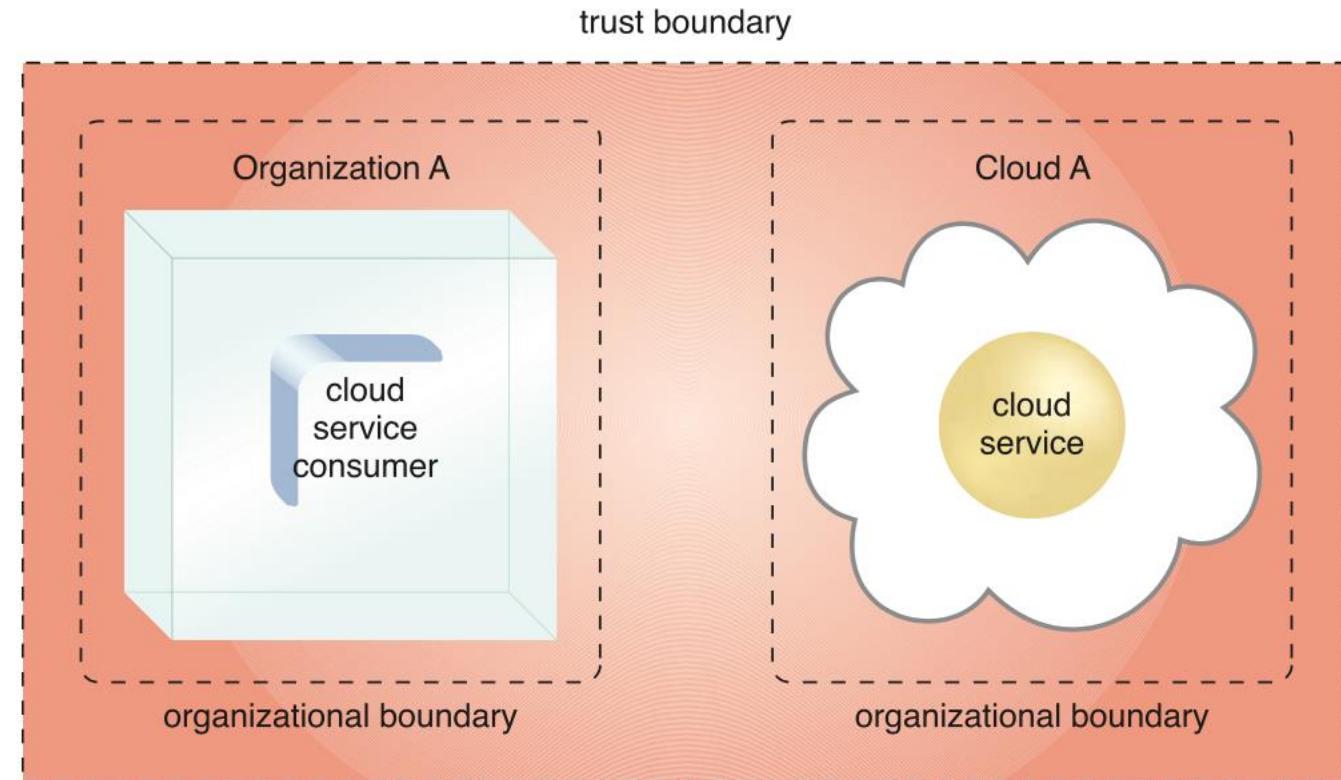
Organizational Boundary



Organizational boundaries of a cloud consumer (left), and a cloud provider (right), represented by a broken line notation.

Trust Boundary

- When an organization assumes the role of cloud consumer to access cloud-based IT resources, it needs to extend its trust beyond the physical boundary of the organization to include parts of the cloud environment.

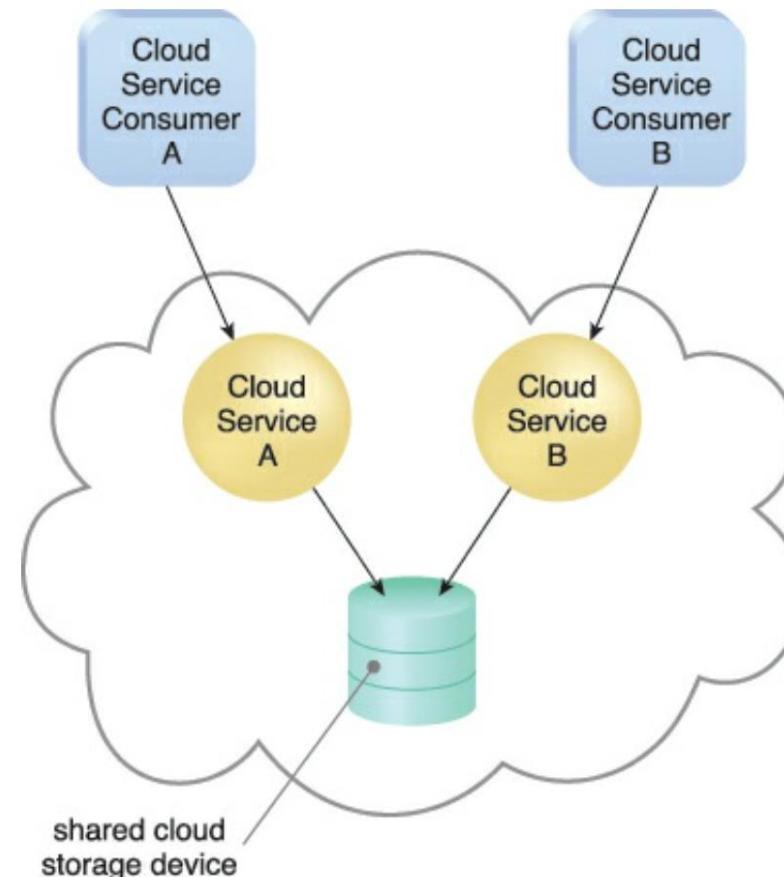


An extended trust boundary encompasses the organizational boundaries of the cloud provider and the cloud consumer.

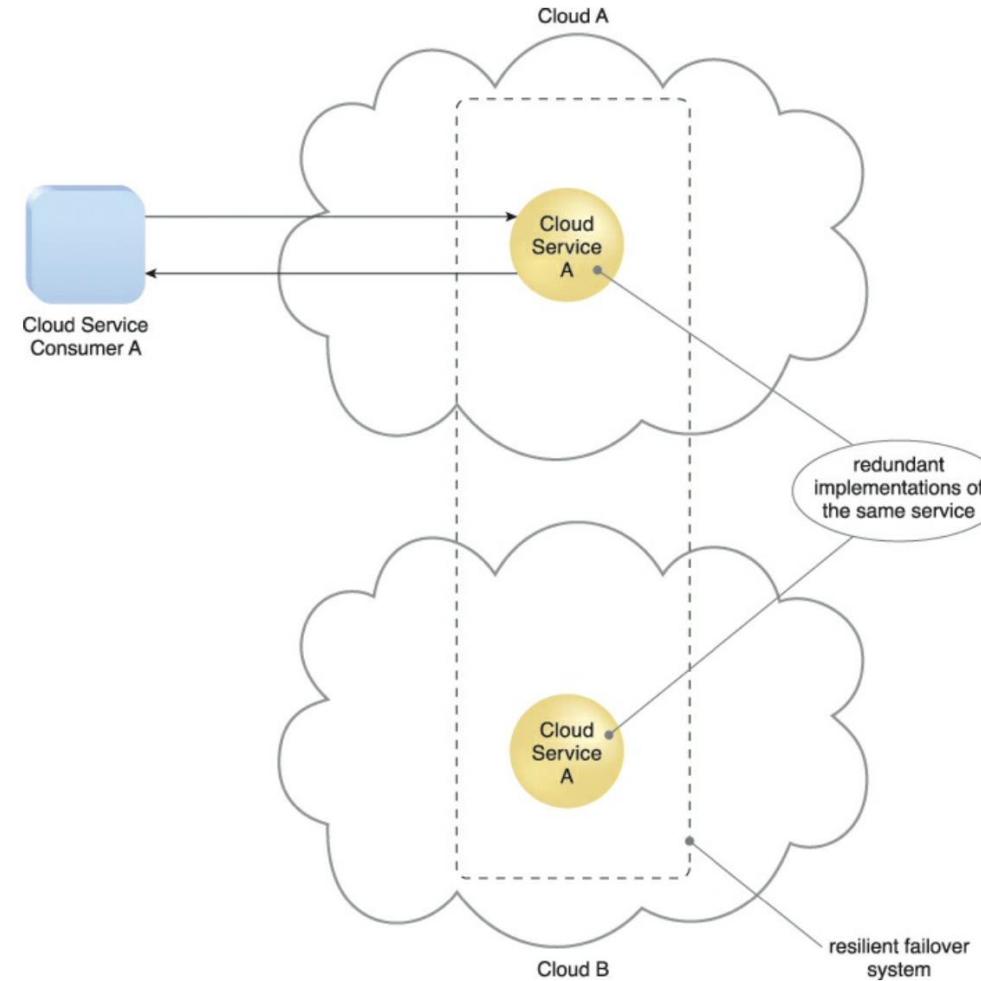
Cloud Characteristics

- The following six specific characteristics are common to most cloud environments:
 1. On-demand usage
 2. Ubiquitous access
 3. Multitenancy(and resource pooling)
 4. Elasticity
 5. Measured usage
 6. Resiliency

Multitenancy



Resiliency



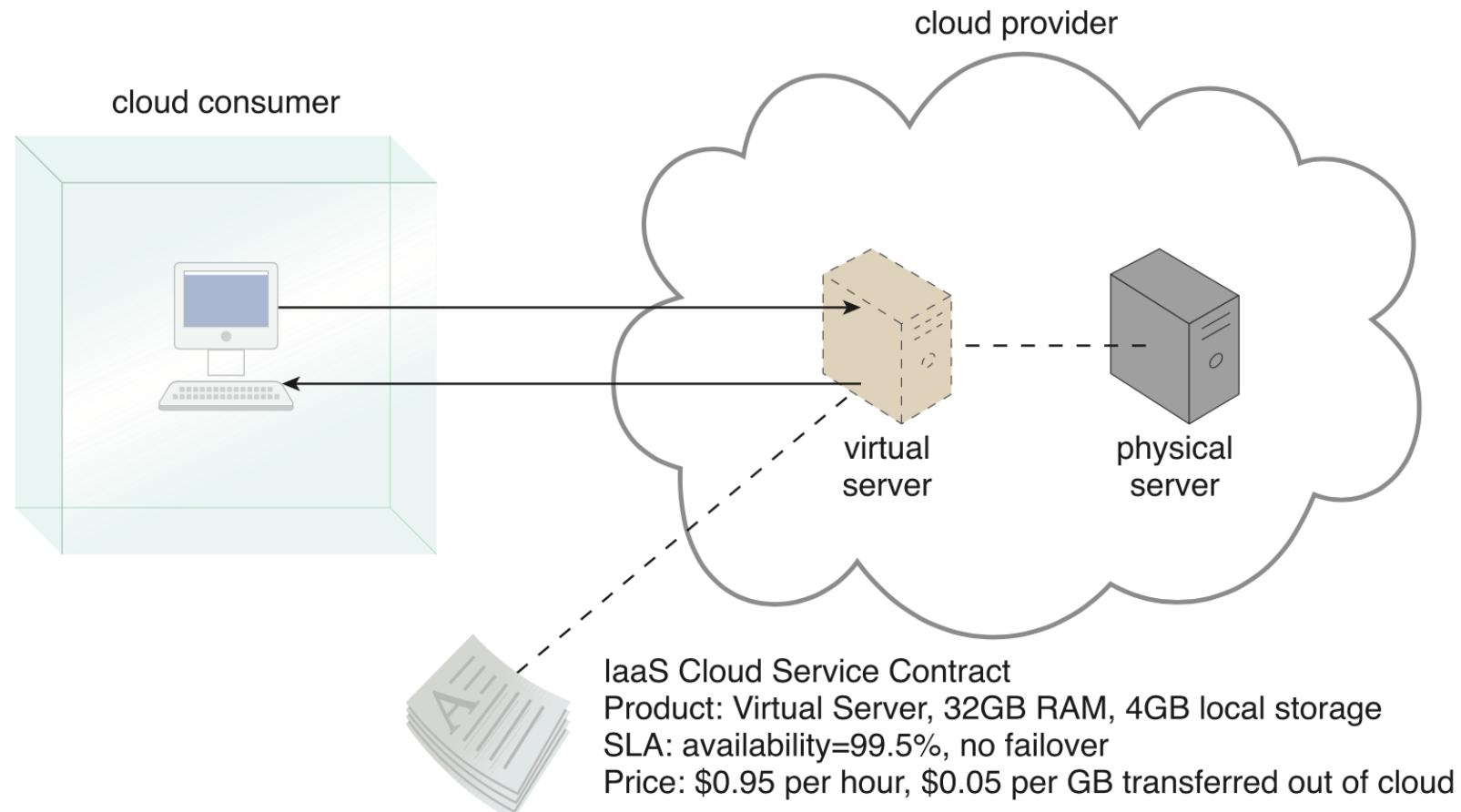
Cloud Delivery Models

- Three common cloud delivery models have become widely established and formalized:
 1. Infrastructure-as-a-Service (IaaS)
 2. Platform-as-a-Service (PaaS)
 3. Software-as-a-Service (SaaS)
- Storage-as-a-Service
- Database-as-a-Service
- Security-as-a-Service
- Communication-as-a-Service
- Integration-as-a-Service
- Testing-as-a-Service
- Process-as-a-Service

Infrastructure-as-a-Service (IaaS)

- 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.
- IaaS environment provide cloud consumers with a high level of control and responsibility over its configuration and utilization.
- Virtual servers are leased by specifying server hardware requirements, such as processor capacity, memory, and local storage space

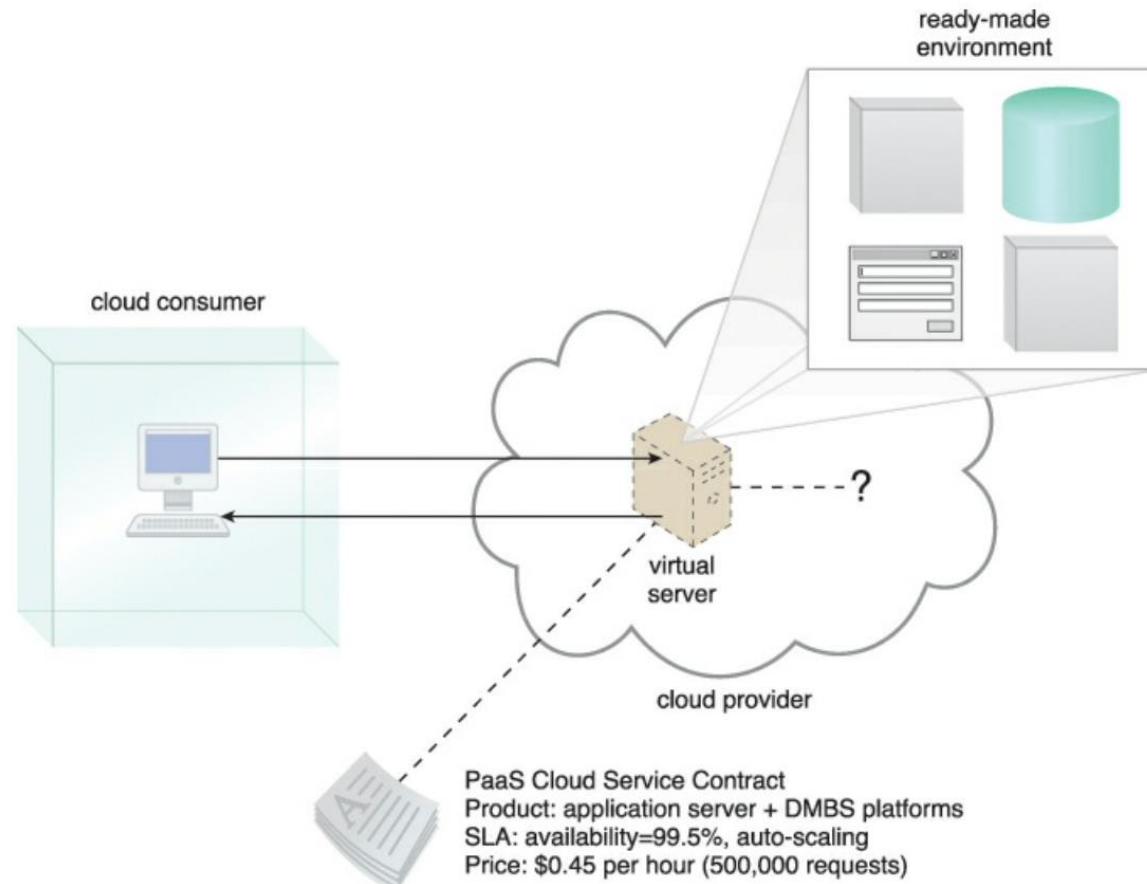
Infrastructure-as-a-Service (IaaS)



Platform-as-a-Service (PaaS)

- The PaaS delivery model represents a pre-defined “ready-to-use” environment comprised of already deployed and configured IT resources.
- Set of pre-packaged products and tools used to support the entire delivery lifecycle of custom applications.
- In PaaS, the cloud consumer is spared the administrative burden of setting up and maintaining the bare infrastructure of IT resources.
- The cloud consumer is granted a lower level of control over the underlying IT resources

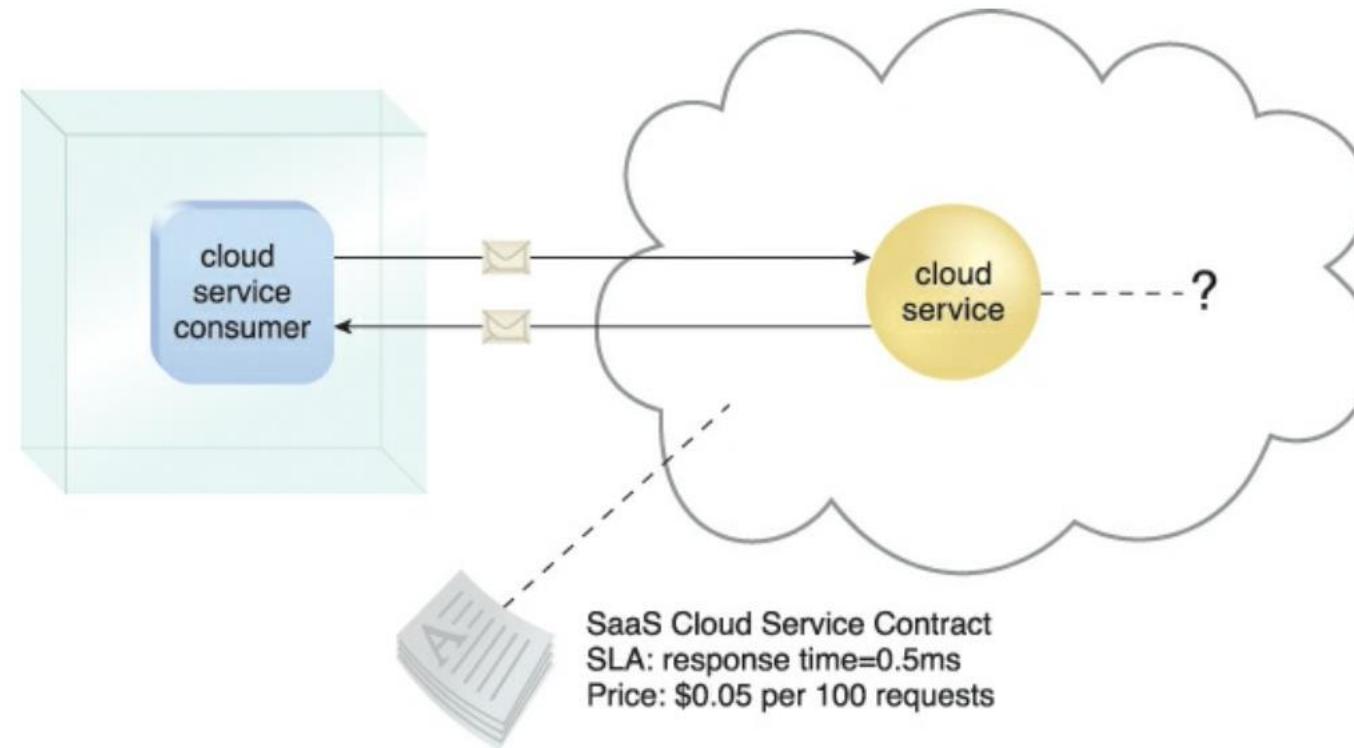
Platform-as-a-Service (PaaS)



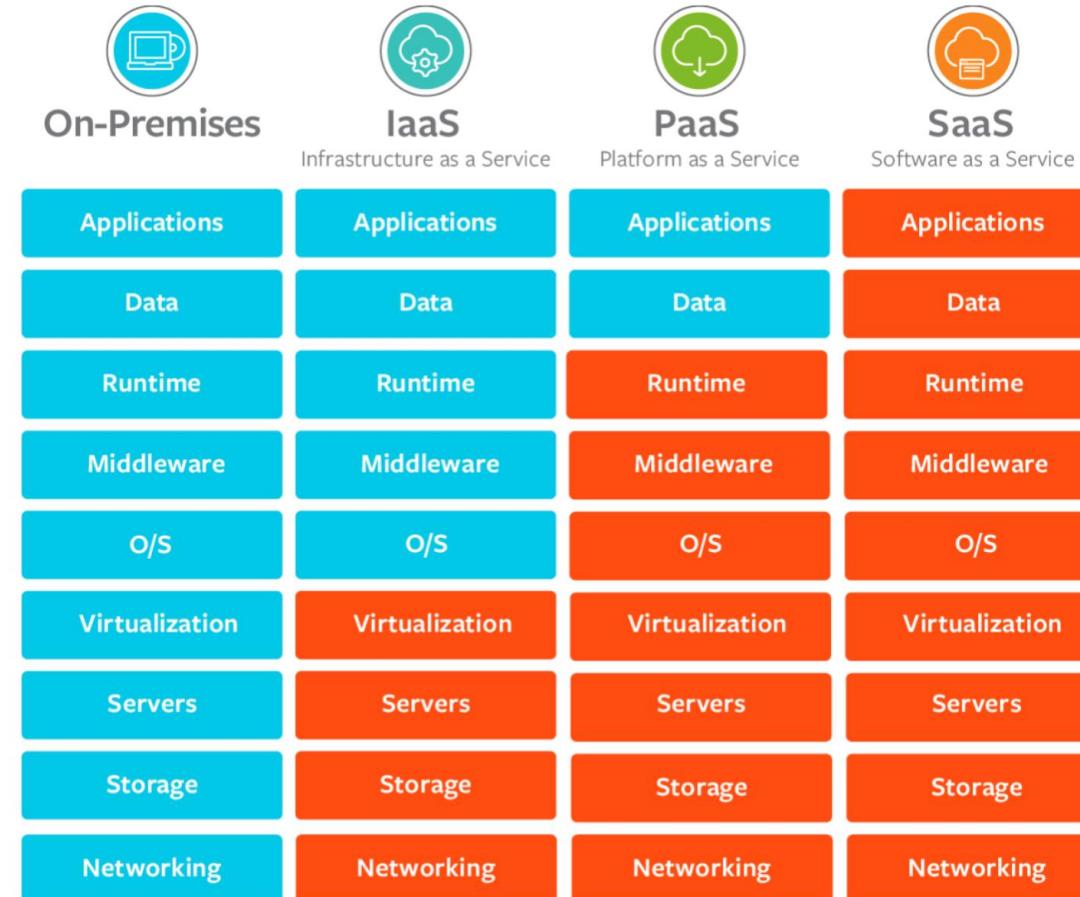
Software-as-a-Service (SaaS)

- A software program hosted in cloud and made available as a “product” represents the typical profile of a SaaS offering.
- A cloud consumer is generally granted very limited administrative control over a SaaS implementation.
- Gmail, Spotify, Amazon Prime, Netflix are all the examples for SaaS Platform.
- The user cannot configure the environment but can manage their data using the platform.

Software-as-a-Service (SaaS)



IaaS vs. PaaS vs. SaaS



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:
 1. Public cloud
 2. Community cloud
 3. Private cloud
 4. Hybrid cloud

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.
- The cloud provider is responsible for the creation and on-going maintenance of the public cloud

Community Clouds

- A community cloud is like 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

Private Cloud

- 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.
- The actual administration of a private cloud environment may be carried out by internal or outsourced staff.

Hybrid Cloud

- A hybrid cloud is a cloud environment comprised of two or more different cloud deployment models.
- 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.

End of Module 1

Cloud Computing – CSE4001

Course Instructor: Dr. Arunkumar Gopu

Senior Assistant Professor – Grade I

School of Computer Science and Engineering (SCOPE)

VIT – AP University, Amaravati, Andhra Pradesh.

Email-id: arunkumar.gopu@vitap.ac.in

Module 2

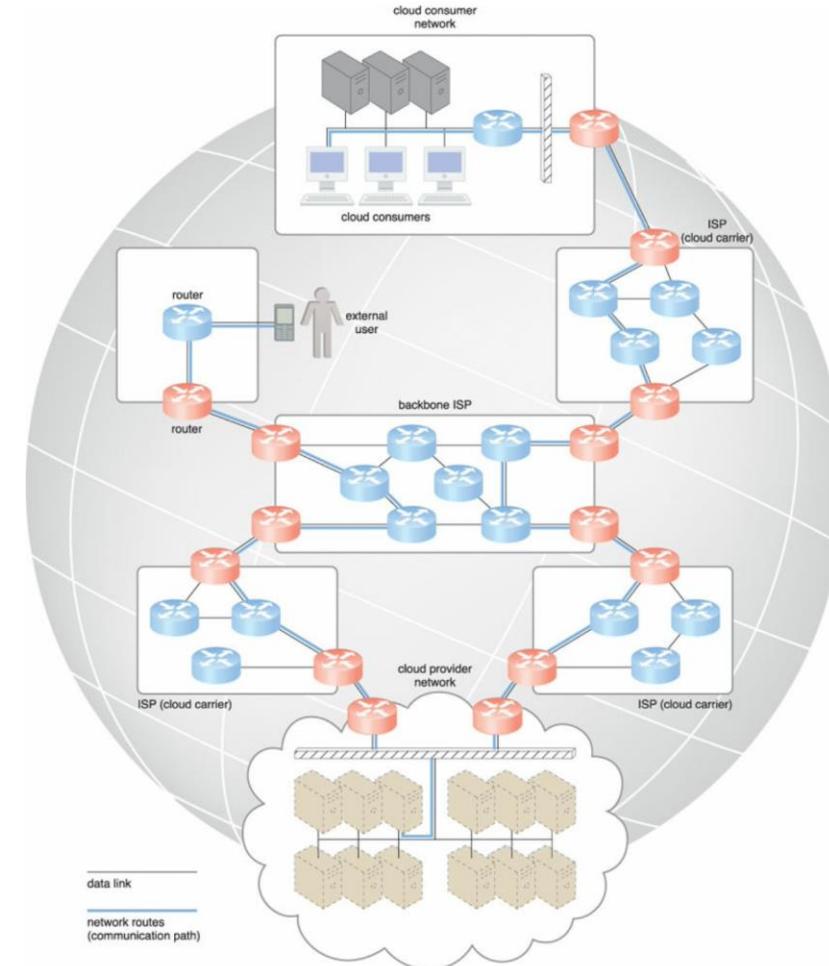
Cloud Enabling Technology

Data centre technology, virtualization technology, web technology, multitenant technology, service technology.

Broadband Networks and Internet Architecture

- All clouds must be connected to a network.
- Cloud consumers have the option of accessing the cloud using private and dedicated network links in LANs,
- Although most clouds are Internet-enabled.
- **Internet Service Providers (ISPs)**
- ISPs, connects largest backbone networks, strategically with core routers that connect the world's multinational networks.

Broadband Networks and Internet Architecture



Broadband Networks and Internet Architecture

- ISPs can deploy, operate, and manage their networks in addition to selecting partner ISPs for interconnection.
- Governmental and regulatory laws dictate the service provisioning conditions for organizations and ISPs both within and outside of national borders.
- The Internet's topology has become a dynamic and complex aggregate of ISPs that are highly interconnected via its core protocols.
- Worldwide connectivity is enabled through a hierarchical topology composed of Tiers 1, 2, and 3 ISPs.

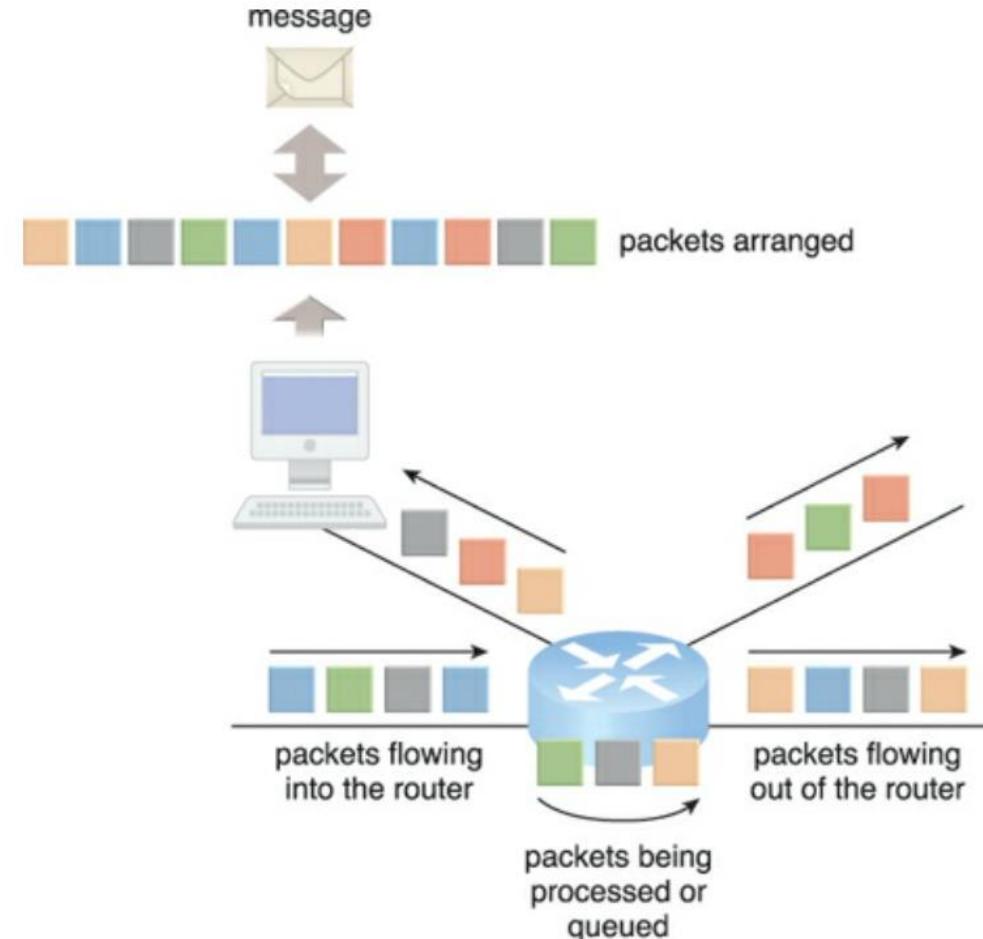
Broadband Networks and Internet Architecture

- **Connectionless Packet Switching (Datagram Networks)**
- End-to-end (sender-receiver pair) data flows are divided into packets of a limited size that are received and processed through network switches and routers.
- Then queued and forwarded from one intermediary node to the next.
- Each packet carries the necessary location information, such as the IP or MAC address, to be processed and routed at every source, intermediary, and destination node.

Broadband Networks and Internet Architecture

- **Router-Based Interconnectivity**
- A router is a device that is connected to multiple networks through which it forwards packets.
- Even when successive packets are part of the same data flow, routers process and forward each packet individually while maintaining the network topology information that locates the next node on the communication path between the source and destination nodes.
- Routers manage network traffic and gauge the most efficient hop for packet delivery.

Broadband Networks and Internet Architecture



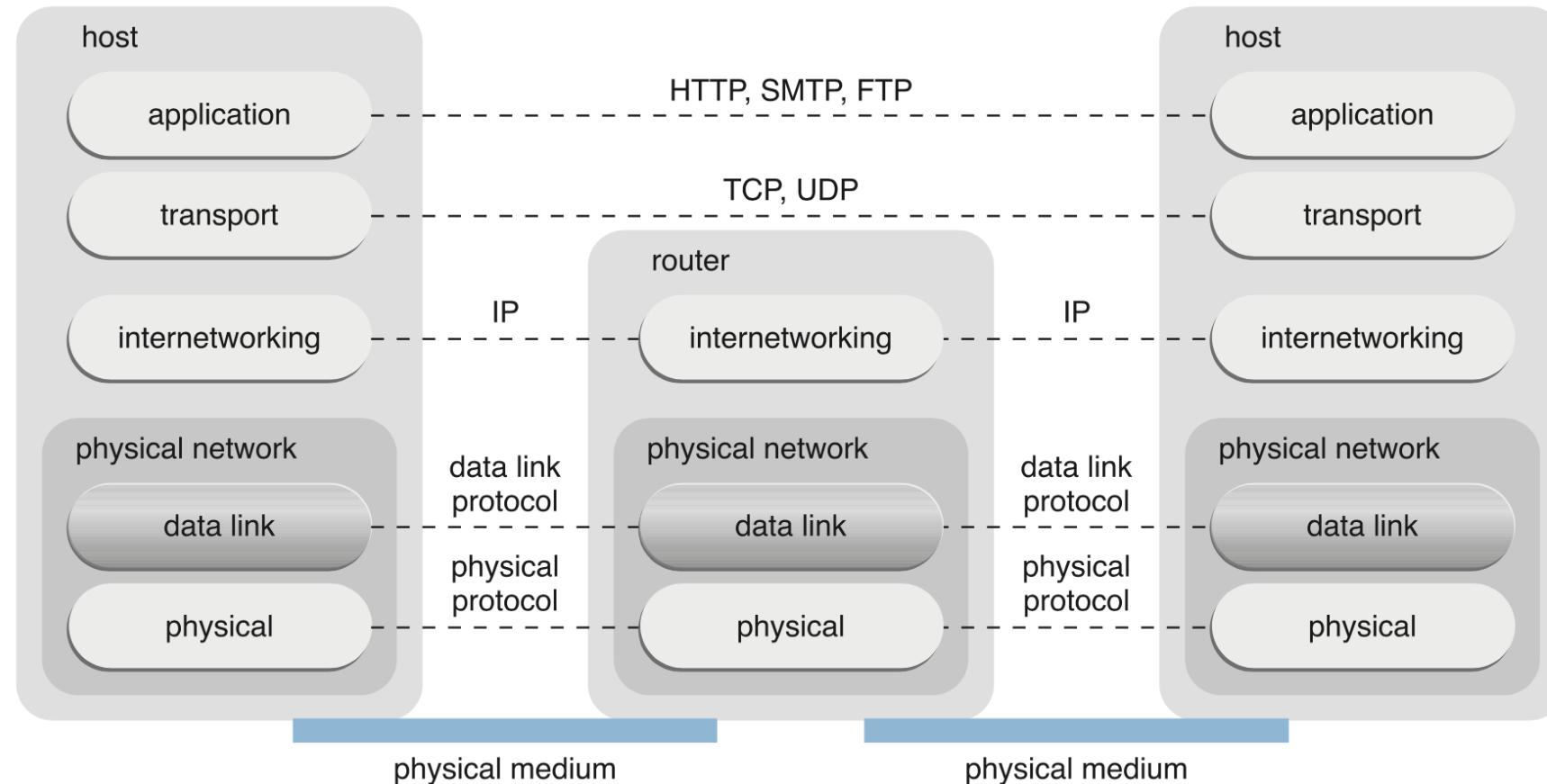
Broadband Networks and Internet Architecture

- Internet's internetworking layer and interact with other network technologies as follows
- **Physical Network**
- IP packets are transmitted through underlying physical networks that connect adjacent nodes, such as Ethernet, ATM network, and the 3G mobile HSDPA.
- Physical networks comprise a data link layer that controls data transfer between neighbouring nodes, and a physical layer that transmits data bits through both wired and wireless media.

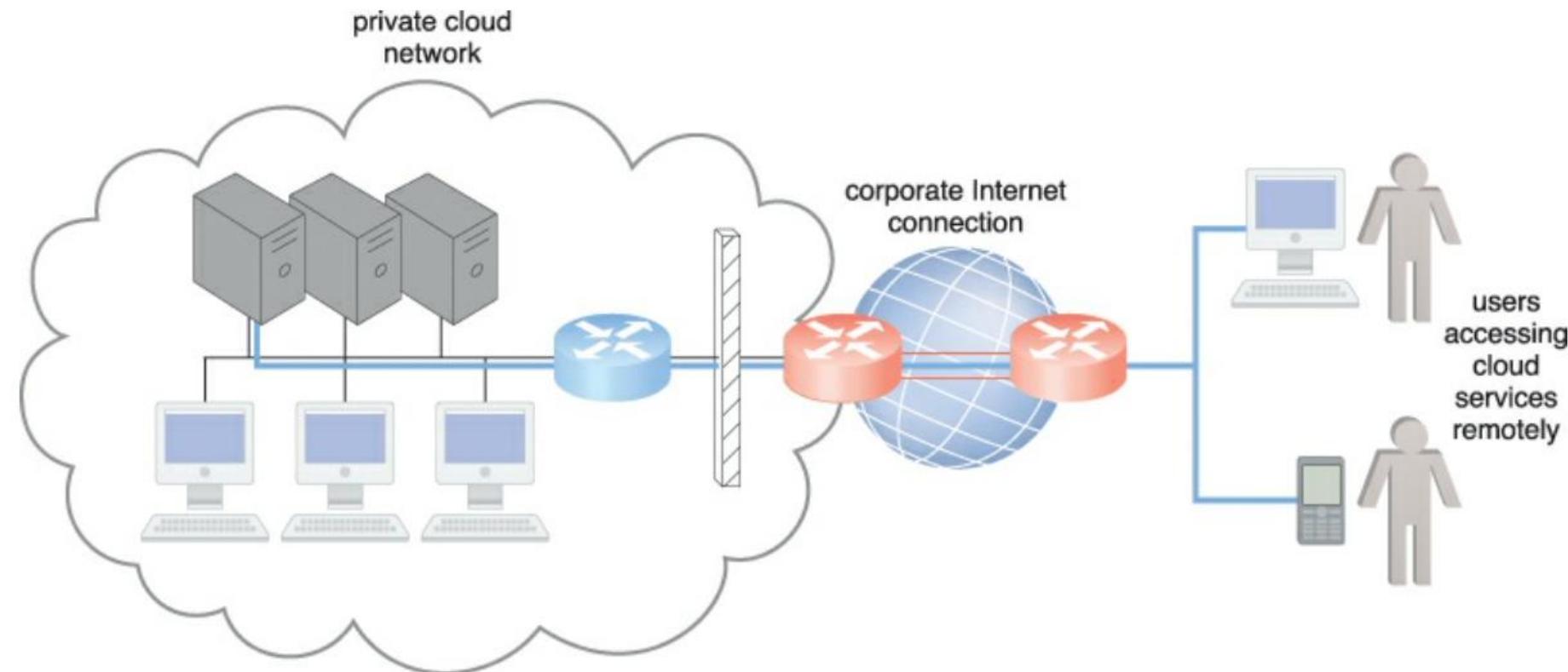
Broadband Networks and Internet Architecture

- **Transport Layer Protocol**
- Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), use the IP to provide standardized, end-to-end communication.
- **Application Layer Protocol**
- Protocols such as HTTP, SMTP for e-mail, BitTorrent for P2P, and SIP for IP telephony use transport layer protocols to enable packet transferring methods over the Internet.
- Many other protocols uses either TCP/IP or UDP as their primary method of data transferring across the Internet and LANs.

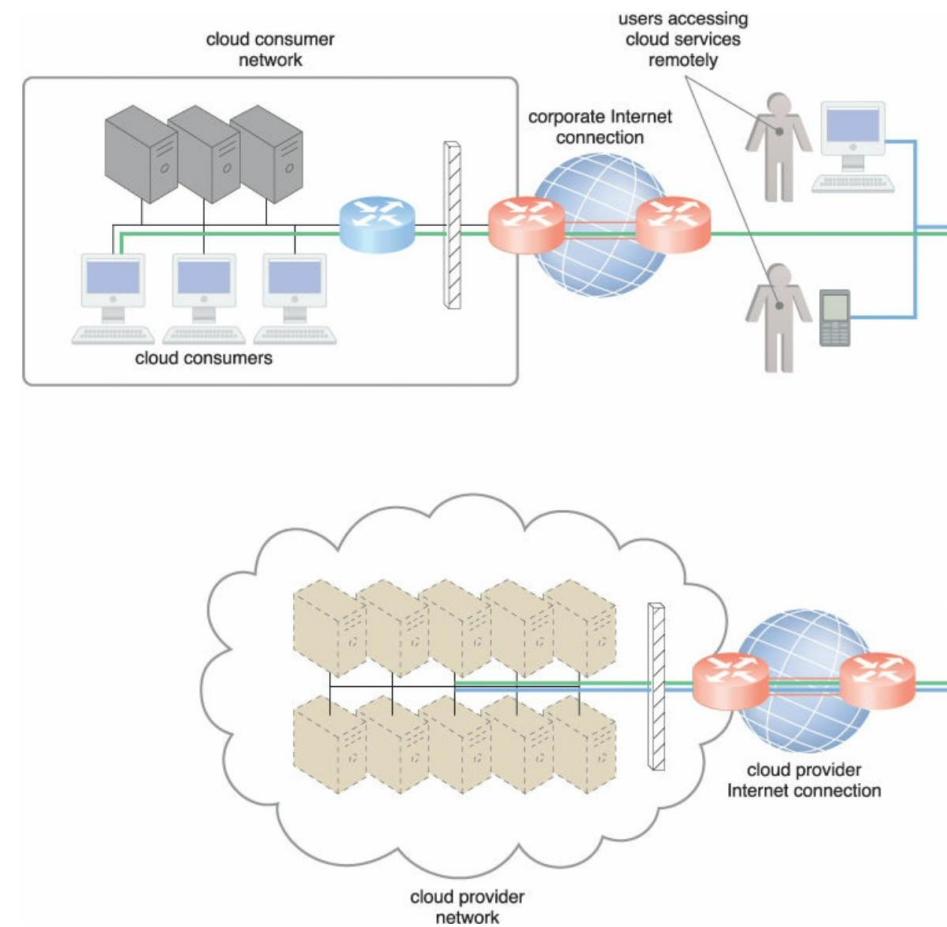
Broadband Networks and Internet Architecture



Technical and Business Considerations



Technical and Business Considerations



Technical and Business Considerations

- **Network Bandwidth and Latency Issues**
- End to end bandwidth is determined by the transmission capacity of the shared data links that connect intermediary nodes.
- Web acceleration technologies, dynamic caching, compression, and pre-fetching, continue to improve bandwidth and connectivity.
- Latency is the amount of time it takes a packet to travel from one data node to another.
- Latency increases with every intermediary node on the data packet's path.
- Transmission queues in the network infrastructure can result in heavy load.

Technical and Business Considerations

- Quality-of-Service (QoS) typically transmit packets on a first-come/first-serve basis.
- Data flows through congested network paths suffer service degradation.
- The nature of packet switching allows data packets to choose routes dynamically as they travel through the Internet's network infrastructure.

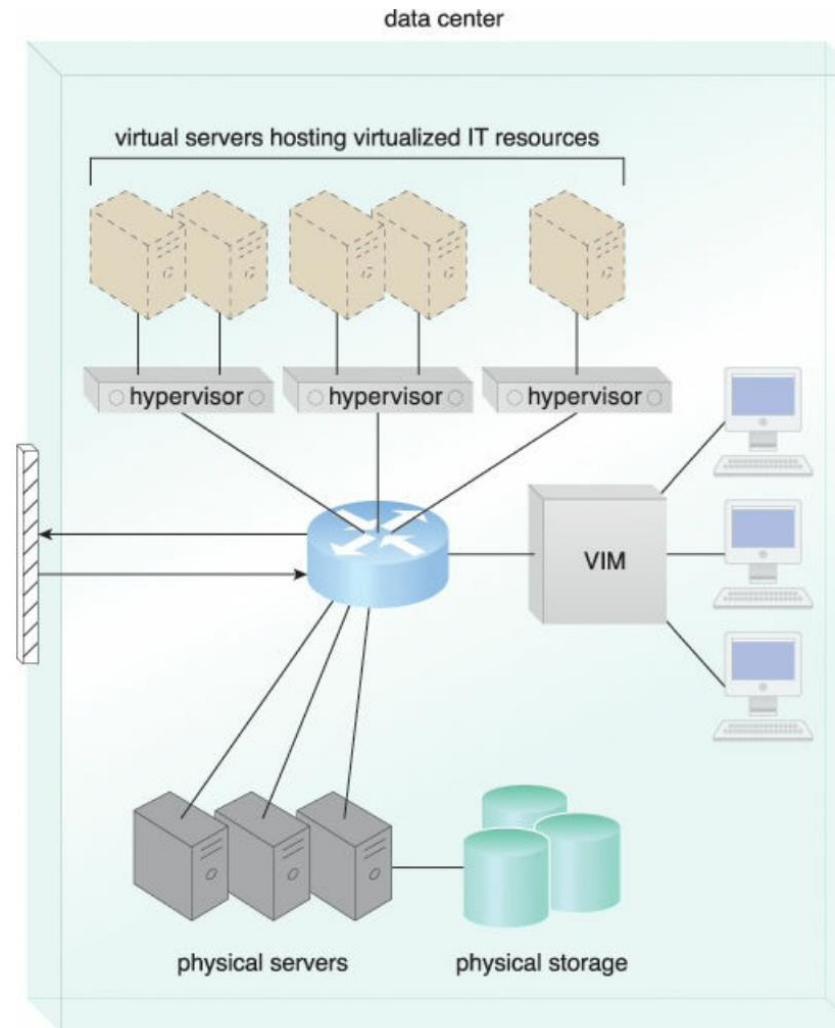
Technical and Business Considerations

- **Cloud Carrier and Cloud Provider Selection**
- The service levels of Internet connections between cloud consumers and cloud providers are determined by their ISPs.
- QoS management across multiple ISPs is difficult to achieve in practice, requiring collaboration of the cloud carriers on both sides to ensure that their end-to-end service levels are sufficient for business requirements.
- Cloud consumers and cloud providers may need to use multiple cloud carriers in order to achieve the necessary level of connectivity and reliability for their cloud applications.

Data Center Technology

- Data centers consist of both physical and virtualized IT resources.
- **Virtualization** layer is comprised of operational and management tools that abstract the physical computing and networking IT resources.
- Virtualized components that are easier to allocate, operate, release, monitor, and control.
- Data centers are built upon standardized commodity hardware and designed with modular architectures.
- **Modularity and standardization** are key requirements for reducing investment and operational costs.

Data Center Technology



Data Center Technology

- Virtualization strategies and the constantly improving capacity and performance of physical devices both favor IT resource consolidation.
- Consolidated IT resources can serve different systems and be shared among different cloud consumers.
- **Automation** tasks like provisioning, configuration, patching, and monitoring without supervision.
- Advances in data center management platforms and tools enables self-configuration and self-recovery.

Data Center Technology

- **Remote Operation and Management**
- Operational and administrative tasks of IT resources in data centers are commanded through the network's remote consoles and management systems.
- Technical personnel are not required to visit the dedicated rooms that house servers, except to perform highly specific tasks, such as equipment handling and cabling or hardware-level installation and maintenance.

Data Center Technology

- **High Availability**
- Data centers are designed to operate with increasingly higher levels of redundancy to sustain availability.
- Data centers usually have redundant, uninterruptable power supplies, cabling, and environmental control subsystems in anticipation of system failure, along with communication links and clustered hardware for load balancing.

Data Center Technology

- **Security-Aware Design, Operation, and Management**
- Requirements for security, such as physical and logical access controls and data recovery strategies, need to be strictly implemented.
- Prohibitive nature of building and operating on-premise data centers, outsourcing data center-based IT resources has been a common industry practice for decades.
- In house datacentre could not provide elasticity, issues that a typical cloud can address via inherent features, such as ubiquitous access, on-demand provisioning, rapid elasticity, and pay-per-use.

Data Center Technology

- **Facilities**
- Data center facilities are custom-designed locations that are outfitted with specialized computing, storage, and network equipment.
- These facilities have several functional layout areas, as well as various power supplies, cabling, and environmental control stations that regulate heating, ventilation, air conditioning, fire protection, and other related subsystems.

Data Center Technology

- **Computing Hardware**
- Processing in data centers is executed by servers that have substantial computing power and storage capacity.
- Several computing hardware technologies are integrated into these modular servers, such as:
 1. Rackmount form factor server design composed of standardized racks with interconnects for power, network, and internal cooling
 2. Support for different hardware processing architectures, such as x86-32bits, x86-64, and RISC

Data Center Technology

- 3. Power-efficient multi-core CPU architecture that houses hundreds of processing cores in a small space as single unit.
- 4. Redundant and hot-swappable components, such as hard disks, power supplies, network interfaces, and storage controller cards.
- These modular architecture supports server hot-swapping, scaling, replacement, and maintenance, which benefits the deployment of fault-tolerant systems that are based on computer clusters.

Data Center Technology



Data Center Technology

- **Storage Hardware**
- *Hard Disk Arrays*—These arrays inherently divide and replicate data among multiple physical drives, increase performance and redundancy.
- I/O Caching - enhance disk access times
- Hot-Swappable Hard Disks
- Storage Virtualization
- Fast Data Replication Mechanisms

Data Center Technology

- **Networked storage devices**
- Storage Area Network(SAN)
- Network-Attached Storage(NAS)
 - Network File System (NFS)
 - Server Message Block (SMB).

Data Center Technology

- **Network Hardware**
- The data center is broken down into five network subsystems.
 1. Carrier and External Networks Interconnection
 - Routing between external WAN connections and the data center's LAN
 2. Web-Tier Load Balancing and Acceleration
 - XML pre-processors, encryption/decryption appliances, and layer 7 switching devices that perform content-aware routing.

Data Center Technology

3. LAN Fabric

- LAN fabric constitutes the internal LAN and provides high-performance and redundant connectivity for all of the data center's network-enabled IT resources.
- VLANs, link aggregation, controlled routing between networks, load balancing, and failover.

4. SAN Fabric

- Implementation of storage area networks (SANs) that provide connectivity between servers and storage systems
- Fibre Channel (FC), Fibre Channel over Ethernet (FCoE)

Data Center Technology

5. NAS Gateways

- Facilitates data transmission between SAN and NAS devices.
- Scalability and high availability are fulfilled using redundant and/or fault-tolerant configurations.
- Other Considerations
 - Hardware lifecycles range five to seven years.
 - On-going need to replace equipment frequently results in a mix of hardware whose heterogeneity can complicate the entire data center's operations and management

Virtualization Technology

- Virtualization is the process of converting a physical IT resource into a virtual IT resource that includes
- *Servers*—A physical server can be abstracted into a virtual server.
- *Storage*—A physical storage device can be abstracted into a virtual storage device or a virtual disk.
- *Network*—Physical routers and switches can be abstracted into logical network fabrics, such as VLANs.
- *Power*—A physical UPS and power distribution units can be abstracted into what are commonly referred to as virtual UPSs.

Virtualization Technology

- Virtual servers use their own guest operating systems, which are independent of the operating system in which they were created.
- Both the guest operating system and the application software running on the virtual server are unaware of the underlying virtualization process.
- Guest operating systems typically require seamless usage of software products and applications that do not need any customization to run in virtual environment.
- Virtualization software runs on *host* or *physical host*.

Virtualization Technology

- Virtualization software is sometimes referred to as a virtual machine manager or a virtual machine monitor (VMM), but most known as a *hypervisor*.
- **Hardware Independence**
- Through hardware independence, virtual servers can easily be moved to another virtualization host, automatically resolving multiple hardware-software incompatibility issues.
- Cloning and manipulating virtual IT resources is much easier than duplicating physical hardware.

Virtualization Technology

- **Server Consolidation**
- Virtualization software allows multiple virtual servers to be simultaneously created in the same virtualization host.
- *Server consolidation* and is commonly used to increase hardware utilization, load balancing, and optimization of available IT resources.
- **Resource Replication**
- Virtual servers are created as virtual disk images that contain binary file copies of hard disk content.

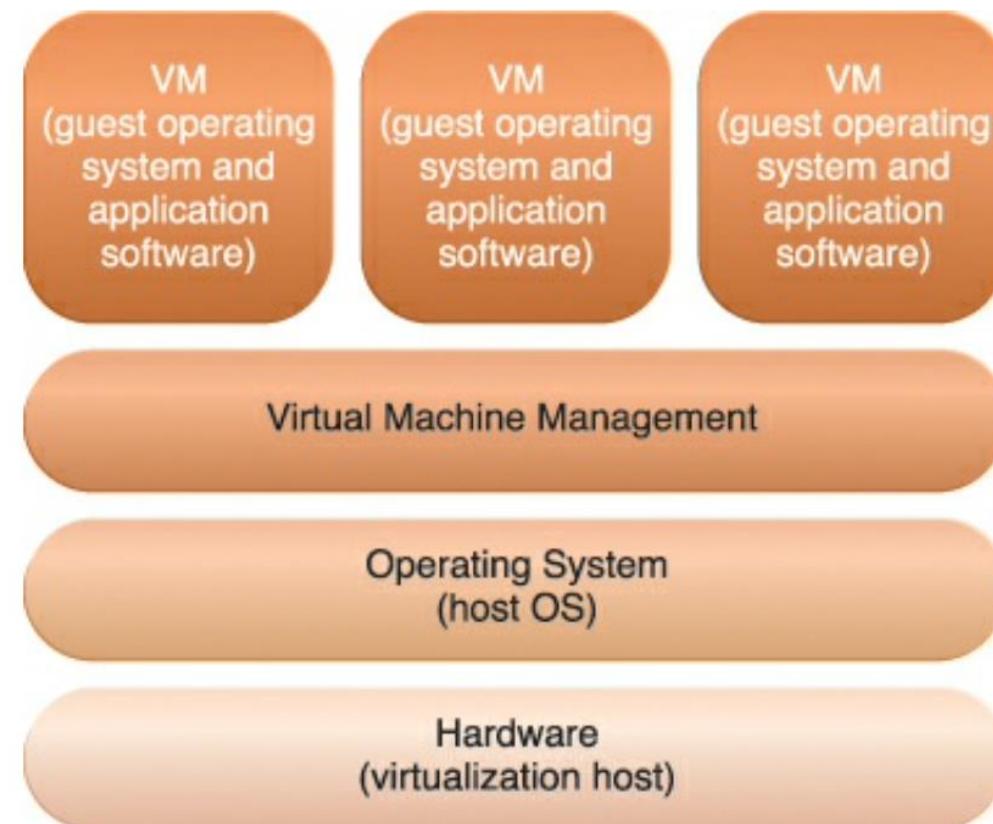
Virtualization Technology

- These virtual disk images are accessible to the host's operating system.
- Pre-packaging in virtual disk images in support of instantaneous deployment.
- Increased agility in the migration and deployment.
- VM snapshots by saving the state of the virtual server's memory and hard disk image to a host- based file.
- Support of business continuity with efficient backup and restoration procedures.

Virtualization Technology

- **Operating System-Based Virtualization**
- User needs to use this application to generate and operate one or more virtual servers.
- The user needs to use virtualization software to enable direct access to any of the generated virtual servers that a host operating system cannot provide.
- With the use of operating system-based virtualization the host OS provides flexibility to control the virtual server running on top of hypervisor.

Virtualization Technology



Virtualization Technology

- **Hardware-Based Virtualization**
- This option represents the installation of virtualization software directly on the physical host hardware without host operating system.
- The host operating system generally makes hardware-based virtualization more efficient.
- Virtualization software is typically referred to as a *hypervisor* for this type of processing.
- A hypervisor has a simple user-interface that requires a negligible amount of storage space.
- It exists as a thin layer of software that handles hardware management functions to establish a virtualization management layer.

Virtualization Technology

- Hardware-based virtualization concerns compatibility with hardware devices.
- Associated device drivers and support software need to be compatible with the hypervisor.
- Hardware device drivers may not be as available to hypervisor platforms as they are to operating systems.
- **Virtualization Management**
- Modern virtualization software provides advanced management functions to automate administration tasks and reduce the overall operational burden.

Virtualization Technology

- Resource management is often supported by virtualization infrastructure management (VIM).
- **Other Considerations**
- Performance Overhead – Virtualization may not be ideal for complex systems that have high workloads with little use for resource sharing and replication.
- A poorly formulated virtualization plan can result in excessive performance overhead.
- Special Hardware Compatibility – Many hardware vendors that distribute specialized hardware may not have device driver versions that are compatible with virtualization software.

Virtualization Technology

- Portability –Portability gaps due to incompatibilities.
- Initiatives such as the Open Virtualization Format (OVF) for the standardization of virtual disk image formats are dedicated to alleviating this concern.

Web Technology

- Web Technology is the implementation medium and the management interface for cloud services.
- **Basic Web Technology**
- The two basic components of the Web are the Web browser client and the Web server.
- Other components, such as proxies, caching services, gateways, and load balancers, are used to improve Web application characteristics
- These additional components reside between the client and the server.

Web Technology

- Three fundamental elements of Web technology architecture are:
 1. Uniform Resource Locator (URL) – A standard syntax used for creating identifiers that point to Web-based resources, the URL is often structured using a logical network location.
 2. Hypertext Transfer Protocol (HTTP) –This is the primary communications protocol used to exchange content and data throughout the World Wide Web.

URLs are typically transmitted via HTTP.

Web Technology

3. Markup Languages (HTML,XML) –Markup languages provide a light weight means of expressing Web-centric data and metadata.

- The two primary markup languages are HTML (which is used to express the presentation of Web pages)
- XML (which allows for the definition of vocabularies used to associate meaning to Web-based data via metadata).

Web Technology

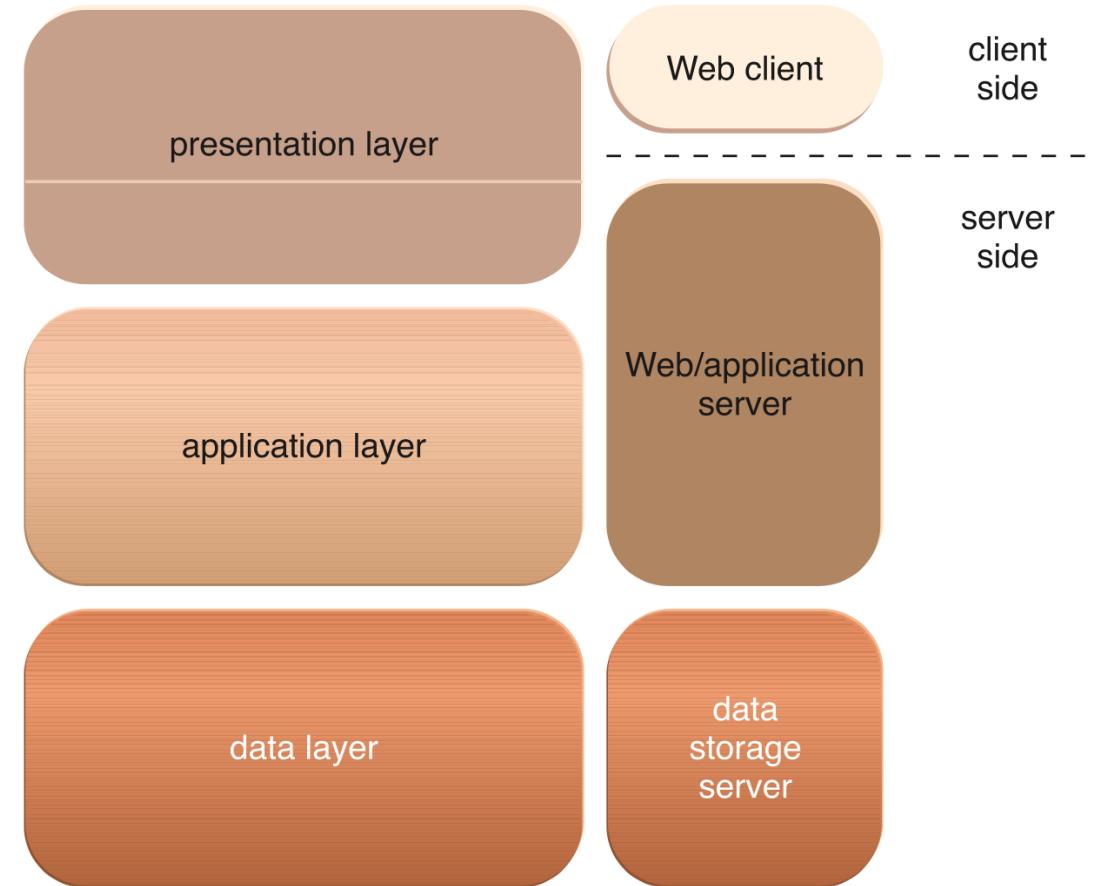
```
<?xml version="1.0" encoding="UTF-8"?>
- <EmployeeData>
  - <employee id="34594">
    <firstName>Heather</firstName>
    <lastName>Banks</lastName>
    <hireDate>1/19/1998</hireDate>
    <deptCode>BB001</deptCode>
    <salary>72000</salary>
  </employee>
  - <employee id="34593">
    <firstName>Tina</firstName>
    <lastName>Young</lastName>
    <hireDate>4/1/2010</hireDate>
    <deptCode>BB001</deptCode>
    <salary>65000</salary>
  </employee>
</EmployeeData>
```

```
{
  "orders": [
    {
      "orderno": "748745375",
      "date": "June 30, 2088 1:54:23 AM",
      "trackingno": "TN0039291",
      "custid": "11045",
      "customer": [
        {
          "custid": "11045",
          "fname": "Sue",
          "lname": "Hatfield",
          "address": "1409 Silver Street",
          "city": "Ashland",
          "state": "NE",
          "zip": "68003"
        }
      ]
    }
  ]
}
```

Web Technology

- **Web Applications**
- A distributed application that uses Web-based technologies (and generally relies on Web browsers for the presentation of user-interfaces) is typically considered a Web application.
- The first tier is called the presentation layer, which represents the user-interface.
- The middle tier is the application layer that implements application logic, while the third tier is the data layer that is comprised of persistent data stores.

Web Technology



Web Technology

- The presentation layer has components on both the client and server-side.
- PaaS ready-made environments enable cloud consumers to develop and deploy Web applications.
- Typical PaaS offerings have separate instances of the Web server, application server, and data storage server environments.

Multitenant Technology

- Enable multiple users (tenants) to access the same application logic simultaneously.
- Each tenant has its own view of the application. Multitenant applications ensure that tenants do not have access to data and configuration information that is not their own.
- Tenant can customize
 1. User Interface—Tenants can define a specialized “look and feel” for their application interface.
 2. Business Process – Tenants can customize the rules, logic, and workflows of the business processes that are implemented in the application.

Multitenant Technology

3. Data Model—Tenants can extend the data schema of the application to include, exclude, or rename fields in the application data structures.
 4. Access Control—Tenants can independently control the access rights for users and groups.
- Multi tenant application architecture is often significantly more complex than single tenant application.
 - Multitenant applications shares various artifacts to multiple users while maintaining security levels that segregate individual tenant operational environments.

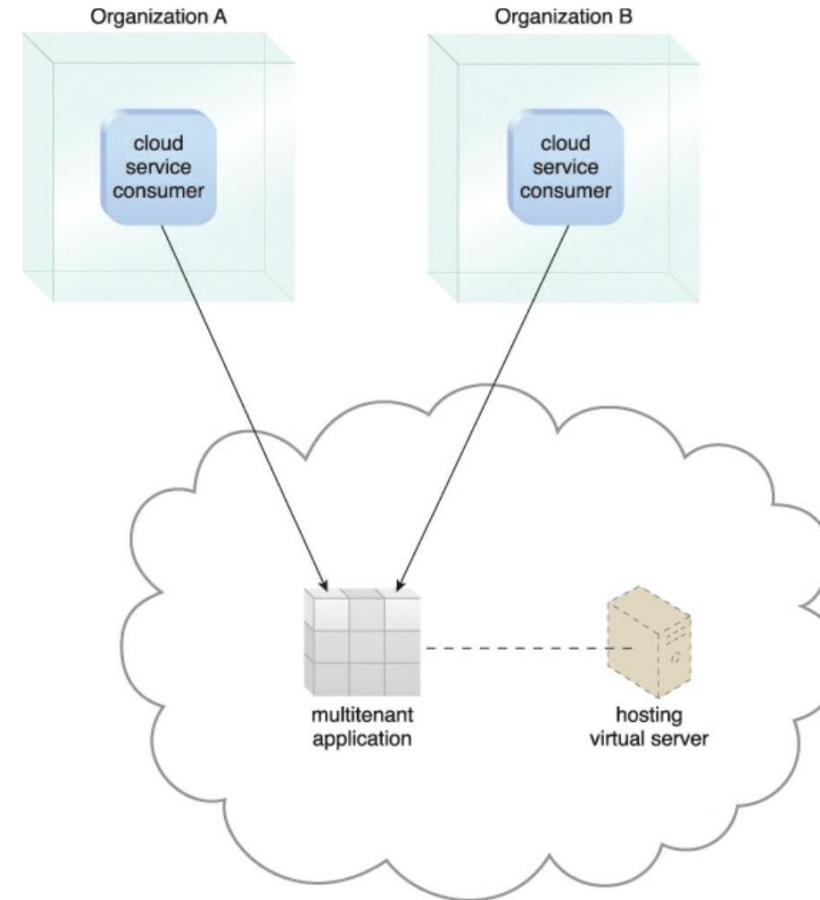
Multitenant Technology

- Common characteristics of multitenant applications include:
 1. Usage Isolation—The usage behavior of one tenant does not affect the application availability and performance of other tenants.
 2. Data Security—Tenants cannot access data that belongs to other tenants.
 3. Recovery—Backup and restore procedures are separately executed for the data of each tenant.
 4. Application Upgrades—Tenants are not negatively affected by the synchronous upgrading of shared software artifacts.

Multitenant Technology

5. Scalability—The application can scale to accommodate increases in usage by existing tenants and/or increases in the number of tenants.
6. Metered Usage—Tenants are charged only for the application processing and features that are consumed.
7. Data Tier Isolation—Tenants can have individual databases, tables, and/or schemas isolated from other tenants.

Multitenant Technology



Multitenant Technology vs. Virtualization

- Multiple virtual copies of the server environment can be hosted by a single physical server. Each copy can be provided to different users, can be configured independently, and can contain its own operating systems and applications.
- With multitenancy: A physical or virtual server hosting an application is designed to allow usage by multiple different users. Each user feels as though they have exclusive usage of the application.

Service Technology

- The field of service technology is a keystone foundation of cloud computing that formed the basis of the “as-a-service” cloud delivery models.
- **Web Service**
- SOAP-based, Web services represent an established and common medium for sophisticated, Web-based service logic.
- Along with XML, the core technologies behind Web services are represented by the following industry standards:

Service Technology

1. Web Service Description Language (WSDL)–This markup language is used to create a WSDL definition that defines the application programming interface (API) of a Web service.
2. XML Schema Definition Language (XML Schema) – Messages exchanged by Web services must be expressed using XML.
 - XML schemas are created to define the data structure of the XML-based input and output messages exchanged by Web services.
 - XML schemas can be directly linked to or embedded within WSDL definitions.

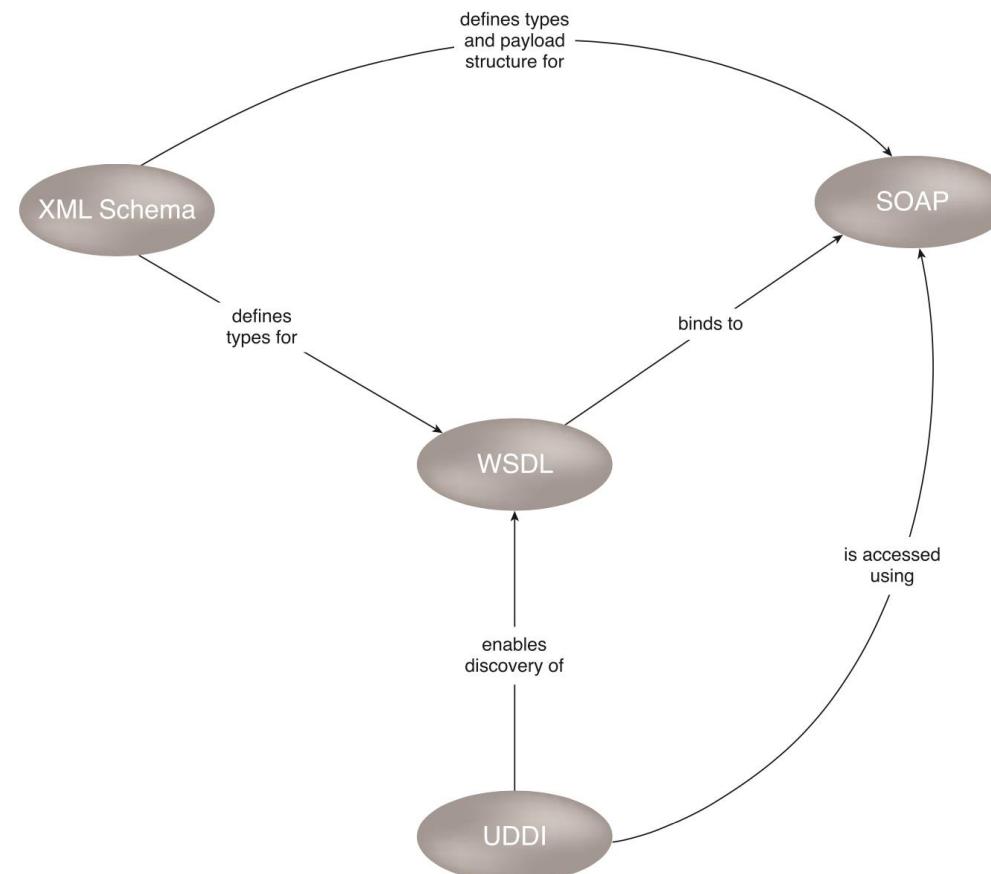
Service Technology

3. SOAP—Formerly known as the Simple Object Access Protocol, this standard defines a common messaging format used for request and response messages exchanged by Web services.

- SOAP messages are comprised of body and header sections.
- The former houses the main message content and the latter is used to contain metadata that can be processed at runtime.

4. Universal Description, Discovery, and Integration(UDDI) – This standard regulates service registries in which WSDL definitions can be published as part of a service catalog for discovery purposes.

Service Technology



Service Technology

- **Service Agents**
- Service agents are event-driven programs designed to intercept messages at runtime.
- There are active and passive service agents, both of which are common in cloud environments.
- Active service agents perform an action upon intercepting and reading the contents of a message. The action typically requires making changes to the message.
- Passive service agents, on the other hand, do not change message contents.

Service Technology

- **Service Middleware**
- Messaging-oriented middleware (MOM) platforms used primarily to facilitate integration, to sophisticated services.
- The two most common types of middleware platforms relevant to services computing are the enterprise service bus (ESB) and the orchestration platform.
- The ESB encompasses a range of intermediary processing features, including service brokerage, routing, and message queuing.
- Orchestration environments are designed to host and execute workflow logic that drives the runtime composition of services.

End of Module 2

Cloud Computing – CSE4001

Course Instructor: Dr. Arunkumar Gopu

Senior Assistant Professor – Grade I

School of Computer Science and Engineering (SCOPE)

VIT – AP University, Amaravati, Andhra Pradesh.

Email-id: arunkumar.gopu@vitap.ac.in

Module 3

Cloud Infrastructure Mechanisms

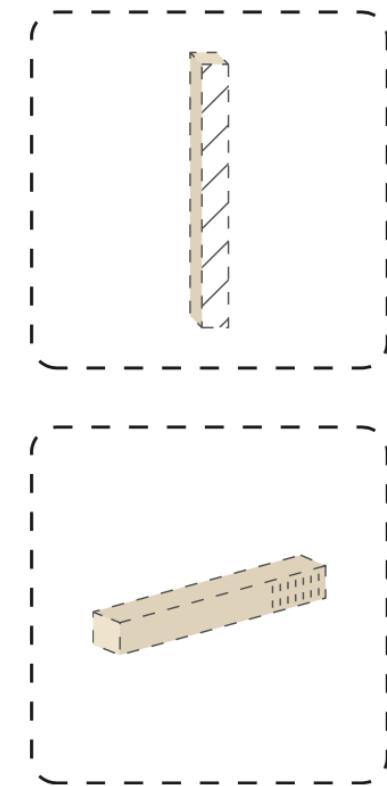
Logical network perimeter, virtual server, cloud storage device, cloud usage monitor, resource replication.

Logical Network Perimeter

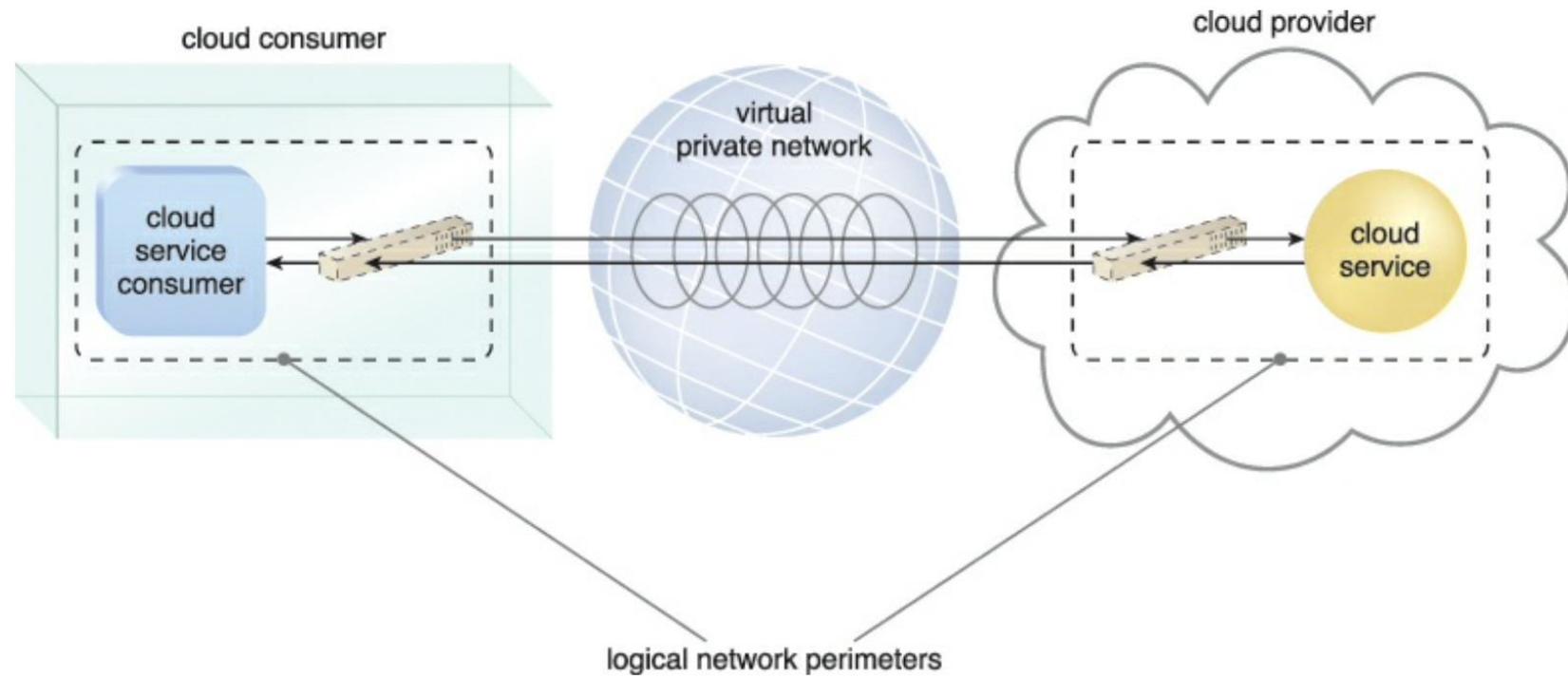
- Logical Network Perimeter is the isolation of a network environment from the rest of a communications network.
- The logical network perimeter establishes a virtual network boundary that can encompass and isolate a group of related cloud-based IT resources.
- This mechanism can be implemented to:
 1. Isolate IT resources in a cloud from non-authorized users.
 2. Isolate IT resources in a cloud from non-users.
 3. Isolate IT resources in a cloud from cloud consumers.
 4. Control the bandwidth that is available to isolated IT resources.

Logical Network Perimeter

- Logical network perimeters are typically established via networking devices
- Virtual Firewall – An IT resource that actively filters network traffic to and from the isolated network while controlling its interactions with the Internet.
- Virtual Network – Usually acquired through VLANs, this IT resource isolates the network environment within the data center infrastructure.

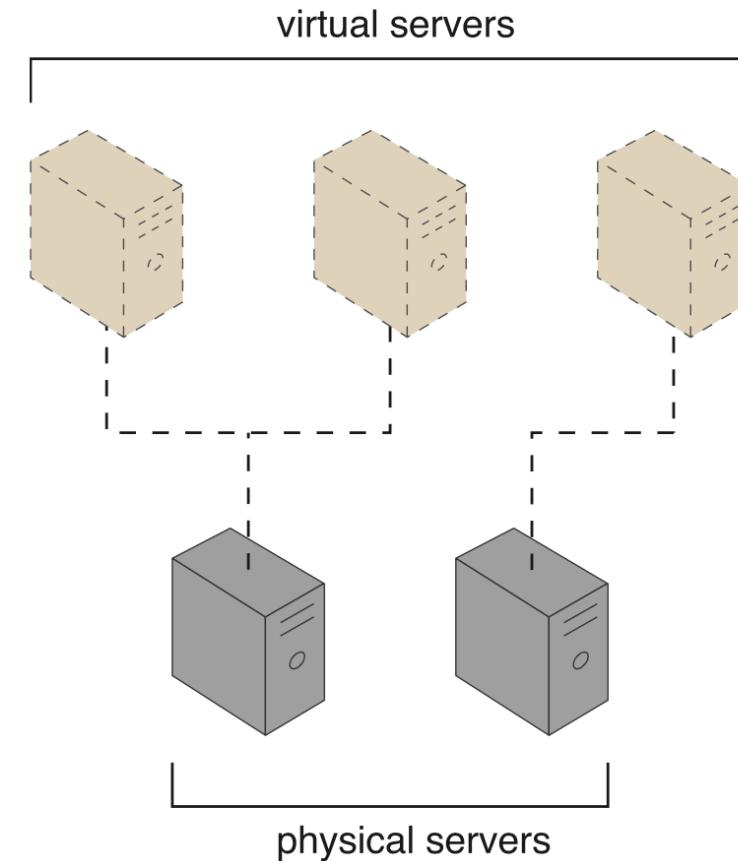


Logical Network Perimeter



Virtual Servers

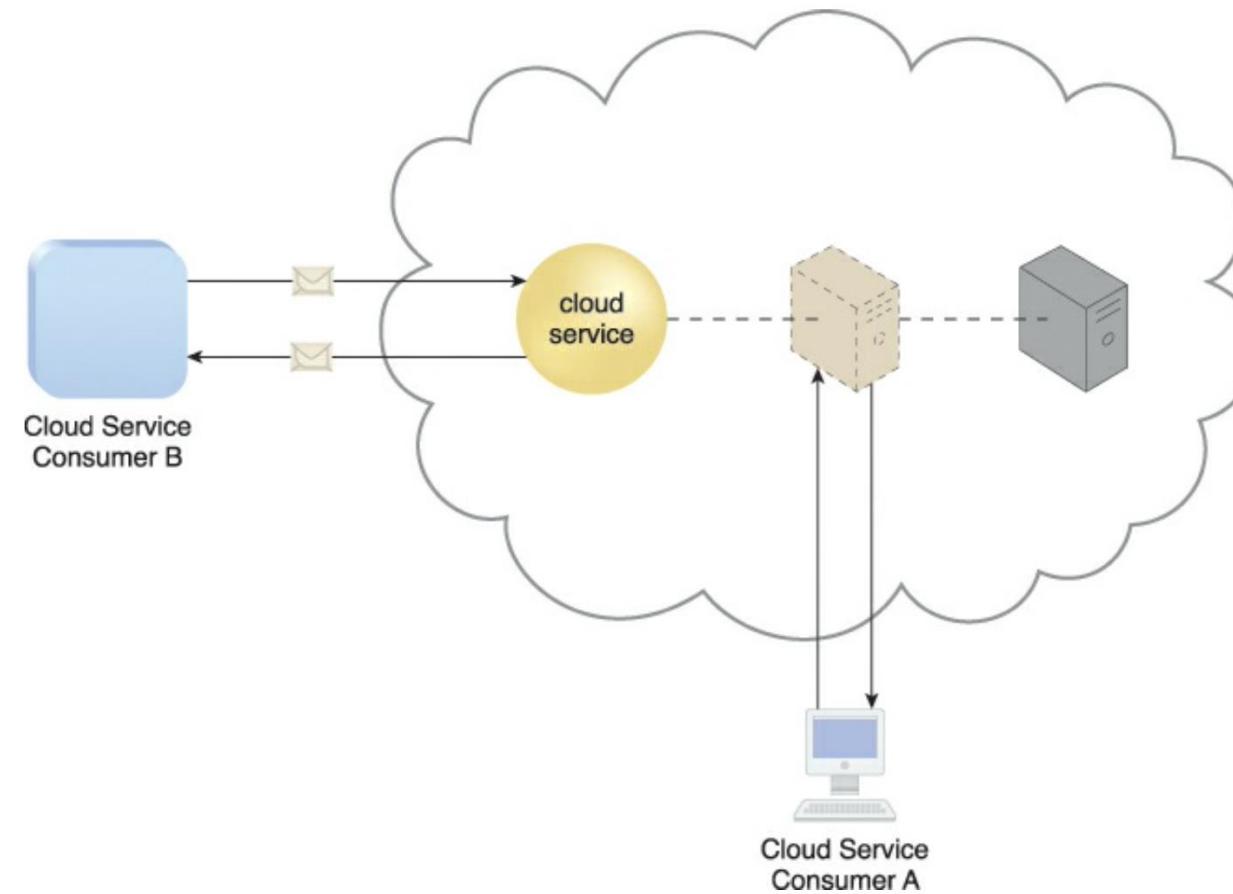
- A virtual server is a form of virtualization software that emulates a physical server.
- Virtual servers are used by cloud providers to share the same physical server with multiple cloud consumers by providing cloud consumers with individual virtual server instances.



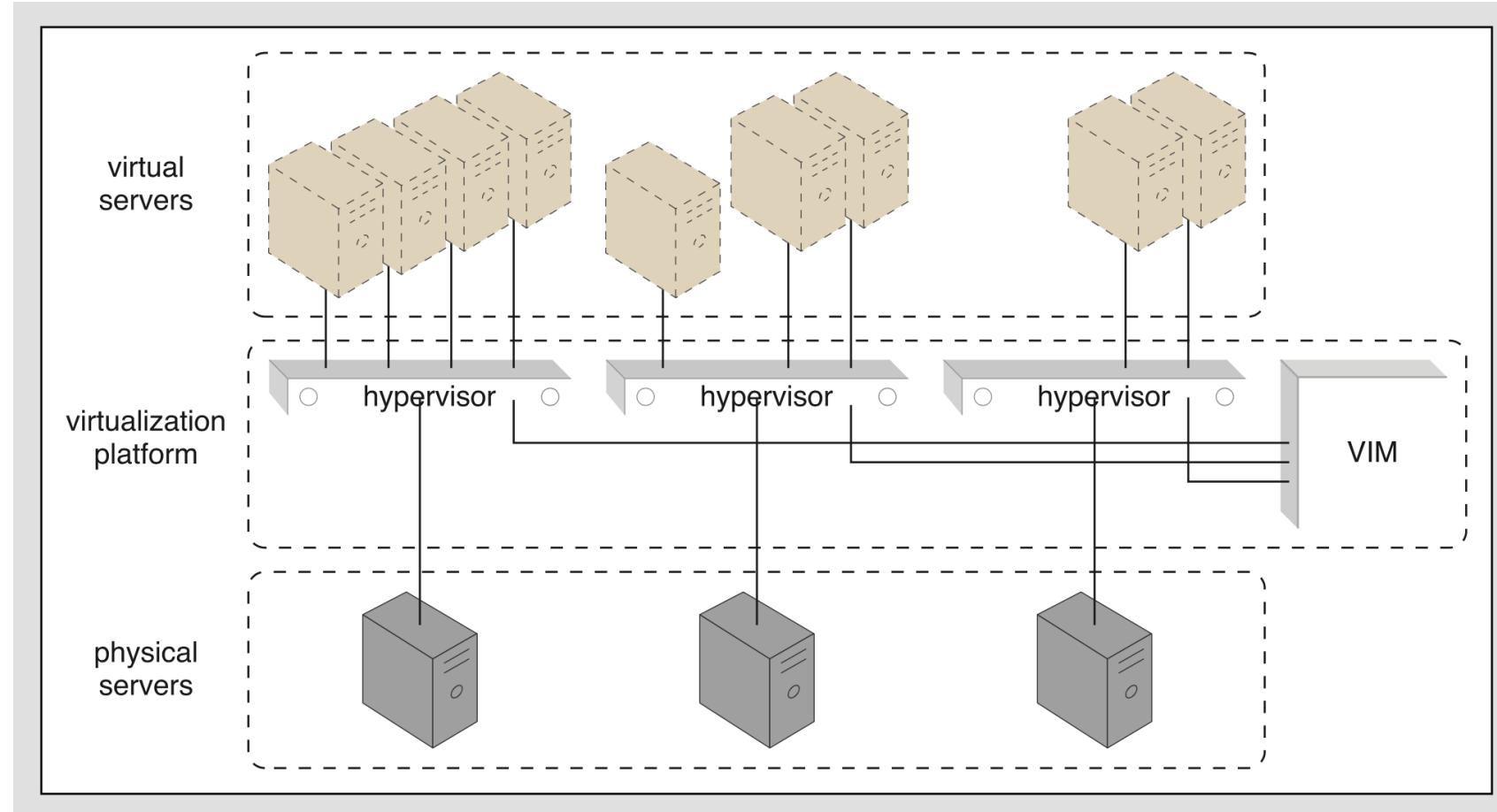
Virtual Servers

- Each virtual server can host numerous IT resources, cloud-based solutions, and various other cloud computing mechanisms.
- The instantiation of virtual servers from image files is a resource allocation process that can be completed rapidly and on-demand.
- Cloud consumers that install or lease virtual servers can customize their environments independently from other cloud consumers.

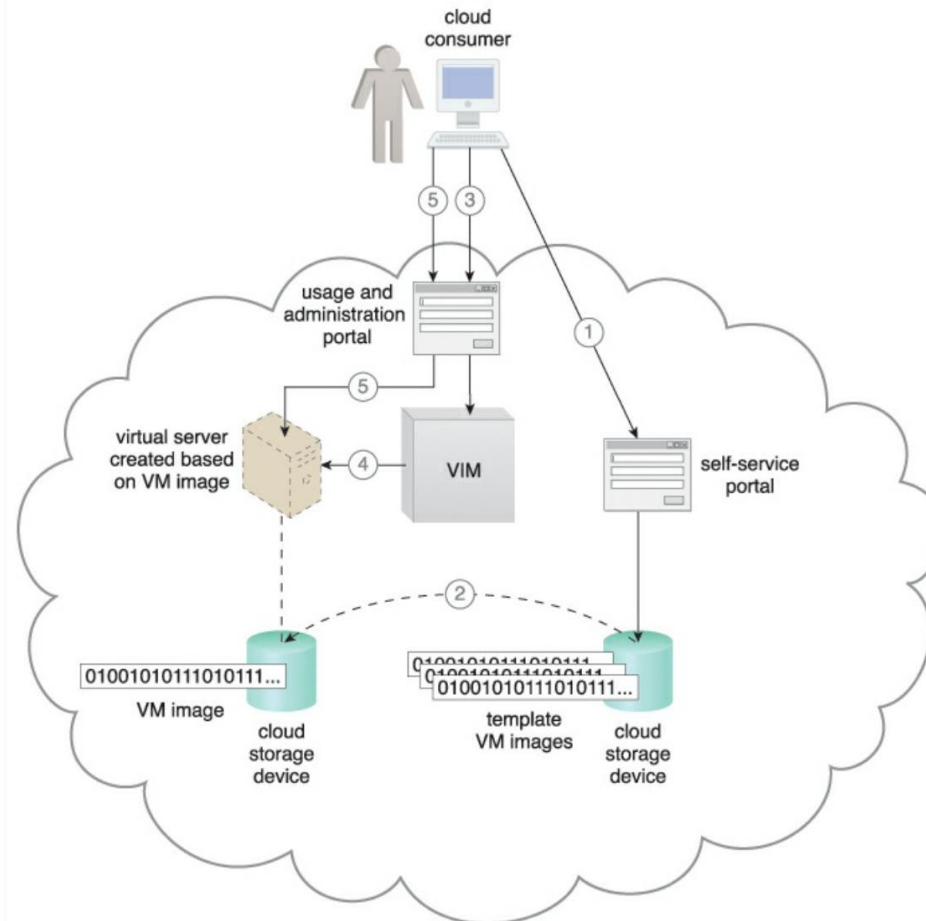
Virtual Servers



Virtual Servers



Virtual Servers



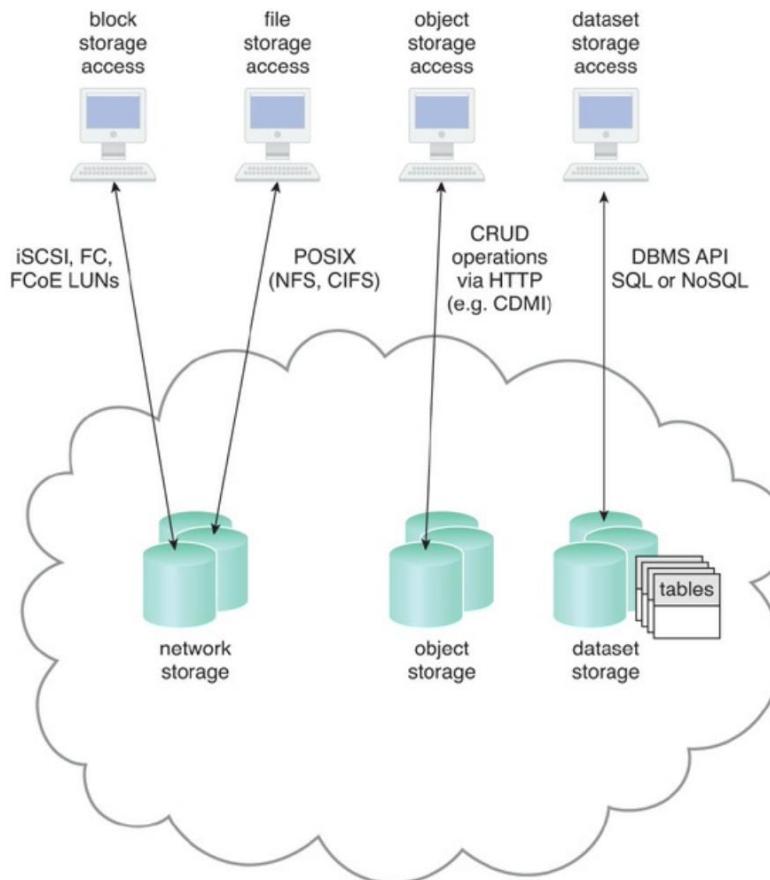
Cloud Storage Device

- The cloud storage device mechanism represents storage devices that are designed specifically for cloud-based provisioning.
- Cloud storage devices are commonly able to provide fixed-increment capacity allocation in support of the pay-per-use mechanism.
- Cloud storage devices can be exposed for remote access via cloud storage services.
- A primary concern related to cloud storage is the security, integrity, and confidentiality of data

Cloud Storage Device

- Another issue applies specifically to the performance of large databases. LANs provide locally stored data with network reliability and latency levels that are superior to those of WANs.
- **Cloud Storage Levels**
- Cloud storage device mechanisms provide common logical units of data storage, such as:
- Files – Collections of data are grouped into files that are located in folders.
- Blocks – The lowest level of storage and the closest to the hardware, a block is the smallest unit of data that is still individually accessible.
- Datasets – Sets of data are organized into a table-based, delimited, or record format.
- Objects – Data and its associated metadata are organized as Web-based resources.

Cloud Storage Device



Cloud Storage Device

- **Network Storage Interfaces**
- Industry standard protocols are used, such as SCSI for storage blocks and the server message block (SMB)
- Common Internet File System (CIFS), and Network File System (NFS) for file and network storage.
- File storage entails storing individual data in separate files that can be different sizes and formats and organized into folders and subfolders.
- Original files are often replaced by the new files that are created when data has been modified.

Cloud Storage Device

- Block storage requires data to be in a fixed format (known as a data block), which is the smallest unit that can be stored and accessed.
- Logical Unit Number (LUN) or virtual volume block-level storage will typically have better performance than file-level storage.
- **Object Storage Interfaces**
- Object storage is based on technologies that can support a range of data and media types.
- Cloud Storage Device mechanisms that implement this interface can typically be accessed via REST or Web service-based cloud services using HTTP as the prime protocol.

Cloud Storage Device

- The Storage Networking Industry Association's Cloud Data Management Interface (SNIA's CDMI) supports the use of object storage interfaces.
- **Database Storage Interfaces**
- Cloud storage device mechanisms based on database storage interfaces typically support a query language in addition to basic storage operations.
- Storage management is carried out using a standard API or an administrative user-interface.
- This classification of storage interface is divided into two main categories according to storage structure.

Cloud Storage Device

- **Relational Data Storage**
- Traditionally, many on-premise IT environments store data using relational databases or relational database management systems (RDBMSs).
- Relational databases rely on tables to organize data into rows and columns.
- Tables can have relationships with each other to give the data increased structure, to protect data integrity, and to avoid data redundancy.
- A cloud storage device mechanism implemented using relational data storage could be based on any number of commercially available database products, such as IBM DB2, Oracle Database, Microsoft SQL Server, and MySQL.

Cloud Storage Device

- Scaling a relational cloud storage device vertically can be more complex and cost-ineffective than horizontal scaling.
- Databases with complex relationships and containing large volumes of data can be afflicted with higher processing overhead and latency.
- **Non-Relational Data Storage**
- NoSQL storage moves away from the traditional relational database model in that it establishes a “looser” structure for stored data.
- The primary motivation for using non-relational storage is to avoid the potential complexity and processing overhead.

Cloud Storage Device

- Non-relational storage can be more horizontally scalable than relational storage.
- The trade-off with non-relational storage is that the data loses much of the native form and validation due to limited or primitive schemas or data models.
- Furthermore, non- relational repositories don't tend to support relational database functions, such as transactions or joins.
- Normalized data exported into a non-relational storage repository will usually become denormalized, meaning that the size of the data will typically grow.

Cloud Storage Device

- Cloud providers often offer non-relational storage that provides scalability and availability of stored data over multiple server environments.
- Non-relational storage mechanisms are proprietary and therefore can severely limit data portability.

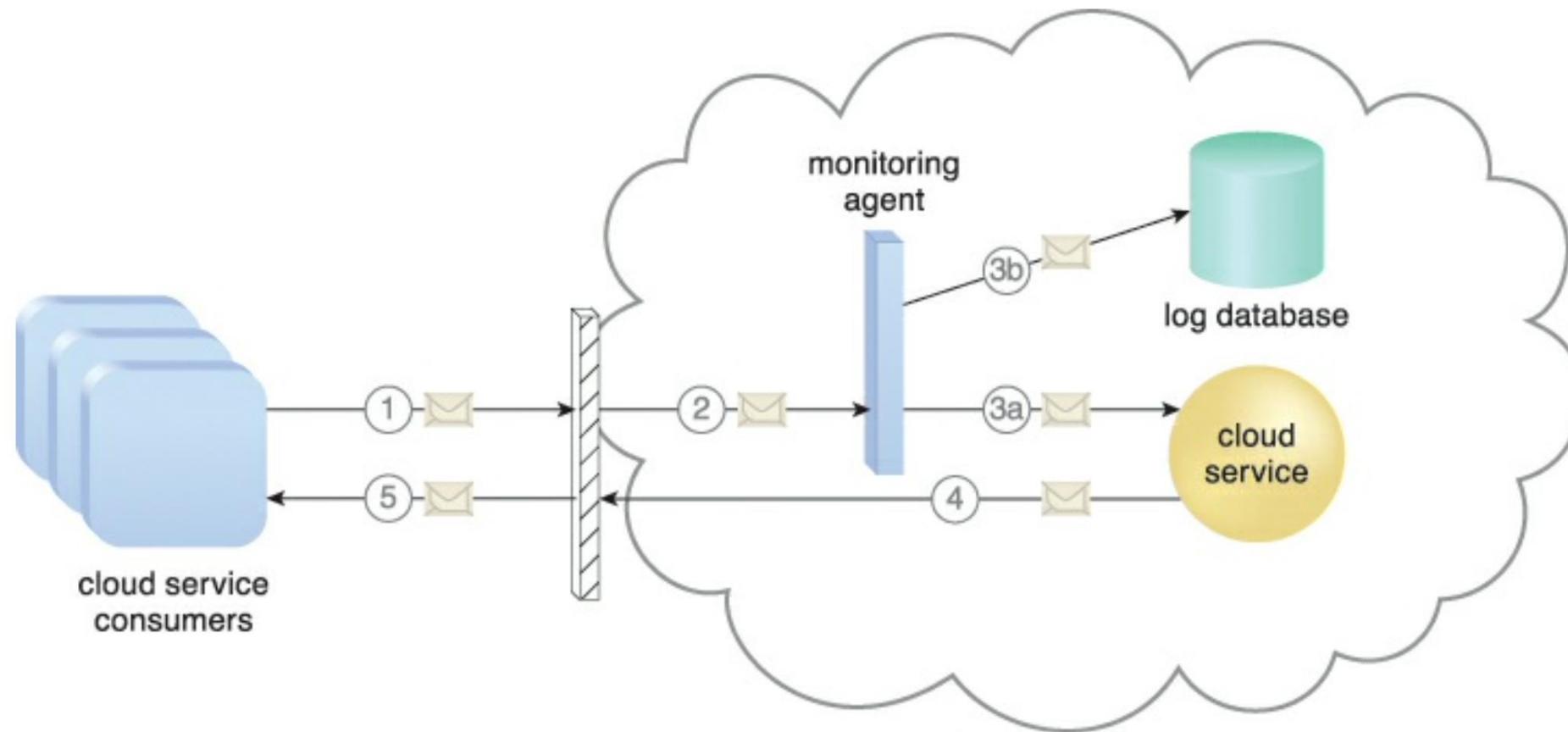
Cloud Usage Monitor

- The cloud usage monitor mechanism is a lightweight and autonomous software program responsible for collecting and processing IT resource usage data.
- Based on the type of usage metrics cloud provider collects usage usage data.
- Cloud usage monitors can exist in different formats. The upcoming sections describe three common agent-based implementation formats..

Cloud Usage Monitor

- **Monitoring Agent**
- A monitoring agent is an intermediary, event-driven program
- This exists as a service agent and resides along existing communication paths to monitor and analyze dataflows.
- This type of cloud usage monitor is commonly used to measure network traffic and message metrics.

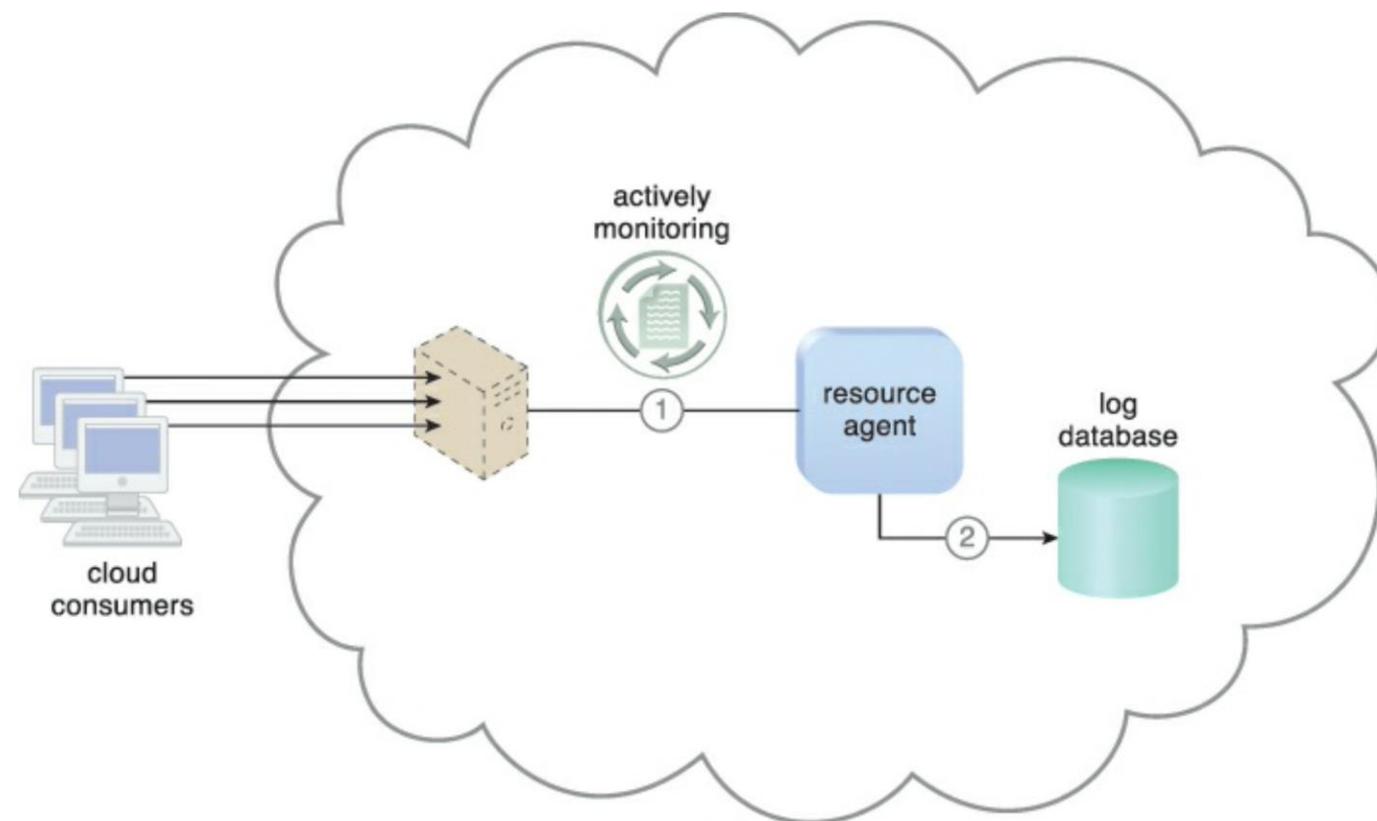
Cloud Usage Monitor



Cloud Usage Monitor

- **Resource Agent**
- A *resource agent* is a processing module that collects usage data by having event-driven interactions with specialized resource software.
- This module is used to monitor usage metrics based on predefined, observable events at the resource software level, such as initiating, suspending, resuming, and vertical scaling.

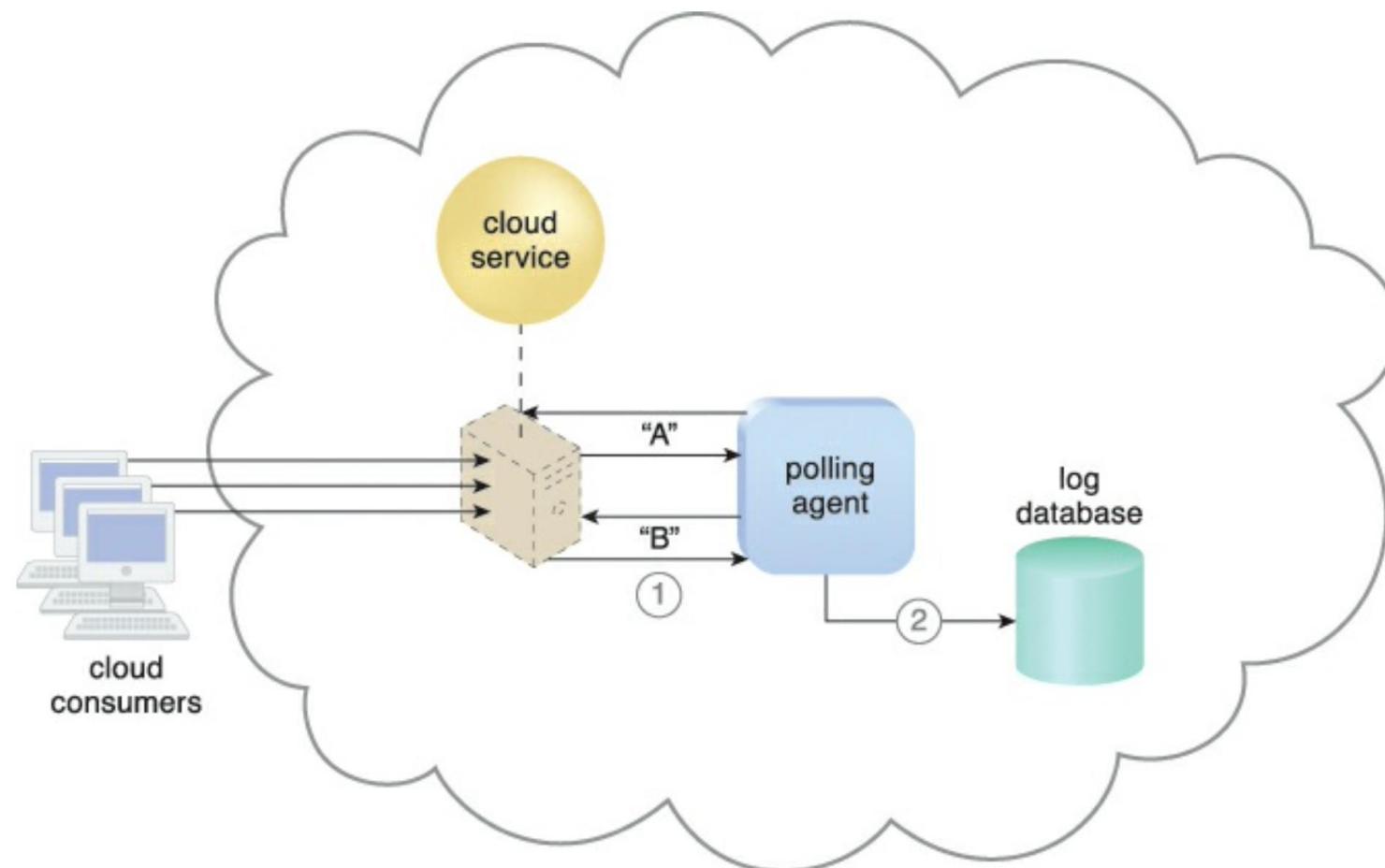
Cloud Usage Monitor



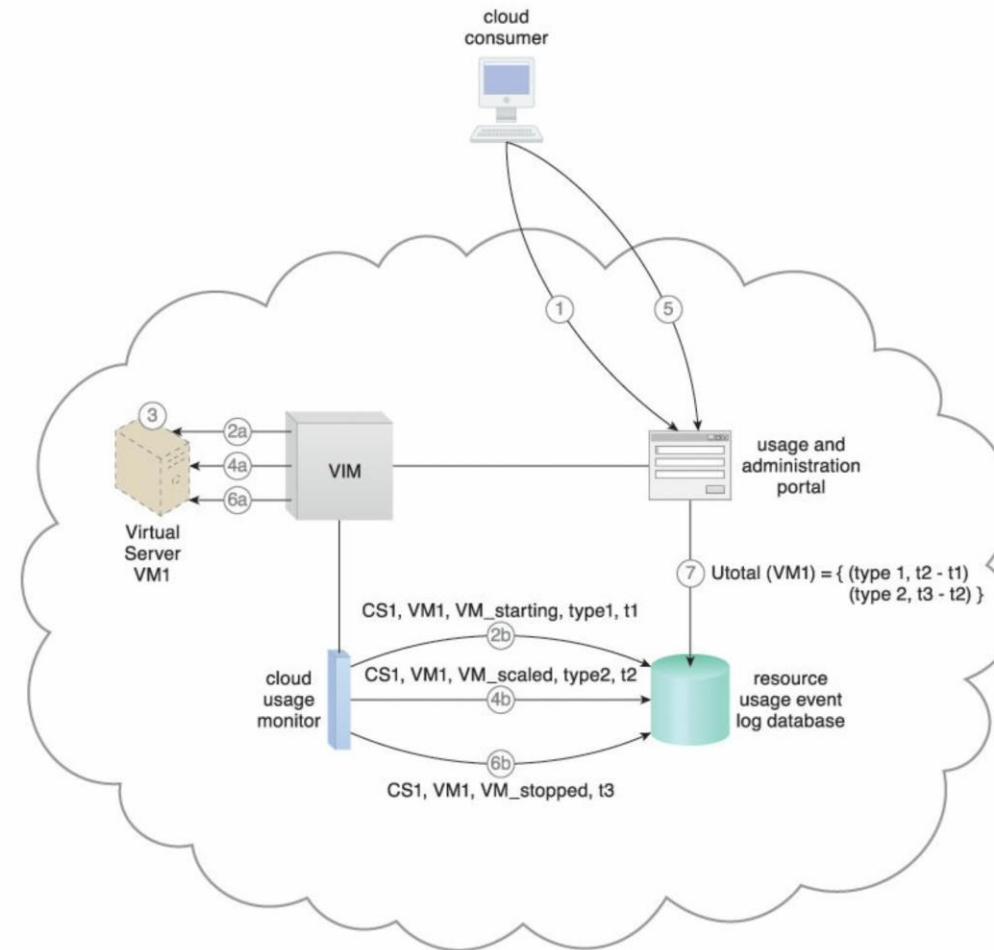
Cloud Usage Monitor

- **Polling Agent**
- A polling agent is a processing module that collects cloud service usage data by polling IT resources.
- This type of cloud service monitor is commonly used to periodically monitor IT resource status, such as uptime and downtime.

Cloud Usage Monitor

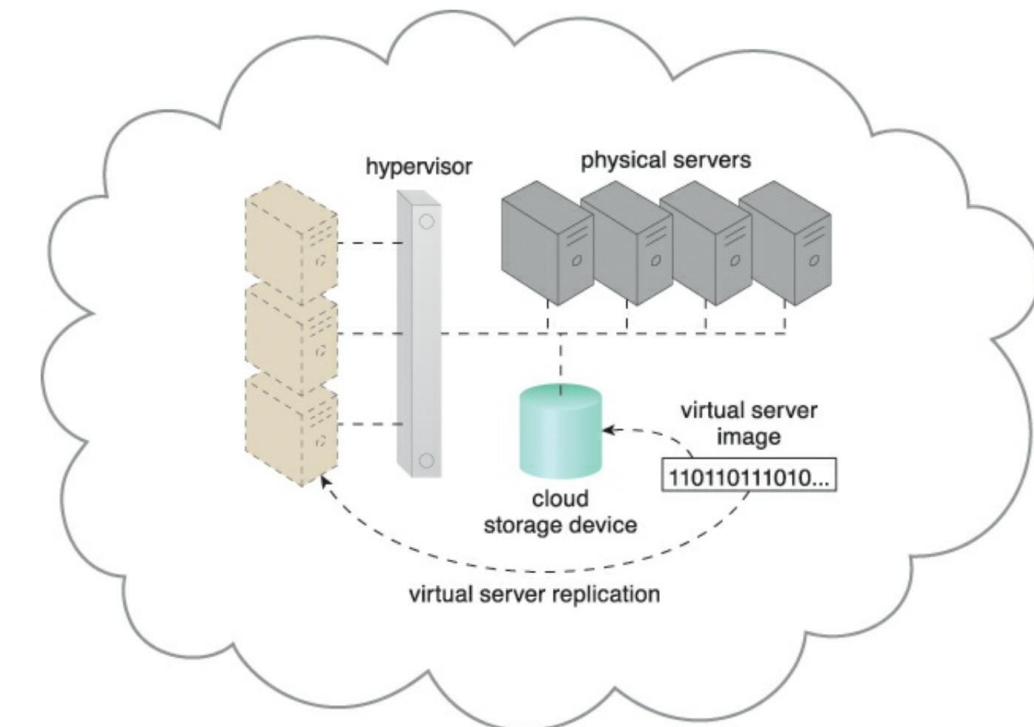


Cloud Usage Monitor



Resource Replication

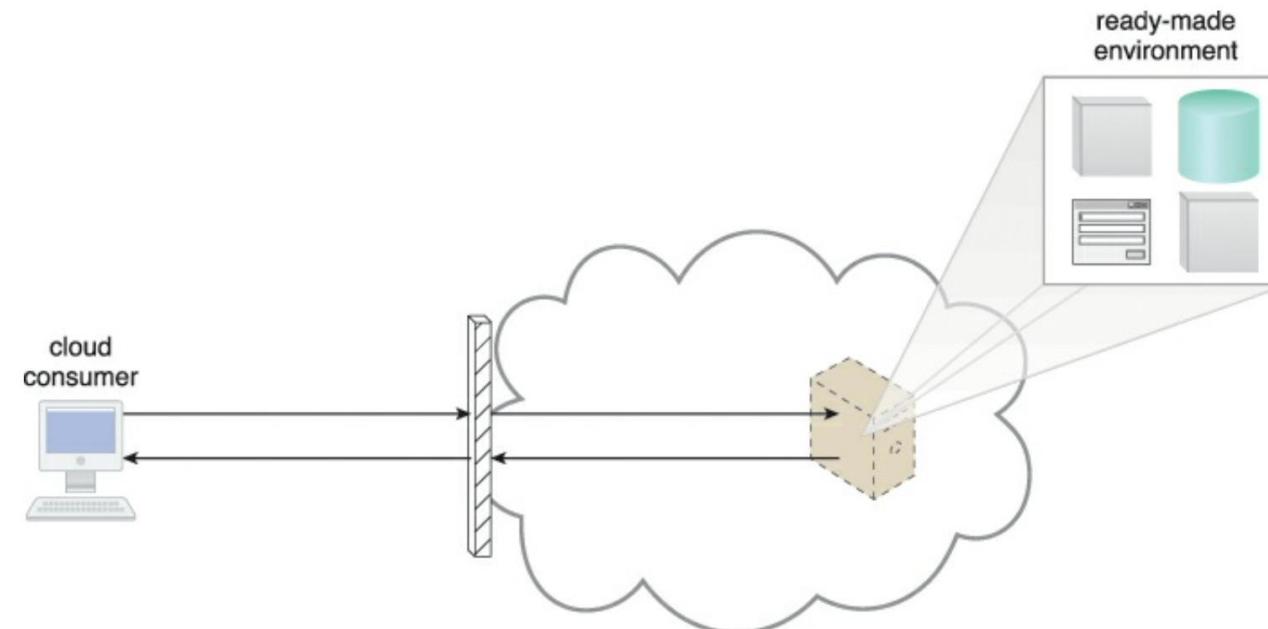
- Creation of multiple instances of the same IT resource
- Replication is typically performed when an IT resource's availability and performance need to be enhanced.
- Virtualization technology is used to implement the resource replication mechanism to replicate cloud-based IT resources.



Ready Made Environment

- The ready-made environment mechanism is a defining component of the PaaS cloud delivery model.
- PaaS represents a pre-defined, cloud-based platform comprised of a set of already installed IT resources, ready to be used and customized by a cloud consumer.
- These environments are utilized by cloud consumers to remotely develop and deploy their own services and applications within a cloud.
- Typical ready-made environments include pre-installed IT resources, such as databases, middleware, development tools, and governance tools.

Ready Made Environment



Ready Made Environment

- A ready-made environment is generally equipped with a complete software development kit (SDK) that provides cloud consumers with programmatic access to the development technologies that comprise their preferred programming stacks.
- Middleware is available for multitenant platforms to support the development and deployment of Web applications.

End of Module 3

Cloud Computing – CSE4001

Course Instructor: Dr. Arunkumar Gopu

Senior Assistant Professor – Grade I

School of Computer Science and Engineering (SCOPE)

VIT – AP University, Amaravati, Andhra Pradesh.

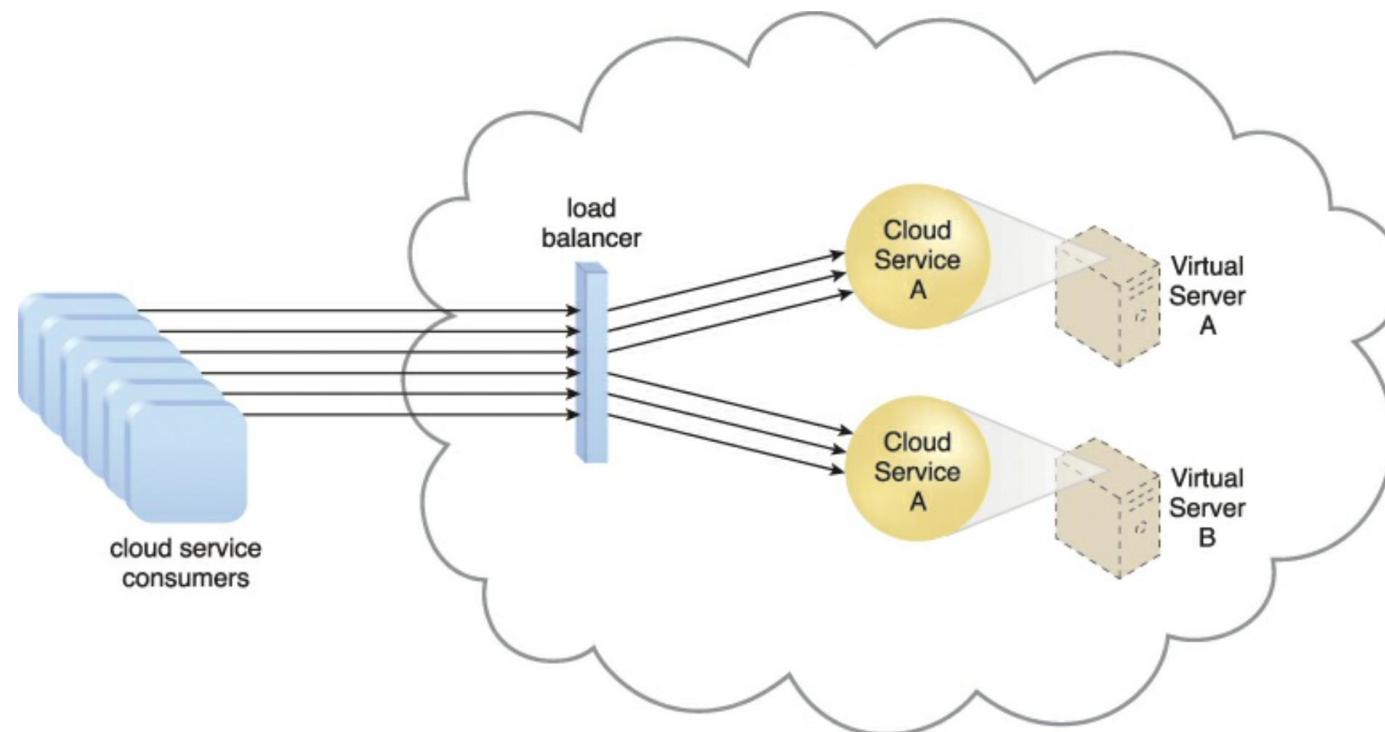
Email-id: arunkumar.gopu@vitap.ac.in

Module 4

Fundamental Cloud Architectures

Workload distribution architecture, resource pooling architecture, dynamic scalability architecture, elastic resource capacity architecture, service load balancing architecture, cloud bursting architecture, elastic disk provisioning architecture, redundant storage architecture, Cloud operations: Migration, static and dynamic Scheduling

Workload Distribution Architecture



Workload Distribution Architecture

- IT resources can be horizontally scaled via the addition of one or more identical IT resources.
- Load balancer provides runtime logic capable of evenly distributing the workload among the available IT resources.
- The resulting workload distribution architecture reduces both IT resource over-utilization and under-utilization to an extent.
- Load Balancing can be applied to any IT resource, with workload distribution commonly among the distributed virtual servers, cloud storage devices, and cloud services.

Workload Distribution Architecture

- The below listed components are involved in workload distributed architecture:
 1. Audit Monitor
 2. Cloud Usage Monitor
 3. Hypervisor
 4. Logical Network Perimeter
 5. Resource Cluster
 6. Resource Replication

Resource Pooling Architecture

- A resource pooling architecture is grouping of identical IT resources and maintained by a system that automatically ensures that they remain synchronized.
- Physical server pools are composed of networked servers that have been installed with operating systems and other necessary programs and/or applications and are ready for immediate use.
- Virtual server pools are configured with several available templates chosen by the cloud consumer during provisioning.

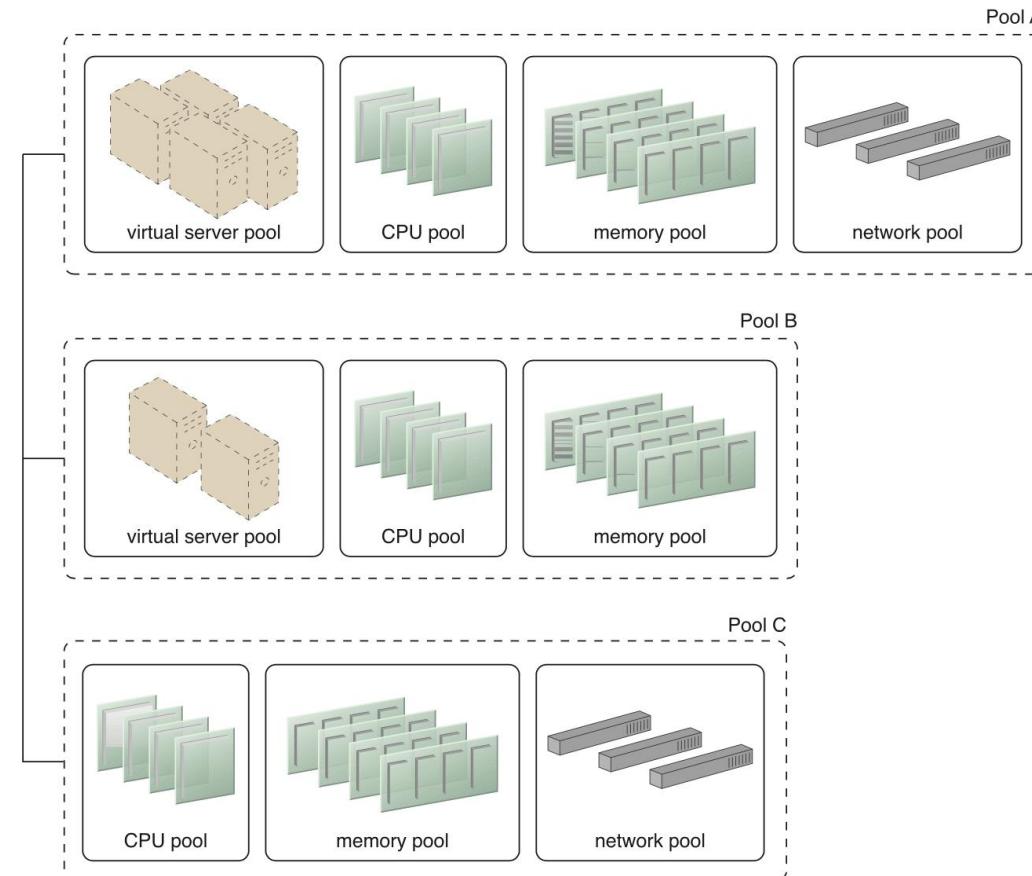
Resource Pooling Architecture

- Storage pools, or cloud storage device pools, consist of file-based or block-based storage structures.
- Network pools (or interconnect pools) are composed of different preconfigured network connectivity devices. For example, a pool of virtual firewall devices or physical network switches can be created for redundant connectivity, load balancing, or link aggregation.
- CPU pools are ready to be allocated to virtual servers and are typically broken down into individual processing cores.

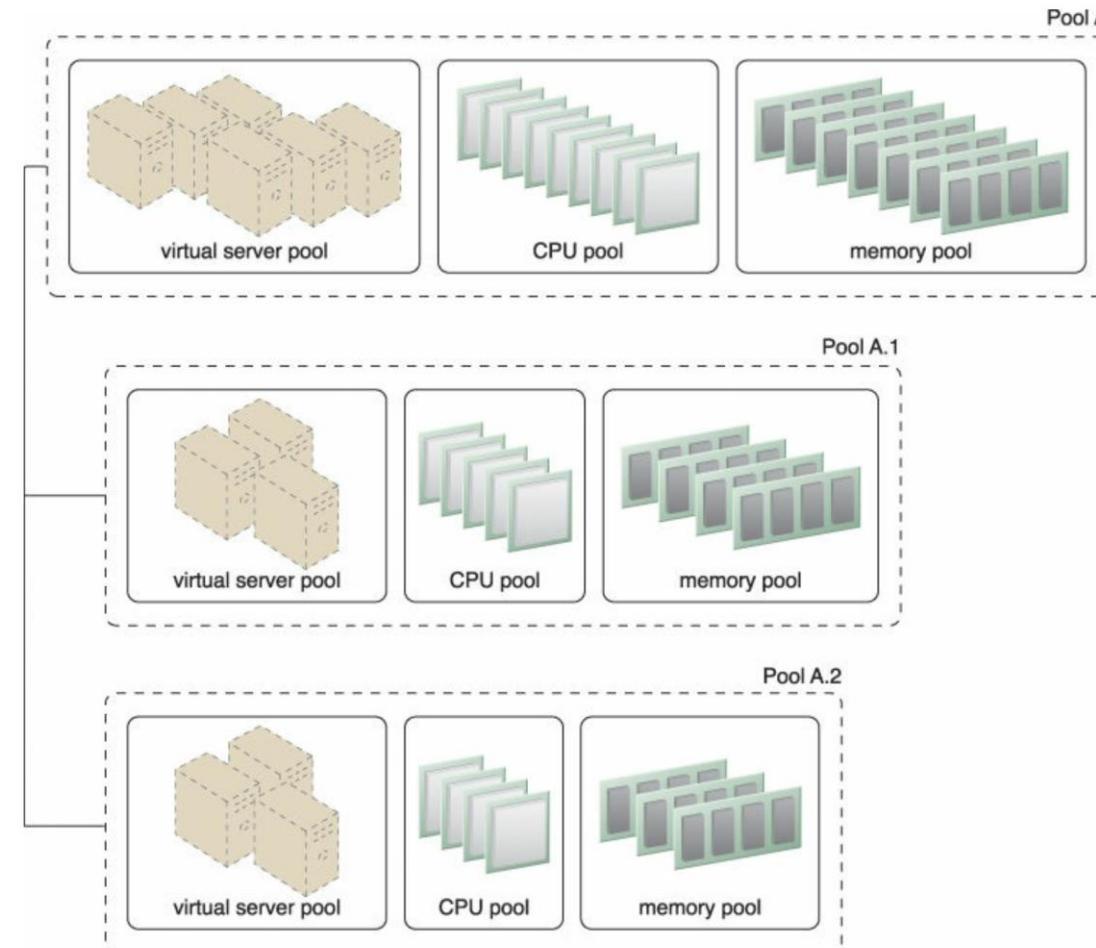
Resource Pooling Architecture

- Pools of physical RAM can be used in newly provisioned physical servers or to vertically scale physical servers.
- Resource pools can become highly complex, with multiple pools created for specific cloud consumers or applications.
- A hierarchical structure can be established to form parent, sibling, and nested pools in order to facilitate the organization of diverse resource pooling requirements.
- In typical case the IT resources pools can spread out over different data centers.

Resource Pooling Architecture



Resource Pooling Architecture



Resource Pooling Architecture

- Audit Monitor
- Cloud Usage Monitor
- Hypervisor
- Logical Network Perimeter
- Pay-Per-Use Monitor
- Remote Administration System
- Resource Management System
- Resource Replication

Dynamic Scaling Architecture

- The dynamic scalability architecture model is based on a **system of pre-defined scaling conditions** that trigger the dynamic allocation of IT resources from resource pools.
- Dynamic allocation enables variable utilization as dictated by usage demand fluctuations, since unnecessary IT resources are efficiently reclaimed without requiring manual interaction.
- The **automated scaling listener** is configured with workload thresholds that dictate when new IT resources need to be added to the workload processing.

Dynamic Scaling Architecture

- This mechanism can be provided with logic that determines how many additional IT resources can be dynamically provided, based on the terms of a given cloud consumer's provisioning contract.
- Dynamic Horizontal Scaling – IT resource instances are scaled out and in to handle fluctuating workloads.
- The automatic scaling listener monitors requests and signals resource replication to initiate IT resource duplication, as per requirements and permissions.

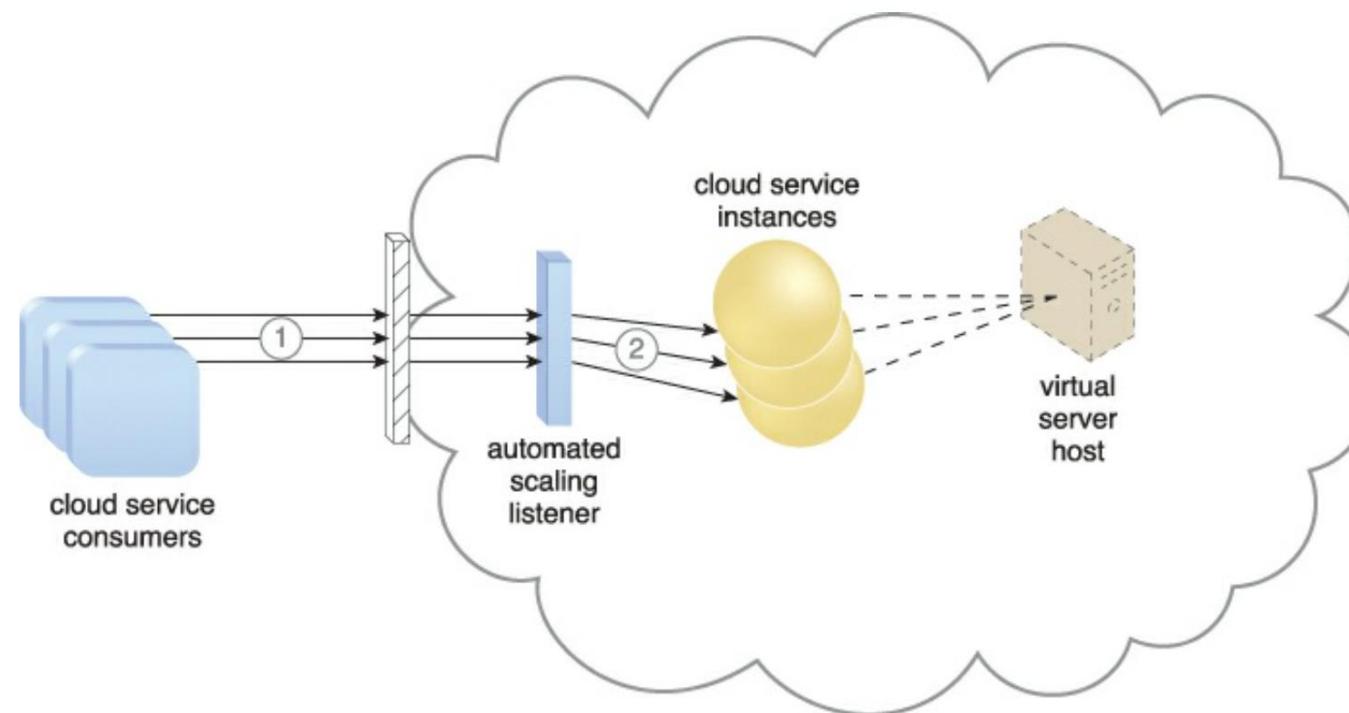
Dynamic Scaling Architecture

- Dynamic Vertical Scaling – IT resource instances are scaled up and down when there is a need to adjust the processing capacity of a single IT resource.
- For example, a virtual server that is being overloaded can have its memory dynamically increased or it may have a processing core added.

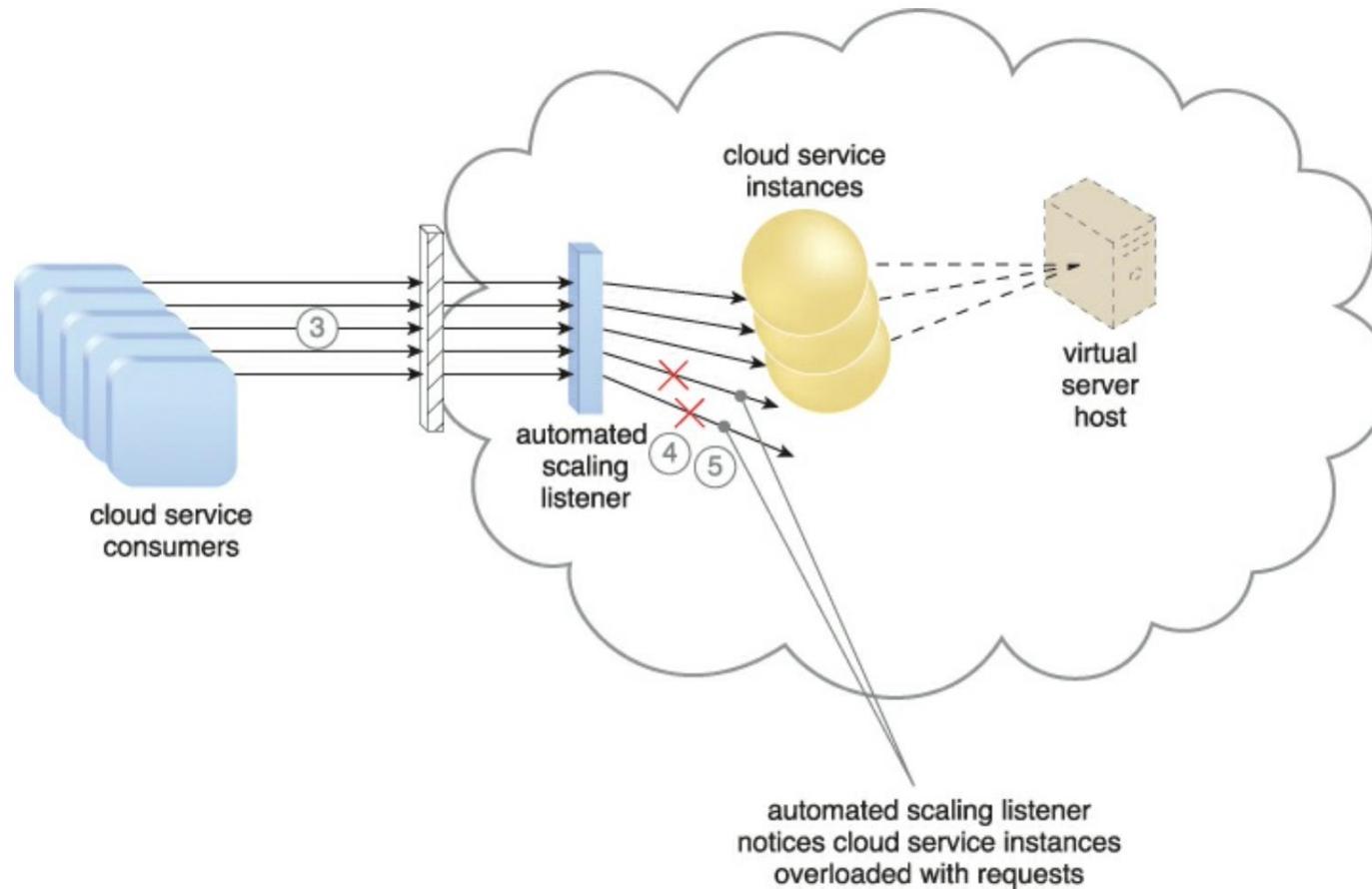
Dynamic Scaling Architecture

- Dynamic Relocation – The IT resource is relocated to a host with more capacity.
- For example, a database may need to be moved from a tape-based SAN storage device with 4 GB per second I/O capacity to another disk-based SAN storage device with 8 GB per second I/O capacity.
- Cloud Usage Monitor
- Hypervisor
- Pay-Per-Use Monitor

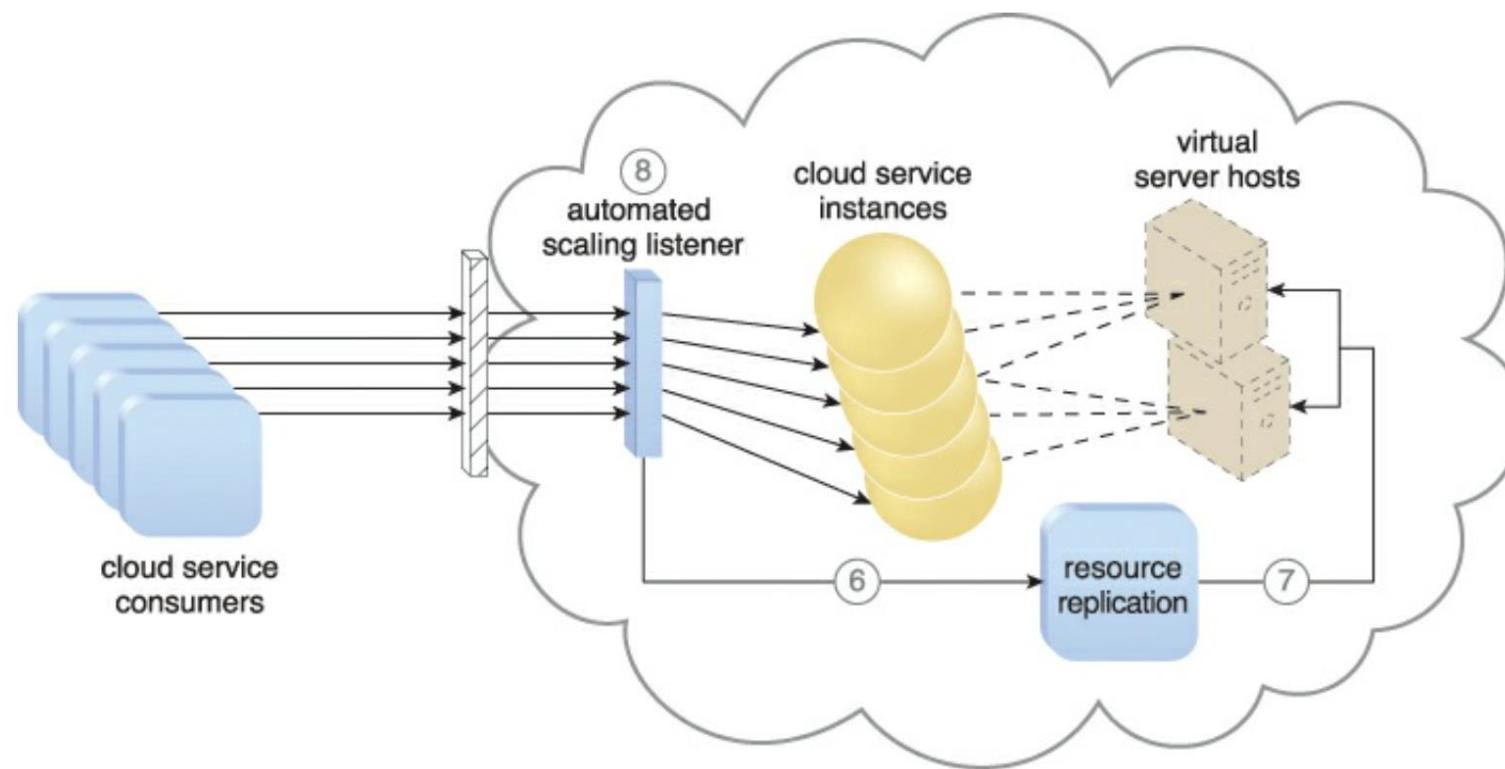
Dynamic Scaling Architecture



Dynamic Scaling Architecture



Dynamic Scaling Architecture

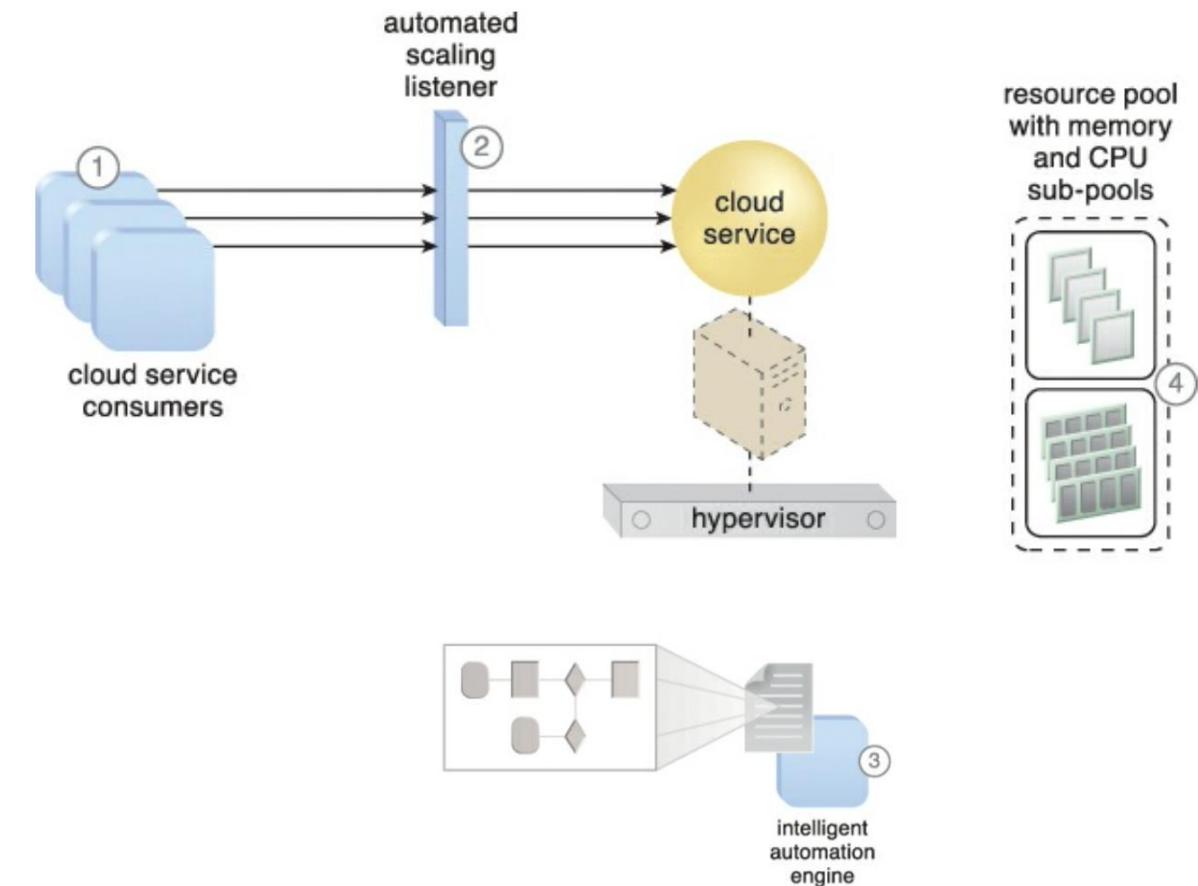


Elastic Resource Capacity Architecture

- The elastic resource capacity architecture is primarily related to the **dynamic provisioning of virtual servers**, using a system that allocates and reclaims CPUs and RAM in immediate response to the fluctuating processing requirements of hosted IT resources.
- Resource pools are used by scaling technology that interacts with the hypervisor and/or VIM to retrieve and return CPU and RAM resources at runtime.

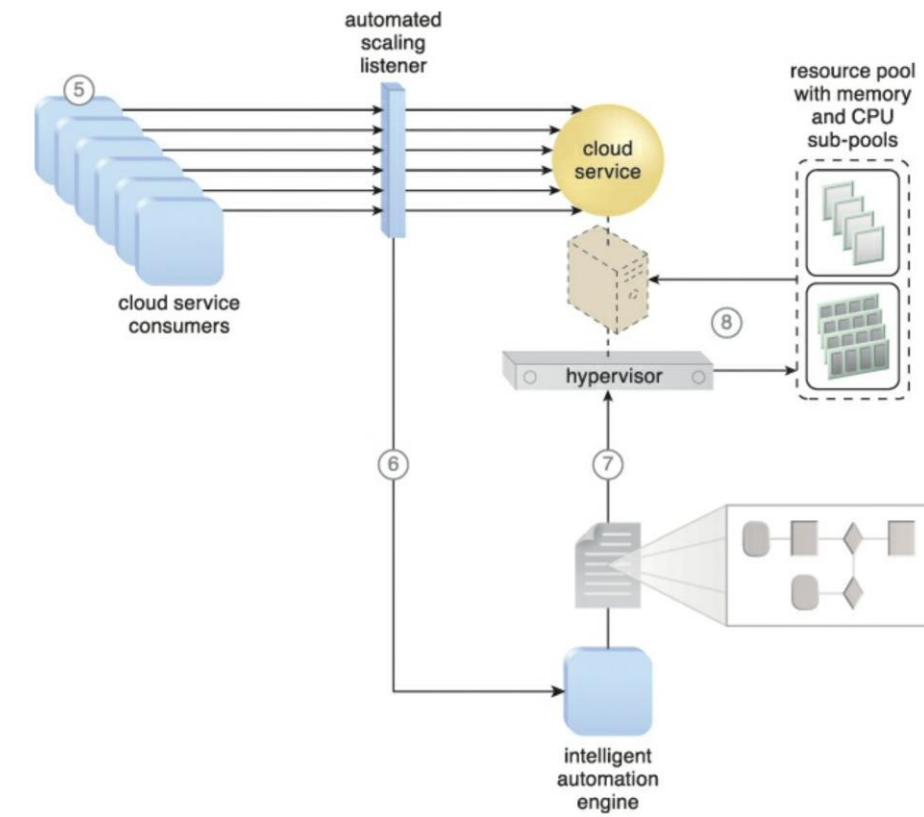
Elastic Resource Capacity Architecture

- Cloud service consumers are actively sending requests to a cloud service (1), which are monitored by an automated scaling listener (2). An intelligent automation engine script is deployed with workflow logic (3) that is capable of notifying the resource pool using allocation requests (4).



Elastic Resource Capacity Architecture

- Cloud service consumer requests increase (5), causing the automated scaling listener to signal the intelligent automation engine to execute the script (6). The script runs the workflow logic that signals the hypervisor to allocate more IT resources from the resource pools (7). The hypervisor allocates additional CPU and RAM to the virtual server, enabling the increased workload to be handled (8).



Elastic Resource Capacity Architecture

- The runtime processing of the virtual server is monitored so that additional processing power can be leveraged from the resource pool via dynamic allocation, before capacity thresholds are met.
- The virtual server and its hosted applications and IT resources are vertically scaled in response.
- This type of cloud architecture can be designed so that the intelligent automation engine script sends its scaling request via the VIM instead of to the hypervisor directly.

Elastic Resource Capacity Architecture

- Virtual servers that participate in elastic resource allocation systems may require rebooting for the dynamic resource allocation to take effect.
- The intelligent automation engine automates administration tasks by executing scripts that contain workflow logic.

Service Load Balancing Architecture

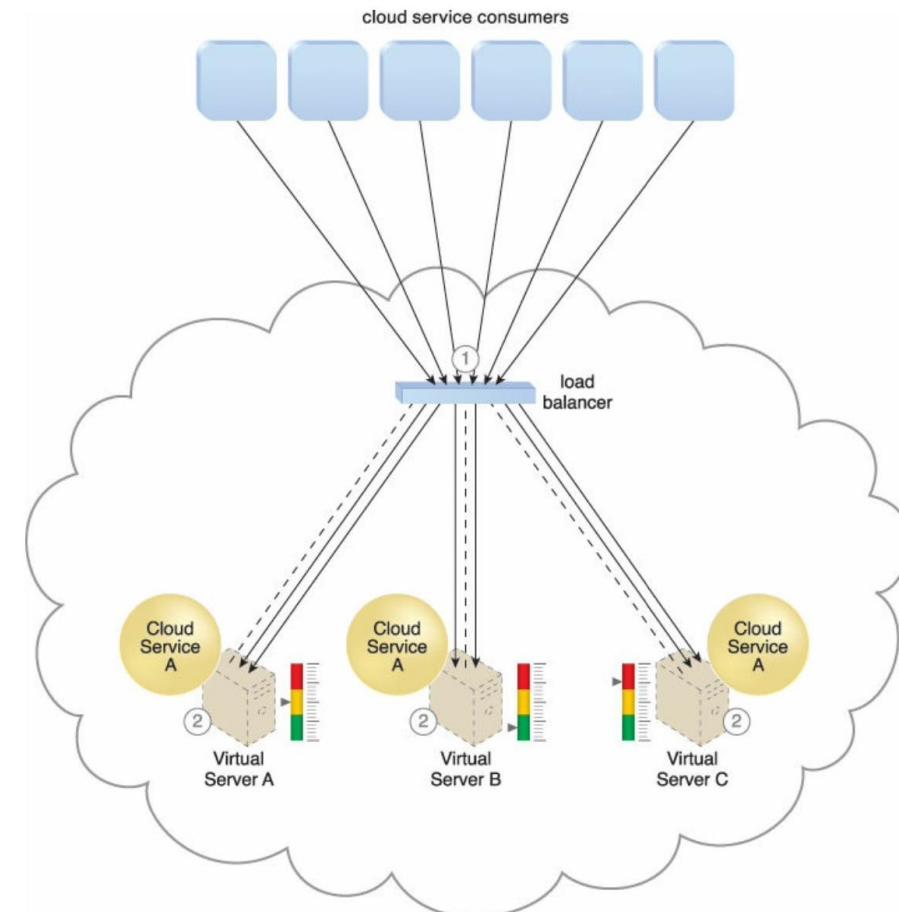
- The service load balancing architecture can be considered a specialized variation of the workload distribution architecture that is geared specifically for scaling cloud service implementations.
- Redundant deployments of cloud services are created, with a load balancing system added to dynamically distribute workloads.
- The duplicate cloud service implementations are organized into a resource pool, while the load balancer is positioned as either an external or built-in component to allow the host servers to balance the workloads themselves.

Service Load Balancing Architecture

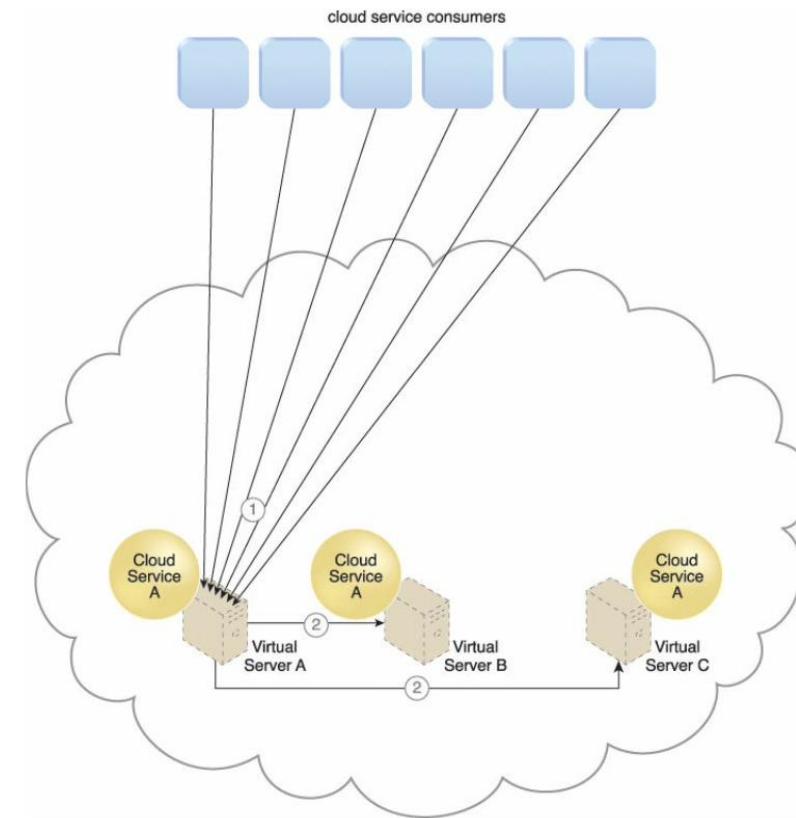
- Depending on the anticipated workload and processing capacity of host server environments, multiple instances of each cloud service implementation can be generated as part of a resource pool that responds to fluctuating request volumes more efficiently.
- The service load balancing architecture can involve the following mechanisms in addition to the load balancer:
 1. Cloud Usage Monitor – Cloud usage monitors may be involved with monitoring cloud service instances and their respective IT resource consumption levels, as well as various runtime monitoring and usage data collection tasks.
 2. Resource Cluster – Active-active cluster groups are incorporated in this architecture to help balance workloads across different members of the cluster.

Service Load Balancing Architecture

3. Resource Replication -
The resource replication mechanism is utilized to generate cloud service implementations in support of load balancing requirements.



Service Load Balancing Architecture



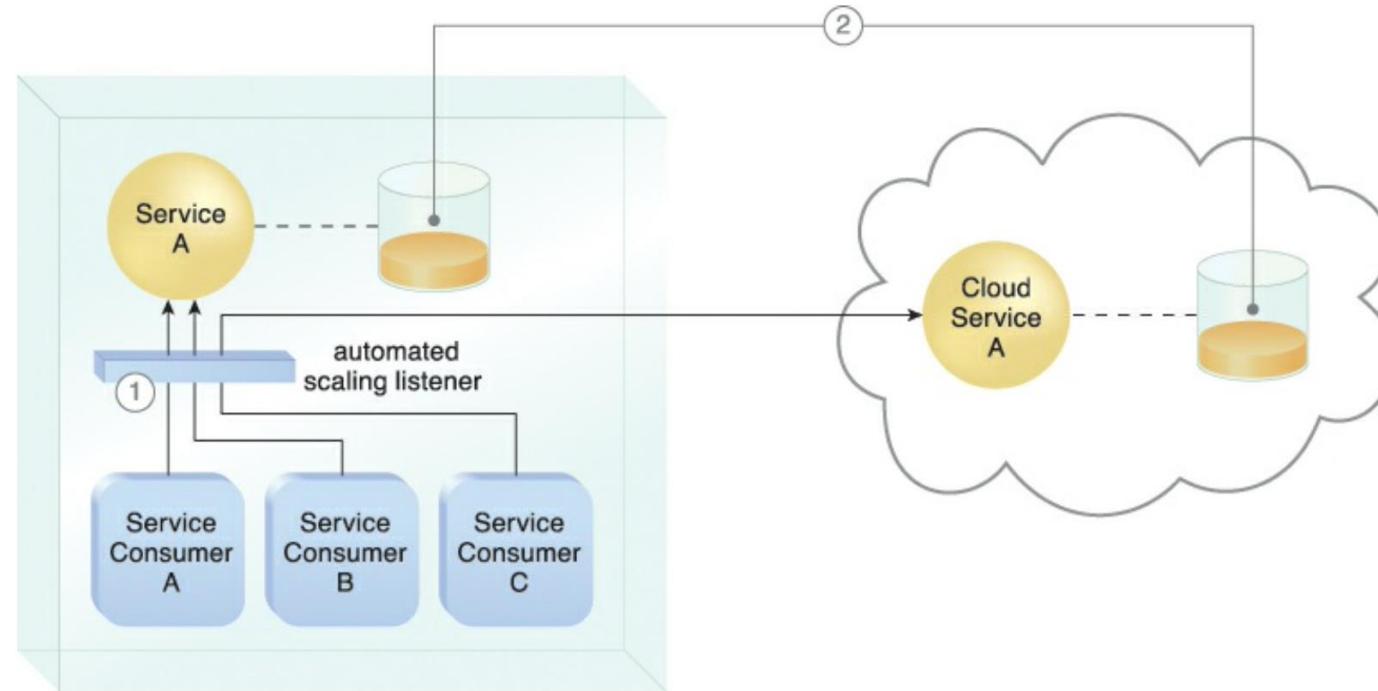
Cloud Bursting Architecture

- The cloud bursting architecture establishes a form of dynamic scaling that scales or “bursts out” on-premise IT resources into a cloud whenever predefined capacity thresholds have been reached.
- The corresponding cloud- based IT resources are redundantly pre-deployed but **remain inactive until cloud bursting occurs**.
- After they are no longer required, the cloud-based IT resources are released and the architecture “bursts in” back to the on-premise environment.

Cloud Bursting Architecture

- Cloud bursting is a flexible scaling architecture that provides cloud consumers with the option of using cloud- based IT resources only to meet higher usage demands.
- The foundation of this architectural model is based on the automated scaling listener and resource replication mechanisms.
- The automated scaling listener determines when to redirect requests to cloud-based IT resources, and resource replication is used to maintain synchronicity between on-premise and cloud-based IT resources in relation to state information.

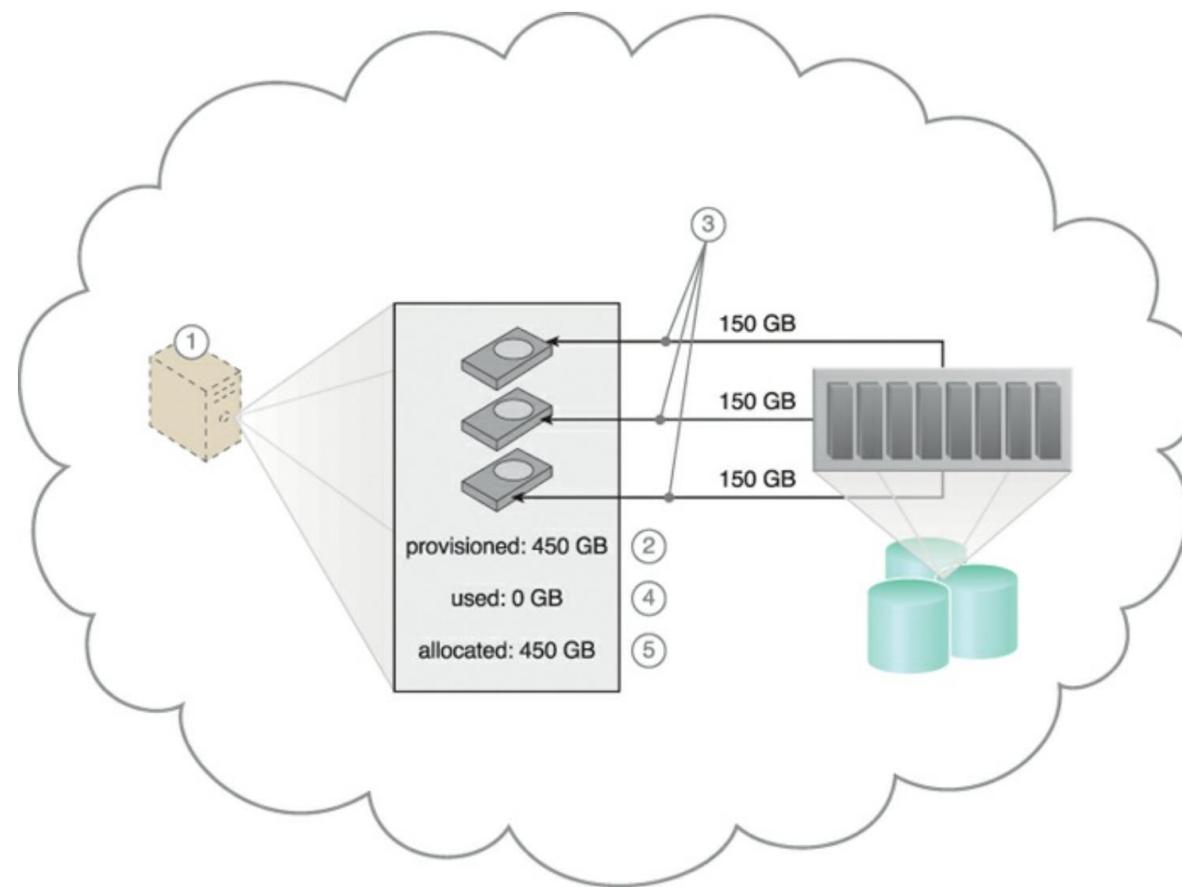
Cloud Bursting Architecture



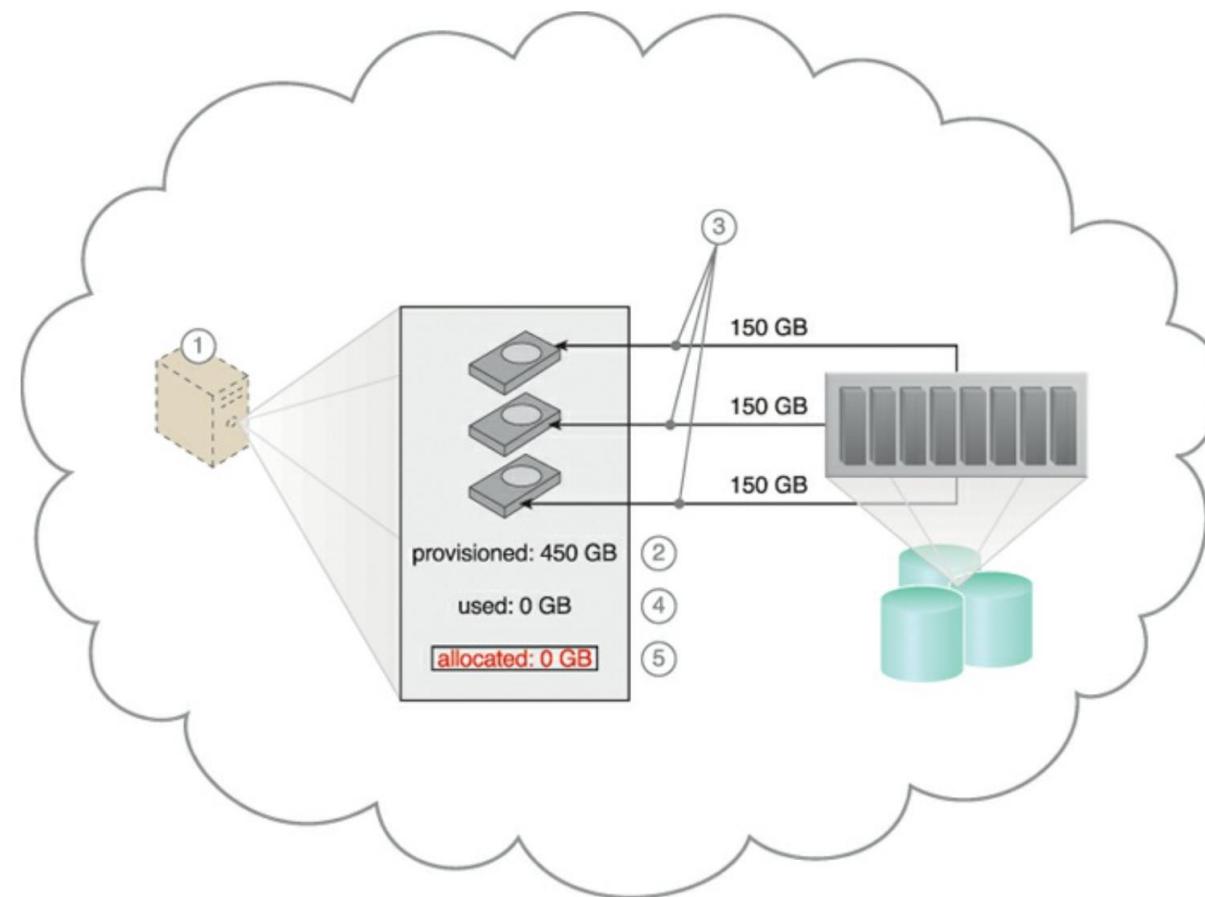
Elastic Disk Provisioning Architecture

- Cloud consumers are commonly charged for cloud-based storage space based on fixed-disk storage allocation, meaning the charges are predetermined by disk capacity and not aligned with actual data storage consumption.
- The elastic disk provisioning architecture establishes a dynamic storage provisioning system that ensures that the cloud consumer is granularly billed for the exact amount of storage that it uses.
- This system uses thin- provisioning technology for the dynamic allocation of storage space, and is further supported by runtime usage monitoring to collect accurate usage data for billing purposes

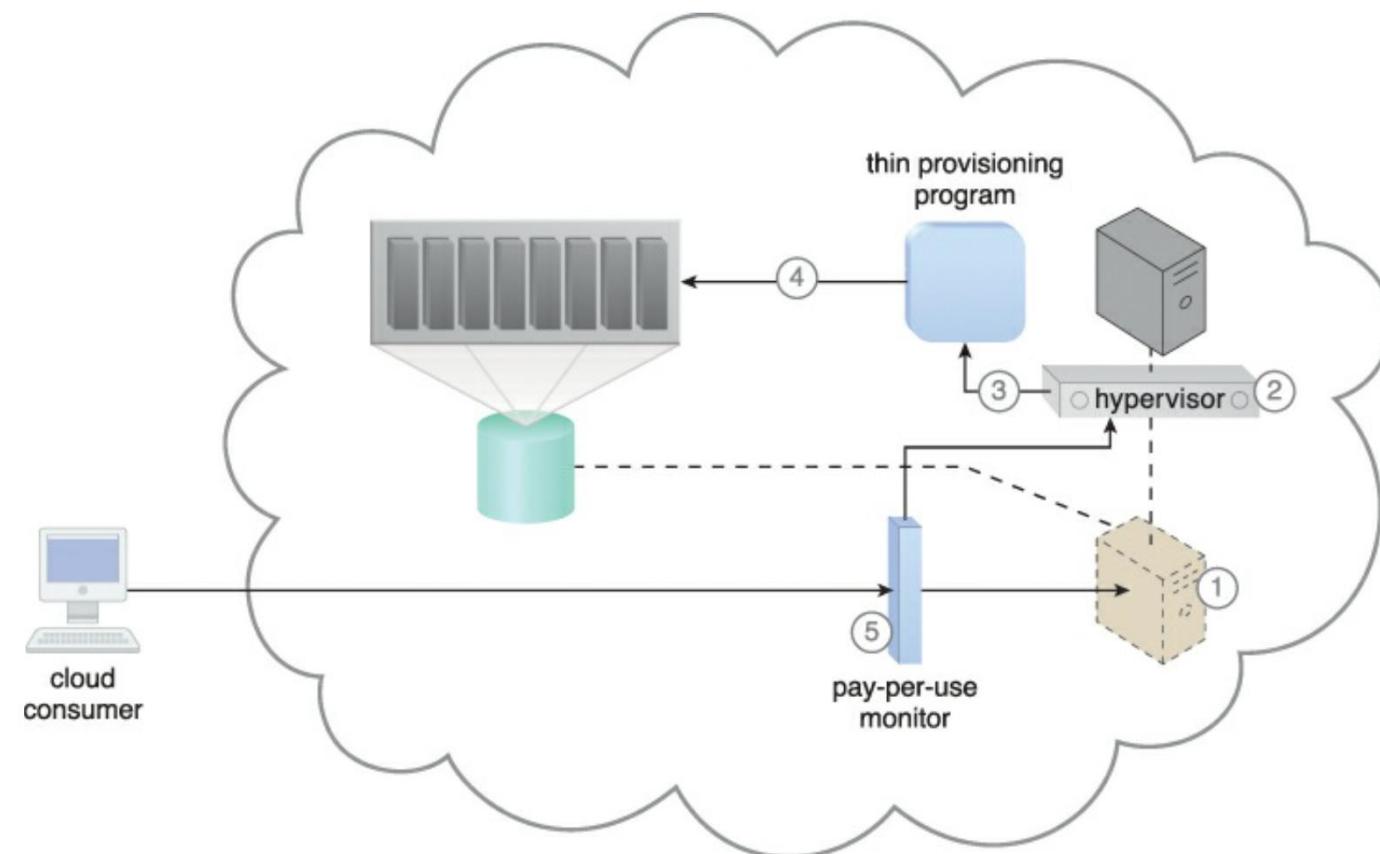
Elastic Disk Provisioning Architecture



Elastic Disk Provisioning Architecture



Elastic Disk Provisioning Architecture

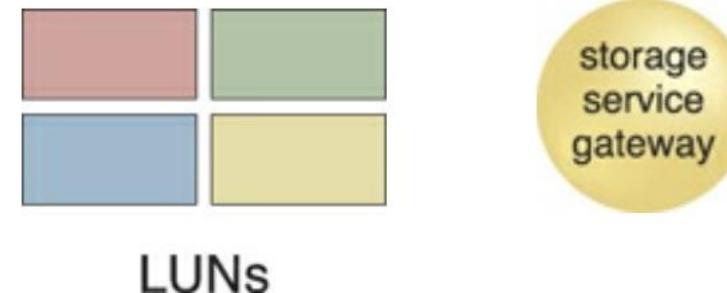


Elastic Disk Provisioning Architecture

- The following mechanisms can be included in this architecture in addition to the
- Cloud storage device.
- Virtual server.
- Hypervisor.
- Pay-per-use monitor.
- Cloud Usage Monitor.
- Resource Replication.

Redundant Storage Architecture

- Cloud storage devices are occasionally subject to failure and disruptions that are caused by network connectivity issues, controller or general hardware failure, or security breaches.
- A compromised cloud storage device's reliability can have a ripple effect and cause impact failure across services and availability is compromised.



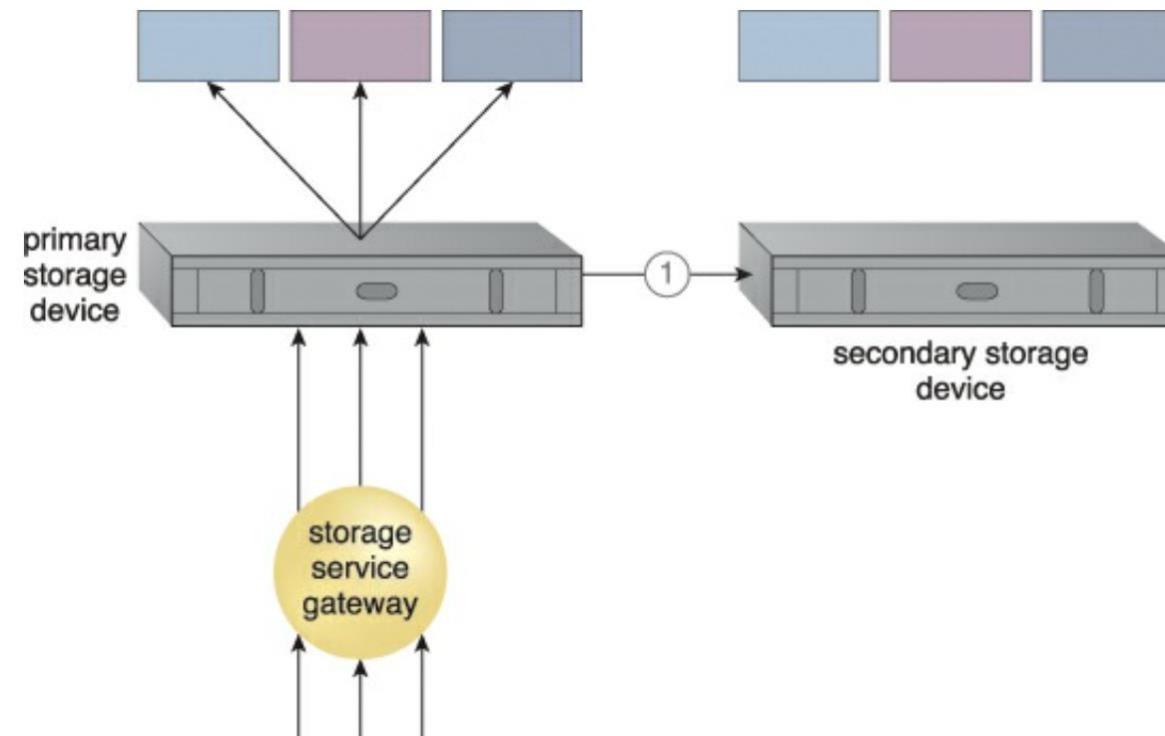
Redundant Storage Architecture

- A logical unit number (LUN) is a logical drive that represents a partition of a physical drive.
- The storage service gateway is a component that acts as the external interface to cloud storage services and is capable of automatically redirecting cloud consumer requests whenever the location of the requested data has changed.
- The redundant storage architecture introduces a secondary duplicate cloud storage device as part of a failover system that synchronizes its data with the data in the primary cloud storage device.

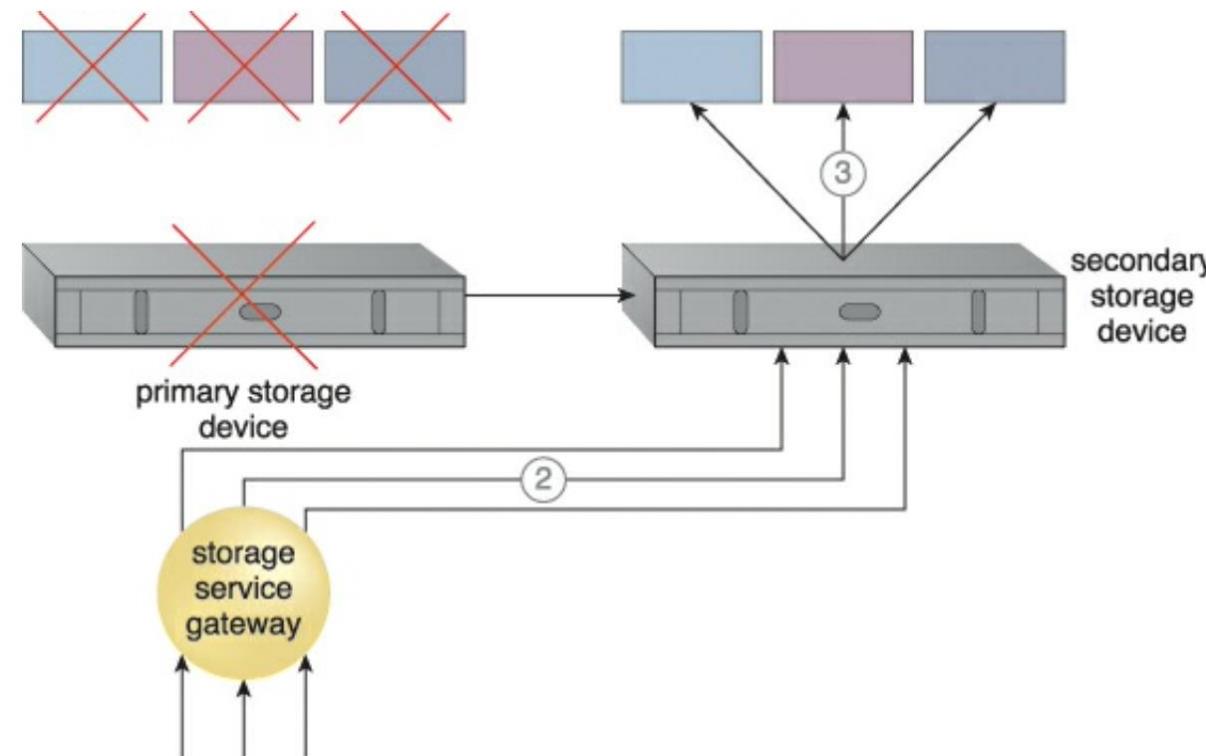
Redundant Storage Architecture

- A storage service gateway diverts cloud consumer requests to the secondary device whenever the primary device fails.
- This cloud architecture primarily relies on a storage replication system that keeps the primary cloud storage device synchronized with its duplicate secondary cloud storage devices

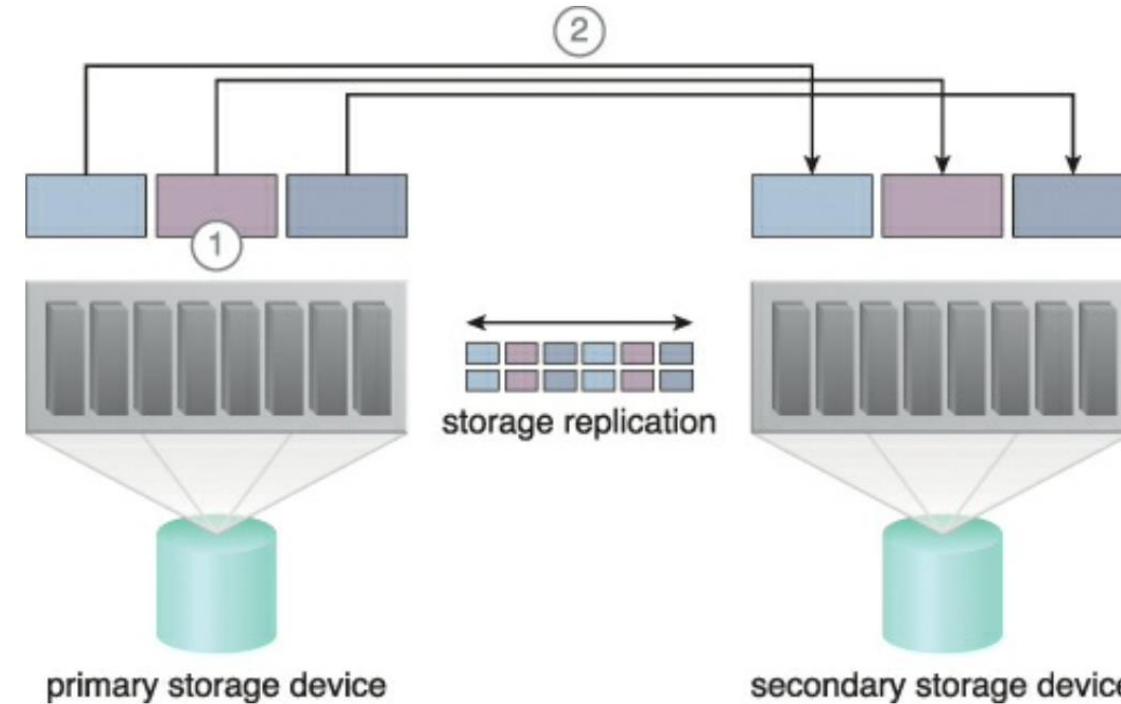
Redundant Storage Architecture



Redundant Storage Architecture



Redundant Storage Architecture



Redundant Storage Architecture

- **Storage Replication**
- Storage replication is a variation of the resource replication mechanisms used to synchronously or asynchronously replicate data from a primary storage device to a secondary storage device.
- It can be used to replicate partial and entire LUNs.
- Cloud providers may locate secondary cloud storage devices in a different geographical region than the primary cloud storage device, usually for economic reasons.
- Legal issues need to be addressed.

End of Module 4

Cloud Computing – CSE4001

Course Instructor: Dr. Arunkumar Gopu

Senior Assistant Professor – Grade I

School of Computer Science and Engineering (SCOPE)

VIT – AP University, Amaravati, Andhra Pradesh.

Email-id: arunkumar.gopu@vitap.ac.in

Module 5

Cloud Delivery Model Considerations

Cloud Delivery Model Considerations: The cloud provider perspective
Building IaaS environments, equipping PaaS environments, optimizing SaaS environments, the cloud consumer perspective, working with IaaS environments, working with PaaS environments, working with SaaS services.

Cloud Delivery Models: The Cloud Provider Perspective

- This module explores the architecture and administration of IaaS, PaaS, and SaaS cloud delivery models from the point of view of the cloud provider.
- The integration and management of these cloud-based environments as part of greater environments and how they can relate to different technologies and cloud mechanism combinations are examined.

Building IaaS Environments

- The virtual server and cloud storage device mechanisms represent the two most fundamental IT resources that are delivered as part of a standard rapid provisioning architecture within IaaS environments.
 - Operating system
 - Primary memory capacity
 - Processing capacity
 - Virtualized storage capacity

Building IaaS Environments

- Memory and virtualized storage capacity is usually allocated with increments of 1 GB to simplify the provisioning of underlying physical IT resources.
- IaaS offerings are pre-emptively assembled by cloud providers via virtual server images that capture the pre-defined configurations.
- Bare-metal provisioning architecture.
- Snapshots can be taken of a virtual server to record its current state, memory, and configuration of a virtualized IaaS environment.

Building IaaS Environments

- The snapshot can be used to duplicate a virtual server.
- Most cloud providers also support importing and exporting options for custom-built virtual server images in both proprietary and standard formats.

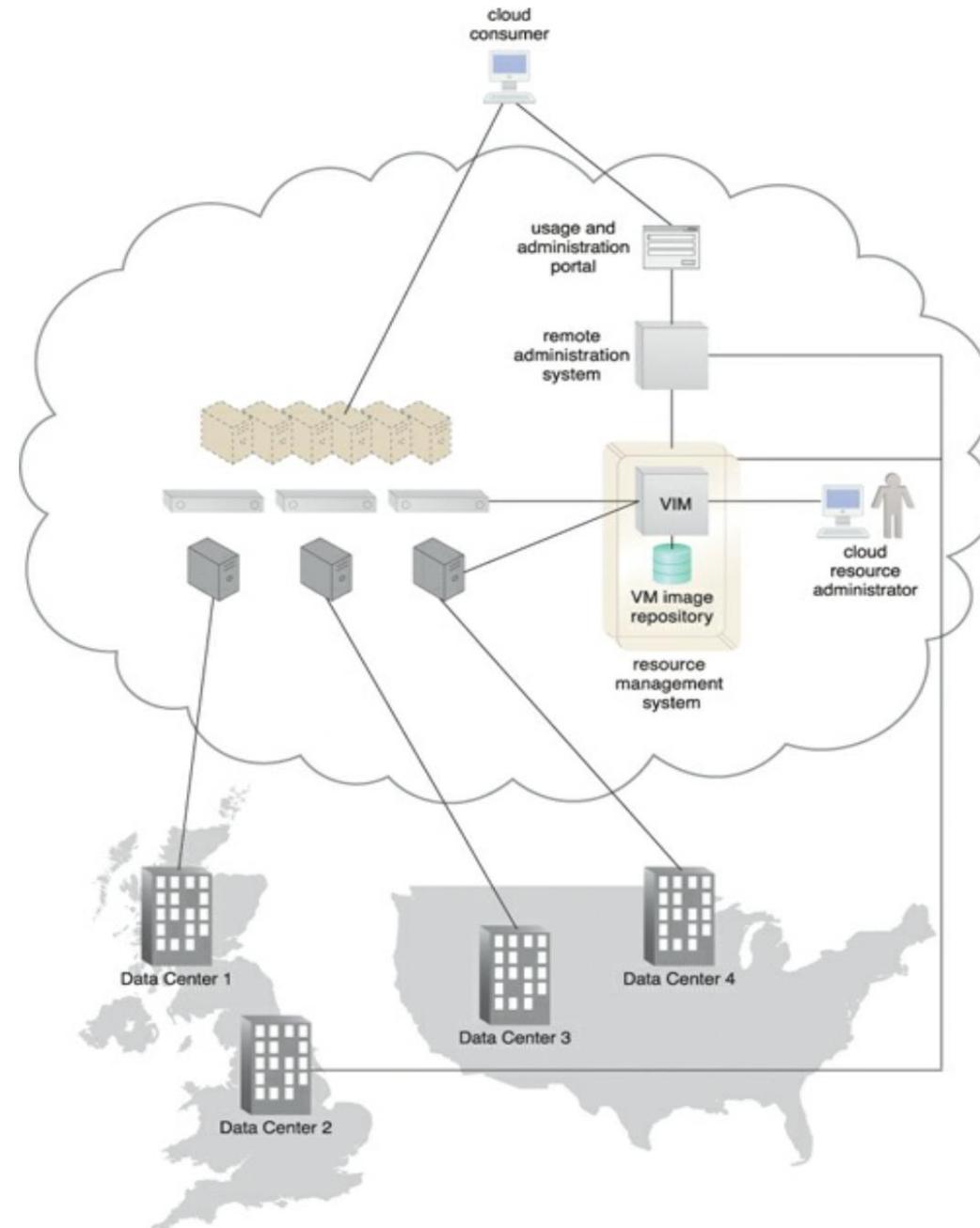
Data Centers

- Cloud providers can offer IaaS-based IT resources from multiple geographically diverse data centers, which provides the following primary benefits:
- Multiple data centers can be linked together for increased resiliency.
- Each data center is placed in a different location to lower the chances of a single failure forcing all of the data centers to go offline simultaneously.
- Connected through high-speed communications networks with low latency.

Data Centers

- Data centers can perform load balancing, IT resource backup and replication, and increase storage capacity, while improving availability and reliability.
- Having multiple data centers spread over a greater area further reduces network latency.
- Data centers that are deployed in different countries make access to IT resources more convenient for cloud consumers that are constricted by legal and regulatory requirements.

Data Centers



Data Centers

- When an IaaS environment is used to provide cloud consumers with virtualized network environments.
- Each cloud consumer is segregated into a tenant environment that isolates IT resources from the rest of the cloud through the Internet.
- VLANs and network access control software collaboratively realize the corresponding logical network perimeters.

Scalability and Reliability

- Within IaaS environments, cloud providers can automatically provision virtual servers via the dynamic vertical scaling type of the dynamic scalability architecture.
- This can be performed through the VIM, if the host physical servers have sufficient capacity.
- The VIM can scale virtual servers out using resource replication as part of a **resource pool architecture**, if a given physical server has insufficient capacity to support **vertical scaling**.

Scalability and Reliability

- The **load balancer mechanism**, as part of a workload distribution architecture, can be used to distribute the workload among IT resources in a pool to complete the **horizontal scaling process**.
- Manual scalability requires the cloud consumer to interact with a usage and administration program to explicitly request IT resource scaling.
- Automatic scalability requires the **automated scaling listener** to monitor the workload and reactively scale the resource capacity.

Scalability and Reliability

- This mechanism typically acts as a monitoring agent that tracks IT resource usage in order to notify the **resource management system** when capacity has been exceeded.
- Replicated IT resources can be arranged in **high-availability configuration** that forms a failover system for implementation via standard VIM features.
- Alternatively, a high- availability/high-performance resource cluster can be created at the physical or virtual server level, or both simultaneously.

Scalability and Reliability

- The **multipath resource access architecture** is commonly employed to enhance reliability via the use of redundant access paths.
- Some cloud providers further offer the provisioning of **dedicated IT resources** via the resource reservation architecture.

Monitoring

- Cloud usage monitors in an IaaS environment can be implemented using the VIM.
- **Virtual Server Lifecycles**—Recording and tracking uptime periods and the allocation of IT resources, for pay-per-use monitors and time-based billing purposes.
- **Data Storage**—Tracking and assigning the allocation of storage capacity to cloud storage devices on virtual servers, for pay-per-use monitors that record storage usage for billing purposes.

Monitoring

- **Network Traffic**—For pay-per-use monitors that measure inbound and outbound network usage and SLA monitors that track QoS metrics, such as response times and network losses.
- **Failure Conditions**—For SLA monitors that track IT resource and QoS metrics to provide warning in times of failure.
- Event Triggers – For audit monitors that appraise and evaluate the **regulatory compliance** of select IT resources.

Security

- Cloud security mechanisms that are relevant for securing IaaS environments include:
- **Encryption, hashing, digital signature, and PKI** mechanisms for overall protection of data transmission.
- **IAM and SSO** mechanisms for accessing services and interfaces in security systems that rely on user identification, authentication, and authorization capabilities.
- Cloud based **security groups** for isolating virtual environments through hypervisors and network segments via network management software.

Security

- Hardened virtual server images for internal and externally available virtual server environments
- Various cloud usage monitors to track provisioned virtual IT resources to detect abnormal usage patterns

Equipping PaaS Environments

- PaaS environments typically is **application development and deployment platforms** in order to accommodate different programming models, languages, and frameworks.
- A **separate ready-made environment** is usually created for each programming stack that contains the necessary software to run applications.
- Each platform is accompanied by a **matching SDK and IDE**, which can be custom-built or enabled by IDE plugins supplied by the cloud provider.
- IDE toolkits can simulate the cloud runtime locally within the PaaS environment and usually include executable application servers.

Equipping PaaS Environments

- The security restrictions that are inherent to the runtime are also simulated in the development environment, including checks for unauthorized attempts to access system IT resources.
- Cloud consumers can create and control customized virtual server images with ready-made environments.
- This mechanism also provides features specific to the PaaS platform, such as managing deployed applications and configuring multitenancy.

Scalability and Reliability

- The scalability requirements of PaaS environments are generally addressed via **dynamic scalability and workload distribution architectures**.
- That rely on the use of native automated scaling listeners and load balancers.
- The resource pooling architecture is further utilized to provision IT resources from resource pools made available to multiple cloud consumers.

Scalability and Reliability

- Cloud providers can evaluate network traffic and server-side connection usage against the instance's workload.
- The reliability of ready-made environments and hosted cloud services and applications can be supported with **standard failover system mechanisms**
- **Non-disruptive service relocation architecture**, to shield cloud consumers from failover conditions.
- The **resource reservation architecture** may also be in place to offer exclusive access to PaaS-based IT resources

Monitoring

- Specialized cloud usage monitors in PaaS environments are used to monitor the following:
- Ready-Made Environment Instances – The applications of these instances are recorded by **pay-per-use monitors** for the calculation of time-based usage fees.
- Data Persistence – This statistic is provided by pay-per-use monitors that record the number of objects, individual occupied storage sizes, and database transactions per billing period.

Monitoring

- Network Usage—**Inbound and outbound network usage** is tracked for pay-per-use monitors and SLA monitors that track network-related QoS metrics.
- Failure Conditions—SLA monitors that track the QoS metrics of IT resources need to capture failure statistics.
- Event Triggers – This metric is primarily used by audit monitors that need to respond to certain types of events.



Security

- The PaaS environment, by default, does not usually introduce the need for new cloud security mechanisms beyond those that are already provisioned for IaaS environments.

Optimizing SaaS Environments

- In SaaS implementations, cloud service architectures are generally based on **multitenant environments** that enable concurrent cloud consumer access.
- SaaS IT **resource segregation** does not typically occur at the infrastructure level.
- SaaS implementations rely heavily on the features provided by the native dynamic scalability and workload distribution architectures.
- Non-disruptive service relocation to ensure that failover conditions do not impact the availability of SaaS-based cloud services.

Optimizing SaaS Environments

- SaaS deployment will bring with it unique architectural, functional, and runtime requirements.
- These requirements are specific to the nature of the business logic the SaaS-based cloud service is programmed with, as well as the distinct usage patterns it is subjected to by its cloud service consumers.
- For example, consider the diversity in functionality and usage of the following recognized online SaaS offerings:
- Collaborative authoring and information-sharing (Wikipedia, Blogger)
- Collaborative management(Zimbra, Google Apps)
- Audio/video communications(Skype, Google Meet)

Optimizing SaaS Environments

- Consider that many of the previously listed cloud services are offered in one or more of the following implementation mediums:
- Mobile Application
- REST service
- Webservice
- Mobile-enabled SaaS implementations are commonly supported by the multi-device broker mechanism.

Optimizing SaaS Environments

- Service Load Balancing—for workload distribution across redundant SaaS-based cloud service implementations
- Dynamic Failure Detection and Recovery—To establish a system that can automatically resolve some failure conditions without disruption in service to the SaaS implementation
- Storage Maintenance Window—to allow for planned maintenance outages that do not impact SaaS implementation availability
- Elastic Resource Capacity/Elastic Network Capacity—to establish inherent elasticity within the SaaS-based cloud service architecture that enables it to automatically accommodate a range of runtime scalability requirements

Optimizing SaaS Environments

- Cloud Balancing—To instill broad resiliency within the SaaS implementation, which can be especially important for cloud services subjected to extreme concurrent usage volumes
- Specialized cloud usage monitors can be used in SaaS environments to track the following types of metrics:
- Tenant Subscription Period—This metric is used by pay-per-use monitors to record and track application usage for time-based billing.
- Application Usage – This metric, based on user or security groups, is used with pay-per-use monitors to record and track application usage for billing purposes.

Optimizing SaaS Environments

- Tenant Application Functional Module – This metric is used by pay-per-use monitors for function-based billing.
- Cloud services can have different functionality tiers according to whether the cloud consumer is free-tier or a paid subscriber.
- SaaS environments are also commonly monitored for data storage, network traffic, failure conditions, and event triggers.



VIT-AP
UNIVERSITY

Security

- SaaS implementations generally rely on a foundation of security controls inherent to their deployment environment.
- Distinct business processing logic will then add layers of additional cloud security mechanisms or specialized security technologies.

The Cloud Consumer Perspective: Working with IaaS

- Virtual servers are accessed at the operating system level using remote terminal applications.
- Remote Desktop Client – for Windows based environments and presents a Windows GUI desktop
- SSH Client – for Mac and other Linux based environments to allow for secure channel connections to text-based shell accounts running on the server OS.

Working with IaaS

- A cloud storage device can be attached directly to the virtual servers and accessed through the virtual servers' functional interface for management by the operating system.
- Alternatively, a cloud storage device can be attached to an IT resource that is being hosted outside of the cloud, such as an on-premise device over a WAN or VPN.
- In these cases, the following formats for the manipulation and transmission of cloud storage data are commonly used:

Working with IaaS

- Networked File System—System based storage access, whose rendering of files is like how folders are organized in operating systems (NFS, CIFS).
- Storage Area Network Devices—Block-based storage access collates and formats geographically diverse data into cohesive files for optimal network transmission (iSCSI, Fibre Channel).
- Web-Based Resources—Object-based storage access by which an interface that is not integrated into the operating system logically represents files, which can be accessed through a Web-based interface (Amazon S3)

Working with PaaS Environments

- A typical PaaS IDE can offer a wide range of tools and programming resources, such as software libraries, class libraries, frameworks, APIs, and various runtime capabilities that emulate the intended cloud-based deployment environment.
- These features allow developers to create, test, and run application code within the cloud or locally (on-premise) while using the IDE to emulate the cloud deployment environment.

Working with PaaS Environments

- Compiled or completed applications are then bundled and uploaded to the cloud and deployed via the ready-made environments. This deployment process can also be controlled through the IDE.
- PaaS also allows for applications to use cloud storage devices as independent data storing systems for holding development specific data (for example in a repository that is available outside of the cloud environment).
- Both SQL and NoSQL database structures are generally supported.

Working with SaaS Environments

- Because SaaS-based cloud services are almost always accompanied by refined and generic APIs, they are usually designed to be incorporated as part of larger distributed solutions.
- A common example of this is Google Maps, which offers a comprehensive API that enables mapping information and images to be incorporated into Web sites and Web-based applications.
- Many SaaS offerings are provided free of charge, although these cloud services often come with data collecting sub-programs that harvest usage data for the benefit of the cloud provider.

Working with SaaS Environments

- When using any SaaS product that is sponsored by third parties, there is a reasonable chance that it is performing a form of background information gathering.
- Reading the cloud provider's agreement will usually help shed light on any secondary activity that the cloud service is designed to perform.
- Cloud consumers using SaaS products supplied by cloud providers are relieved of the responsibilities of implementing and administering their underlying hosting environments.

Working with SaaS Environments

- Customization options are usually available to cloud consumers; however, these options are generally limited to the runtime usage control of the cloud service instances that are generated specifically by and for the cloud consumer.

Cloud Computing – CSE4001

Course Instructor: Dr. Arunkumar Gopu

Senior Assistant Professor – Grade I

School of Computer Science and Engineering (SCOPE)

VIT – AP University, Amaravati, Andhra Pradesh.

Email-id: arunkumar.gopu@vitap.ac.in

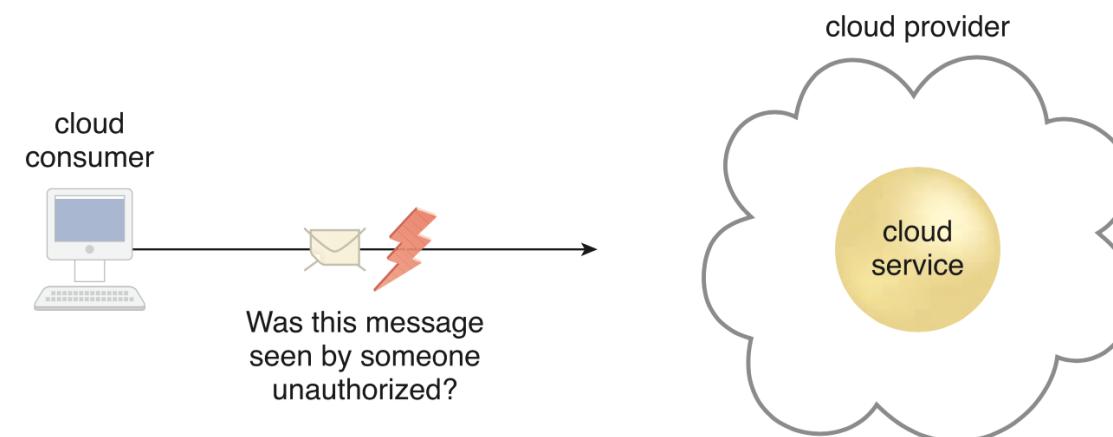
Module 6

Fundamental Cloud Security and Mechanisms

Basic terms and concepts, Threat agents, Cloud security threats, Encryption, Hashing, Digital Signature, Public Key Infrastructure(PKI), Identity and Access Management(IAM), Single Sign-On(SSO), Cloud Based Security Groups, Hardened Virtual Server Images.

Basic Terms and Concepts

- **Confidentiality** is the characteristic of something being made accessible only to authorized parties.
- Within cloud environments, confidentiality primarily pertains to restricting access to data in transit and storage.



Basic Terms and Concepts

- **Integrity** is the characteristic of not having been altered by an unauthorized party.
- An important issue that concerns data integrity in the cloud is whether a cloud consumer can be guaranteed that the data it transmits to a cloud service matches the data received by that cloud service.
- Integrity can extend to how data is stored, processed, and retrieved by cloud services and cloud-based IT resources.

Basic Terms and Concepts

- **Availability** is the characteristic of being accessible and usable during a specified time period.
- In typical cloud environments, the availability of cloud services can be a responsibility that is shared by the cloud provider and the cloud carrier.
- The availability of a cloud-based solution that extends to cloud service consumers is further shared by the cloud consumer.

Basic Terms and Concepts

- A **threat** is a potential security violation that can challenge defences in an attempt to breach privacy and/or cause harm.
- Both manually and automatically instigated threats are designed to exploit known weaknesses, also referred to as vulnerabilities.
- A threat that is carried out results in an attack.

Basic Terms and Concepts

- A **vulnerability** is a weakness that can be exploited either because it is protected by insufficient security controls, or because existing security controls are overcome by an attack.
- IT resource vulnerabilities can have a range of causes, including configuration deficiencies, security policy weaknesses, user errors, hardware or firmware flaws, software bugs, and poor security architecture.



Basic Terms and Concepts

- **Risk** is the possibility of loss or harm arising from performing an activity.
- Risk is typically measured according to its threat level and the number of possible or known vulnerabilities.
- Two metrics that can be used to determine risk for an IT resource are:
 1. The **probability** of a threat occurring to exploit vulnerabilities in the IT resource.
 2. **The expectation of loss** upon the IT resource being compromised.

Basic Terms and Concepts

- **Security controls** are countermeasures used to prevent or respond to security threats and to reduce or avoid risk.
- Details on how to use security countermeasures are typically outlined in the security policy, to enable maximum protection of sensitive and critical IT resources.

Basic Terms and Concepts

- **Security Mechanisms** are countermeasures comprising a defensive framework that protects IT resources, information, and services.
- **Security Policies** establishes a set of security rules and regulations.
- Often, security policies will further define how these rules and regulations are implemented and enforced.
- For example, the positioning and usage of security controls and mechanisms can be determined by security policies.

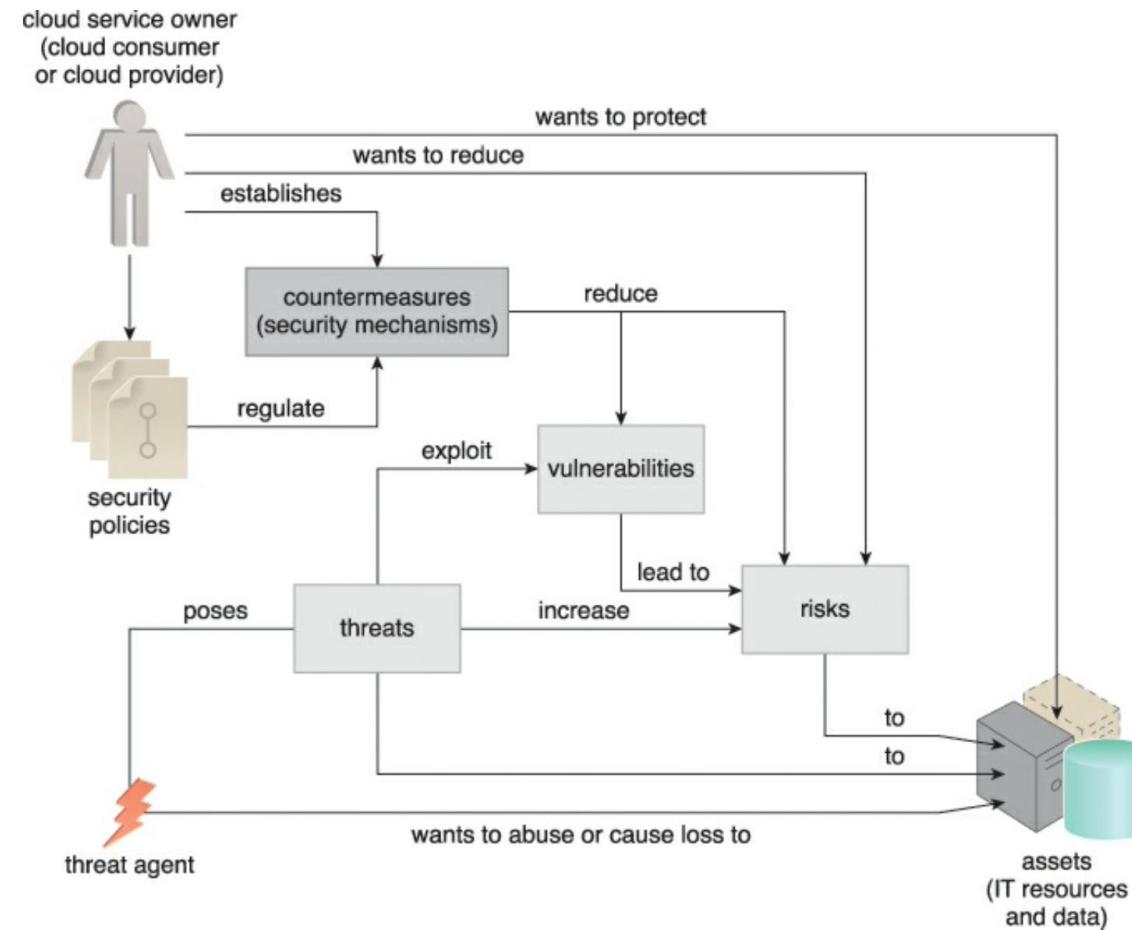
Basic Terms and Concepts

- Confidentiality, integrity, authenticity, and availability are characteristics that can be associated with measuring security.
- Threats, vulnerabilities, and risks are associated with measuring and assessing insecurity, or the lack of security.
- Security controls, mechanisms, and policies are associated with establishing countermeasures and safeguards in support of improving security.

Threat Agents

- A **threat agent** is an entity that poses a threat because it is capable of carrying out an attack.
- Cloud security threats can originate either internally or externally, from humans or software programs.
- Figure illustrates the role a threat agent assumes in relation to vulnerabilities, threats, and risks, and the safeguards established by security policies and security mechanisms.

Threat Agents



Threat Agents

- An **anonymous attacker** is a non-trusted cloud service consumer without permissions in the cloud.
- It typically exists as an external software program that launches network-level attacks through public networks.
- When anonymous attackers have limited information on security policies and defences, it can inhibit their ability to formulate effective attacks.
- Therefore, anonymous attackers often resort to committing acts like bypassing user accounts or stealing user credentials.



Threat Agents

- A **malicious service agent** can intercept and forward the network traffic that flows within a cloud.
- It typically exists as a service agent (or a program pretending to be a service agent) with compromised or malicious logic.
- It may also exist as an external program able to remotely intercept and potentially corrupt message contents.



Threat Agents

- A **trusted attacker** shares IT resources in the same cloud environment as the cloud consumer and attempts to exploit legitimate credentials to target cloud providers and the cloud tenants with whom they share IT resources.
- Unlike anonymous attackers (which are non-trusted), trusted attackers usually **launch their attacks from within a cloud's trust boundaries** by abusing legitimate credentials or via the appropriation of sensitive and confidential information.

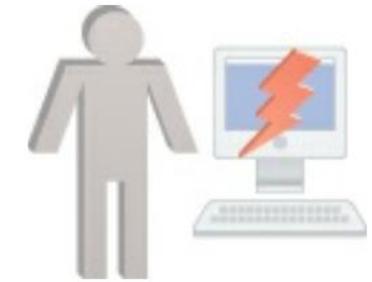


Threat Agents

- **Trusted attackers** (also known as *malicious tenants*) can use cloud-based IT resources for a wide range of exploitations, including the hacking of weak authentication processes, the breaking of encryption, the spamming of e-mail accounts, or to launch common attacks, such as denial of service campaigns.

Threat Agents

- **Malicious Insider** are human threat agents acting on behalf of or in relation to the cloud provider.
- They are typically **current or former employees** or third parties with access to the cloud provider's premises.
- This type of threat agent carries tremendous damage potential, as the malicious insider may have administrative privileges for accessing cloud consumer IT resources.



Threat Agents

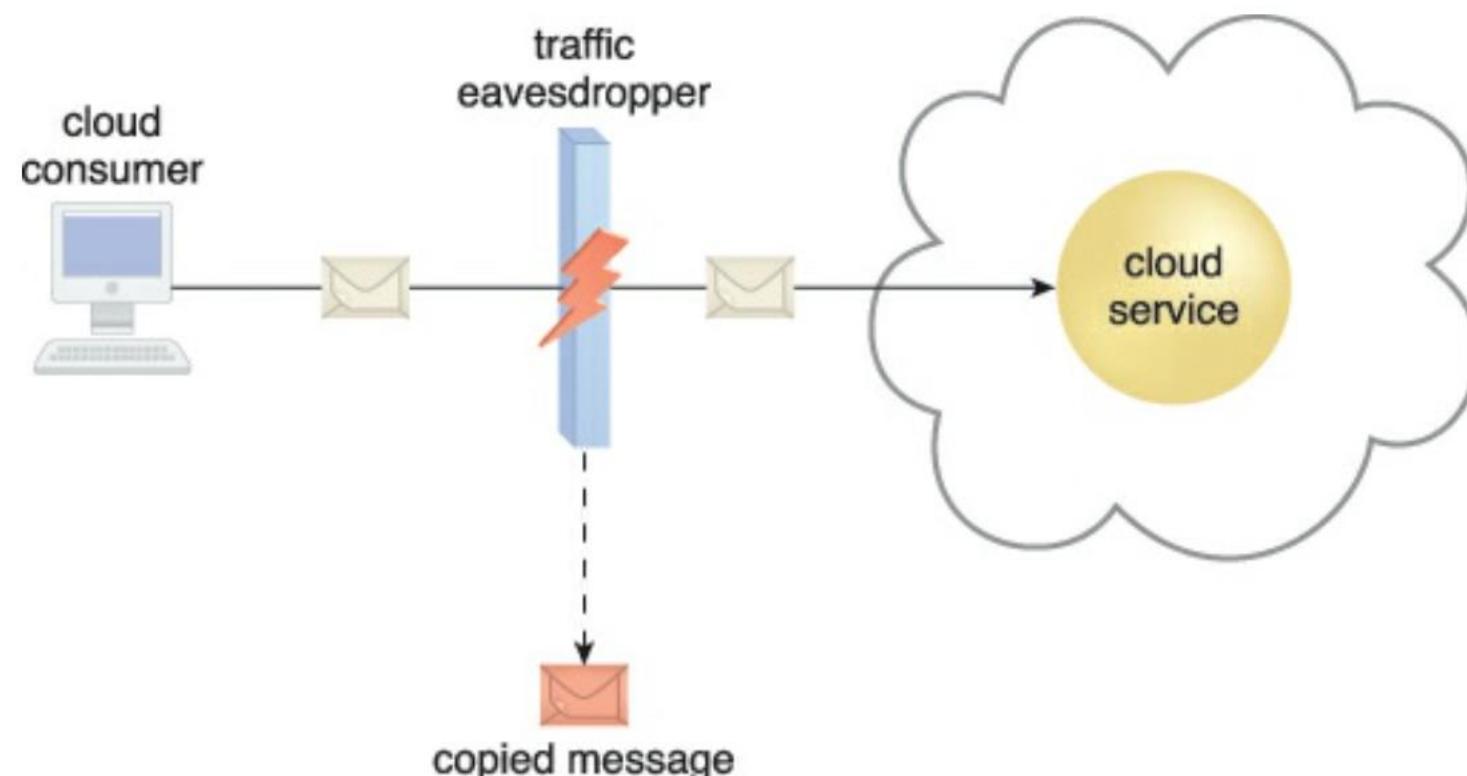
- An anonymous attacker is a non-trusted threat agent that usually attempts attacks from outside of a cloud's boundary.
- A malicious service agent intercepts network communication to maliciously use or augment the data.
- A trusted attacker exists as an authorized cloud service consumer with legitimate credentials that it uses to exploit access to cloud-based IT resources.
- A malicious insider is a human that attempts to abuse access privileges to cloud premises.

Cloud Security Threats

Traffic Eavesdropping

- Traffic eavesdropping occurs when data being transferred to or within a cloud (usually from the cloud consumer to the cloud provider) is passively intercepted by a malicious service agent for illegitimate information gathering purposes.
- The aim of this attack is to directly compromise the confidentiality of the data and, possibly, the confidentiality of the relationship between the cloud consumer and cloud provider.
- Because of the passive nature of the attack, it can more easily go undetected for extended periods of time.

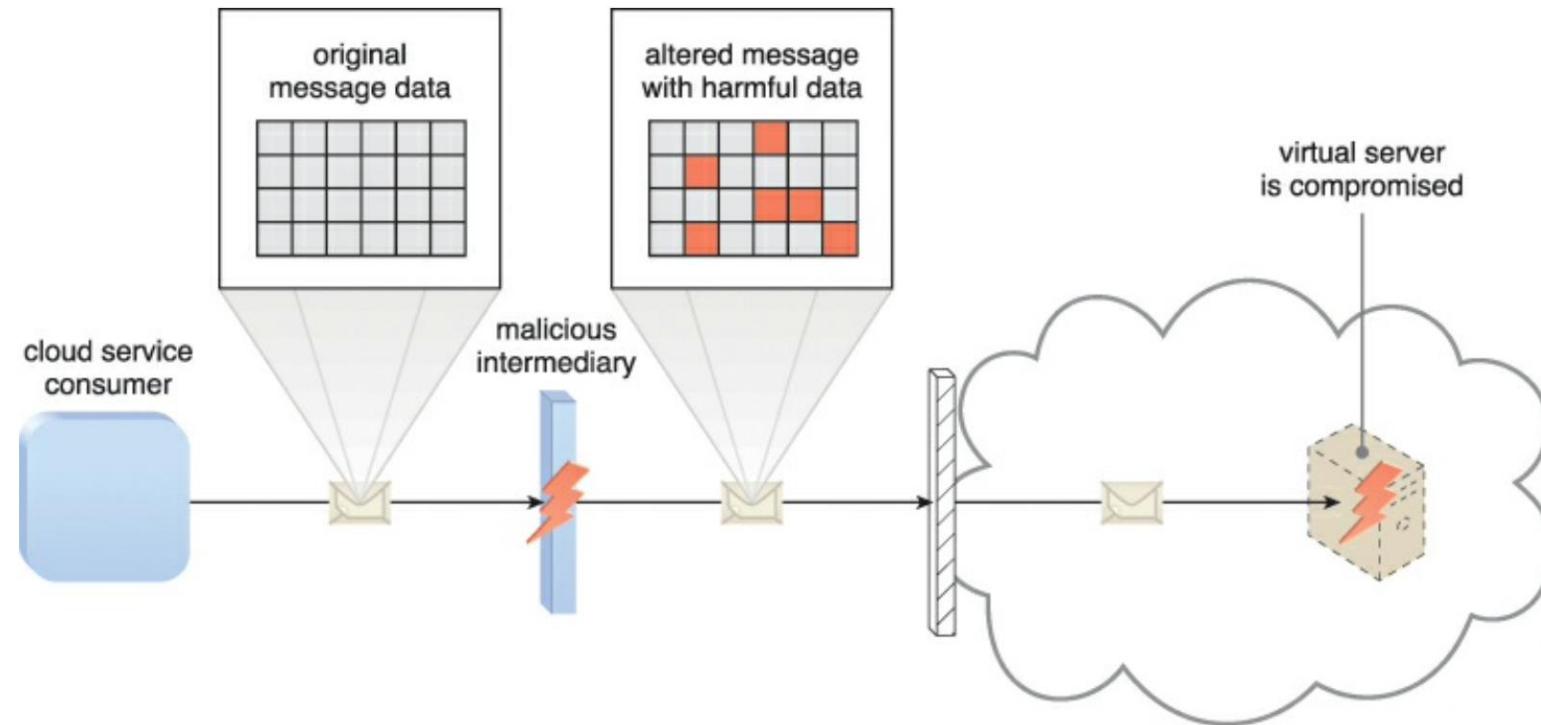
Traffic Eavesdropping



Malicious Intermediary

- The malicious intermediary threat arises when messages are intercepted and altered by a malicious service agent, thereby potentially compromising the message's confidentiality and/or integrity.
- It may also insert harmful data into the message before forwarding it to its destination.

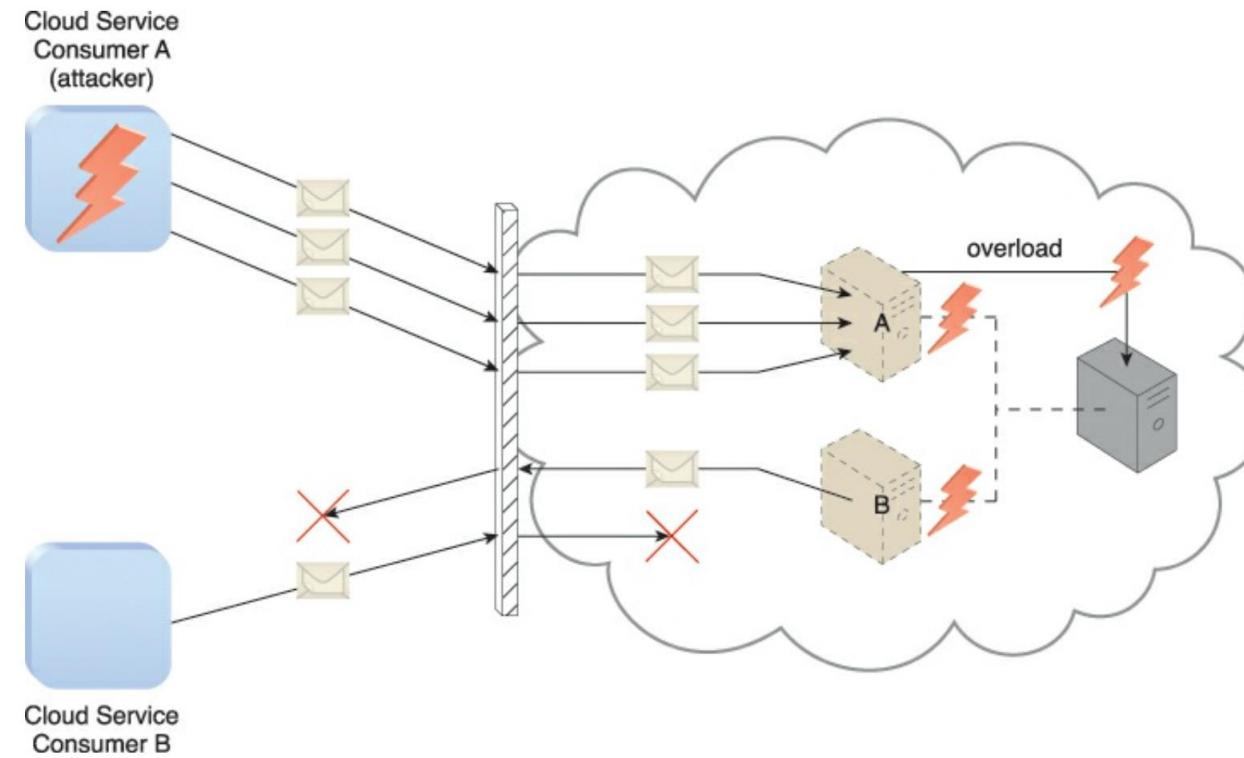
Malicious Intermediary



Denial of Service

- The objective of the denial of service (DoS) attack is to overload IT resources to the point where they cannot function properly.
- This form of attack is commonly launched in one of the following ways:
 1. The workload on cloud services is artificially increased with imitation messages or repeated communication requests.
 2. The network is overloaded with traffic to reduce its responsiveness and cripple its performance.
 3. Multiple cloud service requests are sent, each of which is designed to consume excessive memory and processing resources.

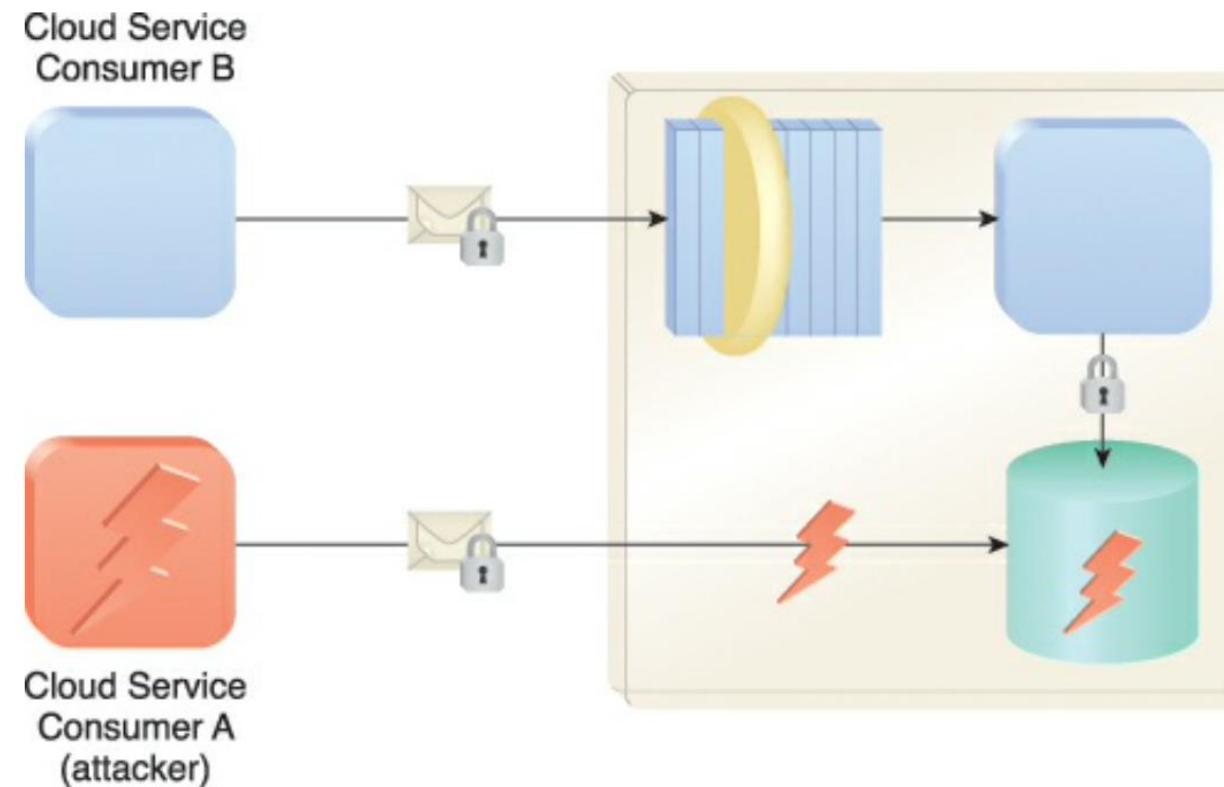
Denial of Service



Insufficient Authorization

- The insufficient authorization attack occurs when access is granted to an attacker erroneously or too broadly, resulting in the attacker getting access to IT resources that are normally protected.
- This is often a result of the attacker gaining direct access to IT resources that were implemented under the assumption that they would only be accessed by trusted consumer programs.

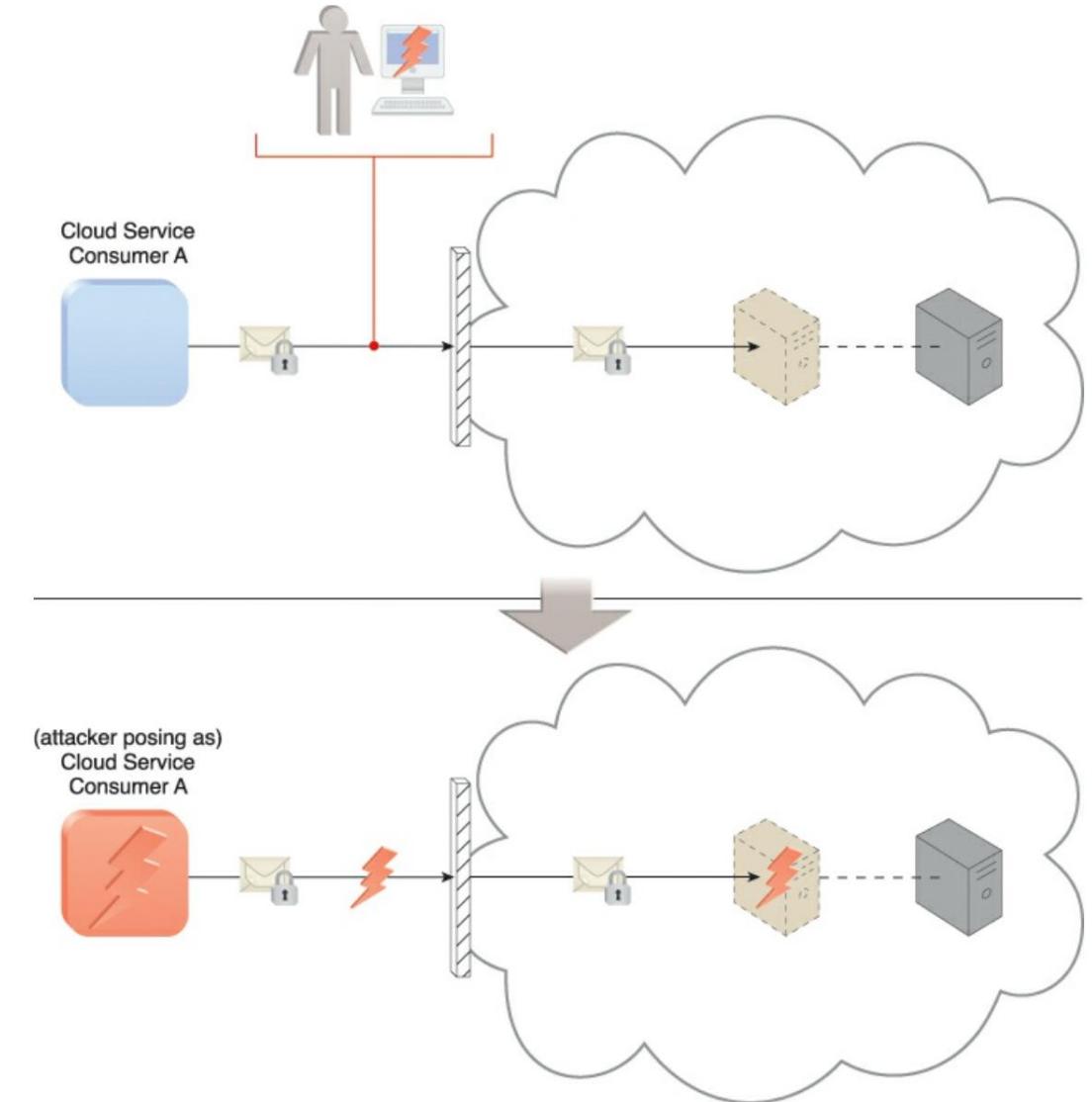
Insufficient Authorization



Insufficient Authorization

- A variation of this attack, known as **weak authentication**, can result when weak passwords or shared accounts are used to protect IT resources.
- Within cloud environments, these types of attacks can lead to significant impacts depending on the range of IT resources.

Insufficient Authorization



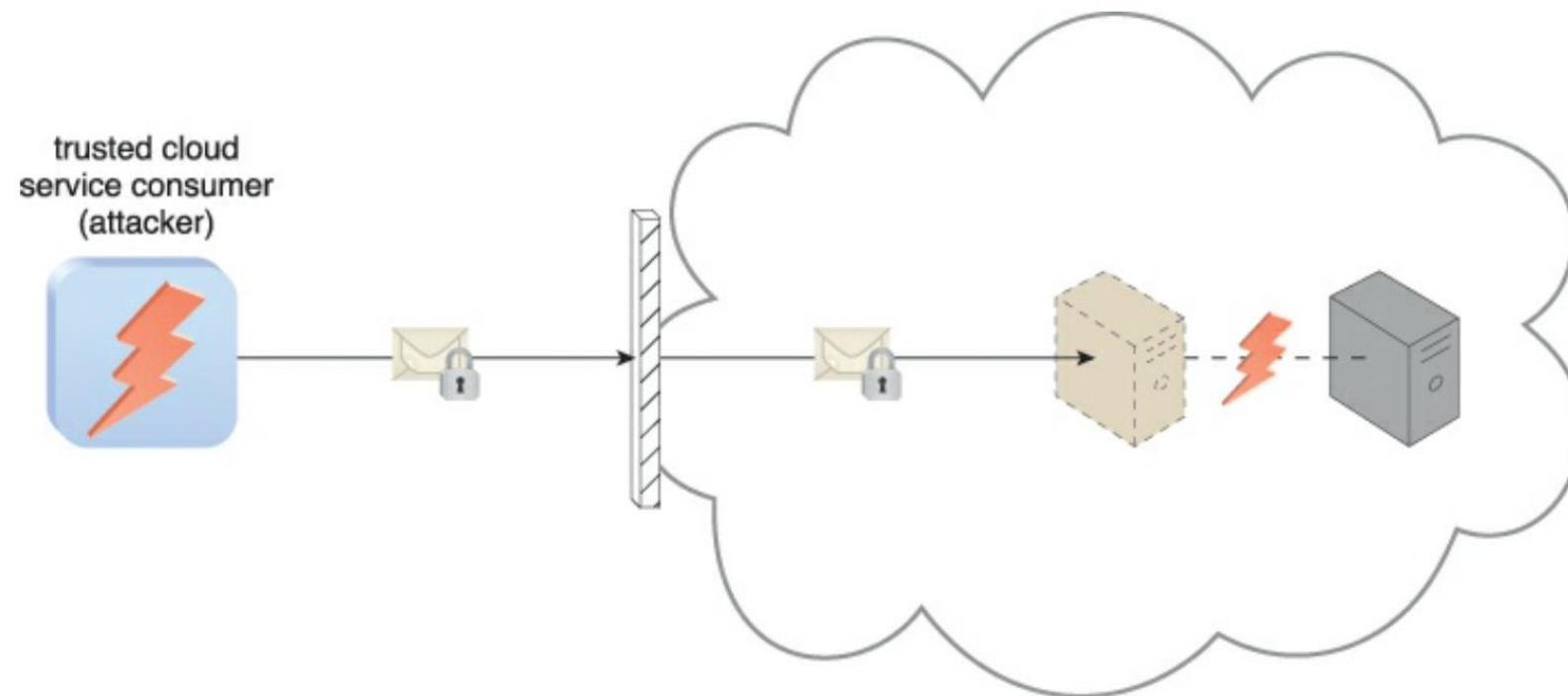
Virtualization Attack

- Virtualization provides multiple cloud consumers with access to IT resources that share underlying hardware but are logically isolated from each other.
- Because cloud providers grant cloud consumers administrative access to virtualized IT resources (such as virtual servers), there is an inherent risk that cloud consumers could abuse this access to attack the underlying physical IT resources.

Virtualization Attack

- A virtualization attack exploits vulnerabilities in the virtualization platform to jeopardize its confidentiality, integrity, and/or availability.
- Where a trusted attacker successfully accesses a virtual server to compromise its underlying physical server.
- With public clouds, where a single physical IT resource may be providing virtualized IT resources to multiple cloud consumers, such an attack can have significant repercussions.

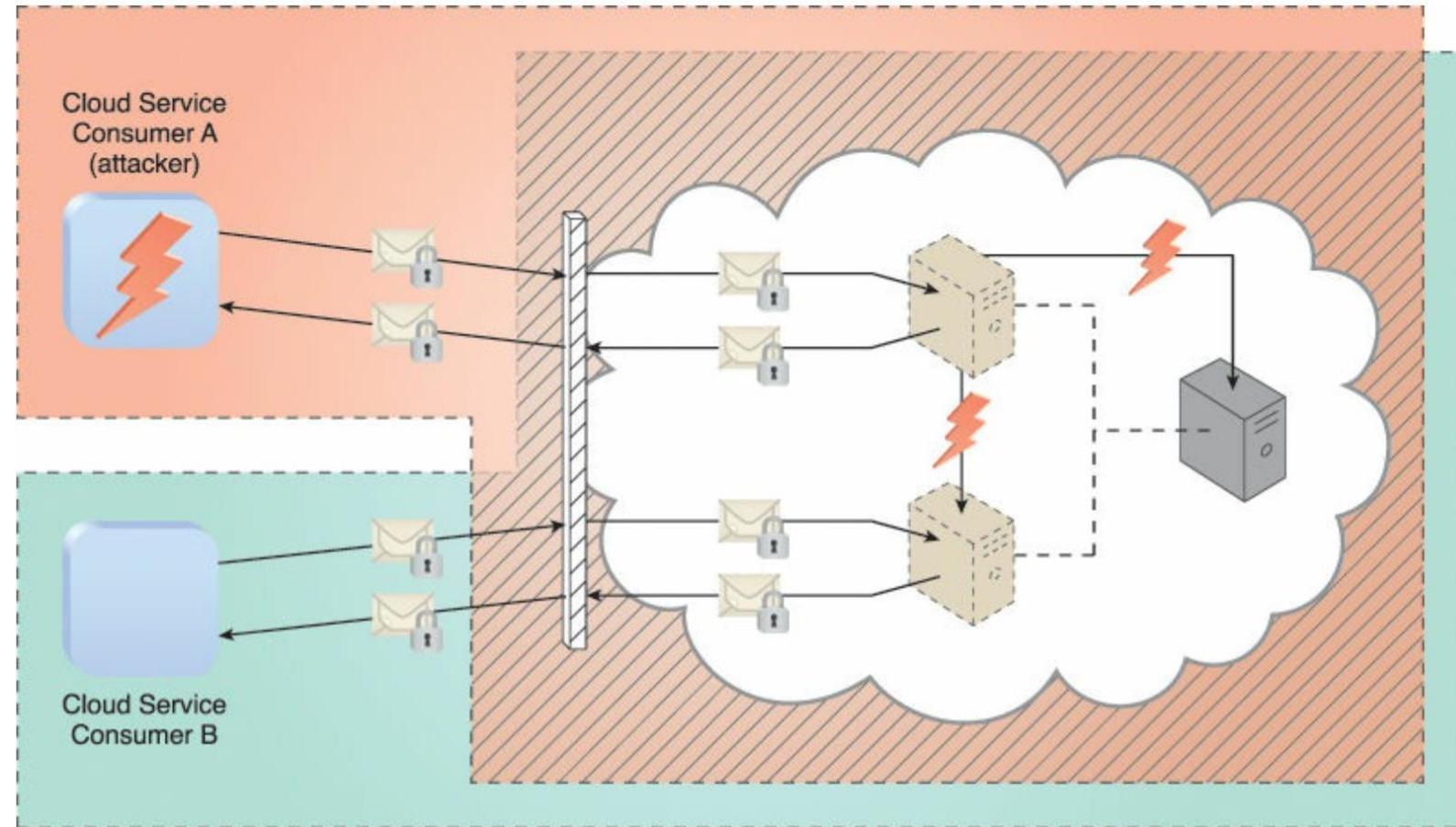
Virtualization Attack



Overlapping Trust Boundaries

- If physical IT resources within a cloud are shared by different cloud service consumers, these cloud service consumers have overlapping trust boundaries.
- Malicious cloud service consumers can target shared IT resources with the intention of compromising cloud consumers or other IT resources that share the same trust boundary.
- The consequence is that some or all of the other cloud service consumers could be impacted by the attack and/or the attacker could use virtual IT resources against others that happen to also share the same trust boundary.

Overlapping Trust Boundaries



Container Attack

- The use of containerization introduces a lack of isolation from the host operating system level.
- Since containers deployed on the same machine share the same host operating system, security threats can increase because access to the entire system can be gained.
- If the underlying host is compromised, all containers running on the host may be impacted.
- Containers can be created from within an operating system running on a virtual server, while other virtual servers (or physical servers) remain intact.

Container Attack

- Another option is a one-service per physical server deployment model where all container images deployed on the same host are the same.
- This can reduce risk without the need to virtualize the IT resources.
- In this case, a security breach to one cloud service instance would only allow access to other instances, and the residual risk could be considered as acceptable.

Encryption

- Data, by default, is coded in a readable format known as plaintext.
- When transmitted over a network, plaintext is vulnerable to unauthorized and potentially malicious access.
- The encryption mechanism is a digital coding system dedicated to preserving the confidentiality and integrity of data.
- It is used for encoding plaintext data into a protected and unreadable format.
- Encryption technology commonly relies on a standardized algorithm called a cipher to transform original plaintext data into encrypted data, referred to as ciphertext.

Encryption

- Access to ciphertext does not divulge the original plaintext data, apart from some forms of metadata, such as message length and creation date.
- When encryption is applied to plaintext data, the data is paired with a string of characters called an encryption key, a secret message that is established by and shared among authorized parties.
- The encryption key is used to decrypt the ciphertext back into its original plaintext format.

Encryption

- The encryption mechanism can help counter the traffic eavesdropping, malicious intermediary, insufficient authorization, and overlapping trust boundaries security threats.
- For example, malicious service agents that attempt traffic eavesdropping are unable to decrypt messages in transit if they do not have the encryption key.
- Symmetric encryption uses the same key for both encryption and decryption, both of which are performed by authorized parties that use the one shared key.

Encryption

- Parties that rightfully decrypt the data are provided with evidence that the original encryption was performed by parties that rightfully possess the key.
- A basic authentication check is always performed, because only authorized parties that own the key can create messages.
- This maintains and verifies data confidentiality.
- Note that symmetrical encryption does not have the characteristic of non-repudiation

Encryption

- Asymmetric encryption relies on the use of two different keys, namely a private key and a public key.
- With asymmetric encryption (which is also referred to as public key cryptography), the private key is known only to its owner while the public key is commonly available.
- A document that was encrypted with a private key can only be correctly decrypted with the corresponding public key.
- Conversely, a document that was encrypted with a public key can be decrypted only using its private key counterpart.

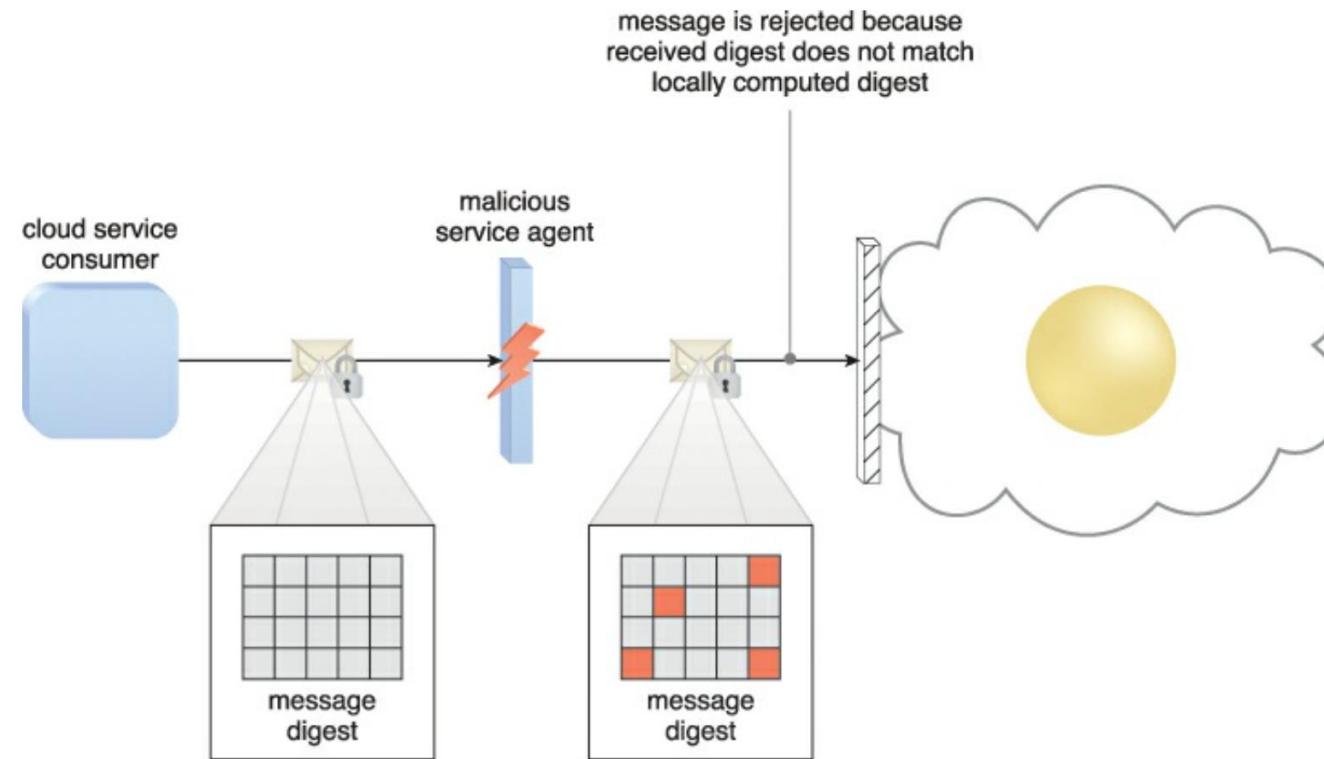
Hashing

- The hashing mechanism is used when a one-way, **non-reversible** form of data protection is required.
- Once hashing has been applied to a message, it is locked and no key is provided for the message to be unlocked.
- A common application of this mechanism is the storage of passwords.
- Hashing technology can be used to derive a hashing code or message digest from a message, which is often of a fixed length and smaller than the original message.

Hashing

- The message sender can then utilize the hashing mechanism to attach the message digest to the message.
- The recipient applies the same hash function to the message to verify that the produced message digest is identical to the one that accompanied the message.
- Any alteration to the original data results in an entirely different message digest and clearly indicates that tampering has occurred.

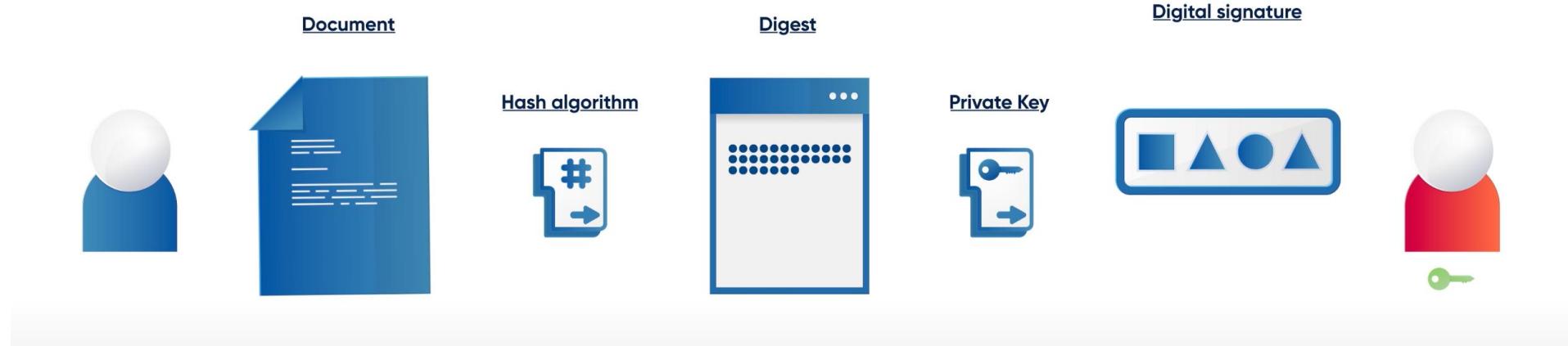
Hashing



Digital Signature

- The digital signature mechanism is a means of providing data authenticity and integrity through authentication and non-repudiation.
- A message is assigned a digital signature prior to transmission, which is then rendered invalid if the message experiences any subsequent, unauthorized modifications.
- A digital signature provides evidence that the message received is the same as the one created by its rightful sender.

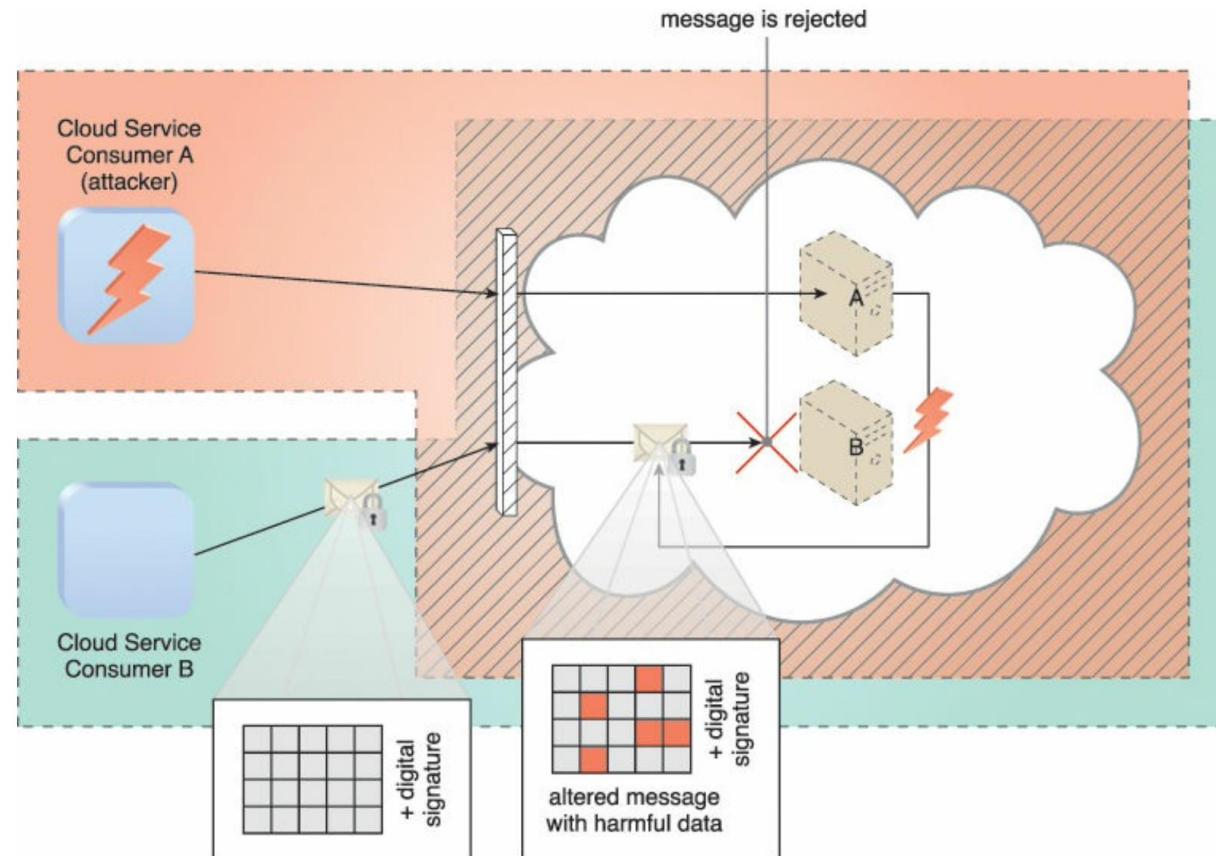
Digital Signature



Digital Signature

- Both hashing and asymmetrical encryption are involved in the creation of a digital signature, which essentially exists as a **message digest that was encrypted by a private key** and appended to the original message.
- The recipient verifies the signature validity and uses the corresponding public key to decrypt the digital signature, which produces the message digest.
- The hashing mechanism can also be applied to the original message to produce this message digest. Identical results from the two different processes indicate that the message maintained its integrity.

Digital Signature



Public Key Infrastructure (PKI)

- A common approach for managing the issuance of asymmetric keys is based on the public key infrastructure (PKI) mechanism, which exists as a system of protocols, data formats, rules, and practices that enable large- scale systems to securely use public key cryptography.
- This system is used to associate public keys with their corresponding key owners (known as public key identification) while enabling the verification of key validity.

Public Key Infrastructure (PKI)

- PKIs rely on the use of digital certificates, which are digitally signed data structures that bind public keys to certificate owner identities, as well as to related information, such as validity periods.

Identity and Access Management (IAM)

- The *identity and access management (IAM)* mechanism encompasses the components and policies necessary to control and track user identities and access privileges for IT resources, environments, and systems.
- Specifically, IAM mechanisms exist as systems comprised of four main components:
- *Authentication* – Username and password combinations remain the most common forms of user authentication credentials managed by the IAM system.

Identity and Access Management (IAM)

- Authentication system also can support digital signatures, digital certificates, biometric hardware (fingerprint readers), specialized software (such as voice analysis programs), and locking user accounts to registered IP or MAC addresses.
- *Authorization* – The authorization component defines the correct granularity for access controls and oversees the relationships between identities, access control rights, and IT resource availability.

Identity and Access Management (IAM)

- *User Management* – Related to the administrative capabilities of the system, the user management program is responsible for creating new user identities and access groups, resetting passwords, defining password policies, and managing privileges.
- *Credential Management* – The credential management system establishes identities and access control rules for defined user accounts, which mitigates the threat of insufficient authorization.

Identity and Access Management (IAM)

- The IAM mechanism is primarily used to counter the insufficient authorization, denial of service, overlapping trust boundaries threats, virtualization attack and containerization attack threats.

Single Sign On

- Propagating the authentication and authorization information for a cloud service consumer across multiple cloud services can be a challenge.
- Especially if numerous cloud services or cloud-based IT resources need to be invoked as part of the same overall runtime activity.
- The *single sign-on (SSO)* mechanism enables one cloud service consumer to be authenticated by a security broker, which establishes a security context that is persisted while the cloud service consumer accesses other cloud services or cloud-based IT resources.

Single Sign On

- Otherwise, the cloud service consumer would need to re-authenticate itself with every subsequent request.
- The SSO mechanism essentially enables mutually independent cloud services and IT resources to generate and circulate runtime authentication and authorization credentials.
- The credentials initially provided by the cloud service consumer remain valid for the duration of a session, while its security context information is shared.
- The SSO mechanism's security broker is especially useful when a cloud service consumer needs to access cloud services residing on different clouds

Cloud based Security Groups

- Cloud resource segmentation is a process by which separate physical and virtual IT environments are created for different users and groups.
- For example, an organization's WAN can be partitioned according to individual network security requirements.
- One network can be established with a resilient firewall for external Internet access, while a second is deployed without a firewall because its users are internal and unable to access the Internet.

Cloud based Security Groups

- Resource segmentation is used to enable virtualization by allocating a variety of physical IT resources to virtual machines.
- It needs to be optimized for public cloud environments, since organizational trust boundaries from different cloud consumers overlap when sharing the same underlying physical IT resources.
- The cloud-based resource segmentation process creates *cloud-based security group* mechanisms that are determined through security policies. Networks are segmented into logical cloud-based security groups that form logical network perimeters.

Cloud based Security Groups

- Each cloud-based IT resource is assigned to at least one logical cloud-based security group. Each logical cloud-based security group is assigned specific rules that govern the communication between the security groups.

Hardened Virtual Server Images

- A virtual server is created from a template configuration called a virtual server image (or virtual machine image).
- Hardening is the process of stripping unnecessary software from a system to limit potential vulnerabilities that can be exploited by attackers.
- Removing redundant programs, closing unnecessary server ports, and disabling unused services, internal root accounts, and guest access are all examples of hardening.

Hardened Virtual Server Images

- A *hardened virtual server image* is a template for virtual service instance creation that has been subjected to a hardening process.
- This generally results in a virtual server template that is significantly more secure than the original standard image.
- Hardened virtual server images help counter the denial of service, insufficient authorization, and overlapping trust boundaries threats.

Hardened Virtual Server Images

