CS 4037
Introduction to Cloud Computing
Lecture 3

Danyal Farhat

FAST School of Computing

NUCES Lahore

Understanding Cloud Computing – Part 1

Lecture’s Agenda

- **Origins and Influences**
- Basic Concepts and Terminologies
- Goals and Benefits
- Risks and Challenges

A Brief History

Computer Scientist John McCarthy 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.”

A Brief History (Cont.)

In 1969, ARPANET's chief scientist Leonard Kleinrock 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."

ARPANET project seeded the Internet.

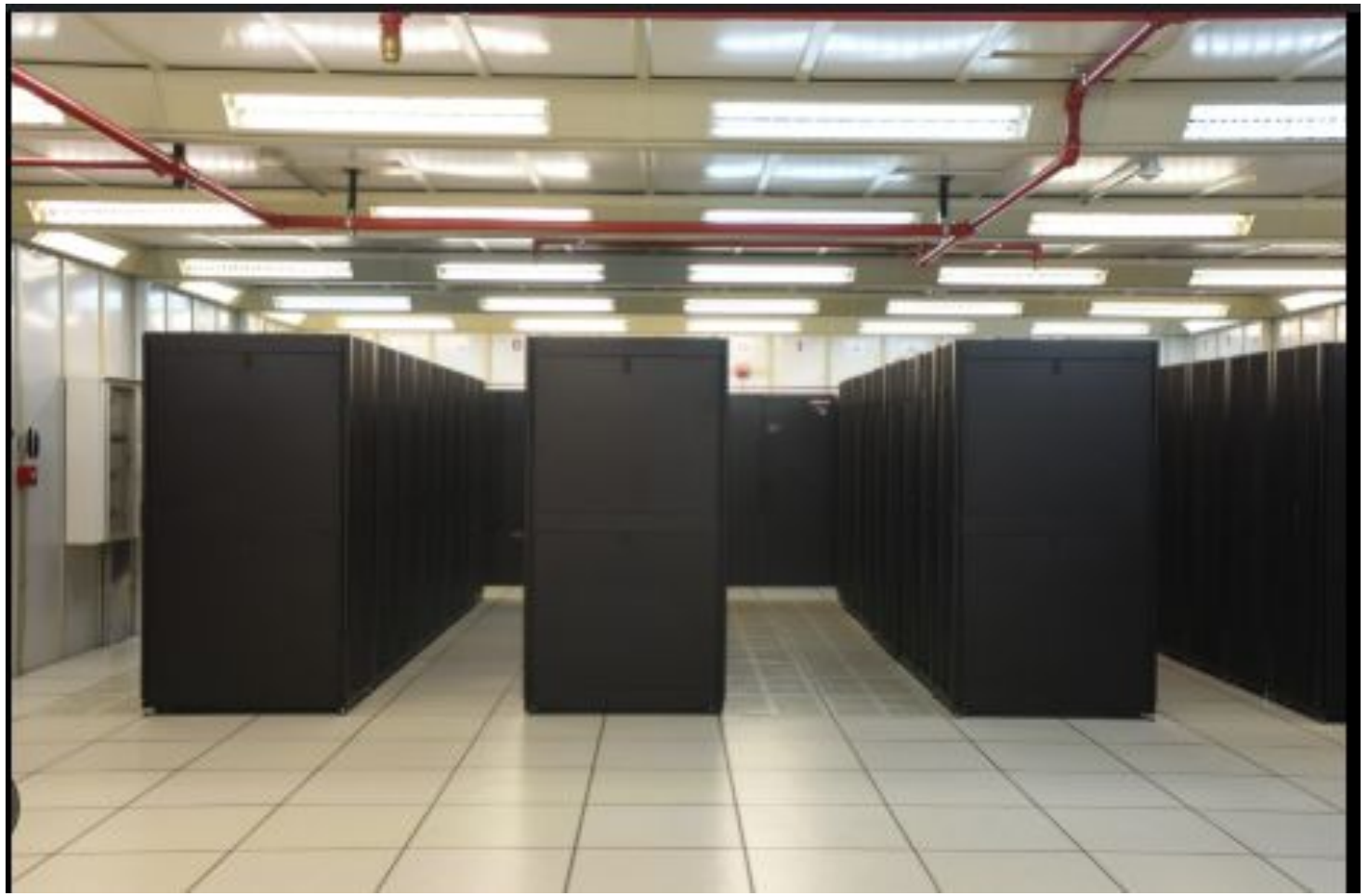
A Brief History (Cont.)

- 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.
- In 2006, the term "cloud computing" emerged in the commercial arena when Amazon launched its Elastic Compute Cloud (EC2) services that enabled organizations to "lease" computing capacity and processing power to run their enterprise applications.

A Brief History (Cont.)

- Google Apps also began providing browser-based enterprise applications in 2006.
- In 2009, the Google App Engine became another historic milestone in cloud computing.
- In 2008, MS Azure was introduced but officially launched in the year 2010.

Data Center



Data Center Components

Rack

EX 4200 Switch

MX 960 Router

Data Center Components

HP Proliant Server

Dell EMC Server

Definitions

Gartner, Inc. is an American technological research and consulting firm. As per a Gartner report:

“Cloud computing is a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.”

Definitions (Cont.)

Forrester is another American research and consulting firm. Forrester Research provided its own definition of cloud computing as:

"A standardized IT capability (services, software, or infrastructure) delivered via Internet technologies in a pay-per-use, self-service way."

Definitions (Cont.)

The definition that received industry-wide acceptance was composed by the US 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."

Definitions (Cont.)

The book provides a more concise definition:

"Cloud computing is a specialized form of distributed computing that introduces utilization models for remotely provisioning scalable and measured resources."

Activity 1

- What is the processing capacity (in GHz) of an HP Gen 10 server having 48 CPUs and 6 cores per CPU?

Business Drivers – Capacity Planning

- "Capacity planning is the process of determining and fulfilling future demands of an organization's IT resources, products, and services."
- Represents the maximum amount of work that an IT resource is capable of delivering in a given period of time.
- Capacity Planning Strategies:
  - 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
    - ✔ For example, add a new node when workload reaches 60% of the total capacity

Business Drivers – Cost Reduction

- Two costs need to be accounted for:
  - The cost of acquiring new infrastructure – Capital Expenditure (CAPEX)
  - The cost of infra's ongoing ownership – Operational Expenditure (OPEX)
- Common forms of infrastructure-related operating overhead:
  - 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
- Migration to cloud usually minimize all the operating overheads

Business Drivers – Organization Agility

- "Organizational agility is the measure of an organization's responsiveness to change."
- Businesses need the ability to adapt and evolve to successfully face change caused by both internal and external factors.
- 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.

Technology Innovations

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.
- Component devices that form a cluster are kept in synchronization through dedicated, high-speed communication links.
- The basic concept of built-in redundancy and failover is core to cloud platforms.

Technology Innovations (Cont.)

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.
- Grid computing systems can involve computing resources that are heterogeneous and geographically dispersed, which is generally not possible with cluster computing-based systems.

Technology Innovations (Cont.)

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.
- As cloud computing evolved, a generation of modern virtualization technologies emerged to overcome the performance, reliability, and scalability limitations of traditional virtualization platforms.
- As a foundation of contemporary cloud technology, modern virtualization provides a variety of virtualization types and technology layers.

Origins and Influences – Key Points

- The primary business drivers that exposed the need for cloud computing and led to its formation include capacity planning, cost reduction, and organizational agility.

- The primary technology innovations that influenced and inspired key distinguishing features and aspects of cloud computing include clustering, grid computing, and traditional forms of virtualization.

Activity 2

- **Draw a simple architecture showing both cluster and grid computing is being used in a single organization having two sites in Lahore, three sites in Karachi and a head office in Islamabad.**

Lecture’s Agenda

- Origins and Influences
- **Basic Concepts and Terminologies**
- Goals and Benefits
- Risks and Challenges

Cloud

- **"A cloud refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources."**

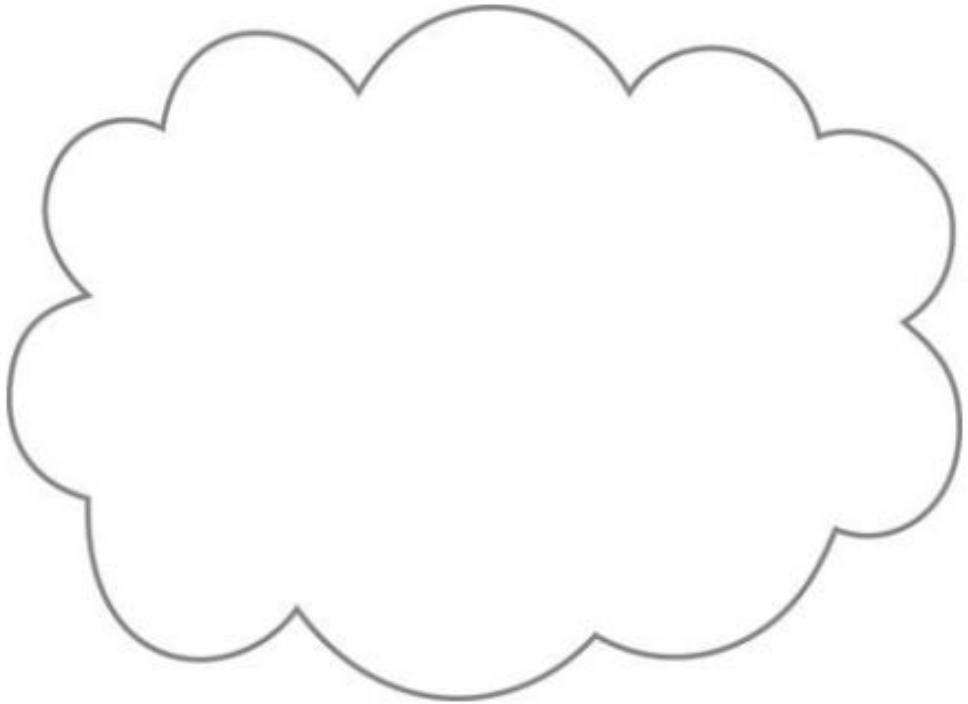


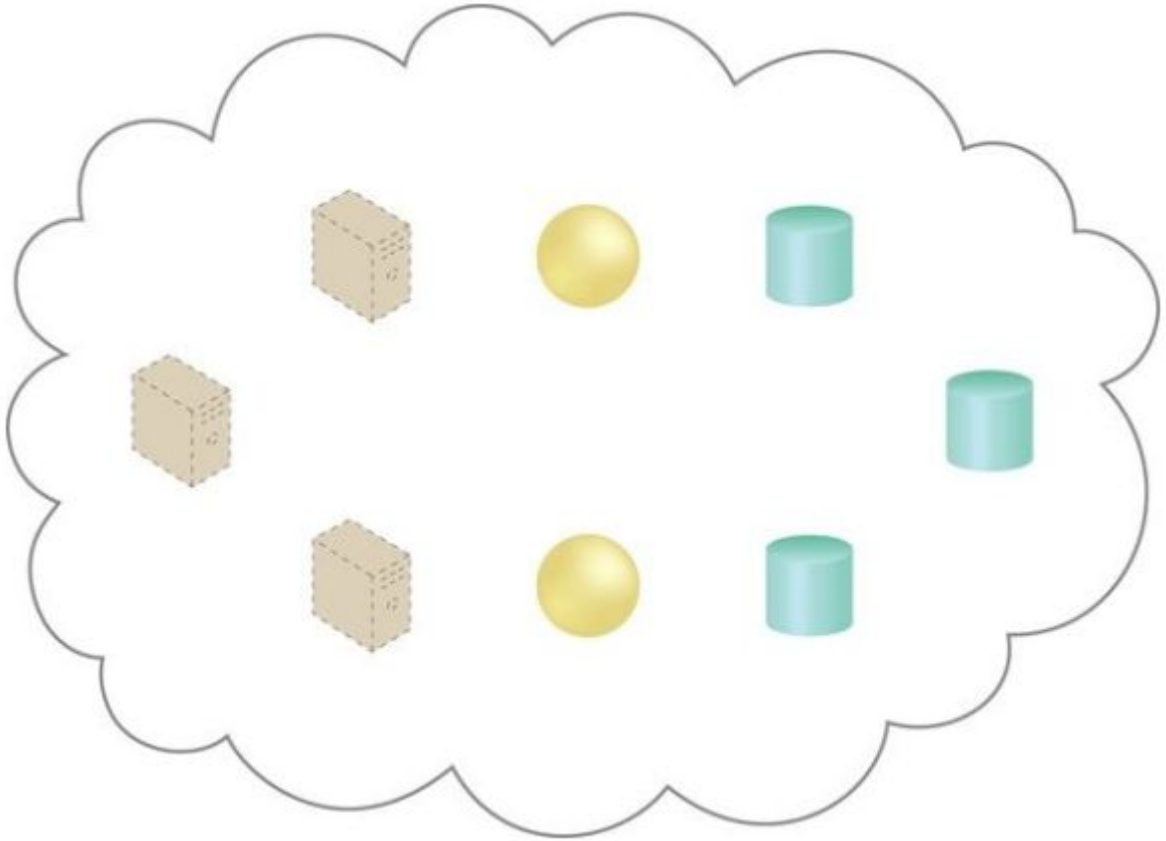
Figure 3.1. The symbol used to denote the boundary of a cloud environment.

Cloud (Cont.)

- IT resources provided by cloud are dedicated to supplying back-end processing capabilities and user-based access to these capabilities.
- A cloud can be based on the use of any protocols that allow for the remote access to its IT resources.
- The party that provides cloud-based IT resources is the cloud provider.
- The party that uses cloud-based IT resources is the cloud consumer.

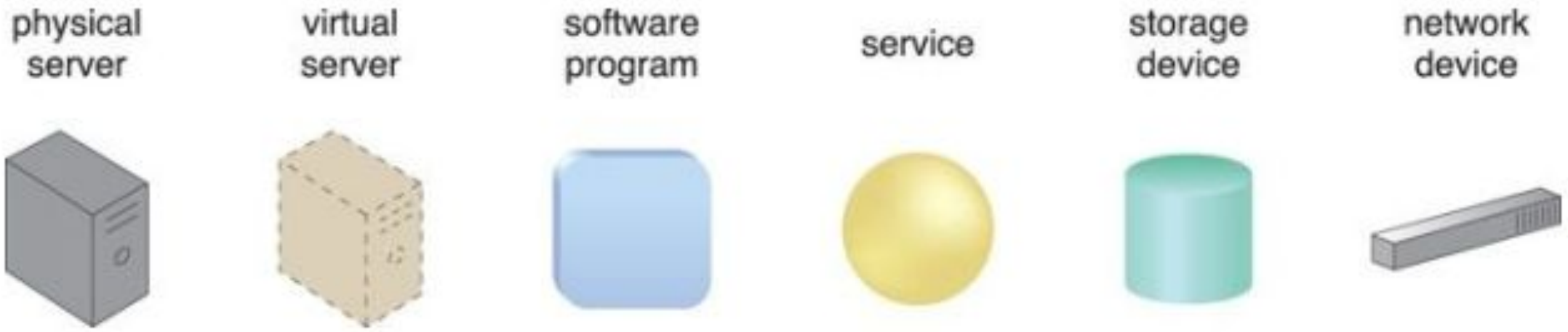
IT Resource (Cont.)

- Figure illustrates how the cloud symbol can be used to define a boundary for a cloud-based environment that hosts and provisions a set of IT resources.



IT Resource

- "An IT resource is a physical or virtual IT-related artifact that can be either software-based (virtual server / custom software program) or hardware-based (physical server / network device)."
- Physical servers or hosts are responsible for hosting virtual servers.



On-Premise

- An IT resource that is hosted in a conventional IT enterprise within an organizational boundary (that does not specifically represent a cloud) is considered to be located on the premises of the IT enterprise.

Key Points:

- An on-premise IT resource can access and interact with a cloud-based IT resource.
- An on-premise IT resource can be moved to a cloud, thereby changing it to a cloud-based IT resource.
- Redundant deployments of an IT resource can exist in both on-premise and cloud-based environments.
  - Example: Primary data on-premises; backup on cloud.

Scaling

- Scaling represents the ability of the IT resource to handle increased or decreased usage demands.

Types of Scaling:

- Horizontal Scaling
  - Scaling out and scaling in
- Vertical Scaling
  - Scaling up and scaling down

Horizontal Scaling

- The horizontal allocation of resources is referred to as scaling out.
- The horizontal releasing of resources is referred to as scaling in.
- Horizontal scaling is a common form of scaling within cloud.

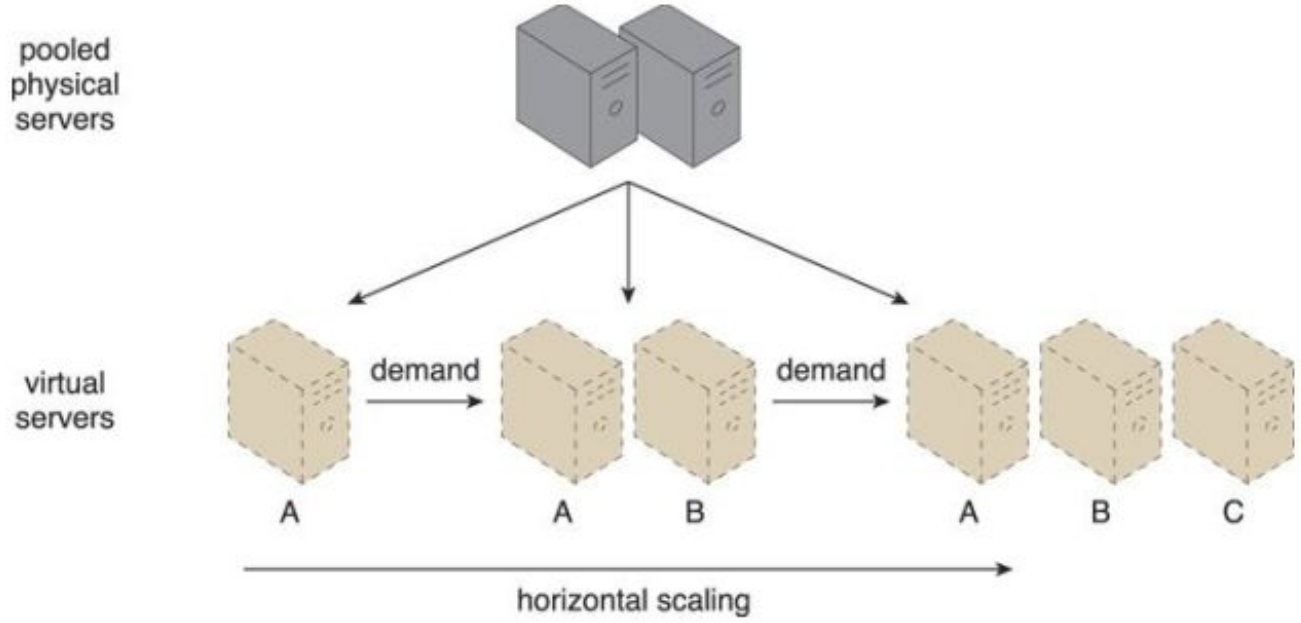


Figure 3.4. An IT resource (Virtual Server A) is scaled out by adding more of the same IT resources (Virtual Servers B and C).

Vertical Scaling

- The replacing of an IT resource with another that has a higher capacity is referred to as scaling up.
- The replacing an IT resource with another that has a lower capacity is referred to as scaling down.
- Vertical scaling is less common in cloud due to the downtime required while the replacement is taking place.

Vertical Scaling (Cont.)

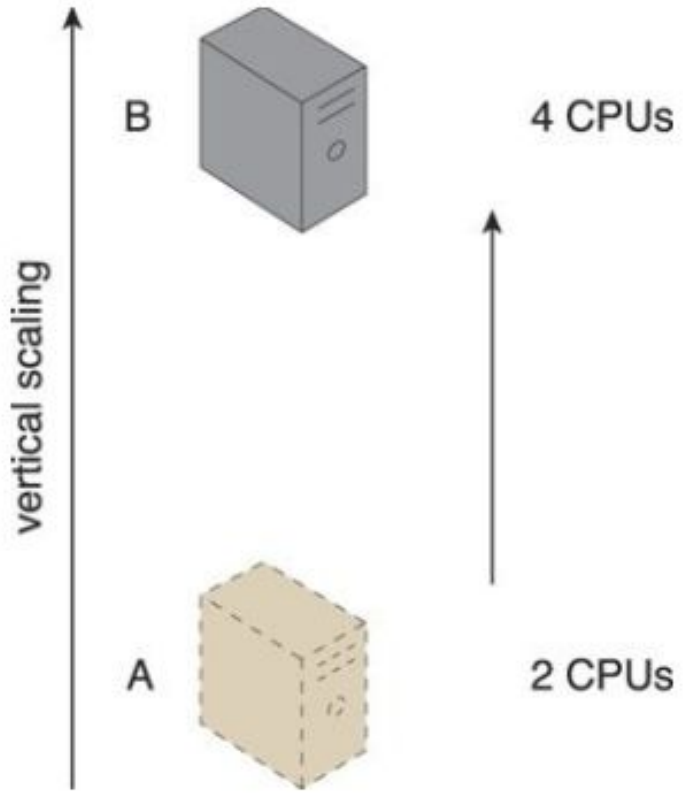


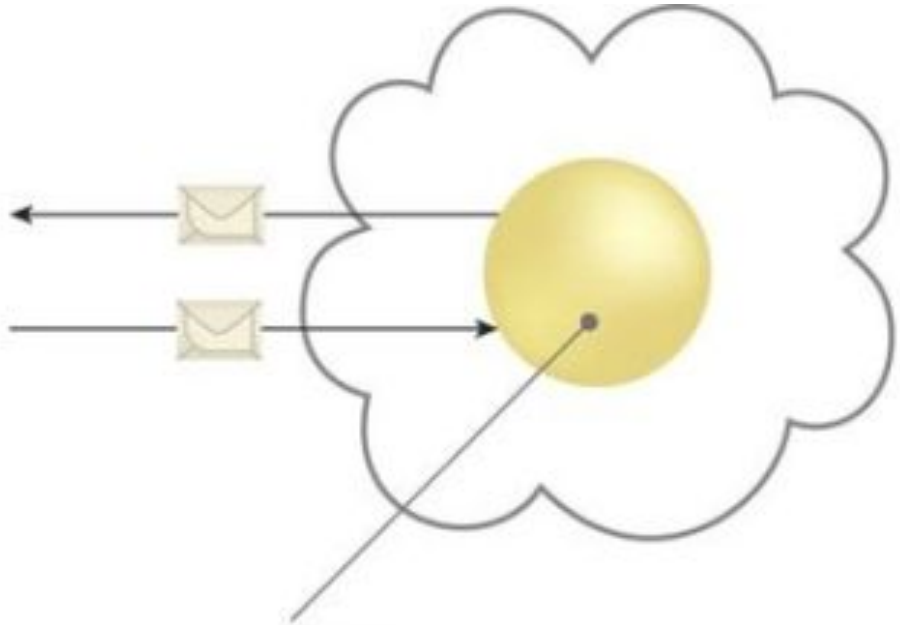
Figure 3.5. An IT resource (a virtual server with two CPUs) is scaled up by replacing it with a more powerful IT resource with increased capacity for data storage (a physical server with four CPUs).

Comparison of Horizontal and Vertical Scaling

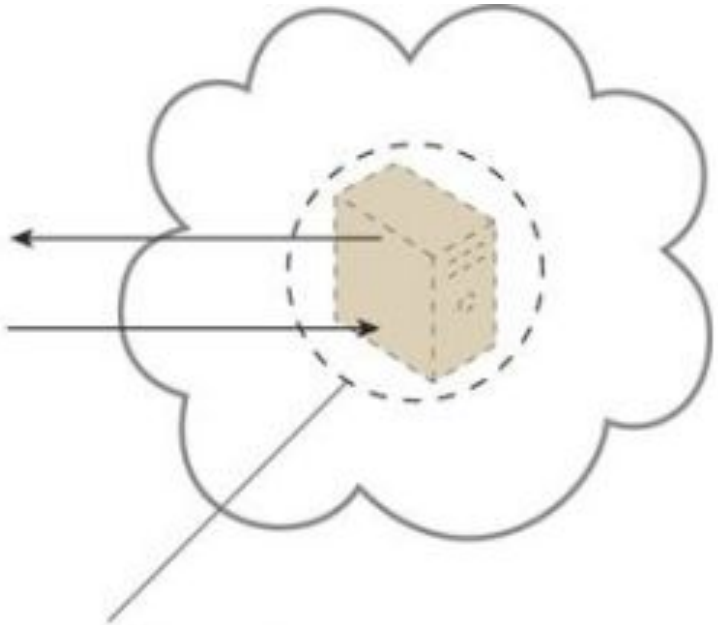
Horizontal Scaling	Vertical Scaling
less expensive (through commodity hardware components)	more expensive (specialized servers)
IT resources instantly available	IT resources normally instantly available
resource replication and automated scaling	additional setup is normally needed
additional IT resources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

Cloud Service

- A cloud service is any IT resource that is made remotely accessible via a cloud.
- The driving motivation behind cloud computing is to provide IT resources as services that encapsulate other IT resources, while offering functions for clients to use and leverage remotely.



remotely accessed Web service
acting as a cloud service



remotely accessed virtual server
acting as a cloud service

Cloud Service (Cont.)

- Cloud service usage conditions are expressed in a service-level agreement (SLA) that is the human-readable part of a service contract between a cloud provider and cloud consumer that describes QoS features, behaviors, and limitations of a cloud-based service.
- An SLA provides details of measurable characteristics related to IT outcomes, such as availability, reliability, pricing and performance.
- Since the implementation of a service is hidden from the cloud consumer, an SLA is a critical specification in the cloud industry.

Cloud Service Consumer

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

Examples of Cloud Service Consumer:

Activity 3

- **Draw an architecture that shows three users are using AWS services.**

  - **Ali is using AWS S3 storage service to upload data on AWS physical storage.**
  - **Eisha is using Elastic Beanstalk service to get a custom web server.**
  - **Talha is using EC2 Instance service to deploy a SQL database server in his cloud environment available on AWS.**

Additional Resources

- **Cloud Computing – Concepts, Technology, and Architecture by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini**

  - **Chapter 3: Understanding Cloud Computing**

Questions?