Unit 1
Cloud Computing Fundamentals

1-Motivation for Cloud Computing

Cloud computing is needed in getting the services of computing resources. Thus, one can say as a one-line answer to the need for cloud computing that it eliminates a large computing investment without compromising the use of computing at the user level at an operational cost. Cloud computing is very economical and saves a lot of money. A blind benefit of this computing is that even if we lose our laptop or due to some crisis our personal computer—and the desktop system—gets damaged, still our data and files will stay safe and secured as these are not in our local machine (but remotely located at the provider's place—machine).In addition, one can think to add security while accessing these remote computing resources as depicted in Figure 2.1.

Figure 2.1 shows several cloud computing applications. The *cloud* represents the Internet-based computing resources, and the accessibility is through some secure support of connectivity. It is a computing solution growing in popularity, especially among individuals and small- and medium-sized companies (SMEs). In the cloud computing model, an organization's core computer power resides offsite and is essentially subscribed to rather than owned.

Thus, cloud computing comes into focus and much needed only when we think about what computing resources and information technology (IT) solutions are required. This need caters to a way to increase capacity or add capabilities on the fly without investing in new infrastructure, training new personnel, or licensing new software. Cloud computing encompasses the subscription-based or pay-per-use service model of offering computing to end users or customers over the Internet and thereby extending the IT's existing capabilities.

FIGURE 2.1
Cloud computing.

2-The Need for Cloud Computing

The main reasons for the need and use of cloud computing are convenience and reliability. In the past, if we wanted to bring a file, we would have to save it to a Universal Serial Bus (USB) flash drive, external hard drive, or compact disc (CD) and bring that device to a different place. Instead, saving a file to the cloud (e.g., use of cloud application Dropbox) ensures that we will be able to access it with any computer that has an Internet connection. The cloud also makes it much easier to share a file with friends, making it possible to collaborate over the web.
While using the cloud, losing our data/file is much less likely. However, just like anything online, there is always a risk that someone may try to gain access to our personal data, and therefore, it is important to choose an access control with a strong password and pay attention to any privacy settings for the cloud service that we are using.

3-Defining Cloud Computing

3.1-NIST Definition of Cloud Computing

The formal definition of cloud computing comes from the 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 config-urable 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 [1].

3.2-Cloud Computing Is a Service

The simplest thing that any computer does is allow us to store and retrieve information. We can store our family photographs, our favorite songs, or even

save movies on it, which is also the most basic service offered by cloud computing.

3.3-Cloud Computing Is a Platform

The World Wide Web (WWW) can be considered as the operating system for all our Internet-based applications. However, one has to understand that we will always need a local operating system in our computer to access web-based applications. The basic meaning of the term *platform* is that it is the support on which applications run or give results to the users. For example, Microsoft Windows is a platform. But, a platform does not have to be an operating system. Java is a platform even though it is not an operating system.
Through cloud computing, the web is becoming a platform. With trends (applications) such as Office 2.0, more and more applications that were originally available on desktop computers are now being converted into web–cloud applications. Word processors like Buzzword and office suites like Google Docs are now available in the cloud as their desktop counterparts. All these kinds of trends in providing applications via the cloud are turning cloud computing into a platform or to act as a platform.

4-Principles of Cloud computing

4.1-Five Essential Characteristics

Cloud computing has five essential characteristics, which are shown in Figure 2.2. Readers can note the word *essential*, which means that if any of these characteristics is missing, then it is not cloud computing:

1. *On-demand self-service*: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.
2. *Broad network access*: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and personal digital assistants [PDAs]).
3. *Elastic resource pooling*: The provider's computing resources are pooled to serve multiple consumers using a multitenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify the location

at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, and network bandwidth.

4. *Rapid elasticity*: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
5. *Measured service*: Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

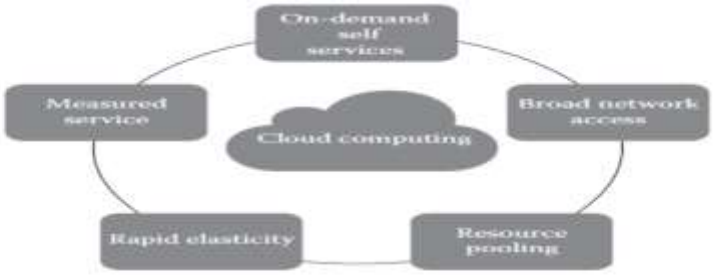

FIGURE 2.2
The essential characteristics of cloud computing.

4.2-Four Cloud Deployment Models

Deployment models describe the ways with which the cloud services can be deployed or made available to its customers, depending on the organizational structure and the provisioning location. One can understand it in this manner too: cloud (Internet)-based computing resources—that is, the locations where data and services are acquired and provisioned to its customers—can take various forms. Four deployment models are usually distinguished, namely, public, private, community, and hybrid cloud service usage:

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

academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

3. *Community cloud*: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

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

4.3- Three Service Offering Models

1. *Cloud SaaS*: The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure, including network, servers, operating systems, storage, and even individual application capabilities, with the possible exception of limited user-specific application configuration settings. The applications are accessible from various client devices through either a thin client

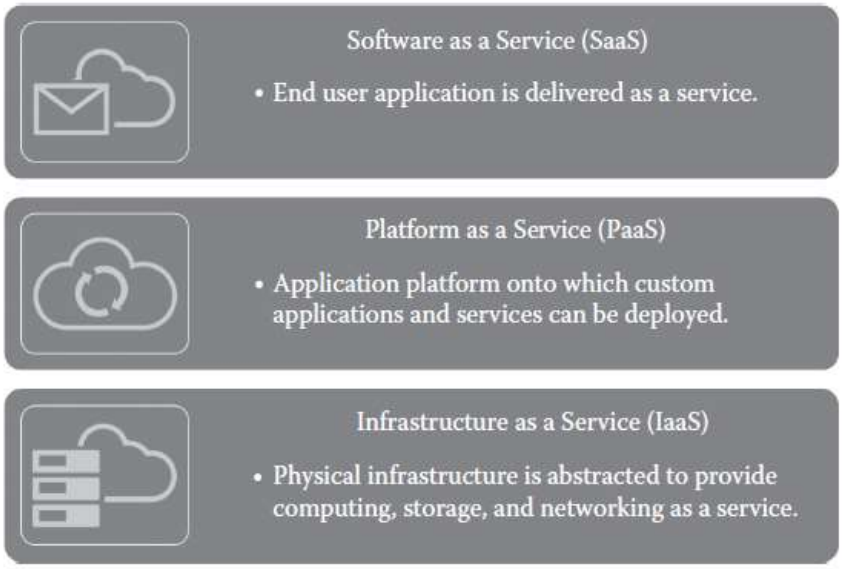

FIGURE 2.3
SPI—service offering model of the cloud.

interface, such as a web browser (e.g., web-based e-mail), or a program interface. The consumer does not manage or control the underlying cloud infrastructure. Typical applications offered as a service include customer

relationship management (CRM), business intelligence analytics, and online accounting software.

2. *Cloud PaaS*: The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure but has control over the deployed applications and possibly configuration settings for the application-hosting environment. In other words, it is a packaged and ready-to-run development or operating framework. The PaaS vendor provides the networks, servers, and storage and manages the levels of scalability and maintenance. The client typically pays for services used. Examples of PaaS providers include Google App Engine and Microsoft Azure Services.
3. *Cloud IaaS*: The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources on a pay-per-use basis where he or she is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over the operating systems, storage, and deployed applications and possibly limited control of select networking components (e.g., host firewalls). The service provider owns the equipment and is responsible for housing, cooling operation, and maintenance. Amazon Web Services (AWS) is a popular example of a large IaaS provider.

The major difference between PaaS and IaaS is the amount of control that users have. In essence, PaaS allows vendors to manage everything, while IaaS requires more management from the customer side. Generally speaking, organizations that already have a software package or application for a specific purpose and want to install and run it in the cloud should opt to use IaaS instead of PaaS.

5-Cloud Ecosystem

Cloud ecosystem is a term used to describe the complete environment or system of interdependent components or entities that work together to enable and support the cloud services. To be more precise, the cloud computing's ecosystem is a complex environment that includes the description of every item or entity along with their interaction; the complex entities include the traditional elements of cloud computing such as software (SaaS), hardware (PaaS and/or IaaS), other infrastructure (e.g., network, storage), and also stakeholders like consultants, integrators, partners, third parties, and anything in their environments that has a bearing on the other components of the cloud.

The cloud ecosystem of interacting components and organizations with individuals, together known as the actors who could be responsible for either providing or consuming cloud services, can be categorized in the following manner:

1. *Cloud service users* (*CSUs*): A consumer (an individual/person), enterprise (including enterprise administrator), and/or government/public institution or organization that consumes delivered cloud services; a CSU can include intermediate users that will deliver cloud services provided by a cloud service provider (CSP) to actual users of the cloud service, that is, end users. End users can be persons, machines, or applications.
2. *CSP*s: An organization that provides or delivers and maintains or manages cloud services, that is, provider of SaaS, PaaS, IaaS, or any allied computing infrastructure.
3. *Cloud service partners* (*CSNs*): A person or organization (e.g., application developer; content, software, hardware, and/or equipment provider; system integrator; and/or auditor) that provides support to the building of a service offered by a CSP (e.g., service integration).

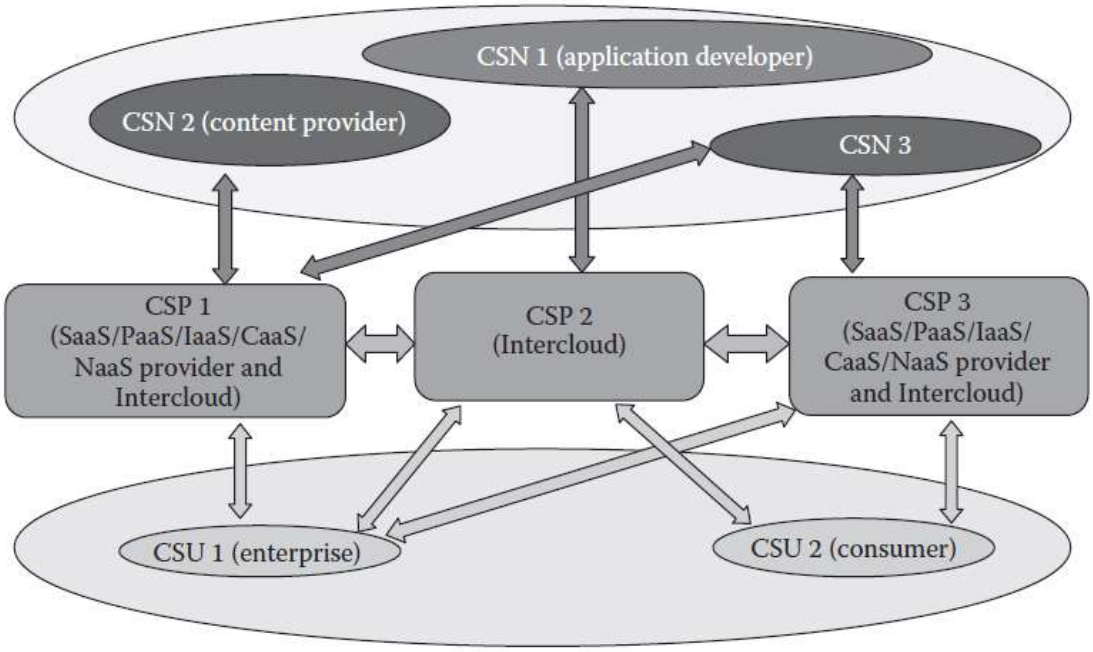

FIGURE 2.4
Actors with some of their possible roles in a cloud ecosystem.

In layman's terms, the cloud ecosystem describes the usage and value of each entity in the ecosystem, and when all the entities in the ecosystem are put together, users are now able to have an integrated suite made up of the best-of-breed solutions. An example of this ecosystem can be a cloud accounting solution such as *Tally*; while this SaaS vendor focuses on their support for accounting and integrated payroll solutions, they can engage (collaborate) with any other third-

party CSPs who could support additional features in the accounting software like reporting tools, dashboards, work papers, workflow, project management, and CRM, covering the majority of a client or customer firm's software needs. And, any other additional requirement that may be essential will likely be added by a partner joining the ecosystem in the near future. Figure 2.4 illustrates the idea of a cloud ecosystem.

6-Requirements for Cloud Services

1. *Multitenancy*: Multitenancy is an essential characteristic of cloud systems aiming to provide isolation of the different users of the cloud system (tenants) while maximizing resource sharing. It is expected that multitenancy be supported at various levels of a cloud infrastructure. As an example, at the application level, multitenancy is a feature that allows a single instance of an application (say, database system) and leverages the economy of scale to satisfy several users at the same time.
2. *Service life cycle management*: Cloud services are paid as per usage and can be started and ended at any time. Therefore, it is required that a cloud service support automatic service provisioning. In addition, metering and charging or billing settlement needs to be provided for services that are dynamically created, modified, and then released in virtual environments.

3. *Security*: The security of each individual service needs to be protected in the multitenant cloud environment; the users (tenants) also support the needed secured services, meaning that a cloud provides strict control for tenants' service access to different resources to avoid the abuse of cloud resources and to facilitate the management of CSUs by CSPs.

4. *Responsiveness*: The cloud ecosystem is expected to enable early detection, diagnosis, and fixing of service-related problems in order to help the customers use the services faithfully.
5. *Intelligent service deployment*: It is expected that the cloud enables efficient use of resources in service deployment, that is, maximizing the number of deployed services while minimizing the usage of resources and still respecting the SLAs. For example, the specific application characteristics (e.g., central processing unit [CPU]-intensive, input/output [IO]-intensive) that can be provided by developers or via application monitoring may help CSPs in making efficient use of resources.
6. *Portability*: It is expected that a cloud service supports the portability of its features over various underlying resources and that CSPs should be able to

accommodate cloud workload portability (e.g., VM portability) with limited service disruption.

7. *Interoperability*: It is expected to have available well-documented and well-tested specifications that allow heterogeneous systems in cloud environments to work together.
8. *Regulatory aspects*: All applicable regulations shall be respected, including privacy protection.
9. *Environmental sustainability*: A key characteristic of cloud computing is the capability to access, through a broad network and thin clients, on-demand shared pools of configurable resources that can be rapidly provisioned and released. Cloud computing can then be considered in its essence as an ICT energy consumption consolidation model, supporting mainstream technologies aiming to optimize energy consumption (e.g., in data centers) and application performance. Examples of such technologies include virtualization and multitenancy.
10. *Service reliability, service availability, and quality assurance*: CSUs demand for their services end-to-end quality of service (QoS) assurance, high levels of reliability, and continued availability to their CSPs.
11. *Service access*: A cloud infrastructure is expected to provide CSUs with access to cloud services from any user device. It is expected that CSUs have a consistent experience when accessing cloud services.
12. *Flexibility*: It is expected that the cloud service be capable of supporting multiple cloud deployment models and cloud service categories.
13. *Accounting and charging*: It is expected that a cloud service be capable to support various accounting and charging models and policies.
14. *Massive data processing*: It is expected that a cloud supports mechanisms for massive data processing (e.g., extracting, transforming, and loading data). It is worth to note in this context that distributed and/ or parallel processing systems will be used in cloud infrastructure deployments to provide large-scale integrated data storage and processing capabilities that scale with software-based fault tolerance.

The expected requirements for services in the IaaS category include the following:

- Computing hardware requirements (including processing, memory, disk, network interfaces, and virtual machines)
- Computing software requirements (including OS and other preinstalled software)
- Storage requirements (including storage capacity)

- Network requirements (including QoS specifications, such as bandwidth and traffic volumes)
- Availability requirements (including protection/backup plan for computing, storage, and network resources)

The expected service requirements for services in the PaaS category include the following:

- Requirements similar to those of the IaaS category
- Deployment options of user-created applications (e.g., scale-out options)

The expected service requirements for services in the SaaS category include the following:

- Application-specific requirements (including licensing options)
- Network requirements (including QoS specifications such as bandwidth and traffic volumes)

7-Cloud Application

A cloud application is an application program that functions or executes in the cloud; the application can exhibit some characteristics of a pure desktop application and some characteristics of a pure web-based application. A desktop application resides entirely on a single device at the user's location (it does not necessarily have to be a desktop computer), and on the other hand, a web application is stored entirely on a remote server and is delivered over the Internet through a browser interface. Like desktop applications, cloud applications can provide fast responsiveness and can work offline. Like web applications, cloud applications need not permanently reside on the local device, but they can be easily updated online. Cloud applications are, therefore, under the user's constant control, yet they need not always consume storage space on the user's computer or communications device. Assuming that the user has a reasonably fast Internet connection, a well-written cloud application offers all the interactivity of a desktop application along with the portability of a web application.

A cloud application can be used with a web browser connected to the Internet. Now, it is possible for the user interface portion of the application to exist on the local device and for the user to cache data locally, enabling full offline mode when desired. Also, a cloud application, unlike a web app, can be used in any sensitive situation where wireless devices—connectivity—are not allowed (i.e., even when no Internet connection is available for some period).

An example of cloud application is a web-based e-mail (e.g., Gmail, Yahoo mail); in this application, the user of the e-mail uses the cloud—all of the emails in their inbox are stored on servers at remote locations at the e-mail service provider.

However, there are many other services that use the cloud in different ways. Here is yet another example: Dropbox is a cloud storage service that lets us easily store and share files with other people and access files from a mobile device as well.

8-Benefits and Drawbacks

A consolidated set of points briefing the benefits of cloud computing can be as follows:

1. *Achieve economies of scale*: We can increase the volume output or productivity with fewer systems and thereby reduce the cost per unit of a project or product.
2. *Reduce spending on technology infrastructure*: It is easy to access data and information with minimal upfront spending in a *pay-as-you-go* approach, in the sense that the usage and payment are similar to an electricity meter reading in the house, which is based on demand.
3. *Globalize the workforce*: People worldwide can access the cloud with Internet connection.
4. *Streamline business processes*: It is possible to get more work done in less time with less resource.
5. *Reduce capital costs*: There is no need to spend huge money on hardware, software, or licensing fees.
6. *Pervasive accessibility*: Data and applications can be accessed anytime, anywhere, using any smart computing device, making our life so much easier.
7. *Monitor projects more effectively*: It is possible to confine within budgetary allocations and can be ahead of completion cycle times.
8. *Less personnel training is needed*: It takes fewer people to do more work on a cloud, with a minimal learning curve on hardware and software issues.
9. *Minimize maintenance and licensing software*: As there is no too much of on-premise computing resources, maintenance becomes simple and updates and renewals of software systems rely on the cloud vendor or provider.
10. *Improved flexibility*: It is possible to make fast changes in our work environment without serious issues at stake.

Drawbacks to cloud computing are obvious. The main point in this context is that if we lose our Internet connection, we have lost the link to the cloud and thereby to the data and applications. There is also a concern about security as our entire working with data and applications depend on other's (cloud vendor or providers) computing power. Also, while cloud computing supports scalability (i.e., quickly scaling up and down computing resources depending on the need), it does not permit the control on these resources as these are not owned by the user or customer. Depending on the cloud vendor or provider, customers may face restrictions on the availability of applications, operating systems, and infrastructure options. And, sometimes, all development platforms may not be available in the cloud due to the fact that the cloud vendor may not aware of such solutions. A major barrier to cloud computing is the interoperabebility of applications, which is the ability of two or more applications that are required to support a business need to work together by sharing data and other business-related resources. Normally, this does not happen in the cloud as these applications may not be available with a single cloud vendor and two different vendors having these applications do not cooperate with each other.