CLOUD COMPUTING

LECTURE NOTES

CLOUD COMPUTING (CS4110PE)

UNIT I

COMPUTING PARADIGMS

Computing Paradigms: High-Performance Computing, Parallel Computing, Distributed Computing, Cluster Computing, Grid Computing, Cloud Computing, Bio computing, Mobile Computing, Quantum Computing, Optical Computing, Nano computing.

HIGH-PERFORMANCE COMPUTING

In high-performance computing systems, a pool of processors (processor machines or central processing units [CPUs]) connected (networked) with other resources like memory, storage, and input and output devices, and the deployed software is enabled to run in the entire system of connected components.

The processor machines can be of homogeneous or heterogeneous type. The legacy meaning of high- performance computing (HPC) is the supercomputers; however, it is not true in present-day computing scenarios. Therefore, HPC can also be attributed to mean the other computing paradigms that are discussed in the forthcoming sections, as it is a common name for all these computing systems.

Thus, examples of HPC include a small cluster of desktop computers or personal computers (PCs) to the fastest supercomputers. HPC systems are normally found in those applications where it is required to use or solve scientific problems. Most of the time, the challenge in working with these kinds of problems is to perform suitable simulation study, and this can be accomplished by HPC without any difficulty. Scientific examples such as protein folding in molecular biology and studies on developing models and applications based on nuclear fusion are worth noting as potential applications for HPC.

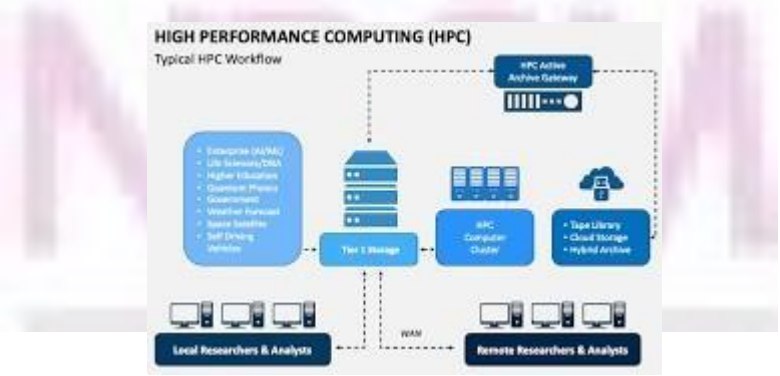


Fig 1 High Performance Computing

PARALLEL COMPUTING

Parallel computing is also one of the facets of HPC. Here, a set of processors work cooperatively to solve a computational problem. These processor machines or CPUs are mostly of homogeneous type. Therefore, this definition is *the same* as that of HPC and is broad enough to include supercomputers that have hundreds or thousands of processors interconnected with other resources. One can distinguish between *conventional* (also known as serial or sequential or Von Neumann) computers and parallel computers in the way the applications are executed.

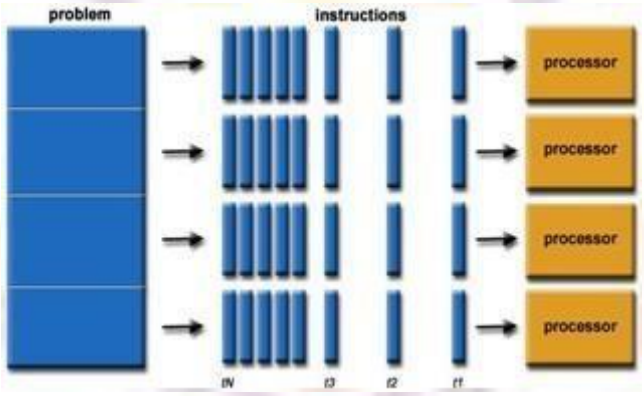


Fig 2 Parallel Computing

In serial or sequential computers, the following apply:

1) It runs on a single computer/processor machine having a single CPU.
2) A problem is broken down into a discrete series of instructions. Instructions are executed one after another.

In parallel computing, since there is simultaneous use of multiple processor machines, the following apply:

1) It is run using multiple processors (multiple CPUs).
2) A problem is broken down into discrete parts that can be solved concurrently.
3) Each part is further broken down into a series of instructions.
4) Instructions from each part are executed simultaneously on different processors.
5) An overall control/coordination mechanism is employed.

DISTRIBUTED COMPUTING

Distributed computing is also a computing system that consists of multiple computers or processor machines connected through a network, which can be homogeneous or heterogeneous, but run as a single system. The connectivity can be such that the CPUs in a distributed system can be physically close together and connected by a local network, or they can be geographically distant and connected by a wide area network. The heterogeneity in a distributed system supports any number of possible configurations in the processor machines, such as mainframes, PCs, workstations, and minicomputers. The goal of distributed computing is to make such a network work as a single computer.

Distributed computing systems are advantageous over centralized systems, because there is a supportfor the following characteristic features:

Scalability: It is the ability of the system to be easily expanded by adding more machines as needed,and vice versa, without affecting the existing setup.

Redundancy or replication: Here, several machines can provide the same services, so that even ifone is unavailable (or failed), work does not stop because other similar computing supports will be available.

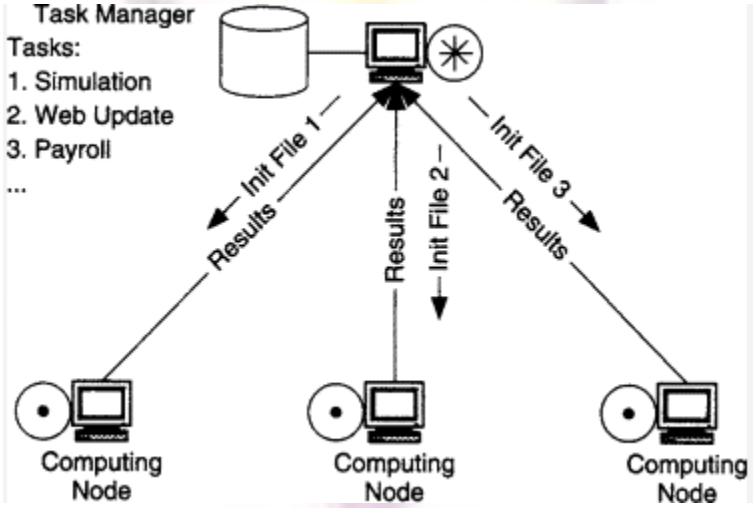


Fig 3 Distributed Computing

CLUSTER COMPUTING

A cluster computing system consists of a set of the same or similar type of processor machines connected using a dedicated network infrastructure. All processor machines share resources such as a common home directory and have software such as a message passing interface (MPI) implementation installed to allow programs to be run across all nodes simultaneously. This is also a kind of HPC category. The individual computers in a cluster can be referred to as *nodes*. The reason to realize a cluster as HPC is due to the fact that the individual nodes can work together to solve a problem larger than any computer can easily solve. And, the nodes need to communicate with one another in order to work cooperatively and meaningfully together to solve the problem in hand.

If we have processor machines of heterogeneous types in a cluster, this kind of clusters become a subtype and still mostly are in the experimental or research stage.

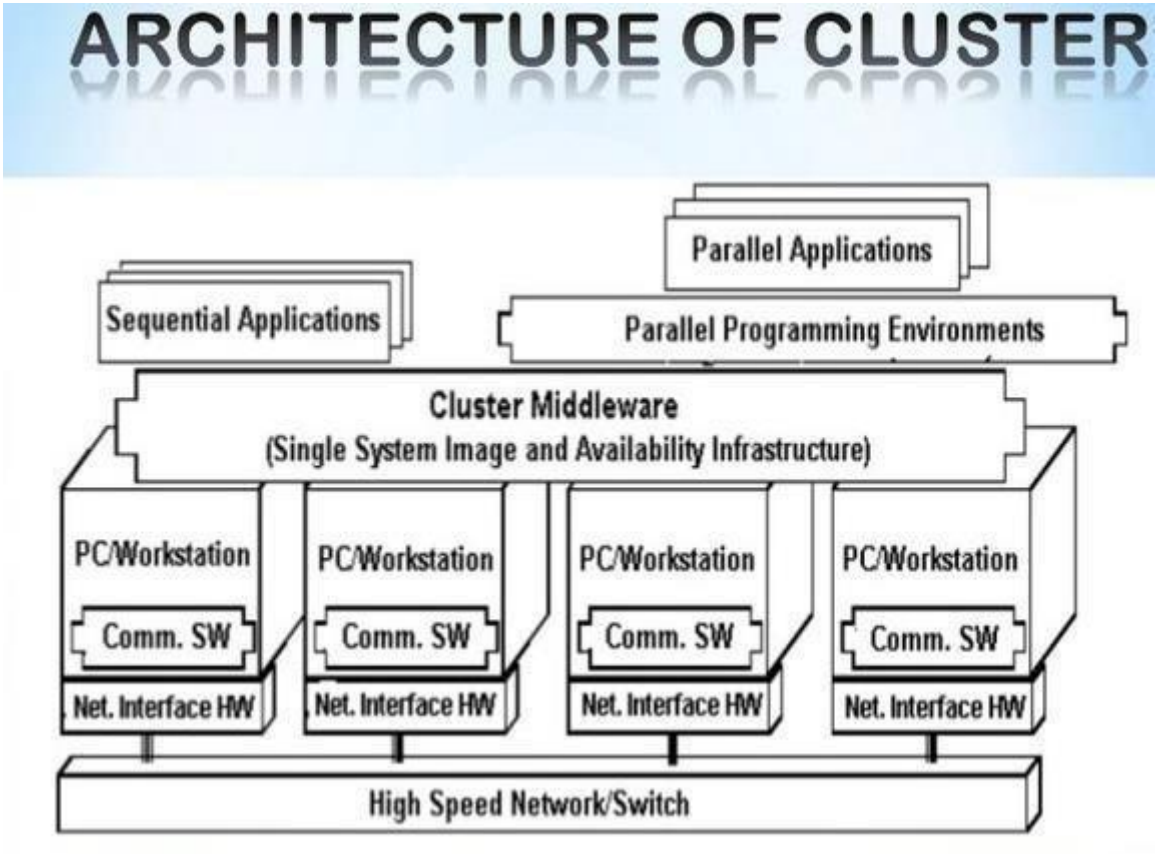


Fig 4 Cluster Computing

GRID COMPUTING

The computing resources in most of the organizations are underutilized but are necessary for certain operations. The idea of grid computing is to make use of such non utilized computing power by the needy organizations, and thereby the return on investment (ROI) on computing investments can be increased. Thus, grid computing is a network of computing or processor machines managed with a kind of software such as middleware, in order to access and use the resources remotely. The managing activity of grid resources through the middleware is called *grid services*. Grid services provide accesscontrol, security, access to data including digital libraries and databases, and access to large-scale interactive and long-term storage facilities.

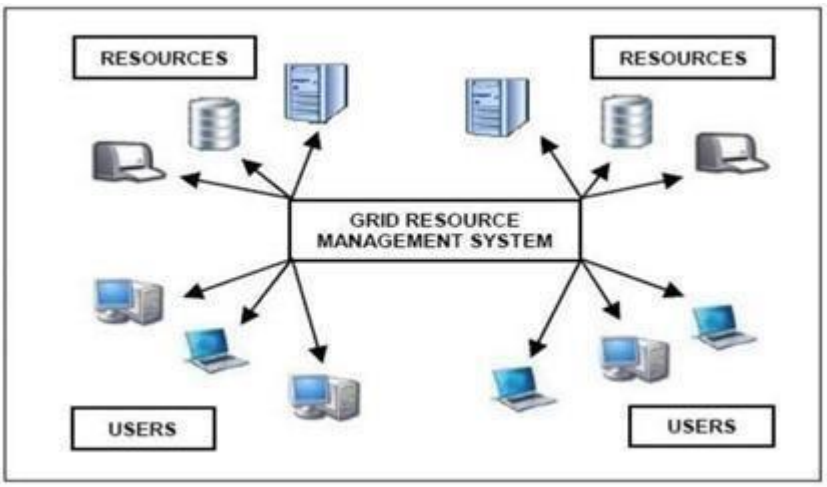


Fig 5 Grid Computing

Grid computing is popular for the following reasons:

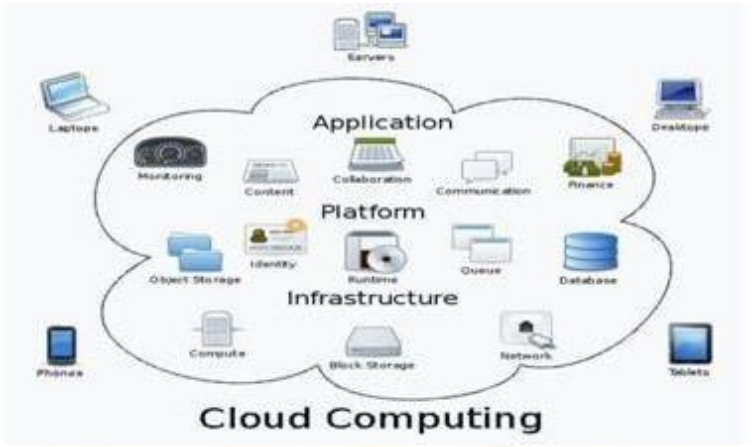
1) Its ability to make use of unused computing power, and thus, it is a cost- effective solution (reducing investments, only recurring costs)
2) As a wayto solve problems in line with any HPC-based application
3) Enables heterogeneous resources of computers to work cooperatively and collaboratively to solvea scientific problem
4) Researchers associate the term *grid* to the way electricity is distributed in municipal areas for the common man. In this context, the difference between electrical power grid and grid computing is worth noting (Table 1.1).

CLOUD COMPUTING

The computing trend moved toward cloud from the concept of grid computing, particularly when large computing resources are required to solve a single problem, using the ideas of computing poweras a *utility* and other allied concepts. However, the potential difference between grid and cloud is that grid computing supports leveraging several computers in parallel to solve a particular application, while cloud computing supports -leveraging multiple resources, including computing resources, to deliver a unified *service* to the end user.

In cloud computing, the IT and business resources, such as servers, storage, network, applications, and processes, can be dynamically provisioned to the user needs and workload.

In addition, while a cloud can provision and support a grid, a cloud can also support non grid environments, such as a three-tier web architecture running on traditional or Web 2.0 applications.



BIO COMPUTING

Bio computing systems use the concepts of biologically derived or simulated molecules (or models) that perform computational processes in order to solve a problem. The biologically derived models aid in structuring the computer programs that become part of the application.

Bio computing provides the theoretical background and practical tools for scientists to explore proteins and DNA. DNA and proteins are nature's building blocks, but these building blocks are not exactly used as *bricks*; the function of the final molecule rather strongly depends on the *order* of these blocks. Thus, the bio computing scientist works on inventing the *order* suitable for various applications mimicking biology. Bio computing shall, therefore, lead to a better understanding of life and the molecular causes of certain diseases.

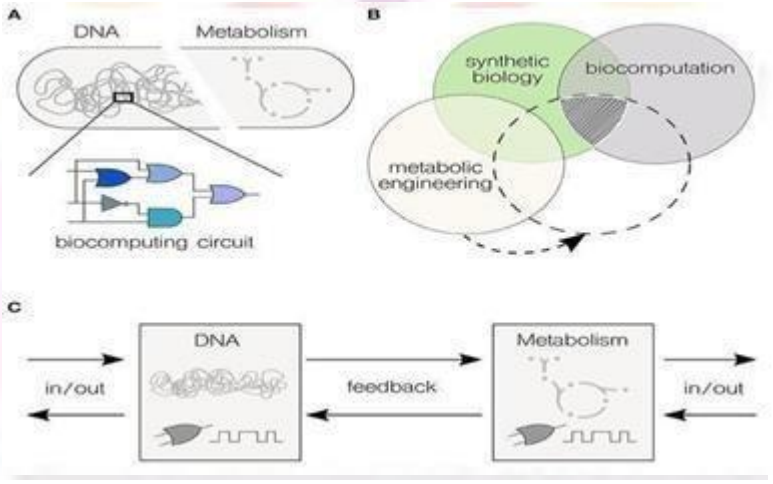


Fig 7 Bio Computing

MOBILE COMPUTING

In mobile computing, the processing (or computing) elements are small (i.e., handheld devices) and the communication between various resources is taking place using wireless media.

Mobile communication for voice applications (e.g., cellular phone) is widely established throughout the world and witnesses a very rapid growth in all its dimensions including the increase in the number of subscribers of various cellular networks. An extension of this technology is the ability to send and receive data across various cellular networks using small devices such as smart phones. There can be numerous applications based on this technology; for example, video call or conferencing is one of the important applications that people prefer to use in place of existing voice (only) communications on mobile phones.

Mobile computing–based applications are becoming very important and rapidly evolving with various technological advancements as it allows users to transmit data from remote locations to other remote or fixed locations.

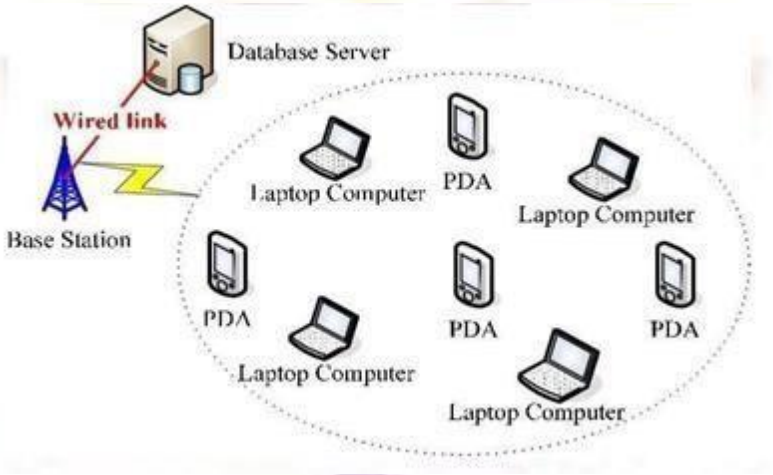


Fig 8 Mobile Computing

QUANTUM COMPUTING

Manufacturers of computing systems say that there is a limit for cramming more and more transistors into smaller and smaller spaces of integrated circuits (ICs) and thereby doubling the processing power about every 18 months.

This problem will have to be overcome by a new *quantum computing*–based solution, wherein the dependence is on quantum information, the rules that govern the subatomic world. Quantum computers are millions of times faster than even our most powerful supercomputers today.

Since quantum computing works differently on the most fundamental level than the current technology, and although there are working prototypes, these systems have not so far proved to be alternatives to today's silicon-based machines.

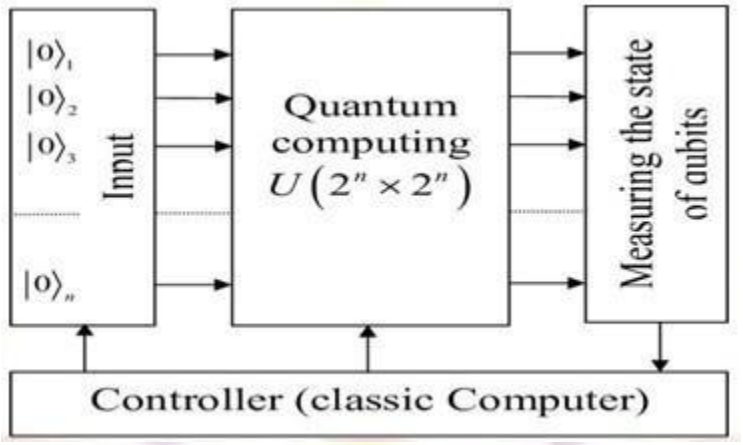


Fig 9 Quantum Computing

OPTICAL COMPUTING

Optical computing system uses the photons in visible light or infrared beams, rather than electric current, to perform digital computations. An electric current flows at only about 10% of the speed of light.

This limits the rate at which data can be exchanged over long distances and is one of the factors that led to the evolution of optical fiber.

By applying some of the advantages of visible and/or IR networks at the device and component scale, a computer can be developed that can perform operations 10or more times faster than a conventional electronic computer.

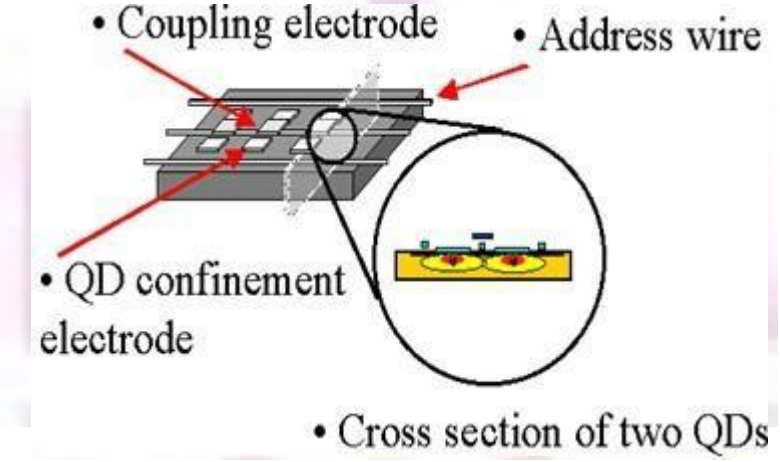


Fig 10 Optical Computing

NANO COMPUTING

Nano computing refers to computing systems that are constructed from nano scale components. The silicon transistors in traditional computers may be replaced by transistors based on carbon nano tubes. The successful realization of nano computersrelates to the scale and integration of these nano tubes or components. The issues of scale relate to the dimensions of the components; they are, at most, a few nanometers in at least two dimensions.

The issues of integration ofthe components are twofold:

1) The manufacture of complex arbitrary patterns may be economically infeasible
2) Nanocomputers may include massive quantities of devices.

Researchers are working on all these issues to bring nano computing a reality.

Fig 11 Devices of Nano Computing

1.11 NETWORK COMPUTING

Network computing is a way of designing systems to take advantage of the latest technology and maximize its positive impact on business solutions and their ability to serve their customers using a strong underlying network of computing resources. In any network computing solution, the client component of a networked architecture or application will be with the customer or client or end user, and in modern days, they provide an essential set of functionality necessary to support the appropriate client functions at minimum cost and maximum simplicity. Unlike conventional PCs, they do not need to be individually configured and maintained according to their intended use. The other end of the client component in the network architecture will be a typical *server* environment to *push* the services of the application to the client end.

Almost all the computing paradigms that were discussed earlier are of this nature. Even in the future, if any one invents a totally new computing paradigm, it would be based on a networked architecture, without which it is impossible to realize the benefits for any end user.

SUMMARY

We are into a post-PC era, in which a greater number and a variety of computers and computing paradigms with different sizes and functions might be used everywhere and with every human being; so, the purpose of this chapter is to illustrate briefly the ideas of all these computing domains, as most of these are ubiquitous and pervasive in its access and working environment.

CLOUD COMPUTING (CS4110PE)

UNIT II
CLOUD COMPUTING FUNDAMENTALS

Cloud Computing Fundamentals: Motivation for Cloud Computing, The Need for Cloud Computing, Defining Cloud Computing, Definition of Cloud computing, Cloud Computing Is a Service, Cloud Computing Is a Platform, Principles of Cloud computing, Five Essential Characteristics, Four Cloud Deployment Models

MOTIVATION FOR CLOUD COMPUTING

Let us review the scenario of computing prior to the announcement and availability of cloud computing: The users who are in need of computing are expected to invest money on computing resources such as hardware, software, networking, and storage; this investment naturally costs a bulk currency to the users as they have to buy these computing resources, keep these in their premises, and maintain and make it operational—all these tasks would add cost. And, this is a particularly true and huge expenditure to the enterprises that require enormous computing power and resources, compared with classical academics and individuals.

On the other hand, it is easy and handy to get the required computing power and resources from some provider (or supplier) as and when it is needed and pay only for that usage. This would cost only a reasonable investment or spending, compared to the huge investment when buying the entire computing infrastructure. This phenomenon can be viewed as *capital expenditure* versus *operational expenditure*. As one can easily assess the huge lump sum required for capital expenditure (whole investment and maintenance for computing infrastructure) and compare it with the moderate or smaller lump sum required for the hiring or getting the computing infrastructure only to the tune of required time, and rest of the time free from that. Therefore, cloud computing is a mechanism of *bringing–hiring or getting the services of the computing power or infrastructure* to an organizational or individual level to the extent required and paying only for the consumed services. One can compare this situation with the usage of electricity (its services) from its producer- cum-distributor (in India, it is the state-

/government-owned electricity boards that give electricity supply to all residences and organizations) to houses or organizations; here, we do not generate electricity (comparable with electricity production– related tasks); rather, we use it only to tune up our requirements in our premises, such as for our lighting and usage of other electrical appliances, and pay as per the electricity meter reading value.

Therefore, 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

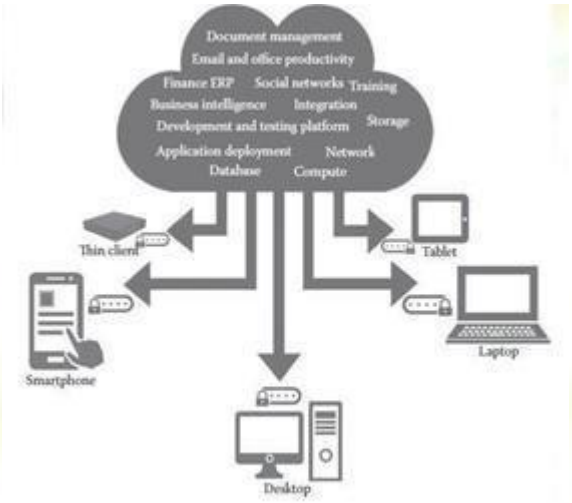


FIGURE 2.1
Cloud computing

Cloud computing 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 subscribedto 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.

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 Drop box) ensures that we will be able to access it with anycomputer that has an Internet connection. The cloud also makes it much easier to share a file with friends, making it possible to collaborate over theweb.

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.

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.

DEFINING CLOUD COMPUTING

In the simplest terms, cloud computing means storing and accessing data and programs over the Internet from a remote location or computer instead of our computer's hard drive. This so called *remote location* has several properties such as scalability, elasticity etc., which is significantly different from a simple remote machine. The cloud is just a metaphor for the Internet. When we store data on or run a program from the local computer's hard drive that is called local storage and computing. For it to be considered *cloud computing*, we need to access our data or programs over

the Internet. The end result is the same; however, with an online connection, cloud computing can be done anywhere, anytime, and by any device.

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 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. It means that the computing resource or infrastructure—be it server hardware, storage, network, or application software.

All available from the cloud vendor or provider's site/premises, can be accessible over the Internet from any remote location and by any local computing device. In addition,- the usage or accessibility is to cost only to the level of usage to the customers based on their needs and demands, also known as the *pay-as-you-go* or *pay-as-per-use* model. If the need is more, more quantum computing resources are made available (provisioning- with elasticity) by the provider. Minimal management effort implies that at the customer's side.

The maintenance of computing systems is very minimal as they will have to look at these tasks only for their local computing devices used for accessing cloud- based resources, not for those computing resources managed at the provider's side. Details of five essential characteristics, three service -models, and four deployment models are provided in the 5-4-3 principles in Section 2.3. Many vendors, pundits, and experts refer to NIST, and both the International Standards Organization (ISO)and the Institute of Electrical and Electronics Engineers (IEEE) back the NIST definition.

Now, let us try to define and understand cloud computing from two other perspectives—as a service and a platform—in the following sections.

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.

Let us look at the example of a popular application called *Flickr* to illustrate the meaning of this section.

While Flickr started with an emphasis on sharing photos and images, it has emerged as a great place tostore those images. In many ways, it is superior to storing the images on your computer:

First, Flickr allows us to easily access our images no matter where we are or what type of device we are using. While we might upload the photos of our vacation from our home computer, later, we can easily access them from our laptop at the office. Second, Flickr lets us share the images. There is no need to burn them to a CD or save them on a flash drive. We can just send someone our Flickr address to share these photos or images.

Third, Flickr provides data security. By uploading the images to Flickr, we are providing ourselves with data security by creating a backup on the web. And, while it is always best to keep a local copy—either on a computer, a CD, or a flash drive— the truth is that we are far more likely to lose the images that we store locally than Flickr is of losing our images.

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 ourcomputer 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 counter-parts. All these kinds of trends in providing applications via the cloud are turning cloud computing into a platform or to act as a platform.

PRINCIPLES OF CLOUD COMPUTING

The 5-4-3 principles put forth by NIST describe (a) the five essential characteristic features that promote cloud computing, (b) the four deployment models that are used to narrate the cloud computing opportunities for customers while looking at architectural models, and (c) the three important and basic service offering models of cloud computing.

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:

***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.

***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]).

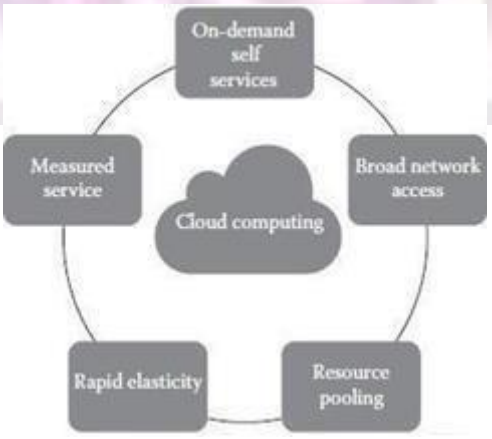


FIGURE 2.2

***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.

***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 anytime.

***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 boththe provider and consumer of the utilized service.

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 bythe organization, a third party.

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 offpremise.

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

THREE SERVICE OFFERING MODELs

The three kinds of services with which the cloud-based computing resources are available to end customers are as follows: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). It is also known as the service–platform– infrastructure (SPI) model of the cloud and is shown in Figure 2.3.

1) **SaaS** is a software distribution model in which applications (software, which is one of the most important computing resources) are hosted by a vendor or service provider and made available to customers over a network, typically the Internet.

2) **PaaS** is a paradigm for delivering operating systems and associated services (e.g., computer aided software engineering [CASE] tools, integrated development environments [IDEs] for developing software solutions) over the Internet without downloads or installation.

3) **IaaS** involves outsourcing the equipment used to support operations, including storage, hardware, servers, and networking components.

***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 or through a web program.

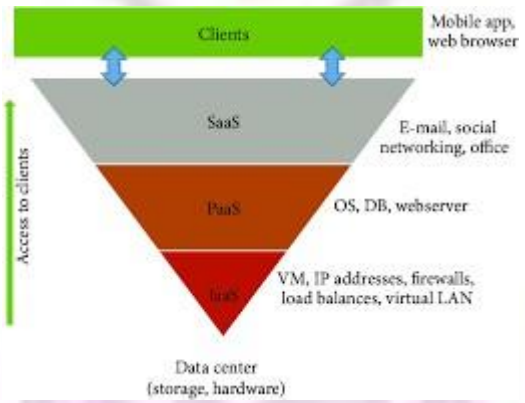


FIGURE 2.3
SPI—service offering model of the cloud.

interface, such as a web browser (e.g., web-based e-mail), or a pro-gram 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.

***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.

***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 includeoperating 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.

CLOUD ECOSYSTEM

Cloud ecosystem is a term used to describe the complete environment or sys-tem 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.

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.

Fig 2.4 Cloud Ecosystem

REQUIREMENTS FOR CLOUD SERVICES

From the concepts illustrated in the earlier sections, one can understand that the cloud services or service offering models require certain features to be exhibited inorder to be considered as *services*. The following are the basic requirements for anything that can be considered as a service by the actors of the cloud computing ecosystem, which can be offered or provisioned through the cloud:

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 infra-structure. 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 pro-vided 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 worktogether.

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 rap-idly 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 accessto 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:

1. Computing hardware requirements (including processing, memory, disk, network interfaces, and virtual machines)
2. Computing software requirements (including OS and other pre installed software)
3. Storage requirements (including storage capacity)
4. Network requirements (including QoS specifications, such as bandwidth and traffic volumes)

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

1) Requirements similar to those of the IaaS category
2) Deployment options of user-created applications (e.g., scale-out options)
3) The expected service requirements for services in the SaaS category include the following:
4) Application-specific requirements (including licensing options)

Network requirements (including QoS specifications such as band-width and traffic volumes)

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 desk-top 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 applicationis 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 froma mobile device as well.

BENEFITS AND DRAWBACKS

One of the attractions of cloud computing is accessibility. If our applications and documents are in the cloud and are not saved on an office server, then we can access and use them at anytime, anywhere for our working, whether we are at work, at home, or even at a friend's house. Cloud computing also enables precisely the right amount of computing power and resources to be used for applications. Cloud computing vendors provide computing-related services as a bundle of computing power and parcel it out on demand.

Customers can draw and make use as much or as little computing power as they need, being charged only for the usage time/computing power; accordingly, this scheme can save money. This also implies that scalability is one of the cloud computing's big benefits.

When we need more computing power, cloud computing can give instant access to exactly what we need. In the cloud model, an organization's core computer power resides offsite and is essentially subscribed to rather than owned. There is no capital expenditure, only operational expenditure.

It also relieves us from the responsibility and costs of maintenance of the entire computing infrastructure and pushes all these to the cloud vendor or provider. The cloud also offers a new level of reliability. The *virtualization* technology enables a vendor's cloud software to automatically move data from a piece of hardware that goes bad or is pulled offline to a section of the system or hardware that is functioning or operational.

Therefore, the client gets seam-less access to the data. Separate backup systems, with cloud disaster recovery strategies, provide another layer of dependability and reliability. Finally, cloud computing also promotes a *green* alternative to paper- intensive office functions. It is because it needs less computing hardware on premise, and all computing-related tasks take place remotely with minimal computing hard-ware requirement with the help of technological innovations such as virtualization and multi tenancy. Another viewpoint on the *green* aspect is that cloud computing can reduce the environmental impact of building, shipping, housing, and ultimately destroying (or recycling) Computer equipment as no one is going to own many such systems in their premises and managing theoffices with fewer computers that consume less energy comparatively.

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 hard-ware, 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 soft-ware 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 notpermit 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 interoperability 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.

UNIT III

Cloud Computing Architecture and Management

Cloud Computing Architecture and Management: Cloud architecture, Layer, Anatomy of the Cloud, Network Connectivity in Cloud Computing, Applications, on the Cloud, Managing the Cloud, Managing the Cloud Infrastructure Managing the Cloud application, Migrating Application to Cloud, Phases of Cloud Migration Approaches for Cloud Migration.

INTRODUCTION

Cloud computing is similar to other technologies in a way that it also has several basic concepts that one should learn before knowing its core concepts. There are several processes and components of cloud computing that need to be discussed. One of the topics of such prime importance is architecture. Architecture is the hierarchical view of describing a technology. This usually includes the components over which the existing technology is built and the components that are dependent on the technology. Another topic that is related to architecture is anatomy. Anatomy describes the corestructure of the cloud. Once the structure of the cloud is clear, the network connections in the cloud and the details about the cloud application need to be known. This is important as the cloud is a completely Internet dependent technology. Similarly, cloud management discusses the important management issues and ways in which the current cloud scenario is managed.

It describes the way an application and infrastructure in the cloud are managed. Management is important because of the quality of service (QoS) factors that are involved in the cloud. These QoS factors form the basis for cloud computing. All the services are given based on these QoS factors.

Similarly, application migration to the cloud also plays a very important role. Not all applications can be directly deployed to the cloud. An application needs to be properly migrated to the cloud to be considered a proper cloud application that will have all the properties of the cloud.

3.1 Cloud architecture

Cloud Computing, which is one of the demanding technology of the current time and which is givinga new shape to every organization by providing on demand virtualized services/resources. Starting from small to medium and medium to large, every organization use cloud computing services for storing information and accessing it from anywhere and any time only with the help of internet.

Cloud Computing Architecture :

The cloud architecture is divided into 2 parts i.e.

1.Frontend

2.Backend

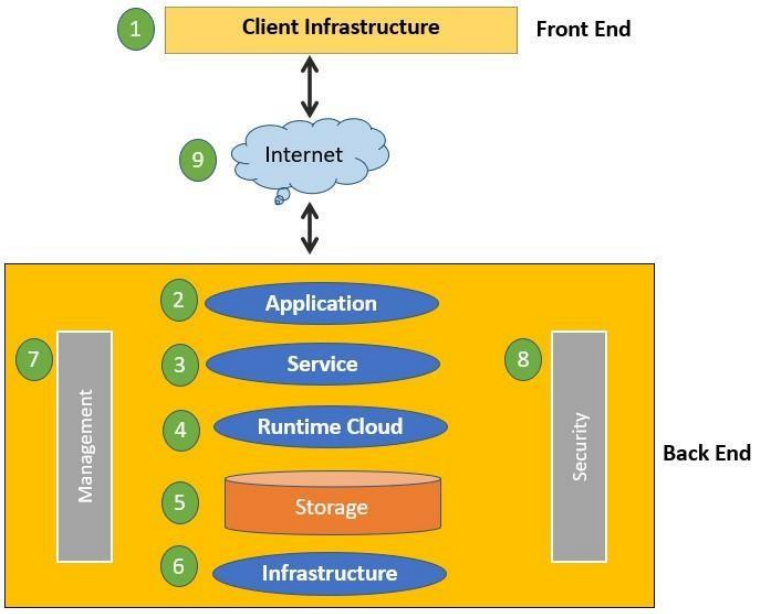


FIGURE 3.1 Cloud Architecture

Front end :

Frontend of the cloud architecture refers to the client side of cloud computing system. Means it contains all the user interfaces and applications which are used by the client to access the cloud computing services/resources. For example, use of a web browser to access the cloud platform.

- **Client Infrastructure –** Client Infrastructure is a part of the frontend component. It contains the applications and user interfaces which are required to access the cloud platform.
- In other words, it provides a GUI(Graphical User Interface) to interact with the cloud.

Back end :

Backend refers to the cloud itself which is used by the service provider. It contains the resources as well as manages the resources and provides security mechanisms. Along with this, it includes huge storage, virtual applications, virtual machines, traffic control mechanisms, deployment models, etc.

1. Application –

Application in backend refers to a software or platform to which client accesses. Means it provides the service in backend as per the client requirement.

2. Service –

Service in backend refers to the major three types of cloud based services like SaaS, PaaS and IaaS. Also manages which type of service the user accesses.

3. Runtime Cloud-

Runtime cloud in backend provides the execution and Runtime platform/environment to the Virtual machine.

4. Storage –

Storage in backend provides flexible and scalable storage service and management of stored data.

5. Infrastructure –

Cloud Infrastructure in backend refers to the hardware and software components of cloud like it includes servers, storage, network devices, virtualization software etc.

6. Management –

Management in backend refers to management of backend components like application, service, runtime cloud, storage, infrastructure, and other security mechanisms etc.

7. Security –

Security in backend refers to implementation of different security mechanisms in the backend for secure cloud resources, systems, files, and infrastructure to end-users.

8. Internet –
Internet connection acts as the medium or a bridge between frontend and backend and establishes the interaction and communication between frontend and backend.

9. Database– Database in backend refers to provide database for storing structured data, such as SQL and NOSQL databases. Example of Databases services include Amazon RDS, Microsoft Azure SQL database and Google CLoud SQL.

10. Networking– Networking in backend services that provide networking infrastructure for application in the cloud, such as load balancing, DNS and virtual private networks.

11. Analytics– Analytics in backend service that provides analytics capabilities for data in the cloud,such as warehousing, business intelligence and machine learning.

Benefits of Cloud Computing Architecture :
- Makes overall cloud computing system simpler.
- Improves data processing requirements.
- Helps in providing high security.

CLOUD Layers

Any technological model consists of an architecture based on which the model functions, which is a hierarchical view of describing the technology. The cloud also has an architecture that describes its working mechanism. It includes the dependencies on which it works and the components that work over it. The cloud is a recent technology that is completely dependent on the Internet for its functioning. Figure 3.2 depicts the layers.

The cloud architecture can be divided into four layers based on the access of the cloud by the user. They are as follows.

LAYER 1 (USER/CLIENT LAYER)

This layer is the lowest layer in the cloud architecture. All the users or client belong to this layer. Thisis the place where the client/user initiates the

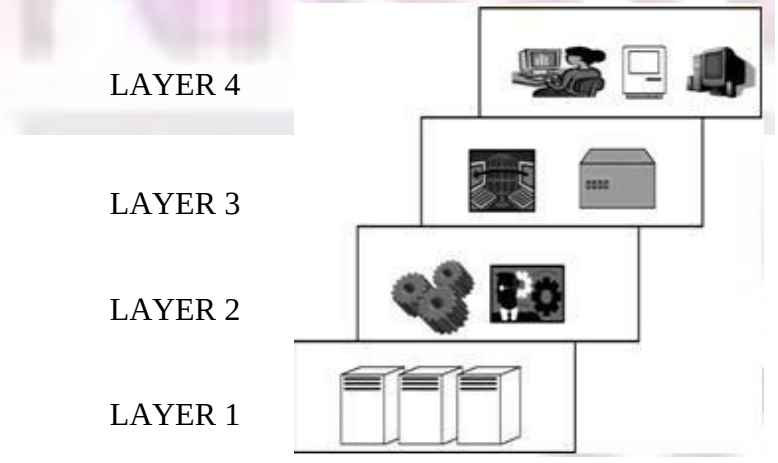


FIGURE 3.2 Cloud Layers.

connection to the cloud. The client can be any device such as a thin client, thick client, or mobile or

any handheld device that would support basic functionalities to access a web application. The thin client here refers to a device that is completely dependent on some other system for its complete functionality. In simple terms, they have very low processing capability. Similarly, thick clients are general computers that have adequate processing capability. They have sufficient capability for independent work. Usually, a cloud application can be accessed in the same way as a web application.But internally, the properties of cloud applications are significantly different. Thus, this layer consists of client devices.

LAYER 2 (NETWORK LAYER)

This layer allows the users to connect to the cloud. The whole cloud infra- structure is dependent on this connection where the services are offered to the customers.

This is primarily the Internet in the case of a public cloud. The public cloud usually exists in a specificlocation and the user would not know the location as it is abstract. And, the public cloud can be accessed all over the world. In the case of a private cloud, the connectivity may be pro-vided by a local area network (LAN).Even in this case, the cloud completely depends on the network that is used. Usually, when accessing the public or private cloud, the users require minimum bandwidth, which is sometimes defined by the cloud providers. This layer does not come under the preview of service-level agreements (SLAs), that is, SLAs do not take into account the Internet connection between the user and cloud for quality of service (QoS).

LAYER 3 (CLOUD MANAGEMENT LAYER)

This layer consists of software that are used in managing the cloud. The software can be a cloud operating system (OS), a software that acts as an interface between the data center (actual resources) and the user, or a management soft- ware that allows managing resources. These software usually allow resource management (scheduling, provisioning, etc.), optimization (server consolidation, storage workload consolidation), and internal cloud governance.

This layer comes under the purview of SLAs, that is, the operations taking place in this layer would affect the SLAs that are being decided upon between the users and the service providers. Any delay in processing or any discrepancy in ser-vice provisioning may lead to an SLA violation.

As per rules, any SLA violation would result in a penalty to be given by the service provider. These SLAs are for both private and public clouds Popular service providers are Amazon Web Services (AWS) and Microsoft Azure for public cloud. Similarly, OpenStack and Eucalyptus allow private cloud creation, deployment, and management.

LAYER 4 (HARDWARE RESOURCE LAYER)

Layer 4 consists of provisions for actual hardware resources. Usually, in the case ofa public cloud, a data center is used in the back end. Similarly, in a private cloud, it can be a data center, which is a huge collection of hardware resources interconnected to each other that is present in a specific location or a high configuration system. This layer comes under the purview of SLAs. This is the most important layer that governs the SLAs. This layer affects the SLAs most in the case of data centers. Whenever a user accesses the cloud, it should be available to the users as quickly as possible and should be within the time that is defined by the SLAs. As mentioned, if there is any discrepancy in provisioning the resources or application, the service provider has to pay the penalty. Hence, the data center consists of a high-speed network connection and a highly efficient algorithm to transfer the data from the data center to the manager.

There can be a number of data centers for a cloud, and similarly, a number of clouds can share a data center.
Thus, this is the architecture of a cloud. The layering is strict, and for any cloud application, this is followed. There can be a little loose isolation between layer 3 and layer 4 depending on the way the cloud is deployed.

ANATOMY OF THE CLOUD

Cloud anatomy can be simply defined as the structure of the cloud. Cloud anatomy cannot be considered the same as cloud architecture. It may not include any dependency on which or over which the technologyworks,

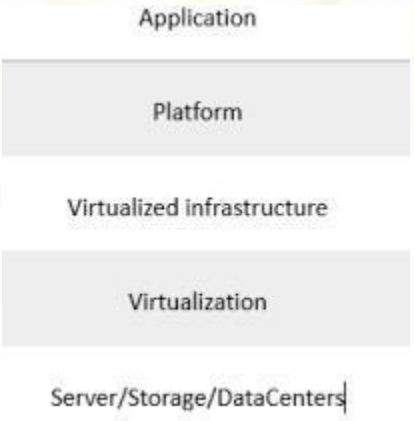


FIGURE 3.3 Anatomy of the Cloud.

where as architecture wholly defines and describes the technology over which it is working. Architecture is a hierarchical structural view that defines the technology as well as the technology over which it is dependent or/and the technology that are dependent on it. Thus, anatomy can be considered as a part of architecture. The basic structure of the cloud is described in Figure 3.2, which can be elaborated, and minute structural details can be given.
Figure 3.2 depicts the most standard anatomy that is the base for the cloud. It depends on the person to choose the depth of description of the cloud. A different view of anatomy is given by Refs. [1,2].

There are basically five components of the cloud:

Application: The upper layer is the application layer. In this layer, any applicationsare executed.
Platform: This component consists of platforms that are responsible for the execution of the application. This platform is between the infrastructure and the application.
Infrastructure: The infrastructure consists of resources over which the other components work. This provides computational capability to the user.
Virtualization: Virtualization is the process of making logical components of resources over the existing physical resources. The logical components are isolated and independent, which form the infrastructure.
Physical hardware: The physical hardware is provided by server and storage units.

NETWORK CONNECTIVITY IN CLOUD COMPUTING

Cloud computing is a technique of resource sharing where servers, storage, and other computing infrastructure in multiple locations are connected by networks. In the cloud, when an application is submitted for its execution, needy and suitable resources are allocated from this collection of resources; as these resources are connected via the Internet, the users get their required results. For many cloud computing applications, network performance will be the key issue to cloud computing performance.

Since cloud computing has various deployment options, we now consider the important aspects related to the cloud deployment models and their accessibility from the viewpoint of network connectivity.

PUBLIC CLOUD ACCESS NETWORKING

In this option, the connectivity is often through the Internet, though some cloud providers may be able to support virtual private networks (VPNs) for customers. Accessing public cloud services will always create issues related to security, which in turn is related to performance. One of the possible approaches toward the support of security is to promote connectivity through encrypted tunnels, so that the information may be sent via secure pipes on the Internet. This procedure will be an overheadin the connectivity, and using it will certainly increase delay and may impact performance. If we want to reduce the delay without compromising security, then we have to select a suitable routing method such as the one reducing the delay by minimizing transit *hops* in the end-to- end connectivity between the cloud provider and cloud consumer.

Since the end-to - end connectivity support is via the Internet, which is a complex federation of interconnected providers (known as Internet service providers [ISPs]),one has to look at the options of selecting the path.

PRIVATE CLOUD ACCESS NETWORKING

In the private cloud deployment model, since the cloud is part of an organizational network, the technology and approaches are local to the in-house network structure. This may include an Internet VPN or VPN service from a network operator. If the application access was properly done with an organizational network— connectivity in a *precloud* configuration—transition to private cloud computing will not affect the access performance.

INTRACLOUD NETWORKING FOR PUBLIC CLOUD SERVICES

Another network connectivity consideration in cloud computing is intra-cloud networking for public cloud services. Here, the resources of the cloud provider and thus the cloud service to the customer are based on the resources that are geographically apart from each other but still connected via the Internet.

Public cloud computing networks are internal to the service provider and thus not visible to the user/customer; however, the security aspects of connectivity Another issue to look for is the QoS in the connected resources worldwide. Most of the performance issues and violations from these are addressed in the SLAs commercially.

PRIVATE INTRACLOUD NETWORKING

The most complicated issue for networking and connectivity in cloud computing is private intracloudnetworking. What makes this particular issue so complex is that it depends on how much intracloud connectivity is associated with the applications being executed in this environment.

Private intra-cloud networking is usually supported over connectivity between the major data center sites owned by the company. At a minimum, all cloud computing implementations will rely on intracloud networking to link users with the resource to which their application was assigned. Once the resource link-age ismade, the extent to which intracloud networking is used depends on whether the application is componentized based on *service-oriented architecture (SOA)* or not, among multiple systems. If the principle of SOA is followed, then traffic may move between components of the application, as well as between the application and the user.

The performance of those connections will then impact cloud computing performance overall. Here too, the impact of cloud computing performance is the differences that exist between the current application and the networkrelationships with the application.
There are reasons to consider the networks and connectivity in cloud computing with newer approaches as globalization and changing network requirements, especially those related to increased Internet usage, are demanding more flexibility in the network architectures of today's enterprises. How are these related to us? The answers are discussed later.

NEW FACETS IN PRIVATE NETWORKS

Conventional private networks have been architected for on-premise applications and maximum Internet security. Typically, applications such as e-mail, file sharing, and *enterprise resource planning* (ERP) systems are delivered to on- premise-based servers at each corporate data center.

Increasingly today, software vendors are offering Software as a Service (SaaS) as an alternative for their software support to the corporate offices, which brings more challenges in the access and usage mechanisms of software from data center servers and in the connectivity of network architectures. The traditional network architecture for these global enterprises was not designed to optimize performance for cloud applications, now that many applications including mission- critical applications are transitioning (moving) from on-premise based to cloud based, wherein the network availability becomes as mission critical as electricity: the business cannot function if it cannot access applications such as ERP and e- mail.

PATH FOR INTERNET TRAFFIC

The traditional Internet traffic through a limited set of Internet gateways poses performance and availability issues for end users who are using cloud-based applications. It can be improved if a more widely distributed Internet gateway infrastructure and connectivity are being supported for accessing applications, as they will provide lower-latency access to their cloud applications. As the volume of traffic to cloud applications grows, the percentage of the legacy network's capacity in terms of traffic to regional gateways increases. Applications such as video conferencing would hog more bandwidth while mission-critical applications such as ERP will consume less bandwidth, and hence, one has to plan a correct connectivity and path between providers and consumers.

APPLICATIONS ON THE CLOUD

The power of a computer is realized through the applications. There are several types of applications. The first type of applications that was developed and used was a stand-alone application. A stand-alone application is
developed- to be run on a single system that does not use network for its functioning. These stand-

alone systems use only the machine in which they are installed. The functioning of these kinds of systems is totally dependent on the resources or features available within the system. These systems do not need the data or processing power of other systems; theyare self-sustaining.

But as the time passed, the requirements of the users changed and certain applications were required, which could be accessed by other users away from the systems. This led to the inception ofweb application.

The web applications were different from the stand-alone applications in many aspects. The main difference was the client server architecture that was followed by the web application. Unlike stand-alone applications, these systems were totally dependent on the network for its working. Here, there are basically two components, called as the client and the server. The server is a high-end machine that consists of the web application installed. This web application is accessed from other client systems. The client can reside anywhere in the network. It can access the web application through the Internet. This type of application was very useful, and this is extensively used from its inception.

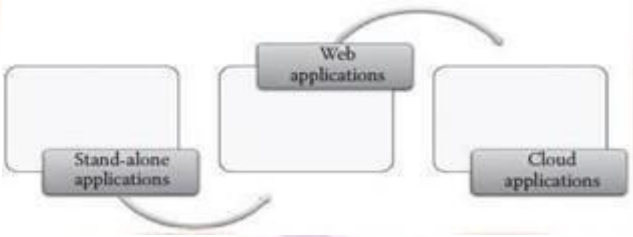


FIGURE 3.3 Computer application evolution

Though this application is much used, there are shortcomings as discussed in the following:

1) The web application is not elastic and cannot handle very heavy loads, that is, it cannot serve highly varying loads.
2) The web application is not multitenant.

The web application does not provide a quantitative measurement of the services that are given to the users, though they can monitor the user.

3) The web applications are usually in one particular platform.
4) The web applications are not provided on a pay-as- you-go basis; thus, a particular service is given to the user for permanent or trial use and usually the timings of user access cannot be monitored.

Due to its non elastic nature, peak load transactions cannot be handled.

Primarily to solve the previously mentioned problem, the cloud applications were developed. Figure depicts the improvements in the applications.

The cloud as mentioned can be classified into three broad access or ser-vice models, Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Cloud application in general refers to a SaaS application.

A cloud application is different from other applications; they have unique features. A cloud application usually can be accessed as a web application but its properties differ. According to NIST[3], the features that make cloud applications unique are described in the following (Figure 3.4 depicts the features of a cloud application):

***Multitenancy*:** Multitenancy is one of the important properties of cloud that make it different from other types of application in which the software can be shared by different users with full independence-. Here, independence refers to logical independence.

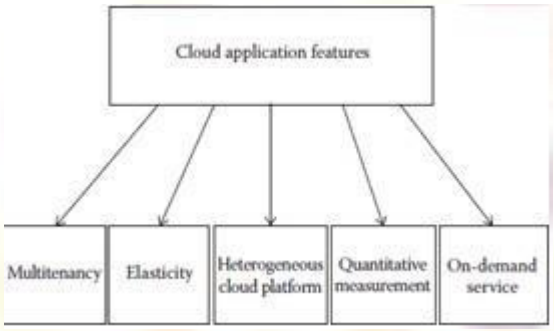


FIGURE 3.4 Features of cloud.

Each user will have a separate application instance and the changes in one application would not affect the other. Physically, the soft-ware is shared and is not independent.

The degree of physical isolation is very less. The logical independence is what is guaranteed. There are no restrictions in the number of applications being shared. The difficulty in providing logical isolation depends on the physical isolation to a certain extent. If an application is physically too close, then it becomes difficult to provide multitenancy. Web application and cloud application are similar as the users use the same way to access both. Figure 3.5 depicts a multitenant application where several users share the same application.

***Elasticity*:** Elasticity is also a unique property that enables the cloud to serve better. According to Herbst et al. [4], elasticity can be defined as the degree to which a system is able to adapt to workload changes by provisioning and deprovisioning resources in an autonomic manner such that at each point in time, the available resources match the current demand as closely as possible. Elasticity allows the cloud providers to efficiently handle the number of users.

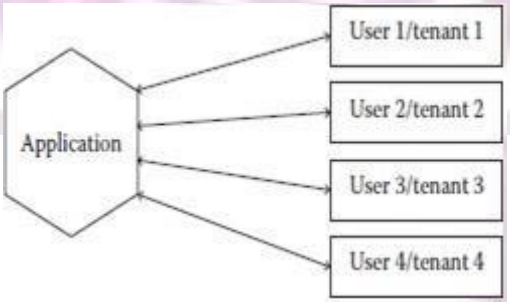


FIGURE 3. Multitenancy.

In addition to this, it supports the rapid fluctuation of loads, that is, the increase or decrease in the number of users and their usage can rapidly change.

***Heterogeneous cloud platform*:** The cloud platform supports heterogeneity, wherein any type of application can be deployed in the cloud. Because of this property, the cloud is flexible for the developers, which facilitates deployment. The applications that are usually deployed can be accessedby the users using a web browser.

***Quantitative measurement*:** The services provided can be quantitatively measured. The user is usually offered services based on certain charges. Here, the application or resources are given as a utility on a pay-per-use basis. Thus, the use can be monitored and measured. Not only the services are measureable, but also the link usage and several other parameters that support cloud applications can be measured. This property of measuring the usage is usually not available in a web application and is a unique feature for cloud-based applications.

***On-demand service*:** The cloud applications offer service to the user, on demand, that is, whenever the user requires it. The cloud service would allow the users to access web applications usually without anyrestrictions on time, duration, and type of device used.

The previously mentioned properties are some of the features that make cloud a unique application platform. These properties mentioned are specific to the cloud hence making it as one of the few technologies that allows application developers to suffice the user's needs seamlessly without any disruption.

MANAGING THE CLOUD

Cloud management is aimed at efficiently managing the cloud so as to maintain the QoS. It is one of the prime jobs to be considered. The whole cloud is dependent on the way it is managed. Cloud management can be divided into two parts:

Managing the infrastructure of the cloud & Managing the cloud application.

MANAGING THE CLOUD INFRASTRUCTURE

The infrastructure of the cloud is considered to be the backbone of the cloud. This component is mainly responsible for the QoS factor. If the infrastructure is not properly managed, then the whole cloud can fail and QoS would be adversely affected. The core of cloud management is resource management. Resource management involves several internal tasks such as resource scheduling, provisioning, and load balancing. These tasks are mainly man-aged by the cloud service provider's core software capabilities such as the cloud OS that is responsible for providing services to the cloud and that internally controls the cloud. A cloud infrastructure is a very complex system that consists of a lot of resources. These resources are usually shared by several users.

Poor resource management may lead to several inefficiencies in terms of performance, functionality, and cost. If a resource is not efficiently managed, the performance of the whole system is affected.

Performance is the most important aspect of the cloud, because everything in the cloud is dependent on the SLAs and the SLAs can be satisfied only if performance is good. Similarly, the basic functionality of the cloud should always be provided and considered at any cost. Even if there is a small discrepancy in providing the functionality, the whole purpose of maintaining the cloud is futile. A partially functional cloud would not satisfy the SLAs.

Lastly, the reason for which the cloud was developed was cost. The cost is a very important criterion as far as the business prospects of the cloud are concerned. On the part of the service providers, if they incur less cost for managing the cloud, then they would try to reduce the cost so as to get a strong user base. Hence, a lot of users would use the services, improving their profit margin. Similarly, if the cost of resource management is high, then definitely the cost of accessing the resources would be high and there is never a lossy business from any organization and so the

serviceprovider would not bear the cost and hence the users have to pay more. Similarly, this would prove costly for service providers as they have a high chance of losing a wide user base, leading to only a marginal growth in the industry. And, competing with its industry rivals would become a big issue. Hence, efficient management with less cost is required.

At a higher level, other than these three issues, there are few more issues that depend on resource management. These are power consumption and optimization of multiple objectives to further reduce the cost. To accomplish these tasks, there are several approaches followed, namely, consolidation of server and storage workloads. Consolidation would reduce the energy consumption and in some cases would increase the performance of the cloud. According to Margaret Rouse [5], server consolidation by definition is an approach to the efficient usage of computer server resources in order to reduce the total number of servers or server locations that an organization requires.

The previously discussed prospects are mostly suitable for IaaS. Similarly, there are different management methods that are followed for different types of service delivery models. Each of the type has its own way of management. All the management methodologies are based on load fluctuation. Load fluctuation is the point where the workload ofthe system changes continuously.

This is one of the important criteria and issues that should be considered for cloud applications. Load fluctuation can be divided into two types: predictable and unpredictable. Predictable load fluctuations are easy to handle. The cloud can be preconfigured for handling such kind of fluctuations. Whereas unpredictable load fluctuations are difficult to handle, ironically this is one of the reasons why cloud is preferred by several users.

This is as far as cloud management is concerned. Cloud governance is another topic that is closely related to cloud management. Cloud governance is different from cloud management. Governance in general is a term in the corporate world that generally involves the process of creating value to an organization by creating strategic objectives that will lead to the growth of the company and would maintain a certain level of control over the company. Similar to that, here cloud organization is involved.

There are several aspects of cloud governance out of which SLAs are one of the important aspects. SLAs are the set of rules that are defined between the user and cloud service provider that decide upon the QoS factor. If SLAs are not followed, then the defaulter has to pay the penalty. The whole cloud is governed by keeping these SLAs in mind. Cloud governance is discussed in detail in further chapters.

MANAGING THE CLOUD APPLICATION

Business companies are increasingly looking to move or build their corporate applications on cloud platforms to improve agility or to meet dynamic requirements that exist in the globalization of businesses and responsiveness to market demands. But this shift or moving the applications to the cloud environment brings new complexities. Applications become more composite and complex, which requires leveraging not only capabilities like storage and database offered by the cloud providers but also third-party SaaS capabilities like e -mail and messaging. So, understanding the avail- ability of an application requires inspecting the infrastructure, the services it consumes, and the upkeep of the application. The composite nature of cloud applications requires visibility into all the services to determine the overall availability and uptime.

Cloud application management is to address these issues and propose solutions to make it possible to have insight into the application that runs in the cloud, as well as implement or enforce enterprise policies like governance and auditing and environment management while the application is deployed in the cloud. These cloud-based monitoring and management services can collect a multitude of events, analyze them, and identify critical information that requires additional remedial actions like adjusting capacity or provisioning new services. Additionally, application management has to be supported with tools and processes required for managing other environments that might coexist, enabling efficient operations.

MIGRATING APPLICATION TO CLOUD

Cloud migration encompasses moving one or more enterprise applications and their IT environments from the traditional hosting type to the cloud environment, either public, private, or hybrid. Cloud migration presents an opportunity to significantly reduce costs incurred on applications. This activity comprises, of different phases like evaluation, migration strategy, prototyping, provisioning, and testing.

PHASES OF CLOUD MIGRATION

Evaluation: Evaluation is carried out for all the components like current infrastructure and application architecture, environment in terms of compute, storage, monitoring, and management, SLAs, operational processes, financial considerations, risk, security, compliance, and licensing needs are identified to build a business case for moving to the cloud.

Migration strategy: Based on the evaluation, a migration strategy is drawn—a hot plug strategy is used where the applications and their data and interface dependencies are isolated and these applications can be operationalized all at once. A fusion strategy is used where the applications can be partially migrated; but for a portion of it, there are dependencies based on existing licenses, specialized server requirements like mainframes, or extensive interconnections with other applications.

Prototyping: Migration activity is preceded by a prototyping activity to validate and ensure that a small portion of the applications are tested on the cloud environment with test data setup.

Provisioning: Pre-migration optimizations identified are implemented. Cloud servers are provisioned for all the identified environments, necessary platform software and applications are deployed, configurations are tuned to match the new environment sizing, and databases and files are replicated. All internal and external integration points are properly configured. Web services, batch jobs, and operation and management software are set up in the new environments.

Testing: Post migration tests are conducted to ensure that migration has been successful. Performance and load testing, failure and recovery testing, and scale- out testing are conducted against the expected traffic load and resource utilization levels.

APPROACHES FOR CLOUD MIGRATION

The following are the four broad approaches for cloud migration that have been adopted effectively by vendors:

1) ***Migrate existing applications*****:** Rebuild or re-architect some or all the applications, taking advantage of some of the virtualization technologies around to accelerate the work.

But, it requires top engineers to develop new functionality. This can be achieved over the course of severalreleases with the timing determined by customer demand.

2) ***Start from scratch*****:** Rather than cannibalize sales, confuse customers with choice, and tie up engineers trying to rebuild existing application, it may be easier to start again. Many of the R&D decisions will be different now, and with some of the more sophisticated development environments, one can achieve more even with a small focused working team.

3) ***Separate company*****:** One may want to create a whole new company with separate brand, management, R&D, and sales. The investment and internet protocol (IP) may come from the existing company, but many of the conflicts disappear once a new *born in the cloud* company is established. The separate company may even be a subsidiary of the existing company. What is important is that the new company can act, operate, and behave like a cloud-based start-up.

4) ***Buy an existing cloud vendor*****:** For a large established vendor, buying a cloud- based competitor achieves two things. Firstly, it removes a competitor, and secondly, it enables the vendor to hit the ground running in the cloud space. The risk of course is that the innovation, drive, and operational approach of the cloud- based company are destroyed as it is merged into the larger acquirer.

UNIT IV

CLOUD SERVICE MODELS

Cloud Service Models: Infrastructure as a Service, Characteristics of IaaS. Suitability of IaaS, Pros and Cons of IaaS, Summary of IaaS Providers, Platform as a Service, Characteristics of PaaS, Suitability of PaaS, Pros and Cons of PaaS, Summary of PaaS Providers, Software as a Service, Characteristics of SaaS, Suitability of SaaS, Pros and Cons of SaaS, Summary of SaaS Providers, Other Cloud Service Models.

INTRODUCTION

Cloud computing is a model that enables the end users to access the shared pool of resources such ascomputer, network, storage, database, and application as an on- demand service without the need to buy or own it. The services are provided and managed by the service provider, reducing the management effort from the end user side.

The essential characteristics of the cloud include on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The National Institute of Standards and Technology (NIST) define three basic service models, namely, IaaS, PaaS, and SaaS, as shown in Figure 4.1.

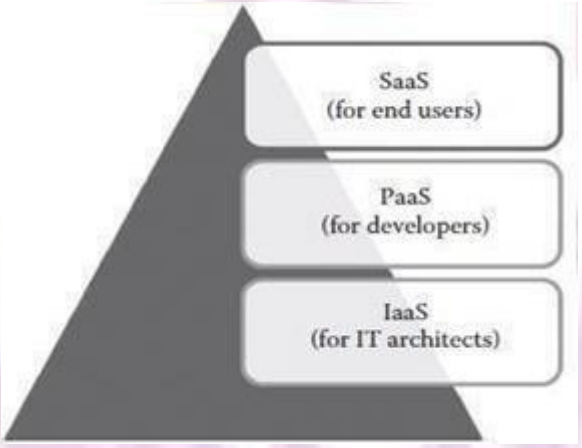


FIGURE 4.1 Basic cloud service models.

The NIST definition of the three basic service models is given as follows:

***IaaS*:** The ability given to the infrastructure architects to deploy or run any software on the computingresources provided by the service provider.

Here, the underlying infrastructures such as computer, network, and storage are managed by the service provider. Thus, the infrastructure architects are exempted from maintaining the data center or underlying infrastructure. The end users are responsible for managing applications that are running on top of the service provider cloud infrastructure. Generally, the IaaS services are provided from the service provider cloud data center.

The end users can access the services from their devices through web command line interface (CLI) or application programming interfaces (APIs) provided by the service providers. Some of the popular IaaS providers include Amazon Web Services (AWS), Google Compute Engine, OpenStack, and Eucalyptus.

***PaaS*:** The ability given to developers to develop and deploy an application on the development platform provided by the service provider. Thus, the developers are exempted from managing the

development platform and underlying infrastructure. Here, the developers are responsible for managing the deployed application and configuring the development environment.

Generally, PaaS services are provided by the service provider on an on-premise or dedicated or hosted cloud infrastructure. The developers can access the development platform over the Internet through web CLI, web user interface (UI), and integrated development environments (IDEs). Some of the popular PaaS providers include Google App Engine, Force.com, Red Hat OpenShift, Heroku, and Engine Yard.

***SaaS*:** The ability given to the end users to access an application over the Internet that is hosted and managed by the service provider. Thus, the end users are exempted from managing or controlling an application, the development platform, and the underlying infra-structure. Generally, SaaS services are hosted in service provider– managed or service provider–hosted cloud infrastructure. The end users can access the services from any thin clients or web browsers. Some of the popular SaaS providers include Saleforce.com, Google Apps, and Microsoft office 365.

The different cloud service models target different audiences. For example, the IaaS model targets the information technology (IT) architects, PaaS targets the developers, and SaaS targets the end users. Based on the services subscribed, the responsibility of the targeted audience may vary as shown in Figure 4.2.

In IaaS, the end users are responsible for maintaining the development platform and the application running on top of the underlying infrastructure. The IaaS providers are responsible for maintaining the underlying hardware as shown in Figure 4.2.

In PaaS, the end users are responsible for managing the application that they have developed. The underlying infrastructure will be maintained by the infrastructure provider as shown in Figure 4.2b. In SaaS, the end user is free from maintaining the infrastructure, development platform, and application that they are using. All the maintenance will be carried out by the SaaS providers as shown Figure 4.2c.

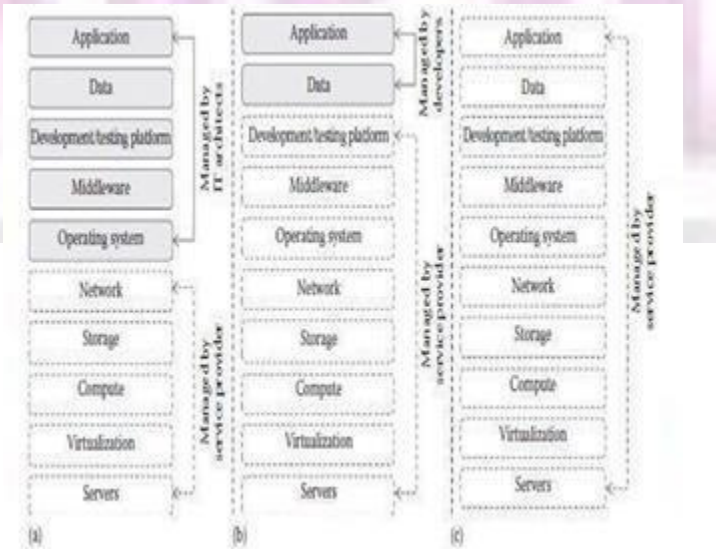


FIGURE 4.2

User and service provider responsibilities of cloud service models: (a) IaaS, (b) PaaS, and (c) SaaS.

The different service models of cloud computing can be deployed and delivered through any one of the cloud deployment models. The NIST defines four different types of cloud deployment models, namely, public cloud, private cloud, community cloud, and hybrid cloud.

The hybrid cloud is anycombination of the public, private, and community clouds. The service delivery of cloud servicesthrough different deployment models is shown in the figure.

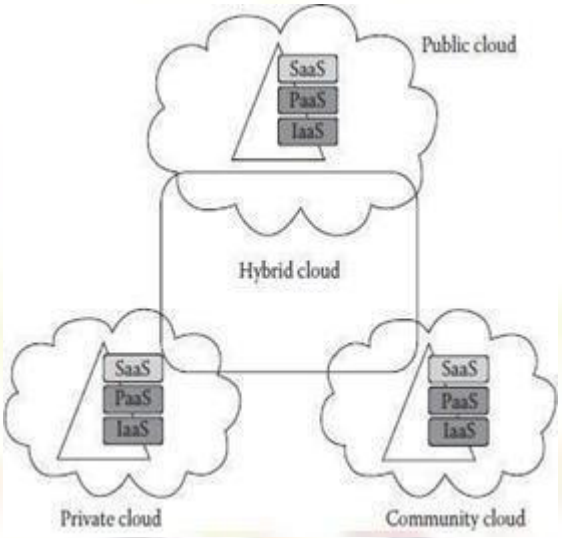


FIGURE 4.3 Deployment and delivery of different cloud service delivery models.

This chapter discusses about the characteristics, suitability, and pros and cons of different cloud service models. Additionally, this chapter gives the summary of popular IaaS, PaaS, and SaaS providers.

INFRASTRUCTURE AS A SERVICE

IaaS changes the way that the compute, storage, and networking resources are consumed. In traditional data centers, the computing power is consumed by having physical access to the infrastructure. IaaS changes the computing from a physical infrastructure to a virtual infrastructure. IaaS provides virtual computing, storage, and network resources by abstracting the physical resources. Technology *virtualization* is used to provide the virtual resources.

All the virtual resources are given to the virtual machines (VMs) that are configured by the service provider. The end users or IT architects will use the infrastructure resources in the form of VMs as shown in Figure 4.4. maintained by the service providers. The physical infrastructure can be maintained by the service providers themselves. Thus, it eliminates or hides the complexity of maintaining the physical infrastructure from the IT architects. A typical IaaS provider may provide the flowing services as shown in Figure 4.5:

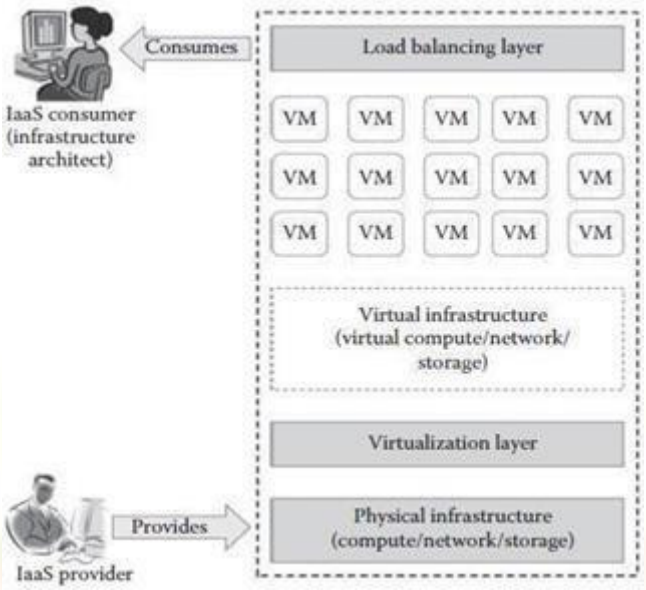


FIGURE 4.4 Overview of IaaS.

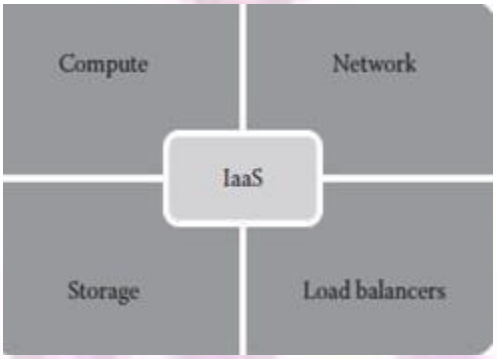


FIGURE 4.5
Services provided by IaaS providers.

***Compute*:** Computing as a Service includes virtual central processing units (CPUs) and virtual main memory for the VMs that are provisioned to the end users.

***Storage*:** STaaS provides back-end storage for the VM images. Some of the IaaS providers also provide the back end for storing files.

***Network*:** Network as a Service (NaaS) provides virtual networking components such as virtual router, switch, and bridge for the VMs.

***Load balancers*:** Load Balancing as a Service may provide load balancing capability at the infrastructure layer.

CHARACTERISTICS OF IAAS

IaaS providers offer virtual computing resources to the consumers on a pay-as- you- go basis. IaaS contains the characteristics of cloud computing such as on- demand self- service, broad network access, resource pooling, rapid elasticity, and measured service. Apart from all these, IaaS has its ownunique characteristics as follows:

***Web access to the resources*:** The IaaS model enables the IT users to access infrastructure resources over the Internet. When accessing a huge computing power, the IT user need not get physical access to the servers.

Through any web browsers or management console, the users can access the required infrastructure.

***Centralized management*:** Even though the physical resources are distributed, the management will be from a single place. The resources distributed across different parts can be controlled from any management console. This ensures effective resource management and effective resource utilization.

***Elasticity and dynamic scaling*:** IaaS provides elastic services where the usage of resources can be increased or decreased according to the requirements. The infrastructure need depends on the load on the application. According to the load, IaaS services can provide the resources. The load on any application is dynamic and IaaS services are capable of proving the required services dynamically.

***Shared infrastructure*:** IaaS follows a one-to-many delivery model and allows multiple IT users to share the same physical infrastructure.

The different IT users will be given different VMs. IaaS ensures high resource utilization.

***Preconfigured VMs*:** IaaS providers offer preconfigured VMs with operating systems (OSs), network configuration, etc. The IT users can select any kind of VMs of their choice. The IT users arefree to configure VMs from scratch. The users can directly start using the VMs as soon as theysubscribed to the services.

***Metered services*:** IaaS allows the IT users to rent the computing resources insteadof buying it. The services consumed by the IT user will be measured, and the users will be charged by the IaaS providers based on the amount of usage.

SUITABILITY OF IAAS

IaaS reduces the total cost of ownership (TCO) and increases the return on investment (ROI) for start-up companies that cannot invest more in buying infrastructure.

IaaS can be used in the following situations:

***Unpredictable spikes in usage*:** When there is a significant spike in usage of computing resources, IaaS is the best option for IT industries. When demand is very volatile, we cannot predict the spikes and troughs in terms of demand of the infrastructure. In this situation, we cannot add or remove infrastructure immediately according to the demand in a traditional infrastructure. If there is an unpredictable demand of infrastructure, then it is recommended to use IaaS services.

***Limited capital investment*:** New start-up companies cannot invest more on buying infrastructure for their business needs. And so by using IaaS, start-up companies can reduce the capital investment on hardware. IaaS is the suitable option for start- up companies with less capital investment on hardware.

***Infrastructure on demand*:** Some organizations may require large infrastructure for a short period of time. For this purpose, an organization cannot afford to buy more on-premise resources. Instead, they can rent the required infrastructure for a specific period of time. IaaS best suits the organizations that look for infrastructure on demand or for a short time period. IaaS helps start-up companies limit its capital expenditure. While it is widely used bystart-upcompanies.

In following situations, IT users should avoid using the IaaS:

***When regulatory compliance does not allow off- premise hosting*:** For some companies, its regulation may not allow the application and data to be hosted on third-party off-premise infrastructure.

***When usage is minimal*:** When the usage is minimal and the available on-premise infrastructure itself is capable of satisfying their needs.

***When better performance is required*:** Since the IaaS services are accessed through the Internet, sometimes the performance might be not as expected due to network latency.

***When there is a need for more control on physical infrastructure*:** Some organizations might require physical control over the underlying- infrastructure. As the IaaS services are abstracted as virtual resources, it is not possible to have more control on underlying physical infrastructure.

PROS AND CONS OF IAAS

Being one of the important service models of cloud computing, IaaS pro-vides lot of benefits to the ITusers. The following are the benefits provided by IaaS:

***Pay-as-you-use model*:** The IaaS services are provided to the customers on a pay-per-use basis. Thisensures that the customers are required to pay for what they have used. This model eliminates the unnecessary spending on buying hardware. *Reduced TCO*: Since IaaS providers allow the IT users to rent the computing resources, they need not buy physical hardware for running their business. The IT users can rent the IT infrastructure rather than buy it byspending large amount.
IaaS reduces the need for buying hardware resources and thus reduces the TCO.

***Elastic resources*:** IaaS provides resources based on the current needs. IT users can scale up or scale down the resources whenever they want. This dynamic scaling is done automatically using some load balancers. This load balancer transfers the additional resource request to the new server and improves application efficiency. *Better resource utilization*: Resource utilization is the most important criteria to succeed in the IT business. The purchased infrastructure should be utilized properly to increase the ROI. IaaS ensures better resource utilization and provides high ROI for IaaS providers.

***Supports Green IT*:** In traditional IT infrastructure, dedicated servers are used for different business needs. Since many servers are used, the power consumption will be high. This does not result in Green IT. In IaaS, the need of buying dedicated servers is eliminated as single infrastructure is shared between multiple customers, thus reducing the number of servers to be purchased and hence the power consumption that results in Green IT.

Even though IaaS provides cost-related benefits to small- scale industries, it lacks in providing security to the data. The following are the drawbacks of IaaS: *Security issues*: Since IaaS uses virtualization as the enabling technology, hypervisors play an important role. There are many attacks that target the hypervisors to compromise it. If hypervisors get compromised, then any VMs canbe attacked easily. Most of the IaaS providers are not able to provide 100% security to the VMs and the data stored on the VMs.

***Interoperability issues*:** There are no common standards followed among the different IaaS providers. It is very difficult to migrate any VM from one IaaS provider to the other. Sometimes, the customers might face the vendor lock- in problem.

***Performance issues*:** IaaS is nothing but the consolidation of available resources from the distributed cloud servers. Here, all the distributed servers are connected over the network.

Latency of the network plays an important role in deciding the performance. Because of latency issues, sometimes the VM contains issues with its performance.
Public IaaS consumers need not consider the host OS as it is maintained by the service provider. In managing the private cloud, the users should see the supported host OS. However, most of the private IaaS supports popular guest OS, fully depending on the hypervisor that the IaaS providers are supporting.

5.5 PLATFORM AS A SERVICE

PaaS changes the way that the software is developed and deployed. In traditional application development, the application will be developed locally and will be hosted in the central location. In stand-alone application development, the applications will be developed and delivered as executables. Most of the applications developed by traditional development platforms result in Licensing based software, whereas PaaS changes the application development from local machine toonline. PaaS providers provide the development PaaS from the data center. The developers canconsume the services over the Internet as shown in Figure 4.6.

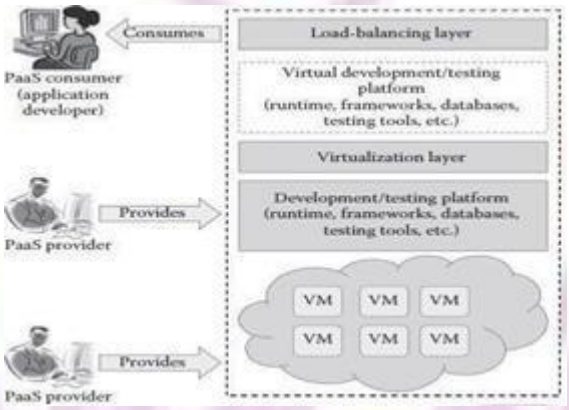


FIGURE 4.6 Overview of PaaS.

PaaS allows the developers to develop their application online and also allows them to deploy immediately on the same platform.
PaaS consumers or developers can consume language runtimes, application - frameworks, databases, message queues, testing tools, and deployment tools as a ser- vice over the Internet. Thus, it reduces the complexity of buying and - maintaining different tools for developing an application. Typical PaaS providers may provide programming languages, application frameworks, databases, and testing tools as shown in Figure 5.7. Some of the PaaS providers also provide build tools, deployment tools, and software load balancers as a service:
***Programming languages*:** PaaS providers provide a wide variety of programming languages for the developers to develop applications. Some of the popular programming languages provided by PaaS vendors are Java, Perl, PHP, Python, Ruby, Scala, Clojure, and Go.
***Application frameworks*:** PaaS vendors provide application frame-works that simplify the application development. Some of the popular application development frameworks provided by a PaaS provider include Node.js, Rails, Drupal, Joomla, WordPress, Django, EE6, Spring, Play, Sinatra, Rack, and Zend. *Database*: Since every application needs to communicate with the

databases, it becomes a must-have tool for every application. PaaS providers are providing databases also with their PaaS plat-forms.

The popular databases provided by the popular PaaS vendors are ClearDB, PostgreSQL, Cloudant, Membase, MongoDB, and Redis.

***Other tools*:** PaaS providers provide all the tools that are required to develop, test, and deployan application.

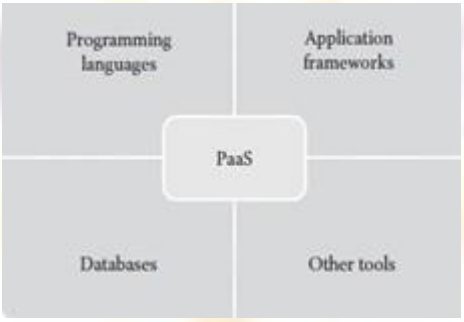


FIGURE 4.7
Services provided by PaaS providers.

CHARACTERISTICS OF PAAS

PaaS development platforms are different from the traditional application development platforms. The following are the essential characteristics that make PaaS unique from traditional development platforms:

***All in one*:** Most of the PaaS providers offer services to develop, test, deploy, host, and maintain applications in the same IDE. Additionally, many service providers provide all the programming languages, frameworks, databases, and other development-related services that make developers choose from a wide variety of development platforms.

***Web access to the development platform*:** A typical development plat-form uses any IDEs for developing applications. Typically, the IDE will be installed in the developer's machines. But, PaaS provides web access to the development platform. Using web UI, any developer can get access to thedevelopment platform. The web- based UI helps the developers create, modify, test, and deploy different applications on the same platform.

***Offline access*:** A developer may not be able to connect to the Internet for a whole day to access the PaaS services. When there is no Internet connectivity, the developers should be allowed to work offline. To enable offline development, some of the PaaS providers allow the developer to synchronize their local IDE with the PaaS services. The developers can develop an application locally and deploy it online whenever they are connected to the Internet.

***Built-in scalability*:** Scalability is an important requirement for the new-generation web or SaaS applications. It is very difficult to enable the dynamic scalability for any application developed using traditional development platforms. But, PaaS services provide built-in scalability to an application that is developed using any particular PaaS. This ensures that the application is capable of handling varying loads efficiently.

***Collaborative platform*:** Nowadays, the development team consists of developers who are working from different places. There is a need for a common platform where the developers can collaboratively work together on the same project.

Most of the PaaS services pro-vide support collaborative development. To enable collaboration

among developers, most of the PaaS providers provide tools for project planning.

***Diverse client tools*:** To make the development easier, PaaS providers provide a wide variety of client tools to help the developer. The client tools include CLI, web CLI, web UI, REST API, and IDE. The developers can choose any tools of their choice. These client tools are also capable of handling billing and subscription management.

SUITABILITY OF PAAS

Most of the start-up SaaS development companies and independent software vendors (ISVs) widely use PaaS in developing an application. PaaS technology is getting attention from other traditional software development companies also. PaaSis a suitable option for the following situations:

***Collaborative development*:** To increase the time to market and development efficiency, there is a need for a common place where the development team and other stakeholders of the application can collaborate with each other. Since PaaS services provide a collaborative development environment, it is a suitable option for applications that need collaboration among developers and other third parties to carryout the development process.

***Automated testing and deployment*:** Automated testing and building of an application are very usefulwhile developing applications at a very short time frame. The automated testing tools reduce the time spent in manual testing tools. Most of the PaaS services offer automated testing and deployment capabilities. The development team needs to concentrate more on development rather than testing and deployment. Thus, PaaS services are the best option where there is a need for automated testing and deployment ofthe applications.

***Time to market*:** The PaaS services follow the iterative and incremental development methodologies that ensure that the application is in the market as per the time frame given. For example, the PaaS services are the best option for application development that uses agile development methodologies. If the software vendor wants their application to be in the market as soon as possible, then the PaaS services are the best option for the development.

PaaS is used widely to accelerate the application development process to ensure the time to market. Most of the start-up companies and ISVs started migrating to the PaaS services. Even though it is used widely, there are some situations where PaaS may not be the best option:

***Frequent application migration*:** The major problem with PaaS services are vendor lock-in. Since there are no common standards followed among PaaS providers, it is very difficult to migrate the application from one PaaS provider to the other.

***Customization at the infrastructure level*:** PaaS is an abstracted service, and the PaaS users do not have full control over the underlying infrastructure. There are some application development platforms that need some configuration or customization of underlying infrastructure.

In these situations, it is not possible to customize the underlying infrastructure with PaaS. If the application development platform needs any configuration at the hardware level, it is not recommended to go for PaaS.

***Flexibility at the platform level*:** PaaS provides template- based applications where all the different programming languages, databases, and message queues are predefined. It is an advantage if the application is a generic application.

***Integration with on-premise application*:** A company might have used PaaS services for some set of applications. For some set of applications, they might have used on- premise platforms.

Since many PaaS services use their own proprietary technologies to define the application stack, it may not match with the on-premise application stack. This makes the integration of application hosted in on-premise plat-form and PaaS platform a difficult job.

PROS AND CONS OF PAAS

The main advantage of using PaaS is that it hides the complexity of maintaining the platform and underlying infrastructure. This allows the developers to work more on implementing the important functionalities of the application. Apart from this, the PaaS has the following benefits:

***Quick development and deployment*:** PaaS provides all the required development and testing tools todevelop, test, and deploy the soft-ware in one place. Most of the PaaS services automate the testing and deployment process as soon as the developer completes the development. This speeds up application development and deployment than traditional development platforms.

***Reduces TCO*:** The developers need not buy licensed development and testing tools if PaaS services are selected. Most of the traditional development platforms requires high- end infrastructure for its working, which increases the TCO of the application development company. But, PaaS allows the developers to rent the software, development platforms, and testing tools to develop, build, and deploy the application. PaaS does not require high-end infrastructure also to develop the application, thus reducing the TCO ofthe development company.

***Supports agile software development*:** Now a days, most of the new-generation applications are developed using agile methodologies. Many ISVs and SaaS development companies started adopting agile methodologies for application development. PaaS services support agile methodologies that the ISVs and other development companies are looking for.

***Different teams can work together*:** The traditional development plat-form does not have extensive support for collaborative development.

PaaS services support developers from different places to work together on the same project. This is possible because of the online common development platform provided by PaaS providers.

***Ease of use*:** The traditional development platform uses any one of CLI- or IDE- based interfaces for development. Some developers may not be familiar with the interfaces provided by the application development platform. This makes the development job a little bit difficult. But, PaaS provides a wide variety of client tools such as CLI, web CLI, web UI, APIs, and IDEs. The developers are free to choose any client tools of their choice. Especially, the web UI–based PaaS services increase the usability of the development platform for all types of developers.

***Less maintenance overhead*:** In on-premise applications, the development company or software vendor is responsible for maintaining the underlying hardware. They need to recruit skilled administrators to maintain the servers. This overhead is eliminated by the PaaS services as the underlying infrastructure is maintained by the infra-structure providers. This gives freedom to developers to work on the application development.

Produces scalable applications: Most of the applications developed using PaaS services are web application or SaaS application. These applications require better scalability on the extra load. For handling extra load, the software vendors need tomaintain an additional server. It is very difficult for a new start-up company to provide extra servers based on the additional load.

But, PaaS services are providing built-in scalability to the application that is developed using the PaaS platform.

PaaS provides a lot of benefits to developers when compared to the traditional development environment. On the other hand, it contains drawbacks, which are described in the following:

***Vendor lock-in*:** The major drawback with PaaS providers are vendor lock-in. The main reason forvendor lock-in is lack of standards. There are no common standards followed among the differentPaaS providers. The other reason for vendor lock-in is proprietary technologies used by PaaSproviders. Most of the PaaS vendors use the proprietary technologies that are not compatible withthe other PaaS providers. The vendor lock-in problem of PaaS services does not allow theapplications to be migrated from one PaaS provider to the other.

***Security issues*:** Like in the other cloud services, security is one of the major issues in PaaS services. Since data are stored in off--premise third-party servers, many developers are afraid to go for PaaS services. Of course, many PaaS providers provide mechanisms to protect the user data, and it is not sufficient to feel the safety of on-premise deployment. When selecting the PaaS provider, the developer should review the regulatory, compliance, and security policies of the PaaS provider with their own security requirements. If not properly reviewed, the developers or users are at the risk of data securitybreach.

***Less flexibility*:** PaaS providers do not give much freedom for the developers to define their own application stack. Most of the PaaS providers provide many programming languages, databases, and other development tools. But, it is not extensive and does not satisfy all developer needs.

Only some of the PaaS providers allow developers-to extend the PaaS tools with the custom or new programming languages. Still most of the PaaS providers do not provide flexibility to the developers.

***Depends on Internet connection*:** Since the PaaS services are delivered over the Internet, the developers should depend on Internet connectivity for developing the application. Even though some of the providers allow offline access, most of the PaaS providers do not allow offline access. With slow Internet connection, the usability and efficiency of the PaaS platform do not satisfy the developer requirements.

SOFTWARE AS A SERVICE

FIGURE 4.8 Services provided by SaaS Providers.

SaaS provider may provide business services, social networks, document management, and mail services as shown in Figure 4.8:

***Business services*:** Most of the SaaS providers started providing a variety of business services that attract start-up companies. The business SaaS services include ERP, CRM, billing, sales, and humanresources.

***Social networks*:** Since social networking sites are extensively used by the general public, many social networking service providers adopted SaaS for their sustainability. Since the number of users of the social networking sites is increasing exponentially, cloud computing is the perfect match for handling the variable load. *Document management*: Since most of the enterprises extensively use electronic documents, most of the SaaS providers started providing services that are used to create, manage, and track electronic documents.

***Mail services*:** E-mail services are currently used by many people. The future growth in e-mail usage is unpredictable. To handle the unpredictable number of users and the load on e-mail services, most of the e- mail providers started offering their services as SaaS services.

CHARACTERISTICS OF SAAS

SaaS services are different and give more benefits to end users than the traditional software. The following are the essential characteristics of SaaS ser-vices that make it unique from traditional software:

***One to many*:** SaaS services are delivered as a one-to- many model where a single instance of the application can be shared by multiple tenants or customers.

***Web access*:** SaaS services provide web access to the software. It allows the end user to access the application from any location if the device is connected to the Internet.

***Centralized management*:** Since SaaS services are hosted and man-aged from the central location, management of the SaaS application becomes easier. Normally, the SaaS providers will perform the automatic updates that ensure that each tenant is accessing the most recent version of the applicationwithout any user-side updates.

***Multidevice support*:** SaaS services can be accessed from any end user devices such as desktops, laptops, tablets, smart phones, and thin clients.

***Better scalability*:** Since most of the SaaS services leverage PaaS and IaaS for its development and deployment, it ensures a better scalability than the traditional software. The dynamic scaling of underlying cloud resources makes SaaS applications work efficiently even with varying loads.

***High availability*:** SaaS services ensure the 99.99% availability of user data as proper backup and recovery mechanisms are implemented at the back end.

***API integration*:** SaaS services have the capability of integrating with other software or service through standard APIs.

SUITABILITY OF SAAS

SaaS is popular among individuals and start-up companies because of the benefits it provides. Most of the traditional software users are looking for SaaS versions of the software as SaaS has several advantages over traditional applications. SaaS applications are the best option for the following:

***On-demand software*:** The licensing-based software model requires buying full packaged software and increases the spending on buying software. Some of the occasionally used software does not give any ROI. Because of this, many end users are looking for a software that they can use as and when they needed. If the end users are looking for on-demand software rather than the licensing-based full-term software, then the SaaS model is the best option.

***Software for start-up companies*:** When using any traditional software, the end user should buy devices with minimum requirements specified by the software vendor. This increases the investment on buying hardware for start- up companies. Since SaaS services do not require high-end infrastructure for accessing, it is a suitable option for start- up companies that can reduce the initial expenditure on buying high-end hardware.

***Software compatible with multiple devices*:** Some of the applications like word processors or mail services need better accessibility from different- devices. The SaaS applications are adaptable with almost all the devices.

***Software with varying loads*:** We cannot predict the load on popular applications such as social networking sites. The user may connect or disconnect from applications anytime. It is very difficult to handle varying loads with the traditional infrastructure. With the dynamic scaling capabilities, SaaS applications can handle varying loads efficiently without disrupting the normal behavior of the application. Most of the traditional software vendors moved to SaaS business as it is an emerging software delivery model that attracts end users. But still many traditional applications do not have its SaaS versions. This implies that SaaS applications may not be the best option for all types of software. The SaaS delivery model is not the best option for the applications mentioned in the following:

***Real-time applications*:** Since SaaS applications depend on Internet connectivity, it may not work better with low Internet speed. If data are stored far away from theend user, the latency issues may delay the data retrieval timings. Real- time applications require fast processing of data that may not be possible with the SaaS applications because of the dependency on high-speed Internet connectivity and latency issues.

***Applications with confidential data*:** Data security, data governance, and data compliance are always issues with SaaS applications. Since data are stored with third- party service providers, there is no surety that our data will be safe. If the stored confidential data get lost, it will make a serious loss to the organization. It is not recommended to go for SaaS for applications that handle confidential data.

***Better on-premise application*:** Some of the on-premise applications might fulfill all the requirements of the organization. In such situations, migrating to the SaaS model may not be the best option.

PROS AND CONS OF SAAS

SaaS applications are used by a wide range of individuals and start-up industries for its cost-related benefits. Apart from the cost-related benefits, SaaS services provide the following benefits:

***No client-side installation*:** SaaS services do not require client-side installation of the software. The end users can access the services directly from the service provider data center without any installation. There is no need of high- end hardware to consume SaaS services.

It can be accessed from thin clients or any handheld devices, thus reducing the initial expenditure on buying high-end hardware.

***Cost savings*:** Since SaaS services follow the utility-based billing or pay-as-you- go billing, it demands the end users to pay for what they have used. Most of the SaaS providers offer different subscription plans to benefit different customers. Sometimes, the generic SaaS services such as word processors are given for free to the end users.

***Less maintenance*:** SaaS services eliminate the additional overhead of maintaining the software from the client side. For example, in the traditional software, the end user is responsible for performing bulk updates. But in SaaS, the service provider itself maintains the automatic updates, monitoring, and other maintenance activities of the applications.

***Ease of access*:** SaaS services can be accessed from any devices if it is connected to the Internet. Accessibility of SaaS services is not restricted to any particular devices. It is adaptable to all the devices as it uses the responsive web UI. *Dynamic scaling*: SaaS services are popularly known for elastic dynamic scaling. It is very difficult for on-premise software to pro-vide dynamic scaling capability as it requires additional hardware. Since the SaaS services leverage elastic resources provided by cloud computing, it can handle any type of varying loads without disrupting the normal behavior of the application.

***Disaster recovery*:** With proper backup and recovery mechanisms, replicas are maintained for every SaaS services. The replicas are distributed across many servers. If any server fails, the end user can access the SaaS from other servers. It eliminates the problem of single point of failure. It also ensures the high availability of the application.

***Multitenancy*:** Multitenancy is the ability given to the end users to share a single instance of the application. Multitenancy increases resource utilization from the service provider side.

The following are the major problems with SaaS services:

***Security*:** Security is the major concern in migrating to SaaS application. Since the SaaS applicationis shared between many end users, there is a possibility of data leakage.

Here, the data are stored in the service provider data center. We cannot simply trust some third-partyservice provider to store our company-sensitive and confidential data. The end user should be careful while selecting the SaaS provider to avoid unnecessary data loss.

***Connectivity requirements*:** SaaS applications require Internet connectivity for accessing it. Sometimes, the end user's Internet connectivity might be very slow. In such situations, the user cannot access the services with ease.

Summary of Popular SaaS Providers

Provider	Services Provided
Salseforce.com	On-demand CRM solutions
Google Apps	Gmail, Google Calendar, Talk, Docs, and Sites
Microsoft Office 356	Online office suite, software, plus services
NetSuite	ERP, accounting, order management, inventory, CRM, professional services automation (PSA), and e-commerce applications
Concur	Integrated travel and expense management solutions
GoToMeeting	Online meeting, desktop sharing, and video-conferencing software
Constant Contact	E-mail marketing, social-media marketing, online survey, event marketing, digital storefronts, and local deals tools
Workday, Inc.	Human capital management, payroll, and financial management
Oracle CRM	CRM applications
Intacct	Financial management and accounting software solutions

Table 4.3

The dependency on high-speed Internet connection is a major problem in SaaS applications.

***Loss of control*:** Since the data are stored in a third-party and off-premise location, the end user does not have any control over the data. The degree of control over the SaaS application and data is lesser than the on-premise application.

SUMMARY OF SAAS PROVIDERS

There are many SaaS providers who provide SaaS services such as ERP, CRM, billing, document management, and mail services. Table 4.3 gives a summary of popular SaaS vendors in the market.

OTHER CLOUD SERVICE MODELS

The basic cloud services such as IaaS, PaaS, and SaaS are widely used by many individual and start-up companies. Now, cloud computing becomes the dominant technology that drives the IT world. Because of the extensive use of basic cloud services, the end users realize the importance and benefits of specific services such as network, storage, and database. The basic cloud service models are the unified models that contain multiple services in it. Now, the end users' expectation changed, and they are expecting the individual services to be offered by service providers.

This makes most of the service providers to think about the separate services that meet end user requirements. Many service providers already started offering separate ser-vices such as network, desktop, database, and storage on demand as given in the following:

NaaS is an ability given to the end users to access virtual network services that are provided by the service provider. Like other cloud service models, NaaS is also a business model for delivering virtual network services over the Internet on a pay-per-use basis.

In on-premise data center, the IT industries spent a lot of money to buy network hardware to manage in-house networks. But, cloud computing changes networking services into a utility-based service. NaaS allows network architects to create virtual networks, virtual network- interface cards (NICs), virtual routers, virtual switches, and other networking components.

Additionally, it allows the network architect to deploy custom routing protocols and enables the designof efficient in-network services, such as data aggregation, stream processing, and caching. Some of the popular services provided by NaaS include virtual private network (VPN), bandwidth on demand (BoD), and mobile network virtualization.

Desktop as a Service (*DEaaS*) is an ability given to the end users to use desktop virtualization without buying and managing their own infrastructure. DEaaS is a pay-per-use cloud service delivery model in which the service provider manages the back-end responsibilities of data storage, backup, security, and upgrades. The end users are responsible for managing their own desktop images, applications, and security.

Accessing the virtual desktop provided by the DEaaS provider is device, location, and network independent. DEaaS ser-vices are simple to deploy, are highly secure, and produce better experience on almost all devices.

STaaS is an ability given to the end users to store the data on the storage services provided by the service provider. STaaS allows the end users to access the files at any time from any place. The STaaS provider provides the virtual storage that is abstracted from the physical storage of any cloud data center. STaaS is also a cloud business model that is delivered as a utility. Here, the customers can rent the storage from the STaaS provider. STaaS is commonly used as a backup storage for efficient disaster recovery.

DBaaS is an ability given to the end users to access the database ser-vice without the need to install and maintain it. The service provider is responsible for installing and maintaining the databases.
The end users can directly access the services and can pay according to their usage. DBaaS automates the database administration process. The end users can access the database services through any API or web UIs provided by the service provider. The DBaaS eases the database administration process. Popular examples of DBaaS include SimpleDB, DynamoDB, MongoDB as a Service, GAE datas- tore, and ScaleDB.

Data as a Service (*DaaS*) is an ability given to the end users to access the data that are provided by the service provider over the Internet. DaaS provides data on demand. The data may include text, images, sounds, and videos. DaaS is closely related to other cloud service models such as SaaS and STaaS.
DaaS can be easily integrated with SaaS or STaaS for providing the composite service. DaaS is highly used in geography data services and financial data services. The advantages of DaaS include agility, cost effectiveness, and data quality.
SECaaS is an ability given to the end user to access the security service provided by the service provider on a pay- per-use basis. In SECaaS, the service provider integrates their security services to benefit the end users. Generally, the SECaaS includes authentication, antivirus, antimalware/spyware, intrusion detection, and security event management. The security services provided by the SECaaS providers are typically used for securing the on-premise or in-house infrastructure and applications. Some of the SECaaS providers include Cisco, McAfee, Panda Software, Symantec, Trend Micro, and VeriSign.
IDaaS is an ability given to the end users to access the authentication infrastructure that is managed and provided by the third-party service provider. The end user of IDaaS is typically an organization or enterprise. Using IDaaS services, any organization can easily manage their employees' identity without any additional overhead. Generally, IDaaS includes directory services, federated services, registration, authentication services, risk and event monitoring, single sign-on services, and identity and profile management.
The different new service models discussed in this section emerged after the introduction of cloud computing. This field still evolves and introduces new service models based on the end user's needs. Many researchers from industry and academia already started introducing their innovative idea to take cloud computing to the next level. Apart from the service models discussed in this chapter, cloud computing researchers are thinking to add more service models.
Now, cloud computing moves to the scenario where everything can be given as a service.

This can be termed as Everything as a Service (XaaS). In the future, we expect many new service models to achieve the goal of XaaS. XaaS may include Backup as a Service (BaaS), Communication as a Service(CaaS), Hadoop as a Service (HaaS), Disaster Recovery as a Service (DRaaS), Testing as a Service (TaaS), Firewall as a Service (FWaaS), Virtual Private Network as a Service (VPNaaS), Load Balancers as a Service (LBaaS), Message Queue as a Service (MQaaS), and Monitoring as a Service(MaaS).

UNIT V

Cloud Service Providers: EMC, EMC IT, Captiva Cloud Toolkit, Google, Cloud Platform, Cloud Storage, Google Cloud Connect, Google Cloud Print, Google App Engine, Amazon Web Services, Amazon Elastic Compute Cloud, Amazon Simple Storage Service, Amazon Simple Queue, service, Microsoft, Windows Azure, Microsoft Assessment and Planning Toolkit, SharePoint, IBM, Cloud Models, IBM Smart Cloud, SAP Labs, SAP HANA Cloud Platform, Virtualization Services Provided by SAP, Sales force, Sales Cloud, Service Cloud: Knowledge as a Service, Rack space, VMware, Manjra soft, Aneka Platform.

INTRODUCTION

Cloud computing is one of the most popular buzzwords used these days. It is the upcoming technology provisioning resources to the consumers in the form of different services like software, infrastructure, platform, and security. Services are made available to users on demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's own on- premise servers. Cloud services are designed to provide easy, scalable access to applications, resources, and services and are fully managed by a cloud service provider. A cloud service can dynamically scale to meet the needs of its users, and because the service provider sup-plies the hardware and software necessary for the service, there is no need for a company to provision or deploy its own resources or allocate information technology (IT) staff to manage the service. Examples of cloud services include online data storage and backup solutions, web-based e-mail services, hosted office suites and document collaboration services, database processing, and managed technical support services.

Cloud services can be broadly classified into three types: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). With growing technologies, many more services are emerging in this field, such as Security as a Service (SeaaS), Knowledge as a Service, and Data Analytics as a Service.

EMC

Many companies have come forward to adapt the cloud environment and ensure that the users as well as the companies benefit from this. Amazon, Microsoft, Google, Yahoo, EMC, Sales force, Oracle, IBM, and many more companies provide various tools and services in order to give cloud support for their customers EMC.

EMC is one of the leading global enterprises that require dynamic scalability and infrastructure agility to meet changing applications as well as business needs. EMC chose cloud computing as the ideal solution to reduce the complexity and optimize the infrastructure. Offering Information Technology as a Service (ITaaS) reduces the energy consumption through resource sharing.

EMC IT

Virtualization is the main concept behind the success of EMC IT. By virtualizing the infrastructure, allocation of the resources on demand is possible. This also helps to increase efficiency and resource utilization.

EMC IT provides its business process units with IaaS, PaaS, and SaaS. Figure 5.1 gives an overview of the services offered by EMC, which are explained in the following:

IaaS offers EMC business units the ability to provision infrastructure components such as network, storage, computing, and operating systems individually or as integrated services.

PaaS provides the secure application and information frameworks on top of application server, web server, database, unstructured content management, and security components as a service to business units from which to develop solutions.

EMC IT offers database platforms (Oracle Database as a Service, SQL Server as a Service, Green plum as a Service) and application platforms (application development, Enterprise Content Management as a Service, Information Cycle Management as a Service, Security PaaS, and Integration as a Service) for the purpose of development.

SaaS provides applications and tools in a service model for business enablement. EMC IT brought together several existing business solutions under the unified architecture named as Business Intelligence as a Service. It also offers Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) as a Service.

User Interface as a Service (UIaaS) provisions user and interface experience, rather than provisioningthe actual device used.

Cloud services by EMC. (Adapted from EMC IT's journey to the private cloud, applications and the cloud experience, White Paper- EMC.

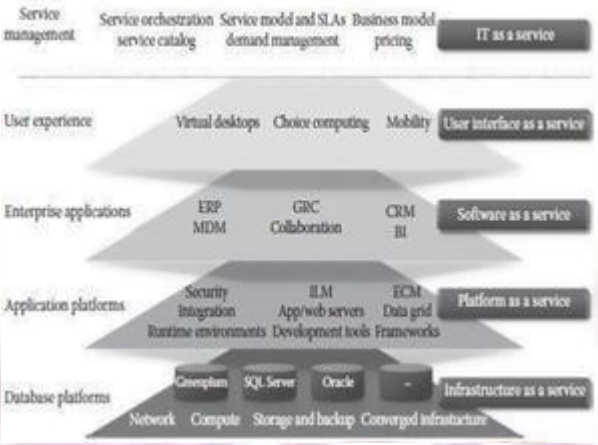


FIGURE 5.1
CAPTIVA CLOUD TOOLKIT

CAPTIVA CLOUD TOOLKIT

EMC offers a tool called *Captiva Cloud Toolkit* to help in the development of softwares. EMC Captiva Cloud Toolkit is a Software Development Kit (SDK) comprised of modules that help web application developers to quickly add scanning and imaging functionality directly to their web-based business applications. It is ideal for document capture vendors, commercial software developers, and enterprises that want to create custom web-based applications that are fully scan enabled, complimenting their business solution offerings.

Using Captiva Cloud Toolkit, developers can quickly create a working scan- enabled web-based business application in as early as 1 week. As a result, time to market is shortened and development, testing, and support costs are greatly reduced. Also, the enterprise's return on investment is quickly achieved, and its ability to compete in an increasingly competitive distributed document capture market is accelerated.

There are a few modules that are commonly used in most of the process development. These are basic modules that import images from various sources like fax, e-mail, or scanner or from any repository. A few of these modules are as follows:

***Scan*:** Scanning is importing activity of documents into Captiva from a scanner. Basically, scanning happens at page level to bring images page by page into Captiva. Scanning is the entry point to Captiva where one can import any kind of document like pdf, tiff, and jpg.

***MDW*:** Multi Directory Watch is another entry point to Captiva. MDW can be pointed to any folder/repository from where Captiva could import documents directly. MDW is very useful if business is getting documents in the form of a soft copy, for example, as an attached file in an e-mail. MDW also acts as a scan module except it does not inter- lock with the scanner.

***IE*:** Image enhancement is a kind of filter or repairing tool for images that are not clear. It enhances the image quality, so it could be processed easily through Captiva. One can configure IE as per business requirement and images being received. The functionalities of IE are deskew, noise removal, etc.

***Index*:** Indexing is a data capturing activity in Captiva through which one can capture key data from various fields. For example, if bank form is being processed, the A/C no. and sort code could be the indexing field. Indexing could be added as per requirement of business. A validation field could be added to avoid unwanted data entry while indexing anydocument.

***Export*:** Export is the exit point of Captiva where images/data are sent to various repositories like file, net, document, or data. The exported data are used for business requirements of various business divisions.

For example, if we are capturing the A/C no. and sort code for a bank application, this could be mapped to anydepartment where it is needed.

***Multi*:** Multi is the last process in Captiva to delete batches that have gone through all modules and exported value successfully. Multi could be configured as per need of business. In the case when it is required to take a backup of batches, this module could be avoided.

Google Cloud Platform

Google is one among the leading cloud providers that offer secure storage of user's data. It provides cloud platform, app engine, cloud print, cloud connect, and many more features that are scalable, reliable, as well as secure. Google offers many of these services for free or at a minimum cost making it user friendly.

Google Cloud Platform enables developers to build, test, and deploy applications on Google's highly scalable and reliable infrastructure. Google has one of the largest and most advanced networks across the globe. Software infra-structures such as MapReduce, BigTable, and Dremel are the innovations for industrial development.

Google Cloud Platform includes virtual machines, block storage, NoSQL datastore, and big data analytics. It provides a range of storage services that allow easy maintenance and quick access of user's data. The cloud platform offers a fully managed platform as well as flexible virtual machines allowing the user to choose as per the requirements. Google also provides easy integration of user's application within the cloud platform.

Applications hosted on the cloud platform can automatically scale up to handle the most demanding workloads and scale down when traffic sub-sides.

The cloud platform is designed to scale like Google's own products, even when there is a huge traffic spike. Managed services such as App Engine or Cloud Datastore provide auto scaling that enables application to grow with the users. The user has to payonly for what he or she uses.

Cloud Storage

Google Cloud Storage is a RESTful online file storage web service for storing and accessing one's data on Google's infrastructure. Representational state transfer (REST) is an architectural style consisting of a coordinated set of architectural constraints applied to components, connectors, and data elements within a distributed system.

The service combines the performance and scalability of Google's cloud with advanced security and sharing capabilities. Google Cloud Storage is safe and secure. Data are protected through redundant storage at multiple physical locations. The following are the few tools for Google Cloud Storage:

Google Developers Console is a web application where one can perform simple storage management tasks on the Google Cloud Storage system.

gsutil is a Python application that lets the user access Google Cloud Storage from the command line.

GOOGLE CLOUD CONNECT

Google Cloud Connect is a feature provided by Google Cloud by integrating cloud and the application programming interface (API) for Microsoft Office. After installing a plug-in for the Microsoft Office suite of programs, one can save files to the cloud. The cloud copy of the file becomes the master document that everyone uses. Google Cloud Connect assigns each file a unique URL that can be shared to let others view the document.

If changes are made to the document, those changes will show up for everyone else viewing it. When multiple people make changes to the same section of a document, Cloud Connect gives chance to theuser to choose which set of changes to keep.

When the user uploads a document to Google Cloud Connect, the service inserts some metadata into the file. Metadata is information about other information. In this case, the metadata identifies the file so that changes will track across all copies. The back end is similar to the Google File System and relies on the Google Docs infrastructure. As the documents sync to the master file, Google Cloud Connect sends the updated data out to all downloaded copies of the document using the metadata to guide updates to the right files.

GOOGLE CLOUD PRINT

Google Cloud Print is a service that extends the printer's function to any device that can connect to the Internet. To use Google Cloud Print, the user needs to have a free Google profile, an app, a program, or a website that incorporates the Google Cloud Print feature, a cloud-ready printer or printer connected to a computer logged on to the Internet when Google Cloud Print is used through an app or website, the print request goes through the Google servers. Google routes the request to the appropriate printer associated with the user's Google account. Assuming the respective printer is on and has an active Internet connection, paper, and ink, the print job should execute on the machine. The printer can be shared with other people for receiving documents through Google Cloud Print.

Because most printers are not cloud ready, most Google Cloud Print users will need to have a computer act as a liaison.

Google Cloud Print is an extension built into the Google Chrome Browser, but it should be enabled explicitly.

Once enabled, the service activates a small piece of code called a connector. The connector's job is to interface between the printer and the outside world. The connector uses the user's computer printer software to send commands to the printer.

If one has a cloud-ready printer, one can connect the printer to the Internet directly without the need for a dedicated computer. The cloud printer has to be registered with Google Cloud Print to take advantage of its capabilities.

Because Google allows app and website developers to incorporate Google Cloud Print into their products as they see fit, there is no standard approach to executing a print job. Google Cloud Print depends on developers incorporating the feature into their products. Not every app or site will have Google Cloud Print built into it, which limits its functionality. Naturally, Google builds the service into its own products, but many people rely on services from multiple sources and may find Google Cloud Print does not have a wide enough adoption to meet all their needs.

Google App Engine

Google App Engine lets the user run web applications on Google's infrastructure. App Engine applications are easy to build, easy to maintain, and easy to scale as traffic and data storage needs grow. With App Engine, there are no servers to maintain: Just upload the application, and it is readyto serve users.

The app can be served from the user's own domain name (such as http:// www.example.com/) using Google Apps. Otherwise, it can be served using a free name on the appspot.com domain.

An application can be shared with the world or limit access to members of an organization. Figure 5.2 shows the different modules in Google App Engine. Integration of cloud computing services with support services and client capabilities is shown in the diagram.

Google App Engine supports apps written in several programming languages. With App Engine's Java runtime environment, one can build one's app using standard Java technologies, including the JVM, the Java servlets, and the Java programming language—or any other language. App Engine also features a Python runtime environment, which includes a fast Python interpreter and the Python standard library. App Engine also features a PHP runtime, with native support for Google Cloud SQL and Google Cloud Storage that works just like using a local MySQL instance and doing local file writes.

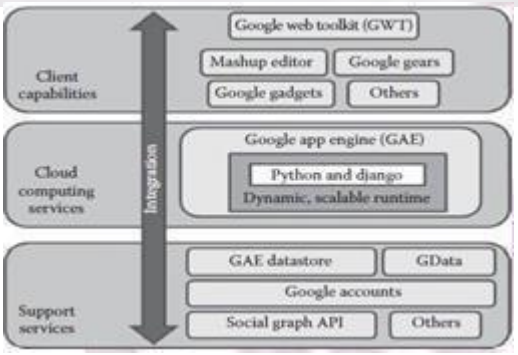


FIGURE 5.2 Google App Engine

GAE provides a Go runtime environment that runs natively compiled Go code. These runtime environments are built to ensure that your application runs quickly, securely, and without interference from other apps on the system.

With App Engine also, the user has to only pay for what he or she uses. There are no setup costs and no recurring fees. The resources used by the application such as storage and bandwidth are measured in gigabyte and billed at competitive rates. One has to control the maximum amount of resources one's app can consume, so it always stays within one's budget.

App Engine costs nothing to get started. All applications can use up to 1 GB of storage and enough CPU and bandwidth to support an efficient app serving around five million page views a month, absolutely free. When billing is enabled for the application, free limits are raised, and one has to only pay for resources one uses above the free levels. The computing capacity (potentially many servers) are much faster and cheaper than building a physical server farm. Amazon's data centers are located at Ashburn, Virginia, Dallas/Fort Worth, Los Angeles, Miami, Newark, New Jersey, Palo, Alto, California, Seattle, St. Louis, Amsterdam, Dublin, Frankfurt, London, Hong Kong, Singapore, Tokyo, etc.

5.9 AMAZON WEB SERVICES

Amazon Web Services (AWS) is a collection of remote computing services (also called web services) that together make up a cloud computing platform, offered over the Internet by Amazon.com. The most central and well known of these services are Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Queue Service (Amazon SQS), and Amazon S3 as shown in Figure 5.3.

Amazon EC2 is a computing service, whereas Amazon SQS and Amazon S3 are support services. The service is advertised as providing a large size of customers.

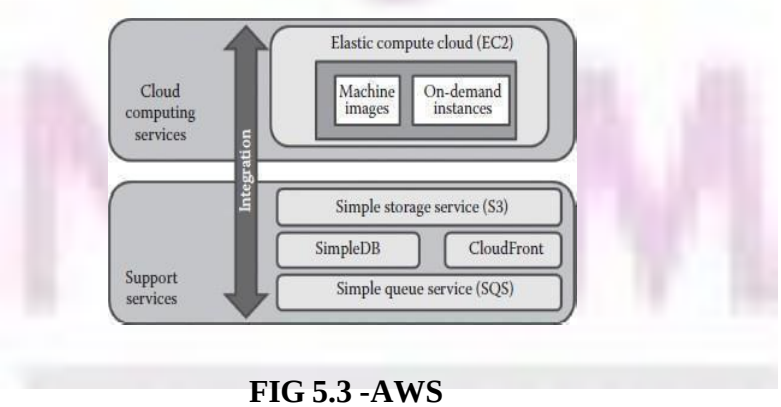


FIG 5.3 -AWS

5.10 AMAZON ELASTIC COMPUTE CLOUD

Amazon EC2 is an IaaS offered by AWS and is the leading provider of IaaS in the current market. Powered by a huge infrastructure that the company has built to run its retail business, Amazon EC2 provides a true virtual computing environment. By providing a variety of virtual machine or instance types, operating systems, and software packages to choose from, Amazon EC2 enables the user to instantiate virtual machines of his choice through a web service interface.

The user can change the capacity and characteristics of the virtual machine by using the web serviceinterfaces, hence named *elastic*. Computing capacity is provided in the form of virtual machines orserver instances by booting Amazon Machine Images (AMI), which can be instantiated by the user. An AMI contains all the necessary information needed to create an instance. The

primary GraphicalUser Interface (GUI) interface is the AWS Management Console(point and click) and a web serviceAPI that supports both Simple Object Access Protocol and Query Requests. The API providesprogramming libraries and resources for Java, PHP, Python, Ruby, Windows, and .Net. Theinfrastructure is virtualized by using Xen hypervisor, and different instance types are provided asfollows:

Standard instances—suitable for most applications

Micro instances—suitable for low-throughput applications

High-memory instances—suitable for high-throughput applications

High-CPU instances—suitable for compute-intensive applications

Cluster compute instances— suitable for high-performance computing (HPC) applications. The instances can be obtained on demand on an hourly basis, thus eliminating the need of forecasting computing needs earlier. Instances can be reserved earlier, and a discounted rate is charged for such instances. Users can alsobid on unused Amazon EC2 computing capacity and obtain instances. Such instances are called as Spot Instances. Those bids that exceed the current Spot Price is provided with the instance, which allows the user toreduce costs. The Spot Price is varying and is decided by the company.

Instances can be placed in multiple locations, which are defined by regions and availability zones. Availability zones are distinct locations that are engineered to be insulated from failures in other availability zones and provide inexpensive, low- latency network connectivity to other availability zones in the same region. Thus, placing the instances in multiple locations enables fault tolerance and failover reliability. The Amazon EC2 instances can be monitored and controlled by the AWS Management Console and the web service API.

However, AWS provides Amazon Cloud Watch, a web service that provides monitoring for AWS cloud resources, starting with Amazon EC2. It provides customers with visibility into resource utilization, operational performance, and overall demand patterns—including metrics such as CPU utilization, disk reads and writes, and network traffic.

Instances are authenticated using a signature- based protocol, which uses key pairs. Another important feature provided is the Amazon Virtual Private Cloud (Amazon VPC). The existing IT infrastructure can be connected to Amazon EC2 via a virtual private network (VPN).

Isolated computing resources are provided in Amazon VPC, and the existing management capabilities such as security services, firewalls, and intrusion detection systems can be extended to isolated resources of Amazon EC2.

Elastic load balancing (ELB) enables the user to automatically distribute and balance the incoming application's traffic among the running instances based on metrics such as request count and request latency. Fault tolerance and automatic scaling can be performed by configuring the ELB as per the specific needs. ELB monitors the health of the instances running and routes traffic away from a failing instance. An instance is stored as long as it is operational and is removed on termination. Persistent storage can be enabled by using either Elastic Block Storage (EBS) or Amazon Simple Storage Service (S3). EBS provides a highly reliable and secure storage, and the storage volumes can be used to boot an Amazon EC2 instance or be attached to an instance as a standard block device. Amazon S3 provides a highly durable storage infrastructure designed for mission- critical and primary data storage. Storage is based on units called objects whose size can vary from one byte to five gigabytes of data. These objects are stored in a bucket and retrieved via a unique, developer- assigned key. It is accessible through a web service interface and provides authentication proceduresto protect against unauthorized access.

AMAZON SIMPLE STORAGE SERVICE

Amazon Simple Storage Service known as Amazon S3, is the storage for the Internet. It is designed to make web-scale computing easier for developers. Amazon S3 provides a simple web service interface that can be used to store and retrieve any amount of data, at any time, from anywhere on theweb. It gives any developer access to the same highly scalable, reliable, secure, fast, inexpensive infrastructure that Amazon uses to run its own global net-work of websites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

Along with its simplicity, it also takes care of other features like security, scalability, reliability, performance, and cost.

Thus, Amazon S3 is a highly scalable, reliable, inexpensive, fast, and also easy to use service that meets design requirements and expectations.

Amazon S3 provides a highly durable and available store for a variety of content, ranging from web applications to media files. It allows users to offload storage where one can take advantage of scalability and pay-as-you-go pricing. For sharing content that is either easily reproduced or where one needs to store an original copy elsewhere, Amazon S3's Reduced Redundancy Storage (RRS) feature provides a compelling solution. It also provides a better solution in the case of storage for data analytics. Amazon S3 is an ideal solution for storing pharmaceutical data for analysis, financial- data for computation, and images for resizing.

Later this content can be sent to Amazon EC2 for computation, resizing, or other large- scale analytics without incurring any data transfer charges for moving the data between the services.

Amazon S3 offers a scalable, secure, and highly durable solution for backup and archiving critical data. For data of significant size, the AWS Import/ Export feature can be used to move large amountsof data into and out of AWS with physical storage devices. This is ideal for moving large quantities of data for periodic backups, or quickly retrieving data for disaster recovery scenarios.

Another feature offered by Amazon S3 is its Static Website Hosting, which is ideal for websites withstatic content, including html files, images, videos, and client-side scripts such as JavaScript.

AMAZON SIMPLE QUEUE SERVICE

Another service of AWS is Amazon SQS. It is a fast, reliable, scalable, fully managed message queuing service. SQS makes it simple and cost effective to decouple the components of a cloud application. SQS can be used to transmit any volume of data, at any level of throughput, without losing messages or requiring other services to be always available.

Amazon SQS is a distributed queue system that enables web service applications to quickly and reliably queue messages that one component in the application generates to be consumed by another component. A queue is a temporary repository for messages that are waiting to be processed.

Amazon SQS offers various features like allowing multiple readers and writers at the same time, providing access control facilities, guaranteeing high availability of sending, and retrieving messages due to redundant infrastructure. It also gives provision for having variable length messages as well as configurable settings for each queue.

MICROSOFT WINDOWS AZURE

Cloud computing provides a new way of looking at IT at Microsoft called Microsoft IT (MSIT). Cloudcomputing is now the preferred and default environment for new and migrated applications at

Microsoft. MSIT has developed a methodology and a set of the best practices for analyzing their current application portfolio for possible candidates to migrate to cloud computing. This analysis enables MSIT to select the ideal cloud -computing–based environment for each application. MSIT has captured these best practices and documented them for other Microsoft customers who wish to migrate their organizations to cloud computing.

Windows Azure Cloud Services (web and worker roles/PaaS) allow developers to easily deploy and manage application services. It delegates the management of underlying role instances and operating system to the Windows Azure platform. The Migration Assessment Tool (MAT) for Windows Azure encapsulates all the information to be aware of before attempting the application migration to Windows Azure.

Based on the response to a series of simple binary questions, the tool generates a report that outlines the amount of development effort involved to migrate the application, or the architecture considerations for a new application.

The Windows Azure Pricing Calculator analyzes an application's potential public cloud requirements against the cost of the application's existing infrastructure. This tool can help to compare current operational costs for an application, against what the operating costs would be on Windows Azure and SQL Azure.

Windows Azure Pack for Windows Server is a collection of Windows Azure technologies available to Microsoft customers at no additional cost for installation into their data center. It runs on top of Windows Server 2012

R2 and System Center 2012 R2 and, through the use of the Windows Azure technologies, it allows you to offer a rich, self-service, multitenant cloud, consistent with the public Windows Azure experience.

MICROSOFT ASSESSMENT AND PLANNING TOOLKIT

The Microsoft Assessment and Planning Toolkit (MAP) is an agent less, automated, multiproduct planning and assessment tool for cloud migration. MAP provides detailed readiness assessment reports, executive proposals, and hardware and software information. It also provides recommendations to help organizations accelerate the application migration process for both private and public cloud planning assessments.

MAP analyzes server utilization data for server virtualization and also server consolidation with Hyper-V.

SHAREPOINT

Microsoft offers its own online collaboration tool called SharePoint. Microsoft SharePoint is a web application platform that comprises a multi-purpose set of web technologies backed by a common technical infrastructure. By default, SharePoint has a Microsoft Office–like interface, and it is closely integrated with the Office suite. The web tools are designed to be usable by nontechnical users. SharePoint can be used to provide intranet portals, document and file management, collaboration, social networks, extranets, websites, enterprise search, and business intelligence. It also has system integration, process integration, and workflow automation capabilities. Unlike Google Cloud Connect, Microsoft SharePoint is not a free tool. But it has additional features that cannot be matched by Google or any other companies.

IBM CLOUD MODEL

IBM is one among the players in the field of cloud computing offering various cloud services to the consumers. IBM cloud computing consists of cloud computing solutions for enterprises as offered by the global IT company IBM. All offerings are designed for business use, marketed under the name IBM Smart Cloud. IBM cloud includes IaaS, SaaS, and PaaS offered through public, private, and hybrid cloud delivery models, in addition to the components that make up those clouds.

IBM offers an entry point to cloud computing whether a client is designing their own virtual private cloud, deploying cloud service, or consuming cloud workload applications. The IBM cloud framework begins with the physical hardware of the cloud. IBM offers three hardware platforms for cloud computing, which offer built- in support for virtualization. The next layer of the IBM framework is virtualization. IBM offers IBM Web sphere application infrastructure solutions that support programming models and open standards for virtualization.

The management layer of the IBM cloud framework includes IBM Tivoli middleware. Management tools provide capabilities to regulate images with auto- mated provisioning and de provisioning, monitor operations, and meter usage while tracking costs and allocating billing.

The last layer of the framework provides integrated workload tools. Workloads for cloud computing are services or instances of code that can be executed to meet specific business needs. IBM offers tools for cloud-based collaboration, development and test, application development, analytics, business- to-business integration, and security.

IBM offers a spectrum of cloud delivery options ranging from solely private cloud to solely public cloud and numerous variations in between. IBM gives the option to build a customized cloud solution out of a combination of public cloud and private cloud elements. Companies that prefer to keep all data and processes behind their own firewall can choose a private cloud solution managed by their own IT staff. A company may also choose pay-as-you-go pricing that allows them to run lower- profile applications on a secure public cloud model. Hybrid cloud options allow for some processes to be hosted and managed by IBM, while others are kept on a private cloud or on a VPN or Virtual Local Area Network. IBM also offers planning and consultation throughout the deployment process. Cloud computing is the best choice for mobile software. IBM offers five different cloud provision models:

Private cloud, owned and operated by the customer. Private cloud, owned by the customer but operated by IBM (or another provider) Private cloud, owned and operated by IBM (or another provider) Virtual private cloud services, based on multitenant support for individual enterprises
Public cloud services, based on the provision of functions to individuals

The majority of cloud users choose a hybrid cloud model, with some work-loads being served by internal systems, some from commercial cloud providers, and some from public cloud service providers.

For enterprise customers who perceive that the security risk of cloud computing adoption is too high, IBM specializes in secure private cloud offerings for building strictly private clouds, IBM offers IBM Workload Deployer and Cloudburst as ready-to-deploy, *cloud in a box*–style solutions.

Cloudburst provides blade servers, middleware, and virtualization for an enterprise to build its own cloud-ready virtual machines. Workload Deployer connects an enterprise's existing servers to virtualization components and middleware in order to help deploy standardized virtual machines

designed by IBM. For customers who prefer to perform their own integration of private clouds, IBM offers a choice of hardware and software building blocks, along with recommendations and reference architecture, leading the way to deployment. Clients may choose from IBM virtualization–enabled servers, middle-ware, and SaaS applications.

IBM SMARTCLOUD

IBM SmartCloud is a branded ecosystem of cloud computing products and solutions from IBM. It includes IaaS, SaaS, and PaaS offered through public, private, and hybrid cloud delivery models.
IBM places these offerings under three umbrellas: Smart Cloud Foundation, SmartCloud Services, and SmartCloud Solutions. Figure 5.4 briefly explains the architecture of IBM SmartCloud.
SmartCloud Foundation consists of the infrastructure, hardware, provisioning, management, integration, and security that serve as the underpinnings of a private or hybrid cloud. Built using those foundational components, PaaS, IaaS, and backup services make up SmartCloud Services. Running on this cloud platform and infrastructure, SmartCloud Solutions consist of a number of collaboration, analytics, and marketing SaaS applications.
Along with IaaS, PaaS, and SaaS, IBM also offers Business Process as a Service (BPaaS). Infrastructure cloud services provide the consumer the provision of processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. In platform cloud services, a consumer can deploy consumer-created or -consumer-acquired applications onto the cloud infrastructure created using programming languages and tools supported by the provider.

Application cloud services allow consumers to use the provider's applications - running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web- based e-mail).
Business process cloud services are any business process (horizontal or vertical) delivered through the cloud service model (multitenant,- self-service provisioning, elastic scaling, and usage metering or pricing) via the Internet with access via web-centric interfaces and exploiting web-oriented cloud architecture. The BPaaS provider is responsible for the related business functions.

SAP LABS

SAP Labs makes enterprise software to manage business operations and customer relations. SAP is the leader in the market of enterprise applications in terms of software and software-related service. The company's best-known software products are its enterprise resource planning application systems and management (SAP ERP), its enterprise data warehouse product-—SAP Business Warehouse (SAP BW), SAP Business Objects software, and most recently, Sybase mobile products and in-memory computing appliance SAP HANA. SAP is one of the largest software companies in the world.

SAP HANA CLOUD PLATFORM

SAP HANA Cloud Platform is an open-standard, Eclipse-based, modular PaaS. In SAP HANA Cloud Platform, applications are deployed via command-line tools to the cloud as web application archive (WAR) files or OSGi bundles. OSGi bundles are normal jar components with extra manifest headers. The applications run within the Java-based SAP HANA Cloud Platform runtime environment. It is powered by SAP HANA and can be maintained using web-based management tools.

The main features of SAP HANA Cloud Platform are as follows:

- Enterprise platform built for developers
- Native integration with SAP and non-SAP software In-memory persistence Secure data platform
- Lightweight, modular runtime container for applications

SAP HANA Cloud Platform lets the users quickly build and deploy business and consumer applications that deliver critical new functionality to meet emerging business needs. It also helps connect users with customers in more engaging experiences. It provides connectivity based on the cloud connectivity service. As a result, the platform streamlines the integration of new applications at the lowest possible total cost of ownership. Support for open programming standards provides a low barrier entry for developers. This makes them productive from the start in building enterprise applications that can integrate with any SAP or non- SAP solution. No new coding skills are required to work with SAP HANA.

VIRTUALIZATION SERVICES PROVIDED BY SAP

ERP virtualization increases a project's return on investment by maximizing hardware utilization. The business benefits of virtualization of ERP applications are shorter development cycles, reduction in IT costs, improved availability, and energy saving. A joint service from SAP and VMware helps in transition to a more open and flexible private cloud platform based on proven virtualization technology.

SALES FORCE

Salesforce.com is a cloud computing and social enterprise SaaS provider based in San Francisco. Of its cloud platforms and applications, the company is best known for its Sales force CRM product, which is composed of Sales Cloud, Service Cloud, Marketing Cloud, Force.com, Chatter, and Work.com. In addition to its products and platforms, Salesforce.com created App Exchange, a custom application building and sharing platform. The company also has consulting, deployment, and training services.

SALES CLOUD

Sales Cloud refers to the *sales* module in Salesforce.com. It includes Leads, Accounts, Contacts, Contracts, Opportunities, Products, Price books, Quotes, and Campaigns (limits apply). It includes features such as web-to-lead to support online lead capture, with auto response rules. It is designed tobe a start-to-end setup for the entire sales process. Sales Cloud manages contact information and integrates social media and real-time customer collaboration through Chatter. The Sales Cloud gives a platform to connect with customers from complete, up-to- date account information to social insights, all in one place and available anytime, anywhere.

Everything is automatically pushed in real time, from contact information to deal updates and discount approvals.

Salesforce.com created the Sales Cloud to be as easy to use as a consumer website like Amazon and built it in the cloud to eliminate the risk and expense associated with traditional software. With its open architecture and automatic updates, the Sales Cloud does away with the hidden costs and drawn-out implementations of traditional CRM software. By continuing to innovate and embrace technologies like mobile, collaboration, and social intelligence, the Sales Cloud has continued to

Pull ahead of the competition.

SERVICE CLOUD: KNOWLEDGE AS A SERVICE

Service Cloud refers to the *service* (as in *customer service*) module in Salesforce.com. It includes Accounts, Contacts, Cases, and Solutions. It also encompasses features such as the public knowledge base, web-to-case, call - center, and self-service portal, as well as customer service automation. It includes a call center–like case tracking feature and a social networking plug-in for conversation and analytics.

The Service Cloud delivers the world's first enterprise-grade knowledge base to run entirely on an advanced, multitenant cloud platform. That means one can get all the cloud computing benefits that Salesforce.com is known for delivering without expensive data centers or software.

Just powerful knowledge management, without the hassle of on-premises software, is provided. Unlike stand-alone applications, this knowledge base is fully integrated with everything else. Service Cloud has to offer all the tools one needs to run the entire service operation. When the consumer's knowledge base is a core part of CRM solution, knowledge as a process can be managed. One can continually create, review, deliver, analyze, and improve the knowledge. And, because it is deliveredby the Service Cloud, user's knowledge is available wherever other customers need it. Agents have the right answers at their finger tips to communicate over the phone, send out through an e-mail, or share via a chat client. The same knowledge base serves up answers to the service website is a part of company's public site. If one wants to take advantage of social channels like Twitter or Facebook, one can easily share knowledge that is tapped into the wisdom of the crowd to capture new ideas or answers. All this is done securely.

The Service Cloud gives the tools that are needed to manage knowledge at enterprise scale. But it also delivers the same great ease of use that Sales force. com is known for. That means user will benefit no matter what size or how complex the business is.

RACKSPACE

Rackspace Cloud, a part of Rackspace, is another player in the cloud computing market. Offering IaaS to clients, it has been used by a large number of enterprises. Rackspace Cloud offers three cloud computing solutions—Cloud Servers, Cloud Files, and Cloud Sites. Cloud Servers provide computational power on demand in minutes;

Cloud Sites are for robust and scalable web hosting, and Cloud Files are for elastic online file storage and content delivery.

Cloud Servers is an implementation of IaaS where the computing capacity is provided as virtual machines that run in the Cloud Servers systems. The virtual machine instances are configured with different amounts of capacities. The instances come in different flavors and images. A flavor is an available hardware configuration for a server. Each flavor has a unique combination of disk space, memory capacity, and priority for CPU time. A varied set of instances are available for the user to choose from.

These virtual machines are instantiated using images. An image is a collection of files used to create or rebuild a server. A variety of prebuilt operating system images are provided by Rackspace Cloud (64-bit Linux distributions— Ubuntu, Debian, Gentoo, CentOS, Fedora, Arch, and Red Hat Enterprise Linux) or Windows Images (Windows Server 2008 and Windows Server 2003). These

images can be customized to the user's choice to create custom images.

The Cloud Servers systems are virtualized using the Xen Hypervisor for Linux and Xen Server for Windows. The virtual machines that are generated come in different sizes and measured based on theamount of physical memory reserved. Currently, the physical memory can vary from 256 MB to

15.5 GB. In the event of availability of extra CPU power, Rackspace Cloud claims to pro-vide extra processing power to the running workloads, free of cost.

Backup schedules can be created to define when to create server images. This is a useful feature, which enables the user to continue work in the event of failures by using the backup images. Custom images are helpful in creating backup schedules. A type of images, referred to as *gold* serverimages, can be produced if the servers of that configuration are to be instantiated frequently.

Cloud Servers can be run through the Rackspace Cloud Control Panel (GUI) or programmatically via the Cloud Server API using a RESTful interface. The control panel provides billing and reporting functions and provides access to support materials including developer resources, a knowledge base, forums, and live chat. The Cloud Servers API was open sourced under the Creative Commons Attribution

3.0 license. Language bindings via high-level languages like C++, Java, Python, or Ruby that adhereto the Rackspace specification will be considered as Rackspace- approved bindings.

The virtual machine instances are authenticated in the API by a token-based protocol that uses the HTTP x-Header. Private/public keys are used to ensure Secured Shell Access.

Cloud Servers scale automatically to balance load. This process is auto-mated and initiated from either the Rackspace Cloud Control Panel or the Cloud Server API. The amount to scale is specified; the Cloud Server is momentarily taken offline; the RAM, disk space, and CPU allotment are adjusted; and the server is restarted. A Cloud Server can be made to act as a load balancer using simple readily available packages from any of the distribution repositories. Rackspace Cloud is working on beta version of the Cloud Load Balancing product, which provides a complete load balancing solution.

Cloud Servers are provided persistent storage through RAID10 disk storage; thus, data persistency isenabled leading to better functioning.

VMWARE

VMware, a leader in virtualization technology, has come up with enterprise cloud computing solutions. Having been a dominating player in the virtualization domain, VMware is currently providing a range of products for the development of private and public clouds and for leveraging the services offered by both as a hybrid cloud, such as VMware vCloud Director, VMware vCloud Datacenter Services, VMwarevSphere, and VMware vShield to name a few.

Private clouds enable the better usage and management of internal IT infra-structure than the traditional methods. Greater operational efficiency, secure, fault-tolerant, well-managed computing environments can be modeled and operated. VMware's private cloud offering provides greater standardization, rapid provisioning, and self- service for all applications and unparalleled cost savings by consolidating their physical infrastructures. VMware's modular technology enables the user to select from a variety of hardware, software, and certified service providers to result in efficient cloud computing. Thus, the family of products offered by VMware promotes compatibility and retains the choice of freedom for the users to obtain desired services.

Private clouds can be created by using the VMware vSphere and VMware vCloud Director. VMware vSphere is a robust virtualization platform used to transform IT infrastructures into virtual storage, compute, and network resources and provide them as a service within the organization. VMware vSphere provides services at both the infrastructure and application levels. At the infrastructure level, it provides options to perform efficient operation and management of the compute, storage, and network resources. At the application level, service-level controls are provided for the applications running on the underlying infrastructures, leading to available, secure, and scalable applications.

The VMware vCloud Director, coupled with VMware vSphere, is a soft-ware solution that enables enterprises to build secure, multitenant private clouds by pooling infrastructure resources into virtual datacenters and exposing them to users through web-based portals and programmatic inter-faces as fully automated, catalog- based services. VMware vCloud Director abstracts the virtual computing environment from the underlying resources and provides a multitenant architecture that features isolated virtual resources, independent LDAP authentication, specific policy controls, and unique catalogs. VMware vShield technologies are used to provide security to these environments by using services like perimeter protection, port-level firewall, NAT and DHCP services, site-to-site VPN, network isolation, and web load balancing. The VMware vCloud Director allows users to catalog infrastructure and application services of the desired configurations and deploy and consumethem as needed. Interactions with the virtual data centers or the catalogs are through a user-friendly web portal or the vCloud API. The vCloud API is an open, REST-based API that provides scripted access, complying with the open virtualization format (OVF). The API can be used along with VMware vCenter Orchestrator to automate and orchestrate operational processes like routine tasks, activities, and workflows.

Public and hybrid cloud solutions are provided by VMware by partnering with other companies, certified as service providers.

VMware vCloud Datacenter Services and VMware vCloud Express offer efficient solutions for utilizing IaaS either as a public cloud or a hybrid cloud. vCloud Datacenter Services provides a scalable environment, where internal resources are augmented with the external resources. vCloud Datacenter Services are built on the same technology and foundations as VMware vCloud Director and VMware vSphere to enable interoperability between cloud environments. Thus, the user is free to burst his private cloud into public cloud of his preferred service provider. vCloud Express is an IaaS offering delivered by leading VMware service provider partners. It is a cobranded service that provides reliable, on-demand, pay-as-you- go infrastructure. The VMware vCloud Express providers are Virtacore vCloud Express, Hosting.com, Melbourne IT, and Terremark's vCloud Express. Instance types, load balancing, storage options, and pricing vary between service providers.

MANJRASOFT

Manjrasoft is one of the non major providers of cloud services. But it has come up with a platform called *Aneka* that provides a set of services that help the development of applications in an easier way. Manjrasoft develops market- oriented cloud computing platforms that allow one to build, accelerate, and manage the applications ultimately saving one's time and money, leading to enhanced business productivity and profit.

ANEKA PLATFORM

Aneka provides a set of services that make enterprise cloud construction and development of applications as easy as possible without sacrificing flexibility, scalability, reliability, and extensibility.

Figure 11.5 gives an overview of the Aneka platform. The key features sup-ported by *Aneka* are as follows:

A configurable and flexible execution platform (container) enabling pluggable services and security implementations. Multiple authentication/ authorization mechanisms such as role-based security and Windows domain–based authentication are considered for this purpose.

Multiple persistence options including Relational Database Management System (RDBMS), Structured Query Language (SQL) Express, MySQL, and flat files.

Software development kit (SDK) supporting multiple programming models including object-oriented thread model, task model for legacy applications, and Map Reduce model for data-intensiveapplications.

Customtools such as Design Explorer for parameter sweep studies.

Easy to use management tool for SLA and Quality of Service (QoS) negotiation and dynamic resource allocation.

Supports deployment of applications on private or public clouds in addition to their seamless integration.

Aneka allows servers and desktop PCs to be linked together to form a very powerful computing infrastructure. This allows companies to become energy efficient and save money without investing in a number of computers to run their complex applications.

Each *Aneka* node consists of a configurable container that includes information and indexing, scheduling, execution, and storage services.

Aneka supports multiple programming models, security, persistence, and communications protocols.

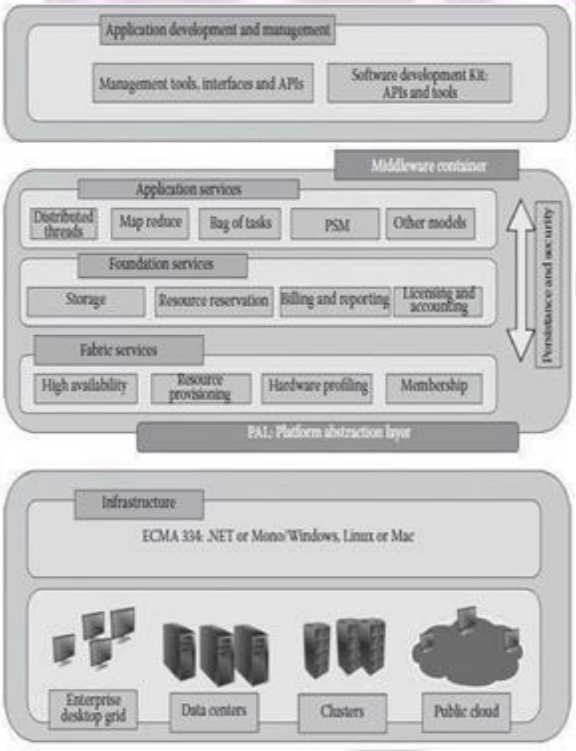


FIGURE 5.5
Overview of the Aneka platform

SUMMARY

Based on the specific requirements, the user has to make a trade -off between open source and closedsource tools/services.

Though there are a number of companies, we have chosen few companies that have progressed a lot in this area. Table 5.1 gives the information about few providers and the prices on a per hour basis along with the service

Basic Cloud Server Pricing for Virtual Servers

	VIRTUAL MACHINE INSTANCE	BANDWIDTH IN	BANDWIDTH OUT	BACKUP	SUPPORT
RackSpace Cloud 1,024 MB/40 GB	$.06/hour or $43.80/ month	$0.08/GB	$0.22/GB	$0.15/GB	Included
Amazon EC2— Small server Linux	$325/year, plus $0.03/hour	$0.10/GB	$0.17/GB	$0.15/GB/ month	$0.015/ instance/ hour
GoGrid 1,024 MB/ 60 GB	$0.19/hour	Free	$0.50/GB	$0.15/GB	Included

Table 5.1