1

What Is Cloud Computing?

Cloud computing is a broad term that simply means delivering computing services – which may include servers, databases, storage, networking, software, data analytics, security solutions, organizational systems, virtual computers, and much more – over the Internet. The term "cloud" came from a symbol in old flow charts and diagrams used to represent the Internet (see Figure 1.1). This symbol suggested resources exist, but we may not necessarily know exactly where they are or how they work. But, through Internet connections and Web-based interfaces, we can set up and configure exactly what we need to use. So, cloud computing becomes the mindset for shared resources that can be used economically from just about anywhere.

> The phrase "Cloud Computing" originates from the "Cloud" symbol used in flow charts and diagrams to represent the Internet. The idea is that any computer with a Web interface has access to an incredible pool of computing resources, power, applications, and files.

Why Cloud Computing?

Cloud computing was the result of the perfect storm to continue the atmospheric paradigm. With the growing number of software systems, databases, security requirements, hardware needs, and so forth, IT specialists found it difficult to keep up. Becoming an expert in so many areas was daunting and beyond the capabilities of many Information Technology (IT) departments. Technologies changed continually. Hardware and software updates needed to be rolled out across organizations that often were spread out geographically. Chief Information Officers (CIOs) and IT managers looked for ways to reduce the burden and provide high end services to end users who needed to do their jobs.

At the same time, the Internet was maturing to the point where high-speed connections meant increasingly complex applications worked remotely. This opened the door to many new possibilities. Experts could be located anywhere that had an Internet connection. Suddenly, cloud computing concepts enabled IT specialists to offer new technologies within their organization without needing in-depth knowledge about, or high levels of expertise with each implementation. The cloud provided access to custom tool sets, even if the tools

Cloud Technologies: An Overview of Cloud Computing Technologies for Managers, First Edition. Roger McHaney.

Companion website: www.wiley.com/go/mchaney/cloudtechnologies

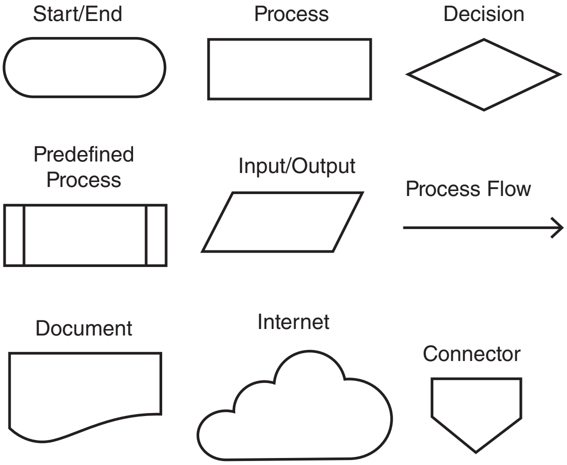

Figure 1.1 Flow chart origins of the cloud.

were used only on rare occasion. This allowed IT departments and users to cut costs and returned the focus to core business needs. Instead of IT driving the business, the business used IT to achieve success with its primary mission and to remove technology-related obstacles.

Cloud Computing's Focus

The benefits of cloud computing helped this new concept win rapid acceptance among Senior IT Managers (although lingering concerns remain which we will cover later). Today, cloud computing focuses on three primary areas: *cost reduction*, *capacity planning*, and *business agility*. All three areas relate to business needs and help make a case for moving to the cloud (see Figure 1.2).

Cost Reduction

IT costs generally come in two categories: (i) cost of new equipment/software (e.g. purchase costs); and (ii) ongoing costs of ownership (e.g. operational costs). As equipment ages and maintenance becomes more time consuming, operational costs can quickly escalate.

In many organizations, it is important to directly tie the costs of IT items to their related business uses. In other words, managers want to know how much it costs to accomplish business outcomes so the profitability of certain activities can be monitored. This alignment between IT costs and business performance can be very difficult to both understand and maintain. Several factors come into play here. First, software and hardware may be used in many ways and for many purposes across an organization. Second, IT expenditures often relate to what internal users expect their maximum usage to eventually become.

Think of it in this way. You have just graduated with your university degree. Your smart boy or girl friend sees the future value of your education and decides to make the relationship more permanent by proposing marriage (they love you as well I am sure!). You think ahead a little and decide to invest in your dream home. Even though it is just the two

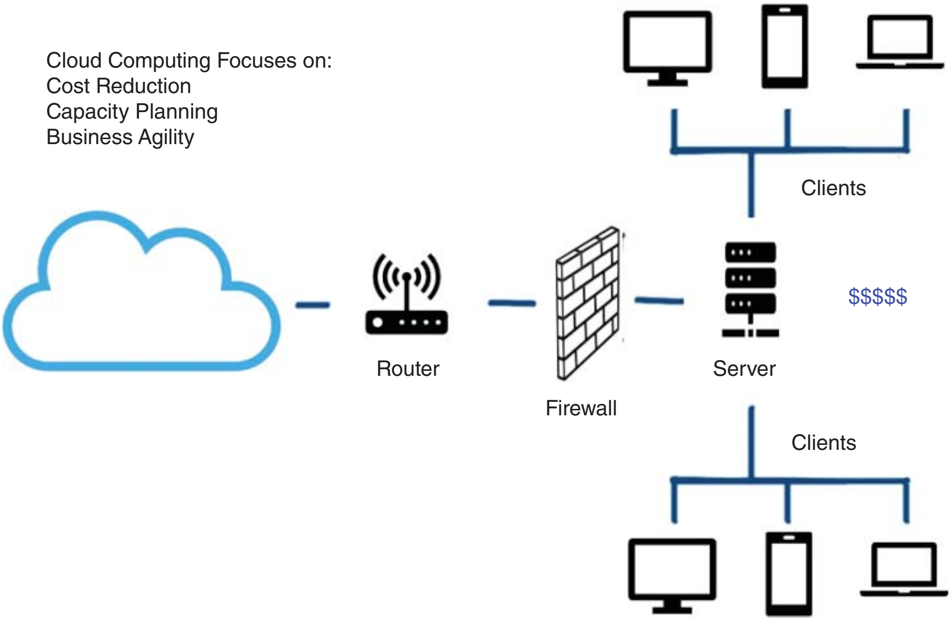

Figure 1.2 The focus of cloud computing.

of you and you could happily live in a single room flat, you decide to buy a four bedroom mini-mansion with a swimming pool. The idea is that in the future, you will need bedrooms for your 2.5 children and a place for your mother-in-law to stay when she visits. If you were trying to allocate the costs, it would be a bit of a challenge. You do not have children yet, but you "bought" space where they will eventually sleep.

Organizations face the same challenge. Do you buy for the future to leverage current purchasing power? Or do you wait until the last minute and face challenges that may include higher costs, slow implementation, and lack of availability? So, you can see the dilemma that many organizations face.

Figure 1.3 illustrates how purchases of software and hardware comprise part of a firm's total IT costs, but the ongoing costs can climb, particularly when a new item is acquired (represented by the stair step in "Purchases"). As items become older, often the cost of maintenance increases, sometimes in ways that were unexpected.

In most cases, operational overhead accounts for a large percent of IT budgets. Over time, these costs probably will exceed up-front investment and purchase costs. Think about everything that goes into operating expenses:

- Technical personnel with high levels of skills. This means regular training, certification and other costs are required.
- Software and hardware upgrades and patches that need testing, user training, and installation time.
- Software leasing costs. Many software packages require a yearly fee to pay for upgrades and so forth.

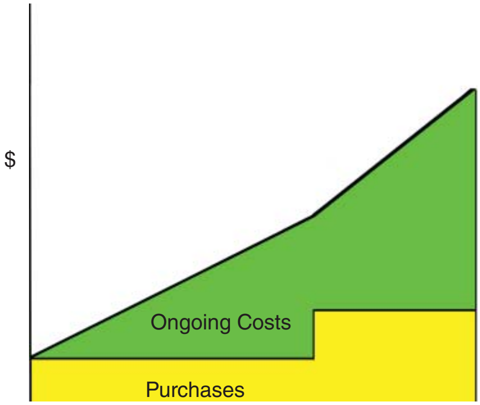

Figure 1.3 Ongoing costs and purchases as part of total cost.

- Operation costs like electric utility bills and investments for cooling systems, surge protection, air filtration systems and so forth.
- The costs of security and access control measures must be in place. These operations are extremely important and include software, network, and physical infrastructure components (e.g. virus protection, intrusion detection, video cameras, locks, and so forth).
- Help desk, staff specialists, and administrative assistants might be required to track licensing, do training and provide daily assistance.
- In addition, many other expenses also exist (like insurance, audits, travel, staff meetings and retreats, et cetera).

The costs of IT, since these are difficult to tie directly to a business activity, may be viewed as a cost sink. This often places IT expenditures right in the crosshairs of managers responsible for cutting costs and reducing budgets. So, IT managers must both innovate and cut costs. That can become a bit of a nightmare. Luckily, cloud computing can help eliminate many of these problems and we will investigate these advantages in Chapter 6.

Capacity Planning

Another area driving IT leaders into the cloud relates to capacity planning. This is the process of deciding how to prepare for the future. In the example of the house purchase, this would include discussions about how many children the happy couple might like to eventually have, how often visitors are expected, and other aspects of how future life is expected to unfold. As you can imagine, often planning and dreaming ends up a long way from reality.

In the IT world, similar activities occur. Capacity planning is the process used to determine and fulfill future demands expected to be placed on an organization's IT resources, duties, operations, and services. Often, the CIO or another senior IT leader is part of an organization's strategic planning group and works to understand the maximum and type of IT demand that may emerge. This figure is used to determine required resources and if existing IT assets are capable of meeting requirements for a specific time. Differences between perceived needs and existing capacity can have several results. If a discrepancy exists, a system might become overwhelmed, also called under-provisioned. Under-provisioned systems cannot meet the needs of the users. If the opposite situation exists and too much IT capacity exists, the system is considered over-provisioned. This is inefficiency and causes waste. Having too much capacity results in waste and money that could have been spent in other, more productive ways. Capacity planning focuses on minimizing any discrepancy, whether it results in over- or under-provisioning and seeks to achieve predictable efficiency and optimal performance levels.

Capacity planning can be approached from a variety of perspectives. Three popular approaches to capacity planning strategy are:

- *Match Strategy* which involves adding IT resources in small increments to match demand as best as possible. This strategy avoids over-provisioning but sometimes fails to take advantage of volume discounts. It also can result in systems that are not able to ramp up quickly enough (e.g. adding staff, training, and so forth).
- *Lead Strategy* means that IT resource capacity is added in anticipation of demand. This strategy can take advantage of volume purchases but is more susceptible to over-provisioning and can result in unused resources that become obsolete before they are needed.
- *Lag Strategy* adds IT resources only after capacity is reached. This approach can suffer from under-provisioning and from being unable to rapidly react to organizational needs. It may be best in situations where demand for IT resources is steady and less susceptible to unexpected changes.

Capacity planning often is considered one of a CIO's most difficult challenges. Finding the sweet spot between too much and too little is difficult and requires understanding the business and its strategic plans and then matching that with knowledge of the IT world, together with a sense of where major systems and capabilities are heading in the future. Many CIOs have lost their jobs because they anticipated things that did not happen or were swept away by broader technology changes. Attending conferences and trade shows can help an IT planner avoid making big mistakes but foretelling the future is never easy. Despite these challenges, a bright spot exists: the cloud makes capacity planning much easier, as we will see in this text.

Organizational Agility

Another major business driver making the cloud approach to computing attractive is organizational agility. The cloud allows businesses to easily acquire and roll out the latest technologies without needing to develop as much internal expertise. More than ever, successful businesses need the ability to change rapidly to innovate and compete. External technology evolves, competitive forces emerge, and internal needs force change on an organization. Agility is the measure of how responsive an organization can be when it comes to these changes.

A big part of organizational agility is called elasticity. IT departments must be ready to respond to business changes by upsizing or downsizing capabilities to match the current situation. This is a little different than capacity planning because it involves a time element. How fast can an IT system resize to meet demand? And, often this is in response to changes in scope that are much different than what was expected or planned.

If we go back to our happy family example, think about the instance where following a visit to the doctor, the birth of an expected daughter or son turned out to be triplets. Agility would be how fast preparations can be made. Do more cribs need to be purchased? Do babysitting services need to be rethought? Is a new house needed? So, capacity planning is ongoing and necessary, and it must be done in a timely fashion (a few months for our example).

Businesses are no different. Perhaps a new product goes viral and demand outstrips production. Or a company website becomes incredibly popular and its servers need to be upgraded to ensure people are not frustrated when they visit. If they are purchasing a product on the site, this becomes even more crucial. The IT department may have little time to respond because viral popularity could come and go fast. And then later, maybe things need to be ramped down again to prevent being over-provisioned which preserves capital and reduces waste.

In many cases, particularly when non-cloud, in-house solutions are used, capacity changes can be difficult, or perhaps not possible. Cloud-based solutions offer businesses more flexibility when it comes to responding to change because of their inherent scalability. If a business does not have the ability to scale its solutions rapidly and tries anyway, the result could be disastrous. For instance, reliability might suffer, customers could be lost and opportunities for new competitors might emerge.

> The term “Scalability” is associated with computing solutions. When a system is scalable, it can grow without being hampered by existing structure or available resources. A scalable system can respond to higher demand with little or no changes required.

How Is Cloud Computing Hosted?

Cloud computing has become the standard way for companies to access IT infrastructure, hardware, and software resources. A cloud can be hosted specifically by an organization for its exclusive use, or an organization can use services that reside on a cloud hosted by an outside vendor. These two deployment approaches, called private and public clouds, each have pros and cons. Some organizations use a third approach called a hybrid cloud.

Private Cloud Deployment

A private cloud, also called an internal or enterprise cloud, is hosted on hardware systems located within an organization’s data center. A private cloud is essentially an intranet where all the data and services are protected by a firewall. Private clouds are generally used by large organizations that may already have data centers or extensive IT expertise. The benefit is that all data is held by the organization on its premises and can offer customized security solutions. This might be preferable to organizations that manage critical and confidential data. The biggest drawback is that an expert, internal IT group is responsible for all data management, software updates, hardware maintenance, and so forth.

Public Cloud Deployment

An organization that opts to use a public cloud has turned the day-to-day management of hosting over to a third-party vendor. Generally, the organization using the public cloud is

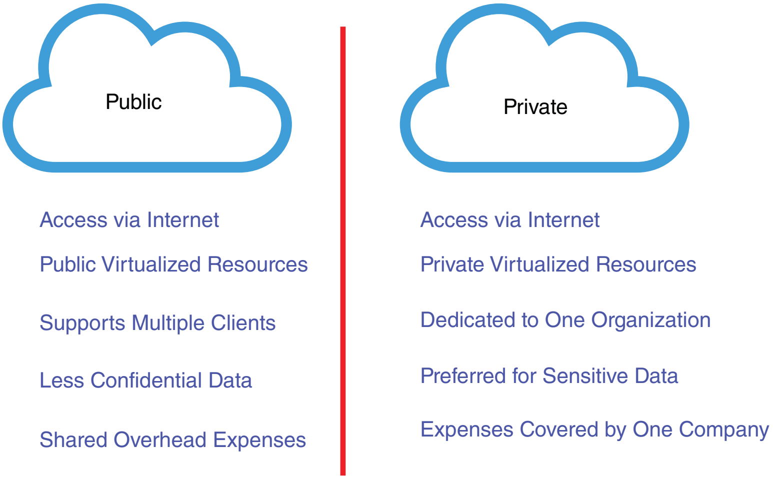
Public
Access via Internet
Public Virtualized Resources
Supports Multiple Clients
Less Confidential Data
Shared Overhead Expenses
Private
Access via Internet
Private Virtualized Resources
Dedicated to One Organization
Preferred for Sensitive Data
Expenses Covered by One Company

Figure 1.4 Public versus private clouds.

not responsible for managing the system, updating software, and maintaining the hardware. Instead, data and applications are stored in the host's data center where all IT operations are managed. Using a public cloud can reduce IT overhead because these functions are taken care of by someone else with expertise in the area, and that shares the costs among several client organizations. Some organizations avoid public clouds because their data is no longer on their own premises and they worry that security problems could result. While this is possible, it is not any more likely than data security breaches on private clouds. Most public cloud hosts have excellent security and maintain separate areas for each of their clients' data and applications (see Figure 1.4).

Hybrid Cloud Deployment

Other organizations need features offered by both public and private clouds. Perhaps their sensitive data must be held on-premise or they use a custom application that requires internal hosting. A hybrid cloud combines features of both public and private clouds, using technology that permits sharing data and applications between them. A hybrid cloud gives a business greater flexibility and the ability to deploy applications and data in more ways.

What Are the Different Types of Cloud Solutions?

Whether a cloud is private or public, it can be used for a variety of functions. It can replace hardware, software, organizational computing infrastructure, and many other IT resources. Cloud computing can be simple. For example, using a service like Dropbox for file sharing. Or it can be complex. For example, a multinational Fortune 500 company using an enterprise resource planning system like SAP to run its operations in 16 different countries. Despite the wide range of possibilities, most solutions fit into one of three general categories. These are Software as a Service (SaaS), Platform as a Service (PaaS),

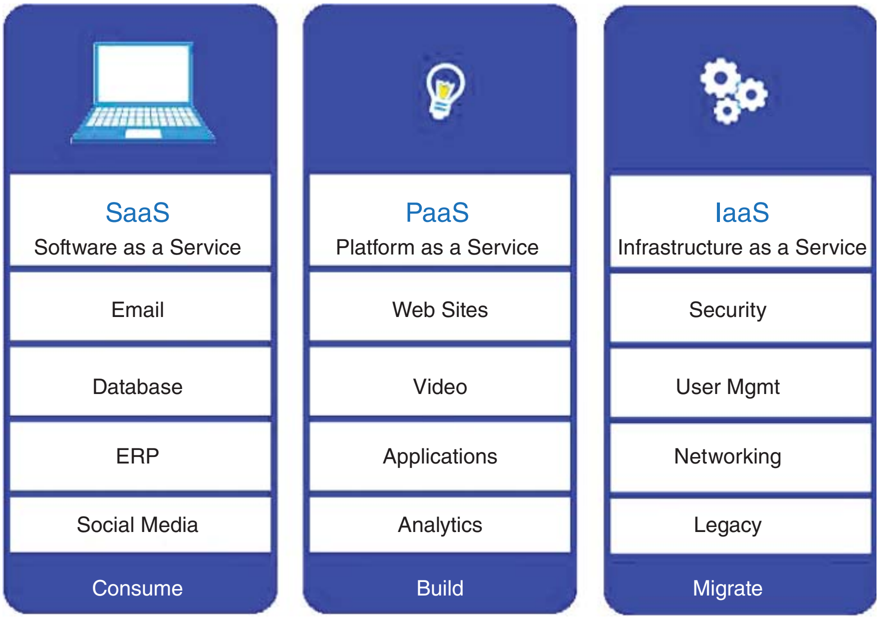

Figure 1.5 Three general cloud solution areas.

and Infrastructure as a Service (IaaS). The type of cloud solution reflects the complexity of the systems and how many functions it can replace. The next sections describe these three general cloud solutions in more detail. Figure 1.5 illustrates our starting point.

Software as a Service (SaaS)

SaaS usually first comes to mind when people talk about cloud computing. It started as an Internet-based software distribution solution. Not too long ago, software was installed from disks, CDs or DVDs directly onto user desktops. As the Internet became more robust and faster, SaaS was the natural next step. Software ran on host computers accessed via the Internet. This ensures customers have the latest edition of software and more easily manage their payment and use options. SaaS is sometimes called on-demand or hosted software.

In the SaaS approach, a software vendor hosts the application at their data center and, in most cases, a customer accesses it via a Web browser or a custom interface that has been downloaded to the user's computer. SaaS applications generally target business users, although more and more we see companies going directly to private consumers. In the business world, examples of SaaS include subscription-based software like Salesforce or SAS; and pay-per-use software like WebEx which charges for online meetings. Other services, like Dropbox charge a yearly fee that is prorated, based on level of service.

SaaS is currently the most popular software distribution model for many business applications, including document management, accounting, human resource management, help desk management, content management, collaboration, and other core business areas. Thousands of SaaS vendors offer 10's of thousands of business applications, made possible by the Internet. Most SaaS vendors use the following approaches to manage their software:

- Software updates applied automatically and transparently to the end-user.
- Users pay a subscription fee based on number of seats, level of service, storage requirements or other criteria.
- The only hardware required by the customer is a desktop, laptop, or mobile device capable of connecting over the Internet. In most cases, no extra hardware is sent to the customer.

Benefits of SaaS Solutions

- Elasticity: Customers can quickly and seamlessly add users, capacity, or capabilities. Often, this is automatic.
- Accessibility: End-users can access the software from anywhere that an Internet connection exists. This is particularly helpful for those who travel or are on the road for their work.
- Cost Savings: Costly infrastructure and in-house expertise costs can be minimized or even eliminated. These are built into subscription fees.

Platform as a Service (PaaS)

PaaS is related to SaaS but is a broader and more complex form of cloud computing generally used by organizations involved in custom application development. Usually, PaaS solutions make it easier to develop, test, collaborate, maintain, track releases and changes, and perform other duties related to software creation. PaaS generally offers a configured sandbox for software testing, project management features, and a deployment environment that enables customers to roll out their cloud-based applications. Examples of commercial PaaS clouds include Amazon Web Services (AWS) and Microsoft Azure.

PaaS usually includes a solution or software stack for development. The term solution stack refers to a set of software components that are required to complete an entire application. This means no further components are needed to create a complete platform. Applications run on top of the resulting platform. Consider a web application as an example. The developer would need to define the stack as the operating system, web server, database, and programming/development language. In the case of business applications, a stack might consist of the operating system, middleware, database, and end-user application. In a development environment, components within the stack often are created by different experts independently.

Commonly used stacks are given acronyms that represent the components. Figure 1.6 provides an example of the WINS stack. WINS comprises Windows Server for the operating system, Internet Information Services as the Web servers, .NET as the framework for software development, and SQL Server to run the required database.

The terms *software stack*, *solution stack* and just *stack* often are used interchangeably by developers and IT specialists. However, the term solution stack may include hardware components in the overall solution. In cloud-based environments, these have become virtualized and integrated into the PaaS cloud solution.

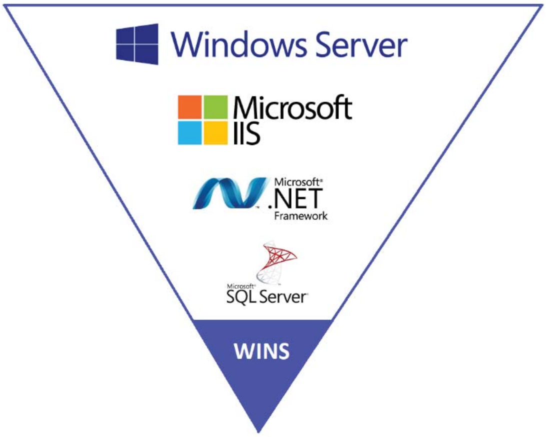

Figure 1.6 WINS solution stack.

Other Example Stacks

WIMP: Windows (operating system); Internet Information Services (web server); MySQL (database); PHP, Perl, or Python (programming language)
WISA: Windows Server (operating system); Internet Information Services (web server); SQL Server (database); ASP.NET (web framework)

PaaS clouds offer several beneficial advantages to organizations. Foremost is the sense of a community formed by an organization's developers. Most modern applications require the interaction and cooperation of many specialists. The idea that an online community is facilitated within the PaaS cloud is an excellent benefit of the technology. Another plus in PaaS, which is common to most cloud services, is that an organization is not required to update its infrastructure software. The PaaS vendor manages all updates, fixes, patches, and regular software maintenance activities. This reduces operational costs. Likewise, upfront investment costs are reduced. Finally, the software development team spends its time building applications rather than maintaining, updating, and working on routine tasks associated with the testing and deployment environment.

Infrastructure as a Service (IaaS)

IaaS can be viewed as the lowest level of cloud solution and focuses more on system configuration issues. Specifically, IaaS clouds host infrastructure components traditionally managed in on-premise data centers. IaaS offers services that include servers, data storage, networking hardware, and virtual machines. IaaS seeks to be a fully outsourced service that replaces a data center. IaaS offers pre-configured hardware (or software) through an interface. Customers install software and services on the IaaS cloud and run/manage their applications as if it were an on-premise data center.

An IaaS model adds value by providing services that accompany infrastructure components. Often, these include billing, activity/load monitoring, access logs, security, load balancing, backup and recovery, replication, and other safety measures. IaaS services seek to be policy-driven, which means customers can automate many operations. Customers access services and are allocated resources using an Internet interface in most cases, although for more secure operations, a wide area network (WAN) or virtual tunnel may be used. An IaaS user may log into the cloud to install their own applications or software stack components or configure virtual machines (VMs), deploy middleware, create backups, and install enterprise applications. Customers also rely on the IaaS provider to monitor application and server performance, track costs, balance network traffic, manage disaster recovery, and monitor security.

The term **middleware** is used to describe the software layer between the operating system and applications on each side of a distributed computing system in a network. Common middleware applications include web servers, application servers, content management systems, databases, and tools, such as ODBC, that enable database integrations.

Examples of organizations that provide IaaS services are Google (with its Google Cloud Platform), RackSpace, and Amazon with AWS. Many cloud providers offer both PaaS and IaaS services, AWS is one such example.

IaaS can be less expensive, faster, and more cost-efficient than developing an in-house data center. IaaS allows businesses to rent or lease infrastructure to avoid over- or under-provisioning their operations. IaaS is particularly useful for organizations that see fluctuations in their workload or projects. For companies that offer software services, being able to expand their infrastructure when a new project is under development and then release that infrastructure when complete, makes more sense than purchasing hardware that soon will become obsolete.

IaaS flexibility is reflected in the payment models used by most vendors. Generally, IaaS clients are charged on a usage basis. Pay-as-you-go models are currently the most common approach with billing based on transactions, time, or virtual space used. For customers deploying in an IaaS environment for the first time, billing can become an issue. Sometimes, the rates sound inexpensive but add up quickly. Since every activity on the system may incur a cost, the overall expense might be more than expected. It is also important for users to ensure their billing is accurate and no unexpected or unwanted services are running on the site.

Benefits of IaaS Solutions

- Reduces capital expenditures and outlays
- Can reduce overall cost of IT function
- Users only pay for the services needed
- Enterprise-grade IT resources and infrastructure are available even to small organizations
- Scalability and elasticity are very easy
- Users maintain control over their own application deployment if critical to their business model

SaaS versus PaaS versus IaaS: A Review

In general, cloud computing allows users to share overhead resources to achieve an economy of scale. In some ways, this is like a utility company which supplies electricity to a variety of customers. Everyone pays a portion of the upkeep expenses while experts ensure electricity is available and so forth.

All forms of cloud computing deliver services – software, databases, storage, servers, networking, and more over the Internet to users' organizations. The three primary forms of services are:

- ***Infrastructure as a Service (IaaS):*** *Hardware is supplied and managed by an external party.*
- ***Platform as a Service (PaaS):*** *Hardware and operating system are supplied and managed by an external party. The hardware functions are transparent to the user.*
- ***Software as a Service (SaaS):*** *Hardware, operating system, and applications are supplied and managed by an external party. Everything but the applications are transparent to the user.*

In a way, PaaS builds on IaaS because in addition to the hardware components, it provides and manages operating systems, middleware, and other runtime services. IaaS is the most flexible, but it also requires the user to have more expertise. PaaS simplifies deployment but can reduce the flexibility to customize IT environments. SaaS is even less flexible since it provides the entire infrastructure including user applications. A SaaS user just logs into an up-and-running system. While applications can be configured to some extent, overall, most IT is handled externally. Users may incorporate their business rules but that is about it.

A good analogy relates to car ownership. Gleb from RubyGarage.org breaks it down this way (Gleb 2020):

- *On-Premise Solutions are like owning a car*. You buy a car and are responsible for its maintenance and upkeep. Upgrading means buying a new car.
- *IaaS is like leasing a car*. You lease a car and choose what you want, drive it where you want but the car belongs to the lease company. If you need to upgrade, you lease a different auto.
- *PaaS is like taking a taxi*. You tell the driver where you want to go, and they get you there.
- *SaaS is like going by bus or train*. There is a fixed route, you share the ride and go where you need with everyone else.

Recovery as a Service (RaaS)

Although not in the same category as SaaS, PaaS, and IaaS, some cloud service vendors offer recovery as a service or RaaS. RaaS includes cloud services that facilitate backup, archives, disaster recovery, and business continuity functions. RaaS ensures an organization has data backed up in multiple locations and can quickly resume operations should a natural disaster or other unexpected event occur. In addition to data backups, RaaS may protect and help recover data centers, servers, middleware, databases, web sites, and many other IT resources. RaaS helps businesses reduce downtime and minimize negative impacts on their clients.

Benefits of RaaS Solutions

- Reduce downtime due to data loss or disasters
- Prevent loss of mission critical company data
- Safeguard IT infrastructure and ensure rapid redeployment of resources
- Give cost-effective backup, recovery, and business continuity planning
- Provide geographic-independence of mission-critical resources
- Offer flexibility and multiple backup options

What Are General Benefits of Cloud Services?

Now that we have defined cloud computing and gotten a high-level overview of ways cloud computing is deployed (see Figure 1.7), it is helpful to review and consolidate the advantages this approach to IT offers. Most apply to all forms of cloud computing.

- Cloud service deployment is fast. For many resources, it takes a very little time to configure an application. If the resource is already in place, it may only take a few moments to upsize its capacity or add more seats for new users.
- Cloud services can be accessed from nearly any computing device attached to the internet including smartphones, tablets, laptops, and desktops.
- Cloud services can be accessed from nearly any geographic location and whether someone is in their office, home, or on the road.
- Cloud services are elastic. As an organization grows or shrinks, subscriptions can be increased or decreased to match organizational needs. Cloud providers enable customers to select the level of service that matches their needs.
- Cloud services facilitate improved efficiency and cost reductions. Cloud services are ideal for small organizations, start-ups, temporary projects, and even large organizations wishing to economize.
- Cloud services provide expertise on IT infrastructure without needing in-house staff.
- Cloud services are billed to subscribers, so they pay only for what they use.

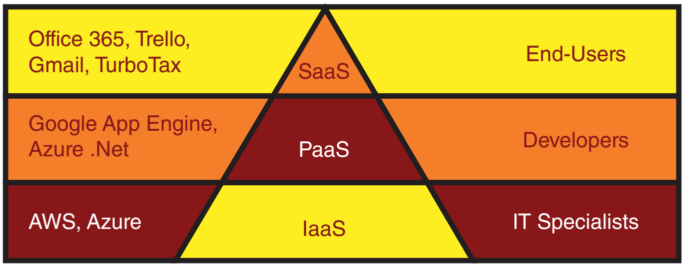
Office 365, Trello, Gmail, TurboTax
SaaS
End-Users
Google App Engine, Azure .Net
PaaS
Developers
AWS, Azure
IaaS
IT Specialists

Figure 1.7 Cloud computing and its users.

What Are General Disadvantages of Cloud Services?

Despite the advantages and compelling reasons to utilize cloud computing, several disadvantages do exist. The following points summarize several which will be explored in more detail throughout this book. Among these are:

- Network connectivity considerations. Cloud servers and data storage generally require network connectivity for resource access. If a network disruption occurs, service can be interrupted.
- Cloud security considerations. Whereas cloud computing generally provides high levels of secure access and data protection, an entirely different set of complex issues emerge. For instance, managing access privileges, protecting secure data, and many other topics become important due to cloud computing's nature.
- Cloud services can be costly. Although upfront investment costs may be lower, cloud computing operating costs can add up. Sometimes using a pay-as-you-go model can result in unexpectedly high billing. It is important to track and manage usage costs.
- Vulnerability to attacks. Cloud services provide an online target for hackers. The scale of damage can be vast if a secure system is compromised.
- Loss of control. In cloud environments, both software resources and data generally are entrusted to a third party. This means that corporate governance policies must be enacted by others which could result in compliance issues.
- Technical problems. If technical problems emerge, the fix might depend on cloud service providers. In organizations using multiple vendors or in those lacking an onsite IT staff, this may become a difficult issue to overcome.

What Is the History Behind Cloud Computing?

Cloud computing did not appear overnight as a sudden invention. Rather, it came about due to gradual changes in technology and the ways computers were used. A historic perspective makes this easier to see.

Mainframe computers provide a good starting point. In the early 1960s commercial mainframes began to appear. These early systems were gigantic, not too reliable, and used large amounts of power. The first systems usually were dedicated to single users and were developed for specific business, government, or scientific tasks. IBM dominated this marketplace and gained fame as the computing platform for businesses. Figure 1.8 shows a mainframe computer from this era.

In the 1970s mainframes become more user-friendly. Instead of solely batch-run tasks often dependent on tapes, punch cards or disks, user-terminals with keyboards and display devices appeared. These computers were time-shared and permitted multiple users. They were more powerful mainframes and used a virtual machine computing model. This meant individuals using the computer were allocated memory, which seemed like a dedicated machine, but was shared memory temporarily addressing their computing requests. Users accessed their virtual machines using thin-clients, usually green or amber screen terminals with a keyboard. The computing power resided on a centralized mainframe, but

Figure 1.8 Mainframe Era computer center. Source: National Center for Toxicological Research (NCTR) which is FDA's internationally recognized research center. (CC BY-SA 2.0). https://www.flickr.com/photos/fdaphotos/7421971854/in/album-72157630240761784/

it did not seem that way to the users, except when other peoples' processing slowed their operations.

In some ways, today's cloud environment resembles mainframe era computing. But, one key element is missing: distribution of users over even broader and larger areas. The cloud makes it possible to have thousands and tens of thousands (or even millions) of machines using the same resource pool. With mainframes, virtualization was driven by lack of computing resources. In the cloud, it is driven more by economies of scale.

The computing world did not jump directly from mainframes to the cloud. An intermediate, technology-driven step occurred. This was precipitated by the appearance of the personal computer. In the 1980s, computing power moved to the desktop. Instead of running tasks from a shared, centralized computing source, applications were developed to run on desktop computers. When these were connected into networks, thick-clients became the norm. Computing power existed both at the user's terminal and on servers where data and applications responded to requests from individual users.

This was known as the client-server era. Mainframes now became back office machines running large corporate transaction processing systems and performing other specialty roles. Although PCs and mainframes were connected by networks, many of these were custom developed and only worked internally to an organization. When networking capability was standardized, suddenly computers on different networks could interact and communicate. In the 1980s, TCP/IP became the protocol of what came to be called the Internet. In the early 1990s, the idea of creating an easy-to-use Internet interface emerged in the form of the World Wide Web (WWW). Hypertext Transfer Protocol (HTTP)

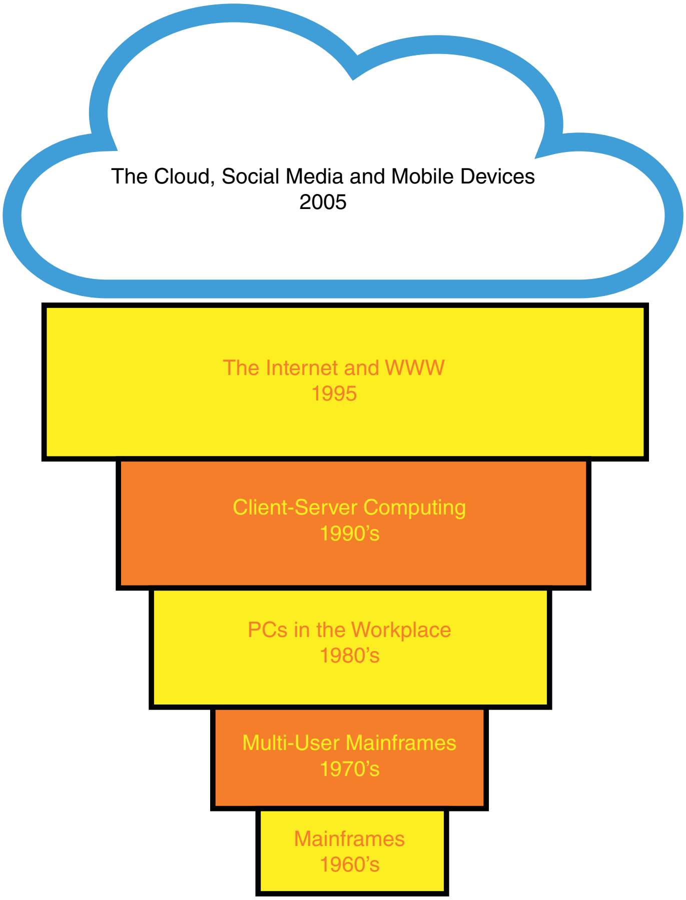

Figure 1.9 General computing eras.

permitted people to access materials stored on web servers without needing to understand the technology.

The WWW was ideal for communication and social media. This came to be known as Web 2.0. Technology companies became clever at distributing software and updates and created online platforms that shared databases among millions of online users. Facebook, Twitter, Instagram, and other social media platforms, enabled by mobile devices, opened computing to even more people.

As in the mainframe days, computing power moved back to shared resources and thin-clients once again became popular. New applications were cloud-based and around 2005, the era of cloud computing, for both individuals and business organizations truly began (see Figure 1.9).

Historic Perspective of Hardware Related to Cloud Computing

Many of the changes driving technology use at both business and individual levels have been brought about by hardware. For instance, the invention of the mainframe opened the door for business computing. The invention of the microprocessor then changed the way computers were built and impacted their sizes. Mobile devices and networking technologies allowed IT specialists to reinvent the way computing services were rolled out.

In general, amazing, and significant changes in hardware capabilities have made the current cloud environment possible. Among these changes were an increase in computing power and low-cost storage capability. For instance, in 1967, a megabyte of storage cost about one million dollars. Currently, that same amount of storage costs less than 2 cents. Similar cost reduction holds true with computing power. In 1985, the Cray-2 Supercomputer was the world's most powerful computer. When the iPad 2 came out in 2011, it offered more power and speed than the Cray-2 for a fraction of the cost.

Network technologies are another area of hardware that have enabled the cloud revolution. Both businesses and individual consumers have access to high-speed broadband. In urban areas, fiber optics make streaming 4K video and other applications possible. Even the most rural locations have satellite Internet capability which offers speeds reasonable enough to access cloud-based resources.

Currently, many cloud providers such as Google, Amazon, Microsoft, Cerner, and others are building and upgrading enormous data centers to power cloud computing. There is no end in sight to the race.

Historic Perspective of Software Related to Cloud Computing

Software has been as important as hardware in the development of cloud-based resources and services. Two major development areas have driven change from the software side. These are: virtualization and SOA.

Virtualization, which came from mainframe environments, commoditizes hardware by taking advantage of its capacity to serve multiple users. VMWare, now owned by Dell Technologies, developed techniques that permitted microprocessors to be virtualized. They found ways to partition and time slice multicore CPUs to enable large numbers of users to share resources. So, when one user was engaged in heavy processing, their application might receive extra resources while another user was demanding less CPU power. Economies of scale allowed better use of existing resources and permitted development of giant, shared server farms.

Operating system virtualization means that software permits hardware resources to run multiple operating system images simultaneously. Each of these virtual operating systems appear real to the user even though they are abstracted away from the actual hardware. In cloud environments, virtualization of servers using a software layer called a hypervisor emulates underlying hardware systems. The emulation may include the CPU's memory, input/output, and networking. Although virtual systems do not run quite as fast as those on actual hardware, for most practical applications, they are fast enough. Greater flexibility and hardware independence offered by virtualization is extremely important in cloud environments.

Another important concept in cloud computing is *Service Oriented Architecture (SOA).* SOA is a term used to describe the philosophy behind how large development projects can be organized through a divide and conquer and communicate approach. As such, SOA describes both an architectural style and vision of how an organization should approach developing, building, and deploying systems. The main mindset of SOA is to develop reusable services that can be integrated to create large scale systems. The days of building gigantic, specific applications have been replaced with creating a set of building blocks used to perform specific functions. SOA is an enterprise-level approach to development and is not meant for single application development.

SOA: A New Way of Doing Old Things

- Code developed in reusable modules that can "talk" to each other.
- Encapsulate design decisions that may change in the future (e.g. all changes can be made in a single, known location).
- Ensure code modules can be combined in different ways depending on specific software deployment needs.
- Draws from fundamental, object-oriented, software design principles.

In many regards, SOA has rebranded software development principles forming the foundation for software engineering practices. A few differences do make SOA unique from the design environments of the 1980s and 1990s. First, SOA is meant to work in networked environments. Software modules "talk" to each other and communicate over a network rather than through traditional procedure calls or parameter passing. SOA uses messages to request responses or provide updates. Another difference is in variable declarations and use. In most cases, SOA architecture does not rely on global variables for tying modules together. If global style variables do exist, the values reside in a database that has been configured to control concurrent access and ensure multiple clients can update or read the values without interfering with each other.

So, SOA can be described as consumer components receiving services from provider components using an infrastructure provided by cloud computing. It is helpful to use an analogy to describe a complicated topic. Many IT consultants use Legos to describe SOA. Here is a version of this analogy:

SOA Explained in Terms of Lego[1] Blocks

From our brief description of SOA, we know that this concept means that software services should be built in ways that are reusable, flexible, and independent. This is like Lego blocks. Here are a few characteristics:

- *Legos are interoperable*. Legos have standard bumps that can fit into any other Lego block. Having a standard interface means that even when new block styles are designed, they can still plug directly into the existing structure without an issue. In SOA, the bumps are the messages that link modules together.
- *Legos are composable*. A single Lego block may be interesting to look at, but it does not do much. Until a structure comprising multiple Legos is created, not much value exists.

[1] *LEGO® is a trademark of the* LEGO Group of companies.

- *Legos are reusable.* A person can build a structure with Legos, then later the same blocks can be reassembled to build something different. The blocks are not wasted, just reused. Some IT leaders like to extend the metaphor and say different colored blocks represent different services. For instance, yellow blocks might represent marketing services and blue blocks accounting services. This enables the composition of different organizational systems based on current business rules.
- *Legos are robust*. Although breaking individual blocks is possible, the blocks remain unlikely to break when structures are reassembled. That same feature applies to SOA modules. They should be developed, tested, and quality approved before ever being used in software development. SOA components must be designed for robustness.

So, what does the Lego analogy mean for businesses? It really represents what organizations want from their IT systems. They do not want to know what is inside a block, but they do want to ensure the blocks perform as needed. They must be interoperable, composable, reusable, and robust. The characteristics of Legos make them flexible and responsive. Software should be the same.

Of course, several downsides do exist in the Lego analogy. First, Lego structures are not glued together and can fall apart. When a Lego structure is complex, it becomes more difficult to manage. It is the same with software systems. Legos link well to other Legos but not to toys from other manufacturers. Legos come in many different sizes. Small blocks are more flexible and permit custom projects but take longer to assemble and require managing more pieces. Larger and specialty blocks make construction faster but offer less flexibility. It is the same with business services. There must be a balance between flexibility, reuse, and practicality. Granularity is a term used to describe the level of service of software components.

Harriet Fryman, the senior director of product marketing for Cognos, provided these seven principles of SOA during an interview with Loraine Lawson published on ITBusinessEdge.com (Lawson 2007).

1) Open and standards based.
2) Platform-neutral. Encapsulated so a service works identically on Linux or Windows.
3) Location-transparent. Should not change based on user or location in the global infrastructure.
4) Peer-to-peer. There should be no primary or secondary (Landau 2020). Every service is created equal. This means services can spawn and scale out without a single point of failure.
5) Loosely coupled. Can enhance capability within a service without impacting another service.
6) Interface-based. Each service is opaque regarding other services.
7) Coarsely grained. Must operate at a business level, not down in the bits and bytes. That makes it more reusable.

The Lego analogy is helpful but not a complete description of SOA. Remember, SOA architecture relies on components “talking” across networks and sending requested data back as a service. Legos physically link together. In either case, having robust, composable, reusable and interoperable services is a worthwhile goal for any organization's software services.

Use of SOA requires both technological and organizational mindset changes. While services become easier, other concerns, like security issues emerge. SOA is the logical extension of the browser-based standardization that was responsible for the emergence of the WWW. When these principles are used with machine-to-machine communication, standardization can emerge. SOA makes it possible to compose independent services into business applications, and this is particularly well suited to make SOA the de-facto architectural model for building virtualized applications running in cloud-based environments (Lawson 2007).

Summary

Chapter 1 has given us an overview of cloud computing and primary concepts that have enabled it to become the leading approach for rolling out enterprise IT resources. Cloud computing is an ideal solution space in the current mobile computing environment. Users have access to remotely hosted data, services, applications, storage, and systems. In the most general sense, cloud computing is a delivery system like a public utility. Clouds can be privately deployed with all resources hosted internally to an enterprise using cloud technologies (e.g. an intranet), publicly from large server farms meant to develop economies of scale by sharing overhead management costs, or in hybrid solutions that retain some data and services on premise while others are hosted off-site. See Table 1.1 for main cloud computing features.

Three main operational paradigms are common in cloud computing. These are:

- SaaS (software as a service)
- PaaS (platform as a service)
- IaaS (infrastructure as a service)

With SaaS, hardware, operating system, and applications are supplied and managed by an external party. Everything but the applications are transparent to the user. With PaaS, hardware and operating system are supplied and managed by an external party. The hardware functions are transparent to the user, but the users supply their own applications. Finally, with IaaS, hardware is supplied and managed by an external party, but the operating systems and applications are managed by the user organization.

Cloud computing delivers storage and applications as a service, rather than a product, offering both cost and business advantages. To do this, cloud computing moves services off-site to a hosting company, or a centralized self-hosted facility. Centralization accomplishes what IT managers seek: increased computing capabilities without having to provide new infrastructure. Cloud capabilities include rapid deployment, scalability, elasticity, access to IT expertise, and cost reductions.

Cloud computing uses the concepts of virtualization and SOA to create and manage resources. Virtualization leverages hardware capabilities by ensuring resources are dynamically allocated to those needing the capacity at the correct times. VM software partitions a physical computing device into multiple virtual devices, which can be used and managed to perform organizational computing tasks. Virtualization offers agility and can speed up IT operations and increase infrastructure utilization. SOA is a philosophy for software

Table 1.1 Cloud computing primary features.

Cloud computing feature	Description
Pooled resources	Available to users with subscription and level of access.
Virtualization	Improved utilization of hardware assets.
Elasticity	Dynamic scaling capability (e.g. scalability) and the ability to downsize when needed.
Automation	Build, configure, and deploy without human intervention required.
Metered billing	Per-usage business model. Pay only for use.

development and deployment where independent, reusable services are developed in ways that permit integration with large scale systems.

The possible uses of cloud computing are plentiful. Users interface with IT resources through their web browser, eliminating the need for installing numerous software applications. Organizational applications interface through cloud services to access resources and exchange information in cost effective and manageable ways.

References

Gleb, B. (2020). IaaS vs PaaS vs SaaS. https://rubygarage.org/blog/iaas-vs-paas-vs-saas (accessed 7 July 2020).

Landau, E. (2020). Tech confronts its use of the labels 'Master' and 'Slave'. https://www.wired.com/story/tech-confronts-use-labels-master-slave (accessed 9 July 2020).

Lawson, L. (2007). The merits and seven principles of SOA. https://www.itbusinessedge.com/cm/community/features/interviews/blog/the-merits-and-seven-principles-of-soa.

Bibliography

Alam, T. (2020). Cloud computing and its role in the information technology. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)* 1 (2): 108–115.

Armbrust, M., Fox, A., Griffith, R. et al. (2010). A view of cloud computing. *Communications of the ACM* 53 (4): 50–58.

Buyya, R., Broberg, J., and Goscinski, A.M. (2010). *Cloud Computing: Principles and Paradigms*. Wiley.

Hayes, B. (2008). Cloud computing. *Communications of the ACM* 51 (7): 9–11.

Hurwitz, J.S. and Kirsch, D. (2020). *Cloud Computing for Dummies*. Wiley.

Jamsa, K. (2012). *Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More*. Jones & Bartlett Publishers.

Mell, P. and Grance, T. (2011). The NIST definition of cloud computing: special Publication 800-145. Gaithersburg, MD. https://csrc.nist.gov/publications/detail/sp/800-145/final.

Rani, D. and Ranjan, R.K. (2014). A comparative study of SaaS, PaaS and IaaS in cloud computing. *International Journal of Advanced Research in Computer Science and Software Engineering* 4 (6).

Singh, A. et al. (2016). Overview of PaaS and SaaS and its application in cloud computing. In: *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, vol. 2016, 172–176.

Velte, T., Velte, A., and Elsenpeter, R. (2009). *Cloud Computing, a Practical Approach*. McGraw-Hill, Inc.