**The lecture 1**

**Cloud principles and delivery mechanisms**

The ***cloud*** is a network of servers, and each server in the network has a different function. Some servers run applications or deliver a service. By using the cloud, you don’t have to have the individual application or services running on the user’s computers. In addition, the cloud allows you to share resources and technology so that they can be accessed by multiple users. From the standpoint of the user, the cloud is a black box that the user accesses. However, the user is not concerned with what happens inside the black box. When choosing to use the cloud, you are shifting certain responsibilities to the cloud so that you can focus on other things—such as your business—and less on the underlying technologies.

Early data centers could consist of hundreds of physical servers, with each server being assigned a workload (such as specific application or service). Unfortunately, most of the resources on an individual sever were often wasted. Eventually, data centers started to consolidate many physical servers to a single server running multiple virtual machines or virtual servers using Microsoft Hyper-V or VMware ESX/ESXi. As a result, there was a significant increase in resource use while reducing overall cost and power consumption.

Cloud computing takes the next step; instead of virtualizing servers, it virtualizes datacenters. It uses a single resource pool containing an infrastructure that delivers infinite computer, network, and storage resources for important services. The cloud is more easily accessible to IT teams and has more accountability features that can be used to figure out cost center‐based chargeback billing.

The advantages in cloud computing are:

* **A virtualized datacenter:** Allows you to access computer services without regard to where, exactly, the data center is located and the hardware that the services are running on. However, you do want to select a data center that is in close proximity to users.
* **Reduced operational costs:** Similar to using virtual machines, cloud computing uses resources more efficiently. In addition, inconsistent availability and high operational costs are reduced by providing pooled resources, elasticity, and virtualization technology.
* **Datacenter/Server consolidation:** A virtual infrastructure helps consolidate servers by hosting multiple virtual machines on a virtualization host. Although the cloud uses a virtual infrastructure, the cloud goes one step further by helping consolidate data centers by moving servers from your current data center to the cloud. In fact, the cloud can also be used to expand current datacenters.
* You can consolidate servers by hosting multiple virtual machines on a virtualization host.
* **Improved resilience and agility:** With the correct applications, the cloud‐computing model improves resiliency and agility.

When looking at the cloud, you should understand the following terms:

* ***Communication as a Service (CaaS)*:** Allows the deployment of communications services through cloud computing without the need to purchase their own equipment. It can include Voice over IP (VoIP), VPN services, and business telephone service that you would find on a private branch exchange (PBX) such as phone menus and voice mails.
* ***Desktop as a Service (DaaS)*:** Provides a desktop or work environment to run applications, access emails, or back up data.
* ***Infrastructure as a Service (IaaS)*:** Provides the infrastructure that the cloud runs on, such as servers, switches, routers, storage area networks, firewalls, and other equipment.
* ***Monitoring as a Service (MaaS)*:** Allows you to monitor software applications so that the correct personnel are notified when it is down or not fully performing as needed.
* ***Network as a Service (NaaS)*:** Offers network services such as network infrastructure/IIAS and Communication services/CaaS.
* ***Platform as a Service (PaaS)*:** Allows you to buy, develop, test, deploy, and manage software applications so that users can access the applications.
* ***Software as a Service (SaaS)*:** Allows the development and provisioning of software for the user, including providing servers on which the software runs on. Typically, SaaS runs on demand through the remote desktop services or through a web browser. Often, the cloud provider owns the software licenses and charges a fee to subscribers.

The cloud can provide your organization with the following services:

* **Productivity services:** Allows users to work and collaborate. An example of productivity services is Office 365, which allows users to create and share documents.
* **Storage services:** Provides a storage platform for data. By storing data on the cloud, the data can be accessed by any user or device. An example of storage services is Azure Storage.
* **Communications services:** Provides communications between users. Examples of communication services include Exchange Online and Skype for Business Online. Exchange Online provides email, calendar, and contact sharing and Skype for Business Online provides instant messaging, computer‐to‐computer audio and video calls, and screen sharing.
* **Search services:** Provides search functionality into custom applications. In addition, it can provide a search engine and storage of data that can be accessed on an Application Programming Interface (API). An example of search services is Azure Search.

Based on the cloud solution that you select, you can also perform self‐service and provide multi‐tenancy. Self‐service provides the ability for an organization end‐user to acquire and manage servers, storage, or other resources without going through the IT operations teams. Users can go to their portals and add more licenses, renew contracts, and reduce the numbers.

***Multi‐tenancy*** allows several companies to use the same cloud products. For example, for an application, multi‐tenancy means that different customers can use the same codebase. However, to keep customers separated from other customers, their configuration and data are stored in separate containers. Multi‐tenancy offers the following benefits:

* Lower costs through economies of scale
* A shared infrastructure or servers lowers costs
* End users don't have to pay costly maintenance fees to perform ongoing maintenance and updates
* Configuring can be done without touching the underlying codebase

Recently, Microsoft developed the ***Enterprise Mobility Suite (EMS)***, which is a comprehensive suite of cloud services that addresses the use of and managing of mobile devices in a corporate environment, including corporate devices and personal devices used on corporate networks. The EMS consists of the following cloud components:

* **Microsoft Azure Active Directory (Azure AD/AAD):** Delivers access management from the cloud and existing on-premises deployment.
* **Microsoft Intune:** Provides advanced device management, including management of Windows, Windows Phones, IOS devices, and Androids. It allows the deployment of policies and software and provides inventory of hardware and software.
* **Microsoft Azure Rights Management (Azure RMS):** Provides protection for company assets with security, compliance, and regulatory requirements by using encryption, identity, and authorization to secure files and email.

In a traditional, on‐premises data center, you will need to pay for the following:

* **Server costs:** All hardware components and the cost of hardware support. Of course, when purchasing servers, don’t forget to design fault tolerance and redundancy, such as clustering of servers, redundant power supplies, and uninterruptable power supplies.
* **Storage costs:** All hardware components and the cost of hardware support. Based on the application and level of fault tolerance, centralized storage can be very expensive. For larger organizations, you can create tiers of storage where more expensive fault‐tolerant storage is used for critical applications and lower priorities use a cheaper form of storage.
* **Network costs:** All hardware components, including cabling, switches, access points, and routers. It also includes WAN connections and Internet connections.
* **Backup and archive costs:** The cost to back up, copy, or archive data to the cloud or data center. Options might include backing up to the cloud or backing up from the cloud.
* **Business continuity and disaster recovery costs:** Along with server fault tolerance and redundancy, you have to think about how to recover from disaster and continue operating should the worst scenario occur. This should consist of creating a data recovery (DR) site. It could also include backup generators.
* **Data center infrastructure costs:** Costs for electricity, floor space, cooling, and building maintenance.
* **Technical personnel:** Based on the technology used, you will need technical expertise and manpower to install, deploy, and manage the systems at the data center.

When using the cloud, many of these costs are shifted to the cloud provider. However, you need to ensure that you have enough bandwidth available for users to connect to the cloud and use the required applications. If you are connecting a data center to the cloud or connecting two clouds together, you have to see how much data needs to be transferred so that you can determine the bandwidth needed. Don’t forget to plan for backup traffic to or from the cloud and replication between data centers or the cloud for data‐recovery purposes.

The ***subscription*** or ***pay‐as‐you‐go*** model is a computing billing method that is aimed at organizations and end‐users. The organization or user is billed for the services used, typically on a recurring basis. You can scale, customize, and provision computing resources, including software, storage and development platforms. For example, when using a dedicated cloud service, you could pay based on server power and usage. When using software on a SaaS—­covered later in this lesson—you lease the software and customized features.

When you use the pay‐as‐you‐go mode, you have to actively manage your subscriptions. You must ensure that users do not misuse the cloud; make sure accounts that are provisioned are actually being used and not wasted. When resources are being provisioned by the provider, billing starts. It is the responsibility of the client to deprovision the resources when they are not in use, so that you can manage costs.

***Capital Expenditures (CapEx)*** are funds used by an organization to acquire or upgrade physical assets, such as servers, networking equipment, and storage. It also includes real estate such as buildings or data center space. Typically, the physical resources are amortized over several years, whereby instead of deducting the full cost of the equipment in the first year, you deduct a smaller portion of it each year.

***Operating Expenses (OpEx)*** are the expenditures that an organization incurs while performing its normal business operations, including the amount of electricity consumed, the cost of employees to manage and support systems, office space, and Internet connections. Management is responsible for minimizing operating expenses without significantly affecting the firm’s operations and its ability to compete in the marketplace. OpEx is expensed each year because you pay for and use the product or service.

When a server needs to be replaced, or a server needs to be added to a data center, you need to use CapEx to pay for the computer. It will affect immediate cash flow because you have to pay for the server up front. Fortunately, however, you can amortize the cost over several years. The expense of running the server and the staff to run the server is an OpEx.

If you lease a server or use the cloud, the cost is based on the pay‐as‐you‐go model. For accounting purposes, the costs are considered an OpEx.

Most organizations will face growth. If you are purchasing a server, you would have to purchase a server that can handle the current demand as well as the demand determined by growth over the next three to five years. The current demand could be based on:

* Total number of log‐ins per hour
* Page response time
* Transaction and process completion time
* Initial load time
* Amount of website traffic and user load based on the average and maximum number of users, peak load, and maximum number of transactions per second (TPS)
* Seasonal trends

When planning capacity, you will must consider memory, CPU (speed and number of core), disks (speed and capacity), and databases (response times and capacity). You will need to perform the following steps:

1. Conduct a demand analysis.
2. Conduct a current capacity analysis.
3. Conduct future capacity planning.

Demand analysis is used to gather all information about the current demand, workload, and trends from all aspects of the infrastructure. Current capacity analysis establishes the threshold and benchmark values so that you can determine when a resource is over utilized or underutilized.

***Scalability*** is the ability of a computer application or product to continue to function as the application or product changes in size or volume in order to meet user need. Since the cloud is based on virtual technology, the ability to scale on demand is the biggest advantage of cloud computers. That could be as simple as increasing or reducing the amount of memory or number of CPU cores or adding another server to a cluster.

***Elasticity*** is the degree in which a system can adapt to workload changes by provisioning or deprovisioning resources automatically. Of course, elastic computing is the dynamic provisioning and deprovisioning of computer resources to meet the varying workload. By offering elasticity, you can increase and decrease cost, quality, and resources.

When a server or service goes down, it most likely causes your organization to lose money. If your network contains an external website or database that controls your sales, ordering, inventory, or production, server downtime can be detrimental to these business needs. If it is an internal server or service, it might not allow your users to perform their jobs. In either case, your company sustains losses in revenue or productivity—and, in some cases, both.

For any service, you need to minimize downtime by identifying potential failures and then taking steps to avoid those failures and to reduce their effects. ***High availability*** is a combination of technology, protocols, and redundant hardware that ensures a certain degree of operational continuity during a given measurement period while resisting disaster and failure. Generally, the term ***downtime*** is used to refer to periods when a system is unavailable. Availability is usually expressed as a percentage of uptime in a given year, as shown in Table 1-1.

|  |  |  |
| --- | --- | --- |
| **Availability %** | **Downtime Per Year** | **Downtime Per Month** |
| 99% ("two nines") | 3.65 days | 7.20 hours |
| 99.9% ("three nines") | 8.76 hours | 43.8 minutes |
| 99.99% ("four nines") | 52.6 minutes | 4.32 minutes |
| 99.999% ("five nines") | 5.26 minutes | 25.9 seconds |
| 99.9999% ("six nines") | 31.5 seconds | 2.59 seconds |

**Table 1-1**

**Availability Guidelines**

When designing servers and the services they provide, they are often assigned ***service‐level agreements (SLAs)***, which state the level of availability those servers or services must maintain. Of course, a server design that can support five or six nines is much more expensive than supporting an availability of 99%. Of course, six nines is going to cost a lot more than two or even three nines.

If there is miscommunication about service‐level expectations between the hosting company and an organization, it could lead to poor business decisions, unsuitable investments, unsuitable service levels, or customer dissatisfaction. Therefore, you need to express availability requirements clearly so that there are no misunderstandings about the implications.

Typically, to make a server more fault tolerant, you should first consider what components are most likely to fail and then implement technology to make a system less likely to fail. Some of the components that are made redundant within a system are usually the following:

* **Disks:** Use some form of RAID and hot spares.
* **Power supplies:** Use redundant power supplies.
* **Network cards:** Use redundant network cards.

Although you can make these components fault tolerant, the entire server still won't be fault tolerant. Instead, you can use a cluster to provide server redundancy. By using the cloud, you don’t have to worry about these details because these details will be handled by the hosting company. When using the cloud, you will need to look at the provided SLAs.

While a hosting company provides six nines for the services it provides, unforeseen disasters could still hinder the availability of those services. Therefore, to provide business continuity in these cases, you should use cloud services that can be accessed from two or more sites. Therefore, if a disaster or mishap impacts the delivery of your agreed services, the necessary services can be provided by the other site or sites.

The cloud can also be used as a DR site for an organization. For example, without establishing a physical DR site, you can deploy backup servers to the cloud. If the primary data center goes down, you can ramp up or switch to the DR site to provide the necessary services.

Public cloud services provide a way to access information from anywhere at any time. Microsoft defines a ***public cloud*** as a web‐based service that is hosted outside of your organization. This means the information technology infrastructure (hardware, servers, software, and so on) is located somewhere other than your office and is managed by a third party (such as when it is hosted). If you use mobile banking—accessing web‐based email or storing your photos online in one of the many services provided—you are interacting with "the cloud."

With the public cloud, you pay only for the resources you consume. For example, with IaaS, you can increase the number of processors, the amount of memory, the amount of network throughput, or the amount of data transfer. With SaaS, you pay for the number of licenses based on the number of users who need to use the services. In addition, the public cloud offers quick deployment, rapid capacity scaling, and all services are delivered with consistent availability, resiliency, security, and manageability.

Using public cloud services such as OneDrive and Office 365 enable you to take advantage of hosted solutions. This means users have the ability to access their information from anywhere at any time across multiple devices. By using cloud‐based services, users can collaborate via calendars, email, and through document sharing. From an administrative perspective, it means you gain access to services and programs without the additional overhead of maintenance and software upgrades.

The public cloud can be organized according to:

* **The *shared public cloud*:** As shown in Figure 1-1, the shared public cloud is used by multiple organizations and is hosted on an infrastructure, where the architecture, customization, and some of the security are designed and managed by the provider.



**Figure 1-1**

**A shared public cloud is used by multiple organizations**

* **The *dedicated public cloud*:** It’s similar to the shared public cloud, but the cloud is delivered on a dedicated physical infrastructure (see Figure 1-2). While cost might be higher than that of the shared public cloud, the dedicated public cloud might offer better security, performance, and customization.



**Figure 1-2**

**A dedicated public cloud is used by one organization**

A ***private cloud*** offers the same features and benefits of a public cloud, but is contained within the corporate network and is controlled by the corporate IT department. The private cloud offers more security and privacy and allows for more control of its resources and data. Newer versions of Microsoft Hyper‐V and VMware ESXi support creating and maintaining a private cloud.

A ***hosted private cloud*** is provided by a third‐party company but is accessible only by users within a specific organization. In other words, the resources are leased or owned by the organization but are managed and located away from the organization. However, the hosted private cloud tends to be cheaper than the private cloud because some of the cost for the building, power, and personnel is distributed across several companies.

A ***hybrid cloud*** is a cloud‐computing environment in which the organization provides and manages some of the resources in‐house and a third party provides the hosted servers for the organization externally. In other words, it utilizes both a public cloud and a private cloud.

When you decide to use the cloud, you need to make sure that your use of the cloud is secure. This includes establishing privacy and compliance and determining how your data is secured.

As more and more customers are relying on online service providers to keep their data safe from loss, theft, or misuse by third parties, other customers, or even the provider's employees, cloud services raise unique privacy questions for businesses. Organizations have legal obligations to ensure the privacy of their employees, customers, and clients.

Laws prohibit some data from being used for a reason other than the purpose for which the data was originally collected. In addition, when you collect and store data in the cloud, you are subject to legal requirements, such as the Health Insurance Portability and Accountability Act (HIPAA) or the Gramm‐Leach‐Bliley Act (GLBA), just as if you were storing data on premise. If you deal with European companies or customers, you will must adhere to EU privacy laws.

Before using the cloud, you should always read the privacy notices that specify how data is accessed by users and how it can be deleted or modified. In addition, you need to know where data is actually kept, how data is backed up, how often data is backed up, and where the backups are stored. In some instances, you might have data that cannot leave the country that it is intended for or cross the borders of other countries.

As previously mentioned, you need to consider how the hosting company handles disaster recovery and business continuity. You must ensure that backups are being done on a regular basis, data is being replicated to another site, and that the services are duplicated on another site.

You also must consider how the hosting company handles security breaches. You should read the cloud provider's disclosure policy and how quickly they will disclose the breach to you. In addition, there are laws that require you to ensure that you are informed promptly of any breaches.

An organization that does not protect its data may be subject to a fine by one or more government or industry regulatory bodies. Some of these fines can be substantial, crippling a small or mid‐sized business.

Laws or regulations typically specify who within an organization should be held responsible for data accuracy and security. For example, the Sarbanes–Oxley Act designates the CFO and CEO as having joint responsibility for the financial data, while the Gramm–Leach–Bliley Act specifies the responsibility for security lies within the entire board of directors. The United States Federal Trade Commission (FTC) requires a specific individual to be accountable for the information security program within a company.

If you store any of your data in the cloud, you must ensure that the cloud service provider follows all legal and regulatory requirements. Although the cloud is hosted by another company, it is still your responsibility to ensure these requirements are met. Before you sign any contract, you need to evaluate the specific needs and requirements. Then after the contract is signed, you need to take steps to ensure that compliance is maintained.

***Encryption*** is the process of converting data into a format that cannot be read by another user. Once a user has encrypted a file, it automatically remains encrypted when the file is stored on disk. ***Decryption*** is the process of converting data from encrypted format back to its original format. To help protect files on a computer, you can use encryption.

***Symmetric encryption*** uses a single key to encrypt and decrypt data. Therefore, it is also referred to as secret‐key, single‐key, shared‐key, and private‐key encryption. To use symmetric key algorithms, you need to initially send or provide the secret key to both the sender and the receiver.

***Asymmetric key***, also known as ***public‐key cryptography***, uses two mathematically related keys. One key is used to encrypt the data and the second key is used to decrypt the data. Unlike symmetric key algorithms, an asymmetric key does not require a secure initial exchange of one or more secret keys to both the sender and the receiver. Instead, you can make the public key known to anyone and use the other key to encrypt or decrypt the data. The public key can be sent to someone or it can be published within a digital certificate via a certificate authority (CA). Secure Socket Layer (SSL)/Transport Layer Security (TLS) and Pretty Good Privacy (PGP) use asymmetric keys.

For example, consider a scenario in which you want a partner to send you data. You send the partner the public key and the partner then encrypts the data with the key and sends you the encrypted message. You then use the private key to decrypt the message. If the public key is used by another user, that user still cannot decrypt the message because the user does not have the private key.

For data that is at rest (sitting on a disk somewhere on the cloud), you should encrypt the disks or files on the disks. If the system is running Windows, you can use EFS to encrypt individual files and folders or you can use BitLocker to encrypt an entire volume. By encrypting the data, you will limit access to only those who have the correct keys to unlock the files. It will also help protect the data if the system is compromised or if the system is being disposed of and the disks will no longer be needed.

For the encryption to be effective, however, you need to use longer keys (2,048 bits minimum, but 4,096 bits would be better). You should also keep your own keys and store the keys off the cloud provider’s premises. Therefore, if the cloud systems are compromised, they will still not have access to the keys. In addition, by using encryption and not providing the keys to the cloud, providers will prevent the cloud provider from accessing the data. Lastly, a cloud provider managing your keys could be compelled to give up your key—if ordered by a government body, court, or law enforcement agency without your knowledge.

When you browse the web, there are times when you need to transmit important data (such as credit card numbers, social security numbers, and so on). You should do so using SSL over http (https), which encrypts the data. By convention, URLs that require an SSL connection start with https instead of http.

***Secure Socket Layer (SSL)*** uses a cryptographic system that uses two keys to encrypt data: a public key known to everyone and a private or secret key known only to the recipient of the message. The public key is published in a digital certificate, which also confirms the identity of the web server.

When you connect to a site that is secured using SSL using Internet Explorer 11, clicking the lock icon displays more information about the site, including the identity of the CA that issued the certificate (see Figure 1-3). For even more information, you can click the View Certificates link to open the Certificate dialog box.



**Figure 1-3**

**Viewing the SSL website identification in Internet Explorer 11**

When visiting certain websites, Internet Explorer might find problems with the digital certificate (such as that the certificate has expired, it is corrupted, it has been revoked, or it does not match the name of the website). When this happens, Internet Explorer blocks access to the site and displays a warning stating that there is a problem with the certificate. You then have a chance to close the browser window or ignore the warning and continue on to the site. Of course, if you chose to ignore the warning, make sure you trust the website and you believe that you are communicating with the correct server.

***Transport Layer Security (TLS)*** is an extension of SSL, which is supported by the Internet Engineering Task Force (IETF) so that it could be an open, community‐supported standard that could then be expanded with other Internet standards. While TLS is often referred to as SSL 3.0, it does not interoperate with SSL. While TLS is usually default for most browsers, it has a downgrade feature that allows SSL 3.0 to run as needed.

***Virtual Private Network (VPN)*** is a technology that uses encrypted tunnels to create secure connections across public networks like the Internet. There are a variety of uses for this technology, but three of the most common are shown in Figure 1-4.



**Figure 1-4**

**Uses for VPN technology**

VPNs are commonly used by remote employees for access to internal networks; VPNs create secure network‐to‐network connections for branch offices or business partner connections. A fourth use is to create secure host‐to‐host connections for additional security and isolation on an internal network. VPNs utilize encryption and authentication to provide confidentiality, integrity, and privacy protection for data. VPN tunnels can be used to connect an organization's data center to the cloud, or to connect one cloud to another.

Of course, when looking at security, you still need to follow the basics. Therefore, you need to consider the following:

* Perimeter security
* Hypervisor security
* Storage security
* Configuration and change management

This includes assessing the security of firewalls and the security of the operating system running the virtual servers. You also need to look at what can connect to a storage logical unit number (LUN), which are the storage units that are used by the virtual machines to create virtual disks. If a program or service has direct access to the LUNs, the program or service also has access to the data stored on the LUNs. Lastly, you need to ensure that you control and review changes to the system by using configuration and change management so that changes do not negatively affect the services.

While you have encryption to help protect the data, you need to establish an audit policy and enable auditing so that when data (individual files and database records) is accessed, a record is created in the logs. You can then review the logs whenever necessary to determine whether data was accessed improperly or without your knowledge.

Lastly, you should analyze data access patterns, looking for anomalies. For example, if you have large amounts of data being pulled at night or weekends (when data traffic should be light), it might indicate that someone has compromised your system and is downloading content. In addition, you should look for network packets that don't belong; it might indicate that someone has accessed your system or is attempting to access your system.

When dealing with transparency, you should assess the following:

* Where does your data actually reside?
* Who has access to your data?
* Do you have visibility to the actual availability of your servers and services, including where there are changes to your service?
* How is the uptime guaranteed?

As part of assessing transparency, you need to know where your data is located. If you have confidential or intellectual property, you may require that data remains within the country. You should also check to see how backups are performed and where those backups are located.

As you recall, a public cloud is used to service multiple customers who are referred to as tenants. Therefore, if the cloud provider has a physical server with 10 different virtual servers for 10 different customers, you need to ensure that one customer cannot access other virtual machines and the data that resides on those virtual machines.

You also need to know if the data is accessible to the cloud provider's employees. For example, the provider might encrypt all data in its cloud environment, which ensures that its employees cannot access the data. Of course, you should also be concerned with how the employees of the provider are screened.

With any cloud technology, you need insurance that the server and services will be available at least 99.9% of the time. Therefore, you should see what tools the provider provides to its tenants to monitor the servers and services on the cloud. One such tool is NewRelic, which can show performance over a period of time (which can be valuable in diagnosing performance issues). To help ensure availability, you might need to use your monitoring tools so that you can be immediately notified when a server or service is not available.

The cloud provider should be transparent regarding service disruptions or poor performance and you must have clear SLAs, including an SLA for availability/uptime, application response time, application throughput, incident‐response time, and problem‐resolution time. The cloud provider should also provide access to logging and root‐cause analysis of service problems at no additional cost. The provider should also have an easy method by which it can escalate issues when problems are found.

**Differentiating Between the Types of Cloud Services and Their Characteristics**

The three primary types of cloud computing are SaaS, PaaS, and IaaS, which are made possible through virtualization. While SaaS is geared toward the end users of your organization, IaaS provides full access to virtual servers without having to maintain the equipment that they are running. PaaS lies between SaaS and IaaS.



**Figure 1-5**

**Comparing SaaS, PaaS, and IaaS**

Software as a Service (SaaS) is the most basic form of cloud computing. It provides software and associated data (such as databases or user documents) that is hosted on the cloud. It is typically accessed by an application (such as a web browser) over the Internet. The users do not manage or control the cloud infrastructure, including network, servers, or storage. They may not even control the individual application's capabilities unless you are using a provider‐defined, user‐specific application. SaaS is sometimes referred to as on‐demand software. The cost of SaaS is typically based on a usage‐based model, where the consumer pays an agreed amount based on the use of the service or a monthly or yearly flat fee per user. Examples of SaaS include Hotmail, Gmail, Office 365, and Dropbox.

Platform as a Service (PaaS) goes one step further than SaaS. Instead of providing the applications, it provides the computing platform (such as networks, servers, and storage) on which the organization can host its own applications. PaaS allows the organization to focus on the developing and improving of the application without worrying about the infrastructure that it runs on. The two primary program languages for PaaS are Java and .NET. The cost of PaaS is typically based on usage of the platform, operating costs of the platform, and the agreed SLA. Examples of PaaS include Microsoft Azure, Amazon Web Services, and Google App Engine.

The Infrastructure as a Service (IaaS) service model provides the hardware for servers, storage, and networking—usually in the form of a standardized virtual server environment—and provides a foundation for PaaS and SaaS. The organization manages the guest operating system, the software, and the database application/servers. Cost is based on an agreed service level. While the hosting company mitigates the risks to the infrastructure, the organization assumes the responsibility for uptime of the cloud. Examples of IaaS include Amazon EC2, Microsoft Azure, Rackspace, and Google Compute Engine.

|  |
| --- |
| A hybrid cloud consists of a private cloud and a public cloud bound together. With the hybrid cloud, you can connect collocations such as an organization data center (private cloud) with a public cloud or by combining two public clouds. For example, you can store sensitive client data on the organization data center but host applications on a public cloud that will access the data center. Adopting the hybrid cloud requires consideration of a number of factors, such as data security and compliance requirements, level of control needed over data, and the application an organization uses. |

Based on your requirements, the cloud can meet the following requirements:

* It can connect a private data center to a public or private cloud environment. By using a connection between the two, you can extend or scale the private data center into the cloud by moving specific workloads from the data center to the cloud (such as when demand for resources suddenly spikes).
* It can connect resources between clouds. By signing up for a SaaS to provide one service, you might need access to resources in another SaaS. For example, you can use customer relationship management product needs to connect to a human resource system or a back‐office accounting system.
* It can implement a service that is hosted or controlled by a partner, such as when an organization needs to connect to a partner that is also using cloud‐based service.
* Although the public clouds offer scalability, you might choose to use the private cloud because you want greater management and control.
* You might want to perform development and testing on one public cloud because of the support and options that it offers, but deploy it on a less expensive public cloud or a public cloud that is more reliable.
* It provides a solution when an organization does not want to become too reliant on one vendor or cloud provider. Therefore, you use another cloud provider to provide service redundancy.

When you are determining whether to implement a public or private cloud, you need to understand the cost associated with changing from the data center to the cloud:

* **Management:** By extending a data center to the cloud, you need to manage multiple environments: the on‐premises data center and the cloud.
* **Data transfer:** The costs to transfer data to or from the cloud. If you have large amounts of data, you might need higher bandwidth to the cloud. In addition, some cloud services might charge based on bandwidth usage.
* **Customization and integration costs:** You might need to pay for the customization of an application so that it can work in the hybrid environment. Some of the applications might need to be rewritten.
* **Storage costs:** You will need to consider long‐term storage costs needed locally and on the cloud.
* **Platform costs:** The cost of licensing for middleware or for software that provides services to software applications. Middleware may include web servers, application servers, and content management severs. You also need to consider how information will be accessed, such as accessing through Extensible Markup Language (XML), Simple Object Access Protocol (SOAP), Web services, SOA, Web 2.0 infrastructure, and Lightweight Directory Access Protocol (LDAP).
* **Software maintenance costs:** The costs for licensing and support of software. Some licensing may be under a global usage deal while others may require a second set of licenses for systems on the cloud.
* **Compliance costs:** The costs to audit cloud services, including checking security and recovery procedures.
* **Server costs:** When you expand to the cloud, you might not significantly reduce the number of servers in your organization’s data center (unless you move several servers to the cloud). However, while you might not reduce the number of servers, expanding to the cloud might allow your organization to avoid purchasing new servers for the data center.
* **Data center infrastructure costs:** When you remove a few servers, you still will not reduce the cost of the floor space used on the data center and will most likely not reduce costs for electricity and cooling.
* **Operational support personnel costs:** Costs can be reduced when you no longer have to pay for personnel because those duties are now being handled by the cloud provider.

To assist in calculating costs, some cloud providers will provide cost calculators to estimate charges and to calculate the total cost of ownership (TCO). Before moving to the cloud, you should determine whether you will actually save money by replacing your existing system or save money by not upgrading or adding to your current systems.

When you integrate a data center with the cloud or from one cloud to another, you might need to use:

* **Cloud‐based tools:** Cloud‐based tools might provide an application or web page that allows you to connect specific applications. For example, you might need database connectivity or you might need to transform data to or from a database or from one database to another.
* **Cloud‐based solutions:** Cloud‐based solutions can be used for data replication or to copy data from one source to another. Cloud‐based solutions can also be used for backups to and from the cloud.

When you are considering a move to the cloud, you should follow these steps:

1. Assess your current IT strategy.
2. Consider your future technology needs.
3. Explore the different cloud computing options, best practices, and vendors.
4. Create a hybrid cloud strategy plan.
5. Plan for implementation.

When you evaluate your current IT strategy, you need to review the servers and services you are currently using and their respective loads. You also need to consider current limitations. You also need to evaluate your recent and current growth trends and assess where your technology needs over the next 3 to 5 years. By knowing where you are today and where you are headed tomorrow, you will be able to determine whether your current system is a good investment.

The next task is to research what is available to you. Evaluate several vendors so that you can compare options, costs, SLAs, and reputations. You should then be ready to design and implement a hybrid cloud strategy.

When working on these steps, be sure to include your business objectives. You should create a task force that includes your organization's IT and management leaders. You should also consider your ultimate objectives with cloud technology and how you will measure the services hosted in the cloud. Lastly, don't rush into the cloud—it's an important decision that must be evaluated and implemented with the greatest of care.