Introduction to the Cloud

From the server room to the board room, there is a lot of talk about the cloud and for good reason. The cloud offers businesses and their information technology staffs several important benefits ranging from increased efficiencies to scalability. Even among IT professionals, however, there's confusion as to what exactly the cloud is.

That is largely because the cloud represents a large ecosystem of diverse models and technologies that are quickly changing as the industry advances. People are making up vocabulary along the way, and every cloud service provider has its own definition of the cloud.

One way to understand the cloud is to start with the definition outlined by the National Institute of Standards and Technology. It attempts to encompass all the various cloud approaches and provides a baseline understanding of the concepts and components involved.

A definition of the cloud

The cloud is not a place or a technology. It is a usage model or philosophy that provides a way to transform business practices.

In Cloud Computing Synopsis and Recommendations (Special Publication 800-146), NIST defines the "cloud" (also known as cloud computing) as:

"... a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (for example, networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[1]

Yet another way to define the cloud, which provides more room for its continuing evolution, is as Shaw Data Centre & Cloud Solutions does: a philosophy of design and execution in which computing resources such as server, network and storage converge into a single infrastructure that is fully managed and delivered as a service.

From an IT infrastructure perspective, the cloud is like shared hosting. Computing resources such as servers, storage and networks are shared across multiple tenants and over a remote connection. However, both the NIST and Shaw Data Centre & Cloud Solutions definitions call out an important difference. The cloud is not a place or a technology. It is a usage model or philosophy that provides a way to transform business practices. Rather than making an up-front investment in IT infrastructure,

Call: 1-866-244-7474
Visit: business.shaw.ca/cloud

Follow us:

Shaw) Data Centre & Cloud Solutions
Powered by Flexential

1 National Institute of Standards and Technology http://csrc.nist.gov/publications/nistpubs/800-146/sp800-146.pdf

a company can draw on IT services that are delivered as a service over a third-party managed network, typically the Internet. The company is charged based on what it uses (like a utility), on a subscription basis or on a reservation system, which ensures access to resources when needed. The costs shift from capital expenditures to operating expenditures.

Cloud characteristics

The cloud can be further defined by specific characteristics. NIST lists out five it considers to be essential, but all these attributes may not be included in every cloud solution. They include:

On-demand self-service. On-demand self-service combines a service catalog and self-service portal to provide users with access, when needed, to a set of standardized cloud services and options. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Broad network access. Capabilities are available over a network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g. mobile phones, tablets, laptops and workstations). The network can be public (the internet) or private (dedicated lines or multiprotocol label switching, infrastructure, for example). Response-time guarantees depend upon this connectivity.

Some cloud vendors offer dedicated links to their data centers and provide appropriate service level agreements for uptime or response time. Others might implement a best-effort guarantee for uptime or response time and provide tools for monitoring application performance and response time, so that users can plan their bandwidth needs.

Resource pooling. The cloud services provider's computing resources are pooled to serve multiple consumers using a multi-tenant model with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state or data center). Examples of resources include storage, processing, memory, and network bandwidth.

Rapid elasticity. Capabilities can be quickly and elastically provisioned, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth and active user accounts). Resource usage can be monitored, controlled and reported, providing transparency for both the provider and consumer of the utilized service.

The virtualization factor

In the IT world, virtualization is the creation of a virtual machine in which to execute workloads.

Another characteristic commonly associated with the cloud is "virtualization." In the IT world, virtualization is the creation of a virtual machine in which to execute workloads. Virtualization software is used to run multiple virtual machines on a single physical server to provide the same functions as multiple physical machines. Known as a hypervisor, this virtualization software performs the abstraction of the hardware to the individual virtual machines. It slips between the physical machine hardware and a guest operating system and gives the illusion that it is

running directly on the hardware. The hypervisor can create this illusion for multiple guests at the same time, allowing multiple virtual machines to share the same physical hardware. Management software can operate in concert with the hypervisor to provide advanced features such as virtual machine migration, load balancing or disaster recovery.

Cloud environments can leverage virtualization as the basis for running services. The consolidation provided by virtualization, along with provisioning automation, creates a high level of utilization and reuse, ultimately yielding a very effective use of capital equipment.

Cloud deployment models

Public cloud enables consumers to pay only for what they use. A hybrid cloud combines two or more cloud types (private, community or public) that remain distinct but share technologies that enable porting of data and applications from one to the other.

In addition to possessing specific characteristics, the cloud is classified by how it is deployed. The most common classifications are public, private, community and hybrid. The different deployment models each present several advantages and disadvantages associated with the scale, cost and availability of cloud resources.

Public cloud. In a public cloud, sometimes called a multitenant cloud, the infrastructure is made available to the public or a large group. It may be owned, operated and managed by a business that provides cloud services, an academic organization, a government organization or by some combination of these organizations. A public cloud exists on the premises of the organization providing the cloud services.

Often referred to as utility computing, a public cloud enables consumers to use computing

resources as they would a utility. They pay only for what they use, and get the impression of unlimited capacity, available on demand. They trust the cloud provider will maintain the necessary infrastructure to run their applications and provide the requested service at their required service level agreement.

Private cloud. Also referred to as an internal or single tenant cloud, a private cloud is deployed for the exclusive use of a single organization. It is accessed from behind the organization's existing firewall. A private cloud may be owned, managed and operated by the organization, a third party or both. It can also exist on premise or off premise. Some companies offer private clouds that are in a multi-tenant environment but employ logical segmentation.

When a third party manages it, it is called a managed private cloud. When it is hosted and operated off premises, it is called a hosted private cloud. Among the reasons companies may choose a private cloud are:

• To leverage existing hardware investments. Consolidated IT resources, automation, self-service and better integrated management tools also reduce total costs and operating expenses.

• Concerns about data security and issues of trust if multiple organizations were sharing the same resources.

• A preference or need for exclusive use of hardware to handle specific workloads or to obtain higher availability of systems and applications during specific times.

Community cloud. The infrastructure of a community cloud is shared by several organizations and supports a specific community that has shared concerns, such as security requirements or policy and compliance considerations. It may be owned, operated and managed by one or more of the organizations, a

third party and/or some combination of them. Like a private cloud, a community cloud can exist on premise or off premise.

Hybrid cloud. Like its name suggests, a hybrid cloud combines two or more cloud types (private, community or public) that remain distinct but share technologies that enable porting of data and applications from one to the other. Companies that use hybrid clouds typically have the security features and customization options of a private cloud but can still take advantage of the scalability and cost-effectiveness of a public cloud. Businesses may use certain applications in a public cloud and then migrate them to a private cloud as their needs change. Many in the industry prefer to refer to this as a hybrid IT strategy. Colocation can also play a role in a hybrid model.

Service models

Another means of distinction regarding the cloud is the variety of cloud service delivery models available. Again, NIST has set some definitions in place to help distinguish the three most common cloud service delivery models: infrastructure as a service, platform as a service and software as a service.

Infrastructure as a service. The IaaS model provides network access to processing, storage, networks and other traditional computing resources that allow the customer to deploy and run arbitrary software. The customer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications; and possibly limited control of select networking components (e.g., host firewalls).[7]

In simple terms, IaaS offers the raw computing, storage and network infrastructure so that customers can load their own software, including operating systems and applications, on to this infrastructure. IaaS providers typically supply resources on-demand from their large resource

pools installed in data centers. For wide-area connectivity, customers can use either the Internet or dedicated virtual private networks.

IaaS offers the greatest degree of control of the three models with scaling and elasticity the responsibility of the customer. Pricing is typically on a usage or reservation basis.

Platform as a service. The PaaS model provides a software platform on which users can build their own applications while hosting them on the PaaS provider's infrastructure. The software platform is used as a development framework to build, debug and deploy applications. It often provides middleware-style services such as database and component services for use by applications. With this model, the customer does not manage or control the underlying cloud infrastructure including network, servers, operating systems or storage, but has control over the deployed applications and possibly application hosting environment configurations.[8]

PaaS is a true cloud model in that applications do not need to worry about the scalability of the underlying hardware and software. When enterprises write their application to run over the PaaS provider's software platform, the elasticity and scalability is guaranteed transparently by the PaaS platform.

Software as a service. In the SaaS model, cloud providers install and operate application software in the cloud. The application is licensed to customers (single or multiple) either as a service-on-demand, through a subscription, in a "pay-as-you-go" model, or at no charge when there is opportunity to generate revenue from streams other than the user.

Cloud users then access the software from cloud clients. The cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on their computers, simplifying maintenance and support.

The customer does not manage or control the underlying cloud infrastructure but does have control over operating systems, storage, deployed applications; and possibly limited control of select networking components (e.g., host firewalls).[9]

SaaS saves the complexity of software installation, maintenance, upgrades and patches for a company's IT team because the software is managed centrally at the SaaS provider's facilities. Monitoring application delivery performance is the responsibility of the SaaS provider. The pricing for the SaaS model is typically on a per-user basis for a fixed bandwidth and storage.

Additional Service Models

While IaaS, PaaS and SaaS are the most common models, in 2012 two additional cloud service models were officially named as basic cloud computing models by the International Telecommunication Union, a specialized agency of the United Nations that is responsible for issues that concern information and communication technologies.[10] They are:

Network as a service. Network as a service provides cloud users with network/transport connectivity services and/or inter-cloud network connectivity services. Services include flexible and extended VPN and bandwidth on demand.

Communications as a service. Communications as a service is an outsourced enterprise communication solution that can be leased from a single vendor. Communications can include voice over IP, instant messaging and collaboration and video conference applications using fixed and mobile devices. The provider is responsible for all hardware and software management and offers guaranteed quality of service. Communications as a service allows businesses to selectively deploy communications devices and modes on a pay-as-you-go, as-needed basis. This eliminates the large capital investment and ongoing overhead for a

system whose capacity may often exceed or fall short of current demand.

With new technologies and business needs emerging, several other cloud service models are taking shape as well. Stay tuned for more information on service models such as:

- Disaster recovery as a service
- Storage as a service
- Security as a service
- Desktop as a service
- Database as a service
- Test environment as a service
- API as a service
- Backend as a service
- Integrated development environment as a service
- Integration platform as a service

Billing models

Just as there are different cloud deployment and service models, there are different ways of billing for cloud services. Each cloud service provider also can have its own variation on these billing models. The two most common billing models, particularly for IaaS, are utility billing and reservation billing.

Utility model. With utility billing, also referred to as "pay-as-you-go" billing, the customer only pays for the resources used, like how customers are charged for using electricity, water or other utilities. The main idea of this pricing configuration is to pay for the actual use with no other commitments. Among the disadvantages of this model is that usage rates often quickly exceed planning parameters, so IT departments can be hit with a huge, unexpected bill.

Reservation model. With the reservation billing model, often referred to as the "use-it-or-lose-it" model, a customer pays for the reserved resources regardless of use. The resources are

always there for the customer's usage. The reservation billing model requires a long-term commitment, but offers discounts based on various factors.

As with the entire cloud ecosystem, cloud billing models will continue to change to meet the needs of both customers and cloud service providers.

Service level agreements

One of the most important aspects of the cloud is the service level agreement. The service level agreement sets the expectation for how cloud resources will be delivered and guarantees that the expectation will be met for uptime and reliability. It creates a strong initial set of ground rules, and a plan for exceptions from the start. This includes the minimum acceptable level of service that must be maintained in worst-case scenarios.

Learn more about the cloud

One of the most important aspects of the cloud is the service level agreement. The service level agreement sets the expectation for how cloud resources will be delivered and guarantees that the expectation will be met for uptime and reliability. It creates a strong initial set of ground rules, and a plan for exceptions from the start. This includes the minimum acceptable level of service that must be maintained in worst-case scenarios. Defining the cloud will continue to be an on-going task, particularly as new cloud service providers enter the market and regulatory,

compliance, security and other issues spur the development of new cloud deployment and service models. However, the Shaw Data Centre & Cloud Solutions Cloud 101

Primer provides an important baseline understanding of the cloud and the various concepts and components that comprise it. Shaw Data Centre & Cloud Solutions can help you learn even more about the cloud, including how the cloud can help your business operate, maintain and improve operations; enhance technology planning; manage capital and operational finance; and even meet regulatory requirements.

Portions reprinted courtesy of NIST, U.S. Department of Commerce. Such portions are not copyrightable.

Shaw Data Centre & Cloud Solutions helps organizations optimize IT transformation while simultaneously balancing cost, scalability, compliance and security. With a focus on building trusted relationships, providing valuable support and delivering tailored solutions and reliable performance, Shaw Data Centre & Cloud Solutions delivers colocation, connectivity, cloud, and managed solutions across the U.S. and Canada.

We'll manage everything for you.

Think of us as a complement to your IT department. Our highly trained team is with you every step of the way.

Call: 1-866-244-7474
Visit: business.shaw.ca/cloud

Follow us:

Shaw) Data Centre & Cloud Solutions

Powered by Flexential