Basics About Cloud Computing

Grace Lewis

September 2010

What is cloud computing and how can an organization decide whether to adopt it? Cloud computing is a distributed computing paradigm that focuses on providing a wide range of users with distributed access to scalable, virtualized hardware and/or software infrastructure over the internet. Despite this rather technical definition, cloud computing is in essence an economic model for a different way to acquire and manage IT resources. An organization needs to weigh the cost, benefits, and risks of cloud computing in determining whether to adopt it as an IT strategy. This paper seeks to help organizations understand cloud computing essentials, including drivers for and barriers to adoption, in support of making decisions about adopting the approach.

Cloud computing is a paradigm for large-scale distributed computing that makes use of existing technologies such as virtualization, service-orientation, and grid computing. It offers a different way to acquire and manage IT resources on a large scale. A simple example of cloud computing is webmail. The webmail provider maintains the server space and provides access; the webmail user just plugs a web address into a browser and submits user information to access an account.

There is growing interest in cloud computing from consumers and providers. For example, cloud service spending worldwide rose by over 20% in 2009 according to onCloudComputing.com, in a year when overall IT spending dropped by about 4%.[1] As of 2010, most cloud consumers are small enterprises, but large enterprises are exploring the paradigm. *Input*, the leading authority on government business, says that federal government spending on cloud computing will grow by a 27% compound annual growth rate between 2009 and 2014.[2] Overall, *Gartner* predicts that by 2012 one in five businesses will not own its own IT assets; one reason for this trend is the move toward cloud computing as a means to reduce IT hardware costs.[3] More adopters will result in more people seeing savings and thus working to reduce barriers to adoption. The growth in cloud com-

Software Engineering Institute
Carnegie Mellon University
4500 Fifth Avenue
Pittsburgh, PA 15213-2612

Phone: 412-268-5800
Toll-free: 1-888-201-4479

www.sei.cmu.edu

1 http://www.oncloudcomputing.com/en/2009/07/list-of-top-cloud-computing-solution-providers-to-watch-in-2009/

2 http://www.input.com/corp/press/detail.cfm?news=1444

3 http://www.cmswire.com/cms/enterprise-20/gartner-top-technology-predictions-for-2010-and-beyond-006390.php

puting consumers will also spur a continuing increase in the number of providers.

Some of the buzz about cloud computing, though, comes from vendor hype that makes it sound like the paradigm is more mature than it is. *Gartner* calculates that cloud computing is at the "Peak of Inflated Expectations," anywhere from two to five years from being adopted in the "mainstream." On the Gartner hype cycle, this technology still needs to pass through the Trough of Disillusionment to get to the Slope of Enlightenment or the Plateau of Productivity.[4]

The emergence of cloud computing and its promise to save money is making it imperative for organizations to examine whether cloud computing makes sense for them. To maneuver through the fog around cloud computing, these organizations first need to know the basics about the technology. This paper gives a solid grounding in the essentials about cloud computing.

Cloud Computing Core Concepts

In the cloud computing model, computing power, software, storage services, and platforms are delivered on demand to external customers over the internet.[5] The access that this technology provides to resources and services can be scaled up or down to meet demand. Cloud computing providers typically charge customers on a pay-per-use model.

Perspectives on Cloud Computing

The types of cloud computing technology can be viewed from two perspectives: capability and access. In this section, we look at three types based on capabilities provided and two based on who can access resources (see Figure 1).

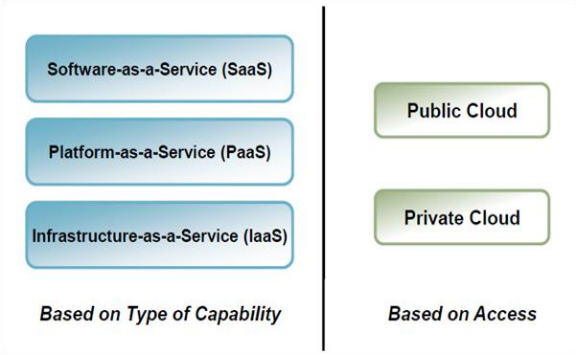

Figure 1: Cloud Computing Types Based on Capability and Access

4 http://www.gartner.com/DisplayDocument?doc_cd= 169368&ref=g_ fromdoc

5 Foster, I., Zhau, Y., Ioan, R., & Lu, S. "Cloud Computing and Grid Computing 360-Degree Compared." Grid Computing Environments Workshop, 2008.2. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4738445

Based on Provided Capabilities

One type of cloud computing capability is called Software-as-a-Service (SaaS). SaaS focuses on providing users with business-specific capabilities, such as e-mail or customer management. In SaaS, organizations and developers can use the business-specific capabilities developed by third parties in the "cloud." Some examples of SaaS providers are

- Google Apps: provides web-based office tools such as e-mail, calendar, and document management
- salesforce.com: provides a full customer relationship management (CRM)[6] application
- zoho.com: provides a large suite of web-based applications, mostly for enterprise use

A second type of cloud computing capability is known as Infrastructure-as-a-Service (IaaS). This capability type provides mainly computational infrastructure available over the internet (e.g., compute cycles or storage). IaaS allows organizations and developers to extend their IT infrastructure on an on-demand basis. Some examples of IaaS providers are

- Amazon Elastic Compute Cloud (EC2): provides users with a special virtual machine (AMI) that can be deployed and run on the EC2 infrastructure
- Amazon Simple Storage Solution (S3): provides users with access to dynamically scalable storage resources
- GoGrid: provides users with access to dynamically scalable computing and storage resources, as well as dedicated servers
- IBM Computing on Demand (CoD): provides users with access to highly configurable servers plus value-added services such as data storage
- Microsoft Live Mesh: provides users with access to a distributed file system; targeted at individual use
- Rackspace Cloud: provides users with access to dynamically scalable computing and storage resources, as well as third-party cloud applications and tools

The third and final type of cloud computing capability is Platform-as-a-Service (PaaS). In this type, application development platforms allow users to leverage the resources of established organizations to create and host applications of a larger scale than an individual or small business would be able to handle.

---

6 http://sites.force.com/appexchange/home

Some PaaS examples include

- Akamai EdgePlatform: provides a large distributed computing platform on which organizations can deploy their web applications; has a large focus on analysis and monitoring
- Force.com: from salesforce.com (an SaaS provider), provides users with a platform to build and run applications and components bought from AppExchange[6] or custom applications
- Google App Engine: provides users with a complete development stack and allows them to run their applications on Google's infrastructure
- Microsoft Azure Services Platform: provides users with on-demand compute and storage services as well as a development platform based on Windows Azure
- Yahoo! Open Strategy (Y!OS): provides users with a means of developing web applications on top of the existing Yahoo! platform and in doing so leveraging a significant portion of the Yahoo! resources

Based on Who Can Access Resources

The two perspectives of cloud computing based on who can access resources can be characterized as public and private.

In public clouds, resources are offered as a service, usually over an internet connection, for a pay-per-usage fee. Users can scale their use on demand and do not need to purchase hardware to use the service. Public cloud providers manage the infrastructure and pool resources into the capacity required by its users.

In private clouds, resources are deployed inside a firewall and managed by the user organization. It is the user organization that owns the software and hardware infrastructure and that manages the cloud and controls access to its resources. Typically, those resources and services are not shared outside the organization.

The National Institute of Standards and Technology (NIST) defines two additional types of cloud deployment models: (1) community clouds that are shared by multiple organizations and support specific needs and concerns of a community and (2) hybrid clouds that are the combination of two or more public, private, and community clouds. However, both community and hybrid cloud are specialties of public and private clouds.

Drivers for Adoption of Cloud Computing

Eight attributes of cloud computing can be seen as drivers for the adoption of cloud computing. The attributes are availability, collaboration, elasticity, lower infrastructure costs, mobility, risk reduction, scalability, and virtualization. Table

1 describes how these attributes can serve as drivers for cloud computing adoption.

Table 1: Cloud Computing Drivers

Attribute	Why It Can Draw an Organization Toward Cloud Computing
Availability	Users have the ability to access their resources at any time through a standard internet connection.
Collaboration	Users begin to see the cloud as a way to work simultaneously on common data and information.
Elasticity	The provider transparently manages a user's resource utilization based on dynamically changing needs.
Lower Infrastructure Costs	The pay-per-usage model allows an organization to only pay for the resources they need with basically no investment in the physical resources available in the cloud. There are no infrastructure maintenance or upgrade costs.
Mobility	Users have the ability to access data and applications from around the globe.
Risk Reduction	Organizations can use the cloud to test ideas and concepts before making major investments in technology.
Scalability	Users have access to a large amount of resources that scale based on their demand.
Virtualization	Each user has a single view of the available resources, independently of how they are arranged in terms of physical devices. Therefore, there is potential from a provider perspective to serve a greater number of users with fewer physical resources.

Barriers to the Adoption of Cloud Computing

Some key organizational concerns can act as barriers to the adoption of cloud computing. These concerns are interoperability, latency, platform or language constraints, regulations, reliability, resource control, and security (see Table 2).

Table 2: Cloud Computing Barriers

Concern	Why It Can Act as a Barrier to Cloud Computing Adoption
Interoperability	A universal set of standards and/or interfaces have not yet been defined, resulting in a significant risk of vendor lock-in.
Latency	All access to the cloud is done via the internet, introducing latency into every communication between the user and the provider.
Platform or Language Constraints	Some cloud providers support specific platforms and languages only.

Concern	Why It Can Act as a Barrier to Cloud Computing Adoption
Regulations	There are concerns in the cloud computing community over jurisdiction, data protection, fair information practices, and international data transfer—mainly for organizations that manage sensitive data.
Reliability	Many existing cloud infrastructures leverage commodity hardware that is known to fail unexpectedly.
Resource Control	The amount of control that the user has over the cloud provider and its resources varies greatly between providers.
Security	The main concern is data privacy: users do not have control or knowledge of where their data is being stored.

Relationship of Cloud Computing to Other Technologies

In this section, we look at the relationship of cloud computing to two other prominent approaches for large-scale, distributed systems: service-oriented architecture (SOA) and Grid Computing.

Cloud Computing and SOA

While it might be tempting to view cloud computing as a substitute for service-oriented architecture (SOA), the two technologies are distinct. SOA is a way of designing, developing, deploying, and managing systems characterized by coarse-grained services that represent reusable functionality. In SOA, service consumers compose applications or systems using the functionality provided by these services through standard interfaces. The services in a cloud can be defined as services in a SOA context.

The two are distinct, then, but it is possible to use elements of SOA in cloud environments. Some cloud environments, for example, offer web service interfaces (one specific implementation of SOA) to their services. From an architecture perspective

- A cloud infrastructure could be built on an SOA infrastructure by adding a layer or virtualization and self-provisioning.
- A service layer could be added on top of cloud resources.

Cloud Computing and Grid Computing

A grid is "a system that uses open, general-purpose protocols to federate distributed resources and to deliver better-than-best-effort qualities of service."[7] Although the distinction with cloud computing is not clear, one differentiator is that grid computing relates exclusively to infrastructure services. A grid infrastruc-

[7] Foster, I. *What is the grid? A three-point checklist*. http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf (2002)

ture provides a set of abstractions and interfaces for access to, and management of, shared resources.[8]

Final Points and Resources

Cloud Computing adoption requires cost/benefit/risk analysis to answer questions such as

- What resources should an organization move to the cloud, if any?
- What situations warrant use of cloud resources, even for one-time situations?
- Which drivers are most important to the organization? Is concern over the security of data in the cloud pre-eminent, for example?
- Which model of access works better for the organization—private or public?
- What risks are associated with using resources on the cloud?

Environment and Tool References

- Akamai EdgePlatform: http://www.akamai.com/html/technology/ edgeplatform.html
- Amazon Elastic Compute Cloud (EC2): http://aws.amazon.com/ec2/
- Amazon Simple Storage Solution (S3): http://aws.amazon.com/s3/
- Force.com:
- GoGrid: http://www.gogrid.com/
- Google App Engine:
- Google Apps: http://www.google.com/apps/intl/en/business/index.html
- IBM Computing On Demand: http://www-03.ibm.com/systems/deepcomputing/cod/
- Microsoft Azure Services Platform: http://www.microsoft.com/azure/
- Microsoft Live Mesh:
- Rackspace:
- Salesforce.com: http://www.salesforce.com/crm/products.jsp/
- Yahoo! Open Strategy (Y!OS): http://developer.yahoo.com/yos/intro/
- Zoho: http://www.zoho.com/

Selected Literature References

- Drdogmus, H. "Cloud Computing: Does Nirvana Hide Behind the Nebula?" *IEEE Software26*, 2 (March/April 2009): 4-6.

[8] Foster, I., Kesselman, C., Nick, J. M., & Tuecke, S. "Grid services for distributed systems integration." *IEEE Computer 35,* 6 (June 2002): 37-46.

- Foster, I., Zhau, Y., Ioan, R., & Lu, S. “Cloud Computing and Grid Computing 360-Degree Compared.” Grid Computing Environments Workshop, 2008.2. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4738445
- Grossman, R. “The Case for Cloud Computing.” *ITPro 11*, 2 (March/April 2009): 23-27.
- Hayes, B. “Cloud Computing.” *Communications of the ACM 51*, 7 (July 2008): 9-11.
- Lin, G., David, F., Jinzy, Z., & Glenn, D. “Cloud Computing: IT as a Service.” *IT Pro 11*, 2 (March/April 2009): 10-13.
- McEvoy, G. & Schulze, B. “Using Clouds to address Grid Limitations,” 1-6. *Proceedings of the 6th International Workshop on Middleware for Grid Computing* (MGC). Leuven, Belgium, December 2008. ACM, 2008. ISBN: 978-1-60558-365-5
- Strowd, Harrison & Lewis, Grace A. *T-Check in System of Systems Technologies: Cloud Computing* (CMU/SEI-2010-TN-009). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. 2010. http:/www.sei.cmu.edu/library/abstracts/reports/10tn009.cfm
- Wang, L., Tao, J., Kunze, M., Castellanos, A., Kramer, D., & Karl, W. “Scientific Cloud Computing: Early Definition and Experience,” 825-830. *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC ’08)*. Dalian, China, September 2008. IEEE, 2008.
- Weiss, A. “Computing in the Clouds.” *networker 11*, 4 (December 2007): 16-25.