CHAPTER I
INTRODUCTION

What Is the Cloud?

The term cloud has been used historically as a metaphor for the Internet. This usage was originally derived from its common depiction in network diagrams as an outline of a cloud. used to represent the transport of data across carrier backbones (which owned the cloud) to an endpoint location on the other side of the cloud. This concept dates back as early as 1961, when Professor John McCarthy suggested that computer time-sharing tech-nology might lead to a future where computing power and even specific applications might be sold through a utility-type business model.1This idea became very popular in the late 1960s, but by the mid-1970s the idea faded away when it became clear that the IT-related technologies of the day were unable to sustain such a futuristic computing model. However, since the turn of the millennium, the concept has been revitalized. It was during this time of revitalization that the term cloud computing began to emerge in technology circles.

This form of computing is growing in popularity, however, as companies have begun to extend the model to a cloud computing paradigm providing virtual servers that IT departments and users can access on demand. Early enterprise adopters used utility computing mainly for non-mission-critical needs. but that is quickly changing as trust and reliability issues are resolved.

Some people think cloud computing is the next big thing in the world of IT. Others believe it is just another variation of the utility computing model that has been repackaged in this decade as something new and cool. However, it is not just the buzzword "cloud computing" that is causing con-fusion among the masses. Currently, with so few cloud computing vendors actually practicing this form of technology and also almost every analyst from every research organization in the country defining the term differ-ently, the meaning of the term has become very nebulous. Even among those who think they understand it, definitions vary, and most of those def-initions are hazy at best. To clear the haze and make some sense of Me new concept, this book will attempt to help you understand just what cloud computing really means, how disruptive to your business it may become in future, and what its advantages and disadvantages are. the cloud is often used as a metaphor for the Internet when "the cloud" is combined with "computing,- it causes a lot of confusion technology vendor define cloud computing

The Global Nature of the Cloud

The cloud sees no borders and thus has made the world a much smaller pike. The Internet is global in scope but respects only established commu-nication paths. People from everywhere now have access to other people hom anywhere else. Globalization of computing assets may be the biggest contribution the doud has made to date. For this reason, the cloud is the subject of many complex geopolitical issues. Cloud vendors must satisfy Myriad regulatory concerns in order to deliver doud services to a global Asko. When the Internet was in its infancy, many people believed cyber-ace was a distinct environment that needed laws specific to

itself. Univer-y computing centers and the ARPANET were, for a time, the encapsulated environments where the Internet existed. It took a while to get Mess to warm up to the idea.

Cloud computing is still in its infancy. There is a hodge-podge of pro-Otitis, both large and small, delivering a wide variety of cloud-based set-tles. For example, there are full-blown applications, support services, mail-Altering services, storage services, etc. IT practitioners have learned to contend with some of the many cloud-based services out of necessity as business needs dictated. However, cloud computing aggregators and integrators are Indy emerging, offering packages of products and services as a single entry point into the cloud.

The concept of cloud computing becomes much more understandable when one begins to think about what modern IT environments always require—the means to increase capacity or add capabilities to their infra-structure dynamically, without investing money in the purchase of new infrastructure. all the while without needing to conduct training for new personnel and without the need for licensing new software. Given a solution to the aforementioned needs, cloud computing models that encompass a subscription-based or pay-per-use paradigm provide a service that can be used over the Internet and extends an IT shop's existing capabilities. Many users have found that this approach provides a return on investment that IT managers are more than willing to accept.

Cloud-Based Service Offerings

Cloud computing may be viewed as a resource available as a service for vir-tual data centers, but cloud computing and virtual data centers are not the same. For example, consider Amazon's S3 Storage Service. This is a data storage service designed for use across the Internet (i.e., the cloud). It is designed to make web-scale computing easier for developers. According to Amazon:

Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

Amazon.com has played a vital role in the development of cloud com-puting. In modernizing its data centers after the dot-corn bubble burst in 2001, it discovered that the new cloud architecture it had implemented resulted in some very significant internal efficiency improvements. By providing access to its systems for third-party users on a utility computing basis, via Amazon Web Services, introduced in 2002, a revolution of sorts mall Web Services began implementing its model by renting sorvice outside a given user's domain, wherever on the ted. This approach modernized a style of could be provided "as a service" to users By allowing their users to access technology-enabled services in the cloud," without any need for knowledge of expertise with, or control over how the technology infrastructure that supports those services worked, Amazon shirted the approach to computing radically. This approach transformed cloud computing into a paradigm whereby data is permanently sotred in remove servers accessible via the Internet and cached temporarily on client devices that may include

desktops, table computers, notebooks. hand-held devices, mobile phones. etc., and is often called *Software as a Service* (SaaS).

SaaS is a type of cloud computing that delivers applications through a lip thousands of customers using a multiuser architecture. The focus is on the end user as opposed to managed services (described below). For the customer, there are no up-front investment costs in servers to licensing. For the service provider, with just one product to maintain, costs arc relatively low compared to the costs incurred with a con-I model. Salesforce.com is by far the best-known example of SaaS computing among enterprise applications. Salesforce.com was founded in 1999 by former Oracle executive Marc Benioff, who pioneered the concept of delivering enterprise applications via a simple web site. Now days, SaaS is also commonly used for enterprise resource planning and tesource applications. Another example is Google Apps, which provides online access via a web browser to the most common office and business applications used today, all the while keeping the software and user data stored on Googlr servers. A decade ago, no one could have predicted the rise of SaaS applications such as these.

Managed service providers (MSPs) offer one of the oldest forms of cloud computing. Basically, a managed service is an application that is accessible to an organization's IT infrastructure rather than to end users. Services includ virus scanning for email, antispam services such as Postini, desktop management services such as those offered by CenterBeam or Everdream, and application performance monitoring. Managed security services that are delivered by third-party providers also fall into this category.

Platform-as-a-Service (PaaS) is yet another variation of SaaS. Some-times referred to simply as web services in the cloud, PaaS is closely related to SaaS but delivers a platform from which to work rather than an applica-tion to work with. These service providers offer application programming interfaces (APIs) that enable developers to exploit functionality over the Internet, rather than delivering hill-blown applications. This variation of cloud computing delivers development environments to programmers, ana-lysts, and software engineers as a service. A general model is implemented finder which developers build applications designed to run on the provider's infrastructure and which are delivered to users in via an Internet browser. The main drawback to this approach is that these services are limited by the vendor's design and capabilities. This means a compromise between free-dom to develop code that does something other than what the provider can provide and application predictability, performance, and integration.

An example of this model is the Google App Engine. According to Google, "Google App Engine makes it easy to build an application that runs reliably, even tinder heavy load and with large amounts of data. The Google App Engine environment includes the following features

Dynamic web serving, with full support for common web technologies

Persistent storage with queries, sorting, and transactions

Automatic scaling and load balancing

APIs for authenticating users and sending email using Google Accounts

A fully featured local development environment that simulates Google App Engine on your computer

Currently, Google App Engine applications are implemented using the Python programming language. The runtime environment includes the full Python language and most of the Python standard library. For extremely lightweight development. cloud based mashup platforms (Ajax modules that arc assembled in code) abound, such as Yahoo Pipes or Dapper.net.

Grid Computing or Cloud Computing?

Grid computing is often confused with cloud computing. Grid computing 'it a form of distributed computing that implements a virtual supercomputer Side up of a duster of networked or Internetworked computers acting in Unison to perform very large tasks. Many cloud computing deployments Say are powered by grid computing implementations and are billed like utilities, but cloud computing can and should be seen as an evolved next Opp away from the grid utility model. There is an ever-growing list of providers that have successfully used cloud architectures with little or no centralised' infrastructure or billing systems, such as the peer-to-peer network liniment and the volunteer computing initiative SETI@home.

Service commerce platforms are yet another variation of SaaS and Pr. This type of cloud computing service provides a centralized service that users interact with. Currently, the most often used application of platform is found in financial trading environments or systems that 1110w users to order things such as travel or personal services from a corn-litun platform (e.g., Expedia.com or Hotels.com), which then coordinates inking and service delivery within the specifications set by the user.

Is the Cloud Model Reliable?

TIte majority of today's cloud computing infrastructure consists of time-lined and highly reliable services built on servers with varying levels of vir-Utilized technologies, which are delivered via large data centers operating under service-level agreements that require 99.99% or better uptime. Corn-lftervial offerings have evolved to meet the quality-of-service requirements ti Customers and typically offer such service-level agreements to their cos-miners. From users' perspective, the cloud appears as a single point of access kr all their computing needs. These cloud-based services are accessible any-where in the world, as long as an Internet connection is available. Open standards and open-source software have also been significant factors in the winvilt of cloud computing.

Benefits of Using a Cloud Model

Because customers generally do not own the infrastructure used in cloud computing environments, they can forgo capital expenditure and consume resources as a service by just paying for what they use. Many cloud comput-ing offerings have adopted the utility computing and billing model described above, while others bill on a subscription basis. By sharing com-puting power among multiple users, utilization rates are generally greatly improved, because cloud computing servers are not sitting dormant for lack of use. This factor alone can reduce infrastructure costs significantly and accelerate the speed of applications development.

A beneficial side effect of using this model is that computer capacity increases dramatically, since customers do not have to engineer their appli-cations for peak times, when processing loads are greatest. Adoption of the cloud computing model has also been enabled because of the greater avail-ability of increased high-speed bandwidth. With greater enablement, though, there are other issues one must consider, especially legal ones.

What About Legal Issues When Using Cloud Models?

The United States—European Union Safe Harbor Act provides a seven-point Gmework of requirements for U.S. companies that may use data from other parts of the world, namely, the European Union. This framework sets forth how companies can participate and certify their compliance and is defined in detail on the U.S. Depart-ment of Commerce and Federal Trade Commission web sites. In summary, the agreement allows most U.S. corporations to certify that they have joined a self-regulatory organization that adheres to the following seven Safe Har-bor Principles or has implemented its own privacy policies that conform with these principles:

1. Notify individuals about the purposes for which information is collected and used.
2. Give individuals the choke of whether their information an be disclosed to a party.
3. Ensure that if it transfers personal information to a third party, that third party also provides the same level of privacy protections)
4. Allow individuals access to their personal information.
5. Take reasonable security precautions to protect collected data from loss, misuse, or disclosure.
6. Take reasonable steps to ensure the integrity of the data col-lected.;
7. Have in place an adequate enforcement mechanism.

Major service providers such as Amazon Web Services cater to a global marketplace, typically the United States, Japan, and the European Union. by deploying local infrastructure at those locales and allowing customers to select availability zones. However, there are still concerns about security and privacy at both the individual and governmental levels. Of major con-cern is the USA PATRIOT Act and the Electronic Communications Pri-vacy Act's Stored Communications Act. The USA PATRIOT Act, more commonly known as the Patriot Act, is a controversial Act of Congress that U.S. President George W. Bush signed into law on October 26. 2001. The contrived acronym stands for "Uniting and Strengthening America by Pro-viding Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001" (Public Law P.L. 107-56). The Act expanded the definition of terrorism to include domestic terrorism, thus enlarging the number of activities to which the USA PATRIOT Act's law enforcement powers could be applied. It increased law enforcement agencies' ability to surveil tele-phone. email communications, medical, financial, and other records and increased the range of discretion for law enforcement and immigration authorities when detaining and deporting immigrants suspected of terror-km-related acts. It lessened the restrictions on foreign intelligence gather-ing within the United States. Furthermore, it expanded the Secretary of the Treasury's authority to regulate financial transactions involving foreign individuals and businesses.

The Electronic Communications Privacy Act's Stored Communications Act is defined in the U.S. Code, Title 18. Part I, Chapter 121, $ 2701, Unlawful Access to Stored Communications. Offenses committed under ibis act include intentional access without authorization to a facility through which an electronic communication service is provided or inten-tionally exceeding an authorization to access that facility in order to obtain. alter, or prevent authorized access to a wire or electronic communication while it is in electronic storage in such a system. Persons convicted under this Act can be punished if the offense is committed for purposes of cAom-mercial advantage, malicious destruction or damage, or private commercial gain, or in furtherance of any criminal or tortious act in violation of the Constitution or laws of the United States or any state by a fine or imprison-ment or both for not more than five years in the case of a first offense. For a second or subsequent offense. the penalties stiffen to fine or imprisonment for not more than 10 years, or both.

What Are the Key Characteristics of Cloud Computing?

There are several key characteristics of a cloud computing environment. Service offerings are most often made available to specific consumers and small businesses that see the benefit of use because their capital expenditure is minimized. This serves to lower barriers to entry in the marketplace, since the infrastructure used to provide these offerings is owned by the cloud ser-vice provider and need not be purchased by the customer. Because users are not tied to a specific device (they need only the ability to access the Inter-net) and because the Internet allows for location independence, use of the cloud enables cloud computing service providers customers to access cloud-enabled systems regardless of where they may be located or what device they choose to use.

Multitenancy enables sharing of resources and costs among a large pool of users. Chief benefits to a multitenancy approach include:

Centralization of infrastructure and lower costs

Increased peak-load capacity

Efficiency improvements for systems that are often underutilized

Dynamic allocation of CPU, storage, and network bandwidth

Consistent performance that is monitored by the provider of the service

Reliability is often enhanced in cloud computing environments because service providers utilize multiple redundant sites. This is attractive to enterprises for business continuity and disaster recovery reasons. The drawback, however, is that IT managers can do very little when an outage occurs.

Another benefit that makes cloud services more reliable is that scalabil-ity can vary dynamically based on changing user demands. Because the ser-vice provider manages the necessary infrastructure, security often is vastly improved. As a result of data centralization, there is an increased focus on protecting customer resources maintained by the service provider. 'lb assure customers that their data is safe, cloud providers are quick to invest in dedi-cated security staff. This is largely seen as beneficial but has also raised con-cerns about a user's

loss of control over sensitive data. Access to data is usually logged, but accessing the audit logs can be difficult or even impossi-ble for the customer.

Data centers, computers, and the entire associated infrastructure needed to support cloud computing are major consumers of energy. Sus-tainability of the cloud computing model is achieved by leveraging improvements in resource utilization and implementation of more energy-efficient system cloud computing ism likely to bring supercomputing capabilities to the masses. Yahoo, Google, Microsoft, IBM, and others are engaged in the creation of online services to give their users even better access to data to aid in daily life issues such as health care, finance, insurance, etc.

Challenges for the Cloud

The biggest challenges these companies face are secure data storage, high-speed access to the Internet, and standardization. Storing large amounts of data that is oriented around user privacy, identity, and application-specific preferences in centralized locations raises many concerns about data protec-tion. These concerns, in turn, give rise to questions regarding the legal framework that should be implemented for a cloud-oriented environment. Mother challenge to the cloud computing model is the fact that broadband penetration in the United States remains far behind that of many other countries in Europe and Asia. Cloud computing is untenable without high-speed connections (both wired and wireless). Unless broadband speeds are available, cloud computing services cannot be made widely accessible. Finally, technical standards used for implementation of the various com-puter systems and applications necessary to make cloud computing work have still not been completely defined, publicly reviewed, and ratified by an oversight body. Even the consortiums that are forming need to get past that hurdle at some point, and until that happens, progress on new products will likely move at a snail's pace.

The reliability of cloud computing has recently been a controversial topic in technology circles. Because of the public availability of a cloud environment, problems that occur in the cloud tend to receive lots of public exposure. Unlike problems that occur in enterprise environments, which often can be contained without publicity, even when only a few cloud computing users have problems, it makes headlines.

Google leads the industry in evolving the cloud computing model to become a part of what is being called Web 3.0—the next generation of Internet.