

How Cloud Computing Works

Introduction to How Cloud Computing Works

Let's say you're an executive at a large corporation. Your particular responsibilities include making sure that all of your employees have the right hardware and software they need to do their jobs. Buying computers for everyone isn't enough -- you also have to purchase software or **software licenses** to give employees the tools they require. Whenever you have a new hire, you have to buy more software or make sure your current software license allows another user. It's so stressful that you find it difficult to go to sleep on your huge pile of money every night.

Soon, there may be an alternative for executives like you. Instead of installing a suite of software for each computer, you'd only have to load one application. That application would allow workers to log into a Web-based service which hosts all the programs the user would need for his or her job. Remote machines owned by another company would run everything from e-mail to word processing to complex data analysis programs. It's called **cloud computing**, and it could change the entire computer industry.

In a cloud computing system, there's a significant workload shift. Local computers no longer have to do all the heavy lifting when it comes to running applications. The network of computers that make up the cloud handles them instead. Hardware and software demands on the user's side decrease. The only thing the user's computer needs to be able to run is the cloud computing system's **interface software**, which can be as simple as a Web browser, and the cloud's network takes care of the rest.

There's a good chance you've already used some form of cloud computing. If you have an e-mail account with a Web-based e-mail service like Hotmail, Yahoo! Mail or Gmail, then you've had some experience with cloud computing. Instead of

running an e-mail program on your computer, you log in to a Web e-mail account remotely. The software and storage for your account doesn't exist on your computer -- it's on the service's computer cloud.

What makes up a cloud computing system?

Although cloud computing is an emerging field of computer science, the idea has been around for a few years. It's called cloud computing because the data and applications exist on a "cloud" of Web servers.

Cloud Computing Architecture

When talking about a cloud computing system, it's helpful to divide it into two sections: the **front end** and the **back end**. They connect to each other through a network, usually the Internet. The front end is the side the computer user, or client, sees. The back end is the "cloud" section of the system.

The front end includes the client's computer (or computer network) and the application required to access the cloud computing system. Not all cloud computing systems have the same user interface. Services like Web-based e-mail programs leverage existing Web browsers like Internet Explorer or Firefox. Other systems have unique applications that provide network access to clients.

On the back end of the system are the various computers, servers and data storage systems that create the "cloud" of computing services. In theory, a cloud computing system could include practically any computer program you can imagine, from data processing to video games. Usually, each application will have its own dedicated server.

A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called **protocols** and uses a special kind of software called **middleware**. Middleware allows networked computers to communicate with each other. Most of the time, servers don't run at full capacity. That means there's unused processing power going to waste. It's possible to fool a physical server into thinking it's actually multiple servers, each running with its own independent operating system. The technique is called server virtualization. By maximizing the output of individual servers, server virtualization reduces the need for more physical machines.

If a cloud computing company has a lot of clients, there's likely to be a high demand for a lot of storage space. Some companies require hundreds of digital storage devices. Cloud computing systems need at least twice the number of storage devices it requires to keep all its clients' information stored. That's because these devices, like all computers, occasionally break down. A cloud computing system must make a copy of all its clients' information and store it on other devices. The copies enable the central server to access backup machines to retrieve data that otherwise would be unreachable. Making copies of data as a backup is called **redundancy**.

What are some of the applications of cloud computing?

Some of the companies researching cloud computing are big names in the computer industry. Microsoft, IBM and Google are investing millions of dollars into research. Some people think Apple might investigate the possibility of producing interface hardware for cloud computing systems.

Cloud Computing Applications

The applications of cloud computing are practically limitless. With the right middleware, a cloud computing system could execute all the programs a normal computer could run. Potentially, everything from generic word processing software to customized computer programs designed for a specific company could work on a cloud computing system.

Why would anyone want to rely on another computer system to run programs and store data? Here are just a few reasons:

- Clients would be able to access their applications and data from anywhere at any time. They could access the cloud computing system using any computer linked to the Internet. Data wouldn't be confined to a hard drive on one user's computer or even a corporation's internal network.
- It could bring hardware costs down. Cloud computing systems would reduce the need for advanced hardware on the client side. You wouldn't need to buy the fastest computer with the most memory, because the cloud system would take care of those needs for you. Instead, you could buy an inexpensive computer terminal. The terminal could include a monitor, input devices like a keyboard and mouse and just enough processing power to run the middleware necessary to connect to the cloud system. You wouldn't need a large hard drive because you'd store all your information on a remote computer.
- Corporations that rely on computers have to make sure they have the right software in place to achieve goals. Cloud computing systems give these organizations company-wide access to computer applications. The companies don't have to buy a set of software or software licenses for every employee. Instead, the company could pay a metered fee to a cloud computing company.
- Servers and digital storage devices take up space. Some companies rent physical space to store servers and databases because they don't have it available on site. Cloud computing gives these companies the option of storing data on someone else's hardware, removing the need for physical space on the front end.
- Corporations might save money on IT support. Streamlined hardware would, in theory, have fewer problems than a network of **heterogeneous** machines and operating systems.
- If the cloud computing system's back end is a grid computing system, then the client could take advantage of the entire network's processing power. Often, scientists and researchers work with calculations so complex that it would take years for individual computers to complete them. On a grid computing system, the client could send the calculation to the cloud for processing. The cloud system would tap into the processing power of all available computers on the back end, significantly speeding up the calculation.

While the benefits of cloud computing seem convincing, are there any potential problems?

There are a few standard hacker tricks that could cause cloud computing companies major headaches. One of those is called **key logging**. A key logging program records keystrokes. If a hacker manages successfully to load a key logging program on a victim's computer, he or she can study the keystrokes to discover user

names and passwords. Of course, if the user's computer is just a streamlined terminal, it might be impossible to install the program in the first place.

Cloud Computing Concerns

Perhaps the biggest concerns about cloud computing are **security** and **privacy**. The idea of handing over important data to another company worries some people. Corporate executives might hesitate to take advantage of a cloud computing system because they can't keep their company's information under lock and key.

The counterargument to this position is that the companies offering cloud computing services live and die by their reputations. It benefits these companies to have reliable security measures in place. Otherwise, the service would lose all its clients. It's in their interest to employ the most advanced techniques to protect their clients' data.

Privacy is another matter. If a client can log in from any location to access data and applications, it's possible the client's privacy could be compromised. Cloud computing companies will need to find ways to protect client privacy. One way is to use **authentication** techniques such as user names and passwords. Another is to employ an **authorization** format -- each user can access only the data and applications relevant to his or her job.

Some questions regarding cloud computing are more philosophical. Does the user or company subscribing to the cloud computing service own the data? Does the cloud computing system, which provides the actual storage space, own it? Is it possible for a cloud computing company to deny a client access to that client's data? Several companies, law firms and universities are debating these and other questions about the nature of cloud computing.

How will cloud computing affect other industries? There's a growing concern in the IT industry about how cloud computing could impact the business of computer maintenance and repair. If companies switch to using streamlined computer systems, they'll have fewer IT needs. Some industry experts believe that the need for IT jobs will migrate to the back end of the cloud computing system.

Introduction to Cloud Computing

Cloud Computing

The term "cloud", as used in this white paper, appears to have its origins in network diagrams that represented the internet, or various parts of it, as schematic clouds. "Cloud computing" was coined for what happens when applications and services are moved into the internet "cloud." Cloud computing is not something that suddenly appeared overnight; in some form it may trace back to a time when computer systems remotely time-shared computing resources and applications. More currently though, cloud computing refers to the many different types of services and applications being delivered in the internet cloud, and the fact that, in many cases, the devices used to access these services and applications do not require any special applications.

Many companies are delivering services from the cloud. Some notable examples as of 2010 include the following:

- **Google** — Has a private cloud that it uses for delivering many different services to its users, including email access, document applications, text translations, maps, web analytics, and much more.

- **Microsoft** — Has Microsoft® Sharepoint® online service that allows for content and business intelligence tools to be moved into the cloud, and Microsoft currently makes its office applications available in a cloud.
- **Salesforce.com** — Runs its application set for its customers in a cloud, and its Force.com and Vmforce.com products provide developers with platforms to build customized cloud services.

But, what is cloud computing? The following sections note cloud and cloud computing characteristics, services models, deployment models, benefits, and challenges.

Characteristics

Cloud computing has a variety of characteristics, with the main ones being:

- ✦ **Shared Infrastructure** — Uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.
- ✦ **Dynamic Provisioning** — Allows for the provision of services based on current demand requirements. This is done automatically using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.
- ✦ **Network Access** — Needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs (for example, ones based on HTTP). Deployments of services in the cloud include everything from using business applications to the latest application on the newest smartphones.
- ✦ **Managed Metering** — Uses metering for managing and optimizing the service and to provide reporting and billing information. In this way, consumers are billed for services according to how much they have actually used during the billing period. In short, cloud computing allows for the sharing and scalable deployment of services, as needed, from almost any location, and for which the customer can be billed based on actual usage.
- ✦ **Service Models:** Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements. The primary service models being deployed (see Figure 1) are commonly known as:
 - **Software as a Service (SaaS)** — Consumers purchase the ability to access and use an application or service that is hosted in the cloud. A benchmark example of this is Salesforce.com, as discussed previously, where necessary information for the interaction between the consumer and the service is hosted as part of the service in the cloud. Also, Microsoft is expanding its involvement in this area, and as part of the cloud computing option for Microsoft® Office 2010, its Office Web Apps are available to Office volume licensing customers and Office Web App subscriptions through its cloud-based Online Services.
 - **Platform as a Service (PaaS)** — Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed.
 - **Infrastructure as a Service (IaaS)** — Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not themselves control the cloud infrastructure. Also known are the various subsets of these models that may be related to a particular industry or market. Communications as a Service (CaaS) is one such subset model used to describe hosted IP telephony services. Along with the move to CaaS is a shift to more IP-centric communications and more SIP trunking deployments. With IP and SIP in place, it can be as easy to have the PBX in the cloud as it is to have it on the premise. In this context, CaaS could be seen as a subset of SaaS.

Figure 1. Service Model Types

Deployment Models

Deploying cloud computing can differ depending on requirements, and the following four deployment models have been identified,

each with specific characteristics that support the needs of the services and users of the clouds in particular ways (see Figure 2).

- **Private Cloud** — The cloud infrastructure has been deployed, and is maintained and operated for a specific organization. The operation may be in-house or with a third party on the premises.
- **Community Cloud** — The cloud infrastructure is shared among a number of organizations with similar interests and requirements. This may help limit the capital expenditure costs for its establishment as the costs are shared among the organizations. The operation may be in-house or with a third party on the premises.
- **Public Cloud** — The cloud infrastructure is available to the public on a commercial basis by a cloud service provider. This enables a consumer to develop and deploy a service in the cloud with very little financial outlay compared to the capital expenditure requirements normally associated with other deployment options.
- **Hybrid Cloud** — The cloud infrastructure consists of a number of clouds of any type, but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud.

Benefits

The following are some of the possible benefits for those who offer cloud computing-based services and applications:

- **Cost Savings** — Companies can reduce their capital expenditures and use operational expenditures for increasing their computing capabilities. This is a lower barrier to entry and also requires fewer in-house IT resources to provide system support.
- **Scalability/Flexibility** — Companies can start with a small deployment and grow to a large deployment fairly rapidly, and then scale back if necessary. Also, the flexibility of cloud computing allows companies to use extra resources at peak times, enabling them to satisfy consumer demands.
- **Reliability** — Services using multiple redundant sites can support business continuity and disaster recovery.
- **Maintenance** — Cloud service providers do the system maintenance, and access is through APIs that do not require application installations onto PCs, thus further reducing maintenance requirements.
- **Mobile Accessible** — Mobile workers have increased productivity due to systems accessible in an infrastructure available from anywhere.

Challenges

The following are some of the notable challenges associated with cloud computing, and although some of these may cause a slowdown when delivering more services in the cloud, most also can provide opportunities, if resolved with due care and attention in the planning stages.

- **Security and Privacy** — Perhaps two of the more “hot button” issues surrounding cloud computing relate to storing and securing data, and monitoring the use of the cloud by the service providers. These issues are generally attributed to slowing the deployment of cloud services. These challenges can be addressed, for example, by storing the information internal to the organization, but allowing it to be used in the cloud. For this to occur, though, the security

mechanisms between organization and the cloud need to be robust and a Hybrid cloud could support such a deployment.

- **Lack of Standards** — Clouds have documented interfaces; however, no standards are associated with these, and thus it is unlikely that most clouds will be interoperable. The Open Grid Forum is developing an Open Cloud Computing Interface to resolve this issue and the Open Cloud Consortium is working on cloud computing standards and practices. The findings of these groups will need to mature, but it is not known whether they will address the needs of the people deploying the services and the specific interfaces these services need. However, keeping up to date on the latest standards as they evolve will allow them to be leveraged, if applicable.
- **Continuously Evolving** — User requirements are continuously evolving, as are the requirements for interfaces, networking, and storage. This means that a “cloud,” especially a public one, does not remain static and is also continuously evolving.
- **Compliance Concerns** — The Sarbanes-Oxley Act (SOX) in the US and Data Protection directives in the EU are just two among many compliance issues affecting cloud computing, based on the type of data and application for which the cloud is being used. The EU has a legislative backing for data protection across all member states, but in the US data protection is different and can vary from state to state. As with security and privacy mentioned previously, these typically result in Hybrid cloud deployment with one cloud storing the data internal to the organization.

Communications in the Cloud

For service developers, making services available in the cloud depends on the type of service and the device(s) being used to access it. The process may be as simple as a user clicking on the required web page, or could involve an application using an API accessing the services in the cloud. Telcos are starting to use clouds to release their own services and those developed by others, but using Telco infrastructure and data. The expectation is that the Telco's communications infrastructure provides a revenue generating opportunity.

Using the Communications Services

When in the cloud, communications services can extend their capabilities, or stand alone as service offerings, or provide new interactivity capabilities to current services.

Cloud-based communications services enable businesses to embed communications capabilities into business applications, such as Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) systems. For “on the move” business people, these can be accessed through a smartphone, supporting increased productivity while away from the office. These services are over and above the support of service deployments of VoIP systems, collaboration systems, and conferencing systems for both voice and video. They can be accessed from any location and linked into current services to extend their capabilities, as well as stand alone as service offerings. In terms of social networking, using cloud-based communications provides click-to-call capabilities from social networking sites, access to Instant Messaging systems and video communications, broadening the interlinking of people within the social circle.

Accessing through Web APIs

Accessing communications capabilities in a cloud-based environment is achieved through APIs, primarily Web 2.0 RESTful APIs, allowing application development outside the cloud to take advantage of the communication infrastructure within it (see Figure 3).

These APIs open up a range of communications possibilities for cloud-based services, only limited by the media and signaling capabilities within the cloud. Today's media services allow for communications and management of voice and video across a complex range of codecs and transport types. By using the Web APIs, these complexities can be simplified and the media can be delivered to the remote device more easily. APIs also enable communication of other services, providing new opportunities and helping to drive Average Revenue per User (ARPU) and attachment rates, especially for Telcos.

Media Server Control Interfaces

When building communications capabilities into the “core of the cloud,” where they will be accessed by another service, the Web 2.0 APIs can be used, as well as a combination of SIP or VoiceXML and the standard media controlling APIs such as MSML, MSCML, and JSR309. The combinations provide different capability sets, but with MediaCTRL being developed in the Internet Engineering Task Force (IETF), it is expected that MediaCTRL will supersede MSML and MSCML and have an upsurge in availability and more developments after it is ratified. JSR309 is a notable choice for those seeking Java development, as it provides the Java interface to media control.

Figure 4 is an example of accessing services in the cloud through Web 2.0 and media control interface APIs.

Whether businesses are deploying communications services for access from outside of or within the cloud, the environment is one that supports the speedy development and rollout of these capabilities.

Communications Scalability

To deliver on the scalability requirements for cloud-based deployments, the communications software should be capable of running in virtual environments. This allows for easily increasing and decreasing session densities based on the needs at the time, while keeping the physical resource requirement on servers to a minimum.