The Cisco Powered Network Cloud: An Exciting Managed Services Opportunity

The cloud computing phenomenon is generating a lot of interest worldwide because of its potential to offer services on demand, at lower cost than current options, and with less complexity, greater scalability, and wider reach. The opportunities for providers of managed services to benefit from this model are significant and exciting. But confusion over cloud terminology, existing services versus vaporware, and evolving service models are widespread. Many potential customers are uncertain about whether to adopt these services and many service providers are unsure of how to best integrate and market the cloud architecture.

This white paper provides a high-level overview of the terminology, concepts, and requirements for cloud computing services. It also highlights components of the intelligent Cisco® IP Next-Generation Network (IP NGN) that are cloud-ready and can be the basis for new services. The Cisco IP NGN delivers the security, quality of service (QoS), high availability, interoperability, service mobility, and other features critical to the successful delivery of cloud computing services.

Cloud Computing Overview

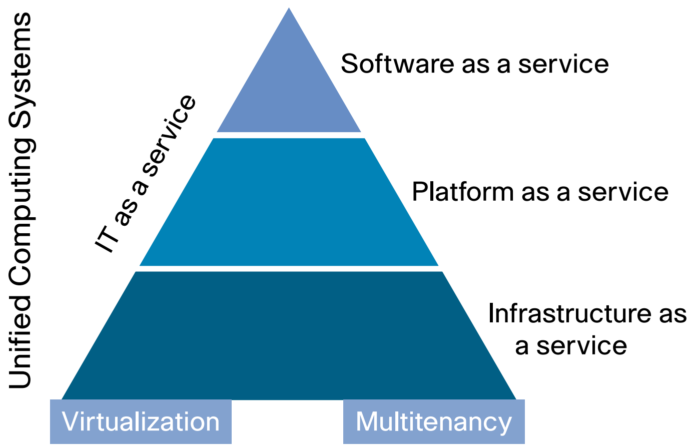
Since the creation of the government-sector ARPANET led to the public domain Internet, a wide range of network connectivity options has evolved for both organizations and consumers. In the last few years, a growing movement has promoted the concept of “cloud computing” to make network, computing, and storage resources available on demand and as easily accessible as electricity. Cloud computing is envisioned as a way to democratize access to resources and services, letting users efficiently purchase as much as they need and can afford. In this sense, some have referred to cloud computing as the fourth utility after water, electricity, and telephone service.

Grid computing – the ability to access computing and storage from a pool of resources, such as multiple autonomous systems – gained prominence beginning in the early 1990s, often for massive academic or scientific computational applications. Cloud computing is an evolution of the same concept; the linking of backend resources to provide web-based services, applications, and storage. Recently, cloud computing has spawned several subcategories, Figure 1, including:

- **IT as a service (ITaaS)** is a service model where an organization or individual contracts with a service provider to obtain network connectivity and either individual or bundled services (for example, network backup, disaster recovery, VPN, VoIP, hosting, video surveillance, and web conferencing)
- **Software as a service (SaaS)** enables service subscribers to access a software application from a software vendor through the web. The SaaS provider hosts and operates the application. Customers do not pay to own the software but instead only pay to use it through a web API. The term SaaS has replaced the older designation for these software vendors, application service providers (ASPs).
- **Platform as a service (PaaS)** makes raw computing power and disk space available from a platform of resources in the network cloud. A recent example is the Google App Engine, a developer tool that enables developers to create scalable web applications and run them on Google’s infrastructure (including 500 MB of persistent storage and bandwidth and CPU to enable five million monthly page views).

- **Infrastructure as a service (IaaS)** refers to the delivery of a virtual computer infrastructure environment as a service. Instead of purchasing servers, software, data center resources, network equipment, and the expertise to operate them, customers can buy these resources as an outsourced service delivered through the network cloud.

Figure 1. Cloud Computing Services



Companies such as Google and Amazon are offering cloud computing services such as Google's App Engine and Amazon's Elastic Compute Cloud 2 (EC2) and Elastic Block Store (EBS). Large data centers provide the infrastructure behind the cloud and the key technologies of virtualization and multi-tenancy make cloud computing resources more efficient and cost-effective for both providers and customers.

- **Virtualization** allows network resources to be available as virtual segments, with devices or portions of resources such as storage repositories accessible as needed, independent of their physical location or physical connection to the network.
- **Multi-tenancy** refers to the architectural principle allowing the sharing of resources (such as software, computing power, and storage repositories) and costs among a large pool of users. The virtual separation of those resources provides enterprise customers with private virtual domains that may provide access to different information and services by departments as well as shared services between departments. Public clouds can provide total isolation of information and services for individual customers.

Cloud Computing Incentives

Cloud computing is coming to prominence due to a convergence of technology and economic developments. Clouds are gaining mindshare and have the potential to be a disruptive technology that will transform service provider business models.

Cloud computing incentives for service providers include:

- Software and middleware applications have evolved so they can take advantage of virtualized resources.
- Cheaper and more scalable computing power allows for the creation of large virtual machines that free customers from having to acquire and provision hardware and instead allows them to provision these virtual machines to run on the hardware within the service provider cloud. Customers can run multiple instances of their virtual machines on different hardware in the cloud or the service provider can scale existing machines by adding more CPU.
- Services delivered through the cloud provide competitive differentiation from other types of hosted and managed services.

- Cloud computing provides the platform for standardized managed services that can be sold into vertical market niches, including smaller customers.
- Cloud computing incentives for customers include:
- Depending upon the type of offering, reduced complexity related to the support of hardware and software components
- Lower total cost of ownership (TCO) with a pay-as-you-go model that gives customers the flexibility to start small and ramp up as required without an initial capital outlay
- Faster and easier acquisition of new services to speed time to market

Standards, Middleware, Interoperability, and Instrumentation

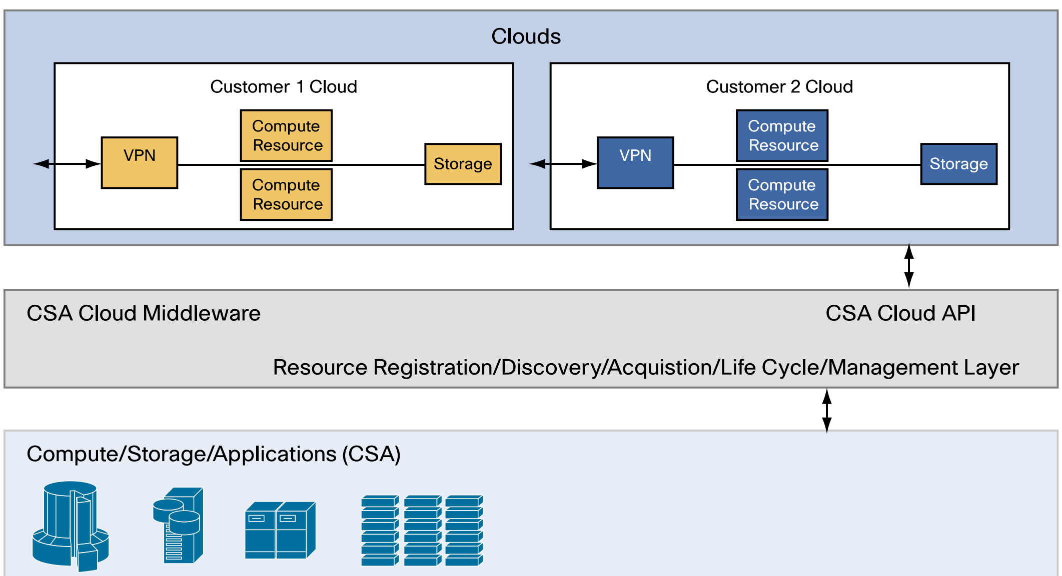
Despite the excitement over cloud computing and its associated services, there are still concerns over creating agreed-upon industry standards and also providing guaranteed service levels. A 2008 IDC survey of 244 IT executives found that 75 percent cited security as a significant challenge with cloud computing. Performance and availability were the next two concerns for 63 percent of respondents.

Standards for security, application mobility between cloud platforms, service-level agreements (SLAs), and other features must be accepted by the cloud industry before customers will feel truly comfortable moving to the cloud model. These standards will enable customers to mix and match applications, platforms, and resources as needed.

Main elements of cloud computing in need of standardization include:

- **Cloud middleware**, also referred to as Cloud OS, is the major system that manages and controls services, Figure 2. Google App Engine and Amazon EC2/S3 are examples of cloud middleware. Using cloud middleware, users should be able to create cloud instances, acquire resources, and perform general resource lifecycle management on demand.

Figure 2. High-Level Cloud Middleware Architecture Example



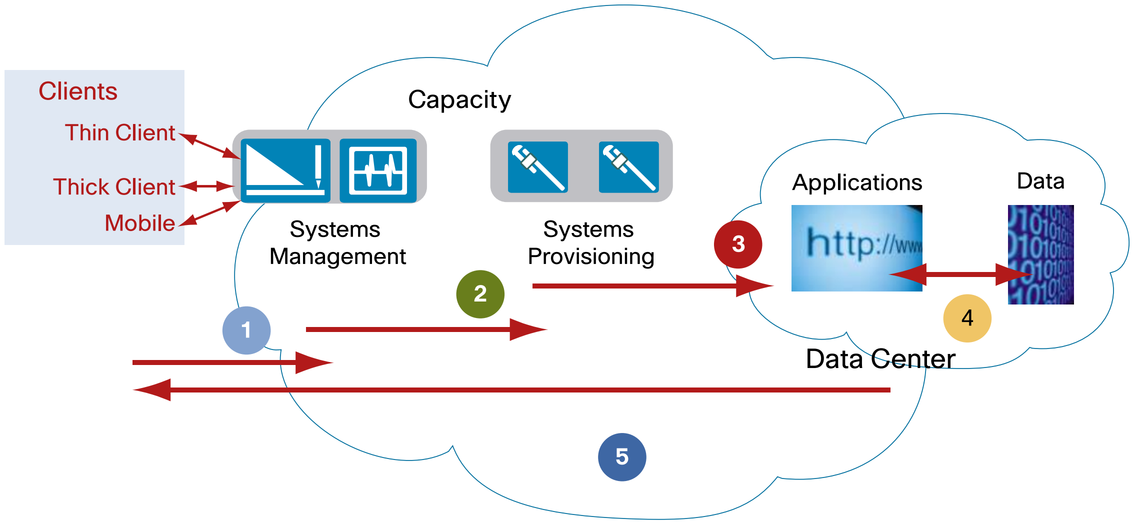
- **APIs** for applications, acquisition of resources such as computing power and storage, and machine image management must be available to make applications suitable for network clouds. Currently, most cloud vendors maintain proprietary APIs that do not allow for the sharing of resources between different clouds. To date, the role of many network vendors has been relegated to providing basic plumbing. There is a market opportunity to more tightly integrate the underlying infrastructure using APIs.

- **Resource management** is a key area requiring development in cloud computing architectures. Typically, network resources are statically provisioned but computing, storage, and application resources in clouds must be capable of being provisioned dynamically and on demand. Additionally, network resources must be capable of being provisioned separately from application resources. In a typical enterprise IT environment, there are many administrative domains in separate silos. Each domain may operate in isolation. But in a cloud environment (where cloud instances must be created on demand and dynamically with minimum turnaround time) the administrative boundaries become a major issue that can increase the provisioning time dramatically. The resource management process or workflow in the cloud must be fully automated across the administrative boundaries, with minimum turnaround time.
- **Virtualization technology** has been around for several years. Adoption within data centers and by service providers is increasing rapidly. Different proprietary virtualization technologies exist and this lack of standardization poses a barrier to an open standards cloud that is interoperable with other clouds and a broad array of computing and information resources.
- **Interoperability between clouds** will require the equivalent of a standardized cloud interoperability control plane that will enable sharing and exchange of cloud resources and communications between clouds that may be owned by multiple service providers. One example is a control plane incorporated into a cloud middleware layer or through Border Gateway Protocol (BGP) extensions. Using protocols such as Extensible Messaging and Presence Protocol (XMPP) over an agreed-upon interface between different service provider clouds enables multiple clouds to check naming and presence details and user policy permissions to provide interoperable services. Cloud 1, for example, could learn what services (such as SPARQL Protocol and RDF Query Language and OWL) are available on Cloud 2 and if these services match a particular user's requirements. Upon receiving a response, Cloud 1 could determine if a user's requirements are met and understand how to hand over the user to Cloud 2. Cloud 1 would also know how to request, provision, and call services in Cloud 2 for the user and how to pass on the user requirements such as billing tariff and SLA. So a mobile user accessing email through a Microsoft Exchange server could be migrated to Cloud 2 without any disruption or changes to the original service agreement.
- **Cross-layer dynamic policy control** is required of applications that run in a cloud as they are in enterprises. When a user creates a cloud instance using a cloud computing service, the user should be able to associate relevant policies for resources within the cloud instance. Network resources must similarly be policy controlled in alignment with application policies.

A simplified vision of the cloud computing architecture, Figure 3, includes this basic flow of information:

1. Client sends service requests
2. System management finds correct resources
3. Systems provisioning finds correct resources
4. Computing resources are found and service request is executed
5. Results of the service requests are sent to the clients

Figure 3. Cloud Computing Workflow



The End-User Perspective

A February 2009 study by the Electrical Engineering and Computer Sciences Department at the University of California, Berkeley, “Above the Clouds: A Berkeley View of Cloud Computing” (a link to the paper is at the end of this white paper), looked at cloud computing from the end-user perspective. It focused on three features that will be significant draws for customers:

- The illusion of infinite computer resources available on demand that eliminate the need for customers to plan far ahead for provisioning
- No need for an up-front commitment by customers, allowing them to start small and increase their use of cloud services as needed
- The pay-as-you-go model that allows customers to buy just what they need, either on a short-term or ongoing basis

The study also emphasized that companies with large batch-oriented computing tasks can get results more quickly and cost-effectively than ever before when using the broad resources of a managed cloud services provider for a short period of time. The on-demand nature of cloud computing is also a boon to companies, allowing them to quickly react to changing market conditions and opportunities. Another proof point for the cost-effectiveness of cloud computing is the estimate that large data centers can decrease the costs for electricity, network bandwidth, operations, software, and hardware by a factor of five. These savings can be shared among the service provider and customers and contribute to green computing trends.

The U.C. Berkeley study also presented examples of applications that are well suited to cloud computing, including:

- **Mobile, interactive applications** that respond in real time to information provided by users or sensors or both must be highly available and rely on large data sets that are most conveniently hosted in large data centers. Services that combine two or more data sources or other services are a good example of interactive applications. The cloud environment is an excellent architecture for these applications, especially for mobile devices that are connected to the cloud nearly all the time.
- **Parallel batch processing** is uniquely suited to cloud computing because users can take advantage of the ability to utilize hundreds or thousands of computers for a short period of time to get the job done.
- **Analytics** is another computing-intensive activity that can be well served in the network cloud. The U.C. Berkeley study noted that a growing share of computing resources are being spent understanding customers, buying habits, and other factors through business analytics.

- **Computing-intensive desktop applications** such as symbolic mathematics that involve a lot of computing per unit of data, image rendering, and 3D animation can be offloaded to the extensive resources of a cloud computing environment (in this case a private cloud) served by a large data center.

Cloud Computing Service Opportunities

Service providers have a major opportunity to provide cloud computing services to customers within organizations of all sizes. According to a 2008 study of cloud computing by Gartner, the cloud computing phenomenon will benefit medium-sized as well as large service providers as virtualization and automation technologies bring services on par for most needs. The costs for cloud computing will go down based on economies of scale, encouraging many businesses to adopt the service model.

Gartner has reported on the growth of SaaS among small businesses due to the economies of scale (for example, smaller organizations now pay more per seat for email services when deployed on-premises as compared to the cloud-based service). Medium-sized companies are expected to adopt SaaS and other cloud computing services in the next few years and by 2012 the cloud model will appeal to the largest enterprises with more than 100,000 users. Those companies will look at SaaS and other cloud computing options to provide segmented user access options, perhaps providing less demanding users with the cloud computing option while offering other users on-premises data center resources and network features.

Much like triple-play service bundles offered to consumers, cloud computing bundles will be capable of bringing a wide array of services and applications to individuals and businesses, including:

- **SaaS opportunities:** Customer relationship management (CRM) and enterprise resource planning (ERP) applications, email, web conferencing, digital content creation
- **ITaaS opportunities:** Storage, backup, unified threat management, security posture analysis, compliance
- **IaaS and PaaS:** Disk space, raw computing power for testing and development

Cisco and Cloud Computing

In general, cloud services must support virtual network resources, expose the resources through APIs, allow manipulation of the resources on demand through APIs, provide some level of application resource visibility within the network, and provide comprehensive and dynamic policy support across the ISO/OSI network layers. Cisco is to providing these capabilities in its data center products, as announced with Cisco Unified Computing Systems

A Cisco technology-powered cloud must support stringent security, privacy, performance, and high-availability requirements for the SLAs for cloud computing. Computing, storage resources, and network resources must be capable of being acquired on demand. Network resources must be virtualized and exposed through APIs so that those resources can be acquired on demand in the process of cloud instance creation.

As the cloud computing phenomenon grows, it will become more apparent that “over-the-top” providers of applications and other resources, such as Google and Amazon, lack the infrastructure needed to provide secure, carrier-class reliability end to end. The Cisco IP NGN, the premier Cisco architecture for service provider networks, has the transport intelligence at network, service, and application layers that will give service providers who deploy managed cloud computing services based on the Cisco IP NGN a competitive advantage, with best-in-class, carrier-class features.

For example, a service provider can offer an on-demand SLA that encompasses delay, jitter, loss, VPN capabilities, and availability. Service-oriented metrics of an SLA could include resources and availability related to each managed service. Application-oriented metrics of an SLA could include application availability, response time, QoS requirements, transactions per second, and other metrics measured end-to-end. These metrics can all be measured and incorporated into SLA with the architecture and many technologies of the Cisco IP NGN, far surpassing the capabilities of over-the-top providers who rely on a "best effort" approach using internetworking infrastructures that they do not control.

Summary

The cloud computing phenomenon is gaining popularity because of its lower TCO, scalability, competitive differentiation, reduced complexity for customers, and faster and easier acquisition of services, even as its infrastructure and standards continue to evolve. Many types of services can be delivered through the network cloud to augment the infrastructure of service providers and end customers.

Cisco is readying for the broader acceptance of cloud computing with an extremely intelligent and flexible network architecture for service providers – the Cisco IP NGN. Together with world-class products and technologies, the Cisco IP NGN allows service providers to cost-effectively utilize their existing infrastructure in new and exciting ways to incorporate cloud assets, fashion new and innovative service bundles, and gain new revenue.

For More Information

Contact your Cisco account manager for more information on the Cisco IP NGN and evolving cloud computing architectures or visit:

Cisco CIO Padmasree Warrior on Cloud Computing
http://blogs.cisco.com/news/comments/cisco_cto_on_cloud_computing/

Services in a Cloud Computing Environment
http://blogs.cisco.com/datacenter/comments/services_in_a_cloud_computing_environment/

Economic Downturn Driving Cloud Computing Evolution
http://blogs.cisco.com/datacenter/comments/economic_downturn_driving_cloud_computing_evolution/

Cisco IP Next-Generation Network
http://www.cisco.com/go/ipngn

"Above the Clouds: A Berkeley View of Cloud Computing", Electrical Engineering and Computer Sciences Department, University of California, Berkeley, February 2009
http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-532553-00 04/09

