Chapter 2

Web Services Delivered from the Cloud

2.1 Chapter Overview

In this chapter we will examine some of the web services delivered from the cloud. We will take a look at Communication-as-a-Service (CaaS) and explain some of the advantages of using CaaS. Infrastructure is also a service in cloud land, and there are many variants on how infrastructure is managed in cloud environments. When vendors outsource Infrastructure-as-a-Service (IaaS), it relies heavily on modern on-demand computing technology and high-speed networking. We will look at some vendors who provide Software-as-a-Service (SaaS), such as Amazon.com with their elastic cloud platform, and foray into the implementation issues, the characteristics, benefits, and architectural maturity level of the service. Outsourced hardware environments (called platforms) are available as Platforms-as-a-Service (PaaS), and we will look at Mosso (Rackspace) and examine key characteristics of their PaaS implementation.

As technology migrates from the traditional on-premise model to the new cloud model, service offerings evolve almost daily. Our intent in this chapter is to provide some basic exposure to where the field is currently from the perspective of the technology and give you a feel for where it will be in the not-too-distant future.

Web service offerings often have a number of common characteristics, such as a low barrier to entry, where services are offered specifically for consumers and small business entities. Often, little or no capital expenditure for infrastructure is required from the customer. While massive scalability is common with these types of offerings, it not always necessary. Many cloud vendors have yet to achieve massive scalability because their user base generally does not require it. Multitenancy enables cost and resource sharing across the (often vast) user base. Finally, device and location independence enables users to access systems regardless of where they are or what device they are using. Now, let's examine some of the more common web service offerings.

2.2 Communication-as-a-Service (CaaS)

CaaS is an outsourced enterprise communications solution. Providers of this type of cloud-based solution (known as CaaS vendors) are responsible for the management of hardware and software required for delivering Voice over IP (VoIP) services, Instant Messaging (IM), and video conferencing capabilities to their customers. This model began its evolutionary process from within the telecommunications (Telco) industry, not unlike how the SaaS model arose from the software delivery services sector. CaaS vendors are responsible for all of the hardware and software management consumed by their user base. CaaS vendors typically offer guaranteed quality of service (QoS) under a service-level agreement (SLA).

A CaaS model allows a CaaS provider's business customers to selectively deploy communications features and services throughout their company on a pay-as-you-go basis for service(s) used. CaaS is designed on a utility-like pricing model that provides users with comprehensive, flexible, and (usually) simple-to-understand service plans. According to Gartner,[1] the CaaS market is expected to total $2.3 billion in 2011, representing a compound annual growth rate of more than 105% for the period.

CaaS service offerings are often bundled and may include integrated access to traditional voice (or VoIP) and data, advanced unified communications functionality such as video calling, web collaboration, chat, real-time presence and unified messaging, a handset, local and long-distance voice services, voice mail, advanced calling features (such as caller ID, three-way and conference calling, etc.) and advanced PBX functionality. A CaaS solution includes redundant switching, network, POP and circuit diversity, customer premises equipment redundancy, and WAN fail-over that specifically addresses the needs of their customers. All VoIP transport components are located in geographically diverse, secure data centers for high availability and survivability.

CaaS offers flexibility and scalability that small and medium-sized business might not otherwise be able to afford. CaaS service providers are usually prepared to handle peak loads for their customers by providing services

1. Gartner Press Release, "Gartner Forecasts Worldwide Communications-as-a-Service Revenue to Total $252 Million in 2007," August 2007, retrieved 13 Jan 2009.

capable of allowing more capacity, devices, modes or area coverage as their customer demand necessitates. Network capacity and feature sets can be changed dynamically, so functionality keeps pace with consumer demand and provider-owned resources are not wasted. From the service provider customer's perspective, there is very little to virtually no risk of the service becoming obsolete, since the provider's responsibility is to perform periodic upgrades or replacements of hardware and software to keep the platform technologically current.

CaaS requires little to no management oversight from customers. It eliminates the business customer's need for any capital investment in infrastructure, and it eliminates expense for ongoing maintenance and operations overhead for infrastructure. With a CaaS solution, customers are able to leverage enterprise-class communication services without having to build a premises-based solution of their own. This allows those customers to reallocate budget and personnel resources to where their business can best use them.

2.2.1 Advantages of CaaS

From the handset found on each employee's desk to the PC-based software client on employee laptops, to the VoIP private backbone, and all modes in between, every component in a CaaS solution is managed 24/7 by the CaaS vendor. As we said previously, the expense of managing a carrier-grade data center is shared across the vendor's customer base, making it more economical for businesses to implement CaaS than to build their own VoIP network. Let's look as some of the advantages of a hosted approach for CaaS.

Hosted and Managed Solutions

Remote management of infrastructure services provided by third parties once seemed an unacceptable situation to most companies. However, over the past decade, with enhanced technology, networking, and software, the attitude has changed. This is, in part, due to cost savings achieved in using those services. However, unlike the "one-off" services offered by specialist providers, CaaS delivers a complete communications solution that is entirely managed by a single vendor. Along with features such as VoIP and unified communications, the integration of core PBX features with advanced functionality is managed by one vendor, who is responsible for all of the integration and delivery of services to users.

2.2.2 Fully Integrated, Enterprise-Class Unified Communications

With CaaS, the vendor provides voice and data access and manages LAN/WAN, security, routers, email, voice mail, and data storage. By managing the LAN/WAN, the vendor can guarantee consistent quality of service from a user's desktop across the network and back. Advanced unified communications features that are most often a part of a standard CaaS deployment include:

- Chat
- Multimedia conferencing
- Microsoft Outlook integration
- Real-time presence
- "Soft" phones (software-based telephones)
- Video calling
- Unified messaging and mobility

Providers are constantly offering new enhancements (in both performance and features) to their CaaS services. The development process and subsequent introduction of new features in applications is much faster, easier, and more economical than ever before. This is, in large part, because the service provider is doing work that benefits many end users across the provider's scalable platform infrastructure. Because many end users of the provider's service ultimately share this cost (which, from their perspective, is miniscule compared to shouldering the burden alone), services can be offered to individual customers at a cost that is attractive to them.

No Capital Expenses Needed

When business outsource their unified communications needs to a CaaS service provider, the provider supplies a complete solution that fits the company's exact needs. Customers pay a fee (usually billed monthly) for what they use. Customers are not required to purchase equipment, so there is no capital outlay. Bundled in these types of services are ongoing maintenance and upgrade costs, which are incurred by the service provider. The use of CaaS services allows companies the ability to collaborate across any workspace. Advanced collaboration tools are now used to create high-quality, secure, adaptive work spaces throughout any organization. This allows a company's workers, partners, vendors, and customers to communicate and collaborate more effectively. Better communication allows organizations to adapt quickly to market changes and to build competitive advantage. CaaS can also accelerate decision making within an organization. Innovative unified communications capabilities (such as presence, instant messaging, and rich media services) help ensure that information quickly reaches whoever needs it.

Flexible Capacity and Feature Set

When customers outsource communications services to a CaaS provider, they pay for the features they need when they need them. The service provider can distribute the cost services and delivery across a large customer base. As previously stated, this makes the use of shared feature functionality more economical for customers to implement. Economies of scale allow service providers enough flexibility that they are not tied to a single vendor investment. They are able to leverage best-of-breed providers such as Avaya, Cisco, Juniper, Microsoft, Nortel and ShoreTel more economically than any independent enterprise.

No Risk of Obsolescence

Rapid technology advances, predicted long ago and known as Moore's law,[2] have brought about product obsolescence in increasingly shorter periods of time. Moore's law describes a trend he recognized that has held true since the beginning of the use of integrated circuits (ICs) in computing hardware. Since the invention of the integrated circuit in 1958, the number of transistors that can be placed inexpensively on an integrated circuit has increased exponentially, doubling approximately every two years.

Unlike IC components, the average life cycles for PBXs and key communications equipment and systems range anywhere from five to 10 years. With the constant introduction of newer models for all sorts of technology (PCs, cell phones, video software and hardware, etc.), these types of products now face much shorter life cycles, sometimes as short as a single year. CaaS vendors must absorb this burden for the user by continuously upgrading the equipment in their offerings to meet changing demands in the marketplace.

2. Gordon E. Moore, "Cramming More Components onto Integrated Circuits," Electronics Magazine, 4, 1965, retrieved 1 Jan 2009.

No Facilities and Engineering Costs Incurred

CaaS providers host all of the equipment needed to provide their services to their customers, virtually eliminating the need for customers to maintain data center space and facilities. There is no extra expense for the constant power consumption that such a facility would demand. Customers receive the benefit of multiple carrier-grade data centers with full redundancy—and it's all included in the monthly payment.

Guaranteed Business Continuity

If a catastrophic event occurred at your business's physical location, would your company disaster recovery plan allow your business to continue operating without a break? If your business experienced a serious or extended communications outage, how long could your company survive? For most businesses, the answer is "not long." Distributing risk by using geographically dispersed data centers has become the norm today. It mitigates risk and allows companies in a location hit by a catastrophic event to recover as soon as possible. This process is implemented by CaaS providers because most companies don't even contemplate voice continuity if catastrophe strikes. Unlike data continuity, eliminating single points of failure for a voice network is usually cost-prohibitive because of the large scale and management complexity of the project. With a CaaS solution, multiple levels of redundancy are built into the system, with no single point of failure.

2.3 Infrastructure-as-a-Service (IaaS)

According to the online reference Wikipedia, Infrastructure-as-a-Service (IaaS) is the delivery of computer infrastructure (typically a platform virtualization environment) as a service.[3] IaaS leverages significant technology, services, and data center investments to deliver IT as a service to customers. Unlike traditional outsourcing, which requires extensive due diligence, negotiations ad infinitum, and complex, lengthy contract vehicles, IaaS is centered around a model of service delivery that provisions a predefined, standardized infrastructure specifically optimized for the customer's applications. Simplified statements of work and à la carte service-level choices make it easy to tailor a solution to a customer's specific application requirements. IaaS providers manage the transition and hosting of selected applications on their infrastructure. Customers maintain ownership and

3. http://en.wikipedia.org/wiki/Infrastructure_as_a_Service, retrieved 11 Jan 2009.

management of their application(s) while off-loading hosting operations and infrastructure management to the IaaS provider. Provider-owned implementations typically include the following layered components:

- Computer hardware (typically set up as a grid for massive horizontal scalability)
- Computer network (including routers, firewalls, load balancing, etc.)
- Internet connectivity (often on OC 192 backbones[4])
- Platform virtualization environment for running client-specified virtual machines
- Service-level agreements
- Utility computing billing

Rather than purchasing data center space, servers, software, network equipment, etc., IaaS customers essentially rent those resources as a fully outsourced service. Usually, the service is billed on a monthly basis, just like a utility company bills customers. The customer is charged only for resources consumed. The chief benefits of using this type of outsourced service include:

- Ready access to a preconfigured environment that is generally ITIL-based[5] (The Information Technology Infrastructure Library [ITIL] is a customized framework of best practices designed to promote quality computing services in the IT sector.)
- Use of the latest technology for infrastructure equipment
- Secured, "sand-boxed" (protected and insulated) computing platforms that are usually security monitored for breaches
- Reduced risk by having off-site resources maintained by third parties
- Ability to manage service-demand peaks and valleys
- Lower costs that allow expensing service costs instead of making capital investments

4. An Optical Carrier (OC) 192 transmission line is capable of transferring 9.95 gigabits of data per second.
5. Jan Van Bon, The Guide to IT Service Management, Vol. I, New York: Addison-Wesley, 2002, p. 131.

- Reduced time, cost, and complexity in adding new features or capabilities

2.3.1 Modern On-Demand Computing

On-demand computing is an increasingly popular enterprise model in which computing resources are made available to the user as needed.[6] Computing resources that are maintained on a user's site are becoming fewer and fewer, while those made available by a service provider are on the rise. The on-demand model evolved to overcome the challenge of being able to meet fluctuating resource demands efficiently. Because demand for computing resources can vary drastically from one time to another, maintaining sufficient resources to meet peak requirements can be costly. Overengineering a solution can be just as adverse as a situation where the enterprise cuts costs by maintaining only minimal computing resources, resulting in insufficient resources to meet peak load requirements. Concepts such as clustered computing, grid computing, utility computing, etc., may all seem very similar to the concept of on-demand computing, but they can be better understood if one thinks of them as building blocks

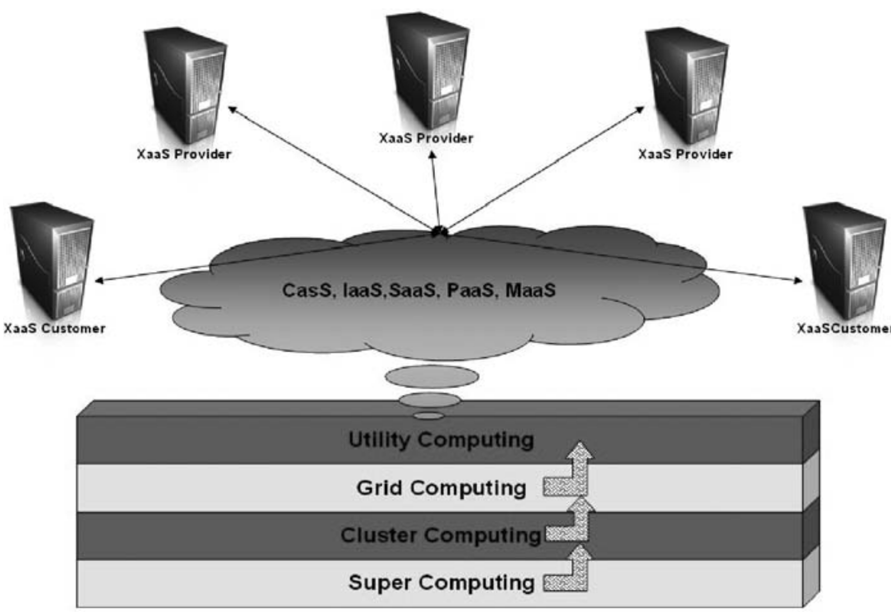

Figure 2.1 Building blocks to the cloud

6. http://searchdatacenter.techtarget.com/sDefinition/0,,sid80_gci903730,00.html#, retrieved 15 Jan 2009.

that evolved over time and with techno-evolution to achieve the modern cloud computing model we think of and use today (see Figure 2.1).

One example we will examine is Amazon's Elastic Compute Cloud (Amazon EC2). This is a web service that provides resizable computing capacity in the cloud. It is designed to make web-scale computing easier for developers and offers many advantages to customers:

- It's web service interface allows customers to obtain and configure capacity with minimal effort.
- It provides users with complete control of their (leased) computing resources and lets them run on a proven computing environment.
- It reduces the time required to obtain and boot new server instances to minutes, allowing customers to quickly scale capacity as their computing demands dictate.
- It changes the economics of computing by allowing clients to pay only for capacity they actually use.
- It provides developers the tools needed to build failure-resilient applications and isolate themselves from common failure scenarios.

2.3.2 Amazon's Elastic Cloud

Amazon EC2 presents a true virtual computing environment, allowing clients to use a web-based interface to obtain and manage services needed to launch one or more instances of a variety of operating systems (OSs). Clients can load the OS environments with their customized applications. They can manage their network's access permissions and run as many or as few systems as needed. In order to use Amazon EC2, clients first need to create an Amazon Machine Image (AMI). This image contains the applications, libraries, data, and associated configuration settings used in the virtual computing environment. Amazon EC2 offers the use of preconfigured images built with templates to get up and running immediately. Once users have defined and configured their AMI, they use the Amazon EC2 tools provided for storing the AMI by uploading the AMI into Amazon S3. Amazon S3 is a repository that provides safe, reliable, and fast access to a client AMI. Before clients can use the AMI, they must use the Amazon EC2 web service to configure security and network access.

Using Amazon EC2 to Run Instances

During configuration, users choose which instance type(s) and operating system they want to use. Available instance types come in two distinct categories, Standard or High-CPU instances. Most applications are best suited for Standard instances, which come in small, large, and extra-large instance platforms. High-CPU instances have proportionally more CPU resources than random-access memory (RAM) and are well suited for compute-intensive applications. With the High-CPU instances, there are medium and extra large platforms to choose from. After determining which instance to use, clients can start, terminate, and monitor as many instances of their AMI as needed by using web service Application Programming Interfaces (APIs) or a wide variety of other management tools that are provided with the service. Users are able to choose whether they want to run in multiple locations, use static IP endpoints, or attach persistent block storage to any of their instances, and they pay only for resources actually consumed. They can also choose from a library of globally available AMIs that provide useful instances. For example, if all that is needed is a basic Linux server, clients can choose one of the standard Linux distribution AMIs.

2.3.3 Amazon EC2 Service Characteristics

There are quite a few characteristics of the EC2 service that provide significant benefits to an enterprise. First of all, Amazon EC2 provides financial benefits. Because of Amazon's massive scale and large customer base, it is an inexpensive alternative to many other possible solutions. The costs incurred to set up and run an operation are shared over many customers, making the overall cost to any single customer much lower than almost any other alternative. Customers pay a very low rate for the compute capacity they actually consume. Security is also provided through Amazon EC2 web service interfaces. These allow users to configure firewall settings that control network access to and between groups of instances. Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly provisioned.

When one compares this solution to the significant up-front expenditures traditionally required to purchase and maintain hardware, either in-house or hosted, the decision to outsource is not hard to make. Outsourced solutions like EC2 free customers from many of the complexities of capacity planning and allow clients to move from large capital investments and fixed costs to smaller, variable, expensed costs. This approach removes the need to overbuy and overbuild capacity to handle periodic traffic spikes. The EC2 service runs within Amazon's proven, secure, and reliable network infrastructure and data center locations.

Dynamic Scalability

Amazon EC2 enables users to increase or decrease capacity in a few minutes. Users can invoke a single instance, hundreds of instances, or even thousands of instances simultaneously. Of course, because this is all controlled with web service APIs, an application can automatically scale itself up or down depending on its needs. This type of dynamic scalability is very attractive to enterprise customers because it allows them to meet their customers' demands without having to overbuild their infrastructure.

Full Control of Instances

Users have complete control of their instances. They have root access to each instance and can interact with them as one would with any machine. Instances can be rebooted remotely using web service APIs. Users also have access to console output of their instances. Once users have set up their account and uploaded their AMI to the Amazon S3 service, they just need to boot that instance. It is possible to start an AMI on any number of instances (or any type) by calling the *RunInstances* API that is provided by Amazon.

Configuration Flexibility

Configuration settings can vary widely among users. They have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows them to select a configuration of memory, CPU, and instance storage that is optimal for their choice of operating system and application. For example, a user's choice of operating systems may also include numerous Linux distributions, Microsoft Windows Server, and even an OpenSolaris environment, all running on virtual servers.

Integration with Other Amazon Web Services

Amazon EC2 works in conjunction with a variety of other Amazon web services. For example, Amazon Simple Storage Service (Amazon S3), Amazon SimpleDB, Amazon Simple Queue Service (Amazon SQS), and Amazon CloudFront are all integrated to provide a complete solution for computing, query processing, and storage across a wide range of applications.

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

Amazon SimpleDB is another web-based service, designed for running queries on structured data stored with the Amazon Simple Storage Service (Amazon S3) in real time. This service works in conjunction with the Amazon Elastic Compute Cloud (Amazon EC2) to provide users the capability to store, process, and query data sets within the cloud environment. These services are designed to make web-scale computing easier and more cost-effective for developers. Traditionally, this type of functionality was provided using a clustered relational database that requires a sizable investment. Implementations of this nature brought on more complexity and often required the services of a database administer to maintain it.

By comparison to traditional approaches, Amazon SimpleDB is easy to use and provides the core functionality of a database (e.g., real-time lookup and simple querying of structured data) without inheriting the operational complexity involved in traditional implementations. Amazon SimpleDB requires no schema, automatically indexes data, and provides a simple API for data storage and access. This eliminates the need for customers to perform tasks such as data modeling, index maintenance, and performance tuning.

Amazon Simple Queue Service (Amazon SQS) is a reliable, scalable, hosted queue for storing messages as they pass between computers. Using Amazon SQS, developers can move data between distributed components of applications that perform different tasks without losing messages or requiring 100% availability for each component. Amazon SQS works by exposing Amazon's web-scale messaging infrastructure as a service. Any computer connected to the Internet can add or read messages without the need for having any installed software or special firewall configurations. Components of applications using Amazon SQS can run independently and do not need to be on the same network, developed with the same technologies, or running at the same time.

Amazon CloudFront is a web service for content delivery. It integrates with other Amazon web services to distribute content to end users with low latency and high data transfer speeds. Amazon CloudFront delivers content using a global network of edge locations. Requests for objects are automatically routed to the nearest edge server, so content is delivered with the best possible performance. An edge server receives a request from the user's computer and makes a connection to another computer called the origin server, where the application resides. When the origin server fulfills the request, it sends the application's data back to the edge server, which, in turn, forwards the data to the client computer that made the request.

Reliable and Resilient Performance

Amazon Elastic Block Store (EBS) is yet another Amazon EC2 feature that provides users powerful features to build failure-resilient applications. Amazon EBS offers persistent storage for Amazon EC2 instances. Amazon EBS volumes provide "off-instance" storage that persists independently from the life of any instance. Amazon EBS volumes are highly available, highly reliable data shares that can be attached to a running Amazon EC2 instance and are exposed to the instance as standard block devices. Amazon EBS volumes are automatically replicated on the back end. The service provides users with the ability to create point-in-time snapshots of their data volumes, which are stored using the Amazon S3 service. These snapshots can be used as a starting point for new Amazon EBS volumes and can protect data indefinitely.

Support for Use in Geographically Disparate Locations

Amazon EC2 provides users with the ability to place one or more instances in multiple locations. Amazon EC2 locations are composed of Regions (such as North America and Europe) and Availability Zones. Regions consist of one or more Availability Zones, are geographically dispersed, and are in separate geographic areas or countries. Availability Zones are distinct locations that are engineered to be insulated from failures in other Availability Zones and provide inexpensive, low-latency network connectivity to other Availability Zones in the same Region.[7] For example, the North America Region may be split into the following Availability Zones: Northeast, East, SouthEast, NorthCentral, Central, SouthCentral, NorthWest, West, SouthWest, etc. By launching instances in any one or more of the separate Availability Zones, you can insulate your applications from a single point of failure. Amazon EC2 has a service-level agreement that commits to a 99.95% uptime availability for each Amazon EC2 Region. Amazon EC2 is currently available in two regions, the United States and Europe.

7. http://developer.amazonwebservices.com/connect/entry.jspa?externalID=1347, retrieved 16 Jan 2009.

Elastic IP Addressing

Elastic IP (EIP) addresses are static IP addresses designed for dynamic cloud computing. An Elastic IP address is associated with your account and not with a particular instance, and you control that address until you choose explicitly to release it. Unlike traditional static IP addresses, however, EIP addresses allow you to mask instance or Availability Zone failures by programmatically remapping your public IP addresses to any instance in your account. Rather than waiting on a technician to reconfigure or replace your host, or waiting for DNS to propagate to all of your customers, Amazon EC2 enables you to work around problems that occur with client instances or client software by quickly remapping their EIP address to another running instance. A significant feature of Elastic IP addressing is that each IP address can be reassigned to a different instance when needed. Now, let's review how the Elastic IPs work with Amazon EC2 services.

First of all, Amazon allows users to allocate up to five Elastic IP addresses per account (which is the default). Each EIP can be assigned to a single instance. When this reassignment occurs, it replaces the normal dynamic IP address used by that instance. By default, each instance starts with a dynamic IP address that is allocated upon startup. Since each instance can have only one external IP address, the instance starts out using the default dynamic IP address. If the EIP in use is assigned to a different instance, a new dynamic IP address is allocated to the vacated address of that instance. Assigning or reassigning an IP to an instance requires only a few minutes. The limitation of designating a single IP at a time is due to the way Network Address Translation (NAT) works. Each instance is mapped to an internal IP address and is also assigned an external (public) address. The public address is mapped to the internal address using Network Address Translation tables (hence, NAT). If two external IP addresses happen to be translated to the same internal IP address, all inbound traffic (in the form of data packets) would arrive without any issues. However, assigning outgoing packets to an external IP address would be very difficult because a determination of which external IP address to use could not be made. This is why implementors have built in the limitation of having only a single external IP address per instance at any one time.

2.3.4 Mosso (Rackspace)

Mosso, a direct competitor of Amazon's EC2 service, is a web application hosting service and cloud platform provider that bills on a utility computing basis. Mosso was launched in February 2008 and is owned and operated by Rackspace, a web hosting provider that has been around for some time. Most new hosting platforms require custom code and architecture to make an application work. What makes Mosso different is that it has been designed to run an application with very little or no modifications. The Mosso platform is built on existing web standards and powered by proven technologies. Customers reap the benefits of a scalable platform for free. They spend no time coding custom APIs or building data schemas. Mosso has also branched out into cloud storage and cloud infrastructure.

Mosso Cloud Servers and Files

Mosso Cloud Servers (MCS) came into being from the acquisition of a company called Slicehost by Rackspace. Slicehost was designed to enable deployment of multiple cloud servers instantly. In essence, it touts capability for the creation of advanced, high-availability architectures. In order to create a full-service offering, Rackspace also acquired another company, JungleDisk. JungleDisk was an online backup service. By integrating JungleDisk's backup features with virtual servers that Slicehost provides, Mosso, in effect, created a new service to compete with Amazon's EC2. Mosso claims that these "cloud sites" are the fastest way for scustomer to put their site in the cloud. Cloud sites are capable of running Windows or Linux applications across banks of servers numbering in the hundreds.

Mosso's *Cloud Files* provide unlimited storage for content by using a partnership formed with Limelight Networks. This partnership allows Mosso to offer its customers a content delivery network (CDN). With CDN services, servers are placed around the world and, depending on where you are located, you get served via the closest or most appropriate server. CDNs cut down on the hops back and forth to handle a request. The chief benefit of using CDN is a scalable, dynamic storage platform that offers a metered service by which customers pay only for what they use. Customers can manage files through a web-based control panel or programmatically through an API.

Integrated backups with the CDN offering implemented in the Mosso services platform began in earnest with Jungle Disk version 2.5 in early 2009. Jungle Disk 2.5 is a major upgrade, adding a number of highly requested features to its portfolio. Highlights of the new version include running as a background service. The background service will keep running even if the Jungle Disk Monitor is logged out or closed. Users do not have to be logged into the service for automatic backups to be performed. There is native file system support on both 32-bit and 64-bit versions of Windows (Windows 2000, XP, Vista, 2003 and 2008), and Linux. A new download resume capability has been added for moving large files and performing restore operations. A time-slice restore interface was also added, allowing restoration of files from any given point-in-time where a snapshot was taken. Finally, it supports automatic updates on Windows (built-in) and Macintosh (using Sparkle).

2.4 Monitoring-as-a-Service (MaaS)

Monitoring-as-a-Service (MaaS) is the outsourced provisioning of security, primarily on business platforms that leverage the Internet to conduct business.[8] MaaS has become increasingly popular over the last decade. Since the advent of cloud computing, its popularity has, grown even more. Security monitoring involves protecting an enterprise or government client from cyber threats. A security team plays a crucial role in securing and maintaining the confidentiality, integrity, and availability of IT assets. However, time and resource constraints limit security operations and their effectiveness for most companies. This requires constant vigilance over the security infrastructure and critical information assets.

Many industry regulations require organizations to monitor their security environment, server logs, and other information assets to ensure the integrity of these systems. However, conducting effective security monitoring can be a daunting task because it requires advanced technology, skilled security experts, and scalable processes—none of which come cheap. MaaS security monitoring services offer real-time, 24/7 monitoring and nearly immediate incident response across a security infrastructure—they help to protect critical information assets of their customers. Prior to the advent of electronic security systems, security monitoring and response were heavily dependent on human resources and human capabilities, which also limited the accuracy and effectiveness of monitoring efforts. Over the past two decades, the adoption of information technology into facility security systems, and their ability to be connected to security operations centers (SOCs) via corporate networks, has significantly changed that picture. This means two important things: (1) The total cost of ownership (TCO) for traditional SOCs is much higher than for a modern-technology SOC; and (2)

8. http://en.wikipedia.org/wiki/Monitoring_as_a_service, retrieved 14 Jan 2009.

achieving lower security operations costs and higher security effectiveness means that modern SOC architecture must use security and IT technology to address security risks.

2.4.1 Protection Against Internal and External Threats

SOC-based security monitoring services can improve the effectiveness of a customer security infrastructure by actively analyzing logs and alerts from infrastructure devices around the clock and in real time. Monitoring teams correlate information from various security devices to provide security analysts with the data they need to eliminate false positives[9] and respond to true threats against the enterprise. Having consistent access to the skills needed to maintain the level of service an organization requires for enterprise-level monitoring is a huge issue. The information security team can assess system performance on a periodically recurring basis and provide recommendations for improvements as needed. Typical services provided by many MaaS vendors are described below.

Early Detection

An early detection service detects and reports new security vulnerabilities shortly after they appear. Generally, the threats are correlated with third-party sources, and an alert or report is issued to customers. This report is usually sent by email to the person designated by the company. Security vulnerability reports, aside from containing a detailed description of the vulnerability and the platforms affected, also include information on the impact the exploitation of this vulnerability would have on the systems or applications previously selected by the company receiving the report. Most often, the report also indicates specific actions to be taken to minimize the effect of the vulnerability, if that is known.

Platform, Control, and Services Monitoring

Platform, control, and services monitoring is often implemented as a dashboard interface[10] and makes it possible to know the operational status of the platform being monitored at any time. It is accessible from a web interface, making remote access possible. Each operational element that is monitored usually provides an operational status indicator, always taking into account

9. A false positive is an event that is picked up by an intrusion detection system and perceived as an attack but that in reality is not.
10. A dashboard is a floating, semitransparent window that provides contextual access to commonly used tools in a software program.

the critical impact of each element. This service aids in determining which elements may be operating at or near capacity or beyond the limits of established parameters. By detecting and identifying such problems, preventive measures can be taken to prevent loss of service.

Intelligent Log Centralization and Analysis

Intelligent log centralization and analysis is a monitoring solution based mainly on the correlation and matching of log entries. Such analysis helps to establish a baseline of operational performance and provides an index of security threat. Alarms can be raised in the event an incident moves the established baseline parameters beyond a stipulated threshold. These types of sophisticated tools are used by a team of security experts who are responsible for incident response once such a threshold has been crossed and the threat has generated an alarm or warning picked up by security analysts monitoring the systems.

Vulnerabilities Detection and Management

Vulnerabilities detection and management enables automated verification and management of the security level of information systems. The service periodically performs a series of automated tests for the purpose of identifying system weaknesses that may be exposed over the Internet, including the possibility of unauthorized access to administrative services, the existence of services that have not been updated, the detection of vulnerabilities such as phishing, etc. The service performs periodic follow-up of tasks performed by security professionals managing information systems security and provides reports that can be used to implement a plan for continuous improvement of the system's security level.

Continuous System Patching/Upgrade and Fortification

Security posture is enhanced with continuous system patching and upgrading of systems and application software. New patches, updates, and service packs for the equipment's operating system are necessary to maintain adequate security levels and support new versions of installed products. Keeping abreast of all the changes to all the software and hardware requires a committed effort to stay informed and to communicate gaps in security that can appear in installed systems and applications.

Intervention, Forensics, and Help Desk Services

Quick intervention when a threat is detected is crucial to mitigating the effects of a threat. This requires security engineers with ample knowledge in the various technologies and with the ability to support applications as well as infrastructures on a 24/7 basis. MaaS platforms routinely provide this service to their customers. When a detected threat is analyzed, it often requires forensic analysis to determine what it is, how much effort it will take to fix the problem, and what effects are likely to be seen. When problems are encountered, the first thing customers tend to do is pick up the phone. Help desk services provide assistance on questions or issues about the operation of running systems. This service includes assistance in writing failure reports, managing operating problems, etc.

2.4.2 Delivering Business Value

Some consider balancing the overall economic impact of any build-versus-buy decision as a more significant measure than simply calculating a return on investment (ROI). The key cost categories that are most often associated with MaaS are (1) service fees for security event monitoring for all firewalls and intrusion detection devices, servers, and routers; (2) internal account maintenance and administration costs; and (3) preplanning and development costs.

Based on the total cost of ownership, whenever a customer evaluates the option of an in-house security information monitoring team and infrastructure compared to outsourcing to a service provider, it does not take long to realize that establishing and maintaining an in-house capability is not as attractive as outsourcing the service to a provider with an existing infrastructure. Having an in-house security operations center forces a company to deal with issues such as staff attrition, scheduling, around the clock operations, etc.

Losses incurred from external and internal incidents are extremely significant, as evidenced by a regular stream of high-profile cases in the news. The generally accepted method of valuing the risk of losses from external and internal incidents is to look at the amount of a potential loss, assume a frequency of loss, and estimate a probability for incurring the loss. Although this method is not perfect, it provides a means for tracking information security metrics. Risk is used as a filter to capture uncertainty about varying cost and benefit estimates. If a risk-adjusted ROI demonstrates a compelling business case, it raises confidence that the investment is likely to succeed because the risks that threaten the project have been considered and quantified. Flexibility represents an investment in additional capacity or agility today that can be turned into future business benefits at some additional cost. This provides an organization with the ability to engage in future initiatives, but not the obligation to do so. The value of flexibility is unique to each organization, and willingness to measure its value varies from company to company.

2.4.3 Real-Time Log Monitoring Enables Compliance

Security monitoring services can also help customers comply with industry regulations by automating the collection and reporting of specific events of interest, such as log-in failures. Regulations and industry guidelines often require log monitoring of critical servers to ensure the integrity of confidential data. MaaS providers' security monitoring services automate this time-consuming process.

2.5 Platform-as-a-Service (PaaS)

Cloud computing has evolved to include platforms for building and running custom web-based applications, a concept known as Platform-as-a-Service. PaaS is an outgrowth of the SaaS application delivery model. The PaaS model makes all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely available from the Internet, all with no software downloads or installation for developers, IT managers, or end users. Unlike the IaaS model, where developers may create a specific operating system instance with home-grown applications running, PaaS developers are concerned only with web-based development and generally do not care what operating system is used. PaaS services allow users to focus on innovation rather than complex infrastructure. Organizations can redirect a significant portion of their budgets to creating applications that provide real business value instead of worrying about all the infrastructure issues in a roll-your-own delivery model. The PaaS model is thus driving a new era of mass innovation. Now, developers around the world can access unlimited computing power. Anyone with an Internet connection can build powerful applications and easily deploy them to users globally.

2.5.1 The Traditional On-Premises Model

The traditional approach of building and running on-premises applications has always been complex, expensive, and risky. Building your own solution has never offered any guarantee of success. Each application was designed to meet specific business requirements. Each solution required a specific set of hardware, an operating system, a database, often a middleware package, email and web servers, etc. Once the hardware and software environment was created, a team of developers had to navigate complex programming development platforms to build their applications. Additionally, a team of network, database, and system management experts was needed to keep everything up and running. Inevitably, a business requirement would force the developers to make a change to the application. The changed application then required new test cycles before being distributed. Large companies often needed specialized facilities to house their data centers. Enormous amounts of electricity also were needed to power the servers as well as to keep the systems cool. Finally, all of this required use of fail-over sites to mirror the data center so that information could be replicated in case of a disaster. Old days, old ways—now, let's fly into the silver lining of todays cloud.

2.5.2 The New Cloud Model

PaaS offers a faster, more cost-effective model for application development and delivery. PaaS provides all the infrastructure needed to run applications over the Internet. Such is the case with companies such as Amazon.com, eBay, Google, iTunes, and YouTube. The new cloud model has made it possible to deliver such new capabilities to new markets via the web browsers. PaaS is based on a metering or subscription model, so users pay only for what they use. PaaS offerings include workflow facilities for application design, application development, testing, deployment, and hosting, as well as application services such as virtual offices, team collaboration, database integration, security, scalability, storage, persistence, state management, dashboard instrumentation, etc.

2.5.3 Key Characteristics of PaaS

Chief characteristics of PaaS include services to develop, test, deploy, host, and manage applications to support the application development life cycle. Web-based user interface creation tools typically provide some level of support to simplify the creation of user interfaces, based either on common standards such as HTML and JavaScript or on other, proprietary technologies. Supporting a multitenant architecture helps to remove developer concerns regarding the use of the application by many concurrent users. PaaS providers often include services for concurrency management, scalability, fail-over and security. Another characteristic is the integration with web services and databases. Support for Simple Object Access Protocol (SOAP) and other interfaces allows PaaS offerings to create combinations of web services (called mashups) as well as having the ability to access databases and reuse services maintained inside private networks. The ability to form and share code with ad-hoc, predefined, or distributed teams greatly enhances the productivity of PaaS offerings. Integrated PaaS offerings provide an opportunity for developers to have much greater insight into the inner workings of their applications and the behavior of their users by implementing dashboard-like tools to view the inner workings based on measurements such as performance, number of concurrent accesses, etc. Some PaaS offerings leverage this instrumentation to enable pay-per-use billing models.

2.6 Software-as-a-Service (SaaS)

The traditional model of software distribution, in which software is purchased for and installed on personal computers, is sometimes referred to as Software-as-a-Product. Software-as-a-Service is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet. SaaS is becoming an increasingly prevalent delivery model as underlying technologies that support web services and service-oriented architecture (SOA) mature and new developmental approaches become popular. SaaS is also often associated with a pay-as-you-go subscription licensing model. Meanwhile, broadband service has become increasingly available to support user access from more areas around the world.

The huge strides made by Internet Service Providers (ISPs) to increase bandwidth, and the constant introduction of ever more powerful microprocessors coupled with inexpensive data storage devices, is providing a huge platform for designing, deploying, and using software across all areas of business and personal computing. SaaS applications also must be able to interact with other data and other applications in an equally wide variety of environments and platforms. SaaS is closely related to other service delivery models we have described. IDC identifies two slightly different delivery models for SaaS.[11] The hosted application management model is similar to an Application Service Provider (ASP) model. Here, an ASP hosts commercially available software for customers and delivers it over the Internet. The other model is a software on demand model where the provider gives customers network-based access to a single copy of an application created specifically for SaaS distribution. IDC predicted that SaaS would make up 30% of the software market by 2007 and would be worth $10.7 billion by the end of 2009.[12]

SaaS is most often implemented to provide business software functionality to enterprise customers at a low cost while allowing those customers to obtain the same benefits of commercially licensed, internally operated software without the associated complexity of installation, management, support, licensing, and high initial cost.[13] Most customers have little interest in the how or why of software implementation, deployment, etc., but all have a need to use software in their work. Many types of software are well suited to the SaaS model (e.g., accounting, customer relationship management, email software, human resources, IT security, IT service management, video conferencing, web analytics, web content management). The distinction between SaaS and earlier applications delivered over the Internet is that SaaS solutions were developed specifically to work within a web browser. The architecture of SaaS-based applications is specifically designed to support many concurrent users (multitenancy) at once. This is a big difference from the traditional client/server or application service provider (ASP)-based solutions that cater to a contained audience. SaaS providers, on the other hand, leverage enormous economies of scale in the deployment, management, support, and maintenance of their offerings.

2.6.1 SaaS Implementation Issues

Many types of software components and applications frameworks may be employed in the development of SaaS applications. Using new technology found in these modern components and application frameworks can drastically reduce the time to market and cost of converting a traditional on-premises product into a SaaS solution. According to Microsoft,[14] SaaS architectures can be classified into one of four maturity levels whose key

11. "Software as a Service Threatens Partner Revenue and Profit Streams, New Partners Emerging, IDC Research Shows," from http://www.idc.com/getdoc.jsp?containerId=prUS20884007, 20 Sep 2007, retrieved 16 Jan 2009.
12. Ibid.
13. http://en.wikipedia.org/wiki/Software_as_a_service, retrieved 11 Jan 2009.
14. http://www.microsoft.com/serviceproviders/saas/saasplatform.mspx, retrieved 14 Jan 2009.

attributes are ease of configuration, multitenant efficiency, and scalability. Each level is distinguished from the previous one by the addition of one of these three attributes. The levels described by Microsoft are as follows.

- **SaaS Architectural Maturity Level 1—Ad-Hoc/Custom.** The first level of maturity is actually no maturity at all. Each customer has a unique, customized version of the hosted application. The application runs its own instance on the host's servers. Migrating a traditional non-networked or client-server application to this level of SaaS maturity typically requires the least development effort and reduces operating costs by consolidating server hardware and administration.
- **SaaS Architectural Maturity Level 2—Configurability.**The second level of SaaS maturity provides greater program flexibility through configuration metadata. At this level, many customers can use separate instances of the same application. This allows a vendor to meet the varying needs of each customer by using detailed configuration options. It also allows the vendor to ease the maintenance burden by being able to update a common code base.
- **SaaS Architectural Maturity Level 3—Multitenant Efficiency.** The third maturity level adds multitenancy to the second level. This results in a single program instance that has the capability to serve all of the vendor's customers. This approach enables more efficient use of server resources without any apparent difference to the end user, but ultimately this level is limited in its ability to scale massively.
- **SaaS Architectural Maturity Level 4—Scalable.** At the fourth SaaS maturity level, scalability is added by using a multitiered architecture. This architecture is capable of supporting a load-balanced farm of identical application instances running on a variable number of servers, sometimes in the hundreds or even thousands. System capacity can be dynamically increased or decreased to match load demand by adding or removing servers, with no need for further alteration of application software architecture.

2.6.2 Key Characteristics of SaaS

Deploying applications in a service-oriented architecture is a more complex problem than is usually encountered in traditional models of software deployment. As a result, SaaS applications are generally priced based on the number of users that can have access to the service. There are often additional fees for the use of help desk services, extra bandwidth, and storage. SaaS revenue streams to the vendor are usually lower initially than traditional software license fees. However, the trade-off for lower license fees is a monthly recurring revenue stream, which is viewed by most corporate CFOs as a more predictable gauge of how the business is faring quarter to quarter. These monthly recurring charges are viewed much like maintenance fees for licensed software.[15] The key characteristics of SaaS software are the following:

- Network-based management and access to commercially available software from central locations rather than at each customer's site, enabling customers to access applications remotely via the Internet.
- Application delivery from a one-to-many model (single-instance, multitenant architecture), as opposed to a traditional one-to-one model.
- Centralized enhancement and patch updating that obviates any need for downloading and installing by a user. SaaS is often used in conjunction with a larger network of communications and collaboration software, sometimes as a plug-in to a PaaS architecture.

2.6.3 Benefits of the SaaS Model

Application deployment cycles inside companies can take years, consume massive resources, and yield unsatisfactory results. Although the initial decision to relinquish control is a difficult one, it is one that can lead to improved efficiency, lower risk, and a generous return on investment.[16] An increasing number of companies want to use the SaaS model for corporate applications such as customer relationship management and those that fall under the Sarbanes-Oxley Act compliance umbrella (e.g., financial recording and human resources). The SaaS model helps enterprises ensure that all locations are using the correct application version and, therefore, that the

15. Erin Traudt and Amy Konary, "2005 Software as a Service Taxonomy and Research Guide," IDC, http://www.idc.com/getdoc.jsp?containerId=33453&pageType=PRINT-FRIENDLY#33453-S-0001, retrieved 11 Jan 2009.
16. http://searchnetworking.techtarget.com/generic/0,295582,sid7_gci1164670,00.html, retrieved 18 Jan 2009.

format of the data being recorded and conveyed is consistent, compatible, and accurate. By placing the responsibility for an application onto the doorstep of a SaaS provider, enterprises can reduce administration and management burdens they would otherwise have for their own corporate applications. SaaS also helps to increase the availability of applications to global locations. SaaS also ensures that all application transactions are logged for compliance purposes. The benefits of SaaS to the customer are very clear:

- Streamlined administration
- Automated update and patch management services
- Data compatibility across the enterprise (all users have the same version of software)
- Facilitated, enterprise-wide collaboration
- Global accessibility

As we have pointed out previously, server virtualization can be used in SaaS architectures, either in place of or in addition to multitenancy. A major benefit of platform virtualization is that it can increase a system's capacity without any need for additional programming. Conversely, a huge amount of programming may be required in order to construct more efficient, multitenant applications. The effect of combining multitenancy and platform virtualization into a SaaS solution provides greater flexibility and performance to the end user. In this chapter, we have discussed how the computing world has moved from stand-alone, dedicated computing to client/network computing and on into the cloud for remote computing. The advent of web-based services has given rise to a variety of service offerings, sometimes known collectively as XaaS. We covered these service models, focusing on the type of service provided to the customer (i.e., communications, infrastructure, monitoring, outsourced platforms, and software). In the next chapter, we will take a look at what is required from the service provider's perspective to make these services available.

2.7 Chapter Summary

In this chapter we have examined the various types of web services delivered from the cloud. Having the ability to leverage reusable software components across a network has great appeal to implementors. Today, the most common and successful example of cloud computing is SaaS, but other functions, including communication, infrastructure, and platforms, are also core components of cloud computing. Because of the extremely low barriers to entry, offerings have been made available to consumers and small businesses as well as mid-sized and large enterprises. This is a key differentiator from many service-oriented architecture (SAO) offerings, which will be covered next.

In the next chapter we will learn how companies build highly automated private cloud networks that can be managed from a single point. We will discuss how server and storage virtualization is used across distributed computing resources. We will describe in basic terms how cloud infrastructure is built. We will provide an overview of the basic approach to SOA as it applies to data center design. And we will examine the role and use of open source software in data centers. Understanding the use and importance of collaboration technologies in cloud computing architectures is fundamental to understanding how requirements of the cloud have evolved.