



Overview of Amazon Web Services

December 2010

Introduction

Managing the unique and groundbreaking changes in both technology and business over the past decade has created an ongoing IT infrastructure challenge for many senior technology executives. Indeed, over the past 10 years, the typical business application architecture has evolved first from a desktop-centric installation, eventually to client/server solutions, and now most prominently to loosely coupled web services and service-oriented architectures (SOA). Each of these evolutionary steps has built upon the previous while adding new challenges, dimensions, and opportunities for IT organizations and their business partners.

Recently, virtualization has become a prevalent and widely accepted way to reduce operating costs and increase the reliability of enterprise IT. In addition, concepts such as grid computing make possible a whole new class of analytics, data crunching, and business intelligence tasks that were previously cost- and time-prohibitive. Along with these technology changes there have also been fundamental changes in how markets work, with the speed of innovation and product introductions increasing at an unprecedented rate. Taken alongside the wide acceptance of Software as a Service (SaaS) offerings, all of these concepts have inspired the latest turn in the IT infrastructure challenge: *cloud computing*.

Amazon Web Services (AWS) provides a flexible, cost-effective, scalable, and easy to use cloud computing platform for businesses of all sizes.

What is “Cloud Computing”?

Cloud computing has become one of the most discussed IT paradigms of recent years. It builds on many of the advances in the IT industry over the past decade and presents significant opportunities for businesses to shorten time to market and reduce costs by consuming shared computing and storage resources rather than building, operating, and improving infrastructure on their own. The speed of change in markets creates significant pressure on the enterprise IT infrastructure to adapt and deliver. As defined by Gartner¹, “Cloud computing is a style of computing where scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.”

Put simply, in much the same way that national electric grids enable homes and businesses to plug into a centrally managed, efficient, and cost-effective energy source, cloud computing enables businesses to obtain flexible, secure, and cost-effective IT infrastructure from a vendor experienced in running large-scale networks and computing environments. When companies were freed from having to create their own electricity, they were able to focus on the core competencies of their line of business and the needs of their customers. Likewise, cloud computing liberates IT organizations from having to devote precious people and budget to activities which don’t directly contribute to their bottom line, while still enabling them to obtain the necessary infrastructure functionality to continue to run their businesses.

¹ “Key Attributes Distinguish Cloud Computing Services”, March 2009. David W. Cearley and David Mitchell Smith, Gartner.

Specifically, these capabilities include compute power, storage, databases, messaging, and other building block services that are subsequently used to run business applications. When coupled with a utility-style pricing and business model, cloud computing promises to deliver enterprise-grade IT infrastructure in a reliable, timely, and cost-effective manner.

To further understand the impact and promise of cloud computing, one may first analyze the significance of and lessons learned from business *outsourcing*. Focusing on a core competency and then shifting the peripheral business tasks to other organizations is a proven business strategy. Today companies outsource business functions such as logistics, HR, payroll, and facilities. In addition, many companies have taken advantage of IT outsourcing as a way to move some capabilities out of their internal organization altogether.

Superficially, at least, cloud computing resembles the trend of business outsourcing, as both provide the benefit of leveraging the expertise of others and being cost efficient. But cloud computing has additional benefits of *flexibility*, *scalability + elasticity* and *reliability*. These additional benefits are why enterprise organizations see cloud computing as a powerful next step in their IT infrastructure evolution.

Amazon and Cloud Computing

Amazon has a long history of leveraging decentralized IT infrastructure. This gave our development teams access to on-demand resources such as compute and storage and increased overall productivity and agility. By 2005, Amazon had spent over a decade and hundreds of millions of dollars building and managing the large-scale, reliable, and efficient IT infrastructure that powered the operation of one of the world's largest online retail platforms. AWS enables all customers to capitalize on and benefit from Amazon's *experience* and investment in running large-scale distributed, transactional IT infrastructure.

With AWS, companies can requisition compute power, storage, and other services in minutes—gaining access to a suite of elastic IT infrastructure services as their business needs them. Using AWS, companies have the flexibility to choose whichever development platform or programming model makes the most sense for the problems they're trying to solve. Companies only pay for what is used, with no up-front expenses or long-term commitments, making AWS a cost-effective way to deliver applications.

Here are some of the examples of how enterprises use AWS today:

- A large enterprise quickly and economically deploys new internal applications, such as HR solutions, payroll applications, inventory management solutions, and online training to its distributed workforce
- An e-commerce web site accommodated sudden demand for a “hot” product caused by viral buzz from Facebook and Twitter without having to upgrade its infrastructure
- A pharmaceutical research firm executes large-scale simulations using computing power provided by AWS
- Media companies serve unlimited video, music, and other media to its worldwide customer base

The Differences That Distinguish AWS

AWS provides unique characteristics among all vendors in the cloud computing landscape, including:

- **Flexible.** AWS enables organizations to use the programming model, operating system, databases, and architectures they are already familiar with. In addition, this flexibility helps companies mix and match architectures in order to serve their diverse business needs.
- **Cost-effective.** With AWS, organizations pay only for what they use, with no up-front commitments or long-term commitments.
- **Scalable and elastic.** Businesses can quickly add and subtract AWS resources to their applications in order to meet customer demand and manage costs.
- **Secure.** In order to provide end-to-end security and end-to-end privacy, AWS builds services in accordance with security best practices, provides appropriate security features in those services, and documents how to use those features.
- **Experienced.** When using AWS, you leverage over fifteen years of Amazon's experience in delivering large-scale, global infrastructure in a reliable, secure fashion.

Flexible

The first key difference between AWS and traditional IT models is the *flexibility* afforded by computing in the cloud. Past attempts to deliver IT solutions often required large investments in new architectures, programming languages, and operating systems. Investments such as these are valuable, but they can also put a throttle on your business, preventing you from quickly responding to changing market dynamics as your teams take the time to adapt to new technologies. When the opportunity to innovate arises, you want to be able to move quickly and not always have to support legacy infrastructure and applications or deal with protracted procurement processes.

In contrast, the flexibility of AWS allows businesses to choose the programming models, languages, and operating systems they are already using or that are best suited for their project. With AWS, developers and other IT professionals can bring their existing skills and knowledge to the platform; they don't have to learn lots of new skills. Flexibility means that migrating legacy applications to the cloud is easier and much less expensive. Instead of re-writing applications, businesses can easily move them to the AWS cloud and tap into advanced computing capabilities.

Indeed, building applications on AWS is very much like building applications using existing hardware resources. Since AWS provides a flexible, virtual IT infrastructure, the services can be used together as a platform or used separately for specific needs. It can be used to run almost anything—from full web applications to batch processing to offsite data back-ups.

In addition to building new applications for AWS to meet time-sensitive business opportunities, companies can begin to move existing SOA-based solutions to the cloud by migrating discrete components of their legacy applications. Typically these are components that benefit from high availability and scalability, or they are self-contained applications with few internal dependencies. Larger companies typically run in a hybrid mode where pieces of the application run in their data center and other portions run in the cloud. Once these companies gain experience with the cloud, they begin transitioning more of their projects to the cloud and begin to appreciate many of the benefits outlined in this document. Ultimately many companies see the unique advantages of the cloud and AWS and make it a permanent part of their IT "mix".

Finally, with AWS, provisioning new services is easy. Instead of the weeks and months it takes a typical project to wind its way through the planning, budgeting, procurement, set-up, deployment, operations, and hiring processes of an

organization, business units can simply sign up for AWS and immediately begin deployment on the cloud equivalent of one, ten, one hundred, or one thousand servers. Whether it's for prototyping an application or hosting a production solution, AWS makes it simple for businesses to get started and stay productive.

In summary, many customers find the flexibility of AWS to be a great asset in improving time to market and overall organizational productivity.

Cost-Effective

Cost represents one of the most complex elements of delivering contemporary IT solutions. It seems that for every advance that will save money there is often a commensurate investment needed to realize that savings. One example of this is the relatively low cost involved in developing and deploying an e-commerce application, but the increased need for hardware and bandwidth that successful deployment can create.

The cloud provides on-demand IT infrastructure that lets users consume exactly the amount of resources they actually need. IT organizations aren't limited to a set amount of storage, bandwidth, or computing resources, and it's often difficult for organizations to predict their actual requirements for each. As a result, they either provision too few resources, in which case their customer satisfaction suffers, or too many resources, in which case they're missing an opportunity for maximizing ROI through full utilization. Striking the right balance requires the type of flexibility already discussed.

As mentioned previously, AWS provides businesses with the increased agility needed to be able to instantly scale their infrastructure up or down based on their unique demands. This business agility can often be a point of cost savings itself. When a business is able to respond quickly to changes, no matter how large or small, it can take on new opportunities and meet business challenges that could drive revenue and reduce costs. With AWS, business can almost instantly provision new resources.

AWS requires no upfront investment, long-term commitment, or minimum spend. IT organizations can get started through a completely self-service experience online, scale up and down as needed, and terminate their relationship with AWS at any time. Unless organizations wish to consult with Amazon for deeper technical support with our systems architects, no salespeople are needed to intercede in the process.

In addition, using a cloud computing platform to host and operate web applications enables businesses to forego many of the often considerable costs of owning and operating their own infrastructure. Paying for items such as power, cooling, real estate, IT administration staff, and so forth are not necessary when choosing the cloud over traditional hosted infrastructure.

Scalable and elastic

In the traditional IT organization, scalability and elasticity were often equated with investment and infrastructure. In the cloud these concepts provide opportunity for savings and ROI. AWS employs the term "elasticity" to describe the power to scale computing resources up and down easily and with minimal friction. Imagine what would happen to an IT shop if traffic to an application doubled or tripled in a short period of time. For example, during benefits open enrollment periods, many corporate users generate significant traffic to internal applications. Businesses need the confidence that their existing infrastructure can handle such a spike in traffic, and that such a spike doesn't interfere with normal

business operations elsewhere in the organization. With AWS, your organization can use, for example, the built-in Elastic Load Balancing and Auto-Scaling features to automatically scale your AWS cloud-based resources up to meet unexpected demand, and scale those resources down when they are no longer needed.

At the same time, the cloud is also useful as a resource for executing mission-critical, short-term jobs. For example, a pharmaceutical company needs to run drug simulations. By using AWS, they can spin up resources in the cloud, and then shut them down when they are no longer needed. Another example would be an enterprise dealing with a natural disaster impacting their data center and the need to tap into new storage and computing resources to accommodate demand. Even tasks such as month-end payroll or invoice processing can be executed in the cloud to preserve computing resources and reduce costs. This elasticity is a key attribute to cloud computing and helps companies avoid having to provision resources up front for projects with variable consumption or short lifetimes.

Further, with AWS, allocating more (or less) resources to your applications involves simple API calls as opposed to the traditional method of acquiring hardware, setting it up, maintaining it, and so on.

Cloud computing allows businesses to tap into massive compute capacity and other resources as needed. The burden of managing the infrastructure falls on the cloud provider, such as AWS, not on the business. Meanwhile, organizations can stay focused on their customers and on solving their business challenges.

Secure

AWS delivers a highly scalable cloud computing platform with high availability and reliability as well as the flexibility to enable customers to build a wide range of applications. In order to provide end-to-end security and end-to-end privacy, AWS builds services in accordance with security best practices, provides appropriate security features in those services, and documents how to use those features. In addition, AWS customers must use those features and best practices to architect an appropriately secure application environment. Enabling customers to ensure the confidentiality, integrity, and availability of their data is of the utmost importance to AWS, as is maintaining trust and confidence.

At a high level, AWS has taken the following approach to reliably secure the cloud infrastructure:

- **Certifications and Accreditations.** AWS has successfully completed a SAS70 Type II Audit, and will continue to obtain the appropriate security certifications and accreditations to demonstrate the security of our infrastructure and services.
- **Physical Security.** Amazon has many years of experience in designing, constructing, and operating large-scale data centers. AWS infrastructure is housed in Amazon-controlled data centers throughout the world. Only those within Amazon who have a legitimate business need to have such information know the actual location of these data centers, and the data centers themselves are secured with a variety of physical barriers to prevent unauthorized access.
- **Secure Services.** Each of the services within the AWS cloud is architected to be secure and contains a number of capabilities that restrict unauthorized access or usage without sacrificing the flexibility that customers demand.
- **Data Privacy.** AWS enables users to encrypt their personal or business data within the AWS cloud and publishes backup and redundancy procedures for services so that customers can gain greater understanding of how their data flows throughout AWS.

For more information on security policies and procedures for AWS, consult the AWS Security Center at aws.amazon.com/security.

Experienced

AWS is designed to provide a low-friction path to cloud computing. But as with any IT project, the move to the AWS cloud should be done thoughtfully. Just as an organization would work closely with any hardware or software vendor to ensure they have the capabilities necessary to support a business now and in the future, organizations should hold their cloud computing partner to the same high standards. The trust that businesses place in their cloud computing vendor will be critical as their business grows and their customers continue to expect the best experience.

The AWS cloud provides levels of scale, security, reliability, and privacy that are often cost-prohibitive for many organizations to meet or exceed. AWS has built an infrastructure based on lessons learned over sixteen years of managing the multi-billion dollar Amazon.com business, and all AWS customers benefit as Amazon continues to hone its infrastructure management skills and capabilities. Today Amazon.com runs a global web platform serving millions of customers and managing billions of dollars worth of commerce every year. AWS has been operating since 2006 and today serves hundreds of thousands of customers worldwide.

Moreover, AWS has a demonstrated track record of listening to its customers and delivering highly innovative new features at a rapid pace. These frequent releases are offered with the same high standards of security and reliability demonstrated in all of the existing AWS infrastructure services. In addition to new services, Amazon constantly hones its operational expertise to ensure the ongoing dependability of AWS. On an ongoing basis, AWS incorporates both industry best practices and proprietary advances into its cloud. Choosing AWS as a cloud computing provider allows businesses to take advantage of all these investments and the ongoing investments Amazon makes into building proven infrastructure.

Overview of Amazon Web Services

AWS is a comprehensive cloud services platform, offering compute power, storage, content delivery, and other functionality that enables businesses to cost-effectively deploy applications and services with greater flexibility, scalability, and reliability. The power of self-service through AWS means you can proactively address your internal plans and react to external demands when you choose and not have to wait for a salesperson to return your call.

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon EC2 is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides various features like Amazon Elastic Load Balancing, Auto scaling, Amazon CloudWatch for monitoring to developers the tools to build failure resilient elastic applications and isolate themselves from common failure scenarios.

Amazon Simple Storage Service (Amazon S3)

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

Amazon Virtual Private Cloud (Amazon VPC)

Amazon VPC is a secure and seamless bridge between a company's existing IT infrastructure and the AWS cloud. Amazon VPC enables enterprises to connect their existing infrastructure to a set of isolated AWS compute resources via a Virtual Private Network (VPN) connection, and to extend their existing management capabilities such as security services, firewalls, and intrusion detection systems to include their AWS resources. Amazon VPC integrates today with Amazon EC2, and will integrate with other AWS services in the future.

Amazon CloudFront

Amazon CloudFront is a web service for content delivery. It integrates with other Amazon Web Services to give developers and businesses an easy way to distribute content to end users with low latency, high data transfer speeds, and no commitments. Amazon CloudFront delivers your content using a global network of edge locations. Requests for your objects are automatically routed to the nearest edge location, so content is delivered with the best possible performance. Amazon CloudFront works seamlessly with Amazon Simple Storage Service (Amazon S3) which durably stores the original, definitive versions of your files.

Amazon Route 53

Amazon Route 53 is a highly available and scalable DNS service designed to give developers and businesses an extremely reliable and cost effective way to route end users to Internet applications by translating human readable names like `www.example.com` into the numeric IP addresses like `192.0.2.1` that computers use to connect to each other. Route 53 is designed to be fast, easy to use, and cost-effective. It answers DNS queries with low latency by using a global network of DNS servers. Queries for your domain are automatically routed to the nearest DNS server, and thus answered with the best possible performance. You pay only for managing domains through the service and the number of queries that the service answers.

Amazon Relational Database Service (Amazon RDS)

Amazon RDS is a web service that makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while managing time-consuming database administration tasks, freeing you up to focus on your applications and business. Amazon RDS gives you access to the full capabilities of a familiar MySQL database. This means the code, applications, and tools you already use today with your existing MySQL databases work seamlessly with Amazon RDS. Amazon RDS automatically patches the database software and backs up your database, storing the backups for a user-defined retention period. You also benefit from the flexibility of being able to scale the compute resources or storage capacity associated with your relational database instance via a single API call.

Amazon SimpleDB

Amazon SimpleDB is a web service providing the core database functions of data indexing and querying in the cloud. By offloading the time and effort associated with building and operating a web-scale database, SimpleDB provides developers the freedom to focus on application development.

A traditional, clustered relational database requires a sizable upfront capital outlay, is complex to design, and often requires extensive and repetitive database administration. Amazon SimpleDB is dramatically simpler, requiring no schema, automatically indexing your data and providing a simple API for storage and access. This approach eliminates the administrative burden of data modeling, index maintenance, and performance tuning. Developers gain access to this functionality within Amazon's proven computing environment, are able to scale instantly, and pay only for what they use.

This service works in close conjunction with Amazon S3 and Amazon EC2, collectively providing the ability to store, process and query data sets in the cloud. Amazon SimpleDB is easy to use and provides the core functionality of a database—real-time lookup and simple querying of structured data—without the operational complexity. Tools available through Amazon Simple DB can help migrate your dataset from existing database solutions.

Amazon Simple Queue Service (Amazon SQS)

Amazon SQS is a reliable, highly scalable, hosted queue for storing messages as they travel between computers. By using Amazon SQS, developers can simply move data between distributed components of their applications that perform different tasks, without losing messages or requiring each component to be always available.

Amazon Simple Notification Service (Amazon SNS)

Amazon SNS is a web service that makes it easy to set up, operate, and send notifications from the cloud. It provides developers with a highly scalable, flexible, and cost-effective capability to publish messages from an application and immediately deliver them to subscribers or other applications. Amazon SNS provides a simple web services interface that can be used to create topics you want to notify applications (or people) about, subscribe clients to these topics, publish messages, and have these messages delivered over clients' protocol of choice (i.e. HTTP, email, etc.). Amazon SNS delivers notifications to clients using a "push" mechanism that eliminates the need to periodically check or "poll" for new information and updates.

Amazon Elastic MapReduce

Amazon Elastic MapReduce is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data. It utilizes a hosted Hadoop framework running on the web-scale infrastructure of Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Simple Storage Service (Amazon S3).

A Plan for Next Steps

As with any IT investment, the first thing to do is make sure your business model is aligned with your IT plan. Knowing when and where to take advantage of cloud resources requires knowing the areas that are your core business competencies and those that are going to be best served through outside infrastructure.

Next, you have to think about some key technology questions. This list will vary depending upon your project and business, but usually includes the following:

- Do you have legacy applications that need greater scalability, reliability, or security than you can afford to maintain in your own environment?
- What are your hardware and bandwidth capacity requirements?
- How will you be prepared to scale up (and down) following deployment?
- How can the cloud advance your IT and business objectives?

As you answer each of these, look at them through the lenses of *flexibility*, *cost effectiveness*, *scalability + elasticity*, and *security*. Being able to take advantage of Amazon Web Services will allow you to focus on your core competencies and leverage the resources and *experience* Amazon provides.