Daniel Firestone, MIT (6.897 Spring 2011)

1 Computation

Amazon Web Services

Amazon Web Services is Amazon's suite of public cloud computing services, a full list of which can be found here. The Elastic Compute Cloud (EC2) is the basic computation service of AWS. The idea is simple - you create a VM image with an OS you'd like to run, or choose a preconfigured image to start with. You can then deploy this image to Amazon's servers, selecting from several types of instances that vary in number of CPUs and quantity of RAM, and select the number of instances you'd like to start. You configure the software you want to run in the VM image, and can purchase static IP addresses through Amazon's Elastic IP Addresses program, which you can dynamically map to your instances. Your instances can run in different Availability Zones (different physical locations) to protect against global failure - an individual machine has no uptime guarantee, however, and software should anticipate individual node failure.

Elastic Load Balancing can be enabled to balance incoming traffic between different machines, and to allow you to elastically start and stop instances in response to current demand. You pay for the number of hours of computation per instance and for bandwidth used. This pricing model is ideal for small companies that don't want to start with a large investment in servers - their fixed infrastructure costs are low and the overall cost grows with demand. A fair amount of technical skill is required to set up VM images and configure machines to run custom software. EC2 is an example of an Infrastructure as a Service (IaaS) - Amazon provides you with a datacenter in which to run VMs, but you choose and install your platform, from the OS (Amazon supports Linux, and Windows for an extra licensing fee) to the runtime.

For distributed compuation, Amazon offers Elastic MapReduce, which runs Hadoop on EC2 and handles imaging and configuration automatically.

Windows Azure

Windows Azure is Microsoft's cloud computing platform. The basic setup of the infrastructure is similar to EC2, and you can upload a custom VM (only Windows is officially supported), but the typical configurations are closer to a Platform as a Service (PaaS). Used in this fashion, Azure allows you to develop applications in .NET with a special API to control Azure services, which would integrate with services like IIS and ASP.NET for web hosting or Dryad for MapReduce-style programming (still in beta). You can then deploy this application either as a public facing Web Role or a background Worker Role.

Just like on EC2, you select how many instances you want to run, and of what type. However, once you deploy the software, an image of a custom tailored Windows Server instance automatically starts with your software running. Your instances are divided into multiple update domains, and when critical security updates (or other large system updates are available) individual domains will be taken down, updated, and restarted, guaranteeing that you have the latest software (both for Windows and for services like IIS and the .NET runtime) without your cluster going down. This can significantly reduce the work required for most cases as compared to a traditional EC2 instance

- you simply upload your application(s) and the system handles administration transparently. At the same time, Azure retains the flexibility of EC2 in allowing you to set up custom instances if you need to run legacy software or don't want to work with .NET - a hybrid of IaaS and PaaS.

Google App Engine

Google App Engine is a purely PaaS solution. It is designed to run Python applications, and a limited subset of Java is also supported - no other languages or applications may be used. There is no notion of individual instances - elasticity and scaling are completely transparent to the user, offering a very simple solution where a user can write an app and deploy it with no further configuration. Pricing is by hour of computation and bandwidth. Google provides a custom web framework in Python, but can also support subsets of other frameworks like Django. Google does not currently support an official MapReduce implementation, but subsets of the MapReduce API are supported through several community projects. Because of a lack of configurability, the platform cannot be used with legacy applications, nor can it integrate with services like a relational database. As a result, most users will end up writing applications from scratch, which are not likely to integrate well with existing infrastructure.

2   Storage

Amazon Web Services

Instances in EC2 come with a free virtual hard drive for temporary or local data. No guarantee is made about this data - it will still be there when an instance is rebooted, but if the instance or its disk fails, the data is lost. AWS provides Extended Block Storage (EBS) as a reliable, off-instance filesystem that can be attached to an individual EC2 instance for more reliability for an extra cost. Amazon's Simple Storage Service (S3) allows a customer to create a filesystem accessible to all of its instances across its network, and is one of AWS's most popular services. Amazon SimpleDB is a highly scalable distributed nonrelational database written in Erlang that can be accessed by any EC2 instance. A scalable version of MySQL called Amazon Relational Database Service (RDS) is also provided. Amazon's Simple Queue Service (SQS) provides queues for messages to coordinate distributed applications, generally between EC2 instances running as a distributed cluster.

Windows Azure

Microsoft provides a similar set of services to AWS. Instances come with virtual hard drives for free, and reliable storage can be accessed from any instance with Azure's BLOB Service. Windows Azure Drive allows you to mount a filesystem to a single instance. Azure's Table Service provides a BigTable style nonrelational database, though common NoSQL databases such as MongoDB are compatible with the platform. A queueing service is also provided for managing distributed systems. Unlike Amazon or Google, which mostly promote their nonrelational database options, Microsoft has heavily promoted SQL Azure, a full featured, scalable version of its standard SQL server. It has promoted this service as a replacement for an internal deployment of SQL server, suggesting that companies use it as a maintenance and administration free SQL service and connect it to their internal applications. APIs for all these services are provided in the Azure development kit, so users developing Azure applications can easily access any of these services without the need for manual configuration.

Google App Engine

Google provides the most limited storage options of any of these services. Since there is no notion of an individual instance, there is no local storage that an app can use. There is a shared datastore for internal files and a blob store for serving files, both of which are billed for by the gigabyte and by transaction. There is no relational database option - instead customers can use BigTable, a nonrelational database that can be queried with Google's GQL language.

3   Other Features

All three services provide Content Distribution Network-like features, though Amazon's is not as robust outside of America and Europe, and Google's is more limited feature-wise. Microsoft has made this a prominent feature, arguing that their caching service (similar to memcached) and CDN, which are the same ones used worldwide for Bing, obviate the need for external services like Akamai. Microsoft says that this allows users to get all their hosting needs in a single package that's easy to configure (since the caching service and CDN integrate with the other Azure services) and less expensive than an external CDN solution. Both Amazon and Microsoft provide cloud identity and notification services, and Amazon has several services designed to facilitate e-commerce and micropayments.

Both Amazon and Microsoft have also added features for businesses to allow for hybrid cloud applications - integrating internal datacenters with the public cloud. Amazon's Virtual Private Cloud provides a service to let you link your EC2 instances to your internal network via a VPN, allowing businesses to potentially add extra VMs for an internal service when a flash load is encountered, or just make a certain portion of their infrastructure operate at variable cost. Microsoft has gone a good deal further, seeing a hybrid cloud model as a way to compete with private cloud services like VMWare. In addition to a service to connect your Azure instances to your private network, they also provide Windows Azure AppFabric - a middleware solution for creating large distributed hybrid or public cloud applications that can also run on regular internal servers. It includes identity and access control services, a service bus for flexible communication, messaging and control between nodes of a distributed system, a caching service, and APIs for developing composite applications.

Many businesses have felt that while the cost savings of a public cloud in terms of reducing IT staff were positive, cloud bandwidth is more expensive than what they pay for in their current datacenters. In addition, companies in the health care, credit card and other industries are often legally obligated to host their data on premises. To target these markets, Microsoft created the Windows Azure Platform Appliance - essentially a datacenter in a (very large) box or crate, soon to be sold by vendors like HP or Dell, that plugs in to your datacenter but runs Windows Azure and requires no administration. This allows companies to pay for their own bandwidth and power, and keep their data on premises, while not needing a full maintenence or IT staff. The fact that the appliances are on premises is transparent to Azure applications, which can be configured to scale dynamically between appliances and the public cloud when necessary.