**Amazon Web Services**

***AWS Module 1.2*** (<https://awseducate.instructure.com/courses/80/pages/module-1-dot-2-aws-overview?module_item_id=4001>)

Cloud computing is the on-demand delivery of compute power, database storage, applications, and other IT resources through a cloud services platform via the internet with pay-as-you-go pricing.

Cloud Computing Basics

Whether you are running applications that share photos to millions of mobile users or you’re supporting the critical operations of your business, a cloud services platform provides rapid access to flexible and low cost IT resources. With cloud computing, you don’t need to make large upfront investments in hardware and spend a lot of time on the heavy lifting of managing that hardware. Instead, you can provision exactly the right type and size of computing resources you need to power your newest bright idea or operate your IT department. You can access as many resources as you need, almost instantly, and only pay for what you use.

How Does Cloud Computing Work?

Cloud computing provides a simple way to access servers, storage, databases and a broad set of application services over the Internet. A Cloud services platform such as Amazon Web Services owns and maintains the network-connected hardware required for these application services, while you provision and use what you need via a web application.

What is AWS? (<https://www.youtube.com/watch?v=mZ5H8sn_2ZI&feature=youtu.be>)

No heavy up-front costs

Build sophisticated, scalable applications over the internet

Services: **Storage** (Amazon Glacier, Amazon S3, Amazon EBS), **Compute** (EC2), **Database** (Amazon Redshift, Amazon DynamoDB, Amazon ElastiCache, Amazon RDS), **Resource Management Tools** (Amazon Cloud Watch, AWS IAM, AWS Cloud Formation, AWS Elastic Beanstalk), **Networking**, **Application Services**, **Deployment & Management**

What is Cloud Computing? (<https://www.youtube.com/watch?v=jOhbTAU4OPI&feature=youtu.be>)

All services available in a traditional data center are available online via AWS

Pay as you go

Services: **Storage**, **Compute**, **Database**, **Resource Management Tools**, **Networking**, **Application Services**, **Deployment & Management**

100,000s of customers, resulting in economies of scale

Businesses bypass IT startup, saving time and money

Easier to experiment and scale up or down

Expand business to global end-users without heavy-lifting or physical expansion due to availability of

AWS in 10 Minutes (<https://www.youtube.com/watch?v=r4YIdn2eTm4&feature=youtu.be>)

What is it?

Global cloud platform, allows businesses to host and manage services on the internet

Hosting provider where you can manage applications on the cloud

Used by 80% of Fortune 500 companies to host infrastructure

Services: Infrastructure services (provide bare servers as a service), Platform (Java, Ruby, PHP), Software (e-mail, queuing, etc.), Cloud Storage (EBS, S3, etc.)

Why so popular?

Billing structure: Pay-as-you-go, transparent billing, easy to sign up

Up and running in minutes without heavy investment of time, money, or other resources to get started

Stable: No regular outages. Any outages in the last 7-8 years (as of 2017) have been few, region specific and only lasted 2-3 hours.

Service Overview

EC2 (Elastic Compute Cloud) – Most commonly used service; gives bare servers on which to run software. Large or small machines are provided based on business needs.

VPC (Virtual Private Cloud) – Chunks of Amazon cloud is rented out for businesses to create networks in the cloud and run servers in those networks

S3 (Simple Storage Service) – Gives users the opportunity to upload and share files

RDS (Relational Database Service) – Allows you to run and manage database on the cloud with things like SQL, Oracle, MySQL, and Aurora.

Route 53 – DNS (Domain Name System) service, where you can point your DNS to Amazon.

* DNS Explained (<https://www.youtube.com/watch?v=72snZctFFtA>)
  + One of the most important parts of the internet
  + The computers that make up the internet communicate via underground wire networks, which are identified by IP addresses.
    - DNS is used to translate an actual name into IP addresses.
  + The Root is the invisible ‘.’ at the end of the web address, which represents the root of the internet’s name space.
  + The browser and OS determine if they know what the IP address is already, which could be configured on your computer or in the memory (cache).
  + Browser and OS ask each other what it is, and if they don’t know, the OS will ask the Resolving Name Server (the workhorse of the DNS lookup), which is configured manually or automatically within the OS.
  + The OS queries the RNS for the root. If it does not have it, it will then query the Top Level Domain servers (TLD).
  + The RNS then takes all of the info from the Root Name Servers, puts it in the cache, then goes to the TLD name servers for another query.
  + If the TLD doesn’t know where to find the IP address, then it points to the direction of the Authoritative Name Servers. The TLD knows which ANS to use by accessing the Domain’s Registrar.
    - When a domain name is purchased, the Registrar is told which ANS to use, which then notifies the Registry (organization responsible for the top level domain), and tells them to update the TLD name server.
    - The Resolving Name Server takes the response from the TLD name server, stores it in cache, then queries the website name server. The ANS will recognize it, give an IP address, then gives it to the Resolving Name Server who will then put it in cache and take it to the computer’s OS.
    - The OS gives this info to the browser, which then makes a connection to the IP address requesting the webpage for the web address.

Elastic Load Balancer – Service that gives the opportunity to load balance incoming traffic to multiple machines

Auto-scaling – Adds capacity on the fly to ELB, so that website or apps are never down due to load

How Much Does it Cost?

* Per hour or per GB month
* Region specific pricing
* Term specific pricing – the longer, the cheaper
* Spot resources – Price based on bids

How Big is It?

15 Regions Globally (No matter where you are in the world, there is a region 1,000 miles of your location)

Multiple availability zones per region

Availability Zone = Massive Data Center

One Massive Data Center has 300K-500K Servers

64 Services Available across infrastructure, software, platform, etc.

Focused on machine learning, SAAS (Software as a Service)

AWS Global Infrastructure (<https://aws.amazon.com/about-aws/global-infrastructure/>)

Regions: 23 (19 + 4 Coming Soon)

Availability Zones: 60

Edge Locations: 109

Getting Started with Amazon Web Services (<https://www.youtube.com/watch?v=VgzzHCukwpc&feature=youtu.be>)

Types of Web Applications Run on AWS

Static Websites: Static Content, Blogs, Basic Product Page

Website Hosting: Medium-sized Website, CMS Systems

Dynamic Web Application: Highly interactive websites, Web Applications, Online Services, Social and Sharing Sites

Scalable

Competitive Pricing Model

Low-Cost

Pay-as-you-go

No long-term contracts

No traditional infrastructure with physical servers

Freedom of choice

Secured storage, premises, user experience, etc.

Intro to Amazon EC2 (<https://www.youtube.com/watch?v=TsRBftzZsQo&feature=youtu.be>)

Access to multiple virtual servers (computer instances) in the Amazon cloud

Range of instance types optimized for CPU, Memory and Storage.

On-Demand Pricing – Pay only for what you use, no commitments or fees

Reserved Instance Pricing – Obtain a significant discount on the on-demand prices by paying up front

Spot Instance Pricing – Name price based on market rates

Instances are located in a VPC (Virtual Private Cloud), which is an isolated network that the user controls. User (or business) can control who has access to it.

Auto-scaling allow custom metrics to determine how much to scale up or down depending on load.

Introduction to Amazon EC2 (<https://www.youtube.com/watch?v=KpVNEzpvaY0&feature=youtu.be>)

EC2 is a service that makes it easier to access elastic server capacity.

Launched in 2006

* Elastic Server Capacity
* Instance Choice
* CPU
* Memory
* Storage
  + Locally attached or “instance storage”
  + Amazon EBS General Purpose (SSD) volumes
  + Amazon EBS Provisioned IOPS (SSD) volumes
  + Amazon s3/Amazon Glacier
* Deployment Options
* OS: Amazon Machine Images (AMIs)
* Applications