**EC2 Basics**

Amazon **Elastic Compute Cloud** (Amazon EC2):

- Provides **scalable computing capacity** in the Amazon Web Services (AWS) cloud.

- Eliminates your need to invest in hardware up front, so you can **develop and deploy applications faster**.

- Enables you to **scale up or down** to handle changes in requirements or spikes in popularity, reducing your need to forecast traffic.

Basic Computer   EC2

Components     v Components

---------------+-----------

O/S            : AMIs

CPU            : Instance Type

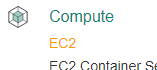
Hard Drive     : EBS

Network Card   : IP Addressing

Firewall       : Security Groups

RAM            : RAM

*Image: Services > Compute > EC2*

[](https://1.bp.blogspot.com/-Sj6ROSgAbUE/WYNCAGfeJpI/AAAAAAAAHSA/W7Ein1fI83Qqjcr1K_0fQWSHYG1FPOzjACLcBGAs/s1600/01%2BCompute%2BEC2.png)

Link: [AWS: Instance Purchasing Options](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html)

Link: [Amazon EC2 Pricing](https://aws.amazon.com/ec2/pricing/)

*Note:****Free Tier****use is available for EC2.*

***EC2 Instance Purchasing Options****(most common)****:***

- On-Demand

- Reserved

- Spot

**On-Demand**(most expense):

* + - Choose any instance type .. Pay fixed rate by the hour with no commitment

- Only charged when the instance is running (billed per hour).

* For users that want flexibility of EC2 w/out up-front payments or long-term commitment
* Best for burst need servers & unpredictable workloads that cannot be interrupted
* Test/Dev for apps running on EC2 for the 1st time.
* Good for apps where compute needs scaling up/down - i.e. usage might increase 10x during certain hours of the day, or certain times of year.
  + Supplement reserved instance servers (for extra temporary server load)

**Reserved** (significant price discount):

* + Provide capacity reservation and offer significant discount on the hourly charge for an instance (1-3 year terms)

- Pay upfront, partial upfront, no upfront

- Charged regardless of how often you use it

* + Applications have steady state, or predictable usage
  + Apps that require reserved capacity
  + Users able to make upfront payments to reduce their total computing costs even further.
  + the longer your contract, the more you save.

Apps that need reserved capacity, steady state or predictable usage

* Domain Controllers
* 1st web server

**Spot** (substantial price discount):

- “Bid” on an instance type, and only pay for and use when the spot price is equal or below your “bid” price.

* + Bid whatever price you want for instance capacity by the hour

- Allows Amazon to sell the use of **unused instances** for short amounts of time

- Prices fluctuate based on supply and demand (billed per hour)

- Instances automatically terminate when spot price > “bid” price..

* + When your bid price is greater than or equal to the spot price, your instance will boot
* When your bid >= spot price, you get a server
  + When the spot price is greater than your bid price, you lose server with 1 hour warning
  + Applications have flexible start and end times
  + Apps that are only feasible at very low compute prices
  + Users with urgent computing needs for large amounts of additional capacity
* Not charged for partial hour if your instance is terminated by AWS.. charged for the FULL hour if YOU terminate your instance.
  + Good for massively parallel computations, or high-compute batch jobs, due to the fact that you can get spot instances for often 50-90% less than on-demand instances, you can massively increase your compute capacity by 2-10x for the same budget.

**How are you charged for using EC2?**

**1) Purchasing Option**

**2) Instance Type** (processing power):

Instance sizing:

* T2 - Lowest Cost General Purpose - Web/Small DBs
* M4 - General Purpose - App Servers
* M3 - General Purpose - App servers
* C4 - Compute Optimized - CPU Intensive Apps/DBs
* C3 - Compute Optimized - CPU Intensive Apps/DBs
* R3 - Memory Optimized - Memory Intensive Apps/DBs
* G2 - Graphics / General Purpose - Video Encoding/Machine Learning/3D App Streaming
* I2 - High Speed Storage - NoSQL DBs, Data Warehousing
* D2 - Dense Storage - Fileservers/Data Warehousing/Hadoop
* D - Density
* I - IOPS
* R - RAM
* T - Cheap General Purpose
* M - Main General Purpose
* C – Compute
* G – Graphics

|  |  |
| --- | --- |
| M | General Purpose i.e. app server (default) |
| T | Micro instances - Low cost, general purpose, web servers |
| C, D, I | CPU/IOPS optimised, good for memory intensive compute |
| R, X | Memory optimised |
| G, P | GPU |
| F | FPGA - Field Programmable Gate Array, for hardware acceleration of code |

Useful mnemoic: Dr Mc Gift Px

- General purpose

- Compute optimized

- GPU optimized

- Memory optimized

- Storage optimized

- EBS optimized (option for higher IOPS performance)

**3) AMI Type** (varies on O/S):

- Linux

- Windows

**4) Data Transfer**(in/out of the instance)

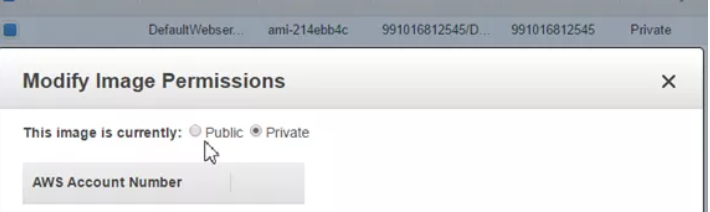
**5) Region**

**5.2) Amazon Machine Images (AMIs)**

**Amazon Machine Image**(AMI) provides the information (template) required to launch an instance (virtual server.)

* AMI = template VM
* AMI's are simply snapshots of a root volume and is stored in S3
* AMI's are regional. You can only launch an AMI from the region in which it was stored.
* You can copy AMI's to other regions using the console, CLI or Amazon EC2 API.

**Contains:**

* Template for root volume for the instance (OS, application servers, apps, etc)
* Launch permissions that control with AWS accounts can use the AMI to launch instances
* Block device mapping that specifies which volumes to attach when launching instance
* When you create an AMI, by default its marked private. You have to manually change the permissions to make the image public or share images with individual accounts
* Hardware Virtual Machines (HVM) AMI's Available
* Paravirtual (PV) AMI's Available
* You can select an AMI based on:
  + Region
  + OS
  + Architecture (32 vs. 64 bit)
  + Launch Permissions
  + Storage for the root device (Instance Store Vs. EBS)

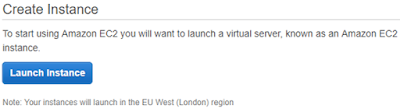
AMIs come in 3 main categories:

**1) Community AMIs:** Free to use, essentially just an OS

**2) AWS Marketplace AMIs:** Pay to use, generally package with additional licensed software

**3) My AMIs**

*Image: Amazon EC2 > Launch Instance*

[](https://4.bp.blogspot.com/-kR7jgtNx_xU/WYNCHObkOkI/AAAAAAAAHSE/zJvW1-F1au894_afoT8SE43X8AQrMOIygCLcBGAs/s1600/02%2BCreate%2BInstance.png)

*Image: AMI categories*

[](https://3.bp.blogspot.com/-JsGUraI0HWY/WYNCLDg9LcI/AAAAAAAAHSI/D4syXj4Y260ZpzqCx3uebpM9MIN5mCcxACLcBGAs/s1600/03%2BAMI%2Bcategories.png)

**5.3) Instance Types**

*When you launch and instance...*

The **instance type** determines the (virtual) hardware used for your instance.

*Instance Type Components:*

**(1) Family***(General Purpose, Compute Optimized, Memory Optimized, Storage Optimized)*

**(2) Type**

**(3) vCPUs**

**(4) Memory (GiB)**

**(5) Instance Storage (GB)**

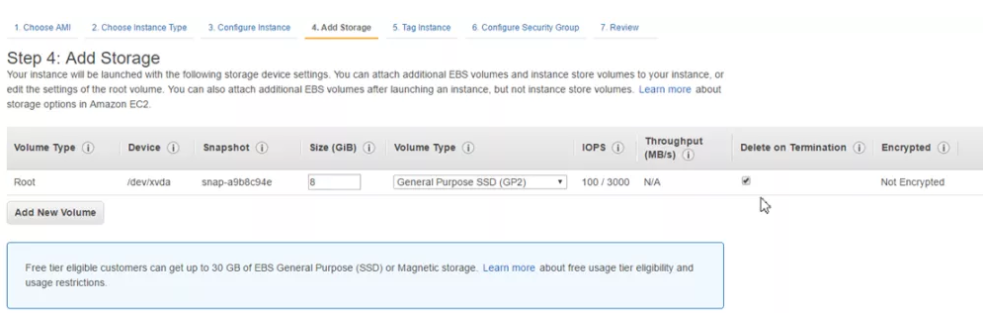
**(6) EBS-Optimized Available**

**(7) Network Performance**

**5.4) Elastic Block Store (EBS)**

\*EBS (Elastic Block Storage) – **Storage volumes that are attached to EC2 instances (think VMDKs).**

* When creating an AMI, on Step 4(Add storage) "Delete on Termination" is checked and not encrypted by default (i.e. Termination protection is turned off by default):
* On an EBS-backed instance, the default action is for the root EBS vol to be deleted when the instance is terminated.



* On an EBS-backed instance, the default action is for the root EBS vol to be deleted when the instance is terminated.
* Root volumes cannot be encrypted by default, you'll need a 3rd party tool (bit locker, etc) to encrypt root vols.

Storage Types:

* Instance Store (Ephemeral):
  + Instances using instance store storage can not be stopped. If they are, data loss would result

(When using an instance store volume, you can not stop the instance (the option to do so will not be available, as the instance moves to another host and and would cause complete data loss)

* + If there is an issue with the underlying host and your instance needs to be moved, or is lost, Data is also lost.
* You can reboot without losing data, if you stop the instance, the data will be wiped.
* Placed in specific AZs & automatically replicated
  + Instance store volumes cannot be detached and reattached to other instances; They exist only for the life of that instance

(Once an Instance has been launched with instance store storage, you can not attach additional instance store volumes after the instance is launched, only EBS volumes.)

* + Best used for scratch storage, storage that can be lost at any time with no bad ramifications, such as a **cache store**
* **The root device for an instance launched from the AMI is an instance store volume created from a template stored in S3**
* EBS (Elastic Block Storage):
* **The root device for an instance launched from the AMI is an EBS volume created from an EBS snapshot**
* Can be stopped, you will not lose data if the underlying host fails.
* Can be detached and reattached to other EC2 instances
  + Elastic Block Storage is persistent storage that can be used to procure storage to EC2 instances.
  + You can NOT mount 1 EBS volume to multiple EC2 instances instead you **must use EFS**
  + Default action for EBS volumes is for the root EBS volume to be deleted when the instance is terminated
* **Termination protection is turned off by default, you must turn it on**
  + By default, ROOT volumes will be deleted on termination, however with EBS volumes only, you can tell AWS to keep the root device volume
  + EBS backed instances can be stopped, you will NOT lose any data.
  + Encryption:
    - Root Volumes cannot be encrypted by default, you need a 3rd party utility
* Other volumes added to an instance can be encrypted and attached to an EC2 instance.
* AWS does not recommend ever putting RAID 5's on EBS

EBS volumes can be detached and reattached to other EC2 instances 3 Types of available EBS volumes can be provisioned and attached to an EC2 instance: EBS 3 Volume Types:

* + - General Purpose SSD (GP2):
      * General Purpose up to 10K IOPS
      * 99.999% availability
      * Ratio of 3 IOPS per GB with up to 10K IOPS and ability to burst
      * Up to 3K IOPS for short periods for volumes under 1GB
    - Provisioned IOPS SSD (I01)
      * Designed for I/O intensive applications such as large relational or No-SQL DBs.
      * Use if need more than 10K IOPS
    - Magnetic (Standard)
      * Ideal for workloads where data is accessed infrequently and apps where the lowest cost storage is important.
* Cheapest/GB.
* Good for infrequently accessed data (fileservers)
* You can poll an instances meta-data by using curl http://169.254.169.254/latest/meta-data/
* You can get an instance's IP address by using curl http://169.254.169.254/latest/meta-data/public-ipv4
* No such thing as user-data, remember its always meta-data not user-data
* When configuring a launch configuration for an auto-scaling group, the Health Check Grace Period is the period of time to ignore health checks while instances or auto-scaled instances are added and booting.

Amazon Elastic Block Store (EBS) provides **block level storage volumes** for use with EC2 instances.

- EBS volumes are **highly available and reliable storage volumes** that can be **attached to any running instance** that is in the **same Availability Zone**.

- EBS volumes that are attached to an EC2 instance are exposed as **storage volumes that persist independently from the life of the instance**.

How to “grow” an EBS volume:

* Detach the original Amazon EBS volume.
* Create a snapshot of the original Amazon EBS volume’s data in Amazon S3.
* Create a new Amazon EBS volume from the snapshot, but specify a larger size than the original volume.
* Attach the new, larger volume to your Amazon EC2 instance in place of the original. (In many cases, an OS-level utility must also be used to expand the file system.)
* Delete the original Amazon EBS volume.

*AWS Definition of IOPs:*

*“IOPs are a unit measure representing****input/output operations per second****. The operations are measured in KiB, and the underlying drive technology determines the maximum amount of data that a volume type counts as a single I/O. I/O size is capped at****256 KiB for SSD****volumes and****1024 KiB for HDD****volumes.”*

EBS volume size determines the amount of IOPS. **Larger size = more IOPS**.

**“Root” vs Additional EBS Volumes**

(1) Every EC2 instance MUST have a “root” volume, which may or may not be EBS

(2) By default, EBS “root” volumes are set to be deleted when the instance is terminated (you can choose to have EBS volume persist)

(3) During the creation of an EC2 instance (and afterwards) you can add additional EBS Volumes to the instance.

(4) Any Additional EBS volume can be attached or detached from the instance at any time, and is not deleted (by default) when the instance is terminated.

**Volumes vs Snapshots**

* Volume
  + A volume is a virtual hard disk (think VMDK)
  + Volumes exist on EBS
  + If you take a snapshot of a volume, this will store that volume on S3
* Snapshot
  + Point in time copy of a volume
  + Exists on S3
  + Are incremental, only the blocks that have changed since the last snap are moved to S3
  + 1st snap takes some time to create
  + Can use snap to create a new volume & change the disk type (magnetic -> GP2 or IO1 or any other combination)
  + If you want to snap a root volume, you should stop the instance before taking snap
    - If you don’t, AWS will stop it prior to taking snap.

**Snapshots:**

Go into EC2 -> Volumes -> create volume (make sure it's in the same AZ as your server!) -> Actions -> attach to server.

Use \_lsblk \_to view disks to confirm new volume attached.

Use \_file –s /dev/xvdf \_to make sure it's clean

Use \_mkfs –t ext4 /dev/xvdf \_to make file system, then \_mkdir /fileserver *to create directory*, & mount /dev/xvdf/fileserver \_to mount

**Volumes vs Snapshots – Security**

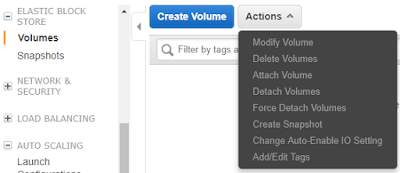
* Snapshots of encrypted volumes are encrypted automatically
* Volumes restored from encrypted snapshots are encrypted automatically
* You can share snapshots but only if they are not encrypted
* Snapshots can be shared with other AWS accounts or made public in the market place again as long as they are NOT encrypted

- A snapshot is an “image” of an EBS volume that can be **stored as a backup** of the volume OR used to **create a duplicate**. It is **NOT an active EBS volume** (cannot attach or detach to an EC2 instance.)

- To **restore**create a **new EBS volume** using the **snapshot as template**.

* You can take a snapshot of a volume, this will store that volumes snapshot on S3
* Snapshots are point in time copies of volumes
* The first snapshot will be a full snapshot of the volume and can take a little time to create
* Snapshots are incremental, which means that only the blocks that have changes since your last snapshot are moved to S3
* **If you are making a snapshot of a root volume, you should stop the instance before taking the snapshot**

*Image: AWS Console > Services > EC2 > Elastic Block Store: (EBS) Volume and Snapshots*

[](https://4.bp.blogspot.com/-LUZhoA-XBGQ/WYNCQvPC9QI/AAAAAAAAHSM/zPWXxsv5BDglwblEWwoUyFy0XeGUngICgCLcBGAs/s1600/04%2BEBS%2BVolumes%2Band%2BSnapshots.png)

**RAID Volumes:**

* If you take a snapshot, the snapshot excludes data held in the cache by applications or OS. This tends to not be an issue on a single volume, however multiple volumes in a RAID array, can cause a problem due to interdependencies of the array
* Take an application consistent snapshot
  + Stop the application from writing to disk
  + Flush all caches to the disk
* Snapshot of RAID array --> 3 Methods:
  + Freeze the file system
  + Unmount the RAID Array
  + Shutdown the EC2 instance --> Take Snapshot --> Turn it back on.

**RAID, Volumes & Snapshots**

* RAID = Redundant Array of Independent Disks
  + RAID 0 – Striped, no redundancy, good performance
  + RAID 1 – mirrored, redundancy
  + RAID 5 – good for reads, bad for writes, **AWS does not recommend ever putting RAID 5’s on EBS**
  + RAID 10 – Striped & Mirrored, good redundancy, good performance
* Why create a RAID in AWS?
  + Not getting Disk I/O that you require from GP2 or IO1 on a single volume.
* Why create a RAID in AWS?
  + Not getting Disk I/O that you require from GP2 or IO1 on a single volume.

**Placement Groups:**

* A logical group of instance in a single AZ
* Using placement groups enables applications to participate in a low latency, 10Gbps network
* Placement groups are recommended for applications that benefit from low network latency, high network throughput or both
* A placement group can't span multiple AZ's so it is a SPoF.
* Then name you specify for a placement group must be unique within your AWS account
* Only certain types of instances can be launched in a placement group. Computer Optimized, GPU, Memory Optimized, and Storage Optimized.
* AWS recommends that you use the same instance family and same instance size within the instance group.
* You can't merge placement groups
* You can't move an existing instance into a placement group
* You can create an AMI from your existing instance and then launch a new instance from the AMI into a placement group

**5.5) Security Groups**

A **security group** acts as a **virtual firewall** that controls the traffic for **one or more instances**.

* You can have any number of EC2 instances associated with a security group

- You associate one or more security groups with an instance.

- Can **modify** the rules for a security group **at any time**.

* If you edit a security group, it takes effect immediately.
* There is an implicit deny any any at the end of the security group rules.
* Can’t add a rule to deny a specific protocol inbound or outbound
* You don't need outbound rules for any inbound request. Rules are stateful meaning that any request allowed in, is automatically allowed out.

Security groups are stateful:

* If you allow a protocol inbound, automatically it’s added to outbound

- We **evaluate all the rules** from **all the security groups** that are associated with the instance.

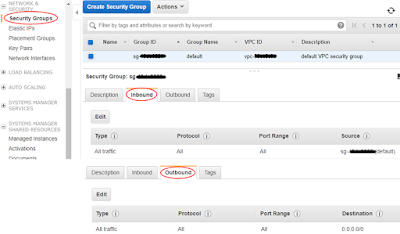
**Inbound and Outbound Rules**

- When you create a new Security Group, **ALL inbound traffic is DENIED**  (except for ssh for listros and rdp for windows) and **ALL outbound traffic is ALLOWED** by default.

- All traffic is DENIED unless there is an EXPLICIT ALLOW rule for it (there are only ALLOW rules.)

**Best Practice: Allow ONLY traffic that is required.**

*Image: AWS Console > Services > EC2 > Network & Security > Security Groups*

[](https://2.bp.blogspot.com/-Ndwpmz0w2sA/WYNCW_ue2SI/AAAAAAAAHSQ/4Tj-I6041kotOBehSVwpDa6Mg2OMf70QACLcBGAs/s1600/05%2BSecurity%2BGroups.png)

*Note: In the image above, inbound just allows traffic from whatever is in the security group.*

**5.6) IP Addressing**

- (By default) ALL EC2 instances have a **private IP address**. *Private IP addresses allow for instances to communicate with each other as long as they are located in the same VPC or broader private network.*

- EC2 instances can be launched with or without a **public IP address**, depending on VPC/subnet settings. *Public IP addresses are REQUIRED for the instance to communicate with the Internet.*

*Note: The “default” VPC and subnets are configured so that any new instance that is provisioned has a public IP address.*

**5.7) Provisioning EC2 (example)**

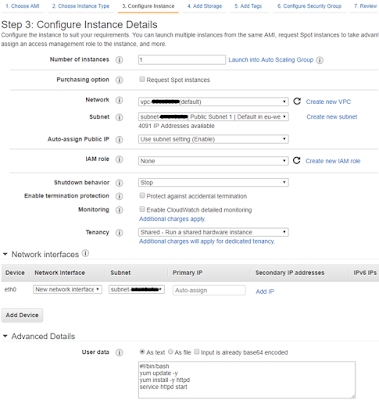
**AWS Console > Services > EC2 > Launch Instance**

Step 1) Quick Start: **Amazon Linux** **(Free tier eligible)** >**Select**

Step 2) Instance Type: **General Purpose t2.micro** **(Free tier eligible)**> **Next: Configure Instance Details**

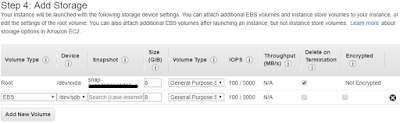
Step 3) Configure Instance Details: **Configure** > **Next: Add Storage**

*Image: Step 3) Configure Instance Details (example)*

[](https://2.bp.blogspot.com/-j0xZSj0L0jI/WYNCc3PvTpI/AAAAAAAAHSU/jafDSvMQMIcI41KDU87y-sdVCCMS-MnrgCLcBGAs/s1600/06%2BConfigure%2BInstance%2BDetails.png)

Step 4) Add Storage: **Configure > Next: Add Tags**

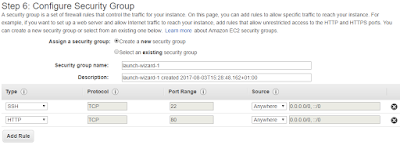
*Image: Step 4) Add Storage (example)*

[](https://2.bp.blogspot.com/-OTo6yByv9Es/WYNClTcnQlI/AAAAAAAAHSY/YAmiuYePpOsxvT29hVPfVGnqOG01nbccgCLcBGAs/s1600/07%2BAdd%2BStorage.png)

Step 5) Add Tag: **Give the ‘Name’ key a value** (optional) > **Next: Configure Security Group**

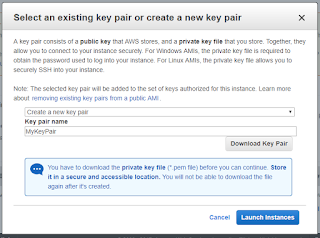
Step 6) Configure Security Group: **Configure** > **Review and Launch**

*Image: Step 6) Configure Security Group (example)*

[](https://2.bp.blogspot.com/-Vro6LLhCZxU/WYNCqItcokI/AAAAAAAAHSc/X0EhEKWpmJ82OpQ8DDWof3mlsrzxIc52ACLcBGAs/s1600/08%2BConfigure%2BSecurity%2BGroup.png)

Step 7) Review Instance Launch: **Launch** > **Create a new key pair** > **Download Key Pair** >**Launch Instance**

*Image: Step 7) Downloading Key Pair prior to Launch Instance(s)*

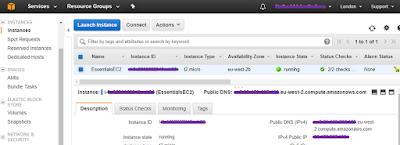
[](https://1.bp.blogspot.com/-HNq5AaGfhOI/WYNCuZn3TrI/AAAAAAAAHSg/auyuNuNujws1EOYZ9HgTYpzVGp-6KJh8ACLcBGAs/s1600/09%2BDownload%2BKey%2BPair.png)

Give it a few minutes and your instance is up and running!

[AWS Docs: Connect to Your Linux Instance](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AccessingInstances.html)

[AWS Docs: Connecting to Your Linux Instance from Windows Using PuTTY](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/putty.html)

*Image: AWS Console > Services > EC2 > Instances: Instances*

[](https://1.bp.blogspot.com/-BbThNX89MKo/WYNC1NehA2I/AAAAAAAAHSk/Ce3WacmGFfkvYWmiwUnKnFs3BzQRYITXACLcBGAs/s1600/10%2BRunning%2BEC2%2BInstance.png)

*Note: If you’re using the Free Tier, remember to stop your instances when you’re not using them.*

Roles:

* You can only assign an EC2 role to an instance on create. You can not assign a role after the instance has been created and/or is running
* You can change the permissions on a role post creation, but can NOT assign a new role to an existing instance
* Role permissions can be changed, but not swapped
* Roles are more secure then storing your access key and secret key on individual EC2 instances
* Roles are easier to manager, You can assign a role, and change permissions on that role at any time which take effect immediately
* Roles can only be assigned when that EC2 instance is being provisioned
* Roles are universal, you can use them in any region
* You can have up to 10 tags per EC2 instance.
* **Resource or Operation Default Limit**
* **Elastic IP addresses for EC2-Classic: 5**
* **Security groups for EC2-Classic per instance: 500**
* **Rules per security group for EC2-Classic: 100**
* **Key pairs: 5000**
* **On-Demand instances: Varies based on instance type**
* **Spot Instances: Varies based on instance type**
* **Reserved Instances: 20 instance reservations per Availability Zone, per month**
* **Dedicated Hosts: Up to 2 Dedicated Hosts per instance family, per region can be allocated**
* **AMI Copies: Destination regions are limited to 50 concurrent AMI copies at a time, with no more than 25 of those coming from a single source region.**
* **Throttle on the emails that can be sent : Throttle applied**
* **Tags per EC2 instance: 10**
* **ELB (Elastic Block Storage Limits)**
* **Resource or Operation Default Limit**
* **Number of EBS volumes: 5000**
* **Number of EBS snapshots: 10,000**
* **Total volume storage of General Purpose SSD (gp2) volumes: 20 TiB**
* **Total volume storage of Provisioned IOPS SSD (io1) volumes: 20 TiB**
* **Total volume storage of Throughput Optimized HDD (st1): 20 TiB**
* **Total volume storage of Cold HDD (sc1): 20 TiB**
* **Total volume storage of Magnetic volumes: 20 TiB**
* **Total provisioned IOPS: 40,000**
* **For additional information about EC2 Limits, see Limits in Amazon EC2**
* **ELB (Elastic Block Storage Limits)**
* **Resource or Operation Default Limit**
* **Number of EBS volumes: 5000**
* **Number of EBS snapshots: 10,000**
* **Total volume storage of General Purpose SSD (gp2) volumes: 20 TiB**
* **Total volume storage of Provisioned IOPS SSD (io1) volumes: 20 TiB**
* **Total volume storage of Throughput Optimized HDD (st1): 20 TiB**
* **Total volume storage of Cold HDD (sc1): 20 TiB**
* **Total volume storage of Magnetic volumes: 20 TiB**
* **Total provisioned IOPS: 40,000**