Welcome to Virtualization

Virtualization

- In computing, virtualization is a broad term that refers to the abstraction of computer resources.
- Virtualization hides the physical characteristics of computing resources from their users, be they applications, or end users.
- This includes making a single physical resource (such as a server, an operating system, an application, or storage device) appear to function as multiple virtual resources; it can also include making multiple physical resources (such as storage devices or servers) appear as a single virtual resource

- In layman's terms virtualization is often:
  - The creation of many virtual resources from one physical resource.
  - The creation of one virtual resource from one or more physical resource.

Types of Virtualization

- Today the term virtualization is widely applied to a number of concepts including:
  - Server Virtualization
  - Client / Desktop / Application Virtualization
  - Network Virtualization
  - Storage Virtualization
  - Service / Application Infrastructure Virtualization

EC2

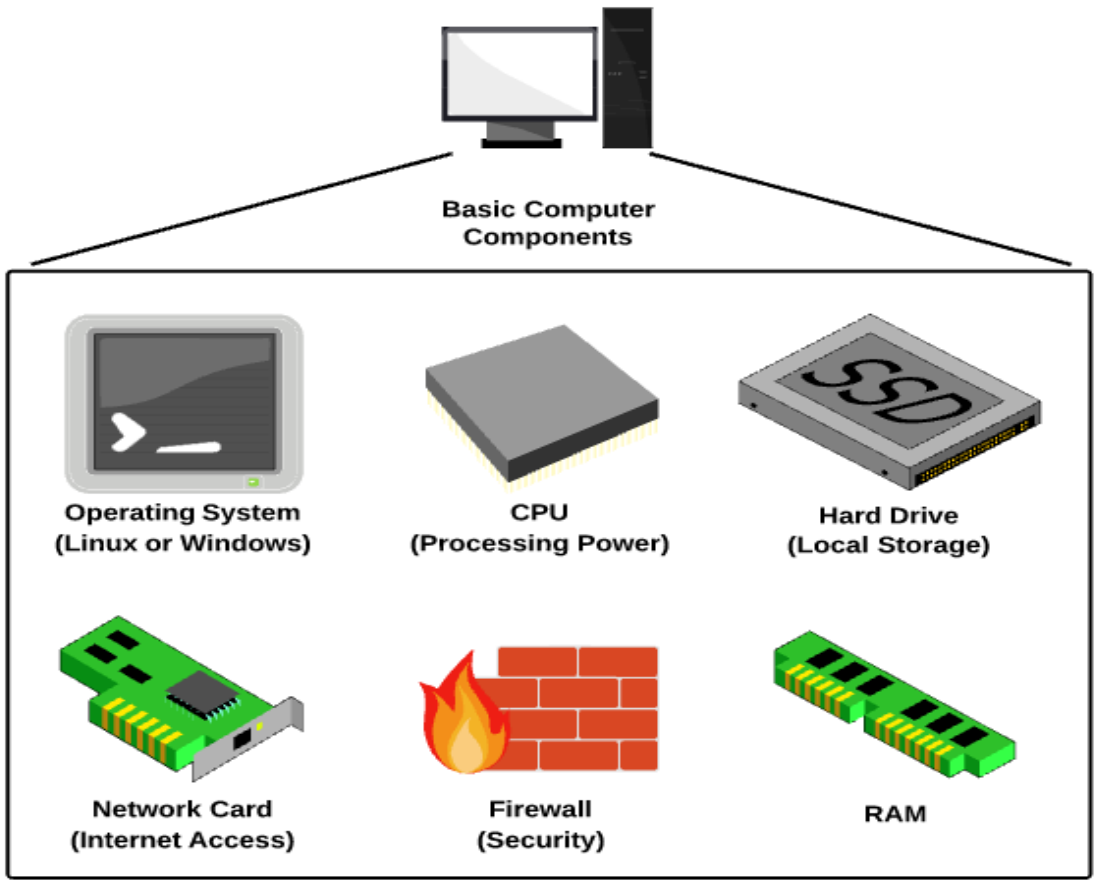
- Think of EC2 as your basic desktop computer
- Amazon Elastic Compute Cloud (EC2) is a web service that provides resizable compute capacity in the cloud.
- Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.
- Configure Security and Networking and Manage Storage
- EC2 uses two different hypervisors, depending on the generation of instance that you are using. The earlier generation instances used a modified version of the **Xen** hypervisor, and the newer generation uses the **Nitro** hypervisor which was developed by AWS themselves, and is based on a customized version of KVM (Linux Kernel-based Virtual Machine (KVM) technology)

[Instance types - Amazon Elastic Compute Cloud]

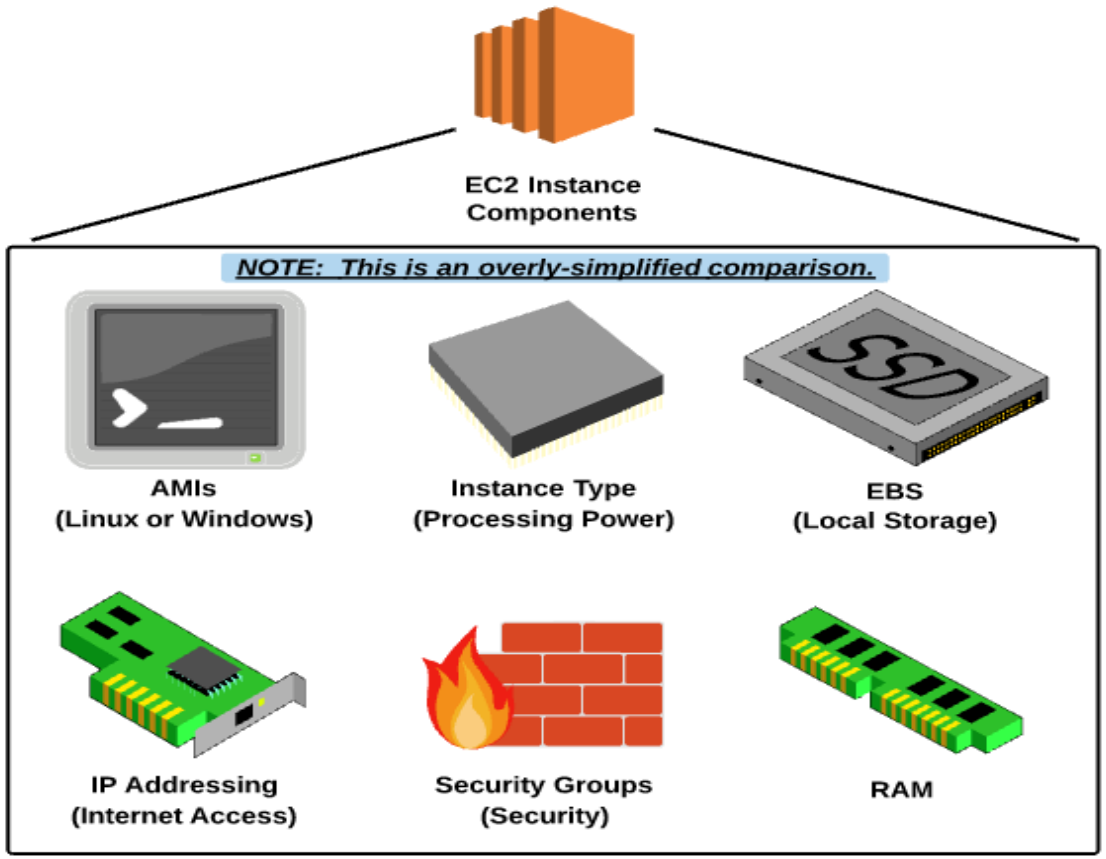
EC2 Basics

EC2 Basics:

Conceptually Understanding EC2:



Conceptually Understanding EC2:



Instance Types

- The instance type is the CPU (Compute Power) of your instance. Each instance type offers different compute, memory and storage capabilities

EC2 instance types

	General Purpose		Compute Optimized	Memory Optimized		Accelerated Computing	Storage Optimized		
Type	t2	m5	c5	r4	x1e	p3	h1	i3	d2
Description	Burstable, good for changing workloads	Balanced, good for consistent workloads	High ratio of compute to memory	Good for in-memory databases	Good for full in-memory applications	Good for graphics processing and other GPU uses	HDD backed, balance of compute and memory	SDD backed, balance of compute and memory	Highest disk ratio
Mnemonic	**t** is for **tiny** or **turbo**	**m** is for **main** or happy **medium**	**c** is for **compute**	**r** is for **RAM**	**x** is for **xtreme**	**p** is for **pictures**	**h** is for **HDD**	**i** is for **IOPS**	**d** is for **dense**

Family	Speciality	Use case
F1	Field Programmable Gate Array	Genomics research, financial analytics, real-time video processing, big data etc
I3	High Speed Storage	NoSQL DBs, Data Warehousing etc
G3	Graphics Intensive	Video Encoding/ 3D Application Streaming
H1	High Disk Throughput	MapReduce-based workloads, distributed file systems such as HDFS and MapR-FS
T2	Lowest Cost, General Purpose	Web Servers/Small DBs
D2	Dense Storage	Fileservers/Data Warehousing/Hadoop
R4	Memory Optimized	Memory Intensive Apps/DBs
M5	General Purpose	Application Servers
C5	Compute Optimized	CPU Intensive Apps/DBs
P3	Graphics/General Purpose GPU	Machine Learning, Bit Coin Mining etc
X1	Memory Optimized	SAP HANA/Apache Spark etc

Key Pairs

- A Key pair consists of a public key that AWS stores, and a private key file that you stores. Together they allow you to connect to your instance securily.
- For Linux AMIs, the private key file allows you to securily SSH into your instance.
- For Windows AMIs, the private key file is required to obtain the password used to log into your instance.

Security Groups

- A Security Group is a Virtual Firewall
- 1 instance can have multiple Security Groups
- Inbound and Outbound rules
- Stateful, anything you added to inbound rules, outbound rules are added automatically
- All inbound Traffic is blocked by default
- All outbound Traffic is Allowed
- Changes to Security Groups take effect Immediately
- You can have any number of EC2 instances within a Security Group
- You cannot block specific IP addresses using Security Groups, instead use Network Access Control Lists.

Elastic Block Storage (EBS)

- Amazon EBS allows you to create Storage volumes and attach them to Amazon EC2 instances.
- Once attached, you can create a file system on top of these volumes, run a database, or use them in any other way you would use a block device.
- Amazon EBS volumes are placed in a specific Availability Zone, where they are automatically replicated to protect you from the failure of a single component.
- Virtual Disk in the Cloud

Before we move any further, we need to learn about IOPS

What are IOPS?
IOPS = Input/Output Operations per Second

Simplified definition:
The amount of data that can be written to or retrieved from EBS per second.

AWS Definition:
"IOPS are a ***unit of 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 1,024 KiB for HDD volumes because SSD volumes handle small or random I/O much more efficiently than HDD volumes."

What does this all mean?
More IOPS means better volume performance (faster read/write speeds)

What determines the amount of IOPS?
EBS volume size. The larger the storage size (in GiB), the more IOPS the volume has.

Elastic IP and Dynamic IP

- Elastic IP address is a public **static IPv4 address** which is reachable from the Internet.
- Elastic IP addresses are used by AWS to manage its dynamic cloud computing services.
- *Elastic IP to an Instance which doesn't change after you stop / start the instance Whereas Dynamic IP will get changed whenever you start and Stop the EC2 Instance*
- In short Elastic IP is a permanent IP for your instance and Dynamic IP is a temporary for your instance

Q & A