Amazon EC2

→ EC2 is one of the most popular of AWS offering
→ EC2 = Elastic Compute Cloud = Infrastructure as a Service
→ It mainly consists in the capability of:
- Renting Virtual Machines (EC2)
- Storing data on Virtual drives (EBS)
- Distributing load across machines (ELB)
- Scaling the services using an auto-scaling group (ASG)

* knowing EC2 is fundamental to understand how the cloud works

* EC2 Sizing & Configuration Options.

→ Operating System (OS): Linux, Windows or MacOS
→ How much compute power & cores (CPU)
→ How much random-access memory (RAM)
→ How much storage space:
- Nlw-attached (EBS & EFS)
- Hardware (EC2 instance Store)
→ Network card: speed of the card, public IP address
→

→ Firewall rules : security group
→ Bootstrap script (configure at first launch): EC2 user data

EC2 User Data

→ It is possible to bootstrap our instances using an EC2 user data script.
→ bootstraping means launching commands when machine starts.
→ that script is only run once at the instance first start.
→ EC2 user data is used to automate boot tasks such as:
  - Installing updates
  - Installing software
  - Downloading common files from the internet
  - Anything you can think of
→ The EC2 User Data Script runs with the root user.

EC2 Instance Types -

① General Purpose :-

① Great for a diversity of workloads such as web servers or code repositories.

② Balance between: Compute
Memory
Networking

③ In the ~~cord~~ course, we will be using the t2.micro which is a general Purpose EC2 instance.

* General purpose instances provide a balance of compute, memory & networking resources, & can be used for variety of diverse workloads. The instances are ideal for applications that use these resources in equal proportions such as web servers & code repositories.

② Compute Optimized :→

* Great for compute-intensive tasks that require high performance processors:

① Batch processing workloads
② Media transcoding
③ High performance web servers
④ High performance computing (HPC)
⑤ Scientific modeling & machine learning
⑥ Dedicated gaming servers.

③ Memory Optimized :-

* Fast performance for workloads that proces large data sets in memory
* Use cases:

① High performance, relational / non-relational DB
② Distributed web scale cache stores
③ In-memory databases optimized for BI
④ Applications performing real-time processing of big unstructured data.

④ Storage Optimized :→

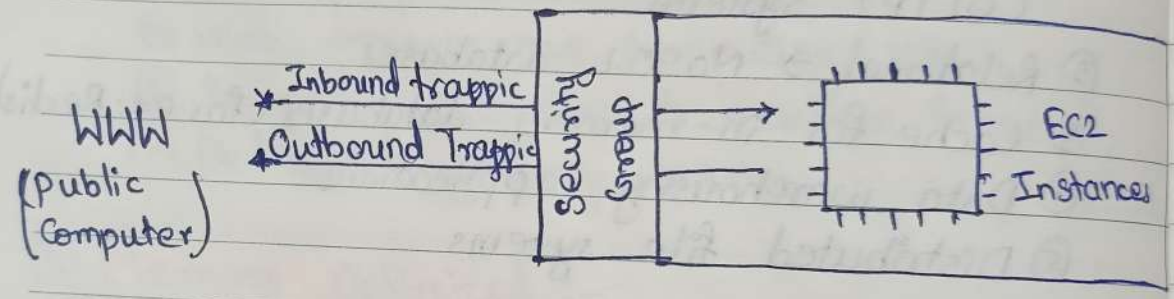
Great for storage-intensive tasks that require high, sequential read & write access to large data sets on local storage.

* Use Cases :

① High frequency online transaction processing (OLTP) systems
② Relational & NoSQL databases.
③ Cache for in-memory databases (for ex: Redis)
④ Data warehousing applications
⑤ Distributed file systems

Introduction to Security Groups :→

① Security Groups are the fundamental of network security in AWS.
② They control how traffic is allowed into or out of our EC2 instances.



③ Security groups only contain allow rules
④ Security groups rules can reperence by Ip or by security groups.

Security Groups :-

* Security groups are acting as a "firewall" on EC2 instances.
* They regulate : ① Access to ports
② Authorised Ip ranges - IPV4 & IV6
③ Control of inbound n/w (from other to the instance)
④ Control of outbound n/w (from the instance to other)

ex of security rule

Type	Protocol	Port Range	source	destination
HTTP	TCP	80	0.0.0.0/0	test HTTP page

Port Range :- Where the traffic can go trough on the instance

source :- which represents an IP address range & 0.0.0.0/0 means everything.

Good to know

① Can be attached to multiple instances
② Locked down to a region /VPC combination
③ Does live "outside" the EC2 - if traffic is blocked, the EC2 instance won't see it.
④ It's good to maintain one seperate security group for ssh access
⑤ If your applications is not accessible (time out) then it's a security group issue
⑥ If your appl" gives a "connection* refused" error, then its an appl" error or it is not launched.

② All inbound trappic is blocked by depault.
③ All outbound trappic is authorised by depa

* Repering

* Reperencing other security groups Diagram

* Suppose we have one EC2 instance and has 2 security group 1 inbound and has ~~two rules~~ Authorising security group 1 & group 2

* Now we have 3 more EC2 instances which has security group 2 all, group 1, group attached to them respectively.

* Now these 3 instances are trying to connect to instance through port 123.

* As, 2 out of 3 instances out of 3 has security group 2 & 1 attached respectively. firewall will allow this connection,
but 3rd instance will get rejected as repered/attached security group to itself is not there in instance to which it is trying to connect

Imp **Classic Ports to know**

① 22 = SSH (Secure Shell) - log into a Linux instance

② 21 = FTP (File transfer Protocol) - upload files into file share

③ 22 = SFTP (Secure file Transfer Protocol) - upload files using SSH

④ 80 = HTTP - access unsecured websites

⑤ 443 = HTTPS - access secured websites

⑥ 3389 = RDP (Remote Desktop Protocol) - log into a windows instance

SSH Overview

How do you connect inside of your servers to perform some maintenance or action:

for this for Linux servers we can use SSH to do secure shell into our servers

	SSH	Putty	EC2 Instance Connect
MAC	✓		✓
Linux	✓		✓
Windows<10		✓	✓
Windows≥10	✓	✓	✓

EC2 instance uses web browser to connect to EC2 instance.
- but it works only with Amazon NX2

EC2 Instances Purchasing Options

① On-Demand Instances:- Short workload, predictable pricing, pay by second

② Reserved (1 & 3 Year) (upto 72% discount)
Reserved Instances - long workloads
Convertible Reserved Instances - Long workloads with flexible instances
↗ (upto 66% discount)

③ Saving Plans (1 & 3 Years) - commitment to an amount of usage, long workloads

④ Spot Instances - Short workloads, cheap, can lose instances (less reliable)

⑤ Dedicated Hosts - book an entire physical server, control instance placement.

⑥ Capacity Reservations - reserve capacity in a specific AZ for any duration.

① EC2 On Demand :-

① Pay for what you use:
- Linus or Windows - billing per second, after the first minute
- All other operating systems - billing per h

② Has the highest cost but no upfront paym

③ No long-term commitment

④ Recommended for short-term & un-interrupt workloads, where you can't predict how appln will behave

② EC2 Reserved Instances

② EC2 Reserved Instances

① Upto 72% discount compared to on-demand

② You reserve a specific instance attributes (Instance type, Region, Tenancy, OS)

③ Reservation Period - 1 year (+ discount) or 3 years (+++ discount)

④ Payment Options - No upfront (+), Partial Upfront (++), All upfront (+++)

⑤ Reserved Instance's scope - Regional or zonal (reserve capacity in an AZ)

⑥ Recommended for steady-state usage appl^n's (think database)

⑦ You can buy & sell in the Reserved Instance Marketplace

* Convertible Reserved Instance :- Can change the EC2 instance type, instance family, OS, scope & tenancy. → Upto 66% discount

③ EC2 Saving Plans :-

① Get discounts based on long-term usage (upto 72% same as RIs)

② Commit to a certain type of usage (

③ Usage beyond EC2 savings Plans is billed at the On-demanded price

④ Locked to a specific instance family & AWS region.

⑤ Flexible across -

i) Instance size (e.g. m5.xlarge, m5.2xlarge)
ii) OS (Linux, Windows)
iii) Tenancy (Host, Dedicated, default)

④ EC2 Spot Instances :-

① can get a discount of upto 90% to compared to on-demand.

② Instances that you can "lose" at any point of time if your max price is less than the current spot price.

③ The most cost-efficient instances in AWS.

④ Useful for workloads that are resilient to failure :- Batch jobs
- data analysis
- Image processing
- Any distributed workloads
- Workloads with flexible start & ending

⑤ Not suitable for critical work.

⑤ EC2 dedicated Host :-

① A physical server with EC2 instance capac fully dedicated to your use.

② Allows you address compliance requirements & use your existing server bound software licenses (per-socket, per-core, pe-VM software diecenses)

③ Purchasing Options:

On-demand - pay per second for active dedicated host

Reserved - 1 or 3 years (

④ Most expensive option

⑤ Useful for software that have complicated licensing model (BYOL - Bring your own license)

⑥ Or for companies that have strong regulatory or compliance needs

⑥ EC2 dedicated Instances :-

① Instances run on hardware that's dedicated to you.

② May share hardware with other instances in same account.

③ No control over instance placement (can move hardware after stop/start)

⑦ EC2 Capacity Reservations :-

① Reserve On-Demand instances capacity in a specific AZ for any duration

② You always have access to EC2 capacity when you need it.

③ No time commitment (create/cancel anytime), no billing discounts

④ Combine with Regional Reserved Instances & savings plans to benefit from billing discount

⑤ You are charged at on-demand rate whether you run instances or not

⑥ Suitable for short-term, uninterrupted workloads that needs to be in a specific AZ

Shared Responsibility Model for EC2

AWS	User
① Infrastructure (global network security)	① Security groups rules
② Isolation on physical hots	② Operating-system patches & updates.
③ Replacing faulty hardware	③ software & utilies installed on the EC2 instance
④ Compliance validation.	④ IAM Roles assigned to EC2 & IAM ~~roles~~ user access management.
	⑤ Data security on your instance.

EC2 Section Summary

- EC2 (Elastic Compute Cloud): Iaas
- EC2 Instance: AMI (OS) + Instance Size (CPU + RAM) + Storage + Security Groups + EC2 User Data
- Security Groups: Firewall attached to the EC2 instance
- EC2 User Data: Script launced at the first start of an instance.
- SSH: start a terminal into our EC2 instances (PORT 22)
- EC2 Instance Role: link to IAM Roles
- Purchasing Options: On-Demand, Spot, Reserved (std, convertible), Dedicated Host, Dedicated Instance.