EC2

Elastic Compute cloud

Budget setup:
- Say 80% of use then trigger.
  Aws account -> Billing -> Budget.

EC2 = IAAS

Mainly consists in the capability of-
- Renting virtual machines (EC2)
- Storing data on virtual drives (EBS)
- Distributing load across machines (ELB)
- Scaling the service using autoscaling group (ASG)

EC2 sizing and configuration option-
- OS - Linux, Windows, MacOS
- CPU - how much compute power & cores
- RAM
- Storage space-
  1) Network attached (EBS & EFS)
  2) hardware (EC2 Instance store)
- Network card: speed of card, public IP address.
- Firewall rules- security group
- Bootstrap script: (configure at first launch): EC2 User Data.

EC2 User Data
- Used to bootstrap our instances at start of machine.
- runs only once at the instance start.
- runs with root user.

EC2 Instance types-

t2.micro. t2.xlarge, c5d.4xlarge etc
free tier (750 hrs)

Configuring EC2 instance-
- Choose AMI (Amazon Machine Image)
- Choose Instance type
- Configure instance
- Add Storage
- Add tags.
- Configure security grp.
- Review

EC2 Instance Types

7 types- optimised for different use cases

Naming convention-

m5.2xlarge

m: instance class

5: generation of instance (hardware)

2xlarge: size within the instance class

General Purpose
Compute Optimized
Memory optimized
Accelerated computing
Storage Optimized
Instance Features
Measuring Instance Performances

} Types of instances in EC2

General Purpose: great for diversity of workloads such as web servers or code repositories.

Balance b/w:

1) Compute
2) Memory
3) Networking.

eg - t2.micro.

Compute Optimized - intensive tasks that require high performance processors. (C)

- Batch processing workloads
- Media transcoding
- High performance web servers.
- High performance computing
- Scientific modeling & machine learning
- Dedicated gaming servers.

eg: C5.C6, etc

Memory Optimized: fast performance for workloads that process large datasets in memory. (R)

Use cases:

- High performance relational / non-relational db
- Distributed web scale cache stores
- In memory db optimized for BI (business intelligence)
- Applications performing real time processing of big unstructured data.

eg R, X1, High memory, z1d

Storage Optimized Instances: great for storage intensive tasks that require high, sequential read and write access to large datasets on local storage.

Use:

- High freq. online transaction processing (OLTP) systems
- Relational and NoSQL db
- cache for in memory db. (Redis)
- Data warehousing apps
- Distributed file system.

Start with. I3, I3en, D2, D3, H1, etc

Security Groups.

- control how traffic is allowed into or out of EC2 instances.
- only contain allow rules.
- rules can reference by IP or security group
- are a firewall on our EC2 instance

They regulate:

- Access to ports
- Authorized IP ranges IPV4 & IPV6
- Control inbound & outbound traffic
- can be attached to multiple instances

0.0.0.0/0 => everything OR ::/0

122.149.196.85/32 => only one

- An instance can have multiple SG.
- Locked down to region/VPC combo.
- Does live outside EC2 => if traffic is blocked the EC2 instance won't even see it
- Timeout => Security grp issue

  connection refused => application error

Default - Inbound - Blocked

Outbound - Allowed

Ports

SSH	22	login to linux instance
FTP	21	upload files into a fileshare
SFTP (Secure FTP)	22	upload files using SSH
HTTP	80	access unsecured websites
HTTPS	443.	access secure websites
RDP	3389	log into a windows instance.

SSH -

	SSH	Putty	EC2 Instance connect
Mac	✓	✗	✓
Linux	✓	✗	✓
Window<10	✗	✓	✓
Windows>=10	✓	✓	✓

Step 1: chmod 0400 EC2KeyPair.pem

Step 2: ssh -i EC2KeyPair.pem ec2-user@192.168.01.01

Key file. Public IP of instance.

EC2 Instance launch types

- On demand instances: short workload, predictible pricing

- Reserved: minimum: 1 year
  1) Reserved Instances for long workloads
  2) Convertible Reserved Instances: long workload, flexible instances
  3) Scheduled Reserved Instances: every monday 3-6pm etc

- Spot Instances: short workloads, cheap, can loose instances (not reliable)

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

EC2 on demand -

- Pay for what you use
  linux/Windows - billing per second, after first minute
  others - per hr
- Has the highest cost but no upfront payment.
- No long term commitment

Use for short term and un-interrupted workloads where u cant predict how the application will behave.

EC2 Reserved Instances:

- Upto 75% discount compared to on demand
- 1yr/3yr more yr more discount. (1yr or 3yr no in between)
- Purchasing options: no upfront | partial = +| All upfront = + + discount
- Reserve a specific instance type
- Use for steady stage usage apps (think db).

Convertible Reserved Instance:
- can change instance type
- upto 54% discount.

Scheduled Reserved Instance:
- time window
- still need commitment 1-3 yrs

Spot Instance

- Highest discount - 90% compared to on demand
- can lose at any time if your spot price is less than current spot price.
- most cost efficient
- Useful for workloads resilient to failure

- Batch jobs
- Data analysis
- Image processing
- Distributed workload.
- workload with flexible start & end time.

- Not suitable for critical jobs or databases.

EC2 Dedicated Host

- is a physical server with EC2 instance capacity fully dedicated to your use.
- can help address compliance requirements and reduce costs by allowing you to use your existing server bound software licenses.
- min 3 yr. reservation.
- more expensive.
- useful when complicated licensing models. (BYOL - Bring Your Own Licen
- Or for companies that have strong regulatory or compliance needs.

EC2 Dedicated Instances:

- You will have dedicated hardware but u cant access it
- Instances running on hardware thats dedicated to you.
- may share hardware with other instance in same account
- no control over instance placement (can move hardware after stop/ start).

Shared Responsibility Model in EC2.

AWS	User.
- Infrastructure (global network security)	- Security grp rules.
- Isolation on physical hosts.	- OS patches and updates.
- Replacing faulty hardware.	- Software and utilities installed on the EC2 instance
- Compliance validation.	- IAM roles assigned to EC2 & IAM user access mgmt.
	- Data security on your instance.