

Amazon EC2

Using EC2 for compute is highly flexible compared to

own servers

→ quick

on premises in a data center that you own

Amazon EC2

- two built datacenters
 - two secured data centers
 - two purchased servers
 - two installed servers
 - the servers are online & ready to use
- } already

→ to use AWS we need to do a request to AWS to launch and boot up, ready to be used within a few minutes.

Stop or terminate the EC2 instance. Once you done, you can easily

→ And you only pay for what you use, not for running.

In EC2 we need to pay for running instance & not for stopped & terminated instances.

→ EC2 runs on top of physical host machines managed by AWS using virtualization technology.

Multitenancy: sharing underlying hardware between virtual machines

Hyper hypervisor is responsible for coordinating multitenancy and it is managed by AWS.

- > ~~loop~~ The hypervisor is responsible for isolating the virtual machines from each other as they share resources from the host host . this means EC2 instances are secure , even though they share resources . one EC2 instance is not aware that another EC2 instance is also on that host . they are secure and separate from each other
- When you provision an EC2 instance you can choose the operating system based on either Windows & Linux
- Beyond the OS , you also configure what software is meant running on the instance whether its your own internal business app , simple web app & complex web apps , databases ; third - party software

Amazon EC2 configurations

1. Windows
2. Linux
3. Internal business app
4. Web app
5. Databases
6. Third - party software

- you have complete control over what happens on that instance
- EC2 are also resizable
- you can also control the networking aspect of Amazon EC2 . means what type of requests make it to your server and if they are publicly or privately accessible it can be decided by us .

Scaling Amazon EC2

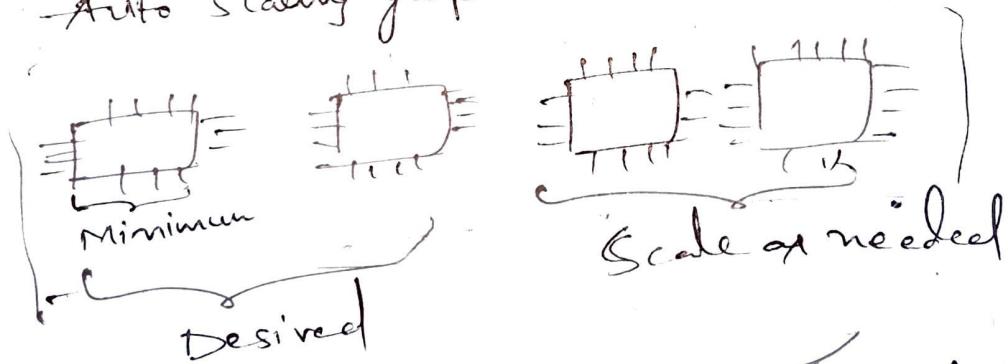
AWS services provide the functionality for Amazon EC2 instance scaling.

Amazon EC2 Auto Scaling • Means Amazon EC2 Auto scaling enables you to automatically add & remove Amazon EC2 instances in response to changing application demand.

Within Amazon EC2 Auto scaling, you can use 2 approaches:

- Scale up → 1. Dynamic scaling: responds to changing demand
- Scale down → 2. Predictive scaling: automatically schedules the right number of Amazon EC2 instances based on predicted demand

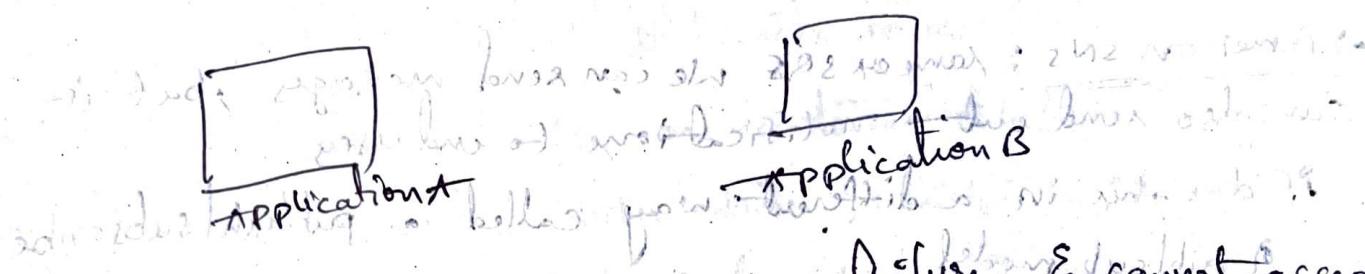
- Auto scaling group



✓ Maximum Amazon EC2 instance

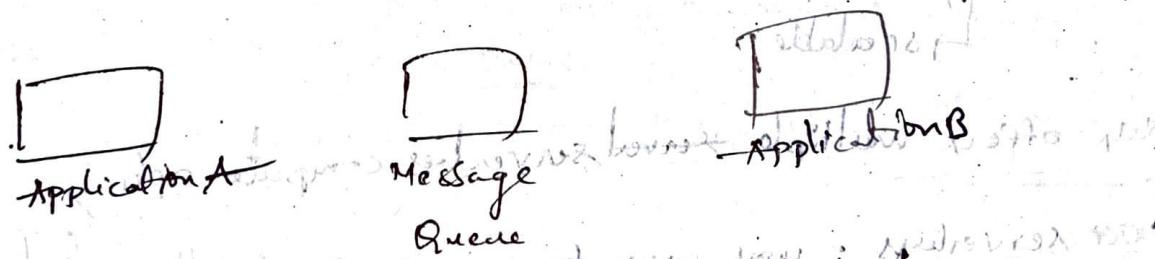
Messaging & queuing

Tightly coupled architecture



→ sending msg directly to B. If B has failure & cannot accept those messages. A will begin to see backlog of messages. This is tightly coupled architecture.

Loosely coupled architecture: Single failure won't cause cascading failures. If component fails, it is isolated & therefore won't cause cascading failure throughout the whole system.



→ if B fails, A doesn't experience any disruption, messages being sent can still be sent to the queue & will remain there until they are eventually processed.

→ AWS Services → Amazon Simple Queue Service (Amazon SQS)

→ Amazon simple Notification Service (Amazon SNS)

→ Amazon SQS (Send message, store messages, receive message between software components at any volume)

Messages on queue board have:

→ person's name

→ coffee order

Data contained with message is

→ type they ordered