

A faint, light gray background graphic on the left side of the slide. It depicts a complex network of interconnected nodes and lines, resembling a web or a data structure. The nodes are represented by small circles, some of which are slightly larger and more prominent than others. The lines connecting them are thin and gray, creating a mesh-like pattern.

# **COMPUTE**

## **EC2, Auto Scaling**

# Agenda

---

- **EC2**

- Introduction of EC2
- Benefits of EC2
- Launching EC2 Instance
- EC2 Limits

- **Auto Scaling**

- Introduction of Auto Scaling
- Auto Scaling Functionality
- Auto Scaling Limits

# Introduction of EC2








- **Resizable** compute capacity
- Complete control of your computing resources
- **Reduces the time required** to obtain and boot new server instances to minutes

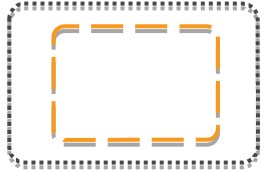



## → Amazon EC2 Facts

- Scale capacity as your computing requirements change
- Pay only for capacity that you actually use
- Choose Linux or Windows
- Deploy across AWS Regions and Availability Zones for reliability

# Elastic Compute Cloud (EC2)

- Virtual computing environments, known as **EC2 instances** 
- Pre-configured templates for EC2 instances, known as **Amazon Machine Images (AMI)** 
- AMI includes the **operating system and additional software**
- Various configurations of CPU, memory, storage, and networking capacity for your instances, known as **Instance types** 
- Secure login information for your instances using **key pairs** 
- **Instance store volumes** are used for temporary data that's deleted when you stop or terminate your instance 
- Persistent storage volumes for data using **Elastic Block Store (EBS)**

# EC2 (Cont..)

- **Regions and Availability Zones** are multiple physical locations for your resources, such as instances and EBS volumes A diagram showing a large dotted rectangle representing a region, with a smaller dashed orange rectangle inside representing an availability zone.
- A firewall to specify the protocols, ports, and source IP ranges that can reach your instances using **Security Groups** A diagram of a red rectangle with a dashed border, representing a security group.
- Static IP addresses, known as **Elastic IP addresses**
- Metadata, known as **tags**, can be created and assigned to EC2 resources A diagram of a blue tag with the word 'TAG' written on it, attached to a small ring.
- Virtual networks that are logically isolated from the rest of the AWS cloud, and can optionally connect to on premise network, known as **Virtual private clouds (VPCs)** A diagram of a yellow cloud shape with the letters 'VPC' written inside in white.

# Benefits of EC2

---

- Elastic Web-Scale Computing
- Completely Controlled
- Flexible Cloud Hosting Services
- Integrated
- Reliable
- Easy to Start
- Inexpensive
- Secure

# EC2 Instance Purchasing Options

- On-Demand instances
- Reserved Instances
- Scheduled Instances
- Spot instance
- Dedicated Hosts

## On-Demand Instance

Pay by the hour

## Reserved Instance

Purchase, at a significant discount, instances that are always available

1- year to 3- year terms.

## Scheduled Instance

Purchase a 1- year recurring period of time

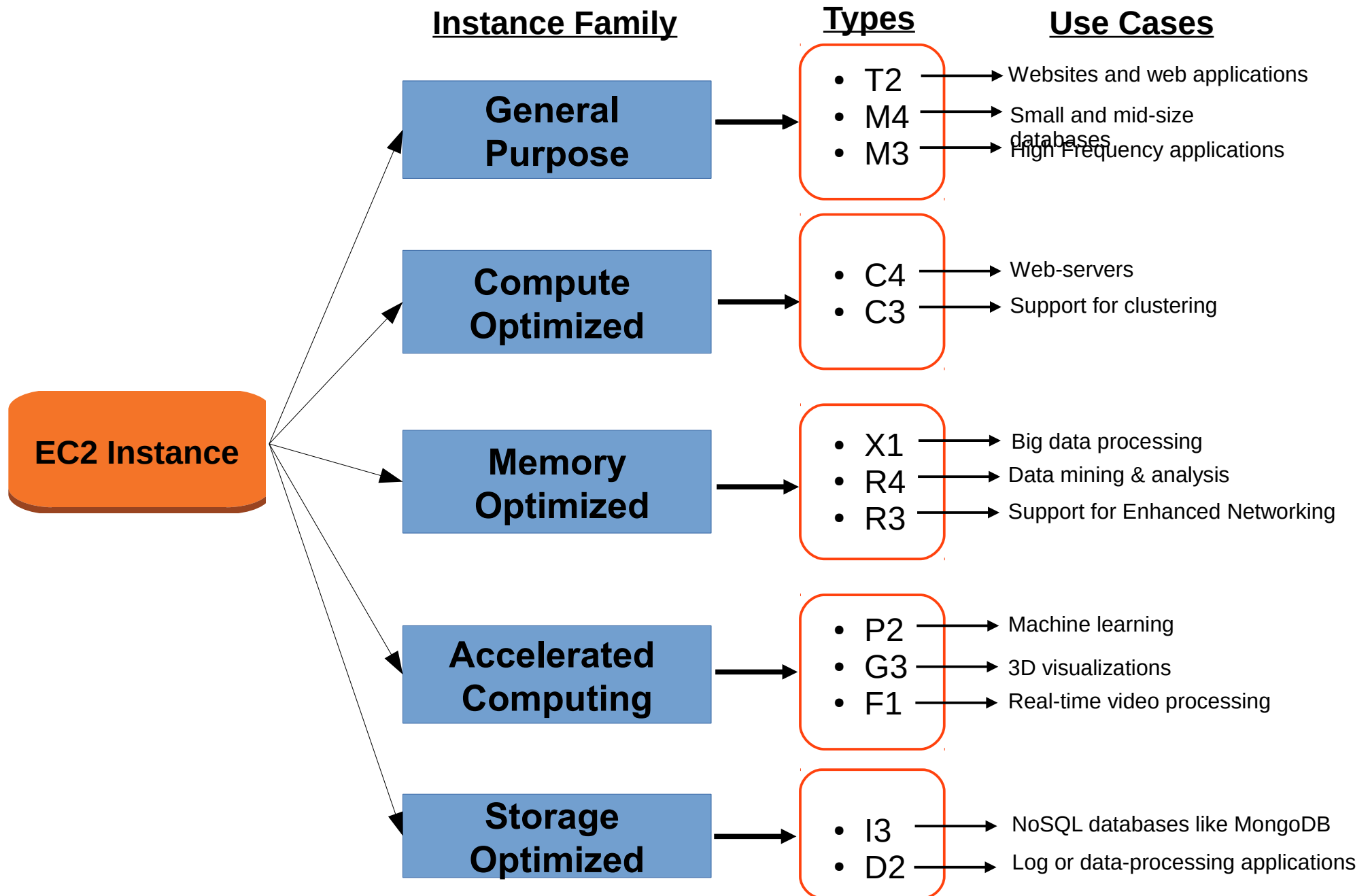
## Spot Instance

Bid on unused instances, which can run as long as they are available and your bid is above the Spot price, at a significant discount.

## Dedicated Hosts

Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs.

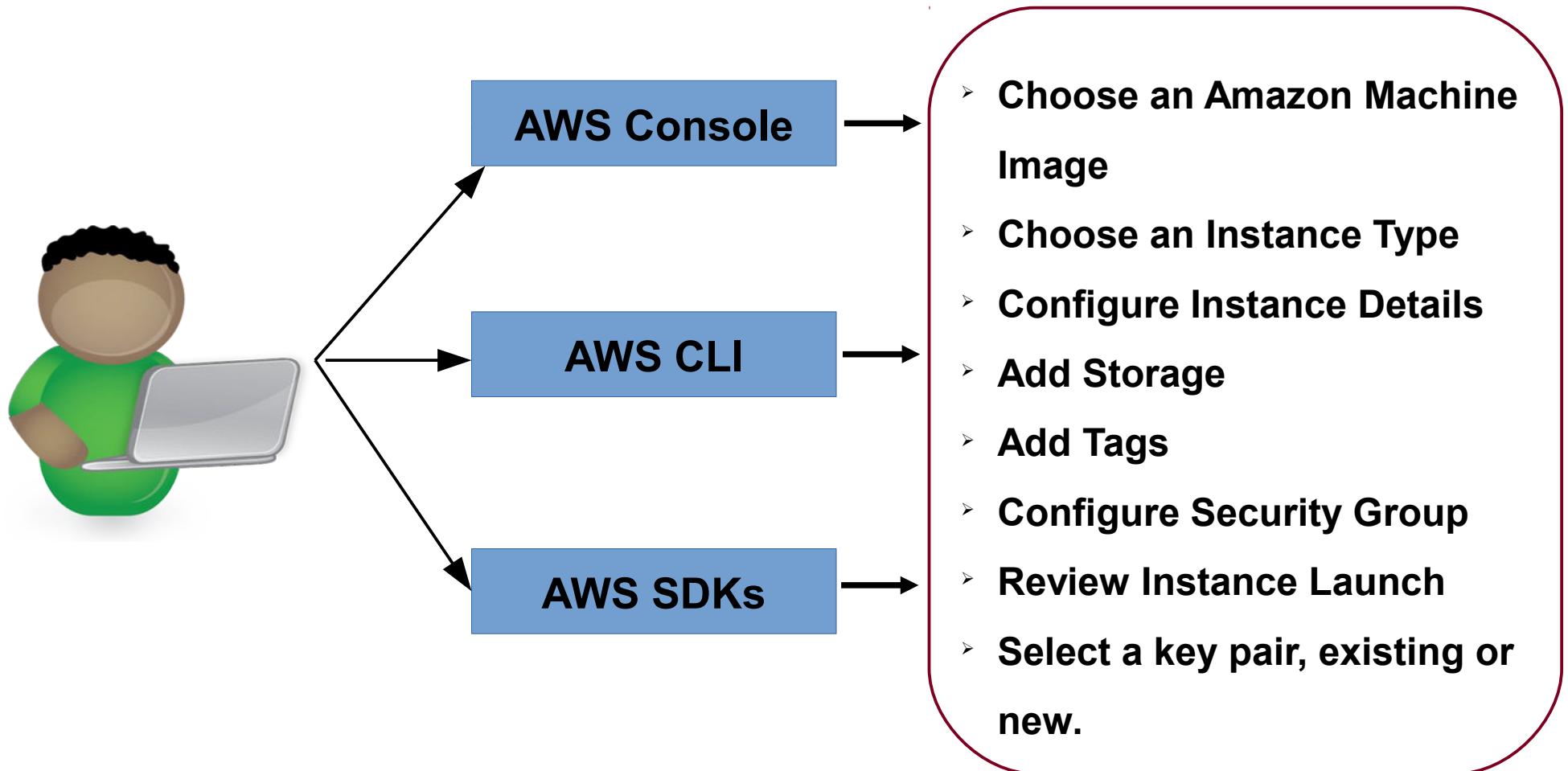
# EC2 Instance Types



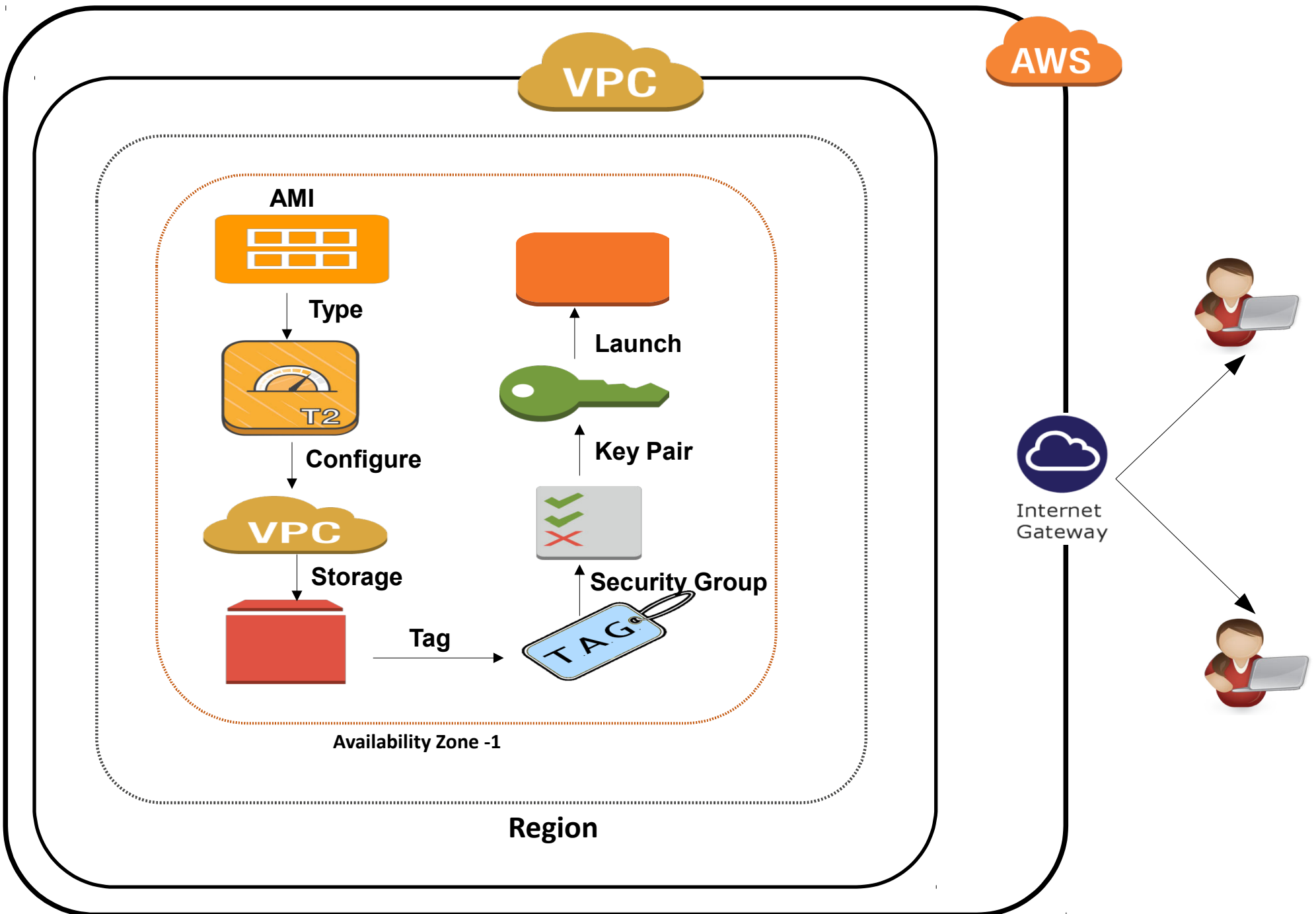


# EC2 Launching Functionality

- Amazon EC2 Launching Functionality



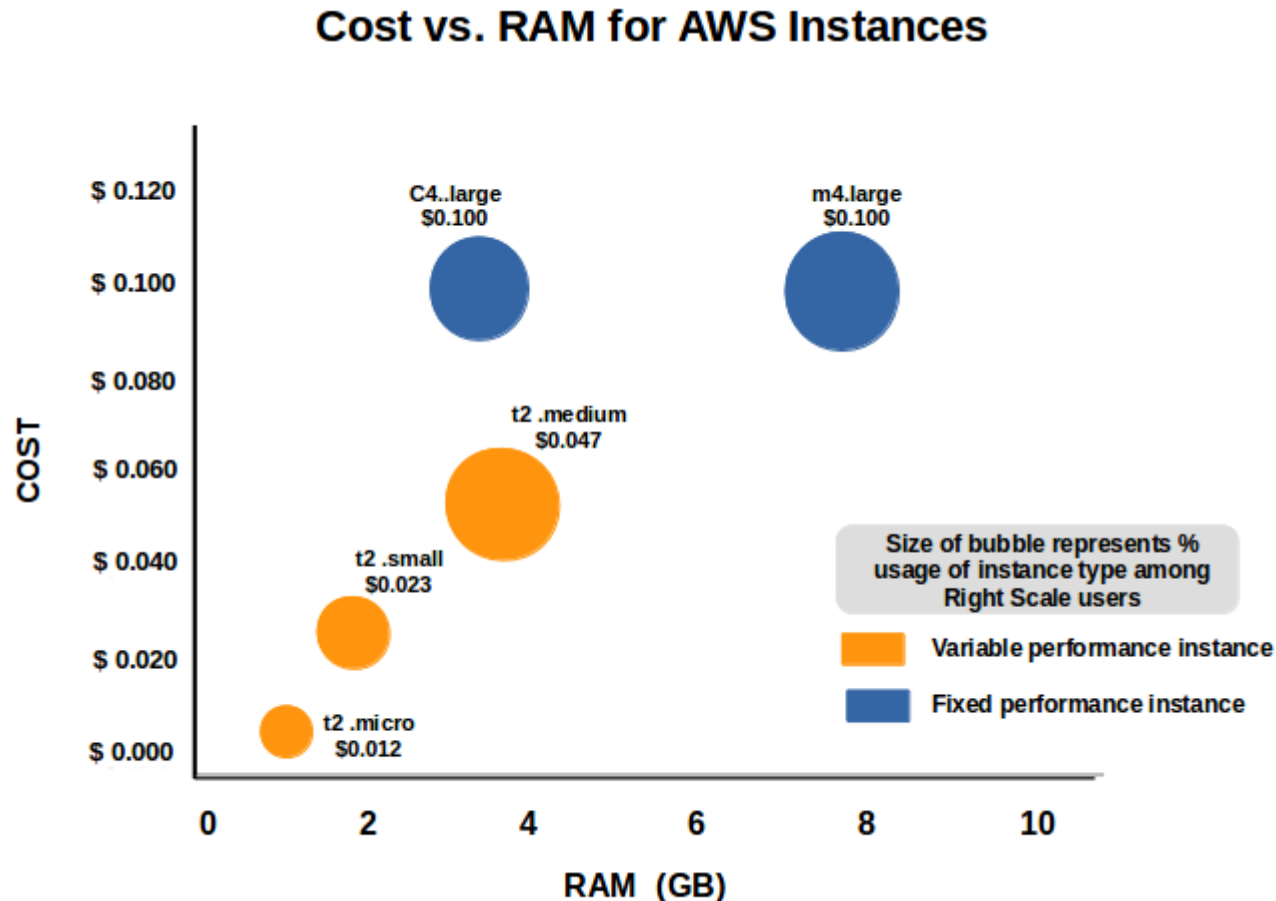
# How to Launch EC2 Instance



# AWS EC2 Limits

S No	Resource	Default Limit
1.	Elastic IP addresses for EC2-Classical	5
2.	Security groups for EC2-Classical per instance	500
3.	Rules per security group for EC2-Classical	100
4.	Key pairs	5,000
5.	Throttle on the emails that can be sent from your Amazon EC2 account	Throttle applied
6.	On-Demand Instances	Limits vary depending on instance type. For more information, see <a href="#">How many instances can I run in Amazon EC2</a> .
7.	Spot Instances	Limits vary depending on instance type, region, and account. For more information, see <a href="#">Spot Instance Limits</a> .
8.	Reserved Instances	20 Reserved Instances per Availability Zone, per month, plus 20 regional Reserved Instances. For more information, see <a href="#">Reserved Instance Limits</a> .
9.	Dedicated Hosts	Up to two Dedicated Hosts per instance family, per region can be allocated.
10.	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.

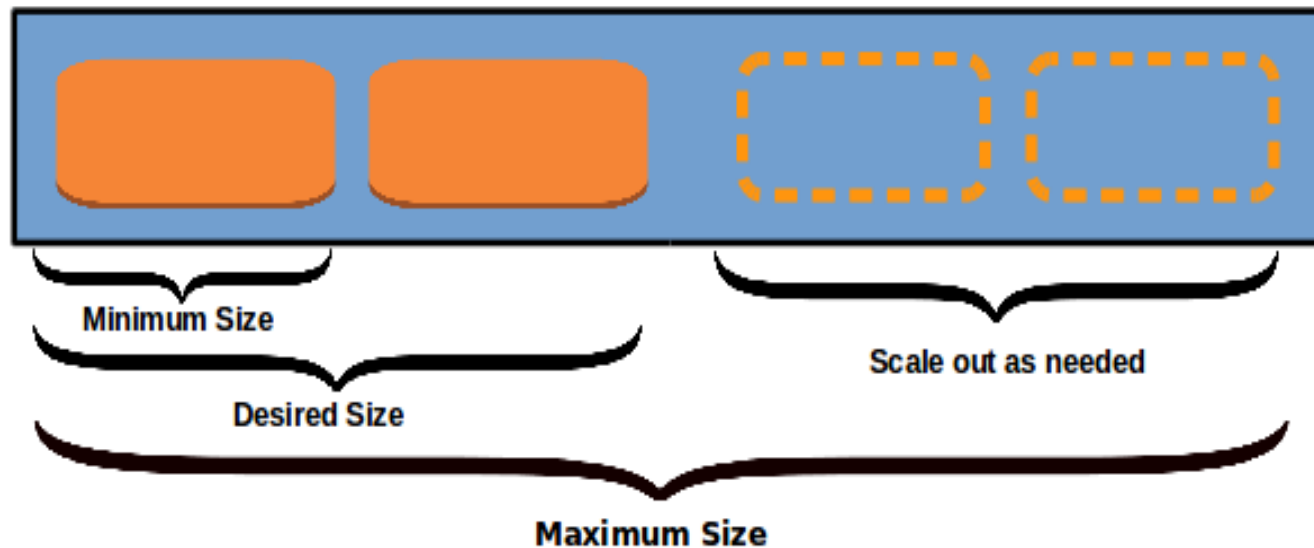
# Cost vs. RAM for AWS Instance



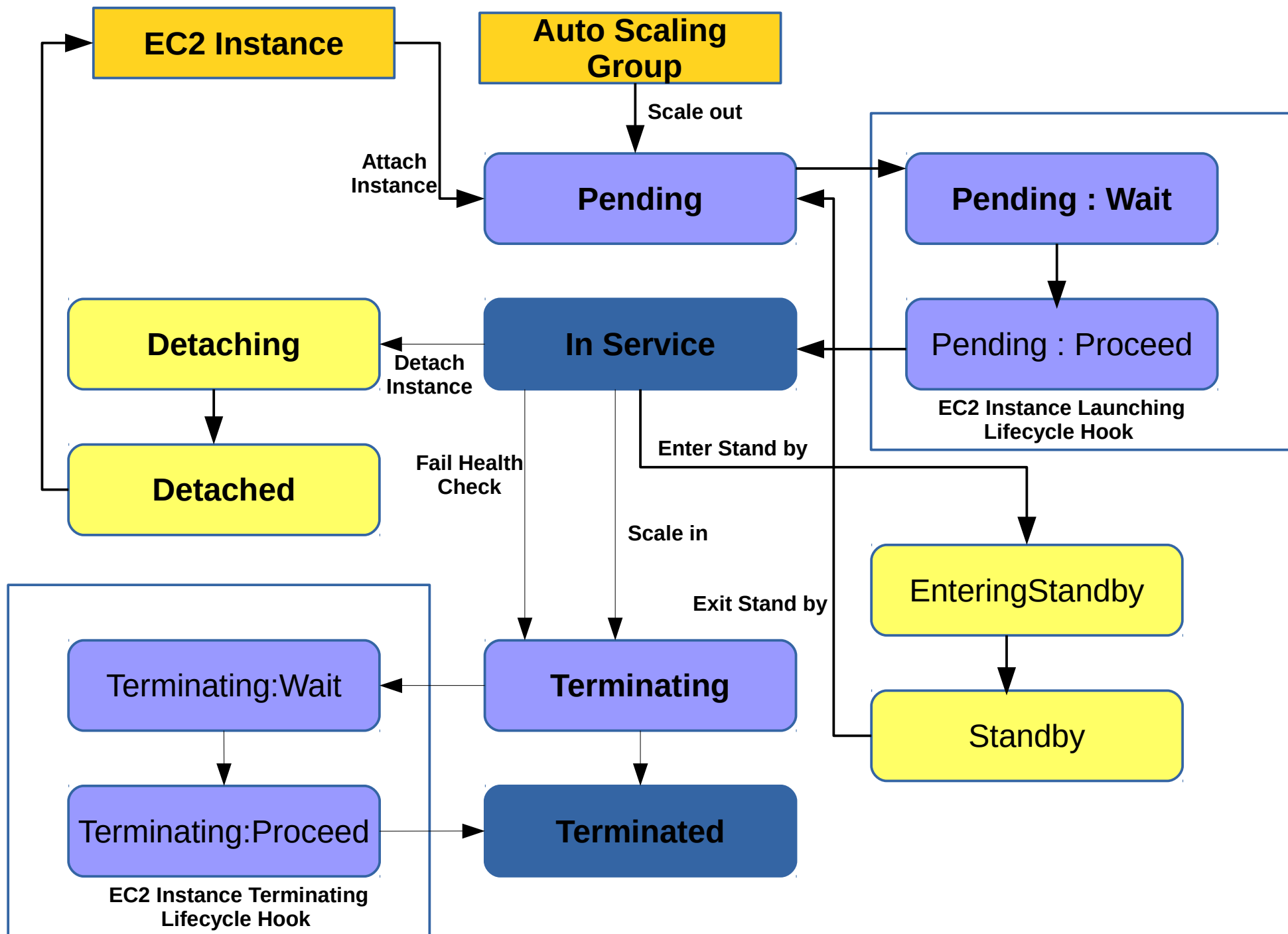
- For Each start and stop price will be calculated
- Pricing calculation of EC2 instance
  - EBS Volume
  - Pricing Option
  - Data Transfer
  - Instance Type
  - AMI Type
  - Region

# Auto Scaling

## Auto Scale Group



- **Auto Scaling** allows **automatically scale of the capacity up** seamlessly during **demand spikes** to maintain performance.
- Auto Scaling **scales down** during **demand spikes** is zero to minimize costs
- Auto Scaling is done in diff AZ bit not diff Regions



# Auto Scaling

---

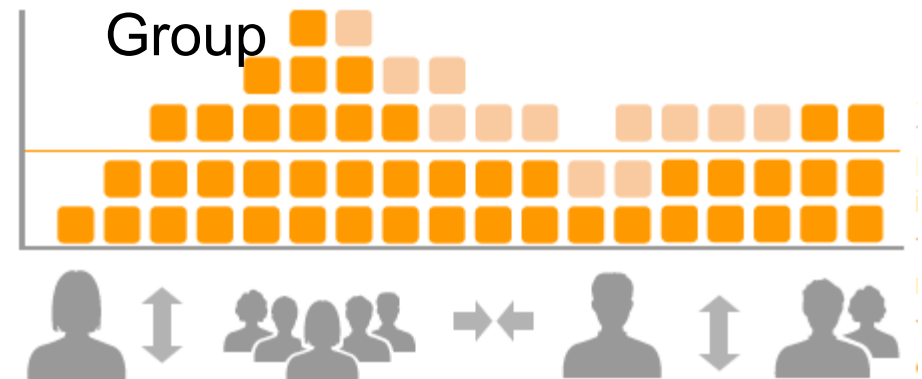
- **Elastic Load Balancing** automatically distributes incoming application traffic across multiple Amazon EC2 instances
- Provide tools to build **failure resilient applications** by launching application instances in **separate AZ's**
- Pay only for resources actually consume, **instance-hours**
- **VM Import/Export** enables you to easily import virtual machine images from your existing environment to Amazon EC2 instances and export them back at any time

# Auto Scaling Functionality

- Create a Launch Configuration



- Create an Auto Scaling Group

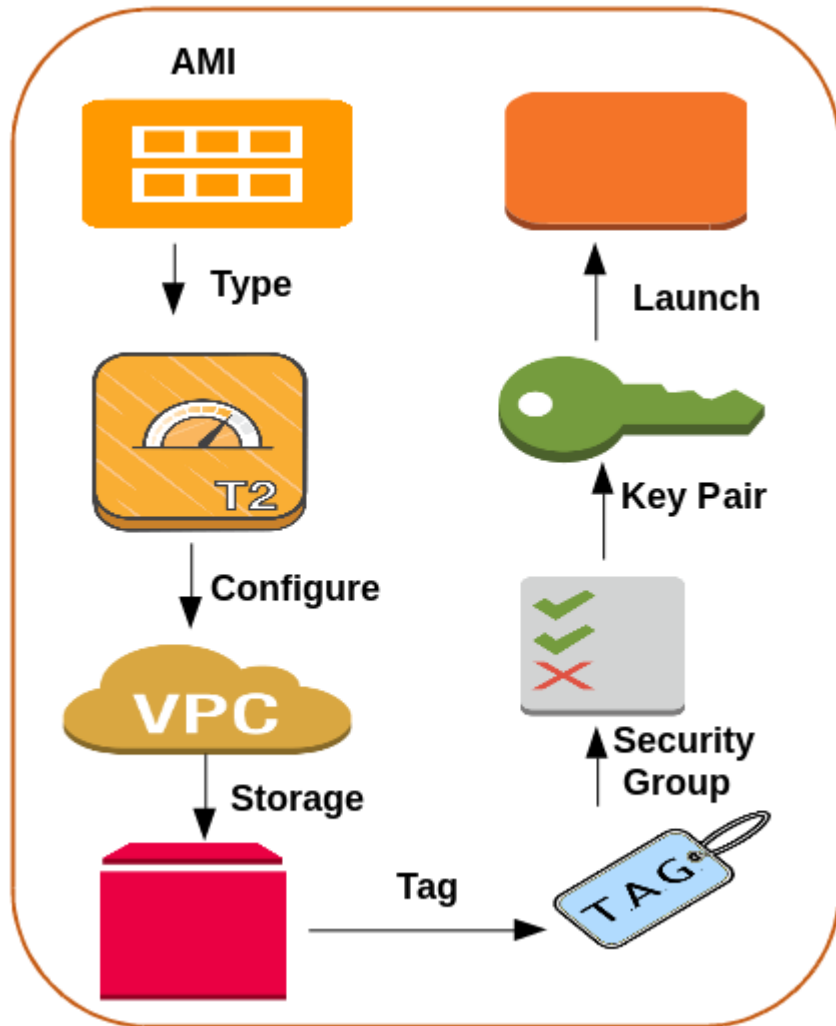


- Verify Your Auto Scaling Group

Details		Activity History	Scaling Policies	Instances	Notifications	Tags	Scheduled Actions
Launch Configuration		my-first-lc					
Load Balancers							
Desired		1					
Min		1					
Max		5					
Health Check Type		EC2					
Health Check Grace Period		300					
Termination Policies		Default					
Creation Time		Tue Jan 26 13:20:17 GMT-800 2016					
Availability Zone(s)		us-west-2a					
Subnet(s)		subnet-cb663da2					
Default Cooldown		300					
Placement Group							
Suspended Processes							
Enabled Metrics							
Instance Protection							



# Auto Scaling Functionality (cont..)

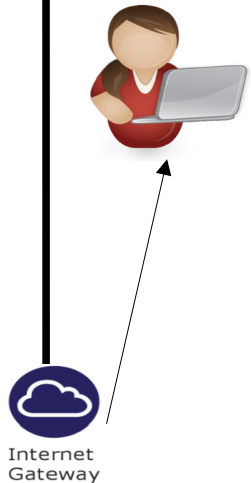


Auto Scale Launch Configuration

1. Configure Auto Scaling group details
  - Group name: --sample--
  - Group size: --3--
  - Network: --vpc123--
  - Subnet: --subnet123--
2. Configure scaling policies
  - Keep this group at its initial size
  - Use scaling policies to adjust the capacity of this group
3. Add Notification
4. Configure Tags
5. Review

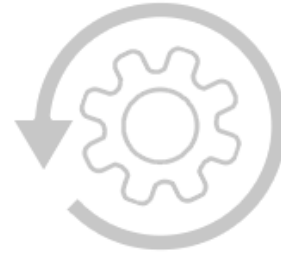
Create Auto Scaling Group

AWS



# Auto Scaling Plans

- Maintain current instance



- Manual scaling

- Scale based on a schedule



- Scale based on demand

# Auto Scaling Limits

S No	Resource	Default Limit
1.	Launch configurations per region	100
2.	Auto Scaling groups per region	20
3.	Scaling policies per Auto Scaling group	50
4.	Scheduled actions per Auto Scaling group	125
5.	Lifecycle hooks per Auto Scaling group	50
6.	SNS topics per Auto Scaling group	10
7.	Classic Load Balancers per Auto Scaling group	50
8.	Target groups per Auto Scaling group	50
9.	Step adjustments per scaling policy	20