# What is Cloud?

- Before cloud came into the picture we had on-prem infrastructure to host our application or website.
- Difficult to auto-scale our application.
- Maintenance.
- Cost may be high in few cases.
- Cloud provides huge space online(collection of servers).
- Buy servers on demand.(Cost effective)

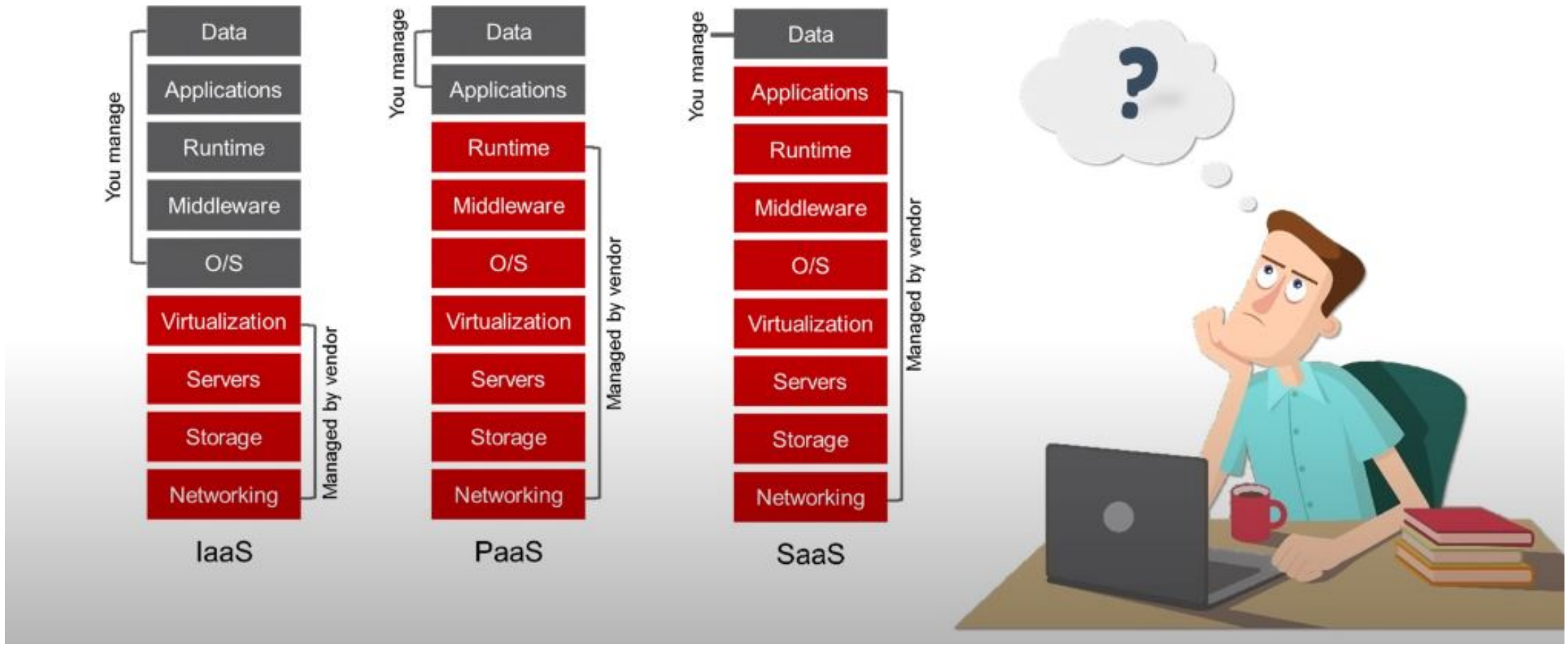
# Cloud Computing

- Store data/Apps on remote servers.
- Process data/Apps on remote servers.
- Access data/Apps on remote servers.

# Cloud service and Types

- Saas – Software as a service
- Paas –Platform as a service
- Iaas –Infrastructure as a service

# Types Of Cloud Services



# AWS Region

- Geographically separated area where AWS has set its resources.
- Each region is independent of each other.
- 24 Available
- 5 announced.
- In India- Mumbai and Hyderabad(coming soon)
- Each region has multiple, isolated and physically separated Az’s.

# Availability Zones

- AZ’s are Datacentres.
- Connected through low latency links.
- In each region at least two AZ’s for fault tolerance.

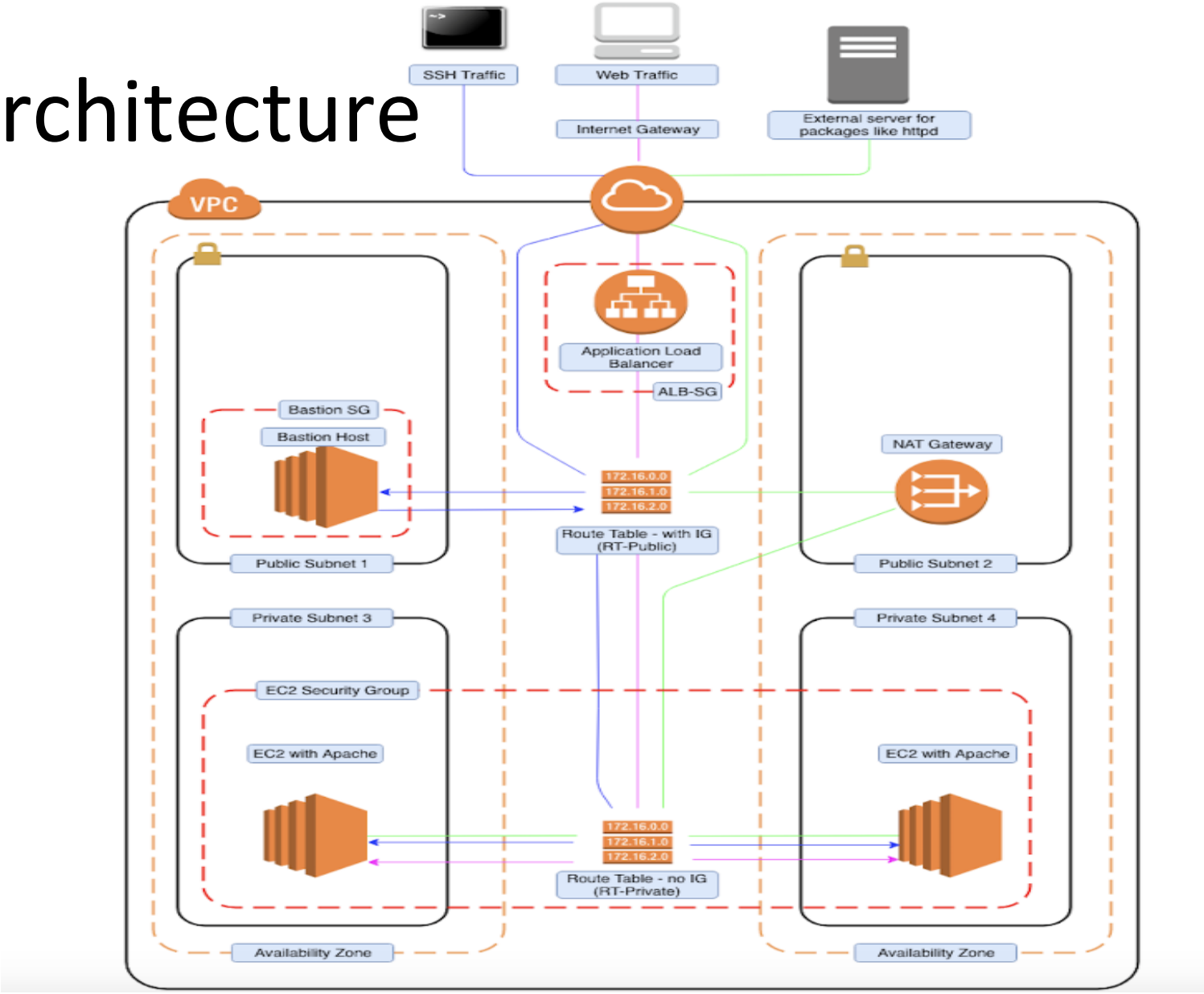
# Edge Locations

- Small Datacentres where end user access services located at AWS.
- Frequently used data will be stored as cache, helps to access faster.
- AWS will not disclose where the AZ’s and edge locations.

# IAM, Roles, Policies

- Identity and Access Management
- Users
- Groups
- Roles
- Policies

# VPC architecture



# VPC limitations(Quotas)

- 5 VPC per region (More available upon request)
- 5 internet gateways per region
- 200 subnets per VPC
- 200 route tables per VPC
- 50 routes/entries per route table
- 2500 security groups per region
- 60 rules per security group

# Reserved IP’s

- AWS reserves the first **four** (4) IP addresses and the last one (1) IP address of every subnet for IP networking purposes.

- 10.0.0.0: Network address.
- 10.0.0.1: Reserved by AWS for the VPC router.
- 10.0.0.2: Reserved by AWS for DNS server
- 10.0.0.3: Reserved by AWS for future use.
- 10.0.0.255: Network broadcast address.

# Cloud Watch

- Mainly used to monitor the resources(services) and applications.
- Basic monitoring-  Collects 5 min data(Metrics)
- Detailed monitoring- 1 min
- Ex- CPU utilization, Memory utilization

# Cloud trail

- Mainly used to track the user activity and API usages.
- We can monitor the accounts activity.

# EC2- Elastic cloud compute

- EC2 instance types

1. General Purpose –T2, M5 and M4
2. Compute Optimized- C4,C5
3. Memory Optimized- X1,R4,R5
4. Accelerated Computing- P2, P3, G3,G4,F1
5. Storage Optimized-H1,I3,D2

# EC2 purchasing types

1. On demand-Short workload
2. Reserved instances- long workload, for a specific period
3. Spot instance- Bid a price and get the instance as long as its price is below that bid price
4. Dedicated instance- Dedicated physical server that we have full control over it.

# EBS- Elastic block storage

- EBS volume can be attached to one EC2 instance at a time.
- By default, EBS volumes are replicated within the availability zones.
- Types

1. General purpose SSD
2. Provisioned IOPS SSD
3. Throughput optimized HDD and cold HDD
4. EBS magnetic

# EFS-Elastic file storage

- EFS is a Cloud based file storage for applications that running AWS

- Types

1. Standard storage class- To store frequently accessed files
2. In frequent access storage class- To store infrequently accessed files

# AMI- Amazon machine image

- AMI is a template that contains software configuration of an instance.
- Required to launch instances.
- We can launch multiple instances using single AMI.
- Backup of AWS instance

# Snapshot

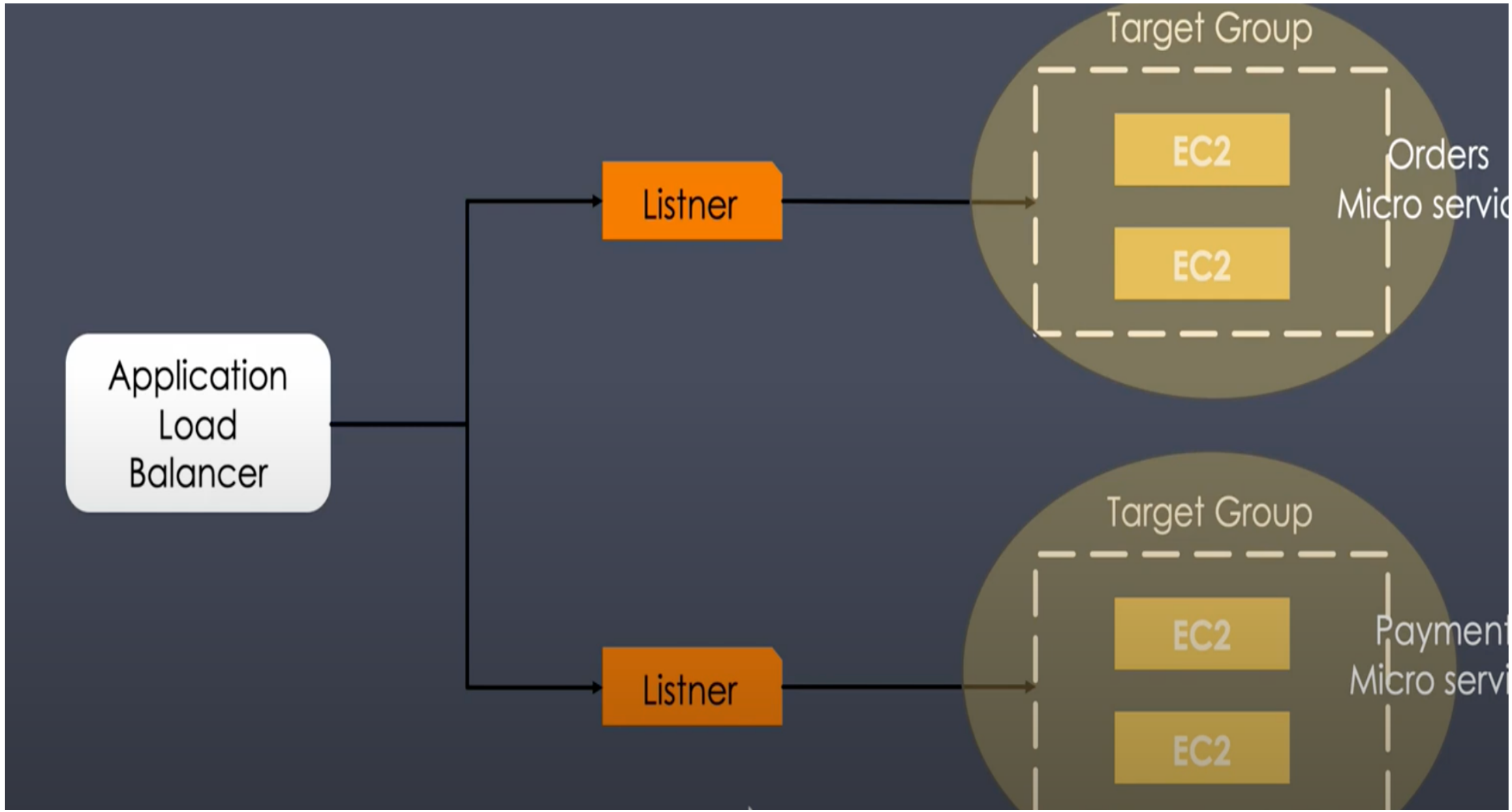
- Backup for AWS volumes.
- We can attach snapshot volume by resizing the storage according to our requirement.

# Load Balancer

1. Application load balancer
2. Network load balancer
3. Classic load balancer

# Application load balancer

- Best suited for load balancing of HTTP and HTTPS traffic.
- Provides advanced request routing targeted at the delivery of modern application architectures, including microservices and containers.


Target Group
EC2
EC2
Orders
Micro servic
Listner
Application
Load
Balancer
Target Group
EC2
EC2
Payment
Micro servi
Listner

# Network load balancer

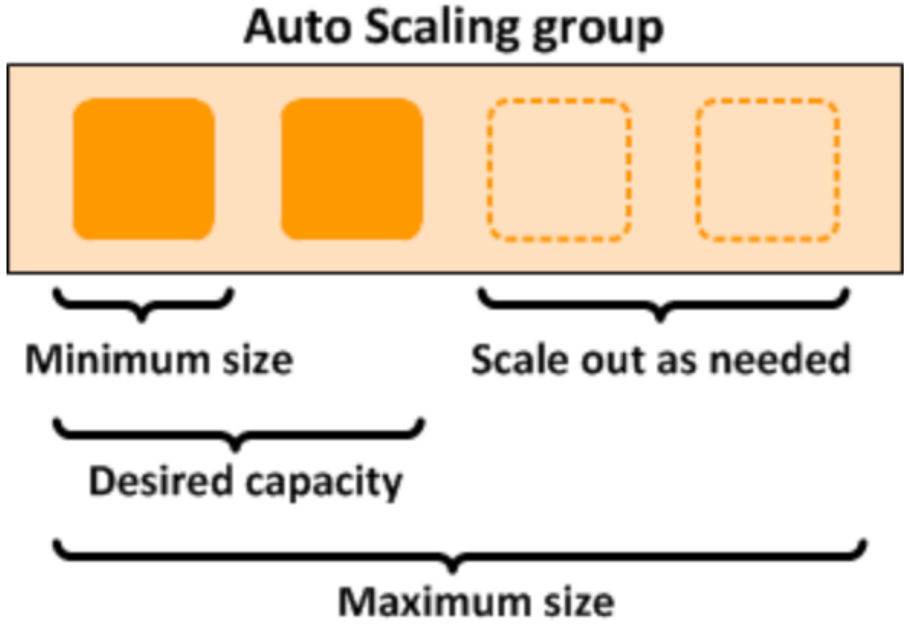
- Best suited for load balancing of Transmission Control Protocol (TCP), User Datagram Protocol (UDP).

# Classic load balancer

- Provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level(Application) and the connection level(Network).

# Autoscaling

- Helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application.
- You create collections of EC2 instances, called Auto scaling groups.



# Auto Scaling components

1. Auto Scaling groups

- A logical group of EC2 instances which participates in Autoscaling.

- Specify its minimum, maximum, and desired number of EC2 instances.

2. Launch configuration

- Template which is used by autoscaling group for adding new instances to the group.

- It contains AMI, security group, EBS volume size, Key pair, IAM role.

# AWS S3

Simple storage service

S3 provides object storage which is built for storing and recovering any amount of data.

Each file can have 5TB in size.

By default, customers can provision up to
100 **buckets** per **AWS** account.

# Benefits of S3

DURABILITY

LOW COST

SCALABILITY

AVAILABILITY

SECURITY

FLEXIBILITY

SIMPLE
DATA
TRANSFER

# Limitations of S3

- Bucket name should be 3 to 63 characters
- Unique across all the AWS accounts.
- Should starts with lowercase or number
- Bucket name can not be an IP address

# Different storage class in S3

- Amazon S3 standard
- Amazon S3 infrequent access
- Amazon Glacier

# Life cycle management

- Moving the data from one storage class to another.

- Transition actions
- Expiration actions

# Features of S3

- Bucket policy
- Data Encryption

# AWS Route 53

- Amazon Route 53 is a highly available and scalable Domain Name System(DNS)  web service.