Elastic Compute Cloud- EC2

EC2- Getting Started

- **E**lastic **C**ompute **C**loud
- Provides **resizable compute** capacity, **Pay** per use
- One of the most used AWS service and core service
- Enables you to build and run applications faster
- Can scale up or down when there is an increase or decrease in website traffic

EC2- Provision Options

- **On demand** – For Short workload, pay fixed rate per hour (per second), no upfront cost, In customer control
- **Reserved Instances** – Reserved capacity in AWS datacenter, duration 1 to 3 years, and flexible payment options, more discount
  - **Standard RI** – Can not change instance attributes like types, OS in between, more discount as compared to Convertible RI
  - **Convertible RI** – Change Instance types, OS etc in between, flexibility in instance types, expensive than standard RI
  - **Scheduled Instances** -Instances that are always available on the specified recurring schedule, for a one-year term e.g. Every Friday 5 PM – 8 PM. Not available in all region
- **Spot** – Bid instances, if application has flexible start and end time, more discount up to 90 % compared to On-Demand, Spot pricing based on offer and capacity availability at AWS
  - https://aws.amazon.com/ec2/spot/pricing/
- **Dedicated hosts**
- **Dedicated Instances**

EC2 – Instance Types

- Comprise varying combinations of CPU, memory, storage, and networking capacity
- Provides flexibility to choose the appropriate mix of resources for your applications

PAGE CONTENT

General Purpose

Compute Optimized

Memory Optimized

Accelerated Computing

Storage Optimized

HPC Optimized

Instance Features

Measuring Instance Performance

Instance Size	vCPU	Memory (GiB)	Instance Storage (GB)	Network Bandwidth (Gbps)***	EBS Bandwidth (Gbps)
c7g.medium	1	2	EBS-Only	Up to 12.5	Up to 10
c7g.large	2	4	EBS-Only	Up to 12.5	Up to 10
c7g.xlarge	4	8	EBS-Only	Up to 12.5	Up to 10
c7g.2xlarge	8	16	EBS-Only	Up to 15	Up to 10
c7g.4xlarge	16	32	EBS-Only	Up to 15	Up to 10
c7g.8xlarge	32	64	EBS-Only	15	10

EC2- AMI

- AMI (Amazon Machine Image)
  - **Software** and **hardware** configuration when instance launched
  - **Sources**
    - New Image from scratch
    - Published by AWS
    - AWS marketplace
    - Generated from existing instance
    - Uploaded Virtual Servers
- **Owned by me** – Specifically owned by your AWS account
- **Public Images** – Shared with all AWS accounts
- **Private images** – Shared with your AWS account and your private ones

EC2- Security Groups

- **Security Groups**
  - **Virtual** firewall
  - Apply on **instance** level
  - **Stateful** – No need to add the rules explicitly in outbound
  - 1 Instance can have many SGs
  - One SG can have many Rules
  - Rules get applied **immediately**
  - Can not add rules to **deny** or **block** specific IP address
  - Its **whitelist** - Anything not listed there will be blocked

EC2 (Security Groups)- LAB

EC2 Lab- Provision Linux VM

- Provision Linux instance with below configuration and Details
  - With additional 5 GB encrypted disk
  - Install OS update patches
  - Create new Security Group which allows SSH and http
  - Install Apache (httpd)
  - Launch simple html web page running on port 80
  - Stop instance
  - Terminate instance

EC2 Lab- Provision Windows VM

- Provision Windows instance with below configuration and Details
  - Run any bootstrap code
  - Add additional Disk
  - Create new Security Group which allows RDP from your IP only
  - Stop instance
  - Terminate instance

EC2 (AMI)- LAB

EC2- Snapshots

- **Snapshots**
  - Kind of backup/recovery
  - Protect your data
  - Data for snapshots stored on S3
  - Need snapshot restore on different region?- then copy the snapshot and create volume
  - Can increase the size of volume during creation
  - Snapshot Life Cycle Policy Or Data Lifecycle Manager

EC2 (Snapshot)- LAB

AWS User Data / Custom Scripts

- Can provide shell script or cloud-directives code
- Option of passing user data to the instance that can be used to perform common **automated configuration tasks**
- Run scripts after the instance **first starts (boots up)**
- Configure instance and install applications **programmatically**
- Also, you can use **AWS CLI** to run the commands
- User data scripts and cloud-init directives run only during the first boot cycle when an instance is launched
- Can configure script to be run at **instance restart**
- On **Windows**, for every restart
  - <persist>true</persist>
- On **Linux**, for every restart
  - #cloud-config cloud_final_modules: - [scripts-user, always]

User Data – Shell Script Example

```
#!/bin/bash
yum update -y
yum install httpd -y
cd /var/www/html
echo "My web server"> index.html
service httpd start
chkconfig httpd on
```

EC2 User Data - LAB

EC2 (Load Balancer)

EC2 Load Balancer – In general

- Virtual Entity or **appliance**
- To **balance** the load or traffic across EC2 instances
- Load balancer serves as the single point of contact for clients
- Distributes incoming application traffic across multiple EC2 instances in multiple Availability Zones
- Enable to achieve **high availability** for application
- Also, scale in and scale out automatically as per demand if you use this with Autoscaling
- Supports HTTP, HTTPS, TCP and SSL protocols
- **Types**
  - **Application**
  - **Network**
  - **Gateway LB (New Launch)**
- https://aws.amazon.com/elasticloadbalancing/features/#compare

Application Load Balancer

- Routing requests to multiple applications on a single EC2 instance
- Can configure specific request goes for specific content to specific instance
- https://docs.aws.amazon.com/elasticloadbalancing/latest/application/introduction.html
- Support multiple listeners with multiple target groups
- Functions at the application layer.
  i.e. 7th layer of OSI model
- Improved performance
- Only support for VPC platform

Application Load Balancer
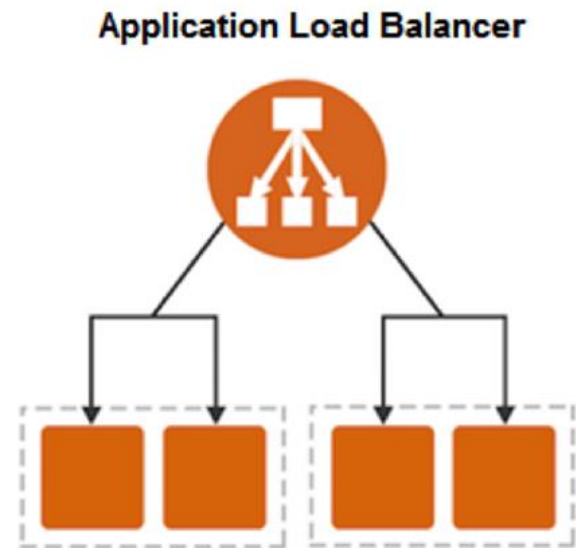
EC2 (Load Balancer)- LAB

Labs

- Provision Win instance with bootstrap code
- Provision Linux instance with bootstrap code
- Elastic IP, Public IP, Private IP Demo
- CLI
  - How to set CLI on your computer
  - How to run CLI directly from AWS instance
  - EC2 operations- start, stop, terminate etc
  - S3- copy from one region to other, download files
  - EC2 Role
- Provision Linux machine. Install Apache. Copy data from S3 and launch webpage without Role.
- Provision Linux with custom scripts with Role

EC2-Autoscaling

- Can also be called as Auto-healing
- Helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application
- Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost
- Specify min and max number of instances in group
- No additional fees with Amazon EC2 Auto Scaling configuration

EC2- Autoscaling
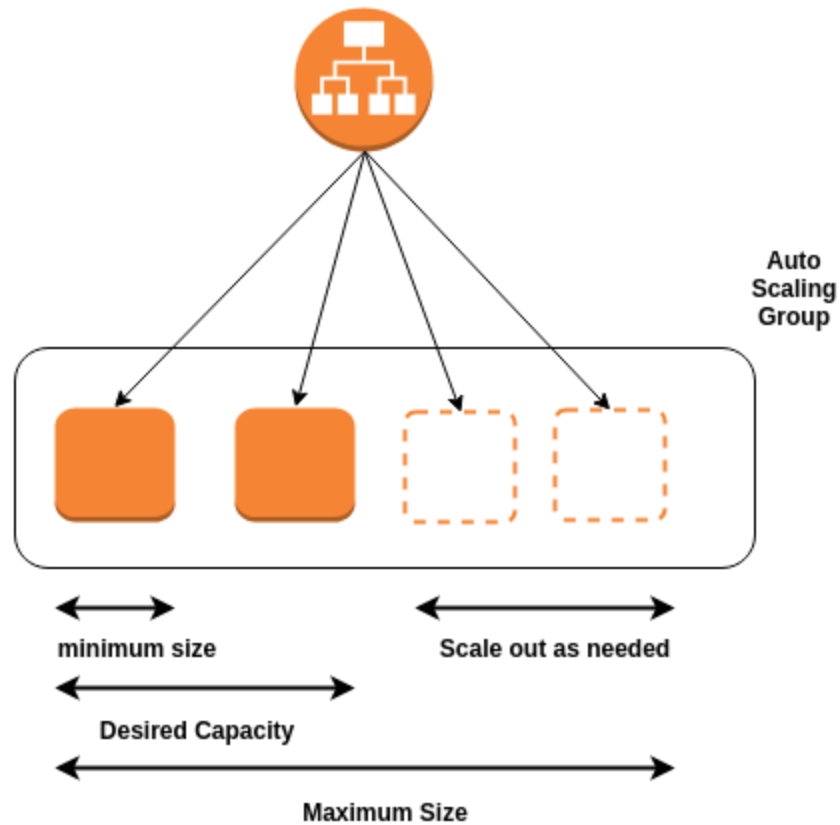
Auto-scaling
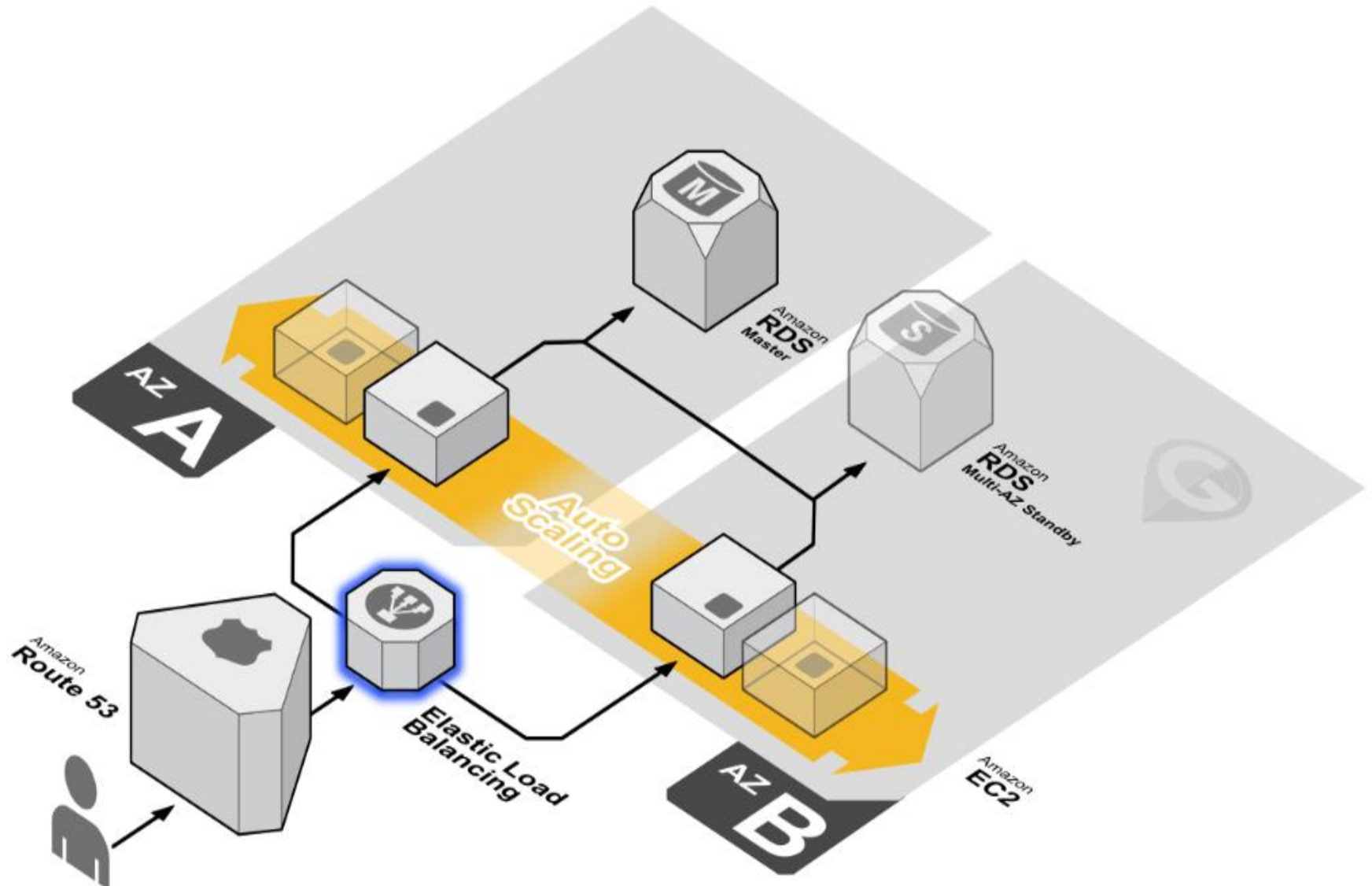
EC2 AutoScaling - LAB

IP – Private/Public/Elastic

- Private
  - Based on vpc/subnet in which EC2 is setup
  - Upon restart, remains same
  - Like your nickname

- Public
  - Allocated from a pool of available public IP's
  - Upon restart, may get change
  - Like your mobile number

- Elastic
  - Static public IP
  - Upon restart, remains same
  - Like your unique Adhaar number

EC2 - Pricing

- Depend on which provisioning model
  - Saving plans, On Demand, RI, Scheduled Instances, Dedicated Instance/Dedicated host, Marketplace, Spot Instance etc
- Pay only for what you use
- Costing Attributes – Running state, storage, data transfer
- Display price an hourly rate – which instance types you are using, hourly or seconds
- Across AZ/region data transfer
- For stop instance – Only volume or storage (EBS) cost, no running and data transfer cost
- AWS Free Tier - 750 hours of Linux and Windows t2.micro instances, (t3.micro for the regions in which t2.micro is unavailable) each month for one year
- https://aws.amazon.com/ec2/pricing/