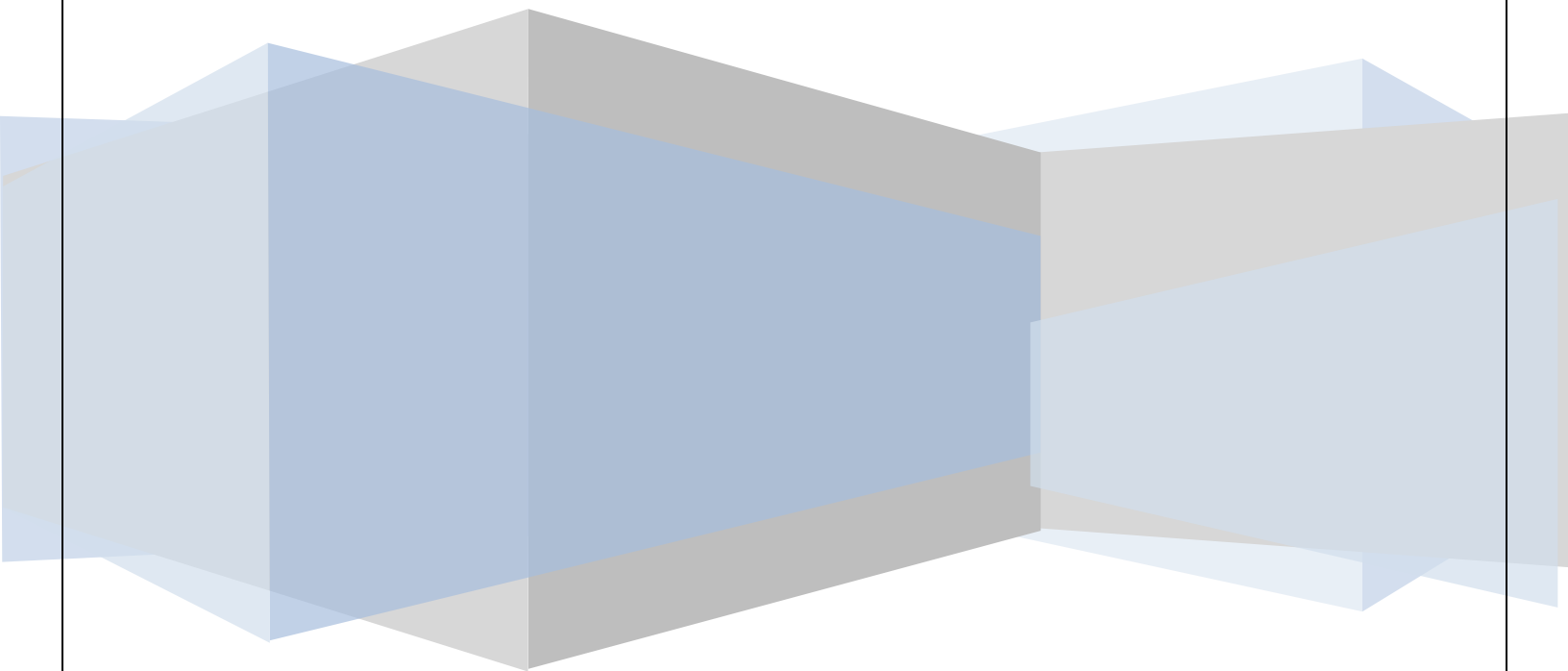AWS Cloud Training
AUTO SCALING

BOOT STRAPPING

When you launch an instance in Amazon EC2, you have the option of passing user data to the instance that can be used to perform common automated configuration tasks and even run scripts after the instance starts.

You can also pass this data into the launch wizard as plain text, as a file (this is useful for launching instances via the command line tools), or as base64-encoded text (for API calls).

Navigate to the **EC2 dashboard** from the AWS Console and select **Instances**, located in the left bar under **INSTANCES.**

Choose **Launch Instance** to create a new instance.
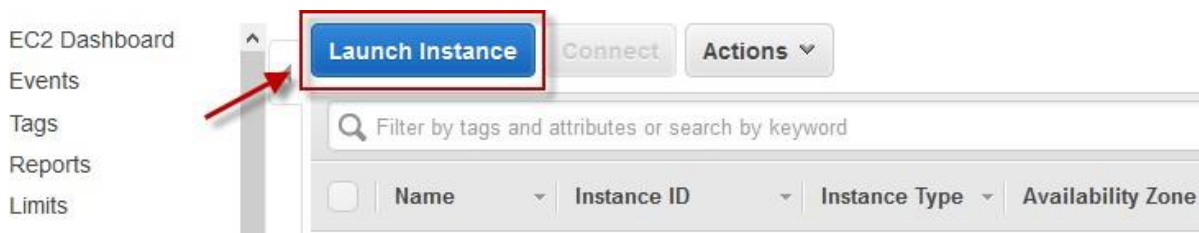
Choose **Amazon Linux AMI** from **choose an Amazon Machine Image** dashboard click on **select.**

Choose instance type from the **Choose an Instance Type** dashboard, then click on Next.

Expand **Advanced Details** section below of the **Configure Instance Details dashboard.**

After expanding Advanced Details section, you will find User data section with a text box.

Enter the below the commands in the text field, then choose Next.

```
#!/bin/bash
yum install httpd php -y
service httpd start
chkconfig httpd on
echo "<?php phpinfo(); ?>" > /var/www/html/index.php
```


Then go with normal process for creating a new instance, make sure you have selected a security group which has 80(HTTP) port open.

Once instance launched, you can browse the IP Address assigned to instance by AWS.
You will be displayed with PHP Info page with all PHP settings.

HORIZONTAL VS. VERTICAL SCALING
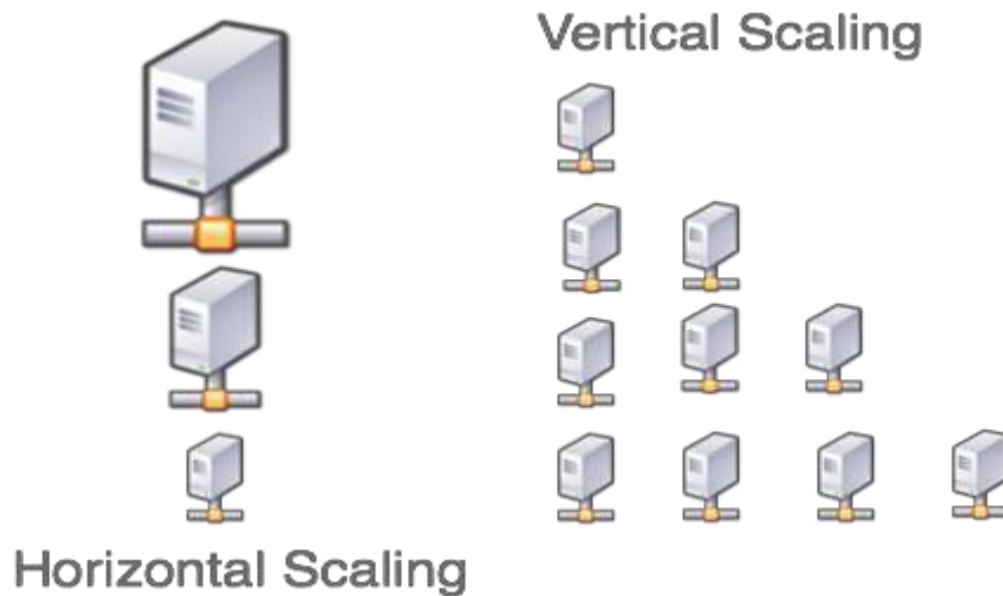
Horizontal scaling means that you scale by adding more machines into your pool of resources.

Vertical scaling means that you scale by adding more power (CPU, RAM) to your existing machine

Navigate to the EC2 dashboard from the AWS Console and select Launch Configurations, located in the left bar under Auto Scaling.
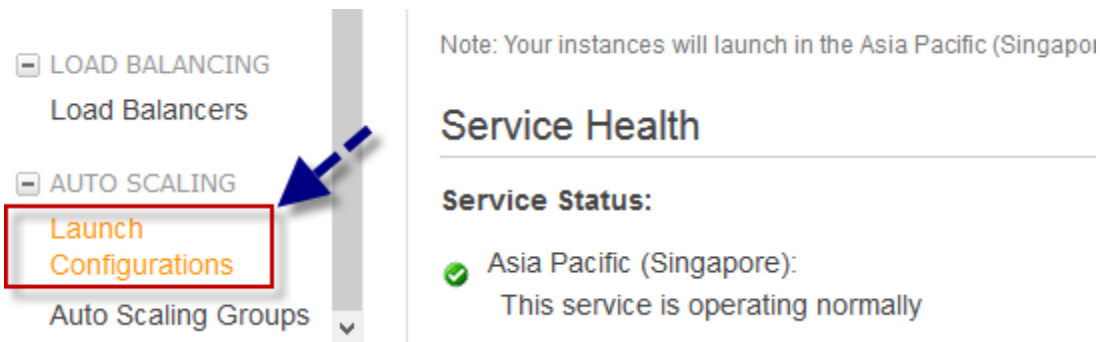
Choose Create Auto Scaling group under Welcome to Auto scaling page.
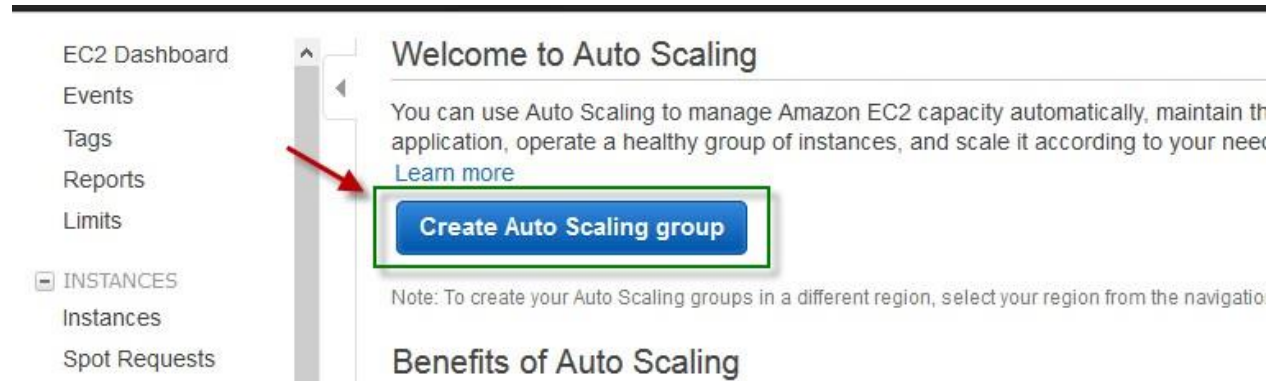
AWS will provide you with a page giving you an overview of Auto Scaling group creation. Click **Create launch configuration**.

Under Create Launch Configuration page, choose My AMIs from left pane, choose your AMI by clicking on Select.

Choose your instance type and click on Next Configure Details.

Specify a name for Launch Configuration, do not check **Request Spot Instances**, and leave the **IAM role** set to **none**. Also, leave **Monitoring** unchecked and choose Next Add Storage.

Create Launch Configuration

Name	(i)	firstlc
Purchasing option	(i)	☐ Request Spot Instances
IAM role	(i)	None
Monitoring	(i)	☐ Enable CloudWatch detailed monitoring Learn more

▸ Advanced Details

💬 Later, if you want to use a different launch configuration, you can create a new one and apply it to any Auto Scaling group. Existing launch configurations cannot be edited.
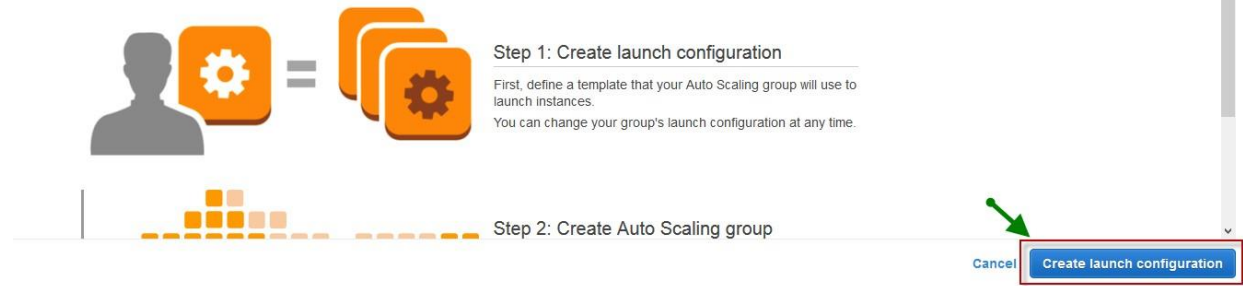
Cancel   Previous   Skip to review   Next: Add Storage

leave everything with the default settings. Go to **Next: Configure Security Group**

Create Launch Configuration

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes.
https://docs.aws.amazon.com/console/ec2/launchinstance/storage about storage options in Amazon EC2.

Volume Type (i)	Device (i)	Snapshot (i)	Size (GiB) (i)	Volume Type (i)	IOPS (i)	Throughput (i)	Delete on Termination (i)	Encrypted (i)
Root	/dev/xvda	snap-4f6201b0	8	General Purpose (SSD)	24 / 3000	N/A	☑	No

Add New Volume

💬 Free tier eligible customers can get up to 30 GB of EBS storage. Learn more about free usage tier eligibility and usage restrictions.

Cancel   Previous   Skip to review   Next: Configure Security Group

Choose existing security group or create a new one by adding required ports and choose **Review.**

Create Launch Configuration

	Security	Name		VPC ID	Description
☐	sg-6895340c	CentOS 6 -x86_64- - with Updates HVM-1602-AutogenByAWSMP-		vpc-adfea0c8	This security group was generated by AWS Marketplace and is based on recommended settings fc
☐	sg-0eb96a6a	default		vpc-adfea0c8	default VPC security group
☐	sg-4d66b929	demo1		vpc-adfea0c8	this is a demo sg
☐	sg-3e70af5a	launch-wizard-1		vpc-adfea0c8	launch-wizard-1 created 2016-04-09T09:09:59.653+05:30
☐	sg-694ce90d	rds-launch-wizard		vpc-adfea0c8	Created from the RDS Management Console
☑	sg-a9b86bcd	test		vpc-adfea0c8	test
☐	sg-2a6bb44e	windows		vpc-adfea0c8	windows

Inbound rules for sg-a9b86bcd Selected security groups: sg-a9b86bcd.

Type ⓘ	Protocol ⓘ	Port Range ⓘ	Source ⓘ
HTTP	TCP	80	10.0.0.10/32
HTTP	TCP	80	0.0.0.0/0

Cancel　Previous　**Review**

AWS will provide you with a page giving you review of all your settings, click **Create launch configuration.**

Create Launch Configuration

t2.micro		Variable		1	1	EBS only	-	Low to Moderate

▼ Launch configuration details　　　　Edit details

Name	firstlc
Purchasing option	On demand
EBS Optimized	No
Monitoring	No
IAM role	None
Tenancy	Shared tenancy (multi-tenant hardware)
Kernel ID	Use default
RAM Disk ID	Use default
User data	
IP Address Type	Only assign a public IP address to instances launched in the default VPC and subnet. (default)

▶ Storage　　　　Edit storage

▶ Security Groups　　　　Edit security groups

Cancel　Previous　**Create launch configuration**

You will be asked for Key pair choose existing or create a new one, then choose **Create launch configuration**.

AWS will provide you with a page to Create Auto Scaling group.
Specify a group name, we will start with 2 instances for high availability.
Select VPC from the VPC drop down list, add subnets under subnet section.
Then expand Advanced Details.

Under Subnet area is an Advanced Details section. Expand this so we can configure load balancing portion of the application. Check Receive traffic from Elastic Load Balancer(s) and in the box below, select the single ELB available. Set the Health Check Type to ELB. You can leave the Health Check Grace Period at the default 300 seconds.
Then choose next configure scaling policies.

We want to **Use scaling policies to adjust the capacity of this group**. You will be presented with two options for actions and alerts: Increasing and Decreasing the group size.
First, we must define the minimum and maximum amount of instances, however. Set it to **Scale between 2 and 4 instances**.
Within the **Increase Group Size** area, press **Add new alarm**.

Create Auto Scaling Group

○ Keep this group at its initial size

● Use scaling policies to adjust the capacity of this group

Scale between [2] and [4] instances. These will be the minimum and maximum size of your group.

Increase Group Size

Name:	Increase Group Size
Execute policy when:	No alarm selected ▾ C Add new alarm
Take the action:	Add ▾ 0 instances ▾
	Add step ⓘ
Instances need:	300 seconds to warm up after each step

Create a simple scaling policy ⓘ

Uncheck the option to send out a notification, and change Whenever to be a Maximum of CPU Utilization [that] is >= 5 Percent. Set for at least to be 1 consecutive period(s) of 1 Minute. Press Create Alarm.

Create Alarm ✕

You can use CloudWatch alarms to be notified automatically whenever metric data reaches a level you define.
To edit an alarm, first choose whom to notify and then define when the notification should be sent.

☐ Send a notification to: demo (trainercollabera@gmail.com) ▾

Whenever: Maximum ▾ of CPU Utilization ▾

Is: >= ▾ 5 Percent

For at least: 1 consecutive period(s) of 1 Minute ▾

Name of alarm: awsec2-firstasg-CPU-Utilization

CPU Utilization Percent

firstasg

Cancel **Create Alarm**

From here, we now need to define the action we want AWS to take when the alarm threshold is hit.
In the Take the action area, we want to Add 1 instance. Set the Instances needed to 300 seconds to warm up after each step.

Increase Group Size

Name: Increase Group Size

Execute policy when: awsec2-firstasg-CPU-Utilization Edit Remove
breaches the alarm threshold: CPUUtilization >= 5 for 60 seconds
for the metric dimensions AutoScalingGroupName = firstasg

Take the action: Add ▾ 1 instances ▾ when 5 <= CPUUtilization < +infinity
Add step ⓘ

Instances need: 300 seconds to warm up after each step

Create a simple scaling policy ⓘ

Under Decrease Group Size, also Add new alarm.
Again, deselect the send notification option.
Set Whenever to a Minimum of CPU Utilization [that] is <= 19 Percent for at least
1 consecutive period of 1 Minute. Create Alarm.

Then set Take the action to Remove 1 instances. Press next configure notifications.

Do nothing on the next page, choose Next: Configure tags.

Create Auto Scaling Group

Configure your Auto Scaling group to send notifications to a specified endpoint, such as an email address, whenever a specified event takes place, including: successful launch of an instance, failed instance launch, instance termination, and failed instance termination.

If you created a new topic, check your email for a confirmation message and click the included link to confirm your subscription. Notifications can only be sent to confirmed addresses.

Add notification

Cancel   Previous   Review   Next: Configure Tags

On the next page, choose Review, as we do not want to add any tags to these instances.

Create Auto Scaling Group

A tag consists of a case sensitive key-value pair that you can use to identify your group. For example, you could define a tag with Key = Environment and Value = Production. You can optionally choose to apply these tags to instances in the group when they launch. Learn more.

Key	Value	Tag New Instances (i)	
		☑	✕

Add tag   9 remaining

Cancel   Previous   Review

AWS will provide you with a page giving you review of all your settings, click **Create Auto Scaling Group.**

Create Auto Scaling Group

Group name	firstasg
Group size	2
Minimum Group Size	2
Maximum Group Size	4
Subnet(s)	subnet-e2595387,subnet-4a585b3d
Load Balancers	demo
Health Check Type	EC2
Health Check Grace Period	300
Detailed Monitoring	No
Instance Protection	None

▼ Scaling Policies   Edit scaling policies

Increase Group Size	With alarm = awsec2-firstasg-CPU-Utilization; Add 1 instances and 300 seconds for instances to warm up
Decrease Group Size	With alarm = awsec2-firstasg-High-CPU-Utilization; Remove 1 instances

▼ Notifications   Edit notifications

▼ Tags   Edit tags

Cancel   Previous   **Create Auto Scaling group**

Then go to Instances page to see the instances which were started creating.

Then go to Load Balancers section, select your load balancer, choose instances tab, you can see instances from two AZ's attached to ELB and status is in service.


Load balancer: demo

| Description | **Instances** | Health Check | Monitoring | Security | Listeners | Tags |

Connection Draining: Enabled, 300 seconds (Edit)

Edit Instances

Instance ID	Name	Availability Zone	Status	Actions
i-13f5c8b7		ap-southeast-1b	InService ⓘ	Remove from Load Balancer
i-790351f7		ap-southeast-1a	InService ⓘ	Remove from Load Balancer

Then select Description tab, you can see the DNS name for ELB, copy and browse your application which is auto scaled and high available.
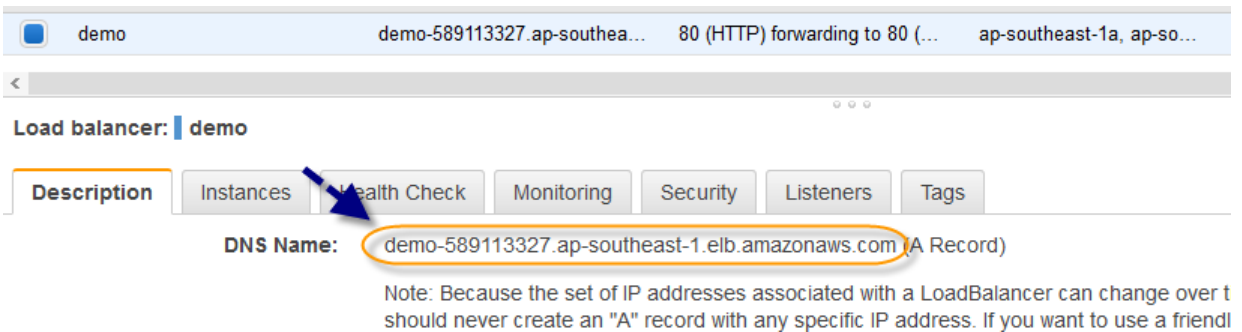
If you want to test self-healing, we can delete one instance which is created by Auto Scaling, Auto Scaling will automatically launch a new instance to meet the minimum requirement which put as 2 instances.