AWS

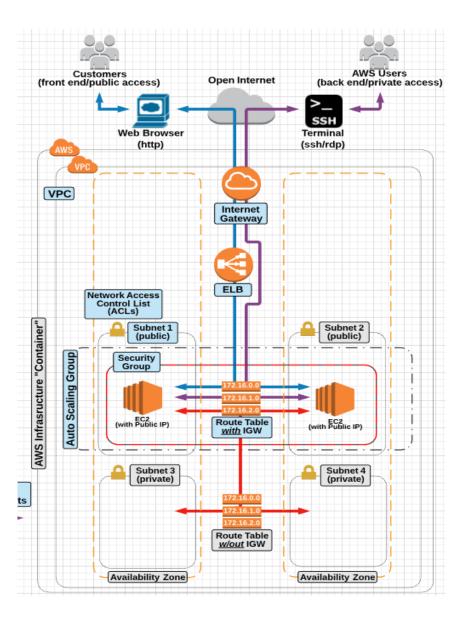
ELB & Auto Scaling Group

By

Keshav Kummari

Elastic Load Balancer

- A load balancer can distribute incoming traffic across your EC2 instances.
- This enables you to increase the availability of your application.
- The load balancer also monitors the health of its registered instances and ensures that it routes traffic only to healthy instances.
- You configure your load balancer to accept incoming traffic by specifying one or more listeners, which are configured with a protocol and port number for connections from clients to the load balancer and a protocol and port number for connections from the load balancer to the instances.
- Elastic Load Balancing supports three types of load balancers: Application Load Balancers, Network Load Balancers, and Classic Load Balancers.
- All Elastic Load Balancing operations are idempotent, which means that they
 complete at most one time.
- If you repeat an operation, it succeeds with a 200 OK response code.



OSI (Open Systems Interconnection) 7 Layer Model

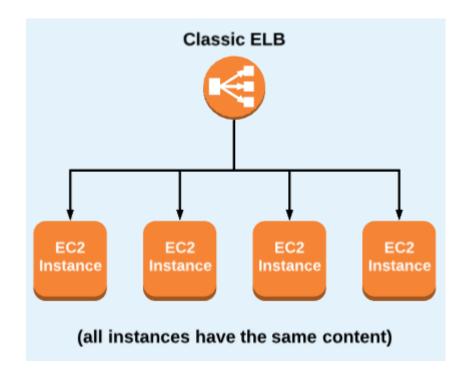
#	Layer	Application	Description	Example
7	Application	Data	Network process to application	DNS, FTP, HTTP, SMTP, Telnet, DHCP
6	Presentation	Data	Data representation and encryption	GIF, JPEG, SSL, MIME
5	Session	Data	Interhost communication	NetBIOS, Sockets, Named Pipes, RPC
4	Transport	Segments	End-to-end connections and reliability	TCP, UDP
3	Network	Packets	Path determination and logical addressing	IP, IPSec, ICMP, BGP
2	Datalinks	Frames	Physical addressing	Ethernet, Wifi, WLAN, MAC
1	Physical	Bits	Media, signal, and binary transmission	CAT5

Elastic Load Balancer(ELB)

- Load Balancing is a common method used for distributing incoming traffic among servers.
- An Elastic Load Balancer is an EC2 service that automates the process of distributing incoming traffic to all the instances that are associated with the ELB.
- Cross-Zone load balancing :
 - An elastic load balancer can load balance traffic to instances located across multiple availability zones.
 - This allows for highly availability and fault tolerant architecture.
- Elastic load balancing can be paired with **Auto Scaling** to enhance high availability and fault tolerance, and allow for automated scalability and elasticity.
- An ELB has it's own DNS record set that allows for direct access from the open internet access.
- Important ELB Facts:
 - An ELB can be **public-facing** or used as an **internal** load balancer and load balancer to internal EC2 instances on Private Subnets(As often done with Multi-Tier Applications).
 - ELB's will automatically stop serving traffic to an instance that becomes unhealthy(Via healthy checks)
 - An ELB or ALB can help reduce compute power or an EC2 instance by allowing for an SSL certificate to be applied directly to the elastic load balancer.

Classic Elastic Load Balancer – OSI Layer 4

- A "Classic" Elastic Load Balancer is designed for SIMPLE balancing of traffic to multiple EC2 instances.
- There are no granular routing "rules" all instances get routed to evenly, and no special routing request can be made based on specific content request from the user.
- Protocols : TCP, SSL HTTP, HTTPS



Application Elastic Load Balancer – OSI Layer 7

- An application load balancer is designed for balancing of traffic to one or more instance target grows using Content-based "rules".
- Content-based rules(setup on the listener) can be configured using:
 - Host-based rules: Route traffic based on the host field of the HTTP header
 - Path-based rules: Route traffic based on the URL path of the HTTP header

• This allows you to structure your application as smaller services, and even monitor/auto-scale based on

traffic to specific "target groups".

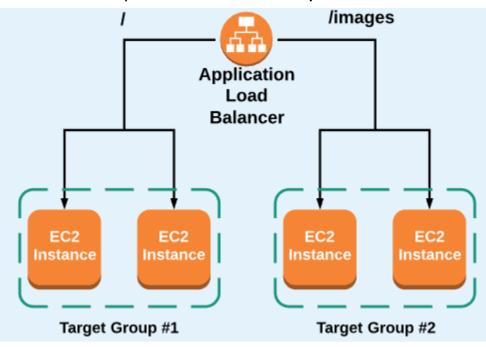
Can balance traffic to multiple ports

An application Load balancer also supports :

ECS and EKS, HTTPS, HTTP/2,

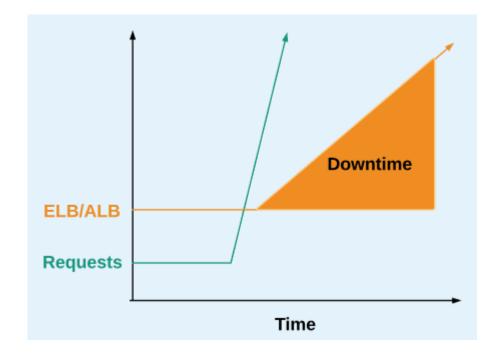
Web Sockets, Access Logs, Sticky Sessions,

and AWS WAF(Web Application Firewall).



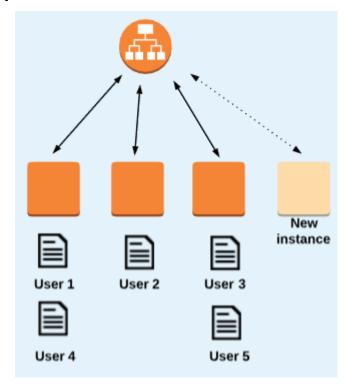
Network Elastic Load Balancer – OSI Layer 3

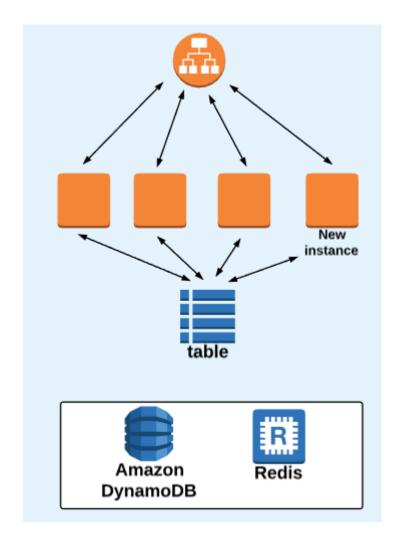
- The Network Load Balancer is designed for extreme performance.
- It does not need to scale to handle large traffic spikes.
- Layer 4 (TCP) load balancing
- Static / Elastic IP address per AZ
- IP Addresses as Targets
- No SSL Offloading



Stateless Architecture

- Store State information Off-Instance
- NoSQL Database
- Shared FileSystem





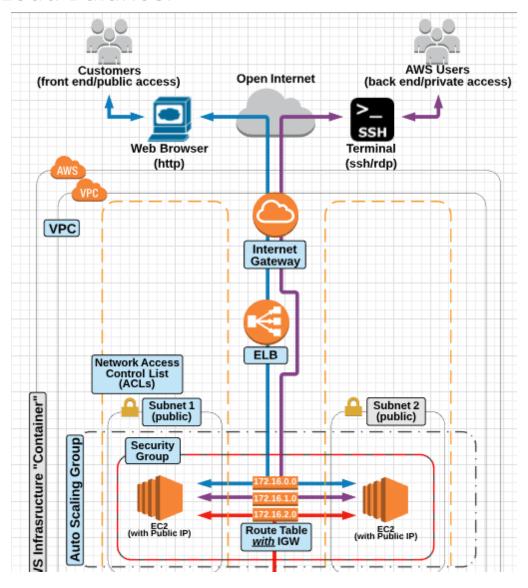
AWS ELB CLI Commands List

https://docs.aws.amazon.com/cli/latest/reference/elb/index.html

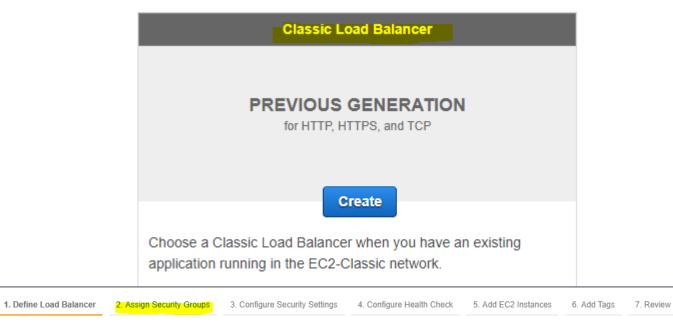
- add-tags
- apply-security-groups-to-load-balancer
- attach-load-balancer-to-subnets
- configure-health-check
- create-app-cookie-stickiness-policy
- create-lb-cookie-stickiness-policy
- create-load-balancer
- create-load-balancer-listeners
- create-load-balancer-policy
- delete-load-balancer
- delete-load-balancer-listeners
- delete-load-balancer-policy
- deregister-instances-from-load-balancer
- describe-account-limits
- describe-instance-health

- describe-load-balancer-attributes
- describe-load-balancer-policies
- describe-load-balancer-policy-types
- describe-load-balancers
- describe-tags
- detach-load-balancer-from-subnets
- disable-availability-zones-for-load-balancer
- enable-availability-zones-for-load-balancer
- modify-load-balancer-attributes
- register-instances-with-load-balancer
- remove-tags
- set-load-balancer-listener-ssl-certificate
- set-load-balancer-policies-for-backend-server
- set-load-balancer-policies-of-listener
- wait

Create Classic Elastic Load Balancer



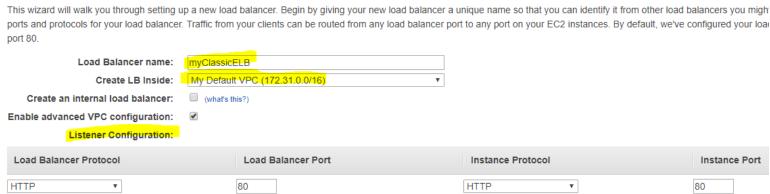
Step-1 Create Classic ELB



Step 1: Define Load Balancer

Basic Configuration

ports and protocols for your load balancer. Traffic from your clients can be routed from any load balancer port to any port on your EC2 instances. By default, we've configured your load



Step-2: Select Subnets

Select Subnets

You will need to select a Subnet for each Availability Zone where you wish traffic to be routed by your load balancer. If you have instances in only one Availat Zones to provide higher availability for your load balancer.

VPC vpc-8b5ba2f0 (172.31.0.0/16)

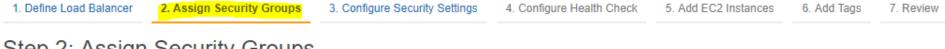
Available subnets

Actions	Availability Zone	Subnet ID	Subnet CIDR
0	us-east-1b	subnet-7647072b	172.31.32.0/20
0	us-east-1d	subnet-e52c60ca	172.31.80.0/20
0	us-east-1e	subnet-88b6f9b7	172.31.48.0/20
0	us-east-1f	subnet-0325ea0c	172.31.64.0/20

Selected subnets

Actions	Availability Zone	Subnet ID	Subnet CIDR
•	us-east-1a	subnet-ea8194a1	172.31.16.0/20
•	us-eäst-1c	subnet-baab9bde	172.31.0.0/20

Step-3: Create Security Groups on Port 80 and assign



Step 2: Assign Security Groups

You have selected the option of having your Elastic Load Balancer inside of a VPC, which allows you to assign security groups to your load balancer. Please select the security groups to assign security groups to your load balancer. Please select the security groups to assign security groups to your load balancer.



Step-4 : Configure Health Check

1. Define Load Balancer

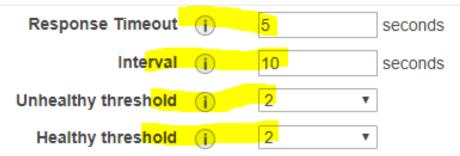
- 2. Assign Security Groups
- Configure Security Settings
- 4. Configure Health Check

Step 4: Configure Health Check

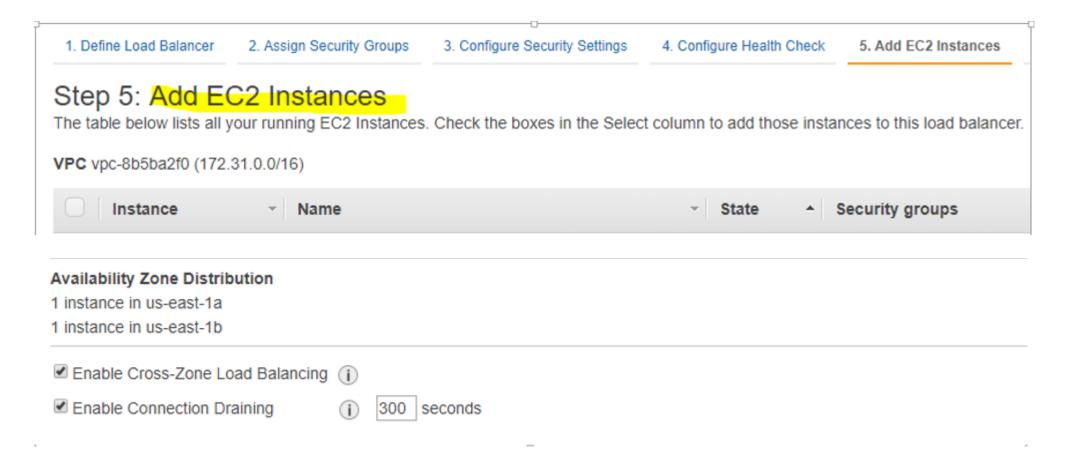
Your load balancer will automatically perform health checks on your EC2 instances and only route traffic to insta balancer. Customize the health check to meet your specific needs.



Advanced Details



Step-5: Add EC2 instances to Elastic Load Balancer using Manual or Auto Scaling



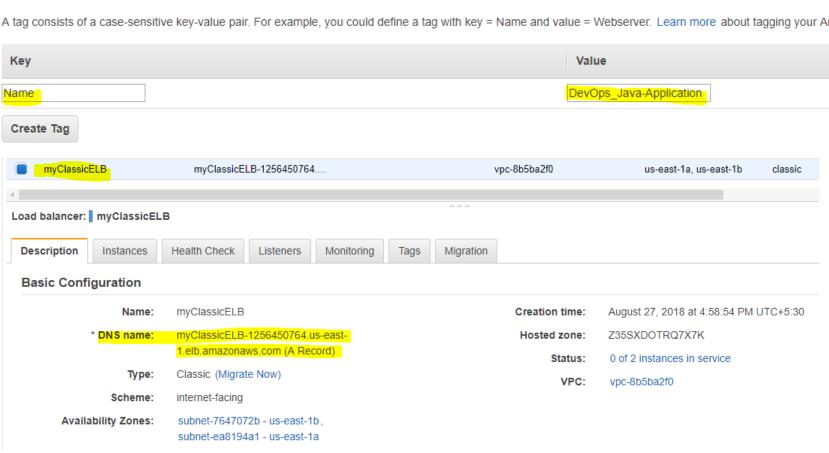
Step-6: Add Tags & Click on Create

1. Define Load Balancer Assign Security Groups Configure Security Settings 4. Configure Health Check Add EC2 Instances 6. Add Tags Review

Step 6: Add Tags

Apply tags to your resources to help organize and identify them.

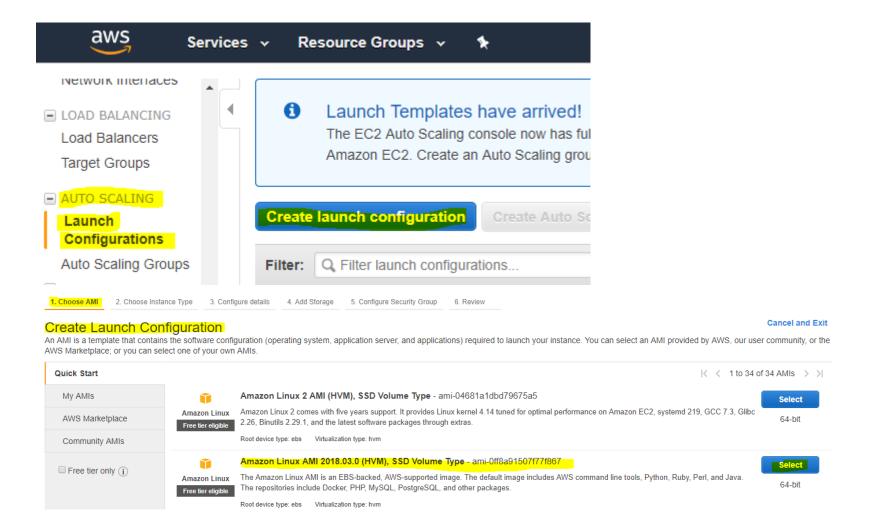
A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. Learn more about tagging your Ai



Cross check the ELB, EC2 instance configuration

- Go to Browser and verify the ELB DNS Name
- http://myclassicelb-1256450764.us-east-1.elb.amazonaws.com/

Auto Scaling Group & Launch Configuration



Select the Instance Type

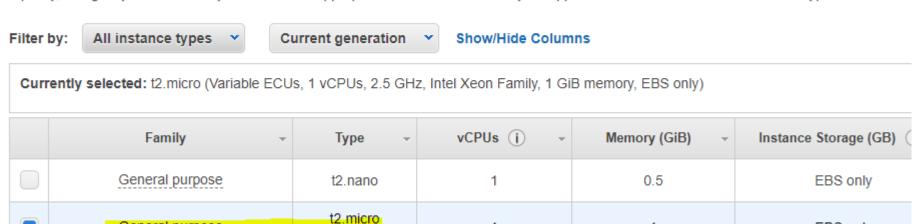
1. Choose AMI 2. Choose Instance Type 3. Configure details 4. Add Storage 5. Configure Security Group 6. Review

Free tier eligible

Create Launch Configuration

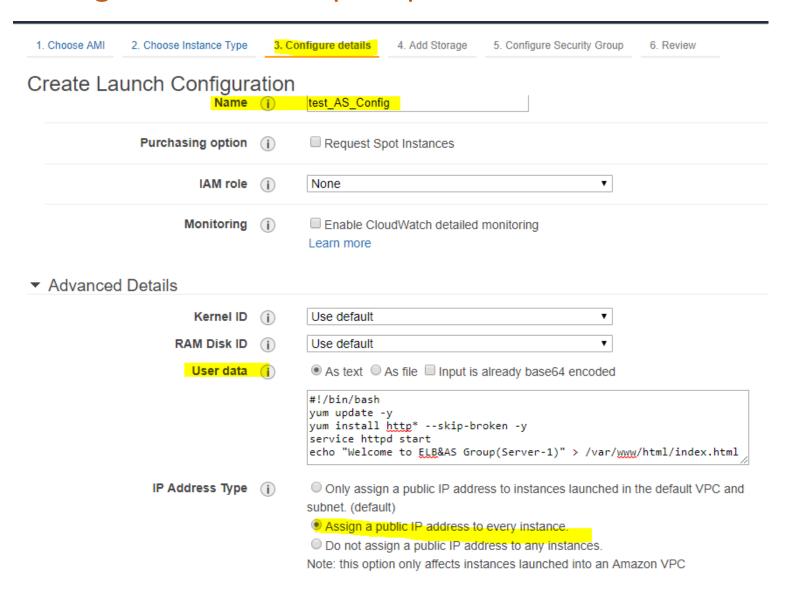
General purpose

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. Learn more about instance types and how the



EBS only

Configure the Bootstrap Script



Add Storage & Create Security Group

Choose AMI

Choose Instance Type

Configure details

4. Add Storage

Configure Security Group

Review

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.



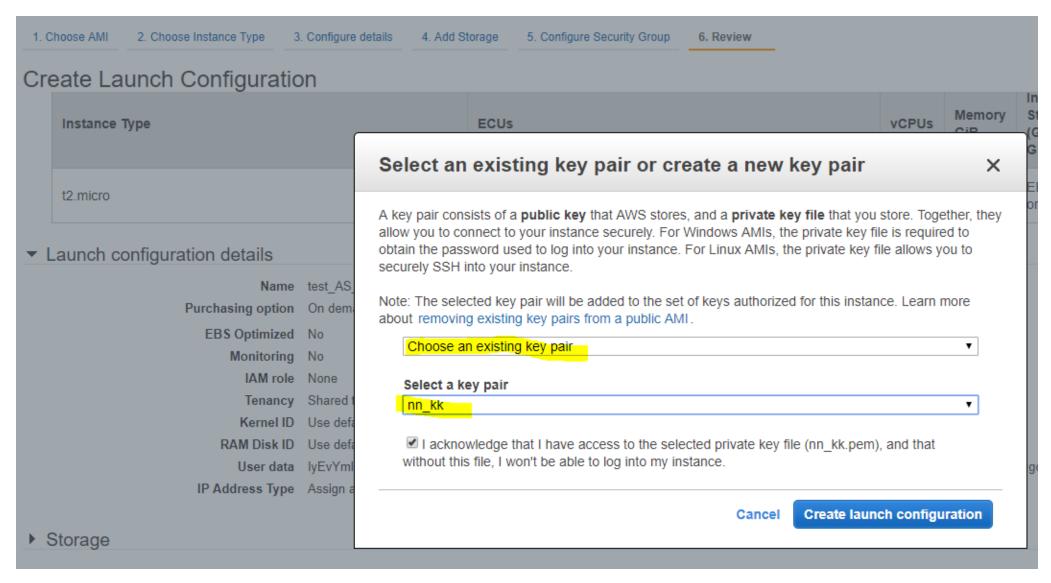
Create Launch Configuration

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. Learn more about

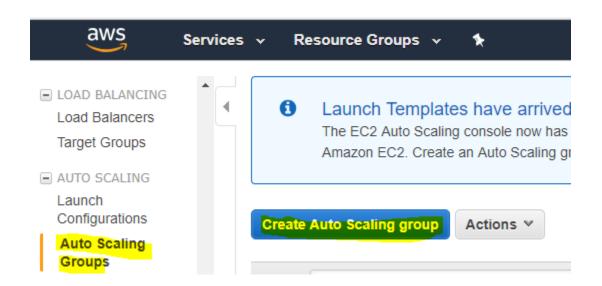


Add Rule

Select the Keypair and Click on Create

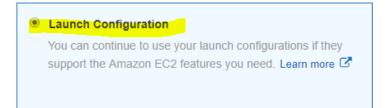


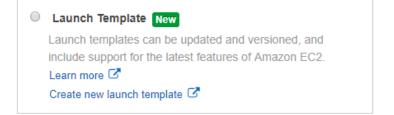
Now, Auto Scaling Group Creation Process



Create Auto Scaling Group

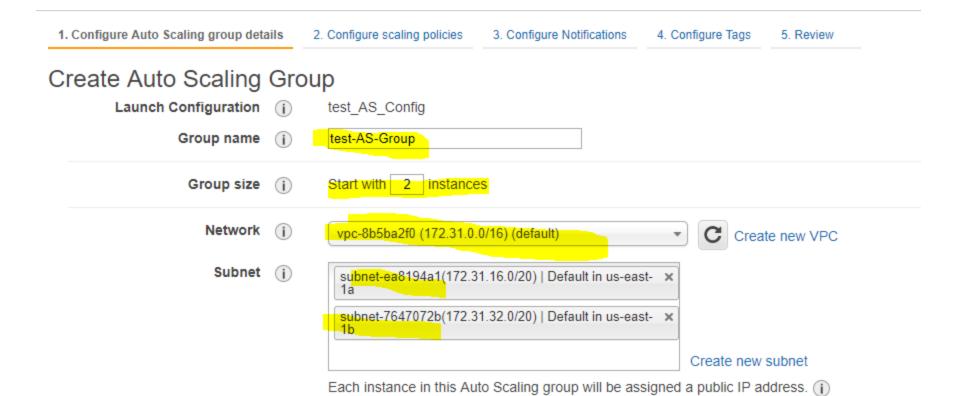
Complete this wizard to create your Auto Scaling group. First, choose either a launch configuration or a launch template to specify the parameters that your Auto Scaling group uses to launch instances.





- Create a new launch configuration
- Use an existing launch configuration





▼ Advanced Details

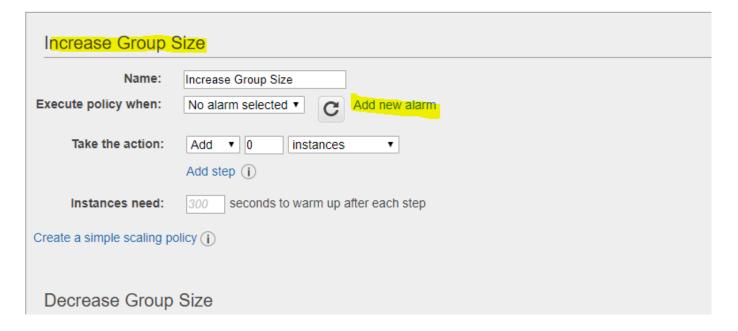
Load Balancing	(i)	Receive traffic from one or more load balancers	Learn about Elastic Load Balancing
Classic Load Balancers	\bigcirc	myClassicELB ×	
Target Groups	(j)		
Health Check Type	(i)	○ ELB <u>• EC2</u>	
Health Check Grace Period	(i)	300 seconds	
Monitoring	Monitoring Amazon EC2 Detailed Monitoring metrics, which are provided at 1 min not enabled for the launch configuration test_AS_Config. Instances la Basic Monitoring metrics, provided at 5 minute frequency. Learn more		Instances launched from it will use
Instance Protection	(i)		
Service-Linked Role	(i)	AWSServiceRoleForAutoScaling	View Role in IAM

Create Auto Scaling Group

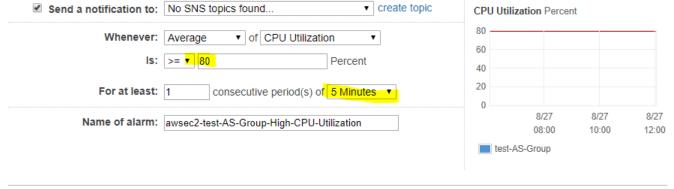
You can optionally add scaling policies if you want to adjust the size (number of instances) of your group automatically. A scaling polic CloudWatch alarm that you assign to it. In each policy, you can choose to add or remove a specific number of instances or a percenta triggers, it will execute the policy and adjust the size of your group accordingly. Learn more about scaling policies.

- Keep this group at its initial size
- Use scaling policies to adjust the capacity of this group

Scale between 2 and 5 instances. These will be the minimum and maximum size of your group.







Cancel

Create Alarm

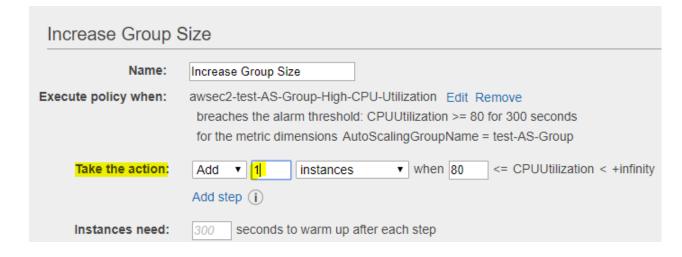
Create Alarm

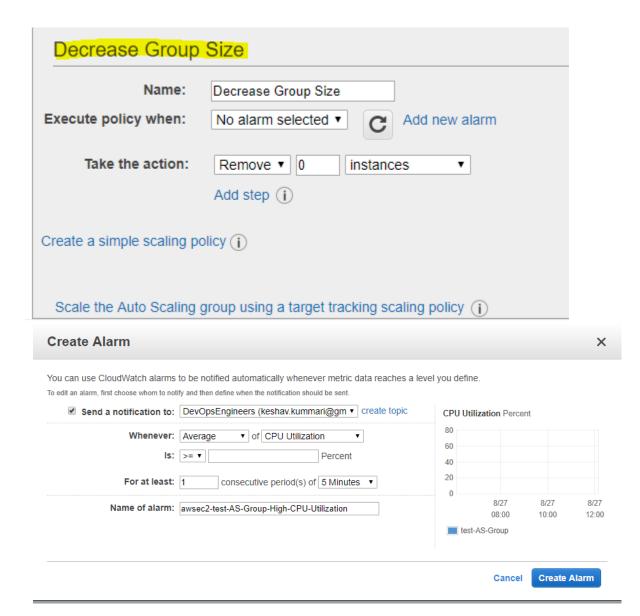
You can use CloudWatch alarms to be notified automatically whenever metric data rearrange $\ensuremath{\text{A}}$

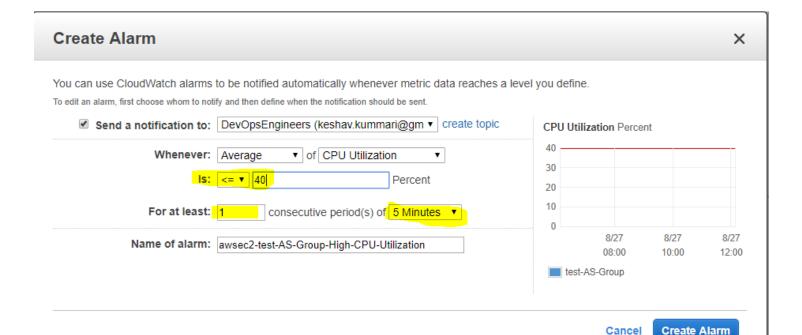
To edit an alarm, first choose whom to notify and then define when the notification should be sent.

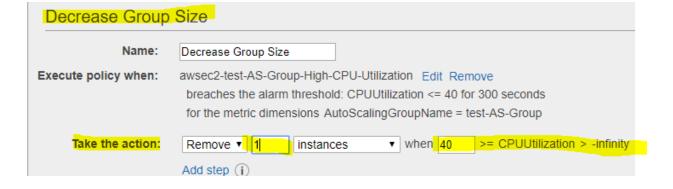
✓ Send a notification to: DevOpsEngineers cancel

With these recipients: keshav.kummari@gmail.com









Add Notification groups if any? Or else skip

1. Configure Auto Scaling group details

2. Configure scaling policies

3. Configure Notifications

Configure Tags

Review

Create Auto Scaling Group

Configure your Auto Scaling group to send notifications to a specified endpoint, such as an email address, whenever a specified 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. N

Add notification

1. Configure Auto Scaling group details

2. Configure scaling policies

3. Configure Notifications

4. Configure Tags

Review

Create Auto Scaling Group

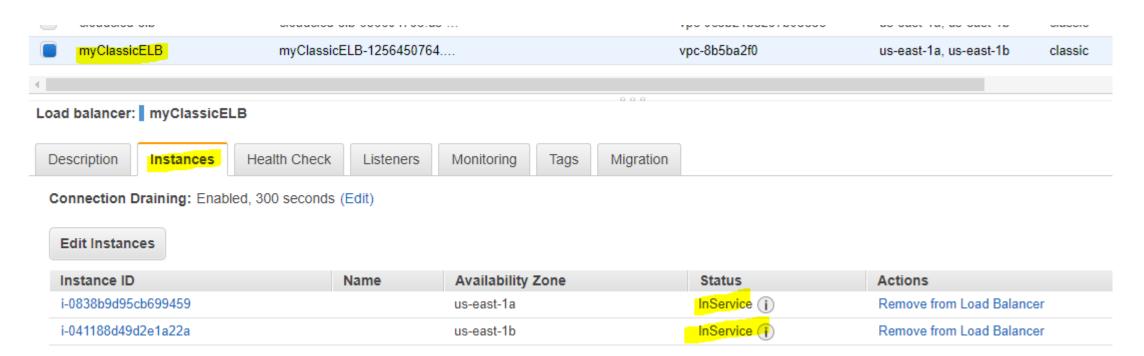
A tag consists of a case sensitive key-value pair that you can use to identify your group. For example, you could define a tag with Key = Environment and Value = Production instances in the group when they launch. Learn more.



Add tag

49 remaining

Verify the ELB & Auto Scaling



Cross check the ELB, EC2 instance configuration

- Go to Browser and verify the ELB DNS Name
- http://myclassicelb-1256450764.us-east-1.elb.amazonaws.com/

Go to Browser and verify the EC2_1 instance DNS Name/IP

http://ec2-54-85-38-140.compute-1.amazonaws.com

http://54.85.38.140

- Go to Browser and verify the EC2_2 instance DNS Name/IP
- http://ec2-54-198-154-106.compute-1.amazonaws.com
- http://54.198.154.106

Auto Scaling Group CLI Commands

• Auto scaling Plans:

https://docs.aws.amazon.com/cli/latest/reference/autoscaling-plans/index.html

- 1. create-scaling-plan
- 2. delete-scaling-plan
- 3. describe-scaling-plan-resources
- 4. describe-scaling-plans
- 5. update-scaling-plan

Application Auto Scaling

https://docs.aws.amazon.com/cli/latest/reference/application-autoscaling/index.html

- 1. delete-scaling-policy
- 2. delete-scheduled-action
- 3. deregister-scalable-target
- 4. describe-scalable-targets
- 5. describe-scaling-activities
- 6. describe-scaling-policies
- 7. describe-scheduled-actions
- 8. put-scaling-policy
- 9. put-scheduled-action
- 10. register-scalable-target

Auto Scaling

https://docs.aws.amazon.com/cli/latest/reference/autoscaling/index.html

- 1. attach-instances
- 2. attach-load-balancer-target-groups
- 3. attach-load-balancers
- 4. batch-delete-scheduled-action
- 5. batch-put-scheduled-update-group-action
- 6. complete-lifecycle-action
- 7. create-auto-scaling-group
- 8. create-launch-configuration
- 9. create-or-update-tags
- 10. delete-auto-scaling-group

- 11. delete-launch-configuration
- 12. delete-lifecycle-hook
- 13. delete-notification-configuration
- 14. delete-policy
- 15. delete-scheduled-action
- 16. delete-tags
- 17. describe-account-limits
- 18. describe-adjustment-types
- 19. describe-auto-scaling-groups
- 20. describe-auto-scaling-instances

- 11. delete-launch-configuration
- 12. delete-lifecycle-hook
- 13. delete-notification-configuration
- 14. delete-policy
- 15. delete-scheduled-action
- 16. delete-tags
- 17. describe-account-limits
- 18. describe-adjustment-types
- 19. describe-auto-scaling-groups
- 20. describe-auto-scaling-instances
- 21. describe-auto-scaling-notification-types
- 22. describe-launch-configurations
- 23. describe-lifecycle-hook-types
- 24. describe-lifecycle-hooks
- 25. describe-load-balancer-target-groups

- 25. describe-load-balancer-target-groups
- 26. describe-load-balancers
- 27. describe-metric-collection-types
- 28. describe-notification-configurations
- 29. describe-policies
- 30. describe-scaling-activities
- 31. describe-scaling-process-types
- 32. describe-scheduled-actions
- 33. describe-tags
- 34. describe-termination-policy-types
- 35. detach-instances
- 36. detach-load-balancer-target-groups
- 37. detach-load-balancers
- 38. disable-metrics-collection
- 39. enable-metrics-collection
- 40. enter-standby

- 41. execute-policy
- 42. exit-standby
- 43. put-lifecycle-hook
- 44. put-notification-configuration
- 45. put-scaling-policy
- 46. put-scheduled-update-group-action
- 47. record-lifecycle-action-heartbeat

- 48. resume-processes
- 49. set-desired-capacity
- 50. set-instance-health
- 51. set-instance-protection
- 52. suspend-processes
- 53. terminate-instance-in-auto-scaling-group
- 54. update-auto-scaling-group

Thank you!