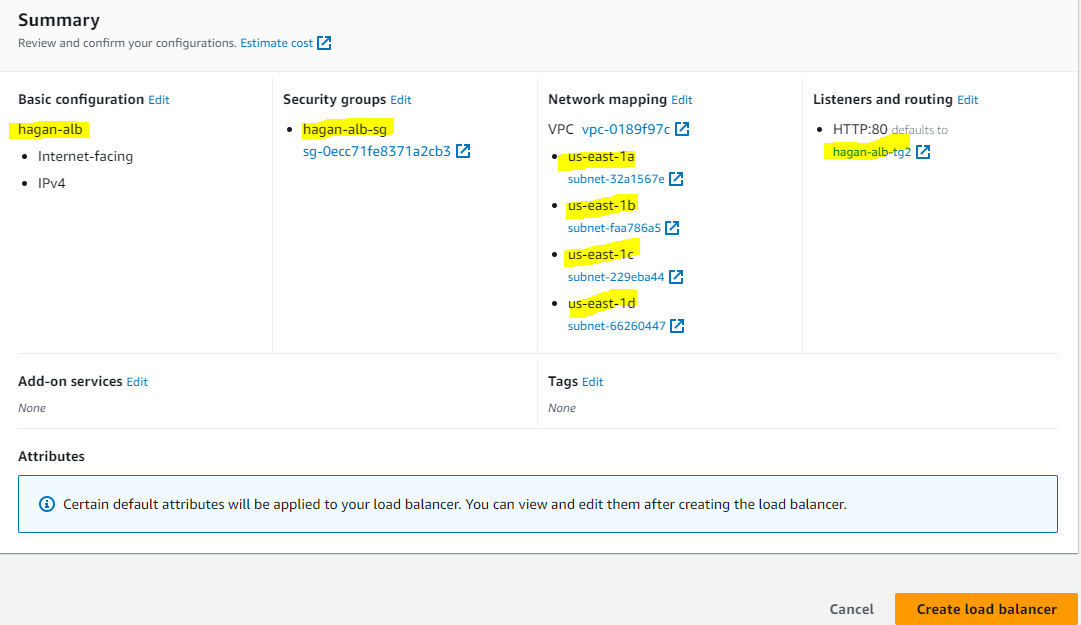
**AWS – ALB with Auto Scaling**

**Create an Application Load Balancer**

1. Navigate to **EC2** > **Load Balancers**.
2. Click **Create Load Balancer**.
3. Click the **Create** button under the **Application Load Balancer** and set the following values:
   * *Name*: **yourlastname-alb**
   * *Scheme*: **internet-facing**
   * *IP address type*: **ipv4**
4. Under **Network Mapping**:
   * Leave Default VPC.
   * Add 4 Availability Zones, like **us-east-1a**, **us-east-1b**, **us-east-1c, and us-east-1d** to your ALB.
5. Under **Security groups**, select to **Create a new security group** (new tab opens) for your ALB, and set the following values:
   * *Name*: **yourlastname-alb-sg**
   * *Description*: **Security group for Application Load Balancer**
   * Click **Add rule**, Type *HTTP* and **Source** *Anywhere-IPv4*
   * Click **Create security group**
   * Go to your browser tabs and find your Load Balancer (should be previous tab)
   * At Security groups, in the dropdown, select the security group you just created. If you don’t see it, click the refresh icon.
   * Click the X to delete the default sg
6. Under **Listeners and routing** click **Create target group** (new tab opens) and enter the following values:
   * *Choose a Target type*: **Instance**
   * *Target group Name*: **yourlastname-alb-tg**
   * Expand **Advanced health check settings**, and reduce the healthy and unhealthy threshold checks down to **2**.
   * Click***Next***
   * Click***Create target group***
   * Go back to the Load Balancer tab and add the target group. Click the refresh icon if it doesn’t show.
7. Scroll down to the bottom an verify the Summary  
   
8. Click **Create load balancer**.
9. Click **View load balancer**.

Make a note of the DNS name associated with the load balancer and open in a new browser tab. You should see a 503 error since we don’t have any operational EC2 instances associated with the load balancer.

**Create a Launch Template**

Create a launch template that will be used by the Auto Scaling group. The launch template defines what the instances are and how they are created. You can use the previous lab ***Creating an EC2 Auto Scaling Group*** with the following information:

1. Navigate to **EC2** > **Instances** > **Launch Templates**.
2. Create a new template, and call it **yourlastname-alb-lt** for the name and description.
3. Check the ***Provide guidance……*** checkbox.
4. Click **Quick Start** under **Application and OS Images** and pick the **Amazon Linux 2 AMI (64-bit x86)**.
5. Set the instance type as **t2.micro**.
6. Select the key pair you created for your webserver in previous labs.
7. Under **Network Settings** for *Subnet* leave “**Don’t include in launch template**”.
8. Select the **webserver-sg** security group you created in an earlier lab.
9. Storage should automatically be populated with a volume, so leave that as default and don’t add anything to the network section.
10. Add Tags of **Key** of ***Name*** and **Value** of ***alb-launch-tmp***.

Expand *Advanced Details*, and paste the following user data in the box.  
  
#!/bin/bash

yum update -y

yum install httpd -y

service httpd start

chkconfig httpd on

cd /var/www/html

instance\_id=$(curl <http://169.254.169.254/latest/meta-data/instance-id>)

availability\_zone=$(curl <http://169.254.169.254/latest/meta-data/placement/availability-zone>)

echo "<html><h1>Welcome to *yourlastname* AutoScaling site</h1>" >index.html

echo "<h1>page generated by instance <strong>$instance\_id</strong> in availability zone <strong>$availability\_zone</strong></h1> " >>index.html

echo "<head> ">>index.html

echo "<meta charset=\"UTF-8\"> ">>index.html

echo "<meta http-equiv=\"X-UA-Compatible\" content=\"IE=edge\"> ">>index.html

echo "<meta name=\"viewport\" content=\"width=device-width, initial scale=1.0\"> ">>index.html

echo "<h1> Time is <span id=\"time\"> </span></h1> ">>index.html

echo "<script src=\"index.js\"> </script> ">>index.html

echo "</head> ">>index.html

echo "</html>">> index.html

echo " \`use strict\` ">index.js

echo "var datetime = new Date(); ">>index.js

echo "console.log(datetime); ">>index.js

echo "document.getElementById(\"time\").textContent = datetime; ">>index.js

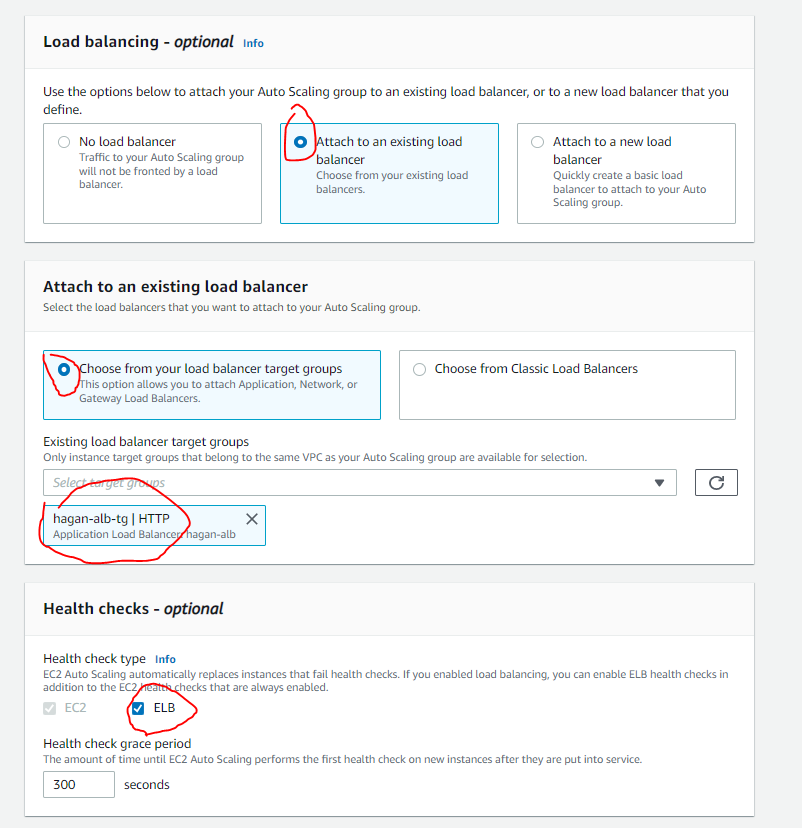
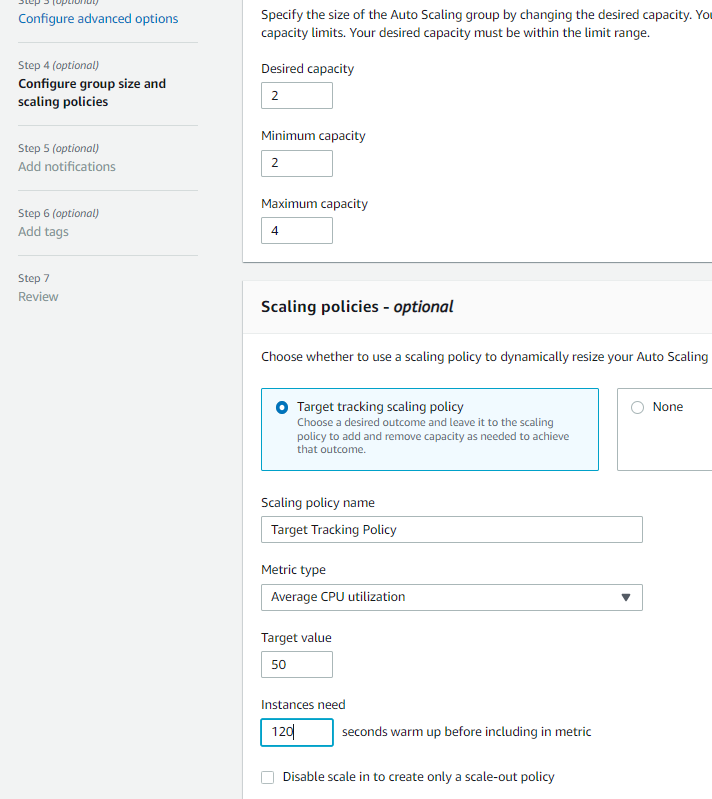
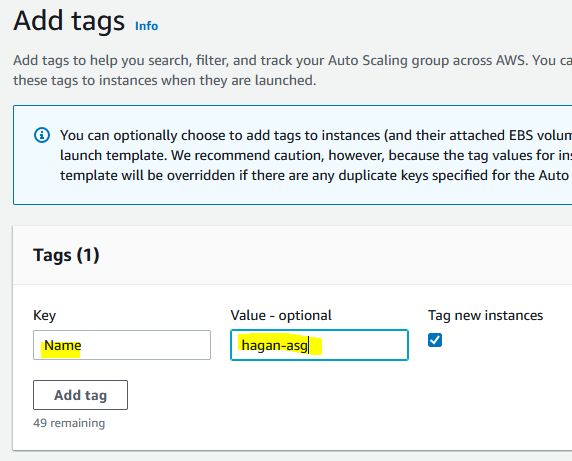
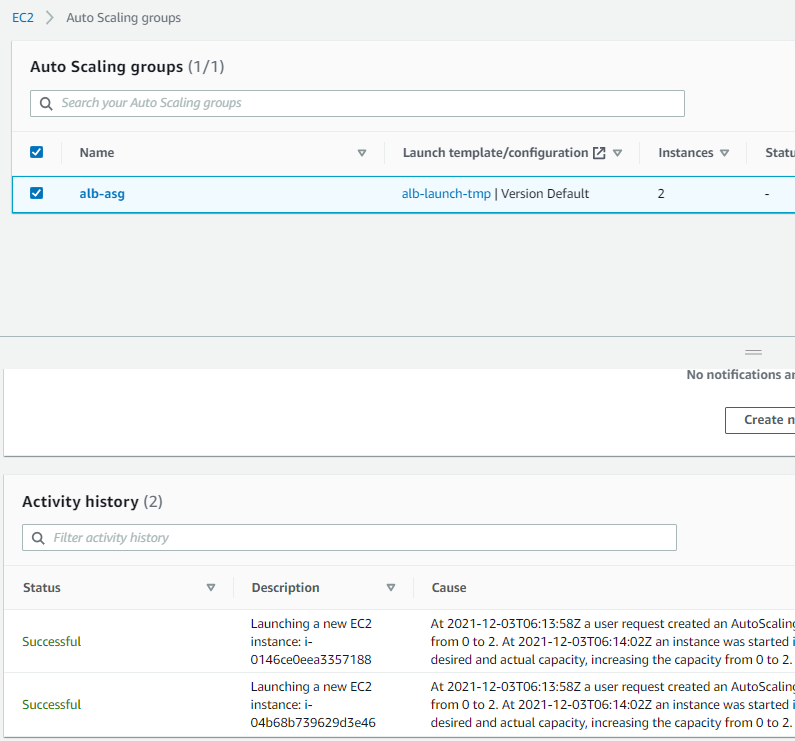
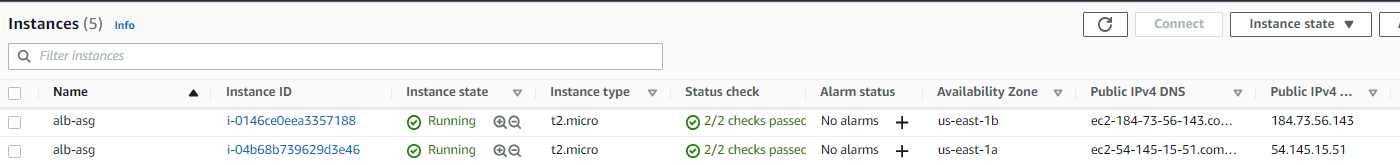
amazon-linux-extras install epel -y

yum install stress -y

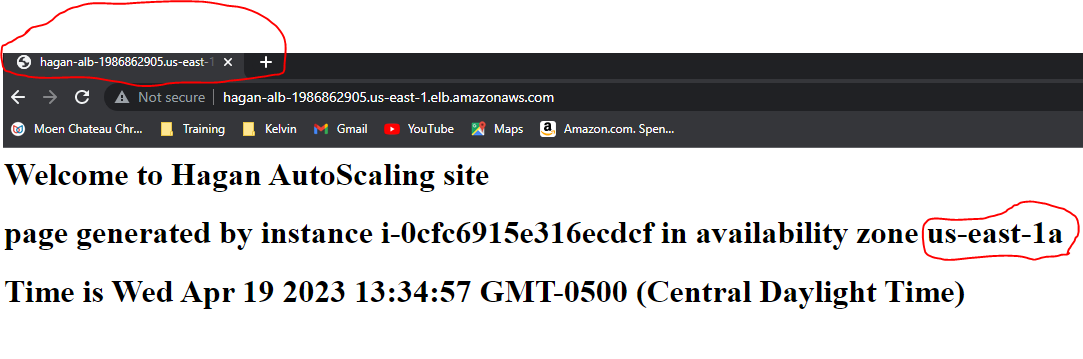
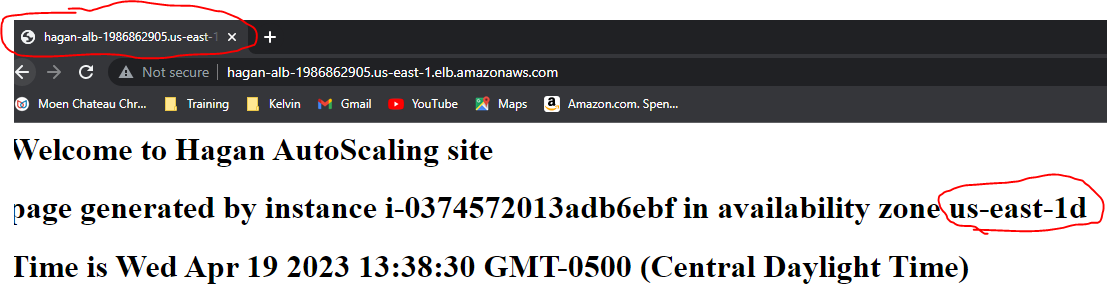
MAKE SURE THIS LINE IS WHERE YOUR CURSOR IS AFTER PASTE. REMOVE THE WORDS

1. Click **Create Launch Template**.
2. Click **View Launch Template**.

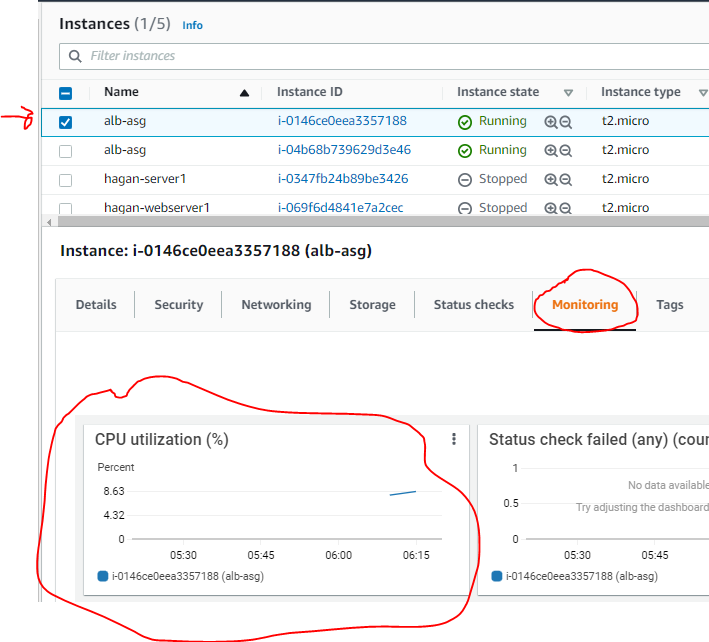
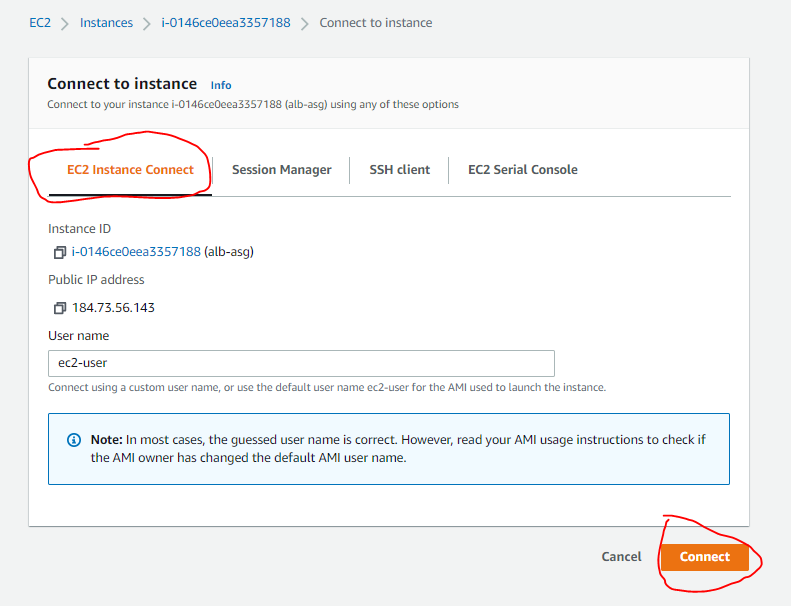
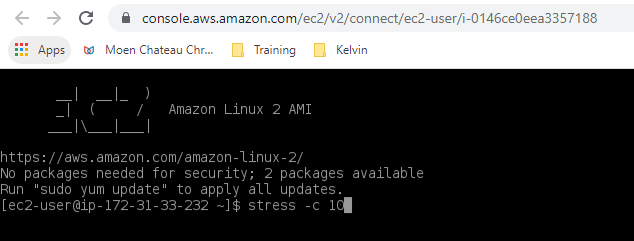
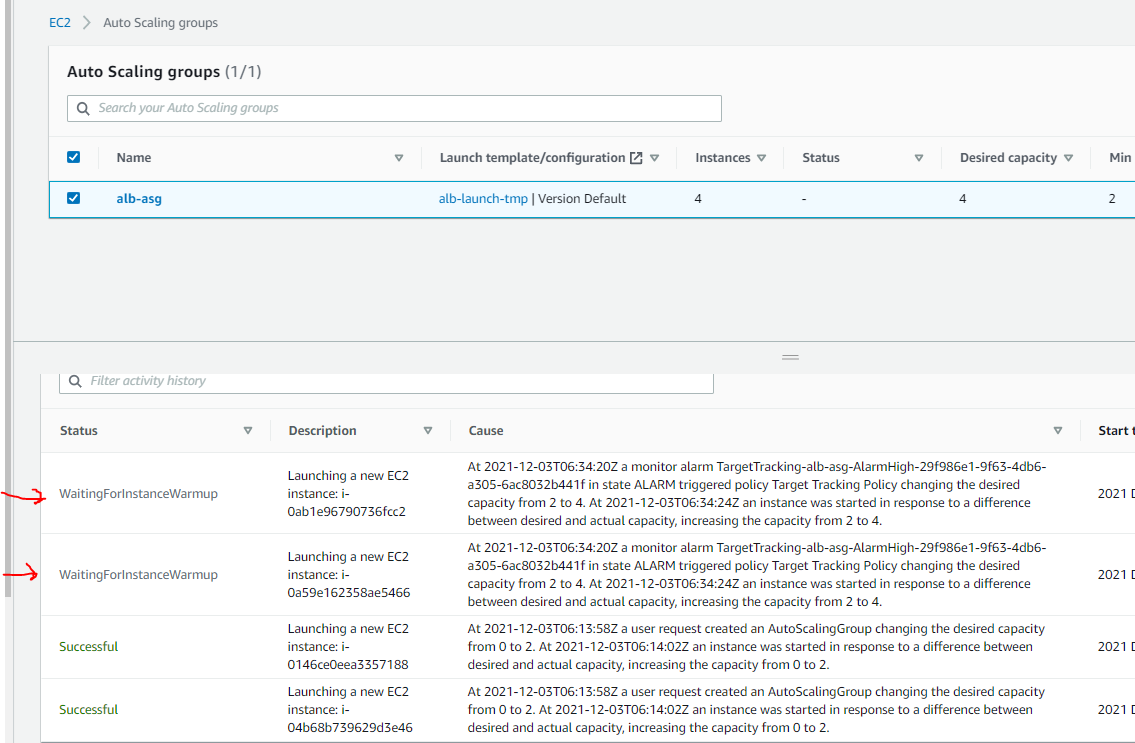
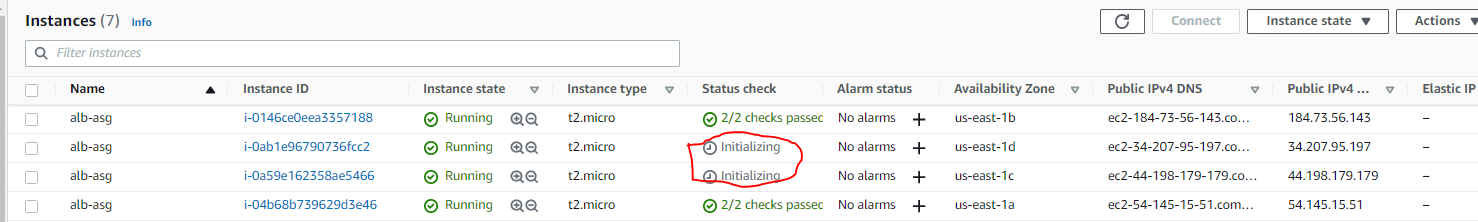
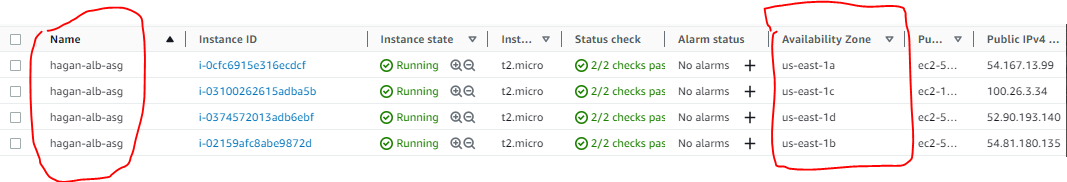
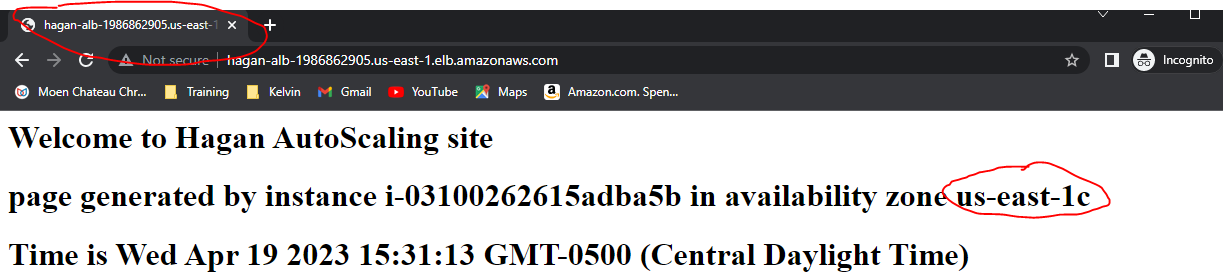
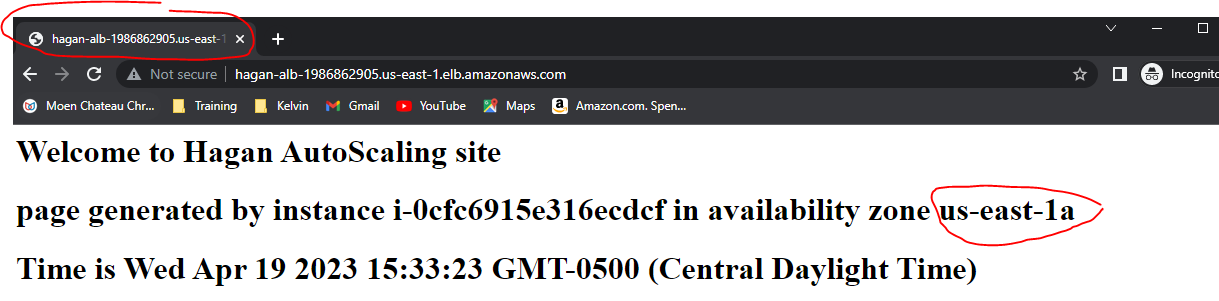
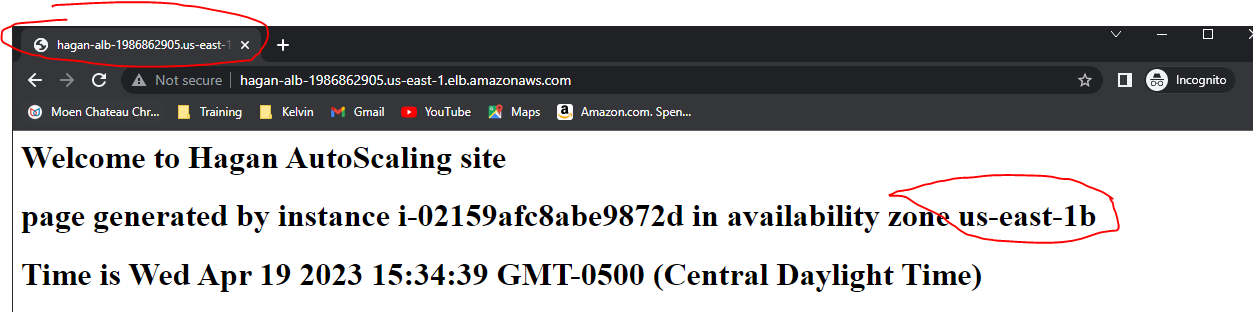
**Create an Auto Scaling Group**You can use the previous lab ***Creating an EC2 Auto Scaling Group*** with the following information:

1. **EC2** > **Auto Scaling** > **Auto Scaling Groups**
2. Click **Create Auto Scaling group**.
3. Call the group **yourlastname-alb-asg**.
4. Select **Launch Template**, and choose the template you just created and click **Next**.
5. Pick the default VPC, select Availability zones **us-east-1a**, **us-east-1b**, **us-east-1c, and us-east-1d** (or the ones selected when creating the Load Balancer) as subnets and click **Next**.
6. Check **Attach to existing load balancing**.
7. Check **Choose from your load balancer target group** and select target group **yourlastname-alb-tg**.
8. Under *Health checks*, check **Turn on Elastic Load Balancing health checks**.  
   
9. Click **Next**.
10. For *Group Size*, enter the following values:
    * *Desired Capacity*: **2**
    * *Minimum Capacity*: **2**
    * *Maximum Capacity*: **4**
11. For *Scaling Policies*, select **Target tracking scaling policy** and enter the following values:
    * *Scaling Policy Name*: **Target Tracking Policy**
    * *Metric type*: **Average CPU utilization**
    * *Target value*: **50**
    * *Instances need*: **120**
12. Click **Next**.
13. For **Add notification**, we can use this to send emails when a new instance is triggered by one the auto scaling metrics selected. Leave the default then choose **Next**.
14. Under **Tags** choose **Add tag**, provide a tag key ***Name*** and value ***yourlastname-alb-asg*** and then choose **Next**.  
    
15. Click **Create Auto Scaling Group**.
16. On the Auto Scaling group page, check the box next to your newly created group.
17. Select the **Activity** tab and scroll down to History to see instances being created.  
    
18. Go to **EC2>Instances** to see you 2 instances coming up. Wait until the Status Check passed.  
    

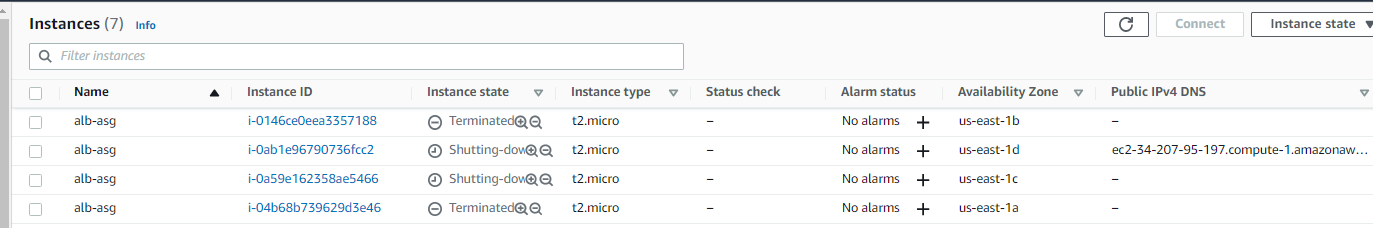
**Attempt to Connect to Website**

1. Now it time to test the Web pages and load balancer. Open a Incognito or Private Window (tab)
2. Go to EC2>Load Balancer and copy the DNS of the load balancer, paste the URL in your private browser, and hit Enter. Screenshot this page and paste into a Word doc (**yourlastname\_ALB-ASG-Lab.docx**) under the heading **Initial Load Balance Test 1**. Note the Availability Zone.  
   
3. Hit the Refresh icon on your browser. If your Load Balancer is working, your AZ should change.   
   Screenshot this page and paste into a Word doc (**yourlastname\_ALB-ASG-Lab.docx**) under the heading **Initial Load Balance Test 2.** If you hit refresh multiple times, it should toggle back and forth between the 2 instances**.**  
   

**Test Horizontal Scaling**

1. In order to trigger scaling, we need to stress the CPUs on the two running instances. First, go to Auto Scaling Groups under EC2>Auto Scaling
2. Check the box next to your need Auto Scaling Group then go to the **Monitoring (EC2)** tab below. This should show the two instances CPU utilizations (they are color coded). Both should be pretty low. The picture below shows the same thing from the **EC2>Instances** screen. Here, one instance is check and CPU monitored.  
   
3. Go to **EC2>Instances** and connect to the EC2 instance in AZ **1a** (or the first instance) by checking the box and hitting the **Connect** button. On the next screen under *EC2 Instance Connect*, click the **Connect** button. A terminal window should open in the browser.  
   
4. In the Terminal window, type **stress -c 10**. This command will start the stress CPU process and cause horizontal scaled when the CPUs crosses 50%.  
     
   
5. Repeat steps 3 & 4 for the instance in **1b (or whatever your second instance is)**.
6. Now go back the **EC2>Instances**, check the boxes next to your 2 running instances, and look at the CPU utilization % under *Monitor tab*. Hit refresh a few times and you’ll see the CPU utilizations rise.
7. Go to the *Activities tab* to see new instances will be added. Hit refresh a few times to see this.  
     
   
8. Now go to EC2>Instances to see the newly created instances starting up.  
   
9. After the Status Checks are green, screenshot this page and paste into a Word doc (**yourlastname\_ALB-ASG-Lab.docx**) under the heading **Horizontal Scaling.**
10. Now return to your browser window and hit the refresh button. You should now see each of the 4 Web server instances. Refresh and screenshot each page showing the different instances and paste into a Word doc (**yourlastname\_ALB-ASG-Lab.docx**) under the heading **Final Load Balancer Test**.  
      
      
      
      
      
      
      
    

**Clean up**

1. After capturing all screenshot, delete the Auto Scaling Group (This will terminate the instances automatically) and the Load Balancer.
2. Make sure the instances are terminated before logging out.  
   
3. Save and upload your Word Doc to Blackboard for credit.