|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
|  | | AWS Lab 6 | | | | |  | |
|  |  | | | | | | |  |
|  | | | |  |  | | | |
|  | | | | Weizhen Chen |  | | | |
|  | | | | —CCNP—Jeffery Mason &Michael Hansen |  | | | |
|  | | |  | | |  | | |

# Lab 6: Scale and Load Balance Architecture

**Elastic Load Balancing** automatically distributes incoming application traffic across multiple Amazon EC2 instances. It enables you to achieve fault tolerance in your applications by seamlessly providing the required amount of load balancing capacity needed to route application traffic.

**Auto Scaling** helps you maintain application availability and allows you to scale your Amazon EC2 capacity out or in automatically according to the conditions you define. You can use Auto Scaling to help ensure that you are running your desired number of Amazon EC2 instances. Auto Scaling can also automatically increase the number of Amazon EC2 instances during demand spikes to maintain performance and decrease capacity during lulls to reduce costs. Auto Scaling is well suited to applications that have stable demand patterns or that experience hourly, daily, or weekly variability in usage.

**The purpose**

Walkthrough using the Elastic Load Balancing (ELB) and Auto Scaling services to load balance and automatically scale your infrastructure. After the lab you can Create an Amazon Machine Image (AMI) from a running instance, create a load balancer, create a launch configuration and an Auto Scaling group, automatically scale new instances within a private subnet, Create Amazon CloudWatch alarms and monitor performance of your infrastructure

**Task 1: Create an AMI for Auto Scaling**

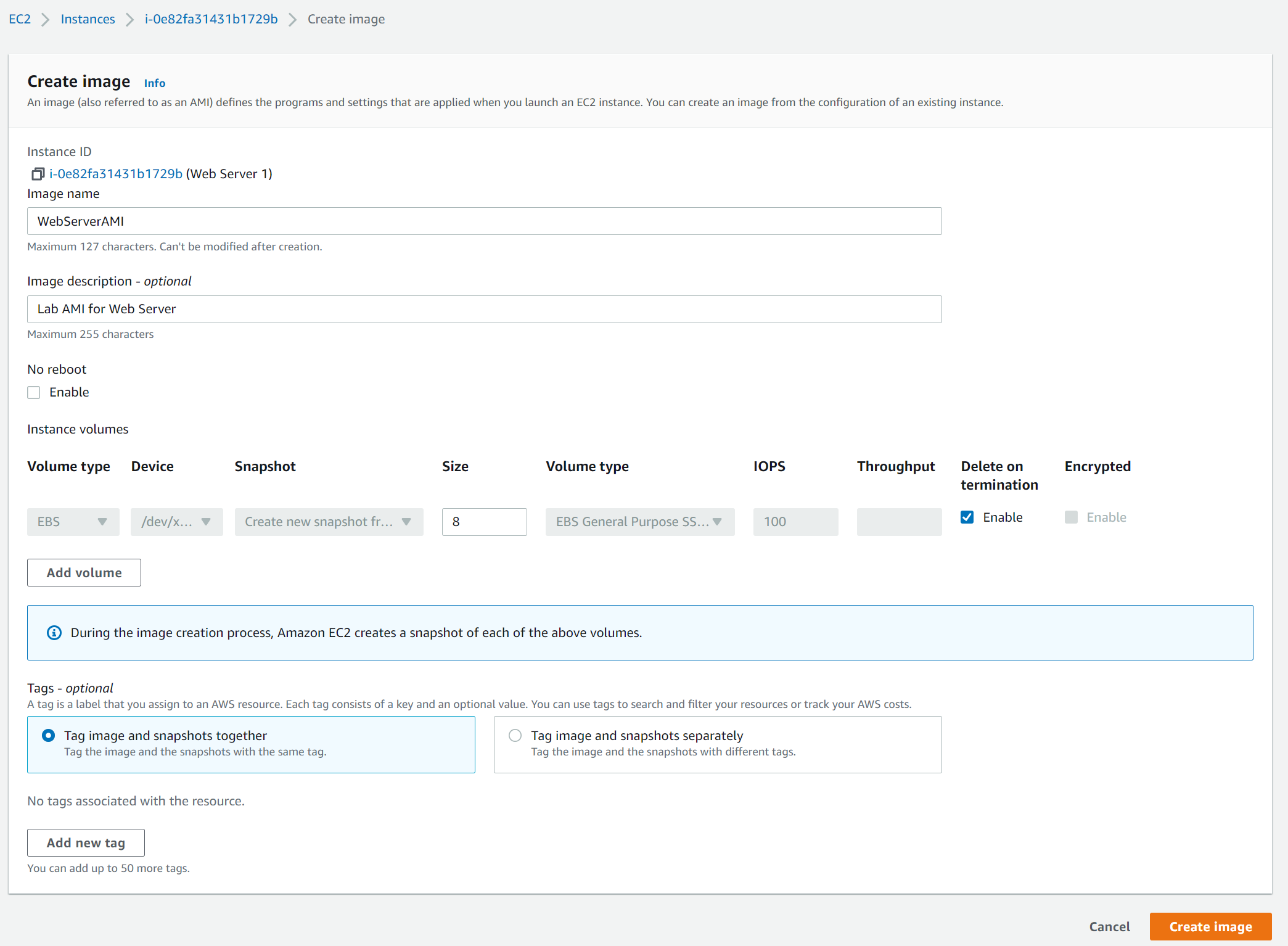
1. In the **AWS Management Console**, on the **Services** menu, click **EC2**.
2. In the navigation pane on the left, choose Instances

Confirm the Instance is running

1. Wait until the **Status Checks** for **Web Server 1** displays 2/2 checks passed
2. Select **Web Server 1** In **Actions** Menu, click **Image and templates** > **Create image**

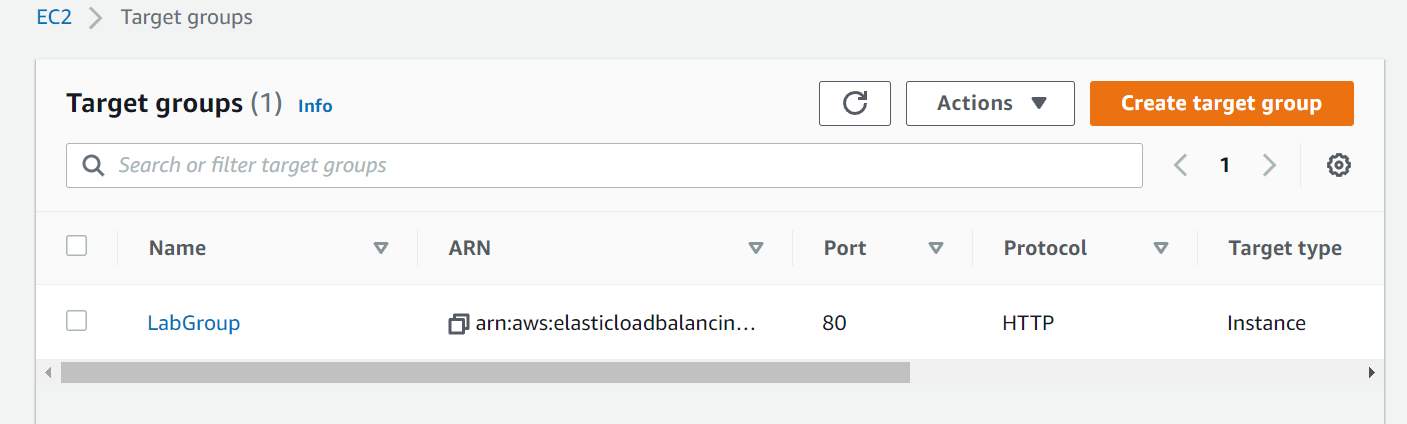
Then configure

* + **Image name:** WebServerAMI
  + **Image description: Lab** AMI for Web Server
  + Select **Lab VPC** from the **VPC** drop-down menu.



## Task 2: Create a Load Balancer

1. In the left navigation pane, choose **Target Groups**.
   * Choose a target type: **Instances**
   * **Target group name**, enter: LabGroup
   * Select **Lab VPC** from the **VPC** drop-down menu.
2. Choose **Next.** The **Register targets** screen appears.
3. Review the settings and choose **Create target group**
4. In the left navigation pane, click **Load Balancers**
5. At the top of the screen, choose **Create Load Balancer**
6. Under **Application Load Balancer**, choose **Create**



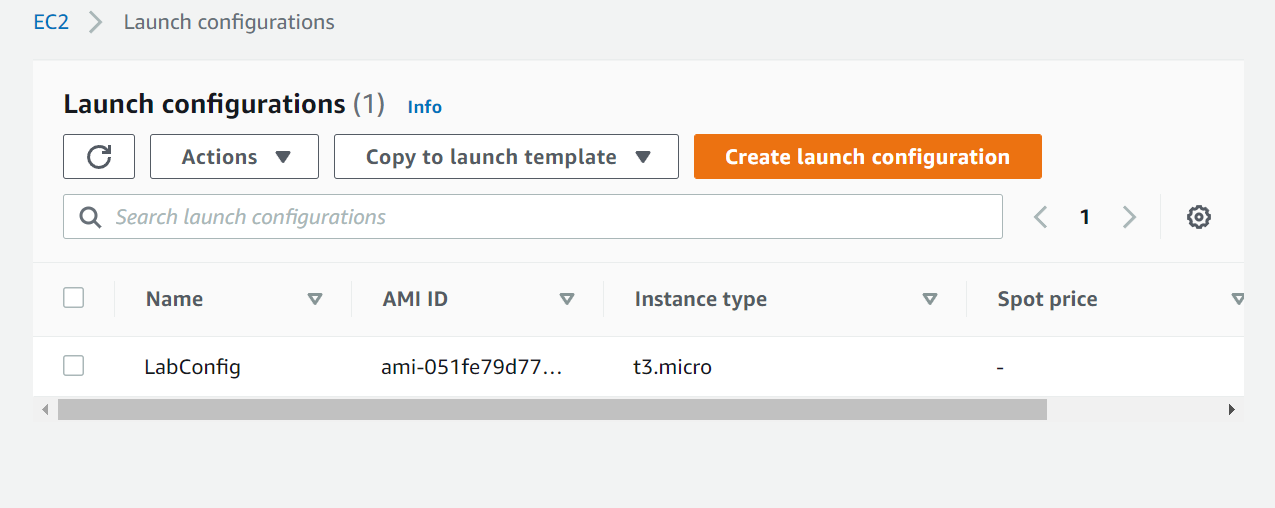
1. Under **Load balancer name**, enter: LabELB
2. Scroll down to the **Network mapping** section, then:

* For **VPC**, select: **Lab VPC, select both public subnets**
* Choose the **first** displayed Availability Zone, then select **Public Subnet 1**
* Choose the **second** displayed Availability Zone, then select **Public Subnet 2**

1. In the **Security groups** section:

* Choose the Security groups drop down menu and select **Web Security Group**
* Below the drop-down menu, choose the **X** next to the default security group to remove it.

1. set the Default action to forward to **LabGroup**.
2. Scroll to the bottom and choose Create load balancer



## Task 3: Create a Launch Configuration and Auto Scaling Group

1. In the left navigation pane, click **Launch Configurations**.
2. Click **Create launch configuration**
3. Configure these settings:

* **Launch configuration name:** LabConfig
* **Amazon Machine Image (AMI)** Choose *Web Server AMI*
* **Instance type:**
  + Choose **Choose instance type**
  + Select *t3.micro*
  + Choose
* **Additional configuration**
  + **Monitoring:** Select **Enable** EC2 instance detailed monitoring within CloudWatch

1. Under **Security groups**, you will configure the launch configuration to use the Web Security Group

* Choose **Select an existing security group**
* Select Web **Security Group**

1. Under **Key pair** configure:

* **Key pair options:** *Choose an existing key pair*
* **Existing key pair:** vockey
* Select I **acknowledge...**
* Click **Create launch configuration**

1. Select the checkbox for the *LabConfig* Launch Configuration.
2. From the **Actions** menu, choose *Create Auto Scaling group*
3. Enter Auto Scaling group name:
   * **Name:** Lab Auto Scaling Group
4. Choose **Next**
5. On the **Network** page configure

* **Network:** *Lab VPC*
* **Subnet:** Select *Private Subnet 1 (10.0.1.0/24)* **and** *Private Subnet 2 (10.0.3.0/24)*

1. Choose **Next**
2. In the **Load balancing - *optional*** pane, choose **Attach to an existing load balancer**
3. In the **Attach to an existing load balancer** pane, use the dropdown list to select *LabGroup*.
4. In the **Additional settings - optional** pane, select **Enable group metrics collection within CloudWatch**
5. Choose **Next**
6. Under **Group size**, configure:

* **Desired capacity:** 2
* **Minimum capacity:** 2
* **Maximum capacity:** 6

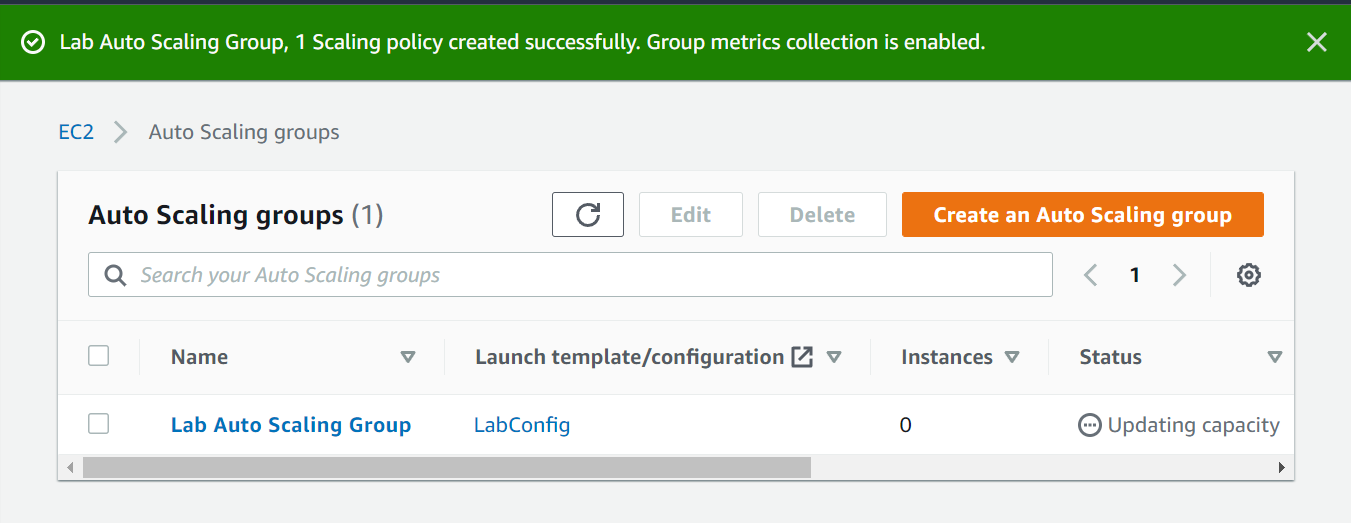
1. Under **Scaling policies**, choose *Target tracking scaling policy* and configure:

* **Lab policy name:** LabScalingPolicy
* **Metric type:** *Average CPU Utilization*
* **Target value:** 60

1. Choose Next two times
2. Choose **Add tag** and Configure the following:

* **Key:** Name
* **Value:** Lab Instance

1. Choose **Next**
2. Review the details of your Auto Scaling group, then click **Create Auto Scaling group**

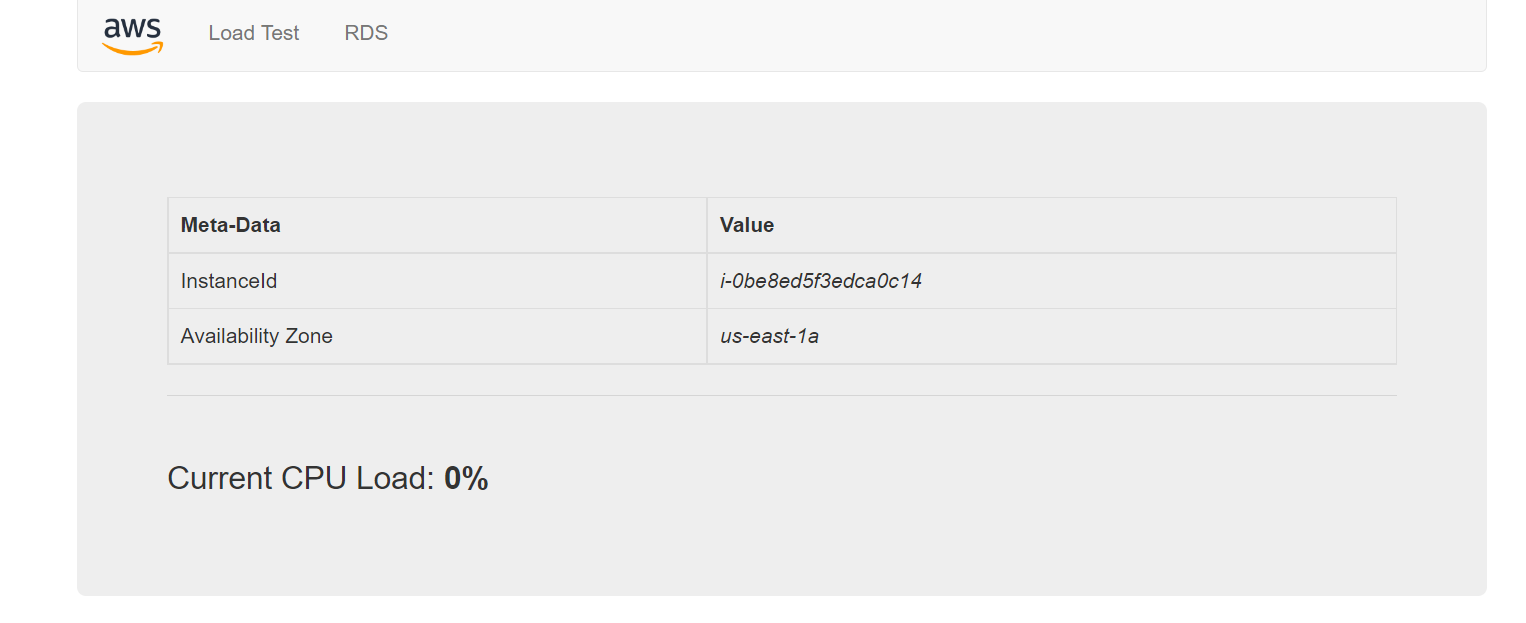


## Task 4: Verify that Load Balancing is Working

1. In the left navigation pane, click **Instances,** you should see two new instances
2. In the left navigation pane, click **Target Groups**
3. Choose *LabGroup*
4. Click the **Targets** tab

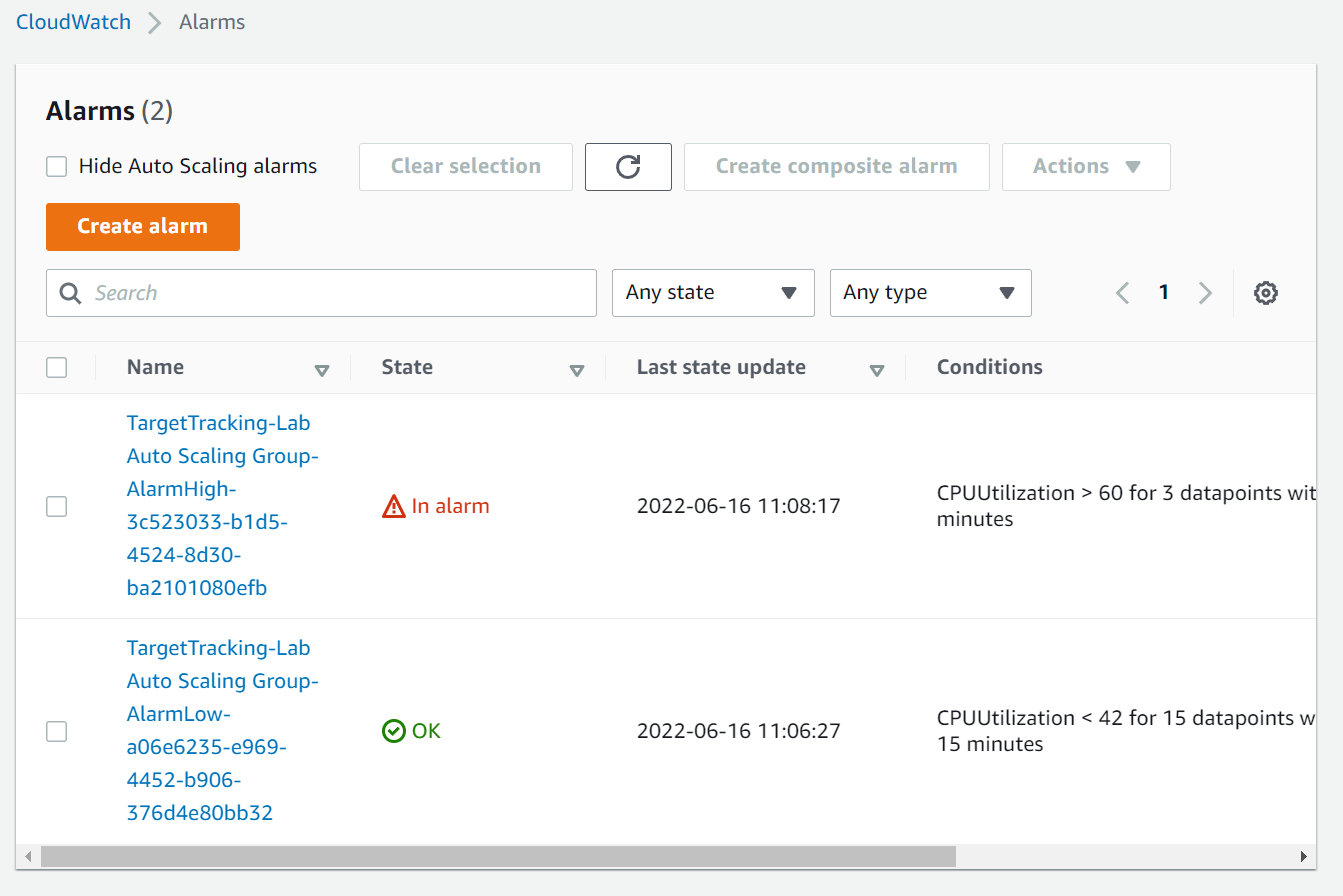
Two **Lab Instance** targets should be listed for this target group.

1. In the left navigation pane, click **Load Balancers**.
2. View the configuration file to see the setting on the last line: cat /etc/fstab
3. Open a new web browser tab, paste the DNS Name you just copied, and press Enter



## Task 5: Test Auto Scaling

1. Return to the AWS management console, but do not close the application tab — you will return to it soon.
2. On the **Services** menu, click **CloudWatch**.
3. In the left navigation pane, choose **All alarms**
4. Click the **OK** alarm, which has *AlarmHigh* in its name.
5. Return to the browser tab with the web application and Click **Load Test** beside the AWS logo.
6. Return to browser tab with the **CloudWatch** console.
7. Wait until the **AlarmHigh** alarm enters the *In alarm* state.
8. On the **Services** menu, click **EC2**.



## Task 5: Test Auto Scaling

1. Select **Web Server 1**
2. In the **instance state** menu, click **instance State > Terminate instance**

1. Choose **Terminate**

