LAB: Automatically Scale Instances with Load Balancing

You need:

- An AWS Account

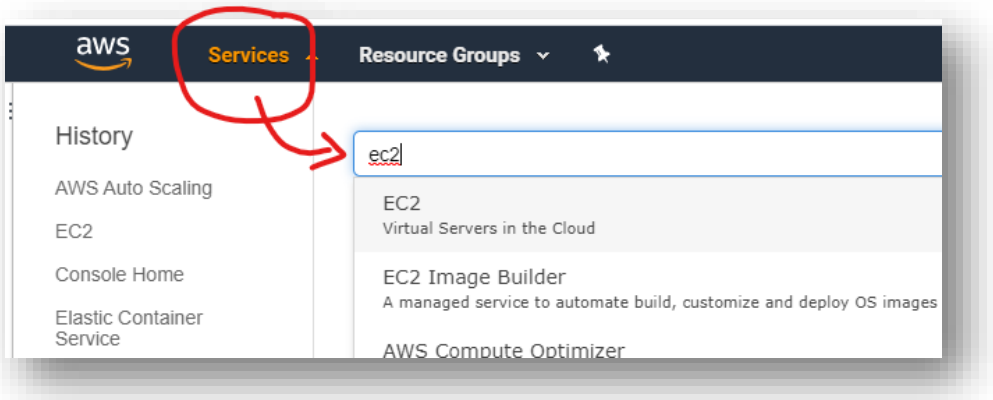
Duration of the Lab: 30 Minutes.

Difficulty: medium

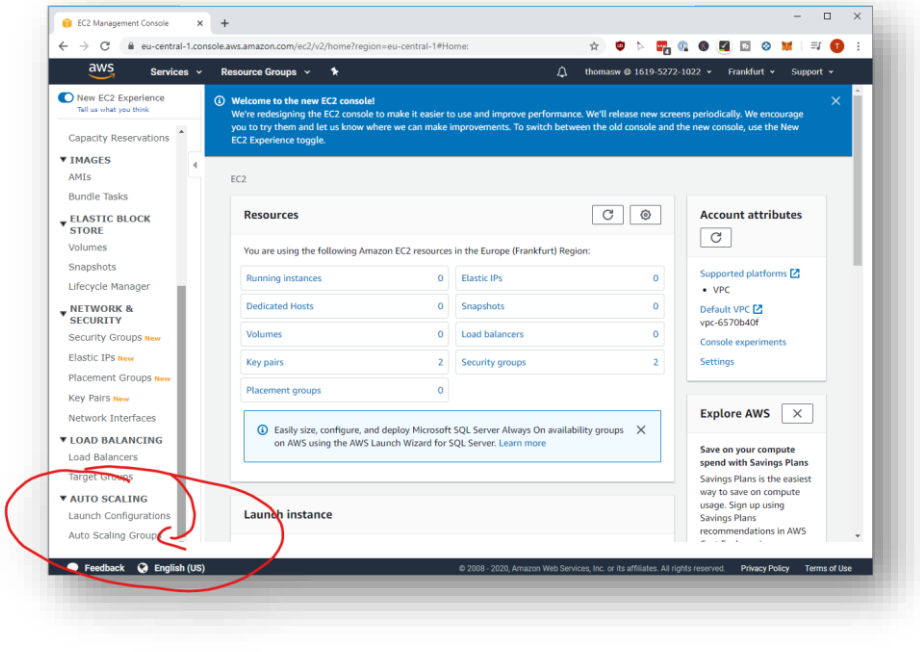
Contents

Create an Auto Scaling Group

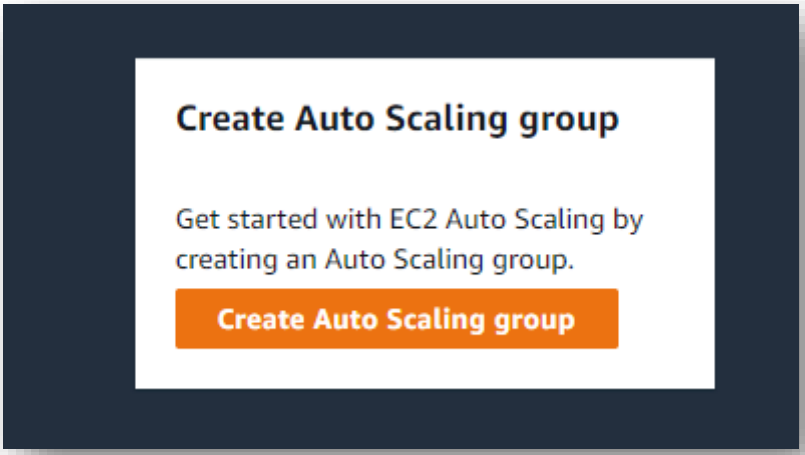
Open your EC2 Dashboard



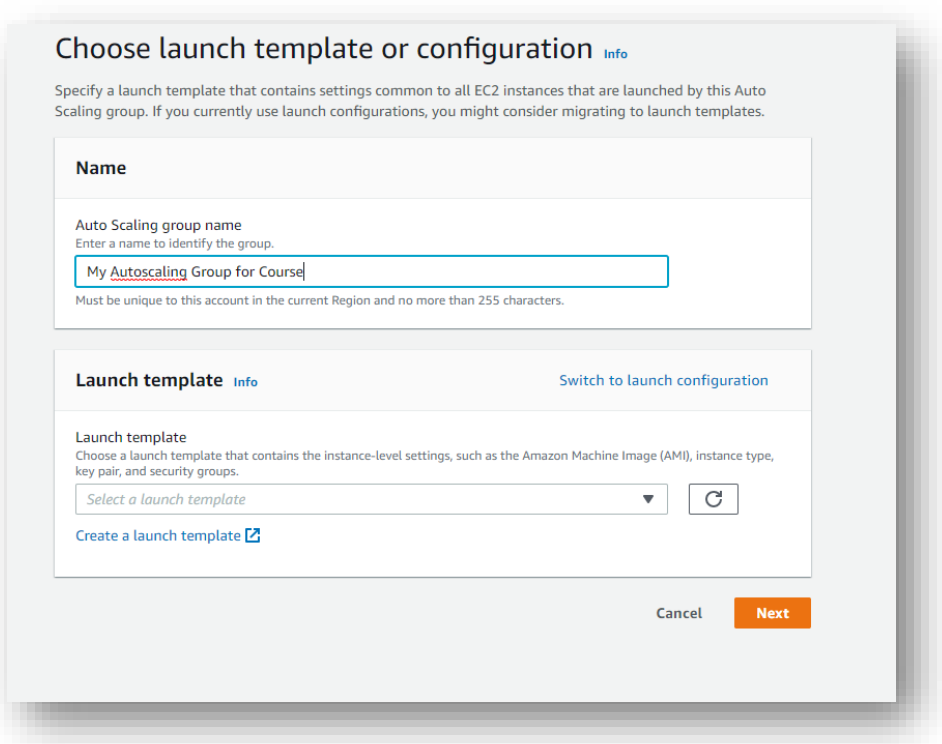
On the bottom of the list, select “Auto Scaling Group”:

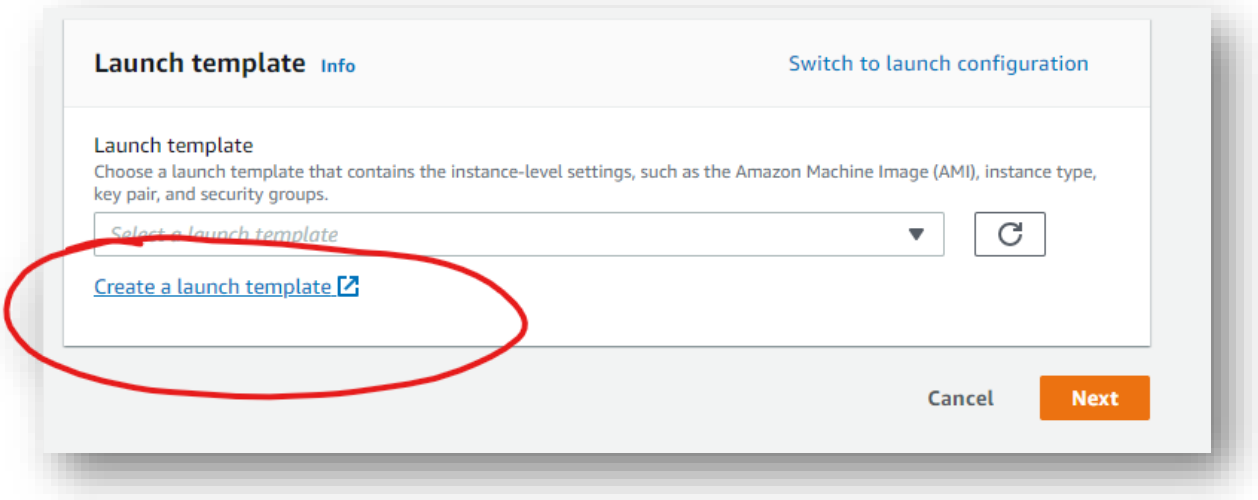


Click on Create Auto Scaling Group



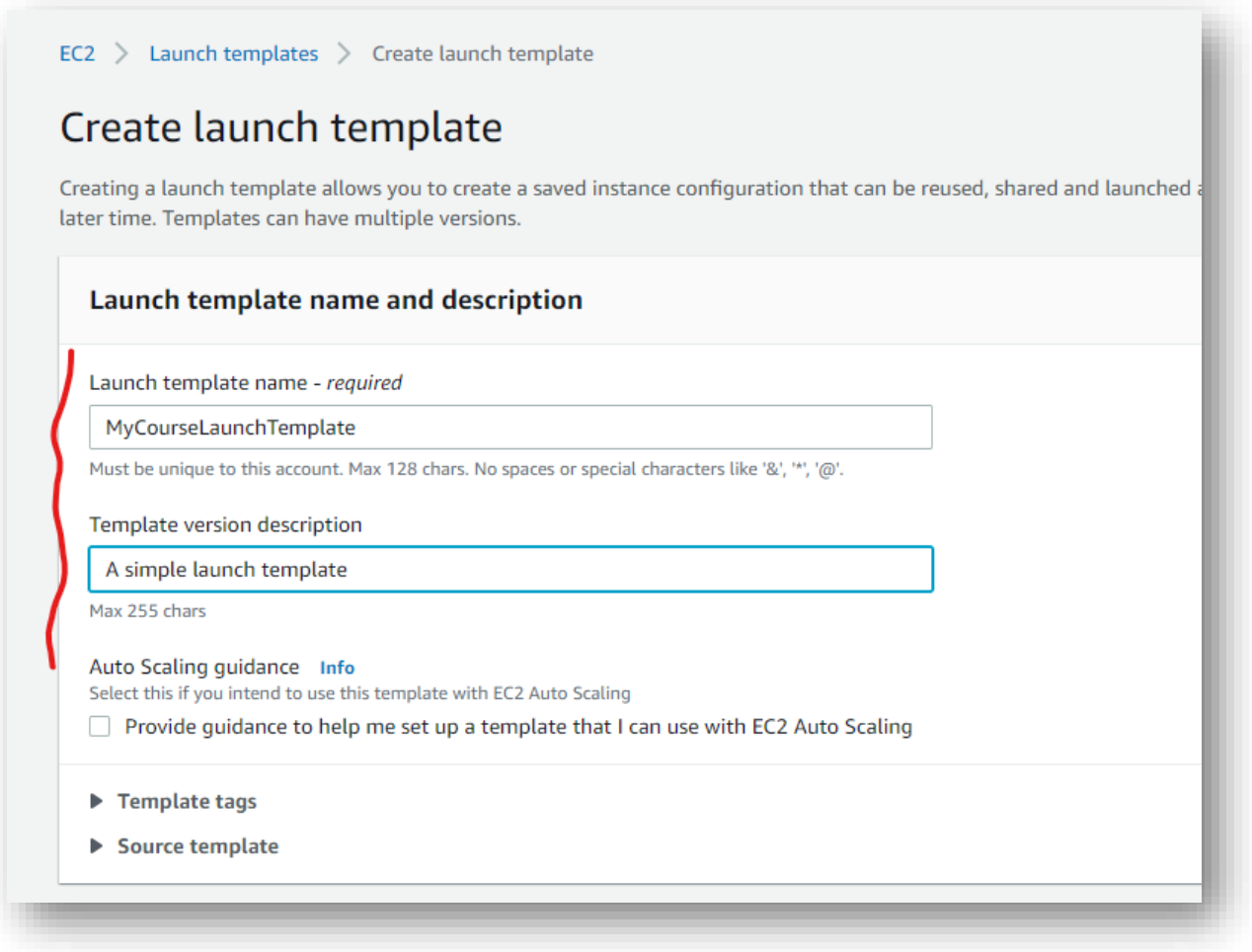
To create an Autoscaling Group you need a launch configuration. This tells the Auto Scaling group how to launch a new instance when it needs to scale out. Give the Autoscaling group a name and then create a launch template:



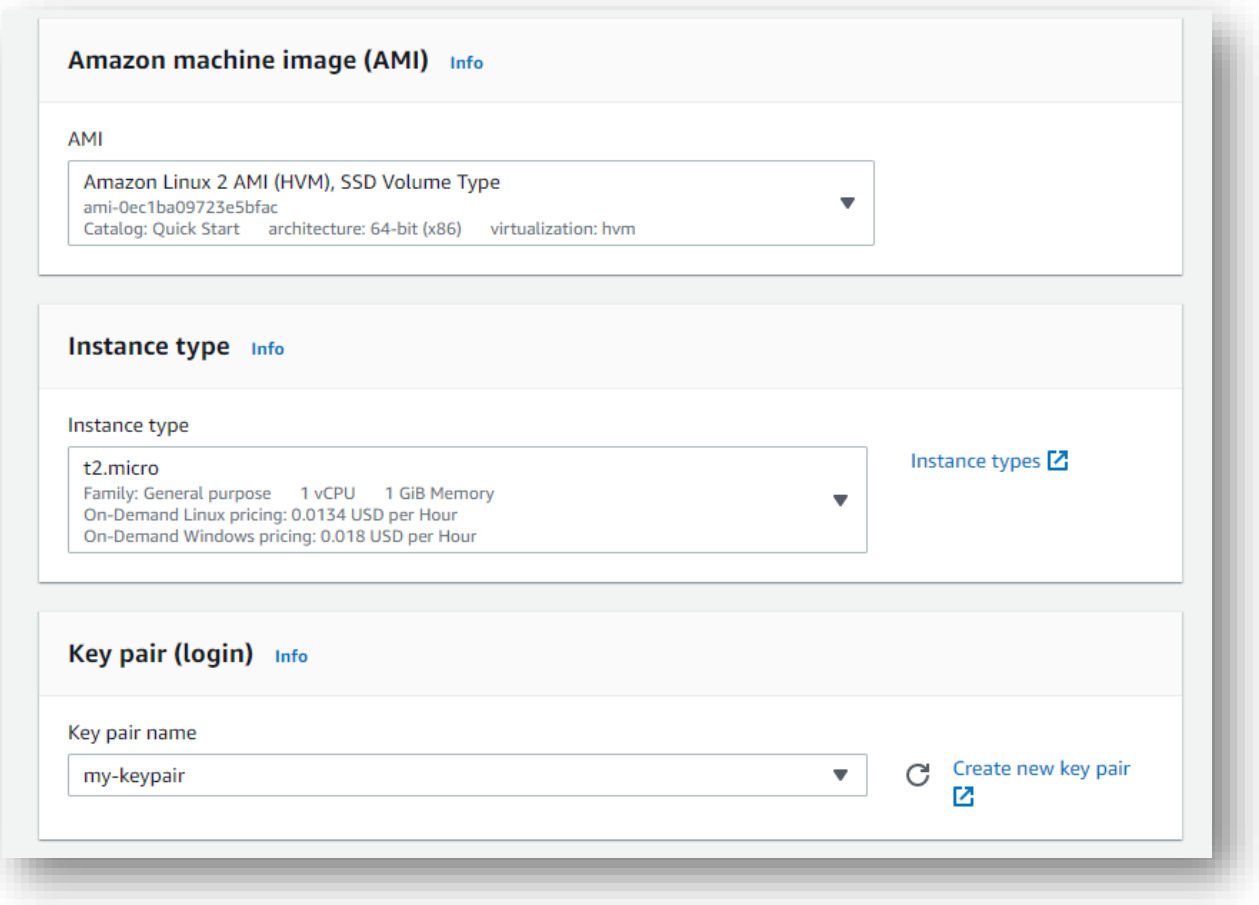


Create an Instance Launch Template

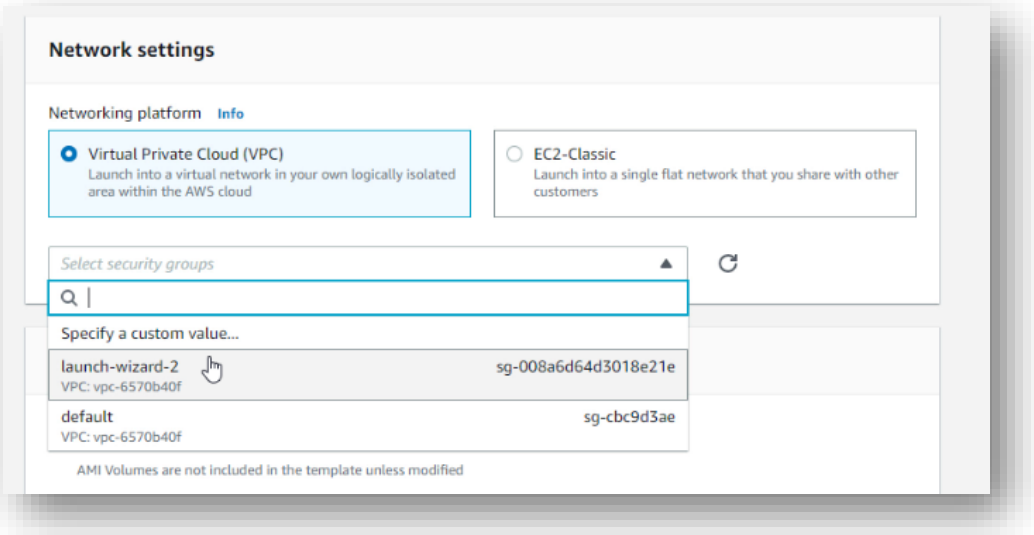
A launch template is a blueprint for AWS how to start a new Instance. Give the Launch Template a name:



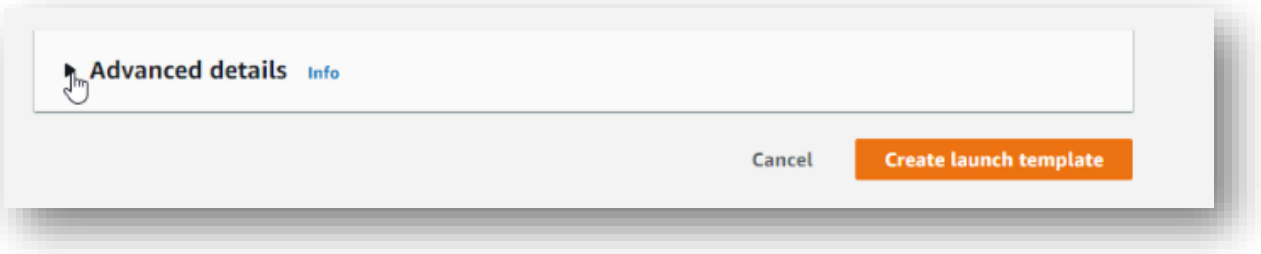
Select an AMI and an Instance Type. As well as a key-pair if you want to be able to login:

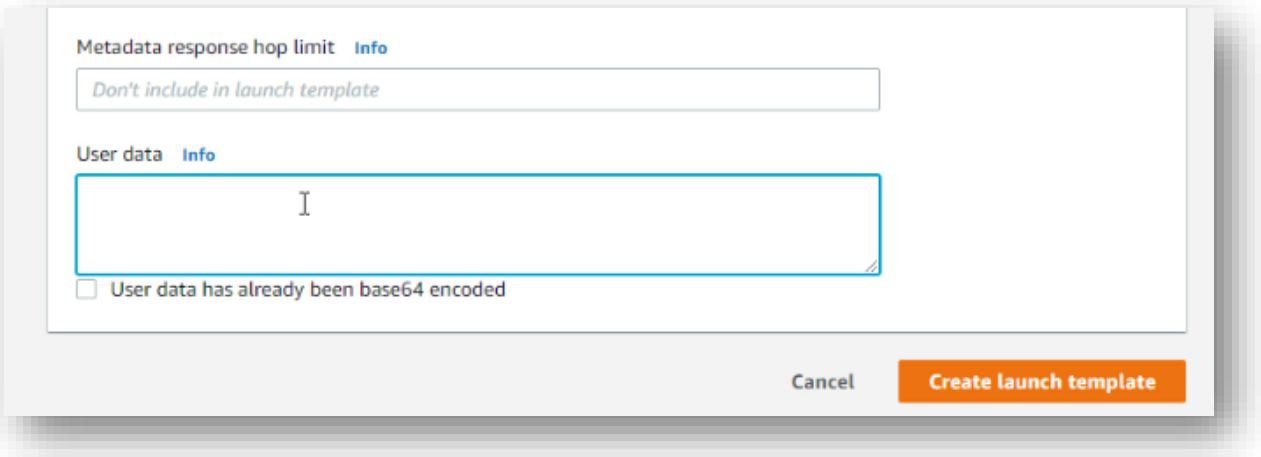


Select the Security group from the previous lectures where port 80 is open:



Under Advanced Setting add some user data to automatically install an apache + php upon launching the instance:

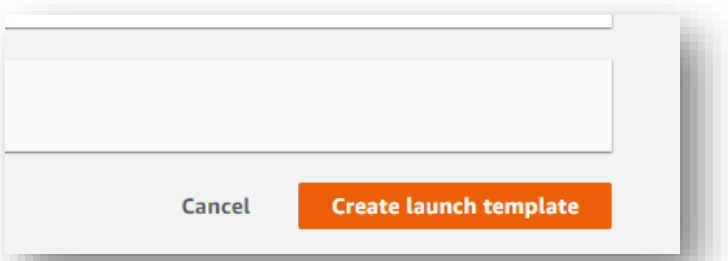


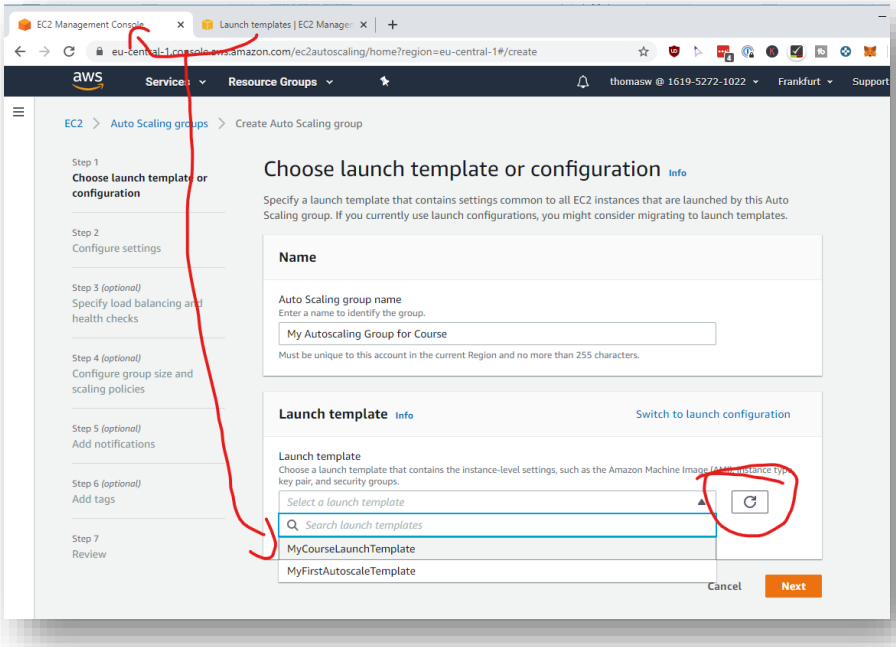


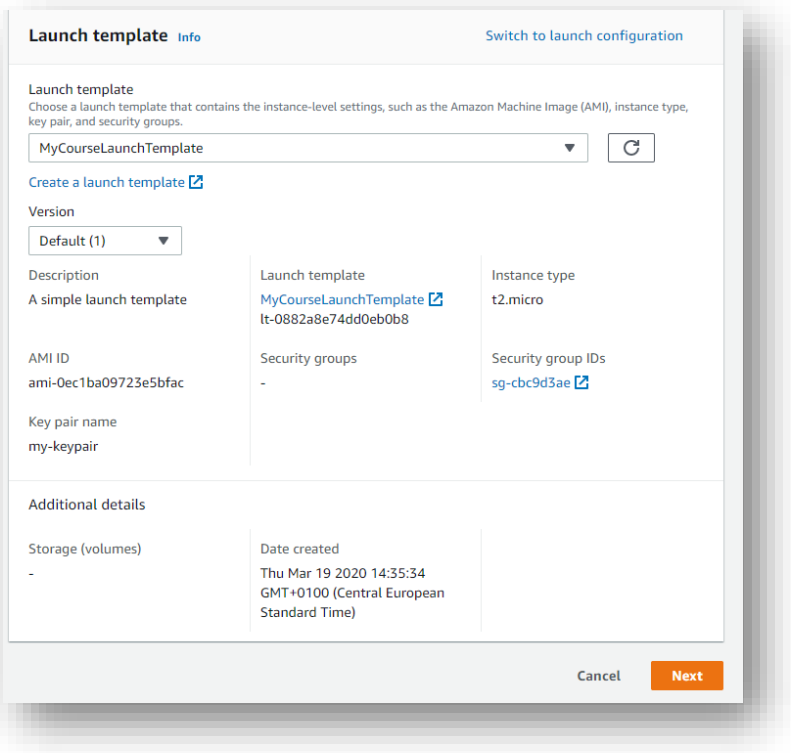
Paste the following script, it installs apache, sets some

```
#!/bin/bash
yum update -y
amazon-linux-extras install -y lamp-mariadb10.2-php7.2 php7.2
yum install -y httpd
systemctl start httpd
systemctl enable httpd
usermod -a -G apache ec2-user
chown -R ec2-user:apache /var/www
chmod 2775 /var/www
find /var/www -type d -exec chmod 2775 {} \;
find /var/www -type f -exec chmod 0664 {} \;
echo "<?php echo
file_get_contents('http://169.254.169.254/latest/meta-data/instance-
id/');" > /var/www/html/index.php
```

Then create the launch template and go back to your Auto Scaling Group Wizard, reload the wizard and select your new Launch Template:

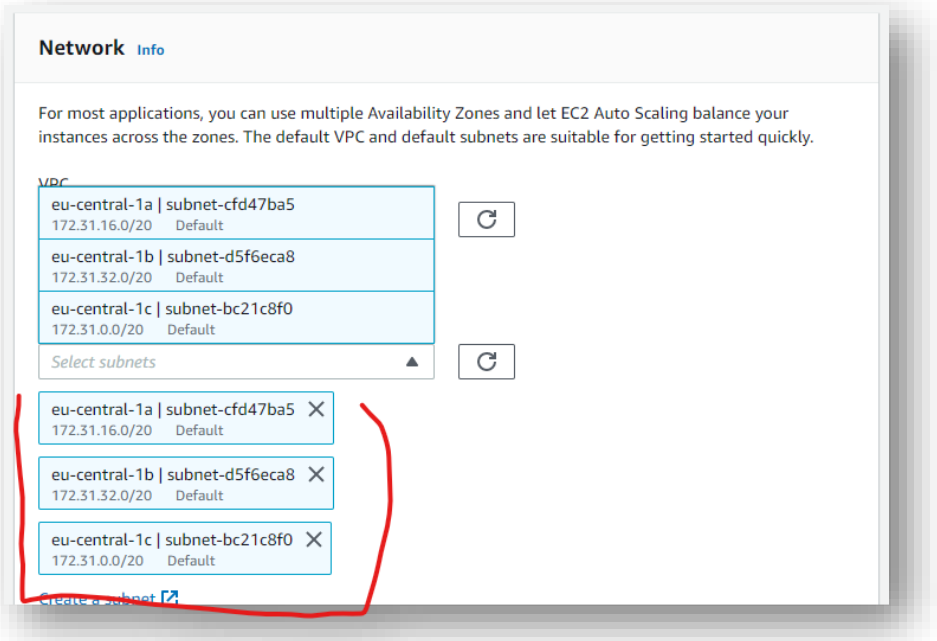






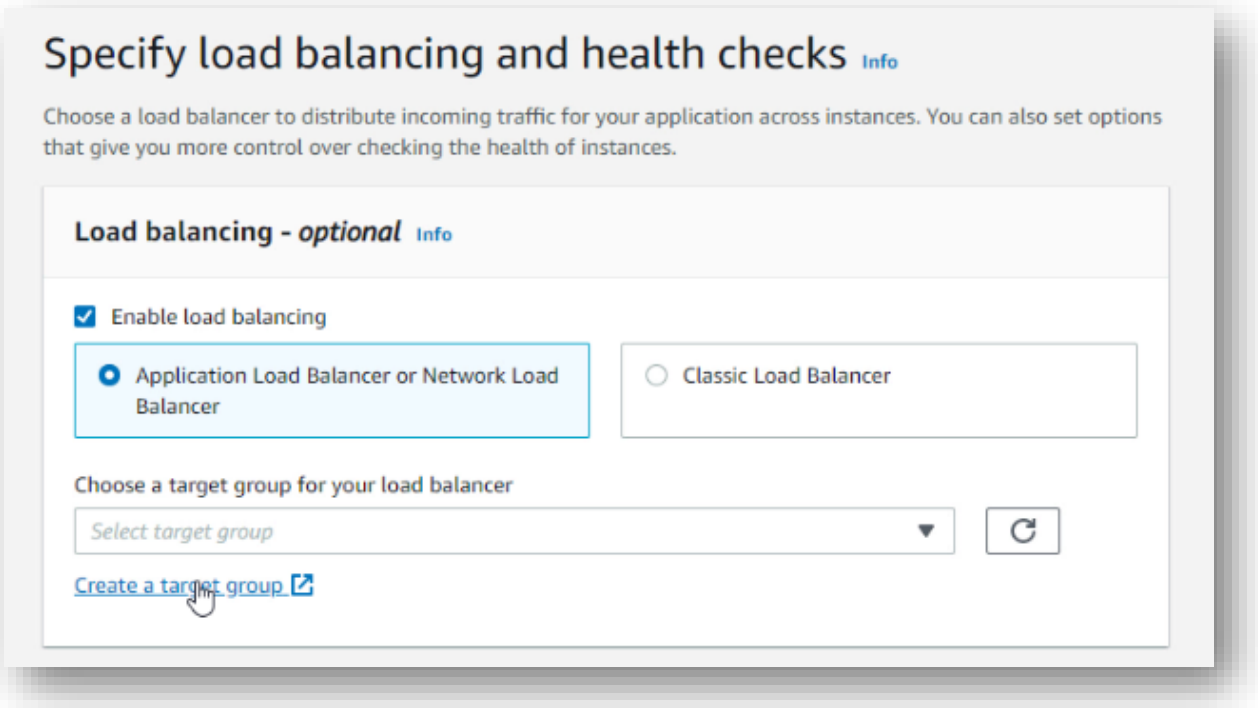
Select Placement

On the next page select the VPC (there should be only one) and all three subnets to place your instances across all three availability zones. If you have only two subnets then select only two.



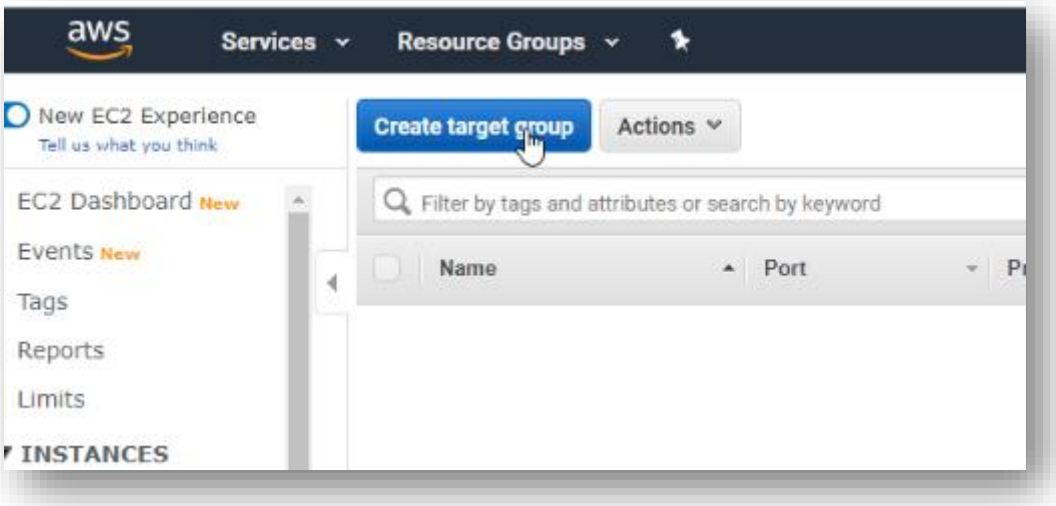
Enable Load Balancing

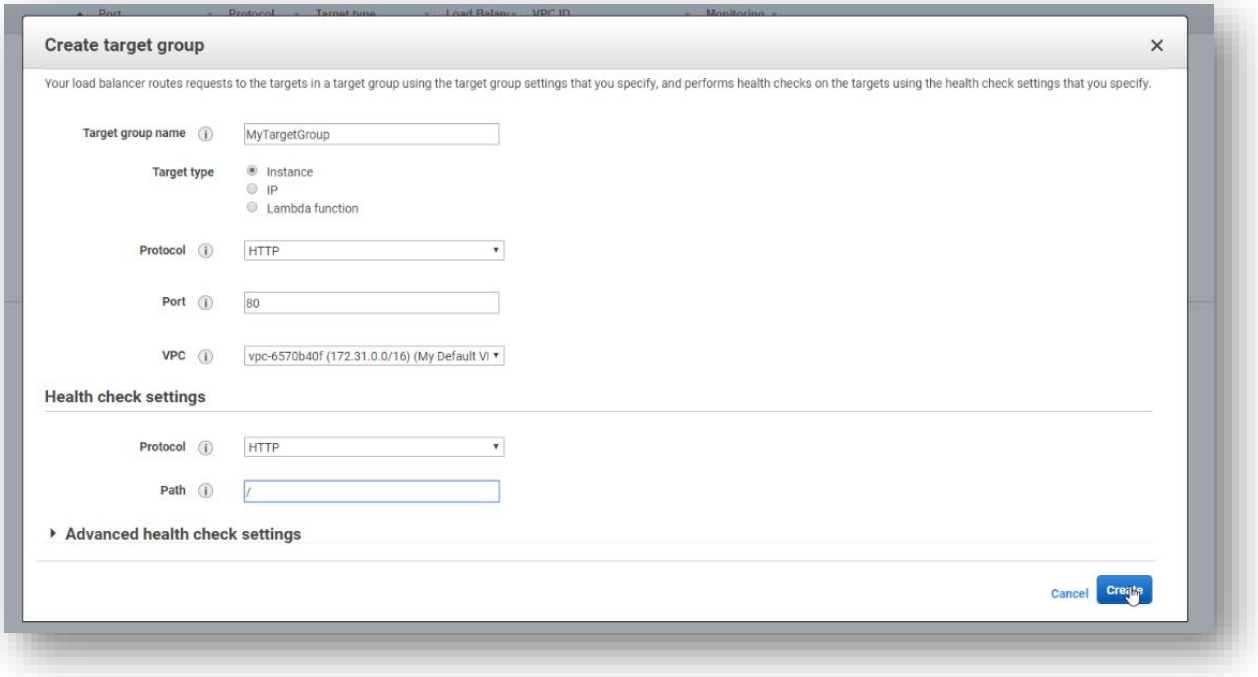
On the next page select that to enable load balancing:



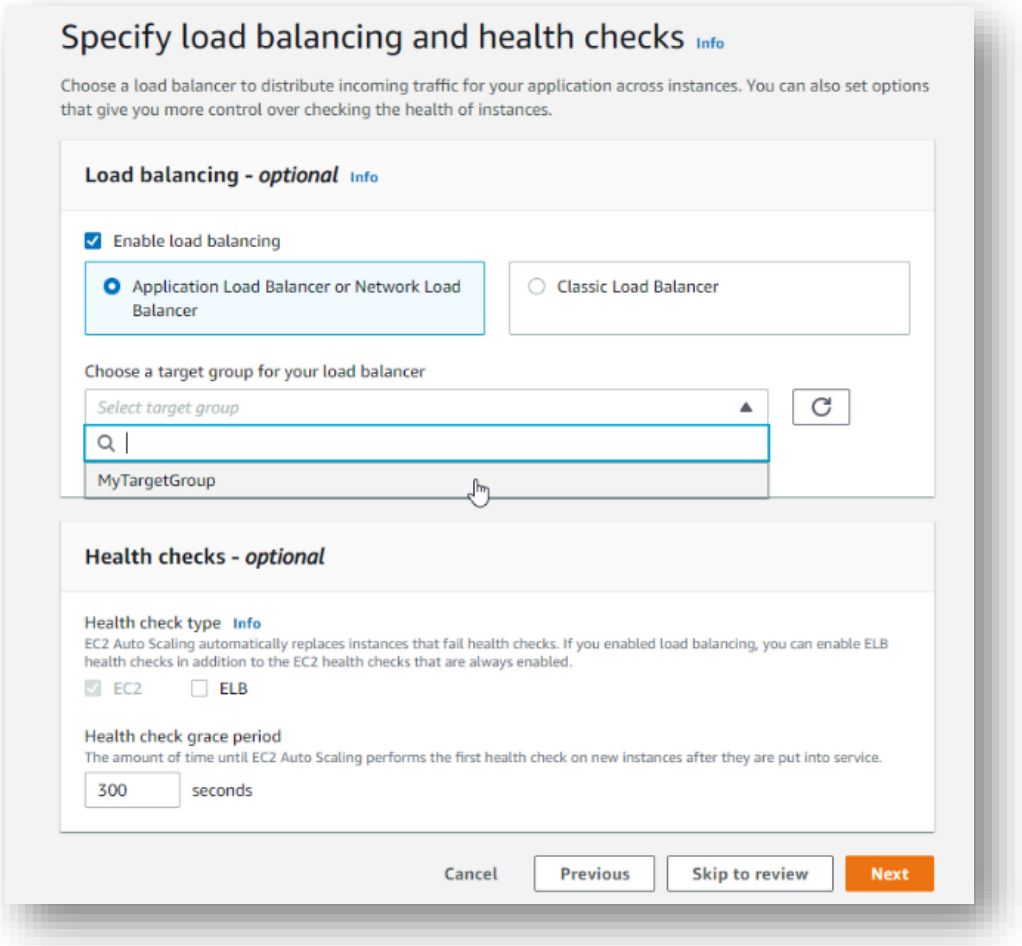
Create a Target Group

As there is no target group yet, create one. Give your target group a name and select as type "Instance":



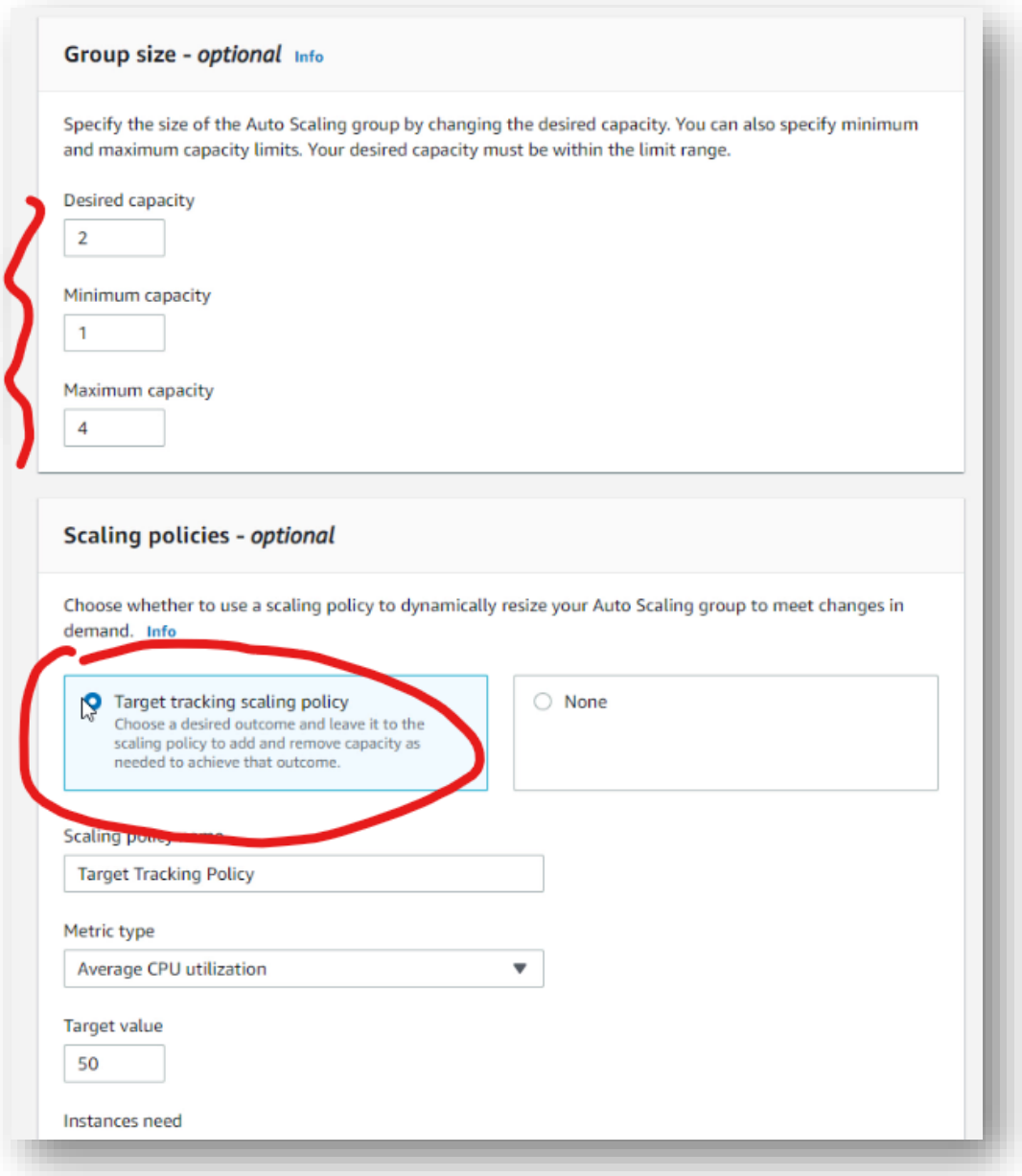


Go back to the other tab and reload the target groups.

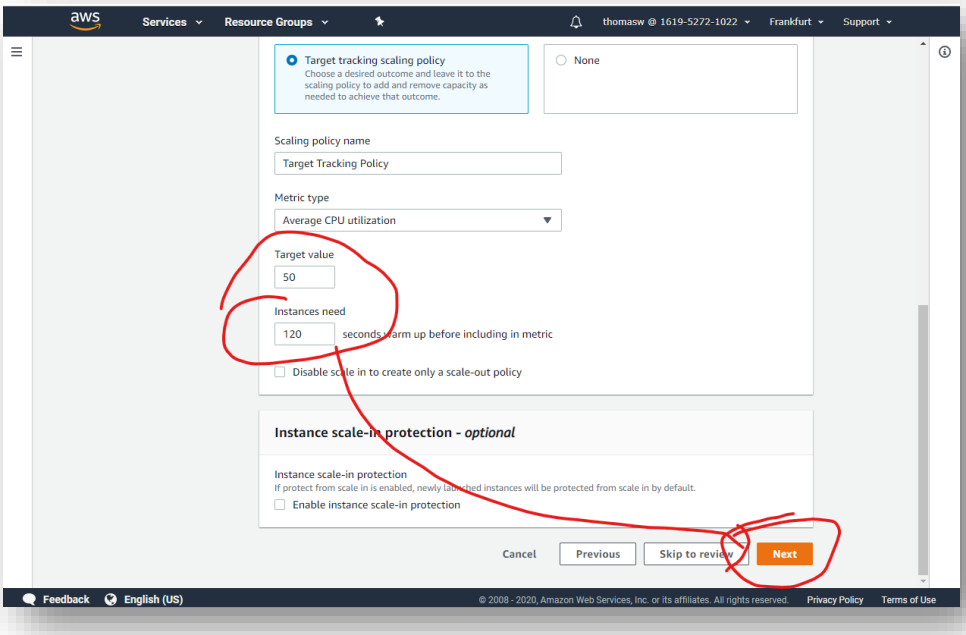


Set the Group Size and Scaling Policies

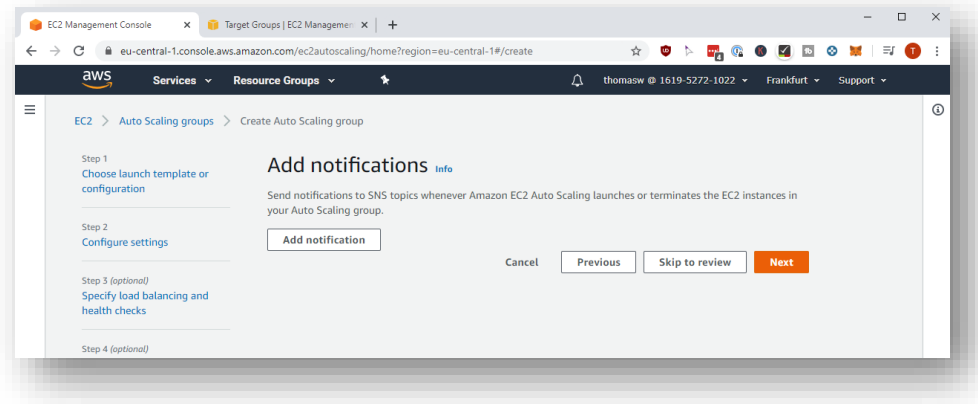
Scaling for your Auto Scaling Group. Select desired instances 2, minimum 1 and maximum 4. For the scaling policy select target tracking scaling policy.



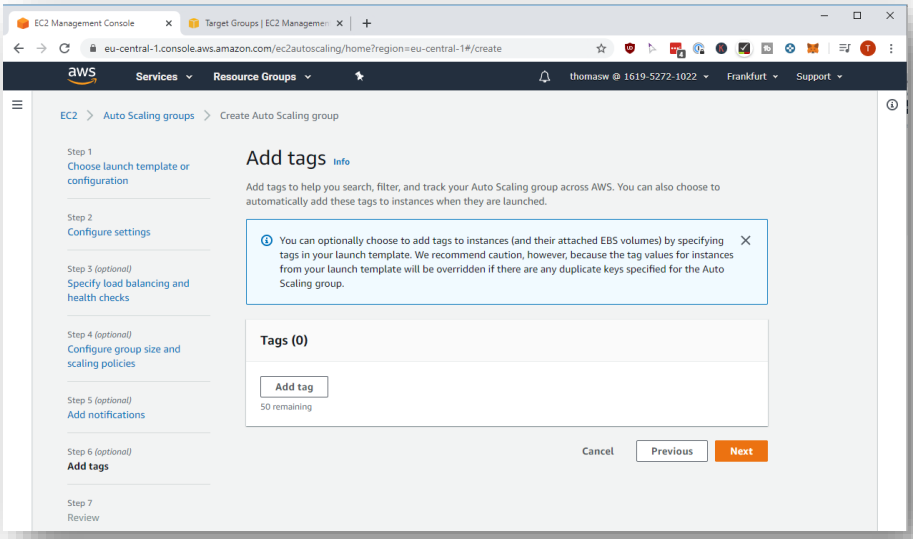
How to scale out or scale in?



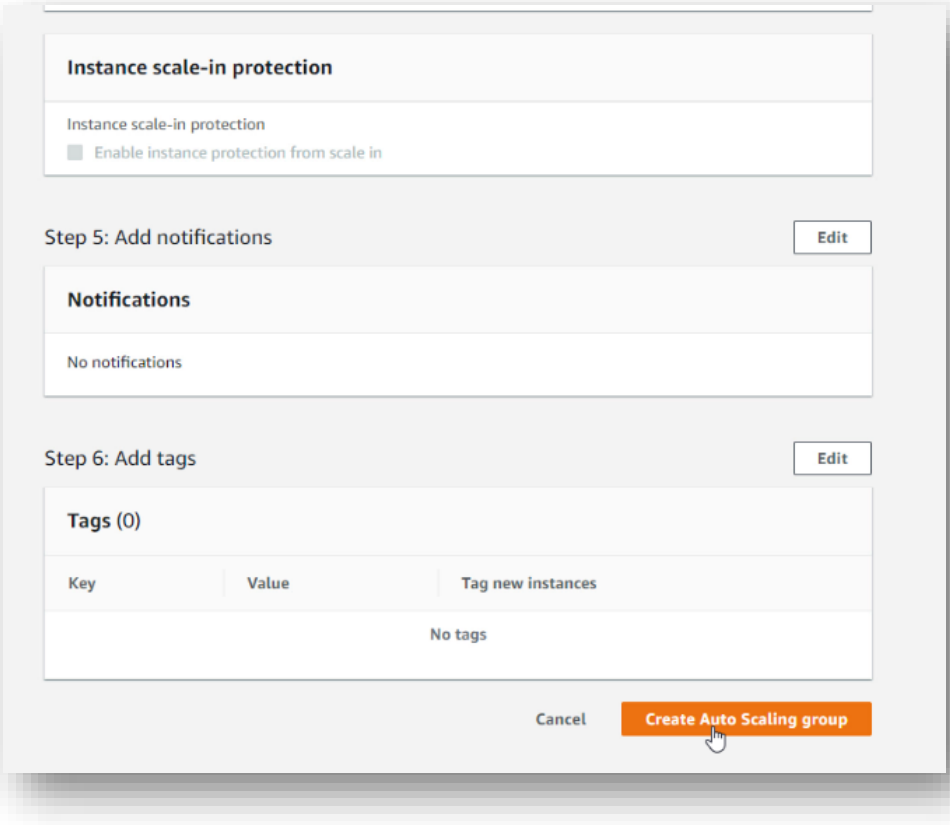
No notifications needed at this point:



No Tags needed at this point:

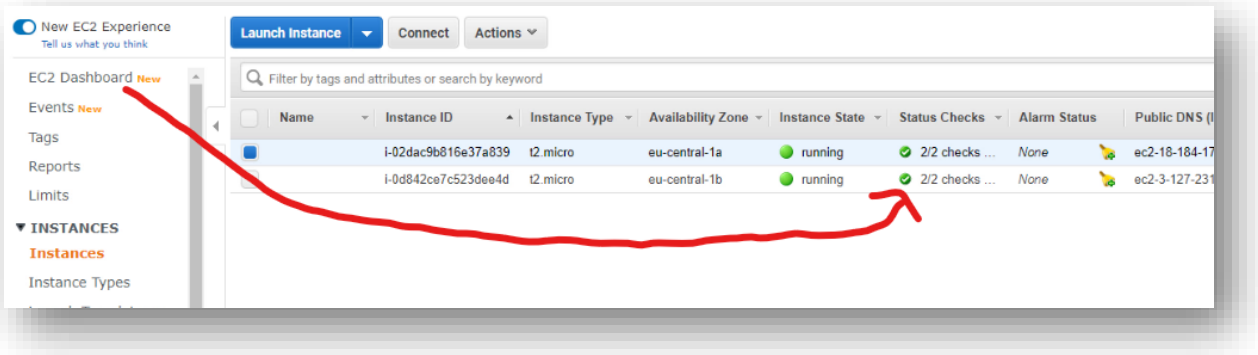


On the last page review everything and click “Create Auto Scaling Group”:

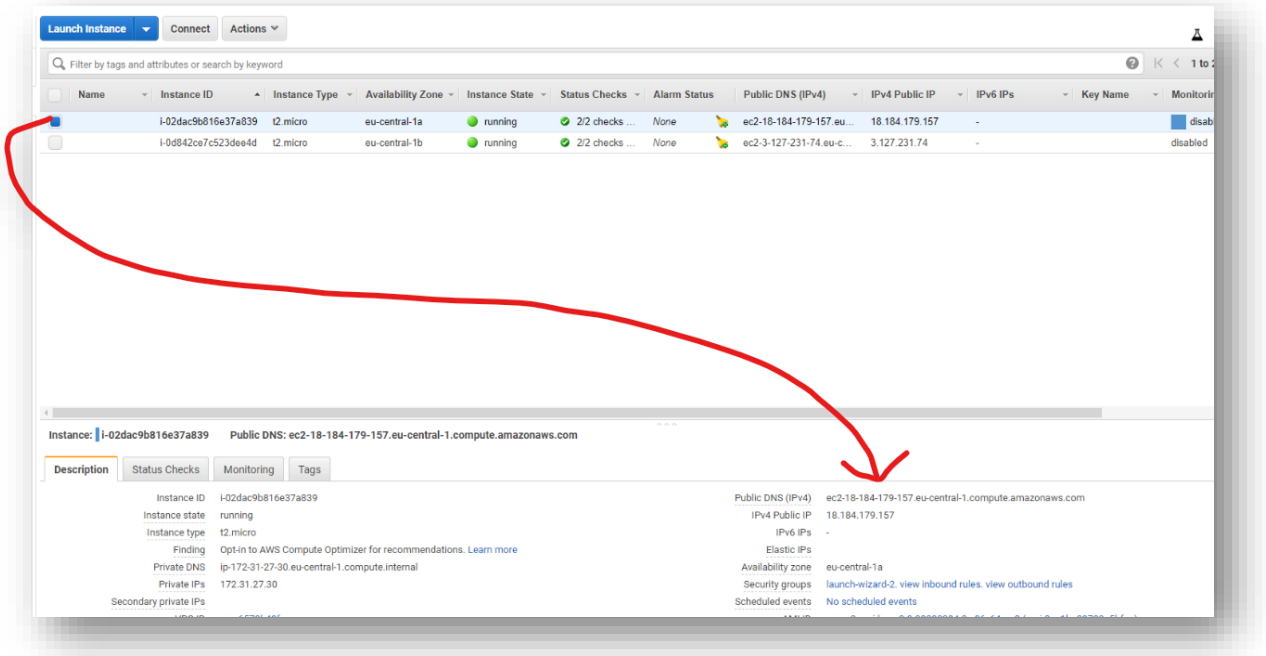


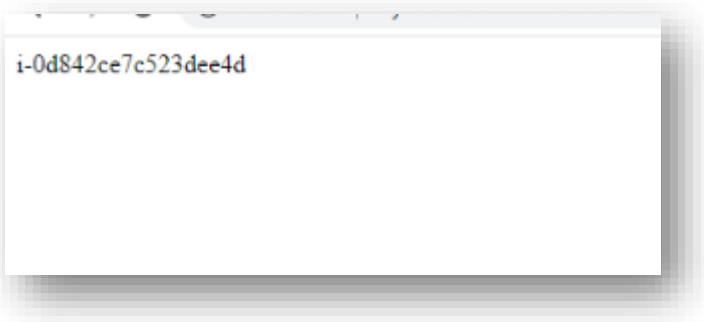
Verify the Instances are Running

Check your EC2 Dashboard and wait until the two instances are ready:



Check if you can reach the instances by opening their public DNS in a new tab:



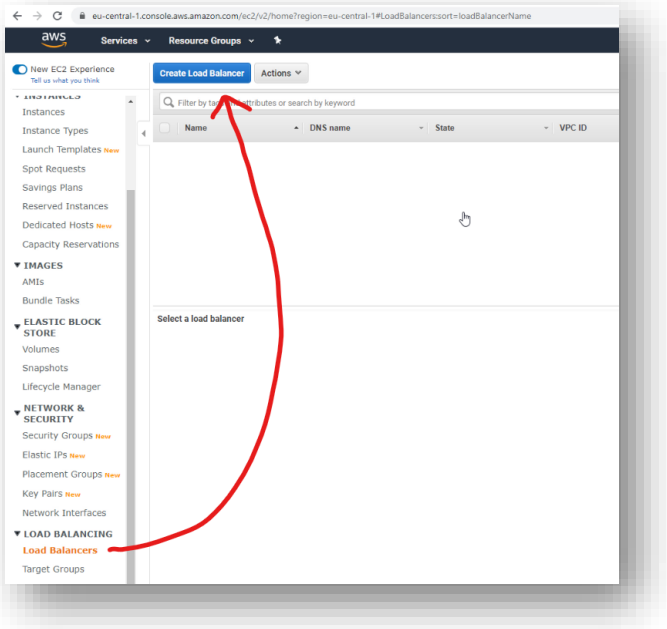


Troubleshooting

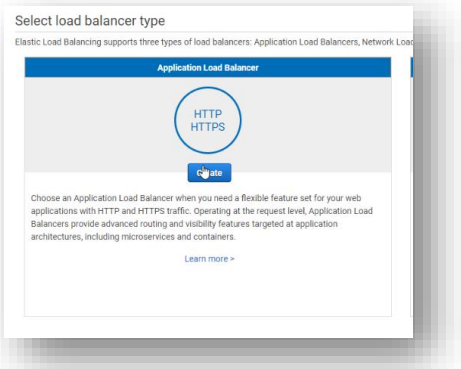
If your instances are up and running, but you can't open the page, make sure the port 80 in the security group is open. Click on the security group inbound rules and add port 80.

Create a Load Balancer

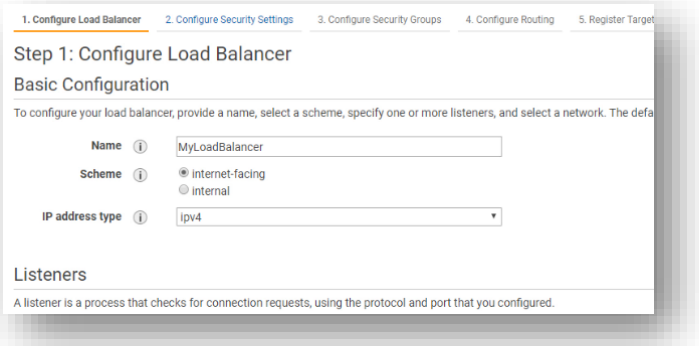
Next, we create a load balancer. You find them on the left side in your EC2 Dashboard under “Load Balancing”:



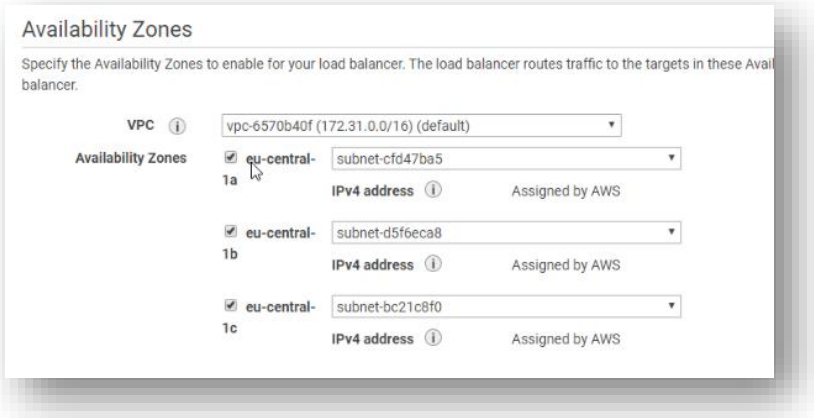
Select the Application Load Balancer



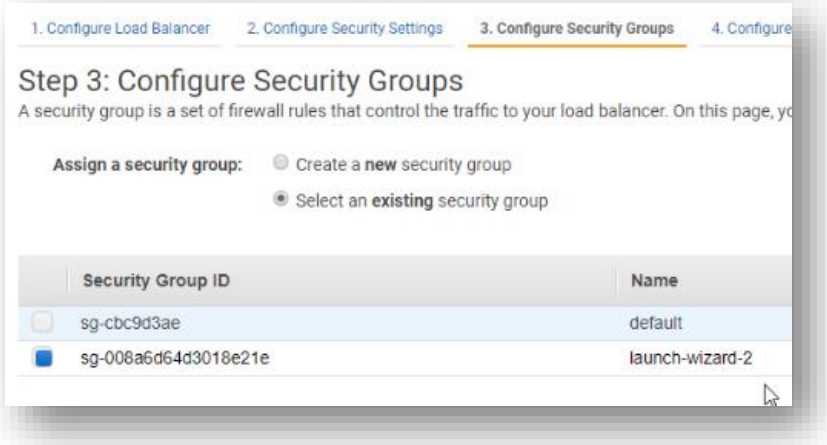
Give the Load Balancer a name:



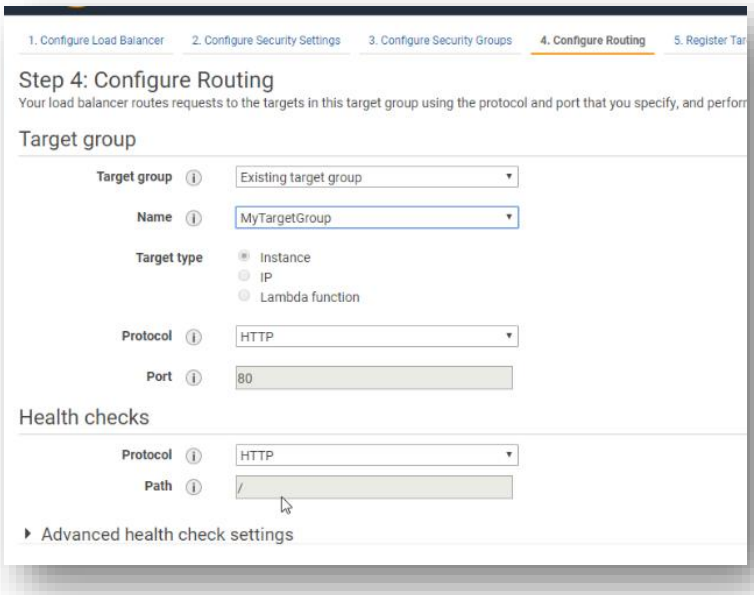
And place the load balancer in all three subnets:



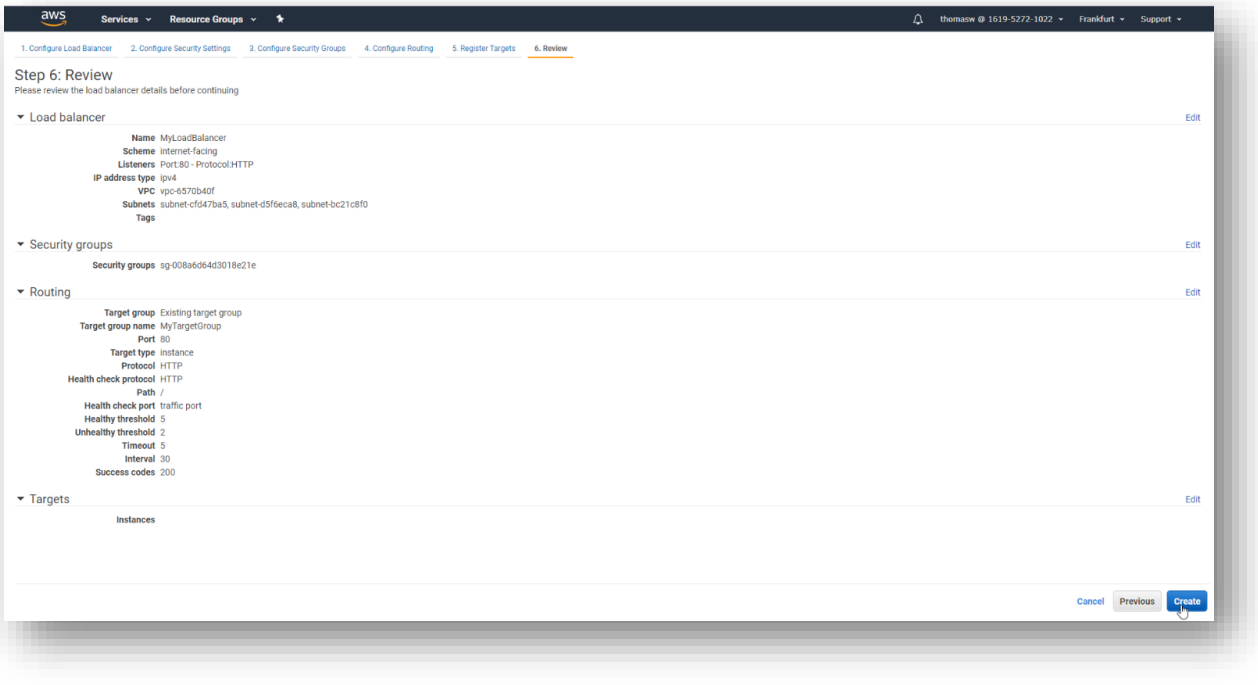
Attach the same security group that has port 80 open:



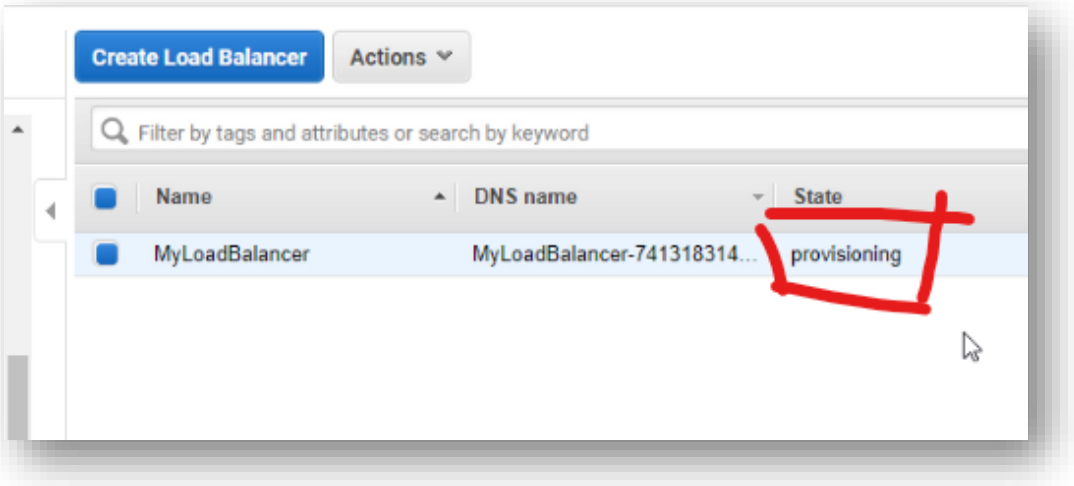
Select the *existing* Target Group from your Auto Scaling Group:



Review everything and click Create:

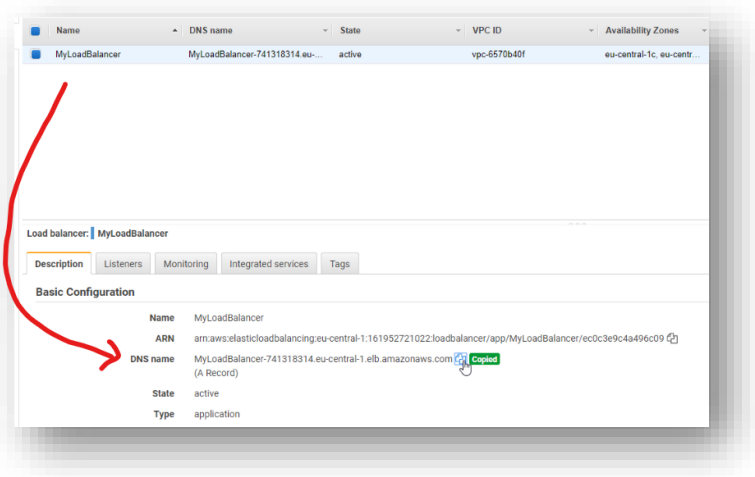


Wait until the Load Balancer is provisioned:

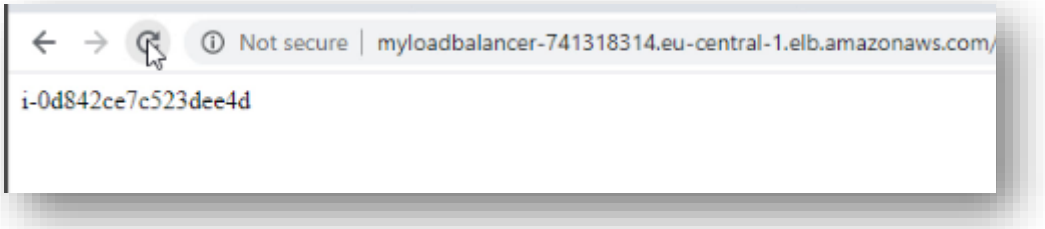


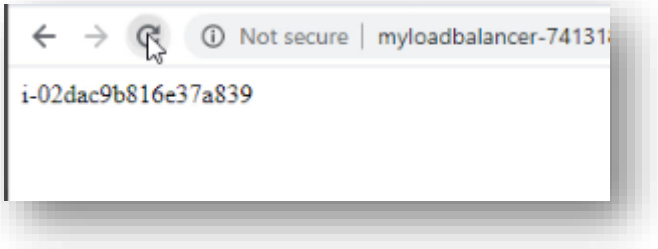
Testing the Load Balancer

Once it's ready copy the address of the load balancer and open it in a new tab:



If you reload the tab, see that the instance id changes, because the load balancer does round robin load balancing:





Tear Down and Clean Up

To save costs, let's remove everything again.

1. Remove the Auto Scaling Group
2. Remove the Load Balancer
3. Remove the Target Group
4. Remove the Instance Launch Template
5. Terminate the remaining EC2 Instances

Lab End
