Lesson 10 Demo 03

Creating an ECS Cluster with Auto Scaling

Objective: To demonstrate how to create an ECS cluster with auto scaling and container Insights for enhanced monitoring and automatic adjustment of resource allocation based on demand

Tools required: AWS Management Console

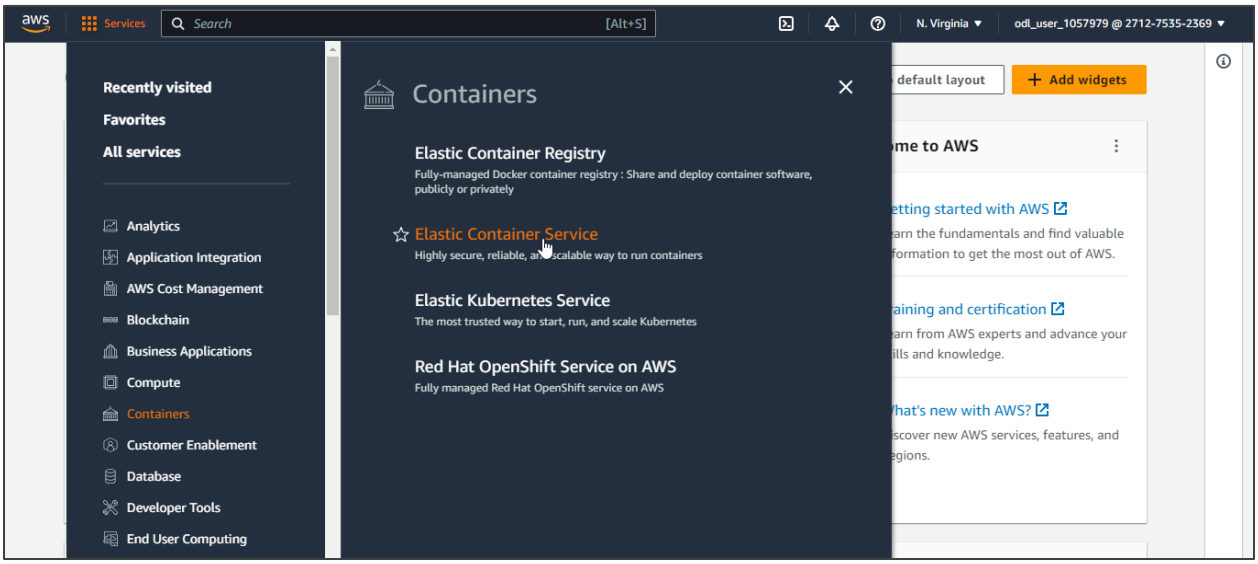
Prerequisites: A key pair for an EC2 instance must be created.

Steps to be followed:

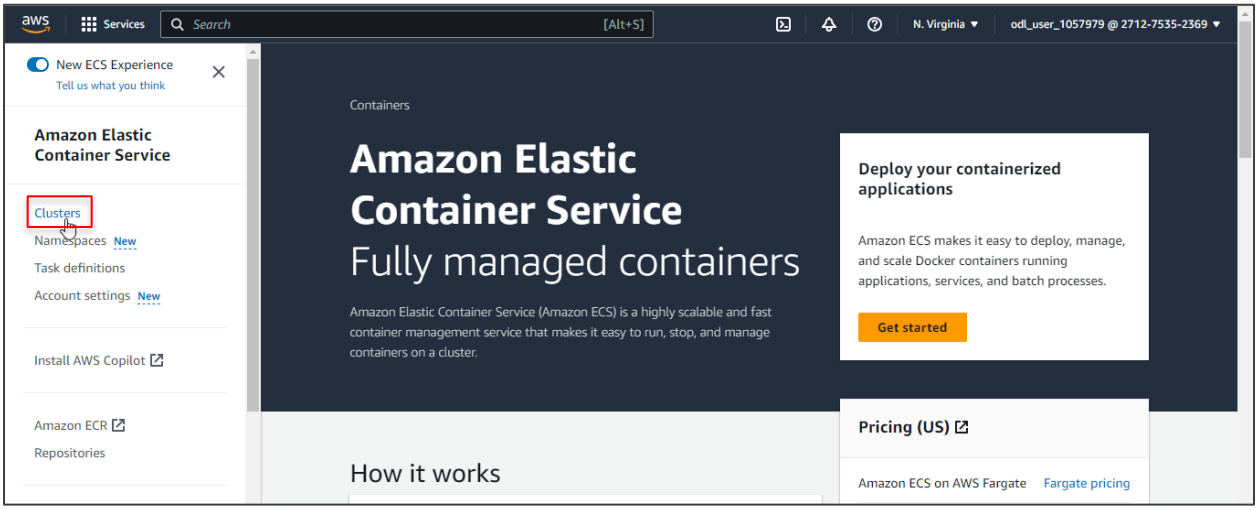
1. Create an ECS cluster
2. Create a task definition
3. Run the task definition on the cluster

Step 1: Create an ECS cluster

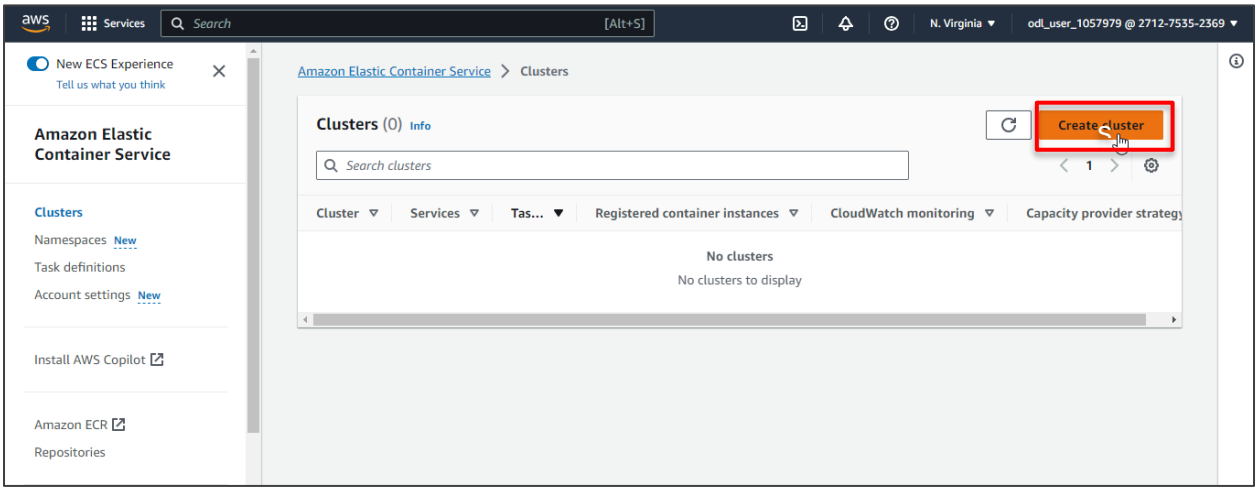
1.1 Navigate to the AWS Console, select **Services** and then choose **ECS** under **Containers**



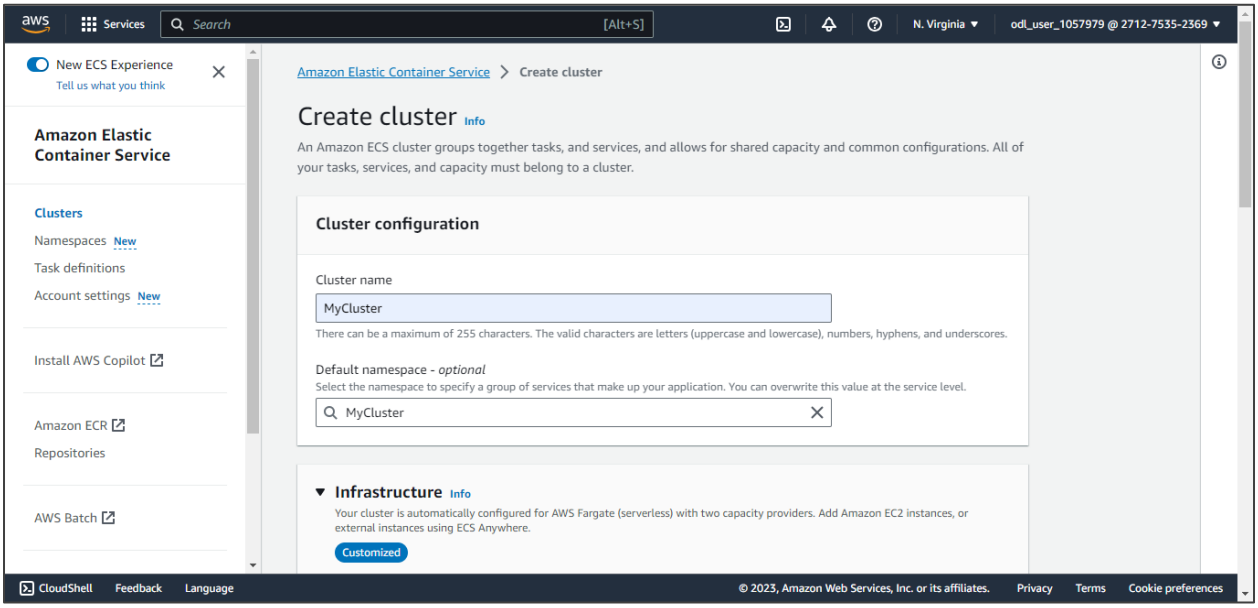
1.2 Click on **Clusters** in the navigation pane

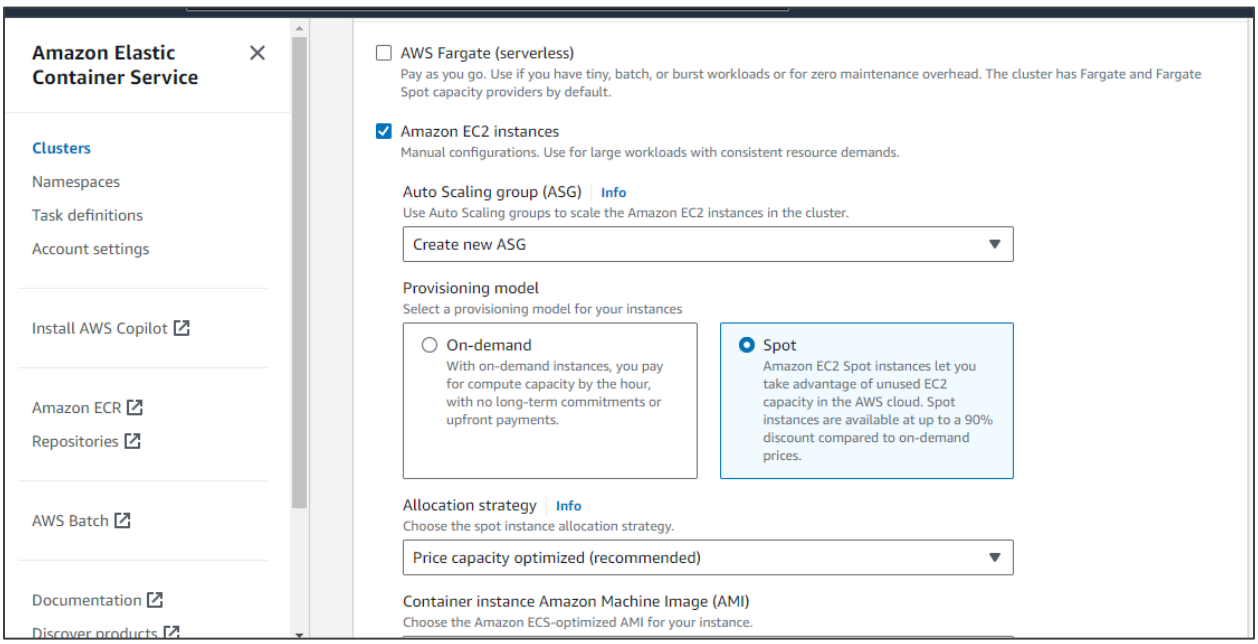


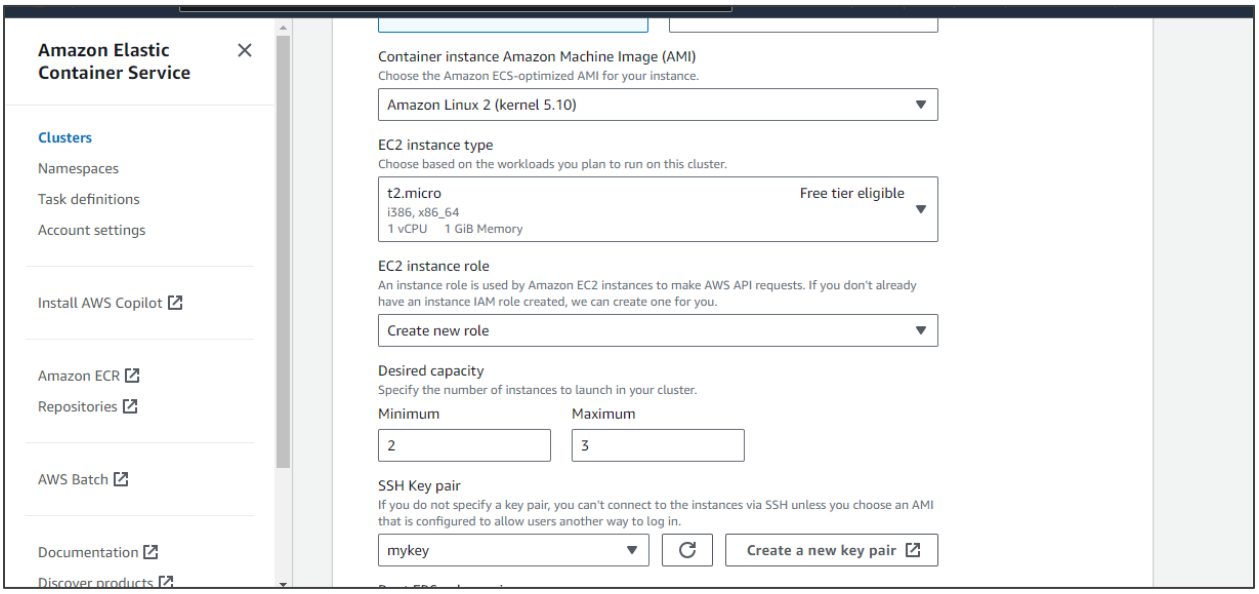
1.3 Click **Create cluster**



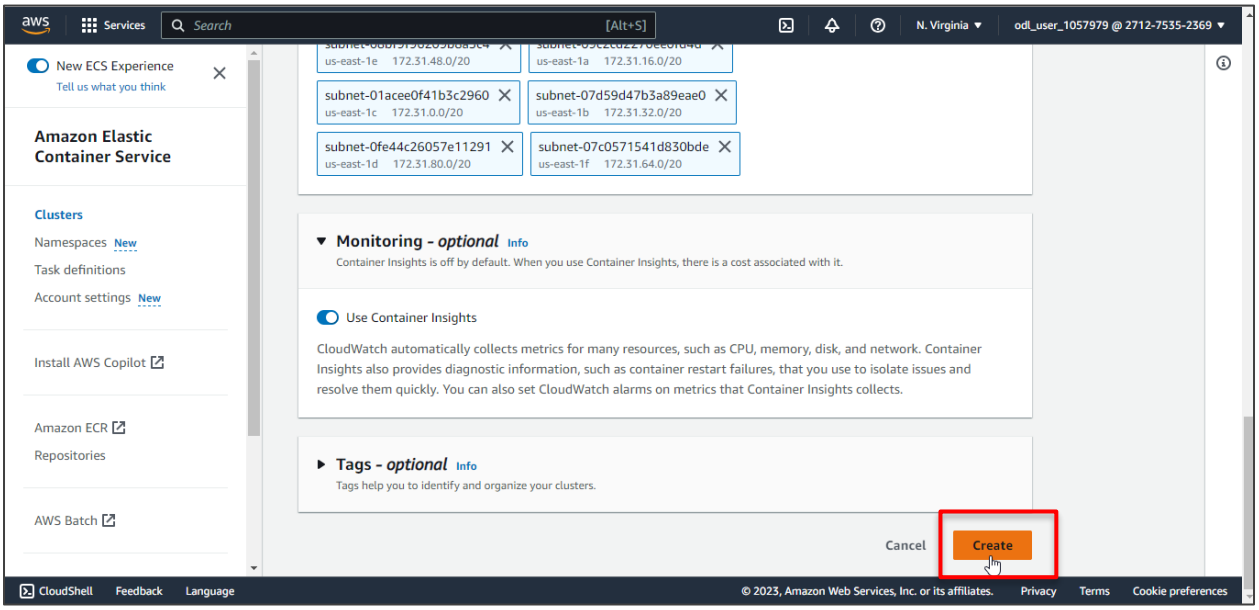
1.4 Create an ECS cluster per the settings shown in the screenshots:



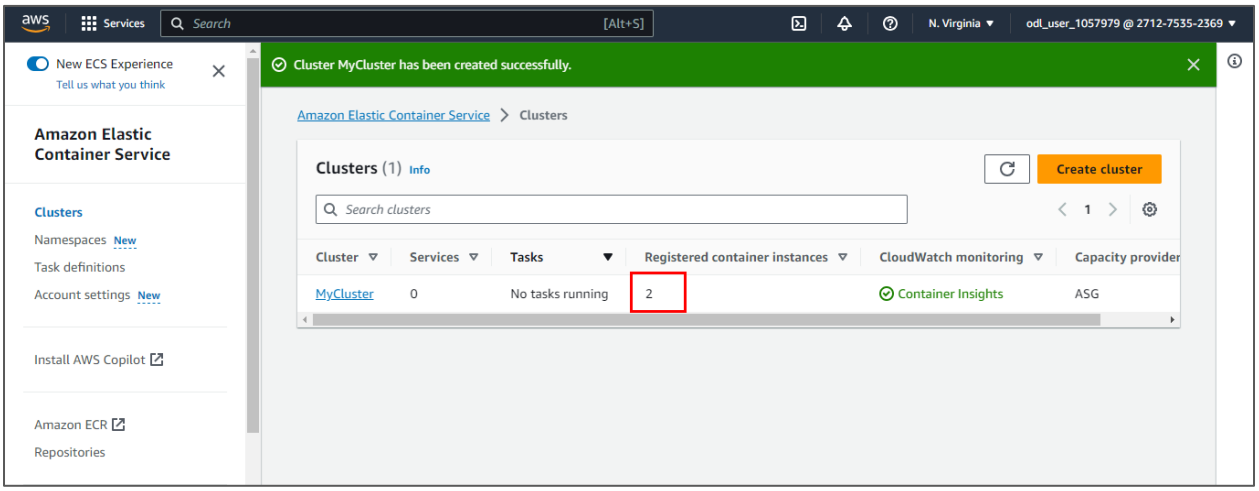




1.5 Click **Create**

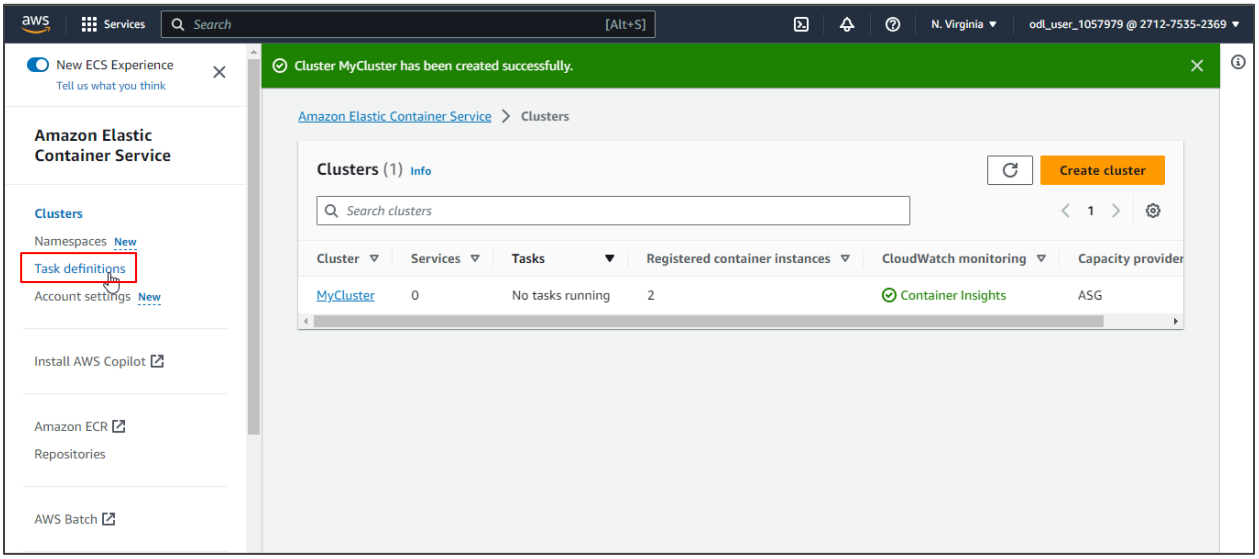


Note: Please wait a few minutes for the EC2 instances to register with the cluster.

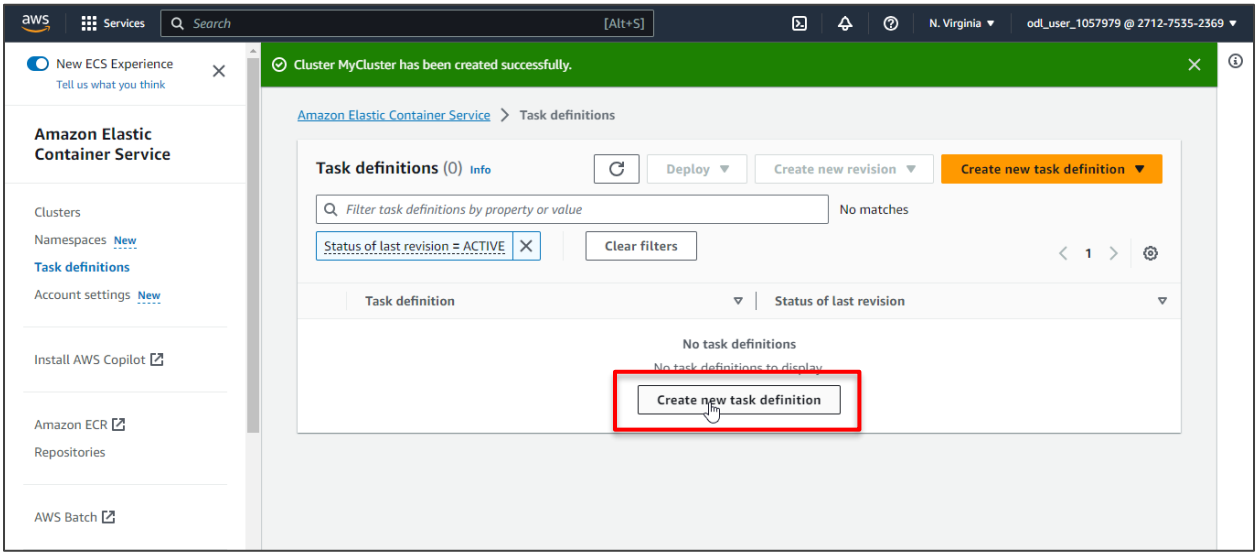


Step 2: Create a task definition

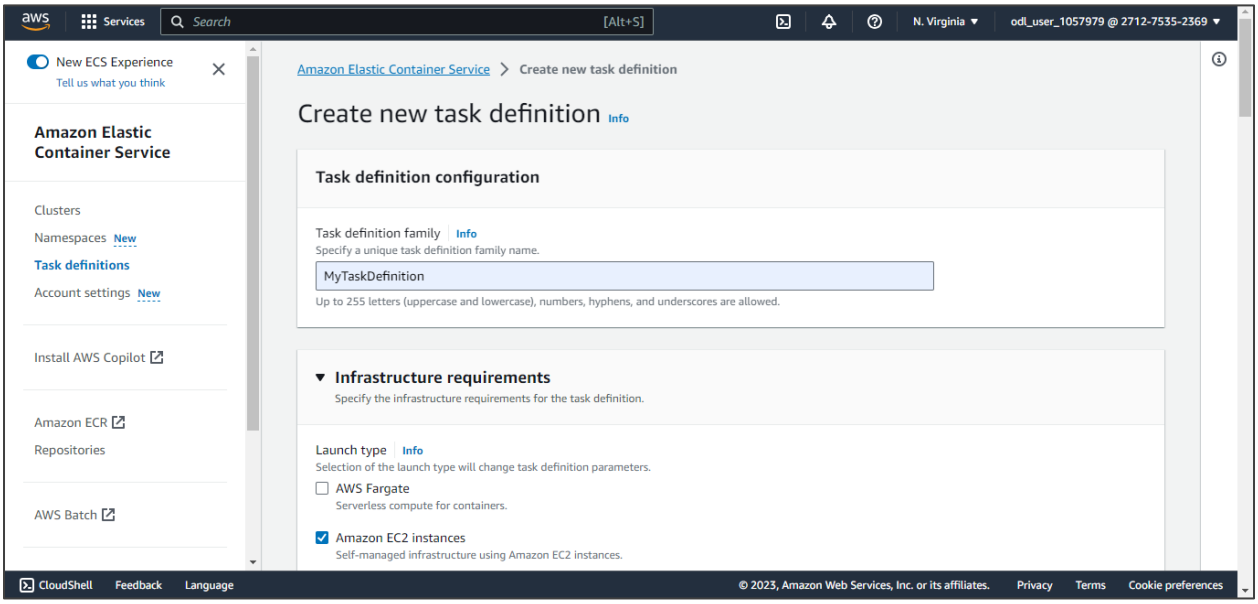
2.1 In the navigation panel, click **Task definitions**

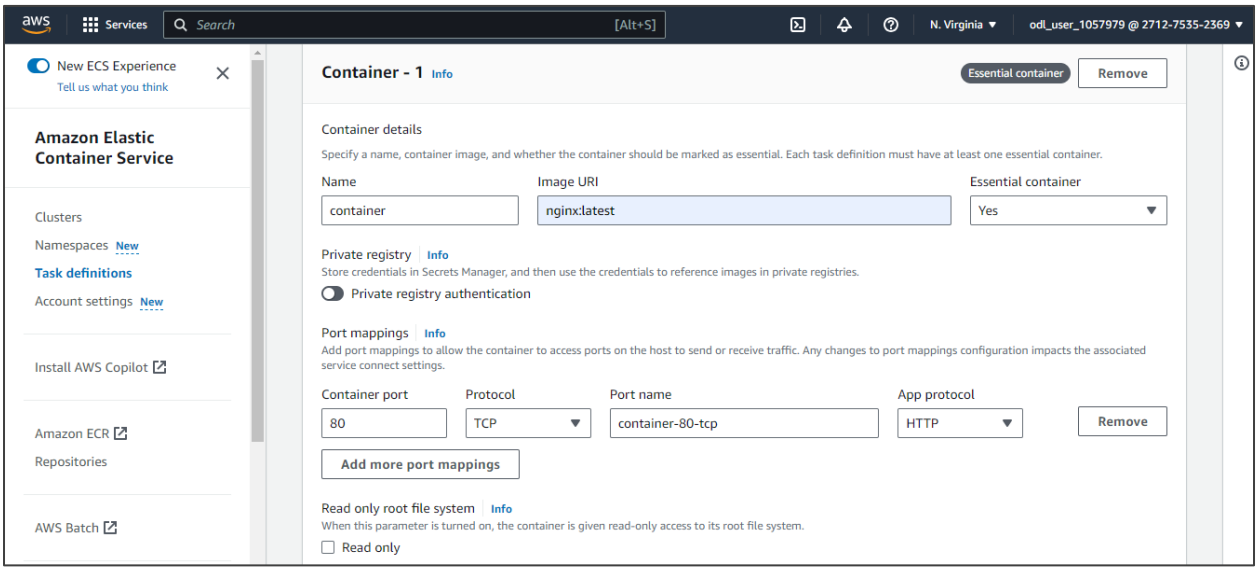


2.2 Click **Create new task definition**

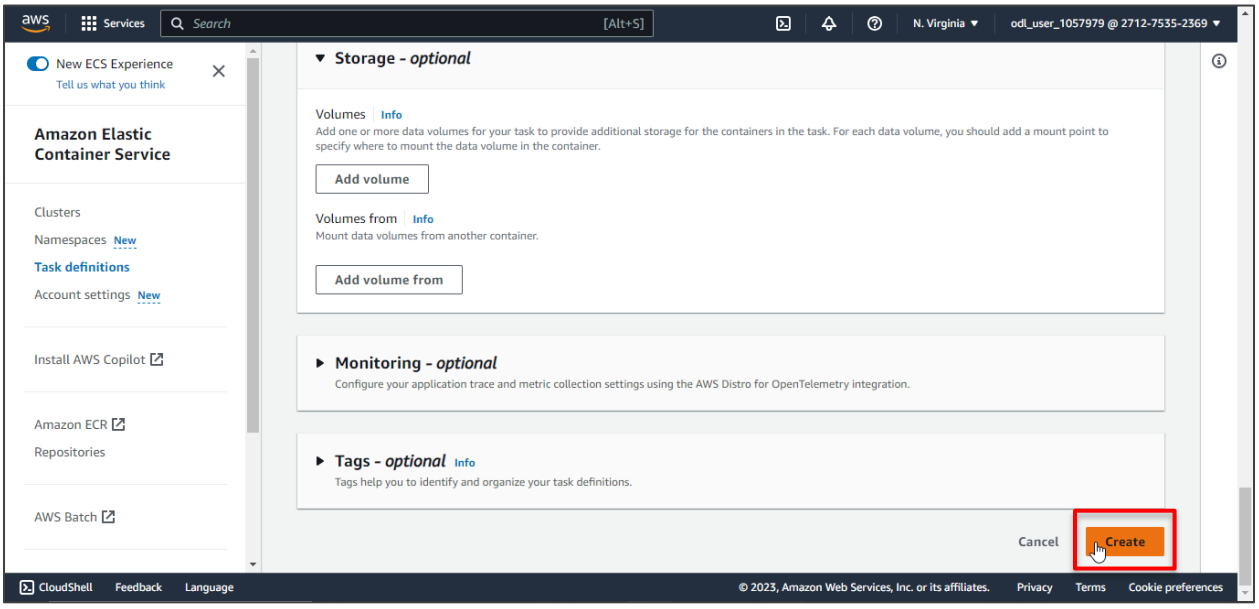


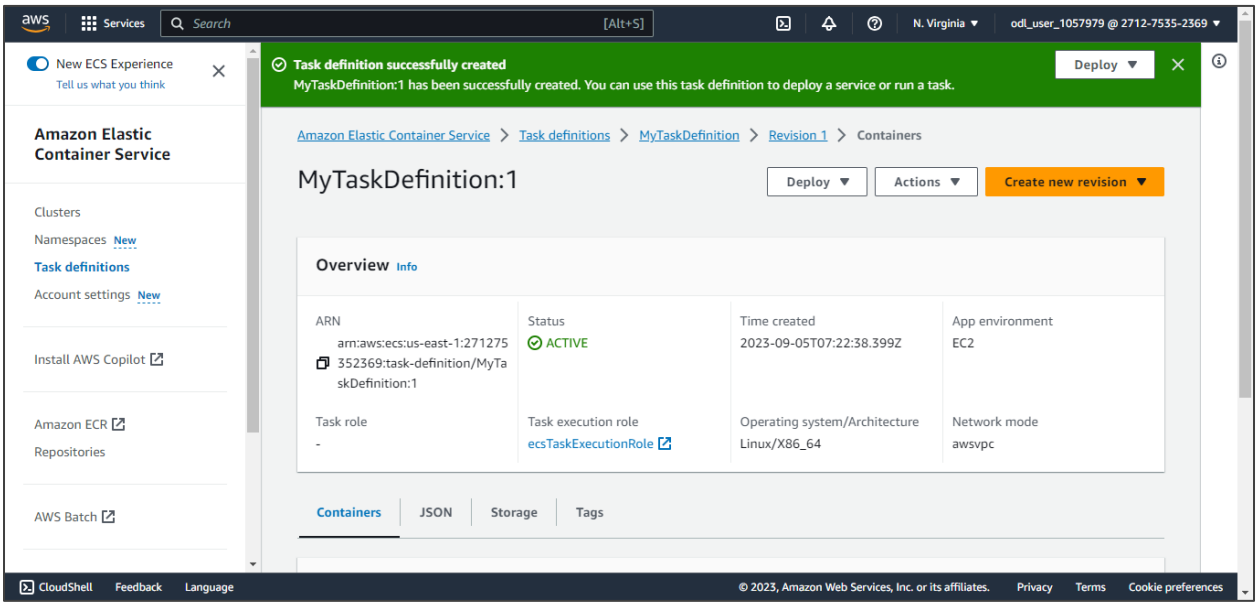
2.3 Create the task definition per the settings shown in the screenshots:





2.4 Click **Create**

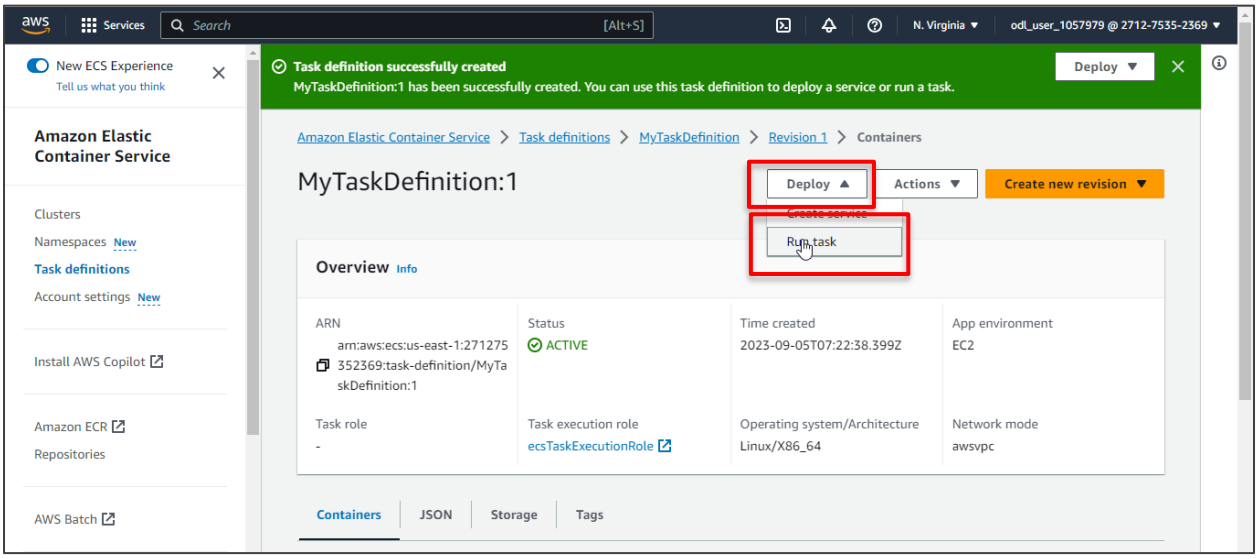




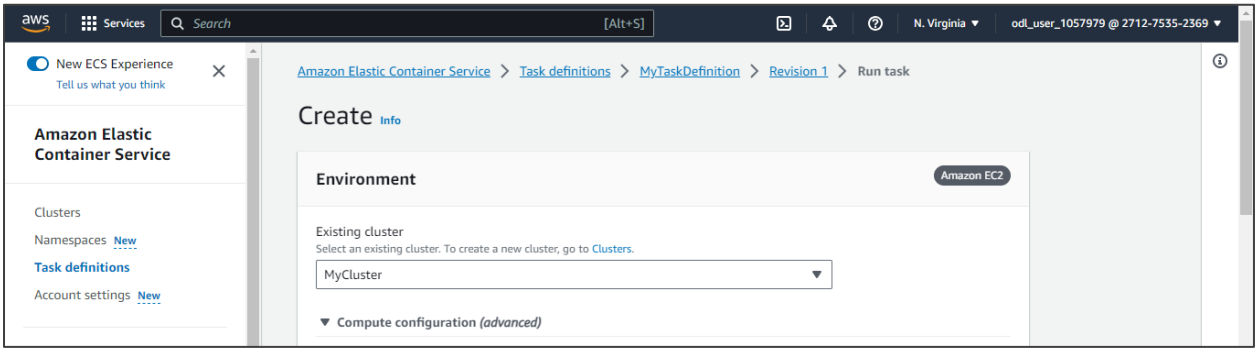
The task definition has been created successfully.

Step 3: Run the task definition on the cluster

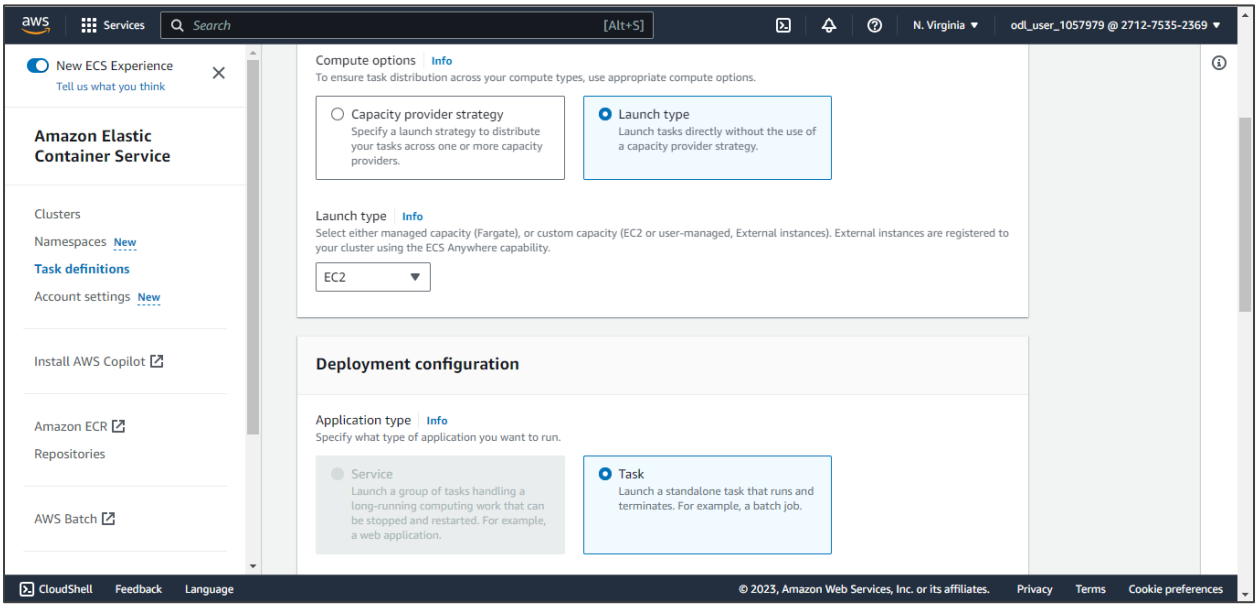
3.1 Select **Deploy** and click **Run task** to run the task definition

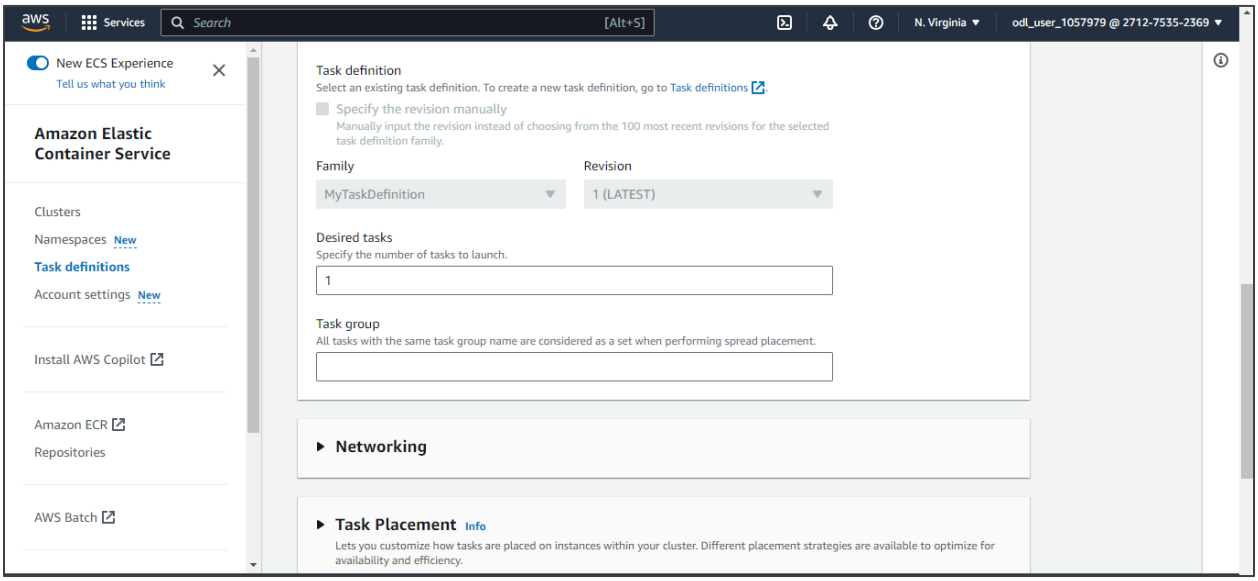


3.2 Choose the cluster that you created in the previous step

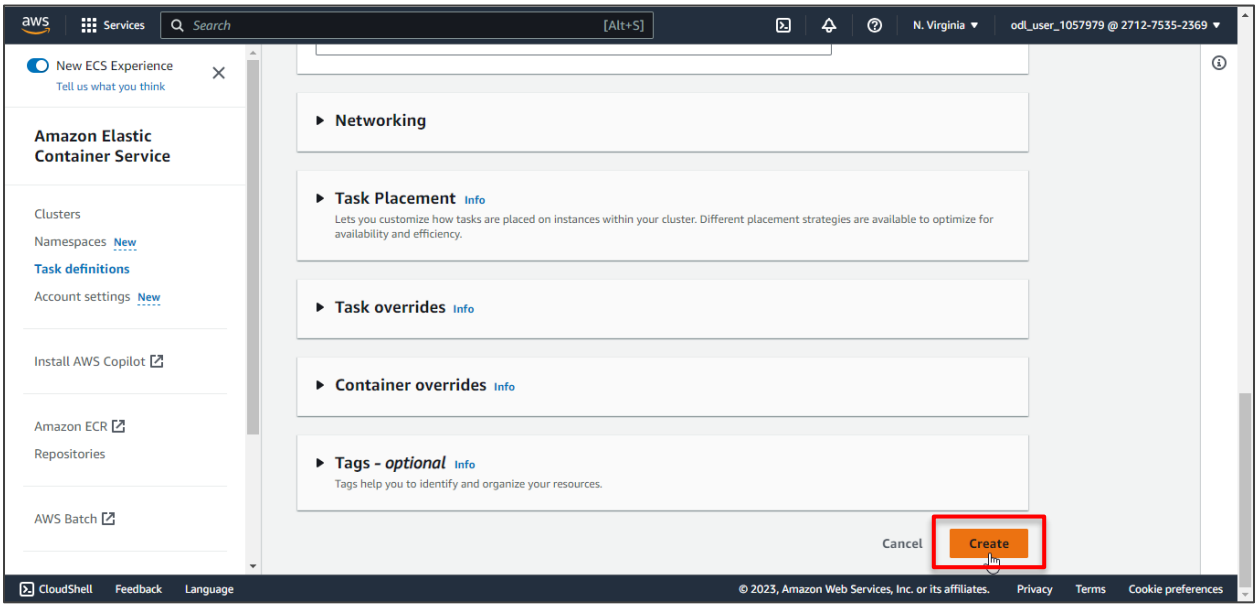


3.3 Configure the other settings per the screenshots:

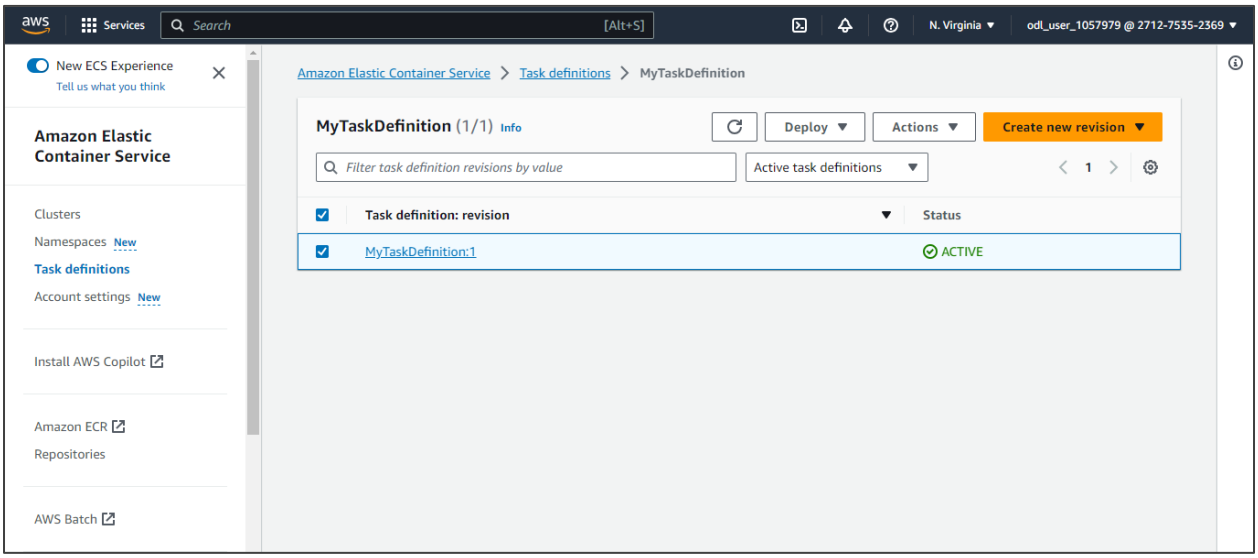


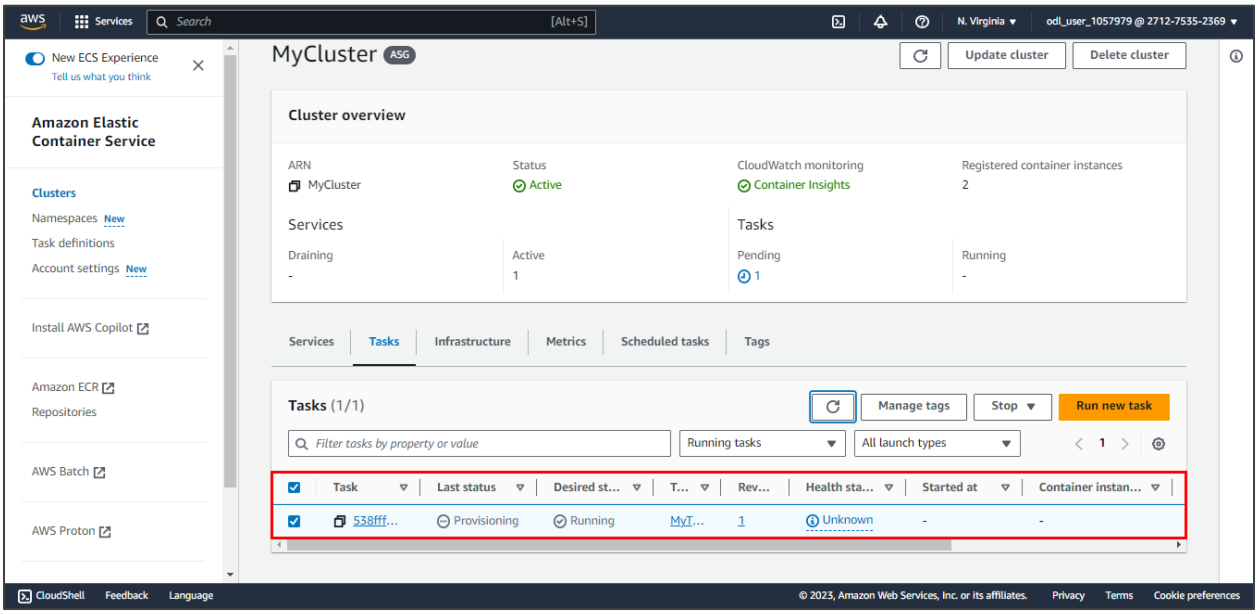


3.4 Choose **Create**



The task definition has been successfully executed.





A task has been added to **MyCluster** and triggered successfully as shown in the screenshot.

By following these steps, you have successfully created an ECS Cluster with Auto Scaling, defined and executed tasks within it.