AUTO-SCALING PRAC

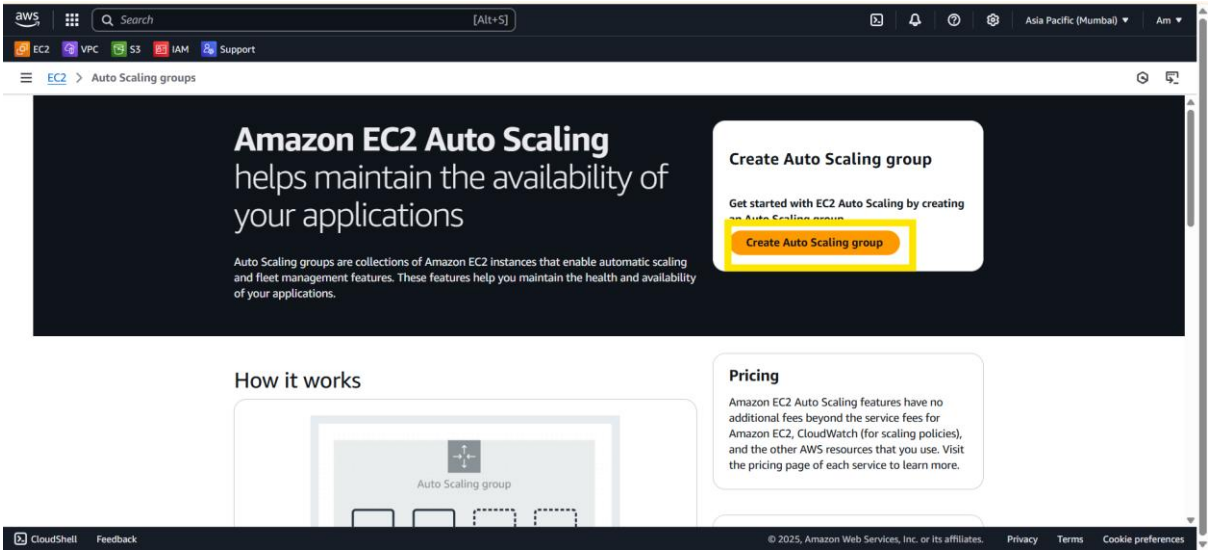
Define Auto-scaling:

system that automatically adjusts your cloud resources (like servers or databases) to match your application's needs.

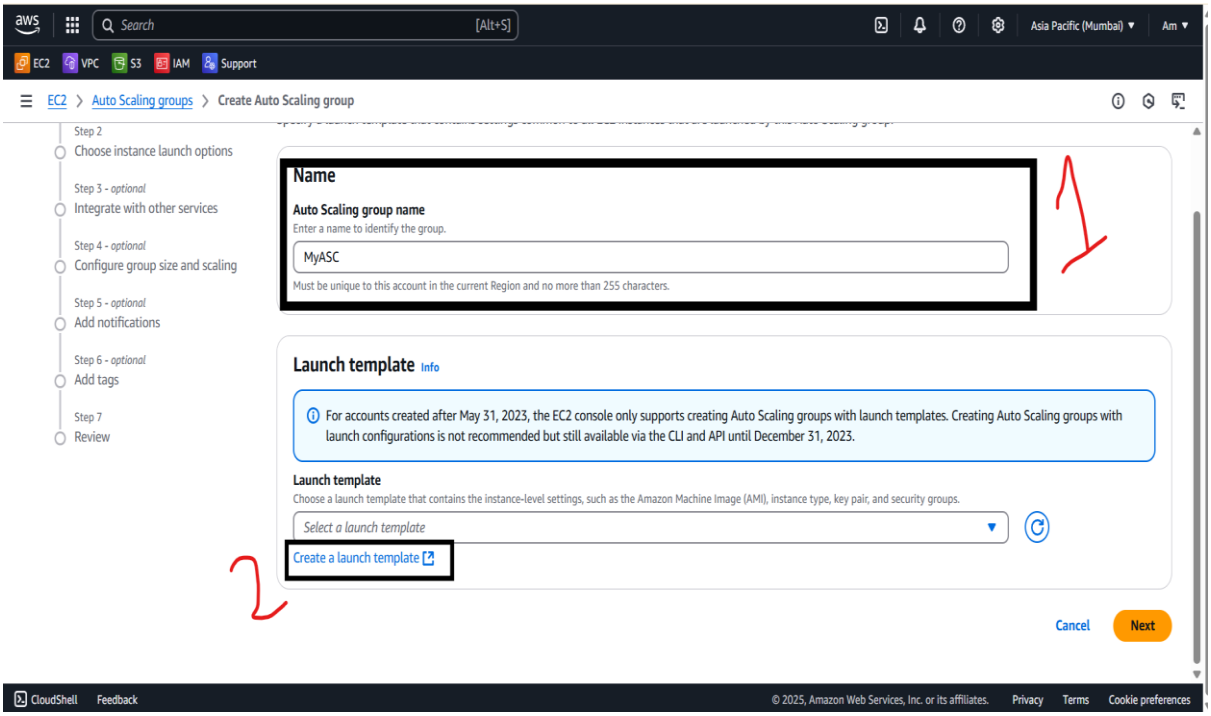
Why use Auto-scaling:

automatically adjusts your resources to match the demands of your application, ensuring optimal performance and cost efficiency.

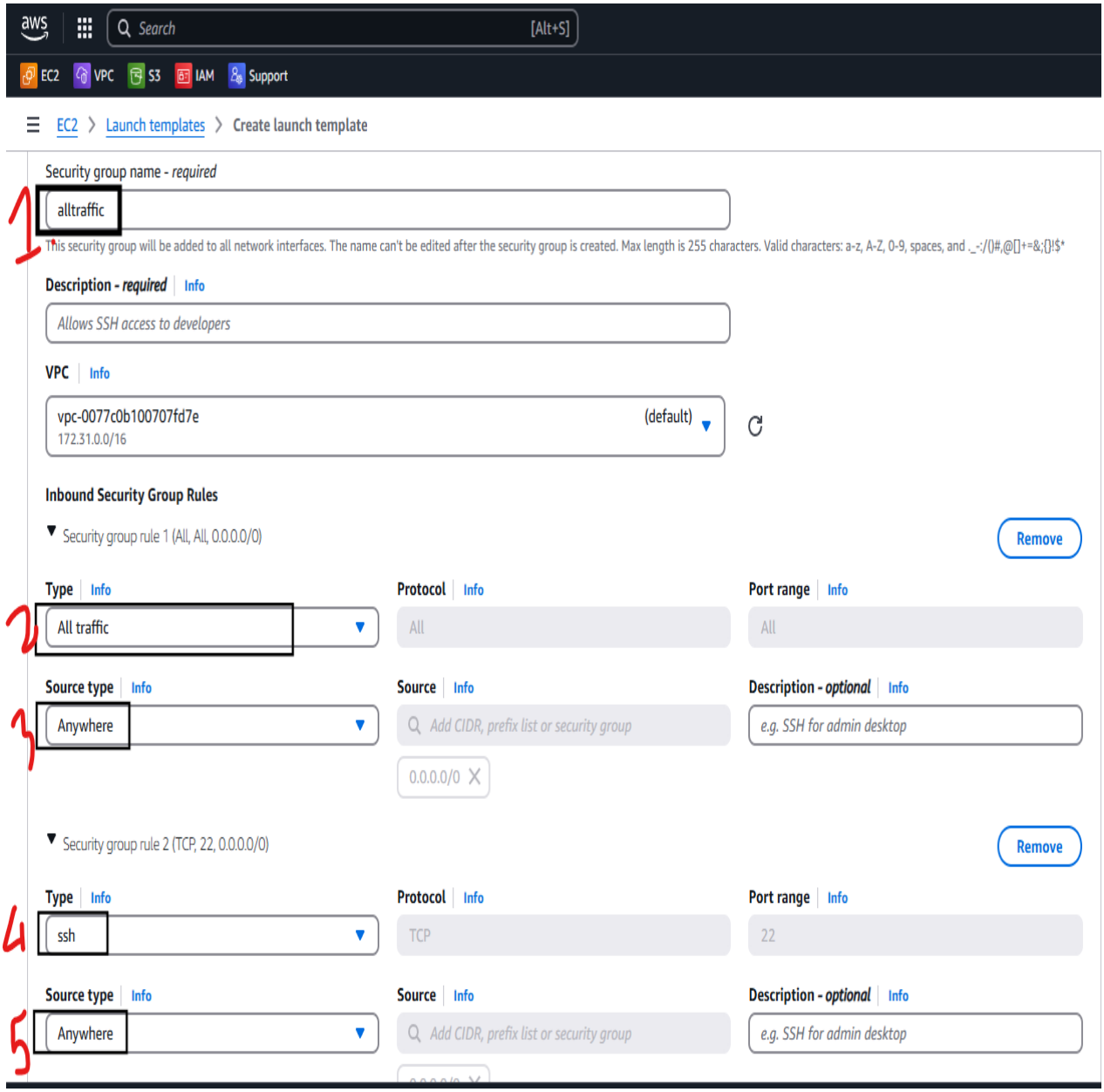
Click on create autoscaling group:-



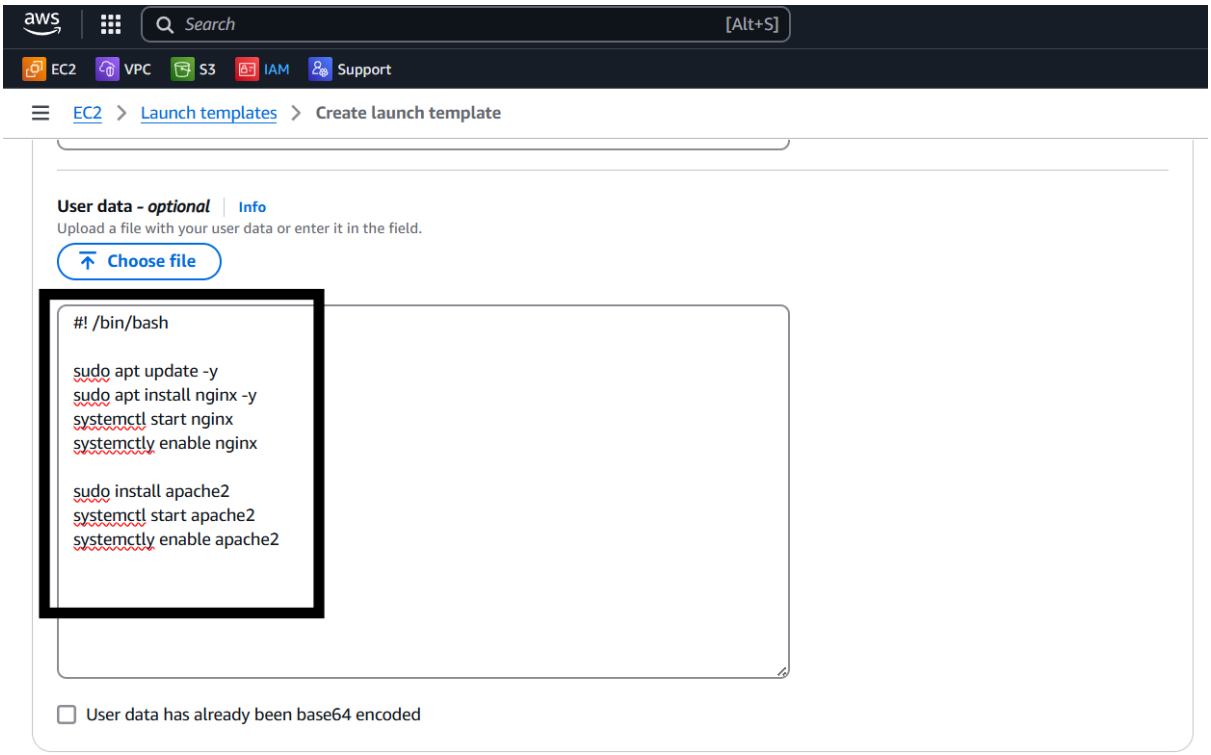
STEP 1: name it and create launch template to launch instance through launch template.



STEP 2: all other things will be same as we create a launch template I have only added a new SG name all traffic, other details can be seen in SC.

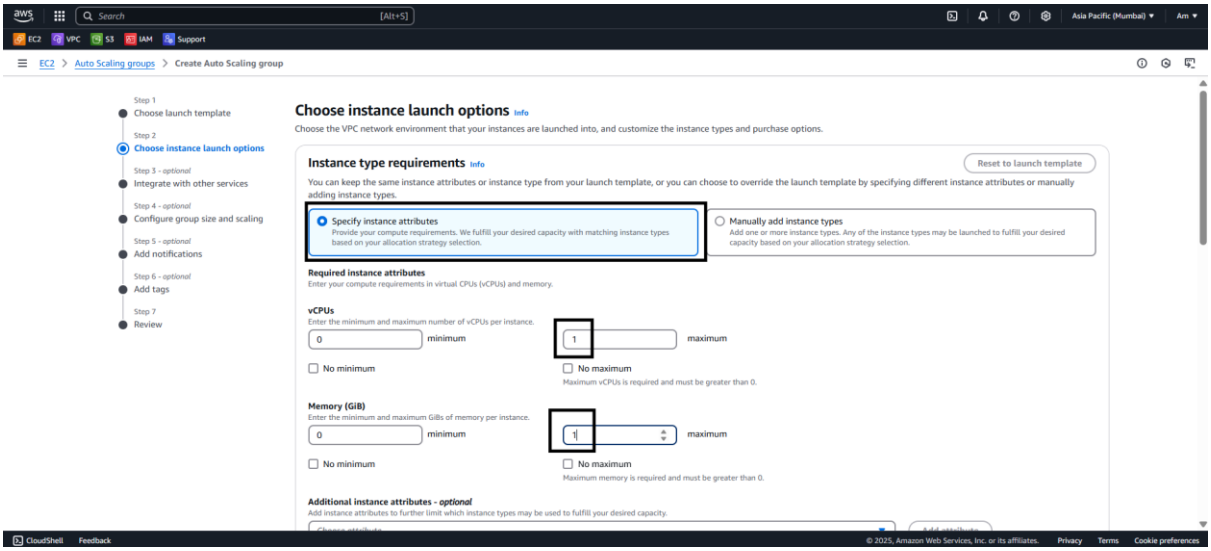


STEP 3:Adding script to install nginx and apache.

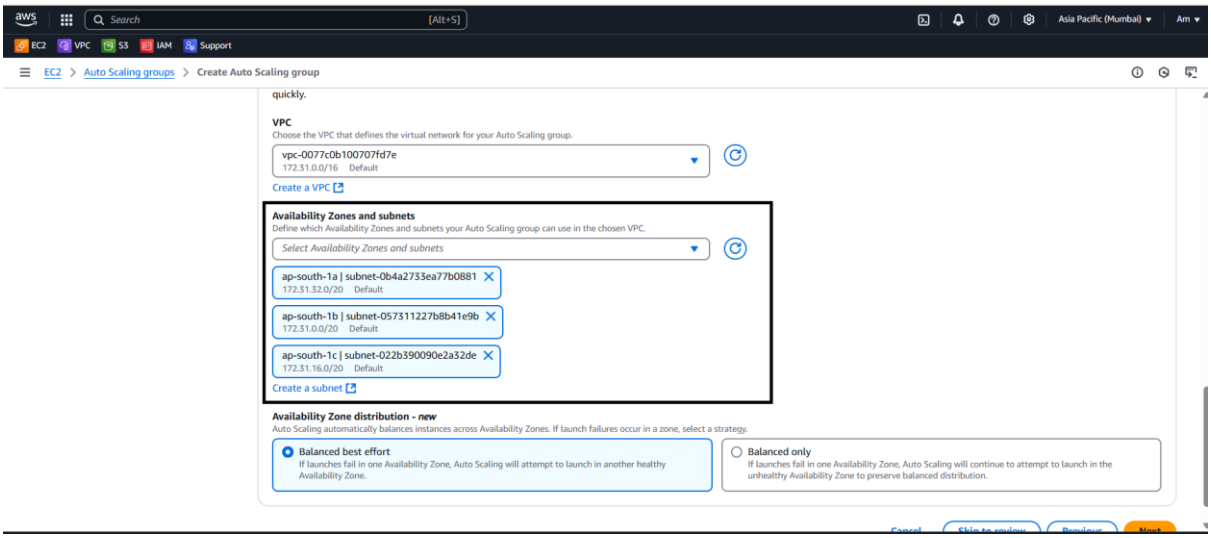


USING THIS SCRIPT SO THAT, WHEN THE INSTANCE IS LAUNCHED WE CAN DIRECTLY ACCESS BOTH THESE PROXY SERVERS BY USING PUBLIC IP OF THE INSTANCE.

STEP 4: Choosing the instance requirements of how many instances to be launched at the initial start of the instance.



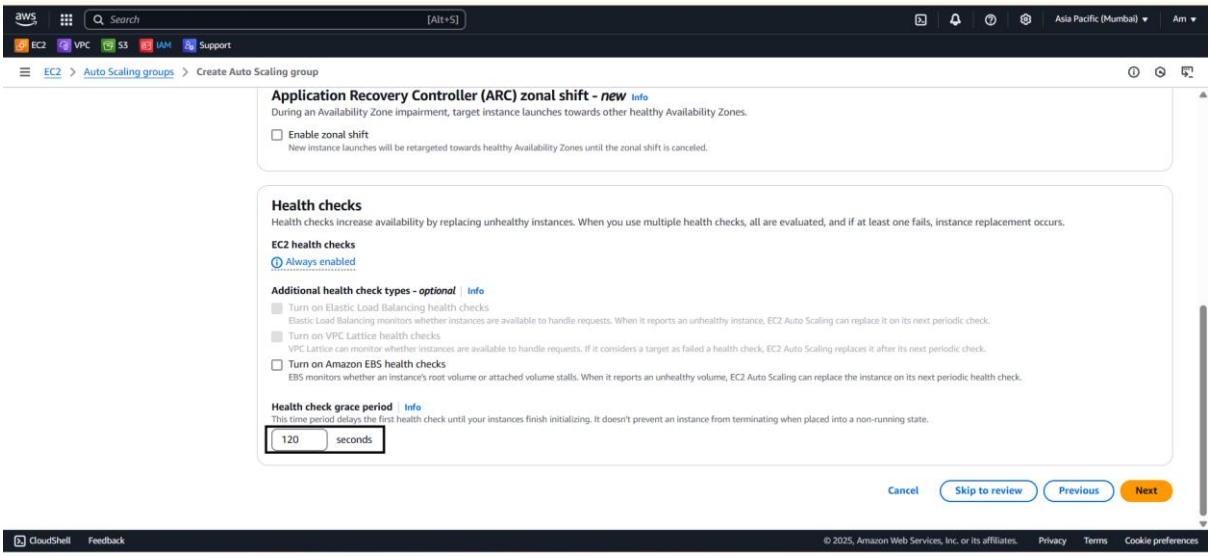
STEP 5: NETWORK CONFIG—Selecting all the AZ's



STEP 6: In health checks we give 120 seconds.

Explain why 120 seconds?

- 120 seconds mean the servers we have taken nginx and apache will refresh after 120 seconds.
- If the load increases on one primary instance, autoscaling will re-direct the traffic to other instance.



STEP 7: Set desired capacity and max. desired capacity.