AWS Auto Scaling Group (ASG) — Step-by-Step Guide

Author: Samarth Mahadik

Date: August 31, 2025

This guide documents the steps taken to create an Auto Scaling Group (ASG) in AWS. It includes creation of security groups, launch template, EC2 instance, ALB integration, creation of the ASG, adding a dynamic scaling policy, and verifying instance health/status. Screenshots captured from the AWS Console are included for each step.

Architecture overview: Client → ALB (Internet-facing) → Auto Scaling Group → EC2 instances (web servers).

Step 1. Create Security Group (web-sg)

– Go to EC2 Console → Security Groups → Create security group.

– Name: web-sg. Description: allow ssh access developer / web traffic.

– Inbound rules: SSH (22) from your IP, HTTP (80) from 0.0.0.0/0.

– Attach this SG to instances launched by ASG or to ALB as needed.

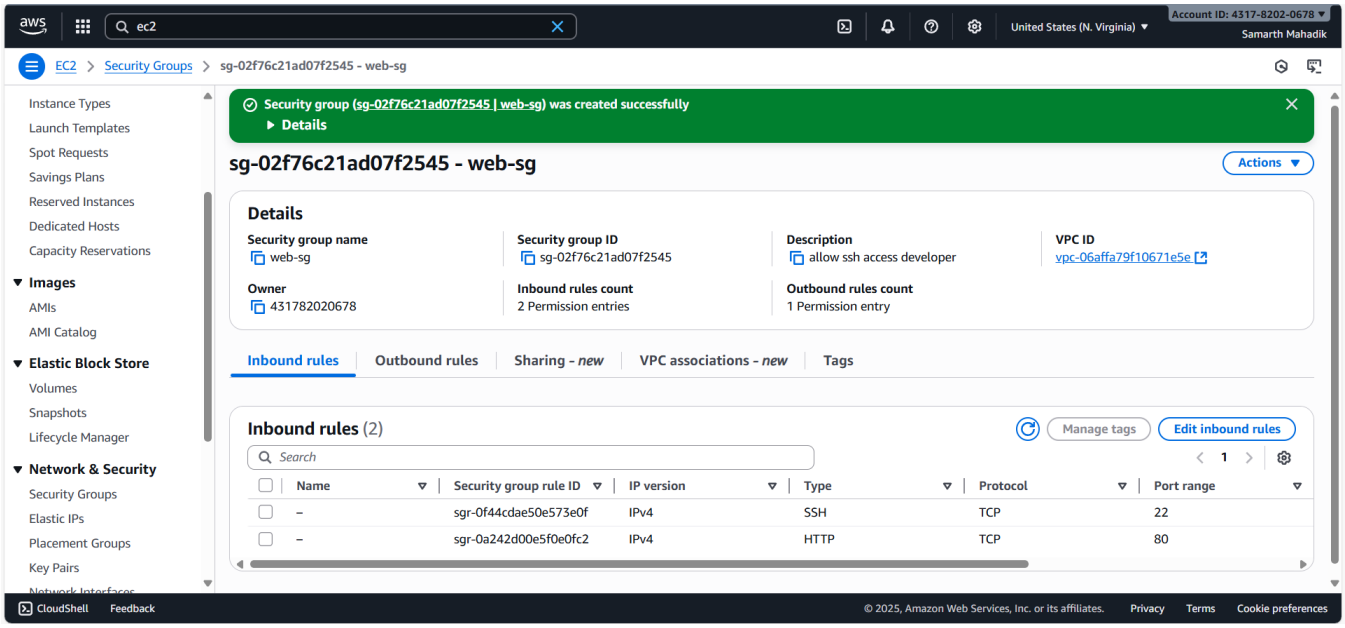

Figure 1: Create Security Group (web-sg)

Step 2. Create Launch Template

– EC2 Console → Launch Templates → Create launch template.

– Choose AMI (e.g., Amazon Linux 2), Instance type (t2.micro), Key pair (optional).

– Select Security group: web-sg. Add user-data script to auto-install web server (optional).

– Save template (example name: asg-LT).

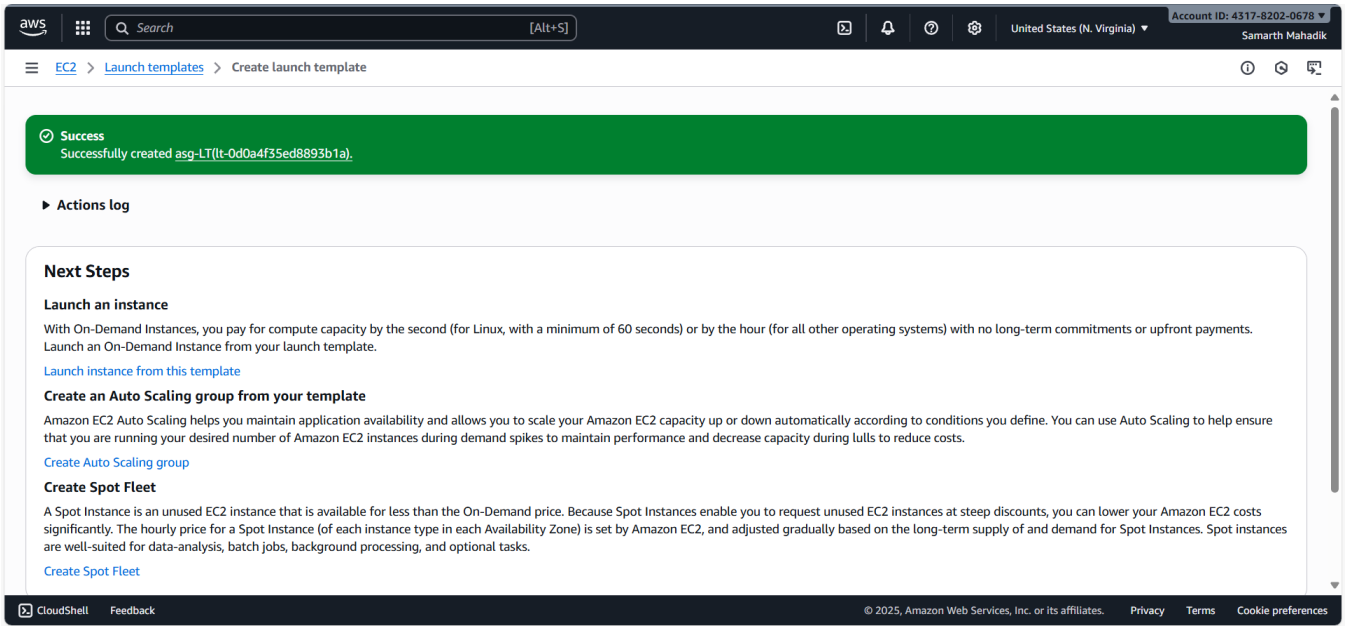

Figure 2: Create Launch Template

Step 3. Launch an EC2 Instance (verify baseline)

– Optionally launch a test instance from the template to verify that the AMI, user-data, and SG work.

– Check instance status and public/private IP. Ensure web server page loads on port 80.

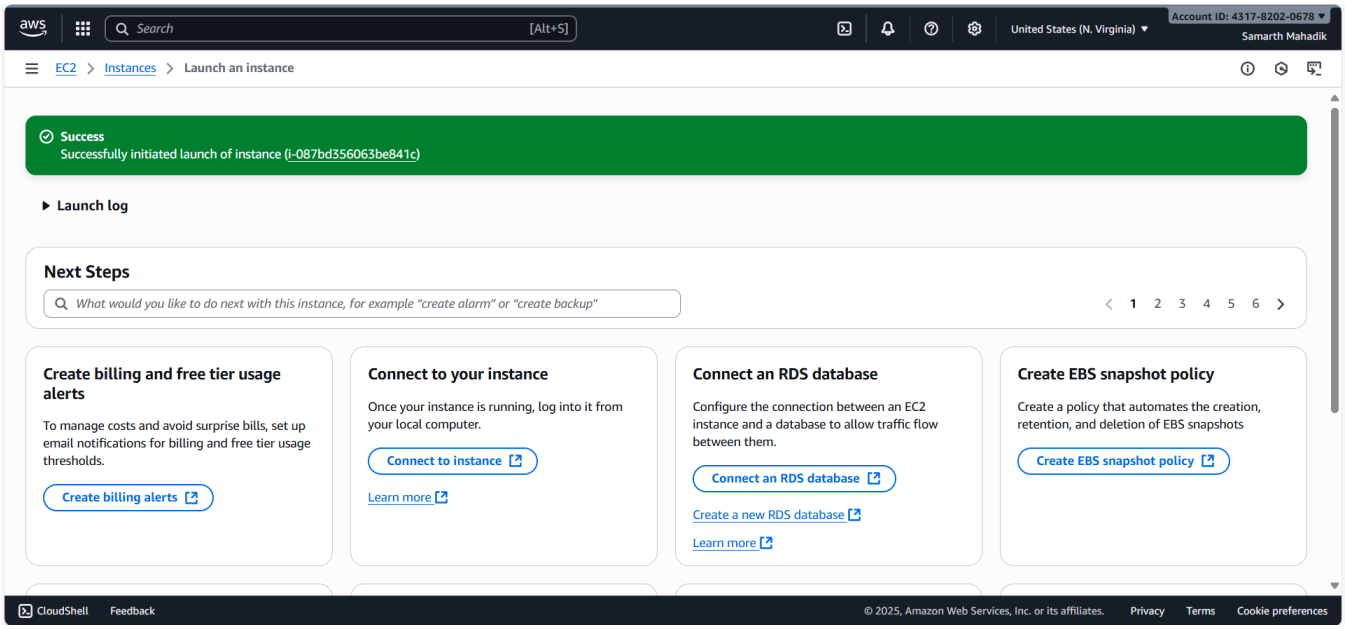

Figure 3: Launch an EC2 Instance (verify baseline)

Step 4. Create ALB (for ASG traffic)

– EC2 Console → Load Balancers → Create Application Load Balancer.

– Name: my-asg-alb (example). Scheme: Internet-facing, IP type: IPv4.

– Select at least two public subnets. Attach appropriate security group (allow HTTP/HTTPS).

– Create and note the ALB DNS name.

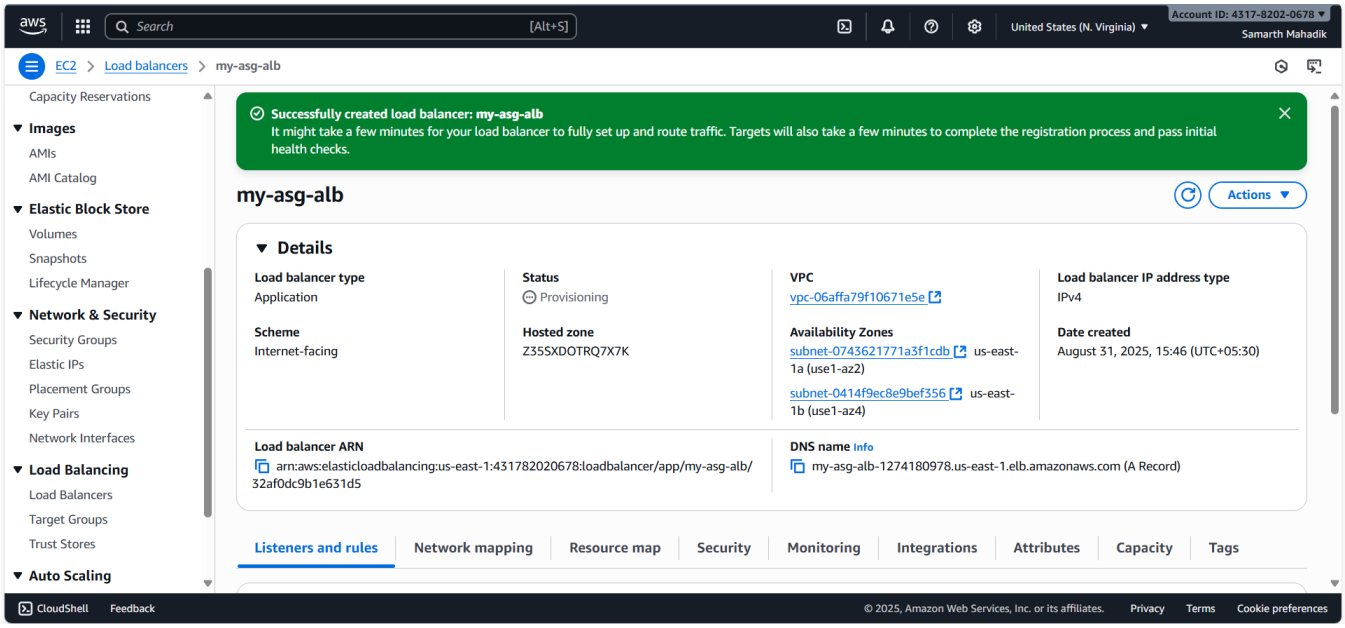

Figure 4: Create ALB (for ASG traffic)

Step 5. Create Auto Scaling Group (ASG)

– EC2 Console → Auto Scaling → Create Auto Scaling group.

– Choose Launch template: asg-LT. Name: ec2-asg (example).

– Set Desired capacity: 1, Min: 1, Max: 3 (example).

– Attach to ALB target group or configure load balancer integration.

– Create the ASG.

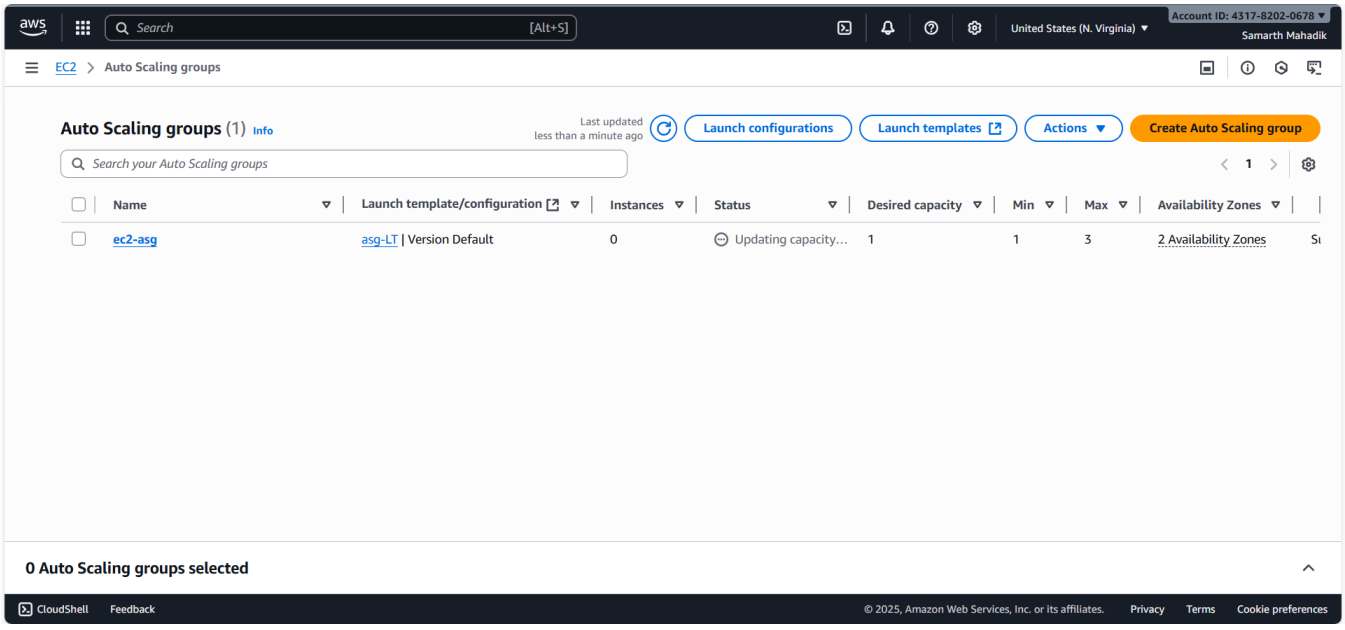

Figure 5: Create Auto Scaling Group (ASG)

Step 6. Create Dynamic (Automatic) Scaling Policy

– In the ASG → Automatic scaling tab → Create dynamic scaling policy.

– Choose policy type (e.g., Target tracking or Step scaling).

– Configure target metric (e.g., average CPU utilization 50%) or CloudWatch alarms.

– Save the scaling policy to enable ASG to scale automatically based on load.

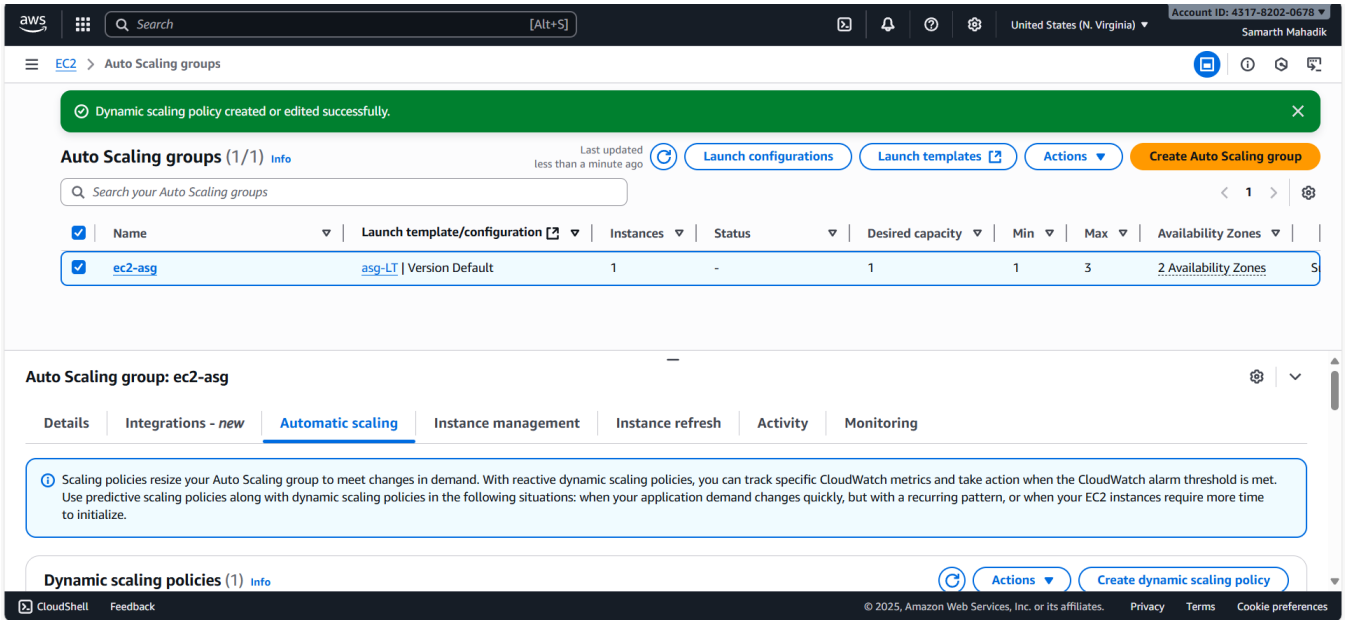

Figure 6: Create Dynamic (Automatic) Scaling Policy

Step 7. Verify ASG Activity & Instance Health

– ASG → Activity / Instances: check events and instance lifecycle actions.

– Verify instances register to the ALB target group and pass health checks.

– Confirm desired capacity matches running instances and status is healthy.

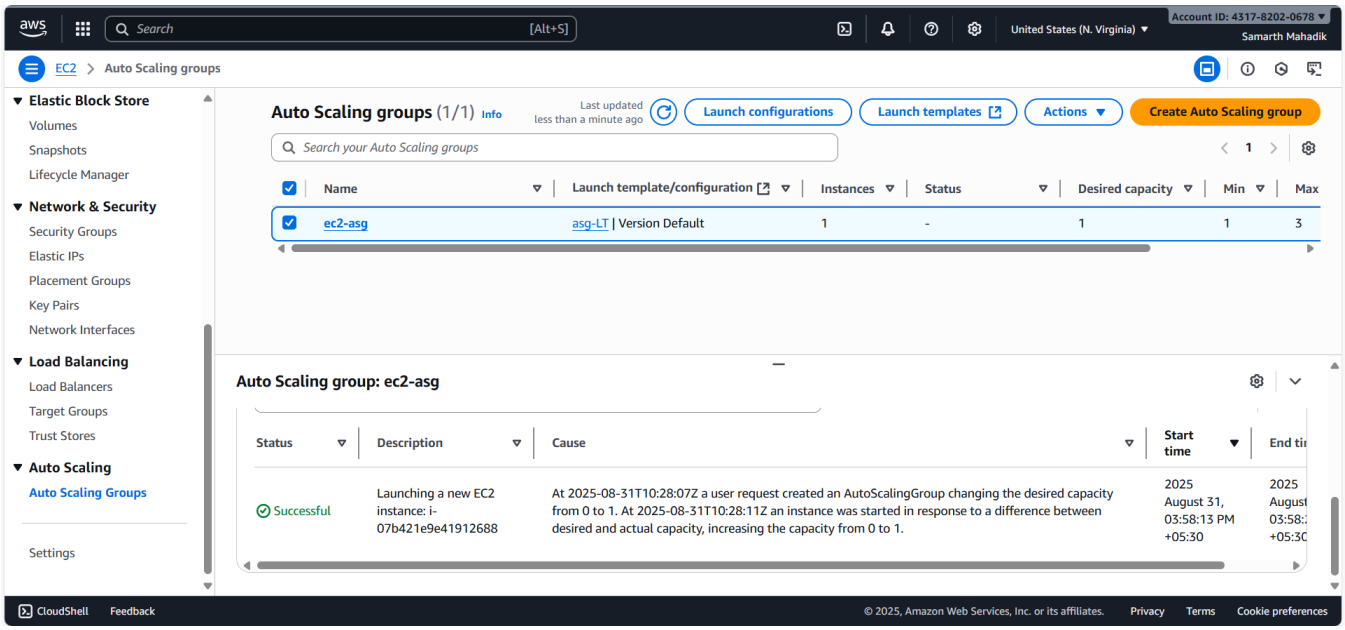

Figure 7: Verify ASG Activity & Instance Health

Next steps & hardening

– Next steps & hardening:

– Configure CloudWatch alarms and notifications for scaling events.

– Use multiple Availability Zones for high availability.

– Add lifecycle hooks if you need custom actions during instance launch/terminate.

– Monitor costs and set budget alerts.