**AutoScaling**

**Benefits:**

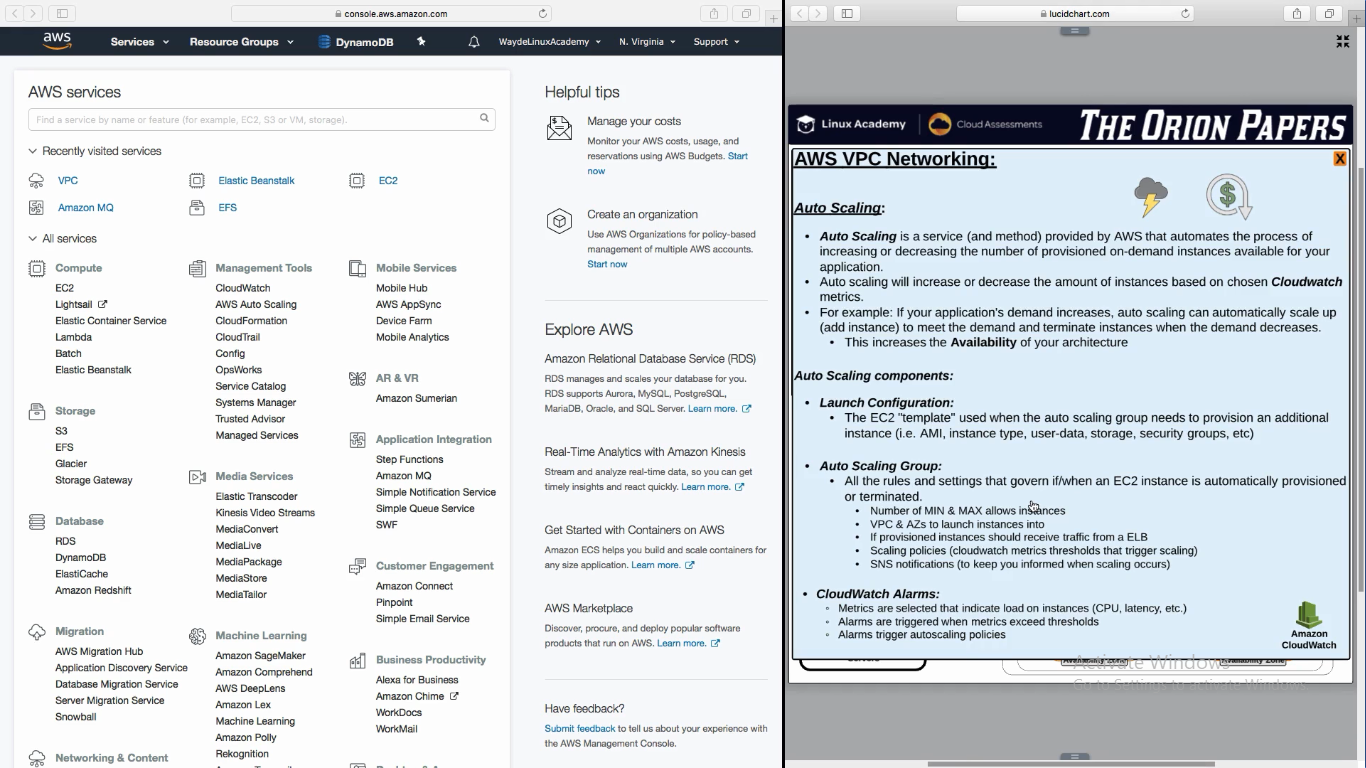
1. Shrink or expand the instances based on load
2. Cost effective as we spin up instance only on demand

CloudWatch used to do monitor of latency or cpu threshold met.

By default above two mechanism is included in cloudwatch except cloud watch alarm

**AutoScaling Components**

1. Launch Configuration
2. Auto scale group
   1. Set ec2 maximum instance above what we expect
3. Cloud Watch Alarm
   1. When this alarm is triggered it will invoke scaling policy which will say what to do



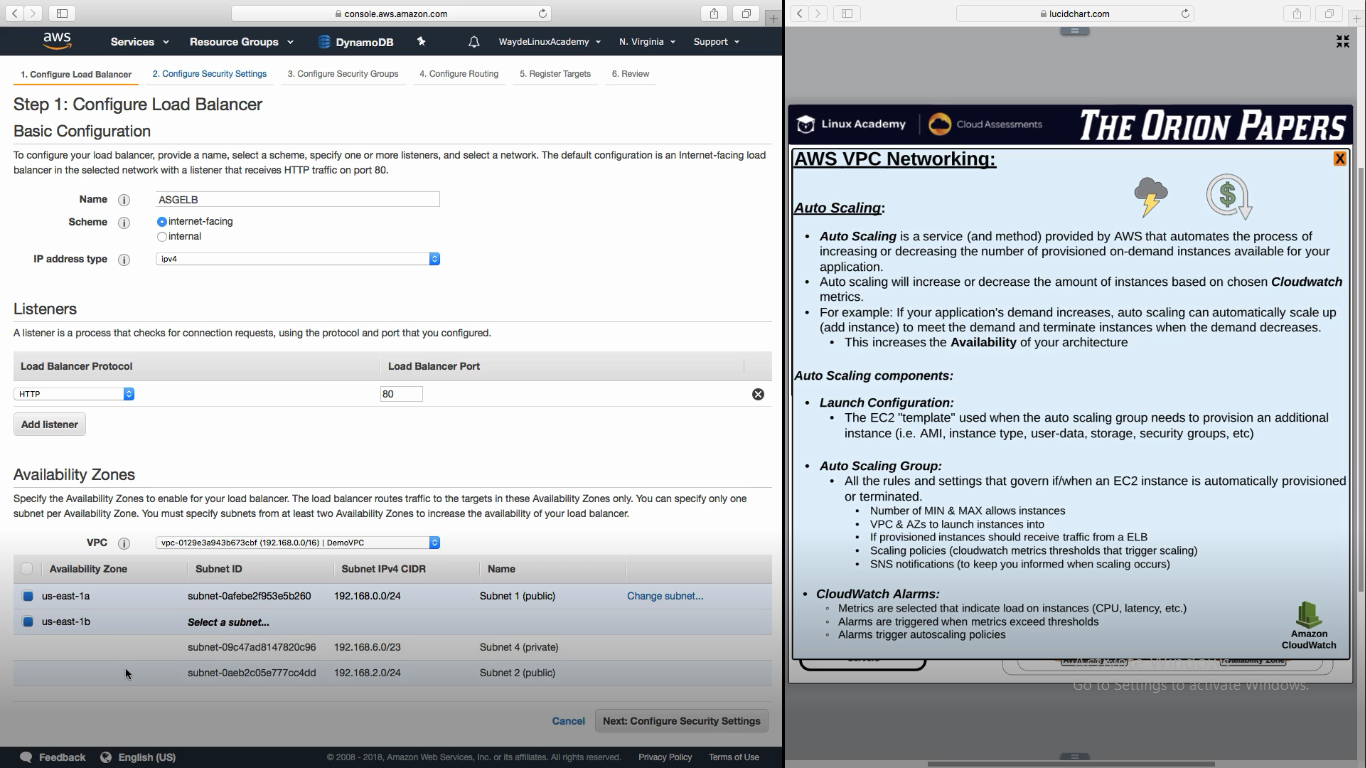
**CloudWatch**

1. Create custom metrics by writing a script
   1. Eg: Number of users logged in

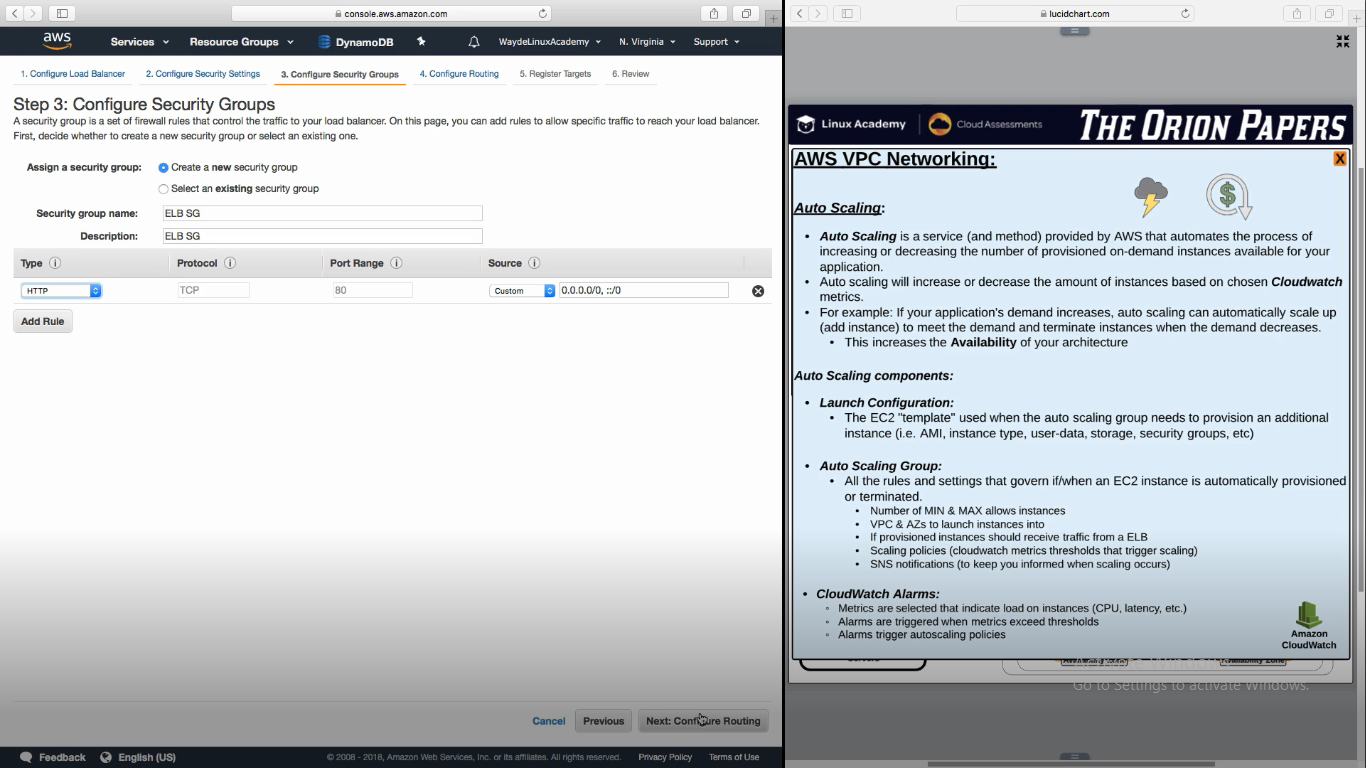
**Auto Scaling Steps:**

1. Create Load Balancer
2. Create Load Balancer Security group
3. Create Target group for load balancer to send the traffic
   1. In advanced section, give threshold level to check if instance is healthy or not, in what frequency it have to check etc
4. Target (here is where we say which instance to be spinned). We will do it in launch configuration in AutoScalling

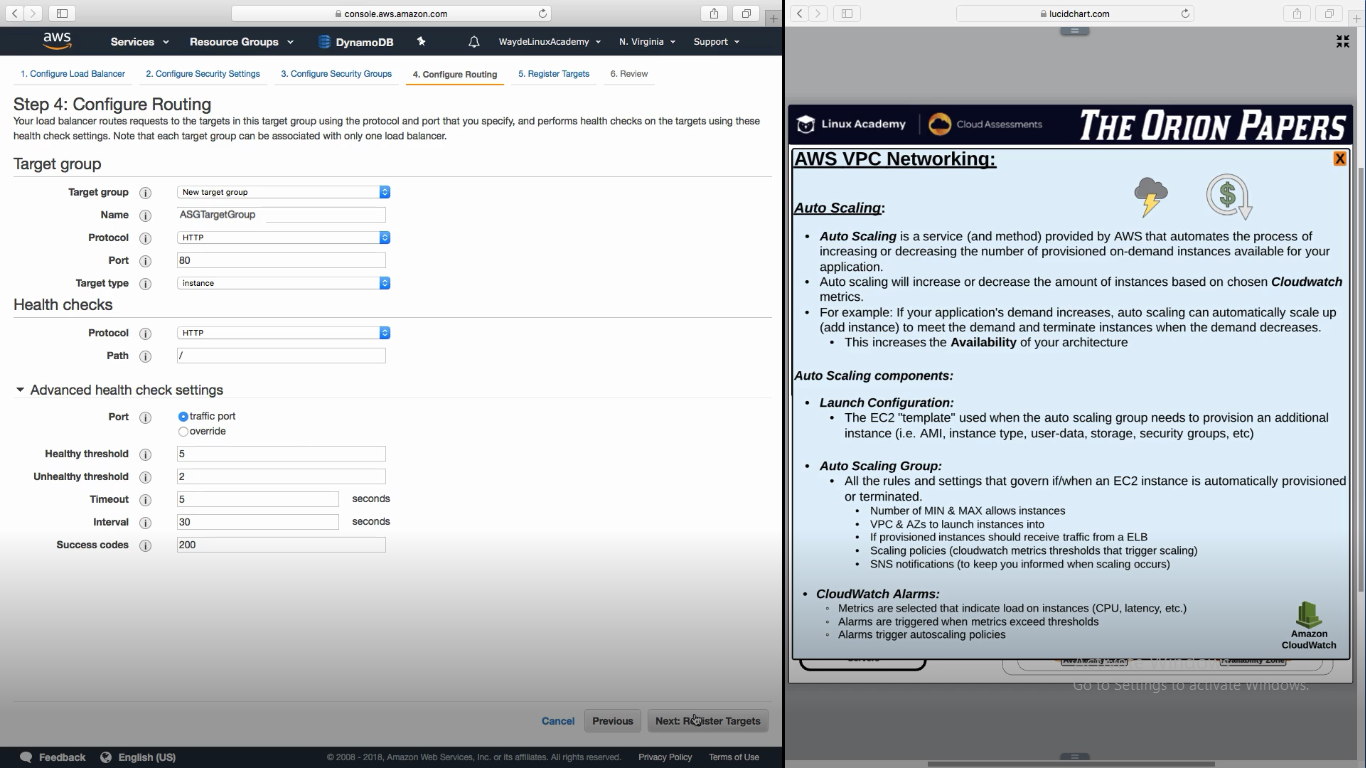
1.



2.



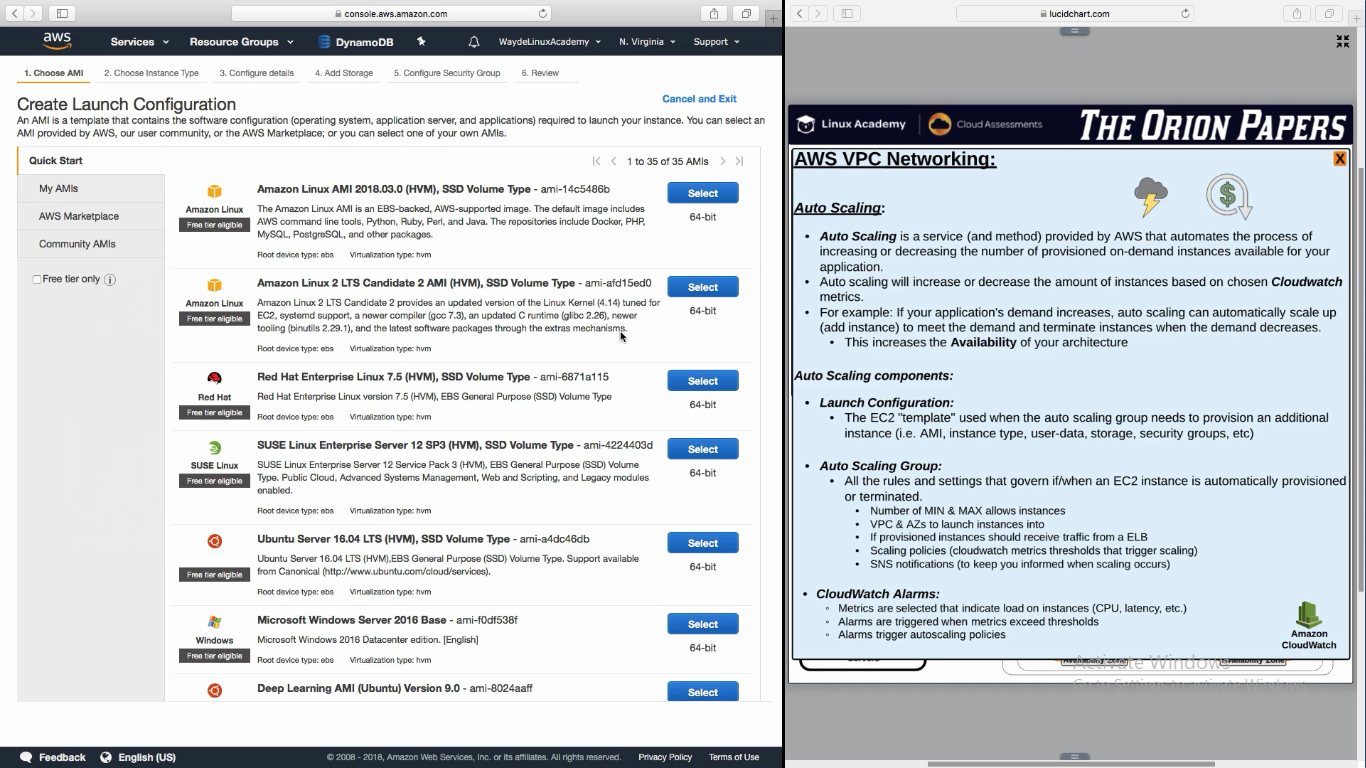
3.



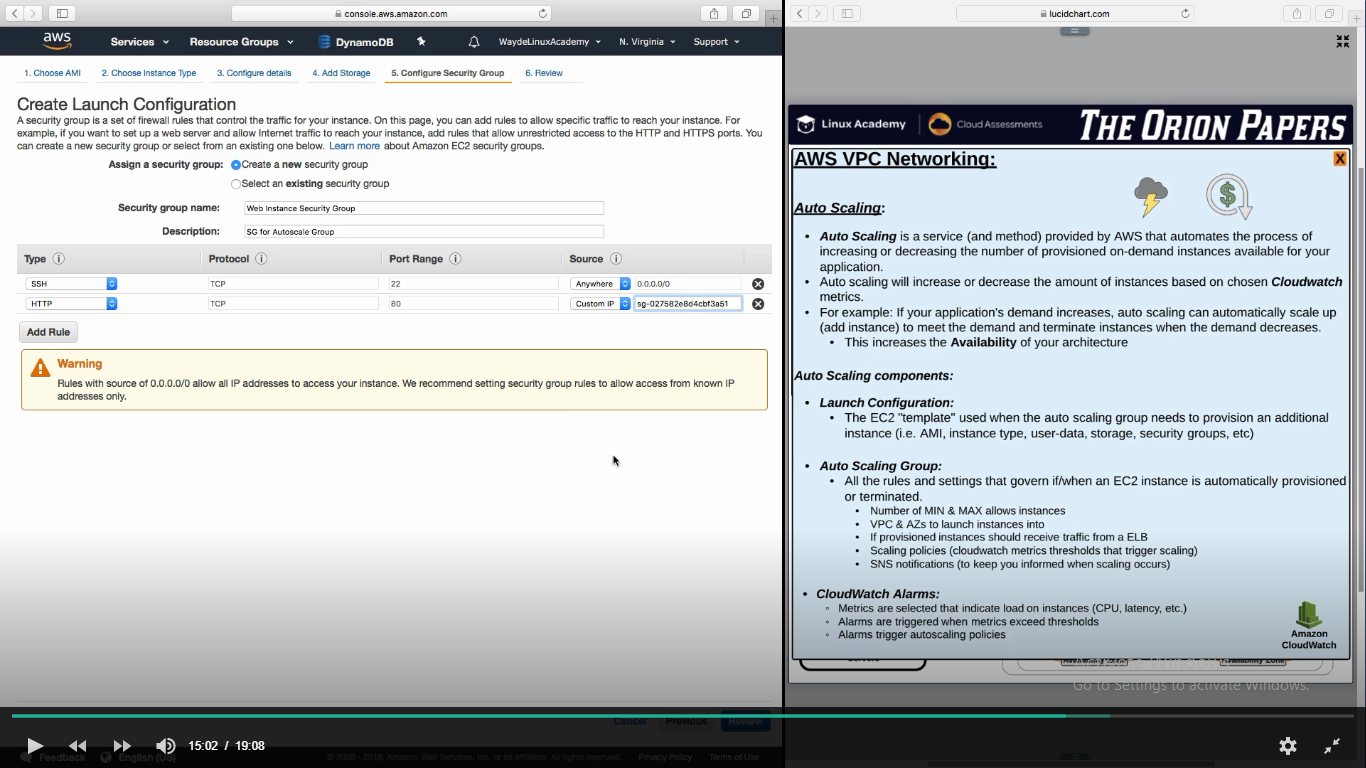
**Create AutoScaling Group**

1. Launch Configuration
2. In security group configure http so that it will accept connection only from load balancer
3. Select how many instance we need, attach it to vpc and select subnet

1.



2.



3.

