Scenario –Using AWS to autoscale webserver layer for Web and API requests when unexpected load is encountered, in addition to not using excessive capacity during the quieter periods
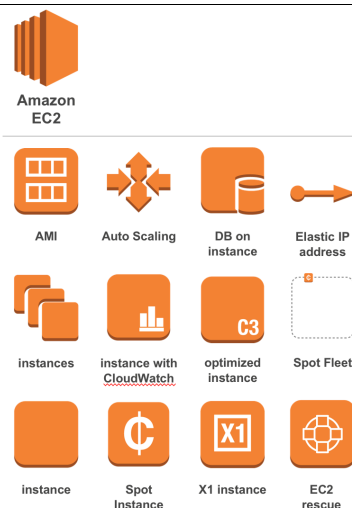
- Standard instance size is t2.medium
- Minimum instances are 2
- Maximum instances are 10
- Cloudwatch Alarm – Healthy Host Count < 2 then add 2 instances
- Scale Group Size – Average CPU Utilisation is 70% add an instance, but with a 3-minute warm up time to take into account spikes

Scenario – Time based scaling of webserver layer for when large campaigns are planned with enhanced capacity in place for certain time periods

- Standard instance size is t2.medium
- Maximum instances are 10
- Recurrence is set to once, these will need to be configured each time.
- Set the specified start time in UTC
- The scaling will then need to be reset once the campaign is over

Elastic Load Balancing*

Classic Load Balancer

Application Load Balancer

Amazon EC2

AMI	Auto Scaling	DB on instance	Elastic IP address
instances	instance with CloudWatch	optimized instance	Spot Fleet
instance	Spot Instance	X1 instance	EC2 rescue

Both scenarios will use the application load balancer, with the relevant security groups

Representing instances in EC2 and scaling are as follows