Setting up Service on AWS that will reply back to HTTP GET request

I chose ECS to accomplish the above task.

I've followed below steps to configure the underlying infrastructure for a container cluster:

* Provision VPC, subnet, routing table and security group.
* Set up IAM roles and instance profile.
* Set up an ALB.
* Set up an Autoscaling group and launch configuration.
* Create the ECS cluster.
* Create the Task definition and service to run on the ECS cluster.

I will brief in short about each of the above steps.

**Provision VPC, subnet, routing table and security group.**

After provisioning VPS & Subnets, an Internet Gateway is added to enable traffic from outside. This Gateway is attached to the above created VPC. A new entry in Main Route Table which is created while VPC is created is added that has Target to the above created Internet Gateway. The subnets are by default placed in this Main Route Table, hence making them Public.

2 Security Groups have been created. First for the Application Load Balancer, which allows port 80 from outside. Second for the EC2 instances, where the Inbound traffic on port 80 is only allowed from the LB. I have also kept port 22 open for debugging purpose.

**Set up an ALB (Application Load Balancer)**

Create a ALB.

Create a Target group under the above VPC created and attach the same with the ALB.

**Set up an Autoscaling group and launch configuration.**

Create a launch configuration that the ASG will use for creating EC2 instances. The configuration includes details of the instance type, image id, etc. Please attach the above created instance profile to the configuration. As these instances needs to be part of the cluster please include user data echo **ECS\_CLUSTER=aws-restAPI-cluster >> /etc/ecs/ecs.config** to the same.

Create a ASG and attach the above launch configuration to it.

**Set up IAM roles and instance profile.**

Setup IAM roles for the Task Role with READ access to ECR and Write to Cloud Watch Logs.

Setup IAM role and instance profile for the ec2 instance.

**Create the ECS cluster.**

Create a ECS cluster with container Insights logs enabled.

**Create the Task definition and service to run on the ECS cluster.**

Created a Task definition with the required Task Tole, the image location present in the ECR. As the service runs on port 5000 inside the container, it has been mapped to port 80 of the EC2 instance.

Created a service to the run the above created tasks. The service runs 2 tasks. The idea behind using service is for High Availability. In case task is stopped due to some reason, new task is initiated by the Service to match the Desired Count.