

AWS ECS (Elastic Container Service)

ECS → Is a Cloud-based container management service that allows you to run and manage Docker containers on a cluster of virtual servers.

It automatically handles

≡ Creation

≡ Management

≡ Updating

→ It allows you to run, stop and manage Docker containers on a cluster of virtual machines by automatically.

Key Components of ECS

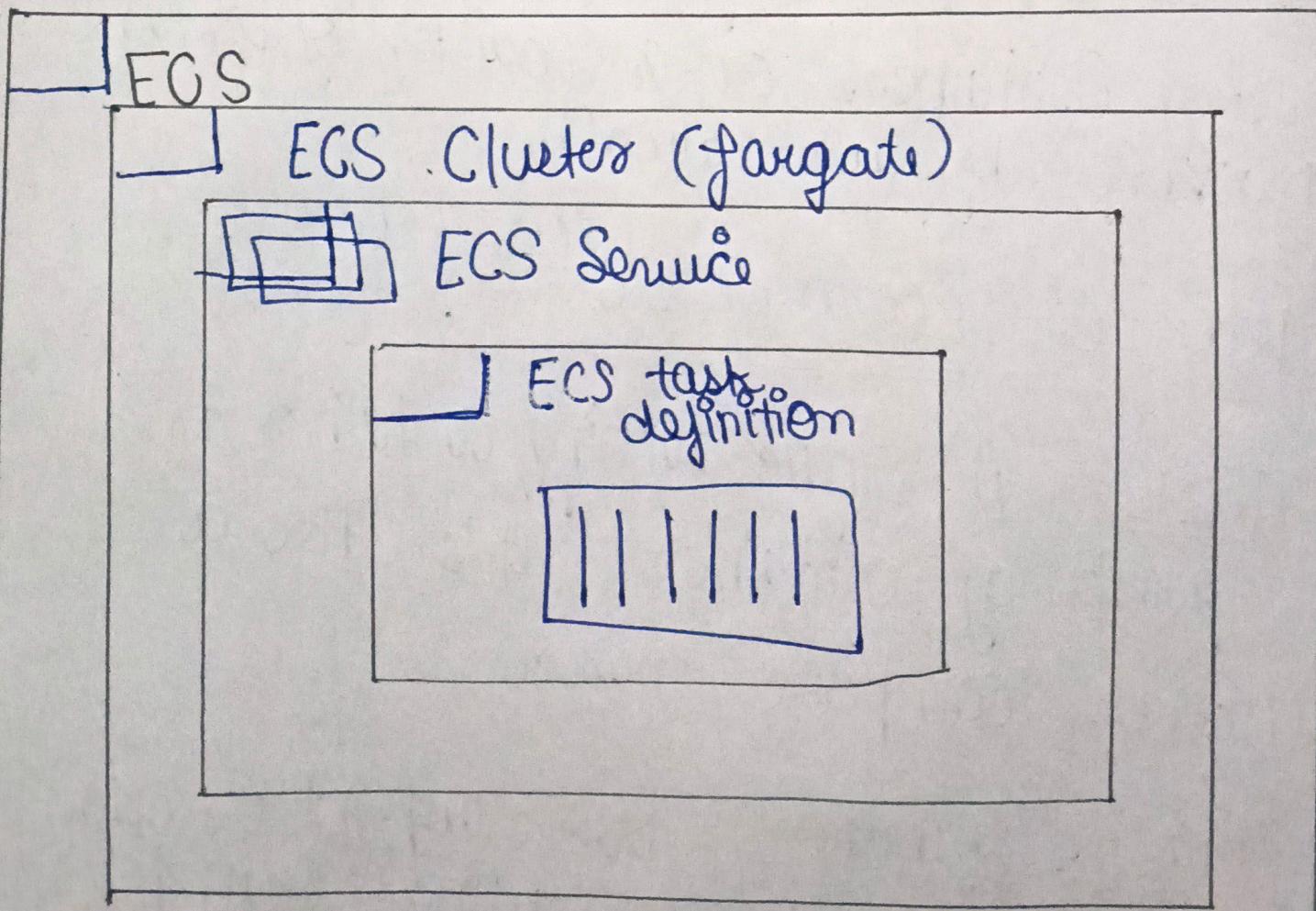
1. Task - A single running container or group of containers. It's like a Docker compose service

2. Task definition - A blueprint (JSON) that describes your container settings (image, CPU, memory, ports etc).

3. Service: Ensure the desired no. of task instances are always running. Supports load balancing and handles scalability.

4. Cluster: A group of EC2 instances or group of tasks and services, host all the resources and infrastructure.

5. Container Agent: Installed on EC2 instances to communicate with ECS.



Two Types Launch

1. Fargate:- Serverless - you don't manage any EC2 instances. just specify GPU, memory, and container image.

2. EC2:- You provision and manage EC2 instances yourself. more control, but more responsibility.

Benefit of ECS:

- = Deep integrate with AWS Services (IAM, Cloudwatch, ELB, VPC etc).
- = Auto Scaling (based on CPU/memory or custom metrics.)
- Secure - IAM roles for tasks, VPC networking
- Supports blue/green deployment via CodeDeploy.

A Step-by-step guide for Fargate on ECS based ECS

Option-1 :- ECS with Fargate (Serverless - NO EC2 to manage).

Step-1 :- Create a docker app

Step-2 :- Push docker image to Amazon ECR

Step-3 :- Create an ECS cluster

Choose : Networking Only (Fargate).

Name :- my-cluster.

Step-4 :- Create Task definition (Fargate)

= ECS > Task definition > Create new

= Launch type : Fargate.

= Add container, image, port

= Network mode : AWSVPC.

. Save.

Create a local Balancer (ALB)

Step-5

= Go to EC2 > Load Balancer > Create ALB

- Target type : ip
- Target port 3000
- Add VPC + subnet.

Step-6 :- Create ECS Service

- ECS > Clusters > Service > Create.
- Launch type : Fargate
- Load balancer : ALB
 - Listen port : 80
 - Target group : that you create.
- Task definition : choose above

Step-7:- Test your app

Open your load Balancer DNS in Browser.

You should see your app running

Option-2:- ECS with EC2

Step-1:- Create Docker App + push to ECR

Step-2:- Create an Cluster

- Choose : EC2 Linux + networking
- Add EC2 instance IAM, instance type, key pair.

Step-3) - Create Task Definition

- = Launch type : EC2
- = Add container with ECR image and ports

Step-4 : Create ECS Service

- Use the task definition
- Launch EC2 type.
- NO. Of task 1 or 2
- Add ALB optional.

Step-5) Security Group (EC2)

- Inbound rule for port 3000 is allowed in SG.

Step-6) Test your app

- Get the EC2 public id or ALB DNS
- Visit the browser