

ECS (Elastic container service)

1) Setup a highly available ecs cluster with load balancer and dynamic port mapping.

NOTE:

Your cluster should maintain atleast 6 tasks at any point of time and should be highly available across multiple AZ's.

Use the below image from deployment:

sabair0509/hiring-app:works

→ Create ECS cluster:

The screenshot displays the AWS Management Console for the 'Create cluster' wizard in the Amazon Elastic Container Service (ECS) console. The interface is in the 'Cluster configuration' step. The cluster name is 'Hiring-App-Dev'. The provisioning model is set to 'On-demand'. The EC2 instance type is 't2.micro'. The desired capacity is set to a minimum of 0 and a maximum of 2. The SSH key pair is 'docker'.

Cluster configuration

Cluster name: Hiring-App-Dev

Cluster name must be 1 to 255 characters. Valid characters are a-z, A-Z, 0-9, hyphens (-), and underscores (_).

► Service Connect defaults - optional

▼ Infrastructure - optional [Info](#) [Customized](#)

Your cluster is automatically configured for AWS Fargate (serverless) with two capacity providers. Add Amazon EC2 instances.

☐ AWS Fargate (serverless)
Pay as you go. Use if you have tiny, batch, or burst workloads or for zero maintenance overhead. The cluster has Fargate and Fargate Spot capacity providers by default.

☒ Amazon EC2 instances
Manual configurations. Use for large workloads with consistent resource demands.

Auto Scaling group (ASG) [Info](#)

Use Auto Scaling groups to scale the Amazon EC2 instances in the cluster.

Create new ASG

Provisioning model
Select a provisioning model for your instances

☒ On-demand
With on-demand instances, you pay for compute capacity by the hour, with no long-term commitments or upfront payments.

☐ Spot
Amazon EC2 Spot instances let you take advantage of unused EC2 capacity in the AWS cloud. Spot instances are available at up to a 90% discount compared to on-demand prices.

Container instance Amazon Machine Image (AMI)
Choose the Amazon ECS-optimized AMI for your instance.

Amazon Linux 2 (kernel 5.10)

EC2 instance type
Choose based on the workloads you plan to run on this cluster.

t2.micro [Free tier eligible](#)
1386, x86_64
1 vCPU 1 GiB Memory

EC2 instance role
An instance role is used by Amazon EC2 instances to make AWS API requests. If you don't already have an instance IAM role created, we can create one for you.

Create new role

Desired capacity
Specify the number of instances to launch in your cluster.

Minimum: 0 Maximum: 2

SSH Key pair
If you do not specify a key pair, you can't connect to the instances via SSH unless you choose an AMI that is configured to allow users another way to log in.

docker [Create a new key pair](#)

Amazon Elastic Container Service

Clusters

Namespaces

Task definitions

Account settings

Install AWS Copilot

Amazon ECR

Repositories

AWS Batch

Documentation

Discover products

Subscriptions

Amazon Elastic Container Service

Create cluster

AMAZON LINUX 2 (ARM64 T3, F3)

EC2 instance type

Choose based on the workloads you plan to run on this cluster.

t2.micro

i386, x86_64

1 vCPU 1 GiB Memory

Free tier eligible

EC2 instance role

An instance role is used by Amazon EC2 instances to make AWS API requests. If you don't already have an instance IAM role created, we can create one for you.

ecsInstanceRole

arn:aws:iam::971422718404:instance-profile/ecsInstanceRole

Desired capacity

Specify the number of instances to launch in your cluster.

Minimum

2

Maximum

2

SSH Key pair

If you do not specify a key pair, you can't connect to the instances via SSH unless you choose an AMI that is configured to allow users another way to log in.

docker

Create a new key pair

Root EBS volume size

You can increase the size of the root EBS volume to allow for greater image and container storage.

30

A min of 30 GiB and a max of 16,384 GiB is allowed.

Amazon Elastic Container Service

Clusters

Namespaces

Task definitions

Account settings

Install AWS Copilot

Amazon ECR

Repositories

AWS Batch

Documentation

Discover products

Subscriptions

Amazon Elastic Container Service

Create cluster

Root EBS volume size

You can increase the size of the root EBS volume to allow for greater image and container storage.

30

A min of 30 GiB and a max of 16,384 GiB is allowed.

External instances using ECS Anywhere can be registered after cluster creation is complete.

Network settings for Amazon EC2 instances

By default Amazon EC2 instances are launched in the default subnets for your default VPC. To use the non-default VPC, specify the VPC and subnets.

VPC

Select a VPC to use for your Amazon ECS resources.

vpc-0f9ee23733eee978a

default

Create a new VPC

Subnets

Select the subnets where your instances are launched and your tasks run. We recommend that you use three subnets for production.

Choose subnets

Clear current selection

subnet-084cd94d0085a7ade

us-east-2a 172.31.0.0/20

subnet-08ece92ad48a60ee4

us-east-2b 172.31.16.0/20

Security group

Choose an existing security group or create a new security group.

Use an existing security group

Create a new security group

Amazon Elastic Container Service

Clusters

Namespaces

Task definitions

Account settings

Install AWS Copilot

Amazon ECR

Repositories

AWS Batch

Documentation

Discover products

Subscriptions

Amazon Elastic Container Service

Create cluster

Security group name

Choose an existing security group.

Choose security groups

sg-0ebf225637d234b6b

default

Auto-assign public IP

Choose whether to auto-assign a public IP to the Amazon EC2 instances

Turn on

Monitoring - optional

CloudWatch Container Insights is a monitoring and troubleshooting solution for containerized applications and microservices.

Encryption - optional

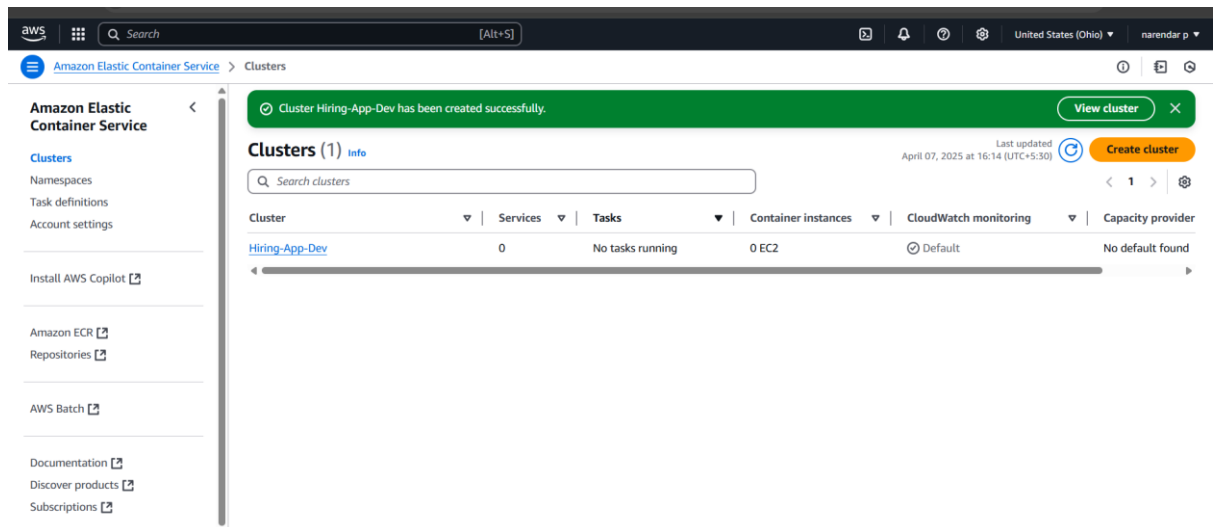
Choose the KMS keys used by tasks running in this cluster to encrypt your storage.

Tags - optional

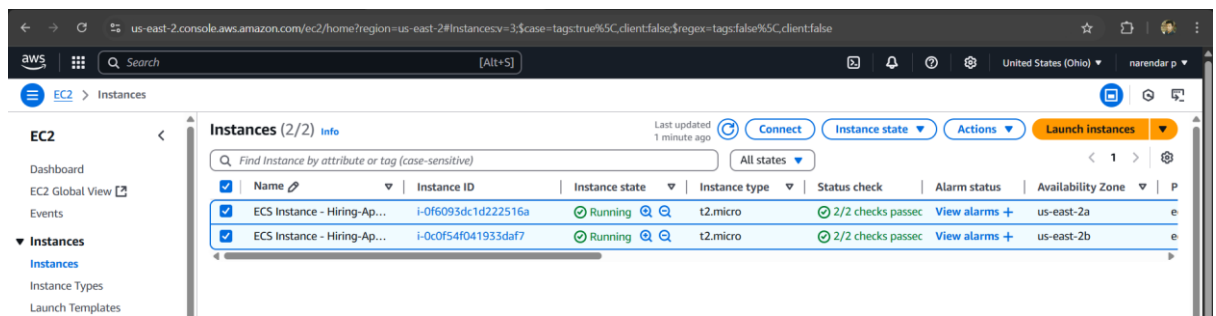
Tags help you to identify and organize your clusters.

Cancel Create

→ created ecs cluster with 2 ec2s:



→ Two ec2s are created:



→ connected to ec2:

```
root@ip-172-31-2-153:~
naren@narendar MINGW64 /c/Users/Public/Downloads
$ ssh -i "docker.pem" ec2-user@ec2-3-15-208-110.us-east-2.compute.amazonaws.com
The authenticity of host 'ec2-3-15-208-110.us-east-2.compute.amazonaws.com (3.15.208.110)' can't be established.
ED25519 key fingerprint is SHA256:yr138H+NLGZXGfy3ccXl+odMZLM3h6exsjyNs18evZg.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-3-15-208-110.us-east-2.compute.amazonaws.com' (ED25519) to the list of known hosts.

  _ _ _ _ _
 _ _ _ _ _ Amazon Linux 2 (ECS Optimized)
 _ _ _ _ _

For documentation, visit http://aws.amazon.com/documentation/ecs
$ package(s) needed for security, out of 11 available
update to the latest ECS-Optimized AMI for all updates, including security
run "sudo yum update --security" to apply security updates in place
run "sudo yum update" to apply all updates in place
ec2-user@ip-172-31-2-153 ~]$ sudo -i
[root@ip-172-31-2-153 ~]# docker --version
Docker version 25.0.8, build 0bab007
[root@ip-172-31-2-153 ~]# docker images
REPOSITORY          TAG                 IMAGE ID            CREATED             SIZE
amazon/amazon-ecs-agent   latest             7ffb112b782a       2 weeks ago        106MB
abs-csi-driver         latest             b8dd02981995       2 weeks ago        57.1MB
ecs-service-connect-agent interface-v1        elf354809f1d       2 months ago       168MB
amazon/amazon-ecs-pause  0.1.0             9dd4685d3644       10 years ago       702kB
[root@ip-172-31-2-153 ~]# docker container ps
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS              PORTS              NAMES
2f1aeb29482b        amazon/amazon-ecs-agent:latest "/agent"           7 minutes ago      Up 7 minutes (healthy)
[root@ip-172-31-2-153 ~]#
```

→ coming to tasks first create Task Definition by adding container (Add nginx image):

The screenshot shows the 'Create new task definition' page in the AWS Management Console. The left sidebar contains the 'Amazon Elastic Container Service' menu with options like Clusters, Namespaces, Task definitions, and Account settings. The main content area is titled 'Create new task definition' and includes an 'Info' link. The 'Task definition configuration' section is active, showing a text input for 'Task definition family' with the value 'Hiring-App-Dev-TD'. Below this, the 'Infrastructure requirements' section is expanded, showing 'Launch type' with 'Amazon EC2 instances' selected. The 'OS, Architecture, Network mode' section shows 'Operating system/Architecture' set to 'Linux/X86_64' and 'Network mode' set to 'awsvpc'. The 'Task size' section is also visible.

Task definition configuration

Task definition family | Info
Specify a unique task definition family name.
Hiring-App-Dev-TD
Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.

▼ Infrastructure requirements
Specify the infrastructure requirements for the task definition.

Launch type | Info
Selection of the launch type will change task definition parameters.

☐ AWS Fargate
Serverless compute for containers.

☒ Amazon EC2 instances
Self-managed infrastructure using Amazon EC2 instances.

OS, Architecture, Network mode
Network mode is used for tasks and is dependent on the compute type selected.

Operating system/Architecture | Info
Linux/X86_64

Network mode | Info
awsvpc

Task size | Info
Specify the amount of CPU and memory to reserve for your task.

The screenshot shows the 'Create new task definition' page in the AWS Management Console, continuing from the previous section. The 'Task roles' section is expanded, showing 'Task role' set to '-'. The 'Task execution role' section is also expanded, showing 'Create new role' as the selected option. The 'Task placement' section is expanded, showing 'Constraint' set to 'Add placement constraint'. The 'Task size' section is also visible.

Operating system/Architecture | Info
Linux/X86_64

Network mode | Info
bridge

Task size | Info
Specify the amount of CPU and memory to reserve for your task.

CPU
1 vCPU

Memory
3 GB

▼ Task roles - conditional

Task role | Info
A task IAM role allows containers in the task to make API requests to AWS services. You can create a task IAM role from the [IAM console](#).
-

Task execution role | Info
A task execution IAM role is used by the container agent to make AWS API requests on your behalf. If you don't already have a task execution IAM role created, we can create one for you.
Create new role

▼ Task placement - optional

Constraint | Info
Task placement constraints allow you to filter the container instances used for the placement of your tasks using built-in or custom attributes. The service scheduler first filters the container instances that match the constraints and then applies the placement strategy to place the task.
Add placement constraint
You can add 10 more placement constraints.

→ create tasks:

This screenshot shows the 'Run task' configuration page in the Amazon Elastic Container Service console. The left sidebar contains the 'Amazon Elastic Container Service' menu with options like Clusters, Namespaces, Task definitions, and Account settings. The main content area is titled 'Existing cluster' and shows the cluster 'Hiring-App-Dev'. Under 'Compute configuration (advanced)', the 'Launch type' is set to 'EC2'. The 'Deployment configuration' section shows the 'Task definition family' as 'Hiring-App-Dev-TD' and the 'Task definition revision' as '1'.

Existing cluster
Hiring-App-Dev

▼ Compute configuration (advanced)

Compute options | Info
To ensure task distribution across your compute types, use appropriate compute options.

☐ Capacity provider strategy
Specify a launch strategy to distribute your tasks across one or more capacity providers.

☒ Launch type
Launch tasks directly without the use of a capacity provider strategy.

Launch type | Info
Select either managed capacity (Fargate), or custom capacity (EC2 or user-managed, External instances). External instances are registered to your cluster using the ECS Anywhere capability.

EC2

Deployment configuration

Task definition family
Select an existing task definition family. To create a new task definition, go to [Task definitions](#).

Hiring-App-Dev-TD

Task definition revision
Select the task definition revision from the 100 most recent entries, or enter a revision. Leave the field blank to use the latest revision.

1

This screenshot shows the 'Create new task definition' page in the Amazon Elastic Container Service console. The left sidebar is the same as the previous screenshot. The main content area is titled 'Create new task definition'. Under 'Resource allocation limits - conditional', the 'CPU' is set to '1', 'GPU' to '1', 'Memory hard limit' to '3', and 'Memory soft limit' to '1'. The 'Environment variables - optional' section has buttons for 'Add environment variable' and 'Add environment file'. The 'Logging - optional' section is also visible.

Resource allocation limits - conditional | Info
Container-level CPU, GPU, and memory limits are different from task-level values. They define how much resources are allocated for the container. If container attempts to exceed the memory specified in hard limit, the container is terminated.

CPU: 1 (in vCPU)
GPU: 1
Memory hard limit: 3 (in GB)
Memory soft limit: 1 (in GB)

▼ Environment variables - optional

Environment variables | Info
Add individually
Add a key-value pair to specify an environment variable.

[Add environment variable](#)

Add from file
Add environment variables in bulk by providing an environment file hosted on Amazon S3.

[Add environment file](#)
You can add 10 more environment files.

▼ Logging - optional

This screenshot shows the 'Run task' configuration page in the Amazon Elastic Container Service console, similar to the first screenshot but with additional changes. The 'Task definition revision' is now '2'. The 'Desired tasks' field is set to '6'. The 'Task group' field is empty.

Compute options | Info
To ensure task distribution across your compute types, use appropriate compute options.

☐ Capacity provider strategy
Specify a launch strategy to distribute your tasks across one or more capacity providers.

☒ Launch type
Launch tasks directly without the use of a capacity provider strategy.

Launch type | Info
Select either managed capacity (Fargate), or custom capacity (EC2 or user-managed, External instances). External instances are registered to your cluster using the ECS Anywhere capability.

EC2

Deployment configuration

Task definition family
Select an existing task definition family. To create a new task definition, go to [Task definitions](#).

Hiring-App-Dev-TD

Task definition revision
Select the task definition revision from the 100 most recent entries, or enter a revision. Leave the field blank to use the latest revision.

2

Desired tasks
Specify the number of tasks to launch.

6

Task group

us-east-2.console.aws.amazon.com/ecs/v2/clusters/Hiring-App-Dev/tasks?region=us-east-2

Amazon Elastic Container Service > Clusters > Hiring-App-Dev > Tasks

Tasks launched

- arn:aws:ecs:us-east-2:971422718404:task/Hiring-App-Dev/3755916898de435989ba41d32517a5f8
- arn:aws:ecs:us-east-2:971422718404:task/Hiring-App-Dev/5341ef18885546ff83e99a371bd5e2ce
- arn:aws:ecs:us-east-2:971422718404:task/Hiring-App-Dev/b28bb7af97d746688be26fd662a02fa7
- arn:aws:ecs:us-east-2:971422718404:task/Hiring-App-Dev/ae929601795b4541994479cf607f0edb
- arn:aws:ecs:us-east-2:971422718404:task/Hiring-App-Dev/39a879ec77c4e41b03d324e659076fa
- arn:aws:ecs:us-east-2:971422718404:task/Hiring-App-Dev/2795dffa04f345d7a00b6f9384acc430

Hiring-App-Dev ASG Last updated April 07, 2025 at 17:48 (UTC+5:30) [Update cluster](#) [Delete cluster](#)

Cluster overview

ARN: [arn:aws:ecs:us-east-2:971422718404:cluster/Hiring-App-Dev](#) Status: Active CloudWatch monitoring: Default Registered container instances: 2

Services: Draining: - Active: - Tasks: Pending: 6 Running: -

Services **Tasks** Infrastructure Metrics Scheduled tasks Configuration Tags

Tasks (6) [Manage tags](#) [Stop](#) [Run new task](#)

→ tasks created:

Dashboard EC2 Global View Events

▼ **Instances**

- Instances
- Instance Types
- Launch Templates
- Spot Requests
- Savings Plans
- Reserved Instances
- Dedicated Hosts
- Capacity Reservations

▼ **Images**

- AMIs
- AMI Catalog

▼ **Elastic Block Store**

- Volumes
- Snapshots
- Lifecycle Manager

Distribution or targets by availability zone (14)

Select values in this table to see corresponding filters applied to the Registered targets table below.

Targets Monitoring Health checks Attributes Tags

Registered targets (5) [Anomaly mitigation: Not applicable](#) [Deregister](#) [Register targets](#)

Target groups route requests to individual registered targets using the protocol and port number specified. Health checks are performed on all registered targets according to the target group's health check settings. Anomaly detection is automatically applied to HTTP/HTTPS target groups with at least 3 healthy targets.

Filter targets

<input type="checkbox"/>	Instance ID	Name	Port	Zone	Health status	Health status details	Admin...	Overri
<input type="checkbox"/>	i-023f35b52b351cd2e	ECS Instance - ...	32769	us-east-1b (us...	Healthy	-	No override	No ove
<input type="checkbox"/>	i-023f35b52b351cd2e	ECS Instance - ...	80	us-east-1b (us...	Unhealthy	Health checks failed	No override	No ove
<input type="checkbox"/>	i-023f35b52b351cd2e	ECS Instance - ...	32768	us-east-1b (us...	Healthy	-	No override	No ove
<input type="checkbox"/>	i-0c8a2d8953f68bc0d	ECS Instance - ...	80	us-east-1a (us...	Unhealthy	Health checks failed	No override	No ove
<input type="checkbox"/>	i-0c8a2d8953f68bc0d	ECS Instance - ...	32768	us-east-1a (us...	Healthy	-	No override	No ove

→ Created target groups for loadbalancer:

us-east-2.console.aws.amazon.com/ec2/home?region=us-east-2#CreateTargetGroup:protocol=HTTP:vpc=vpc-0f9ee23733ee978a

EC2 > Target groups > Create target group

Step 1 Specify group details

Step 2 **Register targets**

Register targets

This is an optional step to create a target group. However, to ensure that your load balancer routes traffic to this target group you must register your targets.

Available instances (2/2)

Filter instances

<input checked="" type="checkbox"/>	Instance ID	Name	State	Security groups	Zone
<input checked="" type="checkbox"/>	i-0f6093dc1d222516a	ECS Instance - Hiring-App-Dev	Running	default	us-east-2a
<input checked="" type="checkbox"/>	i-0c0f54f041933daf7	ECS Instance - Hiring-App-Dev	Running	default	us-east-2b

2 selected

Ports for the selected instances

Ports for routing traffic to the selected instances.

8080

1-65535 (separate multiple ports with commas)

[Include as pending below](#)

us-east-2.console.aws.amazon.com/ec2/home?region=us-east-2#TargetGroup:targetGroupArn=arn:aws:elasticloadbalancing:us-east-2:971422718404:targetgroup/hiring-app-dev-loadb/856667daaea2e3ac

EC2 > Target groups > hiring-app-dev-loadb

Successfully created the target group: hiring-app-dev-loadb. Anomaly detection is automatically applied to all registered targets. Results can be viewed in the Targets tab.

hiring-app-dev-loadb

Details

arn:aws:elasticloadbalancing:us-east-2:971422718404:targetgroup/hiring-app-dev-loadb/856667daaea2e3ac

Target type Instance	Protocol : Port HTTP: 8080	Protocol version HTTP1	VPC vpc-0f9ee23733eee978a
IP address type IPv4	Load balancer None associated		

0 Total targets	0 Healthy	0 Unhealthy	0 Unused	0 Initial	0 Draining
--------------------	--------------	----------------	-------------	--------------	---------------

0 Anomalous

Targets | Monitoring | Health checks | Attributes | Tags

Registered targets (0) [Info](#) [Anomaly mitigation: Not applicable](#) [Deregister](#) [Register targets](#)

Target groups route requests to individual registered targets using the protocol and port number specified. Health checks are performed on all registered targets according to the target group's health check settings. Anomaly detection is automatically applied to HTTP/HTTPS target groups with at least 2 healthy targets.

→load balancer:

us-east-2.console.aws.amazon.com/ec2/home?region=us-east-2#CreateALBWizard:

EC2 > Load balancers > Create Application Load Balancer

Security groups [Info](#)

A security group is a set of firewall rules that control the traffic to your load balancer. Select an existing security group, or you can [create a new security group](#).

Security groups

Select up to 5 security groups

default sg-0ebf225637d234b6b VPC: vpc-0f9ee23733eee978a

Listeners and routing [Info](#)

A listener is a process that checks for connection requests using the port and protocol you configure. The rules that you define for a listener determine how the load balancer routes requests to its registered targets.

Listener HTTP:8080 [Remove](#)

Protocol HTTP	Port 8080 1-65535	Default action Info Forward to hiring-app-dev-loadb Target type: Instance, IPv4 Create target group
-------------------------	--------------------------------	---

Listener tags - optional

Consider adding tags to your listener. Tags enable you to categorize your AWS resources so you can more easily manage them.

[Add listener tag](#)

us-east-2.console.aws.amazon.com/ec2/home?region=us-east-2#LoadBalancer:loadBalancerArn=arn:aws:elasticloadbalancing:us-east-2:971422718404:loadbalancer/app/hiring-app-dev-lb/34bec144d530064c

EC2 > Load balancers > hiring-app-dev-lb

Successfully created load balancer: hiring-app-dev-lb
It might take a few minutes for your load balancer to fully set up and route traffic. Targets will also take a few minutes to complete the registration process and pass initial health checks.

Application Load Balancers now support public IPv4 IP Address Management (IPAM)
You can get started with this feature by configuring IP pools in the Network mapping section. [Edit IP pools](#)

hiring-app-dev-lb

Details

Load balancer type Application	Status Provisioning	VPC vpc-0f9ee23733eee978a	Load balancer IP address type IPv4
Scheme Internet-facing	Hosted zone ZSAADJG6K7TL2	Availability Zones subnet-084cd94d0085a7ade us-east-2a (use2-az1) subnet-08ece92ad48a60ee4 us-east-2b (use2-az2)	Date created April 7, 2025, 18:06 (UTC+05:30)

Load balancer ARN
arn:aws:elasticloadbalancing:us-east-2:971422718404:loadbalancer/app/hiring-app-dev-lb/34bec144d530064c

DNS name [Info](#)
hiring-app-dev-lb-2025809928.us-east-2.elb.amazonaws.com (A Record)

→ Create Service:

The screenshot shows the 'Create service' page in the AWS Management Console for an Amazon Elastic Container Service (ECS) cluster named 'Hiring-App-Dev'. The page is for the 'us-east-2' region. The left sidebar shows the 'Amazon Elastic Container Service' navigation menu with options like Clusters, Namespaces, Task definitions, and Account settings. The main content area is titled 'Task definition revision' and includes a search bar with '2' entered. Below this, there are fields for 'Service name' and 'Service type'. The 'Service type' section has two options: 'Replica' (selected) and 'Daemon'. The 'Replica' option is described as 'Place and maintain a desired number of tasks across your cluster'. Below this, there is a 'Desired tasks' field with a value of '6'. There are also checkboxes for 'Availability Zone rebalancing' (checked) and 'Health check grace period' (set to 0 seconds).

This screenshot shows the scaling options for the 'Create service' page. The 'Target tracking' option is selected, with a description: 'Increase or decrease the number of tasks that your service runs based on a target value for a specific metric.' The 'Step scaling' option is also visible, with a description: 'Increase or decrease the number of tasks that your service runs based on a set of scaling adjustments, known as step adjustments, that vary based on the size of the alarm breach.' Below these options, there are fields for 'Policy name' (set to 'cpu-utilisation'), 'ECS service metric' (set to 'ECSServiceAverageCPUUtilization'), 'Target value' (set to '70'), 'Scale-out cooldown period' (set to '300'), and 'Scale-in cooldown period' (set to '300'). There is also a checkbox for 'Turn off scale-in' which is currently unchecked.

The screenshot shows the 'Devcluster' overview page in the AWS Management Console. The page is titled 'Devcluster' and has a status of 'Active'. It includes a 'Cluster overview' section with the following details:

ARN	Status	CloudWatch monitoring	Registered container instances
arn:aws:ecs:us-east-1:061051216760:cluster/Devcluster	Active	Default	2

Below the overview, there is a 'Services' section with a table showing the status of services:

Draining	Active	Pending	Running
-	1	-	2

The page also includes a 'Services (1)' section with a table showing the details of the service:

ARN	Status	Service...	Created at	Deployments and tasks	Last deployment	Task...
arn:aws:ecs:us-e	Active	REPLICA	5 minutes ago	2/2 Tasks running	Completed	View

→ access the service by using load balancer ARN on the browser:

hello world

|