

## Assignment 2: Rendezvous Server

Team 4: Sam Knisely, Darreion Bailey

1. Navigate to AWS Elastic Container Service (ECS) and select the "rendezvous-tcpunch-fargate-task" from the task definitions

The screenshot shows the AWS Elastic Container Service (ECS) Task Definitions page. The URL is [us-east-1.console.aws.amazon.com/ecs/v2/task-definitions/rendezvous-tcpunch-fargate-task?status=ACTIVE&region=us-east-1](https://us-east-1.console.aws.amazon.com/ecs/v2/task-definitions/rendezvous-tcpunch-fargate-task?status=ACTIVE&region=us-east-1). The page displays a single revision of the task definition named 'rendezvous-tcpunch-fargate-task'. The revision was last updated on July 21, 2025 at 13:02 (UTC-4:00). The status is ACTIVE. The task definition has a single container named 'rendezvous-tcpunch-fargate-task:2'. A note at the top of the page informs users about a change in log driver mode on June 25, 2025.

2. Deploy the task to the Fargate cluster using FARGATE as the launch type

The screenshot shows the AWS Elastic Container Service (ECS) Run Task page. The URL is [us-east-1.console.aws.amazon.com/ecs/v2/task-definitions/rendezvous-tcpunch-fargate-task/run-task?region=us-east-1](https://us-east-1.console.aws.amazon.com/ecs/v2/task-definitions/rendezvous-tcpunch-fargate-task/run-task?region=us-east-1). The page is titled 'rendezvous-tcpunch-fargate-task (1)'. The 'Compute configuration (advanced)' section is open, showing the 'Capacity provider strategy' tab selected. It specifies a launch strategy to distribute tasks across one or more capacity providers. The 'Capacity provider' dropdown is set to 'FARGATE'. The 'Base' field is set to '0' and the 'Weight' field is set to '1'. The 'Platform version' dropdown is set to 'LATEST'.

3. Configure the networking settings with the "open access" security group

us-east-1.console.aws.amazon.com/ecs/v2/task-definitions/rendezvous-tcpunch-fargate-task/run-task?region=us-east-1

Amazon Elastic Container Service > Task definitions > rendezvous-tcpunch-fargate-task > Run task

### Networking

VPC | Info Select a VPC to use for your Amazon ECS resources.

vpc-0a19e74ac58edb30f default

Subnets Choose the subnets within the VPC that the task scheduler should consider for placement.

subnet-0fd60e13b4d7c9065 us-east-1a 172.31.16.0/20 subnet-0942a55798ce989af us-east-1b 172.31.32.0/20

Security group | Info Choose an existing security group or create a new security group.

Use an existing security group  Create a new security group

Security group name Choose an existing security group.

sg-052ff4576743adeaf open access

Public IP | Info Choose whether to auto-assign a public IP to the task's elastic network interface (ENI).

Turned on

us-east-1.console.aws.amazon.com/ecs/v2/clusters/DS5110-Fargate-Cluster/tasks?region=us-east-1

Amazon Elastic Container Service > Clusters > DS5110-Fargate-Cluster > Tasks

On June 25, 2025, Amazon ECS changed the default log driver mode from blocking to non-blocking to improve application availability during CloudWatch outages. [Learn more](#)

Tasks launched arn:aws:ecs:us-east-1:211125778552:task/DS5110-Fargate-Cluster/d94d208e436048e9aed6635fff5b9ed7

### DS5110-Fargate-Cluster

Last updated July 21, 2025 at 13:20 (UTC-4:00)

Cluster overview		Status	CloudWatch monitoring	Registered container instances
ARN	arn:aws:ecs:us-east-1:211125778552:cluster/DS5110-Fargate-Cluster	<input checked="" type="radio"/> Active	<input checked="" type="radio"/> Default	-
Services		Tasks		
Draining	Active 1	Pending 1	Running 11	

Services  Infrastructure Metrics Scheduled tasks Configuration Tags

Tasks (12)

Filter desired status Any desired status Filter launch type Any launch type

Filter tasks by property or value < 1 >

4. Retrieve the public IP address of the deployed task

5. Update the DNS record in Route 53 for "rendezvous.uva-ds5110.com" with your task's IP address

Route 53 < C0333130122PO228CI2XF [Info]

**Change info details**

ID /change/C0333130122PO228CI2XF	Submitted at July 21, 2025, 13:32 (UTC-04:00)
Status INSYNC	Comment -

## 6. Verify the server is accessible by performing a connection test

Step Functions < COSMIC-AI-TEAM4

**Execution started successfully**

**Details**

Arn arn:aws:states:us-east-1:211125778552:stateMachine:COSMIC-AI-TEAM4	Type Standard
IAM role ARN arn:aws:iam::211125778552:role/team4-summer	Status Active
Creation date Jul 21, 2025, 14:34:52 (UTC-04:00)	
X-Ray tracing Disabled	

**Executions** (0/1)

Name	Status	Start Time (local)	End Time (local)	Duration	Version
ab4a347b-3014-496f-9c29-d64cd9014084	Succeeded	Jul 21, 2025, 14:34:58	Jul 21, 2025, 14:35:10	00:00:12.246	-

### Rubric Items:

#### 1-2. Created S3 Bucket (e.g. team4-cosmical); Clone repos

We are actually team4 so we created team4-cosmical and cloned cosmicai-data S3 bucket into it. See screenshot below:

**team4-cosmical** Info

Objects (15)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Name	Type	Last modified	Size	Storage class
100MB/	Folder	-	-	-
10MB/	Folder	-	-	-
25MB/	Folder	-	-	-
50MB/	Folder	-	-	-
75MB/	Folder	-	-	-
Anomaly Detection/	Folder	-	-	-
demo/	Folder	-	-	-
payload.json	json	July 21, 2025, 15:05:04 (UTC-04:00)	3.2 KB	Standard
result-partition-100MB/	Folder	-	-	-
result-partition-10MB/	Folder	-	-	-
result-partition-25MB/	Folder	-	-	-

### 3. Edit Payload – Updating parameters (world size, bucket) in our Cosmic AI state machine in the Lambda Init State

Here we use world size (i.e., file limit = 100) and utilize our team4-cosmical S3 bucket.

**COSMIC-AI-TEAM4** Standard

Actions | Flow | Patterns | Info

```

graph TD
    Start((Start)) --> LambdaInvoke1[Lambda: Invoke Lambda Invoke]
    LambdaInvoke1 --> Map[Map state Distributed]
    Map --> LambdaInvoke2[Lambda: Invoke Model Inference]
    LambdaInvoke2 --> LambdaInvoke3[Lambda: Invoke Summarize]

```

**Lambda Invoke**

View function

**Payload**

The JSON that you want to provide to your Lambda function.

Enter payload

```

1 {
2     "bucket": "team4-cosmical",
3     "file_limit": "100",
4     "batch_size": 512,
5     "object_type": "folder",
6     "S3_object_name": "Anomaly Detection",
7     "script": "/tmp/Anomaly
8 Detection/Inference/inference.py",
9     "result_path": "result-partition-100MB/1GB/3",
10    "data_bucket": "cosmical-data",
11    "data_prefix": "100MB"
}

```

### 4. Successful Step Function Execution

Screenshot of the AWS Step Functions console showing the execution details for a specific workflow. The execution status is Succeeded, and the duration was 0:00:16.073. The graph view shows a sequence of Lambda invoke, Map state, and Model inference steps. The Lambda Invoke step is highlighted, showing its configuration parameters.

```

graph TD
    Start((Start)) --> LambdaInvoke1[AWS Lambda Invoke: Lambda Invoke]
    LambdaInvoke1 --> MapState[Iteration #99<br/>Map state<br/>Distributed<br/>Item source: JSON Payload]
    MapState --> ModelInference[AWS Lambda Invoke: Model Inference]
    ModelInference --> Summarize[AWS Lambda Inv: Summarize]
    Summarize --> End((End))

```

**Lambda Invoke Configuration:**

```

1 v {
  "Type": "Task",
  "Resource": "arn:aws:states::lambda:invoke",
  "Parameters": {
    "FunctionName": "arn:aws:lambda:us-east-1:211125778552:function:data-parallel-init-$LATEST",
    "Payload": {
      "bucket": "team-cosmical",
      "file_limit": "100",
      "batch_size": 512,
      "object_type": "folder",
      "S3_object_name": "Anomaly Detection",
      "script": "/tmp/AnomalyDetection/Inference/inference.py",
      "result_path": "result-partition-100MB/1GB/3",
      "data_bucket": "cosmical-data",
      "data_prefix": "100MB"
    }
  }
}

```

## 5. Cloudwatch logs

Screenshot of the AWS CloudWatch Logs console showing the log group details for /aws/lambda/inference. It displays metrics, subscription filters, and log streams. The log streams table lists recent log entries with their timestamps.

Log stream	Last event time
2025/07/21/[\$LATEST]128e10ce5124815a71beb549698065	2025-07-21 20:00:18 (UTC)
2025/07/21/[\$LATEST]fe4e6875da74b4bf9ddcc1fc11578	2025-07-21 19:53:38 (UTC)
2025/07/21/[\$LATEST]0d6c21c476f149e3a045f1f24cc52952	2025-07-21 19:50:45 (UTC)
2025/07/21/[\$LATEST]10f4e9c294bf4ad388920cd3b564641e	2025-07-21 19:46:58 (UTC)

## 6. Performance Measurement across varying data sizes

World Size	Partition	Requests	Duration (s)	Memory	Cost (\$)
100	100MB	10018	2.64	1.4	0.62
50	100MB	5008	2.86	1.8	0.43
11	100MB	3327	1.87	1.9	0.2