

Project-4-Report

Masinde

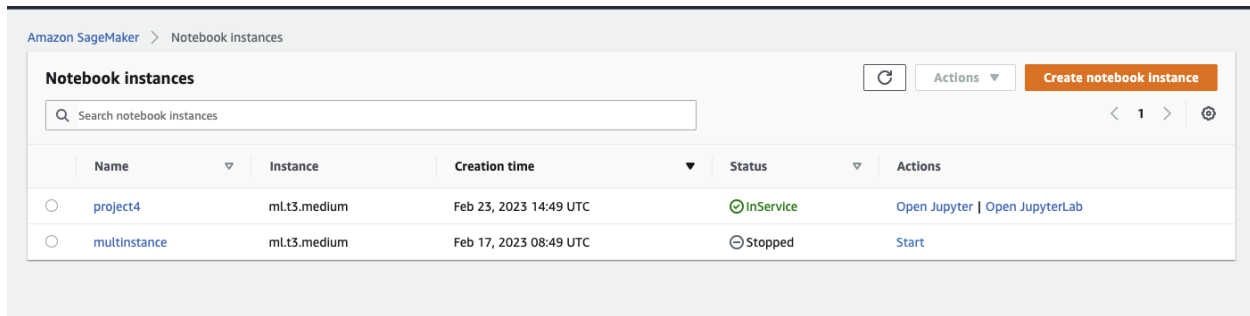
2023-02-24

Contents

Notebook Instance	2
Data Buckets (s3)	2
EC2 Instance	2
The reason for choosing this instance	2
Model training	3
Lambda Functions	3
Lambda security	3
Endpoints	3
End point auto scaling	6

Notebook Instance

The notebook instance where the starter code for the project as well as the project can be seen in figure 1. The instance name project4



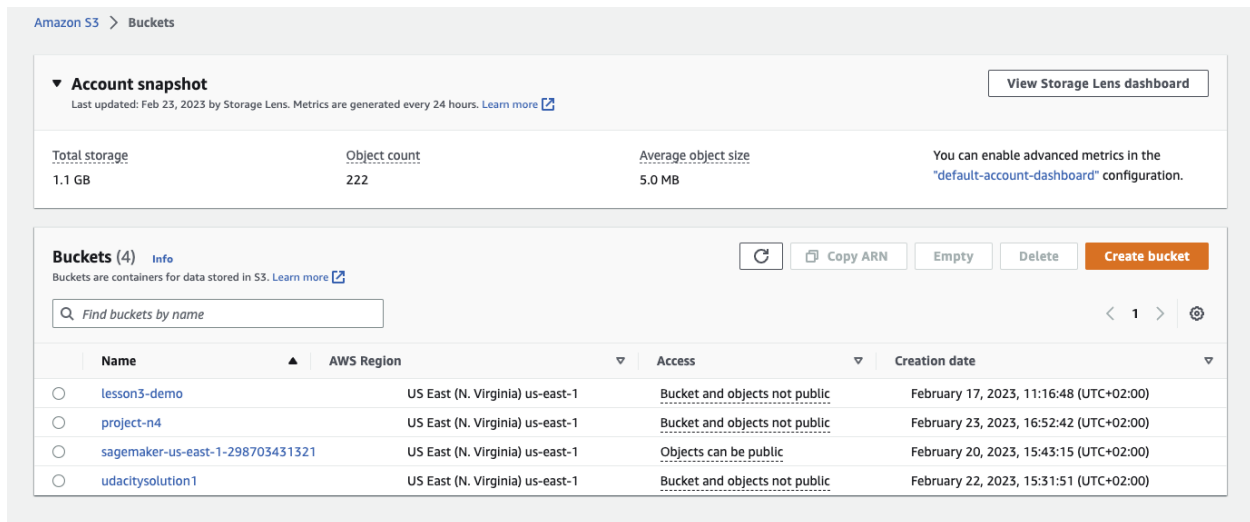
The screenshot shows the Amazon SageMaker Notebook Instances page. At the top, there's a breadcrumb 'Amazon SageMaker > Notebook Instances'. Below it, a 'Notebook instances' header with a search bar and a 'Create notebook instance' button. A table lists two instances: 'project4' (InService) and 'multinstance' (Stopped). The 'project4' instance has links to 'Open Jupyter' and 'Open JupyterLab'.

	Name	Instance	Creation time	Status	Actions
<input type="radio"/>	project4	ml.t3.medium	Feb 23, 2023 14:49 UTC	InService	Open Jupyter Open JupyterLab
<input type="radio"/>	multinstance	ml.t3.medium	Feb 17, 2023 08:49 UTC	Stopped	Start

Figure 1: Note Book Instance

Data Buckets (s3)

The data was stored in the s3 bucket as per project instructions. The name of the s3 bucket is project-n4 can be seen in figure 2.



The screenshot shows the Amazon S3 Buckets page. It includes an 'Account snapshot' section with storage and object count metrics. Below is a 'Buckets (4)' section with a search bar and a table listing four buckets: 'lesson3-demo', 'project-n4', 'sagemaker-us-east-1-298703431321', and 'udacitysolution1'. The 'project-n4' bucket is highlighted.

	Name	AWS Region	Access	Creation date
<input type="radio"/>	lesson3-demo	US East (N. Virginia) us-east-1	Bucket and objects not public	February 17, 2023, 11:16:48 (UTC+02:00)
<input type="radio"/>	project-n4	US East (N. Virginia) us-east-1	Bucket and objects not public	February 23, 2023, 16:52:42 (UTC+02:00)
<input type="radio"/>	sagemaker-us-east-1-298703431321	US East (N. Virginia) us-east-1	Objects can be public	February 20, 2023, 15:43:15 (UTC+02:00)
<input type="radio"/>	udacitysolution1	US East (N. Virginia) us-east-1	Bucket and objects not public	February 22, 2023, 15:31:51 (UTC+02:00)

Figure 2: buckets

EC2 Instance

The instance which was used in this project was Deep Learning AMI GPU PyTorch 1.13.1 (Amazon Linux 2) 20230221 figure 2 shows instance configurations.

The reason for choosing this instance

- pre-configuration of Pytorch environment hence no need to install libraries.
- GPU capability of the instance which enhance faster training

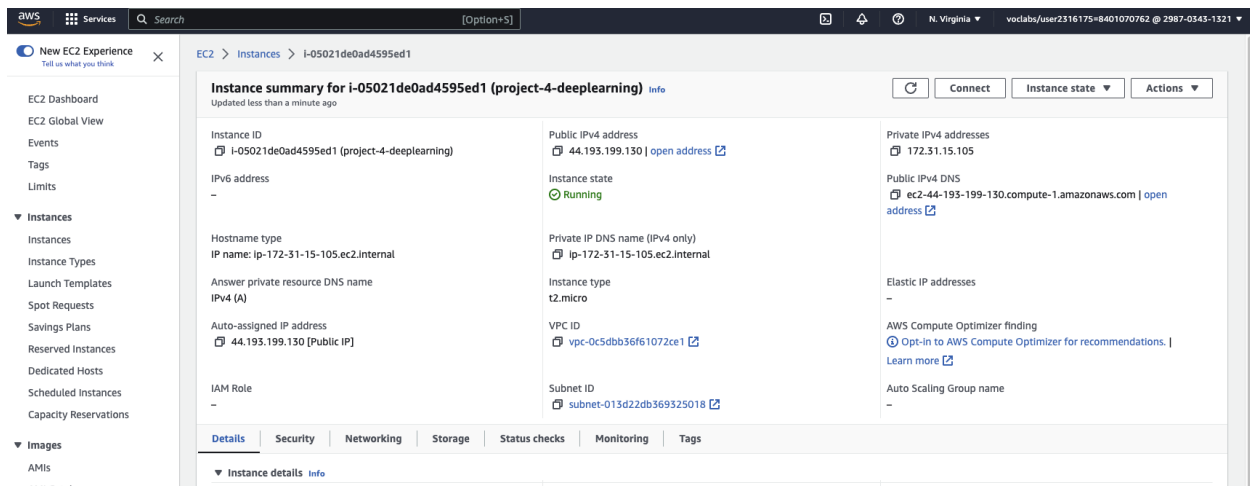


Figure 3: instance for project for 4

Model training

The model was trained as per project instruction. Figure 3 shows model training in the EC2 instance while figure 4 shows completed training.

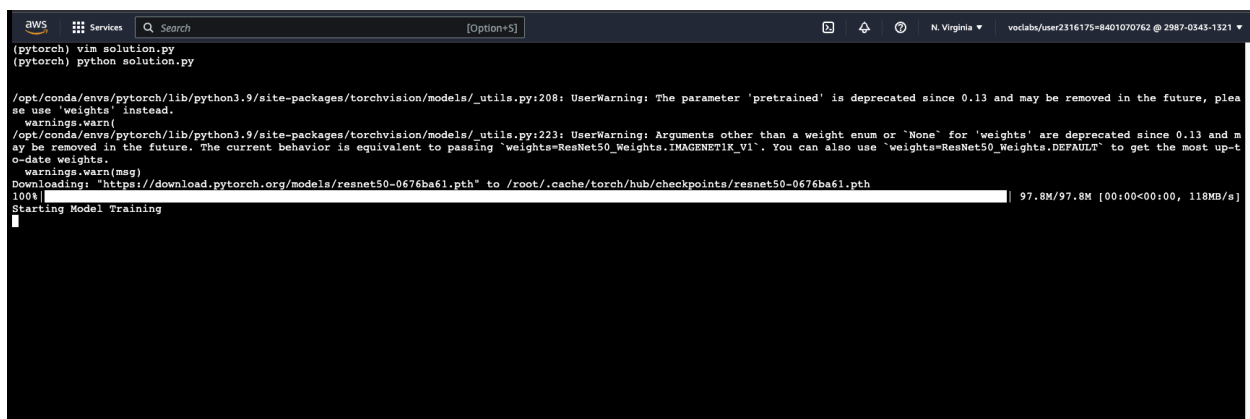


Figure 4: instance training

Lambda Functions

The lambda function which was used in this project to invoke the end point can be seen in figure 6 invoking the end point.

Lambda security

The lambda security can be seen in the following figure 6. In order for the lambda function to invoke sagemaker end point lambda function was given full access to sagemaker.

Endpoints

The deployed end points can be seen in figure 7 where it shows the single instance training as well as multi-instance training. The multi-instance training enhances faster training. The number of instances was 8 for multi-instance training and one for the single instance training.

```
inflatng: dogImages/valid/120.Pharaoh_hound/Pharaoh_hound_07752.jpg
inflatng: dogImages/valid/120.Pharaoh_hound/Pharaoh_hound_07758.jpg
creating: dogImages/valid/121.Plott/Plott_07771.jpg
inflatng: dogImages/valid/121.Plott/Plott_07780.jpg
inflatng: dogImages/valid/121.Plott/Plott_07796.jpg
inflatng: dogImages/valid/121.Plott/Plott_07802.jpg
creating: dogImages/valid/122.Pointer/
inflatng: dogImages/valid/122.Pointer/Pointer_07808.jpg
inflatng: dogImages/valid/122.Pointer/Pointer_07826.jpg
inflatng: dogImages/valid/122.Pointer/Pointer_07831.jpg
inflatng: dogImages/valid/122.Pointer/Pointer_07834.jpg
(pytorch) vim solution.py
(pytorch) python solution.py

/opt/conda/envs/pytorch/lib/python3.9/site-packages/torchvision/models/_utils.py:208: UserWarning: The parameter 'pretrained' is deprecated since 0.13 and may be removed in the future, please use 'weights' instead.
  warnings.warn(
/opt/conda/envs/pytorch/lib/python3.9/site-packages/torchvision/models/_utils.py:223: UserWarning: Arguments other than a weight enum or 'None' for 'weights' are deprecated since 0.13 and may be removed in the future. The current behavior is equivalent to passing 'weights=ResNet50_Weights.IMAGENET1K_V1'. You can also use 'weights=ResNet50_Weights.DEFAULT' to get the most up-to-date weights.
  warnings.warn(msg)
Downloading: "https://download.pytorch.org/models/resnet50-0676ba61.pth" to /root/.cache/torch/hub/checkpoints/resnet50-0676ba61.pth
100%
Starting Model Training
saved
(pytorch)
(pytorch)
(pytorch)
```

Figure 5: Finished training

Lambda > Functions > project-4

project-4

Throttle Copy ARN Actions

Function overview Info

project-4

Layers (0)

+ Add trigger

+ Add destination

Description

-

Last modified

5 minutes ago

Function ARN

arn:aws:lambda:us-east-1:298703431321:function:project-4

Function URL Info

-

Code Test Monitor Configuration Aliases Versions

Code source Info

Upload from

File Edit Find View Go Tools Window Test Deploy

Go to Anything (⌘ P)

lambda_function x Execution result: x

Environment

project-4 -

lambda_function.py

Execution results

Test Event Name

project-4

Response

{

"statusCode": 200,

"headers": {

"Content-Type": "text/plain",

"Access-Control-Allow-Origin": ""

},

"type-result": "<class 'str'>",

"Content-Type-In": "LambdaContext([aws_request_id=bcc2ad59-34a0-4d98-afac-f641c685f4b7, log_group_name=/aws/lambda/project-4, log_stream_name=2023/02/27/[LATEST]1610bf7155704eb58c06c4c3796972, EventId=bcc2ad59-34a0-4d98-afac-f641c685f4b7, ReportRequestId=bcc2ad59-34a0-4d98-afac-f641c685f4b7, Duration: 1094.97 ms, Billed Duration: 1095 ms, Memory Size: 128 MB, Max Memory Used: 68 MB, Init Duration: 307.98 ms])"

}

Function Logs

Loading Lambda function

START RequestId: bcc2ad59-34a0-4d98-afac-f641c685f4b7 Version: \$LATEST

Context::: LambdaContext([aws_request_id=bcc2ad59-34a0-4d98-afac-f641c685f4b7, log_group_name=/aws/lambda/project-4, log_stream_name=2023/02/27/[LATEST]1610bf7155704eb58c06c4c3796972, EventId=bcc2ad59-34a0-4d98-afac-f641c685f4b7, ReportRequestId=bcc2ad59-34a0-4d98-afac-f641c685f4b7, Duration: 1094.97 ms, Billed Duration: 1095 ms, Memory Size: 128 MB, Max Memory Used: 68 MB, Init Duration: 307.98 ms])

END RequestId: bcc2ad59-34a0-4d98-afac-f641c685f4b7

REPORT RequestId: bcc2ad59-34a0-4d98-afac-f641c685f4b7 Duration: 1094.97 ms Billed Duration: 1095 ms Memory Size: 128 MB Max Memory Used: 68 MB Init Duration: 307.98 ms

Request ID

bcc2ad59-34a0-4d98-afac-f641c685f4b7

Figure 6: Lambda Function

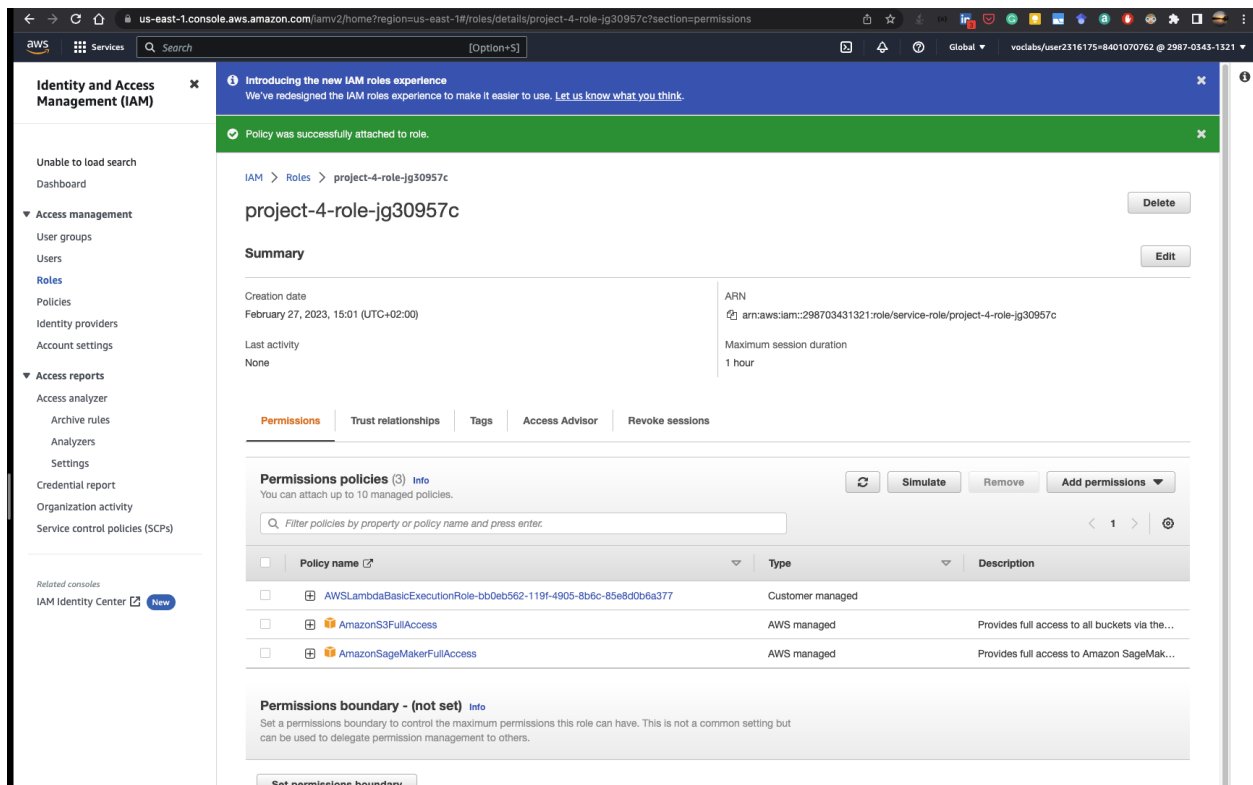


Figure 7: lambda security setup

Amazon SageMaker > Endpoints

Endpoints

Search endpoints

Update endpoint Actions Create endpoint

	Name	ARN	Creation time	Status	Last updated
<input type="radio"/>	pytorch-inference-2023-02-24-11-27-02-675	arn:aws:sagemaker:us-east-1:298703431321:endpoint/pytorch-inference-2023-02-24-11-27-02-675	Feb 24, 2023 11:27 UTC	InService	Feb 24, 2023 11:29 UTC
<input type="radio"/>	pytorch-inference-2023-02-24-10-14-30-382	arn:aws:sagemaker:us-east-1:298703431321:endpoint/pytorch-inference-2023-02-24-10-14-30-382	Feb 24, 2023 10:14 UTC	InService	Feb 24, 2023 10:17 UTC

Figure 8: EndPoints

End point auto scaling

Auto scaling was setup to enable traffic when needed. The scaling up time was 30 seconds when the traffic is high hence another instance is needed, as well as cooling down when the traffic is low. Figure 8 shows the number of instance for high traffic which is 5 max instance.

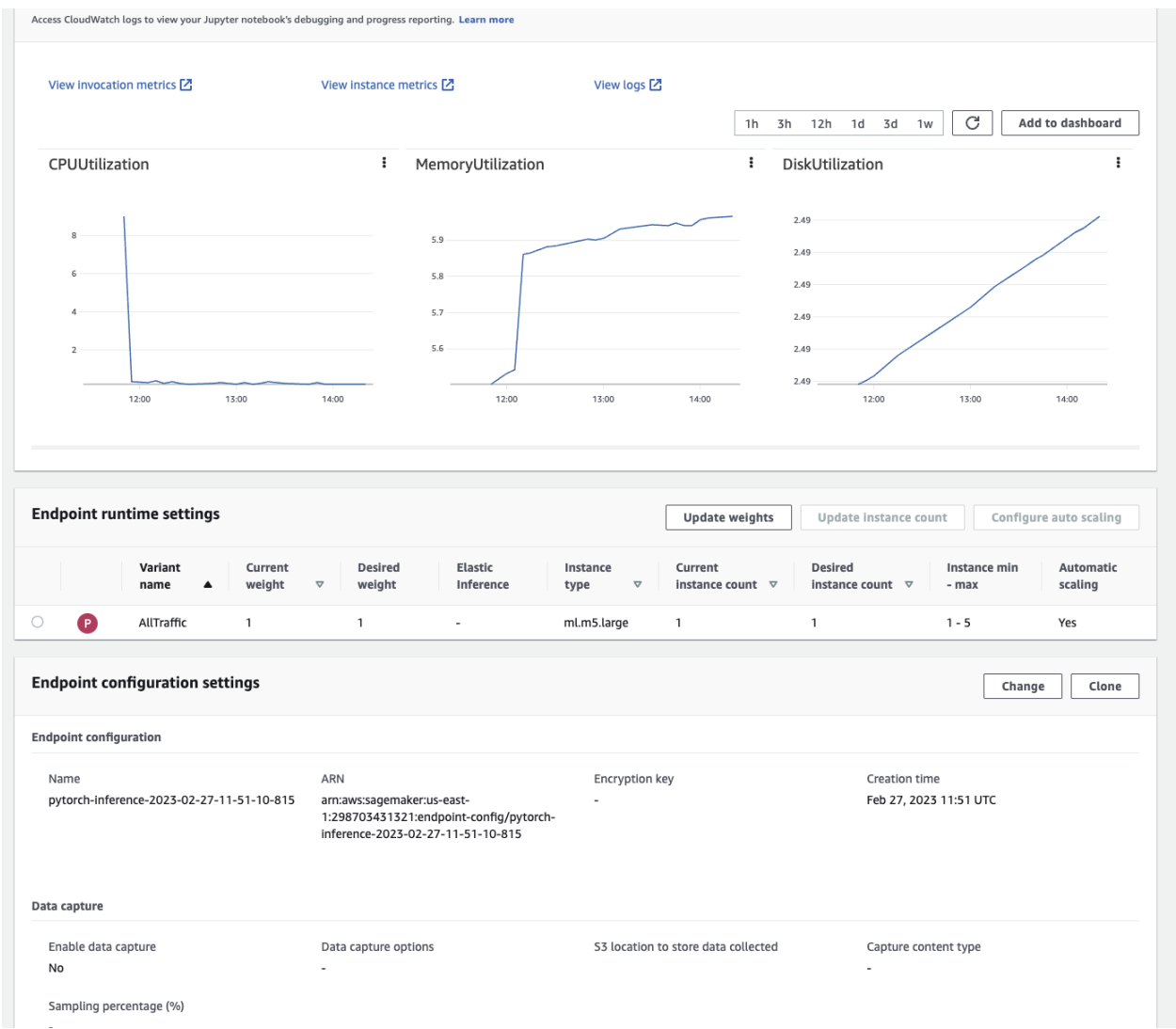


Figure 9: end point auto scaling