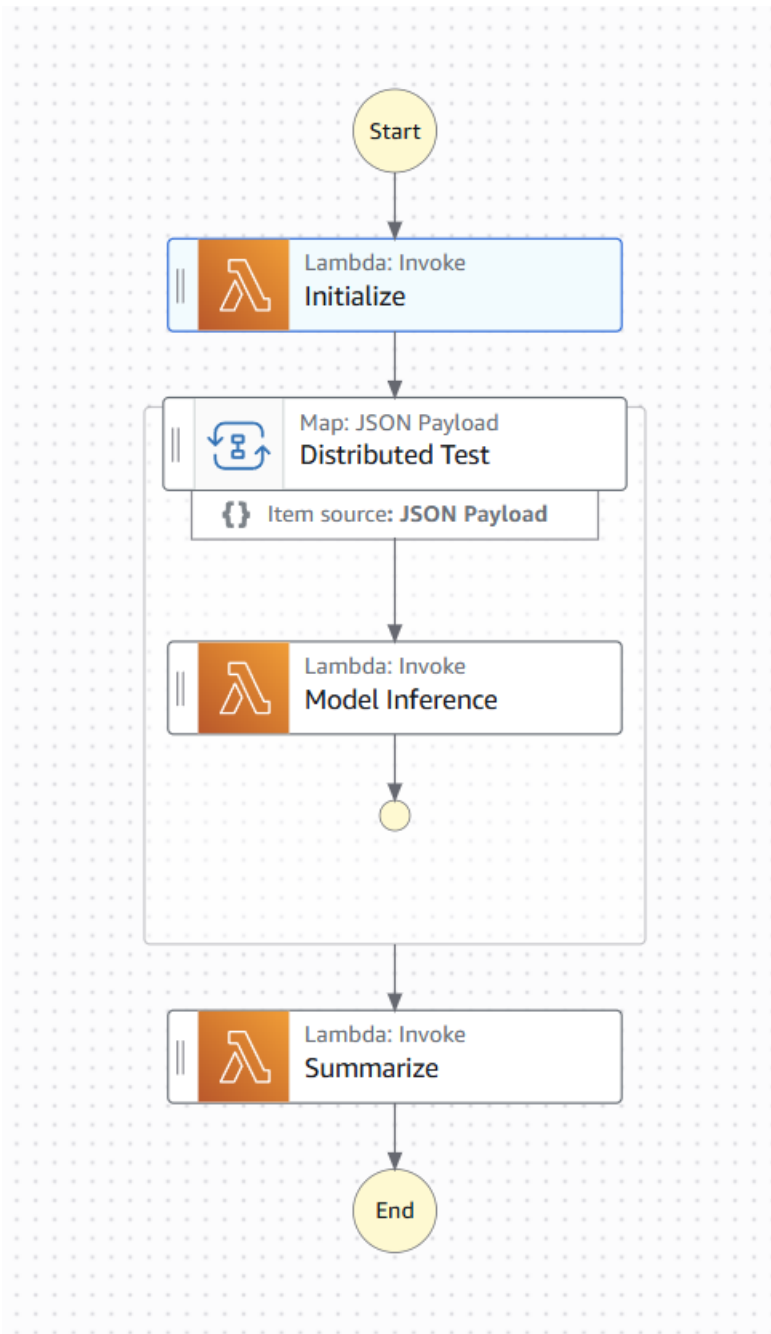


Final Project Step 1

Group 6: Dae Hwang and Pratham Choksi

Screenshots of our designed state machine:



IAM role configurations:

StepFunctions-MyStateMachine-e5ydt2afc-role-z0ymrut89 [Info](#) [Delete](#)

Summary [Edit](#)

Creation date
August 08, 2024, 11:53 (UTC-04:00)

Last activity
 47 minutes ago

ARN
 `arn:aws:iam::211125778552:role/service-role/StepFunctions-MyStateMachine-e5ydt2afc-role-z0ymrut89`

Maximum session duration
1 hour

[Permissions](#) | [Trust relationships](#) | [Tags](#) | [Last Accessed](#) | [Revoke sessions](#)

Permissions policies (6) [Info](#) [Simulate](#) [Remove](#) [Add permissions](#)

You can attach up to 10 managed policies.

Filter by Type
All types

< 1 >

<input type="checkbox"/>	Policy name	Type	Attached entities
<input type="checkbox"/>	AmazonS3FullAccess	AWS managed	14
<input type="checkbox"/>	AWSLambda_FullAccess	AWS managed	5
<input type="checkbox"/>	AWSStepFunctionsFullAccess	AWS managed	2
<input type="checkbox"/>	cloudwatchaccess	Customer inline	0
<input type="checkbox"/>	LambdaInvokeScopedAccessPolicy-...	Customer managed	1
<input type="checkbox"/>	XRayAccessPolicy-fcb083fb-726c-4c...	Customer managed	1

Workflow completion:

[Step Functions](#) > [State machines](#) > [DataParallel-CosmicAI-copy-ab90d7b6-group6](#) > Execution: 6c48c7f3-2913-4faa-90af-b73d809686b6

State transitions [Learn more](#)
5

Execution Logs [Learn more](#)
[CloudWatch Logs](#)

[Graph view](#) | [Table view](#)

Graph view

```
graph TD; Start((Start)) --> Init[AWS Lambda: Initialize]; Init --> Overview[Overview<br/>Map: JSON Payload<br/>Distributed Test<br/>Item source: JSON Payload]; Overview --> Infer[AWS Lambda: Invoke<br/>Model Inference]; Infer --> Sum[AWS Lambda: In<br/>Summarize]; Sum --> End((End));
```

[Actions](#)

In progress Failed Caught error Canceled Succeeded

Screenshot of executions:

Executions (0/14)

View details

Stop execution

Redrive

Start execution

Q Filter executions by property or value

All

Last 15 months

Local timezone

14 matches

< 1 > ⚙

<input type="checkbox"/>	Name	Status	Start Time (local)	End Time (local)	Duration	Info	Version	Alias
<input type="checkbox"/>	01c18276-f4fe-48df-811f-fcd93000ba75	Succeeded	Jul 22, 2025, 00:04:29	Jul 22, 2025, 00:05:23	00:00:53.784		-	-
<input type="checkbox"/>	5f39c394-c5d4-4c7c-8ab1-df8832d739c8	Succeeded	Jul 21, 2025, 23:45:23	Jul 21, 2025, 23:46:16	00:00:53.061		-	-
<input type="checkbox"/>	a08a2603-73c8-46ae-8ebe-1e0c775ec2b6	Succeeded	Jul 21, 2025, 23:28:20	Jul 21, 2025, 23:29:14	00:00:54.002		-	-
<input type="checkbox"/>	ad44882f-8f70-43f0-92c8-1bf4ab292f70	Succeeded	Jul 21, 2025, 23:22:08	Jul 21, 2025, 23:23:01	00:00:52.706		-	-
<input type="checkbox"/>	0ec112e0-6ebd-4c64-a7cf-f7ef7b0945c3	Succeeded	Jul 21, 2025, 22:51:16	Jul 21, 2025, 22:52:07	00:00:51.564		-	-
<input type="checkbox"/>	8edf5533-aea3-42c0-b654-edf717a05868	Succeeded	Jul 21, 2025, 22:43:50	Jul 21, 2025, 22:44:45	00:00:54.660		-	-
<input type="checkbox"/>	5e0e2a75-7edb-4836-be4a-e695276d4b3a	Succeeded	Jul 21, 2025, 22:35:52	Jul 21, 2025, 22:36:49	00:00:56.995		-	-
<input type="checkbox"/>	65ca6f74-5c41-4687-99b1-1d15903459eb	Succeeded	Jul 21, 2025, 02:47:05	Jul 21, 2025, 02:47:59	00:00:53.890		-	-
<input type="checkbox"/>	6c48c7f3-2913-4faa-90af-b73d809686b6	Succeeded	Jul 21, 2025, 02:21:27	Jul 21, 2025, 02:22:26	00:00:58.792		-	-
<input type="checkbox"/>	d01bc6af-a817-4a2b-a979-f63f10e753e1	Succeeded	Jul 20, 2025, 22:39:56	Jul 20, 2025, 22:40:31	00:00:35.421		-	-
<input type="checkbox"/>	9b7cc6b2-c03f-4c31-93c0-5bbc0bb02c21	Succeeded	Jul 20, 2025, 22:37:24	Jul 20, 2025, 22:38:06	00:00:42.362		-	-
<input type="checkbox"/>	9aaddbd1-e788-4132-ba08-4aa3f47350cf	Succeeded	Jul 20, 2025, 22:18:39	Jul 20, 2025, 22:19:44	00:01:05.830		-	-
<input type="checkbox"/>	0aa128f5-bc64-437d-8d2b-d5810a41433b	Succeeded	Jul 20, 2025, 22:16:20	Jul 20, 2025, 22:17:07	00:00:46.206		-	-
<input type="checkbox"/>	e214d2d6-ca95-4862-b9bd-eb681156d4dc	Succeeded	Jul 20, 2025, 19:36:34	Jul 20, 2025, 19:37:16	00:00:41.543		-	-

Performance:

Before we began our experiments, we first investigated the analysis shown on github. While considering workplace settings and what determines a good operational system, we concluded model accuracy and speed is crucial to a well performing application. We needed to make sure our model was efficient while saving cost. For speed, we needed to make sure we were getting high throughput when ingesting data while keeping cost low and fast speed to prevent additional cost. As a result, we decided to stick with 265 batch size since it gave the best throughput-to-inference-time ratio for the 1gb experiment. To compare scaling performances, we chose 8gb and 12.6gb since most datasets on workspace are moderate to relatively large. Below show our results and cost for each experiment.

Partitio n (MB)	Lambda/Fil e Limit	Avg Runtime (s)	Memory (MB)	Data Size (GB)	Cost (\$)	Throughpu t (MB/s)	Batc h Size
25	516	6.403065	1782.90438	12.60	0.0959	1.9674	256
25	327	7.102975	1787.372606	7.98	0.0676	1.1240	256
50	258	11.660477	2626.901905	12.60	0.1287	1.0804	256
50	163	11.521738	2637.320598	7.96	0.0806	0.6908	256
75	173	18.732437	3498.059045	12.67	0.1845	0.6764	256

75	110	17.851546	3491.509553	8.06	0.1116	0.4513	256
100	129	29.539734	4387.226472	12.60	0.2722	0.4265	256
100	82	30.074467	4371.040873	8.01	0.1755	0.2663	256

Different json payloads:

```
{
  "bucket": "cosmicai-data",
  "file_limit": "516",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-25MB/12.6GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "25MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "327",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-25MB/8GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "25MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "258",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-50MB/12.6GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "50MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "163",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-50MB/8GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "50MB"
}
```

```
"data_prefix": "50MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "173",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-75MB/12.6GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "75MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "110",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-75MB/8GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "75MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "129",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-100MB/12.6GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "100MB"
}
{
  "bucket": "cosmicai-data",
  "file_limit": "82",
  "batch_size": 256,
  "object_type": "folder",
  "S3_object_name": "Anomaly Detection",
  "script": "/tmp/Anomaly Detection/Inference/inference.py",
  "result_path": "scaling/result-partition-100MB/8GB/2",
  "data_bucket": "cosmicai-data",
  "data_prefix": "100MB"
}
```