

5110 Big Data

TEAM 1 – STEP FUNCTION

DANIEL ANTHONY, HAROLD HAUGEN, ZACHARY LISMAN, GEORGE SHORIZ

Step Function Execution

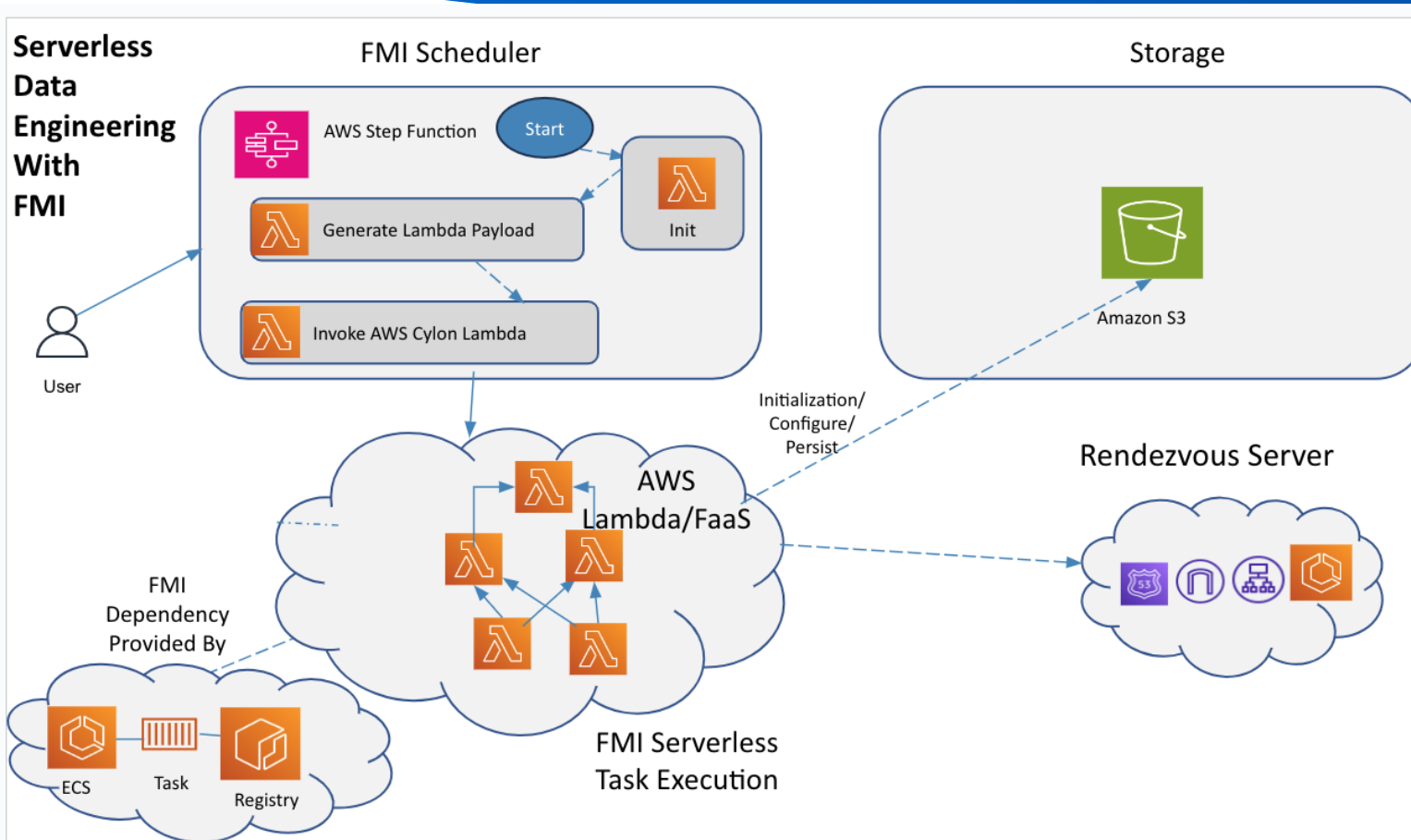
What are Step Functions: Broadly, AWS defines the Step Function process as:

“ With **AWS Step Functions**, you can create workflows, called State machines, to build distributed applications, automate processes, orchestrate microservices, and create data and *machine learning pipelines*. ”

Step Functions is based on state machines and tasks, with State machines being called workflows, which are a series of event-driven steps. Instances of running workflows performing tasks are called executions in Step Functions.

Depending on a use case, you can have Step Functions call AWS services, such as Lambda, to perform tasks. You can also have Step Functions control AWS services, such as AWS Glue, to create extract, transform, and load workflows and create long-running, automated workflows for applications that require human interaction. “

Step Function Execution



The AWS Step Function allows our team to push up Payload Python scripts into AWS Lambda/FaaS for execution.

Step Function Execution

State Machine Execution

Our Team (Team 1) created three State Machine 'executions' as a trial to confirm that:

a) we could create a Workflow in AWS, b) load in a sample Python script to completion, and c) observe how modifications to the 'World_Size' parameter would affect execution time and cost; which will be taken under consideration in future workflow design.

State machines (4)		<input type="checkbox"/> View execution counts		View details	Edit	Copy to new	Delete	
<input type="text" value="Search for state machines"/>		<div>Any type ▼</div>						
	Name ▼	Type ▼	Creation date ▼	Status				
<input type="radio"/>	DS5110StateMachine4	Standard	Sep 25, 2024, 10:48:20 (UTC-04:00)	Active				
<input type="radio"/>	DS5110StateMachine3	Standard	Sep 25, 2024, 10:47:51 (UTC-04:00)	Active				
<input type="radio"/>	DS5110StateMachine2	Standard	Sep 25, 2024, 10:45:29 (UTC-04:00)	Active				
<input type="radio"/>	MyStateMachine-e5ydt2afc	Standard	Aug 8, 2024, 11:53:08 (UTC-04:00)	Active				

We used the 'MyStateMachine-e5ydt2afc' as our primary State machine to build each execution.

Step Function Execution

For each execution, we entered the Lambda Init phase of the workflow and edited the world World_Size parameter.

Team_1_Run3	✓ Succeeded	Oct 6, 2024, 19:18:30	Oct 6, 2024, 19:18:33	00:00:03.554
95827c4c-5016-44ac-9350-54d7831c61e8	✓ Succeeded	Oct 6, 2024, 19:15:58	Oct 6, 2024, 19:16:01	00:00:03.161
b2f46aa0-2a3c-45e6-a912-2c1d8a3b0858	✓ Succeeded	Oct 6, 2024, 19:13:07	Oct 6, 2024, 19:13:12	00:00:05.682

Three Executions performed:

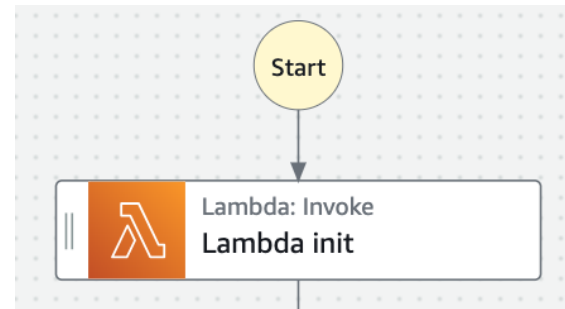
Execution: Team_1_Run3

Details

Execution input and output

Definition

```
1 {
2   "Comment": "A description of my state machine",
3   "StartAt": "Lambda init",
4   "States": {
5     "Lambda init": {
6       "Type": "Task",
7       "Resource": "arn:aws:states:::lambda:invoke",
8       "OutputPath": "$$.Payload",
9       "Parameters": {
10        "FunctionName": "arn:aws:lambda:us-east-1:211125778552:function:fmi_init:$LATEST",
11        "Payload": {
12          "world_size": "10",
13          "bucket": "fmi-lambda-demo",
14          "script": "scripts/fmi-test.py"
15        }
16      }
17    }
18  }
```



For each execution, we updated 'world_size' to values of 2, 10 and 24. Note, for the 'Team_1_Run3' execution, world_size was set to 10. Also, 'bucket' was kept to fmi-lambda-demo'.

Notes:

World size is the number of processes used your training, which is usually the number of GPUs you are using for distributed training.

Rank is a unique ID given to a process, so that other processes know how to identify a particular process. (Stack Overflow)

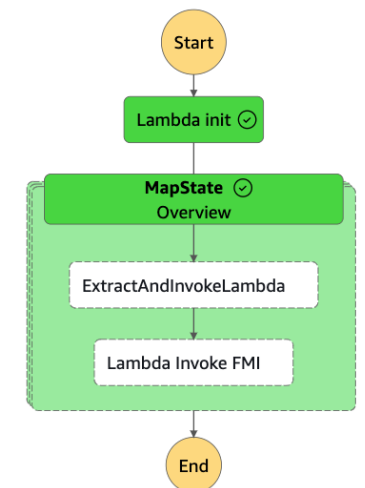
Step Function Execution

Performance Monitoring

Upon monitoring for successful completion (see workflow graph), we reviewed execution metrics captured in the CloudWatch Logs, under the FMI_Executor log group (via Python filters). The following table outlines results for Rank 0 and 1 nodes.

	Exec.	'world_size'	Rank	Request ID	Duration	Billed Duration	Memory Size	Memory Used
1	c61e8	2	0	REPORT RequestId: xxx-6f3a9ce712f0	Duration: 300.54 ms	Billed Duration: 301 ms	Memory Size: 10240 MB	Max Memory Used: 145 MB
2	c61e8	2	1	REPORT RequestId: xxx - 6578018f4698	Duration: 255.44 ms	Billed Duration: 256 ms	Memory Size: 10240 MB	Max Memory Used: 144 MB
3	Team_1_Run3	10	0	REPORT RequestId: xxx-2406456e42b3	Duration: 366.88 ms	Billed Duration: 367 ms	Memory Size: 10240 MB	Memory Used: 145 MB
4	Team_1_Run3	10	1	REPORT RequestId: xxx-f0bc2a87cc74	Duration: 349.07 ms	Billed Duration: 350 ms	Memory Size: 10240 MB	Memory Used: 145 MB
5	b0858	24	0	REPORT RequestId: xxx-ae02f0011f5	Duration: 1045.40 ms Init Duration: 744.94 ms	Billed Duration: 1791 ms	Memory Size: 10240 MB	Max Memory Used: 143 MB
6	b0858	24	1	REPORT RequestId: xxx-180f45b70770	Duration: 1060.50 ms Init Duration: 710.79 ms	Billed Duration: 1772 ms	Memory Size: 10240 MB	Max Memory Used: 143 MB

Monitored each workflow for Successful Completion (with all Green)

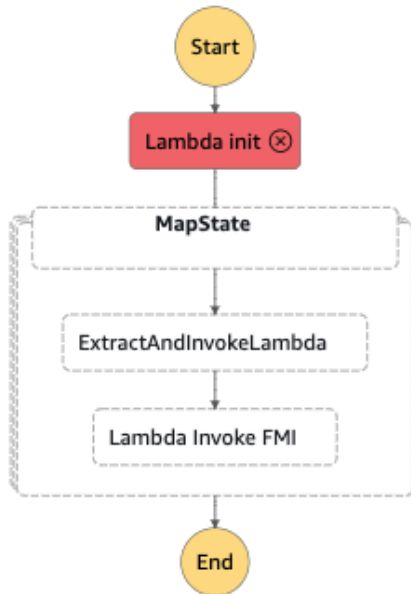


Step Function Execution

Failure Points

During our initial run of the Step Function, we did witness a Failure.

This was apparently caused due to the The Rendezvous server needing to be restarted. A periodic restart should be executed to prevent his issue each week.



Observed Failure

▼ 5	⊗ TaskFailed	Lambda init	00:00:03.477	Oct 4, 2024, 20:36:34.025 (UTC-04:00)
1 ▼	{			
2	"cause": "The cause could not be determined because Lambda did not return an error type. Returned payload {"errorMessage\":\"2024-10-05T00:36:34.010Z 0a1ca044-b598-43ee-902d-91e9243996ca Task timed out after 3.01 seconds\"}",			
3	"error": "Lambda.Unknown",			
4	"resource": "invoke",			
5	"resourceType": "lambda"			
6	}			
▶ 6	⊗ ExecutionFailed		00:00:03.602	Oct 4, 2024, 20:36:34.150 (UTC-04:00)

- ▶ See the Team 1 directory for the README file that describes the project scenario including the following sections.
 - Introduction
 - Data
 - Experimental Design
 - Beyond the original specifications
 - Results
 - Testing
 - Conclusions