

Cloud Computing Lab 9 Report

PES1201800366

Aditeya Baral

1 Configuring AWS EMR Cluster

Summary

Destination

s3://pes1201800366-bucket

Succeeded

2 files, 11.3 MB (100.00%)

Failed

0 files, 0 B (0%)

Files and folders

Configuration

Files and folders (2 Total, 11.3 MB)

Find by name

Name	Folder	Type	Size	Status	Error
food_establishment_data.csv	-	text/csv	11.3 MB	Succeeded	-
health_violations.py	-	text/x-python	2.0 KB	Succeeded	-

Figure 1: S3 bucket with files



Cluster: my-cluster-pes1201800366		Waiting	Cluster ready after last step completed.				
Summary	Application user interfaces	Monitoring	Hardware	Configurations	Events	Steps	Bootstrap actions
Summary			Configuration details				
ID: j-3QTSPNHRCP7KH			Release label: emr-6.2.0				
Creation date: 2021-03-21 09:36 (UTC+5:30)			Hadoop distribution: Amazon				
Elapsed time: 11 minutes			Applications: Spark 3.0.1, Zeppelin 0.9.0				
After last step completes: Cluster waits			Log URI: s3://pes1201800366-bucket/logs/ 				
Termination protection: Off Change			EMRFS consistent view: Disabled				
Tags: -- View All / Edit			Custom AMI ID: --				
Master public DNS: ec2-3-90-105-88.compute-1.amazonaws.com 							
Connect to the Master Node Using SSH							

Figure 2: EMR cluster in running state

2 Managing Amazon EMR Clusters

Cluster: my-cluster-pes1201800366 Waiting Cluster ready after last step completed.

Summary Application user interfaces Monitoring Hardware Configurations Events Steps Bootstrap actions

Concurrency: 1 [Change](#)
After last step completes: Cluster waits

[Add step](#) [Clone step](#) [Cancel step](#)

View Jobs in the Application History Tab

Filter: All steps Filter steps ... 2 steps (all loaded) [Refresh](#)

	ID	Name	Status	Start time (UTC+5:30)	Elapsed time	Log files View logs
<input type="radio"/>	s-2ITG41GG1W8J	Spark application	Completed	2021-03-21 09:53 (UTC+5:30)	36 seconds	View logs

JAR location : command-runner.jar
Main class : None
spark-submit --deploy-mode cluster s3://pes1201800366-bucket/health_violations.py --data_source s3://pes1201800366-bucket/food_establishment_data.csv --output_uri s3://pes1201800366-bucket
Arguments : /myOutputFolder
Action on failure: Continue

Figure 3: Spark job completed

myOutputFolder/ [Copy S3 URI](#)

Objects Properties

Objects (2)
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](#)

[Refresh](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	_SUCCESS	-	March 21, 2021, 09:53:50 (UTC+05:30)	0 B	Standard
<input type="checkbox"/>	part-00000-eafb222c-7b81-4348-9624-414752c5909c-c000.csv	csv	March 21, 2021, 09:53:50 (UTC+05:30)	219.0 B	Standard

Figure 4: Output folder of Spark job