

All instructions for running and compiling the second question are given below:

Q2: Analyzing Social Networks using GraphX/GraphFrame

Steps:

Upload data file and jar file to Amazon S3 bucket

Name	Type	Last modified	Size
j-l4ILX1IZW8GI/	Folder	-	-
output-part2/	Folder	-	-
part2_2.11-0.1.jar	jar	November 16, 2021, 14:02:55 (UTC-06:00)	17.4 KB
Slashdot0902.txt	txt	November 16, 2021, 14:02:52 (UTC-06:00)	10.8 MB

Add a step and run it on the AWS EMR cluster

Cluster: My cluster **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](#)


Filter:

All steps

Filter steps ...

2 steps (all loaded)



	ID	Name	Status	Start time (UTC-6) ▾	Elapsed time	Log files 
<div><div></div><div>▾</div></div>	s-2GNEXVQMG8IEJ	command-runner.jar	Completed	2021-11-16 14:33 (UTC-6)	2 minutes	View logs

JAR location

: command-runner.jar

Main class

: None

Arguments

: spark-submit --deploy-mode cluster --class Part2 s3://bigdata3/part2_2.11-0.1.jar s3://bigdata3/Slashdot0902.txt s3://bigdata3/output-part2

Action on failure

: Continue

Output is obtained in the 'Output' directory in the S3 bucket

Name	Type	Last modified	Size
_SUCCESS	-	November 16, 2021, 14:35:08 (UTC-06:00)	0 B
part-00000	-	November 16, 2021, 14:35:08 (UTC-06:00)	0 B
part-00001	-	November 16, 2021, 14:35:08 (UTC-06:00)	0 B
part-00002	-	November 16, 2021, 14:35:08 (UTC-06:00)	0 B
part-00003	-	November 16, 2021, 14:35:08 (UTC-06:00)	1.3 KB