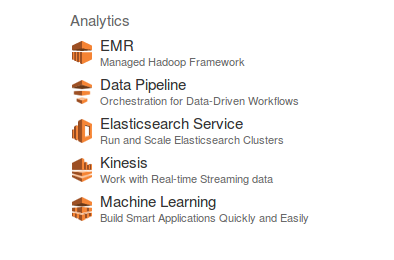


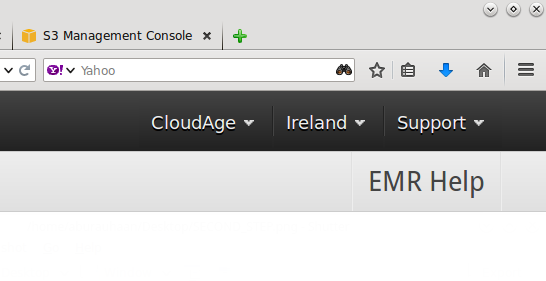
**EMR (ELASTIC MAP REDUCE) Streaming Hadoop Program**

1> Sign in to AWS Console

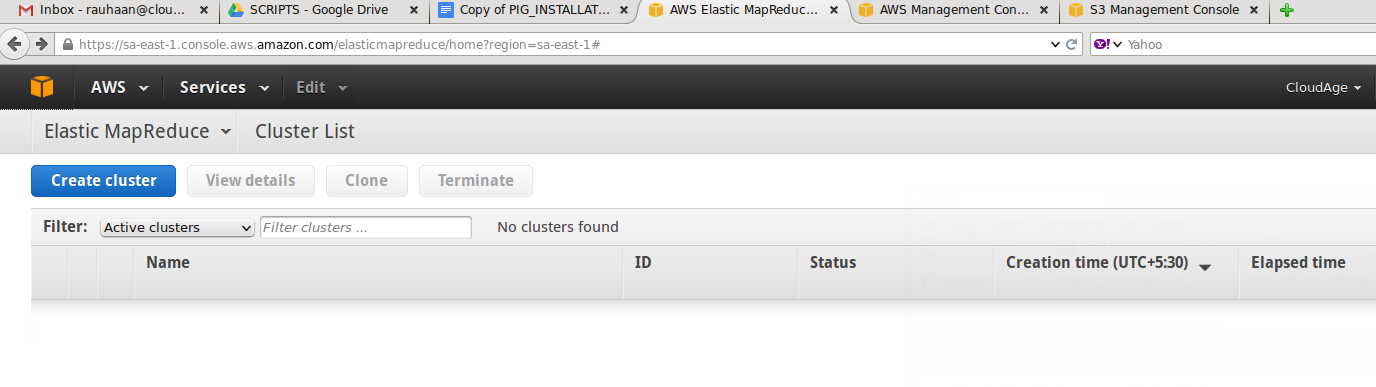
2> Click on EMR on Analytics Section



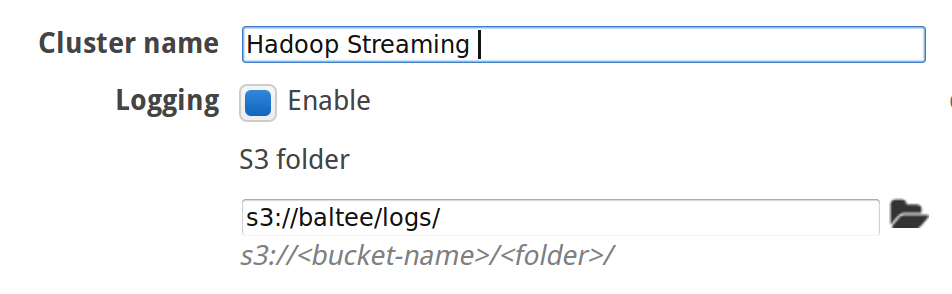
3> Select Data Center



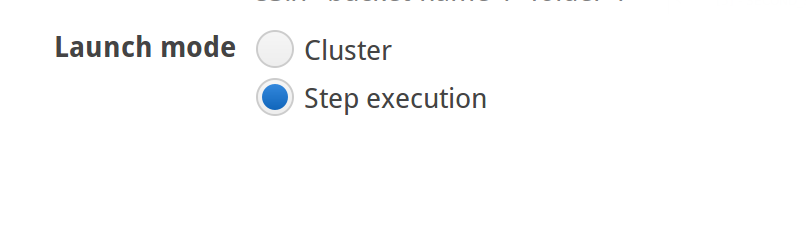
3> Click on Create Cluster



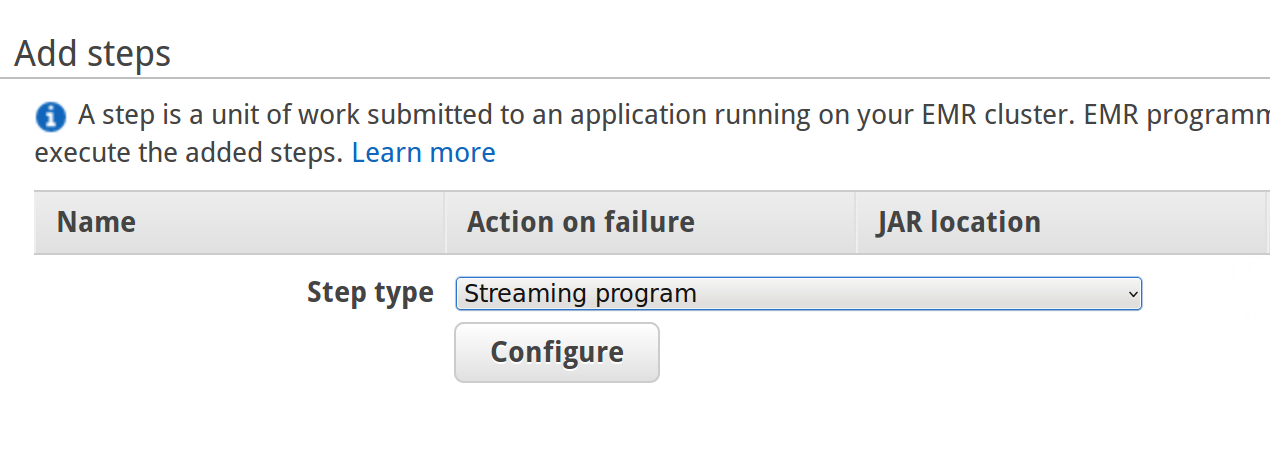
4> Give the Cluster Name Enable Logging and give S3 folder location for Logs



5> Click Step execution as Launch mode



6> Select Step Type as Streaming Program



7> Click on Configure

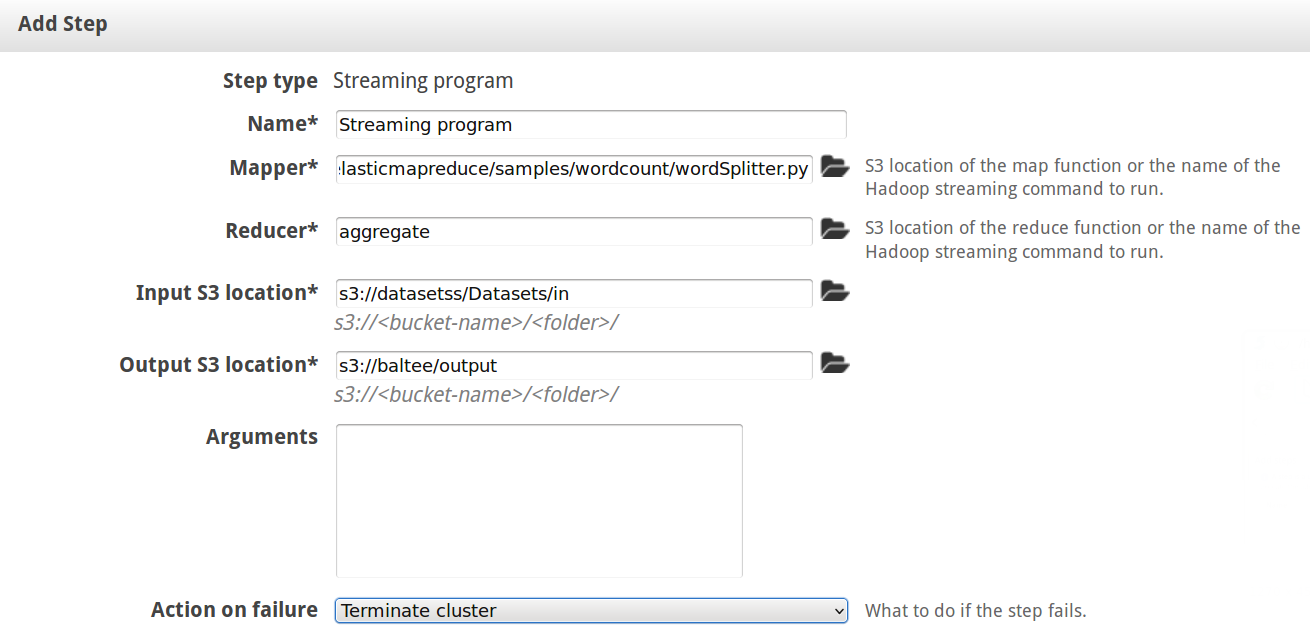
Paste this as Mapper : s3://yourbucket/wordSplitter.py

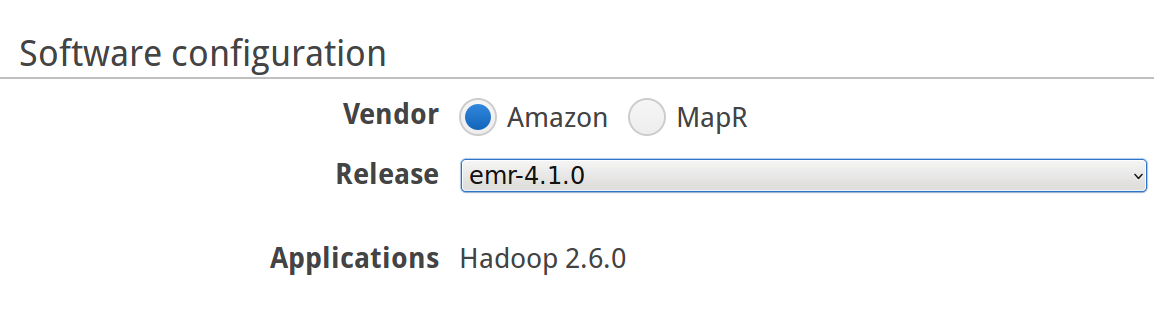
reducer : aggregate

Give Input S3 location (Location of dataset residing on S3)

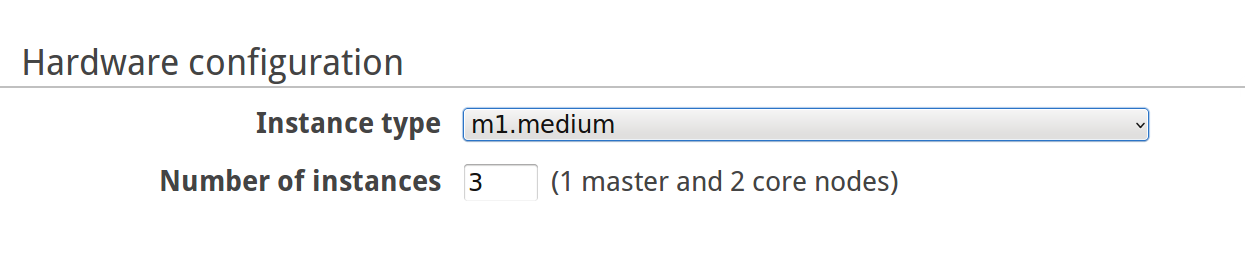
Give Output S3 location (Just give Bucket name and write name of output folder as hadoop will create output Folder. Note: if you select Folder Cluster will terminate with error “Output folder already exists” )

Select Terminate as Action on failure



8> Select Vendor in Software configuration option

9> Select Instance type and Number of instances in Hardware configuration



10> Click on Create Instance