MSDS 7346
Cloud Computing
Mini Project 6 – Log Analysis using AWS EMR

Name: Mooyoung Lee

Log Analysis Using EMR

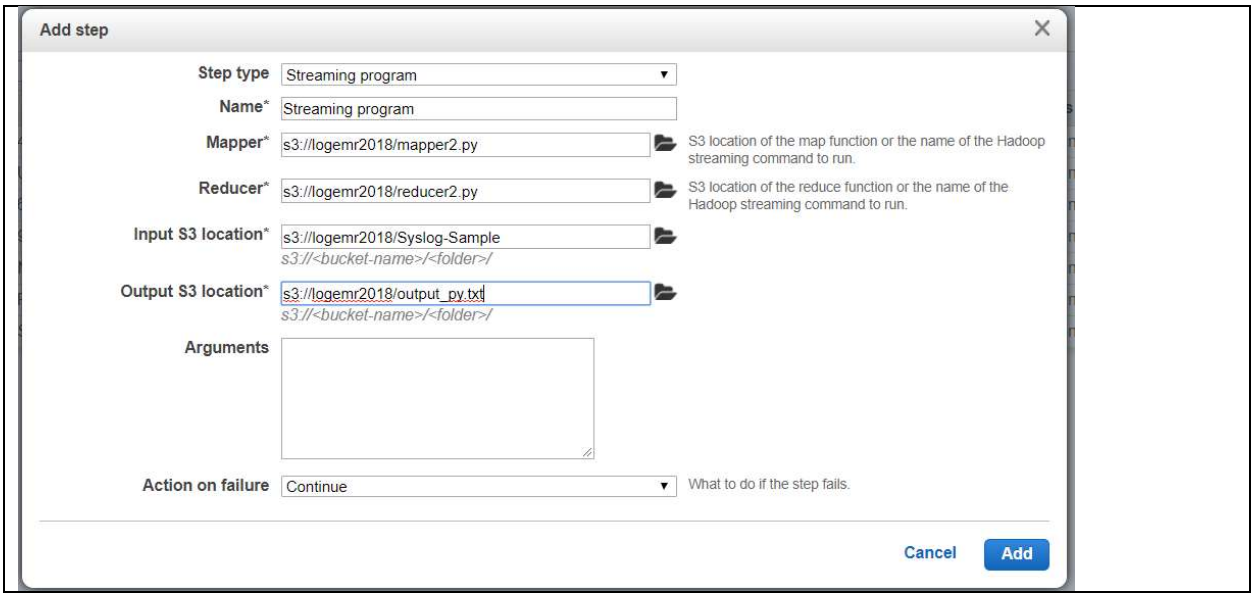
In the live session we will review the building blocks of the MapReduce application. The application should take syslogformatted log records as input and determine the frequency of log events using Amazon EMR to count the number of records per second. I have provided a sample syslog-formatted file as an input.
Although we are keeping this application primarily focused on log analysis, but counting and frequency analysis has many known uses in other data analysis situations. The MapReduce application is performing what is considered a summarization design pattern by simply summing up the values of a common key. Other real-world applications of this technique are:

Load or usage analysis
Many times it is useful to know how many users access a server or a website throughout a time period. Web access logs or application logs that include the time stamps of user events could be imported and processed with a similar MapReduce application to determine usage frequency. At this time we will not be doing load or usage analysis in this lab.

Question 1 : The objective of this lab is to get hands-on experience with developing and deploying MapReduce applications on AWS Elastic Map Reduce. In this lab you are asked to develop two simple MapReduce applications using Pig and Python

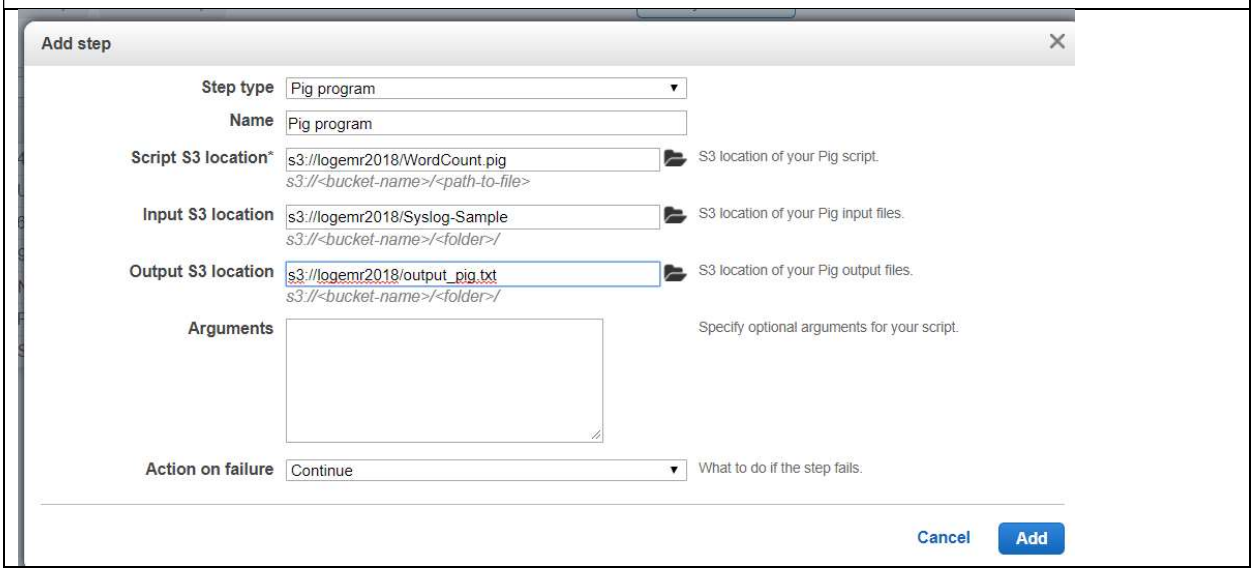
1) Develop MapReduce Application to count number of events per second in the the sample log file (provided) using Python.

- Create an EMR and add Steps.
- Select 'Streaming program' as a Step Type
- Select appropriate files from S3 bucket

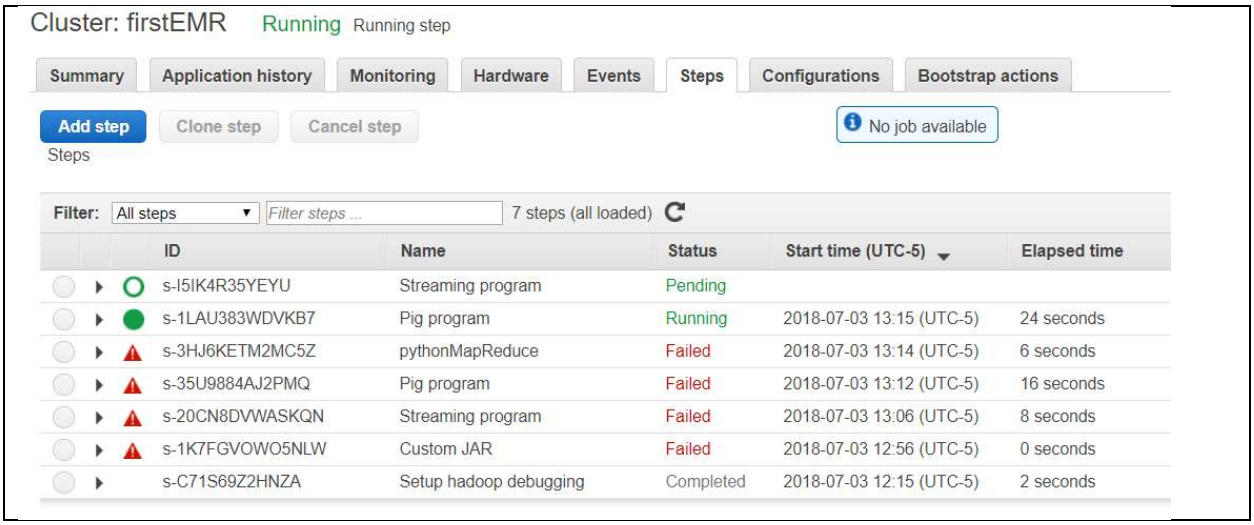


2) Develop MapReduce Application to count number of events per second in the the sample log file (provided) using Pig.

- Add Steps.
- Select 'Pig program' as a Step Type
- Select appropriate files from S3 bucket



3) Deploy both applications in AWS EMR cluster



4) Submit screenshots
Submission: Submit different screen shots to show completion of each steps

Collaborators: None.

Resources: Class lecture