Andrew Abbott

MSDS 7346

Mini-Project 3

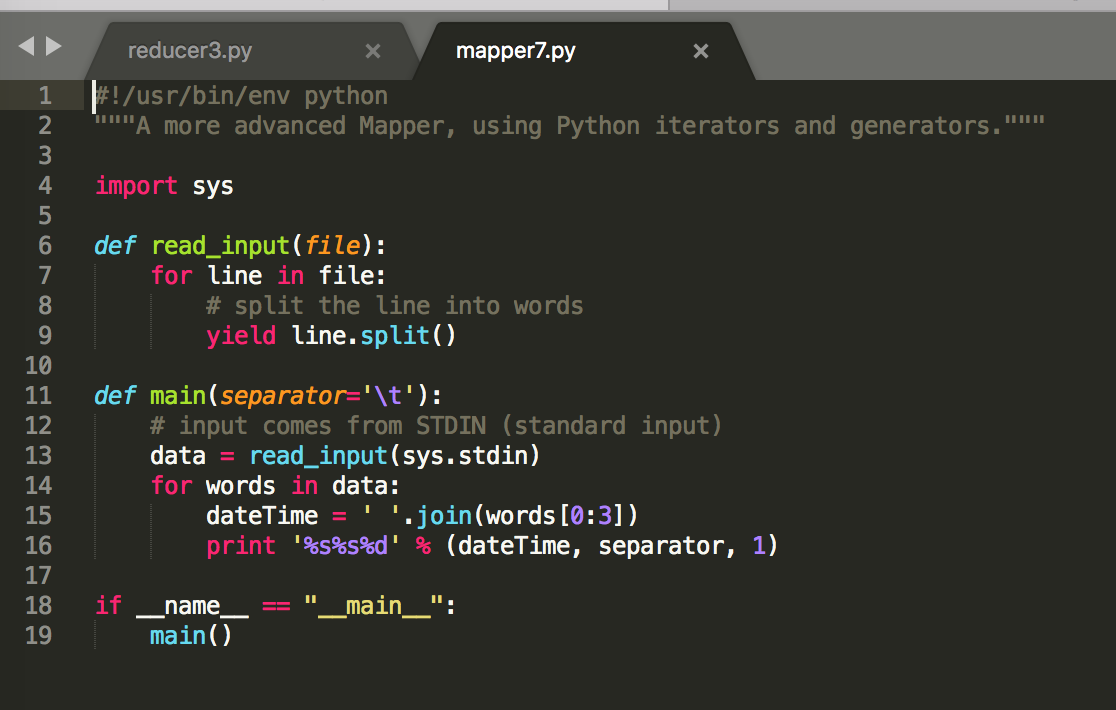
November 12, 2017

Question 1:

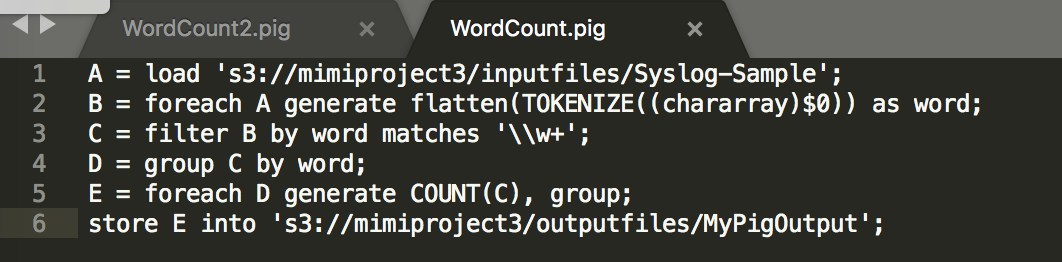
The objective of this lab is to get hands-on experience with developing and deploying MapReduce applications on AWS Elastic Map Reduce. This is simply the extension of what we did in the class. The only new item is enhancing the logic, you all should be familiar with steps of deploying an application in EMR.

In this lab you are asked to develop two simple MapReduce applications using Pig and Python:

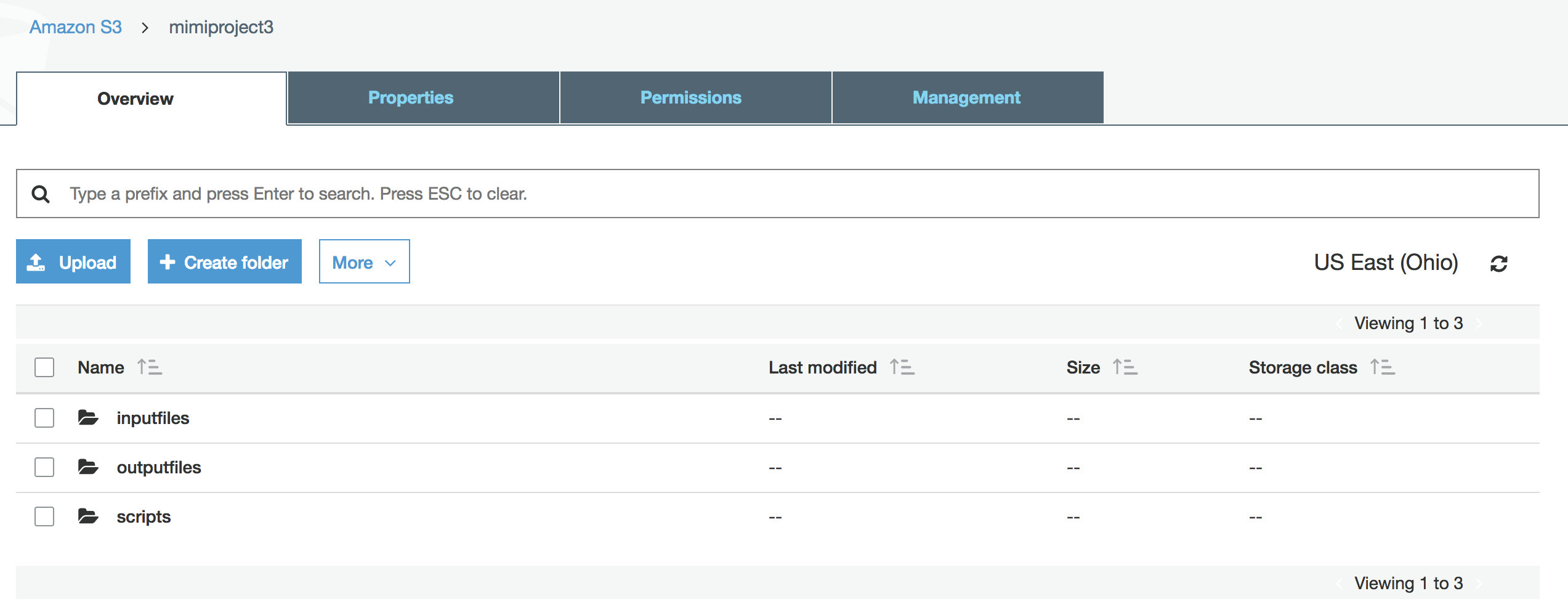
1. Develop MapReduce Application to count number of events per second in the sample log file (provided) using Python.

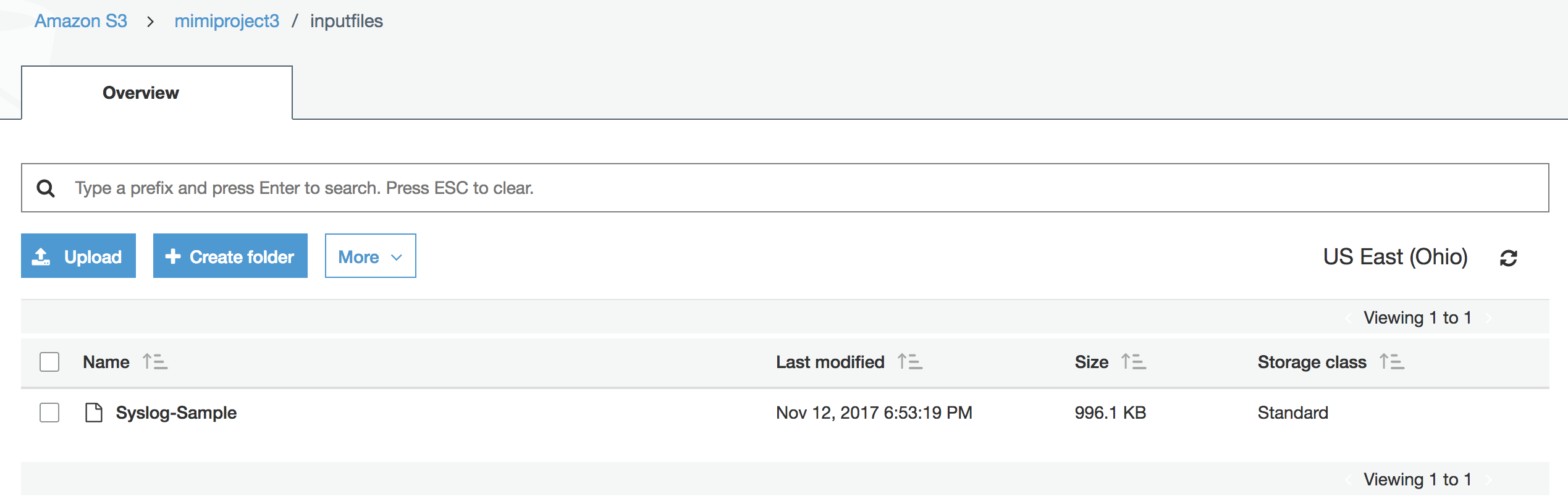


1. Develop MapReduce Application to count number of events per second in the sample log file (provided) using Pig.

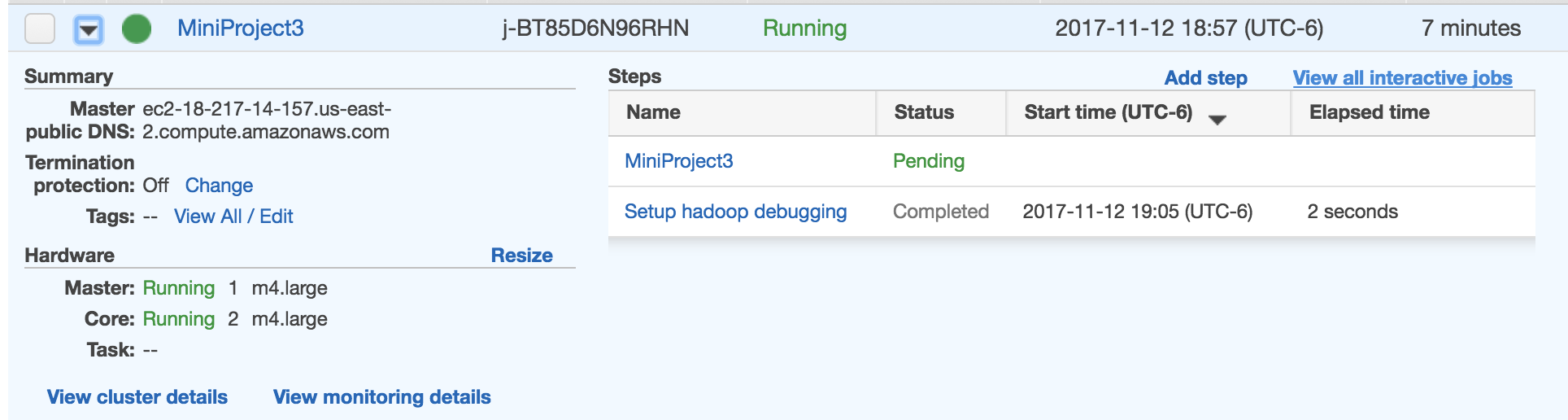
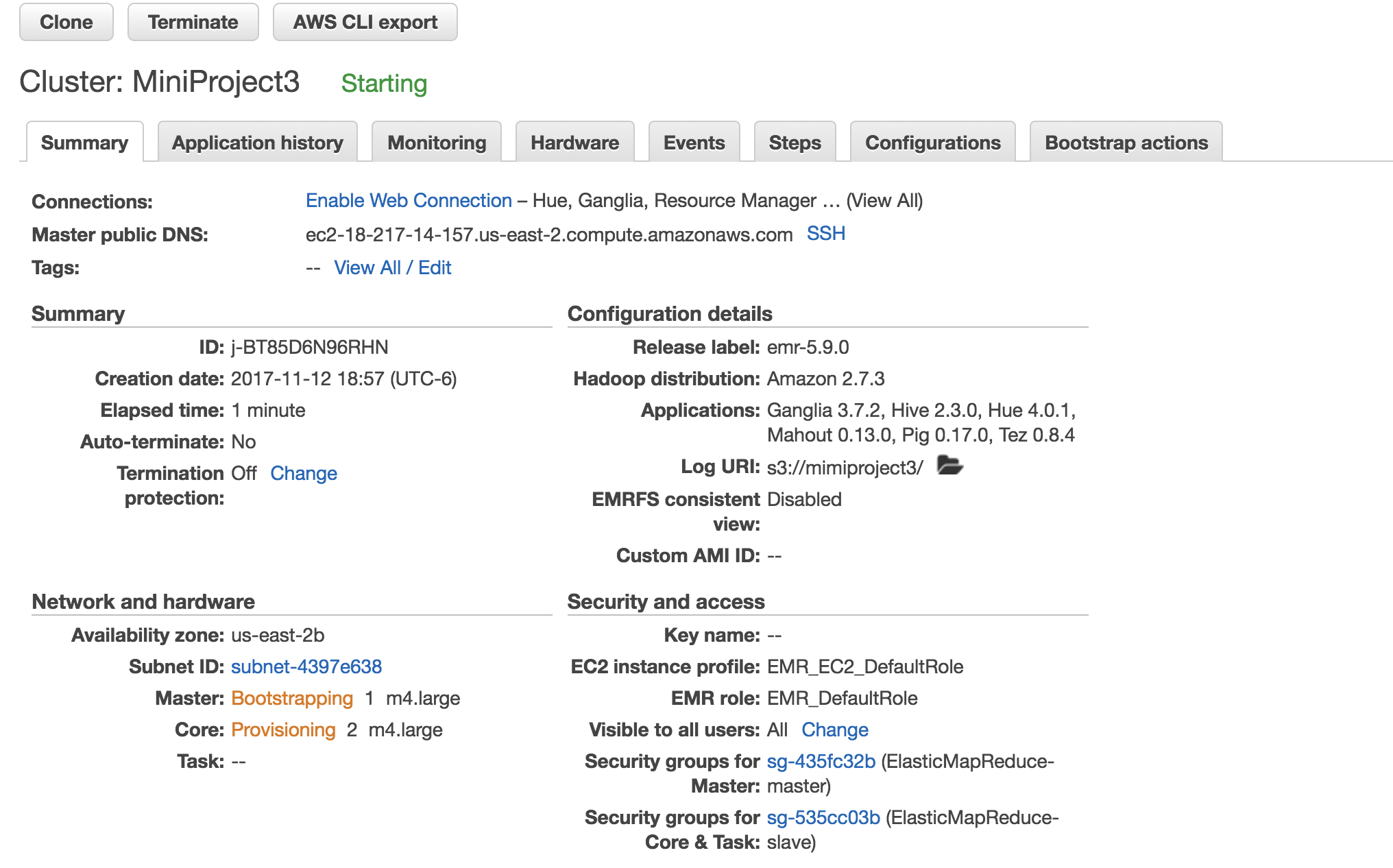


1. Deploy both applications in AWS EMR cluster
   1. Set up the S3

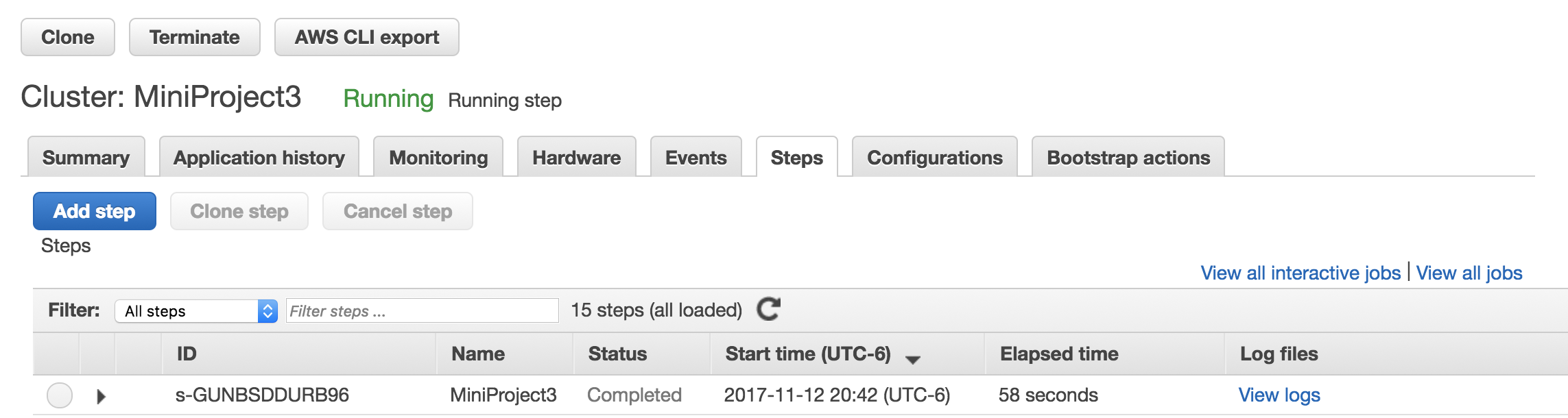




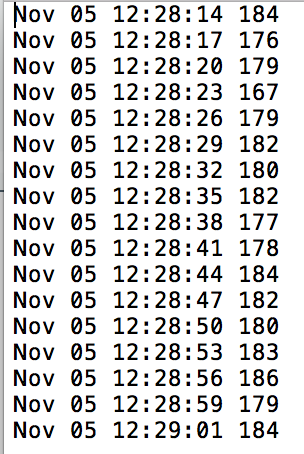
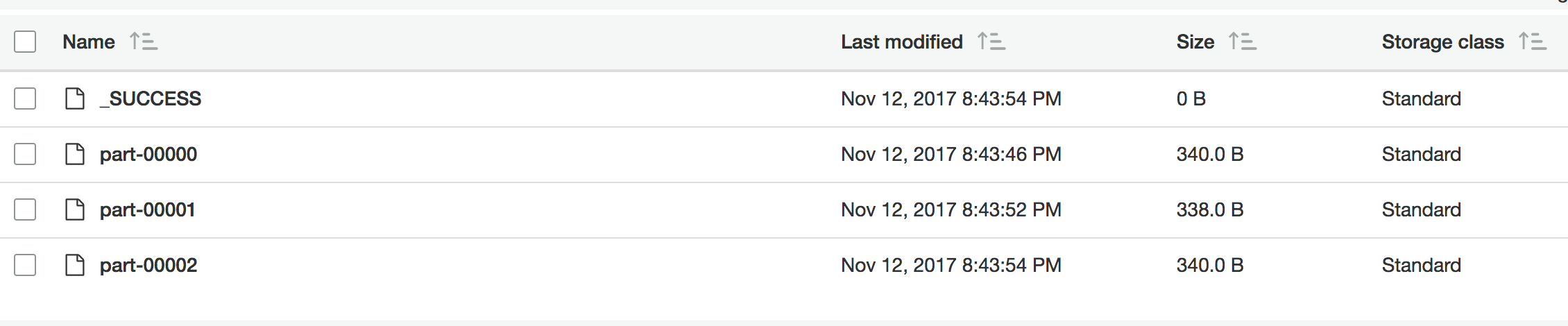
* 1. Set up the EMR Cluster



* 1. Create the step to run the script.



* 1. Output the results:



* 1. Repeat with the Pig script: