Assignment -6

Topic:Hadoop-MapReduce

Team

Abhitej Date   
Rasika Dhanurkar

Sagar Lakhia

To show results obtained from the data set over 25 years to deduce whether the region is getting warmer or dryer over the time span.

We have used the Google cloud to deploy apache Hadoop Mapreduce to demonstrate the whether the region is getting warmer or dryer, and note the time required for a set of mappers and reducers.

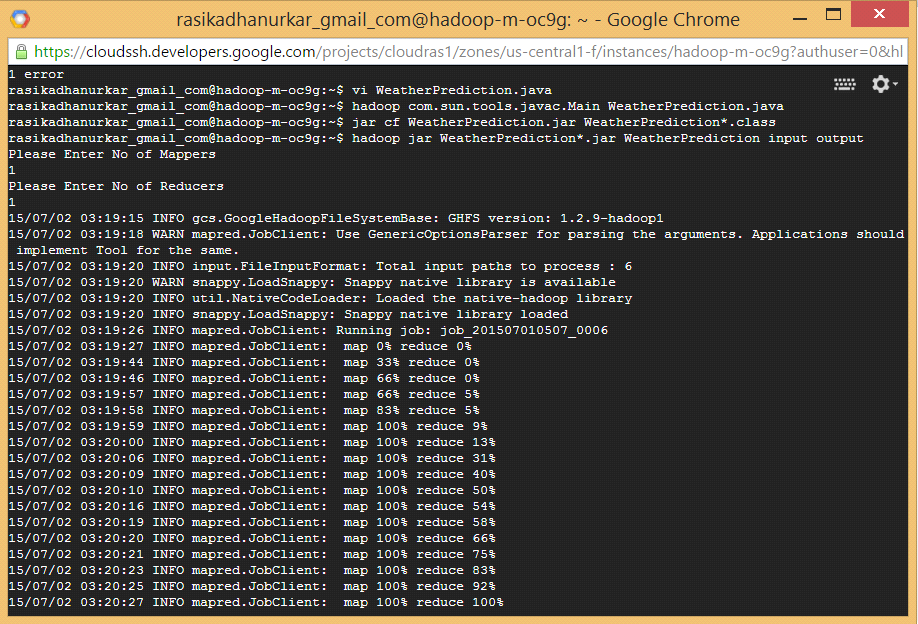
The number of mappers and reducers are taken as input from the users and that input is set as the number of mappers and reducers per transaction. The seasons of the month are sorted in groups of months in spring,summer and winter. 6 input files taken in the bucket.

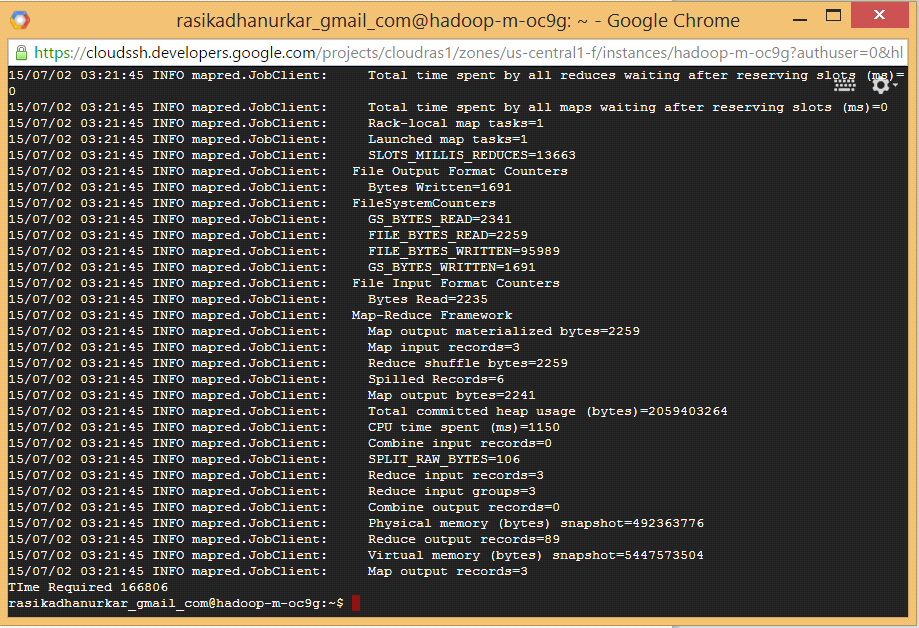
Dataset: ftp://ftp.ncdc.noaa.gov/pub/data/normals/1981-2010/source-datasets/  try picking isdlite-normals.tar

Screenshot:

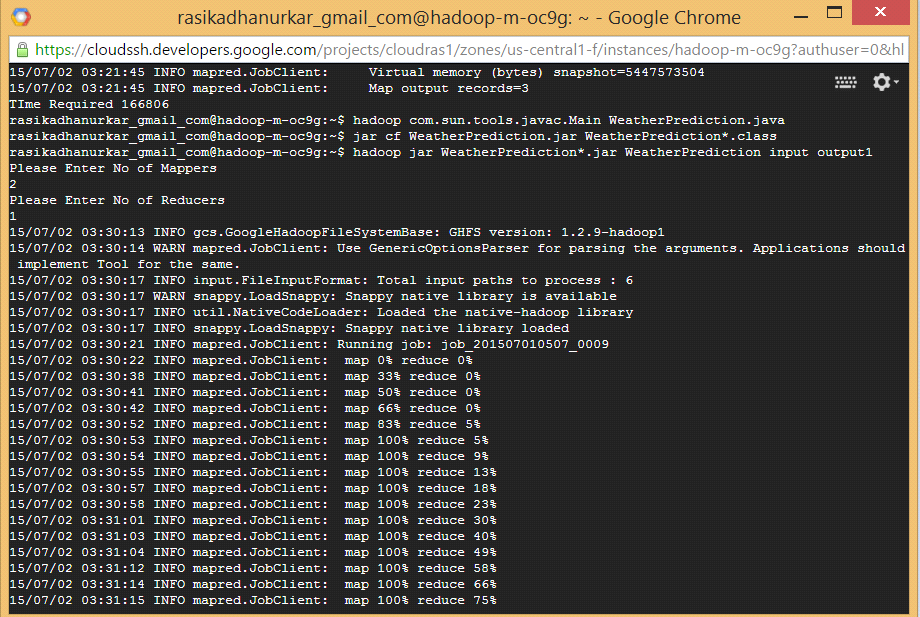
With 2 Instances:

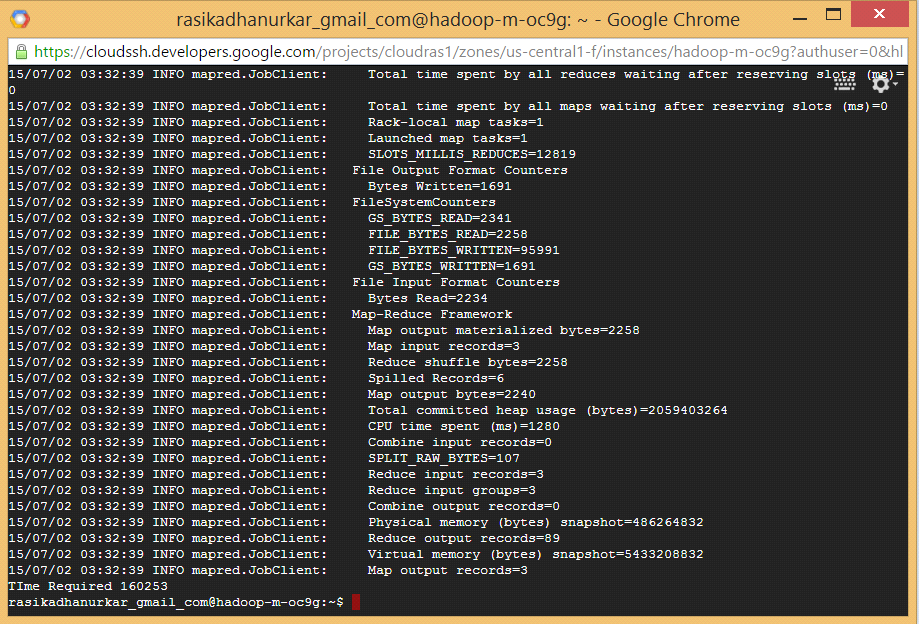
1. For 1 mapper,1 reducer



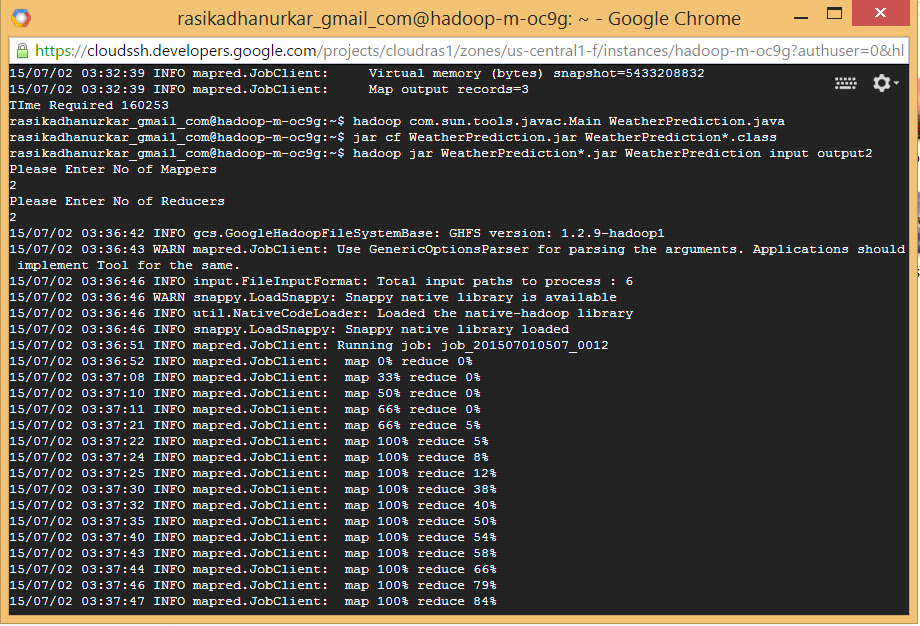


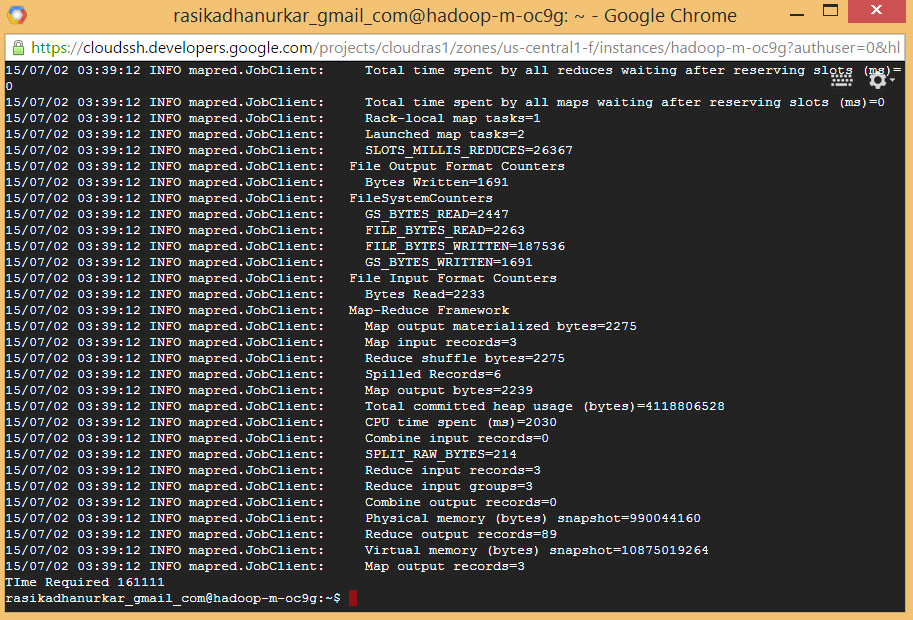
2) For 2 Mappers , 1 Reducer



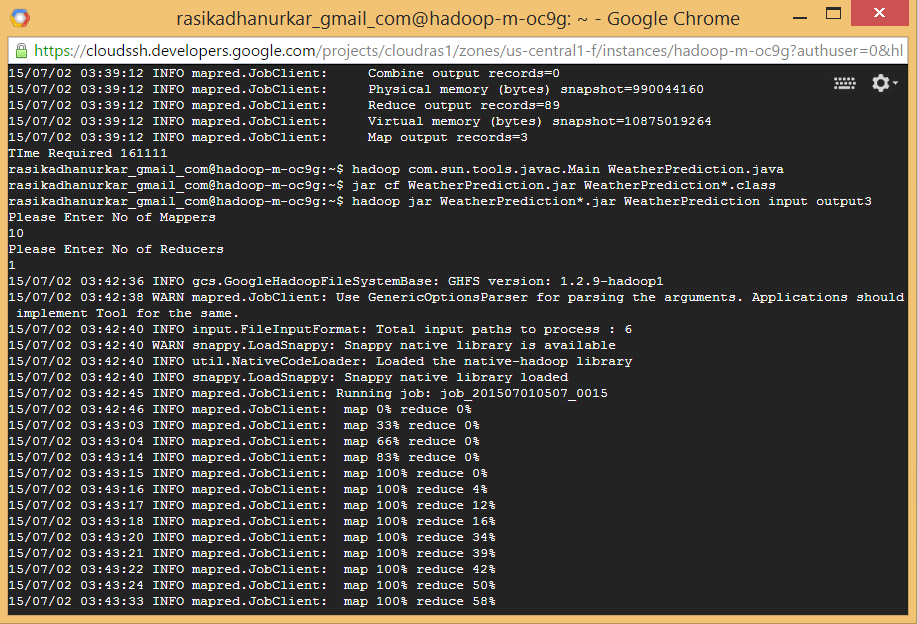


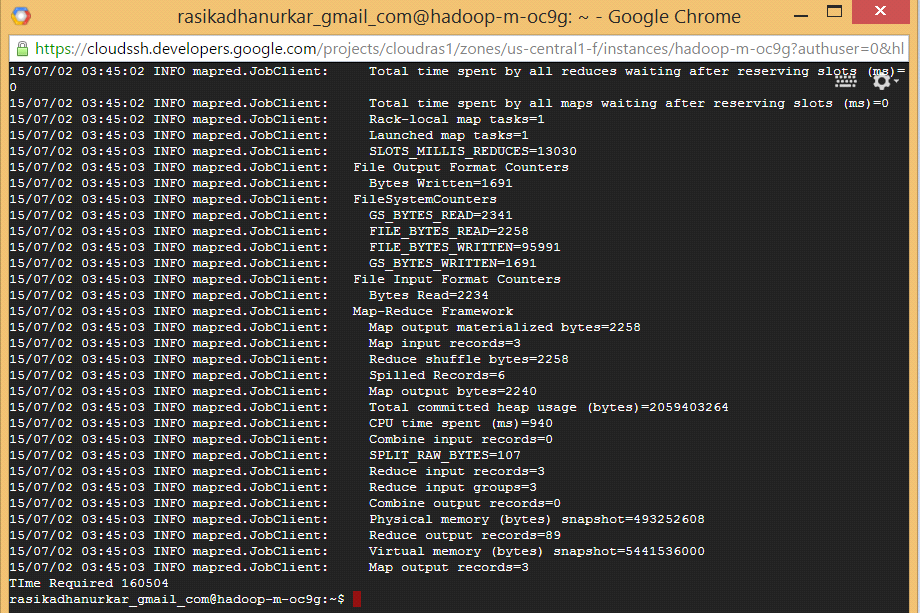
3) For 2 Mappers, 2 Reducers



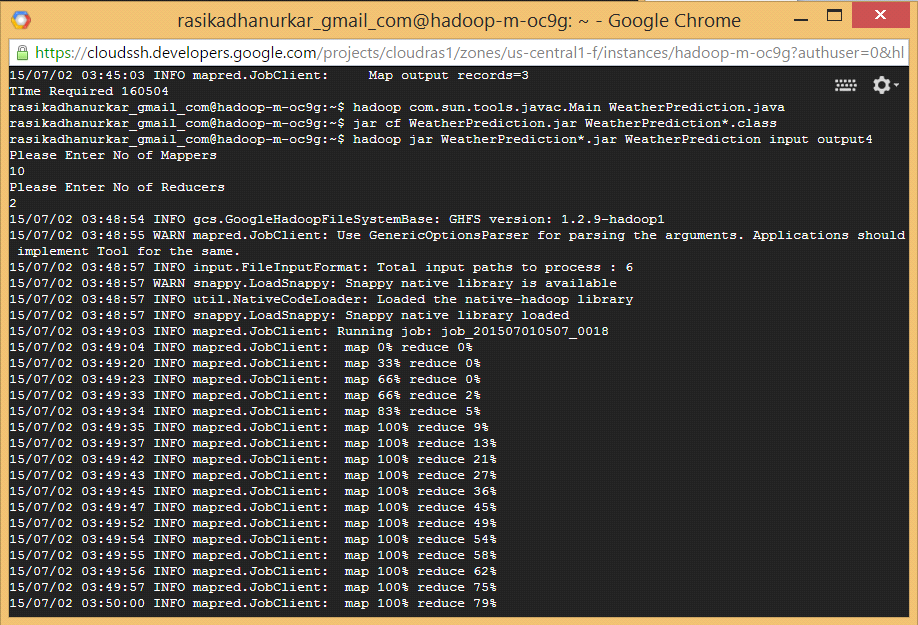


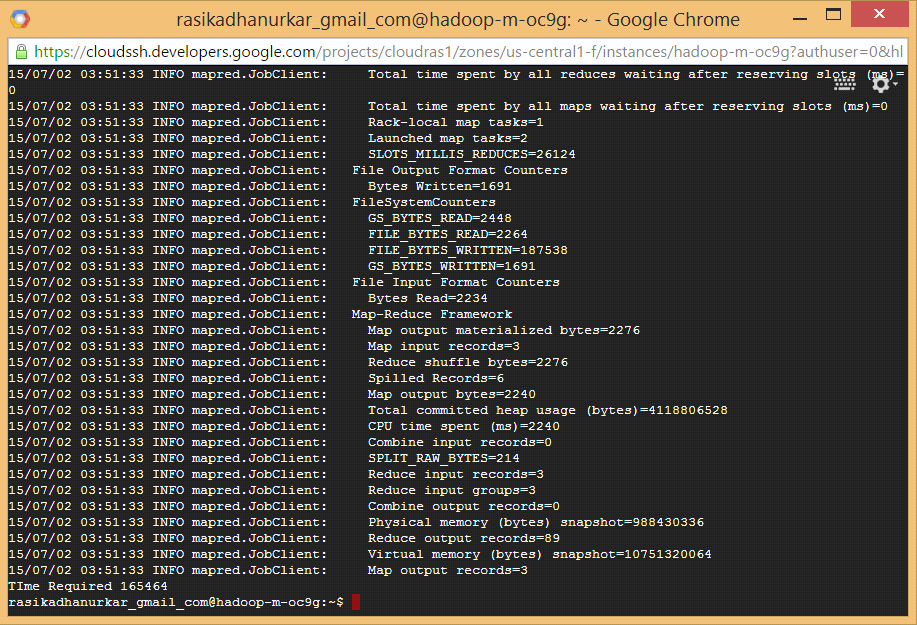
4) For 10 Mappers, 1 Reducers





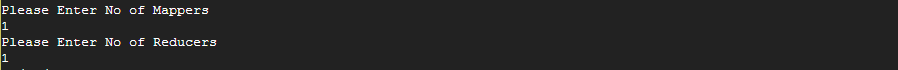
5) For 10 Mappers, 2 Reducers

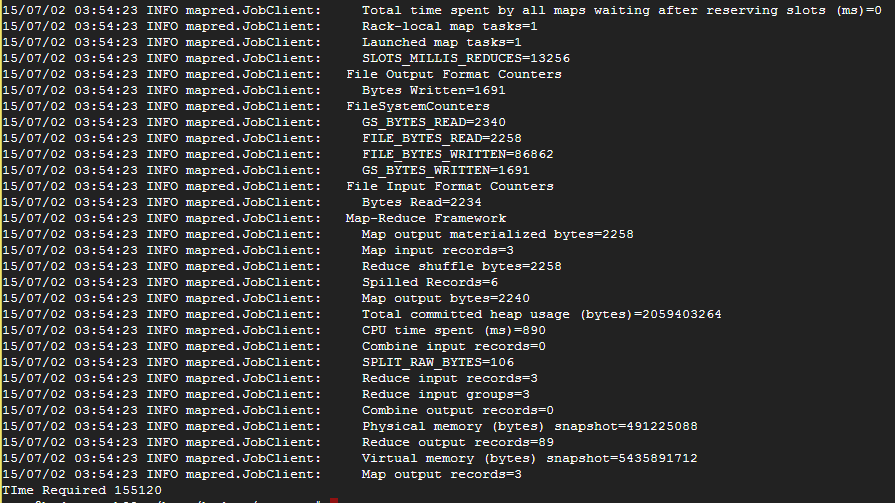




With 1 instance:

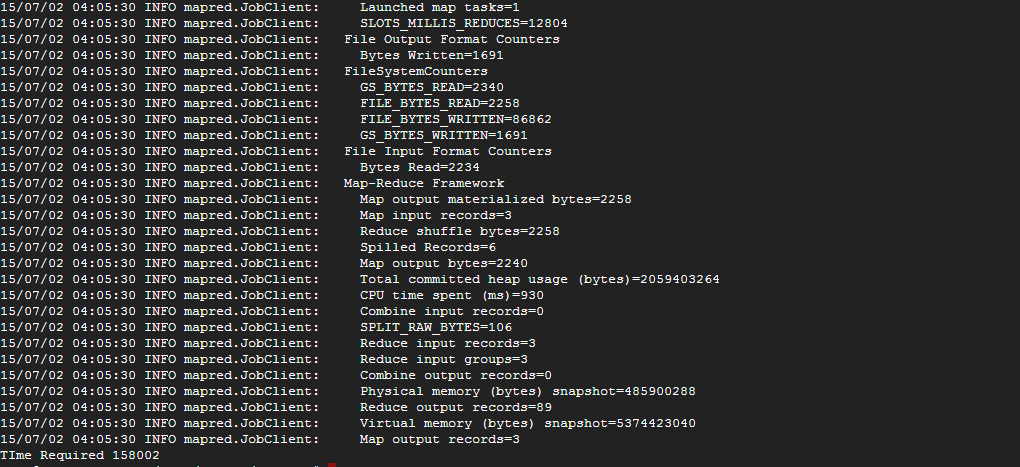
1. For 1 mapper, 1 reducer





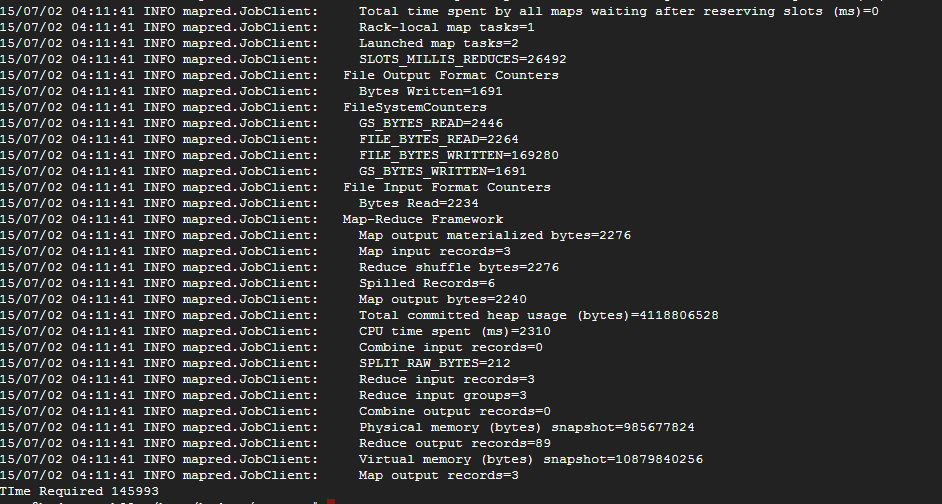
1. For 2 mappers, 1 reducer





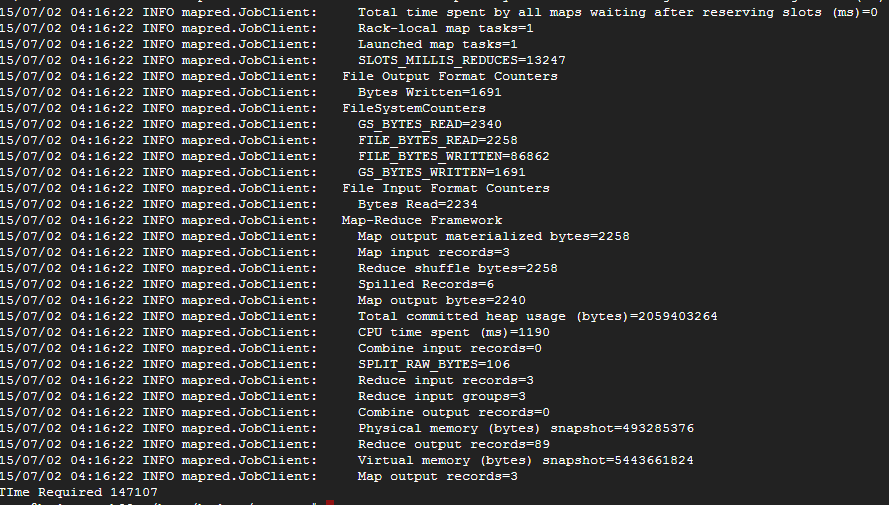
1. For 2 Mappers, 2 Reducers



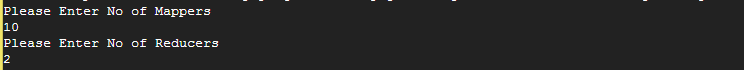


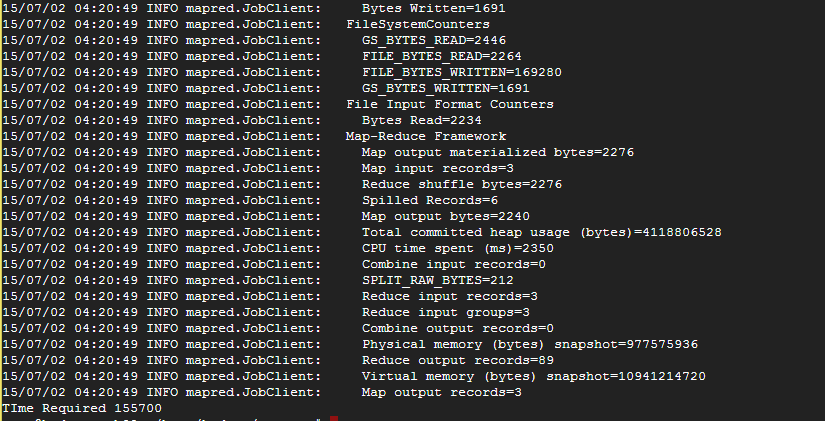
1. For 10 Mappers, 1 Reducer





1. For 10 Mappers, 2 Reducers





References :-

1) Stack Overflow for various doubts and questions

2)https://cloud.google.com/hadoop/getting-started

3)https://www.youtube.com/watch?v=KwW7bQRykHI