**INFSCI 2750 Cloud Computing**

**Mini Project 1**

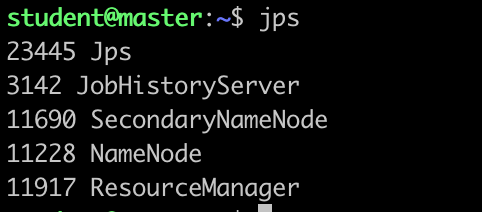
Jing Pang [jip45@pitt.edu](mailto:jip45@pitt.edu)   
Haoyang Qian [haq13@pitt.edu](mailto:haq13@pitt.edu)   
Tian Xue [tix20@pitt.edu](mailto:tix20@pitt.edu)

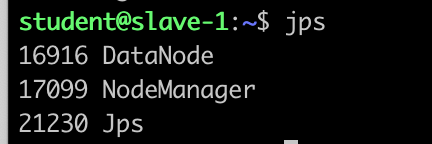
## **Part 1: Hadoop Setup**

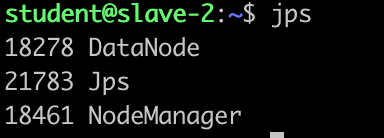
We built the Hadoop cluster following the sequence of preparation steps to install and run Hadoop in our three servers. In this case, we setup our master and slaves in following orders.

export MASTER\_IP=159.65.253.68  
export SLAVE1\_IP=68.183.59.111  
export SLAVE2\_IP=68.183.154.239

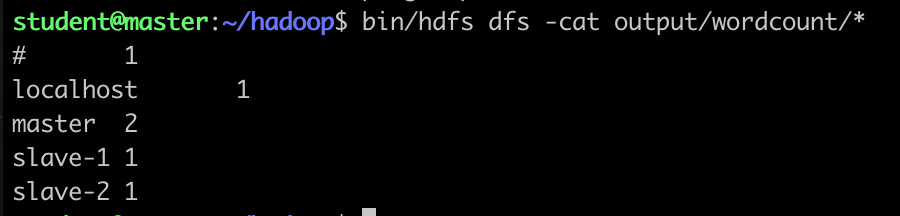
The following graphs show that the Hadoop cluster is successfully starting.





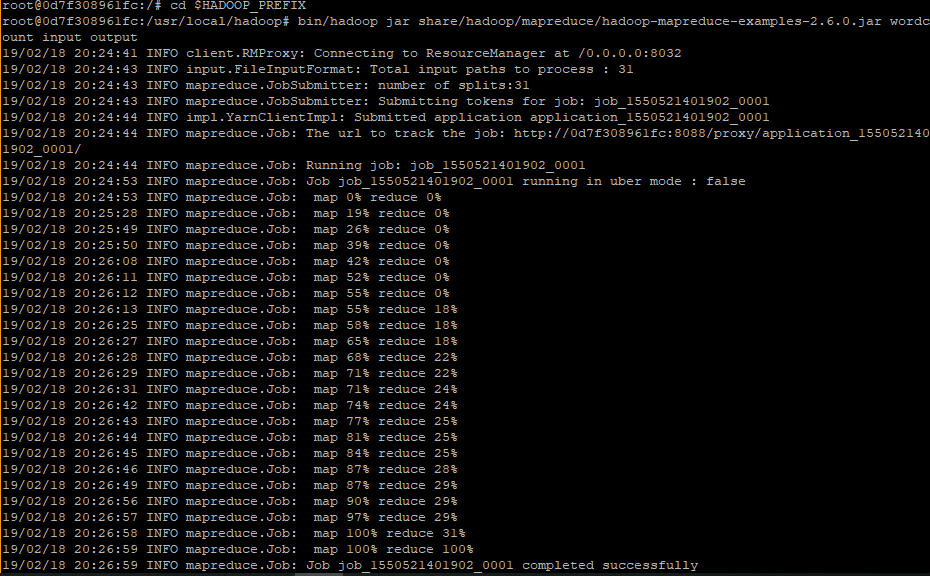


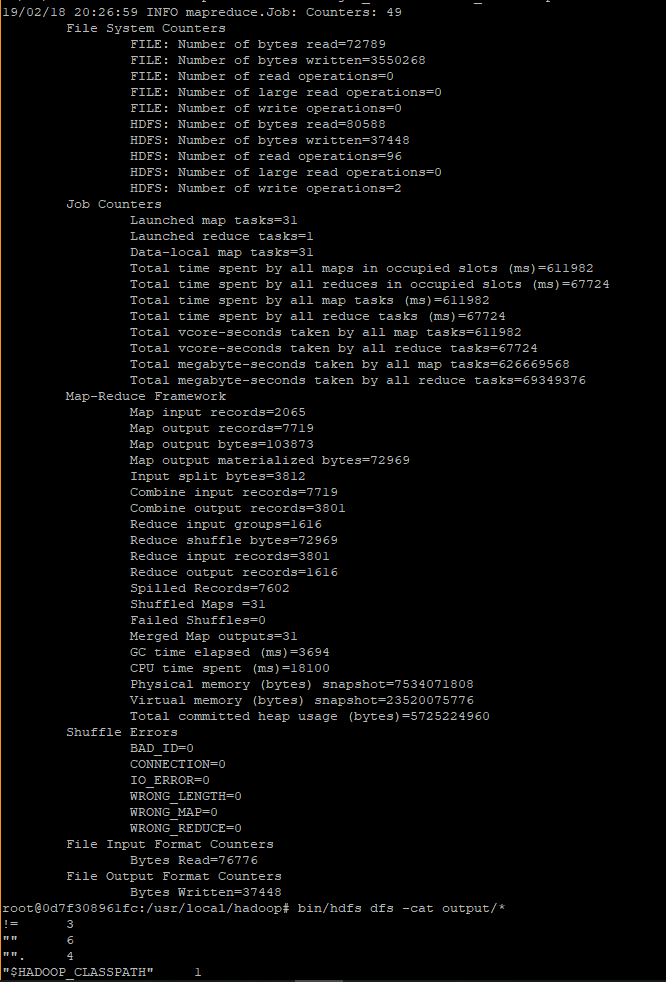
We test our Hadoop cluster with simple example presented by Hadoop default wordcount program.



## **Part 2: Hadoop Docker Image**

In this part, we can build our docker images based on the previous part which can quickly deploy hadoop as we did in the part 1. The support files for the docker image are included in the folder “docker”. We also tested a wordcount job on the built docker image. The result is shown as below.







## **Part 3: Developing Hadoop program (n-gram)**

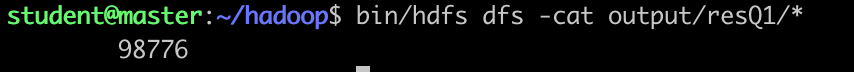
We implemented a mapreduce program to realize the function of n-gram, which is to calculate the n-gram frequencies of the input text file. To run this program, we need to upload a text file to the service, and then input an “n” as a parameter (here we tried 3). Finally, we got part of the output like below:



## **Part 4: Analyzing real logs using Hadoop program**

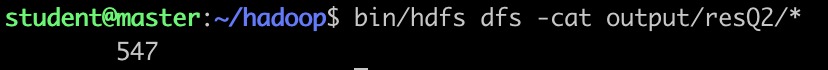
1. How many hits were made to the website item “/assets/img/home-logo.png”?

Answer: 98776



1. How many hits were made from the IP: 10.153.239.5

Answer: 547

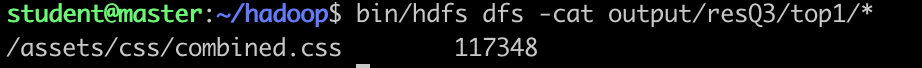


1. Which path in the website has been hit most? How many hits were made to the path?

Answer:

Most hit: “/assets/css/combined.css”

Hits: 117348



1. Which IP access the website most? How many accesses were made by it?

Most IP access the website: 10.216.113.172

# of accesses were made: 158614

