# INFSCI 2750: Cloud Computing Mini Project 1

# Yannan Yuan yay75@pitt.edu

# Xuezhi Xu xux15@pitt.edu

# Ying Zhang yiz165@pitt.edu

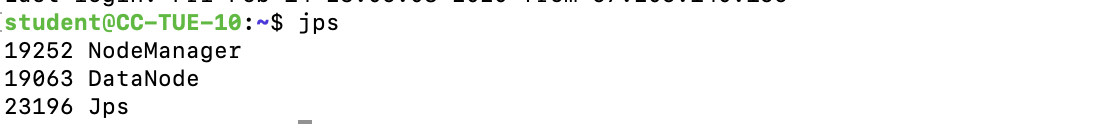
# Part 1: Setting up Hadoop

Set up Hadoop with multinode:

Master:

A close up of a white background

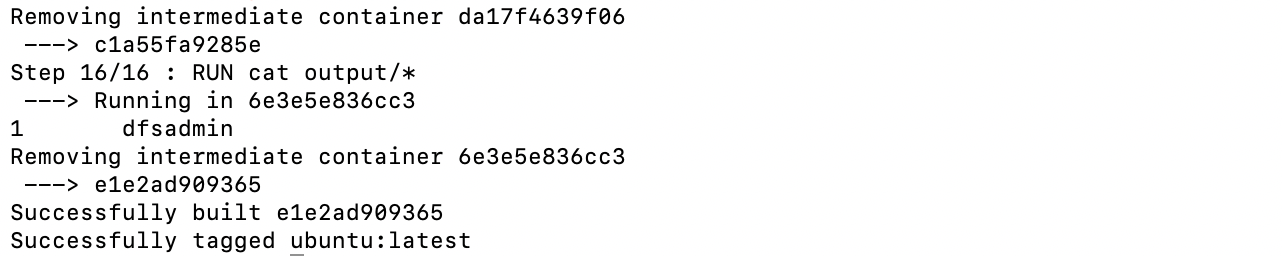
Description automatically generated

Slave: 

# Part 2: Building one Docker image

A close up of text on a white background

Description automatically generated



# Part 3: Developing a Hadoop program

We develop a Hadoop program that produces the n- gram frequencies of the text in a given input file, which includes “helloworld”. We set n to 2. Output is as follow:

A close up of text on a white background

Description automatically generated

A screenshot of a cell phone

Description automatically generated

# Part 4: Developing a Hadoop program to analyze real logs

Solutions:

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

98744

A close up of text on a white background

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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

547

A close up of text on a white background

Description automatically generated

A screenshot of a cell phone

Description automatically generated

# Part 4: Developing a Hadoop program to analyze real logs

Solutions:

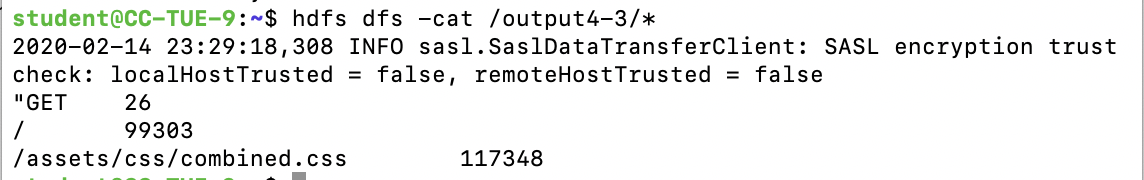
1. Which path in the website has been hit most? How many hits were made to the path?

/assets/css/combined.css

117348

A close up of text on a white background

Description automatically generated



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

10.216.113.172

158614

A close up of text on a white background

Description automatically generated

