**INFSCI 2750: Cloud Computing**

**Mini Project 1**

Pinhao Wang (PIW17) | Zhecheng Qi (zhq27) | Haotian Wu (haw145)

**Part 1:**

The wordcount example has been finished successfully, the output is:

图片包含 文本

描述已自动生成文本

描述已自动生成

Screenshot 1. cat of the output of the worldcount example

**Part 2 Hadoop program - n-gram:**

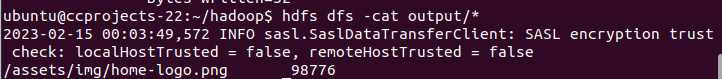
A n-gram Hadoop program has been implemented to produce the n-gram frequencies of the input file with given *n* as a parameter. To test our program, we have performed n-gram with *n* equals to 2 on the input file with simple text “Helloworld”. The *n* must be passed as a parameter in args[2], otherwise an error will occur. The results are shown below:

文本

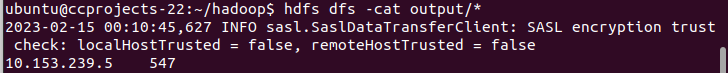
描述已自动生成

Screenshot 2. N-gram result with n = 2

**Part 3 Hadoop program - Log analysis:**

1. How many hits were made to the website item “/assets/img/home- logo.png”?   Answer: 98776 hits 
2. 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: /assets/css/combined.css, with 117348 hits

文本

描述已自动生成

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

Answer:10.216.113.172, with 158614 hits

文本

描述已自动生成

Note: for the questions 3 and 4, we developed Hadoop programs to get top 10 hits by IP/URL