

INFSCI 2750: Cloud Computing

Mini Project 1

Bowen Xu (BOX3), Yesi Xie (YEX4), Wenting Wang (WEW77)

Part 4: Developing a Hadoop program to analyze real logs

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

Answer: 98744 hits were made to the website item “/assets/img/home-logo.png”.

```
Command [genericoptions] [commandoptions]  
[student@CC-AM-27:~/hadoop$ bin/hdfs dfs -cat countwebsite_output/*  
/assets/img/home-logo.png          98744  
student@CC-AM-27:~/hadoop$
```

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

Answer: 547 hits were made from the IP: 10.153.239.5

```
[student@CC-AM-27:~/hadoop$ ~/hadoop/bin/hdfs dfs -cat countIP_output/*  
10.153.239.5      547  
[student@CC-AM-27:~/hadoop$
```

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

Answer: The path “/assets/css/combined.css” has been hit most. 117348 hits were made to the path.

```
student@CC-AM-27:~/hadoop/CountWebsite_classes$ ~/hadoop/bin/hdfs dfs -cat mosthitpath_output/*  
/assets/css/combined.css          117348  
student@CC-AM-27:~/hadoop/CountWebsite_classes$
```

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

Answer: The IP “10.216.113.172” accesses the website most. 158614 accesses were made by it.

```
[student@CC-AM-27:~/hadoop/CountWebsite_classes$ ~/hadoop/bin/hdfs dfs -cat mosthitip_output/*  
10.216.113.172  158614  
[student@CC-AM-27:~/hadoop/CountWebsite_classes$
```