Cloud Computing MiniProject 1 Report

Ruoxi Zhang   ruz39@pitt.edu
Kenny Wu   kew143@pitt.edu
Zhanghaoxiang Yin   zhy77@pitt.edu

Part 3. Ngram results:

We used all .xml files in /etc/hadoop as the input.

```
student@CC-MON-31:~/hadoop$ hdfs dfs -cat ngram_output/*
2020-02-14 20:32:10,083 INFO sasl.SaslDataTransferClient: SA
!--K    1
!--P    5
!--T    1
!--Y    1
");y    1
"*"m    21
".As    1
"1.0    8
"ASI    9
"BAS    9
"Lic    9
"UTF    5
"ali    21
"clu    1
"con    4
"enc    5
"hre    4
"mea    21
"ofa    1
"tex    4
%use    2
'she    1
'tex    1
(ASF    1
(roo    1
(the    9
);yo    1
)und    1
*"me    21
*,/h    1
*</v    42
+20=    1
,/ho    1
,CLA    1
,CPU    1
,HAD    5
,Ver    9
,[us    1
,aft    1
,and    1
,bob    21
,eit    9
,for    2
,ift    1
,nex    1
,sof    9
,the    10
,u:%    1
,use    15
,whe    21
,whi    5
,wil    1
--KM    1
--Pu    5
--Th    1
```

Part 4.

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

```
[student@CC-MON-31:~/hadoop$ hdfs dfs -cat output1/*
 2020-02-14 22:16:18,865 INFO sasl.SaslDataTransferCli
 /assets/img/home-logo.png        98744
```

98744 hits were made to "/assets/img/home-logo.png".

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

```
[student@CC-MON-31:~/hadoop$ hdfs dfs -cat output2/*
 2020-02-14 22:17:00,465 INFO sasl.SaslDataTransferCl
 10.153.239.5     547
```

547 hits were made from IP: 10.153.239.5.

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

```
[student@CC-MON-31:~/hadoop$ hdfs dfs -cat output3/*
 2020-02-14 20:03:59,387 INFO sasl.SaslDataTransferCl
 "GET     26
 /        99303
 /assets/css/combined.css        117348
```

"/assets/css/combined.css" has been hit most in the website.
117348 hits were made to this path.

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

```
[student@CC-MON-31:~/hadoop$ hdfs dfs -cat output4/*
2020-02-14 20:11:31,013 INFO sasl.SaslDataTransferCl
10.1.1.113        1
10.1.1.125        12
10.1.100.183      14
10.1.100.199      35
10.1.109.54       44
10.1.11.81        69
10.1.12.172       93
10.1.122.130      109
10.1.146.91       321
10.1.181.142      10686
10.102.126.202    12298
10.104.62.79      14807
10.118.250.30     18282
10.169.128.121    22516
10.173.141.213    47503
10.216.113.172    158614
```

10.216.113.172 accesses the website most.
158614 accesses were made by it.