

Name: Samaa Khair
SID: 201901481

Assignment One

The image shows two screenshots from a Hadoop environment. The top screenshot is a web browser view of the Hadoop Distributed File System (HDFS) interface at localhost:50070. It displays the 'Browse Directory' page for the path '/pageRank/InlinksFileee1'. The interface includes a navigation bar with tabs like Overview, Datanodes, and a table listing files. The table has columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. Two entries are shown: a file named '_SUCCESS' with a size of 0 B, and a file named 'part-00000' with a size of 1.27 KB. The bottom screenshot shows a file viewer window titled 'part-00000(9)' displaying a list of domain names and their corresponding counts. The domains are listed in ascending order of count, with '1000notes.com' having a count of 1 and 'fahric.com' having a count of 1. The viewer includes a search bar, a 'Save' button, and a status bar at the bottom indicating 'Plain Text', 'Tab Width: 8', and 'Ln 1, Col 1'.

localhost:50070/explorer.html#/pageRank/InlinksFileee1

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/pageRank/InlinksFileee1 Go

Show 25 entries Search:

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|------------|-----------|------------|---------|---------------|-------------|------------|------------|
| -rw-r--r-- | hadoopusr | supergroup | 0 B | Dec 31 15:27 | 1 | 128 MB | _SUCCESS |
| -rw-r--r-- | hadoopusr | supergroup | 1.27 KB | Dec 31 15:27 | 1 | 128 MB | part-00000 |

Showing 1 to 2 of 2 entries Previous 1 Next

Open part-00000(9) Save

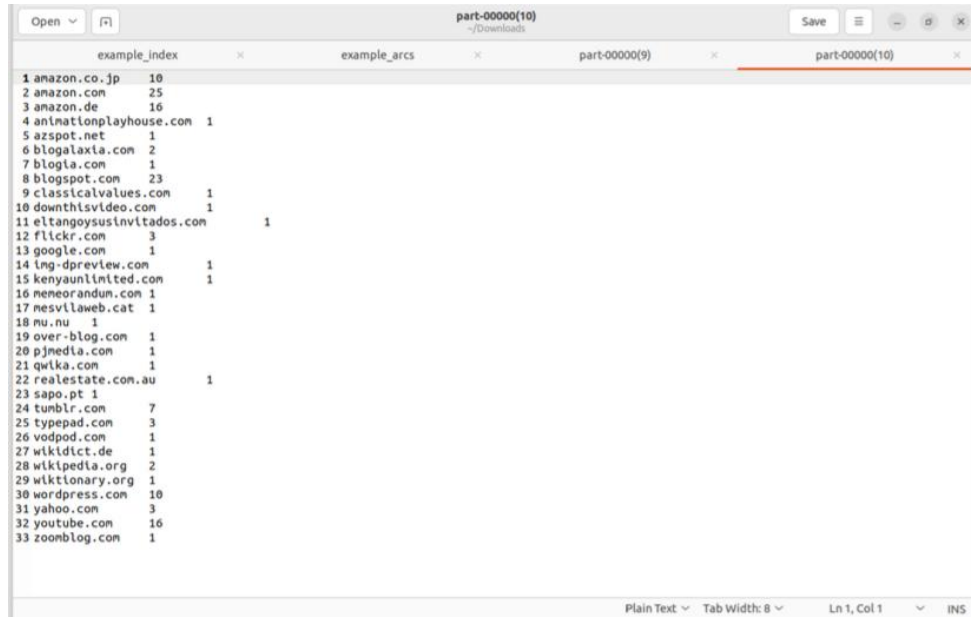
example_index example_arcs part-00000(9)

```
1 1000notes.com 1
2 100500.tv 1
3 abebooks.com 1
4 abebooks.de 1
5 amazon-presse.de 1
6 amazon.ca 3
7 amazon.cn 3
8 amazon.co.jp 2
9 amazon.co.uk 3
10 amazon.com 2
11 amazon.de 2
12 amazon.es 1
13 amazon.fr 3
14 amazon.it 1
15 angrybirds.com 1
16 apple.com 2
17 audible.com 1
18 audible.de 1
19 autonattic.com 1
20 beautybar.com 1
21 blogalaxia.com 1
22 blogg.se 1
23 blogspot.com 14
24 bookdepository.co.uk 2
25 bookdepository.com 1
26 buyvip.com 1
27 canalblog.com 2
28 carambatv.ru 1
29 creativecommons.org 3
30 cubicle17.com 1
31 cyprien.fr 1
32 dlapers.com 1
33 dlsqus.com 1
34 dpreview.com 2
35 drrum.com 1
36 endless.com 1
37 fahric.com 1
```

Plain Text Tab Width: 8 Ln 1, Col 1 INS

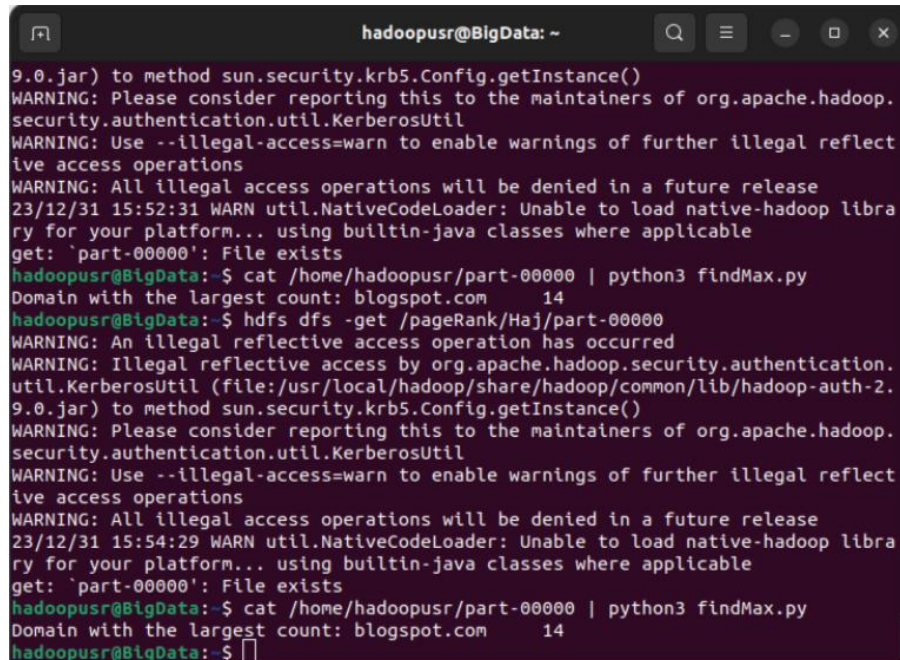
Name: Samaa Khair

SID: 201901481



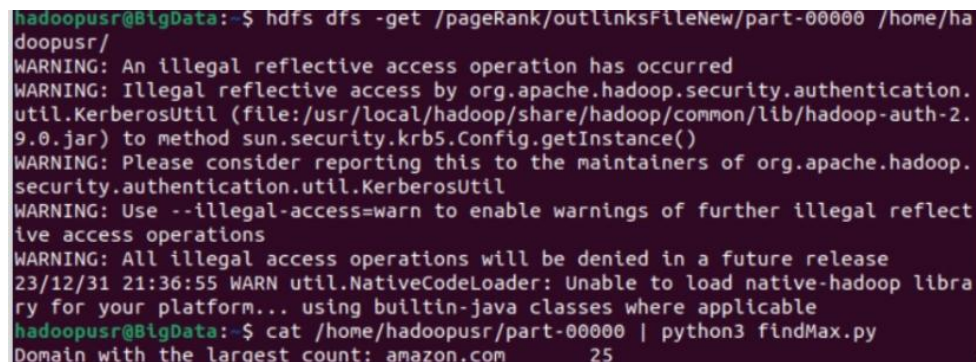
A screenshot of a text editor window titled "part-00000(10)" showing a list of domains and their counts. The list is as follows:

| Domain | Count |
|------------------------------|-------|
| 1 amazon.co.jp | 10 |
| 2 amazon.com | 25 |
| 3 amazon.de | 16 |
| 4 animationplayhouse.com | 1 |
| 5 azspot.net | 1 |
| 6 blogalaxia.com | 2 |
| 7 blogia.com | 1 |
| 8 blogspot.com | 23 |
| 9 classicalvalues.com | 1 |
| 10 downthisvideo.com | 1 |
| 11 eltangoyssusinvitados.com | 1 |
| 12 flickr.com | 3 |
| 13 google.com | 1 |
| 14 img-dpreview.com | 1 |
| 15 kenyaunlimited.com | 1 |
| 16 meneorandun.com | 1 |
| 17 mesvilaweb.cat | 1 |
| 18 nu.nu | 1 |
| 19 over-blog.com | 1 |
| 20 pjmedia.com | 1 |
| 21 qwika.com | 1 |
| 22 realestate.com.au | 1 |
| 23 sapo.pt | 1 |
| 24 tumblr.com | 7 |
| 25 typepad.com | 3 |
| 26 vodpod.com | 1 |
| 27 wikidict.de | 1 |
| 28 wikipedia.org | 2 |
| 29 wiktionary.org | 1 |
| 30 wordpress.com | 10 |
| 31 yahoo.com | 3 |
| 32 youtube.com | 16 |
| 33 zoonblog.com | 1 |



A screenshot of a terminal window titled "hadoopusr@BigData: ~" showing Hadoop commands and output. The output includes warnings about illegal reflective access operations and the results of a findMax.py script.

```
9.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/12/31 15:52:31 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
get: 'part-00000': File exists
hadoopusr@BigData:~$ cat /home/hadoopusr/part-00000 | python3 findMax.py
Domain with the largest count: blogspot.com 14
hadoopusr@BigData:~$ hdfs dfs -get /pageRank/Haj/part-00000
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
9.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/12/31 15:54:29 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
get: 'part-00000': File exists
hadoopusr@BigData:~$ cat /home/hadoopusr/part-00000 | python3 findMax.py
Domain with the largest count: blogspot.com 14
hadoopusr@BigData:~$
```



A screenshot of a terminal window titled "hadoopusr@BigData: ~" showing Hadoop commands and output. The output includes warnings about illegal reflective access operations and the results of a findMax.py script.

```
hadoopusr@BigData:~$ hdfs dfs -get /pageRank/outlinksFileNew/part-00000 /home/ha
doopusr/
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
9.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/12/31 21:36:55 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
hadoopusr@BigData:~$ cat /home/hadoopusr/part-00000 | python3 findMax.py
Domain with the largest count: amazon.com 25
```