Assignment 6

CPS610 – Database 2
Wednesday Section – Group 13
Member #1: Andy Lee (500163559)
Member #2: Sohrab Soltani (500801172)

Table of Contents

20. Counting 'https' Occurrences

Part 1, 2

```
rob@rob-VirtualBox:~/hadoop-3.2.1$ cat hadoop_home_page.html_OUTPUT_2/part-r-00000;
35      https
rob@rob-VirtualBox:~/hadoop-3.2.1$ cat hadoop_home_page.html | grep -o -w 'https' | wc -w;
35
```

Figure 1 Sohrab's Q20 implementation

```
dee@dee-VirtualBox:~/hadoop-3.2.1$ cat hadoop_home_page.html_OUTPUT_2/part-r-00000;
35      https
dee@dee-VirtualBox:~/hadoop-3.2.1$ cat hadoop_home_page.html | grep -o -w 'https' | wc -w;
35
```

Figure 2 Andy's Q20 implementation

21. Count of Each Word in Documents

```
rob@rob-VirtualBox:~/hadoop-3.2.1$ head -n 50 ~/hadoop-3.2.1/hadoop_home_page.html_OUTPUT_1/part-r-00000;
">          4
"AS        8
"License");        8
$('table').addClass('table        1
$(function()       1
&raquo;</a>        3
&rarr;</a></p>     2
'UA-7453027-1',    1
'auto');           1
'pageview');       1
(!doNotTrack)      1
(HDFS™)</strong>:          1
(e.g.      1
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){       1
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new         1
(the       8
-->        8
/          1
/>         1
0.68.3"    1
1.0.0      1
14         1
2          1
2.0        8
2.10       1
2.10.0     1
2.10.0.    1
2.10.0.</p>        2
2.10.1     1
2.x        1
2006-2021          1
21         1
2148       1
218        2
3          1
3.1        1
3.1.3      1
3.1.3.     1
3.1.3.</p>         1
3.1.4      1
3.2        1
3.2.1      1
3.2.1.     1
3.2.1.</p>         2
3.2.2      1
3.2.</p>           1
3.3        1
3.3.0      1
308        2
516        2
rob@rob-VirtualBox:~/hadoop-3.2.1$
```

Figure 3 Sohrab's Q21 implementation

```
dee@dee-VirtualBox:~/hadoop-3.2.1$ ~/hadoop-3.2.1/bin/hdfs dfs -get /user/demo/hadoop_home_page.html_
OUTPUT_1 ~/hadoop-3.2.1/hadoop_home_page.html_OUTPUT_1;
2021-04-09 00:19:30,843 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrust
ed = false, remoteHostTrusted = false
dee@dee-VirtualBox:~/hadoop-3.2.1$ head -n 50 ~/hadoop-3.2.1/hadoop_home_page.html_OUTPUT_1/part-r-00
000;
">      4
"AS     8
"License");     8
$('table').addClass('table       1
$(function()    1
&raquo;</a>     3
&rarr;</a></p>  2
'UA-7453027-1', 1
'auto');        1
'pageview');    1
(!doNotTrack)   1
(HDFS™)</strong>:       1
(e.g.   1
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){    1
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new       1
(the    8
-->     8
/       1
/>      1
0.68.3" 1
1.0.0   1
14      1
2       1
2.0     8
2.10    1
2.10.0  1
2.10.0. 1
2.10.0.</p>     2
2.10.1  1
2.x     1
2006-2021       1
21      1
2148    1
218     2
3       1
3.1     1
3.1.3   1
3.1.3.  1
3.1.3.</p>      1
3.1.4   1
3.2     1
3.2.1   1
3.2.1.  1
3.2.1.</p>      2
3.2.2   1
3.2.</p>        1
3.3     1
3.3.0   1
308     2
516     2
```

Figure 4 Andy's Q21 implementation

22. Mapper and Reducer for Wordcount

```
rob@rob-VirtualBox:~/hadoop-3.2.1$ head ~/hadoop-3.2.1/hadoop_home_page.html_OUTPUT_COUNT/part-00000;
None    0
">      4
"AS     8
"License");     8
$('table').addClass('table      1
$(function()    1
&raquo;</a>     3
&rarr;</a></p>  2
'UA-7453027-1', 1
'auto');        1
rob@rob-VirtualBox:~/hadoop-3.2.1$ ▮
```

Figure 5 Sohrab's Q22 implementation

```
dee@dee-VirtualBox:~/hadoop-3.2.1$ ls ~/hadoop-3.2.1/;
bin                              hadoop_home_page.html_OUTPUT_2  LICENSE.txt  sbin
etc                              include                        logs         share
hadoop_home_page.html            lib                            NOTICE.txt   wordcount_map.py
hadoop_home_page.html_OUTPUT_1   libexec                        README.txt   wordcount_red.py
dee@dee-VirtualBox:~/hadoop-3.2.1$ chmod u+x ~/hadoop-3.2.1/wordcount_map.py;
dee@dee-VirtualBox:~/hadoop-3.2.1$ chmod u+x ~/hadoop-3.2.1/wordcount_red.py;
```

```
dee@dee-VirtualBox:~/hadoop-3.2.1$ head /home/dee/hadoop-3.2.1/hadoop_home_page.html_OUTPUT_COUNT/hadoop_home_pa
ge.html_OUTPUT_COUNT6/part-00000
None    0
">      4
"AS     8
"License");     8
$('table').addClass('table      1
$(function()    1
&raquo;</a>     3
&rarr;</a></p>  2
'UA-7453027-1', 1
'auto');        1
```

Figure 6 Andy's Q22 implementation