



PRACTICA 2

Seminario de Sistemas 2

Comandos para realizar el conteo de palabras en el archivo Correos.txt

```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
Microsoft Windows [Version 10.0.19044.2086]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32\cmd.exe /d /s /c docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
/
Starting sshd: [ OK ]
Starting namenodes on [34d7cdc4cab2]
34d7cdc4cab2: starting namenode, logging to /usr/local/hadoop/logs/hadoop-root-namenode-34d7cdc4cab2.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanode-34d7cdc4cab2.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-root-secondarynamenode-34d7cdc4cab2.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn--resourcemanager-34d7cdc4cab2.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-root-nodemanager-34d7cdc4cab2.out
bash-4.1#
bash-4.1# mkdir resultados
bash-4.1#
bash-4.1# export HADOOP_HOME=/usr/local/hadoop
bash-4.1#
bash-4.1# export export CLASSPATH="$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.0.jar:$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.7.0.jar:$HADOOP_HOME/share/hadoop-common-2.7.0.jar:/resultados/:$HADOOP_HOME/lib/*"
bash-4.1#
bash-4.1# ls $(HADOOP_HOME)
LICENSE.txt NOTICE.txt README.txt bin etc include input lib libexec logs sbin share
bash-4.1#
bash-4.1# cd resultados
bash-4.1# javac -d . WordCount.java
/usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.0.jar(org/apache/hadoop/fs/Path.class): warning: Cannot find annotation method 'value()' in type 'LimitedPrivate': class file for org.apache.hadoop.classification.InterfaceAudience not found
1 warning
bash-4.1#
bash-4.1# cat > Manifest.txt
Main-class: WordCount
bash-4.1#
bash-4.1# jar cfm WordCount.jar Manifest.txt *.class
bash-4.1#
```

1. Comando utilizado para correr la imagen
2. Comando utilizado para crear la carpeta en la que se guardarán los archivos.
3. Comando utilizado para inicializar la variable HADOOP_HOME
4. Comando utilizado para inicializar la variable CLASSPATH
5. Comando utilizado para compilar el archivo .java
6. Comando utilizado para crear el archivo Manifest.txt con el contenido “Main-class: WordCount”
7. Comando utilizado para crear el archivo .jar

```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrapsh -bash
bash-4.1#
bash-4.1# export export CLASSPATH="$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.0.jar:$HADOOP_HOME/share/hadoop-mapreduce/hadoop-mapreduce-client-common-2.7.0.jar:$HADOOP_HOME/share/hadoop-common-2.7.0.jar:/resultados:$HADOOP_HOME/lib/"
bash-4.1#
bash-4.1# ls $(HADOOP_HOME)
LICENSE.txt NOTICE.txt README.txt bin etc include input lib libexec logs sbin share
bash-4.1#
bash-4.1# cd resultados
bash-4.1# javac -d . WordCount.java
/usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.0.jar(org/apache/hadoop/fs/Path.class): warning: Cannot find annotation method 'value()' in type 'LimitedPrivate': class file for org.apache.hadoop.classification.InterfaceAudience not found
1 warning
bash-4.1#
bash-4.1# cat > Manifest.txt
Main-class: WordCount
bash-4.1#
bash-4.1# jar cfm WordCount.jar Manifest.txt *.class
bash-4.1#
bash-4.1#
bash-4.1# mkdir ~/input
bash-4.1# mkdir ~/output
bash-4.1#
bash-4.1# cp Correos.txt ~/input
bash-4.1#
bash-4.1# $(HADOOP_HOME)/bin/hdfs dfs -copyFromLocal ~/input /
bash-4.1#
bash-4.1# $(HADOOP_HOME)/bin/hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 1 root supergroup 31354 2022-10-06 01:45 /input/Correos.txt
bash-4.1#
```

8. Comandos para crear las carpetas input y output en /root
9. Comando utilizado para copiar el archivo de entrada en la carpeta ~/input
10. Comando utilizado para copiar el archivo de entrada en el sistema de archivos de hadoop
11. Comando utilizado para verificar los archivos que fueron copiados en el sistema de archivos de hadoop

```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrapsh -bash
-rw-r--r-- 1 root supergroup 31354 2022-10-06 01:45 /input/Correos.txt
bash-4.1#
bash-4.1# $(HADOOP_HOME)/bin/hadoop jar WordCount.jar /input /output
22/10/06 01:46:47 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/10/06 01:46:47 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/10/06 01:46:48 INFO input.FileInputFormat: Total input paths to process : 1
22/10/06 01:46:48 INFO mapreduce.JobSubmitter: number of splits:1
22/10/06 01:46:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665034689181_0001
22/10/06 01:46:48 INFO impl.YarnClientImpl: Submitted application application_1665034689181_0001
22/10/06 01:46:49 INFO mapreduce.Job: The url to track the job: http://34d7cdc4cab2:8088/proxy/application_1665034689181_0001/
22/10/06 01:46:49 INFO mapreduce.Job: Running job: job_1665034689181_0001
22/10/06 01:46:58 INFO mapreduce.Job: Job job_1665034689181_0001 running in uber mode : false
22/10/06 01:46:58 INFO mapreduce.Job: map 0% reduce 0%
22/10/06 01:47:04 INFO mapreduce.Job: map 100% reduce 0%
22/10/06 01:47:12 INFO mapreduce.Job: map 100% reduce 100%
22/10/06 01:47:12 INFO mapreduce.Job: Job job_1665034689181_0001 completed successfully
22/10/06 01:47:12 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=26259
  FILE: Number of bytes written=282275
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=31461
  HDFS: Number of bytes written=18624
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4135
  Total time spent by all reduces in occupied slots (ms)=5057
  Total time spent by all map tasks (ms)=4135
  Total time spent by all reduce tasks (ms)=5057
  Total vcore-seconds taken by all map tasks=4135
  Total vcore-seconds taken by all reduce tasks=5057
  Total megabyte-seconds taken by all map tasks=4234240
  Total megabyte-seconds taken by all reduce tasks=5178368
Map-Reduce Framework
  Map input records=50
  Map output records=4494
  Map output bytes=49233
  Map output materialized bytes=26259
  Input split bytes=107
  Combine input records=4494
  Combine output records=1923
  Reduce input groups=1923
```

12. Comando utilizado para realizar el conteo de palabras

```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
Combine output records=1923
Reduce input groups=1923
Reduce shuffle bytes=26259
Reduce input records=1923
Reduce output records=1923
Spilled Records=3846
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=71
CPU time spent (ms)=3290
Physical memory (bytes) snapshot=438538240
Virtual memory (bytes) snapshot=1487491872
Total committed heap usage (bytes)=397410304

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=31354
File Output Format Counters
Bytes Written=18624

bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 1 root supergroup 0 2022-10-06 01:47 /output/_SUCCESS
-rw-r--r-- 1 root supergroup 18624 2022-10-06 01:47 /output/part-r-00000

bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -cat /output/Resultado_Correos.txt
cat: /output/Resultado_Correos.txt: No such file or directory

bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -mv /output/part-r-00000 /output/Resultado_Correos.txt
bash-4.1#
```

- 13. Comando utilizado para listar los archivos de salida
- 14. Comando utilizado para ver el contenido del archivo de salida
- 15. Comando utilizado para cambiar el nombre del archivo de salida

```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
cat: /output/Resultado_Correos.txt: No such file or directory
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -mv /output/part-r-00000 /output/Resultado_Correos.txt
bash-4.1#
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -cat /output/Resultado_Correos.txt
1 5
2 1
3 1
4 1
5 1
6 1
7 1
8 1
9 1
10 3
11 1
12 2
13 1
14 1
15 1
16 1
17 1
18 1
19 1
20 1
21 1
22 1
23 1
24 1
25 1
26 1
27 1
28 1
29 1
30 1
31 1
32 1
33 1
34 1
35 1
36 1
37 1
38 1
39 1
40 1
41 1
42 1
43 1
44 1
45 1
46 1
47 1
48 1
49 1
50 1
51 1
52 1
53 1
54 1
55 1
56 1
57 1
58 1
59 1
60 1
61 1
62 1
63 1
64 1
65 1
66 1
67 1
68 1
69 1
70 1
71 1
72 1
73 1
74 1
75 1
76 1
77 1
78 1
79 1
80 1
81 1
82 1
83 1
84 1
85 1
86 1
87 1
88 1
89 1
90 1
91 1
92 1
93 1
94 1
95 1
96 1
97 1
98 1
99 1
100 1
101 1
102 1
103 1
104 1
105 1
106 1
107 1
108 1
109 1
110 1
111 1
112 1
113 1
114 1
115 1
116 1
117 1
118 1
119 1
120 1
121 1
122 1
123 1
124 1
125 1
126 1
127 1
128 1
129 1
130 1
131 1
132 1
133 1
134 1
135 1
136 1
137 1
138 1
139 1
140 1
141 1
142 1
143 1
144 1
145 1
146 1
147 1
148 1
149 1
150 1
151 1
152 1
153 1
154 1
155 1
156 1
157 1
158 1
159 1
160 1
161 1
162 1
163 1
164 1
165 1
166 1
167 1
168 1
169 1
170 1
171 1
172 1
173 1
174 1
175 1
176 1
177 1
178 1
179 1
180 1
181 1
182 1
183 1
184 1
185 1
186 1
187 1
188 1
189 1
190 1
191 1
192 1
193 1
194 1
195 1
196 1
197 1
198 1
199 1
200 1
201 1
202 1
203 1
204 1
205 1
206 1
207 1
208 1
209 1
210 1
211 1
212 1
213 1
214 1
215 1
216 1
217 1
218 1
219 1
220 1
221 1
222 1
223 1
224 1
225 1
226 1
227 1
228 1
229 1
230 1
231 1
232 1
233 1
234 1
235 1
236 1
237 1
238 1
239 1
240 1
241 1
242 1
243 1
244 1
245 1
246 1
247 1
248 1
249 1
250 1
251 1
252 1
253 1
254 1
255 1
256 1
257 1
258 1
259 1
260 1
261 1
262 1
263 1
264 1
265 1
266 1
267 1
268 1
269 1
270 1
271 1
272 1
273 1
274 1
275 1
276 1
277 1
278 1
279 1
280 1
281 1
282 1
283 1
284 1
285 1
286 1
287 1
288 1
289 1
290 1
291 1
292 1
293 1
294 1
295 1
296 1
297 1
298 1
299 1
300 1
301 1
302 1
303 1
304 1
305 1
306 1
307 1
308 1
309 1
310 1
311 1
312 1
313 1
314 1
315 1
316 1
317 1
318 1
319 1
320 1
321 1
322 1
323 1
324 1
325 1
326 1
327 1
328 1
329 1
330 1
331 1
332 1
333 1
334 1
335 1
336 1
337 1
338 1
339 1
340 1
341 1
342 1
343 1
344 1
345 1
346 1
347 1
348 1
349 1
350 1
351 1
352 1
353 1
354 1
355 1
356 1
357 1
358 1
359 1
360 1
361 1
362 1
363 1
364 1
365 1
366 1
367 1
368 1
369 1
370 1
371 1
372 1
373 1
374 1
375 1
376 1
377 1
378 1
379 1
380 1
381 1
382 1
383 1
384 1
385 1
386 1
387 1
388 1
389 1
390 1
391 1
392 1
393 1
394 1
395 1
396 1
397 1
398 1
399 1
400 1
401 1
402 1
403 1
404 1
405 1
406 1
407 1
408 1
409 1
410 1
411 1
412 1
413 1
414 1
415 1
416 1
417 1
418 1
419 1
420 1
421 1
422 1
423 1
424 1
425 1
426 1
427 1
428 1
429 1
430 1
431 1
432 1
433 1
434 1
435 1
436 1
437 1
438 1
439 1
440 1
441 1
442 1
443 1
444 1
445 1
446 1
447 1
448 1
449 1
450 1
451 1
452 1
453 1
454 1
455 1
456 1
457 1
458 1
459 1
460 1
461 1
462 1
463 1
464 1
465 1
466 1
467 1
468 1
469 1
470 1
471 1
472 1
473 1
474 1
475 1
476 1
477 1
478 1
479 1
480 1
481 1
482 1
483 1
484 1
485 1
486 1
487 1
488 1
489 1
490 1
491 1
492 1
493 1
494 1
495 1
496 1
497 1
498 1
499 1
500 1
501 1
502 1
503 1
504 1
505 1
506 1
507 1
508 1
509 1
510 1
511 1
512 1
513 1
514 1
515 1
516 1
517 1
518 1
519 1
520 1
521 1
522 1
523 1
524 1
525 1
526 1
527 1
528 1
529 1
530 1
531 1
532 1
533 1
534 1
535 1
536 1
537 1
538 1
539 1
540 1
541 1
542 1
543 1
544 1
545 1
546 1
547 1
548 1
549 1
550 1
551 1
552 1
553 1
554 1
555 1
556 1
557 1
558 1
559 1
560 1
561 1
562 1
563 1
564 1
565 1
566 1
567 1
568 1
569 1
570 1
571 1
572 1
573 1
574 1
575 1
576 1
577 1
578 1
579 1
580 1
581 1
582 1
583 1
584 1
585 1
586 1
587 1
588 1
589 1
590 1
591 1
592 1
593 1
594 1
595 1
596 1
597 1
598 1
599 1
600 1
601 1
602 1
603 1
604 1
605 1
606 1
607 1
608 1
609 1
610 1
611 1
612 1
613 1
614 1
615 1
616 1
617 1
618 1
619 1
620 1
621 1
622 1
623 1
624 1
625 1
626 1
627 1
628 1
629 1
630 1
631 1
632 1
633 1
634 1
635 1
636 1
637 1
638 1
639 1
640 1
641 1
642 1
643 1
644 1
645 1
646 1
647 1
648 1
649 1
650 1
651 1
652 1
653 1
654 1
655 1
656 1
657 1
658 1
659 1
660 1
661 1
662 1
663 1
664 1
665 1
666 1
667 1
668 1
669 1
670 1
671 1
672 1
673 1
674 1
675 1
676 1
677 1
678 1
679 1
680 1
681 1
682 1
683 1
684 1
685 1
686 1
687 1
688 1
689 1
690 1
691 1
692 1
693 1
694 1
695 1
696 1
697 1
698 1
699 1
700 1
701 1
702 1
703 1
704 1
705 1
706 1
707 1
708 1
709 1
710 1
711 1
712 1
713 1
714 1
715 1
716 1
717 1
718 1
719 1
720 1
721 1
722 1
723 1
724 1
725 1
726 1
727 1
728 1
729 1
730 1
731 1
732 1
733 1
734 1
735 1
736 1
737 1
738 1
739 1
740 1
741 1
742 1
743 1
744 1
745 1
746 1
747 1
748 1
749 1
750 1
751 1
752 1
753 1
754 1
755 1
756 1
757 1
758 1
759 1
760 1
761 1
762 1
763 1
764 1
765 1
766 1
767 1
768 1
769 1
770 1
771 1
772 1
773 1
774 1
775 1
776 1
777 1
778 1
779 1
780 1
781 1
782 1
783 1
784 1
785 1
786 1
787 1
788 1
789 1
790 1
791 1
792 1
793 1
794 1
795 1
796 1
797 1
798 1
799 1
800 1
801 1
802 1
803 1
804 1
805 1
806 1
807 1
808 1
809 1
810 1
811 1
812 1
813 1
814 1
815 1
816 1
817 1
818 1
819 1
820 1
821 1
822 1
823 1
824 1
825 1
826 1
827 1
828 1
829 1
830 1
831 1
832 1
833 1
834 1
835 1
836 1
837 1
838 1
839 1
840 1
841 1
842 1
843 1
844 1
845 1
846 1
847 1
848 1
849 1
850 1
851 1
852 1
853 1
854 1
855 1
856 1
857 1
858 1
859 1
860 1
861 1
862 1
863 1
864 1
865 1
866 1
867 1
868 1
869 1
870 1
871 1
872 1
873 1
874 1
875 1
876 1
877 1
878 1
879 1
880 1
881 1
882 1
883 1
884 1
885 1
886 1
887 1
888 1
889 1
890 1
891 1
892 1
893 1
894 1
895 1
896 1
897 1
898 1
899 1
900 1
901 1
902 1
903 1
904 1
905 1
906 1
907 1
908 1
909 1
910 1
911 1
912 1
913 1
914 1
915 1
916 1
917 1
918 1
919 1
920 1
921 1
922 1
923 1
924 1
925 1
926 1
927 1
928 1
929 1
930 1
931 1
932 1
933 1
934 1
935 1
936 1
937 1
938 1
939 1
940 1
941 1
942 1
943 1
944 1
945 1
946 1
947 1
948 1
949 1
950 1
951 1
952 1
953 1
954 1
955 1
956 1
957 1
958 1
959 1
960 1
961 1
962 1
963 1
964 1
965 1
966 1
967 1
968 1
969 1
970 1
971 1
972 1
973 1
974 1
975 1
976 1
977 1
978 1
979 1
980 1
981 1
982 1
983 1
984 1
985 1
986 1
987 1
988 1
989 1
990 1
991 1
992 1
993 1
994 1
995 1
996 1
997 1
998 1
999 1
1000 1
1001 1
1002 1
1003 1
1004 1
1005 1
1006 1
1007 1
1008 1
1009 1
1010 1
1011 1
1012 1
1013 1
1014 1
1015 1
1016 1
1017 1
1018 1
1019 1
1020 1
1021 1
1022 1
1023 1
1024 1
1025 1
1026 1
1027 1
1028 1
1029 1
1030 1
1031 1
1032 1
1033 1
1034 1
1035 1
1036 1
1037 1
1038 1
1039 1
1040 1
1041 1
1042 1
1043 1
1044 1
1045 1
1046 1
1047 1
1048 1
1049 1
1050 1
1051 1
1052 1
1053 1
1054 1
1055 1
1056 1
1057 1
1058 1
1059 1
1060 1
1061 1
1062 1
1063 1
1064 1
1065 1
1066 1
1067 1
1068 1
1069 1
1070 1
1071 1
1072 1
1073 1
1074 1
1075 1
1076 1
1077 1
1078 1
1079 1
1080 1
1081 1
1082 1
1083 1
1084 1
1085 1
1086 1
1087 1
1088 1
1089 1
1090 1
1091 1
1092 1
1093 1
1094 1
1095 1
1096 1
1097 1
1098 1
1099 1
1100 1
1101 1
1102 1
1103 1
1104 1
1105 1
1106 1
1107 1
1108 1
1109 1
1110 1
1111 1
1112 1
1113 1
1114 1
1115 1
1116 1
1117 1
1118 1
1119 1
1120 1
1121 1
1122 1
1123 1
1124 1
1125 1
1126 1
1127 1
1128 1
1129 1
1130 1
1131 1
1132 1
1133 1
1134 1
1135 1
1136 1
1137 1
1138 1
1139 1
1140 1
1141 1
1142 1
1143 1
1144 1
1145 1
1146 1
1147 1
1148 1
1149 1
1150 1
1151 1
1152 1
1153 1
1154 1
1155 1
1156 1
1157 1
1158 1
1159 1
1160 1
1161 1
1162 1
1163 1
1164 1
1165 1
1166 1
1167 1
1168 1
1169 1
1170 1
1171 1
1172 1
1173 1
1174 1
1175 1
1176 1
1177 1
1178 1
1179 1
1180 1
1181 1
1182 1
1183 1
1184 1
1185 1
1186 1
1187 1
1188 1
1189 1
1190 1
1191 1
1192 1
1193 1
1194 1
1195 1
1196 1
1197 1
1198 1
1199 1
1200 1
1201 1
1202 1
1203 1
1204 1
1205 1
1206 1
1207 1
1208 1
1209 1
1210 1
1211 1
1212 1
1213 1
1214 1
1215 1
1216 1
1217 1
1218 1
1219 1
1220 1
1221 1
1222 1
1223 1
1224 1
1225 1
1226 1
1227 1
1228 1
1229 1
1230 1
1231 1
1232 1
1233 1
1234 1
1235 1
1236 1
1237 1
1238 1
1239 1
1240 1
1241 1
1242 1
1243 1
1244 1
1245 1
1246 1
1247 1
1248 1
1249 1
1250 1
1251 1
1252 1
1253 1
1254 1
1255 1
1256 1
1257 1
1258 1
1259 1
1260 1
1261 1
1262 1
1263 1
1264 1
1265 1
1266 1
1267 1
1268 1
1269 1
1270 1
1271 1
1272 1
1273 1
1274 1
1275 1
1276 1
1277 1
1278 1
1279 1
1280 1
1281 1
1282 1
1283 1
1284 1
1285 1
1286 1
1287 1
1288 1
1289 1
1290 1
1291 1
1292 1
1293 1
1294 1
1295 1
1296 1
1297 1
1298 1
1299 1
1300 1
1301 1
1302 1
1303 1
1304 1
1305 1
1306 1
1307 1
1308 1
1309 1
1310 1
1311 1
1312 1
1313 1
1314 1
1315 1
1316 1
1317 1
1318 1
1319 1
1320 1
1321 1
1322 1
1323 1
1324 1
1325 1
1326 1
1327 1
1328 1
1329 1
1330 1
1331 1
1332 1
1333 1
1334 1
1335 1
1336 1
1337 1
1338 1
1339 1
1340 1
1341 1
1342 1
1343 1
1344 1
1345 1
1346 1
1347 1
1348 1
1349 1
1350 1
1351 1
1352 1
1353 1
1354 1
1355 1
1356 1
1357 1
1358 1
1359 1
1360 1
1361 1
1362 1
1363 1
1364 1
1365 1
1366 1
1367 1
1368 1
1369 1
1370 1
1371 1
1372 1
1373 1
1374 1
1375 1
1376 1
1377 1
1378 1
1379 1
1380 1
1381 1
1382 1
1383 1
1384 1
1385 1
1386 1
1387 1
1388 1
1389 1
1390 1
1391 1
1392 1
1393 1
1394 1
1395 1
1396 1
1397 1
1398 1
1399 1
1400 1
1401 1
1402 1
1403 1
1404 1
1405 1
1406 1
1407 1
1408 1
1409 1
1410 1
1411 1
1412 1
1413 1
1414 1
1415 1
1416 1
1417 1
1418 1
1419 1
1420 1
1421 1
1422 1
1423 1
1424 1
1425 1
1426 1
1427 1
1428 1
1429 1
1430 1
1431 1
1432 1
1433 1
1434 1
1435 1
1436 1
1437 1
1438 1
1439 1
1440 1
1441 1
1442 1
1443 1
1444 1
1445 1
1446 1
1447 1
1448 1
1449 1
1450 1
1451 1
1452 1
1453 1
1454 1
1455 1
1456 1
1457 1
1458 1
1459 1
1460 1
1461 1
1462 1
1463 1
1464 1
1465 1
1466 1
1467 1
1468 1
1469 1
1470 1
1471 1
1472 1
1473 1
1474 1
1475 1
1476 1
1477 1
1478 1
1479 1
1480 1
1481 1
1482 1
1483 1
1484 1
1485 1
1486 1
1487 1
1488 1
1489 1
1490 1
1491 1
1492 1
1493 1
1494 1
1495 1
1496 1
1497 1
1498 1
1499 1
1500 1
1501 1
1502 1
1503 1
1504 1
1505 1
1506 1
1507 1
1508 1
1509 1
1510 1
1511 1
1512 1
1513 1
1514 1
1515 1
1516 1
1517 1
1518 1
1519 1
1520 1
1521 1
1522 1
1523 1
1524 1
1525 1
1526 1
1527 1
1528 1
1529 1
1530 1
1531 1
1532 1
1533 1
1534 1
1535 1
1536 1
1537 1
1538 1
1539 1
1540 1
1541 1
1542 1
1543 1
1544 1
1545 1
1546 1
1547 1
1548 1
1549 1
1550 1
1551 1
1552 1
1553 1
1554 1
1555 1
1556 1
1557 1
1558 1
1559 1
1560 1
1561 1
1562 1
1563 1
1564 1
1565 1
1566 1
1567 1
1568 1
1569 1
1570 1
1571 1
1572 1
1573 1
1574 1
1575 1
1576 1
1577 1
1578 1
1579 1
1580 1
1581 1
1582 1
1583 1
1584 1
1585 1
1586 1
1587 1
1588 1
1589 1
1590 1
1591 1
1592 1
1593 1
1594 1
1595 1
1596 1
1597 1
1598 1
1599 1
1600 1
1601 1
1602 1
1603 1
1604 1
1605 1
1606 1
1607 1
1608 1
1609 1
1610 1
1611 1
1612 1
1613 1
1614 1
1615 1
1616 1
1617 1
1618 1
1619 1
1620 1
1621 1
1622 1
1623 1
1624 1
1625 1
1626 1
1627 1
1628 1
1629 1
1630 1
1631 1
1632 1
1633 1
1634 1
1635 1
1636 1
1637 1
1638 1
1639 1
1640 1
1641 1
1642 1
1643 1
1644 1
1645 1
1646 1
1647 1
1648 1
1649 1
1650 1
1651 1
1652 1
1653 1
1654 1
1655 1
1656 1
1657 1
1658 1
1659 1
1660 1
1661 1
1662 1
1663 1
1664 1
1665 1
1666 1
1667 1
1668 1
1669 1
1670 1
1671 1
1672 1
1673 1
1674 1
1675 1
1676 1
1677 1
1678 1
1679 1
1680 1
1681 1
1682 1
1683 1
1684 1
1685 1
1686 1
1687 1
1688 1
1689 1
1690 1
1691 1
1692 1
1693 1
1694 1
1695 1
1696 1
1697 1
1698 1
1699 1
1700 1
1701 1
1702 1
1703 1
1704 1
1705 1
1706 1
1707 1
1708 1
1709 1
1710 1
1711 1
1712 1
1713 1
1714 1
1715 1
1716 1
1717 1
1718 1
1719 1
1720 1
1721 1
1722 1
1723 1
1724 1
1725 1
1726 1
1727 1
1728 1
1729 1
1730 1
1731 1
1732 1
1733 1
1734 1
1735 1
1736 1
1737 1
1738 1
1739 1
1740 1
1741 1
1742 1
1743 1
1744 1
1745 1
1746 1
1747 1
1748 1
1749 1
1750 1
1751 1
1752 1
1753 1
1754 1
1755 1
1756 1
1757 1
1758 1
1759 1
1760 1
1761 1
1762 1
1763 1
1764 1
1765 1
1766 1
1767 1
1768 1
1769 1
1770 1
1771 1
1772 1
1773 1
1774 1
1775 1
1776 1
1777 1
1778 1
1779 1
1780 1
1781 1
1782 1
1783 1
1784 1
1785 1
1786 1
1787 1
1788 1
1789 1
1790 1
1791 1
1792 1
1793 1
1794 1
1795 1
1796 1
1797 1
1798 1
1799 1
1800 1
1801 1
1802 1
1803 1
1804 1
1805 1
1806 1
1807 1
1808 1
1809 1
1810 1
1811 1
1812 1
1813 1
1814 
```

```
Administrator: Command Prompt - docker run --rm -it -v practica2/source-p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrapsh -bash
years 4
yep 1
yes 2
yikes 1
yoga 1
young 3
zone 1
bash-4.1#
bash-4.1# $(HADOOP_HOME)/bin/hdfs dfs -copyToLocal /output/Resultado_Correo.txt ~/output
22/10/06 01:50:28 WARN hdfs.DFSClient: DFSInputStream has been closed already
bash-4.1# ls
Correo.txt Manifest.txt WordCount$IntSumReducer.class WordCount$TokenizerMapper.class WordCount.class WordCount.jar WordCount.java
bash-4.1#
bash-4.1# $(HADOOP_HOME)/bin/hdfs dfs -copyToLocal /output/Resultado_Correo.txt ~/output
copyToLocal: '/root/output/Resultado_Correo.txt': File exists
bash-4.1#
bash-4.1# cp ~/output/Resultado_Correo.txt /resultados
bash-4.1#
```

17. Comando utilizado para copiar el archivo de salida a la carpeta ~/output

18. Comando utilizado para mover el archivo de salida a la carpeta resultados del contenedor

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19044.2006]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>docker cp "C:\Users\mario\Downloads\Correo.txt" 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados
C:\Windows\system32>docker cp "C:\Users\mario\Downloads\WordCount.java" 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados
C:\Windows\system32>docker cp "C:\Users\mario\Downloads\Resultado_Puntuacion.txt" 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados
C:\Windows\system32>docker cp 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados/Resultado_Puntuacion.txt "C:\Users\mario\OneDrive\Desktop\resultados_semi"
C:\Windows\system32>
```

19. Comando utilizado para copiar el archivo de salida del contenedor a la PC

Browse HDFS – Correos.txt

0798A_202205_LB: Práctico

Browsing HDFS

localhost:50070/explorer.html#/

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Browse Directory

/

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	root	supergroup	0 B	10/5/2022, 11:45:29 PM	0	0 B	input
drwxr-xr-x	root	supergroup	0 B	10/5/2022, 11:49:02 PM	0	0 B	output
drwx-----	root	supergroup	0 B	10/5/2022, 11:46:47 PM	0	0 B	tmp
drwxr-xr-x	root	supergroup	0 B	7/22/2015, 9:17:26 AM	0	0 B	user

Hadoop, 2014.

11:59 PM10/5/2022

0798A_202205_LB: Práctico

Browsing HDFS

localhost:50070/explorer.html#/output

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Browse Directory

/output

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	root	supergroup	18.19 KB	10/5/2022, 11:47:10 PM	1	128 MB	Resultado_Correos.txt
-rw-r--r--	root	supergroup	0 B	10/5/2022, 11:47:10 PM	1	128 MB	_SUCCESS

Hadoop, 2014.

12:00 AM10/6/2022

Comandos para realizar el conteo de palabras en el archivo Puntuaciones.txt

```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
Microsoft Windows [Version 10.0.19044.2006]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
/
Starting sshd: [ OK ]
Starting namenodes on [92820f54b487]
92820f54b487: starting namenode, logging to /usr/local/hadoop/logs/hadoop-root-namenode-92820f54b487.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanode-92820f54b487.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-root-secondarynamenode-92820f54b487.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-resourcemanager-92820f54b487.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-root-nodemanager-92820f54b487.out
bash-4.1#
bash-4.1# mkdir resultados
bash-4.1#
bash-4.1# export HADOOP_HOME=/usr/local/hadoop
bash-4.1#
bash-4.1# export export CLASSPATH="$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.0.jar:$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.7.0.jar:$HADOOP_HOME/share/hadoop-common-2.7.0.jar:/resultados/.$HADOOP_HOME/lib/*"
bash-4.1#
bash-4.1# cd resultados
bash-4.1# rm Correos.txt
bash-4.1# javac -d WordCount.java
/usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.0.jar(org/apache/hadoop/fs/Path.class): warning: Cannot find annotation method 'value()' in type 'LimitedPrivate': class file for org.apache.hadoop.classification.InterfaceAudience not found
1 warning
bash-4.1#
bash-4.1#
bash-4.1# cat > Manifest.txt
Main-class: WordCount
bash-4.1#
```

1. Comando utilizado para correr la imagen de sequenceiq/hadoop-docker
2. Comando utilizado para crear la carpeta en la que se guardarán los archivos.
3. Comando utilizado para inicializar la variable HADOOP_HOME
4. Comando utilizado para inicializar la variable CLASSPATH
5. Comando utilizado para compilar el archivo .java
6. Comando utilizado para crear el archivo Manifest.txt con el contenido “Main-class: WordCount”

Posteriormente se utilizó el comando “jar cfm WordCount.jar Manifest.txt *.class” para crear el archivo .jar pero este no se mostró en pantalla.

```
Administrator Command Prompt - docker run --rm -it -v practica2/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
bash-4.1# mkdir ~/input
bash-4.1# mkdir ~/output
bash-4.1#
bash-4.1# cp Puntuacion.txt ~/input
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -copyFromLocal ~/input /
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 1 root supergroup 18429 2022-10-06 02:11 /input/Puntuacion.txt
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hadoop jar WordCount.jar /input /output
22/10/06 02:13:40 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/10/06 02:13:41 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/10/06 02:13:41 INFO input.FileInputFormat: Total input paths to process : 1
22/10/06 02:13:42 INFO mapreduce.JobSubmitter: number of splits=1
22/10/06 02:13:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665036137066_0001
22/10/06 02:13:42 INFO impl.YarnClientImpl: Submitted application application_1665036137066_0001
22/10/06 02:13:42 INFO mapreduce.Job: The url to track the job: http://92820f54b487:8888/proxy/application_1665036137066_0001/
22/10/06 02:13:42 INFO mapreduce.Job: Running job: job_1665036137066_0001
22/10/06 02:13:51 INFO mapreduce.Job: Job job_1665036137066_0001 running in uber mode : false
22/10/06 02:13:51 INFO mapreduce.Job: map 0% reduce 0%
22/10/06 02:13:57 INFO mapreduce.Job: map 100% reduce 0%
22/10/06 02:14:05 INFO mapreduce.Job: map 100% reduce 100%
22/10/06 02:14:06 INFO mapreduce.Job: Job job_1665036137066_0001 completed successfully
22/10/06 02:14:06 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=46
  FILE: Number of bytes written=229849
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=18539
  HDFS: Number of bytes written=35
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4170
  Total time spent by all reduces in occupied slots (ms)=4365
  Total time spent by all map tasks (ms)=4170
  Total time spent by all reduce tasks (ms)=4365
  Total vcore-seconds taken by all map tasks=4170
```

7. Comandos para crear las carpetas input y output en /root
8. Comando utilizado para copiar el archivo de entrada en la carpeta ~/input
9. Comando utilizado para copiar el archivo de entrada en el sistema de archivos de hadoop
10. Comando utilizado para verificar los archivos que fueron copiados en el sistema de archivos de hadoop
11. Comando utilizado para realizar el conteo de palabras


```
Administrator: Command Prompt - docker run --rm -it -v practica2:/source -p 50070-50080:50070-50080 sequenceiq/hadoop-docker /etc/bootstrap.sh -bash
Bytes Written=35
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 1 root supergroup 0 2022-10-06 02:14 /output/_SUCCESS
-rw-r--r-- 1 root supergroup 35 2022-10-06 02:14 /output/part-r-00000
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -mv /output/part-r-00000 /output/Resultado_Puntuacion.txt
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -cat /output/Resultado_Puntuacion.txt
1 1381
2 1070
3 1245
4 2550
5 2969
bash-4.1#
bash-4.1# ${HADOOP_HOME}/bin/hdfs dfs -copyToLocal /output/Resultado_Puntuacion.txt ~/output
22/10/06 02:17:14 WARN hdfs.DFSClient: DFSInputStream has been closed already
bash-4.1#
bash-4.1# cp ~/output/Resultado_Puntuacion.txt /resultados
bash-4.1#
```

- 12. Comando utilizado para listar los archivos de salida
- 13. Comando utilizado para cambiar el nombre del archivo de salida
- 14. Comando utilizado para ver el contenido del archivo que fue renombrado
- 15. Comando utilizado para copiar el archivo de salida a la carpeta ~/output
- 16. Comando utilizado para mover el archivo de salida a la carpeta resultados del contenedor

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19044.2006]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>docker cp "C:\Users\mario\Downloads\Correos.txt" 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados
C:\Windows\system32>docker cp "C:\Users\mario\Downloads\WordCount.java" 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados
C:\Windows\system32>docker cp "C:\Users\mario\Downloads\Puntuacion.txt" 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:\resultados
C:\Windows\system32>docker cp 92820f54b4872e8dea2a94d9c455063f6546cf95c43884b03901a9686e8e56d2:/resultados/Resultado_Puntuacion.txt "C:\Users\mario\OneDrive\Desktop\resultados_semi"
C:\Windows\system32>
```

- 17. Comando utilizado para copiar el archivo de salida del contenedor a la PC

Browse HDFS – Puntuacion.txt

0798A_202205_LB: Práctico

Browsing HDFS

localhost:50070/explorer.html#/

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Browse Directory

/

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	root	supergroup	0 B	10/6/2022, 12:11:08 AM	0	0 B	input
drwxr-xr-x	root	supergroup	0 B	10/6/2022, 12:16:24 AM	0	0 B	output
drwx-----	root	supergroup	0 B	10/6/2022, 12:13:41 AM	0	0 B	tmp
drwxr-xr-x	root	supergroup	0 B	7/22/2015, 9:17:26 AM	0	0 B	user

Hadoop, 2014.

12:19 AM

10/6/2022

0798A_202205_LB: Práctico

Browsing HDFS

localhost:50070/explorer.html#/output

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Browse Directory

/output

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	root	supergroup	35 B	10/6/2022, 12:14:03 AM	1	128 MB	Resultado_Puntuacion.txt
-rw-r--r--	root	supergroup	0 B	10/6/2022, 12:14:04 AM	1	128 MB	_SUCCESS

Hadoop, 2014.

12:20 AM

10/6/2022

Conclusiones de los resultados

- El documento de los resultados del archivo correos.txt muestra una gran cantidad de palabras y las veces que estas se encuentran repetidas, aunque en algunas ocasiones se puede notar que algunas palabras se repiten, pero con alguna variante como terminar en una coma, esto debido a que la herramienta realiza el conteo utilizando espacios como separador únicamente.
- En documento de resultados del archivo puntuacion.txt muestra una salida más pequeña siendo esto debido a que en el archivo solo se encuentran números del 1 al 5, que son las puntuaciones recolectadas de usuarios, también siendo un resultado más exacto que el anterior.

Uso de Hadoop en Big Data

- Hadoop resulta ser una herramienta de gran ayuda en el tema de Big Data, esto debido a las avanzadas capacidades que esta ofrece en el análisis de datos. Puede almacenar grandes cantidades de información para ser analizada y posteriormente procesada, es útil para mejorar la eficiencia de procesos y operaciones de las organizaciones, mediante esta se pueden descubrir patrones de comportamiento, y debido a su carácter flexible permite modificar el sistema de datos de las empresas u organismos.