WordCount - MapReduce

Computer Specifications:

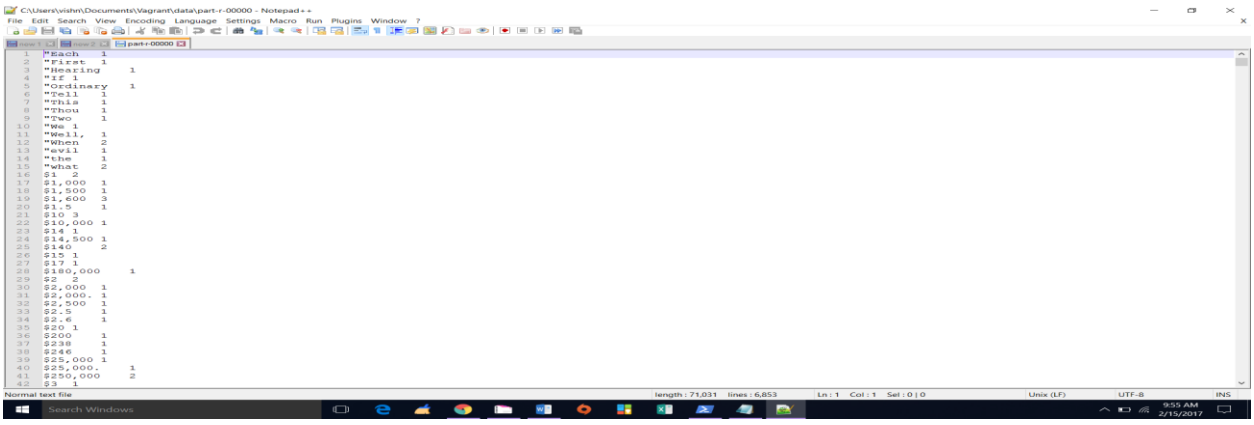
Config:

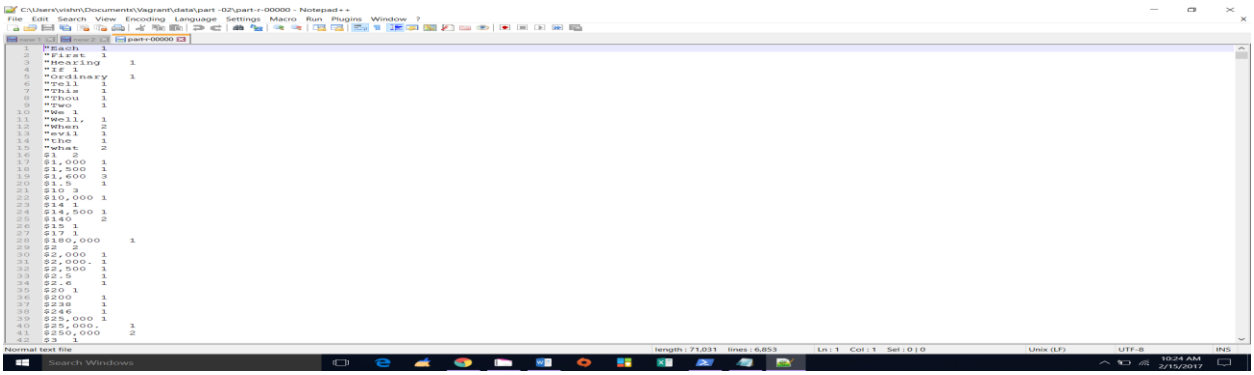
- VRAM – 6144 MB
- PROCESSOR – 2 Core (Virtual machine)

VMachine Used:

- ubuntu/trusty64

Output Screenshot:





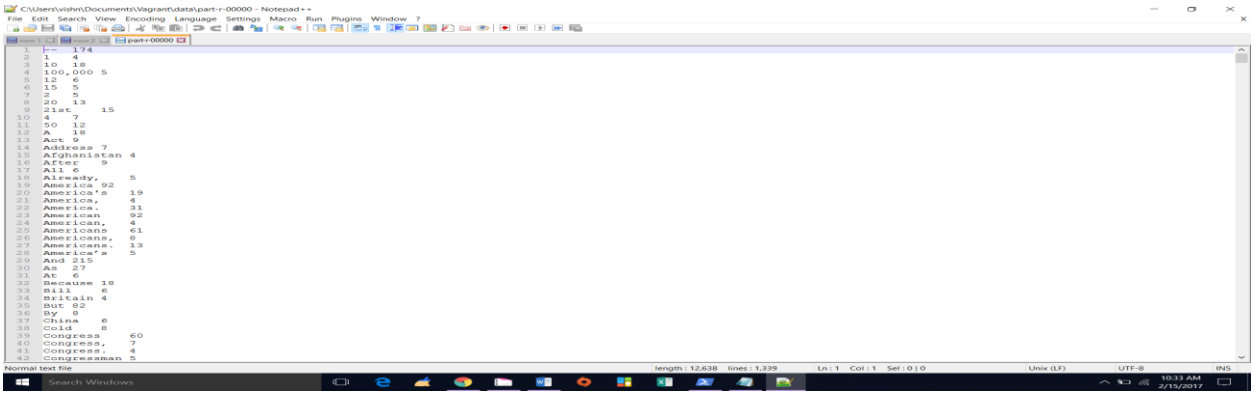
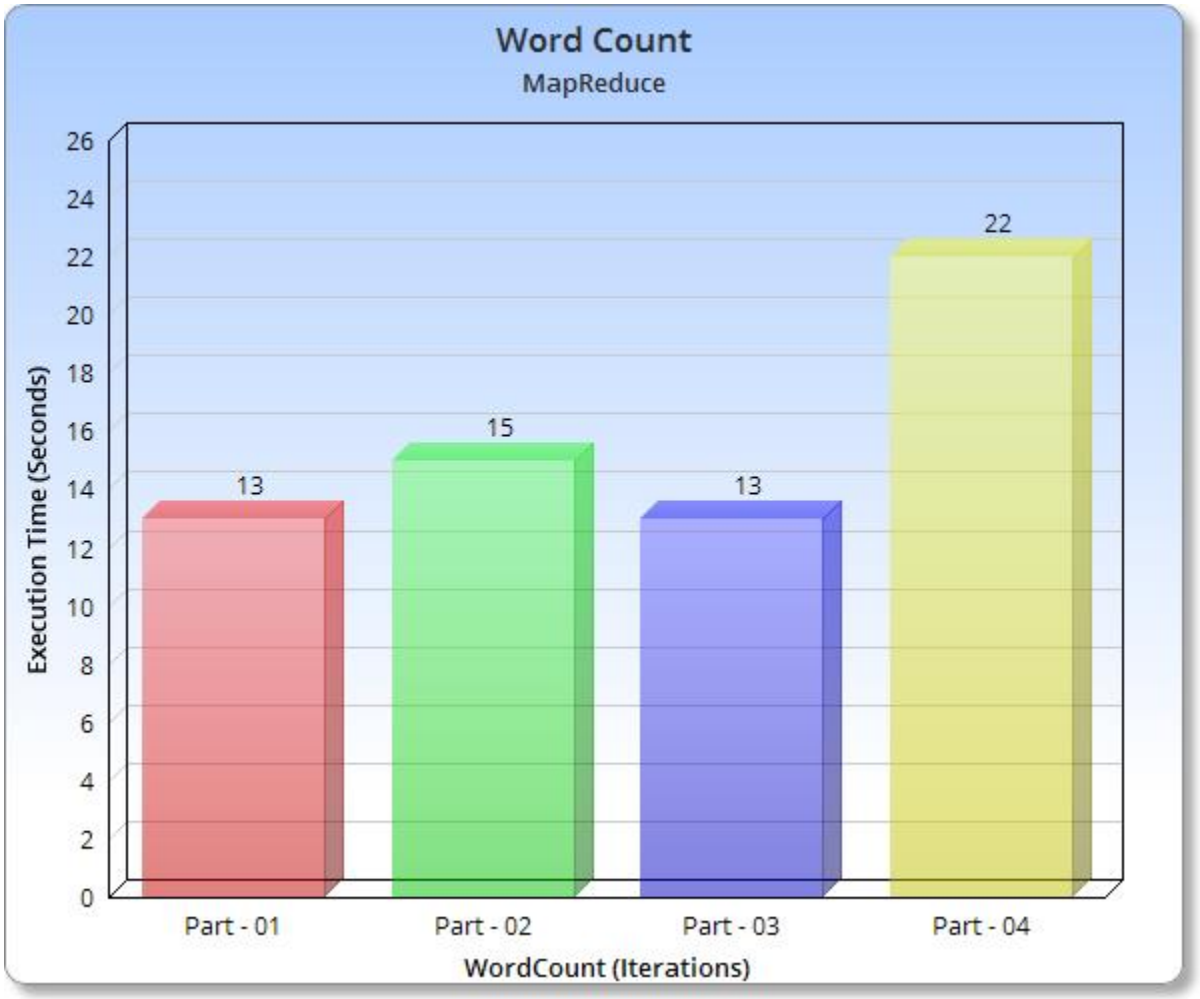


Chart & Graph:

S.No	teration	Total Maps
1	Part - 01 - Run Word Count 1 example on your local psudo-distributed system with supplied text files	1
2	Part - 02 - Run Word Count 2 example on your local psudo-distributed system with supplied text files	1
3	Part - 03 - Modify Wordcount 1 to look for only words that occur more than 4 times	1
4	Part - 04 - Modify Wordcount 2 to modify and use the -skip command	1

S.No	Start Time	End Time	Execution Time	Time Difference
1	15:48:35	15:48:48	***13s***	0s
2	16:20:19	16:20:34	***15s***	299s
3	16:32:23	16:32:36	***13s***	331s
4	16:42:20	16:42:42	***22s***	331s



Additional Runs:

- WordCount1 was executed with words repeating more than 10 times.
- WordCount2 was executed with both Case-Sensitive: ON and Case-Sensitive: OFF. Both the execution took approx. ~14s to finish its run.

Analysis:

- The run times of all the word count files took not more than 25 seconds to execute.
- Word Count 2 improves on Word Count by taking some of the features of MapReduce and implementing it.
- Most noticeable with Word Count 2 is that how DistributedCache can be used to distribute read-only data enhancing functions like case sensitivity, skipping selected patterns or words.
- It also makes use of Hadoop's inline commands providing a more specific result based on a given input.
- Larger datasets related to Word Count files can be set up in a cluster to improve efficiency in reading and parsing through data, provided the dataset is substantial (100GB or more).
- This exercise took just 1 map to run hence it can be seen that the complexity is really small.

Top 10 Words:

The most frequently used words under all 4 tasks can be seen below.

- I 353
- a 757
- and 1217
- for 445
- in 640
- is 393
- of 1142
- our 657
- that 571
- the 1867
- to 1433
- we 560
- will 394