1. How long did it take your job to run?

38sec

1. How many map tasks were used? Why did the system pick this number?

4 tasks，because the task was run against 4 files: Comedies, histories, poems, tragedies.

1. How many reduce jobs were used? Why did the system pick this number?

1, because it is set by default in reducer.

1. How many HDFS bytes were read? Using wc on the original input files, show where this number came from. Hint: Also look at the statistic “Input Split Bytes”.

5,284,706. The four original input files are 1784616, 1479035, 268140and 1752440. The input split bytes is 475. Total is 5284706.

Command:

[training@localhost shakespeare]$ ls

comedies glossary histories poems tragedies

[training@localhost shakespeare]$ wc comedies

59497 320884 1784616 comedies

[training@localhost shakespeare]$ wc histories

46029 261915 1479035 histories

[training@localhost shakespeare]$ wc poems

7308 48661 268140 poems

[training@localhost shakespeare]$ wc tragedies

60292 307776 1752440 tragedies

1. How many HDFS bytes were written? Using an HDFS command, show where this number came from.

299,379.

Command:

hadoop fs -ls wordcounts

-rw-r--r-- 1 training supergroup 0 2015-09-18 17:06 wordcounts/\_SUCCE SS

drwxr-xr-x - training supergroup 0 2015-09-18 17:06 wordcounts/\_logs

-rw-r--r-- 1 training supergroup 299379 2015-09-18 17:06 wordcounts/part-r -00000

1. How many input map records were there? Using wc on the original input files, show where this number came from.

173,126.

[training@localhost shakespeare]$ ls

comedies glossary histories poems tragedies

[training@localhost shakespeare]$ wc comedies

59497 320884 1784616 comedies

[training@localhost shakespeare]$ wc histories

46029 261915 1479035 histories

[training@localhost shakespeare]$ wc poems

7308 48661 268140 poems

[training@localhost shakespeare]$ wc tragedies

60292 307776 1752440 tragedies

59497 + 46029 + 7308 + 60292 = 173126

1. Are the number of map output records equal to the number of reduce input records?

Yes, the records are 964,453.

1. How many reduce output records are there? Using an HDFS command and the wc command, show where this number came from.

29,183

hadoop fs -get wordcounts/part-r-00000 ~/training\_materials/developer/data

[training@localhost data]$ wc part-r-00000

29183 58366 299379 part-r-00000

1. Going back to the main JobTracker status page, how many map and reduce slots (capacity) are available on the VM?

2 and 2

1. How many bytes are there in 128 MB?

128mb = 134217728 bytes