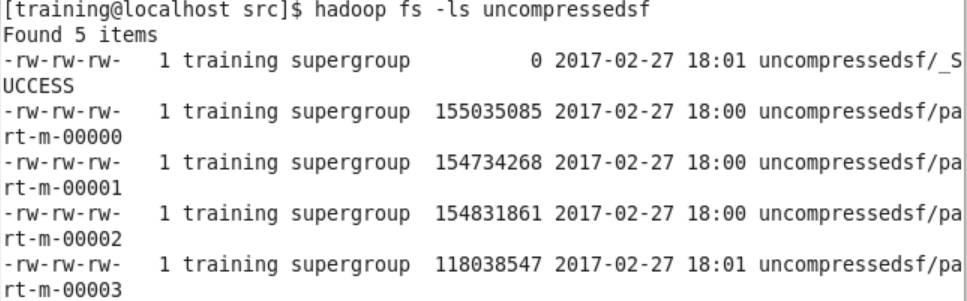
2. (a) 4 files are created.

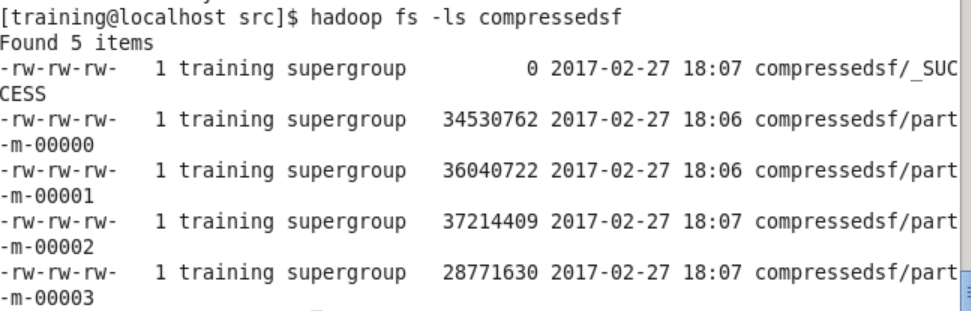


The signature of the file is like:



Which is the sequence with input type of LongWritable and output type of Text. There are four output files were created, because the number of output files depends on the number of mappers that the Hadoop uses in the map only job.

(c)



The signature for the compressed file is like:



Which is the sequence with input type of LongWritable and the output type of Text, also shows that the file is compressed using the SnappyCodec. The main difference is that it shows the information that it is compressed. The total size of the uncompressed file is 555.8MB, and the total size of the compressed file is 130.2MB, so the compression rate is 130.2/555.8 = 0.234.

(d) This output provide us the text information, which is the log file itself, which shows the IP addresses and the log data.

(e) The commend line is:

