1.Estimate minimum Namenode RAM size for HDFS with 1 PB capacity, block size 64 MB, average metadata size for each block is 300 B, replication factor is 3. Provide the formula for calculations and the result.

ram calculation for namenode

1PB/(64MB\*3)\*300B=1.56GB

2. HDDs in your cluster have the following characteristics: average reading speed is 60 MB/s, seek time is 5 ms. You want to spend 0.5 % time for seeking the block, i.e. seek time should be 200 times less than the time to read the block. Estimate the minimum block size.

block size is Y

best block reading time:

5ms\*200 =1s

reading speed is:

60/1s = Y/1sY = 60MB

3. To complete this task use the ['HDFS CLI Playground'](https://www.coursera.org/learn/big-data-essentials/ungradedLti/SGWv1/hdfs-cli-playground) item.

Create text file ‘test.txt’ in a local fs. Use HDFS CLI to make the following operations:

* сreate directory ‘assignment1’ in your home directory in HDFS (you can use a relative path or prescribe it explicitly "/user/jovyan/...")
* put test.txt in it
* output the size and the owner of the file
* revoke ‘read’ permission for ‘other users’
* read the first 10 lines of the file
* rename it to ‘test2.txt’.

4. Provide all the commands to HDFS CLI.

hdfs dfs -mkdir /user/jayasimha/assign-1: create a folder in Hadoop

hdfs dfs -put ~/test.txt assign-1/test.txt: Copy to local file Hadoop

hdfs dfs -ls /user/jayasimha/assign-1/test.txt: check file information

hdfs dfs -chmod o-r /user/jayasimha/assign-1/test.txt: delete read permission of other user

hdfs dfs -cat /user/jayasimha/assign-1/test.txt | head -10: read the first 10 lines

hdfs dfs -mv /user/shanavaz/assign-1/test.txt /user/shanavaz/assign-1/test2.txt: Change file name

5. To complete this task use the ['HDFS CLI Playground'](https://www.coursera.org/learn/big-data-essentials/supplement/ZGk8l/hdfs-cli-playground) item.

Use HDFS CLI to investigate the file ‘/data/wiki/en\_articles\_part/articles-part’ in HDFS:

* get blocks and their locations in HDFS for this file, show the command without an output
* get the information about any block of the file, show the command and the block locations from the output

hdfs fsck /data/wiki/en\_articles\_part/articles-part -files -blocks -locations

hdfs fsck /data/wiki/en\_articles\_part/articles-part