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.

1PB / (64MB * 3) * 300B = 10^15 / (64 * 10^6 * 3) * 300B = 10^9 / (64 * 3) * 300B = 1 562 500 000B = 1.6GB

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.

tread / tseek = 200
tread = 200 * tseek
tseek = 5ms = 0.005s
vread = 60MB/s
tread = size / vread
size = vread * tread = vread * 200 * tseek = 60MB/s * 200 * 0.005s = 60MB

3). Create text file ‘test.txt’ in a local fs. Use HDFS CLI to make the following operations:
create directory ‘assignment1’ in your home directory in HDFS (/user/<your_name>)
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’.
Provide all the commands to HDFS CLI.

hdfs dfs -mkdir -p assignment1
hdfs dfs -touchz assignment1/test.txt
hdfs dfs -ls /assignment1
hdfs dfs -chmod 640 /assignment1/test.txt
hdfs dfs -cat /assignment1/test.txt | head -10
hdfs dfs -mv /assignment1/test.txt /assignment1/test2.txt

Correct variant:
$ hdfs dfs -mkdir assignment1
$ hdfs dfs -put test.txt assignment1/
$ hdfs dfs -ls assignment1/test.txt or hdfs dfs -stat "%b %u" assignment1/test.txt
$ hdfs dfs -chmod o-r assignment1/test.txt

$ hdfs dfs -cat assignment1/test.txt | head -10
$ hdfs dfs -mv assignment1/test.txt assignment1/test2.txt

4). 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 -blockId blk_1073741825
Connecting to namenode via http://localhost:50070/fsck?ugi=jovyan&blockId=blk_1073741825+&path=%2Fdata%2Fwiki%2Fen_articles_part%2Farticles-part
FSCK started by jovyan (auth:SIMPLE) from /127.0.0.1 at Fri Mar 30 12:01:14 UTC 2018

Block Id: blk_1073741825
Block belongs to: /data/wiki/en_articles_part/articles-part
No. of Expected Replica: 1
No. of live Replica: 1
No. of excess Replica: 0
No. of stale Replica: 0
No. of decommissioned Replica: 0
No. of decommissioning Replica: 0
No. of corrupted Replica: 0
Block replica on datanode/rack: c0ff2c2e2b7b/default-rack is HEALTHY

5. Look at the picture of Namenode web interface.
Show the total capacity of this HDFS installation, used space and total data nodes in the cluster.

total capacity: 2.14TB
used space: 242.12GB
total data nodes: 4