**Session 3: HDFS Internals**

**Assignment 2**

**1. Which is not the property of hdfs-site.xml file?**

a. Block size

b. Replication factor

c. Secondary NameNode port address

d. Application server

**Solution:** (d) Application Server

**2. What are the properties that we can edit in hdfs-site.xml file?**

a. Block size

b. Replication factor

c. Block reporting interval

d. All the above

**Solution:** (d) All the above

**3. Underlying storage layers where MapReduce programs are written:**

a. Abstracted

b. Open to all

c. Locked

d. None of these

**Solution:** (a) Abstracted

**3. Which is not a valid method in FSDATA Output Stream?**

a. close( )

b. open( )

c. getpos( )

d. sync( )

**Solution:** (b) open

**5. How many blocks of size 128MB will be allocated for a file of size 524288KB:**

a. 2

b. 3

c. 4

d. 5

**Solution:**

**6. MapReduce Job client calculates the input split by:**

a. Figuring the first and last whole records in the block

b. Figuring only the first block

c. Figuring only the last block

d. None of these

**Solution:** (a) Figuring the first and last whole records in the block

**7. Mappers are directly related to:**

a. Input data

b. Input splits

c. Output data

c. Output splits

**Solution:** (b) Input Splits

**8. After data is written in an HDFS file, HDFS does not provide any guarantee that data are visible to a new reader until the file is \_\_\_\_\_\_**

a. closed

b. opened

c. writing

d. reading

**Solution:** (a) Closed

**9. What mechanism does Hadoop use to make namenode resilient to failure?**

a. Take backup of filesystem metadata to a local disk and a remote NFS mount

b. Store the filesystem metadata in cloud

c. Use a machine with at least 12 CPUs

d. Using expensive and reliable hardware

**Solution:** (d) Using expensive and reliable hardware

**10. All the files in a directory in HDFS can be merged together using:**

a. getmerge

b. putmerge

c. remerge

d. mergeall

**Solution:** (a) getmerge