**Session 3: HDFS Internals**

*Assignment 1 – Try the given quiz questions and provide the answers in a word document*

**1. HDFS is built around the idea that data is written \_\_\_\_\_but read many times.**

a) many

b) twice

c) data already exists

d) once

**Solution:** Once

**2. Hadoop divides input into fixed size pieces called what?**

a) output result

b) input splits

c) input data

d) input blogs

**Solution:** (b) input splits

**3. All the blocks are replicated in other nodes for \_\_\_\_\_\_**

a) security

b) big data

c) pool

d) fault tolerance

**Solution:** (d) Falut tolerance

**4. Block size can be changed using the properties in \_\_\_\_\_\_**

a) core-site.xml

b) Hadoop-env.sh

c) hdfs-site.xml

d) yarn-site.xml

**Solution** ( c) hdfs-site.xml

**5. Hadoop uses the \_\_\_\_\_\_representation of the data stored in the file blocks known as Input splits.**

a) physical

b) logical

c) mechanical

d) none

**Solution:** logical

**6. DFS calls NameNode to create file in file system’s\_\_\_\_\_**

a) dataspace

b) resourcespace

c) namespace

d) nodespace

**Solution:** (c) namespace

**7. Data packets are streamed to first DataNode in the \_\_\_\_\_\_\_\_**

a) handshake

b) pipeline

c) hard disk

d) hdfs

**Solution:** (b) pipeline

**8. The client has finished writing data, it calls \_\_\_\_\_\_\_on the stream.**

a) close()

b) read()

c) open()

d) check()

**Solution:** (a) Close

**9. Blocks are read in order, with the \_\_\_\_\_\_\_\_\_ opening new connections to datanodes as the client reads through the stream.**

a) DFSoutputstream

b) DFSInputStream

c) DFStrackManager

d) DFSStringConcatination

**Solution:** (b) DFSInputStream

**10. If I have 100 input splits, how many maps will run?**

a) 200

b) 50

c) 100

d) 1

**Solution :** (c) 100