Session 3: HDFS Internals

Assignment 1

1. **HDFS is built around the idea that data is written \_\_\_\_\_but read many times.** 
   1. many
   2. twice
   3. data already exists
   4. once Ans

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

a) output result

* 1. input splits Ans
  2. input data
  3. input blogs

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

a) security

* 1. big data
  2. pool
  3. fault tolerance Ans

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

a) core-site.xml

* 1. Hadoop-env.sh
  2. hdfs-site.xml Ans
  3. yarn-site.xml

1. **Hadoop uses the \_\_\_\_\_\_representation of the data stored in the file blocks known as Input splits.** 
   1. physical
   2. logical Ans
   3. mechanical
   4. none

1. **DFS calls NameNode to create file in file system’s\_\_\_\_\_**
2. dataspace
3. Resourcespace
   1. namespace Ans
   2. nodespace

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

a) handshake

* 1. Pipeline Ans
  2. hard disk
  3. hdfs Ans

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

a) close() Ans

* 1. read()
  2. open()
  3. check()

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

* 1. DFSoutputstream Ans
  2. DFSInputStream
  3. DFStrackManager
  4. DFSStringConcatination

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

a) 200

* 1. 50
  2. 100 Ans
  3. 1