**SESSION 3: HDFS INTERNALS**

**ASSIGNMENT 1**

1. HDFS is built around the idea that data is written \_\_\_\_\_but read many times.
2. many
3. twice
4. data already exists
5. **once**
6. Hadoop divides input into fixed size pieces called what?
7. output result
8. **input splits**
9. input data
10. input blogs
11. All the blocks are replicated in other nodes for \_\_\_\_\_\_
12. Security
13. big data
14. pool
15. **fault tolerance**
16. Block size can be changed using the properties in \_\_\_\_\_\_
17. core-site.xml
18. Hadoop-env.sh
19. **hdfs-site.xml**
20. yarn-site.xml
21. Hadoop uses the \_\_\_\_\_\_representation of the data stored in the file blocks known as Input splits.
22. Physical
23. **Logical**
24. Mechanical
25. None
26. DFS calls NameNode to create file in file system’s\_\_\_\_\_
27. Dataspace
28. Resourcespace
29. **Namespace**
30. Nodespace
31. Data packets are streamed to first DataNode in the \_\_\_\_\_\_\_\_
32. Handshake
33. **Pipeline**
34. hard disk
35. hdfs
36. The client has finished writing data, it calls \_\_\_\_\_\_\_on the stream.
37. **close()**
38. read()
39. open()
40. check()
41. Blocks are read in order, with the \_\_\_\_\_\_\_\_\_ opening new connections to datanodes as the client reads through the stream.
42. DFSoutputstream
43. **DFSInputStream**
44. DFStrackManager
45. DFSStringConcatination
46. If I have 100 input splits, how many maps will run?
47. 200
48. 50
49. **100**
50. 1