**SESSION 4: HADOOP CONFIGURATION & ECOSYSTEM**

**ASSIGNMENT 2**

1. The \_\_\_\_\_\_\_\_\_\_is the heart of an HDFS file system. It keeps the metadata such as directory tree of all files in the file system and tracks the cluster where the file data is present. The actual data is stored on \_\_\_\_\_\_\_as HDFS blocks.
2. datanode, datanode
3. resourcemanager, namenode
4. **namenode, datanode**
5. tasktracker, jobtracker
6. When the end of the block is reached, \_\_\_\_\_\_\_closes the connection to the datanode, then finds the best datanode for the next block.
7. DFSOutputStream
8. Client
9. **DFSInputStream**
10. Resourcemanager
11. When the client finishes reading, it calls \_\_\_\_method on the close stream.
12. read()
13. write()
14. shuffle
15. **close()**
16. Which configuration file contains Environmental variable settings used by Hadoop?
17. Core-site.xml
18. mapred-site.xml
19. yarn-site.xml
20. **Hadoop-env.sh**
21. Which MapReduce daemon instantiates user code, and executes map and reduce tasks on a cluster running MapReduce vl (MRvl)?
22. NameNode
23. DataNode
24. **JobTracker**
25. TaskTracker
26. Identify the function performed by the Secondary NameNode daemon on a cluster configured to run with a single NameNode.
27. **In this configuration, the Secondary NameNode performs a checkpoint operation on the files by the NameNode.**
28. In this configuration, the Secondary NameNode is standby NameNode, ready to failover and provide high availability.
29. In this configuration, the Secondary NameNode performs deal-time backups of the NameNode.
30. In this configuration, the Secondary NameNode servers as alternate data channel for clients to reach HDFS, should the NameNode become too busy.
31. Hadoop administrators write a script called Topology script to determine the rack location of nodes. It triggers to know the distance of the nodes to replicate the data and configures this script in \_\_\_\_\_\_\_\_.
32. yarn-site.xml
33. Hadoop-env.sh
34. **core-site.xml**
35. mapred-site.env
36. \_\_\_\_\_\_\_\_\_\_is the master that arbitrates all the available cluster resources and thus helps manage the distributed applications running on the YARN system
37. Node manager
38. data manager
39. **ResourceManager (RM)**
40. Taskmanager
41. \_\_\_\_\_\_\_\_\_take instructions from the ResourceManager and manage resources available on a single node.
42. **NodeManagers**
43. data manager
44. ResourceManager (RM)
45. Taskmanager
46. How does HDFS Federation help HDFS Scale horizontally?
47. HDFS Federation improves the resiliency of HDFS in the face of network issues by removing the NameNode as a single-point-of-failure.
48. HDFS Federation allows the Standby NameNode to automatically resume the services of an active NameNode
49. HDFS Federation provides cross-data center (non-local) support for HDFS, allowing a cluster administrator to split the Block Storage outside the local cluster.
50. **HDFS Federation reduces the load on any single NameNode by using the multiple, independent NameNode to manage individual parts of the filesystem namespace.**