0. Distributed File System (Refer: Data Intensive Text Processing using MapReduce)
1. Introduction to Hadoop (Refer: Slides, Hadoop Definitive Guide, and Tutorial links given below)
2. Commodity Hardware
3. HDFS
4. Design of HDFS
5. HDFS Architecture
6. Hadoop Deamons --> Name Node, Secondary Name Node, Data Node, Resource Manager, and Node Manager
7. Failure cases of Name Node
8. Failure cases of Data Node
9. HDFS Replica and Rack Awareness rules
10. Anatomay of File Read from HDFS
11. Anatomy of File Write in HDFS
12. HDFS Block
13. HDFS Block Caching
14. HDFS Federation
15. HDFS High Availability
16. Introduction to MapReduce
17. Mapper Phase and Reducer Phase
18. Working of MapReduce with Examples
19. Anatomy of MapReduce Job *
20. Failures Handling: Task Failure, Application Master Failure, Node Manager Failure, and Resource Manger Failure *
21. File Input Formats: TextInputFile Format (Discussed in Lab), N-Line File Format and Sequence File Formats (The last two topics will be discussed in next lab)
22. Combiner Concept in Hadoop (Discussed in Lab)
23. Partitioner in Hadoop (Discussed in Lab)
24. Customized Partitioner in Hadoop (Discussed in Lab)
25. Cluster Configuration details

Note: * these concepts are yet to be covered. These will be covered in next 2 classess.

https://data-flair.training/blogs/hadoop-hdfs-tutorial/

https://data-flair.training/blogs/hadoop-mapreduce-tutorial/

https://data-flair.training/blogs/hadoop-tutorial/

https://data-flair.training/forums/topic/what-is-sequencefileinputformat-in-hadoop-mapreduce/

https://www.edureka.co/blog/mapreduce-tutorial/