Module – 3

1. **Explain HDFS architecture with a neat diagram.**

   OR

   Discuss the critical or core components of Hadoop and their working along with a neat diagram.

   OR

   List and explain the technical features of Hadoop.
2. **Discuss the factors considered for design of HDFS and which are areas where HDFS is not good fit today.**
3. **Explain different HDFS concepts in detail.**

   OR

   What are the roles of a Blocks, Name Node, Data Node and Secondary Node.
4. **How do you define "block" in HDFS? What is the default block size in Hadoop-2? Can it be changed? If you have an input file of 350 MB, how many input splits would HDFS create and what would be the size of each input split?**
5. **Explain how HDFS high availability helps to recover the data from failed Name Node.**
6. **With a neat diagram, explain the anatomy of reading data from a file in HDFS.**

   OR

   Describe the sequence of events flow when client reading a file in HDFS with a neat diagram.
7. **Describe the sequence of events flow when client writing data in HDFS with a neat diagram.**

   OR

   With a neat diagram, explain the anatomy of writing data to a file in HDFS.

1. **What is MapReduce? Explain working of various phases of MapReduce.**
2. **What is Map Reduce? Sketch a neat diagram and explain the logical data flow in Map Reduce?**
3. **Write JAVA MapReduce code to find maximum temperature from the weather data set.**

4. **Explain with a diagram, MapReduce data flow with a single reduce task and multiple reduce task.**

   OR

   Explain MapReduce data flow with single reduce task.

   OR

   How does a MapReduce model works with a Single Reduce task. Explain with a neat diagram.
5. **Discuss the combiner function in MapReduce concept.**
6. **Write a note on Input splits on Hadoop.**
7. **Explain the 3 configuration files in developing Hadoop Application.**
8. **Explain in detail the steps involved in running the map reduce program in a cluster.**