UNIT-III

1 Mark Questions

1. Define MapReduce?
2. What are the sub phases in Map?
3. What is the role of RecordReader?
4. Define Combiner?
5. Define Partitioner?
6. Distinguish Combiner and Reducer?
7. What are MapReduce Daemons?
8. Write about Compression in MapReduce?
9. What are the methods used in Job Configuration?
10. What are the optimization techniques in MapReduce? Justify your answer.
11. What is the role of Task Tracker?
12. What are the methods of Mapper class?

Map,MapReduce,run clen setup map

1. What are the methods in Reducer class?Map,Reduce
2. Define Data Serialization?
3. Distinguish “Moving Code to the Data” versus “Moving Data to the Code”.

Moving data processing to sql

1. Write about Thirft?
2. Write about Protocol buffer?
3. Write about AVRO?
4. Write about Sequencefile?
5. Write about XML file?
6. Write about JSON?
7. What are the classes used to read/write XML files in MapReduce?

XMLInputFormat.java, XMLDriver.java, MyMapper.java

1. List the classes used to read/write JSON files in MapReduce?

JsonReader,JsonObect, JsonWriter, InputStream, outputStream

1. What is the need of Data serialization?
2. Why working with JSON files is more complicated compared to XML file?
3. List of real time applications of MapReduce?
4. What is the main purpose of InputFormat class?
5. What are the methods in RecordReader class?

getCurrentValue,getCurrentKey,getProgress,getKey,close

1. Write the methods in RecordWriter?

Close and RecordWrite

1. Write about OutpuFormat class?

**OutputFormat** describes the output-specification for a Map-Reduce job. On the basis of output specification

1. Implement wordcount using MapReduce with partitioner neat sketch.
2. Design MapReduce solution to find the years whose average sales is greater than 3000. 10M

input file format has year, sales of all months and average sales

year jan feb march april may june july august september october november december average

1. Write a MapReduce program to find Dept wise salary.

Empno EmpName Dept Salary

1. Explain Big Data serialization formats.

Unit II

* 1. Distinguish copyFromLocal command and put.
  2. Distinguish copyToLocal versus get.
  3. Give the options of rmr command.
  4. Write about Pig.
  5. Write about Hive.
  6. Write about Mahout.
  7. Write about sqoop.
  8. Write about Flume.
  9. Write the features of YARN.
  10. Resource Manager
  11. Give the functions of Node Manager.
  12. Give the functions of ApplicationManager.
  13. What is the role of ApplicationMaster.
  14. Write the limitations of Hadoop 1.0.
  15. Define TaskTracker.
  16. Define JobTracker.
  17. Give the Features of HDFS.
  18. Distinguish Active Namenode versus Passive Namenode.
  19. Write about Hbase.
  20. moveFromLocal versus moveToLocal.
      1. Explain the following

1. Hadoop Distributors 3M
2. MapReduce Programming architecture 3M
3. HDFS 2 features 2M
4. ApplicationManager 2M
   * 1. Explain Hadoop EcoSystem with neat Sketch.