

Subject - Big Data Hadoop	Max Marks - 100
Program - MCA Plus	Passing Marks - 50
Semester - III	Total Question - 10 (10 Marks Each)

**Important:** Student should attempt every question and submit the answer sheet with response to below questions.

Questions:

1. Do every business unit/organization require Hadoop? If not, describe the situations when distributed (HDFS) is required. What are the cost implications? What are the administrative implications?
2. What is Text Input Format? Explain with appropriate example.
3. State the usage of 'filters', 'group', 'orderBy', 'distinct' keywords in pig scripts.
4. How does bucketing help in the faster execution of queries? Support your answer with programming code related to this.
5. Explain JobConf in MapReduce.
6. What is InputFormat in Hadoop?
7. How Big data is used in education industry. Explain briefly.
8. Explain the three of Vs related to big data generation
9. Why do we want to process big data? Write any give reasons.
10. Explain deletion in Hbase? Mention what are the three types of tombstone markers in Hbase?

**Kindly mention the following details at the top of your Lab Submission Sheet:**

Reference/enrollment number.	
Student Name	
Subject Name	
Semester/Year	

**\*Please Note:** If any of the details are missing your Lab Submission will not be considered for evaluation. University is not responsible for any consequences arising due non submission of details.