

Roll No.

Total No. of Pages : 01

Total No. of Questions : 08

M.Tech.(CSE) (2015 to 2017) EL-I /(IT) PT (2015 & Onwards)(E-3)  
(Sem.-5)

**BIG DATA ANALYSIS**  
Subject Code : MTCS-302  
M.Code : 74152

Time : 3 Hrs.

Max. Marks : 100

INSTRUCTIONS TO CANDIDATES :

1. Attempt any FIVE questions out of EIGHT questions.
2. Each question carries TWENTY marks.
  
1. A. What is Big data? Discuss it in terms of four dimensions, volume, velocity, variety and veracity.  
B. Explain the techniques to optimize MapReduce jobs.
2. A. Briefly explain the steps involved in designing MapReduce.  
B. Show on how a client read and write data in HDFS, give an example code.
3. How NoSQL is different from other models? Discuss features of NoSQL in detail.
- 4 A. Explain flume and Oozie.  
B. What are schema-less databases? Explain in detail.
5. A. Compare and contrast sharing and replication.  
B. Write in brief data visualization techniques and tools used in representation.
6. A. Explain Spark components in detail. Also list the features of spark.  
B. How to customize data with RecordWriter?
7. Explain the need of R scripting language in big data analysis. How to read datasets and export data from R? Explain with the help of example. Also explain the concept of vectors in R. Discuss the following : A. Zookeeper B. Distribution Models C. Yarn D. Packages in R
8. Packages in R

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student