

Roll No.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Total No. of Pages : 01

Total No. of Questions : 08

M.Tech (CSE Engg.) Big data (PIT) (Sem.-2)
FUNDAMENTALS OF BIG DATA ANALYTICS

Subject Code : CSB-211

Paper ID : [51091]

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :**1. Attempt any FIVE questions out of EIGHT questions.****2. Each question carries TWELVE marks.**

- Q1) Why do we need Hadoop for Big Data Analytics? Explain the different features of Hadoop. List various configuration files used in Hadoop Installation. What is use of mapred-site.xml?
- Q2) What is NoSQL? Explain the difference between NoSQL v/s Relational database. What do you understand by "Polyglot Persistence" in NoSQL?
- Q3) Consider the unlabeled two-dimensional data represented as (2,3), (2,-3), (3,9), (-3,3), (1,2), (2,7), (4,3), (-3,5). Using the two points (1,2) & (2,7) as initial centroids.
- a) Draw the clusters obtained after one iteration of the k-means algorithm.
- b) Does your solution change after another iteration of the k-means algorithm?
- Q4) How to build logistic regression model in R? What is over-fitting and under-fitting? How these problems can be resolved in linear and logistic regression?
- Q5) Explain the data import in R language. Explain how to communicate the outputs of data analysis using R language. Difference between library () and require () functions in R language.
- Q6) What is Big Data? Explain how big data processing differs from distributed processing? List various application of big data.
- Q7) Discuss in detail various Phases of Data Analytics Lifecycle. What is the role of a data scientist?
- Q8) What are the various Project Roles of delivery team? Discuss the responsibilities of each of the Project Roles. How to structure a high performance Analytics Team?