[6003]-711

T.E. (I.T.)

DATA SCIENCE AND BIG DATA ANALYTICS

(2019 Pattern) (Semester-II) (314452)

Time : 2½ Hours]

[Max. Marks : 70

Instructions to the candidates:

1) *Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8.*
2) *Figures to the right indicate full marks.*
3) *Neat diagrams must be drawn wherever necessary.*
4) *Assume suitable data, if necessary.*

Q1) a) Explain Google file system and its advantages. **[10]**

  b) Explain Hadoop distributed file system **[8]**

OR

Q2) a) Why map reduce is required in Hadoop? Explain the stages involved in map reduce task with a suitable example? **[9]**

  b) Describe the various types of NoSQL Databases with example and also compare them. **[9]**

Q3) a) Explain Mean, Mode and variance and standard deviation with suitable example. **[9]**

  b) Draw and explain Architecture of HIVE **[8]**

OR

Q4) a) Explain Min-max scaling. For the following dataset carry out min-max Scaling, X=24, 28, 53, 30, 40, 18, 15, 21 **[9]**

  b) What is data Wrangling? Why do you need it? explain data Wrangling methods? **[8]**

Q5) a) Explain any 4 Types of data visualization with example. **[9]**

  b) Explain different data visualization tools. **[9]**

OR

P.T.O.

Q6) a) Explain data visualization with the help of example? What are the advantages of data visualization? **[9]**

b) Explain Data Visulization with Tableau. **[9]**

Q7) a) Explain Big Data Analytics Challenges in brief. **[9]**

b) Explain types of Mobile Analytics. **[8]**

<div align="center">OR</div>

Q8) a) What is Porters valuation creation model? Explain porter's value chain analysis. **[9]**

b) What is social media analytic? Explain the process of social media data analytic. **[8]**

<div align="center">❖ ❖ ❖</div>