[5560]-600

T.E. (I.T.)

DATA SCIENCE AND BIG DATA ANALYTICS

(2015 Pattern) (Semester - II) (End Sem.) (314454)

Time : 2½ Hours]   *[Max. Marks : 70*

Instructions to the candidates:

   1)   *Answer Q1 or Q2, Q3 or Q4, Q5 or Q6, Q7 or Q8, Q9 or Q10.*

   2)   *Neat diagrams must be drawn wherever necessary.*

   3)   *Figures to the right side indicate full marks.*

   4)   *Assume suitable data if necessary.*

Q1) a)   Define Big Data. Enlist the differences and similarity in Big Data and Data Science with example.   **[6]**

      b)   Prove that principle of linearity of Expectation along with the usage related to Big Data.   **[4]**

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

Q2) a)   Determine the distinct element in the following stream using Flajolet Martin algorithm. Also design a suitable hash function which will give accurate result. Input stream - 1, 3, 2, 1, 2, 3, 4, 3, 1, 2, 3, 1.   **[6]**

      b)   Explain the big data infrastructure requirement for mobile computing.**[4]**

Q3) a)   Suppose that an orange juice company controls 20% of the orange juice market. Suppose they hire a market research company to predict the effects of an aggressive ad campaign. Suppose they conclude -   **[6]**

       •   Someone using Brand A will stay with Brand A with probability 90%.

       •   Someone not using Brand A will switch to Brand A with 70% probability.

         i)   What is the initial state matrix.

        ii)   Write down transition probability matrix.

        iii)   Calculate the probability that someone uses Brand A after 1 week.

      b)   Write a short note on Textual ETL.   **[4]**

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

Q4) a) Explain the role of following components in Hadoop system. **[6]**
   - Name node.
   - Secondary name node.
   - Data node.

   b) Discuss advantages and disadvantages of distributed processing. **[4]**

Q5) a) List and explain the different kind of roles for an analytics project. **[8]**

   b) List and explain the steps in discovery phase of data analytics life cycle. **[8]**

OR

Q6) a) Explain the data analytics life cycle in big data. **[8]**

   b) Explain the terms data cleaning and data transformation with the help of example. **[8]**

Q7) a) Discuss different types of data visualization techniques? Why visualization is more important? **[8]**

   b) What are the challenges in visualizing big data and how to overcome them? **[8]**

OR

Q8) a) What are the advantages of data visualization? Write a short note on any three tools used for data visualization. **[8]**

   b) Write short note on : **[8]**
   i) Google chart API
   ii) Scatter plot
   iii) Gephi
   iv) Cloudera

Q9) Write short note on: **[9]**
   a) Social media analytics.
   b) Roles and responsibilities of big data person.
   c) Mobile analytics.

OR

Q10) Write a short note on : **[9]**
   a) Difference between mobile analytics and social media analytics.
   b) Text mining.
   c) Role of Data scientist.

⌘⌘⌘