

Reg. No. :

Question Paper Code : 53227

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2019.

Seventh Semester

Information Technology

IT 6006 — DATA ANALYTICS

(Common to Computer Science and Engineering)

(Regulation 2013)

(Also common to PTIT 6006 – Data Analytics For B.E. (Part-Time)
for Seventh Semester – Computer Science and Engineering – Regulation 2014)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Give some applications of big data.
2. Determine the key factors needed to define a report.
3. Relate the purpose of regression analysis.
4. What do you mean by fuzzy qualitative model?
5. Why do we need Bloom filter in filtering streams?
6. What do you mean by k^{th} moment of the data stream?
7. Define association rule.
8. Identify the alternative rules for controlling hierarchical clustering.
9. Mention some main goals of Hadoop.
10. Outline the main advantages of visual data exploration.

PART B — (5 × 13 = 65 marks)

11. (a) Analyze the role of Analytic Sandbox and its benefits in the Analytic Process. (13)

Or

(b) Outline the various commonly used modern data analytic tools. (13)

12. (a) (i) Explain the steps involved in Bayesian data analysis. (6)

(ii) List and explain the components of time series model. (7)

Or

(b) (i) Define competitive learning. Summarize the basic competitive learning algorithm. (6)

(ii) List out the various techniques for Multivariate analysis. Explain any two. (7)

13. (a) (i) Draw and explain the architecture of data stream model. (9)

(ii) Identify the major issues in data stream query processing. (4)

Or

(b) (i) Discriminate the concept of sampling data in a stream. (9)

(ii) Illustrate the various Real Time Analytic Platform (RTAP) applications. (4)

14. (a) Explain the concept of Apriori algorithm with an illustrative example.

Or

(b) (i) Brief about the working of CLIQUE algorithm. (9)

(ii) Suppose a cluster of three-dimensional points has standard deviations of 2, 3, and 5, in the three dimensions, in that order. Compute the Mahalanobis distance between the origin (0,0,0) and the point (1, -3, 4). (4)

15. (a) (i) Describe the architecture of Hive with its features. (9)

(ii) Brief about the Main components of MapReduce. (4)

Or

(b) Summarize the various visual data analysis techniques.

PART C — (1 × 15 = 15 marks)

16. (a) Categorize the diversity of Intelligent Data Analysis applications.

Or

(b) Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if i divides b with no remainder. Thus, item 1 is in all the baskets, item 2 is in all fifty of the even-numbered baskets, and so on. Basket 12 consists of items {1, 2, 3, 4, 6, 12}, since these are all the integers that divide 12. Answer the following questions:

(i) If the support threshold is 5, which items are frequent? (5)

(ii) What is the confidence of the following association rules? (5 + 5)

{ 5, 7 } → 2

{ 2, 3, 4 } → 5