Roll No. | | | | | | | | | | |

B.Tech (Full Time) DEGREE EXAMINATIONS, APRIL 2014

INFORMATION TECHNOLOGY

Semester VI

IT9043 Data Analytics

(Regulation 2008)

Time: 3 Hours | Answer ALL Questions | Max. Marks 100

PART-A (10 x 2 = 20 Marks)

1. List out the characteristics of big data and challenges in handling big data.
2. What is the role of ETL tool? List the possible operations associated with it.
3. Highlight the uses of regression medeling.
4. List any four commonly used fuzzy membership functions.
5. Write any two examples for stream data.
6. What are the advantages of using real time data in big data analytics?
7. What are the advantages of using data mining techniques?
8. Explain the role of hash tree in association rule discovery.
9. Why the partitions are shuffled in map-reduce?
10. Who is generating big data and what are the ecosystem projects used for processing?

Part – B (5 x 16 = 80 marks)

11. (i) Highlight the features of Hadoop and explain the functionalities of Hadoop cluster. (8)

    (ii) Consider a collection of literature survey made by a researcher in the form of a text document with respect to cloud and big data analytics. Using Hadoop and map reduce, write a program to count the occurrence of predominant key words. (8)

12. (a) (i) Discuss the evolution of big data analytics. (8)

    (ii) Describe the role of sampling distributions in inferential statistics. (8)

OR

(b) (i) Highlight the features of modern data analytics tool. (8)

(ii) List any two possible web data from which effective analysis can be carried out. Justify your answer with an example. (8)

13. (a) (i) Explain the architecture of multi layer feed forward neural network. (8)

(ii) What is a Bayesian network? With an example, explain how this network can be used for analyzing data. (8)

OR

(b) (i) Given a set of person's name and height, Write a member function for calculating the degree of tallness and draw the graph for the same. (8)

(ii) What is a rule based system? Explain the inference associated with it. (8)

14. (a) Explain Apriori algorithm and with an example show how association rules are generated from frequent item sets. (16)

OR

(b) (i) Explain k-means clustering algorithm with an example. (8)

(ii) List the different hierarchical clustering techniques and explain any one. (8)

15. (a) (i) Explain the architecture for processing streaming data. (8)

(ii) explain the system architecture and components of Hive. (8)

OR

(b) (i) Using DGIM method, show how the bits that are flowing through a sliding window can be counted (10)

(ii) Highlight the features of NoSQL. (6)