Roll No.										

B.E. / B. Tech. (FULL TIME) DEGREE END SEMESTER EXAMINATIONS, APRIL/MAY 2013
INFORMATION TECHNOLOGY
VI SEMESTER

IT 9043 – DATA ANALYTICS

(REGULATIONS – 2008)

Time: 3 Hours **Maximum Marks: 100**

Answer All Questions

PART – A (10 x 2 = 20 Marks)

1. List the five essential characteristics of a cloud environment.
2. State the difference between a traditional query and an MPP query.
3. What is stratified sampling technique?
4. How a support vector based classifier is different from that of a linear classifier?
5. State the steps to be followed to maintain DGIM conditions when a new bit comes in and the buckets need to be modified.
6. Give any two real time stream queries.
7. Define support and confidence.
8. Give examples for symmetric and asymmetric binary variables and also state the measure for computing the dissimilarity between asymmetric binary variables.
9. What is CAP theorem? State its significances.
10. What are the components of Hadoop framework?

PART – B (5 x 16 = 80 Marks)

11. a)(i) Discuss about the differences between a traditional analytic and modern data warehouse architecture. (8)
    (ii) Explain how massively parallel processing systems shall be used for data preparation and scoring. (8)

12. a) (i) Describe the steps involved in support vector based inference methodology. (16)

(OR)

b) (i) Write the algorithm for neural network based learning for classification or prediction using back propagation. (8)
(ii) Write a short note on Bayesian inference methodology. (8)

13. a) (i) Explain how to count ones in a window using DGIM algorithm. Also explain how query answering is done using DGIM algorithm. (16)

(OR)

b) Write a short note on the following:
(i) Counting Distinct elements in a stream. (8)
(ii) Finding most popular elements using decaying window. (8)

14. a) (i) A database has five transactions. Let min sup = 60% and min conf = 80%.

TID	**Items**
T100	Milk, Onion, Nuts, Kiwi, Egg, Yoghurt
T200	Dhal, Onion, Nuts, Kiwi, Egg, Yoghurt
T300	Milk, Apple, Kiwi, Egg
T400	Milk, Curd, Kiwi, Yoghurt
T500	Curd, Onion, Kiwi, Icecream, Egg

Find all frequent itemsets using Apriori method. (16)

(OR)

b) (i) Discuss the various steps of PROCLUS clustering algorithm and also give its significances. (16)

15. a) (i) State the significances of MapReduce and discuss about Hadoop distributed file system architecture with a neat diagram. (16)

(OR)

b) Write a short note on the following:
(i) NoSQL Databases and its types. (8)
(ii) Visualization for Big Data. (8)