

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**Question Paper Code : 80594**

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2016.

Seventh Semester

Computer Science and Engineering

IT 6006 — DATA ANALYTICS

(Common to Information Technology)

(Regulations 2013)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. List the characteristics of big data.
2. Tabulate the differences between analysis and reporting.
3. Illustrate a neural network.
4. Define rule induction.
5. Give the advantages of the algorithm used in estimating moments.
6. Differentiate between data stream mining and traditional data mining.
7. State the use of a priori algorithm in data mining.
8. What are the problems faced if clustering exists in non-euclidean space?
9. Summarize the features of Hive.
10. Why is HDFS preferred to RDBMS?



PART B — (5 × 16 = 80 marks)

11. (a) Discuss statistical inference in detail. (16)

Or

(b) (i) What are the issues and challenges related to storage and transport in big data? (8)

(ii) Draw a comparison between the traditional analytic architecture and modern in-database architecture. (8)

12. (a) How is multivariate analysis performed in big data? Show with an illustration. (16)

Or

(b) List all the methods of stochastic search. Elucidate any four in detail. (16)

13. (a) What are streams? Explain stream data model with its architecture. (16)

Or

(b) Taking stock market predictions as a case study, elaborate on the Realtime Analytics Platform (RTAP). Present the assumptions made. (16)

14. (a) Describe the various hierarchical methods of cluster analysis. (16)

Or

(b) Write short notes on the following :

(i) CLIQUE. (8)

(ii) Frequent pattern based clustering methods. (8)

15. (a) Deliberate how data flow takes place in MapReduce framework. (16)

Or

(b) Discuss the different visual data analysis techniques with diagrams and graphs. (16)