

Reg. No. :

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

**Question Paper Code: 31289**

B.E. / B.Tech. DEGREE EXAMINATION, MAY 2017

Elective

Computer Science and Engineering

01UCS917 - MASSIVE DATASET ANALYTICS

(Regulation 2013)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 2 = 20 Marks)

1. List the characteristics of big data and challenges in handling big data.
2. Write any two possible web data from which effective analysis can be carried out.
3. Highlight the uses of regression modeling.
4. Define principal component analysis.
5. Give any two examples for stream data.
6. State how to count the distinct elements in a stream.
7. List the different hierarchical clustering techniques.
8. Define K-Means clustering algorithm.
9. State the significances of Map Reduce.
10. List the components of Hadoop framework.

PART - B (5 x 16 = 80 Marks)

11. (a) (i) Discuss the evolution of big data analytics. (8)  
(ii) Explain in detail about the major resampling techniques. (8)

Or

- (b) (i) Highlight the features of modern data analytics tools. (8)
- (ii) Compare analysis and reporting. (8)
12. (a) (i) What is a Bayesian network? With an example, explain how this network can be used for analyzing data. (8)
- (ii) Describe the steps involved in support vector based inference methodology. (8)

Or

- (b) (i) Explain the architecture of multi layer feed forward neural network. (8)
- (ii) Explain in detail about extracting fuzzy models from data. (8)
13. (a) (i) Explain the architecture for processing streaming data. (8)
- (ii) Discuss the concept of decaying window in detail. (8)

Or

- (b) (i) Explain how to count ones in a window using DGIM algorithm. (10)
- (ii) Describe about any one Real Time Analytics Platform (RTAP) application. (6)
14. (a) Explain Apriori algorithm and with an example show how association rules are generated from frequent item sets. (16)

Or

- (b) (i) Discuss the various steps of PROCLUS clustering algorithm and also give its significances. (8)
- (ii) Describe about Stream clustering and Parallel clustering. (8)
15. (a) (i) Describe Map Reduce framework in detail with neat diagram. (10)
- (ii) Highlight the features of NoSQL. (6)

Or

- (b) (i) Discuss about Hadoop Distributed File System architecture with a neat diagram. (10)
- (ii) Write short notes on Visualization for Big Data. (6)