CS801

USN	1	M	S							

M S RAMAIAH INSTITUTE OF TECHNOLOGY

(AUTONOMOUS INSTITUTE, AFFILIATED TO VTU)

BANGALORE – 560 054

SEMESTER END EXAMINATIONS – JUNE 2015

Course & Branch	:	**B.E.- Computer Science & Engg.**	**Semester**	:	**VIII**
Subject	:	**Big Data and Data Science**	**Max. Marks**	:	**100**
Subject Code	:	**CS801**	**Duration**	:	**3 Hrs**

Instructions to the Candidates:

- Answer one full question from each unit.
- Write the syntax of the function with related to the topic even though not specified in the question.

UNIT – I

1	a)	Define Big Data. Explain the characteristics of Big Data with examples.	(04)
	b)	List and explain the challenges of Big Data.	(08)
	c)	What is cloud computing? Explain the key concepts of cloud computing? How does cloud computing affect Big Data?	(08)
2	a)	How is IoT related to Big Data? List the features in IoT that conform to the Big Data paradigm.	(06)
	b)	What are the development opportunities and challenges that Big Data bring about in the data centers?	(06)
	c)	What is hadoop? What are the advantages of Hadoop that are relevant to the management and analysis of Big Data?	(08)

UNIT – II

3.	a)	Explain how the following data collection methods are used to collect data from the data generation environment i) Log Files ii) Sensors iii) Mobile Equipments	(06)
	b)	What is Data Pre-processing? Explain all the relational Data Pre-processing techniques.	(08)
	c)	Explain any two NoSQL databases which are based on a certain data model with relevant examples.	(06)
4.	a)	Bring out the difference b/w the following Traditional Data Analysis methods: - Cluster Analysis, Factor Analysis, Correlation Analysis, and Statistical Analysis.	(08)
	b)	Explain any three Big Data Analytic Methods. What are the methods used to rapidly extract key information from massive data.	(06)
	c)	List and explain the different analytical Architecture for Big Data Analysis.	(06)

UNIT – III

5.	a)	Describe the responsibilities of the MapReduce Execution framework.	(08)

b) Illustrate the roles and responsibilities of Partitioners and Combiners in the complete MapReduce model. (06)

c) Explain with a neat schematic diagram the architecture of the hadoop distributed file system (06)

6 a) Illustrate the components of the complete Hadoop cluster. (08)

b) Design a MapReduce PageRank algorithm and briefly Explain the map and reduce phases of the algorithm (06)

c) What are the responsibilities of the HDFS namenode? (06)

UNIT – IV

7. a) Explain the following w.r.t R data structures: Vector, Matrix, Array, Data frame. (10)

b) With an example explain why Factors are crucial in R. (05)

c) What are lists in R programming language? Why are lists important R structures? (05)

8. a) What is Product classification? Explain any three common classification methods. (08)

b) What is clustering? List any two common clustering methods and explain their use. (08)

c) Differentiate b/w Linear regression and Logistic regression. (04)

UNIT – V

9. a) What is Text Data Analysis? Elaborate on the different technologies involved in the Text Data Analysis. (10)

b) What is Web Data Analysis? List and explain how the classifications of Web Analysis. (10)

10. a) Explain the following w.r.t the Multimedia Data Analysis :- (10)
i) multimedia summarization
ii) multimedia annotation
iii) multimedia suggestion

b) Write a short note on the following application of Big Data (10)
i) Collective Intelligence
ii) Smart Grid

****