

**M. TECH. –III (COMPUTER ENGINEERING) (CBCS – 2015  
COURSE) : WINTER - 2017**

**SUBJECT: ELECTIVE –I: d) BIG DATA ANALYTICS**

Day: **Tuesday**  
Date: **16/01/2018**

Time: **11.00 AM TO 02.00 PM**  
Max. Marks: 60

**W-2017-2853**

**N.B.:**

- 1) All questions are **COMPULSORY**.
- 2) Figures to the right indicate **FULL** marks.
- 3) Answers to both the sections should be written in **SEPARATE** answer book.
- 4) Assume suitable data if necessary.

**SECTION- I**

**Q.1** A company is having its own data center and have made investments in physical infrastructure. They want to examine the requirements layer-by-layer to implement the reference architecture. What principles of Big Data Should be applied to scale the demands of the business. **(10)**

**OR**

**Q.1** Draw the Process Flow for traditional patient Diagnostic Process and Big Data Leveraged Patient Diagnostic Process and Clearly Compare the two process flows.

**Q.2** Explain the Design Goals of MongoDB. **(10)**

**OR**

**Q.2** A Client needs to maintain database design for a blog website. List the RDBMS scheme for the design of comments, post and tag list. Also write the structure of collection in MongoDB Schema.

**Q.3** The average value per square foot of a new house is \$40 and the average lot sells for \$ 20,000. What is predicted selling price of house of size 'X'? Define a probabilistic model for computing actual selling price and perform simple Regression. **(10)**

**OR**

**Q.3** Car dealers use Blue Book to help them determine value of used cars when customers purchase new cars. An experiment is to be conducted to whether an odometer should be included in Blue Book. Which Regression Model is to be used? Explain with example.

**SECTION- II**

**Q.4.** Explain the process of conversion of unstructured data for dynamic data Records using Sentiment Analysis. **(10)**

**OR**

**Q.4.** How aggregation is used in Data Reduction Techniques?

**Q.5** Derive the K-means clustering algorithm and give its advantages and disadvantages. **(10)**

**OR**

**Q.5.** Find the Clustering Pattern for the data set consisting scores for two variables.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

**Q.6** Explain Distribute File Systems of Hadoop. **(10)**

**OR**

**Q.6** Write Short notes on  
a) MapReduce  
b) NoSQL Databases