

**B.TECH. SEM -VII (COMPUTER) 2014 COURSE (CBCS) : WINTER
- 2017**

SUBJECT: BIG DATA ANALYTICS AND ARCHITECTURE

Day : **Wednesday**
Date : **17/01/2018**

Time : **02.30 PM TO 05.30 PM**
Max Marks : **60**

W-2017-2278

N.B. :

- 1) All Questions are **COMPULSORY**.
- 2) Figures to the **RIGHT** indicate full marks.
- 3) Assume suitable data **WHEREVER** necessary.

Q1. Define Big Data. Give two examples of big data case studies, indicate which V's are satisfied by these case studies. **(10)**

OR

Illustrate Big Data Management life cycle phases with an example. **(10)**

Q2. a) Consider the following data: **(05)**

1% of people have certain genetic defects.

90% of test for the gene detect the defects.

9.6% of the tests are false positive.

If a person gets a positive test results, what is the probability of that person having the genetic defect?

b) What is Biases? Explain different types of biases with example. **(05)**

OR

a) Suppose everyone who visits a retail website gets one promotional offer or no promotional offer at all, we want to see if making a promotional offer makes a difference. What statistical method you recommend for this analysis. Explain. **(05)**

b) The values of x and their corresponding values of y are shown in the table below. **(05)**

x	0	1	2	3	4
y	2	3	5	4	6

i) Find the least square regression line $y = ax + b$.

ii) Estimate the Value of y when $x = 10$

Q3. a) What R commands would you use to: **(05)**

i) Remove NULL values from a Dataset.

ii) Install an additional R Package.

iii) Encode a vector as a category.

iv) Import and export data.

b) In which Data analytics lifecycle phase would the team expect to invest most of the project time? Why? Where would the team expect to spend the least time? **(05)**

OR

a) What kind of tools would be used in the following Data analytics lifecycle phases and for which kinds of use scenarios? **(05)**

i) Data Preparation.

ii) Model Building.

b) What is Type-I error and Type-II error? Is Type-I error always more serious than the Type-II error? Why? **(05)**

P.T.O.

- Q4** What is NoSQL? What are the business drivers for NoSQL? Discuss any two architectural patterns of NoSQL. (10)
- OR**
- Explain in detail the Big Data Technology Stack with neat labeled diagram. (10)
- Q5. a)** What is Map reduce? Describe how Map reduce works? (05)
- b)** Describe the role of Pig and Hive in Hadoop ecosystem. (05)
- OR**
- a)** Enlist and describe the components in Hadoop ecosystem. (05)
- b)** Write Map reduce algorithm for natural join of two relations. (05)
- Q6.** What is Sentimental Analysis? Explain different types of sentimental analysis. Suggest how can sentimental analysis be used in e-commerce and social networking domain? (10)
- OR**
- Explain collaborative filtering base recommendation system. How it is different from content based recommendation system? (10)

* * * *