Total No. of Pages:2

Roll No...

VIIIth SEMESTER

MID SEMESTER EXAMINATION

B.Tech. March-2019

Paper Code: CO414 Time: 1:30 Hours Title: BIG DATA ANALYTICS

Max. Marks:30

Note:

1. Attempt all the questions.

- 2. Assume any suitable value(s) for missing data.
- 3. Marks are indicated against each question.

1. Answer the following questions briefly:

[5x1=5]

- (a) What are the three main characteristics of Big Data?
- (b) What are the three categories of Big Data?
- (c) Define the term "Moments"
- (d) What are the four main libraries in Hadoop?
- (e) Write down the main features of MapReduce?
- 2. (a) Explain the term "Intelligent Data Analysis". Describe various kinds of Data Analysis along with examples. [2.5]
 - (b) Define the term "Resampling". What are the various techniques for resampling? [2.5]
- 3. Consider a random sample of size m = 25 from a normal population with known a = 5.4 and unknown mean μ . The observed sample mean is x' = 128.
 - (a) Test the hypothesis H0: $\mu > = 130$, against H0: $\mu < 130$.

[3]

(b) Calculate the p-value of the observed outcome x' = 128.

[2]

4. Suppose we have a stream of tuples with the schema:

Grades (university, courseID, studentID, grade)

Assume universities are unique, but a courseID is unique only within a university (i.e., different universities may have different courses with

be both hospital

P. T.O.

P.T.O.

the same ID, e.g., "CS101") and likewise, studentID's are unique only within a university (different universities may assign the same ID to different students). Suppose we want to answer certain queries approximately from a 1/20th sample of the data. For each of the queries below, indicate how you would construct the sample.

- (a) For each university, estimate the average number of students in a course.
- (b) Estimate the fraction of students who have a GPA of 3.5 or more.

[2.5+2.5=5]

[2]

- 5. (a) Describe the process of filtering on data streams. Explain Bloom Filter. [3]
 - (b)Write and explain Flajolet-Martin algorithm?
- 6. Describe the architecture of HDFS in detail with the help of a block diagram. [5]

END