

R15

Code No: 127JU

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, November/December - 2018

BIG DATA ANALYTICS

(Common to CSE, IT)

Time: 3 Hours

Max. Marks: 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART- A

(25 Marks)

- 1.a) Discuss some preferred tools used for effective data management strategy and decision making? [2]
- b) What internal procedures impact on when and how data are delivered? [3]
- c) Draw and discuss the architecture of Hive in brief. [2]
- d) What is Spark Unified Stack? Explain using an example. [3]
- e) Why there a sudden requirement for Big Data Analytics today? Explain. [2]
- f) What are the different ways of constructing version stamps? Explain their advantages and disadvantages. [3]
- g) State the relationship between the regression coefficient and correlation coefficient. [2]
- h) State the formulae for the constant term and coefficient in the regression equation. [3]
- i) Draw two example parallel coordinate plots: The first showing positive correlation and the second showing negative correlation. [2]
- j) How to prepare the data for visualization using Tableau tool? Explain. [3]

PART-B

(50 Marks)

2. How does Rackspace implement controls for the physical security and environmental resilience of the underlying hardware? Explain with a suitable example. [10]

OR

3. Describe issues regarding preprocessing of data for classification and prediction. Also explain the process of communication between data providers and data consumers. [10]

4. Explain Map-reduce framework in detail. Draw the architectural diagram for Physical Organization of Compute Nodes. [10]

OR

5. What is Spark? State the advantages of using Apache Spark over Hadoop MapReduce for big data processing with example. [10]

AG AG AG AG AG AG AG A

6. Explain the points to Consider during Data Analysis in detail. What are steps to follow while developing an Analytic Team? Explain. [10]

OR

AG 7. How is Report comparisons and analysis done by a developing team? Explain different types of Analytics. [10] AG A

8. How is a random sample selected? Explain complex random sampling designs. [10]

OR

9. With the help of a block diagram explain the probably approximately correct (PAC) Learning model. [10]

AG 10. Define the concept of a stream surface. Describe a simple method for the generation of a stream surface represented as a triangle mesh. [10] AG A

OR

11. What are the three most important mappings for visualization of a 2D scalar field $d(x,y)$? Give for each mapping a mapping function which defines the relation between the data values and the visual parameters. [10]

AG AG AG AG ~~AG~~ AG AG AG A

AG AG AG AG AG AG AG A

AG AG AG AG AG AG AG A

AG AG AG AG AG AG AG A

AG AG AG AG AG AG AG A