AUTUMN MID SEMESTER EXAMINATION-2023

School of Computer Engineering
Kalinga Institute of Industrial Technology, Deemed to be University
Big Data
[CS-3032]

Time: 1 1/2 Hours **Full Mark: 40**

Answer Any four Questions including Question No. 1 which is compulsory.
The figures in the margin indicate full marks. Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. Answer all the questions. [2 x 5 Marks]

   a) Describe 4 V's characterizing Big Data.
   b) Differentiate between Data-at-Rest and Data-in-Motion.
   c) Provide two advantages of application virtualization?
   d) Calculate the probability of False Positives with table size 10 and number of items to be inserted are 3.
   e) Why Name Node is called the single point of failure?

[10 Marks]

2. A mobile health organization captures patient's physical activities, by attaching various sensors on different body parts. These sensors measure the motion of diverse body parts like acceleration, the rate of turn, magnetic field orientation, etc. You have to build a system for effectively deriving information about the motion of different body parts like chest, ankle, etc. For the above problem statement, identify the tasks/activities to be performed in each stage of the big data analytics life cycle.

[10 Marks]

3. Explain the functions of Ingestion Layer, Platform Management Layer and Storage Layer of Big Data Stack with neat diagram.

[10 Marks]

4. Consider a Bloom Filter of size 11, with integers as stream elements and two hash functions as follows:

   - H1(x) = take odd number of bits from right in the binary representation of 'x'. Subsequently, treat it as an integer i, and result is i modulo 11.
   - H2(x) = same, but take even numbered bits.

   (a) Find the filter after the insertion of elements 25, 15 and 30.
   (b) Check whether the element y=18 exists in the bloom filter or not. Is it the case of False Positive or False Negative? Explain.

[7 +3=10 Marks]

5. (a) Describe the roles of various components of HDFS. How fault tolerance and continuous availability is ensured using HDFS.

   (b) Why we require more memory for node storing the Name Node.

*** Best of Luck ***