Mid Semester Examination

Max. Marks: 30

Duration: 1Hr.

Class: BE

Semester: VII

Course Code -   CS414

Com/IT

Subject: Big Data Analytics and Information Retrieval

Instructions: (1) All questions are compulsory.

(2) Use of a scientific calculator is allowed.

(3) Assume any necessary data but justify the same.

Q. No	Questions	Max Marks	CO
Q.1	Why Facebook, Twitter, Youtube data falls under Big data explain big data characteristics with example ?	8	CO 1
Q.2	Describe the key phases of a MapReduce job and how data is processed in a distributed computing environment. Provide an example scenario where MapReduce can be beneficial for processing large datasets. Explain the MapReduce programming model and how it is utilized in the context of Hadoop.	8	CO 1

1

Q.3	Find the Bloom Filter value for the given elements, 32, 45, 57, 87,113,114,27,33,65,91,99,44,78,71,61,63. hash Functions are as follows size of the bloom filter is 16. H1(x)= x mod 16 H2(x)=(2x +3) mod 16	7	CO4
Q.4	Explain the Flajolet-Martin algorithm for estimating the cardinality of a set(i.e Total Distinct elements). Describe the key steps involved in the algorithm and discuss any factors that can influence the accuracy of the cardinality estimate. Provide an example scenario where the Flajolet-Martin algorithm can be applied effectively.	7	CO4