End Semester Examination

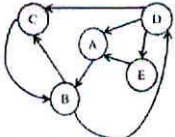|	
Max. Marks: 100	Duration: 3 Hr
Class: B. Tech Computer	Semester: VII
Course Code: CS414	Date: 13/12/2023
Subject: Big Data Analytics and IR	Time: 2.00 p.m-5 pm

Instructions: (1) All questions are compulsory.   *Comp, IT, ETRY EXTC*
(2) Use of scientific calculator is allowed.
(3) Assume any necessary data but justify the same.

Q. No.	Questions	Max Mark	CO	BL
Q.1a	Explain different characteristics of Big Data with example? Explain the challenges of big data and its applications with example?	10	CO 1	L2
Q.1b	What is Hadoop Architecture? Explain Hadoop component with example.	10	CO 2	L2
Q.2a	Differentiate Map- reduce architecture with Spark along with example?	10	CO 2	L4
Q.2b	Wordlist= ['apple', 'banana', 'orange', 'apple', 'grape', 'orange', 'banana', 'apple','orange','grape', 'banana', 'apple', 'grape', 'orange', 'apple', 'grape', 'orange', 'apple','orange','grape'] How you would perform a word count using the MapReduce paradigm. Describe the key steps involved in implementing the MapReduce process to calculate the frequency of each word in the provided list. Additionally, outline the functions and roles of the Mapper and Reducer tasks in the MapReduce framework, illustrating how they collaborate to accomplish the word count task efficiently	10	CO 2	L3
Q.3a	Explain the advantages and limitations of utilizing Spark SQL for processing and analyzing large-scale datasets compared to traditional SQL querying on relational databases? **OR** Explain various components of Spark Projects?	10	CO 3	L3
Q.3b	If the current state of the stream is 101111101110100111100101011100111 create buckets for the given stream and find the final state of the stream if 101100101 bits added in sequence one by one into stream consider bit by bit and show each bit addition separately using DGIM	10	CO 4	L3

	Algorithm ?			
Q.4a	Use N = 11, We have set Bloom Filter to 00000000000. use h1(x) = Take odd number of bits from right of the binary representation of x. h2(x)= Take even number of bits from right of the binary representation of x. The stream elements(X) are 28,358,189,119,196,484,76. What is the final value of the bloom filter?	10	CO 4	L3
Q.4b	Discuss Spark SQL's architectural components, optimizations, and query execution mechanisms that enable enhanced performance and scalability in big data processing scenarios.	10	CO 3	L3
Q.5a	What is PageRank? Define the structure of the web and how top sensitive page ranks are given? Calculate page ranks of all pages given in below graph.	10	CO 5	L3
Q.5b	Discuss in detail the implications of utilizing a property graph database model like Neo4j in a social network setting compared to a traditional relational database. **OR** How Item -Item collaborative filtering and user -user collaborative filtering similarity carried out explain with the help of example of 5 movies and 8 users. Use 5-star ratings?	10	CO 5	L3