KUMARAGURU
college of technology
character is life

B.E DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Fifth Semester

INFORMATION SCIENCE AND ENGINEERING

U18ISI5202**:** Big Data Technologies

COURSE OUTCOMES

CO1: Understand the importance of big data processing and its application

CO2: Gain an understanding of what insights big data can provide by modelling, integrating, and processing big data

CO3: Understand the core tools used to wrangle and analyze big data

CO4: Apply appropriate techniques to solve fundamental problems in Big Data

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions: -
PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1.	State big data era and who benefits from it?	CO1	[K_1]
2.	Recall the 3 types of big data.	CO1	[K_1]
3.	Write the key differences between CSV and JSON.	CO2	[K_2]
4.	List the use of constraints in data models.	CO2	[K_2]
5.	Mention the components of big data pipeline.	CO2	[K_2]
6.	Write the role of Apache Spark in big data.	CO3	[K_2]
7.	Define Hadoop.	CO3	[K_1]
8.	Name the tools to process real-time and high-velocity data.	CO3	[K_2]
9.	Outline the HDFS read and write process.	CO4	[K_1]
10.	Identify the advantages of MapReduce over spark.	CO4	[K_2]

Answer any FIVE Questions: -
PART B (5 x 4 = 20 Marks)
(Answer not more than 80 words)

11.	Why the traditional database is approaches not suitable for big data? Justify the statement.	CO1	[K_2]
12.	Identify the data lake roles and responsibilities in big data.	CO2	[K_3]
13.	Explain the big data integration and its challenges.	CO2	[K_2]
14.	Outline the Apache Tez.	CO3	[K_5]

15.	Distinguish between Hive and Pig.		CO4 [K₄]
16.	How to find the twitter trends using Map Reduce? Analyze.		CO4 [K₄]

Answer any FIVE Questions: -
PART C (5 x 12 = 60 Marks)
(Answer not more than 300 words)

17. a) Explain the following big data terms with example. 12 CO1 [K₂]
     i)   Volume
     ii)   Velocity
     iii)   Variety
     iv)   Veracity

18. a) Compare and contrast structured, semi-structured and unstructured data models 12 CO2 [K₃]
with example.

19. a) Outline the aggregation operations in the big data. 6 CO3 [K₂]

    b) Illustrate the use of Spark in big data. 6 CO3 [K₂]

20. a) Elaborate the main components of HDFS with neat sketch. 12 CO4 [K₃]

21. a) Describe the different types of operators in Pig with examples. 6 CO4 [K₂]

    b) Examine the various data types supported by HIVEQL with an example. 6 CO4 [K₃]

22. a) For the following data, list the input to output from both the map and reduce 12 CO4 [K₄]
functions for getting word count of each word.

Deer, Bear, River, Car, Car, River, Deer, Car, and Bear.

****