Roll No. | | | | | | | | | |

(177) SET-I

COER University

END SEMESTER EXAMINATION, EVEN SEMESTER, 2023-24

Time : 3 hour | Semester : VI | Total Marks : 100

Program Name : B.Tech | Branch/Specialization : CSE,AI&ML

Course Code : SOC306 | Course Name : Big Data Analytics

Note: All questions are compulsory. No student is allowed to leave the examination hall before the completion of the time.

Q. No 1	Attempt Any Four Parts. Each Question Carries 5 Marks.	CO	BL
(a)	Recall challenges faced by Big Data.	CO 1	2
(b)	Explain the significance of Big Data Communications, Media and Entertainment	CO 1	2
(c)	How does Structure, Un-Structured and Semi-Structured data affect the analytic processes?	CO 1	2
(d)	Define Big Data and list its key characteristics	CO 1	2
(e)	Describe Data Warehouse Architecture in detail.	CO 1	2

Q. No 2	Attempt Any Four Parts. Each Question Carries 5 Marks.	CO	BL
(a)	Explain Data Stream Management System Architecture	CO 2	2
(b)	Describe different Sources of Data Stream.	CO 2	2
(c)	Analyze the impact of Lambda Architecture and Kappa Architecture in a data streaming pattern.	CO 2	3
(d)	Evaluate the effectiveness of different Sampling Techniques for Efficient Stream Processing.	CO 2	3
(e)	Can you differentiate between Batch processing and real-time stream processing	CO 2	2

Q. No 3	Attempt Any Four Parts. Each Question Carries 5 Marks.	CO	BL
(a)	Describe the benefits of Hadoop environment.	CO 3	2
(b)	Briefly explain High Level Architecture of HDFS .	CO 3	2
(c)	Implement HDFS to store 3674 MB file and manage dataset in blocks when default block size is 128 MB. Find total replication factor.	CO 3	3
(d)	Evaluate the efficiency Hadoop – Schedulers and explain types of Schedulers.	CO 3	3
(e)	Configure ten commands to invoke HDFS Command-Line Interface.	CO 3	2

Q. No 4	Attempt Any Two Parts. Each Question Carries 10 Marks.	CO	BL
(a)	Explain the services provided by Hive in the context of Big Data analytics. How does Hive facilitate querying and analyzing large datasets stored in Hadoop Distributed File System (HDFS)?	CO 4	3
(b)	Describe Data types and Construction Operators used in Pig Latin.	CO 4	2
(c)	Discuss Storage mechanism in HBase also create difference between HBase and RDBMS.	CO 4	2

Q. No 5	Attempt Any Two Parts. Each Question Carries 10 Marks.	CO	BL
(a)	Discuss Azure Blob data life cycle management tool.	CO 5	2
(b)	Discuss the purpose and benefits of using Hadoop Archives (HAR) for archiving files in HDFS. Explain how HAR files are created and accessed, and provide examples of scenarios where HAR files are useful.	CO 5	3
(c)	Implement Map-Reduce on given data set:(Do not consider "," as a dataset) 1110, 0110, 1111, 0001 1111, 1010, 0110, 1110 0001, 0110, 0001, 1010	CO 5	4

----------End of Paper----------