

DR. BABASAHEB AMBEDKAR TECHNOLOGICAL UNIVERSITY, LONERE Regular and Supplementary Summer Examination – 2024 Course: B. Tech. Semester :VI Branch : Artificial Intelligence and Data Science Engineering And Allied Subject Code & Name: Big Data Analytics (BTAIOE604A) Max Marks: 60 Date: 21/06/2024 Duration: 3 Hr.			
Instructions to the Students: 1. All the questions are compulsory. 2. The level of question/expected answer as per OBE or the Course Outcome (CO) on which the question is based is mentioned in front of the question. 3. Use of non-programmable scientific calculators is allowed. 4. Assume suitable data wherever necessary and mention it clearly.			
		(Level/CO)	Marks
Q. 1	Solve Any Two of the following.		12
A)	Enlist various types of data. Explain with suitable example different types of data.	CO1	6
B)	What is Hadoop? What are the four main Components of Hadoop?	CO1	6
C)	Describe the key components of IBM's big data platform and their respective roles in data processing and analysis.	CO1	6
Q.2	Solve Any Two of the following.		12
A)	Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief.	CO2	6
B)	What are the different types of big data analytics?	CO2	6
C)	What is data serialization? Make a note on how type of data affects data serialization.	CO2	6
Q. 3	Solve Any Two of the following.		12
A)	What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram.	CO3	6
B)	What is the difference between an RDBMS and Hadoop?	CO3	6
C)	Discuss Hadoop YARN in detail with failures in classic Map Reduce.	CO3	6
Q.4	Solve Any Two of the following.		12
A)	Explain working of Hive with proper steps and diagram.	CO4	6
B)	What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example.	CO4	6
C)	Write a short note on Apache Pig. Enlist applications of Apache Pig.	CO4	6
Q. 5	Solve Any Two of the following.		12
A)	What are the major components of Kafka? Explain with neat diagram.	CO5	6
B)	What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset?	CO5	6
C)	What is the difference between Apache Kafka and Apache Spark?	CO5	6
*** End ***			