

Name:
Enrolment No:



End Semester Examination, December 2023

Program Name: BTech Big Data
Course Name: Big Data Ingestion
Course Code: CSBD2006

Semester : 3rd
Time : 3 hrs
Max. Marks : 100

Section A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	What is standardization? Explain the need for standardization in big data.	4	CO1
Q 2	Discuss the features of Hadoop that make it suitable for big data applications.	4	CO2
Q 3	In the healthcare industry, what are the different data sources based on their categories? Explain the need of big data analysis for the same.	4	CO1
Q 4	Define microservice. Explain with an example.	4	CO4
Q 5	Explain the significance of using --fields-terminated-by clause in Apache Sqoop with examples.	4	CO3

Section B
(4Qx10M=40Marks)

Q 6	How does HDFS maintain data integrity despite utilizing commodity hardware, which inherently poses a higher risk of failure? Describe the methods through which HDFS guarantees fault-tolerance within the system. Or Describe the scalability and high availability feature of HDFS. How does a name node determines if a data node is not operational?	10	CO1
Q 7	Illustrate the structure of Map Reduce, highlighting its essential components and their respective functions. Or What is YARN, and how does it improve upon the limitations of the previous Hadoop MapReduce framework? Can you explain the main components of the YARN architecture and their functions?	10	CO2
Q 8	In message queuing system, illustrate the difference between the push model and the pull model. Give a scenario in which the push model is more suitable than the pull model.	10	CO4

Q 9	Define big data ingestion. Explain the difference between batch ingestion and real time ingestion.	10	CO3
Section C (2Qx20M=40Marks)			
Q 10	<p>Using a block diagram, discuss the following components of Kafka architecture:</p> <ul style="list-style-type: none"> i. topics ii. producers iii. consumers iv. brokers <p>Or</p> <p>Explain the working of Kafka producer and Kafka consumer in detail. Also, discuss the use of Kafka in real-time data streaming and processing.</p>	20	CO4
Q 11	<ul style="list-style-type: none"> i. Mention the features of Apache Sqoop. Also, explain the process to perform an incremental data load in Sqoop? ii. Create a table in MySQL. Use that table to illustrate: <ul style="list-style-type: none"> a. Sqoop import of data from particular rows. b. Sqoop import of data in Avro format. c. Sqoop import using a free form query 	20	CO3