Reg. No.															

B.Tech. DEGREE EXAMINATION, JUNE 2023
Sixth Semester

18CSE391T – BIG DATA TOOLS AND TECHNIQUES
(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

Note:
(i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.
(ii) **Part - B & Part - C** should be answered in answer booklet.

Time: 3 hours | Max. Marks: 100

PART – A (20 × 1 = 20 Marks)
Answer **ALL** Questions

	Marks	BL	CO	PO
1. Yarn stands for _____. (A) Yet another resource network (B) Yet another resource negotiator (C) Yet another research node (D) Yet archive resource negotiator	1	1	3	3
2. What would be the number of blocks for storage with HDFS block size 128 MB and disk size 10 GB? (A) 30 (B) 80 (C) 40 (D) 120	1	3	3	3
3. Huge storage and computational power requirement for big data is supported by _____. (A) PaaS (B) IaaS (C) SaaS (D) BDaaS	1	3	1	2
4. _____ is the storage layer of hadoop. (A) Yarn (B) Map reduce (C) HDFS (D) Hive	1	2	2	2
5. HDFS produces a command line interface called _____ to interact with HDFS. (A) HDFS shell (B) FS shell (C) DFS shell (D) Hadoop shell	1	1	3	3
6. _____ supports split table compression. (A) LZO (B) Bzip 2 (C) Gzip (D) ZLO	1	2	2	2
7. The output of the mapper and reducer is in _____ format. (A) Document (B) Key, value (C) Table (D) Tree	1	3	3	3
8. Apache _____ is a serialization framework that produce data in binary format. (A) Kafka (B) Oozie (C) Avro (D) Impala	1	2	3	2

9.	Zookeepers architecture supports high availability through _____ services. (A) Redundant (B) Data (C) Log (D) Access	1	2	4	3
10.	_____ supports data flow from source into hadoop environment. (A) Impala (B) Oozie (C) Flume (D) SQL	1	2	4	5
11.	Hbase and Cassandra are _____ databases. (A) Row oriented (B) Column oriented (C) Document (D) XML	1	1	5	5
12.	_____ pig command is used to display results on the screen. (A) Group (B) Dump (C) Load (D) List	1	3	4	3
13.	Memory computation for interactive processing in spark is supported by _____. (A) Distributed queues (B) Distributed trees (C) Distributed clusters (D) Resilient distribute databases	1	2	5	3
14.	Cassandra uses _____ protocol to discover location and state information. (A) Gossip (B) Goss (C) Inter route (D) Route	1	2	5	3
15.	_____ is a java web application to schedule apache hadoop jobs. (A) Oozie (B) Impala (C) Mahout (D) Hive	1	1	4	2
16.	_____ is a non relational database modeled after Google's big table. (A) Spark (B) Hbase (C) MongoDB (D) Cassandra	1	2	5	5
17.	_____ learning uses labelled training data (A) Supervised (B) Reinforcement (C) Clustering (D) Unsupervised	1	2	6	4
18.	_____ describes a pie chart. (A) Comparison in different categories of data (B) Data trend over time (C) Correlation between different data set (D) Contribution of individual values to a total value	1	4	6	5
19.	_____ supports data visualization in python. (A) Data frame (B) Pandas (C) Numpy (D) Matplotlib	1	2	5	5
20.	Big data size is represented in _____. (A) Giga bytes (B) Mega bytes and Giga bytes (C) Tera bytes (D) Peta bytes and Exa bytes	1	2	2	2

PART – B (5 × 4 = 20 Marks)

Answer **ANY FIVE** Questions

		Marks	BL	CO	PO
21.	Why YARN is introduced in hadoop version 2?	4	4	3	3
22.	Demonstrate data storage process in HDFS.	4	3	3	3
23.	Illustrate space flume architecture.	4	3	4	2
24.	Write the query to create a table in Hive.	4	3	4	3
25.	State the flink methods to transform elements within dataset.	4	1	5	3
26.	Mention the significance of tableau charts for big data visualization.	4	1	5	5
27.	Write the code snippet in python for data visualization.	4	3	6	5

PART – C (5 × 12 = 60 Marks)

Answer **ALL** Questions

		Marks	BL	CO	PO
28. a.	Illustrate the application of cloud computing in a big data environment.	12	3	1	2
	(OR)				
b.	Demonstrate the architecture of YARN and the workflow of job execution in YARN.	12	3	3	2
29. a.	Explain file read and store operations in HDFS with suitable diagrams.	12	1	3	2
	(OR)				
b.	Implement map reduce concept with an example.	12	3	3	2
30. a.	Write Hive queries using		3	4	2
	(i) Limit clause	3			
	(ii) Where conditions	3			
	(iii) Group by column names	3			
	(iv) Order by column names	3			
	(OR)				
b.	Give short notes on		2	4	2
	(i) Pig architecture	4			
	(ii) Pig components	4			
	(iii) Pig data model	4			
31. a.	Discuss the significance of various storage levels of RDD in spark.	12	5	5	2
	(OR)				
b.	Elaborate the role of unstructured databases for big data analytics using MongoDB.	12	2	5	3
32. a.	Explain the different types of tableau charts for big data visualization.	12	2	6	5
	(OR)				
b.	Discuss in detail about the data science solutions in enterprise data science.	12	2	6	3

* * * * *