

B.Tech/ M.Tech (Integrated) DEGREE EXAMINATION, MAY 2024

Fourth Semester

21CSE222T – BIG DATA TOOLS AND TECHNIQUES*(For the candidates admitted from the academic year 2022-2023 onwards)***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** and **Part - C** should be answered in answer booklet.

Time: 3 Hours

Max. Marks: 75

PART – A (20 × 1 = 20Marks)

Marks BL CO PO

Answer ALL Questions

- Which of the following does not come under big data technologies?
(A) Apache Pytarch (B) Apache Hadoop
(C) Apache Spark (D) Apache Kafka
1 1 1 2
- Hadoop is written in _____ language.
(A) Lua (B) Java
(C) Perl (D) Python
1 1 1 2
- Identify the technique which includes the features such as scalability, elasticity, resource pooling, self-service, low cost and fault tolerance.
(A) Power BI (B) Parallel computing
(C) Cloud computing (D) System development
1 2 1 3
- Which of the following is not the characteristic of big data?
(A) Volume (B) Validity
(C) Variety (D) Velocity
1 1 1 2
- During start up, the _____ loads the file system from the FSimage and edit log files.
(A) Data node (B) Name node
(C) Block (D) Action node
1 2 2 2
- YARN command are invoked by _____ scripts.
(A) Hive (B) Bin
(C) Hadoop (D) Home
1 2 2 3
- HDFS workflow related to _____ fashion.
(A) Worker-master fashion (B) Master-slave fashion
(C) Master-worker fashion (D) Slave-master fashion
1 2 2 2
- Which programming model is commonly used in distributed computing frameworks in big data?
(A) Map Reduce (B) Object-oriented programming
(C) Functional programming (D) Procedural programming
1 1 2 2

9. Apache hive is data storage and _____ that stores and organizes data for study. 1 1 3 2
 (A) Spark (B) Querying tool
 (C) Map Reduce (D) YARN
10. Pig operates in mainly on how many nodes? 1 1 3 4
 (A) Two (B) Three
 (C) Four (D) Five
11. In zookeeper, what is "Znode"? 1 2 3 3
 (A) A type of data storage (B) A data processing job
 (C) A hierarchical data node in the zookeeper namespace (D) A unit of data replication
12. Which of the following is not a common source type in apache flume? 1 1 3 3
 (A) Apache Kafka (B) Apache HBase
 (C) Log files (D) Twitter data
13. Which spark component is responsible for scheduling tasks, managing cluster resources and ensuring fault tolerance. 1 1 4 3
 (A) Spark SQL (B) Spark streaming
 (C) Spark MLlib (D) Spark core
14. User accounts may be altered and dropped using the _____ query language. 1 2 4 3
 (A) Hive (B) Cassandra
 (C) Sqoop (D) Flume
15. Which of the following is true about apache Flink? 1 2 4 4
 (A) Apache Flink is a real time processing framework (B) Apache Flink is a stream-processing and batch-processing framework
 (C) Both (A) and (B) (D) Apache flink is a static processing framework
16. Which of the following is a key challenge in enterprise data science? 1 1 4 4
 (A) Lack of data availability (B) Overabundance of data
 (C) Limited computing power (D) Irrelevant data
17. Tableau "Worksheet" is used for _____. 1 1 5 5
 (A) Organizing data sources (B) Creating calculated fields
 (C) Displaying a single visualization or chart (D) Importing data from external databases
18. Machine learning algorithms build a model based on sample data, known as 1 2 5 5
 (A) Training data (B) Transfer data
 (C) Data training (D) Model data
19. Which of the following best describes MongoDB's data model? 1 1 5 5
 (A) Relational (B) Document-oriented
 (C) Key value (D) Graph-based

- | | | | | |
|--|---|---|---|---|
| 20. Which technology is commonly used in enterprise data science for handling and analyzing large dataset? | 1 | 4 | 5 | 5 |
| (A) Virtual reality | | | | |
| (B) Augmented reality | | | | |
| (C) Big data platforms | | | | |
| (D) Blockchain | | | | |

PART – B (5 × 8 = 40 Marks)

Marks BL CO PO

Answer ALL Questions

- | | | | | |
|---|---|---|---|---|
| 21. a.i. Describe any four V's of Big Data. | 5 | 2 | 1 | 3 |
| ii. Write a note on Big Data Analytics. | 3 | 2 | 1 | 3 |
| (OR) | | | | |
| b. Correlate cloud computing and Big data with proper real-time scenario. | 8 | 3 | 1 | 2 |
| 22. a. List out the core components of Hadoop and explain them. | 8 | 2 | 2 | 2 |
| (OR) | | | | |
| b. Demonstrate how job scheduling done by YARN. | 8 | 3 | 2 | 3 |
| 23. a. Explain Apache pig, its key features and how it simplifies data processing tasks in the context of Hadoop. | 8 | 2 | 3 | 2 |
| (OR) | | | | |
| b. Write any five Hive-QL commands with relevant examples. | 8 | 3 | 3 | 4 |
| 24. a. Elaborate the architecture of apache spark. | 8 | 2 | 4 | 3 |
| (OR) | | | | |
| b. Describe the ecosystem of apache flink with a neat diagram. | 8 | 3 | 4 | 5 |
| 25. a. Explain the importance of data visualization in big data. | 8 | 3 | 5 | 4 |
| (OR) | | | | |
| b. Explore the role of enterprise data science in big data analytics. | 8 | 2 | 5 | 4 |

PART – C (1 × 15 = 15 Marks)

Marks BL CO PO

Answer ANY ONE Question

- | | | | | |
|--|----|---|---|---|
| 26. In a large multinational corporation, vast amount of data from various sources including customer transactions, website interactions and sensor readings from manufacturing equipment are been collected. Imagine that you are an employee in that MNC.
Now, how would you explain the importance of Hadoop Distributed File System (HDFS) in addressing the data storage and processing challenges faced by the company? | 15 | 4 | 2 | 5 |
| 27. Discuss the technical elements of Big Data Platforms. | 15 | 3 | 1 | 5 |

* * * * *

