

- b. State about operations of sqoop with its' proper commands. 10 3 3 3,5
29. a. Express the limitation of hadoop and how those limitations are resolved by spark. 10 4 4 3,5
- (OR)**
- b. Justify the reasons for using NOSQL and state about few NOSQL data bases. 10 4 4 3,5
30. a. Expose the use of tableau in big data visualization. 10 3 5 3,5
- (OR)**
- b. Analyse the case study of hadoop in Facebook to solve the data warehousing problem. 10 3 5 3,5

* * * * *

Reg. No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

B.Tech. DEGREE EXAMINATION, MAY 2022

Sixth Semester

18CSE391T – BIG DATA TOOLS AND TECHNIQUES

(For the candidates admitted from the academic year 2018-2019 to 2019-2020)

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** should be answered in answer booklet.

Time: 2½ Hours

Max. Marks: 75

PART – A (25 × 1 = 25 Marks)

Answer **ALL** Questions

- | | Marks | BL | CO | PO |
|--|-------|----|----|-----|
| 1. Identify the platform for constructing data flows for extract, transform and load processing and analysis of large datasets.
(A) Pig (B) Flume
(C) NoSQL (D) Hive | 1 | 2 | 1 | 1,5 |
| 2. Spot the language custom extension supported by hive
(A) PHP (B) Java
(C) Swift (D) Objective C | 1 | 1 | 1 | 1,5 |
| 3. The pig Latin scripting language has operators like _____ apart from higher level data flow language.
(A) JSON (B) Java script
(C) Node JS (D) SQL | 1 | 2 | 1 | 1,5 |
| 4. Hadoop was developed by
(A) Doug cutting (B) Tim berner
(C) Alan Turing (D) Fred brooks | 1 | 1 | 1 | 1 |
| 5. Select the correct statement from the given choices about hadoop.
(A) Hadoop is an ideal environment for extracting and transforming medium volume of data
(B) Hadoop stores data in HDFS and supports data compression / decompression
(C) The graph frameworks is less useful than a map reduce job to solve graph and machine learning
(D) Hadoop has no limitations | 1 | 2 | 1 | 1,5 |
| 6. Which is performing as a slave and responsible for executing a task assigned to it by the name node?
(A) Mapper (B) Reducer
(C) Job tracker (D) Task tracker | 1 | 2 | 2 | 1,5 |
| 7. Which tools allow users to create and run jobs with any executables as the mapper and/or the reducer?
(A) Hadoop streaming (B) Hadoop shuffling
(C) Hadoop sorting (D) Hadoop filter | 1 | 2 | 2 | 1,2 |

8. Choose the pair of phases occur simultaneously in MAPRED
(A) Reduce and sort (B) Shuffle and sort
(C) Map and reduce (D) Filter and map
9. _____ parameter describes destination directory which could hold the archive?
(A) <root> (B) <destinations>
(C) <output> (D) _archivename <name>
10. Pick the CODEC from Google which provides modest compression ratios
(A) LZ0 (B) Bzip2
(C) Snappy (D) File compress
11. _____ holds the site specific configuration details for a hadoop installation.
(A) core-config.xml (B) core-default.xml
(C) core-hadoop.xml (D) core-site.xml
12. To check deprecation of a key, which method is used in hadoop?
(A) ifDeprecated (B) isDeprecated
(C) setDeprecated (D) Deprecatedif
13. _____ is the center of hadoop architecture for multiple data processing engines.
(A) Yarn (B) Sqoop
(C) Zookeeper (D) Flink
14. To run batch mode in pig _____ are being used.
(A) Pig commands (B) Pic config
(C) Pig scripts (D) Pig options
15. Name the shell utility which can be used to run hive queries in either interactive or batch mode.
(A) \$HIVE/ LIB / HIVE (B) \$HIVE_HOME/ BIN / HIVE
(C) \$HIVE/ BIN / HIVE (D) \$HIVE/ S BIN / HIVE
16. Find the java web application used to schedule apache hadoop jobs?
(A) Zoo keeper (B) Apache workflow
(C) Oozie (D) Mahout
17. What is the top component on spark core?
(A) Spark SQL (B) RDDs
(C) Spark ML (D) Spark streaming
18. Apache cassandra is a massively scalable open source _____ database.
(A) NOSQL (B) SQL
(C) CLOUD (D) OOSQL
19. Guess the key which is used by mongoDB instead of primary key
(A) Mongo (B) Default Key_ID
(C) Record ID (D) Column ID

20. H base is what kind of distributed database built on top of the hadoop file system?
(A) Row oriented (B) Object oriented
(C) Column oriented (D) Record oriented
21. Which is not a challenge associated with big data analytics implementation in an enterprise?
(A) Nearly every company is investing in big data, ML and AI (B) Often the company has a corporate mandate
(C) Finding the right use cases (D) Finding resources for big data implementation
22. _____ is oracle's high performance engineered database platform that is designed for resource intensive queries.
(A) Oracle BI (B) Exadata
(C) Oracle ML (D) Tableau
23. _____ from HP is a column oriented massively parallel processing database system.
(A) Vertex (B) Verizon
(C) Vector (D) Vertica
24. Spot the cloud based data warehouse platform from amazon.
(A) Red data (B) Red marker
(C) Red shift (D) Fire base
25. Microsoft's _____ visualization tools enables you to connect with hundreds of data sources, and then publish reports on the web and mobile devices.
(A) Visual analyser (B) QLIK viewer
(C) Lumira (D) Power BI

PART – B (5 × 10 = 50 Marks)

Answer ALL Questions

Marks BL CO PO

26. a. Illustrate about the relationship between cloud computing and big data. 10 3 1 1,5
- (OR)
- b. Describe about the fundamentals of Hadoop components such as core modules, file system and data storage. 10 3 1 1,5
27. a. Analyse about scaling out, data flow and combiner functions in map reduce operation. 10 4 2 3,5
- (OR)
- b. Examine about hadoop streaming process in map reduce. 10 3 2 1,5
28. a. Illustrate about zookeeper architecture, data model and operations. 10 3 3 3,5

(OR)