Q. No.	Question	Marks	BL	CO	PO
27. a.i.	Deduce the syntax for count ( ), arg( ), sum ( ) in SQL and mention its purpose.	6	2	2	2
ii.	Deduce the use of head ( ), tail (), info ( ), describe ( ) in python.	4	2	2	2
	(OR)				
b.	Articulate the various data types involved in python language with code snippets.	10	2	2	2
28. a.i.	Sketch the HDFS architecture.	5	2	3	1
ii.	Deduce the difference between Hadoop 1.X and Hadoop 2.X.	5	2	3	1
	(OR)				
b.	Sketch the architecture of apache hive and articulate the working.	10	2	3	1
29. a.	Illustrate the job run process in map reduce section of Hadoop data analytics.	10	2	4	1
	(OR)				
b.	Deduce the difference between YARN and map reduce processes in Hadoop.	10	2	4	1
30. a.i.	Estimate the challenges involved in security analytics.	5	2	5	1
ii.	Deduce the challenges involved in intrusion of bigdata.	5	2	5	1
	(OR)				
b.	Articulate in detail the techniques of data analysis.	10	2	5	1

* * * * *

Reg. No.															

B.Tech. DEGREE EXAMINATION, NOVEMBER 2022
Sixth and Seventh Semester

18EEE425T – FUNDAMENTALS OF BIG DATA ANALYTICS
(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

Question	Marks	BL	CO	PO
1. Define the data that does not conform to a data model or data schema (A) Structured (B) Unstructured (C) Defined (D) Semi-structured	1	1	1	1
2. Find the range of $\{10, 20, 120, 40, 440\}$ (A) 340 (B) 430 (C) 320 (D) 330	1	1	1	2
3. Outline the use of big data analytics (A) Spread data (B) Analyse data (C) Organize data (D) Collect data	1	1	1	1
4. Find the total V's of big data (A) 3 (B) 4 (C) 5 (D) 6	1	1	1	1
5. Identify the measure of dispersion that is least affected by extreme values (A) Range (B) Mean deviation (C) Standard deviation (D) Quartile deviation	1	1	1	1
6. Display the output of the code Delete from students Where age =16; Roll back; (A) Performs an undo operation on the delete operation (B) Deletes the entire table (C) Deletes the rows (D) Deletes the complete details	1	2	2	2
7. Find the operation of the SQL statement DROP TABLE student; (A) Deletes a table called student (B) Creates a table called student (C) Regenerates the table called student (D) Checks the table called student	1	2	2	2

8. Choose the correct extension of python file 1 2 2 1
(A) .python (B) .pl
(C) .p (D) .py

9. Examine the way numbers in 'R' are treated 1 2 2 1
(A) Single precision real numbers (B) Double precision real numbers
(C) Real precision real numbers (D) Imaginary precision real numbers

10. Deduce the value of the expression (python): 1 2 2 2
4+5%5
(A) 7 (B) 8
(C) 9 (D) 10

11. Deduce the incorrect big data technology 1 2 3 1
(A) Apache kafka (B) Apache pytorch
(C) Apache Hadoop (D) Apache spark

12. Determine the number of nodes in apache PIG 1 2 3 1
(A) 2 (B) 3
(C) 4 (D) 5

13. Determine the minimum amount of data that can be read or written on disk 1 2 3 1
(A) Byte size (B) Block size
(C) Heap (D) Hill

14. Deduce the check point node in HDFS 1 2 3 1
(A) Secondary name node (B) Secondary data node
(C) Name node (D) Data node

15. Determine the class that does the job control in Hadoop 1 2 3 1
(A) Task class (B) Mapper class
(C) Job class (D) Reducer class

16. Articulate that by the total size of _____determines the number of maps. 1 2 4 1
(A) Inputs (B) Outputs
(C) Attributes (D) Tasks

17. Deduce the fixed-size pieces of map reduce jobs 1 2 4 1
(A) Splits (B) Tasks
(C) Maps (D) Records

18. Determine how input key/value pairs to a set of intermediate key/values pairs. 1 2 4 1
(A) Mapper (B) Reducer
(C) Mapper and reducer (D) Tasker

19. Identify the part of map reduce that is responsible for processing one or more chunks of data for producing output 1 2 4 1
(A) Map task (B) Mapper
(C) Task evaluation (D) Map ( )

20. Examine the node that is responsible for executing a task and act as slave 1 2 4 1
(A) Map reduce (B) Mapper
(C) Task tracker (D) Job tracker

21. Articulate the way to intrude data 1 2 5 1
(A) Fall condition (B) Rise condition
(C) Sweep condition (D) Race condition

22. Deduce which is not the major component of intrusion detection 1 2 5 1
(A) Analysis engine (B) Event provider
(C) Alert database (D) Buffer execution

23. Deduce the characteristics of stack based intrusion detection 1 2 5 1
(A) They are integrated closely with the TCP/IP stack and watch packets (B) The host operating system logs in the audit information
(C) It is programmed to interpret a certain series of packets (D) It uses the normal network

24. Identify which is not the strength of host based intrusion detection 1 2 5 1
(A) Attack verification (B) System specific activity
(C) No additional hardware (D) Noise inclusion

25. Determine the way that does not classify intrusion detection 1 2 5 1
(A) Anamoly detection (B) Stack based
(C) Signature based (D) Segment based

PART – B (5 × 10 = 50 Marks) Marks BL CO PO
Answer **ALL** Questions

26. a.i. Find the range of the following frequency distribution of marks scored by class 10 students. 2 1 1 1

Marks interval	No of students
10-20	8
20-30	25
30-40	9

ii. Find the standard deviation for the following data 5 1 1 1
Data = {1,3,6,7,12}

iii. Comment on the characteristics of big data. 3 1 1 1

(OR)

b. Outline the data analytics process in detail. 10 1 1 1