b.	Summarize the architectures of YARN and map reduce using diagram.	12	1	4	1
32. a.	Comment on (i) IDS (ii) classifications	12	1	5	1
	(OR)				
b.	Comment on (i) various detection methods of IDS (ii) challenges in security analytics.	12	1	5	1

* * * * *

Reg. No.														

B.Tech. DEGREE EXAMINATION, JUNE 2023
Seventh Semester

18EEE425T – FUNDAMENTALS OF BIG DATA ANALYTICS
(For the candidates admitted from 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.	Select the data that does not conform to a data model or data schema. (A) Structured data (B) Unstructured data (C) Semi-structured data (D) Multi-structured data	1	1	1	1
2.	What is a symbolic representation of facts or concepts from which information can be obtained? (A) Data (B) Knowledge (C) Program (D) Algorithm	1	1	1	1
3.	Select the true statement (A) Big data can be processed using traditional techniques (B) Big data refers to data sets that are least a petabyte in size (C) Big data analysis does not involve reporting and data mining techniques (D) Big data has low velocity meaning	1	1	1	1
4.	How many total V's in big data? (A) 3 (B) 4 (C) 5 (D) 6	1	1	1	1
5.	Select the way, the R objects with attributes can be (A) Meta data (B) Features (C) Expression (D) Dimensions	1	1	2	1
6.	Find the correct extension of python file. (A) .python (B) .pl (C) .py (D) .p	1	1	2	1

7.	State the output of the following code
	i=1
	while true:
	if i%3= =0:
	break
	print(i+=1)
	(A) 123 (B) 1 2
	(C) Error (D) 3 2 1
		1	1	2	2

8. Show the output of the SQL code given below: 1 1 2 2
Drop Table Student:
(A) Deletes a table called student (B) Creates a table called student
(C) Forms a table called student (D) Destroys a table called student

9. Expand HDFS. 1 1 3 1
(A) Highly Distributed File System (B) Hadoop Directed File System
(C) Highly Distributed File System (D) Hadoop Distributed File System

10. What is the minimum amount of data that a disk can read or write in HDFS? 1 1 3 1
(A) Byte (B) Block
(C) Heap (D) Cluster

11. Identify the language in which hadoop is written 1 1 3 2
(A) C++ (B) Java
(C) Rust (D) Python

12. Which of the following is incorrect big data technology? 1 1 3 1
(A) Apache pytorch (B) Apache kafka
(C) Apache hadoop (D) Apache spark

13. Identify the name given to fixed-size pieces of mapreduce job 1 1 4 1
(A) Splits (B) Tasks
(C) Maps (D) Records

14. Which node acts as slave and is responsible for executing a task assigned to it by the job tracker? 1 1 4 1
(A) Map reduce (B) Mapper
(C) Trask tracker (D) Job tracker

15. Which function is responsible for consolidating the results produced by each of the MAP( ) functions/ tasks? 1 1 4 1
(A) Reduce (B) Map
(C) Reducer (D) Task manage

16. The total size of _____ drives the number of maps. 1 1 4 1
(A) Inputs (B) Outputs
(C) Tasks (D) Map reducers

17. Identify the method that is not used to classify IDS 1 1 5 1
(A) Anomaly detection (B) Signature based misuse
(C) Stack based (D) Signature detection

18. Which is not a method to intrude? 1 1 5 1
(A) Buffer overflow (B) Trade off condition
(C) Race condition (D) Unexpected combination

19. Identify the non-major components of intrusion detection system. 1 1 5 1
(A) Analysis engine (B) Event provider
(C) Alert database (D) Event intruder

20. Predict the characteristics of anomaly based IDS from the following: 1 1 5 1
(A) It models the normal usage of network as a noise characterization (B) It doesn't detect novel attacks
(C) Anything other than noise is not assumed to be intrusion (D) It detects based on signature

PART – B (5 × 4 = 20 Marks) Marks BL CO PO
Answer **ANY FIVE** Questions

21. Articulate on the characteristics of bigdata. 4 1 1 1

22. Discuss on the measures of central tendency and its types. 4 1 1 1

23. State the datatypes of SQL. 4 1 2 1

24. Comment on the usage of head ( ), tail ( ) and sample ( ) in python. 4 1 2 1

25. Comment on the differences between Apache pig and Mapreduce. 4 1 3 1

26. Exemplify the workflow of job run in mapreduce process through diagram. 4 1 4 1

27. Discuss on the benefits of data analytics in security aspects. 4 1 5 1

PART – C (5 × 12 = 60 Marks) Marks BL CO PO
Answer **ALL** Questions

28. a. Exemplify the data analytics process in detail. 12 1 1 1

(OR)

b.i. Determine the mean deviation for $-5, -4, 0, 4, 5$. 6 2 1 2

ii. Determine the variance and standard deviation for 1, 3, 5, 5, 6, 7, 9, 10. 6 2 1 2

29. a.i. Discuss on the operators used in R language with minimum three sample cod/ command for each. 6 2 2 2

ii. Highlight the role of looping in R with code snippet. 3 2 2 2

iii. Highlight the role of functions in R with code snippet. 3 2 2 2

(OR)

b. Articulate the various datatypes used in python with code snippets for each. 12 2 2 2

30. a. Comment on Apache hive architecture with neat diagram. 12 1 3 1

(OR)

b. Highlight the HDFS architecture's operation with neat diagram. 12 1 3 1

31. a. Exemplify the algorithm of mapreduce and the procedure for the same. 12 1 4 1

(OR)