



**SNS COLLEGE OF TECHNOLOGY**  
**DEPARTMENT OF MECHATRONICS ENGINEERING**  
**19MCT401 - DATA ANALYTICS IN AUTOMATION SYSTEM**



**UNIT 1 – INTRODUCTION TO BIG DATA**

1. What are the main components of big data?
  - a) HDFS
  - b) MapReduce
  - c) YARN
  - d) All the above
  
2. On which of the following platforms does Hadoop run?
  - a) Debian
  - b) Cross-platform
  - c) Bare metal
  - d) Unix-like
  
3. Data in \_\_\_\_\_ bytes size is called big data
  - a) Meta
  - b) Giga
  - c) Tera
  - d) Peta
  
4. Transaction of data of the bank is a type of.
  - a) Unstructured data
  - b) Structured data
  - c) Both a and b
  - d) None of the above
  
5. The total forms of big data is \_\_\_\_\_
  - a) 1
  - b) 2
  - c) 3
  - d) 4
  
6. Identify the incorrect big data Technologies.
  - a) Apache Pytorch
  - b) Apache Kafka
  - c) Apache Hadoop
  - d) Apache Spark
  
7. In which language is Hadoop written?
  - a) C++
  - b) Java
  - c) Rust
  - d) Python
  
8. \_\_\_\_\_ is a collection of data that is used in volume, yet growing exponentially with time

- a) Big Database
- b) Big DBMS
- c) Big Datafile
- d) Big Data

9. Identify among the options below which is general-purpose computing model and runtime system for Distributed Data Analytics.

- a) HDFS
- b) MapReduce
- c) Oozie
- d) All of the above

10. Choose the primary characteristics of big data among the following

- a) Value
- b) Variety
- c) Volume
- d) All of the above

## UNIT – II – DATA ANALYTICS

11. Data Analytics uses \_\_\_ to get insights from data.

- a. Statistical figures
- b. Numerical aspects
- c. Statistical methods
- d. None of the mentioned above

12. Amongst which of the following is / are the branch of statistics which deals with the development of statistical methods is classified as \_\_\_.

- A. Industry statistics
- B. Economic statistics
- C. Applied statistics
- D. None of the mentioned above

13. Linear Regression is the supervised machine learning model in which the model finds the best fit \_\_\_ between the independent and dependent variable.

- A. Linear line
- B. Nonlinear line
- C. Curved line
- D. All of the mentioned above

14. Amongst which of the following is / are the types of Linear Regression,

- A. Simple Linear Regression
- B. Multiple Linear Regression
- C. Both A and B
- D. None of the mentioned above

15. Amongst which of the following is / are the true about regression analysis?

- A. Describes associations within the data
- B. Modeling relationships within the data
- C. Answering yes/no questions about the data
- D. All of the mentioned above

16. Linear regression analysis is used to predict the value of a variable based on the value of another variable.

- A. True
- B. False

17. A Linear Regression model's main aim is to find the best fit linear line and the \_\_\_ of intercept and coefficients such that the error is minimized.

- A. Optimal values
- B. Linear line
- C. Linear polynomial
- D. None of the mentioned above

18. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.

- A. True
- B. False

19. The process of quantifying data is referred to as \_\_\_.

- A. Decoding
- B. Structure
- C. Enumeration
- D. Coding

20. Text Analytics, also referred to as Text Mining?

- A. True
- B. False

### UNIT – III – HADOOP

21. Mapper implementations are passed the JobConf for the job via the \_\_\_\_\_ method.

- a) JobConfigure.configure
- b) JobConfigurable.configure
- c) JobConfigurable.configurable
- d) None of the mentioned

22. Point out the correct statement.

- a) Applications can use the Reporter to report progress
- b) The Hadoop MapReduce framework spawns one map task for each InputSplit generated by the InputFormat for the job
- c) The intermediate, sorted outputs are always stored in a simple (key-len, key, value-len, value) format
- d) All of the mentioned

23. Input to the \_\_\_\_\_ is the sorted output of the mappers.

- a) Reducer
- b) Mapper
- c) Shuffle
- d) All of the mentioned

24. The right number of reduces seems to be \_\_\_\_\_

- a) 0.90
- b) 0.80
- c) 0.36
- d) 0.95

25. Point out the wrong statement.

- a) Reducer has 2 primary phases
- b) Increasing the number of reduces increases the framework overhead, but increases load balancing and lowers the cost of failures
- c) It is legal to set the number of reduce-tasks to zero if no reduction is desired
- d) The framework groups Reducer inputs by keys (since different mappers may have output the same key) in the sort stage

26. The output of the \_\_\_\_\_ is not sorted in the Mapreduce framework for Hadoop.

- a) Mapper
- b) Cascader
- c) Scalding
- d) None of the mentioned

27. Which of the following phases occur simultaneously?

- a) Shuffle and Sort
- b) Reduce and Sort
- c) Shuffle and Map
- d) All of the mentioned

28. Mapper and Reducer implementations can use the \_\_\_\_\_ to report progress or just indicate that they are alive.

- a) Partitioner
- b) OutputCollector
- c) Reporter
- d) All of the mentioned

29. \_\_\_\_\_ is a generalization of the facility provided by the MapReduce framework to collect data output by the Mapper or the Reducer.

- a) Partitioner
- b) OutputCollector
- c) Reporter
- d) All of the mentioned

30. \_\_\_\_\_ is the primary interface for a user to describe a MapReduce job to the Hadoop framework for execution.

- a) Map Parameters
- b) JobConf
- c) MemoryConf
- d) None of the mentioned

#### UNIT – IV – MANAGING HADDOP FILE

31. Which of the following class provides access to configuration parameters?

- a) Config
- b) Configuration
- c) OutputConfig
- d) None of the mentioned

32. Point out the correct statement.

- a) Configuration parameters may be declared static
- b) Unless explicitly turned off, Hadoop by default specifies two resources
- c) Configuration class provides access to configuration parameters
- d) None of the mentioned

33. \_\_\_\_\_ gives site-specific configuration for a given hadoop installation.

- a) core-default.xml
- b) core-site.xml
- c) coredefault.xml
- d) all of the mentioned

34. Administrators typically define parameters as final in \_\_\_\_\_ for values that user applications may not alter.

- a) core-default.xml
- b) core-site.xml
- c) coredefault.xml
- d) all of the mentioned

35. Point out the wrong statement.

- a) addDeprecations adds a set of deprecated keys to the global deprecations
- b) configuration parameters cannot be declared final
- c) addDeprecations method is lockless
- d) none of the mentioned

36. \_\_\_\_\_ method clears all keys from the configuration.

- a) clear
- b) addResource
- c) getClass

d) none of the mentioned

37. \_\_\_\_\_ method adds the deprecated key to the global deprecation map.

- a) addDeprecits
- b) addDeprecation
- c) keyDeprecation
- d) none of the mentioned

38. \_\_\_\_\_ checks whether the given key is deprecated.

- a) isDeprecated
- b) setDeprecated
- c) isDeprecatedif
- d) all of the mentioned

39. \_\_\_\_\_ is useful for iterating the properties when all deprecated properties for currently set properties need to be present.

- a) addResource
- b) setDeprecatedProperties
- c) addDefaultResource
- d) none of the mentioned

40. Which of the following adds a configuration resource?

- a) addDeprecation
- b) setDeprecatedProperties
- c) addDefaultResource
- d) addResource

#### UNIT – V – AUTOMATION DATA HANDLING

41. Pictorial representation of data using symbols is known as:

- A. Bar graph
- B. Pictograph
- C. Pie chart
- D. None of these

42. Double bar graphs display \_\_\_\_\_ sets of data simultaneously.

- A. Four
- B. Three
- C. Two
- D. No

43. \_\_\_\_\_ gives the number of times a particular entry occurs.

- A. Organisation of data
- B. Collection of Data
- C. Representation of Data
- D. Frequency distribution table

44. In the grouped data, each of the group is called:

- A. Class interval
- B. Collection of data
- C. Frequency
- D. Grouped frequency distribution

45. If 20-30 is the class interval of grouped data, then the lower class limit is:
- A. 50
  - B. 30
  - C. 20
  - D. 10
46. The difference between the upper-class limit and lower class limit of a class interval is called:
- A. Frequency
  - B. Width
  - C. Grouped data
  - D. Ungrouped data
47. The width or size of the class interval 30-40 is:
- A. 10
  - B. 30
  - C. 40
  - D. 70
48. If a coin is flipped in the air, what is the probability of getting a tail?
- A. 0
  - B.  $\frac{1}{2}$
  - C. 1
  - D. 2
49. A bag has 4 red balls and 4 green balls, what is the probability of getting a red ball randomly?
- A.  $\frac{1}{4}$
  - B.  $\frac{1}{8}$
  - C.  $\frac{1}{2}$
  - D. 0
50. If a die is thrown in the air, the probability of getting even numbers is:
- A.  $\frac{1}{2}$
  - B.  $\frac{1}{6}$
  - C.  $\frac{1}{3}$
  - D.  $\frac{2}{3}$