Choose the best Answer from the given alternative

1. A voluminous amount of structured, semi-structured, and unstructured data that has the potential to be mined for information.
    A. Small Data
    B. Meta Data
    C. Statistical Data
    D. Big Data
2. Combines data from multiple sources into a coherent store
    A. Data integration
    B. Data cleaning
    C. Data reduction
    D. Data transformation
3. A free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment.
    A. Hadoop
    B. Python
    C. R
    D. Apache Groovy
4. Why data reduction?
    A. A database/data warehouse may store terabytes of data.
    B. Complex data analysis may take a very long time to run on the complete data set.
    C. Integrate metadata from different sources.
    D. A and B
5. The branch of data mining concerned with the prediction of future probabilities and trends.
    A. In-memory Analytics
    B. Predictive Analytics
    C. Behavioral Analytics
    D. Big Data Analytics
6. Which is a data reduction strategies

A. Data compression
B. Data Cleaning
C. Data Integration
D. Data transformation

7. The science of examining raw data with the purpose of drawing conclusions about that information.
   A. Data Analytics
   B. In-memory Analytics
   C. Descriptive Analytics
   D. Predictive Analytics
8. An approach to querying data when it resides in a computer's random access memory (RAM), as opposed to querying data that is stored on physical disks.
   A. Deep Analytics
   B. Data Visualization
   C. In-memory Analytics
   D. Data Analytics
9. Leading analyst firm Gartner defines Big Data from three aspects, all starting with the letter V. Which of these are not a part of their consideration of big data?
   A. Value
   B. Varity
   C. Volume
   D. Velocity
10. Which one of the following is used for Data transformation and data discretization
    A. Entity identification problem
    B. Remove redundancies
    C. Normalization
    D. Dimensionality reduction
11. Which of the following is/are correct types of data?
    A. Semi structured
    B. Unstructured
    C. Semi data

D. Both A and B

12. What does SAAS stand for?
    A. System Aero surface actual simulation
    B. System as service
    C. Software acting as service
    D. Software as service
13. Where did Hadoop get its name from?
    A. Fictional character from literature
    B. To elephant
    C. Its acronym
    D. Its acronym for data
14. What is the name of the programming framework originally developed by Google that supports the development of applications for processing large data sets in a distributed computing environment?
    A. Hive
    B. Zookeeper
    C. Hadoop
    D. MapReduce
15. A method of storing data within a system that facilitates the collocation of data in various schemata and structural forms.
    A. Data visualization
    B. Data lake
    C. Big data management
    D. Deep analytic
16. According to a very recent Jaspersoft Survey, what is the most popular big data store?
    A. Relational database
    B. Hadoop HDFS
    C. Mongo DB
    D. Object based database

17. According to a study conducted by IBM, what is the largest single source where data is gathered?
    A. Email
    B. Social Media
    C. Business transaction
    D. Log data
18. Designing and building infrastructure for integrating and managing data from various resources?
    A. Data engineering
    B. Data analysis
    C. Data science
    D. Business Intelligence
19. Hadoop layer that is responsible for data processing?
    A. HDFS
    B. Map Reduce
    C. Master HDFS
    D. All
20. set of strategies, processes, applications, data, products, technologies and technical architectures which are used to support the collection, analysis, presentation and dissemination of business information
    A. Data engineering
    B. Data analysis
    C. Data science
    D. Business Intelligence
21. Which one of the following is NOT the limitations of data Analysis?
    A. Computing time to execute the analysis
    B. How much data can be put in memory per time unit
    C. How much data can be processed at a time
    D. None
22. The differences between Hadoop and RDBMS, which is not true about Hadoop?
    A. Node based flat structure

B. Suitable for Structured, unstructured data, supports variety of formats (xml, json)

C. Suitable for structured data

D. Which does not require any consistent relationships between data

23. Which of the following is the characteristics of distributed big data computing?

A. Concurrency of components

B. Luck of global clock: synchronization

C. Independent of failure of components: fault trounce

D. All

24. Which one of the following is NOT true about assumption and goals of HDFS

A. Hardware failure

B. Stream data access

C. Moving computation is expensive than moving data

D. Portability across heterogeneous hardware and software

25. A type of predictive analytics that shows linear relationships between the independent variables (X-Axis) and the dependent Variable (Y- Axis)?

A. Linear regression

B. Polynomial regression

C. Logistics regression

D. None

26. The data could be presented through a low-latency _____ technology.

A. DBMS

B. NoSQL

C. Data store

D. None of the mentioned above

27. Azure Synapse Analytics provides a managed service for _____.

A. Large-scale

B. Cloud-based

C. Data warehousing

D. All of the mentioned above

28. The goal of most big data solutions is to provide insights into the data through _____.
    A. Hive
    B. HBase
    C. Analysis and reporting
    D. All of the mentioned above
29. To process large data sets quickly, big data architectures use.
    A. Distributed computing
    B. Cluster computing
    C. Parallel computing
    D. All of the above
30. Large _____ of data is considered as big data.
    A. Volume
    B. Veracity
    C. Variety
    D. None of the mentioned above
31. Veracity makes sure that the data is _____.
    A. Inconsistence
    B. Variant
    C. Accurate
    D. Volume
32. _____ are two techniques used in descriptive analytics to discover historical data.
    A. Data ingestion and data mining
    B. Data warehouse and data storage
    C. Data aggregation and data mining
    D. Data ingestion and data storage
33. A data mining technique that classifies each record in a dataset based on a combination of the classes of the k record(s). Sometimes called the k-nearest neighbor technique.
    A. Nearest neighbor method
    B. Decision trees

C. Artificial neural networks

D. Rule induction:

34. Data Mining Techniques that is used to establish a specific predetermined class for each record in a database from a finite set of possible class values.

A. Value prediction

B. Database Segmentation

C. Classification

D. Link Analysis

35. Which is NOT Major Tasks in Data Preprocessing?

A. Data cleaning

B. Data reduction

C. Data discretization

D. All are Correct