MULTIPLE CHOICE QUESTION

1. Data in   bytes size is called Big Data.

A. Tera
B. Giga
C. Peta
D. Meta

2. How many V's of Big Data?

A. 2
B. 3
C. 4
D. 5

3.   Unprocessed data or processed data are observations or measurements that can be expressed as text, numbers, or other types of media.

A. True
B. False

4. In computers, a   is a symbolic representation of facts or concepts from which information may be obtained with a reasonable degree of confidence.

A. Data
B. Knowledge
C. Program
D. Algorithm

5. In Big Data environments, Velocity refers –

A. Data can arrive at fast speed
B. Enormous datasets can accumulate within very short periods of time
C. Velocity of data translates into the amount of time it takes for the data to be processed
D. All of the mentioned above

6. In Big Data environments, Variety of data includes –

A. Includes multiple formats and types of data
B. Includes structured data in the form of financial transactions,
C. Includes semi-structured data in the form of emails and unstructured data in the form of images
D. All of the mentioned above

7. In Big Data environment, Veracity of data refers -

A. Quality or fidelity of data
B. Large size of the data that cannot be process
C. Small size of the data that can easily process
D. All of the mentioned above

8. Which of the following are Benefits of Big Data Processing?

A. Cost Reduction
B. Time Reductions
C. Smarter Business Decisions
D. All of the mentioned above

9. Structured data conforms to a data model or schema and is often stored in tabular form.

A. True
B. False

10. Data that does not conform to a data model or data schema is known as _.

A. Structured data
B. Unstructured data
C. Semi-structured data
D. All of the mentioned above

11. Amongst which of the following is/are not Big Data Technologies?

A. Apache Hadoop
B. Apache Spark
C. Apache Kafka
D. Apache Pytarch

12. __ involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task.

A. Parallel data processing
B. Single channel processing
C. Multi data processing
D. None of the mentioned above

13. Amongst which of the following can be considered as the main source of unstructured data.

A. Twitter
B. Facebook
C. Webpages
D. All of the mentioned above

14. Amongst which of the following shows an example of unstructured data,

A. Students roll number, age
B. Videos
C. Audio files
D. Both B and C

15. Scalability, elasticity, resource pooling, self-service, low cost and fault tolerance are the features of,

A. Cloud computing
B. Power BI
C. System development
D. None of the mentioned above

16. Amongst which of the following is/are the cloud deployment models,

A. Public Cloud
B. Private Cloud
C. Hybrid Cloud
D. All of the mentioned above

17. Virtualization separates resources and services from the underlying physical delivery environment.

A. True
B. False

18. What is a Virtual Machine (VM)?
A. Virtual representation of a physical computer
B. Virtual representation of a logical computer
C. Virtual System Integration
D. All of the mentioned above

19. In the given Virtual Architecture, name the missing layer,

A. Virtualization layer
B. Storage layer
C. Abstract layer
D. None of the mentioned above

20. MongoDB is a   database.
A. SQL
B. DBMS
C. NoSQL
D. RDBMS

21. MongoDB support cross platform and is written in   language.
A. Python
B. C++
C. R
D. Java

22. Amongst which of the following is / are true to run MongoDB?
A. High availability through built-in replication and failover
B. Management tooling for automation, monitoring, and backup
C. Fully elastic database as a service with built-in best practices
D. All of the mentioned above

23. Big data deals with high-volume, high-velocity and high-variety information assets,
A. True
B. False

24.   ______hypervisor runs directly on the underlying host system. It is also known as "Native Hypervisor" or "Bare metal hypervisor".
A. TYPE-1 Hypervisor
B. TYPE- 2 Hypervisor
C. Both A and B
D. None of the mentioned above

25   is also known as "Hosted Hypervisor".
A. TYPE-1 Hypervisor
B. TYPE- 2 Hypervisor
C. Both A and B
D. None of the mentioned above

26. In the layered architecture of Big Data Stack, Interfaces and feeds,

A. Internally managed data
B. Data feeds from external sources.
C. It provides access to each and every layer & components of big data stack
D. All of the mentioned above

27 ___is the supporting physical infrastructure is fundamental to the operation and scalability of big data architecture.

A. Redundant physical infrastructure
B. Integrated System
C. Integrated Database
D. All of the mentioned above

28. Data in bytes size is called Big Data. (GATE 2021)

A. Tera
B. Giga
C. Peta
D. Meta

Answer: C) Peta GATE:, 2021

29. In computers, a is a symbolic representation of facts or concepts from which information may be obtained with a reasonable degree of confidence. (GATE 2021)

A. Data
B. Knowledge
C. Program
D. Algorithm

30. Amongst which of the following represents the Use of Hadoop, (GATE 2019)

A. Robust and Scalable
B. Affordable and Cost Effective
C. Adaptive and Flexible
D. All of the mentioned above

31. Nis a platform for developing data flows for the extraction, transformation, and loading (ETL) of huge datasets, as well as for data analysis.

A. Spark
B. HBase
C. Hive
D. Pig

32. In contrast to relational databases, Hive is a query engine that supports the elements of SQL that are specifically designed for querying data.

A. True
B. False

33. Custom extensions built in the programming language are also supported by Hive.

A. Java
B. C#
C. C
D. C++

34. Amongst which of the following is / are correct,

A. Hive is a relational database that supports SQL queries.
B. Pig is a relational database that supports SQL queries.
C. Both A and B
D. None of the mentioned above

35. In order to analyze all of this Big Data, Hive is a tool that has been developed.

A. True
B. False

36. general-purpose model and runtime framework for distributed data analytics.

A. Mapreduce
B. Spark
C. Hive
D. All of the mentioned above

37. Scalability is prioritized over latency in jobs such as _.

A. HBase
B. HDFS
C. Hive
D. Mapreduce

38. ___node serves as the Slave and is responsible for carrying out the Tasks that have been assigned to it by the JobTracker.

A. TaskReduce
B. Mapreduce
C. TaskTracker
D. JobTracker

39. Apache Hive is data storage and that stores and organizes data for study and querying.

A. Querying tool
B. Mapper
C. MapReduce
D. All of the mentioned above

40. The MapReduce framework is responsible for processing one or more pieces of data and producing the output results as _.

A. Maptask
B. Task execution
C. Mapper
D. All of the mentioned above

41. Apache Hive is a data infrastructure that is built on top of the Hadoop platform.

A. Warehouse
B. Map
C. Reduce
D. None of the mentioned above

42. The Hadoop framework is built in Java, which means that MapReduce applications do not need to be written in _.

A. C#
B. C
C. Java
D. None of the mentioned above

43. ___maps input key/value pairs to a set of intermediate key/value pairs.

A. Reducer
B. Mapper
C. File system
D. All of these

44. HQL is a query language that is used to construct the custom map-reduce framework in Hive, which is written in _.

A. Java
B. PHP
C. C#
D. None of the mentioned above

45. The is the default partitioned in Hadoop, and it offers a method called Get Partition that allows us to partition data.

A. Hash Partitioner
B. Map function
C. Reduce function
D. All of the mentioned above

46. Hadoop is a framework that can be used in conjunction with a number of related products. Among the most common cohorts are _.

A. MapReduce, Hive and HBase
B. Hive, Spark and HBase
C. Spark, Hive and ZooKeeper
D. Spark, HBase and Hive

47. __ is best described as a programming model that is used to construct Hadoop-based applications that can be scaled up and down.

A. Oozie
B. Zookepper
C. MapReduce
D. All of the mentioned above

48. Amongst which of the following is/are the Hive function Meta commands.

A. Show functions
B. Describe function
C. Both A and B
D. None of the mentioned above

Answer: C) Both A and B

49. ___is a shell utility that can be used to run Hive queries in either interactive or batch mode, depending on the situation.

A. $HIVE_HOME/bin/hive
B. $HIVE/bin/
C. $HIVE_HOME/hive
D. All of the mentioned above

50. The tool has the capability of listing all of the possible database schemas.

A. sqoop-list-databases
B. Hbase-list
C. hive schema
D. sqoop-list-columns

51. Amongst which of the following is/are true with reference to User-defined Functions of Hive.

A. function that fetches one or more columns from a row as arguments
B. It returns a single value
C. Both A and B
D. None of the mentioned above

52. Amongst which of the following is/are correct.

A. Default location of Hadoop configuration is in $HADOOP /conf/ HOME
B. If $HADOOP HOME is specified, Sqoop will utilise the default installation location
C. default location of Hadoop configuration is in $HADOOP HOME/conf/
D. Sqoop command-line tool serves as a wrapper for the bin/hadoop script that is included with Hadoop as a base.

53. A serves as the master, and each cluster has just one NameNode. (GATE 2020)

A. Data Node
B. Block Size
C. Data block
D. NameNode

54. HDFS always needs to work with large data sets.

A. True
B. False

55. HDFS operates in a manner.

A. Master-slave architecture
B. Master-worker architecture
C. Worker-slave architecture
D. All of the mentioned above

56. HDFS follows the write-once, read-many.

A. True
B. False

57. Amongst which of the following is not aligns as a characteristic of HDFS? (GATE 2020)

A. HDFS file system is well suited for storing data associated with applications that require low latency data access.
B. HDFS is well-suited for storing data connected to applications that require low-latency data access to be performed.
C. HDFS is not suited for instances in which multiple/simultaneous writes to the same file are required.
D. None of the mentioned above

58. In order to interact with HDFS, a command line interface named is provided. (GATE 2019)

A. HDFS Shell
B. DFS Shell
C. K Shell
D. FS Shell

59. HDFS stores data in a distributed manner, the data can be processed in parallel on a ______of nodes.

A. Cluster
B. Data Node
C. Master Node
D. None of the mentioned above

60. With reference to HDFS, Name Node is the prime node which contains metadata.

A. True
B. False

61. The database which is used to manage and store data in real time is called ___.

A. Traditional database
B. Operational database
C. Database Management System
D. None of the mentioned above

62. Database requirements for operational data includes ___.

A. Indexing and Cataloging, Replication
B. File Storage and Structure, Query Processing
C. Transactions Support
D. All of the mentioned above

63. Indexing and Cataloguing refers to efficiently store data that can be retrieved?

A. True
B. False

64. File Storage and structure is an important function of the operational database to robust enough to sort and store files at relevant locations?

A. True
B. False

Answer: A) True

65. Query processing system refers to the entire process from translating a ___ to the database system.

A. Query
B. Statement
C. Function
D. None of the mentioned above

66. Operational Database with distributed systems and ___ based system can harness the true potential with big data.

A. SQL
B. NoSQL
C. PL / SQL
D. None of the mentioned above

67. ___ a record is created for every search key valued in the database.

A. Primary Index
B. Secondary Index
C. Complex Index
D. None of the mentioned above

68. A non-clustered index tells us where the data lies?

A. True
B. False

69. Data warehouse modeling is the initial stage of building a data warehouse wherein the ___ is designed.

A. Schema
B. Table
C. Both A and B
D. None of the mentioned above

70. An operational database is designed to run the day-to-day operations or transactions of your business?

A. True
B. False

71. As companies move past the experimental phase with Hadoop, many cite the need for additional capabilities, including ______-_-_-_

a) Improved data storage and information retrieval
b) Improved extract, transform and load features for data integration
c) Improved data warehousing functionality
d) Improved security, workload management, and SQL support

72. Point out the correct statement.

a) Hadoop do need specialized hardware to process the data
b) Hadoop 2.0 allows live stream processing of real-time data
c) In the Hadoop programming framework output files are divided into lines or records
d) None of the mentioned

73. According to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop? (GATE 2022)

a) Big data management and data mining
b) Data warehousing and business intelligence
c) Management of Hadoop clusters
d) Collecting and storing unstructured data

74. Hadoop is a framework that works with a variety of related tools. Common cohorts include ______-_-_

a) MapReduce, Hive and HBase
b) MapReduce, MySQL and Google Apps
c) MapReduce, Hummer and Iguana
d) MapReduce, Heron and Trumpet

75. Point out the wrong statement.

a) Hardtop processing capabilities are huge and its real advantage lies in the ability to process terabytes & petabytes of data

b) Hadoop uses a programming model called "MapReduce", all the programs should conform to this model in order to work on the Hadoop platform

c) The programming model, MapReduce, used by Hadoop is difficult to write and test

d) All of the mentioned

76. What was Hadoop named after? (GATE 2020)

a) Creator Doug Cutting's favorite circus act

b) Cutting's high school rock band

c) The toy elephant of Cutting's son

d) A sound Cutting's laptop made during Hadoop development

77. All of the following accurately describe Hadoop, EXCEPT ______________

a) Open-source

b) Real-time

c) Java-based

d) Distributed computing approach

78. __ can best be described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.

a) MapReduce

b) Mahout

c) Oozie

d) All of the mentioned

79. __ has the world's largest Hadoop cluster.

a) Apple

b) Datamatics

c) Facebook

d) None of the mentioned

80. Facebook Tackles Big Data With __based on Hadoop. (GATE 2021)

a) 'Project Prism'

b) 'Prism'

c) 'Project Big'

d) 'Project Data'

81. __ ___is a platform for constructing data flows for extract, transform, and load (ETL) processing and analysis of large datasets.

a) Pig Latin

b) Oozie

c) Pig

d) Hive

82. Point out the correct statement.

a) Hive is not a relational database, but a query engine that supports the parts of SQL specific to querying data
b) Hive is a relational database with SQL support
c) Pig is a relational database with SQL support
d) All of the mentioned

83. __ hides the limitations of Java behind a powerful and concise Clojure API for Cascading.
a) Scalding
b) HCatalog
c) Cascalog
d) All of the mentioned

84. Hive also support custom extensions written in ______________
a) C#
b) Java
c) C
d) C++

85. Point out the wrong statement.
a) Elastic MapReduce (EMR) is Facebook's packaged Hadoop offering
b) Amazon Web Service Elastic MapReduce (EMR) is Amazon's packaged Hadoop offering
c) Scalding is a Scala API on top of Cascading that removes most Java boilerplate
d) All of the mentioned

86. __ ___is the most popular high-level Java API in Hadoop Ecosystem
a) Scalding
b) HCatalog
c) Cascalog
d) Cascading

87. __ ___is general-purpose computing model and runtime system for distributed data analytics.
a) Mapreduce
b) Drill
c) Oozie
d) None of the mentioned

88. The Pig Latin scripting language is not only a higher-level data flow language but also has operators similar to ______.__.__
a) SQL
b) JSON

c) XML
d) All of the mentioned

89. __ jobs are optimized for scalability but not latency.
a) Mapreduce
b) Drill
c) Oozie
d) Hive

90. __ is a framework for performing remote procedure calls and data serialization.
a) Drill
b) BigTop
c) Avro
d) Chukwa

91. Which one is not in Basic analytics for insight

a) Slicing and Dicing of data
b) Reporting & Basic monitoring.
c) Simple Visualizations
d) The business process.

92. Find out which one is under Advanced analytics for insight (GATE 2020)

a) predictive modeling
b) drive revenue.
c) transparency
d) business intelligence

93. Which Analytics type become part of the business process.

a) Operationalized analytics
b) Basic analytics
c) Advanced analytics
d) Monetized analytics

94. Which Analytics type is utilized to directly drive revenue.

a) Basic analytics
b) Operationalized analytics
c) Monetized analytics
d) Advanced analytics

95. Slicing and dicing refers

a) to breaking down your data into smaller sets of data
b) to monitor large volumes of datain real time
c) to identify anomalies
d) provides algorithms for complex analysis

96. Basic monitoring refers

a) to breaking down your data into smaller sets of data
b) to monitor large volumes of datain real time
c) to identify anomalies
d) provides algorithms for complex analysis

97. Anomaly identification refers

a) to breaking down your data into smaller sets of data
b) to monitor large volumes of datain real time
c) to identify anomalies an event where the actual observation differs from what you expected
d) provides algorithms for complex analysis

98. Predictive modeling used to (GATE 2021)

a) to determine future outcomes
b) to find patterns in thatdata
c) calculates the distances between the record and points
d) broke the data into training data anda test data set

99. Advanced analytics can be deployed to find patterns in

a) data & prediction
b) forecasting
c) complex event processing
d) All of the above

100. The process of analyzing unstructured text, extracting relevant information, and transforming it into structured information is called

a) Text analytics
b) data-mining
c) segmentation
d) cluster analysis

101. The potential characteristics of your data

a) It can come from untrusted sources
b) It can be real-time
c) It can be dirty
d) All of the above

102. Big data consists of (GATE 2019)

a) Structured data
b) Semi-structureddata
c) Unstructured data
d) All of the above

103. Dirty data refers

a) inaccurate data

b) incompletedata
c) erroneous data
d) All of the above

104. What are the Infrastructure needed to support big data

a) Integrate technologies
b) Process data in motion
c) Warehouse data
d) All of the above

105. Users of Orbitz perform

a) the company collects hundreds of gigabytes of raw data each day from these searches
b) useful information inthe web log files that it was collecting from its web analytics software
c) Both of (A) & (B)
d) None of these

106. Which one is false statement in the following

a) Hadoop providedthe distributed file system
b) Hive provided an SQL-type interface
c) A series of steps to put the data into Hive. After the data was in Hive, the company used machine learning.
d) None of these

107. Nokia provides

a) multi petabyte platform
b) wireless communication devices and services
c) improve customer retention
d) All of the above

108. A number of vendors on the market today support big data solutions

a) IBM , Oracle
b) SAS , Pentaho
c) Tableau
d) All of the above

109. InfoSphere Streams product is tightly integrated with

a) its Statistical Package for the Social Sciences (SPSS) statistical software to support real-time predictive analytics
b) capability to dynamically update models based on real-time data
c) Both of (A) & (B)
d) None of the above

110. The different kinds of unstructured data are

a) Documents, E-mails
b) Log files , Tweets

c) Face book posts
d) All of the above

111. NLP abbreviated as

a) Numerous Language Processing
b) Natural Language Processing
c) Numerous Language Project
d) Natural Language Project

112. What are the methods exist for analyzing unstructured data

a) Natural Language Processing
b) knowledge discovery & data mining
c) Information retrieval & statis tics
d) All of the above

113. Search is about retrieving a document based on

a) what end users already know they are looking for.
b) discovering information
c) classification of documents
d) None of these above

114. A goal of NLP is

a) To derive meaning from text.
b) Generally makes use of linguistic concepts such as grammatical structures and parts of speech
c) To determine who did what to whom, when, where, how, and why.
d) All of the above

115. Lexical/morphological analysis

a) examines the characteristics of an individual word
b) uses grammatical structure to dissect the text and put individual words into context
c) determines the possible meanings of a sentence.
d) to determine the meaning of text beyond the sentence level.

116. Which one is uses grammatical structure to dissect the text and put individual words into context.

a) Lexical analysis
b) Semantic analysis
c) Syntactic analysis

d) Discourse-level analysis

117. To extract information from various document sources, organiza tions sometimes need to develop rules. These rules can be

a) The name of a person must start with a capital letter.
b) Every course on the college website must follow a three-digit course number and a semicolon.
c) A logo must appear in a certain location on every page.
d) All of the above

118. Sentiment analysis is used to

a) identify viewpoints or emotions in the underlying text
b) organizing information into hierarchical relationships
c) None of these
d) All of the above

119. Text Analytics Tools for Big Data

a) Attensity
b) Clarabridge , OpenText
c) IBM , SAS
d) All of the above

120. NASA is using predictive models to

a) analyze safety data on aircrafts
b) understand whether the introduction of a new technology into an aircraft
c) dealing with a massive amount of data
d) All of the above

121. Which are the major categories of big data integration?

a) Theintegration of multiple big data sources in big data environments
b) The integration of unstructured big data sources with structured enterprise data.
c) Both of these (A) & (B)
d) None of these

122. Which one is not in the stages of Big data analysis?

a) Exploratory stage
b) Codifying stage
c) Integration and incorporation stage
d) None of the above

123. To complete your analysis, you need to move large amounts of data from

a) log files
b) Twitter feeds , RFID tags
c) weather data feeds
d) All of these above.

124. Which is widely used as an underlying building block for capturing and

processing big data

a) Hadoop
b) Twitter feeds , RFID tags
c) weather data feeds
d) All of these above.

125. Which are two primary components of Hadoop

a) Hadoop Distributed File System (HDFS)
b) MapReduce
c) Both (A) & (B)
d) None of the above

126. Traditional integration tools

a) E
T
L
b) S
S
L
c) P
S
L
d) T
T
L

127. Which one is not true for the following?

a) Traditional integration tools such as ETL would not be fast enough to move the large streams ofdata in time to deliver results for analysis.
b) Flume is used to collect large amounts of log data from distributed servers.
c) Flume is designed for scalability and can continually add more resources to a system to handle extremely large amounts of data in an efficient way.
d) None of the above.

128. To codify the relationship between your big data analytics and your operational data, you need to

a) Integrate the data.
b) Split the data
c) Divide the data
d) Explore the data

129. Traditionally, data integration has focused on themovement of data through

a) middleware
b) specifications on message passing
c) application programming interfaces (APIs).
d) All of the above

130. Traditional tools for data integration are evolving to handle the increasing variety of

a) Unstructured data
b) The growing volume
c) Velocity of big data
d) All of the above

131. What are the basic principles apply from specific to individual systems/applications

a) You must create a common understanding of data definitions
b) You must develop of a set of data services to qualify the data and make it consistent and ultimately trustworthy
c) You need a streamlined way to integrate your big data sources and systems of record
d) All of the above

132. What are the new tools used to support integrationof big data environments (GATE 2021)

a) Sqoop
b) Scribe
c) Both (A) & (B)
d) None of the above

133. What are the important functions of ETL which required to get data from one data environment and put it into another data environment.

a) **Extract**: Read data from the source database.
b) **Transform**: Convert the format of the extracted data so that it conforms to the requirements of the target database. Transformation is done by using rules or merging data with other data.
c) **Load**: Write data to the target database
d) All of the above.

134. Customer relationship management [CRM]) used to

a) analyze and report on data relevant to their specific business focus
b) batch processing in data warehouse environments
c) to consolidate information across disparate sources
d) None of the above

135. What are the statements true about Data Transformation?

a) Data transformation is the process of changing the format of data so thatit can be used by different applications.
b) The process of data transformation is made far more complex because of the staggering growth in the amount of unstructured data.
c) Data transformation tools are not designed to work well with unstructured data
d) All of the above

136. Which tools can transform the data in the source or target database without requiring an ETL server? (GATE 2020)

a) ELT (extract, load, and transform)
b) ELT uses structured query language (SQL) to transform the data
c) ETL tools extracted the data to an intermediary location to perform the transformation before loading the data to the data warehouse.
d) Massively parallel processing systems and columnar databases

137. What are the different phase approach followed for Data Quality?

a) Look for patterns in big data without concern for data quality.
b) After you locate your patterns and establish results that are important to the business, apply the same data quality standards that you apply to your traditional data sources.
c) Both (A) & (B)
d) None of the above

138. The quality of data refers to characteristics about the data, including

a) consistency, accuracy
b) reliability, completeness, timeliness
c) reasonableness, and validity

139. All of the aboveData quality software can be used to

a) identify all the variations of the com- pany name in your different data stores
b) ensure that you know everything that this customer purchases from your business.
c) Cleans up or removes redundant data
d) All of the above

140. Data profiling tools are used in the data quality process to help you to understand

a) The content , Structure & .Condition of your data
b) Analyze the data to identify errors and inconsistencies
c) you can ensure that your big data is complete and consistent.
d) All of the above

141. Which Hadoop tools can be used for the transformation process?

a) HiveQL
b) Pig Latin
c) Both (A) & (B)
d) None of the above

142. What are the two techniques for managing the flow of data.

a) **Streaming technology** is closely tied to the volume of the data
b) **Complex event processing** of the volume of data is secondary to the capabilityto match data to rules.
c) Both (A) & (B)
d) None of the above

143. Which statements are true about Complex event processing?

a) CEP is dependent on data streams.
b) CEP is not required for streaming data ,. Like streaming data, CEP relies on analyzing streams of data in motion.
c) Streaming computing is designed to handle a continuous stream of a large amount of unstructured data.
d) All of the above

144. In the Hadoop cluster, data is collected in which mode and then processed. (GATE 2019)

a) Batch
b) Streaming
c) Real-time calculation
d) Cluster

145. Which statement is false in streaming of data?

a) Implicit metadata from unstructured data, it is possible to parse the information using eXtensible Markup Language (XML)
b) XML is a technique for presenting unstructured text files with meaningful tags.
c) Examples of products for streaming data include IBM's InfoSphere Streams, Twitter's Storm, and Yahoo's S4
d) None of the above.

146. IBM InfoSphere Streams used to

a) perform complex analytics of heterogeneous data types
b) perform text, images, audio, voice, VoIP, video, web traffic, e-mail, GPS data, financial transaction data, satellite data, and sensors.
c) provides continuous analysis of massive data volumes.
d) All of the above

147. Twitter's Storm is an open source real-time analytics engine developed bya company called

a) BackType
b) InfoSphere
c) Apache S4
d) None of the above

148. Companies using Storm in their big data implementations include

a) Groupon
b) RocketFuel
c) Navisite and Oolgala
d) All of the above

149. Which statement is not true in CEP?

a) Streams are intended to analyze large volumes of data in real time
b) Complex Event Processing is a technique for tracking, analyzing, and processing data as an event happens
c) CEP is an advanced approach based on simple event processing that collects and combines data from different relevant sources to discover events and patterns that can result in action

d) None of the above

150. **The set of “V” characteristics that are key to operationalizing big data includes**

a) Validity: Is the data correct and accurate for the intended usage?

b) Veracity: Are the results meaningful for the given problem space?

c) Volatility: How long do you need to store this data?

d) All of the above.

ANSWERS:

1	**C**	16	**D**	31	**D**	46	**A**	61	**B**	76	**C**	91	**D**	106	**D**	121	**C**	136	**A**
2	**D**	17	**A**	32	**A**	47	**C**	62	**D**	77	**B**	92	**A**	107	**D**	122	**D**	137	**C**
3	**A**	18	**A**	33	**A**	48	**C**	63	**A**	78	**A**	93	**A**	108	**D**	123	**D**	138	**D**
4	**A**	19	**A**	34	**C**	49	**A**	64	**A**	79	**C**	94	**C**	109	**C**	124	**D**	139	**D**
5	**D**	20	**C**	35	**A**	50	**A**	65	**A**	80	**A**	95	**A**	110	**D**	125	**C**	140	**D**
6	**D**	21	**B**	36	**A**	51	**C**	66	**B**	81	**C**	96	**B**	111	**B**	126	**A**	141	**C**
7	**D**	22	**D**	37	**D**	52	**D**	67	**B**	82	**A**	97	**C**	112	**D**	127	**D**	142	**C**
8	**D**	23	**A**	38	**C**	53	**D**	68	**A**	83	**C**	98	**A**	113	**A**	128	**A**	143	**D**
9	**A**	24	**A**	39	**A**	54	**A**	69	**B**	84	**B**	99	**D**	114	**D**	129	**D**	144	**A**
10	**B**	25	**B**	40	**A**	55	**B**	70	**A**	85	**A**	100	**A**	115	**A**	130	**D**	145	**D**
11	**D**	26	**D**	41	**A**	56	**A**	71	**D**	86	**D**	101	**D**	116	**C**	131	**D**	146	**D**
12	**A**	27	**A**	42	**C**	57	**C**	72	**B**	87	**A**	102	**D**	117	**D**	132	**C**	147	**D**
13	**D**	28	**C**	43	**B**	58	**D**	73	**A**	88	**A**	103	**D**	118	**A**	133	**D**	148	**D**
14	**D**	29	**A**	44	**A**	59	**A**	74	**A**	89	**D**	104	**D**	119	**D**	134	**A**	149	**D**
15	**A**	30	**D**	45	**A**	60	**A**	75	**C**	90	**C**	105	**C**	120	**D**	135	**D**	150	**D**