Total number of questions : 60

12902_DATA SCIENCE AND BIG DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 2. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

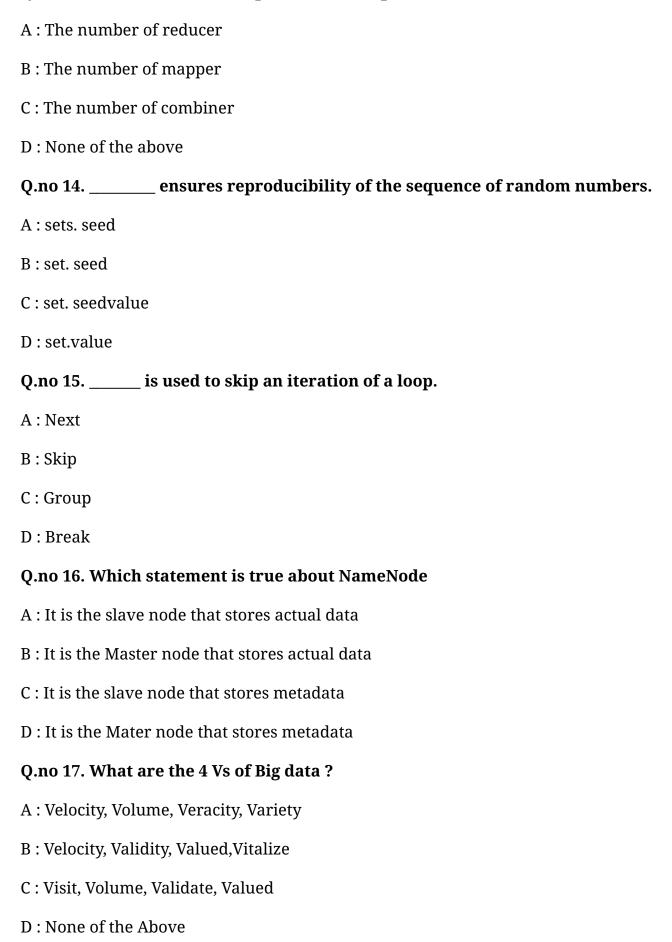C : Sampling

D : None of the above

Q.no 3. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 4. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 5. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 6. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 7. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 8. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 9. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 10. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 11. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 12. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 13. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 14. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 15. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 16. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 17. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 18. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 19. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 20. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 21. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 22. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 23. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 24. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 25. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 26. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 27. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.
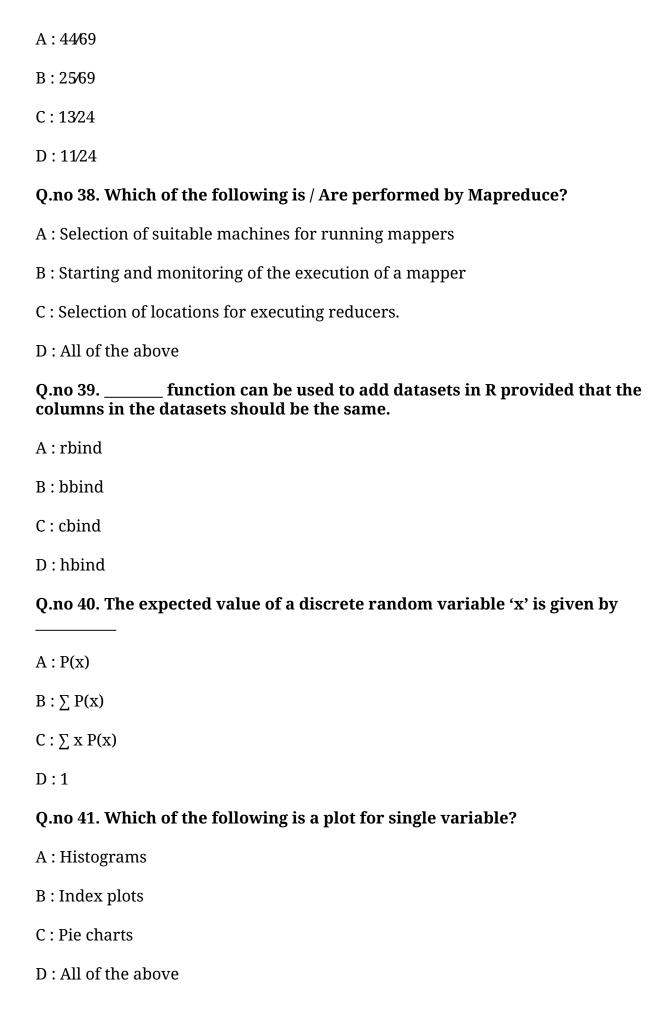
A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 28. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 29. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 30. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 31. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 32. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 33. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 34. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 35. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 36. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 37. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44/69

B : 25/69

C : 13/24

D : 11/24

Q.no 38. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 39. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 40. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 41. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 42. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 43. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 44. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 45. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 46. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 47. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 48. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 49. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 50. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 51. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object

3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 52. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 53. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 54. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2).........p(xn*xn-1)

Q.no 55. ----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 56. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 57. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 58. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 59. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 60. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 1. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 2. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 3. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 4. The total number of partitioner is equal to
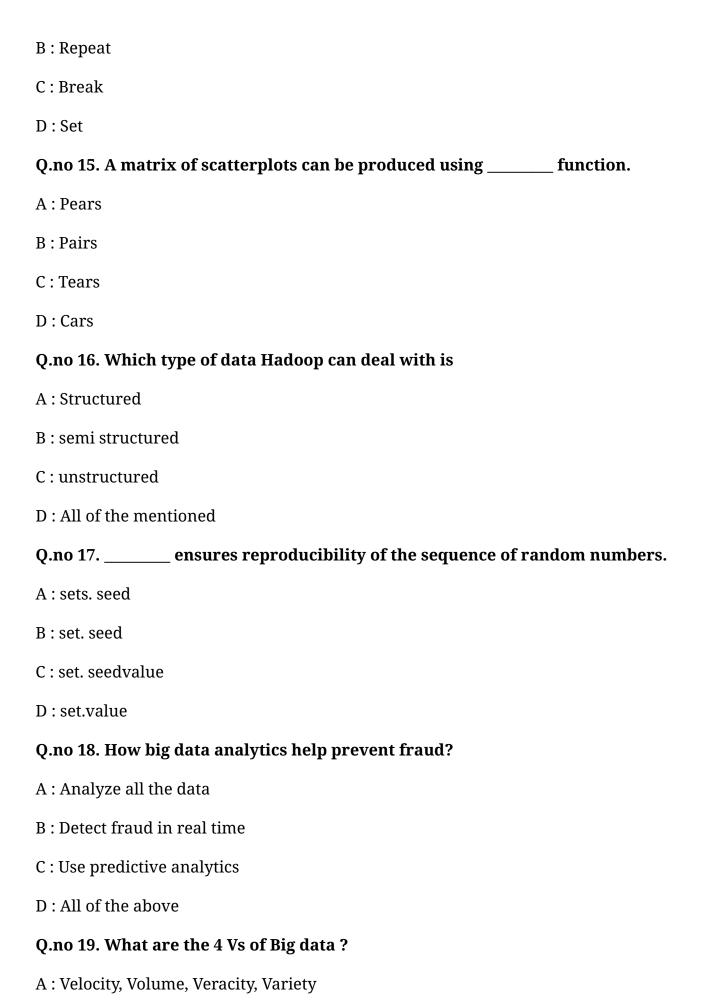
A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 5. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 6. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 7. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 8. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 9. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 10. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 11. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 12. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 13. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 14. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 15. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 16. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 17. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 18. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 19. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 20. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 21. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 22. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 23. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 24. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 25. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 26. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 27. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 28. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 29. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 30. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 31. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 32. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 33. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 34. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 35. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 36. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 37. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

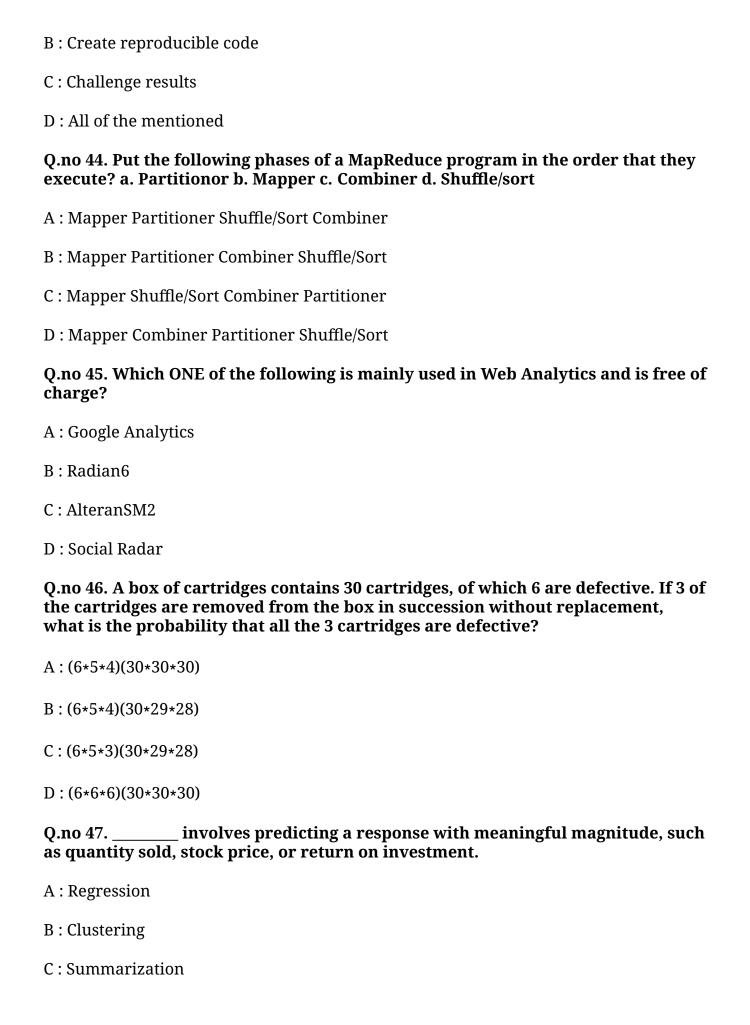C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 38. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 39. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 40. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 41. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 42. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 43. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 44. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 45. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 46. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 47. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 48. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 49. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 50. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 51. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 52. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 53. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 54. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 55. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 56. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 57. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 58. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 59. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 60. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 1. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 2. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 3. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 4. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 5. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 6. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 7. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 8. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 9. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 10. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 11. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 12. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 13. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 14. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 15. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 16. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 17. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 18. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 19. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 20. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 21. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 22. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 23. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 24. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 25. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 26. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 27. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 28. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 29. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 30. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 31. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 32. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 33. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 34. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 35. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 36. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 37. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 38. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 39. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 40. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 41. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 42. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 43. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 44. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 45. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 46. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 47. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 48. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 49. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 50. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 51. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 52. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 53. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 54. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 55. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 56. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 57. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 58. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 59. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 60. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 1. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 2. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 3. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 4. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 5. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 6. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 7. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 8. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 9. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 10. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 11. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 12. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 13. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 14. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 15. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 16. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 17. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 18. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 19. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 20. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 21. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 22. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 23. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 24. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 25. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 26. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 27. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 28. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 29. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 30. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 31. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 32. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 33. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 34. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 35. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 36. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 37. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 38. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 39. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 40. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 41. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 42. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 43. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 44. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 45. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 46. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 47. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 48. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 49. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 50. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 51. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 52. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 53. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 54. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 55. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 56. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

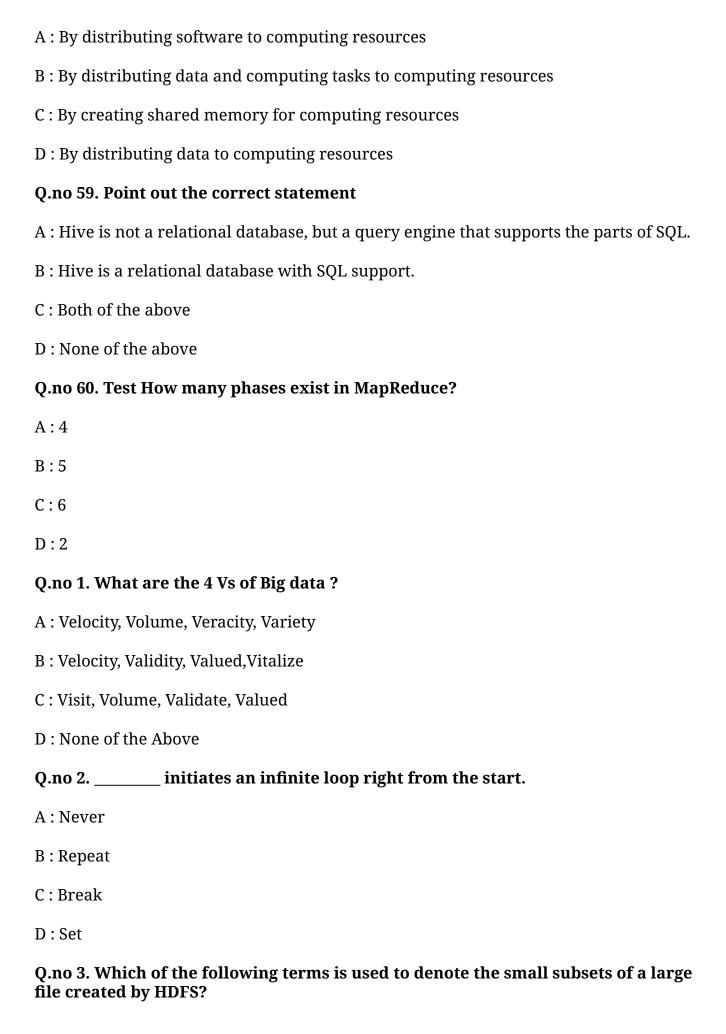**Q.no 57. The objectives for web analytics are likely to concern:**

A : Facebook messages

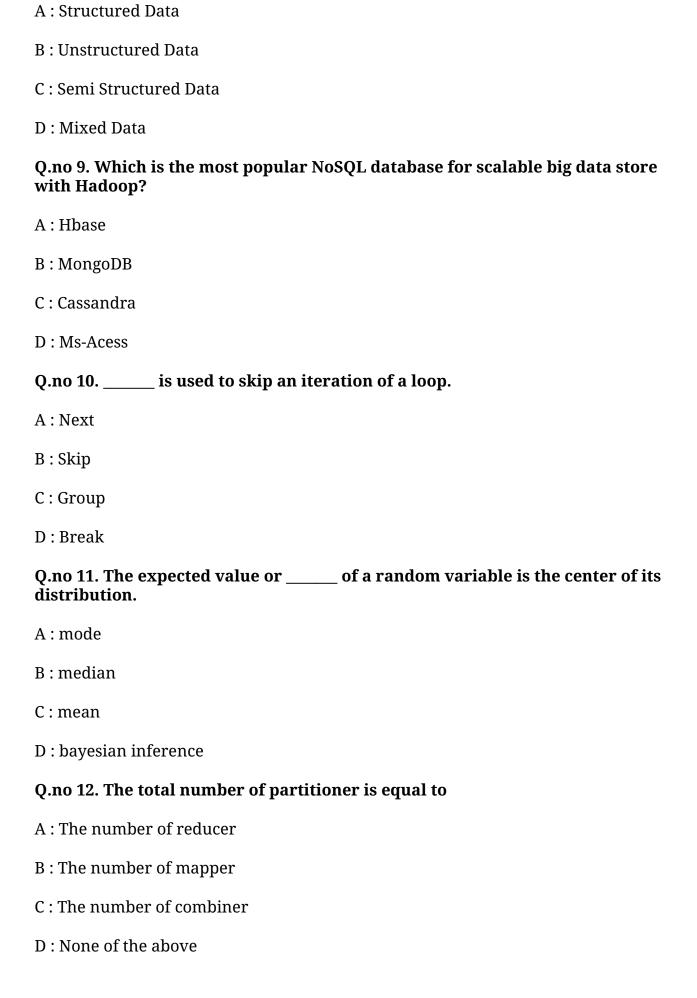B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 58. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 59. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 60. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 1. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 2. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 3. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 4. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 5. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 6. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 7. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 8. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 9. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 10. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 11. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 12. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 13. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 14. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 15. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 16. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 17. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 18. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 19. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 20. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 21. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 22. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 23. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 24. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 25. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 26. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 27. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum x\,P(x)$

D : 1

Q.no 28. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 29. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 30. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 − P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 31. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 32. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 33. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 34. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 35. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 36. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 37. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 38. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 39. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 40. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 41. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 42. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 43. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 44. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 45. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 46. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 47. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 48. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 49. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 50. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 51. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 52. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 53. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 54. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 55. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 56. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 57. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 58. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

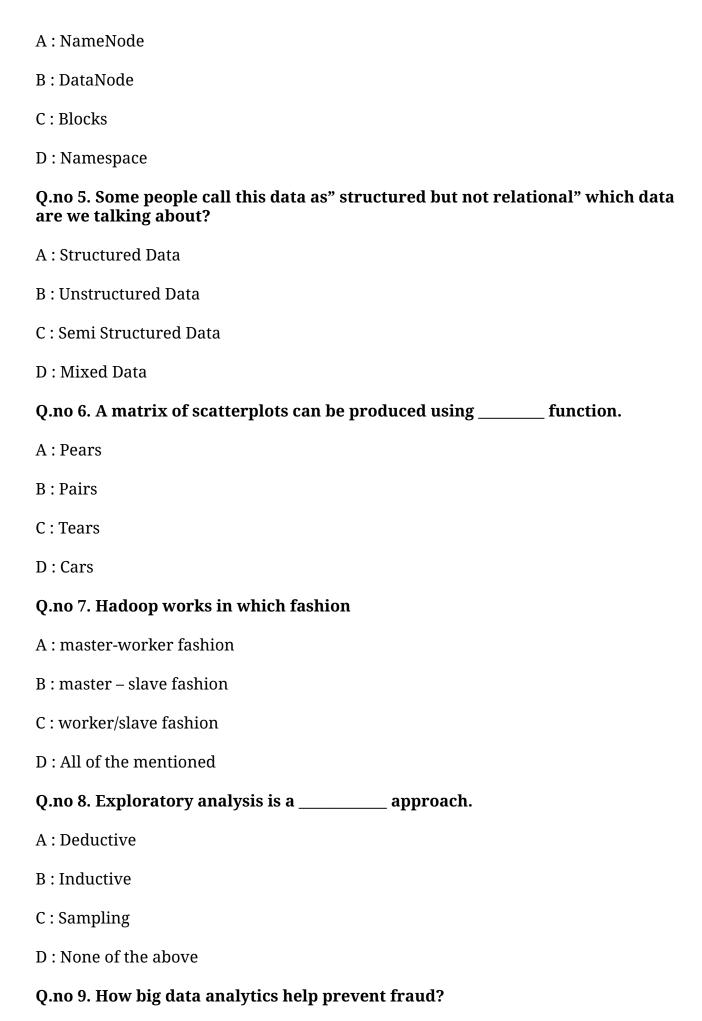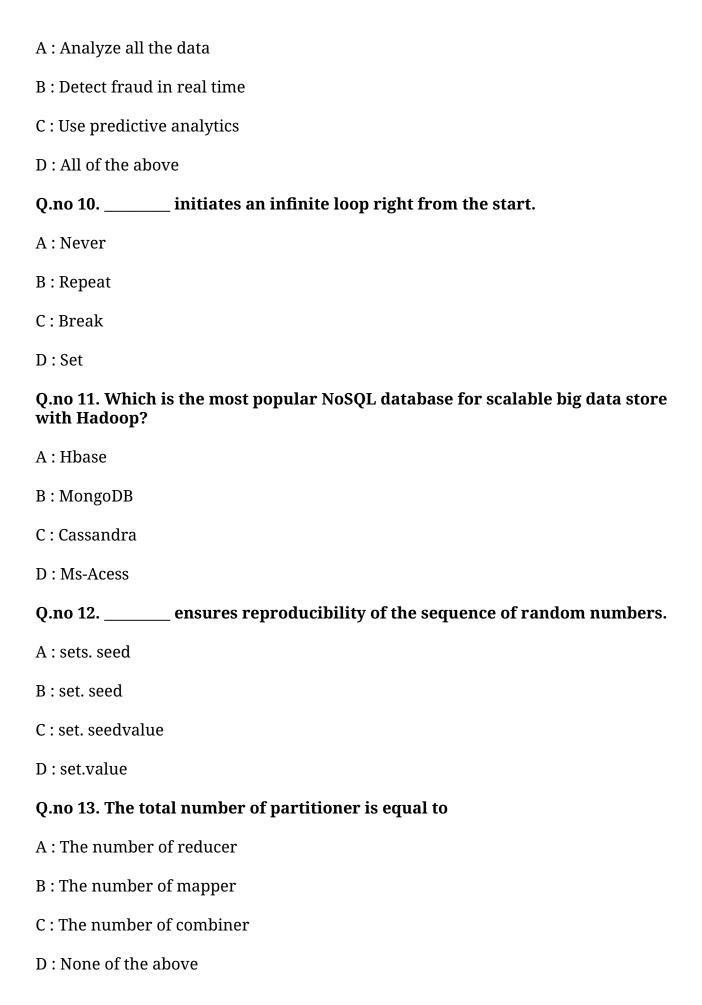D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 59. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2).........p(xn*xn-1)

Q.no 60. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 1. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 2. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 3. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 4. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 5. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 6. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 7. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 8. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 9. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 10. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 11. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 12. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 13. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 14. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 15. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 16. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 17. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 18. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 19. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 20. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 21. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 22. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 23. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 24. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 25. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 26. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 27. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 28. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 29. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 30. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 31. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 32. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 33. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 34. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 35. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 36. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 37. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 38. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 39. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 40. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 41. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 42. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 43. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 44. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 45. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 46. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 47. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 48. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 49. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 50. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 51. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 52. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 53. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 54. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 55. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 56. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 57. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 58. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 59. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 60. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 1. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 2. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 3. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 4. What are the 4 Vs of Big data ?

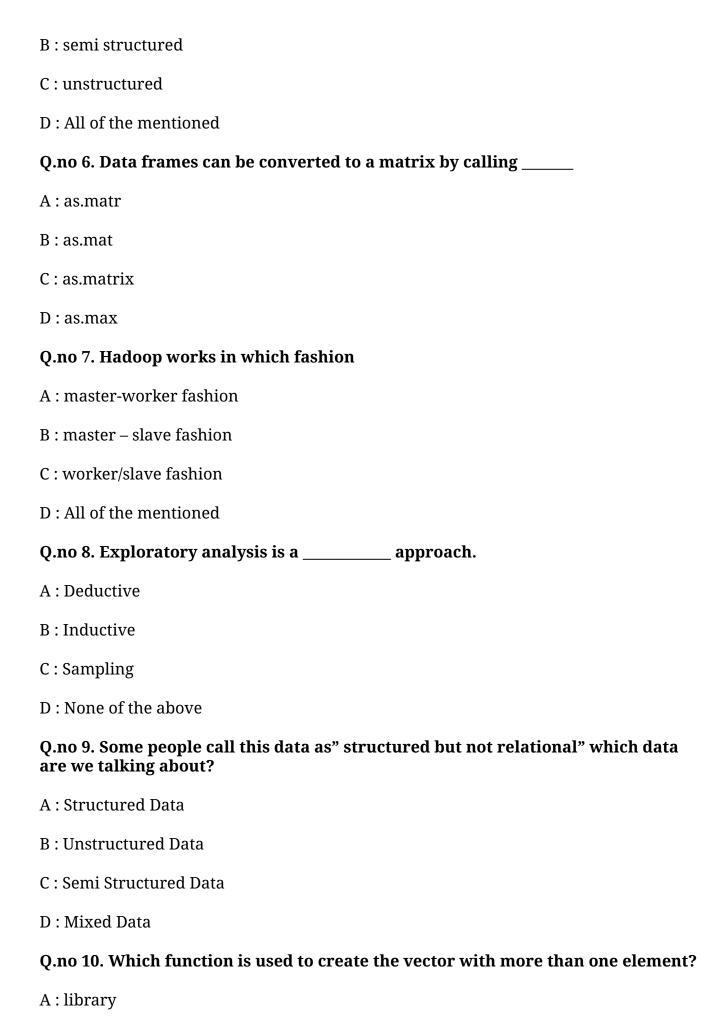A : Velocity, Volume, Veracity, Variety
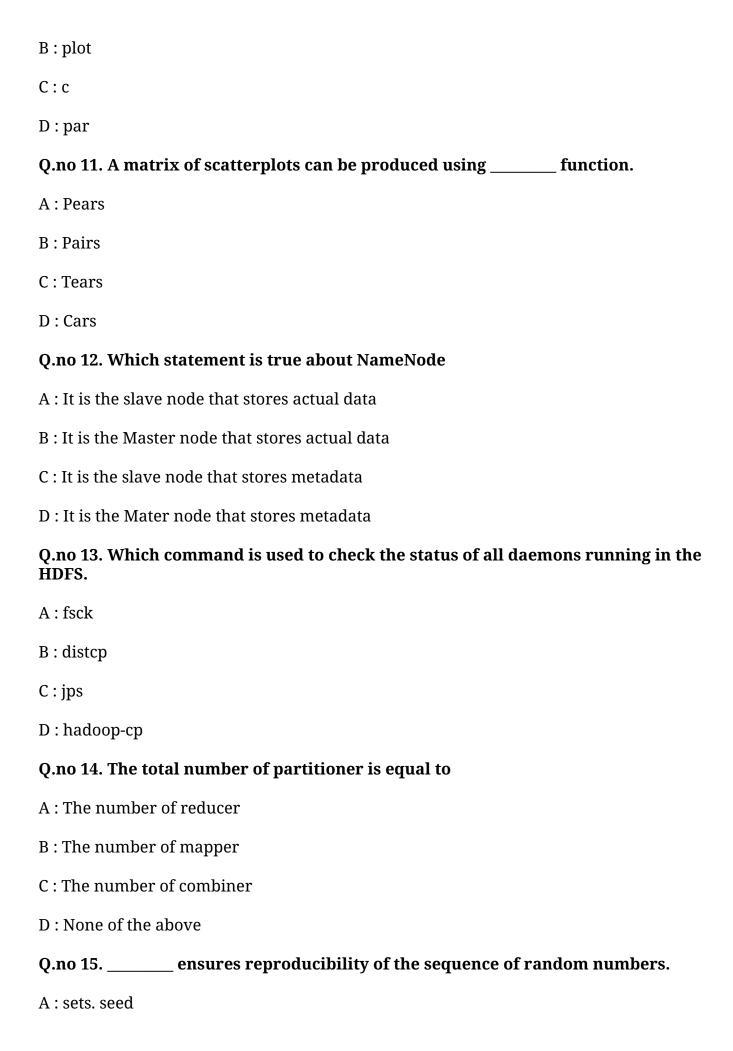
B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

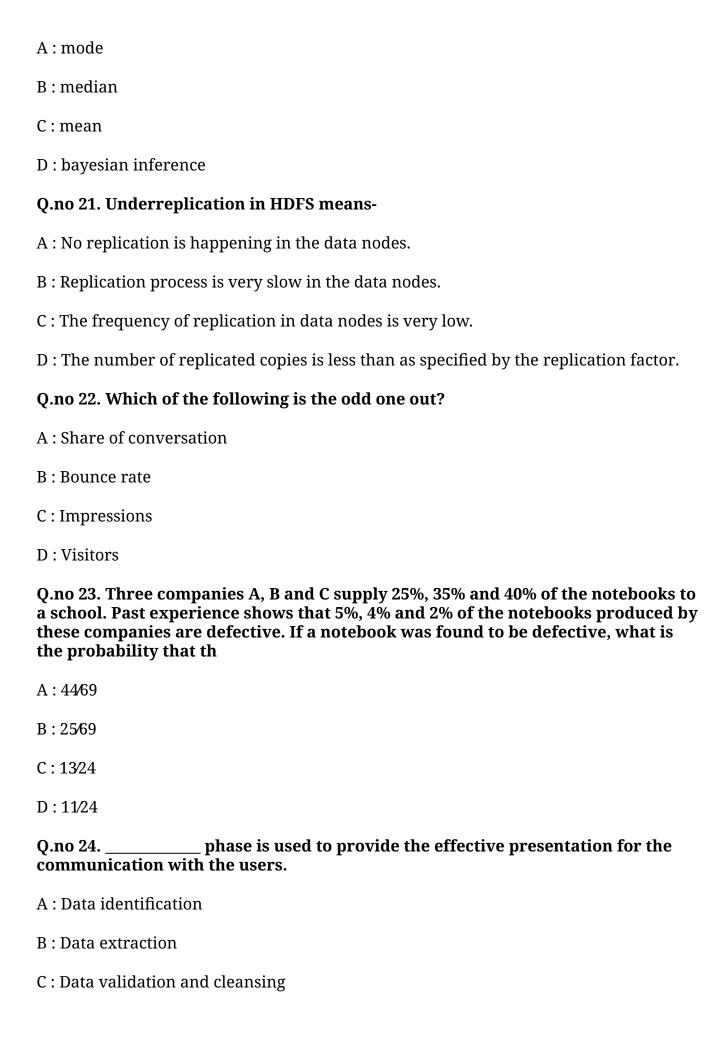**Q.no 5. Which type of data Hadoop can deal with is**

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 6. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 7. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 8. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 9. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 10. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 11. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 12. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 13. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 14. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 15. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 16. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 17. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 18. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 19. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 20. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 21. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 22. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 23. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44/69

B : 25/69

C : 13/24

D : 11/24

Q.no 24. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 25. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 26. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 27. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 28. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 29. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 30. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 31. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 32. The expected value of a discrete random variable 'x' is given by _____

A : $P(x)$

B : $\sum P(x)$

C : $\sum x P(x)$

D : 1

Q.no 33. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 34. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 35. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 36. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 37. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 38. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 39. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 40. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 41. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 42. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 43. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 44. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 45. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 46. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 47. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 48. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 49. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 50. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 51. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 52. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 53. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 54. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 55. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 56. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 57. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 58. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 59. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 60. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 1. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 2. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 3. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 4. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 5. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 6. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 7. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 8. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 9. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 10. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 11. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 12. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 13. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 14. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 15. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 16. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 17. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 18. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 19. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 20. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 21. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 22. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 23. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 24. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 25. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 26. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 27. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 28. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 29. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 30. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 31. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 32. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 33. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 34. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 35. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 36. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 37. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 38. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 39. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 40. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 41. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 42. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 43. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 44. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 45. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 46. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 47. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 48. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 49. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 50. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 51. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 52. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 53. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 54. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 55. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 56. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 57. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 58. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 59. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 60. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 1. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 2. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 3. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 4. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 5. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 6. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 7. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 8. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 9. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 10. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 11. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 12. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 13. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 14. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 15. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 16. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 17. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 18. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 19. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 20. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 21. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 22. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 23. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 24. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 25. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 26. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 27. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 28. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 29. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 30. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 31. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 32. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 33. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 34. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 35. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : $P(G) = 1 / P(H)$

B : $P(G) = 1 - P(H$

C : $P(G) = 1 + P(H)$

D : $P(G) = P(H)$

Q.no 36. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 37. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 38. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 39. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 40. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.
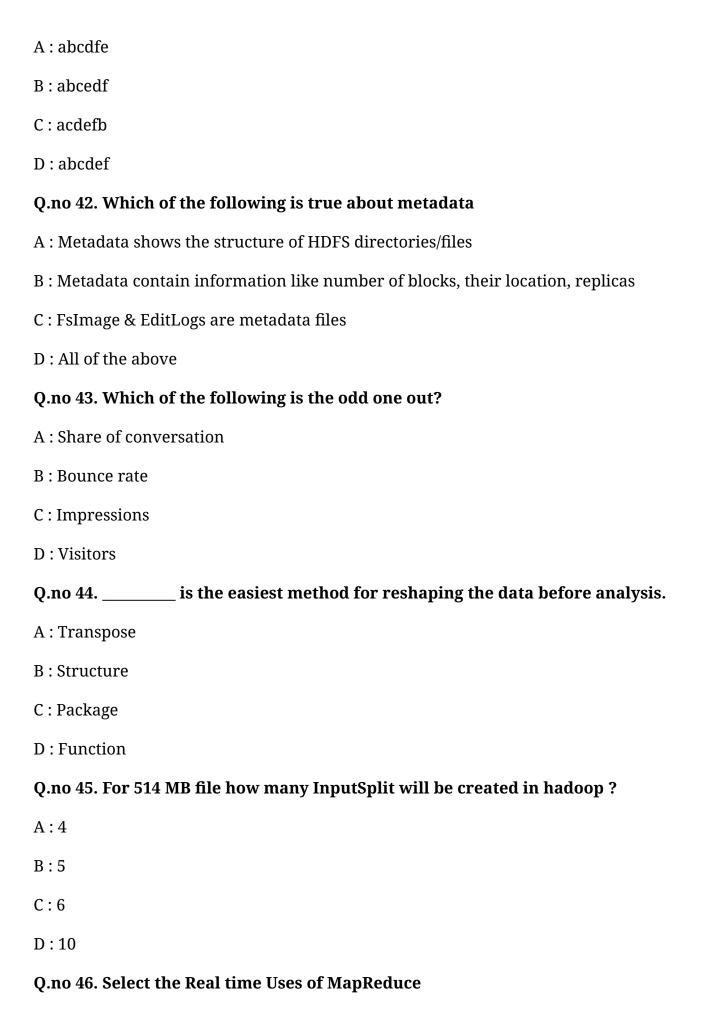
A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 41. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 42. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 43. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 44. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 45. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 46. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 47. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 48. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 49. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 50. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 51. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 52. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 53. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 54. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 55. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 56. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 57. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 58. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 59. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
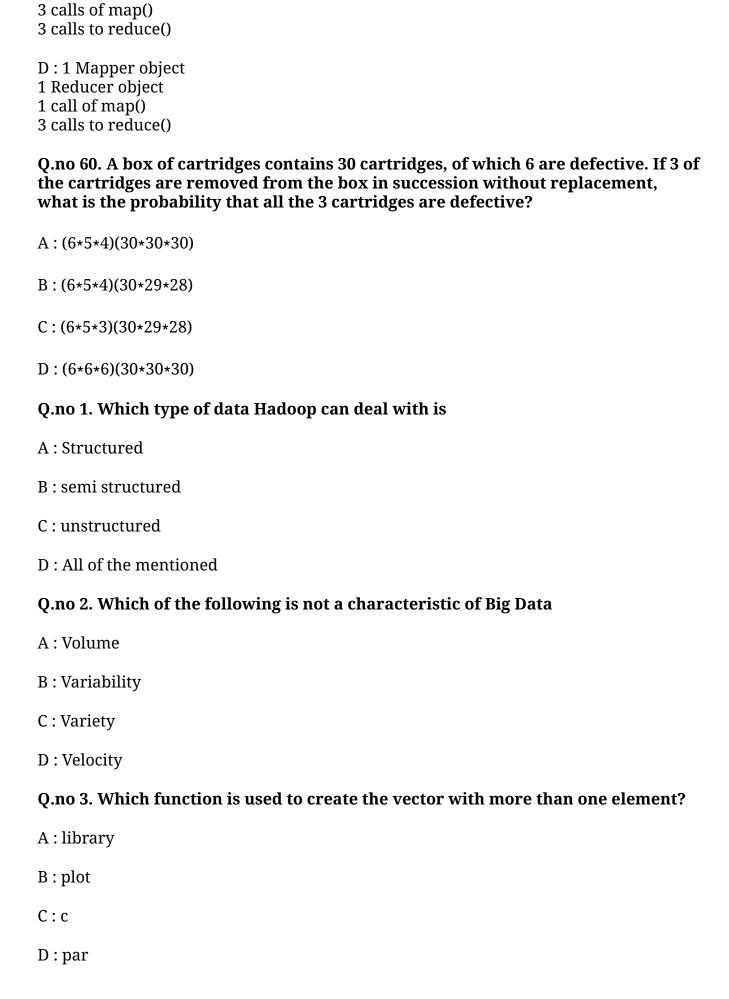1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects

3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 60. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 1. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 2. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 3. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 4. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 5. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 6. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 7. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 8. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 9. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 10. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 11. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 12. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 13. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 14. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 15. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 16. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 17. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 18. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 19. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 20. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 21. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 22. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 23. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 24. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 25. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 26. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 27. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 28. Which of the following is a plot for single variable?
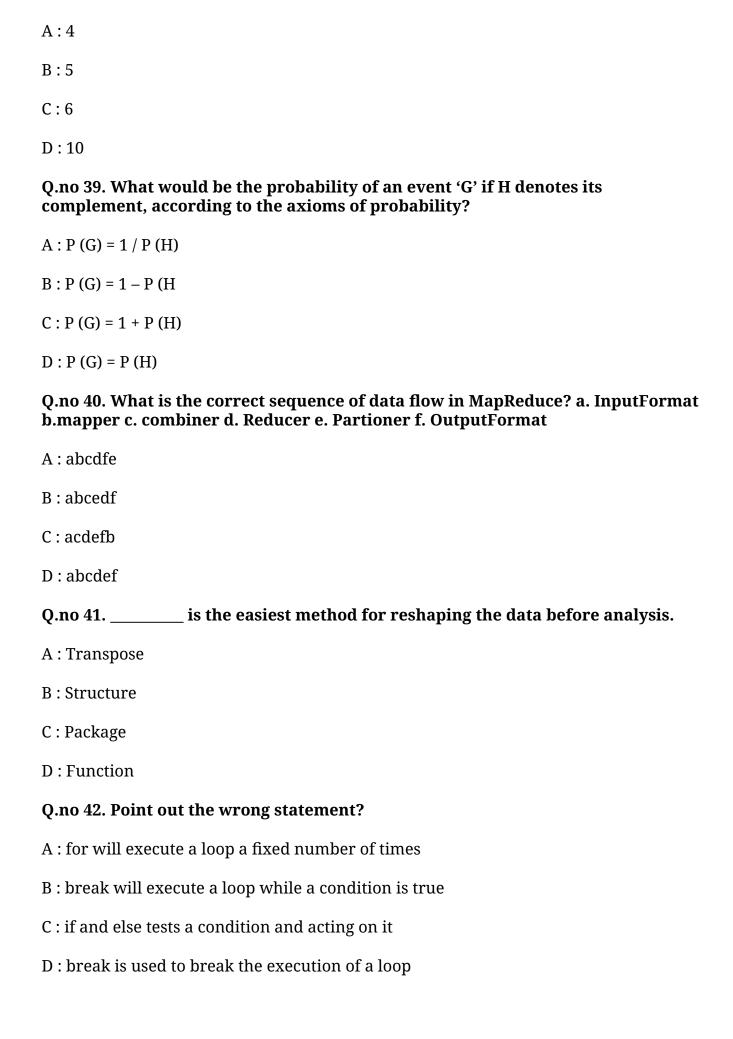
A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 29. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 30. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 31. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 32. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 33. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 34. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 35. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 36. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 37. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 38. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 39. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 40. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 41. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 42. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 43. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 44. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 45. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 46. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 47. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 48. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 49. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 50. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 51. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 52. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 53. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2).........p(xn*xn-1)

Q.no 54. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 55. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 56. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 57. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 58. ----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 59. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 60. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 1. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 2. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 3. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 4. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 5. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 6. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 7. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 8. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 9. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 10. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 11. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 12. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 13. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 14. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 15. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 16. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 17. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 18. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 19. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 20. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 21. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 22. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 23. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 24. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 25. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 26. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 27. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 28. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 29. The expected value of a discrete random variable 'x' is given by _____

A : $P(x)$

B : $\sum P(x)$

C : $\sum x\, P(x)$

D : 1

Q.no 30. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 31. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 32. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 33. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 34. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.
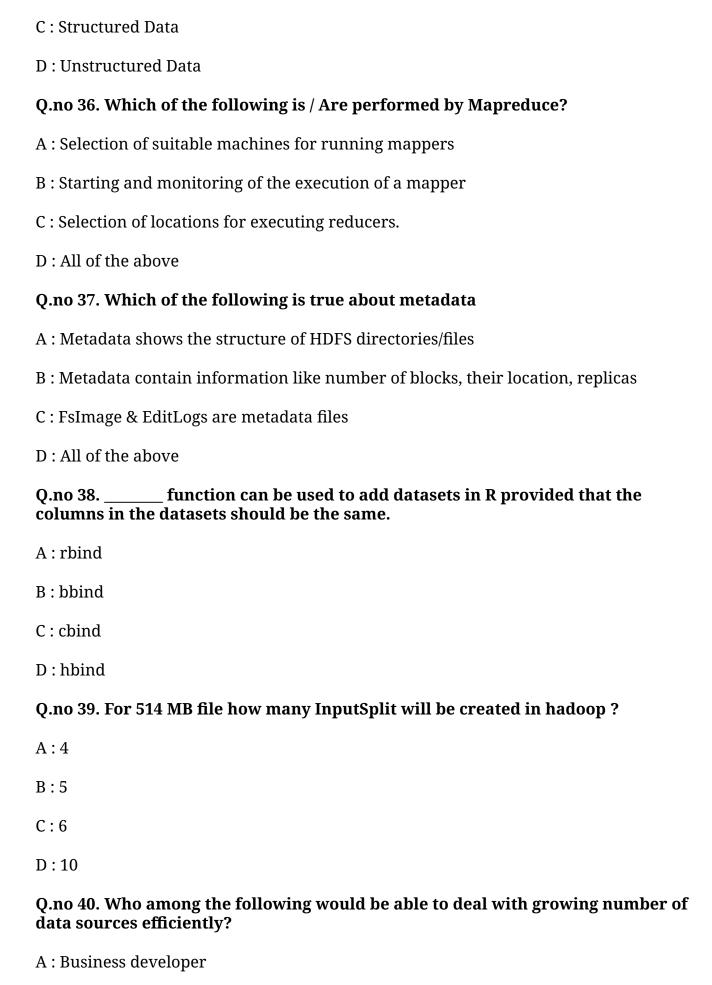
C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 35. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 36. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 37. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 38. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 39. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 40. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 41. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 42. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 43. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 44. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 45. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 46. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 47. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 48. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 49. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 50. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 51. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 52. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 53. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 54. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 55. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 56. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 57. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 58. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 59. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 60. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 1. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 2. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 3. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 4. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 5. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 6. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 7. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 8. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 9. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 10. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 11. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 12. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 13. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 14. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 15. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 16. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 17. The expected value or _____ of a random variable is the center of its distribution.
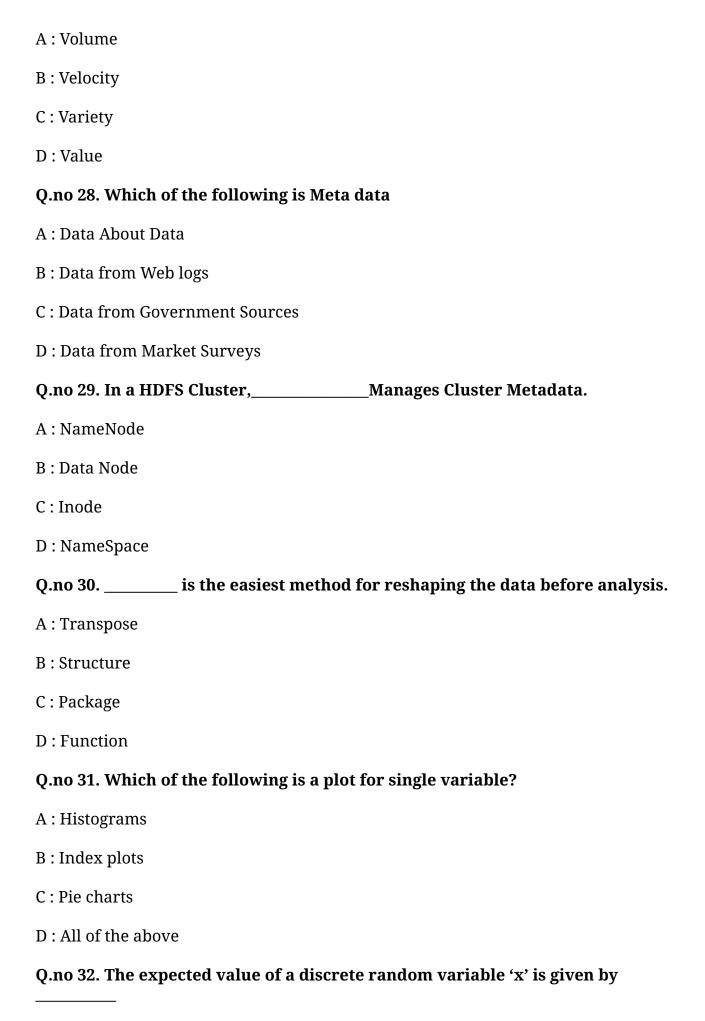
A : mode

B : median

C : mean

D : bayesian inference

Q.no 18. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 19. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 20. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 21. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 22. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 23. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 24. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 25. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 26. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 27. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 28. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 29. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 30. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 31. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 32. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : ∑ P(x)

C : ∑ x P(x)

D : 1

Q.no 33. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 34. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 35. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 36. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 37. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 38. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 39. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 40. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files
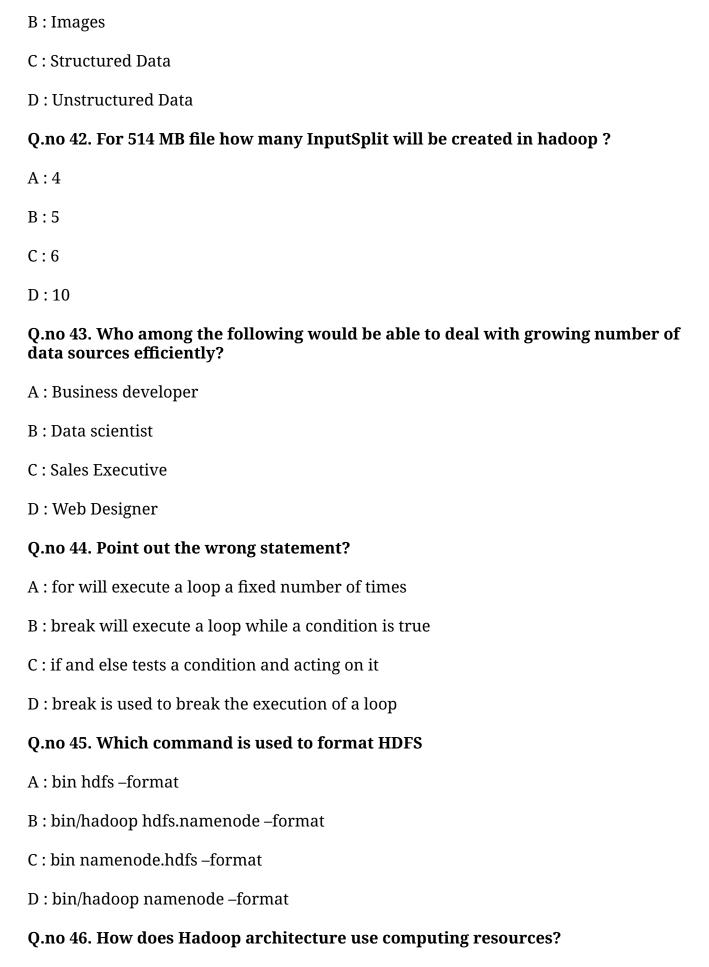
B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 41. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 42. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 43. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 44. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 45. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 46. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 47. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 48. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 49. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 50. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 51. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 52. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 53. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2).........p(xn*xn-1)

Q.no 54. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 55. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 56. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 57. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 58. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 59. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects

3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 60. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 1. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 2. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 3. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 4. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 5. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 6. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 7. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 8. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 9. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 10. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 11. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 12. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 13. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 14. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 15. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 16. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 17. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 18. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 19. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 20. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 21. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 22. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 23. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 24. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 25. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 26. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 27. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 28. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 29. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 30. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 31. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 32. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 33. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 34. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 35. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 36. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 37. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 38. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 39. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 40. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 41. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 42. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 43. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 44. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 45. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 46. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 47. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 48. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 49. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 50. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 51. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 52. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 53. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 54. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 55. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 56. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 57. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 58. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 59. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 60. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 1. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 2. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 3. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 4. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 5. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 6. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 7. Which function is used to create the vector with more than one element?

A : library

B : plot

C : c

D : par

Q.no 8. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 9. Which type of data Hadoop can deal with is

A : Structured

B : semi structured

C : unstructured

D : All of the mentioned

Q.no 10. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 11. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 12. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 13. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 14. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 15. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 16. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 17. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 18. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 19. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 20. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 21. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 22. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 23. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 24. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 25. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 26. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 27. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 28. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 29. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 30. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 31. Put the following phases of a MapReduce program in the order that they execute? a. Partitionor b. Mapper c. Combiner d. Shuffle/sort

A : Mapper Partitioner Shuffle/Sort Combiner

B : Mapper Partitioner Combiner Shuffle/Sort

C : Mapper Shuffle/Sort Combiner Partitioner

D : Mapper Combiner Partitioner Shuffle/Sort

Q.no 32. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 33. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 34. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 35. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 36. _____ phase is used to provide the effective presentation for the communication with the users.

A : Data identification

B : Data extraction

C : Data validation and cleansing

D : Data visualization

Q.no 37. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 38. Which of the following is managed by mapreduce environment?

A : Web logs

B : Images

C : Structured Data

D : Unstructured Data

Q.no 39. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 40. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 41. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 42. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum P(x)$

C : $\sum x\, P(x)$

D : 1

Q.no 43. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 44. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 45. _____ is the easiest method for reshaping the data before analysis.

A : Transpose

B : Structure

C : Package

D : Function

Q.no 46. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 47. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 48. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 49. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 50. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 51. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 52. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 53. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 54. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 55. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 56. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object

1 call of map()
3 calls to reduce()

Q.no 57. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 58. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 59. A box of cartridges contains 30 cartridges, of which 6 are defective. If 3 of the cartridges are removed from the box in succession without replacement, what is the probability that all the 3 cartridges are defective?

A : (6*5*4)(30*30*30)

B : (6*5*4)(30*29*28)

C : (6*5*3)(30*29*28)

D : (6*6*6)(30*30*30)

Q.no 60. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Q.no 1. What is the default size of HDFS Data Block?

A : 16MB

B : 32MB

C : 64MB

D : 128MB

Q.no 2. Which of the following is not a characteristic of Big Data

A : Volume

B : Variability

C : Variety

D : Velocity

Q.no 3. Exploratory analysis is a _____ approach.

A : Deductive

B : Inductive

C : Sampling

D : None of the above

Q.no 4. Which is the most popular NoSQL database for scalable big data store with Hadoop?

A : Hbase

B : MongoDB

C : Cassandra

D : Ms-Acess

Q.no 5. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.

A : MapReduce

B : Mapper

C : TaskTracker

D : JobTracker

Q.no 6. Which statement is true about NameNode

A : It is the slave node that stores actual data

B : It is the Master node that stores actual data

C : It is the slave node that stores metadata

D : It is the Mater node that stores metadata

Q.no 7. _____ is used to skip an iteration of a loop.

A : Next

B : Skip

C : Group

D : Break

Q.no 8. What are the 4 Vs of Big data ?

A : Velocity, Volume, Veracity, Variety

B : Velocity, Validity, Valued,Vitalize

C : Visit, Volume, Validate, Valued

D : None of the Above

Q.no 9. The total number of partitioner is equal to

A : The number of reducer

B : The number of mapper

C : The number of combiner

D : None of the above

Q.no 10. Hadoop works in which fashion

A : master-worker fashion

B : master – slave fashion

C : worker/slave fashion

D : All of the mentioned

Q.no 11. _____ ensures reproducibility of the sequence of random numbers.

A : sets. seed

B : set. seed

C : set. seedvalue

D : set.value

Q.no 12. _____ initiates an infinite loop right from the start.

A : Never

B : Repeat

C : Break

D : Set

Q.no 13. Which of the following terms is used to denote the small subsets of a large file created by HDFS?

A : NameNode

B : DataNode

C : Blocks

D : Namespace

Q.no 14. The expected value or _____ of a random variable is the center of its distribution.

A : mode

B : median

C : mean

D : bayesian inference

Q.no 15. How big data analytics help prevent fraud?

A : Analyze all the data

B : Detect fraud in real time

C : Use predictive analytics

D : All of the above

Q.no 16. Data frames can be converted to a matrix by calling _____

A : as.matr

B : as.mat

C : as.matrix

D : as.max

Q.no 17. Some people call this data as" structured but not relational" which data are we talking about?

A : Structured Data

B : Unstructured Data

C : Semi Structured Data

D : Mixed Data

Q.no 18. A matrix of scatterplots can be produced using _____ function.

A : Pears

B : Pairs

C : Tears

D : Cars

Q.no 19. Which command is used to check the status of all daemons running in the HDFS.

A : fsck

B : distcp

C : jps

D : hadoop-cp

Q.no 20. Which one is not an example of External data source

A : Data from CRM

B : Data from Web logs

C : Data from Government Sources

D : Data from Market surveys

Q.no 21. Which of the following is the odd one out?

A : Share of conversation

B : Bounce rate

C : Impressions

D : Visitors

Q.no 22. Which of the following is Meta data

A : Data About Data

B : Data from Web logs

C : Data from Government Sources

D : Data from Market Surveys

Q.no 23. Which ONE of the following is mainly used in Web Analytics and is free of charge?

A : Google Analytics

B : Radian6

C : AlteranSM2

D : Social Radar

Q.no 24. Which of the following is / Are performed by Mapreduce?

A : Selection of suitable machines for running mappers

B : Starting and monitoring of the execution of a mapper

C : Selection of locations for executing reducers.

D : All of the above

Q.no 25. What would be the probability of an event 'G' if H denotes its complement, according to the axioms of probability?

A : P (G) = 1 / P (H)

B : P (G) = 1 – P (H

C : P (G) = 1 + P (H)

D : P (G) = P (H)

Q.no 26. Which of the following statement is correct

A : Data locality means moving computation to data instead of data to computation

B : Data locality means moving data to computation instead of computation to data

C : Both the above

D : copying data locally to nodes

Q.no 27. Which command is used to format HDFS

A : bin hdfs –format

B : bin/hadoop hdfs.namenode –format

C : bin namenode.hdfs –format

D : bin/hadoop namenode –format

Q.no 28. --------- plot adds a third dimension to the plot where a third variable is mapped to the size of the points.

A : Scatter plot matrix

B : Design plot

C : Bubble plot

D : Histogram

Q.no 29. Which of the following is performed by Data Scientist?

A : Define the question

B : Create reproducible code

C : Challenge results

D : All of the mentioned

Q.no 30. Underreplication in HDFS means-

A : No replication is happening in the data nodes.

B : Replication process is very slow in the data nodes.

C : The frequency of replication in data nodes is very low.

D : The number of replicated copies is less than as specified by the replication factor.

Q.no 31. Which of the following is one of the key data science skills?

A : Statistics

B : Machine Learning

C : Data Visualization

D : All of the mentioned

Q.no 32. Who among the following would be able to deal with growing number of data sources efficiently?

A : Business developer

B : Data scientist

C : Sales Executive

D : Web Designer

Q.no 33. What is the correct sequence of data flow in MapReduce? a. InputFormat b.mapper c. combiner d. Reducer e. Partioner f. OutputFormat

A : abcdfe

B : abcedf

C : acdefb

D : abcdef

Q.no 34. The Data being captured can be in any form or structure. Which characteristics of big data are we talking about?

A : Volume

B : Velocity

C : Variety

D : Value

Q.no 35. Point out the wrong statement?

A : for will execute a loop a fixed number of times

B : break will execute a loop while a condition is true

C : if and else tests a condition and acting on it

D : break is used to break the execution of a loop

Q.no 36. Three companies A, B and C supply 25%, 35% and 40% of the notebooks to a school. Past experience shows that 5%, 4% and 2% of the notebooks produced by these companies are defective. If a notebook was found to be defective, what is the probability that th

A : 44⁄69

B : 25⁄69

C : 13⁄24

D : 11⁄24

Q.no 37. In a HDFS Cluster,_____Manages Cluster Metadata.

A : NameNode

B : Data Node

C : Inode

D : NameSpace

Q.no 38. The expected value of a discrete random variable 'x' is given by _____

A : P(x)

B : $\sum$ P(x)

C : $\sum$ x P(x)

D : 1

Q.no 39. _____ variables are categorical variables which can hold either string or numeric values.

A : Factor

B : Simpler

C : Function

D : Package

Q.no 40. For 514 MB file how many InputSplit will be created in hadoop ?

A : 4

B : 5

C : 6

D : 10

Q.no 41. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

A : independent probabilities

B : posterior probabilities

C : interior probabilities

D : dependent probabilities

Q.no 42. Which of the following is true about metadata

A : Metadata shows the structure of HDFS directories/files

B : Metadata contain information like number of blocks, their location, replicas

C : FsImage & EditLogs are metadata files

D : All of the above

Q.no 43. Which of the following is a plot for single variable?

A : Histograms

B : Index plots

C : Pie charts

D : All of the above

Q.no 44. _____is a open source framework that enables you to store large volumes of data in a distributed manner across multiple machines

A : Hadoop

B : Hive

C : Pig

D : ZooKeeper

Q.no 45. _____ function can be used to add datasets in R provided that the columns in the datasets should be the same.

A : rbind

B : bbind

C : cbind

D : hbind

Q.no 46. You have been assigned the task of reshaping the data wherein you have to convert the wide format data into long format data and vice versa. How will you carry out this operation?

A : melt and dcast

B : dcast and melt

C : melt and rbind

D : dcast and rbind

Q.no 47. How does Hadoop architecture use computing resources?

A : By distributing software to computing resources

B : By distributing data and computing tasks to computing resources

C : By creating shared memory for computing resources

D : By distributing data to computing resources

Q.no 48. Point out the correct statement

A : Hive is not a relational database, but a query engine that supports the parts of SQL.

B : Hive is a relational database with SQL support.

C : Both of the above

D : None of the above

Q.no 49. Which ONE of the following is based on user-generated media, mainly investigating earned media?

A : Web counters

B : Web site analytics

C : Social media monitoring

D : Social Media ROI

Q.no 50. Which of the following is not an example of NoSQL Databases?

A : Hbase

B : MangoDB

C : Allegrograph

D : Oracle

Q.no 51. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A : Regression

B : Clustering

C : Summarization

D : Analytics

Q.no 52. Test How many phases exist in MapReduce?

A : 4

B : 5

C : 6

D : 2

Q.no 53. The objectives for web analytics are likely to concern:

A : Facebook messages

B : Personal Blog activity

C : Social Media ROI

D : Measurement of web site performance

Q.no 54. Hadoop makes system more resilient by

A : Using an effective firewall and antivirus

B : Keeping multiple copies of data

C : Keeping each computing resource isolated

D : Uploading Data to a cloud for back up

Q.no 55. -----------allows to add an additional value in the function without declaring an additional argument.

A : Anonymous function

B : dot argument

C : Optional argument

D : None of the above

Q.no 56. The probability density function of a Markov process is

A : p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

B : p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

C : p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

D : p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

Q.no 57. Consider Hadoop's WordCount program: for a given text, compute the frequency of each word in it. The input is read line by line. As input, you are given one le that contains a single line of text: A Ram Sam Sam How many Mapper objects and Reducer Ob

A : 3 Mapper objects
1 Reducer object
3 calls of map()
1 calls to reduce()

B : 3 Mapper objects
3 Reducer objects,
1 call of map()
1 call to reduce()

C : 1 Mapper object
3 Reducer objects
3 calls of map()
3 calls to reduce()

D : 1 Mapper object
1 Reducer object
1 call of map()
3 calls to reduce()

Q.no 58. Select Which is correct about Hadoop?

A : It is an in-memory tool used in Mahout algorithm computing

B : It is a computing framework used for resource management.

C : It is a framework that allows for distributed processing of large datasets across clusters of commodity computers using a simple programming model

D : It is a search and analytics tool that provides access to analyze data.

Q.no 59. Select the Real time Uses of MapReduce

A : Algorithms,Sorting,Data Mining

B : Search Engine Operations,Enterprise Analytics

C : Gaussian Analysis,Semantic Web 3.0

D : All of these

Q.no 60. The Data generated from a GPS Satellite and Web Logs is classified as _____

A : Structured Data

B : Unstructured Data

C : Both Structured and Unstructured data

D : Semi-Structured Data

Answer for Question No 1. is c

Answer for Question No 2. is b

Answer for Question No 3. is a

Answer for Question No 4. is d

Answer for Question No 5. is b

Answer for Question No 6. is c

Answer for Question No 7. is c

Answer for Question No 8. is b

Answer for Question No 9. is b

Answer for Question No 10. is a

Answer for Question No 11. is c

Answer for Question No 12. is c

Answer for Question No 13. is a

Answer for Question No 14. is b

Answer for Question No 15. is a

Answer for Question No 16. is b

Answer for Question No 17. is a

Answer for Question No 18. is c

Answer for Question No 19. is d

Answer for Question No 20. is d

Answer for Question No 21. is a

Answer for Question No 22. is a

Answer for Question No 23. is d

Answer for Question No 24. is b

Answer for Question No 25. is d

Answer for Question No 26. is b

Answer for Question No 27. is c

Answer for Question No 28. is d

Answer for Question No 29. is d

Answer for Question No 30. is b

Answer for Question No 31. is b

Answer for Question No 32. is a

Answer for Question No 33. is a

Answer for Question No 34. is d

Answer for Question No 35. is a

Answer for Question No 36. is a

Answer for Question No 37. is b

Answer for Question No 38. is d

Answer for Question No 39. is a

Answer for Question No 40. is c

Answer for Question No 41. is d

Answer for Question No 42. is a

Answer for Question No 43. is c

Answer for Question No 44. is d

Answer for Question No 45. is d

Answer for Question No 46. is d

Answer for Question No 47. is c

Answer for Question No 48. is d

Answer for Question No 49. is d

Answer for Question No 50. is a

Answer for Question No 51. is d

Answer for Question No 52. is a

Answer for Question No 53. is d

Answer for Question No 54. is a

Answer for Question No 55. is b

Answer for Question No 56. is c

Answer for Question No 57. is c

Answer for Question No 58. is a

Answer for Question No 59. is c

Answer for Question No 60. is d

Answer for Question No 1. is c

Answer for Question No 2. is b

Answer for Question No 3. is a

Answer for Question No 4. is a

Answer for Question No 5. is c

Answer for Question No 6. is d

Answer for Question No 7. is b

Answer for Question No 8. is c

Answer for Question No 9. is b

Answer for Question No 10. is b

Answer for Question No 11. is a

Answer for Question No 12. is c

Answer for Question No 13. is a

Answer for Question No 14. is b

Answer for Question No 15. is b

Answer for Question No 16. is d

Answer for Question No 17. is b

Answer for Question No 18. is d

Answer for Question No 19. is a

Answer for Question No 20. is c

Answer for Question No 21. is b

Answer for Question No 22. is d

Answer for Question No 23. is a

Answer for Question No 24. is b

Answer for Question No 25. is a

Answer for Question No 26. is c

Answer for Question No 27. is d

Answer for Question No 28. is c

Answer for Question No 29. is a

Answer for Question No 30. is d

Answer for Question No 31. is a

Answer for Question No 32. is b

Answer for Question No 33. is c

Answer for Question No 34. is d

Answer for Question No 35. is d

Answer for Question No 36. is b

Answer for Question No 37. is d

Answer for Question No 38. is d

Answer for Question No 39. is a

Answer for Question No 40. is b

Answer for Question No 41. is a

Answer for Question No 42. is b

Answer for Question No 43. is d

Answer for Question No 44. is d

Answer for Question No 45. is a

Answer for Question No 46. is b

Answer for Question No 47. is a

Answer for Question No 48. is c

Answer for Question No 49. is a

Answer for Question No 50. is d

Answer for Question No 51. is a

Answer for Question No 52. is c

Answer for Question No 53. is c

Answer for Question No 54. is d

Answer for Question No 55. is d

Answer for Question No 56. is a

Answer for Question No 57. is d

Answer for Question No 58. is c

Answer for Question No 59. is d

Answer for Question No 60. is d

Answer for Question No 1. is c

Answer for Question No 2. is c

Answer for Question No 3. is b

Answer for Question No 4. is c

Answer for Question No 5. is b

Answer for Question No 6. is a

Answer for Question No 7. is d

Answer for Question No 8. is a

Answer for Question No 9. is b

Answer for Question No 10. is d

Answer for Question No 11. is c

Answer for Question No 12. is c

Answer for Question No 13. is a

Answer for Question No 14. is d

Answer for Question No 15. is b

Answer for Question No 16. is a

Answer for Question No 17. is b

Answer for Question No 18. is c

Answer for Question No 19. is b

Answer for Question No 20. is c

Answer for Question No 21. is b

Answer for Question No 22. is b

Answer for Question No 23. is d

Answer for Question No 24. is a

Answer for Question No 25. is a

Answer for Question No 26. is b

Answer for Question No 27. is a

Answer for Question No 28. is a

Answer for Question No 29. is b

Answer for Question No 30. is d

Answer for Question No 31. is b

Answer for Question No 32. is d

Answer for Question No 33. is d

Answer for Question No 34. is d

Answer for Question No 35. is a

Answer for Question No 36. is a

Answer for Question No 37. is c

Answer for Question No 38. is d

Answer for Question No 39. is d

Answer for Question No 40. is a

Answer for Question No 41. is a

Answer for Question No 42. is a

Answer for Question No 43. is d

Answer for Question No 44. is b

Answer for Question No 45. is c

Answer for Question No 46. is b

Answer for Question No 47. is b

Answer for Question No 48. is c

Answer for Question No 49. is d

Answer for Question No 50. is a

Answer for Question No 51. is a

Answer for Question No 52. is d

Answer for Question No 53. is d

Answer for Question No 54. is d

Answer for Question No 55. is c

Answer for Question No 56. is c

Answer for Question No 57. is c

Answer for Question No 58. is d

Answer for Question No 59. is d

Answer for Question No 60. is a

Answer for Question No 1. is b

Answer for Question No 2. is c

Answer for Question No 3. is b

Answer for Question No 4. is a

Answer for Question No 5. is c

Answer for Question No 6. is c

Answer for Question No 7. is d

Answer for Question No 8. is a

Answer for Question No 9. is c

Answer for Question No 10. is c

Answer for Question No 11. is a

Answer for Question No 12. is a

Answer for Question No 13. is b

Answer for Question No 14. is d

Answer for Question No 15. is d

Answer for Question No 16. is a

Answer for Question No 17. is b

Answer for Question No 18. is c

Answer for Question No 19. is c

Answer for Question No 20. is b

Answer for Question No 21. is d

Answer for Question No 22. is d

Answer for Question No 23. is c

Answer for Question No 24. is a

Answer for Question No 25. is b

Answer for Question No 26. is b

Answer for Question No 27. is d

Answer for Question No 28. is b

Answer for Question No 29. is a

Answer for Question No 30. is d

Answer for Question No 31. is d

Answer for Question No 32. is b

Answer for Question No 33. is a

Answer for Question No 34. is a

Answer for Question No 35. is d

Answer for Question No 36. is a

Answer for Question No 37. is d

Answer for Question No 38. is c

Answer for Question No 39. is d

Answer for Question No 40. is b

Answer for Question No 41. is a

Answer for Question No 42. is a

Answer for Question No 43. is a

Answer for Question No 44. is b

Answer for Question No 45. is d

Answer for Question No 46. is b

Answer for Question No 47. is a

Answer for Question No 48. is d

Answer for Question No 49. is d

Answer for Question No 50. is a

Answer for Question No 51. is c

Answer for Question No 52. is a

Answer for Question No 53. is c

Answer for Question No 54. is d

Answer for Question No 55. is b

Answer for Question No 56. is d

Answer for Question No 57. is d

Answer for Question No 58. is d

Answer for Question No 59. is a

Answer for Question No 60. is c

Answer for Question No 1. is a

Answer for Question No 2. is b

Answer for Question No 3. is c

Answer for Question No 4. is b

Answer for Question No 5. is b

Answer for Question No 6. is c

Answer for Question No 7. is b

Answer for Question No 8. is c

Answer for Question No 9. is a

Answer for Question No 10. is a

Answer for Question No 11. is c

Answer for Question No 12. is a

Answer for Question No 13. is d

Answer for Question No 14. is c

Answer for Question No 15. is c

Answer for Question No 16. is b

Answer for Question No 17. is b

Answer for Question No 18. is a

Answer for Question No 19. is b

Answer for Question No 20. is c

Answer for Question No 21. is b

Answer for Question No 22. is d

Answer for Question No 23. is d

Answer for Question No 24. is c

Answer for Question No 25. is a

Answer for Question No 26. is d

Answer for Question No 27. is c

Answer for Question No 28. is a

Answer for Question No 29. is b

Answer for Question No 30. is b

Answer for Question No 31. is d

Answer for Question No 32. is d

Answer for Question No 33. is b

Answer for Question No 34. is b

Answer for Question No 35. is c

Answer for Question No 36. is a

Answer for Question No 37. is a

Answer for Question No 38. is b

Answer for Question No 39. is d

Answer for Question No 40. is d

Answer for Question No 41. is a

Answer for Question No 42. is d

Answer for Question No 43. is a

Answer for Question No 44. is d

Answer for Question No 45. is a

Answer for Question No 46. is a

Answer for Question No 47. is c

Answer for Question No 48. is b

Answer for Question No 49. is d

Answer for Question No 50. is d

Answer for Question No 51. is d

Answer for Question No 52. is a

Answer for Question No 53. is a

Answer for Question No 54. is d

Answer for Question No 55. is b

Answer for Question No 56. is c

Answer for Question No 57. is c

Answer for Question No 58. is d

Answer for Question No 59. is a

Answer for Question No 60. is d

Answer for Question No 1. is c

Answer for Question No 2. is a

Answer for Question No 3. is d

Answer for Question No 4. is c

Answer for Question No 5. is c

Answer for Question No 6. is b

Answer for Question No 7. is b

Answer for Question No 8. is b

Answer for Question No 9. is d

Answer for Question No 10. is b

Answer for Question No 11. is a

Answer for Question No 12. is b

Answer for Question No 13. is a

Answer for Question No 14. is c

Answer for Question No 15. is b

Answer for Question No 16. is c

Answer for Question No 17. is d

Answer for Question No 18. is c

Answer for Question No 19. is a

Answer for Question No 20. is b

Answer for Question No 21. is d

Answer for Question No 22. is b

Answer for Question No 23. is a

Answer for Question No 24. is c

Answer for Question No 25. is c

Answer for Question No 26. is a

Answer for Question No 27. is d

Answer for Question No 28. is d

Answer for Question No 29. is b

Answer for Question No 30. is d

Answer for Question No 31. is a

Answer for Question No 32. is b

Answer for Question No 33. is c

Answer for Question No 34. is b

Answer for Question No 35. is a

Answer for Question No 36. is a

Answer for Question No 37. is a

Answer for Question No 38. is a

Answer for Question No 39. is a

Answer for Question No 40. is d

Answer for Question No 41. is b

Answer for Question No 42. is d

Answer for Question No 43. is d

Answer for Question No 44. is b

Answer for Question No 45. is d

Answer for Question No 46. is b

Answer for Question No 47. is c

Answer for Question No 48. is a

Answer for Question No 49. is c

Answer for Question No 50. is d

Answer for Question No 51. is a

Answer for Question No 52. is d

Answer for Question No 53. is c

Answer for Question No 54. is d

Answer for Question No 55. is b

Answer for Question No 56. is d

Answer for Question No 57. is d

Answer for Question No 58. is d

Answer for Question No 59. is c

Answer for Question No 60. is a

Answer for Question No 1. is d

Answer for Question No 2. is c

Answer for Question No 3. is a

Answer for Question No 4. is a

Answer for Question No 5. is d

Answer for Question No 6. is c

Answer for Question No 7. is b

Answer for Question No 8. is b

Answer for Question No 9. is c

Answer for Question No 10. is c

Answer for Question No 11. is b

Answer for Question No 12. is b

Answer for Question No 13. is c

Answer for Question No 14. is a

Answer for Question No 15. is b

Answer for Question No 16. is d

Answer for Question No 17. is a

Answer for Question No 18. is a

Answer for Question No 19. is b

Answer for Question No 20. is c

Answer for Question No 21. is d

Answer for Question No 22. is a

Answer for Question No 23. is b

Answer for Question No 24. is d

Answer for Question No 25. is a

Answer for Question No 26. is b

Answer for Question No 27. is d

Answer for Question No 28. is d

Answer for Question No 29. is d

Answer for Question No 30. is d

Answer for Question No 31. is a

Answer for Question No 32. is c

Answer for Question No 33. is a

Answer for Question No 34. is c

Answer for Question No 35. is b

Answer for Question No 36. is b

Answer for Question No 37. is a

Answer for Question No 38. is b

Answer for Question No 39. is a

Answer for Question No 40. is d

Answer for Question No 41. is c

Answer for Question No 42. is b

Answer for Question No 43. is a

Answer for Question No 44. is d

Answer for Question No 45. is b

Answer for Question No 46. is c

Answer for Question No 47. is a

Answer for Question No 48. is a

Answer for Question No 49. is b

Answer for Question No 50. is a

Answer for Question No 51. is c

Answer for Question No 52. is d

Answer for Question No 53. is c

Answer for Question No 54. is c

Answer for Question No 55. is d

Answer for Question No 56. is d

Answer for Question No 57. is b

Answer for Question No 58. is d

Answer for Question No 59. is d

Answer for Question No 60. is a

Answer for Question No 1. is b

Answer for Question No 2. is d

Answer for Question No 3. is b

Answer for Question No 4. is a

Answer for Question No 5. is d

Answer for Question No 6. is c

Answer for Question No 7. is a

Answer for Question No 8. is c

Answer for Question No 9. is c

Answer for Question No 10. is b

Answer for Question No 11. is b

Answer for Question No 12. is a

Answer for Question No 13. is a

Answer for Question No 14. is c

Answer for Question No 15. is a

Answer for Question No 16. is b

Answer for Question No 17. is c

Answer for Question No 18. is b

Answer for Question No 19. is c

Answer for Question No 20. is d

Answer for Question No 21. is a

Answer for Question No 22. is d

Answer for Question No 23. is d

Answer for Question No 24. is a

Answer for Question No 25. is d

Answer for Question No 26. is d

Answer for Question No 27. is b

Answer for Question No 28. is d

Answer for Question No 29. is d

Answer for Question No 30. is a

Answer for Question No 31. is d

Answer for Question No 32. is b

Answer for Question No 33. is a

Answer for Question No 34. is d

Answer for Question No 35. is b

Answer for Question No 36. is c

Answer for Question No 37. is a

Answer for Question No 38. is b

Answer for Question No 39. is b

Answer for Question No 40. is b

Answer for Question No 41. is a

Answer for Question No 42. is a

Answer for Question No 43. is d

Answer for Question No 44. is b

Answer for Question No 45. is a

Answer for Question No 46. is a

Answer for Question No 47. is d

Answer for Question No 48. is c

Answer for Question No 49. is a

Answer for Question No 50. is b

Answer for Question No 51. is d

Answer for Question No 52. is a

Answer for Question No 53. is c

Answer for Question No 54. is a

Answer for Question No 55. is d

Answer for Question No 56. is d

Answer for Question No 57. is c

Answer for Question No 58. is b

Answer for Question No 59. is c

Answer for Question No 60. is d

Answer for Question No 1. is d

Answer for Question No 2. is b

Answer for Question No 3. is c

Answer for Question No 4. is c

Answer for Question No 5. is c

Answer for Question No 6. is a

Answer for Question No 7. is b

Answer for Question No 8. is d

Answer for Question No 9. is a

Answer for Question No 10. is b

Answer for Question No 11. is b

Answer for Question No 12. is b

Answer for Question No 13. is a

Answer for Question No 14. is c

Answer for Question No 15. is b

Answer for Question No 16. is a

Answer for Question No 17. is c

Answer for Question No 18. is b

Answer for Question No 19. is d

Answer for Question No 20. is a

Answer for Question No 21. is c

Answer for Question No 22. is a

Answer for Question No 23. is d

Answer for Question No 24. is d

Answer for Question No 25. is d

Answer for Question No 26. is d

Answer for Question No 27. is a

Answer for Question No 28. is d

Answer for Question No 29. is c

Answer for Question No 30. is d

Answer for Question No 31. is a

Answer for Question No 32. is b

Answer for Question No 33. is a

Answer for Question No 34. is d

Answer for Question No 35. is b

Answer for Question No 36. is a

Answer for Question No 37. is d

Answer for Question No 38. is b

Answer for Question No 39. is b

Answer for Question No 40. is c

Answer for Question No 41. is b

Answer for Question No 42. is d

Answer for Question No 43. is a

Answer for Question No 44. is a

Answer for Question No 45. is b

Answer for Question No 46. is d

Answer for Question No 47. is d

Answer for Question No 48. is c

Answer for Question No 49. is a

Answer for Question No 50. is a

Answer for Question No 51. is c

Answer for Question No 52. is d

Answer for Question No 53. is a

Answer for Question No 54. is b

Answer for Question No 55. is c

Answer for Question No 56. is a

Answer for Question No 57. is d

Answer for Question No 58. is c

Answer for Question No 59. is d

Answer for Question No 60. is b

Answer for Question No 1. is d

Answer for Question No 2. is b

Answer for Question No 3. is c

Answer for Question No 4. is c

Answer for Question No 5. is d

Answer for Question No 6. is c

Answer for Question No 7. is b

Answer for Question No 8. is c

Answer for Question No 9. is a

Answer for Question No 10. is c

Answer for Question No 11. is b

Answer for Question No 12. is a

Answer for Question No 13. is d

Answer for Question No 14. is a

Answer for Question No 15. is b

Answer for Question No 16. is c

Answer for Question No 17. is a

Answer for Question No 18. is c

Answer for Question No 19. is b

Answer for Question No 20. is b

Answer for Question No 21. is d

Answer for Question No 22. is c

Answer for Question No 23. is a

Answer for Question No 24. is a

Answer for Question No 25. is a

Answer for Question No 26. is d

Answer for Question No 27. is d

Answer for Question No 28. is d

Answer for Question No 29. is d

Answer for Question No 30. is c

Answer for Question No 31. is a

Answer for Question No 32. is b

Answer for Question No 33. is d

Answer for Question No 34. is d

Answer for Question No 35. is a

Answer for Question No 36. is b

Answer for Question No 37. is c

Answer for Question No 38. is b

Answer for Question No 39. is b

Answer for Question No 40. is b

Answer for Question No 41. is a

Answer for Question No 42. is b

Answer for Question No 43. is b

Answer for Question No 44. is a

Answer for Question No 45. is d

Answer for Question No 46. is c

Answer for Question No 47. is d

Answer for Question No 48. is a

Answer for Question No 49. is d

Answer for Question No 50. is d

Answer for Question No 51. is b

Answer for Question No 52. is d

Answer for Question No 53. is a

Answer for Question No 54. is a

Answer for Question No 55. is c

Answer for Question No 56. is c

Answer for Question No 57. is d

Answer for Question No 58. is b

Answer for Question No 59. is a

Answer for Question No 60. is c

Answer for Question No 1. is b

Answer for Question No 2. is d

Answer for Question No 3. is c

Answer for Question No 4. is a

Answer for Question No 5. is b

Answer for Question No 6. is c

Answer for Question No 7. is c

Answer for Question No 8. is d

Answer for Question No 9. is c

Answer for Question No 10. is a

Answer for Question No 11. is c

Answer for Question No 12. is b

Answer for Question No 13. is b

Answer for Question No 14. is a

Answer for Question No 15. is b

Answer for Question No 16. is b

Answer for Question No 17. is b

Answer for Question No 18. is a

Answer for Question No 19. is c

Answer for Question No 20. is c

Answer for Question No 21. is a

Answer for Question No 22. is a

Answer for Question No 23. is d

Answer for Question No 24. is d

Answer for Question No 25. is c

Answer for Question No 26. is d

Answer for Question No 27. is b

Answer for Question No 28. is a

Answer for Question No 29. is c

Answer for Question No 30. is a

Answer for Question No 31. is a

Answer for Question No 32. is a

Answer for Question No 33. is d

Answer for Question No 34. is d

Answer for Question No 35. is d

Answer for Question No 36. is d

Answer for Question No 37. is d

Answer for Question No 38. is a

Answer for Question No 39. is b

Answer for Question No 40. is b

Answer for Question No 41. is b

Answer for Question No 42. is b

Answer for Question No 43. is d

Answer for Question No 44. is b

Answer for Question No 45. is b

Answer for Question No 46. is d

Answer for Question No 47. is c

Answer for Question No 48. is d

Answer for Question No 49. is a

Answer for Question No 50. is d

Answer for Question No 51. is d

Answer for Question No 52. is a

Answer for Question No 53. is a

Answer for Question No 54. is c

Answer for Question No 55. is b

Answer for Question No 56. is d

Answer for Question No 57. is b

Answer for Question No 58. is c

Answer for Question No 59. is a

Answer for Question No 60. is c

Answer for Question No 1. is a

Answer for Question No 2. is a

Answer for Question No 3. is b

Answer for Question No 4. is b

Answer for Question No 5. is c

Answer for Question No 6. is d

Answer for Question No 7. is c

Answer for Question No 8. is c

Answer for Question No 9. is d

Answer for Question No 10. is b

Answer for Question No 11. is c

Answer for Question No 12. is b

Answer for Question No 13. is a

Answer for Question No 14. is b

Answer for Question No 15. is a

Answer for Question No 16. is b

Answer for Question No 17. is c

Answer for Question No 18. is d

Answer for Question No 19. is c

Answer for Question No 20. is a

Answer for Question No 21. is a

Answer for Question No 22. is d

Answer for Question No 23. is a

Answer for Question No 24. is d

Answer for Question No 25. is d

Answer for Question No 26. is a

Answer for Question No 27. is c

Answer for Question No 28. is a

Answer for Question No 29. is a

Answer for Question No 30. is a

Answer for Question No 31. is d

Answer for Question No 32. is c

Answer for Question No 33. is b

Answer for Question No 34. is a

Answer for Question No 35. is c

Answer for Question No 36. is a

Answer for Question No 37. is b

Answer for Question No 38. is b

Answer for Question No 39. is b

Answer for Question No 40. is d

Answer for Question No 41. is d

Answer for Question No 42. is b

Answer for Question No 43. is b

Answer for Question No 44. is b

Answer for Question No 45. is d

Answer for Question No 46. is d

Answer for Question No 47. is a

Answer for Question No 48. is d

Answer for Question No 49. is c

Answer for Question No 50. is d

Answer for Question No 51. is d

Answer for Question No 52. is c

Answer for Question No 53. is a

Answer for Question No 54. is c

Answer for Question No 55. is a

Answer for Question No 56. is d

Answer for Question No 57. is b

Answer for Question No 58. is c

Answer for Question No 59. is d

Answer for Question No 60. is b

Answer for Question No 1. is b

Answer for Question No 2. is c

Answer for Question No 3. is c

Answer for Question No 4. is a

Answer for Question No 5. is b

Answer for Question No 6. is c

Answer for Question No 7. is d

Answer for Question No 8. is b

Answer for Question No 9. is a

Answer for Question No 10. is b

Answer for Question No 11. is b

Answer for Question No 12. is c

Answer for Question No 13. is b

Answer for Question No 14. is c

Answer for Question No 15. is c

Answer for Question No 16. is c

Answer for Question No 17. is a

Answer for Question No 18. is a

Answer for Question No 19. is d

Answer for Question No 20. is b

Answer for Question No 21. is d

Answer for Question No 22. is d

Answer for Question No 23. is a

Answer for Question No 24. is a

Answer for Question No 25. is a

Answer for Question No 26. is a

Answer for Question No 27. is d

Answer for Question No 28. is a

Answer for Question No 29. is d

Answer for Question No 30. is d

Answer for Question No 31. is c

Answer for Question No 32. is a

Answer for Question No 33. is d

Answer for Question No 34. is b

Answer for Question No 35. is d

Answer for Question No 36. is b

Answer for Question No 37. is d

Answer for Question No 38. is a

Answer for Question No 39. is c

Answer for Question No 40. is d

Answer for Question No 41. is b

Answer for Question No 42. is b

Answer for Question No 43. is a

Answer for Question No 44. is c

Answer for Question No 45. is b

Answer for Question No 46. is a

Answer for Question No 47. is a

Answer for Question No 48. is d

Answer for Question No 49. is d

Answer for Question No 50. is d

Answer for Question No 51. is c

Answer for Question No 52. is d

Answer for Question No 53. is a

Answer for Question No 54. is d

Answer for Question No 55. is b

Answer for Question No 56. is c

Answer for Question No 57. is d

Answer for Question No 58. is b

Answer for Question No 59. is c

Answer for Question No 60. is c

Answer for Question No 1. is c

Answer for Question No 2. is b

Answer for Question No 3. is b

Answer for Question No 4. is a

Answer for Question No 5. is d

Answer for Question No 6. is a

Answer for Question No 7. is c

Answer for Question No 8. is b

Answer for Question No 9. is d

Answer for Question No 10. is c

Answer for Question No 11. is c

Answer for Question No 12. is a

Answer for Question No 13. is b

Answer for Question No 14. is b

Answer for Question No 15. is b

Answer for Question No 16. is c

Answer for Question No 17. is d

Answer for Question No 18. is c

Answer for Question No 19. is c

Answer for Question No 20. is a

Answer for Question No 21. is a

Answer for Question No 22. is d

Answer for Question No 23. is a

Answer for Question No 24. is b

Answer for Question No 25. is a

Answer for Question No 26. is a

Answer for Question No 27. is d

Answer for Question No 28. is d

Answer for Question No 29. is a

Answer for Question No 30. is d

Answer for Question No 31. is d

Answer for Question No 32. is a

Answer for Question No 33. is b

Answer for Question No 34. is b

Answer for Question No 35. is a

Answer for Question No 36. is d

Answer for Question No 37. is a

Answer for Question No 38. is d

Answer for Question No 39. is c

Answer for Question No 40. is b

Answer for Question No 41. is d

Answer for Question No 42. is c

Answer for Question No 43. is b

Answer for Question No 44. is b

Answer for Question No 45. is a

Answer for Question No 46. is d

Answer for Question No 47. is a

Answer for Question No 48. is a

Answer for Question No 49. is a

Answer for Question No 50. is d

Answer for Question No 51. is d

Answer for Question No 52. is c

Answer for Question No 53. is b

Answer for Question No 54. is d

Answer for Question No 55. is c

Answer for Question No 56. is d

Answer for Question No 57. is a

Answer for Question No 58. is c

Answer for Question No 59. is b

Answer for Question No 60. is d

Answer for Question No 1. is d

Answer for Question No 2. is b

Answer for Question No 3. is b

Answer for Question No 4. is a

Answer for Question No 5. is c

Answer for Question No 6. is b

Answer for Question No 7. is a

Answer for Question No 8. is a

Answer for Question No 9. is a

Answer for Question No 10. is b

Answer for Question No 11. is b

Answer for Question No 12. is b

Answer for Question No 13. is c

Answer for Question No 14. is c

Answer for Question No 15. is d

Answer for Question No 16. is c

Answer for Question No 17. is c

Answer for Question No 18. is b

Answer for Question No 19. is c

Answer for Question No 20. is a

Answer for Question No 21. is a

Answer for Question No 22. is a

Answer for Question No 23. is a

Answer for Question No 24. is d

Answer for Question No 25. is b

Answer for Question No 26. is a

Answer for Question No 27. is d

Answer for Question No 28. is c

Answer for Question No 29. is d

Answer for Question No 30. is d

Answer for Question No 31. is d

Answer for Question No 32. is b

Answer for Question No 33. is b

Answer for Question No 34. is c

Answer for Question No 35. is b

Answer for Question No 36. is b

Answer for Question No 37. is a

Answer for Question No 38. is c

Answer for Question No 39. is a

Answer for Question No 40. is b

Answer for Question No 41. is b

Answer for Question No 42. is d

Answer for Question No 43. is d

Answer for Question No 44. is a

Answer for Question No 45. is a

Answer for Question No 46. is a

Answer for Question No 47. is d

Answer for Question No 48. is a

Answer for Question No 49. is c

Answer for Question No 50. is d

Answer for Question No 51. is a

Answer for Question No 52. is c

Answer for Question No 53. is d

Answer for Question No 54. is c

Answer for Question No 55. is b

Answer for Question No 56. is a

Answer for Question No 57. is d

Answer for Question No 58. is c

Answer for Question No 59. is d

Answer for Question No 60. is d