



BRAINWARE UNIVERSITY

ODD Semester Examinations 2021- 22

Programme – Bachelor of Technology in Computer Science & Engineering - 2018 [B.Tech.(CSE)]

Course Name – Big Data Analysis

Course Code – PEC-701A

(Semester VII)

Time allotted : 1 Hour 25 Minutes

Full Marks : 70

(Multiple choice type question)

70 x 1 = 70

Choose the correct alternative from the following

- (I) Threads process data
- | | |
|--|-----------------|
| A) local | B) close |
| C) minimize the costs of data access latency | D) All of these |
- (II) Map functions is/are
- | | |
|----------------|-----------------|
| A) computation | B) analysis |
| C) set pairs | D) All of these |
- (III) A _____ serves as the master and there is only one NameNode per cluster.
- | | |
|---------------|----------------|
| A) Data Node | B) NameNode |
| C) Data block | D) Replication |
- (IV) Association Mining is
- | | |
|------------------------------|-----------------|
| A) Finding frequent patterns | B) associations |
| C) correlations | D) All of these |
- (V) Which data base is popular?
- | | |
|------------|----------------------|
| A) MongoDB | B) Oracle |
| C) Mysql | D) None of the above |
- (VI) A ____ is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.
- | | |
|------------------|--------------------|
| A) Decision tree | B) Graphs |
| C) Trees | D) Neural Networks |
- (VII) what are the components of big data
- | | |
|--------------|-----------------|
| A) MapReduce | B) HDFS |
| C) YARN | D) All of these |
- (VIII) Features of Big Data Platform
- | | |
|--|---|
| A) It should support variety of data format | B) It should provide real-time data analysis software |
| C) It should have tools for searching the data through large data sets | D) All of these |
- (IX) BigData Platforms is/ are
- | | |
|---|------------------------|
| A) IBM Open Platform | B) Amazon Web Services |
| C) Both IBM Open Platform and Amazon Web Services | D) None of these |
- (X) Recommender Systems is/are
- | | |
|--------------|------------------|
| A) Content | B) Collaborative |
| C) Knowledge | D) All of these |
- (XI) Real Time Analytics Platform (RTAP)
- | | |
|----------------------|--------------------|
| A) analyzes data | B) correlates data |
| C) predicts outcomes | D) All of these |

- (XII) _____ maps input key/value pairs to a set of intermediate key/value pairs
- A) Mapper
B) Reducer
C) Both Mapper and Reducer
D) None of the above
- (XIII) Hadoop is open source.
- A) ALWAYS True
B) True only for Apache Hadoop
C) True only for Apache and Cloudera Hadoop
D) ALWAYS False
- (XIV) Requirements of cluster analysis
- A) Ability to deal with noisy data
B) Ability to deal with different types of attributes
C) both Ability to deal with noisy data and Ability to deal with different types of attributes
D) None of these
- (XV) Veracity means
- A) Filter
B) Reliable
C) manage data
D) All of these
- (XVI) which of the File system is used by HBase?
- A) Hive
B) Imphala
C) Hadoop
D) Scala
- (XVII) Network provides
- A) persistence of data
B) pipes
C) demanding data
D) All of these
- (XVIII) Big data focus on
- A) Identify the data we have
B) Identify the data we need
C) what you want to achieve
D) All of these
- (XIX) Association rules are sometimes referred to as
- A) market basket analysis
B) itemset filtering
C) frequent item set analysis
D) none of these
- (XX) Which of the following is a characteristic of a NoSQL database?
- A) Uses JSON
B) Needs a schema
C) Requires JOINS
D) Uses tables for storage
- (XXI) Document Stores is/are
- A) computer program
B) designed for storing
C) retrieving and managing document-oriented information
D) All of these
- (XXII) HDFS provides file
- A) permissions
B) authentication
C) Both permissions and authentication
D) None of these
- (XXIII) Hbase tables are
- A) Made read only by setting the read only option
B) Always writeable
C) Always read only
D) Are made read only using the query to the table
- (XXIV) Yahoo wants to know which of its pages are getting unusual number of hits in the past
- A) Mining query stream
B) Mining login stream
C) Mining search stream
D) Mining click stream
- (XXV) k-means
- A) is an analytical technique
B) identifies k clusters
C) Both an analytical technique and identifies k clusters
D) None of these
- (XXVI) The primary aspect of Big Data is to provide
- A) demanding data rapidly
B) demanding data slowly
C) no demanding data
D) None of these
- (XXVII) In Content-based Approach problem is/are
- A) Finding the appropriate features is hard
B) Recommendations for new users

- C) Both Finding the appropriate features is hard and Recommendations for new users
D) None of these
- (XXVIII) Google wants to know which queries are frequent today than yesterday
A) mining query stream
B) mining login stream
C) mining search stream
D) mining click stream
- (XXIX) 4.75 exabytes=
A) genome sequences of 7 billion people
B) genome sequences of 8 billion people
C) genome sequences of 9 billion people
D) genome sequences of 10 billion people
- (XXX) In Filtering Streams _____
A) Accept those tuples in the stream that meet a criterion.
B) Accept data in the stream that meet a criterion.
C) Accept those class in the stream that meet a criterion
D) Accept rows in the stream that meet a criterion.
- (XXXI) What is NoSQL database?
A) NoSQL is a database is an enhanced form of RDBMS.
B) NoSQL is database that is built with enhancements to DBMS.
C) NoSQL is a database that is built on ways and means other than tables and columns.
D) All of the above
- (XXXII) In streaming queries, alter the user when stock crosses over a price point is an example of
A) continuous queries
B) one time queries
C) sampling queries
D) none of these
- (XXXIII) The name node effectively coordinates
A) distributed data nodes
B) data nodes
C) central data nodes
D) None of these
- (XXXIV) By 2027, the volume of data produced digitally will reach to
A) TB
B) YB
C) ZB
D) EB
- (XXXV) The namenode performs
A) Manages the file system namespace
B) Regulates client's access to files
C) renaming
D) All of these
- (XXXVI) Which of the following is a reason to use an SQL database?
A) It's ACID-compliant.
B) It can easily store unstructured data. It's ACID-compliant.
C) It can enable development in the cloud
D) None of above
- (XXXVII) Graph model of NoSQL used in
A) Twitter
B) Facebook
C) Google
D) WhatsAPP
- (XXXVIII) Stock Market Prediction is
A) the act of trying to determine the future value of company stock
B) other financial instrument traded on an exchange.
C) the successful prediction of a stock's future price could yield significant profit
D) All of these
- (XXXIX) Where does the bayes rule can be used?
A) Solving queries
B) Increasing complexity
C) Decreasing complexity
D) Answering probabilistic query
- (XL) Padma=
A) 10^{21}
B) 10^{22}
C) 10^{23}
D) 10^{32}
- (XLI) 422 exabytes=
A) Total digital data created in 2008
B) Total digital data created in 2009
C) Total digital data created in 2010
D) Total digital data created in 2011
- (XLII) _____ takes the grouped key-value paired data as input and runs a Reducer function on each one of them
A) MAPPER
B) REDUCER

C) COMBINER

D) PARTITIONER

(XLIII) Which one of the following is false about Hadoop?

- A) It is a distributed framework
- C) It runs with commodity hardware

- B) The main algorithm used in it is Map Reduce
- D) All are true

(XLIV) Hive was developed by

- A) Facebook
- C) Flipkart

- B) Amazon
- D) None of the above

(XLV) Chance Nodes are represented by ___

- A) Disks
- C) Circles

- B) Squares
- D) Triangles

(XLVI) A _____ store is a simple database that when

- A) document
- C) graph

- B) key-value
- D) simple

(XLVII) NoSQL databases is used mainly for handling large volumes of _____ data

- A) structured
- C) semi-structured

- B) unstructured
- D) None of above

(XLVIII) Any data with unknown form

- A) structured
- C) semi-structured

- B) Un structured
- D) None of these

(XLIX) Which of the following represent column in NoSQL _____

- A) Field
- C) Collection

- B) Database
- D) Document

(L) There are _____ major classification collaborative filtering mechanisms

- A) 1
- C) 3

- B) 2
- D) None of these

(LI) HDFS works in a _____ fashion.

- A) worker-master fashion
- C) master-worker fashion

- B) master-slave fashion
- D) slave-master fashion

(LII) Big Data can be

- A) structured
- C) semi-structured

- B) unstructured
- D) All of these

(LIII) Facebook can generate approximately

- A) 10 billion messages
- C) 20 billion messages

- B) A billion messages
- D) 30 billion messages

(LIV) Satellite gives

- A) huge data
- C) Minimum data

- B) big data
- D) Both a and b

(LV) Which of the following statements about Naive Bayes is incorrect?

- A) Attributes are equally important
- C) Attributes are statistically independent of one another given the class value.

- B) Attributes are statistically dependent of one another given the class value
- D) Attributes can be nominal or numeric

(LVI) Most big data applications achieve their performance through

- A) computing resources
- C) Both collection of storage and resources

- B) collection of storage
- D) None of these

(LVII) Most single node machines have a limit to the amount of

- A) memory
- C) XML file and memory

- B) XML file
- D) None of these

(LVIII) Data Stream is

- A) Large data volume
C) arriving a very high rate
- (LIX) Tabular Stores is/are
A) structured in a tabular form.
C) arranges data elements in horizontal rows
- (LX) In Filtering Streams_____
A) Accept those tuples in the stream that meet a criterion
C) Accept those class in the stream that meet a criterion
- (LXI) What is the consequence between a node and its predecessors while creating bayesian network?"
A) Functionally dependent
C) Conditionally independent
- (LXII) The size of big data
A) GB
C) MB
- (LXIII) Which of the following statements about standard Bloom filters is correct?
A) It is possible to delete an element from a Bloom filter.
C) It is possible to alter the hash functions of a full Bloom filter to create more space.
- (LXIV) Which of the following are the advantage/s of Decision Trees?
A) Possible Scenarios can be added
C) Worst, best and expected values can be determined for different scenarios
- (LXV) semi-structured data is a data represented in an
A) XML file
C) Node JS file
- (LXVI) Requirements of cluster analysis
A) Scalability
C) Minimal requirements for domain knowledge to determine input parameters
- (LXVII) _____recommended items based on similarity measures between users and/or items
A) content based systems
C) collaborative filtering system
- (LXVIII) Petabyte=
A) 10^{15}
C) 10^{18}
- (LXIX) The world's largest hadoop cluster
A) Facebook
C) datamatics
- (LXX) Yottabyte=
A) 10^{21}
C) 10^{23}
- B) likely structured
D) All of these
- B) arranges data elements in vertical columns
D) All of these
- B) Accept data in the stream that meet a criterion.
D) Accept rows in the stream that meet a criterion.
- B) Dependant
D) Both Conditionally dependant & Dependant
- B) KB
D) Peta bytes
- B) A Bloom filter always returns the correct result.
D) A Bloom filter always returns TRUE when testing for a previously added element
- B) Use a white box model, If given result is provided by a model
D) All of Above
- B) HTML file
D) None of these
- B) Ability to deal with noisy data
D) All of these
- B) hybrid system
D) none of these
- B) 10^{17}
D) 10^{19}
- B) apple
D) None of these
- B) 10^{22}
D) 10^{24}