BRAINWARE UNIVERSITY

Term End Examination 2021 - 22
Programme – Bachelor of Technology in Computer Science & Engineering
Course Name – Big Data Analysis
Course Code - PEC-701A
(Semester VII)

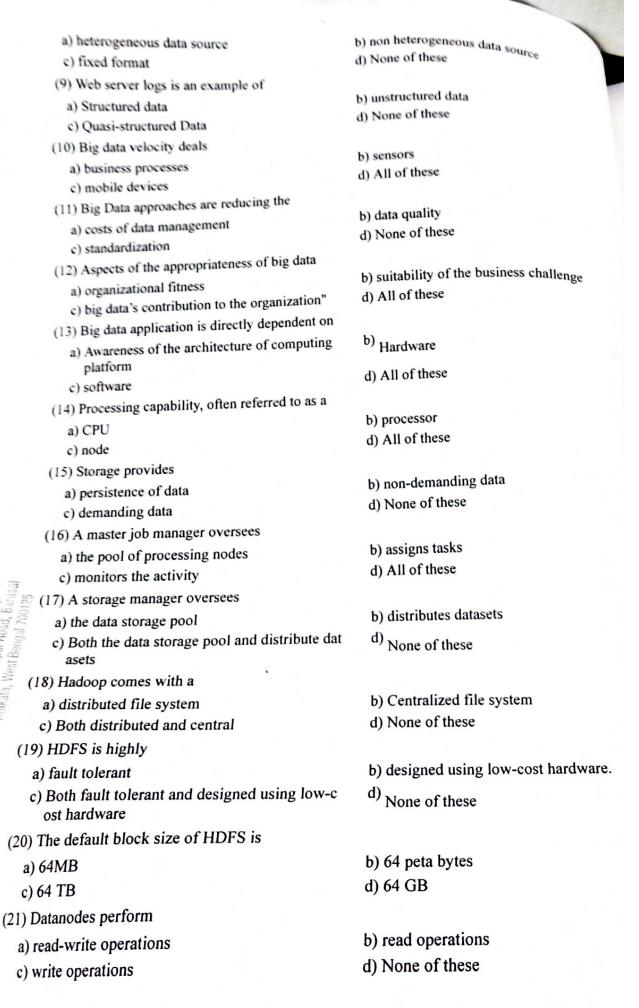Time : 1 Hr.25 Min.

Full Marks : 70

[The figure in the margin indicates full marks.]

Group-A

1 x 70=70

(Multiple Choice Type Question)

Choose the correct alternative from the following :

(1) These data come from many sources like
a) Facebook
b) Google
c) Amazon
d) All of these

(2) Exabyte=
a) 10^15
b) 10^17
c) 10^18
d) 10^19

(3) Zettabyte=
a) 10^15
b) 10^17
c) 10^18
d) 10^21

(4) Padma=
a) 10^21
b) 10^22
c) 10^23
d) 10^32

(5) 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

(6) Facebook can generate approximately
a) 10 billion messages
b) A billion messages
c) 20 billion messages
d) 30 billion messages

(7) Structured data
a) Tabular form
b) Non tabular form
c) variable form
d) None of these

(8) Unstructured data is

a) heterogeneous data source

b) non heterogeneous data source

c) fixed format

d) None of these

(9) Web server logs is an example of

a) Structured data

b) unstructured data

c) Quasi-structured Data

d) None of these

(10) Big data velocity deals

a) business processes

b) sensors

c) mobile devices

d) All of these

(11) Big Data approaches are reducing the

a) costs of data management

b) data quality

c) standardization

d) None of these

(12) Aspects of the appropriateness of big data

a) organizational fitness

b) suitability of the business challenge

c) big data's contribution to the organization"

d) All of these

(13) Big data application is directly dependent on

a) Awareness of the architecture of computing platform

b) Hardware

c) software

d) All of these

(14) Processing capability, often referred to as a

a) CPU

b) processor

c) node

d) All of these

(15) Storage provides

a) persistence of data

b) non-demanding data

c) demanding data

d) None of these

(16) A master job manager oversees

a) the pool of processing nodes

b) assigns tasks

c) monitors the activity

d) All of these

(17) A storage manager oversees

a) the data storage pool

b) distributes datasets

c) Both the data storage pool and distribute datasets

d) None of these

(18) Hadoop comes with a

a) distributed file system

b) Centralized file system

c) Both distributed and central

d) None of these

(19) HDFS is highly

a) fault tolerant

b) designed using low-cost hardware.

c) Both fault tolerant and designed using low-cost hardware

d) None of these

(20) The default block size of HDFS is

a) 64MB

b) 64 peta bytes

c) 64 TB

d) 64 GB

(21) Datanodes perform

a) read-write operations

b) read operations

c) write operations

d) None of these

(22) The name node effectively coordinates

a) distributed data nodes

b) data nodes

c) central data nodes

d) None of these

(23) The tasks that can be executed in map reduce

a) parallel

b) sequential

c) logical

d) None of these

(24) What are the 4 V's in big data

a) volume

b) velocity

c) variety

d) All of these

(25) _____ have a structure but cannot be stored in a database

a) Structured

b) Semi Structured

c) Unstructured

d) None of these

(26) _____ refers to the ability to turn your data useful for business

a) Velocity

b) variety

c) Value

d) Volume

(27) _____ is factors considered before Adopting Big Data Technology

a) Validation

b) Verification

c) Data

d) Design

(28) Choose from the following that are Decision Tree nodes?

a) Decision Nodes

b) End Nodes

c) Chance Nodes

d) All of Above

(29) Decision Nodes are represented by __

a) Disks

b) Squares

c) Circles

d) Triangles

(30) End Nodes are represented by __

a) Disks

b) Squares

c) Circles

d) Triangles

(31) classification is

a) data mining process

b) not data mining process

c) clustering process

d) None of these

(32) How the bayesian network can be used to answer any query?"

a) Full distribution

b) Joint distribution

c) Partial distribution

d) All of these

(33) Node is _____

a) A component of a network

b) In the context of KDD and data mining, this refers to random errors in a database table.

c) One of the defining aspects of a data warehouse

d) None of these

(34) The classification method in which the upper limit of interval is same as of lower class interval is called

a) exclusive method

b) inclusive method

c) mid point method

d) None of these

(35) categories of clustering methods

a) Partitioning methods

b) Hierarchical methods

d) All of these

c) Grid based methods

(36) Bayesian Classifiers are

a) statistical classifiers

b) predict class

d) None of these

c) both

(37) Which of the following is not an example of Social Media?

a) Twitter

b) Google

d) Youtube

c) Insta

(38) Which of the following options is not the example of NoSql ?

a) Google

b) NetFlix

d) CERN

c) Amazon

(39) What are the different features of Big Data Analysis?

a) Open-Source

b) Scalability

d) All the above

c) Data Recovery

(40) Can decision tree be used for clustering?

a) possible

b) impossible

d) None of these

c) impossible in some scenario

(41) _____ recommended items based on similarity measures between users and/or it ems

a) content based systems

b) hybrid system

d) none of these

c) collaborative filtering system

(42) Association rules are sometimes referred to as

a) market basket analysis

b) itemset filtering

d) none of these

c) frequent item set analysis

(43) The number of maps is usually driven by the total size of _____ .

a) task

b) output

d) none

c) input

(44) Which of the following is not an example of a nosql database management system?

a) Cassandra

b) Scylla

d) PostgreSQL

c) Handhoop / Hbase

(45) NoSQL databases are most often referred to as

a) Network

b) Distributed

c) Relational

d) Object-oriented

(46) The core principle of nosql is

a) High availability

b) Low availability

c) both High & Low availability

d) None of above

(47) Which of the following is not a strong feature for nosql databases?

a) Scalability

b) Relational data

c) Faster data access than RDBMS.

d) Data easily held across multiple servers

(48) Which of the following are the simplest NoSQL databases?

a) Key-value

b) Document

c) Wide-column

d) All of the above

(49) What is the aim of nosql?

a) NoSQL is not suitable for storing structured

b) NoSQL databases allow storing non-structur

data.

c) NoSQL is a new data format to store large datasets.

ed data.

d) NoSQL provides an alternative to SQL databases to store textual data.

(50) The Process of describing the data that is huge and complex to store and process is known as

a) Analytics

b) Data mining

c) Big Data

d) Data Warehouse

(51) Check below the best answer to "which industries employ the use of so-called "Big Data" in their day to day operations?

a) Weather forecasting

b) Marketing

c) Healthcare

d) All of the above

(52) Which of the Node serves as the master and there is only one NameNode per cluster.

a) Data Node

b) NameNode

c) Data block

d) Replication

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

a) unstructured

b) structured

c) semi-structured

d) all of the mentioned

(54) Hbase creates a new version of a record during

a) Creation of a record

b) Modification of a record

c) Deletion of a record

d) All the above

(55) Which of the following statements about standard Bloom filters is correct?

a) It is possible to delete an element from a Bloom filter.

b) A Bloom filter always returns the correct result.

c) It is possible to alter the hash functions of a full Bloom filter to create more space.

d) A Bloom filter always returns TRUE when testing for a previously added element

(56) 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.

(57) Which one does not belong to application of data stream?

a) web traffic

b) internet

c) sensor data

d) None of these

(58) Which was not following the data stream concepts?

a) financial applications

b) network monitoring

c) fraud detection

d) web application

(59) The MapReduce algorithm contains two important tasks, namely _____.

a) mapped, reduce

b) mapping, Reduction

c) Map, Reduction

d) Map, Reduce

(60) No SQL means

a) Not SQL

b) No usage of SQL

c) Not only SQL

d) Not for SQL

(61) Which is not example of NoSQL?

a) Google

b) NetFlix

c) Amazon

d) None of these

(62) MongoDB is

a) column based   b) key value based

c) document based   d) graph based

(63) Hive query can be stored in

a) Local file   b) HDFS

c) Both   d) Can not be stored

(64) What is true about Variety in bigdata?

a) high in size   b) speed of data

c) data from   d) data in certain

(65) Problems in recommendation systems

a) poor results   b) poor data

c) Lack of data   d) All of these

(66) Type of recommender systems

a) content   b) collaborative

c) knowledge   d) All of these

(67) Applications of Association rules

a) Basket data analysis   b) cross-marketing

c) clustering   d) All of these

(68) A data stream is a sequence of digitally encoded

a) coherent signals   b) packets of data

c) data packets   d) All of these

(69) A data stream is potentially

a) unbounded in size   b) generated continuously in real time

c) the volume of the data is very large   d) All of these

(70) Data streams are in actions

a) Security applications   b) Telecom call records

c) Financial applications   d) All of these