

These are sample MCQs to indicate pattern, may or may not appear in examination

**University of Mumbai**  
**Online Examination 2020**

Program: BE Computer Engineering

Curriculum Scheme: Revised 2016

Examination: Final Year Semester VII

Course Code:CSDLO7032 and Course Name: Big Data Analytics

Time: 1hour

Max. Marks: 50

**Note to the students:-** All the Questions are compulsory and carry equal marks .

Q1.	Concerning the Forms of Big Data, which one of these is odd?
Option A:	Structured
Option B:	Unstructured
Option C:	Processed
Option D:	Semi structured

Q2.	The feature of big data that refers to the quality of the stored data is _____
Option A:	Variety
Option B:	Volume
Option C:	Velocity
Option D:	Veracity

Q3.	For Hadoop, _____ works as a scheduler
Option A:	Hbase
Option B:	Oozie
Option C:	Pig
Option D:	Flume

Q4.	The _____ operation allows us to find tuples that are present in one relation but not in another relation.
Option A:	Union
Option B:	Set difference
Option C:	Intersection
Option D:	Projection

Q5.	In HDFS, duplicate copies of files are stored at several nodes. This feature is called_____
Option A:	Data Mining
Option B:	Data Consistency
Option C:	Data Replication
Option D:	Data Availability

Q6.	In MapReduce _____ is used between Mapper & Reducer for grouping values with same key.
Option A:	Combiner
Option B:	Splitter
Option C:	Partitioner

Option D:	Shuffler
-----------	----------

Q7.	Cartesian product in relational algebra is ____
Option A:	Unary Operation
Option B:	Binary Operation
Option C:	Ternary Operation
Option D:	Aggregation operation

Q8.	Which of the following is not limitation of Hadoop?
Option A:	Security
Option B:	Not suitable for small files
Option C:	Real time data processing not supported
Option D:	Fault Tolerance

Q9.	NoSQL databases are most often referred to as:
Option A:	Relational
Option B:	Distributed
Option C:	Object-oriented
Option D:	Network

Q10.	JSON stands for _____
Option A:	JavaScript Output Name
Option B:	JavaScript Object Notation
Option C:	Java Standard Output Network
Option D:	Java Source Open Network

Q11.	In Column store database _____ provides date and time when data is inserted in database.
Option A:	Name
Option B:	Timestamp
Option C:	Row key
Option D:	Value

Q12.	Which one of the following is not data stream operation?
Option A:	Data collection
Option B:	Data Filtering
Option C:	Data modeling
Option D:	Data classification

Q13.	_____ queries are not predefined. These are executed according to user requirement.
Option A:	Standing
Option B:	System
Option C:	Ad-hoc
Option D:	Transaction

Q14.	Given following input data stream (4,6,3,5) and the hash function (2X mod 5), Which of the following gives correct hash values?
Option A:	(3,1,2,1)
Option B:	(3,2,1,0)
Option C:	(4,5,2,6)
Option D:	(4,1,3,2)

Q15.	In DGIM algorithm, Which of the following is the correct rule for creating Bucket?
Option A:	Every bucket should have atleast one '0'
Option B:	The left end of bucket should be always 1
Option C:	The size of bucket should be power of 2.
Option D:	On moving from right to left, the size of bucket will decrease.

Q16.	Which one of the following is not algorithm used for frequent itemset mining?
Option A:	FM algorithm
Option B:	PCY algorithm
Option C:	Multi-stage algorithm
Option D:	Multi-hash algorithm

Q17.	Which of the following streaming windows show valid bucket representations according to the DGIM rules?
Option A:	1 1 0 0 1 0 0 1 1 0 1 1 0
Option B:	0 0 1 1 0 1 1 1 0 0 1
Option C:	1 1 1 1 0 0 1 0 0 1 1 0 1
Option D:	0 1 1 0 1 0 0 1 0 1 0 1 1 0 0 1 1

Q18.	Two-dimensional Euclidean space and the points (4, 7) and (7, 3).The L2-norm gives a distance of _____
Option A:	6
Option B:	4
Option C:	5
Option D:	2

Q19.	Jaccard Distance between sets x & y is given by _____
Option A:	1+Jaccard Similarity(x,y)
Option B:	1-Jaccard Similarity(x,y)
Option C:	1*Jaccard Similarity(x,y)
Option D:	2-Jaccard Similarity(x,y)

Q20.	What is Jaccard distance between two sets A & B, where A={4,6,7} and B={6,7,8}?
Option A:	0.5
Option B:	0.6
Option C:	0.2
Option D:	0.8

Q21.	What is the Edit Distance between two strings X and Y if X='ABCD' and Y='ACEDF'?
Option A:	2
Option B:	4
Option C:	3
Option D:	1

Q22.	Which filtering is used to recommend items based on a comparison between content of items and a user profile?
Option A:	Collaborative filtering
Option B:	Category filtering
Option C:	Keyword filtering
Option D:	Content based filtering

Q23.	Which one of the following is not a type of Social Network?
Option A:	Collaborative Network
Option B:	Local Area Network
Option C:	E-mail Network
Option D:	Telephone network

Q24.	Which of the following is not Clustering algorithm used in social network?
Option A:	Hierarchical
Option B:	K-means
Option C:	K-medoid
Option D:	Naive bayes

Q25.	Girvan Newman Method is based on the concept of:
Option A:	Node Betweenness
Option B:	Edge Betweenness
Option C:	Node Clustering Coefficient
Option D:	Node Degree