

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 2012

Examination: Final Year Semester VIII

Course Code: CPE8035 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.	The Data structure of Big data is _____
Option A:	Static
Option B:	Dynamic
Option C:	Semi-structured
Option D:	centralised

Q2.	The Pig Latin scripting language has operators similar to _____
Option A:	SQL
Option B:	JSON
Option C:	XML
Option D:	Python

Q3.	What was Hadoop written in ?
Option A:	C
Option B:	C++
Option C:	Java
Option D:	Python

Q4.	_____ stores are used to store information about networks, such as social connections.
Option A:	Key-value
Option B:	Column Store
Option C:	Document Store
Option D:	Graph Store

Q5.	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

Q6.	What is the property of NoSQL in which the system continues to function even if communication between servers fails?
Option A:	Consistency

Option B:	Availability
Option C:	Partition Tolerance
Option D:	Redundancy

Q7.	HBase uses the _____ File System to store its data
Option A:	Hive
Option B:	Impala
Option C:	Hadoop
Option D:	Scala

Q8.	Which of the following phases occur simultaneously ?
Option A:	Reduce and Sort
Option B:	Shuffle and Sort
Option C:	Shuffle and Map
Option D:	Reduce and Map

Q9.	For YARN, _____ manager performs allocation of resources and scheduling.
Option A:	DataNode
Option B:	NameNode
Option C:	Resource
Option D:	Replication

Q10.	Cartesian product in relational algebra is_____
Option A:	Unary Operation
Option B:	Binary Operation
Option C:	Ternary Operation
Option D:	Aggregation operation

Q11.	Which of the following is not a type of Distance Measure?
Option A:	Euclidean distance
Option B:	Cosine Distance
Option C:	Zero distance
Option D:	Edit Distance

Q12.	What is Jaccard distance between two sets A & B, where A={1,2,3,4} and B={2,3,4,5}?
Option A:	0.2
Option B:	0.6
Option C:	0.4
Option D:	0.8

Q13.	Which of the following streaming windows show valid bucket representations according to the DGIM rules?
------	---

Option A:	1100100110110
Option B:	00110111001
Option C:	1111001001101
Option D:	01101001010110011

Q14.	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)

Q15.	_____ distance is the angle made by two vectors from origin to the two points in the space
Option A:	Jaccard Distance
Option B:	Euclidean Distance
Option C:	Cosine Distance
Option D:	Edit Distance

Q16.	In FM algorithm, We use estimate _____ to count number of distinct elements in the given data stream.
Option A:	$2^R$
Option B:	$3^R$
Option C:	$2R$
Option D:	$3R$

Q17.	In DGIM algorithm, Which of the following is not the rule for creating Bucket?
Option A:	The right end of bucket should be always 1
Option B:	The size of bucket should be power of 2.
Option C:	Every bucket should have atleast one '0'
Option D:	On moving from right to left, the size of bucket will increase.

Q18.	The time complexity for FM algorithm is _____
Option A:	$O(\log m)$
Option B:	$O(2n)$
Option C:	$O(n)$
Option D:	$O(3n)$

Q19.	Number of buckets in DGIM algorithm is given by _____
Option A:	$O(\log N)$
Option B:	$O(\log^2 N)$
Option C:	$O(\log 3N)$

Option D:	O(N)
-----------	------

Q20.	Which algorithm we will implement to know how many distinct users visited the website till now or in last 2 hours.
Option A:	DGIM
Option B:	PCM
Option C:	FM
Option D:	SON

Q21.	In Structure of Web, the components that can reach to Strongly connected components but not reachable from strongly connected components are known as _____
Option A:	Out components
Option B:	In components
Option C:	tendrils in
Option D:	tendrils out

Q22.	Identify the property of frequent itemsets which is defined as follows ' If a set of items in a dataset is frequent then all others are its subsets.'
Option A:	Support
Option B:	Distinct
Option C:	Confidence
Option D:	Monotonicity

Q23.	Which of the following is not stream clustering algorithm?
Option A:	COBWEB
Option B:	BDMMO
Option C:	SON
Option D:	C2ICM

Q24.	A _____ matrix is used by recommendation system in which entities are users and items.
Option A:	Sparse Matrix
Option B:	Utility Matrix
Option C:	Similarity Matrix
Option D:	Comparison Matrix

Q25.	A graph is said to be _____ if it does not possess a cycle of an odd length.
Option A:	Bipartite graph
Option B:	Complete Bipartite graph
Option C:	histogram
Option D:	Trigonometric