University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: BE Computer Engineering
Curriculum Scheme: Rev2012
Examination: Fourth Year Semester VIII
Course Code: CPE8035 and Course Name: Big Data Analytics

Time: 1 hour   Max. Marks: 50

===

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

Q1.	Which of the following is an important characteristic of big data to be considered when selecting storage options?
Option A:	Scalability
Option B:	What is on sale
Option C:	Hardware brand
Option D:	There are no important characteristics
Q2.	Hadoop Framework is written in
Option A:	Python
Option B:	Java
Option C:	C++
Option D:	C#
Q3.	Which of the following is not a NoSQL database?
Option A:	SQL Server
Option B:	MongoDB
Option C:	Cassandra
Option D:	Redis
Q4.	Point out the wrong statement :
Option A:	Non Relational databases require that schemas be defined before you can add data
Option B:	NoSQL databases are built to allow the insertion of data without a predefined schema
Option C:	NoSQL is a set of concepts that allows the rapid and efficient processing of datasets with a focus on scalability, performance
Option D:	Developed to handle large amount of data that need to be frequently accessed and processed.
Q5.	Which of the following is not the phase of Reducer?
Option A:	Shuffle
Option B:	Sort
Option C:	Map
Option D:	Reduce
Q6.	What is the full form of YARN?
Option A:	Yet Another Resource Network
Option B:	Yet Another Relational Negotiator

Option C:	Yet Another Resource Negotiator
Option D:	Yet Another Relational Network
Q7.	Which of the following phases occur simultaneously?
Option A:	Shuffle and Sort
Option B:	Reduce and Sort
Option C:	Shuffle and Map
Option D:	Sort and Map
Q8.	If the two strings are same then Edit Distance between two strings is
Option A:	0
Option B:	1
Option C:	2
Option D:	3
Q9.	A Bloom filter guarantees no
Option A:	false positives
Option B:	false negatives
Option C:	false positives and false negatives
Option D:	false positives or false negatives, depending on the Bloom filter type
Q10.	Which of the following statements about standard Bloom filters is correct?
Option A:	it is possible to delete an element from a Bloom filter.
Option B:	A Bloom filter always returns the correct result.
Option C:	It is possible to alter the hash functions of a full Bloom filter to create more space.
Option D:	A Bloom filter always returns TRUE when testing for a previously added element.
Q11.	Which function will associates the value with the key?
Option A:	Get(key)
Option B:	Put(key, value)
Option C:	Multi-get(key1,key2,….,keyN)
Option D:	Delete(key)
Q12.	The manipulation of pagerank done by spammers is called
Option A:	Link analysis
Option B:	Link Spam
Option C:	Trust rank
Option D:	Term rank
Q13.	Transition Matrix contain information of
Option A:	In links only
Option B:	Out links only
Option C:	Both in links and out links
Option D:	Neither in link nor outline

Q14.	HITS algorithm is mainly used to
Option A:	Provide the Rank of the page only with relevance to the current query posed by a user
Option B:	Provide ranking of all the pages
Option C:	Identify and removal of spider traps
Option D:	Identify spam in the web page
Q15.	What does Apriori algorithm do?
Option A:	It mines all frequent patterns through pruning rules with lesser support
Option B:	It mines all frequent patterns through pruning rules with higher support
Option C:	It mines patterns without support
Option D:	During mining according to patterns
Q16.	What technique cannot be used to improve the efficiency of apriori algorithm?
Option A:	Hash-based techniques
Option B:	Transaction Reduction
Option C:	Partitioning
Option D:	K-means technique
Q17.	Which of the following statement is true?
Option A:	Both apriori and FP-Growth uses horizontal data format
Option B:	Both apriori and FP-Growth uses vertical data format
Option C:	Apriori uses horizontal and FP-Growth uses vertical data format
Option D:	Apriori uses vertical and FP-Growth uses horizontal data format
Q18.	Which clustering procedure is characterized by the formation of a tree like structure?
Option A:	Hierarchical clustering
Option B:	Optimizing partitioning
Option C:	Partition based clustering
Option D:	Density clustering
Q19.	What is clustering?
Option A:	Process of grouping similar objects
Option B:	Process of classifying new object
Option C:	Process of ungrouping similar objects
Option D:	Process of declassifying new objects
Q20.	The memory based collaborative filtering uses:
Option A:	It generally uses Naive bayesian system.
Option B:	This mechanism uses user rating data to compute similarity between users or items.
Option C:	Typical examples of this mechanism are neighborhood based CF and item-based/user-based top-N recommendations.

Option D:	It generally uses Exponential system.
Q21.	Identify the correct type(s) of collaborative filtering algorithm
Option A:	Item based.
Option B:	User based.
Option C:	Memory-based.
Option D:	Model-based and hybrid
Q22.	Which one is not an Item-to-Item Collaborative Filtering methods –
Option A:	It is more scalable than User-Based Collaborative Filtering methods
Option B:	Amazon.com uses this method
Option C:	Scales independently of the catalog size or the total number of customers
Option D:	Acceptable performance by creating the expensive similar-item table offline
Q23.	when Social networks are modeled as graphs, what will be represent node?
Option A:	Entities
Option B:	Connections
Option C:	Characters
Option D:	Attributes
Q24.	If we were to apply standard clustering techniques to a social-network graph, our first step would be
Option A:	distance measure.
Option B:	identify the clustering technique
Option C:	comparison of technique
Option D:	Applying standard clustering method
Q25.	Social networks are naturally modelled as:
Option A:	Directed Graphs
Option B:	Uni-directional Graphs
Option C:	Weighted Graphs
Option D:	Hierarchical Graph

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: BE Computer Engineering
Curriculum Scheme: Rev2012
Examination: Fourth Year Semester VIII
Course Code: CPE8035 and Course Name: Big Data Analytics

Time: 1 hour   Max. Marks: 50

==

Question	Correct Option (Enter either 'A' or 'B' or 'C' or 'D')
Q1.	A
Q2.	B
Q3.	A
Q4	A
Q5	C
Q6	C
Q7	A
Q8.	A
Q9.	B
Q10.	D
Q11.	B
Q12.	B
Q13.	C
Q14.	A
Q15.	A
Q16.	D
Q17.	D
Q18.	A
Q19.	A
Q20.	B
Q21.	D
Q22.	D
Q23.	A
Q24.	A
Q25.	A