University of Mumbai
Examination 2020 under cluster 2 (RGIT)

Program: BE **Computer** Engineering

Curriculum Scheme: Revised 2016

Examination: Final Year Semester VII

Course Code: **CSDLO7032** 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.	Mahout Applications provides
Option A:	Command line to invoke various algorithms
Option B:	Gives a platform for building data flow for ETL
Option C:	Supports user defined functions to accomplish specific needs per
Option D:	Is an open-source non-relational distributed database
Q2.	Which of the following tool is designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases?
Option A:	Apache Sqoop
Option B:	Pig
Option C:	Mahout
Option D:	Flume
Q3.	Consider the following statements: Statement 1: Volume refers to the exponential growth in the data storage Statement 2: Variety refers to the connectedness of big data
Option A:	Only statement 1 is true
Option B:	Only statement 2 is true
Option C:	Both statements are true
Option D:	Both statements are false
Q4.	All the following accurately describe Hadoop, EXCEPT _____
Option A:	Open source

Option B:	Real-time
Option C:	Java-based
Option D:	Distributed computing approach
Q5.	Master node of YARN used in Hadoop 2.0 is called as
Option A:	NameNode
Option B:	Node Manager
Option C:	Data Node
Option D:	Resource Manager
Q6.	How many blocks would be created, if a file of size 514 MB is copied to HDFS when the default size is 128MB?
Option A:	4
Option B:	5
Option C:	3
Option D:	6
Q7.	When a master nodes fails
Option A:	The compute node takes over
Option B:	The compute nodes also fails
Option C:	The entire map-reduce job must be restarted
Option D:	None of the above
Q8.	Which of the following operations can't use Reducer as a combiner?
Option A:	Group by Minimum
Option B:	Group by Maximum
Option C:	Group by Count
Option D:	Group by Average
Q9.	Which architecture is more suitable for NoSQL?
Option A:	Shared Nothing
Option B:	Shared Memory
Option C:	Shared Disk
Option D:	Shared Hardware

Q10.	NoSQL databases useful for any business problem that has complex relationships between objects such as Link analysis and social networking
Option A:	Column Store
Option B:	Row store
Option C:	Document store
Option D:	Graph store
Q11.	A key-value store does not have the following property
Option A:	Simple data format makes write and read operations fast
Option B:	Can perform queries on the values of the table
Option C:	Often scale almost linearly with the number of nodes.
Option D:	Values may contain any data type
Q12.	Disadvantages of NoSQL
Option A:	Distributed Computing
Option B:	Schema flexibility, semi-structure data
Option C:	Scalability
Option D:	Limited query capabilities
Q13.	Unlike ACID system BASE system follows the given principle
Option A:	Block any reports while you are working
Option B:	All reports must be consistent and reliable always
Option C:	Allow the system to be basically available and eventually reach a consistent state.
Option D:	Get transaction details right to maintain consistency
Q14.	Bloom Filter is a
Option A:	complex data type
Option B:	sampling efficient data structure
Option C:	source efficient data structure
Option D:	space efficient data structure
Q15.	DGIM Algorithm which uses how many bits to represent a window of N bits

Option A:	$O(\log_2 N+1)$
Option B:	$O(N \log_2 N)$
Option C:	$O(\log_2 N-1)$
Option D:	$O(\log_2 N)$
Q16.	In a DGIM Algorithm estimation of number of 1s in the window with an error of
Option A:	More than 50%
Option B:	Less than 50%
Option C:	Exact 100%
Option D:	Exact 50%
Q17.	DGIM Algorithm storing the number of 1s needed
Option A:	O(log log N-1)bits
Option B:	O(log log N+1)bits
Option C:	O(log log N)bits
Option D:	O(log logN-2)bits
Q18.	What will be the value of **$L\infty$-norm** Considering the two points (10,4) and (6,7) in the two-dimensional Euclidean space
Option A:	4
Option B:	7
Option C:	5
Option D:	25
Q19.	The edit distance using Longest common subsequence of two strings x=PQRSTU and y=SKTUO is
Option A:	3
Option B:	5
Option C:	6
Option D:	11
Q20.	The angle between two points in Cosine Distance will range from:
Option A:	0 to 90 degrees
Option B:	0 to 180 degrees

Option C:	0 to 360 degrees
Option D:	90 to 180 degrees
Q21.	Which of the following is not true for Distance Metric which maps the pairs of objects to real values
Option A:	It is lower when objects are more alike
Option B:	Minimum distance is zero when comparing an object with itself
Option C:	The upper limit may vary
Option D:	The minimum distance may vary if the objects are similar to themselves
Q22.	The modified equation for calculating PageRank is
Option A:	$v' = Mv + \beta e/n$
Option B:	$v' = \beta Mv + (1-\beta)e/n$
Option C:	$v' = (1-\beta)Mv + \beta e/n$
Option D:	$v' = (1-\beta)Mv + (1-\beta)e/n$
Q23.	Which of the following is the advantage of Collaborative-Filtering system?
Option A:	Cold Start
Option B:	Sparsity
Option C:	First rater
Option D:	Works for any kind of item
Q24.	Girvan and Newman proposed a _____ technique for social graphs that use EB as the distance measure
Option A:	Fuzzy clustering
Option B:	Hierarchical divisive clustering
Option C:	Density based clustering
Option D:	Model based clustering
Q25.	The problem of finding all cliques of a given size in a graph is _____ problem.
Option A:	NP-complete
Option B:	NP-hard
Option C:	P

Option D:	NP