

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: BE Computer Engineering

Curriculum Scheme: Rev2016

Examination: Final Year Semester VII

Course Code: **CSDL07032** 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.	What is big data?
Option A:	A centralized information repository that gathers data from many sources
Option B:	The area of science and technology that deals with information sets that are too large to be handled by traditional means.
Option C:	The thirst to gather more and more information.
Option D:	BIG DATA does not require any computational power
Q2.	HDFS is implemented in _____ language.
Option A:	C
Option B:	Perl
Option C:	Python
Option D:	Java
Q3.	What was Hadoop named after?
Option A:	Creator Doug Cutting's favorite circus act
Option B:	Cutting's high school rock band
Option C:	The toy elephant of Cutting's son
Option D:	A sound Cutting's laptop made during Hadoop development
Q4.	Which function is responsible for consolidating the results produced by each of the Map () functions/tasks?
Option A:	Reduce
Option B:	Map
Option C:	Distributed File System
Option D:	Mapper
Q5.	Interface _____ reduces a set of intermediate values which share a key to a smaller set of values.
Option A:	Mapper
Option B:	Reducer
Option C:	Writable
Option D:	Readable
Q6.	Which of the following is used to assign jobs to Task tracker?
Option A:	SlaveNode
Option B:	MasterNode
Option C:	JobTracker
Option D:	Task Tracker

University of Mumbai
Examination 2020 under cluster 4 (PCE)

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.	Which function maps input key/value pairs to a set of intermediate key/value pairs.
Option A:	Mapper
Option B:	Reducer
Option C:	Both Mapper and Reducer
Option D:	Execute
Q9.	__ Stores are used to store information about networks, such as social connections.
Option A:	Key-value
Option B:	Wide-column
Option C:	Document
Option D:	Graph
Q10.	Which function will associate 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)
Q11.	_____ type database stores entities as well the relations amongst those entities.
Option A:	Key-value
Option B:	Wide-column
Option C:	Document
Option D:	Graph
Q12.	What is Standing queries in Stream data model
Option A:	A question asked once about the current state of a stream or streams.
Option B:	Query which is waiting in the queue to get executed
Option C:	Permanently executing and produce outputs at appropriate times
Option D:	Evaluated once over a point of time snapshot of the data set
Q13.	MongoDB scales horizontally using Sharding for _____ purpose .
Option A:	data balancing,
Option B:	load distribution,
Option C:	memory balancing
Option D:	load balancing.

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Q14.	How many passes required in Flajolet-Martin algorithm to find approximates the number of unique objects in a stream
Option A:	One pass
Option B:	Two passes
Option C:	Three passes
Option D:	n passes
Q15.	What is memory requirement of Flajolet-Martin algorithm in data stream mining
Option A:	$O(\log 2N)$
Option B:	$O(\log(n))$
Option C:	$O(N \log 2N)$
Option D:	$O(N^2 \log 2N)$
Q16.	In FM algorithm we shall use estimate.....for the number of distinct elements seen in the stream.
Option A:	2 to the power R
Option B:	3 to the power R
Option C:	2R
Option D:	3R
Q17.	In DGIM, whenever forming a bucket then
Option A:	Every bucket should have at least one 1 else no bucket can be formed
Option B:	Every bucket should have at least two 1, else no bucket can be formed
Option C:	Every bucket should have at least three 1, else no bucket can be formed
Option D:	Every bucket should have at least four 1, else no bucket can be formed
Q18.	Distance in which, the number of bits which differ between two binary strings called as _____
Option A:	Jaccard Distance
Option B:	Edit Distance
Option C:	Hamming Distance
Option D:	Cosine Distance
Q19.	Identify the large scale clustering algorithm which uses a combination of partition based and hierarchical algorithms
Option A:	FM Algorithm
Option B:	PCY Algorithm
Option C:	SON Algorithm
Option D:	CURE Algorithm
Q20.	A Combiner, also known as
Option A:	a semi-reducer
Option B:	a semi-mapper
Option C:	a full reducer
Option D:	Hive
Q21.	Girwan and Newman proposed a hierachical divisive clustering technique for social graphs that use the:

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Option A:	Edge Betweenness as a distance measure
Option B:	Centrality as a distance measure
Option C:	Jaccard distance as a distance measure
Option D:	Euclidean distance as a distance measure
Q22.	The process of identifying similar users and recommending what similar users like is called
Option A:	collaborative filtering.
Option B:	Content-Based systems
Option C:	Page rank.
Option D:	stream filtering
Q23.	Social networks are naturally modeled as:
Option A:	Directed Graphs
Option B:	Uni-Directional Graphs
Option C:	Weighted Graphs
Option D:	Hierarchical Graph
Q24.	Process that identifies dense sub graphs from social network graphs is called
Option A:	mining data stream
Option B:	Community Detection
Option C:	Clustering
Option D:	Map reduce
Q25.	In page Rank computation in a web a Dead Ends are the pages with no _____ in the web graph
Option A:	Trust Rank
Option B:	In links
Option C:	out links
Option D:	Hub Score

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: BE Computer Engineering

Curriculum Scheme: Rev2016

Examination: Final Year Semester VII

Course Code: **CSDLO7032** 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.	B
Q2.	D
Q3.	C
Q4	A
Q5	B
Q6	C
Q7	A
Q8.	A
Q9.	D
Q10.	B
Q11.	D
Q12.	C
Q13.	D
Q14.	A
Q15.	B
Q16.	A
Q17.	A
Q18.	C
Q19.	D
Q20.	A
Q21.	A
Q22.	A
Q23.	A
Q24.	B
Q25.	C