

Question No.	Question Statement	Options			
		A:	B:	C:	D:
1	Data architecture design considers:	Four design layers with the lowest being identification of internal and external data sources	Four design layers with the lowest being acquisition and next identification of internal and external data sources	Five layers with data consumption for analytics, business processes, business intelligence, data mining, pattern recognition and knowledge discovery.	Five layers with data storage processing and analytics being the highest layer.
2	Hadoop is written in which language?	C	C++	Java	Python
3	Which out of below mentioned example indicates NoSql database?	XML	Table	HTNL	Document
4	In Map phase which tasks happen simultaneously	Shuffle and Map	Shuffle and Sort	Shuffle and Reduce	Shuffle and Archived
5	Euclidean distance measure is	A stage of the KDD process in which new data is added to the existing selection	The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them	The distance between two points as calculated using the Pythagoras theorem	The distance between two cities
6	The DGIM IS also refer as	Data, Graph, Information and Management Algorithm	Dattar, Gionis, Indyk and Motwani Algorithm	Database, Group, Information and Message Algorithm	Data, Graph , Inference ,and Message Algorithm,
7	_____ is group of pages with no links out of the group	Spider traps	Dead ends	Hubs	Authorities
8	If an item set 'XYZP' is a frequent item set, then all subsets of that frequent item set are	Undefined	Not frequent	. Frequent	Depends on the transactions
9	The average data point in a cluster with Euclidean data is called	Centroid	Clusteroid	Radius	Diameter
10	A user views the first 13 seconds of a 5 minute video on YouTube, then browses away. What kind of recommendation input is this?	A Rating	A Recommendation	Implicit input	Indication that the user likes the video
11	What is the full form of GFS?	Grep File System	Giant File System	Group File System	Google File System
12	Which of the following is example of document based store?	Redis	Neo4j	CouchDB	Cassandra
13	The optional part of a hadoop jobs or Map Reduce is	Mapper	Reducer	Driver	Partitioner
14	Which of the following will be Manhattan Distance between the two data point A(1,3) and B(2,3)?	1	2	4	8
15	In DGIM algorithm the size of bucket is	Number of 1's inside the bucket	Number of 0's inside the bucket	Number of 1's inside the stream	Number of 1's inside the stream
16	While Implementing Bloom Filter algorithms why there is need of hash function.	Hash Function improves the speed of Data Stream Process.	In Bloom Filter a hash of the elements is added to the set to improve accuracy.	Hash Function cannot be used in Bloom Filter.	Hash Provides parallel processing.
17	Which of the following Search Engines patented the concept of "TrustRank"?	Yahoo	Google	MSN/Bing	Teoma
18	Significant Bottleneck in the Apriori algorithm is	Finding frequent Pairs	Pruning	Finding frequent triples	Number of iterations
19	The PCY algorithm uses the concept of	Pruning	Hashing	Hashing and Pruning	Distributing
20	CURE algorithm discovers Clusters with	Euclidean	Non Euclidean	Both Euclidean and Non Euclidean	Neither Euclidean or Non Euclidean
21	In the CURE initialization phase	All data is considered	A random sample of the data is considered	One point per cluster is used	Pre defined clusters are used
22	What is InputSplit in MapReduce?	Unit of work for single reduce task	Unit of work for single map task	Unit of work for multiple map task	Unit of work for multiple reduce task
23	If number of reducer is set to 0 then which phase only takes to final output	Mapper	Reducer	Combiner	Partitioner
24	Which of the following serve as the master node?	Data Node	NameNode	Data block	Replication
25	Which of the following statements best distinguished predictions from recommendations?	Predictions are yes/no estimates of whether the user will purchase or consume an item; recommendations include information on whether a user will like the item.	Predictions are about future preferences, recommendations are expressions of past preference.	Predictions are computed using content and summary statistics; recommendations use data from other users.	Predictions are estimates of the rating a user would assign; recommendations are lists of candidate items for the user to consider.