

Scheme	R2016
Semester	VII
Course Code	CSDLO7032
Course Name	Big Data Analytics

Question No.	Question	a	b	c	d
1	Which of the following is not true for 5v?	Volume	variable	Velocity	value
2	Hadoop is a framework that works with a variety of related tools. Common hadoop ecosystem include _____	MapReduce, Hummer and Iguana	MapReduce, Hive and HBase	MapReduce, MySQL and Google Apps	MapReduce, Heron and Trumpet
3	_____ is related with an inconsistency possessed by data and this in turn hampers the data analization process or creates hurdle in the way for those wish to analyze this form of data.	Variability	Variety	Volume	Complexity
4	During start up, the _____ loads the file system state from the fsimage and the edits log file.	Datanode	Namenode	Secondary Namenode	Rack awereness policy
5	Which of the following is responsible for managing the cluster resources and use them for scheduling users' applications?	Hadoop Common	YARN	HDFS	MapReduce
6	Which of the following decides the number of partitions that are created on the local file system of the worker nodes?	Number of map tasks	Number of reduce tasks	Number of file input splits	Number of distinct keys in the intermediate key-value pairs
7	Which of the following is correct option for MongoDB	MongoDB is column oriented data store	MongoDB uses XML more in comparison with JSON	MongoDB is a document store database	MongoDB is a key-value data store
8	Which of the following is a column-oriented database that runs on top of HDFS	Hive	Sqoop	Hbase	Flume
9	Which of the following is a NoSQL Database Type ?	SQL	JSON	Document databases	CSV
10	A _____ query Q is a query that is issued once over a database D, and then logically runs continuously over the data in D until Q is terminated.	One-time Query	Standing Query	Adhoc Query	General Query
11	which of the following is not the characterstic of stream data?	Continuous	ordered	persistant	huge
12	Pick a hash function h that maps each of the N elements to at least log ₂ N bits, Estimated number of distinct elements is	2 ^R	2 ^{√(-R)}	1-(2 ^R)	1-(2 ^{√(-R)})
13	About data streaming, Which of the following statements is true?	Stream data is always unstructured data.	Stream data often has a high velocity.	Stream elements cannot be stored on disk.	Stream data is always structured data.
14	The FM-sketch algorithm can be used to:	Estimate the number of distinct elements.	Sample data with a time-sensitive window.	Estimate the frequent elements.	Determine whether an element has already occurred in previous stream data.
15	Which of the following statements about standard Bloom filters is correct?	It is possible to delete an element from a Bloom filter.	A Bloom filter always returns the correct result.	It is possible to alter the hash functions of a full Bloom filter to create more space.	A Bloom filter always returns TRUE when testing for a previously added element.
16	The police set up checkpoints at randomly selected road locations, then inspected every driver at those locations. What type of sample is this?	Simple Random Sample	Startified Random Sample	Cluster Random Sample	Uniform sampling
17	What is the finally produced by Hierarchical Agglomerative Clustering?	final estimate of cluster centroids	assignment of each point to clusters	tree showing how close things are to each other	Group of clusters
18	Which of the following is not the class of points in BFR algorithm	Discard Set (DS)	Compression Set (CS)	Isolation Set (IS)	Retained Set (RS)
19	The Jaccard similarity of two non-binary sets A and B, is defined by _____	Jaccard Index	Primary Index	Secondary Index	Clustered Index
20	Following is based on grid like street geography of the New York:	Manhattan Distance	Edit Distance	Hamming distance	Lp distance
21	Techniques for fooling search engines into believing your page is about something it is not, are called _____.	term spam	page rank	phishing	dead ends
22	The graphical representation of an SNA is made up of links and _____.	People	Networks	Nodes	Computers

23	Which algorithm is used to find fully connected subgraph in social media mining?	CURE	CPM	SimRank	Girvan-Newman Algorithm
24	Effect of Spider trap on page rank	particular page get the highest page rank	All the pages of web will get 0 page rank	no effect on any page	affects a particular set of pages
25	_____ systems focus on the relationship between users and items for recommendation.	DGIM	Collaborative-Filtering	Content Based and Collaborative Filtering	Content Based