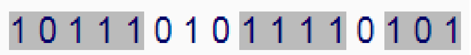
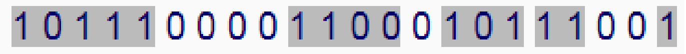
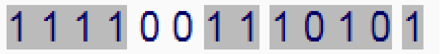
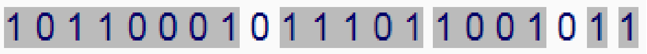
Program: **Computer Engineering**
Curriculum Scheme: Rev2016
Examination: BE Semester VII
Course Code: CSDLO7032 and Course Name: Big Data Analytics

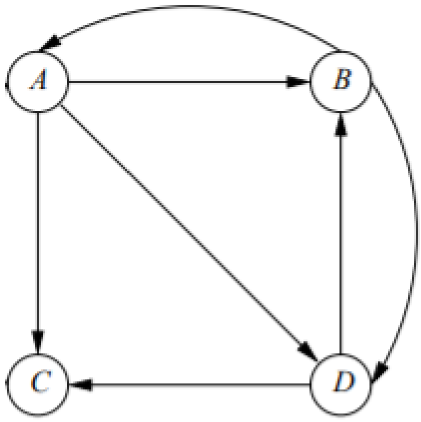
Time: 2 hour Max. Marks: 80

Q1.	**Choose the correct option for following questions. All the Questions are compulsory and carry equal marks**
1.	Which one of the following is not a Hadoop limitation
Option A:	Parallel Processing
Option B:	High Availability
Option C:	Multiple DataCenters
Option D:	Security
2.	"Using Jacard similarity identify how similar these two sets are? A = {0,1,2,5,6} B = {0,2,3,4,5,7,9}"
Option A:	0.33
Option B:	3
Option C:	0
Option D:	0.5
3.	Which of the following for Stream data is true
Option A:	They need not have the same data rates or data types, and the time between elements of one stream need not be uniform.
Option B:	The time between elements of one stream must be uniform
Option C:	They are all of same data types
Option D:	They arrive at same data rate
4.	The source of HDFS architecture in Hadoop originated from which of the following?
Option A:	Facebook distributed file system
Option B:	Yahoo distributed file system
Option C:	Google distributed file system
Option D:	Amazon distributed file system
5.	______ can be used to describe nodes that contain the most amount of information about a network.
Option A:	Social Networks
Option B:	Betweeness Centrality
Option C:	Degree Centrality
Option D:	Broadcasters
6.	________________________ systems recommend items based on similarity measures between users and/or items.
Option A:	Content-based filtering
Option B:	General filtering
Option C:	Collaborative Filtering
Option D:	User-based filtering

7.	PageRank is a function that assigns a _______________
Option A:	number that is number of outgoing links on a page.
Option B:	number that is number of distinct words on a page.
Option C:	real number to each page in the web, based on its importance.
Option D:	number that is number of incoming links on a page.
8.	Which of the following is not a high-level drivers associated with NoSQL movement
Option A:	Agility
Option B:	Velocity
Option C:	Volume
Option D:	Varsity
9.	Suggest appropriate type of distance measure for plagiarism detection?
Option A:	Edit Distance
Option B:	Cosine distance
Option C:	Jaccard Distance
Option D:	Hamming Distance
10.	For Counting Ones in a window which of the following algorithm is used
Option A:	The Datar-Gionis-Indyk-Motwani
Option B:	The Flajolet-Martin Algorithm
Option C:	Bloom's Algorithm
Option D:	The CURE Algorithm
11.	Which of the following is not a reason NoSQL has become a popular solution for some organizations?
Option A:	Better scalability
Option B:	Improved ability to keep data consistent
Option C:	Faster access to data than relational database management systems (RDBMS)
Option D:	More easily allows for data to be held across multiple servers
12.	Which of the following is an example of key value store
Option A:	Azure Table Storage
Option B:	Neo4j
Option C:	Cassandra
Option D:	Accumulo
13.	Which of the following streaming windows show valid bucket representations according to the DGIM rules?
Option A:	 1 0 1 1 1 0 1 0 1 1 1 1 0 1 0 1
Option B:	 1 0 1 1 1 0 0 0 0 1 1 0 0 0 1 0 1 1 1 0 0 1
Option C:	 1 1 1 1 0 0 1 1 1 0 1 0 1
Option D:	 1 0 1 1 0 0 0 1 0 1 1 1 0 1 1 0 0 1 0 1 1

14.	Consider a 2*2 Matrices A and B and whose values are A[1,2,3,4] { here 1, 2 is 1st row values and 3,4 is the 2nd rows values} and B[5,6,7,8] { here 5,6 is the 1st row values and 7,8 is the 2nd rows values}. To perform matrix vector multiplication by map reduce algorithm what will be the correct value will be chosen after evaluating the Map() function when value of k=1,Matrix=B and j=1
Option A:	5
Option B:	7
Option C:	8
Option D:	6
15.	Consider a stream as: S = {1, 2, 1, 3} Let hash function be 2x + 2 mod 4, find the no. of distinct elements.
Option A:	4
Option B:	2
Option C:	8
Option D:	1
16.	Which statement is true about a Stream-Clustering Algorithm
Option A:	size of a bucket is the number of points it represents
Option B:	the points of the stream are partitioned into, and summarized by, buckets whose sizes are a power of two.
Option C:	the sizes of buckets obey the restriction that there are one or two of each size, up to some limit
Option D:	None of the mentioned
17.	The FM uses the no. 0's the binary hash value ends in to make an estimation. Which statement is correct about the hash tail?
Option A:	Any specific bit pattern is equally suitable to be used as hash tail.
Option B:	Only bit patterns with more 0's than 1's equally suitable to be used as hash tails.
Option C:	Only the bit patterns 0000000..00 (list of 0s) or 111111..11 (list of 1s) are suitable hash tails.
Option D:	Only the bit pattern 0000000..00 (list of 0s) is a suitable hash tail.
18.	What is the aim of NoSQL?
Option A:	NoSQL provides an alternative to SQL databases to store textual data.
Option B:	NoSQL databases allow storing non-structured data.
Option C:	NoSQL is not suitable for storing structured data.
Option D:	NoSQL is a new data format to store large datasets.
19.	_______________function processes a key/value pair to generate a set of intermediate key/value pair.
Option A:	Map
Option B:	Reducer
Option C:	Map and Reduce
Option D:	Partition
20.	CURE algorithm
Option A:	Dose not assume anything about the shape of the cluster
Option B:	clusters have a spherical-like shape
Option C:	clusters have a round shape
Option D:	clusters have a square shape

Q2	**Solve any Four out of Six** **5 marks each**
A	Describe the structure of HDFS in a Hadoop ecosystem using a diagram..
B	When it comes to big data how NoSQL scores over RDBMS.
C	What is Recommendation system? Explain Content based recommendation system.
D	Explain with block diagram architecture of data stream management system.
E	What are the Challenges in clustering Data stream. Explain Stream Clustering algorithm.
F	Explain with example Hubs and Authorities in detail

Q3.	**Solve any Two Questions out of Three** **10 marks each**
A	Compute the page rank of each page with teleportation factor beta value β=0.8 
B	What is the role of JobTracker and TaskTracker in MapReduce. Illustrate MapReduce execution pipeline with wordcount example.
C	For the graph given below use Clique percolation method and find all communities.