University of Mumbai
Examination June 2021
Examinations Commencing from 1st June 2021
Program: **Information Technology**
Curriculum Scheme: Rev2016
Examination: BE Semester VIII
Course Code: ITC801 and Course Name: BDA

Time: 2 hour   Max. Marks: 80

==

Q1.	Choose the correct option for following questions. All the Questions are compulsory and carry equal marks
1.	Traditional IT systems provide a foundation for ------------ when they're integrated with big data technologies like Hadoop
Option A:	**Big data management and data mining**
Option B:	Data warehousing and business intelligence
Option C:	Management of Hadoop clusters
Option D:	Collecting and storing unstructured data
2.	State the main component of Big Data
Option A:	**MapReduce**
Option B:	MySQL
Option C:	Google Apps
Option D:	Trumpet
3.	Which of the following platforms does Hadoop run on
Option A:	Bare metal
Option B:	Debian
Option C:	**Cross-platform**
Option D:	Unix-like
4.	The data from the CCTV footage and weather forecast report is
Option A:	structured data
Option B:	**unstructured data**
Option C:	semi- structured data
Option D:	hybrid data
5.	Which is a column-oriented database running on top of HDFS
Option A:	Hive
Option B:	Sqoop
Option C:	**HBase**
Option D:	Flume
6.	Which of the following language is MongoDB not written in?
Option A:	C++.
Option B:	Python
Option C:	Go
Option D:	**MATLAB**
7.	Hive was developed by
Option A:	**Facebook**
Option B:	Google
Option C:	MongoDB.inc
Option D:	Amazon
8.	Which of these is graph based schema

Option A:	Cassandra
Option B:	couch DB
Option C:	**Neo4J**
Option D:	Hbase
9.	The three main Challenges for mining data streams are :
Option A:	Integrity, consistency, durability
Option B:	volume,integrity,isolation
Option C:	**volume,velocity,volatility**
Option D:	concurrency,redundancy,integrity
10.	In Project operation for Map Reduce , the reduce function is used for :
Option A:	**duplicate elimination**
Option B:	selection
Option C:	to include duplicates
Option D:	for joining two columns
11.	State the phenomenon that occurs because of changes in behavior of the data itself
Option A:	concept ACID
Option B:	concept CAP
Option C:	**concept drift**
Option D:	ACID and CAP
12.	The term False Positive means
Option A:	Probably Negative
Option B:	**Probably Positive**
Option C:	Surely Positive
Option D:	Surely Negative
13.	The Space Complexity required for DGIM algorithm is :
Option A:	O(log N +1)
Option B:	O(log N+2)
Option C:	**O (log^2N)**
Option D:	O (log N)
14.	A Bloom filter guarantees no
Option A:	false positives
Option B:	**false negatives**
Option C:	false positives and false negatives
Option D:	false positives or false negatives, depending on the Bloom filter type
15.	This algorithm performs the approximation of the number of single objects in a one-pass path.
Option A:	**Flajolet Martin**
Option B:	DGIM
Option C:	Parallel SVM
Option D:	K-means
16.	Which one among the following is a space efficient probabilistic data structure
Option A:	Flajolet-Martin
Option B:	Sampling of Streams
Option C:	**Bloom Filters**
Option D:	Counting Distinct Elements

17.	The full form of HITS
Option A:	Hyper induced topic search
Option B:	**Hyperlink induced topic search**
Option C:	Highlink induced topic search
Option D:	Hublink induced topic search
18.	Pagerank is developed by Lary page and _____
Option A:	**Sergey Brin**
Option B:	Stev Jobs
Option C:	Ronald Wayne
Option D:	Stev wozniak
19.	HITS algorithm was developed by _____
Option A:	**Jon Kleinberg**
Option B:	Larry page
Option C:	Sergey Brin
Option D:	David Filo
20.	Which of the following search engines patented the concept of "trustrank" as a methodology for ranking websites and pages
Option A:	**Yahoo**
Option B:	Google
Option C:	MSN/bing
Option D:	Teoma

Q2	**Solve any Two Questions out of Three**	**10 marks each**
A	*Explain HDFS architecture with diagram.*	
B	Clearly explain how cure algorithm can be used to cluster big data sets.	
C	Explain PageRank algorithm with suitable example.	

Q3	**Solve any Two Questions out of Three**	**10 marks each**
A	*Explain Hadoop ecosystem with core components.*	
B	Suppose a data stream consist of the integers 1,3,2,1,2,3,4,3,1,2,3,1.Let the hash function being used is h(x)=(6x+1)mod5; estimate the number of distinct in this stream using FM algorithm.	
C	Explain collaborative filtering system. How it is different from content based.	