Program: BE Information Technology
Curriculum Scheme: Revised 2016
Examination: Final Year Semester VIII
Course Code: ITC801 and Course Name: Big Data Analytics
Time: 1 hour
Max Marks:50

1 What is the recommended best practice for managing big data analytics programs?
(a) Adopting data analysis tools based on a laundry list of their capabilities
(b) Letting go entirely of "old ideas" related to data management
(c) Stick with old goals
(d) Focusing on business goals and how to use big data analytics technologies to meet them

2 What percentage of digital information is generated by individual?
(a) 100%
(b) 75%
(c) 50%
(d) 25%

3 For what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop?
(a) Big data management and data mining
(b) Data warehousing and business intelligence
(c) Management of Hadoop clusters
(d) Collecting and storing unstructured data

4 Which is not the feature of Big Data Analytics?
(a) Open-Source
(b) Scalability
(c) Data Recovery
(d) Reduction of data

5 What is not among the V's of Big Data?
(a) Value
(b) Veracity
(c) Volume
(d) Void

6 The examination of large amounts of data to see what patterns or other useful information can be found is known as
(a) Data examination
(b) Information analysis
(c) Big data analytics
(d) Data analysis

7 What makes Big Data analysis difficult to optimize?
(a) Big Data is not difficult to optimize
(b) Both data and cost effective ways to mine data to make business sense out of it
(c) The technology to mine data
(d) MapReduce

8 the best way to accomplish this is by setting aside some of your data in a vault to isolate it from the mining process; once the mining is complete, the results can be tested against the isolated data to confirm the model's _____.
(a) Validity
(b) Integrity
(c) Security
(d) Consistency

9 During business hours, most _____ systems should probably not use parallel execution.
(a) OLAP
(b) DSS
(c) OLTP
(d) Data Mining

10 The goal of ideal parallel execution is to completely parallelize those parts of a computation that are not constrained by data dependencies. The smaller the portion of the program that must be executed _____, the greater the scalability of the computation.
(a) in pipeline
(b) in parallel
(c) sequentially
(d) distributed

11 The goal of ideal parallel execution is to completely parallelize those parts of a computation that are not constrained by data dependencies. The _____ the portion of the program that must be executed sequentially, the greater the scalability of the computation.
(a) Larger
(b) Smaller
(c) Unambiguous
(d) Superior

12 Non uniform distribution, when the data is distributed across the processors, is called _____.
(a) Skew in Partition
(b) Pipeline Distribution
(c) Distributed Distribution
(d) Uncontrolled Distribution

13 Relational databases allow you to navigate the data in _____ that is appropriate using the primary, foreign key structure within the data model.
(a) Only One Direction
(b) Any Direction
(c) Two Direction
(d) Three Direction

14 The need to synchronize data upon update is called
(a) Data Manipulation
(b) Data Replication
(c) Data Coherency
(d) Data Imitation

15 Data Ingestion is
(a) Extraction of data from various sources
(b) Storage of data
(c) Cleaning of data
(d) Partitioning the data
16 Which is not the feature of Big Data Analytics?
(a) Open source
(b) Scalability
(c) Recoverability
(d) Transaction Handling
17 The Problem with HDFS is it cannot handle
(a) Large File
(b) Very Large File
(c) Small Files
(d) Huge Files
18 Following Tool is used for job coordination and workflow management execution in Hadoop
(a) Sqoop
(b) Zookeeper
(c) Mahout
(d) Oozie
19 Which of the following phases occur simultaneously in Map Reduce?
(a) Shuffle and Map
(b) Reduce and Sort
(c) Reduce and Map
(d) Shuffle and Sort
20 HDFS is implemented in following language
(a) Java
(b) C
(c) C++
(d) Hadoop
21 HBase uses the following File System to store its data.
(a) EXT
(b) NTFS
(c) scala
(d) Hadoop
22 Unlike to RDBMS, Hadoop provides following
(a) Provide ACID transactions
(b) Maintain higher data Integrity
(c) Handling unstructured and semi-structured data
(d) Perform read and write many times
23 MongoDB Uses Following formant to store data.
(a) Tree
(b) Graph
(c) Key-Value
(d) Document database
24 Out of following Which is a NoSQL Database Type?
(a) Mysql
(b) Document databases
(c) Oracle
(d) SQL
25 Which of the following are the simplest NoSQL databases?
(a) Sql
(b) Mysql
(c) RDBMS
(d) Key-value
26 The Default HDFS block size is
(a) 128KB
(b) 128MB
(c) 128GB
(d) 128TB
27 Which is the format supported by MongoDB?
(a) BSON
(b) JSON
(c) SQL
(d) MYSQL
28 Following is the best reason to use a NoSQL database
(a) When Retrieval of large quantities of data is needed
(b) When data is not available
(c) When Data is in structured format
(d) When Data is missing
29 Mostly NoSql data is referred as
(a) ORDBMS
(b) Parallel
(c) Sequential
(d) Distributed
30 Apache Hadoop run on following which platforms ?
(a) Cross-platform
(b) Unix
(c) Windows
(d) IOS
31 Following represent column in NoSQL
(a) Field
(b) Row
(c) Tuple
(d) Database

32 Following can be best described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.

(a) Hive
(b) Hbase
(c) Sqoop
(d) Map Reduce

33 Following is not a NoSQL database

(a) Cassandra
(b) MongoDB
(c) SQL Server
(d) Key value

34 Consider a text file with following words:
abcd
acdb
bcda
xyz
yzx
zxy

Suppose we need to find anagrams of words from this file. What will be the output of the map phase?

(a) (abcd, xyz), (abcd,yzx), (abcd,zxy), (xyz,abcd), (xyz,acdb), (xyz,bcda)
(b) (abcd, 1), (acdb,1), (bcda,1), (xyz,1), (yzx,1), (zxy,1)
(c) (abcd, 3), (xyz,3)
(d) (abcd, abcd), (abcd, acdb), (abcd, bcda), (xyz, xyz), (xyz, yzx), (xyz, zxy)

35 Consider a text file with following words:
abcd
acdb
bcda
xyz
yzx
zxy

Suppose we need to find anagrams of words from this file. What will be the input to the reduce function?

(a) (abcd, abcd), (abcd, acdb), (abcd, bcda), (xyz, xyz), (xyz, yzx), (xyz, zxy)
(b) (abcd, xyz), (abcd,yzx), (abcd,zxy), (xyz,abcd), (xyz,acdb), (xyz,bcda)
(c) (abcd, [abcd,acdb,bcda]),(xyz, [xyz,yzx, zxy])
(d) (xyz, [xyz,yzx, zxy]), (abcd, [abcd,acdb,bcda])

36 Consider a text file with following words.
team
meat
tame
but
tub

Suppose we need to find anagrams of words from this file. What will be the output of the map phase?

(a) (team, but), (team, tub), (meat, but), (meat, tub), (tame, but), (tame, tub)
(b) (team, 1), (meat,1), (tame,1), (but,1), (tub,1)
(c) (team, team), (team, meat), (team, tame), (but, but), (but, tub)
(d) (aemt, team), (aemt , meat), (aemt, tame), (btu, but), (btu, tub)

37 Consider a text file with following words.
team
meat
tame
but
tub

Suppose we need to find anagrams of words from this file. What will be the input to the reduce function?

(a) (but, [but, tub]), (team, [team, meat, tame])
(b) (team, [team, meat, tame]), (but, [but, tub])
(c) (aemt, [team, meat, tame]), (btu, [but, tub])
(d) (btu, [but, tub]), (aemt, [team, meat, tame])

38 For each tuple t in R, construct a tuple t' by eliminating from t those components whose attributes are not in S. Output the key-value pair (t', t') in map function. For each key t' produced by any of the Map tasks, there will be one or more key-value pairs (t', t'). After the system groups key-value pairs by key, the Reduce function turns (t', [t', t', … t']) into (t', t'), so it produces exactly one pair (t', t') for this key t'. This logic is applicable to which of the following relational algebra operations?

(a) Projection
(b) Selection
(c) Union
(d) Intersection

39 Consider the following text.

This is BEIT. BEIT has four courses.

In a program to find frequency of occurrence of words in this file, what is the output of the map function?

(a) (1, This), (is,1), (BEIT., 1), (BEIT,1), (has,1), (four,1), (courses.,1)
(b) (This,1), (is,1), (BEIT., 1), (BEIT,1), (has,1), (four,1), (courses.,1)
(c) (1, This), (1, is), (1, BEIT), (1, BEIT), (1, has), (1, four), (1, courses)
(d) (This,1), (is,1), (BEIT,2), (has,1), (four,1), (courses.,1)

40 When using Map-Reduce for matrix vector multiplication if the vector V cannot fit in main memory then
(a) Map Reduce is not a solution
(b) Eliminate the sparse entries and try accommodating the vector
(c) Increase the size of main memory
(d) Divide the matrix into vertical strips of equal width and divide the vector into an equal number of horizontal strips

41 You have a mapper that for each key produces an integer value and the following set of reduce operations:
Reducer A: outputs the sum of the set of integer values
Reducer B: outputs the maximum of the set of values
Reducer C: outputs the mean of the set of values.
Reducer D: outputs the difference between the largest and smallest values in the set.
Which of these reduce operations could safely be used as a combiner?
(a) A and B
(b) A, B and D
(c) C and D
(d) C, B and D

42 In which of the following cases, a combiner can't be used safely?
(a) In finding count of number of words in a file
(b) In finding yearly maximum temperature from given set of temperature values
(c) In finding yearly minimum temperature from given set of temperature values
(d) In finding median of set of prices in a stored file

43 Which of the following partitioner is used by default in Hadoop to partition key space?
(a) BinaryPartitioner
(b) KeyFieldBasedPartitioner
(c) TotalOrderPartitioner
(d) HashPartitioner

44 A partitioner is created when:
(a) There are multiple mappers
(b) There are no mappers
(c) There are multiple reducers
(d) There is a single reducer

45 What will be the output of the Map function for the below given relation to select all the records whose designation is 'Manager'. Emp ID-Name-Designation
1001-AAA-Manager
1002-BBB-VP
1003-CCC-Manager
1004-DDD-AVP
(a) (((1001,AAA,Manager),(1001,AAA,Manager),((1002,BBB,VP),(1002,BBB,VP)),
((1003,CCC,Manager),(1003,CCC,Manager)),((1004,DDD,AVP),(1004,DDD,AVP)))
(b) ((1001,AAA,Manager),(1001,AAA,Manager),((1003,CCC,Manager),(1003,CCC,Manager))
(c) (((1001,AAA,Manager),(1001,AAA,Manager),((1002,BBB,VP),(1002,BBB,VP)),
((1003,CCC,Manager),(1003,CCC,Manager)))
(d) ((1001,AAA,Manager),(1001,AAA,Manager))

46 Consider the MapReduce's WordCount example. Let's now assume that you want to determine the frequency of phrases consisting of 5 words each instead of determining the frequency of single words. Which part of the pseudo-code do you need to adapt?
(a) Only reduce()
(b) Only map()
(c) Both map() and reduce()
(d) Combiner()

47 What will be the output of the Map function for the below given relation to select only all the designations available in an organizations? Emp ID-Name-Designation
1001-AAA-Manager
1002-BBB-VP
1003-CCC-Manager
1004-DDD-AVP
(a) (Manager, Manager), (VP, VP), (Manager, Manager), (AVP, AVP)
(b) (Manager, Manager), (VP, VP), (AVP, AVP)
(c) (1001, Manager), (1002, VP), (1003, Manger), (1004, AVP)
(d) (AAA, Manager), (BBB, VP), (CCC, Manger), (DDD, AVP)

48 Consider Relation R as: Student_id-Student_name-Advisor_id
1-Student1-1
2-Student2-5
5-Student5-3
7-Student7-3 9-Student9-1 10-Student10-5
Consider Relation S as: Advisor_id-Advisor_name 1-Advisor1
3-Advisor3 5-Advisor5 Suppose we need to find student names along with their advisor names using MapReduce. What is the output of map phase?
(a) (1,(R,Student1)), (2,(R,Student2)),(2,(R,Student4)),(3,(R,Student5)),(3,(R,Student7))
(1,(R,Student9)),(3,(R,Student10))(1,(S,Advisor1)),(3,(S,Advisor3)),(5,(S,Advisor5))
(b) (1,(R,Student1)), (2,(R,Student2)),(5,(R,Student5)),(7,(R,Student7))
(9,(R,Student9)),(10,(R,Student10))(1,(S,Advisor1)),(3,(S,Advisor3)),(5,(S,Advisor5))
(c) (1,(R,Student1)), (5,(R,Student2)),(3,(R,Student5)),(3,(R,Student7))
(1,(R,Student9)),(5,(R,Student10))(1,(S,Advisor1)),(3,(S,Advisor3)),(5,(S,Advisor5))
(d) (1,(R,Student1)), (2,(R,Student2)),(5,(R,Student5)),(3,(R,Student7))
(9,(R,Student9)),(5,(R,Student10))(1,(S,Advisor1)),(3,(S,Advisor3)),(5,(S,Advisor5))

49 Consider two matrices in the format
M= $[m_{11},m_{12},m_{21},m_{22}]$ N=$[n_{11},n_{12},n_{21},n_{22}]$ where m_{11} can be interpreted as value in the first row and first column of matrix M and so on. The actual matrices are: M=[1,9,5,4] and N=[4,3,6,7]
In case of multiplication of M and N using 2 phase MapReduce, what will be the output of the first map phase?

(a) (M,1,1),(M,2,5),(M,1,9),(M,2,4),(N,1,4),(N,2,3),(N,1,6),(N,2,7)
(b) (1,(M,1,1),(1,(M,2,5)),(2,(M,1,9)),(2,(M,2,4)),(1,(N,1,4),(1,(N,2,3)),(2,(N,1,6)),(2,(N,2,7))
(c) (1,(M,1,1),(1,(M,2,6)),(2,(M,1,9)),(2,(M,2,4)),(1,(N,1,4),(1,(N,2,3)),(2,(N,1,5)),(2,(N,2,7))
(d) (M,1,1),(M,2,6),(M,1,3),(M,2,4),(N,1,4),(N,2,9),(N,1,5),(N,2,7)

50 Consider two matrices in the format M= [m_{11},m_{12},m_{21},m_{22}] N=[n_{11},n_{12},n_{21},n_{22}] where m_{11} can be interpreted as value in the first row and first column of matrix M and so on. The actual matrices are: M=[1,9,5,4] and N=[4,3,6,7]
In case of multiplication of M and N using 2 phase MapReduce, what will be the output of the first reduce phase?

(a) (1,1,1),(1,2,3),(2,1,20),(2,2,15),(1,1,4),(1,2,63),(2,1,24),(2,2,28)
(b) ((1,1),4),((1,2),3),((2,1),20),((2,2),15),((1,1),54),((1,2),63),((2,1),24),((2,2),28)
(c) ((1,1),1),((1,2),9),((2,1),5),((2,2),4),((1,1),4),((1,2),3),((2,1),6),((2,2),7)
(d) (1,1,4),(1,2,33),(2,1,20),(2,2,15),(1,1,54),(1,2,3),(2,1,24),(2,2,28)

51 Consider two matrices in the format M= [m_{11},m_{12},m_{21},m_{22}] N=[n_{11},n_{12},n_{21},n_{22}] where m_{11} can be interpreted as value in the first row and first column of matrix M and so on. The actual matrices are: M=[1,9,5,4] and N=[4,3,6,7]
In case of multiplication of M and N using 2 phase MapReduce, what will be the output of the second map phase?

(a) ((1,1),4),((1,2),3),((2,1),20),((2,2),15),
((1,1),54),((1,2),63),((2,1),24),((2,2),28)
(b) (2,[(1,1,4),(1,2,3),(2,1,20),(2,2,15)]),(1,[(1,1,54),(1,2,63),(2,1,24),(2,2,28)])
(c) (1,[(1,1,4),(1,2,3),(2,1,20),(2,2,15)]),(2,[(1,1,54),(1,2,63),(2,1,24),(2,2,28)])
(d) ((1,1),24),((1,2),3),((2,1),54),((2,2),15)
((1,1,20),((1.2),63),(2,1),4),((2,2),28)

52 Consider following relation as: Customer_id-Name-Country
1-Alfred-Germany
2-Ana-Mexico
3-Antonio-Mexico
4-Thomas-UK 5-Christina-Sweden
What will be the output of the map function to find all records having Country as Mexico?

(a) (((1,Alfred,Germany), (1,Alfred,Germany),((2,Ana,Mexico),(2,Ana,mexico)),
((3,Antonio,Mexico),(3,Antonio,Mexico)),((4,Thomas,UK),(4,Thomas,UK)),((5,Christina,Sweden),(5,Christina,Sweden))
(b) (((1,Alfred,Germany), (1,Alfred,Germany),((2,Ana,Mexico),(2,Ana,mexico)),
((3,Antonio,Mexico),(3,Antonio,Mexico)))
(c) ((2,Ana,Mexico),(2,Ana,mexico))
(d) ((2,Ana,Mexico),(2,Ana,Mexico)),((3,Antonio,Mexico),(3,Antonio,Mexico))

53 Consider two matrices in the format M= [m_{11},m_{12},m_{21},m_{22}] N=[n_{11},n_{12},n_{21},n_{22}] where m_{11} can be interpreted as value in the first row and first column of matrix M and so on. The actual matrices are: M=[3,6,2,4] and N=[2,3,5,7]
In case of multiplication of M and N using 2 phase MapReduce, what will be the output of the first map phase?

(a) (1,(M,1,3),(1,(M,2,2)),(2,(M,1,6)),(2,(M,2,4)),(1,(N,1,2),(1,(N,2,3)),(2,(N,1,5)),(2,(N,2,7))
(b) (M,1,1),(M,2,5),(M,1,9),(M,2,4),(N,1,4),(N,2,3),(N,1,6),(N,2,7)
(c) (1,(M,1,1),(1,(M,2,5)),(2,(M,1,9)),(2,(M,2,4)),(1,(N,1,4),(1,(N,2,3)),(2,(N,1,6)),(2,(N,2,7))
(d) (M,1,1),(M,2,5),(M,1,9),(M,2,4),(N,1,7),(N,2,6),(N,1,3),(N,2,4)

54 Consider two matrices in the format M= [m_{11},m_{12},m_{21},m_{22}] N=[n_{11},n_{12},n_{21},n_{22}] where m_{11} can be interpreted as value in the first row and first column of matrix M and so on. The actual matrices are: M=[3,6,2,4] and N=[2,3,5,7]
In case of multiplication of M and N using 2 phase MapReduce, what will be the output of the first reduce phase?

(a) (1,1,13),(1,2,9),(2,1,4),(2,2,6),(1,1,30),(1,2,42),(2,1,20),(2,2,28)
(b) ((1,1),6),((1,2),9),((2,1),4),((2,2),6),((1,1),30),((1,2),42),((2,1),20),((2,2),28)
(c) ((1,1),3),((1,2),6),((2,1),2),((2,2),4)]),((1,1),2),((1,2),3),((2,1),5),((2,2),7)
(d) (1,1,63),(1,2,9),(2,1,20),(2,2,28),(1,1,30),(1,2,42),(2,1,4),(2,2,6)

55 Consider two matrices in the format M= [m_{11},m_{12},m_{21},m_{22}] N=[n_{11},n_{12},n_{21},n_{22}] where m_{11} can be interpreted as value in the first row and first column of matrix M and so on. The actual matrices are: M=[3,6,2,4] and N=[2,3,5,7]
In case of multiplication of M and N using 2 phase MapReduce, what will be the output of the second map phase?

(a) (1,[(1,1,6),(1,2,9),(2,1,4),(2,2,6)]),(2,[(1,1,30),(1,2,42),(2,1,20),(2,2,28)])
(b) (2,[(1,1,6),(1,2,9),(2,1,4),(2,2,6)]),(1,[(1,1,30),(1,2,42),(2,1,20),(2,2,28)])
(c) ((1,1),6),((1,2),9),((2,1),4),((2,2),6),((1,1),30),((1,2),42),((2,1),20),((2,2),28)
(d) ((1,1),30),((1,2),9),((2,1),20),((2,2),6),((1,1,6),((1,2),42),(2,1),2),((2,2),28)

56 Consider Relation A as: Roll No.-Name
234-Mark
235-Steve
236-Harry
Consider Relation B as: Roll No.-Name 236-Harry
237-James 238-Jessica Suppose we are finding students present only in Table A and not in Table B using MapReduce. What will be the output of map function?

(a) ((234,Mark),(234,Mark)),((235,Steve),(235,Steve))
(b) ((234,Mark),A),((235,Steve),A),((236,Harry),A),((236,Harry),B),((237,James),B),((238,Jessica),B)
(c) ((234,Mark),B),((235,Steve),A),((236,Harry),A),((236,Harry),B),((237,James),A),((238,Jessica),B)
(d) ((236,Harry),(236,Harry)),((237,James),(237,James))

57 Consider Relation A as: Roll No.-Name
234-Mark
235-Steve
236-Harry
Consider Relation B as: Roll No.-Name 236-Harry
237-James 238-Jessica Suppose we are finding students present in both Table A and Table B using MapReduce. What will be the output of map function?

(a) ((234,Mark),(234,Mark)),((235,Steve),(235,Steve)),((236,Harry),(236,Harry)),((236,Harry),(236,Harry)),((237,James),(237,James),((238,Jessica),(238,Jessica))
(b) ((236,Harry),(236,Harry)), ((236,Harry),(236,Harry))
(c) ((236,Harry),(236,Harry))
(d) Harry

58 Consider Relation A as: Roll No.-Name
234-Mark
235-Steve
236-Harry
Consider Relation B as: Roll No.-Name 236-Harry
237-James 238-Jessica Suppose we are finding students present in either Table A or Table B or both using MapReduce. What will be the output of map function?

(a) ((236,Harry),(236,Harry)), ((235,Steve),(235,Steve)),((234,Mark),(234,Mark))
(b) ((236,Harry),(236,Harry))
(c) Harry
(d) ((234,Mark),(234,Mark)),((235,Steve),(235,Steve)),((236,Harry),(236,Harry)),((236,Harry),(236,Harry)),((237,James),(237,James),((238,Jessica),(238,Jessica))

59 For each tuple (a, b, c) produce the key-value pair (a, b) in map function. Apply the operator θ to the list [b1, b2, …, bn] of B-values associated with key a. The output is the pair (a, x), where x is the result of applying θ to the list in reduce function. This logic applies to which of the following operations?

(a) Intersection
(b) Union
(c) Grouping and Aggregation
(d) Natural Join

60 For each tuple (a, b) of R, produce the key-value pair (b, (R, a)). For each tuple (b, c) of S, produce the key-value pair (b, (S, c)) in map function. Each key value b will be associated with a list of pairs that are either of the form (R, a) or (S, c). Construct all pairs consisting of one with first component R and the other with first component S, say (R, a) and (S, c). The output from this key and value list is a sequence of key-value pairs. The key is irrelevant. Each value is one of the triples (a, b, c) such that (R, a) and (S, c) are on the input list of values for key b in reduce function. This logic applies to which of the following operations?

(a) Natural Join
(b) Intersection
(c) Union
(d) Set Difference

61 A set consist of some elements say 8,10, 12,14…….and so on. Check whether 7 lie in the set or not. Set the array size as 10. Hash functions are: i) 3x+3 mod 6 ii) 3x+7 mod 8

(a) Definitely present
(b) May be present
(c) Surely not present
(d) Cannot say

62 Suppose the pass consist of 1, 2, 3, 1, 2, 3, 4, 1, 2, 4. Hash function is 6X+1 mod 5. How many numbers of distinct elements presents in the pass?

(a) 1
(b) 2
(c) 3
(d) 4

63 Suppose the pass consist of 2,1,6,1,5,2,3,5. Hash function is 2X+3 mod 16. How many numbers of distinct elements presents in the pass?

(a) 1
(b) 2
(c) 3
(d) 4

64 A set consist of some elements say 8,10,…….and so on. Check whether 12 lie in the set or not. Set the array size as 10. Hash functions are:i) 3x+3 mod 6 ii) 3x+7 mod 8 iii) 2x+9 mod 2

(a) May not be present
(b) May be present
(c) Surely not present
(d) Cannot say

65 Suppose the data stream consist of 5,3,9,2. Hash function is 3X+1 mod 32. Estimate the number of distinct elements presents in the stream?

(a) 2
(b) 8
(c) 16
(d) 32

66 Suppose the data stream consist of 2,3,9,5,7,11. Hash function is 3X+1 mod 32. What is the maximum count of zeros(tail length) for the given stream?

(a) 1
(b) 2
(c) 3
(d) 4

67 Data stream model for processing can be based on (i) windows, (ii) relation-oriented tuples, (iii) correlation, (iv) graph
(a) i, ii and iv
(b) i, ii, iii, iv
(c) i, ii, iii
(d) i, ii
68 Suppose the data stream consist of 2,1,6,1,5,9,2,3,5. Hash function is 5X mod 16. Estimate the number of distinct elements presents in the stream?
(a) 1
(b) 2
(c) 3
(d) 4
69 Suppose the data stream consist of 6,1,2,1,5,9,2,3,5. Hash function is 4X+1 mod 16. Estimate the number of distinct elements presents in the stream?
(a) 1
(b) 2
(c) 3
(d) 4
70 What is the main difference between standard reservoir sampling and min-wise sampling?
(a) Reservoir sampling makes use of randomly generated numbers whereas min-wise sampling does not.
(b) Min-wise sampling makes use of randomly generated numbers whereas reservoir sampling does not.
(c) Reservoir sampling requires a stream to be processed sequentially, whereas min-wise does not.
(d) For larger streams, reservoir sampling creates more accurate samples than min-wise sampling.
71 Which of the following statements about standard Bloom filters is correct?
(a) It is possible to delete an element from a Bloom filter.
(b) A Bloom filter always returns the correct result.
(c) It is possible to alter the hash functions of a full Bloom filter to create more space.
(d) A Bloom filter always returns TRUE when testing for a previously added element.
72 Which of the following option is not true about the constraints that must be satisfied for representing a stream by buckets using the DGIM algorithm.
(a) The right end of a bucket always starts with a position with a 1.
(b) Number of 1s must be a power of 2.
(c) Either three or four buckets with the same power-of-2 number of 1s exists.
(d) Buckets do not overlap in timestamps.
73 Suppose the data stream consist of 4,1,3,1,5,9,2,6,5. Hash function is 3X+1 mod 5. Estimate the number of distinct elements presents in the stream?
(a) 1
(b) 2
(c) 3
(d) 4
74 In which type of streaming multimedia file is delivered to the client, but not shared?
(a) Real time streaming
(b) Progressive streaming
(c) Encoding
(d) Compression
75 In teardown state of real time streaming protocol
(a) server resources the client
(b) server delivers the stream to client
(c) server suspends delivery of stream
(d) server breaks down the connection
76 If Multimedia is "Combination of media" then Hypermedia is
(a) Separate Media
(b) Linked Media
(c) Separate Concepts
(d) Linked Concepts
77 _____Audio/Video refers to broadcasting of radio and Tv programs through internet
(a) Streaing live
(b) Interactive
(c) Static
(d) Streaming stored
78 What are closed itemsets?
(a) An itemset for which at least one proper super-itemset has same support
(b) A frequent itemset that is both closed and its support is greater than or equal to minsup.
(c) An itemset for which at least super-itemset has same confidence
(d) An itemset whose no proper super-itemset has same confidence
79 What is association rule mining?
(a) Same as frequent itemset mining
(b) Finding of strong association rules using frequent itemsets
(c) Using association to analyse correlation rules
(d) Finding of minimum association rules using frequent itemsets
80 What is frequent pattern growth?
(a) Same as frequent itemset mining
(b) Use of hashing to make discovery of frequent itemsets more efficient
(c) Mining of frequent itemsets without candidate generation
(d) Finding of minimum association rules using frequent itemsets

81 When is sub-itemset pruning done?
(a) Condition 1: A frequent itemset 'P' is a proper subset of another frequent itemset 'Q'
(b) Condition 2: Support (P) = Support(Q)
(c) When both conditions are true
(d) When condition 1 is true and condition 2 is not true
82 What are Max_confidence, Cosine similarity, All_confidence?
(a) Frequent pattern mining algorithms
(b) Measures to improve efficiency of apriori
(c) Pattern evaluation measure
(d) Same as frequent itemset mining
83 Which of these is not a frequent pattern mining algorithm?
(a) Apriori
(b) FP growth
(c) Decision trees
(d) Eclat
84 Which algorithm requires fewer scans of data?
(a) Apriori
(b) FP growth
(c) Apriori and FP growth
(d) SON and PCY
85 What are maximal frequent itemsets?
(a) a frequent itemset for which none of its immediate supersets are frequent
(b) A frequent itemset whose super-itemset is also frequent
(c) A non-frequent itemset whose super-itemset is frequent
(d) A frequent itemset whose super-itemset is not frequent
86 Why is correlation analysis important?
(a) To make apriori memory efficient
(b) To weed out uninteresting frequent itemsets
(c) To find large number of interesting itemsets
(d) To restrict the number of database iterations
87 Which of the following is finally produced by Hierarchical Clustering?
(a) final estimate of cluster centroids
(b) tree showing how close things are to each other
(c) assignment of each point to clusters
(d) Final estimate ofcluster centroids and assignment of each point
88 What is the minimum no. of variables/ features required to perform clustering?
(a) 0
(b) 1
(c) 2
(d) 3
89 Which of the following can act as possible termination conditions in K-Means?
1. For a fixed number of iterations.
2. Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
3. Centroids do not change between successive iterations.
4. Terminate when RSS falls below a threshold.
(a) 1,2,3
(b) 1,3,4
(c) 2,3,4
(d) 1,2,3,4
90 Which of the following clustering algorithms suffers from the problem of convergence at local optima?
1. K- Means clustering algorithm
2. Agglomerative clustering algorithm
3. Expectation-Maximization clustering algorithm
4. Diverse clustering algorithm
(a) 1,2,3
(b) 1,3
(c) 1,2
(d) 3,4
91 Which of the following clustering algorithm follows a top to bottom approach?
(a) Kmeans
(b) Agglomerative
(c) Divisible
(d) KNN
92 For clustering, we do not require
(a) Labeled data
(b) Unlabeled data
(c) Numerical data
(d) Categorial data
93 Which of the following is not an application of clustering?
(a) Biological Network Analysis
(b) Market Trend Analysis
(c) Topic Modeling
(d) Hash Functioning
94 Bluetooth is an example of
(a) Personal Area network
(b) Virtual Private network
(c) Storage Area Network
(d) Local Area Network
95 In Link Analysis we usually deal with _____ law distributed graphs.
(a) Power
(b) Web
(c) Data
(d) Graph

96 A _____ is a device that forwards packets between networks by processing the routing information included in the packet
(a) Router
(b) Firewall
(c) Bridge
(d) Adapter
97 Which address identifies a process on a host?
(a) Port address
(b) Physical address
(c) Web address
(d) Logical address
98 _____the pioneer in this field with the use of a PageRank measure for ranking Web pages wrt a user query
(a) Google
(b) Facebook
(c) Orkut
(d) LinkedIn
99 Two popular spam indexing techniques includes
(a) Cloaking, use of "Doorway" pages#
(b) Page ranking, Link analysis
(c) Link analysis, Graph analysis
(d) Cloaking and link analysis
100 The techniques used by spammers to fool search engines into ranking useless pages higher are called as
(a) Term Spam
(b) Page rank
(c) Link Analysis
(d) Page Analysis

101 _____is a data analysis technique used in network theory that is used to evaluate the relationships or connections between network nodes. These relationships can be between various types of objects (nodes), including people, organizations and even transactions.
(a) Link analysis
(b) Page analysis
(c) Hadoop analysis
(d) Nosql Analysis
102 Page Rank is a function that defines which one of the following
(a) PageRank is a function that assigns a real number to each page in the Web (or at least to that portion of the Web that has been crawled and its links discovered)
(b) PageRank is a function that assigns a integer number to each page in the Web (or at least to that portion of the Web that has been crawled and its links discovered)
(c) PageRank is a function that assigns a integer number to all pages in the Web (or at least to that portion of the Web that has been crawled and its links discovered)
(d) PageRank is a function that assigns a integer number to each page in the Web (or at least to that portion of the Web that has been crawled)
103 The three primary purposes of Link Analysis are-
(a) Find matches for unknown patterns, find anomalies and find new patterns of interest
(b) Find matches for known patterns, find anomalies and find new patterns of interest
(c) Find matches for known patterns, find regularity and find old patterns of interest
(d) Find matches for unknown patterns, find regularity and find new patterns of interest
104 The structure of the web consist of 3 regions are:
(a) Termination, start, mid
(b) Core, Origination, termination
(c) Start, Core, End
(d) Core, Mid, Termination
105 The purpose of using Search Engine Optimization
(a) An industry that attempts to make a Website attractive to the major search engines and thus increasing rank
(b) An industry that does not attempts to make a Website attractive to the major search engines and thus increase their ranking.
(c) An industry that attempts to make a Website attractive to the major search engines and decreases rank
(d) An industry that does not attempts to make a Website attractive to the major search engines and thus decrease their ranking.
106 PageRank was named after _____, one of the founders of Google
(a) Larry Page
(b) Flajolet-Martin
(c) Bill Gates
(d) Doug Cutting
107 HITS is developed by
(a) Jon Kleinberg
(b) Flajolet-Martin
(c) Bill Gates
(d) Doug Cutting
108 Hub and Authorities is called as
(a) Hyperlink Induced Topic Search
(b) FM algorithm
(c) DGIM Algorithm
(d) Bloom Filter
109 _____is a recommendation system
(a) Surprise
(b) Hadoop
(c) Map Reduce
(d) Nosql Analysis