1. Identify all the possible ways to integrate R into Hadoop Environment

R + Streaming

Rhipe & Rhadoop

library(rhipe)

Your Answer

All of the above

1. When a hash function h apply to a stream element a, the bit string h(a) will end in some number of 0' s, possibly none is called as \_\_\_\_\_\_\_\_\_\_\_\_

**Correct**

**1/1**

Your Answer

Tail Length

Estimate

Blume Filter

None of the Above

1. Consider the following statements and say True or False: Volatility refers to the data velocity relative to timescale of event being studied

**Correct**

**1/1**

True

Your Answer

False

N/A

N/A

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_refers to the connectedness of big data

**Correct**

**1/1**

Value

Velocity

Your Answer

Valence

Veracity

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ensures that key milestones and objectives are met on time and at the expected quality.

**Correct**

**1/1**

Data Engineer

Project Sponsor

Your Answer

Project Manager

None of the Above

1. The characteristics of HDFS includes

**Correct**

**1/1**

Fault tolerant

Write once file access only

Runs with commodity hardware

Your Answer

All the Above

1. Building a successful BDM Architecture includes stages like

**Correct**

**1/1**

Organize

Capture

Analyze

Your Answer

All of the Above

1. An approach to summarizing a stream by look at only the last n elements for some (typically large) n is called as

**Correct**

**1/1**

Your Answer

Fixed-length " window "

Variable Length " window "

Sliding " window "

None of the Above

1. Analytics Sandbox provides

**Correct**

**1/1**

Enables Flexible

High Performance

Reduced Cost

Your Answer

All of the Above

1. Hadoop is used with different modes which include

**X Incorrect**

**0/1**

Cluster Mode, Stand Alone

Stand Alone, Distributed

Your Answer

Stand Alone, Psudo, and Distribted

None of the Above

1. Viscosity refers to the rate of data loss and stable lifetime of data

**Correct**

**1/1**

Your Answer

False

True

N/A

N/A

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is the master node of the MapReduce system, which manages the jobs and resources in the cluster

**Correct**

**1/1**

Your Answer

Job Tracker

Mapper

Reducter

Task Tracker

1. Data Analytics order pf phases is

**Correct**

**1/1**

Model Planning, Discover, Data Preparation, Model Building ..

Data Preparation, Discover, Model Planning, Model Building ..

Your Answer

Discover, Data Preparation, Model Planning, Model Building ..

Discover, Data Preparation, Model Building, Model Planning, ..

1. Batch Processing provides ability to process and analyze data at-rest (stored data)

**Correct**

**1/1**

False

Your Answer

True

N/A

N/A

1. What are the parameters defined to specify window operation ?

**X Incorrect**

**0/1**

Your Answer

Window length, sliding interval

State size, sliding interval

State size, window length

None of the mentioned

1. SPSS Modeler is a commercial tool developed by \_\_\_\_\_\_\_\_\_\_\_ offers methods to build models, explore and analyze data through a GUI.

**Correct**

**1/1**

SAS

Your Answer

IBM

Oracle

Cognus

1. Map-Reduce environment takes care of

**Correct**

**1/1**

Partitioning & Scheduling

Scheduling & Communication

Your Answer

Partitioning, Scheduling & Communication

None of the Above

1. For which of the following operations is NO communication with the NameNode required?

**Correct**

**1/1**

A client requesting the filename of a given block of data.

A client reading a file from the cluster

A client writing a file to HDFS.

Your Answer

A client reading a block of data from the cluster.

1. \_\_\_\_\_\_\_\_\_\_\_\_is the slave/worker node and holds the user data in the form of Data Blocks.

**Correct**

**1/1**

Your Answer

DataNode

NameNode

Data block Replication

Replication

1. Which of the following is a characteristic of Big Data?

**Correct**

**1/1**

Huge volume of data

Speed of data creation and growth

Complexity of data types and structures

Your Answer

All of the mentioned

1. Whenever a new user / new added, then \_\_\_\_\_\_\_\_\_\_\_\_\_ issue will arises

**Correct**

**1/1**

Data Sparsity

Your Answer

Cold Start

Overload

Override

1. Recommender System is a kind of application belongs to

**X Incorrect**

**0/1**

Graph Data

High Dimensional Data

Infinite Data

Your Answer

Machine Learning

1. Importance of web pages can be measured with the help of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ algorithm

**Correct**

**1/1**

Agglomerative Clustering

Association - Rule based

Your Answer

Page Rank

SVM

1. Agglomerative Clustering is a kind of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Correct**

**1/1**

Partitioned based Clustering

Your Answer

Hierarchical based Clustering

Grid Based Clustering

Density based Clustering

1. Decision Tree is a Tree - based algorithm, which is a category of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Correct**

**1/1**

Clustering based algorithm

Association - Rule based algorithm

Prediction - based algorithm

Your Answer

Classification - based algorithm

1. Teleport in the name of " teleport Set " can be used as a topic - specific set of " relevant " pages (teleport set) in

**Correct**

**1/1**

Your Answer

Topic Specific PageRank

Link Analysis

Page Rank

Topic sensitive PageRank

1. The \_\_\_\_\_\_\_\_\_\_\_\_ of an edge (a, b) to be the number of pairs of nodes x and y such that the edge (a, b) lies on the shortest path between x and y.

**Correct**

**1/1**

Your Answer

Betweenness

Volume

Degree

Density

1. Web advertising technique applied in\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**X Incorrect**

**0/1**

High Dimensional Data

Infinite Data

Graph based Data

Your Answer

Web - based Data

1. Importance of web pages can be measured with the help of Page Rank Algorithm

**Correct**

**1/1**

Your Answer

TRUE

FALSE

N/A

N/A

1. Hybrid Filtering based Recommender System is the combination of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Recommender System

**Correct**

**1/1**

Item Based and User based

Your Answer

Content based and Collaborative Filtering based

Personalized and Non - personalized

None of the above

1. If the the sum of the entries in each column is 1 in a matrix, it shows the uniform stationary distribution for stochastic matrix

**X Incorrect**

**0/1**

True

Your Answer

False

N/A

N/A

1. Utility function is defined as u: C × S - > R. In this formula c stands for a set of Customers

**X Incorrect**

**0/1**

True

Your Answer

False

N/A

N/A

1. Content based RS can handle new users easily

**Correct**

**1/1**

TRUE

Your Answer

FALSE

N/A

N/A

1. What is the number of communities we get after removing the edge(s) with highest betweeness score value 4 and above from the figure

**X Incorrect**

**0/1**

3

2

1

Your Answer

None of the Above

1. The Betweenness is computed until no edges are left and edges with highest betweenness is removed

**Correct**

**1/1**

Your Answer

True

False

N/A

N/A

1. Point out the wrong statement?

**X Incorrect**

**0/1**

Your Answer

Setting up a workstation to take full advantage of the customizable features of R is a straightforward thing

q() is used to quit the R program

R has an inbuilt help facility similar to the man facility of UNIX

Windows versions of R have other optional help systems also

1. SPSS Modeler is a commercial tool developed by \_\_\_\_\_\_\_\_\_\_\_ offers methods to build models, explore and analyze data through a GUI.

**Correct**

**1/1**

SAS

Your Answer

IBM

Oracle

Cognus

1. Dimensionality Detection method applied in \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Correct**

**1/1**

Your Answer

High Dimensional Data

Graph Data

Infinite Data

Web - based Data

1. Clustering Algorithm is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of machine learning algorithm

**Correct**

**1/1**

Supervised Model

Your Answer

Unsupervised Model

Association - based Model

Prediction - based Model

1. Training and Testing data supplied in \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Correct**

**1/1**

Clustering based algorithm

Association - Rule based algorithm

Your Answer

Classification - based algorithm

Prediction - based algorithm

1. Express partitioning objectives as a function of the " edge cut " of the partition. What is the value of Cut(A, B) for the given figure?

**Correct**

**1/1**

Your Answer

2

1

3

None of the Above

1. Discovering communities is done by directly by looking for subsets of the nodes that have a relatively large number of edges among them

**Correct**

**1/1**

Your Answer

True

False

N/A

N/A

1. What will be the output of the following R code? y < - c(TRUE, 2)

**Correct**

**1/1**

[1] " TRUE " " 2 "

1] " TRUE " 2

1] " 0 " " 2 "

Your Answer

[1] 1 2

1. From the given figure if we choose S = {H} and T = {A, B, C, D, E, F, G}, then Cut(S, T ),Vol(S) .

**Correct**

**1/1**

Your Answer

1, 1

1, 7

0, 7

None of the Above

1. Seq is a sequence of numbers or character strings

**Correct**

**1/1**

Your Answer

True

False

N/A

N/A

1. The following function can be used to construct matrices by row-binding.

**Correct**

**1/1**

rjoin()

Your Answer

rbind()

rowbind()

rbinding()

1. Data structures in R include

**Correct**

**1/1**

Your Answer

Data frame

Structure

Dictionary

None of the Above

1. The total weight of the edges with at least one endpoint in ?A is called as

**Correct**

**1/1**

Your Answer

Vol (A)

Cut(A)

Density (A)

None of the Above

1. From the given figure , what is the normalized cut for this partition if Cut(S, T ) = 2, Vol(S) = 6, and Vol(T ) = 7?

**Correct**

**1/1**

0.52.

0.23

Your Answer

0.62.

None of the Above

1. From the given figure if we choosebest cut, then what is the calue of Cut(S, T ),Vol(S) and Vol(T) .

**Correct**

**1/1**

Your Answer

2,6,7

3,7,6

2,4,4

1,4,4

1. Which of the following functions in R cannot be used to import a CSV file?

**Correct**

**1/1**

read.table()

read.csv()

Your Answer

read\_excel()

All of the above

1. Collaborative Filtering Systems need data on other users

**Correct**

**1/1**

Your Answer

True

FALSE

N/A

N/A

1. For the given figure what is the volume of A

**Correct**

**1/1**

Your Answer

2

3

1

None of the Above

1. From the given fighure what is the count of positive contributions for A, B, C.

**X Incorrect**

**0/1**

Your Answer

(1, 0, 0)

(0, 0, 1)

(1, 1, 0)

(1, 1, 1)

1. \_\_\_\_\_\_ is a statement that evaluates to a number or character string

**Correct**

**1/1**

Your Answer

Expr

Cond

Seq

None of the above

1. Point out the wrong statement?

**X Incorrect**

**0/1**

Your Answer

Setting up a workstation to take full advantage of the customizable features of R is a straightforward thing

q() is used to quit the R program

R has an inbuilt help facility similar to the man facility of UNIX

Windows versions of R have other optional help systems also

1. Common data mining technique include

**Correct**

**1/1**

Your Answer

Regression

Data Creation

None of the Above

Business Intelligent

1. Identify all the possible ways to integrate R into Hadoop Environment

**Correct**

**1/1**

R + Streaming

Rhipe & Rhadoop

library(rhipe)

Your Answer

All of the above

1. R is mostly used in \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Correct**

**1/1**

Problem solving

Statistics

Probability

Your Answer

All of the mentioned

1. From the given fighure what is the count of negative contributions for A, B, C.

**X Incorrect**

**0/1**

Your Answer

(1, 0, 0)

(0, 0, 1)

(0, 0, 0)

None of the Above