

X

NPTEL

reviewer4@nptel.iitm.ac.in ▼

[Courses](#) » [Big Data Computing](#)[Announcements](#)[Course](#)[Ask a Question](#)[Progress](#)[FAQ](#)

Unit 1 - How to access the portal

[Register for
Certification exam](#)

Course outline

How to access the portal

- ☐ How to access the home page?
- ☐ How to access the course page?
- ☐ How to access the MCQ, MSQ and Programming assignments?
- ☐ How to access the subjective assignments?
- ☐ Quiz : Assignment-0

[Week-1](#)[Week-2](#)[Week-3](#)[Week-4](#)[Week-5](#)[Week-6](#)[Week-7](#)

Assignment-0

The due date for submitting this assignment has passed.

As per our records you have not submitted this assignment.

Due on 2019-02-25, 23:59 IST.

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

1 point

- ☐ Huge volume of data
- ☐ Complexity of data types and structures
- ☐ Speed of data creation and growth
- ☐ All of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

All of the mentioned

2) Concurrent access to shared data may result in _____

1 point

- ☐ Data consistency
- ☐ Data insecurity
- ☐ Data inconsistency
- ☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Data inconsistency

3) The following three conditions must be satisfied to solve the critical section problem

1 point

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



In association with



Funded by

Score: 0**Accepted Answers:***All of the mentioned*

4) Mutual exclusion implies that :

1 point

- ☐ If a process is executing in its critical section, then no other process must be executing in their critical sections
- ☐ If a process is executing in its critical section, then other processes must be executing in their critical sections
- ☐ If a process is executing in its critical section, then all the resources of the system must be blocked until it finishes execution
- ☐ None of the mentioned

No, the answer is incorrect.**Score: 0****Accepted Answers:***If a process is executing in its critical section, then no other process must be executing in their critical sections*

5) In the memory hierarchy, as the speed of operation increases the memory size also increases.

1 point

- ☐ True
- ☐ False

No, the answer is incorrect.**Score: 0****Accepted Answers:***False*

6) To use a _____ network service, the service user first establishes a connection, uses the connection, and terminates the connection.

1 point

- ☐ Connection-oriented
- ☐ Connection-less
- ☐ Service-oriented
- ☐ Service-less

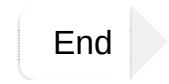
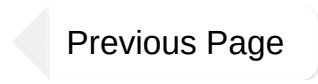
No, the answer is incorrect.**Score: 0****Accepted Answers:***Connection-oriented*

7) Which layer is responsible for the process-to-process delivery ?

1 point

- ☐ Network
- ☐ Transport
- ☐ Application
- ☐ Physical

No, the answer is incorrect.**Score: 0****Accepted Answers:***Transport*



X

NPTEL

reviewer4@nptel.iitm.ac.in ▼

[Courses](#) » [Big Data Computing](#)[Announcements](#)[Course](#)[Ask a Question](#)[Progress](#)[FAQ](#)

Unit 2 - Week-1

[Register for
Certification exam](#)

Course outline

How to access the portal

Week-1

- ☐ Introduction to Big Data
- ☐ Big Data Enabling Technologies
- ☐ Hadoop Stack for Big Data
- ☐ Quiz : Assignment-1
- ☐ Week 1: Lecture material
- ☐ Big Data Computing: Feedback for Week-01
- ☐ Assignment-1 Solution

Week-2

Week-3

Week-4

Week-5

Assignment-1

The due date for submitting this assignment has passed.

As per our records you have not submitted this assignment.

Due on 2019-03-13, 23:59 IST.

1) _____ refers to the biases, noise and abnormality in data, trustworthiness of data. **1 point**

- ☐ Value
- ☐ Veracity
- ☐ Velocity
- ☐ Volume

No, the answer is incorrect.

Score: 0

Accepted Answers:

Veracity

2) _____ refers to the connectedness of big data. **1 point**

- ☐ Value
- ☐ Veracity
- ☐ Velocity
- ☐ Valence

No, the answer is incorrect.

Score: 0

Accepted Answers:

Valence

3) Consider the following statements: **1 point**

Statement 1: Volatility refers to the data velocity relative to timescale of event being studied

Statement 2: Viscosity refers to the rate of data loss and stable lifetime of data

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Funded by

Score: 0**Accepted Answers:***Both statements are false*

4) What are the main components of Hadoop Ecosystem?

1 point

- ☐ MapReduce, HDFS, YARN
- ☐ MLlib, GraphX
- ☐ Gelly, Table, CEP
- ☐ None of the mentioned

**No, the answer is incorrect.****Score: 0****Accepted Answers:***MapReduce, HDFS, YARN*

5) True or False ?

1 point

NoSQL databases store unstructured data with no particular schema.

- ☐ True
- ☐ False

No, the answer is incorrect.**Score: 0****Accepted Answers:***True*

6) Which of the following is guaranteed by Zookeeper?

1 point

- ☐ Interactivity
- ☐ Flexibility
- ☐ Scalability
- ☐ Reliability

No, the answer is incorrect.**Score: 0****Accepted Answers:***Reliability*

7) Which of the following is not a NoSQL database?

1 point

- ☐ HBase
- ☐ SQL Server
- ☐ Cassandra
- ☐ None of the mentioned

No, the answer is incorrect.**Score: 0****Accepted Answers:***SQL Server*

8) _____ is a distributed machine learning framework on top of Spark.

1 point

- ☐ MLlib
- ☐ Spark Streaming
- ☐ GraphX

☐ RDDs

No, the answer is incorrect.

Score: 0

Accepted Answers:

MLlib

9) _____ is a resource management platform responsible for managing compute **1 point** resources in the cluster and using them in order to schedule users and applications.

- ☐ Hadoop Common
- ☐ Hadoop Distributed File System (HDFS)
- ☐ Hadoop YARN
- ☐ Hadoop MapReduce

No, the answer is incorrect.

Score: 0

Accepted Answers:

Hadoop YARN

10) Which of the following tool is designed for efficiently transferring bulk data between Apache **1 point** Hadoop and structured datastores such as relational databases.

- ☐ Apache Sqoop
- ☐ Pig
- ☐ Mahout
- ☐ Flume

No, the answer is incorrect.

Score: 0

Accepted Answers:

Apache Sqoop

11) _____ is a distributed, reliable, and available service for efficiently collecting, **1 point** aggregating, and moving large amounts of log data.

- ☐ Apache Sqoop
- ☐ Pig
- ☐ Mahout
- ☐ Flume

No, the answer is incorrect.

Score: 0

Accepted Answers:

Flume

12) _____ brings scalable parallel database technology to Hadoop and allows users **1 point** to submit low latencies queries to the data that's stored within the HDFS or the Hbase without acquiring a ton of data movement and manipulation.

- ☐ Apache Sqoop
- ☐ Impala
- ☐ Mahout
- ☐ Flume

No, the answer is incorrect.

Score: 0

Accepted Answers:

Impala

Previous Page

End



X

NPTEL

reviewer4@nptel.iit.ac.in ▼

[Courses](#) » [Big Data Computing](#)[Announcements](#)[Course](#)[Ask a Question](#)[Progress](#)[FAQ](#)

Unit 3 - Week-2

[Register for Certification exam](#)

Course outline

[How to access the portal](#)[Week-1](#)[Week-2](#)

- ☐ Hadoop Distributed File System (HDFS)
- ☐ Hadoop MapReduce 1.0
- ☐ Hadoop MapReduce 2.0 (Part-I)
- ☐ Hadoop MapReduce 2.0 (Part-II)
- ☐ MapReduce Examples
- ☐ Quiz : Assignment-2
- ☐ Week-2: Lecture material
- ☐ Big Data Computing: Feedback for Week-02
- ☐ Assignment-2 Solution

Assignment-2

The due date for submitting this assignment has passed.

As per our records you have not submitted this assignment.

Due on 2019-03-13, 23:59 IST.

1) Consider the following statements:

1 point

Statement 1: The Job Tracker is hosted inside the master and it receives the job execution request from the client.

Statement 2: Task tracker is the MapReduce component on the slave machine as there are multiple slave machines.

- ☐ Only statement 1 is true
- ☐ Only statement 2 is true
- ☐ Both statements are true
- ☐ Both statements are false

No, the answer is incorrect.

Score: 0

Accepted Answers:

Both statements are true

2) _____ is the slave/worker node and holds the user data in the form of Data Blocks.

1 point

- ☐ NameNode
- ☐ DataNode
- ☐ Data block
- ☐ Replication

No, the answer is incorrect.

Score: 0

Accepted Answers:

DataNode

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Funded by

Week-7

Week-8

ce De

☐ None of the mentioned**No, the answer is incorrect.****Score: 0****Accepted Answers:***Inputs*

4) _____ function processes a key/value pair to generate a set of intermediate key/value pairs. **1 point**

- ☐ Map
- ☐ Reduce
- ☐ Both Map and Reduce
- ☐ None of the mentioned

No, the answer is incorrect.**Score: 0****Accepted Answers:***Map*

5) True or False ? **1 point**

The main duties of task tracker are to break down the receive job that is big computations in small parts, allocate the partial computations that is tasks to the slave nodes monitoring the progress and report of task execution from the slave.

- ☐ True
- ☐ False

No, the answer is incorrect.**Score: 0****Accepted Answers:***False*

6) Point out the correct statement in context of YARN: **1 point**

- ☐ YARN extends the power of Hadoop to incumbent and new technologies found within the data center
- ☐ YARN is highly scalable
- ☐ YARN enhances a Hadoop compute cluster in many ways
- ☐ All of the mentioned

No, the answer is incorrect.**Score: 0****Accepted Answers:***All of the mentioned*

7) Apache Hadoop YARN stands for: **1 point**

- ☐ Yet Another Reserve Negotiator
- ☐ Yet Another Resource Network
- ☐ Yet Another Resource Negotiator
- ☐ Yet Another Resource Manager

No, the answer is incorrect.**Score: 0****Accepted Answers:***Yet Another Resource Negotiator*

8) Consider the pseudo-code for MapReduce's WordCount example (not shown here). Let's **1 point** now assume that you want to determine the frequency of phrases consisting of 3 words each instead of determining the frequency of single words. Which part of the pseudo-code do you need to adapt?

- ☐ Only map()
- ☐ Only reduce()
- ☐ Both map() and reduce()
- ☐ The code does not have to be changed

No, the answer is incorrect.

Score: 0

Accepted Answers:

Only map()

9) For which of the following operations is NO communication with the NameNode required? **1 point**

- ☐ A client writing a file to HDFS.
- ☐ A client requesting the filename of a given block of data.
- ☐ A client reading a block of data from the cluster.
- ☐ A client reading a file from the cluster.

No, the answer is incorrect.

Score: 0

Accepted Answers:

A client reading a block of data from the cluster.

10) Which of the following components reside on a NameNode?

1 point

- ☐ Filenames, blocks and checksums
- ☐ Blocks and heartbeat messages
- ☐ Filenames, block locations
- ☐ Blocks and block locations

No, the answer is incorrect.

Score: 0

Accepted Answers:

Filenames, block locations

[Previous Page](#)

[End](#)



X

NPTEL

reviewer4@nptel.iitm.ac.in ▼

Courses » Big Data Computing

Announcements

Course

Ask a Question

Progress

FAQ

Unit 6 - Week-5

Register for
Certification exam

Course outline

How to access the portal

Week-1

Week-2

Week-3

Week-4

Week-5

- ☐ Design of HBase
- ☐ Spark Streaming and Sliding Window Analytics (Part-I)
- ☐ Spark Streaming and Sliding Window Analytics (Part-II)
- ☐ Sliding Window Analytics
- ☐ Introduction to Kafka
- ☐ Quiz : Assignment-5
- ☐ Week-5: Lecture material
- ☐ Big Data Computing: Feedback for Week-05
- ☐ Assignment-5 Solution

Week-6

Week-7

Week-8

Assignment-5

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2019-04-03, 23:59 IST

1) Columns in HBase are organized to _____ **1 point**

- ☐ Column group
- ☐ Column families
- ☐ Column list
- ☐ Column base

No, the answer is incorrect.

Score: 0

Accepted Answers:

Column families

2) HBase is a distributed _____ database built on top of the Hadoop file system. **1 point**

- ☐ Row-oriented
- ☐ Column-oriented
- ☐ Tuple-oriented
- ☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Column-oriented

3) A small chunk of data residing in one machine which is part of a cluster of machines holding one Hbase table is known as _____ **1 point**

- ☐ Rowarea
- ☐ Tablearea
- ☐ Region
- ☐ Split

No, the answer is incorrect.

Score: 0

Accepted Answers:

Region

4) Hbase stores data in _____ **1 point**

- ☐ A single filesystem available to all region servers

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM

Funded by

Government of India
Ministry of Human Resource Development

Powered by



Accepted Answers:

A single filesystem available to all region servers

5) Kafka is run as a cluster comprised of one or more servers each of which is called _____ **1 point**

- ☐ cTakes
- ☐ Chunks
- ☐ Broker
- ☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Broker

6) Kafka maintains feeds of messages in categories called _____ **1 point**

- ☐ Chunks
- ☐ Domains
- ☐ Messages
- ☐ Topics

No, the answer is incorrect.

Score: 0

Accepted Answers:

Topics

7) True or False ? **1 point**

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

Statement 2: Stream Processing provides ability to ingest, process and analyze data in-motion in real or near-real-time.

- ☐ Only statement 1 is true
- ☐ Only statement 2 is true
- ☐ Both statements are true
- ☐ Both statements are false

No, the answer is incorrect.

Score: 0

Accepted Answers:

Both statements are true

8) Dstream internally is _____ **1 point**

- ☐ Continuous Stream of DataFrame
- ☐ Continuous Stream of DataSet
- ☐ Continuous Stream of Resilient Distributed Datasets (RDD)
- ☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Continuous Stream of Resilient Distributed Datasets (RDD)

9) What are the parameters defined to specify window operation ? **1 point**

- ☐ Window length, sliding interval
- ☐ State size, window length
- ☐ State size, sliding interval
- ☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Window length, sliding interval

10) Consider the following dataset Customers:

1 point

Name	Date	AmountSpent
Alice	2016-05-01	50
Bob	2016-05-04	29
Bob	2016-05-01	25
Alice	2016-05-04	55
Alice	2016-05-03	45
Bob	2016-05-06	27

Using the Customers table calculate the moving average for following pseudo code using spark streaming fundamentals:

```
val wSpec1 = Window.partitionBy("name").orderBy("date").rowsBetween(-1, 1)
customers.withColumn("movingAvg", avg(customers("amountSpent")).over(wSpec1)).show()
```

Name	Date	AmountSpent	MovingAv
Alice	2016-05-01	50	50
Bob	2016-05-04	29	29
Bob	2016-05-01	25	54
Alice	2016-05-04	55	105
Alice	2016-05-03	45	150
Bob	2016-05-06	27	81

Name	Date	AmountSpent	MovingAv
Alice	2016-05-01	50	47.5
Bob	2016-05-04	29	27
Bob	2016-05-01	25	27
Alice	2016-05-04	55	50
Alice	2016-05-03	45	50
Bob	2016-05-06	27	28

Name	Date	AmountSpent	MovingAv
Alice	2016-05-01	50	Null
Alice	2016-05-03	45	50
Alice	2016-05-04	55	45
Bob	2016-05-01	25	Null
Bob	2016-05-04	29	25
Bob	2016-05-06	27	29

Name	Date	AmountSpent
Alice	2016-05-01	50
Bob	2016-05-04	29
Bob	2016-05-01	25
Alice	2016-05-04	55
Alice	2016-05-03	45
Bob	2016-05-06	27

No, the answer is incorrect.

Score: 0

Accepted Answers:

Name	Date	AmountSpent	MovingAvg
Alice	2016-05-01	50	47.5
Bob	2016-05-04	29	27
Bob	2016-05-01	25	27
Alice	2016-05-04	55	50
Alice	2016-05-03	45	50
Bob	2016-05-06	27	28

Previous Page

End

X

NPTEL

reviewer4@nptel.iitm.ac.in ▼

[Courses » Big Data Computing](#)[Announcements](#)[Course](#)[Ask a Question](#)[Progress](#)[FAQ](#)

Unit 7 - Week-6

[Register for Certification exam](#)

Course outline

[How to access the portal](#)[Week-1](#)[Week-2](#)[Week-3](#)[Week-4](#)[Week-5](#)[Week-6](#)

- ☐ Big Data Machine Learning (Part-I)
- ☐ Big Data Machine Learning (Part-II)
- ☐ Machine Learning Algorithm K-means using Map Reduce for Big Data Analytics
- ☐ Parallel K-means using Map Reduce on Big Data

Assignment-6

The due date for submitting this assignment has passed.

As per our records you have not submitted this assignment. **Due on 2019-04-10, 23:59 IST.**

1) Identify the correct statement in context of Regressive model of Machine Learning. **1 point**

- ☐ Regressive model predicts a numeric value instead of category
- ☐ Regressive model organizes similar item in your dataset into groups
- ☐ Regressive model comes up with a set of rules to capture associations between items or events
- ☐ None of the Mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Regressive model predicts a numeric value instead of category

2) Which of the following is required by K-means clustering ? **1 point**

- ☐ Defined distance metric
- ☐ Number of clusters
- ☐ Initial guess as to cluster centroids
- ☐ All of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

All of the mentioned

3) True or False ? **1 point**

If your model has very low training error but high generalization error, then it is overfitting.

- ☐ True
- ☐ False

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Funded by

Big Data Computing:
Feedback for
Week-06

Assignment-6
Solution

Week-7

Week-8

ce De

You trained a model on training dataset and get the below confusion matrix on validation dataset.

	n=165	
	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Based on the above confusion matrix, choose which option(s) below will give you correct predictions?

1. Accuracy is ~0.91
2. Misclassification rate is ~ 0.91
3. False positive rate is ~0.95
4. True positive rate is ~0.95

- ☐ 1 and 3
☐ 2 and 4
☐ 1 and 4
☐ 2 and 3

No, the answer is incorrect.

Score: 0

Accepted Answers:

1 and 4

- 5) Identify the correct method for choosing the value of 'k' in k-means algorithm?

1 point

- ☐ Dimensionality reduction
☐ Elbow method
☐ Data partitioning
☐ Both Dimensionality reduction and Elbow method

No, the answer is incorrect.

Score: 0

Accepted Answers:

Both Dimensionality reduction and Elbow method

- 6) Identify the correct statement(s) in context of machine learning approaches:

1 point

Statement I: In supervised approaches, the target that the model is predicting is unknown or unavailable. This means that you have unlabeled data.

Statement II: In unsupervised approaches the target, which is what the model is predicting, is provided. This is referred to as having labeled data because the target is labeled for every sample that you have in your data set.

- ☐ Only statement I is true
☐ Only statement II is true
☐ Both statements are true
☐ Both statements are false

No, the answer is incorrect.

Score: 0

Accepted Answers:*Both statements are false*

7) Identify the correct statement(s) in context of overfitting in decision trees:

1 point

Statement I: The idea of Pre-pruning is to stop tree induction before a fully grown tree is built, that perfectly fits the training data.

Statement II: The idea of Post-pruning is to grow a tree to its maximum size and then remove the nodes using a top-bottom approach.

- ☐ Only statement I is true
- ☐ Only statement II is true
- ☐ Both statements are true
- ☐ Both statements are false

No, the answer is incorrect.**Score: 0****Accepted Answers:***Only statement I is true*

8) Which of the following options is/are true for K-fold cross-validation?

1 point

1. Increase in K will result in higher time required to cross validate the result.
2. Higher values of K will result in higher confidence on the cross-validation result as compared to lower value of K.
3. If $K=N$, then it is called Leave one out cross validation, where N is the number of observations.

- ☐ 1 and 2
- ☐ 2 and 3
- ☐ 1 and 3
- ☐ 1, 2 and 3

No, the answer is incorrect.**Score: 0****Accepted Answers:***1, 2 and 3*[Previous Page](#)[End](#)



X

NPTEL

reviewer4@nptel.iitm.ac.in ▼

Courses » Big Data Computing

Announcements

Course

Ask a Question

Progress

FAQ

Unit 8 - Week-7

Register for
Certification exam

Course outline

How to access
the portal

Week-1

Week-2

Week-3

Week-4

Week-5

Week-6

Week-7

- ☐ Decision Trees for Big Data Analytics
- ☐ Big Data Predictive Analytics (Part-I)
- ☐ Big Data Predictive Analytics (Part-II)
- ☐ Quiz : Assignment-7
- ☐ Week-7: Lecture material
- ☐ Big Data Computing: Feedback for Week-07

Assignment-7

The due date for submitting this assignment has passed. **Due on 2019-04-17, 23:59 IST**
As per our records you have not submitted this assignment.

1) In which of the following scenario a gain ratio is preferred over Information Gain?

1 point

- ☐ When a categorical variable has very large number of category
- ☐ When a categorical variable has very small number of category
- ☐ Number of categories is the not the reason
- ☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

When a categorical variable has very large number of category

2) Which of the following is/are true about Random Forest and Gradient Boosting ensemble methods?

1 point

1. Both methods can be used for classification task
2. Random Forest is use for classification whereas Gradient Boosting is use for regression task
3. Random Forest is use for regression whereas Gradient Boosting is use for Classification task
4. Both methods can be used for regression task

- ☐ 1 and 2
- ☐ 2 and 3
- ☐ 2 and 4
- ☐ 1 and 4

No, the answer is incorrect.

Score: 0

Accepted Answers:

1 and 4

3) Given an attribute table shown below, which stores the basic information of attribute a, including the row identifier of instance row_id , values of attribute values (a) and class labels of instances c. 1 point

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -







A project of



In association with



Funded by

Attribute Table			
Outlook	Humidity	Wind	Play
Sunny	High	Weak	No
Sunny	High	Strong	No
Overcast	High	Weak	Yes 
Rain	High	Weak	Yes 
Rain	Normal	Weak	Yes 
Rain	Normal	Strong	No 
Overcast	Normal	Strong	Yes 
Sunny	High	Weak	No 
Sunny	Normal	Weak	Yes
Rain	Normal	Weak	Yes
Sunny	Normal	Strong	Yes
Overcast	High	Strong	Yes
Overcast	Normal	Weak	Yes
Rain	High	Strong	No

Which of the following attribute will first provide the pure subset ?

- ☐ Humidity
☐ Wind
☐ Outlook
☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Outlook

4) True or False ?

1 point

Bagging provides an averaging over a set of possible datasets, removing noisy and non-stable parts of models.

- ☐ True
☐ False

No, the answer is incorrect.

Score: 0

Accepted Answers:

True

5) Hundreds of trees can be aggregated to form a Random forest model. Which of the following is **1 point** true about any individual tree in Random Forest?

- Individual tree is built on a subset of the features
- Individual tree is built on all the features
- Individual tree is built on a subset of observations
- Individual tree is built on full set of observations

- ☐ 1 and 3
- ☐ 1 and 4
- ☐ 2 and 3
- ☐ 2 and 4

No, the answer is incorrect.

Score: 0

Accepted Answers:

1 and 3

6) Boosting any algorithm takes into consideration the weak learners.
Which of the following is the main reason behind using weak learners?

1 point

Reason I- To prevent overfitting

Reason II- To prevent underfitting

- ☐ Both the Reasons
- ☐ Reason I
- ☐ Reason II
- ☐ None of the reasons

No, the answer is incorrect.

Score: 0

Accepted Answers:

Reason I

7) Suppose you are using a bagging based algorithm say a Random Forest in model building.
Which of the following can be true?

1 point

1. Number of tree should be as large as possible .
2. You will have interpretability after using Random Forest.

- ☐ Only 1
- ☐ Only 2
- ☐ Both 1 and 2
- ☐ None of these

No, the answer is incorrect.

Score: 0

Accepted Answers:

Only 1

8) To apply bagging to regression trees which of the following is/are true in such case?

1 point

1. We build the N regression with N bootstrap sample
2. We take the average the of N regression tree
3. Each tree has a high variance with low bias

- ☐ 1 and 2
- ☐ 2 and 3
- ☐ 1 and 3
- ☐ 1,2 and 3

No, the answer is incorrect.

Score: 0

Accepted Answers:

1,2 and 3

◀ Previous Page

End ▶



X

NPTEL

reviewer4@nptel.iitm.ac.in ▼

Courses » Big Data Computing

Announcements

Course

Ask a Question

Progress

FAQ

Unit 9 - Week-8

Register for
Certification exam

Course outline

How to access
the portal

Week-1

Week-2

Week-3

Week-4

Week-5

Week-6

Week-7

Week-8

- ☐ Parameter Servers
- ☐ PageRank Algorithm in Big Data
- ☐ Spark GraphX & Graph Analytics (Part-I)
- ☐ Spark GraphX & Graph Analytics (Part-II)
- ☐ Case Study:

Assignment-8

The due date for submitting this assignment has passed.

As per our records you have not submitted this assignment.

Due on 2019-04-24, 23:59 IST.

1) Match the following:

1 point

- | | |
|--------------------------|-----------------------|
| A. Dataflow Systems | i. Vertex Programs |
| B. Graph Systems | ii. Parameter Servers |
| C. Shared Memory Systems | iii. Guinea Pig |

- ☐ A:ii, B: i, C: iii
- ☐ A:iii, B: i, C: ii
- ☐ A:ii, B: iii, C: i
- ☐ A:iii, B: ii, C: i

No, the answer is incorrect.

Score: 0

Accepted Answers:

A:iii, B: i, C: ii

2) Which of the following statement(s) is/are true in context of Parameter Servers.

1 point

- S1: A machine learning framework
 S2: Distributes a model over multiple machines
 S3: It offers two operations: (i) Pull for query parts of the model (ii) Push for update parts of the model.

- ☐ Only S1 is true
- ☐ Only S2 is true
- ☐ Only S3 is true
- ☐ All of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

All of the mentioned

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



In association with

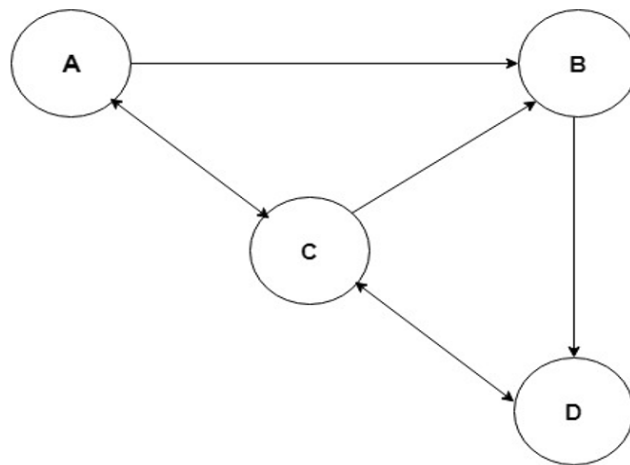


Funded by

material

☐ Big Data Computing: Feedback for Week-08

☐ Assignment-8 Solution



Hint:- The basic PageRank formula is:

$$PR_{t+1}(u) = \sum PR_t(v) / C(v)$$

Where, $PR_{t+1}(u)$: page rank of node u under consideration

$PR_t(v)$: previous page rank of node ' v ' pointing to node ' u '

$C(v)$: outgoing degree of vertex ' v '

- ☐ 1/6
- ☐ 1.5/12
- ☐ 2.5/12
- ☐ 1/3

No, the answer is incorrect.

Score: 0

Accepted Answers:

1/6

4) Which of the following statements are true in the context of Apache Spark GraphX operators ?

1 point

S1: Property operators modify the vertex or edge properties using a user defined map function and produces a new graph.

S2: Structural operators operate on the structure of an input graph and produces a new graph.

S3: Join operators add data to graphs and produces a new graphs.

- ☐ Only S1 is true
- ☐ Only S2 is true
- ☐ Only S3 is true
- ☐ All of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

All of the mentioned

5) True or False ?

1 point

The `outerJoinVertices()` operator joins the input RDD data with vertices and returns a new graph. The vertex properties are obtained by applying the user defined `map()` function to the all vertices, and

includes ones that are not present in the input RDD.

- ☐ True
☐ False

No, the answer is incorrect.

Score: 0

Accepted Answers:

True



6) Which of the following statement(s) is/are true ?

1 point



S1: Apache Spark GraphX provides the following property operators - mapVertices(), mapEdges(), mapTriplets()



S2: The RDDs in Spark, depend on one or more other RDDs. The representation of dependencies in between RDDs is known as the lineage graph. Lineage graph information is used to compute each RDD on demand, so that whenever a part of persistent RDD is lost, the data that is lost can be recovered using the lineage graph information.



- ☐ Only S1 is true
☐ Only S2 is true
☐ Both S1 and S2 are true
☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Both S1 and S2 are true

7) GraphX provides an API for expressing graph computation that can model the _____ **1 point**

- ☐ GaAdt
☐ Spark Core
☐ Pregel
☐ None of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

Pregel

8) Which of the following are provided by spark API for graph parallel computations:

1 point

- i. joinVertices
ii. subgraph
iii. aggregateMessages

- ☐ Only (i)
☐ Only (i) and (ii)
☐ Only (ii) and (iii)
☐ All of the mentioned

No, the answer is incorrect.

Score: 0

Accepted Answers:

All of the mentioned

Previous Page

End

