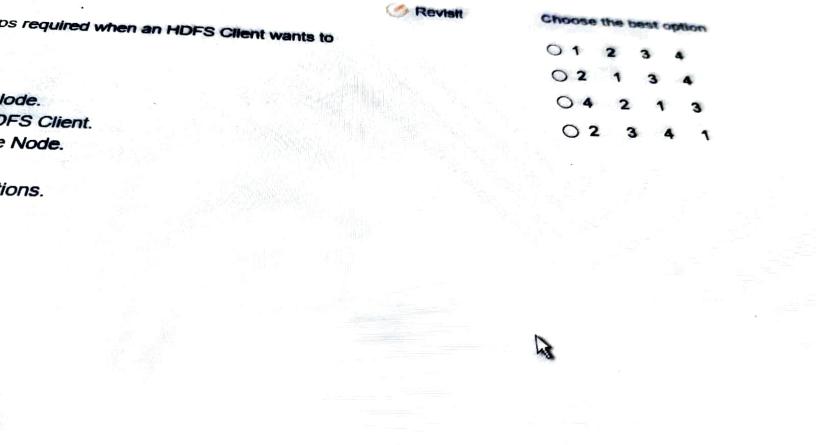
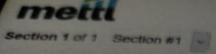
From the below steps, choose right sequence of steps required when an HDFS Client wants to create a File on the Data Node?

- 1) HDFS Client Write the Packet on the Data Node.
- 2) HDFS Client gets the block location from the Name Node.
- 3) Data Node will send acknowledgement back to the HDFS Client.
- 4) HDFS Client will send a create file request to the Name Node.

Choose the most appropriate Sequence from the below options.





8 9 10 11 12 13

Question # 14

C Rev

All MapReduce jobs reading from an HBase table accept their[K1, V1] pair in the form of

Choose the best option

- O rowid:cell value
- O rowkey:scan result
- O column Family:cell value
- O column attribute:scan result

E

When client contact name node to access a file then namenode responds with:

Ӛ

Which of the following is true about HDFS?

Revisit

Choose the best option

O It is a distributed file system.

7 8 9 10 11 12 13 - 242

- It is used for batch processing of data.
- O It is written in Java
- O All of the above

all of the above

Question #1

Which of the following is not a category of NoSQL db?

Choose the best option

Graph

Scientific

Document

PastGressQL

Clear Response

d

True or False: By default streaming context is present in pyspark shell.

Irue

- O Spark
- O Hadoop
- O Kafka
- Kubernetes

posts related to that event. Sports association wanted to know how

Assume that we have a the below Employee data in csv file names "EmployeeData.csv" on HDFS tolder "/SparkDemos"

101,Alex,20000 102,Adam,30000 103, Rose, 40000 101, Linda, 35000 102, Jack, 45000 103, Sam, 50000

Which of the below code fragment can will help create an appropriate Dataframe named "EmployeeDF" for the same?

Stru DSU.



nes "EmployeeData cav" on HDFS folder

aframe named "EmployeeDF" for the

Choose the best option

- Schema! = StructTypeqStructField(Tops (D*, orangesTypes), Trust) StructField("EmpName", StringType(), True):
- StructField("Salary", IntegerType(), True (I)

EmployeeDF =

spark.read.format('com.databricks.spark.csv') option('desmater'; ') schematischematisch

- EmployeeDF = spark.read.("/SparkDemos/EmployeeData cov")
- schema1 = StructType([StructField("Dept_ID", IntegerType(), True)) StructField("EmpName",StringType(),True),\ StructField("Salary",IntegerType(),True)])

EmployeeDF = spark.read.schema(schema1).load("/SparkDemos/EmployeeD

EmployeeDF = spark.read.format('com.databricks.spark.csv').option('delimiter',",").schema /EmployeeData.csv")



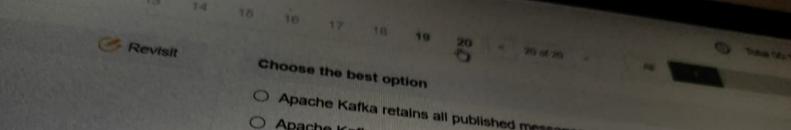
Which of the following statement is True about Apache Kafka?

Bigetata and Harloop Main McG 1 6

(% Peval)

Choose the best option

- Apache Kalka retains all published me
- Apache Kafka does not retain all put
 Apache Kafka retains all published
 - O None of the above



- Apache Kafka retains all published messages regardless of consumption.
- Apache Kafka does not retain all published messages regardless of consumption Apache Kafka retains all published & unconsumed messages
- None of the above

When sport events happen, Tweeter get lots of posts related to that event Sports association wanted to know how Which framework can be used bere?

Choose the best option

O Spark

O Hadoop

O Kafka

Kubernetes

Kafka

Section 1 of 1 Section at 1 Section at 1 Section at 1 Section 2 of 1 Section 2 of

a

C) Hadoop 3.0



Bigdets and Hartop Main MCG 1 @

Of Payet

If we want to see the columns names of the table in HiveQL, the following hive configuration property should be used

Choose the best option

1 1 D 1 4 7 4 5 11 11 12 12 - 44m -

- set hive cli print header-true;
- O set hive cli print header-false ;
- set rive di current db-true;
- o set hive di current do-false ;

Which of the following is true about HDFS?

C Revisit

Choose the best option

6 0 7 0 9 10 11 12 15 - 6 of 20 -

O It is a distributed file system.

It is used for batch processing of data

O It is written in Java

O All of the above

d



Section 1 of 1 Section #1

Question # 7

Ch Revisit

From the below steps, choose right sequence of steps required when an HDFS Client wants to create a File on the Data Node ?

- 1) HDFS Client Write the Packet on the Data Node.
- 2) HDFS Client gets the block location from the Name Node.
- 3) Data Node will send acknowledgement back to the HDFS Client.
- 4) HDFS Client will send a create file request to the Name Node.

Choose the most appropriate Sequence from the below options.

duff Choose the best option

2 2 4 5 0 5 5 5 10 10 10 12 10 - Term -

01 2 3 4

04 2 1 3

02 3 4 1

Which of following file is used to specify the Scheduling Policy

Revisit

Choose the best option

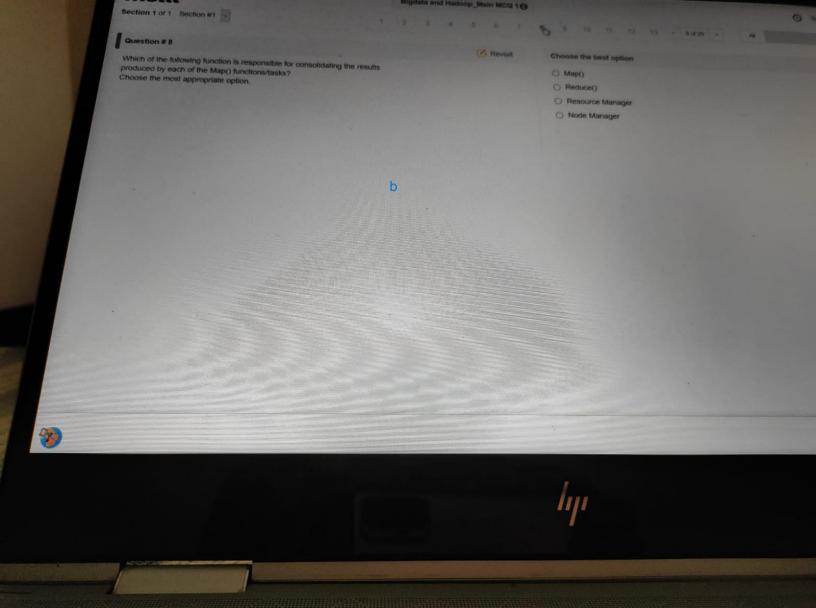
O yarn.xml

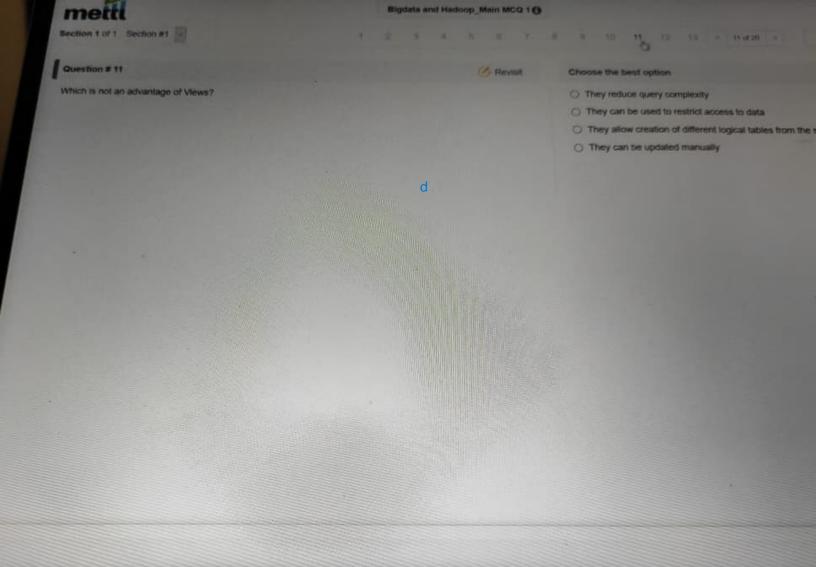
O yarn-site.xml

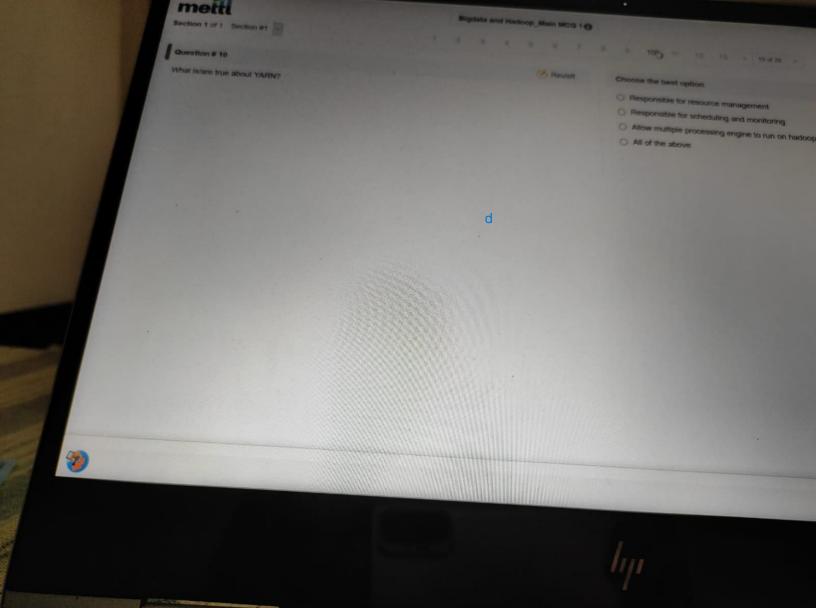
O hdfs-site.xml

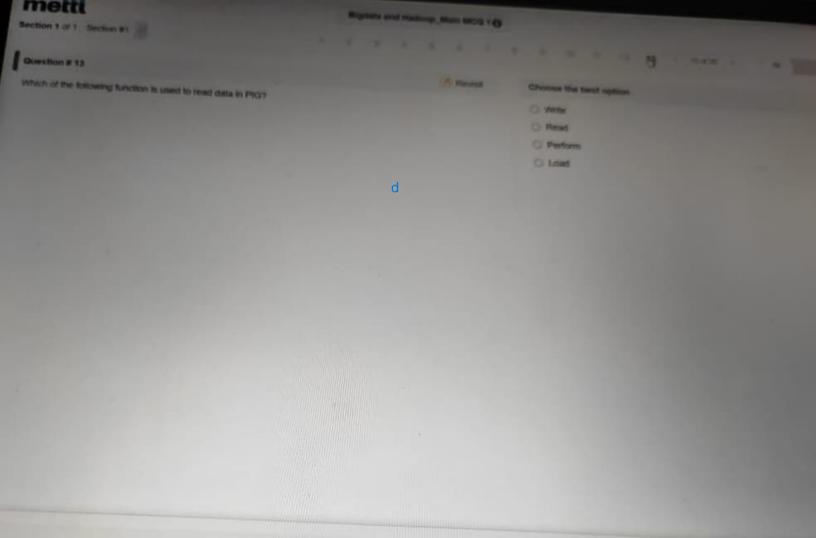
O hdfs.xml

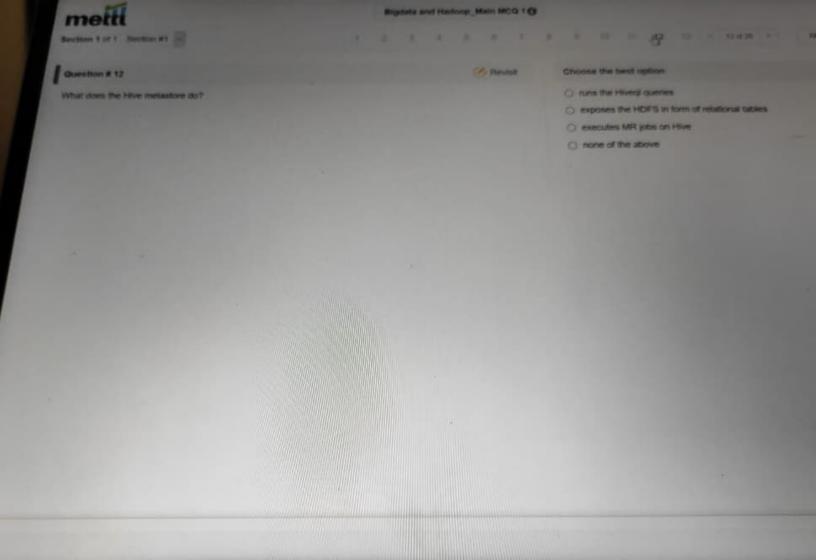
b



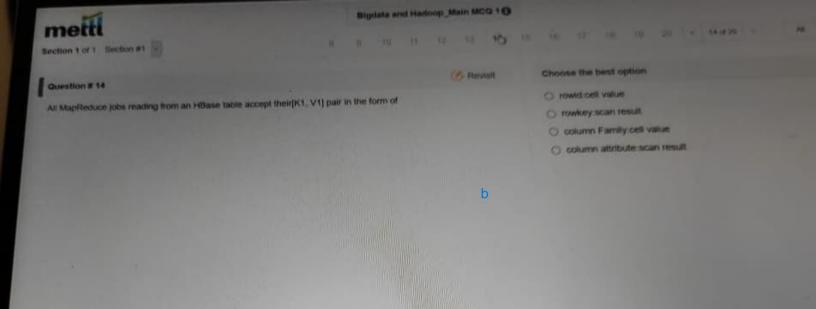








mettl Bigdeta and Hadoop_Main MCG 1.O Section 1 of 1 Section #1 The second secon Question # 17 True or False: By default streaming context is present in pyspark shell. CPS RESIDE Choose the best option C): true O false



Assume that we have a the below Employee data in cav file names "EmployeeData.csv" on HDFS folder

101.Alex.20000 102.Adam,30000 103,Rose,40000 101, Linda, 35000 102, Jack, 45000

103,Sam,50000

Which of the below code fragment can will help create an appropriate Dataframe named "EmployeeDF" for the

Choose the best option.

A. 10 11. 10 15 16 10 16 10 19 10 10 20 1 10 100 100

- schema1 = StructType([StructField:"Dept.ID", IntegerType(), True). StructField("Emphame", StringType(),True).\ StructField("Salary", IntegerType(), True)()
 - EmployeeDF = spark read format/com databricks spark csv/y option/desimater = 1 /EmployeeBata.csv*)
- EmployeeDF = spark read (*/SparkDemos/EmployeeData csv*) schema1 = StructType([StructField("Dept_ID", IntegerType(), T StructField("EmpName",StringType(),True),\ StructField(*Salary*,integerType(),True)])

EmployeeDF = spark.read.schema(schema1).load("/Sparkt

O EmployeeDF = spark.read.format('com.databricks.spark.csv').option('delic /EmployeeData.csv*)

Section 1 of 1 Decision 91

Downton # 15

intends of the following is an output operation on Databases $\boldsymbol{\tau}$

IN HARMS

Change he had output

and test to

SECON.

C) SHIPE

C) restaurablishes



Rigdets and Hadrop, Main McG 1 (5)

Question # 19

Py Davish

Choose the best option.

Communication between the clients and the servers is done with a simple, high-performance, language agreence protocol.

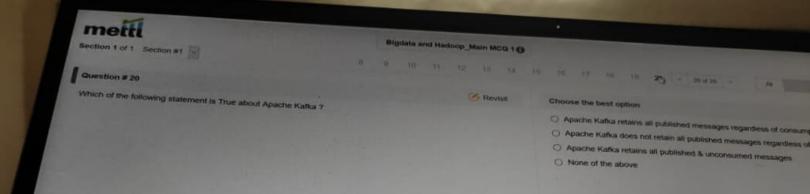
OP.

O TOP

O SMTP

O ICMP

b(not sure)

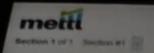


Which of the following is not a category of NoSQL db?

(Revisit

Choose the best option

- O Graph
- Scientific
- O Document
- O PostGressSQL Clear Response



Bigdata and Hartoop, Main MCQ 1 (3)

Question # 18

(Stevent

Your organization want to develop one streaming application which gives latest status of vaccination dose completed in your organization since covid-19 vaccination started. Which method will fulfil requirement?

Choose the pest option

C) reduceByKey

mduceByKeyAndWindow

O updateStateByKey

C) countStateByKey