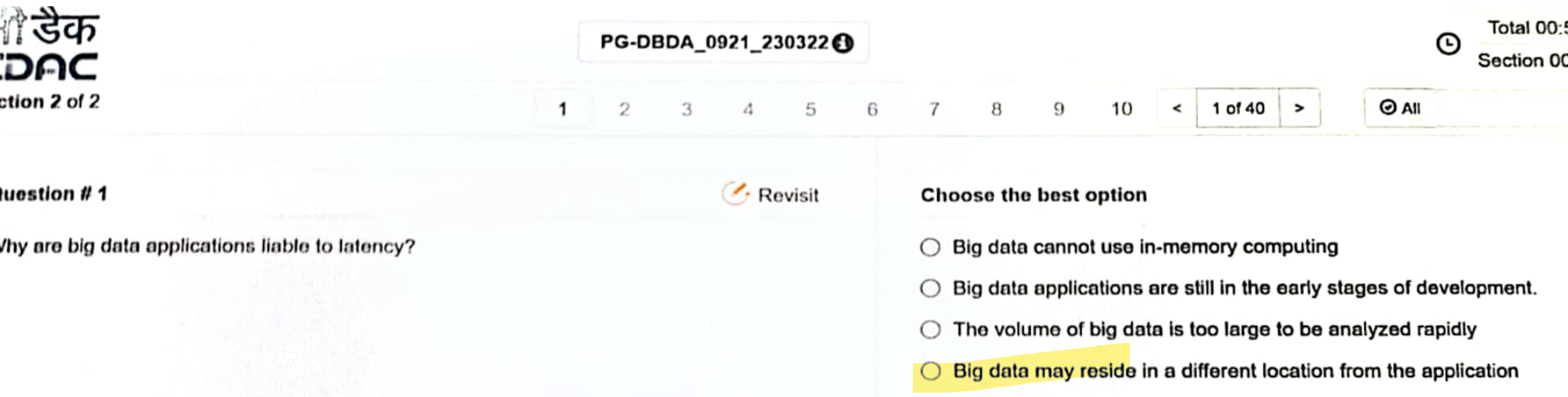


uestion # 1

/hy are big data applications liable to latency?

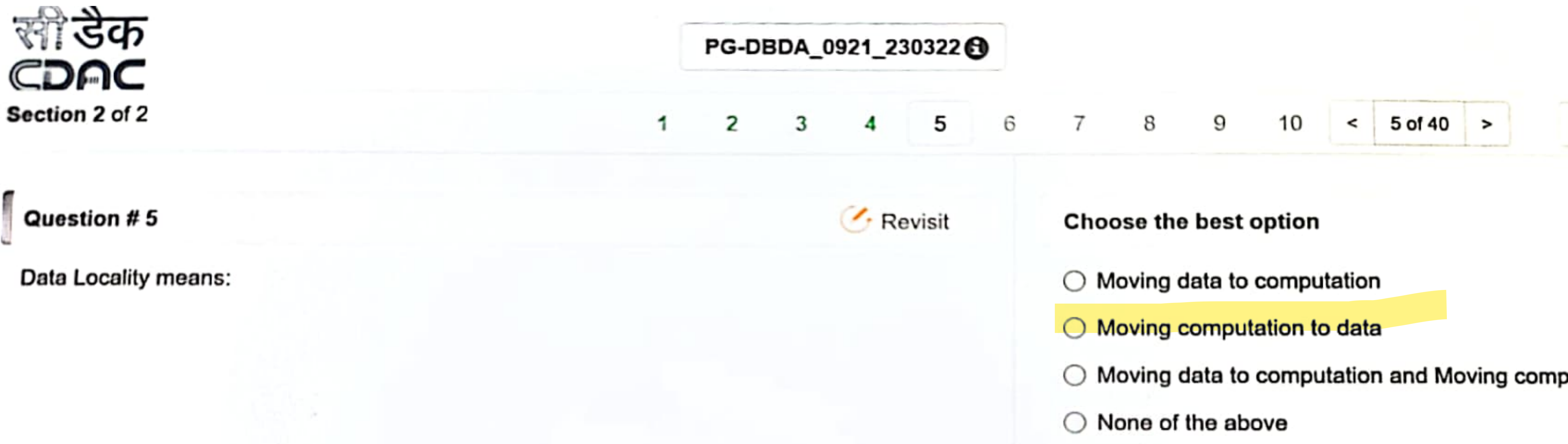
Choose the best option

○ Big data cannot use in-memory computing

○ Big data applications are still in the early stages of development.

○ The volume of big data is too large to be analyzed rapidly

○ [✓ Big data may reside in a different location from the application]



Section 2 of 2

PG-DBDA_0921_230322

1 2 3 4 5 6 7 8 9 10 < 5 of 40 >

Question # 5

Revisit

Data Locality means:

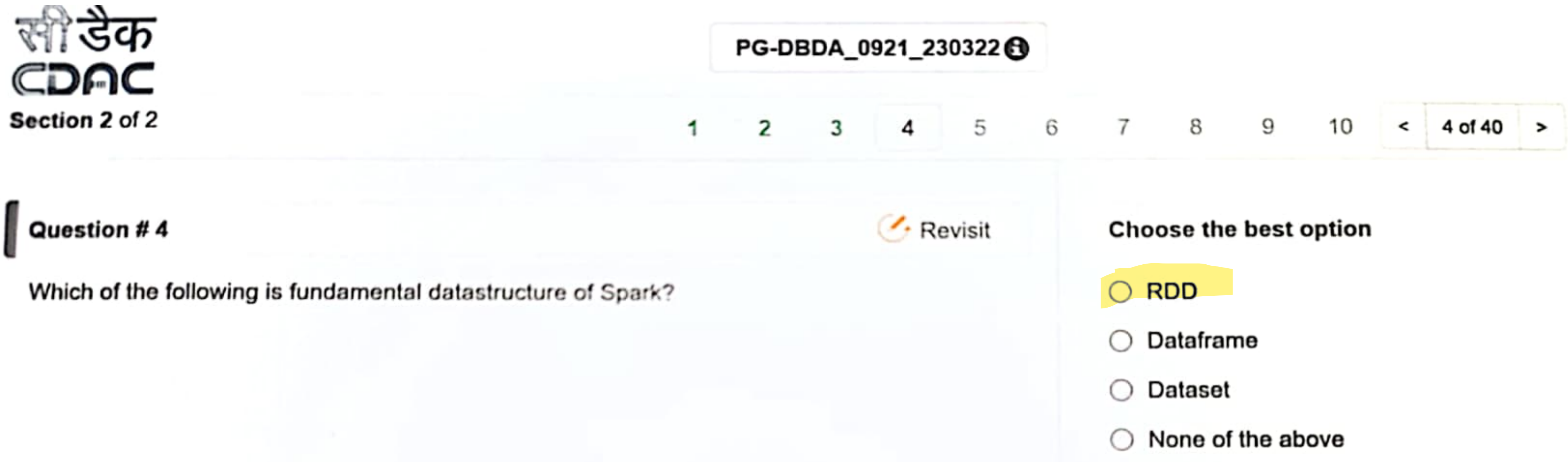
Choose the best option

○ Moving data to computation

[✓ ○ Moving computation to data]

○ Moving data to computation and Moving comp

○ None of the above



सी डैक
CDAC

Section 2 of 2

PG-DBDA_0921_230322

1 2 3 4 5 6 7 8 9 10 < 4 of 40 >

Question # 4

Revisit

Which of the following is fundamental datastructure of Spark?

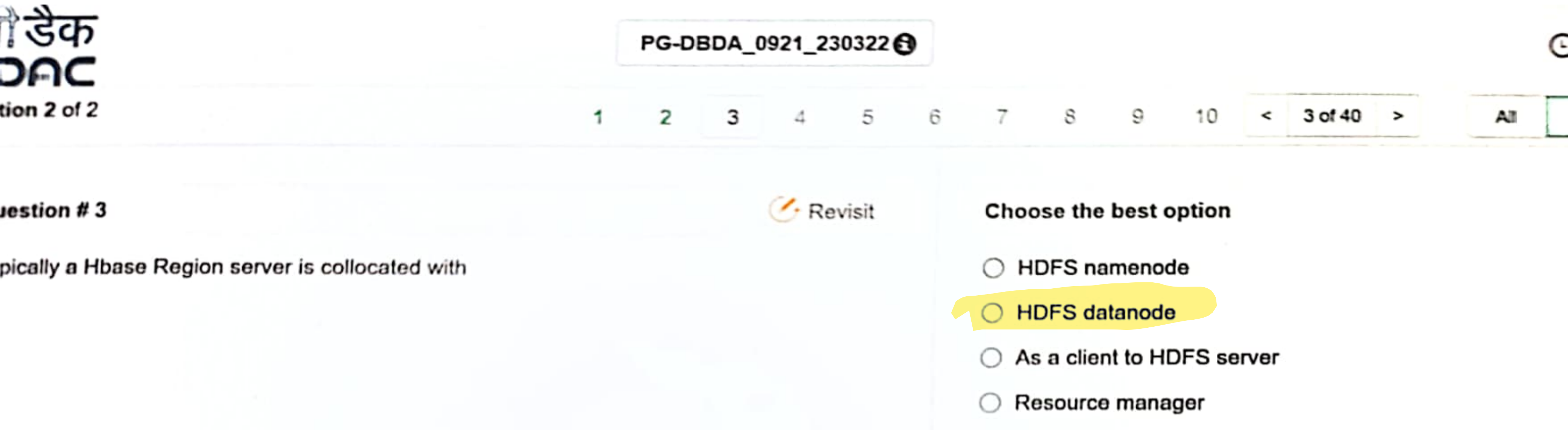
Choose the best option

[✓ ○ RDD]

○ Dataframe

○ Dataset

○ None of the above



PG-DBDA_0921_230322

uestion # 3

pically a Hbase Region server is collocated with

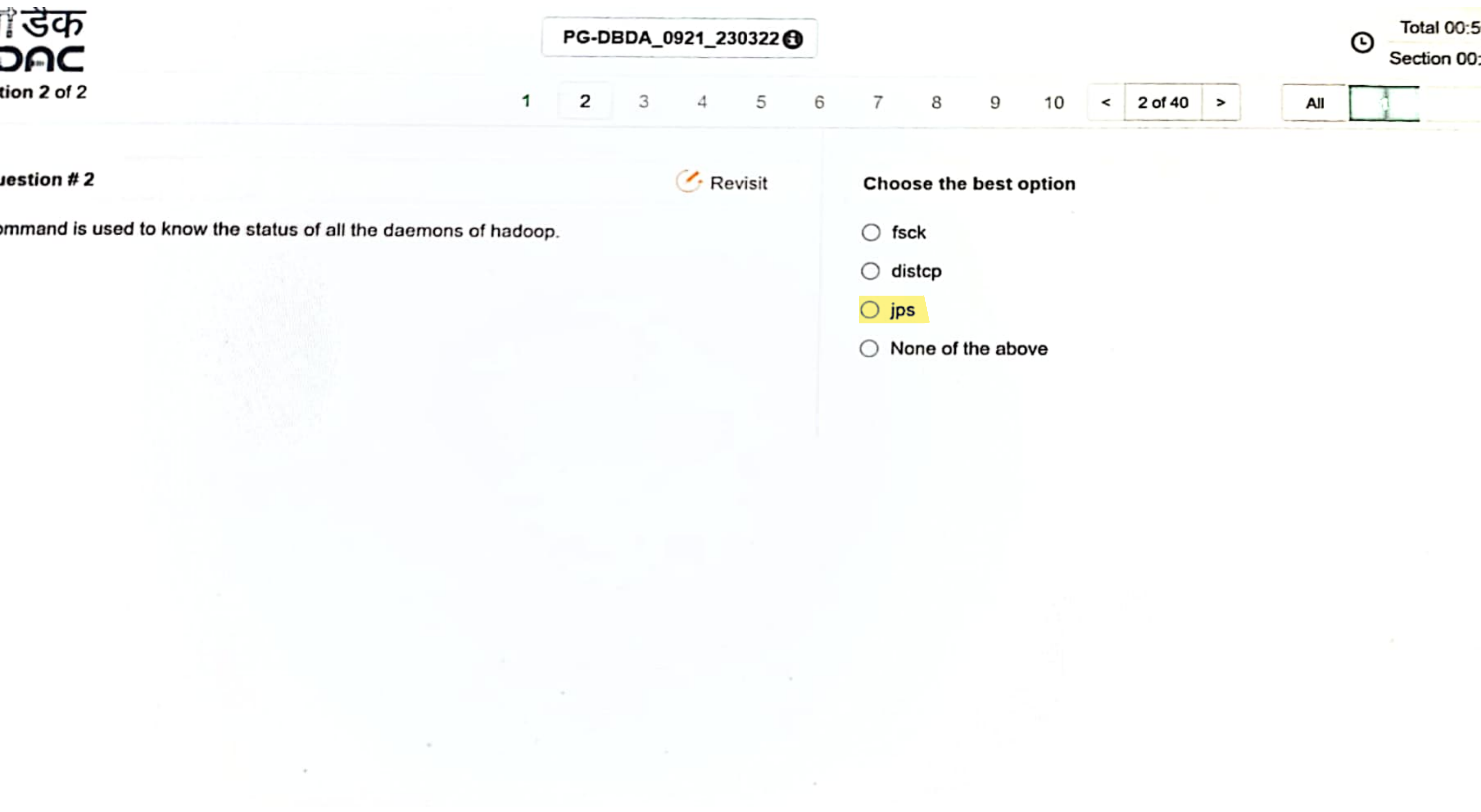
Choose the best option

○ HDFS namenode

[✓ ○ HDFS datanode]

○ As a client to HDFS server

○ Resource manager



uestion # 2

ommand is used to know the status of all the daemons of hadoop.

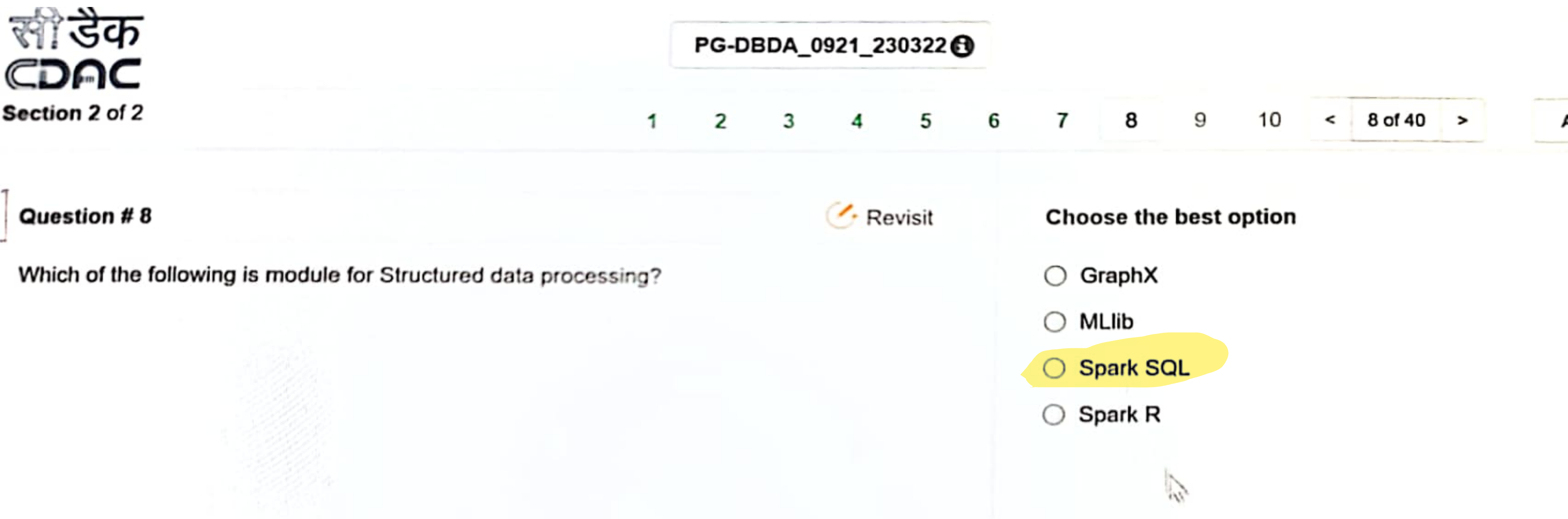
Choose the best option

○ fsck

○ distcp

○ jps

○ None of the above



PG-DBDA_0921_230322

Question # 8

Which of the following is module for Structured data processing?

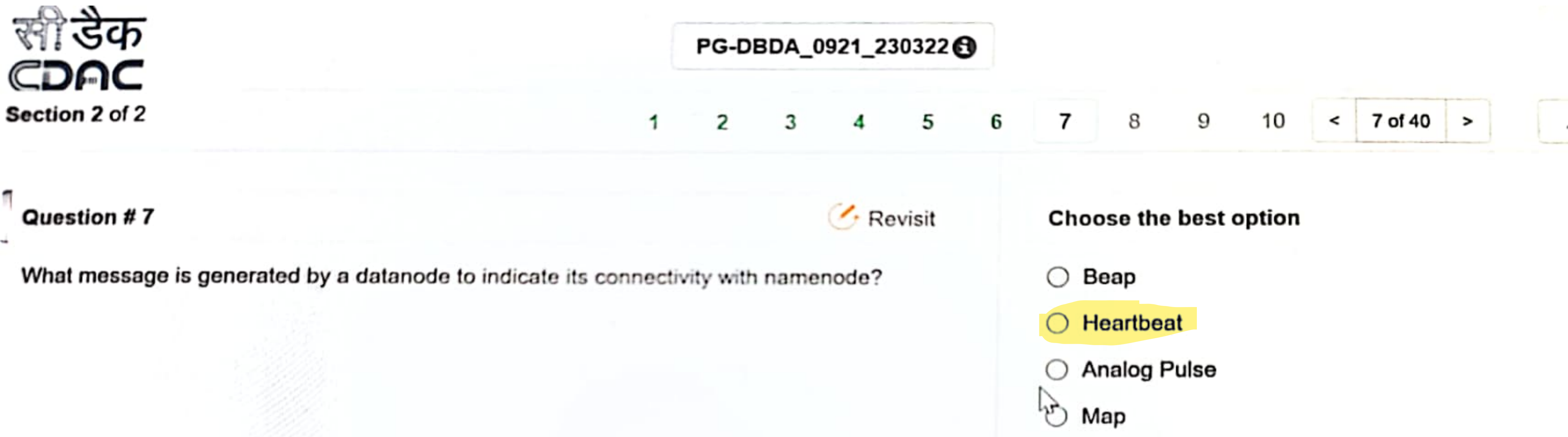
Choose the best option

○ GraphX

○ MLlib

[✓ ○ Spark SQL]

○ Spark R



PG-DBDA_0921_230322

Section 2 of 2

Question # 7

What message is generated by a datanode to indicate its connectivity with namenode?

Choose the best option

○ Beap

○ Heartbeat

○ Analog Pulse

○ Map

Question # 6

__________ is not a component of spark.

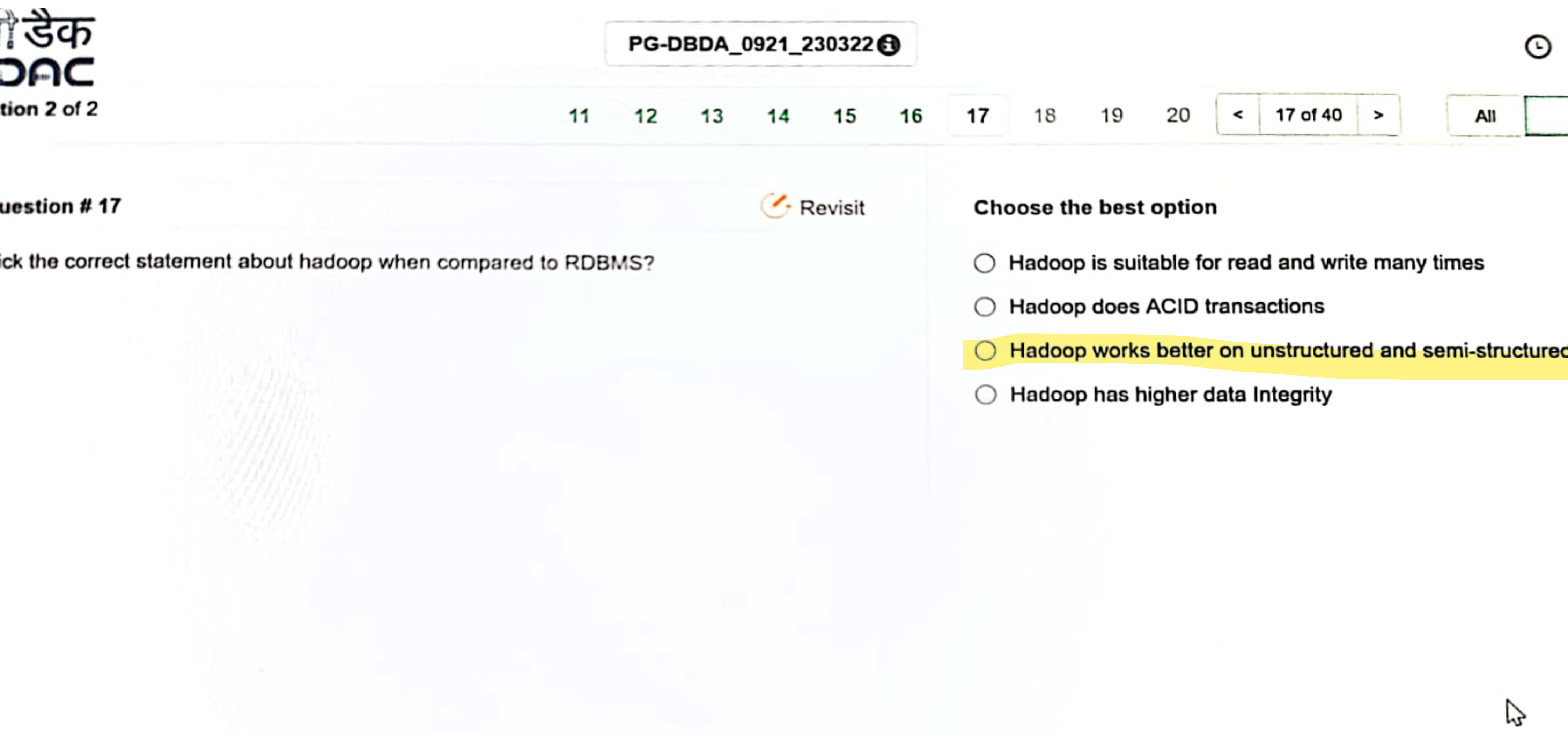
Choose the best option

○ SparkR

○ Mllib

○ GraphX

○ [✓ Sqoop]



uestion # 17

ick the correct statement about hadoop when compared to RDBMS?

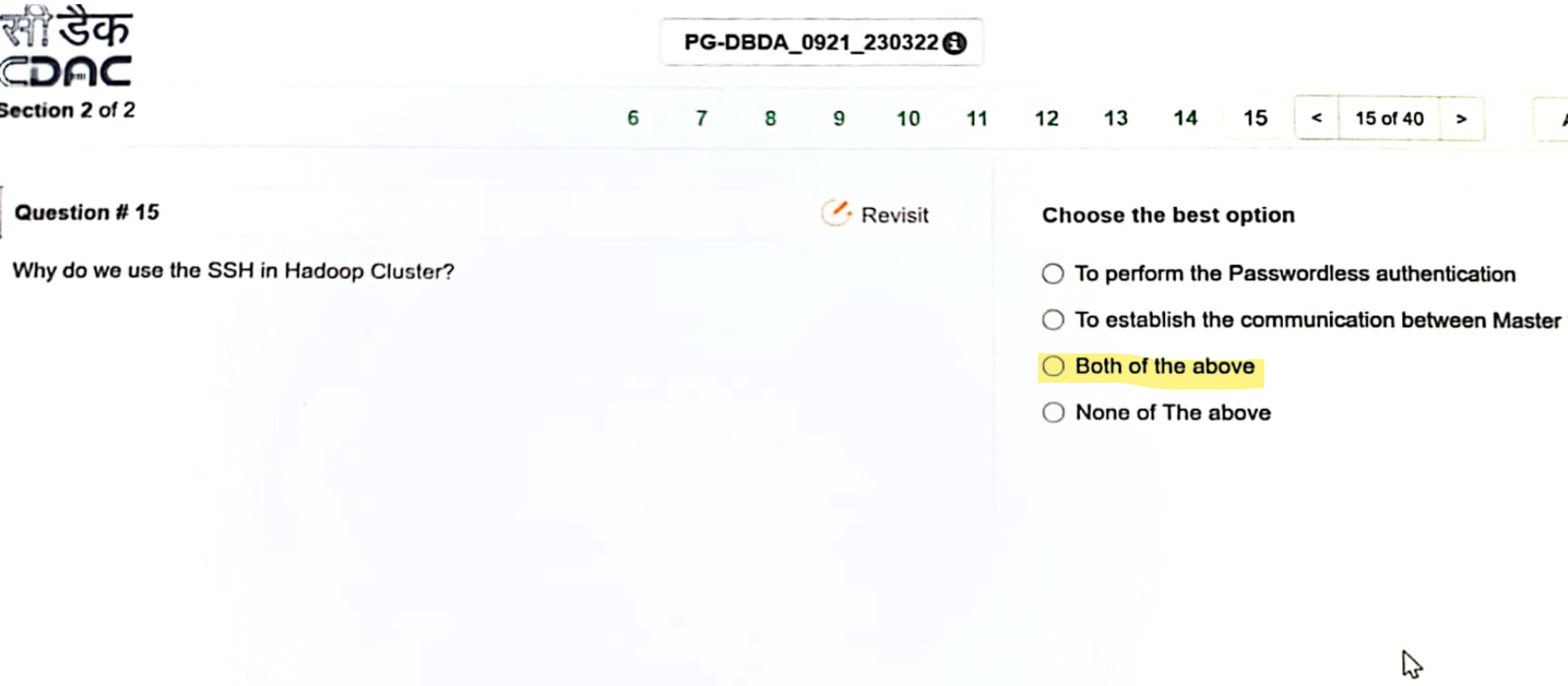
Choose the best option

○ Hadoop is suitable for read and write many times

○ Hadoop does ACID transactions

○ Hadoop works better on unstructured and semi-structurec

○ Hadoop has higher data Integrity



PG-DBDA_0921_230322

Question # 15

Why do we use the SSH in Hadoop Cluster?

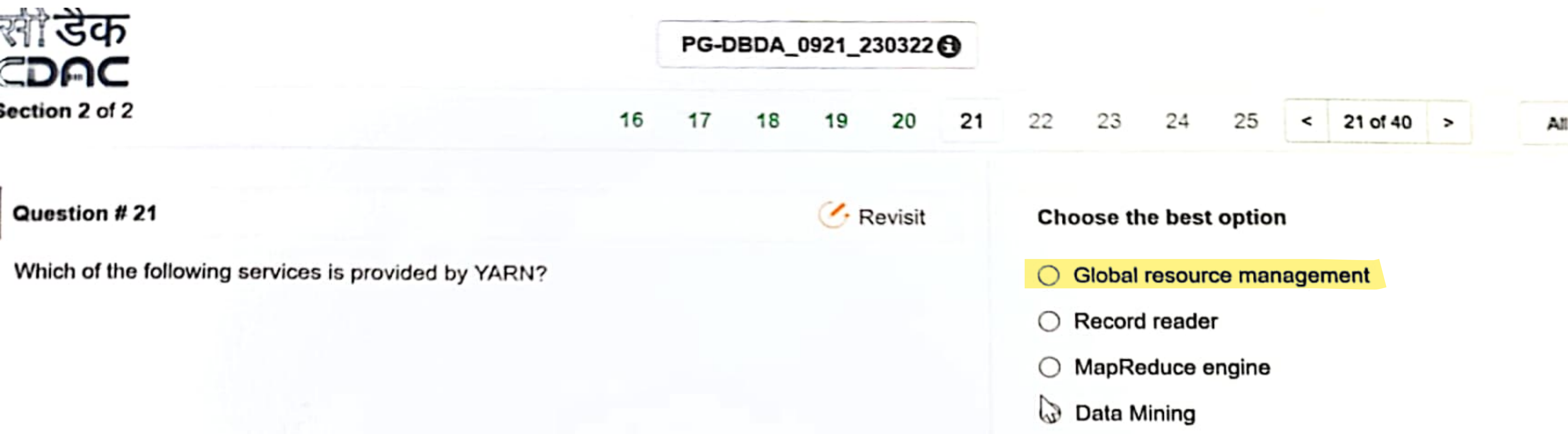
Choose the best option

○ To perform the Passwordless authentication

○ To establish the communication between Master

[✓ ○ Both of the above]

○ None of The above



PG-DBDA_0921_230322

Section 2 of 2

Question # 21

Which of the following services is provided by YARN?

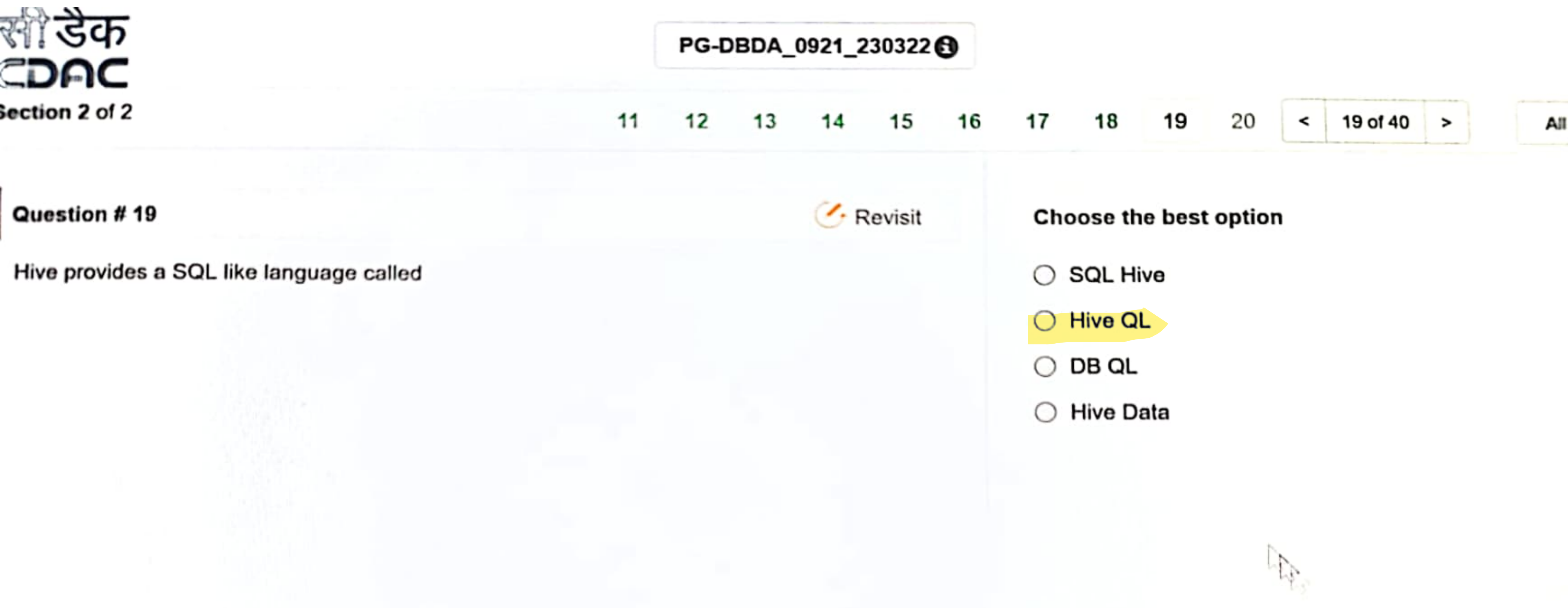
Choose the best option

○ Global resource management

○ Record reader

○ MapReduce engine

○ Data Mining



Question # 19

Hive provides a SQL like language called

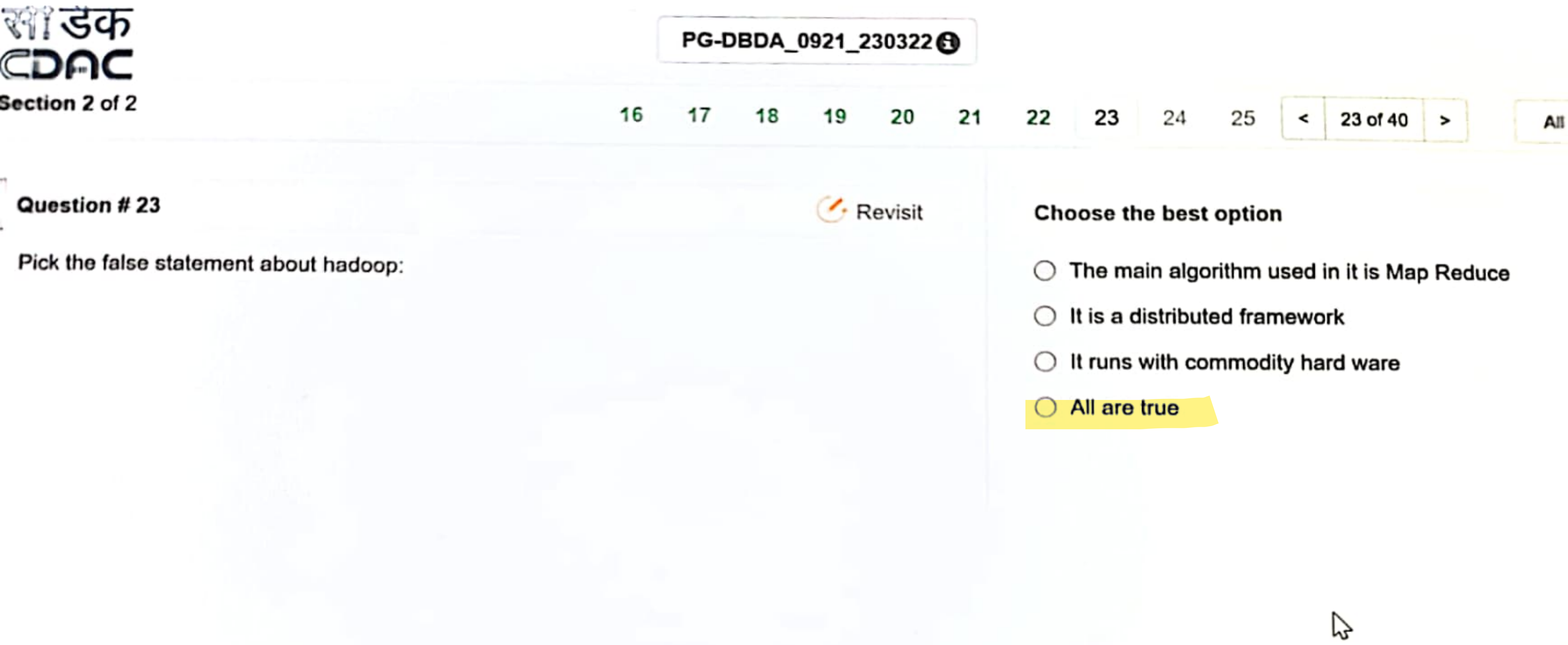
Choose the best option

○ SQL Hive

[✓ ○ Hive QL]

○ DB QL

○ Hive Data



PG-DBDA_0921_230322

Section 2 of 2

Question # 23

Pick the false statement about hadoop:

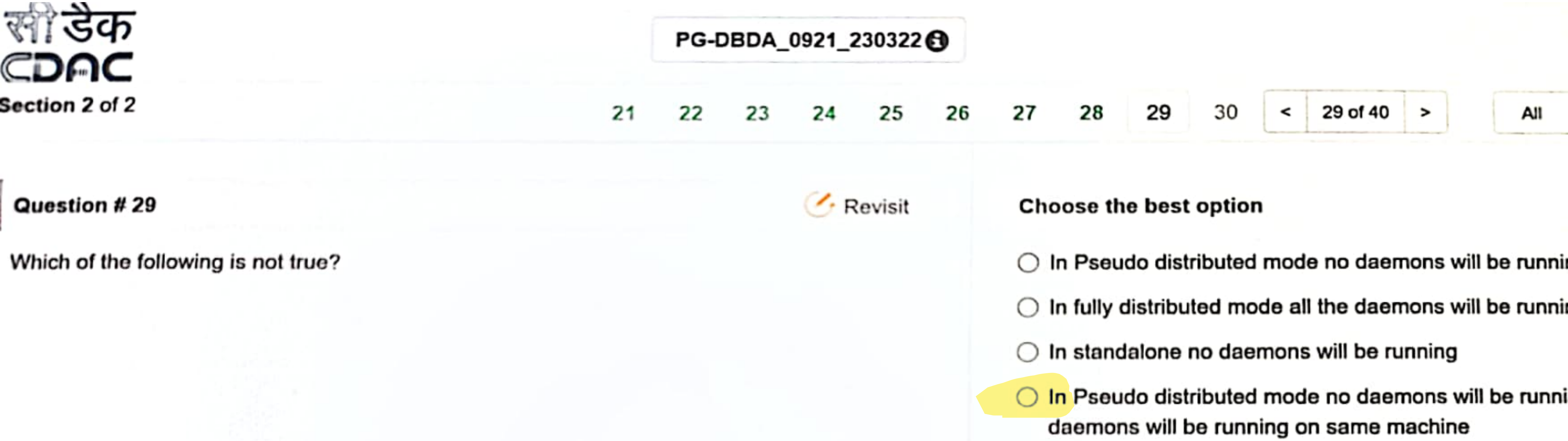
Choose the best option

○ The main algorithm used in it is Map Reduce

○ It is a distributed framework

○ It runs with commodity hard ware

○ All are true



Section 2 of 2

PG-DBDA_0921_230322

Question # 29

Which of the following is not true?

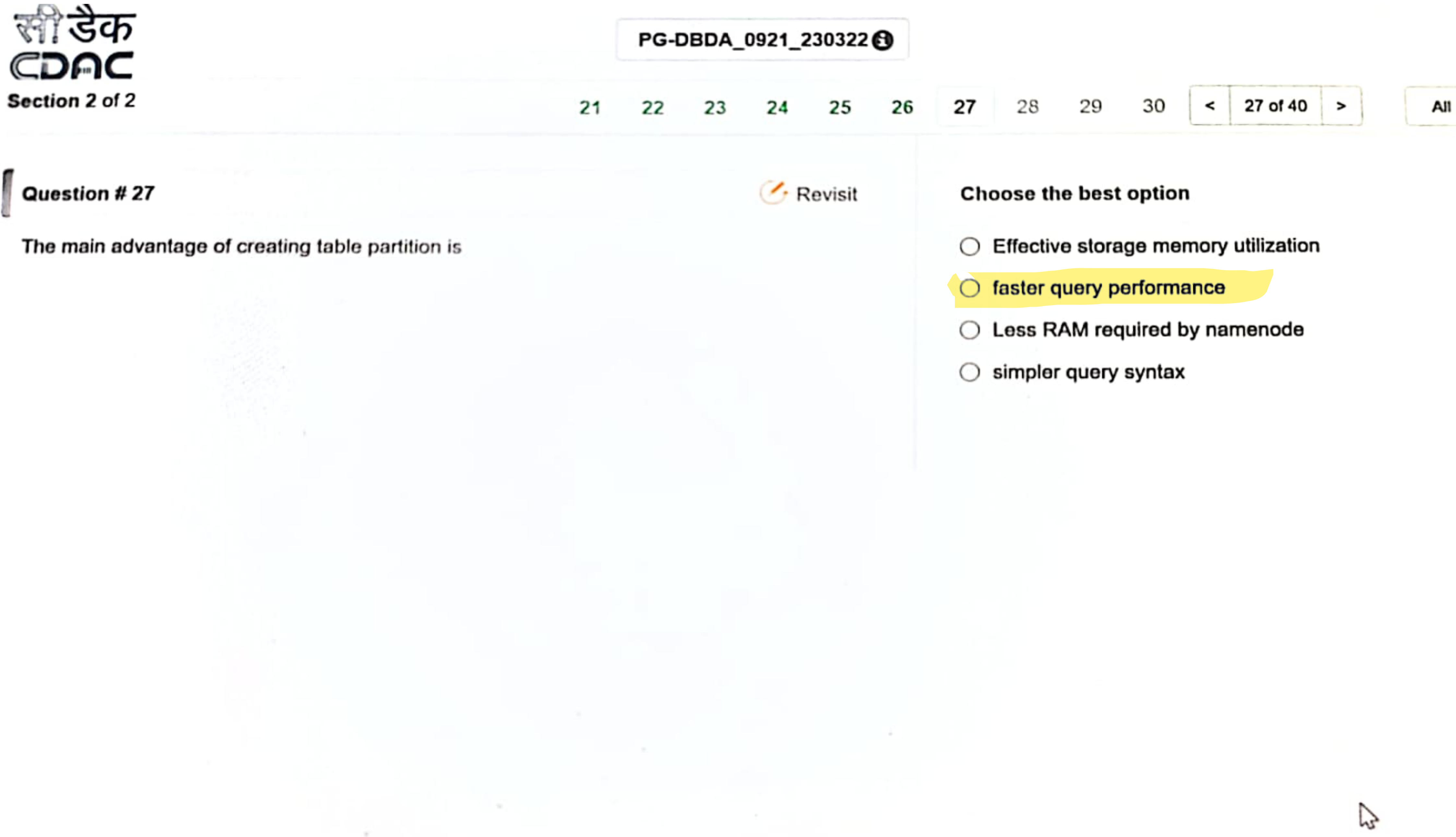
Choose the best option

○ In Pseudo distributed mode no daemons will be runni

○ In fully distributed mode all the daemons will be runni

○ In standalone no daemons will be running

○ In Pseudo distributed mode no daemons will be runni
daemons will be running on same machine



Question # 27

The main advantage of creating table partition is

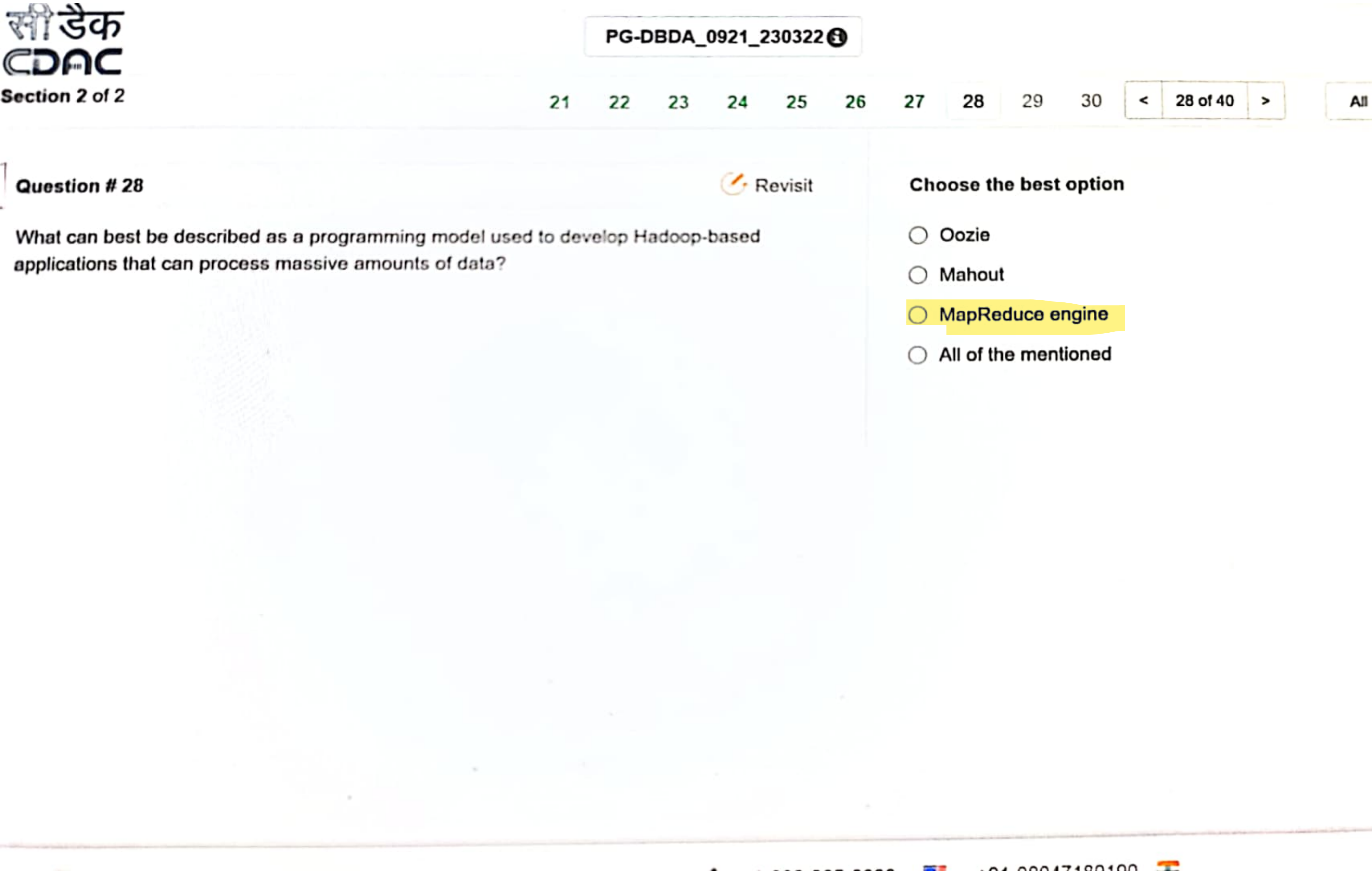
Choose the best option

○ Effective storage memory utilization

○ faster query performance

○ Less RAM required by namenode

○ simpler query syntax



PG-DBDA_0921_230322

Section 2 of 2

21 22 23 24 25 26 27 28 29 30 < 28 of 40 > All

Question # 28

Revisit

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

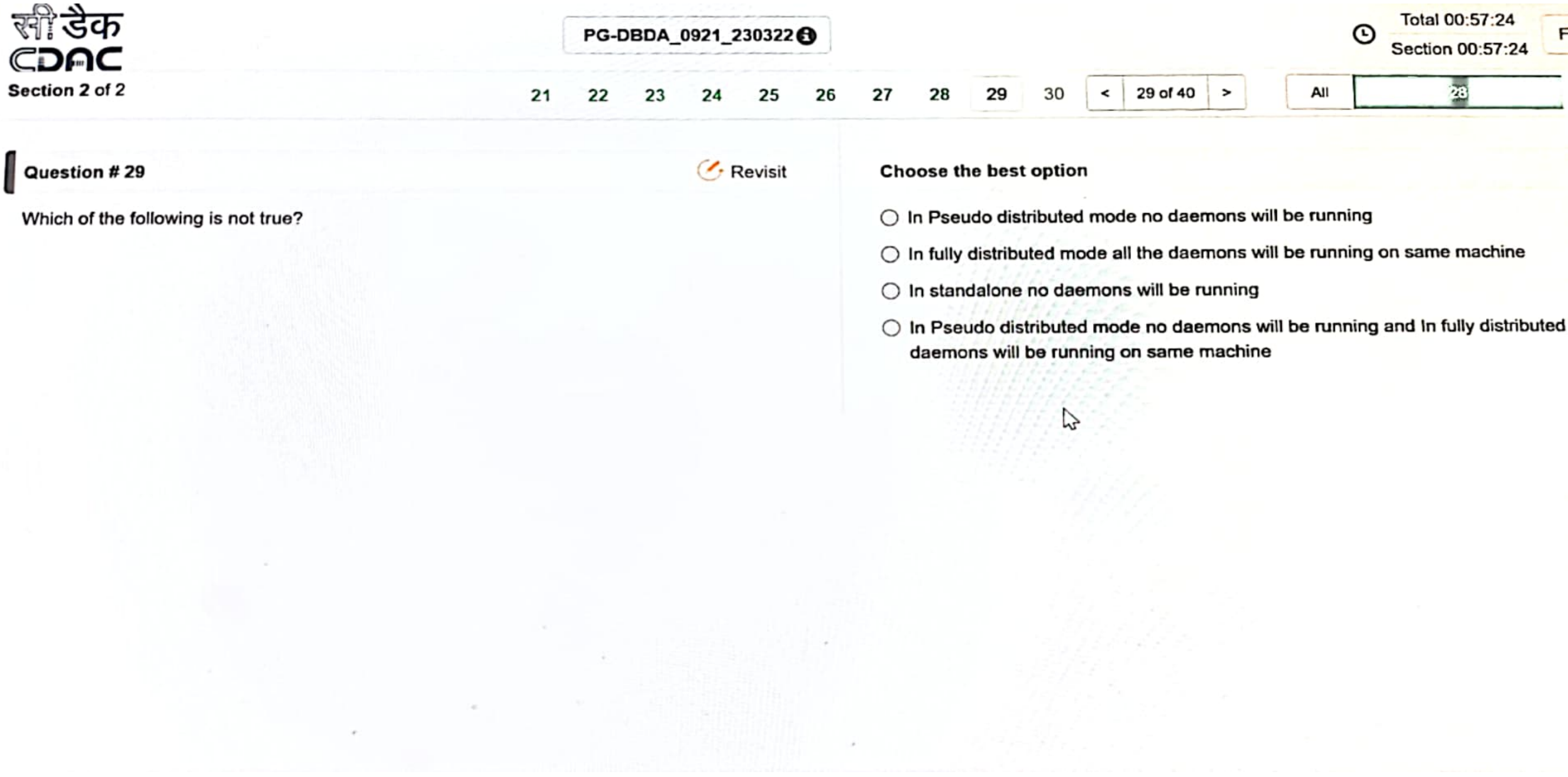
Choose the best option

○ Oozie

○ Mahout

○ [✓ MapReduce engine]

○ All of the mentioned



Question # 29

Which of the following is not true?

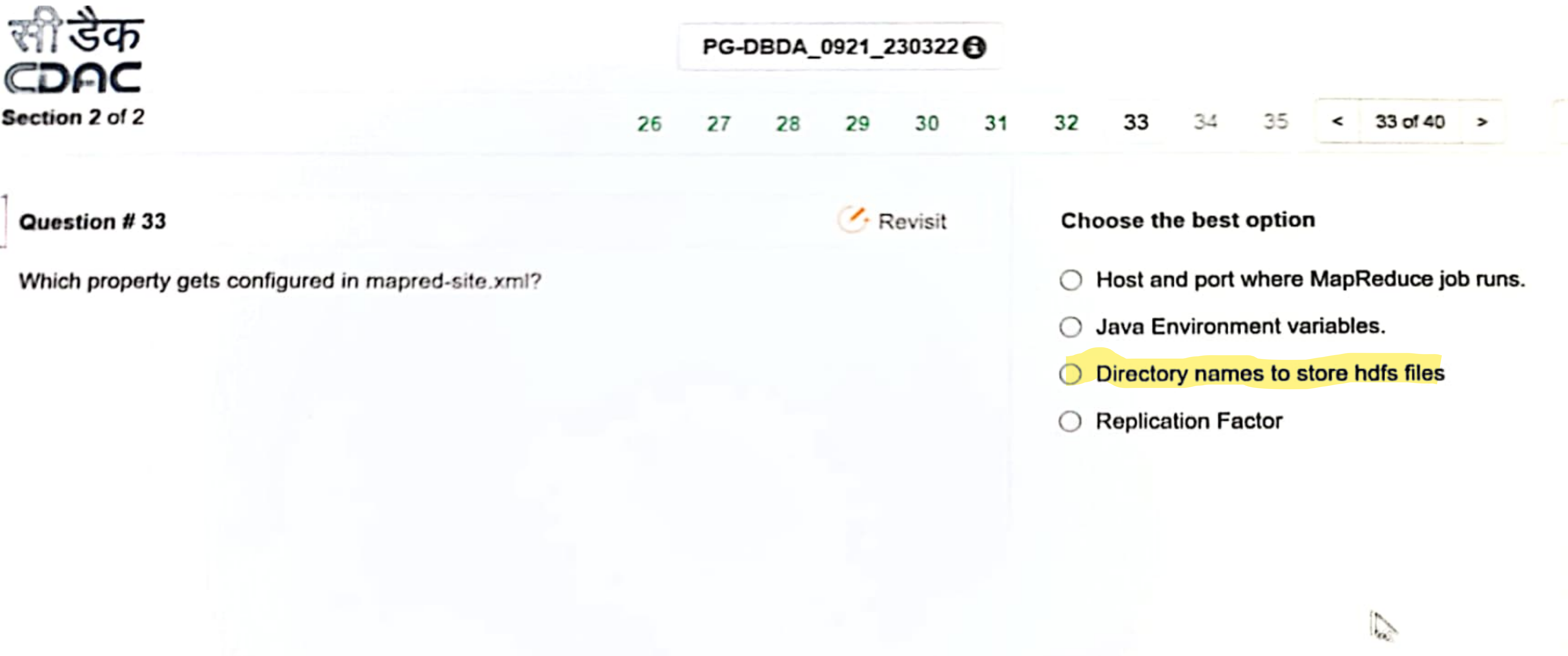
Choose the best option

○ In Pseudo distributed mode no daemons will be running

○ In fully distributed mode all the daemons will be running on same machine

○ In standalone no daemons will be running

○ In Pseudo distributed mode no daemons will be running and In fully distributed daemons will be running on same machine



PG-DBDA_0921_230322

Section 2 of 2

26 27 28 29 30 31 32 33 34 35 < 33 of 40 >

Question # 33

Revisit

Which property gets configured in mapred-site.xml?

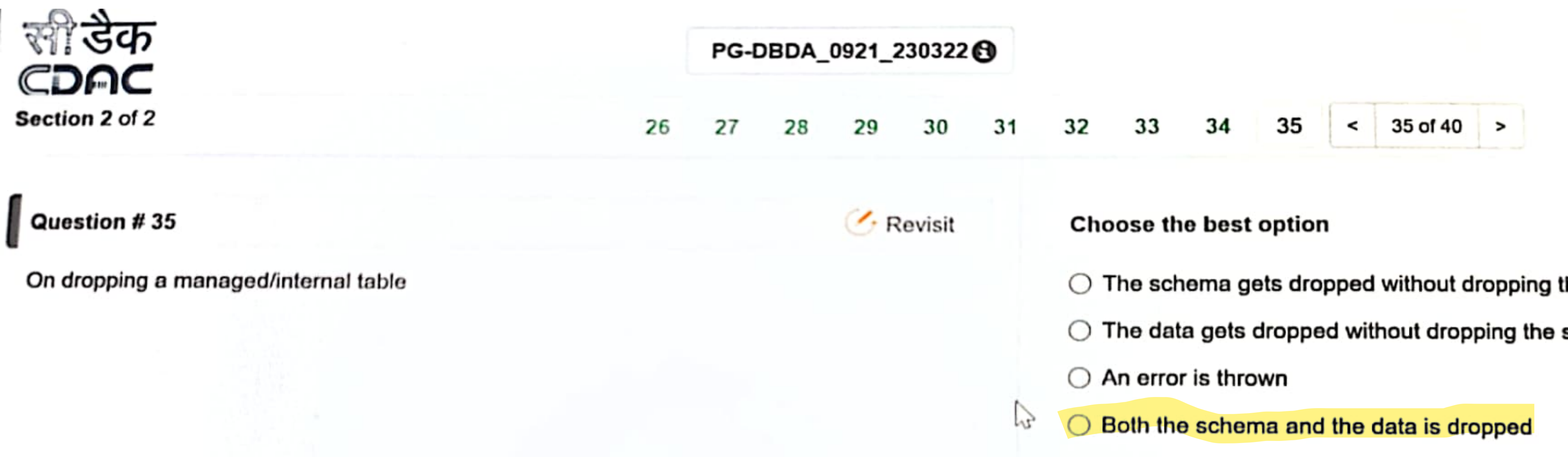
Choose the best option

○ Host and port where MapReduce job runs.

○ Java Environment variables.

[✓ ○ Directory names to store hdfs files]

○ Replication Factor



Section 2 of 2

PG-DBDA_0921_230322

Question # 35

On dropping a managed/internal table

Choose the best option

○ The schema gets dropped without dropping t

○ The data gets dropped without dropping the s

○ An error is thrown

○ Both the schema and the data is dropped

Question # 38

To retrieve all rows of a table in HBase, we use:

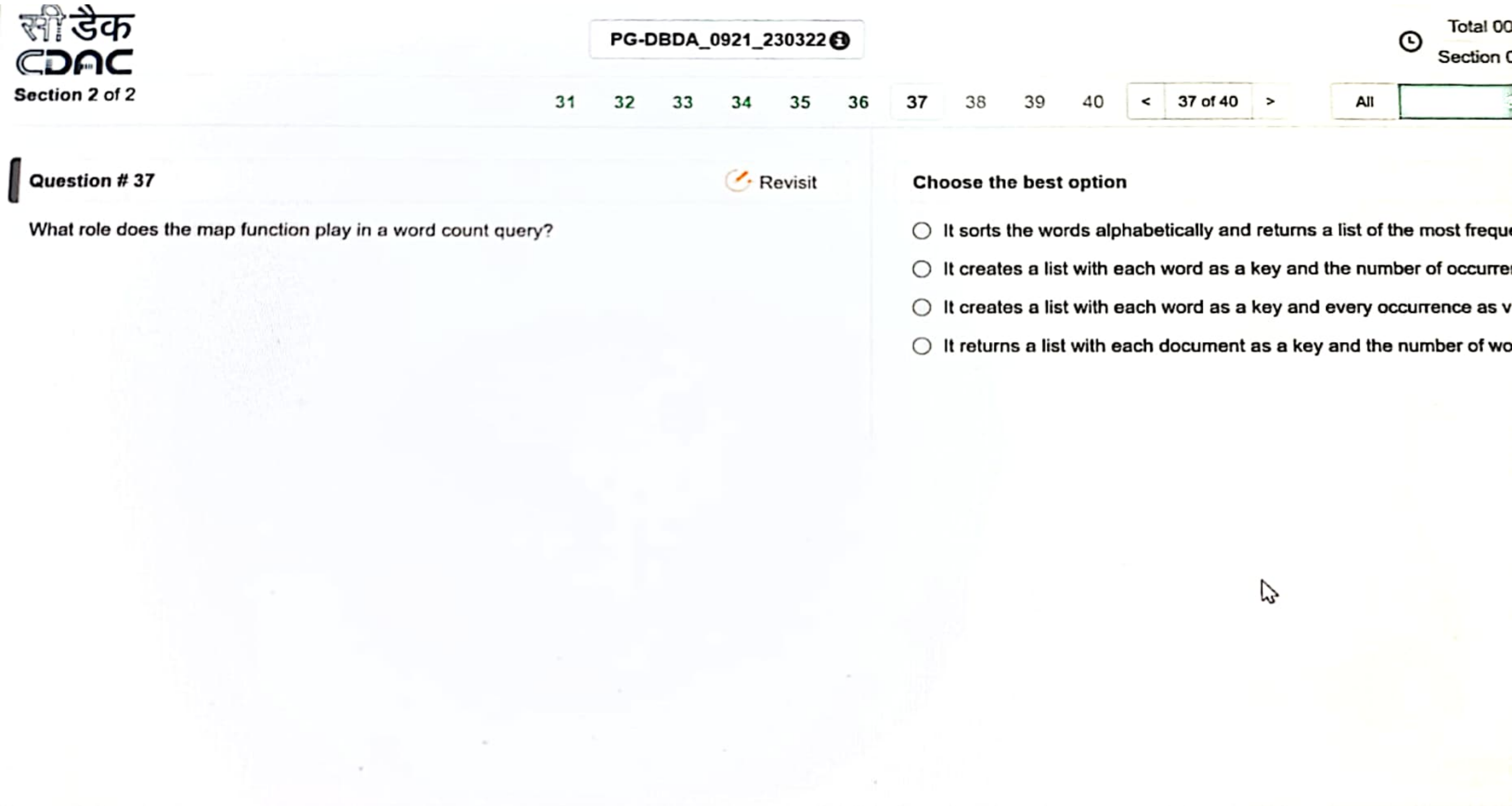
Choose the best option

○ get

○ scan

○ put

○ select



Section 2 of 2

PG-DBDA_0921_230322

Total 00

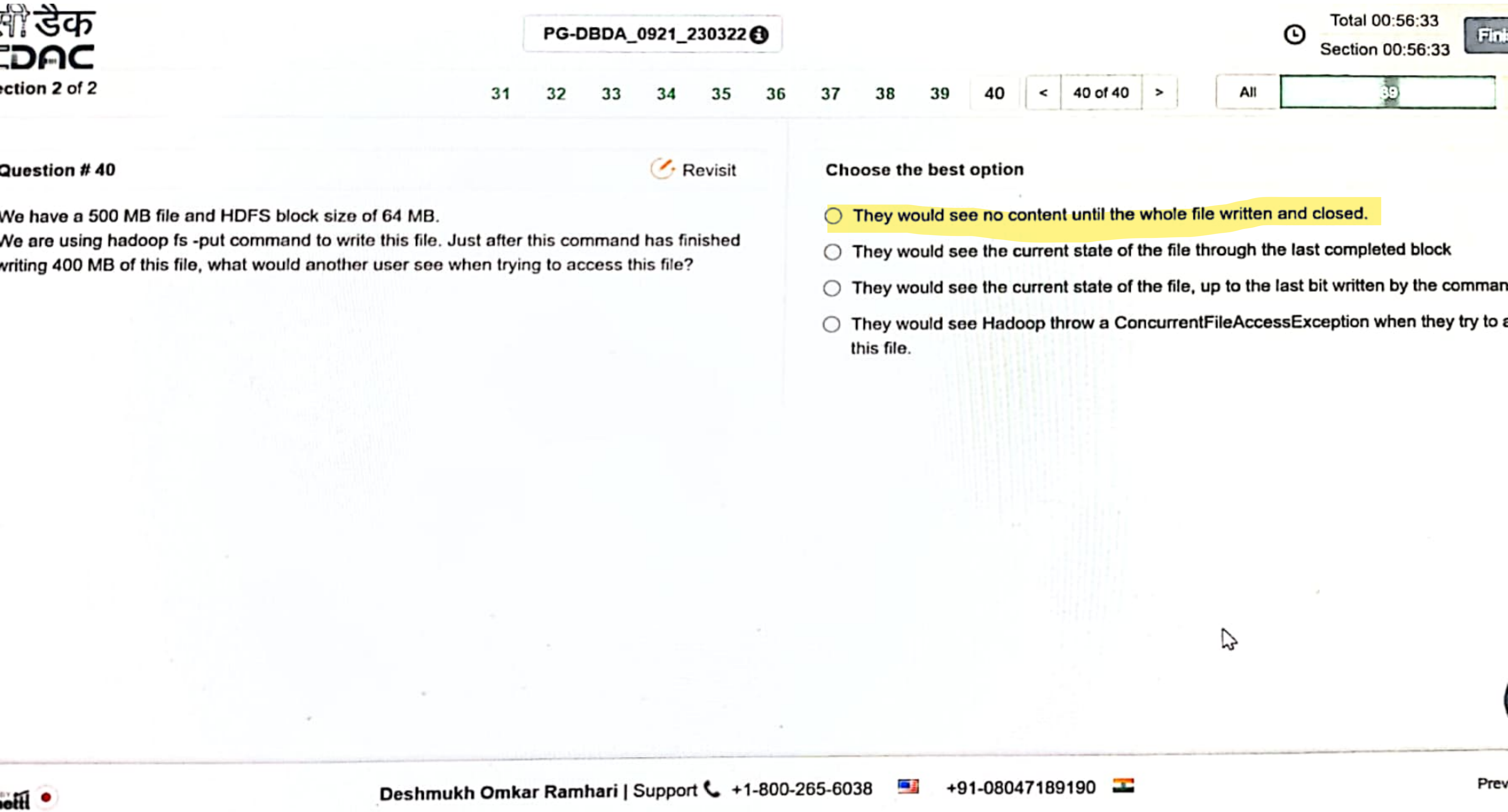
Section 0

Question # 37

What role does the map function play in a word count query?

Choose the best option

- ○ It sorts the words alphabetically and returns a list of the most freque
- ○ It creates a list with each word as a key and the number of occurre
- ○ It creates a list with each word as a key and every occurrence as v
- ○ It returns a list with each document as a key and the number of wo



ection 2 of 2

PG-DBDA_0921_230322

Total 00:56:33

Section 00:56:33

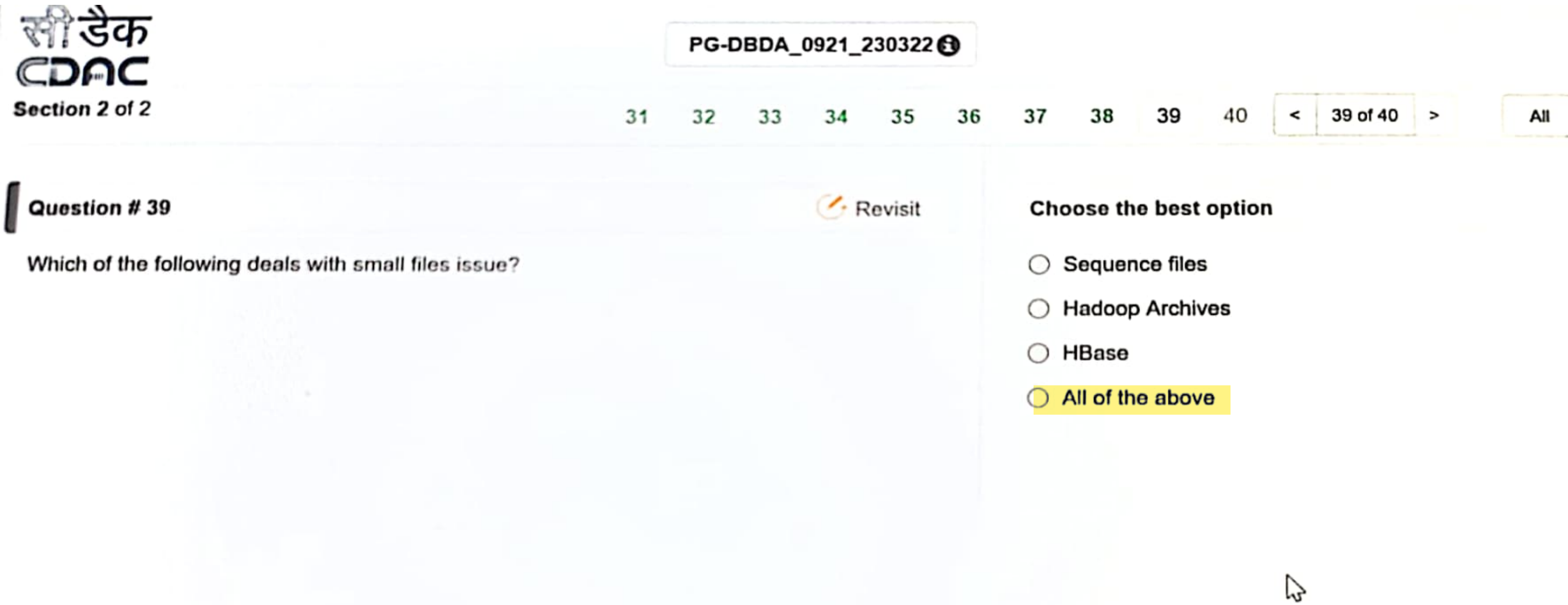
Question # 40

We have a 500 MB file and HDFS block size of 64 MB.
We are using hadoop fs -put command to write this file. Just after this command has finished writing 400 MB of this file, what would another user see when trying to access this file?

Choose the best option

- ○ They would see no content until the whole file written and closed.
- ○ They would see the current state of the file through the last completed block
- ○ They would see the current state of the file, up to the last bit written by the comman
- ○ They would see Hadoop throw a ConcurrentFileAccessException when they try to a this file.

Deshmukh Omkar Ramhari | Support ☎ +1-800-265-6038 +91-08047189190



Section 2 of 2

PG-DBDA_0921_230322

Question # 39

Which of the following deals with small files issue?

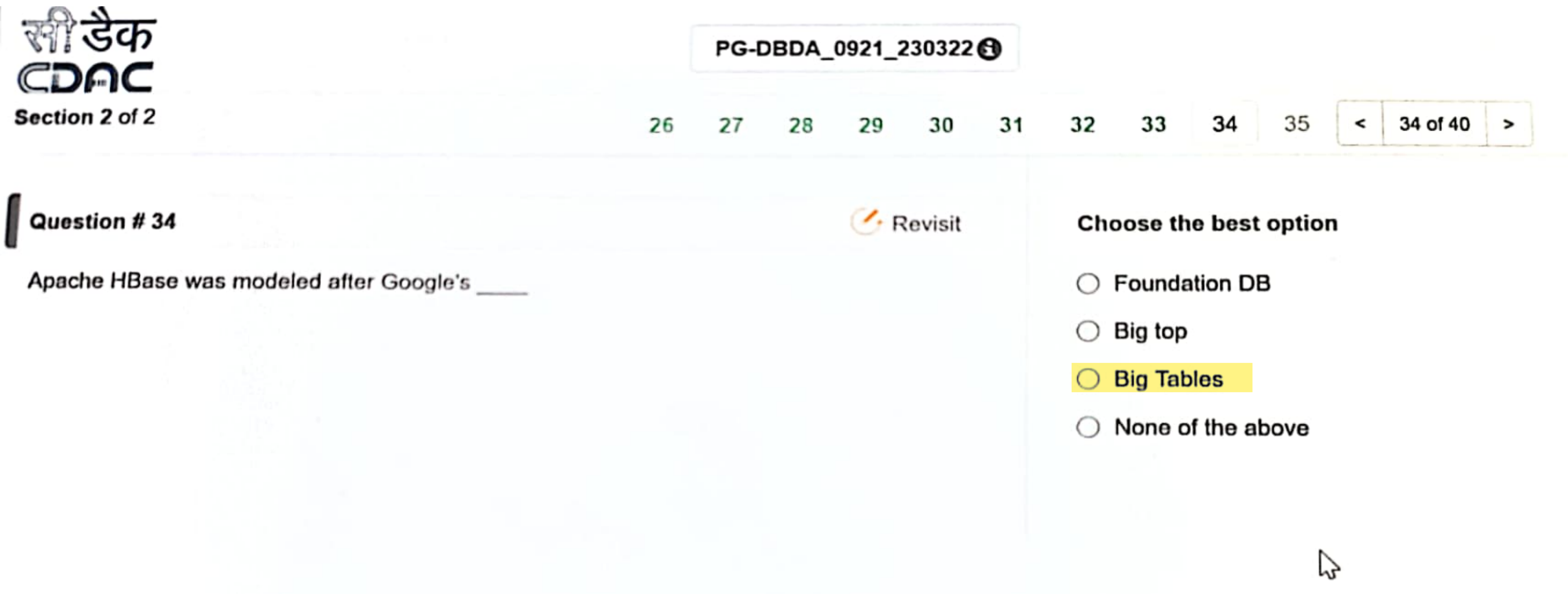
Choose the best option

○ Sequence files

○ Hadoop Archives

○ HBase

○ All of the above



Section 2 of 2

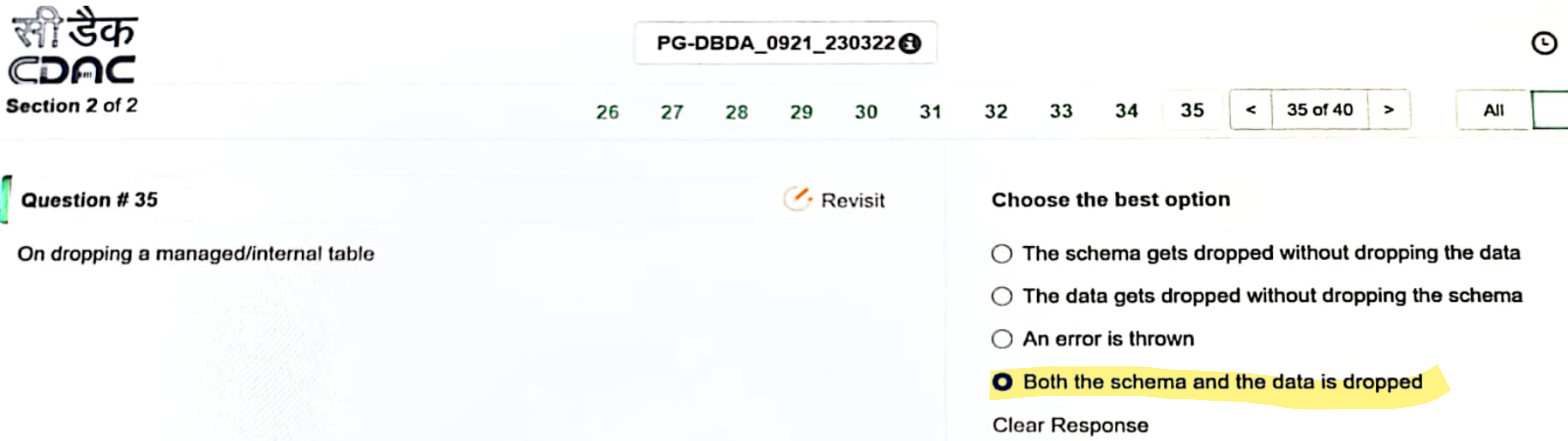
PG-DBDA_0921_230322

Question # 34

Apache HBase was modeled after Google's ____

Choose the best option

- ○ Foundation DB
- ○ Big top
- ○ Big Tables
- ○ None of the above



Section 2 of 2

Question # 35

On dropping a managed/internal table

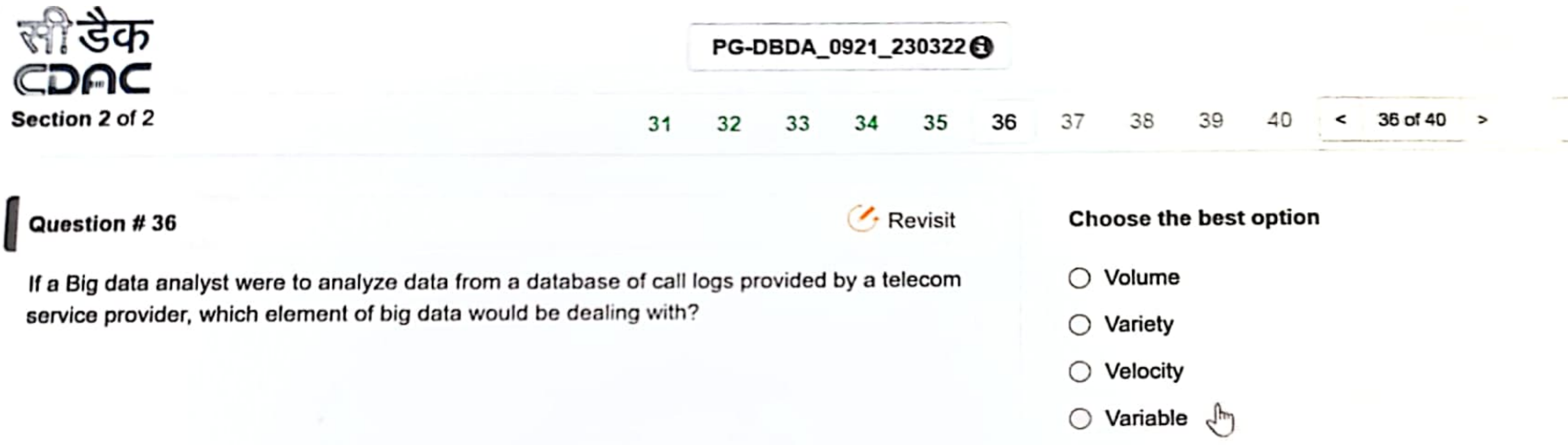
Choose the best option

○ The schema gets dropped without dropping the data

○ The data gets dropped without dropping the schema

○ An error is thrown

● Both the schema and the data is dropped



Section 2 of 2

PG-DBDA_0921_230322

Question # 36

If a Big data analyst were to analyze data from a database of call logs provided by a telecom service provider, which element of big data would be dealing with?

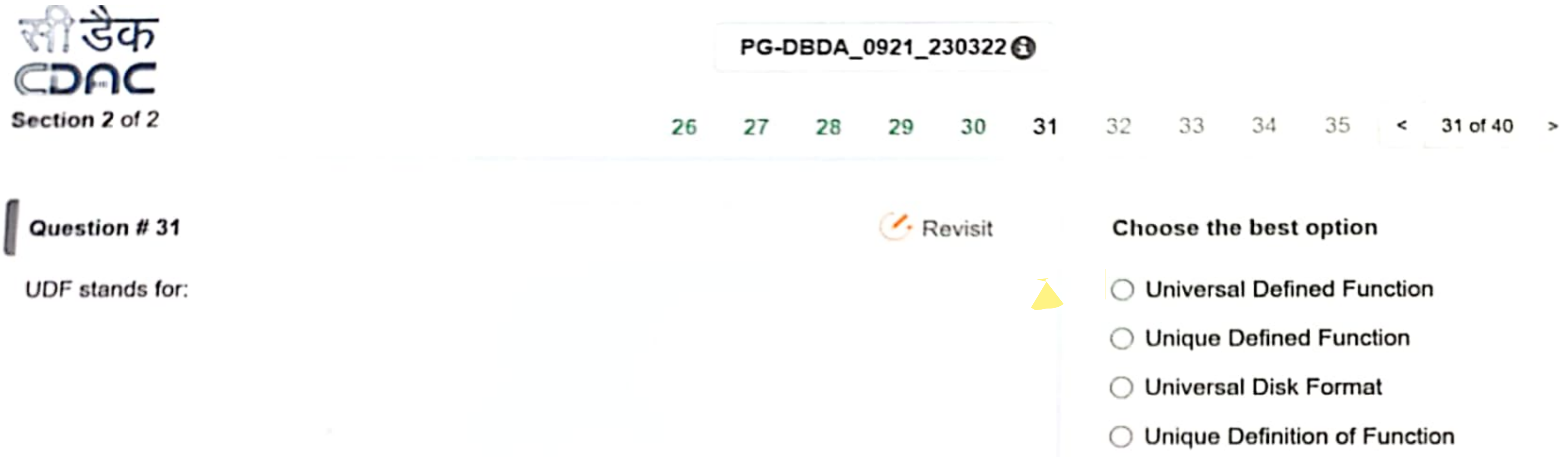
Choose the best option

○ Volume

○ Variety

○ Velocity

○ Variable



Section 2 of 2

PG-DBDA_0921_230322

Question # 31

UDF stands for:

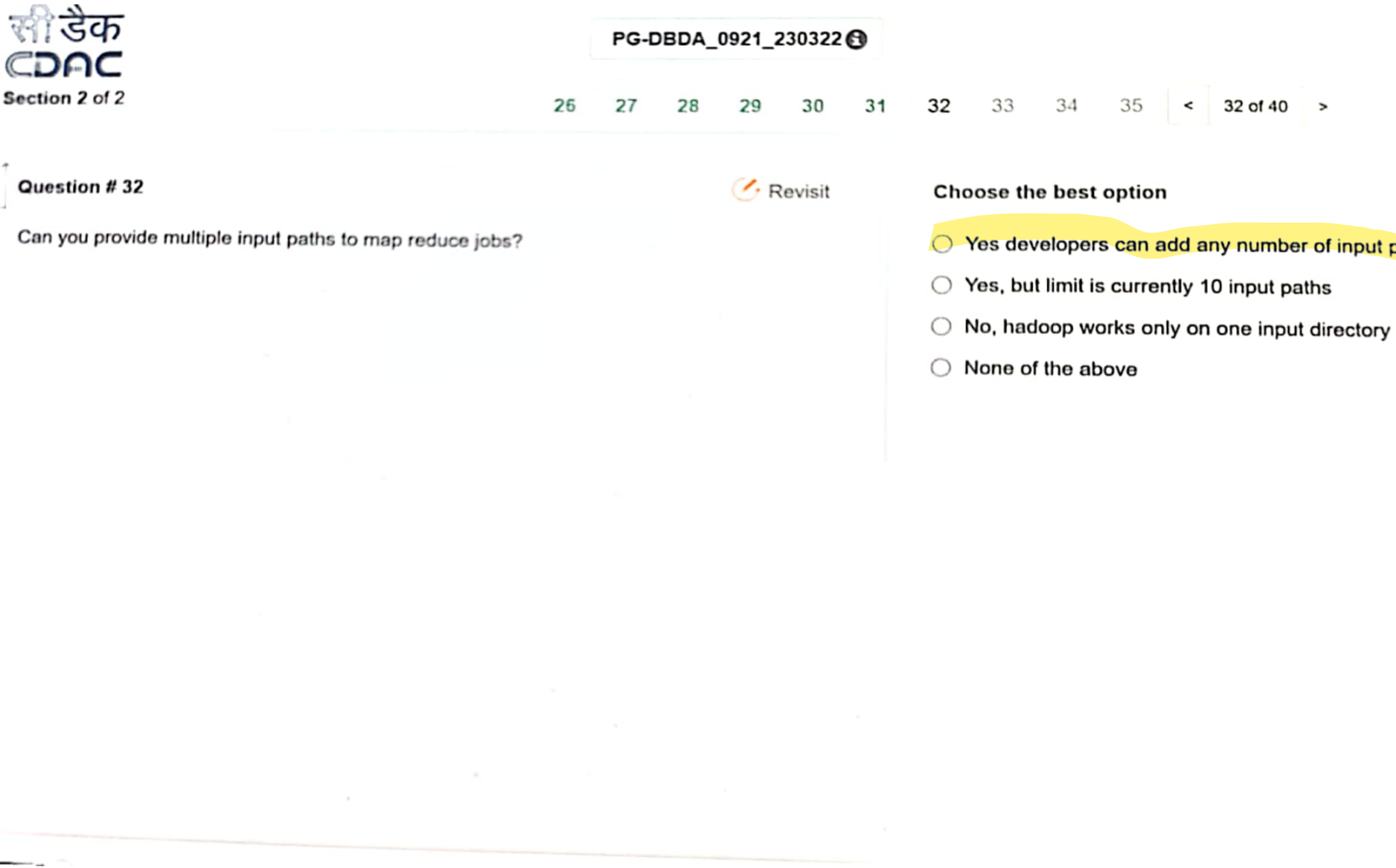
Choose the best option

○ Universal Defined Function

○ Unique Defined Function

○ Universal Disk Format

○ Unique Definition of Function



सी डैक
CDAC
Section 2 of 2

PG-DBDA_0921_230322

26 27 28 29 30 31 32 33 34 35 < 32 of 40 >

Question # 32

Revisit

Can you provide multiple input paths to map reduce jobs?

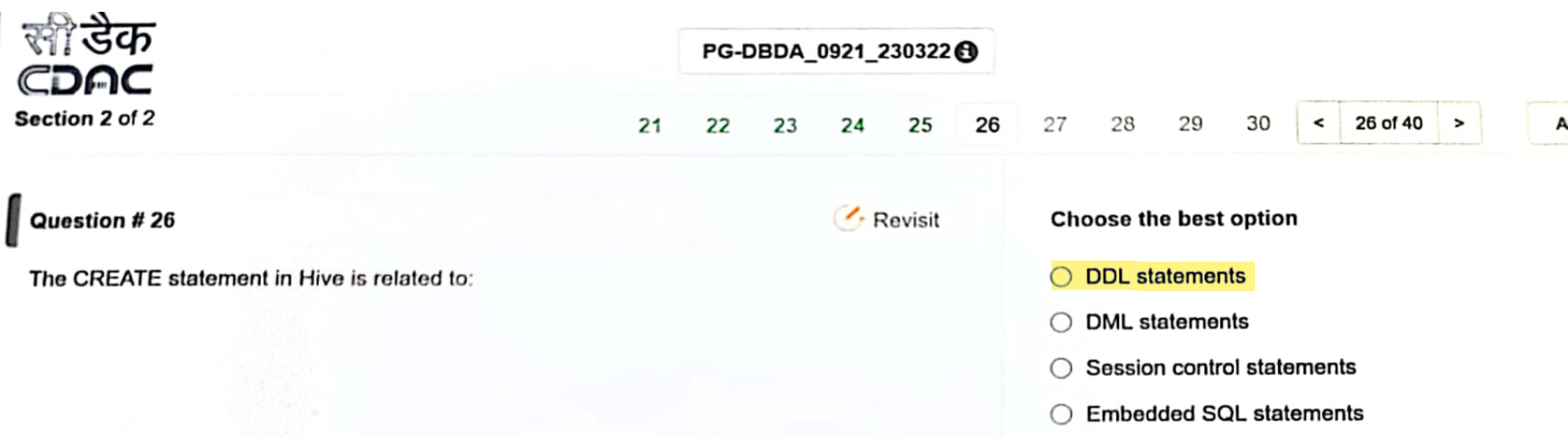
Choose the best option

○ Yes developers can add any number of input p

○ Yes, but limit is currently 10 input paths

○ No, hadoop works only on one input directory

○ None of the above



Section 2 of 2

PG-DBDA_0921_230322

Question # 26

The CREATE statement in Hive is related to:

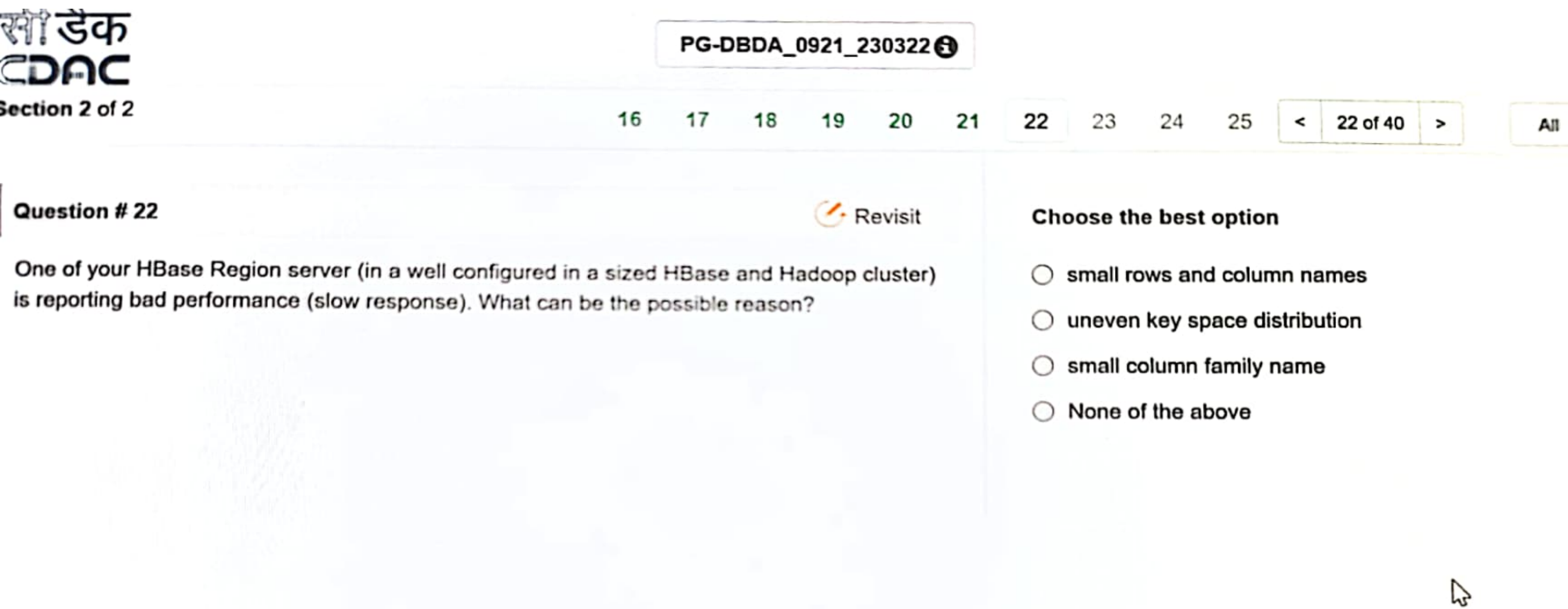
Choose the best option

○ DDL statements

○ DML statements

○ Session control statements

○ Embedded SQL statements



Question # 22

One of your HBase Region server (in a well configured in a sized HBase and Hadoop cluster) is reporting bad performance (slow response). What can be the possible reason?

Choose the best option

○ small rows and column names

○ uneven key space distribution

○ small column family name

○ None of the above

Section 2 of 2

PG-DBDA_0921_230322

Question # 25

Hadoop framework is written in:

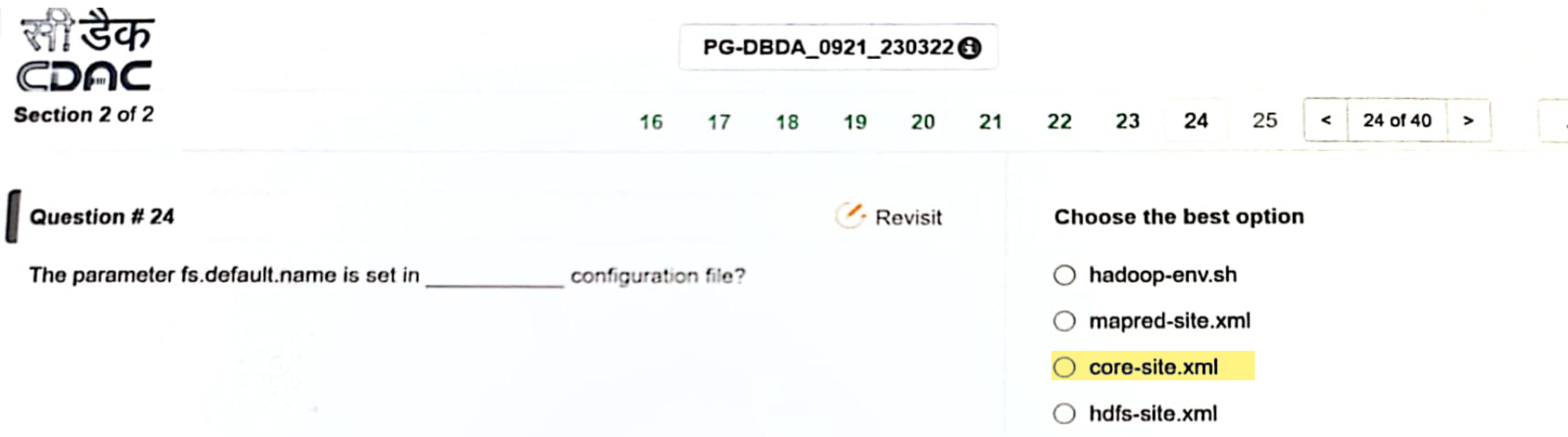
Choose the best option

○ Java

○ Python

○ Scala

○ C++



सी डैक
CDAC

Section 2 of 2

PG-DBDA_0921_230322

Question # 24

The parameter fs.default.name is set in ___________ configuration file?

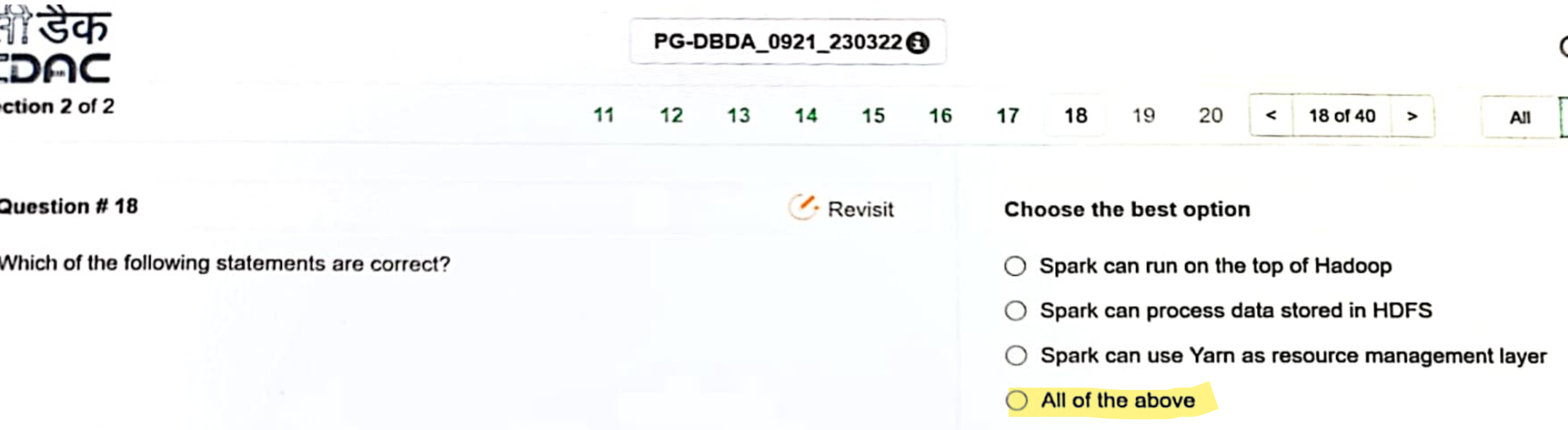
Choose the best option

○ hadoop-env.sh

○ mapred-site.xml

[✓ ○ core-site.xml]

○ hdfs-site.xml



Question # 18

Which of the following statements are correct?

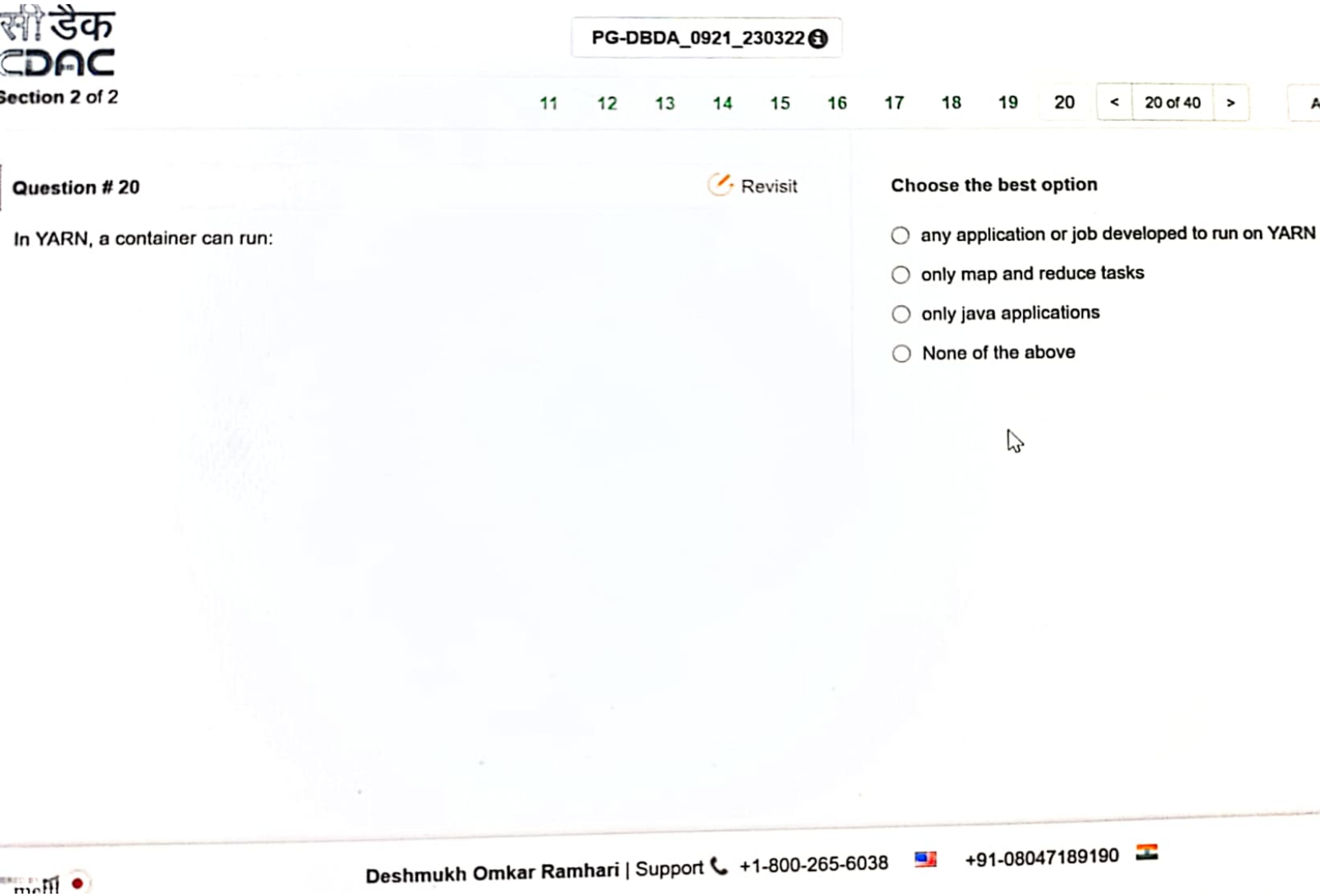
Choose the best option

○ Spark can run on the top of Hadoop

○ Spark can process data stored in HDFS

○ Spark can use Yarn as resource management layer

[✓ ○ All of the above]



Question # 20

In YARN, a container can run:

Choose the best option

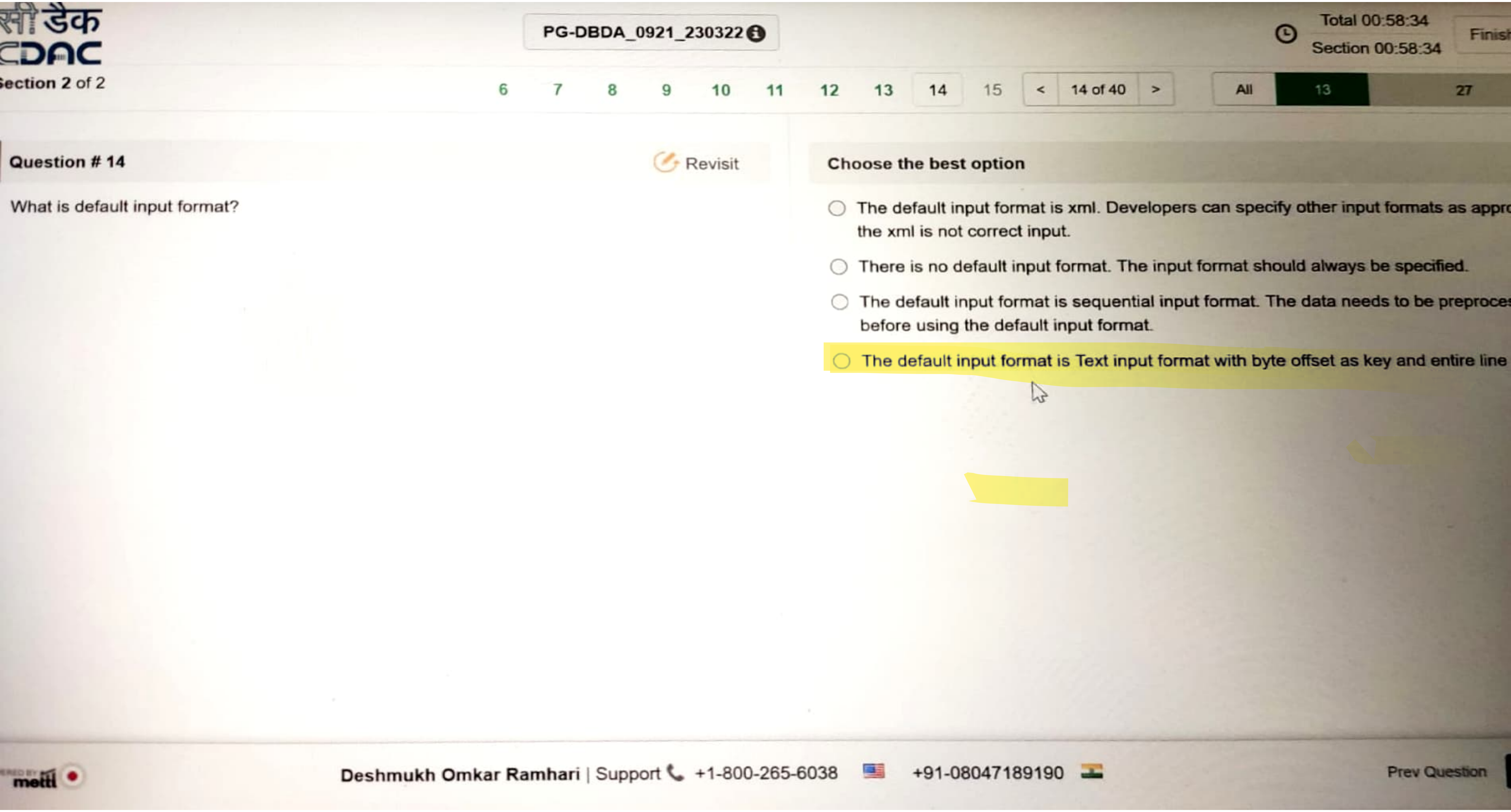
○ any application or job developed to run on YARN

○ only map and reduce tasks

○ only java applications

○ None of the above

Deshmukh Omkar Ramhari | Support ☎ +1-800-265-6038 +91-08047189190



Question # 14

What is default input format?

Choose the best option

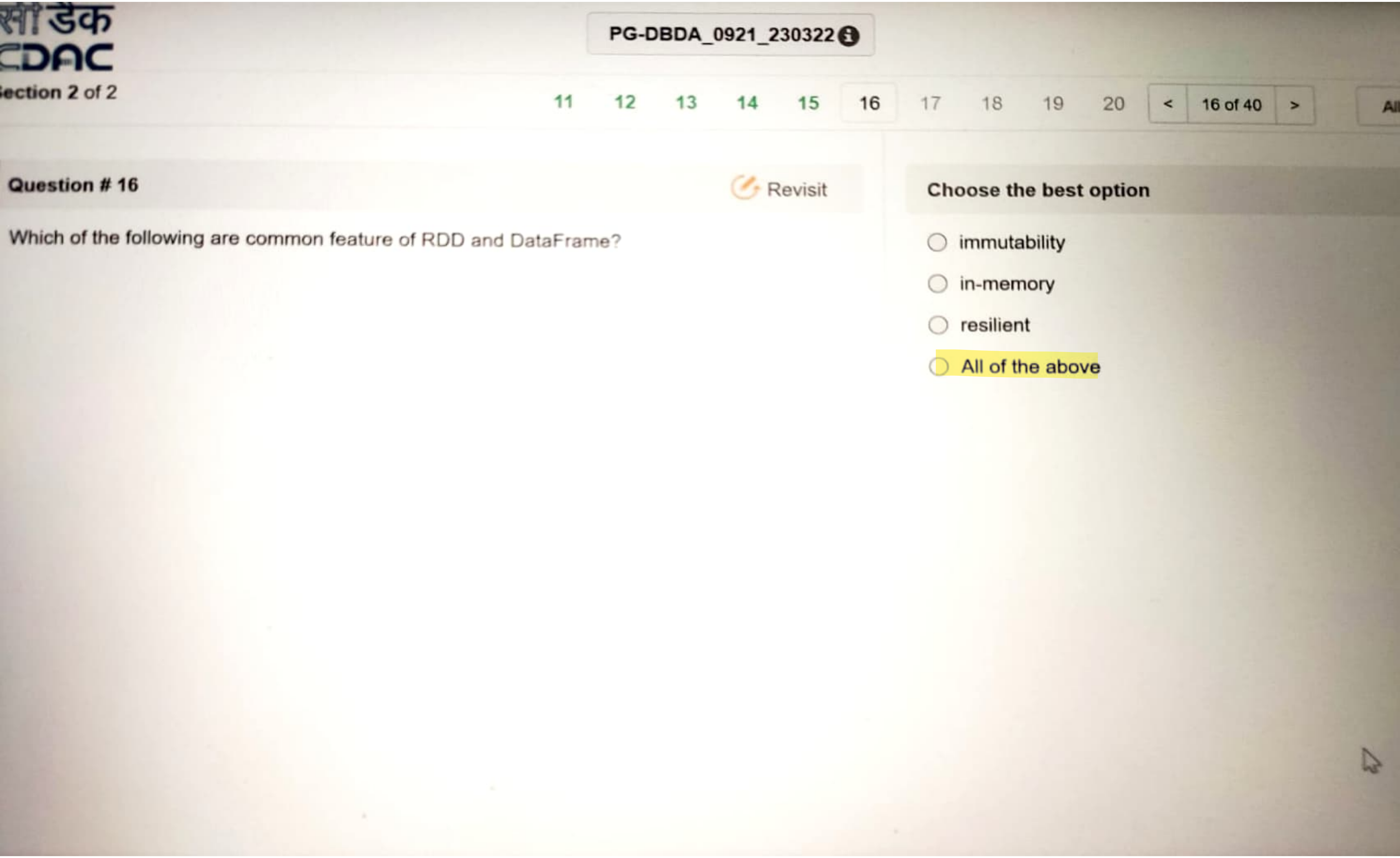
○ The default input format is xml. Developers can specify other input formats as appro
the xml is not correct input.

○ There is no default input format. The input format should always be specified.

○ The default input format is sequential input format. The data needs to be preproce
before using the default input format.

○ The default input format is Text input format with byte offset as key and entire line

Deshmukh Omkar Ramhari | Support ☎ +1-800-265-6038 +91-08047189190



Question # 16

Which of the following are common feature of RDD and DataFrame?

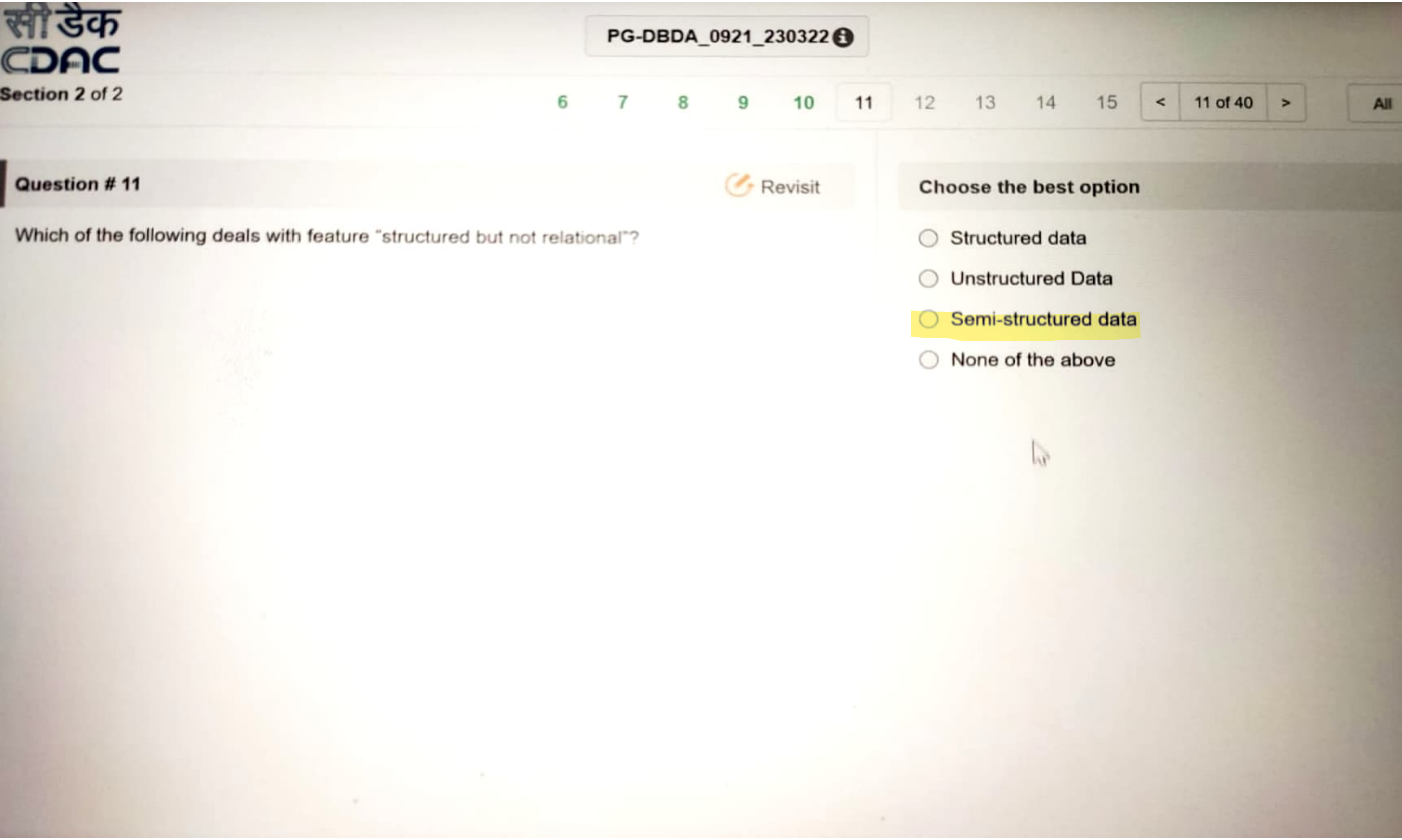
Choose the best option

○ immutability

○ in-memory

○ resilient

○ [✓ All of the above]



सी डैक
CDAC

Section 2 of 2

PG-DBDA_0921_230322

Question # 11

Which of the following deals with feature "structured but not relational"?

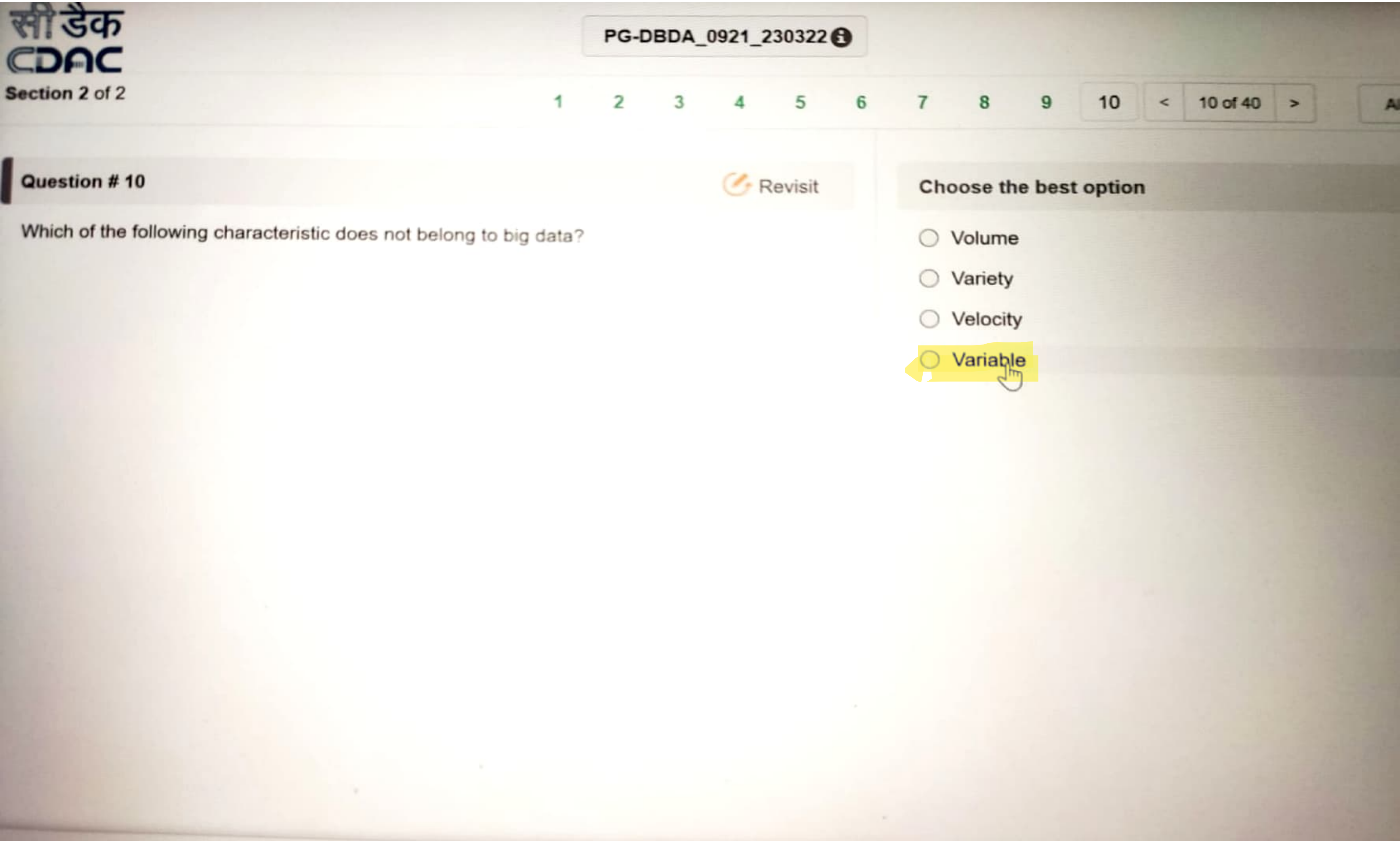
Choose the best option

○ Structured data

○ Unstructured Data

○ Semi-structured data

○ None of the above



Section 2 of 2

Question # 10

Which of the following characteristic does not belong to big data?

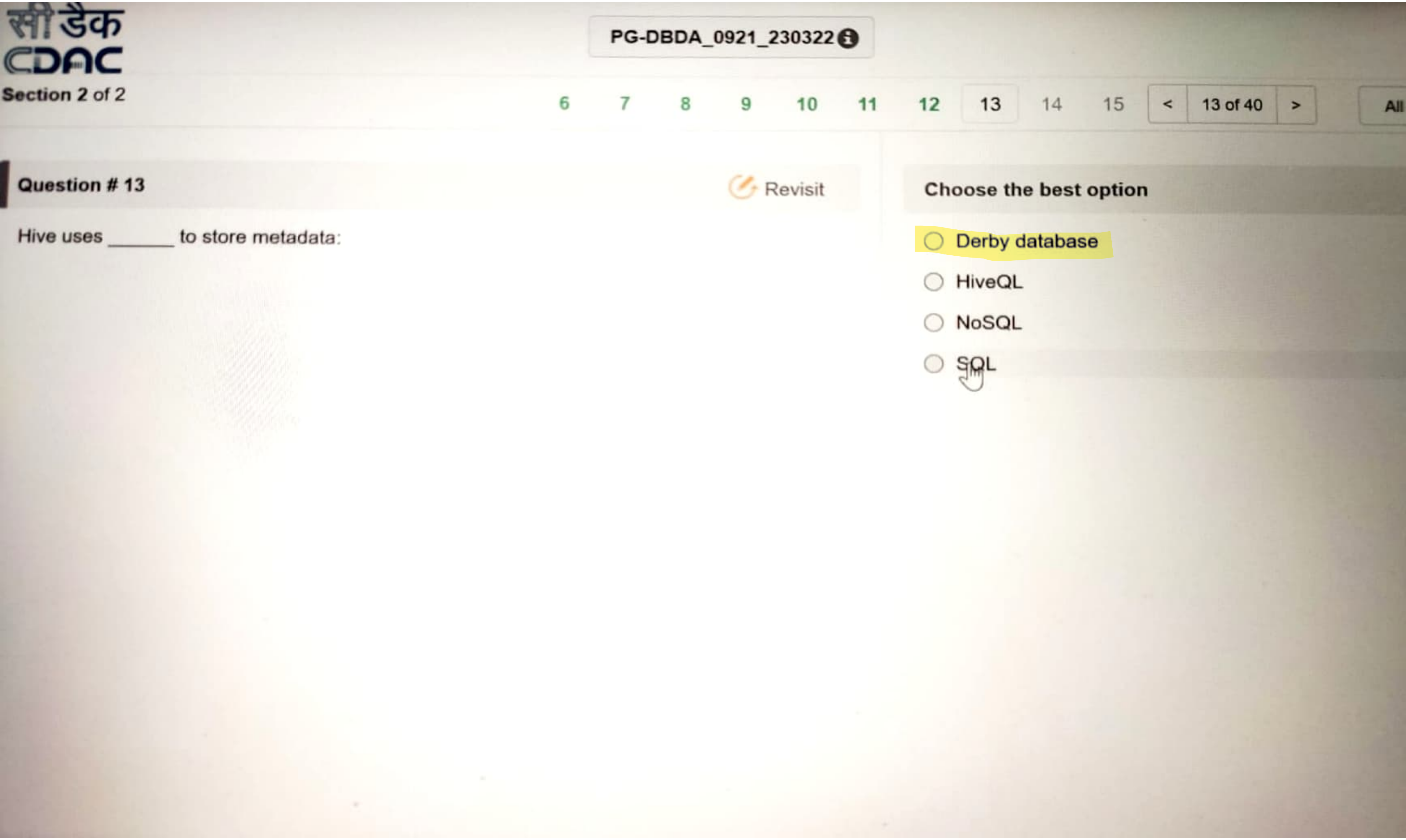
Choose the best option

○ Volume

○ Variety

○ Velocity

[✓ ○ Variable]



PG-DBDA_0921_230322

Section 2 of 2

Question # 13

Hive uses ______ to store metadata:

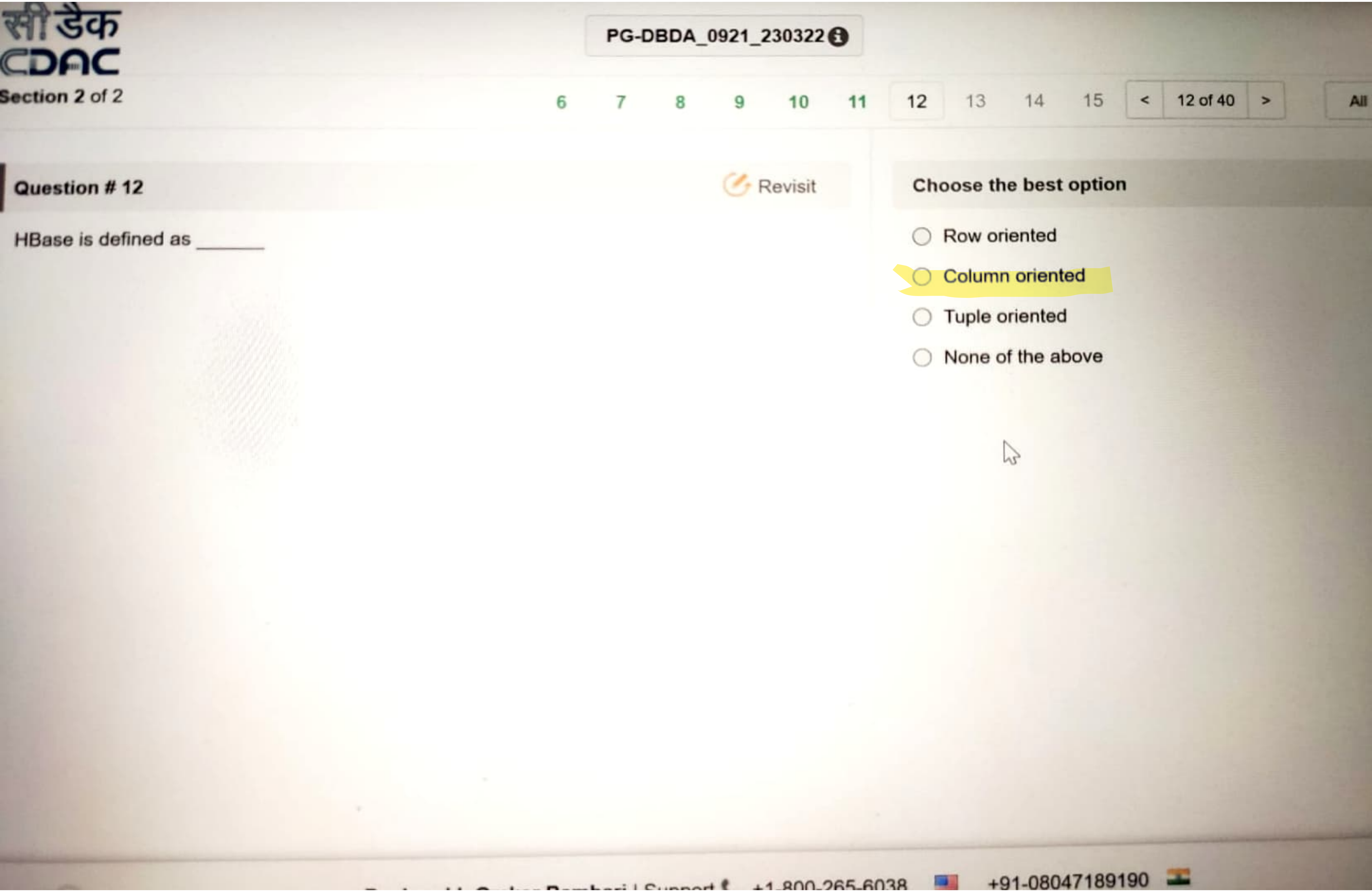
Choose the best option

○ Derby database

○ HiveQL

○ NoSQL

○ SQL



PG-DBDA_0921_230322

Section 2 of 2

Question # 12

HBase is defined as _______

Choose the best option

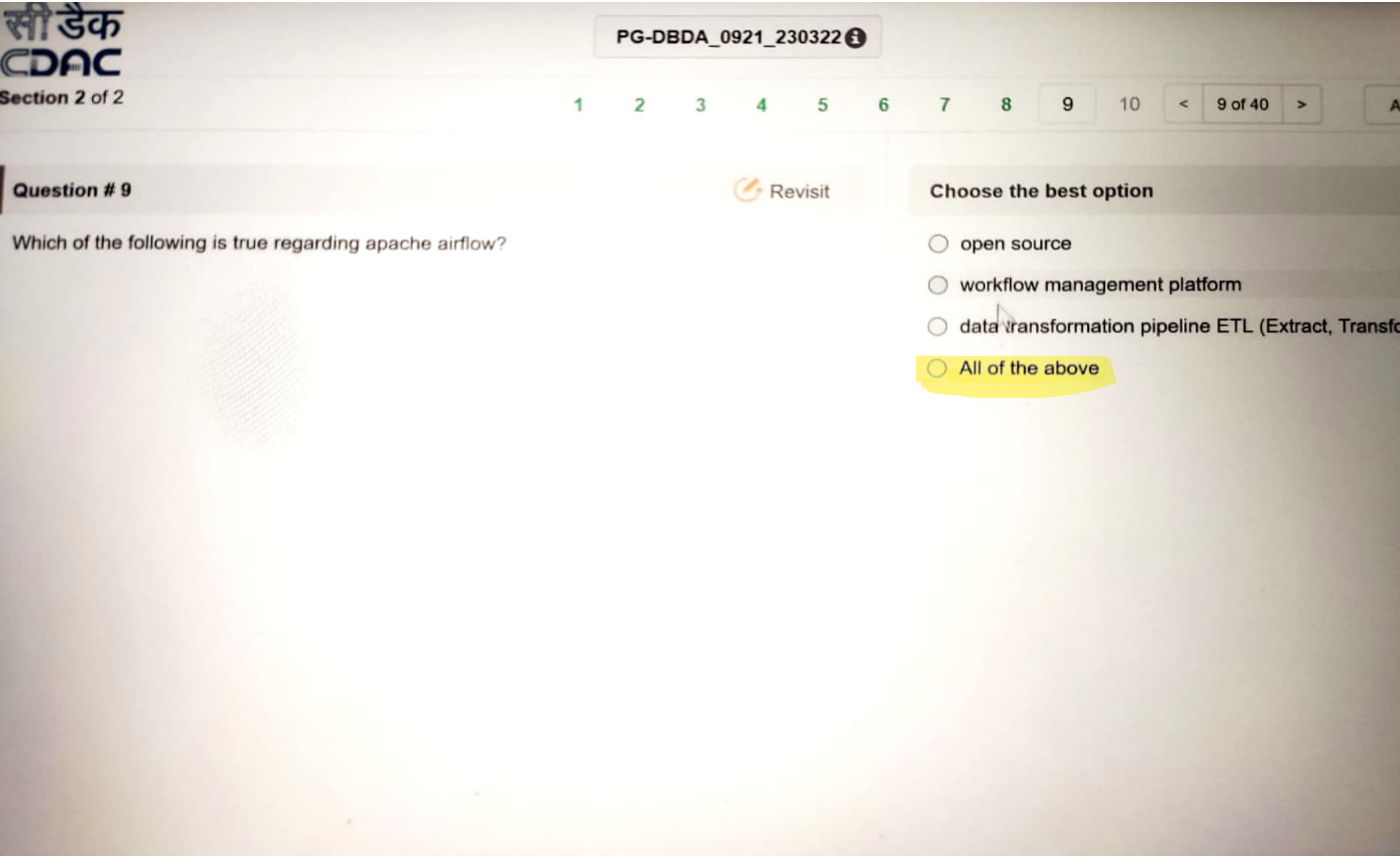
○ Row oriented

○ Column oriented

○ Tuple oriented

○ None of the above

| Support +1-800-265-6038 +91-08047189190



सी डैक
CDAC

PG-DBDA_0921_230322

Section 2 of 2

Question # 9

Which of the following is true regarding apache airflow?

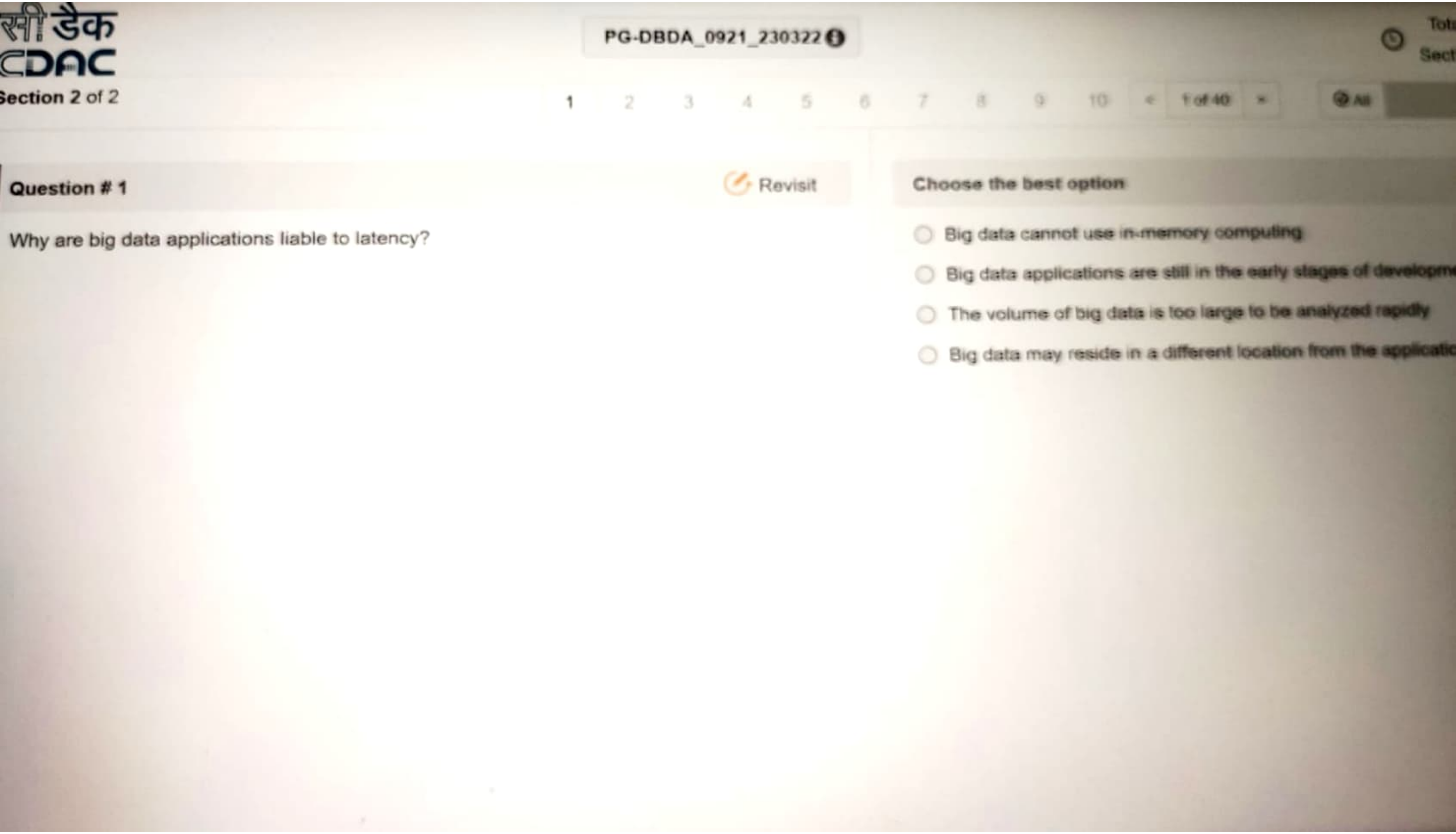
Choose the best option

○ open source

○ workflow management platform

○ data transformation pipeline ETL (Extract, Transfo

[✓ ○ All of the above]



PG-DBDA_0921_230322

Section 2 of 2

Question # 1

Why are big data applications liable to latency?

Choose the best option

○ Big data cannot use in-memory computing

○ Big data applications are still in the early stages of developme

○ The volume of big data is too large to be analyzed rapidly

○ Big data may reside in a different location from the applicatio