

Sample Questions

The following sample questions are not inclusive and do not necessarily represent all of the types of questions that comprise the exams. The questions are not designed to assess an individual's readiness to take a certification exam.

Data Curation for SAS Data Scientists

Question 1

A user is looking at the Basic Properties pan, as shown below.

Basic Properties	
Name	Value
Name	Customer
Description	
Folder Location	/Data Mart Development/Source Data
Checked Out By	
Table Name	CUSTOMER
Library	Source Library (diftodet)
DBMS	SAS Table
Number of Rows	Row count is disabled
Number of Columns	12

The Number of Rows attribute shows Row count is disabled. Why is the row count disabled?

- A. The "Enable row count on basic properties and data viewer for tables" global option is not set.
- B. No physical dataset exists for the Customer table metadata object.
- C. The table's "Enable row count on basic properties and data viewer for tables" option is not set.
- D. The user does not have read permissions on the folder where the physical dataset is stored.

correct answer= "A"

Question 2

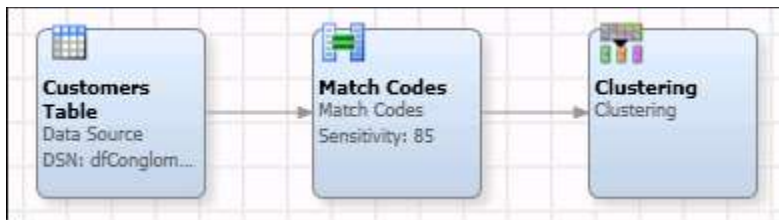
Which is NOT defined as part of the New Library Wizard?

- A. the type of library
- B. the metadata name for the library
- C. the location of the library
- D. the operating system where the library is assigned

correct answer= "D"

Question 3

Consider the partial data job flow shown below:



Match codes fields were generated based on these fields:

NAME
ADDRESS
CITY
STATE

The Clustering node is "over matching". Therefore, it is finding matches where there should NOT be matches.

What can you do in the Clustering node to prevent "over matching"?

- A. Specify a different Cluster ID field.
- B. Specify "Single row clusters only".
- C. Add more conditions perhaps involving more of the match code fields.
- D. Specify "Sort output by cluster number and remove duplicates".

correct answer= "C"

Question 4

You use the PROC DQSRVSVC in a SAS program to pass data to a service that was written in DataFlux Data Management Studio. Where does the service execute and process the data?

- A. DataFlux Data Management Studio
- B. DataFlux Data Management Server
- C. SAS Application Server
- D. SAS Middle Tier Server

correct answer= "B"

Question 5

In Hadoop, what is the purpose of the YARN component?

- A. HDFS file metadata management
- B. data distribution and replication
- C. job scheduling and resource management
- D. HDFS file permissions and security

correct answer= "C"

Question 6

Consider the following three data sets and the Pig program statement. What is the expected behavior?

```
A = { a1:chararray, a2:int}
B = { b1:chararray, b2:int, b3:float}
C = { c1:chararray, c2:int, c3:float, c4:double}
```

Pig Program Statement:

```
D = JOIN A BY (a1, a2),
      B BY (b1, b2, b3),
      C BY (c1, c2, c3);?
```

- A. statement runs successfully
- B. statement gives a compile error JOIN can happen only on two data sets at a time
- C. statement gives a compile error that the JOIN columns have to equal number for all three data sets
- D. statement gives a compile error that the keyword INNER or LEFT OUTER or RIGHT OUTER or FULL is missing

correct answer= "C"

Question 7

After running the following SAS program, what is the format of the HDFS file underlying the Hive table **my.cars**?

```
libname myhdp hadoop server='mysrvr' user='me' password='mine'  
schema='my';  
data my.cars;  
    set sashelp.cars;  
run;
```

- A. Parquet
- B. Text
- C. Sequence
- D. Avro

correct answer= "B"

Question 8

Which directive allows you to write custom DS2 code?

- A. Run a SAS program.
- B. Run a Hive program.
- C. Query or Join Data in Hadoop.
- D. Profile Data.

correct answer= "A"

Question 9

Which statement about writing explicit SQL pass-through code is true?

- A. You can use any SAS programming method, naming DBMS tables as input.
- B. All the code you write will execute in the DBMS instead of in SAS.
- C. You must be able to write native DBMS SQL code.
- D. Automatically generated DBMS SQL might unintentionally cause all processing to happen in SAS instead of the DBMS.

correct answer= "C"

Question 10

SAS data quality functions in SAS Federation Server are implemented as what type of object?

- A. FedSQL views
- B. DS2 methods
- C. QKB definitions
- D. BASE data services

correct answer= "B"