

# 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.*

## SAS Big Data Programming and Loading

### Question 1

In the example below, the input data set is a Hive table accessed using a SAS/ACCESS to Hadoop LIBNAME statement.

```
proc freq data=hivelib.myhivetable;  
  class month;  
  by year;  
run;
```

Which statement is true about this program?

- A. The procedure will fail unless the table HIVELIB.MYHIVETABLE is already stored ordered by YEAR.
- B. SAS will generate a HiveQL query to return the data to SAS ordered by YEAR so that the procedure receives the data ordered by YEAR as required.
- C. BY statements do not require the data be received by the procedure in any specific order.
- D. BY statements are not supported for Hive tables because it is not possible to order data that is distributed on different nodes of the Hadoop cluster.

**correct\_answer = "B"**

### Question 2

When working with data stored in Hadoop, which SAS function is NOT passed to Hive by default?

- A. DATE
- B. HOUR
- C. VAR
- D. UPCASE

**correct\_answer = "A"**

### Question 3

What is an advantage of using a LIBNAME statement to interact with your Hadoop cluster?

- A. It enables you to submit user-written HiveQL code to Hive.
- B. The GENERATE\_PIG\_CODE= option enables you to bypass Hive and generate Pig Latin code.
- C. It enables some SAS procedures to push processing into Hive.
- D. It ensures that Hive will handle all processing.

**correct\_answer = "C"**

#### Question 4

Refer to the log message shown below:

```
58         proc ds2;
59         data test;
60         dcl double date;
61         method run();
62         set work.one;
63         mm=2;
64         end;
65         enddata;
66         run;
ERROR: Compilation error.
ERROR: Parse encountered type when expecting identifier.
ERROR: Parse failed on line 60: dcl double >>> date <<< ;
NOTE: PROC DS2 has set option NOEXEC and will continue to prepare statements.
67         quit;
```

Which of the following changes will fix the errors shown in the log?

- A. Replace line 60 with `dcl double 'date' ;`
- B. Replace line 60 with `dcl double "date" ;`
- C. Replace line 60 with `dcl string date ;`
- D. Replace line 60 with `dcl double 'date'n ;`

**correct\_answer = "B"**

#### Question 5

Which statement creates a temporary array within DS2?

- A. `vararray double a[2];`
- B. `vararray double a(2);`
- C. `array a(2) s1-s2;`
- D. `dcl double a[2];`

**correct\_answer = "D"**

#### Question 6

The following SAS program is submitted:

```
proc ds2;
data work.sasorders;
  dcl timestamp order_timestamp;
  dcl double order_datetime;
  method run();
  set orders;
  order_timestamp = to_timestamp(order_datetime);
  end;
enddata;
run;
quit;
```

What happens when the program is executed?

- A. -- The variable `order_timestamp` is created and processed as an ANSI timestamp value in the DS2 program.  
-- The `order_timestamp` value is converted to a SAS datetime when it is written to the output SAS data set.
- B. -- The variable `order_timestamp` is created and processed as an ANSI timestamp value in the DS2 program.  
-- The output data set stores the values as a SAS timestamp value.
- C. -- The variable `order_timestamp` is converted to a SAS time value.  
-- The output data set stores this as the number of seconds since midnight.
- D. -- The program does not execute because `order_datetime` is a SAS datetime value.

**correct\_answer = "A"**

### Question 7

Which operator is NOT a diagnostic operator for testing a Pig program?

- A. DUMP
- B. ILLUSTRATE
- C. EXPLAIN
- D. SPLIT

**correct\_answer = "D"**

### Question 8

Web server logs are written in an HDFS directory. The following lines indicate the format and an example of the comma-separated values for one line in the log file.

```
# IP address, timestamp, request, status, size  
192.168.12.41,24/Nov/2015:10:09:58 -0500, "GET /services/config.xml HTTP/
```

Which CREATE TABLE statement enables a Hive query to access each of the fields?

- A. `create external table weblogs (ip string, dt string, req string, status int, sz string) row format delimited fields terminated by ',' location '/data/weblogs';`
- B. `create external table weblogs (ip string, dt string, req string, status int, sz string) fields terminated by ',' location '/data/weblogs';`
- C. `create external table weblogs (ip string, rest string) row format delimited fields terminated by ',' location '/data/weblogs';`
- D. `create external table weblogs (ip string, dt string, req string, status int, sz string) fields delimited fields by ',' location '/data/weblogs';`

**correct\_answer = "A"**

### Question 9

Many temporary tables may be created in the LASR server by PROC IMSTAT analysis actions. What happens to temporary tables when a PROC IMSTAT session is terminated?

- A. All temporary tables are saved to the SAS server WORK library.
- B. All temporary tables are purged from the LASR server.
- C. The last temporary table created is saved to the SAS WORK library.
- D. The last temporary table created is saved to storage in the HDFS.

**correct\_answer = "B"**

### Question 10

This question will ask you to provide a line of missing code.

Which line of code would you insert to get the mean and standard deviation of INCOME and AGE, calculated separately for GENDER variable values F (female) and M (male)?

```
proc imstat;  
  table mylasr.bankdata(tag="&TagString");  
  <insert code here>  
run;
```

- A. `summary income age / groupby=gender;`
- B. `crosstab income*gender age*gender;`
- C. `summary income age / by=gender;`
- D. `univariate income age / by=gender;`

**correct\_answer = "A"**