NowExam

Higher Quality , Better Service !

Q&A

Exam : **C2090-101**

Title : IBM Big Data Engineer

Version : DEMO

1.Which statement is TRUE concerning optimizing the load performance?

A. You can improve the performance by increasing the number of map tasks assigned to the load

B. When loading large files the number of files that you load does not impact the performance of the LOAD HADOOP statement

C. You can improve the performance by decreasing the number of map tasks that are assigned to the load and adjusting the heap size

D. It is advantageous to run the LOAD HADOOP statement directly pointing to large files located in the host file system as opposed to copying the files to the DFS prior to load

Answer: A

Explanation:

Reference:

https://www.ibm.com/support/knowledgecenter/en/SSCRJT_5.0.3/com.ibm.swg.im.bigsql.doc/doc/bigsql_loadperf.html

2.Which of the following statements are TRUE regarding the use of Data Click to load data into BigInsights? (Choose two.)

A. Big SQL cannot be used to access the data moved in by Data Click because the data is in Hive

B. You must import metadata for all sources and targets that you want to make available for Data Click activities

C. Connections from the relational database source to HDFS are discovered automatically from within Data Click

D. Hive tables are automatically created every time you run an activity that moves data from a relational database into HDFS

E. HBase tables are automatically created every time you ran an activity that moves data from a relational database into HDFS

Answer: CE

Explanation:

Reference: https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.3.0/com.ibm.swg.im.iis.dataclick.doc/topics/hivetables.html

3.Which of the following statements regarding importing streaming data from InfoSphere Streams into Hadoop is TRUE?

A. InfoSphere Streams can both read from and write data to HDFS

B. The Streams Big Data toolkit operators that interface with HDFS uses Apache Flume to integrate with Hadoop

C. Streams applications never need to be concerned with making the data schemas consistent with those on Hadoop

D. Big SQL can be used to preprocess the data as it flows through InfoSphere Streams before the data lands in HDFS

Answer: D

4.Which of the following is TRUE about storing an Apache Spark object in serialized form?

A. It is advised to use Java serialization over Kryo serialization

B. Storing the object in serialized from will lead to faster access times

C. Storing the object in serialized from will lead to slower access times

D. All of the above

Answer: B

Explanation:

Reference: https://spark.apache.org/docs/latest/rdd-programming-guide.html

5.Which ONE of the following statements regarding Sqoop is TRUE?

A. HBase is not supported as an import target

B. Data imported using Sqoop is always written to a single Hive partition

C. Sqoop can be used to retrieve rows newer than some previously imported set of rows

D. Sqoop can only append new rows to a database table when exporting back to a database

Answer: C

Explanation:

Reference: https://sqoop.apache.org/docs/1.4.1-incubating/SqoopUserGuide.html