

Version: 8.0

Question: 1

You need to place the results of a PigLatin script into an HDFS output directory. What is the correct syntax in Apache Pig?

- A. update hdfs set D as './output';
- B. store D into './output';
- C. place D into './output';
- D. write D as './output';
- E. hdfsstore D into './output';

Answer: B

Use the STORE operator to run (execute) Pig Latin statements and save (persist) results to the file system. Use STORE for production scripts and batch mode processing.

Syntax: STORE alias INTO 'directory' [USING function];

Example: In this example data is stored using PigStorage and the asterisk character (*) as the field delimiter.

```
A = LOAD 'data' AS (a1:int,a2:int,a3:int);
```

```
DUMP A;
```

```
(1,2,3)
```

```
(4,2,1)
```

```
(8,3,4)
```

```
(4,3,3)
```

```
(7,2,5)
```

```
(8,4,3)
```

```
STORE A INTO 'myoutput' USING PigStorage (*);
```

```
CAT myoutput;
```

```
1*2*3
```

```
4*2*1
```

```
8*3*4
```

```
4*3*3
```

```
7*2*5
```

```
8*4*3
```

References: <https://pig.apache.org/docs/r0.13.0/basic.html#store>

Question: 2

How is Oracle Loader for Hadoop (OLH) better than Apache Sqoop?

- A. OLH performs a great deal of preprocessing of the data on Hadoop before loading it into the

database.

B. OLH performs a great deal of preprocessing of the data on the Oracle database before loading it into NoSQL.

C. OLH does not use MapReduce to process any of the data, thereby increasing performance.

D. OLH performs a great deal of preprocessing of the data on the Oracle database before loading it into Hadoop.

E. OLH is fully supported on the Big Data Appliance. Apache Sqoop is not supported on the Big Data Appliance.

Answer: A

Oracle Loader for Hadoop provides an efficient and high-performance loader for fast movement of data from a Hadoop cluster into a table in an Oracle database. Oracle Loader for Hadoop repartitions the data if necessary and transforms it into a database-ready format. It optionally sorts records by primary key or user-defined columns before loading the data or creating output files.

Note: Apache Sqoop(TM) is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases.

Question: 3

Which three pieces of hardware are present on each node of the Big Data Appliance? (Choose three.)

A. high capacity SAS disks

B. memory

C. redundant Power Delivery Units

D. InfiniBand ports

E. InfiniBand leaf switches

Answer: A,B,D

Big Data Appliance Hardware Specification and Details, example:

Per Node:

References: <http://www.oracle.com/technetwork/server-storage/engineered-systems/bigdata-appliance/overview/bigdataappliancev2-datasheet-1871638.pdf>

Question: 4

What two actions do the following commands perform in the Oracle R Advanced Analytics for Hadoop Connector? (Choose two.)

`ore.connect (type="HIVE")`

`ore.attach ()`

A. Connect to Hive.

B. Attach the Hadoop libraries to R.

C. Attach the current environment to the search path of R.

D. Connect to NoSQL via Hive.

Answer: A,C

You can connect to Hive and manage objects using R functions that have an ore prefix, such as ore.connect.

To attach the current environment into search path of R use:

ore.attach()

References: https://docs.oracle.com/cd/E49465_01/doc.23/e49333/orch.htm#BDCUG400

Question: 5

Your customer's security team needs to understand how the Oracle Loader for Hadoop Connector writes data to the Oracle database.

Which service performs the actual writing?

- A. OLH agent
- B. reduce tasks
- C. write tasks
- D. map tasks
- E. NameNode

Answer: B

Oracle Loader for Hadoop has online and offline load options. In the online load option, the data is both preprocessed and loaded into the database as part of the Oracle Loader for Hadoop job. Each reduce task makes a connection to Oracle Database, loading into the database in parallel. The database has to be available during the execution of Oracle Loader for Hadoop.

References: <http://www.oracle.com/technetwork/bdc/hadoop-loader/connectors-hdfs-wp-1674035.pdf>

Question: 6

Your customer needs to manage configuration information on the Big Data Appliance.

Which service would you choose?

- A. SparkPlug
- B. ApacheManager
- C. Zookeeper
- D. Hive Server
- E. JobMonitor

Answer: C

The ZooKeeper utility provides configuration and state management and distributed coordination services to Dgraph nodes of the Big Data Discovery cluster. It ensures high availability of the query processing by the Dgraph nodes in the cluster.

References:

https://docs.oracle.com/cd/E57471_01/bigData.100/admin_bdd/src/cadm_cluster_zookeeper.html

Question: 7

You are helping your customer troubleshoot the use of the Oracle Loader for Hadoop Connector in online mode. You have performed steps 1, 2, 4, and 5.

STEP 1: Connect to the Oracle database and create a target table.

STEP 2: Log in to the Hadoop cluster (or client).

STEP 3: Missing step

STEP 4: Create a shell script to run the OLH job.

STEP 5: Run the OLH job.

What step is missing between step 2 and step 4?

- A. Diagnose the job failure and correct the error.
- B. Copy the table metadata to the Hadoop system.
- C. Create an XML configuration file.
- D. Query the table to check the data.
- E. Create an OLH metadata file.

Answer: C

Question: 8

The `hdfs_stream` script is used by the Oracle SQL Connector for HDFS to perform a specific task to access data.

What is the purpose of this script?

- A. It is the preprocessor script for the Impala table.
- B. It is the preprocessor script for the HDFS external table.
- C. It is the streaming script that creates a database directory.
- D. It is the preprocessor script for the Oracle partitioned table.
- E. It defines the jar file that points to the directory where Hive is installed.

Answer: B

The `hdfs_stream` script is the preprocessor for the Oracle Database external table created by Oracle SQL Connector for HDFS.

References: https://docs.oracle.com/cd/E37231_01/doc.20/e36961/start.htm#BDCUG107