Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)
Course-3: Managing Big Data in Cluster and Cloud Storage
Week-4: Core

Question-1: Which of these are ways to load data into HDFS using Hue on the VM? Check all that apply.
Answer-1: Load data into any HDFS directory through data source assist panel on left side
- Use the File Browser to load data into any HDFS directory
- Use the Table Browser to load data into the storage directory of an existing table.
- Use the Table Browser to load data into the storage directory of a new table at a time when you create the table

Question-2: Which of these tasks can you complete using HDFS shell commands? Check all that apply.
Answer-2: Load data into the storage directory of an existing table
- Load data for a storage directory for a table that hasn't been created yet

Question-3: Which hdfs dfs command option can be used to upload a file from your local filesystem to HDFS?
Answer-3: -put

Question-4: On the course VM, which command could you use to upload the local file /home/ training/training_materials/analyst/data/games.csv to an S3 bucket named bucket-o-games? For purposes of this question, assume you have write access to this bucket.
Answer-4: aws s3 cp /home/training/training_materials/analyst/data/games.csv s3:// bucket-o-games

Question-5: What is the effect of including TBLPROPERTIES ('serialization.null.format' = '') in a CREATE TABLE statement?
Answer-5: Empty strings in the table's data files will appear as NULL in query results

Question-6: This command uses Sqoop to import data from MySQL to HDFS on the VM (with the user training).
$ sqoop import --connect jdbc:mysql://localhost/mydb \
--username training --password training \
--table example_table \
--target-dir /user/hive/warehouse/example_table
Where in HDFS will the data be found?
Answer-6: /user/hive/warehouse/example_table

Question-7: Which are valid uses for Sqoop? Check all that apply.
Answer-7: Exporting data from HDFS to an RDBMS (such as MySQL)
- Importing data to HDFS from an RDBMS (such as MySQL)

Question-8: Which are advantages of using the LOAD DATA SQL statement instead of the hdfs dfs -mv shell command to load data? Check all that apply.
Answer-8: LOAD DATA renames any files that are the same as files that already exist within the directory.
- If you use Impala to execute the LOAD DATA command, the metadata cache is automatically updated

Question-9: Why is it a poor practice to use INSERT commands to load data into a table a few rows at a time?
Answer-9: You get many small files, which can slow big data systems down

Question-10: Which are valid ways to create a copy of an existing table? Check all that apply.
Answer-10: CREATE TABLE newtable LIKE oldtable; then * INSERT INTO newtable SELECT * FROM OLDTABLE;
CREATE TABLE newtable AS SELECT * FROM oldtable;