Caleb Post
Lab 7 Questions

Question 1: In the previous lab, you created a table call logData that had the results of the analysis from Pig. Please specify the following:

a). create table command in mysql that can be used as the target table for exporting the contents of the logData table.

CREATE TABLE logs(blog varchar(25), hitRatio float, errorRatio float, year int(4), day int(2), hour int(2));

b). Sqoop command to export the contents of the logData table to the above mysql table.

sqoop export --connect <jdbc connection> --username <user> --table logs --export-dir /apps/hive/warehouse/<db>.db/table

Question 2: Sqoop exports can fail. List a few reasons for sqoop export failure. Can sqoop export failure result in a partial amount of data being exported to mysql?

An export may fail if there is a loss of connectivity with the hadoop cluster and database, attempting to insert a row violating a consistency constraint, attempt to parse an incomplete or malformed record from the HDFS source data, attempting to parse records using incorrect delimiters, or if there are capacity issues.

Yes it can result in a partial amount of data being exported. Each sqoop export map task is a separate transaction.

Question 3: What is the role of a staging table in a sqoop Export?

It is used to move the data in the sqoop job to a temporary table, and then moves the data in one transaction at the end. This prevents the issue in question 2 where we have a partial export.