

Question 1:

- a. `CREATE TABLE logData (blogName string, hitRatio float, errorRatio float, year int, month int, day int, hour int);`
- b. `sqoop export --connect jdbc:mysql://<mysql location>/<database that contains logData table e.g. wukusil> --username root --password hadoop --table logData --export-dir <location of external table, e.g. /tmp/wukusil/logData> --input-fields-terminated-by '\t'`

Question 2:

Sqoop could come across malformed rows or a JVM could run out of memory or a connection problem could occur.

If sqoop is not using a staging table a failed export can result in partial data transfer.

Question 3:

A staging table can be used as an intermediate step when moving data. The data is exported to the table and then copied from the table to the final destination. This keeps incomplete data from being written to a table if a job fails. It also prevents any jobs that are following the sqoop job from trying to act on an incomplete data set.