Zhihao Xue
CM 3357
Lab7
CSSE490 Dr. Mohan
Thur,Oct 28, 2015

- create table command in mysql that can be used as the target table for exporting

```
CREATE TABLE LogDataExport(
blogName varchar(25) DEFAULT NULL,
hitRatio float DEFAULT NULL,
errRatio float DEFAULT NULL,
year int DEFAULT NULL,
month int DEFAULT NULL,
day int DEFAULT NULL,
hour int DEFAULT NULL,
UNIQUE KEY fuplicateConstraint (blogName)
)
```

- sqoop command

```
sqoop export --connect jdbc:mysql://<mysql path>/<databases name>
    --username root --table logdata -m 1 --export-dir <exprot dir>
    --input-fields-terminated-by '\t' --input-null-string 'This is a Null
    String' --input-null-non-string '\\N'
```

- list few reason for sqoop export fail. Can sqoop export faulure result in a partial amount of dta being exported to mysql?
  Reasons for fail:
  - Loss of connectivity from the Hadoop cluster to the database (either due to hardware fault, or server software crashes)
  - Attempting to INSERT a row which violates a consistency constraint (for example, inserting a duplicate primary key value)
  - Attempting to parse an incomplete or malformed record from the HDFS source data
  - Attempting to parse records using incorrect delimiters
  - Capacity issues (such as insufficient RAM or disk space)

  individual map tasks commit their current transaction periodically. If a task fails, the current transaction will be rolled back. Any previously-committed transactions will remain durable in the database, leading to a partially-complete export.

Zhihao Xue
CM 3357
Lab7
CSSE490 Dr. Mohan
Thur,Oct 28, 2015

---

- What is the role of a staging table in a sqoop Export?
  Since Sqoop breaks down export process into multiple transactions, it is possible that a failed export job may result in partial data being committed to the database. This can further lead to subsequent jobs failing due to insert collisions in some cases, or lead to duplicated data in others. You can overcome this problem by specifying a staging table via the –staging-table option