Integrating MySQL with Hadoop

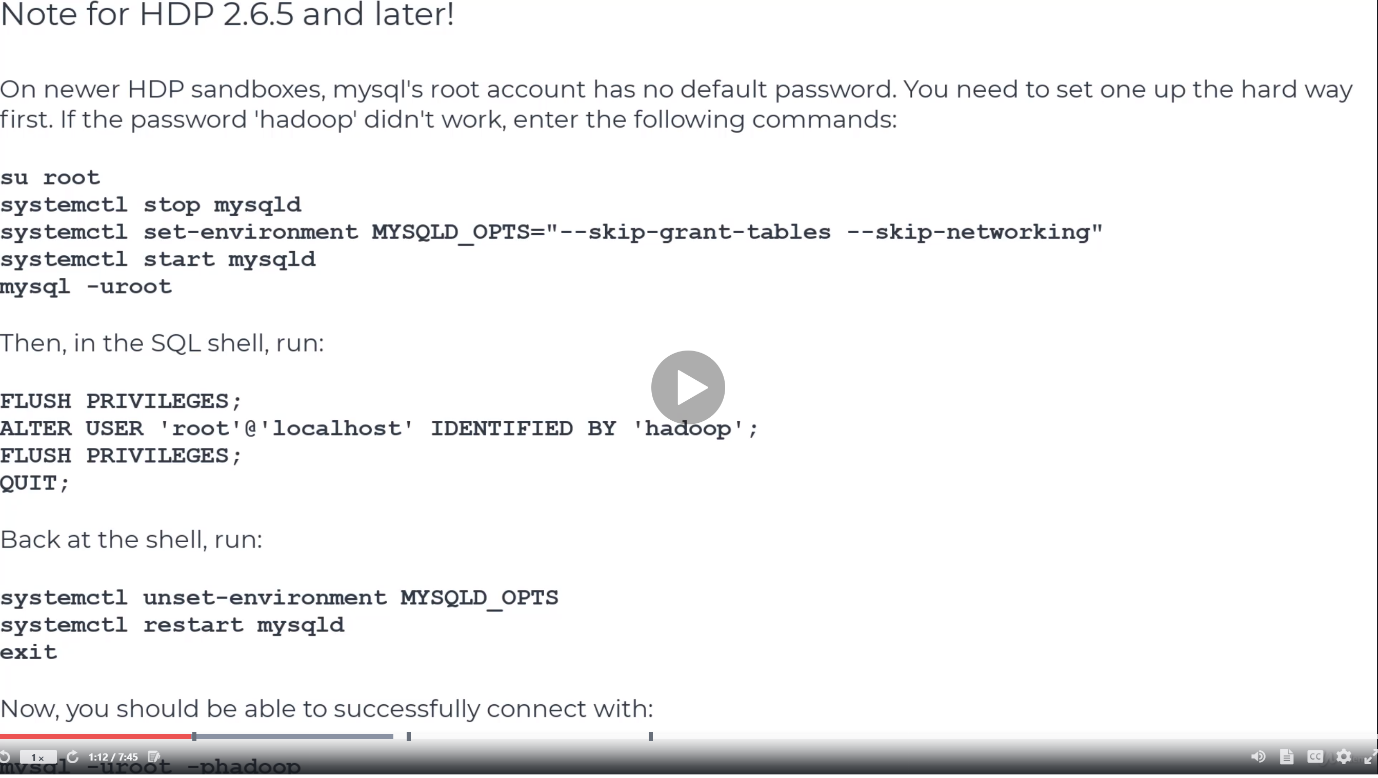
Online Transaction.

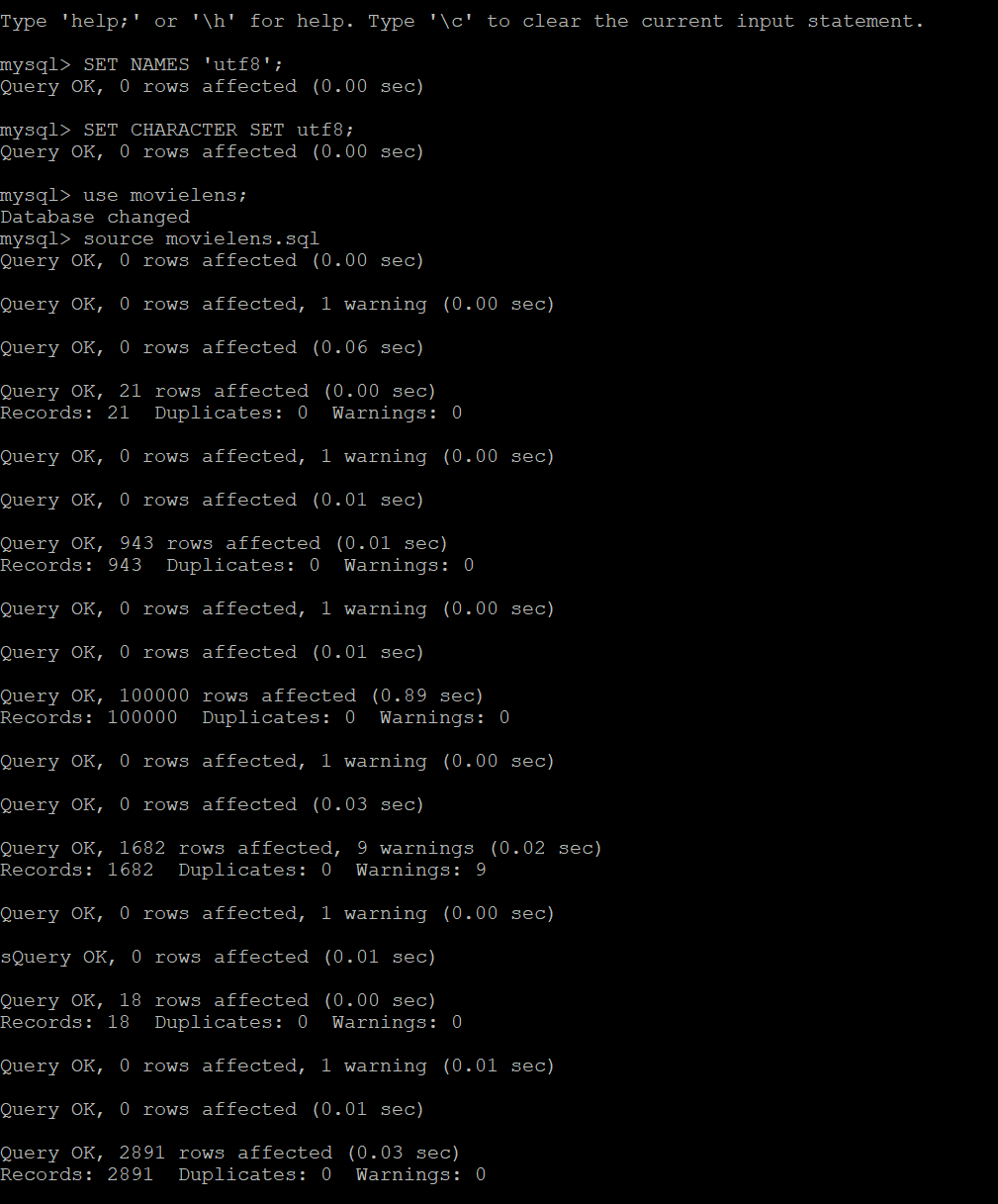
More parallel than it looks. Each individual Mapper will be talking to a block of HDFS.

You can have incrementnal imports.

You can keep your relational database and Hadoop In sync

--check-column and –last-value

Use Putty for the activity

SELECT movies.title, COUNT(ratings.movie\_id) AS ratingCount

FROM movies

INNER JOIN ratings

ON movies.id = ratings.movie\_id

GROUP BY movies.id

ORDER BY ratingCount;

USE scoop to import data from MySQL to HDFS/Hive

Log into mySql and type GRANT ALL PRIVILEGES ON movielens.\* to ‘’@’localhost’;

Exit

Sqoop import --connect jdbc:mysql://localhost/movielens --driver com.mysql.jdbc.Driver --table movies -m 1 --hive-import

It did, files view, user directory for maria\_dev. We have success file. Part-m-0000 because we got oe mapper.

It will input straight into the Hive warehouse. Kafka – Highly distributed tool for broadcasting the data.

HIVE have warehouse, and they are plain old texts. HIVE don’t use comma as delimiter.

Use scoop to import back into MySQL.

Need to ensure that the table exists in mySQL to receive the data

In mysql -> type use movielens;

CREATE TABLE exported\_movies (id INTEGER, title VARCHAR(255), releaseDate DATE);

Exit

Sqoop export --connect jdbc:mysql://localhost/movielens --driver com.mysql.jdbc.Driver --table exported\_movies --export-dir /apps/hive/warehouse/movies --input-fields-terminated-by ‘\0001’

To verify

Use movielens;

Select \* from exported\_movies limit 10;