

Chapter 2 – Loading Data into Hive

Solutions

Part 1

1. `$ tar xvzf ml-data.tar__0.gz`
2. `$ less u.data`
`$ less u.info`
3. `hive> CREATE TABLE UserRatings (userid INT, movieid INT, rating INT, unixtime int) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;`
4. Move the u.data file into HDFS with this command:
`$ hadoop fs -put ml-data/u.data /user/training/u.data`
5. `hive> LOAD DATA INPATH '/user/training/u.data' INTO TABLE UserRatings;`
6. `hive> SELECT * FROM UserRatings LIMIT 10;`

Part 2

1. `$ mysql`
`mysql> use training`
`mysql> DESCRIBE Movies;`
2. `$ sqoop --connect jdbc:mysql://localhost/training --table Movies --hive-import --fields-terminated-by '\001' --lines-terminated-by '\n' --username training`
3. `hive> SHOW TABLES;`
`hive> DESCRIBE Movies;`
`hive> SELECT * FROM Movies LIMIT 10;`
4. `$ hadoop fs -ls /user/hive/warehouse`
`$ hadoop fs -ls /user/hive/warehouse/userratings`
`$ hadoop fs -ls /user/hive/warehouse/movies`