Importing Data With Sqoop

**In this exercise you will import data from a relational database using Sqoop. The data you load here will be used subsequent exercises.**

Consider the MySQL database movielens, derived from the MovieLens project from University of Minnesota. (See note at the end of this exercise.) The database consists of several related tables, but we will import only two of these: movie, which contains about 3,900 movies; and movierating, which has about 1,000,000 ratings of those movies.

**Review the Database Tables**

First, review the database tables to be loaded into Hadoop.

1. Log on to MySQL:

$ mysql --user=[username] --password=[password] movielens

2. Review the structure and contents of the movie table:

mysql> DESCRIBE movie;

. . .

mysql> SELECT \* FROM movie LIMIT 5;

1. Note the column names for the table:

Review the structure and contents of the movierating table:

mysql> DESCRIBE movierating;

…

mysql> SELECT \* FROM movierating LIMIT 5;

1. Note these column names:
2. Exit mysql:

mysql> quit

**Import with Sqoop**

You invoke Sqoop on the command line to perform several commands. With it you can

connect to your database server to list the databases (schemas) to which you have

access, and list the tables available for loading. For database access, you provide a

connect string to identify the server, and -‐ if required -‐ your username and password.

1. Show the commands available in Sqoop:

$ sqoop help

2. List the databases (schemas) in your database server:

$ sqoop list-databases

--connect jdbc:mysql://localhost

--username [username] --password [password]

(NOTE: Instead of entering --password [password] on your command line, you may prefer to enter -P, and let Sqoop prompt you for the password, which is then not visible when you type it.)

3. List the tables in the movielens database:

$ sqoop list-tables

--connect jdbc:mysql://localhost/movielens

--username [username] --password [password]

4. Import the movie table into Hadoop:

$ sqoop import

--connect jdbc:mysql://localhost/movielens

--username [username] --password [password]

--fields-terminated-by '\t' --table movie

5. Verify that the command has worked.

$ hdfs dfs -ls movie

$ hdfs dfs -tail movie/part-m-00000

1. Import the movierating table into Hadoop.

Repeat the last two steps, but for the movierating table.

**END**