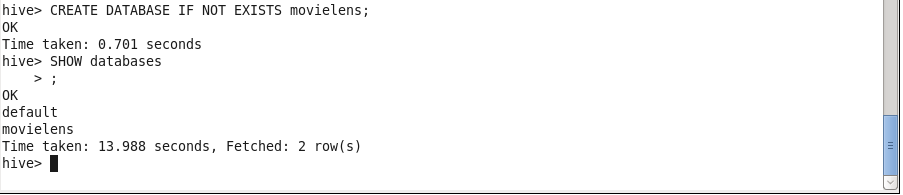
**Assignment 7 Solution**

**STUDENT NAME:**

1. **Create database command line syntax you ran to create a database called “movielens” (Hive SQL you used)**

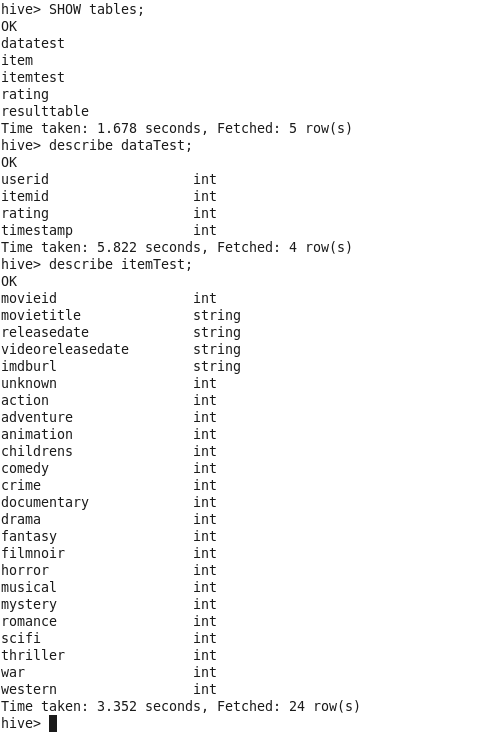


Before creating a “rating” table,

I loaded data from u.data in to *dataTest* table,

Then I loaded data from u.item in to *itemTest* table.

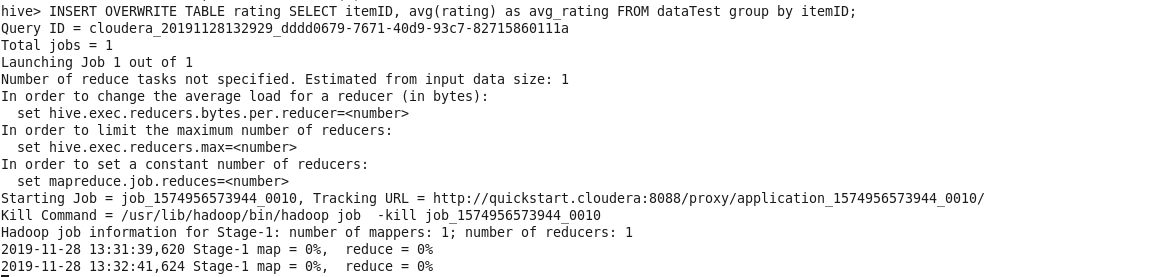
Please see *dataTest* and *itemTest* below with column names:



1. **Create table command line syntax you ran to create a table called “rating” (Hive SQL you used)**

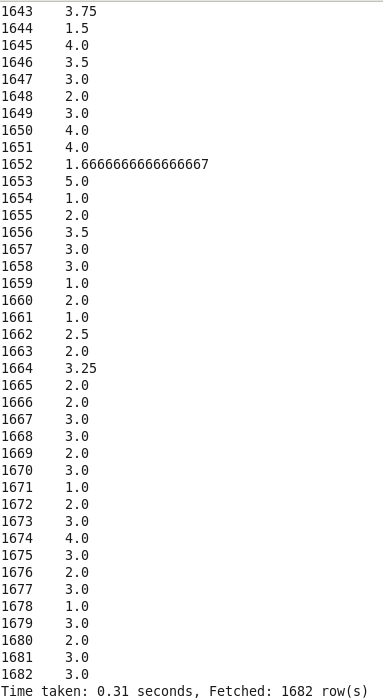
Then I filtered columns from *dataTest* and *itemTest*:

- “itemID” and “rating” from *dataTest*



\*Note: “rating” in *dataTest* is the average rating

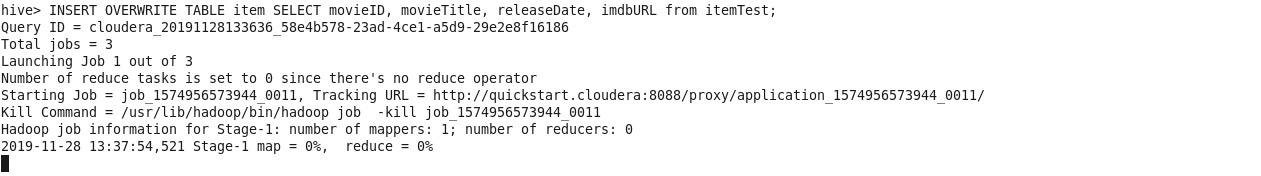
“rating” table contents below:



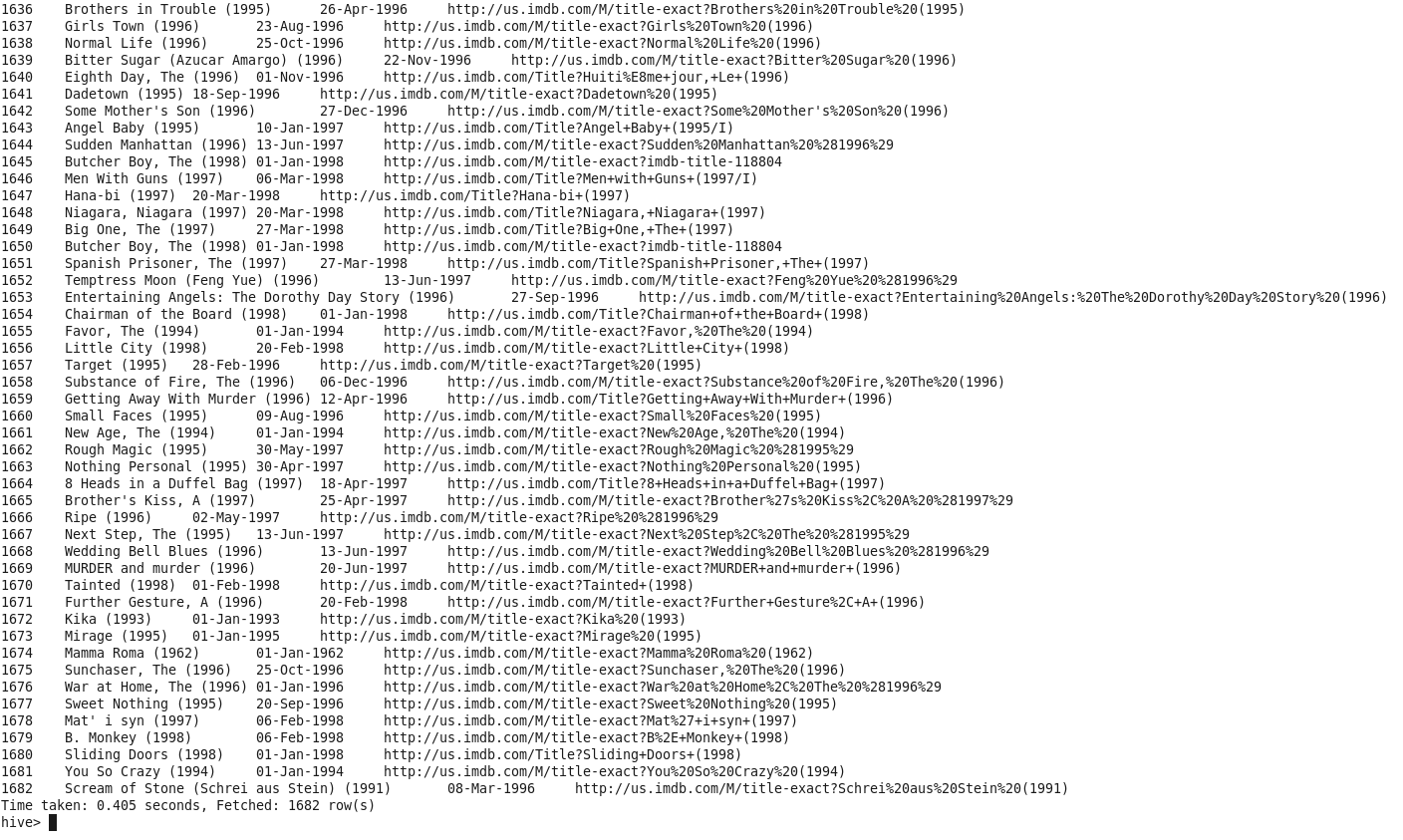
1. **Create table command line syntax you ran to create a table called “item” (Hive SQL you used)**

Then I filtered columns from *itemTest*:

- “movieID”, “movieTitle”, “releaseDate”, “imdbURL”



“item” table contents below:



1. **Syntax you used to load data into rating table. (Hive or whatever syntax you used).**

**INSERT OVERWRITE TABLE rating SELECT itemID, avg(rating) as avg\_rating FROM dataTest group by itemID;**

1. **Syntax you used to load data into item table. (Hive or whatever syntax you used).**

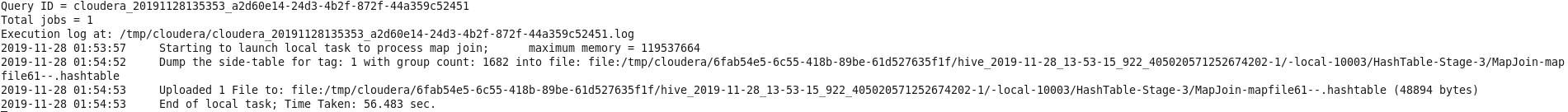
**INSERT OVERWRITE TABLE item SELECT movieID, movieTitle, releaseDate, imdbURL from itemTest;**

1. **Select statement with join and group by to retrieve the records, or whatever statement(s) you used to extract data as described in the assignment (Hive SQL you used)**



1. **Run the program on Hive command line and place the screenshot of successful run/execution right here. You can limit your run to 10 results (if sending output to screen), or you can send the output to a file. At this point show the execution of Hive query to list of all movies rated with the following information per line -- movie id, movie title, release date, IMDB URL, average rating for the movie.**



1.  **A screen shot of the JobHistoryServer when   
    you ran the job. Click on the Job, get into the   
    details, and then take a screen shot.**
2. **For the movie id 376 and 495, what is the movie title, release date, IMDB URL, average rating that you got as result using Hive. Please fill the data grid below:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Movie**  **id** | **Title** | **Release date** | **IMDB URL** | **Average rating** |
| **376** | **Houseguest (1994)** | **01-Jan-1994** | **http://us.imdb.com/M/title-exact?Houseguest%20(1994)** | **2.8333333333333335** |
| **495** | **Around the World in 80 Days (1956)** | **01-Jan-1956** | **http://us.imdb.com/M/title-exact?Around%20the%20World%20in%2080%20Days%20(1956)** | **3.6610169491525424** |