Pig Case Study

Problem Statement:

What are the movie titles that the user has rated?

How many times a movie has been rated by the user?

In question 2 above, what is the average rating given for a movie?

In Line 1: Registering the *piggybank* jar in order to use the CSVExcelStorage class.

In relation **A**, loaded the movies dataset using CSVExcelStorage because of its effective technique to handle double quotes and headers.

In relation **B**, generated the columns that are required for processing and explicitly typecasting each of them.

In relation C, filtered the null values from the "title" column.

In relation **D**, ordered the relation **'C'** in ascending.

In relation **F**, loaded the ratings dataset using CSVExcelStorage because of its effective technique to handle double quotes and headers.

In relation **G**, generated the columns that are required for processing and explicitly typecasting each of them.

In relation **H**, grouped relation '**G**' based on movield.

In relation I, generated the grouped column and count of each.

In relation **J**, filtered the relation '**G**' by using ratings greater than one.

In relation **K**, ordered the relation **'J**' based on movield.

In relation L, cogrouped relations 'D' and 'K' based on movied.

In relation **M**, ordered relation 'L' with first column in ascending.

In relation **N**, generated grouped column from relation '**M**' and average of column rating in relation '**K**'.

In relation **0**, joined relations '**D**' and '**N**'.

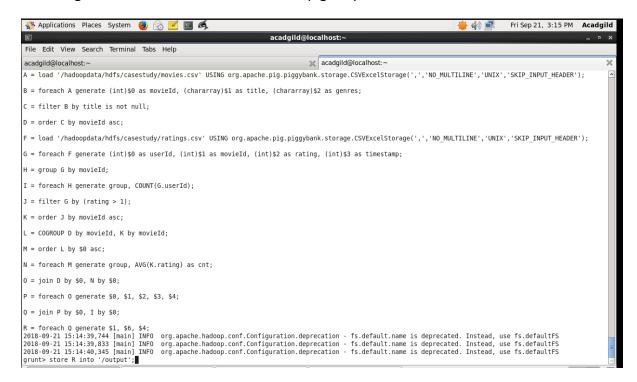
In relation **P**, generated the columns that are required for processing.

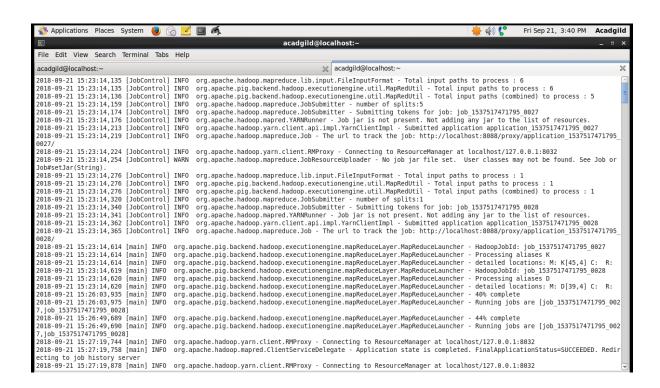
In relation **Q**, joined relations '**P**' and '**I**'.

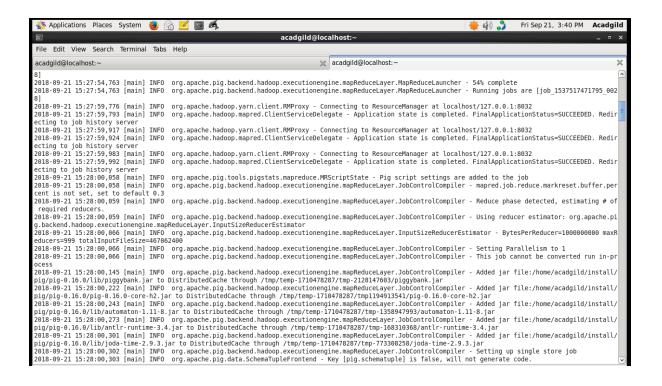
In relation **R**, generated the columns which are required for the output.

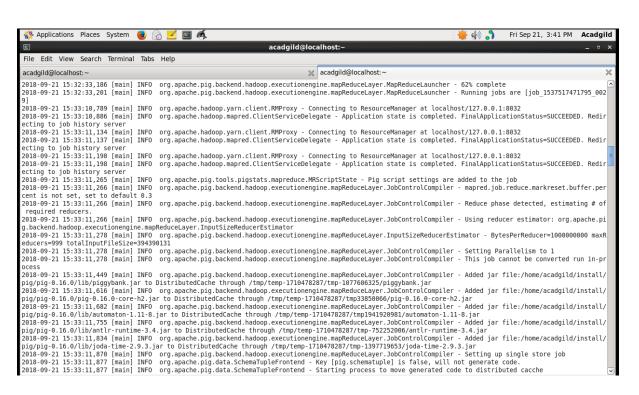
Finally using the store function, the result is dumped into the output folder.

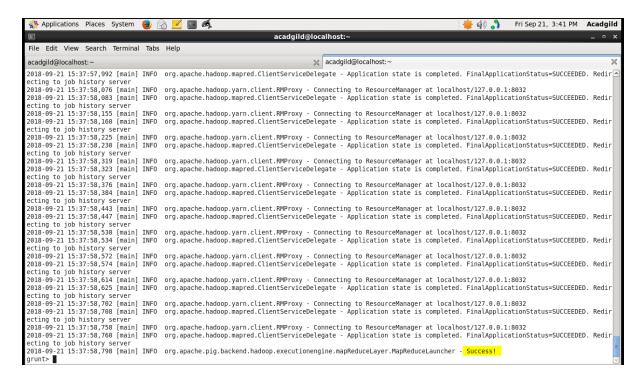
Following screenshot shows the relations of pig script.











The following screenshot shows the required output for the script written.

