

1. How to load the file and print results in Pig Latin script?  
`records = LOAD 'Movie_Log.txt' AS (login_time:String, logout_time:String, user_name:String.....);`  
`Dump records;`
2. How to get the list of movies grouped by ratings.  
`Group_data = GROUP records BY ratings;`
3. What is the primitive “DESCRIBE” in Pig Latin?  
DESCRIBE is to print the information about the file, The describe operator is used to view the schema of a relation.
4. Extend the previous script to process the clickstream data into user sessions.  
`ClickStream = LOAD 'Movie_Log.txt' USING org.apache.pig.piggybank.storage.avro.AvroStorage('no_schema_check', 'schema_file', 'hdfs://bigdatalite.cocaldomain/user/oracle/moviework/schemas/activity.avsv');`
5. How can I use FOREACH statement in Pig Script?  
FOREACH operator is used to generate the specified data transformation based on column data.  
Such as `foreach_data = FOREACH order_by_data GENERATE completed _movie;`
6. How to get the top-rated movies in each group.  
`order_by_data = ORDER records BY rated_point ASC;`  
`foreach_data = FOREACH order_by_data GENERATE completed _movie;`  
`Dump foreach_data;`
7. Select only the clicks, which correspond to starting, browsing, completing, or purchasing movies.  
`SELECT login_time, incompleted _movie, completed _movie, purchased _movie FROM records`

## Exercise 2

1. Describe the structure of students file in Pig Latin script.  
`records = LOAD 'students.csv' AS (.....);`  
`DESCRIBE records;`
2. Write a pig script to assign names to the data fields of the students.csv data. The output file should be called “students\_details”.

```
students_details = JOIN names BY id, records BY id;  
Dump students_details;
```

3. Perform any 3 operations on the file "students\_attendance.csv". The output of the 3rd operation in this case should be called "SA\_details".  
attendance = LOAD ' students\_attendance.csv ' AS (.....);  
First = FILTER attendance BY hours > 5;  
Second = ORDER First BY hours ASC;  
SA\_details = FOREACH Second GENERATE names, hours;
4. Write a script to find the sum of hours attended by each student.  
Hours\_sum = FOREACH SA\_details GENERATE names, hours,  
SUM(hours);
5. Write a script to join StudentID, Name, hours attended.  
Students\_join = JOIN StudentID BY id, Name BY id, hours\_attended BY id;
6. Print the results on the screen.  
Dump Students\_join;