**Step 1:**

**Loading of dataset into movies relation:**

**Command:**

movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

**Screenshot:**

****

**Results:**

****

1. **A. Find the number of movies released between 1950 and 1960.**

**Command:**

movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

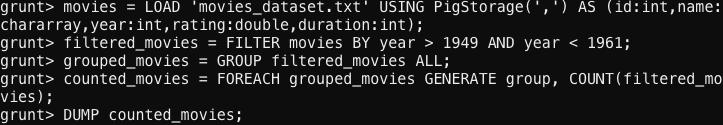
filtered\_movies = FILTER movies BY year > 1949 AND year < 1961;

grouped\_movies = GROUP filtered\_movies ALL;

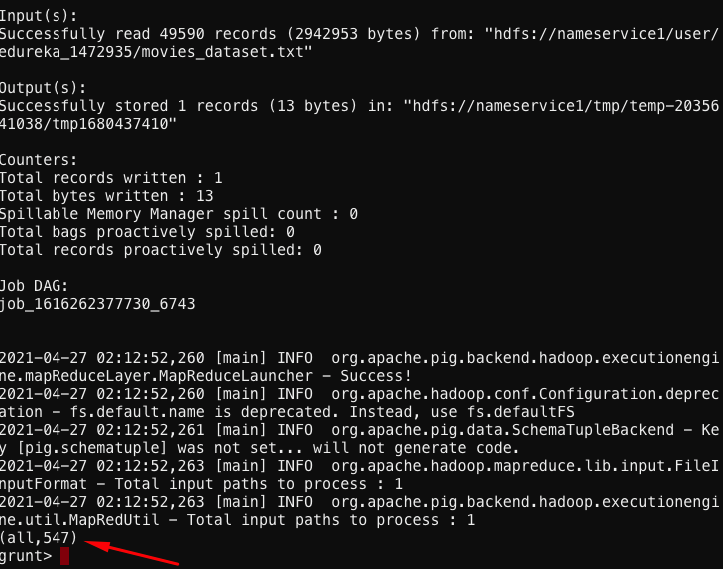
counted\_movies = FOREACH grouped\_movies GENERATE group, COUNT(filtered\_movies);

DUMP counted\_movies;

**Screenshot:**

****

**Results:**

****

1. **B. Find the number of movies having rating more than 4.**

**Command:**

movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

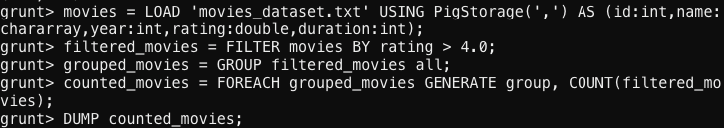
filtered\_movies = FILTER movies BY rating > 4.0;

grouped\_movies = GROUP filtered\_movies all;

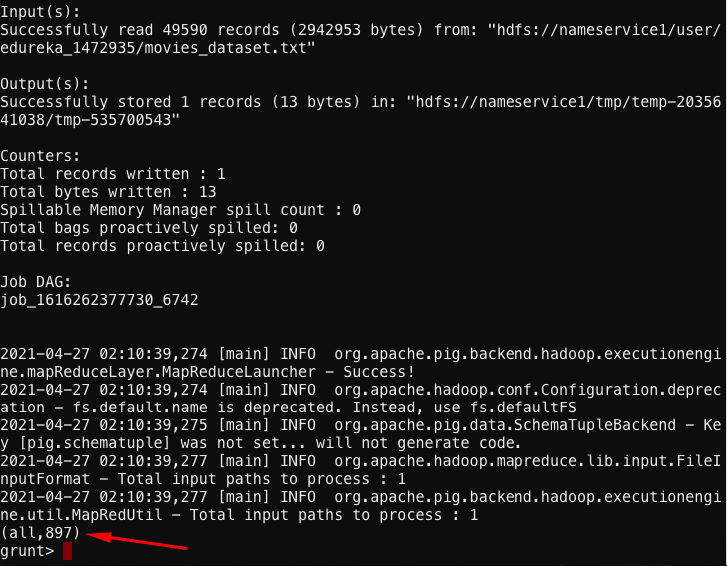
counted\_movies = FOREACH grouped\_movies GENERATE group, COUNT(filtered\_movies);

DUMP counted\_movies;

**Screenshot:**

****

**Result:**

****

1. **C. Find the movies whose rating are between 3 and 4.**

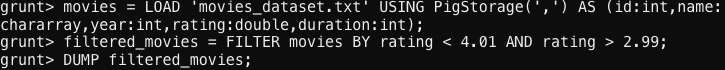
**Command:**

movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

filtered\_movies = FILTER movies BY rating < 4.01 AND rating > 2.99;

DUMP filtered\_movies;

**Screenshot:**

****

**Result:**

****

1. **D. Find the number of movies with duration more than 2 hours (7200 second).**

**Command:**

movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

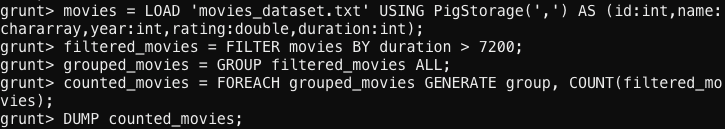
filtered\_movies = FILTER movies BY duration > 7200;

grouped\_movies = GROUP filtered\_movies ALL;

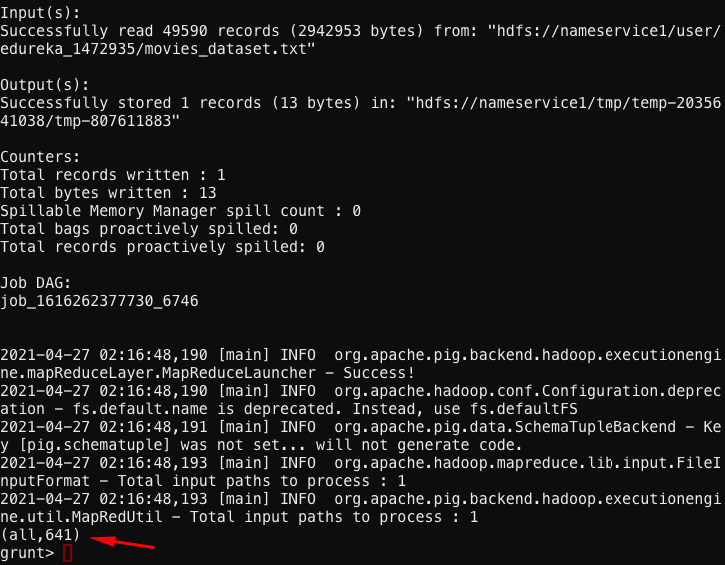
counted\_movies = FOREACH grouped\_movies GENERATE group, COUNT(filtered\_movies);

DUMP counted\_movies;

**Screenshot:**

****

**Result:**



1. **E. Find the list of years and number of movies released each year.**

**Command:**

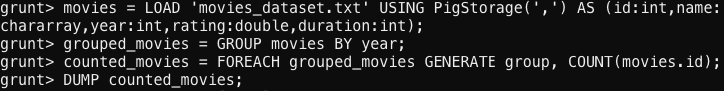
movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

grouped\_movies = GROUP movies BY year;

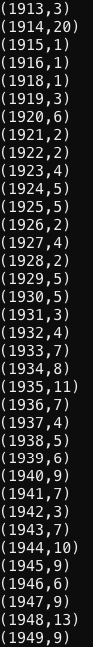
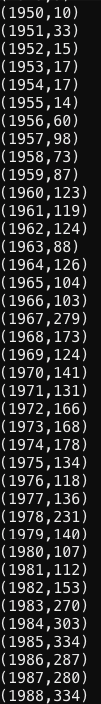
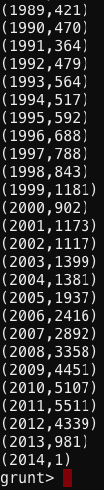
counted\_movies = FOREACH grouped\_movies GENERATE group, COUNT(movies.id);

DUMP counted\_movies;

**Screenshot:**

****

**Result:**

1. **F. Find the total number of movies in the dataset.**

**Command:**

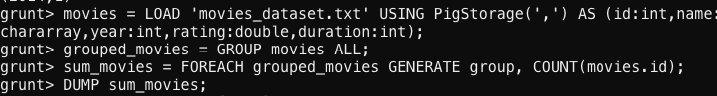
movies = LOAD 'movies\_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);

grouped\_movies = GROUP movies ALL;

sum\_movies = FOREACH grouped\_movies GENERATE group, COUNT(movies.id);

DUMP sum\_movies;

**Screenshot:**



**Result:**

