--18.02.2022

---HIVE

--creating table movies

create table movies(movie\_id int, movie\_name varchar(30),year\_of\_release int,movie\_rating float,duration\_in\_secs long) row format delimited fields terminated by ',';

describe movies;

--loading data into the movies table from local computer

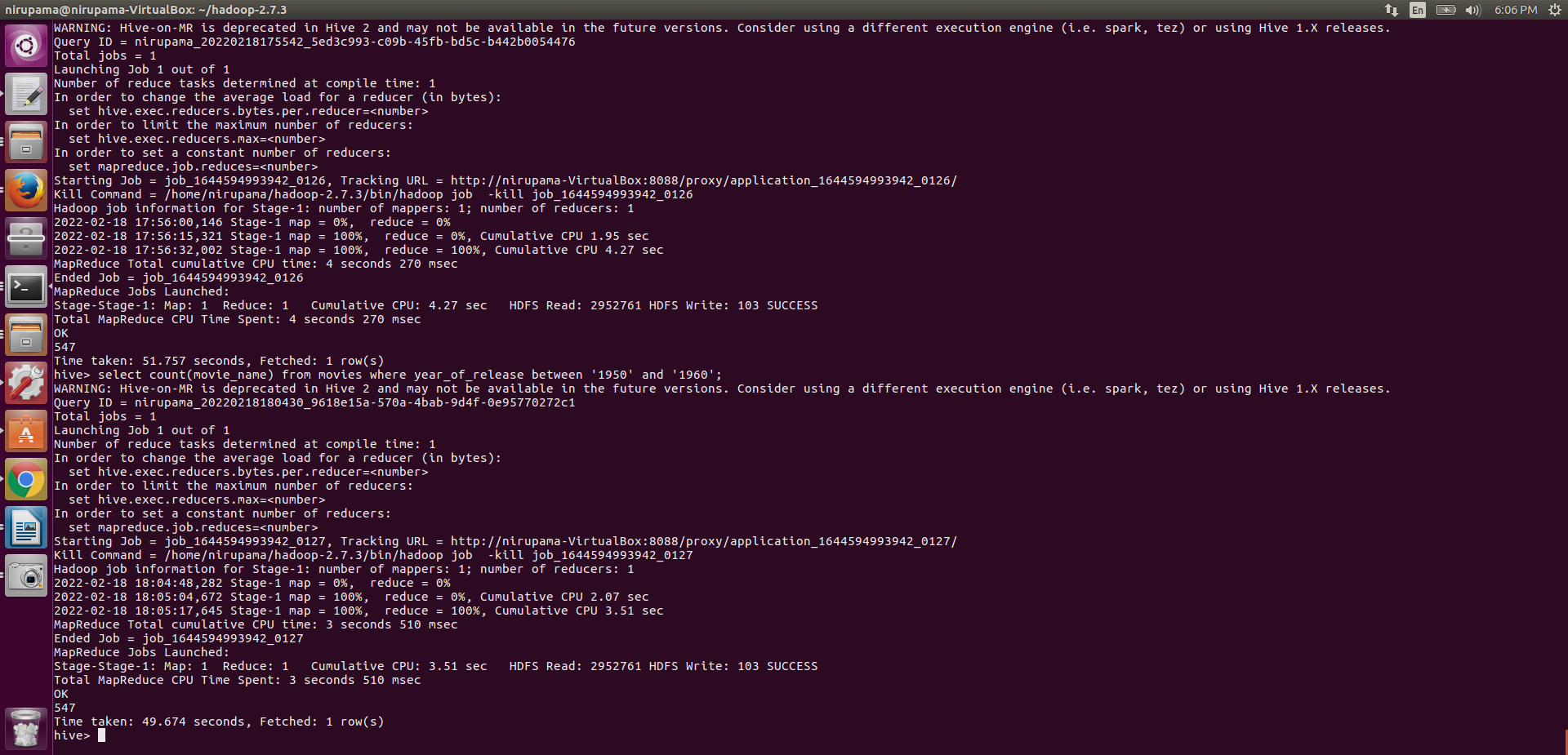
load data local inpath '/home/nirupama/Desktop/Assessment/movies\_dataset.txt' into table movies;

select \*from movies;

PROBLEM STATEMENTS:-

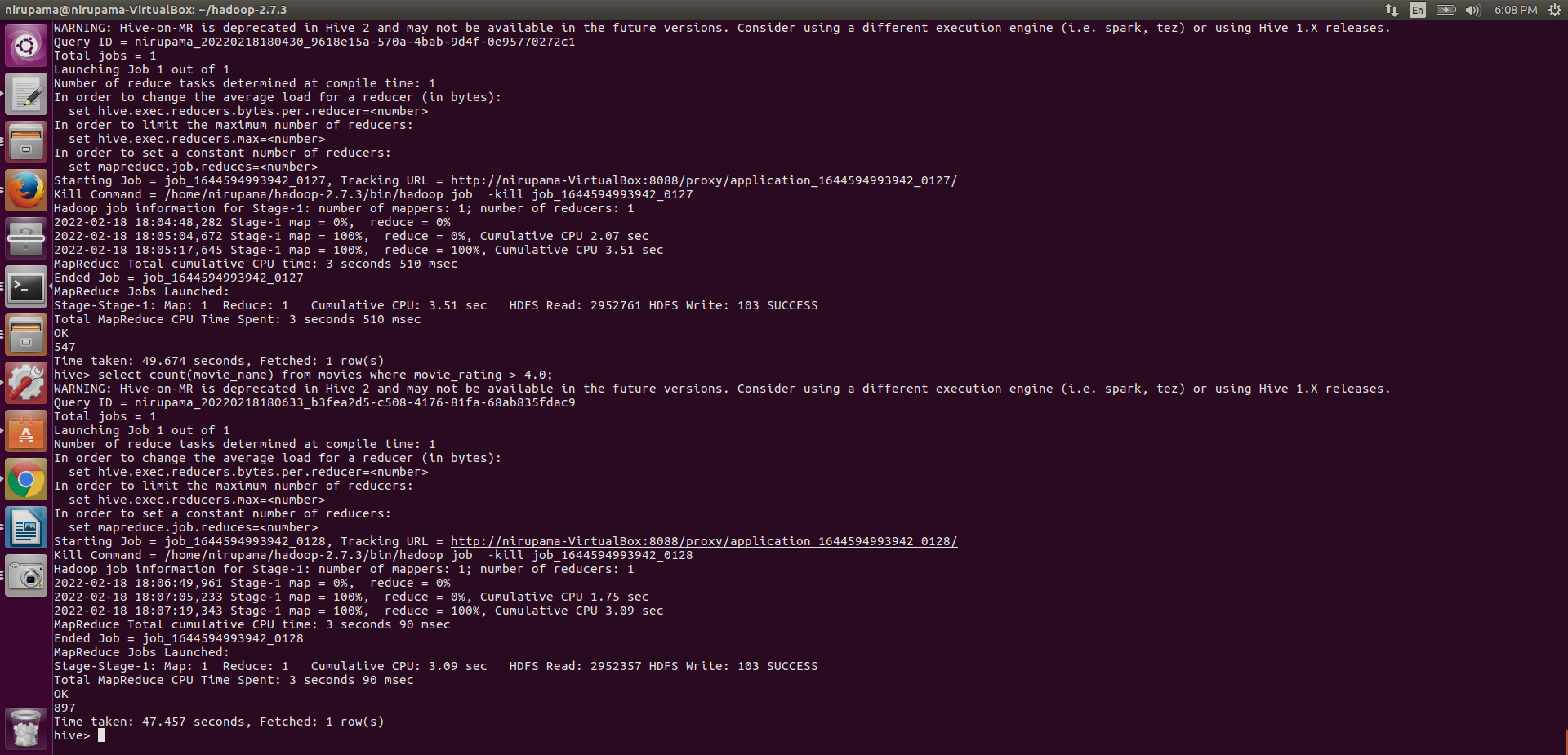
A)

select count(movie\_name) from movies where year\_of\_release between '1950' and '1960';



B)

select count(movie\_name) from movies where movie\_rating > 4.0;



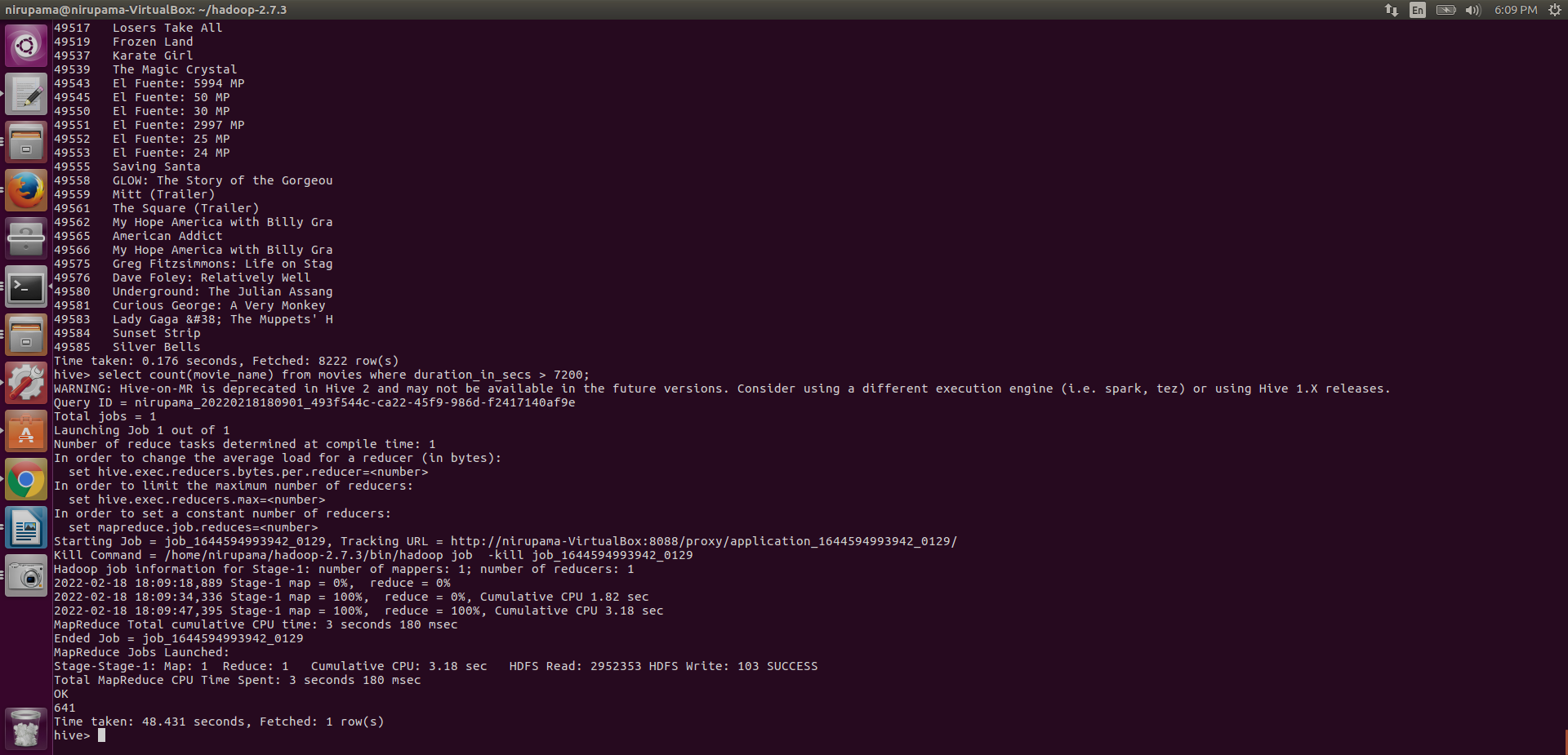
C)

select movie\_id,movie\_name from movies where movie\_rating between 3.0 and 4.0;



D)

select count(movie\_name) from movies where duration\_in\_secs > 7200;



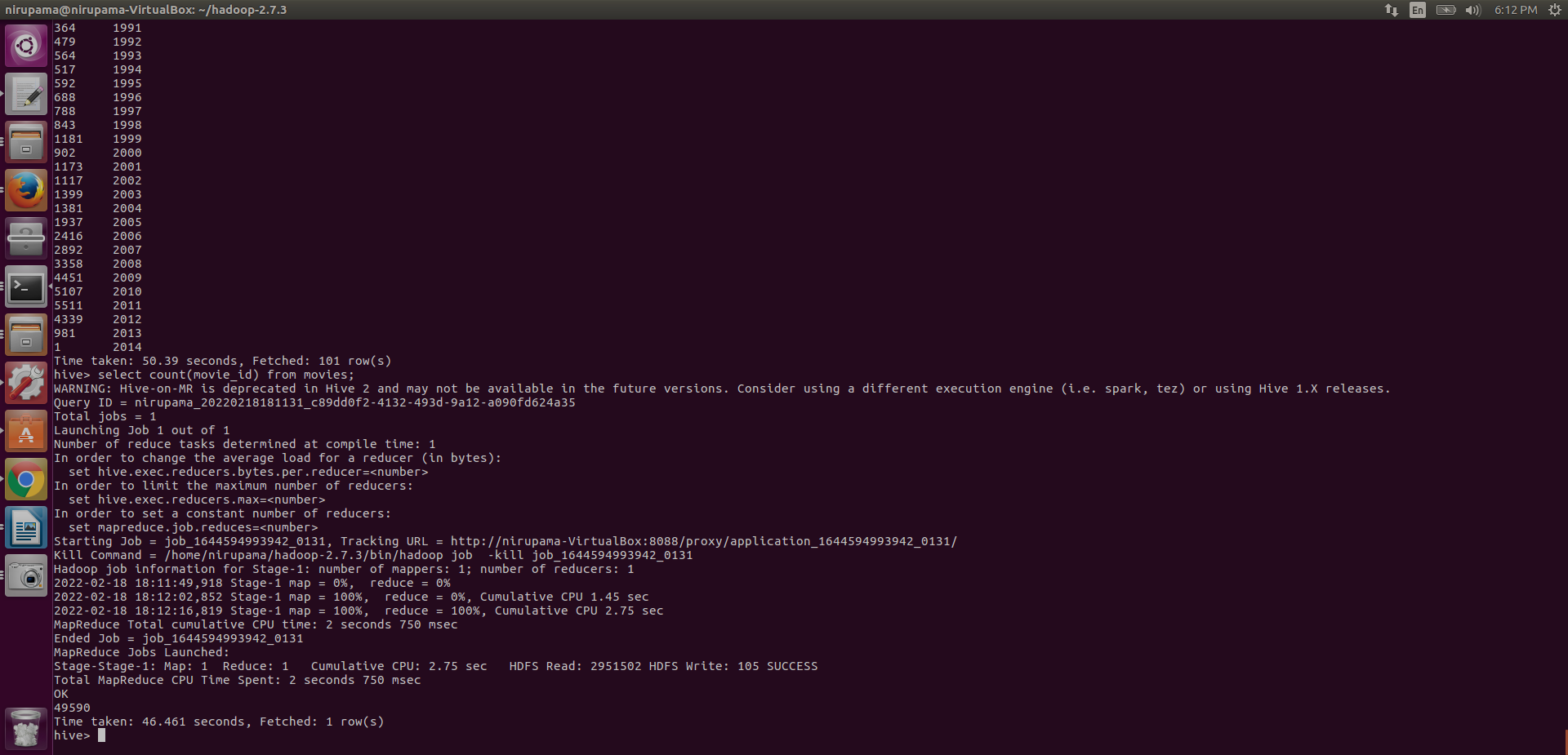
E)

select count(\*), year\_of\_release from movies group by year\_of\_release;



F)

select count(movie\_id) from movies;



--18/02/2022

--creating assignment directory and inserting country.txt file into hdfs

bin/hdfs dfs -mkdir /assessment/w3

bin/hdfs dfs -put /home/nirupama/Desktop/Assessment/movies\_dataset.txt /assessment/w3

--pig

--loading data from hdfs into movies variable

movies = load '/assessment/w3/movies\_dataset.txt' using PigStorage(',') as (movie\_id:int, movie\_name:chararray,year\_of\_release:int,movie\_rating:float,duration\_in\_secs:long);

describe movies; --to give the schema

illustrate movies; -- gives a random value present

--PROBLEM STATEMENT:-

A)

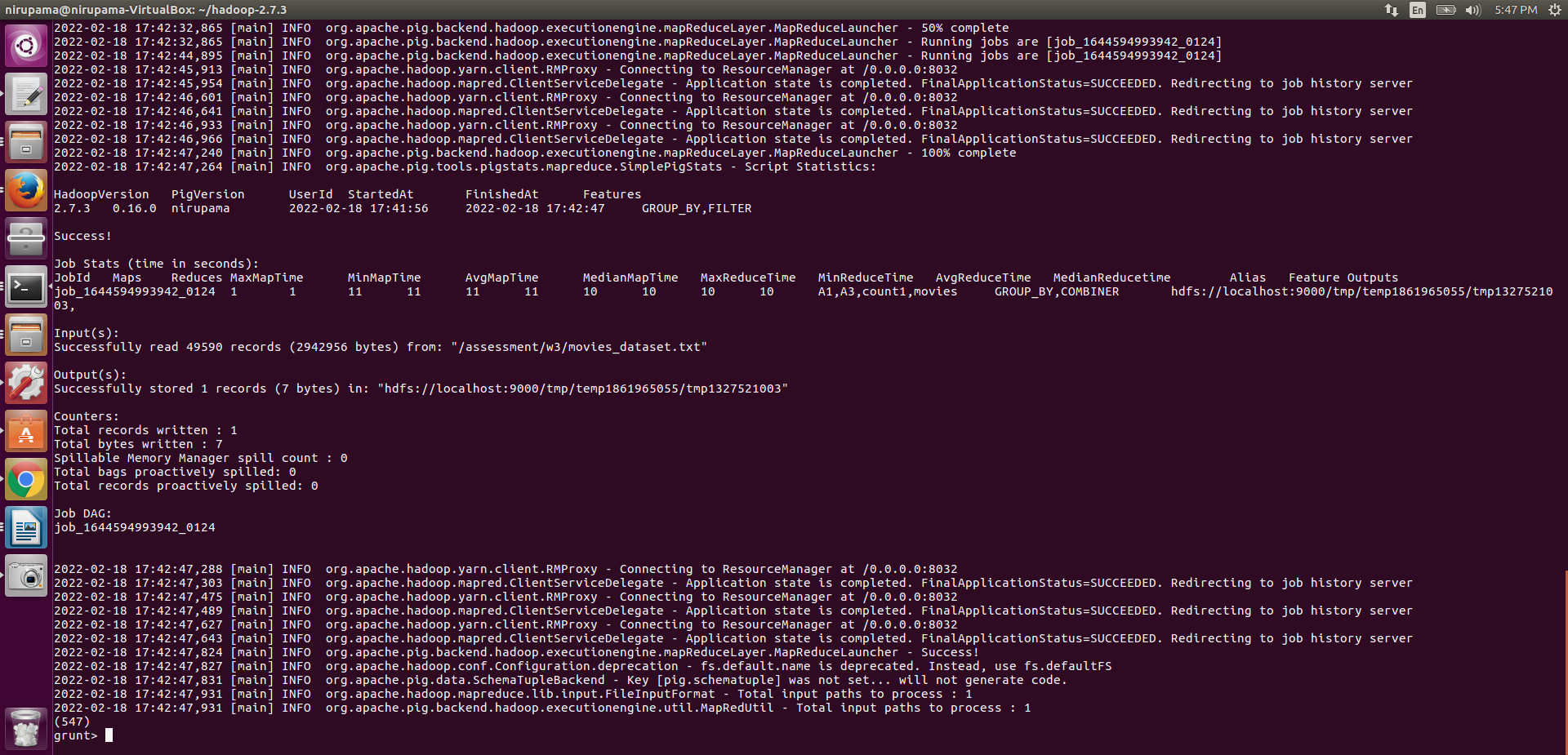
A1 = FILTER movies by year\_of\_release >= 1950;

A2 = FILTER A1 by year\_of\_release <= 1960;

A3 = GROUP A2 All;

count1 = FOREACH A3 GENERATE COUNT(A2);

DUMP count1;



B)

B1 = FILTER movies by movie\_rating > 4.0;

B2 = GROUP B1 All;

Count2 = FOREACH B2 GENERATE COUNT(B1);

DUMP count2;

C)

C1 = FILTER movies by movie\_rating >= 3.0;

C2 = FILTER C1 by movie\_rating <= 4.0;

Mov\_list = FOREACH C2 GENERATE movie\_name;

DUMP Mov\_list;

D)

D1 = FILTER movies by duration\_in\_secs > 7200;

D2 = GROUP D1 All;

count3 = FOREACH D2 GENERATE COUNT(D1);

DUMP count3;

E)

E1 = FOREACH movies GENERATE year\_of\_release;

Yr\_list = GROUP E1 by year\_of\_release ;

List = FOREACH Yr\_list GENERATE group, count(E1) AS Num\_Movies;

DUMP List;

F)

F1 = GROUP movies All;

count4 = FOREACH F1 GENERATE COUNT(movies);

DUMP count4;