|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | movie = load '/pig/movies.txt' USING PigStorage (',') as (movies\_id:int,movie\_name:chararray,yearOfRelease:int,rating:double,duration:int);   bin/hdfs dfs -put /home/sumeeksha/Downloads/Country.txt /pig <-------------- hadoop-2.7.3   A. Number of movies released in 1950 to 1960  A1 = filter movie by yearOfRelease >= 1950; A2 = filter A1 by yearOfRelease <= 1960; A3 = GROUP A2 all; A4 = foreach A3 GENERATE COUNT(A2); Dump A4;  Text  Description automatically generated     B. Rating 4 or greater  B1 = filter movie by rating >= 4; B2 = GROUP B1 by ALL; B3 = FOREACH B2 GENERATE COUNT(B1); DUMP B3;    C. Rating between 4 and 5 C1 = FILTER movie by rating < 4; C2 = FILTER C1 by rating < 4; C3 = FOREACH C2 GENERATE movie\_name; DUMP C3;  Text  Description automatically generated    D. Duration more than 2 hrs D1 = filter movie by duration > 7200; D2 = GROUP D1 All; D3 = foreach D2 GENERATE COUNT(D1); Dump D3; Text  Description automatically generated  E.  E1 = foreach movie by yearOfRelease; E2 = GROUP E1 by yearOfRelease; E3 = foreach E2 GENERATE GROUP, COUNT(E1) AS num\_movies; Dump E3;    F.  F1 = GROUP Movie All; F2 = foreach F1 GENERATE COUNT(Movie); DUMP F2;    Hive Commands  Create table Movie(movie\_id int,movie\_name string,yearOfRelease DATE,rating double,duration int) row format delimited fields terminated by ',';   load data local inpath '/home/sumeeksha/Downloads/movies\_dataset.txt' into table Movie;   A. Number of movies released in 1950 to 1960  Select COUNT(\*) FROM Movie where yearOfRelease >1950 and yearOfRelease <1960;  B. Number of movies having rating more than 4  Select Count(\*) from Movie where rating > 4;  C. Number of movies having rating in 3 to 4  Select movie\_name from Movie (where rating >3 and rating <4);  D. Number of movies whose duration is 7200 sec or 2 hrs  Select COUNT(\*) from Movie where duration>7200;  E.Count number movies per year  Select yearOfRelease, Count(movies\_name) as cnt from Movie group by yearOfRelease order by cnt;  F. Total Number Movies in the dataset  Select COUNT(movie\_name) from Movie;   |  |  | | --- | --- | |  | ReplyForward | |  |