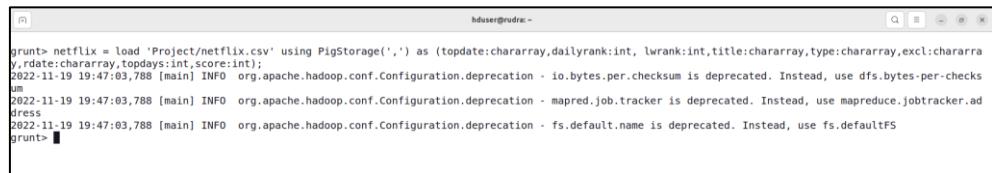


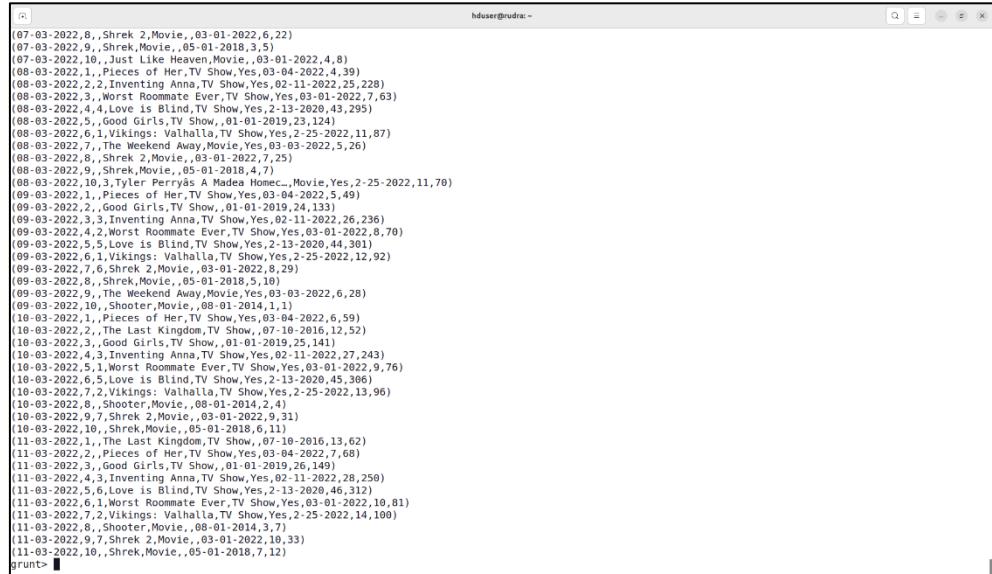
## QUERIES ALONG WITH THEIR SNAPSHOTS

- Query 1** Load the Netflix dataset of Daily Top 10 Shows/Movies into a relation by using Apache Pig

```
netflix = load 'Project/netflix.csv' using PigStorage(',')  
as (topdate:chararray,dailyrank:int,  
lwrank:int,title:chararray,type:chararray,excl:chararray,rda  
te:chararray,topdays:int,score:int);
```



```
hduser@rudra: ~  
grunt> netflix = load 'Project/netflix.csv' using PigStorage(',') as (topdate:chararray,dailyrank:int, lwrank:int,title:chararray,type:chararray,excl:chararray,rda  
te:chararray,topdays:int,score:int);  
2022-11-19 19:47:03,788 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checks  
um  
2022-11-19 19:47:03,788 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.ad  
dress  
2022-11-19 19:47:03,788 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> ■
```



```
hduser@rudra: ~  
(07-03-2022,8,,Shrek 2,Movie,,03-01-2022,6,22)  
(07-03-2022,9,,Shrek,Movie,,05-01-2018,3,5)  
(07-03-2022,10,,Just Like Heaven,Movie,,03-01-2022,4,8)  
(08-03-2022,1,,Pieces of Her,TV Show,Yes,03-04-2022,4,39)  
(08-03-2022,2,,The Last Kingdom,TV Show,,07-10-2016,13,62)  
(08-03-2022,3,,Worst Roommate Ever,TV Show,Yes,03-04-2022,25,228)  
(08-03-2022,4,,Love is Blind,TV Show,Yes,2-13-2020,7,63)  
(08-03-2022,4,4,Love is Blind,TV Show,Yes,03-04-2022,43,295)  
(08-03-2022,5,,Good Girls,TV Show,,01-01-2019,23,124)  
(08-03-2022,6,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,11,87)  
(08-03-2022,7,,The Weekend Away,Movie,Yes,03-03-2022,5,26)  
(08-03-2022,8,,Shrek 2,Movie,,03-01-2022,7,25)  
(08-03-2022,9,,Shrek,Movie,,05-01-2018,4,7)  
(08-03-2022,10,,I Try to Try to Make You Love Me,Movie,Yes,2-25-2022,11,78)  
(09-03-2022,1,,Places of Her,TV Show,Yes,03-04-2022,5,49)  
(09-03-2022,2,,Good Girls,TV Show,,01-01-2019,24,133)  
(09-03-2022,3,3,Inventing Anna,TV Show,Yes,02-11-2022,26,236)  
(09-03-2022,4,2,Worst Roommate Ever,TV Show,Yes,03-01-2022,8,70)  
(09-03-2022,5,5,Love is Blind,TV Show,Yes,2-13-2020,44,301)  
(09-03-2022,6,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,12,92)  
(09-03-2022,7,6,Shrek 2,Movie,,03-01-2022,8,29)  
(09-03-2022,8,,Shrek,Movie,,05-01-2018,5,18)  
(09-03-2022,9,,Shrek 2,Movie,,05-01-2018,5,18)  
(09-03-2022,10,,Shooter,Movie,,08-01-2014,1,1)  
(10-03-2022,1,,Places of Her,TV Show,Yes,03-04-2022,6,59)  
(10-03-2022,2,,The Last Kingdom,TV Show,,07-10-2016,12,52)  
(10-03-2022,3,,Good Girls,TV Show,,01-01-2019,25,141)  
(10-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,27,243)  
(10-03-2022,5,1,Worst Roommate Ever,TV Show,Yes,03-01-2022,9,76)  
(10-03-2022,6,5,Love is Blind,,Show,Yes,2-13-2020,45,306)  
(10-03-2022,7,1,Vikings: Valhalla,TV Show,es,2-25-2022,13,96)  
(10-03-2022,8,,Shooter,Movie,,08-01-2014,2,4)  
(10-03-2022,9,7,Shrek 2,Movie,,03-01-2022,9,31)  
(10-03-2022,10,,Shrek,Movie,,05-01-2018,6,11)  
(11-03-2022,1,,The Last Kingdom,TV Show,,07-10-2016,13,62)  
(11-03-2022,2,,Places of Her,TV Show,Yes,03-04-2022,7,68)  
(11-03-2022,3,,Good Girls,TV Show,,01-01-2019,26,149)  
(11-03-2022,4,2,Love is Blind,,Show,Yes,02-11-2022,28,250)  
(11-03-2022,5,6,Love is Blind,,Show,Yes,2-13-2020,46,312)  
(11-03-2022,6,1,Worst Roommate Ever,TV Show,Yes,03-01-2022,10,81)  
(11-03-2022,7,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,14,100)  
(11-03-2022,8,,Shooter,Movie,,08-01-2014,3,7)  
(11-03-2022,9,7,Shrek 2,Movie,,03-01-2022,10,33)  
(11-03-2022,10,,Shrek,Movie,,05-01-2018,7,12)  
grunt> ■
```

- Query 2** Display title, type and release date of each show/movie

```
rell1 = foreach netflix generate title, type, rdate;
```

```
hduser@rudra: ~
grunt> rel1 = foreach netflix generate title, type, rdate;
grunt>

hduser@rudra: ~
(Shrek 2,Movie,03-01-2022)
(Shrek,Movie,05-01-2018)
(Jurassic Park,Movie,03-01-2022)
(Pieces Of Her,TV Show,03-04-2022)
(Investing Anna,TV Show,02-11-2022)
(Worst Roommate Ever,TV Show,03-01-2022)
(Love Is Blind,TV Show,2-13-2020)
(Good Girls,TV Show,01-01-2019)
(Vikings: Valhalla,TV Show,2-25-2022)
(The Weekend Away,Movie,03-03-2022)
(Shrek 2,Movie,03-01-2022)
(Shrek,Movie,05-01-2018)
(Tyler Perry's Madea Homecoming,Movie,2-25-2022)
(Valhalla,TV Show,03-04-2022)
(Good Girls,TV Show,03-01-2019)
(Investing Anna,TV Show,02-11-2022)
(Worst Roommate Ever,TV Show,03-01-2022)
(Love Is Blind,TV Show,2-13-2020)
(Vikings: Valhalla,TV Show,2-25-2022)
(Shrek 2,Movie,03-01-2022)
(Shrek,Movie,05-01-2018)
(The Weekend Away,Movie,03-03-2022)
(Pieces Of Her,TV Show,03-04-2022)
(The Last Kingdom,TV Show,07-10-2016)
(Good Girls,TV Show,01-01-2019)
(Investing Anna,TV Show,02-11-2022)
(Worst Roommate Ever,TV Show,03-01-2022)
(Love Is Blind,TV Show,2-13-2020)
(Vikings: Valhalla,TV Show,2-25-2022)
(Shooter,Movie,08-01-2014)
(Shrek 2,Movie,03-01-2022)
(Shrek,Movie,05-01-2018)
(The Last Kingdom,TV Show,07-10-2016)
(Pieces Of Her,TV Show,03-04-2022)
(Good Girls,TV Show,01-01-2019)
(Investing Anna,TV Show,02-11-2022)
(Love Is Blind,TV Show,2-13-2020)
(Worst Roommate Ever,TV Show,03-01-2022)
(Vikings: Valhalla,TV Show,2-25-2022)
(Shooter,Movie,08-01-2014)
(Shrek 2,Movie,03-01-2022)
(Shrek,Movie,05-01-2018)
grunt>
```

**Query 3** Display the show/movie details in descending order by Viewership Score

```
order_score = order netflix by score desc;
```

```
hduser@udra: ~
grunt| order_score = order netflix by score desc;
grunt| [{"id": 1, "score": 9.5, "title": "Indian Matchmaking", "type": "TV Show", "year": 2020, "genre": "Romantic"}, {"id": 2, "score": 8.7, "title": "Drunk Parents", "type": "Movie", "year": 2020, "genre": "Comedy"}, {"id": 3, "score": 8.1, "title": "Major Payne", "type": "Movie", "year": 2021, "genre": "Thriller"}, {"id": 4, "score": 8.0, "title": "Training Day", "type": "Movie", "year": 2001, "genre": "Thriller"}, {"id": 5, "score": 7.8, "title": "Burden Inside", "type": "Stand-up Comedy", "year": 2021, "genre": "Comedy"}, {"id": 6, "score": 7.5, "title": "Moesta", "type": "TV Show", "year": 2020, "genre": "Drama"}, {"id": 7, "score": 7.0, "title": "House at the End of the Street", "type": "Movie", "year": 2013, "genre": "Horror"}, {"id": 8, "score": 6.9, "title": "The Dark Knight", "type": "Movie", "year": 2001, "genre": "Action"}, {"id": 9, "score": 6.8, "title": "Shooter", "type": "Movie", "year": 2014, "genre": "Thriller"}, {"id": 10, "score": 6.7, "title": "Game On: A Comedy Crossover", "type": "TV Show", "year": 2020, "genre": "Comedy"}, {"id": 11, "score": 6.5, "title": "The Last Blockbuster", "type": "Movie", "year": 2021, "genre": "Science Fiction"}, {"id": 12, "score": 6.4, "title": "Castaway", "type": "TV Show", "year": 2009, "genre": "Drama"}, {"id": 13, "score": 6.3, "title": "Desperado", "type": "Movie", "year": 2000, "genre": "Action"}, {"id": 14, "score": 6.2, "title": "In the Dark", "type": "TV Show", "year": 2019, "genre": "Mystery"}, {"id": 15, "score": 6.1, "title": "Scary Stories to Tell in the Dark", "type": "Movie", "year": 2022, "genre": "Horror"}, {"id": 16, "score": 6.0, "title": "The Verge TV Show", "type": "Yes", "year": 2021, "genre": "Documentary"}, {"id": 17, "score": 5.9, "title": "Dark Shadows", "type": "Movie", "year": 2022, "genre": "Horror"}, {"id": 18, "score": 5.8, "title": "The Twilight Saga: Eclipse", "type": "Movie", "year": 2011, "genre": "Horror"}, {"id": 19, "score": 5.7, "title": "Double Jeopardy", "type": "Movie", "year": 2016, "genre": "Thriller"}, {"id": 20, "score": 5.6, "title": "The Devil's Advocate", "type": "Movie", "year": 1997, "genre": "Thriller"}, {"id": 21, "score": 5.5, "title": "The Lure", "type": "Movie", "year": 2020, "genre": "Horror"}, {"id": 22, "score": 5.4, "title": "The BFG", "type": "Movie", "year": 2017, "genre": "Family"}, {"id": 23, "score": 5.3, "title": "The Main Event", "type": "Movie", "year": 2020, "genre": "Sports"}, {"id": 24, "score": 5.2, "title": "Sharkdog", "type": "TV Show", "year": 2021, "genre": "Comedy"}, {"id": 25, "score": 5.1, "title": "Daughter From Another Mother", "type": "TV Show", "year": 2021, "genre": "Drama"}, {"id": 26, "score": 5.0, "title": "Wentworth", "type": "TV Show", "year": 2014, "genre": "Drama"}, {"id": 27, "score": 4.9, "title": "Grand Army", "type": "TV Show", "year": 2020, "genre": "Drama"}, {"id": 28, "score": 4.8, "title": "Natalie Christy", "type": "Movie", "year": 2020, "genre": "Drama"}, {"id": 29, "score": 4.7, "title": "Unbreakable Kimmy Schmidt", "type": "TV Show", "year": 2017, "genre": "Comedy"}, {"id": 30, "score": 4.6, "title": "The Queen's Gambit", "type": "TV Show", "year": 2020, "genre": "Drama"}, {"id": 31, "score": 4.5, "title": "His House", "type": "Movie", "year": 2020, "genre": "Thriller"}, {"id": 32, "score": 4.4, "title": "The Witcher", "type": "TV Show", "year": 2019, "genre": "Fantasy"}, {"id": 33, "score": 4.3, "title": "The American Barbecue Showdown", "type": "TV Show", "year": 2020, "genre": "Food & Beverage"}, {"id": 34, "score": 4.2, "title": "The Impossible", "type": "Movie", "year": 2002, "genre": "Disaster"}, {"id": 35, "score": 4.1, "title": "The Blacklist", "type": "TV Show", "year": 2014, "genre": "Drama"}, {"id": 36, "score": 4.0, "title": "Midnight at the Magnolia", "type": "Movie", "year": 2020, "genre": "Romantic"}, {"id": 37, "score": 3.9, "title": "The Girl in the Layby", "type": "Movie", "year": 2022, "genre": "Thriller"}, {"id": 38, "score": 3.8, "title": "60 Days In TV Show", "type": "Yes", "year": 2020, "genre": "Documentary"}, {"id": 39, "score": 3.7, "title": "The Take", "type": "Movie", "year": 2020, "genre": "Thriller"}, {"id": 40, "score": 3.6, "title": "Munich: The Edge of War", "type": "Movie", "year": 2022, "genre": "Thriller"}, {"id": 41, "score": 3.5, "title": "Link Master", "type": "TV Show", "year": 2014, "genre": "Comedy"}, {"id": 42, "score": 3.4, "title": "Community", "type": "TV Show", "year": 2020, "genre": "Comedy"}, {"id": 43, "score": 3.3, "title": "Home Game", "type": "TV Show", "year": 2020, "genre": "Sports"}]
```

## Query 4

Display details of 100 show/movies from a sample of 10% of the above dataset

```
sample.netflix = sample netflix 0.1;

limit_sample = limit sample.netflix 100;
```



```
hduser@rudra:~
```

```
2022-11-19 19:53:00,330 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
3.3.3 0.16.0 hduser 2022-11-19 19:52:59 2022-11-19 19:53:00 FILTER
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A
job_local189176324_0006 1 0 n/a n/a n/a 0 0 0 netflix,sample.netflix MAP_ONLY file:/tmp/temp-1183967787/tmp1051952090,
```

Input(s): Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"

Output(s): Successfully stored 734 records in: "file:/tmp/temp-1183967787/tmp1051952090"

Counters:

- Total records written : 734
- Total bytes written : 0
- Spilled Memory Manager spill count : 0
- Total bytes proactively spilled: 0
- Total records proactively spilled: 0

Job DAG:

```
job_local189176324_0006
```

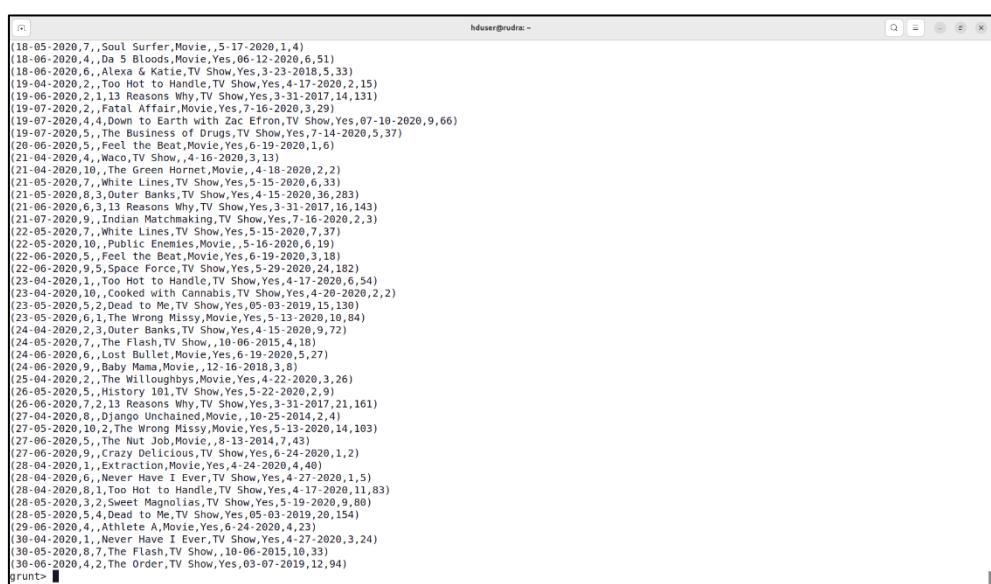
```
2022-11-19 19:53:00,331 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 19:53:00,332 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 19:53:00,333 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 19:53:00,334 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE CONVERSION FAILED 3968 time(s).
2022-11-19 19:53:00,334 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

```
2022-11-19 19:53:00,335 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checks
2022-11-19 19:53:00,335 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-11-19 19:53:00,335 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 19:53:00,344 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 19:53:00,344 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
[01-04-2020 3 3 All American TV Show, 3-28-2019,9,76]
```



```
hduser@rudra:~
```

```
2022-11-19 19:53:00,335 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-11-19 19:53:00,335 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 19:53:00,344 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 19:53:00,344 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
[01-04-2020 3 3 All American TV Show, 3-28-2019,9,76]
```



```
hduser@rudra:~
```

```
(18-05-2020,7,,Soul Surfer,Movie,,5-17-2020,1,4)
(18-06-2020,4,,Da 5 Bloods,Movie,Yes,06-12-2020,6,51)
(18-06-2020,6,,Alexa & Katie,TV Show,Yes,3-23-2018,5,33)
(19-04-2020,2,,Too Hot to Handle,TV Show,Yes,4-17-2020,2,15)
(19-06-2020,2,1,13 Reasons Why,TV Show,Yes,3-31-2017,14,131)
(19-07-2020,2,,Fatal Affair,Movie,Yes,7-16-2020,3,29)
(19-08-2020,2,,The Good Fight,Zooey Deschanel,TV Show,Yes,8-10-2020,9,66)
(19-07-2020,5,,The Business of Drugs,TV Show,Yes,7-14-2020,5,37)
(20-06-2020,5,,Feel the Beat,Movie,Yes,6-19-2020,1,6)
(21-04-2020,4,,Waco,TV Show,,4-16-2020,3,13)
(21-04-2020,10,,The Green Hornet,Movie,,4-18-2020,2,2)
(21-05-2020,7,,White Lines,TV Show,Yes,5-15-2020,6,33)
(21-05-2020,8,3,Outer Banks,TV Show,Yes,4-15-2020,36,283)
(21-06-2020,6,3,13 Reasons Why,TV Show,Yes,3-31-2017,16,143)
(21-07-2020,5,2,The White Lotus,TV Show,Yes,8-03-2020,2,3)
(22-05-2020,7,,White Lines,TV Show,Yes,5-15-2020,7,37)
(22-05-2020,10,,Public Enemies,Movie,,5-16-2020,6,19)
(22-06-2020,5,,Feel the Beat,Movie,Yes,6-19-2020,3,18)
(22-06-2020,9,5,,Space Force,TV Show,Yes,5-29-2020,24,182)
(23-04-2020,1,,Too Hot to Handle,TV Show,Yes,4-17-2020,6,54)
(23-04-2020,10,,Cooked with Cannabis,TV Show,Yes,4-28-2020,2,2)
(23-05-2020,5,2,Dead to Me,TV Show,Yes,05-03-2019,15,130)
(23-05-2020,5,2,The Morning Show,TV Show,Yes,4-13-2020,10,84)
(24-04-2020,2,3,Outer Banks,TV Show,Yes,4-15-2020,9,72)
(24-05-2020,7,,The Flash,TV Show,,10-06-2015,4,18)
(24-06-2020,6,,Lost Bullet,Movie,Yes,6-19-2020,5,27)
(24-06-2020,9,,Baby Mama,Movie,,12-16-2018,3,8)
(25-04-2020,2,,The Willoughbys,Movie,Yes,4-22-2020,3,26)
(26-05-2020,5,,History 101,TV Show,Yes,5-22-2020,2,9)
(26-06-2020,7,2,13 Reasons Why,TV Show,Yes,3-31-2017,21,161)
(27-04-2020,8,,Django Unchained,Movie,,10-25-2012,4,4)
(27-04-2020,9,,The Nut Job,TV Show,Yes,4-17-2020,20,163)
(27-06-2020,5,2,Dead to Me,TV Show,Yes,05-03-2019,20,163)
(27-06-2020,9,,Crazy Delicious,TV Show,Yes,6-24-2020,1,2)
(28-04-2020,1,,Extraction,Movie,Yes,4-24-2020,4,40)
(28-04-2020,6,,Never Have I Ever,TV Show,Yes,4-27-2020,1,5)
(28-04-2020,8,1,Too Hot to Handle,TV Show,Yes,4-17-2020,11,83)
(28-05-2020,3,2,,Sweet Magnolias,TV Show,Yes,5-19-2020,9,88)
(28-05-2020,5,4,Dead to Me,TV Show,Yes,05-03-2019,20,154)
(29-04-2020,1,,Athlete,Movie,Yes,6-24-2020,2,2)
(30-04-2020,1,,Never Have I Ever,TV Show,Yes,4-27-2020,3,24)
(30-05-2020,8,7,The Flash,TV Show,,10-06-2015,16,33)
(30-06-2020,4,2,The Order,TV Show,Yes,03-07-2019,12,94)
[01-04-2020 3 3 All American TV Show, 3-28-2019,9,76]
```

## Query 5 Display the details of movies only

```
movies.netflix = filter netflix by LOWER(type)=='movie';
```

The image shows two terminal windows side-by-side. Both windows have a title bar 'hduser@brudra: ~'. The left window contains the command 'grunt> movies.netflix = filter netflix by LOWER(type)=='movie'' followed by 'grunt>'. The right window shows the output of the command, displaying a large list of movie entries. Each entry consists of a unique identifier (e.g., 02-03-2022,4), the title ('Tyler Perry's A Madea Homec...'), the type ('Movie'), the release date ('Yes, 2-25-2022'), and the ID ('5,44'). There are approximately 20 such entries listed.

```
(02-03-2022,4,,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,5,44)
(02-03-2022,6,,Shrek 2,Movie,,03-01-2022,1,5)
(02-03-2022,7,,Battleship,Movie,,03-01-2022,1,4)
(02-03-2022,8,,Just Like Heaven,Movie,,03-01-2022,1,3)
(02-03-2022,10,,Shrek 2,Movie,,03-01-2022,1,18)
(03-03-2022,1,,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,6,51)
(03-03-2022,7,,Shrek 2,Movie,,03-01-2022,2,9)
(03-03-2022,8,,Just Like Heaven,Movie,,03-01-2022,2,6)
(03-03-2022,9,,Against the Ice,Movie,Yes,03-02-2022,1,2)
(03-03-2022,10,,Battleship,Movie,,03-01-2022,1,2)
(04-03-2022,4,,The Weekend Away,Movie,Yes,03-03-2022,1,7)
(04-03-2022,5,,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,7,57)
(04-03-2022,6,,Shrek 2,Movie,,03-01-2022,3,1)
(04-03-2022,9,,Against the Ice,Movie,Yes,03-02-2022,2,4)
(04-03-2022,10,,Battleship,Movie,,03-01-2022,3,6)
(05-03-2022,6,,The Weekend Away,Movie,Yes,03-03-2022,2,12)
(05-03-2022,7,1,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,8,61)
(05-03-2022,8,,Shrek 2,Movie,,03-01-2022,4,16)
(05-03-2022,10,,Shrek,Movie,,05-01-2018,1,1)
(06-03-2022,6,,The Weekend Away,Movie,Yes,03-03-2022,3,17)
(06-03-2022,7,,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,9,65)
(06-03-2022,8,,Shrek 2,Movie,,03-01-2022,5,19)
(06-03-2022,9,,Shrek,Movie,,05-01-2018,2,3)
(06-03-2022,10,,Just Like Heaven,Movie,,03-01-2022,3,7)
(07-03-2022,6,,The Weekend Away,Movie,Yes,03-03-2022,4,22)
(07-03-2022,7,2,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,10,69)
(07-03-2022,8,,Shrek 2,Movie,,03-01-2018,6,22)
(07-03-2022,9,,Shrek,Movie,,05-01-2018,3,5)
(07-03-2022,10,,Just Like Heaven,Movie,,03-01-2022,4,8)
(08-03-2022,6,,The Weekend Away,Movie,Yes,03-03-2022,5,26)
(08-03-2022,8,,Shrek 2,Movie,,03-01-2022,7,25)
(08-03-2022,9,,Shrek,Movie,,05-01-2018,4,7)
(08-03-2022,10,3,Tyler Perry's A Madea Homec...,Movie,Yes,2-25-2022,11,70)
(09-03-2022,7,6,Shrek 2,Movie,,03-01-2022,8,29)
(09-03-2022,8,,Shrek,Movie,,05-01-2018,5,10)
(09-03-2022,9,,Shrek 2,Movie,,03-01-2022,6,28)
(09-03-2022,10,,Shooter,Movie,,08-01-2014,1,1)
(10-03-2022,6,,Shooter,Movie,,08-01-2014,2,4)
(10-03-2022,9,7,Shrek 2,Movie,,03-01-2022,9,31)
(10-03-2022,10,,Shrek,Movie,,05-01-2018,6,11)
(11-03-2022,8,,Shooter,Movie,,08-01-2014,3,4)
(11-03-2022,9,7,Shrek 2,Movie,,03-01-2022,10,33)
(11-03-2022,10,,Shrek,Movie,,05-01-2018,7,12)
grunt> █
```

## Query 6 Display the details of movies having viewership score more than 75

```
viewer.netflix = filter netflix by score>75;
```

```

hduser@rudra: ~
grunt> viewer netflix = filter netflix by score>75;
grunt>

hduser@rudra: ~
(28-02-2022,4,2,Love is Blind,TV Show,Yes,2-13-2020,35,239)
(28-02-2022,6,4,Sweet Magnolias,TV Show,Yes,5-19-2020,43,337)
(28-02-2022,7,7,Ozark,TV Show,Yes,7-21-2017,88,532)
(28-02-2022,9,,Despicable Me 2,Movie,,02-01-2022,26,85)
(01-03-2022,2,1,Inventing Anna,TV Show,Yes,02-11-2022,18,173)
(01-03-2022,4,2,Love is Blind,TV Show,Yes,2-13-2020,36,246)
(01-03-2022,6,3,Sweet Magnolias,TV Show,Yes,5-19-2020,44,342)
(01-03-2022,7,8,Ozark,TV Show,Yes,7-21-2017,88,536)
(01-03-2022,9,1,Despicable Me 2,Movie,,02-01-2022,26,88)
(01-03-2022,9,2,Love is Blind,TV Show,Yes,5-19-2020,44,347)
(02-03-2022,3,1,Inventing Anna,TV Show,Yes,02-11-2022,19,181)
(02-03-2022,5,2,Love is Blind,TV Show,Yes,2-13-2020,37,252)
(02-03-2022,9,3,Sweet Magnolias,TV Show,Yes,5-19-2020,45,344)
(03-03-2022,3,1,Inventing Anna,TV Show,Yes,02-11-2022,20,189)
(03-03-2022,5,2,Love is Blind,TV Show,Yes,2-13-2020,38,258)
(04-03-2022,5,3,Inventing Anna,TV Show,Yes,02-11-2022,21,207)
(04-03-2022,6,2,Love is Blind,TV Show,Yes,2-13-2020,36,263)
(05-03-2022,1,2,Love is Blind,TV Show,Yes,2-13-2020,49,273)
(05-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,22,204)
(06-03-2022,3,4,Love is Blind,TV Show,Yes,2-13-2020,41,281)
(06-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,23,211)
(06-03-2022,5,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,9,76)
(07-03-2022,3,3,Inventing Anna,TV Show,Yes,02-11-2022,24,219)
(07-03-2022,4,4,Love is Blind,TV Show,Yes,2-13-2020,42,288)
(07-03-2022,5,4,Good Girls,TV Show,01-01-2019,23,182)
(08-03-2022,2,2,Inventing Anna,TV Show,Yes,02-11-2022,25,228)
(08-03-2022,4,4,Love is Blind,TV Show,Yes,2-13-2020,43,295)
(08-03-2022,6,1,Vikings: Valhalla,TV Show,01-01-2019,23,124)
(08-03-2022,6,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,11,87)
(09-03-2022,2,2,Good Girls,TV Show,01-01-2019,24,133)
(09-03-2022,3,3,Inventing Anna,TV Show,Yes,02-11-2022,26,236)
(09-03-2022,5,5,Love is Blind,TV Show,Yes,2-13-2020,44,361)
(09-03-2022,6,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,12,92)
(10-03-2022,3,3,Good Girls,TV Show,01-01-2019,25,141)
(10-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,27,243)
(10-03-2022,5,1,Worst Roommate Ever,TV Show,Yes,03-01-2022,9,76)
(10-03-2022,6,5,Worst Roommate Ever,TV Show,Yes,2-13-2020,45,306)
(10-03-2022,7,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,13,96)
(11-03-2022,3,,Good Girls,TV Show,01-01-2019,26,149)
(11-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,28,259)
(11-03-2022,5,3,Love is Blind,TV Show,Yes,2-13-2020,46,313)
(11-03-2022,6,1,Worst Roommate Ever,TV Show,Yes,03-01-2022,10,81)
(11-03-2022,7,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,14,108)
grunt>
```

**Query 7** Display the title and type of the shows/movies present in Netflix Exclusive and being in top 10 for more than 7 times (and with two conditions)

```
topexclusive = foreach (filter netflix by excl=='Yes' and
topdays==7) generate title, type;
```

```

hduser@rudra: ~
grunt> topexclusive = foreach (filter netflix by excl=='Yes' and topdays==7) generate title, type;
grunt> ■

hduser@rudra: ~
(Maya and the Three,TV Show)
(Army of Thieves,Movie)
(The Harder They Fall,Movie)
(Catching Killers,TV Show)
(Love Hard,Movie)
(Narcos: Mexico,TV Show)
(Red Notice,Movie)
(The Queen of Flow,TV Show)
(Arcane,TV Show)
(Cowboy Bebop,TV Show)
(Hellbound,TV Show)
(True Story,TV Show)
(Bruised,Movie)
(A Boy Called Christmas,Movie)
(A Girl Like Her,Movie)
(Lost in Space,TV Show)
(The Power of the Dog,Movie)
(The Unforgivable,Movie)
(Back to the Outback,Movie)
(The Witcher,TV Show)
(TwentySomethings: Austin,TV Show)
(Selling Tampa,TV Show)
(Don't Look Now,Movie)
(The Silence,Sea,TV Show)
(Stay Close,TV Show)
(Cheer,TV Show)
(Brazen,Movie)
(Archive 81,TV Show)
(The Royal Treatment,Movie)
(All of Us Are Dead,Movie)
(The House in the House Acro.,TV Show)
(In From the Cold,TV Show)
(Home Team,Movie)
(Raising Dion,TV Show)
(The Tinder Swindler,Movie)
(Disenchantment,TV Show)
(Inventing Anna,TV Show)
(The Cuphead Show,TV Show)
(Jeepers... A Bloody Trilogy,TV Show)
(Vikings: Valhalla,TV Show)
(Tyler Perry's A Madea Home...,Movie)
(Worst Roommate Ever,TV Show)
(Pieces of Her,TV Show)
grunt>
```

**Query 8** Categorize the titles according to their type

```
group_titles = group netflix by type;
```

**Query 9** Display the count of the titles and the average of their viewership score type-wise

```
group_count = foreach group_titles generate group,  
COUNT(netflix), AVG(netflix.score);
```

```

hduser@rudra:~$ grunt> group count = foreach group titles generate group, COUNT(netflix), AVG(netflix.score);
hduser@rudra:~$ 

hduser@rudra:~$ success!
Job Stats (time in seconds)
JobId Maps Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime     MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime   A
alias    Feature Outputs
job_local347141730 0018 1          n/a           n/a           n/a           n/a           n/a           n/a           n/a           group_count,group_titles,netflix   GROUP_BY,COMBI
NER      file:/tmp/temp-1183967787/tmp1295143579,
Input(s):
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"
Output(s):
Successfully stored 4 records in: "file:/tmp/temp-1183967787/tmp1295143579"
Counters:
Total records written : 4
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local347141730_0018

2022-11-19 20:45:00,786 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 20:45:00,788 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 20:45:00,789 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 20:45:00,791 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION FAILED 3968 times(s).
2022-11-19 20:45:00,791 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-19 20:45:00,791 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checks
[...]
2022-11-19 20:45:00,792 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.adress
2022-11-19 20:45:00,792 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 20:45:00,792 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-19 20:45:00,801 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 20:45:00,801 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Movie,2611,38.948678667177326)
(TV Show,4446,172.8540269986862)
(Concert/Film...,2,12,0)
(Stand-up Comedy,41,38.58536585365854)
grunt> 

```

**Query 10** Convert each field of a tuple to an individual tuple such that those should be in a bag (ToBag)

```

bag.netflix = foreach netflix generate TOBAG(topdate,
dailyrank, lwrank, title, type, excl, rdate, topdays,
score);

```

```

hduser@rudra:~$ grunt> bag.netflix = foreach netflix generate TOBAG(topdate, dailyrank, lwrank, title, type, excl, rdate, topdays, score);
hduser@rudra:~$ 

hduser@rudra:~$ 
((07-03-2022),(8),(1),(Shrek 2),(Movie),(1),(03-01-2022),(6),(22)))
((07-03-2022),(9),(1),(Just Like Heaven),(Movie),(1),(05-01-2018),(3),(5))
((07-03-2022),(10),(1),(Pieces of Her),(TV Show),(Yes),(03-04-2022),(4),(8))
((08-03-2022),(1),(1),(Inventing Anna),(TV Show),(Yes),(02-11-2022),(25),(228))
((08-03-2022),(2),(2),(Worst Roommate Ever),(TV Show),(Yes),(03-01-2022),(7),(63))
((08-03-2022),(3),(1),(Love is Blind),(TV Show),(Yes),(2-13-2020),(43),(295))
((08-03-2022),(4),(1),(Good Girls),(TV Show),(1),(01-01-2019),(23),(124))
((08-03-2022),(5),(1),(Vikings: Valhalla),(TV Show),(Yes),(2-25-2022),(11),(87))
((08-03-2022),(6),(1),(The Weekend Away),(Movie),(Yes),(03-03-2022),(5),(26))
((08-03-2022),(8),(1),(Shrek 2),(Movie),(1),(03-01-2022),(7),(25))
((08-03-2022),(9),(1),(Shrek),(Movie),(1),(05-01-2018),(4),(7))
((08-03-2022),(10),(3),(Tyler Perry's A Madea Home...),(Movie),(Yes),(2-25-2022),(11),(70))
((09-03-2022),(1),(1),(Pieces of Her),(TV Show),(Yes),(03-04-2022),(5),(49))
((09-03-2022),(2),(1),(Love is Blind),(TV Show),(Yes),(01-01-2019),(24),(132))
((09-03-2022),(3),(1),(Inventing Anna),(TV Show),(Yes),(02-11-2022),(6),(236))
((09-03-2022),(4),(2),(Worst Roommate Ever),(TV Show),(Yes),(03-01-2022),(8),(70))
((09-03-2022),(5),(5),(Love is Blind),(TV Show),(Yes),(2-13-2020),(44),(301))
((09-03-2022),(6),(1),(Vikings: Valhalla),(TV Show),(Yes),(2-25-2022),(12),(92))
((09-03-2022),(7),(6),(Shrek 2),(Movie),(1),(03-01-2022),(8),(29))
((09-03-2022),(8),(1),(Shrek),(Movie),(1),(05-01-2018),(5),(10))
((09-03-2022),(9),(1),(The Weekend Away),(Movie),(Yes),(03-03-2022),(6),(28))
((09-03-2022),(10),(1),(Shooter),(Movie),(1),(08-01-2014),(2),(4))
((09-03-2022),(11),(1),(Love is Blind),(TV Show),(Yes),(03-01-2022),(6),(59))
((09-03-2022),(12),(1),(The Last Kingdom),(TV Show),(1),(01-01-2019),(25),(141))
((09-03-2022),(13),(1),(Good Girls),(TV Show),(1),(01-01-2019),(26),(149))
((09-03-2022),(14),(3),(Inventing Anna),(TV Show),(Yes),(02-11-2022),(27),(243))
((09-03-2022),(15),(1),(Worst Roommate Ever),(TV Show),(Yes),(03-01-2022),(9),(76))
((09-03-2022),(16),(5),(Love is Blind),(TV Show),(Yes),(2-13-2020),(45),(306))
((10-03-2022),(7),(2),(Vikings: Valhalla),(TV Show),(Yes),(2-25-2022),(13),(96))
((10-03-2022),(8),(1),(Shooter),(Movie),(1),(08-01-2014),(2),(4))
((10-03-2022),(9),(1),(Love is Blind),(TV Show),(Yes),(03-01-2022),(6),(31))
((10-03-2022),(10),(1),(Shrek),(Movie),(1),(05-01-2018),(6),(11))
((11-03-2022),(1),(1),(The Last Kingdom),(TV Show),(1),(07-10-2016),(13),(62))
((11-03-2022),(2),(1),(Pieces of Her),(TV Show),(Yes),(03-04-2022),(7),(68))
((11-03-2022),(3),(1),(Good Girls),(TV Show),(1),(01-01-2019),(26),(149))
((11-03-2022),(4),(3),(Inventing Anna),(TV Show),(Yes),(02-11-2022),(28),(250))
((11-03-2022),(5),(6),(Love is Blind),(TV Show),(Yes),(2-13-2020),(46),(312))
((11-03-2022),(6),(1),(Worst Roommate Ever),(TV Show),(Yes),(03-01-2022),(10),(81))
((11-03-2022),(7),(2),(Vikings: Valhalla),(TV Show),(Yes),(2-25-2022),(14),(100))
((11-03-2022),(8),(1),(Shooter),(Movie),(1),(08-01-2014),(3),(7))
((11-03-2022),(9),(7),(Shrek 2),(Movie),(1),(03-01-2022),(10),(33))
((11-03-2022),(10),(1),(Shrek),(Movie),(1),(05-01-2018),(7),(12))
((11-03-2022),(11),(1),(Shrek),(Movie),(1),(05-01-2018),(7),(12))

```

**Query 11** Display all the titles which starts with ‘The’ and contains ‘of’ within it.

```
the_titles = filter netflix by STARTSWITH(title, 'The') and  
title matches '.*of.*';
```

**Query 12** Display the titles, type and score of type TV Shows after changing it to TV Series

```
series.netflix = foreach netflix generate title,  
REPLACE(type, 'Show', 'Series'), score;
```

```

hduser@rudra: ~
grunt> series.netflix = foreach netflix generate title, REPLACE(type,'Show','Series'), score;
grunt> ■

hduser@rudra: ~
(Shrek 2,Movie,22)
(Shrek,Movie,5)
(Just Like Heaven,Movie,8)
(Pieces of Her,TV Series,39)
(Inventing Anna,TV Series,228)
(Worst Roommate Ever,TV Series,63)
(Love Is Blind,TV Series,295)
(Good Girls,TV Series,124)
(Vikings: Valhalla,TV Series,87)
(The Weekend Away,Movie,26)
(Shrek 2,Movie,25)
(Shrek,Movie,7)
(Tyler Perry's A Madea Homec...,Movie,78)
(Pieces of Her,TV Series,149)
(Good Girls,TV Series,133)
(Inventing Anna,TV Series,236)
(Worst Roommate Ever,TV Series,70)
(Love Is Blind,TV Series,301)
(Vikings: Valhalla,TV Series,92)
(Shrek 2,Movie,29)
(Shrek,Movie,10)
(The Weekend Away,Movie,28)
(Shrek 2,Movie,11)
(Pieces of Her,TV Series,59)
(The Last Kingdom,TV Series,52)
(Good Girls,TV Series,141)
(Inventing Anna,TV Series,243)
(Worst Roommate Ever,TV Series,76)
(Love Is Blind,TV Series,306)
(Vikings: Valhalla,TV Series,96)
(Shooter,Movie,4)
(Shrek 2,Movie,31)
(Shrek,Movie,11)
(The Last Kingdom,TV Series,62)
(Pieces of Her,TV Series,68)
(Good Girls,TV Series,149)
(Inventing Anna,TV Series,250)
(Love Is Blind,TV Series,312)
(Worst Roommate Ever,TV Series,81)
(Vikings: Valhalla,TV Series,100)
(Shooter,Movie,7)
(Shrek 2,Movie,33)
(Shrek,Movie,12)
grunt>

```

**Query 13** Generate four relations containing titles of each type TV Show, Movies, Concert and Stand-up Comedy (each in one relation) and store each one of them

```

split netflix into tvshows if(type=='TV Show'), movies
if(type=='Movie'), comedy if(type matches '.*Comedy.*'),
concert if (type matches '.*Concert.*');

store tvshows into 'TV Shows' using PigStorage(',');
store movies into 'Movies' using PigStorage(',');
store comedy into 'Comedy' using PigStorage(',');
store concert into 'Concert' using PigStorage(',');

```

```
hduser@rudra:~  
grunt> split netflix into tvshows if(type=='TV Show'), movies if(type=='Movie'), comedy if(type matches '.*Comedy.*'), concert if (type matches '.*Concert.*')  
;  
grunt>
```

```
hduser@rudra:~  
2022-11-19 21:46:36,446 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases netflix,tvshows  
2022-11-19 21:46:36,447 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: netflix[1,10],net  
flix[1,-1],tvshows[3,19] C: R:  
2022-11-19 21:46:36,455 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2022-11-19 21:46:36,460 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2022-11-19 21:46:36,461 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2022-11-19 21:46:36,468 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete  
2022-11-19 21:46:36,468 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:  
  
HadoopVersion PigVersion UserId StartedAt FinishedAt Features  
3.3.3 0.16.0 hduser 2022-11-19 21:46:35 2022-11-19 21:46:36 UNKNOWN  
  
Success!  
  
Job Stats (time in seconds):  
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A  
job_local664818409_0007 1 0 n/a n/a n/a 0 0 0 0 netflix,tvshows MAP_ONLY file:///home/hduser/TV  
Shows,  
  
Input(s):  
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"  
  
Output(s):  
Successfully stored 4446 records in: "file:///home/hduser/TV Shows"  
  
Counters:  
Total records written : 4446  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local664818409_0007  
  
2022-11-19 21:46:36,473 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2022-11-19 21:46:36,475 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2022-11-19 21:46:36,477 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2022-11-19 21:46:36,479 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE  
CONVERSION FAILED 3968 time(s).  
2022-11-19 21:46:36,480 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
grunt> ■
```

```
hduser@rudra:~  
(04-03-2022,8,3,jeen-yuh: A Kanye Trilogy,TV Show,Yes,2-16-2022,11,56)  
(05-03-2022,1,2,Love is Blind,TV Show,Yes,2-13-2020,40,273)  
(05-03-2022,2,,Pieces of Her,TV Show,Yes,03-04-2022,1,9)  
(05-03-2022,3,,Worst Roommate Ever,TV Show,Yes,03-01-2022,4,37)  
(05-03-2022,4,,Inventing Anna,TV Show,Yes,02-11-2022,25,204)  
(05-03-2022,5,4,Vikings: Valhalla,TV Show,Yes,2-25-2022,6,70)  
(05-03-2022,9,5,jeen-yuh: A Kanye Trilogy,TV Show,Yes,2-16-2022,12,58)  
(06-03-2022,1,,Pieces of Her,TV Show,Yes,03-04-2022,2,19)  
(06-03-2022,2,,Worst Roommate Ever,TV Show,Yes,03-01-2022,5,46)  
(06-03-2022,3,4,Love is Blind,TV Show,Yes,2-13-2020,41,281)  
(06-03-2022,3,4,Inventing Anna,TV Show,Yes,02-11-2022,23,211)  
(06-03-2022,5,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,9,76)  
(07-03-2022,1,,Placing Bets,TV Show,Yes,04-04-2022,3,29)  
(07-03-2022,2,,Worst Roommate Ever,TV Show,Yes,02-11-2022,6,55)  
(07-03-2022,3,3,Inventing Anna,TV Show,Yes,02-11-2022,24,219)  
(07-03-2022,4,4,Love is Blind,TV Show,Yes,2-13-2020,42,288)  
(07-03-2022,5,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,10,82)  
(08-03-2022,1,,Pieces of Her,TV Show,Yes,03-04-2022,4,39)  
(08-03-2022,2,2,Inventing Anna,TV Show,Yes,02-11-2022,25,228)  
(08-03-2022,3,,Worst Roommate Ever,TV Show,Yes,03-01-2022,7,63)  
(08-03-2022,4,4,Love is Blind,TV Show,Yes,2-13-2020,43,295)  
(09-03-2022,1,1,Good Girls,TV Show,Yes,02-11-2022,27,243)  
(09-03-2022,6,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,11,87)  
(09-03-2022,1,,Pieces of Her,TV Show,Yes,03-04-2022,5,49)  
(09-03-2022,2,,Good Girls,TV Show,,01-01-2019,24,133)  
(09-03-2022,3,3,Inventing Anna,TV Show,Yes,02-11-2022,26,236)  
(09-03-2022,4,2,Worst Roommate Ever,TV Show,Yes,03-01-2022,8,70)  
(09-03-2022,5,5,Love is Blind,TV Show,Yes,2-13-2020,44,301)  
(09-03-2022,6,1,Vikings: Valhalla,TV Show,Yes,2-25-2022,12,92)  
(10-03-2022,1,,Placing Bets,TV Show,Yes,04-04-2022,6,59)  
(10-03-2022,3,,The Last Kingdom,TV Show,Yes,07-10-2016,12,52)  
(10-03-2022,3,,Good Girls,TV Show,,01-01-2019,25,141)  
(10-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,27,243)  
(10-03-2022,5,1,Worst Roommate Ever,TV Show,Yes,03-01-2022,9,76)  
(10-03-2022,6,5,Love is Blind,TV Show,Yes,2-13-2020,45,306)  
(10-03-2022,7,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,13,96)  
(11-03-2022,1,,The Last Kingdom,TV Show,,07-10-2016,13,62)  
(11-03-2022,2,1,Good Girls,TV Show,Yes,02-11-2022,26,78)  
(11-03-2022,3,,Good Girls,TV Show,,01-01-2019,26,149)  
(11-03-2022,4,3,Inventing Anna,TV Show,Yes,02-11-2022,28,250)  
(11-03-2022,5,6,Love is Blind,TV Show,Yes,2-13-2020,46,312)  
(11-03-2022,6,1,Worst Roommate Ever,TV Show,Yes,03-01-2022,10,81)  
(11-03-2022,7,2,Vikings: Valhalla,TV Show,Yes,2-25-2022,14,100)  
grunt>
```

```

hduser@rudra: ~
2022-11-19 21:47:33,800 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases movies,netflix
2022-11-19 21:47:33,800 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: netflix[1,10].net
flix[-1,-1],movies[3,10],C: R
2022-11-19 21:47:33,806 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - JobTracker metrics system already initialized!
2022-11-19 21:47:33,810 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:47:33,816 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:47:33,828 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-11-19 21:47:33,828 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
3.3.0 0.16.0 hduser 2022-11-19 21:47:33 2022-11-19 21:47:33 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A
alias Feature Outputs job_local1596611455_0008 1 0 n/a n/a n/a 0 0 0 0 movies,netflix MAP_ONLY file:///home/hduser/Movies,

Input(s):
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"

Output(s):
Successfully stored 2611 records in: "file:///home/hduser/Movies"

Counters:
Total records written : 2611
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1596611455_0008

2022-11-19 21:47:33,830 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:47:33,832 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:47:33,834 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:47:33,836 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION FAILED 3968 times(s).
2022-11-19 21:47:33,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>

```

```

hduser@rudra: ~
(02-03-2022,4,,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,5,44)
(02-03-2022,6,,Shrek The Third,Movie,Yes,03-01-2022,1,5)
(02-03-2022,7,,Just Like Heaven,Movie,03-01-2022,1,4)
(02-03-2022,8,,Just Like Heaven,Movie,03-01-2022,1,3)
(02-03-2022,10,,Restless,Movie,Yes,2-25-2022,4,18)
(03-03-2022,4,,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,6,51)
(03-03-2022,7,,Shrek 2,Movie,,03-01-2022,2,9)
(03-03-2022,8,,Just Like Heaven,Movie,,03-01-2022,2,6)
(03-03-2022,9,,Against the Ice,Movie,Yes,03-02-2022,1,2)
(03-03-2022,10,,Battleship,Movie,,03-01-2022,2,5)
(04-03-2022,4,,Shrek The Third,Movie,Yes,03-03-2022,1,7)
(04-03-2022,5,,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,7,57)
(04-03-2022,7,,Shrek 2,Movie,,03-01-2022,3,13)
(04-03-2022,9,,Against the Ice,Movie,Yes,03-02-2022,2,4)
(04-03-2022,10,,Battleship,Movie,,03-01-2022,3,6)
(05-03-2022,6,,The Weekend Away,Movie,Yes,03-03-2022,2,12)
(05-03-2022,7,1,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,8,61)
(05-03-2022,8,,Shrek 2,Movie,,03-01-2022,4,16)
(05-03-2022,10,,Shrek 2,Movie,,03-01-2022,1,1)
(06-03-2022,7,,The Weekend Away,Movie,Yes,03-03-2022,3,17)
(06-03-2022,7,1,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,9,65)
(06-03-2022,8,,Shrek 2,Movie,,03-01-2022,5,19)
(06-03-2022,9,,Shrek,Movie,,05-01-2018,2,3)
(06-03-2022,10,,Just Like Heaven,Movie,,03-01-2022,3,7)
(07-03-2022,6,,The Weekend Away,Movie,Yes,03-03-2022,4,22)
(07-03-2022,7,2,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,10,69)
(07-03-2022,8,,Shrek 2,Movie,,03-01-2022,4,22)
(07-03-2022,9,,Shrek,Movie,,05-01-2018,3,5)
(07-03-2022,10,,Just Like Heaven,Movie,,03-01-2022,4,8)
(08-03-2022,7,,The Weekend Away,Movie,Yes,03-03-2022,5,26)
(08-03-2022,8,,Shrek 2,Movie,,03-01-2022,7,25)
(08-03-2022,9,,Shrek,Movie,,05-01-2018,4,7)
(08-03-2022,10,3,Tyler Perry's A Madea Home...,Movie,Yes,2-25-2022,11,78)
(09-03-2022,7,6,Shrek 2,Movie,,03-01-2022,8,29)
(09-03-2022,8,,Shrek 2,Movie,,03-01-2022,5,28)
(09-03-2022,9,,The Weekend Away,Movie,Yes,03-03-2022,6,28)
(09-03-2022,10,,Shooter,Movie,,09-01-2014,1,1)
(10-03-2022,8,,Shooter,Movie,,08-01-2014,2,4)
(10-03-2022,9,7,Shrek 2,Movie,,03-01-2022,9,31)
(10-03-2022,10,,Shrek,Movie,,05-01-2018,6,11)
(11-03-2022,8,,Shooter,Movie,,08-01-2014,3,7)
(11-03-2022,9,7,Shrek 2,Movie,,03-01-2022,10,33)
(11-03-2022,10,,Shrek,Movie,,05-01-2018,7,12)
grunt>

```

```

hduser@rudra: ~
2022-11-19 21:51:32,757 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases comedy,netflix
2022-11-19 21:51:32,757 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: netflix[1,10].net
flix[-1,-1],comedy[3,74] C: R
2022-11-19 21:51:32,764 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:51:32,766 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:51:32,771 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:51:32,773 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-11-19 21:51:32,773 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
3.3.0 0.16.0 hduser 2022-11-19 21:51:32 2022-11-19 21:51:32 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A
alias Feature Outputs job_local1085400113_0011 1 0 n/a n/a n/a 0 0 0 0 comedy,netflix MAP_ONLY file:///home/hduser/Comedy,

Input(s):
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"

Output(s):
Successfully stored 41 records in: "file:///home/hduser/Comedy"

Counters:
Total records written : 41
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1085400113_0011

2022-11-19 21:51:32,774 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:51:32,775 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:51:32,776 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:51:32,777 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION FAILED 3968 times(s).
2022-11-19 21:51:32,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>

```

```
hduser@rudra:~
2022-11-19 21:44:37,400 [main] INFO org.apache.hadoop.mapred.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 21:44:37,400 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(15-04-2020_4,_Chris D'Elia: No Pain,Stand-Up Comedy,Yes,4-14-2020,2,14)
(17-04-2020_9,_Chris D'Elia: No Pain,Stand-Up Comedy,Yes,4-14-2020,3,16)
(06-05-2020_6,_Jerry Seinfeld: 23 Hours to.,Stand-Up Comedy,Yes,05-05-2020,1,5)
(07-05-2020_3,_Jerry Seinfeld: 23 Hours to.,Stand-Up Comedy,Yes,05-05-2020,2,13)
(08-05-2020_5,_Jerry Seinfeld: 23 Hours to.,Stand-Up Comedy,Yes,05-05-2020,3,19)
(09-05-2020_8,_Jerry Seinfeld: 23 Hours to.,Stand-Up Comedy,Yes,05-05-2020,4,22)
(10-05-2020_10,_Jerry Seinfeld: 23 Hours to.,Stand-Up Comedy,Yes,05-05-2020,5,23)
(01-07-2020_3,_George Lopez: We'll Do It To.,Stand-Up Comedy,Yes,6-30-2020,1,8)
(02-07-2020_4,_George Lopez: We'll Do It To.,Stand-Up Comedy,Yes,6-30-2020,2,15)
(03-07-2020_7,_George Lopez: We'll Do It To.,Stand-Up Comedy,Yes,6-30-2020,3,19)
(04-07-2020_8,_George Lopez: We'll Do It To.,Stand-Up Comedy,Yes,6-30-2020,4,22)
(05-07-2020_9,_George Lopez: We'll Do It To.,Stand-Up Comedy,Yes,6-30-2020,5,24)
(18-11-2020_2,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,1,9)
(19-11-2020_3,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,2,18)
(20-11-2020_3,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,3,26)
(21-11-2020_3,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,4,34)
(22-11-2020_2,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,5,43)
(23-11-2020_2,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,6,52)
(24-11-2020_3,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,7,60)
(25-11-2020_5,2_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,8,66)
(26-11-2020_7,2_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,9,70)
(27-11-2020_10,_Kevin Hart: Zero F***ks Given,Stand-Up Comedy,Yes,11-17-2020,10,73)
(06-06-2021_10,_Bo Burnham: Inside,Stand-Up Comedy,Yes,5-30-2021,1,1)
(10-06-2021_10,_Bo Burnham: Inside,Stand-Up Comedy,Yes,5-30-2021,2,2)
(07-10-2021_4,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,1,7)
(08-10-2021_4,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,2,14)
(09-10-2021_3,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,3,22)
(10-10-2021_3,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,4,30)
(11-10-2021_3,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,5,38)
(12-10-2021_3,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,6,46)
(13-10-2021_3,_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,7,54)
(14-10-2021_4,4_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,8,61)
(15-10-2021_4,4,2_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,9,68)
(16-10-2021_5,3_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,10,74)
(17-10-2021_6,3_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,11,79)
(18-10-2021_8,3_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,12,82)
(19-10-2021_9,3_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,13,84)
(21-10-2021_10,4_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,14,85)
(22-10-2021_9,4_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,15,87)
(23-10-2021_10,5_Dave Chappelle: The Closer,Stand-Up Comedy,Yes,10-05-2021,16,88)
[23-run]>
```

```
hduser@rudra:~
2022-11-19 21:52:03,070 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases concert,netflix
2022-11-19 21:52:03,070 [main] INFO org.apache.pig.backend.hadoop.executionengine.MapReduceLauncher - detailed locations: M: netflix[1,10].net
file[x-1,1],concert[3,112] C: R
2022-11-19 21:52:03,073 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:52:03,075 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:52:03,076 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:52:03,077 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-11-19 21:52:03,078 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
3.3.3 0.16.0 hduser 2022-11-19 21:52:02 2022-11-19 21:52:03 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A
list Feature Outputs job_local1217761740_0012 1 0 n/a n/a n/a n/a 0 0 0 concert,netflix MAP_ONLY file:///home/hduser/Concert,
hduser@Concert,
Inputs:
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"
Outputs:
Successfully stored 2 records in: "file:///home/hduser/Concert"

Counters:
Total records written : 2
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1217761740_0012

2022-11-19 21:52:03,080 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:52:03,082 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:52:03,084 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:52:03,086 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE CONVERSION FAILED 3968 time(s).
2022-11-19 21:52:03,086 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> ■
```

```
hduser@rudra:~
3.3.3 0.16.0 hduser 2022-11-19 21:44:50 2022-11-19 21:44:50 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A
list Feature Outputs job_local1237122132_0006 1 0 n/a n/a n/a n/a 0 0 0 concert,netflix MAP_ONLY file:/tmp/temp-215458488/tmp-1987279998,
hduser@Concert,
Inputs:
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"
Outputs:
Successfully stored 2 records in: "file:/tmp/temp-215458488/tmp-1987279998"

Counters:
Total records written : 2
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1237122132_0006

2022-11-19 21:44:50,696 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:44:50,697 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:44:50,699 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 21:44:50,700 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE CONVERS_ FAILER 3968 time(s).
2022-11-19 21:44:50,700 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-19 21:44:50,701 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-19 21:44:50,701 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.adress
2022-11-19 21:44:50,702 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 21:44:50,702 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-19 21:44:50,703 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 21:44:50,712 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(22-12-2020_2,_Ariana Grande: Excuse Me; I_,Concert/Perf_-,Yes,12-21-2020,1,9)
(23-12-2020_5,_Ariana Grande: Excuse Me; I_,Concert/Perf_-,Yes,12-21-2020,2,15)
grunt> ■
```

**Query 14** Group the relations of TV Shows and Movies on the basis of if these are present in Netflix Exclusive or not, and show the count of each

```

movies_group = group movies by excl;

tvshows_group = group tvshows by excl;

movies_exl_count = foreach movies_group generate group,
COUNT(movies);

tv_exl_count = foreach tvshows_group generate group,
COUNT(tvshows);

```

```

hduser@rudra: ~
grunt> movies = group movies by excl;
2022-11-19 22:33:07,051 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 22:33:07,051 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> tvshows = group tvshows by excl;
2022-11-19 22:33:07,886 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 22:33:07,886 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> movies_exl_count = foreach movies_group generate group, COUNT(movies);
2022-11-19 22:33:07,922 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 22:33:07,922 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> tv_exl_count = foreach tvshows_group generate group, COUNT(tvshows);
2022-11-19 22:33:09,502 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 22:33:09,502 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt>

```

```

hduser@rudra: ~
3.3.3 0.16.0 hduser 2022-11-19 22:34:40 2022-11-19 22:34:40 GROUP_BY
Success!
Job Stats (time in seconds):
JobID Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A
job_local1394555963_0020 1 1 n/a n/a n/a n/a n/a n/a n/a movies,movies_exl_count,movies_group,netflix f
ile:/tmp/temp-215458488/tmp584021593,
Input(s):
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"
Output(s):
Successfully stored 2 records in: "file:/tmp/temp-215458488/tmp584021593"
Counters:
Total records written : 2
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1394555963_0020

2022-11-19 22:34:40,870 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 22:34:40,872 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 22:34:40,876 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 22:34:40,879 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 3968 time(s).
2022-11-19 22:34:40,879 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-19 22:34:40,880 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-19 22:34:40,880 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-11-19 22:34:40,880 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 22:34:40,893 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-19 22:34:40,893 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 22:34:40,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Yield:1469)
(1,1142)
grunt>

```

**Query 15** Rank the titles on the basis of Viewership Score and display their rank, titles and type

```

ranked_netflix = rank netflix by score;

ranked_titles = foreach ranked_netflix generate $0, title,
type;

```

```

hduser@rudra: ~
grunt> ranked.netflix = rank.netflix by score;
2022-11-19 22:41:41,352 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 2 time(s).
2022-11-19 22:41:43,352 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> ranked.titles = foreach ranked.netflix generate $0, title, type;
2022-11-19 22:41:44,310 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 2 time(s).
2022-11-19 22:41:44,310 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> ■

```

```

hduser@rudra: ~
(7058,Cocomelon,TV Show)
(7059,Cocomelon,TV Show)
(7060,Cocomelon,TV Show)
(7061,Cocomelon,TV Show)
(7062,Cocomelon,TV Show)
(7063,Cocomelon,TV Show)
(7064,Cocomelon,TV Show)
(7065,Cocomelon,TV Show)
(7066,Cocomelon,TV Show)
(7067,Cocomelon,TV Show)
(7068,Cocomelon,TV Show)
(7069,Cocomelon,TV Show)
(7070,Cocomelon,TV Show)
(7071,Cocomelon,TV Show)
(7072,Cocomelon,TV Show)
(7073,Cocomelon,TV Show)
(7074,Cocomelon,TV Show)
(7075,Cocomelon,TV Show)
(7076,Cocomelon,TV Show)
(7077,Cocomelon,TV Show)
(7078,Cocomelon,TV Show)
(7079,Cocomelon,TV Show)
(7080,Cocomelon,TV Show)
(7081,Cocomelon,TV Show)
(7082,Cocomelon,TV Show)
(7083,Cocomelon,TV Show)
(7084,Cocomelon,TV Show)
(7085,Cocomelon,TV Show)
(7086,Cocomelon,TV Show)
(7087,Cocomelon,TV Show)
(7088,Cocomelon,TV Show)
(7089,Cocomelon,TV Show)
(7090,Cocomelon,TV Show)
(7091,Cocomelon,TV Show)
(7092,Cocomelon,TV Show)
(7093,Cocomelon,TV Show)
(7094,Cocomelon,TV Show)
(7095,Cocomelon,TV Show)
(7096,Cocomelon,TV Show)
(7097,Cocomelon,TV Show)
(7098,Cocomelon,TV Show)
(7099,Cocomelon,TV Show)
(7100,Cocomelon,TV Show)
grunt> ■

```

## Query 16 Display the details of titles released in 2020

```

recent.netflix = filter.netflix by
GetYear(ToDate(rdate,'MM-dd-yyyy'))==2020;

```

```

hduser@rudra: ~
grunt> recent.netflix = filter.netflix by GetYear(ToDate(rdate,'dd/MM/yyyy'))==2020;
2022-11-19 22:55:10,959 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 2 time(s).
2022-11-19 22:55:10,959 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> ■

```

```

hduser@rudra: ~
(18-02-2022,3,1,Sweet Magnolias,TV Show,Yes,5-19-2020,33,270)
(19-02-2022,2,3,Love is Blind,TV Show,Yes,2-13-2020,26,162)
(19-02-2022,2,3,Sweet Magnolias,TV Show,Yes,5-19-2020,34,277)
(19-02-2022,6,Space Force,TV Show,Yes,5-19-2020,26,180)
(20-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,27,171)
(20-02-2022,4,3,Sweet Magnolias,TV Show,Yes,5-19-2020,35,284)
(20-02-2022,6,Space Force,TV Show,Yes,5-29-2020,27,193)
(21-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,28,180)
(21-02-2022,4,3,Sweet Magnolias,TV Show,Yes,5-19-2020,36,291)
(21-02-2022,5,Space Force,TV Show,Yes,5-29-2020,28,199)
(22-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,28,189)
(22-02-2022,3,3,Sweet Magnolias,TV Show,Yes,5-19-2020,31,299)
(22-02-2022,6,Space Force,TV Show,Yes,5-29-2020,29,284)
(23-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,30,198)
(23-02-2022,3,3,Sweet Magnolias,TV Show,Yes,5-19-2020,38,307)
(23-02-2022,6,3,Space Force,TV Show,Yes,5-29-2020,30,209)
(24-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,31,207)
(24-02-2022,4,3,Sweet Magnolias,TV Show,Yes,5-19-2020,39,314)
(24-02-2022,8,Space Force,TV Show,Yes,5-29-2020,31,212)
(25-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,28,216)
(25-02-2022,4,3,Sweet Magnolias,TV Show,Yes,5-19-2020,46,321)
(25-02-2022,9,Space Force,TV Show,Yes,5-29-2020,32,214)
(25-02-2022,10,,Cocomelon,TV Show,,06-01-2020,427,1472)
(26-02-2022,2,2,Love is Blind,TV Show,Yes,2-13-2020,33,225)
(26-02-2022,6,4,Sweet Magnolias,TV Show,Yes,5-19-2020,41,326)
(27-02-2022,4,2,Love is Blind,TV Show,Yes,2-13-2020,34,232)
(27-02-2022,4,2,Love is Blind,TV Show,Yes,2-13-2020,42,332)
(28-02-2022,4,2,Love is Blind,TV Show,Yes,2-13-2020,42,238)
(28-02-2022,6,4,Sweet Magnolias,TV Show,Yes,5-19-2020,43,337)
(01-03-2022,4,2,Love is Blind,TV Show,Yes,2-13-2020,36,246)
(01-03-2022,6,3,Sweet Magnolias,TV Show,Yes,5-19-2020,44,342)
(01-03-2022,9,3,Cocomelon,TV Show,,06-01-2020,428,1474)
(02-03-2022,5,2,Love is Blind,TV Show,Yes,2-13-2020,37,252)
(02-03-2022,9,3,Sweet Magnolias,TV Show,Yes,5-19-2020,45,344)
(03-03-2022,5,2,Love is Blind,TV Show,Yes,2-13-2020,38,250)
(04-03-2022,1,2,Love is Blind,TV Show,Yes,2-13-2020,40,253)
(05-03-2022,1,2,Love is Blind,TV Show,Yes,2-13-2020,49,273)
(06-03-2022,3,4,Love is Blind,TV Show,Yes,2-13-2020,41,281)
(07-03-2022,4,4,Love is Blind,TV Show,Yes,2-13-2020,42,288)
(08-03-2022,4,4,Love is Blind,TV Show,Yes,2-13-2020,43,295)
(09-03-2022,5,5,Love is Blind,TV Show,Yes,2-13-2020,44,301)
(10-03-2022,6,5,Love is Blind,TV Show,Yes,2-13-2020,45,306)
(11-03-2022,5,6,Love is Blind,TV Show,Yes,2-13-2020,46,312)
grunt> ■

```

**Query 17** Display the titles and type which were released before three years and are not present in Netflix Exclusive

```
before_titles = filter netflix by  
YearsBetween(CurrentTime(),ToDate(rdate,'MM-dd-yyyy'))>3  
and excl is null;
```

The screenshot shows two terminal windows. The top window displays the command and its execution results:

```
hduser@rudra: ~  
grunt> before_titles = filter netflix by YearsBetween(CurrentTime(),ToDate(rdate,'MM-dd-yyyy'))>3 and excl is null;  
2022-11-20 03:03:55,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 2 time(s).  
grunt>
```

The bottom window shows the output of the query, listing numerous titles and their details:

```
hduser@rudra: ~  
25-10-2021,5,4, Shameless, TV Show, 0-30-2010,48,246)  
(26-10-2021,6,,Riverdale, TV Show, 5-19-2017,29,132)  
(26-10-2021,9,4, Shameless, TV Show, 6-30-2016,41,248)  
(27-10-2021,6,9,Riverdale, TV Show, 5-19-2017,30,137)  
(27-10-2021,9,4, Shameless, TV Show, 6-30-2016,42,250)  
(28-10-2021,9,8,Riverdale, TV Show, 5-19-2017,31,139)  
(29-10-2021,10,7,Riverdale, TV Show, 5-19-2017,32,148)  
(30-10-2021,7,8,The Great British Baking Show,TV Show,,10-03-2015,20,70)  
(02-11-2021,9,6,Riverdale, TV Show,,5-19-2017,33,142)  
(03-11-2021,10,1,Shameless, TV Show, 0-30-2010,48,246)  
(06-11-2021,8,7,The Great British Baking Show,TV Show,,10-03-2015,21,73)  
(13-11-2021,5,8,The Great British Baking Show,TV Show,,10-03-2015,22,79)  
(20-11-2021,7,5,The Great British Baking Show,TV Show,,10-03-2015,23,83)  
(27-11-2021,9,7,The Great British Baking Show,TV Show,,10-03-2015,24,85)  
(08-12-2021,10,,Ink Master, TV Show, 9-15-2014,1,1)  
(09-12-2021,8,,Ink Master,TV Show, 9-15-2014,2,4)  
(18-12-2021,8,,Ink Master,TV Show, 9-15-2014,3,7)  
(05-01-2022,6,,The Longest Yard,Movie,,04-01-2015,1,2)  
(06-01-2022,7,,The Longest Yard,Movie,,04-01-2015,2,6)  
(07-01-2022,6,,The Longest Yard,Movie,,04-01-2015,3,11)  
(08-01-2022,7,,The Longest Yard,Movie,,04-01-2015,4,15)  
(09-01-2022,7,,The Longest Yard,Movie,,04-01-2015,5,19)  
(10-01-2022,7,,The Longest Yard,Movie,,04-01-2015,6,23)  
(11-01-2022,9,,The Longest Yard,Movie,,04-01-2015,7,25)  
(12-01-2022,10,,St. Vincent,Movie,,08-10-2016,1,1)  
(16-02-2022,10,,St. Vincent,Movie,,08-10-2016,2,11)  
(17-02-2022,8,,St. Vincent,Movie,,08-10-2016,2,41)  
(18-02-2022,7,,St. Vincent,Movie,,08-10-2016,3,8)  
(21-02-2022,10,,St. Vincent,Movie,,08-10-2016,4,9)  
(22-02-2022,10,,St. Vincent,Movie,,08-10-2016,5,10)  
(23-02-2022,10,,St. Vincent,Movie,,08-10-2016,6,11)  
(05-03-2022,9,,Shrek,Movie,,05-01-2018,1,1)  
(06-03-2022,9,,Shrek,Movie,,05-01-2018,2,3)  
(07-03-2022,9,,Shrek,Movie,,05-01-2018,3,5)  
(08-03-2022,9,,Shrek,Movie,,05-01-2018,4,7)  
(09-03-2022,8,,Shrek,Movie,,05-01-2018,5,10)  
(09-03-2022,10,,Shooter,Movie,,08-01-2014,1,1)  
(10-03-2022,2,,The Last Kingdom,TV Show,,07-18-2016,12,52)  
(10-03-2022,8,,Shooter,Movie,,08-01-2014,2,4)  
(10-03-2022,10,,Shrek,Movie,,05-01-2018,6,11)  
(11-03-2022,1,,The Last Kingdom,TV Show,,07-18-2016,13,62)  
(11-03-2022,8,,Shooter,Movie,,08-01-2014,3,7)  
(11-03-2022,10,,Shrek,Movie,,05-01-2018,7,12)  
grunt>
```

**Query 18** Display the titles and their type which were released in the month July and are in top 3 in July of 2020

```
july_titles = filter netflix by GetMonth(ToDate(rdate,'MM-  
dd-yyyy'))==7 and dailyrank==3 and  
GetMonth(ToDate(topdate,'dd-MM-yyyy'))==7;
```

```

hduser@rudra:~
```

```

grunt> July titles = filter netflix by GetMonth(ToDate(rdate, 'MM-dd-yyyy'))==7 and dailyrank==3 and GetMonth(ToDate(topdate,'dd-MM-yyyy'))==7;
2022-11-19 23:29:42,218 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 11 time(s).
2022-11-19 23:29:42,218 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 23:29:42,218 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt>
```

```

hduser@rudra:~
```

```

2022-11-19 23:29:50,736 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 23:29:50,736 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-19 23:29:50,744 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 23:29:50,744 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(02-07-2020,3,,Patriots Day,Movie,,07-01-2020,1,8)
(04-07-2020,3,,Patriots Day,Movie,,07-01-2020,3,25)
(05-07-2020,3,,The Baby-Sitters Club,TV Show,Yes,07-03-2020,2,15)
(06-07-2020,3,,The Baby-Sitters Club,TV Show,Yes,07-03-2020,3,23)
(07-07-2020,3,,The Baby-Sitters Club,TV Show,Yes,07-03-2020,4,31)
(08-07-2020,3,,The Baby-Sitters Club,TV Show,Yes,07-03-2020,5,39)
(09-07-2020,3,,Dr. Seuss's The Lorax,Movie,,07-07-2020,2,9)
(10-07-2020,3,,Warrior Nun,TV Show,Yes,07-02-2020,8,69)
(11-07-2020,3,,Dr. Seuss's The Lorax,Movie,,07-07-2020,4,26)
(12-07-2020,3,,Dr. Seuss's The Lorax,Movie,,07-07-2020,5,34)
(13-07-2020,3,,Down to Earth with Zac Efron,TV Show,Yes,07-10-2020,3,20)
(14-07-2020,3,,Dr. Seuss's The Lorax,Movie,,07-07-2020,7,49)
(15-07-2020,3,,Dr. Seuss's The Lorax,Movie,,07-07-2020,8,57)
(16-07-2020,3,,Down to Earth with Zac Efron,TV Show,Yes,07-10-2020,6,46)
(17-07-2020,3,,The Business of Drugs,TV Show,Yes,7-14-2020,3,24)
(18-07-2020,3,1,The Old Guard,Movie,Yes,07-10-2020,8,77)
(19-07-2020,3,1,The Old Guard,Movie,Yes,07-10-2020,9,85)
(20-07-2020,3,1,The Old Guard,Movie,Yes,07-10-2020,10,93)
(21-07-2020,3,,Fatal Affair,Movie,Yes,7-16-2020,5,46)
(22-07-2020,3,,Fatal Affair,Movie,Yes,7-16-2020,6,54)
(23-07-2020,3,,Fear City,Movie,Yes,7-19-2020,1,8)
(24-07-2020,3,,The Last Dance,TV Show,Yes,7-19-2020,3,42)
(25-07-2020,3,2,Cursed,TV Show,Yes,7-17-2020,8,76)
(26-07-2020,3,1,Cursed,TV Show,Yes,7-17-2020,9,84)
(27-07-2020,3,4,The Last Dance,TV Show,Yes,7-19-2020,8,64)
(28-07-2020,3,,Fear City: New York vs. The..,TV Show,Yes,7-22-2020,6,53)
(29-07-2020,3,,Animal Crackers,Movie,Yes,7-24-2020,5,29)
(30-07-2020,3,,Last Chance U,TV Show,Yes,7-29-2016,2,12)
(31-07-2020,3,,The Clickers,Movie,Yes,7-29-2020,1,9)
(03-07-2021,3,,Fear Street Part 1: 1994,Movie,Yes,07-02-2021,1,8)
(05-07-2021,3,,Fear Street Part 1: 1994,Movie,Yes,07-02-2021,3,25)
(06-07-2021,3,,Fear Street Part 1: 1994,Movie,Yes,07-02-2021,4,33)
(08-07-2021,3,,Mother's Day,Movie,,07-01-2021,5,33)
(09-07-2021,3,,Resident Evil: Infinite Darkness,TV Show,Yes,07-08-2021,1,8)
(15-07-2021,3,,Gunpowder Milkshake,Movie,Yes,7-14-2021,1,8)
(16-07-2021,3,,Gunpowder Milkshake,Movie,Yes,7-14-2021,2,16)
(17-07-2021,3,,Fear Street Part 3: 1866,Movie,Yes,7-16-2021,1,8)
(18-07-2021,3,,Gunpowder Milkshake,Movie,Yes,7-14-2021,4,31)
(29-07-2021,3,,Blood Red Sky,Movie,Yes,7-23-2021,6,53)
grunt>
```

## Query 19 Display the count of TV Shows year-wise

```

tvyear_group = group tvshows by GetYear(ToDate(rdate, 'MM-
dd-yyyy')) ;
```

```

tvyear_count = foreach tvyear_group generate group,
COUNT(tvshows) ;
```

```

hduser@rudra: ~
grunt> tyear_group = group tvshows by GetYear(ToDate(rdate,'MM-dd-yyyy'));
2022-11-19 23:33:33,672 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 11 time(s).
2022-11-19 23:33:33,672 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 23:33:33,672 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> tyear_count = foreach tyear_group generate group, COUNT(tvshows);
2022-11-19 23:33:34,934 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 11 time(s).
2022-11-19 23:33:34,934 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2022-11-19 23:33:34,934 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt>

hduser@rudra: ~
file:/tmp/temp-215458488/tmp-348596877,
Input(s):
Successfully read 7100 records from: "file:///home/hduser/Project/netflix.csv"
Output(s):
Successfully stored 10 records in: "file:/tmp/temp-215458488/tmp-348596877"
Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1992817692_0044

2022-11-19 23:34:02,012 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 23:34:02,014 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 23:34:02,015 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-19 23:34:02,018 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION FAILED 3968 time(s).
2022-11-19 23:34:02,018 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-19 23:34:02,019 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-19 23:34:02,019 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-11-19 23:34:02,019 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-19 23:34:02,019 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-19 23:34:02,033 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-19 23:34:02,033 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2007,1)
(2014,44)
(2015,88)
(2016,94)
(2017,251)
(2018,309)
(2019,385)
(2020,1840)
(2021,1248)
(2022,184)
grunt>

```

**Query 20** Display the titles which were released in between of 10<sup>th</sup> and 30<sup>th</sup> week of year 2015 to 2018

```

week_year_titles = filter netflix by
GetWeek(ToDate(rdate,'MM-dd-yyyy'))>=10 and
GetWeek(ToDate(rdate,'MM-dd-yyyy'))<=30 and
GetYear(ToDate(rdate,'MM-dd-yyyy'))>=2015 and
GetYear(ToDate(rdate,'MM-dd-yyyy'))<=2018;

```

```

lhduser@rudra: ~
grunts> week_year_titles = filter.netflix.filter((rdate,'MM-dd-yyyy'))>=10 and GetWeekToDate(rdate,'MM-dd-yyyy'))<=2018
2022-11-19 23:39:07,850 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO LONG 11 time(s).
2022-11-19 23:39:07,850 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 2 time(s).
2022-11-19 23:39:07,850 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt>

lhduser@rudra: ~
(27-01-2022,1,9,Ozark,TV Show,Yes,7-21-2017,56,,337)
(28-01-2022,1,6,Ozark,TV Show,Yes,7-21-2017,57,,347)
(29-01-2022,1,1,Ozark,TV Show,Yes,7-21-2017,58,,357)
(30-01-2022,2,1,Ozark,TV Show,Yes,7-21-2017,59,,366)
(31-01-2022,3,1,Ozark,TV Show,Yes,7-21-2017,60,,376)
(01-02-2022,3,1,Ozark,TV Show,Yes,7-21-2017,61,,382)
(02-02-2022,4,1,Ozark,TV Show,Yes,7-21-2017,62,,389)
(03-02-2022,4,1,Ozark,TV Show,Yes,7-21-2017,63,,396)
(04-02-2022,4,1,Ozark,TV Show,Yes,7-21-2017,64,,403)
(05-02-2022,5,2,Ozark,TV Show,Yes,7-21-2017,65,,409)
(06-02-2022,5,2,Ozark,TV Show,Yes,7-21-2017,66,,415)
(07-02-2022,3,3,Ozark,TV Show,Yes,7-21-2017,67,,423)
(08-02-2022,3,4,Ozark,TV Show,Yes,7-21-2017,68,,430)
(09-02-2022,3,4,Ozark,TV Show,Yes,7-21-2017,69,,439)
(10-02-2022,3,4,Ozark,TV Show,Yes,7-21-2017,70,,447)
(11-02-2022,4,4,Ozark,TV Show,Yes,7-21-2017,71,,454)
(12-02-2022,6,5,Ozark,TV Show,Yes,7-21-2017,72,,459)
(13-02-2022,6,5,Ozark,TV Show,Yes,7-21-2017,73,,464)
(14-02-2022,7,3,Ozark,TV Show,Yes,7-21-2017,74,,468)
(15-02-2022,5,3,Ozark,TV Show,Yes,7-21-2017,75,,474)
(16-02-2022,5,3,Ozark,TV Show,Yes,7-21-2017,76,,480)
(17-02-2022,5,4,Ozark,TV Show,Yes,7-21-2017,77,,486)
(18-02-2022,5,4,Ozark,TV Show,Yes,7-21-2017,78,,492)
(19-02-2022,8,6,Ozark,TV Show,Yes,7-21-2017,79,,495)
(20-02-2022,7,6,Ozark,TV Show,Yes,7-21-2017,80,,499)
(21-02-2022,7,7,Ozark,TV Show,Yes,7-21-2017,81,,503)
(22-02-2022,8,5,Ozark,TV Show,Yes,7-21-2017,82,,506)
(23-02-2022,5,5,Ozark,TV Show,Yes,7-21-2017,83,,512)
(24-02-2022,5,5,Ozark,TV Show,Yes,7-21-2017,84,,518)
(25-02-2022,7,8,Ozark,TV Show,Yes,7-21-2017,85,,524)
(26-02-2022,8,8,Ozark,TV Show,Yes,7-21-2017,86,,527)
(27-02-2022,10,7,Ozark,TV Show,Yes,7-21-2017,87,,528)
(28-02-2022,7,7,Ozark,TV Show,Yes,7-21-2017,88,,532)
(01-03-2022,7,8,Ozark,TV Show,Yes,7-21-2017,89,,536)
(05-03-2022,10,,Shrek,Movie,,05-01-2018,1,1)
(06-03-2022,9,,Shrek,Movie,,05-01-2018,2,3)
(07-03-2022,9,,Shrek,Movie,,05-01-2018,3,5)
(08-03-2022,9,,Shrek,Movie,,05-01-2018,4,7)
(09-03-2022,8,,Shrek,Movie,,05-01-2018,5,10)
(10-03-2022,2,,The Last Kingdom,TV Show,,07-10-2016,12,52)
(10-03-2022,10,,Shrek,Movie,,05-01-2018,6,11)
(11-03-2022,1,,The Last Kingdom,TV Show,,07-10-2016,13,62)
(11-03-2022,10,,Shrek,Movie,,05-01-2018,7,12)
grunt>

```

## CASE STUDY 1

### Most occurred Start Letter

---

Analyze large text file ‘shakespeare.txt’, and

- a. Find the most occurred start letter
- b. Calculate the wordcount for each word.

Loading the file ‘shakespeare.txt’ into a relation ‘rel’:

```
rel = load 'Project/shakespeare.txt' as (line:chararray);
```

Tokenizes each word from the ‘line’ field of above relation ‘rel’ as word:

```
words = foreach rel generate flatten(TOKENIZE(line)) as word;
```

Extracting first letter of each word:

```
letters = foreach words generate SUBSTRING(word, 0, 1) as letter;
```

Removing the tuples if any contains NULL values:

```
new_letters = filter letters by letter is not NULL;
```

Grouping the letters on the basis of their similarity:

```
grp = group new_letters by letter;
```

Counting the letter present in each bag of the groups:

```
count_letter = foreach grp generate group, COUNT(new_letters);
```

Sorting the letter in descending order:

```
order_letter = order count_letter by $1 desc;
```

Generating the letter having count maximum among all

```
grp2 = group order_letter all;  
x = foreach grp2 {top= TOP(1,1,order_letter); generate top;};
```

Display the count of occurrences of all letters and letter having maximum occurrence

```
dump x
```

```
dump count_letter
```

```
hduser@rudra: ~
(?,7)
(A,21100)
(B,10903)
(C,11882)
(D,6168)
(E,8262)
(F,7970)
(G,6813)
(H,10872)
(I,29873)
(J,1749)
(K,3631)
(L,7198)
(M,9456)
(N,4944)
(O,9285)
(P,8346)
(Q,1045)
(R,3858)
(S,13853)
(T,21995)
(U,1485)
(V,1684)
(W,14620)
(X,14)
(Y,3960)
(Z,18)
(I,2082)
(L,17)
(.,2)
(a,63466)
(b,34521)
(c,23365)
(d,23436)
(e,10355)
(f,28729)
(g,14699)
(h,50495)
(i,32163)
(j,1593)
(k,5784)
(l,22322)
(m,46190)
(n,21775)
(o,33981)
```

## Occurrences of each Letter

```
hduser@rudra: ~
Successfully read 124505 records from: "file:///home/hduser/Project/shakespeare.txt"
Output(s):
Successfully stored 1 records in: "file:/tmp/temp869450440/tmp1737659775"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local989686850_0006 -> job_local937540274_0007,
job_local937540274_0007 -> job_local1057578523_0008,
job_local1057578523_0008 -> job_local685958702_0009,
job_local685958702_0009

2022-11-20 14:47:34,090 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,091 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,092 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,095 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,096 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,097 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,097 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,101 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,103 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,104 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,107 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,108 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,110 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-11-20 14:47:34,111 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning UDF_WARNING_2 9454 times(s).
2022-11-20 14:47:34,111 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-20 14:47:34,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-20 14:47:34,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-11-20 14:47:34,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-20 14:47:34,112 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-20 14:47:34,127 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-20 14:47:34,127 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
{{(t,101298)}}
grunt> ■
```

## Most Occurred Start Letter and it's occurrence

## CASE STUDY 2

### Word Count

---

Removal of stopwords and punctuation from a large text file. Calculate the occurrence of each word from the file afterwards.

Loading the ‘shakespeare.txt’ file into relation ‘wordfile’ and ‘stop-word-list.csv’ file into a relation ‘stopword’:

```
wordfile = load 'Project/shakespeare.txt' as (line:chararray);  
stopword = load 'Project/stop-word-list.csv' using  
PigStorage();
```

Tokenize each word after removing punctuations and changing those to Lower case from ‘wordfile’ relation:

```
words = foreach wordfile generate  
flatten(TOKENIZE(REPLACE(LOWER(TRIM(line)), '[\p{Punct},\p{Cn  
trl}]', ' '))) as word;
```

Removing any tuple if containing NULL values from ‘words’ relation:

```
new_words = filter words by word is not NULL;
```

Tokenize each word from ‘stopword’ relation:

```
stopword = foreach stopword generate flatten(TOKENIZE($0)) as  
(sword:chararray);
```

Combine ‘wordfile’ and ‘stopword’ relation on the basis of words:

```
join_word = join words by word left outer, stopword by sword;
```

Removing the tuples where ‘stopword’ field ‘sword’ is NULL:

```
filter_word = filter join_word by stopword::sword is NULL;
```

Now extracting the words remaining in the ‘words’ relation:

```
words1 = foreach filter_word generate words::word as word;
```

Group the words on the basis of their similarity:

```
grp = group words1 by word;
```

Count the occurrences of the word present:

```
count_word = foreach grp generate group,COUNT(words1);  
dump count_word
```

```
hduser@rudra: ~
(understandings,1)
(unpremeditated,1)
(unquestionable,1)
(unthankfulness,3)
(voluptuousness,2)
(circumscription,1)
(disproportioned,1)
(distemperatures,1)
(distinguishment,1)
(enfranchisement,6)
(excommunication,1)
(extraordinarily,2)
(flibbertigibbet,2)
(gloucestershire,1)
(gloucestershire,11)
(impossibilities,2)
(interchangeably,4)
(interchangement,1)
(interrogatories,2)
(merchantability,1)
(misconstruction,1)
(northumberlands,1)
(notwithstanding,22)
(particularities,2)
(perpendicularly,1)
(portotartarossa,1)
(praeclarissimus,1)
(prognostication,2)
(superstitiously,1)
(transformations,1)
(uncompassionate,1)
(uncomprehensive,1)
(undistinguished,1)
(unreconciliable,1)
(incomprehensible,1)
(northamptonshire,1)
(superserviceable,1)
(anthropophaginian,1)
(indistinguishable,1)
(indistinguishable,2)
(honorificabilitudinitatibus,1)
(,0)
grunt>
grunt> ■
```

## Word count – Occurrence of each Word