

Below are the steps followed to execute the solution-

Below shows the input files in HDFS

```
[acadgild@localhost Casestudy1]$ hadoop fs -ls /
18/08/06 22:49:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 23 items
drwxr-xr-x   - acadgild supergroup          0 2018-07-08 23:05 /assignment-3
-rw-r--r--   1 acadgild supergroup         317 2018-07-22 16:50 /custompiginput.txt
drwxr-xr-x   - acadgild supergroup          0 2018-07-08 23:21 /deviation
-rw-r--r--   1 acadgild supergroup 1056906185 2018-07-08 17:42 /file.txt
drwxr-xr-x   - acadgild supergroup          0 2018-07-15 17:19 /hadoopdata
drwxr-xr-x   - acadgild supergroup          0 2018-07-30 23:27 /hbase
drwxr-xr-x   - acadgild supergroup          0 2018-07-09 12:58 /invalidcount
-rw-r--r--   1 acadgild supergroup    2283410 2018-08-06 22:47 /movies.csv
drwxr-xr-x   - acadgild supergroup          0 2018-07-09 14:37 /onidastatecount
-rw-r--r--   1 acadgild supergroup 709550327 2018-08-06 22:49 /ratings.csv
drwxr-xr-x   - acadgild supergroup          0 2018-07-22 21:38 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxr-xr-x   - acadgild supergroup          0 2018-07-22 22:08 /sqoopout2
drwxr-xr-x   - acadgild supergroup          0 2018-07-20 23:36 /sqoopoutbyid
drwxr-xr-x   - acadgild supergroup          0 2018-07-18 22:31 /sqoopoutbyid2
-rw-r--r--   1 acadgild supergroup          733 2018-07-09 11:47 /television.txt
-rw-r--r--   1 acadgild supergroup          358 2018-07-08 16:59 /test.txt
drwxrwx---   - acadgild supergroup          0 2018-07-22 16:54 /tmp
drwxr-xr-x   - acadgild supergroup          0 2018-07-09 13:51 /unitcount
drwxr-xr-x   - acadgild supergroup          0 2018-07-15 21:44 /user
drwxr-xr-x   - acadgild supergroup          0 2018-07-22 16:08 /wordcount
drwxr-xr-x   - acadgild supergroup          0 2018-07-08 23:14 /wordmean
drwxr-xr-x   - acadgild supergroup          0 2018-07-15 16:04 /wordoutput
```

Below screenshot shows the command to execute jar-

```
[acadgild@localhost Casestudy1]$ hadoop jar casestudy1.jar /movies.csv /ratings.csv /CaseStudy1out
18/08/06 22:53:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/06 22:53:18 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/06 22:53:26 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application through ToolRunner to remedy this.
18/08/06 22:53:29 INFO input.FileInputFormat: Total input paths to process : 1
18/08/06 22:53:29 INFO input.FileInputFormat: Total input paths to process : 1
18/08/06 22:53:29 INFO mapreduce.JobSubmitter: number of splits:7
18/08/06 22:53:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533572985892_0001
18/08/06 22:53:34 INFO impl.YarnClientImpl: Submitted application application_1533572985892_0001
18/08/06 22:53:35 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533572985892_0001/
18/08/06 22:53:35 INFO mapreduce.Job: Running job: job_1533572985892_0001
```

```
18/08/06 22:59:07 INFO mapreduce.Job: map 17% reduce 0%
18/08/06 22:59:09 INFO mapreduce.Job: map 18% reduce 0%
18/08/06 22:59:11 INFO mapreduce.Job: map 19% reduce 0%
18/08/06 22:59:13 INFO mapreduce.Job: map 21% reduce 0%
18/08/06 22:59:17 INFO mapreduce.Job: map 22% reduce 0%
18/08/06 22:59:18 INFO mapreduce.Job: map 23% reduce 0%
18/08/06 22:59:21 INFO mapreduce.Job: map 24% reduce 0%
18/08/06 22:59:24 INFO mapreduce.Job: map 25% reduce 0%
18/08/06 22:59:25 INFO mapreduce.Job: map 26% reduce 0%
18/08/06 22:59:29 INFO mapreduce.Job: map 27% reduce 0%
18/08/06 22:59:33 INFO mapreduce.Job: map 28% reduce 0%
18/08/06 22:59:36 INFO mapreduce.Job: map 29% reduce 0%
18/08/06 22:59:37 INFO mapreduce.Job: map 30% reduce 0%
18/08/06 22:59:39 INFO mapreduce.Job: map 31% reduce 0%
18/08/06 22:59:40 INFO mapreduce.Job: map 32% reduce 0%
18/08/06 22:59:44 INFO mapreduce.Job: map 33% reduce 0%
18/08/06 22:59:47 INFO mapreduce.Job: map 34% reduce 0%
18/08/06 22:59:48 INFO mapreduce.Job: map 35% reduce 0%
18/08/06 22:59:54 INFO mapreduce.Job: map 36% reduce 0%
18/08/06 22:59:58 INFO mapreduce.Job: map 37% reduce 0%
18/08/06 23:00:02 INFO mapreduce.Job: map 38% reduce 0%
18/08/06 23:00:10 INFO mapreduce.Job: map 39% reduce 0%
18/08/06 23:00:14 INFO mapreduce.Job: map 40% reduce 0%
18/08/06 23:00:27 INFO mapreduce.Job: map 45% reduce 0%
18/08/06 23:01:11 INFO mapreduce.Job: map 46% reduce 0%
18/08/06 23:01:16 INFO mapreduce.Job: map 47% reduce 0%
18/08/06 23:01:21 INFO mapreduce.Job: map 48% reduce 0%
18/08/06 23:01:24 INFO mapreduce.Job: map 49% reduce 0%
18/08/06 23:01:25 INFO mapreduce.Job: map 50% reduce 0%
18/08/06 23:01:28 INFO mapreduce.Job: map 51% reduce 0%
18/08/06 23:01:30 INFO mapreduce.Job: map 52% reduce 0%
18/08/06 23:01:34 INFO mapreduce.Job: map 53% reduce 0%
18/08/06 23:01:37 INFO mapreduce.Job: map 54% reduce 0%
18/08/06 23:01:40 INFO mapreduce.Job: map 55% reduce 0%
18/08/06 23:01:43 INFO mapreduce.Job: map 56% reduce 0%
18/08/06 23:01:47 INFO mapreduce.Job: map 57% reduce 0%
18/08/06 23:01:49 INFO mapreduce.Job: map 58% reduce 0%
```

```
18/05/24 23:43:59 INFO mapreduce.Job: map 100% reduce 98%
18/05/24 23:44:01 INFO mapreduce.Job: map 100% reduce 100%
18/05/24 23:44:01 INFO mapreduce.Job: Job job_1527185406334_0001 completed successfully
18/05/24 23:44:01 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=961694264
    FILE: Number of bytes written=1457486118
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=711855894
    HDFS: Number of bytes written=1669001
    HDFS: Number of read operations=24
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=8
    Launched reduce tasks=1
    Data-local map tasks=8
    Total time spent by all maps in occupied slots (ms)=683050
    Total time spent by all reduces in occupied slots (ms)=69542
    Total time spent by all map tasks (ms)=683050
    Total time spent by all reduce tasks (ms)=69542
    Total vcore-milliseconds taken by all map tasks=683050
    Total vcore-milliseconds taken by all reduce tasks=69542
    Total megabyte-milliseconds taken by all map tasks=699443200
    Total megabyte-milliseconds taken by all reduce tasks=71211008
  Map-Reduce Framework
    Map input records=26070134
    Map output records=26070132
    Map output bytes=442789828
    Map output materialized bytes=494930141
    Input split bytes=1677
    Combine input records=0
    Combine output records=0
```

```
[acadgild@localhost ~]$ hadoop fs -ls /CaseStudyIOut
18/05/24 23:59:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
re applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-05-24 23:44 /CaseStudyIOut/_SUCCESS
-rw-r--r--  1 acadgild supergroup 1669001 2018-05-24 23:43 /CaseStudyIOut/part-r-00000
[acadgild@localhost ~]$
```

Below shows the output-

```
[acadgild@localhost ~]$ hadoop fs -cat /CaseStudyIOut/part-r-00000
18/05/25 00:01:02 WARN util.NativeCodeLoader: Unable to load native-hadoop
re applicable
Toy Story (1995)      66008    3.888157
GoldenEye (1995)     32534    3.431841
City Hall (1996)     4436     3.232304
Curdled (1996)    217      3.099078
"Comic 1            4.000000
Up in Smoke (1957)    3         3.666667
First Daughter (1999) 3         3.333333
"Flaw 14           3.714286
Battle of Los Angeles (2011) 44      2.522727
Jason Becker: Not Dead Yet (2012) 9      3.444444
Chicago Massacre: Richard Speck (2007) 2      2.500000
Keep the Lights On (2012) 25      3.100000
Beauty Is Embarrassing (2012) 15      3.600000
Girl Model (2011)    32      3.281250
Crossfire Hurricane (2012) 18      3.388889
Middle of Nowhere (2012) 11      3.454545
True Blue (2001)     3         3.000000
"Guns of Fort Petticoat 3         3.333333
Human Planet (2011)  197      4.271574
Madagascar (2011)  26      3.769231
Omar Killed Me (Omar m'a tuer) (2011) 9      3.166667
Enola Gay and the Atomic Bombing of Japan (1995) 1      3.500000
Red Hook Summer (2012) 11      2.045455
Stella Maris (1918)  2         3.750000
Die (2010)          10      2.550000
Patrice O'Neal: Elephant in the Room (2011) 22      3.204545
Sunny (Sseo-ni) (2011) 26      3.576923
My Way (Mai Wei) (2011) 30      3.716667
```