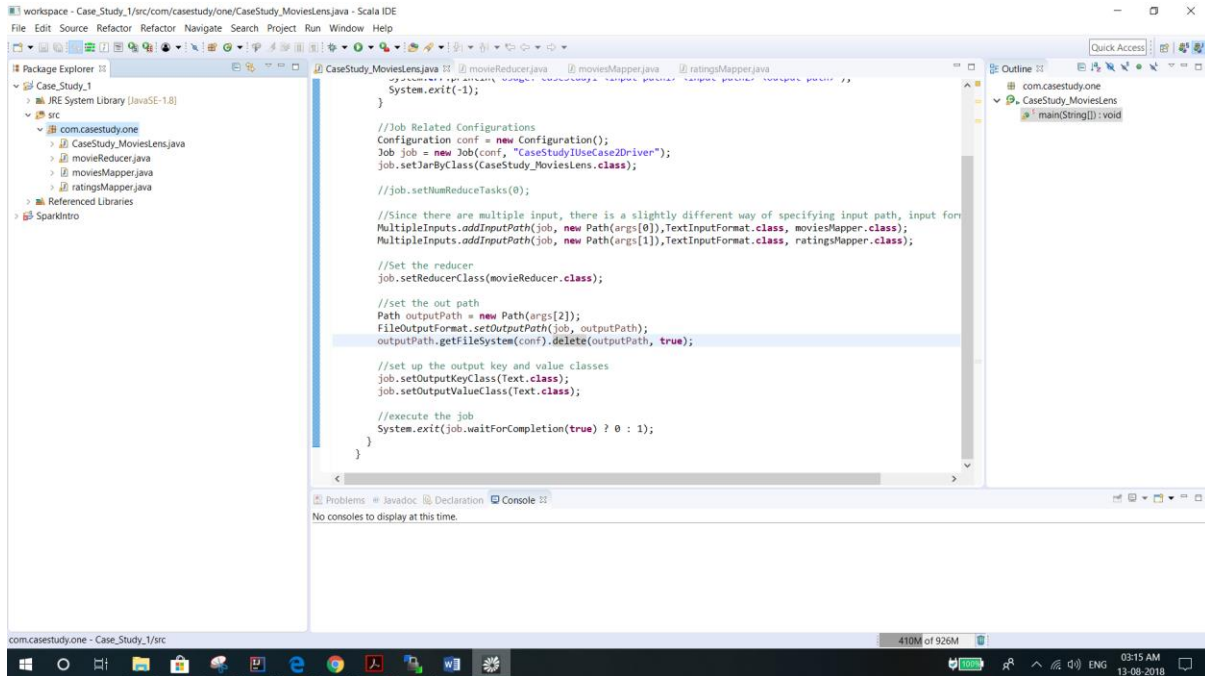


Case Study – I

Working with Movie ratings

- Project Setup Created as below
- Created jar file & ftp'd to acadgild VM.
- Please find the attached codes & jar file in git for reference.

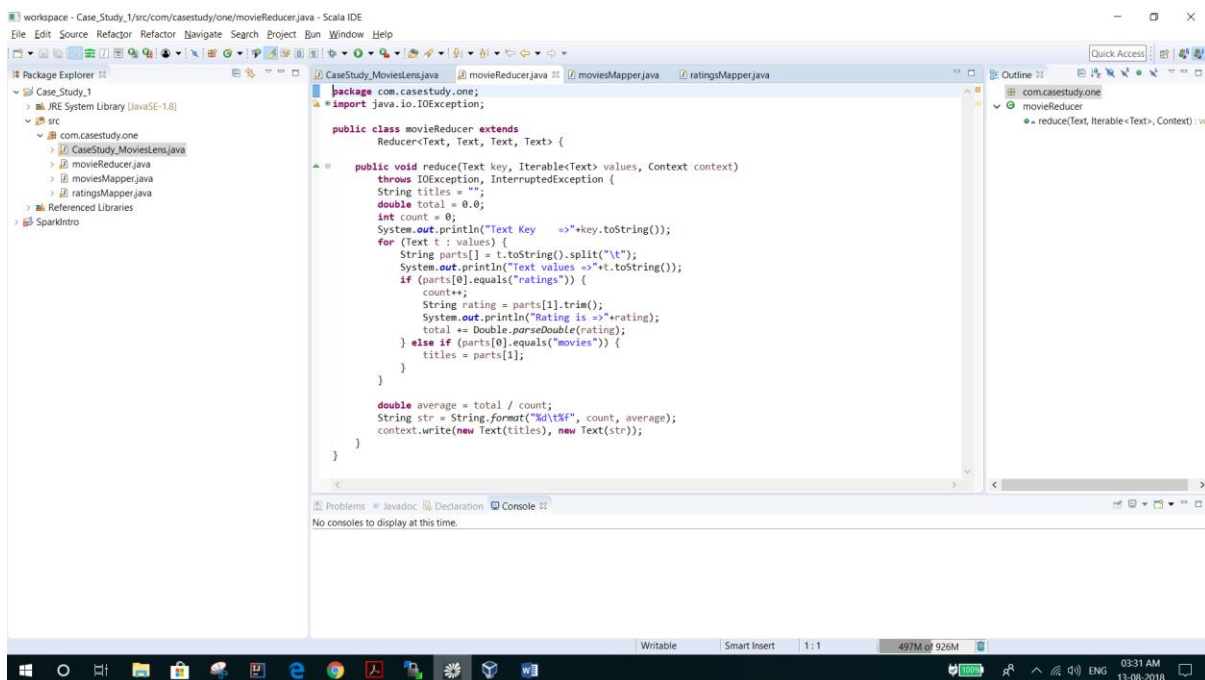
1. What are the movie titles that the user has rated?



2 How many times a movie has been rated by the user?

3 In question 2 above, what is the average rating given for a movie?

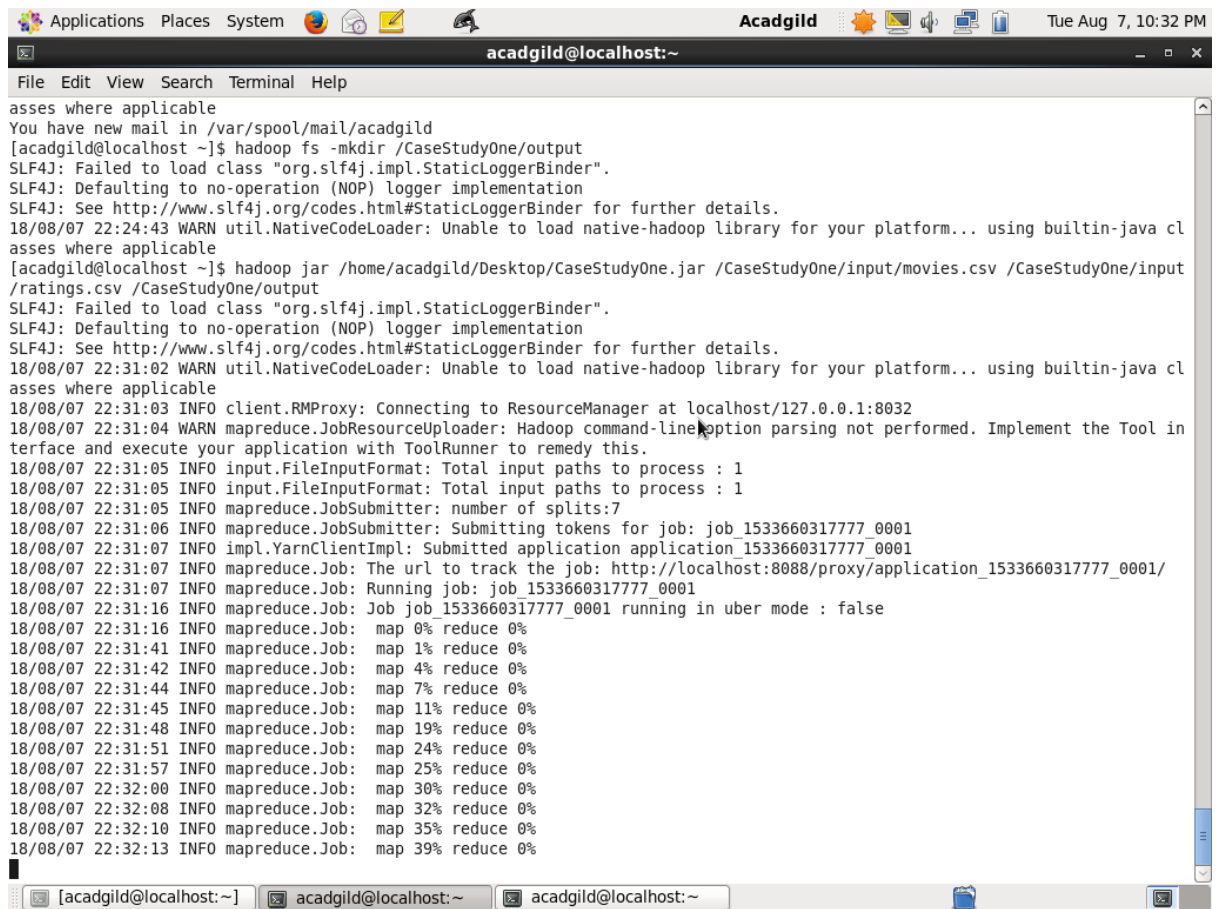
Here we will in reducer class we need count the ratings of each user & calculate the average rating for each movie by diving with count



Case Study – I

Working with Movie ratings

- Submitted job using Hadoop jar command < both input paths > <output path>
- Please find the screenshots below



The screenshot shows a terminal window titled 'acadmild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the following output:

```
asses where applicable
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ hadoop fs -mkdir /CaseStudyOne/output
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
18/08/07 22:24:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadmild@localhost ~]$ hadoop jar /home/acadmild/Desktop/CaseStudyOne.jar /CaseStudyOne/input/movies.csv /CaseStudyOne/input
/ratings.csv /CaseStudyOne/output
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
18/08/07 22:31:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/08/07 22:31:03 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/07 22:31:04 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/08/07 22:31:05 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 22:31:05 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 22:31:05 INFO mapreduce.JobSubmitter: number of splits:7
18/08/07 22:31:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533660317777_0001
18/08/07 22:31:07 INFO impl.YarnClientImpl: Submitted application application_1533660317777_0001
18/08/07 22:31:07 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533660317777_0001/
18/08/07 22:31:07 INFO mapreduce.Job: Running job: job_1533660317777_0001
18/08/07 22:31:16 INFO mapreduce.Job: Job job_1533660317777_0001 running in uber mode : false
18/08/07 22:31:16 INFO mapreduce.Job: map 0% reduce 0%
18/08/07 22:31:41 INFO mapreduce.Job: map 1% reduce 0%
18/08/07 22:31:42 INFO mapreduce.Job: map 4% reduce 0%
18/08/07 22:31:44 INFO mapreduce.Job: map 7% reduce 0%
18/08/07 22:31:45 INFO mapreduce.Job: map 11% reduce 0%
18/08/07 22:31:48 INFO mapreduce.Job: map 19% reduce 0%
18/08/07 22:31:51 INFO mapreduce.Job: map 24% reduce 0%
18/08/07 22:31:57 INFO mapreduce.Job: map 25% reduce 0%
18/08/07 22:32:00 INFO mapreduce.Job: map 30% reduce 0%
18/08/07 22:32:08 INFO mapreduce.Job: map 32% reduce 0%
18/08/07 22:32:10 INFO mapreduce.Job: map 35% reduce 0%
18/08/07 22:32:13 INFO mapreduce.Job: map 39% reduce 0%
```

The terminal window has three tabs at the bottom, all titled 'acadmild@localhost:~'. The system tray at the bottom right shows a clock icon and a network icon.

Case Study – I

Working with Movie ratings

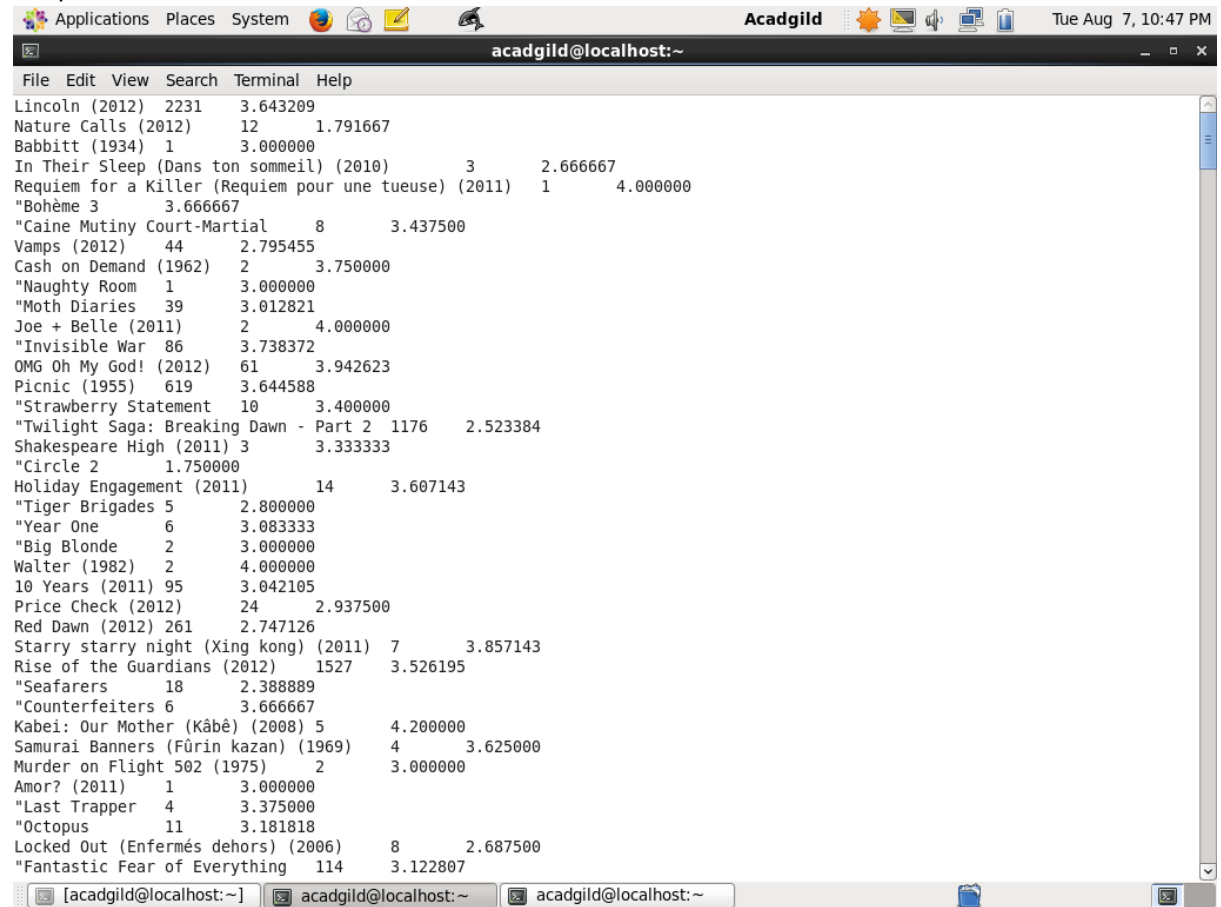
```
Applications Places System Acadgild Tue Aug 7, 10:45 PM
acadgild@localhost:~
File Edit View Search Terminal Help
18/08/07 22:38:21 INFO mapreduce.Job: map 100% reduce 95%
18/08/07 22:38:33 INFO mapreduce.Job: map 100% reduce 96%
18/08/07 22:38:42 INFO mapreduce.Job: map 100% reduce 97%
18/08/07 22:38:51 INFO mapreduce.Job: map 100% reduce 98%
18/08/07 22:39:03 INFO mapreduce.Job: map 100% reduce 99%
18/08/07 22:39:15 INFO mapreduce.Job: map 100% reduce 100%
18/08/07 22:39:25 INFO mapreduce.Job: Job job_1533660317777_0001 completed successfully
18/08/07 22:39:25 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=961694264
    FILE: Number of bytes written=1457486366
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=711856021
    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)=769610
    Total time spent by all reduces in occupied slots (ms)=402949
    Total time spent by all map tasks (ms)=769610
    Total time spent by all reduce tasks (ms)=402949
    Total vcore-milliseconds taken by all map tasks=769610
    Total vcore-milliseconds taken by all reduce tasks=402949
    Total megabyte-milliseconds taken by all map tasks=788080640
    Total megabyte-milliseconds taken by all reduce tasks=412619776
  Map-Reduce Framework
    Map input records=26070134
    Map output records=26070132
    Map output bytes=442789828
    Map output materialized bytes=494930141
    Input split bytes=1804
    Combine input records=0
    Total time spent by all reduce tasks (ms)=402949
    Total vcore-milliseconds taken by all map tasks=769610
    Total vcore-milliseconds taken by all reduce tasks=402949
    Total megabyte-milliseconds taken by all map tasks=788080640
    Total megabyte-milliseconds taken by all reduce tasks=412619776
  Map-Reduce Framework
    Map input records=26070134
    Map output records=26070132
    Map output bytes=442789828
    Map output materialized bytes=494930141
    Input split bytes=1804
    Combine input records=0
    Combine output records=0
    Reduce input groups=45843
    Reduce shuffle bytes=494930141
    Reduce input records=26070132
    Reduce output records=45843
    Spilled Records=76775523
    Shuffled Maps =7
    Failed Shuffles=0
    Merged Map outputs=7
    GC time elapsed (ms)=5990
    CPU time spent (ms)=469690
    Physical memory (bytes) snapshot=2032889856
    Virtual memory (bytes) snapshot=16670748672
    Total committed heap usage (bytes)=1542455296
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=1669001
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Case Study – I

Working with Movie ratings

Output:-

Output is as shown below in screenshots:-



The screenshot shows a terminal window titled 'Acadgild' with the prompt 'acadgild@localhost:~'. The window displays a list of movies and their ratings, organized into columns. The data is as follows:

Movie Title	Year	Count	Rating
Lincoln	2012	2231	3.643209
Nature Calls	2012	12	1.791667
Babbitt	1934	1	3.000000
In Their Sleep (Dans ton sommeil)	2010	3	2.666667
Requiem for a Killer (Requiem pour une tueuse)	2011	1	4.000000
"Bohème	3	3.666667	
"Caine Mutiny Court-Martial	8	3.437500	
Vamps	2012	44	2.795455
Cash on Demand	1962	2	3.750000
"Naughty Room	1	3.000000	
"Moth Diaries	39	3.012821	
Joe + Belle	2011	2	4.000000
"Invisible War	86	3.738372	
OMG Oh My God!	2012	61	3.942623
Picnic	1955	619	3.644588
"Strawberry Statement	10	3.400000	
"Twilight Saga: Breaking Dawn - Part 2	1176	2.523384	
Shakespeare High	2011	3	3.333333
"Circle 2	1.750000		
Holiday Engagement	2011	14	3.607143
"Tiger Brigades	5	2.800000	
"Year One	6	3.083333	
"Big Blonde	2	3.000000	
Walter	1982	2	4.000000
10 Years	2011	95	3.042105
Price Check	2012	24	2.937500
Red Dawn	2012	261	2.747126
Starry starry night (Xing kong)	2011	7	3.857143
Rise of the Guardians	2012	1527	3.526195
"Seafarers	18	2.388889	
"Counterfeiters	6	3.666667	
Kabei: Our Mother (Kâbê)	2008	5	4.200000
Samurai Banners (Fûrin kazan)	1969	4	3.625000
Murder on Flight 502	1975	2	3.000000
Amor?	2011	1	3.000000
"Last Trapper	4	3.375000	
"Octopus	11	3.181818	
Locked Out (Enfermés dehors)	2006	8	2.687500
"Fantastic Fear of Everything	114	3.122807	

Case Study – I

Working with Movie ratings

The screenshot shows a Linux terminal window with the title bar 'Acadgild' and the date 'Tue Aug 7, 10:47 PM'. The terminal displays a list of movie ratings with columns for movie title, year, number of ratings, and average rating. The list includes movies like 'Cargo (2011)', 'Jesse Stone: No Remorse (2010)', 'Jesse Stone: Innocents Lost (2011)', 'Jesse Stone: Benefit of the Doubt (2012)', 'Diana Vreeland: The Eye Has to Travel (2011)', 'Fine-Tune (2008)', 'Sangre de mi sangre (Padre Nuestro) (2007)', 'Milarepa (2006)', 'Marked for Murder (1993)', '2 Days in the Valley (1996)', '0 Panishyros Megistanas Ton Ninja (2008)', 'Morgan Pålsson - världsreporter (2008)', 'Renoir (2012)', 'Typhoon (Tae-poong) (2005)', 'Last Stand (2013)', 'Mama (2013)', 'Candidate (2013)', 'Upstream Color (2013)', 'Fast Life (1932)', 'Careless Love (2012)', 'Special When Lit (2009)', 'Year of the Tiger (2012)', 'Outpost: Black Sun (2012)', 'Preston Sturges: The Rise and Fall of an American Dreamer (1990)', 'Double Indemnity (1973)', 'London After Midnight (2002)', 'Broken City (2013)', 'Nursery University (2008)', 'Counterfeit Coin (2012)', 'Nameless Gangster (Bumchoiwai junjaeng) (2012)', 'Ginger & Rosa (2012)', 'First Shot (2002)', 'Marina Abramovic: The Artist Is Present (2012)', 'Bonsai (2011)', 'Shadow Dancer (2012)', 'Thale (2012)', 'It's a Disaster (2012)', and 'Texas Across the River (1966)'. The terminal also shows a Hadoop command being executed: `[acadgild@localhost ~]$ hadoop fs -cat /CaseStudyOne/output/part-r-00000`. The output of the command is displayed below the command, showing the same list of movie ratings as above. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The status bar at the bottom shows 'Current workspace: "Workspace 1"'. The terminal window is titled 'acadgild@localhost:~'.

```
File Edit View Search Terminal Help
Cargo (2011) 3 2.333333
Jesse Stone: No Remorse (2010) 30 3.266667
Jesse Stone: Innocents Lost (2011) 24 3.354167
Jesse Stone: Benefit of the Doubt (2012) 27 3.148148
Diana Vreeland: The Eye Has to Travel (2011) 17 3.617647
Fine-Tune (2008) 1 2.000000
Sangre de mi sangre (Padre Nuestro) (2007) 2 3.250000
Milarepa (2006) 3 2.666667
Marked for Murder (1993) 0 NaN
2 Days in the Valley (1996) 3260 3.286043
0 Panishyros Megistanas Ton Ninja (2008) 4 2.125000
Morgan Pålsson - världsreporter (2008) 7 2.071429
Renoir (2012) 24 3.083333
Typhoon (Tae-poong) (2005) 13 3.038462
Last Stand (2013) 617 3.092382
Mama (2013) 686 3.131195
Candidate (2013) 5 3.200000
Upstream Color (2013) 629 3.580286
Fast Life (1932) 1 3.000000
Careless Love (2012) 2 3.000000
Special When Lit (2009) 10 3.650000
Year of the Tiger (2012) 1 2.500000
Outpost: Black Sun (2012) 21 2.142857
Preston Sturges: The Rise and Fall of an American Dreamer (1990) 1 4.500000
Double Indemnity (1973) 4 2.375000
London After Midnight (2002) 5 3.600000
Broken City (2013) 224 2.973214
Nursery University (2008) 6 4.166667
Counterfeit Coin (2012) 12 3.666667
Nameless Gangster (Bumchoiwai junjaeng) (2012) 36 3.611111
Ginger & Rosa (2012) 54 3.138889
First Shot (2002) 3 3.666667
Marina Abramovic: The Artist Is Present (2012) 65 3.576923
Bonsai (2011) 8 3.687500
Shadow Dancer (2012) 63 3.079365
Thale (2012) 44 2.977273
It's a Disaster (2012) 97 3.489691
Texas Across the River (1966) 6 3.500000
[acadgild@localhost ~]$
```

Current workspace: "Workspace 1"

```
File Edit View Search Terminal Help
Sangre de mi sangre (Padre Nuestro) (2007) 2 3.250000
Milarepa (2006) 3 2.666667
Marked for Murder (1993) 0 NaN
2 Days in the Valley (1996) 3260 3.286043
0 Panishyros Megistanas Ton Ninja (2008) 4 2.125000
Morgan Pålsson - världsreporter (2008) 7 2.071429
Renoir (2012) 24 3.083333
Typhoon (Tae-poong) (2005) 13 3.038462
Last Stand (2013) 617 3.092382
Mama (2013) 686 3.131195
Candidate (2013) 5 3.200000
Upstream Color (2013) 629 3.580286
Fast Life (1932) 1 3.000000
Careless Love (2012) 2 3.000000
Special When Lit (2009) 10 3.650000
Year of the Tiger (2012) 1 2.500000
Outpost: Black Sun (2012) 21 2.142857
Preston Sturges: The Rise and Fall of an American Dreamer (1990) 1 4.500000
Double Indemnity (1973) 4 2.375000
London After Midnight (2002) 5 3.600000
Broken City (2013) 224 2.973214
Nursery University (2008) 6 4.166667
Counterfeit Coin (2012) 12 3.666667
Nameless Gangster (Bumchoiwai junjaeng) (2012) 36 3.611111
Ginger & Rosa (2012) 54 3.138889
First Shot (2002) 3 3.666667
Marina Abramovic: The Artist Is Present (2012) 65 3.576923
Bonsai (2011) 8 3.687500
Shadow Dancer (2012) 63 3.079365
Thale (2012) 44 2.977273
It's a Disaster (2012) 97 3.489691
Texas Across the River (1966) 6 3.500000
[acadgild@localhost ~]$ hadoop fs -cat /CaseStudyOne/output/part-r-00000
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
18/08/07 22:48:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost ~]$
```

Case Study – I

Working with Movie ratings

The screenshot shows a Linux terminal window with a taskbar at the top. The taskbar includes icons for Applications, Places, System, and a search icon, along with the system clock showing 'Tue Aug 7, 10:48 PM'. The terminal window title is 'acadgild@localhost:~'. The terminal content displays a list of movie ratings with columns for movie title, year, count, and rating. The list is sorted by rating in descending order. The first few entries are: 'Sangre de mi sangre (Padre Nuestro) (2007)' with a rating of 3.250000, 'Milarepa (2006)' with 2.666667, and 'Marked for Murder (1993)' with a rating of NaN. The list continues with various other movies and their ratings. At the bottom of the terminal, there are Hadoop commands and their output: '[acadgild@localhost ~]\$ hadoop fs -cat /CaseStudyOne/output/part-r-00000' followed by SLF4J-related warnings and a warning from util.NativeCodeLoader about the native-hadoop library.

```
File Edit View Search Terminal Help
Sangre de mi sangre (Padre Nuestro) (2007) 2 3.250000
Milarepa (2006) 3 2.666667
Marked for Murder (1993) 0 NaN
2 Days in the Valley (1996) 3260 3.286043
0 Panishyros Megistanas Ton Ninja (2008) 4 2.125000
Morgan Pålsson - världsreporter (2008) 7 2.071429
Renoir (2012) 24 3.083333
Typhoon (Tae-poong) (2005) 13 3.038462
"Last Stand 617 3.092382
Mama (2013) 686 3.131195
"Candidate 5 3.200000
Upstream Color (2013) 629 3.580286
Fast Life (1932) 1 3.000000
Careless Love (2012) 2 3.000000
Special When Lit (2009) 10 3.650000
"Year of the Tiger 1 2.500000
Outpost: Black Sun (2012) 21 2.142857
Preston Sturges: The Rise and Fall of an American Dreamer (1990) 1 4.500000
Double Indemnity (1973) 4 2.375000
London After Midnight (2002) 5 3.600000
Broken City (2013) 224 2.973214
Nursery University (2008) 6 4.166667
"Counterfeit Coin 12 3.666667
Nameless Gangster (Bumchoiwaui junjaeng) (2012) 36 3.611111
Ginger & Rosa (2012) 54 3.138889
First Shot (2002) 3 3.666667
Marina Abramovic: The Artist Is Present (2012) 65 3.576923
Bonsái (2011) 8 3.687500
Shadow Dancer (2012) 63 3.079365
Thale (2012) 44 2.977273
It's a Disaster (2012) 97 3.489691
Texas Across the River (1966) 6 3.500000
[acadgild@localhost ~]$ hadoop fs -cat /CaseStudyOne/output/part-r-00000
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
18/08/07 22:48:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
```

[acadgild@localhost:~] [acadgild@localhost:~] [acadgild@localhost:~]

Applications Places System Acadgild Tue Aug 7, 10:49 PM

```
File Edit View Search Terminal Help
Cargo (2011) 3 2.333333
Jesse Stone: No Remorse (2010) 30 3.266667
Jesse Stone: Innocents Lost (2011) 24 3.354167
Jesse Stone: Benefit of the Doubt (2012) 27 3.148148
Diana Vreeland: The Eye Has to Travel (2011) 17 3.617647
Fine-Tune (2008) 1 2.000000
Sangre de mi sangre (Padre Nuestro) (2007) 2 3.250000
Milarepa (2006) 3 2.666667
Marked for Murder (1993) 0 NaN
2 Days in the Valley (1996) 3260 3.286043
0 Panishyros Megistanas Ton Ninja (2008) 4 2.125000
Morgan Pålsson - världsreporter (2008) 7 2.071429
Renoir (2012) 24 3.083333
Typhoon (Tae-poong) (2005) 13 3.038462
"Last Stand 617 3.092382
Mama (2013) 686 3.131195
"Candidate 5 3.200000
Upstream Color (2013) 629 3.580286
Fast Life (1932) 1 3.000000
Careless Love (2012) 2 3.000000
Special When Lit (2009) 10 3.650000
"Year of the Tiger 1 2.500000
Outpost: Black Sun (2012) 21 2.142857
Preston Sturges: The Rise and Fall of an American Dreamer (1990) 1 4.500000
Double Indemnity (1973) 4 2.375000
London After Midnight (2002) 5 3.600000
Broken City (2013) 224 2.973214
Nursery University (2008) 6 4.166667
"Counterfeit Coin 12 3.666667
Nameless Gangster (Bumchoiwaui junjaeng) (2012) 36 3.611111
Ginger & Rosa (2012) 54 3.138889
First Shot (2002) 3 3.666667
Marina Abramovic: The Artist Is Present (2012) 65 3.576923
Bonsái (2011) 8 3.687500
Shadow Dancer (2012) 63 3.079365
Thale (2012) 44 2.977273
It's a Disaster (2012) 97 3.489691
Texas Across the River (1966) 6 3.500000
[acadgild@localhost ~]$ hadoop fs -cat /CaseStudyOne/output/part-r-00000
```

[acadgild@localhost:~] [acadgild@localhost:~] [acadgild@localhost:~]