

Detailed steps to execute the program with screenshots of the execution results:

Create the Project Directory:

On your machine, create a directory structure as follows: Docker/Labs/Project

Place the following files in the Project directory:

- docker-compose.yml
- hadoop.env

Add Required Files:

Place the following additional files in the Project directory:

- hadoop-project-1.0-SNAPSHOT.jar
- ml-25m folder (containing movies.csv and ratings.csv)

Copy Files to HDFS:

Follow these steps to copy files from your local file system to the Hadoop Distributed File System (HDFS), this is an example with movies file:

- Create the target directory in HDFS for the dataset (ml-25m):

```
hdfs dfs -mkdir -p /ml-25m
```

- Copy the movies.csv file to the ml-25m directory in HDFS:

```
hdfs dfs -put /Docker/Labs/Project/ml-25m/movies.csv /ml-25m/
```

```
root@e0f7d2118a11:/hadoop/labs# hdfs dfs -mkdir -p /ml-25m
root@e0f7d2118a11:/hadoop/labs# ls
current docker-compose.yml hadoop-project-1.0-SNAPSHOT.jar hadoop.env in_use.lock ml-25m test
root@e0f7d2118a11:/hadoop/labs# hdfs dfs -put /hadoop/labs/ml-25m/ratings.csv /ml-25m/
put: '/ml-25m/ratings.csv': File exists
root@e0f7d2118a11:/hadoop/labs# hdfs dfs -put /hadoop/labs/ml-25m/movies.csv /ml-25m/
put: '/ml-25m/movies.csv': File exists
root@e0f7d2118a11:/hadoop/labs# hdfs dfs -put /hadoop/labs/ml-25m/movies.csv /ml-25m/
```

Navigate to the Docker/Labs/Project directory in your terminal and run docker compose up

Open a new terminal and execute the following command to enter the namenode container:

```
docker exec -it namenode bash
```

Once inside the container, navigate to the labs directory using:

```
cd hadoop/labs
```

```
PS C:\Users\maroi> docker exec -it namenode bash
root@e0f7d2118a11:/# ls
KEYS  boot  entrypoint.sh  hadoop      home  lib64  mnt  proc  run  sbin  sys  usr
bin   dev   etc            hadoop-data lib  media  opt  root  run.sh  srv  tmp  var
root@e0f7d2118a11:/# cd hadoop/labs
root@e0f7d2118a11:/hadoop/labs# ls
current                hadoop-project-1.0-SNAPSHOT.jar  in_use.lock  test
docker-compose.yml     hadoop.env                       ml-25m
root@e0f7d2118a11:/hadoop/labs# |
```

Run the HighestRatedMovie Hadoop Job:

HighestRatedMovie.java

This class finds the highest-rated movie for each user:

- RatingMapper: Processes input data in CSV format, skips header, and emits (userId, movieId:rating) pairs
- RatingReducer: For each user, finds the movie with the highest rating and outputs (userId, movieId)

Execute the following command to run:

```
hadoop dfs -rm -r /output1 ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar
com.example.hadoop.HighestRatedMovie /ml-25m/ratings.csv /output1
```

Execution results :

```
root@0f7d2118a11:/hadoop/labs# hadoop dfs -rm -r /output1 ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar com.example.hadoop.HighestRatedMovie /ml-25m/ratings.csv /output1
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Deleted /output1
2024-12-06 23:20:50,542 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.18.0.4:8032
2024-12-06 23:20:50,649 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.6:10200
2024-12-06 23:20:50,736 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-12-06 23:20:50,747 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1733526206953_0002
2024-12-06 23:20:50,802 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:20:50,895 INFO input.FileInputFormat: Total input files to process : 1
2024-12-06 23:20:50,921 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:20:51,020 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:20:51,506 INFO mapreduce.JobSubmitter: number of splits:5
2024-12-06 23:20:51,564 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:20:51,616 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1733526206953_0002
2024-12-06 23:20:51,617 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-06 23:20:51,706 INFO conf.Configuration: resource-types.xml not found
2024-12-06 23:20:51,707 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-12-06 23:20:51,938 INFO impl.YarnClientImpl: Submitted application application_1733526206953_0002
2024-12-06 23:20:51,957 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8080/proxy/application_1733526206953_0002/
2024-12-06 23:20:51,958 INFO mapreduce.Job: Running job: job_1733526206953_0002
2024-12-06 23:20:57,039 INFO mapreduce.Job: Job job_1733526206953_0002 running in uber mode : false
2024-12-06 23:20:57,043 INFO mapreduce.Job: map 0% reduce 0%
2024-12-06 23:21:11,137 INFO mapreduce.Job: map 20% reduce 0%
2024-12-06 23:21:13,144 INFO mapreduce.Job: map 40% reduce 0%
2024-12-06 23:21:14,148 INFO mapreduce.Job: map 60% reduce 0%
2024-12-06 23:21:25,188 INFO mapreduce.Job: map 80% reduce 0%
2024-12-06 23:21:26,194 INFO mapreduce.Job: map 100% reduce 0%
2024-12-06 23:21:39,223 INFO mapreduce.Job: map 100% reduce 58%
2024-12-06 23:21:45,241 INFO mapreduce.Job: map 100% reduce 74%
2024-12-06 23:21:51,259 INFO mapreduce.Job: map 100% reduce 90%
2024-12-06 23:21:52,262 INFO mapreduce.Job: map 100% reduce 100%
2024-12-06 23:21:54,277 INFO mapreduce.Job: Job job_1733526206953_0002 completed successfully
2024-12-06 23:21:54,351 INFO mapreduce.Job: Counters: 55
  File System Counters
    FILE: Number of bytes read=180117054
    FILE: Number of bytes written=271551740
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=678277891
    HDFS: Number of bytes written=1660209
    HDFS: Number of read operations=20
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Killed map tasks=1
    Launched map tasks=5
    Launched reduce tasks=1
    Rack-local map tasks=5
    Total time spent by all maps in occupied slots (ms)=225268
    Total time spent by all reduces in occupied slots (ms)=211760
    Total time spent by all map tasks (ms)=56317
    Total time spent by all reduce tasks (ms)=26470
    Total vcore-milliseconds taken by all map tasks=56317
    Total vcore-milliseconds taken by all reduce tasks=26470
```

You can see the content of output1 using command:

```
hadoop fs -cat /output1/part-r-00000
```

Let's just expose 20 lines:

```
hadoop fs -cat /output1/part-r-00000 | head -n 20
```

```
root@0f7d2118a11:/hadoop/labs# hadoop fs -cat /output1/part-r-00000 | head -n 20
2024-12-06 23:31:54,025 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
1      2632
2      318
3       50
4     5618
5       32
6     608
7     306
8       47
9        2
10     50
11     372
12     16
13     231
14     608
15     296
16     858
17    59784
18     592
19     741
20     260
```

MovieJoin.java:

This class performs a join operation between user-movie data and movie metadata:

- UserMovieMapper: Processes user-movie viewing data, emits (movieId, "USER:userId")
- MovieMapper: Processes movie metadata, skips header, emits (movieId, "MOVIE:movieName")
- JoinReducer: Joins the data by matching movies with users, outputs (userId, movieName) pairs

To run the MovieJoin job, execute the following command:

```
hadoop dfs -rm -r /output2 ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar  
com.example.hadoop.MovieJoin /output1 /ml-25m/movies.csv /output2
```

```
root@e0f7d118a11:/hadoop/labs# hadoop dfs -rm -r /output2 ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar com.example.hadoop.MovieJoin /output1 /ml-25m/movies.csv /output2
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Deleted /output2
2024-12-06 23:35:50.435 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.18.0.4:8032
2024-12-06 23:35:50.596 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.6:10200
2024-12-06 23:35:50.698 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-12-06 23:35:50.708 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1733526206953_0003
2024-12-06 23:35:50.772 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:35:50.895 INFO input.FileInputFormat: Total input files to process : 1
2024-12-06 23:35:50.916 INFO input.FileInputFormat: Total input files to process : 1
2024-12-06 23:35:50.930 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:35:50.992 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:35:51.031 INFO mapreduce.JobSubmitter: number of splits:2
2024-12-06 23:35:51.138 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-12-06 23:35:51.206 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1733526206953_0003
2024-12-06 23:35:51.286 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-06 23:35:51.381 INFO conf.Configuration: resource-types.xml not found
2024-12-06 23:35:51.381 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-12-06 23:35:51.951 INFO impl.YarnClientImpl: Submitted application application_1733526206953_0003
2024-12-06 23:35:51.974 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1733526206953_0003/
2024-12-06 23:35:51.975 INFO mapreduce.Job: Running job: job_1733526206953_0003
2024-12-06 23:35:56.022 INFO mapreduce.Job: Job job_1733526206953_0003 running in uber mode : false
2024-12-06 23:35:56.024 INFO mapreduce.Job: map 0% reduce 0%
2024-12-06 23:36:01.071 INFO mapreduce.Job: map 50% reduce 0%
2024-12-06 23:36:02.085 INFO mapreduce.Job: map 100% reduce 0%
2024-12-06 23:36:05.097 INFO mapreduce.Job: map 100% reduce 100%
2024-12-06 23:36:05.104 INFO mapreduce.Job: Job job_1733526206953_0003 completed successfully
2024-12-06 23:36:05.171 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=1712800
    FILE: Number of bytes written=4068761
    FILE: Number of read operations=8
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4698821
    HDFS: Number of bytes written=4887349
    HDFS: Number of read operations=11
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Rack-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=12868
    Total time spent by all reduces in occupied slots (ms)=12480
    Total time spent by all map tasks (ms)=3217
    Total time spent by all reduce tasks (ms)=1560
    Total vcore-milliseconds taken by all map tasks=3217
    Total vcore-milliseconds taken by all reduce tasks=1560
    Total megabyte-milliseconds taken by all map tasks=13176832
    Total megabyte-milliseconds taken by all reduce tasks=12779520
  Map-Reduce Framework
    Map input records=224965
    Map output records=224964
```

You can see the content of output2 using command:

```
hadoop fs -cat /output2/part-r-00000
```

Let's just expose 20 lines:

```
hadoop fs -cat /output2/part-r-00000 | head -n 20
```

```
2024-12-06 23:43:40,272 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
128304 Toy Story (1995)
128300 Toy Story (1995)
127653 Toy Story (1995)
127664 Toy Story (1995)
128283 Toy Story (1995)
128269 Toy Story (1995)
127715 Toy Story (1995)
18404 Toy Story (1995)
127729 Toy Story (1995)
128253 Toy Story (1995)
127746 Toy Story (1995)
18393 Toy Story (1995)
18385 Toy Story (1995)
128233 Toy Story (1995)
18136 Toy Story (1995)
18379 Toy Story (1995)
127774 Toy Story (1995)
18366 Toy Story (1995)
18151 Toy Story (1995)
18363 Toy Story (1995)
```

MovieCountGrouping.java:

This class performs two sequential MapReduce jobs:

- The first job counts how many users watched each movie:
 - MovieCountMapper: Takes input data and emits (movie_name, 1) pairs
 - MovieCountReducer: Sums up all the counts for each movie
- Second job groups movies by their view count:
 - GroupingMapper: Takes the results from first job and inverts key-value pairs
 - GroupingReducer: Groups all movies with the same view count together

To run the MovieJoin job, execute the following command:

```
hadoop dfs -rm -r /final_output ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar
com.example.hadoop.MovieCountGrouping /output2 /final_output
```

```

root@e9f7d218a11:/hadoop/laabs$ hadoop dfs -rm -r /final_output ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar com.example.hadoop.MovieCountGrouping /output2 /final_output
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

rm: /final_output: No such file or directory
2024-12-06 23:48:45,701 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.18.0.4:8032
2024-12-06 23:48:45,401 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.6:10200
2024-12-06 23:48:45,501 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-12-06 23:48:45,511 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1733526206953_0004
2024-12-06 23:48:45,570 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:45,761 INFO input.FileInputFormat: Total input files to process : 1
2024-12-06 23:48:45,792 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:45,860 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:45,896 INFO mapreduce.JobSubmitter: number of splits:1
2024-12-06 23:48:45,993 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:46,050 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1733526206953_0004
2024-12-06 23:48:46,050 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-06 23:48:46,156 INFO conf.Configuration: resource-types.xml not found
2024-12-06 23:48:46,156 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-12-06 23:48:46,393 INFO impl.VarClientImpl: Submitted application application_1733526206953_0004
2024-12-06 23:48:46,417 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8080/proxy/application_1733526206953_0004/
2024-12-06 23:48:46,418 INFO mapreduce.Job: Running job: job_1733526206953_0004
2024-12-06 23:48:50,472 INFO mapreduce.Job: Job job_1733526206953_0004 running in uber mode : false
2024-12-06 23:48:50,478 INFO mapreduce.Job: map 0% reduce 0%
2024-12-06 23:48:55,599 INFO mapreduce.Job: map 100% reduce 0%
2024-12-06 23:48:59,637 INFO mapreduce.Job: map 100% reduce 100%
2024-12-06 23:48:59,667 INFO mapreduce.Job: Job job_1733526206953_0004 completed successfully
2024-12-06 23:48:59,719 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=87100
    FILE: Number of bytes written=631979
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4887455
    HDFS: Number of bytes written=189902
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=6076
    Total time spent by all reduces in occupied slots (ms)=10840
    Total time spent by all map tasks (ms)=1519
    Total time spent by all reduce tasks (ms)=1355
    Total vcore-milliseconds taken by all map tasks=1519
    Total vcore-milliseconds taken by all reduce tasks=1355

WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=4887349
File Output Format Counters
  Bytes Written=189902
2024-12-06 23:48:59,744 INFO client.RMProxy: Connecting to ResourceManager at resourcemanager/172.18.0.4:8032
2024-12-06 23:48:59,744 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.6:10200
2024-12-06 23:48:59,752 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-12-06 23:48:59,752 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1733526206953_0005
2024-12-06 23:48:59,778 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:59,842 INFO input.FileInputFormat: Total input files to process : 1
2024-12-06 23:48:59,849 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:59,912 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:59,960 INFO mapreduce.JobSubmitter: number of splits:1
2024-12-06 23:48:59,973 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remotHostTrusted = false
2024-12-06 23:48:59,984 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1733526206953_0005
2024-12-06 23:48:59,984 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-06 23:49:00,250 INFO impl.VarClientImpl: Submitted application application_1733526206953_0005
2024-12-06 23:49:00,270 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8080/proxy/application_1733526206953_0005/
2024-12-06 23:49:00,270 INFO mapreduce.Job: Running job: job_1733526206953_0005
2024-12-06 23:49:09,359 INFO mapreduce.Job: Job job_1733526206953_0005 running in uber mode : false
2024-12-06 23:49:09,359 INFO mapreduce.Job: map 0% reduce 0%
2024-12-06 23:49:14,397 INFO mapreduce.Job: map 100% reduce 0%
2024-12-06 23:49:17,416 INFO mapreduce.Job: map 100% reduce 100%
2024-12-06 23:49:18,451 INFO mapreduce.Job: Job job_1733526206953_0005 completed successfully
2024-12-06 23:49:18,488 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=54581
    FILE: Number of bytes written=566949
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=189918
    HDFS: Number of bytes written=99979
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Rack-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=4636
    Total time spent by all reduces in occupied slots (ms)=9904
    Total time spent by all map tasks (ms)=1159
    Total time spent by all reduce tasks (ms)=1248
    Total vcore-milliseconds taken by all map tasks=1159
    Total vcore-milliseconds taken by all reduce tasks=1248
    Total megabyte-milliseconds taken by all map tasks=4747264
    Total megabyte-milliseconds taken by all reduce tasks=10223616
  Map-Reduce Framework
    Map input records=4280
    Map output records=4280
    Map output bytes=116166
    Map output materialized bytes=54573
    Input split bytes=116
    Combine input records=0

```

You can see the content of final_output using command :

```
hadoop fs -cat /final_output/part-r-00000
```



```
root@b97d311a11:/hadoop/lsab# hadoop fs -cat /final_output/part-r-00000
2024-12-06 21:52:42,319 INFO SASLDataTransferClient: SASL encryption trust check: LocalHostTrusted = false, remoteHostTrusted = false
1 Great Performances (1998) Cats (1998) [REC] (2007) Zulu (1964) Zindagi Na Milegi Dobara (2011) Ziggy Stardust and the Spiders from Mars (1972) Zero Effect (1998) Zeitgeist: Addendum (2008) Zazie dans le
metro (1960) Zathura (2003) Zach and Mihi Have a Borne (2008) Zach Galifianakis: Live at the Purple Onion (2006) Zabriskie Point (1970) Youth (2013) Your Sister's Sister (2011) Young Guns II (1990) Young Elms
tein (1988) Young & Beautiful (2012) You Might Be the Killer (2018) You Can't Take It with You (1938) Year One (2009) Kanada (1980) XXV (2007) Wrong Side Up (Pribely obyejehno silenstv) (2005) Wow! A Talking
Fish (1983) Wavelength (2009) World War Z (2013) Woodstock (1970) Woman of the Year (1962) Woman in the Dunes (Suna no onna) (1964) Woman in Gold (2015) Witness for the Prosecution (1957) Without a
Paddle (2004) Wit (2001) Wish You Were Here (1987) Winter's Bone (2010) Winter Sleepers (Winterschläfer) (2000) Winter Light (Nattvardsgästerna) (1963) Winnie the Pooh and Tigger Too (1974) Wings (1927) Windo
w to Paris (Omo v Parizu) (1994) Winchester '73 (1950) Willard (2003) Wild Hogs (2007) Wild Child (2008) Wide Awake (1998) Why We Fight (2005) Who's Harry Crumb? (1989) White Oleander (2002) White Men Can't J
ump (1992) White Heat (1949) Where the Wild Things Are (2002) Where the Heart Is (1990) Where in the World is Osama Bin Laden? (2008) Where To Invade Next (2015) Whatever Works (2009) What's Your Number? (2011
) Wet Hot American Summer (2011) West Beirut (West Beyrouth) (1998) Wend Kuuni (a.k.a. God's Gift) (1982) Welcome to Sarajevo (1997) Weekend at Bernie's II (1993) We Need to Talk About Kevin (2011) We Bought a
Zoo (2011) We Are Marshall (2006) We Are the World (1978) Washington Square (1997) Weasels (2011) Weirlock (1989) Warcraft (2016) War for the Planet of the Apes (2017) War Room (2015) War Photographer (2001) War Machine
(2017) Walking and Talking (1996) Walkabout (1971) Waiting for 'Superman' (2010) Von Ryan's Express (1965) Vive L'Amour (Ai qing wan sui) (1994) Virus (1999) Virtuosity (1995) Videodrome (1983) Victory (a.k.
a. Escape to Victory) (1981) Victoria (2015) Vice (2018) Vera Drake (2004) Venom (2018) Veer-Zaara (2004) Valley Girl (1983) Valkyrie (2008) V/H/S/2 (2013) Urban Legend (1998) Up in the Air (2008) Until the E
nd of the World (Bis ans Ende der Welt) (1991) Unthinkable (2018) Universal Soldier: The Return (1999) Unrequited: Twenty-Five Years After Willowbrook (1986) Unfaithfully Yours (1983) Unfaithful (2002) Underwe
rld: Blood Wars (2016) Underworld (1996) Undercover Brother (2002) Under the Sand (2008) Under the Hawthorn Tree (2010) Uncle Buck (1989) UHF (1989) U2: Rattle and Hum (1988) The Ni
ght Stand (2014) Two Girls and a Guy (1997) Twin Peaks (1990) Twin Falls Idaho (1999) Twin Dragons (Shuang Long hui) (1992) Twelve O'Clock High (1999) Turtles Can Fly (Lamposhtia han parvaz niknand) (2004) Tu
rner & Hoach (1993) Turbo: A Power Rangers Movie (1997) Turbo (2013) Tulip Fever (2017) Tuck Everlasting (2002) Tsotsi (2005) Truth or Dare (2010) Trouble the Water (2008) Trixie (2006) Tristan & Isolde (2006)
Trick (1999) Triangle (2009) Trial and Error (1997) Trees Lounge (1996) Trainwreck (2015) Train of Life (Train de vie) (1998) Toys (1992) Touching the Void (2003) Touch of Pink (2004) Torn Curtain (1966) Torc
h Song Trilogy (1988) Tony-Tony (1999) Top Secret (1984) Tomboy (2011) Tokyo Story (Tokyo monogatari) (1953) Tokyo Godfathers (2003) Today's Special (2009) To Sir with Love (1967) Titus (1999) Tito and Me (
Tito i ja) (1992) Titanic (1992) Tinker Bell (2008) Tin Men (1987) Time Lapse (2014) Thunderbolt and Lightfoot (1974) Thumbsucker (2005) Throne of Blood (Kumonosu jo) (1957) Three Men to Tango (1999) Three Season
s (1999) This Means War (2012) This Boy's Life (1993) Thirteen Ghosts (a.k.a. Thirteen Ghosts) (2001) Think Like a Man Too (2014) Think Like a Man (2012) Thesis (Tesis) (1996) Them (11c) (2006) The Way He Look
s (2014) The Thorn (1971) The Square (2017) The Show (1995) The Shallows (2016) The Scorpion King (2002) The Salesman (2016) The Rosenbergs: An Intimate History (2014) The Raid 2: Berandal (2014) The Price of
Gold (2014) The Players Club (1998) The Night Manager (2016) The Nacine Habits of the Earthbound Human (1995) The Man from Nowhere (2010) The Little World of Don Camillo (1952) The Little Prince (2015) The Lig
ht Between Oceans (2016) The Killing of a Sacred Deer (2017) The Jungle Book (2016) The Invisible Guest (2016) The Importance of Being Earnest (1952) The Ino (1961) The Huntsman Winter's War (2016) The Hunger
Games: Mockingjay – Part 2 (2015) The Hundred-Foot Journey (2014) The Greatest Showman (2017) The Great Raid (2005) The Good Mother (2013) The Good Dinosaur (2015) The Glass Shield (1994) The Girls (1961) The
Girl with All the Gifts (2016) The Garden of Sinners – Chapter 4: gaze-no-dou, (The Hollow Shrine) (2008) The Favourite The Fate of the Furious (2017) The Fantastic Flying Books of Mr. Morris Lessmore (2011)
The End of the Tour (2015) The Emoji Movie (2017) The Diary of a Teenage Girl (2015) The Death of Stalin (2017) The Darkest Minds (2018) The Conjuring 2 (2016) The Cloverfield Paradox (2018) The Breadwinner (2
017) The Big Sick (2017) The Best of Me (2014) The Age of Adaline (2015) The 8th Wave (2016) That Obscure Object of Desire (Cet obscur objet du desir) (1977) Terminal Velocity (1990) Tenacious D in the Pick of
Destiny (2006) Teenage Mutant Ninja Turtles: Out of the Shadows (2016) Ted (2012) Team America: World Police (2004) Teaching Mrs. Tingle (1999) Taxi (1998) Taste of Cherry (Ta'm-e guila
s) (1997) Tarzan and the Lost City (1998) Tape (2001) Talvar (2015) Talladega Nights: The Ballad of Ricky Bobby (2008) Tales from Earthsea (Gedo Senki) (2006) Taken 3 (2015) Take Me Home Tonight (2011) Take C
are of My Cat (Goyangilul butashae) (2001) T2: Trainspotting (2017) Sympathy for Mr. Vengeance (Boksuneun naui geot) (2002) Sydney White (2007) Switchblade Sisters (1976) Swiss Army Man (2016) Suing Time (199
6) Swing Kids (1993) Swept from the Sea (1997) Sweet November (1968) Sweeney Todd: The Demon Barber of Fleet Street (2007) Suspiria (2017) Superstar (1999) Superman Returns (2006) Supercop 2 (Project S) (Chao
ji ji hua) (1993) Super 8 (2011) Super (2010) Sunshine (1999) Sunrise: A Song of Two Humans (1927) Sunday Bloody Sunday (1971) Summer of '42 (1971) Sully (2016) Suicide Squad (2016) Sugar
& Spice (2001) Stuck in Love (2012) Stroszek (1977) Striking Distance (1993) Street Thief (2006) Street Fighter (1994) Street Fight (2005) Stray Dog (2014) Stranger by the Lake (L'inconnu du lac) (2013) Strang
e Brew (1983) Straight Outta Compton (2015) Stop! Or My Mom Will Shoot (1992) Stolen Misses (Baisers volés) (1968) Stevie (2002) Steamboat Willie (1928) Stay (2009) Starkey & Hutch (2004)
Stardom (2008) Standard Operating Procedure (2008) Spy Kids 3-D: Game Over (2003) Spud (2014) Spread (2009) Spice World (1997) Spectre (2015) Species II (1998) Spanking the Monkey (1994) Spanglish (2004) Space
Cowboys (2008) Space Grapes (1998) Sonic Outlaws (1995) Song of the South (1946) Song of the Sea (2014) Son of God (2014) Son of Frankenstein (1939) Son of Flubber (1963) Sommersby (1993) Something's Gotta Gi
ve (2008) Someone's Daughter (2012) Somebody is Waiting (1986) Some Mother's son (1996) Some Kind of Wonderful (1997) Solomon Kane (2009) Soler: A Star Wars Story (2018) Solas (1999) Soundbitch
(1991) So Close (Chik Yeung Tin Sai) (2002) Snoeflake (2017) Snow-White and Rose-Red (1979) Snow Falling on Cedars (1999) Smurf!: The Lost Village (2017) Sleepwalk with Me (2012) Sleepever (2004) Slam (1998)
Six-String Samurai (1998) Six Days Seven Nights (1998) Simon of the Desert (Simón del desierto) (1965) Sidelwits (Medianeras) (2011) Shrek the Third (2007) Shortbus (2006) SherryBaby (2006) Shelter (2007) Sho
's so Lovely (1997) Shut of My League (2018) Shaolin Soccer (Shu lan jiu hau) (2001) Shaq (1989) Shadow of the Vampire (2000) Shadow of the Thin Man (1941) Sexual Predator (2001) Sex and the City 2 (2010)
Sex Is Comedy (2002) Seven Brides for Seven Brothers (1954) Seven Beauties (Pasqualine Settebellezza) (1976) Session 7: The Contenders (2001) Senki szigete (2014) Selma (2014) Selena (1997) Sea
rching for Sugar Man (2012) Seance on a Wet Afternoon (1964) Sea of Love (1989) Scream of Stone (Carro torpe: Schrei aus Stein) (1991) Scarface (1922) Saw VI (2009) Saw II (2006) Saving Grace
(2008) Saturday Night Fever (1977) Sanjiv (Tsubaki Sanjiv) (1962) San Andreas (2015) Salvador (1986) Salt (2010) Safety Not Guaranteed (2012) Saboteur (1942) Saboteur (1936) S.P.I.L.: Kill Zone (Sant po long)
(2008) Rust and Bone (De rouille et d'os) (2012) Running on Empty (1988) Running Scared (2006) Ruby & Quentin (Tais-tout!) (2003) Rocketta (1999) Room 237 (2012) Romeo and Juliet (1968) Romeo Must Die (2008) Rol
e Models (2008) Rocky Balboa (2006) Rocket Science (2007) Robots (2005) Robots 2 (1999) Robin Williams: Weapons of Self Destruction (2009) Robin Hood (2010) Ringo (1998) Riding in Cars with
h Boys (2001) Ridicule (1996) Ride the High Country (1962) Richard Pryor: Live in Concert (1979) Richard Pryor Live on the Sunset Strip (1982) Revolutionary Road (2008) Revenge of the Nerds (1984) Return to th
e Blue Lagoon (1991) Return to Never Land (2002) Return of the Killer Tomatoes! (1988) Retrospective (1997) Resurrection Man (1998) Resident Evil: Retribution (2012) Resident Evil: Afterlife (2010) Resident Evil
(2002) Requiem for a Heavyweight (1962) Rent-a-Kid (1995) Renaissance Man (1994) Remember Me (2010) Relax... It's Just Sex (1998) Reign Over Me (2007) Regret to Inform (1998) Reel Injun (2009) Red Scorpion (Ho
ng ga liang) (1987) Red Sonja (1985) Red Dawn (1986) Recess: School's Out (2001) Real Steel (2011) Ready to Wear (Prêt-à-porter) (1998) Ready Player One Re-Animator (1985) Ravenous (1999) Rare Exports: A Chr
istmas Tale (Barns jule (2010) Rapa Nui (1992) Rango (2011) Rambo III (1988) Raise Your Voice (2004) Race to Mars (2007) R.I.P.D. (2013) Queen of the Damned (2002) Queen of Hearts (1999) Quatermass and the
Pit (1967) Quadrophonia (1979) Pushing Hands (Tui shou) (1992) Pumpkin (2002) Pump Up the Volume (1998) Public Enemy (2009) Prozac Nation (2001) Proof of Life (2006) Promises (2008) Project X (2012) Project
Almanac (2013) Project X: The Movie (2012) Princes of the Mountains (Kovdarsky plenilo) (1960) Prince of Persia: The Sands of Time (2010) Prince of Darkness (1987) Prick Up Your Ears (1987)
Premium Rush (2012) Prefontaine (1997) Prayers for Bobby (2009) Porky's II: The Next Day (1983) Popstar: Never Stop Never Stopping (2016) Pontypool (2008) Polyester (1981) Pollyanna (1968) Polish Wedding (1998
) Police Academy (1989) Police Academy 5: Assignment: Miami Beach (1988) Police Academy 3: Back in Training (1986) Pokémon: The First Movie (1998) Pokémon the Movie 2000 (2000) Point Break
(1991) Point Blank (A bout portant) (2001) Playing God (1997) Play Mixy for Me (Next Day) (2011) plane of the Ases (2001) Planet Terror (2007) Pitch Perfect 2 (2015) Pirates of the Caribbean: On Stranger Tides (2011) p
```

- 360 Saving Private Ryan (1998)
- 363 Interview with the Vampire: The Vampire Chronicles (1994)
- 370 American Beauty (1999)
- 378 Ace Ventura: When Nature Calls (1995)
- 385 "Fugitive
- 391 Crimson Tide (1995)
- 392 Dumb & Dumber (Dumb and Dumber) (1994)
- 415 Good Will Hunting (1997)
- 423 Before Sunrise (1995)
- 429 2001: A Space Odyssey (1968)
- 458 One Flew Over the Cuckoo's Nest (1975)
- 472 Die Hard: With a Vengeance (1995)
- 494 Happy Gilmore (1996)
- 495 Grumpier Old Men (1995)
- 496 Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)
- 510 Monty Python and the Holy Grail (1975)
- 511 Casablanca (1942)
- 617 Sabrina (1995)
- 638 "Princess Bride
- 655 Jurassic Park (1993)
- 685 "Lord of the Rings: The Fellowship of the Ring
- 702 "Postman
- 739 Star Wars: Episode V - The Empire Strikes Back (1980) Clueless (1995)
- 745 Terminator 2: Judgment Day (1991)
- 774 Clerks (1994)
- 786 "Lion King
- 844 Jumanji (1995)
- 851 Fargo (1996)
- 854 Fight Club (1999)
- 883 Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1964)
- 898 Mr. Holland's Opus (1995)
- 1079 "City of Lost Children
- 1116 GoldenEye (1995)
- 1136 Get Shorty (1995)
- 1202 "American President
- 1205 "Matrix
- 1339 Blade Runner (1982)
- 1349 "Silence of the Lambs
- 1357 Dead Man Walking (1995)
- 1419 Casino (1995)
- 1425 Léon: The Professional (a.k.a. The Professional) (Léon) (1994)
- 1616 Leaving Las Vegas (1995)
- 1968 Apollo 13 (1995)
- 2080 "Godfather
- 2130 Schindler's List (1993)
- 2142 Babe (1995)
- 2177 Taxi Driver (1976)
- 3053 Heat (1995)
- 3057 Sense and Sensibility (1995)
- 3109 Forrest Gump (1994)
- 4146 Twelve Monkeys (a.k.a. 12 Monkeys) (1995)
- 5032 Seven (a.k.a. Se7en) (1995)
- 5251 Braveheart (1995)
- 5925 Pulp Fiction (1994)
- 6524 Star Wars: Episode IV - A New Hope (1977)
- 7167 "Usual Suspects
- 8090 Toy Story (1995)
- 8158 "Shawshank Redemption

You can access the final file by saving it locally using the command:

```
hdfs dfs -get /final_output/part-r-00000 /hadoop/labs/part-r-00000
```

```
root@e0f7d2118a11:/hadoop/labs# hdfs dfs -get /final_output/part-r-00000 /hadoop/labs/part-r-00000
2024-12-06 23:55:39,264 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
```

You will find it in your Directory under the name part-r-00000 (results)

current	06/12/2024 14:46	File folder	
ml-25m	05/12/2024 08:26	File folder	
test	06/12/2024 22:31	File folder	
docker-compose.yml	06/12/2024 14:41	Yaml Source File	3 KB
hadoop.env	06/12/2024 14:25	ENV File	3 KB
hadoop-project-1.0-SNAPSHOT.jar	06/12/2024 22:20	JAR File	17 KB
in_use.lock	07/12/2024 00:02	LOCK File	1 KB
part-r-00000	07/12/2024 00:55	File	98 KB

N.B:

If you get any error related to a temporary file when launching a Hadoop command, just delete it. Here is an example of such situation:

```
root@e0f7d2118a11:/hadoop/labs# hadoop dfs -rm -r /final_output ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar com.example.hadoop.MovieCountGrouping /output2 /final_output
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Deleted /final_output
2024-12-08 11:13:27,493 INFO client.RMPProxy: Connecting to ResourceManager at resourcemanager/172.18.0.6:8032
2024-12-08 11:13:27,594 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.5:10200
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://namenode:9000/final_output_temp already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:164)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:277)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1730)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1588)
    at com.example.hadoop.MovieCountGrouping.main(MovieCountGrouping.java:124)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
root@e0f7d2118a11:/hadoop/labs# hadoop dfs -rm -r /final_output_temp
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Deleted /final_output_temp
root@e0f7d2118a11:/hadoop/labs# hadoop dfs -rm -r /final_output ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar com.example.hadoop.MovieCountGrouping /output2 /final_output
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

rm: '/final_output': No such file or directory
2024-12-08 11:14:19,856 INFO client.RMPProxy: Connecting to ResourceManager at resourcemanager/172.18.0.6:8032
2024-12-08 11:14:19,964 INFO client.AHSProxy: Connecting to Application History server at historyserver/172.18.0.5:10200
2024-12-08 11:14:20,052 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-12-08 11:14:20,063 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1733656172959_0003
2024-12-08 11:14:20,117 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
```


In this case, we try to execute the following command:

```
hadoop dfs -rm -r /final_output ; hadoop jar hadoop-project-1.0-SNAPSHOT.jar  
com.example.hadoop.MovieCountGrouping /output2 /final_output
```

The error shows the following:

```
Output directory hdfs://namenode:9000/final_output_temp already exists
```

We proceed with the deletion of the temporary file:

```
hadoop dfs -rm -r /final_output_temp
```

Re-execute your command and it will be launched correctly.