

# MOVIES DATASET

[from MovieLens; a movie recommendation service]

## MAP-REDUCE IN HADOOP

MapReduce is a framework using which we can write applications to process huge amounts of data, in parallel, on large clusters of commodity hardware in a reliable manner.

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes.

**MAP:-** Hadoop's Mapper class implements org.apache.hadoop.mapreduce.Mapper. It has input key type, input value type, output key type and output value type. From org.apache.hadoop.io package, data types are optimized for network serialization. Main work is done by the map() method. A line of text is extracted from the given file, converted to a Java String type, and then parsed to get the required.

**REDUCE:-** Hadoop's Reducer class implements org.apache.hadoop.mapreduce.Reducer. Its input key and value types must match Mapper output key and value types. Work is done by reduce() method. It inputs values passed as iterable list, goes over all values to find/calculate the required, and the result pair is written by using the Context onto the HDFS(Hadoop's Distributed File System).

## DATASET:- Totally consists of 9742 movie records.

Jun 13 03:33 movies.csv - LibreOffice Calc

1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
2	Jumanji (1995)	Adventure Children Fantasy
3	Grumpier Old Men (1995)	Comedy Romance
4	Death Becomes Her (1993)	Comedy Drama Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action Crime Thriller
7	Sabrina (1995)	Comedy Romance
8	Tom andam (1995)	Action Adventure Thriller
9	One Fine Day (1995)	Comedy Drama Romance
10	GoldenEye (1995)	Comedy Horror
11	American President, The (1995)	Adventure Animation Children
12	Dracula: Dead and Loving It (1995)	Drama
13	13 Going on 30 (1995)	Action Adventure Romance
14	Never Say Never Again (1995)	Crime Drama
15	Cuthbert Island (1995)	Drama Romance
16	Casino (1995)	Crime Drama
17	Death Becomes Her (1995)	Comedy Romance
18	Pearl Harbor (1995)	Comedy
19	Ace Ventura: When Nature Calls (1995)	Action Comedy Crime Drama Thriller
20	Money Train (1995)	Comedy Crime Thriller
21	Get Shorty (1995)	Comedy Drama Horror Mystery Thriller
22	Coyote Ugly (1996)	Action Comedy Thriller
23	Assault on Precinct 13 (1995)	Drama Sci-Fi
24	Provocer (1995)	Drama Romance
25	Leaving Las Vegas (1995)	Drama
26	Othello (1995)	Children Drama
27	Now and Then (1995)	Drama Romance
28	Concussion (1995)	Adventure Drama Fantasy Mystery Sci-Fi
29	City of Lost Children, The (Clé des enfants perdus, La) (1995)	Crime Drama
30	Shanghai Triad (Yao a yao diao weipo qiao) (1995)	Drama
31	Dangerous Minds (1995)	Mystery Sci-Fi Thriller
32	12 Monkeys (a.k.a. 12 Monkeys) (1995)	Children Drama
33	48 (1995)	Crime Drama
34	Dead Man Walking (1995)	Crime Drama
35	It Takes Two (1995)	Children Comedy
36	Clueless (1995)	Comedy Romance
37	40 Days and 40 Nights (1995)	Drama War
38	Richard III (1995)	Action Crimetime Drama
39	42 Dead Presidents (1995)	Drama
40	Restoration (1995)	Action Adventure Fantasy
41	44 French Kiss (1995)	Comedy Drama Thriller
42	To Die For (1995)	Drama Romance
43	How to Make an American Quilt (1995)	Mystery Thriller
44	Seven (a.k.a. Se7en) (1995)	Animation Children Drama Musical Romance
45	Pocahontas (1995)	Drama Romance
46	40 Days and 40 Nights (1995)	Crime Mystery Thriller
47	50 Usual Suspects, The (1995)	Drama Romance
48	55 Georgia (1995)	Comedy Drama Romance
49	Larry Crowne (1994)	Adventure Drama
50	54 Big Green, The (1995)	Children Comedy
51	55 Georgia (1995)	Drama
52	57 Home Alone 2: Lost in New York (1995)	Comedy Drama Romance
53	58 Postman, The (Postmen, The) (1994)	Adventure Children Fantasy
54	60 Indian in the Cupboard, The (1995)	Drama Thriller
55	61 Eye for an Eye (1996)	Drama
56	62 Mr. Holland's Opus (1995)	Comedy Crime
57	63 Don't Be a Menace to South Central While Drinking Your Juice in the Hood (1996)	Comedy Romance
58	64 If These Walls Could Talk (1996)	Comedy
59	65 Big Dumb (1996)	Action Sci-Fi Thriller
60	66 Lawncutter Man 2: Beyond Cyberspace (1996)	Comedy Romance
61	68 French Twist (Gazon maudit) (1995)	Comedy
62	69 Friday (1995)	Action Comedy Horror Thriller
63	70 Friday on My Mind, Till Dawn (1996)	Action
64	71 Fair Game (1995)	Comedy Drama
65	72 Kicking and Screaming (1995)	Drama Romance
66	73 Mistakes, Les (1995)	Comedy Drama
67	74 Color of Money, The (1996)	Drama Romance
68	75 Big Bubba (1996)	Comedy Drama
69	76 Screamers (1995)	Action Sci-Fi Thriller
70	77 Nico Icon (1995)	Documentary
71	78 Crossing Jordan, The (1995)	Action Comedy Drama Thriller
72	79 The Siege (1998)	Drama Thriller
73	80 White Balloon, The (Bakhonake seifid) (1995)	Children Drama
74	81 Things to Do in Denver When You're Dead (1995)	Crime Drama Romance
75	82 Antonito's Line (Antonito) (1995)	Comedy Drama
76	83 Color of Money, The... When We Were Colored (1995)	Drama Romance
77	85 Angels and Insects (1995)	Drama Romance
78	86 White Squid (1996)	Action Adventure Drama
79	87 Dunston Checks In (1996)	Children Comedy
80	88 Black Sheep (1996)	Comedy
81	89 Nick of Time (1995)	Action Thriller

Jun 13 03:33 movies.csv - LibreOffice Calc

34	36 Dead Man Walking (1995)	Crime Drama
35	38 It Takes Two (1995)	Children Comedy
36	39 Clueless (1995)	Comedy Romance
37	40 Cry, the Beloved Country (1995)	Drama
38	41 Richard III (1995)	Drama War
39	42 Dead Presidents (1995)	Action Crimetime Drama
40	43 Restoration (1995)	Drama
41	44 French Kiss (1995)	Action Adventure Fantasy
42	45 To Die For (1995)	Comedy Drama Thriller
43	46 How to Make an American Quilt (1995)	Drama Romance
44	47 Seven (a.k.a. Se7en) (1995)	Mystery Thriller
45	48 Pocahontas (1995)	Animation Children Drama Musical Romance
46	49 Mystery of the Living (1995)	Drama Romance
47	50 Usual Suspects, The (1995)	Crime Mystery Thriller
48	52 Mystery of the Elite (1995)	Comedy Drama Romance
49	53 Laramie (1994)	Adventure Drama
50	54 Big Green, The (1995)	Children Comedy
51	55 Georgia (1995)	Drama
52	57 Home Alone 2: Lost in New York (1995)	Comedy Drama Romance
53	58 Postman, The (Postmen, The) (1994)	Adventure Children Fantasy
54	60 Indian in the Cupboard, The (1995)	Drama Thriller
55	61 Eye for an Eye (1996)	Drama
56	62 Mr. Holland's Opus (1995)	Comedy Crime
57	63 Don't Be a Menace to South Central While Drinking Your Juice in the Hood (1996)	Comedy Romance
58	64 If These Walls Could Talk (1996)	Comedy
59	65 Big Dumb (1996)	Action Sci-Fi Thriller
60	66 Lawncutter Man 2: Beyond Cyberspace (1996)	Comedy Romance
61	68 French Twist (Gazon maudit) (1995)	Comedy
62	69 Friday (1995)	Action Comedy Horror Thriller
63	70 Friday on My Mind, Till Dawn (1996)	Action
64	71 Fair Game (1995)	Comedy Drama
65	72 Kicking and Screaming (1995)	Drama Romance
66	73 Mistakes, Les (1995)	Comedy Drama
67	74 Color of Money, The (1996)	Drama Romance
68	75 Big Bubba (1996)	Comedy Drama
69	76 Screamers (1995)	Action Sci-Fi Thriller
70	77 Nico Icon (1995)	Documentary
71	78 Crossing Jordan, The (1995)	Action Comedy Drama Thriller
72	79 The Siege (1998)	Drama Thriller
73	80 White Balloon, The (Bakhonake seifid) (1995)	Children Drama
74	81 Things to Do in Denver When You're Dead (1995)	Crime Drama Romance
75	82 Antonito's Line (Antonito) (1995)	Comedy Drama
76	83 Color of Money, The... When We Were Colored (1995)	Drama Romance
77	85 Angels and Insects (1995)	Drama Romance
78	86 White Squid (1996)	Action Adventure Drama
79	87 Dunston Checks In (1996)	Children Comedy
80	88 Black Sheep (1996)	Comedy
81	89 Nick of Time (1995)	Action Thriller

## GITHUB LINK:-

[https://github.com/1NT18IS155/1NT18IS155\\_somya\\_C\\_bdLab/tree/main/bdlab/Movies\\_MapReduce](https://github.com/1NT18IS155/1NT18IS155_somya_C_bdLab/tree/main/bdlab/Movies_MapReduce)

## PROGRAM CODE:-

```
package moviesGenre;

import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;

public class moviesGenre
{
    //MAPPER CODE

    public static class Map extends MapReduceBase implements Mapper<LongWritable, Text,
    Text, IntWritable>
    {
        private final static IntWritable one = new IntWritable(1);

        public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable>
        output, Reporter reporter) throws IOException
        {
            String myMovies = value.toString();

            String myGen = myMovies.substring((myMovies.lastIndexOf(","))+1); /*Extracting
            only the substring after the last comma(to ensure that it takes only the genre, and not a part of
            movie name separated by a comma)*/
            String[] filGens = myGen.split("\\|");
        }
    }
}
```

```

        for(int i=0;i<filGens.length;i++) //segregating "action|thriller" into "action" and
        "thriller"
    {
        output.collect(new Text(filGens[i]), one); //Comedy,1      Action,1
    }

}

```

#### //REDUCER CODE

```

public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable,
Text, IntWritable>
{
    public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text,
IntWritable> output, Reporter reporter) throws IOException
    {
        int finalMovieCount = 0 ;

        Text mykey = key ;

        while(values.hasNext())
        {
            IntWritable value = values.next();

            finalMovieCount += value.get(); //reducing it to total count per particular
genre

        }

        output.collect(mykey, new IntWritable(finalMovieCount)); //comedy,3
action,5 (example)

    }
}

```

```
//DRIVER CODE

public static void main(String[] args) throws Exception
{
    JobConf conf = new JobConf(moviesGenre.class);

    conf.setJobName("moviesGenre");

    conf.setOutputKeyClass(Text.class);

    conf.setOutputValueClass(IntWritable.class);

    conf.setMapperClass(Map.class);

    conf.setCombinerClass(Reduce.class);

    conf.setReducerClass(Reduce.class);

    conf.setInputFormat(TextInputFormat.class);

    conf.setOutputFormat(TextOutputFormat.class);

    FileInputFormat.setInputPaths(conf, new Path(args[0]));

    FileOutputFormat.setOutputPath(conf, new Path(args[1]));

    JobClient.runJob(conf);

}

}
```

## OUTPUT SNAPSHOTS:-

```
Activities Terminal Jun 13 02:53
hadoop@sonyarsrv16:~/hadoop-3.2.2/sbin$ hdfs dfs -mkdir /movleshadoop
hadoop@sonyarsrv16:~/hadoop-3.2.2/sbin$ hdfs dfs -copyFromLocal /home/sonya/bdlab/movies_dataset/movies.csv /movleshadoop
hadoop@sonyarsrv16:~/hadoop-3.2.2/sbin$ hadoop fs -copyFromLocal /home/sonya/bdlab/moviesGenre /home/hadoop/hadoop-3.2.2/sbin
2021-06-13 02:50:32.326 INFO Client: RMProxy: Connecting to ResourceManager at /127.0.0.1:8033
2021-06-13 02:50:32.749 INFO Client: RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-06-13 02:50:33.964 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-06-13 02:50:33.514 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-13 02:50:33.514 INFO mapred.FileInputFormat: Input file path: /tmp/hadoop-yarn/staging/hadoop/_staging/job_1623524194233_0008/movies.csv
2021-06-13 02:50:34.099 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/_staging/job_1623524194233_0008
2021-06-13 02:50:33.514 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-13 02:50:34.101 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1623524194233_0008
2021-06-13 02:50:34.429 INFO conf.Configuration: resource-types.xml not found
2021-06-13 02:50:34.430 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-06-13 02:50:34.529 INFO impl.YarnClientImpl: Submitted application[application_1623524194233_0008]
2021-06-13 02:50:34.560 INFO mapreduce.Job: Running job: job_1623524194233_0008
2021-06-13 02:50:42.747 INFO mapreduce.Job: Job: job_1623524194233_0008 running in uber mode : false
2021-06-13 02:50:42.750 INFO mapreduce.Job: map 0% reduce 0%
2021-06-13 02:50:48.883 INFO mapreduce.Job: map 100% reduce 0%
2021-06-13 02:50:54.946 INFO mapreduce.Job: map 100% reduce 100%
2021-06-13 02:50:54.946 INFO mapreduce.Job: Job: job_1623524194233_0008 completed successfully
2021-06-13 02:50:57.119 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=57
FILE: Number of bytes written=705045
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=488957
HDFS: Number of bytes written=253
HDFS: Number of read operations=1
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
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=8103
Total time spent by all reduces in occupied slots (ms)=3581
Total time spent by all map tasks (ms)=8103
Total time spent by all reduce tasks (ms)=3581
Total vcore-milliseconds taken by all map tasks=8103
Total vcore-milliseconds taken by all reduce tasks=3581
Total megabyte-milliseconds taken by all map tasks=297472
Total megabyte-milliseconds taken by all reduce tasks=3666944
Map-Reduce Framework
Map Input records=2742
Map output records=22084
Map output bytes=252958
Map output materialized bytes=563
Input split bytes=194
Combine input records=2084
Combine output records=39
Reduce input groups=20
Reduce input records=563
Reduce output records=20
Reduce output bytes=578
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=262
CPU time spent (ms)=1650
Physical memory (bytes) snapshot=884854784
Virtual memory (bytes) snapshot=673647104
Total committed heap usage (bytes)=748158976
Peak Map Physical memory (bytes)=348975104
Peak Map Virtual memory (bytes)=2556149760
Peak Reduce Physical memory (bytes)=190398464
Peak Reduce Virtual memory (bytes)=2562764800
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
MOVED=0
MOVED_TO_OTHER_FRAME=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=488763
File Output Format Counters
Bytes Written=253
hadoop@sonyarsrv16:~/hadoop-3.2.2/sbin$
```

```
Activities Terminal Jun 13 02:53
hadoop@sonyarsrv16:~/hadoop-3.2.2/sbin$ hadoop fs -copyFromLocal /home/sonya/bdlab/moviesGenre /home/hadoop/hadoop-3.2.2/sbin
2021-06-13 02:50:32.326 INFO Client: RMProxy: Connecting to ResourceManager at /127.0.0.1:8033
2021-06-13 02:50:32.749 INFO Client: RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-06-13 02:50:33.964 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-06-13 02:50:33.514 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-13 02:50:33.514 INFO mapred.FileInputFormat: Input file path: /tmp/hadoop-yarn/staging/hadoop/_staging/job_1623524194233_0008/movies.csv
2021-06-13 02:50:34.099 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/_staging/job_1623524194233_0008
2021-06-13 02:50:33.514 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-13 02:50:34.101 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1623524194233_0008
2021-06-13 02:50:34.429 INFO conf.Configuration: resource-types.xml not found
2021-06-13 02:50:34.430 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-06-13 02:50:34.529 INFO impl.YarnClientImpl: Submitted application[application_1623524194233_0008]
2021-06-13 02:50:34.560 INFO mapreduce.Job: Running job: job_1623524194233_0008
2021-06-13 02:50:42.747 INFO mapreduce.Job: Job: job_1623524194233_0008 running in uber mode : false
2021-06-13 02:50:42.750 INFO mapreduce.Job: map 0% reduce 0%
2021-06-13 02:50:48.883 INFO mapreduce.Job: map 100% reduce 0%
2021-06-13 02:50:54.946 INFO mapreduce.Job: map 100% reduce 100%
2021-06-13 02:50:54.946 INFO mapreduce.Job: Job: job_1623524194233_0008 completed successfully
2021-06-13 02:50:57.119 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=57
FILE: Number of bytes written=705045
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=488957
HDFS: Number of bytes written=253
HDFS: Number of read operations=1
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
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=8103
Total time spent by all reduces in occupied slots (ms)=3581
Total time spent by all map tasks (ms)=8103
Total time spent by all reduce tasks (ms)=3581
Total vcore-milliseconds taken by all map tasks=8103
Total vcore-milliseconds taken by all reduce tasks=3581
Total megabyte-milliseconds taken by all map tasks=297472
Total megabyte-milliseconds taken by all reduce tasks=3666944
Map-Reduce Framework
Map Input records=2742
Map output records=22084
Map output bytes=252958
Map output materialized bytes=563
Input split bytes=194
Combine Input records=2084
Combine output records=39
Reduce input records=39
Reduce input groups=20
Reduce shuffle bytes=563
Reduce input records=39
Reduce output records=20
Reduce output bytes=578
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=262
CPU time spent (ms)=1650
Physical memory (bytes) snapshot=884854784
Virtual memory (bytes) snapshot=673647104
Total committed heap usage (bytes)=748158976
Peak Map Physical memory (bytes)=348975104
Peak Map Virtual memory (bytes)=2556149760
Peak Reduce Physical memory (bytes)=190398464
Peak Reduce Virtual memory (bytes)=2562764800
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
MOVED=0
MOVED_TO_OTHER_FRAME=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=488763
File Output Format Counters
Bytes Written=253
hadoop@sonyarsrv16:~/hadoop-3.2.2/sbin$
```

Activities Firefox Web Browser ▾ Somya Srivastava - MAPR ▾ Browsing HDFS Jun 13 03:51

localhost:9870/explorer.html#/ Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

### Browse Directory

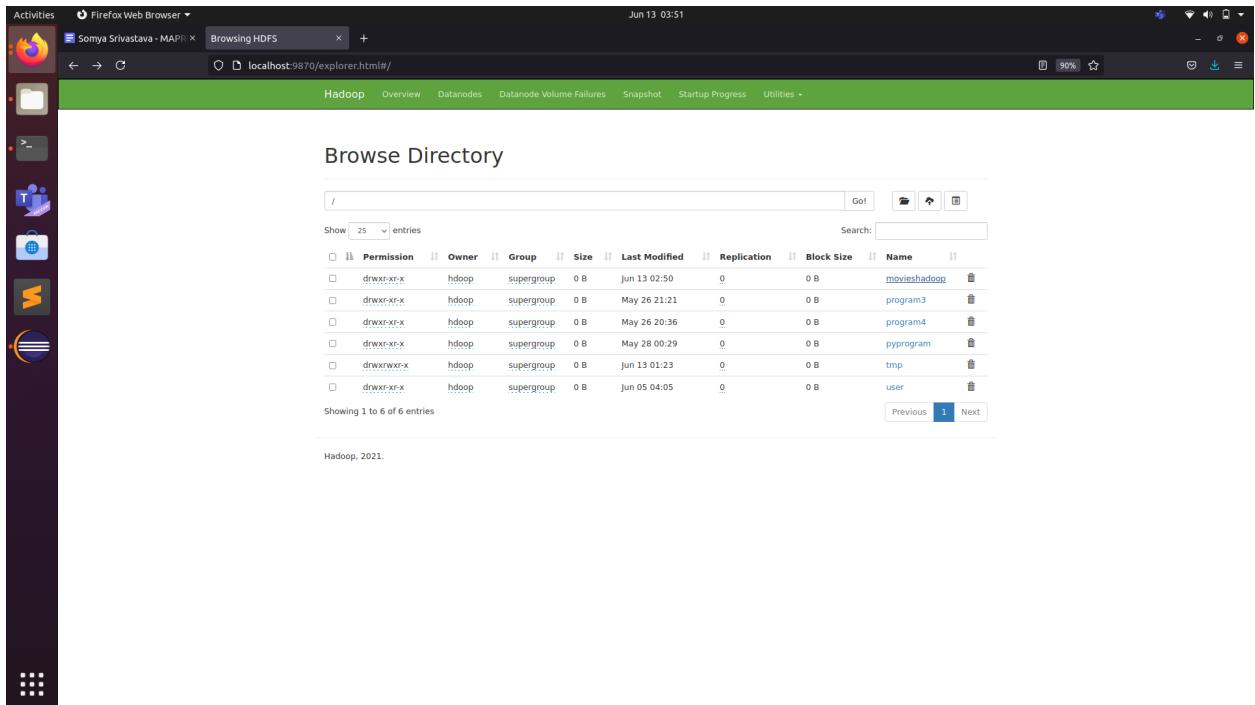
/

Show 25 entries Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	hdooop	supergroup	0 B	Jun 13 02:50	0	0 B	movieshadoop
□	drwxr-xr-x	hdooop	supergroup	0 B	May 26 21:21	0	0 B	program3
□	drwxr-xr-x	hdooop	supergroup	0 B	May 26 20:36	0	0 B	program4
□	drwxr-xr-x	hdooop	supergroup	0 B	May 28 00:29	0	0 B	pyprogram
□	drwxrwxr-x	hdooop	supergroup	0 B	Jun 13 01:23	0	0 B	tmp
□	drwxr-xr-x	hdooop	supergroup	0 B	Jun 05 04:05	0	0 B	user

Showing 1 to 6 of 6 entries Previous 1 Next

Hadoop, 2021.



Activities Firefox Web Browser ▾ Somya Srivastava - MAPR ▾ Browsing HDFS Jun 13 03:52

localhost:9870/explorer.html#/movieshadoop Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

### Browse Directory

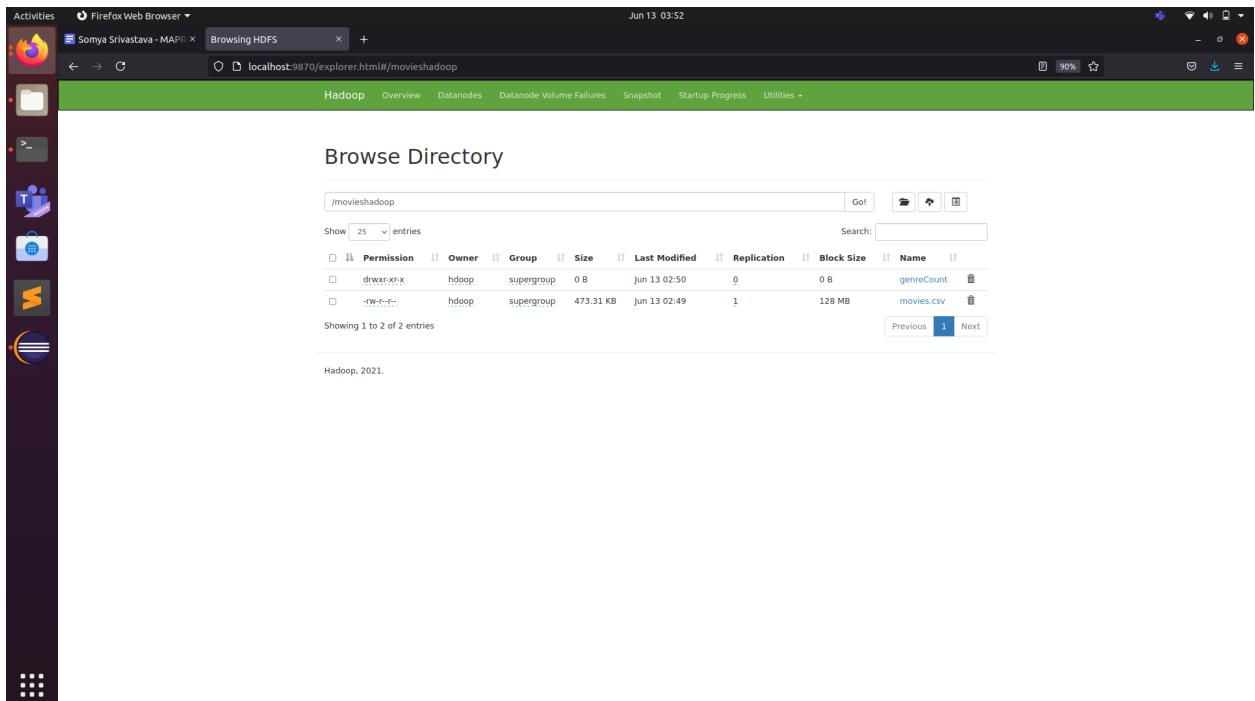
/movieshadoop

Show 25 entries Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	hdooop	supergroup	0 B	Jun 13 02:50	0	0 B	genreCount
□	-rw-r--r--	hdooop	supergroup	473.31 KB	Jun 13 02:49	1	120 MB	movies.csv

Showing 1 to 2 of 2 entries Previous 1 Next

Hadoop, 2021.



Activities Firefox Web Browser Somya Srivastava - MAPR Browsing HDFS Jun 13 03:52

localhost:9870/explorer.html#/movieshadoop

Hadoop Overview Datanodes

Browse Director

/movieshadoop

Show 25 entries

Permission Owner

drwxr-xr-x hdoop

drwxr-xr-x hdoop

Showing 1 to 2 of 2 entries

Hadoop, 2021.

File information - movies.csv

Download Head the file (first 32K) Tail the file (last 32K)

Block information - Block 0

Block ID: 1073741942

Block Pool ID: BP-436010109-127.0.1.1-1619457295229

Generation Stamp: 1118

Size: 48467

Availability:

- somyasriv16

File contents

1.Toy Story (1995).Adventure|Animation|Children|Comedy|Fantasy  
2.Jumanji (1995).Adventure|Comedy|Fantasy  
3.The Lion King (1994).Animation|Comedy|Romance  
4.Walking to Exhale (1995).Comedy|Drama|Romance  
5.Father of the Bride Part II (1995).Comedy  
6.Heat (1995).Action|Crime|Thriller  
7.Sabrina (1995).Comedy|Romance  
8.Tom and Huck (1995).Adventure|Children

Close

This screenshot shows the HDFS browser interface in a Firefox window. The main pane displays a directory listing for '/movieshadoop'. A modal window is open for the file 'movies.csv', providing detailed information about its first block. The file's size is 48,467 bytes, and it was last modified on June 13, 2021. The file contents are listed below, showing a sample of movie titles and genres.

Activities Firefox Web Browser Somya Srivastava - MAPR Browsing HDFS Jun 13 03:52

localhost:9870/explorer.html#/movieshadoop/genreCount

Hadoop Overview Datanodes

Browse Director

/movieshadoop/genreCount

Show 25 entries

Permission Owner

drwxr-xr-x hdoop

drwxr-xr-x hdoop

Showing 1 to 2 of 2 entries

Hadoop, 2021.

File information - part-00000

Download Head the file (first 32K) Tail the file (last 32K)

Block information - Block 0

Block ID: 1073741949

Block Pool ID: BP-436010109-127.0.1.1-1619457295229

Generation Stamp: 1125

Size: 253

Availability:

- somyasriv16

File contents

(no genres listed) 34

Genre	Count
Action	1828
Adventure	1263
Animation	611
Comics	664
Comedy	3756
Crime	1199
Documentary	440

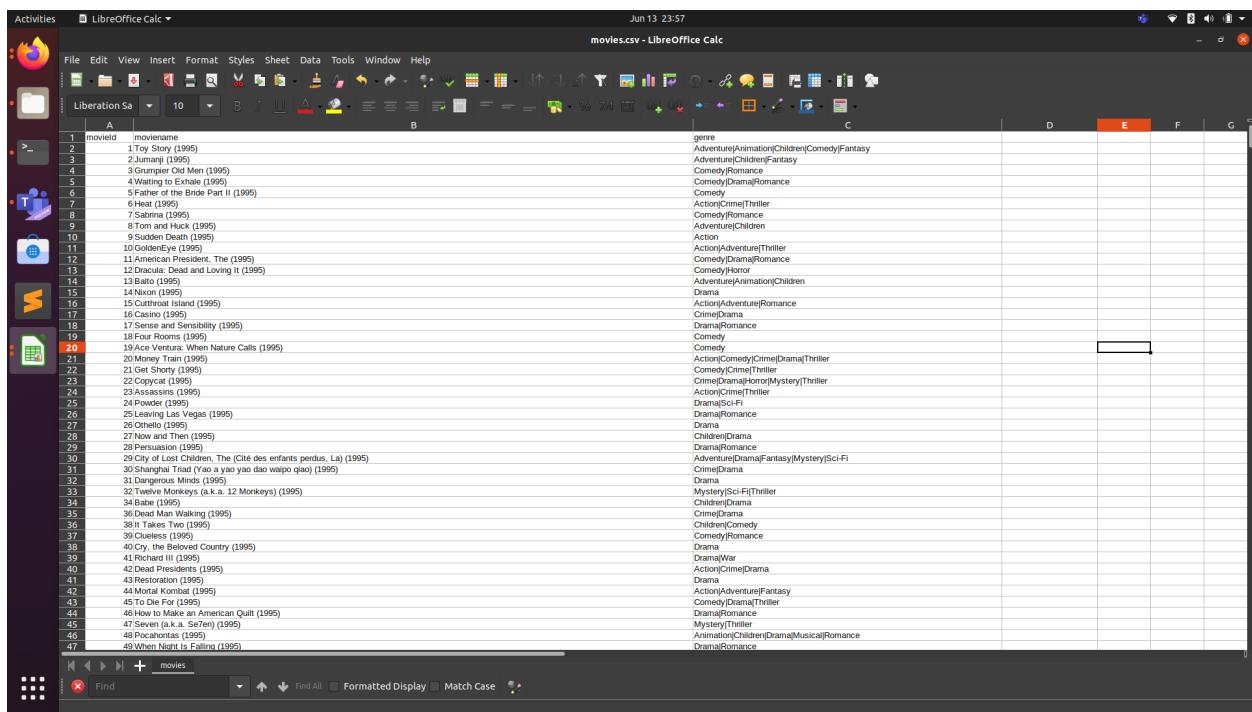
Close

This screenshot shows the HDFS browser interface in a Firefox window. The main pane displays a directory listing for '/movieshadoop/genreCount'. A modal window is open for the file 'part-00000', providing detailed information about its first block. The file's size is 253 bytes, and it was last modified on June 13, 2021. The file contents are listed below, showing the count of various movie genres.

# HIVE

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy. Apache Hive supports analysis of large datasets stored in Hadoop's HDFS and compatible file systems such as Amazon S3 filesystem and Alluxio. It provides a SQL-like query language called HiveQL with schema on read and transparently converts queries to MapReduce, Apache Tez and Spark jobs.

## DATASET\_FILE1\_movies.csv:-



The screenshot shows a LibreOffice Calc spreadsheet titled "movies.csv - LibreOffice Calc". The data consists of two columns: "imdbID" and "genre". The "imdbID" column contains movie IDs from 1 to 49, and the "genre" column lists the genres for each movie. The genres include Adventure, Animation, Children, Comedy, Fantasy, Drama, Thriller, Romance, Horror, Mystery, Sci-Fi, and War. The spreadsheet has a dark theme and is displayed on a desktop screen with other icons visible in the background.

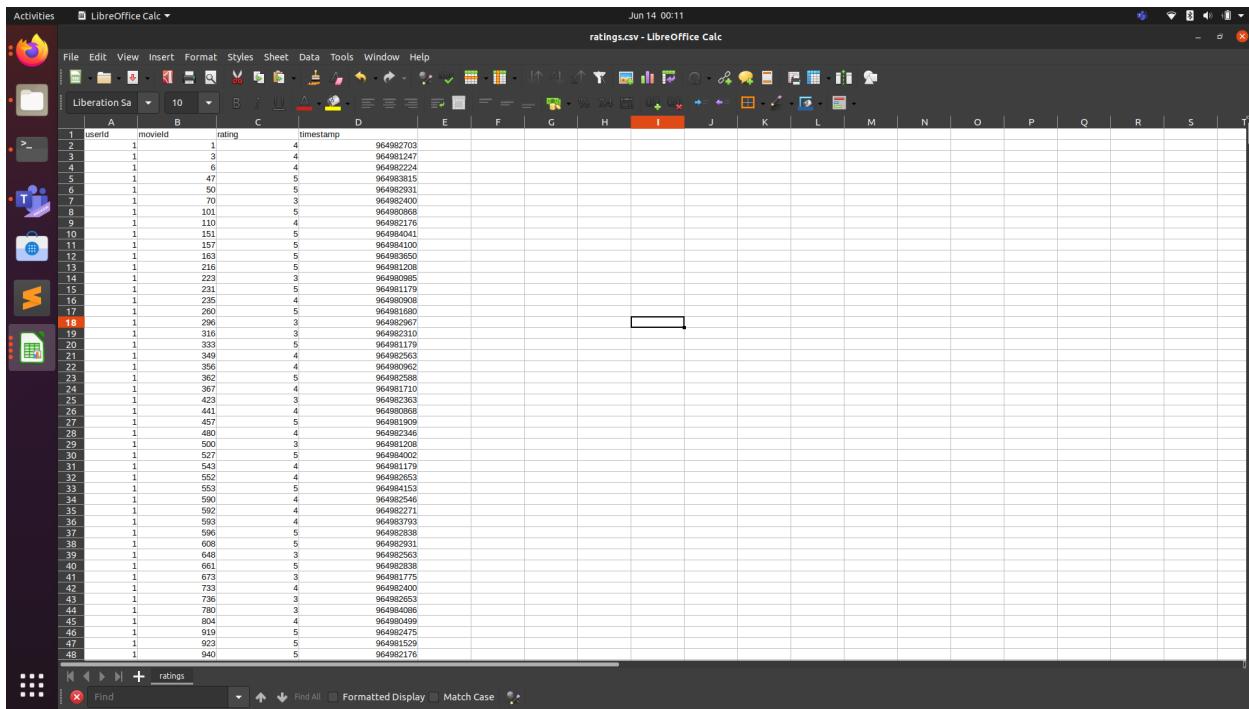
imdbID	genre
1	Adventure Animation Children Comedy Fantasy
2	Adventure Children Fantasy
3	Comedy Romance
4	Comedy Drama Romance
5	Comedy
6	Action Crime Thriller
7	Comedy Romance
8	Adventure Children
9	Action
10	Action Adventure Thriller
11	Comedy Drama Romance
12	Comedy Horror
13	Adventure Animation Children
14	Drama
15	Action Adventure Romance
16	Crime Drama
17	Drama Romance
18	Comedy
19	Action Comedy Crime Drama Thriller
20	Comedy Crime Thriller
21	Crime Drama Horror Mystery Thriller
22	Action Crime Mystery Thriller
23	Drama Sci-Fi
24	Drama Romance
25	Drama
26	Children Drama
27	Children Romance
28	Children Drama
29	Adventure Drama Fantasy Mystery Sci-Fi
30	Crime Drama
31	Drama
32	Mystery Sci-Fi Thriller
33	Children Drama
34	Crime Drama
35	Children Comedy
36	Comedy Romance
37	Drama
38	War
39	Action Crime Drama
40	Drama
41	Action Adventure Fantasy
42	Comedy Drama Thriller
43	Drama Horror
44	Mystery Thriller
45	Animation Children Drama Musical Romance
46	Drama Romance
47	
48	
49	

## **DATASET\_FILE2\_tags.csv:-**

## **DATASET\_FILE3\_links.csv:-**

	instvid	instvid
1	1	114709
2	2	113497
3	3	113228
4	4	114985
5	5	113041
6	6	113277
7	7	113119
8	8	112302
9	9	114576
10	10	113189
11	11	112346
12	12	112896
13	13	112453
14	14	113987
15	15	112700
16	16	112441
17	17	114388
18	18	113101
19	19	5
20	20	112281
21	21	113496
22	22	113611
23	23	112722
24	24	112401
25	25	114168
26	26	112665
27	27	114057
28	28	114011
29	29	114117
30	30	116982
31	31	115112
32	32	112792
33	33	114746
34	34	112431
35	35	112110
36	36	113442
37	37	112697
38	38	112749
39	39	34615
40	40	114279
41	41	31174
42	42	112119
43	43	114272
44	44	113855
45	45	114681
46	46	113447
47	47	114369
48	48	114448
49	49	114016
50	50	8391
48		629

## DATASET\_FILE4\_ratings.csv:-



The screenshot shows a LibreOffice Calc spreadsheet titled "ratings.csv - LibreOffice Calc". The window title bar indicates the file was last modified on Jun 14 00:11. The spreadsheet has four columns: A (userid), B (movieid), C (rating), and D (timestamp). The data starts at row 2 and continues down to row 48. The timestamp column contains values starting from 964982703 and ending at 964982176. Row 18 is highlighted with a red background.

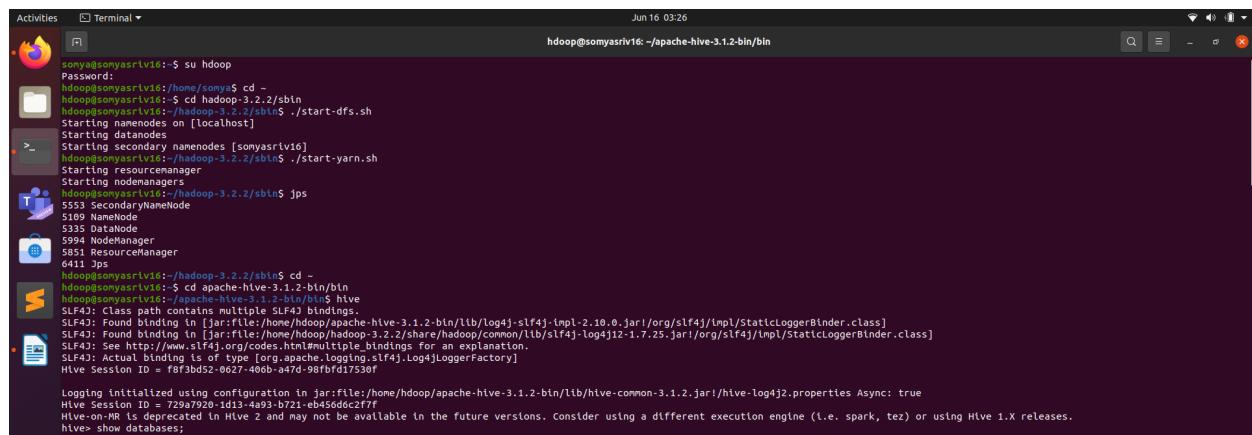
	A	B	C	D
1	userid	movieid	rating	timestamp
2	1	1	1	964982703
3	1	3	4	964981247
4	1	6	4	964982224
5	1	47	5	964982085
6	1	50	5	964982931
7	1	70	3	964982400
8	1	101	5	964980668
9	1	110	4	964982176
10	1	151	5	964981041
11	1	157	5	964981010
12	1	183	5	964983650
13	1	216	5	964981208
14	1	222	5	964982865
15	1	231	5	964981179
16	1	235	4	964980908
17	1	260	5	964981680
18	1	296	3	964982967
19	1	316	3	964982030
20	1	333	5	964981179
21	1	349	4	964982563
22	1	356	4	964980902
23	1	362	5	964982868
24	1	357	4	964982719
25	1	423	3	964982303
26	1	441	4	964980668
27	1	457	5	964981908
28	1	480	4	964982240
29	1	500	3	964981208
30	1	527	5	964984002
31	1	543	4	964981179
32	1	552	4	964982653
33	1	553	5	964981155
34	1	590	4	964982546
35	1	592	4	964982271
36	1	593	4	964983793
37	1	596	5	964982838
38	1	639	5	964982931
39	1	648	3	964982563
40	1	661	5	964982838
41	1	673	3	964981775
42	1	733	4	964982400
43	1	738	3	964982653
44	1	780	3	964984008
45	1	804	4	964980499
46	1	919	5	964982475
47	1	923	5	964981529
48	1	940	5	964982176

## GITHUB LINK:-

[https://github.com/1NT18IS155/1NT18IS155\\_somya\\_C\\_bdLab/tree/main/bdlab/Movies\\_hive](https://github.com/1NT18IS155/1NT18IS155_somya_C_bdLab/tree/main/bdlab/Movies_hive)

## DATA PREPARATION:-

### Accessing Hive:-

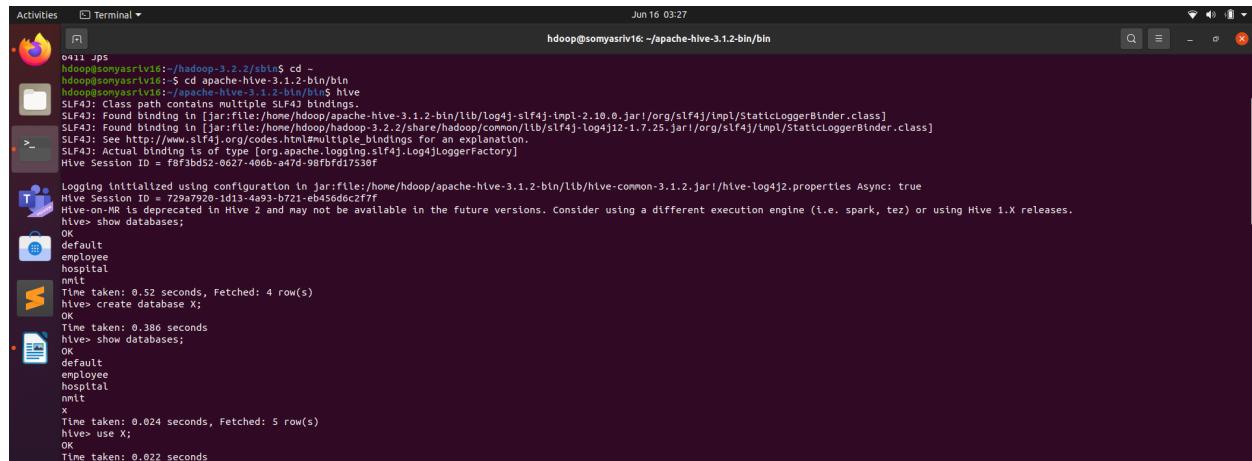


```
somyas@sonyasriv16:~$ su hdoop
Password:
hdoop@somyasriv16:~/home/somyas$ cd ~
hdoop@somyasriv16:~$ cd hadoop-3.2.2/sbin
hdoop@somyasriv16:~/hadoop-3.2.2/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting secondary namenodes [somyasriv16]
hdoop@somyasriv16:~/hadoop-3.2.2/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hdoop@somyasriv16:~/hadoop-3.2.2/sbin$ jps
5553 SecondaryNameNode
5109 NameNode
5335 DataNode
5994 NodeManager
5851 ResourceManager
6011 Jps
hdoop@somyasriv16:~/hadoop-3.2.2/sbin$ cd ~
hdoop@somyasriv16:~$ cd apache-hive-3.1.2-bin/bin
hdoop@somyasriv16:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/.m2/repository/org/apache/hive/hive-3.1.2-bin/lib/lbl/Log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/.m2/repository/hadoop/hadoop-3.2.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = f8f3bd52-0627-40cb-a47d-98fbfd1d7530f

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 729a7920-1d13-4ab3-b721-eb456ddc277f
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show databases;
hive>
```

### Creation of database:-

```
hive> show databases;
hive> create database X;
hive> show databases;
hive> use X;
```

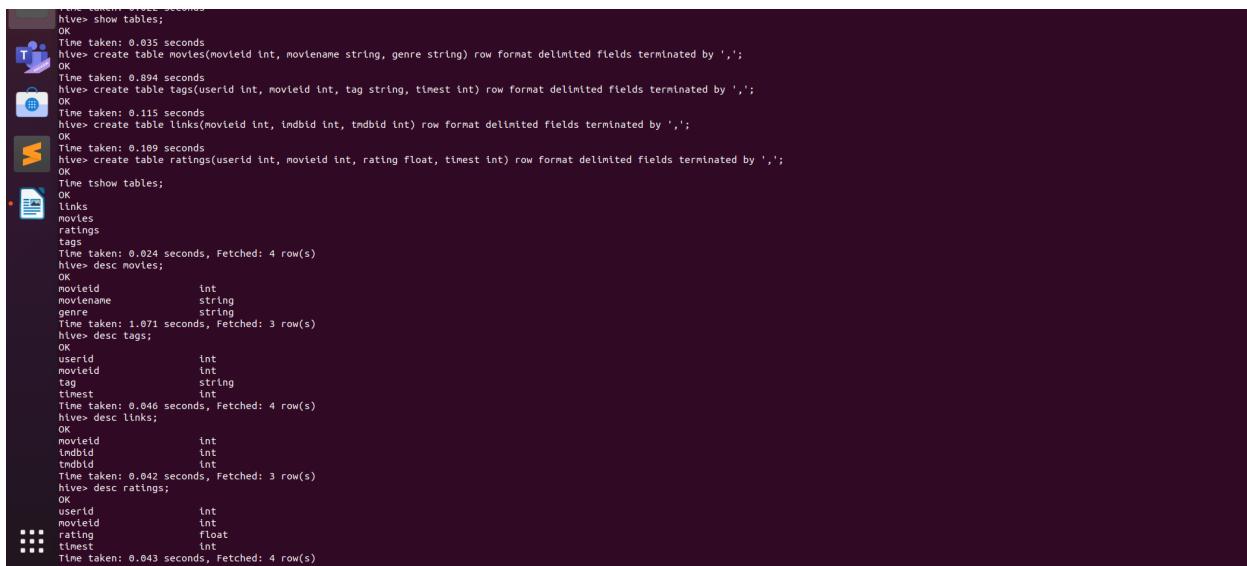


```
Activities Terminal Jun 16 03:27
hdoop@somyasriv16:~/hadoop-3.2.2/sbin$ cd ~
hdoop@somyasriv16:~$ cd apache-hive-3.1.2-bin/bin
hdoop@somyasriv16:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/.m2/repository/org/apache/hive/hive-3.1.2-bin/lib/lbl/Log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/.m2/repository/hadoop/hadoop-3.2.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = f8f3bd52-0627-40cb-a47d-98fbfd1d7530f

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 729a7920-1d13-4ab3-b721-eb456ddc277f
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show databases;
OK
default
employee
hospital
nmit
Time taken: 0.52 seconds, Fetched: 4 row(s)
hive> create database X;
OK
Time taken: 0.386 seconds
hive> show databases;
OK
default
employee
hospital
nmit
X
Time taken: 0.024 seconds, Fetched: 5 row(s)
hive> use X;
OK
Time taken: 0.022 seconds
hive>
```

## **Creation of tables:-**

```
hive> show tables;
hive> create table movies(movieid int, moviename string, genre string) row format delimited
fields terminated by ',';
hive> create table tags(userid int, movieid int, tag string, timest int) row format delimited fields
terminated by ',';
hive> create table links(movieid int, imdbid int, tmdbid int) row format delimited fields terminated
by ',';
hive> create table ratings(userid int, movieid int, rating float, timest int) row format delimited
fields terminated by ',';
hive> show tables;
hive> desc movies;
hive> desc tags;
hive> desc links;
hive> desc ratings;
```



```
hive> show tables;
OK
Time taken: 0.035 seconds
hive> create table movies(movieid int, moviename string, genre string) row format delimited fields terminated by ',';
OK
Time taken: 0.894 seconds
hive> create table tags(userid int, movieid int, tag string, timest int) row format delimited fields terminated by ',';
OK
Time taken: 0.115 seconds
hive> create table links(movieid int, imdbid int, tmdbid int) row format delimited fields terminated by ',';
OK
Time taken: 0.189 seconds
hive> create table ratings(userid int, movieid int, rating float, timest int) row format delimited fields terminated by ',';
OK
Time taken: 0.024 seconds, Fetched: 4 row(s)
hive> desc movies;
OK
movieid          int
moviename        string
genre            string
Time taken: 1.071 seconds, Fetched: 3 row(s)
hive> desc tags;
OK
userid           int
movieid          int
tag              string
timest           int
Time taken: 0.046 seconds, Fetched: 4 row(s)
hive> desc links;
OK
movieid          int
imdbid          int
tmdbid          int
Time taken: 0.042 seconds, Fetched: 3 row(s)
hive> desc ratings;
OK
userid           int
movieid          int
rating           float
timest           int
Time taken: 0.043 seconds, Fetched: 4 row(s)
```

## **Insertion of dataset manually:-**

```
hive> insert into movies values(193610,"Rang De Basanti", "Comedy|Romance|Historical
Fiction");
hive> insert into movies values(193611,"Neerja", "Drama");
hive> insert into movies values(193612,"Gully Boy", "Drama");
hive> insert into movies values(193613,"Piku", "Drama");
hive> insert into movies values(193614,"Zindagi Na Milegi Dobara",
"Adventure|Drama|Comedy");
hive> desc movies;
```

```
hive> select * from movies;
```

```
hive> desc movies;
OK
moviedb          int
moviename        string
genre            string
Time taken: 0.032 seconds, Fetched: 3 row(s)
hive> select * from movies;
OK
193610 Rang De Basanti Comedy|Romance|Historical Fiction
193611 Neerja Drama
193612 Gully Boy     Drama
193613 Piku      Drama
193614 Zindagi Na Milegi Dobara Adventure|Drama|Comedy
Time taken: 0.207 seconds, Fetched: 5 row(s)
```

## Loading of dataset:-

```
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/movies.csv' into table movies;
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/tags.csv' into table tags;
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/links.csv' into table links;
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/ratings.csv' into table ratings;
```

```
hive> desc movies;
OK
moviedb          int
moviename        string
genre            string
Time taken: 0.032 seconds, Fetched: 3 row(s)
hive> select * from movies;
OK
193610 Rang De Basanti Comedy|Romance|Historical Fiction
193611 Neerja Drama
193612 Gully Boy     Drama
193613 Piku      Drama
193614 Zindagi Na Milegi Dobara Adventure|Drama|Comedy
Time taken: 0.212 seconds, Fetched: 5 row(s)
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/movies.csv' into table movies;
Loading data to table x.movies
OK
Time taken: 0.07 seconds
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/tags.csv' into table tags;
Loading data to table x.tags
OK
Time taken: 0.07 seconds
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/links.csv' into table links;
Loading data to table x.links
OK
Time taken: 0.035 seconds
hive> LOAD DATA local INPATH '/home/somya/bdlab/Movies_hive/ratings.csv' into table ratings;
Loading data to table x.ratings
OK
Time taken: 1.485 seconds
hive> []
```

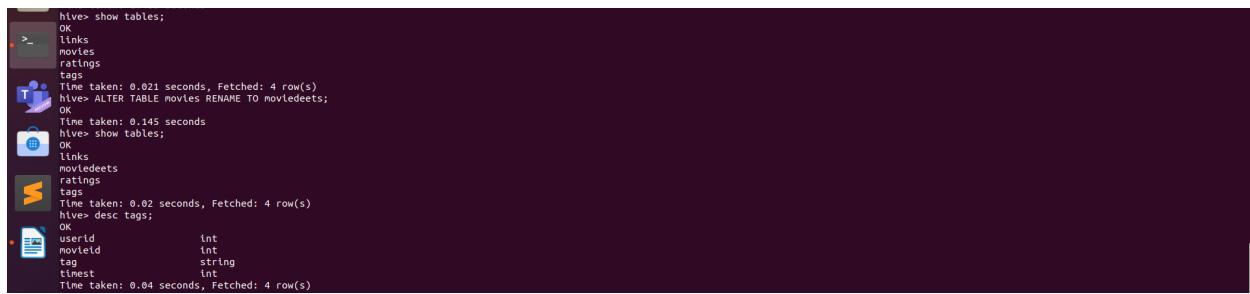
## QUERIES:-

### 1) Renaming table “movies” to “moviedeets”

```
ALTER TABLE movies RENAME TO moviedeets;
```

```
show tables;
```

```
desc tags;
```



```
hive> show tables;
OK
links
movies
ratings
tags
Time taken: 0.021 seconds, Fetched: 4 row(s)
hive> ALTER TABLE movies RENAME TO moviedeets;
OK
Time taken: 0.145 seconds
hive> show tables;
OK
links
moviedeets
ratings
tags
Time taken: 0.02 seconds, Fetched: 4 row(s)
hive> desc tags;
OK
userid          int
movieid         int
tag             string
timest          int
Time taken: 0.04 seconds, Fetched: 4 row(s)
```

### 2) Altering column name “tag” to “comment” in table “tags”

```
ALTER TABLE tags CHANGE tag comment string;
```

```
desc tags;
```



```
hive> desc tags;
OK
userid          int
movieid         int
tag             string
timest          int
Time taken: 0.04 seconds, Fetched: 4 row(s)
hive> ALTER TABLE tags CHANGE tag comment string;
OK
Time taken: 0.159 seconds
hive> desc tags;
OK
userid          int
movieid         int
comment         string
timest          int
Time taken: 0.033 seconds, Fetched: 4 row(s)
```

### 3) AND, OR, IN, NOT IN, SUBSTR, CONCAT, Case Operators

```
select * from tags where userid='62' AND movieid='110';
```

```

hive> select * from tags where userid='62' AND movieid='110';
OK
62    110    beautiful scenery      1528152541
62    110    epic                1528152532
62    110    historical          1528152523
62    110    inspirational        1528152527
62    110    Medieval            1528152528
62    110    mel gibson          1528152521
62    110    revenge (best Cinematography) 1528152539
62    110    sword fight          1528152535
Time taken: 0.19 seconds, Fetched: 9 row(s)

```

select \* from tags where userid='103' OR movieid='296';

```

hive> select * from tags where userid='103' OR movieid='296';
OK
103   260    EPIC      1431954312
103   260    great soundtrack 1431954337
103   296    good dialogue 1431954555
103   296    movie soundtrack 1431954555
103   296    non-linear 1421954559
424   296    cult film 1457844546
424   296    drugs 1457844550
424   296    Quentin Tarantino 1457844541
424   296    Tarantino 1457844557
476   296    men 1137285000 1498456537
599   296    1990s 1498456537
599   296    achronological 1498456475
599   296    action 1498456356
599   296    action packed 1498456535
599   296    aggressive 1498456533
599   296    amoral 1498456678
599   296    hearing dialogue 1498456473
599   296    anthology 1498456531
599   296    assassin 1498456382
599   296    atmospheric 1498456376
599   296    AWESOME 1498456559
599   296    bad 1498456461
599   296    bad ass 1498456461
599   296    bad boy 1498456469
599   296    bad-ass 1498456469
599   296    bibble 1498456675
599   296    biblical references 1498456529
599   296    big boys with guns 1498456673
599   296    big budget movies 1498456673
599   296    Black Comedy 1498456350
599   296    black humor 1498456481
599   296    black humour 1498456670
599   296    blood 1498456448
599   296    blood splatters 1498456467
599   296    bloodbath 1498456461
599   296    bone wills 1498456352
599   296    brutality 1498456543
599   296    casual violence 1498456669
599   296    character development 1498456527
599   296    character 1498456466
599   296    classic 1498456324
599   296    classic movie 1498456667
599   296    coke 1498456666
599   296    comedy 1498456378
599   296    conversation 1498456524
599   296    cool 1498456495
599   296    cool girl 1498456664
599   296    crime 1498456353
599   296    crime scene scrubbing 1498456663
599   296    cult 1498456391
599   296    cult classic 1498456522
599   296    cult film 1498456345
599   296    cult movie 1498456666
599   296    dancing 1498456664
599   296    dark 1498456404
Time taken: 0.19 seconds, Fetched: 9 row(s)

```

```

hive> select * from tags where userid='103' OR movieid='296';
OK
599   296    notable soundtrack 1498456585
599   296    offensive 1498456583
599   296    organised crime 1498456585
599   296    organised crime 1498456583
599   296    original 1498456479
599   296    original plot 1498456455
599   296    out of order 1498456586
599   296    Palme d'Or 1498456690
599   296    parody 1498456591
599   296    rape 14984565590
599   296    pop culture references 1498456478
599   296    psychological 1498456593
599   296    pulp 1498456484
599   296    Quentin Tarantino 1498456338
599   296    quirky 1498456363
599   296    Quirkie 1498456446
599   296    rids disturbing violent content including rape 1498456582
599   296    rids disturbing violent images 1498456595
599   296    r:graphic sexuality 1498456597
599   296    r:some violence 1498456599
599   296    r:strong bloody violence 1498456601
599   296    r:strong language 1498456452
599   296    r:stultained strong stylized violence 1498456603
599   296    r:violence 1498456424
599   296    random 1498456593
599   296    rape 1498456417
599   296    retro 1498456442
599   296    Royal Avenue 1498456682
599   296    Samuel L. Jackson 1498456342
599   296    sarcasm 1498456612
599   296    satire 1498456614
599   296    sexy 1498456615
599   296    sexual titling 1498456617
599   296    sophisticated 1498456618
599   296    soundtrack 1498456439
599   296    splatter 1498456620
599   296    Steve Buscemi 1498456694
599   296    stylized writing 1498456375
599   296    stylized 1498456354
599   296    suspense 1498456680
599   296    Tarantino 1498456358
599   296    thought-provoking 1498456638
599   296    thrill 1498456415
599   296    volta 1498456639
599   296    twist ending 1498456639
599   296    Uma Thurman 1498456370
599   296    unique 1498456641
599   296    unpredictable 1498456476
599   296    unusual 1498456661
599   296    very 1498456434
599   296    violence 1498456347
599   296    violent 1498456366
599   296    witty 1498456437
Time taken: 0.118 seconds, Fetched: 183 row(s)

```

```
select * from moviedeets where genre IN ("Crime","Thriller");
```

```
Activities Terminal Jun 16 03:58
hive> select * from moviedeets where genre IN ("Crime","Thriller");
OK
132 Jade (1995) Thriller
198 Safe (1995) Thriller
240 Hideaway (1995) Thriller
373 Red Rock West (1992) Thriller
452 The Hunted (1995) Thriller
457 Fugitive (The) (1993) Thriller
490 Malice (1993) Thriller
548 Silver (1993) Thriller
662 Fear (1996) Thriller
1055 Shadow Conspiracy (1997) Thriller
1101 Silence of the Lambs (1991) Thriller
1343 Cape Fear (1991) Thriller
1447 Gridlock'd (1997) Crime
1662 Gang Related (1997) Crime
1718 Stranger in the House (1997) Thriller
1794 Hush (1998) Thriller
1939 Paper Chase (1998) Thriller
2118 Dead Zone The (1983) Thriller
2178 Frenzy (1972) Thriller
2179 Topaz (1969) Thriller
2180 Torn Curtain (1966) Thriller
2181 I Confess (1953) Thriller
2182 Sabotage (1942) Thriller
2211 Secret Agent (1936) Thriller
2280 Clay Pigeons (1998) Crime
2298 Strangelands (1998) Thriller
2425 General T. (1998) Crime
2767 Arlington Road (1999) Thriller
2826 I Want to Hold You (1965) Thriller
2983 Inpress File The (1965) Thriller
3005 Bone Collector The (1999) Thriller
3015 Come (1978) Thriller
3055 Felicia's Journey (1999) Thriller
3070 Dead Calm (1988) Thriller
3220 Odessa File The (1974) Thriller
3230 Eye of the Beholder (1999) Thriller
3484 Skulls Th (2000) Thriller
3634 Seven Days in May (1964) Thriller
3797 In Crowd The (2000) Thriller
3835 Crush The (2000) Thriller
3837 Blood Diamond (2006) Thriller
4020 Gift The (2000) Thriller
4083 Best Seller (1987) Thriller
4093 Cop (1988) Thriller
4110 Fourth Protocol The (1987) Thriller
4167 15 Minutes (2001) Thriller
4170 Glass House Th (2001) Thriller
4814 Don't Say a Word (2001) Thriller
4885 Domestic Disturbance (2001) Thriller
4945 Enforcer Th (1976) Crime
5266 Panic Room (2002) Thriller
5282 High Crimes (2002) Thriller
5364 Swinman (2002) Thriller
5590 Mack The (1973) Crime
Activities Terminal Jun 16 03:58
hive> select * from moviedeets where genre IN ("Crime","Thriller");
OK
4093 Cop (1988) Thriller
4110 Fourth Protocol The (1987) Thriller
4167 15 Minutes (2001) Thriller
4770 Glass House Th (2001) Thriller
4814 Don't Say a Word (2001) Thriller
4885 Domestic Disturbance (2001) Thriller
4945 Enforcer Th (1976) Crime
5266 Panic Room (2002) Thriller
5282 High Crimes (2002) Thriller
5564 Swinman (2002) Thriller
5590 Mack The (1973) Crime
5707 Eyes of a Stranger Th (1988) Thriller
5949 Intact (Intacto) (2001) Thriller
6809 Tightrope (1984) Thriller
6909 Eye The (Gin gwat) (Jian gu) (2002) Thriller
7321 Twisted (2004) Thriller
7349 Spartan (2004) Thriller
7349 Paper Boy New (1992) Thriller
7871 Manchurian Candidate The (2004) Thriller
39427 Stay (2005) Thriller
79536 Helsinki (Röperi) (2009) Crime
86960 Across the Hall (2009) Thriller
103235 Best Offer The (Migliore offerta La) (2013) Thriller
108125 Bad Boys II (2003) Thriller
108949 Cat of the Steel The (2013) Crime
111384 Blue Ruin (2013) Thriller
112940 A Most Wanted Man (2014) Thriller
113186 Felony (2013) Thriller
113862 Guest The (2014) Thriller
114239 Mr. & Mrs. Smith (Agenten für Stichen) (2014) Thriller
127319 The Loft (2014) Thriller
131439 Kill Me Three Times (2014) Thriller
134158 Return to Sender (2015) Thriller
136786 Careful What You Wish For (2015) Thriller
139857 Colonia (2016) Thriller
140038 Special Agent (2015) Thriller
143081 Anti-Social (2015) Crime
147328 The Adventures of Sherlock Holmes and Dr. Watson: Bloody Signature (1979) Crime
147330 Sherlock Holmes and Dr. Watson: Acquaintance (1979) Crime
148709 Mojave (2015) Thriller
149598 Standoff (2016) Thriller
150177 10 Cloverfield Lane (2016) Thriller
156726 Hatch (2016) Thriller
160840 The Maid's Room (2014) Thriller
161634 Don't Breathe (2016) Thriller
161966 Elle (2016) Thriller
162605 The Girl on the Train (2016) Thriller
165835 The Girl on the Train (2016) Thriller
165840 Miss Sloane (2016) Thriller
169180 American Fable (2017) Thriller
173751 Tiger Raid (2016) Thriller
173941 Atomic Blonde (2017) Thriller
185135 Sherlock - A Study in Pink (2010) Crime
Time taken: 0.114 seconds, Fetched: 96 row(s)
hive> 
```

```
select distinct(movieid, rating) from ratings where rating NOT IN  
(1.0,1.5,2.0,2.5,3.0,3.5,4.0,4.5,5.0);
```

Activities Terminal Jun 16 04:29 hdoop@somyasrlv16: ~/apache-hive-3.1.2-bin/bin

```
({"col1":134246,"col2":0.5}, {"col1":134393,"col2":0.5}, {"col1":134401,"col2":0.5}, {"col1":135216,"col2":0.5}, {"col1":135536,"col2":0.5}, {"col1":135861,"col2":0.5}, {"col1":135887,"col2":0.5}, {"col1":136026,"col2":0.5}, {"col1":136034,"col2":0.5}, {"col1":136684,"col2":0.5}, {"col1":137517,"col2":0.5}, {"col1":138186,"col2":0.5}, {"col1":138798,"col2":0.5}, {"col1":139385,"col2":0.5}, {"col1":139402,"col2":0.5}, {"col1":139717,"col2":0.5}, {"col1":140711,"col2":0.5}, {"col1":141668,"col2":0.5}, {"col1":141994,"col2":0.5}, {"col1":144594,"col2":0.5}, {"col1":145274,"col2":0.5}, {"col1":145951,"col2":0.5}, {"col1":149352,"col2":0.5}, {"col1":149354,"col2":0.5}, {"col1":151315,"col2":0.5}, {"col1":151323,"col2":0.5}, {"col1":151745,"col2":0.5}, {"col1":152063,"col2":0.5}, {"col1":152079,"col2":0.5}, {"col1":156607,"col2":0.5}, {"col1":156609,"col2":0.5}, {"col1":157172,"col2":0.5}, {"col1":157173,"col2":0.5}, {"col1":158783,"col2":0.5}, {"col1":158872,"col2":0.5}, {"col1":159858,"col2":0.5}, {"col1":160872,"col2":0.5}, {"col1":161001,"col2":0.5}, {"col1":165181,"col2":0.5}, {"col1":165645,"col2":0.5}, {"col1":166461,"col2":0.5}, {"col1":166534,"col2":0.5}, {"col1":167296,"col2":0.5}, {"col1":172345,"col2":0.5}, {"col1":173307,"col2":0.5}, {"col1":175303,"col2":0.5}, {"col1":175475,"col2":0.5}, {"col1":175485,"col2":0.5}, {"col1":184235,"col2":0.5}, {"col1":184471,"col2":0.5}, {"col1":184641,"col2":0.5}) Time taken: 25.13 seconds, Fetched: 1066 row(s) hive: []
```

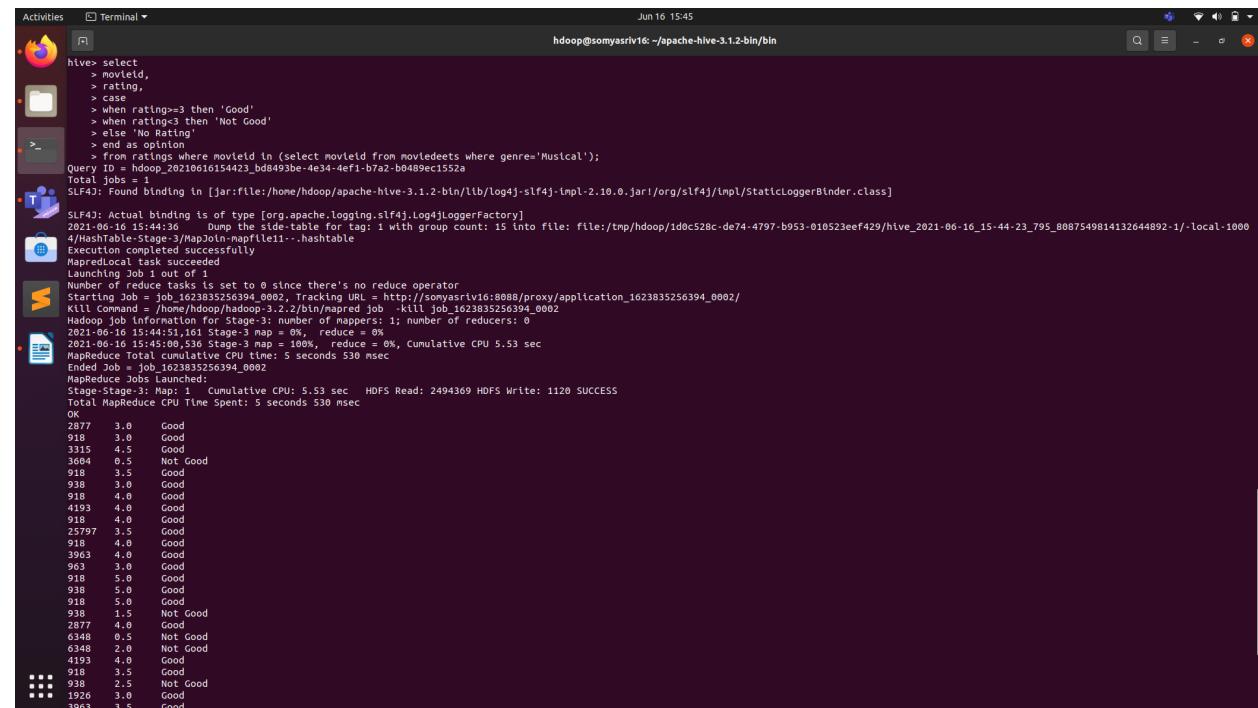
```
select substr(moviename,-6) from moviecharts where genre='Musical';
```

```
hive> select substr(moviename,-6) from moviecharts where genre='Musical';
OK
(1944)
(1958)
(1949)
(1952)
(1975)
(1951)
(1962)
(1964)
(1995)
(1959)
(1957)
(1946)
(1933)
(1988)
(1998)
Time taken: 0.144 seconds, Fetched: 15 row(s)
hive> □
```

```
select movieid, moviename, genre, concat(movieid,'.',moviename,'.',genre) from moviecharts
where genre='Film-Noir';
```

```
hive> select movieid, moviename, genre, concat(movieid,'.',moviename,'.',genre) from moviecharts where genre='Film-Noir';
OK
2066    Out of the Past (1947)  Film-Noir      2066,Out of the Past (1947),Film-Noir
4426    Kiss Me Deadly (1955)  Film-Noir      4426,Kiss Me Deadly (1955),Film-Noir
7735    Pickup on South Street (1953)  Film-Noir    7735,Pickup on South Street (1953),Film-Noir
Time taken: 1.179 seconds, Fetched: 3 row(s)
hive> □
```

```
hive> select
> movieid,
> rating,
> case
> when rating>=3 then 'Good'
> when rating<3 then 'Not Good'
> else 'No Rating'
> end as opinion
> from ratings where movieid in (select movieid from moviecharts where genre='Musical');
```



```
Activities Terminal ▾ Jun 16 15:45
hive> select
> movieid,
> rating,
> case
> when rating>=3 then 'Good'
> when rating<3 then 'Not Good'
> else 'No Rating'
> end as opinion
> from ratings where movieid in (select movieid from moviecharts where genre='Musical');
Query ID = hdoop_20210616154423_bdd8493be-4e34-4ef1-b7a2-bd489ec1552a
Total Jobs = 1
SLF4J: Found binding in [jar:file:/home/hdoop/apache-hive-3.1.2-bin/lib/Log4jLoggerFactory]
SLF4J: Actual binding is of type [org.apache.logging.log4j.Log4jLoggerFactory]
2021-06-16 15:44:36   Dump the slide-table for tag: 1 with group count: 15 into file: file:/tmp/hdoop/1d0c528c-de74-4797-b953-010523eeff429/hive_2021-06-16_15-44-23_795_8087549814132644892-1/-local-1000
4/HashTable-Stage-3/MapJoin->mapfile11--hashtable
Execution completed successfully
MapredLocal task succeeded
Launching job 1623835256394_0002
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1623835256394_0002, Tracking URL = http://somyasriv16:8088/proxy/application_1623835256394_0002/
Kill Command = /home/hdoop/hadoop-3.2.2/bin/nmapred job -kill job_1623835256394_0002
Hadoop job Information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-16 15:45:51.161 Stage-3 map = 0%, reduce = 0%
2021-06-16 15:45:51.161 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.53 sec
MapReduce Total cumulative CPU time: 5 seconds 530 msec
Ended Job = job_1623835256394_0002
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1  Cumulative CPU: 5.53 sec  HDFS Read: 2494369 HDFS Write: 1120 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 530 msec
0 rows
0.000000s
2877  3.0  Good
918  3.0  Good
3315  4.5  Good
3664  0.5  Not Good
918  3.5  Good
936  4.0  Good
918  4.0  Good
4193  4.0  Good
918  4.0  Good
25797  3.5  Good
918  4.0  Good
3063  4.0  Good
983  3.0  Good
918  5.0  Good
938  5.0  Good
918  5.0  Good
938  1.5  Not Good
2347  4.0  Good
6348  0.5  Not Good
6348  2.0  Not Good
4193  4.0  Good
918  3.5  Good
938  2.5  Not Good
1926  3.0  Good
3963  3.5  Good
```

```

Activities Terminal Jun 16 15:45
hadoop@somysriv16: ~/apache-hive-3.1.2-bin/bin

MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1623835256394_0002, Tracking URL = http://somysriv16:8088/proxy/application_1623835256394_0002/
Kill Command = /home/hadoop/hadoop-3.2.2/bin/apred -kill job_1623835256394_0002
Hadoop Job Information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-16 15:44:51,161 Stage-3 map = 0%, reduce = 0%
2021-06-16 15:45:00,536 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.53 sec
MapReduce Total cumulative CPU time: 5 seconds 530 msec
Ended Job = job_1623835256394_0002
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Reduce: 0 Cumulative CPU: 5.53 sec HDFS Read: 2494369 HDFS Write: 1120 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 530 msec
OK
2877 3.0 Good
918 3.0 Good
3315 4.5 Good
6348 4.5 Not Good
918 3.5 Good
938 3.0 Good
918 4.0 Good
4193 4.0 Good
918 4.0 Good
4193 4.0 Good
918 4.0 Good
3963 4.0 Good
963 3.0 Good
918 5.0 Good
938 5.0 Good
938 1.5 Not Good
2877 4.0 Good
6348 0.5 Not Good
6348 2.0 Not Good
4193 4.0 Good
918 3.5 Good
938 2.5 Not Good
1926 3.0 Good
3963 3.5 Good
6396 3.5 Good
6391 3.5 Good
918 5.0 Good
938 4.5 Good
2877 4.0 Good
918 4.0 Good
938 3.0 Good
1926 5.0 Good
2877 1.0 Not Good
2877 4.0 Good
2877 2.0 Not Good
6348 2.5 Not Good
26639 3.0 Good
2877 3.0 Good
Time Taken: 40.002 seconds, Fetched: 39 row(s)
hive> 
```

## 4) Aggregate Functions(COUNT,MIN,MAX,SUM,AVG)

`select count(*) from moviedeets;`

```

hive> select count(*) from moviedeets;
Query ID = hdoop_20210616154922_95805760-9ad8-4298-bfb2-e55b1da27212
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set mapreduce.job.reduces=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623835256394_0003, Tracking URL = http://somysriv16:8088/proxy/application_1623835256394_0003/
Kill Command = /home/hadoop/hadoop-3.2.2/bin/apred -kill job_1623835256394_0003
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-16 15:49:42,122 Stage-1 map = 0% reduce = 0%
2021-06-16 15:49:52,481 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.24 sec
MapReduce Total cumulative CPU time: 6 seconds 880 msec
Ended Job = job_1623835256394_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.88 sec HDFS Read: 503529 HDFS Write: 104 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 880 msec
OK
9748
Time Taken: 32.089 seconds, Fetched: 1 row(s)
hive> 
```

select min(rating) from ratings;

```
hive> select min(rating) from ratings;
Query ID = hdoop_20210616155957_74afa042-b14-4471-bd43-0adb694a1add
Total Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623835256394_0004, Tracking URL = http://somyasriv16:8088/proxy/application_1623835256394_0004/
Kill Command = /home/hadoop/hadoop-3.2.2/bin/apred job -kill job_1623835256394_0004
Hadoop Job Info for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-16 16:00:11,702 Stage-1 map = 0%, reduce = 0%
2021-06-16 16:00:17,931 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.26 sec
2021-06-16 16:00:26,261 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.44 sec
MapReduce Total cumulative CPU time: 8 seconds 440 msec
Ended Job = job_1623835256394_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.44 sec HDFS Read: 2496454 HDFS Write: 103 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 440 msec
OK
0.5
Time taken: 32.977 seconds, Fetched: 1 row(s)
```

select max(rating) from ratings;

```
hive> select max(rating) from ratings;
Query ID = hdoop_20210616160042_a4c9846e-95bc-ac09-8192-177bc8ccda3f
Total Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623835256394_0005, Tracking URL = http://somyasriv16:8088/proxy/application_1623835256394_0005/
Kill Command = /home/hadoop/hadoop-3.2.2/bin/apred job -kill job_1623835256394_0005
Hadoop Job Info for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-16 16:00:58,461 Stage-1 map = 0%, reduce = 0%
2021-06-16 16:01:05,161 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.74 sec
2021-06-16 16:01:12,461 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.47 sec
MapReduce Total cumulative CPU time: 8 seconds 470 msec
Ended Job = job_1623835256394_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.47 sec HDFS Read: 2496454 HDFS Write: 103 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 470 msec
OK
0.0
Time taken: 32.076 seconds, Fetched: 1 row(s)
```

select sum(rating) from ratings;

```
hive> select sum(rating) from ratings;
Query ID = hdoop_20210616160916_d76bb0e8-f0ee-404c-9bc5-c502cd57ad64
Total Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623835256394_0006, Tracking URL = http://somyasriv16:8088/proxy/application_1623835256394_0006/
Kill Command = /home/hadoop/hadoop-3.2.2/bin/apred job -kill job_1623835256394_0006
Hadoop Job Info for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-16 16:09:34,181 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 3.74 sec
2021-06-16 16:09:41,355 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.41 sec
MapReduce Total cumulative CPU time: 7 seconds 410 msec
Ended Job = job_1623835256394_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.41 sec HDFS Read: 2496457 HDFS Write: 108 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 410 msec
OK
353083.0
Time taken: 26.397 seconds, Fetched: 1 row(s)
```

select avg(rating) from ratings;

```
hive> select avg(rating) from ratings;
Query ID = hdoop_20210616161022_55f3a3ee-2842-43e2-b3c2-a1521b2bba27
Total Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623835256394_0007, Tracking URL = http://somyasriv16:8088/proxy/application_1623835256394_0007/
Kill Command = /home/hadoop/hadoop-3.2.2/bin/apred job -kill job_1623835256394_0007
Hadoop Job Info for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-16 16:10:33,191 Stage-1 map = 0%, reduce = 0%
2021-06-16 16:10:40,466 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.35 sec
2021-06-16 16:10:48,000 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.6 sec
MapReduce Total cumulative CPU time: 9 seconds 600 msec
Ended Job = job_1623835256394_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.6 sec HDFS Read: 2498604 HDFS Write: 117 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 600 msec
OK
3.501556983616962
Time taken: 28.228 seconds, Fetched: 1 row(s)
hive> 
```

## 5) Creation of View

```
hive> create view if not exists musical_movies as  
> select * from moviedeets where genre='Musical';
```

```
hive> create view if not exists musical_movies as  
  > select * from moviedeets where genre='Musical';  
OK  
Time taken: 1.971 seconds  
hive> show tables;  
OK  
moviedeets  
musical_movies  
ratings  
tags  
Time taken: 0.055 seconds, Fetched: 5 row(s)  
hive> select * from musical_movies;  
OK  
918 Meet Me in St. Louis (1944) Musical  
938 Gigi (1958) Musical  
963 Inspector General The (1949) Musical  
1926 Broadway Melody The (1929) Musical  
2871 Tommy (1950) Musical  
3115 Holiday (1933) Musical  
3664 Gypsy (1962) Musical  
3963 Unsinkable Molly Brown The (1964) Musical  
4193 Fantasticks The (1995) Musical  
6340 Breakin' 2: Electric Boogaloo (1984) Musical  
6805 Sweet Smell of Success (1957) Musical  
6911 Jolson Show The (1946) Musical  
25797 Gold Diggers of 1933 (1933) Musical  
26630 Moonwalker (1988) Musical  
107780 Cats (1998) Musical  
Time taken: 0.234 seconds, Fetched: 15 row(s)  
hive> desc musical_movies;  
OK  
# Name          Type  
movieid        int  
moviename      string  
genre          string  
Time taken: 0.057 seconds, Fetched: 3 row(s)
```

## 6) GROUP BY

```
select rating,count(*) from ratings group by rating;
```

```
hive> select rating,count(*) from ratings group by rating;  
Query ID = hdoop_20210616173047_ded33602-956f-4d0e-ab06-5121745dcfc8c  
Total MapReduce CPU Time Spent: 0 msec  
Launching Job 1 out of 1  
Number of reduce tasks not specified. Estimated from input data size: 1  
In order to change the average load for a reducer (in bytes):  
  set hive.exec.reducers.bytes.per.reducer=<number>  
In order to limit the maximum number of reducers:  
  set hive.exec.reducers.max=<number>  
In order to set a constant number of reducers:  
  set mapreduce.job.reduces=<number>  
Starting Job = job_1623835256394_0012, Tracking URL = http://somyasrividig:8088/proxy/application_1623835256394_0012/  
Kill Command = /home/hadoop/hadoop-3.2.2/bin/mapred job -kill job_1623835256394_0012  
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1  
2021-06-16 17:30:58,155 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.2 sec  
2021-06-16 17:31:03,303 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.04 sec  
MapReduce Total cumulative CPU time: 4 seconds 40 msec  
Ended Job = job_1623835256394_0012  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.04 sec HDFS Read: 2497031 HDFS Write: 318 SUCCESS  
Total MapReduce CPU Time Spent: 4 seconds 40 msec  
OK  
NULL      1  
0.5      1370  
1.0      2811  
1.5      7074  
2.0      7551  
2.5      5558  
3.0      20047  
3.5      13136  
4.0      26818  
4.5      858  
5.0      1321  
Time taken: 18.339 seconds, Fetched: 11 row(s)
```

## 7) Joins (Outer, left outer, right outer)

```
select m.movieid, m.moviename, m.genre, r.rating from moviedeets m full outer join ratings r on  
(m.movieid=r.movieid);
```

```
Activities Terminal Jun 16 10:22 hdoop@somyasriv16: ~apache-hive-3.1.2-bin/bin
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 5.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 5.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 1.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 1.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 4.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 4.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 4.0
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 4.5
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi 3.0
187593 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi 4.0
187593 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi 3.5
187593 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi 3.0
187593 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi 4.0
187593 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi 5.0
187717 Won't You Be My Neighbor? (2018) Documentary 5.0
188189 Sorry To Bother You (2018) Comedy|Fantasy|Sci-Fi 4.5
188301 Ant-Man and the Wasp (2018) Action|Adventure|Comedy|Fantasy|Sci-Fi 4.0
188301 Ant-Man and the Wasp (2018) Action|Adventure|Comedy|Fantasy|Sci-Fi 3.0
188301 Ant-Man and the Wasp (2018) Action|Adventure|Comedy|Fantasy|Sci-Fi 4.0
188305 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi 4.0
188675 Dogman (2018) Crime|Drama 3.5
188751 Mama Mia: Here We Go Again! (2018) Comedy|Romance 4.5
188797 Tag (2018) Comedy 4.0
188833 The Man Who Killed Don Quixote (2018) Adventure|Comedy|Fantasy 4.5
188833 The Man Who Killed Don Quixote (2018) Comedy|Drama 2.5
189311 Big Little Lies (2018) Drama|Thriller 3.0
189333 Mission: Impossible - Fallout (2018) Action|Adventure|Thriller 3.5
189333 Mission: Impossible - Fallout (2018) Action|Adventure|Thriller 4.0
189381 Superfly (2018) Action|Crime|Thriller 2.5
189547 Iron Soldier (2018) Action|Sci-Fi 1.0
189717 BlackKklansman (2018) Comedy|Crime|Drama 2.5
190183 The Darkest Minds (2018) Sci-Fi|Thriller 3.5
190207 Title (2011) Drama|Romance 1.5
190209 Jeff Ross Roasts the Border (2017) Comedy 4.0
190213 John From (2015) Drama 1.0
190215 Liquid Truth (2017) Drama 1.0
190216 The Last Face (2016) Drama|Romantic 1.0
190221 Hommage à Zougou (et salut à Sabine Mamou) (2002) Documentary 1.0
191005 Gintama (2017) Action|Adventure|Comedy|Sci-Fi 4.5
193565 Gintama: The Movie (2010) Action|Animation|Comedy|Sci-Fi 3.5
193567 anohana: The Flower We Saw That Day - The Movie (2013) Animation|Drama 3.0
193571 Silver Spoon (2014) Comedy|Drama 4.0
193573 The School for Good and Evil (2015) Animation 4.0
193579 Jon Stewart Has Left the Building (2015) Documentary 3.5
193581 Black Butler: Book of the Atlantic (2017) Action|Animation|Comedy|Fantasy 4.0
193583 No Game No Life: Zero (2017) Animation|Comedy|Fantasy 3.5
193585 Flint (2019) Drama 3.5
193588 The King's Gambit (2020) Adventure|Sci-Fi 4.0
193609 Andrew Dice Clay: Dice Serves (1991) Comedy 4.0
193610 Rang De Basanti Comedy|Romance|Historical Fiction NULL
193611 Neerja Drama NULL
193612 Gully Boy Drama NULL
193613 Piku Drama NULL
193614 Madagad Na Mlekti Dobarra Adventure|Drama|Comedy NULL
Time taken: 23.419 seconds, Fetched: 100000+ rows(s)
hive>
```

```
select m.movieid,m.movieName,m.genre,t.comment from movieDeets m left outer join tags t on (m.movieid=t.movieid);
```

```
hive> select m.movieid,m.movieName,m.genre,t.comment from movieDeets m left outer join tags t on (m.movieid=t.movieid);
Query ID = hdoop_20210610183111_0b775b2e-6f81-4d30-96a9-11b9b23c2ff6
Total Jobs = 1
SLF4J: Found binding in [jar:file:/home/hdoop/apache-hive-3.1.2-bin/lbl/log4j-slf4j-impl-2.10.0.jar!/org/slf4j.impl.StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hdoop/hadoop-3.2.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j.impl.StaticLoggerBinder.class]
SLF4J: see http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2021-06-10 18:31:11,111 Starting to local task to process map join: maximum memory = 230075328
2021-06-10 18:31:12,000 Dump the side-table for tag: with group count: 1573 into file: file:/tmp/hdoop/1d0c528c-de74-4797-b953-010523ee429/hive_2021-06-10_18-31-11_409_5853731087791542178-1-local-10
004/HashTable-Stage-3/MapJoin-mapfile21--hashtable
2021-06-10 18:31:20 End of local task; Time Taken: 1.151 sec.
Execution completed successfully
MapReduce tasks succeeded
  Running Job 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = Job_1623835256394_0014, Tracking URL = http://sonyaslrv16:8088/proxy/application_1623835256394_0014/
Kill Command = /home/hdoop/hadoop-3.2.2/bin/nmapred job -kill job_1623835256394_0014
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-10 18:31:23,598 Stage-3 map = 0%, reduce = 0%
2021-06-10 18:31:23,598 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 3.23 sec
MapReduce Total cumulative CPU time: 3 seconds 230 msec
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 3.23 sec HDFS Read: 499958 HDFS Write: 803202 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 230 msec
Ended Job = job_1623835256394_0014
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 3.23 sec HDFS Read: 499958 HDFS Write: 803202 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 230 msec
193610 Rang De Basanti Comedy|Romance|Historical Fiction NULL
193611 Neerja Drama NULL
193612 Gully Boy Drama NULL
193613 Piku Drama NULL
193614 Koi Miliagi Na Milegi Dobara Adventure|Drama|Comedy NULL
NULL movieName genre NULL
1 Toy Story (1995) Adventure|Animation|Children|Comedy|Fantasy pixar
1 Toy Story (1995) Adventure|Animation|Children|Comedy|Fantasy pixar
1 Toy Story (1995) Adventure|Animation|Children|Comedy|Fantasy fun
2 Jumanji (1995) Adventure|Children|Fantasy fantasy
2 Jumanji (1995) Adventure|Children|Fantasy movie board game
2 Jumanji (1995) Adventure|Children|Fantasy Robin Williams
2 Jumanji (1995) Adventure|Children|Fantasy game
3 Grumpier Old Men (1995) Comedy|Romance moldy
3 Grumpier Old Men (1995) Comedy|Romance old
4 Waiting to Exhale (1995) Comedy|Drama|Romance NULL
5 Father of the Bride Part II (1995) Comedy pregnancy
5 Father of the Bride Part II (1995) Comedy remake
6 Heat (1995) Action|Crime|Thriller NULL
7 Sabrina (1995) Comedy|Romance remake
8 Tom and Huck (1995) Adventure|Children NULL
9 Sudden Death (1995) Action NULL
10 Ed Wood (1994) Mystery|Thriller NULL
11 American President The (1995) Comedy|Drama|Romance politics
11 American President The (1995) Comedy|Drama|Romance president
12 Dracula: Dead and Loving It (1995) Comedy|Horror NULL
13 Balto (1995) Adventure|Animation|Children NULL
14 Nixon (1995) Drama politics
14 Nixon (1995) Drama president
15 Cutthroat Island (1995) Action|Adventure|Romance NULL
16 Casino (1995) Crime|Drama Mafia
16
```

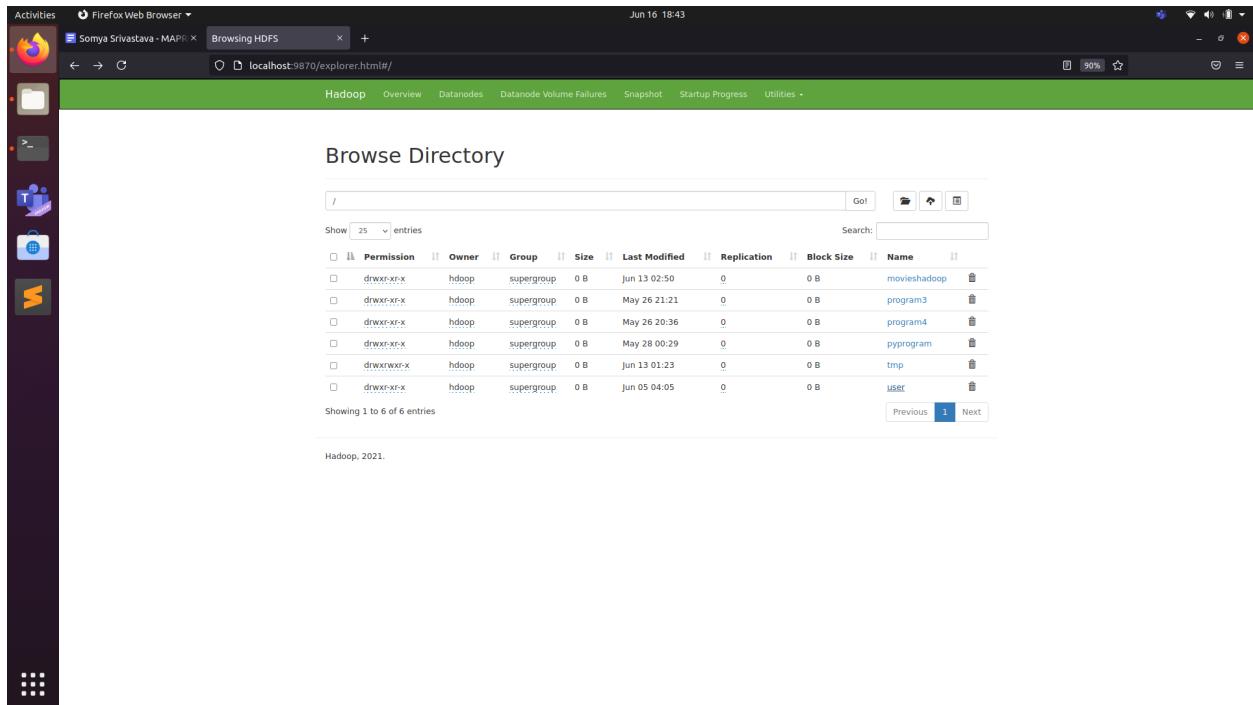
```
hive> select m.movieid,m.movieName,m.genre,t.comment from movieDeets m left outer join tags t on (m.movieid=t.movieid);
Query ID = hdoop_20210610183111_0b775b2e-6f81-4d30-96a9-11b9b23c2ff6
Total Jobs = 1
SLF4J: Found binding in [jar:file:/home/hdoop/apache-hive-3.1.2-bin/lbl/log4j-slf4j-impl-2.10.0.jar!/org/slf4j.impl.StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hdoop/hadoop-3.2.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j.impl.StaticLoggerBinder.class]
SLF4J: see http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2021-06-10 18:31:11,111 Starting to local task to process map join: maximum memory = 230075328
2021-06-10 18:31:12,000 Dump the side-table for tag: with group count: 18859 into file: file:/tmp/hdoop/1d0c528c-de74-4797-b953-010523ee429/hive_2021-06-10_18-31-11_409_5853731087791542178-1-local-10
004/HashTable-Stage-3/MapJoin-mapfile21--hashtable
2021-06-10 18:31:20 End of local task; Time Taken: 1.151 sec.
Execution completed successfully
MapReduce tasks succeeded
  Running Job 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = Job_1623835256394_0014, Tracking URL = http://sonyaslrv16:8088/proxy/application_1623835256394_0014/
Kill Command = /home/hdoop/hadoop-3.2.2/bin/nmapred job -kill job_1623835256394_0014
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-10 18:31:23,598 Stage-3 map = 0%, reduce = 0%
2021-06-10 18:31:23,598 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 24.262 sec
MapReduce Total cumulative CPU time: 24.262 msec
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 24.262 msec HDFS Read: 118859 HDFS Write: 118859 SUCCESS
Total MapReduce CPU Time Spent: 24.262 msec
Ended Job = job_1623835256394_0014
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 24.262 msec HDFS Read: 118859 HDFS Write: 118859 SUCCESS
Total MapReduce CPU Time Spent: 24.262 msec
184791 Fred Armisen: Standup for Drummers (2018) Comedy NULL
184931 Death Wish (2018) Action|Crime|Drama|Thriller NULL
184937 A Million Ways to Die in the West (2014) Action|Comedy|Drama|Sci-Fi NULL
184997 Love Simon (2018) Comedy|Drama NULL
185029 A Quiet Place (2018) Drama|Horror|Thriller NULL
185031 Alpha (2018) Adventure|Thriller NULL
185033 I Kill Giants (2018) Drama|Fantasy|Thriller NULL
185135 Sherlock - A Study In Pink (2010) Crime NULL
185136 The Devil Wears Prada (2006) Drama|Comedy NULL
185472 Blockers (2018) Comedy NULL
185585 Pacific Rim: Uprising (2018) Action|Fantasy|Sci-Fi NULL
186587 Rampage (2018) Action|Adventure|Sci-Fi NULL
187031 Jurassic World: Fallen Kingdom (2018) Action|Adventure|Drama|Sci-Fi|Thriller NULL
187541 Incredibles 2 (2018) Action|Adventure|Animation|Children NULL
187549 Deadpool (2016) Action|Comedy|Sci-Fi Jos Brolin
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi Ryan Reynolds
187593 Deadpool 2 (2018) Action|Comedy|Sci-Fi sarcasm
187595 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi Emilia clarke
187595 Solo: A Star Wars Story (2018) Action|Adventure|Children|Sci-Fi star wars
187717 Won't You Be My Neighbor? (2018) Documentary NULL
188307 Son of Rambow (2007) Comedy|Fantasy|Sci-Fi NULL
188381 Ant-Man and the Wasp (2018) Action|Adventure|Comedy|Fantasy|Sci-Fi NULL
188675 Dogman (2018) Crime|Drama NULL
188751 Mamma Mia: Here We Go Again! (2018) Comedy|Romance NULL
188797 Tag (2018) Comedy NULL
188833 The Man Who Killed Don Quixote (2018) Adventure|Comedy|Fantasy NULL
188949 Bourne Ultimatum (2007) Action|Thriller NULL
189115 Viral (2018) Documentary NULL
189333 Mission: Impossible - Fallout (2018) Action|Adventure|Thriller NULL
189381 SuperFly (2018) Action|Crime|Thriller NULL
189547 Iron Soldier (2010) Action|Sci-Fi NULL
189713 BlackKKlansman (2018) Comedy|Crime|Drama NULL
189733 The Shape of Water (2017) Drama|Thriller|Romantic|Sci-Fi|Thriller NULL
190207 Tilt (2011) Drama|Romance NULL
190209 Jeff Ross Roasts the Border (2017) Comedy NULL
190213 John From (2015) Drama NULL
190215 Liquid Truth (2017) Drama NULL
190219 Bunny (1998) Animation NULL
190229 The Zookeeper's Wife (A True Story About a Woman and Her Husband) (2016) Documentary NULL
191085 Gintama (2017) Action|Adventure|Comedy|Sci-Fi NULL
193565 Gintama: The Movie (2010) Action|Animation|Comedy|Sci-Fi anime
193565 Gintama: The Movie (2010) Action|Animation|Comedy|Sci-Fi comedy
193565 Gintama: The Movie (2010) Action|Animation|Comedy|Sci-Fi gintama
193565 Gintama: The Movie (2010) Action|Animation|Comedy|Sci-Fi remaster
193571 I Want to Know What We Saw That Day - The Movie (2013) Animation|Drama NULL
193571 Silver Spoon (2014) Comedy|Drama NULL
193573 Love Live! The School Idol Movie (2015) Animation NULL
193579 Jon Stewart Has Left the Building (2015) Documentary NULL
193581 Black Butler: Book of the Atlantic (2017) Action|Animation|Comedy|Fantasy NULL
193581 Game No Life: Zero (2017) Animation|Comedy|Fantasy NULL
193585 Flawless (2017) Drama NULL
193587 Bungo Stray Dogs: Dead Apple (2018) Action|Animation NULL
193609 Andrew Dice Clay: Dice Rules (1991) Comedy NULL
Time taken: 24.262 seconds, Fetched: 118859 rows(s)
hive> []
```

```
select m.movieid,m.movieName,m.genre,t.comment from movieDeets m right outer join tags t on (m.movieid=t.movieid);
```

```
hive> select m.movieid,m.movieName,m.genre,t.comment from movieDeets m right outer join tags t on (m.movieid=t.movieid);
Total Jobs = 1
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
004/HashTable$Stage-3/MapJoin.mapfile30:available for tag: 0 with group count: 9748 into file: file:/tmp/hadoop/b953-010523eef429/hive_2021-06-16_18-36-02/_27_700093055876880370-1-local-10
2021-06-16 18:36:08 End of local task; Time Taken: 1.0 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
No reduce operator is set to 0 since there's no reduce operator
Starting Job Job_1623835256394_0015, Tracking URL = http://somyasriv16:8080/proxy/application_1623835256394_0015/
Kill Command = /home/hadoop/hadoop/bin/mapred job -kill job_1623835256394_0015
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-16 18:36:15 693 Stage-3 map = 0%, reduce = 0%
2021-06-16 18:36:27 859 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.87 sec
2021-06-16 18:36:27 Job 1623835256394 finished in 2 seconds 870 msec
Ended Job > job_1623835256394_0015
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.87 sec HDFS Read: 127708 HDFS Write: 275946 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 870 msec
Or
08756 NULL NULL tag
Step Brothers (2008) Comedy funny
60756 Step Brothers (2008) Comedy Highly quotable
60756 Step Brothers (2008) Comedy will ferrell
89774 Warrior (2011) Drama Boxing story
89774 Warrior (2011) Drama MMA
89774 Warrior (2011) Drama Tom Hardy
106782 Wolf of Wall Street The (2013) Comedy|Crime|Drama drugs
106782 Wolf of Wall Street The (2013) Comedy|Crime|Drama Leonardo DiCaprio
106782 Wolf of Wall Street The (2013) Comedy|Crime|Drama Martin Scorsese
48516 Departed The (2006) Crime|Drama|Thriller way too long
431 Carlito's Way (1993) Crime|Drama Al Pacino
431 Carlito's Way (1993) Crime|Drama gangster
431 Carlito's Way (1993) Crime|Drama mafia
1221 Godfather: Part II The (1974) Crime|Drama Al Pacino
1221 Godfather: Part II The (1974) Crime|Drama Mafia
5994 Pianist The (2002) Drama|War holocaust
5994 Pianist The (2002) Drama|War true story
54665 Little Children (2001) Crime|Drama|Mystery twist ending
52604 Fracture (2007) Crime|Drama|Mystery|Thriller Anthony Hopkins
52604 Fracture (2007) Crime|Drama|Mystery|Thriller courtroom drama
52604 Fracture (2007) Crime|Drama|Mystery|Thriller twist ending
88094 Upside Down: The Creation Records Story (2010) Documentary bridget
88094 Upside Down: The Creation Records Story (2010) Documentary indie record label
88094 Upside Down: The Creation Records Story (2010) Documentary music
144210 Just Eat It: A Food Waste Story (2014) Documentary dumpster diving
144210 Just Eat It: A Food Waste Story (2014) Documentary sustainability
1569 My Best Friend's Wedding (1997) Comedy|Romance romantic comedy
1569 My Best Friend's Wedding (1997) Comedy|Romance wedding
118985 Big Eyes (2014) Drama painter
118985 Big Eyes (2014) Drama painter
109487 Interstellar (2014) Sci-Fi|IMAX black hole
109487 Interstellar (2014) Sci-Fi|IMAX sci-fi
```

```
hive> select m.movieid,m.movieName,m.genre,t.comment from movieDeets m right outer join tags t on (m.movieid=t.movieid);
Total Jobs = 1
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
2959 Fight Club (1999) Action|Crime|Drama|Thriller classic
2959 Fight Club (1999) Action|Crime|Drama|Thriller clever
2959 Fight Club (1999) Action|Crime|Drama|Thriller complicated
2959 Fight Club (1999) Action|Crime|Drama|Thriller consumerism
2959 Fight Club (1999) Action|Crime|Drama|Thriller crime
2959 Fight Club (1999) Action|Crime|Drama|Thriller dark
2959 Fight Club (1999) Action|Crime|Drama|Thriller dark comedy
2959 Fight Club (1999) Action|Crime|Drama|Thriller David Fincher
2959 Fight Club (1999) Action|Crime|Drama|Thriller disturbing
2959 Fight Club (1999) Action|Crime|Drama|Thriller double life
2959 Fight Club (1999) Action|Crime|Drama|Thriller Edward Norton
2959 Fight Club (1999) Action|Crime|Drama|Thriller fight
2959 Fight Club (1999) Action|Crime|Drama|Thriller great acting
2959 Fight Club (1999) Action|Crime|Drama|Thriller helena bonham carter
2959 Fight Club (1999) Action|Crime|Drama|Thriller imaginary friend
2959 Fight Club (1999) Action|Crime|Drama|Thriller imbd top 250
2959 Fight Club (1999) Action|Crime|Drama|Thriller mental illness
2959 Fight Club (1999) Action|Crime|Drama|Thriller mind-blowing
2959 Fight Club (1999) Action|Crime|Drama|Thriller mindfuck
2959 Fight Club (1999) Action|Crime|Drama|Thriller narrated
2959 Fight Club (1999) Action|Crime|Drama|Thriller Nudity (Topless)
2959 Fight Club (1999) Action|Crime|Drama|Thriller Palahnukic
2959 Fight Club (1999) Action|Crime|Drama|Thriller philosophical
2959 Fight Club (1999) Action|Crime|Drama|Thriller philosophical
2959 Fight Club (1999) Action|Crime|Drama|Thriller postmodern
2959 Fight Club (1999) Action|Crime|Drama|Thriller powerful ending
2959 Fight Club (1999) Action|Crime|Drama|Thriller psychological
2959 Fight Club (1999) Action|Crime|Drama|Thriller psychological thriller
2959 Fight Club (1999) Action|Crime|Drama|Thriller psychology
2959 Fight Club (1999) Action|Crime|Drama|Thriller quirky
2959 Fight Club (1999) Action|Crime|Drama|Thriller satirical
2959 Fight Club (1999) Action|Crime|Drama|Thriller schizophrenia
2959 Fight Club (1999) Action|Crime|Drama|Thriller social commentary
2959 Fight Club (1999) Action|Crime|Drama|Thriller societal criticism
2959 Fight Club (1999) Action|Crime|Drama|Thriller surreal
2959 Fight Club (1999) Action|Crime|Drama|Thriller surreal
2959 Fight Club (1999) Action|Crime|Drama|Thriller TERRORISM
2959 Fight Club (1999) Action|Crime|Drama|Thriller thought-provoking
2959 Fight Club (1999) Action|Crime|Drama|Thriller twist
2959 Fight Club (1999) Action|Crime|Drama|Thriller twist ending
2959 Fight Club (1999) Action|Crime|Drama|Thriller violence
2959 Fight Club (1999) Action|Crime|Drama|Thriller violence
273 Mary Shelley's Frankenstein (Frankenstein) (1994) Drama|Horror|Sci-Fi gothic
1357 Shine (1996) Drama|Romance music
1940 Tom Jones (1963) Adventure|Comedy|Romance British
3578 Gladiator (2006) Action|Adventure|Drama Romans
5010 Star Wars: Episode I (1999) Comedy|Science Fiction space
6187 Night of the Shooting Stars (Notte di San Lorenzo La) (1982) Drama|War World War II
7382 I'm Not Scared (Io non ho paura) (2003) Drama|Mystery|Thriller for katie
7936 Shame (Skammen) (1968) Drama|War austere
3265 Hard-Bolled (Lat sau san taam) (1992) Action|Crime|Drama|Thriller gun fu
3265 Hard-Bolled (Lat sau san taam) (1992) Action|Crime|Drama|Thriller heroic bloodshed
100240 John Wick: Chapter Two (2017) Action|Crime|Thriller Heroic Bloodshed
Time Taken: 20.25 seconds, Fetched: 3684 row(s)
```

## SNAPSHOTS FROM HADOOP CLUSTER:-

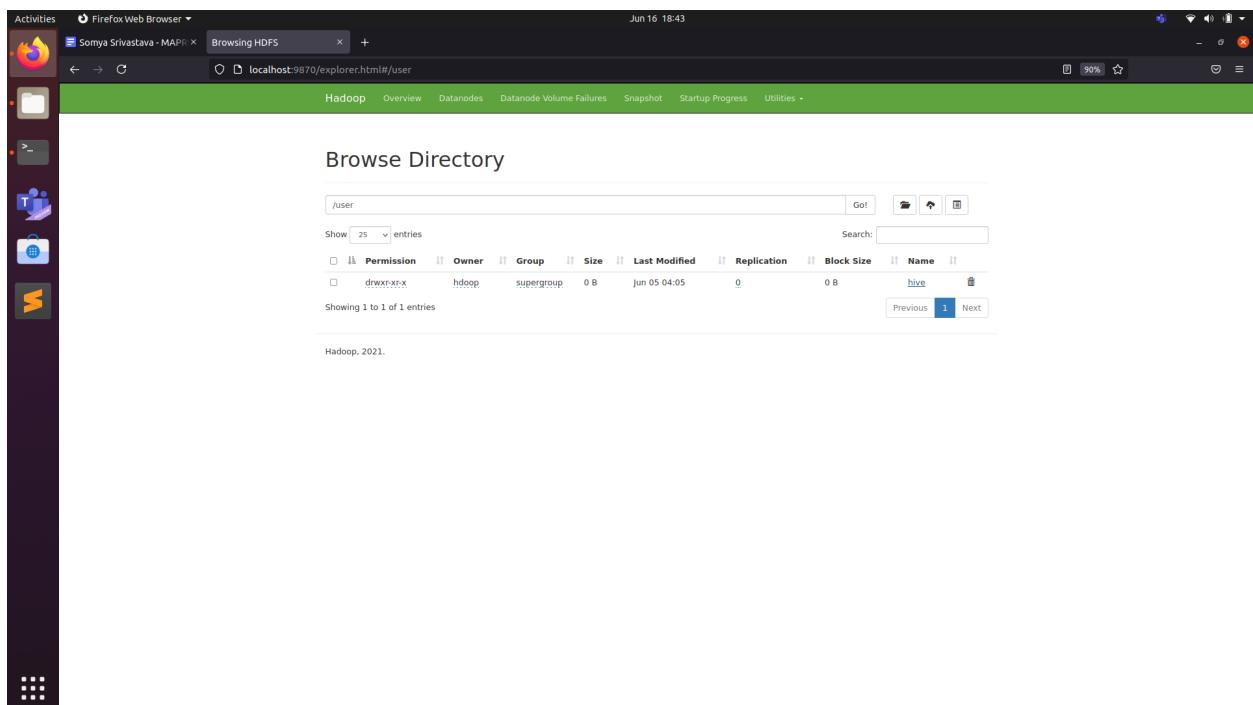


A screenshot of a Firefox browser window titled "Browsing HDFS" at "localhost:9870/explorer.html#/". The title bar also shows "Somya Srivastava - MAPR" and the date "Jun 16 18:43". The browser interface has a green header bar with tabs for "Hadoop", "Overview", "Datanodes", "Datanode Volume Failures", "Snapshot", "Startup Progress", and "Utilities". On the left is a dark sidebar with icons for file operations. The main content area is titled "Browse Directory" and shows a table of files under the root directory "/". The table includes columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. The entries are:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hdooop	supergroup	0 B	Jun 13 02:50	0	0 B	movieshadoop
drwxr-xr-x	hdooop	supergroup	0 B	May 26 21:21	0	0 B	program3
drwxr-xr-x	hdooop	supergroup	0 B	May 26 20:36	0	0 B	program4
drwxr-xr-x	hdooop	supergroup	0 B	May 28 00:29	0	0 B	pyprogram
drwxrwxr-x	hdooop	supergroup	0 B	Jun 13 01:23	0	0 B	tmp
drwxr-xr-x	hdooop	supergroup	0 B	Jun 05 04:05	0	0 B	user

Showing 1 to 6 of 6 entries.

Hadoop, 2021.



A screenshot of a Firefox browser window titled "Browsing HDFS" at "localhost:9870/explorer.html#/user". The title bar shows "Somya Srivastava - MAPR" and the date "Jun 16 18:43". The browser interface has a green header bar with tabs for "Hadoop", "Overview", "Datanodes", "Datanode Volume Failures", "Snapshot", "Startup Progress", and "Utilities". On the left is a dark sidebar with icons for file operations. The main content area is titled "Browse Directory" and shows a table of files under the "/user" directory. The table includes columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. The entries are:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hdooop	supergroup	0 B	Jun 05 04:05	0	0 B	hive

Showing 1 to 1 of 1 entries.

Hadoop, 2021.

Activities Firefox Web Browser Somya Srivastava - MAPR Jun 16 18:43 + localhost:9870/explorer.html#/user/hive

Browsing HDFS Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Browse Directory

/user/hive Go! Search:

Show 25 entries Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxrwxr-x	hdoop	supergroup	0 B	Jun 16 03:12	0	0 B	warehouse

Showing 1 to 1 of 1 entries Previous 1 Next

Hadoop, 2021.

Activities Firefox Web Browser Somya Srivastava - MAPR Jun 16 18:43 + localhost:9870/explorer.html#/user/hive/warehouse

Browsing HDFS Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Browse Directory

/user/hive/warehouse Go! Search:

Show 25 entries Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	hdoop	supergroup	0 B	Jun 13 21:41	0	0 B	employee.db
□	drwxr-xr-x	hdoop	supergroup	0 B	Jun 13 14:10	0	0 B	hospital.db
□	drwxr-xr-x	hdoop	supergroup	0 B	Jun 09 16:37	0	0 B	nmit.db
□	drwxr-xr-x	hdoop	supergroup	0 B	Jun 16 03:53	0	0 B	x.db

Showing 1 to 4 of 4 entries Previous 1 Next

Hadoop, 2021.

Activities Firefox Web Browser Somya Srivastava - MAPR Browsing HDFS Jun 16 18:43

localhost:9870/explorer.html#/user/hive/warehouse/x.db

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Browse Directory

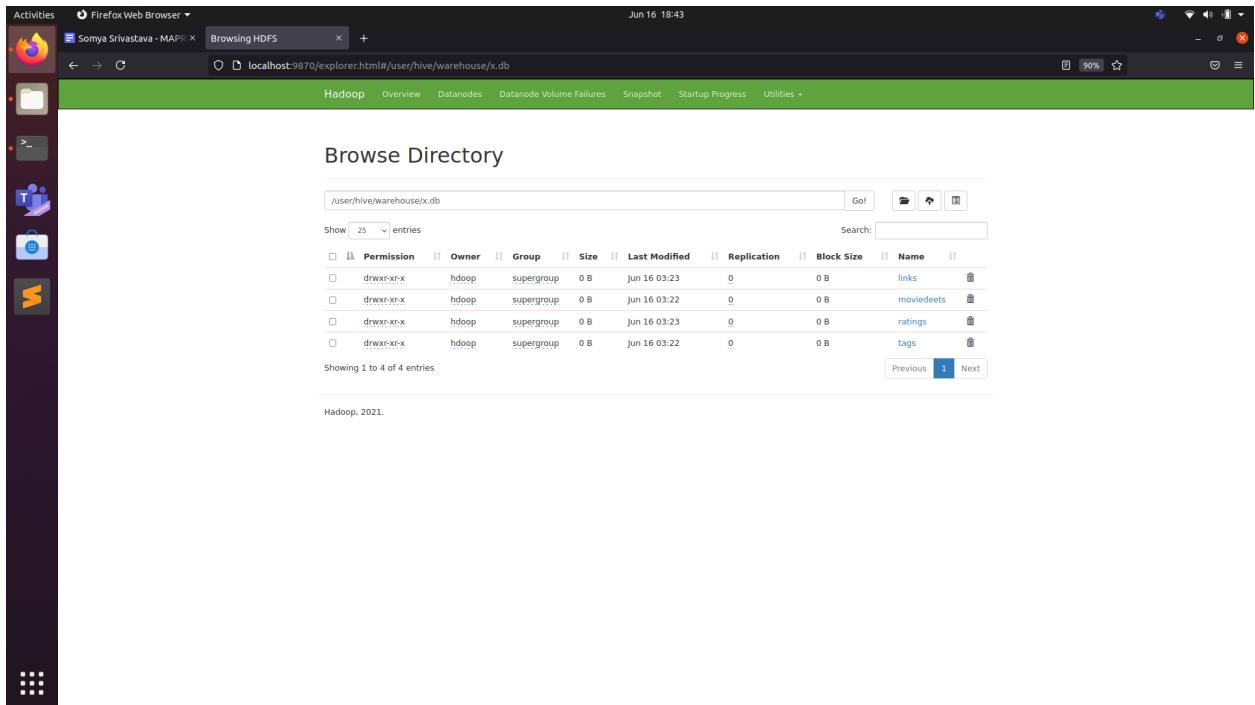
/user/hive/warehouse/x.db

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	Jun 16 03:23	0	0 B	links
drwxr-xr-x	hadoop	supergroup	0 B	Jun 16 03:22	0	0 B	movie��ets
drwxr-xr-x	hadoop	supergroup	0 B	Jun 16 03:23	0	0 B	ratings
drwxr-xr-x	hadoop	supergroup	0 B	Jun 16 03:22	0	0 B	tags

Showing 1 to 4 of 4 entries Previous 1 Next

Hadoop, 2021.



Activities Firefox Web Browser Somya Srivastava - MAPR Browsing HDFS Jun 16 18:44

localhost:9870/explorer.html#/user/hive/warehouse/x.db/links

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Browse Directory

/user/hive/warehouse/x.db/links

Show 25 entries Search:

Block Size	Name
28 MB	links.csv

File information - links.csv

Download Head the file (first 32K) Tail the file (last 32K)

Block information - Block 0

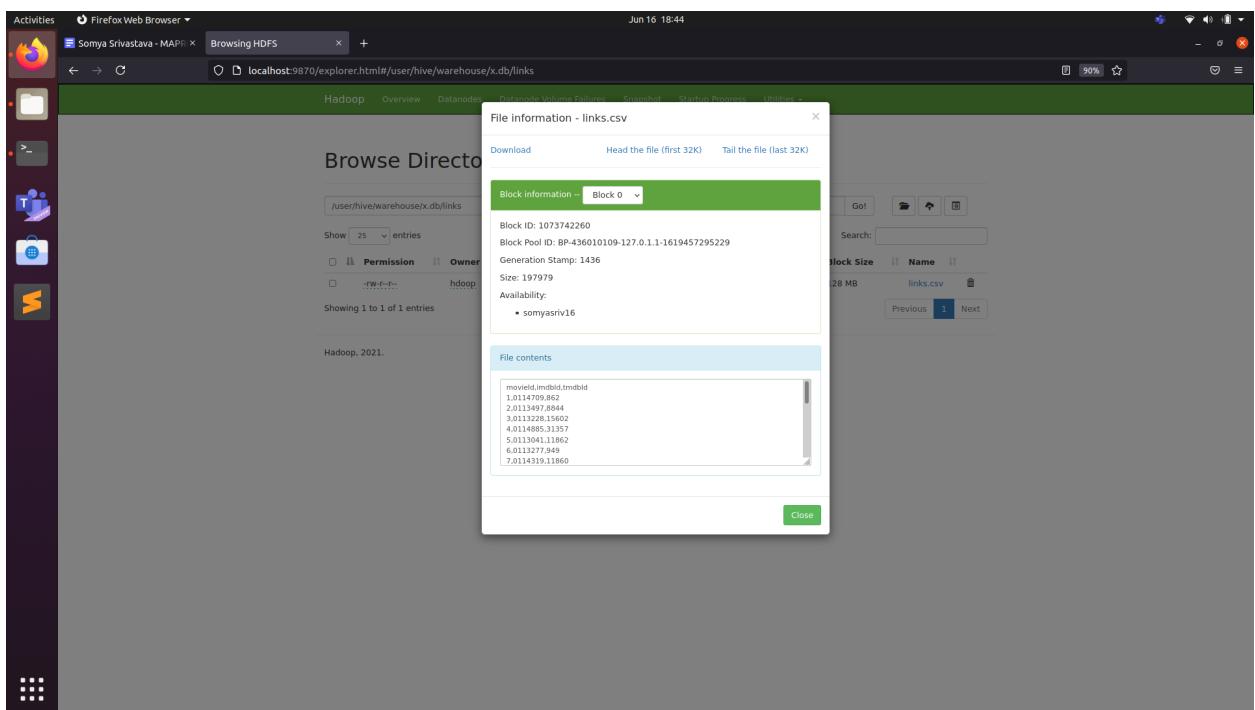
Block ID: 1073742260  
Block Pool ID: BP-436010109-127.0.1.1-1619457295229  
Generation Stamp: 1436  
Size: 197979  
Availability:

- somyasriv16

File contents

```
mvnvid印edt.lmdbd
1.0114709.862
2.0113497.8844
3.0113228.15602
4.0114885.31357
5.0113228.15602
6.0113277.949
7.0114319.11860
```

Close



Activities Firefox Web Browser Jun 16 18:44

Somya Srivastava - MAPR Browsing HDFS + localhost:9870/explorer.html#/user/hive/warehouse/x.db/moviedeets

Hadoop Overview Datanodes DataNodes Utilities +

Browse Directory /user/hive/warehouse/x.db/moviedeets

Show 25 entries

Permission	Owner
-rw-r--r--	hadoop

Showing 1 to 6 of 6 entries

Hadoop, 2021.

File information - movies.csv

Download Head the file (first 32K) Tail the file (last 32K)

Block information - Block 0

Block ID: 1073742258  
Block Pool ID: BP-436010109-127.0.1.1-1619457295229  
Generation Stamp: 1434  
Size: 490275  
Availability: somysriv16

File contents

```
moviedb/moviedata/movies.csv
1.Toy Story (1995).Action|Animation|Children|Comedy|Fantasy.
2.An Officer and a Gentleman (1982).Action|Drama|Romance.
3.Gangs of New York (2002).Action|Crime|Drama|Thriller.
4.Waiting to Exhale (1995).Comedy|Drama|Romance.
5.Father of the Bride Part II (1995).Comedy.
6.Hair (1995).Action|Crime|Thriller.
7.Sabrina (1995).Comedy|Romance.
```

Close

Detailed description: This screenshot shows the HDFS browser interface in Firefox. A modal window titled 'File information - movies.csv' is open, displaying details about the first block of the file. It includes the block ID, pool ID, generation stamp, size (490275), and owner (somyasriv16). Below this, the 'File contents' section shows the first few lines of the CSV file, listing various movies with their genres. The background shows a list of other files in the directory, all owned by 'hadoop'.

Activities Firefox Web Browser Jun 16 18:45

Somya Srivastava - MAPR Browsing HDFS + localhost:9870/explorer.html#/user/hive/warehouse/x.db/ratings

Hadoop Overview Datanodes DataNodes Volume Failures Snarls Startups Processes Utilities +

Browse Directory /user/hive/warehouse/x.db/ratings

Show 25 entries

Permission	Owner
-rw-r--r--	hadoop

Showing 1 to 1 of 1 entries

Hadoop, 2021.

File information - ratings.csv

Download Head the file (first 32K) Tail the file (last 32K)

Block information - Block 0

Block ID: 1073742261  
Block Pool ID: BP-436010109-127.0.1.1-1619457295229  
Generation Stamp: 1437  
Size: 2483723  
Availability: somysriv16

File contents

```
user_id,movie_id,rating,timestamp
1,3,4.0,964980703
1,3,4.0,964981247
1,6,4.0,964982224
1,47,5.0,964983815
1,101,5.0,964984231
1,70,3,0.964982400
1,101,5.0,964980868
```

Close

Detailed description: This screenshot shows the HDFS browser interface in Firefox. A modal window titled 'File information - ratings.csv' is open, displaying details about the first block of the file. It includes the block ID, pool ID, generation stamp, size (2483723), and owner (somyasriv16). Below this, the 'File contents' section shows the first few lines of the CSV file, listing user IDs, movie IDs, ratings, and timestamps.

