

NoSQL Assignment 2 – Part B

IMT2021008 – Sheikh Muteeb

IMT2021003 – Keshav Chandak

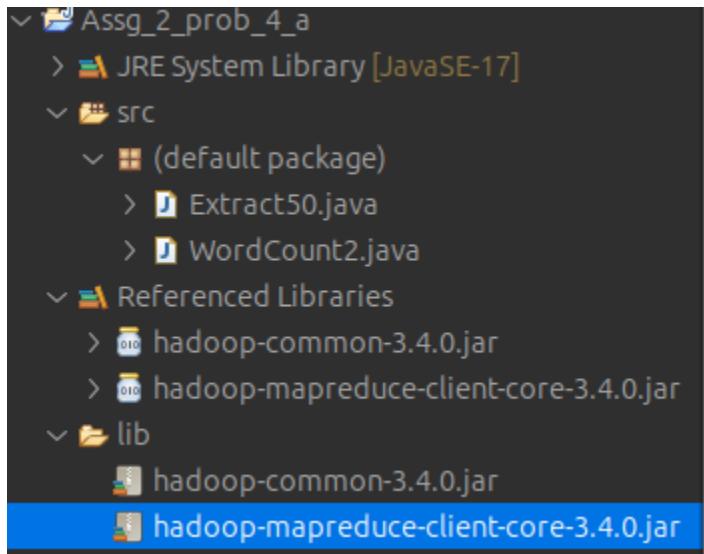
IMT2021007 – Sunny Kaushik

IMT2021076 – Devendara Rishi Nelapati

Problem 4

A

1. File Structure:



2. Outputs:

9291	d	658336
0 54086	s	375301
00 2754	pos	274361
000 13789	quot	268974
0000 35	w	222021
00000 5	m	173453
000000 91	l	169347
0000000 1	wi	138141
000000000000 1	re	129404
000000001 1	r	113910
000000017 1	g	109626
00000001__ 1	p	93316
00000002 2	cy	90251
0000000221 1	c	89608
00000003 3	n	85713
	h	78773

Output from WordCount2.java

Output from Extract50.java (first few lines of the top 50 words)

3. Execution Time:

3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/AS hadoop jar prob_4_a.jar WordCount2 ../Wikipedia-EN-20120601_ARTICLES ./output -skippatterns ../stopwords.txt
2025-03-23 07:10:50,397 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
```



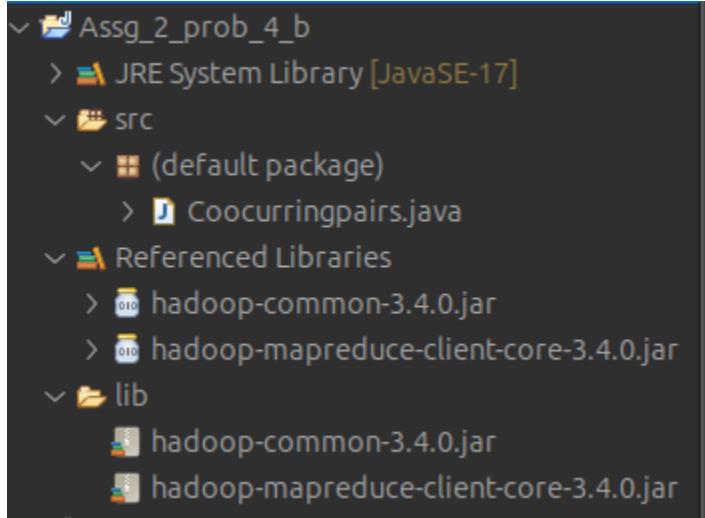
3.2. Execution end time and successful completion.

```
2025-03-23 07:18:23,801 INFO mapred.LocalJobRunner: Finishing task: attempt_local707555935_0001_r_000000_0
2025-03-23 07:18:23,801 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 07:18:24,003 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 07:18:25,003 INFO mapreduce.Job: Job job_local707555935_0001 completed successfully
2025-03-23 07:18:25,234 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1698241974654
    FILE: Number of bytes written=1500201987969
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=19260391
    Map output bytes=166481643
    Map output materialized bytes=205062438
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=608968
    Reduce shuffle bytes=205062438
    Reduce input records=19260391
    Reduce output records=608968
    Spilled Records=38520782
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2922
    Total committed heap usage (bytes)=41234122932224
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=626483
```

3.3 Total Execution Time: **455 seconds.**

B

1. File Structure:



2. Outputs:

0,0	2786	0,0	6053	0,0	10457	0,0	14772
0,1	3383	0,1	6042	0,1	10598	0,1	13562
0,10	503	0,10	887	0,10	1247	0,10	1571
0,2	2430	0,2	3908	0,2	6968	0,2	8546
0,2010	156	0,2010	549	0,2010	723	0,2010	1004
0,2011	44	0,2011	117	0,2011	230	0,2011	368
0,3	1776	0,3	2710	0,3	5300	0,3	6419
0,4	1320	0,4	2136	0,4	4544	0,4	5421
0,5	1343	0,5	2108	0,5	4433	0,5	5254
0,b	56	0,b	120	0,b	192	0,b	291
0,c	186	0,c	310	0,c	397	0,c	496

d = 1

d = 2

d = 3

d = 4

3. Execution Time:

3.1. d = 1

3.1.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/B$ hadoop jar prob_4_b.jar Cooccurringpairs ../Wikimedia-EN-20120601_ARTICLES ./output/d1 ..//A/top50.txt -d 1
2025-03-23 06:41:58,781 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 06:41:58,880 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 06:41:59,880 INFO impl.MetricsSystemImpl: Jobtracker metrics system started
2025-03-23 06:41:59,931 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to
edy this.
2025-03-23 06:41:59,932 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 06:41:59,301 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 06:41:59,517 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 06:41:59,623 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local120537615_0001
2025-03-23 06:41:59,623 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 06:41:59,781 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local120537615_0001_fbbce161-eea3-4c4a-8b85-c4aa4ac06db6/top50.txt
/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/B/top50.txt
2025-03-23 06:41:59,781 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/l
/job_local120537615_0001_fbbce161-eea3-4c4a-8b85-c4aa4ac06db6/top50.txt
2025-03-23 06:41:59,848 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 06:41:59,849 INFO mapreduce.Job: Running job: job_local120537615_0001
2025-03-23 06:41:59,849 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 06:41:59,854 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 06:41:59,855 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 06:41:59,855 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 06:41:59,856 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.1.2. Execution end time and successful completion.

```
2025-03-23 06:44:22,348 INFO mapred.LocalJobRunner: Finishing task: attempt_local120537615_0001_r_000000_0
2025-03-23 06:44:22,348 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 06:44:22,979 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 06:44:23,980 INFO mapreduce.Job: Job job_local120537615_0001 completed successfully
2025-03-23 06:44:24,221 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697809666863
    FILE: Number of bytes written=46778178434
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=214461
    Map output bytes=2145993
    Map output materialized bytes=2634915
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=955
    Reduce shuffle bytes=2634915
    Reduce input records=214461
    Reduce output records=955
    Spilled Records=428922
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2204
    Total committed heap usage (bytes)=41234873712640
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=8126
```

3.1.3. Total Execution Time: 146 seconds.

3.2. d = 2

3.2.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/B5 hadoop jar prob_4_b.jar Cooccurringpairs ..//Wikipedia-EN-20120601_ARTICLES ./output/d2 ..//A/top50.txt -d 2
2025-03-23 06:46:27,049 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 06:46:27,149 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 06:46:27,150 INFO impl.MetricsSystemImpl: Jobtracker metrics system started
2025-03-23 06:46:27,197 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to
do this.
2025-03-23 06:46:27,198 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
```

[]

3.2.2. Execution end time and successful completion.

```
2025-03-23 06:48:58,112 INFO mapred.LocalJobRunner: Finishing task: attempt_local794901536_0001_r_000000_0
2025-03-23 06:48:58,113 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 06:48:58,242 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 06:48:59,242 INFO mapreduce.Job: Job job_local794901536_0001 completed successfully
2025-03-23 06:48:59,461 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697813866897
    FILE: Number of bytes written=61986245594
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=385954
    Map output bytes=3903024
    Map output materialized bytes=4734932
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1041
    Reduce shuffle bytes=4734932
    Reduce input records=385954
    Reduce output records=1041
    Spilled Records=771908
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2041
    Total committed heap usage (bytes)=41233351180288
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9127
```

3.2.3. Total Execution Time: **152 seconds.**

3.3. d = 3

3.3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/B$ hadoop jar prob_4_b.jar Cooccurringpairs ../Wikipedia-EN-20120601_ARTICLES ./output/d3 ./A/top50.txt -d 3
2025-03-23 06:50:42,941 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 06:50:43,034 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 06:50:43,034 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 06:50:43,085 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to
do this.
2025-03-23 06:50:43,086 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
```



3.3.2. Execution end time and successful completion.

```
2025-03-23 06:53:21,953 INFO mapred.LocalJobRunner: Finishing task: attempt_local1709763063_0001_r_000000_0
2025-03-23 06:53:21,953 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 06:53:22,190 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 06:53:23,191 INFO mapreduce.Job: Job job_local1709763063_0001 completed successfully
2025-03-23 06:53:23,416 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697818364481
    FILE: Number of bytes written=78383881580
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=572168
    Map output bytes=5779388
    Map output materialized bytes=6983724
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1076
    Reduce shuffle bytes=6983724
    Reduce input records=572168
    Reduce output records=1076
    Spilled Records=1144336
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2260
    Total committed heap usage (bytes)=41239346937856
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9619
```

3.3.3. Total Execution Time: **161 seconds.**

3.4. d = 4

3.4.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/B$ hadoop jar prob_4_b.jar Cooccurringpairs ..../Wikipedia-EN-20120601_ARTICLES ./output/d4 ..../A/top50.txt -d 4
2025-03-23 06:59:50,841 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 06:59:50,940 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 06:59:50,940 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 06:59:50,987 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to
do this.
2025-03-23 06:59:50,988 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
```

□

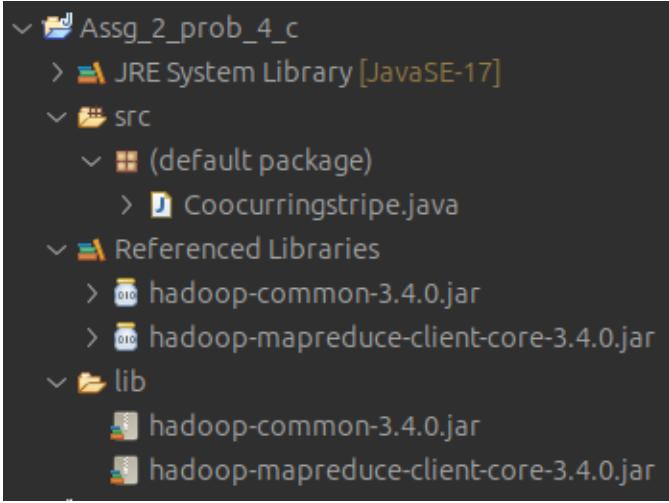
3.4.2. Execution end time and successful completion.

```
2025-03-23 07:02:27,773 INFO mapred.LocalJobRunner: Finishing task: attempt_local896282142_0001_r_000000_0
2025-03-23 07:02:27,773 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 07:02:28,192 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 07:02:29,193 INFO mapreduce.Job: Job job_local896282142_0001 completed successfully
2025-03-23 07:02:29,397 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697822316463
    FILE: Number of bytes written=92410102266
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=737167
    Map output bytes=7425381
    Map output materialized bytes=8959715
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1099
    Reduce shuffle bytes=8959715
    Reduce input records=737167
    Reduce output records=1099
    Spilled Records=1474334
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2110
    Total committed heap usage (bytes)=41238696820736
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9950
```

3.4.3. Total Execution Time: 159 seconds.

C

1. File Structure:



2. Outputs:

```
0 [2786, 1676, 226, 1009, 137, 23, 736, 608, 730, 16, 55, 0, 0, 15, 55, 0, 34, 6, 14, 15, 0, 101, 0, 17, 0, 58, 6, 0, 118, 61, 143, 63, 0, 123, 4, 4, 30, 2, 1, 10, 0, 0, 0, 17, 35, 60, 0, 0, 3, 10]
1 [1707, 2162, 390, 2340, 375, 250, 1169, 1017, 1045, 54, 86, 3, 2, 30, 64, 1, 69, 26, 53, 52, 1, 225, 0, 72, 0, 35, 13, 4, 134, 131, 228, 316, 0, 378, 60, 8, 54, 3, 16, 50, 1, 42, 0, 51, 57, 76, 0, 0, 8, 57]
10 [283, 357, 388, 313, 244, 205, 244, 267, 308, 25, 45, 1, 0, 13, 30, 2, 18, 10, 12, 10, 0, 277, 0, 12, 0, 6, 1, 0, 12, 52, 75, 27, 0, 140, 1, 6, 20, 3, 4, 1, 0, 10, 0, 11, 2, 57, 0, 0, 1, 36]
2 [1421, 1550, 276, 1254, 268, 212, 1355, 969, 910, 101, 114, 0, 1, 71, 43, 1, 87, 20, 51, 91, 3, 252, 0, 49, 1, 33, 10, 2, 109, 128, 170, 219, 0, 603, 59, 7, 109, 9, 17, 37, 1, 65, 2, 22, 45, 42, 0, 0, 5, 79]
20 [19, 129, 706, 72, 168, 211, 57, 42, 37, 1, 10, 0, 1, 10, 7, 5, 8, 16, 3, 15, 1, 1450, 0, 9, 0, 9, 2, 0, 52, 8, 14, 150, 0, 296, 8, 1, 13, 0, 10, 3, 0, 83, 4, 49, 4, 7, 1, 0, 17, 14]
2010 [21, 99, 744, 80, 46, 186, 62, 30, 59, 5, 4, 0, 0, 3, 7, 5, 5, 20, 3, 7, 0, 1585, 0, 5, 0, 7, 3, 0, 26, 4, 12, 103, 0, 483, 3, 6, 9, 0, 14, 12, 1, 47, 1, 26, 5, 3, 0, 0, 18, 2]
2011 [2140, 954, 277, 1003, 335, 1061, 998, 906, 19, 74, 0, 0, 85, 27, 4, 27, 7, 30, 33, 0, 291, 0, 19, 8, 1, 49, 116, 82, 110, 0, 281, 14, 4, 66, 2, 8, 9, 0, 19, 0, 4, 15, 35, 0, 0, 7, 32]
3 [712, 761, 219, 732, 289, 202, 723, 673, 823, 16, 60, 0, 38, 17, 2, 23, 12, 37, 36, 0, 171, 0, 8, 1, 13, 2, 0, 39, 97, 51, 104, 0, 225, 10, 8, 28, 1, 6, 17, 0, 24, 1, 5, 11, 36, 0, 0, 28, 37]
4 [712, 761, 219, 732, 289, 202, 723, 673, 823, 16, 60, 0, 38, 17, 2, 23, 12, 37, 36, 0, 171, 0, 8, 1, 13, 2, 0, 39, 97, 51, 104, 0, 225, 10, 8, 28, 1, 6, 17, 0, 24, 1, 5, 11, 36, 0, 0, 28, 37]
5 [613, 691, 306, 590, 280, 217, 565, 512, 633, 23, 80, 0, 1, 8, 24, 1, 16, 11, 22, 23, 0, 197, 0, 23, 0, 43, 3, 1, 26, 177, 66, 105, 0, 240, 0, 10, 20, 1, 9, 18, 0, 10, 0, 7, 24, 36, 0, 0, 29, 30]
```

d = 1

```
0 [6053, 2866, 368, 1783, 442, 52, 1130, 963, 1035, 40, 116, 0, 0, 47, 107, 1, 78, 17, 80, 28, 1, 232, 0, 51, 0, 99, 11, 0, 169, 158, 235, 102, 0, 299, 23, 6, 89, 5, 7, 39, 0, 20, 0, 30, 74, 151, 0, 0, 49, 27]
1 [3176, 5265, 777, 4287, 785, 605, 2129, 1695, 1669, 117, 173, 3, 2, 98, 133, 9, 141, 183, 137, 188, 1, 635, 0, 176, 1, 130, 19, 5, 246, 447, 499, 453, 0, 886, 115, 13, 256, 3, 28, 102, 1, 98, 0, 99, 110, 143, 0, 0, 127, 173]
10 [519, 677, 866, 504, 598, 374, 391, 374, 415, 45, 77, 1, 0, 22, 56, 5, 33, 25, 22, 30, 0, 756, 0, 44, 3, 19, 5, 0, 293, 0, 293, 8, 131, 2, 0, 175, 108]
2 [2125, 2766, 513, 3665, 656, 554, 2564, 1669, 1437, 146, 195, 0, 3, 154, 142, 2, 137, 63, 111, 166, 4, 530, 0, 136, 1, 132, 16, 2, 288, 379, 335, 331, 0, 965, 104, 14, 282, 9, 36, 86, 1, 98, 3, 106, 89, 126, 1, 0, 128, 283]
2010 [107, 296, 985, 178, 696, 320, 158, 197, 108, 20, 23, 0, 1, 27, 17, 16, 20, 67, 18, 26, 1, 23, 15, 1753, 0, 26, 0, 19, 3, 0, 65, 20, 34, 165, 0, 233, 29, 6, 126, 0, 18, 8, 1, 136, 6, 68, 15, 20, 1, 0, 1020, 20]
2011 [65, 252, 1052, 184, 145, 702, 127, 98, 126, 19, 17, 2, 0, 23, 20, 9, 11, 88, 11, 11, 0, 1777, 0, 21, 22, 10, 0, 33, 18, 21, 122, 2, 573, 15, 15, 85, 0, 24, 22, 1, 71, 8, 48, 10, 9, 1, 0, 1084, 5]
3 [1580, 1773, 443, 1845, 567, 492, 2339, 1816, 1427, 55, 132, 0, 1, 119, 66, 9, 63, 46, 82, 99, 3, 564, 0, 57, 0, 93, 6, 1, 182, 306, 166, 283, 0, 655, 44, 11, 198, 3, 22, 31, 1, 39, 1, 34, 47, 96, 1, 0, 174, 127]
4 [1173, 1368, 383, 1410, 472, 499, 1277, 1369, 1396, 37, 111, 0, 0, 65, 65, 2, 46, 30, 61, 82, 0, 414, 0, 46, 1, 69, 5, 0, 64, 250, 118, 172, 0, 513, 42, 11, 163, 2, 14, 53, 1, 38, 2, 24, 24, 94, 1, 0, 85, 100]
5 [1073, 1221, 531, 1131, 486, 546, 1013, 1204, 1287, 39, 113, 1, 2, 36, 62, 4, 56, 35, 43, 60, 0, 533, 0, 43, 0, 100, 4, 61, 299, 123, 157, 0, 455, 21, 12, 100, 2, 20, 43, 0, 393, 1, 35, 39, 82, 1, 0, 153, 116]
```

d = 2

```
0 [10457, 5591, 521, 3981, 536, 88, 2995, 2886, 2890, 73, 156, 0, 2, 73, 146, 2, 134, 38, 186, 78, 1, 306, 0, 79, 0, 122, 16, 2, 288, 228, 361, 160, 0, 431, 52, 13, 218, 7, 14, 62, 0, 31, 0, 33, 95, 194, 0, 0, 115, 53]
1 [5007, 8171, 1231, 6152, 863, 719, 3345, 2715, 2652, 185, 270, 3, 3, 182, 227, 13, 220, 186, 284, 166, 2, 1029, 0, 323, 1, 178, 30, 6, 303, 686, 678, 583, 0, 1167, 188, 26, 524, 4, 39, 179, 1, 138, 1, 153, 154, 188, 0, 0, 370, 205]
10 [726, 934, 1441, 739, 737, 639, 617, 564, 575, 65, 189, 1, 1, 38, 85, 9, 57, 60, 44, 59, 0, 1854, 0, 66, 4, 30, 6, 0, 66, 255, 195, 100, 1, 434, 25, 14, 278, 8, 23, 31, 0, 98, 1, 48, 10, 153, 3, 0, 459, 117]
2 [2987, 4052, 944, 5539, 802, 666, 3610, 2231, 1972, 216, 298, 1, 6, 234, 215, 5, 208, 113, 176, 247, 5, 912, 0, 228, 1, 184, 27, 2, 265, 519, 498, 428, 0, 1354, 186, 18, 487, 15, 53, 151, 2, 153, 5, 134, 111, 168, 1, 0, 287, 235]
2010 [187, 487, 1085, 294, 1850, 976, 239, 172, 191, 37, 144, 0, 1, 53, 29, 16, 42, 119, 26, 38, 3, 283, 0, 51, 1, 26, 4, 1, 73, 45, 57, 190, 1, 549, 41, 12, 379, 2, 33, 13, 2, 166, 10, 119, 23, 32, 1, 0, 1246, 24]
2011 [142, 391, 1151, 278, 320, 1988, 199, 145, 179, 35, 30, 2, 0, 51, 30, 12, 22, 142, 23, 24, 0, 2567, 0, 40, 1, 37, 13, 0, 46, 52, 44, 15, 2, 25, 17, 15, 2, 0, 1221, 10]
3 [2305, 2699, 768, 2630, 680, 545, 3494, 2591, 1894, 94, 208, 3, 1, 185, 99, 14, 97, 87, 118, 171, 4, 845, 0, 92, 0, 129, 11, 3, 138, 392, 221, 296, 1, 945, 92, 23, 326, 6, 35, 76, 1, 73, 2, 52, 82, 172, 1, 0, 341, 144]
4 [1738, 2018, 651, 1937, 552, 552, 1818, 2065, 1946, 56, 163, 0, 0, 108, 102, 3, 73, 62, 88, 116, 0, 681, 0, 73, 1, 113, 8, 1, 87, 324, 178, 235, 0, 731, 83, 19, 290, 3, 21, 73, 1, 66, 6, 54, 32, 134, 1, 0, 213, 106]
5 [1543, 1781, 766, 1644, 570, 596, 1403, 1566, 2016, 65, 161, 1, 3, 59, 90, 7, 90, 67, 62, 101, 1, 733, 0, 68, 0, 115, 7, 1, 104, 381, 162, 203, 0, 666, 49, 20, 598, 4, 27, 49, 1, 424, 2, 58, 63, 187, 2, 0, 354, 122]
b [119, 245, 85, 304, 81, 72, 166, 124, 69, 783, 1028, 3, 1, 511, 255, 6, 193, 28, 188, 162, 1, 97, 5, 198, 30, 15, 103, 266, 297, 205, 0, 616, 335, 16, 424, 9, 11, 211, 13, 71, 0, 21, 98, 147, 0, 0, 47, 84]
```

d = 2

```
0 [10457, 5591, 521, 3981, 536, 88, 2995, 2886, 2890, 73, 156, 0, 2, 73, 146, 2, 134, 38, 186, 78, 1, 306, 0, 79, 0, 122, 16, 2, 288, 228, 361, 160, 0, 431, 52, 13, 218, 7, 14, 62, 0, 31, 0, 33, 95, 194, 0, 0, 115, 53]
1 [5007, 8171, 1231, 6152, 863, 719, 3345, 2715, 2652, 185, 270, 3, 3, 182, 227, 13, 220, 186, 284, 166, 2, 1029, 0, 323, 1, 178, 30, 6, 303, 686, 678, 583, 0, 1167, 188, 26, 524, 4, 39, 179, 1, 138, 1, 153, 154, 188, 0, 0, 370, 205]
10 [726, 934, 1441, 739, 737, 639, 617, 564, 575, 65, 189, 1, 1, 38, 85, 9, 57, 60, 44, 59, 0, 1854, 0, 66, 4, 30, 6, 0, 66, 255, 195, 100, 1, 434, 25, 14, 278, 8, 23, 31, 0, 98, 1, 48, 10, 153, 3, 0, 459, 117]
2 [2987, 4052, 944, 5539, 802, 666, 3610, 2231, 1972, 216, 298, 1, 6, 234, 215, 5, 208, 113, 176, 247, 5, 912, 0, 228, 1, 184, 27, 2, 265, 519, 498, 428, 0, 1354, 186, 18, 487, 15, 53, 151, 2, 153, 5, 134, 111, 168, 1, 0, 287, 235]
2010 [187, 487, 1085, 294, 1850, 976, 239, 172, 191, 37, 144, 0, 1, 53, 29, 16, 42, 119, 26, 38, 3, 283, 0, 51, 1, 26, 4, 1, 73, 45, 57, 190, 1, 549, 41, 12, 379, 2, 33, 13, 2, 166, 10, 119, 23, 32, 1, 0, 1246, 24]
2011 [142, 391, 1151, 278, 320, 1988, 199, 145, 179, 35, 30, 2, 0, 51, 30, 12, 22, 142, 23, 24, 0, 2567, 0, 40, 1, 37, 13, 0, 46, 52, 44, 15, 2, 25, 17, 15, 2, 0, 1221, 10]
3 [2305, 2699, 768, 2630, 680, 545, 3494, 2591, 1894, 94, 208, 3, 1, 185, 99, 14, 97, 87, 118, 171, 4, 845, 0, 92, 0, 129, 11, 3, 138, 392, 221, 296, 1, 945, 92, 23, 326, 6, 35, 76, 1, 73, 2, 52, 82, 172, 1, 0, 341, 144]
4 [1738, 2018, 651, 1937, 552, 552, 1818, 2065, 1946, 56, 163, 0, 0, 108, 102, 3, 73, 62, 88, 116, 0, 681, 0, 73, 1, 113, 8, 1, 87, 324, 178, 235, 0, 731, 83, 19, 290, 3, 21, 73, 1, 66, 6, 54, 32, 134, 1, 0, 213, 106]
5 [1543, 1781, 766, 1644, 570, 596, 1403, 1566, 2016, 65, 161, 1, 3, 59, 90, 7, 90, 67, 62, 101, 1, 733, 0, 68, 0, 115, 7, 1, 104, 381, 162, 203, 0, 666, 49, 20, 598, 4, 27, 49, 1, 424, 2, 58, 63, 187, 2, 0, 354, 122]
b [119, 245, 85, 304, 81, 72, 166, 124, 69, 783, 1028, 3, 1, 511, 255, 6, 193, 28, 188, 162, 1, 97, 5, 198, 30, 15, 103, 266, 297, 205, 0, 616, 335, 16, 424, 9, 11, 211, 13, 71, 0, 21, 98, 147, 0, 0, 47, 84]
```

d = 3

```
0 [14772, 7227, 694, 4772, 737, 157, 3557, 3243, 3267, 185, 285, 1, 2, 103, 183, 7, 168, 92, 149, 137, 1, 944, 0, 120, 0, 158, 21, 2, 249, 335, 484, 606, 1, 775, 74, 23, 322, 9, 25, 93, 2, 48, 0, 50, 109, 226, 0, 0, 161, 79]
1 [6335, 1186, 1666, 7563, 1143, 863, 4189, 3322, 3165, 259, 374, 4, 6, 259, 274, 17, 279, 293, 259, 218, 4, 1452, 0, 424, 1, 236, 36, 7, 409, 752, 914, 767, 1, 1566, 262, 40, 822, 4, 60, 230, 2, 229, 3, 197, 205, 236, 1, 0, 624, 248]
10 [877, 1237, 1963, 969, 890, 736, 805, 757, 743, 87, 140, 1, 2, 69, 136, 11, 82, 93, 68, 66, 1, 1293, 0, 86, 5, 41, 8, 0, 78, 325, 227, 130, 1, 579, 48, 20, 450, 8, 33, 55, 0, 114, 6, 71, 17, 183, 4, 0, 648, 143]
2 [3774, 5400, 1271, 7089, 997, 771, 4567, 2834, 2480, 271, 301, 1, 7, 296, 280, 12, 264, 170, 221, 340, 6, 1139, 0, 297, 3, 224, 36, 6, 346, 649, 635, 526, 0, 1746, 258, 24, 754, 21, 70, 202, 4, 197, 8, 179, 160, 212, 1, 0, 533, 279]
2010 [267, 647, 1228, 377, 2460, 1238, 315, 215, 308, 52, 69, 0, 1, 76, 44, 17, 59, 287, 31, 48, 3, 3232, 0, 63, 3, 45, 5, 3, 83, 69, 88, 211, 1, 723, 52, 19, 629, 13, 43, 26, 2, 196, 18, 142, 31, 43, 2, 0, 1998, 60]
2011 [211, 502, 1243, 365, 452, 2610, 216, 250, 45, 45, 2, 0, 73, 43, 14, 42, 225, 34, 32, 4, 2935, 0, 61, 1, 46, 19, 0, 53, 68, 73, 151, 2, 927, 53, 30, 58, 504, 4, 45, 47, 3, 112, 14, 103, 27, 2, 0, 1786, 85]
3 [2862, 3450, 1034, 3493, 845, 641, 4741, 3304, 2403, 146, 265, 5, 1, 230, 141, 16, 134, 142, 157, 207, 4, 1046, 0, 134, 0, 173, 15, 6, 193, 481, 306, 376, 1, 1238, 132, 29, 518, 12, 42, 97, 5, 126, 4, 83, 92, 211, 1, 0, 503, 166]
4 [2178, 2631, 901, 2531, 699, 625, 2391, 2751, 2453, 86, 201, 3, 0, 135, 136, 5, 91, 120, 115, 155, 1, 823, 0, 99, 2, 150, 16, 1, 120, 395, 284, 267, 1, 979, 112, 25, 420, 4, 33, 86, 1, 97, 9, 82, 48, 183, 1, 0, 308, 124]
5 [1987, 2208, 992, 2067, 690, 656, 1927, 1962, 2617, 95, 193, 2, 5, 92, 118, 11, 107, 105, 78, 129, 1, 0, 153, 10, 1, 131, 459, 202, 239, 1, 905, 67, 27, 769, 10, 39, 64, 2, 457, 5, 81, 73, 136, 3, 0, 485, 135]
b [186, 305, 106, 379, 91, 79, 209, 144, 87, 1155, 1177, 3, 3, 666, 305, 6, 302, 45, 236, 215, 1, 133, 0, 287, 6, 265, 36, 19, 149, 345, 369, 271, 1, 763, 404, 20, 581, 10, 21, 200, 17, 94, 10, 24, 118, 177, 0, 8, 58, 97]
```

d = 4

3. Execution Time:

3.1. d = 1

3.1.1. Execution start time.

```
* sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/C$ hadoop jar prob_4.c.jar Cooccurringstripe ..../Wikipedia-EN-20120601_ARTICLES ./output/d1 ..../A/top50.txt -d 1
2025-03-23 07:32:29.082 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 07:32:29.173 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 07:32:29.173 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 07:32:29.215 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to r
eady this.
2025-03-23 07:32:29.217 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
```

3.1.2. Execution end time and successful completion.

```
2025-03-23 07:35:11,265 INFO mapred.LocalJobRunner: Finishing task: attempt_local891736858_0001_r_000000_0
2025-03-23 07:35:11,265 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 07:35:11,500 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 07:35:12,500 INFO mapreduce.Job: Job job_local891736858_0001 completed successfully
2025-03-23 07:35:12,747 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697892910647
    FILE: Number of bytes written=350034213758
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map -Reduce Framework
    Map input records=10000
    Map output records=214461
    Map output bytes=43553424
    Map output materialized bytes=44256807
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=44256807
    Reduce input records=214461
    Reduce output records=49
    Spilled Records=428922
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2132
    Total committed heap usage (bytes)=41239103668224
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=8963
```

3.1.3. Total Execution Time: **163.7 seconds.**

3.2. d = 2

3.2.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/C$ hadoop jar prob_4_c.jar Cooccurringstripe ../Wikipedia-EN-20120601_ARTICLES ./output/d2 ..//A/top50.txt -d 2
2025-03-23 07:38:54,412 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 07:38:54,500 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 07:38:54,500 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 07:38:54,550 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to ready this.
2025-03-23 07:38:54,551 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
```

3.2.2. Execution end time and successful completion.

```
2025-03-23 07:41:37,370 INFO mapred.LocalJobRunner: Finishing task: attempt_local1914290772_0001_r_000000_0
2025-03-23 07:41:37,370 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 07:41:37,854 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 07:41:38,855 INFO mapreduce.Job: Job job_local1914290772_0001 completed successfully
2025-03-23 07:41:39,096 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697963584277
    FILE: Number of bytes written=604506174570
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=385954
    Map output bytes=78375760
    Map output materialized bytes=79593622
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=79593622
    Reduce input records=385954
    Reduce output records=49
    Spilled Records=771908
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2999
    Total committed heap usage (bytes)=41222632636416
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9474
```

3.2.3. Total Execution Time: **164.6 seconds.**

3.3. d = 3

3.3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/C$ hadoop jar prob_4_c.jar Cooccurringstripe .. /Wikipedia-EN-20120601_ARTICLES ./output/d3 .. /A/top50.txt -d 3
2025-03-23 07:43:11,424 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 07:43:11,519 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 07:43:11,520 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 07:43:11,564 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to r
eady this.
2025-03-23 07:43:11,565 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 07:43:11,955 INFO InputFormat: Total input files to process : 10000
2025-03-23 07:43:12,182 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 07:43:12,315 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local108203982_0001
2025-03-23 07:43:12,498 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local108203982_0001_5d4ce2e0-219e-425f-a486-c32fb123a0b7/top50.txt + /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/C/top50.txt
2025-03-23 07:43:12,492 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/loc
/job_local108203982_0001_5d4ce2e0-219e-425f-a486-c32fb123a0b7/top50.txt
2025-03-23 07:43:12,555 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 07:43:12,556 INFO mapreduce.Job: Running job: job_local108203982_0001
2025-03-23 07:43:12,556 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 07:43:12,564 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 07:43:12,566 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 07:43:12,566 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 07:43:12,566 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.3.2. Execution end time and successful completion.

```
2025-03-23 07:46:03,079 INFO mapred.LocalJobRunner: Finishing task: attempt_local108203982_0001_r_000000_0
2025-03-23 07:46:03,079 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 07:46:03,706 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 07:46:04,707 INFO mapreduce.Job: Job job_local108203982_0001 completed successfully
2025-03-23 07:46:04,925 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1698040310797
    FILE: Number of bytes written=882201644916
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=572168
    Map output bytes=116180378
    Map output materialized bytes=117956882
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=117956882
    Reduce input records=572168
    Reduce output records=49
    Spilled Records=1144336
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2358
    Total committed heap usage (bytes)=41238967353344
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9810
```

3.3.3. Total Execution Time: 168.6 seconds.

3.4. d = 4

3.4.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/C$ hadoop jar prob_4_c.jar Coocurringstripe ../Wikimedia-EN-20120601_ARTICLES ./output/d4 ..//A/top50.txt -d 4
2025-03-23 07:47:04,171 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 07:47:04,274 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 07:47:04,274 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 07:47:04,324 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to re
edy this.
2025-03-23 07:47:04,325 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 07:47:04,698 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 07:47:04,924 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 07:47:05,027 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local280947590_0001
2025-03-23 07:47:05,027 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 07:47:05,176 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local280947590_0001_02d18a1a-afef-4f86-979a-c770726c1091/top50.txt <-
/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/C/top50.txt
2025-03-23 07:47:05,180 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/j
ob_local280947590_0001_02d18a1a-afef-4f86-979a-c770726c1091/top50.txt
2025-03-23 07:47:05,218 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 07:47:05,219 INFO mapred.LocalJobRunner: Running job: job_local280947590_0001
2025-03-23 07:47:05,219 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 07:47:05,223 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 07:47:05,224 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 07:47:05,224 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 07:47:05,225 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

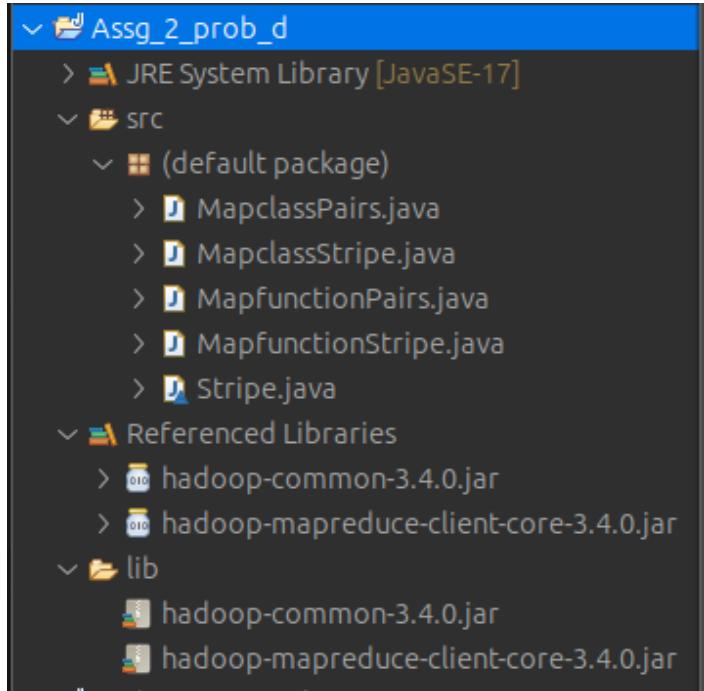
3.4.2. Execution end time and successful completion.

```
2025-03-23 07:50:32,265 INFO mapred.LocalJobRunner: Finishing task: attempt_local280947590_0001_r_000000_0
2025-03-23 07:50:32,265 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 07:50:32,487 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 07:50:33,487 INFO mapreduce.Job: Job job_local280947590_0001 completed successfully
2025-03-23 07:50:33,782 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1698108275327
    FILE: Number of bytes written=1122040217335
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=737167
    Map output bytes=149667646
    Map output materialized bytes=151939147
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=151939147
    Reduce input records=737167
    Reduce output records=49
    Spilled Records=1474334
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2943
    Total committed heap usage (bytes)=41242331185152
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=10668
```

3.4.3. Total Execution Time: 209.7 seconds.

D

1. File Structure:



2. Outputs:

2.1. Map-class with Pairs approach:

0,0	2786	0,0	6053	0,0	10457	0,0	14772
0,1	3383	0,1	6042	0,1	10598	0,1	13562
0,10	503	0,10	887	0,10	1247	0,10	1571
0,2	2430	0,2	3908	0,2	6968	0,2	8546
0,2010	156	0,2010	549	0,2010	723	0,2010	1004
0,2011	44	0,2011	117	0,2011	230	0,2011	368
0,3	1776	0,3	2710	0,3	5300	0,3	6419
0,4	1320	0,4	2136	0,4	4544	0,4	5421
0,5	1343	0,5	2108	0,5	4433	0,5	5254
0,b	56	0,b	120	0,b	192	0,b	291
0,c	186	0,c	310	0,c	397	0,c	496

d = 1

d = 2

d = 3

d = 4

2.2. Map-class with Stripe approach:

```

0 [2786, 1676, 220, 1069, 137, 23, 736, 608, 730, 16, 55, 0, 0, 15, 55, 0, 34, 6, 14, 15, 0, 101, 0, 17, 0, 58, 6, 0, 118, 61, 143, 63, 0, 123, 4, 4, 30, 2, 1, 10, 0, 0, 0, 17, 35, 60, 0, 0, 3, 10]
1 [1707, 2162, 390, 2340, 375, 250, 1169, 1017, 1045, 54, 86, 3, 2, 30, 64, 1, 69, 26, 53, 52, 1, 225, 0, 72, 0, 35, 13, 4, 134, 131, 228, 316, 0, 378, 68, 8, 54, 3, 16, 50, 1, 42, 0, 51, 57, 76, 0, 8, 57]
10 [283, 357, 388, 313, 244, 285, 244, 267, 308, 25, 45, 1, 0, 13, 30, 2, 18, 10, 0, 277, 0, 12, 0, 6, 1, 0, 12, 52, 75, 27, 0, 140, 1, 6, 26, 3, 4, 1, 0, 10, 0, 11, 2, 57, 0, 0, 1, 36]
2 [1421, 1550, 276, 1254, 268, 212, 1355, 909, 910, 101, 114, 0, 1, 71, 43, 1, 87, 20, 51, 91, 3, 252, 0, 49, 1, 33, 10, 2, 109, 128, 219, 0, 603, 59, 7, 109, 9, 17, 37, 1, 65, 2, 22, 45, 42, 0, 0, 5, 79]
2010 [19, 129, 786, 72, 168, 211, 57, 42, 37, 1, 10, 0, 1, 10, 7, 5, 8, 16, 3, 15, 1, 1450, 0, 9, 0, 9, 2, 0, 52, 8, 14, 150, 0, 296, 8, 1, 13, 0, 10, 3, 0, 83, 4, 49, 4, 7, 1, 0, 17, 14]
2011 [21, 99, 744, 80, 46, 186, 62, 30, 59, 5, 4, 0, 0, 3, 7, 5, 13, 20, 3, 7, 0, 1585, 0, 5, 0, 7, 3, 0, 26, 4, 12, 103, 0, 483, 3, 6, 9, 0, 14, 12, 1, 47, 1, 26, 5, 3, 0, 0, 18, 2]
3 [1040, 954, 277, 1003, 335, 186, 1061, 998, 906, 19, 74, 0, 0, 65, 27, 4, 27, 7, 38, 33, 0, 291, 0, 19, 0, 31, 1, 1, 49, 116, 82, 110, 0, 281, 14, 4, 66, 2, 8, 9, 0, 19, 0, 4, 15, 35, 0, 0, 7, 32]
4 [712, 761, 219, 732, 289, 282, 723, 673, 823, 16, 60, 0, 0, 38, 17, 2, 23, 12, 37, 36, 0, 171, 0, 8, 1, 13, 2, 0, 39, 97, 51, 104, 0, 225, 10, 8, 28, 1, 6, 17, 0, 24, 1, 5, 11, 36, 0, 0, 20, 37]
5 [613, 691, 306, 590, 280, 217, 565, 512, 633, 23, 80, 0, 1, 8, 24, 1, 16, 11, 22, 23, 0, 197, 0, 23, 0, 43, 3, 1, 26, 177, 66, 105, 0, 240, 0, 10, 20, 1, 9, 18, 0, 10, 0, 7, 24, 36, 0, 0, 29, 30]
b [40, 76, 16, 140, 9, 3, 43, 31, 8, 188, 547, 3, 0, 110, 88, 3, 82, 0, 59, 48, 0, 29, 0, 129, 2, 107, 8, 11, 43, 78, 111, 60, 0, 186, 189, 1, 135, 6, 4, 84, 0, 13, 0, 11, 48, 78, 0, 0, 0, 33]
c [131, 151, 19, 166, 16, 13, 78, 64, 58, 96, 366, 10, 2, 253, 173, 1, 451, 0, 81, 196, 0, 60, 0, 91, 4, 196, 2, 5, 38, 130, 116, 92, 0, 195, 162, 5, 167, 0, 3, 109, 1, 57, 1, 3, 51, 83, 0, 0, 0, 25]

```

d = 1

```

0 [6053, 2866, 368, 1783, 442, 52, 1130, 963, 1035, 40, 116, 0, 0, 47, 107, 1, 78, 17, 80, 28, 1, 232, 0, 51, 0, 99, 11, 0, 169, 158, 235, 102, 0, 299, 23, 6, 89, 5, 7, 39, 0, 20, 0, 30, 74, 151, 0, 0, 49, 27]
1 [3176, 5265, 777, 4287, 705, 605, 2129, 1695, 1695, 117, 173, 3, 2, 98, 133, 9, 141, 103, 137, 198, 1, 635, 0, 176, 1, 138, 19, 5, 246, 447, 499, 453, 0, 806, 115, 13, 256, 3, 28, 182, 1, 98, 0, 99, 110, 143, 0, 0, 127, 173]
10 [519, 677, 866, 504, 598, 574, 391, 374, 415, 45, 77, 1, 0, 22, 56, 5, 33, 25, 22, 30, 0, 756, 0, 44, 3, 19, 5, 0, 44, 176, 151, 66, 0, 293, 9, 10, 137, 5, 14, 11, 0, 54, 0, 33, 8, 131, 2, 0, 175, 108]
2 [2125, 2760, 513, 3665, 656, 554, 2564, 1669, 1437, 146, 195, 0, 3, 154, 142, 2, 137, 63, 111, 166, 4, 530, 0, 136, 1, 132, 16, 2, 288, 379, 335, 331, 0, 965, 104, 14, 282, 9, 36, 86, 1, 98, 3, 106, 89, 126, 1, 0, 128, 203]
2010 [187, 296, 985, 178, 690, 378, 150, 97, 108, 20, 23, 0, 1, 27, 17, 10, 26, 67, 18, 26, 3, 1753, 0, 26, 0, 19, 3, 0, 65, 20, 24, 165, 0, 363, 29, 6, 126, 0, 18, 8, 1, 136, 0, 68, 15, 20, 1, 0, 1020, 20]
2011 [65, 252, 1052, 184, 145, 782, 127, 90, 126, 19, 17, 2, 0, 23, 20, 9, 11, 88, 11, 11, 0, 1777, 0, 21, 1, 22, 10, 0, 33, 21, 22, 2, 573, 15, 15, 85, 10, 22, 1, 71, 8, 48, 10, 9, 1, 0, 1084, 5]
3 [1580, 1773, 443, 1845, 567, 492, 2339, 1816, 1427, 55, 132, 0, 1, 119, 66, 9, 63, 46, 82, 99, 3, 564, 0, 57, 0, 93, 6, 1, 102, 306, 166, 293, 0, 655, 44, 11, 189, 3, 22, 31, 1, 39, 1, 34, 47, 96, 1, 0, 174, 127]
4 [1173, 1308, 383, 1410, 472, 499, 1277, 1360, 1396, 37, 111, 0, 0, 65, 65, 2, 46, 30, 61, 82, 0, 414, 0, 46, 1, 69, 5, 0, 64, 250, 118, 172, 0, 513, 42, 11, 163, 2, 14, 53, 1, 38, 2, 24, 24, 94, 1, 0, 85, 108]
5 [1673, 1221, 531, 1131, 496, 546, 1013, 1204, 1287, 39, 113, 1, 2, 36, 62, 4, 56, 35, 43, 66, 0, 53, 0, 43, 3, 0, 109, 4, 1, 61, 299, 123, 157, 21, 18, 487, 15, 53, 151, 2, 153, 5, 134, 111, 168, 1, 0, 153, 116]
b [86, 138, 30, 193, 65, 63, 90, 63, 23, 462, 845, 3, 1, 365, 139, 3, 149, 16, 130, 102, 0, 75, 0, 180, 4, 150, 13, 13, 66, 195, 239, 126, 0, 461, 273, 14, 249, 6, 0, 155, 11, 43, 0, 15, 78, 123, 0, 0, 18, 65]
c [194, 273, 39, 262, 33, 26, 117, 97, 88, 226, 1411, 21, 2, 707, 358, 1, 612, 14, 156, 363, 0, 123, 0, 137, 5, 252, 12, 6, 52, 273, 178, 162, 1, 403, 318, 5, 365, 2, 10, 176, 5, 95, 1, 6, 79, 125, 0, 0, 40, 44]

```

d = 2

```

0 [6053, 2866, 368, 1783, 442, 52, 1130, 963, 1035, 40, 116, 0, 0, 47, 107, 1, 78, 17, 80, 28, 1, 232, 0, 51, 0, 99, 11, 0, 169, 158, 235, 102, 0, 299, 23, 6, 89, 5, 7, 39, 0, 20, 0, 30, 74, 151, 0, 0, 49, 27]
1 [3176, 5265, 777, 4287, 705, 605, 2129, 1695, 1695, 117, 173, 3, 2, 98, 133, 9, 141, 103, 137, 198, 1, 635, 0, 176, 1, 138, 19, 5, 246, 447, 499, 453, 0, 806, 115, 13, 256, 3, 28, 182, 1, 98, 0, 99, 110, 143, 0, 0, 127, 173]
10 [519, 677, 866, 504, 598, 574, 391, 374, 415, 45, 77, 1, 0, 22, 56, 5, 33, 25, 22, 30, 0, 756, 0, 44, 3, 19, 5, 0, 44, 176, 151, 66, 0, 293, 9, 10, 137, 5, 14, 11, 0, 54, 0, 33, 8, 131, 2, 0, 175, 108]
2 [2125, 2760, 513, 3665, 656, 554, 2564, 1669, 1437, 146, 195, 0, 3, 154, 142, 2, 137, 63, 111, 166, 4, 530, 0, 136, 1, 132, 16, 2, 288, 379, 335, 331, 0, 965, 104, 14, 282, 9, 36, 86, 1, 98, 3, 106, 89, 126, 1, 0, 128, 203]
2010 [187, 296, 985, 178, 690, 378, 150, 97, 108, 20, 23, 0, 1, 27, 17, 10, 26, 67, 18, 26, 3, 1753, 0, 26, 0, 19, 3, 0, 65, 20, 24, 165, 0, 363, 29, 6, 126, 0, 18, 8, 1, 136, 0, 68, 15, 20, 1, 0, 1020, 20]
2011 [65, 252, 1052, 184, 145, 782, 127, 90, 126, 19, 17, 2, 0, 23, 20, 9, 11, 88, 11, 11, 0, 1777, 0, 21, 1, 22, 10, 0, 33, 21, 22, 2, 573, 15, 15, 85, 10, 22, 1, 71, 8, 48, 10, 9, 1, 0, 1084, 5]
3 [1580, 1773, 443, 1845, 567, 492, 2339, 1816, 1427, 55, 132, 0, 1, 119, 66, 9, 63, 46, 82, 99, 3, 564, 0, 57, 0, 93, 6, 1, 102, 306, 166, 293, 0, 655, 44, 11, 189, 3, 22, 31, 1, 39, 1, 34, 47, 96, 1, 0, 174, 127]
4 [1173, 1308, 383, 1410, 472, 499, 1277, 1360, 1396, 37, 111, 0, 0, 65, 65, 2, 46, 30, 61, 82, 0, 414, 0, 46, 1, 69, 5, 0, 64, 250, 118, 172, 0, 513, 42, 11, 163, 2, 14, 53, 1, 38, 2, 24, 24, 94, 1, 0, 85, 108]
5 [1673, 1221, 531, 1131, 496, 546, 1013, 1204, 1287, 39, 113, 1, 2, 36, 62, 4, 56, 35, 43, 66, 0, 53, 0, 43, 3, 0, 109, 4, 1, 61, 299, 123, 157, 21, 18, 487, 15, 53, 151, 2, 153, 5, 134, 111, 168, 1, 0, 153, 116]
b [86, 138, 30, 193, 65, 63, 90, 63, 23, 462, 845, 3, 1, 365, 139, 3, 149, 16, 130, 102, 0, 75, 0, 180, 4, 150, 13, 13, 66, 195, 239, 126, 0, 461, 273, 14, 249, 6, 0, 155, 11, 43, 0, 15, 78, 123, 0, 0, 18, 65]
c [194, 273, 39, 262, 33, 26, 117, 97, 88, 226, 1411, 21, 2, 707, 358, 1, 612, 14, 156, 363, 0, 123, 0, 137, 5, 252, 12, 6, 52, 273, 178, 162, 1, 403, 318, 5, 365, 2, 10, 176, 5, 95, 1, 6, 79, 125, 0, 0, 40, 44]

```

d = 3

```

0 [10457, 5591, 521, 3981, 536, 88, 2995, 2806, 2890, 73, 156, 0, 2, 73, 146, 2, 134, 38, 106, 78, 1, 306, 0, 79, 0, 122, 16, 2, 208, 228, 361, 168, 0, 431, 52, 13, 218, 7, 14, 62, 0, 31, 0, 33, 95, 194, 0, 0, 115, 53]
1 [5807, 8171, 1231, 6152, 863, 719, 3345, 2715, 2652, 185, 270, 3, 3, 182, 227, 13, 228, 186, 204, 166, 2, 1829, 0, 323, 1, 178, 38, 6, 303, 606, 678, 583, 0, 1167, 188, 26, 524, 4, 39, 179, 1, 138, 1, 153, 154, 188, 0, 0, 370, 205]
10 [726, 934, 1441, 739, 737, 639, 617, 564, 575, 65, 109, 1, 1, 38, 85, 9, 57, 60, 44, 59, 0, 1054, 0, 66, 4, 30, 6, 0, 66, 255, 195, 100, 1, 434, 25, 14, 278, 8, 23, 31, 0, 88, 1, 48, 10, 153, 3, 0, 459, 117]
2 [2987, 4052, 944, 5539, 802, 666, 3610, 2231, 1972, 216, 298, 1, 6, 234, 215, 5, 208, 113, 176, 247, 5, 71, 5, 265, 519, 498, 428, 0, 1354, 186, 18, 487, 15, 53, 151, 2, 153, 5, 134, 111, 168, 1, 0, 127, 235]
2010 [187, 487, 1095, 294, 1850, 976, 239, 172, 191, 37, 44, 0, 1, 53, 29, 16, 42, 119, 26, 38, 3835, 0, 51, 2, 16, 4, 1, 73, 45, 57, 190, 1, 549, 41, 12, 379, 2, 33, 13, 2, 166, 10, 119, 23, 32, 1, 0, 1248, 24]
2011 [142, 391, 1151, 278, 320, 1988, 199, 145, 179, 35, 38, 2, 0, 51, 30, 12, 22, 142, 23, 24, 0, 2567, 0, 40, 1, 37, 31, 0, 46, 52, 44, 135, 2, 752, 36, 23, 318, 1, 35, 32, 1, 91, 10, 86, 17, 15, 2, 0, 1221, 10]
3 [2305, 2699, 768, 2630, 680, 545, 3494, 2591, 1894, 94, 299, 3, 185, 99, 14, 97, 87, 118, 171, 4, 845, 0, 92, 0, 129, 11, 3, 138, 392, 221, 296, 1, 945, 92, 23, 326, 6, 35, 76, 1, 73, 2, 52, 82, 172, 1, 0, 341, 144]
4 [1738, 2018, 651, 1937, 552, 552, 1810, 2065, 1946, 56, 163, 0, 0, 100, 102, 3, 73, 62, 88, 116, 0, 681, 0, 73, 1, 113, 8, 1, 87, 324, 178, 235, 0, 731, 83, 19, 298, 3, 21, 73, 1, 66, 5, 54, 32, 134, 1, 0, 213, 106]
5 [1543, 1781, 766, 1644, 570, 596, 1403, 1566, 2016, 65, 161, 1, 3, 59, 90, 7, 90, 62, 101, 1, 73, 0, 68, 0, 115, 1, 57, 49, 20, 598, 0, 666, 49, 20, 598, 4, 27, 49, 1, 424, 2, 58, 63, 107, 2, 0, 354, 122]
b [119, 245, 85, 304, 81, 72, 166, 124, 69, 783, 1028, 3, 1, 511, 255, 6, 193, 28, 188, 162, 1, 97, 0, 231, 5, 198, 30, 15, 103, 266, 297, 205, 0, 616, 335, 16, 424, 9, 11, 211, 13, 71, 0, 21, 96, 147, 0, 0, 47, 84]
c [241, 341, 65, 347, 62, 66, 156, 139, 135, 327, 1890, 26, 6, 861, 449, 4, 784, 27, 206, 438, 0, 191, 0, 186, 6, 308, 15, 9, 88, 381, 263, 235, 1, 548, 392, 6, 598, 2, 16, 243, 6, 116, 4, 12, 110, 173, 0, 0, 83, 78]

```

d = 4

```

0 [14772, 7227, 694, 4772, 737, 157, 3557, 3243, 3267, 185, 205, 1, 2, 103, 183, 7, 168, 92, 149, 137, 1, 944, 0, 128, 0, 158, 21, 2, 249, 335, 484, 606, 1, 775, 74, 23, 322, 9, 25, 93, 2, 48, 0, 50, 109, 226, 0, 0, 161, 79]
1 [6335, 11186, 1666, 7563, 1143, 863, 4189, 3322, 3165, 259, 374, 4, 6, 259, 274, 17, 279, 293, 259, 218, 4, 1452, 0, 424, 1, 236, 36, 7, 409, 752, 914, 767, 1, 1566, 262, 40, 822, 4, 60, 230, 2, 229, 3, 197, 205, 236, 1, 0, 624, 248]
10 [877, 1237, 1963, 989, 890, 736, 805, 757, 743, 87, 148, 1, 2, 69, 136, 11, 82, 93, 68, 86, 1, 1293, 0, 86, 5, 41, 8, 0, 78, 325, 227, 138, 1, 579, 48, 20, 458, 8, 33, 55, 0, 114, 6, 71, 17, 183, 4, 0, 648, 143]
2 [3774, 5480, 1271, 7089, 997, 771, 4567, 2834, 2480, 271, 381, 1, 7, 296, 280, 12, 264, 178, 221, 340, 6, 1139, 0, 297, 3, 224, 36, 6, 346, 649, 635, 526, 0, 1746, 258, 24, 754, 21, 70, 202, 4, 197, 8, 179, 160, 212, 1, 0, 533, 279]
2010 [267, 647, 1228, 377, 2468, 1238, 315, 215, 308, 52, 69, 0, 1, 76, 44, 17, 59, 207, 31, 48, 3, 3232, 0, 63, 3, 45, 5, 3, 83, 69, 88, 211, 1, 723, 52, 19, 629, 13, 43, 26, 2, 196, 10, 142, 31, 43, 2, 0, 1998, 60]
2011 [211, 502, 1243, 365, 452, 2610, 261, 210, 250, 45, 45, 2, 0, 73, 43, 14, 42, 225, 34, 32, 4, 2903, 0, 61, 1, 46, 19, 0, 53, 68, 73, 151, 2, 927, 53, 30, 584, 4, 45, 47, 3, 112, 14, 103, 27, 27, 3, 0, 1786, 85]
3 [2862, 3450, 1034, 3493, 845, 641, 4741, 3304, 2403, 146, 265, 5, 1, 238, 141, 16, 134, 142, 157, 207, 4, 1048, 0, 134, 0, 173, 15, 6, 193, 481, 306, 376, 1, 1238, 132, 29, 518, 12, 42, 97, 5, 126, 4, 83, 92, 211, 1, 0, 503, 166]
4 [2178, 2631, 901, 2531, 699, 625, 2391, 2751, 2453, 96, 201, 3, 0, 135, 136, 5, 91, 120, 115, 155, 1, 023, 0, 99, 2, 150, 16, 1, 120, 395, 284, 267, 1, 979, 112, 25, 420, 4, 33, 86, 1, 97, 9, 82, 40, 183, 1, 0, 388, 124]
5 [1987, 2280, 992, 2067, 690, 656, 1927, 1962, 2617, 95, 193, 2, 5, 92, 118, 11, 107, 105, 78, 129, 1, 853, 0, 102, 1, 135, 10, 1,
```

2.3. Map-function with Pairs approach:

0,0	2786	0,0	6053	0,0	10457	0,0	14772
0,1	3383	0,1	6042	0,1	10598	0,1	13562
0,10	503	0,10	887	0,10	1247	0,10	1571
0,2	2430	0,2	3908	0,2	6968	0,2	8546
0,2010	156	0,2010	549	0,2010	723	0,2010	1004
0,2011	44	0,2011	117	0,2011	230	0,2011	368
0,3	1776	0,3	2710	0,3	5300	0,3	6419
0,4	1320	0,4	2136	0,4	4544	0,4	5421

d = 1

d = 2

d = 3

d = 4

2.4. Map-function with Stripe approach:

0	[2786, 1676, 220, 1009, 137, 23, 736, 608, 730, 16, 55, 0, 0, 15, 55, 0, 34, 6, 14, 15, 0, 101, 0, 17, 0, 58, 6, 0, 118, 61, 143, 63, 0, 123, 4, 4, 30, 2, 1, 10, 0, 0, 0, 17, 35, 60, 0, 0, 3, 10]
1	[1707, 2162, 390, 2340, 375, 250, 1169, 1017, 1045, 54, 86, 3, 2, 30, 64, 1, 69, 26, 53, 52, 1, 225, 0, 72, 0, 35, 13, 4, 134, 131, 228, 316, 0, 378, 60, 8, 54, 3, 16, 50, 1, 42, 0, 51, 57, 76, 0, 0, 8, 57]
10	[283, 357, 308, 313, 244, 205, 244, 267, 308, 25, 45, 1, 0, 13, 30, 2, 18, 10, 12, 10, 0, 277, 0, 12, 0, 6, 1, 0, 12, 52, 75, 27, 0, 140, 1, 6, 28, 3, 4, 1, 0, 10, 0, 11, 2, 57, 0, 0, 1, 36]
2	[1421, 1556, 276, 1254, 268, 212, 1355, 909, 910, 101, 114, 0, 1, 71, 43, 1, 87, 28, 51, 91, 3, 252, 0, 49, 1, 33, 10, 2, 109, 128, 170, 219, 0, 603, 59, 7, 109, 9, 17, 37, 1, 65, 2, 22, 45, 42, 0, 0, 5, 79]
2010	[19, 129, 706, 72, 168, 211, 57, 42, 37, 1, 10, 0, 1, 10, 7, 5, 8, 16, 3, 15, 1, 1456, 0, 9, 0, 52, 8, 14, 150, 0, 296, 8, 1, 13, 0, 10, 3, 0, 83, 4, 49, 4, 7, 1, 0, 17, 14]

d = 1

0	[6053, 2866, 368, 1783, 442, 52, 1130, 963, 1035, 40, 116, 0, 0, 47, 107, 1, 78, 17, 80, 28, 1, 232, 0, 51, 0, 99, 11, 0, 169, 158, 235, 182, 0, 299, 23, 6, 89, 5, 7, 39, 0, 20, 0, 30, 74, 151, 0, 0, 49, 27]
1	[3176, 5265, 777, 4287, 705, 605, 2129, 1695, 1669, 117, 173, 3, 2, 98, 133, 9, 141, 103, 137, 108, 1, 635, 0, 176, 1, 130, 19, 5, 246, 447, 499, 453, 0, 806, 115, 13, 256, 3, 28, 102, 1, 98, 0, 99, 110, 143, 0, 0, 127, 173]
10	[519, 677, 866, 504, 598, 574, 391, 374, 415, 45, 77, 1, 0, 22, 56, 5, 33, 25, 22, 38, 0, 0, 756, 0, 44, 3, 19, 5, 0, 44, 178, 151, 66, 0, 293, 9, 10, 137, 5, 14, 11, 0, 54, 0, 33, 8, 131, 2, 0, 175, 108]
2	[2125, 2768, 513, 3665, 656, 554, 2564, 1669, 1437, 146, 195, 0, 3, 154, 142, 2, 137, 63, 111, 166, 4, 530, 0, 136, 1, 132, 16, 2, 288, 379, 335, 331, 0, 965, 104, 14, 282, 9, 36, 86, 1, 98, 3, 106, 89, 126, 1, 0, 128, 203]
2010	[107, 296, 985, 178, 696, 378, 150, 97, 108, 20, 23, 0, 1, 27, 17, 10, 20, 0, 67, 18, 26, 3, 1753, 0, 26, 0, 19, 3, 0, 65, 20, 34, 165, 0, 363, 29, 6, 126, 0, 18, 8, 1, 136, 0, 68, 15, 20, 1, 0, 1020, 20]

d = 2

0	[10457, 5591, 521, 3981, 536, 88, 2995, 2806, 2890, 73, 156, 0, 2, 73, 146, 2, 134, 38, 106, 78, 1, 306, 0, 73, 0, 122, 16, 2, 208, 228, 361, 160, 0, 431, 52, 13, 218, 7, 14, 62, 0, 31, 0, 33, 95, 194, 0, 0, 115, 53]
1	[5007, 8171, 1231, 6152, 863, 719, 3345, 2715, 2652, 185, 270, 3, 3, 182, 227, 13, 220, 186, 204, 166, 2, 1829, 0, 323, 1, 178, 30, 6, 303, 606, 678, 583, 0, 1167, 188, 26, 524, 4, 39, 179, 1, 138, 1, 153, 154, 188, 0, 0, 370, 205]
10	[726, 934, 1441, 739, 737, 639, 617, 564, 575, 65, 109, 1, 1, 38, 85, 9, 57, 68, 44, 59, 0, 1054, 0, 66, 4, 30, 6, 0, 66, 255, 195, 100, 1, 434, 25, 14, 278, 8, 23, 31, 0, 80, 1, 48, 10, 153, 3, 0, 459, 117]
2	[2987, 4052, 944, 5539, 802, 666, 3610, 2231, 1972, 216, 298, 1, 6, 234, 215, 5, 208, 113, 176, 247, 5, 912, 0, 228, 1, 184, 27, 5, 265, 519, 498, 428, 0, 1354, 186, 18, 487, 15, 53, 151, 2, 153, 5, 134, 111, 168, 1, 0, 287, 235]
2010	[187, 487, 1085, 294, 1850, 976, 239, 172, 191, 37, 44, 0, 1, 53, 29, 16, 42, 119, 26, 38, 3, 2835, 0, 51, 1, 26, 4, 1, 73, 45, 57, 190, 1, 549, 41, 12, 379, 2, 33, 13, 2, 166, 10, 119, 23, 32, 1, 0, 1248, 24]

d = 3

0	[14772, 7227, 694, 4772, 737, 157, 3557, 3243, 3267, 105, 205, 1, 2, 103, 183, 7, 168, 92, 149, 137, 1, 944, 0, 120, 0, 158, 21, 2, 249, 335, 404, 606, 1, 775, 74, 23, 322, 9, 25, 93, 2, 40, 0, 50, 109, 226, 0, 0, 161, 79]
1	[6335, 11186, 1666, 7563, 1143, 863, 4189, 3322, 3165, 259, 374, 4, 6, 259, 274, 17, 279, 293, 259, 218, 4, 1452, 0, 424, 1, 236, 36, 7, 409, 752, 914, 767, 1, 1566, 262, 40, 822, 4, 60, 230, 2, 229, 3, 197, 205, 236, 1, 0, 624, 248]
10	[877, 1237, 1963, 989, 890, 736, 805, 757, 745, 87, 146, 1, 2, 69, 136, 11, 82, 93, 68, 86, 1, 1293, 0, 86, 5, 41, 8, 0, 78, 325, 227, 130, 1, 579, 48, 20, 450, 8, 33, 55, 0, 114, 6, 71, 183, 4, 0, 648, 143]
2	[3774, 5460, 1271, 7089, 997, 771, 4567, 2834, 2480, 271, 381, 1, 7, 296, 286, 12, 264, 221, 340, 6, 1139, 0, 297, 3, 224, 36, 6, 346, 649, 635, 526, 0, 1746, 258, 24, 754, 21, 78, 202, 4, 197, 8, 179, 166, 212, 1, 0, 533, 279]
2010	[267, 647, 1228, 377, 2460, 1238, 315, 215, 308, 52, 69, 0, 1, 76, 44, 17, 59, 207, 31, 48, 3, 3232, 0, 63, 3, 45, 5, 3, 83, 69, 88, 211, 1, 723, 52, 19, 629, 13, 43, 26, 2, 196, 10, 142, 31, 43, 2, 0, 1998, 60]

d = 4

3. Execution Time:

3.1. Map-class with Pairs approach:

3.1.1. d = 1

3.1.1.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapclassPairs ..../Wikipedia-EN-20120601_ARTICLES ./output/map_class_pairs/d1 ..../A/top50.txt
2025-03-23 08:14:29,711 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:14:29,807 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:14:29,807 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:14:29,857 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:14:29,859 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:14:30,260 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:14:30,554 INFO mapreduce.JobSubmitter: number of splits:10000
```

□

3.1.1.2. Execution end time and successful completion.

```
2025-03-23 08:17:11,946 INFO mapred.LocalJobRunner: Finishing task: attempt_local211927060_0001_r_000000_0
2025-03-23 08:17:11,946 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:17:12,112 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:17:13,112 INFO mapreduce.Job: Job job_local211927060_0001 completed successfully
2025-03-23 08:17:13,410 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697806600817
    FILE: Number of bytes written=34780833628
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=91288
    Map output bytes=859316
    Map output materialized bytes=1101892
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=955
    Reduce shuffle bytes=1101892
    Reduce input records=91288
    Reduce output records=955
    Spilled Records=182576
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ns)=2255
    Total committed heap usage (bytes)=41221613420544
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=8126
```

3.1.1.3. Total Execution Time: **163.7 seconds.**

3.1.2. d = 2

3.1.2.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapClassPairs .../Wikipedia-EN-20120601_ARTICLES ./output/map_class_pairs/d2 ../A/top50.txt
xt -d 2
2025-03-23 08:21:02,002 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:21:02,093 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:21:02,093 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:21:02,137 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:21:02,138 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:21:02,481 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:21:02,679 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:21:02,779 INFO mapreduce.JobSubmitter: Submitting token for job: job_local1886708848_0001
2025-03-23 08:21:02,779 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:21:02,914 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local1886708848_0001_b17e79cf-0d7f-44cf-b720-06ec4bbc47af/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 08:21:02,916 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local1886708848_0001_b17e79cf-0d7f-44cf-b720-06ec4bbc47af/top50.txt
2025-03-23 08:21:02,954 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 08:21:02,954 INFO mapreduce.Job: Running job: job_local1886708848_0001
2025-03-23 08:21:02,960 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:21:02,962 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 08:21:02,962 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:21:02,962 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.1.2.2. Execution end time and successful completion.

```
2025-03-23 08:23:34,363 INFO mapred.LocalJobRunner: Finishing task: attempt_local1886708848_0001_r_000000_0
2025-03-23 08:23:34,363 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:23:35,083 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:23:36,083 INFO mapreduce.Job: Job job_local1886708848_0001 completed successfully
2025-03-23 08:23:36,314 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697808122253
    FILE: Number of bytes written=39899343362
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map Input records=10000
    Map output records=156478
    Map output bytes=1489654
    Map output materialized bytes=1862610
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1041
    Reduce shuffle bytes=1862610
    Reduce input records=156478
    Reduce output records=1041
    Spilled Records=312956
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2326
    Total committed heap usage (bytes)=41218283143168
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9127
```

3.1.2.3. Total Execution Time: 154.3 seconds.

3.1.3. d = 3

3.1.3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/D$ hadoop jar prob_4_d.jar MapclassPairs .../Wikipedia-EN-20120601_ARTICLES ./output/map_class_pairs/d3 ../A/top50.txt
XT -d 3
2025-03-23 08:26:35,944 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:26:36,063 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:26:36,063 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:26:36,110 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:26:36,111 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:26:36,499 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:26:36,761 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:26:36,901 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1332942001_0001
2025-03-23 08:26:36,905 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:26:37,091 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local1332942001_0001_b6a0ccc7-48d4-4bf3-8ddd-ee01552fbe40/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/D/top50.txt
2025-03-23 08:26:37,094 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/D/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local1332942001_0001_b6a0ccc7-48d4-4bf3-8ddd-ee01552fbe40/top50.txt
2025-03-23 08:26:37,165 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 08:26:37,166 INFO mapreduce.Job: Running job: job_local1332942001_0001
2025-03-23 08:26:37,166 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 08:26:37,174 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:26:37,175 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 08:26:37,175 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:26:37,175 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.1.3.2. Execution end time and successful execution.

```
2025-03-23 08:29:21,113 INFO mapred.LocalJobRunner: Finishing task: attempt_local1332942001_r_000000_0
2025-03-23 08:29:21,113 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:29:21,398 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:29:22,398 INFO mapreduce.Job: Job job_local1332942001_0001 completed successfully
2025-03-23 08:29:22,656 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697889496565
    FILE: Number of bytes written=44509188078
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=217176
    Map output bytes=2055414
    Map output materialized bytes=2549766
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1076
    Reduce shuffle bytes=2549766
    Reduce input records=217176
    Reduce output records=1076
    Spilled Records=434352
    Shuffled Maps =10800
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2207
    Total committed heap usage (bytes)=41193694035968
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9619
```

3.1.3.3. Total Execution Time: 166.7 seconds.

3.1.4. d = 4

3.1.4.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapclassPairs ../Wikipedia-EN-20120601_ARTICLES ./output/map_class_pairs/d4 ./A/top50.txt
xt -d 4
2025-03-23 08:30:25,530 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:30:25,638 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:30:25,638 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:30:25,680 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:30:25,681 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:30:26,048 INFO InputFormat: Total input files to process : 10000
2025-03-23 08:30:26,272 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:30:26,378 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local452293174_0001
2025-03-23 08:30:26,378 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:30:26,529 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local452293174_0001_361c15f7-22cd-4cf8-875d-ce473a178bd2/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 08:30:26,532 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local452293174_0001_361c15f7-22cd-4cf8-875d-ce473a178bd2/top50.txt
2025-03-23 08:30:26,576 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 08:30:26,576 INFO mapreduce.Job: Running job: job_local452293174_0001
2025-03-23 08:30:26,577 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 08:30:26,583 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:30:26,584 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 08:30:26,584 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:30:26,585 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.1.4.2. Execution end time and successful completion.

```
2025-03-23 08:33:02,012 INFO mapred.LocalJobRunner: Finishing task: attempt_local452293174_0001_r_000000_0
2025-03-23 08:33:02,012 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:33:02,071 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:33:03,072 INFO mapreduce.Job: Job job_local452293174_0001 completed successfully
2025-03-23 08:33:03,296 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697810569939
    FILE: Number of bytes written=47976756423
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=264228
    Map output bytes=2497997
    Map output materialized bytes=3086453
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1099
    Reduce shuffle bytes=3086453
    Reduce input records=264228
    Reduce output records=1099
    Spilled Records=528456
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2698
    Total committed heap usage (bytes)=41228949258240
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9950
```

3.1.4.3. Total Execution Time: 157.5 seconds.

3.2. Map-class with Stripe approach:

3.2.1. d = 1

3.2.1.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapclassStripe ../Wikipedia-EN-20120601_ARTICLES ./output/map_class_stripe/d1 ../A/top50.txt -d 1
2025-03-23 08:35:11,062 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:35:11,136 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:35:11,136 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:35:11,173 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:35:11,174 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:35:11,497 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:35:11,694 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:35:11,777 INFO mapreduce.JobSubmitter: Submitting token for job: job_local346930622_0001
2025-03-23 08:35:11,777 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:35:11,899 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local346930622_0001_339a923c-4feb-4c59-a4c5-b697c01d8db1/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 08:35:11,904 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local346930622_0001_339a923c-4feb-4c59-a4c5-b697c01d8db1/top50.txt
2025-03-23 08:35:11,939 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 08:35:11,940 INFO mapreduce.Job: Running job: job_local346930622_0001
2025-03-23 08:35:11,940 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 08:35:11,942 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:35:11,943 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 08:35:11,943 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:35:11,943 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.2.1.2. Execution end time and successful completion.

```
2025-03-23 08:37:47,418 INFO mapred.LocalJobRunner: Finishing task: attempt_local346930622_0001_r_000000_0
2025-03-23 08:37:47,418 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:37:48,127 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:37:49,128 INFO mapreduce.Job: Job job_local346930622_0001 completed successfully
2025-03-23 08:37:49,341 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697831865223
    FILE: Number of bytes written=114015025867
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=66475
    Map output bytes=13474670
    Map output materialized bytes=13734095
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=13734095
    Reduce input records=66475
    Reduce output records=49
    Spilled Records=132950
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ns)=2179
    Total committed heap usage (bytes)=41237446918144
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=8963
```

3.2.1.3. Total Execution Time: 158.3 seconds.

3.2.2. d = 2

3.2.2.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/D$ hadoop jar prob_4_d.jar MapclassStripe ../Wikimedia-EN-20120601_ARTICLES ./output/map_class_stripe/d2 ./A/top50
.txt -d 2
2025-03-23 08:38:47,882 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:38:47,969 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:38:47,969 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:38:48,014 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to ready this.
2025-03-23 08:38:48,015 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:38:48,351 INFO InputFormat: Total input files to process : 10000
2025-03-23 08:38:48,558 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:38:48,652 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1600129603_0001
2025-03-23 08:38:48,652 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:38:48,841 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local1600129603_0001_d9a771bd-ab60-43c4-b703-02958d537f7f/top50.txt <-- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/D/top50.txt
```

□

3.2.2.2. Execution end time and successful completion.

```
2025-03-23 08:41:21,313 INFO mapred.LocalJobRunner: Finishing task: attempt_local1600129603_0001_r_000000_0
2025-03-23 08:41:21,313 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:41:22,039 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:41:23,040 INFO mapreduce.Job: Job job_local1600129603_0001 completed successfully
2025-03-23 08:41:23,292 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697842473707
    FILE: Number of bytes written=145288286879
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=92257
    Map output bytes=18701566
    Map output materialized bytes=19038337
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=19038337
    Reduce input records=92257
    Reduce output records=49
    Spilled Records=184514
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2217
    Total committed heap usage (bytes)=41228187992064
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9474
```

3.2.2.3. Total Execution Time: 155.5 seconds.

3.2.3. d = 3

3.2.3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapclassStripe ..//Wikipedia-EN-20120601_ARTICLES ./output/map_class_stripe/d3 ..//A/top50
.txt -d 3
2025-03-23 08:42:14,475 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:42:14,569 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:42:14,569 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:42:14,612 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:42:14,613 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:42:15,009 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:42:15,211 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:42:15,314 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1321211459_0001
2025-03-23 08:42:15,314 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:42:15,465 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local1321211459_0001_65237bd3-6d71-443c-a3a1-e13ebd14eab7/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 08:42:15,469 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local1321211459_0001_65237bd3-6d71-443c-a3a1-e13ebd14eab7/top50.txt
2025-03-23 08:42:15,509 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 08:42:15,510 INFO mapreduce.Job: Running job: job_local1321211459_0001
2025-03-23 08:42:15,510 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 08:42:15,516 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:42:15,517 INFO output.FileOutputCommitter: File Output Committer Algoithm version is 2
2025-03-23 08:42:15,517 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:42:15,518 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.2.3.2. Execution end time and successful completion.

```
2025-03-23 08:45:15,272 INFO mapred.LocalJobRunner: Finishing task: attempt_local1321211459_0001_r_000000_0
2025-03-23 08:45:15,272 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:45:15,710 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:45:16,711 INFO mapreduce.Job: Job job_local1321211459_0001 completed successfully
2025-03-23 08:45:16,980 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697850662883
    FILE: Number of bytes written=168997591815
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map -Reduce Framework
    Map input records=10000
    Map output records=112163
    Map output bytes=22736436
    Map output materialized bytes=23132925
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=23132925
    Reduce input records=112163
    Reduce output records=49
    Spilled Records=224326
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2559
    Total committed heap usage (bytes)=41241949503488
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9810
```

3.2.3.3. Total Execution Time: 182.6 seconds.

3.2.4. d = 4

3.2.4.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapclassStripe ../Wikipedia-EN-20120601_ARTICLES ./output/map_class_stripe/d4 ./A/top50.txt -d 4
2025-03-23 08:47:38,442 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:47:38,531 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:47:38,532 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:47:38,576 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:47:38,577 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:47:38,937 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:47:39,134 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:47:39,247 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local394643651_0001
2025-03-23 08:47:39,247 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:47:39,393 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local394643651_0001_3e77fefb-53a2-492e-b616-cadccfd76aa8/top50.txt ->/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt
2025-03-23 08:47:39,393 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local394643651_0001_3e77fefb-53a2-492e-b616-cadccfd76aa8/top50.txt
2025-03-23 08:47:39,451 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 08:47:39,451 INFO mapreduce.LocalJobRunner: OutputCommitter set in config null
2025-03-23 08:47:39,451 INFO mapreduce.Job: Running job: job_local394643651_0001
2025-03-23 08:47:39,458 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:47:39,458 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 08:47:39,458 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:47:39,459 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.2.4.2. Execution end time and successful completion.

```
2025-03-23 08:50:25,719 INFO mapred.LocalJobRunner: Finishing task: attempt_local394643651_0001_r_000000_0
2025-03-23 08:50:25,719 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:50:26,632 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:50:27,633 INFO mapreduce.Job: Job job_local394643651_0001 completed successfully
2025-03-23 08:50:27,899 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697855569627
    FILE: Number of bytes written=182128130963
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=124087
    Map output bytes=25151036
    Map output materialized bytes=25586297
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=25586297
    Reduce input records=124087
    Reduce output records=49
    Spilled Records=248174
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2151
    Total committed heap usage (bytes)=41226466230272
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=10068
```

3.2.4.3. Total Execution Time: 169.5 seconds.

3.3. Map-function with Pairs approach:

3.3.1. d = 1

3.3.1.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/BS hadoop jar prob_4_d.jar MapfunctionPairs ../Wikipedia-EN-20120601_ARTICLES ./output/map_function_pairs/d1 ..//A/t
op50.txt -d 1
2025-03-23 08:53:56,534 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:53:56,647 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:53:56,647 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:53:56,699 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:53:56,701 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:53:57,111 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:53:57,401 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:53:57,515 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local699758527_0001
2025-03-23 08:53:57,515 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 08:53:57,708 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local699758527_0001_de7de818-5dbd-4544-b9bf-20913227cda3/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/BS/top50.txt
2025-03-23 08:53:57,711 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/BS/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local699758527_0001_de7de818-5dbd-4544-b9bf-20913227cda3/top50.txt
2025-03-23 08:53:57,758 INFO mapreduce.Job: The url to track the job: http://localhost:8008/
2025-03-23 08:53:57,758 INFO mapreduce.Job: Running job: job_local699758527_0001
2025-03-23 08:53:57,759 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 08:53:57,764 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 08:53:57,766 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 08:53:57,766 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 08:53:57,767 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.3.1.2. Execution end time and successful completion.

```
2025-03-23 08:57:13,742 INFO mapred.LocalJobRunner: Finishing task: attempt_local699758527_0001_r_000000_0
2025-03-23 08:57:13,742 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 08:57:14,030 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 08:57:15,031 INFO mapreduce.Job: Job job_local699758527_0001 completed successfully
2025-03-23 08:57:15,360 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=16978066000817
    FILE: Number of bytes written=34781013646
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=91288
    Map output bytes=859316
    Map output materialized bytes=1101892
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=955
    Reduce shuffle bytes=1101892
    Reduce input records=91288
    Reduce output records=955
    Spilled Records=182576
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2554
    Total committed heap usage (bytes)=41241557336064
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=8126
```

3.3.1.3. Total Execution Time: 198.9 seconds.

3.3.2. d = 2

3.3.2.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapfunctionPairs ..../Wikipedia-EN-20120601_ARTICLES ./output/map_function_pairs/d2 ..//A/t
op50.txt -d 2
2025-03-23 08:59:07,306 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 08:59:07,413 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 08:59:07,413 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 08:59:07,465 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 08:59:07,466 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 08:59:07,795 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 08:59:08,005 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 08:59:08,127 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local359573959_0001
```

3.3.2.2. Execution end time and successful completion.

```
2025-03-23 09:02:04,729 INFO mapred.LocalJobRunner: Finishing task: attempt_local359573959_0001_r_000000_0
2025-03-23 09:02:04,729 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:02:05,574 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:02:06,574 INFO mapreduce.Job: Job job_local359573959_0001 completed successfully
2025-03-23 09:02:06,841 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697808122253
    FILE: Number of bytes written=39865149943
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=156478
    Map output bytes=1489654
    Map output materialized bytes=1862610
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1041
    Reduce shuffle bytes=1862610
    Reduce input records=156478
    Reduce output records=1041
    Spilled Records=312956
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2251
    Total committed heap usage (bytes)=41236387856384
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9127
```

3.3.2.3. Total Execution Time: 179.5 seconds.

3.3.3. d = 3

3.3.3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapfunctionPairs ..../Wikipedia-EN-20120601_ARTICLES ./output/map_function_pairs/d3 ..//A/t
os50.txt -d 3
2025-03-23 09:02:55,356 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 09:02:55,450 INFO input.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 09:02:55,450 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 09:02:55,496 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 09:02:55,497 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 09:02:55,844 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 09:02:56,041 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 09:02:56,140 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1659604268_0001
2025-03-23 09:02:56,141 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 09:02:56,260 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local1659604268_0001_4424da67-4b18-484a-8814-ededad3f0f0/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 09:02:56,264 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local1659604268_0001_4424da67-4b18-484a-8814-ededad3f0f0/top50.txt
2025-03-23 09:02:56,301 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 09:02:56,302 INFO mapreduce.Job: Running job: job_local1659604268_0001
2025-03-23 09:02:56,302 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 09:02:56,307 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 09:02:56,308 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 09:02:56,308 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 09:02:56,309 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.3.3.2. Execution end time and successful time.

```
2025-03-23 09:05:46,363 INFO mapred.LocalJobRunner: Finishing task: attempt_local1659604268_0001_r_000000_0
2025-03-23 09:05:46,363 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:05:46,439 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:05:47,440 INFO mapreduce.Job: Job job_local1659604268_0001 completed successfully
2025-03-23 09:05:47,666 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697889496565
    FILE: Number of bytes written=44509368096
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=217176
    Map output bytes=2055414
    Map output materialized bytes=2549766
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1076
    Reduce shuffle bytes=2549766
    Reduce input records=217176
    Reduce output records=1076
    Spilled Records=434352
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2222
    Total committed heap usage (bytes)=41234385076224
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9619
```

3.3.3.3. Total Execution Time: 172.4 seconds.

3.3.4. d = 4

3.3.4.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapfunctionPairs ../Wikipedia-EN-20120601_ARTICLES ./output/map_function_pairs/d4 ../A/t
op50.txt -d 4
2025-03-23 09:10:56,872 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 09:10:56,978 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 09:10:56,978 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 09:10:57,021 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 09:10:57,022 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 09:10:57,374 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 09:10:57,582 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 09:10:57,683 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1262914467_0001
2025-03-23 09:10:57,683 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 09:10:57,829 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local1262914467_0001_50215f34-446e-48f8-ae42-c70b2699188d/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 09:10:57,833 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local1262914467_0001_50215f34-446e-48f8-ae42-c70b2699188d/top50.txt
2025-03-23 09:10:57,886 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 09:10:57,886 INFO mapreduce.Job: Running job: job_local1262914467_0001
2025-03-23 09:10:57,886 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 09:10:57,892 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 09:10:57,892 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 09:10:57,892 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 09:10:57,893 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.3.4.2. Execution end time and successful completion.

```
2025-03-23 09:13:14,554 INFO mapred.LocalJobRunner: Finishing task: attempt_local1262914467_0001_r_0_00000_0
2025-03-23 09:13:14,554 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:13:14,984 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:13:15,984 INFO mapreduce.Job: Job job_local1262914467_0001 completed successfully
2025-03-23 09:13:16,193 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697810569939
    FILE: Number of bytes written=48011309878
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=264228
    Map output bytes=2497997
    Map output materialized bytes=3086453
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=1099
    Reduce shuffle bytes=3086453
    Reduce input records=264228
    Reduce output records=1099
    Spilled Records=528456
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=1813
    Total committed heap usage (bytes)=41228808749056
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9950
```

3.3.4.3. Total Execution Time: 139.4 seconds.

3.4. Map-function with Stripe approach:

3.4.1. d = 1

3.4.1.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapfunctionStripe ..//Wikipedia-EN-20120601_ARTICLES ./output/map_function_stripe/d1 .../A
/top50.txt -d 1
2025-03-23 09:14:48,075 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 09:14:48,178 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 09:14:48,178 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 09:14:48,229 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to rem
edy this.
2025-03-23 09:14:48,230 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 09:14:48,591 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 09:14:48,801 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 09:14:48,894 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local602422497_0001
2025-03-23 09:14:48,895 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 09:14:49,051 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local602422497_0001_9241de85-bb0d-498e-9542-ec93f47bf35f/top50.txt <-
/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 09:14:49,051 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local
/job_local602422497_0001_9241de85-bb0d-498e-9542-ec93f47bf35f/top50.txt
2025-03-23 09:14:49,115 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 09:14:49,116 INFO mapreduce.Job: Running job: job_local602422497_0001
2025-03-23 09:14:49,116 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 09:14:49,120 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 09:14:49,121 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 09:14:49,121 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 09:14:49,122 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.4.1.2. Execution end time and successful completion.

```
2025-03-23 09:17:04,466 INFO mapred.LocalJobRunner: Finishing task: attempt_local602422497_0001_r_000000
2025-03-23 09:17:04,466 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:17:05,203 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:17:06,203 INFO mapreduce.Job: Job job_local602422497_0001 completed successfully
2025-03-23 09:17:06,467 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697831865223
    FILE: Number of bytes written=114015205885
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=66475
    Map output bytes=13474670
    Map output materialized bytes=13734095
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=13734095
    Reduce input records=66475
    Reduce output records=49
    Spilled Records=132950
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=1707
    Total committed heap usage (bytes)=41239172874240
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=8963
```

3.4.1.3. Total Execution Time: 138.5 seconds.

3.4.2. d = 2

3.4.2.1. Execution start time.

```
sheikh@sheikh-latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapfunctionStripe ../Wikipedia-EN-20120601_ARTICLES ./output/map_function_stripe/d2 ..A
./top50.txt -d 2
2025-03-23 09:17:54,577 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 09:17:54,654 INFO input.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 09:17:54,654 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 09:17:54,692 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 09:17:54,693 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 09:17:55,023 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 09:17:55,207 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 09:17:55,294 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local439285302_0001
2025-03-23 09:17:55,294 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 09:17:55,419 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local439285302_0001_192618cc-a84d-4e0f-ae81-e57ceb85e946/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 09:17:55,423 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local439285302_0001_192618cc-a84d-4e0f-ae81-e57ceb85e946/top50.txt
2025-03-23 09:17:55,481 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 09:17:55,481 INFO mapreduce.LocalJobRunner: OutputCommitter set in config null
2025-03-23 09:17:55,481 INFO mapreduce.Job: Running job: job_local439285302_0001
2025-03-23 09:17:55,486 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 09:17:55,487 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-03-23 09:17:55,487 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 09:17:55,487 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.4.2.2. Execution end time and successful completion.

```
2025-03-23 09:20:17,224 INFO mapred.LocalJobRunner: Finishing task: attempt_local439285302_0001_r_000000_0
2025-03-23 09:20:17,225 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:20:17,591 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:20:18,592 INFO mapreduce.Job: Job job_local439285302_0001 completed successfully
2025-03-23 09:20:18,807 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=16978424737097
    FILE: Number of bytes written=145254093460
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=92257
    Map output bytes=18701566
    Map output materialized bytes=19038337
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=19038337
    Reduce input records=92257
    Reduce output records=49
    Spilled Records=184514
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=2005
    Total committed heap usage (bytes)=41239988666368
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9474
```

3.4.2.3. Total Execution Time: 144.2 seconds.

3.4.3 d = 3

3.4.3.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D$ hadoop jar prob_4_d.jar MapfunctionStripe ../Wikipedia-EN-20120601_ARTICLES ./output/map_function_stripe/d3 .../A/top50.txt -d 3
2025-03-23 09:21:40,917 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 09:21:41,017 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 09:21:41,018 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 09:21:41,066 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to run
eddy this.
2025-03-23 09:21:41,067 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 09:21:41,396 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 09:21:41,616 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 09:21:41,712 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local248990572_0001
2025-03-23 09:21:41,712 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 09:21:41,874 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local248990572_0001_e66c27c7-cedc-49da-9312-a9ae36ebe7dd/top50.txt <-
/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/D/top50.txt
2025-03-23 09:21:41,876 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment_2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/
job_local248990572_0001_e66c27c7-cedc-49da-9312-a9ae36ebe7dd/top50.txt
2025-03-23 09:21:41,923 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 09:21:41,924 INFO mapreduce.Job: Running job: job_local248990572_0001
2025-03-23 09:21:41,924 INFO Mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 09:21:41,930 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 09:21:41,931 INFO output.FileOutputCommitter: File Output Committor Algoithm version is 2
2025-03-23 09:21:41,931 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 09:21:41,932 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.4.3.2. Execution end time and successful completion.

```
2025-03-23 09:23:53,147 INFO mapred.LocalJobRunner: Finishing task: attempt_local248990572_0001_r_000000_0
2025-03-23 09:23:53,147 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:23:54,016 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:23:54,016 INFO mapreduce.Job: Job job_local248990572_0001 completed successfully
2025-03-23 09:23:54,244 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697850662883
    FILE: Number of bytes written=168963398396
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=112163
    Map output bytes=22736436
    Map output materialized bytes=23132925
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=23132925
    Reduce input records=112163
    Reduce output records=49
    Spilled Records=224326
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=1736
    Total committed heap usage (bytes)=41239665704960
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=9810
```

3.4.3.3. Total Execution Time: 133.4 seconds.

3.4.4. d = 4

3.4.4.1. Execution start time.

```
sheikh@sheikh-Latitude-3420:~/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/D$ hadoop jar prob_4_d.jar MapfunctionStripe ../Wikipedia-EN-20120601_ARTICLES ./output/map_function_stripe/d4 ..//A
/top50.txt -d 4
2025-03-23 09:24:46,126 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-03-23 09:24:46,203 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-03-23 09:24:46,203 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-03-23 09:24:46,244 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-23 09:24:46,245 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-03-23 09:24:46,580 INFO input.FileInputFormat: Total input files to process : 10000
2025-03-23 09:24:46,785 INFO mapreduce.JobSubmitter: number of splits:10000
2025-03-23 09:24:46,883 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local472740214_0001
2025-03-23 09:24:46,883 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-23 09:24:47,011 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-sheikh/mapred/local/job_local472740214_0001_4d68d5db-3e69-4d03-9fc5-8c2311b69040/top50.txt <- /home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/A/top50.txt
2025-03-23 09:24:47,016 INFO mapred.LocalDistributedCacheManager: Localized file:/home/sheikh/Muteeb_Laptop/Semester_8/NoSQL/Assignment 2b/A/top50.txt as file:/tmp/hadoop-sheikh/mapred/local/job_local472740214_0001_4d68d5db-3e69-4d03-9fc5-8c2311b69040/top50.txt
2025-03-23 09:24:47,067 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-03-23 09:24:47,068 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-03-23 09:24:47,068 INFO mapreduce.Job: Running job: job_local472740214_0001
2025-03-23 09:24:47,074 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-03-23 09:24:47,075 INFO output.FileOutputCommitter: File Output Committer Algoithm version is 2
2025-03-23 09:24:47,075 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-03-23 09:24:47,076 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

3.4.4.2. Execution end time and successful completion.

```
2025-03-23 09:27:03,793 INFO mapred.LocalJobRunner: Finishing task: attempt_local472740214_0001_r_000000_0
2025-03-23 09:27:03,793 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-03-23 09:27:04,198 INFO mapreduce.Job: map 100% reduce 100%
2025-03-23 09:27:05,198 INFO mapreduce.Job: Job job_local472740214_0001 completed successfully
2025-03-23 09:27:05,411 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=1697855569627
    FILE: Number of bytes written=182128310981
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
  Map-Reduce Framework
    Map input records=10000
    Map output records=124087
    Map output bytes=25154036
    Map output materialized bytes=25586297
    Input split bytes=1686997
    Combine input records=0
    Combine output records=0
    Reduce input groups=49
    Reduce shuffle bytes=25586297
    Reduce input records=124087
    Reduce output records=49
    Spilled Records=248174
    Shuffled Maps =10000
    Failed Shuffles=0
    Merged Map outputs=10000
    GC time elapsed (ms)=1836
    Total committed heap usage (bytes)=41224515878912
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=166222646
  File Output Format Counters
    Bytes Written=10068
```

3.4.4.3. Total Execution Time: 139.3 seconds.