

Join Operation

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -mkdir /posts
```

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -put
```

```
E:\Desktop\sampleposts.tsv \posts
```

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -mkdir /users
```

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -put
```

```
E:\Desktop\sampleusers.tsv \users
```

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -cat \posts\sampleposts.tsv
```

```
"2312" "Feedback on Audio Quality" "cs101 production audio"
```

```
"100005361" "<p>We are looking for feedback on the audio in our videos. Tell  
us what you think and try to be as <em>specific</em> as possible.</p>"
```

```
"question"
```

```
"\N" "\N" "2012-02-23 00:28:02.321344+00" "2" "" "\N"
```

```
"201398145" "2014-01-14 17:18:35.613939+00" "2960" "\N" "\N" "524"
```

```
"f"
```

```
"2014856" "" "cs101 " "100022094" "<p>I also would like to know  
the answer to this question. An 'open exam' sounds great, but on the other hand  
it also seems pretty easy to cheat now: solutions have been posted and anybody  
only interested in a certificate wouldn't have much of a problem getting the  
highest distinction. So where is the catch??</p>" "answer" "2014706"
```

```
"2014706" "2012-07-01 10:32:36.302782+00" "0" "" "\N"
```

```
"100022094" "2012-07-01 10:32:36.302782+00" "2020501" "\N" "\N"
```

```
"0" "f"
```

```
"2004004" "" "cs101 " "100018705" "<p>But then why even the  
new variable q? Why not just modify the variable p?</p>" "comment"
```

```
"2003997" "2003993" "2012-05-03 21:07:52.028935+00" "2" ""
```

```
"\N" "100018705" "2012-05-03 21:07:52.028935+00" "2005150" "\N"
```

```
"\N" "0" "f"
```

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -cat \users\sampleusers.tsv
```

```
"100006402"  "18"  "0"  "0"  "0"
"100022094"  "6354" "4"  "12" "50"
"100018705"  "76"  "0"  "3"  "4"
"100005361"  "36134" "73"  "220" "333"
```

```
c:\hadoop_new\share\hadoop\mapreduce>hadoop jar E:\Desktop\Join.jar
JoinDriver \join \users \joinOutput
2022-06-12 12:28:44,441 INFO client.RMProxy: Connecting to
ResourceManager at /0.0.0.0:8032
2022-07-03 12:28:45,518 INFO client.RMProxy: Connecting to
ResourceManager at /0.0.0.0:8032
2022-07-03 12:28:46,975 INFO mapreduce.JobResourceUploader: Disabling
Erasure Coding for path:
/tmp/hadoop-yarn/staging/Admin/.staging/job_1623480742672_0001 2022-07-
03 12:28:47,543 INFO mapred.FileInputFormat: Total input files to process : 1
2022-07-03 12:28:47,635 INFO mapred.FileInputFormat: Total input files to
process : 1
2022-07-03 12:28:48,092 INFO mapreduce.JobSubmitter: number of splits:4
2022-07-03 12:28:53,031 INFO Configuration.deprecation:
yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead,
use yarn.system-metrics-publisher.enabled
2022-07-03 12:28:53,944 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1623480742672_0001
2022-07-03 12:28:53,947 INFO mapreduce.JobSubmitter: Executing with tokens:
[]
2022-07-03 12:28:54,424 INFO conf.Configuration: resource-types.xml not found
2022-07-03 12:28:54,426 INFO resource.ResourceUtils: Unable to find
'resource-types.xml'.
2022-07-03 12:28:55,154 INFO impl.YarnClientImpl: Submitted application
application_1623480742672_0001
2022-07-03 12:28:55,293 INFO mapreduce.Job: The url to track the job:
http://DESKTOP-CK14PFH:8088/proxy/application_1623480742672_0001/
2022-07-03 12:28:55,295 INFO mapreduce.Job: Running job:
job_1623480742672_0001
2022-07-03 12:29:19,847 INFO mapreduce.Job: Job job_1623480742672_0001
running in uber mode : false
```

2022-07-03 12:29:19,874 INFO mapreduce.Job: map 0% reduce 0%
2022-07-03 12:31:53,514 INFO mapreduce.Job: map 67% reduce 0% 2022-07-03 12:31:59,518 INFO mapreduce.Job: map 83% reduce 0%
2022-07-03 12:32:00,667 INFO mapreduce.Job: map 100% reduce 0%
2022-07-03 12:33:23,194 INFO mapreduce.Job: map 100% reduce 100% 2022-07-03 12:33:32,307 INFO mapreduce.Job: Job job_1623480742672_0001 completed successfully
2022-07-03 12:33:32,532 INFO mapreduce.Job: Counters: 53

File System Counters

FILE: Number of bytes read=155
FILE: Number of bytes written=1071678
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=2595
HDFS: Number of bytes written=71
HDFS: Number of read operations=17
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters

Launched map tasks=4
Launched reduce tasks=1
Data-local map tasks=4
Total time spent by all maps in occupied slots (ms)=630042
Total time spent by all reduces in occupied slots (ms)=80087
Total time spent by all map tasks (ms)=630042
Total time spent by all reduce tasks (ms)=80087
Total vcore-milliseconds taken by all map tasks=630042
Total vcore-milliseconds taken by all reduce tasks=80087
Total megabyte-milliseconds taken by all map tasks=645163008
Total megabyte-milliseconds taken by all reduce tasks=82009088

Map-Reduce Framework

Map input records=7
Map output records=7
Map output bytes=135

Map output materialized bytes=173
Input split bytes=750
Combine input records=0
Combine output records=0
Reduce input groups=4
Reduce shuffle bytes=173
Reduce input records=7
Reduce output records=3
Spilled Records=14
Shuffled Maps =4
Failed Shuffles=0
Merged Map outputs=4
GC time elapsed (ms)=903
CPU time spent (ms)=15864
Physical memory (bytes) snapshot=990265344
Virtual memory (bytes) snapshot=1415651328
Total committed heap usage (bytes)=663224320
Peak Map Physical memory (bytes)=219955200
Peak Map Virtual memory (bytes)=296644608
Peak Reduce Physical memory (bytes)=204709888
Peak Reduce Virtual memory (bytes)=246579200
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=71

c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -ls \joinOutput
Found 2 items

```
-rw-r--r-- 1 Admin supergroup      0 2022-07-03 12:33 /joinOutput/_SUCCESS
-rw-r--r-- 1 Admin supergroup    71 2022-07-03 12:33 /joinOutput/part-00000
```

```
c:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -cat \joinOutput\part-00000
```

```
"100005361"  "2"      "36134"
```

```
"100018705"  "2"      "76"
```

```
"100022094"  "0"      "6354"
```