

# How to Run MapReduce (Jar file) in Windows 10

All black Text are taken from run-time environment and blue text are augmented later for learning purposes.

Dr. Supriya Chakraborty

Download MapReduce program from the link :

<https://github.com/MuhammadBilalYar/Hadoop-On-Window/blob/master/MapReduceClient.jar>

Download the Input file from the link: [https://github.com/MuhammadBilalYar/Hadoop-On-Window/blob/master/input\\_file.txt](https://github.com/MuhammadBilalYar/Hadoop-On-Window/blob/master/input_file.txt)

---

- Go to the sbin directory in Hadoop to start:

```
C:\>cd bigdata
```

```
C:\bigdata>dir
```

Volume in drive C has no label.

Volume Serial Number is D6BF-FB97

Directory of C:\bigdata

```
27-01-2021 15:12 <DIR>      .
27-01-2021 15:12 <DIR>      ..
22-01-2021 16:53 <DIR>      data
22-01-2021 17:07 <DIR>      hadoop-3.2.1
21-01-2021 00:12    359,196,911 hadoop-3.2.1.tar.gz
09-10-2019 06:53 <DIR>      winutils-master
22-01-2021 16:17    20,758,178 winutils-master.zip
```

```
2 File(s) 379,955,089 bytes
```

```
5 Dir(s) 73,227,182,080 bytes free
```

```
C:\bigdata>cd hadoop-3.2.1
```

```
C:\bigdata\hadoop-3.2.1>start-all.cmd
```

'start-all.cmd' is not recognized as an internal or external command,  
operable program or batch file.

```
C:\bigdata\hadoop-3.2.1>cd sbin
```

- Under SBIN directory, start all the required services of hadoop

```
C:\bigdata\hadoop-3.2.1\sbin>start-all.cmd
```

This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd

starting yarn daemons

- Create a directory

```
C:\bigdata\hadoop-3.2.1\sbin>hadoop fs -mkdir /input_dir
```

- Transfer the input file into the HDFS

```
C:\bigdata\hadoop-3.2.1\sbin>hadoop fs -put c:/input_file.txt /input_dir
```

2022-04-18 15:17:44,666 INFO sasl.SaslDataTransferClient: SASL encryption trust check:  
localhostTrusted = false, remoteHostTrusted = false

```
C:\bigdata\hadoop-3.2.1\sbin>cd input_dir
```

The system cannot find the path specified.

```
C:\bigdata\hadoop-3.2.1\sbin>dir
```

Volume in drive C has no label.

Volume Serial Number is D6BF-FB97

Directory of C:\bigdata\hadoop-3.2.1\sbin

23-01-2021 22:52 <DIR> .

23-01-2021 22:52 <DIR> ..

10-09-2019 21:31 2,756 distribute-exclude.sh

```

22-01-2021 15:09 <DIR>      FederationStateStore
10-09-2019 21:27      1,983 hadoop-daemon.sh
10-09-2019 21:27      2,522 hadoop-daemons.sh
10-09-2019 21:34      1,542 httpfs.sh
10-09-2019 21:28      1,500 kms.sh
10-09-2019 22:06      1,841 mr-jobhistory-daemon.sh
10-09-2019 21:31      2,086 refresh-namenodes.sh
10-09-2019 21:27      1,779 start-all.cmd
10-09-2019 21:27      2,221 start-all.sh
10-09-2019 21:31      1,880 start-balancer.sh
10-09-2019 21:31      1,401 start-dfs.cmd
10-09-2019 21:31      5,170 start-dfs.sh
10-09-2019 21:31      1,793 start-secure-dns.sh
10-09-2019 22:03      1,571 start-yarn.cmd
10-09-2019 22:03      3,342 start-yarn.sh
10-09-2019 21:27      1,770 stop-all.cmd
10-09-2019 21:27      2,166 stop-all.sh
10-09-2019 21:31      1,783 stop-balancer.sh
10-09-2019 21:31      1,455 stop-dfs.cmd
10-09-2019 21:31      3,898 stop-dfs.sh
10-09-2019 21:31      1,756 stop-secure-dns.sh
10-09-2019 22:03      1,642 stop-yarn.cmd
10-09-2019 22:03      3,083 stop-yarn.sh
23-01-2021 22:58 <DIR>      test
10-09-2019 21:27      1,982 workers.sh
10-09-2019 22:03      1,814 yarn-daemon.sh
10-09-2019 22:03      2,328 yarn-daemons.sh

      26 File(s)      57,064 bytes
      4 Dir(s) 73,210,441,728 bytes free

```

```
C:\bigdata\hadoop-3.2.1\sbin>dir input_dir
```

Volume in drive C has no label.

Volume Serial Number is D6BF-FB97

Directory of C:\bigdata\hadoop-3.2.1\sbin

File Not Found

C:\bigdata\hadoop-3.2.1\sbin>hadoop fs -mkdir /input\_dir

mkdir: `/input\_dir': File exists

C:\bigdata\hadoop-3.2.1\sbin> dir input\_dir

Volume in drive C has no label.

Volume Serial Number is D6BF-FB97

Directory of C:\bigdata\hadoop-3.2.1\sbin

File Not Found

- Check existence of the created directory into HDFS

C:\bigdata\hadoop-3.2.1\sbin>hadoop fs -ls /input\_dir/

Found 1 items

-rw-r--r-- 1 IT supergroup 1888 2022-04-18 15:17 /input\_dir/input\_file.txt

- Check existence of the input file into the HDFS

C:\bigdata\hadoop-3.2.1\sbin>hadoop dfs -cat /input\_dir/input\_file.txt

DEPRECATED: Use of this script to execute hdfs command is deprecated.

Instead use the hdfs command for it.

2022-04-18 15:23:48,327 INFO sasl.SaslDataTransferClient: SASL encryption trust check:  
localhostTrusted = false, remoteHostTrusted = false

23 23 27 43 24 25 26 26 26 26 25 26 25

26 27 28 28 28 30 31 31 31 30 30 30 29

31 32 32 32 33 34 35 36 36 34 34 34 34

39 38 39 39 39 41 42 43 40 39 38 38 40

38 39 39 39 39 41 41 41 28 40 39 39 45  
23 23 27 43 24 25 26 26 26 26 25 26 25  
26 27 28 28 28 30 31 31 31 30 30 30 29  
31 32 32 32 33 34 35 36 36 34 34 34 34  
39 38 39 39 39 41 42 43 40 39 38 38 40  
38 39 39 39 39 41 41 41 28 40 39 39 45  
23 23 27 43 24 25 26 26 26 26 25 26 25  
26 27 28 28 28 30 31 31 31 30 30 30 29  
31 32 32 32 33 34 35 36 36 34 34 34 34  
39 38 39 39 39 41 42 43 40 39 38 38 40  
38 39 39 39 39 41 41 41 28 40 39 39 45  
23 23 27 43 24 25 26 26 26 26 25 26 25  
26 27 28 28 28 30 31 31 31 30 30 30 29  
31 32 32 32 33 34 35 36 36 34 34 34 34  
39 38 39 39 39 41 42 43 40 39 38 38 40  
38 39 39 39 39 41 41 41 28 40 39 39 45  
23 23 27 43 24 25 26 26 26 26 25 26 25  
26 27 28 28 28 30 31 31 31 30 30 30 29  
31 32 32 32 33 34 35 36 36 34 34 34 34  
39 38 39 39 39 41 42 43 40 39 38 38 40  
38 39 39 39 39 41 41 41 28 40 39 39 45  
23 23 27 43 24 25 26 26 26 26 25 26 25  
26 27 28 28 28 30 31 31 31 30 30 30 29  
31 32 32 32 33 34 35 36 36 34 34 34 34  
39 38 39 39 39 41 42 43 40 39 38 38 40  
38 39 39 39 39 41 41 41 28 40 39 39 45

### ■ Run the map reduce program

```
C:\bigdata\hadoop-3.2.1\sbin>hadoop jar c:/MapReduceClient.jar wordcount /input_dir /output_dir
```

```
2022-04-18 15:25:24,516 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
```

```
2022-04-18 15:25:25,752 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path:  
/tmp/hadoop-yarn/staging/IT/.staging/job_1650275077275_0001
```

2022-04-18 15:25:26,040 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-04-18 15:25:26,372 INFO input.FileInputFormat: Total input files to process : 1

2022-04-18 15:25:26,651 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-04-18 15:25:26,850 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-04-18 15:25:26,912 INFO mapreduce.JobSubmitter: number of splits:1

2022-04-18 15:25:27,602 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-04-18 15:25:27,730 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1650275077275\_0001

2022-04-18 15:25:27,730 INFO mapreduce.JobSubmitter: Executing with tokens: []

2022-04-18 15:25:28,056 INFO conf.Configuration: resource-types.xml not found

2022-04-18 15:25:28,057 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2022-04-18 15:25:28,564 INFO impl.YarnClientImpl: Submitted application application\_1650275077275\_0001

2022-04-18 15:25:28,666 INFO mapreduce.Job: The url to track the job: [http://AUKLT0364:8088/proxy/application\\_1650275077275\\_0001/](http://AUKLT0364:8088/proxy/application_1650275077275_0001/)

2022-04-18 15:25:28,668 INFO mapreduce.Job: Running job: job\_1650275077275\_0001

2022-04-18 15:25:41,906 INFO mapreduce.Job: Job job\_1650275077275\_0001 running in uber mode : false

2022-04-18 15:25:41,908 INFO mapreduce.Job: map 0% reduce 0%

2022-04-18 15:25:48,820 INFO mapreduce.Job: map 100% reduce 0%

2022-04-18 15:25:57,969 INFO mapreduce.Job: map 100% reduce 100%

2022-04-18 15:25:58,990 INFO mapreduce.Job: Job job\_1650275077275\_0001 completed successfully

2022-04-18 15:25:59,241 INFO mapreduce.Job: Counters: 50

#### File System Counters

FILE: Number of bytes read=195

FILE: Number of bytes written=454761

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=1998  
HDFS: Number of bytes written=120  
HDFS: Number of read operations=8  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=4729  
Total time spent by all reduces in occupied slots (ms)=6384  
Total time spent by all map tasks (ms)=4729  
Total time spent by all reduce tasks (ms)=6384  
Total vcore-milliseconds taken by all map tasks=4729  
Total vcore-milliseconds taken by all reduce tasks=6384  
Total megabyte-milliseconds taken by all map tasks=4842496  
Total megabyte-milliseconds taken by all reduce tasks=6537216

#### Map-Reduce Framework

Map input records=30  
Map output records=390  
Map output bytes=2730  
Map output materialized bytes=195  
Input split bytes=110  
Combine input records=390  
Combine output records=21  
Reduce input groups=21  
Reduce shuffle bytes=195  
Reduce input records=21  
Reduce output records=21  
Spilled Records=42

Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=109  
CPU time spent (ms)=0  
Physical memory (bytes) snapshot=0  
Virtual memory (bytes) snapshot=0  
Total committed heap usage (bytes)=338690048

#### 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=1888

#### File Output Format Counters

Bytes Written=120

### ▪ Check the output of the map reduce

```
C:\bigdata\hadoop-3.2.1\sbin>hadoop dfs -cat /output_dir/*
```

DEPRECATED: Use of this script to execute hdfs command is deprecated.

Instead use the hdfs command for it.

2022-04-18 15:27:06,747 INFO sasl.SaslDataTransferClient: SASL encryption trust check:  
localhostTrusted = false, remoteHostTrusted = false

23 12

24 6

25 18

26 36

27 12

28 24



29	6
30	24
31	24
32	18
33	6
34	30
35	6
36	12
38	24
39	66
40	18
41	24
42	6
43	12
45	6

C:\bigdata\hadoop-3.2.1\sbin>

## Find The WordCount Code and Generate the Jar file according to your Configuration

Find the WordCount implementation with Java in Hadoop according to your platform. Identify the classes like mapper, reducer and main class. Just try to identify what each class is doing, not how they are doing. The detail of working principal of each class will be done in the theory class.

Generate the jar file of the classes according to Hadoop.