

Labsheet-2

Mapreduce

```
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

```
cd $HADOOP_HOME/sbin
```

```
./start-all.sh
```

```
* Connection #0 to host localhost left intact
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ jps
33810 NodeManager
34884 Jps
33653 ResourceManager
32057 NameNode
32461 SecondaryNameNode
32239 DataNode
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ |
```

1 *hdfs dfs -mkdir /bigdata*

```
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ hdfs dfs -mkdir /Bigdata
mkdir: '/Bigdata': File exists
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x - hadoop supergroup 0 2025-02-22 23:44 /Bigdata
drwxr-xr-x - hadoop supergroup 0 2025-02-22 23:41 /bigdata
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ |
```

2 *hdfs dfs -touchz /bigdata/input*

```
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ hdfs dfs -touchz /bigdata/input
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ hdfs dfs -ls /bigdata/
Found 1 items
-rw-r--r-- 1 hadoop supergroup 0 2025-02-22 23:45 /bigdata/input
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ |
```

3 *hdfs dfs -appendToFile -/bigdata /input*

aaaaa bbbb cc aaaa cc

cc bbbb zzzz ctrl+d ctrl+ d

```
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ echo "aaaaa bbbb cc aaaa cc\ncc bbbb zzzz" | hdfs dfs -appendToFile - /bigdata/input
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ hdfs dfs -cat /bigdata/input
aaaaa bbbb cc aaaa cc\ncc bbbb zzzz
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ |
```

5 *cd \$HADOOP_HOME*

6 *ls*

```

hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ cd $HADOOP_HOME
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0$ ls
LICENSE-binary LICENSE.txt NOTICE-binary NOTICE.txt README.txt bin etc include lib libexec licenses-binary logs sbin share
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0$ |

```

7 *cd share*

8 *ls*

9 *cd Hadoop*

10 *ls*

11 *cd mapreduce*

12 *ls*

hadoop-mapreduce-client-app-3.3.0.jar

hadoop-mapreduce-client-common-3.3.0.jar

hadoop-mapreduce-client-core-3.3.0.jar

hadoop-mapreduce-client-hs-3.3.0.jar

hadoop-mapreduce-client-hs-plugins-3.3.0.jar

hadoop-mapreduce-client-jobclient-3.3.0.jar

hadoop-mapreduce-client-jobclient-3.3.0-tests.jar

hadoop-mapreduce-client-nativetask-3.3.0.jar

hadoop-mapreduce-client-shuffle-3.3.0.jar

hadoop-mapreduce-client-uploader-3.3.0.jar

hadoop-mapreduce-examples-3.3.0.jar

jdifff

lib-examples

sources

```

hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/sbin$ cd $HADOOP_HOME
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0$ ls
LICENSE-binary LICENSE.txt NOTICE-binary NOTICE.txt README.txt bin etc include lib libexec licenses-binary logs sbin share
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0$ cd share
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share$ ls
doc hadoop
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share$ cd hadoop
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share/hadoop$ ls
client common hdfs mapreduce tools yarn
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share/hadoop$ cd mapreduce
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share/hadoop/mapreduce$ ls
hadoop-mapreduce-client-app-3.4.0.jar      hadoop-mapreduce-client-hs-3.4.0.jar      hadoop-mapreduce-client-jobclient-3.4.0.jar  hadoop-mapreduce-client-uploader-3.4.0.jar  sources
hadoop-mapreduce-client-common-3.4.0.jar  hadoop-mapreduce-client-hs-plugins-3.4.0.jar  hadoop-mapreduce-client-nativetask-3.4.0.jar  hadoop-mapreduce-examples-3.4.0.jar
hadoop-mapreduce-client-core-3.4.0.jar    hadoop-mapreduce-client-jobclient-3.4.0-tests.jar  hadoop-mapreduce-client-shuffle-3.4.0.jar    jdifff
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share/hadoop/mapreduce$ |

```

13 *hadoop jar hadoop-mapreduce-examples-3.3.0.jar wordcount /bigdata /out*

```
hadoop@DESKTOP-O7BK13U: ~$ hadoop jar hadoop-mapreduce-examples-3.4.0.jar wordcount /bigdata /out
2025-02-22 23:56:02,320 INFO client.DefaultHARMAFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-02-22 23:56:03,111 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1740244122322_0001
2025-02-22 23:56:03,497 INFO input.FileInputFormat: Total input files to process : 1
2025-02-22 23:56:03,580 INFO mapreduce.JobSubmitter: number of splits:1
2025-02-22 23:56:03,919 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1740244122322_0001
2025-02-22 23:56:03,919 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-02-22 23:56:04,282 INFO conf.Configuration: resource-types.xml not found
2025-02-22 23:56:04,282 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-02-22 23:56:04,933 INFO impl.YarnClientImpl: Submitted application application_1740244122322_0001
2025-02-22 23:56:05,004 INFO mapreduce.Job: The url to track the job: http://DESKTOP-O7BK13U:8088/proxy/application_1740244122322_0001/
2025-02-22 23:56:05,005 INFO mapreduce.Job: Running job: job_1740244122322_0001
2025-02-22 23:56:15,264 INFO mapreduce.Job: Job job_1740244122322_0001 running in uber mode : false
2025-02-22 23:56:15,266 INFO mapreduce.Job: map 0% reduce 0%
2025-02-22 23:56:21,455 INFO mapreduce.Job: map 100% reduce 0%
2025-02-22 23:56:29,544 INFO mapreduce.Job: map 100% reduce 100%
2025-02-22 23:56:30,575 INFO mapreduce.Job: Job job_1740244122322_0001 completed successfully
2025-02-22 23:56:30,726 INFO mapreduce.Job: Counters: 54

File System Counters
  FILE: Number of bytes read=62
  FILE: Number of bytes written=617003
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=137
  HDFS: Number of bytes written=36
  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
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=3810
  Total time spent by all reduces in occupied slots (ms)=4492
  Total time spent by all map tasks (ms)=3810
  Total time spent by all reduce tasks (ms)=4492
  Total vcore-milliseconds taken by all map tasks=3810
  Total vcore-milliseconds taken by all reduce tasks=4492
  Total megabyte-milliseconds taken by all map tasks=3901440
  Total megabyte-milliseconds taken by all reduce tasks=4599808

Map-Reduce Framework
  Map input records=1
  Map output records=7
  Map output bytes=65
  Map output materialized bytes=62
  Input split bytes=100
  Combine input records=7
  Combine output records=5
  Reduce input groups=5
  Reduce shuffle bytes=62
  Reduce input records=5
  Reduce output records=5
  Spilled Records=10
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=70
```

14 *hdfs dfs -ls /out*

Found 2 items

<i>-rw-r--r--</i>	<i>1</i>	<i>hadoop supergroup</i>	<i>0</i>	<i>2023-04-03 15:20</i>
<i>/output/_SUCCESS</i>				
<i>-rw-r--r--</i>	<i>1</i>	<i>hadoop supergroup</i>	<i>63</i>	<i>2023-04-03 15:20</i>
<i>/output/part-r-00000</i>				

```
hadoop@DESKTOP-O7BK13U: ~/hadoop-3.4.0/share/hadoop/mapreduce$ hdfs dfs -ls /out
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2025-02-22 23:56 /out/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 36 2025-02-22 23:56 /out/part-r-00000
hadoop@DESKTOP-O7BK13U: ~/hadoop-3.4.0/share/hadoop/mapreduce$ |
```

15 *hdfs dfs -cat /out/part**

aaaaa 2

bbbb 2

cc 3

zzzz 1

```
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share/hadoop/mapreduce$ hdfs dfs -cat /out/part*
aaaaa 2
bbbb 2
cc 1
cc\ncc 1
zzzz 1
hadoop@DESKTOP-07BKI3U:~/hadoop-3.4.0/share/hadoop/mapreduce$ |
```