

1. Follow the tutorial to upload the file to cluster:

Name	Size	Changed	Rights	Owner
..		2/5/2019 1:11:34 PM	rw-r--r--	root
perl5		2/12/2019 3:35:28 PM	rw-r--r--	dyang33
wiki-micro.txt	5,512 KB	2/12/2019 9:34:13 PM	r-----	dyang33
WordCount.java	3 KB	2/12/2019 3:52:14 PM	r-----	dyang33

2. After the source code and text file is uploaded, go to step 3, using the command to create a directory on the HDFS:

\$ `hadoop fs -mkdir /user/<username>/input`

3. To see if you create the directory or not, using the following command. If you see this, it means you have created the directory.

```
drwxr-xr-x  - dyang33 dyang33      0 2019-02-12 15:57 input
```

4. Now put the text file on HDFS by using:

`hadoop fs -put wiki-micro.txt /user/<username>/input`

`$ hadoop fs -ls /user/<username>/input`

if you can see this, you have uploaded the file to HDFS.

```
[dyang33@mba-i2 ~]$ hadoop fs -put wiki-micro.txt /user/dyang33/input
[dyang33@mba-i2 ~]$ hadoop fs -ls /user/dyang33/input
Found 1 items
-rw-r--r--   3 dyang33 dyang33    5644081 2019-02-24 21:01 /user/dyang33/input/wiki-micro.txt
```

5. Compile the WordCount class:

`$ mkdir build`

`$ javac -cp /opt/cloudera/parcels/CDH/lib/hadoop/*:/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/* WordCount.java -d build -Xlint`

Then you will see the following warning. These warning are fine. You have compiled the source code.

```
[dyang33@mba-i2 ~]$ mkdir build
[dyang33@mba-i2 ~]$ javac -cp /opt/cloudera/parcels/CDH/lib/hadoop/*:/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/* WordCount.java -d build -Xlint
warning: [path] bad path element "/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/jaxb-api.jar": no such file or directory
warning: [path] bad path element "/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/activation.jar": no such file or directory
warning: [path] bad path element "/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/jsr173_1.0_api.jar": no such file or directory
warning: [path] bad path element "/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/jaxb1-impl.jar": no such file or directory
4 warnings
```

6. Then create a JAR file using:

jar -cvf wordcount.jar -C build/ .

```
[dyang33@mba-i2 ~]$ jar -cvf wordcount.jar -C build/ .
added manifest
adding: org/(in = 0) (out= 0)(stored 0%)
adding: org/myorg/(in = 0) (out= 0)(stored 0%)
adding: org/myorg/WordCount$Map.class(in = 2165) (out= 955)(deflated 55%)
adding: org/myorg/WordCount$Reduce.class(in = 1647) (out= 692)(deflated 57%)
adding: org/myorg/WordCount.class(in = 2029) (out= 999)(deflated 50%)
```

7. Run the application using

**hadoop jar wordcount.jar org.apache.hadoop.examples.WordCount
/user/<username>/input /user/<username>/output**

Then the application starts processing the input data:
























```
[dyang33@mba-i2 ~]$ hadoop jar wordcount.jar org.apache.hadoop.examples.WordCount  
t /user/<username>/input /user/<username>/output  
-bash: username: No such file or directory  
[dyang33@mba-i2 ~]$ hadoop jar wordcount.jar org.apache.hadoop.examples.WordCount  
t /user/dyang33/input /user/dyang33/output  
19/02/24 21:08:35 INFO hdfs.DFSCClient: Created token for dyang33: HDFS_DELEGATIO  
N_TOKEN owner=dyang33@ITS.UNCC.EDU, renewer=yarn, realUser=, issueDate=155106051  
5883, maxDate=1551665315883, sequenceNumber=155, masterKeyId=54 on ha-hdfs:DSBA  
19/02/24 21:08:35 INFO security.TokenCache: Got dt for hdfs://DSBA; Kind: HDFS_D  
ELEGATION_TOKEN, Service: ha-hdfs:DSBA, Ident: (token for dyang33: HDFS_DELEGATI  
ON_TOKEN owner=dyang33@ITS.UNCC.EDU, renewer=yarn, realUser=, issueDate=15510605  
15883, maxDate=1551665315883, sequenceNumber=155, masterKeyId=54)  
19/02/24 21:08:36 INFO input.FileInputFormat: Total input paths to process : 1  
19/02/24 21:08:36 INFO mapreduce.JobSubmitter: number of splits:1  
19/02/24 21:08:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15  
47055584507_0102  
19/02/24 21:08:36 INFO mapreduce.JobSubmitter: Kind: HDFS_DELEGATION_TOKEN, Serv  
ice: ha-hdfs:DSBA, Ident: (token for dyang33: HDFS_DELEGATION_TOKEN owner=dyang3  
3@ITS.UNCC.EDU, renewer=yarn, realUser=, issueDate=1551060515883, maxDate=155166  
5315883, sequenceNumber=155, masterKeyId=54)  
19/02/24 21:08:36 INFO impl.YarnClientImpl: Submitted application application_15  
47055584507_0102  
19/02/24 21:08:36 INFO mapreduce.Job: The url to track the job: http://mba-hml.u  
ncc.edu:8088/proxy/application_1547055584507_0102/  
19/02/24 21:08:36 INFO mapreduce.Job: Running job: job_1547055584507_0102  
19/02/24 21:08:43 INFO mapreduce.Job: Job job_1547055584507_0102 running in uber mode : false  
19/02/24 21:08:43 INFO mapreduce.Job: map 0% reduce 0%
```

8. Download the output from HDFS using

hadoop dfs -get /user/<username>/output

And then check all the output on WinSCP under /output directory. Zip up all the output files, then you have finished task 1!!

/users/dyang33/output/

Name	Size	Changed	Rights	Owner
 ^		2/24/2019 9:11:23 PM	rwxr-x---	dyang33
 _SUCCESS	0 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00000	34 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00001	34 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00002	33 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00003	35 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00004	33 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00005	35 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00006	36 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00007	33 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00008	34 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00009	35 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00010	34 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00011	34 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00012	34 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00013	35 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00014	35 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00015	36 KB	2/24/2019 9:11:23 PM	rw-r--r--	dyang33
 part-r-00016	34 KB	2/24/2019 9:11:24 PM	rw-r--r--	dyang33
 part-r-00017	35 KB	2/24/2019 9:11:24 PM	rw-r--r--	dyang33
 part-r-00018	34 KB	2/24/2019 9:11:24 PM	rw-r--r--	dyang33
 part-r-00019	35 KB	2/24/2019 9:11:24 PM	rw-r--r--	dyang33
 part-r-00020	34 KB	2/24/2019 9:11:24 PM	rw-r--r--	dyang33