

Running WordCount v1.0 (#topic_5_2)

You can run WordCount v1.0 in your own Hadoop environment. If you do not have one available, you can download and install the [Cloudera QuickStart VM](http://www.cloudera.com/content/cloudera-content/cloudera-docs/DemoVMs/Cloudera-QuickStart-VM/cloudera_quickstart_vm.html) (http://www.cloudera.com/content/cloudera-content/cloudera-docs/DemoVMs/Cloudera-QuickStart-VM/cloudera_quickstart_vm.html).

1. Before you run the sample, you must create input and output locations in HDFS. Use the following commands to create the input directory /user/cloudera/wordcount/input in HDFS:

```
$ sudo su hdfs
$ hadoop fs -mkdir /user/cloudera
$ hadoop fs -chown cloudera /user/cloudera
$ exit
$ sudo su cloudera
$ hadoop fs -mkdir /user/cloudera/wordcount /user/cloudera/wordcount/input
```

2. Create sample text files to use as input, and move them to the /user/cloudera/wordcount/input directory in HDFS. You can use any files you choose; for convenience, the following shell commands create a few small input files for illustrative purposes. The Makefile also contains most of the commands that follow.

```
$ echo "Hadoop is an elephant" > file0
$ echo "Hadoop is as yellow as can be" > file1
$ echo "Oh what a yellow fellow is Hadoop" > file2
$ hadoop fs -put file* /user/cloudera/wordcount/input
```

3. Compile the WordCount class.

To compile in a package installation of CDH:

```
$ mkdir -p build
$ javac -cp /usr/lib/hadoop/*:/usr/lib/hadoop-mapreduce/* WordCount.java -d build -Xlint
```

To compile in a parcel installation of CDH:

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

4. Create a JAR file for the WordCount application.

```
$ jar -cvf wordcount.jar -C build/ .
```

5. Run the WordCount application from the JAR file, passing the paths to the input and output directories in HDFS.

```
$ hadoop jar wordcount.jar org.myorg.WordCount /user/cloudera/wordcount/input /user/cloudera/wordcount/output
```

When you look at the output, all of the words are listed in UTF-8 alphabetical order (capitalized words first). The number of occurrences from all input files has been reduced to a single sum for each word.

```
$ hadoop fs -cat /user/cloudera/wordcount/output/*
Hadoop      3
Oh          1
a           1
an          1
as          2
be          1
can         1
elephant    1
fellow      1
is           3
what         1
yellow       2
```

6. If you want to run the sample again, you first need to remove the output directory. Use the following command.

```
$ hadoop fs -rm -r /user/cloudera/wordcount/output
```

[\(../categories/hub_tutorial.html\)](#) | [WordCount 1.0](#) ([\(../categories/hub_wordcount_1.0.html\)](#)) | [All Categories](#)
[\(../categories/hub.html\)](#)

- [About Cloudera](#) (<http://www.cloudera.com/about-cloudera.html>)
- [Resources](#) (<http://www.cloudera.com/resources.html>)
- [Contact](#) (<http://www.cloudera.com/contact-us.html>)
- [Careers](#) (<http://www.cloudera.com/about-cloudera/careers.html>)
- [Press](#) ([/about-cloudera/press-center.html](#))
- [Documentation](#) ([/documentation.html](#))

United States: +1 888 789 1488

Outside the US: +1 650 362 0488

© 2016 Cloudera, Inc. All rights reserved. [Apache Hadoop](#) (<http://hadoop.apache.org>) and associated open source project names are trademarks of the [Apache Software Foundation](#) (<http://apache.org>). For a complete list of trademarks, [click here](#). ([/legal/terms-and-conditions.html](#))

-  (<https://www.linkedin.com/company/cloudera>)
-  (<https://www.facebook.com/cloudera>)
-  (<https://twitter.com/cloudera>)
-  ([/contact-us.html](#))

[Terms & Conditions](#) ([/legal/terms-and-conditions.html](#)) | [Privacy Policy](#) ([/legal/privacy-policy.html](#))

Page generated June 29, 2016.