

Step 4: Running a Sample MapReduce Job

1. Upload Sample Data to HDFS:

Inside the Docker container, create an input directory and upload sample files to HDFS:

```
hdfs dfs -mkdir -p /user/hadoop/input
```

```
hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
```

○

2. Run the WordCount Job:

Run a built-in Hadoop example to test the setup:

bash

```
hadoop jar
```

```
$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar
```

```
wordcount /user/hadoop/input /user/hadoop/output
```

○

3. Check the Output:

Display the results of the WordCount job:

```
hdfs dfs -cat /user/hadoop/output/part-r-00000
```

○

Step 5: Exiting the Container

1. Stop the Container:

If you want to stop the Hadoop container without deleting it, run:

```
docker stop hadoop-cluster
```

○

2. Restart the Container:

To restart the container later, use:

```
docker start -i hadoop-cluster
```