MapReduce程序测试

一、Hadoop自带的wordcount程序测试：

参考文档：<https://www.cnblogs.com/xiaomila-study/p/4973662.html>

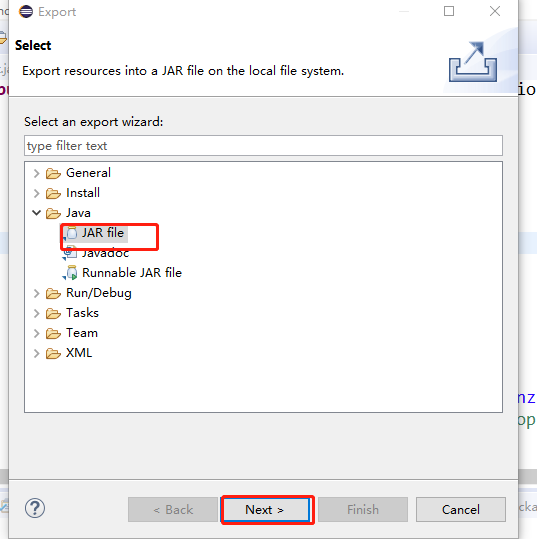
<http://hadoop.apache.org/docs/r1.0.4/cn/mapred_tutorial.html>

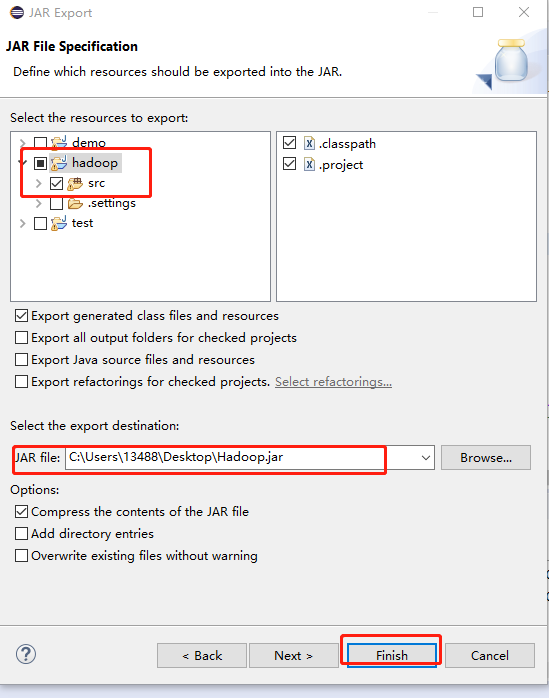
二、排序算法的测试：

1、本地调试完成后，将项目打成jar包：

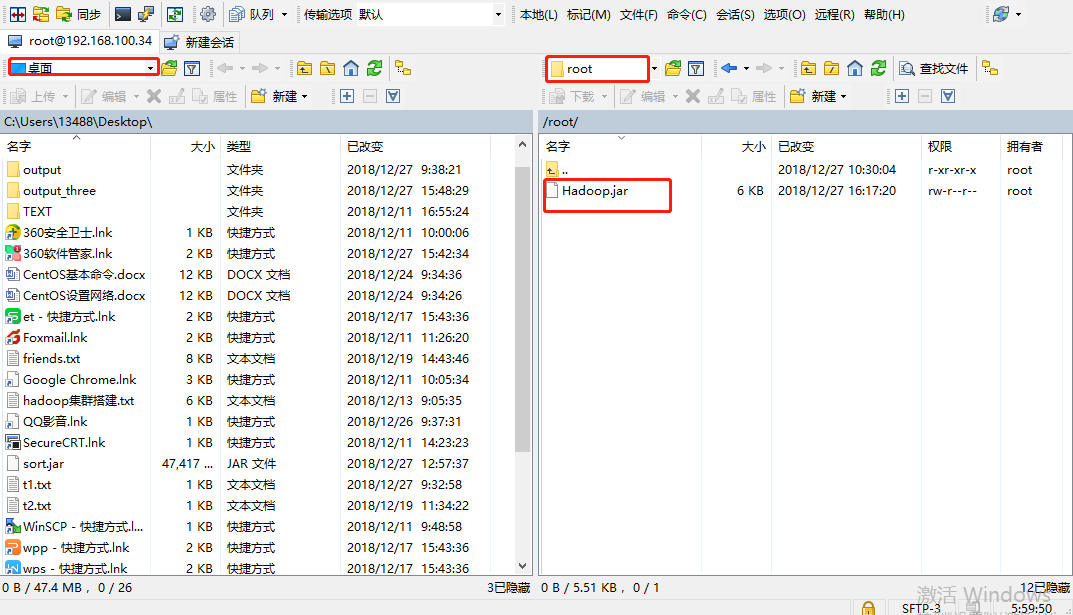
打jar包的方法：

右击项目名称----->export





2、将jar包上传到服务器：



3、将输入文件上传到服务器：

vi /root/t1.txt

4、将本地文件上传到hdfs上

hadoop fs -put /root/t1.txt /zhaoyanzi/input/

5、Hdfs操作命令：

查看：

hdfs dfs -ls /zhaoyanzi/

查看所有：

hdfs dfs -ls -R /zhaoyanzi/

查看文档：

hdfs dfs -cat /zhaoyanzi/output\_four/part-r-00000

跑jar包：

hadoop jar /root/Hadoop.jar com.hadoop.example1.MyText /zhaoyanzi/input/t1.txt /zhaoyanzi/input/t2.txt /zhaoyanzi/output/

删除非空文件夹：

hdfs dfs -rm -r /zhaoyanzi/output\_four

sudo -u hdfs hdfs dfs -rm -r /zhaoyanzi

三、测试排序合并代码遇到的问题：

1. JDK版本不兼容，集群是jdk1.7，本地是jdk1.8，解决办法：将本地改成jdk1.7
2. 代码里面的问题：

报错信息：

WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).

Error:java.lang.RuntimeException: java.lang.ClassNotFoundException:

Class com.hadoop.example1.MyText$MyMapper not found

解决办法：在代码里面加入：

conf.set("mapred.jar", "Hadoop.jar");

1. 对于排序完成结果分为三个文件的解决办法：

将reduce设置为1：

job.setNumReduceTasks(1);

1. 针对代码中路径写死问题的修改：

FileInputFormat.*addInputPath*(job, **new** Path(args[0]));

FileOutputFormat.*setOutputPath*(job, **new** Path(args[1]));