

*A friendly  
place for  
programming  
greenhorns!*

## The Big Moose Saloon

[Search](#) | [Java FAQ](#) | [Recent Topics](#)  
[Register / Login](#)

[JavaRanch](#) » [Java Forums](#) » [Java](#) » [Distributed Java](#)

[Post Reply](#)

[Bookmark It](#)

[Watch Topic](#)

[New Topic](#)

Author

## Hadoop key mismatch

**Larry Homes**

Greenhorn

Joined: Jan 18, 2009  
Posts: 25

posted Feb 24, 2010 12:48:55



Hello,

Hope this is the correct forum for a hadoop question.

I have a file with a bunch of lines like this:

```
education:43 Alabama
education:11 Alaska
education:44 Arizona
```

It continues on for all 50 states, then there is another word like politics:30 Virginia ... etc.

I want to do a distributed sort on this using mapreduce. I know mapreduce sorts between the map and reduces stages, so I just want to emit from map, then from reduce without processing, but it is not working. Here is my map and reduce function:

```
public static class SortMap extends Mapper<Text, Text, Text, Text>
{
    public void map(Text key, Text value, Context context) throws IOException, InterruptedException
    {
        context.write(key, value);
    }
}
public static class SortReduce extends Reducer<Text, Text, Text, Text>
{
```

Here is my main class

```
public static void main(String[] args) throws Exception
{
    Configuration conf = new Configuration();

    Job sort = new Job(conf, "sort");
    sort.setMapperClass(SortMap.class);
    sort.setCombinerClass(SortReduce.class);
    sort.setReducerClass(SortReduce.class);
    sort.setInputFormatClass(CountInputFormat.class);
```

And here is the inputformat class i wrote since FileInputFormat would always fail


```
class CountInputFormat extends FileInputFormat<Text, Text>
{
    public RecordReader<Text, Text> createRecordReader(InputSplit is, TaskAttemptContext tac) throws IOException,
    InterruptedException
    {
        CountReader cr = new CountReader();
        cr.initialize(is, tac);
        return cr;
    }
}
```

Here is the error

```
Feb 24, 2010 2:48:40 PM org.apache.hadoop.metrics.jvm.JvmMetrics init
INFO: Initializing JVM Metrics with processName=JobTracker, sessionId=
Feb 24, 2010 2:48:40 PM org.apache.hadoop.metrics.jvm.JvmMetrics init
INFO: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Feb 24, 2010 2:48:40 PM org.apache.hadoop.mapred.JobClient configureCommandLineOptions
WARNING: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
Feb 24, 2010 2:48:40 PM org.apache.hadoop.mapred.JobClient configureCommandLineOptions
WARNING: No job jar file set. User classes may not be found. See JobConf(Class) or JobConf#setJar(String).
Feb 24, 2010 2:48:40 PM org.apache.hadoop.mapreduce.lib.input.FileInputFormat listStatus
INFO: Total input paths to process : 1
```

Thanks

**Larry Homes**  
Greenhorn

posted Feb 28, 2010 14:27:26  PM

 Quote

Joined: Jan 18, 2009  
Posts: 25

Thought I would post the solution I found. It was an incredibly dumb error on my part. In my main class, I named the Job instance sort, but then when setting the mapOutputKey, mapOutputValues, outputKey and outputValue, I use the identifier job. That identifier was from a previous mapreduce in the chain and I had just copied and pasted the code without remembering to change the job identifier.



subject: Hadoop key mismatch

All times are in JavaRanch time: GMT-6 in summer, GMT-7 in winter



[Contact Us](#) | Powered by [JForum](#) | Copyright © 1998-2011 [Paul Wheaton](#)