

# JobConfFile

## Intro

This is a job configuration with defaults. Only the properties that are specific to a job need to be included in the job configuration file.

## Sample job config file

```
<configuration>

<property>
<name>mapred.job.name</name>
<value></value>
<description>The name of the job.</description>
</property>

<property>
<name>mapred.mapper.class</name>
<value>org.apache.hadoop.mapred.lib.IdentityMapper</value>
<description>The full class name of the mapper.</description>
</property>

<property>
<name>mapred.combiner.class</name>
<value></value>
<description>The full class name of the combiner.</description>
</property>

<property>
<name>mapred.reducer.class</name>
<value>org.apache.hadoop.mapred.lib.IdentityReducer</value>
<description>The full class name of the reducer.</description>
</property>

<property>
<name>mapred.jar</name>
<value>No default.</value>
<description>The full path to the jarfile containing all the needed classes.</description>
</property>

<property>
<name>mapred.map.tasks</name>
<value>1</value>
<description>The default number of map tasks per job. Typically set to a prime several times greater than
number of available hosts. Ignored when mapred.job.tracker is "local". </description>
</property>

<property>
<name>mapred.reduce.tasks</name>
<value>1</value>
<description>The default number of reduce tasks per job. Typically set to a prime close to the number of
available hosts. Ignored when mapred.job.tracker is "local". </description>
</property>

<property>
<name>mapred.input.dir</name>
<value></value>
<description>A comma separated list of input directories.</description>
</property>

<property>
<name>mapred.output.dir</name>
<value></value>
<description>A comma separated list of output directories.</description>
</property>

<property>
```

```
<name>mapred.input.format.class</name>
<value>org.apache.hadoop.mapred.TextInputFormat</value>
<description>The full class name of the InputFormat class to be used for obtaining the input to the mapper.<
/description>
</property>

<property>
<name>mapred.output.format.class</name>
<value>org.apache.hadoop.mapred.TextOutputFormat</value>
<description>The full class name of the OutputFormat class to be used for saving the output of the reducer.<
/description>
</property>

<property>
<name>mapred.input.key.class</name>
<value>org.apache.hadoop.io.LongWritable</value>
<description>The full classname of the input key.</description>
</property>

<property>
<name>mapred.input.value.class</name>
<value>org.apache.hadoop.io.UTF8</value>
<description>The full classname of the input value.</description>
</property>

<property>
<name>mapred.output.key.class</name>
<value>org.apache.hadoop.io.LongWritable</value>
<description>The full classname of the output key.</description>
</property>

<property>
<name>mapred.output.value.class</name>
<value>org.apache.hadoop.io.UTF8</value>
<description>The full classname of the output value.</description>
</property>

<property>
<name>mapred.partitioner.class</name>
<value>org.apache.hadoop.mapred.lib.HashPartitioner</value>
<description>The full classname of the partitioner class.</description>
</property>

<property>
<name>user.name</name>
<value>Dr. Who</value>
<description>The name of the user running the job.</description>
</property>

<property>
<name>mapred.combine.buffer.size</name>
<value>100000</value>
<description>The number of entries the combining collector caches before combining them and writing to disk.<
/description>
</property>

<property>
<name>mapred.speculative.execution</name>
<value>true</value>
<description>If true, then multiple instances of some map tasks may be executed in parallel.</description>
</property>

<property>
<name>mapred.min.split.size</name>
<value>0</value>
<description>The minimum size chunk that map input should be split into. Note that some file formats may have
minimum split sizes that take priority over this setting.</description>
</property>

</configuration>
```

