

# Hadoop 作业管理

上海交通大学 OmniLab

<http://omnilab.sjtu.edu.cn>

2013 年 11 月 27 日更新

## 目录

<b>1 Hadoop 作业提交的一般形式</b>	<b>2</b>
<b>2 TODO</b>	<b>2</b>
<b>3 参考资料</b>	<b>2</b>

## 1 Hadoop 作业提交的一般形式

hadoop命令的一般形式:

```
$ hadoop -h
Usage: hadoop [--config confdir] COMMAND
      where COMMAND is one of:
      fs                run a generic filesystem user client
      version           print the version
      jar <jar>         run a jar file
      checknative [-a|-h] check native hadoop and compression libraries availability
      distcp <srcurl> <desturl> copy file or directories recursively
      archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
      classpath         prints the class path needed to get the
                        Hadoop jar and the required libraries
      daemonlog         get/set the log level for each daemon
or
      CLASSNAME         run the class named CLASSNAME

Most commands print help when invoked w/o parameters.
```

提交作业使用的主要子命令是`jar`，其主要附加参数如下：

- `-files`
- `-libjars`
- `-archives`

## 2 TODO

- 改进关于`hadoop`命令的说明，增加 MapReduce 作业运行过程的说明。
- 在命令行，而不是代码中指定这些参数：作业名、并行 Mapper/Reducer 数量；

## 3 参考资料

- Apache Hadoop Commands Guide [http://hadoop.apache.org/docs/r1.2.1/commands\\_manual.html](http://hadoop.apache.org/docs/r1.2.1/commands_manual.html)