Hadoop是一个开源的框架，可编写和运行分布式应用处理大规模数据。

mrjob is the easiest route to writing Python programs that run on Hadoop. If you use mrjob, you’ll be able to test your code locally without installing Hadoop or run it on a cluster of your choice.