**Hadoop Notes**

May 4th

1. VirtualBox/VMware
2. When connected wirelessly, use Network -> Host only, or master machine cannot connect to the virtual machine
3. <http://www.powerxing.com/install-hadoop/>

May 5th

1. 关于mapper.py reducer.py输入的理解

mapper.py的输入是文件每一行的读取结果，按照streaming形式

reducer.py的输入： mapper的输出，shuffle到reducer，在reducer本地sort之后，按照streaming的形式输入

而输入和输出，不需要我们编程控制？。。。