## 一、前提条件

首先本地需要安装好 jdk 、 maven 、 python的基础环境，如果安装过程有问题可以参考百度对应的教程，这里就不多说了。

下面是我的安装环境：



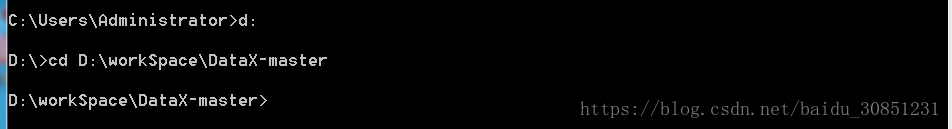


è¿éåå¾çæè¿°

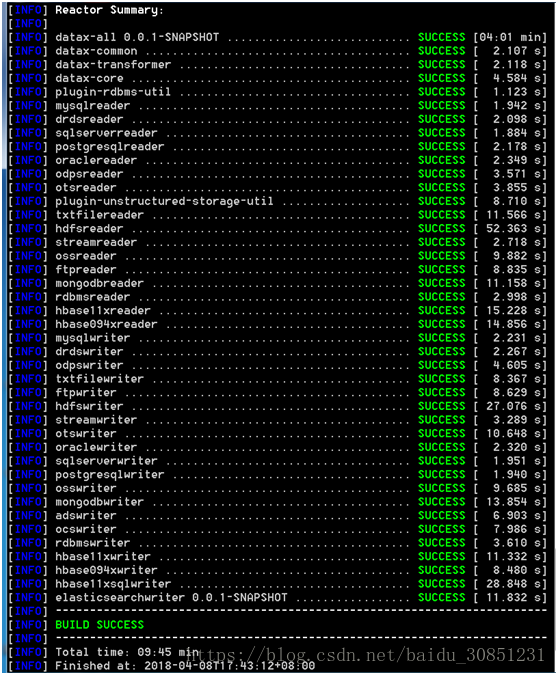
## 二、源码安装

1、DataX源码下载https://github.com/alibaba/DataX/tree/master路径，下载压缩包解压到本地目录。

2、进入命令行模式（cmd），切换到对应的源码路径（我的是 “D:\workSpace\DataX-master”）

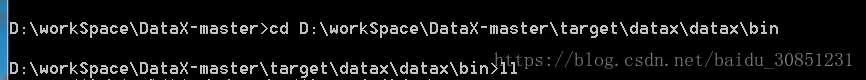


3、输入 “ mvn -U clean package assembly:assembly -Dmaven.test.skip=true”命令进行maven 打包，最终的打包效果如下图：



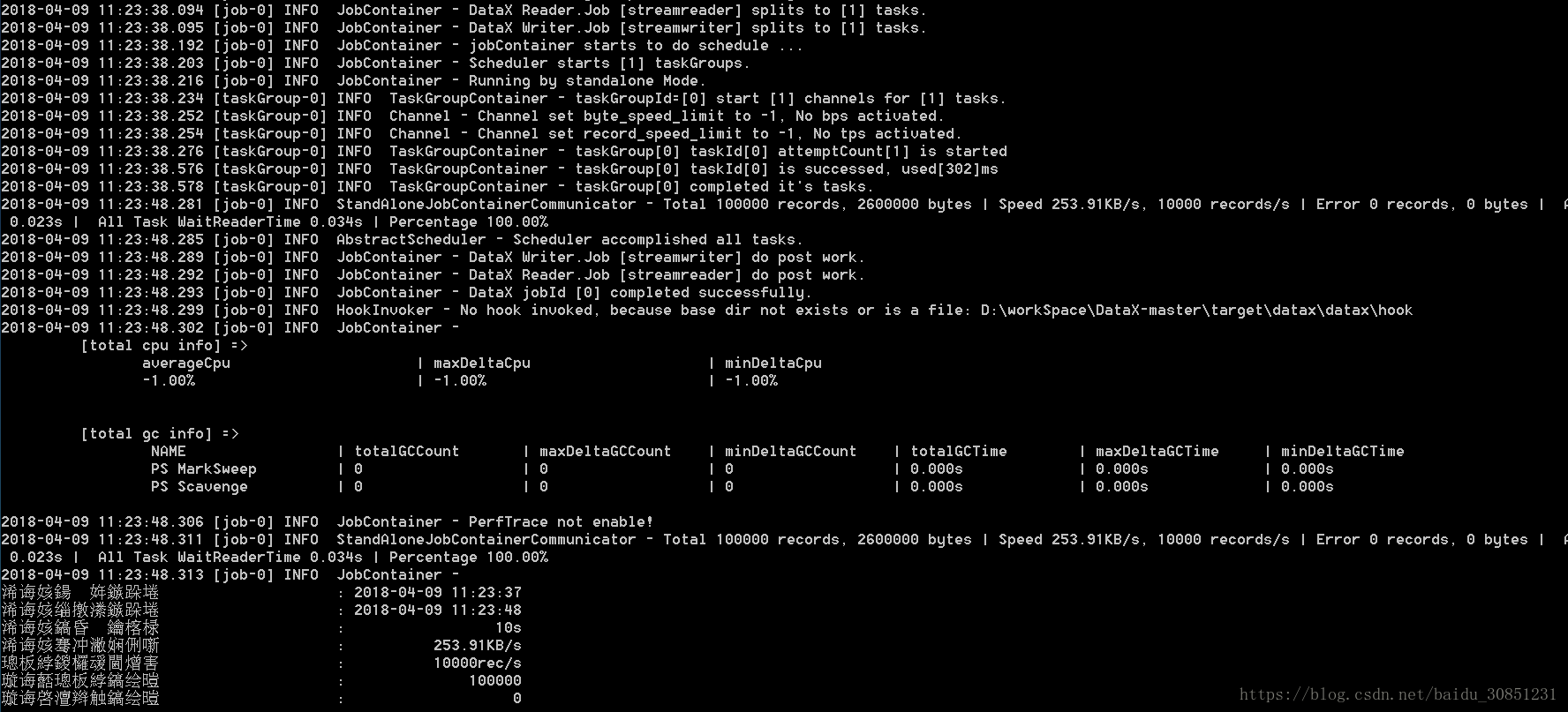
4、验证DataX是否成功

打包成功后会在源目录中多一个target 目录，切换到对应打包后的bin路径



输入命令 ： “python datax.py ../job/job.json” 最终输出结果如下图，则证明项目运行正常：





如果乱码，可以在cmd中输入：

CHCP 65001

同时，github也提供了快速使用DataX的方法，有打包好的datax.tar.gz ,可以下载解压后直接使用

http://datax-opensource.oss-cn-hangzhou.aliyuncs.com/datax.tar.gz

---------------------

作者：Session-在水一方

来源：CSDN

原文：https://blog.csdn.net/baidu\_30851231/article/details/79863779

版权声明：本文为博主原创文章，转载请附上博文链接！

## 三、配置测试样例

下面我们配置一组 从mysql数据库到另一个mysql数据库

**第一步、创建作业的配置文件（json格式）**   
可以通过命令查看配置模板： python datax.py -r {YOUR\_READER} -w {YOUR\_WRITER}

****

**第二步、根据配置文件模板填写相关选项**   
命令打印里面包含对应reader、writer的文档地址，以及配置json样例，根据json样例填空完成配置即可。根据模板配置json文件(mysqlTomysql.json)如下



**第三步：启动DataX**

[root@slave1 bin]# python datax.py ../mysql2mysql.json

同步结束，显示日志如下：

任务启动时刻 : 2016-12-10 07:47:16

任务结束时刻 : 2016-12-10 07:47:27

任务总计耗时 : 11s

任务平均流量 : 320B/s

记录写入速度 : 2rec/s

读出记录总数 : 25

读写失败总数 : 0

## 四、job的配置

（1）、Job基本配置

Job基本配置定义了一个Job基础的、框架级别的配置信息，包括：



（2） Job Setting配置



● job.setting.speed(流量控制)

Job支持用户对速度的自定义控制，channel的值可以控制同步时的并发数，byte的值可以控制同步时的速度

● job.setting.errorLimit(脏数据控制)

Job支持用户对于脏数据的自定义监控和告警，包括对脏数据最大记录数阈值（record值）或者脏数据占比阈值（percentage值），当Job传输过程出现的脏数据大于用户指定的数量/百分比，DataX Job报错退出。

---------------------

作者：Xlucas

来源：CSDN

原文：https://blog.csdn.net/paicMis/article/details/53563824

版权声明：本文为博主原创文章，转载请附上博文链接！