1.执行MR的命令：

hadoop jar <jar在linux的路径> <main方法所在的类的全类名> <参数>

例子：

hadoop jar /root/wc1.jar cn.itcast.d3.hadoop.mr.WordCount hdfs://itcast:9000/words /out2

2.MR执行流程

(1).客户端提交一个mr的jar包给JobClient(提交方式：hadoop jar ...)

(2).JobClient通过RPC和JobTracker进行通信，返回一个存放jar包的地址（HDFS）和jobId

(3).client将jar包写入到HDFS当中(path = hdfs上的地址 + jobId)

(4).开始提交任务(任务的描述信息，不是jar, 包括jobid，jar存放的位置，配置信息等等)

(5).JobTracker进行初始化任务

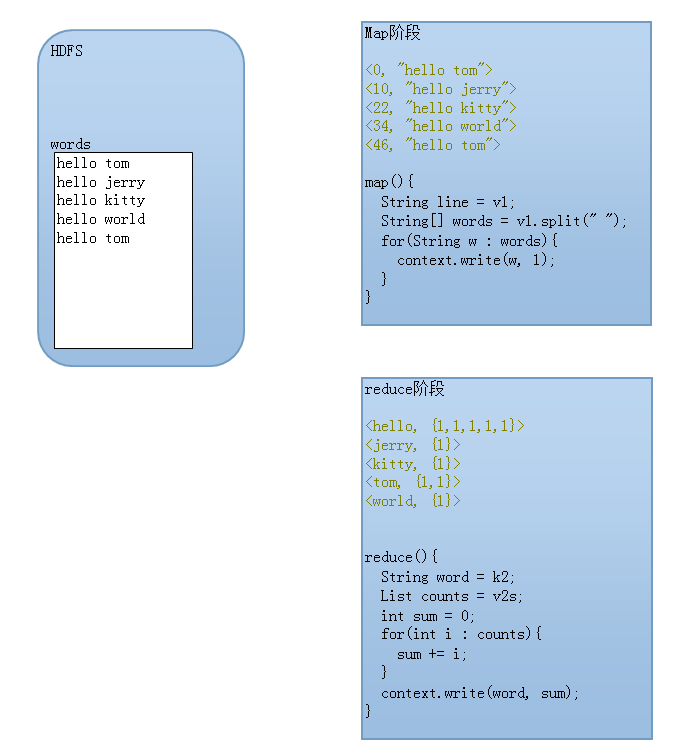
(6).读取HDFS上的要处理的文件，开始计算输入分片，每一个分片对应一个MapperTask

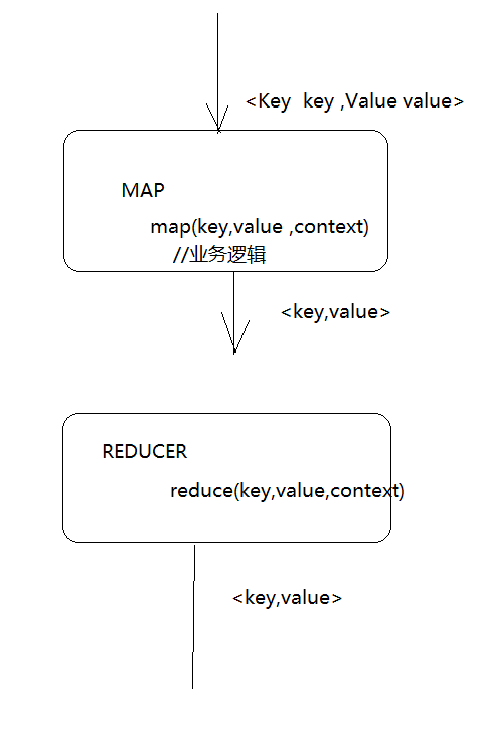
(7).TaskTracker通过心跳机制领取任务（任务的描述信息）

(8).下载所需的jar，配置文件等

(9).TaskTracker启动一个java child子进程，用来执行具体的任务（MapperTask或ReducerTask）

(10).将结果写入到HDFS当中





3.MR程序的本地运行模式 //注意：自己运行会报错，需要改配置。3-05最后。Winutils

Could not locate executable null\bin\winutils.exe in the Hadoop binaries.

4.MR程序的集中提交运行模式

本地模型运行

1）在windows的eclipse里面直接运行main方法，就会将job提交给本地执行器localjobrunner执行

----输入输出数据可以放在本地路径下（c:/wc/srcdata/）

----输入输出数据也可以放在hdfs中(hdfs://weekend110:9000/wc/srcdata)

2）在linux的eclipse里面直接运行main方法，但是不要添加yarn相关的配置，也会提交给localjobrunner执行

----输入输出数据可以放在本地路径下（/home/hadoop/wc/srcdata/）

----输入输出数据也可以放在hdfs中(hdfs://weekend110:9000/wc/srcdata)

集群模式运行0

1）将工程打成jar包，上传到服务器，然后用hadoop命令提交 hadoop jar wc.jar cn.itcast.hadoop.mr.wordcount.WCRunner

2）在linux的eclipse中直接运行main方法，也可以提交到集群中去运行，但是，必须采取以下措施：

----在工程src目录下加入 mapred-site.xml 和 yarn-site.xml

----将工程打成jar包(wc.jar)，同时在main方法中添加一个conf的配置参数　conf.set("mapreduce.job.jar","wc.jar");

3）在windows的eclipse中直接运行main方法，也可以提交给集群中运行，但是因为平台不兼容，需要做很多的设置修改

----要在windows中存放一份hadoop的安装包（解压好的）

----要将其中的lib和bin目录替换成根据你的windows版本重新编译出的文件

----再要配置系统环境变量 HADOOP\_HOME 和 PATH

----修改YarnRunner这个类的源码

5.序列化/反序列化

Hadoop序列化机制不会传递对象的继承结构信息。

Write/readFields

6.多个job在一个main方法中提交。一般分开写。