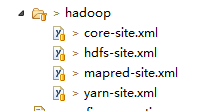
分析平台相关文档：

1. 配置：
   1. 项目中resources下的/hadoop目录的配置文件是关于hadoop的配置文件，其中的配置文件内容和linux系统下安装的hadoop中配置文件保持一致就行。更改namenode，datanode的地址。更改mapper的地址等。
   2. 在/hadoop外面的配置文件中hbase-site.xml与安装hbase中的配置文件保持一致。Hadoop.priperties.xml中配置文件系统hdfs的路径和端口：和mapreduce程序任务执行的地址：而在config.properties中主要是配置远程调用的RMI url 和端口的配置。这里的url 和端口要和web 平台的保持一致才能完成远程的调用。

DataServer中配置RMI：

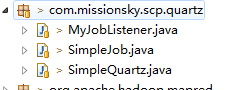


Web平台配置RMI：



* 1. Jar包：各个项目中关于hadoop和hbase的jar包。必须和linux集群上安装的hadoop还有hbase的版本保持一致。如果没有保持一致的话，会导致程序出错没有办法正确的连接到hadoop和hbase。
  2. 更改算法路径：



1. 运行过程:
2. 在web平台对任务进行订制过后（RMI要已经可以进行正常的通信：分析平台先绑定RMI，web平台在进行连接）。
3. 分析平台定时框架：去判断web平台的任务的执行的类型（runOneTime，runDaily，runWeek）。当web平台点击run后任务进入redeay状态。等待执行时间的到来。选择执行的类型，调用simplejob类：

，

进入到任务装配线（BasicTaskAssemblyLine）：对任务进行装配。

1. 在装配中先对/hadoop目录下的配置文件进行加载：



1. 取得标准文件中的内容content：



1. 确定算法执行后输出文件的路径：



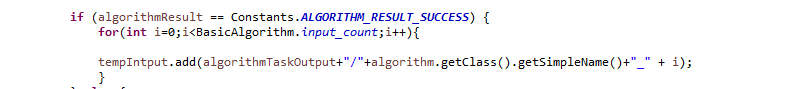
1. 执行算法：



在执行算法的过程中调用了另外的一个算法的run方法。



1. 判断算法执行的结果：修改下次算法执行时候的输入路径，避免重名的出现：



1. 将数据往hbase中存放：（StoreToHBaseAlgorithm）



通过工具类在hbase中建立表，还有族。

新建一个往hbase中存放的算法。

算法执行



在run中通过mapreduce进行数据往hbase的存放。

1. RMI远程调用机制：
2. 在web平台中通过DataClient 类中进行与分析平台的通信。（用RMI对象调用datautil的方法）：



1. 分析平台中调用IDataUtil的接口：



然后DataUtil继承IDataUtil去调用datautil中的方法进行算法和hbase中数据的操作。

比如：

