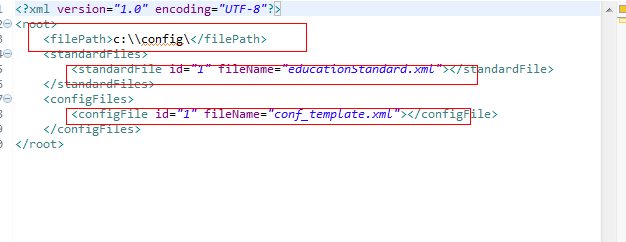
数据适配器项目的配置，部署和运行步骤：

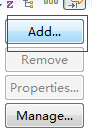
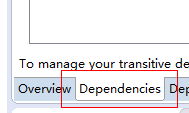
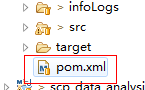
1. 配置文件.
2. 修改hdfs-site.xml，core-site.xml：这2个配置文件需要与hadoop集群上的hdfs-site.xml，core-site.xml文件中的值一致。Hadoop集群上配置文件的位置在/user/hadoop/etc/hadoop
3. 修改config.properties.xml文件，将其中的改为当前你的config.xml文件的路径。
4. 修改config.xml。修改配置文件路径filePath ，修改写入hdfs的配置文件standardFile，修改数据源的配置文件configFile。
5. 在配置文件中的各个数据的名称要和DataDetcherDispernser中的一致：

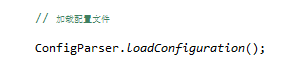


1. 升级项目中的jar包。

安装hadoop的时候选用的是hadoop的最新版本。

程序的hadoop相关jar包于hadoop安装中的版本jar包不匹配报错。

1. 然后通过maven升级hadoop-client，hadoop-hdfs。
2. 运行过程：



Main



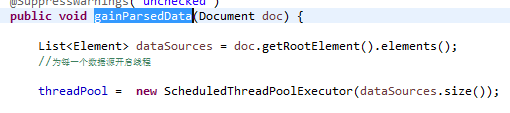


通过数据源中的format进行适配器选择

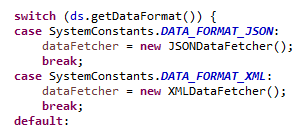
加载配置文件并为单例模式赋值（ConfigParser）

对数据源配置文件解析（XmlUtil）



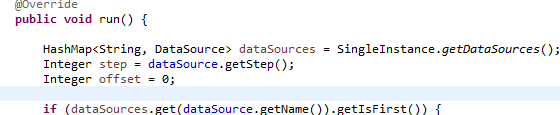


将解析后的doc转发（gainParsedData）



周期执行执行线程（传入适配器类型，相应数据源）



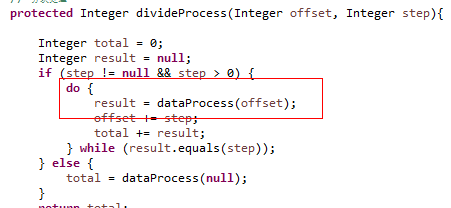


将数据进行分块divideprocess

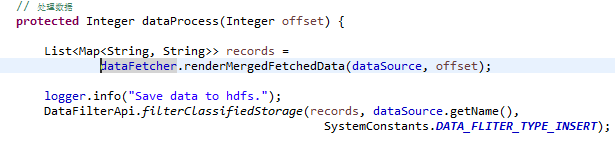
ExcuteFecthThread

调用线程中的run（）ExcuteFecthThread



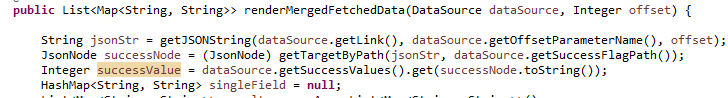


对数据进行处理dataprocess ExcuteFecthThread

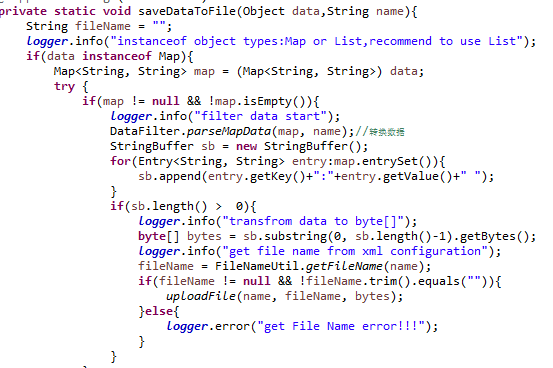


通过各类适配器读取数据

XXXDataFetcher



向hadoop中的hdfs中写入文件DataFilterApi



运行过程说明：

1. 在对适配器进行选择的时候是通过数据源中的配置文件进行选择的。
2. 写入hdfs中的标准配置文件中的标准文件名和配置文件名一致。避免web平台在加入配置文件后运行导致文件名不一致，在hdfs中找不到。
3. 在往hdfs中写入数据的时候会先判断文件是否存在，如果存在的话。会先删除，在往里面写入数据。