|  |  |  |  |
| --- | --- | --- | --- |
| **中科院软件所**  **软件工程中心** | **文 档 编 号** | **产品版本** | **密级** |
| XK-DN-2000-10-11-16 | **V 1.0** | 内部 |
| **产品名称： Haflow** | | **共 页** |

**(仅供内部使用)**

**文 档 作 者： \_\_\_\_陈茜\_\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_/\_\_\_/\_\_\_**

**说 明 书 校 对： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_/\_\_\_/\_\_\_**

**产 品 经 理： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_/\_\_\_/\_\_\_**

**C:\Users\dawn\Desktop\logo.png**

**版权所有 不得复制**

文件修改记录

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 修改日期 | 版本 | 修改页码、章节、条款 | 修改描述 | 作者 |
| 2013-12-2 | 1 |  | 创建文档 | 陈茜 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**目录**

[文件修改记录 2](#_Toc373756217)

[1 1引言 4](#_Toc373756218)

[1 .1 编写目的 4](#_Toc373756219)

[2 组件开发步骤 4](#_Toc373756220)

[2 .1 环境配置 4](#_Toc373756221)

[2 .2 组件接口定义 5](#_Toc373756222)

[2 .3 组件实现 7](#_Toc373756223)

[3 组件打包 8](#_Toc373756224)

## 1引言

### 编写目的

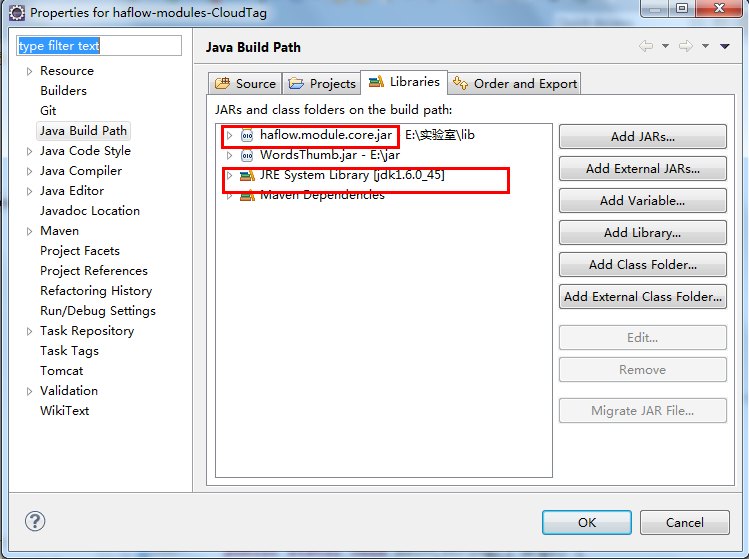
编写本说明的目的是帮助组件开发者更好的开发组件。

## 组件开发步骤

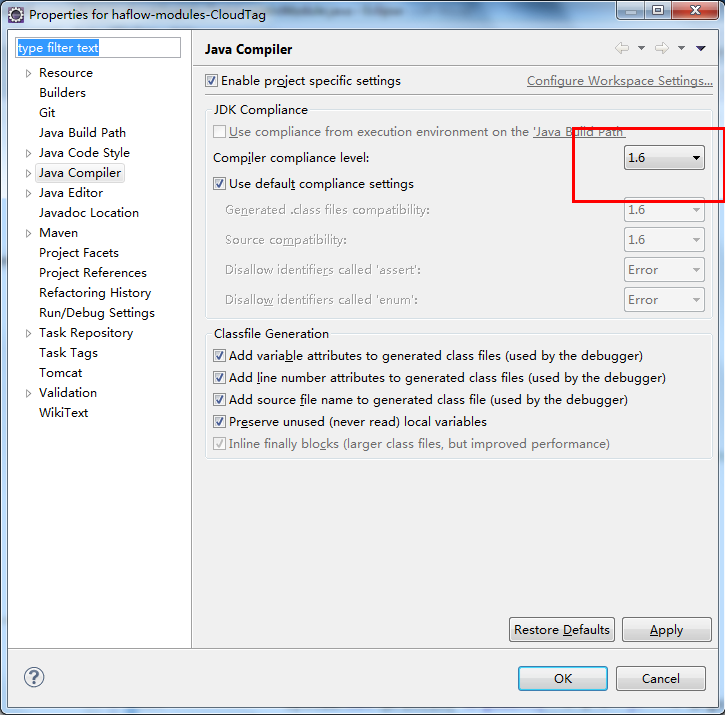
### 环境配置

支持用java, R, python 等可以运行在jvm上的语言编写的组件，这里只讲述如何开发java组件。

* 在eclipse里新建一个maven项目；
* 引入外部的jar包（haflow.module.core.jar）（如图1）；
* JRE System Library设置为jdk1.6版本（如图1）；
* JRE Compiler同样设置为jdk1.6（如图2）。



图片 1



图片 2

### 组件接口定义

#### 包结构

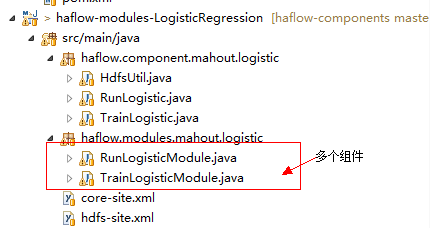
组件实现包的组织接口定义成下图：



组件应包含组件定义package：haflow.module.XXX.

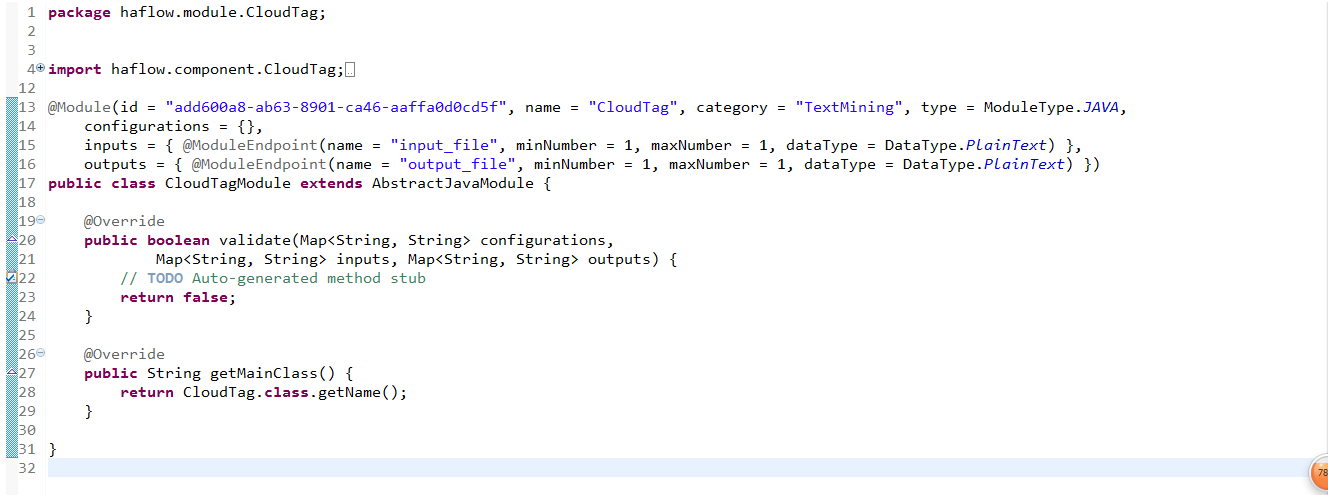
该package中定义要展现在界面上的module类，注意：有几个组件写几个module类。

如下图:



#### 组件接口定义

例如：对于一个只有输入输出的组件，module类代码如下：



* 第4行： 引入实现组件功能的类；
* 第13-16行：注解该组件的配置参数。一个module包含以下参数：
  + Id：每个组建一个，不重复，16进制数表示。
  + Name：显示名称。
  + Category：类别。
  + Type：编写语言。
  + Configurations：组件运行参数。
  + Inputs：输入节点。
    - 输入节点可以是数据节点(文件),也可以是上个组件产生的输出;
    - 输入节点的name字段值要跟程序中主函数命令行参数中的某个key对应.
    - 输入节点的dataType可以是*PlainText*, *Csv*, *Arff*, *Image*, *Jpg*, *Model*, *ALL*, *AUTO*其中之一.
  + Outputs：输出节点。
* 第17行：继承AbstractJavaModule。
* 第18行到结束：重写其中的validate和getMainClass方法。
  + Validate：对特殊的参数做验证。
  + GetMainClass：【实现功能的类】.class.getName().

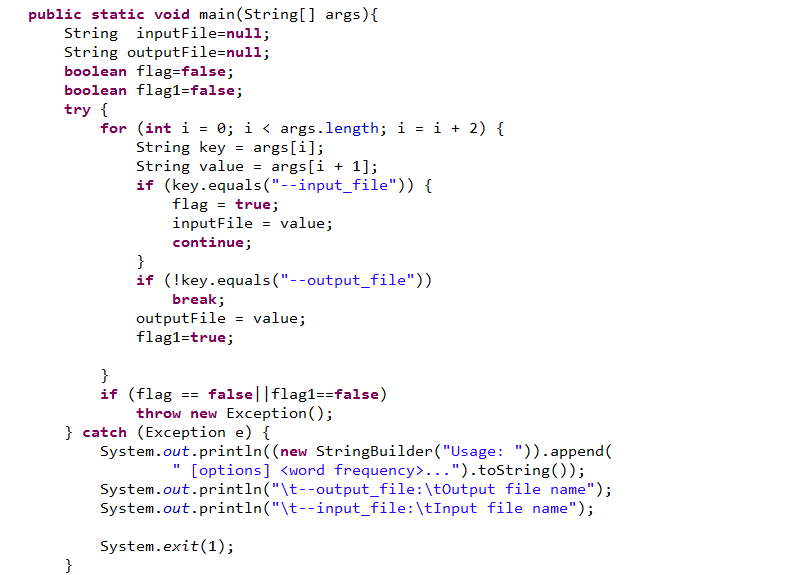
### 组件实现

可以采用map-reduce 结构，可以用R，python实现。唯一要求是，文件的读写应该在hdfs上，即输入输出都是hdfs的文件。

Component包中组件的定义需要满足如下规范:

*组件运行需要用户设定的参数,来自于main函数的命令行参数，是一些<key,value>值对。其中key与module模块中的定义的相同。*

例如：与上文中module对应的组件实现类，其参数部分代码如下：



* Main函数的命令行参数是module类中输入节点的name加上“—”。

## 组件打包

编写好的组件要打成jar包，并将引用的外部jar包或者文本文件放在jar包里。Jar包不能是嵌套jar包结构，要将引用的外部jar（如果引用apache等机构的jar可不打进包里），以文件目录的形式打到包中。

例如：CloudTag组件引用了一个外部的jar包，WordsThumb.jar,要将其以目录结构的形式打进CloudTag包中。

