**风控系统批处理开发规范**

目 录

[1. 数据文件名命规范 1](#_Toc9881)

[1.1. 源文件名命 1](#_Toc15063)

[2. HDFS文件系统使用规范 1](#_Toc12505)

[2.1. 源文件目录 1](#_Toc26579)

[2.2. 贴源表分区目录 1](#_Toc9300)

[3. 数据表名命规范 1](#_Toc1331)

[3.1. 贴源表名命 1](#_Toc12125)

[3.2. 中间表名命 1](#_Toc23230)

[4. Kettle文件和SQL文件名命规范 2](#_Toc27250)

[4.1. Kettle文件名命 2](#_Toc6011)

[4.2. sql文件名命 2](#_Toc1759)

[5. 定时任务配置规范 2](#_Toc29800)

[5.1. 名称 2](#_Toc23324)

[5.2. 类名 2](#_Toc12741)

[5.3. 参数 3](#_Toc877)

版本信息

|  |  |  |  |
| --- | --- | --- | --- |
| **版本号** | **修订日期** | **修订内容** | **编辑人** |
| V1.0 | 2019.11.21 | 新建 | 李德和 |
| V2.0 | 2020.01.10 | 细化名命规则 | 李德和 |
|  |  |  |  |
|  |  |  |  |

# 数据文件名命规范

## 源文件名命

源文件指待处理的文件，可以由外部系统提供，也可由风控系统生成。一般是从源表中导出的数据文件。

源文件名命按照如下格式：

表名\_数据日期.格式后缀

例如：custinfo\_20191114.txt

# HDFS文件系统使用规范

## 源文件目录

源文件均存放在/rawData/ 目录下

## 贴源表分区目录

如果源文件是从某个表中导出的，则数据表的分区目录为：

/rrms/tableName/dataDate/

例如：假设表名字为custinfo，导出的数据是20191114的，那么贴源表分区目录为：

/rrms/custinfo/20191114/

# 数据表名命规范

## 贴源表名命

如果源文件是从数据表导出的，那么在保证不与已有数据表名冲突的情况下，应当与贴源表名命与源表明相同。如custinfo\_20191114.txt对应的贴源表名字应当为custinfo。

## 中间表名命

在贴源表名字的基础上，添加“\_逻辑标记”，例如保存统计custinfo表的记录总数的表可以名命为custinfo\_count。

如果是多个贴源表关联查询，推荐将关联的表名缩写在中间表名中。

# Kettle文件和SQL文件名命规范

## Kettle文件名命

kjb文件和ktr文件的名字要一致

一次性作业的kettle文件名字末尾为Init，增量作业的kettle文件名字末尾为Main

如果涉及到一个kjb调用多个ktr，则按照调用顺序在ktr文件名字最后再添加阿拉伯数字，从1开始。

一次性作业名命案例：

testInit.kjb

testInit1.ktr

testInit2.ktr

增量作业名命案例：

testMain.kjb

testMain1.ktr

testMain2.ktr

## sql文件名命

如果作业用于建表，那么sql文件应当名命为create.sql。否则按需求自行名命，与kettle文件中的脚本调用的sql文件名保持一致即可。

# 定时任务配置规范

## 名称

任务名称描述任务的功能，必须是唯一的。

## 类名

使用开发好的通用类即可：org.jeecg.modules.quartz.job.BatchJob

如果这个类不满足用户需求，请参考jeecg文档开发新的类。

## 参数

定时任务的参数为在可视化界面编辑好的kettle文件完整文件名，如xxx.kbj。需要注意的是，填写的是kjb文件，不是ktr文件。

# 结果文件存储规范

结果文件指从hdfs上取回服务器本地待后续使用的文件，存放在用户目录下的resultFile目录下。