## Kettle

### 介绍

Kettle是一款国外开源的ETL工具，纯java编写，可以在Windows、Linux、Unix上运行，[数据抽取](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%8A%BD%E5%8F%96/218221)高效稳定。Kettle 中文名称叫水壶，该项目的主程序员MATT 希望把各种数据放到一个壶里，然后以一种指定的格式流出。

Kettle中有两种[脚本文件](https://baike.baidu.com/item/%E8%84%9A%E6%9C%AC%E6%96%87%E4%BB%B6)，transformation和[job](https://baike.baidu.com/item/job)，transformation完成针对数据的基础转换，job则完成整个工作流的控制。

作为[Pentaho](https://baike.baidu.com/item/Pentaho)的一个重要组成部分，现在在国内项目应用上逐渐增多。

Kettle家族目前包括4个产品：Spoon、Pan、CHEF、Kitchen。

**SPOON**允许你通过图形界面来设计ETL转换过程（Transformation）。

**PAN**允许你批量运行由Spoon设计的ETL转换 (例如使用一个时间调度器)。Pan是一个后台执行的程序，没有图形界面。

**CHEF**允许你创建任务（Job）。 任务通过允许每个转换，任务，脚本等等，更有利于自动化更新数据仓库的复杂工作。任务通过允许每个转换，任务，脚本等等。任务将会被检查，看看是否正确地运行了。

**KITCHEN**允许你批量使用由Chef设计的任务 (例如使用一个时间调度器)。KITCHEN也是一个后台运行的程序

### 工作如何涉及到kettle的

开发智能数据治理平台数据接入模块，涉及到的内容。

**部署** 安装kettle应用到一台linux服务器上面，并且设置好用户和密码；

项目中到kettle的应用包里面获取到kettle核心的jar包（//todo）；

**模块功能** 指定元数据库表信息（包含目标表和源表），同时指定两张表字段之间的映射关系。

任务调度参数是由**AZKABAN**控制的，这个做另外一个技术文档进行总结。

按照任务设置好的参数根据设计好的ktr步骤生成一个任务特有的ktr脚本。

主要代码可参考com.sailing.dmp.dataaccess包下数据接入模块相关功能

DaDfJobController，DaDfJobTaskController

其中dataaccess包下transfer包里面的是kettle api使用的主要封装