

ETL 项目说明书

1. 硬件需求：

机器数量	CPU	RAM	HD	Network Bandwidth
1	4Core	8GB	100GB	1MB

2. 软件环境

2.1 JDK1.8

2.2 Kettle7.1

2.3 MongoDB

2.4 Mysql5.6

3. 模块介绍

3.1 抽取模块

抽取现有项目 MongoDB 中指定的表满足条件的数据。

3.2 转换模块

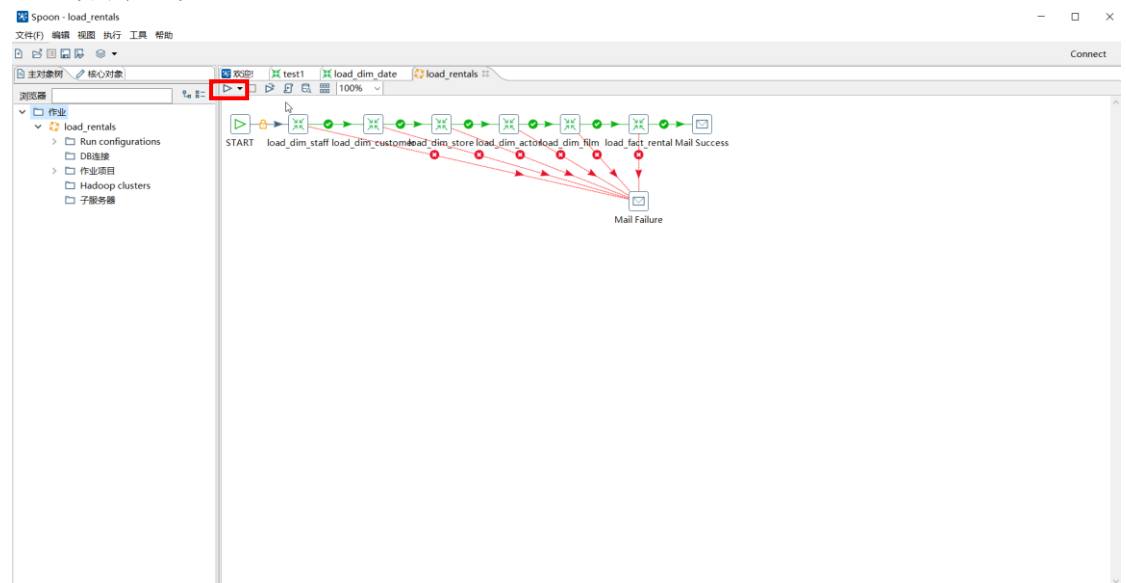
将步骤 3.1 中抽取的数据转换为可以插入 Mysql 中临时表的临时存储结构中。

3.3 加载模块

将步骤 3.2 中临时存储结构中的数据加载到 Mysql 目标表中，更新或插入。

4. 启动步骤

4.1 界面启动,点击界面三角按钮



4.2 命令行启动

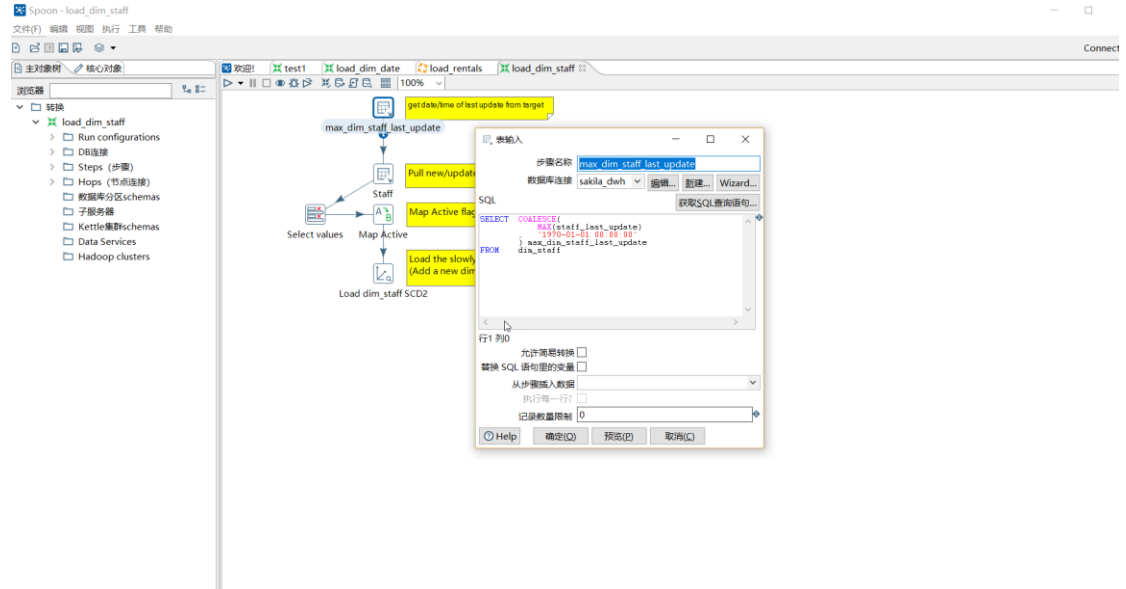
4.2.1 调用脚本 ETL_Sch_day.bat \$last_extract_date

5. 参数说明

5.1 全量抽取不需要参数

5.2 增量数据抽取设置

5.2.1 界面参数设置



5.2.2 命令行参数设置

上次抽取日期\$last_extract_date

6. 故障排除

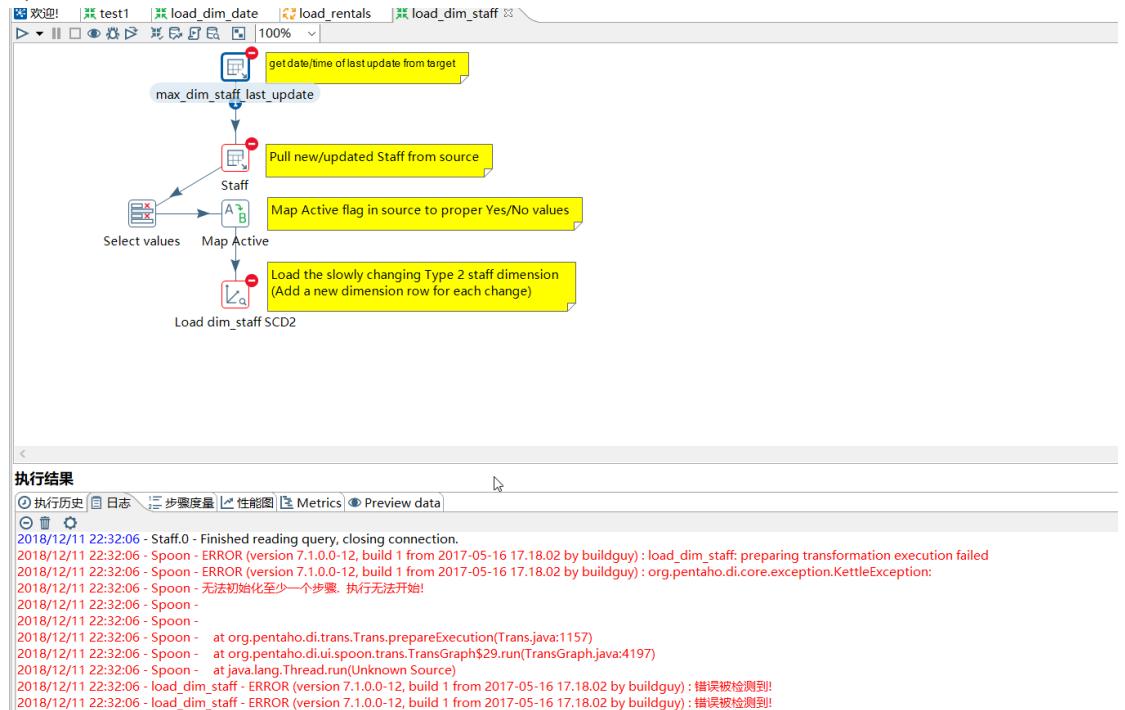
6.1 MongoDB 连接故障

6.2 Mysql 连接故障

6.3 邮件通知故障

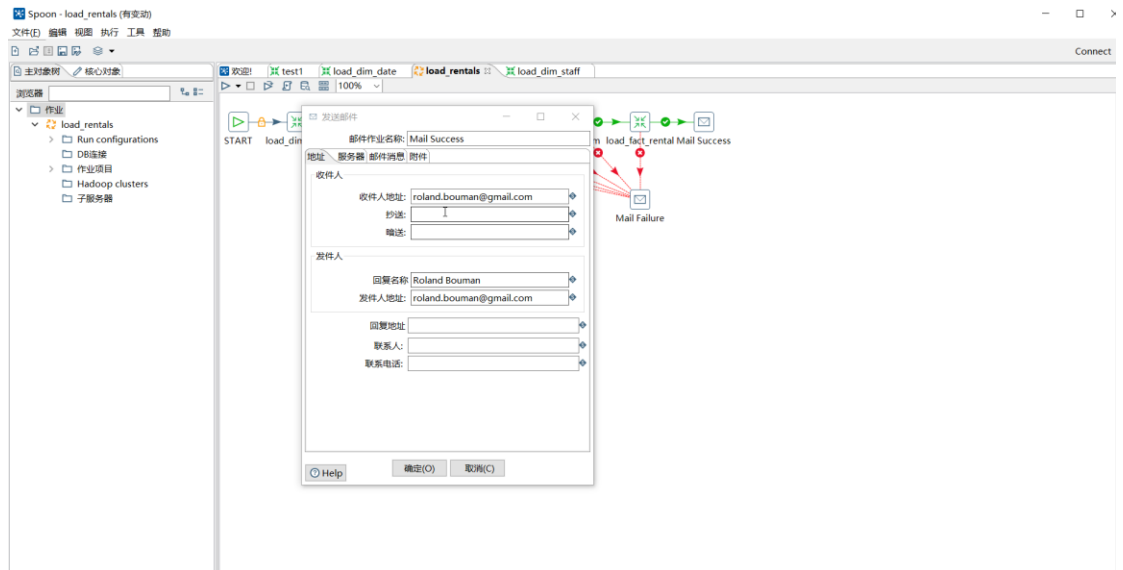
7. 监控方式及执行状态

7.1 Spoon 界面工具



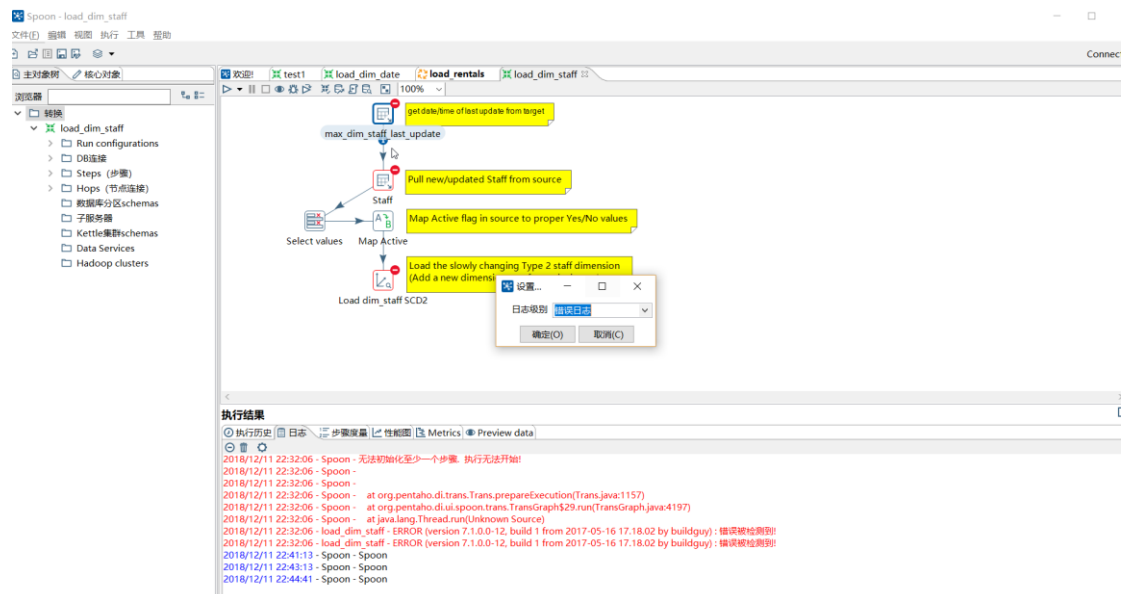
7.2 脚本执行 log 文件 data-integration\logs\spoon.log

7.3 执行成功邮件通知

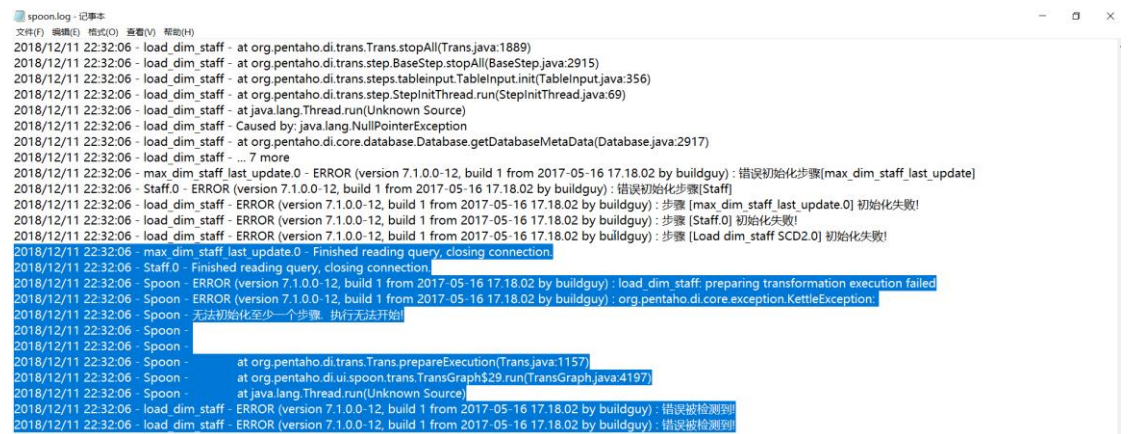


8. 异常记录

8.1 界面错误日志



8.2 错误日志 data-integration\logs\spoon.log



9. ETL 测试步骤
 - 9.1 数据质量屏幕设计
 - 9.2 数据质量屏幕执行
10. 程序中断后恢复
 - 10.1 全部数据重新刷新
 - 10.2 增量数据局部刷新
11. 继续由中断处续写
支持单独执行某以转换，比如整个 Job 包含