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 |

1. 软件环境
   1. JDK1.8
   2. Kettle7.1
   3. MongDB
   4. Mysql5.6
2. 模块介绍
   1. 抽取模块

抽取现有项目MongDB中指定的表满足条件的数据。

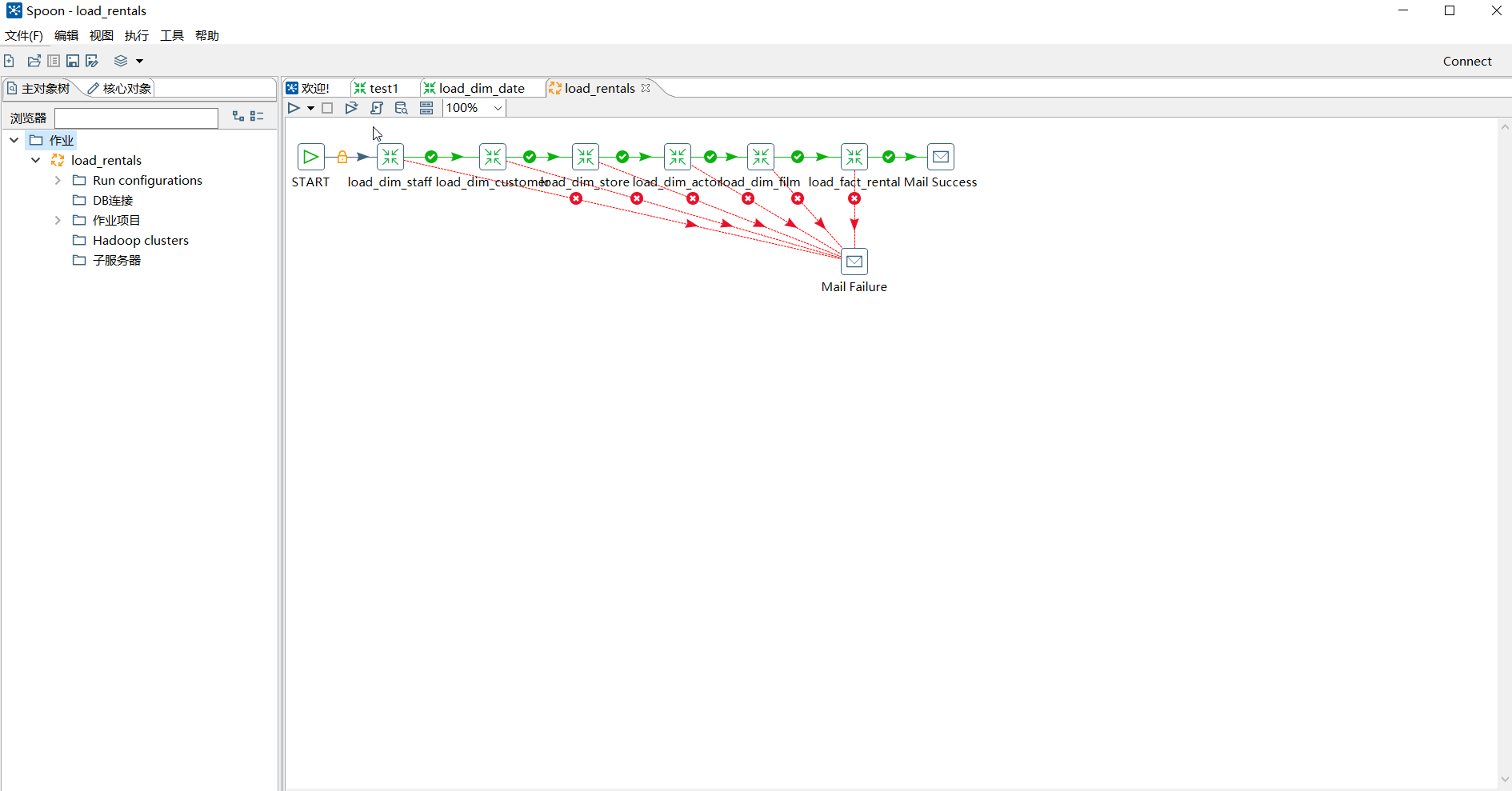
* 1. 转换模块

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。

* 1. 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。

1. 启动步骤
   1. 界面启动,点击界面三角按钮



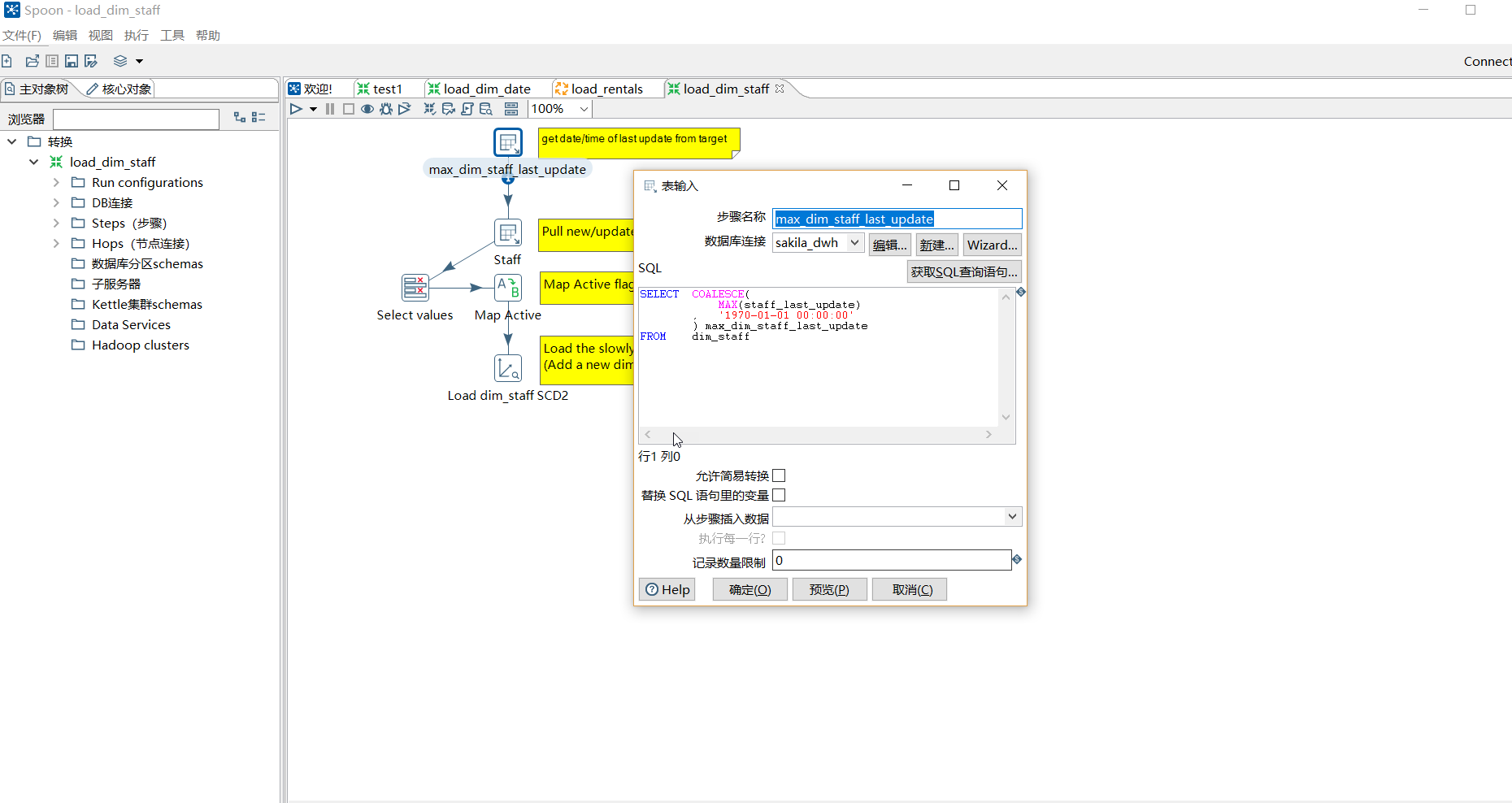
* 1. 命令行启动
     1. 调用脚本ETL\_Sch\_day.bat $last\_extract\_date

1. 参数说明

全量抽取设置抽取日期为last\_extract\_date

2000-01-01

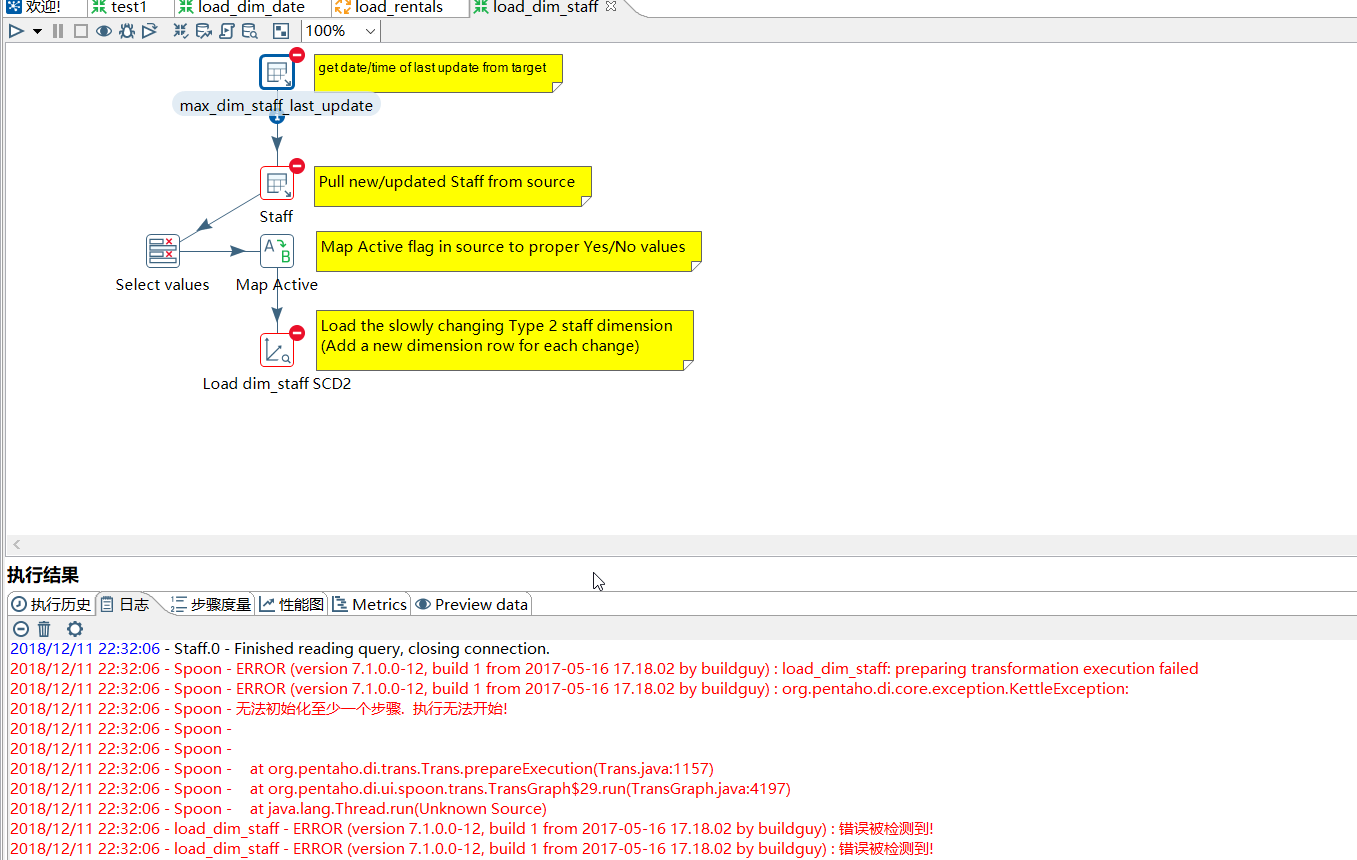
* 1. 增量数据抽取设置
     1. 界面参数设置



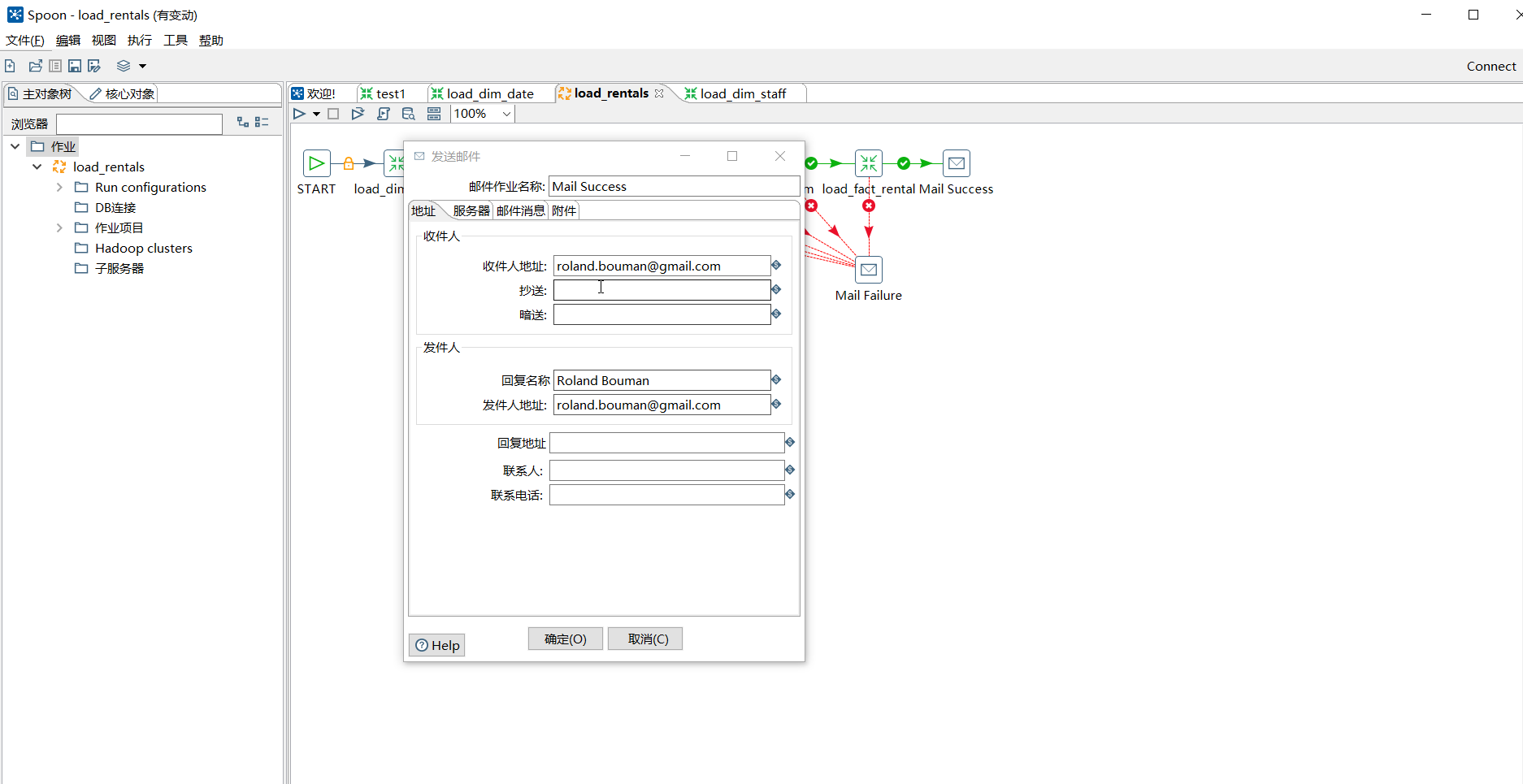
* + 1. 命令行参数设置

上次抽取日期$last\_extract\_date

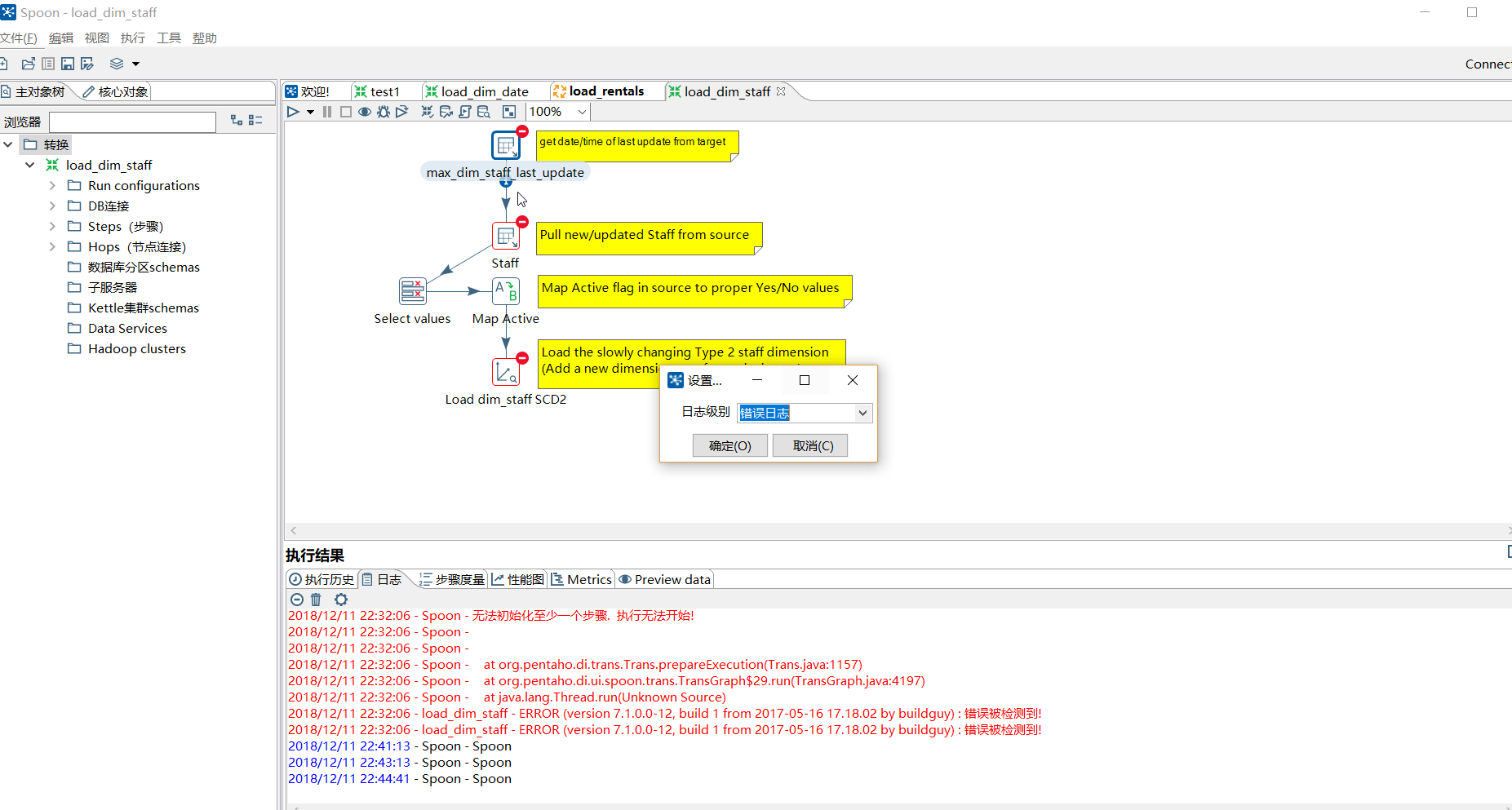
1. 故障排除
   1. MongDB连接故障
   2. Mysql连接故障
   3. 邮件通知故障
2. 监控方式及执行状态
   1. Spoon界面工具



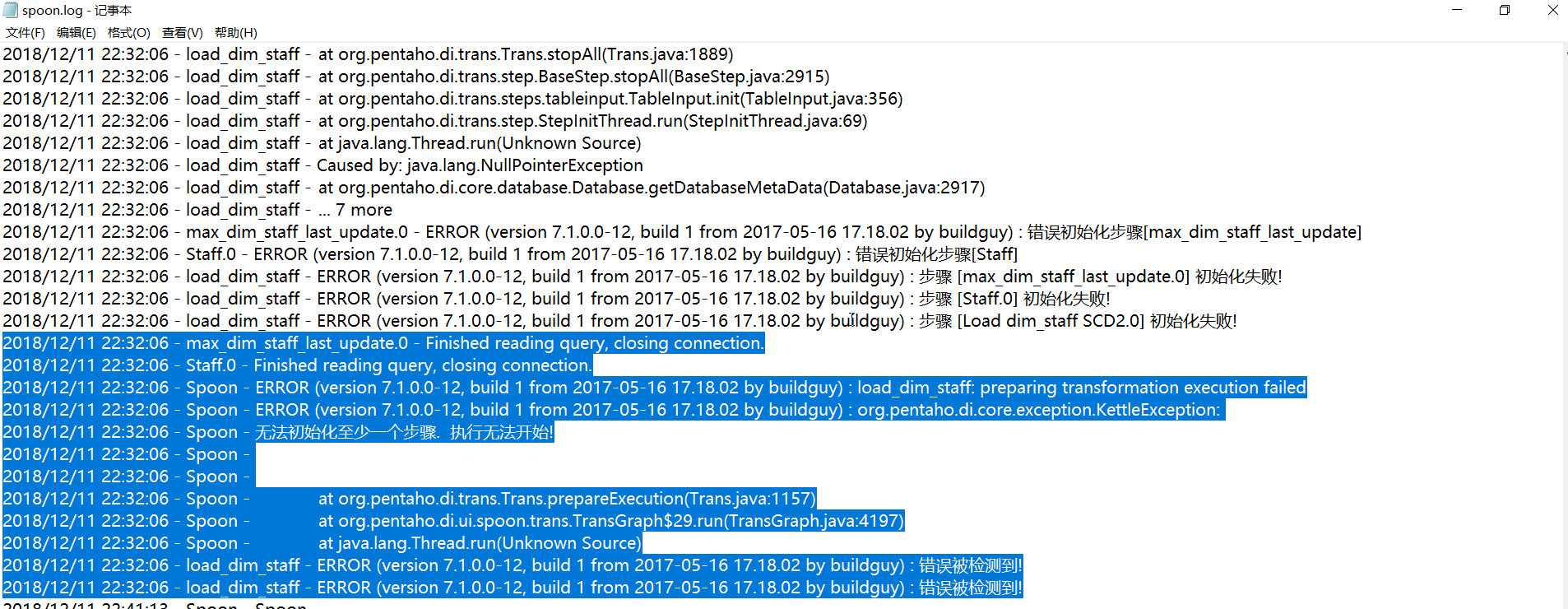
* 1. 脚本执行log文件data-integration\logs\spoon.log
  2. 执行成功邮件通知



1. 异常记录
   1. 界面错误日志



* 1. 错误日志data-integration\logs\spoon.log



1. ETL测试步骤
   1. 数据质量屏幕设计CASE
      1. 新建数据加载测试
         * 新增用户
         * 新增捕鱼角色
         * 新增子弹信息
      2. 更新数据加载测试
         * 用户信息更新
         * 捕鱼角色信息更新
         * 子弹信息更新
      3. 删除数据加载测试
         * 用户信息删除
         * 捕鱼角色删除
         * 子弹信息删除
   2. 数据质量屏幕执行
      * + 应用端做相应的用户，捕鱼角色和子弹信息增加，更新和删除操作
        + 运行ETL，加载数据到目标库中
        + 确认应用端与目标库端数据是否一致
2. 程序中断后恢复

运行全量数据重新刷新，保证没有数据丢失

1. 继续由中断处续写

界面上运行Run Next Entries in Parallel

