

kettle入门(七) 之kettle增量方案(一)全量比对取增量-根据唯一标示 - xiaohai798的专栏 - CSDN博客

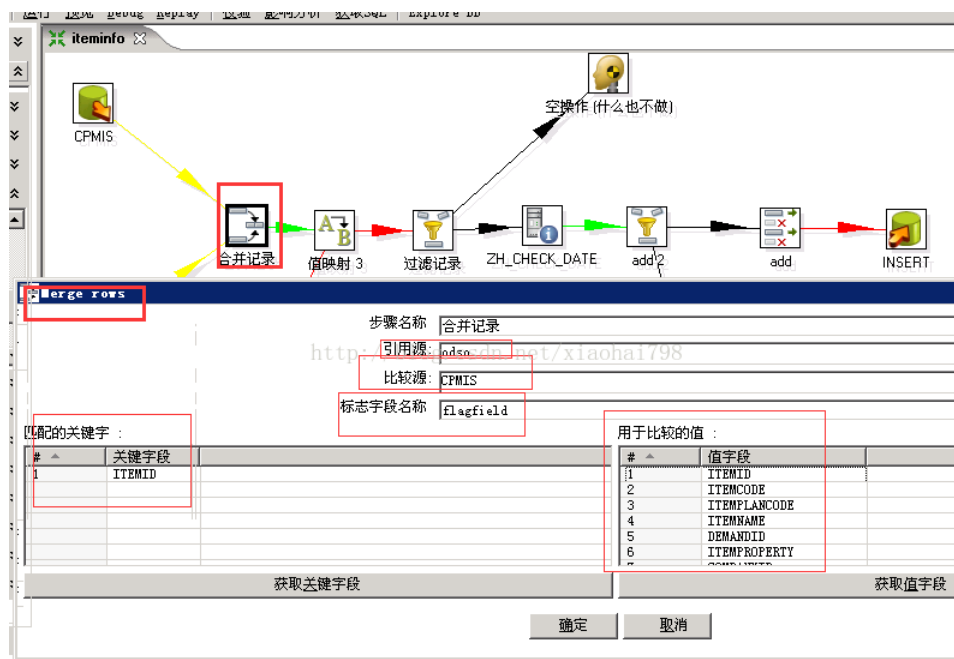
[sql]view plaincopy

引:

ods有个工程表来自于上游系统，数据量不大 十几万，下游系统需要此数据，并且需要每天提供截止当天的增量数据 要求每条数据给出数据变化时间及标示，即数据若是插入 有插入时间和插入标示 若是修改 有修改时间和修改标示 若是删除需逻辑删除、有删除标示且有删除时间

解决:

kettle的转换ktr里有一个图元叫做合并记录，可以把两个表输入分为源和目的根据唯一标示 进行全量比对。由此，我们得到增量的数据流，再写入到RDBMS里，即可以实现该需求。实现功能的ktr如下图：



1 输入源 合并记录

上图，最左侧是两个表输入，上面一个cpmis 是上游系统 下面一个被 merge rows挡住了，名字是odso

cpmis

```
SELECT
ITEMID,
ITEMCODE,
ITEMPLANCODE,
ITEMNAME,
DEMANDID,
ITEMPROPERTY,
```

```

COMPANYID,
MANAGELEVELID,
SPECIALSMALLTYPEID,
PLANTYPEID,
CONSTRUCTPROPERTYID,
AREAID,
ASSISTANTTYPEID,
CONSTRUCTTYPEID,
FEEDEPARTMENTID,
..... FROM PROJECT_AH.ITEMINFO order by ITEMID

```

odso

```

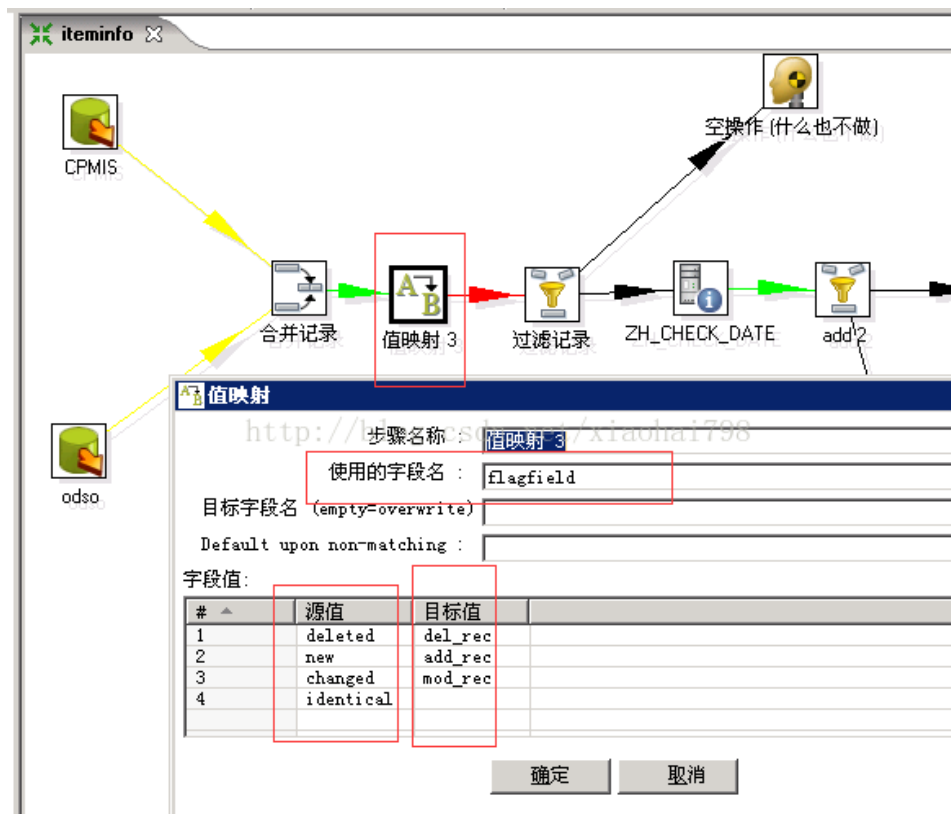
SELECT
ITEMID,
ITEMCODE,
ITEMPLANCODE,
ITEMNAME,
DEMANDID,
ITEMPROPERTY,
COMPANYID,
MANAGELEVELID,
SPECIALSMALLTYPEID,
PLANTYPEID,
CONSTRUCTPROPERTYID,
AREAID,
ASSISTANTTYPEID,
CONSTRUCTTYPEID,
FEEDEPARTMENTID,
..... FROM odso.ITEMINFO order by ITEMID

```

2 值映射

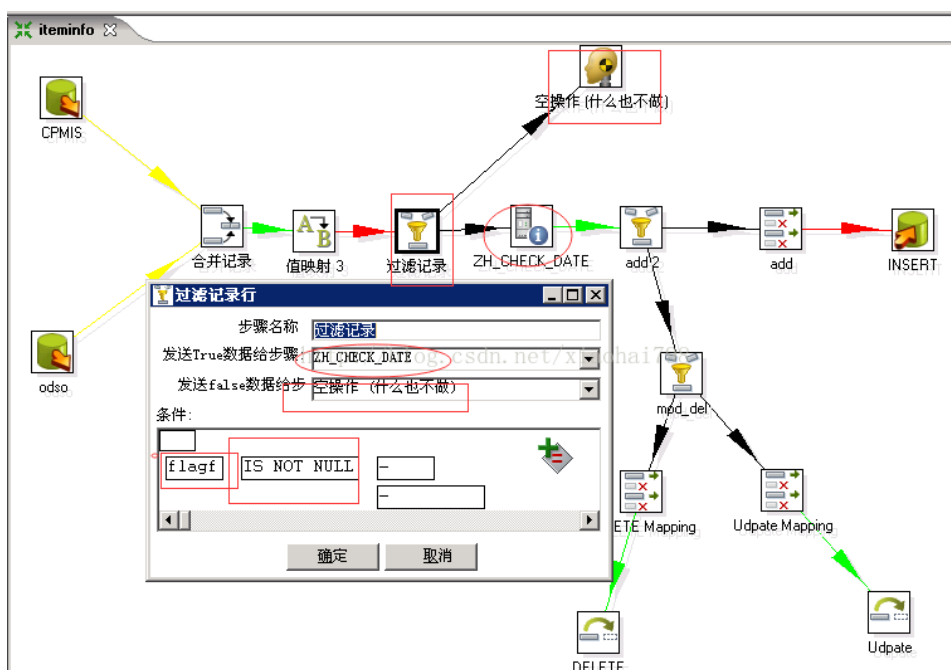
如下图， 将输入源与目的源的每个字段数据根据唯一字段比较后 到值映射图元，使用字段名 为起的后面用到的一个变量名（可任意起），源值列为系统默认

1 代表输入源与目标源比较后删除的数据标志 2 是输入源新增 3 是输入源更新 4是不变 目标值 是自己起的名字 可以根据需要不变 或修改



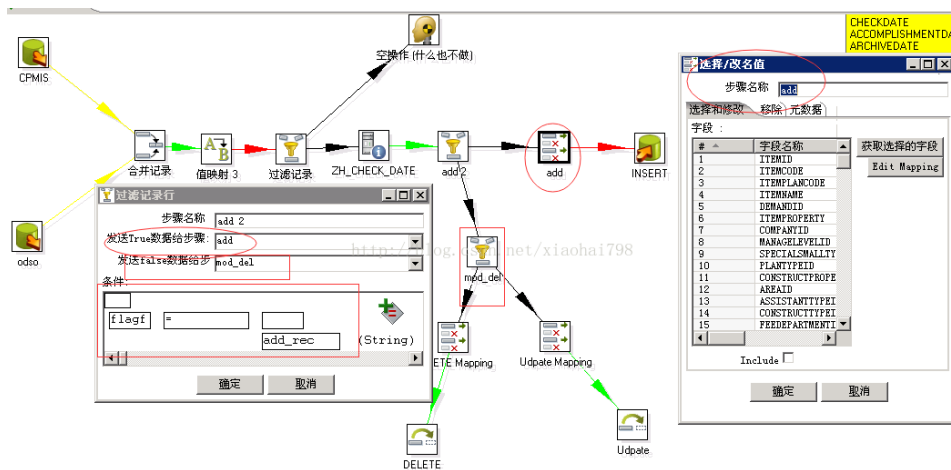
3 第一次过滤无效记录

如下图，条件 flagfield is not null （后面没显示完），若条件成立发送给下一步zh_check_date，若不成立发送给空操作。



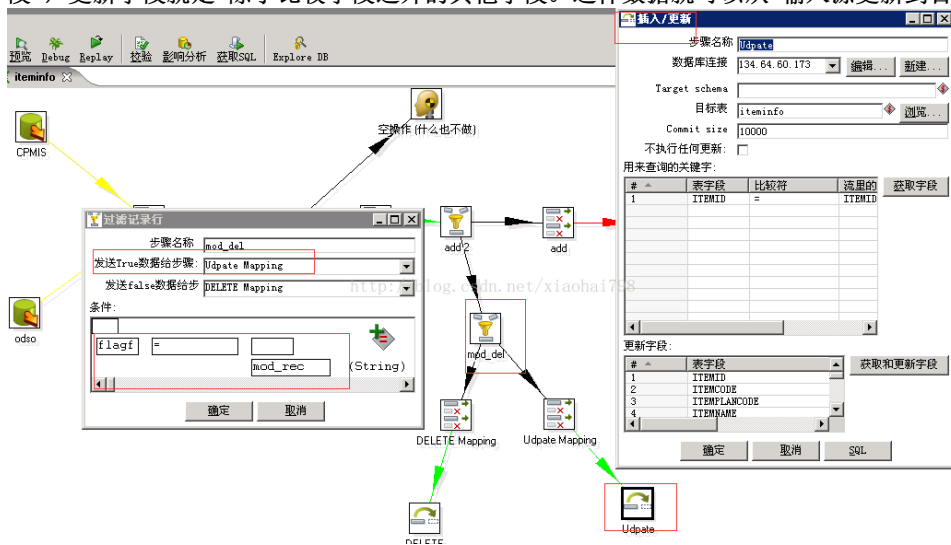
4 新增数据判断add2

如下图，zh_check_date 为获取当前的系统时间变量。add2 图元打开为 画圈的图 左侧的地方，条件 flagfield = add_rec，若成立及发送数据到中间画圈的add图元，若不成立则 发送数据到mod_del图元（矩形红框） 假设为true数据到 add图元，打开 即是下图右侧 部分 填写需要插入的数据字段 再到insert图元，即可把输入源比目标源多的新数据更新到目标表来 并且加上时间戳。



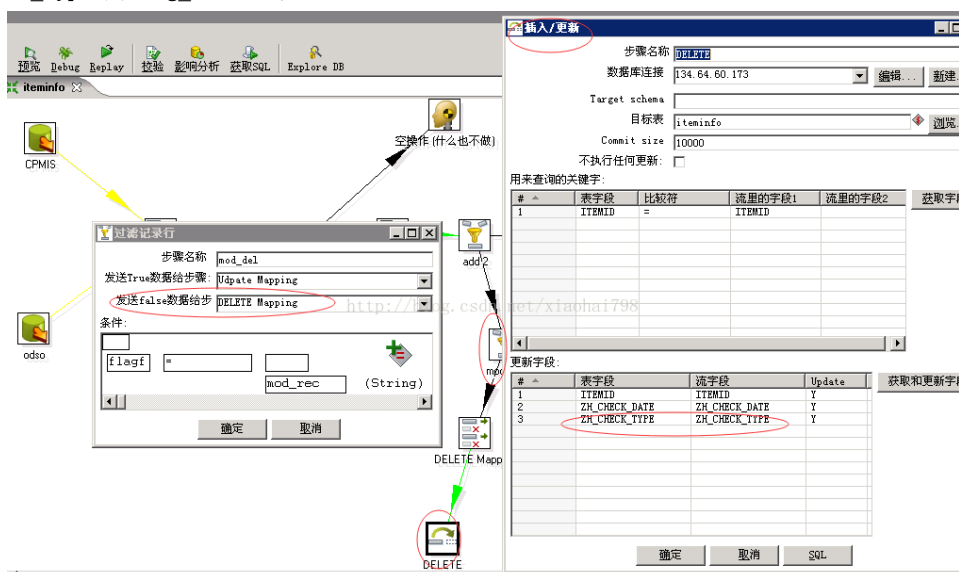
5 修改或删除 mod_rec

如下图，假设数据从add2发送而来，打开矩形框 mod_del 条件flagfield = mod_rec 若true 则发送到 update mapping 若false 则发送到delete mapping，假设是更新，则右侧的查询关键字是更新的比较字段即是一开始合并记录的比较字段，更新字段就是除了比较字段之外的其他字段。这样数据就可以从输入源更新到目标源。



6 删除数据 delete mapping

如下图，假设数据流到了 delete mapping，则只需要根据比较字段把目标表的时间戳更新和状态更新为del_rec即可，下图zh_check_type 为flag_field的值



7 数据结果

zh_check_type 和check_date已更新为

```
select * from odso.iteminfo t
```

	DEMANDCREATEMANID	DEMANDMANAGER	ZH_CHECK_TYPE	ZH_CHECK_DATE	ITEMMANAGER	ITEMMANAGEDEPT
1			mod_rec	2014-8-26 13:03:04		需求管理中心
2			mod_rec	2014-8-26 13:03:04		需求管理中心
3			mod_rec	2014-8-26 13:03:04		需求管理中心
4			mod_rec	2014-8-26 13:03:04		需求管理中心
5			del_rec	2015-4-29 2:49:56		需求管理中心
6			mod_rec	2014-3-24 11:40:51		需求管理中心
7			mod_rec	2014-3-24 11:40:51		需求管理中心
8			mod_rec	2014-3-24 11:40:51		需求管理中心
9			mod_rec	2014-3-24 11:40:51		需求管理中心
10			del_rec	2015-4-29 2:49:56		需求管理中心
11			mod_rec	2014-8-26 13:03:04		需求管理中心
12			mod_rec	2014-8-26 13:03:04		需求管理中心
13			del_rec	2015-4-29 2:49:56		需求管理中心
14			del_rec	2015-4-29 2:49:56		需求管理中心