iUAP DI维度查询和更新简要解析

作者：魏剑龙

目录

[1. 介绍 2](#_Toc427825169)

[2. iUAP DI第1种方法的支持 3](#_Toc427825170)

[2.1. 维度更新 3](#_Toc427825171)

[2.2. 事实表载入加载维度外键 5](#_Toc427825172)

[3. iUAP DI第2种方法的支持 6](#_Toc427825173)

[4. 事实表载入加载维度外键 8](#_Toc427825174)

# 介绍

处理缓慢变化维的方法通常分为三种方式

1. 当某个维度成员的数据发生变化时，最新的列值将覆盖以前的维度记录，从而清除了该维度成员的历史记录

1. 当某个维度成员的数据发生变化时，最新的列值将存储为维度中的新记录，从而提供了一个维度成员的多个实例，这样便保留了历史记录

1. 当某个维度成员的列数据发生变化，而数据仓库要保留该变化列的最后一个版本时，原始数据将移到该维度记录的最后一个版本列中，并且所有新维度信息将覆盖现有列。

以上三种方式中，第1种处理最简单，直接在原来维度的基础上进行更新,不会产生新的记录，但是将会导致分析结果的失真，比如当一个销售员所在的部门发生变化的时候，如果采用第一种方式，在处理部门和销售员之间的上钻、下钻操作时，会导致聚合误差，将销售员在原来部门的销售业绩带入到新的部门中。

|  |  |  |  |
| --- | --- | --- | --- |
| emp\_rid | emp\_id | emp\_name | position |
| 更新前： | | | |
| 101212 | 12345 | Jack | Developer |
| 更新后： | | | |
| 101212 | 12345 | Jack | Manager |

采用第2 种方法的时候，需要加入一个代理键，不修改原有的数据，重新产生一条新的记录,这样就可以追溯所有的历史记录。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| emp\_rid | emp\_id | emp\_name | position | dim\_version | start\_date | end\_date |
| 更新前： | | | | | | |
| 101212 | 12345 | Jack | Developer | 1 | 2010-02-15 | 2011-02-15 |
| 更新后： | | | | | | |
| 101212 | 12345 | Jack | Developer | 1 | 2010-02-15 | 2011-02-15 |
| 105178 | 12345 | Jack | Manager | 2 | 2011-02-16 | 2199-02-18 |

第3种方式由于可能会涉及到动态改变数据仓库模型，这样会带来比较大的开销，所以在实际的数据仓库实现中并不经常使用。

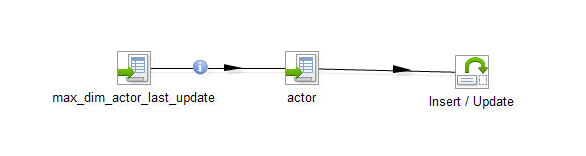
|  |  |  |  |
| --- | --- | --- | --- |
| emp\_id | emp\_name | position | old\_position |
| 更新前： | | | |
| 12345 | Jack | Developer | null |
| 更新后： | | | |
| 12345 | Jack | Manager | Developer |

一般情况下，采用第1、第2 种方法结合使用的策略，对于维度中一些描述性的属性，采用直接覆盖的方式，而对于那些将会出现在层次中，可能会参与聚合计算的属性，则采用第二种方式。在下文中着重主要介绍第一种和第二中方法的支持，包括维度的更新和事实表加载。

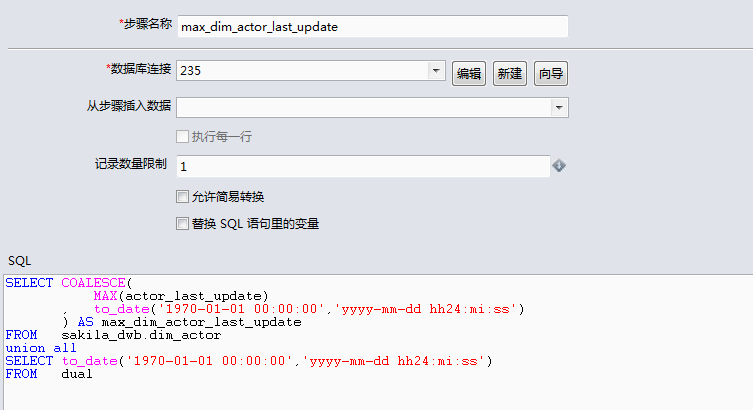
# iUAP DI第1种方法的支持

## 维度更新

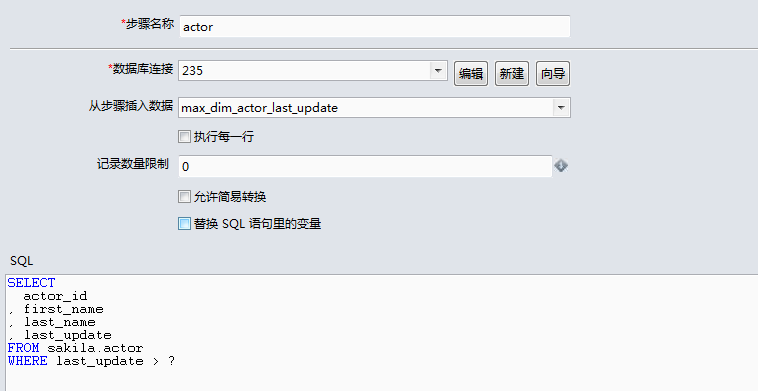
在转换中使用”插入\更新”,例如下面转换设置:



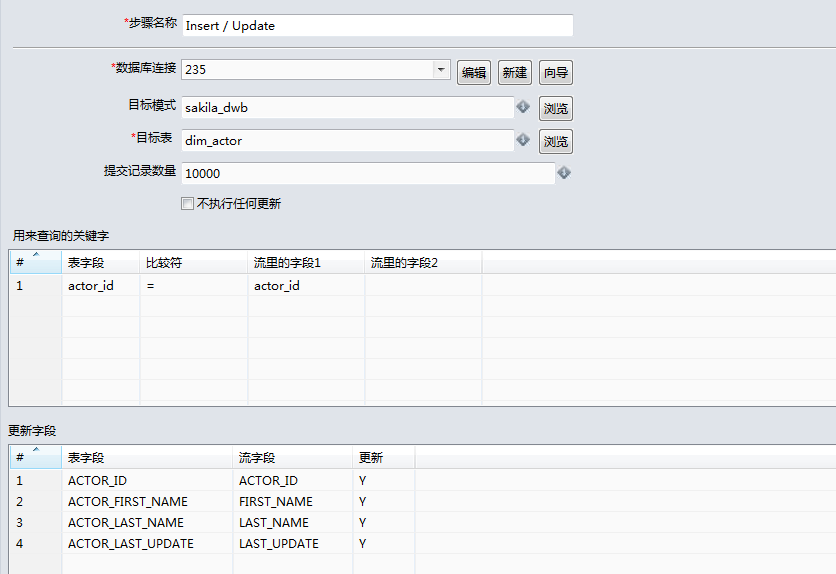
其中第一个步骤用来查询演员维度的最大时间戳



第二个步骤根据最大时间戳从源表中加载演员数据:



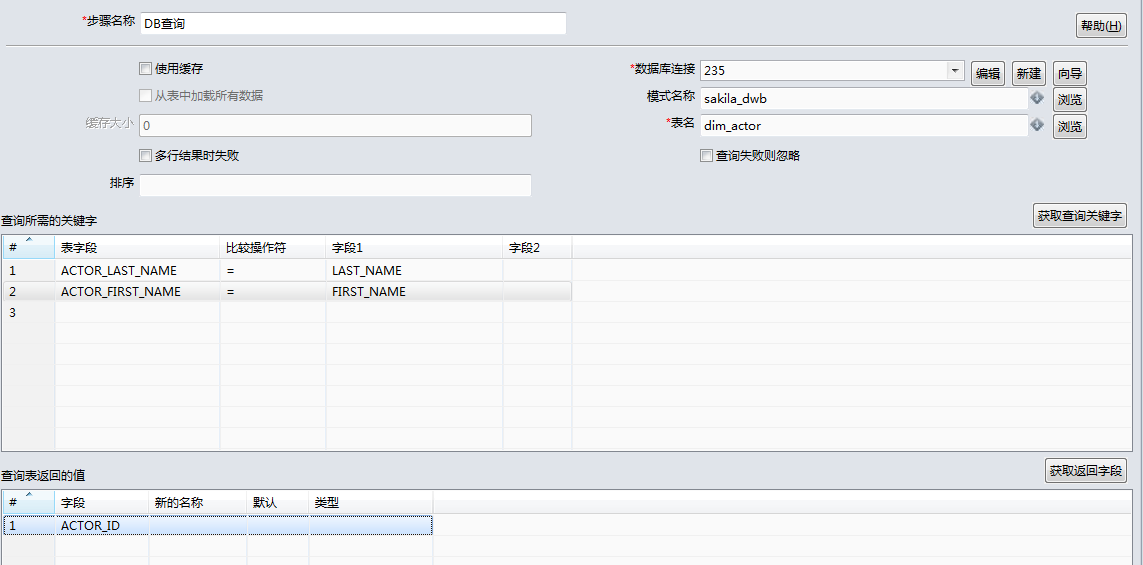
第三个步骤在维度表中插入\更新数据



运行以上转换,发现自上次有更新的记录,就会同步到维度表中;

## 事实表载入加载维度外键

使用AE DI转换”查询”分类中”DB查询”或”meta查询”即可实现使用关键字查询返回维度表的主键,如果维度表的主键和维度源表的主键相同,可以直接使用源事实表的外键值.



# iUAP DI第2种方法的支持

**维度更新**

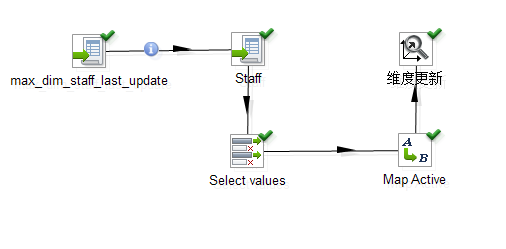
使用转换中的”维度更新”步骤,实现维度更新.

首先在为维度表中使用代理主键staff\_key,保存源表的主键staff\_id,同时要多出三个字段

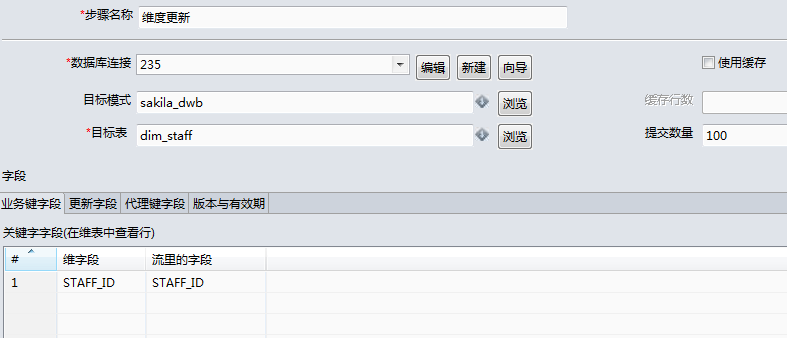
STAFF\_VERSION\_NUMBER NUMBER(10) //版本

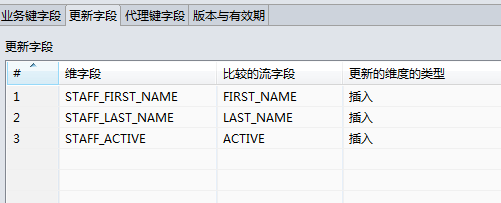
STAFF\_VALID\_FROM TIMSTAMP //有效起始日期

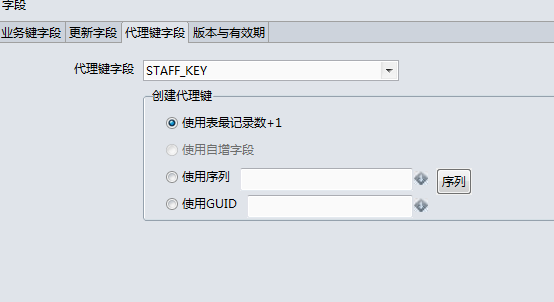
STAFF\_VALID\_THROUGH TIMSTAMP //有效截止日志

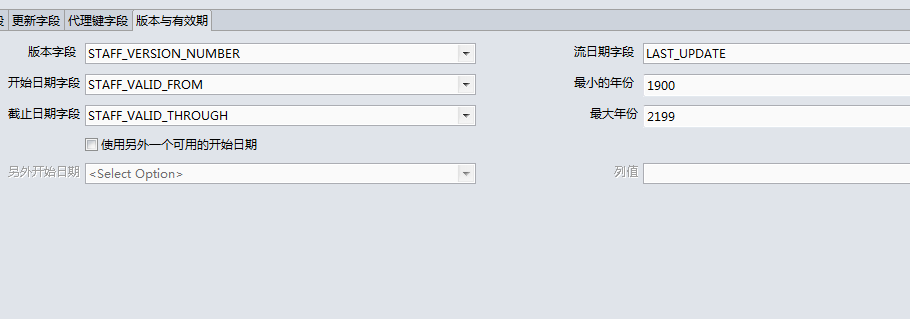


维度更新的设置截图如下:

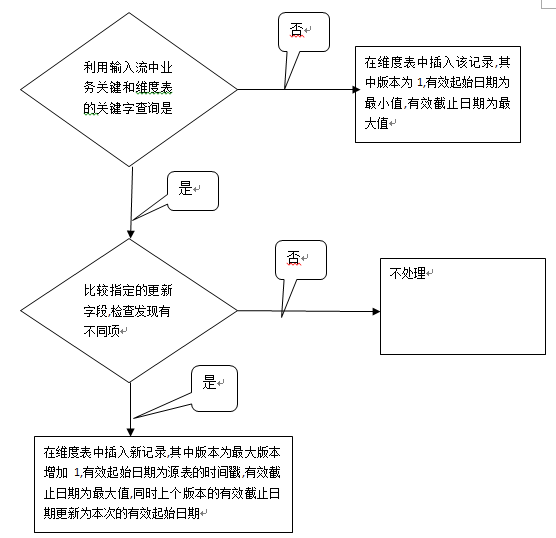






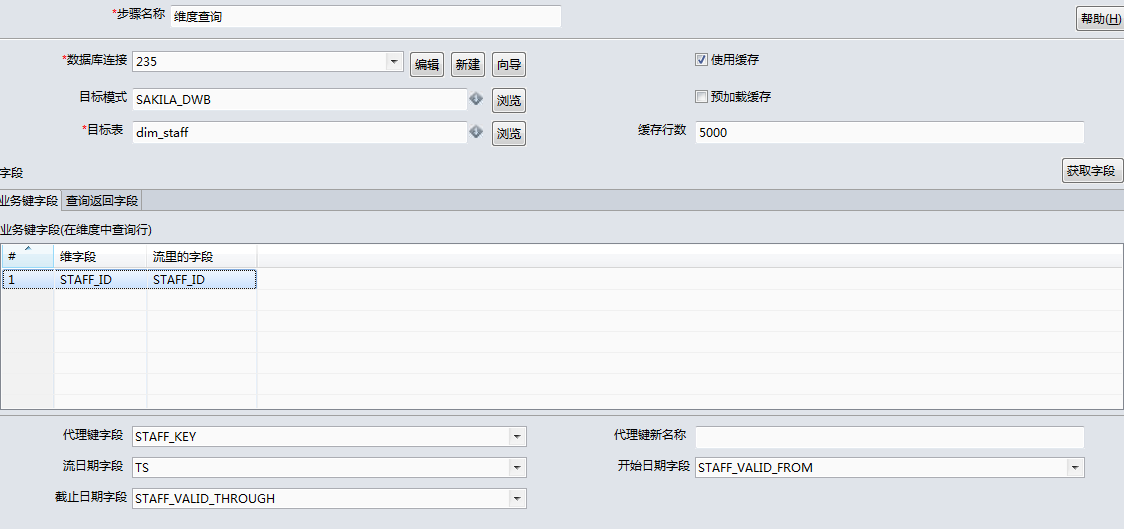


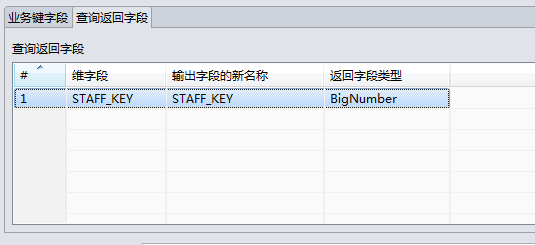
其原理就是如下图:



# 事实表载入加载维度外键

使用转换中的维度查询载入,设置截图如下:





其原理是利用业务ID在维度表中查询出维度表的数据,再利用业务时间和维度表的有效起始和截止时间进行比较,找出符合条件的维度记录.