BDI地址：[http://10.161.35.65:1821/portal/pure/Frame.action#](http://10.161.35.65:1821/portal/pure/Frame.action)

按照实现以下需求为例进行讲解：

## 一、建表和扫描表

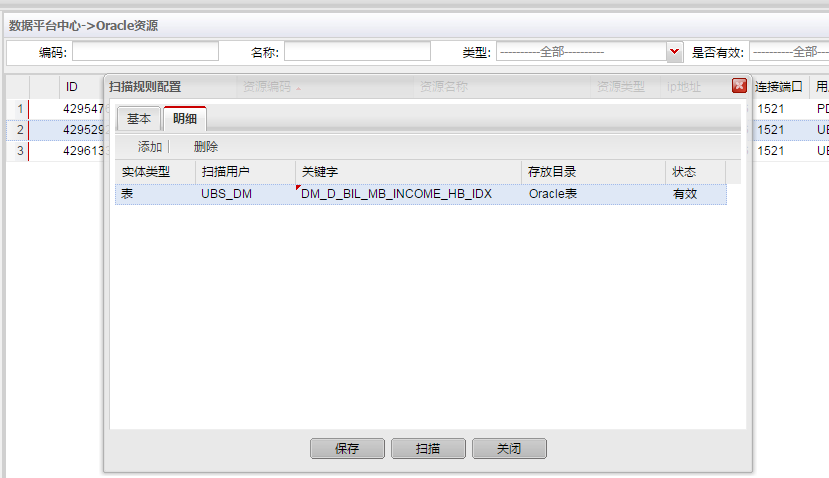
首先在Oracle和Hive上分别建立表dm\_d\_bil\_mb\_income\_hb\_idx，



由于字段较多，收入包含bs和cb两部分，用到的来源表均不同，故在Hive上建立dm\_d\_bil\_mb\_income\_hb\_idx\_bs、dm\_d\_bil\_mb\_income\_hb\_idx\_cb ，分别计算出bs和cb侧的当日收入，然后再根据这两张表计算本月累计、上月同期、环比以及移网收入等其他指标。

建完表后需要将Oracle和Hive的表分别扫描到BDI里，

点击“我的应用”—“元数据”—“数据中心”，点击“Oracle资源”，选择“Oracle-UBS\_DM”右击选择“实体扫描”，切换到“明细”页，如下图：



实体类型：选择表；

扫描用户：对应Oracle上的用户名；

关键字：直接填写表名；

存放目录：选择在bdi里建好的目录，放在Oracle表目录下；

然后点击“扫描”即可，右击点击“扫描日志”可以查看扫描是否成功；



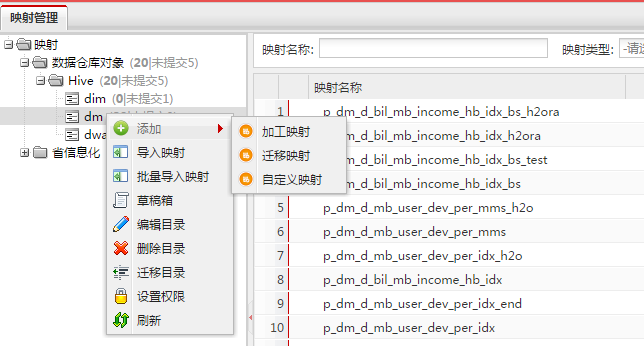
同理将Hive的表扫描到BDI里，扫描用户填写Hive的用户open\_017\_dm，存放目录选择建立好的目录即可。

## 二、映射

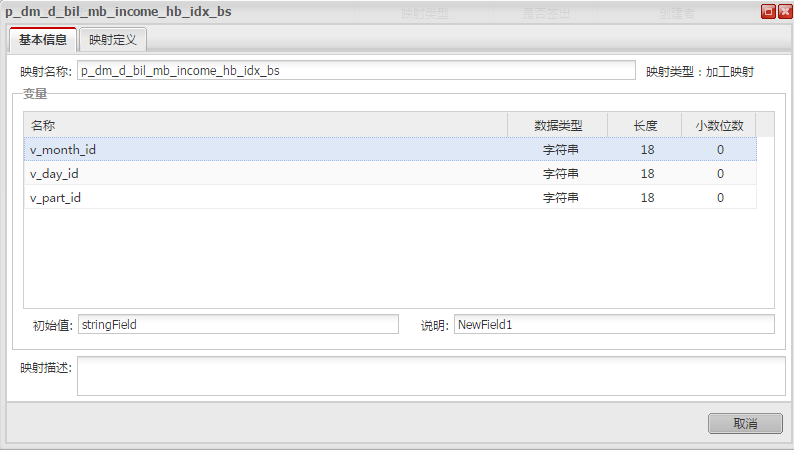
映射相当于存储过程，数据处理逻辑。编辑完的映射需要签入后工作流才可以引用，未签入的映射会保存在草稿箱里，要编辑映射需要先把映射签出。

### （一）加工映射

点击“我的应用”—“云化ETL”—“映射管理”，打开映射的界面

**BS侧映射p\_ dm\_d\_bil\_mb\_income\_hb\_idx\_bs**

添加“加工映射”，映射名称“p\_表名”，添加用到的变量参数，参数也可以后续添加



点击确定后如下图所示：



根据相应的sql算法用到什么节点直接拖到右侧即可。

select substr(service\_type,1,2) service\_type,

sum(nvl(total\_fee,0))/10000 total\_fee

from (select \* from dwa\_v\_d\_cus\_mb\_user\_info

where month\_part='201701' and day\_part='01'

and is\_stat='1') a,

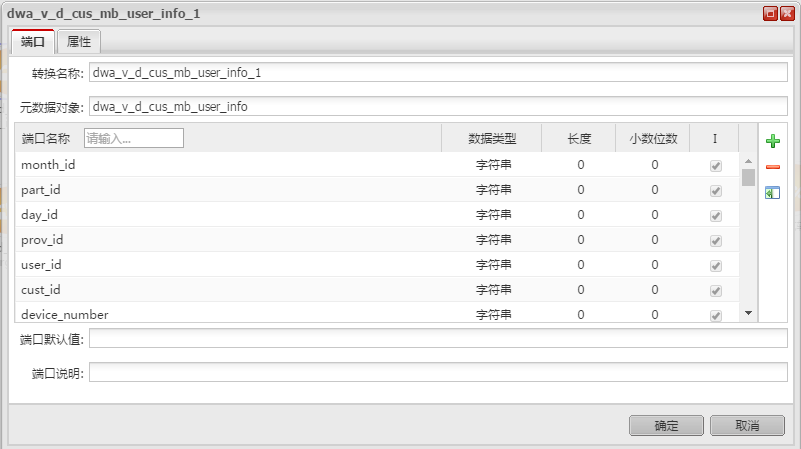
(select \* from dwa\_v\_d\_cus\_al\_bcharge b

where part\_id='1' and day\_id='01' ) b

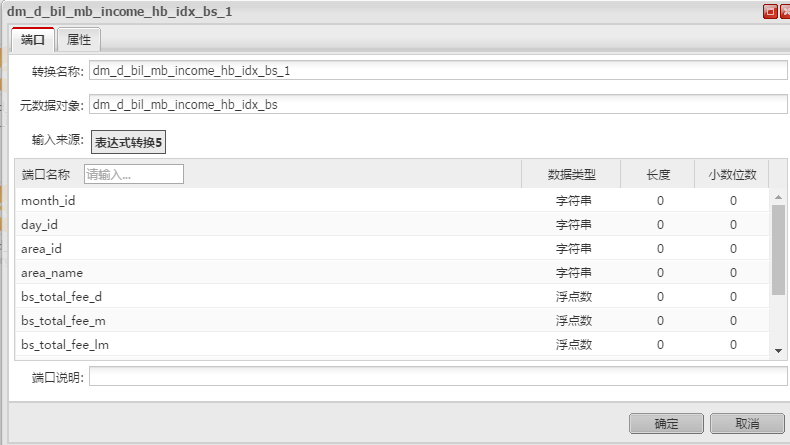
where a.user\_id=b.user\_id

group by substr(service\_type,1,2)

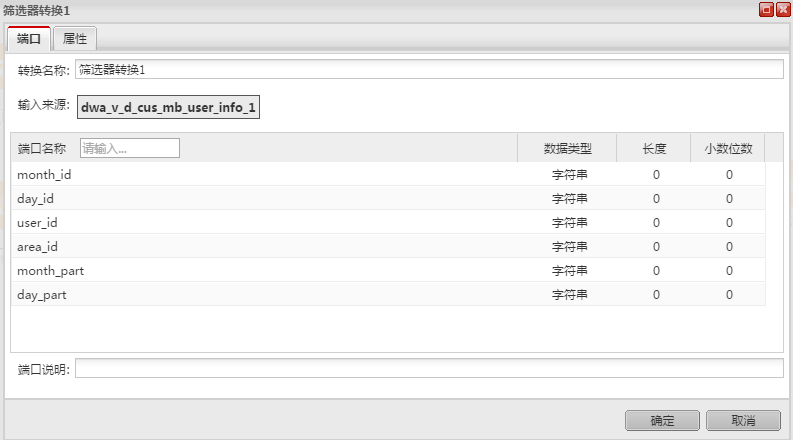
C:\Users\SGQ\AppData\Roaming\Tencent\Users\398003507\QQ\WinTemp\RichOle\6PAPB37GX@2~8MU]{RW$]C8.png：来源表，映射的开始节点。

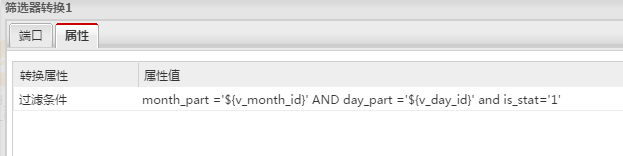
如图：

：目标表，映射的结束节点，将数据插入到表中。

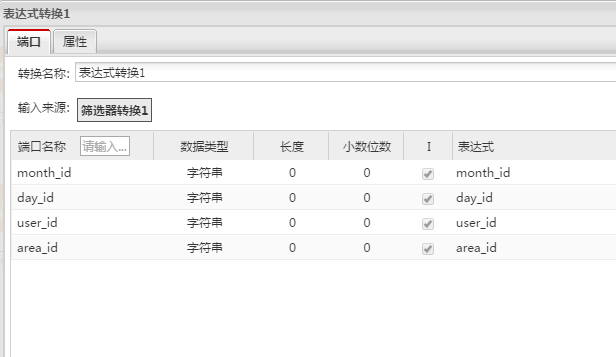
如图：输入来源是表达式5

：紧跟在来源表后面，条件限制，相当于sql中的where，在此组件中可以选择必要的字段，属性里填写各种条件，如图：

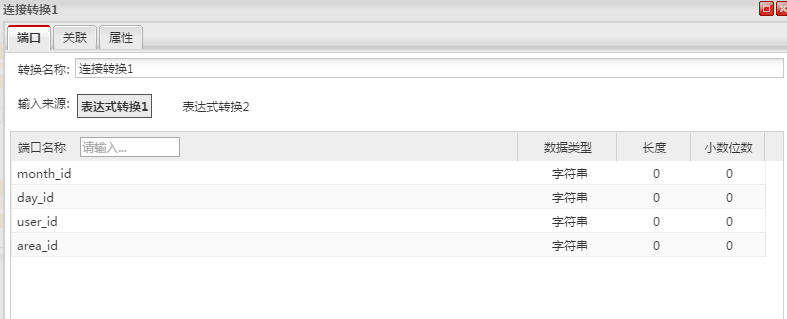


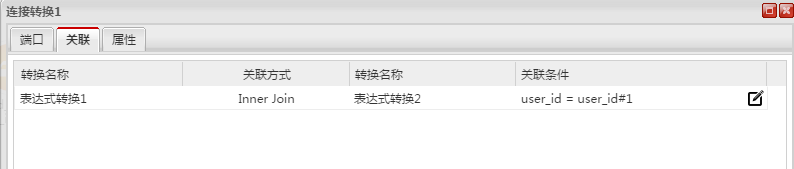


：表达式，一般进行相关的转换，比如case when 语句、字符串截取、Nvl函数等，如图：



：连接转换，用于两个表的连接，可以选择Inner Join、Left Join、Right Join、Full join等，如图：





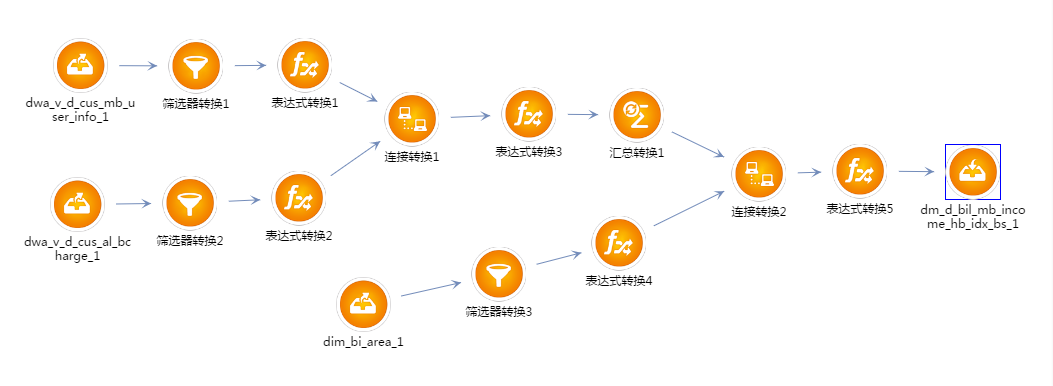
表达式转换1和表达式转换2通过user\_id 进行等值连接。

：用于分组的情况，表达式里用到sum、count等函数需要分组时，如图：



：就是合并，类似union all，两部分的字段要保持一致。

画完后的bs表映射如下图所示：



实现的sql功能如下：

select substr(service\_type,1,2) service\_type,

sum(nvl(total\_fee,0))/10000 total\_fee

from (select \* from dwa\_v\_d\_cus\_mb\_user\_info

where month\_part='201701' and day\_part='01'

and is\_stat='1') a,

(select \* from dwa\_v\_d\_cus\_al\_bcharge b

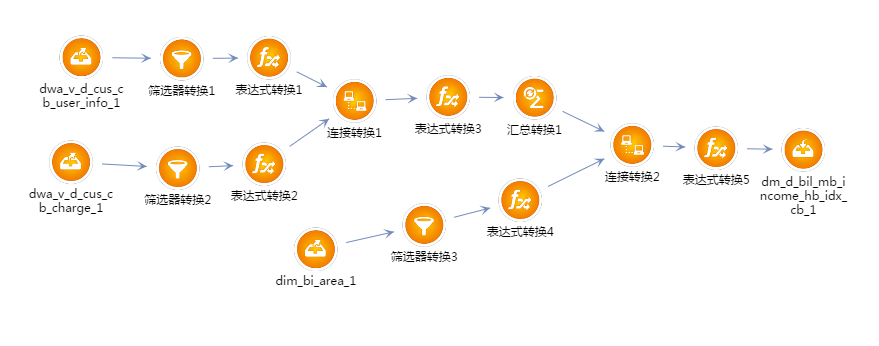
where part\_id='1' and day\_id='01' ) b

where a.user\_id=b.user\_id

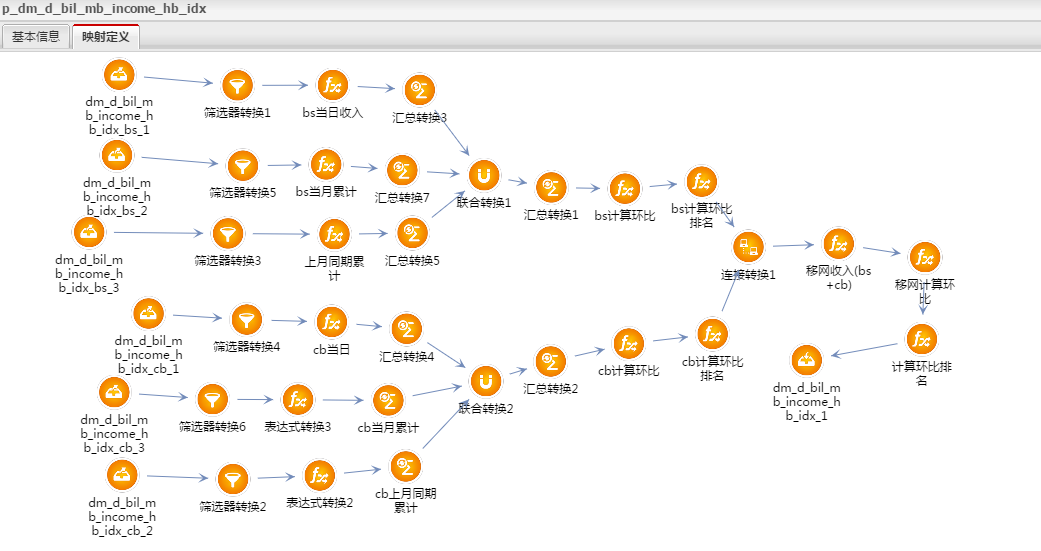
group by substr(service\_type,1,2)

**CB侧映射p\_ dm\_d\_bil\_mb\_income\_hb\_idx\_cb**

Cb侧的映射同理，如图：

**bs侧和cb侧合并p\_dm\_d\_bil\_mb\_income\_hb\_idx**

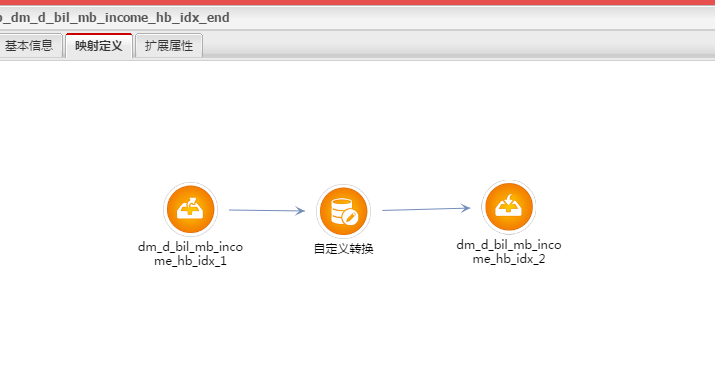
映射如下：根据bs侧表和cb侧表分别得到当日收入、本月累计收入、上月同期收入、环比以及排名，然后union all，最后再计算移网收入以及排名



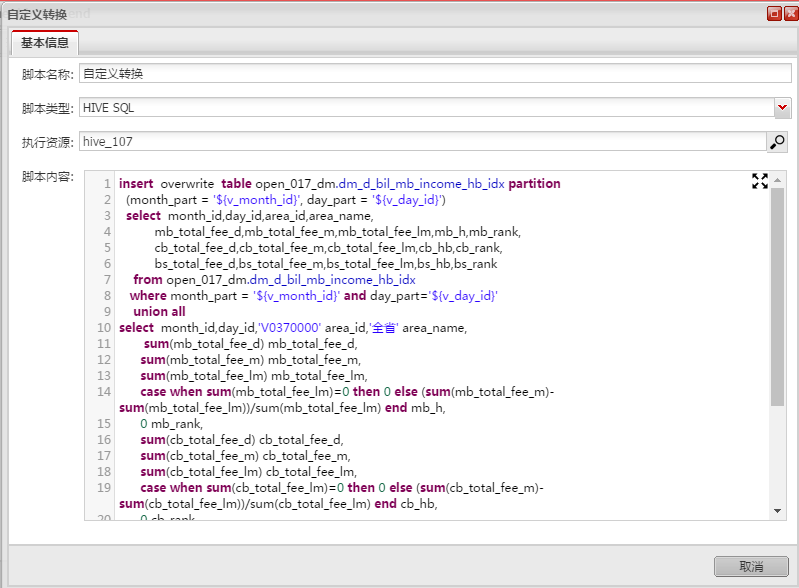
### （二）自定义映射

**加工映射计算全省数据p\_dm\_d\_bil\_mb\_income\_hb\_idx\_end**

右击“添加”选择“自定义映射”，如图：

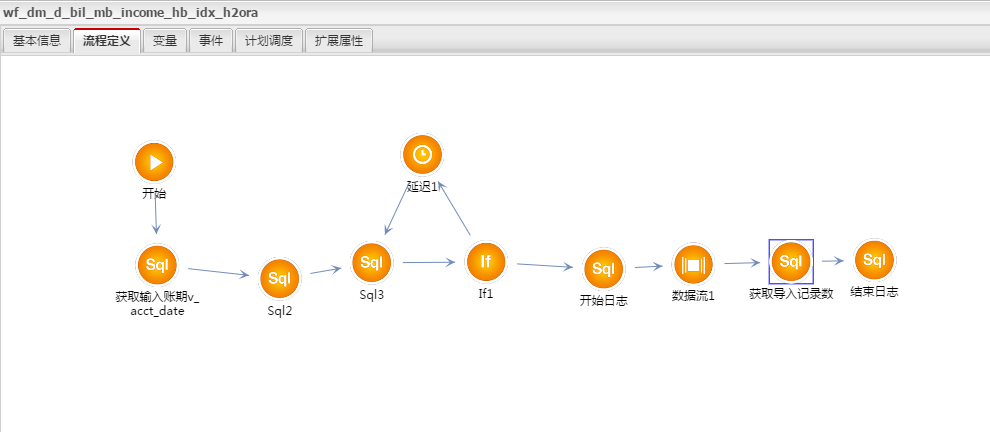


自定义转换需要自己写sql，计算一条全省的数据。

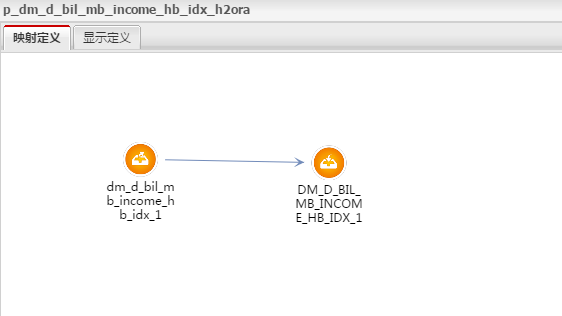


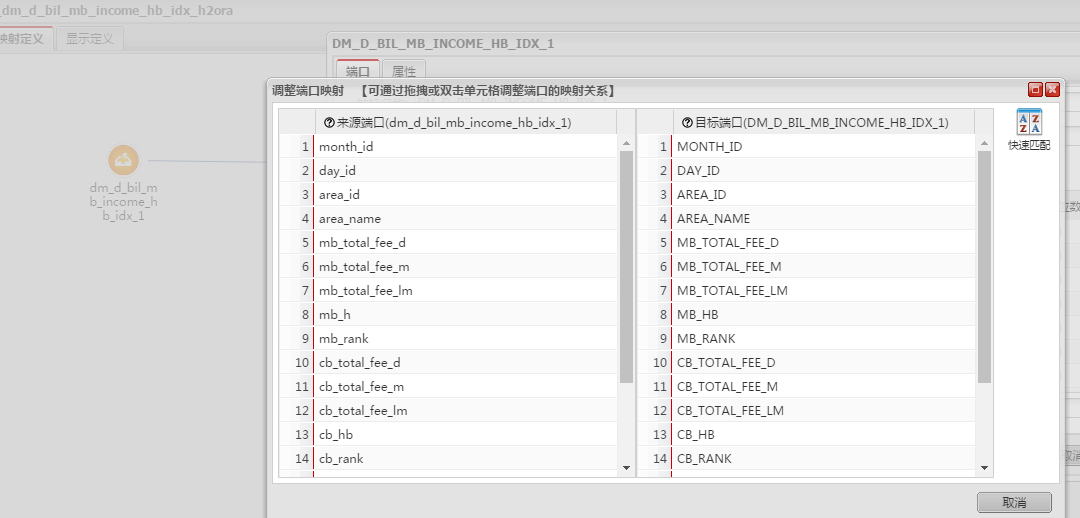
### （三）迁移映射(迁Oracle行云等)

Hive的表数据迁到oracle或者行云等。wf\_dm\_d\_bil\_mb\_income\_hb\_idx\_h2ora



判断迁移的表当前账期是否有数，满足条件后执行迁移流程。

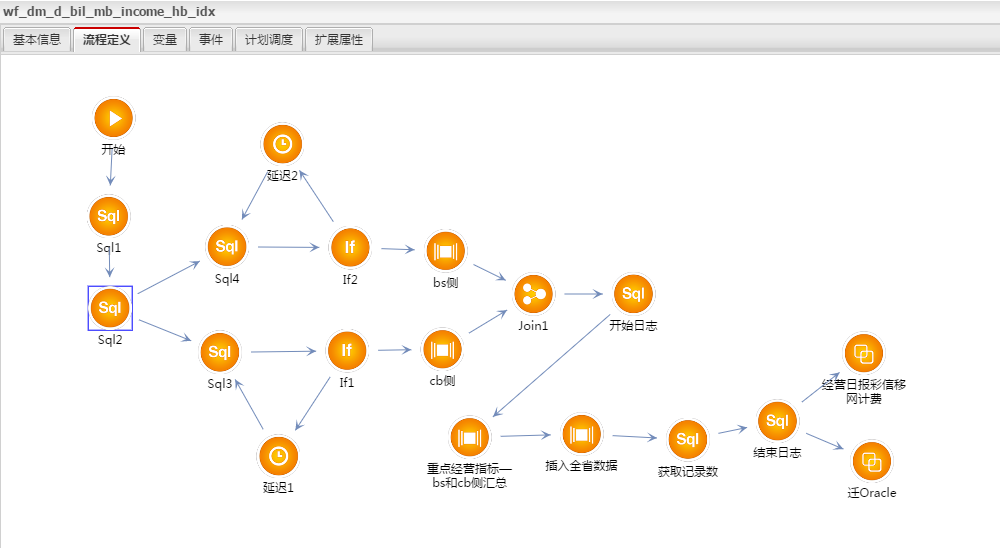




来源表与目标表的字段对应正确即可。

## 三、工作流

工作流主要就是将各种日期参数的获取计算、查询日志前置条件是否满足、以及开始结束日志、映射流程等串起来。



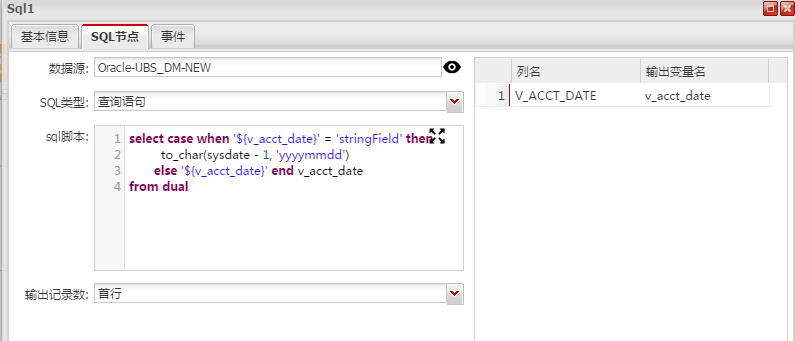
节点Sql1：获取日期参数，默认是前一天，否则是手工输入的日期。

select case when '${v\_acct\_date}' = 'stringField' then

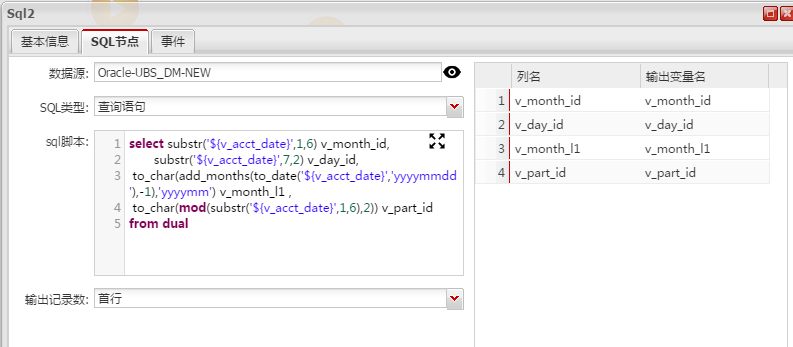
to\_char(sysdate - 1, 'yyyymmdd')

else '${v\_acct\_date}' end v\_acct\_date

from dual



节点Sql2：加工日期参数，获取6位的月份和2位的天，上个月，part\_id等。



select substr('${v\_acct\_date}',1,6) v\_month\_id,

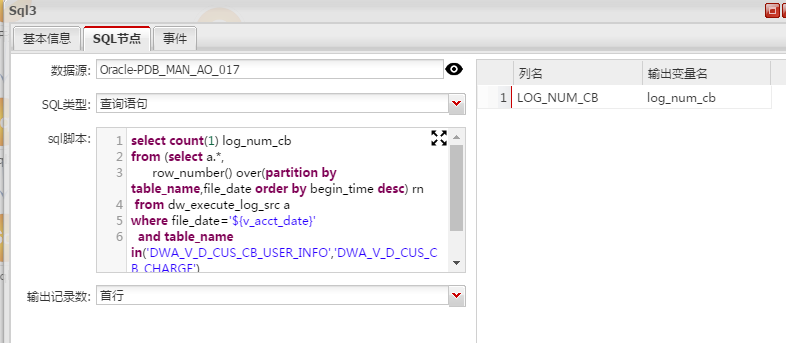
substr('${v\_acct\_date}',7,2) v\_day\_id,

to\_char(add\_months(to\_date('${v\_acct\_date}','yyyymmdd'),-1),'yyyymm') v\_month\_l1 ,

to\_char(mod(substr('${v\_acct\_date}',1,6),2)) v\_part\_id

from dual

节点Sql3、If1、延迟1：判断cb侧的前置条件是否满足，满足就执行cb侧流程，不满足就执行延迟1。



select count(1) log\_num\_cb

from (select a.\*,

row\_number() over(partition by table\_name,file\_date order by begin\_time desc) rn

from dw\_execute\_log\_src a

where file\_date='${v\_acct\_date}'

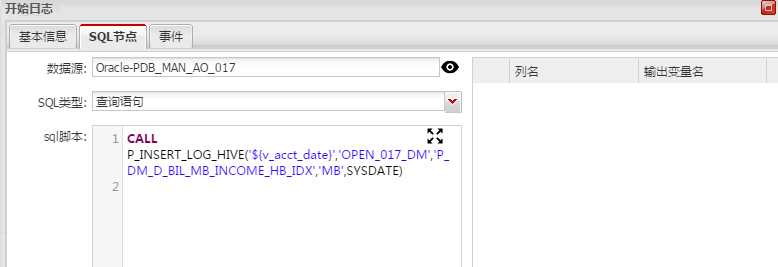
and table\_name in('DWA\_V\_D\_CUS\_CB\_USER\_INFO','DWA\_V\_D\_CUS\_CB\_CHARGE')

and file\_status = 'SUCCESS')

where rn = 1

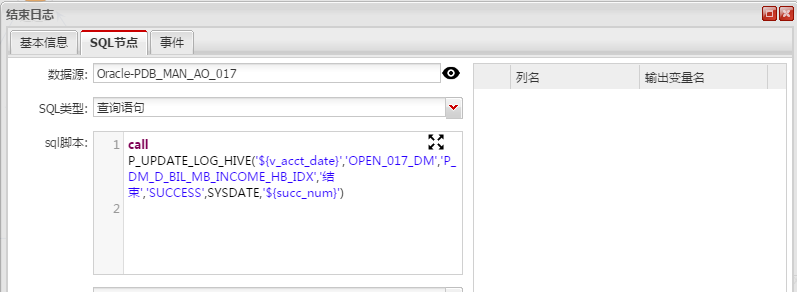


join节点：bs侧和cb侧两者需要同时完成后再向下执行插入开始日志，



CALL P\_INSERT\_LOG\_HIVE('${v\_acct\_date}','OPEN\_017\_DM','P\_DM\_D\_BIL\_MB\_INCOME\_HB\_IDX','MB',SYSDATE)

调用bs侧和cb侧和汇总流程，调用插入全省的流程，插入结束日志，



call P\_UPDATE\_LOG\_HIVE('${v\_acct\_date}','OPEN\_017\_DM','P\_DM\_D\_BIL\_MB\_INCOME\_HB\_IDX','结束','SUCCESS',SYSDATE,'${succ\_num}')

最后调用迁移的流程迁移到oracle。

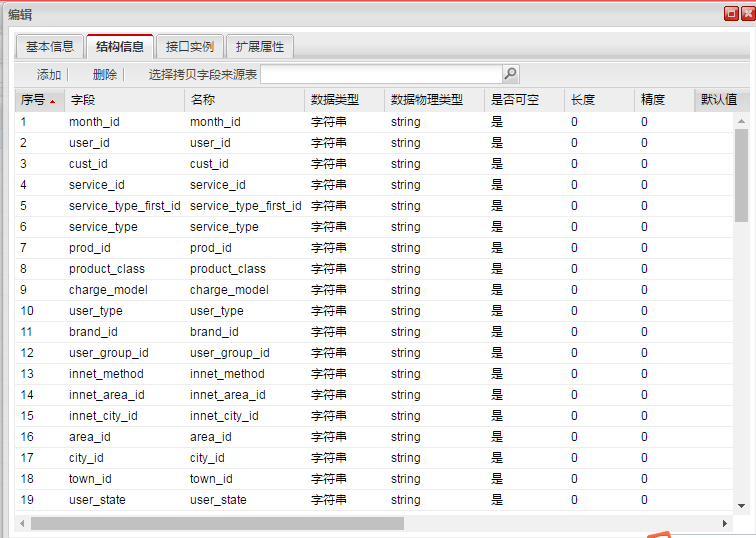
## 四、文件入库

### 1、配置接口文件

依次点击“我的应用”—“元数据”—“数据中心”，选择接口文件，如图所示：

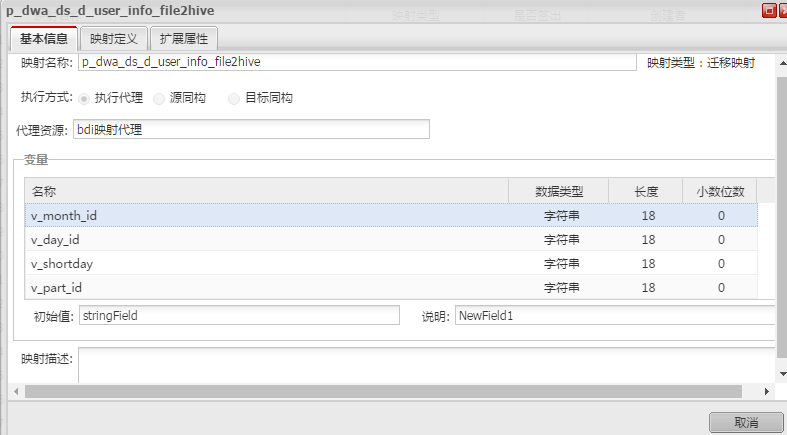


新增接口规范，主要定义接口的文件名、FTP目录、行列分隔符以及是否压缩是否分包、对应的表名字段信息等。



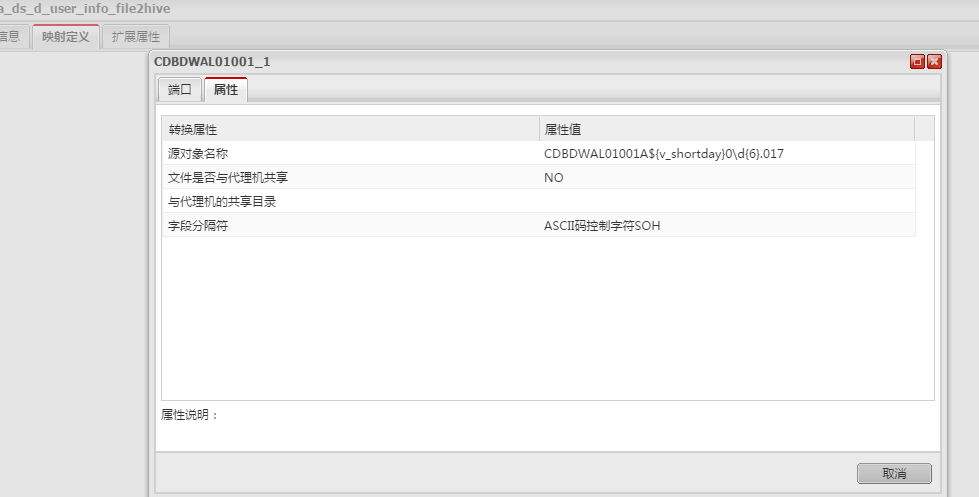
### 2、画映射

新建“迁移映射”p\_dwa\_ds\_d\_user\_info\_file2hive，以p\_表名\_file2hive命名



源选择定义的接口规范，





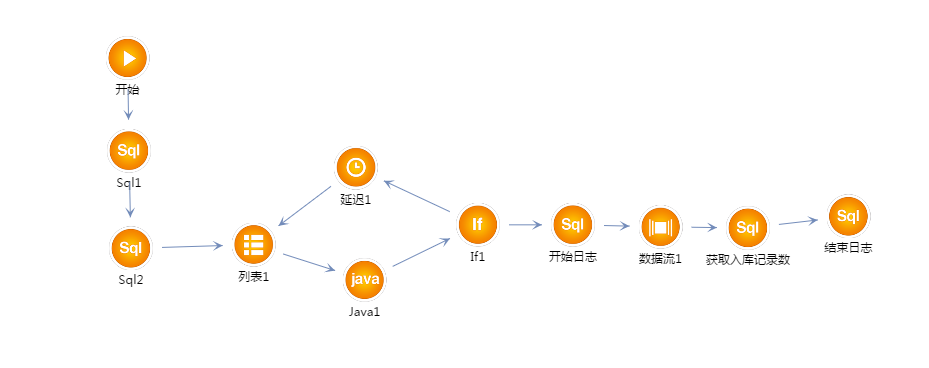
目标表选择相应的hive的表名，字段一一对应





### 3、画流程

Wf\_dwa\_ds\_d\_user\_info\_file2hive



Sql1:获取日期参数

select case when '${v\_acct\_date}' = 'stringField' then

to\_char(sysdate - 1, 'yyyymmdd')

else '${v\_acct\_date}' end v\_acct\_date

from dual

Sql2：获取其他参数

select substr('${v\_acct\_date}',1,6) v\_month\_id,

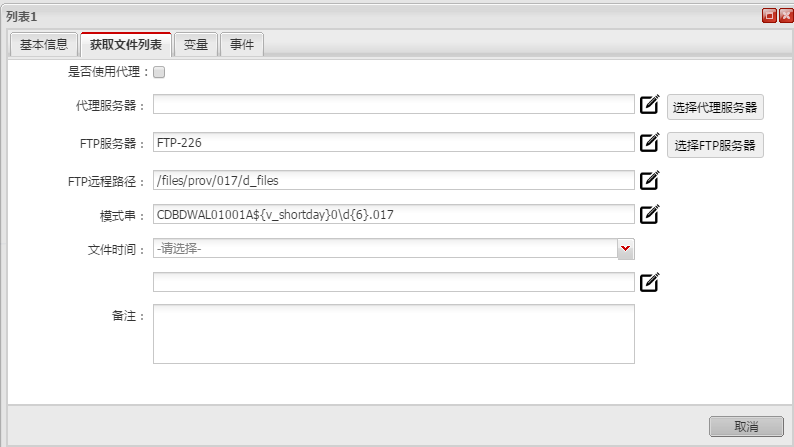
substr('${v\_acct\_date}',7,2) v\_day\_id,

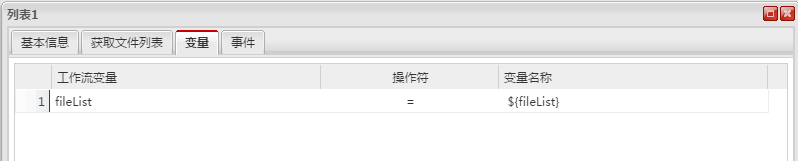
'${v\_acct\_date}' v\_shortday,

to\_char(mod(substr('${v\_acct\_date}',1,6),2)) v\_part\_id

from dual

列表1：获取FTP上的文件列表





Java1：判断文件是否上传完成，根据文件名分包数判断文件个数

boolean ok = false;

System.out.println( "Hello World!" );

List l = fileList; //把获取到的文件名list 赋值给l

if(l!=null&&l.size()>0){

Map map = l.get(0);

String s = map.get("FILE\_NAME").toString();

s = s.substring(s.indexOf(".") - 6, s.indexOf(".") - 3);

System.out.println(s);

int n = Integer.parseInt(s);

if(n==l.size()){

ok = true;

}

}

if(ok){

v\_file\_exist = "1";

}

else{

v\_file\_exist = "0";

} //把ok的值传给需要赋值的变量

If节点:



最后插入开始日志，调用映射，获取入库记录数，结束日志。