在iUAP DI中进行数据连接

数据连接是两个结果集水平方向的按照一定条件合并，连接是关系代数中8中运算的一种。连接类型有内连接、左外连接、右外连接、全连接。根据两个结果集存储的实际情况，还可以分为异构连接和同构连接。在iUAP DI中对上述的数据连接都提供了良好的支持。

1. **同构连接**

如果两个表在同一个数据库中，最好使用SQL进行连接，关系型数据库对连接操作提供了很好的性能优化。同时要注意连接字段上要建立索引，这样可以大幅提高查询速度。

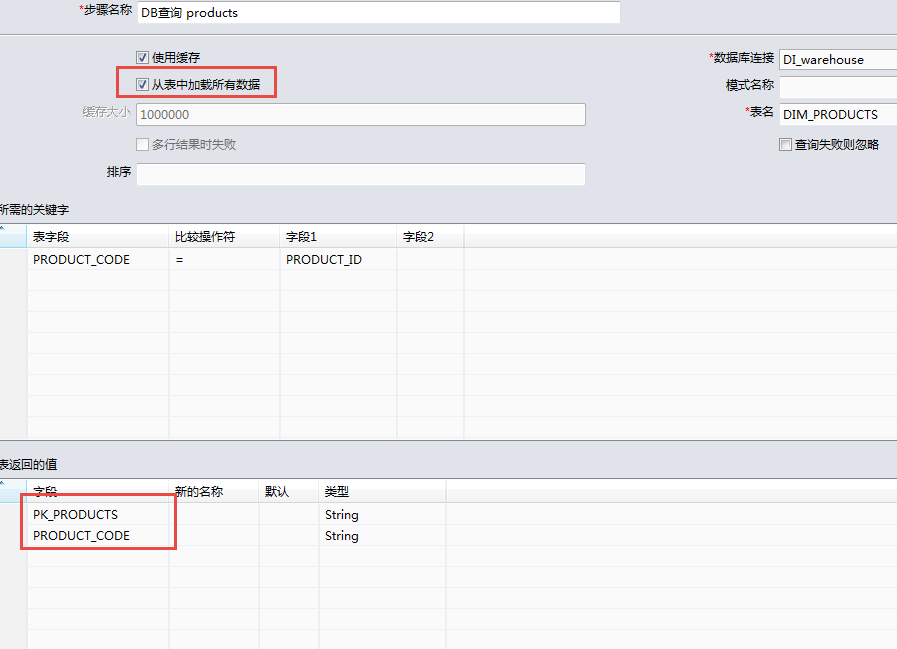
如果两个表在同种类型的数据库，但不是在同一个数据库时，可以使用数据库的连接工具建立数据链接，例如Oracle的DB Link，DB2联邦数据库等。当然也可以使用异构连接，在下文中会有方案描述。

1. **异构连接**

异构连接是指两张表不在同一个数据库的情况，实际情况有很多种，两个表可以在同类型数据但不是同一个数据库上，可以是不同类型的数据上，甚至是在平面文件上。在实践中，两个表的连接可以分为两个大表连接，一张大表和一张小表连接，两张小表连接（这里指的大表是记录数在50万以上的，反之为小表），两张小表连接既可以按照两张大表连接处理，也可以按照一大一小处理，这里不再赘述，这里主要讲解一大一小和两个大表的处理方法。

* 1. “一大一小”连接

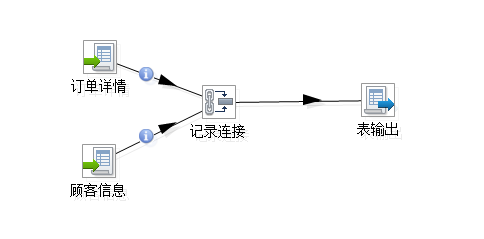
在ETL开发过程经常涉及清洗和补充数据，例如利用编码查询出名称等，此时使用Meta查询或DB查询是一个很好的选择，前者是针对AE元数据的封装，后者更适用于一般数据库。在下例中是DB查询的一个设置截图，其目的是利用编码查询返回其主键，通过DB查询可以是数据流中的字段多出了“表返回的值”的两个字段。如果数据量小于50万，建议使用开启缓存并一次性从表中加载所有的数据。这样可以大幅度提高效率。



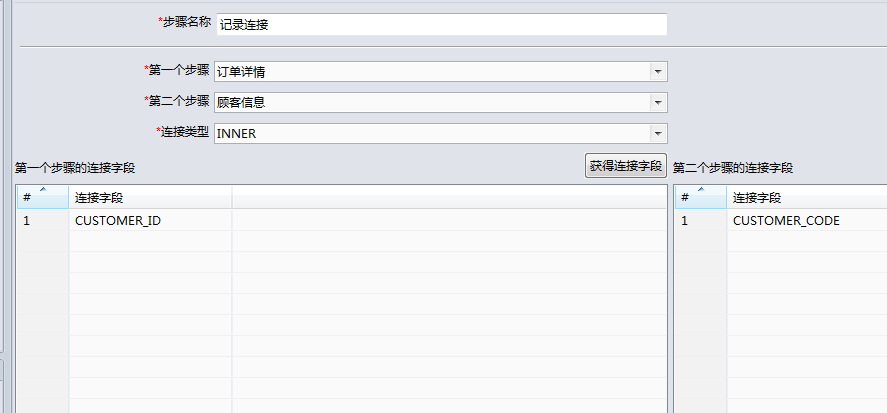
DB查询配置截图

* 1. “两张大表”连接

如果是大表使用数据查询，由于不能缓存全部数据，可能需要多次查询数据库，这样运行可能就效率不高。在这种情况下比较合适使用记录连接就。在下面的DI转换中订单详情和顾客信息都是比较大的数据表（都超过了100万），使用数据连接进行了连接操作。注意，两个表输入必须按照连接字段排序，否则连接的结果不正确。



转换截图



记录连接配置截图