1.进行字段选择时，在弹出窗口的第1、3个标签中尽量填入所有字段，以免出现找不到字段的问题。

2.两个数据流join前必须先进行排序操作。

3.数据流从计算器或join操作流出，并将流入另一join操作之前，需要插入一个字段选择插件，否则会报错。

4.表输入插件中，填写查询语句时，语句末尾不能加分号。

5.sql脚本插件中，填写sql语句时，语句末尾必须加分号，否则不会执行。

6.引用变量时，尽量采用${变量名}的方式引用，以免出现问题。此种引用方式还可以在变量内容后接其他字符串，例如：${key}\_20151217会被解析为：value\_2015217。

7.job中执行多个trans时，如何确定trans执行的先后顺序？

根据连接线的先后顺序执行，如果需要优先执行则先连接从开始到该trans的连线即可。

8.如果源数据中有空值，但是目标数据中不希望出现空值：

直接设置目标数据库各列的默认值为空字符串是不行的，因为只有不对某列赋值才会填入默认值，而NULL是有赋值的，只是赋的是NULL。

在表输入中将NULL值替换为空字符串也是不行的，因为Kettle默认在数据流传输过程中会将空字符串转换为NULL进行传输，因此，到了写入目标表时，传入的是NULL值。

因此，正确的解决方案是：

        首先，在表输入中将NULL值替换为空字符串；

        其次，将kettle的配置文件中配置以下内容：KETTLE\_EMPTY\_STRING\_DIFFERS\_FROM\_NULL=Y

        配置文件位置：/opt/kettle/data-integration/.kettle/kettle.properties

这样，表输入中原来的NULL会被‘’替换，而且在数据流传输中不会变回NULL，插入目标表中即为空字符串。

PS：hive中如何用空字符串替换NULL值？

       coalesce函数，返回列表中的第一个非空元素，如果元素都为空则返回NULL，采用coalesce(列名，'')。