面试总结：

1、Kettle有哪些控件：

表输出，表输入、自定义常量数据、获取系统信息、文本文件输入、从CSV文件输入

XML文件输入、JSON文件输入、excel输出、插入/更新、数据同步、MYSQL批量加载，

SQL文件输出、XML输出。

2、kettle上的锁、×、√是什么意思：

锁：表示无论上一步执行成功还是失败，都一定会执行下一步；

√：表示上一步执行成功才会执行下一步；

×：表示上一步执行失败执行下一步。

3、一个表数据量比较大，怎么提高其查询效率？

建立表分区

4、数据仓库的架构？

传统数仓：贴源层、数仓层、集市层。

贴源层：主要是做数据的抽取，转换、加载，通过调度工具（ctm）按照每天/每月，定期从原系统抽取到ODS层。减少对业务系统的影响。

数仓层：主要是从ODS层同步数据到数据仓库，根据业务的需要对事实表和维度表进行建模（雪花、星型模型）。

DM层主要是为报表展示提供数据支撑，会从dw层同步数据到dm层，同时对指标提供数据或者通过函数计算得到报表需要的数据指标。

建模原则：从业务系统增量抽取、保留时间由业务需求决定、可分表进行周期存储，数据不做清洗转换与业务系统模型保持一致，按主题逻辑划分。

5、项目指标及字段。

这个下来背。

6、hints

7、锁表

8、游标

9、储存过程与函数的区别。