平安智慧城市

1、一个表数据很多，有新增的数据，怎么才能知道新增了那些？

答：插入的时候添加时间撮，根据时间就能知道最新，添加系统时间，根据这个时间筛选；

2、问接上题如果没有时间撮，没有主键，怎么看？

回答不上；

3、说下数据仓库的架构？

数据仓库分三层，

ODS层：面向主题的，数据源中数据经过抽取、清洗、传输，就是ETL之后装入本层，总体上还是按照源头业务系统分类方式而分类的；装入前需要去重、去噪（去掉明显偏离正常水平的银行刷卡信息）、提脏（银行卡被盗十分钟内两笔订单分别在中国、日本刷）、业务提取、砍字段（用于支撑前段，但数据挖掘中不需要的字段）；

DW层：从ODS层中获得的数据按照主题建立数据模型，星型模型、雪花模型；例如研究客户所在区域可以按照IP地址、常刷卡城市进行结合分析，产生数据集；把需要的数据按照事实、维度、指标来进行的；

DM层：数据集层，主要做数据分析和报表数据展示；根据DW变数据按照各种维度或多维度组合把需要查询的一些事实字段进行汇总统计作为单独的列进行存储，满足一定的特性查询及数据挖掘应用；

4、kettle使用的是什么版本？

随便说了个7，说回答的不对

5、具体做的一些项目的细节介绍？

6、都使用过哪些调度工具，如何实现调度？调度适用的场景？

答：定时调度，适用于数据实时更新；

7、为什么要对数据仓库分层？

分层管理可以简化数据清洗过程，相当于把一个工作分层了几个简单的工作，那一层出现问题能更好的发现解决，

诚迈科技

1、对过往项目中业务的了解程度？

介绍过往项目；

2、会数据建模吗？

从ODS层中获得的数据按照主题建立数据模型，有星型模型、雪花模型，例如研究客户所在区域可以按照IP地址、常刷卡城市进行结合分析，产生数据集；把需要的数据按照事实、维度、指标来进行的；但建模目前没有让我们做，都是项目经理在做建模；

3、对建模工具使用的熟悉程度？

DDM

4、对数据库，数据库工具的了解？

常用Orscle数据库，PL/SQL、Kettle、帆软

5、对新业务的理解和学习能力？