文思海辉面试总结

1. 你先自我介绍一下

。。。（前面省略教育背景），我原来公司是做项目外包的，公司有几个项目组，公司比较小，我做的项目里面最长的工期只有六个月，公司承接的业务比较偏向金融这一块。我做的最后一个项目是一个证券公司的驾驶舱项目，主要负责的是报表开发这一块。

1. 你平时的工作内容

因为我做的是报表开发，也会有负责集市层这一块的工作内容。我的主要工作就是写存储过程，数据同步，还有根据需求开发报表将数据可视化。因为资历不够，表结构这一块就没有参与设计。

1. 数据在数据库怎么存放

数据库一般分为三层，贴源层、数仓层、集市层嘛，贴源层主要是同步各个系统的数据，尽量呢不改变数据的原貌，通过一对一的关系同步数据；数仓层的话根据需求将数据按主题分类存放，其中从贴源层同步数据过来数仓层的时候还涉及了数据的清洗等操作；最后的集市呢按我的理解就是数仓层的一部分，根据需求从数仓层取出相应的数据并根据星型模型或者雪花模型建立事实表和维度表来存放数据。

1. 数据溢出怎么解决

可能一次性同步的数据量太大了，那可以考虑分批同步

1. 如何优化

优化的话我们常见的方式就是建索引或者表分区，要么就是按F5看看执行计划找到耗时长的步骤然后考虑优化语句呗

1. 数仓和集市的区别？

就像存放数据的不同一样，数仓层记录了所有的历史数据而且按要求是不能更改的，集市层的话因为是从数仓层取出数据，所以只有数仓的部分数据，而且根据要求可能还要在集市层中计算并存放一些指标以供报表使用

1. 去重方式有多少种？

Diatinct\union\group by \伪列

1. 有两个表，A一个有100条数据，B有70条，B的数据在A中都能找到，请问如何找出剩下的30条？

数据量小的话用in 或者not in都行，数据量大的话用exist,或者用补集的函数也行

1. ABC三个表，AB有一样的字段，BC有一样的字段，请问如何找出A中与C对应的字段？

这个就要用到关联了，先BC关联成一张宽表，再关联A，最后用A中关键字段找

1. 你会HIVE吗？

会，语法结构和SQL类似，但是我们工作中会用的机会不多

1. 你聊一下分布式存储结构

此处自由发挥。。。