杭州博彦科技

1.有没有参加过设计模型，有没有带过团队全项目执行？

2.开发过最长的代码逻辑

3.一个大表+一个小表怎么连接，为什么？

4.Kettle为什么要先排序，再去重？

5.用过kettle哪些组件？

6.一个表50条，另一个10条，left join之后的结果条数范围

上海天正智能（宁波区域）

1.自我介绍

2.做到的项目有哪些指标，哪些表？

3.hive sql跟 saprk sql的区别？、

4.数据倾斜怎么发生的，怎么解决？

5.对于MR的理解？

6.项目难点，怎么解决，思路和工具？

7.项目中，不是自己的问题怎么看待？怎么做，后果如何？