**---北京 中电金信 腾讯会议语音**

最近两年做了哪些项目，主要负责的工作内容，在项目中充当什么角色

介绍一下最近一个项目的整体架构，也就是项目的整体流程

你们是用开源的HIVE平台吗

数仓应该分几层，每层的工作内容是什么

大数据项目的数据从哪里来的，也是oracle吗

数据加载和清洗有哪些规则

去重的关键字是什么？怎么查看哪些数据是重复？去重有哪些方式

源数据抽取过来之后除了去重这一项，还有哪些可以验证数据清洗规则的

你的项目里写到spark-sql对hive脚本进行改写，可以简单说一下吗

Spark-sql和hql有什么区别

Hive优化有哪些

数据倾斜是怎么样产生的，怎么样解决

简单说一下对分区分桶的理解，分区分桶的原理是什么

常用hive的分区方式

Hive的原理是什么

你们大数据项目中用什么调度工具来调度

在hive中，表连接过程中on和where有什么区别

索引失效有哪些原因

哪些函数能导致索引失效

Not in和not exist 有什么区别

说一个对视图的了解

有用过模型吗，知道哪些模型，星型模型和雪花模型的区别

有没有用过银行里面的四大主题模型

平时工作中是沟通多些还是开发多一些

最熟悉的银行业务有哪些

简单说下项目里提到的贷款和存款