易佰网络面试总结

1. 做下自我介绍
2. 了不了解Hadoop，有哪些组件？

了解一些，HDFS分布式文件系统，一个服务器为主服务器，若干个子服务器为子节点，hive用的是类SQL，sqoop实现数据跨库，spark实时数据

（感觉回答的不是特别全面）

1. Hive底层的基本原理

不是特别清楚

1. 报表工具可不可以设定指定的背景颜色

报表工具只是做一些图表展示，做不到那么精细化，工具所能做到的有限，可能有些达不到

（不知道怎么回答）

1. Linux有哪些命令，查内存，CPU，还有一个和内存有关的不记得了

Cd ls vi top free

1. Linux怎么实现两个系统之间取文件

不太清楚

1. SQL调优

首先是查看SQL语句能不能优化，可以优化的先优化一下SQL语句，查看执行计划看消耗，然后建索引看消耗，还有hints优化，指定关联方式或者用并行

1. 表分区有没有用过，追问有没有用过索引分区

表分区用过，主要是按时间还有地区啊这些分区，索引分区没有用过

1. Kettle工具是怎么用的

里面有转换和作业，转换里面有表输入和表输出，可以同步数据，作业的话是把整个过程做成一个流程

1. 有没有编写过存储过程

有过，用的还挺多的

9.有没有想问我的

赛意电话面试

介绍项目

1. 有没有写过存储过程

有写过

1. 讲一下写的存储过程的结构

Create or replace 存储过程 有游标的话定义游标 begin end；中间写逻辑，后面加上一些异常处理，返回多行或者找不到数据

1. 存储过程和函数有没有区别

存储过程没有返回值，函数有返回值

调用方式不一样，存储过程是用begin end 调用，函数是通过select调用，

工作中一些应用场景不一样，增量更新和全量更新用存储过程，一些计算用函数

1. 游标用到循环时，在校验一批数据时，发现数据出错了，想停止这个游标，要怎么停止游标呢

刚开始说的是会报错吧，然后面试官说肯定会报错，继续追问我怎么结束游标，我说end loop，还是不对，后来就没答上来了（百度了答案，在循环中用break跳出就可以了declare @i int
set @i=100
while @i>1
begin
select @i
set @i=@i-1
if @i=90
begin
break
end
end）

1. 有写过程序包吗

有写过，我们一般都是把存储过程和自定义函数放在包里的

1. 全局变量和局部变量有什么区别—没回答出来
2. 另一个包想调这个包的变量，这个变量要声明在哪里另一个包才能调用到呢

写在外面，写在包头里面（应该是回答错了）

1. 会报哪些异常

找不到数据，返回多行，等待资源超时，试图使用一个无效游标

1. union和union all 的区别，哪个性能比较好一点

union会去重，union all 不会去重

1. 处理大数据的时候怎么处理，比如说数据量很大，比如说从某一代游标里面有很大的数据量，想插到另一种表里面，一般是怎么处理的

要用到表分区—不是表分区，从语法方面写怎么写—我的回答是在后面加where条件把他分开来，用union all连起来 分批插入—感觉回答的不对

1. 有做过SQL的性能调优，出现性能有异常的数据量有多大

 回答的是SQL调优的方法

1. 一个存储过程或者是一个程序包特别慢，怎么定位是哪一段慢

用日志定位，

问了好多技术上的问题，前面有的不会，后面就有点慌了，下次不会也没事，要淡定！！！