

爪哇科技

sqoop遇到过什么问题？

sqoop采集怎么保证数据一致性？

hive sql执行原理的？

实时采集碰到过什么问题？

建行一面

flink算子计算10个股票的平均价，大概思路

flink一致性语义

上交所黄埔区

sqoop分隔符用的什么？

hive sql有哪些优化

shell命令查询文件内容 awk

awk是干什么的

查看动态日志

极兔速递

spark内存模型

hive sql如何处理小文件

spark执行流程

spark master分配worker的规则时什么？

taskset中分配task的规则时什么？

hive再第8张表join的时候出现数据倾斜，怎么定位？

paypal (1面+2面)

yarn的client模式和cluster模式

hive行列转换

spark shuffle？

spark宽展依赖，stage怎么划分

sql

1.连续登录三天的用户

2.用户表user，字段id，登录日期，求出

- 每天活跃用户数
- 次日留存用户数
- 次日留存率

3.把以下数据变为一张表，字段为id，value

```
[{"id": 360000300367, "value": null},  
{"id": 46900768, "value": null},  
{"id": 30590838, "value": null},  
{"id": 53094627, "value": null},  
{"id": 53426788, "value": null},  
{"id": 53094667, "value": null}]
```

4.删除一张表里面的数据

5.删除一张表的分区

上交所浦东

1. 统计每个性别的人数sql怎么写？reduce个数多少个？
2. Hive中reduce个数怎么决定的？
3. hive sql替换字符串用什么函数？Hive sql切分字段用什么函数
4. 一张表7个分桶，另一张表5个分桶，这两张表可以join吗？
5. 除了Key，Null，还有什么会导致数据倾斜？
6. Hive的join类型有哪些？（没懂题意）