58面试

sql 采样

什么是有向无环图DAG 的理解

map的4中遍历方法

介绍一下kafka kafka的API写过吗？

手写mr 的join mr中小表join大表 怎么优化

sql 语句 怎样查询非关联字段

项目中 你们的任务是定时跑吗？ 那如果你的数据需要A表的数据 A表今天没有跑完 你怎么解决 （ps 不能退给测试）

foreach 和map d 区别 foreachRDD 用过吗？

你们数据仓库是怎么建的

介绍项目 项目有多少人 怎么分工

linux

马蜂窝

线程池 了解多少 线程池的参数 多线程了解多少

flume 的channel 用到是那种 （）

hashmap的底层实现 链表 hash冲突

项目中按时间抽取数据的意义 以及实现

flume中向kafka拉取 数据 数据丢失怎么办

我设定consume 中从头开始消费kafka的数据 如果consumer 关掉 在启动时 数据怎么消费

数据一致性怎么保证