哗啦啦面试问题总结

1 stage 的划分

mr的流程。hdfs读写流程

spark submint 参数设直 级任务提交流程

produce向kafka中发送数据产生的offset 怎么算（给你传入几条大小的消息 求offset是多少）

项目中遇到的问题 及解决

项目的数据怎么跑的 ，测试过程出现问题怎么查看 （死扣项目）

yarn 的参数 mr参数 hive设置过什么参数

flatMap 算子怎么压平的

reduceBykey 能否替换掉 groupBykey (不能：那就问你 哪些情况不能”)

手写 wordCount scala 分 几个stage