1. KNN和K均值的流程简述和区别联系
2. 从文本词句上看，余弦相似性和Jaccord的区别；余弦相似性计算流程简述
3. HBase中被写入数据如何到达某一个数据点的（Rigion怎么分片并分发 怎么映射到HDFS）；具有容错和恢复的是哪一层，如何做到的
4. AdaBoost简单计算误差率、AdaBoost的优点
5. Spark的窄 宽的概念；判断某一是窄是宽；找出图中Stage的划分的一处错误
6. Aprior计算 ppt原题 找出所有频繁项集；根据（1）中的最大频繁项集，任意找出一个规则，并判断是不是强规则。