一:Hive的本质是什么

1. Hive是分布式数据仓库，同时又是查询引擎，所以Spark Sql取代的只是Hive查询引擎，在企业的实际生产环境下Hive + Spark Sql是目前最为经典的数据分析组合
2. Hive本身就是一个简单单机版本的软件，主要负责:
3. 把HQL翻译成Mapper(s)-Reducer-Mapper(s)的代码；并且可能产生很多MapReduce的job
4. 把生成的MapReduce代码及相关资源打包成为jar并发布到Hadoop集群中且进行运行（这一起都是自动的)
5. Hive本身的架构如下所示：

二:Hive安装和配置实战

1. 我们使用Hive 1.2.1版本
2. Hive默认情况下放元数据的是Derby是单用户的，所以在生产环境下一般会采用支持多用户的数据库来进行Metastore，且进行Master-Slaves主从读写分离和备份
3. SogouQ3.txt的数据进行移动了
4. 外部表的方式，放在hdfs上，外部表关联的时候，只是把元数据放在hive中，删除表的时候是不会被删除掉的。而内部表示把表放在hive上，删除表的时候会删除掉

三.1.项目分析

2.Hive动手实战