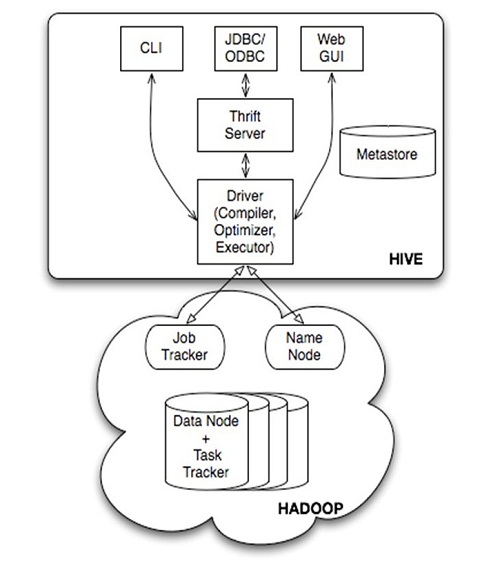
一:Hive的历史价值

1. Hive是Hadoop上的Killer Application,Hive是Hadoop上的数据仓库，Hive同时兼具有数据仓库中的存储引擎和查询引擎的作用，而Spark Sql是一个更加出色和高级的查询引擎，所以在现在企业级应用中SparkSQL+Hive成为了业界使用大数据最为高效和流行的趋势
2. Hive是facebook推出的，主要是为了让不懂Java代码编程的人员也能够通过SQL来驾驭Hadoop集群进行分布式数据的多维度分析，甚至可以通过Web界面来直接操作Hive(通过Hive来驾驭Hadoop)，这对于市场营销人员，系统管理员，数据分析师至关重要
3. Hive的核心是把Hive自己的SQL语言即HQL翻译成为MapReduce代码，然后交给Hadoop集群来执行，也就是说Hive本身是一个单机版本的软件
4. 由于是通过写HQL来完成业务需求的，所以相对于编程MapReduce而言，非常的简易和灵活，能够非常轻易的满足业务的需求和多变的场景；
5. Hive几乎存在与一切使用的大数据的公司中

二：Hive的架构设计



1. 安装Hive的时候，只需要在一台机器上安装，内嵌模式，本地独立模式，远程模式，模式是指Hive后面的引擎是什么模式就是什么模式，Hive中的数据只是元数据，所谓元数据就是Hive要操作的数据的描述信息，该Metastore默认存储在Derby中，但是Derby不能被生产环境下的多用户去使用，所有生成环境下一般最常用的是使用mySql来存储Hive元数据
2. Hive要操作的数据是由Hive的配置文件来决定的，在生产环境下该数据位于HDFS下，其实也就是HDFS上的普通文件而已，只不过按照Hive的方式进行组织
3. 从Hive的角度来看，数据就是一张张的Table,我们的操作就是基于SQL的多维度来查询Table
4. 人们一直努力用Hive来取代传统的数据仓库，但是以失败来告终，因为Hive太慢啦，所以业界目前趋势上黄金组合是Hive(数据仓库的存储引擎）+Spark SQL(分析查询引擎)
5. HQL会被HIVE解释，优化并生成查询计划，一般情况下查询计划会被转化为MapReduce任务
6. 但是形如:select \* from table则不会转化为Mapreduce任务
7. Hive有索引吗?没有!Hbase有索引

