1. 思考Hive数据仓库在Hadoop中起到的作用是什么？

答：（1）可以将结构化的数据文件映射为一张数据库表，并提供完整的SQL查询功能；

1. 通过自己的SQL查询分析需要的内容,这套SQL简称Hive SQL,使不熟悉mapreduce的用户可以很方便地利用SQL语言查询、汇总和分析数据。而mapreduce开发人员可以把自己写的mapper和reducer作为插件来支持hive做更复杂的数据分析。
2. 它还提供了一系列的进行数据提取转化加载，用来存储、查询和分析存储在Hadoop中的大规模数据集，并支持UDF（User-Defined Function）、UDAF(User-Defnes AggregateFunction)和USTF（User-Defined Table-Generating Function），也可以实现对map和reduce函数的定制，为数据操作提供了良好的伸缩性和可扩展性。
3. 简述Hive清洗数据清洗的特点。

答：（1）收集数据到Hadoop hdfs；

1. 使用ETL（MapReduce）进行数据清洗；
2. 更新元数据 target；
3. Hive关联外部表；
4. Hive元数据库是用来做什么的，主要存存储什么信息？

答：通过查询Hive元数据可以确定用户编写SQL和最终操作文件之间的关系。本质上只是用来存储hive中有哪些数据库，哪些表，表的模式，目录，分区，索引以及命名空间。为数据库创建的目录一般在hive数据仓库目录下。