Hive:

1. hive有哪些表，以及表的作用和特点。
2. hive文件的存储格式和特点。

Hive的文件存储格式：

textFile

textFile为默认格式

存储方式：行存储

缺点：磁盘开销大；数据解析开销；hive无法对文件进行合并和拆分sequencefile

二进制文件，以<key,value>的形式序列化到文件中

存储方式：行存储

优点：可分割、压缩，一般选择block压缩，与hadoop api中的mapfile是互相兼容的。Rcfile

存储方式：数据按行分块 每块按照列存储

压缩快 快速列存取

读记录尽量涉及到的block最少

读取需要的列只需要读取每个row group 的头部定义。

读取全量数据的操作 性能可能比sequencefile没有明显的优势

ORC(0.11以后出现)

存储方式：数据按行分块 每块按照列存储

压缩快 快速列存取

效率比rcfile高，是rcfile的改良版本

自定义格式

用户可以通过实现inoutformat和outputformat来定义输入输出格式。

总结：

textfile 存储空间消耗比较大，并且压缩的text 无法分割和合并 查询的效率最低,可以直接存储，加载数据的速度最高

sequencefile 存储空间消耗最大,压缩的文件可以分割和合并 查询效率高，需要通过text文件转化来加载

rcfile 存储空间最小，查询的效率最高 ，需要通过text文件转化来加载，加载的速度最低

orc  效率比rcfile高，是rcfile的改良版本

1. UDF函数