面经：

|  |
| --- |
| Sql的索引：   1. 自己答出的聚簇索引和非聚簇索引。   实现原理是：B+树， |
| Top k问题很适合采用MapReduce框架解决，用户只需编写一个map函数和两个reduce 函数，然后提交到Hadoop（采用mapchain和reducechain）上即可解决该问题。对于map函数，采用hash算法，将hash值相同的数据交给同一个reduce task；对于第一个reduce函数，采用HashMap统计出每个词出现的频率，对于第二个reduce 函数，统计所有reduce task输出数据中的top k即可。  <http://dongxicheng.org/big-data/select-ten-from-billions/> |
| <http://blog.csdn.net/v_JULY_v/article/details/6279498>  大数据十道题 答hash和堆排序一定没错。 |
|  |
| <http://www.knowsky.com/890083.html>  HashMap环路。 |