由于大数据技术太多，如果分太多文件的话可能不好复习，所以整理在一个文件，会分模块。比如hadoop、mapreduce、spark、kafka、flink等。

## 模块1：hadoop&mapreduce&hdfs

1. hadoop1.x中namenode保存的元数据有什么？

文件名、权限、大小、被分成哪些块，注意磁盘上保存的并没有块分布在哪些node上，而是在启动集群的时候datanode上报的，在内存中。

1. secondarynamenode是如何合并fsimage和edit log的？

## 模块2：spark&rdd&sparksqlstreamingstructured

## 模块3：其他技术工具如hive&kafka&zookeeper&ELK等