个人心得

这一次实训第一次接触SVN，第一次接触HDFS，第一次接触Hbase列存储数据库。这次实训中我负责的是hdfs文件上传与下载、数据库的设计与代码编写，以及测试。第一次碰到列式存储的数据库根本摸不着头脑，和以往的关系数据库完全不一样，什么行健、列族、cells以往都没有接触过，更不要说对这种数据库的设计了。最后根据需求模模糊糊建立了多个表，不过最近再来看这个表，发现有挺多问题，例如一个cell它有许多的版本，我可以行健直接设置为视频的MD5码，然后每有一个用户上传了这个视频，就往这个行健为MD5码插入一条记录。不用把行健设计的很复杂。再编写数据库接口代码的时候，不要轻易相信网上的代码，因为可能更新了。直接去github上找到这个模块的源码，看它的测试代码以及源码。

对于hdfs，第一次接触这个，在这个实训中，学会简单的使用它的webhdfs的rest接口，也学会了它的一些基础命令 例如 hadoop fs –rm 之类的命令，也了解了一下hdfs的框架，一个namenode，多个datanode的框架模式。在编写从hdfs上传下载文件到客户端的过程中，一开始不知道外网的机器连不上内网的hdfs服务器。于是找到搭VPN的方法，但是老师拒绝了这种使用方法，那么想的就是把文件传到应用服务器上，在应用服务器上再把文件传到客户端上。这个过程中，本来使用的是webhdfs的rest接口，但是下载文件从这个接口传回来的数据，它每一段传过来的数据都会附带一些其他的数据，这些数据又处理不了，还有上传文件到hdfs，webhdfs接口curl –T filename -i PUT <http://namenode:50070/webhdfs/v1/filename1?op=create>

要先把文件上传到ftp服务器上，在传过去，感觉有点麻烦，最后选择了一个很早之前写完被抛弃的webhdfs模块,可以直接从hdfs下载文件到应用服务器，应用服务器上传文件到hdfs上。然后还可以传hdfs上的输入文件流来实现在线播放的功能。

测试学到的东西有，要搭建好测试环境，测试环境与开发环境分离。然后就是svn代码的更新一定要告诉组员，一定要先更新再去更改再及时更新，不然真的会出现很多问题，很多实现好的，测试好的问题，代码更新后，就崩溃了，给测试带来许多不便之处。