夏季学期企业实训个人总结

周萌 0171123083

在数据时代已经悄然到来，如何应对大数据时代带来的挑战与机遇，是我们当代大学生特别是我们计算机类专业的大学生的一个必须面对的严峻课题。在这次夏季学期中我便选了大数据课程，希望能学一些东西。

在为期3周的大数据实训中，我在老师的教导下学习到了很多知识，在合作中也进行了实践。

在培训的头几天，老师讲了很多关于大数据技术的理论知识，这对以前没有接触过大数据的我来说有点难。第一周的理论教学结束后，我跟着老师的操作，安装了centos7虚拟机、Hadoop、hive等工具，搭建了项目框架。根据老师的讲解，通过简单的实验，我们了解了MapReduce的基本过程，包括从HDFS下载文件、上传文件到HDFS、使用Hadoop进行简单的文本处理。之后，我们开始分组进行大数据实验。经过调查，我们小组决定对世界杯的大数据进行分析，并进行分工。我负责分析结果的可视化开发。当其他团队成员在做MapReduce分析数据时，我也在学习D3和echarts可视化开发。在之前对HTML、CSS等语言研究的基础上，利用D3和echart对MapReduce的分析结果进行可视化，达到了预期的可视化目的，得到了分析结论。通过三周短暂的实训，我对大数据开发的基本流程有了更加清晰的认识，了解了进行一个完整的大数据项目开发需要用到的工具以及需要做的工作，第一次接触了Hadoop，hive，Map Reduce，d3,echarts等相关技术的知识并加以使用，对以后的大数据学习打下了坚实的基础，也树立了更加明确的目标。

在团队合作方面，我深刻的感受到了团队紧密合作的重要性，虽然这次实训只是一个小项目，但对于第一次接触大数据的我们来说，想要在一周多的时间里完成是不可能的，正是因为有了组员之间紧密的协作、相互的沟通，才能使我们更好的、更快的、更高质量的完成这次实训项目！

我在实训中学习到的具体技术主要如下：

1、hadoop集群的搭建：主要的搭建过程分为以下几步：第一步，配置hosts列表；第二步，安装JDK；第三步，免密钥登录配置；第四步，Hadoop配置部署；第五步，将主节点的配置文件复制到从节点；第六步，启动Hadoop集群。在这个过程中我认为需要注意的问题主要有一下几点：

第一点，注意普通用户和root用户的切换。比如配置hosts列表、安装JDK是在root用户下，而免密钥登录配置、Hadoop配置部署都是在普通用户下。在安装的过程中出现了很多由于用户不对而出现错误。这个问题不太大，但在学习前期特别容易忽略。

第二点，就是配置过程中的细节问题。在设置环境变量的时候，将路径要写正确，还有就是在配置文件中是严格区分大小写的，这点也容易出现错误。在部署hadoop集群的时候，需要更改很多的配置文件，在整个配置过程中，出现问题最多的地方就是这些配置文件，所以在配置的过程中需要仔细一点。

2、d3.js和echarts的使用：d3更底层，但也更灵活，而echarts使用更方便。