暑期企业实训个人总结

--白文亮 0171121683

通过这次大数据实训，我学到了很多有关大数据的知识。在这十几天中，老师讲了很多关于大数据技术的理论知识，在老师的教学与文档的帮助下，我先后安装好了CentOS虚拟机以及hadoop，hive等工具。同时也学会了MapReduce工作的基本流程，包括从HDFS下载文件、向HDFS上传文件、利用hadoop进行简单的文本处理和数据分析等。

之后我们开始进行大数据实例开发，我们组选择了分析1950-2014年历届世界杯比赛场次。我负责了数据收集和MapReduce分析数据，通过MapReduce工具，成功分析出了一系列数据，同时，组员也将分析过的结果成功的可视化，达到了预期的可视化目的，并且得到了分析结论。

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，MapReduce，前端开发等相关技术的知识并加以使用，对以后的学习和工作都有极大的帮助，使我受益匪浅。

同时由于的项目开发是分组进行，在开发过程中我们共同努力，，一起进步。组员各司其职，遇到困难共同商讨，相互帮助，最终才能完成这个项目，这使我培养出了团队协作能力，为以后团队学习和工作上有极大的帮助。

这次实训中，我收获了很多，学习了的若干的新知识：

1. 掌握了大数据计算平台相关系统的安装部署，包括CentOS7虚拟机、Hadoop完安装、Hadoop开发插件安装、Hive的安装。
2. 了解并掌握了数据分析之前的数据清洗的主要目的及基本流程。
3. 理解了大数据MapReduce计算模型，掌握了简单的MapReduce程序开发。
4. 学习了前端页面的开发与使用

这次开发过程中，也发现了一些不足之处：

1. 开发过程中没有对数据进行太过复杂的分析，也没有实现预测功能。
2. 分析数据过程中没有将收集的数据全部利用。
3. 审美不行，导致可视化不够美观。
4. 开发技术不行，导致项目复杂度不够。