实验五 数据分析系统的设计与实现

一、实验目的

随着 Hadoop 与 Spark 产生的影响越来越深，各种基于 Hadoop 与 Spark 平台的数据分析系统也随之出现。本次实验要求各位同学利用之前实验以及所学知识实现一个基于 Hadoop、Spark 或其他大数据平台的数据分析系统，理解其中的实现细节以及各种算法的原理。

二、实验平台

1）操作系统：Linux（实验室版本为 Ubuntu17.04，集群环境为 centos6.5）；

2）Hadoop 版本：2.9.0；

3）JDK 版本：1.8；

4）Java IDE：Eclipse 3.8。

6）Spark 版本：实验室版本为 2.1.0，集群环境为 2.3.0；

三、实验内容

自行准备数据集，使用 Hadoop、Spark 或其他大数据分析平台完成一个小的数据分析系统，如网站日志分析系统，电影推荐系统等。

例如网站日志分析系统，通过对服务器日志.log 文件进行分析，清楚地得知用户以什么 IP、什么浏览器、什么时间、什么操作系统访问了你网站的哪一个网页，是否访问成功，同时还能记录网站的流量，时间集中度(24 小时网站访问高峰)等信息。

四、实验要求

基本要求：实现的数据分析系统要有对数据分析结果以及各种功能的图形化、图表化展示界面。

高级要求：在数据分析系统中应用算法解决一些实际问题，例如采用某个推荐系统算法实现产品推荐，或某个挖掘算法产生数据的深度分析结果，算法都是基于大数据系统的并行化算法。

五、实验报告

计算机科学与技术学院 大数据管理与分析 课程实报告

实验题目：		学号：201500000000
日期：2018.3.20	班级：2015 级 1 班/菁英班	姓名：张三
Email：zhangsan@qq.com		

实验目的：
实验软件和硬件环境：
实验原理和方法：
实验步骤：（不要求罗列完整源代码）
结论分析与体会：
就实验过程中遇到和出现的问题，你是如何解决和处理的，自拟 1－3 道问答题：