## 数据挖掘算法与行业应用案例分析系统

## 一、系统技术要求

## 1. 数据挖掘系统平台云桌面设备基本技术要求

1. 工作站基本性能要求：CPU：E5-2603、内存：64G、硬盘：1TB SSD固态硬盘
2. 服务要求：3年上门
3. 采购数量：1台

## 2. 数据挖掘系统平台云桌面软件需求：

1、提供云桌面功能，每台设备可支承20学生虚拟桌面5个科研桌面资源  
2、系统采用基于Openstack的全定制化虚拟化方案，基于X86架构的VT功能的单个CPU核可运行多个虚拟机。虚拟化系统支持多种操作系统，包括Linux和Windows的不同版本，每个虚拟机有私有的硬件，包括网卡、磁盘以及图形适配卡等，能模拟至接近真实电脑的速度。  
3、★系统实现从虚拟机资源到大数据实验与科研全过程统一管理，根据大数据实验过程的特征，解决虚拟机动态部署与调度过程中统一资源管理问题，保证虚拟机的服务质量，提高资源利用效率。实验管理人员可通过页面实现虚拟资源的关联，学生可通过页面无缝访问所分配的虚拟计算资源。  
4、★系统架构包括物理硬件层、虚拟化层、管理层和应用层。结合大数据实验流程的云计算平台管理软件，将所有的硬件层整合，把统一的硬件资源抽象出来组成一个统一的云计算资源池平台。资源管理与实验流程无缝结合。  
 系统对虚拟化平台具有良好的开放性和兼容性，为了适应大数据分析实验与科研工作的要求，必须支持市场主流的Windows版本和Linux操作系统。  
5、系统构建分布式存储系统，用于虚拟机镜像存储和应用数据存储。对于传统的专用存储设备，可以纳入分布式存储系统，进行统一管理。  
6、系统实现虚拟化镜像的自动管理，用户可通过对实验流程的管理自动实现虚拟机镜像的启动、关闭等管理流程。  
7、系统支持虚拟机镜像的扩充与增加，随着今后实验内容的升级，系统可支持虚拟机镜像的增加与修改以支持新的实验环境。  
8、可实现对于物理资源池的资源的有效复用，即创建虚拟机的CPU与内存资源大于物理资源池的资源。  
9、★能够对分配的虚拟机资源进行生命周期管理，可以根据实验流程计算资源的使用时间，当可使用时间结束时，能够自动的释放计算资源，包括处理器、内存、存储空间等。  
10、可支持后续的系统资源扩展，通过增加硬件资源可实现支持实验人数的线性增加。  
11、★支持多用户的并发快速启动，实验所支持的学生虚拟机可同时在3分钟内完成启动。

## 3. 数据挖掘系统平台教师管理系统软件需求

系统采用B/S结构，支持在线注册功能，分为超级管员和普通教师管理功能，系统包括班级管理、实训设置、实训内容管理、实验报告管理、虚拟机管理、云计算集群管理功能；  
详细功能：  
 1.班级管理：包括添加班级、添加学生；可对班级和学生信息进行编辑、删除、清空学生；学生信息里要包括学号、学生姓名、电话、邮箱、所在班级；可批量添加学生信息。  
 2.实训设备：可以按教学大纲及教学情况新建课程；新建课程可以选择参与实验班级、选择要开设的实验、实验开始时间，实验的结束时间等；  
3. 实训内容管理：包括新建实验、新建任务、实验管理、任务管理、模板管理。  
新建实验中包括：实验名称、实验分类、实验目标、实验描述、实验原理、实验报告模板、实验指导书、实验资源等。系统可对所有的实验进行增加、修改、删除等功能。  
 4.实验报告管理：包括报告批改、分数查询、报告下载等功能。  
 5.虚拟机管理：要求可演示实验名称、实验类型、实验时间、可要看详细内容、查看实验虚拟机、关闭实验虚拟机等功能。  
 6.云计算集群管理：包括存储管理、Hadoop管理、Spark管理。要求可显示正常节点数、运行作业数、内存使用情况、CPU使用情况、作业名称、用户名、开始时间、作业计算状态等功能。

7. 采购数量：1套

## 4. 数据挖掘算法演示类实验软件需求：

1、演示类实验：  
1.1、统计与建模方法演示  
1.1.1 探索性数据分析演示  
1.1.2 常用概率分布和渐进性演示  
1.1.3 置信区间和假设检验演示  
1.1.4 线性回归模型演示  
1.1.5 广义线性回归模型演示  
★1.2、数据挖掘方法演示  
1.2.1 分类预测基本流程演示  
1.2.2 数据预处理演示  
1.2.3 分类方法演示  
1.2.4 聚类分析演示  
1.2.5 关联分析演示

## 5. 数据挖掘算法Excel类实验软件需求：

2、Excel类实验  
2.1、Excel数据处理  
2.1.1 Excel基本操作  
2.2.2 Excel数据可视化  
2.2.3 Excel函数与公式  
2.2.4 Excel透视表  
2.2.5 Excel数据分析  
2.2、Excel高级编程  
2.2.1 VBA程序基础  
2.2.2 VBA数据类型  
2.2.3 VBA流程控制  
2.2.4 VBA综合应用

## 6.数据挖掘算法MySQL类实验软件需求：

3、MySQL类实验

3.1 MySQL数据库操作

3.1.1 MySQL数据定义语句

3.1.2 MySQL数据操作语句

3.1.3 MySQL数据查询基础语句

3.1.4 MySQL数据查询高级语句

3.1.5 MySQL视图和索引

## 7. 数据挖掘算法R语言类实验软件需求

4、R语言类实验  
4.1 R语言编程基础  
4.1.1 R语言数据类型  
4.1.2 R语言子集和控制语句  
4.1.3 R语言数据导入导出  
4.1.4 R语言数据可视化  
4.1.5 R语言探索性数据分析  
4.2 R语言统计与建模  
4.2.1 R语言常用概率分布和渐进性  
4.2.2 R语言置信区间和假设检验  
4.2.3 R语言单元线性回归模型  
4.2.4 R语言多元线性回归模型  
4.2.5 R语言广义线性回归模型  
4.3 R语言数据挖掘  
4.3.1 R语言分类预测基本流程  
4.3.2 R语言数据预处理  
4.3.3 R语言决策树分类方法  
4.3.4 R语言高级分类方法  
4.3.5 R语言聚类分析  
4.3.6 R语言关联分析  
★4.4 R语言数据分析综合应用  
4.4.1 R语言建立营销响应模型  
4.4.2 R语言预测股票价格  
4.4.3 R语言建立信用评分模型  
4.4.4 R语言预测门店销售额  
4.4.5 R语言人口教育情况分析

## 8. 数据挖掘算法Hadoop类实验软件需求

5、Hadoop类实验

5.1 Hadoop大数据分析

5.1.1 HDFS基本操作

5.1.2 MapReduce词频统计

5.1.3 MapReduce高级特性

5.1.4 迭代式MapReduce程序开发

5.1.5 Hive基本操作

★5.2 Spark大数据分析

5.2.1 Spark基本操作

5.2.2 Spark词频统计

5.2.3 Spark网络日志分析

5.2.4 Spark实体解析

5.2.5 Spark电影推荐

★5.3 大数据分析综合应用

5.3.1 车辆GPS位置信息分析

5.3.2 超市零售数据分析

5.3.3 微博消息分析

5.3.4 用户网上行为分析

5.3.5 电影评分分析

## 9. 数据挖掘在行业应用案例分析系统需求

## 9.1营销响应模型案例库

数据描述：

数据包含1,932个匿名字段，共145,232个样本以上。

## 9.2预测股票价格大数据案例库：

数据描述：

数据包含上证50组合中从2007年1月1日至20xx年12月31日每日的股票开盘价、最高价、最低价、收盘价、成交量和调整价以上。

## 9.3建立信用评分模型大数据案例库

数据描述：

数据包含1111个字段，共250,000250,000个样本以上

## 9.4车辆GPS位置大数据信息案例库：

数据描述：

数据包含XX市出租车从2010年9月1日到2日的GPS位置数据，包含了7726辆出租车的共33,042,225条位置记录以上

## 9.5 门店销售额大数据案例库

数据描述：

数据包含了14个字段，共1,017,2091,017,209个样本以上

## 9.6人口教育情况分析大数据案例库

数据描述：

数据包含了283个字段，共1,017,209个样本以上，探索教育程度（本科、硕士和博士）对就业情况和收入的影响。

## 二、项目实施与售后服务要求

## 2.1技术支持与售后服务承诺

1. 供货单位应提供三年的免费的技术支持和服务，承诺三年行业实战案例库免费升级维护。
2. 本项目所要求的技术服务包含但不限定于上述技术服务项目，供货方可以根据自己的服务经验及能力提出更多的适合于本项目的技术服务内容。

## 三、项目其他相关要求

1. **为了证明产品完全符合招标要求，使用单位有权要求中标单位在中标公示期内提供所投产品到用户指定地点测试，如果发现所投产品实际功能与招标文件技术要求存在明显不符合，有权取消中标资格；**
2. **为了证明产品功能的真实性，投标单位提供数据挖掘相关的软件著作权证书，如不提供在评委评分是则扣去相应的分值；**
3. **免费提供根据实验环境要求的实验指导书，如不提供在评委评分是则扣去相应的分值；**