

一、投标响应报价明细表

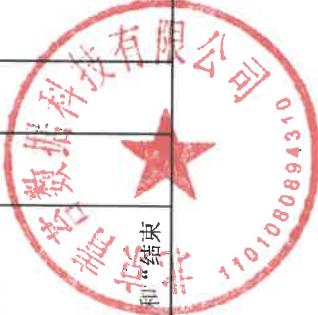
项目名称：江苏工程职业技术学院大数据实验室建设项目

项目编号：JXXM-Z22013

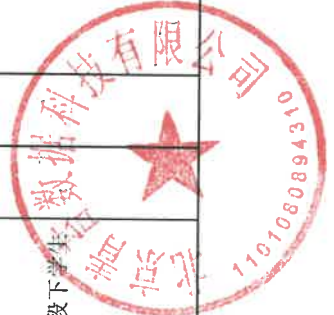
序号	设备名称	技术参数	数量	单位	品牌型号	单价 (元)	合价 (元)	备注
1	大数据实验室后台支撑平台	<p>技术参数</p> <p>△1. 支持 100 个以上的学生同时进行大数据实验支持 100 个学生并发使用公共资源进行大数据采集加工处理、数据复制、数据可视化 and 机器学习建模；支持 100 个学生并发使用私有云大数据平台进行各种大数据实验和机器学习实验。支持混合云架构部署</p> <p>△2. 提供负载均衡功能</p> <p>1) 通过对多台云服务器进行均衡的流量分发调度,消除单点故障提升应用系统的可靠性与吞吐力。负载均衡需要包含面向 4 层 (TCP/UDP) 的传统型负载均衡和面向 7 层 (HTTP/HTTPS) 的应用型负载均衡。</p> <p>2) 提供公网负载均衡；支持 TCP/UDP/HTTP/HTTPS 协议。</p> <p>3) 内网负载均衡；支持 TCP/UDP 协议。</p> <p>△3. 提供离线分布式计算平台公有云服务</p> <p>1) 计算安全性要求：离线分布式大数据计算平台所有计算在沙箱中运行。</p> <p>2) 系统安全性要求： I) 提供多种权限管理方式、灵活数据访问控制策略。 II) 角色分离，基于 ACL 的权限控制，支持列级授权，保证数据安全。</p> <p>3) 易用性要求：</p>	1	套	青苔 V1.0	42000	42000	零税率，免税



6)	预警机制要求 2																																																																
a)	通过简单配置赋予智能监控系统自行决策“是否报警、何时报警、如何报警、给谁报警”的能力，以实现复杂工作流的全链路监控。																																																																
b)	仅需对最终产出任务配置基线即可实现全链路监控，告别对成百上千任务单独设置监控的繁琐。																																																																
c)	仅过滤出可能影响最终任务产出的事件进行报警推送，有的放矢地作出提醒。																																																																
7)	大数据安全管理要求：提供大数据安全管理，提供数据资产识别、敏感数据发现、数据分类分级、脱敏、访问监控、风险发现预警与审计能力。																																																																
8)	敏感数据识别要求：基于自学习的模型算法，自动识别企业拥有的敏感数据，并以直观的形式展示具体类型、分布、数量等信息；同时支持自定义类型的数据识别。																																																																
9)	数据脱敏要求：																																																																



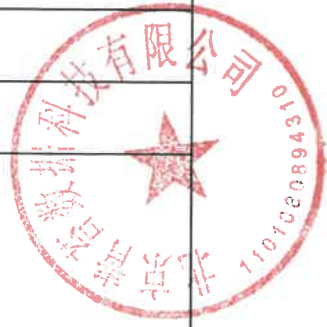
<p>实验”功能:</p>	<p>2) 实验数据: 显示当前实践课程中的所有实践数据列表, 每个实践数据的具体信息包括“格式”、“实验数据名称”、“点击下载”等内容;</p>	<p>3) 教学文档: 显示当前实践课程中的所有实验所对应的实验手册列表, 每个实验手册的具体信息包括“格式”、“实验手册名称”、“查看文档”等内容;</p>	<p>4) 实验报告: 显示当前实践课程中的实验报告列表, 每个实验报告的具体信息包括“实验报告名称”、“更新时间”等内容;</p>	<p>5) 课程测验: 显示已经下发的考试课程列表, 具备创建题库和下发考试功能;</p>	<p>6) 实验问答: 显示当前实践课程中所有学生提交的实验问题, 如: 点击“我要回答”按钮, 跳出“文本编辑器”窗口, 内容填写完成, 点击“提交”按钮;</p>	<p>7) 考勤管理: 教师可以对学生的考勤管理包含序号 学号 姓名 班级 已到 请假 旷课 迟到 早退等信息, 支持对考勤列表的导出功能;</p>	<p>8) 常见问题: 显示当前实践课程中的所有常见问题;</p>	<p>11. 提供教师入口的学生管理</p>	<p>1) 学生搜索: 包含学生个人账号信息及相关的课程信息;</p>	<p>2) 查看学生: 教师可查看某实践课下所有参与学生信息列表;</p>	<p>3) 添加学生: 对已经加入排课计划中实践课, 可导入学生到该实践课中, 邮件发送用户账户信息;</p>	<p>4) 移除学生: 教师可对目前存在在实践课里的学生进行移除;</p>	<p>5) 学生分组: 教师可以根据学生人数进行分组操作;</p>	<p>6) 成绩自动打分: 教师可以根据学生的作业完成情况自动进行手动打分;</p>	<p>7) 成绩手动打分: 教师可以根据学生的平时表现进行手动打分;</p>	<p>8) 平时成绩管理: 对每阶段成绩列表可以进行总览查看及细化查看, 了解该阶段下学生提交明细;</p>	<p>9) 导出成绩管理: 教师可进行成绩导出;</p>	<p>10) 重置学生密码: 将学生的密码进行重置。</p>																																																



12. 提供教师入口的统计报表功能	1) 学生成绩汇总：统计学生成绩支持导出功能；	2) 考勤报表：统计学生的考勤状况以报表形式展现；	13. 提供学生入口的实践课管理	1) 实验列表：显示当前实验课程中的所有实验，每个实验都具有“开始实验”和“结束实验”功能；	2) 实验数据：显示当前实践课程中的所有实践数据列表，每个实践数据的具体信息包括“格式”、“实验数据名称”、“点击下载”等内容；	3) 教学文档：显示当前实践课程中的所有实验所对应的实验手册列表，每个实验手册的具体信息包括“格式”、“实验手册名称”、“查看文档”等内容；	4) 实验报告：显示当前实践课程中的实验报告列表，每个实验报告的具体信息包括“实验报告名称”、“更新时间”等内容；	5) 课程测验：提供对课程相关的习题管理，题型包含单选、多选、判断题等类型；	6) 小组成员：学生可以查看自己小组成员的详细信息（序号 姓名 学号）；	7) 学习成绩：显示当前实验课程的学习成绩；	8) 实验问答：显示当前实验课程中的所有提交的实验问题，如：点击“我要提问”按钮，跳出“文本编辑器”窗口，内容填写完成，点击“提交”按钮；	9) 常见问题：显示当前实验课程中的所有常见问题。	14. 提供学生入口的个人中心功能	1) 个人信息：查看、修改个人信息；	2) 我的考勤：学生可以查看自己的考勤情况；	3) 查看成绩：序号 姓名 学号 课程名称 授课教师 课程成绩 平时成绩；	4) 我的消息：学生可以查看管理员发布的信息；	5) 修改密码：修改实验室密码等；	6) 退出：退出当前账号。																																												



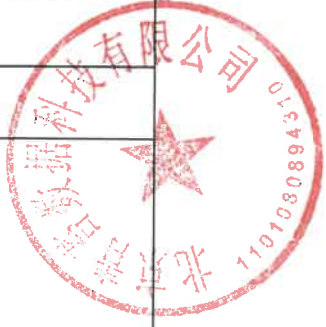
	<p>△支持 Apache 开源组件 需要包含以下组件版本或更高版本： Avro 1.8.2、Flume1.8.0、Hadoop3.0.0、HBase2.1.1、Hive2.1.1、fluc 4.3.0、Impala3.1.0、Kafka 2.0.0、Kudu1.8.0、Solr 7.4、Oozie5.0.0、Spark2.4、Sqoop1.4.7、ZooKeeper 3.4.5 计算引擎支持 计算引擎需要支持 MapReduce, Hive on MapReduce, Hive on Spark, Spark, Oozie 和 Impala 等 提供纠删码支持。 支持纠删码 (Erasure Coding, EC)。EC 是 HDFS 默认三副本的替代方式。当 HDFS 集群启用 EC 后, 不需要额外的数据副本复制, 数据会被条带化成 block 然后加密生成校验 block。如果有 block 丢失或者损坏, HDFS 会在后台使用剩下的原始数据和校验 block 重新生成丢失的部分。这种方式跟 3 副本方式提供相同的数据持久性, 但是会大大降低存储成本。 支持非安全集群到安全集群的数据复制。 支持使用 Backup and Disaster Recovery (BDR) 将数据从非 Kerberos 集群复制到 Kerberos 集群。 支持使用 HDFS 快照差异报告更快地进行增量复制。 支持比较两个 HDFS 快照从而获得需要复制的文件, 进而减少扫描的文件数。如果有大量文件不需要合并, 这可以显著提升复制性能。 提供离线分布式计算集群可视化管理功能。</p>	1	套	青苔 V1.0	1000	1000	1000	零税率, 免税
2	<p>离线分布式计算 平台 (私有云部署)</p>							



3	<p>分布式机器学习平台（私有云部署）</p>	<p>支持国际流行的分布式计算平台使用的分布式机器学习平台需要是以下三个国际流行分布式机器学习平台之一： TensorFlow、Spark、PML5。 支持 Keras 深度学习库 支持由纯 Python 编写 Keras 深度学习库，用户可以使用 Keras 进行简易和快捷的原型设计，从而轻松的完成模型构建。 多种语言支持 支持 Python、R、Swift、JavaScript、C++、Java、Go、Julia、C#等 支持 CPU 和 GPU 计算 支持 CPU 和 GPU 计算，支持分布式计算。 支持 DNN（深度神经网络算法） 支持各种 DNN（深度神经网络算法），DNN 是深度学习的基础。 支持卷积神经网络（CNN，有时被称为 ConvNet）。CNN 是一类包含卷积计算且具有深度结构的前馈神经网络（Feedforward Neural Networks），是深度学习（deep learning）的代表算法之一。 支持 RNN（循环神经网络） 支持循环神经网络（Recurrent Neural Network, RNN），其是一类以序列（sequence）数据为输入，在序列的演进方向进行递归（recursion）且所有节点（循环单元）按链式连接的递归神经网络（recursive neural network）。RNN 是深度学习（deep learning）的代表算法之一。 支持 MNIST 分类器 支持以分布式方式训练完整的 MNIST 分类器 支持深度 Q 网络（DQN） 支持深度 Q 网络（DQN）。深度 Q 网络（DQN）是将 Q learning 和卷积神经网络（CNN）结合在一起，由 Mnih 等人在 2013 年首次提出。 支持 DeepDream 网络</p>	1 套	晋苔 V1.0	1000	1000	零税率，免税
---	-------------------------	---	-----	---------	------	------	--------



	支持 DeepDream 网络, DeepDream 网络是 Google 开源了用来分类和整理图像的 AI 程序 Inceptionism。					
4.1	<p>△大数据基础实践软件包和课程包: 此软件包和课程包主要手把手地教授学员熟悉大数据平台的各种基础操作, 基于一个个的用例 (use cases) 从数据采集、加工和治理, 到数据开发、管理、BI 报表和机器学习等全方位培训学员, 全面提升学员基于图形化方式构建 BI 报表和算法建模能力。</p> <p>课程需提供授课讲义、课件、相关视频教程、教学所需的多行业真实项目案例教程和案例数据, 学生实验所需的实验报告、数据、标准答案、实验步骤以及参考标准等。</p> <p>包含的大数据课程体系需由浅入深, 层次渐进, 涵盖绝大部分主流的大数据技术领域。要求至少能满足以下课程的教学,</p> <p>实验一: 概述</p> <p>1.1 离线分布式计算引擎简介</p> <p>1.1.1 离线分布式计算引擎发展历程</p> <p>1.1.2 离线分布式计算引擎组件</p> <p>1.2 什么是大数据开发套件</p> <p>1.2.1 功能与特点</p> <p>1.2.2 基本概念</p> <p>实验二: 大数据开发套件 :</p> <p>1.1 项目空间</p> <p>1.2 表</p> <p>1.3 视图</p> <p>实验三: SQL 基础 :</p> <p>3.1 插入数据</p> <p>3.2 查询数据</p> <p>实验四: 数据开发管理示例 :</p>	1	套	青苔 V1.0	45000	45000
						零税率, 免税



4.1	使用向导创建商品目录表 (作业)
4.2	使用 DDL 语句创建商品目录表 (作业)
4.3	导入本地数据 (作业)
4.4	数据表管理：商品目录表
4.5	数据同步
实验五：分区操作	
5.1	分区
5.1.1	创建分区表
5.1.2	查看表的详细信息
5.1.3	添加分区
5.1.4	列出已有分区
5.1.5	查看单个分区信息
5.1.6	插入数据到分区表
5.1.7	删除分区
5.1.8	修改分区值
5.1.9	修改分区数据最后修改时间
实验六：SQL 进阶	
6.1	插入数据
6.1.1	实验准备
6.1.2	多路输出 (MULTI INSERT)
6.1.3	输出到动态分区 (DYNAMIC PARTITION)
6.2	查询数据
6.2.1	UNION ALL
6.2.2	子查询
6.2.3	JOIN 操作
6.2.4	MAP JOIN HINT



	<p>7.7 用户自定义函数</p> <p>7.7.1 用户自定义标量值函数 (UDF)</p> <p>7.8 其他函数</p> <p>7.8.1 UUID</p> <p>7.8.2 CAST</p> <p>7.8.3 COALESCE</p> <p>7.8.4 GREATEST</p> <p>7.8.5 LEAST</p> <p>7.8.6 ORDINAL</p> <p>7.8.7 CASE WHEN 表达式</p> <p>7.8.8 IF</p> <p>实验八: TUNNEL 开发</p> <p>8.1 环境部署</p> <p>8.1.1 获取 AK</p> <p>8.1.2 Project 的获取</p> <p>8.1.3 安装 JDK</p> <p>8.1.4 安装客户端 odpscmd</p> <p>8.1.5 安装 Eclipse 开发环境</p> <p>8.2 TUNNEL 命令行上传下载数据</p> <p>8.2.1 Tunnel Upload</p> <p>8.2.2 Tunnel Show</p> <p>8.2.3 Tunnel Resume</p> <p>8.2.4 Tunnel Download</p> <p>8.2.5 Tunnel Purge</p> <p>8.3 TUNNEL SDK 上传下载数据</p> <p>8.3.1 Tunnel SDK 上传数据</p>
--	---



8.3.2 Tunnel SDK 下载数据							
实验九: JAVA SDK 开发							
9.1 介绍							
9.1.1 AliyunAccount							
9.1.2 离线分布式计算引擎							
9.1.3 Projects							
9.1.4 Project							
9.1.5 SQLTask							
9.1.6 Instances							
9.1.7 Instance							
9.1.8 Tables							
9.1.9 Table							
9.1.10 Resources							
9.1.11 Resource							
9.1.12 Functions							
9.1.13 Function							
9.2 使用 ODPS JDBC 接入 ODPS							
9.2.1 数据查询							
9.2.2 批量数据导入							
9.2.3 第三方客户端工具							
9.2.4 程序开发							
9.2.5 常见问题							
9.3 利用 ODPS 计算圆周率 PI							
9.3.1 原理							
9.3.2 表设计与数据上传							
9.3.3 计算圆周率 pi							



10.1 前言
10.2 图加载阶段
10.2.1 输入分解
10.2.2 加载点和边
10.2.3 规约点和边
10.3 编程接口
10.3.1 GraphLoader
10.3.2 VertexResolver
10.4 举例
10.4.1 边类型数据
10.4.2 点类型数据
10.4.3 迭代计算
10.4.4 迭代终止
10.4.5 案例介绍
实验十一：MAPREDUCE 开发
11.1 基本介绍
11.2 测试准备
11.3 代码演示
11.4 资源测试
实验十二：PYTHON SDK 开发
12.1 介绍
12.2 快速开始
12.2.1 项目空间
12.2.2 表操作
12.2.3 表分区



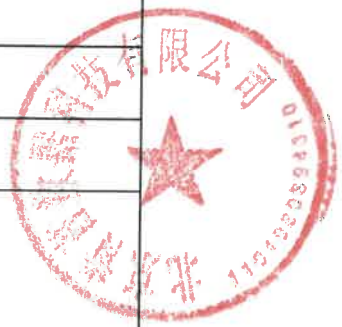
12.2.4 SQL									
12.2.5 Resource									
12.2.6 DataFrame									
12.2.7 Configuration									
实验十三：数据开发管理示例									
13.1 使用向导创建商品目录表									
13.2 使用 DDL 语句创建商品目录表									
13.3 导入本地数据									
13.3.1 从本地文件导入数据到商品目录表									
13.3.2 查看商品目录表数据									
13.4 数据表管理：商品目录表									
13.5 数据同步									
实验十四：机器学习实践：贷款拖欠预测									
14.1 创建基础数据表									
14.1.1 创建贷款拖欠预测表									
14.2 导入基础数据									
14.2.1 导入数据到贷款拖欠预测表									
14.3 贷款拖欠预测建模									
14.3.1 新建实验									
14.3.2 读入数据									
14.3.3 数据预处理									
14.3.4 逻辑回归									
14.3.5 切分比例参数调整									
14.3.6 随机森林									
14.3.7 模型训练结束语									
实验十五：机器学习实践：超市商品购买关联分析									



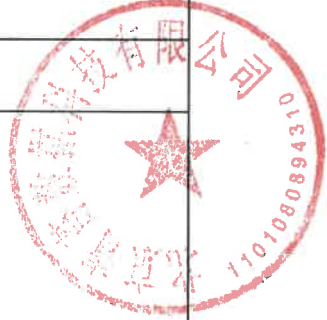
15.1 创建基础数据表					
15.1.1 创建用户购物行为数据表					
15.1.2 创建用户购物行为更新表					
15.2 导入基础数据					
15.2.1 导入数据到用户购物行为数据表					
15.2.2 导入数据到用户购物行为更新表					
15.3 超市商品购买关联分析建模					
15.3.1 新建实验					
15.3.2 读入数据					
15.3.3 数据预处理					
15.3.4 协同过滤					
15.3.5 数据合并					
15.3.6 统计分析					
15.3.7 读入数据					
15.3.8 数据预处理					
15.3.9 统计分析					
15.3.10 模型训练结束语					
实验十六：机器学习实践：学生考试成绩预测					
16.1 创建基础数据表					
16.1.1 创建学生成绩预测表					
16.2 导入基础数据					
16.2.1 导入数据到学生成绩预测表					
16.3 学生考试成绩预测建模					
16.3.1 新建实验					
16.3.2 读入数据					
16.3.3 数据预处理					



16.3.4	逻辑回归					
16.3.5	预测					
16.3.6	评估					
16.3.7	模型训练结束语					
实验十七：人口普查统计分析						
17.1	创建基础数据表					
17.1.1	创建人口普查表					
17.2	导入基础数据					
17.2.1	导入数据到人口普查表					
17.3	统计分析建模					
17.3.1	新建实验					
17.3.2	全表统计					
17.3.3	数据视图					
17.3.4	直方图（多字段）					
17.3.5	数据预处理					
17.3.6	百分位					
17.3.7	散点图					
17.3.8	箱线图					
实验十八：名著阅读文本分析						
18.1	创建基础数据表					
18.1.1	使用 DDL 语句创建全文数据表					
18.1.2	使用向导创建自定义字典表					
18.1.3	使用向导创建中文停用词表					
18.2	导入基础数据					
18.2.1	导入全文数据表数据					
18.2.2	导入自定义字典表数据					



		<p>18.2.3 导入中文停用词表数据</p> <p>18.3 文本分析实验建模</p> <p>18.3.1 新建实验</p> <p>18.3.2 读入数据</p> <p>18.3.3 Split Word</p> <p>18.3.4 词频统计</p> <p>18.3.5 停用词过滤</p> <p>18.3.6 关键词抽取</p> <p>18.3.7 句子拆分</p> <p>18.3.8 文本摘要</p> <p>18.3.9 语义向量距离</p> <p>△此软件包和课程包主要用来讲述并带领用户学习如何基于 Java 和 Python 进行分布式开发，深入学习大数据离线分布式平台，并进一步学习 MapReduce、图计算等相关开发。课程需提供授课讲义、课件、相关视频教程、教学所需的多行业真实项目案例教程和案例数据，学生实验所需的实验报告、数据、标准答案、实验步骤以及参考标准等。</p> <p>包含的大数据课程体系需由浅入深，层次渐进，涵盖绝大部分主流的大数据技术领域。要求至少能满足以下课程的教学：</p> <p>实验一 开发大数据应用</p> <p>1.1 工具安装</p> <p>1.1.1 安装 JDK</p> <p>1.1.2 安装 IntelliJ IDEA</p> <p>1.1.3 安装 Studio 插件</p> <p>1.2 项目空间连接</p> <p>1.3 本地数据上传下载</p> <p>1.3.1 本地数据导入</p> <p>1.3.2 数据导出本地</p>					
4.2	大数据开发实践 软件包和课程包		1	套	青苔 V1.0	45000 45000	零税率， 免税



3.4.1	经纬度								
3.4.2	计算地球上两点间的距离								
3.4.3	使用 Geolash 算法								
3.5	项目开发								
3.5.1	开发环境								
3.5.2	建表并上传数据								
3.5.3	创建离线分布式计算引擎 Java 项目								
3.5.4	编写代码								
3.5.5	本地测试								
3.5.6	离线分布式计算引擎上测试								
	实验四 搭建社交好友推荐系统								
4.1	介绍								
4.2	实现方式								
4.2.1	好友推荐的实现								
4.2.2	数据描述								
4.2.3	期望推荐结果								
4.2.4	MR 介绍								
4.2.5	分析实现								
4.2.6	在阿里云上的实现方式								
4.3	搭建好友推荐系统								
4.3.1	任务说明								
4.3.2	本地开发测试								
4.3.3	云端部署								
	实验五 基于数加实现电商在线交易分析								
5.1	业务场景								
5.2	数据源								



5.3	业务需求及分析					
5.4	准备工作					
5.4.1	进入离线分布式计算引擎项目					
5.4.2	创建在线交易表					
5.4.3	创建用户信息表					
5.4.4	创建交易分析存储表					
5.4.5	创建用户数汇总表					
5.4.6	创建工作流任务					
5.5	数据导入					
5.5.1	导入本地用户信息表数据					
5.5.2	导入本地在线交易表数据					
5.6	数据加工					
5.6.1	数据加工“交易分析”					
5.6.2	数据加工“用户数”					
5.6.3	配置任务调度					
5.6.4	测试工作流					
5.7	数据展现					
5.7.1	添加数据源					
5.7.2	创建数据集					
5.7.3	在线交易分析可视化					
实验六：餐饮外卖大数据分析						
6.1	业务整体介绍					
6.2	业务需求及分析					
6.2.1	源系统数据收集					
6.2.2	数据汇总表建模设计					
6.2.3	选取数据分析工具					



	<p>6.2.4 场景分析及展现层设计</p> <p>6.3 数据装载</p> <p>6.3.1 进入离线分布式计算引擎项目</p> <p>6.3.2 建立 ODS 层模型</p> <p>6.3.3 数据导入</p> <p>6.4 数据加工</p> <p>6.4.1 建立 CDM 层和 ADS 层模型</p> <p>6.4.2 CDM 层小时数据汇总</p> <p>6.4.3 CDM 层日数据汇总</p> <p>6.4.4 CDM 层和 ADS 层月数据汇总</p> <p>6.4.5 套餐打分月数据汇总</p> <p>6.4.6 套餐配送时间段月数据汇总</p> <p>6.5 数据展现</p> <p>6.5.1 获取 Access Key ID 和 Access Key Secret</p> <p>6.5.2 添加数据源</p> <p>6.5.3 创建数据集</p> <p>6.5.4 客户画像</p> <p>6.5.5 套餐打分</p> <p>6.5.6 套餐配送时间段匹配</p>																		<p>零税率， 免税</p>									
4.3	<p>Hadoop 开发实践 软件包和课程包</p>																		<p>青苔 V1.0 套</p>	45000	45000	45000						



<p>数据，学生实验所需的实验报告、数据、标准答案、实验步骤以及参考标准等。 包含的大数据课程体系需由浅入深，层次渐进，涵盖绝大部分主流的大数据技术领域。要求至少能满足以下课程的教学：</p>	<p>章节一：大数据计算框架概述</p>	<p>1.1 HADOOP 概述</p>	<p>1.1.1 Hadoop 发展历史</p>	<p>1.1.2 Hadoop 发行版本</p>	<p>1.1.3 Hadoop 优势</p>	<p>1.1.4 Hadoop 组成</p>	<p>1.2 HDFS 概述</p>	<p>1.2.1 写文件</p>	<p>1.2.2 读文件</p>	<p>1.2.3 可靠性</p>	<p>1.2.4 命令工具</p>	<p>1.3. YARN 概述</p>	<p>1.3.1 ResourceManager</p>	<p>1.3.2 NodeManager</p>	<p>1.3.3 ApplicationMaster</p>	<p>1.3.4 Container</p>	<p>1.3.5 Failover</p>	<p>1.4 MAPREDUCE</p>	<p>1.4.1 读取数据</p>	<p>1.4.2 Mapper</p>	<p>1.4.3 Shuffle</p>	<p>1.5 HIVE 概述</p>	<p>1.5.1 Hive 架构原理</p>																																																																																																		



1.5.2	Hive 优缺点							
1.5.3	Hive 和关系型数据库区别							
1.6	ZOOKEEPER 概述							
1.6.1	Zookeeper 元素介绍							
1.6.2	Zookeeper 特点							
1.6.3	Zookeeper 数据结构							
1.6.4	Zookeeper 典型场景							
1.7	HBASE 概述							
1.7.1	HBase 优势							
1.7.2	HBase 表的特点							
1.7.3	HBase 逻辑模型							
1.7.4	HBase 物理模型							
1.7.5	HBase 架构							
1.7.6	HBase 典型场景							
1.7.7	HBase 和关系型数据库区别							
	实验二: HDFS 操作							
2.1	命令行操作							
2.1.1	查看可执行参数							
2.1.2	查看 HDFS 目录文件							
2.1.3	创建目录							
2.1.4	复制本地文件到 HDFS							
2.1.5	复制 HDFS 文件到本地文件							
2.1.6	剪切本地文件到 HDFS							
2.1.7	追加文件							
2.1.8	显示文件内容							
2.1.9	显示文件末尾							



2.1.10	复制								
2.1.11	移动								
2.1.12	合并下载								
2.1.13	删除文件/文件夹								
2.1.14	删除空目录								
2.1.15	统计文件夹大小								
2.1.16	统计文件系统可用空间								
2.1.17	设置HDFS文件副本数								
2.2	API操作								
2.2.1	获取HDFS文件系统								
2.2.2	文件上传								
2.2.3	文件下载								
2.2.4	创建目录								
2.2.5	删除文件夹								
2.2.6	更改文件名								
2.2.7	查看文件详情								
2.2.8	文件/文件夹判断								
2.3	IO流操作								
2.3.1	HDFS文件上传								
2.3.2	HDFS文件下载								
2.3.3	文件下载合并								
2.4	数据流								
2.4.1	解析文件读取								
2.4.2	解析文件写入								
2.4.3	一致性模型								
	实验三: MAPREDUCE简介								



实验五: ZOOKEEPER 使用	5.1 命令行操作	5.1.1 启动服务器端	5.1.2 启动客户端	5.1.3 显示所有操作	5.1.4 查看当前 znode 中包含内容	5.1.5 查看当前节点数据	5.1.6 创建普通节点	5.1.7 获取节点值	5.1.8 创建短节点	5.1.9 创建带序号节点	5.1.10 修改节点数据值	5.1.11 监听节点 11	5.1.12 监听节点的子节点	5.1.13 删除节点	5.1.14 递归删除节点	5.1.15 查看节点状态	5.2 API 操作	5.2.1 创建 Maven Project	5.2.2 创建 Zookeeper 客户端	5.2.3 创建子节点	5.2.4 获取子节点	5.2.5 判断 znode 是否存在	5.2.6 监听服务器节点动态上下线	实验六: HIVE 的基本概念	6.1 HIVE 的特点																																								



6.2	HIVE 架构						
6.3	HIVE 与 HADOOP 的关系						
6.4	HIVE 的基本操作						
6.5	数据类型						
6.5.1	基本数据类型						
6.5.2	集合数据类型						
6.6	DDL 数据定义	7					
6.6.1	创建/删除/修改/使用数据库						
6.6.2	创建/删除/清空表						
6.6.3	修改表/分区/列						
6.6.4	创建/删除/修改视图						
6.6.5	创建/删除/修改索引						
6.6.6	创建/删除/授予/撤销角色						
6.6.7	授予/撤销权限						
6.7	DML 数据操作						
6.7.1	导入数据						
6.7.2	导出数据						
实验七: HBASE 使用							
7.1	术语说明						
7.1.1	主键						
7.1.2	列族						
7.1.3	时间戳与存储单元						
7.2	HBASE HDFS 目录						
7.2.1	WAL 讲解						
7.2.2	Region 讲解						
7.2.3	HDFS 目录						



7.3 HBASE SHELL 操作	7.3.1 基本操作	7.3.2 表的操作	7.3.3 组命令	7.4 JAVA API 操作	7.4.1 新建 class	7.4.2 判断表是否存在	7.4.3 创建表	7.4.4 删除表	7.4.5 插入数据	7.4.6 删除多行数据	7.4.7 查询某一行数据	7.4.8 获取某一行指定“列族:列”的数据	7.4.9 查询全部数据	7.4.10 自定义 HBase 的 MR	实验八: IMPALA 使用	8.1 IMPALA 的概述	8.1.1 Impala 服务器的优点和不足	8.1.2 配置 Impala 的资源管理	8.1.3 Impala 服务器的架构及组件	8.2 IMPALA 的基本操作	8.2.1 Impala 支持熟悉的 SQL 注释风格	8.2.2 Impala 的数据类型	8.2.3 复杂类型	8.2.4 Impala 的 SQL 运算符	8.2.5 Impala 的 SQL 语句																																																											



8.3 IMPALA 的函数	8.3.1 数学函数	8.3.2 类型转换函数	8.3.3 日期和时间函数	8.3.4 条件函数	8.3.5 聚合函数	8.3.6 Impala 用户自定义函数	8.3.7 使用 HIF UDF 与 Impala	实验九: FLUME 使用	9.1 FLUME 概述	9.2 FLUME 运行机制	9.3 FLUME 采集系统结构图	9.1.1 简单结构	9.1.2 多代理流程 (多个 agent 顺序连接)	9.1.3 多个 Agent 的数据汇聚到同一个 Agent	9.1.4 多路复用流 (多级流)	9.1.5 load balance 功能	9.4 相关概念	9.1.6 Source	9.1.7 Channel	9.1.8 Sink	9.5 KAFKA 消费者编程模型	9.5.1 分区消费模型	9.5.2 组 (Group) 消费模型	9.5.3 两种消费模型对比	9.6 KAFKA 消费者的 PYTHON 和 JAVA 客户端实现																																																																					



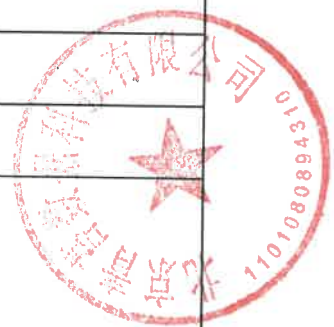
9.6.1	Python 客户端实例讲解					
9.6.2	Python 客户端参数调优					
9.6.3	Java 客户端实例讲解					
9.6.4	Java 客户端参数调优					
9.7	KAFKA 生产者编程模型					
9.7.1	同步生产模型					
9.7.2	异步生产模型					
9.7.3	两种生产模型伪代码描述					
9.7.4	两种生产模型对比					
9.8	KAFKA 生产者的 PYTHON 和 JAVA 客户端实现					
9.8.1	Python 客户端实例讲解					
9.8.2	Python 客户端参数调优					
9.8.3	Java 客户端实例讲解					
9.8.4	Java 客户端参数调优					
	实验十: SQOOP 使用					
10.1	SQOOP 的概述					
10.2	SQOOP 安装					
10.2.1	安装步骤					
10.2.2	示例					
10.3	SQOOP 导入					
10.3.1	全量导入 mysql 表数据到 HDFS					
10.3.2	导入数据: (带 where 条件)					
10.3.3	查询指定列					
10.3.4	导入表数据子集 (query 查询)					
10.3.5	增量导入					
10.3.6	全量导入 mysql 表数据到 HIVE					



	10.4 SQOOP 导出							
	10.4.1 默认模式导出 HDFS 数据到 mysql							
	10.4.2 更新导出 (updateonly 模式)							
	10.4.3 更新导出 (allowinsert 模式)							
	10.5 SQOOP JOB 作业							
	10.5.1 sqoop job 作业介绍							
	10.5.2 创建作业 (- create)							
	10.5.3 验证作业 (- list)							
	10.5.4 检查作业 (- show)							
	10.5.5 执行作业 (- exec)							
	10.5.6 job 的免密输入							
	10.5.7 创建 sqoop job							
	10.5.8 执行 job							
	实验十一: KAFKA 使用							
	11.1 KAFKA 使用背景							
	11.2 相关概念							
	11.2.1 AMQP 协议							
	11.2.2 kafka 支持的客户端语言							
	11.2.3 kafka 架构							
	11.3 KAFKA 消费者编程模型							
	11.3.1 分区消费模型							
	11.3.2 组(Group)消费模型							
	11.3.3 两种消费模型对比							
	11.4 KAFKA 消费者的 PYTHON 和 JAVA 客户端实现							
	11.4.1 Python 客户端实例讲解							
	11.4.2 Python 客户端参数调优							



		<p>11.4.3 Java 客户端实例讲解</p> <p>11.4.4 Java 客户端参数调优</p> <p>11.5 KAFKA 生产者编程模型</p> <p>11.5.1 同步生产模型</p> <p>11.5.2 异步生产模型</p> <p>11.5.3 两种生产模型伪代码描述</p> <p>11.5.4 两种生产模型对比</p> <p>11.6 KAFKA 生产者的 PYTHON 和 JAVA 客户端实现</p> <p>11.6.1 Python 客户端实例讲解</p> <p>11.6.2 Python 客户端参数调优</p> <p>11.6.3 Java 客户端实例讲解</p> <p>11.6.4 Java 客户端参数调优</p>				
4.4	Spark 基础实践 软件包和课程包	<p>△此软件包和课程包主要用来带领学生学习 Spark 的各种相关知识，具体包括 Spark SQL、Spark RDD/Dataset/DataFrame、Spark MapReduce、Spark Streaming、Spark GraphX 等，并进一步学习如何基于 Java 和 Python 进行开源 Spark 的离线分布式开发，Spark 集群环境采用 Cloudera CDH 中部署的企业 Spark 版本，通过手把手的教学，让学生熟悉如何基于 Spark SQL 进行结构化数据管理，如何基于 Spark Streaming 进行流计算的开发、如何基于 Spark GraphX 进行图并行计算等。</p> <p>课程需提供授课讲义、课件、相关视频教程、教学所需的多行业真实项目案例教程和案例数据，学生实验所需的实验报告、数据、标准答案、实验步骤以及参考标准等。</p> <p>包含的大数据课程体系需由浅入深，层次渐进，涵盖绝大部分主流的大数据技术领域。要求至少能满足以下课程的教学：</p> <p>实验一：Scala 概述 Windows 下搭建 Scala 开发环境 Linux 上安装 Scala（实验环境为 centos7） 实验二：Scala 基础</p>	1 套	青苔 V1.0	45000	45000 零税率， 免税



2.10 Scala 模式匹配

- 2.10.1 类型匹配
- 2.10.2 样例类匹配
- 2.11 异常处理

实验三: Scala 进阶

- 3.1 类和对象
- 3.2 Scala 继承
- 3.3 Scala 单例对象
- 3.4 Scala Trait(特征)
- 3.5 隐式转换
- 3.6 Scala 正则表达式
- 3.7 Scala I/O

实验四 SPARK 简介

- 4.1. 什么是 SPARK
- 4.2. SPARK 生态圈
 - 4.2.1 Spark Core
 - 4.2.2 SparkStreaming
 - 4.2.3 Spark SQL
 - 4.2.4 BlinkDB
 - 4.2.5 MLBase/MLli
 - 4.2.6 GraphX
- 4.3. SPARK 的特点
- 4.4. SPARK 2.0 的介绍
- 4.5. SPARK 成功的案例

实验五 SPARK 集群环境的安装与配置

- 5.1 配置 Hadoop 2.8.2 环境



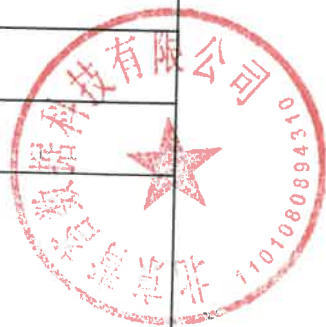
5.2. JDK 配置					
5.3. HADOOP 配置					
5.3.1 文件准备					
5.3.2 环境配置					
5.4. 修改配置文件					
5.4.1 修改 core-site.xml					
5.4.2 修改 hadoop-env.sh					
5.4.3 修改 hdfs-site.xml					
5.4.4 修改 mapred-site.xml					
5.4.5 修改 yarn-site.xml 文件					
5.4.6 修改 slaves					
5.5. HADOOP 启动					
5.6. 配置 SPARK 2.2.3 环境					
5.6.1 下载 Spark2.2.3 安装包					
5.6.2 文件准备					
5.6.3 环境配置					
5.6.4 修改 slaves.template 文件					
5.6.5 修改 spark-env.sh.template 文件					
5.6.6 配置从节点					
5.6.7 Spark 启动					
5.6.8 执行第一个 spark 程序					
5.6.9 启动 Spark Shell					
5.6.10 在 Spark Shell 中编写 WordCount 程序					
5.6.11 在 IDEA 中编写 WordCount 程序					
实验六 在 SPARKSHELL 中编写 WORDCOUNT 程序					
6.1 SPARK 启动					



6.2 执行第一个 SPARK 程序							
6.2.1 Spark 应用模型							
6.2.2 Spark 执行模型							
6.2.3 启动 Spark Shell							
6.2.4 在 Spark Shell 中编写 WordCount 程序							
6.2.5 在 Eclipse 中编写 WordCount 程序							
实验七 RDD 介绍与使用							
7.1. RDD 基础							
7.1.1 RDD 的属性							
7.2. 创建 RDD							
7.3. RDD 的操作							
7.3.1 转化操作							
7.3.2 行动操作							
7.4. 向 SPARK 传递函数							
7.4.2 Java							
7.5. 常见的 RDD 转化操作和 ACTION 操作							
7.6. RDD 的缓存机制							
7.7. RDD 的 CHECKPOINT (检查点) 机制: 容错机制							
实验八 键值对的操作							
8.1. 创建 PAIR RDD							
8.2. PAIR RDD 的转化操作							
8.3. 聚合操作							
8.4. 数据分组							
8.5. 连接							
8.6. 数据排序							
实验九 数据读取与保存							



9.1. 文本文件							
9.2. JSON							
9.3. 逗号分隔值与制表符分隔值							
9.3.1 SequenceFile							
9.3.2 对象文件							
9.3.3 Hadoop 输入输出格式							
9.3.3 文件压缩							
9.4. 文件系统							
9.4.1 本地/“常规”文件系统							
9.4.2 Amazon S3							
9.4.3 HDFS							
9.5. SPARK SQL 中的结构化数据							
9.5.1 Apache Hive							
9.5.2 JSON							
9.6. 数据库							
9.6.1 Java 数据库连接							
9.6.2 Cassandra							
9.6.3 HBase							
9.6.4 Elasticsearch							
实验十 SPARK SQL							
10.1. 连接 SPARK SQL							
10.2. 在应用中使用 SPARK SQL							
10.2.1 初始化 Spark SQL							
10.2.2 基本查询示例							
10.2.3 SchemaRDD							
10.2.4 缓存							



10.3.	读取和存储数据						
10.3.1	Apache Hive						
10.3.2	JSON						
10.3.3	基于RDD						
10.4.	JDBC/ODBC 服务器						
10.4.1	使用 Beeline						
10.4.2	长生命周期的表与查询						
10.5.	用户自定义函数						
10.5.1	Spark SQL UDF						
10.5.2	Hive UDF						
10.6.	SPARK SQL 性能						
10.6.1	性能调优选项						
	实验十一 SPARK STREAMING						
11.1.	SPARK STREAMING 简介						
11.2.	一个简单的例子						
11.3.	架构与抽象						
11.4.	转化操作						
11.4.1	无状态转化操作						
11.4.2	有状态转化操作						
11.5.	输出操作						
11.6.	输入源						
11.6.1	核心数据源						
11.6.2	附加数据源						
11.6.3	多数据源与集群规模						
11.7.	/7 不间断运行						
11.7.1	检查点机制						



11.7.2	驱动器程序容错					
11.7.3	工作节点容错					
11.7.4	接收器容错					
11.7.5	处理保证					
11.8.	STREAMING 用户界面					
11.9.	性能调优					
11.9.1	批次和窗口大小					
11.9.2	并行度					
11.9.3	垃圾回收和内存使用					
	实验十二：王者荣耀英雄分析	5				
12.1	简介	5				
12.2	项目分析	6				
12.2.1	启动 spark-shell	6				
12.2.2	读取数据并展示 (read 和 show)	7				
	实验十三：货品交易分析					
13.1	数据说明					
13.2	加载数据					
13.2.1	加载 tbStock					
13.2.2	加载 tbStockDetail					
13.2.3	加载 tbDate					
13.2.4	注册表					
13.3	货品交易分析					
13.3.1	计算所有订单中每年的销售单数、销售总额					
13.3.2	计算所有订单每年最大金额订单的销售额					
13.3.3	计算所有订单中每年最畅销货品					



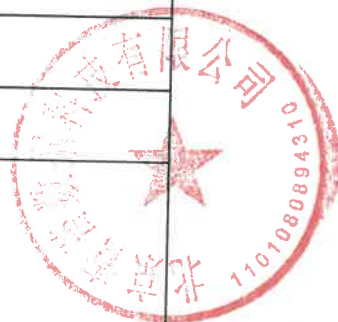
4.5	<p>电商广告精准营销包和课程</p> <p>△此软件包和课程包主要带领用户学习第三代电商广告精准营销的相关技术，在课程中，学员将会学习如何对每个用户进行360度画像；对某用户个性化推荐，预测客户打开推荐链接的点击通过率（CTR）；预测打开推荐链接的客户购买该商品转化率是多少（CVR）；根据预测结果，对高购买倾向客户进行精准营销推荐。</p> <p>课程需提供授课讲义、课件、相关视频教程、教学所需的多行业真实项目案例教程和案例数据，学生实验所需的实验报告、数据、标准答案、实验步骤以及参考标准等。</p> <p>包含的大数据课程体系需由浅入深，层次递进，涵盖绝大部分主流的大数据技术领域。要求至少能满足以下课程的教学：</p> <p>实验一：电商领域实践：精准营销概述</p> <ol style="list-style-type: none"> 1.1 业务理解 1.2 原始数据文件 <ol style="list-style-type: none"> 1.2.1 用户信息数据文件 1.2.2 推荐分析数据文件 1.2.3 用户行为数据文件 1.2.4 待预测用户信息数据文件 <p>实验二：电商领域实践：精准营销数据分析处理</p> <ol style="list-style-type: none"> 2.1 创建基础数据表 <ol style="list-style-type: none"> 2.1.1 创建用户信息表 2.1.2 创建推荐分析表 2.1.3 创建用户行为表 2.1.4 创建待预测用户信息表 2.2 导入基础数据 <ol style="list-style-type: none"> 2.2.1 导入数据到用户信息表 2.2.2 导入数据到推荐分析表 2.2.3 导入数据到用户行为表 2.2.4 导入数据到待预测用户信息表 	1	套	青苔 V1.0	45000	45000	零税率，免税
-----	---	---	---	---------	-------	-------	--------



		<p>2.3 更新用户信息</p> <p>2.3.1 创建用户信息更新表</p> <p>2.3.2 向用户信息更新表中插入数据</p> <p>2.3.3 查询导入到用户信息更新表的数据</p> <p>2.4 更新推荐状态</p> <p>2.4.1 创建推荐分析更新表</p> <p>2.4.2 向推荐分析更新表中插入数据</p> <p>2.4.3 查询导入到推荐分析更新表的数据</p> <p>2.5 用户信息推荐分析宽表状态</p> <p>2.5.1 创建用户信息推荐分析宽表</p> <p>2.5.2 向用户信息推荐分析宽表中插入数据</p> <p>2.5.3 查询导入到用户信息推荐分析宽表的数据</p> <p>实验三：电商领域实践：数据特征工程</p> <p>3.1 数据特征工程</p> <p>3.1.1 新建实验</p> <p>3.1.2 读数据表</p> <p>3.1.3 SQL 脚本</p> <p>3.1.4 全表统计</p> <p>3.1.5 直方图</p> <p>3.1.6 类型转换</p> <p>实验四：电商领域实践：逻辑回归二分类模型</p> <p>4.1 逻辑回归二分类模型</p> <p>4.1.1 归一化</p> <p>4.1.2 拆分</p> <p>4.1.3 逻辑回归二分类</p> <p>4.1.4 预测</p>			
--	--	--	--	--	--



4.1.5	评估					
	实验五：电商领域实践：模型参数调整					
	5.1 切分比例参数调整					
	5.2 最小收敛误差调整					
	实验六：电商领域实践：随机森林模型					
	6.1 随机森林模型					
	6.1.1 拆分					
	6.1.2 随机森林					
	6.1.3 预测					
	6.1.4 评估					
	实验七：电商领域实践：GBDT 二分类模型					
	7.1 GBDT 二分类模型					
	7.1.1 GBDT 二分类					
	7.1.2 预测					
	7.1.3 评估					
	实验八：电商领域实践：模型参数调整					
	8.1 最大叶子数调整					
	8.2 叶节点最少样本数调整					
	实验九：电商领域实践—线性支持向量机模型					
	9.1 线性支持向量机模型					
	9.1.1 线性支持向量机					
	9.1.2 预测					
	9.1.3 评估					
	实验十：PS-SMART 二分类模型					
	10.1 PS-SMART 二分类模型					
	10.1.1 PS-SMART 二分类					



		<p>10.1.2 预测</p> <p>10.1.3 评估</p> <p>实验十一：推荐预测</p> <p>11.1 模型训练结语</p> <p>11.2 进行推荐预测</p>						
5	<p>软件包 MA 服务、</p> <p>培训服务和现场</p> <p>运维服务</p>	<p>提供的各个产品软件包、相关组件和服务，具体包括：大数据实验室后台支撑平台、离线分布式计算平台（私有云部署）、分布式机器学习平台（私有云部署）、大数据基础实践软件包和课程包、大数据开发实践软件包和课程包、Hadoop 开发实践软件包和课程包、Spark 基础实践软件包和课程包、电商广告精准营销软件包和课程包、服务器、交换机、机柜、UPS 稳压电源以及实验室所需的弹性计算服务器、带宽、离线分布式计算资源和数据集成资源等，对应订单的软件产品包自交付之日起，须包含有【三年】的维护服务，维护服务具体包括：缺陷修改、故障及漏洞解决和版本升级等。</p> <p>供各个软件包各个模块的培训，1000 课时*人次； 3 年内有效</p>	4	年	青苔	0	0	
			合计（元）269000 元					

投标供应商：（盖章）北京青苔数据科技有限公司

法定代表人或被授权人：（签字或盖章）

日期：2022 年 9 月 5 日



- 注：1、本表式，可根据实际报价的明细需要自行添加。
- 2、如果不提供详细分项报价将视为未实质性响应招标文件要求。