## 湄洲湾职业技术学院大数据产教融合实训平台产品技术参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **项目分类** | **货物名称** | **技术参数要求** | **单位** | **数量** |
| 大数据云实验平台 | 云实验平台 | **(一)开发技术**  1.系统基于Java EE标准，使用spring技术栈开发。  2.前端展示使用Bootstrap框架提供良好的响应式布局能力。  3.使用RabbitMQ进行业务解耦，提高系统吞吐量。  4.使用Mysql5.6存储数据，保证数据的安全，可靠。  5.实验环境使用Docker作为不同实验环境的技术支撑。  **(二)功能要求**  **一、云环境管理**  1.B/S架构：采用B/S架构，兼容谷歌chrome、火狐等支持HTML5技术的现代浏览器。支持平台用户使用Web页面（B/S架构）方式访问容器桌面。  2.容器化实验环境：支持基于Docker的容器化实验环境，提供轻量级、快速响应的实验环境，满足基于Linux操作系统的实验。  3.裸机配置管理：支持配置容器化服务器的CPU、内存等资源使用率限制。  4.网络配置管理：支持IP地址段配置，可查看具体IP地址的占用状态。  5.图形化资源监控:提供图形化的云实验室平台资源监控页面，包括服务器数量、CPU、内存、存储等资源的使用情况。  6.服务器监控:提供服务器监控功能，以图表的形式展示每台服务器的IP地址、状态、CPU使用情况、内存使用情况。  7.★Docker管理：提供Docker管理功能，可查看Docker主机状态；镜像的查看、检索、删除操作；容器查看、检索、删除。  8.系统资源回收：系统后台自动监控符合回收条件的容器化环境，释放IP地址，减轻服务器负载。回收条件包括：已过期的实验环境、教师定制实验环境时产生的容器、已过期的学生申请的实验环境。  9.日志管理：支持服务器异常日志的查询、检索、删除。有效帮助管理员快速定位系统问题。  **二、教学管理**  1.实验环境管理：系统支持三种实验环境的分类管理,由系统预置的基础实验环境、教师基于基础实验环境自定义的在线定制实验环境、学生基于基础实验环境申请的预约实验环境组成。  2.★在线定制实验环境：教师基于基础实验环境自定义，定义后的实验环境同基础实验环境一样，可以用于学生实验使用。  3.★预约实验环境：学生基于课程相关实验环境申请预约。由学生发起申请，经由管理员、教师审批通过，设置可使用时长后，供学生使用。  4.课程管理：支持管理员/教师开课及创建新课。支持对系统内的课程进行管理、检索、分类。支持对课程进行创建、修改、删除。支持在课程下依据实验模板关联实验，配置实验课相关信息。  5.实验模板管理：支持管理员为课程配置实验模板，可设置相关的实验文档、配置多个实验环境、实验报告。  6.★多种实验环境分配：支持至少3种实验环境分配机制，应对不同类型的实验。实验环境分配的最小粒度为用户组,按组分配的实验环境属于组内所有成员共有的实验资源。支持控制每一个实验环境的CPU、内存资源的限额。支持为每一个实验环境配置不同的实验界面,至少支持SSH界面、VNC图形界面、CloudIDE界面以及Notebook界面中的三种。  7.实验报告模板管理：支持在线编写或者上传word、excel等格式文件作为实验报告模板管关联至实验供学生使用。  8.班级和授权管理：支持单独或成批方式进行班级信息的添加和导入。为学生分配班级，从而建立教师与课程和授课班级的关系。  9.软件资源库管理：支持上传实验资源如软件安装包等至软件资源库，学生可以在实验环境中下载使用。  10.课堂答疑：支持学生与教师基于课堂答疑功能进行互动，留存记录。  11.批阅实验报告：教师可以查看学生提交的实验报告并进行打分，下载留存。  **三、实验管理**  1.实验情况统计：支持教师可以教学数据的统计汇总、查看运行中的实验以及各个用户组实验环境启动情况。  2.远程协助：支持教师通过浏览器直接进入学生的实验环境界面，远程协助学生解决问题，实时查看学生实验情况。  3.管理实验资源：支持教师一键下课，关闭所有学生的实验环境，支持教师关闭课程，释放所有的实验资源、IP地址。  4.★课程连续性管理：教师可以创建不同类型的课程；连续性课程：在课程有效时间内，学生的实验环境（含操作）将一直被保存，并且可以保存至少两个实验的回滚节点；在当前实验课时结束时，实验坏境会记录最后实验状态并保存到硬盘，释放此实验所占用的全部计算资源；保存的实验环境不会被删除，当下一课时被开启时，系统会自动开启之前的实验环境，保证课程的连续性。非连续性课程：此课程下的所有实验的实验环境，每次结束实验课后会被销毁，节省资源。  5.限定实验时间：支持实验环境在指定时间内使用，超出时间回收实验环境，中途可以进行延时操作。  6.实验环境连接切换：支持多实验环境的场景，在同页面下可以进行切换实验环境连接。  7.实验界面切换：支持多实验界面的场景，在同页面下可以进行切换实验界面连接。  8.实验页面可拖动：支持实验页面内拖动，调整实验文档、实验操作部分页面占比。  9.剪切板功能：支持基于剪切板功能，进行学生本地机器与实验环境剪切板通信。  10.代码下载：支持学生下载实验环境内代码至本地机器。  11.★实验环境生命周期管理：支持学生对实验环境的生命周期管理：初始化、保存回滚点、回滚、关闭等操作。  12.查看实验记录：支持查看实验上机记录。  13.填写实验报告：支持根据指定的实验报告模板提交实验报告。  **四、用户管理**  1.用户与角色管理：平台提供了不同的角色与权限，包括资源管理员、授课教师、学生三种角色。其中，资源管理员可以管理资源池、为整个平台建立用户及分配权限，查看整体资源分配情况；授课教师可以对授课进行管理、分配授课时间等；学生可以参加已被分配的在线课程、申请自定义实验环境、进行在线实验、记录实验报告、查看实验结果等。  2.用户组管理：支持用户组的管理，实现逻辑单元的用户管理。实验课授权的最小粒度即为用户组。  3.用户单点登录：平台支持基于CAS协议的单点登录功能，进行用户的统一身份认证。可以快速接入学校已存在的用户体系。  **（三）实施安装**  供货方应提供详细的系统设计方案和实施文档。完成软件的安装调试工作，负责解决系统集成中的相关技术问题。  **（四）技术培训**  供货方在项目实施完毕后在用户指定地点对使用用户及网管人员组织现场培训，时间不低于1天，使用能够掌握所提供产品的操作与基本维护；提供对产品的长期保修、维护、技术支持服务。  **（五）售后服务**  在服务及保修期内，提供产品7\*8小时的电话技术服务，根据实际情况提供远程技术支持服务；如遇系统出现故障，需要现场服务时，供货方专业技术人员根据目的地远近在48小时内到达现场服务。 | 套 | 1 |
| 大数据云实验平台（项目资源包） | Hadoop | 1. **项目需求**   通过此实验可以识记Linux基本命令，识记SSH免密登录的原理以及进行SSH免密登录的原因，理解Hadoop集群的工作原理，运用SSH工具进行集群中各个主机的免密登录配置，运用Hadoop集群安装软件进行集群的安装与配置。   1. **项目内容**   搭建由4个节点组成的一个Hadoop分布式集群，在一个主节点上部署NameNode服务，在三个子节点上分别部署DataNode服务，启动或关闭整个集群，在客户端运行基于集群部署的实例程序。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| HDFS | 1. **项目需求**   实验的目的是帮助学员了解HDFS设计思想、基本概念，了解HDFS体系结构及数据存储策略，掌握HDFS工作原理，熟练完成HDFS程序的编写，熟悉HDFS Shell接口，掌握小文件处理。   1. **项目内容**   HDFS实验手册--基础理论、HDFS实验手册--HDFS Shell、Hdfs实验手册--上传文件、HDFS实验手册--读写文件内容、HDFS实验手册--hdfs维护、HDFS实验手册--小文件处理   1. **资源参数**   （1）实验手册6个  （2）配套实验6个 | 套 | 1 |
| MapReduce课程 | 1. **项目需求**   实验的目的是帮助学员了解Mapreduce设计思想、掌握Mapreduce框架的执行过程、能够独立完成Mapreduce程序的编写、掌握Mapreduce Combiner、数据去重、排序、TopK等方法。   1. **项目内容**   MapReduce实验手册--MapReduce理论基础、MapReduce实验手册--词频统计、MapReduce实验手册—Combiner、MapReduce实验手册--数据排序、MapReduce实验手册--两数据集比较、MapReduce实验手册--数据去重、MapReduce实验手册—TopK、MapReduce实验手册--倒排索引MapReduce实验手册--MapReduce连接。   1. **资源参数**   （1）实验手册9个  （2）配套实验9个 | 套 | 1 |
| Hive基础教学和实践 | 1. **项目需求**   实验的目的是帮助学员了解Hive的作用、工作原理、工作场景，掌握Hive数据仓库的HQL使用方法。   1. **项目内容**   Hive数据仓库--hive理论基础、Hive数据仓库--Hive安装配置、Hive数据仓库--hive数据定义、Hive数据仓库--Hive数据操作、Hive数据仓库--用户自定义函数、Hive数据仓库--用户行为分析。   1. **资源参数**   （1）实验手册6个  （2）配套实验6个 | 套 | 1 |
| Kafka消息队列 | 1. **项目需求**   实验的目的是帮助学员了解kafka消息队列的作用及工作原理，掌握kafka单节点及集群的安装配置，能够深入了解kafka中Producers、broker、consumer、topic、Partition、message等各个组件，能够使用kafka进行消息的发布与订阅，能够进行kafka的Java编程。   1. **项目内容**   完成单节点-单代理配置、完成单节点-多代理配置、完成多节点-多代理集群配置、完成主题操作、使用Java客户端发布、使用消息、完成kafka+flume+hdfs实时日志流系统实验。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| Scala类实验包 | 1. **项目需求**   实验的目的是能够使学员了解Scala基础知识、能够搭建scala开发环境、掌握Scala数据类型和变量、控制和循环语句、字符串与数组、函数与异常、Scala集合、类和对象、文件操作，为后续的Spark学习打下坚实的基础。   1. **项目内容**   Scala环境配置、Scala基础语法、Scala字符串与数组、Scala函数、Scala集合、Scala类和对象、Scala文件操作。   1. **资源参数**   （1）实验手册7个  （2）配套实验7个 | 套 | 1 |
| Spark计算框架 | 1. **项目需求**   实验的目的是帮助学员掌握SparkRDD的特性、编译和运行Spark技术应用程序、理解Spark内核架构、运行原理等内容,为编写Spark应用程序提供所需要的知识和技能。   1. **项目内容**   Spark理论基础、Spark安装部署、Spark编程基础、Spark SQL基础、Spark SQL多种数据源的处理、Spark streaming基础。   1. **资源参数**   （1）实验手册6个  （2）配套实验6个 | 套 | 1 |
| 数据库 | 1. **项目需求**   通过此实验的学习使学员能够独立地安装及创建数据库，掌握常用的函数及SQL语句，包括表结构的设计及数据的增删改查操作、各种查询操作、内置函数、触发器等。   1. **项目内容**   数据库实验手册-MySQL安装配置、数据库实验手册-数据库和数据表管理、数据库实验手册-用户和权限管理、数据库实验手册-简单查询、数据库实验手册-条件查询和排序、数据库实验手册-统计查询、数据库实验手册-多表连接查询、数据库实验手册-子查询、数据库实验手册-MySQL内置函数、数据库实验手册-容灾备份、数据库实验手册-MySQL存储过程、数据库实验手册-触发器、数据库实验手册-自定义函数。   1. **资源参数**   （1）实验手册13个  （2）配套实验13个 | 套 | 1 |
| 数据可视化 | 1. **项目需求**   实验的目的是帮助学员使学员能够了解Echarts特性、了解Echarts API，掌握常用的个性化图表的样式、异步数据加载和更新、能够独立完成饼图绘制、南丁格尔图的绘制、异步数据加载、Loading动画、Ajax实现异步加载和更新，理解数据区域缩放组件及其使用。   1. **项目内容**   完成Echarts环境搭建、完成饼图绘制、完成南丁格尔图的绘制、完成异步数据加载、完成Loading动画、完成Ajax实现异步加载和更新、完成数据区域缩放可视化实验手册。   1. **资源参数**   （1）实验手册7个  （2）配套实验7个 | 套 | 1 |
| Redis内存数据库 | 1. **项目需求**   实验的目的是帮助学员了解Redis内存数据库应用场景、能够进行Redis安装配置、掌握Redis各种数据类型的使用、能够使用Java对Redis进行编程，掌握Redis主从复制、事务控制、Redis持久化。   1. **项目内容**   Redis基础操作、Redis各种数据类型、Redis高级操作。   1. **资源参数**   （1）实验手册3个  （2）配套实验3个 | 套 | 1 |
| Flume数据采集 | 1. **项目需求**   实验的目的是帮助学员了解flume数据采集的作用及工作原理，掌握flume的安装配置，能够深入了解flume中source、channel、sink各个组件支持的常用类型，能够使用flume进行数据的采集，使用Flume拦截器进行数据的过滤。   1. **项目内容**   完成Flume安装部署、Flume NetCat（监听端口）操作、通道 Channel：内存通道、文件通道、HDFS接收器、Exec源、Spooling Directory源、avro源、通道选择器、Flume拦截器（Host Interceptor）、Flume拦截器（Static Interceptor）。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| Mahout | 1. **项目需求**   实验的目的是帮助学员了解了解数据挖掘与机器学习的基础知识，理解推荐算法的原理与应用场景，掌握推荐算流程与程序设计方法，掌握基于Mahout数据挖掘框架开发数据挖掘应用的方法。   1. **项目内容**   完成Mahout安装部署、完成基于物品的推荐(用户访问数据)、完成基于物品的推荐(用户评分数据)、完成基于用户的推荐（用户访问数据）、完成基于用户的推荐（用户评分数据）。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| Sqoop数据导入导出 | 1. **项目需求**   实验的目的是帮助学员掌握Sqoop工具的使用，了解Sqoop工作原理，掌握Sqoop安装配置，能够使用Sqoop进行mysql与hdfs、hbase、hive之间的数据导入导出。   1. **项目内容**   完成Sqoop安装部署、使用Sqoop将数据从Mysql导入到HDFS、使用Sqoop将Mysql数据并发导入到HDFS、使用Sqoop从HDFS将数据导出到Mysql、使用Sqoop从将mysql数据导入到hive、使用Sqoop将Hive数据导出到Mysql、使用Sqoop将Mysql数据导入到hbase。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| Zookeeper | 1. **项目需求**   实验的目的是帮助学员了解zookeeper基本概念、工作原理、安装配置。   1. **项目内容**   完成 zookeeper 安装部署。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| Python机器学习 | 1. **项目需求**   实验的目的是是能够使学员了解机器学习基本知识、掌握KNN.、线性回归、决策树等算法。   1. **项目内容**   机器学习概述、KNN算法、线性回归、梯度下降法、多项式回归、逻辑回归、决策树、聚类分析、贝叶斯分类器、随机森林、AdaBoost、SVM支持向量机、深度前馈神经网络。   1. **资源参数**   （1）实验手册13个  （2）配套实验12个 | 套 | 1 |
| Python编程 | 1. **项目需求**   实验的目的是能够使学员了解Python基础知识、能够搭建Python开发环境、掌握Python数据类型和变量、控制和循环语句、字符串与数组、函数、匿名函数、与异常、列表、元组、字典、集合、类和对象，为后续的Python高级课程的学习打下坚实的基础。   1. **项目内容**   Python安装、Python基础语法、Python数据类型、Python函数。   1. **资源参数**   （1）实验手册4个  （2）配套实验4个 | 套 | 1 |
| Python网络爬虫 | 1. **项目需求**   实验的目的是帮助学员了解Scrapy架构和原理、各个模块的配合、Xpath、解析网站数据等。   1. **项目内容**   爬取股票信息、爬取房源信息、爬取招生章程信息。   1. **资源参数**   （1）实验手册3个  （2）配套实验3个 | 套 | 1 |
| 波士顿房价预测 | 1. **项目需求**   案例利用马萨诸塞州波士顿郊区的房屋信息数据训练和测试一个模型，并对模型的性能和预测能力进行测试。通过该数据训练好的模型可以被用来对房屋做特定预测。   1. **项目内容**   技术点：多元线性回归模型、梯度下降法、特征缩放、均方差损失函数、python、tensorflow2.1。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| MNIST手写体识别 | 1. **项目需求**   MNIST手写数字数据集来源于是美国国家标准与技术研究所，是著名的公开数据集之一，通常这个数据集都会被作为深度学习的入门案例。数据集中的数字图片是由250个不同职业的人纯手写绘制；通过构造单层神经网络模型来完成手写体的识别功能。   1. **项目内容**   技术点：单层的神经网络系统、识别模型、逻辑回归、Softmax回归、交叉熵损失函数、python、tensorflow2.1。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| CNN图像识别 | 1. **项目需求**   图像识别是人工智能的一个重要领域。伴随着近几年深度学习算法的发展，尤其是对深度卷积神经网络的不断研究，使得图像识别的错误率连续下降，甚至在一些特定的数据集上超越了人类。因此本案例基于CIFAR-10数据集，使用卷积神经网络来学习解决图像识别问题的基本方法。   1. **项目内容**   技术点：卷积神经网络、反向传播算法、池化、卷积神经网络的加速计算方法、python、tensorflow2.1。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| 交通时空大数据项目 | 1. **项目需求**   研发交通时空大数据分析系统，进行城市人群出行分析，解决该市交通拥堵问题。   1. **项目内容**   技术点：MySQL数据库、Sqoop数据导入导出、HDFS分布式文件系统、Hive数据仓库、kafka消息队列、Spark计算框架（重点）、Python入门编程。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| 智慧交通大数据  （智慧交通套牌车） | 1. **项目需求**   现在已有很多面向智能交通管理的车辆异常行为检测方法，但是在公共安全领域的异常行为检测研究不足．需要一种基于车牌识别系统车辆轨迹的行为异常检测机制，通过车牌识别系统获取抓拍记录，分析各个车辆在系统卡口的历史通行记录，提取车辆轨迹的时间空间特征，在卡口设备识别率不理想的情况下有效地提高异常检测率。   1. **项目内容**   技术点：MySQL数据库、Sqoop数据导入导出、HDFS分布式文件系统、Hive数据仓库、Spark计算框架（重点）。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| 智慧交通大数据（智慧交通伴随车） | 1. **项目需求**   通过一定方法和技术从海量车牌识别数据中挖掘出存在伴随关系的车辆。   1. **项目内容**   技术点：MySQL数据库、Sqoop数据导入导出、HDFS分布式文件系统、Hive数据仓库、Spark计算框架（重点）。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 |  |  |
| 互联网电影智能推荐 | 1. **项目需求**   推荐系统的任务就是联系用户和内容，通过分析用户对电影的评分行为，对用户兴趣建模，用数据挖掘的方法实现对用户行为分析，从而预测用户的兴趣并给用户推荐。   1. **项目内容**   技术点：Sqoop数据导入导出、HDFS分布式文件系统、MapReduce并行计算。   1. **资源参数**   （1）实验手册1个  （2）配套实验1个 | 套 | 1 |
| 硬件平台 | 管理节点服务器 | 1.规格：2U机架式服务器，标配原厂导轨  2.处理器：CPU配置≥2颗英特尔至强银牌及更高（CPU核心≥12核心，主频≥2.4GHz，支持超线程、虚拟化技术）  3.内存：内存可扩展≥24个内存插槽；实配内存≥128G DDR4 2666MHZ  4.阵列卡：配置1个Raid阵列卡,带2GB缓存，支持Raid0/1/5  5.硬盘：配置6块600GB 15K RPM SAS硬盘  6.网络：配置至少4个千兆接口  7.支持8个以上PCI-e 3.0可用插槽，支持2个双宽GPU。整机具备4个USB接口  8.电源：配置2个热插拔冗余电源 | 台 | 1 |
| 计算节点服务器 | 1.规格：2U机架式服务器，标配原厂导轨  2.处理器：CPU配置≥2颗英特尔至强金牌及更高（CPU核心≥24核心，主频≥2.2GHz，支持超线程、虚拟化技术）  3.内存：实配内存≥512G DDR4 2666MHZ  4.阵列卡：配置1个Raid阵列卡,带2GB缓存，支持Raid0/1/5  5.硬盘：配置8块2.4TB 10K RPM SAS硬盘  6.网络：配置4个千兆接口  7.支持8个以上PCI-e 3.0可用插槽，支持2个双宽GPU。整机具备4个USB接口  8.电源：配置2个热插拔冗余电源 | 台 | 4 |
| 交换机 | 1：48个10/100/1000 Base-T以太网端口， 4个万兆SFP+万兆以太网端口。支持交换机横向虚拟化。最大交换容量≥6Tbps，整机包转发率≥300Mpps，配置双电源。 | 台 | 2 |