附件：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 名称 | 设备内容 | | 性能要求 | 功能要求 | 单位 | 数量 |
| 1 | 大数据开发部署实训平台 | 大数据教学一体机 | 典型配置为每台4节点，每个节点支持2颗intel E5-2600v3，v4系列的CPU/16根内存插槽，最大支持512G内存/支持6个2.5寸硬盘/双千兆网口/一个IPMI口;2000W冗余电源/导轨。CPU：8 \*Intel® Xeon® Processor E5-2620 v4 （20M Cache, 2.10 GHz） 8核；内存：32 \* 16G DDR4RECC；硬盘：24 \* 1T 2.5'sata企业级硬盘； | | 能够完成基于Hadoop的大数据部署综合教学，搭建Hadoop运行环境，支持学生Hadoop分布式计算环境安装、配置实验及Mapreduce开发等基础实验；提供课程相应的讲义、教材、ppt、实验项目及实验指导书，实验项目包含： Hadoop大数据分析实验：HDFS基本操作实验、MapReduce基础实验、MapReduce高级特性实验、迭代式MapReduce程序开发实验、Hive基本操作实验等。 | 台 | 1 |
| 2 | 大数据分析数据集及案例库 | 基本数据集 | 1.提供图形化拖拽工具进行数据模型构建、训练和预测检验在工具中 2.支持算子与模型管理，降低数据挖掘开发门槛。 3.支持不少于30人并发使用 4.支持自定义分布式数据挖掘算法，用户可以根据生产需求自定义开发分布式数据挖掘算法，包括： （1）提供R语言开发接口支持。 （2）提供Python语言开发接口支持。 （3）提供API开发接口支持。 | | 1.整合主流深度学习框架，包括Tensorflow、MXNet，通过将深度学习框架与Hadoop体系融合，帮助数据科学家以及分析人员敏捷构建DNN，CNN等模型。 2.使用全量数据样本进行模型训练，提高算法的精准度。 3.提供可用于实践教学的开源大数据训练集、测试集等基本数据集。 4.提供国内外政府统计口脱敏数据。 | 套 | 1 |
| 3 | 行业数据集 | 提供包括金融、保险、通信、电商、交通、医疗在内不低于6类行业专项数据和应用案例（仅限教学使用）。 | | 1.提供实际行业应用项目中的真实脱敏数据，如商圈聚类中的刷卡数据、担保关系数据、电商交易记录数据等， 2.提供脚本产生模拟数据，供算法、测试集及应用案例选用。 | 套 | 1 |
| 4 | 课程开发 | 教学资源 | 提供课程相应的讲义、教材、ppt、实验项目及实验指导书 | | 1、大数据技术基础实验模块： Hadoop大数据分析实验：HDFS基本操作实验、MapReduce基础实验、MapReduce高级特性实验、迭代式MapReduce程序开发实验、Hive基本操作实验等。 2、数据分析与挖掘模块： 实现基于大数据环境下的数据分析和挖掘的算法，使用R和Python两种语言实现，提供源代码，分析结果以可视化方式输出；可在相应编译环境内修改调试源代码，并提供优化后的算法与原算法的对比分析结果以及可视化图表展示。 （1）数据探索与预处理实验：数据质量分析实验、数据特征分析实验、统计特征函数实验、统计作图函数实验、数据清洗实验、数据集成实验、数据变换实验、数据规约实验等。 （2）R语言数据挖掘实验：R语言分类预测基本流程实验、R语言数据预处理实验、R语言决策树分类方法实验、R语言高级分类方法实验、R语言聚类分析实验、R语言关联分析实验等。 （3）Python数据分析实验：Python数据降维实验、Python聚类分析实验、Python K近邻分类实验、Python回归分析实验、Python决策树分类实验等。 | 套 | 1 |