**《Hadoop大数据技术》实验课程教学大纲**

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**

**一、课程简介**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程中文名** | **Hadoop大数据技术** | | | | | | | | |
| **课程英文名** | **Hadoop Big Data Technology** | | | | | | **双语授课** | | □是 ☑否 |
| **课程代码** | 10114084 | **课程学分** | | 2.5 | **总学时数** | | | 60 | |
| **课程类别** | **□**专业基础课程  **☑**专业核心课程  **□**专业选修课程  **□**其他 | **课程性质** | | ☑必修  □选修  □其他 | **课程形态** | | | □线上  ☑线下  □线上线下混合式  □社会实践  □虚拟仿真实验教学 | |
| **考核方式** | □闭卷 □开卷 □课程论文 ☑课程作品 □汇报展示 □报告  ☑课堂表现 ☑阶段性测试 ☑平时作业 □其他（可多选） | | | | | | | | |
| **开课学院** | 大数据与智能工程学院 | | **开课**  **系(教研室)** | | | 数据科学与大数据技术 | | | |
| **面向专业** | 数据科学与大数据技术 | | **开课学期** | | | 第4学期 | | | |
| **课程负责人** | 程东东 | | **审核人** | | | 黄金龙、孙建乐 | | | |
| **先修课程** | Java程序设计 | | | | | | | | |
| **后续课程** | 无 | | | | | | | | |
| **选用教材** | 1. 林子雨. 大数据技术原理与应用:概念、存储、处理、分析与应用(第3版)[M]. 北京：人民邮电出版社, 2021年. | | | | | | | | |
| **参考书目** | 1. Tom White等. Hadoop权威指南:大数据的存储与分析（第4版）[M]. 北京：清华大学出版社, 2017年  2. 杨治明, 许桂秋. Hadoop大数据技术与应用[M]. 北京：人民邮电出版社, 2019年. | | | | | | | | |
| **课程资源** | 大数据技术原理与应用. 中国大学MOOC（厦门大学）. | | | | | | | | |
| **课程简介** | Hadoop大数据技术课程是一门综合实践性课程，旨在让学生掌握Hadoop这一主流的大数据技术和工具，具备处理大规模数据的能力。课程内容包括Hadoop分布式文件系统、MapReduce、HBase、Hive、Pig等，同时还涉及大数据分析和处理的实践技能。通过该课程学习，让学生掌握Hadoop的安装、配置和使用，以及如何处理不同类型的数据、进行大数据分析和处理。 | | | | | | | | |

**二、课程目标**

**表1 课程目标**

|  |  |
| --- | --- |
| **序号** | **具体课程目标** |
| **课程目标1** | 掌握Hadoop大数据处理基本原理，能够理解和应用Hadoop生态系统中的多个组件，包括Hadoop分布式文件系统、MapReduce、Hive、Pig等，掌握Hadoop的安装、配置和使用，根据不同项目要求设计目标，采集和处理不同数据。 |
| **课程目标2** | 能够设计和实现基于Hadoop的大数据解决方案，以满足实际应用的需求。能够利用Hadoop技术原理进行案例分析和讨论，将所学知识应用于解决实际问题的实践中，培养创新思维和团队协作能力。 |
| **课程目标3** | 能够理解和遵守职业道德规范，以负责任的态度应用技术，尊重并保护个人和团队的知识产权，并能够适应技术和行业的快速发展。 |

**表2 课程目标与毕业要求对应关系**

| **毕业要求** | **指标点** | **课程目标** |
| --- | --- | --- |
| **毕业要求3：**设计/开发解决方案：能够针对计算机复杂工程问题的解决方案，设计并实现满足特定需求的软件系统，并能够在设计和实现环节中体现创新意识，考虑社会、健康、安全、法律、文化以及环境等因素。【M】 | 3.1 能够归纳描述用户的需求，确定设计目标，并能选择正确的计算机软件系统设计开发方法。× | 2 |
| **毕业要求4：**研究：能够基于科学原理并采用科学方法对计算机复杂工程问题进行研究，设计实验、分析与解释数据，并能通过信息综合得到合理有效结论。【H】 | 4.2 能够根据实验设计的目标，有效地提取和搜索实验数据，保证数据的真实性和完整性。 | 1 |
| **毕业要求6：**工程与社会：能够基于工程背景知识进行合理分析，评价计算机工程实践和复杂工程问题解决方案对社会、健康、安全、法律以及文化的影响，并理解应承担的责任。【L】 | 6.2 能够综合评价计算机软件、硬件、网络和相应的复杂工程问题对社会、健康、安全、法律、文化等方面的影响，明确自己的社会责任。 | 3 |

**三、课程教学内容与方法**

**表3课程目标、教学内容和方法对应关系**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **项目名称** | **项目来源** | **教学目标（观测点、重难点）** | **学时数** | **项目类型** | **要求** | **每组人数** | **教学方法** | **课程目标** |
| 1 | 实验1：初识Hadoop大数据技术 | 实验教材 | 1.理解熟悉大数据理论 | 6 | 验证性 | 必做 | 1 | 讲授法 | 课程目标1 |
| 2.掌握大数据相关问题的解决方法（重点） |
| 3.掌握Hadoop技术原理（难点） |
| 4.完成Hadoop应用案例 | 课程目标3 |
| 2 | 实验2：Hadoop环境的安装与伪分布式配置 | 实验教材 | 1.掌握Hadoop三种模式：单机、伪分布式和完全分布式 | 6 | 验证性 | 必做 | 1 | 讲授法  实验指导 | 课程目标3 |
| 2.掌握Hadoop伪分布式安装原理，并完成相关步骤（重点） |
| 3 | 实验3：分布式文件系统HDFS | 实验教材 | 1.理解HDFS组成与架构 | 14 | 设计研究 | 必做 | 1 | 讲授法  案例教学  实验指导 | 课程目标1 |
| 2.掌握HDFS工作机制（重点） |
| 3.掌握并熟练应用HDFS操作命令（重点） | 课程目标2 |
| 4.理解熟悉HDFS高级功能（难点） |
| 4 | 实验4: YARN组件 | 实验教材 | 1.理解YARN组件架构 | 4 | 验证性 | 必做 | 1 | 讲授法 | 课程目标2 |
| 2.熟悉YARN调度组件（难点） |
| 5 | 实验5：MapRecue原理及编程 | 实验教材 | 1.理解MapReduce编程模型理论 | 14 | 设计研究 | 必做 | 1 | 讲授法  案例教学  实验指导 | 课程目标1 |
| 2.掌握MapReduce工作机制（重点） |
| 3.完成MapReduce 编程案例（难点） | 课程目标3 |
| 4.熟悉MapReduce编程进阶 |
| 6 | 实验6：HBase基本原理与编程 | 实验教材 | 1.掌握HBase基本知识 | 14 | 设计研究 | 必做 | 1 | 讲授法  案例教学  实验指导 | 课程目标1 |
| 2.掌握HBase的环境搭建步骤 |
| 3.掌握HBase Shell基本操作 |
| 4.掌握HBase客户端API应用开发 | 课程目标2 |
| 5.掌握基于HBase的MapReduce项目开发 |

## 四、课程考核

**（一）考核内容与考核方式**

**表4-1 课程目标、考核内容与考核方式对应关系**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **课程目标** | **考核内容** | **所属**  **学习项目** | **考核占比** | **考核方式** |
| 课程  目标 1 | 大数据相关概念理论和相关问题的解决方法 | 项目1 | 52% | 课堂表现、实验作业、阶段测试、课程设计 |
| Hadoop技术原理 |
| HDFS组成和架构及其工作机制 | 项目3 |
| MapReduce编程模型理论及其工作机制 | 项目5 |
| HBase基本原理、环境搭建及其shell基本操作 | 项目6 |
| 课程  目标 2 | HDFS的常用操作命令及部分高级功能 | 项目3 | 40% | 课堂表现、实验作业、阶段测试、课程设计 |
| YARN组件构架及其调度原理 | 项目4 |
| HBase客户端API应用开发 | 项目6 |
| 基于HBase的MapReduce项目开发 |
| 课程  目标 3 | Hadoop应用领域及其相关案例的实现 | 项目1 | 8% | 课堂表现、课程设计 |
| Hadoop伪分布式原理及其实现 | 项目2 |
| MapReduce API编程案例实现及其进阶 | 项目5 |

**表4-2 课程目标与考核方式矩阵关系**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 课程  目标 | 考核方式 | | | | 考核占比 |
| 课程设计成绩比例60% | 课堂表现成绩比例10% | 实验作业成绩比例20% | 阶段性测验成绩比例10% |
| 课程目标1 | 50% | 40% | 50% | 80% | 52% |
| 课程目标2 | 40% | 40% | 50% | 20% | 40% |
| 课程目标3 | 10% | 20% | 0% | 0% | 8% |

## （二）成绩评定

**1.平时成绩评定**

**（1）课堂表现（25%）：**通过学生在课堂上发言、回答提问情况，评价学生的课程参与能力。

**（2）实验作业完成情况（50%）：**围绕课程的学习目标进行作业的设计。如让学生简述大数据的认识，考核学生对于大数据相关概念的理解情况，帮助学生将定义转化为自己的理解。

**（3）阶段性测验（25%）**：学生在平时测试、测验中掌握课程的情况。

**2.期末成绩评定**

期末考试主要考察学生对Hadoop框架HDFS、MapReduce、HBase、Hive、Pig、Flume、Sqoop等组件的原理、特点及使用方法掌握程度。方式为实验设计+实验报告。

**3.总成绩评定**

总成绩应由平时考核成绩和期末考核成绩构成

总成绩（100%）=平时成绩（40%）+期末成绩（60%）

## （三）评分标准

**表5 评分标准（非试卷考核项目）**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **考核项目** | **评分标准** | | | | |
| **优秀**  **(100>x≥90)** | **良好**  **(90> x≥80)** | **中等**  **(80> x≥70)** | **及格**  **(70> x≥60)** | **不及格**  **(x <60)** |
| 课堂表现 | 课上严格遵守课堂纪律，按要求完成课堂练习且正确率高，积极参与课堂讨论，讨论观点有新意，有见解，实验课不迟到早退，与老师积极互动。 | 课上严格遵守课堂纪律，按要求完成课堂练习且正确率较高，积极参与课堂讨论，讨论观点有一定新意和见解，实验课不迟到早退，与老师积极互动。 | 课上严格遵守课堂纪律，按要求完成课堂练习且有一定正确率，参与课堂讨论，讨论观点有一定新意和见解，实验课不迟到早退，与老师互动。 | 课上严格遵守课堂纪律，按要求完成课堂练习且基本正确率，参与课堂讨论，讨论观点基本的新意和见解，实验课不迟到早退，与老师偶尔互动。 | 课上不遵守课堂纪律，不参与课堂讨论，未按要求完成课堂练习或正确率低。实验课缺课。 |
| 实验作业 | 实验作业按照实验目标完成，步骤完整、数据详实、有深刻的实验反思，并按时提交。 | 实验作业基本按照实验目标完成，步骤完整、数据详实，有实验反思，并按时提交。 | 实验作业基本按照实验目标完成，步骤完整，缺乏数据支撑，有实验反思，并按时提交。 | 实验作业基本按照实验目标完成，步骤描述不完整，缺乏数据支撑，缺乏实验反思，并按时提交。 | 实验作业未按照实验目标完成，缺乏实验步骤、数据支撑和实验反思，未按时提交。 |
| 阶段性测验 | 掌握物Hadoop分布式文件系、MapReduce、Hive、Pig等相关知识，能熟练使用相关知识并解决问题。 | 掌握物Hadoop分布式文件系、MapReduce、Hive、Pig等相关知识，能使用相关知识并解决问题。 | 掌握物Hadoop分布式文件系、MapReduce、Hive、Pig等相关知识，能在一定程度上使用相关知识并解决问题。 | 基本掌握物Hadoop分布式文件系、MapReduce、Hive、Pig等相关知识，能部分使用相关知识并解决问题。 | 未掌握物Hadoop分布式文件系、MapReduce、Hive、Pig等相关知识，无法使用相关知识并解决问题。 |
| 课程设计 | （1）课程设计符合课程性质，选题范围适中，具有较高的研究价值和意义，表现出很强的问题意识。（2）课程设计思路清晰且具有创新性。（3）理解课程设计原理，且代码能够正常运行（4）课程报告结构合理，组织严密，连贯一致。语言表达准确，叙述清楚，符合专业术语规范。（5）课程报告符合学术规范。 | （1）课程设计选题恰当合理，具有较高的研究价值和意义，表现出较强的问题意识。（2）课程设计思路较为清晰且具有一定创新性。（3）基本理解课程设计原理，且代码能够正常运行（4）课程报告结构较为合理，语言表达较为准确，叙述清楚，所使用的专业术语较为规范。（5）课程报告基本符合学术规范，无明显错误。 | （1）课程设计选题较为合理，具有一定的研究价值和意义，表现出一定的问题意识。（2）课程设计思路基本清晰且具有一定创新性（3）在一定程度上理解课程设计原理，且代码能够正常运行（4）课程报告结构较为合理，语言表达较为准确，叙述清楚，所使用的专业术语较为规范。  （5）课程报告基本符合学术规范，无明显错误。 | （1）课程设计选题具有一定的研究价值和意义，但凝练不够，问题意识欠佳。（2）课程设计思路基本清晰，但创新性不足（3）在一定程度上理解课程设计原理，且代码能够正常运行（4）课程报告结构基本合理，语言表达基本清楚，所使用的教育专业术语基本规范。（5）课程报告基本符合学术规范，有部分错误。 | （1）课程设计选题不符合课程性质，或主题不明确（2）课程设计思路不清晰（3）不理解课程设计原理，代码不能正常运行（4）课程报告结构不合理，语言不通顺，所使用的教育专业术语不规范。（5）课程报告明显不符合学术规范，或存在抄袭现象。 |

## 五、其它说明

本课程大纲依据2023版数据科学与大数据技术专业人才培养方案，由大数据与智能工程学院（部）数据科学与大数据技术教学系（教研室）讨论制定，大数据与智能工程学院（部）教学工作委员会审定，教务处审核批准，自2023级开始执行。