**《Hadoop大数据技术综合实训》实训课程教学大纲**

**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大数据技术综合实训 |
| **课程英文名** | Comprehensive Training of Hadoop Big Data Technology | **双语授课** | □是 ☑否 |
| **课程代码** | 10114090 | **课程学分** | 2 | **周（学时）** | 2周 |
| **课程类别** | □专业认知实习□专业见习☑工程实训□毕业实习□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** | 黄金龙 | **审核人** | 张素兰 |
| **先修课程** | Java程序设计 |
| **后续课程** | 无 |
| **选用教材** | 1. 陆嘉恒. Hadoop实战. 机械工业出版社，2012.
 |
| **参考书目** | 1. [安俊秀](http://www.dangdang.com/author/%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD_1)、[王鹏](http://www.dangdang.com/author/%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD_1)、[靳宇倡](http://www.dangdang.com/author/%EF%BF%BD%EF%BF%BD%EF%BF%BD%EE%B3%AB_1). Hadoop大数据处理技术基础与实践. [人民邮电出版社](http://www.dangdang.com/publish/%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD%CA%B5%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD%EF%BF%BD_1)，2015 .
2. lex Holmes 梁李印、宁青、杨卓荦 译. Hadoop硬实战. 电子工业出版社，2015.
 |
| **课程资源** | 1. 厦门大学公开课：Hadoop集群的部署和使用

http://open.163.com/newview/movie/free?pid=MEKD4Q6GG&mid=MEKDKSSEG |
| **课程简介** | Hadoop大数据技术综合实训是数据科学与大数据技术专业必修的综合实践课程。该课程内容包括Hadoop平台的安装及其工作流程，Pig组件、Map Reduce原理、HBase原理、Hive组件、Sqoop组件、Zookeeper组件等。使学生能够选择不同的Hadoop组件和技术方法，解决国内主流的大数据应用场景中存在的问题。 |

**二、课程目标**

**表1 课程目标**

|  |  |
| --- | --- |
| **序号** | **具体课程目标** |
| **课程目标1** | 能够应用计算机科学和软件工程、操作系统等领域的知识，在Hadoop部署的构思与设计阶段，通过文献研究、实验试验等方法，识别、表达Hadoop工程部署的问题及其解决方法。 |
| **课程目标2** | 能够应用Hadoop的理论与方法，以及计算机及操作系统等相关领域的专业知识、技能与工具，能够设计针对Hadoop部署问题的解决方案，解决Hadoop组件部署过程中的问题，并在设计环节中体现创新意识。 |
| **课程目标3** | 能够针对大数据挖掘，大数据系统开发过程中涉及的复杂工程问题，选择和使用恰当的大数据处理技术和工具，进行信息处理、表达、建模、设计、模拟、验证，并能够在实践中了解国内主流的Hadoop平台在各个应用中的优势和局限性；树立使用国产软件、保护版权的意识，激发民族自豪感；坚持自主创新，为建设世界科技强国而奋斗。 |

**表2-1 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9：**具有良好的综合素质，能够在多学科背景下的团队中理解与承担个体、团队成员以及负责人角色，并发挥相应的作用。【M】 | 9.2能够适应多学科背景的团队合作方式，具备一定的组织管理能力，并能综合团队成员的意见，进行合理决策。 | 课程目标1 |
| **毕业要求10：**具备一定的国际视野和跨文化沟通能力，能够就大数据相关应用领域的复杂工程问题与业界同行及公众进行有效沟通，包括撰写报告、陈述发言、清晰表达等。【L】 | 10.1能够就大数据相关应用领域中的复杂工程问题与业界同行及社会公众进行交流，通过书面报告、设计文档、编写代码和口头陈述清晰地表达团队或个人观点与设计理念。 | 课程目标2 |
| **毕业要求11：**能够理解并掌握工程管理与经济决策方法，熟悉大数据应用领域工程项目管理的基本方法和技术，并能在多学科环境中应用。【M】 | 11.1理解从事大数据应用领域工程实践活动所需的管理与经济因素，掌握计算机工程管理和经济决策方法。 | 课程目标3 |

**三、教学内容及要求**

**（一）学习内容**

A. 通过本次课程设计，进一步掌握和应用Hadoop平台及其相关组件的知识，需掌握的内容包括：

1) Hadoop平台的工作流程，开发环境搭建，常用开发工具，开发环境的关键配置

2) Hadoop平台安装模式：单机模式、伪分布式模式、完全分布式模式

3) Pig组件： Pig环境参数，Pig安装配置流程，及操作方式

4) MapReduce原理：MapReduce的基本原理及其程序的基本编写，Map函数和Reduce函数的输入输出。

5) HBase组件：HBase组件在Hadoop框架中的地位和作用，HBase的配置方法，及HBase的常规曾删改查等命令操作。

6) Hive组件：Hive组件在Hadoop框架中的地位和作用，Hive的配置方法，及Hive的常规命令操作。

7) Sqoop组件：Hadoop的安装配置、数据仓库的基本原理，Sqoop进行数据采集的功能特点。

8) Zookeeper组件：理解Zookeeper在Hadoop框架中的地位与作用，Zookeeper的一致性服务原理。

B.从下列国内主流的大数据应用场景中选择并根据情况设定项目名称，按表中所列步骤完成该课程：

1) 电子商务大数据分析系统的设计与实现

2) 金融大数据分析系统的设计与实现

3) 交通大数据分析系统的设计与实现

4) 人脸大数据分析系统的设计与实现

5) 视频大数据分析系统的设计与实现

6) 音频大数据分析系统的设计与实现

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 项目名称 | 学习目标 | 学习内容 | 主要仪器设备 |
| 1 | Hadoop平台环境配置及安装 | 掌握Hadoop在Linux系统下不同模式（单机、伪分布式、完全分布式）的安装配置，掌握SSH配置方法。  | Hadoop大数据框架 | 计算机，及以下软件：Linux虚拟机、Hadoop及其相关组件等 |
| 2 | Hadoop上Pig部署 | 了解Pig环境参数，掌握Pig安装配置流程，熟悉MapReduce的基本原理及其程序的基本编写，熟悉Map函数和Reduce函数的输入输出 | 1. Pig的安装与配置
2. Pig的常用命令操作
3. Pig的编程实现
 | 计算机，及以下软件：Linux虚拟机、分布式的Hadoop，HBase，Eclipse等 |
| 3 | Hadoop上HBase的部署 | 掌握Hadoop的安装配置、理解分布式数据库的基本原理及安装配置流程、掌握HBase组件在Hadoop框架中的地位和作用，掌握HBase的配置方法，及HBase的常规命令操作，例如曾删改查等命令。 | 1. HBase的安装与配置2. HBase的Shell命令3. HBase的编程实现 | 计算机，及以下软件：Linux虚拟机、伪分布式的Hadoop，HBase，Eclipse等 |
| 4 | Hadoop上Hive部署 | 了解Hive组件在Hadoop框架中的地位和作用，掌握Hive的配置方法，及Hive的常规命令操作。 | 1. Hive的安装与配置
2. Hive的常用命令操作
3. Hive的编程实现
 | 计算机，及以下软件：Linux虚拟机、伪分布式的Hadoop，HBase，Eclipse等 |
| 5 | Hadoop上Sqoop部署 | 理解数据仓库的基本原理、掌握MySQL的基本操作，掌握Sqoop进行数据采集的功能特点。 | 1. Sqoop的安装与配置
2. Sqoop的常用命令操作
3. Sqoop的编程实现
 | 计算机，及以下软件：Linux虚拟机、伪分布式的Hadoop，HBase，Eclipse等 |
| 6 | Hadoop上Zookeeper部署 | 理解Zookeeper在Hadoop框架中的地位与作用，掌握Zookeeper的一致性服务原理。 | 1. Zookeeper任务管理方式
2. Zookeeper一致性原理
 | 计算机，及以下软件：Linux虚拟机、伪分布式的Hadoop，HBase，Eclipse等 |

**（二）时间安排**

|  |  |  |
| --- | --- | --- |
|  | 上午 | 下午 |
| 周一 | 教师讲解 | Hadoop平台环境配置及安装 |
| 周二 | 教师讲解 | Hadoop上Pig部署 |
| 周三 | Hadoop上Pig常用命令操作 | Hadoop上Pig编程实现 |
| 周四 | 教师讲解 | Hadoop上HBase的部署 |
| 周五 | Hadoop上HBase常用命令操作 | Hadoop上HBase编程实现 |
| 周一 | 教师讲解  | Hadoop上Hive部署 |
| 周二 | Hadoop上Hive常用命令操作 | Hadoop上Hive编程实现 |
| 周三 | 教师讲解 | Hadoop上Sqoop部署 |
| 周四 | Hadoop上Sqoop常用命令操作 | Hadoop上Sqoop编程实现 |
| 周五 | 教师讲解 | Hadoop上Zookeeper部署 |

**（三）工作流程**

1)技术讲解：教师应用讲授法、演示法、实验法、任务驱动法、讨论法、自主学习法，讲解综合实训应用的知识点,及各个软件的使用要点。

2)问题分析：通过对项目需要解决的问题进行分析，确定各个步骤的技术方案，书写对问题分析的文档。

3）技术方案设计：对项目所需要的总体技术方案进行设计。

4) Hadoop平台及其组件安装配置：确定Hadoop框架组织结构，并在实验过程中将pig、HBase、Hive、Sqoop和Zookeeper等组件配置安装完成。

5)Hadoop平台及其组件应用：使用Hadoop平台及其组件，通过Java或Python语言编程实现特定功能或系统。

6)撰写总结材料：总结实验过程，并形成实验报告，分析收获和不足。

7)评分：根据评分标准对每个同学的项目的设计、完成度、项目演示、文档撰写等结果进行综合评分。

**（四）业务指导**

在综合课程设计期间，由校内老师1名和企业老师1名，每天深入机房，随时指导解答问题。

## 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** | **所属环节** | **考核占比** | **考核方式** |
| 课程目标 1 | MapReduce函数原理及应用 | Hadoop上Pig部署 | 46% | 课堂表现实验报告团队协作期末考核 |
| Pig的编程实现 |
| HBase的编程实现 | Hadoop上HBase的部署 |
| Hive的编程实现 | Hadoop上Hive部署 |
| Sqoop的编程实现 | Hadoop上Sqoop部署 |
| Zookeeper一致性原理 | Hadoop上Zookeeper部署 |
| 课程目标 2 | Pig的常用命令操作 | Hadoop上Pig部署 | 35% | 课堂表现实验报告团队协作期末考核 |
| HBase的Shell命令 | Hadoop上HBase的部署 |
| Hive的常用命令操作 | Hadoop上Hive部署 |
| Sqoop的常用命令操作 | Hadoop上Sqoop部署 |
| 课程目标 3 | 国内主流的Hadoop平台及其组件的安装与配置 | 全部学习模块 | 19% | 课堂表现实验报告团队协作期末考核 |
| 国内主流的Hadoop平台在各个应用场景中的有适合局限性 |

**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 |
| 平时考核（40%） | 期末考核（60%） |
| 课堂表现（10%） | 实验报告（20%） | 团队协作（10%） |
| 课程目标1 | 60% | 60% | 40% | 40% | 46%=10%\*60%+20%\*60%+10%\*40%+60%\*40% |
| 课程目标2 | 30% | 20% | 40% | 40% | 35%=10%\*30%+20%\*20%+10%\*40%+60%\*40% |
| 课程目标3 | 10% | 20% | 20% | 20% | 19%=10%\*10%+20%\*20%+10%\*20%+60%\*20% |

**（二）成绩评定**

**1.平时成绩评定**

1）课堂表现（25 分）：通过学生在课堂上的表现情况、提问与回答、与教师互动情况，评价学生相关的能力。（包括理想信念、家国情怀、社会责任、学习态度、学习兴趣、团队合作与终身学习意识等）

2）实验报告（50分）：学生在过程中协调、沟通等的情况。

3）团队写作（25分）：由教师采取课堂随机点名、电子点名或课堂目检等方法，区分不同情况，采用不同记载符号进行记载。

**2.期末成绩评定**

综合课程设计过程中，根据学生的设计报告、实验的参与和完成情况，以及其他能够评价学生设计学习情况的因素评定成绩。期末成绩的评定为作品设计成绩，即系统代码实现。

**3.总成绩评定**

总成绩（100%）=平时成绩（40%）+期末成绩（60%）

**（三）评分标准**

评分依据：

1)优秀（90-100分）

按设计任务书要求圆满完成规定任务；综合运用知识能力和实践动手能力强，软件设计方案合理，实验效果好；设计态度认真，独立工作能力强，并具有良好的团队协作精神。设计报告条理清晰、论述充分、图表规范、符合设计报告文本格式要求。 答辩过程中，思路清晰、论点正确、对设计方案理解深入，问题回答正确。

2)良好（80-89分）

按设计任务书要求完成规定设计任务；综合运用知识能力和实践动手能力较强，软件设计方案较合理，实验效果较好；设计成果质量较高；设计态度认真，有一定的独立工作能力，并具有较好的团队协作精神。设计报告条理清晰、论述正确、图表较为规范、符合设计报告文本格式要求。 答辩过程中，思路清晰、论点基本正确、对设计方案理解较深入，主要问题回答基本正确。

3)中等（70-79分）

按设计任务书要求完成规定设计任务；能够一定程度的综合运用所学知识，软件设计基本合理，有一定的实践动手能力，设计成果质量一般；设计态度较为认真，设计报告条理基本清晰、论述基本正确、文字通顺、图表基本规范、符合设计报告文本格式要求，但独立工作能力较差；答辩过程中，思路比较清晰、论点有个别错误，分析不够深入。

4)及格（60-69分）

在指导教师及同学的帮助下，能按期完成规定设计任务；综合运用所学知识能力及实践动手能力较差，设计方案基本合理，设计成果质量一般；独立工作能力差；或设计报告条理不够清晰、论述不够充分但没有原则性错误、文字基本通顺、图表不够规范、符合设计报告文本格式要求；或答辩过程中，主要问题经启发能回答，但分析较为肤浅。

5)不及格（60分以下）

未能按期完成规定设计任务。不能综合运用所学知识，实践动手能力差，设计方案存在原则性错误，计算、分析错误较多；或设计报告条理不清、论述有原则性错误、图表不规范、质量很差；或答辩过程中，主要问题阐述不清，对设计内容缺乏了解，概念模糊，问题基本回答不出。

## 五、其他说明

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。

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**