**《 大数据开发技术（Hadoop）》课程教学大纲**

一、课程基本信息

课程代码：18100093

课程名称：大数据开发技术

英文名称：Big data development technology

课程类别：专业课

学 时：48

学 分：3

适用对象:数据科学与大数据技术

考核方式：考试

先修课程：高级程序语言，数据库，算法分析

二、课程简介

《大数据开发技术》是数据科学与大数据技术专业的必修课。该专业学生在具有对大型数据库使用能力的基础上，必须掌握一种或多种大数据处理工具的使用。本课程是大体量数据有关批处理，离线处理的基础课程。

Big data development technology is an optional course for the major of data science and big data technology. Students of this major must master the use of one or more big data processing tools on the basis of their ability to use large databases. This course is a basic course on batch processing and offline processing of large volume data.

三、课程性质与教学目的

《大数据开发技术》大数据专业选修课。通过学习课程使得学生掌握大数据分析的主要思想和基本步骤，并通过编程练习和典型应用实例加深了解；同时对Hadoop平台应用与开发的一般理论有所了解，如分布式数据收集、分布式数据存储、分布式数据计算、分布式数据展示。

开设本学科的目的是让学生掌握如何使用大数据分析技术解决特定业务领域的问题。完成本课程学习后能够熟练的应用大数据技术解决企业中的实际生产问题。围绕课程思政元素，强调自主创新、工匠精神、科技报国等精神内涵。

1. 教学内容及要求

**第一章 初始Hadoop**

1. 目的与要求
2. 了解什么是大数据及其特征
3. 熟悉大数据的典型应用
4. 了解Hadoop的发展历史及其版本
5. 掌握Hadoop的生态体系
6. 教学内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 知识点 | 了解 | 掌握 | 重点 | 难点 |
| 什么是大数据 | √ |  |  |  |
| 大数据的特征 | √ |  |  |  |
| 研究大数据的意义 | √ |  |  |  |
| 大数据的应用场景 |  | √ |  |  |
| Hadoop的发展历史 | √ |  |  |  |
| Hadoop的优势 |  | √ | √ |  |
| Hadoop的生态体系 |  | √ | √ | √ |
| Hadoop的版本 | √ |  |  |  |

1. 思考与实践

本章主要内容为介绍性，在对相关内容的讲述基础上，引导学生思考当前有哪些情形可以用到这类工具。如在疫情防控中，大数据分析会怎样起到作用。技术该如何为国家服务，为人民服务。

1. 教学方法与手段

介绍本章教学主要采用课堂讲授、多媒体教学、网络辅助教学、团队合作、分组讨论、课堂讨论等。

1. **HDFS分布式文件系统**
2. 目的与要求
3. 了解HDFS演变
4. 掌握HDFS特点
5. 掌握HDFS的架构和原理
6. 掌握HDFS的Shell和Java Api操作
7. 教学内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 知识点 | 了解 | 掌握 | 重点 | 难点 |
| HDFS的演变 | √ |  |  |  |
| HDFS的基本概念 |  | √ |  |  |
| HDFS的特点 |  | √ | √ |  |
| HDFS架构和原理 |  | √ | √ | √ |
| HDFS的Shell操作 |  | √ |  |  |
| HDFS的Java API操作 |  | √ |  |  |

1. 思考与实践

本章在完成对相关内容的讲述基础上，需要学生动手完成相关实验，实验有一定难度。针对不同的情形，需要提出相应的解决方案。这是“具体问题，具体分析”方法的一个具体落实。

1. 教学方法与手段

介绍本章教学主要采用课堂讲授、多媒体教学、网络辅助教学、团队合作、分组讨论、课堂讨论等。

**第四章 MapReduce分布式计算系统**

1. 目的与要求
2. 理解MapReduce的核心思想
3. 掌握MapReduce的编程模型
4. 掌握MapReduce的工作原理
5. 掌握MapReduce常见编程组件的使用
6. 教学内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 知识点 | 了解 | 掌握 | 重点 | 难点 |
| MapReduce核心思想 |  | √ |  |  |
| MapReduce编程模型 |  | √ | √ |  |
| MapReduce编程实例——词频统计 |  | √ |  |  |
| MapReduce工作过程 |  | √ | √ | √ |
| MapTask工作原理 |  | √ | √ |  |
| ReduceTask工作原理 |  | √ | √ |  |
| Shuffle工作原理 |  | √ | √ |  |
| MapReduce编程组件 |  | √ | √ | √ |
| MapReduce运行模式 |  | √ |  |  |
| MapReduce性能优化策略 | √ |  |  |  |
| MapReduce经典案例——倒排索引 |  | √ |  |  |
| MapReduce经典案例——数据去重 |  | √ |  |  |
| MapReduce经典案例——TopN |  | √ |  |  |

1. 思考与实践

本章在完成对相关内容的讲述基础上，需要学生动手完成相关实验，实验有一定难度。针对不同的情形，需要提出相应的解决方案。这是“具体问题，具体分析”方法的一个具体落实。

1. 教学方法与手段

介绍本章教学主要采用课堂讲授、多媒体教学、网络辅助教学、团队合作、分组讨论、课堂讨论等。

**第五章 Zookeeper分布式协调服务**

1. 目的与要求
2. 了解Zookeeper的概念和特性
3. 理解Zookeeper数据模型
4. 掌握Zookeeper的Watch机制和选举机制
5. 掌握Zookeeper的集群部署
6. 掌握Zookeeper的Shell操作和Java API操作
7. 熟悉Zookeeper的应用场景
8. 教学内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 知识点 | 了解 | 掌握 | 重点 | 难点 |
| Zookeeper的简介 | √ |  |  |  |
| Zookeeper的特性 | √ |  |  |  |
| Zookeeper集群角色 |  | √ |  |  |
| Zookeeper的数据模型 |  | √ |  |  |
| Zookeeper的Watch机制 |  | √ | √ |  |
| Zookeeper的选举机制 |  | √ | √ |  |
| Zookeeper分布式集群部署 |  | √ | √ | √ |
| Zookeeper Shell操作 |  | √ |  |  |
| Zookeeper Java API操作 |  | √ |  |  |
| Zookeeper典型应用场景 |  | √ |  |  |

1. 思考与实践

本章在完成对相关内容的讲述基础上，需要学生动手完成相关实验，实验有一定难度。该组件是如何保证工作顺利进行的，这值得学生思考和总结。对应到实际生活中，理解我国民主选举方式的优越性。

1. 教学方法与手段

介绍本章教学主要采用课堂讲授、多媒体教学、网络辅助教学、团队合作、分组讨论、课堂讨论等。

**第六章 Hadoop2.0新特性**

1. 目的与要求
2. 掌握YARN的体系结构和工作流程
3. 掌握HDFS的高可用架构
4. 会搭建Hadoop高可用集群
5. 教学内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 知识点 | 了解 | 掌握 | 重点 | 难点 |
| Hadoop2.0改进与提升 |  | √ |  |  |
| YARN体系结构 |  | √ | √ |  |
| YARN工作流程 |  | √ |  | √ |
| HDFS HA的搭建方式 |  | √ |  |  |
| Hadoop的高可用架构 |  | √ |  |  |
| 启动Hadoop HA方式 |  | √ |  |  |

1. 思考与实践

本章在完成对相关内容的讲述基础上，需要学生动手完成相关实验，实验有一定难度。在完成某项工作后，不应该止步不前，要思考如何改进，要有精益求精的工匠精神。

1. 教学方法与手段

介绍本章教学主要采用课堂讲授、多媒体教学、网络辅助教学、团队合作、分组讨论、课堂讨论等。

**第七章 Hive数据仓库**

1. 目的与要求
2. 了解Hive的相关功能和特点
3. 熟悉Hive的简单安装和配置
4. 掌握HiveQL的相关操作

1. 教学内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 知识点 | 了解 | 掌握 | 重点 | 难点 |
| 数据仓库简介 | √ |  |  |  |
| 数据仓库的结构 | √ | √ |  |  |
| 数据仓库数据模型 | √ |  |  | √ |
| Hive简介 |  | √ |  |  |
| Hive系统架构 |  | √ | √ |  |
| Hive工作原理 |  | √ | √ | √ |
| Hive数据模型 |  | √ | √ |  |
| Hive安装模式 |  | √ |  |  |
| Hive的管理方式 |  | √ |  |  |
| Hive内置数据类型 |  | √ |  | √ |
| Hive的操作方式 |  | √ |  |  |

1. 思考与实践

本章在完成对相关内容的讲述基础上，需要学生动手完成相关实验，实验有一定难度。在大数据情形下，过去不见的问题有可能会出现，这是量变引起质变原理的一个现实体现。学生不仅是需要学习处理问题的技术，更应重视背后蕴含的哲学原理。

1. 教学方法与手段

介绍本章教学主要采用课堂讲授、多媒体教学、网络辅助教学、团队合作、分组讨论、课堂讨论等。

五、各教学环节学时分配

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **教学环节**  **教学时数**  **课程内容** | **讲**  **课** | **习**  **题**  **课** | **讨**  **论**  **课** | **实验** | **实习** | **其他教学环节** | **小**  **计** |
| 第一章 | 4 |  |  |  |  |  | 4 |
| 第二章 | 4 |  | 2 | 4 |  |  | 10 |
| 第三章 | 4 |  |  |  |  |  | 4 |
| 第四章 | 6 |  | 2 | 6 |  |  | 14 |
| 第五章 | 4 |  |  | 2 |  |  | 6 |
| 第六章 | 2 |  |  | 2 |  |  | 4 |
| 第七章 | 4 |  |  | 2 |  |  | 6 |
| 合计 | 28 |  | 4 | 16 |  |  | 48 |

六、课程考核

（一）考核方式

考试

（二）成绩构成

平时成绩占比： 40% 期末考试占比：60%

（三）成绩考核标准

七、推荐教材和教学参考资源

八、其他说明

大纲修订人： 邹杨 修订日期：

大纲审定人： 审定日期：