河北师范大学《Hadoop架构》课程教学大纲

（理论课程）

课程代码：32201090

课程名称：Hadoop架构

英文名称：Hadoop architecture

授课语言：中文

开课单位：软件学院

大纲制定人：王伟

大纲审定人：王伟

一、课程说明

**1.课程类别/性质：**专业课程/选修课

**2.学分/学时：** 3.5/80

**理论学时：** 32 **实践学时：**48

**3.适用专业：** 软件工程

**4.先修课程：** 无

**5.教材及参考书目：**

参考书目：

[1]深入理解大数据，黄宜华，机械工业出版社，2015年6月，ISBN:9787111473251，否

**6.课程考核方式：** 闭卷考试

**7.主要实践教学环节：**

上机实践主要内容:

搭建Hadoop环境，包括HDFS、Hive、HBase等。使用命令和Java编程完成对HDFS文件的操作，实现HBase数据的维护，编写MapReduce程序实现大数据的分析和挖掘。

二、课程简介

大数据处理是现在最为流行的数据挖掘扩展领域。在数据挖掘课程的基础上延伸在大数据领域的应用是计算科学专业的一大研究方向。本课是计算机科学与技术专业的一门重要的专业实践课，也是软件工程专业的一门院专业选修课。课程主要包括两部分内容：大数据处理的基本概念及Hadoop的基本概念知识及常见算法讲解。

本课程的主要特色是理论内容和实践操作相结合。在理论学习中掌握方法，理解原理，在实践过程中巩固理论知识，并掌握相应的开发工具进行编码。此外，还鼓励学生进一步思考，结合科创等实践活动，解决自己感兴趣的问题，提高学生发现问题和解决问题的能力。

三、课程目标

1.掌握大数据的基本概念

2.掌握HDFS的知识和编码

3.掌握MapReduce的思想和编码

4.掌握Hive的使用

5.掌握HBase的使用

6.了解机器学习的知识

四、课程目标与毕业要求的对应关系

|  |  |  |  |
| --- | --- | --- | --- |
| **课程目标** | **对应章节** | **支撑毕业要求** | **备注** |
| 课程目标1 | 第1章 | 毕业要求2、12 |  |
| 课程目标2 | 第2章 | 毕业要求3、4、5 |  |
| 课程目标3 | 第3、4章 | 毕业要求3、4、5 |  |
| 课程目标4 | 第5章 | 毕业要求3、5 |  |
| 课程目标5 | 第6章 | 毕业要求3、5 |  |
| 课程目标6 | 第7章 | 毕业要求2、3 |  |

五、教学内容及要求

**章节名称：**1. 大数据和Hadoop概述

**主要内容：**

大数据的背景、概念

大数据案例

大数据技术介绍

Hadoop简介

**基本要求：**

掌握大数据的概念

了解大数据的案例

了解大数据技术

了解Hadoop

**重点：**

掌握大数据的概念

了解大数据技术

了解Hadoop

**难点：**

了解大数据的案例

了解大数据技术

**章节名称：**2.HDFS和HA

**主要内容：**

HDFS概述

HDFS交互操作

HDFS编程

HA

**基本要求：**

理解HDFS组成和原理

掌握HDFS常用命令

掌握HDFS编程

了解HA

**重点：**

理解HDFS组成和原理

掌握HDFS常用命令

掌握HDFS编程

**难点：**

掌握HDFS编程

**章节名称：**3.MapReduce原理和Yarn

**主要内容：**

MapReduce原理

MapReduce执行流程

MapReduce开发

Yarn

**基本要求：**

理解MapReduce原理

理解MapReduce执行流程

掌握MapReduce开发

了解Yarn

**重点：**

理解MapReduce原理

理解MapReduce执行流程

掌握MapReduce开发

**难点：**

理解MapReduce执行流程

掌握MapReduce开发

**章节名称：**4.MapReduce程序设计

**主要内容：**

MapReduce的三个层次

MapReduce的组件

基本MapReduce算法

**基本要求：**

理解MapReduce的三个层次

掌握MapReduce的组件

了解MapReduce的几个基本算法

**重点：**

理解MapReduce的三个层次

掌握MapReduce的组件

**难点：**

理解MapReduce的三个层次

**章节名称：**5.Hive

**主要内容：**

Hive简介

Hive的部署与安装

Hive的使用

Hive JDBC编程

**基本要求：**

了解Hive

掌握Hive的部署与安装

掌握Hive的使用

掌握Hive JDBC编程

**重点：**

理解Hive工作原理

掌握Hive的使用

**难点：**

理解Hive的工作原理

掌握Hive的使用

**章节名称：**6.HBase

**主要内容：**

HBase简介

HBase核心知识

HBase安装部署

HBase的开发

**基本要求：**

理解组件和工作原理

理解HBase物理结构和数据逻辑结构

掌握HBase常用命令

掌握HBase开发

**重点：**

理解组件和工作原理

理解HBase物理结构和数据逻辑结构

掌握HBase常用命令

掌握HBase开发

**难点：**

理解HBase物理结构和数据逻辑结构

掌握HBase开发

**章节名称：**7.Mahout

**主要内容：**

Mahout介绍

机器学习介绍

Mahout算法介绍

**基本要求：**

了解Mahout

了解机器学习

了解Mahout算法及实现

**重点：**

了解Mahout算法及实现

**难点：**

了解Mahout算法及实现

**章节名称：**8.ZooKeeper

**主要内容：**

ZooKeeper简介

ZooKeeper命令

ZooKeeper编程

**基本要求：**

了解ZooKeeper

掌握ZooKeeper编程接口

**重点：**

掌握ZooKeeper编程接口

**难点：**

掌握ZooKeeper编程接口

六、实践教学环节

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **实验/设计 名称** | **实验/设计 内容与要求** | **学时/周** | **每组人数** | **备注** |
| 1 | HDFS编码 | 使用Java编码，对HDFS中文件进行管理 | 3/8 | 1 | 综合 |
| 2 | MapReduce编码 | 使用Java编码，编写MapReduce程序 | 6/8 | 1 | 综合 |
| 3 | Hive | 使用Hive命令对数据进行报表分析 | 3/9 | 1 | 综合 |
| 4 | Hbase | 使用HBase命令和Java编码，对数据实现增删改查等操作 | 3/9 | 1 | 综合 |
| 5 | Mahout | 使用Mahout包里的内容，实现推荐算法 | 3/10 | 1 | 综合 |

七、学时分配

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **章节内容** | **理论** | **实验** | **课程设计** | **…** | **总学时** |
| 1 | 大数据和Hadoop概述 | 2 | 3 |  |  | 5 |
| 2 | HDFS和HA | 4 | 6 |  |  | 10 |
| 3 | MapReduce原理和Yarn | 2 | 3 |  |  | 5 |
| 4 | MapReduce程序设计 | 2 | 3 |  |  | 10 |
| 5 | 综合练习（1-4章） | 4 | 6 |  |  | 10 |
| 6 | Hive | 2 | 3 |  |  | 5 |
| 7 | HBase | 4 | 6 |  |  | 10 |
| 8 | Mahout | 4 | 6 |  |  | 10 |
| 009 | ZooKeeper | 2 | 3 |  |  | 5 |
| 10 | 综合练习（5-8章） | 4 | 6 |  |  | 5 |
| 11 | 课程总结 | 2 | 3 |  |  | 5 |
| **合 计** | | 32 | 48 |  |  | 80 |