面向计算机大类的大数据方向基础课程模块化教学研究

李实于鸣

东北林业大学信息与计算机工程学院, 中国·黑龙江 哈尔滨 150040

【摘要】大数据及相关技术的发展给企业获取、使用和管理信息的方式带来了范式级别的转变。这就要求每一个IT相关专业的学生都应该适应这种新范式的转变。本文研究为特定的核心课程开发大数据相关内容学习模块,使得一些这些计算机基础类课程的覆盖后面大数据类专业课程的基础范围。

【关键词】大数据; 计算机基础课; 计算机大类

【项目】黑龙江省高等教育教学改革资助项目(SJGY20190031)。

1 介绍

在全球信息化快速发展的大背景下,数据已成为国家重要的基础性战略资源,大数据正日益渗透到社会生活和经济发展的各方面。目前大数据人才培养的挑战在于,大数据和云计算范式的许多工具和技术只是在最近几年才出现,许多院校围绕这些指导方针和要求开发计算机科学课程,因此无法将这些当代主题纳入密集课程的核心课程。特别是目前很多高校的数据科学与大数据专业属于计算机大类,到第四学期甚至第五学期才能进入真正的专业课程学习,同时一些其他计算机大类专业也要学习大数据类的基础知识,所以相关主题需要整合到计算机本科课程的多个核心课程中,而不是创建额外的独立核心课程或选修课程。

本文正式为了解决以上问题,考虑将大数据和云计算概念和技能的覆盖到未来从事大数据方向工作的计算机类本科生,专业教师需要更多地、逐步地对学生进行辅导,并且应将大数据相关知识和技术整合到课程的多个核心课程中,特别是计算机大类基础课。所以本研究探讨了在计算机大类的平台课中开发一系列简短、独立的学习模块,使得传统的课程覆盖范围适应大数据从业者的需求。

2 模块详细信息

建议的模块设计为简短,并倾向自主学习,这样更容易操作。因此,每个模块跨越两个课时(大概90分钟),包括具体的学习目标、课堂讲稿和评估资源。评估资源包括测验问题、教程、示例程序、实践项目等。这些模块预计将在实验室环境中进行,第一节通常用于讨论新概念和技术,第二节用于向学生介绍动手项目以及完成项目所需的必要工具和资源。还有一类课下作业项目包含多个任务,通常有给学生两周的时间去自己完成。

2.1 “计算机硬件基础”(CH)模块

本模块整合到“计算机硬件”类课程中,旨在让学生了解虚拟化和云计算基础知识。在我们学校,这是计算机科学、软件工程、数据科学与大数据技术专业的共同基础课,大部分是大一和大二学生。具体的学习内容是:

- (1) 认识云计算的关键特性、技术、优势和挑战。
- (2) 认识Linux系统的初步认识,和基础的操作shell命令。

方法:首先对云计算重要特性的进行初步讨论,如可伸缩性、按需访问、可测量服务和弹性。同时解释资源共享和虚拟化,还有其重要方面如:迁移、分时、隔离等。还要探讨在企业内部利用云服务的好处以及采用相关技术所需要迎接的挑战,如数据机密性、性能不可预测性;介绍Linux和shell命令的基本知识。

2.2 “算法分析”(AA)模块

本模块是为需要进行算法课程学习的学生设计的,主要介绍流行的大数据分析引擎,如Hadoop和Spark。本课程基于已完成数据结构课程,要求熟悉并熟练使用Java或Python等编程语言的计算机大类学生,主要是大二的学生。

本模块的学习内容如下:

(1) 认识MapReduce和Spark框架的关键特性、技术、优势和挑战。

(2) 使用Hadoop和HDFS构建基于MapReduce的应用程序。

(3) 使用Spark/Hadoop平台在云平台上开发大数据分析的基本方法。

方法:首先探讨并行计算,并介绍MapReduce框架,它可以通过将大型数据集拆分为单独的并行处理模块来处理这些数据集。然后用示例介绍MapReduce编程的基础知识。Apache Hadoop作为MapReduce模型的一个具体实现,在课程中会介绍它的分布式文件系统HDFS。同时还要介绍Hadoop集群的概念以及框架。最后,简要讨论MapReduce编程中存在的问题和Spark的运行环境。

这一模块的第二部分为实验课,老师需要提供一些示范和指导,让学生跟着学。包括设置HDFS,编译和执行MapReduce应用程序,并且基于Spark会做一个数据分析的实际案例。

2.3 “数据库管理”(DB)模块

目前,许多应用场景都需要以可扩展和分布式的方式处理非常大的数据集。为了迎接这一挑战目前出现了各种类型的大数据系统。数据库管理这一模块旨在让学生接触各种类型的大数据系统,并在这些系统中集成SQL的学习。这将集成到计算机大类学生的数据库类平台必修课,通常由大二和大三学生参加。

本模块的学习内容如下:

(1) 认识重要的大型数据库管理系统(BDBMS)的关键属性、优势和局限性,如MapReduce、No-SQL;

(2) 在Spark框架内使用SQL加载和查询大型数据集的基础实践技能。

方法:在两个学时的教学中讨论关系数据库系统的局限性,以及各种大数据管理系统(如MapReduce、No SQL)的关键属性、优势和局限性。

3 结论

本研究的目的是探讨将大数据相关知识和技术三个模块整合进计算机大类的核心课程中。在计算机类学生未进入专业学习之前就打下大数据和云计算的一些基础,接触到这些前沿的知识和技术,而无需开发额外的核心课程或选修课。

参考文献:

[1] 夏大文, 王林, 张乾, 魏嘉银, 冯夫健, 李华青. 大数据应用技术课程教学改革与实践. 大数据. 2020; 6(4).

[2] 林子雨. 大数据技术原理与应用课程建设经验分享. 大数据. 2018; 4(6).

作者简介:

李实(1976.1-)女,黑龙江省哈尔滨,博士,数据科学与大数据技术专业主任,副教授,研究领域:数据挖掘商务智能。

于鸣(1981.4-)男,黑龙江,硕士,系副主任,副教授,研究方向:智慧农业。