文章编号：1672-5913(2017)02-0014-04　　中图分类号：G642

面向系统观培养的大数据系统课程建设

李　弋，张为华，赵文耘
（复旦大学 软件学院，上海 201203）

摘　要：分析计算机系统基础（ICS）课程教学中存在的问题——内容缺乏综合性、未结合最新技术、两学期教学不利于学生的系统认知等，提出强化系统观培养的 ICS 后继课程——大数据系统教学设计，包括课程的目标、理念，内容、实验设计以及它和 ICS 之间的关系。

关键词：系统教育；计算机系统基础；大数据；系统观

1　背　景

随着信息技术的发展，计算机系统处理的对象呈现网络化、多媒体化、大数据化、智能化等特征。计算模式发生了改变，形成嵌入式计算、移动计算、并行计算、服务计算等多种计算模式。这些变化对计算机专业人才的系统知识结构、大局观和创新能力提出了更高的要求。

为了应对挑战，ACM/IEEE CS2013[1] 在课程体系中增加了新的系统课程。教育部高等学校计算机类专业教学指导委员会（以下简称教指委），组织开展对计算机专业学生能力培养和实践教学体系的研究 [2]，总结了计算机专业高级人才应具有的四大专业基本能力：计算思维能力、算法设计与分析能力、程序设计与实现能力和系统能力，其中系统能力占总能力的 75%。通过分析计算技术的发展，教指委明确系统能力表现“在掌握计算机系统基本原理的基础上，熟悉如何进一步开发构建以计算技术为核心的应用系统”[2]。

根据系统能力培养的内涵和需求，教指委提出了课程体系设置的总体思路，建议开设基础课程计算机系统基础（以下简称 ICS）[2]。ICS 课程由卡耐基 • 梅隆大学（CMU）的 Bryant 等人 [3] 开发，综合介绍计算机系统的基础概念。全球近 300 家机构，包括中国内地的 20 多所高校开设了 ICS 课程。

国内的 ICS 教学会结合实际情况对授课形式和内容进行调整。普遍地，课程分两学期讲授，内容的顺序也有所调整。ICS 课程的本地化取得了很好的教学效果，但教学中也存在一些不足，具体包括：课程组织松散，欠缺整体性；课程实验综合性不够，不能促进学生全面理解系统；内容没有体现计算机系统的技术发展趋势。

为了弥补 ICS 教学的不足，强化对系统概念的综合理解，培养系统观，笔者提出将大数据处理建设成一门系统课程，并将其作为 ICS 的后继课程。课程由理论和实验组成：理论以热门的大数据处理为载体，结合大数据系统的前沿技术，强化 ICS 的相关知识，促进学生对系统的整体认知；课堂实验使学生熟悉相关主要系统和工具，课后综合实验提升学生理解和运用概念的能力，培养学生处理复杂系统的能力。

2　计算机系统基础（ICS）

2.1　CMU 的 ICS 课程介绍

ICS 的目标是“解释计算机系统的本质概念，并展示这些概念对应用程序正确性、性能和可用性的影响”[4]。学习后，学生应理解应用的运行，并编写高性能的应用程序。

教材从程序员的角度出发，分 3 部分介绍计算机系统：

（1）程序结构和执行：理解程序和硬件间的关系。描述程序及信息在硬件上的表示，介绍程

第一作者简介：李弋，男，讲师，研究方向为计算机系统软件和体系结构，liy@fudan.edu.cn。

序运行的硬件——处理器和存储系统，结合硬件特性优化程序。

（2）程序在系统中的运行：理解程序和操作系统间的交互。介绍可执行程序的生成；描述操作系统管理处理器和内存资源的机制——进程和虚存，支持应用程序的运行。

（3）程序间的通信和交互：理解程序和外界的交互。系统 I/O 和网络通信提供基本的交互机制；并发支持多种交互模式，更加高效地利用资源。

ICS 共设计了 9 个课后实验，覆盖对应章节的基本概念。这些实验提高了学生对课程的兴趣，促进学生对基本概念的理解。

2.2 问题分析

ICS 已成为国内计算机专业系统能力培养的核心基础课程，普遍在二年级分两学期讲授。ICS 对计算机系统的综合介绍，为后续课程的学习奠定了坚实基础；促进了系统方向教学的改革。然而，现有课程内容和授课组织等方面也存在以下不足：

（1）内容组织相对松散，概念间结合不够，不利于综合理解。两学期教学虽然让学生有充裕的时间学习和理解，但破坏了连续性，妨碍学生对计算机系统的整体认识。

（2）实验只覆盖部分基础概念，缺乏贯穿系统的综合性实验，不利于培养学生的系统观和解决复杂系统问题的能力。

（3）内容以经典的 PC 系统为主，没有结合当前主流技术的发展。学生日常生活多与大数据和移动计算相关，理论和生活的差异弱化了学习兴趣。

ICS 课程培养学生对计算机系统的基本认知；操作系统、计算机网络、体系结构和编译原理等课程，则聚焦方向的专门知识。在这种教学体系下，培养学生的系统观需要一门综合课程，弥补 ICS 教学的不足。该课程具有以下特点：

（1）基础性：课程面向计算机专业，构建在 ICS 基础之上，开设在三年级。

（2）承接性和强化性：ICS 的相关知识内容在该课程中都有对应知识，是 ICS 课程相关知识的强化、递进和综合运用。

（3）面向技术潮流：内容贴近日常生活，选自主流技术；并展现系统对应用的影响，促进学生系统理解和运用系统基础知识。

3 大数据系统课程

大数据系统课程基于 ICS，结合大数据处理的特点建立课程知识体系，介绍领域内前沿进展，从 3 个方面强化系统观的培养：①课堂讲授大数据系统对 ICS 扩展的核心知识；②通过课堂实验让学生了解大数据系统的各种软件架构，掌握主流软件框架和相关工具的使用；③通过课后实验培养学生综合运用系统知识解决复杂问题的能力。

3.1 总 述

随着数据规模的不断扩大和智能技术的发展，大数据计算成为热点。大数据系统把任务分配到多台机器上执行，是一种分布式的系统。与 PC 系统相比，大数据系统具有多层次、高复杂性和综合性强的特点。

大数据系统通过高速互联网络组成计算机集群，共享每个节点的计算和存储资源。程序并行运行，访问的存储层次更复杂；程序运行和优化涉及更多的因素。类比 ICS 课程，大数据系统的 3 层抽象逻辑如下：

（1）大数据系统程序结构和执行：大数据应用通常为并行执行，使用 ICS 课程的程序表示和网络编程。计算机系统包括单节点系统和高速互联部分。存储系统中的本地存储结构更复杂，还涉及远程访问等因素。在系统优化时，除了性能优化外，还涉及功耗优化。应用 Amdahl 定律时，也要进行必要的调整。

（2）程序在大数据系统中的运行：大数据系统除涵盖 ICS 的知识外，还涉及计算机集群运行时环境和资源管理的概念，同时还包括运行时的监控、调度和优化等。

（3）程序间的通信和交互：编程框架隐藏了 ICS 讲授的实现细节，MapReduce、内存计算和图计算扩展了并发编程的模型。

大数据系统在单个节点上应用 ICS 的知识，并扩展了每层的内涵和外延。表 1 给出了大数据系统与 ICS 课程的对比，大数据系统不仅延续了 ICS 课程的内容，还有很多相关知识的扩展和外

表 1　大数据系统对 ICS 的扩展

抽象层	ICS 课程	大数据系统的扩充
程序的结构与执行	处理器和存储器 信息和程序的计算机表示 结合硬件特性的优化	多机并行的复杂计算系统 分布式存储，复杂的存储层次 分布式程序的结构 多机环境下的应用优化
程序的运行	可执行程序的生成 进程和虚拟存储系统	分布式程序的构成和加载 多机任务的协调、调度和资源管理
程序间通信和交互	系统 I/O 和网络通信 并发编程	编程框架对系统和网络 I/O 的隐藏 编程模型和应用的适配

延，同时体现了 IT 技术发展的主流趋势。因此，大数据系统较适合作为 ICS 的后继课程，大数据系统对相关知识的集中运用，强化了计算机系统的整体理解，锻炼了学生处理复杂系统的能力，有助于培养学生的系统观。

3.2　理论内容

理论内容主要讲授大数据系统的核心理论框架，整体可分为 3 层，包括硬件平台（数据中心计算机）、软件平台（大数据系统的运行时）和大数据编程框架（开发大数据应用）。

数据中心计算机主要包括分布式计算和存储系统，以及高速互联系统。同时，这部分还涉及可扩展性、容错和低功耗等内容。硬件的特性、组织结构和系统目标，影响软件的构造和表现。因此，硬件平台相关的授课内容将围绕这些方面展开。

软件平台介绍大数据应用运行的基础服务，主要包括分布式文件系统、非结构化分布式数据库、资源的管理和监控以及任务的调度等。了解系统软件运行时的功能、特性和工作原理，有利于实现高效的大数据应用。

编写程序处理大规模数据非常具有挑战性。大数据系统的编程框架，极大地简化了这类应用的编程。根据应用的特点选择编程框架是开发高效应用的关键。因此，课程将选择代表性的大数据编程框架，如 MapReduce，讲授编程框架的特点和工作原理。

3.3　实践设计

实践分为课堂和课后实验：课堂实验促进学生理解知识体系，熟悉典型的环境和工具。课后实验强化理论与实践的结合，要求学生运用理论知识解释应用的表现，培养学生的综合能力。

3.3.1　课堂实验

大数据系统的核心软件架构如图 1 所示，包括分布式文件系统、数据存储、公共服务、访问模型、计算模型和数据分析等组件；基于不同的设计目标，每个组件有多种实现。课堂实验基于主流大数据系统的典型组件开展。

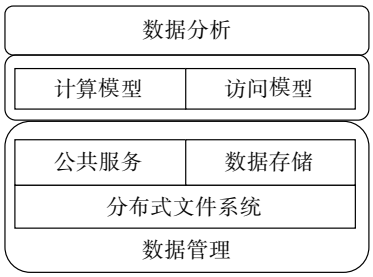


图 1　典型大数据系统的软件架构

Hadoop 是开源的大数据计算平台，广泛应用于工业和学术界。Hadoop 具有很好的兼容性和伸缩性，可支持一个组件的多种实现。课程可以根据教学需求构建实验系统。

通过课堂演示和动手实践，学生配置并使用 Hadoop 系统的核心典型组件。主要内容包括：

（1）掌握分布式文件系统 HDFS 的框架、操作命令和访问接口；

（2）掌握典型的 NoSQL 数据库 HBase 的框架、操作命令和访问接口；

（3）用 Hive 访问 HBase 中的数据；

（4）掌握基于编程框架的应用开发，学习编程框架 Spark 和 Storm；

（5）了解典型的大数据处理工具，如统计软件 RHadoop；

（6）学习 Hadoop 核心组件的配置，理解不同配置对系统性能的影响；特别地，通过制造故障展示容错机制对系统的意义。

在实验时，教师要引导学生观察应用的表现，结合理论分析现象的原因，帮助学生理解大数据处理平台的运行，并强化基本概念的运用。

3.3.2　课后实验

通过课后实验的训练，学生能初步综合运用大数据系统提供的服务和相关的系统概念，独立

实现大数据应用。一方面，课后实验要覆盖大数据系统的各层面，以培养学生的综合能力。另一方面，实验不能太复杂，要兼顾不同能力的学生。

一般而言，学生此时已经修完数据结构，课后实验可选用海量数据排序 TeraSort，要求学生分阶段实现并优化。学生在第一阶段实现单机上内存和外存排序的串行版本，并优化性能；再实现并行多线程版本。在第二阶段，学生基于 MapReduce 和 Spark 实现不同的大数据版本。在第三阶段，学生分析应用的瓶颈，探索提高分布式应用程序性能的方法。

在实践中，学生还应该理解：

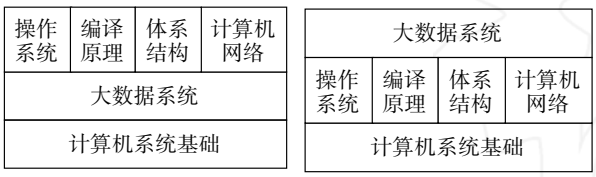
（1）编写单机和分布式程序的差异；

（2）数据的分布和读写方式，以及网络架构对应用表现的影响；

（3）系统对硬件的使用方式可显著影响应用的性能，如 MapReduce 和 Spark。

4　讨　论

作为 ICS 的后继，大数据系统课程有多种开设方式，如图 2 所示。课程可紧接在 ICS 之后开设，也可在其他系统专业课程（如体系结构、操作系统、计算机网络和编译原理等）之前、平行或之后开设。甚至，本课程可作为一门大规模的实践课程独立开设。大数据系统课程要根据与其他课程的先后关系和实际情况调整内容，以适应学生的知识结构。



(a) 作为 ICS 的直接后继　(b) 系统方向的综合课程

图 2　典型的开课方式

（1）一些高校基于 Patt[5] 或南大 [6] 的教材开设 ICS 课程，没有涉及网络和并发编程，因此，大数据系统课程需要增加这两方面的知识。

（2）如果专业方向不用修读系统专业课，课程应引入更多现代系统的概念并增加课时，提高系统能力培养的强度。

（3）针对不同层次的教育，授课内容和方式可进一步调整。普通高校可减少理论内容，强化课堂实验，并结合实验讲授系统概念。

（4）针对学生的不同水平，综合实验部分可采取分层评分的方式。

数据中心计算机 [7] 可作为体系结构的参考。大数据技术还在持续发展，各方面知识需结合最新的研究进展和软件版本作调整。讲课时，课程内容要根据实践条件和学生接受情况进行调整。在课程知识体系稳定后，可进行教材的整理和编写。

由于大数据是各个领域研究的热点，课程可结合最新研究进展引入研讨环节。例如，结合近年主流会议或期刊的论文，组织学生深入研讨一些开放性的内容。课程实验内容也可根据各校具体情况进行调整和增强，如结合异构平台的数据处理等。

5　结　语

针对现有 ICS 教学和系统课程培养知识体系的特点和不足，我们设计了“大数据系统”课程。课程理论结合实践，通过构建大数据应用系统，学生既了解了当前主流系统体系和最新研究进展，更强化了学生系统观的培养。今后，我们可以参考大数据系统课程，建设 Android 移动平台相关系统课程。该课程将作为我们的未来系统课程改革和建设的方向之一。

参考文献：

[1] Computer science curricula 2013[EB/OL]. [2016-07-03]. http://www.acm.org/education/CS2013-final-report.pdf.
[2] 王志英, 周兴社, 袁春风, 等. 计算机专业学生系统能力培养和系统课程体系设置研究[J]. 计算机教育, 2013(9): 1-6.
[3] Bryant R E, O' Hallaron D R. Introducing computer systems from a programmer's perspective[J]. ACM SIGCSE Bulletin, 2001,33(1): 90-94.
[4] Bryant R E, O' Hallaron D R. Computer systems: A programmer's perspective[M]. 3rd. ed. Boston: Pearson, 2015.
[5] Patt Y N, Patel S J. Introduction to computing systems: From bits and gates to C and beyond[M]. 2nd ed. Berkeley: McGraw-Hill Higher Education, 2004.
[6] 袁春风. 计算机系统基础[M]. 北京: 机械工业出版社, 2014.
[7] Barroso L A, Clidaras J, Holzle U, et al. The datacenter as a computer: An introduction to the design of warehouse-scale machines, 2nd ed. [J]. Synthesis Lectures on Computer Architecture, 2013, 8(3): 1-154.

（编辑：彭远红）