

大数据时代数据库系统课程建设研究与实践

戴牡红

(湖南大学信息科学与工程学院, 湖南 长沙 410082)

摘 要: 针对新一代NoSQL和NewSQL技术, 以及软件工程专业本科生数据库课程教学存在授课内容陈旧、学生数据库应用系统分析与设计能力不够、课程讲授与实验一直采用国外数据库产品等现状, 重构数据库教学中的讲授内容和课程实验内容。将思政教育元素融入数据库教学中, 构建新一代数据库教学课程知识体系。阐述大数据时代的数据库系统课程建设研究与实践, 包括教学目标、课程内容、课程结构、教学计划和课程评价方法。实践结果表明, 学生们对最新一代的数据库技术很感兴趣, 课程效果良好, 有效提高了学生数据库系统分析、设计与实现能力。

关键词：大数据；NoSQL；NewSQL；数据库；课程

中图分类号: TP392 文献标识码: A

Research and Practice on the Construction of the Database System Course in Big Data Era

DAI Muhong

(College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China)

Abstract:Aiming at the new-generation NoSQL and NewSQL technology as well as the outdated teaching contents in database course for software engineering undergraduates,insufficient student's database application system analysis and design ability,course teaching and experiment depending on foreign database products,this paper reconstructs the contents of database teaching by means of integrating the elements of ideological and political education into the database teaching,and builds a new-generation database teaching course system.The paper describes the research and practice of database system course construction in the era of big data,including teaching objectives,course content,course structure,teaching plans and course assessment methods.Practical results show that students are very interested in the latest generation database technology and the course effect is good,which effectively improves student's database system analysis,design and implementation capabilities.

Keywords:big data;NoSQL;NewSQL;database;course

1 引言(Introduction)

数据库技术在过去5到10年内经历了前30多年来最重大的变化，非关系型的大规模分布式数据库系统不断出现，这些新技术提供了前所未有的大数据处理能力。目前国内许多大学的数据库系统课程主要讲授的内容是关系代数、SQL语言和E-R建模，对以NoSQL、NewSQL等复杂数据的管理技术不做讨论，不能适应大数据时代企业的数据管理与数据分析需求。

2 数据库技术的发展(Development of database technology)

自1970年以来,数据库技术发展到了第四代,如图1所示^[1]。第一代是20世纪60—70年早期的层次和网状数据库系统,第二代是70年代中期开始的关系数据库系统,第三代是80年出现的对象数据库系统,以及90年出现的对关系模型进

行扩展,使关系数据库系统具有面向对象功能。

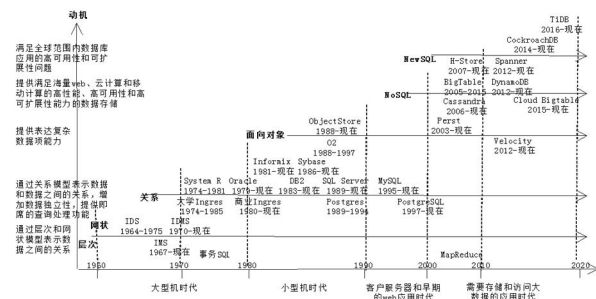


图1 数据库技术和主要数据库产品的发展

Fig.1 Database technology and development of major database products

自2000年以来，Web技术、社交网络、移动计算和物

联网的进步导致全球范围应用程序生成的半结构化、非结构化数据呈现爆炸式增长态势, 此类应用程序产生了各种不同的需求, 包括水平可扩展能力、支持大规模海量数据处理能力, 系统具有高可用性和容错能力以应对硬件和软件故障, 并且可以通过利用额外资源响应客户端请求, 从而提高查询处理速率, 这就导致了第四代数据库技术NoSQL和NewSQL技术的出现。

NoSQL数据库的关键特征是: (1)采用图形、树、键值、文档等表格模型以外的数据模型; (2)数据库设计时没有固定的模式定义; (3)提供对水平可扩展性的隐式支持; (4)具有最终一致性而不是强一致性^[2]。NoSQL的另一个重要概念是CAP定理, 它对分布式系统至关重要的属性提供了一些约束, 即一致性、可用性和分区容错性。CAP定理表明, 分布式系统在任何给定时间最多只能满足其中两个属性。

NewSQL是最新一代的数据库系统, 它不仅提供NoSQL数据库系统具有的可扩展性和可用性, 还保留了传统关系数据库的ACID特性、关系数据模型和SQL查询语言^[3,4]。

3 数据库教学中存在的问题(Problems in database course teaching)

云计算、并行处理和分布式存储技术等大数据开发技术的飞速发展, 彰显数据库课程教学中存在问题^[5]。

(1)数据库课程教学以基于关系代数、SQL语言和E-R数据库设计模型, 讲授数据管理的方法和技术。这些技术与方法难以高效处理海量的非结构化数据。随着新一代WEB技术的广泛应用, 关系型数据库系统不适用的环境越来越多, 应运而生了许多新的非关系型数据库处理技术NoSQL和NewSQL。因此, 数据库教学中应增加有关NoSQL、NewSQL等新一代的数据管理技术。

(2)目前的数据库系统课程教学主要采用MySQL、SQL Server等国外的数据库产品。近年来, 国产数据库正在以开源面貌焕发新一代的生机, 新一代的NoSQL、NewSQL等国产数据库软件正在崛起。在数据库系统课程的教学和课程实验中, 采用主流的、最新的国产数据库软件, 对培养学生的爱国主义情怀非常有必要。

(3)高等教育的目标是立德树人^[6], NewSQL数据库技术对非关系型数据处理技术的优势及其广泛应用为当前思政教育的革新指出了新的方向。将思政教育融入到数据库系统等专业课程的教学, 建立思政素材数据库, 将思政素材潜移默化地引入到数据库系统的课程教学和实验教学中, 重新设计数据库课程的SQL案例和数据库实验数据集, 采用思政素材, 非常有必要^[7]。

4 课程目标(Course objectives)

目前NoSQL数据库和NewSQL数据库并没有类似SQL的ISO/IEC标准, 它们有自己独特的功能, 都是针对特定问题而开发的, 并带有它自己的基础概念和交互语言。课程中

我们涵盖数据建模、SQL语言。学生将能够达成以下目标和能力: (1)了解关系数据库和NewSQL数据库的基本存储结构和访问技术。(2)使用关系数据库和NewSQL数据库特有的特征, 可以依据应用场景开发关系型和非关系型数据库应用系统。(3)识别新数据库与先前已知数据库之间的语义相似性和差异。(4)区分常用的一致性模型。(5)根据要存储的数据和系统的可用性、可伸缩性和一致性要求, 确定要部署的NewSQL/SQL数据库系统。

5 课程建设内容(Course construction contents)

为了上述课程目标的达成, 课程中将讨论如下主题: 关系数据库体系结构、关系数据模型、关系代数与SQL语言、事务处理、非关系型数据建模; NewSQL中的数据库定义与查询处理技术、应用软件系统的开发。所有这些知识的掌握通过课堂测验和课程实验达成。课程建设的主要内容包括:

(1)基础知识讲授, 数据库概念、E-R模型、SQL语言、事务处理与故障恢复技术、并发控制, 以及数据库安全访问技术。

(2)NewSQL探究性知识讲授, 非结构化数据表示方法; NewSQL数据建模方式, 主要有key-value、Column-oriented、Document等; 主流的NoSQL、NewSQL数据库系统概念; NoSQL的CAP理论。

(3)设计新的数据库系统课程实验, 以思想政治素材数据库的设计贯穿数据库系统课程实验中, 设计新的数据库系统课程实验, 实验的数据采用思想政治素材数据库中的数据。

(4)采用TiDB NewSQL进行教学与实验, 培养学生的爱国主义情操。TiDB是国内PingCAP团队开发的一个分布式NewSQL数据库系统, TiDB支持传统RDBMS和NoSQL的特性。在数据库领域, 国产正在崛起! 国产数据库的探索者和专家们, 为我们呈现数据库的隐秘世界和国产力量, 我们将带领学生跟随国产数据库领路人的步伐, 去发现国产数据库的强大。

6 课程组织结构(Course organization)

课程要求学生在掌握数据库基础知识的同时, 具有自主学习能力和工程项目能力, 我们将这一理念贯穿在数据库课程教学的整个环节, 通过课程讲授、课程测验、小班讨论、课程实验和小组项目开发, 培养学生的工程能力和团队协作精神。

6.1 课程讲授和课程测验

上课讲授核心的基础知识和应用开发技能, 并且针对课堂上介绍的各种概念、技术, 设计各种小的课堂测验, 以便了解学生们的掌握程度。

6.2 小班讨论

课程设计了4次、8个学时的小班讨论。每个小班分成由3—4个学生组成的讨论小组, 就数据库系统中的完整性约束、查询与统计、存储过程、数据库设计、事务管理、

NewSQL等关键问题和项目开发进行研讨,由学生宣讲,其他同学参与讨论,老师进行综述与点评。

6.3 课程实验

数据库技术的掌握,需要学生动手、动手再动手。实验指导书中给出了软件开发公司所使用的最新技术说明,通过实际的操作和实验,巩固课堂教学内容,让学生掌握数据库建模技术、SQL定义和查询处理技术,培养学生实际动手能力。每次实验前,教师向学生讲清实验的整体要求及实现的目标任务,实验1人一组,在规定的时间内,由学生独立完成,出现问题时,教师引导学生独立分析和解决,所有实验为一整体,有其延续性,每一次实验必须在实验课堂上认真完成。

6.4 小组项目

每个班级在课程的小班讨论中组成的项目小组必须利用课余时间完成一个使用NewSQL系统开发的应用项目,满足最低的功能复杂性要求。小组项目的目标是让学生学习和使用NewSQL数据库应用系统开发的分析、建模和实现能力,培养学生自主学习能力、知识获取和团队协作能力,提高创新能力,使之较早满足企业需求。

7 教学计划(Teaching plan)

课程总共72学时,48学时授课,8学时小班讨论,16学时课程实验,总计是16周,表1是16周的教学计划。

表1 课程教学计划

Tab.1 Course teaching plan

周次	课程讲授	周次	课程讲授
1	数据库系统概论、关系数据模型	9	数据库建模E-R模型、UML模型
2	关系代数	10	函数依赖理论
3,4	SQL-DDL	11	事务处理与故障恢复
4,5	SQL-DML	11,12	NoSQL、NewSQL
5	查询与统计	13	NoSQL中的CAP理论和BASE概念
6	SQL-DCL、数据库安全技术	14	WEB编程接口ODBC/JDBC技术
7	视图、触发器、存储过程、函数	15	关系数据库+NewSQL数据库混合编程技术
8	期中考试	16	课程总复习

8 课程评价方法与实施效果(Course assessment method and implementation effect)

数据库系统设计课程教学过程中,强调认知数据库的特征与特性、面临的问题及其解决方案,力求让学生对数据库知识的认知达到“当前面临这些问题,自然要这样做”的感性高度。课程以案例作为载体,展示数据库的特征与特性,然后引导学生进行观察和分析,把专业知识通俗化、形象化、直观化和具体化。课程将工具的使用作为实验教学的重点,强化学生将课程知识与解决具体问题相结合的训

练,培养学生的学习兴趣。学生通过解决具体实际问题,对数据库理论知识有了更好的理解和掌握。课程的总评成绩计算方法如下:总评成绩=课堂测试10%+小班讨论10%+期中考试10%+课程实验10%+作业10%+期末考试30%+小组项目20%。在教学中,我们强调的是数据库的设计和应用开发技能的培养,课程实验,老师自己编写实验指导书,每一个实验要求学生独立完成,并且每一个实验完成后必须发到老师的邮箱,老师对学生的每一次实验都要检查。这门课程学生的到课率是很高的,对每一次作业老师都会认真批改,许多学生认真完成了作业,许多学生认真做了实验,对作业中出现的普遍性问题在习题课上重点讲解,并且对所有的作业都给学生参考答案。所有的小班讨论课,全部是由老师组织,没有交给助教。任课教师希望20多年数据库应用系统开发的经验和技术分享给学生们。每一次课程的SQL代码,讲解的内容在课后4个小时内都会上传到课程网站。学生们普遍反映能够学到东西,成绩效果良好。

9 结论(Conclusion)

NewSQL数据库越来越多地用于大数据时代的Web应用系统之中,目前的本科数据库课程教学主要侧重于关系型数据库系统,对各种类型NewSQL和面向NewSQL的新数据模型的掌握是目前迫切需要学生掌握的技能。我们将新的技术融合到数据库课程的教学之中进行了有意义的探索,取得了一定成效。目前课程总评成绩的计算由课堂测试、小班讨论、期中考试、课程实验、作业、期末考试和小组项目等过程组成,各个部分所占比重如何合理?将是未来的教学中着重探究的问题。

参考文献(References)

- [1] Ali Davoudian,Liu chen,Mengchi Liu.A Survey on NoSQL Stores[J].ACM Computer Survey,2018,51(2):401-406.
- [2] Jaroslav Pokorný.Database Technologies in the World of Big Data[J].Proc of International Conference on Computer Systems and Technologies,2015:1-12.
- [3] Yasin N. Silva.SQL:From Traditional Databases to Big Data[J].Proc of SIGCSE,TN,USA,2016:413-418.
- [4] 徐述,汪彦,曾海洋.大数据应用下的新型分布式数据库NewSQL[J].数字技术与应用,2018,36(8):51-52.
- [5] 岳昆.大数据时代本科数据库课程体系改革设想[J].计算机教育,2015(11):104-107.
- [6] 赵厚宝,曾井泉.研发类课程开展课程思政教育的探索与实践[J].高教学刊,2019(2):188-190.
- [7] 余江涛,王文起,徐晏清.专业教师实践“课程思政”的逻辑及其要领——以理工科课程为例[J].学校党建与思想教育,2018(568):64-66.

作者简介:

戴壮红(1964-),男,硕士,研究员.研究领域:数据科学。