《分析型大数据系统》教学大纲

课程编号：071593B

课程类型：□通识教育必修课 □通识教育选修课

□学科基础课 □专业核心课

☑专业提升课 □专业拓展课

总学时：48 讲课学时：32 实验（上机）学时：16

学　分：3

适用对象：数据科学与大数据技术专业（信息技术）

先修课程：数据结构、数据库原理与应用、数据科学导论

一、教学目标

数据库是按照数据结构来组织、存储和管理数据的仓库，主要有关系型数据库和非关系型数据库两大类，分析型大数据系统即为非关系型数据库，分析型大数据系统是数据科学与大数据技术专业（信息技术）重要的专业选修课。通过本课程的学习，使学生对分析型大数据系统的基本概念和基础理论有系统的理解，掌握分析型大数据系统的基本操作、应用的基本技术以及数据库技术的最新发展，为学生增强理论基础，为后续课程学习及将来从事分析型大数据系统的设计、开发、应用及相关理论、实践研究奠定坚实的基础。为实现专业教育与思政教育的同向同行，发挥协同效应，本课程在建设过程中坚持分析型大数据系统理论知识和应用技术与思政教育一体化设计，从课程教学目标、教学内容、教学方法和评价考核上多个维度融入思政元素，帮助学生塑造正确的世界观、人生观和价值观，实现从专业知识传授、能力培养延伸到价值塑造，达到“润物细无声”的效果。

目标1：系统掌握分析型大数据系统的整体结构、基本概念和基础理论；

目标2：熟悉分析型大数据系统的主要实现方法；

目标3：能够熟练掌握分析型大数据系统的基本操作以及应用的基本技术。

二、教学内容及其与毕业要求的对应关系

本课程主要讲授分析型大数据系统的基本原理、基本特点以及主要实现方法等，并以MongoDB为例讲解分析型大数据系统的基本的增、删、改、查等基本操作以及MongoDB自身的一些机制。

本课程以《NoSQL指南》、《MongoDB大数据处理权威指南》等教材为基础，注重理论与实践相结合，采取实例教学法、小组教学法、模拟教学法等多种教学方法进行授课，加强学生对分析型大数据系统理论知识的理解，提高学生应用能力，培养学生的创新精神。要求学生在实验环节掌握分析型大数据系统的基本原理、基本操作及相关领域的实战应用。要求学生课后完成基于课堂和实验教学内容的扩展实践练习，进一步熟悉分析型大数据系统的基本原理和基本操作。

通过本课程的学习从以下方面促进了毕业要求的实现：熟练掌握分析型大数据系统的核心理论及技术，并能运用这些理论、知识、原理和方法解决社会经济运行各个领域中出现的实际问题；掌握分析型大数据系统下搜集、整理、分析和呈现数据的基本技能、工具和方法，具有较宽的知识面和知识结构；具有自主学习和终身学习意识，有创新创业能力及不断学习与适应社会发展与改革创新的能力。

课堂授课内容和实践环节如何紧密贴合，让学生真正理解和掌握分析型大数据系统的基本思想，并将之运用于适合的应用场景。

三、各教学环节学时分配

以表格方式表现各章节的学时分配，表格如下：

**教学课时分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 章节内容 | 讲课 | 实验 | 其他 | 合计 |
| 第1章  绪论 | 数据库概述  关系型数据库理论简述  分布式系统理论  存储算法简述 | 2 | 0 |  | 2 |
| 第2章NoSQL和事务处理 | 事务处理  CAP定理  服务器端一致性  非关系型数据库简述 | 4 | 2 |  | 6 |
| 第3章  列式数据库 | 列式数据库的历史  技术原理  多用户和硬件  执行一个ALTER语句  数据仓库和列式数据库 | 4 | 2 |  | 6 |
| 第4章  图数据库与文本数据库 | 图论基础  RDBMS与图数据库  凯文贝肯问题的六度  经典文档管理系统  文本挖掘与理解 | 4 | 2 |  | 6 |
| 第5章  大数据和云计算 | 对大数据和云计算的疑问  大数据和数据挖掘 | 4 | 2 |  | 6 |
| 第6章  分析型数据库 | 数据立方体  Codd博士的OLAP规则  Codd博士理论的基础特性  MOLAP  ROLAP  HOLAP | 4 | 2 |  | 6 |
| 第7章  MongoDB基础 | MongoDB简介  数据模型  使用数据 | 3 | 1 |  | 4 |
| 第8章  Python和MongoDB | 在Python中使用文档  使用Python模块  连接和断开  插入、搜索、修改、删除数据  在两个文档之间创建链接 | 4 | 2 |  | 6 |
| 第9章  数据库管理 | 使用管理工具  备份MongoDB服务器  备份大数据库  将数据导入MongoDB  从MongoDB导出数据 | 4 | 2 |  | 6 |
| 合计 |  | 33 | 15 |  | 48 |

四、教学内容

以“章节”为单位说明本章节的教学内容、教学重点、难点、课程的考核要求和复习思考题等，各章节格式如下：

第1章 绪论

第一节 数据库概述

1. 理解数据库的概念与特点；
2. 了解数据库的发展历程。

第二节 关系型数据库理论简述

1. 理解关系型数据库的概念与特点；
2. 了解关系型数据库在各个行业领域的应用；

第三节 分布式系统理论

1. 理解分布式系统理论；
2. 了解分布式存储的问题；

3. 分布式存储算法简述。

教学重点、难点：重点理解数据库的概念与特点以及分布式系统理论，难点在于关系型数据库与非关系型数据库的区别和特点。

课程的考核要求：掌握数据库的概念和特点以及理解分布式系统理论。

复习思考题：

1. 谈谈对数据库的理解，以及对业界竞争关系的分析和未来发展方向的判断。
2. 请针对你所了解的数据库系统，讨论有哪些属于分布式系统。

第2章 NoSQL和事务处理

第一节 事务处理

1. 掌握批处理中的数据库事务处理；
2. 掌握磁盘处理中的事务处理；

第二节 CAP定理和服务器端一致性

1. 了解CAP定理；
2. 理解服务器端一致性；

第三节 非关系型数据库简述

1. 为什么这些场景下SQL不能发挥作用；
2. 理解非关系型数据库的概念和特点。

教学重点、难点：重点掌握非关系型数据库的概念和特点，难点在于理解关系型数据库和非关系型数据库的适用场景。结合非关系数据库技术特点，通过数据库学习与项目开发实践，挖掘思政元素，培养学生项目开发的全局观和系统思维，拓展学习内容的广度，融入数据安全、工程伦理和团队精神等思政元素。

课程的考核要求：能熟练掌握非关系型数据库的概念和特点，理解两种类型数据库的适用场景。

复习思考题：

1. 谈谈对非关系型数据库的理解。
2. 讨论关系型数据库和非关系型数据库有哪些适用场景。

第3章 列式数据库

第一节 列式数据库的历史、技术原理

1. 了解列式数据库的历史及发展历程；
2. 理解列式数据库的技术原理；

第二节 多用户和硬件

1. 了解多用户和硬件；
2. 执行一个ALTER语句；

第三节 数据仓库和列式数据库

1. 了解数据仓库的概念和特点；
2. 掌握列式数据库的概念和特点；

教学重点、难点：重点掌握列式数据库的概念和特点，难点在于理解数据仓库和列式数据库。

课程的考核要求：能熟练掌握列式数据库的概念和特点。

复习思考题：

1. 谈谈对列式数据库和数据仓库的理解。
2. 执行ALTER语句。

第4章 图数据库与文本数据库

第一节 图论基础

第二节 RDBMS与图数据库

1. 了解RDBMS基本原理和思想；
2. 掌握图数据库的概念与特点；

第三节 凯文贝肯问题的六度

第四节 经典文档管理系统

1. 了解经典文档管理系统；
2. 理解文本挖掘；

教学重点、难点：重点掌握图数据库的概念与特点，难点在于理解凯文贝肯问题的六度。了解我国数据库特别是图数据库与文本数据库的发展史、主要流派及代表人物，理解并培养学生在进行设计及构建时所需的工匠精神、团队精神和职业操守、潜移默化进行价值引领，实现课程育人目标。

课程的考核要求：熟练掌握图数据库的概念与特点，并深刻理解凯文贝肯问题的六度。

复习思考题：

1. 谈谈对图数据库的理解。
2. 讨论凯文贝肯问题的六度。

第5章 大数据和云计算

第一节 对大数据和云计算的疑问

第二节 大数据和数据挖掘

1. 理解大数据的概念和基本特点；
2. 理解数据挖掘的基本概念和原理；

教学重点、难点：重点掌握大数据和数据挖掘的基本概念和原理，难点在于使学生能应用大数据的思想解决适用问题。结合典型事迹，在讲解存储过程在数据库后台工作的原理时，以王坚院士十年如一日开发“阿里云”，实现了我国数据库云平台从0到1突破为例，向学生讲解默默无闻的科技工作者的自主创新精神，培养学生家国情怀和民族自豪感。

课程的考核要求：能熟练掌握大数据和数据挖掘的基本概念和原理。

复习思考题：

1. 谈谈对大数据的理解，列举生活中的大数据实例
2. 讨论常用的数据挖掘方法，以及大数据背景下的数据挖掘方法。

第6章 分析型数据库

第一节 数据立方体

第二节 Codd博士的OLAP规则

1. 理解Codd博士的OLAP规则的特点；
2. 掌握Codd博士的OLAP规则的基本框架；

第三节 Codd博士理论的基础特性

1. 了解Codd博士理论；
2. 掌握Codd博士理论的基础特性；

第四节 MOLAP、ROLAP、HOLAP

1. 了解MOLAP、ROLAP、HOLAP的概念；
2. 掌握MOLAP、ROLAP、HOLAP的原理和特点；

教学重点、难点：重点掌握MOLAP、ROLAP、HOLAP的原理和特点，难点在于使学生能根据不同应用场景选择合适的方法来解决实际问题。通过在抗疫过程中，人工智能和软件应用案例，让学生切身感受到我国抗疫的制度优势的同时，着力启迪学生思考其后的数据库设计、存储和工作原理。

课程的考核要求：能熟练掌握MOLAP、ROLAP、HOLAP的原理和特点；能根据不同的应用场景选择合适的方法解决遇到的实际问题。

复习思考题：

1. 谈谈对Codd博士理论的理解。
2. 讨论MOLAP、ROLAP、HOLAP各自的适用场景。

第7章 MongoDB基础

第一节 MongoDB简介

1.了解MongoDB；

2.掌握MongoDB的安装；

第二节 数据模型

1. 了解如何设计数据库；
2. 掌握如何构建索引以及使用地理空间索引；
3. 掌握如何在实际问题中使用MongoDB；

第三节 使用数据

1. 掌握数据的插入、查询、更新、删除等操作；

2. 掌握如何引用数据库以及使用索引相关的函数；

教学重点、难点：重点掌握数据的插入、查询、更新、删除等操作，难点在于理解数据模型。

课程的考核要求：熟练掌握数据的插入、查询、更新、删除等操作；难点在于理解数据模型。

复习思考题：

1. 请给出MongoDB安装的一般流程。
2. 请在library数据库中实现数据的插入、查询、更新、删除等操作。
3. 请在library数据库中创建索引以及使用命令浏览数据库。

第8章 Python和MongoDB

第一节 在Python中使用文档；

第二节 使用Python模块；

第三节 数据操作；

第四节 在两个文档间创建链接；

教学重点、难点：重点掌握数据的插入、查询、更新、删除等操作，难点在于理解MongoDB与Python模块。关注当前技术热点和舆论环境，介绍国产数据库发展现状，使学生理解数据库作为我国信创产业等底层技术之一，解决技术卡脖子问题等紧迫性。在讲解数据库权限时，以ebay、万豪等21世纪著名的数据泄露事件，讲解数据库应用和开发中涉及数据安全和隐私等道德伦理的内容。

课程的考核要求：熟练掌握数据的插入、查询、更新、删除等操作；难点在于理解MongoDB与Python模块。

复习思考题：

1. 请给出使用Python模块的一般过程。

2. 请在library数据库中实现数据的插入、查询、更新、删除等操作。

3. 请在两个文档间创建链接。

第9章 数据库管理

第一节 使用管理工具

1. 了解MongoDB控制台；

2. 学会使用第三方管理工具；

第二节 备份MongoDB服务器及大数据库

1.掌握备份单个数据库、集合的方法；

2.掌握使用日志文件系统创建快照；

第三节 数据导入、导出

1.将数据导入MongDB；

2.从MongoDB导出数据；

教学重点、难点：重点掌握数据库的备份、数据导入导出等操作，难点在于运用数据库管理工具。

课程的考核要求：熟练掌握数据库的备份、数据导入导出等操作。

复习思考题：

1. 请备份library数据库。

2. 请在library数据库中导入数据，并将数据导出。

五、考核方式、成绩评定

本课程将采用设计和大作业的方式进行考核；成绩评定分为两部分：平时上机作业占比40%，期末大作业占比60%。

六、主要参考书及其他内容

列出主要参考书目，所列条目及其顺序如下：

[1] Joe Celko. NoSQL指南. 北京：人民邮电出版社. 2016.7

[2] David Hows,Eelco Plugge等. MongoDB大数据处理权威指南(第2版). 北京：清华大学出版社. 2015.1

执笔人：高迎

教研室主任： 范烺

系教学主任审核签名：