《数据库》教学大纲

一、课程基本信息

课程名称:	数据库	英文名称:	Database
课程类型：	本科	学分/学时:	4/64
主要面向专业：	计算机科学	授课语言：	中英文
先修课程：	算法与数据结构，操作系统 I		
开课单位:	信息科学与技术学院	课程代码:	CS150A

二、课程简介

本课程涵盖数据库设计和数据库在应用程序中的使用，并介绍关系数据库引擎的内部以及基本的数据挖掘方法。本课程的主要内容包括以下主题：

- 关系数据模型，关系代数和 SQL。
- 查询优化和事务处理简介。
- 基于依赖关系和规范形式的数据库设计和关系设计原则。
- 数据库服务，包括保护，完整性控制和备用数据视图。
- 数据挖掘，机器学习和数据库技术之间的集成。
- 现代数据库系统，包括用于大规模数据处理的并行数据库，NoSQL，Hadoop 和 Spark。

三、课程教学目标

- 知识认知能力：能掌握数据库相关的基本知识，包括关系数据模型、SQL 和关系代数的数学模型及概念，关系型数据库的层次结构和功能，查询优化和事务处理的基本概念、逻辑及并行化和分布式处理方法、数据库设计的基本方法和原则，数据挖掘和数据库技术之间的联系和集成，以及现代数据库系统（NoSQL，Hadoop 和 Spark）的原理和构架。
- 综合素质能力：能理解数据库管理系统方面的工程职业道德和规范，具备科学精神和工程师的基本素养，具备科技报国的家国情怀和使命担当；能进行团队协作，具备合作精神和人际沟通能力。使学生全面深入了解和掌握数据库和数据库管理系统的基本概念和

分析方法，为以后学习数据库和数据科学等相关专业课以及从事与之有关的工程项目和科学研究打下坚实的理论基础。

四、课程教学方法

- 课堂讲授与讨论：数据库课程知识点基本以课堂讲授为主，在讲解基本知识点的基础上，关注课程重点难点内容的讲授，采用启发式教学方法，引导学生对问题展开思考和讨论，使学生从数学概念、物理概念及工程概念出发分析和解决数据库领域的相关问题。
- 线下及线上的讨论和答疑：课程答疑会在线下和线上（piazza）同时进行，为课程讲授和讨论提供补充，并及时解答学生们个性化的问题。
- 课程实践：数据库课程的实践教学环节以课程 project 的方式完成。通过数据库和数据挖掘相关的工程实际案例的分析和解决，使学生在掌握课程基本理论和方法的同时，理解课程知识在工程中的实际应用，激发学生的研究兴趣，启迪学生创新思维。

五、课程教学内容与安排

以教学周方式安排教学内容

教学周	章节名称	主要教学内容 （主要知识点）	学时安排	教学方法 （仅列名称）
1	数据库介绍 SQL I	• 课程介绍 • 关系型数据库管理系统 • 数据定义语言 • 数据操作语言	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
2	SQL II 硬盘，缓冲区和文件 I	• SQL 语法 • DBMS 构架 • 磁盘与闪存 • 磁盘空间管理器	4	线下：授课+阅读 RDBMS+讨论课+小测试

				线上：piazza + blackboard + gradescope
3	硬盘，缓冲区和文件 II 文件组织	• 数据库文件结构 • 页面布局与记录布局 • 成本模型和分析	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
4	**国庆节** 索引和 B+树	• 索引格式 • 静态高扇出搜索树—ISAM • 动态高扇出搜索树—B+树 • B+树中的插入、删除与批量载入	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
5	缓冲区管理 关系代数	• 脏页处理 • 页面替换准则—LRU 和 MRU • 关系代数运算子	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
6	排序和哈希 迭代和联接 I	• 双缓冲与分而治之 • 外部排序与哈希 • 并行化排序与哈希 • 联接运算	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope

		▪ 块嵌套循环联接 ▪ 索引嵌套循环联接 ▪ 排序-合并联接		
7	迭代和联接 II 查询优化 I	• 联接运算 ▪ Grace 哈希联接 • 计划空间 • 关系代数等价 • Materializing	4	线下：授课+阅读RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
8	查询优化 II **期中考试**	• 成本和搜索 • System R优化器 • 动态规划	4	线下：授课+阅读RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
9	事务和并发 I 事务和并发 II	• 事务与事务管理器 • 事务的属性：ACID • 事务调度--串行化与冲突串行化	4	线下：授课+阅读RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
10	恢复 I 恢复 II	• 两阶段锁定（2PL)与严格两阶段锁定 • 锁管理器（LM）	4	线下：授课+阅读RDBMS+讨论课+小测试

		▪ 死锁与死锁检测 ▪ 多重锁定粒度 • Steal/No-Force 准则 • 预写日志记录协议（WAL） • 崩溃恢复—分析、重做与撤销		线上：piazza + blackboard + gradescope
11	数据库设计 I 数据库设计 II	• 数据独立 • 实体关系模型 ▪ 实体与关系 ▪ 关键约束与参与约束 • 函数依赖（FD）和归一化	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
12	并行查询处理 I 并行查询处理 II	• 并行化数据库操作 ▪ 操作间并行化 ▪ 操作内并行化 ▪ 查询内并行化	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
13	分布式事务 I	• 分布与并行	4	

	分布式事务 II	• 分布式数据库 • 分布式锁定 • 分布式死锁检测 • 两阶段提交（2PC）		线下：授课+阅读RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
14	数据挖掘和机器学习 I 数据挖掘和机器学习 II	• 在线事务处理（OLTP）与在线分析处理(OLAP) • 数据仓库、数据湖与数据沼泽 ▪ 多维数据模型 • 有监督学习 ▪ 线性回归与分类 • 无监督学习 ▪ 聚类与降维	4	线下：授课+阅读RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope
15	NoSQL I NoSQL II	▪ 复制数据 ▪ 放松事务约束 ▪ 松散一致性与NoSQL ▪ 无线性化与无串行化 ▪ 弱隔离性 ▪ 快照隔离	4	线下：授课+阅读RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope

16	MapReduce 和 Spark **课程复习**	• MapReduce ▪ 用户定义 map 和 reduce 函数 • Spark ▪ 预定义的关系运算符	4	线下：授课+阅读 RDBMS+讨论课+小测试 线上：piazza + blackboard + gradescope

六、考核方式和成绩评定方法

- 作业 (20%) + 小测试 (10%)
- 课程project (25%)
- 期中考试 (20%)
- 期末考试 (25%)

七、教材和参考书目

（一）推荐教材

- 数据库系统概念(第 6 版)，Abraham Silberschatz，Henry Korth 和 S. Sudarshan (S/K/S)，机械工业出版社，ISBN：9787111375296
- Database Management Systems (3rd edition) by Raghu Ramakrishnan and Johannes Gehrke (R/G), ISBN: 9780072465631

（二）参考书目

- Database Systems: The Complete Book (2nd edition) by Garcia-Molina, Jeffrey Ullman and Jennifer Widom (G/U/W), ISBN: 9780131873254
- Fundamentals of Database Systems (7th edition) by Ramez Elmasri and Shamkant Navathe (E/N), ISBN: 9780133970777

八、学术诚信教育

本课程高度重视学术诚信，严禁抄袭、作弊等行为。

“在学习、科研、实习实践等活动中，学生应恪守学术道德，坚守学术诚信，保护知识产权，坚持勇于创新、求真务实的科学精神，努力培养自己严谨求实、诚实自律、真诚协作的科学态度，成为良好学术风气的维护者、严谨治学的力行者、优良学术道德的传承者。”

九、其他说明（可选）

评分注意事项：

- 由于考试是我们掌握成绩的主要指标，因此我们保留根据考试成绩调整最终字母等级的权利。 特别是，那些未在考试中达到及格分数的学生将不会获得课程的及格分数。
- 提交的作业必须是自己的（或者对于两个人的项目，是团队的）。 我们将使用标准的软件对提交的编程作业进行查重。
- 我们将遵循SIST关于学术诚信的政策，因此请确保熟悉它。

Database Syllabus

1. Basic Course Information

Course Name:	Database		
Course Level:	Undergraduate	Credit/Contact Hour:	4 / 64
Major:	CS	Teaching Language:	Chinese / English
Prerequisite:	Algorithms and Data Structures, Operating Systems I		
School/Institute:	SIST	Course Code:	CS150A

2. Course Introduction

This course covers database design and the use of databases in applications, with an introduction to the internals of relational database engines as well as basic data mining methods. The main content of this course includes the following topics:

- Relational data model, relational algebra, and SQL.
- Query optimization and introduction to transaction processing.
- Database design and relational design principles based on dependencies and normal forms.
- Database services including protection, integrity control, and alternative views of data.
- The integration between data mining, machine learning and database technology.
- Modern database systems, including Parallel Databases, NoSQL, Hadoop and Spark for large-scale data processing.

3. Learning Goal

- Cognitive competence: Ability to master basic knowledge related to databases, including relational data models, SQL and relational algebra, hierarchy of RDBMS, basic concepts of query optimization and transaction processing, the principles of database design, the integration between data mining and database technology, and the principles and architecture of modern database systems (NoSQL, Hadoop and Spark).

- Comprehensive qualities: have the scientific spirit and basic qualities as an engineer for DBMS, have the national sentiment and mission of serving the country with science and technology; be able to work in a team, have a spirit of cooperation and communication skills. It enables students to fully understand and master the basic concepts and analysis methods of databases and DBMS, and lay a theoretical foundation for the future study of related professional courses related to database and data science, as well as engineering projects and scientific research.

4. Instructional Pedagogy

- Teaching and discussion: The knowledge points of this courses are basically introduced via on-site teaching. We pay attention to the teaching of key and difficult points of the course, and use heuristic teaching methods to guide students to think and discuss the problems, and encourage students to analyze and solve related problems in the field of database in terms of mathematics, physics and engineering.
- Offline/online discussion and Q&A: Course Q&A will be conducted both offline and online (piazza) to provide supplements for course teaching and discussion, and to answer students' personalized questions in a timely manner.
- Curriculum practice: The practical teaching of this course is conducted based on the course project. Through analyzing and solving the practical project on database and data mining, students can understand the course knowledge from the viewpoint of engineering. It stimulates the research interest and innovative thinking of students.

5. Course Content and Schedule

Course Structure by Week

Week	Chapter	Teaching Contents	Contact Hours	Teaching Modes
1	Introduction to DBMS SQL I	• Course Introduction • Relational DBMS	4	• Offline: teaching,

		• Data definition language • Data manipulation language		discussion, quiz • Online: piazza, Blackboard , Gradescope
2	SQL II Disk, Buffers, Files I	• SQL syntax • DBMS architecture • Disk and flash memory • Disk Space Manager	4	• Offline: teaching, discussion, quiz • Online: piazza, Blackboard , Gradescope
3	Disk, Buffers, Files II File Organization	• Database file structure • Page layout and record layout • Cost model and analysis	4	• Offline: teaching, discussion, quiz • Online: piazza, Blackboard , Gradescope
4	Indexes and B+ Trees Buffer Management	• Index format • Static high fan-out search tree—ISAM • Dynamic high fan-out search tree—B+ tree	4	• Offline: teaching, discussion, quiz • Online: piazza,

		• Insert, delete and batch load in B+ tree		Blackboard , Gradescope
5	**National Day** Relational Algebra	• Dirty page • Page Replacement Guidelines—LRU and MRU • Relational algebra operators	4	• Offline: teaching, discussion, quiz • Online: piazza, Blackboard , Gradescope

6	Sorting and Hashing
	Iterations and Joins I
		• Double buffering and divide and conquer
		• External sorting and hashing
		• Parallel sorting and hashing
		• Join operation
		  • Block nested loop join
		  • Index nested loop join
		  • Sort-merge join
			4	• Offline: teaching, discussion, quiz
				• Online: piazza, Blackboard , Gradescope

7	Iterations and Joins II Query Optimization I	• Join operation • Grace hash join • Planning space	4	• Offline: teaching, discussion, quiz

		• Relational algebra equivalence • Materializing		• Online: piazza, Blackboard , Gradescope
8	Query Optimization II **Midterm**	• Cost and search • System R optimizer • Dynamic programming	4	• Offline: teaching, discussion, quiz • Online: piazza, Blackboard , Gradescope
9	Transactions and Concurrency I Transactions and Concurrency II	• Transaction and Transaction Manager • ACID • Transaction scheduling- serialization and conflict serialization	4	• Offline: teaching, discussion, quiz • Online: piazza, Blackboard , Gradescope

10	Recovery I
	
	Recovery II
		• 2PL and strict 2PL
		• Lock Manager
		  • Deadlock and deadlock detection
		  • Multiple locking granularity
			4	• Offline: teaching, discussion, quiz
				• Online: piazza, Blackboard , Gradescope

		• Steal/No-Force criterion • Write Ahead Logging Protocol (WAL) • Crash recovery--analysis, redo and undo		

11	Database Design I
	Database Design II
		• Data independence
		• Entity relationship model
		  • Entity and relationship
		  • Key constraints and participation constraints
		• Functional dependence (FD) and normalization
			4	• Offline: teaching, discussion, quiz
				• Online: piazza, Blackboard , Gradescope

12	Parallel Query Processing I
	Parallel Query Processing II
		• Parallelize database operations
		  • Parallelization between operations
		  • In-operation parallelization
			4	• Offline: teaching, discussion, quiz
				• Online: piazza, Blackboard , Gradescope

		• In-query parallelization		
13	Distributed Transactions I Distributed Transactions II	• Distribution and parallelism • Distributed database • Distributed locking • Distributed deadlock detection • Two-phase commit (2PC)	4	• Offline: teaching, discussion, quiz • Online: piazza, Blackboard , Gradescope

14	Data Mining & Machine Learning I
	Data Mining & Machine Learning II
		• OLTP and OLAP
		• Data warehouse, data lake and data swamp
		• Multidimensional data model
		• Supervised learning
		  • Linear regression and classification
		• Unsupervised learning
		  • Clustering and dimensionality reduction
			4	• Offline: teaching, discussion, quiz
				• Online: piazza, Blackboard , Gradescope

15	NoSQL I NoSQL II	• Copy data • Relax transaction constraints	4	• Offline: teaching, discussion, quiz

		• Loose consistency and NoSQL • Wireless and no serialization • Weak isolation • Snapshot isolation		• Online: piazza, Blackboard , Gradescope

16	MapReduce and Spark
	Course Review
		• MapReduce
		  • User-defined map and reduce functions
		• Spark
		  • Predefined relational operators
			4	• Offline: teaching, discussion, quiz
				• Online: piazza, Blackboard , Gradescope

6.Grading Policy

- Problem Sets (20%) + Quizzes (10%)
- Course project (25%)
- Midterm (20%)
- Final exam (25%)

7. Textbook & Recommended Reading

(1) Textbook

- Database Management Systems (3rd edition) by Raghu Ramakrishnan and Johannes Gehrke (R/G), ISBN: 9780072465631

- 数据库系统概念(第 6 版)，Abraham Silberschatz，Henry Korth 和 S. Sudarshan (S/K/S)，机械工业出版社，ISBN：9787111375296

(2) Recommended Reading

- Database Systems: The Complete Book (2nd edition) by Garcia-Molina, Jeffrey Ullman and Jennifer Widom (G/U/W), ISBN: 9780131873254
- Fundamentals of Database Systems (7th edition) by Ramez Elmasri and Shamkant Navathe (E/N), ISBN: 9780133970777

8. Academic Integrity

This course highly values academic integrity. Behaviors such as plagiarism and cheating are strictly prohibited. Please list more if you have more specific requirements.

9. Other Information (Optional)

Notes on grading:

- Since exams are the main indicators we have of individual grasp of the material, we reserve the right to adjust final letter grades based on exam performance. In particular, students who do not achieve a passing average on the exams will not receive a passing grade in the class.
- Work that you submit must be your own (or for two-person projects, the teams). We will run the standard software duplication checkers on submitted assignments.
- We will be following the SIST policy on Academic Honesty, so be sure you are familiar with it.