

Database System Principles

《数据库系统原理》

关佶红/Jihong Guan

教授/Professor

Email: jhguan@tongji.edu.cn

张毅超/Yichao Zhang

副教授/Associate Professor

Email: yichaozhang@tongji.edu.cn

李文根/Wengen Li

副教授/Associate Professor

Email: lwengen@tongji.edu.cn

先进数据与机器智能系统实验室 (ADMIS)

<https://admis.tongji.edu.cn/main.htm>



同濟大學
TONGJI UNIVERSITY

电子与信息工程学院 计算机科学与技术系

College of Electronics and Information Engineering Department of Computer Science and Technology

- **课程学分：** 3
- **理论课时：** 51 (理论与课堂研讨)
- **实验课时：** 16 (8次实验课)
- **实践课时：** 2024年秋课程设计
- **课程教师：** 李文根、关佶红、张毅超
- **助教：**
 - 彭 瀚：18010699229, penghan@tongji.edu.cn
 - 金 唱：17721293653, jinchang1223@tongji.edu.cn
 - 汪宇晗：18365131329, 2331925@tongji.edu.cn
 - 宋子涵：13386457589, 2331902@tongji.edu.cn
 - 蒋旭东：18234867925, xdjiang@tongji.edu.cn

李文根

课程时间: Mon.13:30~15:05(G207), Wed.10:00~11:35 (单周, G207)

答疑时间: Tue. 13:30-16:30, Thu.13:30-16:30 (智信馆410)

Tel: 15000621786; Email: lwengen@tongji.edu.cn

Online: QQ, Email

研究方向: 时空数据智能管理与分析, 多模态智能, 知识图谱

[Homepage](#)

Prof. Guan Jihong (关佶红)

ADMIS Lab: Room 429B, Zhixin Building, Research: Data Management, Data Mining, Big Data, Machine Learning, AI, Bioinformatics, et al

Lecture hours: Mon.10:00~11:35, Wed.10:00~11:35(Odd)

Office: Room 429B/458, Zhixin Building, Jiading campus

Office Hour: Mon.12:30-16:00, Wed.13:00-15:00

Online: Tencent Meeting, ML/AI/BD/Bioinformatics

Tel: 186-1610-2875; **Email:** jhguan@tongji.edu.cn

Homepage: <http://admis.tongji.edu.cn/>

Associate Prof. Yichao Zhang (张毅超)

Lecture Hours: Mon.13:30~15:05, Wed.10:00~11:35 (单周)

Office Hours: Tue. 13:30-16:30, Thu.9:30-11:30 (智信馆458)

Tel: 15021825768; **Email:** yichaozhang@tongji.edu.cn

Online: Email, Wechat

Research interests: Graph neural networks, link prediction, weighted network modeling, random diffusion, network games, financial data analysis, and urban disaster prevention.

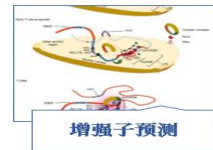
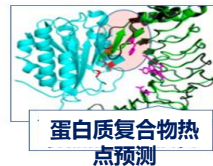
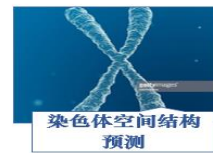
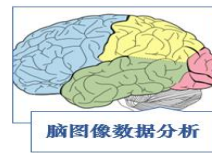
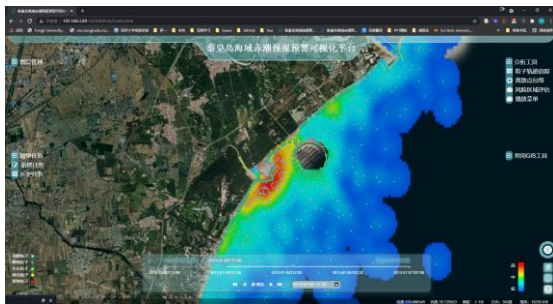
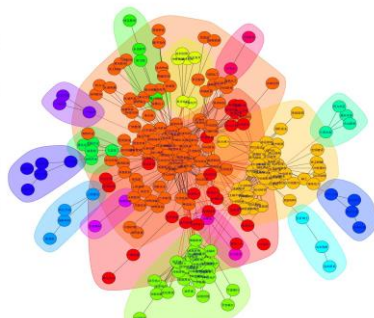
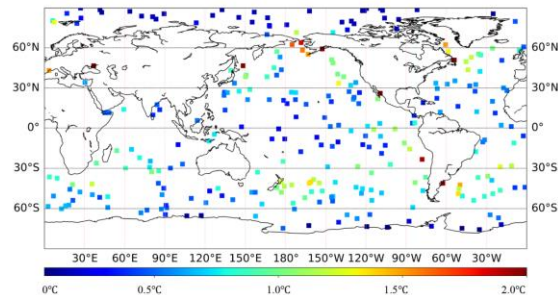
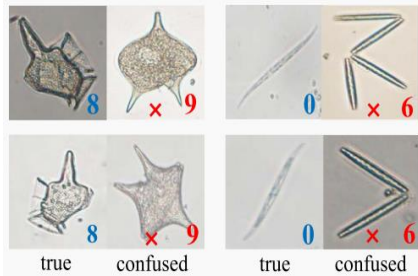
[Homepage](#)

► ADMIS课题组研究方向



同济大学
TONGJI UNIVERSITY

- **Spatio-temporal data analytics**: urban computing, ocean computing
- **Multimodal intelligence**: sentiment analysis, knowledge graph
- **Bioinformatics**: protein data analytics, cell data analytics, drug discovery



- **什么是数据?**

- 在现实生活, 数据是可识别的抽象符号, 即描述事物的符号记录
- 在计算机中, 数据是所有能被计算机处理的符号的总称
- 在数据库中, 数据是数据库中存储的基本对象

- **数据的种类:**

- 数字, 例如 1、2、3
- 文本, 例如张三、李四
- 时间, 例如2022年2月22日
- 向量, 例如图片或文本的表征
- 图片, 例如logo
- 音频, 例如电视语音
- ...

- **结构化数据**：关系数据（表）
- **半结构化数据**
 - 键值对Key-Value
 - XML、JSON
 - 图
 - 向量
- **非结构化数据**
 - 文本文档、电子邮件、图像、音频、视频

► 关系数据 (Relational Data)



Sno (学号)	Sname (姓名)	Sgender (性别)	Sage (年龄)	Sdept (系别)
2021310721	李博	男	17	CS
2021310722	赵宇	男	19	CS
2021310723	张敏	女	18	CS
2021310724	王勇	男	18	MA
2021310725	刘佳	女	17	MA

Sno (学号)	Cno (课程号)	Grade (成绩)
2021310721	5	98
2021310722	1	87
2021310723	1	92
2021310723	5	76
2021310724	7	84
2021310725	4	95

Cno (课程号)	Cname (课程名)	Cpno (先修课)	Ccredit (学分)
1	数据库	2	4
2	数据结构与算法	6	4
3	操作系统	2	3
4	高等数学		4
5	软件工程	6	2
6	程序设计		3
7	数值分析	4	2

关系模型：二维表格（表）来表示数据

- 表中的每一行代表一个记录（实体或元组）
- 表中的每一列代表一个属性（字段）

查询赵宇数据库课程成绩

- 1) Student表查找赵宇学号
- 2) Course表查找数据库课程号
- 3) SC表查找成绩

全球大数据热度变化



同济大学
TONGJI UNIVERSITY

全球 ▼

2009/1/1 - 2024/2/24 ▼

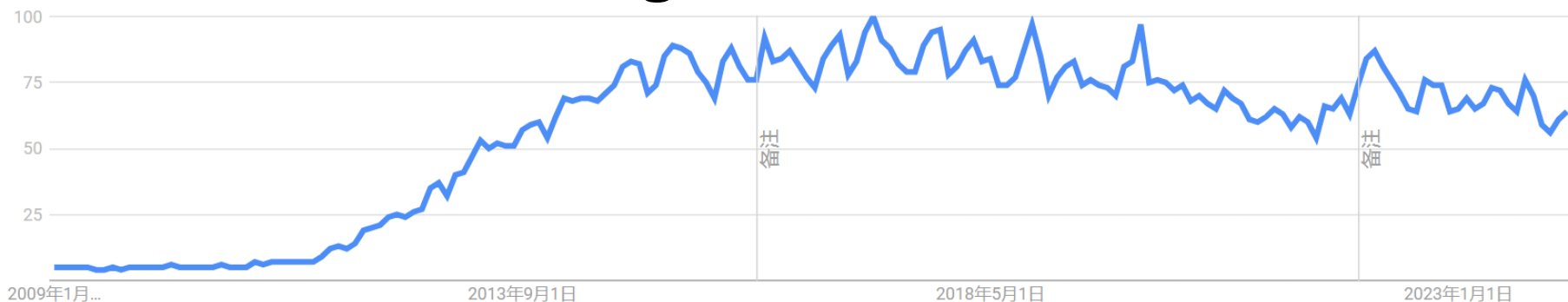
所有类别 ▼

Google 网页搜索 ▼

热度随时间变化的趋势 ?



Big Data



From Google trends (<https://trends.google.com/trends>), 2024/2/24

Big Data: A Joke

某比萨店的电话铃响了，客服人员拿起电话。

客服：XXX比萨店。您好，请问有什么需要我为您服务？

顾客：你好，我想要一份.....

客服：先生，烦请先把您的会员卡号告诉我。

顾客：16846146***。

客服：陈先生，您好！**您是住在泉州路一号12楼120x室**，请问您想要点什么？

顾客：我想要一个海鲜比萨.....

客服：陈先生，海鲜比萨不适合您。

顾客：为什么？

客服：**根据您的医疗记录，您的血压和胆固醇都偏高。**

顾客：那你们有什么可以推荐的？

客服：您可以试试我们的低脂健康比萨。

顾客：你怎么知道我会喜欢吃这种的？

客服：**您上星期一在中央图书馆借了一本《低脂健康食谱》。**

顾客：好。那我要一个家庭特大号比萨，要付多少钱？

客服：**99元，这个足够您一家六口吃了。但您母亲应该少吃，她上个月刚刚做了心脏搭桥手术，还处在恢复期。**

顾客：那可以刷卡吗？

客服：陈先生，对不起。请您付现款，因为**您的信用卡已经刷爆了，您现在还欠银行4807元，而且还不包括房贷利息**

顾客：那我先去附近的提款机提款。

客服：陈先生，**根据您的记录，您已经超过今日提款限额。**

顾客：算了，你们直接把比萨送我家吧，家里有现金。你们多久会送到？

客服：大约30分钟。如果您不想等，可以自己骑车来。

顾客：为什么？

客服：**根据我们全球定位系统的车辆行驶自动跟踪系统记录。您登记有一辆车号为XX-748的摩托车，而目前您正在解放路东段华联商场右侧骑着这辆摩托车。**

顾客：.....

家庭住址

医疗记录

借阅记录

医疗记录

借贷记录

定位跟踪

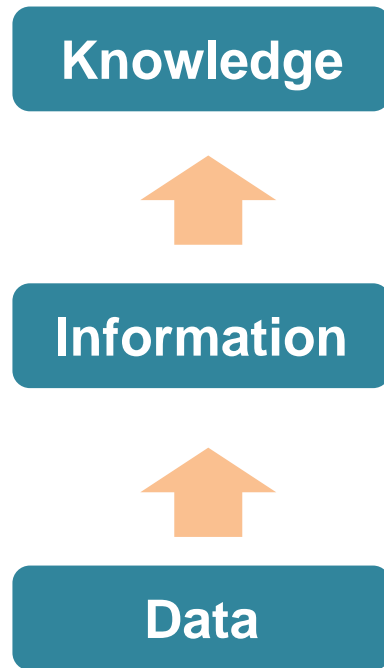
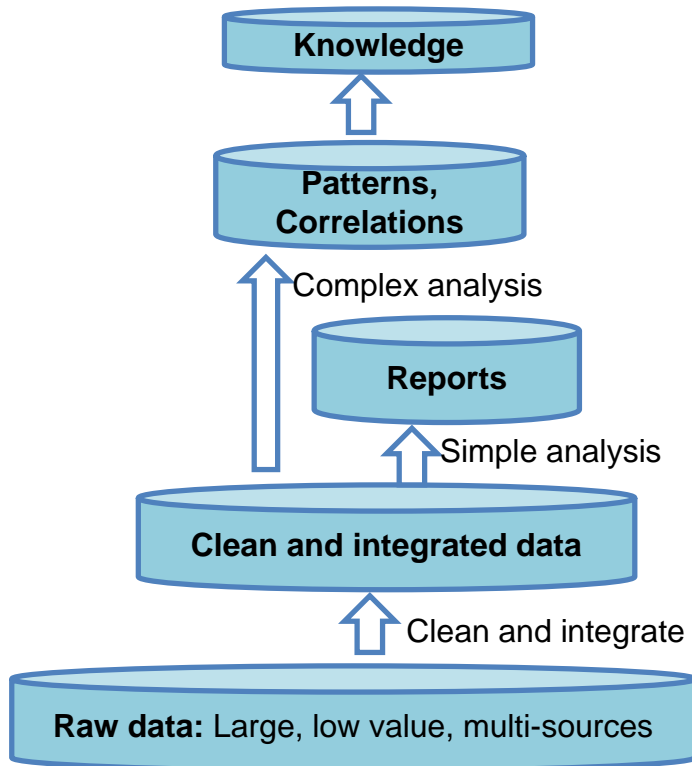
► 多源、异构、多模态大数据

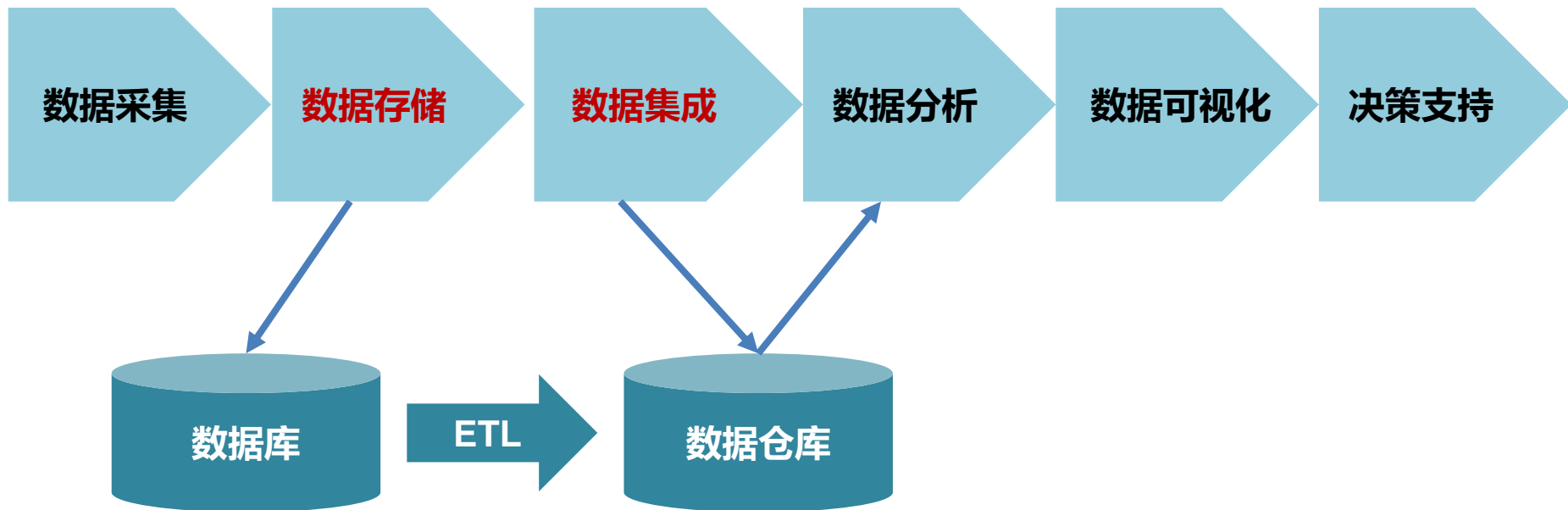


同济大学
TONGJI UNIVERSITY



► Data, Information & Knowledge





ETL: 抽取(Extract)、转换(Transform)、加载(Load)

- **什么是数据库?**
 - 一组相互有关联的数据集合
 - 长期储存在计算机中，有组织、可管理和可共享
- **数据库的基本特征**
 - 数据按一定的数据模型组织、描述和储存
 - 支持数据的增删改查
 - 支持并发查询处理

- **什么是数据库系统?**

- 数据库系统是指由数据库管理系统和相关工具组成的软件系统，用于管理和操作大量数据
- 一般包括
 - 数据库
 - 数据库管理系统
 - 开发工具、应用系统
 - 数据库管理员和终端用户

- **什么是数据库管理系统?**

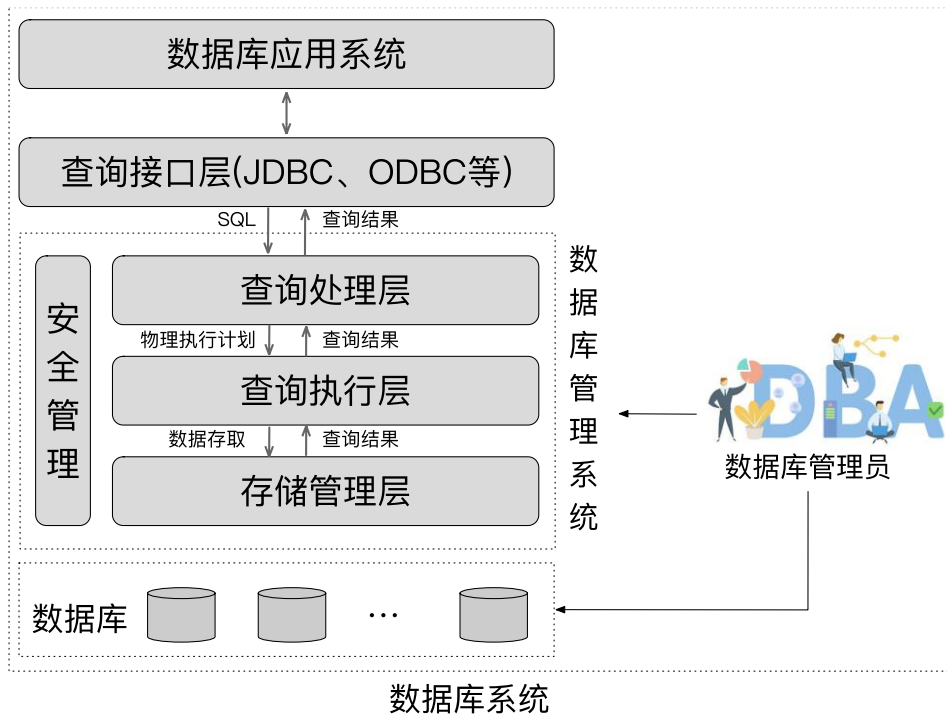
- 管理数据库的软件

什么是数据库管理系统?

- **定义1:** 用户 (应用程序) 与操作系统之间的数据库管理软件
- **定义2:** 一个管理数据的大型复杂基础软件系统

数据库管理系统的用途

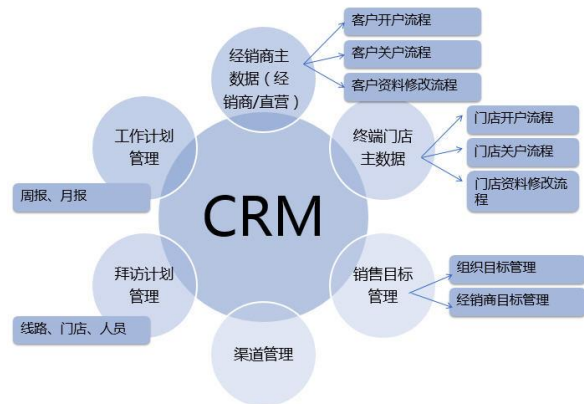
- 优雅查询和数据抽象
- 高效组织和存储数据
- 正确一致的并发更新
- 低时延高吞吐的查询
- 并行高效的有序执行
- 可用性和高可靠保证
- 安全可信的统一控制
- 方便易用的用户接口



► 为什么要学习数据库?



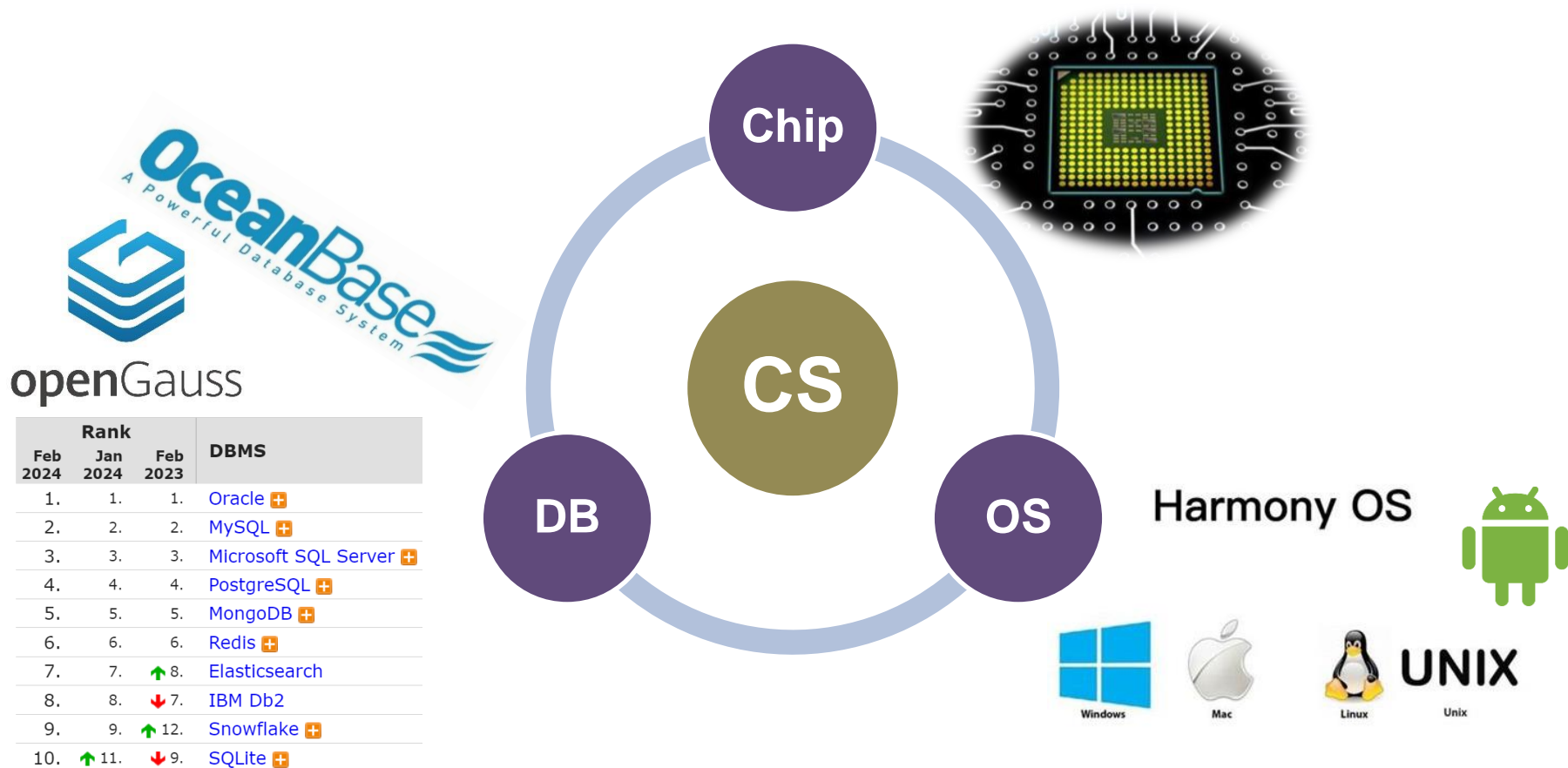
- **Need for DB has exploded in the past years**
 - Customer Relationship Mgmt (CRM), Supply Chain Mgmt, Enterprise Resource Planning (ERP), Business intelligence(BI), etc.
 - Industry 4.0 (工业4.0), Made in China 2025(中国制造2025)



► 为什么要学习数据库?



同济大学
TONGJI UNIVERSITY



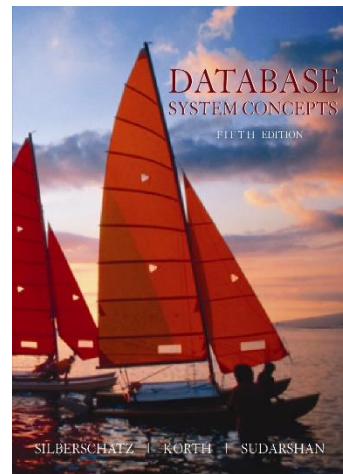
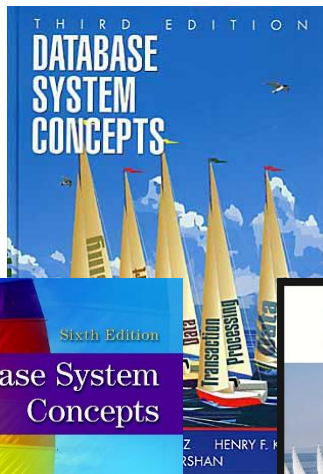
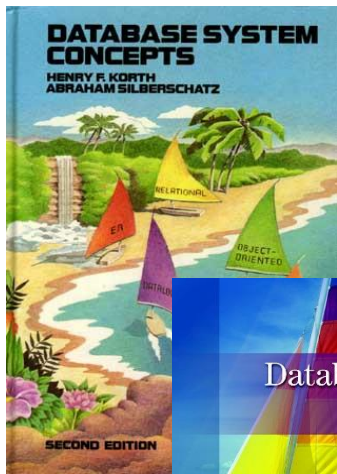
- **目标一**

- 学习掌握数据库设计、存储管理、查询处理与优化、事务管理等基础知识
- 针对特定数据管理需求，具备设计和开发数据库解决方案的能力
- 了解数据库的前沿发展动态和相关先进技术

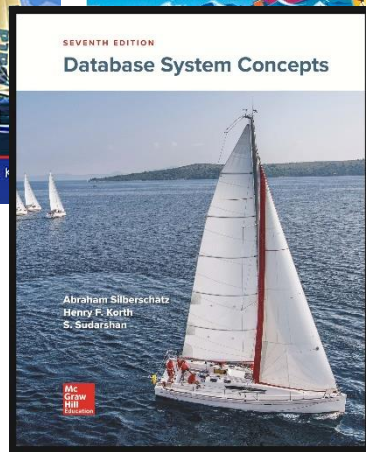
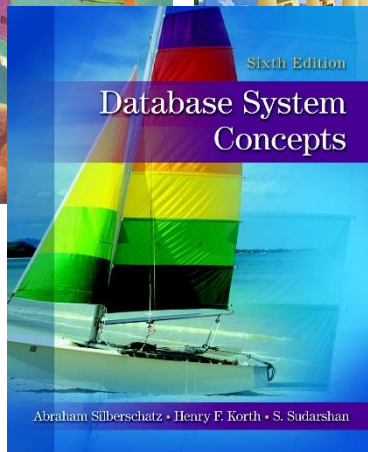
- **目标二**

- 通过小组合作，提高团队协助能力
- 通过课程实验和大作业，提高个人实践能力
- 通过前沿探索，培养科研素养

- Abraham Silberschatz(耶鲁), Henry F. Korth(理海大学), and S. Sudarshan(印度理工学院), **Database System Concepts**



2010



2019

- 中文版（**原书第7版**），北京大学杨冬青等译, 2021年



- **Part 0: Overview**
 - Ch1: Introduction
- **Part 1 Relational Databases**
 - Ch2: Relational model (data model, relational algebra)
 - Ch3&4: SQL(Structured Query Language)
 - Ch5: Advanced SQL
- **Part 2 Database Design**
 - Ch6: Database design based on E-R model
 - Ch7: Relational database design
- **Part 3 Application Design & Development**
 - Ch8: Complex data types
 - Ch9: Application development
- **Part 4 Big data analytics**
 - Ch10: Big data
 - Ch11: Data analytics
- **Part 5 Data Storage & Indexing**
 - Ch12: Physical storage system
 - Ch13: Data storage structure
 - Ch14: Indexing
- **Part 6 Query Processing & Optimization**
 - Ch15: Query processing
 - Ch16: Query optimization
- **Part 7 Transaction Management**
 - Ch17: Transactions
 - Ch18: Concurrency control
 - Ch19: Recovery system
- **Part 8 Parallel & Distributed Database**
 - Ch20: Database system architecture
 - Ch21-23: Parallel & distributed storage, query processing & transaction processing
- **Part 9**
 - DB Platform: **OceanBase**, MongoDB, Neo4J

- **出勤和课堂练习**：10%
- **期末考试**：30%
- **课程大作业**：40%
 - 数据库设计大作业报告
- **其他**：20%
 - 数据库前沿调研报告与分享 (蚂蚁OceanBase, 华为OpenGauss等)
 - 课程实验
 - 课程作业
- **科研、竞赛**：额外加分，上限10%

- **研究对象：蚂蚁科技OceanBase或华为OpenGauss**
 - **存储**：Traditional + Cloud
 - **索引**：Traditional + Spatial
 - **查询**：Traditional + Complex
 - **优化**：Different levels, strategies, techniques
 - **事务**：事务处理、并发、恢复
 - **分布式DB**：分布式存储、查询处理、事务
- **Advanced topics**
 - NoSQL数据库, HTAP数据库, 内存数据库, 云原生数据库, 新硬件数据库
 - BlockChain, AI4DB, DB4AI
 - ...

- **系统开发 (70%)**
 - System (50%): an application system, and a system report
 - Oral presentation (20%): mid term(10%) + final(10%)
- **前沿报告 (20%)**
 - Case studies for 蚂蚁OceanBase, 华为OpenGauss, PostgreSQL, MySQL, NoSQL, etc.
- **考勤 (10%)**
- **科研、竞赛**
 - 额外加分, 上限10%

- **数据库国际学术会议**
 - SIGMOD/PODS, VLDB, ICDE
 - CIKM, ICDT, EDBT, ER, DASFAA, SSTD, etc.
- **数据库相关学术期刊**
 - ACM Trans. on Database Systems (TODS)
 - IEEE Trans. on Knowledge and Data Engineering (TKDE)
 - VLDB Journal
 - Data and Knowledge Engineering (DKE)
- **其他网上资源**
 - DBLP: <http://dblp.uni-trier.de/>
 - Google Scholar, Citeseer, etc.
- **Wechat (微信)关注**
 - 数据分析精选
 - 大数据
 - 大数据文摘
 - 互联网分析沙龙
 - 网络数据大全
 - 数据分析
 - 战略前沿技术
 - 大数据魔方
 -

• 课程QQ群

- 教辅答疑
- 发布通知
- 交流分享



群名称:2024数据库系统原理课程群
群 号:836132046

课程总群: 836132046



群名称:2024-数据库系统原理课程群
群 号:693461114

班级群: 693461114

► Punishment Policy



- **Plagiarism or cheating** in assignments, course projects, and examinations is absolutely unacceptable. Once found, your grade will be definitely set to **Fail**

遵守规定的时间

守时



生活中遵守作息時間，学习中遵守課表時間，日常生活中遵守約定時間。

遵守秩序

守序



集体活动中遵守排队秩序，日常生活中维护公共场所秩序，校外活动中遵守公共秩序。

信守承诺

守信



爱惜自己的信用记录，遵守承诺的事情，做诚信学子。



守纪

遵守纪律

遵守各项校规校纪，特别是日常行为规范和考试纪律。



五守

同济大学学生行为规范



守法

遵守国家法律

在校期间遵守法律规定的各项要求，在走出校园后，恪守法律法规。