

数据库开发技术

南京大学软件学院（2024年度）

Instructor

- 刘嘉
 - Email : liujia@nju.edu.cn
 - Office : 910B

Objectives

- 高性能数据库开发原则
- SQL中的优化
 - 优化SQL提高数据库应用效率
 - 优化索引提高数据库应用效率
 - 优化物理结构提高数据库应用效率
 - 优化数据库设计提高数据库应用效率
- 数据库的其它专题
 - 并发、大数据量、分布式等专题
- 课程实践
- 重点增加数据库的实现部分

第二门数据库课程

Grading

- Practice 40%
 - SQL Exam 20%
 - Implementation of miniDataBase 20%
- Final Exam 60%

SQL Exam

- 关注教学系统，完成注册，完成 Sample
- 共4-8次练习，每次2-4个SQL练习

- 林哲远 zheyuanlin@smail.nju.edu.cn

SQL EXAM

手机号码 注册时使用的手机号

请输入密码

登录

注册

忘记密码?

SQL EXAM

手机号码

验证码

发送验证码

姓名

学号 真实学号

邮件地址 请使用学校邮箱 学号@smail.nju.edu.cn

密码

确认密码

Sign up

Reference URL

- 中国大学MOOC，数据库开发技术（刘嘉）

The screenshot shows the course page for '数据库开发技术' (Database Development Technology) by Liu Jia (刘嘉) from Nanjing University (南京大学). The page includes a header with navigation links, a main banner with the course title and instructor, and a sidebar with course details. The course is currently in its 4th week of 12 weeks. The main content area features a '课程概述' (Course Overview) section, which states that database development is a core technology in cloud computing, big data, and AI, and that the course focuses on performance and practical application. The sidebar also lists the course schedule and a list of 1 instructor.

This screenshot shows the course content page for '数据库开发技术' (Database Development Technology) by Liu Jia (刘嘉) from Nanjing University (南京大学). The page displays a list of course topics under the heading '1. 开发中的数据库问题' (1. Database Development Problems). The topics include: 1.1 黑盒的问题 (1.1 Black Box Problems), 1.2 不同数据库的差异——并发控制 (1.2 Differences between different databases——Concurrency Control), 1.3 对程序员的影响 (1.3 Impact on programmers), and 1.4 为性能而设计 (1.4 Design for performance). The page also includes a sidebar with navigation links and a bottom section for '第一章测验' (Chapter 1 Quiz).

Reference Books

- Database Management Systems [Raghu Ramakrishnan ; Johannes Gehrke]
- Expert Oracle Database Architecture [Thomas Kyte]
- High Performance MySQL [Baron Schwartz , Peter Zaitsev , Vadim Tkachenko]
- Database Design & Relational Theory [C.J.Date]
- SQL CookBook [Anthony Molinaro]
- The Art of SQL [STÉPHANE FAROULT , Peter Robson]
- Database Internals: A Deep Dive into Distributed Data Systems Work[Alex Petrov]
- SQL编程思想：基于5种主流数据库代码实现 董旭阳

未必要选这门数据库开发课程

- 这门课可能性价比没那么高
 - 你们大三太辛苦了，对于实际上对此负责的我来说，深表歉意
 - 学分本身都是等价的，虽然不同的知识对每一个具体的个体来说价值未必相同
- 软件工程的技能重要性在发生变化
 - 如何与GPT相处、协同、做基友，可能是未来十年最重要的技能之一
 - 问出好问题远比答案重要的多，问出好问题的关键是底层的专业能力
- 全球化的学习资源非常丰富，自我学习能力可能是未来最重要的能力

全球化的数据库学习资源

- UCB CS186 : Introduction to Database System (150h , 6Project)
 - SQL能力，用Java实现支持SQL并发查询、B+树Index和故障恢复的关系型数据库
- CMU 15-445 : Database Systems (100h , 大牛Andy Pavlo)
 - 教学资源完全开源，基于教学用数据库bustub，4 Project (内存管理、存储引擎、算子优化器、并发控制)
- Caltech CS 122: Database System Implementation (150h)
- Stanford CS 346: Database System Implementation (150h)
- CMU 15-799: Special Topics in Database Systems (80h)
-

你应该了解的

- 关系代数
 - 选择 (select)、投影 (project)、连接 (join)、联合 (union)、差 (difference)、乘积 (product)
- DB , DBMS , 基于数据库的应用程序
- 数据库的基本特性 (表、KEY、完整性约束、锁、视图、事务...)
- SQL : 基本的DDL , DML , 触发器 , 存储过程等的语法和基本用法
- 数据库设计的基本原则

数据库课程的基本主线

- 数据库行业是一个“夕阳”透着“朝阳”的行业
 - 关系型数据库起步于20 世纪 70 年代
 - DB-Engines Ranking (<https://db-engines.com/en/ranking>)
 - 但是，MongoDB 市值翻了 10 倍，Splunk /Elastic 高速成长
- “数据库” 这个词汇和 20 世纪已经完全不一样了
 - PB 级的数据、成百上千台机器的分布式集群
 - 不同类型的数据存储
 - 不同产品针对不同业务场景的“最佳实践”

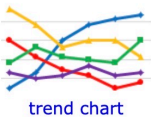
[Ranking](#) > Complete Ranking

[RSS](#) [RSS Feed](#)

DB-Engines Ranking

The DB-Engines Ranking ranks database management systems according to their popularity. The ranking is updated monthly.

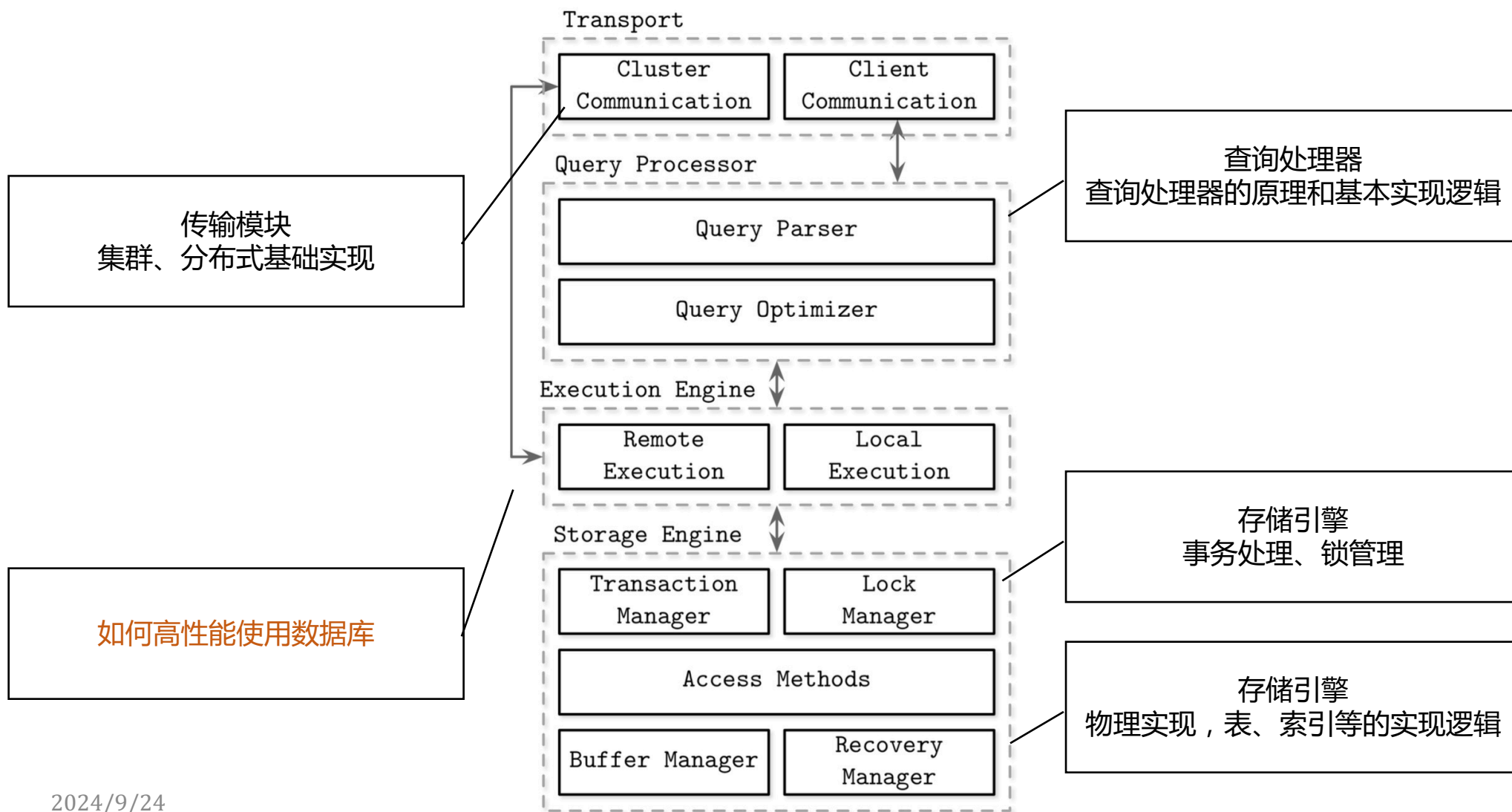
Read more about the [method](#) of calculating the scores.



417 systems in ranking, February 2024

Rank			DBMS	Database Model	Score		
Feb 2024	Jan 2024	Feb 2023			Feb 2024	Jan 2024	Feb 2023
1.	1.	1.	Oracle +	Relational, Multi-model	1241.45	-6.05	-6.08
2.	2.	2.	MySQL +	Relational, Multi-model	1106.67	-16.79	-88.78
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	853.57	-23.03	-75.52
4.	4.	4.	PostgreSQL +	Relational, Multi-model	629.41	-19.55	+12.90
5.	5.	5.	MongoDB +	Document, Multi-model	420.36	+2.88	-32.41
6.	6.	6.	Redis +	Key-value, Multi-model	160.71	+1.33	-13.12
7.	7.	↑ 8.	Elasticsearch	Search engine, Multi-model	135.74	-0.33	-2.86
8.	8.	↓ 7.	IBM Db2	Relational, Multi-model	132.23	-0.18	-10.74
9.	9.	↑ 12.	Snowflake +	Relational	127.45	+1.53	+11.80
10.	↑ 11.	↓ 9.	SQLite +	Relational	117.28	+2.08	-15.38
11.	↓ 10.	↓ 10.	Microsoft Access	Relational	113.17	-4.50	-17.86
12.	12.	↓ 11.	Cassandra +	Wide column, Multi-model	109.27	-1.77	-6.95
13.	13.	13.	MariaDB +	Relational, Multi-model	97.23	-2.00	+0.42
14.	14.	14.	Splunk	Search engine	91.65	-1.07	+4.57
15.	↑ 16.	15.	Amazon DynamoDB +	Multi-model	82.90	+1.96	+3.21
16.	↓ 15.	16.	Microsoft Azure SQL Database	Relational, Multi-model	79.56	-1.51	+0.81
17.	17.	↑ 19.	Databricks +	Multi-model	76.91	-3.62	+16.58
18.	18.	↓ 17.	Hive	Relational	65.81	-1.15	-6.31
19.	19.	↑ 22.	Google BigQuery +	Relational	63.63	+0.15	+11.17
20.	20.	↓ 18.	Teradata	Relational, Multi-model	51.24	-1.94	-11.79
21.	21.	21.	FileMaker	Relational	50.48	-1.56	-2.32
22.	22.	↓ 20.	Neo4j +	Graph	46.61	-1.57	-8.82
23.	23.	23.	SAP HANA +	Relational, Multi-model	45.22	-1.22	-4.45
24.	24.	24.	Solr	Search engine, Multi-model	44.60	-0.53	-1.28
25.	25.	25.	SAP Adaptive Server	Relational, Multi-model	39.34	-0.09	-3.71
26.	26.	26.	HBase	Wide column	33.95	+0.09	-4.46
27.	27.	27.	Microsoft Azure Cosmos DB +	Multi-model	31.99	-1.48	-4.52
28.	28.	↑ 29.	InfluxDB +	Time Series, Multi-model	28.48	+0.92	-0.96
29.	↑ 30.	↓ 28.	PostGIS	Spatial, Multi-model	25.88	-0.78	-3.93
30.	↑ 31.	↑ 34.	Microsoft Azure Synapse Analytics	Relational	25.01	-1.27	+3.23

数据库课程的基本主线



数据库课程的基本主线

- 存储和一致性
 - SQL 已经不是最主要的了（不是不重要，而是不是最重要的）
 - 所有数据库最大的差异来源于存储格式、索引数据结构、数据一致性保障
 - 核心的逻辑和技术实现
- 选项和权衡点
 - 选择和平衡，成为不同数据库设计理念的巨大分野
 - 为可用是否可以/怎么牺牲一致性；为读取是否可以/怎么牺牲写入性能，反过来呢？
- 扩展和分布式
 - 解决扩展性和搞可用性的唯一方式，目前是分布式系统

Practice in class 1-1

- 打开你手头的数据库，去了解
 - 这个数据库的版本演变；
 - 存储的特征
 - SQL引擎
 - 事务处理级别划分
 - 综合以上的，称之为——体系结构及背后的考量、权衡、最佳实践场景