

# 数据库系统概论



计算机与信息学院  
人工智能学院

➤ 数据库系统概论（第六版），王珊等，高等教育出版社，2023

➤ 课件参考

➤ [李国良 @ 清华大学 \(tsinghua.edu.cn\)](http://nigongliang@tsinghua.edu.cn)

➤ [http://dbgroun.cs.tsinghua.edu.cn/ligl/courses\\_cn.html](http://dbgroun.cs.tsinghua.edu.cn/ligl/courses_cn.html)

· 数据库管理系统

◦ 课程简介

- 在信息化时代，大量数据被生产、传输、存储、查询和分析，这些数据都需要被有效地管理起来。数据库管理系统则是用来有效管理数据的核心基础软件，被称为IT系统皇冠上的明珠。数据库管理系统是数字化时代的核心基础设施，已经广泛地应用于各行各业（例如金融、电信、制造、能源等），对社会的发展和人类进步产生了深远的影响。学习数据库不是简单的学会“用”数据库，更重要的是学习如何“造”数据库（自主研发数据库）。本课程主要讲述数据库管理系统核心技术，不仅介绍数据库的基本原理，更重要的是介绍数据库的构建方法。

# 1 基本概念

- 数据 (Data)
- 数据库 (Database, DB)
- 数据库管理系统 (Database Management System, DBMS)
- 数据库系统 (Database System, DBS)

➤ 数据 (Data)

➤ 数据库 (Database, DB)

➤ 数据库管理系统 (Database Management System, DBMS)

➤ 数据库系统 (Database System, DBS)

- 描述事物的符号
- 数据库中存储的基本对象

种类：数字、文本、时间、向量、图像、音视频.....

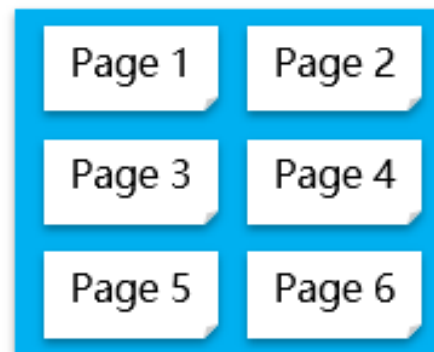
数据是有语义的

# 基本概念

- 数据 (Data)
- 数据库 (Database, DB)
- 数据库管理系统 (Database Management System, DBMS)
- 数据库系统 (Database System, DBS)

- 长期储存在计算机中的有组织的、可管理和可共享的数据集合。
- 数据按一定的数据模型组织、描述和储存

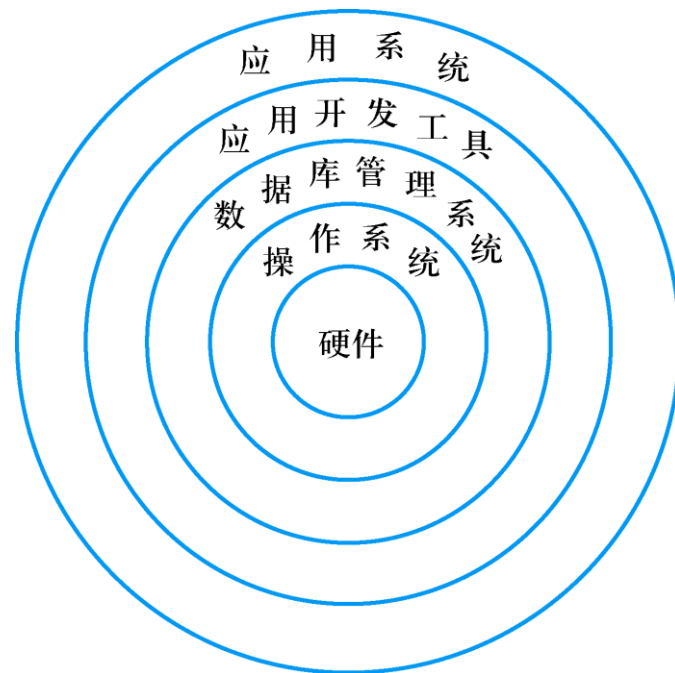
磁盘

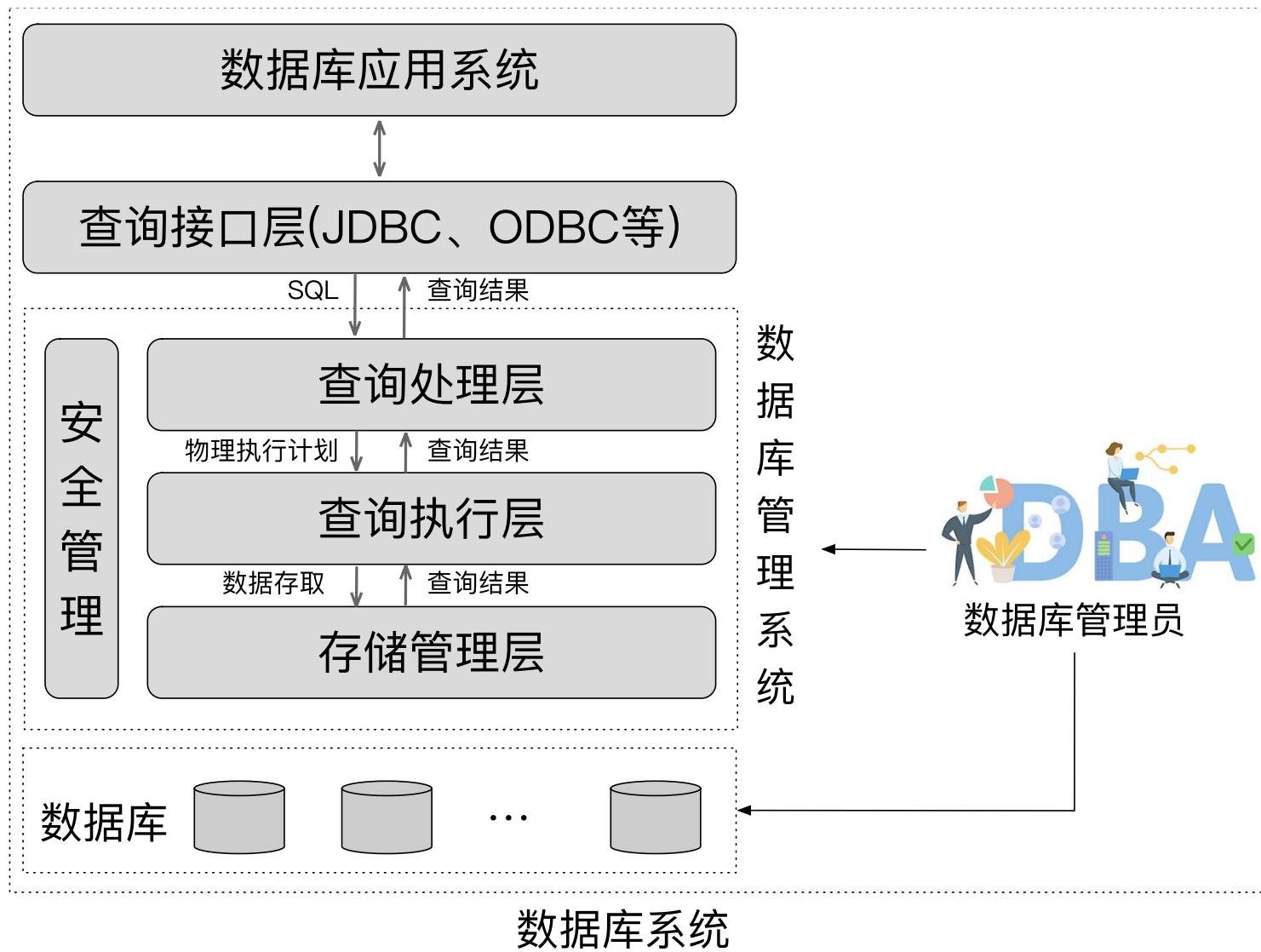


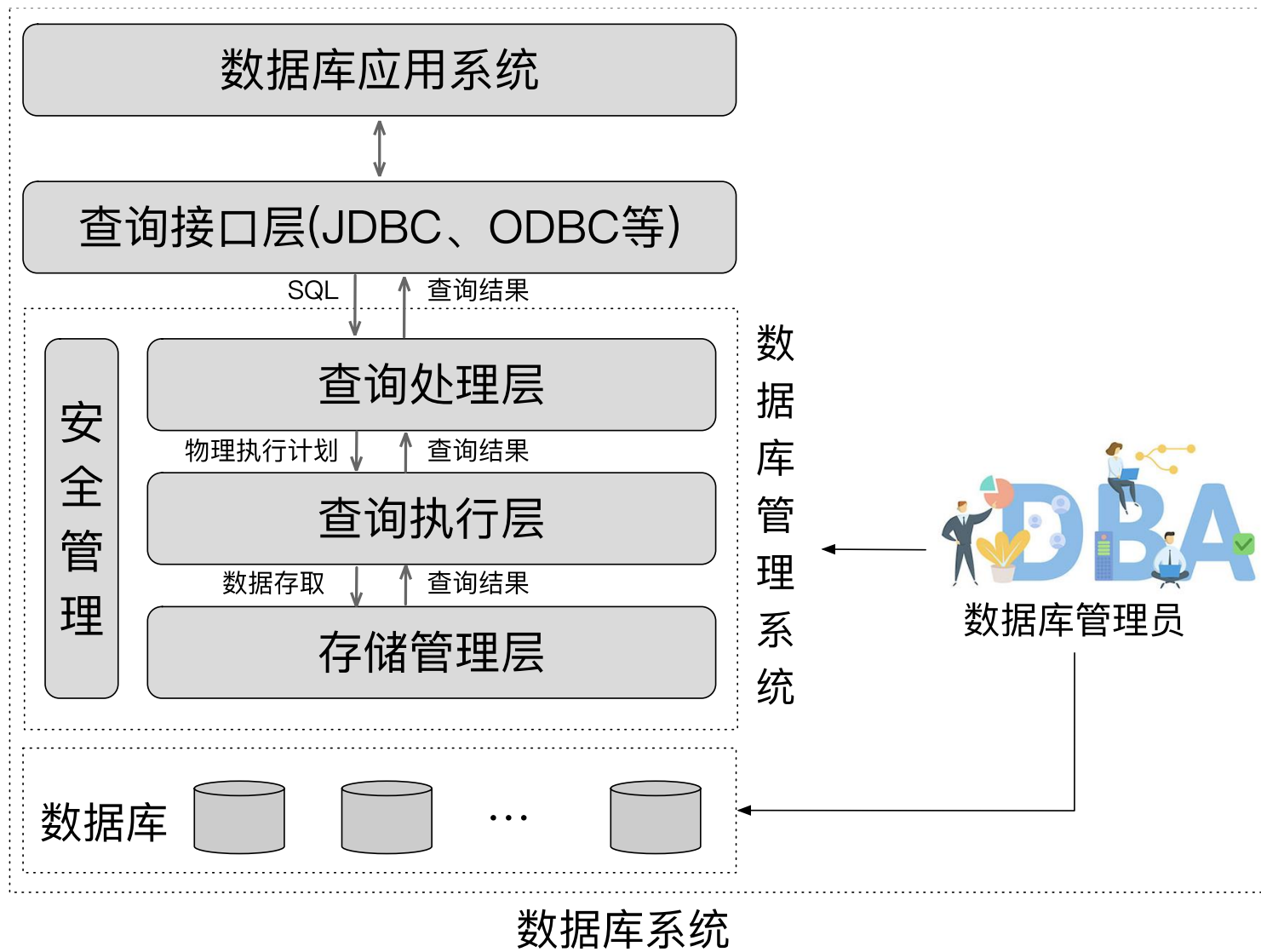
# 基本概念

- 数据 (Data)
- 数据库 (Database, DB)
- 数据库管理系统 (Database Management System, DBMS)
- 数据库系统 (Database System, DBS)

- 应用程序与操作系统之间的**数据管理软件**
- 一个管理数据的大型复杂基础软件系统









- 数据 (Data)
- 数据库 (Database, DB)
- 数据库管理系统 (Database Management System, DBMS)
- 数据库系统 (Database System, DBS)

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

- ◆ 数据库
- ◆ 数据库管理系统
- ◆ 开发环境/工具
- ◆ 数据库管理员和终端用户

➤ 数据库系统概论（第六版），王珊等，高等教育出版社，2023

➤ 课件参考

➤ [李国良 @ 清华大学 \(tsinghua.edu.cn\)](http://nigjia@tsinghua.edu.cn)

➤ [http://dbgroun.cs.tsinghua.edu.cn/ligl/courses\\_cn.html](http://dbgroun.cs.tsinghua.edu.cn/ligl/courses_cn.html)

· 数据库管理系统

◦ 课程简介

- 在信息化时代，大量数据被生产、传输、存储、查询和分析，这些数据都需要被有效地管理起来。数据库管理系统则是用来有效管理数据的核心基础软件，被称为IT系统皇冠上的明珠。数据库管理系统是数字化时代的核心基础设施，已经广泛地应用于各行各业（例如金融、电信、制造、能源等），对社会的发展和人类进步产生了深远的影响。学习数据库不是简单的学会“用”数据库，更重要的是学习如何“造”数据库（自主研发数据库）。本课程主要讲述数据库管理系统核心技术，不仅介绍数据库的基本原理，更重要的是介绍数据库的构建方法。

## 2 数据库产品

➤ 数据库排名权威网站DB-Engines

➤ 墨天轮-国产数据库排名

# DB Engine Ranking

([DB-Engines Ranking](#) - popularity ranking of database management systems)

424 systems in ranking, February 2025

Rank			DBMS	Database Model	Score		
Feb 2025	Jan 2025	Feb 2024			Feb 2025	Jan 2025	Feb 2024
1.	1.	1.	Oracle	Relational, Multi-model	1254.82	-3.93	+13.38
2.	2.	2.	MySQL	Relational, Multi-model	999.99	+1.84	-106.67
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	786.87	-11.69	-66.70
4.	4.	4.	PostgreSQL	Relational, Multi-model	659.62	-3.79	+30.21
5.	5.	5.	MongoDB	Document, Multi-model	396.63	-5.87	-23.73
6.	7.	6.	Redis	Key-value, Multi-model	157.91	+4.55	-2.80
7.	6.	9.	Snowflake	Relational	155.58	+1.68	+28.13
8.	8.	7.	Elasticsearch	Multi-model	134.63	-0.29	-1.11
9.	9.	8.	IBM Db2	Relational, Multi-model	125.44	+2.46	-6.79
10.	10.	10.	SQLite	Relational	113.82	+7.13	-3.47

中国数据库排行 - 墨天轮 (modb.pro)

排行	上月	半年前	名称	模型 ∨	数据处理 ∨	部署方式 ∨	商业模式 ∨	专利	论文	案例	资质	书籍	得分
🏆1	1	↑ 2	OceanBase	关系型	HP	☒ ☒	☒ ☒	151	40	55	16	2	692.51
🏆2	2	↓ 1	PolarDB	关系型	TP HP	☒ ☒ ☒	☒ ☒	592	78	85	14	2	686.61
🏆3	↑ 4	↑ ↑ ↑ 8	GaussDB	关系型	TP HP	☒ ☒ ☒	☒	950	64	36	17	5	682.74
4	↓ 3	↑ ↑ ↑ 9	GoldenDB	关系型	HP	☒ ☒	☒	806	54	98	12	2	663.27
5	5	↓ 4	TiDB	关系型	HP	☒ ☒	☒ ☒	40	28	19	10	1	587.06
6	6	↓ 5	金仓数据库	关系型	TP AP	☒ ☒ ☒	☒	576	1	64	12	4	570.72
7	7	7	GBASE	关系型	AP TP	☒ ☒ ☒	☒	277	0	84	13	0	551.27
8	8	↓ ↓ 6	达梦数据库	关系型	TP	☒ ☒ ☒	☒	518	0	21	8	12	517.56
9	9	↓ ↓ ↓ 3	openGauss	关系型	TP	☒	☒	570	20	34	7	6	422.66
10	10	10	TDSQL	关系型	HP	☒ ☒	☒ ☒	122	23	21	13	0	358.84

2025年1月排行榜前十产品得分情况

### 3 Data Model (数据模型)

定义： 对数据结构，数据语义、数据之间的关系、数据约束等特征的抽象描述

A collection of tools for describing

Data

Data relationships

Data semantics

Data constraints

## 用二维表来描述数据

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

**行 (row) : 记录**

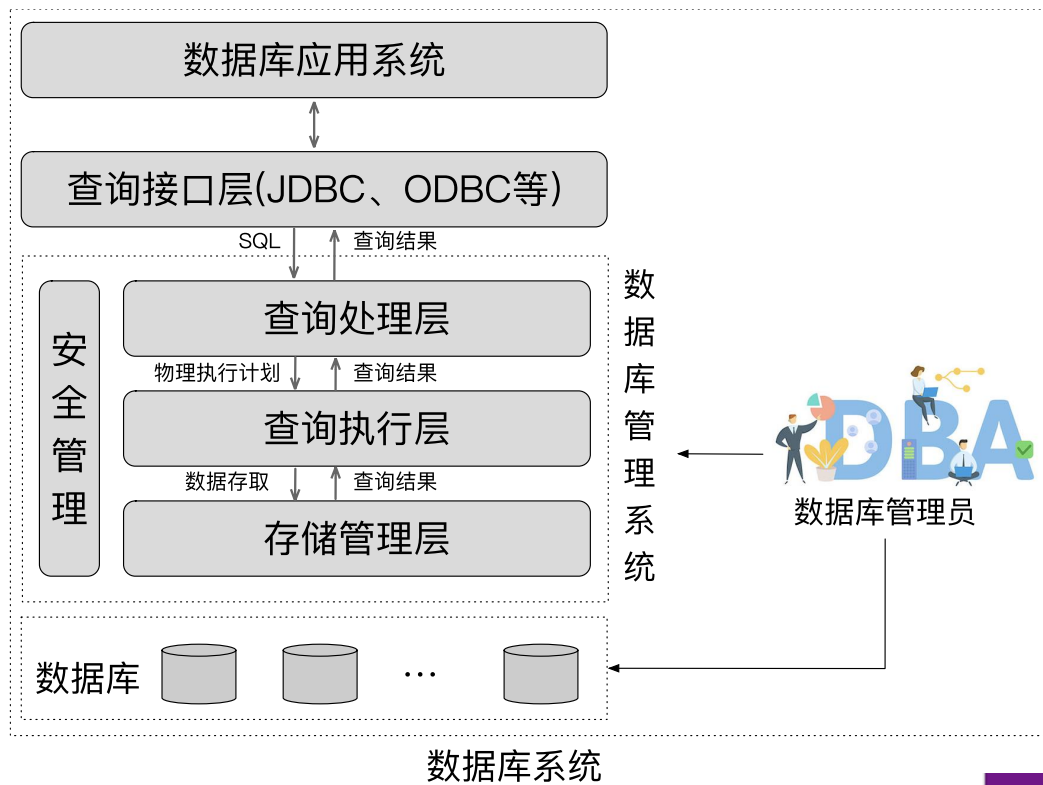
**列 (column) : 属性**

**约束**

DBMS采用的, 便于应用程序和开发人员使用的数据模型:

- ✓ 关系模型
- ✓ 层次模型、网状模型
- ✓ 面向对象模型
- ✓ 对象关系模型
- ✓ 半结构化模型: XML (Extensible Markup Language)、JSON (JavaScript Object Notation )
- ✓ 图模型
- ✓ 矩阵向量模型





Data model:关系模型

<u>Sno</u> (学号)	<u>Sname</u> (姓名)	<u>Sgender</u> (性别)	<u>Sage</u> (年龄)	<u>Sdept</u> (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

## ➤ DBMS的主要功能

- 数据定义（提供数据定义语言DDL，定义数据库中的数据对象）
- 数据操纵（提供数据操纵语言DML，对数据进行增、删、修等操作）
- 数据组织、存储、备份
- 查询执行、事务管理、并发控制

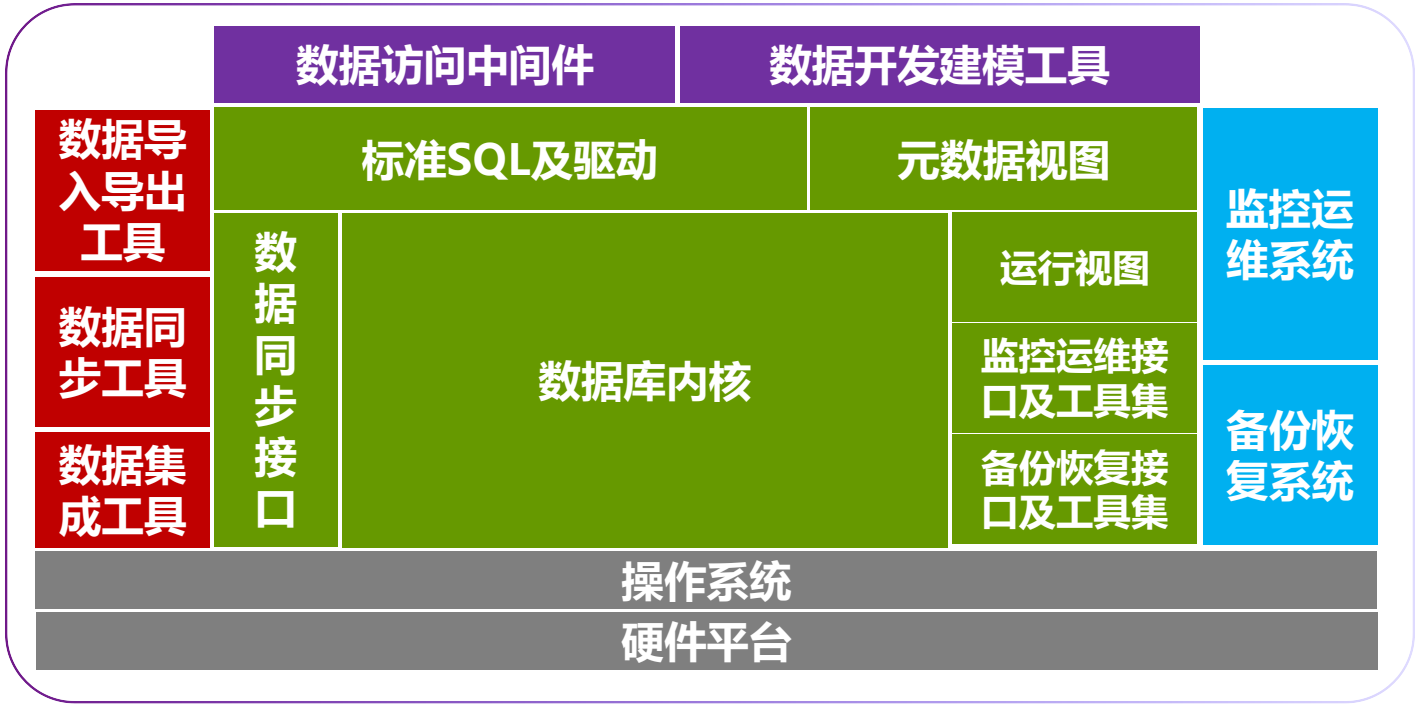
# DBMS周边工具

## 北向（应用用得最好）：

- 数据访问中间件：业务访问数据库的通道；
- 数据开发建模工具：建模、设计、开发、调试业务数据访问代码；

## 西向（数据进的来出得去）：

- 导入导出工具：数据离线迁移；
- 同步工具：数据在线迁移；
- 集成工具：数据导出到分析系统；



## 东向（数据库运行得好）：

- 监控运维系统：配置、告警、统计、巡检等；
- 备份恢复系统：数据的备份与恢复；

## 南向（数据库运行的基础平台）：

- 操作系统：运行的软件平台；
- 硬件平台：运行的硬件平台；

# DBMS发展

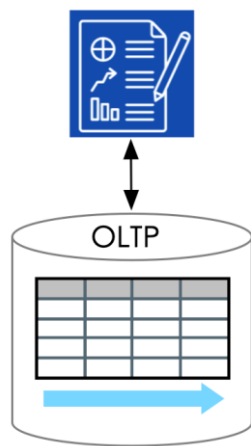


排行	上月	半年前	名称	模型 ∨	数据处理 ∨	部署方式 ∨	商业模式 ∨	专利	论文	案例	资质	书籍	得分
	1	↑ 2	OceanBase	关系型	HP			151	40	55	16	2	692.51
	2	↓ 1	PolarDB	关系型	TP HP			592	78	85	14	2	686.61
	↑ 4	↑ ↑ ↑ 8	GaussDB	关系型	TP HP			950	64	36	17	5	682.74
4	↓ 3	↑ ↑ ↑ 9	GoldenDB	关系型	HP			806	54	98	12	2	663.27
5	5	↓ 4	TiDB	关系型	HP			40	28	19	10	1	587.06
6	6	↓ 5	金仓数据库	关系型	TP AP			576	1	64	12	4	570.72
7	7	7	GBASE	关系型	AP TP			277	0	84	13	0	551.27
8	8	↓ ↓ 6	达梦数据库	关系型	TP			518	0	21	8	12	517.56
9	9	↓ ↓ ↓ 3	openGauss	关系型	TP			570	20	34	7	6	422.66
10	10	10	TDSQL	关系型	HP			122	23	21	13	0	358.84

2025年1月排行榜前十产品得分情况

## 数据库系统的业务应用场景

✓ OLTP ( Online Transaction Processing ) : 在线事务处理

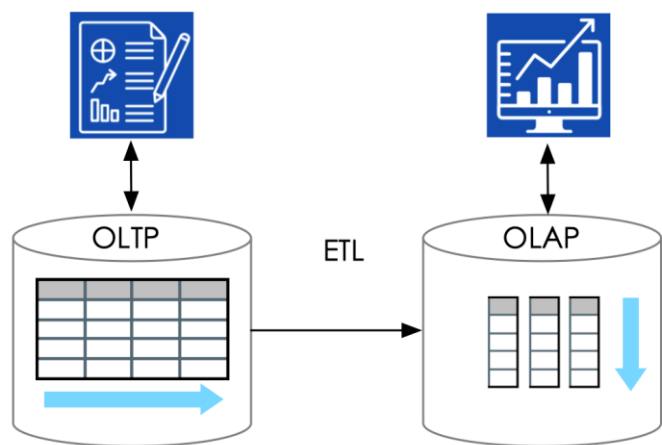


- 处理业务交易数据，用于支持实时、并发的用户事务，如网上银行、购物和订单管理等

- ✓ 数据正确（事务管理）
- ✓ 高并发、低延迟、高吞吐量
- ✓ 数据存储和操作通常针对单行或少量行进行

## 数据库系统的业务应用场景

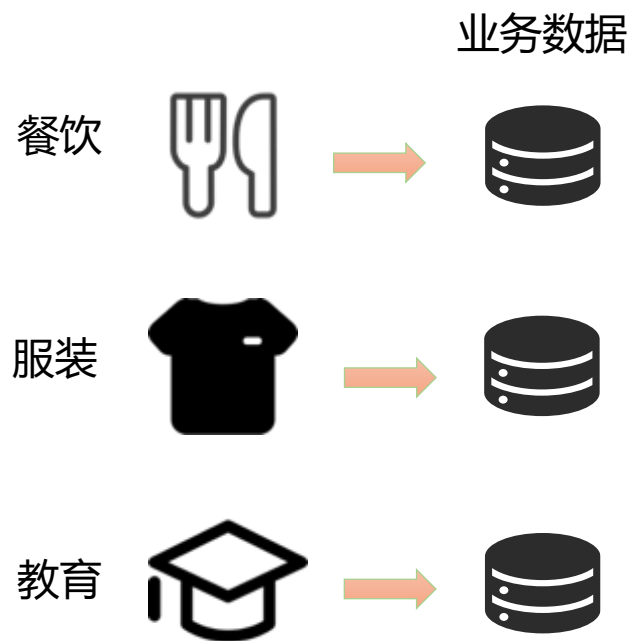
✓ OLAP ( Online Analytical Processing ) : 在线分析处理



- 从大量数据中做数据分析，挖掘数据的价值

- ✓ 数据收集
- ✓ 清理、汇总和存储数据
- ✓ 从多维数据集中进行查询并生成报告

## OLTP场景



汇总



分析

决策

## OLAP场景



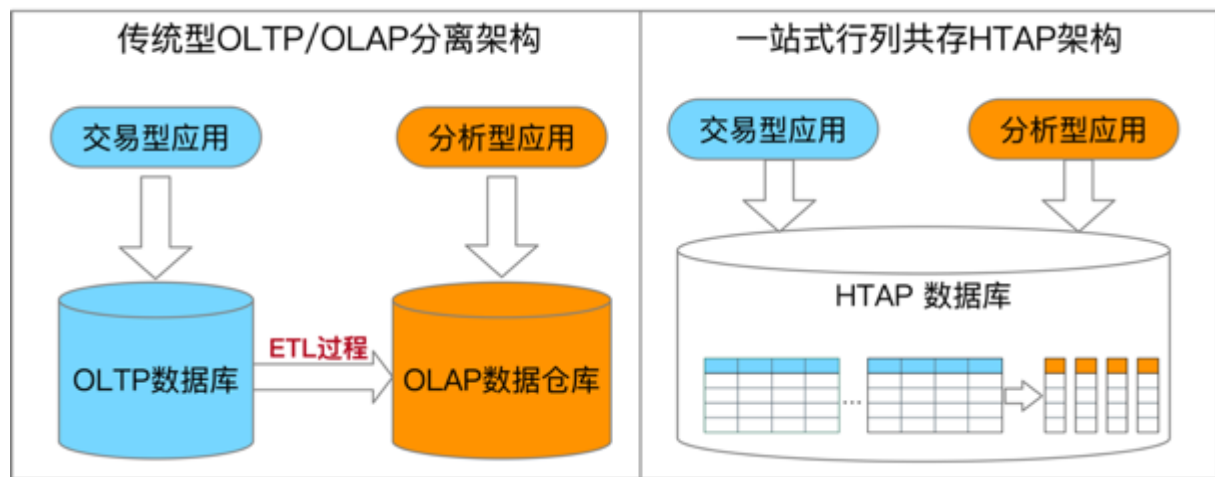
评估：每秒执行的Transaction以及SQL的数量

例如：GuassDB 100 (GuassDB T)

评估：磁盘系统的吞吐量

例如：GuassDB 200 (GuassDB A)

- HTAP（Hybrid Transactional/Analytical Processing）混合事务分析处理
  - ✓ 统一支持OLTP和OLAP场景：OLTP与OLAP负载都可以被快速执行与响应
  - ✓ 常用场景：在线电商平台、金融反欺诈系统等



GuassDB 300：企业级分布式HTAP数据库



# 课程内容

## DBMS基础

- ✓ 关系模型
- ✓ 关系代数
- ✓ 基础SQL（增删修、查询）
- ✓ 高级SQL（存储过程、游标、触发器）

## DBMS内核

- ✓ 存储和索引
- ✓ 查询处理和优化
- ✓ 事务（数据备份与恢复、并发控制）

## DBS设计

- ✓ ER模型
- ✓ 关系数据理论（范式）
- ✓ DBS设计的步骤
- ✓ 数据库编程接口

# 课程安排

- 课堂：40学时
- 实验：8学时 (SQL)
- 【课程设计】:24学时 (DBS设计、数据库接口、SQL)

# 数据管理的三个阶段

— 20世纪50年代中期

20世纪50年代中期 — 60年代中期

20世纪60年代末 — 至今

人工管理

文件系统管理

数据库管理

	人工管理	文件系统管理	数据库管理
应用背景	满足科学计算	满足科学计算和数据管理	大规模数据管理
硬件条件	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
软件条件	没有操作系统	有文件系统	数据库管理系统
处理方式	批处理	联机实时处理、批处理	联机实时处理、批处理、分布式处理
实现方式	通过卡带、卡片和磁带等计算机间接存取设备实现	通过磁盘和磁鼓等计算机直接存取设备实现	通过数据库管理系统与数据库进行交互以实现数据管理操作
特点	<ul style="list-style-type: none"><li>人力维护成本极高</li><li>数据与应用程序深度绑定</li><li>数据间缺乏组织和独立性</li><li>数据无共享、冗余度极大</li></ul>	<ul style="list-style-type: none"><li>人力维护成本很高</li><li>数据文件与应用程序耦合</li><li>数据共享性差、冗余度大</li><li>难以支持多用户并发访问</li></ul>	<ul style="list-style-type: none"><li>使用数据模型描述，无需应用程序定义，组织程度和共享程度高</li><li>编程友好，无需关心数据库文件的物理操作和系统控制</li><li>支持多用户并发访问</li></ul>

# 作业

见学习通