

数据库的前世今生

管长龙

KCN 数据库运营

2021.01.19

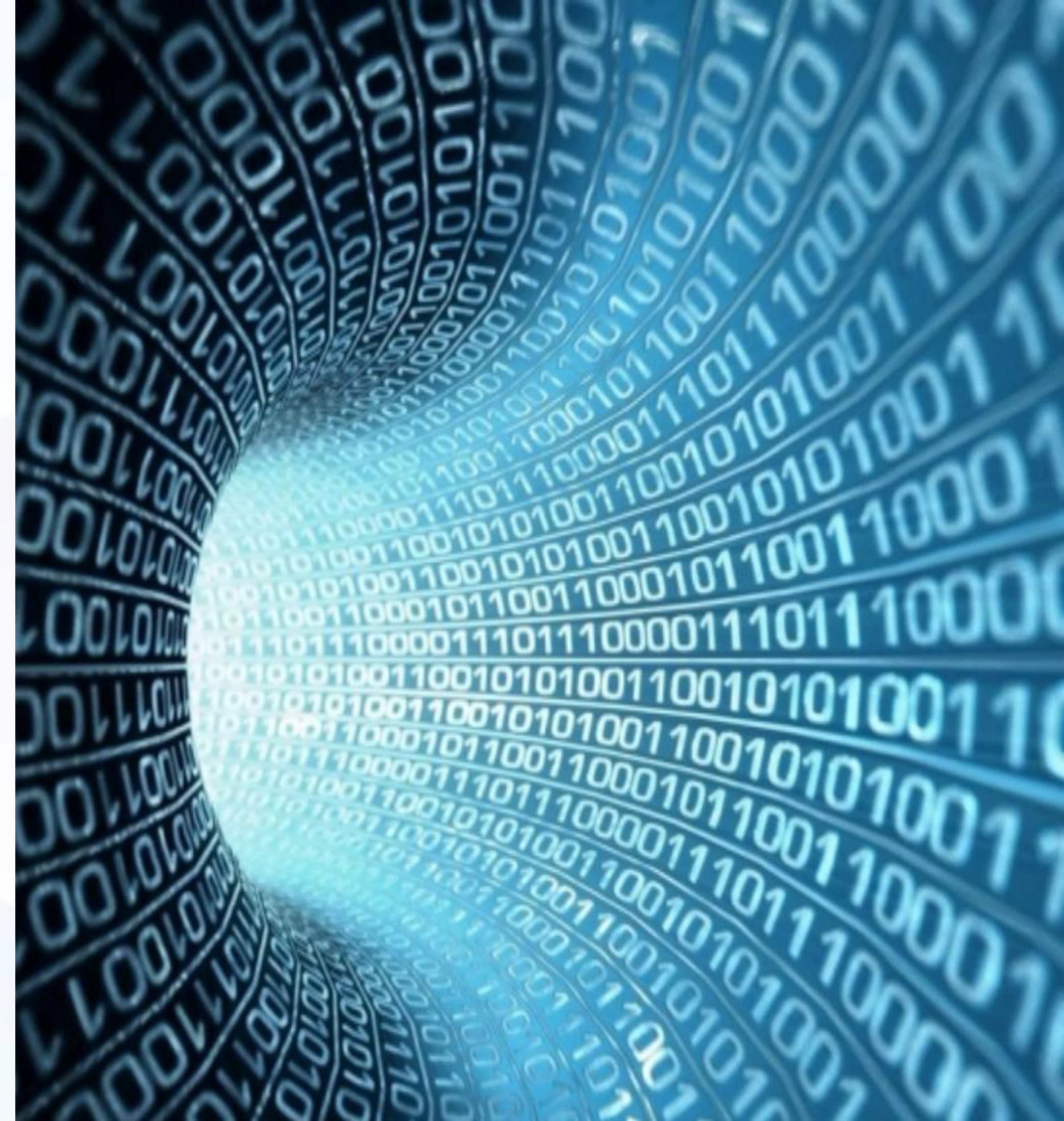
| 大纲

- 数据库概述
- 数据库发展史
- 数据库的分类

什么是数据库?

数据

数据 (data) 在 **计算机科学** 中，数据是所有能输入计算机并被计算机程序处理的符号的介质的总称，是用于输入电子计算机进行处理，具有一定意义的 **数字、字母、符号和模拟量** 等的通称。



数据库

数据库 (database) 是一个长期存储在计算机内的、有组织的、可共享的、统一管理的 **大量数据的集合**。



数据库管理系统

数据库管理系统（Database Management System）操纵和管理数据库的 **大型软件**，简称 DBMS。

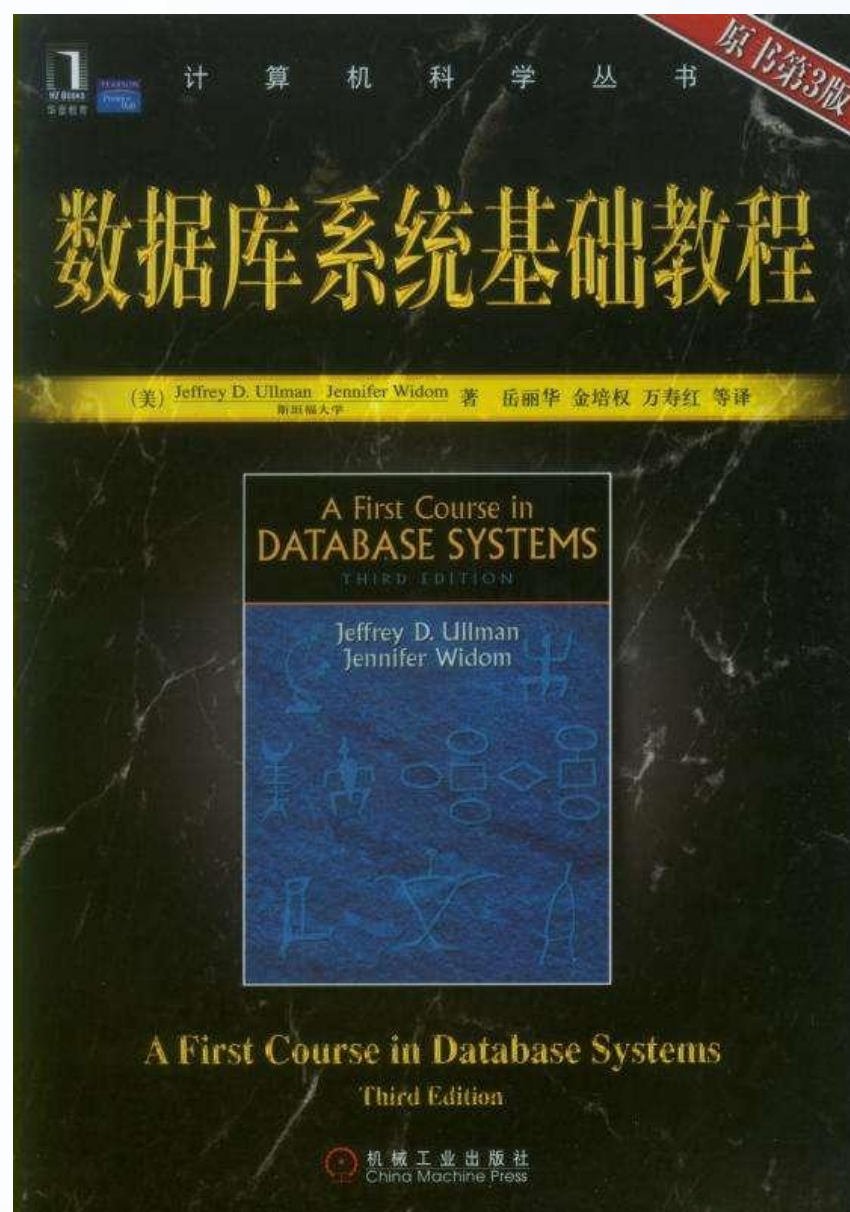
对 **数据库** 进行统一的管理和控制，以保证数据库的安全性和完整性。

用户通过应用等方式访问数据库中的数据，数据库管理员（DBA）也通过 DBMS 进行数据库的维护工作。

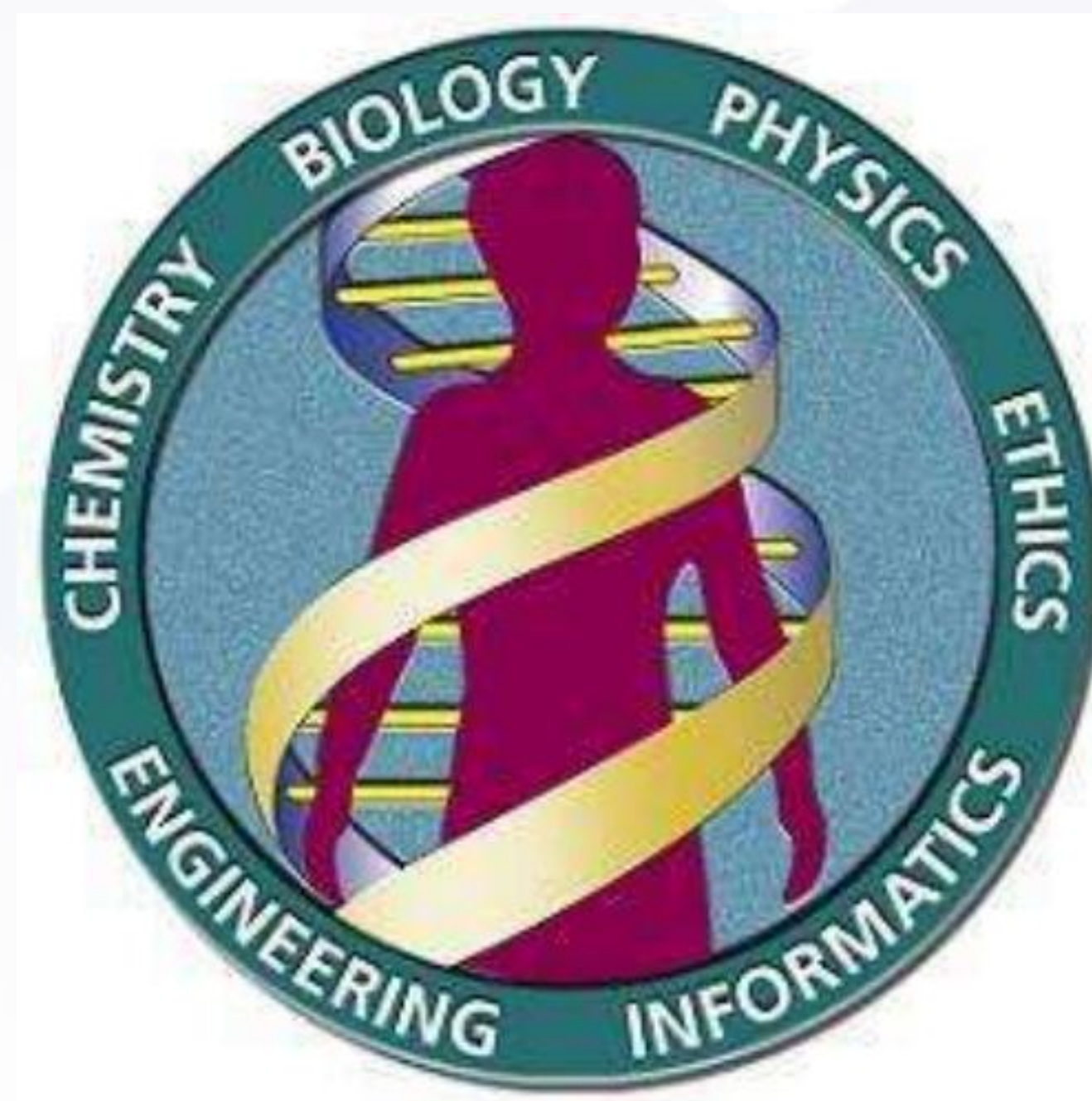
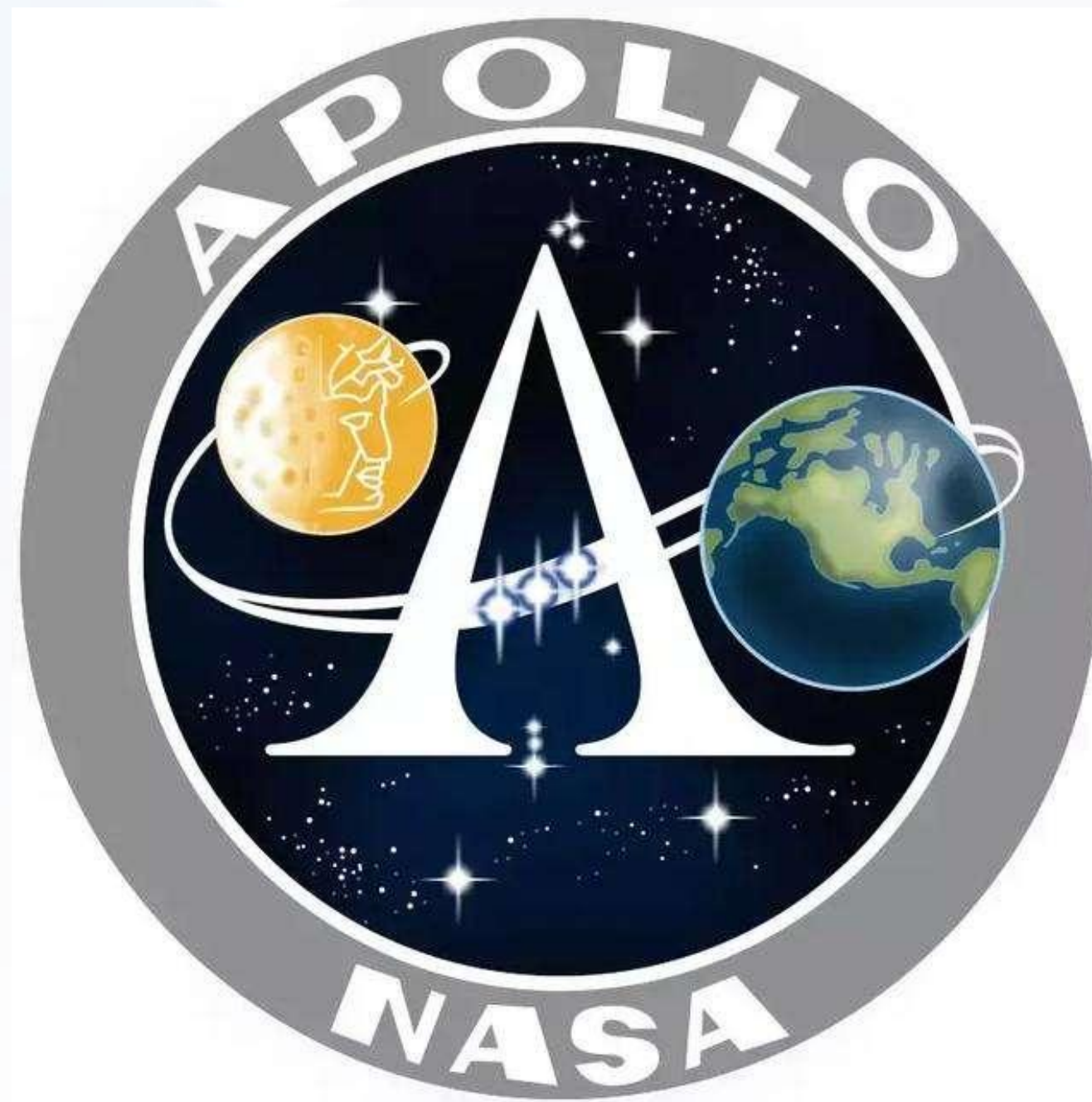
* 人们习惯将 **数据库管理系统** 简称为 **数据库**

数据库概述

现实类比

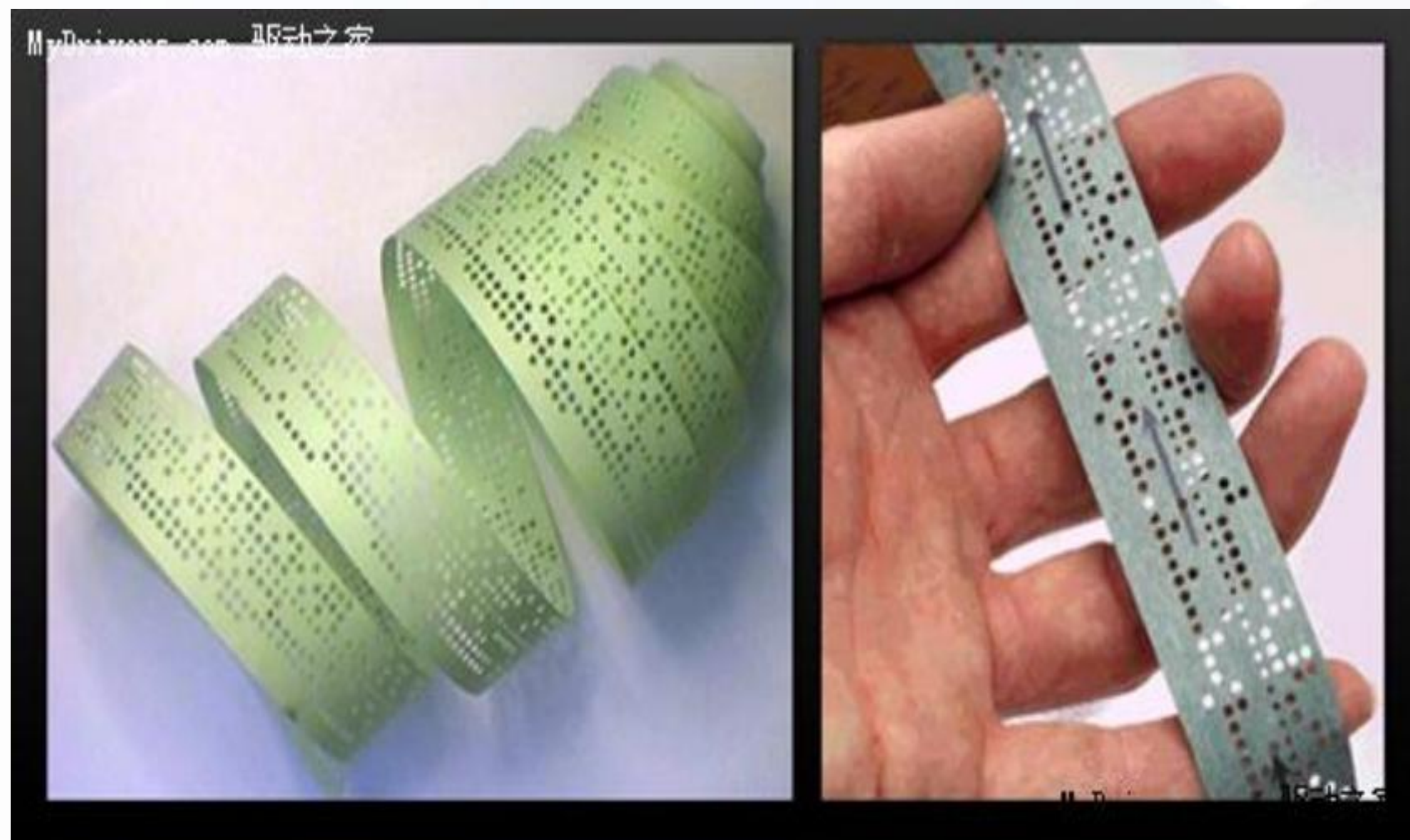
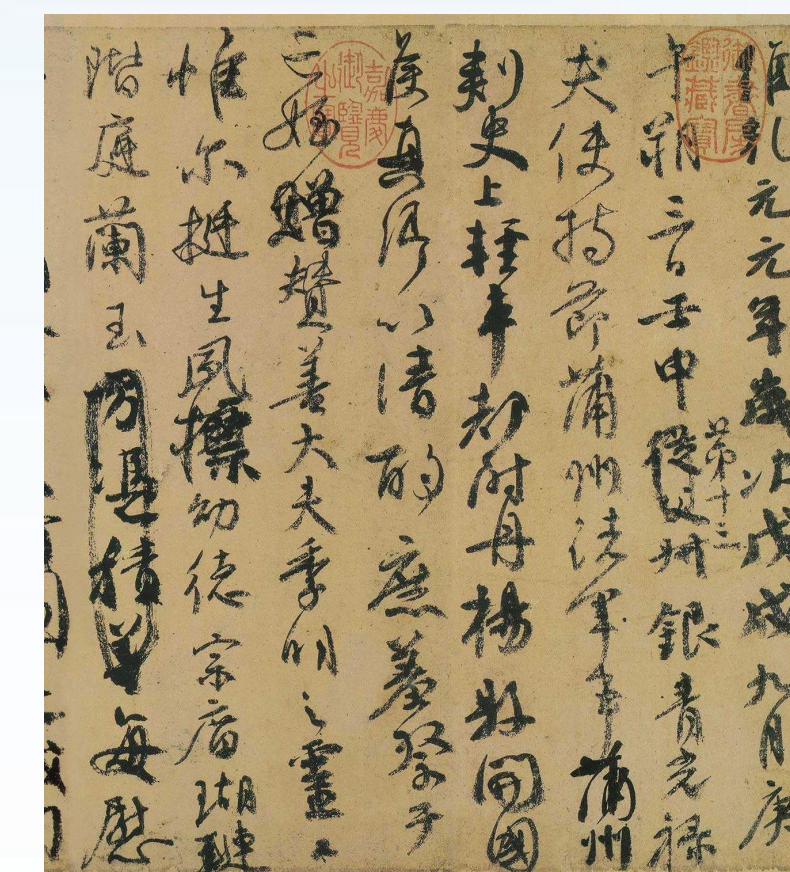


数据库出现的历史必然性



数据库概述

存在哪？



数据库概述

怎么存取？

[illegible]

| 数据库发展史

层次型、网络型

1950s

科学研究阶段

代表：
IMS (IBM 公司)

关系型

1970s

商业处理

代表：
DB2 (IBM 公司)
Oracle (甲骨文)

关系型、数据仓库

1990s

企业信息化、数字化

代表：
PostgreSQL (开源)
SQLServer (微软)

开源、NoSQL

2000s

电子商务、社交、媒体

代表：
MySQL (开源)
BigData (Google)
MongoDB (开源)

云数据库

2015s

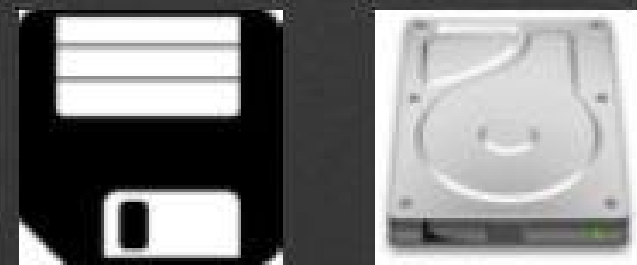
云计算、物联网、金融

代表：
AWS RDS (亚马逊)
阿里云 RDS (阿里巴巴)
.....

未来.....

至80年代中期

File Centric



KB - MB

至10年代末期

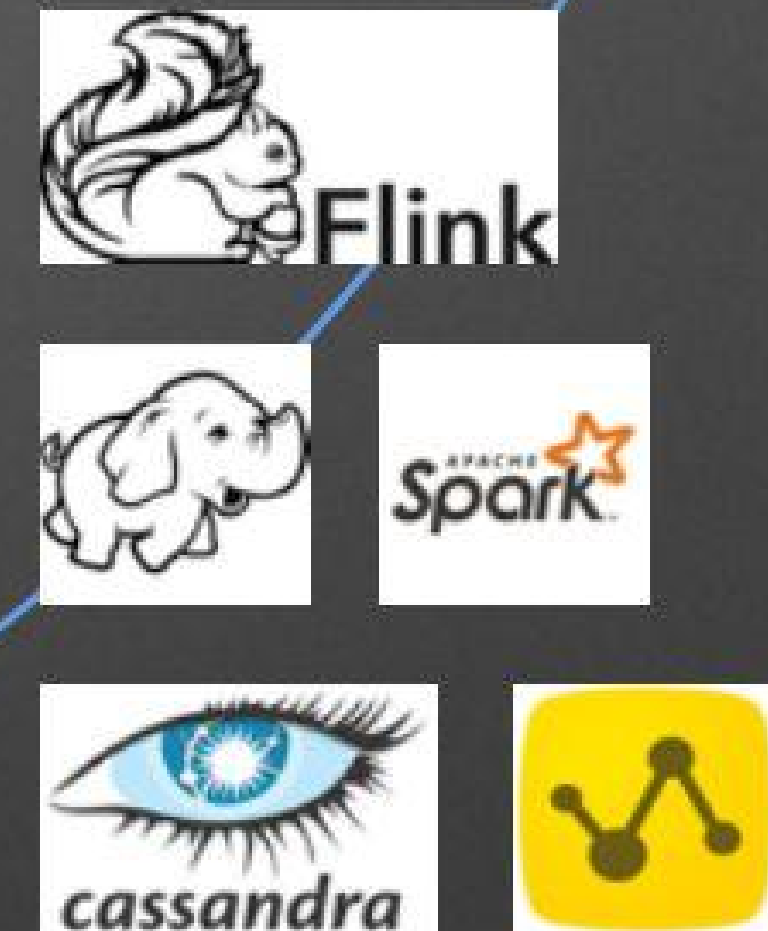
DB Centric



MB - GB

至今

Bigdata Centric



100GB - PB

云栖社区 yq.aliyun.com

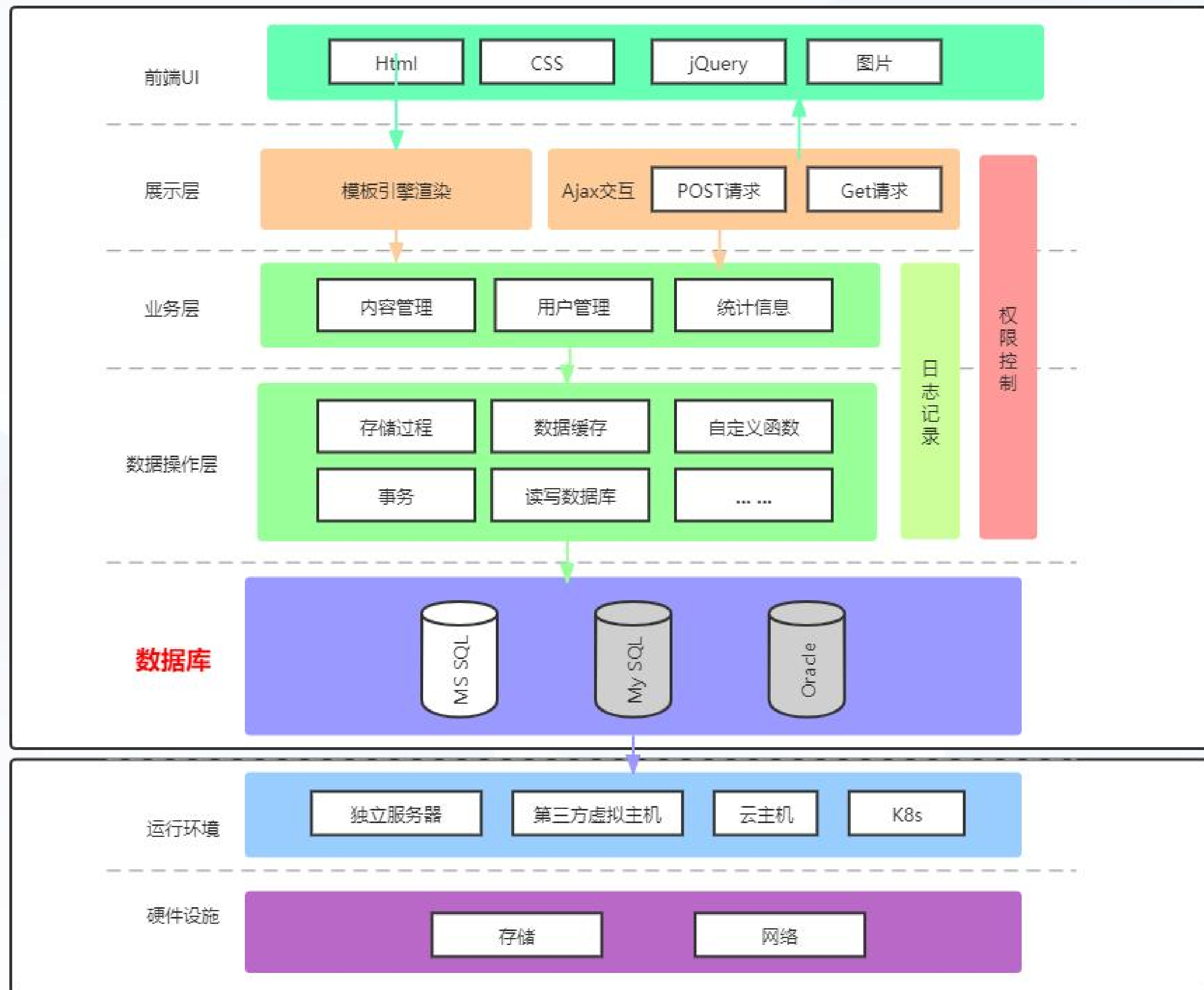
数据库在架构中的具体位置？

| 数据库发展史

数据库的位置

一般位于整体应用的 **最底层**。

属于应用价值的 **核心**！



数据库有哪些分类？

有哪些代表数据库？

不同的分类角度

数据结构：关系型、NoSQL（文档、键值、图、时序等）、NewSQL（TiDB）

业务类型：事务类（OLTP）、分析类（OLAP）-> HTAP

架构复杂度：单机、分布式

部署（交付）方式：本地部署、云上部署 -> 云原生

商业模式：开源、商业

<https://db-engines.com/en/article/Relational+DBMS?ref=RDBMS>

| 数据库的分类

关系型数据库 (RDBMS)

以 **二维表** 的形式来描述数据。

使用结构化查询语言 **SQL** (Structured Query Language) 来进行操作!

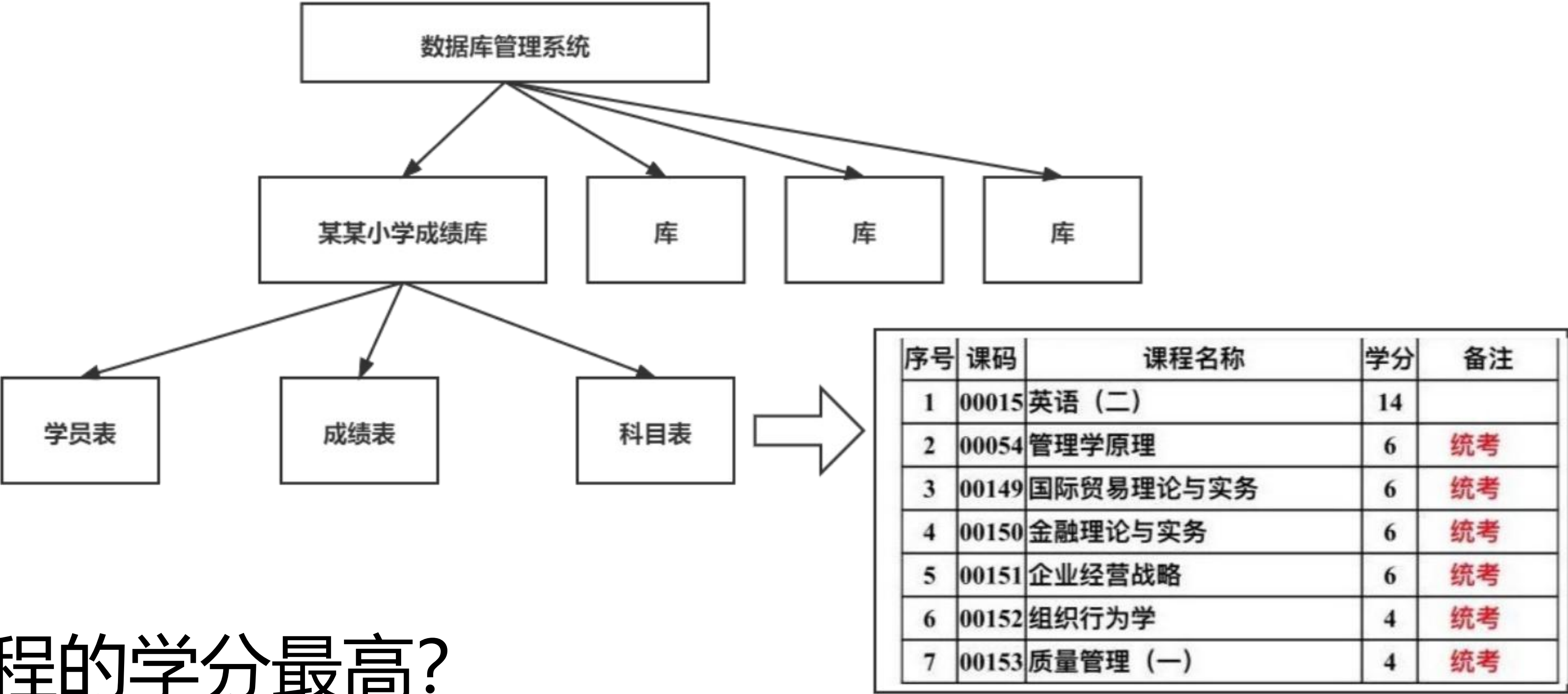
是当今世界主流的数据库结构化模式。

代表数据库:

Oracle、**MySQL**、**SQL Server**、**PostgreSQL**

<https://db-engines.com/en/article/Relational+DBMS?ref=RDBMS>

关系型数据库的数据存储结构示例



哪个课程的学分最高？

SQL：SELECT 课程名称 FROM 科目表 ORDER BY 学分 DESC LIMIT 1；

结果：英语（二）

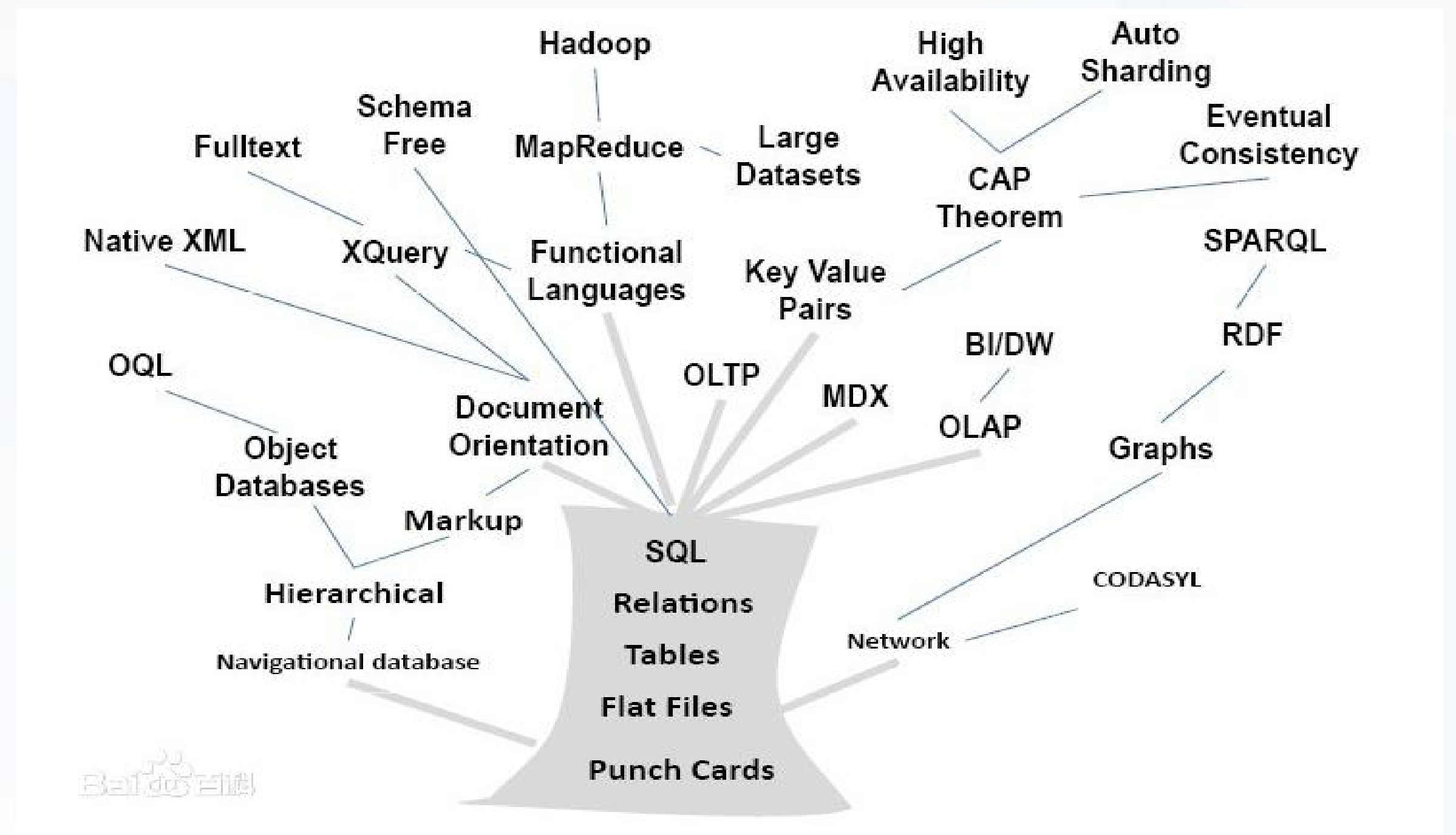
| 数据库的分类

非关系型数据库 (NoSQL)

不使用结构化查询语言 **SQL** (Structured Query Language) 来进行操作！针对特定的使用场景。

主要类型及代表数据库：

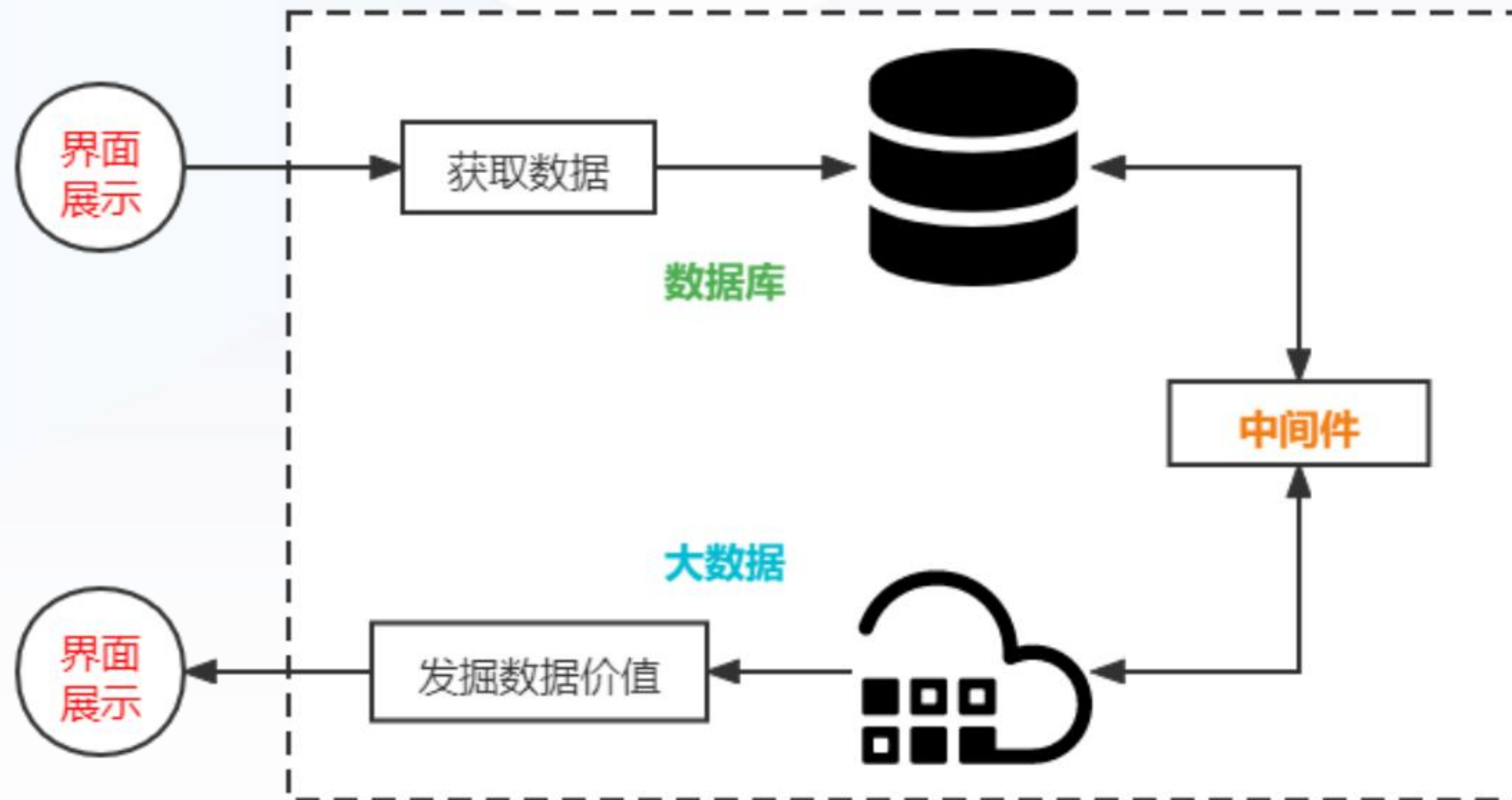
- 文档型：MongoDB
- 键值：Redis、Memcached
- 搜索引擎：Elasticsearch
- 图：Neo4j、Nebula Graph
- 时序：InfluxDB、ClickHouse (可实现)



事务类与分析类 (OLTP & OLAP)

	OLTP	OLAP
业务类型	日常交易和读写	统计、分析
数据类型	当前的、最新的	历史的、聚集的
数据规模	GB	TB ~ PB
使用者	用户	管理员

数据库跟大数据是什么关系？



兼具事务与分析 (HTAP)

能够统一支持事务处理 (TP) 和工作负载分析 (AP) 的数据库成为众多企业的需求。

在此背景下，由 Gartner 提出的 **HTAP**（混合事务 / 分析处理，Hybrid Transactional/Analytical Processing）的概念，行业认同持疑。

代表数据库：

TiDB MySQL + ClickHouse

| 数据库的分类

单机、集群、分布式

单机结构:



(累如狗)

全栈老实人

集群结构:



+



(凑合过)

全栈老实人

全栈老实人

分布式:



+



(男女搭配)

后端攻城狮

前端攻城狮

知

| 数据库的分类

单机、集群、分布式



单机



集群



分布式

| 数据库的分类

本地、云上

略：同 自建机房 和 云服务器 的区别。

云原生数据库

弹性、高可用、弹性扩展、开放生态

- Shared Everything: 共享硬件资源（一体化）
- Share-Nothing: 独立单元，通过协议互相关联（物理分割）

代表数据库及云环境：

Aurora (AWS)、Spanner (Google)、Socrates (微软)

TiDB (PingCAP)、PolarDB (阿里)、CynosDB (腾讯)

下期预告

- 主流数据库及对应使用场景
- 《2021 中国数据库研究报告》解读



changlongguan@unify.com