

数据库原理

The Theory of Database System

第一章 绪论



中国矿业大学计算机学院



中国矿业大学数据库原理精品课程

第一章 绪论

- 1.1 数据库、数据库管理系统和数据库系统
- 1.2 数据库系统的产生与发展
- 1.3 数据模型
- 1.4 数据库系统结构
- 1.5 数据库管理系统
- 1.6 数据库应用系统常见的几种结构



1.4 数据库系统结构

数据库系统内部的模式结构

从数据库管理系统角度看

数据库系统外部的体系结构

从数据库最终用户角度看



1.4.1 数据库系统的模式结构

- 数据库系统模式的概念
- 数据库系统的三级模式结构
- 数据库的二级映象功能与数据独立性
- 小结



数据库系统模式的概念

➤ “型” 和 “值” 的概念

— 型 (Type)

对某一类数据的结构和属性的说明

— 值 (Value)

是型的一个具体赋值

例如：学生记录

记录型：

(学号，姓名，性别，系别，年龄，籍贯)

该记录型的一个记录值：

(**900201**，李明，男，计算机，**22**，江苏)



数据库系统模式的概念（续）

➤ 模式（Schema）

- 数据库逻辑结构和特征的描述
- 是型的描述
- 反映的是数据的结构及其联系
- 模式是相对稳定的

➤ 实例（Instance）

- 模式的一个具体值
- 反映数据库某一时刻的状态
- 同一个模式可以有很多实例
- 实例随数据库中的数据更新而变动



关系模式

学生（学号，姓名，性别，院系，班级）

实例

S						
	SNO	CLASS	SNAME	SEX	AGE	DEPT
+	02501	计02-5	蔡广文	男	19	信电
+	02502	计02-5	陈昶铮	男	20	信电
+	02503	计02-5	冯仰家	男	20	信电
+	02504	计02-5	郭然	男	20	信电
+	02505	计02-5	黄杰	男	20	信电
+	02507	计02-5	李刚	男	20	计算机
+	02508	计02-5	李小川	男	20	计算机
+	02510	计02-5	刘平	男	20	计算机
+	02511	计02-5	卢旭	男	20	计算机
+	02512	计02-5	马运勇	男	20	计算机
+	02513	计02-5	钱军	男	20	计算机
+	02515	计02-5	田德生	男	20	机电
+	02516	计02-5	王晋	男	20	机电
+	02517	计02-5	王少帅	男	20	机电
+	02518	计02-5	王瑜	男	20	机电
+	02519	计02-5	谢凯峰	男	20	机电
+	02520	计02-5	杨万学	男	20	计算机



关系模式 (Relation Schema)

关系模式是关系中信息内容结构的描述。

$$R(U, D, DOM, I, \Sigma)$$

R: 是关系名

U: 是组成关系**R**的全部属性的集合

D: 是**U**中属性取值的值域

DOM: 是属性列到域的映射

I: 是一组完整性约束条件

$\Sigma(F)$: 是属性集间的一组数据依赖

简写: $R(U)$ 或 $R(U, F)$



设有一个关系模式**DS**，其中

U=(导师，专业，研究生)

F=(导师->专业，研究生->导师，
研究生->专业)

D= { **D1**= {张清枚，刘易，李勇，
刘晨，王敏}，

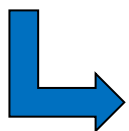
D2= {计算机专业、信息专业}
}



$D = \{D1 = \{\text{张清枚, 刘易, 李勇, 刘晨, 王敏}\},$
 $D2 = \{\text{计算机专业、信息专业}\} \}$



$\text{dom}(\text{导师}) = \text{dom}(\text{研究生}) = D1$



域名及属性向域的映像常说明为属性的
类型、长度



导师	专业	研究生
张清枚	信息专业	李勇
张清枚	信息专业	刘晨
刘易	计算机专业	王敏

导师	研究生
张清枚	李勇
张清枚	刘晨
刘易	王敏

导师	专业
张清枚	信息专业
刘易	计算机专业

研究生	专业
李勇	信息专业
刘晨	信息专业
王敏	计算机专业

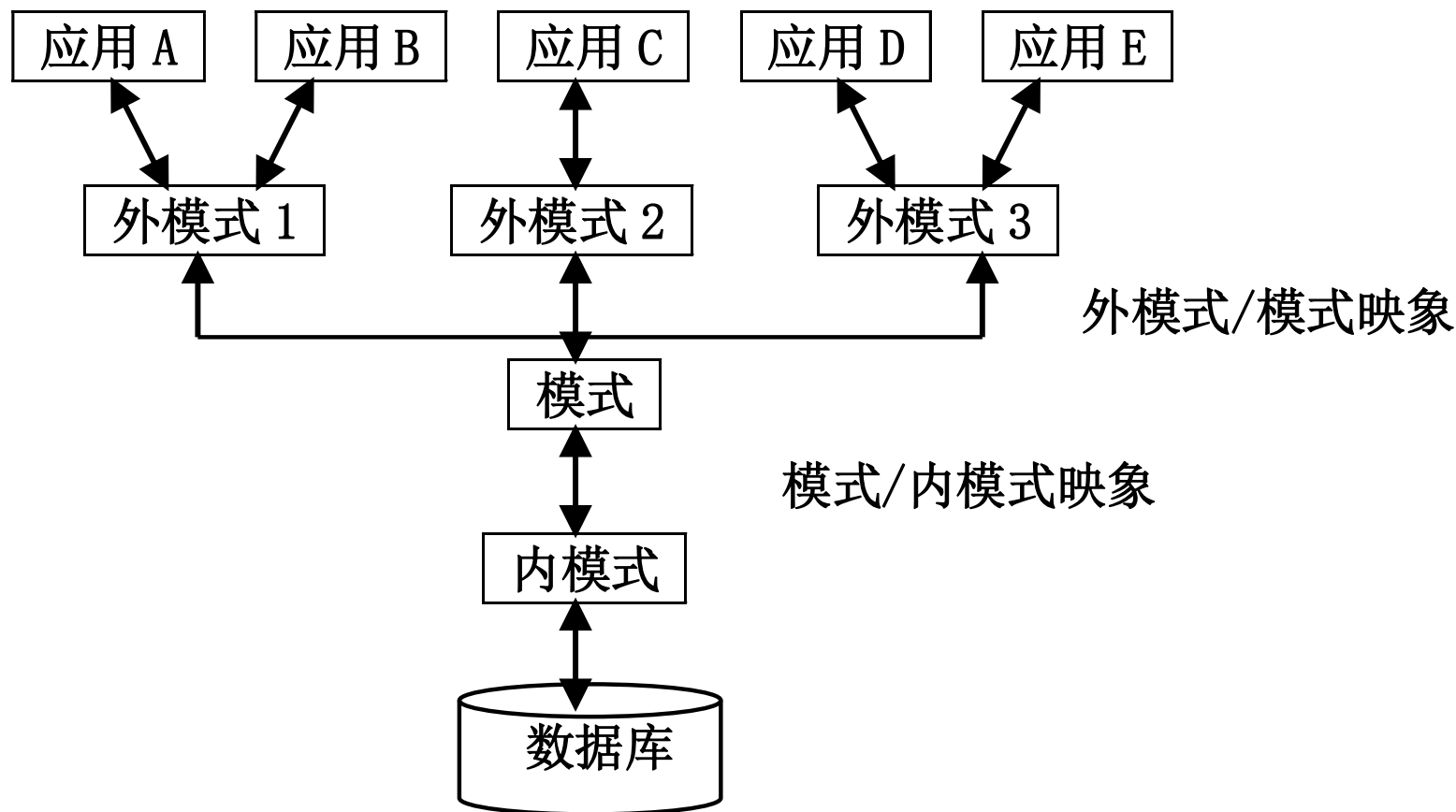


1.4 数据库系统的模式结构

- 数据库系统模式的概念
- 数据库系统的三级模式结构
- 数据库的二级映像功能与数据独立性
- 小结



数据库系统的三级模式结构



1. 模式 (Schema)

➤ 模式 (也称逻辑模式)

- 数据库中全体数据的逻辑结构和特征的描述
- 所有用户的公共数据视图, 综合了所有用户的需求
- 一个数据库只有一个模式

➤ 模式的地位: 是数据库系统模式结构的中间层

- 与数据的物理存储细节和硬件环境无关
- 与具体的应用程序、开发工具及高级程序设计语言无关

➤ 模式的定义

- 数据的逻辑结构 (数据项的名字、类型、取值范围等)
- 数据之间的联系
- 数据有关的安全性、完整性要求



教务系统

- 学生（学号，姓名，班级，专业，院系）
- 教师（职工号，姓名，职称，院系）
- 课程（课程号，课程名，学分，开课院系）
- 选课（学号，课程号，职工号，平时成绩，期中成绩，期末成绩，实验，备注.....）



2. 外模式 (External Schema)

- 外模式 (也称子模式或用户模式)
 - 数据库用户 (包括应用程序员和最终用户) 使用的局部数据的逻辑结构和特征的描述
 - 数据库用户的数据视图, 是与某一应用有关的数据的逻辑表示



外模式（续）

➤ 外模式的地位：介于模式与应用之间

– 模式与外模式的关系：一对多

- 外模式通常是模式的子集
- 一个数据库可以有多个外模式。反映了不同的用户的应用需求、看待数据的方式、对数据保密的要求
- 对模式中同一数据，在外模式中的结构、类型、长度、保密级别等都可以不同

– 外模式与应用的关系：一对多

- 同一外模式也可以为某一用户的多个应用系统所使用，
- 但一个应用程序只能使用一个外模式。



教务系统中的成绩管理

教师端界面

学号	姓名	平时 (30%)	期末 (70%)	总评	备注
----	----	----------	----------	----	----

学生端界面

课程名称	学分	成绩	成绩备注	考试性质	绩点
------	----	----	------	------	----



外模式（续）

➤ 外模式的用途

保证数据库安全性的一个有力措施。

每个用户只能看见和访问所对应的外模式中的数据



设置外模式优点

- 方便用户使用，简化了用户接口
- 保证数据的独立性
- 有利于数据共享
- 有利于数据安全和保护



3. 内模式 (Internal Schema)

➤ 内模式 (也称存储模式)

- 是数据物理结构和存储方式的描述
- 是数据在数据库内部的表示方式
 - 记录的存储方式 (顺序存储, 按照**B**树结构存储, 按**hash**方法存储)
 - 索引的组织方式
 - 数据是否压缩存储
 - 数据是否加密
 - 数据存储记录结构的规定

➤ 一个数据库只有一个内模式



三级模式举例

外模式 (oracle)	外模式 (cobol)
学生 学号 char (6) 院系 char (4)	学生 学号 PIC X(6) 年龄 PIC X(4)
概念模式	S SNO character(6) DEMP character(4) AGE numeric(5)
内模式	STU PREFIX TYPE=BUYES(6) , offset=0 SNO# TYPE=BUYES(6) , offset=6 DEPT# TYPE=BUYES(4) , offset=12 AGE TYPE= FULLWORD, offset=16



1.4 数据库系统的模式结构

- 数据库系统模式的概念
- 数据库系统的三级模式结构
- 数据库的二级映象功能与数据独立性
- 小结

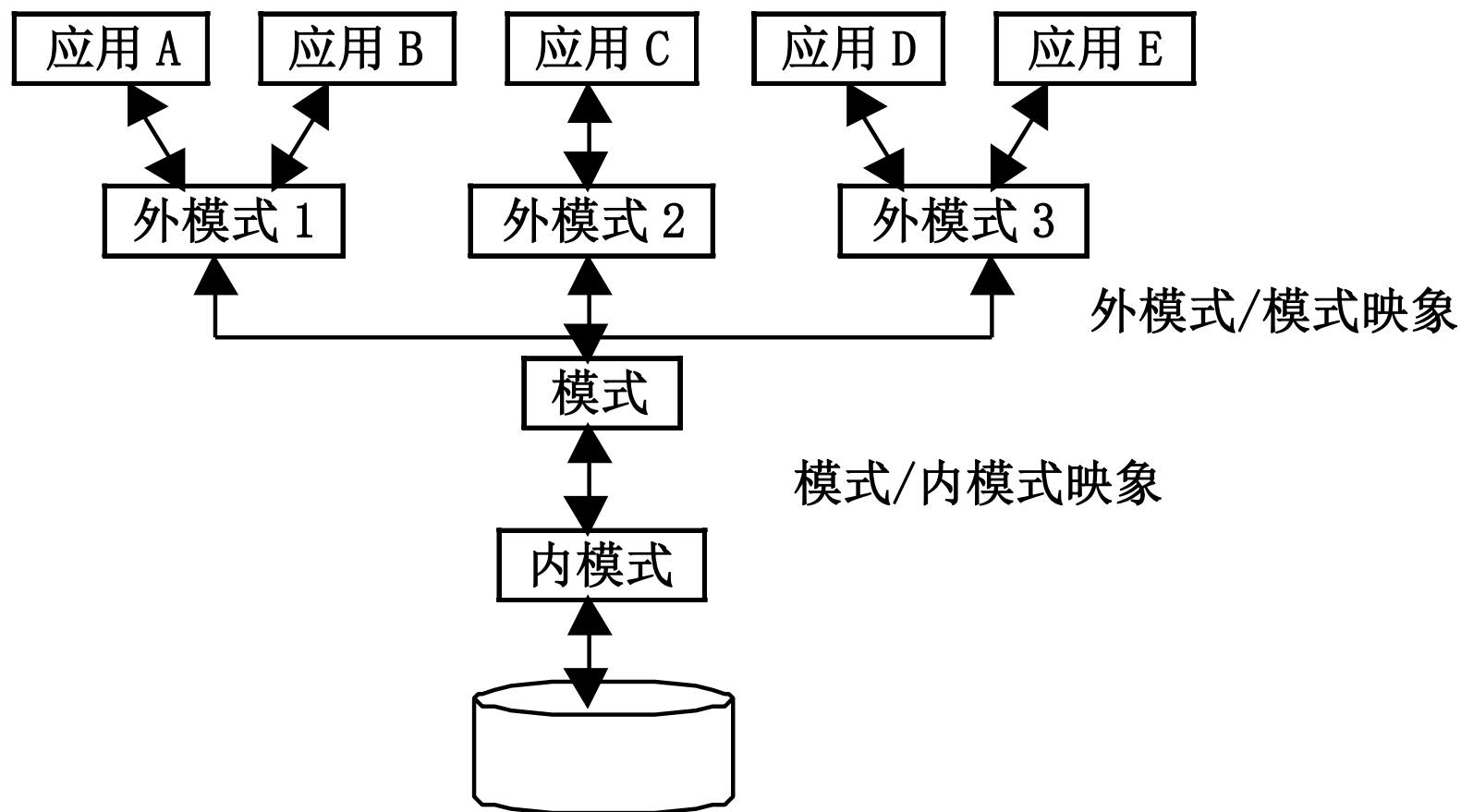


三级模式与二级映像

- 三级模式是对数据的三个抽象级别
- 二级映像在**DBMS**内部实现这三个抽象层次的联系和转换



数据库系统的三级模式结构



1. 外模式 / 模式映像

- 定义外模式与模式之间的对应关系
- 每一个外模式都对应一个外模式 / 模式映像
- 映像定义通常包含在各自外模式的描述中



外模式 / 模式映象的用途

保证数据的逻辑独立性

- 当模式改变时，数据库管理员修改有关的外模式 / 模式映象，使外模式保持不变。
- 应用程序是依据数据的外模式编写的，从而应用程序不必修改，保证了数据与程序的逻辑独立性，简称数据的逻辑独立性。



2. 模式 / 内模式映像

- 模式 / 内模式映像定义了数据全局逻辑结构与存储结构之间的对应关系。例如，说明逻辑记录和字段在内部是如何表示的。
- 数据库中模式 / 内模式映像是**唯一**的
- 该映像定义通常包含在模式描述中



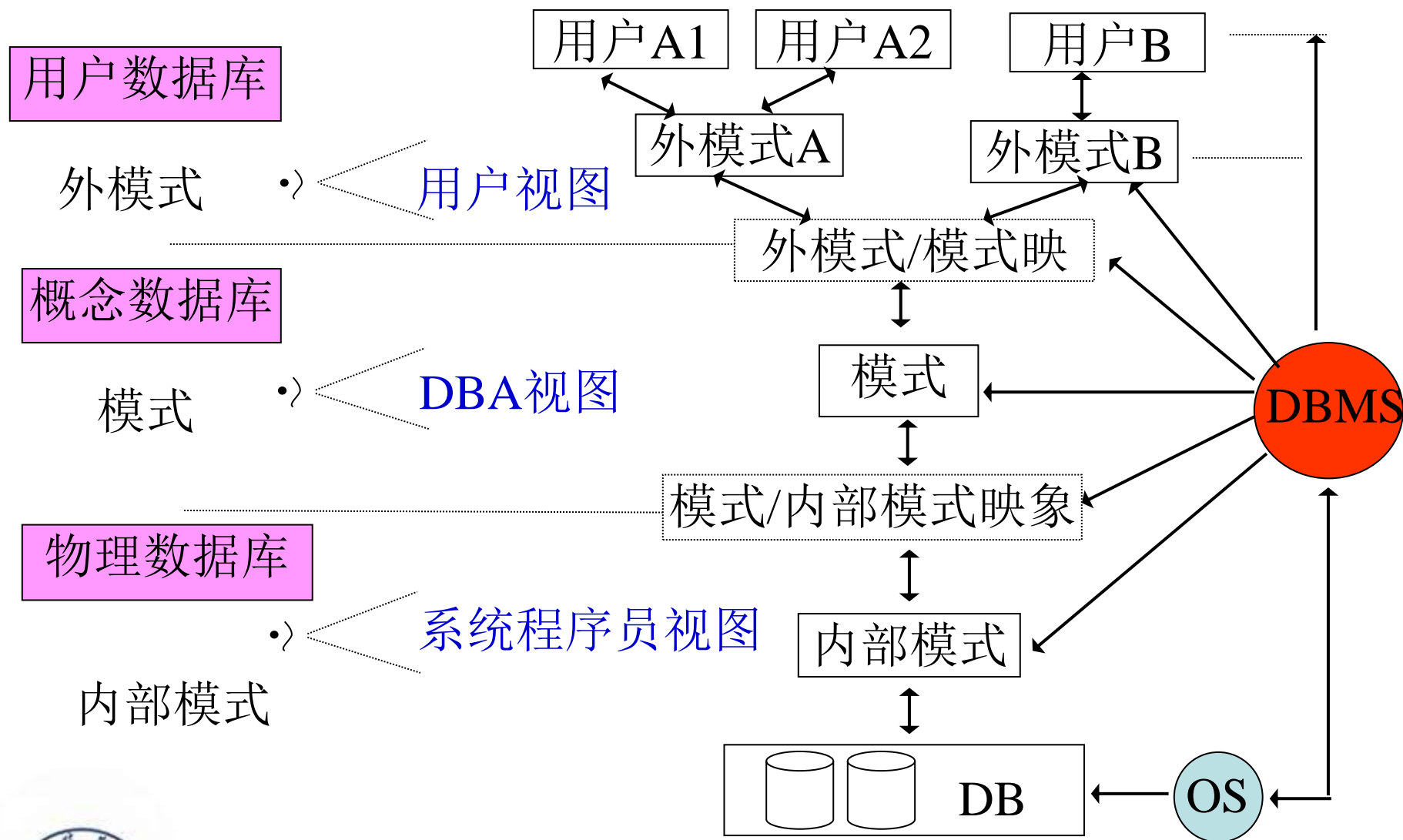
模式 / 内模式映象的用途

保证数据的物理独立性

- 当数据库的存储结构改变了（例如选用了另一种存储结构），数据库管理员修改模式 / 内模式映象，使模式保持不变。
- 应用程序不受影响。保证了数据与程序的物理独立性，简称数据的物理独立性。



数据库的抽象层次



分层抽象的数据库结构

- 模式是独立于数据库其他层次结构的描述
- 内模式依赖于模式，但独立于数据库的用户视图即外模式，也独立于具体的存储设备
- 外模式面向应用程序，定义在逻辑模式之上，独立于内模式和存储设备
- 特定的应用程序是在外模式描述的数据结构上编制的，依赖于特定的外模式，与数据库的模式和存储结构独立
- 关系数据库系统中的视图与外模式相对应
- 数据独立性，使数据的定义和描述可以从应用程序中分离出去



软件框架模式

MVC

控制器

Co

event

视图

View

MVP

管理器

Presenter

视图

View

MVVM

视图模型

View-Model

binding

视图

View

notify

operate

模型

Model

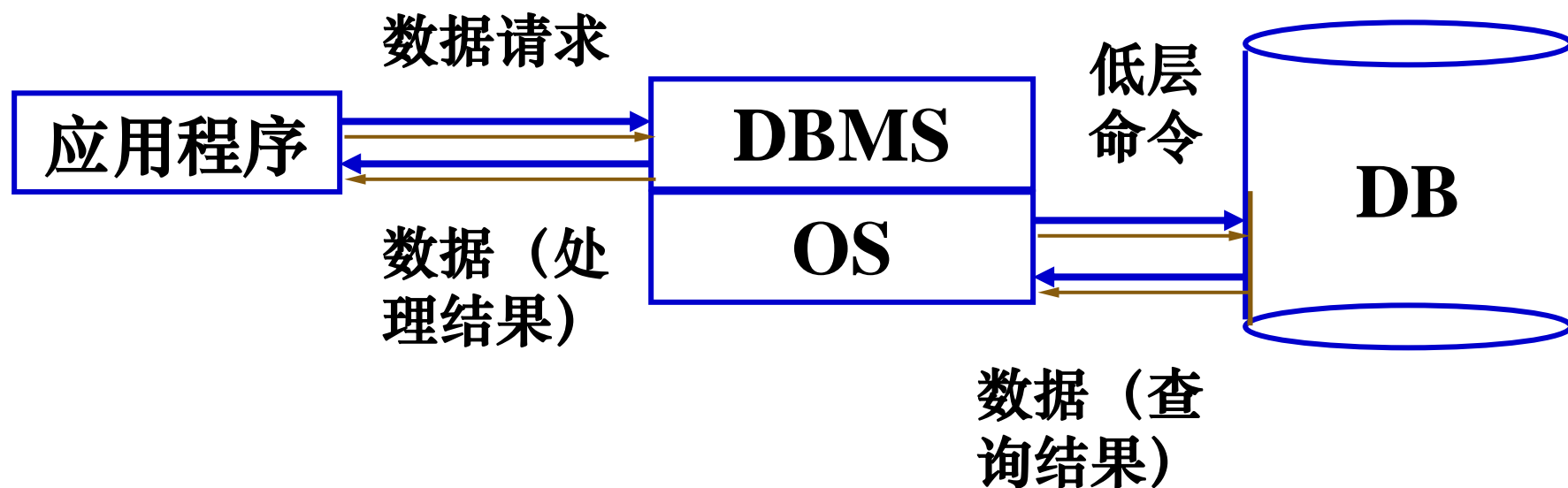


第一章 绪论

- 1.1 数据库、数据库管理系统和数据库系统
- 1.2 数据库系统的产生与发展
- 1.3 数据模型
- 1.4 数据库系统结构
- 1.5 数据库管理系统**
- 1.6 数据库应用系统常见的几种结构



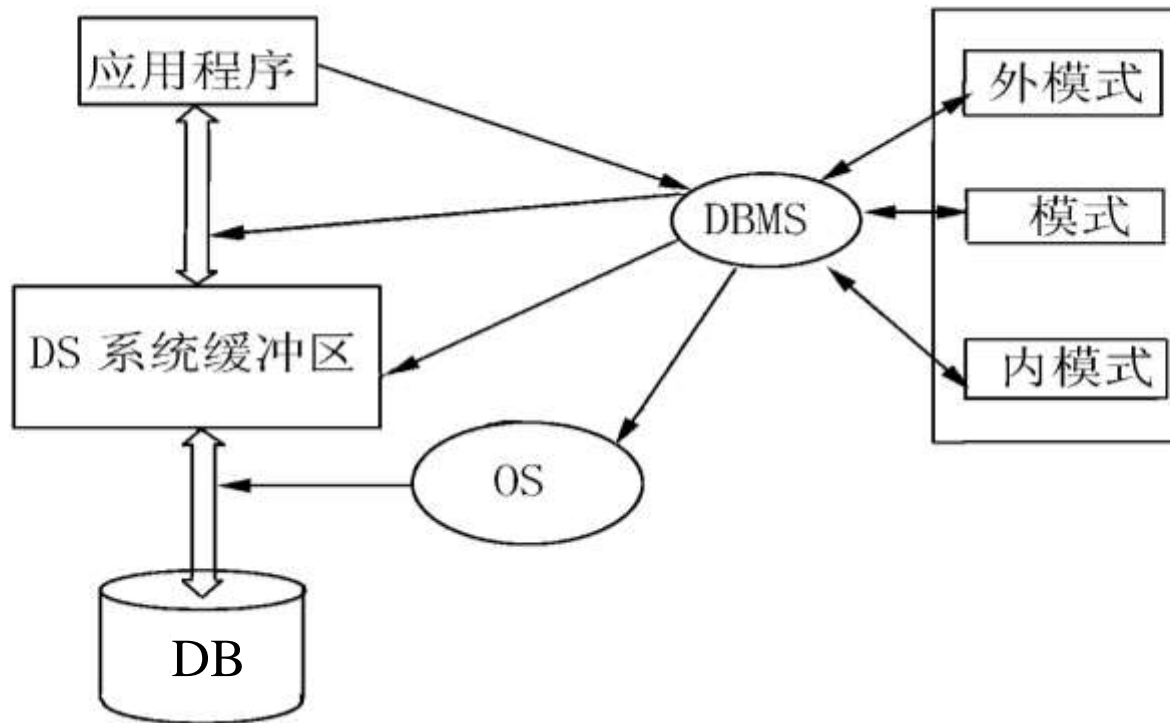
1.5 数据库管理系统



数据库管理系统是位于用户与操作系统之间的一层数据管理软件。它主要功能是建立和维护数据库，接受和完成用户访问数据库的各种请求。



DBMS工作过程



DBMS工作过程

- (1) 接受应用程序的数据请求。
- (2) DBMS对用户的操作请求进行分析
- (3) DBMS向操作系统发出操作请求。
- (4) 操作系统接到命令后，对数据库中的数据进行处理，将结果送到系统缓冲区，并发出读完标志。
- (5) DBMS接到回答信号后，将缓冲区的数据经过模式映射，变成用户的逻辑记录送到用户工作区，同时给用户回答成功与否的信息。

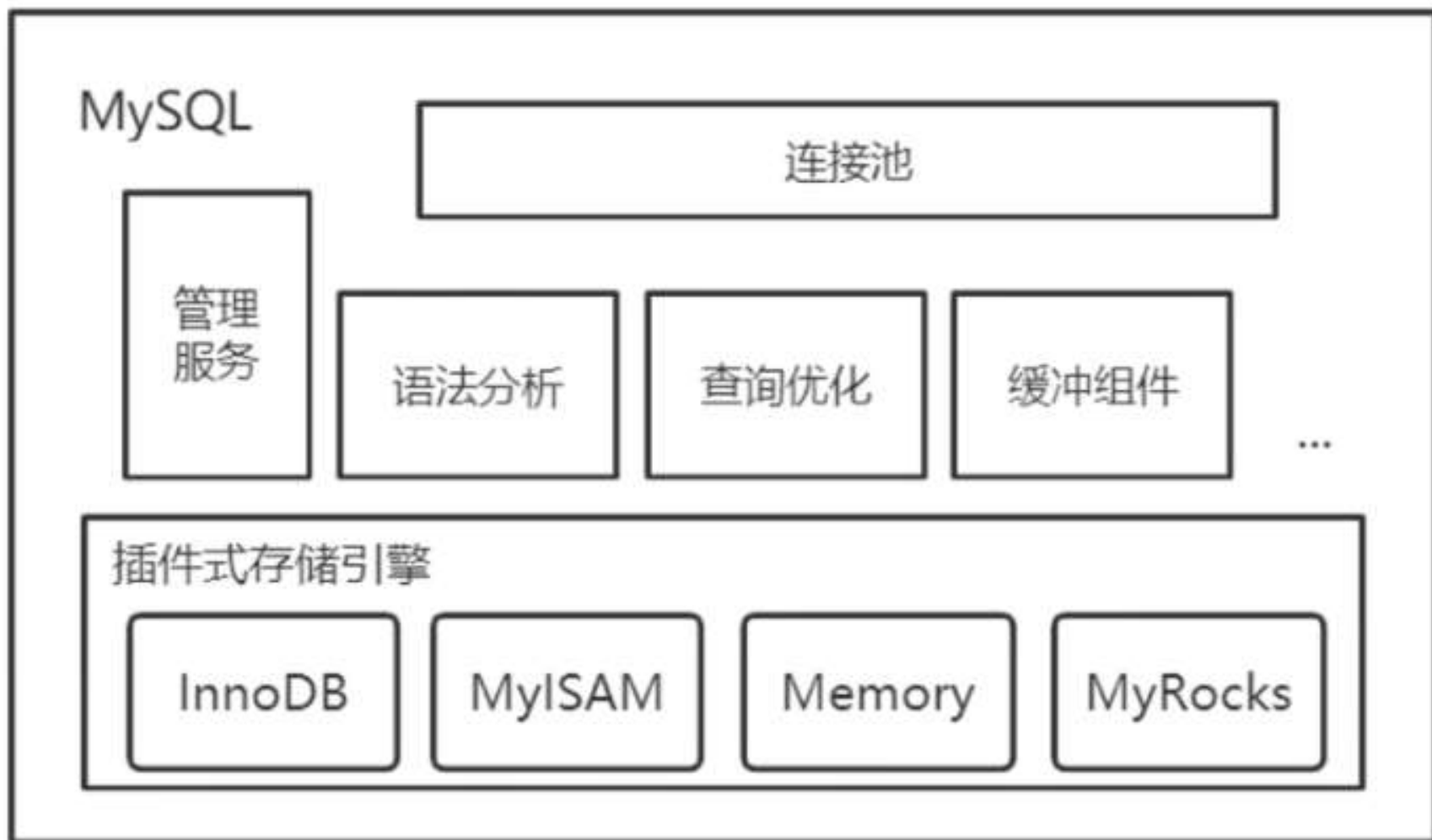


DBMS主要功能

1. 数据定义功能
2. 数据库的建立和维护功能
3. 数据库管理功能
4. 数据组织、存储和管理功能
5. 通信功能



MySQL主要模块





在数据库技术发展的赛道上，中国一直都处于落后地位



自主可控——国产数据库当自强



一旦发生不可预知的危险情况，产品不具备自主可控能力，这对国家信息安全将造成无法挽回的重大损失。



国产数据库的破局之路

自主可控，国产数据库如何破局？

- 核心代码自研，摒弃“拿来主义”
- 另辟蹊径，探索新赛道



自主研发之路

- 信息学院创始人萨师煊老师在1978年将数据库学科引入中国。经过20多年的技术攻关， 1999年，萨师煊、王珊老师创建人大金仓。
- 人大金仓数据库最早起源于中国数据库学科发源地——中国人民大学信息学院。



2009 年，阿里的 Oracle RAC 集群节点数达到了创记录的 20 个。但是 Oracle 并没有弹性扩展的功能，只能按照峰值流量购买需要的小型机和数据库。



2010 年，阿里开始去 IOE。阿里数据库工程师们根据开源 MySQL 搭建了 AliSQL。



2019 年8月，OceanBase 通过TPC-C 测试，打破了Oracle 的垄断地位。



国产数据库排行

墨天轮中国数据库流行度排行

排行	上月	半年前	名称	模型▽	属性▽	三方评测▽	生态▽	专利	论文	得分
	1	1	OceanBase +	关系型	HP 百分 图 图			151	22	701.02
	2	2	TiDB +	关系型	HP 百分 图 图			26	44	647.12
	3	3	openGauss +	关系型	TP 图 图 图 图			562	65	616.85
4	4	4	达梦 +	关系型	TP 图 图 图			381	0	515.97
5	5	↑↑ 7	人大金仓 +	关系型	TP 图 图 图			232	0	495.43
6	6	↓ 5	GaussDB +	关系型	百分 图 HP			562	65	488.04
7	7	↓ 6	PolarDB +	关系型	百分 图 图 HP 图			592	71	469.39
8	↑ 9	↑ 9	GBase +	关系型	TP AP 图 图 图			152	0	307.06
9	↓ 8	↓ 8	TDSQL +	关系型	百分 图 图 HP			39	10	305.10
10	10	10	AnalyticDB +	关系型	AP 图 图 图			480	54	225.70



国产数据库

OceanBase	Relational, Multi-model
TiDB	Relational, Multi-model
OpenGauss	Relational, Multi-model





OceanBase是由蚂蚁集团完全自主研发的国产原生分布式数据库，始创于2010年。已连续10年平稳支撑双11，创新推出“三地五中心”城市级容灾新标准，是一个在TPC-C和TPC-H测试上都刷新了世界纪录的国产原生分布式数据库。





TiDB 是 PingCAP公司自主设计、研发的开源分布式关系型数据库，是一款同时支持在线事务处理与在线分析处理（Hybrid Transactional and Analytical Processing, HTAP）的融合型分布式数据库产品，具备水平扩容或者缩容、金融级高可用、实时 HTAP、云原生的分布式数据库





openGauss

openGauss是一款开源的关系型数据库管理系统，它具有多核高性能、全链路安全性、智能运维等企业级特性。 openGauss内核早期源自开源数据库PostgreSQL，融合了华为在数据库领域多年的内核经验，结合**鲲鹏硬件**优化，在架构、事务、存储引擎、优化器及ARM架构上进行了适配与优化。



我们能够为国产数据库发展做些什么？

- 牢固掌握经典关系数据理论
- 扩展分布式架构、云计算架构的相关知识
- 参与数据库技术相关的各类竞赛
- 参与实际的应用系统开发



第一章 绪论

- 1.1 数据库、数据库管理系统和数据库系统
- 1.2 数据库系统的产生与发展
- 1.3 数据模型
- 1.4 数据库系统结构
- 1.5 数据库管理系统
- 1.6 数据库应用系统常见的几种结构



1.6 数据库系统结构（了解）

- 数据库系统内部的模式结构
 - 从数据库管理系统角度看
- 数据库系统外部的体系结构
 - 从数据库最终用户角度看



1.6 数据库系统外部的体系结构

- 单用户结构
- 集中式结构
- 分布式结构
- 客户/服务器结构
- 浏览器/应用服务器/数据库服务器结构



1. 单用户数据库系统

- 整个数据库系统(应用程序、DBMS、数据)装在一台计算机上，为一个用户独占，不同机器之间不能共享数据。
- 早期的最简单的数据库系统

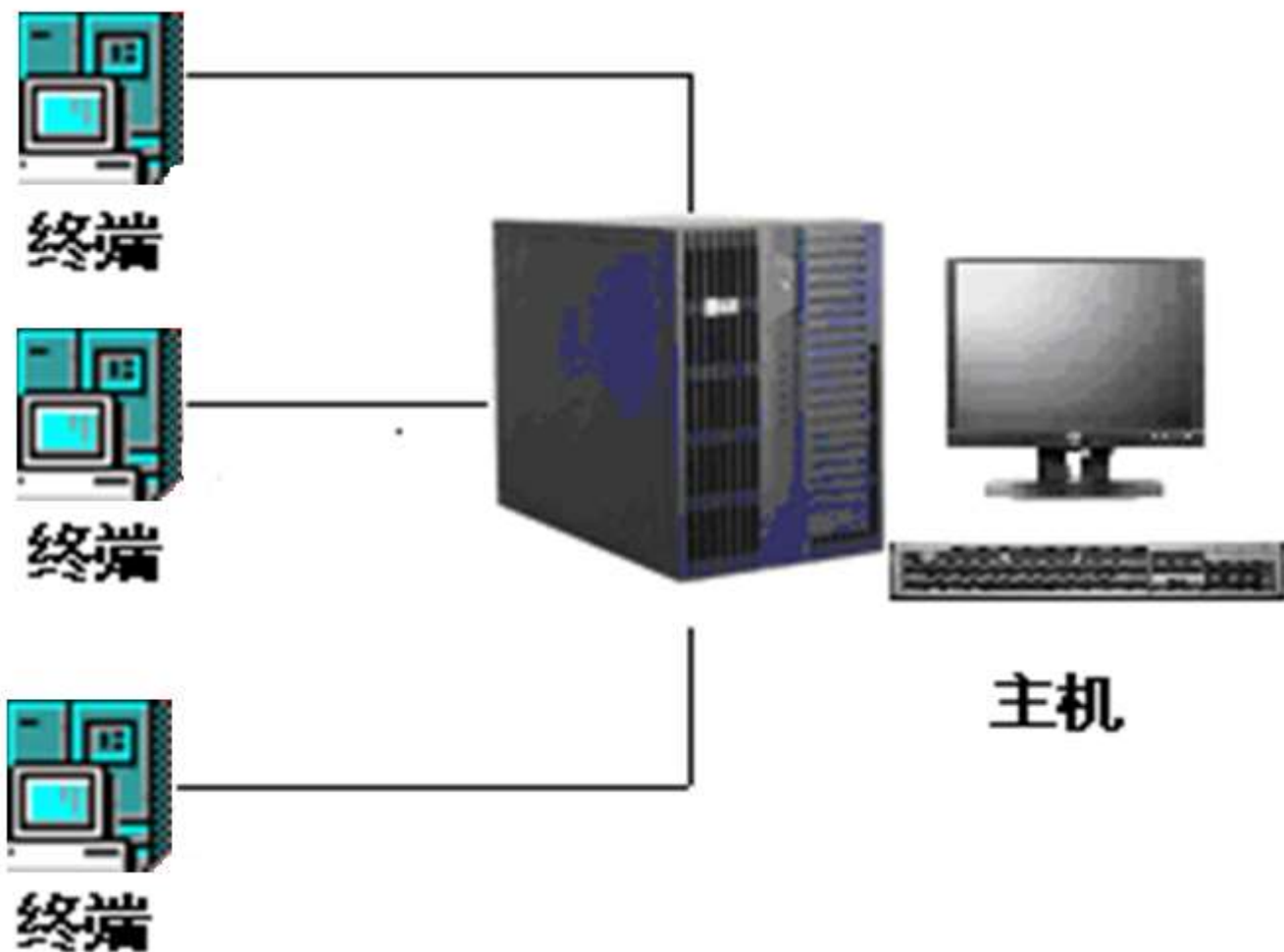


2. 集中式结构的数据库系统

- 一个主机带多个终端的多用户结构
 - 数据库系统，包括应用程序、DBMS、数据，都集中存放在主机上，所有处理任务都由主机来完成
 - 各个用户通过主机的终端并发地存取数据库，共享数据资源



集中式结构的数据库系统



集中式结构的数据库系统(续)

➤ 优点

- 易于管理、控制与维护。

➤ 缺点

- 当终端用户数目增加到一定程度后，主机的任务会过分繁重，成为瓶颈，从而使系统性能下降。
- 系统的可靠性依赖主机，当主机出现故障时，整个系统都不能使用。



3. 客户 / 服务器结构的数据库系统

➤ 把DBMS功能和应用分开

- 网络中某个（些）结点上的计算机专门用于执行DBMS功能，称为数据库服务器，简称服务器
- 其他结点上的计算机安装DBMS的外围应用开发工具，用户的应用系统，称为客户机



客户 / 服务器数据库系统的种类

➤集中的服务器结构

一台数据库服务器，多台客户机

➤分布的服务器结构

- 在网络中有多台数据库服务器
- 分布的服务器结构是客户 / 服务器与
分布式数据库的结合



客户 / 服务器结构的优点

- 客户端的用户请求被传送到数据库服务器，数据库服务器进行处理后，只将结果返回给用户，从而显著减少了数据传输量
- 数据库更加开放
 - 客户与服务器一般都能在多种不同的硬件和软件平台上运行
 - 可以使用不同厂商的数据库应用开发工具



客户 / 服务器结构的缺点

“胖客户”问题：

- 系统安装复杂，工作量大。
- 应用维护困难，难于保密，造成安全性差。
- 相同的应用程序要重复安装在每一台客户机上，从系统总体来看，大大浪费了系统资源。



4. 浏览器/应用服务器/数据库服务器结构

➤ 客户端：

浏览器软件、用户界面

浏览器的界面统一，广大用户容易掌握
大大减少了培训时间与费用。

➤ 服务器端分为两部分：

– **Web**服务器、应用服务器

– 数据库服务器等

大大减少了系统开发和维护代价
能够支持数万甚至更多的用户



5. 分布式结构的数据库系统

- 数据库中的数据在逻辑上是一个整体，但物理地分散在计算机网络的不同结点上。
 - 网络中的每个结点都可以独立处理本地数据库中的数据，执行局部应用
 - 同时也可以同时存取和处理多个异地数据库中的数据，执行全局应用



分布式结构的数据库系统（续）

➤ 优点

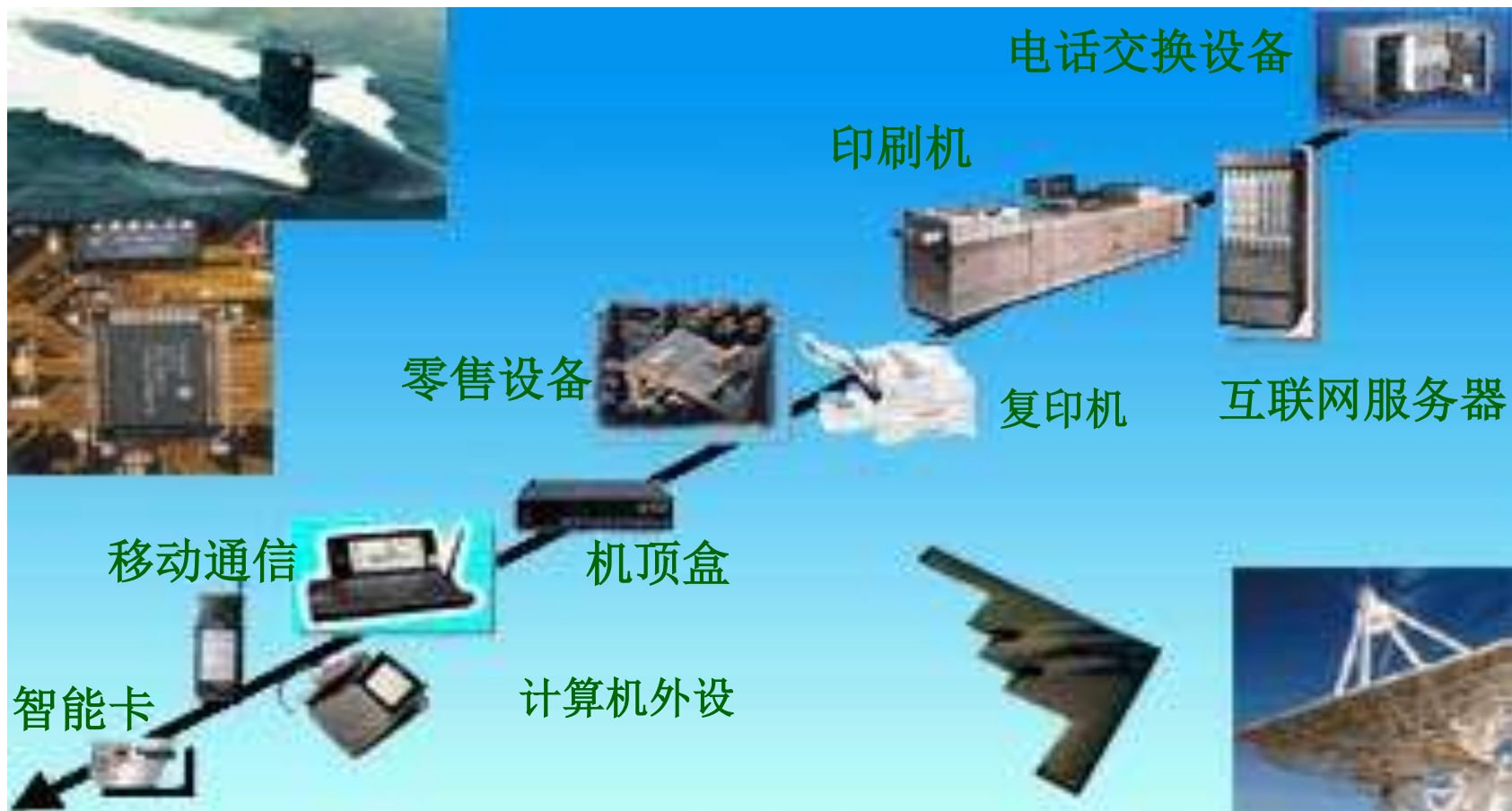
- 适应了地理上分散的公司、团体和组织对于数据库应用的需求。

➤ 缺点

- 数据的分布存放给数据的处理、管理与维护带来困难。
- 当用户需要经常访问远程数据时，系统效率会明显地受到网络传输的制约。



6. 嵌入式数据管理系统的结构



小结

➤ 数据库系统概述

- 数据库的基本概念
- 数据管理的发展过程
- 数据库系统的研究领域

➤ 数据模型

- 数据模型的三要素
- 概念模型： E-R 模型
- 逻辑数据模型： 关系数据模型



小结(续)

➤ 数据库系统的结构

- 数据库系统三级模式结构
- 数据独立性
- 数据库系统的体系结构

➤ 数据库系统的组成

