

数据库系统原理

引用中国人民大学信息学院原版PPT

华中科技大学计算机学院左琼修改版

School of Computer Science and Technology , HUST
2020



第一章 绪论

1.1 数据库系统概述

1.2 数据模型

1.3 数据库系统的结构

1.4 数据库系统的组成

1.5 小结

1.3 数据库系统的结构

■ 数据库中数据如何存储？

数据库不仅存放数据的值，也需存放数据的语义。

数据的语义是通过存储数据的结构来实现的，库中数据有**值**与**型**两部分，分别存放在**数据文件**和**字典**中。

■ 数据模型中“型”和“值”的区别：

- **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**记录型**： (学号，姓名，性别，系别，年龄，籍贯)

一个**记录值**： (900201，李明，男，计算机，22，江苏)

1.3.1 数据库系统模式的概念

- 从数据库应用开发人员角度看，数据库系统通常采用**三级模式**结构，是数据库系统**内部**的系统结构。
- **模式 (Schema)**
 - 数据库中全体数据的逻辑结构和特征的描述。
 - 仅仅是**型**的描述，与具体的值无关。
 - 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；
 - 实例随数据库中数据的更新而**变动**。

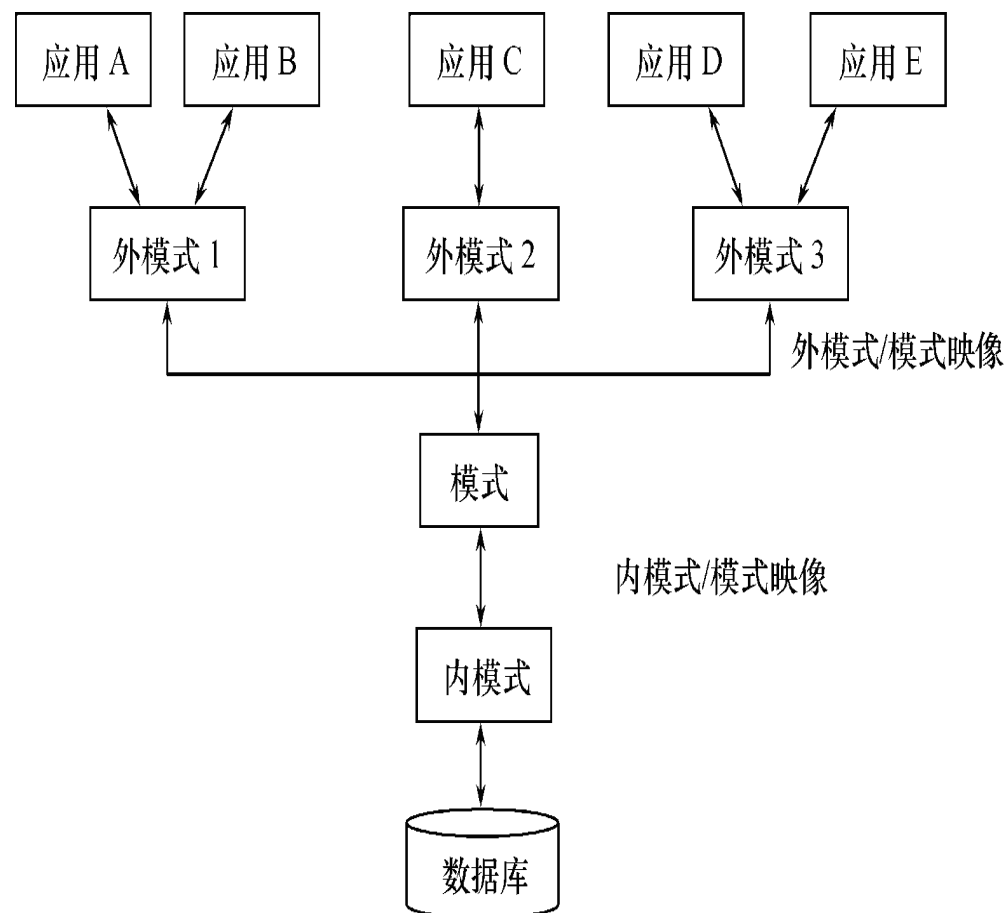
1.3.2数据库系统的三级模式

■数据库系统三级模式结构

- CODASYL (Conference On Data System Language, 美国数据系统语言协会) 提出**模式、外模式、内模式**三级模式的概念, 三级模式之间有**两级映像**。

■模式的分级

- 从数据库用户的观点, 即**用户看到的数据库**, 与数据库的物理方面, 即**实际存储的数据库**区分开来, 数据库系统的模式是分级的。
- **提高了数据的逻辑独立性和物理独立性。**



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——全局逻辑视图**, 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**, **独立于**数据库的其他层次。
 - 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。

1. 模式 (Schema)

■ 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。

■ 模式一般由多个“记录”组成，包含数据库的所有信息。

如：某课程选修信息系统的数据库模式包括下列记录：

学生（学号，姓名，性别，系别，年龄）

课程（课程号，课程名）

选修（学号，课程号，成绩）

- 设计数据库模式结构时应**首先**确定数据库的逻辑模式。
- 模式的作用是为了支持数据的**少冗余共享**。

2. 外模式(external schema)

- 外模式（也称**子模式**subschema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**，面向具体的应用程序，是与某一应用有关的数据的逻辑表示；
 - 定义在逻辑模式之上；
 - 独立于存储模式和存储设备；
 - 当应用需求发生较大变化，相应外模式不能满足其视图要求时，该外模式就得做相应改动；
 - 设计外模式时应充分考虑到应用的扩充性。

2. 外模式

- 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。

1.3.2 数据库的三级模式结构

外模式



劳资科

学号	姓名	系别	补贴
----	----	----	----



房产科

学号	姓名	性别	系别	住址
----	----	----	----	----



学籍科

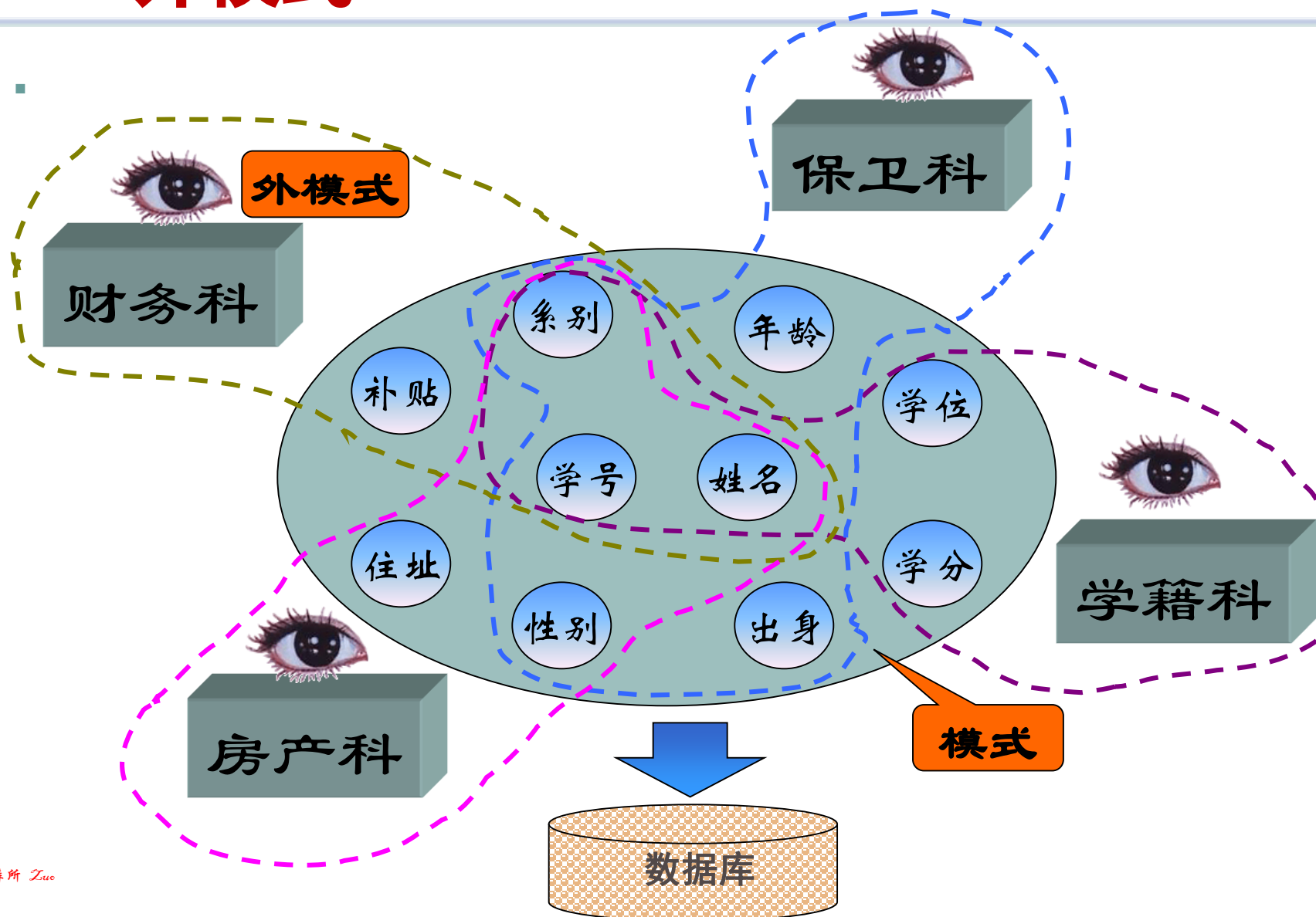
学号	姓名	系别	学分	学位
----	----	----	----	----



人事科

学号	姓名	性别	系别	年龄	学位	出身
----	----	----	----	----	----	----

模式 vs. 外模式



模式 vs. 外模式

又如，数据库**模式**如下：

学生（学号，姓名，性别，系别，年龄）

课程（课程号，课程名）

选修（学号，课程号，成绩）

可定义**子模式**如下：

①**单**关系子模式：student_1（学号，姓名，系别）

②**多**关系子模式：SC（学号，姓名，课程号，课程名，成绩）

■ 外模式的用途：

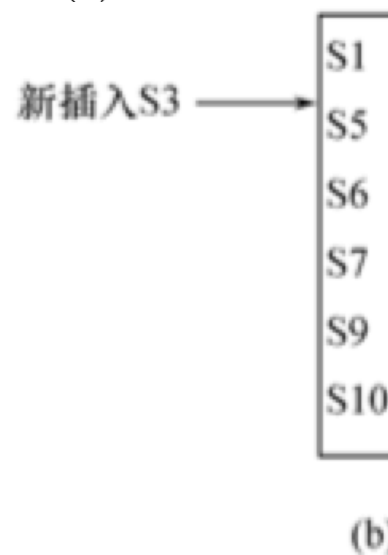
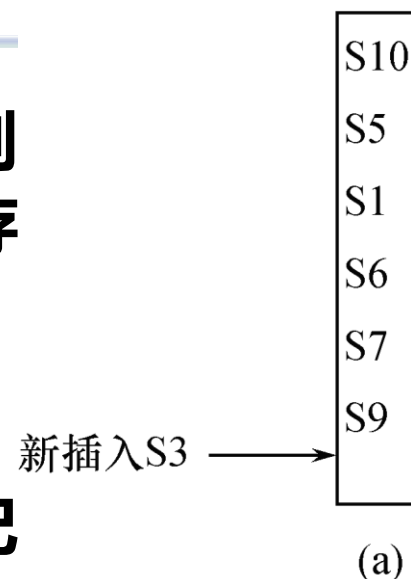
- 1) 支持不同用户建立适应局部应用特征的结构，**每个用户只能看见和访问所对应的外模式中的数据**；
- 2) 简化应用处理；
- 3) 提高安全性，**保证数据库安全性的一个有力措施**。

3. 内模式 (Internal schema)

- 内模式（也称**存储模式storage schema**或**物理模式**）：
 - 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，例：
 - 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。

3. 内模式

- 例如：学生记录，如果按**堆**存储，则插入一条新记录总是放在学生记录存储的**最后**，如图(a)所示。
- 如果按**学号升序**存储，则插入一条记录就要找到它应在的位置插入，如图(b)所示。



3. 内模式

- 如果按照学生**年龄聚簇**存放，假如新插入的S3是16岁，则应插入的位置如图(c)所示。

内模式的作用是支持用户建立适应需求（如存取效率、空间效率、数据安全）的物理结构。

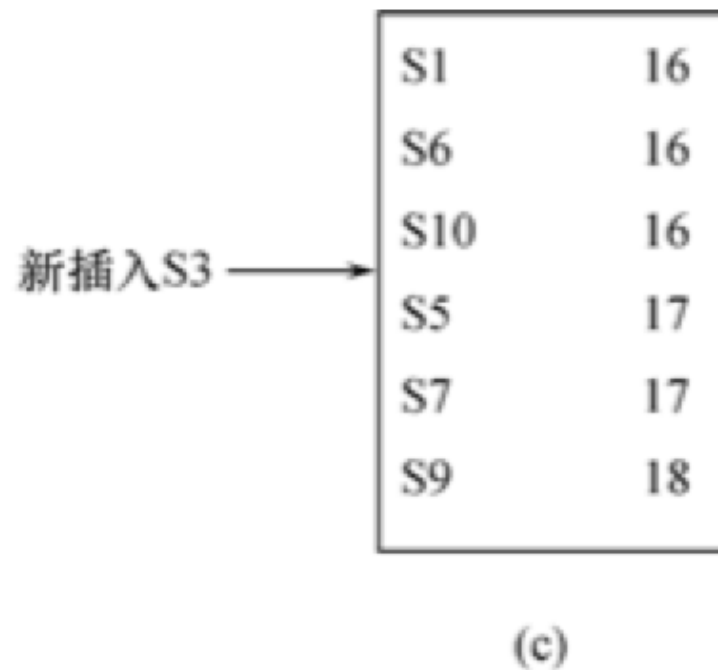


图 记录不同的存储方式示意图

内模式依赖于它的全局逻辑结构；
独立于数据库的用户视图，即外模式。

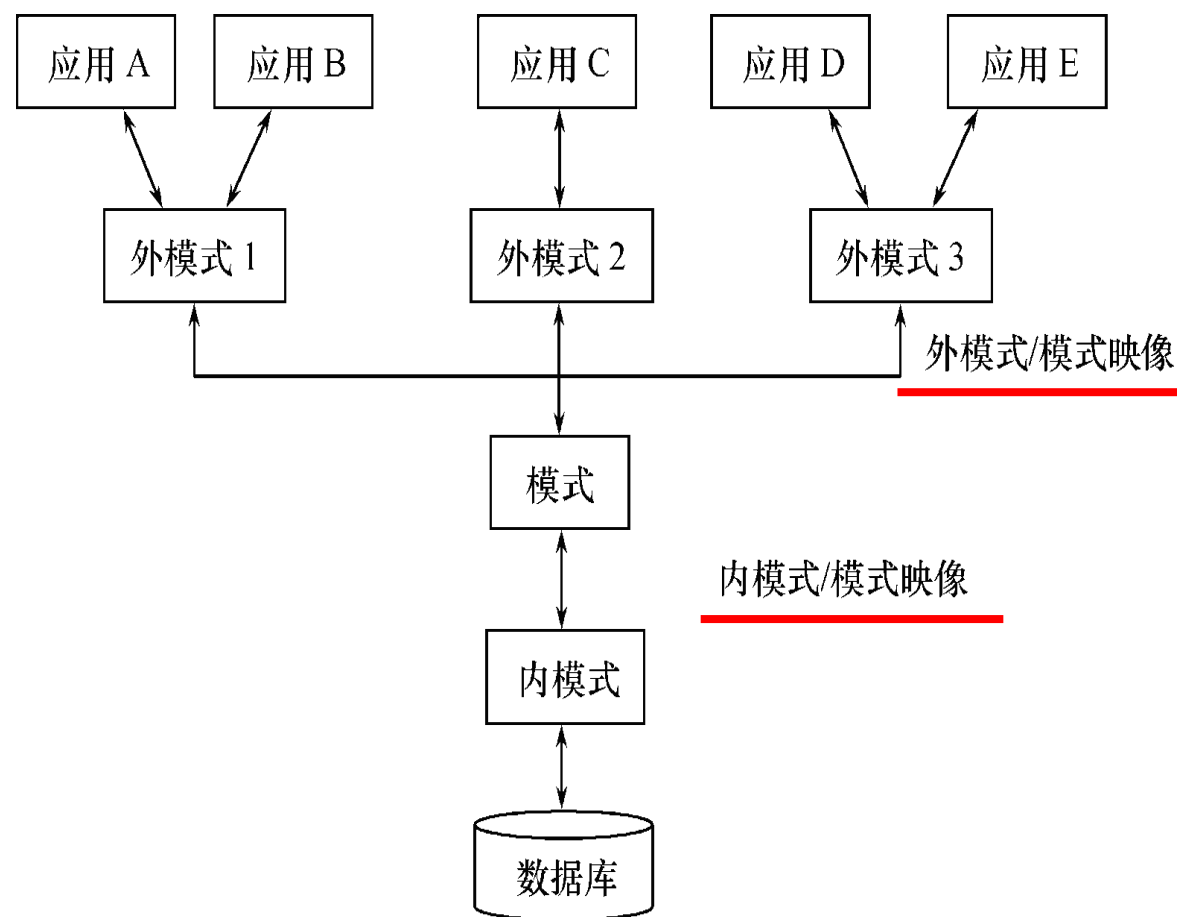
1.3.2 数据库系统的三级模式

■ 三级模式小结：

- 内模式是整个数据库实际存储的表示，反映的是数据的**存储观**；
- 概念模式是整个数据库的抽象表示，反映的是数据的**全局观**；
- 外模式是概念模式的某一部分的抽象表示，反映的是数据的**用户观**。

1.3.3 数据库的二级映像与数据独立性

- 三级模式是对数据的三个抽象级别;
- 二级映像是在DBMS内部实现这三个抽象层次的联系和转换:
 - 外模式 / 模式映像
 - 模式 / 内模式映像



1. 外模式 / 模式映象

- 外模式/模式映像定义**某一个外模式和模式之间的对应关系。**
- 作用：**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。**

1. 外模式 / 模式映象

student

NO	NAME	AGE	SEX
301	赵钱	20	男
302	钱	21	女



Create View Stud(学号,姓名,性别,年龄)
As Select NO, NAME, AGE, SEX,
From student



stud

学号	姓名	性别	年龄
301	赵钱	男	20
302	钱	女	21



Create View Stud(学号,姓名,性别,年龄)
As Select XH,XM,XB,
datediff(year,CSRQ,getdate())
From student

student

XH	XM	XB	CSRQ
301	赵钱	男	95.07.01
302	钱	女	94.03.07



2. 模式 / 内模式映像

- 模式 / 内模式映象定义了**数据全局逻辑结构与存储结构之间的对应关系**。
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 / 内模式映象是**唯一的**。
- 该映象定义通常包含在**模式描述**中。
- **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；
 - 存储结构变化的影响被限制在模式之下，这使数据的存储结构和存储方法独立于应用程序应用程序不受影响。保证了数据与程序的物理独立性，简称数据的物理独立性。

三级模式结构的优点

- **保证数据的独立性**

- 将模式和内模式分开，保证了数据的物理独立性；
- 将外模式和模式分开，保证了数据的逻辑独立性。

- **简化了用户接口**

按照外模式编写应用程序或敲入命令，而不需了解数据库内部的存储结构，方便用户使用系统。

- **有利于数据共享**

在不同的外模式下可有多多个用户共享系统中数据，减少了数据冗余。

- **利于数据的安全保密**

在外模式下根据要求进行操作，不能对限定的数据操作，保证了其他数据的安全。



第一章 绪论

1.1 数据库系统概述

1.2 数据模型

1.3 数据库系统结构

1.4 数据库系统的组成

1.5 小结

1.4 数据库系统的组成

1. 硬件平台及数据库
2. 软件
3. 人员

1.硬件平台及数据库

■ 数据库系统对硬件资源的要求:

(1) 足够大的内存

操作系统

DBMS的核心模块

数据缓冲区

应用程序

(2) 足够大的外存

磁盘或磁盘阵列

➤ 数据库

光盘、磁带

➤ 数据备份

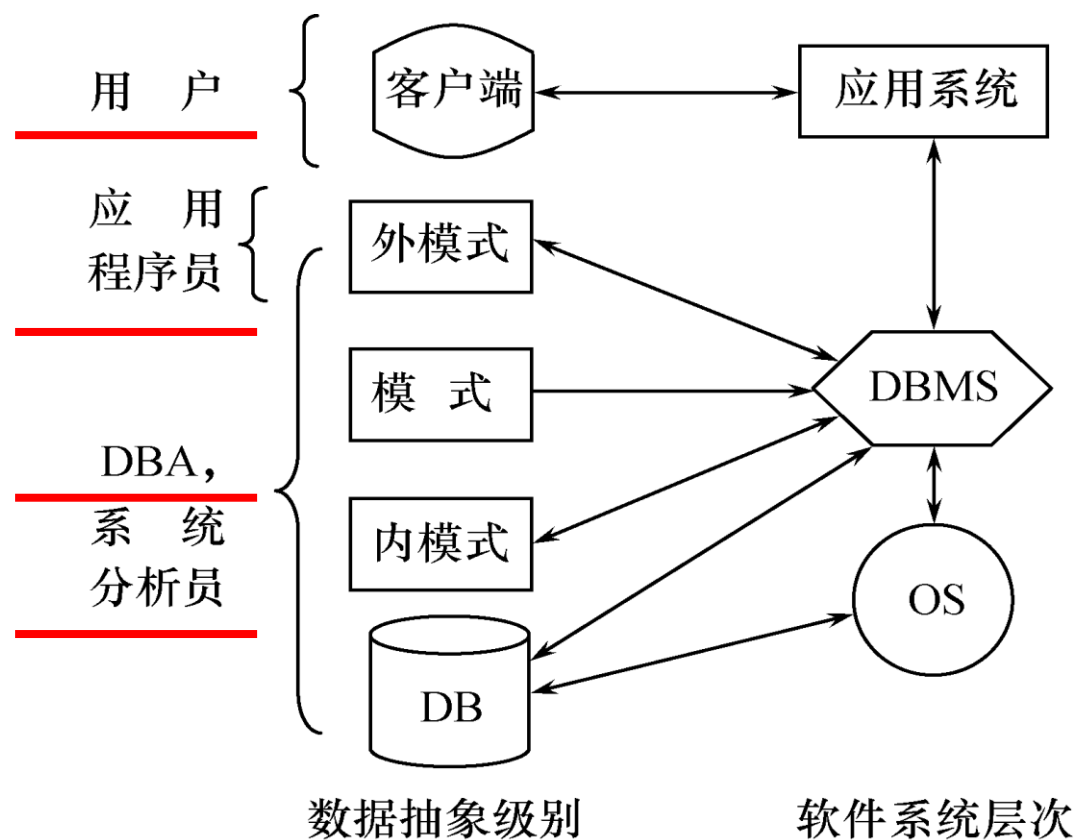
(3) 较高的通道?能力, 提高数据传送率

2. 软件

- 1) DBMS
- 2) 支持DBMS运行的操作系统
- 3) 与数据库接口的高级语言及其编译系统
- 4) 以DBMS为核心的应用开发工具
- 5) 为特定应用环境开发的数据库应用系统

3. 人员

- 不同的人员涉及不同的数据抽象级别，具有不同的数据视图，如图。



3. 人员:数据库管理员(DBA)

1) **DBA**: 负责全面管理和控制数据库系统。

具体职责:

- (1) 决定数据库中的信息内容和结构——DB设计;
- (2) 决定数据库的存储结构和存取策略——内模式设计;
- (3) 定义数据的安全性要求和完整性约束条件;
- (4) 监控数据库的使用和运行:
 - 周期性转储数据库: 数据文件、日志文件;
 - 系统故障恢复; 介质故障恢复;
 - 监视审计文件。
- (5) 数据库的改进和重组
 - 性能监控和调优
 - 定期对数据库进行重组织, 以提高系统的性能
 - 需求增加和改变时, 数据库须需要重构造

3. 人员

2) 系统分析员

- 负责应用系统的需求分析和规范说明
- 与用户及DBA协商，确定系统的硬软件配置
- 参与数据库系统的概要设计

3) 数据库设计人员

- 参加用户需求调查和系统分析
- 确定数据库中的数据
- 设计数据库各级模式

4) 应用程序员

- 设计和编写应用系统的程序模块
- 进行调试和安装

5) 用户是指最终用户（End User）。

- 最终用户通过应用系统的用户接口使用数据库。

3.人员：用户

1) 偶然用户

- 不经常访问数据库，但每次访问数据库时往往需要不同的数据库信息
- 企业或组织机构的高中级管理人员

2) 简单用户

- 主要工作是查询和更新数据库
- 银行的职员、机票预定人员、旅馆总台服务员

3) 复杂用户

- 工程师、科学家、经济学家、科技工作者等
- 直接使用数据库语言访问数据库，甚至能够基于数据库管理系统的API编制自己的应用程序

1.5 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

■ 数据模型

- 数据模型的三要素
- 概念模型, **E-R 模型**
- 三种主要数据库模型 (了解各种模型的主要特性)

■ 数据库系统的结构

- 数据库系统**三级模式**结构
- 数据库系统**两层映像**系统结构

■ 数据库系统的组成

作业:

P34 3, 17

(请在超星平台提交作业)

回顾本章主要内容

- 什么是数据库系统?
- 产生DBMS的动机是什么?
- 如何设计数据库?
- 数据库系统的组成?
- 数据库系统的结构?

数据库系统是由DB、DBMS、应用系统、DBA组成的存储、管理、处理和维护数据的系统。

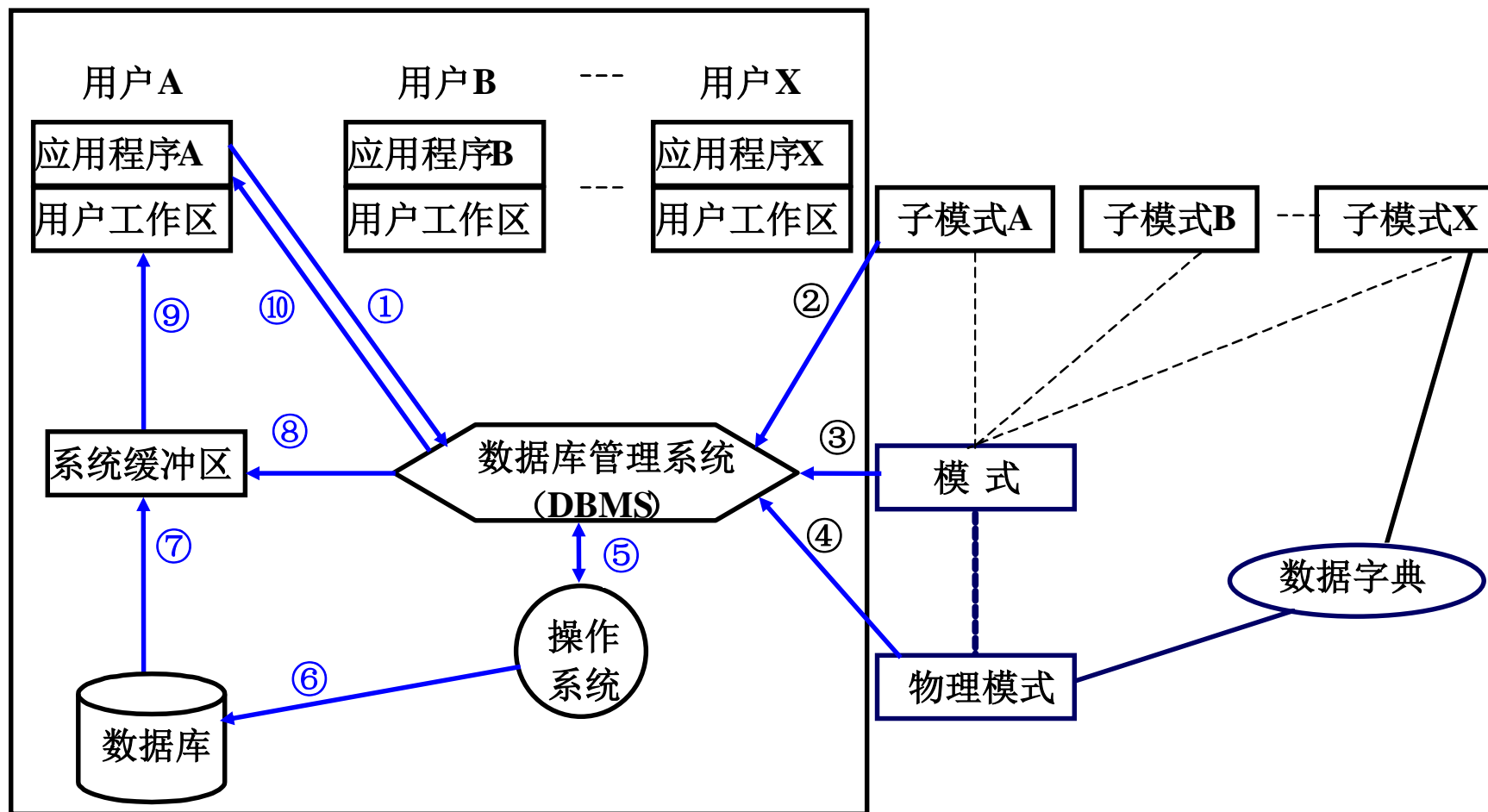
P7 1.1

数据模型，ER图

硬件平台+DB，软件，人员

三级模式，二级映像

补：DBMS的工作流程



DBMS工作过程

在数据库系统中，当一个应用程序或用户需要存取数据库中的数据时，应用程序、DBMS、操作系统、硬件等几个方面必须协同工作，共同完成用户的请求。

应用程序从数据库读取一个数据通常需要以下步骤：

1. 应用程序A向DBMS发出从数据库中读数据记录的命令；
2. DBMS对该命令进行语法检查、语义检查，并调用应用程序A对应的子模式，检查A的存取权限，决定是否执行命令，如果拒绝执行，则向用户返回错误信息；
3. 在决定执行该命令后，DBMS调用模式，依据子模式/模式映象的定义，确定应读入模式中的哪些记录；

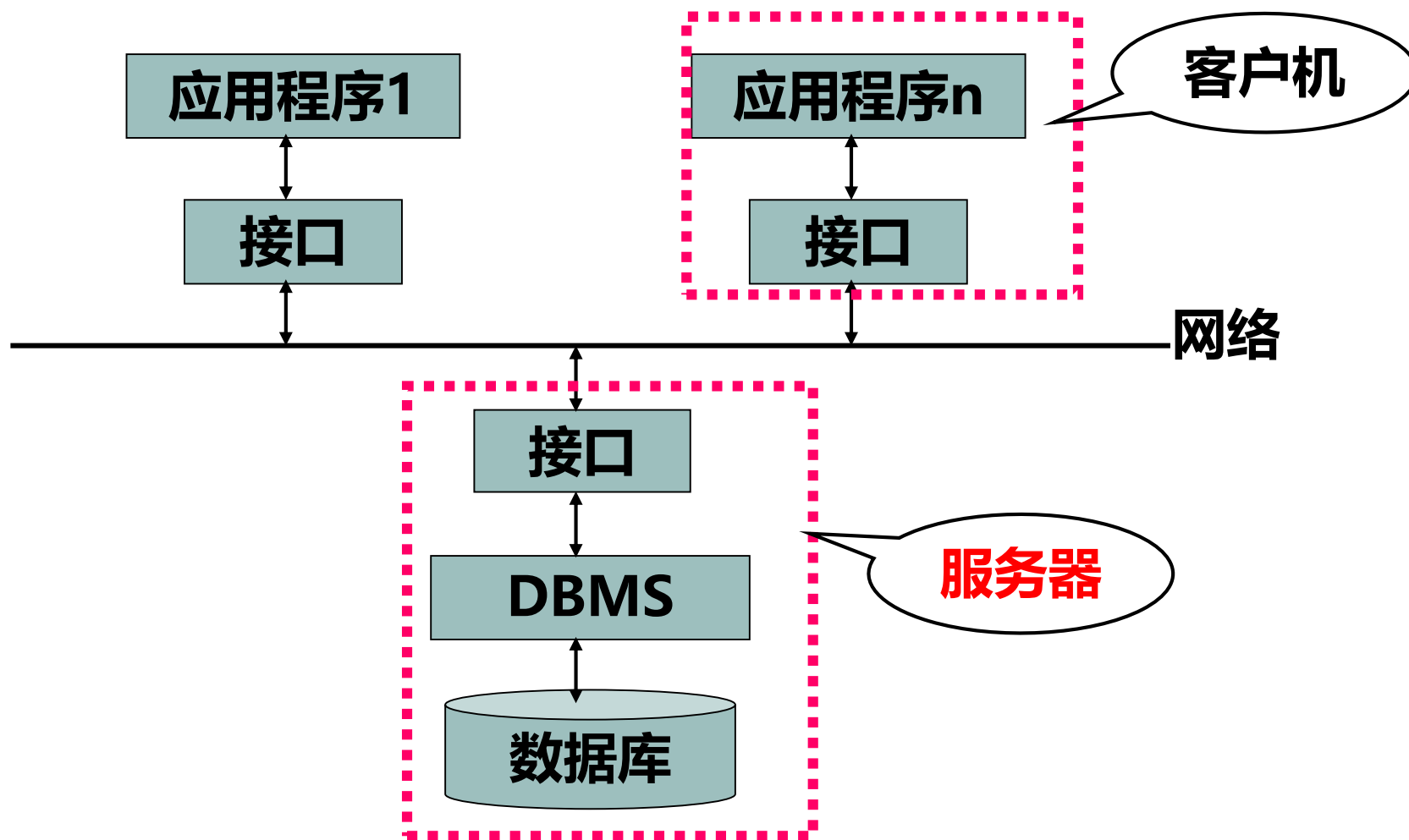
DBMS工作过程 (续)

4. DBMS调用物理模式，依据模式/物理模式映象的定义，决定从哪个文件、用什么存取方式、读入哪个或哪些物理记录；
5. DBMS向操作系统发出执行读取所需物理记录的命令；
6. 操作系统执行读数据的有关操作；
7. 操作系统将数据从数据库的存储区送到系统缓冲区；
8. DBMS依据子模式/模式映象的定义，导出应用程序A所要读取记录的格式；
9. DBMS将数据记录从系统缓冲区传送到应用程序A的用户工作区；
10. DBMS向应用程序返回命令执行情况的状态信息。

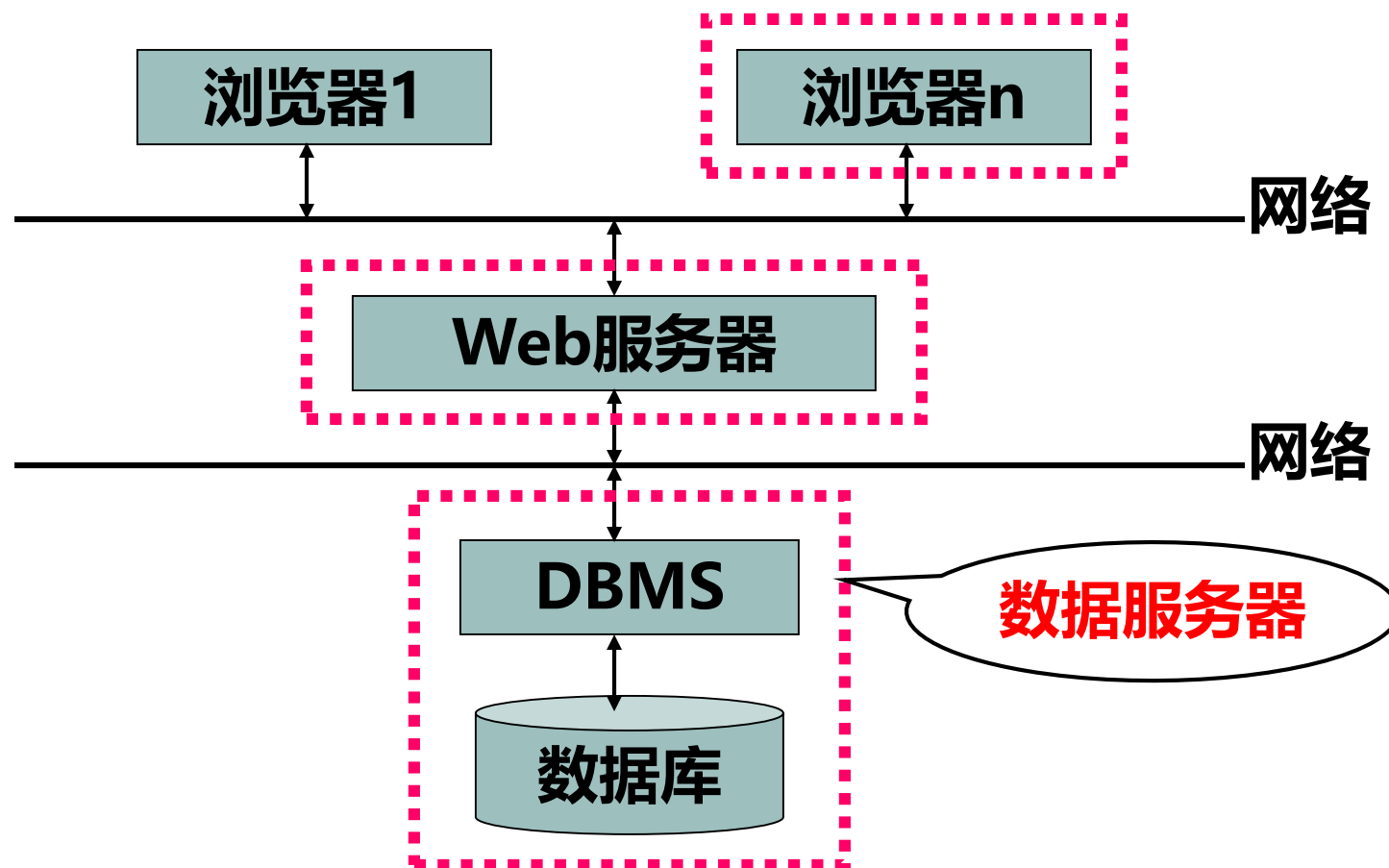
PS. 数据库系统的体系结构

- 从数据库最终用户角度看（数据库系统外部的体系结构），数据库系统的体系结构分为：
 - 单用户结构
 - 客户 / 服务器（Client/Server，简称C/S模式）
 - 浏览器 / 应用服务器 / 数据库服务器多层结构（Browser/Server，简称B/S模式）
 - 主从式结构
 - 分布式结构
 - 并行数据库系统

客户/服务器模式的数据库系统结构



浏览器/服务器模式的数据库系统结构



分布式数据库应用

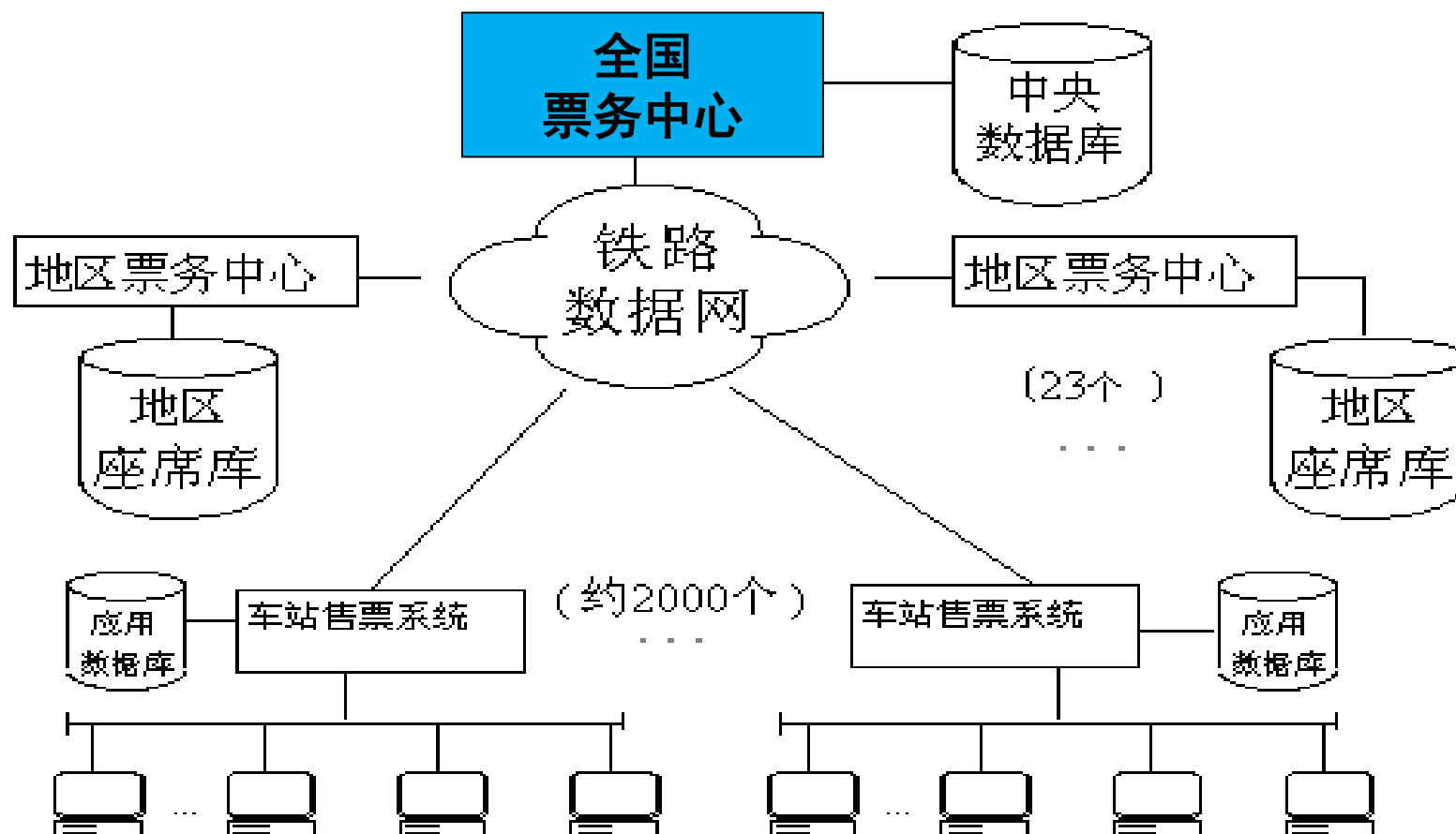
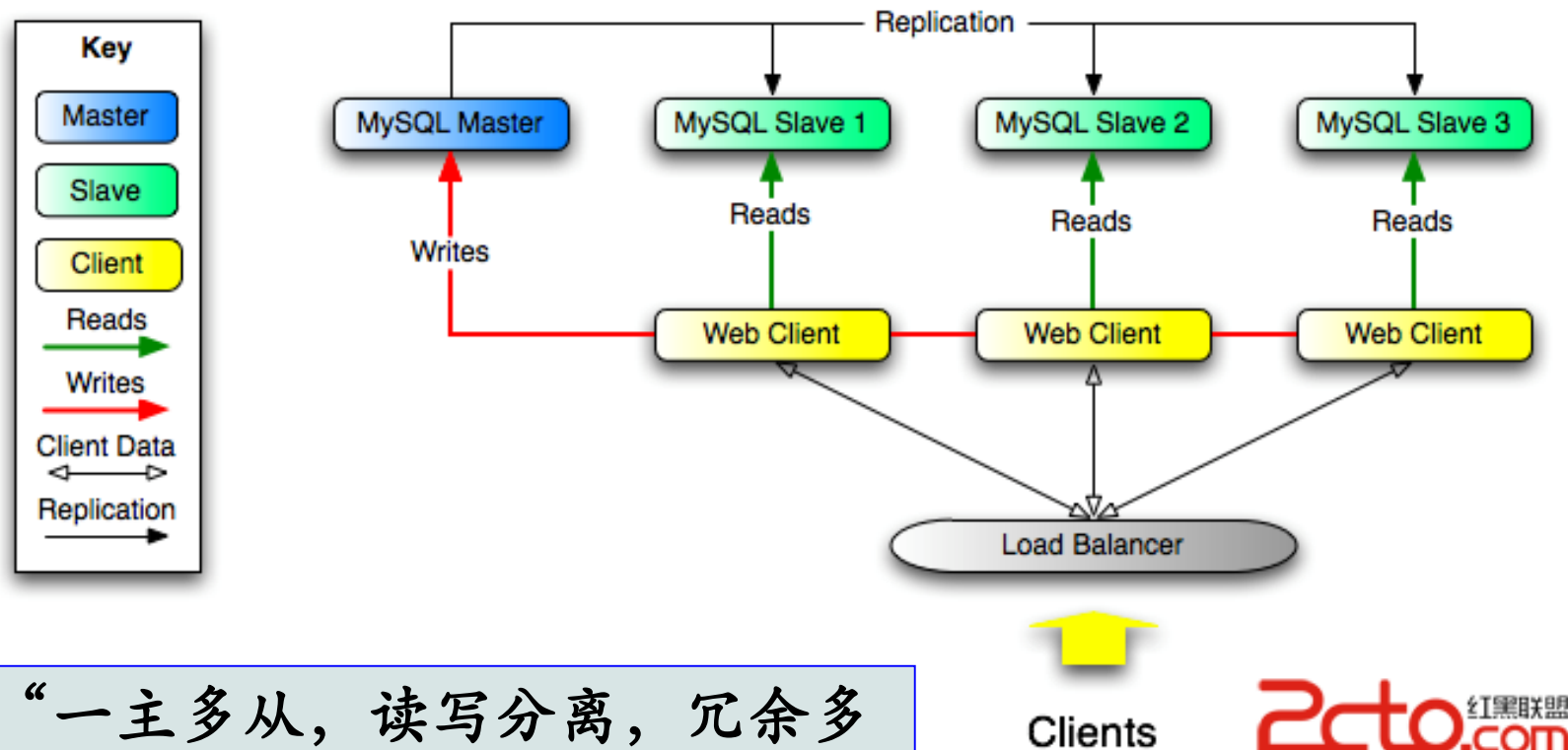


图 一个分布式数据库应用系统实例

MySQL的主从架构



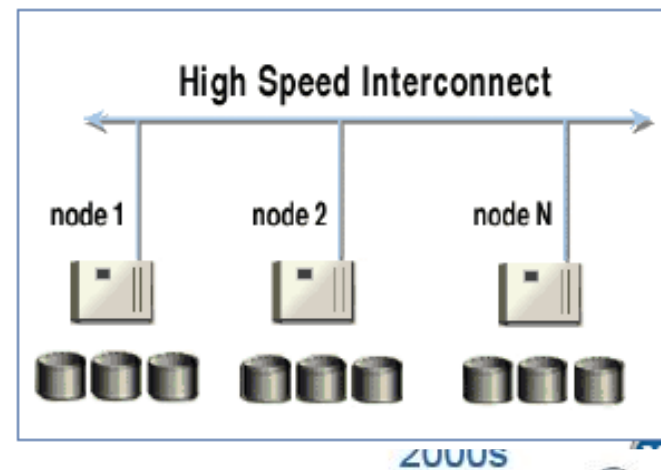
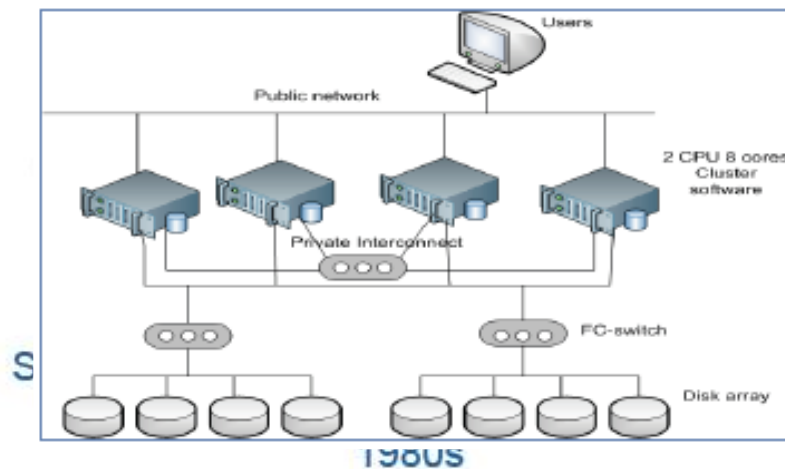
“一主多从，读写分离，冗余多个读库”的数据库架构来提升数据库的读性能。

2cto.com 红黑联盟

RDBMS进展-for scalability

- 单节点设备到多节点机器
- DB 共享存储
- 数据仓库应用和列存储数据库
- 并行数据库
- 分布式数据库
- 数据库非共享集群

主要用于：
mission critical ,
high transaction
OLTP systems



分布式数据库

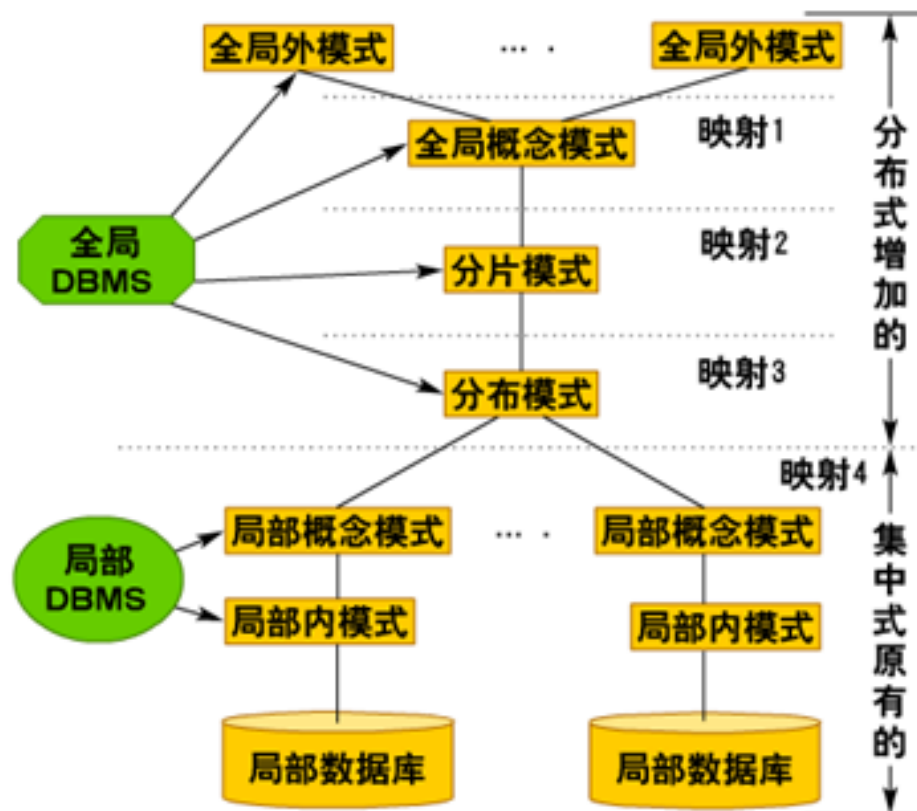
■ 分布式数据库

是地理上分布在计算机网络的不同结点，逻辑上属于同一系统的数据库系统。

■ 分布式数据库的主要特点是：

- 数据是分布的
- 数据是逻辑相关的
- 结点的自治性
- 支持分布式查询
- 支持分布式事务

DM分布式数据库

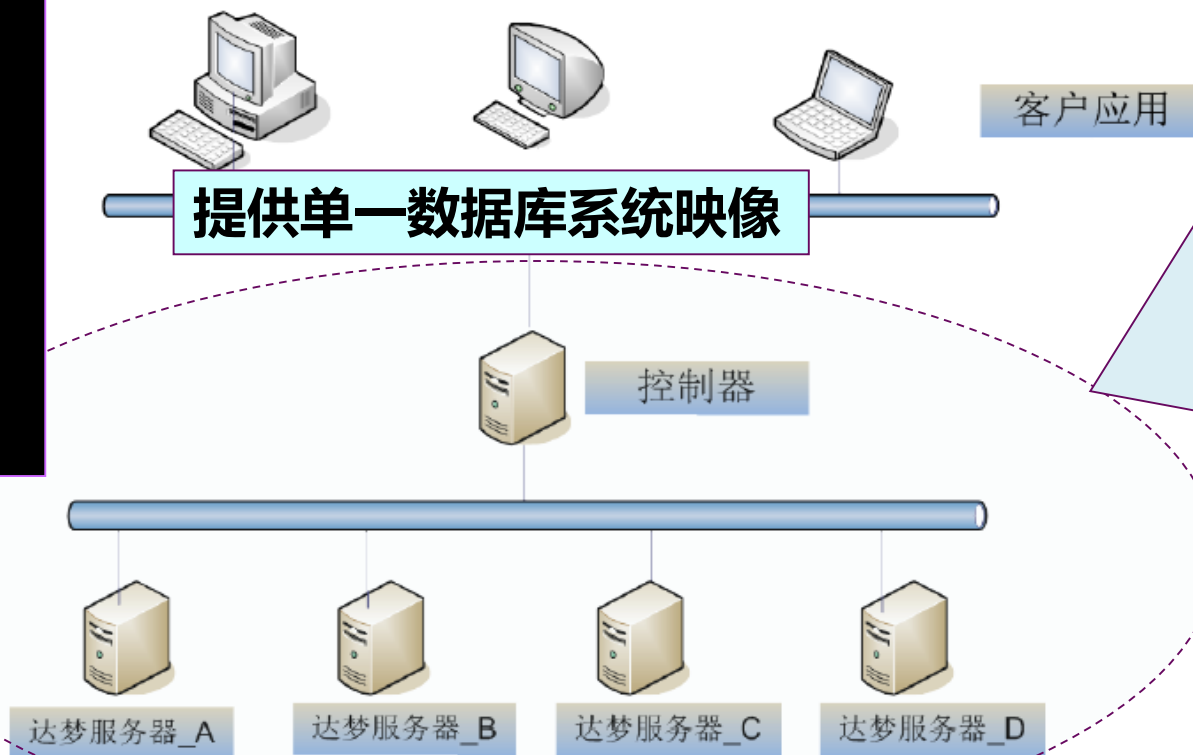


分布式数据库系统的模式结构

DM非共享集群——对称无共享节点HA

关键技术:

- 集群管理
- 负载均衡
- 心跳检测
- 数据同步复制
- 自动故障转移



- 1) 实现集群服务器节点的查询负载均衡;
- 2) 管理服务器节点加入集群和退出集群;
- 3) 处理服务器节点出现故障的情况、容灾;
- 4) 保持服务器节点的数据的一致性。

图 15.2.1 DM 数据库集群体系结构

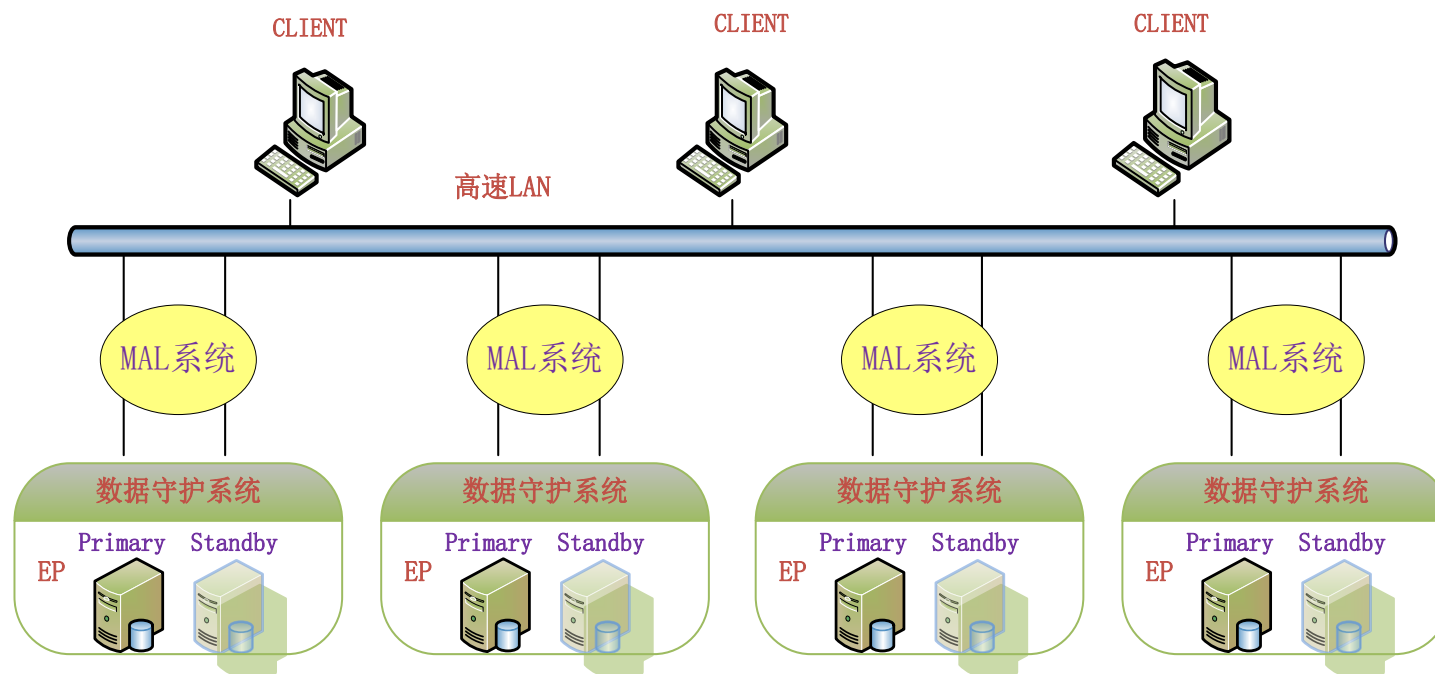
能有效保证系统的高可用性

DM7 MPP集群

DM MPP体系架构:

- 完全对等无共享
- 客户端: 可连接任一节点 (EP)
- 内部通信: 高速邮件子系统
- 负载均衡: 按列HASH散列划分数据, 存储和查询自动化并行
- 容灾: 每EP可配置数据守护系统 (主备)

提供基于数据划分的高性能大规模并行集群功能, 多个节点能够并行处理同一个请求, 提升系统性能, 并且通过节点间的守护技术来保证系统中的任意节点出现故障时, 仍然能够提供不间断的服务。



集成完全不共享架构优点, 无需专用硬件, 无单点问题

Chapter 1 ends。 。 。

追求

休息一会儿。。。

