



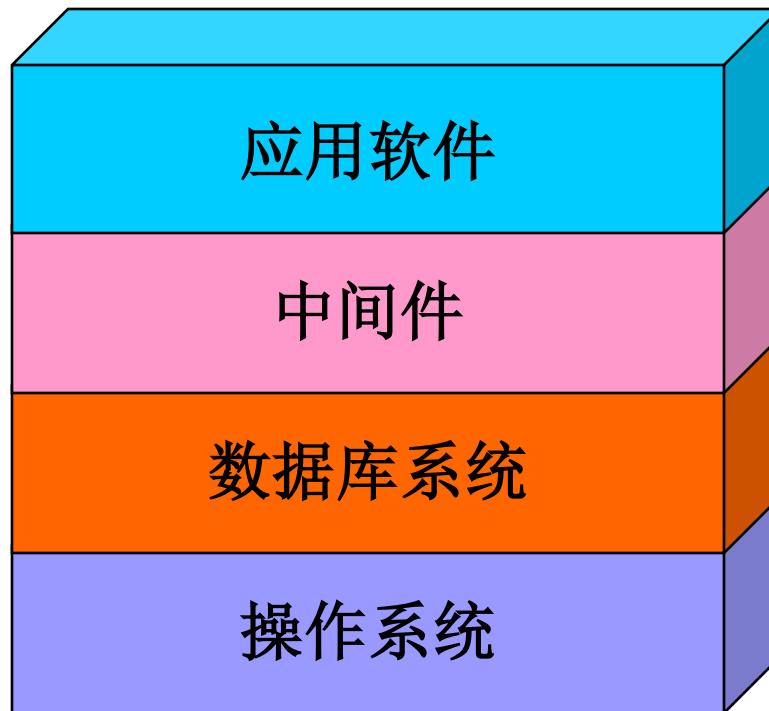
软件与数据组织

西安电子科技大学计算机学院—School
of Computer Science and Technology,
Xidian University, China

讲授内容

- 软件结构
- 操作系统
- 程序设计
- 数据组织
- 数据库

软件结构



系统软件与应用软件

■ 系统软件用于扩展系统(包括硬件)的功能

- 操作系统

- 支撑软件

 - 设备驱动程序

 - 语言翻译程序

 - ...

■ 应用软件用来解决用户的特定问题



操作系统

西安电子科技大学计算机学院—School
of Computer Science and Technology,
Xidian University, China

操作系统的发展过程

■ 无操作系统时代

计算机资源是**以时间为单位整体划分**给用户的。早期的输入设备主要是**纸带机**、**卡片机**和**磁带机**。

将**程序和运行程序**所需的数据表示在输入设备上，再把表示在纸带上的程序和数据输入到计算机内存，然后运行程序得到输出结果。

用户在划分给自己的时间内，对计算机的所有硬件资源是**全部占有**的，当用户通过输入设备输入程序和数据时，计算机的其余硬件资源都处于空闲状态。

操作系统的发展过程

■ 单道批处理系统

单道批处理系统的核心是一个专门管理和调度用户作业的、独立于用户程序的程序，这个程序被称为**监督程序**。

监督程序是管理和调度用户作业的。**作业**是指用户要求计算机完成的工作，包括完成用户某个任务的程序和必要的数据。

此时出现了专门负责管理计算机以及帮助用户完成作业的称为**操作员**的人员。所有用户需要把自己的程序和数据都交给操作员，由操作员负责执行这些用户程序。对作业一般采用**队列**进行管理！

串行

操作系统的发展过程

■ 多道批处理系统

由于输入/输出设备的速度相对于CPU的速度慢很多，而计算机在输入/输出数据时CPU是空闲的，为解决这一问题，提高计算机的使用效率，在20世纪60年代中期又出现了多道批处理系统。

多道批处理系统的核心是调度程序，调度程序把若干个作业从作业队列中调入内存，并将CPU资源分配给一个作业，让它开始运行。若当前正处理的作业要进行输入/输出操作，就应释放对CPU的占有权。调度程序则从其他作业中重新选择一个运行。

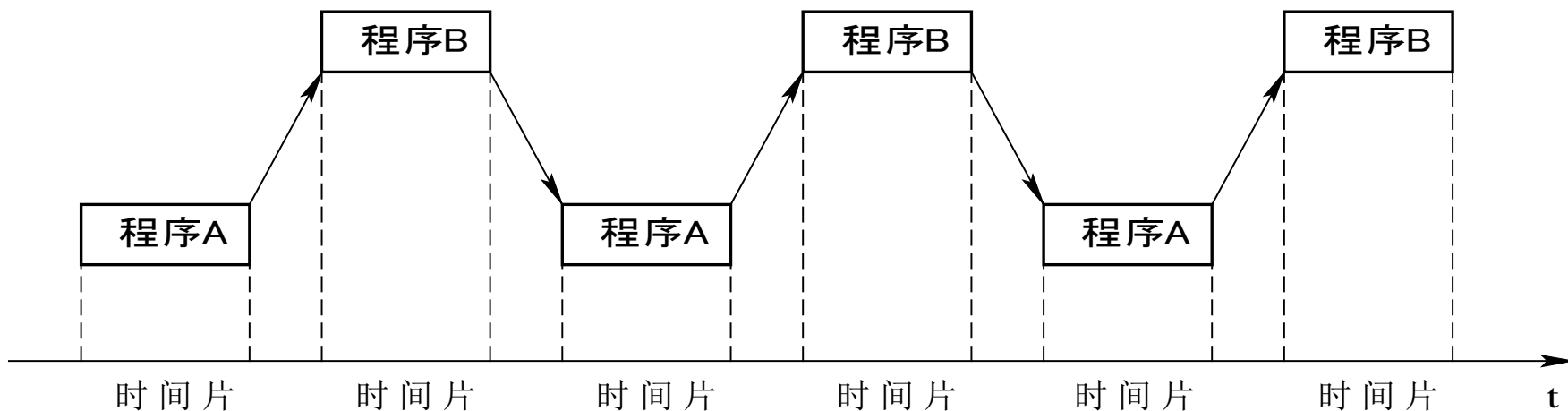
优点：在作业程序进行输入/输出操作时CPU也不会闲置。

操作系统的发展过程

■ 分时系统

□ 分时系统的核心是把CPU的使用时间划分成非常短的时间片，多个用户的程序可**同时驻留在内存**，当轮到某个用户程序使用CPU时，该程序只能在限定的时间片内运行。当用户的时间片用完时，操作系统就暂停该用户程序的运行，并按某种调度策略调出内存中的另一个用户程序开始运行。**干预计算任务**

□ 优点：方便人机交互频繁的程序



操作系统的新发展

■ 实时系统

- 满足实时控制和实时信息处理的需求

■ 通用操作系统

■ 多处理机操作系统

■ 网络操作系统

- 网络用户与计算机网络之间的接口

■ 分布式操作系统

- 实现信息交流和资源共享，使多个用户协同工作

推动操作系统发展的因素

- 提高计算机资源利用率的需要
- 方便用户使用的需要
- 不断扩大的新的应用方式和应用领域的需要
- 硬件技术不断发展的需要

如：多操作系统和高并发系统的要求，促进了虚拟化软件的发展：
vmware workstatoin、oracle virtualbox、Microsoft的hyper-v、
linux kvm、Citrix的开源xen

操作系统的基本功能

- 管理计算机中的**资源**；
- 提供**用户界面**，以方便各种人使用；
- 提供**系统功能调用**（System functions call）支持，以方便应用程序的开发；
- 加载（Load）并运行（Run）**应用程序**（可同时运行多个）：
 - 确定应用程序在外存或网络上的**位置（地址）**，在内存中分配相应的空间，把应用程序和必要的**数据**读入该空间；
 - 从应用程序的**第一条机器指令**（程序入口，Program entry）开始执行它；
 - 按照既定的**调度策略**（如时间片、优先级等），暂时中止运行中的程序及继续执行它，执行结束后释放内存空间。

操作系统的核心功能

- 存储管理

 - 管理内存

- CPU管理

 - 程序/任务调度

- 设备管理

- 文件管理

- 用户接口

操作系统的启动

- 开机后操作系统是怎样自动装入内存？
- 任何计算机中都必须包括一个称作**自举程序**的程序，自举程序的功能是：首先把操作系统从磁盘装入内存，然后启动操作系统开始运行。
 - ROM类型的内存单元在制造时就把自举程序固化在其中。
- 操作系统的启动过程分为两步：
 - (1) 计算机开机时，硬件设计成自动执行ROM中的自举程序，自举程序把操作系统从磁盘装入内存；
 - (2) 计算机开始运行操作系统。

PC自举过程

冷启动

加电自检测试

查找装有OS的磁盘驱动器

A 盘上有吗？

是

软盘启动

否

C 盘上有吗？

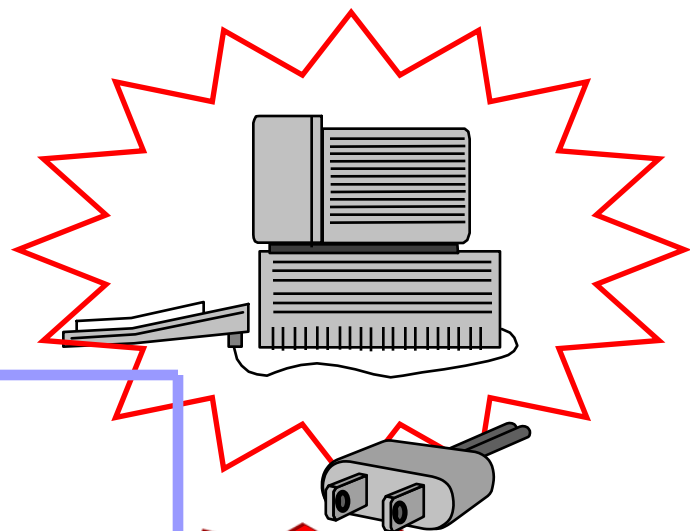
是

硬盘启动

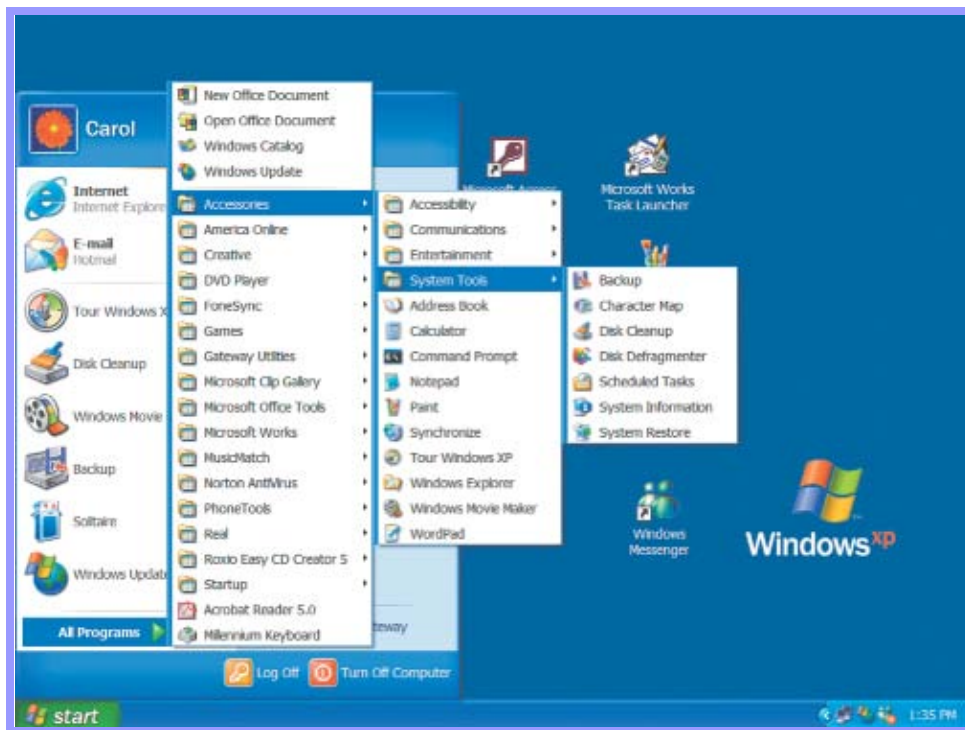
系统总清

热启动

Ctrl+Alt+Del

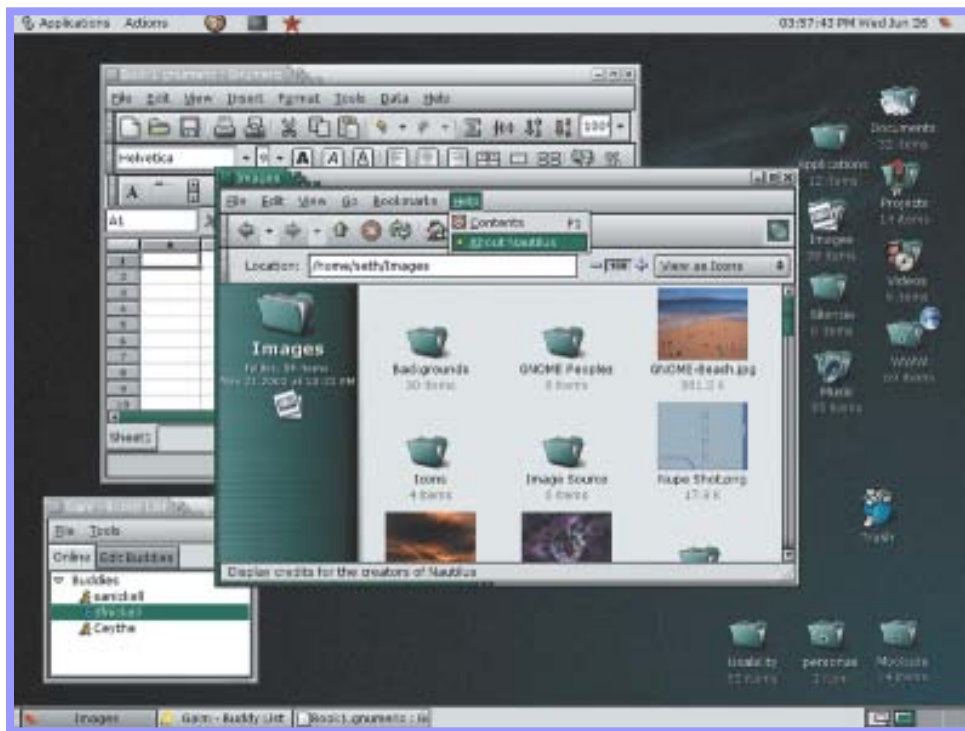


Microsoft Windows



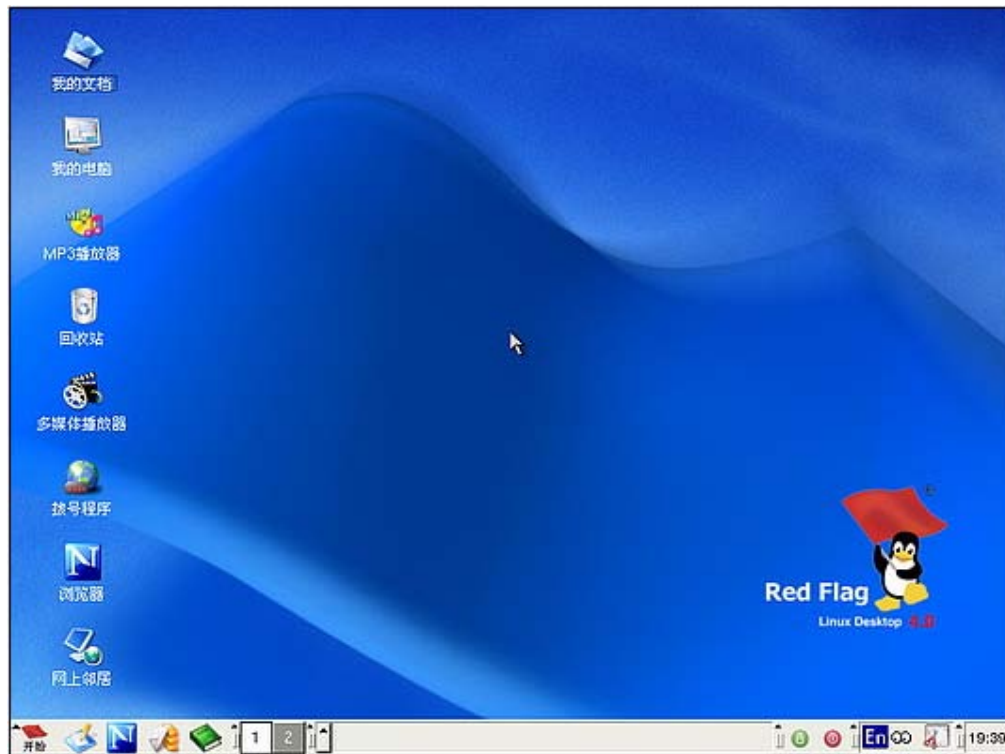
- 目前最流行的微机操作系统。
- 要求 CPU 是 Intel /AMD 微处理器。
- 具有简单、易学、易用的图形用户界面
(Graphical User Interface, GUI)。

Unix



- 原本主要是在大型机、小型机和工作站上运行，现在可在各种 CPU 、各种档次的计算机上运行。
- 运行稳定、可靠。
- 较强的安全性控制。
- 目前的大型网络服务器主要采用的操作系统。
- 标准化程度很高。
- GUI 和命令行界面。

Linux



- 免费使用和自由传播的类 Unix 操作系统
- 继承了 Unix 以网络为核心的设计思想，稳定、可靠、安全。
- 有很多发行版本，较流行的有：
 - ☐ RedHat Linux
 - ☐ Debian Linux
 - ☐ RedFlag Linux，第一个国产的操作系统版本

Apple MacOS



- 在 Apple公司的 Macintosh 计算机上运行。
- Apple公司拥有最先进的 GUI技术
- 要求 CPU 是 PowerPC 微处理器。
- 在影视、广告等行业被广泛使用。

Linux的前世今生

- Linux 内核最早由Linux Torvalds在1991年赫尔辛基大学上学时出于个人爱好编写的，仅有10000行代码

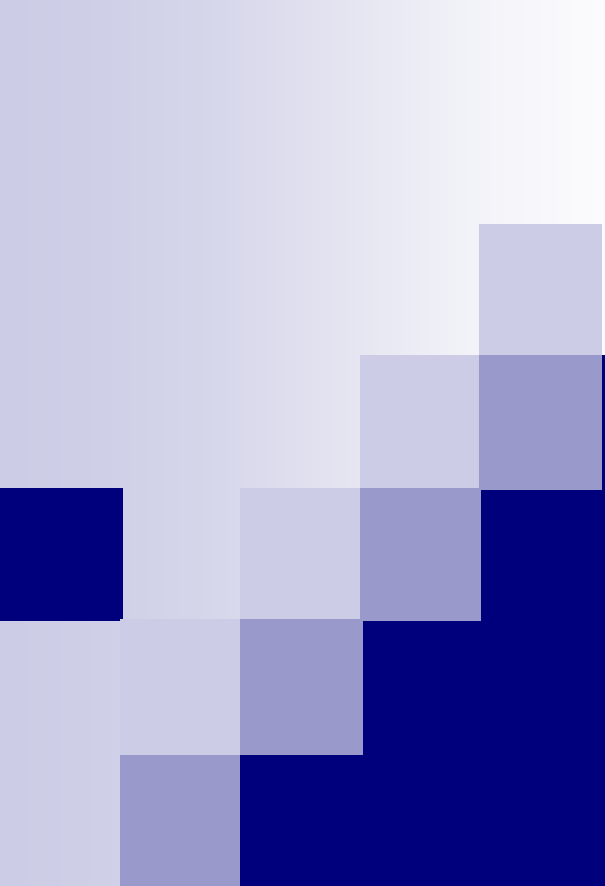
□ Hello everybody out there using minix-
I'm doing a (free) operation system
(just a hobby, won't be big and
professional like gnu) for 386(486) AT
clones.

- 开源与共享
- 超级电脑TOP 500中占有率达到87.8%
- 企鹅Tux



GNU/Linux
The Soft Revolution





程序设计

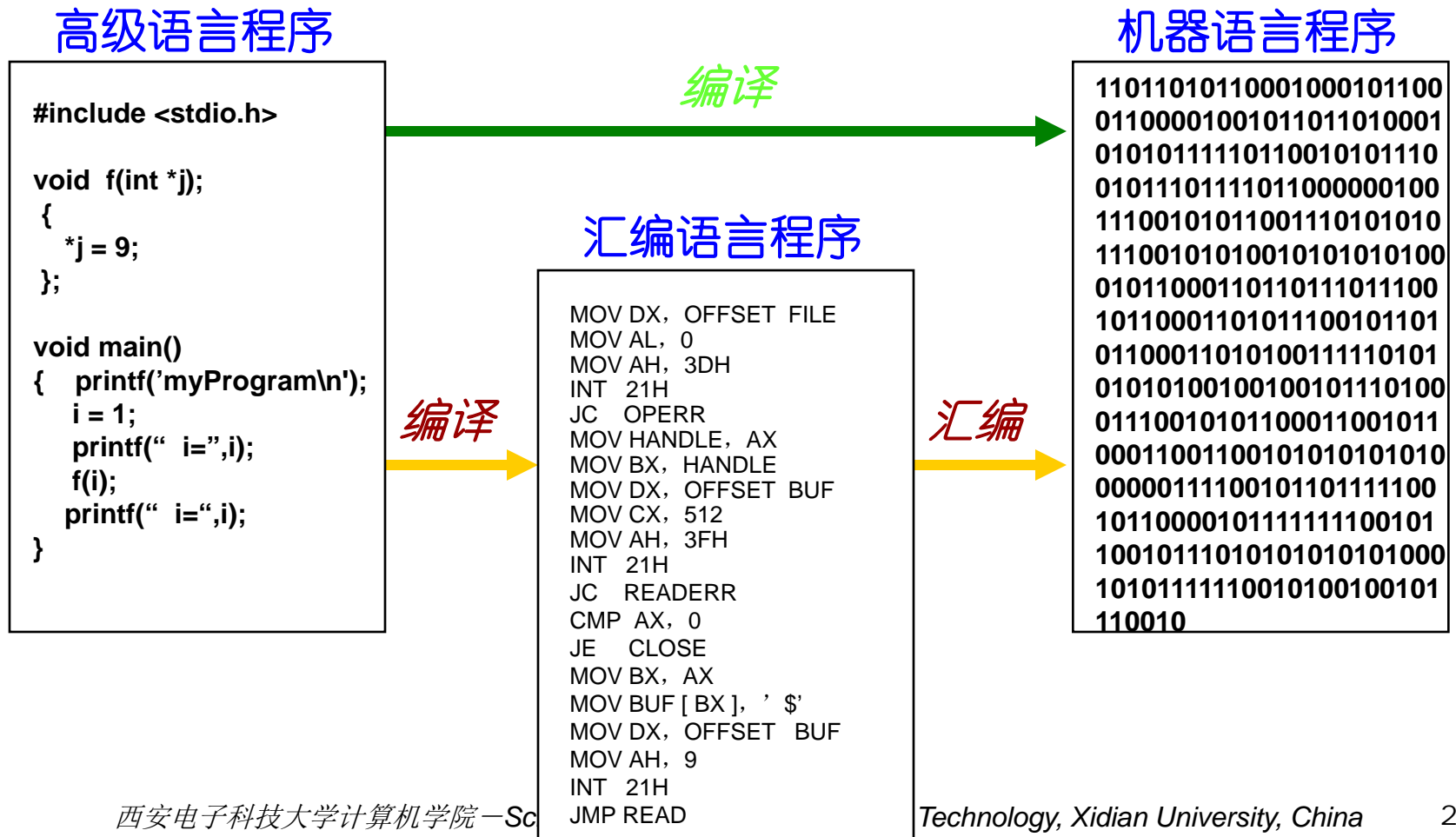
西安电子科技大学计算机学院—School
of Computer Science and Technology,
Xidian University, China

程序和程序设计语言

- 软件的核心内容是程序，而程序是指令的序列。
- 起初，程序员用计算机能够直接执行的二进制码（机器语言，Machine language）来写程序。工作效率非常低，而且容易出错、不容易查错。
- 随后的思路：先用英文字母和数字按照一定规则来写程序，再由另一个已经可以执行、具有翻译能力的程序把它翻译成等价的二进制表示，交给计算机执行。
- 因此，程序实际上是用特定语言表示的指令的序列。
- 这样的语言叫做程序设计语言（Programming languages）。
 - 汇编语言（Assemble language）
 - 高级语言（High-level programming language）

程序和程序设计语言

■ 对程序进行自动翻译:



语言翻译程序

- 将用高级语言或汇编语言编写的程序（源程序，Source programs），翻译成等价的机器语言程序。
- 翻译时要对源程序进行检查，来保证翻译后的程序是可执行的。
 - 语言翻译程序无法保证程序的执行结果是正确的，这要靠程序设计人员自己来判断。
 - 在正确地表达了一个错误的源程序时，语言翻译程序只能肯定源程序的表达方式（语法，Syntax）是正确的，而没有能力判断出程序的内涵（语义，Semantics）是错误的。

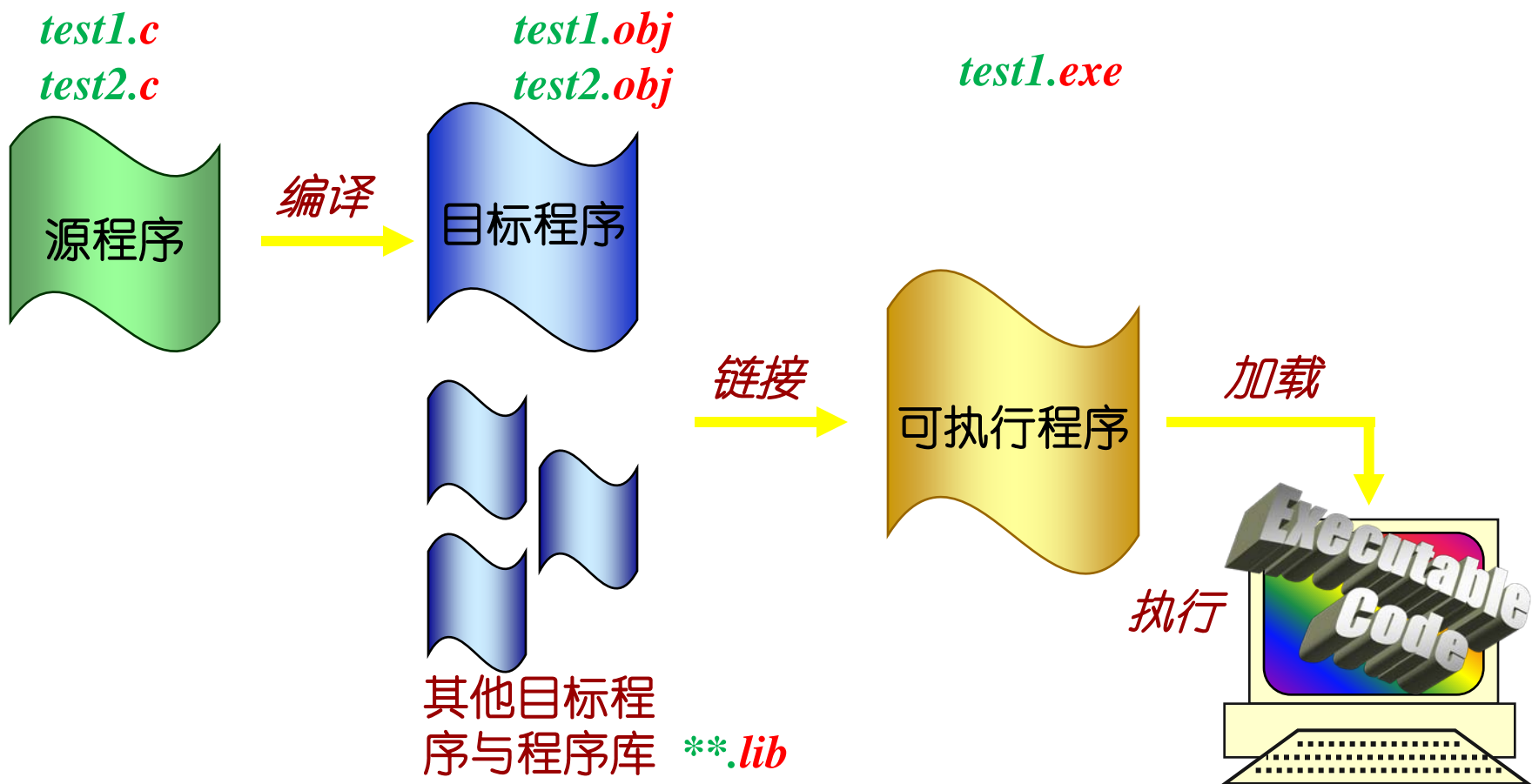
语言翻译程序

■ 类别:

- **编译程序** (Compiler) : 将**高级语言程序**翻译成对应的**汇编语言程序**, 或者直接翻译成对应的机器语言程序 (称为**目标程序** (Object program), 存储于外存, 不能直接执行) 。
- **汇编程序** (Assembler) : 将**汇编语言程序**翻译成对应的**目标程序**。
- **链接程序** (Linker) : 将一个或多个目标程序与一个或多个相关的程序库 (Library, 也是由一些目标程序构成) 组织在一起, 产生**可执行代码** (Executable code) 并存入外存, 在需要运行时再由操作系统加载后执行。
- **解释程序** (Interpreter) : 对高级语言程序**逐句地进行翻译**, 产生对应的机器语言指令序列并执行之。

语言翻译程序

■ 编译—链接—加载—执行的过程：



语言翻译程序

- 不同的高级语言有不同的编译程序或解释程序。目前主要的高级语言有：
 - ☐ C 、 C++、 C#
 - ☐ Java
 - ☐ Visual Basic
 - ☐ PASCAL、 Delphi
 - ☐ Ada
 - ☐ Fortran
 - ☐ COBOL
- 计算机的机器指令不同，汇编程序就不同。
- 链接程序通常与操作系统有关，可以链接同一操作系统中不同语言的编译程序或汇编程序所产生的目标程序。



数据组织

数据结构

数学模型有**定量模型**和**定性模型**两类，定量模型：可以用数值方程表示的一类模型；定性模型：是指非数值性的数据结构（如表、树和图等）及其运算。

在计算领域中，**数据结构 (Data Structure)** 指的是一类定性数学模型，它是计算机算法设计的基础，它在计算机科学中占有十分重要的地位。

瑞士著名计算机科学家尼可莱·沃思 (Nikiklaus Wirth) 在1976年曾提出这样一个公式：

算法+数据结构=程序

数据结构的基本概念

■ 数据结构的组成

数据结构是一类定性的数学模型，它由数据的**逻辑结构**、数据的**存储结构**（或称物理结构）及其**运算**3部分组成。

■ 数据**逻辑结构**的形式化定义

$DS = \langle D, R \rangle$

其中：

D表示数据的集合；

R表示数据D上关系的集合。

数据的存储结构

- 数据的存储结构是指在反映数据逻辑关系的原则下，数据在存储器中的存储方式。数据存储结构的基本组织方式有顺序存储结构和链式存储结构：
 - 顺序存储结构：借助元素在存储器中的相对位置来表示数据元素的逻辑关系。
 - 链式存储结构：借助指针来表示数据元素之间的逻辑关系，通常在数据元素上增加一个或多个指针类型的属性来实现这种表示方式。

数据结构的基本运算

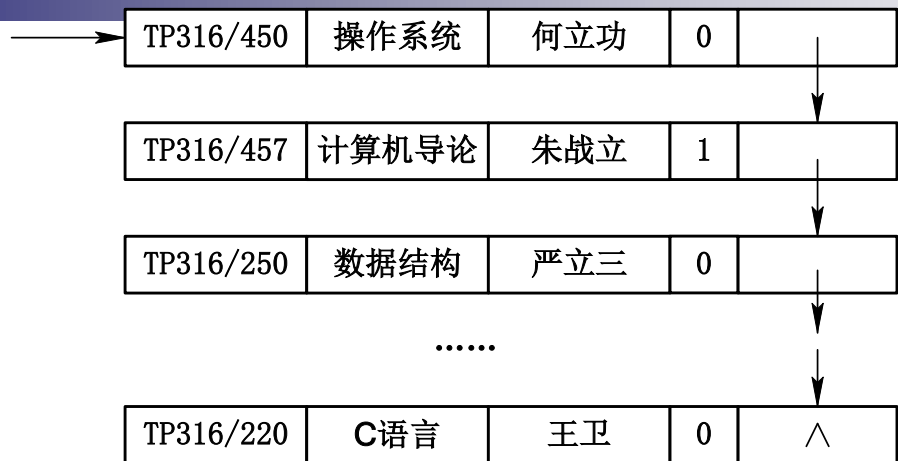
- 建立数据结构;
- 清除数据结构;
- 插入数据元素;
- 删除数据元素;
- 更新数据元素;
- 查找数据元素;
- 按序重新排列;
- 判定某个数据结构是否为空, 或是否已达到最大允许的容量;
- 统计数据元素的个数。

数据结构的基本概念（例）

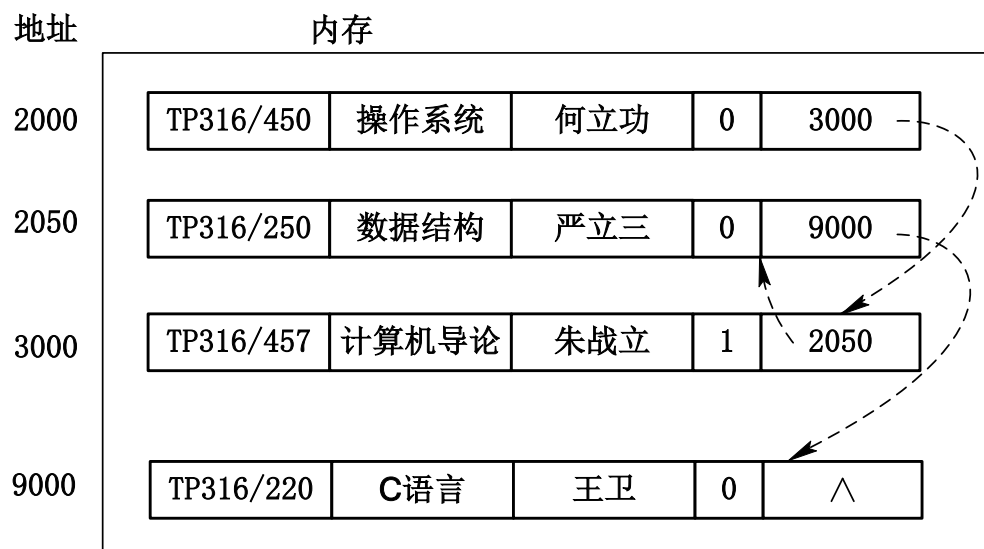
- 计算机进行数据输入和数据输出时使用的最小单位称为**数据元素**。在图书馆图书的计算机管理问题中，一本图书的描述数据由**书号**、**书名**、**作者名**、**状态**等组成，一本书的这些数据在输入或输出时必须作为一个整体进行，所以这个问题的数据元素就由书号、书名、作者名、状态等数据组成。
- 要用计算机来管理图书馆图书的借阅，需要分析和建立图书数据的模型。所有数据元素应组织成一种**一个数据元素后接另一个数据元素**的结构。任何一次借阅的计算机管理过程实际就是对其中某一个数据元素的某一个(或某几个)**数据项数据值的修改过程**。

TP316/450	操作系统	何立功	0
TP316/457	计算机导论	朱战立	1
TP316/250	数据结构	严立三	0
.....			
TP316/220	C语言	王卫	0

用数组存储图书信息表



(a)



(b)

(a) 链接方式的图书信息表 (b) 链表在内存中的存放示意

- 对于图书馆图书的计算机管理问题来说，图书信息表的**存储结构不同，算法的实现方法必然不同**。
- 例如，任何一次图书借阅管理过程首先需进行所借阅或归还图书的查找，若图书信息表采用数组来存储，为了提高查找算法的效率，可以先按图书的书号对该数组进行排序，然后可以按折半查找算法来设计图书的查找算法。
- 若图书信息表采用链式存储结构来存储，就不能用折半查找算法来设计查找算法，因为在链式存储结构下，每一个结点的地址都是当前运行计算机的操作系统临时分配的，因此，无法计算出链表中最中间结点在内存中的存储地址。

常用的几种数据结构

- 线性表
- 数组
- 树、二叉树
- 图
- 堆栈
- 队列
- ...



数据库

西安电子科技大学计算机学院—School
of Computer Science and Technology,
Xidian University, China

数据

- 信息系统五要素之一就是数据。
- 数据是关于人、地点、事物、事件等的事实。
- 计算机所存储、管理和处理的数据有着非常大的数量与相当多的种类。
- 在计算机中表示和存储数据，需要有许多种格式。
- 如何来看数据？
 - **物理视图** (Physical view)：看到的是数据的外在特征。如格式、位置等。
 - **逻辑视图** (Logical view)：看到的是数据的内在特征。如涵义、内容等。

数据的组织

- 对数据必须进行精心的组织。
- 数据的组织分为以下层次：

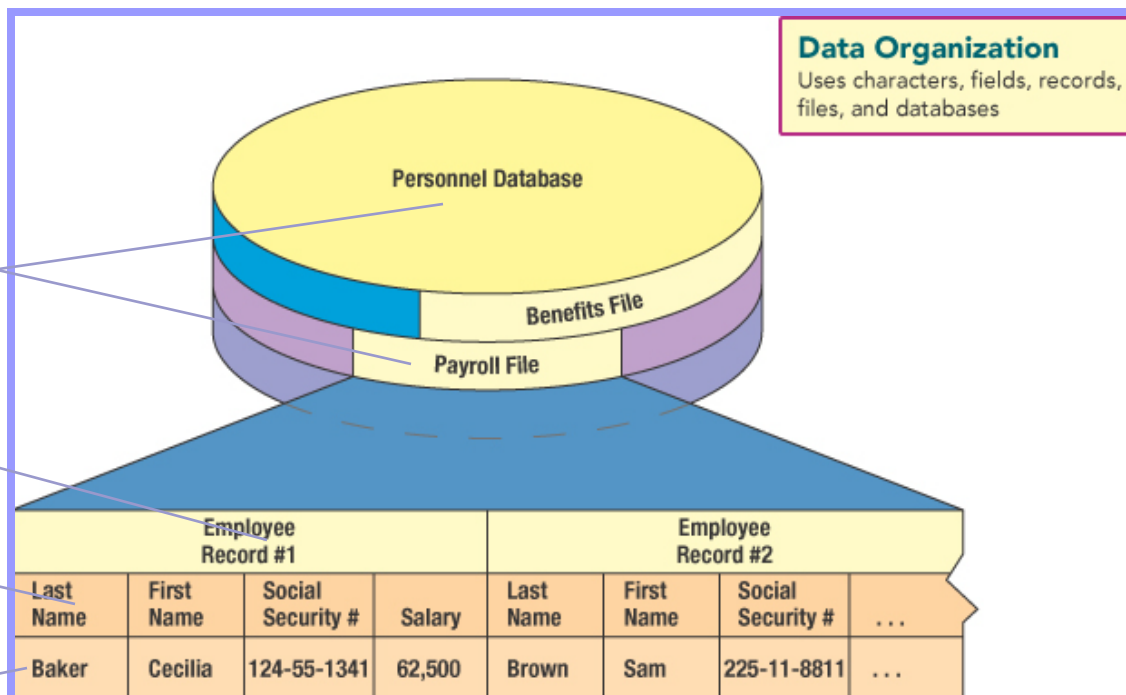
- ☐ 字符 (Characters)

- ☐ 域 (Fields)

- ☐ 记录 (Records)

- ☐ 文件

- ☐ 数据库
(Database)



为什么要用数据库来组织和管理数据？

- 数据是信息系统中最宝贵、最脆弱的财富：
 - 程序被丢失或破坏，还可以重新产生。
 - 数据是大量的最终用户日积月累产生的，如果被丢失或破坏，要付出很大代价才能恢复（有些甚至无法恢复）。
- 应用软件几乎都要涉及数据的组织和管理，在实现上要做大量的工作。
- 文件系统很难有效的管理、组织数据。
- 数据库统一承担了数据的组织和管理工作，其优点是：
 - 与具体的应用程序无关，因而减少了应用程序的开发工作量，并使得多个应用程序共享所需要的或要处理的数据。
 - 容易保证数据的安全性。
 - 减少了数据的冗余（Redundancy），节省了存储空间。
 - 保证了数据的完整性（Integrity）。

用什么来管理数据库？

- 有专门的软件：**数据库管理系统**（Database Management System, DBMS）。其职责是：为产生、修改、访问数据库提供有效的支持。
- DBMS 一般包含以下这些子系统（Subsystems）：
 - ☐ DBMS 引擎（DBMS engine）
 - ☐ 数据定义子系统（Data definition subsystem）
 - ☐ 数据操纵子系统（Data manipulation subsystem）
 - ☐ 应用程序生成子系统（Application generation subsystem）
 - ☐ 数据管理子系统（Data administration subsystem）

数据模型

- 数据模型（Data models）是数据库管理和组织数据时所采用的结构模型。
- 已有的数据模型：
 - 关系（Relational）模型
 - 层次（Hierarchical）模型
 - 网状（Network）模型
 - 多维（Multidimensional）模型
 - 面向对象（Object-oriented）模型
- 目前的 DBMS 主要采用的是关系模型，但已经开始向多维模型（亦称为对象—关系模型）演变。

数据库语言与产品

- 目前的数据库语言基本上都服从 SQL (Structured Query Language) 标准。SQL 语言可以完成数据定义、数据操纵 (包括数据的查询 (Query)、插入 (Insert)、删除 (Delete)、修改 (Update) 等操作) 和数据管理等功能。
- 目前常见的 DBMS:
 - ☐ Oracle
 - ☐ SQL Server
 - ☐ UDB / DB2
 - ☐ MySQL
 - ☐ Sybase
 - ☐ Informix

课后作业

- 1、熟悉有哪些主流的操作系统。
- 2、了解C语言的编译过程。
- 3、了解目前主流的数据库有哪些。